æ

PRECALCULUS:

MODELING

by Ath. Kehagias

American College of Thessaloniki

February 1997

Foreword

I wrote these notes in response to an exceptional class I taught at ACT in the Fall of 1996. They were rather comfortable with the usual mathematical content of Math 100, but they wanted “more examples”. So I thought that I would package the mathematical techniques in a more applications-oriented wrap. Hence the idea of centering the course around modeling. In the process of teaching the course and writing the notes I realized how far one can go in terms of real (or nearly-real) world examples with relatively simple mathematical tools.

The notes were written with the intention of using the computer for experimentation, and this is how the class was taught. While there was considerable resistance to this idea from the students’ part, I believe in the end they were convinced that the effort required to get the necessary computer skills (that is, the use of Microsoft Excel) paid off in terms of increased understanding of the mathematical *concepts*, unencumbered by the burden of numerical computation by hand.

I want to thank Panos Vlachos and Kostis Vezerides for long and enlightening discussions that helped to improve these notes and my state of mind in general. I dedicate these notes to Math 100 F, Fall 1996, for a pleasantly unforgettable semester.

Contents

1 Introduction

The problem we wil consider is finding a *model* for a given *data set*. A data set is simply a table of numbers, such as the following. In Table 1.1 we have a row of x-values and a row of y-values.

x	y
1	3
2	5
3	7
4	9

Table 1.1

These values are the data, also called *observations* or *data points*. The term "points" is used because we can consider the data to be x and y coordinates of points in the $x - y$ plane. In fact, when building a model, the first step is to make a graph of our data. For Table 1.1, the corresponding graph is displayed in Figure 1.1. Finding a model means finding a relationship

Figure 1.1

between the x and y values; in other words a formula that connects every x value with the corresponding y value. $y = ax + b$ is a simple example for such a relationship, but many other possibilities exist. In fact, in this particular example, it is easy to see that $y = 2x + 1$, in other words $a = 2$ and $b = 1$. In other cases the formula may be any of the following

1. more complicated,
2. harder to find,
3. *approximate* (in other words it does not *fit* exactly to the data).

What is the use of modelling a data set?

First, a model results in "economy" of description. Table 1.1 contains 4 x-values and 4 y-values, a total of 8 numbers. But to recreate the table, all these numbers are not necessary. In fact we only need two numbers: $a = 2$ and $b = 1$ from the equation $y = ax + b$. Given these two numbers, we can generate the complete table: for every x value that we choose we can *compute* the corresponding y value by using the formula $y = 2x + 1$. For example, we can take $x = 1$ and find $y = 2 \cdot 1 + 1 = 3$. In fact we can expand the table, by using x values not in it. For example, take $x = 5$ and find $y = 2 \cdot 5 + 1 = 11$. The numbers a and b are usually called the *parameters* of *coefficients* of the model. A model with few parameters is generally considered better than a model with many parameters.

Second, the model (i.e. the formula) may describe an underlying physical situation, which was initially unknown to us. In general, when we create models, we are faced with one of two situations.

1. We may be presented with a bunch of numbers, without any understanding of the process from which these numbers came. For instance, we may have a table of the daily price of a stock; we do not know why the price changes from day to day. In this case, finding a relationship between day (call it t) and price (call it $p(t)$), will make it possible to *predict* the price $p(t)$ for day nr. t.

2. We may know the physical situation that creates the data, and even a general formula that describes this situation, but the parameters of the formula may be unknown, or not easy to determine. For example, in a problem of motion with constant speed, we know that at time t the position $s(t)$ is given by $s(t) = at + b$ (a is the speed and b is the initial position) , but we may have trouble in finding a and b; however we can easily create a table of times and positions. The table can be used to find a and b and this will give us a physical understanding, namely what the speed and initial position are.

At any rate, suppose that (for whatever reason) our goal is to find a model of a data set. Hence we will study methods of *model building*. Out of the many existing possibilities, we will consider just a few, which can be categorized by the type of models used. We will consider

1. linear models: $y = ax + b$;

2. polynomial models: $y = a_M x^M + a_{M-1} x^{M-1} + ... + a_1 x + a_0$;

3. rational models: $y = \frac{ax+b}{cx+d}$;

4. recursive models: $y(x) = a_1 y(x-1) + a_2 y(x-2) + ... + a_M y(x-M) + b$.

In every case we will: (a) review mathematical properties of the models, (b) present methods for discovering the models and (c) apply these methods to particular data sets.

2 Linear and Piecewise Linear Models

The first type of models which we will use is based on the *linear function* $y(x) = ax + b$. This is probably the simplest possible function. It has two parameters: a, b Let us first recall its basic properties.

2.1 Linear Functions

The following are the basic formulas associated with the linear function.

1. Equation of a straight line: $y(x) = ax + b$.

2. More general equation of a straight line: $Ax + By = C$.

3. a is the *slope* of the line; if the line passes through points (x_1, y_1) and (x_2, y_2) we have $a = \frac{y_2 - y_1}{x_2 - x_1}$.

   An alternative interpretation of a is in terms of *rate of change*. Suppose that x was initially x_1 and then changes to $x_2 = x_1 + h$; the "absolute change" in x is $\Delta x = x_2 - x_1 = (x_1 + h) - x_1 = h$. What is the respective absolute change of $y(x)$? Obviously it is

   $$\Delta y = y(x_2) - y(x_1) = y(x_1 + h) - y(x_1) = a \cdot (x_1 + h) + b - (ax_1 + b) = a \cdot h. \tag{1}$$

   But in many cases what we are interested in is not absolute, but *relative* change, in other words the change of y compared to the change of x. To find this we can use the *rate of change* $\frac{\Delta y}{\Delta x}$. We then have

   $$\frac{\Delta y}{\Delta x} = \frac{a \cdot h}{h} = a. \tag{2}$$

   From eq.(2) it follows that $\Delta y = a \cdot \Delta x$. Hence, an interpretation of a is this: whenever x changes by Δx, $y(x)$ changes by $a \cdot \Delta x$. Hence we call a the *rate of change* of $y(x)$.

4. Equation of a straight line passing through points (x_1, y_1) and (x_2, y_2): $\frac{y - y_1}{x - x_1} = \frac{y_2 - y_1}{x_2 - x_1}$.

2.2 Exact Linear Models

The simple linear model is just the formula $y = ax + b$. This is probably the simplest possible model to use; in many cases it is sufficient. By looking at the data table or the data graph, we may realize that a the data fit to a linear model; in other words that the relationship between x and y is simply $y = ax + b$. In this case, it is required to find the coefficients a and b. To do this we select two points from the table (i.e. two pairs of x and y values) and pass a straight line through these points. Then we check that the straight line also passes through the remaining points.

Example Find the straight line passing through points:

x	y
1	-1
3	3
6	9
8	13

Solution We start by graphing the data. The graph looks like this:

Figure 2.1

It appears that the data lie on a straight line. To find its equation, choose two points from the table, say $(1, -1)$ and $(6, 9)$. Using these two with the equation $y = ax + b$, we get the equations

$$\left.\begin{array}{rrl} (-1,1) & -1 = & a \cdot 1 + b \\ (6,9) & 9 = & a \cdot 6 + b \end{array}\right\} \Rightarrow \left.\begin{array}{rl} a + b = & -1 \\ 6a + b = & 9 \end{array}\right\} \Rightarrow$$

$$(6a + b) - (a + b) = 9 - (-1) \Rightarrow 5a = 10 \Rightarrow a = 2 \Rightarrow 2 + b = -1 \Rightarrow b = -3.$$

Then we have $y = 2x - 3$; we check these with the remaining points (3,3) and (8,13): $3 = 2 \cdot 3 - 3$ and $13 = 2 \cdot 8 - 3$ are true. Hence the straight line passing through the points in the data set is $y = 2x - 3$. •

2.2.1 Supply, Demand and Price

It is rather obvious that when the price of a product increases, the sales decrease. Let us consider a simple model of this dependence. Suppose that the market research has established that google price p is related to number of googles sold q according to the following table (prices are in \$).

p	q
40	30
50	25
60	20
70	15

The Google Makers company considers raising the price of googles, to 80 but they would like to know how many googles will be sold at that price.

To answer this question, let us first graph the data of the table. We get the following graph.

Figure 2.2

It is obvious that the data points lie on a straight line. We expect to have a relationship of the form $q = ap + b$ and what is required is to find a and b. We select two points from the table, say (40,30), (70,15). We have the following equations.

$$(40, 30): \quad 30 = 40a + b,$$

$$(70, 15): \quad 15 = 70a + b.$$

Subtracting the second equation from the first we get

$$15 = -30a \Rightarrow a = -\frac{1}{2}.$$

Then we substitute the value of a in the first equation, to get

$$30 = 40 \cdot (-\frac{1}{2}) + b \Rightarrow b = 50.$$

It follows that $q = -\frac{1}{2} \cdot p + 50$. Now, to answer the original question, substituting $p = 80$, we get $q = -\frac{1}{2} \cdot 80 + 50 = 10$. So, at a price of 80 \$ per google, 10 googles will be sold.

Looking at the problem in this way, *demand* for googles determines price. An alternative way of looking at this situation, is to assume that when a given quantity of googles is produced and thrown into the market, the price at which googles sell is determined by the laws of the "free market economy". From this point of view, the phenomenon is described by a function of the form $p = aq + b$ and we want to find a and b (which will be different from the previous case!) using the same data. Using the same two points as before (but with the order of numbers now reversed) we have teh following equations.

$$(30, 40): \quad 40 = 30a + b,$$

$$(15, 70): \quad 70 = 15a + b.$$

Subtracting the second equation from the first we get

$$-30 = 15a \Rightarrow a = -2.$$

Then we substitute the value of a in the first equation, to get

$$40 = 30 \cdot (-2) + b \Rightarrow b = 100.$$

It follows that $p = -2 \cdot q + 100$. If, for example, 45 googles are placed in the market, the price at which they will be sold will be $p = -2 \cdot 45 + 100 = 10$.

This point of view expresses demand (as described by price p) as a function of supply q (that's why $p = -2 \cdot q + 100$ is sometimes called a *demand* function) while the previous formulation expresses supply q as a function of demand (i.e. price, p). Both expressions describe a law of supply and demand.

There is no reason to attach more importance to p or q. We can also formulate the law of supply and demand by expressing the dependence between p and q by a general linear function. Do this on your own, using the data above, and discover that the linear function is

$$p + 2q = 100.$$

This is a rather simplistic law of supply and demand; a more complicated, non-linear relationship may hold between p and q.

2.2.2 Alex Takes a Walk

Alex goes for a walk in the countryside. He starts at point A and moves in a straight line. Taking his starting time as $t = 0$ and expressing his position as distance from point A, we observe that at various times he is at the following positions.

t	s
0	0
5	25
10	50
12	60

What is the function of Alex's position s as a function of time t? Graphing the data we have the following figure.

Figure 2.3

It is obvious that the data points lie on a straight line. We expect to have a relationship of the form $s = at + b$ and what is required is to find a and b. We select two points from the table, say (0,0), (12,60). We have the following equations.

$$(0,0): \quad 0 = 0a + b,$$

$$(12,60): \quad 60 = 12a + b.$$

From the first equation it follows immediately that $b = 0$. Substituting in the second equation we get

$$60 = 12a \Rightarrow a = 12.$$

It follows that $s = 12t$. a=12 is Alex's *speed*.

2.3 Exact Piecewise Linear Model

The piecewise linear model consists of several formulas of the form $y = ax + b$. By looking at the data table or, especially, at the data graph, we may realize that different parts of the data fit to separate linear models fit different parts of the data. For example, the first half of the data fit to $y = a_1x + b_1$ and the second half of the data fit to $y = a_2x + b_2$. In this case, it is required to find the coefficients a_1 and b_1, a_2 and b_2. To do this we split the table into two sub-tables; for each one we select two points from the table (i.e. two pairs of x and y values) and pass a straight line through these points. Then we check that the straight line also passes through the remaining points.

Formulas To find the intersection of two straight lines, say $y = a_1x + b$ and $y = a_2x + b$, we need to solve the system

$$\left.\begin{array}{rl} y = & a_1x + b_1 \\ y = & a_2x + b_2 \end{array}\right\} \Rightarrow a_1x + b_1 = a_2x + b_2.$$

Example Find a piecewise linear model passing through points:

x	y
1	3
2	5
3	7
4	6.5
5	5.5
6	4.5
7	3.5
8	2.5

Solution We start by graphing the data. The graph is given in Figure 2.4.

Figure 2.4

It appears that the data lie on two straight lines, one for the first three points, the other for the remaining five. To find the first line, choose two points from (1,3), (2,5), (3,7); say we choose (1,3), (2,5). We have the system of equations

$$\left.\begin{array}{lll}(1,3) & 3= & a\cdot 1+b\\ (2,5) & 5= & a\cdot 2+b\end{array}\right\}\Rightarrow \left.\begin{array}{ll}a+b= & 3\\ 2a+b= & 5\end{array}\right\}\Rightarrow$$

$$(2a+b)-(a+b)=5-3\Rightarrow a=2\Rightarrow 2+b=3\Rightarrow b=1.$$

Then we have $y=2x+1$; we check these with the remaining point (3,7): $7=2\cdot 3+1$ is true. Hence the straight line passing through the first three points in the data set is $y=2x+1$.

We now proceed in the same manner for the remaining five points. We choose two of these, say $(4,6.50)$, $(5,5.50)$. From these we obtain the equations

$$4a+b=6.50,\quad 5a+b=5.50;$$

solving these we obtain $a=-1$ and $b=10.50$. We check the remaining points: (6,5.50) gives $4.50=-6+10.50$, (7,3.50) gives $3.50=-7+10.50$, (8,2.50) gives $2.50=-8+10.50$, all of which are true. Hence our piecewise linear model is

$$y=\left\{\begin{array}{ll}2x+1 & x\leq 3\\ -x+6.50 & x\geq 4\end{array}\right.$$

For the interval $3\leq x\leq 4$ we can extend either of the straight lines, or find their point of intersection. The graph of the data with the piecewise linear model is given in 2.5 and the solution of the problem is complete. •

Figure 2.5

2.3.1 An Example: Alex and Basil Take a Walk

Alex and Basil are running after each other. To be exact, this is how things happened. At 12 o' clock Alex starts moving in a straight line, with constant speed of 10 m/sec; at 12:10 Basil goes after him, in the same straight line, with a speed of 20 m/sec. What is required is to find a model of Alex's position as a function of time and a model of Basil position as a function of time.

We are talking here about linear models of the form $s(t) = at + b$. To find a and b, let us first build a table of time, Alex's position and Basil's position. Time is measured in seconds.

t	$s_A(t)$	$s_B(t)$
0	0	0
60	600	0
600	6000	0
1200	12000	12000
2400	24000	36000

Figure 2.6

Graphing the data, in Figure 2.6, they appear to lie on straight lines. For Alex, we will use one straight line, so $s_A(t) = at + b$. Pick the first two points (0,0), (60,600) to get the equations

$$y = \begin{array}{lll} (0,0) & 0 = & a \cdot 0 + b \\ (60,600) & 600 = & a \cdot 60 + b \end{array} \Big\} \Rightarrow$$

b=0, a=10, $s_A(t) = 10t$. This is just what we expected, namely a linear model (because Alex is moving with constant speed) with a being Alex's speed.

For Basil, it appears that we will need two straight lines, so $s_B(t)$ will be a piecewise linear model. For the interval $0 \leq t \leq 600$, pick the first two points (0,0), (60,0) to get the equations

$$y = \begin{array}{lll} (0,0) & 0 = & a \cdot 0 + b \\ (60,0) & 0 = & a \cdot 60 + b \end{array} \Big\} \Rightarrow$$

b=0, a=0, $s_B(t) = 0$, which holds until time $t = 600$. This is just what we expected, since Basil does not start moving until $t = 600$, i.e. ten minutes after Alex.

For the interval $600 \leq t$, pick the points (600,0), (1200, 12000) to get the equations

$$y = \begin{array}{lll} (600,0) & 0 = & a \cdot 600 + b \\ (1200,12000) & 12000 = & a \cdot 1200 + b \end{array} \Big\} \Rightarrow$$

b=-12000, a=20, $s_B(t) = 20t - 12000$, which holds after time $t = 600$. •

2.3.2 The Salesman's Commissions

A salesman works for the Google Makers company under the following agreement. He has a monthly base salary of 1,000 \$. In addition, he gets a 5% commission on the total monthly sales, if the total sales are up to 10,000 \$; for sales over 10,000 \$, he gets a 7% commission. The price of one google is 40 \$. What is the salesman's salary per month as a function of monthly google sales?

We are looking for a function $s = f(q)$, where s is the salesman's salary, q is the amount of googles sold and $f(q)$ is a yet unknown function. To determine $f(q)$, let us first *create* a table with hypothetical sales and the respective salaries. For example, suppose the salesman sells no googles; then he gets the base salary of 1,000 \$. So a data point is (0,1000). Suppose he sells, 50 googles; this means sales are $50 \cdot 40 = 2,000$, on which he gets 5% commission, i.e. $2,000 \cdot 0.05 = 100$. This is *in addition* to the 1,000 \$ base salary, so the total salary is 1,000+100=1,100 \$. So a new data point is (50,1100). Similarly, for sales of 100 googles, we get data point (100, 1200). Suppose now the salesman sells 300 googles. Total sales are now $300 \cdot 40$= 12,000 \$, and, at 7%, his comission is $12,000 \cdot 0.07$= 840 \$, so he gets a total salary of 1,840 \$. Simiarly, for 600 googles sold, he gets a total of 2,680 \$. So we build a table as follows.

q	s
0	1,000
50	1,100
300	1,840
600	2,680

Figure 2.7

It appears that the salary is given by a piecewise linear function. Graph the data and determine the form of this function (i.e. how many linear pieces it requires) and find the required parameters $(a_1, b_1, a_2, b_2, ...)$. •

2.4 Approximate Linear Model

By looking at the data table or the data graph, we may realize that the data do not fit to a linear model. It is always possible to use a piecewise linear model to model the relationship of x and y; if nothing else works, we can connect every pair of consecutive points with a straight line. But each straight line used requires two parameters (a and b); if the total number of parameters is comparable to the number of data, then we have not created an "economical" description of our data. In this case, we may decide to use a different approach which requires only one formula $y = ax + b$. However, rather than finding an exact formula, so that every y equals exactly $ax + b$, we find a formula which is always wrong, but *not very wrong*.

But what does *not very wrong mean*? How do we judge if an approximation is good or bad? To understand this, take a specific example. Consider the data of the next table.

Table 2.8

x	y
1	1.80
2	3.20
3	3.80
4	5.20

Let us graph the data in Figure 2.8.

Figure 2.8

It looks rather obvious that no single straight line passes through all the data. A piecewise linear model, would require three straight lines, such as indicated by the dashed line in the graph. This requires a total of six parameters; compared with the eight data, it does not look like an economic model.

Now consider two straight lines that appear to be close to the data points, namely $y = x+1$ and $y = 1.1x + 0.95$. Which one is a "better" model of the data? What does "better" mean?

To evaluate these two lines, consider Table 2.9. For the time being, consider only the first *six* rows of the table. They contain the original x and y data, the *estimates* of the true y values as given by $y = x+1$ and $y = 1.1x + 0.95$ and the error of each estimate. In the final column, sums of errors are presented.

x	y	$x+1$	$y-(x+1)$	$1.1x+0.95$	$y-(1.1x+0.95)$	$[y-(x+1)]^2$	$[y-(1.1x+0.95)]^2$
1	1.80	2.00	-0.20	2.05	0.25	0.04	0.06
2	3.20	3.00	0.20	3.15	-0.05	0.04	0.00
3	3.80	4.00	-0.20	4.25	0.45	0.04	0.20
4	5.20	5.00	0.20	5.35	0.15	0.04	0.02
			0.00		0.80	0.16	0.29

Table 2.9

First, we see that for some points the first line gives smaller errors, for some other points the second. Hence we cannot compare by using one error at a time. Let us use the *total error*; this shown in the final column of the table. Using the total error, it appears that the first line ($y = x+1$) is better than the second, since 0.00 is less than 0.80. But something is wrong here! It appears that the first line has zero total error. This is certainly not true; it is just the result of positive and negative errors cancelling each other. We must find a way for making all errors positive. A convenient way for doing this, is by taking the *square errors*. This is done in the last two columns of Table 2.9; at the last row we show the *total square errors*. Using the criterion of the total square error, it appears that the first line is better than the second: $0.16 < 0.29$.

The conclusion is that we can use total square error as a criterion of how well a straight line (or other model) fits our data. This still does not tell us which line is the *best* according to this criterion; we have just compared *two* possible lines to use. Fortunately, there is a method to find the best of all possible lines: the method of *least squares* or *linear regression*. This is now described.

To describe the method of least squares, we need the following notation (compare with Table 2.9). When I write $\sum x$ I mean the sum of all x values: $\sum x = 1+2+3+4$. Similarly, $\sum y = 1.80 + 3.20 + 3.80 + 5.20$ and so on for $\sum xy$, $\sum x^2$ etc. Keeping this notation in mind, and in addition calling N the total number of data points, we have the following formulas.

$$a = \frac{N \cdot (\sum xy) - (\sum x) \cdot (\sum y)}{N \cdot (\sum x^2) - (\sum x)^2}. \tag{3}$$

$$b = \frac{(\sum y) - a \cdot (\sum x)}{N}. \tag{4}$$

These formulas give the a and b parameters for the straight line that approximates the data set with *least total square error*.

In summary, to find an approximate linear model for a data set, we use the method of linear regression and take the following steps.

1. As usual, first we graph the data.

2. If the data appear to be approximately on a straight line, we apply formulas (3) and (4) to find a and b.

3. Using a and b we compute the error for each point, the square error for each point and the total square error.

4. If the total square error is small, than we have a good approximate linear model for our data.

Example Apply the method of least squares to find the line that best approximates the data set of Table 2.9.
Solution We collect the relevant information in the following table. For the time being, consider only the first four rows of the table.

x	y	x^2	xy	$\hat{y} = ax + b$	$[y - \hat{y}]^2$
1	1.80	1.00	1.80	1.88	0.0064
2	3.20	4.00	6.40	2.96	0.0576
3	3.80	9.00	11.40	4.04	0.0576
4	5.20	16.00	20.80	5.12	0.0064
10	14.00	30.00	40.40		0.1300

Now write the necessary sums out: $\sum x = 10.00$, $\sum y = 14.00$, $\sum x^2 = 30.00$, $\sum xy = 40.40$; also recall that $N = 4$. Substitute these values in the a and b equations, that is (3) and (4) to get

$$a = \frac{4.00 \cdot 40.40 - 10.00 \cdot 14.00}{4.00 \cdot 30.00 - (10.00)^2} = 1.08$$

$$b = \frac{14.00 - 1.08 \cdot 10.00}{4.00} = 0.80.$$

In other words $\hat{y} = ax + b = 1.08 \cdot x + 0.80$ is the straight line of least total square error. The expression $\hat{y}$ is used, rather than y, because the numbers we get from $1.08 \cdot x + 0.80$ are not exactly equal to the y-numbers. Now we can complete the last two rows of Table 2.10 and actually compute the total square error: it is 0.13, which is smaller than the errors of the two linespreviously discussed (for $y = x + 1$, total square error is 0.16; for $y = 1.1 \cdot x + 0.95$, total square error is 0.21). Finally, we can graph our data and the least squares line in Figure 2.8. •

Figure 2.8

Regarding the method of least squares, the following points are worth noting.

1. If the data are exactly on a straight line, then the method of least squares still works and gives the true a and b.

2. From the plot of the data it may appear more likely that *two* straight lines approximate the data; in this case we split the table into two subtables and apply the method of least squares separately to each table.

3. The method of least squares requires considerable computation. Therefore it is most suitable in case a computer is available.

2.4.1 Predict ACT Enrollment

The number of students enrolled at ACT changes every semester. In the following table are presented the numbers for every semesetr starting with Fall 1991 and going up to Fall 1996.

Semester	Nr. of Students
1991 Fall	326
1992 Spring	326
1992 Fall	366
1993 Spring	381
1993 Fall	398
1994 Spring	401
1994 Fall	475
1995 Spring	490
1995 Fall	552
1996 Spring	548

Practically all of ACT's semester revenue comes from tuition, which is approximately the number of students enrolled, multiplied by 500,000 drs (tuition for one student). The budget of 1997-98 will be done in April 1997, and obviously requires to know the revenue avauilable for this year. So it is required to *predict* enrollment for the semesters Fall 1997 and Spring 1998. How can we do this on the basis of the above data?

What is required is to find a model $\hat{y} = f(t)$ and use $t = 1997$ to find $\hat{y} = f(1997)$. Let us, as usual, first graph the data; we get the following figure.

Figure 2.9

These data do not fall clearly on a straight line. However, since we have not studied other types of models so far, let us try to fit a straight line model anyway. This model will be of the form $\hat{y} = at + b$, where $\hat{y}$ is *estimated* enrollment and t is time. It is clear what are the e values, but what should we take as t? This is a decision you will have to make for yourselves. Once you get your data straight, you must use equations (3) and (4) to compute a and b. If you get these, you can substitute the appropriate t values to get the respective $\hat{y}$'s and from these compute the expected revenue.

2.5 Problems

1. Find the straight line that passes through points (2,3), (6,4). Graph.

2. Find the straight line that passes through points (1.50,4), (3.50,-1). Graph.

3. Find a straight line that fits the following data. Graph.

x	0	2	3	6
y	3	7	9	15

4. Find a straight line that fits the following data. Graph.

x	1.10	2.40	3.20	4.30
y	0.62	1.05	1.32	1.68

5. Find a piecewise linear model for the following data. Graph.

x	1	2	3	4	5	6	7	8
y	-2	-1	0	1	2	1	0	-1

6. Find a piecewise linear model for the following data. Graph.

x	1	2	3	4	5	6	7	8
y	-1.5	-0.5	0.5	1.5	1.5	0.5	-0.5	-1.5

7. Find a least squares line that fits the following data. Graph.

x	1	2	3	4
y	-1.1	-1.9	-3.2	-3.9

8. Find a least squares line that fits the following data. Graph.

x	-1	1	2	4
y	0.1	1.8	3.5	4.9

9. Find a linear model (exact or approximate; linear or piecewise linear) that fits the following data. Graph.

x	1.0	2.0	3.0	4.0	5.0
y	1.1	2.1	3.1	3.9	4.9

10. Find a linear model (exact or approximate; linear or piecewise linear) that fits the following data. Graph.

x	0.00	1.00	4.00	7.00
y	-2.42	-1.45	1.54	4.55

11. Alex starts from point A and travels in a straight line with speed of 10 m/sec. Basil starts from same point A 40 seconds later and travels in the same direction as Alex, with a speed of 12 m/sec. Write the positions of Alex and Basil as functions of time, graph them and find the time and position where they meet.

12. Alex starts from point A and travels in a straight line with speed of 20 m/sec. Basil starts from same point A 30 seconds later and travels in the same direction as Alex, with a speed of 18 m/sec. Write the positions of Alex and Basil as functions of time, graph them and find the time and position where they meet.

13. A clothing company produces a type of shirt; each shirt costs 2000 drs to produce. In addition the company has fixed costs equal to 500000 drs per month. What is the cost of producing 100 shirts in a month? 200 shirts? What is the cost of producing x shirts in a month? Graph monthly costs as a function of shirts produced.

14. Continuing from the previous problem, suppose that each shirt can be sold at a price of 3000 drs. How many shirts must the company produce in a month to *break even*, i.e. to have income equal to its costs? Make a graph that illustrates the situation.

15. The following table lists the population of the United States for the year 1986-1989. What will the population be in year 2000? The solution may be obtained using a least squares line. The calculations however can be quite lengthy. Can you think of a way to simplify the claculations?

x	1986	1987	1988	1989
y	241	244	245	248

16. The following table lists the enrollment of freshman class at ACT for the years 1990-1996. What will the enrollment be in year 2000? The solution may be obtained using a least squares line. The calculations however can be quite lengthy. Can you think of a way to simplify the calculations?

x	1991	1991.5	1992	1992.5	1993	1993.5	1994	1994.5	1995	1995.5	1996
y	326	326	366	381	398	401	475	490	552	548	598

17. The following table lists observed costs and revenue of a company per units sold. Use this information to perform a *break- even analysis*.

Units (in thousands)	1	2	3	4	5	6	7
Cost (in million drs.)	1.010	1.250	1.400	1.650	1.840	2.000	2.100
Revenue(in million drs.)	0.700	1.500	2.400	3.300	4.300	5.300	6.350

Also do book problems 2.3.1-.4, 2.3.7-.9, 2.3.11-.28, 2.3.59-.60, 4.1.12-.17, 4.1.21-.30, 4.1.41-.45.

3 Quadratic Models

We now move on to a new family of models, which make use of the *quadratic* function $y(x) = ax^2 + bx + c$. This function has three parameters (a, b, c), as compared to the two parameters (a, b) of the linear function $y = ax + b$. Hence the quadratic offers more opportunities for modelling. Let us first recall its basic algebraic properties.

3.1 Quadratic Functions

There are three equivalent ways of writing the quadratic function .

1. The first form is

$$y(x) = ax^2 + bx + c, \tag{1}$$

   which is the most natural one to write a formula involving x^2, x and a constant term. The graph of $y(x)$ has one of two forms of Figure 3.1, depending on the sign of a. This shape is called the *parabola*.

   Figure 3.1

   Example $y(x) = 2x^2 - 6x + 4$; $a = 2, b = -6, c = 4$. •

2. The following form is of interest when we are considering the solutions to the equation $ax^2 + bx + c$. The following equality holds

$$y(x) = ax^2 + bx + c == a \cdot (x - x_1) \cdot (x - x_2). \tag{2}$$

   Hence, $ax^2 + bx + c = 0$ is equivalent to $a \cdot (x - x_1) \cdot (x - x_2) = 0$, which is true when either $x = x_1$ or $x = x_2$. Hence x_1, x_2 are called the *roots* (i.e. solutions) of the equation $ax^2 + bx + c = 0$. The x_1, x_2 are related to a, b, c as follows.

$$x_1 = \frac{-b - \sqrt{D}}{2a}, x_2 = \frac{-b + \sqrt{D}}{2a}, \tag{3}$$

where the quantity $D = b^2 - 4ac$ is the *discriminant*, a very important quantity that characterizes ("discriminates") the behavior of the quadratic. In particular, the representation of eq.(3) is possible only when $D \geq 0$, so that $\sqrt{D}$ is an acceptable expression. The equation $ax^2 + bx + c = 0$ can be solved if and only if $D \geq 0$. Geometrically, the graph of $y(x) = ax^2 + bx + c$ will cross the x-axis at points x_1, x_2; these can be two distinct points (when $D > 0$, or a single point (when $D = 0$), or a non-existent point (i.e. the graph does not cross the axis), when $D < 0$. The situation is illustrated in Figure 3.2, for $a > 0$ (a similar situation holds for $a < 0$, but with the parabola reversed).

Figure 3.2

Example Write $y(x) = 2x^2 - 6x + 4$ in the form $a \cdot (x - x_1) \cdot (x - x_2)$.

Solution $a = 2, b = -6, c = 4$; $D = (-6)^2 - 4 \cdot 2 \cdot 4 = 36 - 32 = 4$. Then, $x_1 = \frac{-(-6)-\sqrt{4}}{2 \cdot 2} = \frac{6-2}{4} = 1$, and $x_2 = \frac{-(-6)+\sqrt{4}}{2 \cdot 2} = \frac{6+2}{4} = 2$. Hence

$$y(x) = 2x^2 - 6x + 4 = 2 \cdot (x - 1) \cdot (x - 2)$$

and $2x^2 - 6x + 4 = 0$ has solutions $x_1 = 1$, $x_2 = 2$.

3. A final form of the quadratic function is

$$y(x) = ax^2 + bx + c = B + a \cdot (x - A)^2. \tag{4}$$

where A, B are related to a, b, c by

$$A = -\frac{b}{2a}, B = \frac{4ac - b^2}{4a}. \tag{5}$$

This form tells us something about the maximum and minimum value of $y(x)$. Take the case $a > 0$. Then $a \cdot (x - A)^2$ is a nonegative quantity and

$$y(x) = B + a \cdot (x - A)^2 \geq B. \tag{6}$$

In other words, $y(x)$ is never smaller than B. The minimum value of $y(x)$ is attained when the term $a \cdot (x - A)^2$ equals zero (this happens for $x = A$) and is B. •

Geometrically, the minimum value of $y(x)$ corresponds to the *tip* of the graph of $y(x)$; this is attained for $x = A = \frac{b}{2a}$. Hence the form $y(x) = B + a \cdot (x - A)^2$ gives us the x and y coordinates of the tip of the parabola. The situation is illustrated in the Figure 3.3 (for $a > 0$; a similar situation holds for $a < 0$, but with the parabola reversed).

Figure 3.3

Example Write $y(x) = 2x^2 - 6x + 4$ in the form $B + a \cdot (x - A)^2$.

Solution $a = 2, b = -6, c = 4$; $D = (-6)^2 - 4 \cdot 2 \cdot 4 = 36 - 32 = 4$. Then, $A = -\frac{-6}{2 \cdot 2} = -\frac{-6}{4} = \frac{3}{2}$, and $B = \frac{4 \cdot 2 \cdot 4 - (-6)^2}{4 \cdot 2} = \frac{-4}{8} = -\frac{1}{2}$. Hence

$$y(x) = 2x^2 - 6x + 4 = -\frac{1}{2} + 2 \cdot (x - \frac{3}{2})^2 \geq -\frac{1}{2}.$$

It follows that $y(x) \geq -\frac{1}{2}$, and $y_{min} = -\frac{1}{2}$ is attained for $x = \frac{3}{2}$. •

In the case of a linear function we have computed the rate of change $\frac{\Delta y}{\Delta x}$ and found it to be equal to a. Let us now perform a similar calculation for the quadratic function. For x changing from x_1 to $x_2 = x_1 + h$, $\Delta x = h$. We also have $y(x_1) = ax_1^2 + bx_1 + c$ and $y(x_2) = y(x_1 + h) = a \cdot (x_1 + h)^2 + b \cdot (x_1 + h) + c$. Hence

$$\frac{\Delta y}{\Delta x} = \frac{y(x_1 + h) - y(x_1)}{h} = \frac{(a \cdot (x_1 + h)^2 + b \cdot (x_1 + h) + c) - (ax_1^2 + bx_1 + c)}{h} =$$

$$\frac{a \cdot (x_1^2 + 2x_1 h + h^2) + b \cdot (x_1 + h) + c - a \cdot x_1^2 - b \cdot x_1 - c}{h} = \frac{2a \cdot x_1 h + b \cdot h}{h} = 2ax_1 + b + h. \quad (7)$$

In other words $\frac{\Delta y}{\Delta x} = 2ax_0 + b + h$ is the rate of change of $y(x)$. There is a significant difference between the linear and quadratic case. For the linear, the rate of change is always the same, namely a. For the quadratic the rate of change $2ax_1 + h$ depends on both x_1 and the increment h.

As was the case with linear models, quadratic models can also be used to model a data-set. If we think of data as points on the $x - y$ plane, the task of finding a model is equivalent to finding a parabola that passes through the points. We can use

1. a single parabola, which fits all the points exactly;
2. several parabolas, each of which fits exactly a subset of the data;
3. a parabola that fits the data approximately.

This is exactly analogous to the linear case. We will now examine the first and third case in more detail. The second case ("piecewise quadratic models") is similar to the linear case and will not be discussed.

3.2 Exact Quadratic Models

Let us now return to the problem of fitting a quadratic function to some data. In the case of the linear model, two data points (namely two (x, y) pairs) are necessary and sufficient to find the parameters a, b: each point yields an equation to end up with a system of two equations in two unknowns (a and b). Geometrically, this corresponds to the fact that a straight line is determined by two points.

Similarly, in the case of the quadratic model, *three* data points (namely three (x, y) pairs) are necessary and sufficient to find the parameters a, b, c: each point yields an equation to end up with a system of three equations in three unknowns. Geometrically, this corresponds to the fact that a parabola is determined by three points.

This means that to find a quadratic model we must solve a system of three equations. This is generally more difficult than a system of two equations and will require a discussion of systems of equations in general. However, before starting this discussion, let us consider a some cases where a quadratic model is relatively easy to find.

Example Fit a quadratic model to the following data.

x	y
0	10
1	0
2	0

Solution Let us first graph the data in Figure 3.4.

Figure 3.4

The data appear to lie on a parabola. When we have two zeros in the y row, it is best to use the quadratic in the form $y(x) = a \cdot (x - x_1)(x - x_2)$. This is because we know that $y(x_1) = 0$ and $y(x_2) = 0$. Hence, in our example $x_1 = 1$, $x_2 = 2$. We have already used two of the three points provided. Now we use the third point as follows.

$$(0, 10): \quad 10 = a \cdot (0 - 1) \cdot (0 - 2) \Rightarrow 10 = a \cdot 2 \Rightarrow a = \frac{10}{2} = 5.$$

Hence the quadratic function is of the form $y(x) = 5 \cdot (x - 1) \cdot (x - 2)$. If we want to get it in the form $ax^2 + bx + c$, we just do the multiplications to get $y(x) = 5x^2 - 15x + 10$. •
Example Fit a quadratic model to the following data.

x	y
0	1
2	2
3	4

Solution Let us graph the data in Figure 3.5.

Figure 3.5

If we have one zero in the x row, it is fairly easy to find the quadratic in the form $y(x) = ax^2 + bx + c$. To do this, start with the $x = 0$ point to get

$$(0, 1): \quad 1 = a \cdot 0^2 + b \cdot 0 + c \Rightarrow c = 1.$$

Now use the remaining two points.

$$\left.\begin{array}{rl} (2, 2): \ 2 = & a \cdot 2^2 + b \cdot 2 + 1 \\ (3, 4): \ 4 = & a \cdot 4^2 + b \cdot 4 + 1 \end{array}\right\} \Rightarrow \begin{array}{rl} 4a + 2b = & 1 \\ 16a + 4b = & 3 \end{array}.$$

Solving the linear system we get $a = \frac{1}{8}$, $b = \frac{1}{4}$. So we finally get $y(x) = \frac{1}{8}x^2 + \frac{1}{4}x + 1$. •

In the previous examples we got lucky: we only had to solve a second order system. But in the general case a third order system will be involved, as in the following example.
Example Fit a quadratic model to the following data.

x	y
1	1
2	3
3	2

Solution We have the following equations:

$$\left.\begin{array}{lrl} (1,1): & 1 = & a\cdot 1^2 + b\cdot 1 + c \\ (2,3): & 3 = & a\cdot 2^2 + b\cdot 2 + c \\ (3,2): & 2 = & a\cdot 3^2 + b\cdot 3 + c \end{array}\right\} \Rightarrow \begin{array}{rl} a+b+c = & 1 \\ 4a+2b+c = & 3 \\ 9a+3b+c = & 2 \end{array}.$$

We will discuss methods of solving this system in the next section. For now, just suppose that it has been solved to obtain $a = -3/2$, $b = 13/2$, $c = -4$. Hence the required quadratic model is $y(x) = -\frac{3}{2}x^2 + \frac{13}{2}x - 4$ and the solution is complete. •

There is another way in which the previous examples are simplified: we have not had to select the points to use for finding a, b and c. This is because we were always given three points, which is the minimum required. But in general we will have more than three points, and part of the problem is to find the ones that yield the solution most easily.

Example Fit a quadratic model to the following data.

x	y
-2	-15
-1	-8
0	-3
1	0
2	1
3	0
4	-3

Solution Let us graph the data in Figure 3.6.

Figure 3.6

This is a case where two zeros are given in the y row, so we use the form $y(x) = a \cdot (x - x)1) \cdot (x - x_2)$. Hence, in our example $x_1 = 1$, $x_2 = 3$. We have already used two points, namely (1,0) and (3,0). Now we use the third point, (0,−3) as follows.

$$(0,-3): \quad -3 = a\cdot(0-1)\cdot(0-3) \Rightarrow -3 = a\cdot 3 \Rightarrow a = \frac{-3}{3} = -1.$$

Hence the quadratic function is of the form $y(x) = -(x-1)\cdot(x-3)$. If we want to get it in the form ax^2+bx+c, we just do the multiplications to get $y(x) = -x^2+4x-3$. •
Example Fit a quadratic model to the following data.

x	y
-2	-1.25
-1	0.75
0	0.75
1	-1.25
2	-5.25
3	-11.25
4	-19.25

Solution Let us graph the data in Figure 3.7.

Figure 3.7

This is a case where one zero is given in the x column, so we use the form $y(x) = a\cdot x^2 + b\cdot x + c$. It is fairly easy to find the quadratic in the form $y(x) = ax^2+bx+c$. Start with the $x=0$ point to get

$$(0, 0.75): \quad 0.75 = a\cdot 0^2 + b\cdot 0 + c \Rightarrow c = 0.75.$$

Now use the remaining two points.

$$\left.\begin{array}{lrr} (-1,0.75): & 0.75 = & a\cdot(-1)^2 + b\cdot(-1) + 0.75 \\ (1,-1.25): & -1.25 = & a\cdot 1^2 + b\cdot 1 + 0.75 \end{array}\right\} \Rightarrow \begin{array}{rr} a-b= & 0 \\ a+b= & -2 \end{array}.$$

Solving the linear system we get $a=-1$, $b=-1$. So we finally get $y(x) = -x^2 - x + 0.75$. •

3.2.1 Profit Maximization

Let us now consider an application of quadratic modelling. Consider a factory producing *googles*. The first issue that we will consider is the relationship between the price of googles and the amount sold.

Market observations of the quantity q of googles sold, depending on the price p, are described by Table 3.8.

p	q
0	1000
100	800
200	600
500	0

Table 3.8

The table reflects the fact that when the price is lowered, more googles are sold. Graphing the data, they appear to be on a straight line (see Figure 3.8; hence we expect a relationship between p and q of the form $q(p) = ap + b$. Solving by the usual methods we find $a = -2$, $b = 1000$ and so $q(p) = -2p + 1000$. This equation reflects the law of *supply and demand*.

Figure 3.8

The law of supply and demand can be put in a different form, where the independent variable is quantity q and the dependent variable is price p. In this case, the idea is that a quantity of googles is thrown in the market, and price is determined by market forces. If we accept this interpretation, we need a graph of the following form. The data again appear to lie on a straight line (naturally!) of the form $p(q) = ap + b$. Now solving for a and b we obtain $p(q) = -\frac{1}{2}q + 500$.

The target of the google maker is to maximize profit. Now, profit is the difference between *revenue* (income from google sales) and *cost*. Both of these are functions of the quantity q produced. We use $P(q)$ to denote profit, $R(q)$ to denote revenue and $C(q)$ to denote cost. We have

$$P(q) = R(q) - C(q). \tag{8}$$

$R(q) = q \cdot p(q) = q \cdot (-\frac{1}{2}q + 500)$, as already computed. The google maker knows that his cost is given by $C(q) = 10000 + 100q$, where 10000 is *fixed* cost (independent of level of production) and

$100q$ is variable cost (say that raw materials etc. cost 100 drs per google). Finally, the profit is given by

$$P(q) = q \cdot (-\frac{1}{2}q + 500) - (10000 + 100q) = -\frac{1}{2}q^2 + 400q - 10000. \tag{9}$$

This is a quadratic function; it is graphed in Figure 3.9.

Figure 3.9

We will rewrite $P(q)$ using the form $B + a \cdot (x - A)^2$. In this case, x is replaced by q; $a = -\frac{1}{2}$, $b = 400$, $c = 10000$; $A = -\frac{b}{2a} = -\frac{400}{2\cdot(-1/2)} = 400$ and $B = \frac{4ac-b^2}{4a} = \frac{4\cdot(-\frac{1}{2})\cdot 10000-(400)^2}{4\cdot(-\frac{1}{2})} = 70000$. So we finally have

$$P(q) = 70000 - \frac{1}{2} \cdot (q - 400)^2 \leq 70000. \tag{10}$$

It follows that the profit attains a maximum value of 70000 drs, when level of production is q=400 googles.

To practice the computation of rate of change, let us determine $\frac{\Delta P}{\Delta q}$, $\frac{\Delta R}{\Delta q}$, $\frac{\Delta C}{\Delta q}$. We have

$$\frac{\Delta P}{\Delta q} = \frac{P(q+h) - P(q)}{h} =$$

$$\frac{(-\frac{1}{2} \cdot (q+h)^2 + 400 \cdot (q+h) - 10000) - (-\frac{1}{2} \cdot q^2 + 400 \cdot q - 10000)}{h} =$$

$$\frac{(-\frac{1}{2} \cdot (q^2 + 2qh + h) + 400 \cdot (q+h) - 10000) - (-\frac{1}{2} \cdot q^2 + 400 \cdot q - 10000)}{h} =$$

$$\frac{(-\frac{1}{2} \cdot (2qh + h) + 400 \cdot h}{h} = -q + 400 + h.$$

Similarly we compute

$$\frac{\Delta R}{\Delta q} = -q + 500 + h, \quad \frac{\Delta C}{\Delta q} = 100. \tag{11}$$

$\frac{\Delta P}{\Delta q}$ can be interpreted as follows. suppose that currently q googles are being produced and now it is decided to produced an additional h googles: the *average* change of profit is $\frac{\Delta P}{\Delta q} = -q + 400 + h$, in other words the change of profit *per google*. This depends both on the current level of production q and the contemplated increase h. Similarly, $\frac{\Delta R}{\Delta q} = -q + 500 + h$ is the change in revenue *per google* for a contemplated h increase above current level of production q. And, $\frac{\Delta R}{\Delta q} = 100$ is the change in cost *per google* for a contemplated h increase above current level of production q.

Now, since these are *average* rates of change, and the average is being taken over h googles, it would be nice if the rates of change were independent of the actual increment h. We can approximate this situation, if we take h to be very small, in fact *tending to zero*. E.g. take h to be 0.1, or 0.01 etc. Then, *in the limit* we get $\frac{\Delta P}{\Delta q} = -q + 400$, $\frac{\Delta R}{\Delta q} = -q + 500$ and $\frac{\Delta R}{\Delta q} = 100$. Of course one can not produce 0.01 google (they only come in whole numbers), but taking h very small approximates the rate of change over very small changes. Now, the rates of change still depend on the current level of production q. For some values of q, an increase in production yields a higher increase in revenue, than in profit. E.g., at $q = 100$, $\frac{\Delta R}{\Delta q} = -100 + 500 = 400$, and $\frac{\Delta R}{\Delta q} = 100$. While an increase in production increases both revenue *and* cost, the increase in revenue is bigger than that of cost ($400 > 100$). Hence, it makes sense to increase production to more than 100 googles. On the other hand, for $q = 4500$, $\frac{\Delta R}{\Delta q} = -450 + 500 = 50$, and $\frac{\Delta R}{\Delta q} = 100$. While an increase in production increases both revenue *and* cost, the increase in revenue is smaller than that of cost ($400 > 100$). Hence, it makes sense to decrease production to less than 450 googles.

You may have guessed it already: there is a level of production q, where increases in revenue are exactly balanced by increases in cost. At this level it does not make sense to either increase q or decrease q (why?!). To get this q, we must have $\frac{\Delta R}{\Delta q} = \frac{\Delta R}{\Delta q}$. In our example, this level is given by $\frac{\Delta R}{\Delta q} = -q + 500 = \frac{\Delta C}{\Delta q} = 100$, i.e. $-q + 500 = 100 \Rightarrow q = 400$, which is exactly the level previously computed.

3.3 Interlude: Systems of Linear Equations

In this section we discuss the solution of *systems of linear equations*. By this we mean several equations of the form $ax + by + ... = c$, which must be satisfied simultaneously. Here is an example, which originated in the previous section.

$$\begin{array}{rrrl} x & +y & +z & = 1 \\ 4x & +2y & +z & = 3 \\ 9x & +3y & +z & = 2 \end{array} .$$

Here we have three equations and three unknowns (x, y, z); we call this a *third order* linear system. In this particular example, the choices $x = -3/2$, $y = 13/2$ and $z = -4$ satisfy all three equations. Hence the solution is $(x, y, z) = (-3/2, 13/2. - 4)$. Note that *one* solution consists of *three* numbers!.

How did we find the particular x, y, z values? Several methods are available; we will discuss one in particular, which makes use of *determinants*; these must now be defined.

A determinant is a number, associated with a square *table* of values. For instance, for the table

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

the determinant is -13. For the general case of a two-by-two table

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

the determinant is $ad - cb$. For the general case of a three-by-three table

$$\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$$

the determinant is $aei - ahf - bdi + bgf + cdh - ceg$. There is a certain symmetry in the way the parameters $a, b, c, ...$ appear (permutations). If you think you have understood this symmetry, try to find the determinant of four-by-four table.

We will not give the rule for computing determinants of higher order. The reason is that this is a complicated numerical calculation which you will only do by computer, if you ever have to do it. In case you use a computer, the determinant function will probably be built-in in the program you are using (for instance EXCEL). Now, supposing that you know to compute determinants, how are they used ins solving systems of equations?

The method is simple and mechanical, like a recipe. First we give an example involving two equations and two unknowns.

Example Solve the system

$$\begin{array}{lll} 2x & +3y & = 4 \\ 5x & +y & = 10 \end{array}.$$

Solution First form the table of coefficients (it has two rows and *three* columns).

$$\begin{bmatrix} 2 & 3 & 4 \\ 5 & 1 & 10 \end{bmatrix}.$$

Next form *three* tables of two rows and two columns each, as follows

$$\begin{bmatrix} 2 & 3 \\ 5 & 1 \end{bmatrix}, \quad \begin{bmatrix} 4 & 3 \\ 10 & 1 \end{bmatrix}, \quad \begin{bmatrix} 2 & 4 \\ 5 & 10 \end{bmatrix}.$$

The first table includes the coefficients of the unknowns; the second table is identical to the first, except that the column of x-coefficients is replaced by the column of constant terms; the third table is identical to the first, except that the column of y-coefficients is replaced by the column of constant terms.

Second, compute the determinant of each table: $D = 2 \cdot 1 - 5 \cdot 3 = 2 - 15 = -13$; $D_x = 4 \cdot 1 - 10 \cdot 3 = 4 - 30 = -26$; $D_y = 2 \cdot 10 - 5 \cdot 4 = 20 - 20 = 0$.

Third set $x = \frac{D_x}{D} = \frac{-26}{-13} = 2$, $y = \frac{D_y}{D} = \frac{0}{-13} = 0$.

Finally, check that $x = 2$, $y = 0$ satisifies the original equations:

$$\begin{array}{lll} 2 \cdot 2 & +3 \cdot 0 & = 4 \\ 5 \cdot 2 & +0 & = 10 \end{array}.$$

And we are done. •

Next we consider an example with three equations and three unknowns, the one that originated in the problem of quadratic fit.

Example Solve the system

$$\begin{array}{llll} x & +y & +z & = 1 \\ 4x & +2y & +z & = 3 \\ 9x & +3y & +z & = 2 \end{array} .$$

Solution The tables and determinants required are

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 4 & 2 & 1 & 3 \\ 9 & 3 & 1 & 2 \end{bmatrix}.$$

$$\begin{bmatrix} 1 & 1 & 1 \\ 4 & 2 & 1 \\ 9 & 3 & 1 \end{bmatrix}$$

$$D = +1 \cdot 2 \cdot 1 - 1 \cdot 3 \cdot 1 - 1 \cdot 4 \cdot 1 + 1 \cdot 9 \cdot 1 + 1 \cdot 4 \cdot 3 - 1 \cdot 9 \cdot 2 = 2 - 3 - 4 + 9 + 12 - 18 = -2.$$

$$\begin{bmatrix} 1 & 1 & 1 \\ 3 & 2 & 1 \\ 2 & 3 & 1 \end{bmatrix}$$

$$D_x = +1 \cdot 2 \cdot 1 - 1 \cdot 3 \cdot 1 - 1 \cdot 3 \cdot 1 + 1 \cdot 2 \cdot 1 + 1 \cdot 3 \cdot 3 - 1 \cdot 2 \cdot 2 = 2 - 3 - 3 + 2 + 9 - 4 = 3.$$

$$\begin{bmatrix} 1 & 1 & 1 \\ 4 & 3 & 1 \\ 9 & 2 & 1 \end{bmatrix}$$

$$D_y = +1 \cdot 3 \cdot 1 - 1 \cdot 2 \cdot 1 - 1 \cdot 4 \cdot 1 + 1 \cdot 9 \cdot 1 + 1 \cdot 4 \cdot 2 - 1 \cdot 9 \cdot 3 = 3 - 2 - 4 + 9 + 8 - 27 = -13.$$

$$\begin{bmatrix} 1 & 1 & 1 \\ 4 & 2 & 3 \\ 9 & 3 & 2 \end{bmatrix}$$

$$D_z = +1 \cdot 2 \cdot 2 - 1 \cdot 3 \cdot 3 - 1 \cdot 4 \cdot 2 + 1 \cdot 9 \cdot 3 + 1 \cdot 4 \cdot 3 - 1 \cdot 9 \cdot 2 = 4 - 9 - 8 + 27 + 12 - 18 = 8.$$

Then $x = \frac{D_x}{D} = -\frac{3}{2}$, $y = \frac{D_y}{D} = \frac{13}{2}$, $z = \frac{D_z}{D} = -\frac{8}{2} = 4$, as announced in previous section. •

By now you have probably got the idea; let us describe the method for solving linear systems of any order using determinants (incidentally, this is called *Cramer's Rule*).

Cramer's Rule

To solve a system of N equations and N unknowns (call them x_1, x_2, ... , x_N you must compute $N+1$ determinants. The following steps are used.

1. Arrange the parameters of the equations (without the constant terms) in a table of N rows, N columns. Compute the determinant of the table, call it D.

2. In the original table, substitute the first column with the column of constant terms. Compute the determinant, call it D_1.

3. In the original table, substitute the second column with the column of constant terms. Compute the determinant, call it D_2.

4. Continue with all remaining columns of the original table.

5. Compute the solution by $x_1 = \frac{D_1}{D}$, $x_2 = \frac{D_2}{D}$, ... , $x_N = \frac{D_N}{D}$.

It must be pointed out that Cramer's rule works only in case that the original determinant D is different from zero. In this course we will only deal with this case; when $D = 0$, there are other methods to solve a system of linear equations. However, to get an idea of when this situation might arise, as well as to practice determinants computation, try the following example and try to find what goes wrong.

Example Solve the system

$$\begin{array}{rrrl} 2x & +3y & -z & = 3 \\ 4x & +6 & -2z & = 6 \\ -x & -2y & +z & = 1 \end{array} .$$

Solution Do it yourself!!! •

3.4 Approximate Quadratic Model

Let us now turn to the case of approximate quadratic models. The situation is exactly analogous to that of approximate linear models. In other words, we have a number of data points and each one gives an equation to be satisfied by the parameters a, b, c. If there is more than three data points, we have more equations than unknowns and, generally, not all equations will be satisfied exactly. Hence we try to find the a, b, c values which minimize total square error. It turns out that there are values of a, b, c which achieve this; just like in the linear case, we will give equations for a, b, c (without, for the time being) explaining their derivation.

However, rather than giving explicit formulas for a, b, c (which we did for the linear model) we will give a system of three linear equations which a, b, c must satisfy. Since we have three equations and three unknowns, the system can be solved exactly, by use of Cramer's rule. The equations are as follows

$$\begin{array}{rrrl} Nc & +(\sum x)b & +(\sum x^2)a & = (\sum y) \\ (\sum x)c & +(\sum x^2)b & +(\sum x^3)a & = (\sum xy) \\ (\sum x^2)c & +(\sum x^3)b & +(\sum x^4)a & = (\sum x^2 y) \end{array} .$$

At this point it is worthwhile to also give the equations for a, b in the linear model. These equation, when solved give the formulas for a, b that we used in the previous chapter. While the final formulas are easier to use, the equations show more clearly the analogy between the linear and quadratic model.

$$\begin{array}{rrl} Nb & +(\sum x)a & = (\sum y) \\ (\sum x)b & +(\sum x^2)a & = (\sum xy) \end{array} .$$

Do you see the analogy? The equations for the quadratic model actually *contain* the ones for the linear model (in the upper right corner). If you forget the names a, b, c and think in terms of the coefficient of x and the constant term, you realize that the equations are the same for both linear and quadratic model. This is actually a deep observation and will be discussed again in the following chapter.

Example Find a quadratic model for the following data.

x	y
-3	20.1
-2	11.9
-1	6.2
1	-2.2
2	0.1
3	2.1

Solution Let us graph the data in Figure 3.10.

Figure 3.10

We complete the x, y table as follows

x	y	x^2	x^3	x^4	xy	x^2y
-3	20.1	9	-27	81	-60.3	180.09
-2	11.9	4	-8	16	-23.8	47.6
-1	6.2	1	-1	1	-6.2	6.2
1	-2.2	1	1	1	-2.2	-2.2
2	0.1	4	8	16	0.2	0.4
3	2.1	9	27	81	6.3	18.9
0	38.2	28	0	196	-86	251.8

From this table we write the equations of the regression

$$\begin{array}{rrrl} 6c & +0b & +28a & = 38.2 \\ 0c & +28b & +0a & = -86. \\ 28c & +0b & +196a & = 251.8 \end{array} .$$

This system can still be solved by hand, because we can find b immediately, but the prospect is not very pleasing. Let us solve using Cramer's rule, where the determinants have been computed by EXCEL: $D = -10976$, $D_a = -12353.6$, $D_b = 33712$, $D_c = -12230.4$; $a = 1.12551$, $b = -3.07143$, $c = 1.114286$; the approximate quadratic fit (of least square error) is $y(x) = 1.12551 \cdot x^2 - 3.07143 \cdot x + 1.114286$. •

3.4.1 Predict ACT Enrollment

We have already seen the data of ACT enrollment by. It is reproduced below.

Semester	Nr. of Students
1991 Fall	326
1992 Spring	326
1992 Fall	366
1993 Spring	381
1993 Fall	398
1994 Spring	401
1994 Fall	475
1995 Spring	490
1995 Fall	552
1996 Spring	548

We have discussed the reasons why enrollment prediction is of interest. In the previous chapter you performed a linear regression on these data. Now try to find a quadratic regression of the form $e = at^2 + bt + c$, using the method of the previous section. Compare the two predictions. Which one would you use?

3.5 Systems of Inequalities

In this section we present a method for solving inequalities such as

$$x - 3 < 0,$$

$$x^2 - 3x + 2 \geq 0,$$

$$\frac{x-3}{x+1} > 0$$

etc. The main tools used are: (a) factorization and (b) a table to summarize information about factors.

It is well known how to solve a *linear* inequality, such as $x-3 < 0$. The solution is immediately found to be $x < 3$, or, in *interval notation* $x \in (-\infty, 3)$. Note that the solution is an interval (or more) of numbers, rather than just one number. This is a difference between inequalities and equations.

Let us next solve $x^2 - 3x + 2 \geq 0$. The first step is to *factorize* $x^2 - 3x + 2$, i.e. to write as the product of linear terms. We know how to do this: first we compute the roots of $x^2 - 3x + 2 = 0$, which are $x_1 = 1$ and $x_2 = 2$. Then, using the formula $ax^2 + bx + c = a \cdot (x - x_1)(x - x_2)$, we get

$$x^2 - 3x + 2 \geq 0 \Rightarrow 1 \cdot (x-1)(x-2) \geq 0.$$

The point of the factorization, is that it is easy to tell the sign of a linear term for every value of x. Now consider the following table.

x	$-\infty$		1		2		$+\infty$
x-1		$-$	0	$+$	0	$-$	
x-2		$-$	0	$-$	0	$-$	
(x-1)(x-2)		$+$	0	$-$	0	$+$	

Look first at the rows. The first row in the table has x and all "interesting" x values, i.e. values where some of the factors changes sign. Each of the following two rows refers to one of the factors and the final row refers to the product of the factors, in other words to the original epxression of the inequality. Now look at the columns. Each column refers to one interval (as indicated at the top) and gives the sign of each factor; the last row of each column gives the sign of the product of the factors. For instance, in the first column we see that in $(-\infty, 1)$, $x-1$ is negative, $x-2$ is negative and $(x-1)(x-2)$ is positive (because it is the product of two negative factors).

From the last row of the table, we find immediately the solution to $x^2 - 3x + 2 \geq 0$. It is $(-\infty, 1] \cup [2, +\infty)$. Notice that the intervals are *closed* at 1 and 2, because the inequality sign is $\geq$, i.e. the possibility of equality is included. If we had $x^2 - 3x + 2 > 0$, then the solution would be $(-\infty, 1) \cup (2, +\infty)$.

A similar table can be applied to solving the inequality $\frac{x-2}{x+1} \leq 0$. It looks like this.

x	$-\infty$		-1		2		$+\infty$
x-2		$-$	0	$+$	0	$-$	
x+1		$-$	0	$-$	0	$-$	
$\frac{x-2}{x+1}$		$+$	*	$-$	0	$+$	

The first three rows are created in exactly the same manner as in the previous table. As for the final row, *dividing* two negative factors gives a positive number (exactly the same way as *multiplying* two positive factors). In the last row, first of all notice the *: it shows that at $x = -1$, the denominator of the fraction is 0, hence the fraction is not defined. Hence, the value $x = -1$ must not be included in the solution. The solution then is $(-\infty, -1) \cup [2, +\infty)$.

The above method can be extended to inequalities with more factors. For example, to solve $x^3 - 3x^2 + 2x < 0$, we first observe that

$$x^3 - 3x^2 + 2x < 0 \Rightarrow x \cdot (x^2 - 3x + 2) < 0 \Rightarrow x \cdot (x-1) \cdot (x-2) < 0.$$

Now we set up the table of signs as follows:

x	$-\infty$		0		1		2		$+\infty$
x-1		$-$		$+$	0	$+$		$+$	
x		$-$	0	$-$		$+$		$+$	
x-2		$-$		$-$		$-$	0	$+$	
$x(x-1)(x-2)$		$-$	0	$+$	0	$-$	0	$+$	

The solution is $(-\infty, 0] \cup [1, 2]$.

3.6 Problems

1. Solve the following equations: (a) $x^2 - 8x + 12 = 0$, (b) $-2x^2 + 6x - 4 = 0$ (c) $x^2 + x + 1 = 0$, (d) $2x^2 - 16x + 32 = 0$, (e) $x^3 - 5x^2 + 6x = 0$, (f) $-2x^3 + 6x^2 - 4x = 0$.

2. Find the maximum and minimum value of the following functions: (a) $x^2 - 8x + 12 = 0$, (b) $-2x^2 + 6x - 4 = 0$ (c) $x^2 + x + 1 = 0$, (d) $2x^2 - 16x + 32 = 0$.

3. Find the level of production which achieves maximum profit, as well as this maximum profit for the following demand and cost functions: (a) $p(q) = -q + 1000$, $C(q) = 20000 + 200q$, (b) $p(q) = -2q + 1000$, $C(q) = 10000 + 200q$, (c) $p(q) = -q + 2000$, $C(q) = 20000 + 200q$, (d) $p(q) = -5q + 6000$, $C(q) = 20000 + 1000q$.

4. Find an exact quadratic fit to the following data.

x	-3	-2	-1	1	2	3
y	0	-5	-8	-8	-5	0

5. Find an exact quadratic fit to the following data.

x	-2	-1	0	1	2	3
y	3	1	1	3	7	13

6. Find an exact quadratic fit to the following data.

x	1	2	3	4	5	6
y	3	7	13	21	31	43

7. Find a least squares quadratic fit to the following data.

x	-3	-2	-1	1	2	3
y	20.2	11.8	6.1	-2.1	0.15	2.2

8. Find a least squares quadratic fit to the following data.

x	1	2	3	4	5	6
y	2.1	4.9	9.8	17.3	26.1	37.1

9. Solve the following linear system.

$$\begin{array}{rrl} 3x & +2y & = 2 \\ -5x & +y & = -1 \end{array} .$$

10. Solve the following linear system.

$$\begin{array}{rrl} -x & +3y & = 4 \\ 2x & -4y & = 10 \end{array} .$$

11. Solve the following linear system.

$$\begin{array}{rrrl} x & +y & -z & = 2 \\ x & -y & +2z & = 2 \\ -x & +2y & +z & = 2 \end{array} .$$

12. Solve the following linear system.

$$\begin{array}{rrrl} x & +y & -z & = 0 \\ x & -y & +2z & = 5 \\ -x & +2y & +z & = 4 \end{array} .$$

13. Write a linear system of three equations and three unknowns and solve it.

14. Solve the following ineualities.

   (a) $x^2 - 6x + 8 < 0$.

   (b) $x^3 - 6x^2 + 8x \geq 0$.

   (c) $\frac{x^2-6x+8}{x^2-4} \leq 0$.

   (d) $\frac{x^3-3x^2+2x}{x^2-9} \geq 0$.

Also do book problems 1.4.25-.36, 4.2.1-.6, 4.2.25-30, 4.3.3-.4, 4.3.7-.10, 4.3.12, 4.3.40, 4.3.43, 4.4.1.-.5, 4.4.9, 4.4.18-.21, 4.4.48, 4.4.52, 10.5.14-.18, 10.5.33-.41, 1.6.1-.25, 1.6.51-62.

4 Multilinear Models

We have seen so far formulas that give the regression parameters for the linear and quadratic regression. These are special cases of what we call *multiple* regression. Consider first the following example. We have *three input variables*, x_1, x_2, x_3, and an *output variable* y. The data set which consists of several combinations of $x_1, x_2, ..., x_M, y$ values; these can be arranged in a table.

x_1	x_2	x_3	y
1	2	3	5
1	3	1	0
2	1	1	3
2	2	4	8
3	3	2	4
3	4	1	1

We want to find coefficients a_1, a_2, a_3 which connect x_1, x_2 , x_3 and y in the following manner.

$$y = a_1 x_1 + a_2 x_2 + a_3 x_3.$$

In fact, you can check easily that in this case we have $y = x_1 - x_2 + 2x_3$, i.e. $a_1 = 1$, $a_2 = -1$, $a_3 = 2$.

In the previous example the parameters are easy to find and the fit is exact. In general, neither of the above is true.

In the general case we have some *input variables*, call them x_1, x_2, ... , x_M, and an *output variable* y. We aso have a data set which consists of several combinations of $x_1, x_2, ..., x_M, y$ values Now we want to find parameters a_1, a_2, ... , a_M which connect x_1, ... , x_M and y in the following manner.

$$y = a_1 x_1 + a_2 x_2 + ... + a_M x_M.$$

Furthermore, we want to chooce a_1, a_2, ... , a_M so as to minimize *total square error*, i.e. the quantity

$$\sum (y - a_1 x_1 - a_2 x_2 - ... - a_M x_M)^2.$$

This is the goal of this chapter; but before we attain it, we must discuss *matrices*, which are a convenient tool to simplify the development of regression.

Before moving on, note that in the previous example M=3 and N=6. Also note that the linear and quadratic regression are special type of multiple regression. For example, in the linear regression $x_1 = 1$, $x_2 = x$; in the quadratic regression we have $x_1 = 1$, $x_2 = x$, $x_3 = x^2$.

4.1 Prelude: Matrices

The theory of matrices was developed to have a convenient method of performing complicated calculations. A matrix is, simply, a table. It can be written as

$$A = \begin{bmatrix} a_{11} & a_{12} & ... & a_{1N} \\ a_{21} & a_{22} & ... & a_{2N} \\ ... & ... & ... & ... \\ a_{M1} & a_{M2} & ... & a_{MN} \end{bmatrix}.$$

Here M is the number of rows, N is the number of columns and the matrix is of *size* M-by-N.
Example

$$A = \begin{bmatrix} 2 & 7 & -4 & 13 \\ 4 & 1 & 3 & -5 \\ 10 & -87 & 14 & 20 \end{bmatrix};$$

here the matrix A is 3-by-4, i.e. M=3 and $N = 4$; $a_{11} = 2$, $a_{21} = 4$, $a_{12} = 7$ and so on. $\bullet$.

Of particular interest are matrices of size M-by-1 (i.e. a column) and matrices of size 1-by-N (i.e. a row). Such matrices are also called *vectors*. For instance

$$B = \begin{bmatrix} 2 & 7 & -4 & 13 \end{bmatrix}$$

is a *row* vector of size 1-by-4.

Also of interest is a matrix of size 1-by-1, which is simply a number; for instance $C = [3]$. Such matrices are called *scalars*.

Matrices are in many ways similar to numbers; there are also a few significant differences. First of all, for two matrices to be equal, they must be of the same size and every element of the same position is the same for both matrices. I.e.

$$A = B \Rightarrow a_{11} = b_{11},\ \ a_{12} = b_{12},\ \ ...,\ \ a_{MN} = b_{MN}.$$

Generally, we have *matrix operations* just like we have number (or scalar) operations. For instance, matrix addition is defined as follows: if

$$A = \begin{bmatrix} a_{11} & a_{12} & ... & a_{1N} \\ a_{21} & a_{22} & ... & a_{2N} \\ ... & ... & ... & ... \\ a_{M1} & a_{M2} & ... & a_{MN} \end{bmatrix}, B = \begin{bmatrix} b_{11} & b_{12} & ... & b_{1N} \\ b_{21} & b_{22} & ... & b_{2N} \\ ... & ... & ... & ... \\ b_{M1} & b_{M2} & ... & b_{MN} \end{bmatrix},$$

then

$$A + B = \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} & ... & a_{1N} + b_{1N} \\ a_{21} + b_{21} & a_{22} + b_{22} & ... & a_{2N} + b_{2N} \\ ... & ... & ... & ... \\ a_{M1} + b_{M1} & a_{M2} + b_{M2} & ... & a_{MN} + b_{NN} \end{bmatrix},$$

i.e. we add the elements of the two matrices according to their position.
Example

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

then

$$A + B = \begin{bmatrix} 1+2 & 3-1 & -1+6 \\ 3+1 & 2-1 & -3+7 \\ -5+3 & 4+3 & 3+5 \end{bmatrix} = \begin{bmatrix} 3 & 2 & 5 \\ 4 & 1 & 4 \\ -2 & 7 & 8 \end{bmatrix}. \bullet$$

Matrix subtraction is defined in the same manner. Note that both for matrix addition and subtraction to be possible, the matrices we want to add or subtract must have the same size.

When we deal with number addition and subtraction, we have a special number which leaves the result of these operations unchanged. This number is, of course, 0. For example, $a + 0 = a$, $b - 0 = 0$. There is a special matrix like this, too, which we symbolize by $\mathbf{0}$ and has the same property: $A + \mathbf{0} = A$, $B - \mathbf{0} = B$. Actually there are many such matrices, one for each choice of size. When we are talking about M-by-N matrices, then $\mathbf{0}$ is an M-by-N matrix with each element equal to 0. It is obvious that $A+ \mathbf{0}{=}A$.

Next we get to *scalar multiplication*. This operation requires a scalar (i.e. a number), call it c and a matrix, call it A. Then

$$c \cdot A = \begin{bmatrix} c \cdot a_{11} & c \cdot a_{12} & ... & c \cdot a_{1N} \\ c \cdot a_{21} & c \cdot a_{22} & ... & c \cdot a_{2N} \\ ... & ... & ... & ... \\ c \cdot a_{M1} & c \cdot a_{M2} & ... & c \cdot a_{MN} \end{bmatrix}.$$

Example ,

$$2 \cdot \begin{bmatrix} 1 & 3 & -1 \\ 3 & 2 & -3 \\ -5 & 4 & 3 \end{bmatrix} = \begin{bmatrix} 2 & 6 & -2 \\ 6 & 4 & -6 \\ -10 & 8 & 6 \end{bmatrix}. \bullet$$

So far we have not *done* much with matrices. To understand why matrices are useful, consider now *matrix multiplication*. This involves two matrices A and B, which are of sizes L-by-M and M-by-N respectively. That is, the first matrix has the same column size as the second matrix has row size. How are A and B multiplied? To explain this, consider first the multiplication of a matrix by a *column vector*. That is, take $N = 1$. For example, suppose that

$$A = \begin{bmatrix} 3 & 2 & 1 \\ -2 & 4 & -6 \\ -7 & 12 & 10 \end{bmatrix}, B = \begin{bmatrix} 11 \\ 22 \\ 16 \end{bmatrix}.$$

Then

$$A \cdot B = \begin{bmatrix} 3 \cdot 11 + 2 \cdot 22 + 1 \cdot 16 \\ -2 \cdot 11 + 4 \cdot 22 - 6 \cdot 16 \\ -7 \cdot 11 + 12 \cdot 22 + 10 \cdot 16 \end{bmatrix} = \begin{bmatrix} 93 \\ -30 \\ 347 \end{bmatrix}.$$

This may look very strange, but it allows us, first of all, to write some complicated expressions very quickly. For example, consider the system of linear equations

$$\begin{array}{cccc} x_1 & +x_2 & -x_3 & = 2 \\ x_1 & -x_2 & +2x_3 & = 3 \\ -x_1 & +2x_2 & +x_3 & = 5 \end{array}.$$

This can be wriiten very quickly as $Ax = b$, where

$$A = \begin{bmatrix} 1 & 1 & -1 \\ 1 & -1 & 2 \\ -1 & 2 & 1 \end{bmatrix}, x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, b = \begin{bmatrix} 2 \\ 3 \\ 5 \end{bmatrix}.$$

Now, let us define the product of two matrices in general. This involves two matrices A and B, which are of sizes L-by-M and M-by-N. That is, the first matrix has the same column size as the second matrix has row size. The new matrix $A \cdot B$ is of size L-by-N, i.e. it has as many rows as A and as many columns as B. Each column of $A \cdot B$ is the product of A with the corresponding column of B. That is

$$A \cdot B = \begin{bmatrix} a_{11} \cdot b_1 + a_{12} \cdot b_2 + ... + a_{1N} \cdot b_N \\ a_{21} \cdot b_2 + a_{22} \cdot b_2 + ... + a_{2N} \cdot b_N \\ ... \\ a_{M1} \cdot b_1 + a_{M2} \cdot b_2 + ... + a_{MN} \cdot b_N \end{bmatrix}.$$

Example

$$\begin{bmatrix} 1 & 2 & 1 \\ 4 & 0 & 2 \end{bmatrix} \cdot \begin{bmatrix} 3 & -4 \\ 1 & 5 \\ -2 & 2 \end{bmatrix} = \begin{bmatrix} 1\cdot 3 + 2\cdot 1 + 1\cdot(-2) & 1\cdot(-4) + 2\cdot 5 + 1\cdot 2 \\ 4\cdot 3 + 0\cdot 1 + 2\cdot(-2) & 4\cdot(-4) + 0\cdot 5 + 2\cdot 2 \end{bmatrix} = \begin{bmatrix} 3 & 8 \\ 8 & -12 \end{bmatrix} . \bullet$$

In essence, writing the product of two matrices, we write in a compact way a large amount of scalar (number) products. Why this is useful, will become apparent later, but it is important to emphasize that it is mostly a matter of *notation*.

Note that matrix multiplication is in some ways different from number multiplication. For instance, $A \cdot B$ is different from $B \cdot A$, i.e. order of multiplicands makes a difference. Also, in matrix multiplication (unlike number multiplication), we can have $A \neq \mathbf{0}$, $B \neq \mathbf{0}$, but $AB = \mathbf{0}$.

Example

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} .$$

Just like for numbers, there is a unit for matrices, too. For numbers, we have $a \cdot 1 = 1 \cdot a = a$. For matrices, for every N we have an N-by-N matrix I, such that for all N-by-N matrices A we have $IA = AI = A$. The matrix I is of the form

$$I = \begin{bmatrix} 1 & 0 & ... & 0 \\ 0 & 1 & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & 1 \end{bmatrix} .$$

Example consider

$$\begin{bmatrix} 1 & 2 & 1 \\ 4 & 0 & 2 \\ 3 & 2 & 5 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} =$$

$$\begin{bmatrix} 1\cdot 1 + 2\cdot 0 + 1\cdot 0 & 1\cdot 0 + 2\cdot 1 + 1\cdot 0 & 1\cdot 0 + 2\cdot 0 + 1\cdot 1 \\ 4\cdot 1 + 0\cdot 0 + 2\cdot 0 & 4\cdot 0 + 0\cdot 1 + 2\cdot 0 & 4\cdot 0 + 0\cdot 0 + 2\cdot 1 \\ 3\cdot 1 + 2\cdot 0 + 5\cdot 0 & 3\cdot 0 + 2\cdot 1 + 5\cdot 0 & 3\cdot 0 + 2\cdot 0 + 5\cdot 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 1 \\ 4 & 0 & 2 \\ 3 & 2 & 5 \end{bmatrix} . \bullet$$

Just like with numbers, we can have powers of a matrix: A^2=$A \cdot A$, $A^3 = A \cdot A \cdot A$ etc. Also, we can define $A^0 = I$. Hence we have defined A^n, $n = 0, 1, 2, ...$ What about negative exponents? This leads us to *matrix division*.

Matrix division is obtained by an implicit argument. Recall that for numbers, c/a= $a^{-1}c$. Now, considering the analogous situation for matrices. it seems reasonable to use $A^{-1}C$; this implements division in terms of multiplication. But of course we must first *define* A^{-1}. This must be a matrix that has the properties $AA^{-1} = A^{-1}A = I$ (i.e. it is analogous to the scalar multiplication property $aa^{-1} = a^{-1}a = 1$).

It turns out that we can define A^{-1} in a consistent way for every *square* matrix A (i.e. a matrix of size N-by-N), *subject to one condition*: the determinant of A must be non-zero. [1] If A has non-zero determinant than it has a unique *inverse* matrix, written as A^{-1}. If A is a 2-by-2 matrix, then the formula for the invesre is easy to state and check. For

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

[1]Here we introduce the idea of the *determinant of a matrix*; this is exactly the same as the detrminant of a table, discussed in the previous chapter.

the inverse is given by

$$A^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

Example , take

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

Then

$$A^{-1} = \frac{1}{2 \cdot 4 - 3 \cdot 1} \begin{bmatrix} 4 & -3 \\ -1 & 2 \end{bmatrix} = \begin{bmatrix} 4/5 & -3/5 \\ -1/5 & 2/5 \end{bmatrix}.$$

We can check very easily that

$$\begin{bmatrix} 2 & 3 \\ 1 & 4 \end{bmatrix} \cdot \begin{bmatrix} 4/5 & -3/5 \\ -1/5 & 2/5 \end{bmatrix} = \begin{bmatrix} 2 \cdot 4/5 + 3 \cdot -1/5 & 2 \cdot -3/5 + 3 \cdot 2/5 \\ 1 \cdot 4/5 + 4 \cdot -1/5 & 1 \cdot -3/5 + 4 \cdot 2/5 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Hence,

$$A^{-1} = \begin{bmatrix} 4/5 & -3/5 \\ -1/5 & 2/5 \end{bmatrix}. \bullet$$

Several things must be said regarding the inverse.

1. We see that for every square matrix we can compute its determinant (called $det(A)$); this is expected, since the matrix is really a table.

2. We also see that not all matrices have inverses. Compare this with the fact that every scalar *except* 0 has an inverse. But there are many matrices which have zero determinant and hence no inverse. For instance, when

$$A = \begin{bmatrix} 1 & 3 \\ 2 & 6 \end{bmatrix},$$

   $det(A) = (1 \cdot 6 - 2 \cdot 3) = 0$ and A has no inverse. This situation corresponds to the fact that the number $a = 0$ has no inverse and division by 0 is not possible.

3. Regarding the actual computation of an inverse matrix, we have only given the formula for the 2-by-2 case. Actually, for any N, given a N-by-N matrix A, we can find A^{-1} using a method which essentially solves a system of N^2 linear equations. But we will not discuss this here; instead for the computation of inverses we refer the reader to computer methods; for instance, EXCEL has a function for inverting a matrix.

4. Finally, having defined the inverse matrix A^{-1}, we can also define $A^{-2}= A^{-1} \cdot A^{-1}$, A^{-3} and so on. The important thing is that the usual properties of powers are preserved. For example, $A^{-1}A^2= A^{-1}AA= IA = A = A^1$. Generally, $A^m \cdot A^n= A^{m+n}$, for integers m, n.

Having defined the inverse matrix, we are ready to see another advantage of using matrices: we can solve a system of linear equations very simply. Consider the analogy with the scalar case: $ax = b \Rightarrow x = a^{-1}b$, where a, b, x are numbers. Now consider $Ax = b$, where A is an N-by-N matrix, x and b are N-by-1 matrices. By analogy tothe scalar case, we can write the solution as $x = A^{-1}b$. This gives the solution of the system in one line. Of course, we also have to compute

A^{-1}. If $det(A) \neq 0$, then A^{-1} exists and is computed by EXCEL. So now we can solve a system of linear equations in the computer, using matrix inverse, in manner which is more convenient than using Cramer's rule.
Example Consider the system

$$\begin{aligned} 2x_1 + 3x_2 &= 1 \\ x_1 + 4x_2 &= 0 \end{aligned}.$$

This can be written as $Ax = b$, with

$$A = \begin{bmatrix} 2 & 3 \\ 1 & 4 \end{bmatrix} .x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, b = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

We have already seen that

$$A^{-1} = \begin{bmatrix} 4/5 & -3/5 \\ -1/5 & 2/5 \end{bmatrix}.$$

Hence we can solve the system using

$$x = A^{-1}b = \begin{bmatrix} 4/5 & -3/5 \\ -1/5 & 2/5 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 4/5 \\ -1/5 \end{bmatrix}. \bullet$$

In the case of multiple regression, we have a similar situation. Here too we have a number of equations that we want to satisfy, namely

$$y = a_1x_1 + a_2x_2 + ... + a_Mx_M.$$

In matrix notation this can be written as $Xa = Y$. To understand this, consider an example.
Example Take the following data set

x_1	x_2	x_3	y
1	1	1	2.1
2	4	8	4.9
3	9	27	9.8
4	16	64	17.3
5	25	125	26.1
6	36	216	37.1

Then, if we define

$$X = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 4 & 8 \\ 3 & 9 & 27 \\ 4 & 16 & 64 \\ 5 & 25 & 125 \\ 6 & 36 & 216 \end{bmatrix}, a = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}, Y = \begin{bmatrix} 2.1 \\ 4.9 \\ 9.8 \\ 17.3 \\ 26.1 \\ 37.1 \end{bmatrix},$$

we can write the desired equations in compact form as $Xa = Y$. Here a is the column matrix of unknowns; X, Y are constant matrices, obtained from the data. If we could solve the equation $Xa = Y$, we would also have solved the problem of multiple regression. But now the matrix X (which represents constant parameters) is *not square*; it has $N = 6$ rows and $M = 3$ columns, where $N > M$. So, while we can use matrix notation to write the equations, we cannot apply the method of inverse matrix to solve them *because we have not defined the inverse of a non-square matrix!*. •

However, matrices prove useful in the solution of the multiple regression problem, too. To understand why, we need one final concept from matrix theory, that of the transpose. The transpose of X is written as X' and is obtained by making every column of X a row of X'.

Example

x_1	x_2	x_3	y
1	2	3	5.09
1	3	1	0.09
2	1	1	3.06
2	2	4	8.03
3	3	2	4.06
3	4	1	1.09

We want to find coeffcients a_1, a_2, a_3 which connect x_1, x_2 , x_3 and y in the following manner.

$$y = a_1x_1 + a_2x_2 + a_3x_3.$$

We set up the required matrices as follows.

$$X = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 3 & 1 \\ 2 & 1 & 1 \\ 2 & 2 & 4 \\ 3 & 3 & 2 \\ 3 & 4 & 1 \end{bmatrix} Y = \begin{bmatrix} 5.09 \\ 0.09 \\ 3.06 \\ 8.03 \\ 4.06 \\ 1.09 \end{bmatrix}, a = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}.$$

Here is a case where the transpose is useful (to simplify notation). Suppose that d is a column matrix; then d' is a row matrix and the product $d' \cdot d$ is well defined. In fact

$$d' \cdot d = d_1 \cdot d_1 + d_2 \cdot d_2 + ... + d_N \cdot d_N = d_1^2 + d_2^2 + ... + d_N^2.$$

This is related to multiple regression, because it can be used to write the total square error. The *error matrix* is $Y - Xa$; and the toal square error is

$$(Y - Xa)' \cdot (Y - Xa).$$

Note that Y, Xa, $Y - Xa$ are all N-by-1; hence Y', Xa', $(Y - Xa)'$ are all 1-by-N and $(Y - Xa)'(Y - Xa)$ is 1-by-1, i.e. a *number* (as the total square error ought to be). •

So we have written the total square error of a multiple regression in a really compact form. This is not all; now that we have the expression for total square error, we can also find the solution, i.e. an expression for a, the matrix of regression parameters, that minimizes totalsquare error. This will be done in the following section.

4.1.1 Computing Class Grades

You may have already asked yourself why define matrix multiplication in such a crazy way. However, it actually makes sense. Perhaps this example will convince you.

Here is how I compute your grades. There are five components: solo assignments (30% of final), group assignments (10% of the final), two midterm exams (20% each) and a final exam (20% of final). For each component I consider the average grade (e.g. the total points you get from solo assignments, divided by the number of solo assignments). For instance a student might get the following average grades: [90 85 100 92 87]. That is, he got an average of 90 in the solo assignments, an average of 85 in the group assignments, 100 in the first midterm, 92 in the second midterm and 87 in the final midterm. His final grade will be $90 \cdot 0.3 + 85 \cdot 0.1 + 100 \cdot 0.2 + 92 \cdot 0.2 + 87 \cdot 0.2 = 91.3$. Another student got the following grades: [75 80 70 85 92]. Her final grade will be $75 \cdot 0.3 + 80 \cdot 0.1 + 70 \cdot 0.2 + 85 \cdot 0.2 + 92 \cdot 0.2 = 79.9$. And so on. Another two students got [100 100 78 70 95] and [35 68 55 45 30]. Let us suppose that there are only these four students in class; let us organize their grades in a matrix as follows.

$$X = \begin{bmatrix} 90 & 85 & 100 & 92 & 87 \\ 75 & 80 & 70 & 85 & 92 \\ 100 & 100 & 78 & 70 & 95 \\ 35 & 68 & 55 & 45 & 30 \end{bmatrix}.$$

Now consider also the column matrix

$$a = \begin{bmatrix} 0.3 \\ 0.1 \\ 0.2 \\ 0.2 \\ 0.2 \end{bmatrix}.$$

And now take the matrix product

$$Y = X \cdot a = \begin{bmatrix} 90 & 85 & 100 & 92 & 87 \\ 75 & 80 & 70 & 85 & 92 \\ 100 & 100 & 78 & 70 & 95 \\ 35 & 68 & 55 & 45 & 30 \end{bmatrix} \cdot \begin{bmatrix} 0.3 \\ 0.1 \\ 0.2 \\ 0.2 \\ 0.2 \end{bmatrix} = \begin{bmatrix} 91.3 \\ 79.9 \\ 88.6 \\ 43.3 \end{bmatrix}.$$

You see that by doing this multiplication I get the right answer. Namely, the column matrix Y has four elements, each giving the final grade of the respective student. Of course, this is simple to explain: the calculations inherent in the matrix multiplication are exactly the ones we need to perform to compute the final grades. And this is exactly the point: this type of calculations occurs frequently enough in practice to justify introducing a special operation (matrix multiplication) that performs the required calculations all in one step.

4.1.2 Input-Output Economies

Here is another case where a lot of calculations can be expressed easily by the use of matrix products. The problem we will deal with involves the model of a "closed" economy.

Consider three industries: Industry X, Industry Y and Industry Z. Industry X produces *xebecs*, Industry 2 produces *yarrows* and Industry Z produces *zebus*. Of course, each of these industries needs some *raw materials*. It just so happens that to produce xebecs, the raw materials required are yarrows, zebus and xebecs. In fact, to produce one Kg of xebecs are required 0.2

Kgs of xebecs, 0.3 Kgs of yarrows and 0.2 Kgs of zebus. It is also true that the daily demand of xebecs is 10 Kgs. This is the demand for personal consumption of xebecs, apart from demand for industry consumption.

Similarly, to produce one Kg of yarrows are required 0.4 Kgs of xebecs, 0.1 Kgs of yarrows and 0.2 Kgs of zebus. It is also true that the daily demand of yarrows is 5 Kgs. This is the demand for personal consumption of yarrows, apart from demand for industry consumption.

Finally, to produce one Kg of zebus are required 0.1 Kgs of xebecs, 0.3 Kgs of yarrows and 0.2 Kgs of zebus. It is also true that the daily demand of zebus is 6 Kgs. This is the demand for personal consumption of zebus, apart from demand for industry consumption.

Now, each of the xebec, yarrow and zebu makers produces exactly enough of his product to satisfy the needs of the industry consumption and the personal cosumption, neither more nor less. The question that we must answer is: *what is the quantity of xebecs, yarrows and zebus produced*? Do you think we have enough information to answer this question?

OK, suppose that x Kgs of xebecs, y Kgs of yarrows and z Kgs of zebus are produced. Consider the production of xebecs. Since 1 Kg of xebecs requires 0.2 Kgs of xebecs, 0.3 Kgs of yarrows and 0.2 Kgs of zebus, it follows that x Kgs of xebecs require $0.2{\cdot}x$ Kgs of xebecs, $0.3{\cdot}x$ Kgs of yarrows and $0.2{\cdot}x$ Kgs of zebus. Also, 10 Kgs of xebecs are required to satisfy personal consumption. Since the amount produced must be equal to the amount consumed, we have:

$$x = 0.2 \cdot x + 0.3 \cdot y + 0.2 \cdot z + 10.$$

The left side is the *output* of the xebec industry and the right side is the *input*. Reasoning in a similar manner, we get the following equations for yarrows and zebus:

$$y = 0.4 \cdot x + 0.1 \cdot y + 0.2 \cdot z + 5.$$

$$z = 0.1 \cdot x + 0.3 \cdot y + 0.2 \cdot z + 6.$$

We can write the three equations in matrix form:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0.2 & 0.3 & 0.2 \\ 0.4 & 0.1 & 0.2 \\ 0.1 & 0.3 & 0.2 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \begin{bmatrix} 10 \\ 5 \\ 6 \end{bmatrix}.$$

So we have a system of three equations and three unknowns. Let us define the matrices

$$u = \begin{bmatrix} x \\ y \\ z \end{bmatrix}, \quad d = \begin{bmatrix} 10 \\ 5 \\ 6 \end{bmatrix}, \quad A = \begin{bmatrix} 0.2 & 0.3 & 0.2 \\ 0.4 & 0.1 & 0.2 \\ 0.1 & 0.3 & 0.2 \end{bmatrix}.$$

Then the system can be written as

$$u = Au + d \Rightarrow Iu = Au + d \Rightarrow Iu - Au = d \Rightarrow (I - A)u = d \Rightarrow u = (I - A)^{-1}d.$$

Since

$$I = \begin{bmatrix} 1.0 & 0.0 & 0.0 \\ 0.0 & 1.0 & 0.0 \\ 0.0 & 0.0 & 1.0 \end{bmatrix}, \quad A = \begin{bmatrix} 0.2 & 0.3 & 0.2 \\ 0.4 & 0.1 & 0.2 \\ 0.1 & 0.3 & 0.2 \end{bmatrix} \Rightarrow I - A = \begin{bmatrix} 0.8 & -0.3 & -0.2 \\ -0.4 & 0.9 & -0.2 \\ -0.1 & -0.3 & 0.8 \end{bmatrix}.$$

Then we can compute

$$(I-A)^{-1} = \frac{1}{0.384} \cdot \begin{bmatrix} 0.66 & 0.30 & 0.24 \\ 0.34 & 0.62 & 0.24 \\ 0.21 & 0.27 & 0.60 \end{bmatrix} \Rightarrow$$

$$(I-A)^{-1} \cdot d = \frac{1}{0.384} \cdot \begin{bmatrix} 0.66 & 0.30 & 0.24 \\ 0.34 & 0.62 & 0.24 \\ 0.21 & 0.27 & 0.60 \end{bmatrix} \cdot \begin{bmatrix} 10 \\ 5 \\ 6 \end{bmatrix} = \begin{bmatrix} 24.84 \\ 20.68 \\ 18.36 \end{bmatrix},$$

i.e., $x = 24.84$ Kgs of xebecs, $y = 20.68$ of yarrows, and z=18.36 Kgs of zebus are produced daily.

Several points must be noted withe regard to this problem.

1. First of all, perhaps surprisingly, from the information given us, we can find the answer, namely the daily production of each industry.

2. Once we use the information, we set up a system of three linear equations in three unknowns. It is rather expected that these equations will admit one solution.

3. However, this is not the whole story. The solution, namely x, y, z, cannot be just any numbers. Since they are quantities produced, they must be *positive*. (In fact, they must be greater than (respectively) 10, 5 and 6 - can you justify this claim?)

4. So we find ourselves in a new situation. We have a system of linear equations which must have *positive* solutions. Is it guaranteed that any input - output system that we can think of will have positive solutions? Obviously not, since we know of systems of linear equations that have negative solutions. Can you think of necessary conditions that an input - output system must satisfy so that it has positive solutions? (*Hint*: Look at the form of matrix A, especially at the sums of columns. What particular property *must* they have?)

4.2 Multiple Regression: the Equations

Let us repeat the multiple regression problem. We want to find parameters a_1, a_2, ... , a_M which connect x_1, ... , x_M and y in the following manner.

$$y = a_1 x_1 + a_2 x_2 + ... + a_m x_m$$

and so as to minimize total square error:

$$\sum (y - a_1 x_1 + a_2 x_2 + ... + a_m x_m)^2.$$

The equations can be found using an analogy with the method of completing the square, but now with matrices instead of numbers. To do this, first define the matrices: a, X and Y; this was done in the previous section. Then, the regression equations are written as $Xa = Y$; this is a *matrix* equation: one equation describes the whole table (several scalar equations). Now, the regression error is $Y - Xa$ and the total square error is $(Y - Xa)'(Y - Xa)$. [2] This is where we

[2]Once again: note that Y, Xa, $Y - Xa$ are all N-by-1; hence Y', Xa', $(Y - Xa)'$ are all 1-by-N and $(Y - Xa)'(Y - Xa)$ is 1-by-1, i.e. a *number* (as the total square error ought to be).

start to develop the equations for a. First of all, note that if $Xa = Y$, then the error becomes zero. But also note that $Xa = Y$ is a total of N number equations, and the unknowns are a_1, ..., a_M, a total of M unknowns. When $N > M$, we cannot solve $Xa = Y$, and we cannot make the error zero. Hence we need another approach, to *minimize* the error. We will start with an observation. *Suppose* $Xa = Y$; then, also,

$$X'Xa = X'Y.$$

But, $X'X$ is an $M - by - M$ matrix and $X'Xa = X'Y$ is M equations with M unknowns (the a's. This equation *can* be solved for a; we get

$$a = (X'X)^{-1}X'Y. \tag{1}$$

And this gives the required expression for a, the regression parameters which minimize total square error. Of course, this just a conjecture. That $a = (X'X)^{-1}X'Y$ does indeed minimize the error, will be proved a little later. For the time being, how does one *use* eq.(1)? To understand this, consider the following example.

Example Suppose we want to perform a multiple regression for the data set

x_1	x_2	x_3	y
1	2	3	5.09
1	3	1	0.09
2	1	1	3.06
2	2	4	8.03
3	3	2	4.06
3	4	1	1.09

We want to find coeffcients a_1, a_2, a_3 which connect x_1, x_2 , x_3 and y in the following manner.

$$y = a_1x_1 + a_2x_2 + a_3x_3.$$

We set up the required matrices as follows.

$$X = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 3 & 1 \\ 2 & 1 & 1 \\ 2 & 2 & 4 \\ 3 & 3 & 2 \\ 3 & 4 & 1 \end{bmatrix}, X' = \begin{bmatrix} 1 & 1 & 2 & 2 & 3 & 3 \\ 2 & 3 & 1 & 2 & 3 & 4 \\ 3 & 1 & 1 & 4 & 2 & 1 \end{bmatrix}, Y = \begin{bmatrix} 5.09 \\ 0.09 \\ 3.06 \\ 8.03 \\ 4.06 \\ 1.09 \end{bmatrix}, a = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}.$$

Now we have

$$X'X = \begin{bmatrix} 28 & 32 & 23 \\ 32 & 43 & 28 \\ 23 & 28 & 32 \end{bmatrix}, X'Y = \begin{bmatrix} 42.82 \\ 46.12 \\ 59.78 \end{bmatrix},$$

Using EXCEL, we find

$$(X'X)^{-1} = \begin{bmatrix} 0.26 & -0.17 & -0.04 \\ -0.17 & 0.16 & -0.02 \\ -0.04 & -0.02 & 0.08 \end{bmatrix}.$$

And finally,

$$a = (X'X)^{-1}X'Y = \begin{bmatrix} 0.26 & -0.17 & -0.04 \\ -0.17 & 0.16 & -0.02 \\ -0.04 & -0.02 & 0.08 \end{bmatrix} \cdot \begin{bmatrix} 42.82 \\ 46.12 \\ 59.78 \end{bmatrix} = \begin{bmatrix} 1.00 \\ -0.97 \\ 2.00 \end{bmatrix}.$$

In other words, $a_1 = 1.00$, $a_2 = -0.97$, $a_3 = 2.00$ and the estimated $\hat{y}$ is $\hat{y} = 1.00 \cdot x_1 - 0.97 \cdot x_2 + 2.00 \cdot x_3$. Using this formula, we get for $\hat{y}$; let us compare these to the true y values:

y	$\hat{y}$
5.09	5.06
0.09	0.08
3.06	3.02
8.03	8.06
4.06	4.08
1.09	1.10

which gives quite good agreement. This completes the solution of the problem of multiple regression. •

It is a useful exercise to try and find the form of the regression equations without using matrix theory. To obtain these equations, we need to expand the matrix equation

$$X'Xa = X'Y.$$

After a serious amount of algebra, we obtain

$$X'X = \begin{bmatrix} (\sum x_1x_1) & (\sum x_1x_2) & ... & (\sum x_1x_M) \\ (\sum x_2x_1) & (\sum x_2x_2) & ... & (\sum x_2x_M) \\ ... & ... & ... & ... \\ (\sum x_Mx_1) & (\sum x_Mx_2) & ... & (\sum x_Mx_M) \end{bmatrix}, X'Y = \begin{bmatrix} (\sum x_1y) \\ (\sum x_2y) \\ ... \\ (\sum x_My) \end{bmatrix}.$$

In other words, the required system is

$$\begin{array}{lllll} (\sum x_1x_1)a_1 & +(\sum x_1x_2)a_2 & ... & +(\sum x_1x_M)a_M & = (\sum x_1y) \\ (\sum x_2x_1)a_1 & +(\sum x_2x_2)a_2 & ... & +(\sum x_2x_M)a_M & = (\sum x_2y) \\ ... & ... & ... & ... & = ... \\ (\sum x_Mx_1)a_1 & +(\sum x_Mx_2)a_2 & ... & +(\sum x_Mx_M)a_M & = (\sum x_My) \end{array}.$$

To *obtain* this system, we can proceed similarly to the method we used for linear and quadratic regression, i.e. build tables of x_1^2, x_2^2, ... , x_1x_2, x_1x_3, ... , x_1y, x_2y, ... and so on and compute the sums in the final row of these tables. Then the system has been obtained; it can be *solved* using Cramer's rule or some other method. This is multiple regression *without matrices*;

On the other hand, multiple regression *with matrices* we just compute $X'X$, $X'Y$, $(X'X)^{-1}$, $(X'X)^{-1}\cdot$ $X'Y$ using the EXCEL transpose, multiplication and inversion formulas, resulting in considerable economy of time and space.

What is left, is to prove that a as given by eq.(1) minimizes the total square error. To check this, we will need two properties of the transpose, which we give without proof (they are easy to check):

1. $(AB)' = B'A'$;

2. if A and B are *symmetric* square matrices, then $A'B = B'A$.

If we accept this, let us now take a as given by eq.(1) and also take some other regression parameters, say b, different from a. Let us then compute the total square error for b. We have

$$(Y - Xb)'(Y - Xb) = (Y - Xa + Xa - Xb)'(Y - Xa + Xa - Xb) =$$

$$(Y - Xa)'(Y - Xa) + (Y - Xa)'(Xa - Xb) + (Xa - Xb)'(Y - Xa) + (Xa - Xb)'(Xa - Xb) =$$
$$(Y - Xa)'(Y - Xa) + 2(Y - Xa)'(Xa - Xb) + (Xa - Xb)'(Xa - Xb).$$

Now let us use the fact that $a = (X'X)^{-1}X'Y$. Putting this in the previous equation we get

$$(Y - Xa)'(Y - Xa) + 2(Y - Xa)'(Xa - Xb) + (Xa - Xb)'(Xa - Xb). \tag{2}$$

Let us concentrate on the term $(Y - Xa)'(Xa - Xb)$:

$$(Y - Xa)'(Xa - Xb) = Y'Xa - a'X'Xa - YXb + a'X'Xb =$$

(using the fact that $a = (X'X)^{-1}X'Y$)

$$Y'X[(X'X)^{-1}X'Y] - [(X'X)^{-1}X'Y]'X'X[(X'X)^{-1}X'Y] - YXb + [(X'X)^{-1}X'Y]'X'Xb =$$

$$Y'X(X'X)^{-1}X'Y - Y'X(X'X)^{-1}X'X(X'X)^{-1}X'Y - YXb + Y'X(X'X)^{-1}X'Xb =$$
$$Y'X(X'X)^{-1}X'Y - Y'X(X'X)^{-1}X'Y - YXb + Y'Xb = 0.$$

OK, that's it. Summarizing, we proved that $(Y - Xa)'(Xa - Xb) = 0$. But then, substituting in eq.(2) we get

$$(Y - Xb)'(Y - Xb) =$$
$$(Y-Xa)'(Y-Xa)+2\cdot 0+(Xa-Xb)'(Xa-Xb).(Y-Xa)'(Y-Xa)+(Xa-Xb)'(Xa-Xb). \tag{3}$$

But we know that $(Xa - Xb)'(Xa - Xb) \geq 0$. Hence eq.(3) implies that

$$(Y - Xb)'(Y - Xb) \geq (Y - Xa)'(Y - Xa). \tag{4}$$

Now, note that in $(Y - Xb)'(Y - Xb)$, we have X and Y fixed, b variable. Here we have a function $F(b) = (Y - Xb)'(Y - Xb)$. What we have established is that

$$F(b) = F(a) + (Xa - Xb)'(Xa - Xb) \geq F(a) \tag{5}$$

Hence $F(a)$ is the minimum value of $F(b)$, in other words $F(b)$ achieves its minimum value for $b = a$. In still other words: the total square error $F(b)$ achieves its minimum value for $b = a$, where

$$a = (X'X)^{-1}X'Y.$$

4.3 Polynomial and Multiple Regression

General discussion of relationship between polynomial and multiple regressions: We have seen the equations for *first order* regression (i.e. $y = ax + b$); a and b are the solutions of the system

$$\begin{array}{rll} Nb & +(\sum x)a & = (\sum y) \\ (\sum x)b & +(\sum x^2)a & = (\sum xy) \end{array}.$$

We have also seen the equations for *second order* regression (i.e. $y = ax^2 + bx + c$); a, b and c are the solutions of the system

$$\begin{array}{rlll} Nc & +(\sum x)b & +(\sum x^2)a & = (\sum y) \\ (\sum x)c & +(\sum x^2)b & +(\sum x^3)a & = (\sum xy) \\ (\sum x^2)c & +(\sum x^3)b & +(\sum x^4)a & = (\sum x^2 y) \end{array}.$$

At this point you can probably guess the equations that give coefficients for higher order regressions (i.e. third, fourth etc.). But can you guess the equations for *zero* order regression?

4.3.1 Zero-Order Regression

Zero order regression means that $y = a$. I.e., we are looking for a constant that approximates the data points with minimum total square error. Before we dwell on the meaning of this, read and consider the attached passage from Thucydides. It refers to the siege of Plataea by the Peloponessians. The Peloponessians have built a wall around Plataea and the Plataeans want to break out.

> ...There remained about 220 volunteers who persisted in the idea of venturing out. Their method was as follows: they constructed ladders to reach the top of the enemy's wall, and they did this by calculating the height of the wall from the number of the layers of bricks at a point which was facing their way and had not been plastered. The layers were counted by a lot of people at the same time, and though some were likely to get the figure wrong, the majority would get it right. ... Thus, guessing what the thickness of a single brick was, they calculated how long their ladders would have to be ... (Thucydides, *The Peloponessian War*, Book 3, 20-21.)

Figure 4.1

The passage is not clear, but it appears that the Plataeans got several (say N) measurements of the number of layers; call these y; then they estimated the *true* number of layers a by the formula $a = \frac{\sum y}{N}$.

OK, let us go back to zero order regression of y and give the equation for a, which minimizes $\sum (y-a)^2$. You probably have figured it out yourselves.

$$N \cdot a = (\sum y) \Rightarrow$$

$$a = \frac{\sum y}{N}.$$

This is simply the *average* of the y's. It is exactly what the Plataeans used, and maybe you now also see the rationale behind it: the true data is just a constant, but the observations of the constant are *noisy*, or they have an observation error. The average of the observations is an estimate that minimizes total square error.

Can you actually *prove* that the average minimizes total square error? Yes, you can , using simply the techniques of quadratic functions. Here is how. Define $F(a) = \sum (y-a)^2$. We have

$$F(a) = \sum [y^2 - 2a \cdot y + a^2] =$$

$$\sum [y^2 - \sum (2a \cdot y) + a^2] = (\sum y^2) - 2a(\sum y) + (\sum a^2) =$$

$$N \cdot a^2 - 2a \cdot (\sum y) + (\sum y^2) = N \cdot [a^2 - 2a\frac{\sum y}{N} + \frac{\sum y^2}{N}] =$$

$$N \cdot [a^2 - 2a\frac{\sum y}{N} + (\frac{\sum y}{N})^2 - (\frac{\sum y}{N})^2 + \frac{\sum y^2}{N}] =$$

$$N \cdot (a - \frac{\sum y}{N})^2 + (\sum y^2 - \frac{(\sum y)^2}{N}) \geq (\sum y^2 - \frac{(\sum y)^2}{N})$$

You see, $F(a)$ is simply a quadratic function of a; it has minimum value equal to $\sum y^2 - \frac{(\sum y)^2}{N}$, which is attained for $a = \frac{\sum y}{N}$.

The Plataeans recognized, like most of us do today, that taking the average is the "reasonable thing"; we have seen here that it also has a theoretical justification. Now consider the minimization of total absolute error.

Squares are mathematically more convenient than absolute values. Is it true that $a = \frac{\sum y}{N}$ also minimizes the total absolute error $\sum |y-a|$? If it is, can you prove it?

4.4 More on matrices

In the previous sections we have used matrices in the context of slvong linear equations, exactly or approximately. But matrices are useful for a lot mre than this. Here are given a few hints to alternative interpretations of matrices. These start with a more algebraic point of view and later become more geometric.

Because this section is outside the main theme of this book, it is organized as a problem set. "Do" as many of the problems as you can, where "doing" means: prove, give examples, discuss or whatever else comes to your mind.

1. We know that *when* $-1 < a < 1$, then $\frac{1}{1-a} = (1-a)^{-1}$ equals $1 + a + a^2 + ...$ Can you think of a similar formula for $(I - A)^{-1}$? You should find the formula, perhaps by analogy with the case of $(1-a)^{-1}$, and verify it is correct.

2. Consider a N-by-N matrix A as a collection of N rows, i.e. $A = [r_1, r_2, ..., r_N]$. Suppose you want to *permute* its rows. For instance, starting with $A = [r_1, r_2, ..., r_N]$, you want to get $B = [r_2, r_1, ..., r_N]$. Can you do this by multiplying with a new matrix P having the property that its elements are only 0's and 1's, and in addition every row and every row has *exactly one* 1? Such matrices P are called *permutation* matrices. Which permutation matrix permutes rows r_m with r_n? How about permutation of columns; i.e. how can i permute columns?

3. Consider a N-by-N matrix A as a collection of N columns, i.e. $A = [c_1, c_2, ..., c_N]$. Then I make the following **Claim:** *$det(A)$ is the volume of the parallelepiped which is defined by the begiinning of axes and points c_1, c_2, ..., c_N.*

   For example, if a
   $$A = \begin{bmatrix} 1 & 1 & 0 \\ 2 & 0 & 1 \\ 0 & 3 & 1 \end{bmatrix}$$
   then $det(A)$ is the volume of the parallelepiped defined by points (0,0,0), (1,2,0), (1,0,3), (0,1,1).

   Can you justify / prove the above claim? It holds for papallelepiped of dimension N=2 (parallelograms), $N = 3$ and even for any dimension N. It may help you to work with a few examples.

4. From the previous discussion, it appears that we can identify a N-by-N matrix A with a collection of points in R^N. Can you explain this statement? Give examples.

5. Take the claim that determinant is volume for granted; use it to prove the following properties of the determinant.

   (a) $det([c_1, c_2, ..., c_N])$=0 if c_m= c_n for some m, n.

   (b) $det([c_1, c_2, ..., c_m, ..., c_n, ..., c_N])$= $det([c_1, c_2, ..., c_n, ..., c_m, ..., c_N])$ for some m, n.

   (c) $det([c_1, c_2, ..., \lambda c_m, ..., c_N])$= $\lambda det([c_1, c_2, ..., c_m, ..., c_N])$ for some m, n and any $\lambda \in R$.

   (d) $det([c_1, c_2, ..., c_m + b_m, ..., c_N])$= $det([c_1, c_2, ... , c_m, ..., c_N]) + det([c_1, c_2, ..., b_m, ..., c_N])$ for some m, n.

6. Show that $det(AB)$= $det(A) \cdot det(B)$. Now use this and the previous claim (determinant is volume) to justify the fact that A and $B = PA$ have the same determinant when P is a permutation matrix.

   Example: consider the permutation matrix to be
   $$P = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}.$$

If I take a 3-by-3 matrix $A = [r_1, r_2, r_3]$ (where r_1, r_2, r_3 are rows) what is the result of the multiplication PA? What is the geometric meaning? To explain this, you should also think of A in terms of columns, i.e. $A = [c_1, c_2, c_3]$ and each column as a point; what happens to the coordinates of these points? Why should A and PA have the same volume / determinant.

7. Now, given a N-by-1 column x, we can consider it as a point in R^N. Take a N-by-N matrix A; we can consider $f(x) = Ax$ as a new point in R^N. This gives a geometric meaning to $f(x) = Ax$: we are given an *initial* point x and a final point $f(x)$. In fact, if we take $X = [x_1, x_2, ..., x_N]$, and $Y = [y_1, y_2, ..., y_N]$ where the columns x_1, x_2, ... , x_N and y_1, y_2, ... , y_N correspond to points in R^N, we see that the parallelepiped X (what does this mean?) is transformed to parallelepiped $Y = AX$ (why?). In addition, if A has an inverse, the parallelepiped Y (what does this mean?) is transformed to parallelepiped $X = A^{-1}Y$. Now, if $det(X) \neq 0$ and $det(A) = 0$, then the volume of X is nonzero, but the volume of Y is zero (why?). Can you relate this geometric fact with the algebraic fact that when $det(A) = 0$, A has no inverse?

8. Given a N-by-N matrix A, one can write the equation $Ax = a \cdot x$, where a is a number and x a column matrix. Note that here we have N equations and $N + 1$ unknowns (x_1, ... , x_N *and* a). Show that it is possible to solve this system of equations. Give an example with N=2 and another with N=3. What is the geometric meaning of the above equation?

9. Suppose we have variables x_1, x_2, ... , x_N, arranged in the column matrix $x = [x_1, x_2, ... , x_N]'$. Also consider the *fixed* N-by-N matrix A. Then the product Ax defines a *function of N variables*, namely $f(x_1, x_2, ... , x_N) = Ax$. Can you find the derivative of this function in matrix form? (The derivative with respect to *many* variables is a column matrix, where the n-th element is the derivative with respect to the n-th variable (i.e. the derivative considering x_n variable and all other x's fixed).

   Do the same thing for the function $g(x_1, x_2, ... , x_N) = x'Ax$. Do you see any similarity with the case of one-variable functions?

4.5 Problems

1. Find the products $A \cdot B$ and $B \cdot A$ (when possible!) for the following cases of A and B.

2.

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

3.

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

4.

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

5.

$$A = \begin{bmatrix} 5 & -6 & 10 \end{bmatrix}, B = \begin{bmatrix} 1 & 12 & -61 \\ 15 & -71 & 12 \\ 23 & -22 & -3 \end{bmatrix}.$$

6.

$$A = \begin{bmatrix} 1 & 1 & -1 \\ 1 & -1 & 2 \\ -1 & 2 & 1 \end{bmatrix}, B = \begin{bmatrix} 1 & 1 & -1 \\ -1 & 2 & 1 \end{bmatrix}.$$

7.

$$A = \begin{bmatrix} 1 & -1 \\ 1 & 2 \\ -1 & 1 \end{bmatrix}, B = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}.$$

8. In every case verify that $A \cdot B = B \cdot A = I$. Hence $A^{-1} = B$.

(a)

$$A = \begin{bmatrix} 6 & 2 \\ 0 & 4 \end{bmatrix}, B = \frac{1}{24}\begin{bmatrix} 4 & -2 \\ 0 & 6 \end{bmatrix}.$$

(b)

$$A = \begin{bmatrix} 1 & 3 & 3 \\ 1 & 3 & 4 \\ 1 & 4 & 3 \end{bmatrix}, B = \begin{bmatrix} 7 & -3 & -3 \\ -1 & 0 & 1 \\ -1 & 1 & 0 \end{bmatrix}.$$

(c)

$$A = \begin{bmatrix} 2 & 3 & 1 \\ 1 & 2 & 3 \\ 3 & 1 & 2 \end{bmatrix}, B = \frac{1}{18}\begin{bmatrix} 1 & -5 & 7 \\ 7 & 1 & -5 \\ -5 & 7 & 1 \end{bmatrix}.$$

9. For the following data compute a multiple regression.

x_1	1	2	3	1	3	3
x_2	1	2	3	2	2	3
x_3	1	2	3	3	1	3
y	3.1	5.9	9.2	6.3	6.1	8.5

10. For the following data compute a multiple regression.

x_1	1	1	1	2	2	2
x_2	1	1	2	2	1	2
x_3	1	2	1	2	1	2
y	0.9	-0.1	1.8	4.3	1	2.1

11. For the following data compute a first and second order autoregression. Compare their accuracy.

t	1	2	3	4	5	6
$y(t)$	1	1.1	1.2	1.35	1.6	1.8

12. For the following data compute a first and second order autoregression.

t	1	2	3	4	5	6
$y(t)$	1.1	1.8	1.52	1.75	1.6	1.7

Also do book problems 10.3.25-.51, 10.3.55, 10.3.59-.60, 10.4.5-12, 10.4.20-.21, 10.4.27-.31.

5 Dynamic Models

Consider the following expressions. They involves constants a_1, a_2, ... and a function $y(t)$, where $t = 0, 1, 2, ...$

$$y(t) = a_1 y(t-1), \tag{1}$$

$$y(t) = a_1 y(t-1) + a_2 y(t-2), \tag{2}$$

$$...$$

$$y(t) = a_1 y(t-1) + a_2 y(t-2) + ... + a_M y(t-M). \tag{3}$$

These equations are called *difference equations*. The first equation is a *first order* difference equation, the second a *second order* difference equation and the last one is an *M-th order* difference equation.

What exactly do these equations tell us? They involve a function $y(t)$, which so far is unknown. Now, in general, when we use the term "equation", it is implied that the equation can be solved and the unknown determined. Is this true of difference equations? To make things specific, consider the equation

$$y(t) = y(t-1) + y(t-2).$$

(In this case we have $a_1 = 1$ and a_2=1.) This formula is not sufficient by itself to tell us what $y(t)$ is. But if we also knew the *initial values* $y(1)$ and $y(2)$, then $y(t)$ is determined for every t. For instance, with $y(1) = 1$ and $y(2) = 1$, we have

$$y(3) = y(2) + y(1) = 1 + 1 = 2,$$

$$y(4) = y(3) + y(2) = 2 + 1 = 3,$$

$$y(5) = y(4) + y(3) = 3 + 2 = 5...etc.$$

The first question of interest is whether equations of the form (3) really define a function. The answer can be obtained from the previous examples. Given the difference equation and initial values, we can compute $y(t)$ for any t; hence we have a function. But the definition of the function is *implicit*: rather than having a formula that uses t to compute $y(t)$, we have a sequence of formulas that gives $y(t)$ from previous values $y(t-1)$, $y(t-2)$, ... Can we find a formula for $y(t)$ in terms of t? In some cases it is quite easy, as we will see in the following sections.

The second question of interest is: what can we do with a model that uses a difference equation. Is it any "better" than a model which directly gives us the formula of the function, e.g. $y(t) = 1 + t^2$?

In the following sections we will discuss the above two questions.

5.1 First Order Scalar Difference Equations

In general, suppose that we have the difference equation

$$y(t) = a \cdot y(t-1), \quad y(0) = y_0. \tag{4}$$

Then, we can obtain the *solution* to (4); it is

$$y(t) = a^t \cdot y_0. \tag{5}$$

Eq. (4) is an *homogeneous linear first order difference equation.* An *inhomogeneous linear first order difference equation* has the form

$$y(t) = a \cdot y(t-1) + b, \quad y(0) = y_0. \tag{6}$$

In other words, we have an additional constant b. Proceeding as previously, we compute

$$y(1) = ay_0 + b,$$

$$y(2) = a \cdot (ay_0 + b) + b = a^2 y_0 + (a+1) \cdot b,$$

$$y(3) = a \cdot (a^2 y_0 + ab + b) + b = a^3 y_0 + (a^2 + a + 1) \cdot b$$

and in general

$$y(t) = a^t \cdot y_0 + (a^{t-1} + a^{t-2} + ... + a + 1) \cdot b. \tag{7}$$

This is the *solution* to (6).

In general, difference equations are appropriate to describe phenomena where the increase of quantity $y(t)$ is proportional to $y(t)$; capital increasing according to interest is such a case. Other cases include increase of the population of a country, decrease of a radioactive substance etc.

So far we have talked about linear difference equations, i.e. $y(t)$ is given as a linear function of $y(t-1)$. But we can also have difference equations that are *nonlinear*, e.g.

$$y(t) = [y(t-1)]^2.$$

In other words, in a nonlinear difference equation, we have $y(t)$ given as a nonlinear expression of $y(t-1)$. It is clear that in such a case we can also find $y(1)$, $y(2)$, ... , $y(t)$, ... from the difference equation and an initial condition $y(0) = y_0$. But, generally, it is difficult or impossible to find a direct expression of $y(t)$ as a function of time. This is a difference between a nonlinear diference equation and a linear one, where $y(t)$ can be given by an explicit formula, such as (9).

Finally, we can have difference equations which involve more than one functions. For instance

$$y(t) = 0.9 \cdot y(t-1) - 0.1 \cdot z(t-1), \;\; z(t) = 0.2 \cdot y(t-1) + 0.2 \cdot z(t-1). \tag{8}$$

This is a linear difference equation. We can also have nonlinear difference equations with several functions. Both in the linear and nonlinear case, we can find $y(1)$, $z(1)$, $y(2)$, $z(2)$, ..., $y(t)$, $z(t)$ ... from the difference equation (8) and initial conditions $y(0) = y_0$, $z(0) = z_0$. In addition, in the case of linear difference equations, we can find explicit formulas for $y(t)$, $z(t)$, using matrices. This will be done a little later; let us first see some applications of difference equations.

5.1.1 Capital and Interest

Suppose that you put 1000 drachmas in a bank account with yearly interest rate 10%. Then, at the end of the first year you get interest 1000·0.10= 100 drs., which is *compounded* with the initial capital of 1000 drs to give total capital 1000+100 drs=1100 drs. At the end of the second year you get interest rate of 1100·0.10= 110 drs. , which is *compounded* with the initial capital of 1000 drs to give total capital 1100+110 drs=1210 drs. In general, if capital at year $t-1$ is $y(t-1)$, at year t capital is

$$y(t) = 1.1 \cdot y(t-1);$$

in particular we have

$$y(1) = 1.1 \cdot y(0);$$

$$y(2) = 1.1 \cdot y(1) = 1.1 \cdot 1.1 \cdot y(0) = 1.1^2 \cdot y(0);$$

$$...$$

$$y(t) = 1.1 \cdot y(t-1) = 1,1^t y(0).$$

This, combined with the *initial value* of capital, $y(0)$= 1000, gives the formula

$$y(t) = (1.1)^t \cdot 1000.$$

In a similar manner, if the initial value of capital is $y(0) = y_0$ and the interest rate is r, we have

$$y(t) = (1+r)^t \cdot y_0.$$

This formula holds for *annual compounding*. Suppose that interest is compounded with the capital n times a year. Then, the formula for total capital at year t is given by

$$y(t) = (1 + \frac{r}{n})^{n \cdot t} \cdot y_0. \tag{9}$$

The fraction $\frac{r}{n}$ appears because, with a yearly interest rate r, the actual interest rate in periods of $\frac{1}{n}$-th of the year is $\frac{r}{n}$. The power $n \cdot t$ appears because in one year $n \cdot t$ compoundings take place.

5.1.2 Dynamic Ajustment of Price According to Supply and Demand

Here is another application involving a first order difference equation. We have already mentioned how the sales of googles, denoted by q, is a function of the going price p:

$$q = 100 - 2p. \tag{10}$$

Under the previous analysis of the problem, we have implicitly assumed that whatever amount q is produced, it is also sold. This, however, is not necessarily true. In fact, we can differentitate between q, which is the amount of googles *sold*, and is given by the previous equation, and r, which is the amount of googles *produced*, and is related to price according to

$$r = p + 400. \tag{11}$$

We can call q, as given by eq.(10), *demand* and r, as given by eq.(11), *supply*. You see that q, as given by (10) is a *decreasing* function of price, i.e. when price increases demand go down; while r as given by (11) is a *increasing* function of price, i.e. when price increases supply goes up. Does this make sense to you?

Now, we have so far assumed that price, production and sales are fixed in time. But, in fact, all of these quantities are *time-variable*, that is they change in time. So it is better tow rite $p(t)$, $q(t)$, $r(t)$. In this case we will have the following equations.

$$q(t) = 100 - 2p(t). \tag{12}$$

$$r(t) = p(t) + 40. \tag{13}$$

Now, what we would like to know, is how price, demand and supply evolve in time. For example, suppose that we start at time t=0 with an initial price $p(0)$=10. Then we will have a supply

$$r(0) = p(0) + 40 = 50, \tag{14}$$

and demand

$$q(0) = 100 - 2p(0) = 100 - 2 \cdot 10 = 80. \tag{15}$$

So less googles are produced (50) than demanded (80). There is a *scarcity* of googles (80-50=30). What will the surplus do to the price? It will tend to increase it. But by how much? Let as *assume* that price changes at times t=1, 2, 3 ... and that the following relationship holds:

$$p(t) - p(t-1) = 0.1 \cdot (q(t-1) - r(t-1)). \tag{16}$$

Eq.(16) tells us that when there is more demand for googles than supply (i.e. $q(t-1) > r(t-1) \Rightarrow q(t-1) - r(t-1) > 0$), then the price tends to *increase* ($p(t) - p(t-1) = 0.1 \cdot (q(t-1) - r(t-1)) > 0.$). Conversely, when there is less demand for googles than supply (i.e. $q(t-1) < r(t-1) \Rightarrow q(t-1) - r(t-1) < 0$), then the price tends to *increase* ($p(t) - p(t-1) = 0.1 \cdot (q(t-1) - r(t-1)) < 0.$). The 0.1 is just a *constant of proportion.*

If we accept (16), then the price at time t=1, $p(1)$ will be given by

$$p(1) - p(1-1) = 0.1 \cdot (q(1-1) - r(1-1)) \Rightarrow p(1) - p(0) = 0.1 \cdot (q(0) - r(0)) \Rightarrow$$

$$p(1) - 10 = 0.1 \cdot 30 \Rightarrow p(1) = 10 + 3 = 13. \tag{17}$$

So the price of googles at time t=1 becomes 13. Now supply at time $t = 1$ is given by

$$r(1) = p(1) + 40 = 13 + 40 = 53. \tag{18}$$

and demand becomes

$$q(1) = 100 - 2p(1) = 100 - 2 \cdot 13 = 74. \tag{19}$$

We see several things happening:

1. the price of googles increased ($p(1) = 13 > p(0) = 10$);
2. the supply of googles increased ($r(1) = 53 > r(0) = 50$);
3. the demand for googles decreased ($q(1) = 74 < r(0) = 80$).

All of these things make sense. Now, looking at $t = 2$, there is a new price for googles:

$$p(2) - p(2-1) = 0.1 \cdot (q(2-1) - r(2-1)) \Rightarrow p(2) - p(1) = 0.1 \cdot (q(1) - r(1)) \Rightarrow$$

$$p(2) - 13 = 0.1 \cdot 21 \Rightarrow p(2) = 13 + 2.1 = 15.1. \tag{20}$$

The price increased some more. This will in turn mean an increase in supply and a decrease in demand. And this process will keep going until ... Until what? What will happen in the long run? This is the problem that we want to answer. Here is how we go about it.

Let us rewrite the equations relating $p(t)$, $q(t)$, $r(t)$.

$$q(t) = 100 - 2p(t), \tag{21}$$

$$r(t) = p(t) + 40, \tag{22}$$

$$p(t) - p(t-1) = 0.1 \cdot (q(t-1) - r(t-1)). \tag{23}$$

Substitute eq.(21) and eq.(22) in eq.(23) to get

$$p(t) - p(t-1) = 0.1 \cdot (100 - 2p(t-1) - (p(t-1) + 40)) = 0.1 \cdot (60 - 3p(t-1)) \Rightarrow$$

$$p(t) = p(t-1) + 0.3 \cdot (20 - p(t-1)). \tag{24}$$

We can also write this in more familiar form.

$$p(t) = 0.7 \cdot p(t-1) + 6. \tag{25}$$

Recall also the initial condition:

$$p(0) = 10. \tag{26}$$

Eqs.(25) and (26) give a first order scalar difference equation, which we know how to solve. The theoretical solution is (with a=0.7, b=6):

$$p(t) = 0.7^t \cdot p(0) + (1 + 0.7 + 0.7^2 + ... + 0.7^{t-1}) \cdot b =$$

$$p(t) = 0.7^t \cdot 10 + (1 + 0.7 + 0.7^2 + ... + 0.7^{t-1}) \cdot 6 =$$

When $t \to \infty$, $0.7^t \to 0$. Also,

$$(1 + 0.7 + 0.7^2 + ... + 0.7^{t-1}) \cdot 6 \to (1 + 0.7 + 0.7^2 + ...) \cdot 6 = \frac{1}{1 - 0.7} \cdot 6 = 20.$$

So we can say that the *equilibrium* value is $p(\infty)$=20. Note that when $p(\infty)$=20, then

$$q(\infty) = 100 - 2p(\infty) = 100 - 2 \cdot 20 = 60, \tag{27}$$

$$r(\infty) = p(\infty) + 40 = 20 + 40 = 60, \tag{28}$$

that is, supply becomes equal to demand; that's why we say that $p(\infty)$ is the equilibrium value.

All of these things are also verified numerically: we can start with eq.(26) and repeatedly apply (25) (EXCEL will do this for us) to obtain numerically $p(t)$ for t=1, 2, 3 ... . Then we can use eq.(21) to also obtain $q(t)$ and eq.(22) to obtain $r(t)$ for t=1, 2, ... Finally, we can graph these functions and get Figures 5.1, 5.2, 5.3.

Figure 5.1

Figure 5.2

Figure 5.3

5.1.3 Epidemics

Let us now consider an example from biology. What is the mechanism by which a disease is spread in a population? To make things concrete, suppose we have a population of 1000 people. Let the number of infected people at time t be $n(t)$. Suppose that initially (i.e. at time t=0), there is one ifected person, i.e.

$$n(0) = 1. \tag{29}$$

Now, how is the number of people infected at time t determined? A simple but plausible mechanism, is that

$$n(t) = n(t-1) + 0.001 \cdot n(t-1) \cdot (1000 - n(t-1)). \tag{30}$$

The reasoning behind eq.(30) is the following: we assume that people who were infected in the previous period remain infected. This is the term $n(t-1)$. In addition, some more people get infected. The number of new infections depends on the number of already infected people and the number of as yet healthy people; obviously when more infected people are around the number of healthy persons getting infected must increase; also when fewer healthy people are around, the number of new infections becomes smaller. This is the term $0.1 \cdot n(t-1) \cdot (1000 - n(t-1))$; the 0.011 is simply a *constant of proportionality.*

Eqs. (29) and (30) constitute a first ordre scalar difference equation. This equation is different from the ones we have already seen; in particular it does not have the form

$$y(t) = a \cdot y(t-1) + b.$$

So we have not seen a way for obtaining the formula for $n(t)$. We will not try to obtain such a formlua; but we can still find a numerical solution. Using EXCEL we tabulate and graph $n(t)$ for $t = 1, 2, ...$ The graph is presented in Figure 5.4. Note how initially the number of infected people rises very fast, but as it gets closer to 1000, there are few healthy people to get infected and so the rate of increase gets smaller and smaller. In the limit, everybody will get infected.

Figure 5.4

5.1.4 Investment and Consumption (or: "Money does not make you Happy!!!")

Consider a man who initially has 100000 drs that he can use in one of two ways:

1. invest it, with annual interest rate 10%, compounded annually; or

2. consume it immediately.

In fact he has a third option: he can invest part of his money and consume the rest.

Our man is basically interested in spending his money. If he spends a sum of money, say c this year, he will get c^2 "units of happiness"; if he spends the same money next year he will get $\frac{c^2}{2}$ "units of happiness"; if he spends it two years later he will get $\frac{c^2}{4} = \frac{c^2}{2^2}$, and, in general, if he spends the amount c in t years from now he will get $\frac{c^2}{2^t}$ "units of happiness".

What are "units of happiness"? Happiness probably cannot be measured and almost certainly not with money. We just use this term to show that spending money has some value for a person. The more money one spends, the happier he is. Why does het c^2 (rather than c or $\sqrt{c}$) units of happiness when he spend c drs? Because the more money you spend the happier you are: spending 100000 drs is not just ten times better than spending 10000 drs.

Anyway, back to our problem. What should our man do? Go out and spend all his money immediately? Probably not. There may be an advantage in investing his money now and spending them next year. The reason is, of course, that by deferring his spending, the man may accumulate more money and obtain more happiness later, as he has more to spend.

On the other hand, the man should not keep saving for ever. Every time he defers spending his money for a year, the pleasure he will get from the spending is divided by half. There must be a balance between spending everything immediately and invesing everything he has, to spend later.

Let us suppose that our man plans with a ten year horizon. In other words, for the next ten years, every year the man has $y(t)$ drs. Of these, he will invest $i(t)$ drs and spend $s(t)$ drs. His goal is to maximize the happiness he will get from this schedule. Furthermore, suppose he will always spend and invest according to the following schedule: the money he will invest will be a % of his total money; the money he will spend, will obviously be $100 - a$ of his total money. How should he choose a to maximize his happiness?

We can set this problem up as a difference equation. In the first year, the man invests $y(0)$ drs and spends $s(0)$ drs . We know the following things: first that he has a total of $y(0)$ drs. where

$$y(0) = 1000;$$

second that he spends $s(0)$ and invests $i(0)$ drs, where

$$s(0) = y(0) \cdot \frac{a}{100}, \quad i(0) = y(0) \cdot \frac{100 - a}{100}.$$

Of course $s(0) + i(0) = y(0)$. Now, the $i(0)$ drs investement will, after a year yield a total of

$$y(1) = i(0) \cdot (1 + \frac{10}{100}) = 1.1 \cdot i(0) = 1.1 \cdot \frac{100 - a}{100} \cdot y(0).$$

So the man will have

$$y(1) = 1.1 \cdot \frac{100 - a}{100} \cdot y(0);$$

out of which he will spend $s(1)$ and invest $i(1)$ drs, where

$$s(1) = y(1) \cdot \frac{a}{100}, \quad i(1) = y(1) \cdot \frac{100 - a}{100}.$$

Again $s(1) + i(1) = y(1)$. Continuing in this way, we get that at time t the man will have

$$y(t) = 1.1 \cdot \frac{100 - a}{100} \cdot y(t-1);$$

$$s(t) = y(t) \cdot \frac{a}{100}, \quad i(t) = y(t) \cdot \frac{100 - a}{100},$$

and this for t=1, 2, ... , 9. For t=10, that is at the end of his investment plan, there is no point in saving any money (we are all going to die sooner or later!), so he will spend everything:

$$s(10) = y(10), \quad i(0) = 0.$$

Remember that our man's plan is to maximize his happiness. This is given by

$$[s(0)]^2 + \frac{[s(1)]^2}{2} + \frac{[s(2)]^2}{4} + ... + \frac{[s(10)]^2}{2^{10}}.$$

This can be written in terms of y:

$$[y(0) \cdot \frac{a}{100}]^2 + \frac{[y(1) \cdot \frac{a}{100}]^2}{2} + \frac{[y(0) \cdot \frac{a}{100}]^2}{4} + ... + \frac{[y(9) \cdot \frac{a}{100}]^2}{2^9} + \frac{[y(10)]^2}{2^{10}}. \tag{31}$$

So, the man's happiness depends on his total money $y(t)$; on the other hand, this depends on a, since

$$y(t) = 1.1 \cdot \frac{100 - a}{100} \cdot y(t-1). \tag{32}$$

Do not forget the initial condition

$$y(0) = 1000. \tag{33}$$

In short: by choosing a value for a, say a= 50%, the man will get some amount of happiness over the ten-year period; by choosing some other a, say a=30%, he will get a different amount of happiness. Question: what a should our man choose to maximize his happiness?

It is not easy to solve this problem theoretically; but you should be able to solve it numerically in EXCEL. For example, I chose a=50% and used eqs.(33), (32) to compute $y(1)$, $y(2)$, ... I plot these in Figure 5.5.

Figure 5.5

Then I used $y(0)$, $y(1)$, ... , $y(10)$ to compute the man's happiness. You should do the same thing for various values of a and compute theman's happiness $h(a)$. As you see, I write happiness as a function of a; h is really a function, since for every a I get a value of happiness. Compute this function numerically, for various a values and graph a, $h(a)$. Where do you see the maximum happiness?

5.2 First Order Matrix Difference Equations

We return to the subject of first order, linear difference equations. We have already considered the equation

$$\begin{aligned} y(t) &= 0.9 \cdot y(t-1) - 0.1 \cdot z(t-1) \\ z(t) &= 0.2 \cdot y(t-1) + 0.3 \cdot z(t-1). \end{aligned} \tag{34}$$

Let us rewrite this equation, using matrix notation. Introduce a new function $x_1(t) = y(t)$, $x_2(t) = z(t)$. Define

$$x(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}, A = \begin{bmatrix} 0.9 & -0.1 \\ 0.2 & 0.3 \end{bmatrix}.$$

Then we can write (34) as

$$x(t) = A \cdot x(t-1). \tag{35}$$

This is a simple (one-function)linear homogeneous difference equation, but now it involves a *matrix function.* Here we see again the power of matrix notation. (35) can be solved very simply, if we are given an initial condition $x(0) = x_0$. Namely, we have

$$x(1) = A \cdot x(0) = A \cdot x_0,$$

$$x(2) = A \cdot x(1) = A \cdot A \cdot x_0 = A^2 x_0$$

and, in general,

$$x(t) = A^t x_0,$$

which is the explicit solution of the *linear homogeneous difference matrix equation* (35). The *linear inhomogeneous difference matrix equation* is written as

$$x(t) = A \cdot x(t-1) + b \quad x(0) = x_0 \tag{36}$$

and its solution is

$$x(t) = A^t \cdot x_0 + (A^{t-1} + A^{t-2} + ... + A + I) \cdot b; \tag{37}$$

this is obtained in exactly the same way as in the case of the scalar difference equation.

It is clear that this formulation also includes, as a special case, the difference equation with one variable, discussed earlier.

5.2.1 Dynamic Input-Output Economies

For an example of a matrix difference equation, recall the input-output model of Section 4.1.2. There we had expressed supply of xebecs, yarrows and zebus (x, y and z, respectively) by the following equation. We can write the three equations in matrix form:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0.2 & 0.3 & 0.2 \\ 0.4 & 0.1 & 0.2 \\ 0.1 & 0.3 & 0.2 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \begin{bmatrix} 10 \\ 5 \\ 6 \end{bmatrix}.$$

Or, by defining the matrices

$$u = \begin{bmatrix} x \\ y \\ z \end{bmatrix}, \quad d = \begin{bmatrix} 10 \\ 5 \\ 6 \end{bmatrix}, \quad A = \begin{bmatrix} 0.2 & 0.3 & 0.2 \\ 0.4 & 0.1 & 0.2 \\ 0.1 & 0.3 & 0.2 \end{bmatrix}. \tag{38}$$

we can write the system as

$$u = Au + d.$$

Now consider a slightly different situation: at every time t the producers do not know in advance what quantity will be demanded on them. Rather, they supply enough to cover the demand of the previous time step. So, for example, at time t, the xebec producer produces $x(t)$ xebecs that will be enough to cover the industry consumption and the personal consumption of the *previous* time $t-1$. In other words, he produces

$$x(t) = 0.2 \cdot x(t-1) + 0.3 \cdot y(t-1) + 0.2 \cdot z(t-1) + 10. \tag{39}$$

The yarrow and zebu producers behave in an analogous way. So, there is a *time lag* between demand and supply. This is somewhat simlar to the supply and demand problem we saw in Section 5.1.2, except that we do not introduce a price mechanism to enforce equality of supply and demand. Now, a little thought will convince you that the following equation holds:

$$u(t) = Au(t-1) + d, \tag{40}$$

where

$$u(t) = \begin{bmatrix} x(t) \\ y(t) \\ z(t) \end{bmatrix},$$

with $x(t)$ the supply of xebecs at time t, $y(t)$ the supply of yarrows at time t and $z(t)$ the supply of zebus at time t. A and d are the same as in eq.(38).

The question that arise is: under such a mechanism of supply and demand determination, what will happen to this economy? Will it reach equilibrium, i.e. will we get a situation where the amounts of xebecs, yarrows and zebus supplied is exactly equal to the ones demanded? This is one possibility. Another possibilit is that the producers will keep producing more and more, in response to ever increasing demand, and there will be no equilibrium.

You recognize that eq.(40) is a matrix difference equation. So we can solve it, if an initial condition is provided. To be specific, suppose that

$$u(0) = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}.$$

We have

$$u(t) = A^t \cdot u(0) + (I + A + A^2 + ... + A^t) \cdot d.$$

If there is an equilibrium value, it will be $u(\infty)$, given by

$$u(\infty) = A^\infty \cdot u(0) + (I + A + A^2 + A^3 ...) \cdot d.$$

Here, A^∞ means $\lim_{t\to\infty} A^t$. Now suppose that Suppose that $A^t \to 0$ when $t \to \infty$ (this means that every element of A^t goes to zero). Then we will have $A^\infty = 0$ and so

$$u(\infty) = 0 \cdot u(0) + (I + A + A^2 + A^3 ...) \cdot d = 0 \cdot u(0) + (I + A + A^2 + A^3 ...) \cdot d = (I + A + A^2 + A^3 ...) \cdot d.$$

What is $I + A + A^2 + A^3 + ...$ equal to? I claim that it is equal to $(I - A)^{-1}$. In fact I can prove it easily. Check the property of the inverse:

$$(I + A + A^2 + A^3 ...) \cdot (I - A) = (I + A + A^2 + A^3 ...) - A \cdot (I + A + A^2 + A^3 ...) =$$

$$I + A + A^2 + A^3 ... - A - A^2 - A^3 = I.$$

Similarly I can prove that $(I - A) \cdot (I + A + A^2 + A^3 ...) = I$. Since the two properties of the inverse are satisfied, $(I + A + A^2 + A^3 ...) = (I - A)^{-1}$. So I finally have

$$u(\infty) = (I + A + A^2 + A^3 ...) \cdot d = (I - A)^{-1} \cdot d.$$

In the particular example we have

$$u(\infty) = \frac{1}{0.384} \cdot \begin{bmatrix} 0.66 & 0.30 & 0.24 \\ 0.34 & 0.62 & 0.24 \\ 0.21 & 0.27 & 0.60 \end{bmatrix} \cdot \begin{bmatrix} 10 \\ 5 \\ 6 \end{bmatrix} = \begin{bmatrix} 24.84 \\ 20.68 \\ 18.36 \end{bmatrix}.$$

Notice that this is the same solution as the one we obtained for the "static' case, in Section 5.1.2

Of course, I can also solve eq.(40) numerically. When I did this with EXCEL, I got the following graph.

Figure 5.6

5.3 Exponential and Logarithmic Functions

In eq.(9) appears the quantity $(1 + \frac{r}{n})^{n \cdot t}$. This quantity defines a function of t, namely

$$f(t) = (1 + \frac{r}{n})^{n \cdot t} = a^t, \tag{41}$$

where $a = (1 + \frac{r}{n})^n$. This motivates us to consider the function $f(t) = a^t$, which is called an *exponential function*. Take an example: for $a = 2$, $f(t) = 2^t$. Saying that $f(t) = 2^t$ is a function of t, means that 2^t is defined not only for $t = 1, 2, ...$, but also for $t = 0$, $t = -1, -2, ...$, for $t = 1/2, 1/3$ etc. In general whenever I give you a value of t (any value, positive or negative, integer, fraction or whatever) $f(t) = 2^t$ gives back another value, as displayed in the table below. So $f(t) = 2^t$ is really a function. The same holds for $f(t) = a^t$, any t.

t	1	2	3	0	-1	1/2
$f(t) = 2^t$	2^1=2	4	8	1	1/2	$\sqrt{2}$=1.414...

Here we call a the *base* and t the *exponent*. For exponential functions the usual properties of powers hold:

1. $a^t \cdot a^s = a^{t+s}$.
2. $\frac{a^t}{a^s} = a^{t-s}$.
3. $(a^t)^s = a^{t \cdot s}$.
4. $a^{\frac{1}{t}} = \sqrt[t]{a}$.
5. $a^0 = 1$.
6. $a^1 = a$.

Let us go back to (41). Of particular interest is the case where n in (41) becomes very large, in fact it goes to infinity. In this case, two factors are influencing the value of $a = (1+\frac{r}{n})^n$. First, as n gets bigger, $\frac{r}{n}$ and consequently $1+\frac{r}{n}$, gets smaller. Second, however, $1+\frac{r}{n}$ is multiplied by itself n, i.e. more, times; since r is positive, $(1 + \frac{r}{n})^n$ gets bigger. The final result of these two opposing factors is that as n goes to infinity, $(1 + \frac{r}{n})^n$ goes to a limiting value. In particular, if $r = 1$, then $(1 + \frac{r}{n})^n = (1 + \frac{1}{n})^n$, which goes to a limiting value equal to 2.718... . For this value we use a special symbol: $e = 2.718...$. In other words

$$\lim_{n\to\infty}(1 + \frac{1}{n})^n = e. \tag{42}$$

The number e has several interesting properties. First,

$$\lim_{n\to\infty}(1 + \frac{r}{n})^n = e^r. \tag{43}$$

Note that, while (42) is the *definition* of e, (43) is a *property* of e. I.e., it *follows* from (42). Can you prove this?

Using (42), we can prove that

$$\lim_{n\to\infty}(1 + \frac{r}{n})^{n \cdot t} = e^{rt}. \tag{44}$$

In particular for $r = 1$ we have

$$e^t = \lim_{n\to\infty}(1 + \frac{1}{n})^{n \cdot t} = exp(t), \tag{45}$$

where we have defined the function $exp(t) = e^t$; this is *the* exponential function. In other words, among all exponential functions of the form a^t, e^t (where the *base* is equal to e) is the most important one. That e^t is so special follows from a property involving the rate of change, which will be discussed presently. However, let us first note that $exp(t)$ has all the usual properties of the exponential functions.

1. $e^t \cdot e^s = e^{t+s}$.

2. $\frac{e^t}{e^s} = e^{t-s}$.

3. $(e^t)^s = e^{t \cdot s}$.

4. $e^{\frac{1}{t}} = \sqrt[t]{e}$.

5. $e^0 = 1$.

6. $e^1 = a$.

It must also be stressed that the variable in e^t is t; once this is fixed we get a number. Consider the following table.

t	1	2	0
$exp(t) = e^t$	e^1=2.718...	e^2=7.389...	e^0=1

As already mentioned, in addition to the usual properties of an exponential function, $exp(t)$ has a special property, which is related to its rate of change. So let us first compute this rate. We have

$$\frac{\Delta exp(t)}{\Delta t} = \frac{exp(t+h) - exp(t)}{h} =$$
$$\frac{e^{t+h} - e^t}{h} = \frac{e^t \cdot (e^h - 1)}{h} = e^t \cdot \frac{e^h - 1}{h}.$$

Now let us consider the behavior of $e^t \frac{e^h - 1}{h}$ as h goes to *zero*. In this case, e^h tends to $e^0 = 1$. So the numerator goes to $1 - 1 =$; of course the denominator is h which also goes to 0. It turns out (this requires considerable analysis) that the numerator and denominator go to 0 at the same rate, so that

$$\lim_{h \to 0} \frac{e^h - 1}{h} = 1;$$

hence

$$\frac{\Delta exp(t)}{\Delta t} = \lim_{h \to 0} \frac{e^{t+h} - e^t}{h} = e^t \cdot \lim_{h \to 0} \frac{e^h - 1}{h} = e^t \cdot 1 = exp(t).$$

In other words, the limiting rate of change of $exp(t)$ is $exp(t)$; this is a very special property, which is used extensively in *Calculus*. Note one aspect of this property: for all t $exp(t)$ is positive, so $exp(t)$ has a positive rate of change, so it is increasing; when it increases some, its rate of change, which is the same $exp(t)$ increases too, which means that $exp(t)$ increases even faster, which means that its rate of change increasees even more, so $exp(t)$ increases even more ... . This leads to the so called *exponential growth*; you may have heard this term as a colloquial expression describing very fast growth. You can see this growth in Figure 5.7, which is a graph of $exp(t)$. In fact $exp(t)$ grows faster than, say, t, t^2, ... , t^{10}, ... and in fact faster than t^n for any n.

Figure 5.7

This completes the discussion of exponential functions. At this point we must discuss briefly a related type of functions, the so-called logarithmic functions. A logarithmic function is defined by $y = \log_a(x)$, which means

$$y = \log_a(x) \Leftrightarrow x = a^y.$$

From this definition it follows immediately that

$$\log_a(a^x) = x;$$

indeed this is a tautology. The number a is called the *base* of logarithms. In case we use $a = e$ we obtain the *natural* logarithms, denoted especially by the symbol ln:

$$y = \ln(x) = \log_e(x) \Leftrightarrow x = e^y;$$

It is easy to see that the functions $exp(.)$ and $\log(.)$ are in an *inverse* relationship. Namely

$$\log_e(exp(x)) = \log_e(e^x) = x;$$

$$exp(\log_e(x)) = e^{\log_e(e^x)} = x.$$

In other words, the one "undoes" the effect of the other. We can look at this "inversion" from a different point of view, too. Consider

$$y = e^x \Leftrightarrow x = \log_e(y).$$

Here the variables x and y are in a functional relationship; determining the one determines the other. We usually think of x as the independent variable and of y as the dependent one; when we graph $y = e^x$ we place the x-axis horizontally and the y-axis vertically. Now, if we kept the same graph, but considered the y variable as independent and the x variable as dependent, the axes would be flipped around. This can be seen in Figures 5.8 and 5.9. These two graphs are combined in Figure 5.10, where we place both curves (exponential and logarithmic); then we note that they are symmetric around the line $y = x$. Can you explain this fact? From this symmetry also follows that while e^x increases very fast, $\log_e(x)$ increases very slowly.

Figure 5.9

Figure 5.9

Figure 5.10

5.3.1 The Game of Joker

A game similar to the one described is currently (Winter 1996) being played in Greece. There is a central office that coordinates the game; in particular it keeps track of the *generation* of players. The rules of the game are as follows.

1. The player who starts the game is a player of generation 1.

2. A player of generation N can introduce up to three players in the game.

3. If a player X, of generation N, introduces a new player Y to the game, the new player gets generation $N + 1$. Y pays to X 30,000 drs. to be introduced in the game.

4. Player X gives to player Y a *Coupon*. X keeps 10,000 drs. to himself, sends 10,000 drs to the central office and deposits the "remaining" 10,000 drs. to a certain bank account;

5. This bank account belongs to player Z who introduced the 4 intermediate players up to X.

6. As soon as the central office receives the 10,000 drs, it sends player Y three Coupons and the number of *his personal* bank account.

7. A player of generation N, can introduce a player of generation 0 to the game by selling him a coupon; hence he cannot introduce to the game more than three new players.

These are the rules of the game. Is it sensible to play this game? Will you break even if you play? Will you win money?

To analyze the game, it is useful to form the following Figure 5.11.

Figure 5.11

This figure is called a *tree*. Each circle denotes a player. The players below a player are called his *descendants*. The players directly below a player (i.e. these connected to him by arrows) are called *direct descendants*. A player of generation N can have up to three direct descendants. Each of these can, in turn, have up to three descendants and so on.

To break even one has to introduce three players in the game. In this case he gets 10,000 drs. and breaks even, since he originally paid 30,000 drs.

After one introduces three players in the game, he will make no more money until he obtains descendnants of the *5th order*. I.e. if he has generation N, his must get descendants of generation $N+5$. How many such descendants does a player have? He has three descendants of the first order; each of these has three descendants of the first order, so the original player has $3 \cdot 3 = 3^2 = 9$ descendants of the second order. He has $3^3 = 27$ descendants of the third order, $3^4 = 81$ descendants of the fourth order and $3^5 = 243$ descendants of the fifth order. So he may get a maximum of $241 \cdot 10,000 = 2,410,000$ drs. This will happen if the tree is not *broken* at any link, i.e. every new player introduces three additional players.

This game does not create any new money. It just redistributes money between people. If somebody gets in the game early enough, there will be plenty of people that will enter after him, so there is a good chance that he will get all his 241 fifth order descendants. But if the game goes on for a while, almost everybody will have played and may be unwilling to play for a second time. For example, for a player of the 10-th generation to make money, he must get all his fifth order descendants; they must be players of the fifteenth generation. How many of them are there? 3^{15}=13,997,521, which is more than the total population of Greece. It is unlikely that a player of tenth generation will win any money; at any rate not everybody can win money in this game,because thismoney must come from somebody else.

Probably the most profitable role in this game is played by the central office, that collects 10,000 drs from every player. What does this office *do* to earn this money? Does it *sell* a product or a service? This is not a mathematical question.

5.4 Second Order Scalar Difference Equations

Recall the following difference equation:

$$y(t) = y(t-1) + y(t-2).$$

This is an example fo a *second order* difference equation. In general these have the form

$$y(t) = a_1 y(t-1) + a_2 y(t-2);$$

we are given a_1, a_2 and, in addition, *two* initial conditions: $y(0) = y_0$ and $y(1) = y_1$. From these we can compute $y(t)$ for every t.

In the case of first order difference equations, in addition to a computational procedure for finding $y(t)$, we also found the solution as a function of t (namely $y(t) = y_0 a^t$). We will do the same for the second order case, as well. Let us start with an example.

Example Consider the difference equation

$$y(t) = 5y(t-1) - 6y(t-2), \tag{46}$$

with initial conditions $y(0) = 1$, $y(1) = 4$.

Now rewrite the equation as

$$y(t) - 2y(t-1) = 3 \cdot (y(t-1) - 2y(t-2)), \tag{47}$$

with t=2, 3, 4, ... Now, *if*

$$y(t) = 2y(t-1) \tag{48}$$

for every t, then (46) will also be true for every t.

But there is another possibility, too. Rewrite (46) as

$$y(t) - 3y(t-1) = 2 \cdot (y(t-1) - 3y(t-2)), \tag{49}$$

with t=2, 3, 4, ... Now, *if*

$$y(t) = 3y(t-1) \tag{50}$$

for every t, then (46) will also be true for every t.

Hence from (46) we got two possibilites: (48) and (50). From (48) follows that $y(t) = A \cdot 2^t$ and from (50) follows that $y(t) = B \cdot 3^t$, where A and B are *some* numbers. We can actually *check* this. In (46) we can put $y(t) = A \cdot 2^t$, $y(t-1) = A \cdot 2^{t-1}$, $y(t-2) = A \cdot 2^{t-2}$. Then we get

$$A \cdot 2^t = 5A \cdot 2^{t-1} - 6 \cdot 2^{t-2} \Leftrightarrow$$

$$A \cdot 2^t - 5A \cdot 2^{t-1} + 6 \cdot 2^{t-2} = 0 \Leftrightarrow$$

$$A \cdot 2^{t-2} \cdot (2^2 - 5A \cdot 2^1 + 6 \cdot 2^0) = 0.$$

The last equation is true for any t and A, because 2 is a root of the equation

$$x^2 - 5x + 6 = 0.$$

Similarly, in (46) we can put $y(t) = B \cdot 3^t$, $y(t-1) = B \cdot 3^{t-1}$, $y(t-2) = B \cdot 3^{t-2}$. Then we get

$$B \cdot 3^t = 5B \cdot 3^{t-1} - 6 \cdot 3^{t-2} \Leftrightarrow$$

$$B \cdot 3^t - 5B \cdot 3^{t-1} + 6 \cdot 3^{t-2} = 0 \Leftrightarrow$$

$$B \cdot 3^{t-2} \cdot (3^2 - 5B \cdot 3^1 + 6 \cdot 3^0) = 0.$$

The last equation is true for any t and B, because 3 is also a root of the equation

$$x^2 - 5x + 6 = 0.$$

You must have noted that the numbers 2 and 3 are not arbitrary. Starting with the *difference* equation

$$y(t) = 5y(t-1) - 6y(t-2),$$

which is equivalent to

$$y(t) - 5y(t-1) + 6y(t-2) = 0,$$

we associated to it a *quadratic equation*, namely

$$x^2 - 5x + 6 = 0.$$

It is the roots of this last equation which were used to express the solutions of the difference equation.

But now we can also see one more possibility. Suppose $y(t) = A \cdot 2^t + B \cdot 3^t$. Then $y(t-1) = A \cdot 2^{t-1} + B \cdot 3^{t-1}$, $y(t-2) = A \cdot 2^{t-2} + B \cdot 3^{t-2}$. Now substitute these to (46). After some algebra, we finally get that the original equation is equivalent to

$$A \cdot 2^{t-2} \cdot (2^2 - 5A \cdot 2^1 + 6 \cdot 2^0) + B \cdot 3^{t-2} \cdot (3^2 - 5B \cdot 3^1 + 6 \cdot 6^0) = 0.$$

The last equation is true for any t, A and B, because 2 and 3 are the roots of $x^2 - 5x + 6 = 0$. They are the *only* roots of this quadratic equation, so the general solution must be written only in terms of 2^t and 3^t.

So far we have not said anything about the initial conditions $y(0) = 1$, $y(1) = 5$. We said that $y(t) = A2^t + B3^t$, but this holds for t=2, 3, ... If we also made this formula hold for t=0 and 1, then we would have the solution of (46) expressed by one formula which holds for every t.

But note that we still have to unspecified numbers, A and B. So we can say that

$$1 = y(0) = A2^0 + B3^0 = A + B$$

and

$$5 = y(0) = A2^1 + B3^1 = 2A + 3B.$$

In short, we have the system

$$A + B = 1, \;\; 2A + 3B = 5 \Rightarrow A = -2, B = 3.$$

From this we finally we get that

$$y(t) = -2 \cdot 2^t + 3 \cdot 3^t.$$

This holds for all t and is the solution to the difference equation

$$y(t) = 5y(t-1) - 6y(t-2), \tag{51}$$

with initial conditions $y(0) = 1$, $y(1) = 5$. This concludes the solution of the example.

From the example we can also see the general solution method for an *homogeneous linear difference equation.* Say we start with the equation

$$y(t) = a_1 y(t-1) + a_2 y(t-2)$$

with $y(0) = y_0$, $y(1) = y_1$. The solution goes like this:

1. Form the quadratic equation
$$x^2 - a_1 x - a_2 = 0.$$
2. Find its roots, say x_1, x_2.
3. Find A and B from the system
$$x_1^0 A + x_2^0 B = y_0 \quad x_1^1 A + x_2^1 B = y_1.$$
4. the solution is
$$y(t) = A \cdot x_1^t + B \cdot x_2^t.$$

5.5 Autoregression

A special, and particularly useful case of multiple regression is *autoregression*. This applies in cases where our data evolve *in time*. The situation is best understood by an example.
Example Consider a data set of the following form.

t	y
1	8.00
2	9.62
3	11.58
4	13.93
5	16.78
6	20.16

Here t is time and y is the dependent variable. We could do a regression of the form $y = at + b$, $y = at^2 + bt + c$ etc. But, looking at the data, we note that

$$9.62 \simeq 8.00 \cdot 1.2; \; 11.58 \simeq 9.62 \cdot 1.2 \;\; 13.93 \simeq 11.58 \cdot 1.2 \;\; etc.$$

In other words, it appears that each y is obtained by the previous one, multiplied by 1.2. I.e., current y is the previous y, increased by 20% of the previous y. Now, the 1.2 value does not give exact agreement with the data. But maybe there is an a such that we have the following model:

$$y(t) = a \cdot y(t-1). \tag{52}$$

. In this case, we build the following table of data.

$y(t-1)$	$y(t)$
8.00	9.62
9.62	11.58
11.58	13.93
13.93	16.78
16.78	20.16

In terms of multiple regression, the first row in the table above plays the role of x_1 and the second the role of y. Now we want to perform a regression to find a. The equations that we want to satisfy are

$$\begin{aligned} 9.62 &= a \cdot 8.00 \\ 11.58 &= a \cdot 9.62 \\ 13.93 &= a \cdot 11.58 \\ 16.78 &= a \cdot 13.93 \\ 20.16 &= a \cdot 16.78 \end{aligned}$$

We can write these equations in matrix form, as $Xa = Y$, where

$$X = \begin{bmatrix} 8.00 \\ 9.62 \\ 11.58 \\ 13.93 \\ 16.78 \end{bmatrix}, Y = \begin{bmatrix} 9.62 \\ 11.58 \\ 13.93 \\ 16.78 \\ 20.16 \\ 37.1 \end{bmatrix}$$

and a is just a number. In this case, $X'X$= [8.00 9.62 11.58 13.93 16.78] $\cdot$[8.00 9.62 11.58 13.93 16.78]$'$= 830.11, simply a number (since X is 1-by-5 and X' 5-by-1, $X'X$ is 1-by-1). Also $X'Y$= [8.00 9.62 11.58 13.93 16.78] $\cdot$[9.62 11.58 13.93 16.78 20.16]$'$= 985.56. Finally, a= $(X'X)^{-1}X'Y$= $830.11^{-1}\cdot$ 985.56= 985.56/830.11= 1.1873, pretty close to the 1.20 which we suspected. Using a=1.1873, we get the following $\hat{y}$ values:

t	y	$\hat{y}$
2	9.62	9.50
3	11.58	11.43
4	13.93	13.75
5	16.78	16.54
6	20.16	19.92

This is a pretty good fit.

This is the idea of autoregression. It is generally applied to *time data*. It is called autoregression because the time variable is essentially thrown out, and regression is performed only on the y data. The model, in general, has the form

$$y(t) = a_1 y(t-1) + a_2 y(t-2) + ... + a_M y(t-M).$$

I.e.

$$\begin{array}{rrrrr} y(M+1) = & a_1 y(M)+ & a_2 y(M-1)+ & ... & a_1 y(1) \\ ... = & ... & ... & ... & ... \\ y(N) = & a_1 y(N-1)+ & a_2 y(N-2)+ & ... & a_1 y(N-M+1) \end{array}$$

or, in matrix form $Xa = Y$, where

$$X = \left[\begin{array}{rrrr} y(M) & y(M-1) & ... & y(1) \\ y(M+1) & y(M) & ... & y(2) \\ ... & ... & ... & ... \\ y(N-1) & y(N-2) & ... & y(N-M+1) \end{array} \right], a = \left[\begin{array}{c} a_1 \\ a_2 \\ ... \\ a_M \end{array} \right], Y = \left[\begin{array}{r} y(M+1) \\ y(M+2) \\ ... \\ y(N) \end{array} \right].$$

From the above we can also obtain $X'X$ and $X'Y$ and find $a = (X'X)^{-1}X'Y$.

Let us try one more example, with a *second order* autoregression.

Example Consider a data set of the following form.

t	y
1	3.00
2	2.50
3	2.80
4	2.62
5	2.76
6	2.68
2	2.77

Maybe we have the following model:

$$y(t) = a_1 \cdot y(t-1) + a_2 \cdot y(t-2). \tag{53}$$

. In this case, we build the following table of data.

$y(t-2)$	$y(t-1)$	$y(t)$
3.00	2.50	2.80
2.50	2.80	2.62
2.80	2.62	2.76
2.62	2.76	2.68
2.76	2.68	2.77

In terms of multiple regression, the first row in the table above plays the role of x_2, the second the role of x_1 and the third the role of y. Now we want to perform a regression to find a. In matrix form, we want $Xa = Y$, where

$$X = \begin{bmatrix} 2.50 & 3.00 \\ 2.80 & 2.50 \\ 2.62 & 2.80 \\ 2.76 & 2.62 \\ 2.68 & 2.76 \end{bmatrix}, Y = \begin{bmatrix} 2.80 \\ 2.62 \\ 2.76 \\ 2.68 \\ 2.77 \end{bmatrix}$$

and $a = [a_1\ a_2]$. In this case

$$X'X = \begin{bmatrix} 44.77 & 42.47 \\ 42.47 & 41.56 \end{bmatrix}, X'Y = \begin{bmatrix} 43.90 \\ 42.34 \end{bmatrix},$$

$$(X'X)^{-1} = \begin{bmatrix} 0.73 & -0.74 \\ -0.74 & 0.78 \end{bmatrix},$$

$$(X'X)^{-1} \cdot X'Y = \begin{bmatrix} 0.46 \\ 0.55 \end{bmatrix}.$$

Using a=[0.46 0.55], we get the following $\hat{y}$ values:

t	y	$\hat{y}$
3	2.80	2.80
4	2.62	2.66
5	2.76	2.74
6	2.68	2.71
7	2.77	2.75

This is a pretty good fit.

5.5.1 ACT Enrollment

We have already seen the data of ACT enrollment by. It is reproduced below.

Semester	Nr. of Students
1991 Fall	326
1992 Spring	326
1992 Fall	366
1993 Spring	381
1993 Fall	398
1994 Spring	401
1994 Fall	475
1995 Spring	490
1995 Fall	552
1996 Spring	548

We have discussed the reasons why enrollment prediction is of interest. In the previous chapters you performed a linear and a quadratic regression on these data. Now try to find an autoregression for the same data. Compare the predictions you obtain to these of linear and quadratic regression. Which one would you use?

5.6 Problems

1. Solve the following scalar difference equations.

   (a)
   $$y(0) = 10, \quad y(t) = 0.5 \cdot y(t-1).$$

   (b)
   $$y(0) = 12, \quad y(t) = \frac{1}{2} \cdot y(t-1).$$

   (c)
   $$y(0) = 3, \quad y(t) = 0.5 \cdot y(t-1) + 1.$$

   (d)
   $$y(0) = 5, \quad y(t) = \frac{1}{3} \cdot y(t-1) + 5.$$

2. Solve the following matrix difference equations.

   (a)
   $$\begin{bmatrix} x_1(0) \\ x_2(0) \end{bmatrix} = \begin{bmatrix} 3 \\ 7 \end{bmatrix}, \quad \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} 0.1 & 0 \\ 0.2 & 0.3 \end{bmatrix} \cdot \begin{bmatrix} x_1(t-1) \\ x_2(t-1) \end{bmatrix}.$$

   (b)
   $$\begin{bmatrix} x_1(0) \\ x_2(0) \end{bmatrix} = \begin{bmatrix} 3 \\ 7 \end{bmatrix}, \quad \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} 0.1 & 0.2 \\ -0.5 & 0.3 \end{bmatrix} \cdot \begin{bmatrix} x_1(t-1) \\ x_2(t-1) \end{bmatrix}.$$

(c)

$$\begin{bmatrix} x_1(0) \\ x_2(0) \end{bmatrix} = \begin{bmatrix} 3 \\ 7 \end{bmatrix}, \quad \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} 0.1 & 0 \\ 0.2 & 0.3 \end{bmatrix} \cdot \begin{bmatrix} x_1(t-1) \\ x_2(t-1) \end{bmatrix} + \begin{bmatrix} 1 \\ 3 \end{bmatrix} +$$

3. You put \$ 1500 in the bank. In each of the following cases, determine how much money you will have after five years.
   (a) Annual interest rate 10%, annual compounding.
   (b) Annual interest rate 12%, compounding four times a year.
   (c) Annual interest rate 18%, monthly compounding.

4. Find the equilibrium level of production for the following input / ouput system:
   (a) Xebecs: initial supply: 10 Kgs per day; consumer demand: fixed at 5 Kgs per day; raw materials per Kg production: 0.1 Kgs pf xebecs, 0.5 Kgs of yarrows and 0.2 Kgs of zebus.
   (b) Yarrows: initial supply: 15 Kgs per day; consumer demand: fixed at 7 Kgs per day; raw materials per Kg production: 0.3 Kgs pf xebecs, 0.1 Kgs of yarrows and 0.2 Kgs of zebus.
   (c) Zebus: initial supply: 12 Kgs per day; consumer demand: fixed at 15 Kgs per day; raw materials per Kg production: 0.1 Kgs pf xebecs, 0.2 Kgs of yarrows and 0.3 Kgs of zebus.

5. If the following equation describes the spread of an epidemic, in a population of 2000 individuals, determine in how many time steps 50%, 70%, 95% of the population will be infected.

$$n(0) = 5, \quad n(t) = n(t-1) + 0.05 \cdot n(t-1) \cdot (2000 - n(t-1)).$$

6. Do a first order autoregression on the following data.

t	y
1	1.00
2	2.50
3	3.90
4	6.50
5	10.10
6	14.10

7. Do a second order autoregression on the following data.

t	y
1	8.00
2	9.10
3	11.00
4	9.80
5	11.10
6	10.10

Also do book problems 5.5.1-.12, 5.6.1-.18.

6 Probabilistic Models

The models we have considered so far are *deterministic* models. This means that when we fix our independent variable(s), the dependent variable is *determined* unambiguously. For example, in the production level - price mode, when I fix the level of production at q=10, the price at which my product will be sold is fixed at p=100-10=90. There is no uncertainty about the price.

However, there are many situations where the final outcome of a situation cannot be known in advance, even if all the independent variables are known. A classic example is tossing a coin: even if I know the weight of the coin, the force I apply when I toss it and so on, I cannot say with certainty that the result of the toss will be heads. The best I can do is to say that *there is 50% probability* that the result will be heads (provided that the coin is *fair*).

While I cannot say anything certain about a single toss of a coin, I can still desribe the process of coin tossing and give useful information about it. For example, I can say that if a coin is fair and it is tossed 1000 times, it is *very likely* that I will get close to 500 heads and 500 tails. This is a rather simple example of a question that is answered in *probabilistic* terms, i.e. by describing not an outcome itself, but the probability of the outcome. There are many similar, but more complicated questions, e.g. what is the probability that in ten tosses of a fair coin I will get two tails and eight heads. Questions of this type can be answered in terms of probabilities. So, this last chapter is devoted to the study of models that can be used to answer probabilistic questions.

6.1 Choose K objects out of N

What is the probability of getting K heads in N tosses of a coin? What is the coefficient of $p^K q^{N-K}$ in the expansion of $(p+q)^N$? These two questions are, perhaps surprisingly, related as we will see in ths section.

Let us start with a coin which is repeatedly tossed. Every toss can give one of two result: either heads or tails. If the coin is *fair*, heads will come up 50% of the time and tails 50% of the time. We can express this also in terms of *probability*: the probability of heads is $p = 0.5$ and of tails $q = 0.5$. Other values of p and q are also possible: for a 75% heads and 25% tails coin (this is clearly an *unfair* coin) we would have $p = 0.75$ and $q = 0.25$. It is clear that p and q are not unrelated. We must have $0 \leq p \leq 1$, $0 \leq q \leq 1$ and $p+q = 1$. What is the probability of getting heads in one toss? By definition, it is p. What is the probability of getting heads in two tosses, one after the other? Call this result HH. If the result of each coin toss is *independent* of the other toss, the probability of HH is $p \cdot p = p^2$; in other words the *joint probability* of *independent events* taking place together ("jointly") is the product of the probability of each event. Similarly, the probability of getting in two tosses first heads, then tails (i.e. HT) is $p \cdot q$; the probability of getting first tails, then heads (i.e. TH) is $q \cdot p$; finally, the probability of getting tails at both tosses (i.e. TT) is $q \cdot q = q^2$.

What is the probability of getting one toss of heads and one toss of tails in two coin tosses? There are two possibilities here: we could get first heads, then tails (HT), with probability pq; or we could get first tails, then heads (TH), with probability qp. The sum, $pq + qp = 2pq$, is *total probability* of the two *mutually exclusive events*: HT and TH. Here we see that the total probability of mutually exclusive events is the sum of the probability of each event.

In the preceding analysis we used several concepts of probability theory. First, we have used the ideas of *independent events* and *mutually exclusive events*.

1. Some events are called independent if the probability of one event taking place is the same whether the remaining events took place or not. Example: obtaining heads in the first coin toss is independent of obtaining heads in the second toss.

2. Some events are called mutually exclusive if the occurence of one excludes the occurence of all the others. Example: The events of obtaining first heads and then tails (in two coin tosses) is mutually exclusive with the event of obtaining first heads, then heads.

Second, we have used two fundamental laws of probability.

1. **Law of joint probability:** the probability of several *independent* events take place, is equal to the *product* of the individual probabilities.

2. **Law of total probability:** the probability of any one of several *mutually exclusive* events taking place, is the sum of the probabilities of the separate events.

6.1.1 Pascal's Triangle

Now let us compute the probability of having K heads in N tosses of a coin, when p is the probability of heads in one toss and q is the probability of tails in one toss. To compute this probability, first consider the expression

$$(p+q)^N = (p+q)\cdot(p+q)\cdot\ldots\cdot(p+q)(\text{there are } N \text{ terms in the product}). \tag{1}$$

If we expand the product above, we will get a sum of products, each having the form $p^n \cdot q^{N-n}$, $n = 1, 2, ..., N$. For example

$$(p+q)^1 = p + q = p^1q^0 + p^0q^1;$$

$$(p+q)^2 = (p+q)\cdot(p+q) = pp + pq + qp + qq = p^2q^0 + p^1q^1 + q^1p^1 + p^0q^2.$$

Now, each term in these expressions, represents a particular sequence of tosses. For instance, pp represents a HH toss; this has (by the law of joint probability) probability $pp = p^2$. pq represents a HT toss (with probability pq); qp represents a TH toss (with probability qp); qq represents a TT toss (with probability $qq = q^2$). The total probability of a toss with one heads outcome and one tails outcome, is (by the law of total probability) the sum of probabilities of HT and TH, i.e. $pq + qp = 2pq$.

Taking one more example, consider the expression

$$(p+q)^3 = ppp + ppq + pqp + pqq + qpp + qpq + qqp + qqq =$$

$$p^3 + 3p^2q + 3pq^2 + q^3 = p^3q^0 + 3p^2q^1 + 3p^1q^2 + p^0q^3.$$

The term p^3q^0 is the probability of three heads in three tosses (i.e. HHH), the term $3p^2q$ is the probability of two heads and one tails in three tosses (in any sequence, i.e. HHT, HTH or THH), the term $3p^1q^2$ is the probability of one heads and two tails in three tosses (in any sequence, i.e. HTT, THT or TTH), and the term q^3 is the probability of three tails in three tosses (i.e. TTT).

It appears plausible from the above analysis that, to compute the probability of getting (*in any order*) K heads in N coin tosses, we must look at $(p+q)^N$, locate the term p^Kq^{N-K} and take it together with its coefficient.

This becomes even more convincing if we think in the following terms. Getting K heads in N tosses can be done in a number of different ways; call this number $C(N,K)$. This number is (for the time being) unknown. But we can think of it in the following way: suppose we start with N empty boxes, like this:

$$\sqcup \ \sqcup \ ... \sqcup \ \sqcup \, .$$

We have K H's to place in the N boxes; the remaining boxes will be filled with T's. This will give a particular sequence of heads and tails in N coin tosses. The total number of ways we can place the H's and T's is $C(N,K)$.

But we can also think of having K p's to place in the boxes

$$\sqcup \ \sqcup \ ... \sqcup \ \sqcup \, ;$$

the remaining boxes will be filled with q's. This will give a particular sequence of p's and q's (of length N); and the product of the p's and q's is always of the form $p^K q^{N-K}$. The total number of ways we can place the p's and q's is again $C(N,K)$.

The $p^K q^{N-K}$ product is the probability of a particular sequence of K H's and $N-K$ T's; all sequences have the same probability. The total probability is the sum of the individual probabilities. Since there are $C(N,K)$ terms, all equal to $p^K q^{N-K}$, their sum is

$$\text{"Probability of } K \text{ H's and } N-K \text{ T's in any order"} = C(N,K)p^K q^{N-K}.$$

So we see that the required probability is the product of two terms:

1. $p^K q^{N-K}$, which can be computed, once p, $q = 1-p$, N and K are given, and
2. $C(N,K)$ which is a yet unknown.

We also see that $(p+q)^N$ can be written in terms of $p^K q^{N-K}$ products:

$$(p+q)^N =$$

$$C(N,N)p^N q^0 + C(N,N-1)p^{N-1}q^1 + ... + C(N,K)p^K q^{N-K} + ... + C(N,1)p^1 q^{N-1} + C(N,0)p^0 q^N. \tag{2}$$

It must be remarked that we do know some values of $C(N,K)$, because we know the expansion of $(p+q)^N$ for several values of N.

1. For $N = 1$, we have
$$(p+q)^1 = p^1 + q^1,$$
so $C(1,1) = 1$, $C(1,0) = 1$.

2. For $N = 2$, we have
$$(p+q)^2 = p^2 + 2p^1 q^1 + q^2,$$
so $C(2,2) = 1$, $C(2,1) = 2$, $C(2,0) = 1$.

3. For $N = 3$, we have
$$(p+q)^3 = p^3 + 3p^2 q^1 + 3p^1 q^2 + q^3,$$
so $C(3,3) = 1$, $C(3,2) = 3$, $C(3,1) = 3$, $C(3,0) = 1$.

From the examples above, we see that $C(N,N)=1$ and $C(N,0)$ always. This makes sense. There is only one N-long sequence of p's (namely $ppp...p$) and only one N-long sequence of q's (namely $qqq...q$). So we can say that $C(N,N)=1$ and $C(N,)=1$ for every N.

How about $C(N,K)$ for K=1, 2, ... , $N-1$? We can figure this out as follows. Consider

$$(p+q)^N = (p+q)\cdot(p+q)^{N-1}. \tag{3}$$

Now let us use (2) to write both $(p+q)^N$ as

$$(p+q)^N =$$
$$C(N,N)p^Nq^0 + C(N,N-1)p^{N-1}q^1 + ... + C(N,K)p^Kq^{N-K} + ... + C(N,1)p^1q^{N-1} + C(N,0)p^0q^N. \tag{4}$$

and $(p+q)^{N-1}$ as

$$(p+q)^{N-1} =$$
$$C(N-1,N-1)p^{N-1}q^0 + C(N-1,N-2)p^{N-2}q^1 + ...+$$
$$C(N-1,K)p^Kq^{N-1-K} + ... + C(N-1,1)p^1q^{N-2} + C(N-1,0)p^0q^{N-1}. \tag{5}$$

Substitute (4) and (5) in (3) to get

$$C(N,N)p^Nq^0 + C(N,N-1)p^{N-1}q^1 + ... + C(N,K)p^Kq^{N-K} + ... + C(N,1)p^1q^{N-1} + C(N,0)p^0q^N =$$
$$(p+q)\cdot[C(N-1,N-1)p^{N-1}q^0 + C(N-1,N-2)p^{N-2}q^1 + ...+$$
$$C(N-1,K-1)p^{K-1}q^{N-K} + C(N-1,K)p^Kq^{N-1-K} + ... + C(N-1,1)p^1q^{N-2} + C(N-1,0)p^0q^{N-1}] \tag{6}$$

Consider the term p^Kq^{N-K}. On the left side of (6) it has coeefficient $C(N,K)$. On the right side of (6), we will get *two* p^Kq^{N-K} terms: one from multiplying $C(N-1,K-1)p^{K-1}q^{N-K}$ (from $(p+q)^{N-1}$) with p (from $(p+q)$) and one from multiplying $C(N-1,K)p^Kq^{N-1-K}$ (from $(p+q)^{N-1}$) with q (from $(p+q)$). So on the right side we have $[C(N-1,K-1)+C(N-1,K)]\cdot p^Kq^{N-K}$. Since the right and left sides of (6) are equal, the coefficient of every term must be equal; in particular p^Kq^{N-L} must have equal coefficients. This finally gives us teh following relationship for the C's:

$$C(N,K) = C(N-1,K-1) + C(N-1,K). \tag{7}$$

"And what is the use of (7)?" , are you asking. Actually, (7), togther with $C(N,N)=1$, $C(N,0)$ allows us to compute $C(N,K)$ for every N and K. Consider that we already know $C(N,K)$ for N=1,2,3 and $K=N$, $N-1$, ... , 1, 0. Now we will compute $C(4,K)$ for K=4, 3, 2, 1, 0.

1. For $K=4$, we have $C(4,4)=1$.
2. For $K=3$, we have $C(4,3)=C(3,2)+C(3,3)=3+1=4$.
3. For $K=2$, we have $C(4,2)=C(3,1)+C(3,2)=3+3=6$.
4. For $K=1$, we have $C(4,1)=C(3,1)+C(3,0)=1+3=4$.
5. For $K=0$, we have $C(4,0)=1$.

Now that we know $C(4, K)$ for K=4, 3, 2, 1, 0, we can compute $C(5, K)$ for K=5, 4, 3, 2, 1, 0 and so on. Actually, this computation can be organized in the followin g triangular array.

	K=N	K=N-1	K=N-2	K=N-3	K=N-4	K=N-5	...
N=1	1	1					...
N=2	1	2	1				...
N=3	1	3	3	1			...
N=4	1	4	6	4	1		...
N=5	1	5	10	10	5	1	...
...	...	...	...	...	...	...	...

We obtain every row by setting the first element equal to 1, and the following ones by adding the following elements of the previous row: the one exactly above and the one above and to the left. This implements exactly the computation of (7).

And so we have, in one step (admittedly a long one), computed the coefficients of $p^K q^{N-K}$ in $(p+q)^N$, as well as the probability of getting K heads in a sequence of N row tosses.
Example The expansion of $(p+q)^5$ is

$$(p+q)^5 = p^5 + 5p^4q + 10p^3q^2 + 10p^2q^3 + 5pq^4 + q^5.$$

The probability of getting 4 heads in 5 tosses of a fair coin is ($N = 5$, $K = 4$, $p = 0.5 = \frac{1}{2}$, $q = 0.5 = \frac{1}{2}$) $C(5,4)p^4q^1 = 5 \cdot (\frac{1}{2})^4(\frac{1}{2}) = \frac{5}{2^5} = \frac{5}{32}$. •

One last point. Eq.(7) is good for computation, especially in connection with the triangle), but one would like to know a direct formula for $C(N, K)$, i.e. a formula where, given N and K we get *directly* $C(N, K)$. Such a formula does exist; it is

$$\frac{(N-K+1)\cdot(N-K+2)\cdot\ldots\cdot(N-1)\cdot N}{1\cdot 2\cdot\ldots\cdot K}\cdot p^K q^{N-K}.$$

We will not explain *how* the formula was obtained. But if we suppose somebody proposed it to us, we observe that it satisfies (7):

$$C(N-1, K-1) + C(N-1, K) =$$

$$\frac{(N-1-K+2)\cdot(N-1-K+3)\cdot\ldots\cdot(N-2)\cdot(N-1)}{1\cdot 2\cdot\ldots\cdot(K-1)} =$$

$$\frac{(N-1-K+1)\cdot(N-1-K+2)\cdot\ldots\cdot(N-2)\cdot(N-1)}{1\cdot 2\cdot\ldots\cdot K} =$$

$$(\frac{N-1-K+1}{K}+1)\cdot\frac{(N-1-K+2)\cdot(N-1-K+3)\cdot\ldots\cdot(N-2)\cdot(N-1)}{1\cdot 2\cdot\ldots\cdot(K-1)} =$$

$$(\frac{N-K}{K}+1)\cdot\frac{(N-1-K+2)\cdot(N-1-K+3)\cdot\ldots\cdot(N-2)\cdot(N-1)}{1\cdot 2\cdot\ldots\cdot(K-1)} =$$

$$\frac{(N-K+1)\cdot(N-K+2)\cdot\ldots\cdot(N-1)\cdot N}{1\cdot 2\cdot\ldots\cdot K} = C(N, K).$$

So we have a formula which satisfies (7), i.e. gives the same results as (7) for every value of N and K; so we can use the formula in place of (7).

6.1.2 Gambler's Profit

Suppose that you go to a casino to play a very simple gambling game. The game has the following rules.

1. At *every turn* you bet; the amount you bet is always one thousand drs..
2. After you bet a coin is tossed; the coin is fair, i.e. it has 50% probability of coming heads up and 50% probability of coming tails up.
3. If the coin comes heads up, you get twice your bet, i.e. two thousand drs.; if it comes tails up, you lose your bet.

The goal of this section is to study the expected outcome of this game. The analysis will use *probability theory* and matrix difference equations. It is assumed you have *infinite capital*, i.e. you can keep losing money without ever exhausting your resources, in other words you can keep playing forever.

Let us first compute what is the chance of winning N thousand drs. in a row. Write $p = 0.5$ and $q = 1 - p = 0.5$ for the probabilites of winning and losing one thousand drs., respectively. The probability of winning two thousand drs. in a row, is $p \cdot p = p^2 = 0.25$. The probability of winning three thousand drs. in a row is $p \cdot p \cdot p = p^3$ and so on; the probability of winning N thousand drs. in a row is p^N. Note that a different way of saying the same thing, is to say that the probability of winning N thousand drs. in N steps is p^N (why?).

Next consider the probability of winning M thousand drs. in N steps. To make this concrete, let us compute the probability of winning 1 thousand drs. in 3 steps (i.e. M=1, $N = 3$). There are three ways you can do this.

1. You can win at the first two steps and lose at the last; this has probability $ppq = p^2q$.
2. Or you can win at the first and third step and lose at the second; this has probability $pqp = p^2q$.
3. Or you can lose at the first step and win at the next two; this has probability $qpp = p^2q$.

The probability of each case is p^2q; the *total probability* is $p^2q + p^2q + p^2q = 3p^2q$.

Let us return to the probability of winning M thousand drs. in N coin tosses. The analysis presented above used $M = 1$ and $N = 3$. Let us now compute the probability for any M and N. For this to happen, we must win K times and lose L times, where the following relationships must hold:

$$K + L = N \quad K - L = M.$$

From this it follows that $K = \frac{N+M}{2}$ and $L = \frac{N-M}{2}$. But there is an important detail: both $\frac{N-M}{2}$ and $\frac{N+M}{2}$ must be integers, because they represent numbers of coin tosses. Keeping this in mind, we need to know the probability of getting K heads and $L = N - K$ tails in N coin tosses. This, as already explained, turns out to be

$$\frac{(N-K+1)\cdot(N-K+2)\cdot\ldots\cdot(N-1)\cdot N}{1\cdot 2\cdot\ldots\cdot K}\cdot p^K q^{N-K}.$$

Then, recalling that $K = \frac{N+M}{2}$ and $N-K = \frac{N-M}{2}$, substitute in the previous expression and we finally get the following expression for the probability of winning M thousand drs. in N steps:

$$\begin{cases} \frac{(\frac{N-M}{2}+1)\cdot(\frac{N-M}{2}+2)\cdot\ldots\cdot(\frac{N-M}{2}-1)\cdot(\frac{N-M}{2})}{1\cdot 2\cdot\ldots\cdot\frac{N+M}{2}} \cdot p^{\frac{N+M}{2}} q^{\frac{N-M}{2}} & \text{If } \frac{N+M}{2} \text{ is integer} \\ 0 & \text{If } \frac{N+M}{2} \text{ is not integer} \end{cases}$$

A final remark: we started with the assumption that p=q, i.e. a fair game. But the above analysis and results remain exactly the same in case $p \neq q$.

6.2 Markov Chains

We have discussed in some detail the problem of coin tossing from the gambler's point of view. We have so far paid attention to the outcome of coin tossing and used this to estimate the gambler's profit. But an alternative way of looking at the problem is to concentrate at the amount of money that the gambler has at every time step.

Now, there is an important difference between looking at coin tosses and amount of money. It is the following: what the outcome of each coin toss will be, is *independent* of the outcomes of previous tosses. On the other hand, the amount of money the gambler has at a given time, *depends* on the amount of money he had at the previous time. In other words, there is a dependence through time. For example, if the gambler had 1000 drs at the previous time, there is no way he can have 10000 drs at this time.

This idea of time dependence can be used to introduce a new type of model, the so-called *Markov chain*, which can give quite powerful results. Before we trun back to the gambling problem, let us look at a simpler example, involving our old friends: Alex and Basil.

6.2.1 Basil goes Looking for Alex

It is 12 midnight right now and Alex is at Home. He decides to go out; he has two choices: Nani and Vareladiko.

There is a 0.25 probability of going from Home to Vareladiko and 0.75 probability of going from Home to Nani. If he goes to Nani, he will stay there for one hour; then he will decide what to do next. There is 0.75 probability of going from Nani to Vareladiko and 0.25 probability of *staying to Nani for another hour*. If he goes to Vareladiko, he will stay there for one hour; then he will think where to go next. There is 0.25 probability of going from Vareladiko to Home, 0.25 probability of going from Vareladiko to Nani and 0.50 probability of *staying at Vareladiko for another hour*. As soon as he goes Home, the same story goes one once again: he stays there for onehour and then has a 0.25 probability of going from Home to Vareladiko and 0.75 probability of going from Home to Nani. And so it goes: at the end of every hour, Alex decides what to do for the next hour, always with the same probabilities, which depend on where he at the current hour.

Basil went someplace else (never mind where). At 5 0'clock in the morning he decides to go look for Alex. Where should he go first? In other words, where does he have the maximum probability of finding Alex?

To answer this question, we must compute the probabilities of Alex being at: Home, Nani, Vareladiko at 5 o'clock in the morning. To answer this question, it is useful to make a diagram like this.

Figure 6.1

We will call the circles *states*; they represent Home, Nani and Vareladiko and are numbered 1, 2, 3. The arrows represent possible paths that Alex can take and the numbers represent the probabilities of taking each of these paths. For example, the arrow from 2 to 3, with number 0.75 tells us that Alex has 0.75 probability of going from Nani to Vareladiko, *in one step* (of course Alex can go from Nani to Home and then to Vareladiko, i.e go from Nani to Vareladiko in *two* steps). the arrow from 2 to 2, with number 0.25 tells us that Alex has 0.25 probability of going from Nani to Nani in one step. Since there is no arrow from 2 to 1, it is impossible (there is 0 probability) of going from Nani to Home in one step.

Since Alex starts at Home, we can put our finger at circle 1 and trace possible paths that Alex could take; e.g. $1 \to 2$, $2 \to 2$, $2 \to 3$, $3 \to 1$ etc. The problem is that there are too many paths to consider. So we need another method to attack the problem.

Let us introduce a matrix $y(t)$, defined as follows:

$$y(t) = \begin{bmatrix} y_1(t) \\ y_2(t) \\ y_3(t) \end{bmatrix},$$

where

1. $y_1(t)$= "Probability of being at state 1, at hour t",
2. $y_2(t)$= "Probability of being at state 2, at hour t",
3. $y_3(t)$= "Probability of being at state 3, at hour t".

What we really want is $y(5)$, because then we would know the probabilities of being at Home, Nani, Vareladiko at hour 5, i.e. at 5 o' clock, and then Basil should go to the place of highest probability.

We do not know $y(5)$. But we *do* know $y(0)$, i.e. the probabilities of Alex being at Home, Nani, Vareladiko at 12 o' clock. In fact

$$y(0) = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix},$$

since we know that Alex is *for sure* at Home at 12 o' clock.

What we will now do, is to derive a formula that gives $y(t)$ when $y(t-1)$ is known. The formula will have the form

$$y(t) = A \cdot y(t-1),$$

i.e. it will be a matrix difference equation. If we get this, and we also determine the matrix A, then from $y(0)$ we will get $y(1)$, from $y(1)$ we will get $y(2)$ and so on, until we get to the required $y(5)$.

OK, let's do it. Suppose we are at hour t. There are three possibilities: Alex is at Home, Nani or Vareladiko.

1. What is the probability that he is at Home? There is only one possibility

   (a) at time $t-1$ he was at Vareladiko (which has probability $y_3(t-1)$ of having happened) and then he went Home (which has probability 0.25 of having happened).

   So the probability of being at Vareladiko at $t-1$ *and* going Home, is the *product* of the two independent probabilities. Namely

   $$y_1(t) = 0.25 \cdot y_3(t-1).$$

2. What is the probability that he is at Nani? There are three possibilities:

   (a) at time $t-1$ he was at Home (which has probability $y_1(t-1)$ of having happened) and then he went to Nani (which has probability 0.75 of having happened); joint probability of these two things happening is $0.75 \cdot y_1(t-1)$;

   (b) or, at time $t-1$ he was at Nani (which has probability $y_2(t-1)$ of having happened) and then he stayed at Nani (which has probability 0.25 of having happened); joint probability of these two things happening is $0.25 \cdot y_2(t-1)$;

   (c) or, at time $t-1$ he was at Vareladiko (which has probability $y_3(t-1)$ of having happened) and then he went to Nani (which has probability 0.25 of having happened); joint probability of these two things happening is $0.25 \cdot y_3(t-1)$.

   So the *total* probability of being at Nani at t is the *sum* of the three probabilities. Namely

   $$y_2(t) = 0.75 \cdot y_1(t-1) + 0.25 \cdot y_2(t-1) + 0.25 \cdot y_3(t-1).$$

3. What is the probability that he is at Vareladiko? There are three possibilities:

   (a) at time $t-1$ he was at Home (which has probability $y_1(t-1)$ of having happened) and then he went to Vareladiko (which has probability 0.25 of having happened); joint probability of these two things happening is $0.25 \cdot y_1(t-1)$;

   (b) or, at time $t-1$ he was at Nani (which has probability $y_2(t-1)$ of having happened) and then he went to Vareladiko (which has probability 0.75 of having happened); joint probability of these two things happening is $0.75 \cdot y_2(t-1)$;

   (c) or, at time $t-1$ he was at Vareladiko (which has probability $y_3(t-1)$ of having happened) and then he stayed at Vareladiko (which has probability 0.50 of having happened); joint probability of these two things happening is $0.50 \cdot y_3(t-1)$.

So the probability of being at Vareladiko at t, is the *sum* of the three probabilities. Namely

$$y_3(t) = 0.25 \cdot y_1(t-1) + 0.75 \cdot y_2(t-1) + 0.50 \cdot y_3(t-1).$$

Let us put the three equations together.

$$y_1(t) = 0.25 \cdot y_3(t-1).$$

$$y_2(t) = 0.75 \cdot y_1(t-1) + 0.25 \cdot y_2(t-1) + 0.25 \cdot y_3(t-1).$$

$$y_3(t) = 0.25 \cdot y_1(t-1) + 0.75 \cdot y_2(t-1) + 0.50 \cdot y_3(t-1).$$

These can be wriiten in matrix form:

$$\begin{bmatrix} y_1(t) \\ y_2(t) \\ y_3(t) \end{bmatrix} = \begin{bmatrix} 0.00 & 0.00 & 0.25 \\ 0.75 & 0.25 & 0.25 \\ 0.25 & 0.75 & 0.50 \end{bmatrix} \cdot \begin{bmatrix} y_1(t-1) \\ y_2(t-1) \\ y_3(t-1) \end{bmatrix}.$$

This is the required difference equation that connects $y(t)$ and $y(t-1)$. The matrix A is , of course,

$$A = \begin{bmatrix} 0.00 & 0.00 & 0.25 \\ 0.75 & 0.25 & 0.25 \\ 0.25 & 0.75 & 0.50 \end{bmatrix}.$$

Do you see the connection between the trasnition *diagram* and A? More specifically, if you are given the transition diagram, can you get the *transiitn matrix* A, without going throught the probabilistic argument above? We will talk more about this a little later.

For the time being, let us return to the computation of $y(5)$. Now it is only a question of calculations. For example, to find $y(1)$, we have

$$\begin{bmatrix} y_1(1) \\ y_2(1) \\ y_3(1) \end{bmatrix} = \begin{bmatrix} 0.00 & 0.00 & 0.25 \\ 0.75 & 0.25 & 0.25 \\ 0.25 & 0.75 & 0.50 \end{bmatrix} \cdot \begin{bmatrix} y_1(0) \\ y_2(0) \\ y_3(0) \end{bmatrix} = \begin{bmatrix} 0.00 & 0.00 & 0.25 \\ 0.75 & 0.25 & 0.25 \\ 0.25 & 0.75 & 0.50 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.00 \\ 0.75 \\ 0.25 \end{bmatrix}.$$

Now we can compute $y(2)$:

$$\begin{bmatrix} y_1(2) \\ y_2(2) \\ y_3(2) \end{bmatrix} = \begin{bmatrix} 0.00 & 0.00 & 0.25 \\ 0.75 & 0.25 & 0.25 \\ 0.25 & 0.75 & 0.50 \end{bmatrix} \cdot \begin{bmatrix} y_1(1) \\ y_2(1) \\ y_3(1) \end{bmatrix} = \begin{bmatrix} 0.00 & 0.00 & 0.25 \\ 0.75 & 0.25 & 0.25 \\ 0.25 & 0.75 & 0.50 \end{bmatrix} \cdot \begin{bmatrix} 0.00 \\ 0.75 \\ 0.25 \end{bmatrix} = \begin{bmatrix} 0.0625 \\ 0.2500 \\ 0.6875 \end{bmatrix}.$$

And so we go, until we get that

$$y(5) = \begin{bmatrix} y_1(5) \\ y_2(5) \\ y_3(5) \end{bmatrix} = \begin{bmatrix} 0.1318 \\ 0.3183 \\ 0.5499 \end{bmatrix}.$$

Hence, the probability at 5 o' clock of Alex being at Home is 0.1318, of being at Nani is 0.3183, and of being at Vareladiko is 0.5499 . The highest probability is that Alex is at Vareladiko and that's where Basil should look for him.

6.2.2 Gambler's Ruin

The previous analysis can be extended to the problem of Gambler's ruin, as follows. We will change the setup somewhat, by assuming finite capital, and also introduce some new tools, to make the analysis easier.

Let us first say that we go to the casino with M thousand drs. and our target is to reach a total of N thousand drs. (i.e. to win $N-M$ rachmas, where $M<N$). What is the probability of this happening? This is the question to which we will now turn.

To study this question, we will introduce a graphical aid, illustrated in Figure 6.2.

Figure 6.2

This figure is related to a specific example, used so as to make things concrete, In the example M=2, N=4. Every circle in the figure is associated with an amount of money: 0, 1 , 2, 3 or 4 thousand drs.. The arrows indicate *one-step transitions* from one amount to another. The idea is that we start at circle 2 (we have 2 thousand drs.); from there we can move to circle 3 (if we win 1 thousand drs. at the next toss) or to circle 1 (if we lose 1 thousand drs.). Moving from 2 to 3 happens with probability p; moving from 2 to 1 happens with probability $q=1-p$. After taking one of these steps ($2 \rightarrow 3$ or $2 \rightarrow 1$) we start from our new amount and again make a transition to one of the two neighboring circles; the game stops when we get either to circle 4 (i.e. finally reach a total capital of 4 thousand drs.) or to 0 (i.e. lose all our money).

Now, what is the probability of having n thousand drs.? This depends on the time we are talking about. At the initial step, we have probability 1 (certainty) of having 2 thousand drs. and probability 0 of having any other amount. Define the following quantities:

$$\begin{array}{rl} y_0(0) = & \text{Prob("Having 0 thousand drs. at time 0")} \\ y_1(0) = & \text{Prob("Having 1 thousand drs. at time 0")} \\ y_2(0) = & \text{Prob("Having 2 thousand drs. at time 0")} \\ \ldots & \qquad\qquad\qquad\qquad\qquad\qquad\qquad\ldots \end{array};$$

and place them in a row matrix:

$$y(0) = \begin{bmatrix} y_0(0) \\ y_1(0) \\ y_2(0) \\ ... \\ y_4(0) \end{bmatrix}.$$

Similarly, define

$$\begin{array}{rl} y_0(1) = & \text{Prob("Having 0 thousand drs. at time 1")} \\ y_1(1) = & \text{Prob("Having 1 thousand drs. at time 1")} \\ y_2(1) = & \text{Prob("Having 2 thousand drs. at time 1")} \\ ... & \qquad\qquad\qquad\qquad\qquad\qquad ... \end{array};$$

and place them in a column matrix:

$$y(1) = \begin{bmatrix} y_0(1) \\ y_1(1) \\ y_2(1) \\ ... \\ y_4(1) \end{bmatrix}.$$

Continue in this manner, defining a row matrix $p(t)$ for every time t=0, 1, 2, 3, ... Each such row matrix contains the probabilities of having 0, 1, 2, 3 or 4 thousand drs. at time t.

These probabilities are related in a very specific way. To understand this, take any time t and consider the probability of having n thousand drs. at time t, i.e. the probability $y_n(t)$. There are several possibilities.

1. If n=0, then there are two possibilities.

   (a) Either at time $t-1$ we had 1 thousand drs. (this has probability $y_1(t-1)$ and then lost one thousand drs. (this has probability q): this has probability $y_1(t-1) \cdot q$;

   (b) or at time $t-1$ we had 0 thousand drs. (this has probability $y_0(t-1)$), in which case we already lost and it is certain (probability 1) that we stay at 0 drachams: this has probability $y_0(t) \cdot 1$.

   So the total probability of having 0 thousand drs. at time t is $y_0(t) = y_0(t-1) \cdot 1 + y_1(t-1) \cdot q$.

2. If n=1, then at time $t-1$ we had 2 thousand drs. (this has probability $y_2(t-1)$ and then lost one thousand drs. (this has probability q). So $y_1(t) = y_2(t-1) \cdot q$.

3. If n=2, then there are two possibilities:

   (a) at time $t-1$ we had 3 thousand drs. (this has probability $y_3(t-1)$ and then lost one thousand drs. (this has probability q): this has probability $y_3(t-1) \cdot q$; or

   (b) at time $t-1$ we had 1 thousand drs. (this has probability $y_1(t-1)$ and then won one thousand drs. (this has probability p): this has probability $y_1(t-1) \cdot p$.

   So the total probability of having 2 thousand drs. at time t is $y_2(t) = y_1(t-1) \cdot p + y_3(t-1) \cdot q$.

4. If n=3, then at time $t-1$ we had 2 thousand drs. (with probability $y_2(t-1)$) and then we won one thousand drs. (with probability p): this has probability $y_3(t) = y_2(t) \cdot p$.

5. Finally, if $n = 4$, there are two are two possibilities.

   (a) Either at time $t-1$ we had 3 thousand drs. (this has probability $y_3(t-1)$ and then won one thousand drs. (this has probability p): this has probability $y_3(t-1) \cdot p$;

   (b) or at time $t-1$ we had 4 thousand drs. (this has probability $y_4(t-1)$), in which case we already won and it is certain (probability 1) that we stay at 4 drachams: this has probability $y_4(t) \cdot 1$.

   So the total probability of having 4 thousand drs. at time t is $y_0(t) = y_3(t-1) \cdot p + y_4(t-1) \cdot 1$.

At this point it all looks confusing; actually we are pretty close to a very neat formulation. Let us recapitulate the previous equations; they form a linear system.

$$
\begin{array}{rlllll}
y_0(t) = & y_0(t-1)\cdot 1+ & y_1(t-1)\cdot q & & & \\
y_1(t) = & & & y_2(t-1)\cdot q & & \\
y_2(t) = & & y_1(t-1)\cdot p+ & & y_3(t-1)\cdot q & \\
y_3(t) = & & & y_2(t-1)\cdot p & & \\
y_4(t) = & & & & y_3(t-1)\cdot p+ & y_4(t-1)\cdot 1
\end{array}
.
$$

We can write this in matrix notation as

$$
\begin{bmatrix} y_0(t) \\ y_1(t) \\ y_2(t) \\ y_3(t) \\ y_4(t) \end{bmatrix}
=
\begin{bmatrix} 1 & q & 0 & 0 & 0 \\ 0 & 0 & q & 0 & 0 \\ 0 & p & 0 & q & 0 \\ 0 & 0 & p & 0 & 0 \\ 0 & 0 & 0 & p & 1 \end{bmatrix}
\cdot
\begin{bmatrix} y_0(t-1) \\ y_1(t-1) \\ y_2(t-1) \\ y_3(t-1) \\ y_4(t-1) \end{bmatrix}.
$$

Or, much shorter, $y(t) = Ay(t-1)$, where

$$
y(t) = \begin{bmatrix} y_0(t) \\ y_1(t) \\ y_2(t) \\ y_3(t) \\ y_4(t) \end{bmatrix}, \quad
A = \begin{bmatrix} 1 & q & 0 & 0 & 0 \\ 0 & 0 & q & 0 & 0 \\ 0 & p & 0 & q & 0 \\ 0 & 0 & p & 0 & 0 \\ 0 & 0 & 0 & p & 1 \end{bmatrix}.
$$

You can see the pattern clearly enough in the matrix; we can also do a state transition diagram. Consider the m-th row of matrix P to be related to the m-th circle in the diagram. Then, the n-th element in the m-th row gives the probability of going from circle m to circle n. Check that this works for every element of matrix A.

Figure 6.3

The advantage of all this formulation (diagram plus matrix) is that it generalizes very easily for any target capital N. We will always have a figure such as Figure 6.3 and a related matrix A of the form

$$A = \begin{bmatrix} 1 & q & 0 & 0 & ... & 0 & 0 & 0 \\ 0 & 0 & q & 0 & ... & 0 & 0 & 0 \\ 0 & p & 0 & q & ... & 0 & 0 & 0 \\ ... & ... & ... & ... & ... & ... & ... & ... \\ 0 & 0 & 0 & 0 & ... & 0 & q & 0 \\ 0 & 0 & 0 & 0 & ... & p & 0 & 0 \\ 0 & 0 & 0 & 0 & ... & 0 & p & 1 \end{bmatrix}$$

with $N+1$ rows and $N+1$ columns. In addition, for every time t, we will have a row matrix

$$y(t) = [y_0(t) \;\; y_1(t) \;\; ... \;\; y_{N-1}(t) \;\; y_N(t)]$$

where $y_n(t)$ is the probability of having n thousand drs. at time t. Finally, the evolution of these probabilities will be given by

$$y(t) = Ay(t-1). \tag{8}$$

The formulation that we presented above is quite powerful and can be used to answer several questions regarding the gambling game.

For instance, suppose that we want to compute the probability of getting to capital N, *in any number of steps*, given that we started with capital M. Note that once we get our N thousand drs., say at time t, we play no longer, so we will have N thousand drs. at times $t+1$, $t+2$ etc. So, the probability of getting to N thousand drs. *in any number of steps*, is the same as the probability of having N thousand drs. at time $t=\infty$, in other words to $y_N(\infty)$, or better to $\lim_{t\to\infty} y_N(t)$. Now, this is very easy to compute. We have:

$$y(1) = Ay(0),$$

$$y(2) = Ay(1) = AAy(0) = A^2y(0),$$

$$...$$

$$y(t) = A^t y(0).$$

In the final expression, let t go to infinity. Then we get

$$\lim_{t\to\infty} y(t) = \lim_{t\to\infty} A^t y(0).$$

If we compute $\lim_{t\to\infty} y(t)$, which is a column matrix, the last element of this column matrix is $\lim_{t\to\infty} y_N(t)$, which is what we were looking for. To compute $\lim_{t\to\infty} p(t)$ we must compute $\lim_{t\to\infty} A^t y(0)$; starting with M thousand drs. means that $y_M(0) = 1$ and $y_m(0) = 0$ for all $m \neq M$; as for $\lim_{t\to\infty} A^t$, we just need to compute P^T with a very large T, say $T-$ 100 or 200.

In fact, computing $\lim_{t\to\infty} y(t)$ not only gives us the probability of getting the target N thousand drs., but also the probability of getting any amount of n thousand drs., n=0, 1, 2, ... A very interesting conclusion of the analysis is that in the long run (as t goes to infinity) the probability of having a capital of n thousand drs. is zero for every n except n=0 or N. In other words, in the long run we will either get our target capital or be ruined ("gambler's ruin"). To see this, recall the equations for $y_n(t)$, as they follow from $y(t) = Ay(t-1)$. We have

$$y_0(t) = 1 \cdot y_0(t-1) + q \cdot y_1(t-1),$$

$$y_1(t) = q \cdot y_2(t-1),$$
$$y_2(t) = p \cdot y_1(t-1) + q \cdot y_3(t-1),$$
$$\ldots$$
$$y_{N-2}(t) = p \cdot y_{N-3}(t-1) + q \cdot y_{N-1}(t-1),$$
$$y_{N-1}(t) = p \cdot y_{N-2},$$
$$y_N(t) = 1 \cdot y_N(t-1) + q \cdot y_{N-1}(t-1).$$

Now, as t goes to infinity, $\lim_{t\to\infty} y_n(t) = \lim_{t\to\infty} y_n(t-1) = y_n$, for every n. Using this fact in the above equations, we get

$$y_0 = 1 \cdot y_0 + q \cdot y_1,$$
$$y_1 = q \cdot y_2,$$
$$y_2 = p \cdot y_1 + q \cdot y_3,$$
$$\ldots$$
$$y_{N-2} = p \cdot y_{N-3} + q \cdot y_{N-1},$$
$$y_{N-1} = p \cdot y_{N-2},$$
$$y_N = 1 \cdot y_N + q \cdot y_{N-1}.$$

From the first equation it follows that y_1=0; then from this and the second equation follows that $y_2 = 0$ and so on, until the equation before the last, which tells us that $y_{N-1} = 0$. So only y_0 and y_N are different from zero: we either win our target capital or are ruined. As already mentioned, the actual probabilities y_0 and y_N can be computed numerically from $p(t)P^t$ as t goes to infinity.

We have examined this example, the so-called gambler's ruin, in great detail. In itself it is quite amusing; in addition it is an example of a general theory, the theory of *Markov Chains.* Gambler's ruin is a very special type of Markov Chain. In particular, it is an *absorbing* Markov chain; this means that there are *states* (i.e. circles in our diagram) where you stay forever, once get there. Another important type of Markov chain is the case where from every circle you can get to any other circle; these are called *ergodic* Markov chains.

6.3 Problems

1. (a) What is the probability of getting exactly three heads in five tosses of a fair coin?
   (b) What is the probability of getting more than three heads in five tosses of a fair coin.
   (c) What is the probability of getting exactly three heads in five tosses of a coin which has probability of heads $p = \frac{2}{3}$.
   (d) What must be the probability of heads for a coin, so that your probability of getting two heads in three tosses is bigger than 0.9?

2. I want to call my grandmother and I remember the first five digits of her phone number, but I forgot the last digit. So I will call as many times as necessary until I get her. Assuming that my grandmother is always at home, what is the probability that I will not need more than three phone calls?

3. You are to play three games of chess; if you win *two games in a row* you will receive 1000000 drs. Either you play the first and third game with Gary Kasparov and the second game with your randmother, or the first and third game with your grandmother and the second game with Gary Kasparov. In which do you have higher probability of receivin the one million?

4. (a) Draw the state transition diagram that corresponds to the following transition matrix.

$$A = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0.5 & 0 & 0 \\ 0 & 0.5 & 0.2 & 0 \\ 0 & 0 & 0.8 & 1 \end{bmatrix}.$$

   Now suppose that this matrix corresponds to Alex's movements, where state 1 is Home, state 2 is Vareladiko, state 3 is Nani and state 4 is ACT. Where should Basil look for Alex after t=200 time steps?

   (b) Alex starts at home. He has the following *transition* probabilities: from home to Nani 0.5, from home to vareladiko 0.5; from Nani to Nani 0.5, from Nani to Vareladiko 0.5; from Vareladiko to Vareladiko 0.5, from Vareladiko to home 0.5. Where shoudl Basil look for Alex(where is the maximum probability to find him) at time t=5.

   (c) Alex starts at home. He has the following *transition* probabilities: from home to Nani 0.75, from home to vareladiko 0.25; from Nani to Nani 0.25, from Nani to Vareladiko 0.75; from Vareladiko to Vareladiko 0.75, from Vareladiko to home 0.25. Where shoudl Basil look for Alex(where is the maximum probability to find him) at time t=5.

   (d) Alex starts at home. He has the following *transition* probabilities: from home to Nani 0.66, from home to vareladiko 0.34; from Nani to Nani 0.66, from Nani to Vareladiko 0.34; from Vareladiko to Vareladiko 0.34, from Vareladiko to home 0.66. Where shoudl Basil look for Alex(where is the maximum probability to find him) at time t=7.

5. In a certain exotic island, the weather conditions are as follows.

   (a) If on one day it rains, for the next day there is a 80% chance of sunshine and a 20% chance of snow.

   (b) If on one day it snows, for the next day there is a 50% chance of sunshine and a 50% chance of rain.

   (c) If on one day the sun shines, for the next day there is a 70% chance of sunshine and a 30% chance of rain.

   A traveller arrives on a sunny day. What is the probability that three days later it will snow?

6. Coma Cola started an advertising campaign. Because of this campaign, if a person was not drinking Coma Cola this week, there is a 50% chance that he will start drinking it next week. On the other hand, a person who drinks Coma Cola this week, has a 10% chance of not drinking it next week. Suppose that currently 30% of the population drinks Coma Cola. What percentage of the population will be drinking Coma Cola three weeks from now? ten weeks? one year from now?

7. If the gambler goes to the casino with 10,000 drs., if he wants to win an additional 5,000 drs, and if he plays a fiar game, betting 1,000 ds. at a time, what are the chances that he will lose all his money?

8. The gambler goes to the casino and plays an *unfair* game, where his probability of win is 30%. He starts with 10,000 drs. and he wants to win an additional 5,000 drs. He can, at every turn, bet any amount of money. Is it better for him to bet 1,000 drs. at a time, 2,000 drs. or 5,000 drs?

9. A drunk man always walks on a straight line, either to the left or to the right; he takes a step to the left or the right with equal probabilities. Suppose with every step he takes he covers one meter; further suppose that he starts in a position with a wall three meters to his right and another wall two meters to his left. What is the probability that he will bump against the wall to his left?

10. The same drunk now has a wall three meters to his left, and no wall to his right. What is the probability that he will bump against the wall to his left?

11. What is the probability that in a group of ten people, at least two have the same birthday? In a group of thirty people? In a group of one hundred people?

12. Alex, Basil and Chris start shooting each other until only one person remains alive. Alex shoots first, Basil second, Crhis thrid, and continue in this sequence (of course a dead man loses his turn). Alex has 20% chance of hitiing his target, Basil a 50% and Chris never misses. Who should Alex shoot to maximize his probability of survival?

13. A secretary must put 10 letters to ten envelopes, already stamped and addressed. She puts letters in envelopes completely randomly. On the average, how many letters will be placed in the correct envelopes?

14. I write the numbers from one to ten, each on one piece of paper; I put these in a box and shuffle them. You can keep drawing pieces of paper (without looking in the box) as long as there are any left in the box; you can declare that you stop any time you choose. The number on the last paper you have drawn is the amount of money (in thousands of drs.) that I will give you. What should your strategy be so as to maximize your profit?

Afterword

7 Additional Topics

7.1 Linear Models

7.1.1 Linear Discount and Depreciation

7.1.2 EXCEL Instructions

7.2 Quadratic Models

7.2.1 Σ Notation

7.3 Dynamic Models

7.3.1 Trees, Logarithms and Information Theory

7.3.2 The Spread of Rumors and Internet

7.3.3 Fibonacci Numbers

Limiting properties: the Golden Ratio. Applications to art, continued fractions.

7.3.4 Predict your Grades

7.4 Probabilistic Models

7.4.1 Probabilistic and Fuzzy Reasoning

7.4.2 Prisoner's Dilemma

7.4.3 The Story of an Island

7.4.4 Language Generation

7.4.5 Estes' Learning Model

7.4.6 States of Learning

7.4.7 States of Love Stories

7.4.8 Gambler's Ruin: Probability of Ruin

7.4.9 The Game of Joker with Probabilistically Broken Links

7.4.10 VSSA and the Prisoner's Dilemma

7.4.11 Markov Fields and Voting

7.4.12 Pecking Orders and Dominance

7.4.13 Grading: Asymptotic, Proportional, Curving, Ordinal

7.4.14 Markovian Traffic

7.5 Miscellaneous

7.5.1 Zeno's paradox and limits

7.5.2 The Game of Life