13th International Congress on Mathematical Education
Hamburg, 24-31 July 2016

REPRESENTATIONS IN PROBABILITY PROBLEMS

Maria Nascimento[1,2], Eva Morais[1], J. Alexandre Martins[3]

[1]UTAD, Portugal, [2]CIDTFF/LabDCT-UTAD, Portugal, [3]UDI/IPG, IPG, Portugal

Probabilities and particularly conditional probabilities, total probability theorem, and Bayes' formula are a theme of higher education Probability and Statistics courses in Engineering degrees. The students' difficulties in learning these probability concepts were already discussed in previous researches. In this work, we present a first attempt to use a reference matrix to analyze students' representations in a probability problem. In what concerns the results to our first analysis, when considering only the students with correct or partially correct answers, in some representation, they revealed a better performance in the written language representation. Amongst students with a better overall performance in the correct and partially correct answers, the worst results were the answers in the symbolic representation. The tests analysis needs additional development including the wrong cases and the reference matrix of analysis needs further improvement so that in future students may overcome their difficulties in learning these concepts.

INTRODUCTION

Probabilities and specifically conditional probabilities, total probability theorem, and Bayes' formula, are a main part of higher education, HE, Probability and Statistics courses in engineering degrees. Nevertheless, students show many difficulties in studying these concepts. Amongst others, Díaz and Batanero (2009) refer the confusion between conditioning and causation, the fallacy of the transposed conditional, the fallacy of the time axis, the base rate fallacy, and the conjunction fallacy. Previously, Tversky and Kahneman (1982) had already discussed the conjunction fallacy and errors in applying Bayes' theorem and they state that understanding Bayes' formula requires cognitive effort and it is therefore not easy for students to interpret it accurately. Given this research topics, we have tried to view the difficulties of the students in another perspective, the students' uses of representations. Although different from some common engineering representations sketches, solid models, prototypes (e.g. Rachel, Shraddha & Summers, 2012) are some of the engineering representations commonly used by engineers and, some of those, in mathematics. In our view, not only mathematical representations can aid students' understanding of mathematics. In statistics, also teachers use several representations in order to explain the concepts and think they may aid students understanding concepts, but students can sometimes have difficulty in interpreting representations correctly. This question arises in parallel with the several difficulties briefly mentioned above. Goldin (2002, p. 208) wrote that

> "[i]n the most general sense, a representation is a configuration that can represent something else in some manner. For example, a word can represent a real-life object, a numeral can represent the cardinality of a set, or the same numeral can represent a position on a number line. (...) Written words, numerals, graphs, or algebraic equations are examples of external representations".

Goldin (2002) refers external and internal representations. The external representations have a physical presence, a sheet of paper, an electronic device, whether another support any (e.g., symbols representing numbers and their operations, the algebraic notation, the Cartesian graphs, several diagrams) and internal representations cannot be directly observable, as much they can be inferred through observable behaviors of the person or through their interaction with the external representations. The mathematical representations can be various (Bruner, 1999): active, iconic and symbolic. Active representations are those where although we have knowledge of many things, it is very difficult to teach them in words, diagrams, or images. The iconic representations (IcR) are those which relate to the look and use of images to explain the contents, e.g. Venn diagrams. Finally, the symbolic representations are those that make use of symbols, words and language. Our option was to analyze the texts written by the students' written language (WL) and the use of mathematical symbols as symbolic representations (SR). We studied the HE engineering students' representations in probability problems from a test within the continuous assessment, therefore nearby the class and study time devoted by students to learn. The aim of this study is to define a first approach to a reference matrix of analysis so that in future we may assess their use crossed with the students' difficulties in learning these probability concepts.

METHOD

We considered a sample of 45 engineering (Energy, E, Mechanics, M, and Civil, C, respectively 10 – 22%, 11 – 24%, and 24 – 54%) students from a public university in the northeastern Portugal in the school year of 2012/2013. Students were in the second year of their degree and we analyzed the answers given in a first written test of a Probability and Statistics course. Probabilities were the subject of the test, but we only studied the representations of the conditional probabilities problem that included total probability theorem and the Bayes' formula (presented in the poster of Morais, Nascimento & Martins, 2015). The problem presented to E and M students had the same text, but C students had a similar problem text. In the first question, students had to identify the events (1.1), the given probabilities (1.2), use symbolic notation to define a partition (1.3), and sketch a Venn diagram (1.4.1 for the given partition and 1.4.2 for the conditional event). In the second one, they had to compute the total probability, identifying the required probability (2.1), using a table or a tree diagram (2.2), and finally computing the probability (2.3). In the final question, they had to use the Bayes' formula, identifying the required probability (3.1), using the Bayes' formula (3.2), and finally computing the probabilities (3.3). Within the students' answers, we also analyzed the representations as written language (WL), symbolic language (SR), and iconic representation (IcR). In face of the answers of the students, we decided to consider the following categories of analysis, correct, partially correct, wrong, and no answer. The last two categories were only raised in the overall analysis of the students' answers by question. In Table 1 we present the grid of the analysis done to students' answers, where the gray cells indicate that the representation was not considered in the respective question.

R \ Q	1.1	1.2	1.3	1.4.1	1.4.2	2.1	2.2	2.3	3.1	3.2	3.3
WL											
SR											
IcR											

Table 1: Grid of the representations (R) analysis by question (Q). An example is presented in Morais, Nascimento and Martins (2015)

RESULTS AND DISCUSSION

In the overall analysis of the students' answers to the questions, we observed that the majority of the students (58%) identified correctly the events of the problem (1.1) and the given probabilities were identified correctly by 76% of the students (1.2). Only 9% of the students were able to identify correctly/partially correctly a partition (1.3). This was a surprising result, since it was a mandatory aspect in the students' resolutions of problems in classes and homework. The Venn diagram of the given partition (1.4.1) was correctly sketched by 47% of the students, and 40% were able to sketch the event involved in the Bayes' formula (1.4.2). In the question of the total probability, 51% of the students were able to identify correctly the required probability (2.1), and 47% used correctly a table or a tree diagram (2.2), but only 29% were able to compute correctly the required probability (2.3). The reasons for this result are the well-known difficulties in correctly computing conditional probabilities, and also confusing those with the probabilities of independent events (e.g. Díaz & Batanero, 2009). In the final question, they had to use the Bayes' formula, 51% identified correctly the required probability (3.1), 31% used/identified correctly the Bayes' formula (3.2), but only 18% computed correctly the probabilities (3.3). The same reasons referred above were also the origin of these results (e.g. Tversky and Kahneman, 1982).

In what concerns the analysis of the representations (Table 1) in the 1st question, the WL in all the items 1.1 and 1.3 were correctly/partially correctly answered by 49% of the students; in the SR (all items 1.1 to 1.4.2), 33% were correct/partially correct; in the items of IcR (1.4.1 and 1.4.2), 51% were correct/partially correct. These results are in line with the overall analysis to question 1, and confirm the difficulties that engineering students have in "reading problems" and "translating" them to the symbolic language. In the second question, the WL was correctly/partially correctly answered by 42% of the students (items 2.1 to 2.3); 36% correctly/partially correctly answered in the SR (items 2.1 to 2.3); in the IcR, 56% correctly/partially correctly answered (items 2.2 and 2.3). According to the results, the main difficulties of the students still are the use of SR. The same remark as above is valid for the final question, since 31% of the students were correct/partially correct in the WL and 27% were correct/partially correct in IcR (items 3.1 to 3.3).

Analyzing the relationship between the answers given in different representations for each question, the majority of students that answered correctly/partially correctly (C/PC) in one of the representations (R1), also answered correctly/partially correctly in a different representation (R2) for the same question. Those results are presented in Table 2. For example, in question 1 we had 12 students that answered correctly/partially correctly in the symbolic representation (R2) from the 22 students that answered correctly/partially correctly in the written language (R1) resulting in the 55% that appears in the 1st row and in the 2nd column. Also, note that we do not need to compare the representation WL in both R1 and R2 representations, and so on, thus the grey cells. The other grey cells refer to impossible comparisons as it may be inferred from Table 1.

	R1 \ R2	WL	SR	IcR
Question 1	WL		55%	68%
	SR	80%		65%
	IcR	65%	100%	
Question 2	WL		79%	100%
	SR	94%		100%
	IcR	76%	64%	
Question 3	WL		86%	

	SR	100%		

Table 2: Percentages of students that gave a C/PC answer in R2 among the students that gave a C/PC answer in R1

The results of the 1st column of Table 2 showed us that within the SR and IcR (when existing) not all the students presented a C/PC answer in the WL, except in the SR in question 3. This result is intriguing since we should expect students to master written abilities in order to solve the proposed problems. In the 2nd column within the WL and IcR not all the students presented a C/PC answer in the SR, except in the IcR in question 1. This results were not so surprising since the SR is usually more difficult to be C/PC used by the students. In the 3rd column within the WL and SR not all the students presented a C/PC answer in the IcR of question 1, but the 100% C/PC answers in question 2 were due to the students' support on the IcR, table or tree diagram. In this first analysis of the results we didn't present other cross tables since this was a first attempt to understand the difficulties of the students dealing with the different representations after the teaching of these probabilities concepts. Furthermore, even in the C/PC answers of these representations we found some of the difficulties already mentioned above and in the literature.

FINAL REMARKS

This first analysis of the results revealed that the students' worst global performance was in the written language representations in comparison with the symbolic and iconic representations. Even though, when considering only the students that answered C/PC in some representation, they revealed a better performance in the written language representation. Amongst students with a better overall performance (C/PC), the worst results were the answers in the symbolic representation. This work is still ongoing and much more has to be explored in the students' answers, namely presenting cases that may exemplify the use of the devised representations matrix, and also the analysis of the wrong ones in parallel and in comparison to the C/PC ones.

The reference matrix of analysis needs further improvement, namely to devise a way to take in to account the difficulties presented by the students in learning these concepts already discussed in previous researches (e.g. Díaz and Batanero, 2009; Tversky & Kahneman, 1982).

REFERENCES

Bruner, J. (1999). *Para uma teoria da educação* [Toward a Theory of Instruction]. Lisboa: Relógio d'Água.

Díaz, C., & Batanero, C. (2009). University students' knowledge and biases in conditional probability reasoning. *International Electronic Journal of Mathematics Education*, *4*(3), 131–162.

Goldin, G. (2002). Perspectives on representation in mathematical learning and problem solving. In L. English (Ed.), *Handbook of International Research in Mathematics Education* (pp. 178–203). New York, NY: Routledge.

Morais, E., Nascimento, M. & Martins, J. A. (2015). Representations in Probability Problems: Some examples . Submited Poster to ICME-13.

Rachel H., Shraddha, J. & Summers, J. D. (2012). A user study of interpretability of engineering design representations. *Journal of Engineering Design*, *23*(6), 443-468.

Tversky, A., & Kahneman, D. (1982). On the psychology of prediction. In D. Kahneman, P. Slovic y A. Tversky (Eds.), *Judgement under uncertainty: Heuristics and biases* (pp. 69–83). Cambridge, MA: Cambridge University Press.