

## BAYESIAN REASONING: CONNECTING ARITHMETIC, ALGEBRA AND TREE DIAGRAMS. A LONGITUDINAL RESEARCH

Cindy N. Morgado Hernández and Gabriel Yáñez Canal  
 Universidad Industrial de Santander, Colombia  
 cindy.morgado@correo.uis.edu.co

*This paper shows some of the results of the research conducted with the purpose of knowing the Bayesian reasoning of college students doing their first course in probability and statistics. For this there three tests that were designed and implemented at different times during the semester. The results showed three stages in the reasoning of the students: the first, which coincides with the first test, is characterized by the use of mathematical arguments proportions; the second, which coincides with the second test is characterized by the joint use of tree diagrams and algebraic expression of Bayes which had been taught by teachers; the third, which coincides with the third test, is characterized by the failed recall Bayes rule and incomplete use of the tree diagram attempt.*

### INTRODUCTION

Bayesian reasoning refers to the calculation of conditional probabilities inverse Bayes' theorem by:

$$P(B_i | A) = \frac{P(A | B_i)P(B_i)}{\sum_{j=1}^n P(A | B_j)P(B_j)}$$

The occurrence of event  $A$  is conditional on the realization of some of the events  $B_j$  which constitute a partition of the sample space, ie, a set of possible "causes" that may give rise to the event  $A$ . Considering that the event  $A$  happened and has occurred and inquire into the probability that  $B_i$  has occurred,  $P(B_i|A)$  is the posterior probability and  $P(B_i)$  is the prior probability; the values  $P(A|B_i)$  are the likelihood or probability of  $A$  occurred when it happened  $B_i$ .

Bayes theorem holds conceptual elements as a conditional probability, total probability theorem and product rule, which have been the subject of research in the field of statistical education. These studies have been directed to the difficulties presented by the students to face problems of conditional probability on the insights and ideas that form on the associated concepts (Diaz & de la Fuente, 2007) and on the representations used in solving problems. Recently the research interest has been directed to make the effect display formats of information about success in problem solving (Sedlmeier, 2002). In addition, some research performed by Corter & Zahner (2007), in the same direction Yáñez (2001) suggest that the schematic representations as tree diagrams should be taught for certain types of problems and stress that a certain representation is useful depending on the structural aspect of the problem, so that future research should explore and better document the success of such practices. It should be noted that these studies have been mainly of transversal court. In fact, in the literature reviewed any research on Bayesian reasoning over time was found. This was the reason that motivated us to conduct an investigation whose purpose is to see how well students confront problematic situations related to conditional probability and Bayes' theorem at different times of the semester, as well as identify insights, fallacies or were biases that remained and disappeared as a result of the instruction process.

### FRAMEWORK OF REFERENCE

A framework of reference is taken in the literature of conditional probability and, in particular, the Bayes theorem. One of the first studies was that conducted by Kahneman and Tversky, (1972) who showed that students mostly have no a priori probabilities to calculate the inverse probability. Also, it has been shown that students confuse the conditioning event with conditioned leading to the fallacy of the transposed conditional, i.e. confuse  $P(A|B)$  con la  $P(B|A)$  (Falk, 1986).

On the other hand, Gras and Totohasina (1995) highlight a temporary conception that students awarded him the conditional probability in the sense that the conditioning event always precedes B in time to event A. Now, the biggest problem at the moment of problem solving situations associated with the conditional probability is the difficulty of distinguishing between conditional and probabilities and joint probabilities (Pollatsek et. al., 1987), also in the correct identification of the conditioning event and in the correct partition of the sample space (Díaz de la Fuente, 2006).

Other studies have focused on the forms of representation of information, as performed by Gigerenzer and Hoffrage (1995) shows that Bayesian computations are simpler when the information is given in natural frequencies, instead of using probabilities or percentages. Later, Sedlmeier (2002) in addition to reaffirm what to Gigerenzer, suggests that problems must be presented in a way that invokes the pre-existing intuitions that is, they tend to be spontaneous for trial or to act in a certain way.

On the use of representations in problem solving Yáñez (2001) suggests that the coordination of several registers of representation are required to exploit the organizational and operational advantages they possess. However, he concludes that the only record that allows complete resolution of any problem of conditional probability is algebraic, the tables and trees have limited applications. Corter & Zahner (2007), investigating the use of visual representations (trees, lists of results, reorganization of information, charts and Venn diagram) finding that the difficulty in solving probability problems is to select the appropriate representation for the structure of the problem and that adequate representation is determined by the scheme underlying the problem.

## METHOD

A study was developed over time with college engineering students whom were doing their first course in probability and statistics to see how well these students confront problematic situations related to the Bayes theorem.

Three tests were designed to evaluate the students on Bayesian reasoning every five weeks. As Bayes rule includes conditional probability, marginal probability, product rule and total probability theorem, these tests were developed in a way to include items designed to evaluate each of these concepts. Each test consisted of three items with the same level of difficulty. The first item assessed the Bayesian reasoning in the context of channels. The second item assessed the domain of Bayes' theorem in the context of manufacturing machines. The last item is a problem in a context of sampling without replacement and consists of four subsections evaluating direct conditional probability, conditional probability and Bayes rule reverse, the product rule for dependent events and the theorem of total probability.

The sample consisted of 76 students in industrial engineering from the Universidad Industrial de Santander, Bucaramanga, Colombia, who were pursuing a course in statistics I in the fifth semester of their career in three different groups with three different teachers, the classes were masterful type: basically outlining the topics on the board, performed some examples and solved exercises form some guides texts (Montgomery & Runger, 2010 and Navidi, 2006). The first test was applied in the second week of the course; the second was applied one week after the teachers taught Bayes' theorem; the third test was applied in the penultimate week of class. Note that the teachers considered the results of these tests to the final grade.

## RESULTS

Of the 76 engineering students, 57 of them presented the three tests. Next, in Table 1 the success rates of each item is.

Table 1. Percentage of correct answers according to the content of the items for the three tests students

Context	Content	First Test	Second Test	Third Test
Canals	Bayes Theorem	0%	9%	7%

<b>Manufacture of articles of two machines</b>	Bayes Theorem	2%	75%	56%
	Conditional direct	61%	54%	53%
<b>Urn without replacement</b>	Conditional Reverse or Bayes rule	14%	2%	4%
	Product Rule	11%	40%	44%
	Total Probability Theorem	4%	11%	23%

Table 1 presents several facts that are worth highlighting:

- The enormous difficulty that students represented the problems of channels, the success rates were always below 10% in all three tests.
- The high success rates obtained in the second and third test on the item relating to the machines as opposed to the very low success rate of the first test. However, a decrease is observed success in the third test relative to the second.
- In problem urn, which can be considered free from the influences of contexts, results call attention were presented: the questions of conditional probability were best answered in the first test when students barely knew the classical way of calculating probabilities, those with high percentages of low probability and direct the reverse percentages probability. The item that asked by a joint probability not fail to increase your success rate over the three tests. Something similar occurred, although with lower percentages, with the item that asked by a marginal probability in the second extraction. Then we analyze the responses of students in each of the proposed contexts.

*Item in the context of channels.* This item is aimed at assessing the Bayes theorem. Most students in the first and third test responded  $\frac{1}{2}$  (Figure 1-a), i.e., assumed the conditional questions as a joint: pass through the canal I and exit R. In the second test, where the channel was more complicated because the teachers had resolved in his classroom the first problem, answers which assumed a total probability of presented  $\frac{1}{2} + \frac{1}{4} + \frac{1}{8}$  (Figure 1-b) assuming that the question inquired by the marginal probability that the ball leaves by R. In the third test some students made the tree diagram and nothing more; other students incorrectly raised the Bayes formula (Figure 1-c). Draws attention to the responses of some students who made the tree diagram without more respondent's  $\frac{1}{2}$  in the second and in the third test.

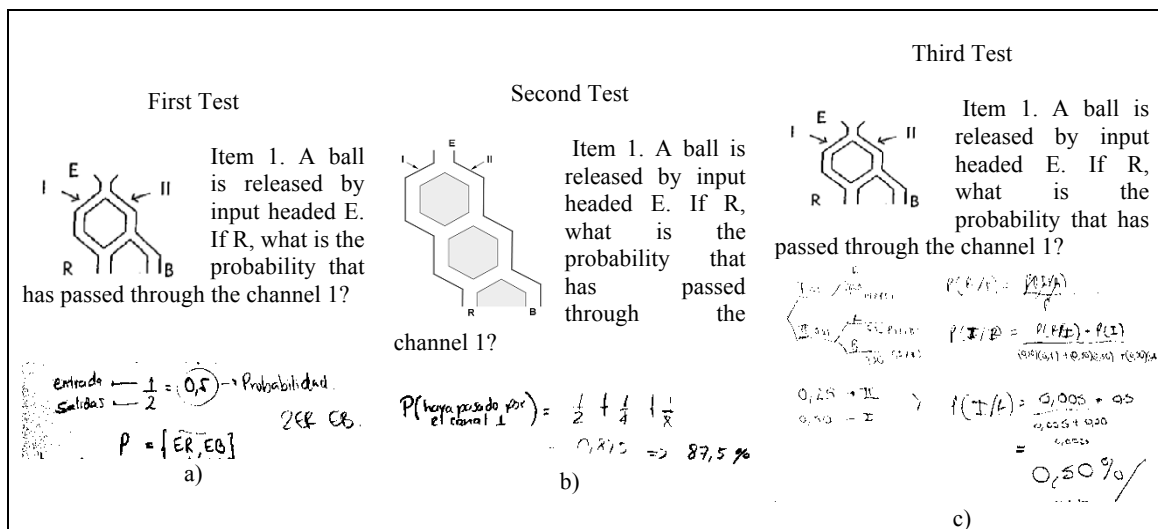


Figure 1. Responses of students to item 1 in the three tests

*Item in a context of manufacture of articles of two machines.* The relative ease with which students answered this item may perhaps be associated with full equivalence between the problem statement and the tree diagram as explained Yáñez (2001). A student is special correctly answered the item in the first test without using Bayes' formula simply did rule of three and played probability as a ratio between 5% to 40% and the sum of 5% to 40% more 1% 60% (Figure 2-a), this student in the second test misapplied Bayes's formula taking the wrong percentages and in the third test again responded incorrectly, showing that confrontation between an arithmetic intuition schematic and a new modeling that fails to grasp.

In the first test some students took a joint probability ( $P = 0.4 \times 0.05$ ), while in the second test was answered correctly by using Bayes' formula, but in the third test again answer a joint probability, which shows the difficulty of distinguishing between conditional probabilities and joint probabilities (Pollatsek et. al., 1987). In addition, some students made the tree diagram correctly but did not solve the problem (Figure 2-b).

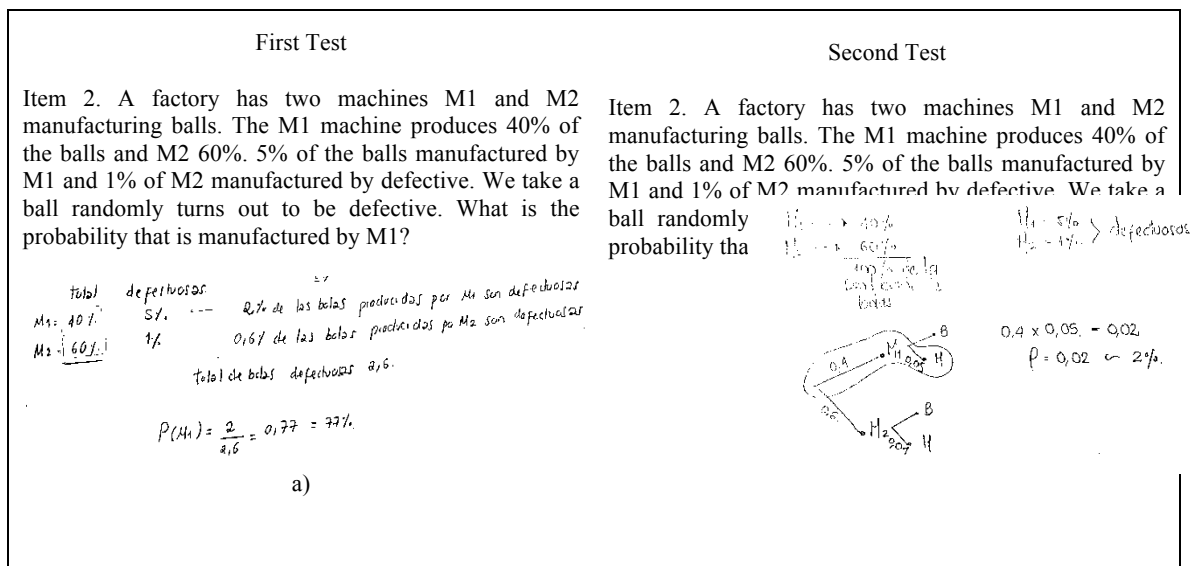


Figure 2. Responses of students to item 2 in the first and second test

*Item in a context urn without replacement.* With respect to subsection evaluating direct conditional, in the second and third test some students took as a joint probability (Figure 3-a). In subsection reverse conditional temporary concept of conditional probability (Gras & Totahasina, 1995) was evident when many students responded in all tests with  $\frac{1}{2}$  (Figure 3-b). The problem that investigates the probability of getting a white ball in the second extraction, typical application of the theorem of total probability, it is observed that students who do not respond correctly use the strategy of "dependent" as reported by Yáñez (2003) and reflecting the design approach of the result of isolated (Konold 1991).

Second Test	Third Test
<p>Item 3. An urn contains two white balls and two black balls. Blind draw two balls from the urn, one after, without returning the first to the urn. What is the probability of drawing a white ball in the second extraction if you know that a black ball was extracted in the first extraction?</p>	<p>Item 3. An urn contains two white balls, two black and two red balls. Blind draw two balls from the urn, one after, without returning the first to the urn. What is the probability of having drawn a red ball first, knowing that we extracted a black ball second?</p>
$P(A) = \left(\frac{2}{6}\right)\left(\frac{2}{5}\right) = 0,1333$ $\approx 13,3\%$ <p style="text-align: center;">a)</p>	$P(\text{negro. 1.ª extracción}) = \frac{2}{4} = 0,5 \approx 50\%$ <p style="text-align: center;">c)</p>

Figure 3. Responses of students to item 3 in the first and second test

## CONCLUSIONS

The results of this study highlight the difficulty when trying to modify primary intuitions by others using teaching methodologies focused primarily on the algebraic aspects of the models without even trying to bind these primary intuitions, Arithmetic character in the case of probability conditional probability with purely algebraic expressions court.

Also noteworthy is the enormous challenge represented for the students the problem of the channels that were hardly able to solve it. This problem where information is not explicit but is contained in the graph presented and arguably should reflect a clear Bayesian reasoning in the sense that additional information modifies an initial probability, showed clear signs of poor teaching process on these issues.

The research also found the difficulty in overcoming biases that are classics to find in all investigations carried out with the conditional probability: the confusion between conditional and joint and temporal concept of conditional probability.

These results reflect how innocuous it may be teaching these topics conditional probability when a strategy based on the algebraic representations and the tree diagram without any connection with arithmetic schema with students is assumed.

## REFERENCES

- Corter, J. E., & Zahner, D. C. (2007). Use of external visual representations in probability problem solving. *Statistics Education Research Journal*, 6(1), 22-50.
- Díaz, C., & de la Fuente, I. (2007a) Assessing psychology students' difficulties with conditional probability and bayesian reasoning. *International Electronic Journal of Mathematics Education*, 2 (2), 128-148.
- Díaz, C., & de la Fuente, I. (2006). Dificultades en la resolución de problemas que involucran el teorema de Bayes, un estudio exploratorio en estudiantes españoles de psicología. *Educación matemática*, 18, 75-94.
- Falk, R. (1986). Conditional probabilities: insights and difficulties. En R. Davidson y J. Swift (Eds.), *Proceedings of the Second International Conference on Teaching Statistics*. University of Victoria. 292 – 297.
- Gras, R., & Totohasina, A. (1995). Chronologie et causalité, conceptions sources d'obstacles épistémologiques à la notion de probabilité conditionnelle. *Recherches en Didactique des Mathématiques*. 15(1), 49 – 95.
- Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, 102, 684 – 704.
- Kahneman & Tversky (1972) Subjective probability: A judgment of representativeness. *Cognitive Psychology* 3, 430–454.
- Konold, C. (1991). Understanding Students' Beliefs about Probability. In E. Von Glasersfeld (Ed.), *Radical Constructivism in Mathematics Education*, Kluwer Academic Publishers, 139-156.

- Montgomery, D. C., & Runger, G.C. (2010). *Probabilidad y Estadística aplicadas a la Ingeniería*. Limusa Wiley.
- Navidi, W. (2006). *Estadística para Ingenieros y Científicos*. McGraw-Hill, Mexico 2006.
- Pollatsek, A., Well, A. D., Konold, C., & Hardiman, P. (1987). Understanding conditional probabilities. *Organization, Behavior and Human Decision Processes*, 40, 255– 269.
- Sedlmeier, P. (2002). Improving statistical reasoning by using the right representational format. *In Proceedings of the Sixth International Conference on Teaching Statistics*. Cape Town: International Statistical Institute (ISBN: 085590 782 7).
- Yáñez, G. (2001). El álgebra, las Tablas y los Árboles en Problemas de Probabilidad Condicional. *Iniciación a la investigación en didáctica de la matemática. Homenaje al profesor Mauricio Castro*. En Gómez, P., y Rico, L. Granada, España, 355-37.
- Yáñez, G. (2003). *Estudios sobre el papel de la simulación computacional en la comprensión de las secuencias aleatorias, la probabilidad y la probabilidad condicional*. Tesis doctoral no publicada. Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, México.