CONNECTING THEORETICAL PROBABILITY AND EXPERIMENTAL PROBABILITY IN A MODELING ENVIRONMENT

Santiago Inzunsa

Universidad Autónoma de Sinaloa (México)

We report results of a study with a group of social science students taking an introductory course in probability. The purpose was to explore the reasoning linked to the connections between the theoretical probability and experimental probability in an environment modeling. The results show that the modeling process was not a complex activity for most students when the probabilities of the problem are known (from theoretical approach toward the frequency approach), but in the opposite direction (from frequency approach toward the theoretical approach), the variability of the results was one of the main obstacles to identify the theoretical model underlying, particularly in those students who used 1,000 or fewer runs. The students who used 5,000 or more runs developed an intuitive understanding of the law of large numbers and adjusted the frequencies to the correct theoretical model that generated.

INTRODUCTION

Among the main causes that have been identified as limiting the development of probabilistic reasoning of students is the excessive emphasis that some teachers make in the classic approach (theoretical probability) and the use of routine process and combinatorial techniques to calculate probabilities; leaving aside other approaches with very important applications such as the frequency approach to probability (empirical or experimental probability). Some researchers propose that the teaching of probability emphasizes the conceptual connections between the theoretical and experimental concepts using different approaches to probability (e.g. Fischbein & Gazit, 1984) using technology to build probabilistic models and explore random situations (e.g. Konold & Kazak, 2008). The probabilistic models that are assumed in both approaches and the results generated can be compared through a bidirectional process from classical approach to frequency approach and vice versa (Prodromou, 2012). In this process it comes into play important concepts (e.g. randomness, variability, sample size, sample space, law of large numbers) that can help students develop their probabilistic reasoning and understand the role of models in the prediction of the uncertainty and randomness. The modeling perspective achieves a synthesis between these two approaches (classical and frequentist) and recent didactic studies have led to teaching based on this modeling process using simulations of models in statistics (Chaput, Girard & Henry, 2011; Batanero, Henry & Parzysz, 2005). The literature shows few studies on the perspective of modeling at the university level and the conceptual connection between theoretical and experimental probability. Specifically we propose the following research questions:

- Can students establish a correct conceptual connection between the theoretical probability and experimental probability?
- What relationships the students establish between models, probabilities and frequencies?

CONCEPTUAL FRAMEWORK AND METHODOLOGY

In the theoretical approach, the probability is obtained by the fraction of outcomes favorable to this event in the simple space; this makes use of an implicit assumption of equal likelihood of all single outcomes of the simple space. By contrast, a frequentist approach to probability defines the probability of an event as the probability estimated from relative frequencies used to determine the likelihood of an event from the outcomes of a random experiment that is repeated a sufficiently large number of times under the same conditions. In the context of the probability, Henry (1997, p. 78) states, "a model is an abstract, simplified and idealized of a real object, a system of relations or an evolutionary process within a description of reality representation". The process of building a model and their simulation consists of different stages: a) develop a pseudo-concrete model and description of reality (putting empirical observations into a working model), b) mathematization and formalization (translating working to hypotheses to design probability model) and c) validation and interpretation in the context (checking fit of a probability model to data).

The study took place with a group of fifteen students in an introductory course in probability. The orientation of the course was designed to complementary develop the classical and frequency approaches. The background in probability of students was very basic and they only had superficial ideas about the classical approach to probability. The results presented were obtained from an activity in the first part of a larger study, the activity was developed in two sessions of 1.5 hours each one. The first part of the activity was intended to explore the connection frequency approach toward the theoretical approach and estimate the probabilities of a hidden model that generated the results, after the students proposed a model that would generate similar results. The second part was intended to explore the connection from theoretical approach toward the frequency approach, by building three equivalent models from known probabilities, simulating and validating its equivalent model based on the observed frequencies, and finally calculating theoretically the probabilities and comparing with the corresponding frequencies. Additional interviews were realized with some students for more detail about their reasoning.

RESULTS AND DISCUSSION

Connections from frequency approach (data) toward the theoretical approach (model).

To explore these connections the activity included three frequency distributions obtained when passing through revision lines A, B and C; 100, 500 and 1,000 passengers. It requested that based on this information, the students estimate the probability that the system is assigning to each one of the lines review. After the students opened a file containing the hidden model that generated the results, the students made several runs and then build an equivalent model. All students identified that line A is much more likely than the lines B and C, and both have the same probability; however, only four students did not have problems with the variability and identified the convergence to the correct values 50%-25%-25% when increased the number of runs. Anaid was a student whose answer showed adequate understanding of the variability when she observed the three graphs.

> "Of the three graphs is observed that as more passengers who pass, the frequencies are more accurate, and appear 50% -25% -25%, because 51% -25% -24% which is observed in the last graph has a range of variability because is random"

In the exploring of the model most students used 1000 runs; only three students used 5,000 and 10,000 and they made a very accurate estimate. Only four models were correct and the others had inconsistencies. Some incorrect models assigned equal probabilities to the three lines and other assigned probabilities to the results obtained but with some variation.

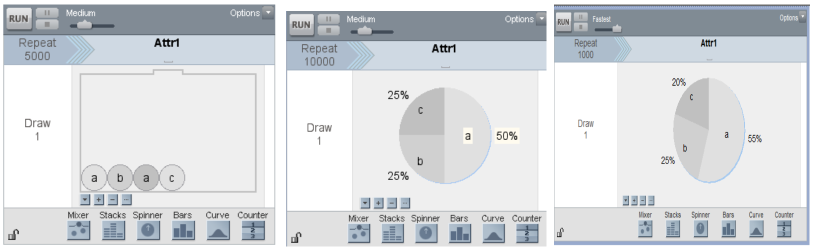

Figure 1: a) Correct model b) Correct model c) Incorrect model

The students who generated a model similar to model c (incorrect), have not reached to overcome the barrier of variability, even when the software allows them to make a lot of runs repeated times to identify the pattern of frequencies; such was the case of Katya who made three runs of 1000 cases and proposed a model approximate to the correct model.

Interviewer: How did you get the model 54% -25% -21%?

Katya: I made three runs of 1000 and I always got C lower frequencies, so I put a 4% below 25%. My model considers a small variation in each outcome because there is always 50% -25% -25%.

Naxely executed the model with 10000 runs and answer the following questions:

Interviewer: What was your basis to propose the model 50% -25% -25%?

Naxely: In the last graph it has 50% -25% -25% frequencies.

Interviewer: What would happen if you run 10,000 simulations repeatedly?

Naxely: It will maintain the same close numbers. For example line A is near to 50% and other lines 25% and 25%.

Connections from theoretical approach (model) toward the frequency approach (data).

To analyze these connections we used an extension of the activity. The probabilities of the revision lines (A, B, C) now are known, considering that passengers had to pass a traffic light duty with an 80% chance for the green and 20% for red. Unlike the previous session it was required to adjust the data to a model that generated it, starting from probabilities known of the events was easier because 12 students built the model correctly.

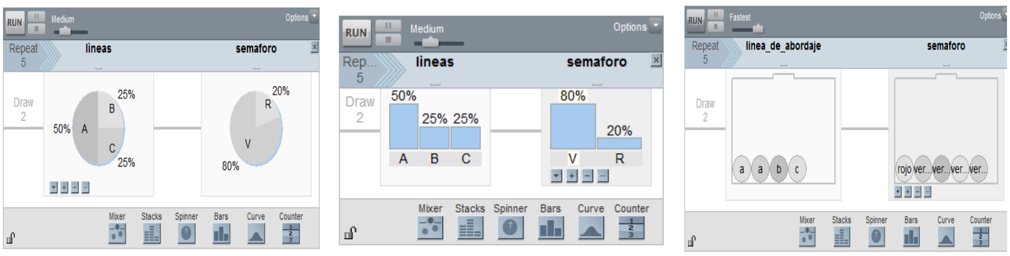

Figure 2: Examples of correct models

Interviewer: Do you think that the three models produce the same results?

Fernanda: Yes, because I wrote the same percentages in each section of the spinners and each bar, but the in mixers it did not let me write the percentages and I doubted whether

	the model was right, but when I ran the models it gave the same results and I understood that the three models were equivalent.
Interviewer:	Were they exactly the same results?
Fernanda:	There were differences of more or less 1%, with very little variability.

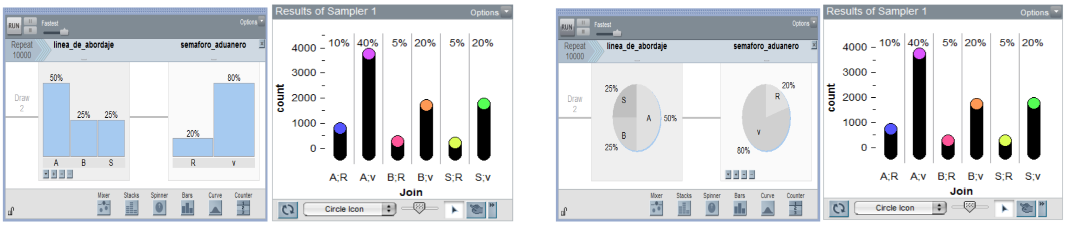

Fig. 3: Two models built by Fernanda

CONCLUSIONS

In the connection from frequency approach toward the theoretical approach, the variability of the results was one of the main obstacles to identify the theoretical model in those students who used 1,000 or fewer runs. In this sense, putting empirical observations into a working model was a difficult task for these students. Students who used more than 5000 runs noted that the variability of the frequencies decreased as the number of runs increased and frequencies converge to the theoretical probabilities, thereby developing a correct intuitive idea of the law of large numbers. These students were successful in identifying the correct model that generated frequencies. In the connection from theoretical approach toward the frequency approach, the students modeled a compound event successful. It seems that it was easier to start from probabilities known of the events to build a model that identify from the observed frequencies. The construction of three models of same problem served as a validation mechanism for many students, who noted the similarity in the behavior pattern of frequencies distributions. Although the results come from a single activity, which was realized in the first topics of the probability course, is observed that the modeling process was not a complex activity for most students, because the representational software features that becomes concrete, abstract aspects of modeling.

References

Batanero, C., M. Henry & B. Parzysz (2005), The nature of chance and probability. In G. Jones (Ed.), *Exploring probability in school: challenges for teaching and learning,* Springer Science+Business Media, 15-37.

Chaput, B., Girard, J. C. & Henry, M. (2011). Frequentist approach: Modelling and simulation in statistics and probability teaching. In C. Batanero, G. Burrill & C. Reading (Eds.), *Teaching Statistics in School Mathematics-Challenges for Teaching and Teacher Education: A Joint ICMI/IASE Study, Heidelberg:* Springer

Fischbein, E. & Gazit, A. (1984). Does the teaching of probability improve probabilistic intuitions?. *Educational Studies in Mathematics, 15*, 1-24.

Henry, M. (1997): Notion de modéle et modélization en l'enseignement. *En Enseigner les probabilités au lycée,* Commission Inter-IREM, Reims, 77-84

Konold, C. & Kazak, S. (2008). Reconnecting data and chance. *Technology Innovations in Statistics Education, 2*(1). Retrieved from http://escholarship.org/uc/item/38p7c94v

Prodromou, T. (2012). Connecting experimental probability and theoretical probability. *ZDM: The International Journal on Mathematics Education, 44*(7), 855-868.