On Conditional Probability Problem Solving Research – Structures and Contexts

Huerta, M. Pedro
Universitat de València
Valencia, Spain
Manuel.P.Huerta@uv.es

Summary

In this paper we summarize our research on solving problems of conditional probability we have recently carried out. We investigate a particular world of school word problems we call ternary problems of conditional probability. We classify them into four families and twenty subfamilies depending on data structure and investigated students' behaviour in solving problems, those with one conditional probability as known data. We identify four types of thinking processes related to data format and sense of use of data. With the help of a mathematical object, the trinomial graphs, and the analysis and synthesis method, we provide a didactical and phenomenological analysis of ternary problems of conditional probability in a particular context by identifying situations and contexts of use. Here we will illustrate our approach by the diagnostic test situation and the particular context of health.

The main purpose of our work is to improve secondary school students' understanding of conditional probability and their probability literacy by proposing a teaching approach based on problem solving in contexts.

Introduction

If we think of preparing our students in conditional probability in secondary school, at least two questions arise: 'why?' and 'how?' 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 actually is involved. If we do think in this way, we will determine not only what kind of competences secondary schools students should have connected to conditional probability, but also what 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.

Of course, problem solving, particularly in connection to probability and conditional probability is a topic that is usually taught around the world with greater or lesser degree of success. Some time ago, Shaughnessy (1992) pointed out the difficulties of teaching probability by relating it to teaching problem-solving because teaching probability and statistics *is* teaching problem solving, he said. On the other hand, in our country (BOE, 2006; BOE, 2007) and also in other countries (NCTM, 2000), curricular standards suggest that in general mathematics, and also probability and consequently probability problem solving, should be taught in context, including mathematical context, connecting school mathematics to experimental reality of students. In recent works some new notions are considered in probability education. These notions are Probability in Context (Watson, 2005; Gal, 2005; Carles & Huerta, 2007) and Probability Literacy (Gal, 2005). Both are also related to a new notion of chance as precursor to probability (Watson 2006).

In other pieces of work (Carles & Huerta, 2007; Cerdán & Huerta, 2007), the subject of our research, i.e., conditional probability problems and ternary problems' of conditional probability, has already been defined.

A mathematical reading of these problems allows us to classify them by means of a three-component vector (x, y, z) which represents the known data in the text of the problems: *x* represents the number of absolute probabilities, *y* the number of intersection probabilities, and *z* the number of conditional probabilities with $x + y + z = 3$. By choosing *x*, *y*, and *z* in a suitable way we are theoretically able to identify 9 types of conditional probability problems. Lonjedo (2007) showed that there are some types of problems that are not included in secondary school textbooks. Cerdán & Huerta (2007) use trinomial graphs in order to study ternary problems of conditional probability by means of the analysis and synthesis method. These problems have been modelled using a trinomial graph as shown in Figure 1, where dashed lines represent ternary additive relationships, whereas solid lines represent multiplicative relationships, and nodes are labelled by the probabilities of the related "events". In the graph, dark nodes represent known data and white nodes signify unknown data. Solving a problem in the graph consists in transforming white nodes into dark nodes by means of an algorithm. This algorithm, that we call destruction algorithm of the graph, allows users familiar to it to find the solution of a posed problem.

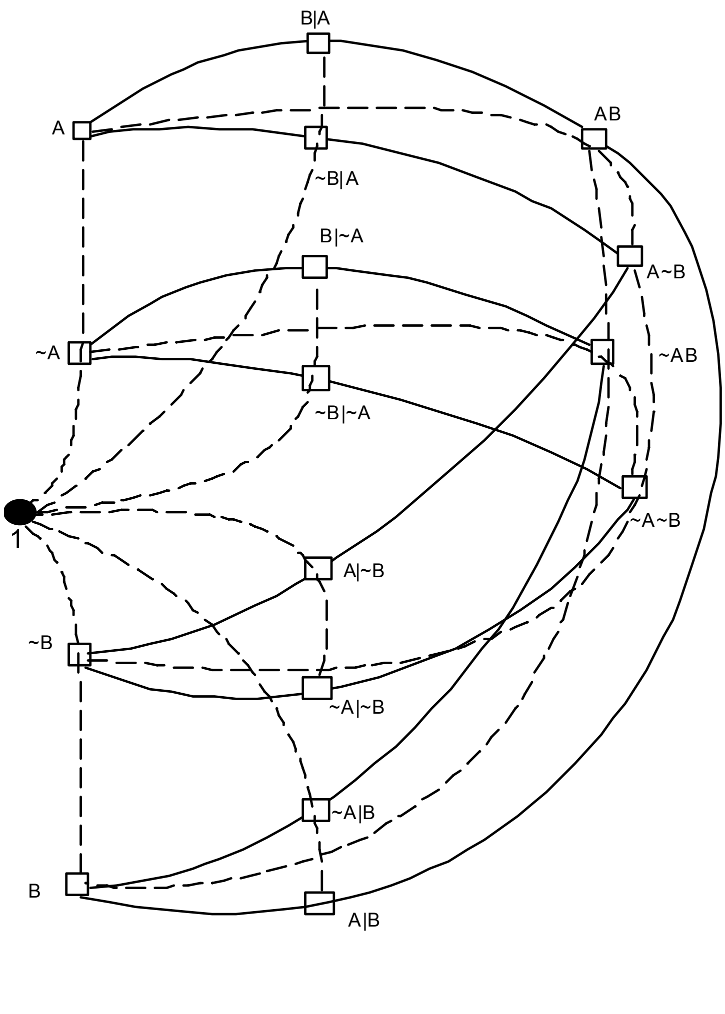

Fig. 1: Trinomial Graph of ternary problems of conditional probability

Ternary Problems of Conditional Probability have to fulfil at least the following conditions:

1. One conditional probability is involved, either as data or as a question or both.
2. Three quantities are known.
3. All quantities, known or unknown, are related by ternary relationships such as:

   $P(A) + P(noA) = 1$

   $P(A \cap B) + P(A \cap noB) = P(A)$

   (additive relationships); and

   $P(A|B) \times P(B) = P(A \cap B)$

   (multiplicative relationships)
4. One unknown quantity is asked for, which is related to other quantities by means of more than one of the relationships above.

It would be misleading to comprehend these conditions 1-4 as *characterizing* ternary problems. The accent lies on fulfilling *at least* these conditions. They could well fulfil more than these conditions and would still remain to be ternary problems. There are many school problems with conditional probabilities with more than three known data and they indeed *are* ternary problems. We call them extra-dimensioned problems because there are more known data than necessary to find out each unknown data. The term stems from the related graphs, which are extra-dimensioned in this case. For example, problems, which are presented in a contingency table format, are extra-dimensioned: *all* four data are known.

Evans and others (2000), Girotto and González (2001), Hoffrage and others (2002) agree that the presentation format of the data in the formulation of the problems is an influential factor on students' success even if the contextual information given in the problems is the same (Huerta & Lonjedo, 2006; Huerta & Lonjedo, 2007). In general, if data in the text of the problems are expressed in natural frequencies (Hoffrage, Gigerenzer, Krauss & Martignon, 2002), students are more likely to succeed in conditional probability problems. Estrada & Díaz (2006) report how pre-service teachers have difficulties in reading a contingency table with natural frequencies. Maury (1984) and Ojeda (1996) study whether both linguistic and contextual factors have an influence on students' success in solving conditional probability problems

But, we do not know of any work in which the purpose of analysis is focused on the study of structural, contextual, and phenomenological aspects (according to Freudenthal, 1983) of probability problems. We think that this kind of study is necessary in order to find out whether these factors also have an influence on problem solving processes or, which structures and contexts we should use in order to teach conditional probability to enhance students' understanding of the subject.

On Conditional Probability Problem Solving Research – Structures and Contexts

Lonjedo (2007) shows that in order to classify ternary problems of conditional probability it is necessary to take also into account the question in the text of the posed problem as unknown data. She consequently considers

- Level (N) of a problem as the number of known conditional probabilities in text of the problem. There are four levels, corresponding to 0, 1, 2, and 3 known conditional probabilities.
- Related to each level, she defines Category (C) as the number of known absolute probabilities in the formulation of the problem. Depending on each level, the category could be 0, 1 or 2.
- Finally, Type (T) of a problem represents the unknown data in the problem. There are three possible types: $T_1 \equiv$ conditional probability, $T_2 \equiv$ absolute probability and $T_3 \equiv$ intersection probability.

So, every ternary problem of conditional probability belongs to an N-family of problems described by means of a vector like (N_2, C_2, T_3). If a problem belongs to a (N_2, C_2, T_3)-subfamily, this means that in the formulation of the problem there are three known quantities that read themselves in a probabilistic sense; there should be one conditional probability (N_2), one absolute probability (C_2), and consequently one intersection probability, with one quantity being unknown, namely an intersection probability (T_3). See Table 1 for the result of the classification of all ternary problems of conditional probability into four families and twenty subfamilies. The symbol ∅ means that there is no ternary problem belonging to that subfamily.

	N_1			N_2			N_3			N_4		
C_1	C_1T_1	∅	∅	C_1T_1	C_1T_2	C_1T_3	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_2	C_2T_3	C_2T_1	C_2T_2	C_2T_3	∅	∅	∅
C_3	C_3T_1	∅	∅	C_3T_1	∅	C_3T_3	∅	∅	∅	∅	∅	∅

Table 1. Classification of ternary problems of conditional probability into families and subfamilies.

To each family and subfamily of problems we can associate a trinomial graph that represents the structure of data and relationships between data. These graphs are useful to state isomorphism between problems, complexities of problems and so. By working systematically on them by the analysis and synthesis method, we can derive the solution of all ternary problems.

In other papers (Huerta & Lonjedo, 2006 and 2007) we have investigated the influence of the nature and the data format in the text of ternary problems of conditional probability on students' problem solving behaviour. Depending on the data format, students perform better if data are expressed in frequency rather than in probability format. Related to this fact, and as a consequence, we observe that students find the assignment strategy more efficient than the calculation strategy when data in the text of the problems are expressed in frequency format even if students belong to upper levels of achievement (Huerta & Lonjedo, 2006). We may also conclude that students using the assignment strategy do not use data in the text of the problems in a probabilistic sense.

Following the first work, we (Lonjedo, 2007; Huerta & Lonjedo, 2007) continued to investigate this new concept whether it allows us to identify students' modes of reasoning depending on both the structure of data and data format. We used problems of the N_2-family with students from secondary school (14-18 years old) and from math college with different competences in mathematics. Only students from upper secondary school and university are instructed about conditional probability, whereas students from lower secondary school are not.

Because we were not only interested in students' success in solving problems but also in their solving processes, successful or not, we designed a set of descriptors to analyze problem solving behaviour (Huerta & Lonjedo, 2007). From your results we may distinguish four types of thinking in those students who worked successfully on the problems. We call these types as follows: *exclusively* arithmetic, *mostly* arithmetic, *basically* probabilistic and *exclusively* probabilistic, depending on students' problem solving approaches: using data in the text of the problems without probabilistic sense, in transition towards, or with probabilistic sense.

We also observe that in students' solving processes without success their difficulties are related to semantic variables, mainly in the used expressions about conditionality, so that these students misinterpreted data referring to conditional probability. According to our observations of the thinking process, their problem solving behaviour is strongly related to format, in which the data are presented in the tasks. With data given as frequencies, successful students use mainly arithmetic reasoning; with data presented in percentages or probabilities, successful students use probabilistic reasoning.

In general, in solving these problems students use data explicitly mentioned without translation from one format into another. Only in a few cases did students translate frequencies into probabilities and use probabilistic reasoning. These were all mathematics students but not all of them used such transformations. On the other hand, the percentage of students that succeed in solving structurally isomorphic problems increases if data in the formulation of the problem are expressed in frequencies and percentages and expressions related to conditionality are avoided (e.g. *and* and *also* are words, which refer to the intersection of events).

The investigations mentioned above do not pay attention to the context in which N_2-problems are stated. They investigate whether students' behaviour is dependent on problem structure and data format. So, we have knowledge about problem structure, data format and students' reasoning depending on both. But, in order to teach students about probability concepts through problem solving, we also need to have knowledge about situations and contexts in which problems are stated. A word problem could be seen as an instance of something more general. We use the term 'situation' for this more general and 'context' for a particular situation that is responsible for a restriction of the semantic field of a notion or concept (Puig & Cerdán, 1988). The semantic field related to a concept contains mathematical and non-mathematical meanings that allow us to make sense of events, event operations, and probability relationships.

Of course, in general we are interested in school *word* problems. This does not mean that we are not also interested in ternary problems formulated in a *symbolic* format, although this is not within the scope of this research. When ternary problems are analysed in such a symbolic context we would complete the analyses of ternary problems in *every* context they could be formulated in and we would consequently use the term "ternary problems" without the adjective "word". To simplify language here, we use ternary problems denoting "word ternary problems" and *not* in the general sense, which could include symbolic ternary problems, too.

In order to identify situations and context in which ternary problems of conditional probability are stated (Carles & Huerta, 2007; Carles, 2007) we base the research on the analysis of a didactical phenomenology (Freudenthal, 1983) of those problems. Due to the fact that we investigate problems in teaching-learning environments, our information comes from several sources, from textbooks in colleges, from the Internet, introducing a word chain in a searcher as follows: Probability, Conditional Probability, Conditioned Probability, Bayes' Theorem, and so on. In both cases, the main item in the search was conditional probability problems that have been used in teaching during the school year 2005-2006.

By analyzing the aforementioned documents we are able to classify them according to the following criteria:

- Context (in which the problem is formulated),
- phenomena referring to involved events (that is to say, organized by means of events),
- phenomena referring to involved probabilities (that is to say, organized by means of probabilities),
- specific terminology, classification (referring to the structure of the data in the text of the problem and the presentation format of the data), and
- specific teaching environment or reference.

The aforementioned criteria may be defined 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 such a context. In our work, this context not only is thought for problems like those in this page but for all those situations where something has to be tested in order to determine if it fulfils specific quality or health control criteria. Generally, tests are never completely reliable. Therefore test results are always afflicted with risks, which are usually expressed by probabilities.

Phenomena (referring to events): In a particular context, 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 can not be organized by means of a reference set.

Phenomena (referring to probabilities): In a particular context, apart from quantities, we refer to signs, words, and 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, which we consider, it acquires a probabilistic meaning. So, the phenomenon of *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.

As an example, in the next table, we will show the results of a phenomenological analysis of the following problems:

P2. 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 in fact is diabetic if his/her test is positive? What is the probability that a person is not diabetic if his/her test is negative?

P3. 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?

P4. 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.

Using the items for the analysis mentioned above, we present the results for the listed problems in Table 2. The problems we mentioned could be analyzed with the help of tables like Tables 3 and 4 and the graph in Figure 2. These tables organize the phenomenological analysis made in the Diagnostic Test Context in Health and the graph supports us to apply the analysis-synthesis method to make analytical readings of the problems.

Problem	**Con-text**	**Setting or area**	**Phenomena referring to**		**Specific terms**	**Classification and data format**
			events	**conditional probabilities**		
P2	Diagnostic test	Health setting	- To be diabetic - A positive person tested can suffer from diabetes - A negative person tested cannot suffer from diabetes	- False Positive Coefficient (FPC) - False negative Coefficient (FNC)	- FPC - FNC - Prevalence of diabetes - Test is positive - Test is negative - Diagnostic Test	$N_3C_2T_1$family - (1,0,2) - a) p(D|+) - b) p(~D|–) - Percentages
P3			- Not suffer from uterine cancer - Suffer from uterine cancer - To test positive in diagnostic procedure without uterine cancer - To test negative in diagnostic procedure with uterine cancer - A person tested positive can not suffer from uterine cancer	- FPC - FNC	- FPC - FNC - Pre-test probability - Negative result in test - Diagnostic Test	$N_3C_2T_1$family - (1,0,2) - p(~D|–) - Probability
P4			- Prevalence of the tuberculosis - To be infected by tuberculosis - To be not infected by tuberculosis	- Sensitivity - Specificity - False positive	- Sensitivity - Specificity - False-positive - Prevalence of disease - Tuberculin Test	$N_4C_1T_2$family - (0,0,3) - p(D) - Probability

Table 2. Aspects of the phenomenological analysis of the problems 2 to 4.

In Table 3 we show the reference sets in this context, which may serve as means of organization of the phenomena we describe, particularly in the health setting.

<table>
<tr><th colspan="4">Pre-test o pre-testing</th><th colspan="4">Post-test</th></tr>
<tr><td>D</td><td>noD</td><td>+</td><td>–</td><td>D∩+</td><td>NoD∩+</td><td>D∩–</td><td>NoD∩–</td></tr>
<tr><td rowspan="2">To be ill

To be infected

To suffer from a specific disease</td><td rowspan="2">Not to be ill

Not to be infected

Not to suffer from the disease</td><td>To test positive</td><td>To test negative</td><td>To suffer from a specific disease and to test positive</td><td>Not to suffer from the disease, and yet test positive</td><td>To suffer from a specific disease and to test negative</td><td>Not to suffer from the disease, and to test negative</td></tr>
<tr><td colspan="2">in the diagnostic procedure, regardless of the person's health status</td><td colspan="4">(or similar phrases referring to the conjunction of the two phenomena)</td></tr>
</table>

Table 3. Reference sets in the Diagnostic Context in a Health Setting.

In table 4, phenomena have been organized by means of probabilities of the aforementioned reference sets, now events in a mathematical sense, both by means of absolute and conditional probabilities. From that we necessarily read the data in a probabilistic sense.

Differences between Table 4 and another table for another setting, such as the quality control context for example, must be located in phenomena and specific terms but not in organization means and format of expression of data. Introducing organization means as a column in Table 4 is based on the relationships between phenomena and organization means that Freudenthal (1983, p. 32) identified some years ago, like an interplay of the pair phenomena-organization means. In our case, this supposes reference sets for "to suffer from a specific disease", "not to suffer from the disease", "to test positive in the diagnosing procedure", "to test negative in the diagnosing procedure", designed by capital letters or specifics signs: respectively "D", "noD", "+", "– ". These sets describe people for whom some of the phenomena are present. These sets, for themselves are probabilistically not very important unless we consider them as Borel sets and we can operate with them, that is, we can consider complements, unions and intersections of them (Freudenthal, 1983, p. 43). On the other hand, it is supposed to express judgements about these reference sets in terms of probabilities, like this: "the probability of D is…" that is usually expressed by means of signs such as $p(\mathrm{D})$.

The next graph represents the world of ternary problems of conditional probability (Carles & Huerta, 2007) in the context we are considering. The graph shows all relationships between events and quantities that give sense to both of them. Lines representing ternary relationships in the graph above are displayed in order to allow global and local analysis to be made. The conditional probabilities: sensitivity, specificity, false positive coefficient and false negative coefficient act on prevalence of D or prevalence of noD in order to update them after people have been tested by a diagnostic test. We talk about updating for example an absolute probability $p(\mathrm{D})$ by means of conditional probabilities, $p(\mathrm{D}|+)$ and $p(\mathrm{D}|-)$, resulting from this updating either positive and negative predictive, or false positive and false negative values. This updating involves a net of relationships (see next to the Fig. 2).

Phenomena (under uncertainty or referring to probabilities)	Specific Terms	Organization means	Data Format
Mistakes produced by test	FPC	$p(+ \mid \text{no } E)$	Necessarily in percentages and probabilities
	FNC	$p(- \mid E)$	
Success produced by test “VALIDITY”	Sensitivity	$p(+ \mid E)$	
	Specificity	$p(- \mid \text{no } E)$	
Mistakes produced in the diagnostic procedure “DIAGNOSTIC ERRORS”	False negative	$p(E \mid -)$	
	False positive	$p(\text{no } E \mid +)$	
Success produced in the diagnostic procedure “PREDICTIVE VALUES”	PPV or Positive Predictive Value	$p(E \mid +)$	
	NPV or Negative Predictive Value	$p(\text{no } E \mid -)$	
to have the disease (pre-test)	Prevalence of the disease	$p(E)$	Reasonably in natural numbers or absolute or natural frequencies
not to have the disease (pre-test)	Prevalence of no disease	$p(\text{no } E)$	
Results from the diagnostic test	To test positive	$p(+)$	
	To test negative	$p(-)$	
to have the disease and to test positive in the diagnostic procedure	We didn’t find out or it does not exist	$p(E \cap +)$	
not to have the disease and to test positive in the diagnostic procedure		$p(\text{no}E \cap +)$	
to have the disease and to test negative in the diagnostic procedure		$p(E \cap -)$	
not to have the disease and to test negative in the diagnostic procedure		$p(\text{no}E \cap -)$	

Table 4. Results of the phenomenological analysis in Diagnostic Test context in Health Setting.

We have been able to make a reading of the mistakes and success mentioned in Table 4 using probabilities. Therefore, in this sense, we have 8 conditional probabilities and 4 absolute or marginal probabilities, complementary in pairs. Some of these complementary relationships are important in this context:

FPC + specificity = 1 FNC + sensitivity = 1 probability (–) + probability (+) = 1

Now, the updated probability is also expressed next to Fig. 2. In order to talk about updating of the prevalence of a disease in a country that is tested by a diagnostic test, it would be necessary to take into account the pre-test prevalence of the disease and the specificity and sensitivity of the diagnostic test. We have called False Negative and Predictive Positive Value to those updating, terms that refer to post-test percentages or probabilities to suffer from a disease. Of course, all the problems can be read in a similar way as we will do with the next problem.

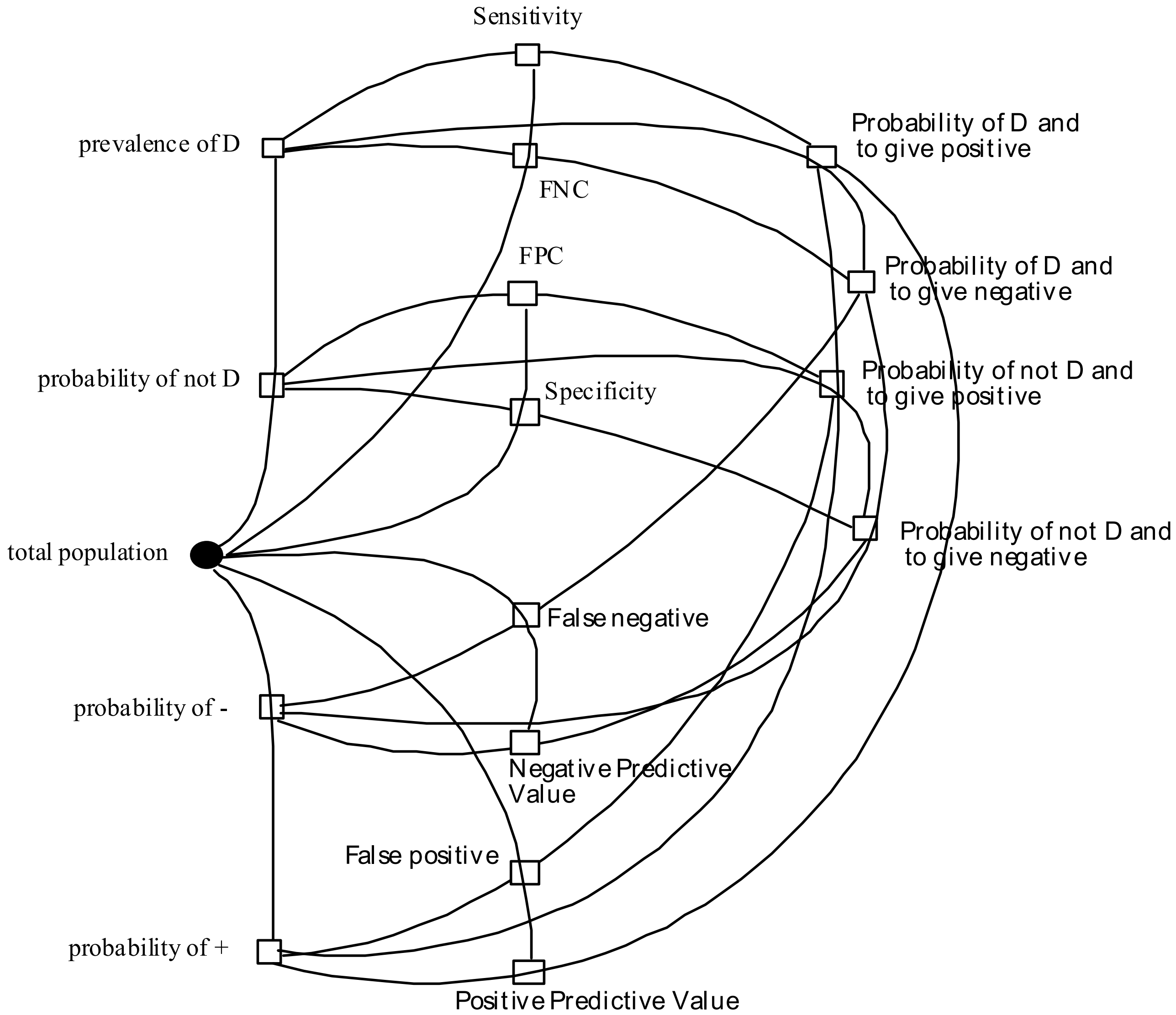

Fig. 2. Graph of the World of ternary problems concerning conditional probability in the Diagnostic test in health settings

Updating prevalence of D: yields

$$False\,negative = \frac{prevalence\ \mathrm{D} \times FNC}{probability\,(-)} \qquad False\,negative = \frac{prevalence\ \mathrm{D} \times (1 - sensitivity)}{1 - probability\,(+)}$$

$$PPV = \frac{prevalence\ \mathrm{D} \times sensitivity}{probability\,(+)} \qquad PPV = \frac{prevalence\ \mathrm{D} \times sensitivity}{probability\,(+)}$$

Substituting

$$probability\ (-) = 1 - probability\ (+)$$

$$FNC = 1 - sensitivity$$

P1. 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 yields a positive result in 97% of cases. When he/she is not tubercular, only 0.01% of the cases yield positive results. If the test is positive for that person, what is the probability that he/she is tubercular?

In the text, the expression "one out of every 100 citizens is a tubercular person" could be read as follows: in a town the prevalence of the tuberculosis is one out of 100 citizens". A mathematical reading of the same supposes to choose a reference set for the set of citizens that suffer from tuberculosis, let us denote that by T. Without a prior probabilistic sense we can affirm only few things about T, unless the number of citizens living in this town is known. On the contrary, with a prior probabilistic sense it is possible to affirm that $p(\mathrm{T}) = 0.01$.

The sensitivity of the test is communicated by means of the expression "If a person is tubercular, the test gives a positive result in 97% of cases", and the false positive coefficient by the expression "When he/she is not tubercular, only 0, 01% of the cases give positive results". That is, the sensitivity is 97% and the CFP is 0,01%. The corresponding mathematical reading of the data, events and probabilities is possible only if this is done with a probabilistic sense, because nothing is known about samples of population in which T is included. The probability of being positive in the test has to be updated from new information about T and noT, resulting that now the sensitivity is read $p(+|\mathrm{T}) = 0.97$ and the CFP, $p(+|\mathrm{noT}) = 0.0001$. The question in problem 1 asks about the positive predictive value (PPV) of the test by means of the expression "If the test is positive for these people, what is the probability that he/she is tubercular?", that is, about an updating of the prevalence of T from the additional information of the results of the test. In a probabilistic sense, it asks about $p(\mathrm{T}|+)$. Then, if we introduce in the graph in Figure 2 the three known probabilities and apply the analysis-synthesis method, we obtain:

$$PPV = \frac{prevalence\ \mathrm{T} \times sensitivity}{prevalence\ \mathrm{T} \times sensitivity + (1 - prevalence\ \mathrm{T}) \times FPC},$$

that is, PPV is a ratio between the quantities: the rate of the prevalence of tubercular and positive people related to the total of positives, which mathematically is the well known Bayes' Formula:

$$p(\mathrm{T}|+) = \frac{p(\mathrm{T}) \times p(+|\mathrm{T})}{p(\mathrm{T}) \times p(+|\mathrm{T}) + [(1 - p(\mathrm{T})] \times p(+|no\mathrm{T})}$$

Conclusion

The problems analyzed in this study may be considered as problems of application for the concept of conditional probability, that is to say, problems that are usually posed after instruction of the formal concept. But, teaching based on an exploration of the phenomena involving its application in various contexts is precisely the opposite point of view compared to formal teaching. Hence, as Freudenthal (1983) suggests, teaching the topic of conditional probability in secondary school might begin with solving problems that give students an opportunity to explore the pertinent phenomena and would only then be followed by teaching the formal concept of conditional probability as a means of organizing these phenomena. The problems we have analyzed in this paper might serve as a paradigm for this approach. These problems should be dealt with in secondary school education including the related context. The formal concepts of conditional probability should be introduced not before college level; after the basis with the phenomena is already laid, the formal concept could then serve as a means of modelling the phenomena and problems within a context.

One of the most commonly used contexts has been analyzed in this work. We used the term "diagnostic test". It can be recognized in various settings: in textbooks, in research on students' behaviour in solving conditional probability problems and so on. Generally, data and the relationships between data are not previously analyzed in relation to the context. But, if, when we think about teaching conditional probability we first analyze problems as we suggest in this work, we can determine what type of problems can reasonably be proposed to our students at every level of education in which the subject matter is taught and we can suggest in which context those problems should be stated in order to improve students' understanding of conditional probability.

This world of diagnostic tests exists and could bring together a professional application with an educational usage. The question is which teaching model we choose in order to improve students' conditional probability literacy: that based on a formal approach, that is, first Bayes' Formula and then applications via solving problems like 1 to 4 or a phenomenology-based model that focuses first on exploring phenomena via solving problems and only after that searching for means of organization of all of those phenomena. The position we defend here is clearly the second, following ideas of Freudenthal.

References

B.O.E.: 2006, Real Decreto 1513/2006, por el que se establecen las enseñanzas mínimas correspondientes a la Educación Primaria, *Boletín Oficial del Estado*, nº 293. Madrid, Ministerio de Educación y Ciencia.

B.O.E.: 2007, Real Decreto 1631/2006, por el que se establecen las enseñanzas mínimas correspondientes a la Educación Secundaria Obligatoria, *Boletín Oficial del Estado*, nº 5. Madrid, Ministerio de Educación y Ciencia.

Carles, M.: 2007, *Estudios preliminares de los problemas de probabilidad condicional en contexto. El caso del test de diagnóstico*, Memoria para el examen de DEA. (Preliminary studies on conditional probability problems in context. The case of diagnostic test, Protocol of the pre-doctoral DEA exam), Universidad de València.

Carles, M., Huerta, M. P.: 2007, Conditional probability problems and contexts. The diagnostic test context, in Pitta-Pantazi, D. & Philippou, G. (eds.) *Proc. Fifth Congress of the European Society for Research in Mathematics Education, CERME 5,* 702-710.

Cerdán, F. & Huerta, M. P.: 2007, Problemas ternarios de probabilidad condicional y grafos trinomiales. *Educación Matemática* 19 (1), 27-62.

Estrada, A. & Díaz, C.: 2006, *Computing probabilities from two way tables: an exploratory study with future teachers*. Paper presented at ICOTS 7.

Evans, J., Handley, S. J., Perham, N., Over, D. E., & Thompson, V. A.: 2000, Frequency versus probability formats in statistical word problems. *Cognition* 77*,* 197-213.

Freudenthal, H.: 1983, *Didactical phenomenology of mathematical structures*, Reidel, Dordrecht.

Gal, I. 2005, Towards "probability literacy" for all citizens: Building blocks and instructional dilemmas, in Jones, G. A. (ed.), *Exploring Probability in School: Challenge for Teaching and Learning*, Springer, Berlin, 39-63.

Girotto, V., González, M.: 2001, Solving probabilistic and statistical problems: a matter of information structure and question form. *Cognition* 78, 247-276.

Hoffrage, U., Gigerenzer, G., Krauss, S., Martignon, L.: 2002, Representation facilities reasoning: what natural frequencies are and what they are not, *Cognition* 84, 343-352.

Huerta, M. P., Lonjedo, Mª A.: 2006, The nature of the quantities in a conditional probability problem. Its influence on the problem solving behaviour, in Bosch, M. (ed.), *European Research in Mathematics Education IV. Proc. Fourth Congress of the European Society for Research in Mathematics Education (CERME 4)*, 528-538.

Huerta, M.P., Lonjedo, M.A.: 2007, The same problem in three presentation formats: Different percentages of success and thinking processes, in Pitta-Pantazi, D. & Philippou, G. (eds.) *Proc. Fifth Congress of the European Society for Research in Mathematics Education, CERME 5,* 732-741.

Lonjedo, M. A.: 2007, *Análisis de los problemas ternarios de probabilidad condicional de enunciado verbal y de sus procesos de resolución* (Analysis of word ternary problems on conditional probability and their resolution processes). Doctoral Dissertation. Universitat de València.

Maury, S.: 1984, La quantification des probabilités: analyse des arguments utilises par les élèves de classe de seconde, *Recherches en Didactique des Mathématiques* 5 (2), 187-214.

NCTM: 2000, *Principles and Standard for school mathematics*, Reston, VA; N.C.T.M. Online: http://standards.nctm.org/

Ojeda, A. M.: 1996, Contextos, representaciones y la idea de probabilidad condicional, in Hitt, F. (ed.) *Investigaciones en Matemática Educativa*, Grupo Editorial Iberoamericana, México, 291-310.

Puig, L., Cerdán, F.: 1988, *Problemas aritméticos escolares*, Síntesis, Madrid.

Shaughnessy, J. M.: 1992, Research in probability and statistics: Reflections and directions. In Grouws, D. (Ed.), *Handbook of Research on Mathematics Teaching and Learning,* MacMillan, New York, 465-494.

Watson, J.: 2005, The probabilistic reasoning of middle school students. In Graham A. Jones (ed.), *Exploring Probability in School: Challenge for Teaching and Learning*, Springer, Berlin, 145-169.

Watson, J.: 2006, Chance—Precursor to Probability, in Watson, J. 2006, *Statistical Literacy at School*, Lawrence Erlbaum, Philadelphia, 127-185.