Simulations in Mathematics - Probability and Computing

Gary Kader - Boone, North Carolina, USA

1. Introduction

"Simulations in Mathematics-Probability and Computing" (SIM-PAC) (Perry, 1989), is a three-year project (1987-1990) funded by the United States' National Science Foundation's Materials Research and Development Program (Grant #MDR 8751110). Through this project, instructional strategies for the teaching and learning of probability through the study of computer simulations have been developed, tested and evaluated. The materials developed include microcomputer software with accompanying documentation, student activities, and teacher materials. The materials are designed for the advanced secondary or undergraduate level student.

The SIM-PAC approach is based on the belief that the learning of probability concepts may be facilitated by the study of computer simulations and simulation model-building is an important problem-solving approach. The exploration of simulation models is a key ingredient of the instructional strategy.

This paper describes the background and rationale for the project, its goals and objectives, and the instructional strategy utilised by SIM-PAC. An example of a typical learning activity and the capabilities of the software are illustrated. The ICOTS presentation included a demonstration of two learning activities.

2. The SIM-PAC Project

2.1 *Background and rationale*

The types of problems considered in the typical introductory probability course are often limited by the mathematical prerequisites. Interesting problems may never be considered because the mathematical models are too sophisticated. Most introductory presentations of probability treat the subject in a static fashion. They focus on the behaviour of a single trial of the experiment and the behaviour of the random variable over repetitions of the experiment is frequently ignored or treated in a superficial manner. Probability theory has traditionally been taught as a "formal" textbook mathematics course with emphasis on the mathematical model built from axioms.

The convergence theories which describe the behaviour of random experiments over a long sequence of trials often get cursory treatment and students cannot relate the concepts being studied to real experiences. It is not unusual for a student to be able to compute a probability but not be able to provide a meaningful interpretation of the probability obtained. The theory is only applied to trivial experiments such as tossing coins or rolling dice.

2.2 *Goals and objectives*

Recent developments in microcomputing technology have led to applications for both learning and research mathematics. A major goal of SIM-PAC is to effect changes in the teaching styles and the learning environment of the concepts in probability and statistics. Using graphics and symbol manipulation technology, SIM-PAC provides strategies and activities for the teaching and learning of these concepts and for developing problem-solving skills. Using computer simulations, a laboratory environment is created that allows students to experience probability concepts in a dynamic fashion.

The understanding of probability concepts and the simulation modelling process are the primary learning objectives of SIM-PAC. Five fundamental probability concepts are emphasised:

(i) relative frequency interpretation of probability;
(ii) a random variable;
(iii) variability in a random variable;
(iv) probability distribution of a random variable;
(v) expected value of a random variable.

SIM-PAC is an extension of the project "A Program to Improve Quantitative Literacy in the Schools" (Schaeffer, 1986). In this project, a joint committee of the American Statistical Association and the National Council of Teachers of Mathematics developed curriculum materials to assist teachers in the proper presentation of statistical and probabilistic concepts. These include *The Art and Techniques of Simulation* (Gnanadesikan et al., 1986) which introduces an Eight-Step simulation modelling process. In SIM-PAC modelling, these eight steps are modified as follows:

(1) State the problem clearly.
(2) Formulate the problem.
    (a) Describe the random experiment.
    (b) Describe the random variable and determine its possible outcomes.
(3) State the underlying assumptions.
(4) Develop a simulation model for:
    (a) the key component;
    (b) a trial of the experiment.

(5) Conduct a simulated trial.
(6) Record the outcome of the random variable.
(7) Repeat steps (5) and (6) a "large number of times".
(8) Summarise the results on the outcomes of the random variable.

2.3 *SIM-PAC programs, student activities, and teacher materials*

SIM-PAC software, with documentation, has been developed for sixteen simulation models. Some of the more traditional models include the Sum of Dice, the Binomial and Geometric random variables, and the Birthday problem. There are also models for many non-traditional problems such as collecting tokens, the study of queues, placing objects at random into boxes, random walks, and an equilibrium problem. Several models deal with statistical ideas such as estimation and acceptance sampling.

Each program provides a simulation of a random experiment and statistical summaries on at least one random variable. Statistical summaries include graphs, tables, and numerical measures. There is a standard user format and most features and capabilities are shared by all programs. Each program allows four active simulations, editing of parameters, zooms of graphs, comparisons of graphs, a summary table, printer output, and disk storage of data. SIM-PAC software is implemented for IBM-PC or Macintosh computers. With a PC compatible microcomputer (256k) a graphics card is required. The programs were written and compiled with Turbo Pascal (TM) version 5.0.

For each simulation model, there is at least one student activity. A student activity is developed in four phases:

(i) Present a problem.
(ii) Develop a hand simulation for modelling the problem using the Eight-Step modelling process.
(iii) Develop a computer simulation for modelling the problem.
(iv) Explore the simulation model.

In Phase (iv), carefully designed sequences of activities direct exploration of the model and encourage students to think in terms of the five fundamental probability concepts. The most important type of SIM-PAC exploration is studying the effects that changing parameters in a model have on the simulation results. Parameter playing and its resulting comparisons are employed as a strategy to teach statistical concepts, study mathematical relationships, and as a ploy to encourage question-asking and hypothesis making.

Teacher materials include "Program Descriptions and User Documentation" and "An Overview of the Basic Ideas of Probability and Modelling". Teacher versions of the student activities include guidelines for using the activities as well as solutions and analysis. Where appropriate, a discussion of the mathematical model is included.

The software and student activities have had extensive classroom testing (10 sites) and have been reviewed by teachers and independent reviewers; thus their designs are influenced by their intended classroom uses and contain features which facilitate the suggested learning activities.

3. Illustration

The following illustration shows the nature of the student activities, how the SIM-PAC simulation outputs facilitate exploration, and how statistical concepts are experienced in the process. Keep in mind that this is not a complete activity.

3.1 *Description and objectives*

Many games or hobbies are played by collecting a set of "tokens". The objective is to collect a complete set of tokens. Tokens are obtained by purchasing a "pack" containing one or more tokens. In the simplest case, there is only one token in a pack and the token you receive may be one you already have. In the more complex case, there are several tokens in a pack and the tokens within a pack are not necessarily all different. The occurrence of tokens within packs is assumed to be equally likely.

The problem to be studied is how many packs must be purchased before a complete set of tokens is obtained.

The objectives are:

(i) Investigate the probability distributions for the number of packs required when there are n tokens in a complete set and tokens are purchased one at a time (n = 4, 8, 16 and 32). Explore the relationship between the means, medians and modes as n increases and tokens are purchased one at a time.

(ii) See that as n increases and tokens are purchased one at a time that the mean number of packs required increases faster than n.

(iii) See that as n increases and tokens are purchased one at a time that the standard deviation in the number of packs required increases faster than n.

(iv) Investigate the probability distributions for the number of packs required when there are n tokens in a complete set and the number of tokens in a pack, k, increases proportionally with n (n = 4 / k = 1; n = 8 / k = 2; n = 16 / k = 4; n = 32 / k = 8).

(v) Explore the relationship between the means, medians and modes as n increases and k increases proportionally with n.

(vi) See that as n increases and k increases proportionally with n that the mean number of packs required increases slower than n.

3.2 *Procedures*

First, the problem of collecting four different tokens obtained one at a time is simulated as a *class activity* by tossing a four-sided die. The die is tossed until each side has occurred at least once. Using data collected from the entire class, several different probabilities associated with this problem are estimated.

The class then passes to *computer activity* and considers the following formulation of the problem. "If there are four tokens in a complete set and tokens are purchased one at a time, then how many purchases are typically required to obtain a complete set?"

This is first studied in *visual mode*, which provides for a transition from the hand activity to the computer simulation. In the visual mode the effect that one pack of tokens has on the problem of collecting all tokens is visually displayed.

Figure 1 shows the results from a typical trial. The graph at the top shows the number of times each of the different tokens is obtained. The graph at the bottom shows the proportion of all tokens obtained plotted against the number of packs. Notice the length of the horizontal line becomes longer after three tokens are obtained. This illustrates that after obtaining almost all tokens, it becomes more difficult to obtain the last few tokens to complete the set. *The Number of Packs* required for a complete set, the *Number (of packs) to go* and the *Proportion (of tokens) Obtained* are also displayed in the visual mode.

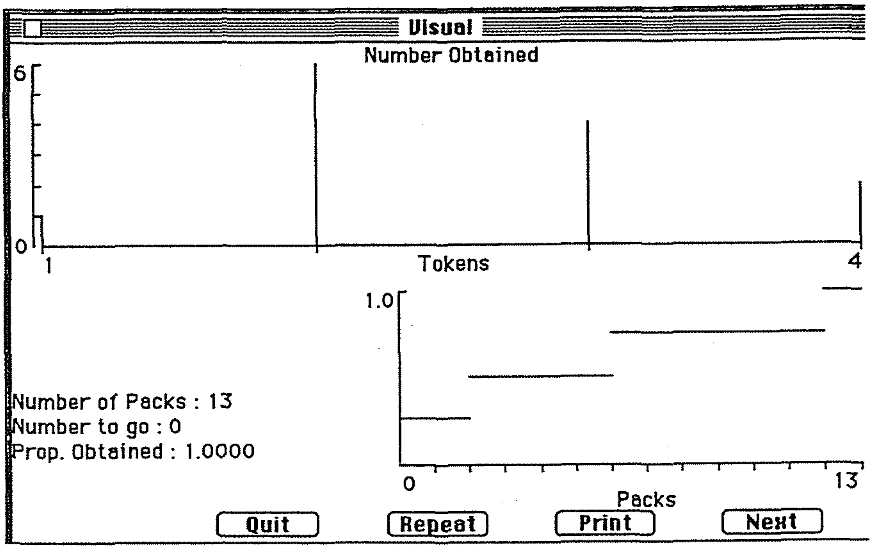

FIGURE 1
Visual display of typical trial

While "stepping" through several trials in the visual mode, the student is asked to think about the following ideas:

(i) Each time the number on the token is different from all previous tokens, by what amount does the proportion obtained increase?

(ii) After the first pack, what is the probability the next pack contains a token you don't have?

(iii) For each pack, what is the probability that the next token is different from the tokens already obtained?

(iv) After obtaining three different tokens, what is the probability the next pack contains the token you need?

After experimenting in the visual mode, the simulation continues on the *Main Screen*. Figures 2 and 3 show typical main screens after completing 100 and 5000 trials of the simulation. The main screen provides numerical information on the settings for the parameters in the simulation model and updates on several summary statistics. The bar graph at the bottom shows the relative frequency of each outcome of the random variable (number of packs until full set). The graph at the top is called the *Mean Plot*. This graph shows the convergence of the mean as the number of trials increases and gives a serial plot of the outcomes about the mean. Up to 100 outcomes can be displayed on this plot.

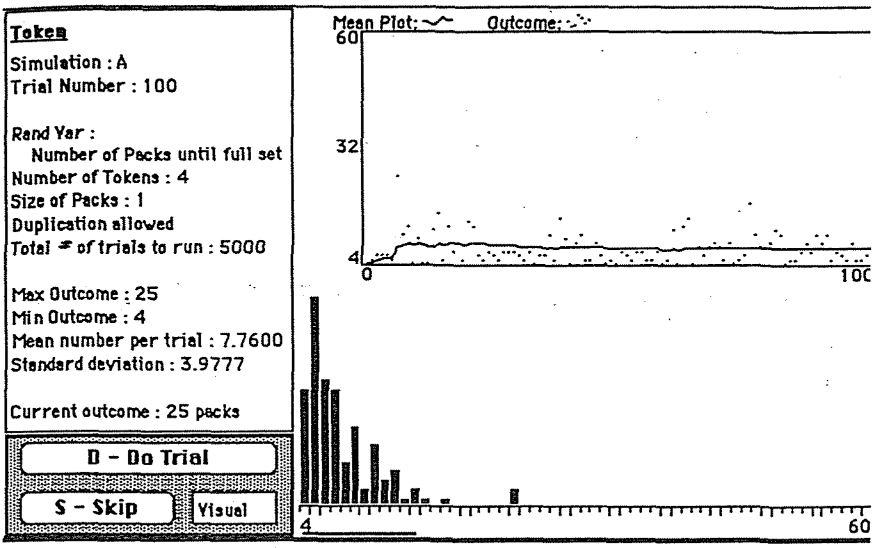

FIGURE 2
Main screen after 100 trials

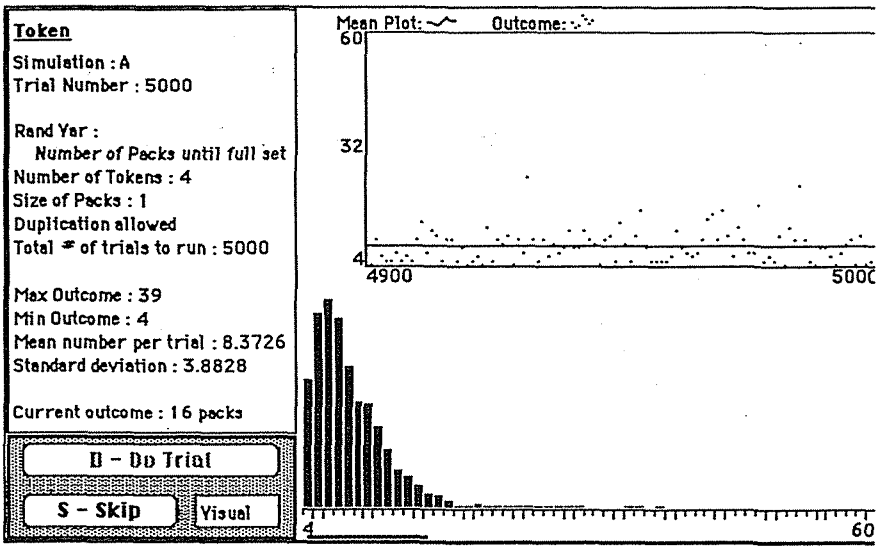

FIGURE 3
Main screen after 5000 trials

After completing the simulation, the student is asked several questions based on the results from the simulation. For example, how often was a complete set obtained after (1) exactly four packs? (2) 20 or more packs? The computer also displays the partial frequency, relative frequency and cumulative relative frequency table resulting from a simulation of 5000 trials.

In the next part of the activity, the effect of increasing the number of tokens in a complete set on the number of packs required is studied. Specifically, the results on the number of packs required for n = 4, 8, 16 and 32 are compared when each pack contains 1 token.

After completing the four simulations, the results are compared. Figure 4 is a comparison of the four bar graphs. These graphs show that as n increases there is (1) a shift in the overall location of the outcomes, (2) less skewness in the distribution, and (3) an increase in the overall variability. Similar conclusions about location and variability are reached based on comparisons of the mean plots.

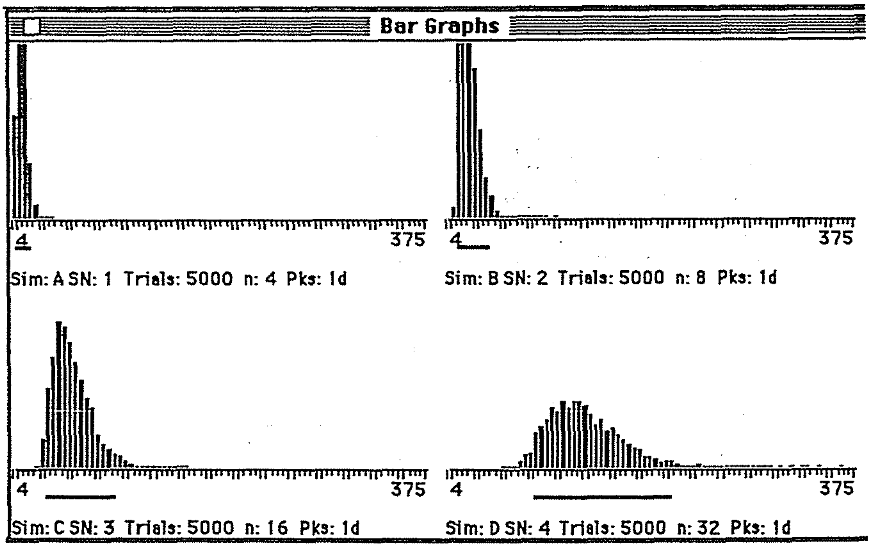

FIGURE 4
Comparisons of bar graphs

Some hand activities based on the simulation include studying the relationship between n and (1) the mean number of packs, and (2) the standard deviation in the number of packs. Table 1 shows the Summary Table resulting from the four simulations. A plot of the means against n would show that as n increases the mean is increasing at a faster rate than n. A plot of the standard deviations against n would show that as n increases the standard deviation is increasing at a faster rate than n.

TABLE 1
Summary table

Summary Table

Display Subset | Click to select. Click again to deselect | Print Subset

Ser. No.	n	Pack	Reps	Mean	Max	Min	Std. Dev.
1	4	1d	5000	8.3726	39	4	3.8828
2	8	1d	5000	21.5418	100	8	8.7972
3	16	1d	5000	54.6686	158	20	18.6893
4	32	1d	5000	130.6702	364	50	39.1036

In the final part of the activity, the number of packs required when n (the number of tokens in a complete set) and k (the number of tokens per pack) increase proportionally is studied. Simulations are performed for the following situations: (n = 4; k = 1); (n = 8; k = 2); (n = 16; k = 4); (n = 32; k = 8).

Each simulation begins in the visual mode. Figure 5 shows a typical visual for n = 32 and k = 8. Notice that the lengths of the horizontal lines in this plot tend to become longer as more tokens are obtained. Again, this demonstrates that after obtaining many different tokens, it becomes more difficult to obtain the last few tokens to complete the set.

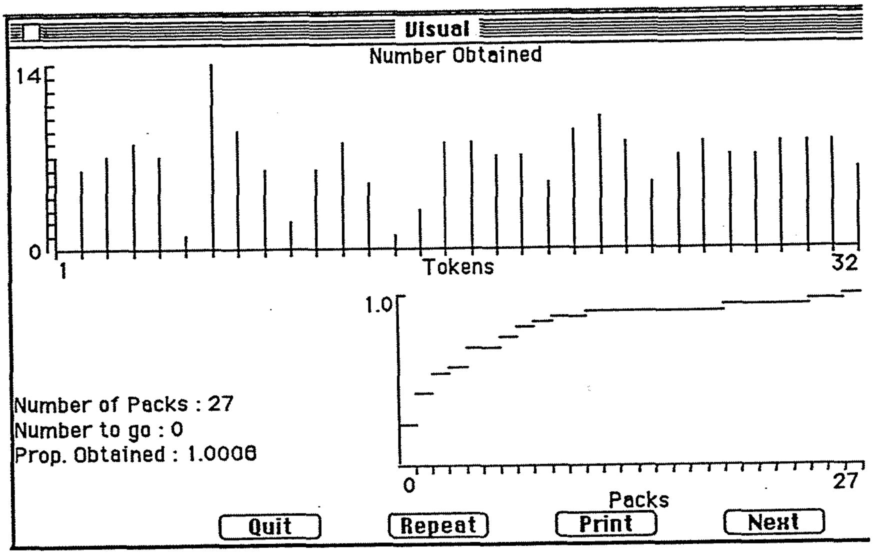

FIGURE 5
Visual display for typical trial

Later displays show comparisons of the mean plots and box plots after completing all simulations. These graphs typically show that as n increases there is (1) a shift in the overall location of the outcomes, (2) less skewness in the distribution, and (3) little change in the overall variability. A Summary Table allows for a comparison of the mean number of packs required. A plot of the means against n would show that as n increases the mean is increasing at a slower rate than n.

References

Perry, M (1989) Learning probability concepts and modelling with computer simulations. *Proceedings of the National Educational Computing Conference.* International Council on Computers for Education, Eugene, Oregon.

Schaeffer, R (1986) The Quantitative Literacy Project. *Teaching Statistics* **8(2)**.

Gnanadesikan, M R, Schaeffer, R and Swift, J (1986) *The Art and Techniques of Simulation* (the Quantitative Literacy Series - student and teacher editions). Dale Seymour Publications, Palo Alto, California.