2 Mathematical Prerequisites

2.1 INTRODUCTION

This book is aimed at practitioners and students in transport modelling and planning. Some of these may have a sound mathematical background; they may skip this chapter without loss of continuity. Other readers may have a weaker mathematical background or may simply welcome the opportunity to refresh ideas and notation. This chapter is addressed to these readers. It aims only to outline the most important mathematical prerequisites needed to benefit from this book.

Most of the mathematical prerequisites, however, are not that demanding: the reader can get by with little more than school algebra and some calculus. We introduce first the idea of functions and some specialised notation together with the idea of plotting functions in Cartesian (orthogonal) coordinates. After introducing the concept of series we treat the very important topic of matrix algebra; this is particularly important in transport as we often deal with trip and other matrices. Elements of calculus come next, including differentiation and integration. Logarithmic and exponential functions deserve some special attention as we will find them often in transport models. Finding maxima and minima of functions plays an important role in model development and the generation of solution algorithms. Finally, a few statistical concepts are introduced in the last section of this chapter. Statistics play a key part in contemporary transport modelling techniques and this section provides only an elementary entry point to the subject. A few other statistical concepts and techniques will be introduced in subsequent chapters as needed.

There are several books available as reference works for the more informed reader and as first textbooks for readers needing to brush up their mathematical background. These include those by Morley (1972), Stone (1966), and Wilson and Kirby (1980). We see the future of transport modelling practice moving steadily away from experience through shortcuts and ‘fudge factors’, and increasingly adopting models with sounder theoretical backing. This trend results from the need to provide consistent advice to decision makers; this advice should not depend on an arbitrarily chosen number of iterations or starting point, nor on models likely to produce pathological results when used to forecast completely new options. This increased rigour will rely on better mathematical and statistical representations of problems and therefore requires further reading in these areas.
2.2 ALGEBRA AND FUNCTIONS

2.2.1 Introduction

Elementary algebra consists of forming expressions using the four basic operations of ordinary mathematics on letters which stand for numbers. It is useful to distinguish between variables (generally denoted by letters such as $x, y$ and $z$), which represent measured quantities, and constants or parameters (generally denoted by letters such as $a, b, c, \ldots, k, m, n, \ldots$), or by letters from the Greek alphabet. The value of a constant is supposed to remain invariant for the particular situation examined.

Variables, and constants, are related through equations such as:

$$y = a + bx$$

(2.1)

and if we were interested in $x$, we could 'solve' (2.1) for $x$, obtaining:

$$x = \frac{(y - a)}{b}$$

(2.2)

The variables $x$ and $y$ in (2.1) and (2.2) are related by the '=' sign. However, in algebra we may also have inequalities of the following four types:

- which means less than
- which means less than or equal to
- which means greater than, and
- which means greater than or equal to

and which are used to constrain variables, for example:

$$x + 2y \leq 5$$

(2.3)

This expression, unlike an equation, cannot be 'solved' for $x$ or $y$, but note that both variables can take only a restricted range of values. For example, if we restrict them further to be positive integers, it can easily be seen that $x$ cannot be greater than 3 and $y$ cannot be greater than 2.

It is possible to manipulate inequalities in much the same way as equations, thus:

- we can add or subtract the same quantity to/from each side;
- we can also multiply or divide each side by the same quantity, but if the number which is being multiplied or divided is negative, the inequality is reversed.

For example, if we subtract 5 on both sides of (2.3) we get

$$x + 2y - 5 \leq 0$$

which is certainly the same constraint. However, if we multiply it by $-2$ we obtain:

$$-2x - 4y \geq -10$$

which can be checked by the reader to provide the same constraint as (2.3).

The use of different letters to denote each variable is only convenient up to a certain point. Soon it becomes necessary to use indices (e.g. subscripts or superscripts) to define additional variables, as in $x_1, x_2, x_3, \ldots, x_n$, which we can conveniently summarise as $x_i, i = 1, 2, \ldots, n$; it does not matter if we use another letter for the index if it has the same numerical range. For example, we could have defined also $x_k, k = 1, 2, \ldots, n$.

The use of indices facilitates a very convenient notation for summations and products:

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + x_3 + \ldots + x_n$$

(2.4)

or

$$\prod_{j=1}^{m} y_j = y_1y_2y_3 \ldots y_m$$

(2.5)

In certain cases a single index is not enough and two or more may be used. For example we could define the following six variables, $T_1, T_2, T_{21}, T_{22}, T_{31}, T_{32}$ as $T_{ij}, i = 1, 2, 3$, and $j = 1, 2$. With two-subscript variables we can have double summations or double products, as in:

$$\sum_{i=1}^{3} \sum_{j=1}^{3} T_{ij} = \sum_{i=1}^{3} (T_{i1} + T_{i2}) = T_{11} + T_{12} + T_{21} + T_{22} + T_{31} + T_{32}$$

(2.6)

2.2.2 Functions and Graphs

We have already referred to variables as being related by equations and inequalities; in general these can be called functional relations. A particular function is some specific kind of relationship between two or more variables. For example, the power function:

$$y = \phi x^n$$

(2.7)

yields values of the dependent variable $y$, given values of the parameters $\phi$ and $n$, and of the independent variable $x$; a function requires that for every value of $x$ in some range, a corresponding value of $y$ is specified. Often we do not wish to refer to a particular function, but only to state that $y$ is 'some function of $x$' or vice versa; this can be written as:

$$y = f(x)$$

(2.8)

A large range of functions exists and readers should familiarise themselves with these as they arise. It is usually convenient to plot functions graphically on a Cartesian coordinate system (see Figure 2.1).
A dependent variable may be a function of several independent variables, for example:

\[ y = f(x_1, x_2, \ldots, x_n) \]  

(2.9)

but this would require \( n + 1 \) dimensions to represent it (\( n \) for the independent variables and 1 for \( y \)). Cartesian coordinates can be used in three or more dimensions, in the case of three dimensions the orientation of the third axis is out of this side of the paper in Figure 2.1. More than three dimensions cannot be easily visualised physically but are dealt with algebraically in just the same way as one, two and three dimensions. For example, in the case of \( n = 2 \) the function can be represented by a surface over the relevant part of the \((x_1, x_2)\) plane.

Generally, any equation for an unknown quantity \( x \) can be put in the form \( f(x) = 0 \); for example, the linear equation:

\[ ax = b \]

is equivalent to

\[ ax - b = 0 \]

where \( f(x) = ax - b \). Solving the equation is therefore equivalent to finding the points on the curve \( y = f(x) \) which meet the \( x \) axis. These points are called real solutions or zeros of \( f(x) \); for example, \( x_1 \) and \( x_2 \) in Figure 2.2.

We are sometimes interested in what happens to the value of a function \( f(x) \), as \( x \) increases indefinitely \( (x \to \infty) \); it can easily be seen that the possibilities are only the following:

- tend to infinity (e.g. when \( f(x) = x^2 \))
- tend to minus infinity (e.g. when \( f(x) = -x \))
- oscillate infinitely (e.g. when \( f(x) = (-1)^x \sqrt{x^2} \))
- tend to a finite limit (e.g. \( f(x) = 1 + 1/x \)).

For more complex functions some ingenuity may be required to find out if they tend to a finite limit when \( x \to \infty \).

We may also be interested in finding the limit when \( x \) approaches a finite value. For example, if \( f(x) = 1/(x + 3) \), it can easily be seen that the limit when \( x \to 0 \) is \( 1/3 \). If for some value \( a \) we have that \( f(x) \to \infty \) as \( x \to a \), the curve \( y = f(x) \) is said to have an asymptote \( x = a \) (see Figure 2.3).

One of the most important functions is the straight line, shown in Figure 2.4 and whose general equation is (2.1). It can easily be seen that \( b \) is the value of \( y \) when \( x = 0 \); this is usually called the intercept on the \( y \) axis. The constant \( a \) is called the gradient and it can be shown to be given by:
Solving for \( x \) in (2.11b) and replacing this value (i.e. \(-y + 4\)) in (2.11a) we get that the solution is point \( P \) with coordinates \( (x = 1, y = 3) \).

### 2.2.3 Sums of Series

A series is simply defined as a sequence of numbers \( u_n, n = 1, 2, \ldots, N \). In many cases it may be interesting to find its sum given by:

\[
S_N = u_1 + u_2 + \ldots + u_N = \sum_n u_n
\]  

(2.12)

In some cases, such as the arithmetic progression given by:

\[
u_n = u_{n-1} + d
\]

(2.13)

it can be shown that the series has a sum to \( N \) terms. For example, if the first term is \( b \) the sum can be shown to be:

\[
S_N = N b + N(N - 1)d / 2
\]  

(2.14)

The geometric progression (2.15), formed by multiplying successive terms by a constant factor \( r \), also has an expression for the sum of \( N \) terms. If \( b \) is again the first term, the sum can be shown to be given by (2.16)

\[
u_n = r u_{n-1}
\]  

(2.15)

\[
S_N = b (1 - r^n) / (1 - r)
\]  

(2.16)

In other cases the series may have a simple expression for its sum, such as in:

\[
u_n = n, \text{ where the sum is given by } S_N = N(N + 1)/2.
\]

or \( u_n = x^n \), where it is given by \( S_N = x (1 - x^n) / (1 - x) \) for \( x \) different from 1; but still diverge (i.e. \( S_N \) keeps increasing indefinitely when \( N \) tends to infinity). That happens to \( u_n = n \); it also happens to \( u_n = x^n \) if \( x > 1 \) above; however, the latter converges to \( S_N = x / (1 - x) \) for the range \( 0 < x < 1 \).

### 2.3 MATRIX ALGEBRA

#### 2.3.1 Introduction

Any variable with two subscripts can be called a matrix. We will denote matrices by the notation \( B = \{b_{ij}\} \), where the variables \( b_{ij}, \ i = 1, 2, \ldots, N; \ j = 1, 2, \ldots, M \) are the elements of \( B \). This can be written as follows:
\[
\mathbf{B} = \begin{pmatrix}
B_{11} & B_{12} & B_{13} & \cdots & B_{1M} \\
B_{21} & B_{22} & B_{23} & \cdots & B_{2M} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
B_{N1} & B_{N2} & B_{N3} & \cdots & B_{NM}
\end{pmatrix}
\]

(2.17)

As can be seen, the matrix has \(N\) rows and \(M\) columns; for this reason it is known as a \(N \times M\) matrix. A \textit{vector} is an important special case, being a one-dimensional array or a \(N \times 1\) matrix. In these cases the second index is redundant, so we write:

\[
\mathbf{V} = \begin{pmatrix}
V_1 \\
V_2 \\
V_3 \\
\vdots \\
V_N
\end{pmatrix} = \begin{pmatrix}
V_1 \\
V_2 \\
\vdots \\
V_N
\end{pmatrix}
\]

(2.18)

Formally, a non-indexed variable, or even a constant, can be thought of as a \(1 \times 1\) matrix and it is known as a \textit{scalar}.

If we interchange rows and columns we obtain an \(M \times N\) matrix known as the \textit{transpose} \(\mathbf{B}^T\) of \(\mathbf{B}\), which is given by:

\[
\mathbf{B}^T = \begin{pmatrix}
B_{11} & B_{12} & B_{13} & \cdots & B_{1N} \\
B_{21} & B_{22} & B_{23} & \cdots & B_{2N} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
B_{M1} & B_{M2} & B_{M3} & \cdots & B_{MN}
\end{pmatrix}
\]

(2.19)

Similarly, the transpose of an \(N \times 1\) vector (also known as a \textit{column} vector) is a \textit{row} vector:

\[
\mathbf{V}^T = [V_1 \ V_2 \ V_3 \ \cdots \ V_N]
\]

(2.20)

A \textit{square} matrix \(\mathbf{S}\) is one where \(N = M\); a square matrix such that \(\mathbf{S} = \mathbf{S}^T\) is called \textit{symmetric}. A \textit{diagonal} matrix \(\mathbf{D} = \{D_{ij}\}\) is one where \(D_{ij} = 0\) unless \(i = j\). The \textit{unit} matrix is a square diagonal matrix with each diagonal element equal to 1, that is:

\[
\mathbf{I} = \begin{pmatrix}
1 & 0 & 0 & \cdots & 0 \\
0 & 1 & 0 & \cdots & 0 \\
0 & 0 & 1 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & 1
\end{pmatrix}
\]

(2.21)

2.3.2 Basic Operations of Matrix Algebra

We will define the operations between two matrices \(\mathbf{A}\) and \(\mathbf{B}\) by setting a new matrix \(\mathbf{C}\) which will represent the combination required. First matrix \textit{addition}:

\[
\mathbf{C} = \mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}
\]

(2.22)

is defined by \(C_{ij} = A_{ij} + B_{ij}\) and requires that both matrices being combined are of the same size, say both \(N \times M\) matrices; then \(\mathbf{C}\) is also an \(N \times M\) matrix. This is also a requirement for matrix subtraction:

\[
\mathbf{C} = \mathbf{A} - \mathbf{B}
\]

(2.23)

similarly defined as \(C_{ij} = A_{ij} - B_{ij}\). An operation which is unique to matrix algebra is \textit{multiplication} by a \textit{scalar}:

\[
\mathbf{C} = k\mathbf{A}
\]

(2.24)

defined by \(C_{ij} = kA_{ij}\), where obviously the new 'grossed up' matrix has the same size as the old one.

Matrix \textit{multiplication} is more complex, as:

\[
\mathbf{C} = \mathbf{AB}
\]

(2.25)

is defined by \(C_{ij} = \sum A_{ik} B_{kj}\), where \(\mathbf{A}\) is an \(N \times M\) matrix and \(\mathbf{B}\) is any \(M \times L\) matrix (i.e. the number of columns in \(\mathbf{A}\) must equal the number of rows in \(\mathbf{B}\) but there are no other restrictions). In this case \(\mathbf{C}\) is an \(N \times L\) matrix.

It is easy to see that in general \(\mathbf{AB}\) is not equal to \(\mathbf{BA}\), i.e. the operation is non-commutative, as opposed to elementary algebra. However, this is not the case with the unit matrix \(\mathbf{I}\); in fact, it can easily be checked that:

\[
\mathbf{IA} = \mathbf{AI} = \mathbf{A}
\]

(2.26)

Thus, although it is possible to define the product of any number of matrices, order must always be preserved. In fact we refer to \textit{pre-multiplication} of \(\mathbf{A}\) by \(\mathbf{B}\) to form the product \(\mathbf{BA}\), and to \textit{post-multiplication} to form \(\mathbf{AB}\).

To define \textit{division} it is convenient to use the concept of inverse of a matrix. Unfortunately this only exists for square matrices and then not always. If the inverse exists, it is denoted as \(\mathbf{B}^{-1}\) and is the matrix that satisfies:

\[
\mathbf{B}^{-1}\mathbf{B} = \mathbf{BB}^{-1} = \mathbf{I}
\]

(2.27)

In this case \(\mathbf{B}\) is said to be \textit{non-singular}. We will not give a procedure for the calculation of the elements of the inverse matrix as it is fairly complicated. It is sufficient to know that under suitable conditions it exists. Division is then just \textit{pre} or \textit{post-multiplication} by \(\mathbf{B}^{-1}\).

In this book matrices and vectors are mostly used to provide a shorthand notation for such things as sets of simultaneous equations and for obtaining their solution in terms of the inverse matrix.
2.4 ELEMENTS OF CALCULUS

The two main branches of calculus are differentiation and integration; their basic nature can be intuitively identified by reference to the function \( y = f(x) \) depicted in Figure 2.6. Consider the points P and Q and the straight line (chord) connecting them. Differentiation is concerned with the calculation of the gradient of a curve at a point. To do this, it is useful to consider Q approaching P; in the limit the chord PQ becomes the tangent to the curve at P = Q (i.e. when their horizontal 'distance' \( h \) is 0) and by definition its gradient is equal to that of the curve.

Integration, on the other hand, is concerned with calculating the area under a curve. Say the shaded area in Figure 2.6; as we will see below, these two operations are closely related.

2.4.1 Differentiation

Using (2.10) the gradient of the chord PQ in Figure 2.6 can be written as:

\[
\delta(x) = \frac{[f(x_0 + h) - f(x_0)]}{h}
\]

If the limit of \( \delta(x) \) when \( h \to 0 \) exists and is the same whether \( h \) tends to zero from above or below, it is called the derivative of \( y \), or of \( f(x) \), with respect to \( x \) at \( x_0 \) and it is often written as \( f'(x_0) \) or \( \frac{dy}{dx}|_{x_0} \). The process of finding the derivative is called differentiation.

If \( f(x) \) is given as an expression in \( x \), it is usually not difficult to find \( f'(x) \) as a function of \( x \) using the results in Table 2.1, plus others we will give below.

Since derivatives are themselves functions of \( x \), we can also define second- and higher-order derivatives (i.e. \( f''(x) \) or \( \frac{d^2y}{dx^2} \) and so on). For example, if we differentiate the first derivative of \( y = x^b \) in Table 2.1, we get:

\[
\frac{d^2y}{dx^2} = b(b-1)x^{b-2}
\]  (2.28)

![Figure 2.6 Gradient at a point and area under a curve](image)

Table 2.1 Common derivatives

<table>
<thead>
<tr>
<th>Function ( f(x) )</th>
<th>Derivative ( f'(x) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( k )</td>
<td>( 0 )</td>
</tr>
<tr>
<td>( ax )</td>
<td>( a )</td>
</tr>
<tr>
<td>( x^n )</td>
<td>( nx^{n-1} )</td>
</tr>
<tr>
<td>( ax^n )</td>
<td>( anx^{n-1} )</td>
</tr>
<tr>
<td>( \sin x )</td>
<td>( \cos x )</td>
</tr>
<tr>
<td>( \cos x )</td>
<td>( -\sin x )</td>
</tr>
<tr>
<td>( \tan x )</td>
<td>( \sec^2 x )</td>
</tr>
<tr>
<td>( \frac{1}{x} )</td>
<td>( -\frac{1}{x^2} )</td>
</tr>
<tr>
<td>( \sqrt{x} )</td>
<td>( \frac{1}{2x^{1/2}} )</td>
</tr>
<tr>
<td>( e^x )</td>
<td>( e^x )</td>
</tr>
<tr>
<td>( \log x )</td>
<td>( \frac{1}{x} )</td>
</tr>
</tbody>
</table>

2.4.2 Integration

This is the reverse of differentiation; if we know the gradient of some curve at every point then the equation of the curve itself is known as the integral. For example, if \( g = g(x) \) is the gradient, the equation of the curve is written

\[
y = \int g(x) \, dx
\]

and this result is always arbitrary up to an additive constant; for example, if \( g = bx^{b-1} \) we know from Table 2.1 that the indefinite integral of \( g(x) \) is given by:

\[
y = G(x) = \int x \, dx = \frac{1}{2}x^2 + C
\]

where \( C \) is an arbitrary constant of integration (i.e. the derivative of \( \frac{1}{2}x^2 + C \) is \( bx^{b-1} \) no matter the value of \( C \)). The most practical elementary use of integration is to obtain the area under a curve as the definite integral, as shown in Figure 2.7.

\[
\text{Area abed} = [F(x)]_a^b = F(b) - F(a) = \int_a^b f(x) \, dx
\]

![Figure 2.7 Areas under a curve: (a) general case, (b) line parallel to x axis](image)
For example, if we take the simple case of a straight line parallel to the $x$ axis at height $h$ and want to integrate between the values $a$ and $b$ (see Figure 2.7b), we get:

$$y = f(x) = h$$

and

$$F(x) = hx + C$$

then

$$\text{Area} = F(b) - F(a) = h(b - a)$$

which is indeed the area of the shaded rectangle in the figure.

Table 2.1 can be used in reverse to help to find indefinite integrals. In particular, if

$$\int u(x)dx = U(x) + C_1 \quad \text{and} \quad \int v(x)dx = V(x) + C_2$$

then

$$\int u(x)v(x)dx = U(x)V(x) + C_3$$

and

$$\int U(x)V(x)dx = U(x)V(x) - \int u(x)V(x)dx$$

Of course not all functions, even some that are deceptively simple in appearance, have indefinite integrals which are similarly simple expressions. However, for those that do not it is still possible to evaluate definite integrals numerically.

2.4.3 The Logarithmic and Exponential Functions

Among the functions we have considered so far, the simplest one with no indefinite integral is the inverse function $f(x) = 1/x$, depicted in Figure 2.8.

The integral of this function has been defined as the natural logarithm of $x$, or $\log_e(x)$, where $e$ is Neper's constant. Its value at approximately 2.7183 corresponds to the point on the $x$ axis of Figure 2.8 such that the shaded area is 1, i.e., $\log_e(e) = 1$. As in this book we will only use natural logarithms, we will drop the base $e$ from our notation.

In common with any other logarithm, $\log(x)$ has the following properties:

$$\log(1) = 0;$$
$$\text{As } t \to \infty, \log(t) \to \infty;$$
$$\text{As } t \to 0, \log(t) \to -\infty;$$
$$\log(uv) = \log(u) + \log(v).$$

Another useful related function is the exponential function $\exp(x)$ or $e^x$ for short, defined as the number $w$ such that $\log(w) = x$. Then, as expected of a power function, we have:

$$\exp(x+y) = e^xe^y,$$

moreover,

$$\exp(\log(x)) = x.$$

Both functions $\log(x)$ and $\exp(x)$ are easy to differentiate; by definition:

$$\frac{d}{dx} \log(x) = \frac{1}{x} \quad (2.31)$$

and it is not difficult to show that:

$$\frac{d}{dx} (e^x) = e^x \quad (2.32)$$

Thus the exponential is the function which remains unaltered under differentiation.

2.4.4 Finding Maximum and Minimum Values of Functions

This is one important use of differentiation. Consider Figure 2.9 for example; the function shown has a maximum at $x_1$ and a minimum at $x_2$. Both are characterized by the gradient of the curve being zero at those points, so the first step in finding them is to solve the equation $f'(x) = 0$. 
introduce the complementary notions of convexity and concavity. These result not only in a more powerful (although more restrictive) theory, but also provide an interesting geometric interpretation of the second-order conditions (2.33) to (2.35).

Figure 2.11 presents some examples of convex and non-convex functions. Geometrically, a function is convex if the line joining two points on its graph lies nowhere below the graph, as shown in Figure 2.11a; in two dimensions, a convex function would have a bowl-shaped graph. Similarly and simply, a function $y$ is said to be concave if the function $f = -g$ is convex. A nice property of convex functions is that the sum of two such functions is also convex.

\[ f'''(x) < 0 \]  
(2.33)

For a minimum we need:

\[ f''(x) > 0 \]  
(2.34)

and for a point of inflexion,

\[ f''''(x) = 0 \]  
(2.35)

These cases are illustrated in Figure 2.10, which suggests a good mnemonic. Consider the function as a cup of water; if it is facing downwards as in the case of the maximum, the liquid will drop (i.e. a minus sign). Conversely if its is facing upwards (e.g. a minus sign).

In order to develop a theory directed toward characterising global, rather than local, minimum (or maximum) points mathematicians have found it necessary to

\[ y = f(x_1, x_2, \ldots, x_n) \]  
(2.36)

Then the derivative of $y$ with respect to one of these variables may be calculated assuming the other variables remain constant during the operation. This is known as a partial derivative and is written $\frac{\partial y}{\partial x_i}$. For example, if:

\[ y = 2x_1 + x_2 x_3 \]

then

\[ \frac{\partial y}{\partial x_1} = 2 \]
\[ \frac{\partial y}{\partial x_2} = 3x_2 x_3 \]
\[ \frac{\partial y}{\partial x_3} = x_2 \]
It can be shown that maxima and minima of a function such as (2.36) can be found by setting all the partial derivatives to zero:

$$\frac{\partial y}{\partial x_i} = 0, \quad i = 1, 2, \ldots, n$$  \hspace{1cm} (2.37)

which gives a set of simultaneous equations to solve. A particularly interesting case is that of the restricted maximum or minimum. Assume we wish to maximise (2.36) subject to the following restrictions:

$$r_1(x_1, x_2, \ldots, x_n) = b_1$$
$$r_2(x_1, x_2, \ldots, x_n) = b_2$$
$$\vdots$$
$$r_K(x_1, x_2, \ldots, x_n) = b_K$$  \hspace{1cm} (2.38)

This can be done by defining Lagrangian multipliers $\lambda_1, \lambda_2, \ldots, \lambda_K$ for each of the equations (2.38) in turn, and maximising

$$L = f(x_1, x_2, \ldots, x_n) + \sum K \lambda_k [r_k(x_1, x_2, \ldots, x_n) - b_k]$$  \hspace{1cm} (2.39)

as a function of $x_1, x_2, \ldots, x_n$ and $\lambda_1, \lambda_2, \ldots, \lambda_K$. Thus, we solve:

$$\frac{\partial L}{\partial x_i} = 0, \quad i = 1, 2, \ldots, n$$  \hspace{1cm} (2.40)

and

$$\frac{\partial L}{\partial \lambda_k} = 0, \quad k = 1, 2, \ldots, K$$  \hspace{1cm} (2.41)

The equations (2.41) are simply the restrictions (2.38) in another form; the device of introducing the multipliers as additional variables enables the restricted maximum to be found.

2.4.6 Multiple Integration

In the case of integration, multiple integrals can be defined. For example, given (2.36) we might have:

$$V = \int \int \ldots \int f(x_1, x_2, \ldots, x_n) \, dx_1 \, dx_2 \ldots dx_n$$  \hspace{1cm} (2.42)

with $n$ integral signs. In order to get an intuitive feeling of its meaning it is useful to consider the two-dimensional case. The function

$$S = f(x_1, x_2)$$  \hspace{1cm} (2.43)

may be considered as defining a surface in a three-dimensional Cartesian system. Therefore,

$$V = \int \int f(x_1, x_2) \, dx_1 \, dx_2$$  \hspace{1cm} (2.44)

measures a volume under this surface, in a similar way to the single variable measuring an area under a curve.

2.4.7 Elasticities

The elasticity of a dependent variable $y$ with respect to another variable $x_i$ in a function such as (2.9), is given by the expression:

$$E(y, x_i) = \frac{\frac{\partial y}{\partial x_i}}{\frac{x_i}{y}}$$  \hspace{1cm} (2.45)

and can be interpreted as the percentage change in the dependent variable with respect to a given percentage change in the relevant independent variable.

In econometrics we will be often interested in the elasticities of a given demand function with respect to changes in the values of some explanatory variables or attributes. We will generally distinguish between direct- and cross-elasticities, the first relate to attributes of the service or good under consideration and the second to attributes of competing options or goods. For example, it is often stated that the elasticity of public transport demand to fares is around $-0.33$; this means that if we increase fares by 1% we should expect patronage to decrease by approximately 0.33%.

2.4.8 Series Expansions

It is sometimes necessary to estimate the values of a function $f(x)$ in the neighbourhood of a particular value $x_0$ of $x$, in terms of the values of the function and its derivatives at this value. For suitable functions this can be done by means of Taylor's series expansion; first we require to define the concept of a factorial number $(n!)$ which applies to non-negative integers:

$$n! = n(n-1)(n-2) \ldots 3 \cdot 2 \cdot 1$$  \hspace{1cm} (2.46)

$$0! = 1$$

A Taylor's series expansion is defined as:

$$f(x_0 + h) = f(x_0) + hf'(x_0) + \frac{h^2}{2!}f''(x_0) + \frac{h^3}{3!}f'''(x_0) + \ldots$$  \hspace{1cm} (2.47)

and it is most useful when $h$ is small enough for the higher-order terms to become rapidly smaller so that a good approximation is obtained by stopping the summation after just a few terms—even just after two terms.
as \( n_1 + n_2 + \ldots = n \). An alternative view of the expected probability of the result can be expressed in terms of a fair bet. If a person regards a fair a bet in which they win $35 if \( e_i \) happens and loses $x if it does not, then their estimate of \( p_i \) is \( x/(x + 35) \). This is so because they have solved the following equation which makes their expected gains or losses equal to zero, i.e. a fair bet:

\[
35p_i - x(1 - p_i) = 0
\]

On many occasions the probabilities of certain experiments are not simple to calculate. It is convenient to define an event \( (E) \) as a subset of the set of possible results of an experiment, \( E = \{e_1, \ldots, e_n\} \). The probability of an event is the sum of the probabilities of the results it is composed of,

\[
P(E) = \sum_i p_i, \quad e_i \in E
\]

**Example 2.1:** The event \( E \): (to obtain at least two heads in three throws of a coin) includes (the first) four results out of the eight possible ones: \((H, H, H), (H, H, T), (H, T, H), (T, H, H), (T, H, T), (H, T, T)\) and \( (T, T, T) \). As each result has a probability of 1/8 (if the probabilities of getting heads and tails are equal), the probability of the event is 1/2.

For combinations of events (i.e. two heads but such that not all throws give the same result) it becomes necessary to work with the concepts of union \((\cup)\) and intersection \((\cap)\) of set theory as presented in Figure 2.12. The rectangle in the figure represents the event space and \( A \) and \( B \) are events within it.

In general, it is true that

\[
P(A \cup B) = P(A) + P(B) - P(A \cap B)
\]

and if \( A \) and \( B \) are mutually exclusive,

\[
P(A \cup B) = P(A) + P(B)
\]
The conditional probability \(P(A/B)\) of \(A\) happening given that \(B\) is true, is:

\[
P(A/B) = \frac{P(A \cap B)}{P(B)}
\]

(2.54)

An event \(F\) is statistically independent of another event \(E\), if and only if (iff) \(P(F|E) = P(F)\). Therefore, for independent events we have:

\[
P(E \cap F) = P(E)P(F)
\]

(2.55)

which we applied intuitively when estimating event probability in Example 2.1.

2.5.2 Random Variables

These can be defined as those which take values following a certain probability distribution (see Figure 2.13).

![Random variable mapping from sample space](image)

**Figure 2.13** Random variable mapping from sample space

Example 2.2: The experiment ‘spinning a coin twice’, can yield only the following results (sample space): \(S = \{HH, HT, TH, TT\}\). If we define the random variable \(X\) = number of heads, it is easy to see that it can only take the following three values: \(X(HH) = 2, X(HT) = X(TH) = 1\) and \(X(TT) = 0\). Therefore, an advantage of the random variable concept becomes immediately apparent: the set of results (sample space) is reduced to a smaller, more convenient, numerical set (the range of the variable). The probabilities of \(X\) are as follows:

\[
P(X = 1) = P(HT \cup TH) = P(HT) + P(TH) = 1/2
\]

\[
P(X = 2) = P(X = 0) = 1/4.
\]

Random variables may be discrete or continuous. In the former case they can take values from a finite set with probabilities belonging to a set \(P(X)\) which satisfy (2.51) and \(p(x) \geq 0\). In the latter case it is necessary to define a probability density function \(f(x)\), such that:

\[
\int_{s} f(x) \, dx = 1
\]

\[
f(x) \geq 0, \quad \forall x
\]

(2.56)

2.5.3 Descriptive Statistics

When dealing with statistical data, summary information may be provided conveniently by specifying certain key features rather than the whole of a distribution. For example, the distribution of a random variable might be described with reference to its mean value and the dispersion around it. Descriptive statistics can be used to make simple comparisons between distributions without going into full details. More interestingly, certain standard distributions can be completely specified by just a few descriptive statistics.

The most usual descriptive statistic attempts to indicate the ‘middle’ of the distribution; three measures of central tendency are commonly defined to do this:

1. The mean (or expectation) \(E(X)\), given by:

\[
E(X) = \sum_{i} x_{i} p(x_{i}), \quad \text{discrete case}
\]

\[
E(X) = \int_{a}^{b} x f(x) \, dx, \quad \text{continuous case}
\]

(2.57)

where \(f(x)\) is defined for the range \((a \leq x \leq b)\). The mean is normally found by direct application of the expectation operator to the random variable \(X\). It can be applied also to functions of random variables. The operator has the important property of linearity, whereby for any random variables \(X\) and \(Y\), and constants \(a, b\) and \(c\) we have:

\[
E(a + bX + cY) = a + bE(X) + cE(Y)
\]

(2.58)

2. The mode \(X^*\) is the value of \(X\) which maximises \(p(x_i)\).
3. The median \(X_{0.5}\) is the value of \(X\) below which lies half of the distribution, that is:

\[
P(X_{0.5}) = \sum_{i} p(x_{i}) = 0.5 \quad \text{discrete case}
\]

\[
P(x < X_{0.5}) = \int_{x}^{X_{0.5}} f(x) \, dx = 0.5 \quad \text{continuous case}
\]

(2.59)

Neither the median nor the mode can be found by direct calculation, but need the solution of a problem.

Another important feature of a distribution is its spread or width; statistics commonly used for this purpose are:

1. The variance, which is the expectation of the square of the deviation from the mean:

\[
\text{Var}(X) = E[(X - E(X))^2]
\]

(2.60)
but it is normally calculated from the following result:

\[
\Var(X) = E[X^2] - 2E(X)E(X) + [E(X)]^2
\]

\[
= E(X^2) - 2[E(X)]^2 + [E(X)]^2
\]

\[
= E(X^2) - [E(X)]^2
\]

(2.61)

Unlike expectation, the variance is not a linear operator:

- \( \Var(a + bX) = b^2 \Var(X) \), i.e. adding a constant does not affect the spread of the distribution.
- \( \Var(aX + bY) = a^2 \Var(X) + b^2 \Var(Y) + 2ab \Cov(X, Y) \), where the covariance of \( X \) and \( Y \) is given by:

\[
\Cov(X, Y) = E(XY) - E(X)E(Y)
\]

(2.62)

Thus, the covariance of two mutually independent random variables is 0.

2 The standard deviation \( \sigma(X) \), which is the square root of the variance. This, in contrast with the variance, has the same dimensions as the random variable \( X \) and the measures of central tendency.

3 The coefficient of variation \( CV \), which is the ratio of the standard deviation to the mean, and is a dimensionless measure of spread.

3 Data and Space

This chapter is devoted to issues in data collection and their representation for use in transport modelling. We will consider four subjects which are a prerequisite for the rest of the book. Firstly, we will provide a brief introduction to statistical sampling theory, which will complement the elementary concepts discussed in section 2.5. Interested readers are advised that there is a complete book on the subject (Stephner and Meyburg 1979) which should be consulted for more details. In section 3.2 we will discuss the nature and importance of errors which can arise both during model estimation and when forecasting with the aid of models; the interesting question of data accuracy versus model complexity and cost is also addressed.

In section 3.3 we will consider the various types of surveys used in applied transport planning; we will be particularly interested in problems such as the correction, expansion and validation of survey data. We will also discuss issues involved in the collection of both stated preferences and longitudinal (e.g. panel) data. Finally, section 3.4 considers the important practical problems of network representation and zoning design; this is where the 'spatial capabilities' of the model are actually decided. Poor network representations or too coarse zoning systems may invalidate the results of even the most theoretically appealing model.

3.1 BASIC SAMPLING THEORY

3.1.1 Statistical Considerations

Statistics may be defined as the science concerned with gathering, analysing and interpreting data in order to obtain the maximum quantity of useful information. It may also be described as one of the disciplines concerned with decision making under uncertainty; its goal would be in this case to help determine the level of uncertainty associated with measured data in order to support better decisions.

Data usually consist of a sample of observations taken from a certain population of interest which is not economically (or perhaps even technically) feasible to observe in its entirety. These observations are made about one or more attributes of each member of the population (say income). Inferences can be made then about the mean value of these attributes, often called parameters of the population. Sample design aims at ensuring that the data to be examined provide the greatest amount of useful information about the population of interest at the lowest possible cost; the problem remains of how to use
the data (i.e. expand the values in the sample) in order to make correct inferences about this population. Thus two difficulties exist:

- how to ensure a representative sample; and
- how to extract valid conclusions from a sample satisfying the above condition.

Neither of these would constitute a problem if there was no variability in the population. To solve the second difficulty, a well-established procedure exists which does not present major problems if certain conditions and assumptions hold. The identification of a representative sample, however, may be a more delicate task in certain cases, as we shall see below.

### 3.1.1 Basic Definitions

**Sample** The sample is defined as a collection of units which has been especially selected to represent a larger population with certain attributes of interest (i.e. height, choices, options). Three aspects of this definition have particular importance: first, which population the sample seeks to represent; second, how large the sample should be; and third, what is meant by 'especially selected'.

**Population of Interest** This is the complete group about which information is sought; in many cases its definition stems directly from the study objectives. The population of interest is composed of individual elements; however, the sample is usually selected on the basis of sampling units which may not be equivalent to these individual elements as aggregation of the latter is often deemed necessary. For example, a frequently used sampling unit is the household while the elements of interest are individuals residing in it.

**Sampling Method** Most of the acceptable methods are based on a form of random sampling. The key issue in these cases is that the selection of each unit is carried out independently, with each unit having the same probability of being included in the sample. The more interesting methods are:

- **Simple random sampling**, which is not only the simplest method but constitutes the basis of all the rest. It consists in first associating an identifier (number) to each unit in the population and then selecting these numbers at random to obtain the sample; the problem is that far too large samples may be required to ensure sufficient data about minority options of particular interest. For example, it may well be that sampling households at random in a developing country would provide little information on multiple car ownership.

- **Stratified random sampling**, where *a priori* information is first used to subdivide the population into homogeneous strata (with respect to the stratifying variable) and then simple random sampling is conducted inside each stratum using the same sampling rate. The method allows the correct proportions of each stratum in the sample to be obtained; thus it may be important in those cases where there are relatively small subgroups in the population as they could lack representation in a simple random sample.

It is also possible to stratify with respect to more than one variable, thus creating an $n$-dimensional matrix of group cells. However, care must be taken with the number of cells created as it increases geometrically with the number of strata; large figures may imply that the average number of sampling units per cell is too small. Nevertheless, even stratified sampling does not help when data are needed about options with a low probability of choice in the population; in these cases a third method called *choice-based sampling* actually a subset of the previous one, is required. The method consists in stratifying the population based on the result of the choice process under consideration. This method is fairly common in transport studies, as we will see in section 3.3. A major advantage is that data may be produced at a much lower cost than with the other sampling methods; its main drawback is that the sample thus formed may not be random and therefore the risk of bias in the expanded values is greater.

**Sampling Error and Sampling Bias** These are the two types of error that might occur when taking a sample; combined, they contribute to the measurement error of the data. The first is simply due to the fact that we are dealing with a sample and not with the total population, i.e. it will always be present due to random effects. This type of error does not affect the expected values of the means of the estimated parameters; it only affects the variability around them, thus determining the degree of confidence that may be associated to the means; it is basically a function of sample size and of the inherent variability of the parameter under investigation.

The sampling bias, on the other hand, is caused by mistakes made either when defining the population of interest, or when selecting the sampling method, the data collection technique or any other part of the process. It differs from the sampling error in two important respects:

- It can affect not only the variability around the mean of the estimated parameters but the values themselves; therefore it implies a more severe distortion of the survey results;

- While the sampling error may not be avoided (it can only be reduced by increasing sample size), the sampling bias may be virtually eliminated by taking extra care during the various stages of sampling design and data collection.

**Sample Size** Unfortunately, there are no straightforward and objective answers to the calculation of sample size in every situation. This happens, in spite of the fact that sample size calculations are based on precise statistical formulae, because many of their inputs are relatively subjective and uncertain; therefore they must be produced by the analyst after careful consideration of the problem in hand.

Determining sample size is a problem of trade-offs, as:

- a much too large sample may imply a data-collection and analysis process which is too expensive given the study objective and its required degree of accuracy; but
- a far too small sample may imply results which are subject to an unacceptably high degree of variability reducing the value of the whole exercise.

Somewhere between these two extremes lies the most efficient (in cost terms) sample size given the study objective. In what follows it will be assumed that this consists in estimating certain population parameters by means of statistics calculated from sample
data; as any sample statistics are subject to sampling error, it is also necessary to include an estimate of the accuracy that may be associated to its value.

3.1.1.2 Sample Size to Estimate Population Parameters

This depends on three main factors: variability of the parameters in the population under study, degree of accuracy required for each, and population size. Without doubt the first two are the most important; this may appear surprising at first sight because, to many, it seems intuitively necessary to take bigger samples in bigger populations in order to maintain the accuracy of the estimates. However, as will be shown below, the size of the population does not significantly affect sample size except in the case of very small populations.

The Central Limit Theorem, which is at the heart of the sample size estimation problem, postulates that the estimates of the mean of a sample tend to become distributed Normal as the sample size (n) increases. This holds for any population distribution if n is greater than or equal to 30; the theorem holds even in the case of smaller samples, if the original population has a Normal-like distribution.

Consider a population of size N and a specific property which is distributed with mean μ and variance σ². The Central Limit Theorem states that the distribution of the mean (x̄) of successive samples is distributed Normal with mean μ and standard deviation se (x̄), known as the standard error of the mean, and given by:

\[
\text{se} (x̄) = \sqrt{\frac{(N - n)\sigma^2}{n(N - 1)}}
\]  

(3.1)

If only one sample is considered, the best estimate of μ is x̄ and the best estimate of σ² is S² (the sample variance); in this case the standard error of the mean can be estimated as:

\[
\text{se} (x̄) = \sqrt{\frac{(N - n)S^2}{nN}}
\]  

(3.2)

and, as mentioned above, it is a function of three factors: the parameter variability (S²), the sample size (n) and the size of the population (N). However, for large populations and small sample sizes (the most frequent case) the factor (N - n)/N is very close to 1 and equation (3.2) reduces to:

\[
\text{se} (x̄) = \frac{S}{\sqrt{n}}
\]  

(3.3)

Thus, for example, quadrupling sample size will only halve the standard error, i.e. it is a typical case of diminishing returns of scale. The required sample size may be estimated solving equation (3.2) for n, it is usually simpler to do it in two stages, first calculating n from equation (3.3) such that:

\[
n = \frac{S^2}{\text{se} (x̄)^2}
\]  

(3.4)

then correcting for finite population size, if necessary, by:

\[
n = \frac{n'}{1 + \frac{n'}{N}}
\]  

(3.5)

Although the above procedure appears to be both objective and relatively trivial it has two important problems that impair its application: estimating the sample variance S² and choosing an acceptable standard error for the mean. The first one is obvious: S² can only be calculated once the sample has been taken, so it has to be estimated from other sources. The second one is related with the desired degree of confidence to be associated with the use of the sample mean as an estimate of the population mean; normal practice does not specify a single standard error value, but an interval around the mean for a given confidence level. Thus, two judgements are needed to calculate an acceptable standard error:

• First a confidence level for the interval must be chosen; this expresses how frequently the analyst is prepared to make a mistake by accepting the sample mean as a measure of the true mean (e.g. the typical 95% level implies a willingness to err in 5% of cases).

• Second, it is necessary to specify the limits of the confidence interval around the mean, either in absolute or relative terms; as the interval is expressed as a proportion of the mean in the latter case, an estimate of this is required to calculate the absolute values of the interval. A useful option considers expressing sample size as a function of the expected coefficient of variation (CV = s/μ) of the data.

For example, if a Normal distribution is assumed and a 95% confidence level is specified, this means that a maximum value of 1.96 se (x̄) would be accepted for the confidence interval (i.e. μ ± 1.96 se contains 95% of the Normal probability distribution); if a 10% error is specified we would get a (μ ± 1.0μ) interval and it may be seen that:

\[
\text{se} (x̄) = 0.1\mu / 1.96 = 0.051\mu
\]  

and replacing this value in (3.4) we get:

\[
n' = (S/0.051\mu)^2 = 384CV^2
\]  

(3.6)

Note that if the interval is specified as (μ ± 0.05μ), i.e. with half the error, n' would increase fourfold to 1536 CV².

To complete this point it is important to emphasise that the above exercise is relatively subjective; thus, more important parameters may be assigned smaller confidence intervals and/or higher levels of confidence. However, each of these actions will result in smaller acceptable standard error and, consequently, bigger samples and costs. If multiple parameters need to be estimated the sample may be chosen based on the requirement of a larger sample size.
3.1.3 Obtaining the Sample

The last stage of the sampling process is the extraction of the sample itself. In some cases the procedure may be easily automated, either on site or at the desk (in which case care must be taken that it is actually followed on the field), but it must always be conducted with reference to a random process. Although the only truly random processes are those of a physical nature (i.e. roll of a dice or flip of a coin), they are generally too time consuming to be useful in sample selection. For this reason pseudo-random processes, capable of generating easily and quickly a set of suitable random-like numbers, are usually employed in sampling.

Example 3.1: Consider a certain area whose population may be classified in groups according to: automobile ownership (with and without a car); and household size (up to four and more than four residents).

Let us assume that m observations are required by cell in order to guarantee a 95% confidence level in the estimation of, say, trip rates; assume also that the population can be considered to have approximately the following distribution (i.e. from historic data).

<table>
<thead>
<tr>
<th>Car ownership</th>
<th>Household size</th>
<th>% of population</th>
</tr>
</thead>
<tbody>
<tr>
<td>With car</td>
<td>Four or less</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>More than four</td>
<td>16</td>
</tr>
<tr>
<td>Without car</td>
<td>Four or less</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>More than four</td>
<td>50</td>
</tr>
</tbody>
</table>

There are two possible ways to proceed:

1. Achieve a sample with m observations by cell by means of a random sample. In this case it is necessary to select a sample size which guarantees this for each cell, including that with the smallest proportion of the population. Therefore, the sample size would be:

\[ n = \frac{100m}{9} = 11.1m \]

2. Alternatively, one can undertake first a preliminary random survey of 11.1m households where only cell membership is asked for; this low-cost survey can be used to obtain the addresses of m households even in the smallest group. Subsequently, as only m observations are needed by cell, it would suffice to randomly select a (stratified) sample of 3m households from the other groups to be interviewed in detail (together with the m already detected for the most restrictive cell).

As can be seen, a much higher sample is obtained in the first case; its cost (approximately three times more interviews) must be weighed against the cost of the preliminary survey.

3.1.2 Conceptualisation of the Sampling Problem

In this part we will assume that the final objective of taking the sample is to calibrate a choice model for the whole population. Following Lerman and Manski (1976) we will denote by \( P \) and \( f \) population and sample characteristics respectively. We will also assume that each sampled observation may be described on the basis of the following two variables:

\[ i = \text{observed choice of the sample individual (e.g. took a bus);} \]

\[ X = \text{vector of characteristics (attributes) of the individual (age, sex, income, car ownership) and of the alternatives in his choice set (walking, waiting and travel times, cost).} \]

We will finally assume that the underlying choice process in the population may be represented by a model with parameters \( \theta \); in this case, the joint distribution of \( i \) and \( X \) is given by:

\[ P(i, X/\theta) \]

and the probability of choosing alternative \( i \) among a set of options with attributes \( X \) is:

\[ P(i/X, \theta) = \text{Prob of choosing } i \text{ given person type } X \text{ and model } \theta \]

Depending on the form in which each observation is extracted, the sample will have its own joint distribution of \( i \)'s and \( X \)'s which we will denote by \( f(i, X/\theta) \). On the basis of this notation the sampling problem may be formalised as follows (Lerman and Manski, 1979).

3.1.2.1 Random Sample

In this case the distribution of \( i \) and \( X \) in the sample and population should be identical, that is:

\[ f(i, X/\theta) = P(i, X/\theta) \tag{3.7} \]

3.1.2.2 Stratified Sample sampling by categories defined by \( X \)

In this case the sample is not random with respect to certain independent variables of the choice model (e.g. a sample with 50% low-income households and 50% high-income households is stratified if and only if a random sample is taken inside each stratum). The sampling process is defined by a function \( f(X) \), giving the probability of finding an observation with characteristics \( X \); in the population this probability is of course \( P(X) \). The distribution of \( i \) and \( X \) in the sample is thus given by:

\[ f(i, X/\theta) = f(X)P(i, X, \theta) \tag{3.8} \]
It is simple to show that a random sample is just a special case of stratified sample
where \( f(X) = P(X) \), because:

\[
f(i, X|\theta) = P(X|\theta)P(i|X, \theta) = P(i, X|\theta)
\]  

(3.9)

3.12.3 Choice-based Sample

Sampling by categories defined by \( i \)

In this case the sampling procedure is defined by a function \( f(i) \), giving the probability
of finding an observation that chooses option \( i \). Now the distribution of \( i \) and \( X \) in the
sample is given by:

\[
f(i, X|\theta) = f(i)P(X|i, \theta)
\]  

(3.10)

We had not defined this latter probability, but we may obviate it on the basis of a
Bayes theorem stating:

\[
P(X|i, \theta) = \frac{P(i|X, \theta)P(X)}{P(i|\theta)}
\]  

(3.11)

The expression in the denominator, which has not been defined either, may be
obtained assuming discrete \( X \) from:

\[
P(i|\theta) = \sum_X P(i|X, \theta)P(X)
\]  

(3.12)

Therefore the final expression for the joint probability of \( i \) and \( X \) for a choice-based
sample is clearly more complex:

\[
f(i, X|\theta) = f(i)\frac{P(i|X, \theta)P(X)}{\sum_X P(i|X, \theta)P(X)}
\]  

(3.13)

and it serves to illustrate not only that choice-based sampling is intuitively more
problematic than the others, but also that it has higher bias potential in what really
concerns us: choice.

Thus, each sampling method yields a different distribution of choices and character-
istics in the sample, and there are no a priori reasons to expect that a single parameter
estimation method would be applicable in all cases.

**Example 3.2:** Assume that for the purposes of a transport study the population of a
certain area has been classified according to two income categories, and that there are
only two means of transport available (car and bus) for the journey to work. Let us also
assume that the population distribution is given by:

<table>
<thead>
<tr>
<th></th>
<th>Low income</th>
<th>High income</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bus user</td>
<td>0.45</td>
<td>0.15</td>
<td>0.60</td>
</tr>
<tr>
<td>Car user</td>
<td>0.20</td>
<td>0.20</td>
<td>0.40</td>
</tr>
<tr>
<td>Total</td>
<td>0.65</td>
<td>0.35</td>
<td>1.00</td>
</tr>
</tbody>
</table>
3.1.3 Practical Considerations in Sampling

3.1.3.1 The Implementation Problem

Stratified (and choice-based) sampling requires random sampling inside each stratum; to do so it is first necessary to isolate the relevant group and this may turn out to be not easy in practice. Consider for example a case where the population of interest consists of all potential travellers in a city. Thus if we stratify by area of residence, it may be relatively simple to isolate the subpopulation of residents inside the city (e.g. using data from a previous survey); the problem is that it is extremely difficult to isolate and take a sample of the rest, i.e. those living outside the city.

An additional problem is that in certain cases even if it is possible to isolate all subpopulations and conforming strata, it may still be difficult to ensure a random sample inside each stratum. For example, if we are interested in taking a mode choice-based sample of travellers in a city we will need to interview bus users and for this it is first necessary to decide which routes will be included in the sample. The problem is that certain routes might have, say, higher than average proportions of students and/or older age pensioners, and this would introduce bias.

3.1.3.2 Finding the Size of Each Subpopulation

This is a key element in determining how many people will be surveyed. Given a certain stratification, there are several methods available to find out the size of each subpopulation, such as:

1. Direct measurement. This is possible in certain cases. Consider a mode choice-based sample of journey-to-work trips; the number of bus and metro tickets sold, plus traffic counts during the peak hour in an urban corridor, may yield an adequate measure (although imperfect as not all trips during peak are to work) of the number of people choosing each mode. If we have a geographical (i.e. zonal) stratification, on the other hand, the last census may be used to estimate the number of inhabitants in each zone.

2. Estimation from a random sample. If a random sample is taken, the proportion of observations corresponding to each stratum is a consistent estimator of the fraction of the total corresponding to each subpopulation. It is important to note that the cost of this method is low as the only information sought is that necessary to establish the stratum to which the interviewee belongs.

3. Solution of a system of simultaneous equations. Assume we are interested in stratifying by chosen mode and that we have data about certain population characteristics (e.g. mean income and car ownership). Taking a small on-mode sample we can obtain modal average values of these variables and postulate a system of equations which has the subpopulation fractions as unknowns.

Finally, the ‘failure rate’ of different types of surveys must be considered when designing sampling frameworks. The sample size discussed above corresponds to the number of successful and valid responses to the data-collection effort. Some survey
procedures are known to generate low valid response rates (e.g. some postal surveys), but they may still be used because of their low cost.

**Example 3.3:** Assume the following information is available:

- Average income of population (I): 33 600 $/year
- Average car ownership (CO): 0.44 cars/household

Assume also that small on-mode surveys yield the following:

<table>
<thead>
<tr>
<th>Mode</th>
<th>I ($/year)</th>
<th>CO (cars/household)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Car</td>
<td>78 000</td>
<td>1.15</td>
</tr>
<tr>
<td>Bus</td>
<td>14 400</td>
<td>0.05</td>
</tr>
<tr>
<td>Metro</td>
<td>38 400</td>
<td>0.85</td>
</tr>
</tbody>
</table>

If $F_i$ denotes the subpopulation fraction of the total, the following system of simultaneous equations holds:

\[
33,600 = 78,000F_1 + 14,400F_2 + 38,400F_3 \\
0.44 = 1.15F_1 + 0.05F_2 + 0.85F_3 \\
1 = F_1 + F_2 + F_3
\]

the solution of which is:

\[
F_1 = 0.2451 \\
F_2 = 0.6044 \\
F_3 = 0.1505
\]

This means that if the total population of the area was 180,000 inhabitants, there would be approximately 44,100 car users, 108,800 bus users and 27,100 metro users.

### 3.2 ERRORS IN MODELLING AND FORECASTING

The statistical procedures normally used in (travel demand) modelling *assume not only* that the correct functional specification of the model is known *a priori*, but also that the data used to *estimate the model parameters have no errors*. In practice, however, these conditions are often violated; furthermore, even if they were satisfied, model forecasts are usually subject to errors due to inaccuracies in the values assumed for the explanatory variables in the design year.

The ultimate goal of modelling is often forecasting (i.e. the number of people choosing given options); an important problem model designers face is to find which combination of model complexity and data accuracy fits best the required forecasting precision and study budget. To this end, it is important to distinguish between different types of errors, in particular:

- those that could cause even correct models to yield incorrect forecasts, e.g. errors in the prediction of the explanatory variables, transference and aggregation errors; and
- those that actually cause incorrect models to be estimated, e.g. measurement, sampling and specification errors.

In the next section consideration is given first to the types of errors that may arise with the broad effects they may cause; then the trade-off between model complexity and data accuracy is examined with particular emphasis on the role of simplified models in certain contexts.

#### 3.2.1 Different Types of Error

Consider the following list of errors that may arise during the processes of building, calibrating and forecasting with models.

##### 3.2.1.1 Measurement Errors

These occur due to the inaccuracies inherent in the process of actually measuring the data in the base year, such as: questions badly registered by the interviewee, answers badly interpreted by the interviewer, network measurement errors, coding and digitising errors, and so on. These errors tend to be higher in less developed countries but they can always be reduced by improving the data-collection effort (e.g. by appropriate use of computerised interview support) or simply by allocating more resources to data quality control; however, both of these cost money.

Measurement error, as defined here, should be distinguished from the difficulty of defining the variables that ought to be measured. The complexity that may arise in this area is indicated in Figure 3.1; ideally, modelling should be based on the information perceived by individual travellers but whilst reported data may give some insight into perception, its use raises the difficult question of how to forecast what users are going to perceive in the future. So it appears inevitable that models will be endowed with perception errors which tend to be greater for non-chosen alternatives due to the existence of self-selectivity bias (i.e. the attributes of the chosen option are perceived as better and those of the rejected option as worse than they are, such as to reinforce the rationality of the choice made).

##### 3.2.1.2 Sampling Errors

These arise because the models must be estimated using finite data sets. Sampling errors are approximately inversely proportional to the square root of the number of observations (i.e. if the sample is doubled, it is necessary to quadruple sample size); thus, reducing them may be costly. Deganzo (1990) has examined the problem of defining optimal sampling strategies in the sense of refining estimation accuracy.