Revista Latinoamericana de Investigación en Matemática Educativa
ISSN: 1665-2436
relime@clame.org.mx
Comité Latinoamericano de Matemática Educativa
Organismo Internacional

LONJEDO VICENT, M. ÁNGELES; HUERTA PALAU, M. PEDRO; CARLES FARIÑA, MARTA

CONDITIONAL PROBABILITY PROBLEMS IN TEXTBOOKS AN EXAMPLE FROM SPAIN

Revista Latinoamericana de Investigación en Matemática Educativa, vol. 15, núm. 3, noviembre, 2012, pp. 319-337
Comité Latinoamericano de Matemática Educativa
Distrito Federal, Organismo Internacional

M. Ángeles Lonjedo Vicent, M. Pedro Huerta Palau, Marta Carles Fariña

CONDITIONAL PROBABILITY PROBLEMS IN TEXTBOOKS AN EXAMPLE FROM SPAIN

RESUMEN

En este artículo, que forma parte de los resultados de un estudio más amplio, presentamos un método para identificar, clasificar y analizar los problemas ternarios de probabilidad condicional en los libros de texto de matemáticas. Este método se basa tanto en la estructura de los problemas como en un enfoque realista de la enseñanza de las matemáticas basado en la resolución de problemas en contexto. Presentamos algunos resultados de este análisis realizado con libros de texto de matemáticas españoles.

PALABRAS CLAVE:

- *Probabilidad condicional*
- *Resolución de problemas*
- *Análisis de libros de texto*

ABSTRACT

In this paper we show part of the results of a larger study[1] that investigates conditional probability problem solving. In particular, we report on a structure-based method to identify, classify and analyse ternary problems of conditional probability in mathematics textbooks in schools. This method is also based on a realistic mathematics education approach to mathematics teaching through problem solving in context. We report some results of the analysis of mathematics textbooks from Spain.

KEY WORDS:

- *Conditional probability*
- *Solving problems*
- *Analysis of textbook*

RESUMO

Neste artigo, que é parte dos resultados de um estudo maior, apresentamos um método para identificar, classificar e analisar problemas de probabilidade condicional em livros didáticos de matemática. Este método baseia-se tanto na estrutura dos problemas quanto em uma abordagem realista para o ensino da matemática, com base na resolução de problemas em contexto. Apresentamos alguns resultados desta análise, realizada nos livros didáticos espanhóis de matemática.

PALAVRAS CHAVE:

- *Probabilidade condicional*
- *Resolução de problemas*
- *Análise de livros didáticos*

[1] Proyecto EDU2008-03140/EDU del Ministerio de Ciencia e Innovación, España.

Revista Latinoamericana de Investigación en Matemática Educativa (2012) *15* (3): 319-337.
Recepción: Junio 14, 2011 / Aceptación: Junio 11, 2012.

RÉSUMÉ

Cet article reproduit une partie des résultats d'une étude plus vaste consacrée à la résolution de problèmes de probabilité conditionnelle avec, en particulier, la description d'une méthode basée sur une structure visant à identifier, classer et analyser des problèmes ternaires relevant de la probabilité conditionnelle dans les manuels scolaires de mathématiques. Cette méthode est également fondée sur une approche réaliste de l'enseignement des mathématiques par le biais de résolutions de problèmes dans leur contexte. C'est pourquoi plusieurs résultats concernant l'analyse des manuels scolaires de mathématiques en Espagne sont aussi reproduits.

MOTS CLÉS:

- *Probabilité conditionnelle*
- *Résolution des problèmes*
- *Analyse de manuels scolaires*

1. Introduction

There are two main questions that guide our research. Why and how should we prepare our students in conditional probability at secondary school? The answer to the first takes into account students' future, both as students at university and as citizens. From this perspective, it is necessary to explore contexts and phenomena in which conditional probability is actually involved. If we think in this way, we will determine not only the kind of skills secondary schools students should have in relation to conditional probability, but also the type of problems they should be able to solve and in which contexts they will face such problems. The answer to the second question may be found in a phenomenological and realistic approach to teaching conditional probability through problem solving.

In schools, it is quite common to consider problem solving as a particular scenario in which to apply mathematical knowledge. Thus, problems are considered as examples in which mathematical knowledge can be applied in order to find an answer to the question asked. This point of view regarding the role of problem solving is very widespread in probabilistic education, on every school level. Firstly, formal teaching of the theory of probability and secondly solving application problems seems to be the most widespread teaching model. This mechanistic approach (Streefland, 1991, p. 16) was already criticized by Freudenthal (1974) in reference to the design of a course on probability.

However, another teaching model is possible if a school problem is seen as an instance of something more general that could be called a *situation* and the aim of teaching consists of modelling those situations in which conditional probability is involved. For example, the following problem (the disease problem):

> *A diagnostic test for uterine cancer has a false positive coefficient of 0.05 and a false negative coefficient of 0.01. A woman with a probability of 0.15 of having the disease prior to the test has a negative result in her test. Calculate the probability that she is not ill.*

This is a well-known problem that refers to a general situation, that of the Diagnostic test (Carles & Huerta 2007; Huerta, 2009), in a particular context, the Health context, in which it is formulated. Therefore, every school problem could also be considered as an instance of a particular situation and, in order to model phenomena that are present in it, certain probability theory is required. In this respect, our approach to teaching conditional probability is based on exploring realistic situations by means of problem solving in context. This approach fulfills the Realistic Mathematics Education (RME) approach since "in RME context problems play a role from the start onwards" (Gravemeijer & Doorman, 1999, p. 111). In order to do that, a prior phenomenological-based analysis of problems is required. Depending on this analysis, we can decide if a particular problem is a good instance for a particular context or if a particular family of problems enables students to develop skills in conditional probability in a particular situation. Which problem does and which one does not. And, also, which situation students should explore and which they should not, depending, among other things, on their prior mathematical skills and on how interesting the situation is for improving students' mathematical competence.

In this paper we are going to hypothesize that a great amount of the mathematical content that is actually taught in schools is carried out by means of school mathematical textbooks. If this is the case, we need to analyze those textbooks in order to obtain a global view of the reality of the teaching of the probability in schools.

In this paper, we will demonstrate a method of organizing and classifying conditional probability problems for further analysis. We will identify a particular family of problems in textbooks, which we call ternary problems of conditional probability, and we will illustrate their method of analysis. The results we obtain may allow teachers and researchers to decide which problems students should solve and which they should not, and whether textbooks provide students with opportunities for exploring meaningful situations that help them to improve their skills in conditional probability. Finally, we will show the results of research on textbooks from Spain.

2. Objectives

Conditional probability problems[2] in textbooks can be divided into two groups: problems that use Bayes' Formula for two basic events; and problems in which the generalized Bayes' Formula is required. Here, our concern relates to the former; however, we consider them as precursors of the latter in every teaching model of conditional probability considered. Therefore, we feel that any research question we formulate about them will be interesting for a wide sample of probabilistic education practitioners. In particular:

Two methodological questions we will try to answer:

- How can we classify school probability problems into families and subfamilies of ternary problems of conditional probability?
- How can we identify situations and contexts in which school problems are presented?

The answers to these two questions could give us an idea about the structural-mathematical complexity of problems related to contextual difficulties inherent in a situation or context.

As an example, here are two questions resulting from the experimental study of textbooks from Spain:

- What are the families of problems that textbooks provide to be solved?
- What are the situations and contexts in which problems in textbooks are presented?

3. Background

Given two events A and B, their complementary events, $\overline{A}$ and $\overline{B}$, and the corresponding intersection events, we can define a world of conditional probability problems generated by these events and their probabilities. Here is an example:

> *In a school, the probability of success in Philosophy is 0.6 and in Mathematics, 0.7. Choosing a student at random among those that passed Mathematics, the probability that he/she also passes Philosophy is 0.8. If Juan passed Philosophy, what is the probability that he also passed Mathematics?*

[2] They are probability problems in which at least one conditional probability is involved, either known or unknown.

We define conditional probability problems as ternary problems if they belong to the above world of problems and verify the following conditions: (1) One conditional probability is involved, either with known or unknown data or both; (2) Three probabilities are known; (3) All probabilities, both known and unknown, are connected by ternary relationships[3].

Our example is a ternary problem because:

- The conditional probabilities (condition 1) involved are:
  - known data: *Choosing a student at random among those that passed Mathematics, the probability that he/she also passes Philosophy is 0.8*. $p(P \mid M)$
  - unknown data: *If Juan passed Philosophy, what is the probability that he also passed Mathematics?* $p(M \mid P) = 0.8$.
- The three known probabilities (condition 2): *The probability of success in Philosophy is* $p(P) = 0.6$ *and in Mathematics,* $p(M) = 0.7$, *and the conditional probability in the previous section* $p(P|M) = 0.8$
- All probabilities, both known and unknown, are connected by ternary relationships:
  - $p(M \cap P) = p(P \mid M) \,.\, p(M)$
  - $p(M \cap P) = p(M \mid P) \,.\, p(P)$
  - $p(M \cap P) + p(\overline{M} \cap P) = p(P)$

We also distinguish between school problems and ternary (school) problems. Each time we have three known data and one (asked) unknown data, we say we have a ternary (school) problem. Hence, every school problem may be formulated by means of one or more ternary problems.

Since, in ternary problems, all data is related, in order to classify problems into families and subfamilies we must define the following variables (Lonjedo, 2007; Huerta, 2009):

- The level (L) of a problem is the number of known conditional probabilities in the text of the problem. There are four levels, corresponding to 0, 1, 2, and 3 known conditional probabilities. This variable characterizes the family to which a problem belongs.

3 That is, by means of relationships like these: $p(A) + p(\overline{A}) = 1$; $p(A \cap B) + p(A \cap \overline{B}) = p(A)$ or $p(A \mid B) \times p(B) = p(A \cap B)$, $p(B) \neq 0$.

- In relation to each level, we define Category (C) as the number of known absolute probabilities in the text of the problem. Depending on each level, the category could be 0, 1 or 2. Together with Level L, the category divides the four families into 12 subfamilies of problems.
- Finally, we define Type (T) of a problem as the unknown data in the problem. There are three possible types: T_1-conditional probability, T_2-absolute probability and T_3-intersection probability.

Therefore, every ternary problem of conditional probability belongs to an L - family of problems and to a subfamily described by means of a vector like this (L_1, C_1, T_3). In other words, if a problem belongs to a (L_1, C_1,T_3) - subfamily, this means that, in the formulation of the problem, there are three known quantities which, when read in a probabilistic sense, lead to one conditional probability (L_1), one absolute probability (C_1), and, consequently, one intersection probability, and one unknown quantity, namely the intersection probability (T_3). The example we provide above is a problem belonging to the L_1,C_2,T_1 - subfamily.

Therefore, we have four families of conditional probability problems and, taking into account the category and type, they are categorized into twenty sub-families, as follows (table i[4]):

TABLE I
Classification of ternary problems of conditional probability into families and subfamilies.

	L_0			L_1			L_2			L_3		
C_0	C_0T_1	∅	∅	C_0T_1	C_0T_2	C_0T_3	C_0T_1	C_0T_2	C_0T_3	C_0T_1	C_0T_2	C_0T_3
C_1	C_1T_1	∅	∅	C_1T_1	C_1T_2	C_1T_3	C_1T_1	C_1T_2	C_1T_3	∅	∅	∅
C_2	C_2T_1	∅	∅	C_2T_1	∅	C_2T_3	∅	∅	∅	∅	∅	∅

In order to carry out analytical readings of problems, in other words the reading of problems that only pay attention to data (either events or probabilities) and relationships between data, Cerdán and Huerta, (2007) introduce the trinomial graphs (see also Huerta, 2009). In Figure 1 (to the left), we show the trinomial graph of the world of ternary problems of conditional probability in a particular context, the mathematical context, in which events and probabilities are represented in a symbolic format. In Figure 1 (to the right), we also show a similar graph but, this

4 The symbol ∅ means that there are no ternary problems belonging to that sub-family.

time, in a different context, the Diagnostic test context in Health, in which events and probabilities are represented in a format that is the corresponding format in this context.

In the graphs, nodes or vertexes show the probabilities of the "events" they are representing; dotted lines represent ternary additive relationships between probabilities while solid lines show multiplicative relationships. With the help of an algorithm for moving throughout the graph[5], we can determine the minimal graph of a problem, representing one of the possible resolutions of the problem that involves the minimum number of intermediate data and relationships required to find the answer to the problem, and therefore its structural complexity.

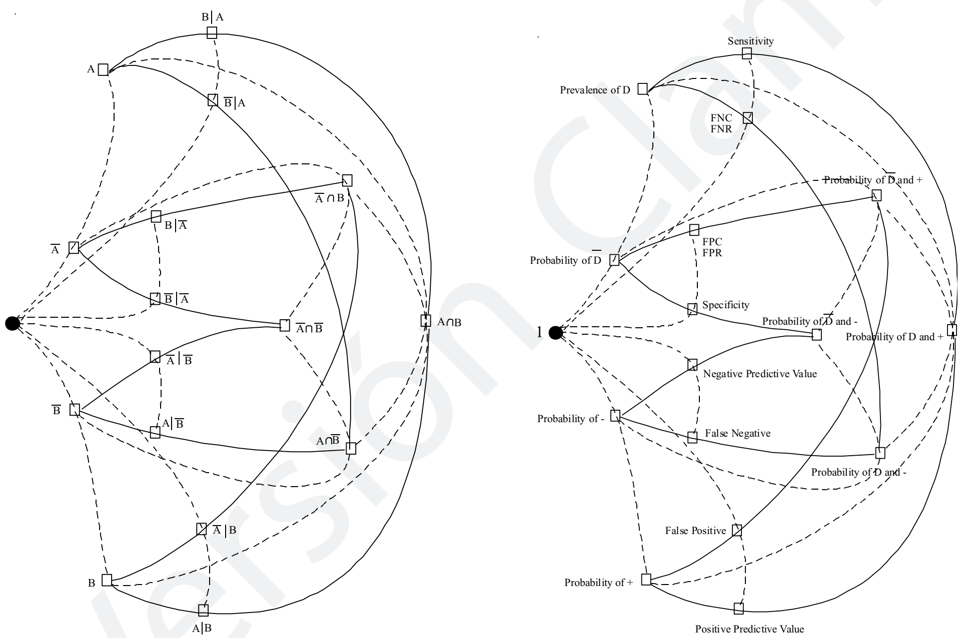

Figure 1. Trinomial Graph of ternary problems of conditional probability in two different contexts.

On the other hand, in order to identify situations and contexts instanced by ternary problems of conditional probability, Carles and Huerta (2007) suggest an analysis of didactical phenomenology (Freudenthal, 1983) of those problems based on the following criteria:

5 We call it the destruction algorithm of the graph (Huerta, 2009).

- Context (in which the problem is formulated),
- Phenomena referring to events (i.e. organized by means of events),
- Phenomena referring to probabilities (i.e. organized by means of probabilities),
- Specific terminology, Classification (in accordance with the structure of the data in the text of the problem and the presentation format of the data) and
- Specific teaching environment or reference.

Carles and Huerta (2007) describe the *Test of diagnostic* situation in the particular context of Health. This situation is represented by means of a trinomial graph that represents all possible relationships between known and unknown data in that situation (Figure 1 to the right).

Since they are relevant criteria for the analysis of problems formulated in every context, it is necessary to recall some of the aforementioned criteria as follows:

Context: A particular situation in which problems are put forward. In a context, a particular concept such as conditional probability has a specific meaning or is used with a specific sense. For example, a Diagnostic Test is a context. In our work, the Diagnostic Test Context is not only designed for problems like those in the annex but for any situation in which something has to be tested in order to determine whether it meets quality or health control parameters. Generally, tests are not completely reliable. Therefore tests results have to be assumed with certain risk, with probability.

Phenomena (referring to events): In a particular context, those statements that can be recognized as having an uncertain possible outcome are phenomena. These statements can be organized by means of reference sets (Freudenthal, 1983, p.41), events in a probabilistic and mathematical sense and operations between events. For example, "being ill", "being ill and having a negative diagnostic"...

Neither of these phenomena will be recognized as a "conditional event" even though it is possible to talk about them as if they were. For example, "knowing that he / she has a negative diagnostic, being ill" is sometimes said but it cannot be organized by means of a reference set.

Phenomena (referring to probabilities): In a particular context, apart from quantities, we refer to signs, words, sets of words and statements that express a measurement or the need for a measurement regarding the uncertainty of a phenomenon. For example, "sensitivity" is a term that refers to probabilities.

By means of the sensitivity of a test, usually in percentages or probabilities, we express the probability that the result of a test be positive if the patient in fact has the disease.

Prevalence of a disease is another example of phenomena we are referring to. The encyclopaedic meaning of the word "prevalence" is not related to probabilities. However, in the particular context it acquires a probabilistic meaning. Therefore, the phenomenon of the prevalence of a disease can be described by a probability and is usually expressed by percentages or a number between 0 and 1. In some problems, this number might be unknown.

4. Methodology

Our data comes from conditional probability problems in textbooks. We analyse them by means of two types of related analysis. First, we classify problems into families and subfamilies, paying attention to data structure; and, second, we analyse these problems paying attention this time to the situations and contexts in which they appear to be formulated. Thus, we methodologically proceed as follows:

1. We take into account almost all editorials of mathematics textbooks, including those more popular among high schools (see Table ii), which include conditional probability problems in their lessons. This fact makes us quite sure that a majority of secondary school students (12-18 years old) use at least one of the analysed textbooks.
2. We divide textbooks into two classes: textbooks for compulsory secondary school (for 12-16 year old students) and textbooks for post-compulsory secondary school (16-18 year olds).

TABLE II
Frequency and percentage of editorials and textbooks observed

School level	Number of editorials	Number of textbooks
12-16 years old	7 (38.9%)	8 (25%)
16-18 years old	11 (61.1%)	24 (75%)

3. We identify conditional probability problems in textbooks. These problems are classified into two groups: ternary problems of conditional probability and non-ternary problems of conditional probability. We select ternary problems for the analysis.

4. We read every ternary problem in a probabilistic sense. In other words, every problem is translated into symbolic language of the conditional probability and consequently every known and unknown quantity in a problem is read as a probability. We translate the problem into the graph in Figure 1 (to the left) and determine the minimal graph of the problem or simply the graph of the problem. On the basis of this graph, we determine the complexity of the problem.

5. By means of a three-component vector (L_i, C_j, T_k), $i = 0, 1, 2, 3$; $j = 0, 1, 2$; $k = 1, 2, 3$, every problem is codified. This coding allows us to include them in one of the families and sub-families described in Table 1. In order to do so, some decisions have to be previously taken into account. For example, as we already said, given a set of known data there is a problem each time an unknown value is asked for. Therefore, school problems formulated with more than one question could be classified into more than one family and/or sub-family.

6. A situation could be inferred or suggested from a set of problems. A context could be inferred from problems within a situation. For example, the following problems refer to the same (mathematical) situation but in different contexts we call symbolic and non-symbolic context:

   - *Given $p(A) = 0.3$; $p(B) = 0.2$ and $p(A/B)= 0.5$, calculate $p(B/A)$*
   - *It is known that every day the probability of reading paper A is 0.3 and paper B is 0.2. If it is also known that the probability of reading paper A knowing that paper B has been read is 0.5, calculate the probability of reading paper B knowing that paper A has been read.*

   Problems are classified according to the three-component vector and the situation and context in which they are formulated.

7. Every problem is finally analyzed from the point of view of the didactical phenomenology.

5. Some results

In the 32 Spanish mathematics textbooks we finally considered, we identify and analyse 106 ternary problems of conditional probability, 19 of which (17.9%) are problems in compulsory secondary school textbooks. Very few textbooks at the (12-16) level include ternary problems of conditional probability. Almost all problems are, however, in textbooks at the (16-18) level. The next table (Table iii) shows the distribution of these problems according to school level and the level (L) of problems.

There is something common to both school levels. Problems with three conditional probabilities (L_3) are not formulated in any textbook and the presence of problems with only one conditional probability (L_1) is not relevant at any school level. Therefore, almost all problems (84%) are formulated either without conditional probabilities as known data or with two conditional probabilities.

TABLE III
Distribution of frequencies and percentages of problems in accordance with the school level and problems' level

	Level of problems				
School level of the textbooks	L_0	L_1	L_2	L_3	Total}
Compulsory Secondary School (12 – 16 year olds)	17 89.5% *40.5%*	0	2 11.5% *4.2%*	0	19 17.9%
Post-compulsory Secondary School (16 – 18 year olds)	25 28.7% *59.5%*	17 19.5% *100%*	45 51.7% *95.7%*	0	87 82.1%
Total	42	17	47	0	106

Problems without conditional probabilities as known data (L_0) are mostly formulated with three or four intersection probabilities at the 12-16 year old school level (in other words, without known absolute probabilities, C_0) but with two absolute probabilities (C_2) at the 16-18 school level. In both cases, the unknown data is always a conditional probability (T_1) (see Tables iv and v). This is the general property of those problems that are formulated including contingency tables as a data presentation.

TABLE IV
Distribution of the percentages of ternary problems of conditional probability in compulsory secondary school textbooks, according to families and subfamilies.

	L_0= 17			**L_1= 0**			**L_2= 2**			**L_3= 0**		
C_0	C_0T_1 70.6%	∅	∅	0	0	0	0	0	0	0	0	0
C_1	C_1T_1 11.8%	∅	∅	0	0	0	C_1T_1 100%	0	0	∅	∅	∅
C_2	C_2T_1 17.6%	∅	∅	0	∅	0	∅	∅	∅	∅	∅	∅

The fact that there are no problems formulated with two conditional probabilities (L_2) and without absolute probabilities (C_0) as known data is also common in textbooks. They are always formulated with one known absolute probability and mostly ask for another absolute probability (T_2) or one conditional probability (T_1) (Table v). The presence of problems at the L_3-level is not relevant at the 12-18 year old level.

TABLE V
Percentages of ternary problems of conditional probability in post-compulsory secondary school textbooks in families and subfamilies.

	L_0= 25			**L_1= 17**			**L_2= 45**			**L_3= 0**		
C_0	C_0T_1 8%	∅	∅	0	0	0	0	0	0	0	0	0
C_1	0	∅	∅	C_1T_1 11.8%	C_1T_2 11.8%	C_1T_3 5.9%	C_1T_1 57.8%	C_1T_2 31.1%	C_1T_3 11.1%	∅	∅	∅
C_2	C_2T_1 92%	∅	∅	C_2T_1 58.8%	∅	C_2T_3 11.8%	∅	∅	∅	∅	∅	∅

Finally, looking at table iv and v, we can also observe that problems with one known conditional probability (L_1) are only formulated for the 16-18 year old level. They have a relative presence in these textbooks (16%). The majority of them are formulated by adding two absolute probabilities (C_2) to the conditional probability data and asking for one new conditional probability, T_1. It is also common for there to be no L_iC_0-problems, $i \neq 0$, in any of the textbooks we considered.

Therefore, what is theoretically possible at school (Table i) is not reflected in textbooks in practice. In fact, these results clearly show the conditional probability teaching approach in textbooks. It is mechanistic because students at the 12-16 level exclusively solve problems from the L_0-family, problems in which one conditional probability is asked for, with absolute and intersection probabilities being the known quantities. With 70.6 % of problems (Table iv), the $L_0C_0T_1$ sub-family is the most popular among textbooks. All three known quantities are intersection probabilities, easily presentable in 2x2 contingency tables. The formula for conditional probability solves the question. However, no questions about Bayes' theorem are possible.

In order to introduce Bayes' rule and theorem, problems from L_1 or/and L_2 families are required. This time conditional probabilities act as known data in problems. If problems are also T_1, as is usual in the 16-18 age group (Table v), then they are formulated in textbooks with the intention of providing examples of Bayes' rule, applying the corresponding formula.

The results we mention above only deal with data structure. However, as we know, problems are usually formulated in non-mathematical contexts (83% of observed problems). In Table vi we describe situations and contexts that can be inferred or suggested from problems in textbooks and the frequency in every situation and context.

Almost all problems (73.6%) describe what we call statistical situations and are formulated in several contexts, such as a situation in which known data is taken from the result of surveys or enquiries comparing characteristics of a sample of individuals or objects. For example, to be male or female and to smoke or not to smoke. We can find these problems formulated at all L-levels, but basically they are L_0 and L_2 levels (63 out of 78 problems in Table vii).

TABLE VI
Situations and contexts identified in secondary school textbooks; frequency of the problems.

Situation	Context	12-16 year olds Number of problems	16-18 year olds Number of problems	Total
Statistical Surveys (opinion or others)	Urban Business Anthropometric Social	13	65	78
Random choice Random draw	Urns	0	1	1
Mathematical	Symbolic	4	5	9
	Non-Symbolic	0	9	9
Diagnostic Test	Health	0	2	2
	Quality Control	2	0	2
	Surveys/ Opinion	0	1	1
	Law	0	4	4

The fact that most of these problems have been classified into the L_1C_j - sub-families, $j = 0, 1, 2$, is probably due to data format in problems. This situation in different contexts needs to be reasonably described by means of data in percentages and/or absolute frequencies organized in contingency tables. Conditional probabilities in situations like this are expressed using conditional sentences and expressions similar to this: If a person chosen at random is female, what is the probability that she smokes, which expresses a hypothesis and its consequence in terms of probability. This, of course, is a possible use of conditional probability.

There are other situations that problems are not exploring. In our opinion, one of the most important situations that needs to be explored at secondary school is the Diagnostic test. This situation is responsible for generating many realistic problems in various contexts: Health, quality control, forensic biology, and so on. However, textbooks as a whole only devote 8.5% of problems to exploring this situation. Therefore, updating known probabilities by means of Bayes' Theorem is not explored at all. As a result, more context problems in situations like this are required, exploring phenomena related to cause-effect relationships between two basic events: to be ill and to be tested, for example.

Something similar occurs with mathematical situations, either in symbolic or non-symbolic contexts. Only 17% of problems explore mathematical situations. This fact probably confirms our hypothesis that problems in textbooks are problems for applying learned mathematical content and not as a means of organization of the phenomena related to events and probabilities (in the sense of vertical mathematization after horizontal mathematization, Treffers, 1987).

TABLE VII
Distribution of the observed problems according to situations and contexts and the family and subfamily they belong to.

Situation	**Context**	**L_0=42**			**L_1=19**						**L_2=47**		
		C_0T_1 14	**C_1T_1 2**	**C_2T_1 26**	**C_0T_2 0**	**C_1T_1 2**	**C_1T_2 2**	**C_1T_3 1**	**C_2T_1 10**	**C_2T_3 2**	**C_1T_1 28**	**C_1T_2 14**	**C_1T_3 5**
Statistical Surveys (opinion or others)	Urban, Business, Anthropometric, Social, Health,... 78	13	2	19	0	2	2	1	9	1	12	12	5
Random choice	Urns, Coins, Lotteries, Cards,... 1	0	0	0	0	0	0	0	0	0	1	0	0
Mathematical	Symbolic 9	1	0	4	0	0	0	0	1	1	2	0	0
	Non-Symbolic 9	0	0	3	0	0	0	0	0	0	5	1	0
Diagnostic Test	Health 2	0	0	0	0	0	0	0	0	0	2	0	0
	Quality Control 2	0	0	0	0	0	0	0	0	0	2	0	0
	Surveys /Enquires 1	0	0	0	0	0	0	0	0	0	0	1	0
	Law 4	0	0	0	0	0	0	0	0	0	4	0	0

Problems in context could be analyzed by means of the phenomenological analysis described before. In Table vii, we refer to the phenomenological analysis we presented at other conferences (Carles & Huerta, 2007; Huerta, 2008). It is pertinent to cite it here because it allows us to conclude the study of problems in textbooks. Problems in Table viii could be others, but what we are showing here is an example of the result of the analysis of four problems in a Diagnostic Test situation, formulated in a Health context.

TABLE VIII

Aspects of the phenomenological analysis of the problems 1 to 4 showed in the Annex.

Problem	**Situation**	**Context**	**Classification and data format**	**Phenomena referring to**		**Specific terms**
				events	**conditional probabilities**	
P1			$L_2C_1T_1$ family - (1,0,2) Known data - p (D|+) Unknown data - Format: Probabilities Rates Percentages	- Be tubercular - Not be tubercular - To give a result	- If a person is tubercular, the test yields a positive result with a high probability (in % format) - If a person is not tubercular, the test yields a positive result with a small probability (in % format) - If the test is positive, there is a probability (<1) that the person actually is tubercular	- Positive result in tests - Diagnostic test
P2			L_2C_1 family - (1,0,2) - a) p (D|+): T_1 - b) p(~D|+): T_1 - Percentages	- Be diabetic - A person does not suffer from diabetes but is positive on test - A person suffers from diabetes but is negative on test	- FPC or FPR (False Positive Coefficient or Rate) - FNC or FNR (False Negative Coefficient or Rate)	- FPC or FPR - FNC or FNR - Prevalence of diabetes - Test is positive - Test is negative - Diagnostic test

P3			$L_2C_1T_1$ family - (1,0,2) - p(~D|-) - Probabilities	- Not suffer from uterine cancer - Suffer from uterine cancer - Positive result in diagnostics without uterine cance - Negative result in diagnostics with uterine cance - A person tested negative who does not suffer from uterine cancer	- False positive coefficient or rate (FPC or FPR) - False negative coefficient or rate (FNC or FNR)	- False positive coefficient or rate - False negative coefficient or rate - Pre-test probability - Negative result in test - Diagnostic test
P4	Diagnostic test	Health context	$L_3C_0T_2$ family - (0,0,2) - p (D) - Probability	- Be infected by tuberculosis - Not be infected by tuberculosis	- Sensitivity - Specificity - False Positive	- Sensitivity - Specificity - False Positive - Prevalence of the disease - Tuberculin test

6. Conclusion

In general, it does not seem that conditional probability problems in Spanish mathematics textbooks are formulated in order to explore as many situations as possible in a wide range of contexts. It can be said that only one, among those possible, is being explored, which we call a statistical situation. In this situation, the meaning associated with the conditional probability in every context relates to the hypothesis formulated in the problem with conditional clauses. The uncertainty created by the reliability of tests is not present in these problems, as is the case in the Diagnostic test situation. Consequently, students will only be competent with one meaning of conditional probability.

As a rule, we do not know how many problems students need to solve in order to be competent in a situation and its contexts and neither of these problems need to belong to every family and sub-family we theoretically identify. A situation with just a few problems will probably be enough, depending on to what extent the situation is meaningful to students and on the different contexts in which problems could be formulated. In other situations, however, a great amount of problems will be required. Therefore, teachers need to know which situations and contexts are meaningful to students, which problems formulated in these situations and contexts are interesting for students to solve, which data structure they have to be formulated with, and so on. The answer to these questions needs to be found in relation to the following question: what skills should students acquire in conditional probabilities by the end of secondary school, either as citizens or students at college?

References

Carles, M & Huerta, M.P. (2007) Conditional probability problems and contexts. The diagnostic test context. In D. Pitta-Panzati & G. Philippou (Eds.), *Proceedings of the 5th Congress of the European Society for Research in Mathematics Education* (pp. 702-710). Larnaca: Cyprus.

Cerdán, F. y Huerta, M.P. (2007). Problemas ternarios de probabilidad condicional y grafos trinomiales. *Educación Matemática 19*(1), 27-61.

Freudenthal, H. (1974). The crux of course design in probability. *Educational Studies in Mathematics 5*(1), 261-277. DOI: 10.1007/BF00684702.

Freudenthal, H. (1983). *Didactical Phenomenology of Mathematical Structures.* The Netherlands, Dordrecht: Kluwer Academic Press.

Gravemeijer, K. & Doorman, M. (1999). Context problems in Realistic Mathematics Education: A calculus course as an example. *Educational Studies in Mathematics 39*(1-3), 111-129. DOI: 10.1023/A:1003749919816

Huerta, M. P. (2008). On Conditional Probability Problem Solving Research – Structures and Contexts. Available at *http://tsg.icme11.org/tsg/show/14*

Huerta, M. P. (2009). On Conditional Probability Problem Solving Research – Structures and Context. *International Electronic Journal of Mathematics Education 4*(3), 163-194.

Lonjedo, M. A. (2007), *Análisis de los problemas ternarios de probabilidad condicional de enunciado verbal y de sus procesos de resolución*. Doctoral Dissertation, Universitat de València. Servei de publicacions de la Universitat de València.

Streefland, L. (1991). *Fractions in Realistic Mathematics Education: A Paradigm of Developmental Research.* The Netherlands, Dordrecht: Kluwer Academic Press.

Treffers, A. (1987). *Three dimensions: A model of Goal and Theory Description in Mathematics Instruction – The Wiskobas Project.* The Netherlands, Dordrecht: Kluwer Academic Press.

Annex

Q1. It is known that in a certain city one out of every 100 citizens is a tubercular person. A test was administered to a citizen. When a person is tubercular the test gives a positive result in 97% of cases. When he/she is not tubercular, only 0.01% of the cases give positive results. If the test is positive for these people, what is the probability that he/she is tubercular?

Q2. A diagnostic test for diabetes has an FPC of 4% and an FNC of 5%. If the prevalence of diabetes in a town is 7%, what is the probability that a person is diabetic if his/her test was positive? What is the probability that a person is not diabetic if his/her test was negative?

Q3. A diagnostic test for uterine cancer has a false positive coefficient of 0.05 and false negative of 0.01. A woman with a pre-test probability of 0.15 of having the disease has a negative result in her test. Calculate the probability that she is not ill?

Q4. The tuberculin test can test whether a person is infected by tuberculosis or not. The sensitivity and specificity of the test is very high, 0.97 and 0.98 respectively. If in a certain town there is a very high proportion of false positives, exactly 0.9, calculate the prevalence of the disease.

Autores:

M. Ángeles Lonjedo Vicent. Intituto de Educación Secundaria 26 Misericordia, València.
tomas_lonjedo@telefonica.net

M. Pedro Huerta Palau. Universitat de València. Manuel.P.Huerta@uv.es

Marta Carles Fariña. Intituto de Educación Secundaria Gata de Gorgos, Alicante.
marcarfa220@hotmail.com