Reprint from: W. Blum, J.S. Berry, R. Biehler, I.D. Huntley, G. Kaiser-Messmer, L. Profke (eds.): Applications and Modelling in Learning and Teaching Mathematics. Chichester: Ellis Horwood 1989, 361 - 367

58

Computer Simulation as Tool and Object of Teaching and Learning Probability and Statistics

R. Biehler
University of Bielefeld, FRG

SUMMARY

From a theoretical and epistemological point of view, the role of computer simulation in secondary probability and statistics education is critically discussed. As a background, the experiences with non-computer simulation in probability education and the theoretical debates on the role of simulation in science education are taken into account. The relation of simulation to learning about real systems and with real data is analysed.

1. INTRODUCTION

Computer simulation in teaching probability and statistics can be approached from two different perspectives:

(1) **as an increasingly important scientific tool**: computers are used in the context of probability modelling. Computer simulation is considered as a method which may extend or, at least partly, substitute other methods of modelling, namely physical simulation and classical analytical methods for representing and studying models. A main educational argument for considering computer simulation as new curriculum content is in its possibilities to handle more realistic situations, when analytical methods are beyond the scope of the students.

(2) **as an educational tool or medium**: in particular, its potential for providing environments where students can obtain stochastic experiences or transform their probabilistic intuitions is often underlined. Examples which are often suggested are:
- visualising random walks in the plane;
- structures in series of random numbers/data;

- decision situations in probability games or other situations;
- the central limit theorem, law of large numbers;
- random sampling.

In there extreme educational versions, both approaches seem fairly opposite: (1) often includes the construction of the model; whereas in (2) the underlying model is often concealed. But there is in fact a lot of overlapping and fruitful interaction between these two approaches.

2. COMPUTER SIMULATION FROM THE CONTENT PERSPECTIVE

2.1 Physical simulation

Physical simulation (with urns, dice and spinners etc.) has quite a tradition in (at least the theory of) secondary probability education (see, e.g., Travers & Gray 1981). Computer simulation at that level should be discussed against this background. I will summarise and distinguish the following major arguments for physical simulation:

(1) Students can express models and think in concrete terms like urns, spinners **(representational aspect)**.
(2) Processing the generated data may be easier for students than using analytical/ combinatorical methods and the only way for teaching more complex or realistic probability problems **(computational aspect)**.
(3) When confronted with 'word problems' students first have to design an experimental set-up and think of a model instead of direct experimentation or starting with more or less blind calculations. This may serve as a first step towards a theoretical approach and an introduction of the model concept **(concept of model aspect)**.
(4) Relating different random situations to each other and recognising similarities may be one important step in developing the probability concept. Physical simulation can contribute to developing a cognitive tool kit of adaptable models for new situations **(students' cognitive development aspect)**.

The major drawback of physical simulation consists, of course, in the processing time and the amount and quality of physical equipment (e.g. a collection of good spinners) which is necessary. Using computers can have advantages in several respects:

(1) Increasing the **number of repetitions**: with that, uncertainty and variation in the results can be reduced; new kinds of patterns become detectable.
(2) **Extensive explorations**: changing assumptions of the model or rules of the game, making further experiments, calculating further quantities, changing the way generated data are analysed, etc., have become options that are not easily available without computers. In this respect, computer simulation or computational modelling can be much closer to the hypothetical and exploratory features of traditional mathematical modelling than physical modelling is.
(3) New and more flexible **representational possibilities**: new symbol systems for expressing models and processes, graphical displays for instance for data, etc.

Limited to this perspective, an early use of computer simulation might strengthen the applied nature of probability learning. On the other hand, there is a risk of jeopardising the achievements of physical simulation.

2.2 The problem of computer generated random numbers

A basic theoretical and practical problem consists of course in an adequate educational perspective on the **use of computer generated random numbers.** What distinguishes the computer as a generator of random numbers from other random number generators like spinners? Usually, dice and spinners are, in the first place, not considered as generators of numbers; rather their physical properties such as symmetry and homogeneity give additional meaning and understanding to the concept of equal chances. These properties are taken into account when a decision has to be taken, whether a random device is 'good enough'. Looking only at the data is usually not sufficient. But with a 'black box' random generator on a computer, there is no direct correlation to the physical symmetry or some other physical feature. The analysis of data – which is highly theory dependent – plays the central role for judging the quality of the generator. Therefore, a cautious use of computer generators is often recommended for a later stage when experience with physical simulation can be exploited for a metaphorical understanding of computer generators. The other side of the coin has seldom been explored, i.e. integrating the computer generator in the 'zoo' of random devices from the beginning, trying to aim at a co-evolution of mutually related understanding, where the computer generators can play a particular role and may shed new light on the other physical generators, especially with regard to the data aspect. In a deep sense, urns and spinners are nearly as black a box as are computer generators.

2.3 Programming and simulation

One of the basic motives behind educational uses of physical simulation is the striving for reducing unnecessary or meaningless symbolism. Using computers in the shape of programming is therefore often regarded and experienced as counterproductive. But in 'user-friendly' educational software, some of the pioneering basic ideas behind the marriage between programming and simulation (see Engel, 1975) were neglected. Distinguishing the following aspects will be helpful: (1) concept of algorithm; (2) programming languages as new representational possibilities; (3) the metaphor of 'constructing probability machines' on a computer.

(1) Although the concept of simulation algorithm can be closely linked to the process of physical simulation, a confinement on this aspect may become an obstacle for interpreting computer models as interactive, explorable environments.
(2) Despite the problems with integrating the learning of BASIC or another general purpose language into teaching probability, the rich potential and variety of symbol systems including domain specific programming environments for probability and statistics should be further explored.
(3) The metaphor of constructing probability machines should be further explored from the perspective of reinforcing the concept of model aspect, because computer models are situated between physical and mathematical models. Besides, the idea of **(re-)construction** by means of a **construction set** (RND-commands as basic

primitives) is fundamental, and establishes a new conceptual framework which should be further explored with regard to students' cognitive development.

2.4 Software

An elementary modelling and simulation tool for secondary stochastics education which combines the flexibility and adaptability of programming languages with the advantages of a user-friendly interface seems not to exist at the moment (see Biehler *et al.*, 1988). Some of the deficiences of current software are: selection of models instead of construction; limited options for displaying and exploring the generated data; law degrees of interactivity: e.g. preselected graphs; no elaborated link to external inputs/outputs, e.g. for real data; no elaborated (iconic) interfaces which would support styles of thinking that secondary students apply in physical simulation environments. Commercial statistical packages may give support if they have enough options for generating and transforming random data and are programmable so that they can be adapted to educational purposes. Usually, however, many available systems in their raw versions are too complicated and expensive for general use in secondary education (see Biehler *et al.*, 1988).

3. SIMULATION AS A TOOL OF TEACHING AND LEARNING

3.1 Experientalising the abstract chance set up

Simulation is one of the main types of educational uses of computers, on the other hand, it is controversial because people are suspecting and criticising that 'real experience' is substituted by experience in 'artificial worlds'. The discussion on the role of computer simulation in probability should be embedded in this broad discussion.

In two recent papers, diSessa (1986, 1987) argues with this criticism, explaining his arguments with the case of the DYNATURTLE. His basic theoretical starting point is that physics is not directly about the physical world as we naturally perceive it but about 'abstractions that have been put together with great effort over hundreds of years [and] which happen to be very powerful once we have learned to interpret the world in their terms' (diSessa, 1986, p. 210). The DYNATURTLE is more of an Newtonian object than any real system. It is not a cheap substitute for the real thing, its significance lies, not least, in its difference from real systems. It is more of a temporary replacement of an abstract theoretical object: an attempt of **experientialising a theoretical object**, where the new representational and interactional possibilities of the technology are being exploited. Gaining an experience is seen as *part of* a strategy for transforming preconceptions in science which have proved difficult to affect by traditional ways of teaching. diSessa underlines that the DYNATURTLE cannot be the whole story and that a direct application of theoretical knowledge to real systems cannot be expected, but interacting with experiental software may be an important step, especially for developing 'qualitative knowledge'.

Interestingly, in probability and statistics education we find a comparable discussion on preconceptions, misconceptions or primary intuitions in probability, including a diagnosis of similar difficulties with traditional teaching and some research evidence that experientalisation by simulation might help. Many ideas for using simulation in stochastics education could be interpreted as aiming at a kind of qualitative knowledge, e.g. concerning:

- the role of sample size;
- a qualitative knowledge of the possible variation in small sample sizes ('law of small numbers');
- the role of independence and predictability;
- the idea of embedding a set of experimental data in a series of hypothetical repetitions.

And, similar to the DYNATURTLE, the computer can be programmed as a much better ideal chance-set-up with much quicker interactive and experimental opportunities than any real system. It can effectively be used for supporting the classical objectivist–frequentist interpretation of probability and statistics (repeatability under same conditions, long run properties). Nevertheless, the relation of experiencing the abstract chance-set-up with learning about real systems and real data in statistics and probability has to be analysed closely.

3.2 Relations to learning about real systems

Two problem areas which will be called the **temptation** and the **change** generate part of a problem space where future solutions have to be sought.

The temptation

The idea of experiencing an abstract system without paying to much attention to its relation to real systems is of course very attractive for mathematics education and its tradition, where the role of observation, experimentation and measurement is not well established as compared to science education. We therefore find proposals for a nearly universal role of simulation in teaching probability and statistics where hardly any systematic role for real data and 'direct experience' is left over (see e.g., Rade, 1983). We find software that explicitly introduces itself as an economic substitute of real Galton boards and other random generators. The rare examples of real data in probability text-books, e.g. diagrams showing the 'empirical law of large numbers' with real data get replaced by their pseudorandom counterparts with their dynamical displays.

The change

On the other hand, the scientific development of probability and statistics itself has been striving for adapting to practical demands and for overcoming the artificial world of simple probability models and ideal statistical inference processes. The developments around **exploratory data analysis, bootstrap methods**, and the **theory of complex systems** are evidence for this claim. These developments are closely related to the spread of information technology and often considered as highly relevant for education.

Let us now ask what may be an adequate perspective on simulation as a tool for teaching and learning. Many papers which discuss the role of simulation do not really consider whether using real data would be more appropriate or an essential extension for the objective they have in mind, but this attitude should become standard. One of the obstacles for this, namely the insufficient access to large data bases and flexible data analysis software in schools, will probably disappear in the future. But probably there are also deeper, more or less 'philosophical' obstacles on the teachers' side. Let us briefly look at two examples.

(1) **The laws of large numbers and the central limit theorem** are very wide-spread standard examples, where the use of simulation including dynamical display is suggested. But, if we intend to establish that the laws of large numbers represent regularities which can be observed in real systems to a certain extent, the first thing to do would be to explore sequences of real data and continually return to real data sets during theory development. With adequate software, it may become possible to take up some important suggestions which Freudenthal made years ago, i.e. introducing the square-root-n-law by analysing sets of real data first. The only concrete proposal for discussing the central limit theorem using real data has been made by Collis (1983): she suggests analysing frequencies of letters and words in texts with computer support for this purpose. These opportunities have not been explored very much in secondary education.

(2) **The Galton Board** and any other classical material device (urn, spinners) play a dual role in education: physical system and embodiment of abstract structure. Very often, the first aspect plays only a limited role: If one aims only at a combinatorical analysis of the paths in Galton boards, a simulated board on the screen would be enough, but a real Galton board can be used for many other purposes (see *Mathematiklehren*, 1985), e.g.:

- looking for explanations for the fact of the favourable middle;
- testing whether prediction of paths is possible;
- testing whether the path of the ball can be influenced;
- discussing which physical conditions may be responsible for equal chances and 'local' and 'global' independence;
- using this knowledge to build simple Galton boards;
- modelling and simulating a board – naively and more advanced: how are the balls really distributed (cf. Steinbring, 1985)?

In summary, Galton boards and other physical random devices are often only used for experiencing abstract chance-set-ups and not as real systems with a value in themselves. Handling this dual nature in probability education should have always been a problem. The computer has increased the gravity of this situation but also the possibilities for further differentiating and extending the two dual functions with educational value: as a tool for modelling and real-data analysis and as a medium for experiencing abstract chance-set-ups and processes.

REFERENCES

Biehler, R. (1988). Computers in Probability Education. To appear in: Kapadia, R. (ed.): *Chance Encounters: a critical review of research and pedagogical approaches to probability education*. Amsterdam: Reidel.

Biehler, R., Rach, W. & Winkelmann, B. (1988). Bewertung von Software für den Mathematikunterricht: Beurteilungsdimensionen und Ergebnisse einer internationalen Recherche. In: *Neue Medien im Unterricht*: Mathematik 1987/1988. Soest: Soester Verlagskontor, pp. 29–120.

Collis, B. (1983). Teaching descriptive and Inferential Statistics using a Classroom Microcomputer. *Mathematics Teacher*, 76, 318–322.

diSessa, A. (1986). Artificial Worlds and Real Experience. *Instructional Science*, 14, 207–227.

diSessa, A. (1987). The Third Revolution in Computers and Education. *J. Res. Sci. Teach.*, 24, 343–367.

Engel, A. (1975). Computing and probability. In: *Statistics at the School Level*. Stockholm: Almqvist and Wiksell, pp. 95–120.

mathematiklehren (1985). Galton–Brett, issue 12/1985.

Rade, L. (1983). Stochastics at the School Level in the Age of the Computer. In: Grey, D.R. *et al.* (eds.): *Proceedings of the First International Conference on Teaching Statistics*. Sheffield: University of Sheffield, vol. 1, pp. 19–33.

Travers, K.L. & Gray, K.G. (1981). The Monte Carlo Method: a fresh approach to teaching probabilistic concepts, *The Mathematics Teacher,* 74, 327–334.

Steinbring, H. (1985). Wie verteilen sich die Kugeln beim Galton–Brett wirklich? *Mathematiklehren,* 12, 31–38.