ISSN No:-2456-2165

Exploring Students' Probabilistic Difficulties using Numerical Simulations of Random Experiment

Hamid EL AMRANI
Faculty of Sciences Dhar El Mahraz – Sidi Mohammed Ben Abdellah University,
BP: 1796, 30000, Fez, Morocco

Moncef ZAKI
Faculty of Sciences Dhar El Mahraz – Sidi Mohammed Ben Abdellah University,
BP: 1796, 30000, Fez, Morocco

Youssef NAFIDI
Faculty of Sciences Dhar El Mahraz – Sidi Mohammed Ben Abdellah University,
BP: 1796, 30000, Fez, Morocco

Interdisciplinary Laboratory of Research in Didactics of Sciences and Technology (ILRDST)

Abstract:- In this paper, we will see the interest of the simulation of random experiments in a study of mathematics education, especially in probability modeling. Through simulations accompanied by a questionnaire, we were able to explore various conceptual difficulties among students concerning the teaching of probability theory at the university. Indeed, on the one hand, these difficulties revolve around the basic probabilistic notions, contributing to the modeling through the random variable notion. For example, difficulties to adopt the frequency approach to estimate the fortune of the gambling player, difficulties in identifying a random experiment, difficulties in constructing the random variable appropriate to a probabilistic situation, etc. On the other hand, this simulation innovation also allowed students to explore a cognitive model for two probabilities to have a “heads” p and $1-p$ of intuitive estimation of “the fortune of the player”. Moreover, many students arrived to put the frequency approach of probability on availability thanks to the observation of the repetitions whose simulation allowed them. The challenge of this work is to integrate numerical simulations into teaching in order to explore the learning difficulties of students and to evaluate their learning. In this way, students will be able to construct their lively concepts, which will be available to engage in probabilistic modeling.

***Keywords**:- Simulation, conceptual difficulties, random experiments, random variable, probabilistic modeling.*

I. INTRODUCTION

Digital treatments have become indispensable in various fields of our daily life [1]. These treatments solve very complicated problems through the study of massive data “big data” [2], as well as through qualitative processing of statistics that focuses on reduced samples. In our didactic study, we integrated a digital support that relies on the generation of pseudo-randomness in order to build a random experiment that we will explore with the conceptual difficulties that hinder probabilistic modeling of students.

In general, the teaching of probabilities at the university level is accompanied by enormous conceptual difficulties of the students, because the notion of probability is often introduced from its axiomatic definition. In fact, this definition marks a major break with the equiprobability model, a model previously used by students through high school education which is essentially limited to typical probabilistic situations using finite possibilities universes. The difficulties become more widespread when they are confronted with problem situations that require in their resolution the development of a probabilistic model using random variables. It is this type of situation that we have chosen for our experimentation, with the integration of the simulation of random experiments.

Our choice of conducting random experiments in the form of computer simulation has been guided by the following reasons.

- The possibility of repeating a very large number of times with the same random experiment, under the same conditions, in a fairly limited period of time. Such a possibility of repetition of a (pseudo) random experiment is in favor of a frequency approach to probabilities, necessary for a probabilistic apprehension, but also for the application of the law of large numbers in an experimental setting [3] [4].

- The development of probabilistic intuition for probabilistic modeling, through the identification of random variables that are relevant to the probabilistic situation being treated [5] [6] [7] [8].

- The possibility of changing the parameters of the probabilistic situation treated in the simulation, in this case the advantage of the player in each game, the number of game parts, and to assess the impact on the question being dealt with, namely the fortune of the player. It is, in fact, flexibility in the choice of the conditions of the random experiment [9].

The problematic of the present research is to explore, through the computerized simulation of random experiment of the different conceptual difficulties among students during probabilistic modeling.

II. EXPERIMENTATION: SIMULATION AND QUESTIONNAIRE

The student handover on the questionnaire, designed in an open manner, was done individually accompanied by the simulation of random experiment, over an overall average duration of two hours. The sample of 27 students surveyed, all volunteers, was extracted from the population of students who validated[1] the probability module in the third semester of the six-semester of third year (B. A). The choice of students having validated the module of probabilities, to better assert the credibility of the results obtained.

In addition to the choice of use of computer simulation, the choice of treated refers to a situation of gambling, and this for obvious epistemological reasons.

The situation-problem is stated first to the students in all its generality is as follows.

≪ A player throws successively and independently N times a coin. The probability of getting at each throw the "heads" side is p, with $0 < p < 1$. Each time he gets a heads he wins 1 DH, otherwise (if he gets a "tails") he loses 1 DH. Calculate the fortune of the player at the end of the game ≫.

The experimental protocol of the probabilistic treatment of this game situation is divided into the following steps:

- Identification of the parameters of the game situation, in this case is the probability p of winning at each part of the game and its intuitive interpretation [10] in terms of the equitable game.

Under what condition(s), the game is:			
Duration of a game	equitable	Favorable to the player.	Unfavorable to the player
$N = 2$	………	………	………
$N = 3$	………	………	………
$N = 5$	………	………	………
$N = 10$	………	………	………
$N = 100$	………	………	………

Table 1 : Intuitive exploration of game settings

- The treatment of the problem with simple values of its parameters, in this case the parameter N representing the duration of the game.

- The simulation of the game for various values of the parameters of the game $0 < p < 1$ which represents the probability of gain to a part of the game, and N the duration of the game.

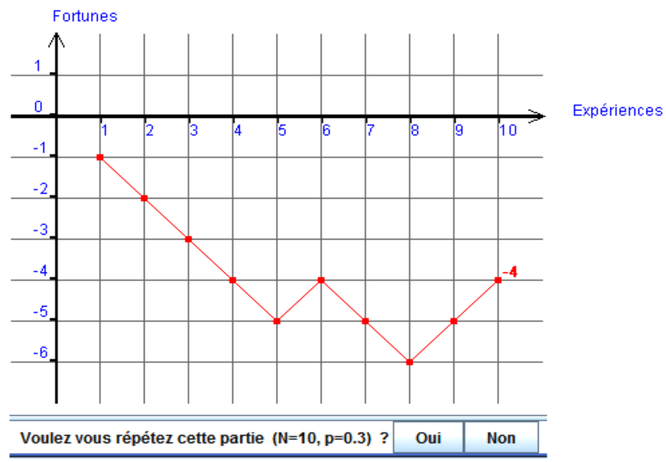

Fig 1:- Simulation of the game for $N = 10$ and $p = 0.3$

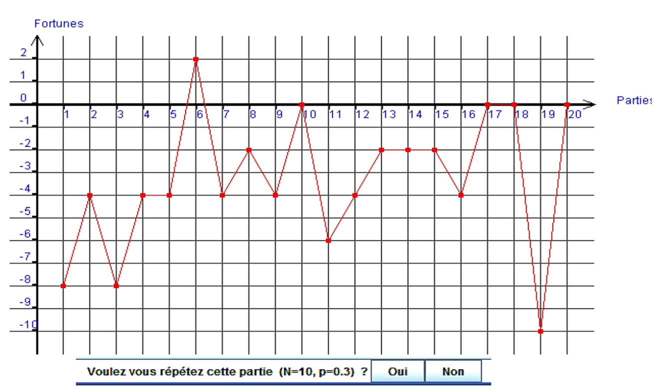

Fig 2 :- Repetition of the previous simulation a 20 times

[1] These students have eaither validated this module with an average greater than or equal to 10/20, or by compensation with other modules of the third semester of the basic licence

ISSN No:-2456-2165

- The comparison of the results of the simulation with those obtained during the treatment of the problem with simplified data of the parameters.

The experiment was conducted with 27students all volunteers, and lasted on average 2 hours with each of them.

The analysis of students' productions was done according to a qualitative procedure in order to obtain a classification of the conceptual representations identified. Indeed, the methodological approach is exploratory, with a reduced number of students (27 students), with in part individual observations of relatively long duration; it was relevant to qualitatively analyze students' answers one by one [11], by looking for common conceptual elements that would make it possible to categorize the students' conceptual representations. Thus, the analysis of the productions will be elaborated in the form of a classification of student conceptual representations about the notion of random variable and probabilistic modeling through a random variable concept.

III. RESULTS AND DISCUSSIONS

A. *Intuitive Exploitation of the Situation With Identification and Interpretation of the Parameters of the Problem*

Through this step, we put the student interviewed in front of an intuitive exploration of this game situation where he will be led to identify the parameters of the game with an intuitive interpretation of each of its parameters.

Question 1: On which condition (s) is the game equitable?

Question 2: Complete the table 1 with your answers.

This game is equitable if $p = 1/2$, favorable to the player if $p > 1/2$ and unfavorable to this player if $p < 1/2$.

In fact, 88% of students surveyed were able to give this interpretation:

Approach to equitable play conditions	Correct	Incorrect
Percentage	78%	22%

Table 2 : the intuitive exploration results

In question 2 of this step, the students surveyed also felt intuitively the conditions when the "Favorable game to the player" or "Unfavorable game to the player". Here, the results obtained in question 2:

Approach conditions when the game is favorable and unfavorable to the player	Correct	Incorrect
Percentage	21%	79%

Table 3: Intuitive exploration results when the game is favorable or unfavorable to the player

The results obtained on this question are very surprising! There is a complete reversal between the success and failure rates between questions 1 and 2. The fact of having integrated the parameter N, the duration of a game, explicitly in question 2 completely disrupted the intuitive approach. For example, the answers of the student E5:

Duration of the game	Equitable	Favorable to the player	Unfavorable to the player
$N = 2$	$p = 1/2$	$p = 1$	$p = 0$
$N = 3$	impossible	$p > 2/3$	$p < 1/3$
$N = 5$	impossible	$p > 3/5$	$p < 2/5$
$N = 10$	$p = 1/2$	$p > 6/10$	$p < 4/10$
$N = 100$	$p = 1/2$	$p > 51/100$	$p < 49/100$

Table 4: The Answers of the Student E5

B. *Problem solving for $N = 2$ then for $N = 3$, with $0 < p < 1$*

A tree procedure (using a weighted binary tree) or a complete counting procedure [3] [12] [5], both make it very easy to calculate the fortune f of the player at the end of a game:

For $N = 2$, we obtain $f = 2\,[p^2 - (1-p)^2] = 2(2p-1)$.

For $N = 3$, we get

$$f = 3[p^3 + p^2(1-p) - p(1-p)^2 - (1-p)^3] = 3(2p-1)$$

It is a question here of doing a complete combinatorial treatment, in the absence of a modeling of the problem by means of a relevant choice of random variables.

In this question, students are led to implement probabilistic modeling using random variables, in the absence of complete combinatorial processing. Moreover, it should be noted that prior to this question, the previous questions were intended to have a first intuitive input into this problem, by identifying the parameters of the problem, and their possible effect on the equitable game. The students' responses showed that 70% of them resorted to the notion of random variable, and that the remaining 30% were limited to simple probability calculations. We shared student responses into five categories.

Category1 (48%): "Modeling the situation-problem using a random variable designating the fortune of the player"

Approach	Correct	Incorrect
Percentage	54%	46%

Table 5: Category 1 approaches

Category2 (7%): "Modeling the situation-problem using a random variable denoting the gain of a player"

Approach	Correct
Percentage	100%

Table 6: Category 2 approaches

Category3 (15%): "Confusion between the player's fortune and the player's winnings"

Approach	Incorrect
Percentage	100%

Table 7: Category 3 approaches

Category 4 (11%): "Modeling through probabilistic treatment"

Approach	Incorrect
Percentage	100%

Table 8: Category 4 approaches

Category 5 (19%): "Others"

Approach	Incorrect
Percentage	100%

Table 9: Category 1 approaches

The fact of having proposed the treatment of the situation-problem in the particular cases $N = 2$ and $N = 3$, prior to simulation of the game, was doubly important in our experimental protocol:

- After identification and interpretation of the parameters involved in the situation-problem (previous questions), we put students in a probabilistic modeling situation in two simple cases ($N = 2$ et $N = 3$).
- After dealing with the problem in the two simple cases $N = 2$ and $N = 3$, we put the students in confrontation with the results provided in the first step, then allow them to perform simulations of the game with various values of the parameters of the problem, in the objective to prepare them for a better entry for the treatment of the general case of the problem.

Nevertheless, in some probabilistic treatments, especially in categories 3 and 4, we have again found conceptual difficulties in the interpretation of a mathematical expectation of a random variable, and especially in its application from its mathematical formulation.

C. *Comparison of the results between the intuitive exploration of the situation and the treatment of its cases N = 2 and N = 3*

This step is necessary in the protocol of our experimentation: the confrontation of the intuitive approach of the students to recognize the situation of equity of the game in the face of the results obtained as for the fortune of the player, at the end of the game in the particular cases $N = 2$ and $N = 3$.

Indeed, all of the students who correctly formulated the conditions of equitable game in questions 1 and 2 of step 1 of the experimental protocol, including those who gave extreme conditions, ie a total of 30% of students interviewed, concluded that there was a compatibility between the results in questions 1 and 2 and those in question 3: this is simply explained by the fact that all students have correctly resolved the problem in particular cases $N = 2$ and $N = 3$.

However, for the remaining students, the breakdown is as follows:

- A majority of 59% of students who recognize that there is an incompatibility between their results in questions 1 and 2 (step 1) and their result obtained in question 3: this is not surprising because all these students have given irrelevant answers to the three questions 1, 2 and 3. However, some of them continue to want to question the fact that an odd number of experiments in the game, is "certainly a cause" that could "explain" this difference between their intuitive approach and their result from question 3. Step 3 of the experimental protocol, will most likely be discriminatory for this category of students, for further processing of the problem situation "Fortune of a player."
- A minority of 11% who did not give an answer to question 3: these same students did not give answers to questions 1 and 2!

D. *Simulation of the game with different values of the parameters N and p for the estimation of the fortune of the player and implicit identification of the variables intervening in the modeling of the problem*

In this step, we try to explore the simulation of this game situation for different values of the parameters N and p, on the one hand to estimate the fortune of the player using a frequency approach at the end of a part of the game, and secondly to analyze and interpret the evolution of the game through representations provided during the simulation: the role of the simulation is to allow students to judge the effect of the parameters N and p on the fortune of the player, and especially to identify the relevant random variables for modeling the problem.

In the simulation of the random experiments of the situation-problem "Fortune of a player", we noticed essentially two difficulties with the students one of conceptual order and the other relating to the interpretation of the question posed by the chosen game situation.

A large number of students simulated the game for different values of the parameters N and p of the problem, without repeating the random experiment a sufficient number of times to be able to draw a suitable estimate of the expectation of the gain of the player.

For example, for the student E20, the number of simulation repetition of a random experiment under the same conditions does not exceed 5 times which is not in favor of good estimates of the "fortune of the player ".

Duration of a game	p	Estimation of Fortune	p	Estimation of Fortune	p	Estimation of Fortune
$N = 2$	0.5	1/2	0.4	-2	0.6	2
$N = 3$	0.5	1	0.3	-1	0.7	1
$N = 10$	0.5	2	0.25	0	0.75	-1
$N = 50$	0.5	-1	0.2	2	0.8	0
$N = 100$	0.5	1	0.1	1	0.9	2

Table 10: the estimation of the fortune of the player through the simulation by the student E20

The difficulty put forward by these students is certainly that of the unavailability of the frequency approach in the context of random experiments, which is essential for making estimates, especially for a mathematical expectation.

Moreover, students' lack of familiarity with (pseudo) random experiments (using a computer simulation) must also have had an effect on their experimental approach: multiply games, changing each time the values of the parameters N and p.

The unavailability of a frequency approach in a situation of simulation of random experiments is a conceptual difficulty for students who do not have a priori. The frequency approach represents a necessary (and minimal) condition in terms of conceptual representation in probabilistic situations, so that students can derive optimal benefit by repeating the random experiment (under the same conditions) a sufficient number of times for an exploration (for various estimates) of the results provided by the repetition of random experiment.

Note also that the practice of random experiments, as it is posited in the situation-problem of ruin of a player that we propose, is also an opportunity for students to empirically treat the results of some fundamental theorems, in probability theory, such as the law of large numbers and the central-limit theorem.

IV. CONCLUSION

The student simulation approach to estimate the "Player's Fortune" for various values of the N and p parameters, showed that not all students naturally had a frequency approach in a (pseudo) situation of random experiment. Some, such as students E15, E20 and E26, have operated according to an approach that does not translate the frequency aspect to make estimates, with one extreme case, which is that of the students E26 who does not have a frequency approach in probabilistic situation: it performs only one random experiment!

That said, we also found that as they performed simulations, students generally increased the number of repetitions of (pseudo) random experiments (with fixed N and p): this is therefore in favor of the development of their frequency apprehension of a probabilistic situation.

This repetition of (pseudo) random experiments was also, as was noticed (and expected) by our experimental protocol, a means (empirical tool) to better identify the effect of each of the parameters and on, first of all, the random evolution of a game situation, and then on "the fortune of the player".

Regarding the estimation of the "player's fortune" itself, except for students who have a frequency approach and who appropriately estimate "the fortune of the player", there remains a significant proportion (56%) of students who continue to erroneously estimate "the player's gain", by excluding certain results of random experiments that they consider "impossible", especially for $p = \frac{1}{2}$ and odd p.

However, in such situations, they see well evolution (pseudo) randomness of the game, as well as the fortune of the player at the end of the game!

This last estimate approach is of course wrong, especially when students do not refer to "the fortune of the player" but to the gain of the player at the end of the experiment (number of head).

Finally, we note that the students E1 and E3, using their simulations, have been able to conclude that the estimate of "the fortune of the player" in a game where the advantage of the player is p, must be the opposite of a game where the player's advantage would be $1 - p$.

The majority of students who have advanced in the exploration of the situation-problem, and who in this case have succeeded thanks to the simulation to identify and interpret in a relevant way the random variables involved in the modeling of this problem.

REFERENCES

[1] V. K. M. Abhinava, V. P. R. Enugala and A. K. M, “A Study on Digitalization Effects of Students Harassment by Educational Institutions in the Telangana State”, International Journal of Innovative Science and Research Technology, vol. 2, no. 11, pp. 220-223, 2017.

[2] P. Priyanka and D. A. J. Agrawal, “Smart Farming System Using Data Mining”, International Journal of Innovative Science and Research Technology, vol. 2, no. 6, pp. 376-380, 2017.

[3] M. ZAKI and F. Pluvinage, “Démarche de résolution et de simulation face au problème de la ruine d’un joueur”, Educational Studies in Mathematics, vol. 22, no. 2, pp. 149-181, Avril 1991.

[4] B. Parzysz, “De l’expérience aléatoire au modèle, via la simulation”, REPERES - IREM. N° 74, pp. 91-103, 2009.

[5] M. Zaki, “Traitement de problèmes de probabilités en situation de simulation, Thèse”, Université de Strasbourg, 1990.

[6] J. Bordier, “Définition, construction et expérimentation de logiciels de simulation sur ordinateur, pour l’enseignement de la probabilité”, Université Paris VII, Paris, 1991.

[7] C. d. Q. e. S. Coutinho, “Introduction aux situations aléatoires dès le collège: de la modélisation à la simulation d’expériences de Bernoulli dans l’enseignement informatique Cabri Géomètre II”, Grenoble, 2001.

[8] F. Gaydier, “Simulation informatique d’expérience aléatoire et acquisition de notions de probabilité au lycée”, Paris, 2011.

[9] H. Amrani and M. ZAKI, “Student’s Conceptual Difficulties With respect To the Notion of Random Variable”, International Journal of Education, Learning and Development, vol. 3, no. 9, pp. 65-81, 2015.

[10] M.-P. LECOUTRE and E. FISCHBEIN, “Évolution avec l’âge de « Misconceptions » dans les intuitions probabilistes en France et Israël”, vol. 18, no. 3, pp.311-332, 1998.

[11] M. Zaki, “Acquis et applications de la didactique des mathématiques : quelques résultats méthodologiques et de recherches au niveau universitaire”, Annales de didactique et de sciences cognitives, vol. 9, pp. 61-66, 2004.

[12] M. ZAKI, “Acquis et applications de la didactique des mathématiques : quelques résultats méthodologiques et de recherches au niveau universitaire”, Annales de didactique et de sciences cognitives, vol. 9, pp. 61-66, 2004.