F.J. Díez

Dept. Artificial Intelligence
Universidad Nacional de Educación a Distancia
Madrid, Spain

Research and Education

Teaching probabilistic medical reasoning with the Elvira software

Abstract. This paper briefly describes the main features of the UNED (Spanish National University for Distance Education) and the course on *Probability and Statistics in Medicine*. Then it introduces Bayesian networks and influence diagrams, two of the methods taught in this course. Finally, it explains how Elvira, a software package, can help the students to understand some difficult probabilistic concepts.

1. UNED, the Spanish National distance university

The UNED (Universidad Nacional de Educación a Distancia) was founded in 1972. Its purpose was to offer high-level training to those students that could not attend a regular university because of their jobs, the lack of a university near their homes, or any other reason. Currently, the UNED is the biggest Spanish university, with 200,000 pre-graduate students. Its structure consists of a Central Office and a network of support centers. The Central Office (*Sede Central*), located in Madrid, comprises the management offices and eleven schools that gather a total of 1,200 teachers, 62% of which are doctors—the highest rate among Spanish universities. Additionally, the UNED has 156 support centers in Spain and 15 centers in other countries, in which the students can perform administrative procedures, receive tutorial assistance, access bibliographical and computational resources and be examined three times a year: in February, June, and September. The UNED has special programs for students who are abroad, students in prison (there are currently 600), and handicapped students.

The teaching materials made up by the UNED consist basically of printed stuff (guides, text books, addendas, etc.) and audiovisual materials (videos, cassettes, multimedia contents on CD-ROM, and radio and TV programs).

The departments teaching at the School of Computer Science (*Escuela Técnica Superior de Ingeniería Informática*) have been using the Internet as a mean of contact with their students at least since 1995. Currently the UNED is in the process of implementing a program that aims at offering Internet support for all its studies. Of course, the students can still contact their teachers at the Central Office and their tutors at the support centers by means of telephone, fax, conventional mail, or personal meetings.

2. The course on *Probability and Statistics in Medicine*

In 1996 two professors of the UNED organized a course on *Probability and Statistics in Medicine*. Prof. Francisco Javier Díez, a specialist in artificial intelligence and probabilistic reasoning applied to medicine, was in charge of the part on probability and Prof. Pedro Juez, a specialist in health economy, was in charge of the part on statistics. Table 1 shows the contents of the current edition of the course.

The purpose of the course is to offer to health professionals the basic probabilistic and statistical formation for clinical practice, biomedical research and health management. As a result, students should be able to apply different methods with some of the existing software tools. More specifically, the teaching of probability is based on Elvira (see below) and the teaching of statistics on SPSS,[1] but it will not be difficult for the students to use other packages after learning how to use these. Given that the student is not supposed to do the numerical analysis by him/herself, but by means of such software tools, the course does not dive into the computational complexities of the algorithms involved. It rather concentrates on the conceptual properties of each method and on the interpretation of the results returned

---

[1] Elvira is a free software that can be dowloaded from the Internet (see below). SPSS is a commercial package available in many hospitals and research centers. The students of the UNED can also acquire a one-year licence at a reduced price.

Table 1. Contents of the course on ***Probability and Statistics in Medicine***

First part. Evidence-based medicine	
Chapter 1.	EBM: objective and qualitative medicine
Second part. Probability	
Chapter 2.	Fundamentals of probability
Chapter 3.	Bayesian networks
Chapter 4.	Influence diagrams and decision trees
Chapter 5.	Subjective estimation of probabilities
Chapter 6.	Probabilistic foundations of statistical inference
Third part. Statistics	
Chapter 7.	Design of epidemiological studies
Chapter 8.	Computation of the sample size
Chapter 9.	Descriptive statistics
Chapter 10.	Linear regression
Chapter 11.	Contingency tables
Chapter 12.	Logistic regression
Chapter 13.	Statistical tests
Chapter 14.	Discriminant analysis
Chapter 15.	Survival analysis and Cox regression
Chapter 16.	Joint analysis

by the computer, with special attention to the conditions that justify the application of each method to particular problems, since the results of the computer would unreliable or even completely wrong if such conditions were not fulfilled.

In the first seven editions of the course over 1,000 health professionals have registered and over half of them have passed the exams and obtained the degree. We currently are studying the possibility of imparting this course in Mexico, in collaboration with the ITESM.

2.1. Methodology

After registering, the student receives a study guide, two books, a collection of solved problems, a collection of exercises for the evaluation of the course, and a CD-ROM by ordinary mail. There are four four-hour sessions that students can attend either physically in Madrid, by videoconference transmitted to some of the support centers of the UNED, or by Internet. There is also a work group at aLF, a collaborative learning environment developed by the UNED.[2] This Internet tool offers web pages, group e-mail, a forum (similar to a news group, but much more powerful), a download area, etc. The CD-ROM contains the Elvira program, some documents, and the next year it will also contain a recording of the videoconferences (in MPEG format, for instance). The students can contact the instructors by e-mail, ordinary mail, phone, fax, or personally in Madrid.

3. Probabilistic graphical models

After a brief first part that introduces evidence-based medicine, the second part of the course studies four probabilistic methods: the naïve Bayes method, Bayesian networks, decision trees, and influence diagrams. The naïve Bayes is a particular case of a Bayesian network in which there is a root node that is the parent of all the other nodes (see below). Both methods are used for diagnosis and prognosis, while decision trees and influence diagrams are decision-analysis tools that can be applied, in the case of medicine, for therapy selection and the economic evaluation of medical technologies. There are other graphical probabilistic models, such as Markov decision graphs, not studied in this course.

3.1. Bayesian networks

A Bayesian network [1,2] consists of an acyclic directed graph, whose nodes represent random variables, together with a family of conditional probability distributions (CPDs) for each node. In the case of discrete variables, each family of CPDs is given by a table of probabilities. The product of all the CPDs gives the joint probability for the variables of the network. This distribution satisfies the *d*-separation} property [1], which is equivalent to the Markov property [3] (see below). Any other marginal or conditional probability can be obtained from it. In particular, probabilistic diagnosis consists of computing the posterior probability of the unobserved variables given the evidence. (The evidence is the set of findings, obtained from the clinical history, physical examination, laboratory tests, etc.) This computation is usually called *evidence*

---

[2] The web site for aLF is www.innova.uned.es.

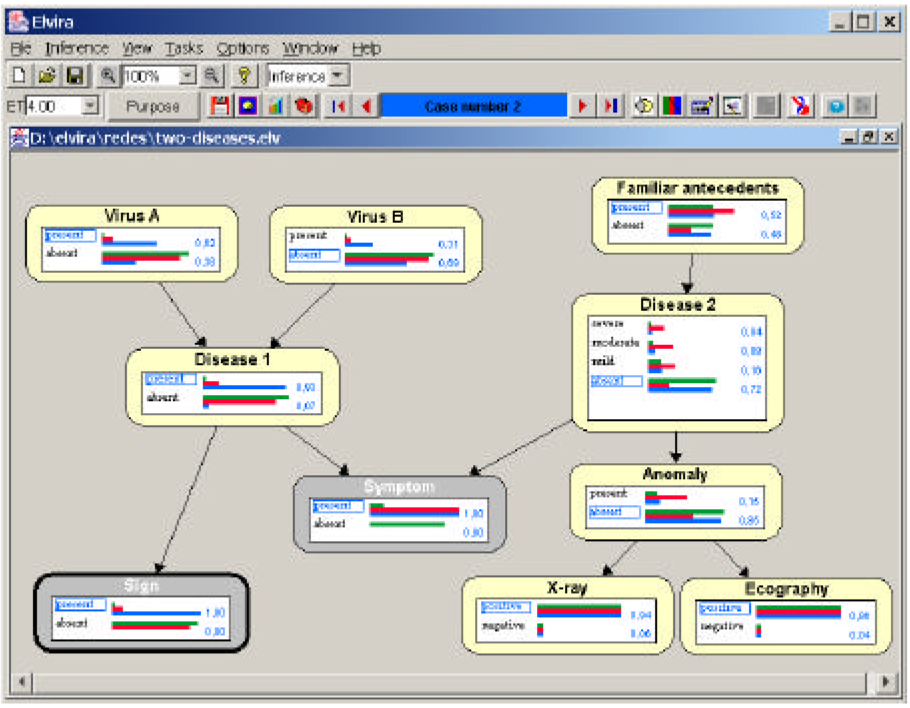

Fig. 1. A Bayesian network for the differential diagnosis of two diseases.

propagation, and is based, more or less explicitly, in the application of Bayes theorem.[3]

For instance, Figure 1 shows how the probabilities of two diseases increase or decrease when several findings are introduced. The finding that the patient has a Symptom increases the probability of both Disease 1 and Diseased 2. The presence of a certain Sign confirms Disease 1 and "explains away" (reduces our suspicion of) Disease 2.

There are two ways of building a Bayesian network (BN). The *automatic method* consists of using a database and applying some of the many algorithms available, which give both the structure and the CPDs of the network. In this case, the resulting network is a black box model whose links do not have a direct interpretation. It is similar, therefore, to other black-box diagnostic methods, such as neural networks or logistic regression. The automatic construction of BNs from databases is a very active line of research, producing several new algorithms every year [2]. However, the lack of high-quality databases makes the automatic construction of Bayesian networks unfeasible in many cases, especially in medicine.

The *manual method*, in contrast, consists of first building the graph of the network by drawing causal links among the nodes, according with experts' knowledge of causal mechanisms, and then obtaining the conditional probabilities from epidemiological studies, databases, medical literature or experts' estimates [4,5].

The appearance of Bayesian networks in the 1980's allowed the practitioners of artificial intelligence to overcome the main limitations of previous probabilistic methods. In fact, the naïve Bayes method, used in the 1960's for building the first artificial-intelligence diagnostic systems, assumed that the diagnoses were mutually exclusive and that the findings were conditionally independent given the diseases. Although those systems performed quite well in small diagnostic problems, the method was unfeasible for larger domains. The possibility of diagnosing multiple diseases in a natural way and the use of graphs —in general, causal graphs— to represent the dependences and independencies among variables led to a renewed interest for probabilistic methods in artificial intelligence. Unfortunately, in the field of medicine there is still a general unawareness of the existence of Bayesian networks, and even the articles and books that advocate the application of probabilistic methods in clinical practice limit themselves to mentioning the old naïve Bayes method, without offering any alternative to overcome its serious drawbacks—see for instance [6].

3.2. Decision trees and influence diagrams

A probabilistic decision tree [7] contains two types of nodes: *decision nodes*, drawn as squares, represent the decisions to be made, and *random nodes*, drawn as circles, represent random variables, i.e., variables that are not under the control of the decision maker, such as diseases and symptoms. The *root node* is the only one that has no parents. Each decision node gives birth to several branches, corresponding to the options for that decision. In turn, each random node gives birth to several branches, one for each possible outcome. In the usual representation of decision trees, the root node is drawn on the left and branches expand to the right. The right-most branches have associated *utilities*, which represent the benefit or cost associated to each set of decisions and outcomes.

---

[3] This is the origin of the term *Bayesian network*. We would like to underline that BNs are not tied to a subjectivistic interpretation of probability, as in the case of *Bayesian* statistics. In fact, the probabilities associated with a BN can be seen as subjective opinions (Bayesian interpretation) or as real-world frequencies (frequentist interpretation).

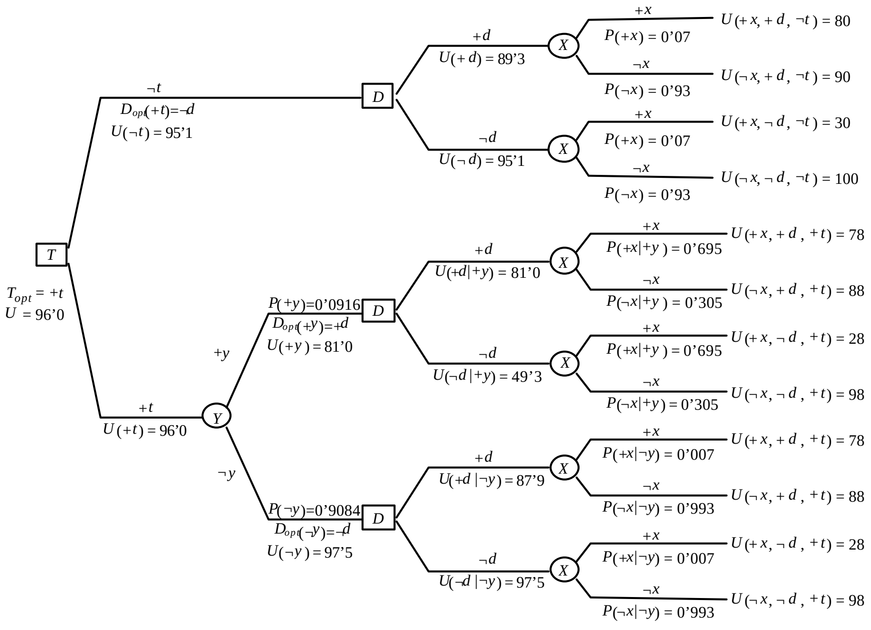

Fig. 2. Decision tree for deciding whether it is beneficial to perform test ***Y.***

For instance, the decision tree in Figure 2 contains two decisions —whether to perform a test (decision *T*) and whether to apply a certain therapy (decision *D*)— and two random variables: *X* represents the presence of a certain disease and *Y* represents the result of the test.

The evaluation of a decision tree is performed from right to left. The utility associated with a *random node* is computed as the weighted average of the utilities of its branches, where the weight of each branch is the probability of the corresponding outcome. The utility associated with a *decision node* is the maximum of the utilities of its branches; the option that yields the highest utility is the optimal decision for that node.

Given that the construction of decision trees is very difficult in many practical problems, influence diagrams were developed as an alternative representation method for decision analysis [1,2,8]. Each decision tree is equivalent to an influence diagram, and vice versa, but in general (causal) influence diagrams are much easier to build, because they contain an explicit representation of causal dependencies and independencies, because they use cause-effect probabilities, which in general can be obtained much more easily than the effect-cause probabilities required by decision trees, and because different utilities are represented separately. Please compare Figures 2 and 3.

Influence diagrams can be viewed as an extension of Bayesian networks: both use a graph - generally a causal graph- and a set of conditional probability distributions. The main difference is that Bayesian networks only contain *random nodes*, while influence diagrams also contain *decision nodes* and *utility nodes* (the two bottom nodes in Figure 3 are utility nodes).

Another advantage of influence diagrams with respect to decision trees is that there exist methods for the evaluation of influence diagrams that are much more efficient than the expansion and evaluation of the associated decision tree [2,9].[4]

Unfortunately, influence diagrams, developed in the 1980s [8,9], are still

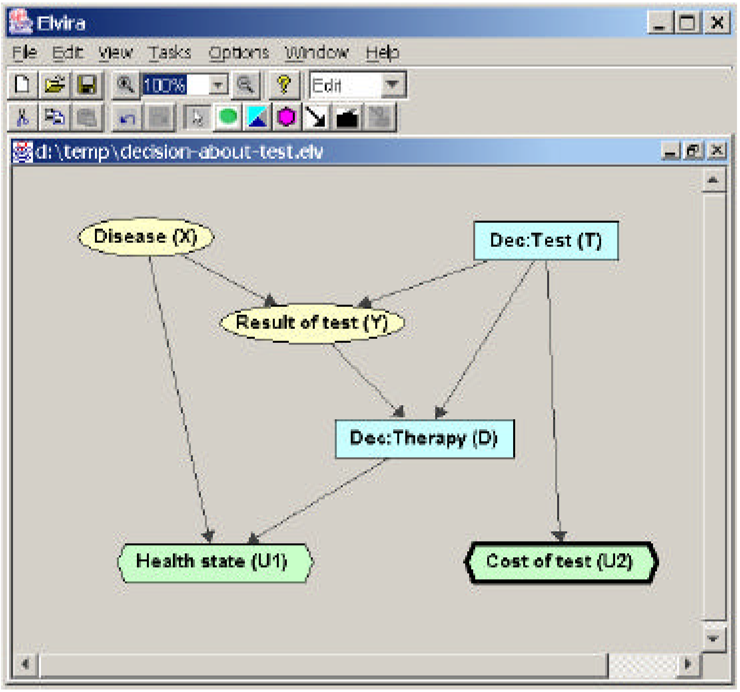

Fig. 3. Influence diagram corresponding to the decision tree in Figure 2.

---

[4] In the past, there were some problems in which decision trees were more efficient than influence diagrams. However, recent advances in algorithms for the evaluation of influence diagram have made them at least as efficient as decision trees for all kinds of problems, and much more efficient in general.

rarely used in medicine, even in medical research—see for instance any issue of the *Medical Decision Making* journal. Most articles and books only use decision trees, which were developed in the late 1950s [7]. As a remedy, our course teaches our students how to build influence diagrams for medical problems and how to evaluate them by either expanding the associated decision tree or, which is much more efficient, by using the Elvira software.

4. Elvira

4.1. The Elvira project and the Elvira program

Elvira is a free software package developed as a joint project of several Spanish universities, which began in 1997 [10]. Elvira has its own format for reading and writing Bayesian networks and influence diagrams, a parser, a graphical interface for building and evaluating the networks, exact, stochastic and approximate algorithms for the evaluation of the models, advanced explanation facilities, and several methods for learning Bayesian networks from databases.

In addition to the wide set of methods for inference and learning and the advanced interface, the main quality of Elvira is that it is implemented in Java, and therefore it runs on several platforms. Another advantage is that Elvira is freely available on the Internet, at www.ia.uned.es/~elvira, including the source code, which consists of around 115,000 lines. The graphical interface can be displayed in Spanish or in English, and, given that it uses the Java internationalization facilities, it will be easy to add other languages in the future. On the other hand, the main weaknesses of Elvira are its relative inefficiency, partially due to the use of Java, the lack of on-line help, and the poor documentation of the source code.

Elvira has been used for building and debugging several medical Bayesian networks [11-13], and is currently used in the construction of influence diagrams for cost-effectiveness analysis in three medical problems: infectious diseases in terminal patients, lung cancer, and head injuries. The latter two projects are developed in collaboration with two former students of the course on *Probability and Statistics in Medicine*.

4.2. Use of Elvira as a pedagogical tool

Elvira has proved to be an efficient tool for illustrating some probabilistic concepts and properties that were difficult to understand for our students, even for those that had a certain mathematical background. For instance, the relation of Bayes theorem with probabilistic parameters, such as prevalence, sensitivity, specificity, odds, and likelihood ratios, or the relationship between causality, graphical models, and conditional independence, can be intuitively understood by "playing" with some Bayesian networks in Elvira.

For instance, the Markov property [2,3] is very difficult to understand when stated in mathematical terms: "A node is conditionally independent of its non-descendants given its parents". However, it becomes quite intuitive when we open in Elvira the causal network displayed in Figure 1, introduce some findings, and observe which probabilities vary and which do not.

As mentioned above, this network also shows that the presence of the Symptom increases the probability of both Disease 1 and Diseased 2, and the presence of the Sign confirms Disease 1 and "explains away" (reduces our suspicion of) Disease 2, while the absence of this Sign or a positive echocardiographic result will tend to confirm Disease 2 and discard Disease 1, thus performing a differential diagnosis based just on the application of Bayes theorem on a graphical probabilistic model. This and other examples show our students that the numerical results obtained from simple diagnostic models agree with common-sense reasoning.

In the same way, the construction of simple decision trees and influence diagrams, such as those explained in Section 3.2 (Figures 2 and 3), shows our students the applicability of decision theory to medicine. As in the case of probabilistic diagnosis, we try to explain them the close relation of common-sense decision making with the mathematical principles and methods of decision theory. In particular, Elvira has been especially useful for analyzing how the variation of the numerical parameters of a medical influence diagram (prevalence and severity of a disease, sensitivity, specificity, and cost of a test, etc.) may affect the therapeutic policy. This way, we demonstrate that influence diagrams are much more flexible and powerful than static *clinical practice guidelines*, for three reasons: first, because influence diagrams can uncover the optimal policy in many cases in which human experts can not; second, because, contrary to static guidelines, influence diagrams can be easily adapted to new situations—for instance, by setting the prevalence of a disease to that of a different country, by updating the economic costs, by adding new diagnostic or therapeutic techniques, etc.; and third, because instead of assuming what it is best for a "universal" patient, influence diagrams can explicitly take into account the preferences of each *real* patient.

This way, medical doctors are not passive appliers of guidelines developed by others but, on the contrary, active decision analysts—and much more, of course. In fact, the adaptation of an influence diagram to a specific setting and to an individual patient requires a lot of medical knowledge, expertise, communication skills, common sense, and intuition.

5. Conclusion

In the probabilistic part of the course on *Probability and Statistics in Medicine* we try to show our students that medical reasoning is based on probabilistic concepts. In particular, Bayesian networks can be used to model and solve many diagnostic problems, and influence diagrams are a substitute or a complement of decision trees for the analysis of medical decision problems. The use of Elvira, a software tool for Bayesian networks and influence diagrams, has been very useful for our students to intuitively understand probabilistic and decision-theoretic concepts, and to build their own models. We have used Elvira in the construction of several medical Bayesian networks and are currently conducting two research projects of cost-effectiveness analysis in collaboration with former students of this course.

Acknowledgments

The Elvira project was financed by the Spanish Ministry of Science and Technology under grants TIC97-1135-C04 and TIC2001-2973-C05.

References

1. Pearl J. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. San Mateo, CA: Morgan Kaufmann; 1988.
2. Jensen FV. Bayesian Networks and Decision Graphs. New York: Springer; 2001.
3. Neapolitan RE. Probabilistic Reasoning in Expert Systems: Theory and Algorithms. New York: Wiley-Interscience; 1990.
4. Druzdzel MJ, van der Gaag LC. Building probabilistic networks: “Where do the numbers come from?” (Guest editors’ introduction). IEEE Transactions on Knowledge and Data Engineering 2000;12:481-6.
5. Druzdzel M, Díez FJ. Combining knowledge from different sources in probabilistic models. J Machine Learning Research, 2003;4:295-316.
6. Sackett DL, Richardson WS, Rosenberg W, Haynes RB. Evidence-based Medicine. How to Practice & Teach EBM. 2nd ed. New York: Churchill Livingstone; 2000.
7. Raiffa H, Schlaifer RO. Applied Statistical Decision Theory. Cambridge, MA: Harvard Business School; 1961.
8. Howard RA, Matheson JE. Influence diagrams. In: Howard RA, Matheson JE, editors. Readings on the Principles and Applications of Decision Analysis. Menlo Park, CA: Strategic Decisions Group; 1984. p. 719-762.
9. Olmsted SM. On Representing and Solving Decision Problems. PhD thesis. Dept. Engineering-Economic Systems. Stanford University, CA; 1983.
10. The Elvira Consortium. Elvira: An environment for creating and using probabilistic graphical models. Proceedings of the First European Workshop on Probabilistic Graphical Models (PGM’02); 2002 Nov; Cuenca, Spain.
11. Lacave C, Onisko A, Díez FJ. Debugging medical Bayesian networks with Elvira’s explanation facility. Conference on Artificial Intelligence in Medicine (AIME-2001), Workshop on Bayesian Models in Medicine; 2001 Jul; Cascais, Portugal.
12. Galán SF, Aguado F, Díez FJ, Mira J. NasoNet. Modelling the spread of nasopharyngeal cancer with temporal Bayesian networks. Artificial Intelligence in Medicine 2002;25:247-254.
13. Lacave C, Onisko A, Díez FJ. Knowledge acquisition in Prostanet, a Bayesian network for diagnosing prostate cancer. Seventh International Conference on Knowledge-Based Intelligent Information & Engineering Systems (KES-2003); 2003 Sep; Oxford, UK.

Address of the author:
Francisco J. Díez
Dept. Artificial Intelligence
Universidad Nacional de Educación
a Distancia
Senda del Rey, 9
28040 Madrid, Spain
E-mail: fjdiez@dia.uned.es