

# Discrete Stochastic Modelling

## Introduction to Simulation and Random Walks

This is the first of three lab sheets to accompany computer laboratory sessions which illustrate computational techniques of use in the study of discrete stochastic models. Lab sheets will be available for downloading from the course website [www.staff.city.ac.uk/r.j.gerrard/courses/2dsm/](http://www.staff.city.ac.uk/r.j.gerrard/courses/2dsm/) shortly after publication. Attendance at the laboratory sessions is not compulsory, but you will find the coursework difficult if you do not understand the techniques introduced here.

The lab sheet was developed using Microsoft Excel for Office 97, but should work on subsequent versions as well.

### 1 The Linear Congruential Generator

There are various problems with using Excel's pseudo-random number generator:

- Excel insists on recalculating all its random numbers each time you do anything;
- the generator is not very sophisticated and does not rate highly in comparison with other generators;
- it is hard to set seeds in such a way that a sequence of pseudo-random numbers can be reproduced.

Each of these problems can be addressed within Excel, but it is easier to write our own generator from scratch. The format of the Linear Congruential Generator is

$$x_n = (a x_{n-1} + c) \pmod{m}, \quad u_n = x_n/c,$$

where  $u_n$  is the  $n$ th pseudo-random number returned. The parameters of this model are  $a$ ,  $c$ ,  $m$  and the seed which gives the starting value.

In fact we are going to short-cut this procedure somewhat, as on the first tutorial exercise sheet.

Open a new workbook and rename one of the sheets "Generator" (double-click on the tab at the bottom of the sheet). Use A1 for a title, A2, A3 and A4 for the words "Factor", "Summand" and "Seed". Use values 3893, 0.25248 and 0.6 (or something similar) for these parameters in B2-B4 and use Insert | Name | Define to ensure that the labels in column A are used for the names of the cells in column B.<sup>1</sup> In C5 put the formula =Factor\*B4+Summand, and in B5=C5-int(C5). These two formulas can be copied down to give a sequence of 200 pseudo-random uniform variables all generated by your original seed. Give the name "Random" to the range B5:B204.

### 2 The Simple Random Walk

Start a new worksheet and call it "SRW". The simple random walk has a single parameter,  $p$ , so set aside a cell to hold the value, and name the cell  $p$ . To start off with, use the value 0.5.

Column A will hold the values of the random walk, column B the increments (jumps). So, in cell B5, enter the formula =IF(Random<p, 1, -1), and copy this down into the next 50 cells. Remember that the IF statement takes the form IF(Condition, value if condition true, value if condition false).

In A4 enter the starting value (0 will do), then in A5 the formula =A4+B5, which can again be copied down for 50 rows. While this column is still selected, click on Chart Wizard and go through the steps required to obtain a Line chart of the SRW.

---

<sup>1</sup> Excel variable names are not allowed to contain spaces. If you enter a label containing spaces and then try to use the label as the name of a variable, the program will not let you use it unchanged. Better to use a space-free label instead.

Change the value of  $p$  by a small amount such as  $\pm 0.01$  and observe what happens to the SRW as shown on the chart. Since we are using the same seed all the time, changes are relatively small. Now change the seed on the Generator sheet (choose any number in the range 0 to 1) and see how the chart changes.

### 3 Barriers

We shall use columns C and D to construct another SRW, but this time with barriers. We shall need to store the location and type of each barrier. Type labels UpperLocation, LowerLocation, UpperType and LowerType into cells C1, C2, E1 and E2 and use Insert | Name | Define as before. Now decide where you want the barriers to be — 0 and 5 would be simple to begin with, so enter these in D2 and D1.

The barrier type will be a numerical value:  $-1$  for a downwards reflecting barrier,  $0$  for an absorbing barrier and  $1$  for an upwards reflecting barrier. Choose whatever types you like for your barriers and enter them in F1 and F2. Make sure that cells D1, D2, F1 and F2 are named.

The increments, which will be in column D starting from D5, are now generated using a nested IF statement:

```
=IF(C4=UpperLocation,UpperType,IF(C4=LowerLocation,LowerType,IF(Random<p, 1, -1)))
```

Check that you understand how this works, then enter the formula  $=C4+D5$  in cell C5 and copy both formulas down for 50 rows. Select the values in the C column, press Copy, then click on the chart and choose Paste Special (new series) from the Edit menu to get the new data series added to the chart along with the original random walk.

Try changing the barrier locations to see how much difference they make, both in the case where  $p = 0.5$  and in the case where  $p$  is different from  $0.5$ . See if you can predict what you are going to observe before you actually observe it.

### 4 Continuous distributions for increments

Excel has built-in routines for simulating observations from quite a large number of continuous distributions, including Normal,  $t$ ,  $F$ , chi squared and some others. As an example of how this works, let us construct a RW with Normal increments, using the same “SRW” worksheet as before.

Set aside cells for  $\mu$  and  $\sigma$ . If you want to compare the Normal RW with the SRW, it makes sense to use the same mean and variance for the increments, which means that  $\mu$  should be  $2p-1$  and  $\sigma$  should be the square root of  $4p(1-p)$ .

The formula to use for the increments this time is  $=\mu + \sigma * \text{NORMSINV}(\text{Random})$ . Simulate 50 of these and plot the resulting RW on the same chart as the others.

Try changing the seed a few times to see how the chart changes. If you find it inconvenient to keep switching from one worksheet to another, type "Seed" in I2 on the SRW worksheet, enter the value in J2 and replace the value in B4 of the Generator worksheet by a reference " $=\text{SRW!J2}$ ": now you can change the seed without switching to the Generator worksheet. (Notice how the syntax "SheetName!" can be used to refer to values on different worksheets.)

If you have time, you might like to try increments with a double exponential distribution ( $J = I_1 - I_2$ , where  $I_1$  and  $I_2$  are exponentially distributed).

### 5 General discrete distributions for increments

Although Excel can be used to simulate from the Binomial and Poisson distributions, we are going to show how to use it to simulate from a more general discrete distribution.

To begin with, we need to set up a table listing the possible values and their probabilities. Start a new worksheet and name it "Discrete". At the top left corner of the new worksheet construct a region like this:

Prob	Cum Prob	Value
0.1	0	-2
0.25	=A2+B2	-1
0.3	=A3+B3	0
0.25	=A4+B4	1
0.1	=A5+B5	2
	=A6+B6	

and give the name "ProbTable" to the range B2:C6. This is called a Lookup Table: we are going to produce random numbers between 0 and 1, look them up in the Cumulative Probability column and return the corresponding quantity from the Value column. The worksheet function to use is VLOOKUP. The syntax of the VLOOKUP function is =VLOOKUP(Lookup value, Lookup table, Column number), where the column number refers to the column of the table to look in, starting from the first column of the table which is counted as column 1. In this case the lookup values are in the first column of ProbTable (column B) and the corresponding values of the variable are in the second column (column C). Therefore a suitable increment for the random walk will be generated by the formula

$$=VLOOKUP(\text{Random}, \text{ProbTable}, 2).$$

Simulate 50 of these and add the corresponding random walk to the diagram.

## 6 Controls

(Omit this section if you are short of time.)

It is a bit of a nuisance changing the value of  $p$ : even for small changes, we have to click on the cell, type in a new value and press Enter. Here we introduce a simple method of getting  $p$  to change at the click of a button.

In the View menu choose Toolbars and click on Forms. This produces a new toolbar with a number of buttons on it. One of the buttons shows a grey rectangle containing upward and downward pointing arrows. This is the "scroll bar" control. Click it.

Nothing happens, except for a message appearing in the status bar. The program wants to know where to put the scroll bar. So click on the top left of a suitable area and drag the pointer across to the bottom right. A scrollbar will appear.

To move the scrollbar around or resize it you can right-click on it and drag it around the screen or drag the corners to where you want them to be. If you want to delete the scrollbar, right-click then press the Delete key.

At the moment the scrollbar has no connection with the sheet. We want to arrange it so that the scrollbar reflects the value of  $p$  used in constructing the SRW:  $p=1$  at the right hand end,  $p=0$  at the left. Right-click on the scrollbar and look at the bottom of the menu which appears. If we were going to use "Assign macro..." it would be best to have written a Visual Basic macro in advance, so let us pass on to "Format control...".

When you choose "Format control..." the dialogue box which appears has 4 tabs: the most useful of these is the one labelled "Control". Associated with the scrollbar is an integer value which reflects the position of the slider along the bar, from "Minvalue" at the left to "Maxvalue" at the right hand end. You can set these values, but both have to be integers. To get non-integer values (suitable for a probability like  $p$ ) you must take the integer value from the scrollbar and divide it by "Maxvalue".

Let us use the default settings of 100 for Maxvalue, 0 for Minvalue. Choose a suitable out-of-the-way cell to hold the integer value associated with the scrollbar, then click on the input box labelled "Cell link" and click on your selected cell. Now, whenever the slider is moved, the value in the cell is updated to reflect the slider position. In order to have  $p$  updated as well, all you have to do is to enter a formula in the cell named  $p$ , setting it equal to one hundredth of the value of the cell linked to the scrollbar.

Once you have done that, move the slider and observe how  $p$  changes and how the SRW changes in response.

## 7 Multiple runs

Suppose we are interested in a quantity whose distribution cannot easily be calculated, such as the position at time 50 of the RW whose increments follow the general discrete distribution introduced in section 5. Even if we can't find the exact distribution by calculation, we can simulate the process a few times and calculate the sample mean, or draw a histogram of the results, or something similar.

The Excel feature which we shall use is called a Data Table. Its purpose is to allow the user to perform a long calculation several times, using several different values for an input variable, without having to make a huge number of copies of the worksheet. In our case the input variable we want to change is the seed, and the output variable we want to study is the position of the RW at time 50.

The one drawback of using the Data Table is that it requires both the input and output variables to be located on the same worksheet as the Data Table itself. So let us start by inserting a new worksheet to hold the data table, name it "DataTable", and write labels Input in B1 and Output in C1.

The value in the Output cell C2 must be equal to the position of the RW at time 50, so a formula like =SRW!E54 will do the job OK.

The Input cell works the other way round: we need to ensure that a value entered in the Input cell is then copied across into the Seed cell on the Generator sheet. Type in cell B2 a value (0,1) for the seed (between 0 and 1), then go back to the Generator sheet and replace the seed value with the formula =DataTable!B2.

Now set up the data table on the DataTable worksheet: in cells A5 to A20 enter the formula =RAND( ). (This is our collection of different seeds.) Then in B4 use the formula =C2, so that the output value appears there. Next, select the range from A4 to B20, choose Data | Table; leaving the Row Input Cell blank, enter \$B\$2 in the Column Input Cell and press Enter.

The right-hand column of the table now fills with the values which the output cell *would* contain *if* the input cell B2 contained the value shown in the left-hand column. Since the input cell contains only the seed for the random number generator, these are the output values for a number of independent runs of the simulation. It goes without saying that a 'proper' simulation requires a good many more runs than 16, but naturally the data table can be made as large as you like, within the limits imposed by Excel (maximum is 65,000 or so).

Once you have a collection of independently-generated output values you can analyse it in the way you would analyse any data set: find the sample mean and variance, plot a histogram of the output values, test for normality, etc.