STRATEGIES AND DIFFICULTIES THAT UNIVERSITY STUDENTS DEVELOP THROUGH THE MODELING OF RANDOM PHENOMENA BY SIMULATION

Santiago Inzunsa
Universidad Autónoma de Sinaloa
sinzunza@uas.uasnet.mx

Mario Romero
Universidad Autónoma de Sinaloa
mario.romero@info.uas.edu.mx

This paper reports the results of a research about the strategies and difficulties developed by university students in the process of modeling and simulating of random phenomena in an environment of a spreadsheet. The results indicate that students had difficulties to identify key components of the problems, which are crucial to formulate a simulation model. We have identified three different schemes to generate the results of the key components, which only one of them is correct; this scheme is based in the generation of random numbers. In consequence during this investigation it was observed that the process of the instrumentation of the spreadsheet to simulate random phenomena it is complex.

Keywords: Spreadsheet; Modeling; Probability

Purposes of the Study

The computer simulation has been suggested for many researchers and organizations as a pedagogical tool for the probability and statistics teaching, but above all, the statistics software, spreadsheets and computer technology have been penetrating even more inside laboratories and schools (Biehler, 1991; NCTM, 2000). In this way, the modeling of random phenomena by simulation has been constituted as an important part of research in statistics education during the last years. As a consequence of this, some studies were undertaken in order to know their effectiveness in probability and statistics teaching and learning (Maxara & Biehler, 2006; Lee & Mojica, 2008; Chaput, Girard & Henry, 2011).

The computer simulation integrates different aspects that are important in mathematics teaching, and particularly, in probability teaching:

1. It requires an activity of mathematical modeling in which students develop some skills, such as making assumptions to simplify the problem, identify and symbolize variables and parameters, as well as formulate the model by taking into account the assumptions and conditions, to finally, solve them and interpret the results.
2. When it is possible an analytic solution of the problem, the experimental results that are generated by the simulation can be contrasted with theoretical results. In some cases, in which the analytic solution is not possible or complex, the simulation is an important and fundamental tool.
3. It allows to work with abstract issues in concrete words, and above all, when the simulation is performed in computer environments that are equipped with representations (graphics, symbols, numbers) bound together, which make possible a visualization and feedback of the different parts of the model.

Among the advantages of using simulation as a method to solve problems of probability and statistics, Biehler (1991) mentions:

1. The possibility of formulating models in concrete terms instead of expressing the ideas by means of symbolic models (representational aspect).
2. Students can process the data generated more easily than data generated using analytical and combinatory methods (computational aspect).
3. It is possible to begin with the design of the experimental environment instead of starting with the calculations (concept-model aspect).

There are different proposals to formulate a simulation model in a computer environment. Some examples are Gnanadesikan, Scheaffer, and Swift (1987) that propose a complete and detailed a process:

Van Zoest, L. R., Lo, J.-J., & Kratky, J. L. (Eds.). (2012). *Proceedings of the 34th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education.* Kalamazoo, MI: Western Michigan University.

1. State the problem clearly.
2. Define the key components.
3. State the underlying assumptions.
4. Select a model to generate the outcomes for a key component.
5. Define and conduct a trial.
6. Record of observation of interest.
7. Repeat steps 5 and 6 a large number of times.
8. Summarize the information and draw conclusions.

On the other side, Albright (2010) proposes a three step process:

1. Construct a model that uses random numbers.
2. Evaluate the model many times using different random numbers each time.
3. Analyze the results statistically.

The implementation of the simulation in probability and statistics teaching could be given in different ways. NCTM (2000) recommends that the probability's problems could be first investigated through simulations in order to get an estimated result, and after that, to use a theoretical model in order to find the exact result.

One of the most important aspects in the implementation of simulation as a pedagogical tool is the computer tool utilized, because its design determinate some potentialities and constraints to the mathematical activity that students develop through the interaction with it. A special kind of tool in which it's easy to simulate different random phenomena is the spreadsheet. However, its use in mathematics education remains still relegated, despite its potential to handled quantitative information (Haspekian, 2005).

Although, the advantages that simulation offers, it is necessary to do a deep analysis about its didactic potentially. Our interest in this work has been the research of the potential that a spreadsheet can have in order to modeling random phenomena in a basic probability university course, and the difficulties that the students present in the different stages of the process of modeling. Particularly, we ask the following questions: (a) what are the potentialities and constraints that the spreadsheets have to the simulation of random phenomena? (b) what are the strategies that students can develop? and (c) what are the difficulties that students can find in the process of modeling?

Theoretical Framework

In the development of this work, we have adopted an instrumental approach of the mathematical cognition (Artigue, 2002). The most basic notions of this approach consist in the meanings of *instrument* and *artifact*. The artifact is a material or abstract object that is available for certain activities. Examples of this are the language, the calculator and the spreadsheet. On the other hand, the instrument is a personal construct that can be developed by handling an artifact in a progressive way.

An artifact becomes a instrument when the subject achieves to appropriate of the artifact and establishes meaningful relationships for doing a specific kind of work (a mathematical one, in this case), this means that the subject can use and control this artifact to achieve their goals, and to integrate it to their activities (Verillón & Rabardel, 2005). The process of the transformation of the artifact into an instrument is called *instrumental genesis*.

The process of instrumental genesis evolves in two interrelated directions. The first one is directed to the artifact, loading it progressively with potentialities, and eventually transforming it for specific uses; this is called the *instrumentalization* of the artifact. Secondly, instrumental genesis is directed towards the subject, leading to the development or appropriation of schemes of instrumented action, which progressively take shape as techniques that permit an effective response to given tasks. The latter direction is properly called *instrumentation*. In order to understand and promote instrumental genesis for learners, it is necessary to identify the constraints induced by the instrument (Artigue, 2002). The restrictions are the result of the tool's design. In this way, the use of a tool is not a unidirectional process, but a dialectic process between the subject that acts over the instrument, and the instrument, which acts over the thinking of the subject.

In the process of instrumental genesis, the user develops mental schemes for specific tasks. In these schemes, the technical knowledge or in other words, the skills to use the artifact and the knowledge of the specific domain of the mathematical content, are intertwined or complimented (Drijvers & Trouche, 2008). An instrument is a mixed entity constituted by one piece of artifact and other part with the personal schemes that the users develop through doing specific tasks. In the case of a mathematical task, a mental scheme involves the global strategy of solution, as well as the technical resources that the artifact offers, and the mathematical concepts in which underpin the strategy.

Potentialities and Constraints of the Spreadsheet to Random Phenomena Simulation

Excel spreadsheet possesses diverse potentialities and a framework of representational aspects of calculation and communication. In the communication aspect, a spreadsheet requires that the students work with an interactive algebra-like language, which focuses their attention on a rigorous syntax. This is why it is said that spreadsheets help to translate a problem by means and algebraic code (Haspekian, 2005). Into the representational aspect, the spreadsheet has multiple representations that allows several semiotic registers that can be presented in simultaneous ways on screen, such as the case of the formulas register to express relations between cells, numerical register to represent data or results of calculations, graphics register that allows user several types of graphical representations dynamically linked to the numerical data. And finally, for the aspect of calculation, Excel has an extended range of formulas that make possible formulating models, generating data and making calculations. Other important element is the numerical feedback obtained when working with a formula, which allows students to experiment, speculate, and help them to find mistakes. On the specific case of simulation of random phenomena, Excel spreadsheet has several commands to generate pseudorandom numbers. Under the case of discrete random phenomena which are simulated through models of urns (as it is in our case), Excel has two commands that generates numbers provided from a uniform distribution: *rand() and randbetween(bottom,top)*. The function *rand()* returns a random number between 0 and 1, meanwhile the function *randbetween(bottom,top)* returns an entire random number between the limits specified. The figure 1 and 2, show those formulas and the resulted generates are shown:

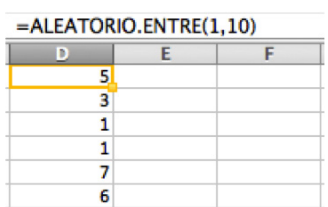

Figure 1: randbetween(1,10)

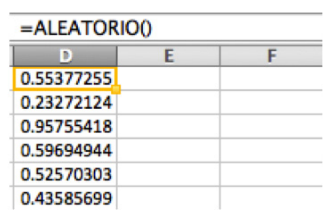

Figure 2: rand()

The pseudorandom numbers generated by *rand()* or *randbetween(inf,sup)* have two properties that make them comparable to in fact random numbers:

1. Any number between 0 and 1 has the same probability to be generated.
2. The numbers generated are independent between each other.

Excel spreadsheet, also as we know, works with arithmetic formulas, conditionals and statistics that make possible the conditions of a problem and realize the analysis of the results. The copy/paste option and F9 permits to evaluate the model as many times doable and therefore a quick feedback.

Van Zoest, L. R., Lo, J.-J., & Kratky, J. L. (Eds.). (2012). *Proceedings of the 34th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education.* Kalamazoo, MI: Western Michigan University.

Methodology

This study was done with 22 students between 19–20 years old of the undergraduate program of Information Systems. The students were enrolled in a basic course of probability and statistics in the first semester of the school year 2011–2012. This course emphasized the simulation as complement of the probability theoretical approach in some themes. Students weren't trained in a previous course about use of Excel to simulation and the study started with the their basic knowledge of Excel spreadsheet. Simple examples of simulation were raised in the class and some assignments to be complement it took place, therefore contrast results of the simulation with the theory achievement. The present work shows the results from the following three activities:

Activity 1: Simulate the rolling of two dices:

a) *Add of points resulting of faces up and determine the probability that the result is 7.*
b) *Subtract points resulting of faces up and determine the probability that the difference of points is 3.*
c) *Multiplicate the points of the faces up and calculate the probability that the product of points is bigger than 10.*

The purpose of this activity was to introduce students to simulation environment in the Excel´s spreadsheet, to observe with what formulas they will be working for and what difficulties they will be confronted with to finally formulate the model. According with the instrumental theory, this activity represents the starting point for the analysis of the instrumentation developed by the students in the spreadsheet.

Activity 2: If a friend of yours thinks in a number between 1 and 100 ¿what is the probability of making it divisible by 6 or 10?

This activity require the application of the rule of the addition of the probabilities; this is it, $\mathrm{P(AUB)} = \mathrm{P(A)} + \mathrm{P(B)} - \mathrm{P(A \cap B)}$, where A is an event that represents divisible numbers by 6, B is an event that represents divisible numbers by 10 and $\mathrm{P(A \cap B)}$ is event that represents divisible numbers by 6 and by 10.

Activity 3: There are 3 urns with black and white balls (Urn 1: 3 white and 2 black. Urn 2: 1 white and 3 black. Urn 3: 6 white and 2 black). There is selected one ball from one urn. What is the probability that a ball color white shows up? There is an assumption of equal probabilities to select any of the urns.

The purpose of this activity was to formulate a model of simulation to the calculus of the total probability of a event; this is it, $P(B) = P(U1)P(B/U1) + P(U2)P(B/U2) + P(U3)/P(BU3)$, where U1, U2 and U3 are events Urn 1, Urn 2, Urn 3 respectively, and B is the event that represents to which white ball is selected. Formulate the model requires identify 2 stages; the first one consist in decide the urn in a random way, and second stage; select the ball from that urn.

In the analysis of the information we have present the strategies and the elements of the spreadsheet that were used by students in each phase of the process of the modeling described by Gnanadesikan, Scheaffer, and Swift (1987) and Albright (2010).

Results

Formulation of the Models through Random Nnumbers

The formulation of a simulation's model of random phenomena contemplates the comprehension of the problem, identify the key components, make assumptions and build a symbolic expression through one or more commands to obtaining results.

In the context of the activity 1, the key components are the results of the dices. The assumption is that the dices are symmetrical, because each face of them is equally possible. One result is obtained adding, subtracting or multiplying the points of two dices -referring that case-. In this activity the students didn't show difficulties in identify the key components, which are essential part to make the simulation correctly. Figures 3 and 4 show the model made it by two students on the case by adding points:

E3 =SI(D3=7,1,0)+E2

LANZAMIENTOS	DADO 1	DADO 2	SUMA	FRECUENCIA
1	1	5	6	0
2	2	5	7	1
3	3	4	7	2
4	6	5	11	2
5	4	6	10	2

Figure 3: Model constructed by Luis

=SI(E3=7, 1, 0)

Lanzamientos	Dado 1	Dado 2	Suma	Frecuencia
1	5	4	9	0
2	4	2	6	0
3	3	2	5	0
4	2	6	8	0
5	4	3	7	1
6	6	1	7	1
7	3	5	8	0
8	3	5	8	0
9	2	3	5	0
10	5	2	7	1

Figure 4: Model constructed by Silvia

In the case of the activity 2, the key component is the number that is thought. There is the assumption of any number between 1 and 100 are equally probable to happen, therefore the result from the key component is obtained generating random numbers between 1 and 100. In this case, students had difficulties to identify the key component and most of them started with favorable results (divisible by 6 or by 10 or both), when the correct one was to generate firstly the key component and lately to identify favorable results.

In other hand, in the activity 3, the key components are the selection of the urn (urn1, urn2 or urn3) and the selection of one of the balls (black or white). The assumptions are that each urn has the same probability to be selected and for the balls option is that each ball has the same assumption of equal probability. As the same way that it was in the other activities, here was observed that students do showed difficulties to identify the key components, however some of them identified the balls selection (second part) but omitted the urn selection (first part).

In conclusion, we have identified three schemes for the formulation of the model:

1. Students, who use the space of the spreadsheet like a notebook, make calculus in manual manner and write it down the results. They don't use the potential of the spreadsheet.
2. Students, who use formulas to calculate probabilities, but do not request to generate random number how it is supposed to be in a model of simulation.
3. Students, who use formulas with random numbers to simulate the key components.

In the following figures of one the students who participated in the activity 2 which results generated from a random command (Figure 5) and another one where no random formula results to accomplish the divisibility condition (Figure 6).

=F2/1000

Numero	Divisible 6	Total Divisible	Divisible 10	Total Divisible	Interseccion	Probabilidad
5	0	175	0	117	33	
82	0	0.175	0	0.117	0.033	0.259
52	0		0			
50	0		1			
1	0		0			

Figure 5: Model constructed by José

F12 =F3+F6-F9

	A	B	C	D	E	F
1						
2	Espacio Muestral	Divisible 6	Divisble 10	AUB		Probabilidad de que sea divisible entre 6
3	1	6	10	30		0.16
4	2	12	20	60		
5	3	18	30	90		Probabilidad de que sea divisible entre 10
6	4	24	40	Total		0.1
7	5	30	50	3		
8	6	36	60			Probabilidad de que sea divisible entre 6 & 10
9	7	42	70			0.03
10	8	48	80			
11	9	54	90			Probablidad de que sea divisible entre 6 o 10
12	10	60	100			0.23
13	11	66	Total			
14	12	72	10			
15	13	78				
16	14	84				
17	15	90				
18	16	96				
19	17	Total				
20	18	16				

Figure 6: Model constructed by Ana

For his part in Figure 7, the student Angel shows his work in activity 3. Where, first at all, he made calculations in manual method using formula of total probability, then simulates extraction of balls and calculates frequencies to 10,000 cases. Results of both approaches match up given certainty to precision of his problem.

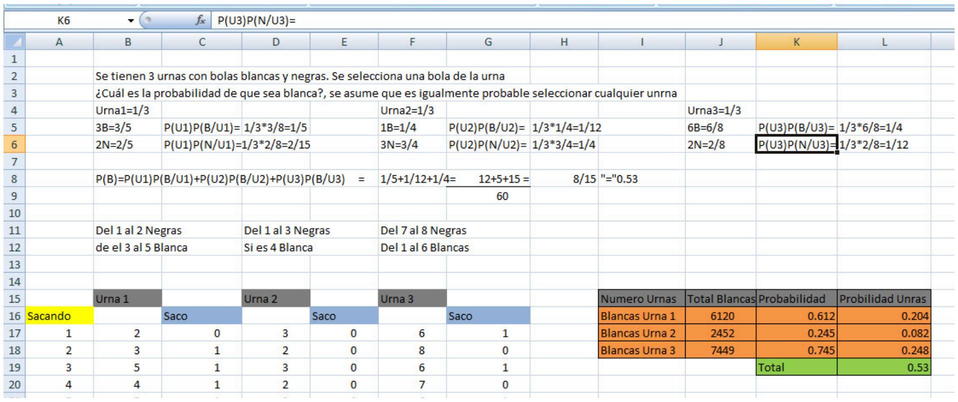

Figure 7: Work developed by Angel in activity 3

The Table 1, show the results of the different schemes developed by the students and the commands utilized to formulate the model of simulation.

Table 1: Classification of Students by Schemas and Commands Utilized

Activity	*Frequency of students by schema*			*Commands used to generate the model*
	1	*2*	*3*	
1	8	0	14	Randbetween(1,6) A1+A2, Sum(A1,A2), A1-A2, A1*A2
2	7	9	6	Randbetween (1,100)
3	3	14	5	To the selection of the urn: Randbetween(1,3) To the selection of the balls: Randbetween (1,5), Randbetween (1,4), Randbetween (1,8).

Evaluation of the Model

The evaluation of the model consist in generate a case and record the outcome of interest, and then repeat it many times and see if the results satisfy the conditions of the problem. In the first activity, the analysis reveals that the students who generated the results of the key components though a random function (scheme 3), use the function "copy of formulas" to generate many cases (hundreds or thousands) as it is shown in figures 1 and 2. Other students used F9 in order to repeat the case in the same cell, and afterwards, they copied again. The students who used scheme 1 (data direct introduction) had the difficulty to generate many cases. It is also important to highlight that some students found out some mistakes in their model when the result obtained was not what they were waiting for. This important function of the spreadsheet allows students to monitor a problem solution process.

In activity 2 the students who did not use a random formula to generate the results of the key components (thought number), could not evaluate their model as a case of simulation, however, they got correct results through probability's classic formula. In the case of the students, who generated randomly the results (e.g., José, Figure 5), identified the favorable cases through the formula *residue(number,number_divisor),* combined with the conditional formula *if(logical test,value_if_true_value_if_false)*. Afterwards, when students observed that some correct results were produced, they copied the formula to solve many cases (see Figure 5).

On the other side, in activity 3, only 5 students used random numbers in order to build a model and as mentioned before, the simulation was partial (the extraction of the ball in the urn selected), and this is the reason why the evaluation of the model consisted in identifying if the random generated number corresponded to the white or black ball, for which they used the conditional formula *if(logical test, value_if_true_value_if_false)*.

Table 2: Commands Utilized to Obtain Conditions and to Evaluate the Model

Activity	*Commands used to identify favorable results*
1	To additions=7: IF(D6=7,1,0 To substractions=3: IF(G6=3,1,0) To products > 10: IF(J6>10,1,0)
2	To divisors of 6: IF(RESIDUE(B1,6)=0,1,0) To divisors of 10: IF(RESIDUE(B1,10)=0,1,0) To divisors 6 or 10: =COUNTIF(C2:C1000,1,E2:E1000,1),

3	IF(B17=1,0,IF(B17=2,0,1))
	IF(D17=4,1,0)
	IF(F17=7,0,IF(F17=8,0,1))

Statistic Analysis of the Results

In this phase, we focus on summarize the information obtained by the model and draw conclusions. In specific, it means register observations of interest (favorable events), and calculation of the relative frequencies.

In activity 1, most of the students who generated the model through of a random command followed a schema of identify and accumulate the favorable cases in the same column, what means introduce a recursive formula starting at the second line, how is showed in case Luis (Figure 1). This part was cause of difficulties for some students, who only can identify the favorable cases but no one accumulated the results as is showed in Silvia's work. A schema more simple was identified the favorable cases and make additions in determined cells, nevertheless, any student didn't show that kind of schema.

In the activity 2, the register of the favorable results and calculation of the frequencies did it by the students using a random phenomena through formulas like *countif(range,criteria)* and *countif(range,criteria1,range,criteria2,)*. In other hand, activity 3, using 0 and 1 as variable indicator to identify favorable results, it helped to calculate the frequency through the formula *countif(range,criteria)* or *sum(number1,number2)*

Table 3: Commands Used for the Register Favorable Results and Calculation of Frequencies

Activity	*Commands used to count favorable results*
1	=IF(D5=7,1,0)+E4
2	=COUNTIF(E2:E1000,1)
3	=COUNTIF(C17:C10016,1)

Conclusions

The results obtained in this research have showed that the process of the instrumentation of the spreadsheet in the modeling of random phenomena it is not an easy process, even for the students with a previous programming knowledge, as it was the case of the subjects involved in our study. The process of simulation works with a methodology that demands from students the use of intertwined strategies between the knowledge they already have of the topic and the technical knowledge of the resources that the spreadsheet has.

As it was observed the empty knowledge of the functions of the spreadsheet (potentialities) and constraints, from where depends the level of instrumentation, by taking in consideration spreadsheet wasn´t conceived as an educational tool. In other hand, the theoretical approach that the students are used in probability courses since secondary school is part of the influence that relegate the use of the Excel potentialities to the typical calculus functions, because for them results more simple and easy use it in that way.

The incorrect schemas developed for the formulation of the model show the difficulties that the students had in identify the key components of the problems, therefore they opted to use other strategies as to introduce in a direct way the data or generating through no random formulas to accomplish the problem. According with this our conclusion states that the competences to develop models of simulation in a spreadsheet, requires a planned process to show to the students the methodology of simulation, as well it is needed a better knowledge of the potentialities and constraints of the tool.

References

Albright, B. (2010). *Mathematical modeling with Excel*. Sudbury, MA: Jones and Bartlett.

Artigue, M. (2002). Learning mathematics in a CAS environment: The genesis of a reflection about instrumentation and the dialectics between technical and conceptual work. *International Journal of Computers for Mathematical Learning, 7*(3), 245–274.

Biehler, R. (1991). Computers in probability education. In R. Kapadia & M. Borovcnik (Eds.), *Chance encounters: Probability in education. A review of research and pedagogical perspectives* (pp. 169–212). Dordrecht, The Netherlands: Kluwer Academic.

Chaput, B., Girard, J., & Henry, M. (2011). Frequentist approach: Modeling 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*. The 18th ICMI Study (pp. 85–95). Springer Science+Business Media.

Drijvers, P., & Trouche, L. (2008), From artifacts to instruments: A theoretical framework behind the orchestra metaphor. In K. Heid & G. Blume (Eds.), *Research on technology and the teaching and learning of mathematics, Vol. 2. Cases and perspectives* (pp. 363–392). Charlotte, NC: Information Age.

Gnanadesikan, Scheaffer, & Swift. (1987). *The art and techniques of simulation*. Quantitative Literacy Series. Dale Seymour.

Haspekian, M. (2005). An "instrumental approach" to study the integration of a computer tool into mathematics teaching: The case of spreadsheets. *International Journal of Computers for Mathematical Learning, 10*, 109–141.

Lee, & Mojica. (2008). Examining how teachers' practices support statistical investigations. In C. Batanero, G. Burrill, C. Reading, & A. Rossman (Eds.), *Proceedings of the ICMI Study 18 and 2008 IASE Round Table Conference*. Monterrey, México. Retrieved from http://www.ugr.es/~icmi/iase_study/

Maxara, C., & Biehler, R. (2006). Students' probabilistic simulation and modeling competence after a computer-intensive elementary course in statistics and probability. *Proceedings of the Seventh International Conference on Teaching Statistics,* Salvador Bahía. Voorburg, The Netherlands: International Statistical Institute. Retrieved from www.stat.auckland.ac.nz/~iase/publications.php

NCTM. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.

Verillon, P., & Rabardel, P. (1995). Cognition and artifacts: A contribution to the study of thought in relation to instrumented activity. *European Journal of Psychology of Education, 10*(1), 77–101.