STUDENTS' DIFFICULTIES AND STRATEGIES IN SOLVING CONDITIONAL PROBABILITY PROBLEMS WITH COMPUTATIONAL SIMULATION

Gabriel Yáñez
Cinvestav - IPN, México
UIS, Bucaramanga, Colombia

This paper reports some difficulties and strategies the students have in solving conditional probability problems with computer simulation. The difficulties are related to programming in the software used and with the trend of the relative frequencies graphic representation when the number of cases increases. The main strategy for estimation was taking the last value of the relative frequencies as the requested probability.

INTRODUCTION

The use of simulation in probability teaching is often suggested. Besides of being an universal problems solver (Batanero, 2001) the computer simulation has some interesting features for its implementation in the classroom. Some of them are:

- It allows overcome, as Biehler (1991, p. 170) says, the "lack of experience" that the students have about random experiments, providing representations of the real phenomenas and increasing the number of repetitions in short time.
- Solving problems by computer simulation increases the range of problems and realistic situations accessible to students' activities overcoming the "concept-tool gap" as Biehler says (p. 170), i.e. the gap between the intended generality of the probability concept and the system of operations and tools students actually use.
- Acting as a laboratory, checking the results obtained theoretically. In this sense, allows an exploration, changing the assumptions of the model, making further experiments, etc.

With these perspectives raise the question about the effects might have the computer simulation on the understanding of the probabilistic concepts. There are studies that document the use of the simulation as a method for modeling probability problems even though it produces good results, exist evidence that points out the random experiences are not enough neither for improving the probabilistic methods of the students nor for helping to develop the inductive methods. However, it's useful to highlight their misconceptions and give the students better explanations about the reality (Zaki & Pluvinage, 1991; Bordier, 1991; McClintock & Jiang, 1997; Bordier & Bergeron, 1998).

From other point of view, the use of computer in order to adopt the frequentist approach in probability raise questions about its use that are worthy to search. In this sense, Coutinho (2001) reports some difficulties the students had when several situations were given for building urns models for simulating some probabilistic games. These difficulties are associated with the use of the software, the resistance to use the simulation for solving a problem that can be solved directly by calculation and the difficulty on accept simulation data that has not obtained by themselves in order to estimate probabilities.

With these ideas in mind, we are developing a searching for knowing the understanding that the students generate as a result of an instruction supported on the use of the computer simulation. The first part of this work, besides of looking for students' intuitive ideas about probability, look for the spontaneous ideas they have about the frequentist approach and, in particular, the difficulties they face solving probability problems using the computer simulation. Furthermore, we pretended find the strategies they used for estimating the values of the requested probabilities. Some of the results obtained are the content of this paper (Yáñez, 2001).

DESCRIPTION OF THE STUDENTS AND THE ACTIVITY

Twelve university engineer students in Mexico City participated in this study. The activity was made in four sessions of three hours each. In the two first sessions after a short introduction to the frequentist approach we started the work with Fathom (Finzer, Erickson & Binker, 2000) solving some problems with computer simulation. In the third session the students worked in pairs for solving three problems of conditional probability; in the fourth session the students individually solved three problems of conditional probability. During the solving problems sessions the students worked alone, without help of the researcher and without discussions between them neither during nor after the sessions. The idea was to detect the spontaneous work and the conceptions that the students had respect the simulation and the frequentist probability and to know the strategies they used in solving problems.

RESULTS AND DISCUSSION

We present the solutions that were given for some of the students that participated in the study.Carlos for solving the urn problem (See the Appendix) simulated the drawings of the two balls with substitution. For answering the first question he made the graphic of the Figure 1. He wrote: *From this exercise I can say that the stability is between 0.2 and 0.3 and its frequency is 0.23 with 500 cases. This for the question 1, and for the question 2 with 500 cases its stability almost no change staying between 0.2 and 0.3 and a frequency of 0.274*

However Carlos answered in the paper in the following way:

(i) *Probability 2nd. Ball is white = 1/3 because one ball has been drawn.*

(ii) *1/3 counting all the possible balls and knowing that the second ball drawn was white.*

It is interesting that Carlos in the diagnosis test gave the following answer to the same problem:

(i) *P(2nd white ball) = 1/2 , i.e. one white and two blacks.*

(ii) *P(1st white) = 2/2. I say it because it is possible that the two black balls could have drawn, because they are the same number of balls.*

Carlos believes that the stability of the relative frequencies is a different concept that the value of the probability requested. Estimate the probability with the last value of the relative frequency.

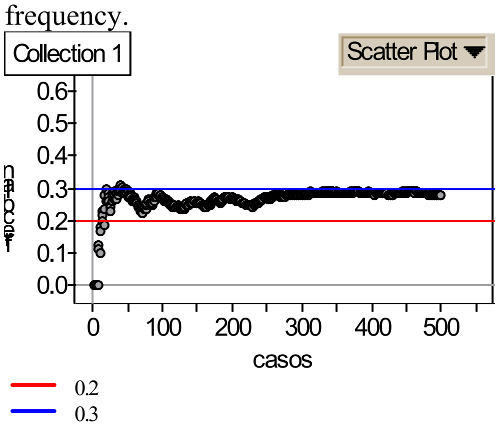

Figure 1. Carlos's Graphic for the First Question of the Urn Problem.

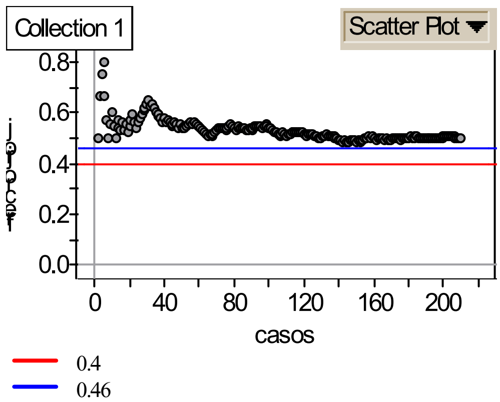

Figure 2. José's Graphic for the Chips Problem.

For solving the problem of the chips (see appendix) José made the program that allows calculate the probability of drawing the red side of the chip that has two different colors. He made the graphic of the Figure 2 and wrote: *I generated 210 cases and with the help of the graphic I can say that its stability is between 0.4 and 0.6 and a frequency of 0.46667.* In the paper he wrote: *1/2 because one chip is red on both sides and other one is red on one side and blue on the other.*

José as well as Carlos, used the strategy of the last value for estimating the probability. Although the two answers given are the same, he assumes them as different, one is the result of the simulation and the other one is the result of the theoretical analysis. His theoretical answer shows the mistake reported by Falk (1986) selecting the chip as the conditioning event.

Rafael for solving the taxi problem (Appendix) made a program for calculating the probability of the conjunction of picking a blue taxi and that the witness correctly identified the color of the taxi as independents events. He wrote the following comments: *For this experiment I conclude the following: 1. When you increase the number of cases you obtain a different probability, however when I generated 4500 cases the probability was between .19 and .22 that* *was very stable, different samples were always contained in this interval. I think that this kind of problems are not easy because it is very difficult for me transfer to the computer my theoretical analysis. For this reason sometimes I don't trust the results of the simulation.*

Rafael, besides of his intuition about the estimation by intervals, show us the content of his strategy that is to transfer his theoretical analysis to the computer simulation language, instead of modeling the random experiment. Luis and Valentín in the chips problem, estimated the probability of drawing the chip with different colors and to obtain the red side. They generated 170 cases and his estimated value was 0.51 using the mode of the results: *because in the majority of the cases that value was obtained.* They made the graphic of the Figure 3.

Marcos y Jesús working in a problem about dishonest dice that are thrown, for answering a question about of probability of obtaining the sum 7, did a good program, but only generated 110 cases, made the graphic of the Figure 4 that they did not well interpreted and chose the strategy of the last case for estimating the probability with the value 0.0818.

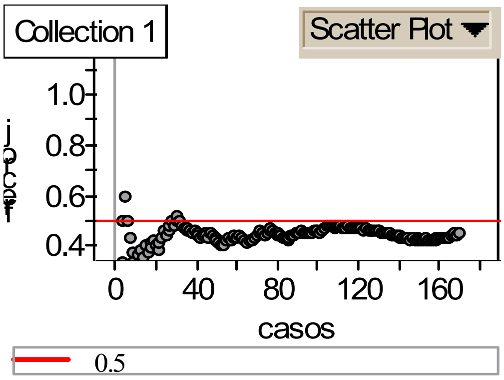

Figure 3. Graphic Made by Luis and Valentín in the Chips Problem.

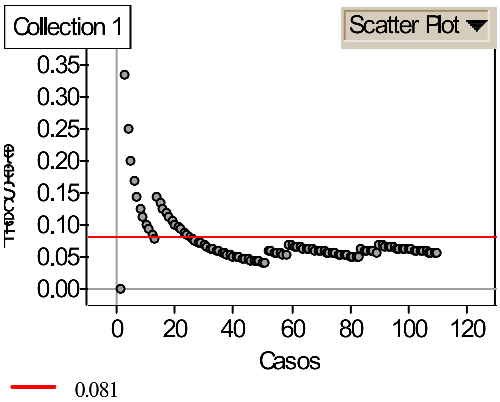

Figure 4. Graphic Made by Marcos and Jesús in the Dishonest Dice Problems.

CONCLUSIONS

For the results obtained, it is seen that the students have difficulties in the modeling of the random experiment and in its programming in the simulation computer language. Other difficulty deals with the interpretation of the graphics of the relative frequencies for estimating the probabilities. Almost all of the students used the *last value strategy* that identify the last value of the relative frequency with the probability requested. Another strategy deals with modeling for solving only the questions and not the random experiment, trying to transfer the theoretical analysis to the simulation computer language.

It might be that the students make distinction between a theoretical probability an simulated one, even that they believe there are as many probabilities as generated values of the relatives frequencies. Between the theoretical and the experimental value they chose the theoretical, because the lack of confidence in the simulation method and because the estimation by simulation based in the reading of the trend in a graphic in a infinity process. For some students, even, the probability requested is the mode of the values of the generated relatives frequencies. Other thing to mention is the small quantity of cases generated showing the believe in "the law of small numbers" and therefore a lack of understanding about the necessity for a large number to achieve the stability of the relative frequencies to be able to estimate the probability.

REFERENCES

Batanero, C. (2001). *Didáctica de la estadística.* Grupo de Educación Estadística Universidad de Granada. Departamento de Didáctica de la Matemática, Universidad de Granada, Granada, España.

Biehler, R. (1991). Computers in probability education. In R. Kapadia and M. Borovcnik, (Eds.), *Chance Encounters: Probability in Education*, Volume 6 (pp. 169-211). Kluwer Academic Publishers.

Bordier, J. (1991). Un modèle didactique, utilisant la simulation sur ordinateur, pour l'enseignement de la probabilité. *Doctoral Thesis*. L'Université Paris VII, France.

Bordier, J., & Bergeron, A. (1998). Simulation and formal manipulations: Complementary approaches for studying random events with computers. In *Proceedings of the 5th International Conference on Teaching of Statistics,* ICOTS 5, pp. 835-840. Singapore.

Coutinho, C. (2001). Introduction aus situations aléatoires dés le Collége: de la modélisation à la simulation d'experiences de Bernoulli dans l'environment informatique Cabri-géometre-II. *Doctoral Thesis*. University of de Grenoble.

Falk, R. (1986). Conditional Probabilities: Insights and Difficulties. In *Proceedings of the 2nd International Conference on the Teaching of Statistics,* ICOTS 2, pp. 992-997. Quebec, Canadá.

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

McClintock, E., & Jiang, Z. (1997). Spreadsheets: powerful tools for probability simulations. *Mathematics Teacher*, *90(7)*, 572-579.

Zaki, M., & Pluvinage, F. (1991). Demarches de resolution et de simulation face au probleme de la ruine d'un joueur. *Educational Studies in Mathematics, 22*, 149-181.

Yáñez, G. (2001). Estudios sobre el papel de la simulación computacional en la comprensión de la probabilidad condicional. *Unpublished doctoral project*. Cinvestav-IPN, México.

APPENDIX

Urn problem: An urn contains two white balls and two black balls. We shake the urn thoroughly, and blindly draw out two balls, one after the other, without replacement. (i) Suppose we know that the first drawn was white. What is the probability that the second ball is also white? (ii) Suppose we know that the second ball drawn is white. What is the probability that the first ball is also white?

Chip problem: Three chips are in a hat. One is blue on both sides, one is red on both sides, and one is blue on one side and red on the other. We draw one chip blindly and put it on the table as it comes out. It shows a red face up. What is the probability that the hidden side is also red?

Taxi problem: A cab was involved in a hit and run accident at night. Two cab companies, the Green and the Blue, operate in the city. You are given the following data:

(a) 75% of the cabs in the city are Green and 25% are Blue.

(b) A witness identified the cab as blue. The court tested the reliability of the witness under the same circunstances that existed on the night of the accident and concluded that the witness correctly identified each one of the two colors 80% of the time and failed 20% of the time. What is the probability that the cab involved in the accident was Blue rather than Green?