Computational Simulation and Conditional Probability Problem Solving

Ernesto Sánchez Sánchez, Gabriel Yánez Canal
Cinvestav-IPN, Departamento de Matemática a Educativa
Av. Instituto Politecnico Nacional 2508 Col San Pedro Zacatenco
C.P. 07360, México D.F., México
esanchez@mail.cinvestav.mx, gyanez@mail.cinvestav.mx

1. Introduction

This paper we are presenting is part of a project which aims to explore the possibilities of simulation as a tool in solving statistics and probability problems. We have adopted the Hawkins' (1966) point of view: "Introducing technology effectively requires exactly the same kind of planning and understanding about how students learn, and how best to teach them, that we should use to plan any other nontechnologically-based teaching" (p. 6). We ask these questions: How must the content of a simulation-based course be? And: How must this content be organized (particularly at the end of high school and at the beginning of college) in order to be a useful instrument when solving problems? We suggest an answer in advance, carrying out an *a priori* analysis; then, we'll give an example that points out to the relevance of our proposition. In that example, two students work in solving problems of conditional probability.

2. Simulation as a tool box

The toolbox metaphor to build models related to random situations clearly reflects how we have come to conceive simulation. The simulation process links situations and problems with software commands through models. Building an appropriate **model** to carry out the simulation is crucial (Biehler, 1991); but the concept of a model is a broad one, so if we stop to analyze it, it will help us to answer the question asked. For this, we have looked-up an old article by Fischbein (1977) in which he defines a *model* as "a simplified version of the original, which permits an easier and more complete control of a set of variables". The essential roles of a model are: a) to facilitate the interpretation of certain given facts, and b) to help solve problems according to the original facts.

Fischbein defines some other features of good models: a model is a *heuristic* one if it is possible to pose problems within the model and if, additionally, it is possible to find its solutions without leaving the model. Such problems must have a meaning in the original problem. A model is a *generative* one, if it is able to represent an unlimited number of different situations, using a limited number of elements and rules. A model is a *consistent* one, if its elements are interpreted in the same way for every different situation, and naturally there are not contradictions in its operation and progress. A model is *well structured*, if it possesses an internal structure that is isomorphic with the original structure. Such structure is supplied by the properties and relationships of the elements of the model and they are not externally imposed.

Even though the simulation process cannot be conceived as a model by itself, it is clear that in the computational domain it is an instrument to build models. How can the models built by means of simulation have some of the properties of Fischbein's models?

The features *heuristic* and *well structured* that Fischbein asks for the models, give some insights to improve the power of simulation in solving problems. Particularly, when a student is working in the solution of a problem by means of simulation but he/she frequently turns to a theoretical model to get an outcome out of the simulation domain or the software language, and he/she wants to make a "simulation" *a posteriori* , he/she really hasn't built a good model; i.e., for him/her it is not a heuristic and well structured model. We think that this is due to the fact that simulation techniques are not taught to show an apparent internal structure. A member of our research team (Yáñez, 2002) has made a proposal to teach simulation in a structured way, making use of the theoretical construction carried out by von Mises (1957), which is based in a frequential focusing of probability. This focusing by von Mises is interesting because he defines operations made on collectives: *Selection, Mixture, Partition and Combination*. The basic objects of a simulation using Fathom software constitute a sequence of trials that show some attributes and these can be identified directly with the von Mises' *collectives*. So, the operations defined by von Mises are translated into operations on sequences in Fathom. We can't explain in an exhaustive way in this space, what are such operations and their corresponding operations in Fathom but we are preparing a more extensive paper on this purpose. We can only state that the theoretical work of von Mises provides a consistent structure for simulation so that, continuing with the metaphor, the toolbox acquires more power as it permits to build good models.

3. Problems of Conditional Probability

In the following example we can see that when students begin to understand and use the von Mises' operations on simulation, they can solve problems. This example refers particularly to conditional probability and the operation of *Partition* (according to von Mises' definition).

One kind of paradigmatic problems about conditional probability involves those in which a partition $\{B_1, B_2, \ldots, B_n\}$ of the Sample Space and an event A is proposed. Also the proportions p_1, p_2, …, p_n of the part of A that is in each B_i is known. The problems consist in determining the probability of A and the conditional probabilities type $P(B_i|A)$.

Constructing a solving procedure for this kind of problems is very difficult for students, even in the simplest situations when the context does not overshadow the conditional nature of the problem and where the construction of the sample space and the events of the partition can be produced easily.

In a probability course with a strong focus on simulation, the students were asked to solve the following problem with the help of Fathom software and the simulation techniques:

In a school, 40% of the students are male sex. Assuming that 10% of male students and 15% of female students are myopic,

a) What is the probability that, selecting a student at random, the student is myopic?

b) What is the probability that, selecting a female student, she is myopic?

Using Fathom, the solution requires to use the "If…" command. This command gets the structure of the problem and we think that, when a student uses it to solve the problem, he/she is finding a way to model a broad kind of situations. Next, we present a segment of transcriptions from Carlos, and then, from Laura, in their attempts to model the problem:

Carlos: "*I think we are confused*"

He wrote on the board: *M* *40%* → *10%*

F *60%* → *15%*

And he said: "*I took what was 10% of these 40 and 15% of these 60. After making this relationship I made an attribute called Students: Randompick ("M", "M", "F", "F", "F") and another attribute Myopic: RP("Fm", "Mm", "Fnm", "Mnm").*

Carlos knows how to generate columns of cases according to certain possibilities, but he does not know yet how to make basic operations with those columns. In this case, the required operation is *partition*; i. e., to consider separately two sub-sequences of the first sequence and to generate a second column, keeping in mind their differences; this operation must be identified with the "If…" command. Let's see how Laura shows that understanding. First, she observes that the procedure proposed by Carlos is not the proper one and, then, she proposes her solution:

Researcher: *What do you think about that planning?* [Carlos']

Laura: [Pointing to attribute Myopic] *There he is considering all the population. He is wrong.*

Researcher: *Why?*

Laura: *Because there he is considering that 25% of all women are myopic, the other 25% men myopic and the 50%, … let's say…*

After the discussion, Laura goes away for a moment, works alone for some few minutes and then:

Laura: [goes to the board and writes:

Sexo: RandomPick ("H", "H", "M", "M", "M")

$$\text{Miopía: if (Sexo = "H")} \begin{cases} \text{RandomPick ("Hm" , 9"Hnm"} \\ \text{RandomPick (3 "Hm" ,17" Mnm"} \end{cases} \Bigg]$$

If sex equals man, according to the whole population, 10% are myopic, so I put 1 Hm and here I put 9 Hnm. On the contrary, if it wasn't a Man it was a Woman. It says that, of that population, 15% are myopic women; so, I put 3 Mm, myopic women (in Spanish M = mujer [for woman]; H = hombre [for man]) *and 17 Mnm, women not myopic. And I came until this point. But to calculate the probability of selecting a student at random, I didn't do that.*

Afterwards, she calculates what was asked.

Laura has begun to give structure to the simulation elements; now she possesses a fundamental operation: the *partition*, which along the course it assured her success with other problems. Laura didn't made the attempt to represent the problem with a tree, table or algebraic language before proposing her solution, but she was able to use directly the software commands to represent the problem conditions. The operations that she did without the software, were of arithmetic nature: to find numbers in proportion to 10-90, 3-17. This type of reduction is not necessary because those proportions could be directly represented using the *RandomInteger* command instead the *RandomPick* one.

4. Conclusions

The operations *Selection, Mixture, Partition, Combination, Sampling*, defined by von Mises to be applied to Collectives, can now be constructed with the help of software tools like Fathom. The possibility of generating sequences of random trials (collectives) with the software and the operations that could be made with them could be seen as a toolbox useful to build models of random situations. Such models will have many possibilities to be good didactical models, in the sense of Fischbein.

Models built with that toolbox can easily possess the properties stressed by Fischbein for good didactical models. Indeed, the von Mises work, as well as the software structure make highly probably that the model be *heuristic, generative, consistent* and *well structured* (except for planning errors).

Particularly, we see that the initial complexity in problems of conditional probability (Total Probability and Bayes) is more manageable when the subjects begin to carry out the *partition* operation (associated with the "If…" command).

So, we have conceived the possibility to design a probability course with a focusing in simulation, with an organized structure linked to the von Mises proposal that, in turn, breaks with the classical organization of probability courses, which depend on the axiomatic structure of Kolmogorov.

REFERENCES

Biehler, R. (1991). Computers in Probability Education, en A. Hawkins, R. Kapadia (eds), *Chance encounters, Probability in Education,* Kluwer Academic Publishers, London.

Fischbein, E. (1977). Image and Concept in Learning Mathematics, *Educational Studies in Mathematics* 8, 153-165.

Finzer, W. Erickson, T., Binker, J.(2000). *Fathom*, Key Curriculum Press.

Hawkins, A. (1996). Myth–Conceptions!, en J. B. Garfield & G. Burril, Research on the role of technology in teaching and learning statistics,[Internet version] *International Statistical Institute, Voorburg,* The Netherlands, pp. 1-14.

Von Mises, R. (1957). *Probability, Statistics and Truth.* New York: Dover Publications. 2nd Ed. in English prepared and revised by Hilda Geiringer.

Sánchez, E. (2002). Teachers' beliefs about usefulness of simulation with the educational software fathom for developing probability concepts in statistics classroom, *Proceedings of the Sixth International Conference on Teaching Statistics*.

Yáñez, G. (2002). Propuesta Didáctica de la simulación computacional utilizando Fathom, *Hipótesis Alternativa*, 6.

RÉSUMÉ

Dans ce papier, nous présentons une expérience menée avec des élèves de DEUG, dans laquelle nous nous demandons si ces élèves seraient plus à l'aise pour inventer une simulation, plutôt que de répondre à des questions de probabilité. Nous pensons que les élèves pourraient prendre conscience des différents niveaux de simulation dans plusieurs problèmes en réalisant eux-mêmes des simulations. Nous décrivons les différents niveaux de compréhension que les élèves révèlent en résolvant des problèmes de probabilité.

Cette contribution fait partie d'un projet dont le but est d'explorer les possibilités de la simulation comme outil pour résoudre des problèmes de statistique et de probabilités. Nous avons adopté le point de vue de Hawkins (1966): "Introduire effectivement la technologie suppose de comprendre exactement les mêmes choses sur l'apprentissage et sur la réalisation du meilleur enseignement possible, que si nous devions préparer tout autre enseignement non basé sur la technologie" (p.6). Nous posons ces questions : quel doit être le contenu d'un cours basé sur la simulation ? Et : comment ce contenu doit-il être organisé (particulièrement à la fin de l'enseignement secondaire et au début de l'enseignement supérieur) pour être un instrument utile pour résoudre des problèmes ? Nous avançons d'abord une réponse à partir d'une analyse a priori; puis nous donnons un exemple qui souligne la pertinence de notre proposition. Dans cet exemple, deux élèves travaillent à résoudre des problèmes de probabilités conditionnelles.