LEARNING PROBABILITY CONCEPTS USING MICROSOFT EXCEL

Chris du Feu, Queen Elizabeth's High School, UK
Neville Hunt, Coventry University, UK
Rob Lees, The Database (Nottingham) Ltd., UK

The difficulties encountered by students of all ages in understanding probability and randomness have been widely documented. This paper describes an attempt to teach probability concepts at secondary school level using widely available spreadsheet software. The 10PLUS (Ten Probability Lessons Using Spreadsheets) project seeks to improve students' grasp of randomness and elementary probability using dynamic spreadsheet simulation and subsequent analysis of ten probability experiments. In each simulation there are several parameters to be changed, encouraging an investigative approach. The paper describes the underlying philosophy of the project and how it links in with the UK National Curriculum.

INTRODUCTION

The many and varied applications of spreadsheets to teaching statistical concepts have been demonstrated by Hunt(1995) through the DISCUS project – Discovering Important Statistical Concepts Using Spreadsheets. Among the attractions of a spreadsheet approach are:

- the familiar user interface (namely Microsoft Excel)
- the natural, interactive, dynamic, graphical working environment
- the widespread availability of Excel in schools and, increasingly, homes
- the longevity of materials produced, despite rapid technological advances
- the ability to develop IT skills whilst learning about Statistics

Although written primarily with an undergraduate audience in mind, DISCUS materials have proved extremely popular among secondary-school teachers in the UK. Feedback from teachers indicated that the probability simulations in DISCUS were being used as the basis of investigations to "stretch" able pupils as young as 12 years. It was in the light of this clear demand and interest that the 10PLUS project was born – Ten Probability Lessons Using Spreadsheets – building on the DISCUS principles, but concentrating on probability and aimed towards pupils at Key Stage 3 (aged 11-14). Green (1987), among many other authors, has drawn attention to misconceptions held by children regarding probability and randomness. For example, there is a widely held belief in a "law of small numbers" which leads one to expect long-term odds to apply even in a small number of trials. It is therefore important that the school curriculum contains a coherent strategy for developing probabilistic ideas. Such a strategy needs to go far beyond the handling of equally likely, countable events and mechanical drawing of tree diagrams which often passes as an “introduction to probability”. The 10PLUS project aims to give pupils first hand experience of a wide variety of random processes, to facilitate their analysis of seemingly haphazard outcomes, and hence to discover underlying models.

The UK National Curriculum dictates what is taught to pupils up to the age of 16. Statistics and probability are taught compulsorily within Mathematics. Whilst this ensures that every pupil meets statistical ideas, there is evidence of a mechanistic approach to the subject and it is debatable to what extent statistical literacy results. Many in the UK statistical education community would like to see Statistics taught as a discipline in its own right, outside Mathematics, but the curriculum is too full for this to be a realistic possibility. One way of ensuring a greater exposure to statistical ideas is to use Statistics as a vehicle for delivering Information Technology. All schools are keenly aware of the need to develop IT skills and the majority of pupils are equally enthusiastic. Hence IT can be harnessed as a convenient “sugar coating” to wrap around the (perceived) “nasty medicine” of Statistics.

10PLUS MODULES

The basic 10PLUS module consists of an Excel workbook containing a large number of worksheets. The opening sheet contains a “map” of the module with “hotword” buttons providing links to each activity. The user is returned to this map upon completion of each activity. Each activity occupies several worksheets containing either text-based information, an interactive simulation, or a sheet for recording and analysing results. Navigation between sheets is via consistently labelled buttons.

Each 10PLUS module is built around a simulated random process. The first stage in each module is devoted to introducing the process, considering the possible outcomes and providing any relevant background information. Hawkins et al (1992, p.62) stress the need for this “formulation and enumeration” in any probability experiment. As Green (1990) observed, use of computer simulations is not without its dangers. It is important that pupils do not view the simulation as a computer game, since this might raise suspicions that the results are “fixed” or being controlled in some way. For this reason, each 10PLUS module begins with an experiment using real apparatus. The results of these hand experiments are recorded and analysed using exactly the same spreadsheet tools as are later applied to the results of computer simulations. The simulation is thus embedded in reality. However, the hand experiment can legitimately be small-scale, given that much larger samples can be generated rapidly by the computer simulation. The first computer simulation always has fixed values of parameters so that any variation in pupils' results is clearly attributable to random causes and not due to changes in the process itself. In subsequent simulations pupils are progressively allowed to change more characteristics of the process. The pupils' spreadsheet skills are expected to develop as they work through a module, progressing from data entry to plotting charts, cutting and pasting, use of simple formulae, sorting, and use of menu options.

The ten modules planned are listed in Table 1. The modules are being produced as part of the teaching company scheme between Nottingham Trent University and The Database (Nottingham) Ltd. described by Davies et al (1998). This scheme brings together professional software producers, statistical education specialists and serving schoolteachers working as a team. At the time of writing, the first module is being finalised for testing in schools. The next five modules are nearing completion.

Table 1 List of 10PLUS modules

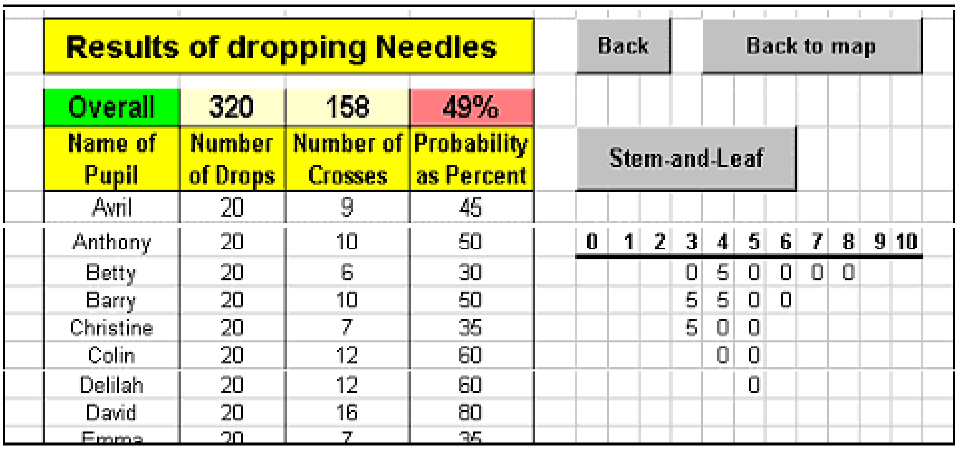
MODULE	RANDOM PROCESS	PRIMARY CONCEPT
1	Buffon's Needle	Relative frequency
2	Coupon Collecting	Random sequences
3	Lotteries	Improbable events
4	Epidemics	Spatial randomness
5	Play Your Cards Right (TV quiz game)	Degrees of dependence
6	Yahtzee (dice game)	Expectation
7	Birthday Problems	Language of probability
8	Battleships	Random pattern generation
9	Gambler's Ruin	Random walks
10	Queues	Probability distributions

BUFFON'S NEEDLE

As an example of the 10PLUS design principles, we consider the contents of the Buffon's Needle module.

1. Introduction: Following a brief look at Count Buffon himself and the historical background, pupils are introduced to the famous needle experiment. They are confronted with the uncountable number of outcomes and the need for an experimental approach to estimate the probability of a “crossing”. Pupils are invited to make a subjective estimate of the probability for a given needle and floor. Answers may reveal that some pupils have misunderstood either probability measure or some aspect of the experiment, which can be sorted out at this early stage.

2. Hand Experiment: In planning the experiment pupils are encouraged to consider the practical difficulties of dropping needles “at random”, of judging whether a needle has “crossed” a gridline, and of deciding what sample size is sufficient. The use of relatively small samples at this stage is advisable, both to avoid boredom and to magnify the sampling variation. For obvious reasons pupils should drop matchsticks onto cardboard grids rather than needles on the floor! Data are entered into a spreadsheet template as shown in Figure 1. A needle:width ratio of 3:4 conveniently makes the probability 0.48, leading pupils to believe the “answer” to be 0.5 at this point!



Overall	320	158	49%
Name of Pupil	Number of Drops	Number of Crosses	Probability as Percent
Avril	20	9	45
Anthony	20	10	50
Betty	20	6	30
Barry	20	10	50
Christine	20	7	35
Colin	20	12	60
Delilah	20	12	60
David	20	16	80
Emma	20	7	35

0	1	2	3	4	5	6	7	8	9	10
			0	5	0	0	0	0		
			5	5	0	0				
			5	0	0					
				0	0					
					0					

Figure 1. Recording results onto template

3. Simulation 1 - needle dropped on planks (fixed values of parameters): A screen display of the computer simulation is shown in Figure 2. Up to 1000 needles can be dropped at the click of a button, but it is actually more instructive for the pupil to build up

the number of drops gradually and record the cumulative results. An optional automatic plot is available to demonstrate the convergence graphically.

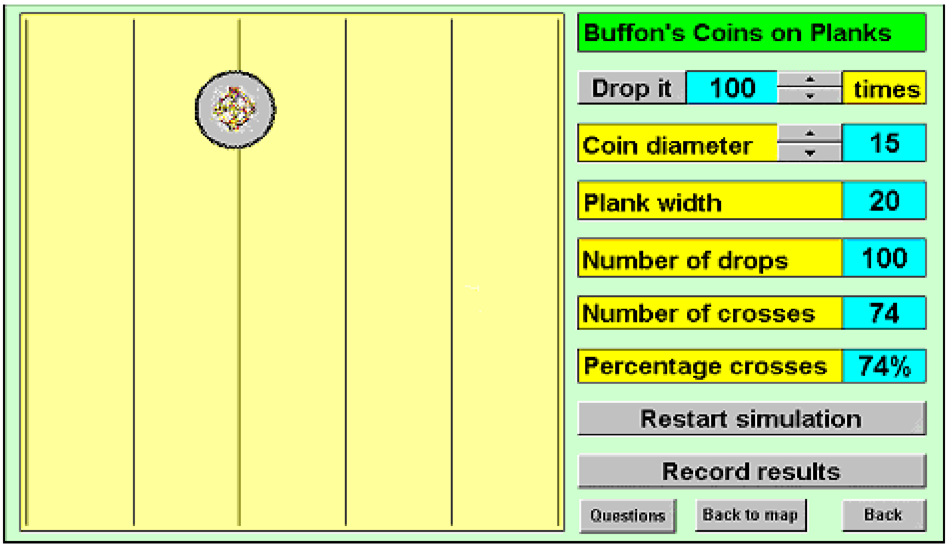


Figure 2. Fixed parameter simulation

4. Simulation 2 - needles dropped on planks (vary needle length): Here a revised results sheet is provided to allow the needle length to be recorded. An automatic scatterplot shows the (non-linear) relationship between probability estimate, P, and needle length, L, blurred of course by sampling variation. No attempt is made at this stage to model this relationship formally.

5. Simulation 3 - needles dropped on tiles (vary needle length): Only the more able pupils are expected to advance to this stage where no automatic plots or statistics are provided. It is suggested that the pupil calculates and plots new columns of data - for example, P against L^2 or P/L against L, to seek to reduce the relationship to a linear form.

6. Simulation 4 - coins dropped on planks (vary coin diameter): Having come so far, pupils are rewarded with a more tractable model. For coin diameter, D, and plank width of W, the theoretical probability of a “cross” is D/W. In the simulation W is fixed at 20. A (now routine) plot of P against D reveals the linear relationship. The linear model can be estimated by eye or using Excel’s Add Trendline facility, as in Figure 3. Sufficient hints are given to enable the more able pupil to verify this result using a geometric proof (see Glaister,1997).

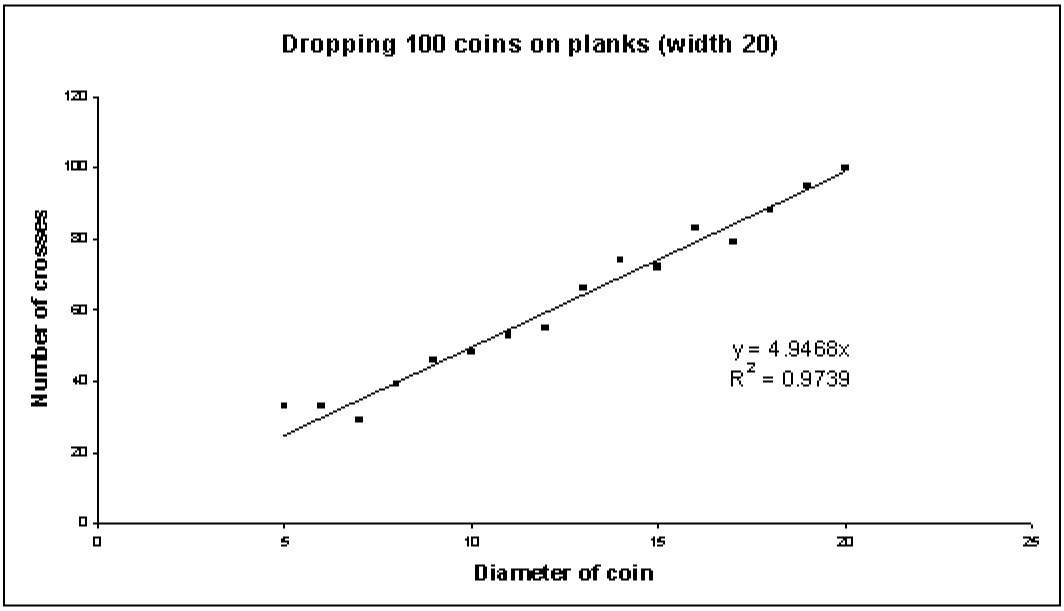


Figure 3. Fitting trendlines to simulation results

7. Simulation 5 - coins dropped on tiles (vary coin diameter): In the final activity pupils are "on their own" to hypothesise the relationship between P, D and W. The spreadsheet for results and analysis is left entirely blank. The underlying model is quadratic, beyond the scope of the National Curriculum but no more difficult to fit in Excel.

8. Finishing off: A summary is given of what has been learned and suggestions made for further work - for example, dropping playing cards onto different shaped tiles.

CONCLUSION

The authors believe that the 10PLUS modules provide teachers with the tools to investigate and solve seemingly intractable probability problems. The modules establish the vital links between experimentation, probability, data and modelling. It remains to be seen from our testing whether we are successful in countering pupils' misconceptions.

REFERENCES

Davies, N., Lees, R., Smith, S. and O'Neill, R. (1998). Integrating Microsoft Windows applications with statistics modules. *Proceedings of ICOTS5.*

Glaister, P. (1997). Playing Cards with Buffon. *Mathematical Spectrum*, *30(1)*, 8-9.

Green, D. R. (1987). Children's understanding of randomness: report of a survey of 1600 children aged 11-16 years. *Proceedings of ICOTS2.*

Green, D. R. (1990). Using computer simulation to develop statistical concepts. *Teaching Mathematics and its Applications*, *9(2)*, 58-62.

Hawkins, A., Jolliffe, F. and Glickman, L. (1992). *Teaching Statistical Concepts*. Longman.
Hunt, D. N. (1995) Teaching Statistics Using Spreadsheets. *Proceedings of 1995 ASLU Conference*, pp. 1-3. Teaching Statistics Trust.