International Journal of Education, Learning and Development
Vol.3, No.9, pp.82-98, December 2015
Published by European Centre for Research Training and Development UK (www.eajournals.org)

STUDENT'S CONCEPTUAL DIFFICULTIES WITH RESPECT TO THE NOTION OF RANDOM VARIABLE

Hamid Amrani and Moncef Zaki

Interdisciplanary Laboratory of Research in Didactics of Sciences and Technology Faculty of Sciences Dhar El Mahraz – Sidi Mohammed Ben Abdellah University, Fez, Morroco

ABSTRACT: *In this article, we are interested in an introductory teaching of the probabilistic formalism at university level, in particular around the notion of random variable. Our research hypothesis is that a teaching based on a formal approach, even if it is intended for second year students of the Bachelor of Science degree, can be doomed to a didactic failure. Our study with a small number of students, but over a long duration of observations, has allowed to raise various conceptual difficulties and obstacles around the definition and production of random variable examples. The difficulties that impede the availability of this object are mainly due to conceptual confusions between the concept of random variable and the notions of image universe, random experiment, or law of probabilities. A quantitative analysis of the productions of students showed that the relevance of the formal approach was without effect on the production of example, whereas that of the intuitive approach had an effect on the validity of the production of example of random variable. These results encourage thus the adoption of a dialectical formalism/intuition in the introduction of the probabilistic formalism; such an approach of teaching would seem to be a priori quite in favor of a good apprehension by the students.*

KEYWORDS: Didactics of Mathematical Sciences, Random Variables, University, Formal and Intuitive Approaches, Conceptual Difficulties.

INTRODUCTION AND PROBLEMATIC

The teaching of the theory of probabilities is exempted at university from the beginning of the Bachelor of Science degree (BS degree) in the form of lectures (2 hours/week) and directed working sessions (2.5 hours/week), during all the first semester of the second year of the BS degree (15 weeks/semester). The definition of a probability is then systematically introduced in an axiomatic way, such as it is advocated by the official university programs. That generates thus from the start enormous conceptual difficulties for the students.

The latter do not manage to easily establish a natural and intuitive link between the random situations of real life and the theory of probabilities (H. Steinbring, 1991). In addition, probabilistic modeling in terms of random variable, being closely related to the axiomatic definition of a probability, generates even more difficulties for the students at the time of treatments of the random variables.

The probabilistic modeling in the BS degree cycle, through random variables defined on probabilized spaces, represents a fundamental basis in the course of the students, because the final objective is to make students acquire, at the conclusion of the BS degree, the mathematical tools for the treatment of forecasts, the estimate of parameters by confidence intervals, or the treatment of decision through the notion of statistical tests.

ISSN 2054-6297(Print), ISSN 2054-6300(Online)

The concept of random experiment is not often privileged in the fundamental teaching of probabilities at university, in situation of probabilistic modeling (see in this respect also, Inter-IREM Commission, 2001 and M. Henry, 2009). The epistemological analysis of the genesis of the definition of probabilities and the notion of random variable heavily supports this difficulty encountered in their didactic transposition in teaching situation (C.S. Castro, 1998). The historical example of the famous *Chevalier de Méré*'s Problem, although it is elementary, is a perfect illustration in terms of epistemological difficulty in probabilistic modeling (M. Zaki, 1992).

The formal standardization of the teaching of the concepts of random variable and law of probabilities, reinforced by exercises during directed working sessions focusing on the formal aspect of these concepts, if not calculative, caused the students to encounter various conceptual difficulties, some of which are the following:

- Students talk about laws of probabilities without specifying beforehand the corresponding random variable.
- Students make confusions between the universe on which a random variable is defined and the image universe by the so called random variable.
- Students have many difficulties to construct a random variable and identify its law of probabilities vis-à-vis a probabilistic situation.

The problematic of this research relates to the identification and analysis of difficulties encountered by students regarding the notions of random variable and law of probabilities. Thus, we call in question the very formal approach to the University of the teaching of these two concepts, whose didactic consequences enable us to formulate the two followings hypothesis of research:

First hypothesis of research: the probabilistic formalism of the notions of random variable and law of probabilities are complex and difficult to understand and interpret.

Second hypothesis of research: The lack of reference and even the absence of probabilistic situations in the university education, making use of random experiments privileging an intuitive approach of the concepts of random variable and law of probabilities, impede a good apprehension by the students of these two concepts, and thus a good appropriation of the probabilistic formalism (see also in this regard B. Greer, 2001).

In the following paragraph we will see that the epistemological analysis of the genesis of the concepts of random variables and laws of probabilities, largely corroborate these two hypothesis.

Epistemological analysis

The definition of the concepts of random variables and of laws of probabilities witnessed an evolution intimately linked not only to that of the definition of a probability, but also to the development of the concept of function and its properties whose emergence of rigorous definitions appeared only around the second half of the 19th century especially thanks to Weierstrass in 1861 (A. Benbachir, 2002). In his thesis, A. Benbachir retraces all the epistemological difficulties and obstacles of the emergence of rigorous definitions about of function concepts, of continuity and derivability, of which the emergence took place around

ISSN 2054-6297(Print), ISSN 2054-6300(Online)

the end of the 19th century thanks in particular to works of Weierstrass, Bolzano, Cauchy and Riemann.

The epistemological study also showed that this evolution has been highly dependent on the nature of the contexts and the problems which followed one another to lead to the genesis of the concept of probability, of random variable and law of probabilities (I. Todhenter, 1865; N. Meusnier, 2006).

During the *initial period of the theory of probabilities* (early 17th century until the late 19th century: Huygens to Laplace), the context was primarily that of the games of chance, in which one sought to solve problems related to the distribution of stakes. The notion of mathematical expectation was present without for as much making a reference to an established concept of random variable; it is in particular the case for Pascal (1623-1662) and Fermat (1601-1665) in the problem of points. In his book *De ratiociniis in ludo aleae*, devoted to the games of chance, Huygens (1629-1695) manages to give a rather precise definition of the expected gain of a player that represents for this latter: "*the sum that one must pay to enter a fair game giving the same results as the initial game*". The definition is rather interesting, only the problems dealt by Huygens depend on the concept of *fair game*, and the concept of random variable is not explicit in it, it is precisely replaced by the consideration of fair game. Despite this "insufficiency" of probabilistic modeling, Huygens manages to state calculation rules of the value of expectation, and gives applications for the calculation of *life annuities*.

Laplace (1749-1827), on the other hand, proceeded to a synthesis of all knowledge on the probabilities of his time: he has the merit to have made an important contribution to the development of the probabilities, comparable with that of Euclide for geometry. Although he remained in the context of the games of chance, he succeeded in giving a definition of expectation that is more elaborate than that of Huygens, and more direct, because it does not refer to the concept of fair game, it "*generally expresses the advantage of someone who waits for an unspecified good in suppositions which are only probable. This advantage, in the theory of the chance, is the product of the sum expected by the probability of obtaining it: it is the partial sum which must return when one does not want to run the risks of the event, supposing that the distribution is done proportionally with the probabilities (…) When the advantage depends on several events, one obtains it by taking the sum of the products of the probability of each event by the welfare attached on its arrival*" Laplace (p. XIX).

It is clear that in his definition of the mathematical expectation, Laplace does not explicitly define the concepts of random variable and law of probabilities; nevertheless, all the mathematical ingredients of these concepts appear in this definition; moreover, the reasoning which Laplace holds for example in the treatment of the problem of the "*tournament*" or that of the "*parties*", lend themselves easily to a treatment using the concept of random variable.

During this initial period of the theory of probability, we find that the treatment of the problems of chance was done primarily in a *context of games*, with problems referring to the *gain* and the *expectation of gain*: in this type of context, these notions remain *intuitive* and *natural*. Thus, the treatment of these notions was done in an *intuitive way*, without requiring any explicit way of *formal mathematical definitions* of notions of random variables and law of probabilities. It is finally worth noting that at the same time, the notion of function also was very intuitive and not yet established, and this has certainly also contributed to the difficulty of immediate emergence of the *formalism of the theory of probability*.

ISSN 2054-6297(Print), ISSN 2054-6300(Online)

At the end of the 19th century, the theory of probability witnessed a historical turning point thanks to the development of the theory of measurement, in particular by the introduction of the integral of Lebesgue (1875-1941) and the measurable sets by Borel (1871-1956). In fact, the context of theory of probability evolved towards situations of measurement: errors of measuring of physical quantities and calculations of uncertainties, polls and forecasts,… and the notion of probability distribution was imposed by itself; it is thus not any more a question of studying the individual events, but rather the whole systems of random parameters. Thus, the theory of probability gained a new conceptual point of view, in which the concepts of random variable and law of probabilities had absolutely their place.

In 1901, Lebesgue presents his theory of integration which generalizes that of Riemann (1826-1866), where the concepts of random variable X and mathematical expectation $E(X)$ are well established. Borel supplements this work, by publishing in 1909 a research on the countable probabilities, where a probability is likened with a measurement having the property of □-additivity. This same work allowed to Lomnicki and Steinhaus to publish respectively in 1923, the works "*Nouveaux fondements du calcul des probabilités*" and "*Les probabilités dénombrables et leur rapport à la théorie de la mesure*". Lastly, at the same time Paulo Levy (1886-1971) and Alexandre Khintchine (1894-1959) founded the modern theory of the random variables.

In 1933, Kolmogorov (1903-1987) publishes his work of "*Bases of the theory of probability*", where he presents an *axiomatic theory of probabilities*: the axiomatic definition of the probabilities has thus completely freed the theory of probability from the problems of the current life. In his turn, Kolmogorov played for the probabilities the role which Hilbert played for the geometry (for the anecdote, in 1954, Kolmogorov and his pupil Arnold Vladimir have succeeded in solving the 13th problem of Hilbert "the resolution of an algebraic equation of degree 7 by means of functions of two variables": the answer is no). Since then, thanks to the formal character of this axiom, other mathematical theories related to the theory of probability were able to develop: the stochastic processes by Kolomogorov and Joseph Doob (1910-2004), the theory of potential by Kolmogorov and Paul André Meyer (1934-2003), or in mathematical statistics, tests of hypothesis by Karl Pearson (1857-1936) and tests of significance by Fisher (1890-1962).

Thus, the axiom of Kolmogorov, through which the concepts of random variable and law of probabilities have benefited from a formal mathematical definition, made it possible to develop mathematical theories of *modeling of all the random situations*, where the *law of great numbers*, the *central-limit theorem* and the *law of the iterated logarithm*, played a fundamental role. Thus, within the framework of the axiomatic theory of Kolmogorov, the situations of games of chance are no more than simple particular cases of probabilistic situations.

This epistemological analysis shows the great importance of the intake of the intuitive ground to the emergence of the notions of random variable and law of probabilities under their formal aspect, and this, during almost three centuries. This strongly reinforces both hypothesis of research of our problematic; a teaching which would make a dead end on the intuitive aspects and which only privileges the formal aspect of these notions will be generally doomed to a didactic failure. Thus, contrary to models of teaching which are more based on the intuitive representations, a teaching based on the presupposed theoretical ones cannot always guarantee the availability of the notions of random variable and law of

ISSN 2054-6297(Print), ISSN 2054-6300(Online)

probability (see also in this regard Marie-Paule Lecoutre, 1998). The results of our experimentation conducted on the students will to show the founded good of our hypotheses.

Experimentation: methodological approach and questionnaire

Our research question results initially from effective observations on the ground in connection with the probabilistic difficulties encountered by our students trained at the university. In addition, the epistemological study enabled us to make a first analysis of the origin of these difficulties and to assert our hypothesis of research. We have thus privileged an exploratory approach during our research to identify well and better analyze these difficulties. We have then used a questionnaire, in which we questioned the students on different aspects related to the notions of random variable and law of probabilities, their natures, on the meaning of the parameters of a random variable and their interpretation; until exploring the availability of these notions in situation of elementary probabilistic modeling.

In this article, we will limit ourselves to the analysis of the part of the questionnaire related to **aspects concerning the definition of a random variable, as well as to production of examples of random variables**: this part alone is already very rich in information regarding our general research question of probabilistic modeling, and in particular in connection with the difficulties and conceptual representations of the students vis-à-vis the statute of a random variable.

The handover of the questionnaire to the students, designed in an opened way, was done in an individual way, in a global average duration of two hours. The sample of the **25 questioned students**, all volunteers, was extracted from among the students who have validated the module of probabilities in the third semester of the BS degree cycle (consisting of six semesters): these students have either validated this module with an average superior or equal to 10/20, or by compensation with other modules of the third semester of the BS degree.

The choice of students having validated the module of probabilities is not anodyne. Indeed, in order to better assert the results of our analyses, and while remaining in conformity with our hypothesis of research, we seek to analyze the implementation of a formal teaching about the random variables on the conceptual representations dealing with this notion, of students who have validated the aforementioned probabilities module.

Lastly, to locate the contents of teaching of this module, we will specify that the latter is synthetically established starting from the probabilized spaces introduced by the means of the axiomatic definition of a probability, of random variables as measurable functions defined on probabilized spaces, of laws of probabilities as images-probabilities, the classical definitions of the parameters of random variables and their properties, with a great part devoted to the investigation of the usual discrete and continuous laws of probabilities.

The analysis of the productions of the students was done according to a qualitative procedure, in order to obtain a classification of the raised conceptual representations. Indeed, as the methodological approach is exploratory, at a reduced number of students (25 students), with on the other hand individual observations of relatively long durations; it was relevant to qualitatively analyze one by one the answers of the students (M. Zaki, 2004), by seeking the common conceptual elements which would make it possible to proceed to a categorization of the conceptual representations of the students: thus, the analysis of the productions will be elaborated in the form of a classification of the conceptual representations of the students in connection with the notion of random variable.

ISSN 2054-6297(Print), ISSN 2054-6300(Online)

Analysis of the productions of the students: classification of the conceptual representations vis-à-vis the notion of random variable

Category 1: «The random variable is an application»

Four out of the twenty five questioned students have defined a random variable simply as "*an application*". The distribution of the examples provided by these students to illustrate a random variable is the following one:

Table 1: Distribution of the examples illustrating a random variable according to the definition «an application»

		Nature of answer	**Total**
Definition	The random variable is an application	Fausse	**4**
Examples	« One throws a balanced coin, that is to say X the random variable which defines the number of tails obtained in the k^{th} launch » « Throw of dice for which one can obtain an even number»	Correct	2
	« One can give as a random variable the fact of launching a coin»	False	1
	No answer	No answer	1

The definition of a random variable here originates in a conceptual representation which refers to a *formal approach*: an application. This definition is very partial, and thereby, it is erroneous. Only two students out of four provide an example in a correct way. This example, although it refers to a random experiment, remained intimately related to the character of application of a random variable that is to say, still according to a *formal approach*.

The second provided example assimilates a random variable to a random experiment: this confusion clearly translates a conceptual difficulty as to the representation of this student to the notion of random variable (see also, Inter-IREM Commission, 2001).

The fourth student does not provide an example.

Category 2: «The random variable is a measurable application»

Six students out of the twenty five questioned defined a random variable as "*a measurable application*". The distribution of the examples provided by these students to illustrate a random variable is the following one:

ISSN 2054-6297(Print), ISSN 2054-6300(Online)

Table 2: Distribution of the examples illustrating a random variable according to the definition «a measurable application»

		Nature of answer	Total
Definition	The random variable is an measurable application	False	**6**
Examples	«One throws two dice, one considers the random variable which associates each launch of the two dice to the sum of the obtained points»	Correct	1
	«One has one discrete random variable, and one continuous random variable» «For a throw of dice, the number of the points which it carries. The whole of the possible results is a random variable» « Let X be a real random variable and let $a \in \mathbb{R}$ $(X = a) = \{\omega \in \Omega / X(\omega) = a\}$ $(X \leq a) = \{\omega \in \Omega / X(\omega) \leq a\}$ $(a \leq X \leq b) = \{\omega \in \Omega / b \leq X(\omega) \leq a\}$ » « Throw a coin twice, X: the number of heads; $\Omega = \{0; 1; 2\}$ » «One launches a coin with $\Omega = \{heads; tails\}$ with $X(heads) = \frac{1}{2}$ and $X(tails) = \frac{1}{2}$»	False	5

The definition of a random variable originates here in a conceptual representation which refers to a *formal approach*: a measurable application. This definition is partially correct, because it does not refer to a probabilized space on which a random variable must be defined. As with previous category 1, a student out of six will provide the same type of correct example, which occurs according to a *formal approach* related to the character of application of a random variable, which moreover also finds himself in the definition provided by these students.

Five other examples, which are false, were provided by the rest of the students of category 2:

- The first one refers to the discrete and continuous random variables: it is here a confusion between an example of random variable and the nature of a random variable. One notes that the very formal character of the exempted teaching prevented the availability of the random variable object in terms of example.

- The last four examples translate coarse conceptual confusions between a random variable and its image-universe, the events which it can generate, or the probability of probabilized space on which the random variable is defined. It is once again the fruit of a very formal teaching.

ISSN 2054-6297(Print), ISSN 2054-6300(Online)

Category 3: «The random variable is a measurable application defined on a universe of probabilized possibilities»

Five students out of the twenty five questioned defined a random variable as "*a measurable application defined on a universe of probabilized possibilities*". The distribution of the examples provided by these students to illustrate a random variable is as follows:

Table 3: Distribution of the examples illustrating a random variable according to the definition «a measurable application defined on a universe of probabilized possibilities»

		Nature of answer	Total
Definition	The random variable is a measurable application defined on a universe of probabilized possibilities.	Correct	**5**
Examples	«The consumption of milk in liters per family is a random variable» «An example of the random variable it is $Random(x)$ »	Correct	2
	«One considers two balanced dice that one throws only once, having: $\Omega = \{1; 2; 3; 4; 5; 6\}^2$; $\mathcal{F} = \mathcal{P}(\Omega)$, one is interested in the sum S of the values indicated by the two dice, S takes its values in: $\Omega' = \{2; 3; 4; 5; 6; 7; 8; 9; 10; 11; 12\}$, S is then an application of Ω in Ω' defined by $S(\omega, \omega') = \omega + \omega'$ so $S: (\Omega, \mathcal{P}(\Omega)) \to (\Omega', \mathcal{P}(\Omega'))$ is a random variable» « One launches two dice balanced simultaneously, one is interested in the sum of the points obtained, $\Omega = \{1; 2; 3; 4; 5; 6\}^2$ et $\Omega' = \{2; 3; ...; 12\}$, so one notes $X: \Omega \to \Omega'$ with $X(x_1, x_2) = x_1 + x_2$ »	partially correct	2
	« One launches a balanced dice, then one has six possible values $\{1; 2; 3; 4; 5; 6\}$, one defines a random variable: $X : X \in \{1; 2; 3; 4; 5; 6\}$ »	False	1

Contrary to the two preceding categories 1 and the 2, where the provided definitions were partial and thus erroneous, here the definition of a random variable is correct, because the students specify that it is a measurable application defined on a probabilized space that they express by «a universe of probabilized possibilities». However, the definition is once again formal.

Two correct examples, provided according to an intuitive approach, accompany this definition (2 students out of five):

ISSN 2054-6297(Print), ISSN 2054-6300(Online)

- The function «*random (x)*», that originates in a function of computer applications, does constitute well an intuitive example of random variable. Nevertheless, the student which presents this example does not show a true availability of examples of random variable arising from his formal definition (see also, Inter-IREM Commission, 2001).

- The second example related to the «milk consumption per family» is intuitively correct. It does not at all reflect the formal aspect of the definition of a random variable such as it was given by this student: this student shows a hybrid availability of the notion of random variable, both formal and intuitive.

- Two students out of five suggest the same partially correct example, where the students do not specify the definition of a probability of probabilistic space on which the random variable is defined. This example is presented in a formal way, and returns directly to the contents of the formal definition provided in this category.

- The last example which is provided is false. It presents as we have already seen in category 2, a conceptual confusion between a random variable and its image universe.

Category 4: «The random variable binds the results of a random experiment to real values»

Three students out of the twenty-five questioned give the definition "the random variable binds the results of a random experiment to real values". The distribution of the examples provided by these students to illustrate a random variable is the following one:

Table 4: Distribution of the examples illustrating a random variable according to the definition «The random variable binds the results of a random experiment to real values»

		Nature of answer	**Total**
Definitions	«The random variable is a real number $X \in \mathbb{R}$ which binds the results of an experiment with a real number, $X : \Omega \to \mathbb{R}$ with Ω is the set of probabilized space»	Correct	**1**
	«A random variable is a magnitude which depends on the result of the experiment» «The random variable is a variable which takes unforeseen values»	partially correct	**4**
Examples	«X is the number of times that one can obtain an ace after three iterations» «The sum of numbers of two randomly	Correct	2

ISSN 2054-6297(Print), ISSN 2054-6300(Online)

	drawn numbered balls »		
	«Size, Color »	partially correct	1
	«One launches a dice in the air, let X the random variable which reads the number obtained, so $X = \{x_1, x_2, x_3, x_4, x_5, x_6\} = X(\Omega)$; $X(\Omega) = \sum_{i=1}^{6} x_i$ » «Number of heads obtained for an infinite number of throws, the random variable X can take values in the set $F = \{0; 1; 2; \dots; +\infty\}$»	False	2

The only correct definition of a random variable provided here, is founded on an *intuitive interpretation*, which is formulated according to a *formal approach*. Otherwise, the two other definitions, which originate in an *intuitive approach*, are partially correct; because they do not specify the nature of the probabilized space to which should belong the universe on which the random variable is a priori defined.

In reference to this intuitive definition, two students provide two correct examples, drawn from situations of relevant random experiments. A third student cites the example of "size and color", which is partially correct: initially because the color refers to a qualitative character; otherwise for "the size", the student does not give any precision on the nature of the random experiment, and in particular on the probabilized space on which the «size» random variable would a priori be defined.

Finally, the last two students, although they specified the random experiment of reference, their examples of random variables are false: in the first example, the student in question makes several conceptual confusions of a random variable, in particular with the image-universe; otherwise, for the second provided example, this one does not correspond to a real random variable.

Category 5 : «The random variable represents the results during an experiment which proceeds randomly and spontaneously» – «The random variable is an arbitrary variable»

Two students out of the twenty five questioned give the definitions «the random variable represents the results during an experiment which proceeds randomly and spontaneously» – «The random variable is an arbitrary variable». The distribution of the examples provided by these students to illustrate a random variable is the following one:

ISSN 2054-6297(Print), ISSN 2054-6300(Online)

Table 5: Distribution of examples illustrating a random variable according to the definitions «The random variable represents the results during an experiment which proceeds randomly and spontaneously» - «The random variable is an arbitrary variable»

		Nature of answer	Total
Definitions	« A random variable is a variable which represents the results during an experiment which proceeds randomly and spontaneously» « A random variable is an arbitrary variable that is chosen arbitrarily»	False	**2**
Examples	« For example, if one throws a dice, there are 6 possible results $\{1; 2; 3; 4; 5; 6\}$ » X is a random variable	False	2

The first definition is erroneous, because it does not give any precision on the nature of application of the random variable, and does not make any reference to the probabilized space on which it is defined. The first example, which is partially correct, is given by the same student who provided this definition: one notes that he reproduces exactly its definition of throw of dice.

As for the second definition, which is also false, it translates a conceptual confusion concerning the "random" term in the "random variable" expression, and which the student translated with the term of "arbitrary" term. The example provided by this student is not surprising: "X". It is indeed a letter commonly used in mathematics to denote an arbitrary value!

Category 6: «The random variable is a continuous function»

One student out of the twenty five questioned gives the definition "The random variable is a continuous function". The example provided by this student to illustrate a random variable is the following one:

Table 6: Example illustrating a random variable according to the definition« The random variable is a continuous function »

		Nature of answer	Total
Definition	The random variable is a continuous function	Partially correct	**1**
Example	« The diameters of the tubes in a factory»	Correct	1

The definition of a random variable as "continuous function" is partially correct, because it is only one special case of random variable; and the example provided here, is a direct representation of this. One could allot the answers of this student to an effect of teaching,

ISSN 2054-6297(Print), ISSN 2054-6300(Online)

which remains formal, and in which the productions of random variables are often continuous functions. Moreover, the example mentioned by this student is only a simple reflection!

Category 7: «The random variable serves to calculate a probability»

Two students out of the twenty five questioned give the definition «The random variable serves to calculate a probability». The distribution of the examples provided by these students to illustrate a random variable is the following one:

Table 7: Distribution of examples illustrating a random variable according to the definition «The random variable serves to calculate a probability»

		Nature of answer	Total
Definitions	«The random variable is a problematic element or question and contains a certain number of solutions or probabilities of emergences» «A random variable is a variable that helps us calculate a particular probability»	False	**2**
Examples	«$X \sim (\mu, \sigma)$, $\tilde{B} \sim (p, n)$ $Binomial$»	Partially correct	1
	« Let X be a random variable that determines the number of natural childbirths among women»	False	1

The two provided definitions, which originate from an intuitive approach, return rather to the distribution of a random variable, and so the approach of this definition reveals a certain conceptual confusion between a random variable and its law of probabilities; moreover, the two provided examples confirm this ambiguity for these two students.

Availability of the object «random variable» : production of examples

The previous analysis shows that the production of definitions and examples by the students originates from one of either approaches, namely formal or intuitive. The question is to know if the nature of the approaches, formal and intuitive, has an effect on the relevance of production of examples of random variables. For this reason, we will carry out a quantitative analysis on a relevant availability of the random variable object, by crossing the answers of the students about their productions of definitions and examples of random variables.

Effect of the (formal or intuitive) approach of the definition on the relevance of production of example of random variable

The first observation to be made is that 76% of the whole of the questioned students have provided a false definition of a random variable, among whom 58% have adopted a formal approach of the definition. A priori, the formal approach seems to have an effect on the treatment of the random variables and in particular on the availability of this mathematical object. Independently of the (correct or false) validity of the definition provided by a student,

ISSN 2054-6297(Print), ISSN 2054-6300(Online)

we initially sought to know whether the nature of his approach (formal or intuitive) in the definition could have an effect on the (correct or false) validity of the example which he provided. Among all the 25 questioned students, after a dividing in two under populations making respectively a reference to the students who provided a formal definition and to those who provided an intuitive definition, we crossed the two qualitative variables, the D qualitative variable : «Approach of the definition of a random variable», by assigning it two modalities, «formal» and «intuitive», with the E qualitative variable: «Validity of the provided example», by assigning it two modalities «right» and «false». We thus obtained the following table:

Example / Definition	False	Right	
Formal	14	7	21
Intuitive	4	0	4
	18	7	25

Table 8: Crossing of variables «Approach of the definition of a random variable» and «Validity of the example provided» by 25 questioned students

The vast majority (84%) of the students has provided a formal definition. That can be explained by an effect of teaching which, let us recall it, is primarily focused on a formal approach. In addition, it is noted that 33% among the students who provided a formal definition, produced a correct example; whereas none the students who have provided an intuitive definition, has produced a correct example. It is then normal to be favorable to the hypothesis H_1 «A formal approach in the definition has an effect on the production of a correct example of random variable». We then considered H_0 as a null hypothesis, the opposite of the H_1 hypothesis, which is to say H_0 «The approach adopted in the definition does not have an effect on the validity of the produced example». Thus, under the hypothesis H_0 and at fixed margins, knowing that the theoretical numbers are not all higher than 5, we used the exact test of comparison of Fischer. We thus calculated the probability p of obtaining such a distribution of answers (see table 8):

$$p = \frac{C_{18}^{14}.C_{7}^{7}}{C_{25}^{18}} = 0{,}0063 < 0{,}01$$

It is worth noting that one can only consider this single distribution of answers for this calculation of probability, since there is none that is more extreme under the null hypothesis H_0.

Thus, the exact test of Fischer is very significant at the level of 1%, and thus we can reject the null hypothesis H_0. However, if this test confirms a certain effect of the formal approach in the definition of a random variable on the production of a correct example, it does not inform us similarly on the following question: Which is the effect of the relevance or irrelevance of the formal approach, respectively intuitive, of the production of example, on the (false or correct) nature of the validity of such an example. Indeed, the formal or intuitive approach held in the definition produced by the students was not systematically renewed in their approach of production of their examples: some students, who adopted a formal approach in the definition, have produced examples according to an intuitive approach (see for example category 3 of §4), and conversely (see for example category 6 of §4). We will thus refine our quantitative analysis even more, by studying this question distinctly with both under-

ISSN 2054-6297(Print), ISSN 2054-6300(Online)

Published by European Centre for Research Training and Development UK (www.eajournals.org)

populations: the one which provided examples according to a formal approach and the one which provided examples according to an intuitive approach.

Effect of the relevance of the (formal or intuitive) approach of the example on the (correct or false) nature of its validity

Effect of the formal approach in the production of example on its validity

Sixteen out of the twenty five questioned students (64 %) have adopted a formal approach in their production of example. Their formal approach, relevant or not compared to the provided example, produced an example whose validity could be correct or false. We thus considered two qualitative variables: F «Formal approach» in two modalities «Relevant» and «Not relevant», and E «Example» in two modalities as well «Right» and «False». The crossing of these two variables is the following:

Table 9: Crossing of the variables «Formal approach» and «Example» with the 16 students who have adopted a formal approach of production of example

Example / Formal Approach	False	Right	
Relevant	3 3,125	2 1,875	5
Not Relevant	7 6,875	4 4,125	11
	10	6	16

The comparison of the observed numbers to the theoretical numbers, under the H_1 hypothesis of independence of these two variables (in terms of cause with effect), is a priori in favor of this hypothesis. The crossing relates to the same population of students who have adopted a formal approach of production of example (16 students), consequently the application of a test of independence is completely justified. The fact that some theoretical sums are lower than 5, does not allow to apply directly a test of Chi2; consequently we will use once again an exact test of Fischer. We will retain for null hypothesis H_0 the opposite of H_1 that is to say H_0 «The relevance of the formal approach of production of example has an effect on the validity of the example». Let us then calculate, under this hypothesis H_0, at fixed margins, the probability p of the observed distribution (see table 9) and of those which are more extreme.

$$p = \frac{C_{10}^{3}.C_{6}^{2} + C_{10}^{2}.C_{6}^{3} + C_{10}^{1}.C_{6}^{4} + C_{10}^{0}.C_{6}^{5}}{C_{16}^{5}} = 0{,}006 < 0{,}01.$$

The test is very significant at the level of 1%, and thus the null hypothesis H_0 is rejected: the relevance of the formal approach in the production of example has no effect (is independent) on its validity.

Effect of the intuitive approach in the production of example on its validity

ISSN 2054-6297(Print), ISSN 2054-6300(Online)

International Journal of Education, Learning and Development
Vol.3, No.9, pp.82-98, December 2015
Published by European Centre for Research Training and Development UK (www.eajournals.org)

Nine out of the twenty five questioned students (36 %) have adopted an intuitive approach in their production of example. Again, their intuitive approach relevant or not compared to the provided example, produced an example whose validity could be correct or false. We thus considered two qualitative variables: I «Intuitive approach» in two modalities «Relevant» and «Not relevant", and E «Example» in two modalities «Right» and «False». The crossing of these two variables is the following:

Table 10: Crossing of variables «Intuitive approach» and «Example» with 9 students who have adopted an intuitive approach to production example

Example / Intuitive approach	False	Right	
Relevant	1 / 0,78	0 / 0,22	1
Not Relevant	6 / 6,82	2 / 1,78	8
	7	3	9

Under the hypothesis H_1 «The relevance of the intuitive approach in the production of an example is independent of the validity of the example», the observed numbers are close to the theoretical numbers, which is in favor of the hypothesis H_1. We will thus nullify hypothesis H_0, the opposite of H_1, namely H_0 «The relevance of the intuitive approach has of the effect on the validity of the example». Due to the theoretical numbers which are very weak, we used again the exact test of Fischer. We thus calculated under the null hypothesis H_0, at fixed margins, the probability p of obtaining the observed distribution (see table 10) and one that is more extreme (there is just one):

$$p = \frac{C_7^1.C_2^0 + C_7^0.C_2^1}{C_9^1} = 1$$

The test is far from being significant, we cannot thus reject the null hypothesis, and consequently, the relevance of the intuitive approach in the production of an example has some effect on the validity of the example.

CONCLUSION

The introduction to the calculation of probabilities in the second year of BS degree at the Moroccan university aims to introduce the students to the probabilistic formalism. All the probabilistic contents are introduced in a formal way, in conformity with the official program. Only during courses and tutorials, we noted that the students have enormous difficulties in their probabilistic apprehensions, undoubtedly due to a very formal teaching.

A first epistemological analysis on the emergence of the notion of random variable showed all the conceptual complexity that had to be overcome before reaching the mathematical

ISSN 2054-6297(Print), ISSN 2054-6300(Online)

formulation of its formal definition, independently of the situations of games which often were more privileged in the probabilistic treatments. Nevertheless, it should be recognized that the contribution of the intuitive ground, thanks in particular to the games of chance, has been of a great interest in the emergence of notions of random variable and law of probabilities under their formal aspect.

This epistemological analysis corroborates the hypotheses of our research question: the notion of random variable is very complex, and therefore, a teaching which would make a dead end on the intuitive aspects and which only privileges the formal aspect of this notion will be generally doomed to a didactic failure.

Indeed, our experimentation with a reduced number of students (25 volunteers), but which on the other hand was conducted over a long period (a 2 hours average per student, a total of about fifty hours of observations), allowed to detect various conceptual difficulties in the production of the definition of a random variable. Moreover, the vast majority of the students (84%) adopted a formal approach rather than an intuitive approach for the production of this definition, with a very important rate of failure: 76% of the questioned students have provided a false definition of a random variable.

The analysis of the productions of the students shows that this failure is essentially due to erroneous conceptual representations, which originate in various conceptual confusions:

- the random variable is assimilated with a random experiment,
- the random variable is confused with its image universe (the most common confusion),
- the random variable refers to an arbitrary variable,
- the random variable is confused with its law of probabilities.

We have besides also spotted a conceptual obstacle, where the student will represent the object "random variable" by referring quite simply to the nature of a random variable, namely “discrete” or “continuous”.

These conceptual difficulties and obstacles are essentially due to the formal teaching provided to these students. Nevertheless, the quantitative analyses that we performed, have allowed us to better emphasize this effect of teaching: the formal approach, correct or not, of the definition of a random variable has a particular effect on the production of relevant examples of random variables. However, if one is interested in the approach itself, formal or intuitive, adopted in the production of an example and its effect on the (correct or false) validity of the example provided, one notices that the answer to this question delimits the effect of the formal approach of the definition of a random variable on the (correct) availability of object “random variable”. Indeed, the quantitative analysis that we performed on the whole of the productions of examples provided by the students showed that:

- The relevance of a formal approach in the production of an example of random variable is independent of the validity (correct or false) of the produced example.
- The relevance of an intuitive approach in the production of an example of random variable has an effect on the validity (correct or false) of the produced example.

ISSN 2054-6297(Print), ISSN 2054-6300(Online)

These results show that an introductory teaching to the probabilistic formalism should not remain based on theoretical and formal presuppositions; quite the contrary, it should be accompanied by situations and examples which are based on the intuitive ground. Only such a hybrid teaching can lead the students to a better apprehension of probabilistic modeling, because it is through a dialectical formalism/intuition that the student can build a relevant probabilistic representation, especially around the notion of random variable.

REFERENCES

Benbachir, A. (2002) Reconnaissance et construction de contre-exemples en analyse mathématique des fonctions: étude de cas en première année d'université, Thèse de Doctorat ès-sciences, Département de Mathématiques, Faculté des Sciences Dhar El Mahraz, Université Sidi Mohammed Ben Abdellah, Fès, Maroc.

Castro, C. S. (1998) Teaching probability for conceptual change, Educational Studies in Mathematics, Vol. 35, Issue 3, pp. 233-254.

Commission Inter-IREM, Statistique et Probabilités, Sous la Coordiantion de Henry, M. (2001) Autour de la modélisation en probabilités : Notions d'expériences aléatoires. Vocabulaire et modèle probabiliste, pp. 161-171. Presses universitaires de Franche-Comté – Besançon, France.

Greer, B. (2001) Understanding probabilistic: the legacy of Efraim Fischbein, Educational Studies in Mathematics, Vol. 45, pp. 15-33.

Henry, M. (2009) Emergence de la probabilité et enseignement : définition classique, approche fréquentiste et modélisation, Repères IREM (75), pp. 76-89.

Lecoutre, M.-P. (1998) Evolution avec l'âge de « misconceptions » dans les intuitions probabilistes en France et en Israël, Recherches en Didactique des Mathématiques (RDM), 18(3), 311-332.

Meusnier, N. (2006) Sur l'histoire de l'enseignement des probabilités et statistique, Electronic Journ@l for History of probability and statistics, www.jehps.net, Vol.2, n°2, December 2006.

STEINBRING, H. (1991) The concept of chance in everyday teaching: aspects of social epistemology of mathematical knowlegde, Educational Studies Mathematics, Vol. 22, pp. 503-522.

Todhunter, I. (1865) A history of the mathematical theory of probability, from the time of Pascal to that Laplace, Cambridge and London, Macmillan.

Zaki, M. (1992) Evolution d'appréhensions probabilistes en situation de simulation, Thèse de doctorat, Université de Mons-Hainaut, Institut de Mathématique et d'informatique, Faculté des Sciences, Mons, Mons-Hainaut – Belgique.

Zaki, M. (2004) Acquis et applications de la didactique des mathématiques : quelques résultats méthodologiques et de recherches au niveau universitaire, Annales de Didactique et de Sciences Cognitives, Revue Internationale de Didactique des Mathématiques, IREM de Strasbourg,Vol. 9, pp. 61 – 66.

ISSN 2054-6297(Print), ISSN 2054-6300(Online)