

Criteria for Combining Knowledge from Different Sources in Probabilistic Models

Marek J. Druzdzel

Decision Systems Laboratory
School of Information Sciences
University of Pittsburgh
Pittsburgh, PA 15260, U.S.A.
marek@sis.pitt.edu

F. Javier Díez

Departamento de Inteligencia Artificial
Universidad Nacional de Educación a Distancia
Senda del Rey, 9
28040 Madrid, Spain
fjdiez@dia.uned.es

Abstract

Building probabilistic and decision-analytic models requires a considerable knowledge engineering effort in which obtaining numerical parameters is especially daunting. Often knowledge engineers combine various sources of information, such as information reported in textbooks and professional literature, available statistics, and data collected in practical settings. We show that combining probabilistic knowledge that originates from different sources requires utmost care. In particular, we demonstrate that even such seemingly population-independent characteristics as sensitivity and specificity of medical symptoms can vary within a population, depending purely on how the data are collected. We offer guidelines for detecting when different sources of data can be safely combined. Our analysis shows that a knowledge engineer should exercise much care in building practical models.

1 Introduction

Development of probabilistic graphical models, such as Bayesian networks [Pearl, 1988] and closely related influence diagrams [Howard and Matheson, 1984], has caused a considerable interest in applying probability theory and decision theory in intelligent systems (see [Henrion *et al.*, 1991] for an accessible overview of decision-analytic methods in AI). Graphical probabilistic models have been successfully applied to a variety of problems, including medical diagnosis, prognosis, and therapy planning, machine diagnosis, user interfaces, natural language interpretation, planning, vision, robotics, data mining, and many others (for examples of successful real world applications of Bayesian networks, see March 1995 special issue of

the *Communications of ACM*). While in the sequel we focus on Bayesian networks, our discussion applies equally well to influence diagrams, which could be viewed as Bayesian networks enhanced with an explicit representation of decisions and utilities over the outcomes of the decision process.

A Bayesian network consists of a qualitative part, reflecting the structure of the model in terms of an acyclic directed graph encoding probabilistic influences among a domain's variables, and a quantitative part, encoding the joint probability distribution over these variables. While building the structure of a model is in itself a challenging task that needs much care, most practitioners consider it doable. In our experience, most medical experts, for example, either give similar graphical structures or converge on the same structure after some discussion [Díez *et al.*, 1997, Oniško *et al.*, 1999]. Directed graphical models built in practice usually mimic the causal structure of a domain, which, given the fundamental role of causality in scientific understanding, explains expert agreement on the structure of models. The structure of a directed graphical model is subsequently quantified with numerical probabilities. Quantification of a Bayesian network consists of prior probability distributions over those variables that have no predecessors in the network and conditional probability distributions over those variables that have predecessors. Typically, a large model needs hundreds or even thousands of numbers and the task of quantifying a network is daunting. Since the expert time is scarce and, therefore, costly, knowledge engineers utilize various sources of information. These include, for example, existing textbook knowledge, available statistics, and databases of cases.

In this paper we focus on dangers related to combining various sources of numerical data in the same model. We show that lack of attention to whether the sources are compatible and whether they can be combined can lead to erroneous behavior of the model. We show that data that seemingly capture the properties of the

same population of individuals can differ, depending on how they were collected. The problem that we are describing has nothing to do with problems related to small databases, missing data, or unreliable expert judgment. Neither is it an instance of selection bias, even though the two are related. It occurs when data about the same population are collected in a different way and combined into one model. Even such population-independent parameters as sensitivity and specificity of medical symptoms and tests vary within the same population as a function of the way the observations were collected. These statistical effects go beyond simple and obvious differences in populations. We study the problem on three simple examples and offer guidelines for detecting when different sources of data can be safely combined.

This work was inspired by our practical experiences in building medical diagnostic systems. We encountered a puzzling phenomenon that led to an initial disagreement between us. We have subsequently analyzed the problem gaining insight that escaped each of us despite our field experience. We suspect that many knowledge engineers face similar problems, often not realizing them. We hope that this paper will make them aware of the problem and prepare them better for dealing with it in practice. We base the paper on medical examples, but the principles applied and the conclusions of this analysis are general.

We will use upper case letters, such as V , to denote variables and lower-case letters, such as v , to denote their outcomes. When a variable V is propositional, $+v$ and $-v$ will denote the truth and falsity of the proposition respectively.

The remainder of this paper is structured as follows. Section 2 presents a motivating imaginary example in which we show a problem related to using knowledge from different sources. Section 3 analyzes the example in depth and explains statistical reasons for the somewhat surprising fact that combining the knowledge is incorrect. Section 4 shows two more examples, generalizes this analysis, and shows that the effect is due to probabilistic conditioning on unmodeled variables. It also gives guidelines for dealing with this problem in practice. Finally, Section 5 discusses the implications of our analysis for knowledge engineering.

2 A Motivating Example

Two internal medicine residents decided to build a simple diagnostic decision support system for a certain disease D . In the first version of the system, they decided to model only D and its most important symptom S . They started by creating a Bayesian network consisting of two nodes, D and S (see Figure 1).

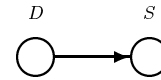


Figure 1: Example model under construction.

In the second stage, the residents focused on obtaining numerical parameters for their network. These parameters consist of $\Pr(+d)$, the prevalence of D , and the conditional probabilities of the symptom given the disease, $\Pr(+s|+d)$ and $\Pr(-s|-d)$, also known as *sensitivity* and *specificity* of the symptom. They decided to obtain these parameters from a data set of previous patient cases collected at their hospital. While there was no disagreement about sensitivity and specificity of the symptom, the residents had different opinion about the prevalence. One of them said that they need to take the prevalence as observed in their hospital, while the other suggested that they should take the prevalence for general population, so that their system remains unbiased. After all, the second resident argued, one of the reasons why people are admitted to the hospital is because of symptom S , so if they used the hospital prevalence rate, the evidence would be double-counted. They ended up using the latter.

While the reader may disagree with the arguments made by either of the residents, it is easy to imagine obtaining the same model by the sheer fact that prevalence of a disease in the general population is often easy to find in a statistics yearbook or a morbidity table and the sensitivity and specificity may be in practice obtained from hospital records or elicited from an expert with clinical experience, i.e., one who has seen a large number of cases in clinical settings.

Let the prevalence of the disease, $\Pr(+d)$, taken from an epidemiological study performed in the town in question, be $\Pr(+d) = 0.01597$. Let the hospital data be summarized in the following contingency table.

N	$+d$	$-d$	Total
$+s$	729	63	792
$-s$	1	174	175
Total	730	237	967

Sensitivity and specificity extracted from this table are

$$\Pr(+s|+d) = 729/730 = 0.99863 \quad (1)$$

$$\Pr(-s|-d) = 174/237 = 0.73418 \quad (2)$$

The most important type of reasoning in Bayesian networks is known as *belief updating*, and amounts to computing the probability distribution over variables of interest conditional on other, observed variables. In other words, the probability distribution over the variables of interest is adjusted for a particular case, in

which some other model variables assume given values. A medical diagnostic system based on a Bayesian network, for example, focuses on the probability distribution over possible diseases conditional on various risk factors, symptoms, and results of medical tests that constitute a patient case.

In our example model, the variable of interest is D and we are interested in the posterior probability distribution over D given an observed value of S . According to the thus constructed model, the probability that a patient presenting with symptom S suffers from D is

$$\begin{aligned} & \Pr(+d|+s) \\ &= \frac{\Pr(+s|+d) \cdot \Pr(+d)}{\Pr(+s|+d) \cdot \Pr(+d) + \Pr(+s|\neg d) \cdot \Pr(\neg d)} \\ &= 0.05748. \end{aligned} \quad (3)$$

Leaving aside a possible estimation error in obtaining the probabilities from the database, the procedure followed seems to be correct. Nevertheless, we are going to show in the next section that this model and the posterior probability computed by it, $\Pr(+d|+s) \approx 6\%$, are incorrect.

3 Analysis of the Problem

To understand this problem, we should model the variable H , hospital admission, explicitly. Figure 2 shows a graph modeling variables D , S , and H . From the point of view of disease D , admission to the hospital depends directly only on observing the symptom S . In other words, H is independent of D given S , or $\Pr(h|s,d) = \Pr(h|s)$. Given the symptom S , knowing whether the patient is in the hospital does not influence our belief in the presence of the disease, i.e.,

$$\Pr(d|s,h) = \Pr(d|s), \quad (4)$$

which means that, if we know whether the patient has the symptom or not, the information that the patient has been admitted to the hospital does not affect the diagnosis.

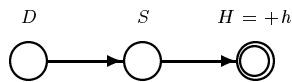


Figure 2: A causal model for the hospital data set. Note that since each data point is collected for a hospital patient, the model is effectively conditioned on $H = +h$, presence in the hospital. This is denoted by a double circle in the model.

The second resident in our example suggested that using the prevalence of D observed in the hospital would

not be appropriate because it would double-count the evidence from observing S . In order to demonstrate that the argument behind the resident's reasoning is fallacious, we will first assume that the population of the town in question is distributed as follows:

N	$+d$	$\neg d$	Total
$+s$	972	84	1,056
$\neg s$	532	92,568	93,100
Total	1,504	92,652	94,156

If a patient presenting with S is admitted to the hospital with probability $\Pr(+h|+s) = 0.75$ and a patient not presenting with S is admitted with probability $\Pr(+h|\neg s) = 1/532 = 0.00188$, the frequencies in the database faithfully correspond to this probabilistic model.

The prevalence of D is $\Pr(+d) = 0.01597$, in agreement with the result of the epidemiological study. Nevertheless, sensitivity and specificity of the symptom S in the general population are

$$\begin{aligned} \Pr(+s|+d) &= 972/1,504 = 0.64628 \\ \Pr(\neg s|\neg d) &= 92,568/92,652 = 0.99909, \end{aligned}$$

which is quite different from the sensitivity and specificity among the hospital patients (see Equations 1 and 2). This difference can be attributed purely to the effect of conditioning on the patient population, i.e., looking only at those patients who are in the hospital. Given our assumption that random variation was the only factor explaining presence or absence of the symptom and admission to the hospital, these patients are in every respect identical to individuals in the general population. We can compute $\Pr(+d|+s)$ by applying Bayes theorem or directly from the table:

$$\Pr(+d|+s) = \frac{972}{1,056} = 0.92045. \quad (5)$$

This result differs over an order of magnitude from the value $\Pr(+d|+s) \approx 0.06$ obtained in Equation 3. What is the explanation of this paradox?

The answer is that the frequencies contained in the database do not reflect the probabilities $\Pr(d,s)$ but rather $\Pr(d,s|+h)$. For this reason, Equations 1 and 2 are wrong: they do not represent $\Pr(+s|+d)$ and $\Pr(\neg s|\neg d)$ but rather $\Pr(+s|+d,+h)$ and $\Pr(\neg s|\neg d,+h)$ respectively. If we apply Bayes theorem properly, then

$$\Pr(d|s,h) = \frac{\Pr(s|d,h) \cdot \Pr(d|h)}{\Pr(s|+d,h) \cdot \Pr(+d|h) + \Pr(s|\neg d,h) \cdot \Pr(\neg d|h)}.$$

From the hospital database, we obtain the prevalence of D among the hospital patients $\Pr(+d|+h) =$

730/967 = 0.75491. Hence

$$\Pr(+d|+s,+h) = \frac{0.99863 \cdot 0.75491}{0.99863 \cdot 0.75491 + 0.26582 \cdot 0.24509} = 0.92045. \quad (6)$$

Comparing this result with Equation 5, we verify that $\Pr(+d|+s) = \Pr(+d|+s,+h)$, in agreement with Equation 4.

To better understand the mistake made in Equations 1 and 2, we can profit from the fact that the variables involved in this model are binary and write Bayes theorem in its odds-likelihood form,

$$OR_{\text{post}}(D) = OR_{\text{pre}}(D) \cdot LR_D(s), \quad (7)$$

where $OR_{\text{pre}}(D)$ is the prior odds of D

$$OR_{\text{pre}}(D) = \frac{\Pr(+d)}{\Pr(-d)},$$

$OR_{\text{post}}(D)$ is the posterior odds of D given the evidence $S = s$

$$OR_{\text{post}}(D) = \frac{\Pr(+d|s)}{\Pr(-d|s)},$$

and $LR_D(s)$ is the likelihood ratio for D given $S = s$

$$LR_D(s) = \frac{\Pr(s|+d)}{\Pr(s|-d)}.$$

The probability $\Pr(+d|s)$ can be computed from $OR_{\text{post}}(D)$ by the following formula

$$\Pr(+d|s) = \frac{OR_{\text{post}}(D)}{1 + OR_{\text{post}}(D)}.$$

We also have

$$OR_{\text{post}}^H(D) = OR_{\text{pre}}^H(D) \cdot LR_D^H(s),$$

which only differs from Equation 7 in that all the probabilities involved are conditioned on $+h$, presence in the hospital. It is possible to show that

$$OR_{\text{post}}^H(D) = OR_{\text{post}}(D)$$

(this equation is equivalent to Equation 4), as shown in Table 1.

Nevertheless, if we mix general population probabilities with hospital probabilities, we obtain

$$OR_{\text{post}}^{\text{wrong}}(D) = OR_{\text{pre}}(D) \cdot LR_D^H(s).$$

In particular, when $S = +s$, $OR_{\text{post}}^{\text{wrong}}(D) = 0.06098$ and $\Pr^{\text{wrong}}(+d|+s) = 0.05748$, which is the erroneous result obtained in Equation 3.

	General population	Hospital database
$\Pr(+d)$	0.01597	0.75491
$OR_{\text{pre}}(D)$	0.01623	3.0802
$LR_D(+s)$	712.84	3.7568
$OR_{\text{post}}(D)$	11.571	11.571
$\Pr(+d +s)$	0.92045	0.92045

Table 1: Probabilities and ratios. The quantities in the right column are all conditioned on $+h$.

We can also deduce that

$$\frac{OR_{\text{post}}^{\text{wrong}}(D)}{OR_{\text{post}}(D)} = \frac{OR_{\text{pre}}(D)}{OR_{\text{pre}}^H(D)}.$$

In our example,

$$\frac{OR_{\text{pre}}(D)}{OR_{\text{pre}}^H(D)} = 5.27 \cdot 10^{-3} \approx \frac{1}{200}.$$

Please note that this ratio is the same for $+s$ and $-s$. It means that if the hospital prevalence is much higher than the prevalence in the general population, then the posterior probability due to a wrong combination of probabilities will be much smaller than the true probability, as shown in the example above.

This ratio is also equal to

$$\begin{aligned} \frac{OR_{\text{post}}^{\text{wrong}}(D)}{OR_{\text{post}}(D)} &= \frac{\Pr(+d)/\Pr(-d)}{\Pr(+d|+h)/\Pr(-d|+h)} \\ &= \frac{\Pr(+h|-d)}{\Pr(+h|+d)} = \frac{1}{LR_D(+h)}. \end{aligned}$$

Therefore, the key parameter affecting the magnitude of the error is $LR_D(+h)$, i.e., the proportion between $\Pr(+h|+d)$ and $\Pr(+h|-d)$.

It follows from this equation that

$$\begin{aligned} \Pr^{\text{wrong}}(+d|s) &= \frac{\Pr(+d|s)}{\Pr(+d|s) + (1 - \Pr(+d|s)) \cdot LR_D(+h)} \end{aligned}$$

or, equivalently,

$$\begin{aligned} \frac{\Pr^{\text{wrong}}(+d|s)}{\Pr(+d|s)} &= (\Pr(+d|s) + (1 - \Pr(+d|s)) \cdot LR_D(+h))^{-1}. \end{aligned}$$

In summary, when building a model for this problem, we can take the values of prevalence, sensitivity, and specificity either from the general population or from the hospital data. In both cases, the model will predict correctly the posterior probability of disease D .

If we mix data from these two sources, however, the model will be incorrect. The assertion that combining data from two different sources is dangerous may seem trivial. Nevertheless, when building probabilistic models, knowledge engineers tend to concentrate on choosing the right prevalence, because they are aware that it is strongly population-dependent. They often do not realize that sensitivity and specificity are also population-dependent. In fact, it is quite normal for medical literature to offer values of sensitivity and specificity without an explanation of how they were obtained, because they are assumed to be invariant. However, selection of sensitivity and specificity deserves as much attention as selection of prevalence. In principle, there is no wrong choice of prevalence as far as the sensitivity and specificity are chosen coherently.

4 Extending the Model

The problem observed above can be explained in terms of conditional independence. Conditional independence is captured in causal networks by a property of the interaction between causality and probability known as *causal Markov condition* [Spirtes *et al.*, 1993]. This is related to a condition known as *d-separation* [Geiger *et al.*, 1990], which originates from the pioneering work of Dawid [1979] in the domain of statistics. We will discuss here two other similar cases where the role of conditional independence becomes apparent.

In our example,

$$\Pr(d|s) = \Pr(d|s, h) .$$

Therefore, $\Pr(d|s)$ could be computed from population data,

$$\Pr(d|s) = \alpha \cdot \Pr(s|d) \cdot \Pr(d) ,$$

or from hospital data,

$$\begin{aligned} \Pr(d|s) &= \Pr(d|s, +h) \\ &= \alpha \cdot \Pr(s|d, +h) \cdot \Pr(d|+h) . \end{aligned}$$

That is, we can remove node H and replace each population probability by the corresponding hospital probability. From the point of view of diagnosis, both models will be equivalent.

However, when the model is extended to include more variables, we must be more careful. As an example, let us consider two symptoms, S_1 and S_2 , of a certain disease D , such that both S_1 and S_2 may make the patient go to the hospital (Figure 3).

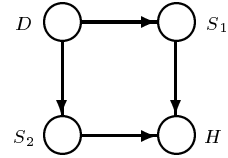


Figure 3: Example model with two symptoms.

Also in this case,

$$\Pr(d|s_1, s_2, +h) = \Pr(d|s_1, s_2) , \tag{8}$$

which means that when we know whether each symptom is present or absent it is not necessary to take into account the fact that the patient is at the hospital, because it does not affect the diagnosis. However, in this case we cannot remove node H and replace each probability by the corresponding probability conditioned on $+h$, as we did in the previous example, because in general

$$\begin{aligned} \Pr(d|s_1, s_2, +h) \\ \neq \alpha \cdot \Pr(d|+h) \cdot \Pr(s_1|d, +h) \cdot \Pr(s_2|d, +h) . \end{aligned}$$

The reason is that when the value of H is known, S_1 and S_2 are no longer conditionally independent given D , i.e.,

$$\Pr(s_1, s_2|d, +h) \neq \Pr(s_1|d, +h) \cdot \Pr(s_2|d, +h) .$$

Therefore, in this case, a model that uses hospital data but does not explicitly represent the variable H , is wrong. It will compute the probability $\Pr(d|s_1, s_2)$ as if the symptoms were conditionally independent given the disease. On the other hand, if the model explicitly includes H , as shown in Figure 3, it then requires four probability distributions: $\Pr(d)$, $\Pr(s_1|d)$, $\Pr(s_2|d)$, and $\Pr(h|s_1, s_2)$, the last of which cannot be extracted from the hospital database. In this case it is simply not possible to build an accurate causal model based only on the hospital data. Equation 8 suggests that it may be not worth to include H in the model.

To counterbalance this discouraging result, we will show in the next example that in some cases it is possible to combine data from different sources into a single model. Let us go back to the first example and assume that, in addition to a symptom S , there is also a test T performed at the hospital. Again, variable H indicates whether the patient has been admitted to the hospital (Figure 4). Please note that there is no link $T \rightarrow H$ because in this example the value of T does not affect the probability that the patient is admitted to the hospital and, hence, is included in the database. Therefore, $\Pr(t|d, h) = \Pr(t|d)$, which is equivalent to

saying that the sensitivity and specificity of T with respect to D do not depend on H .

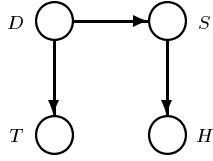


Figure 4: Example model with a symptom and a test.

As a consequence of the structure of the graph, S and T are conditionally independent given D , even when the value of H is known, i.e.,

$$\Pr(s, t|d, h) = \Pr(s|d, h) \cdot \Pr(t|d),$$

which implies that

$$\begin{aligned} \Pr(d|s, t, h) \\ = \alpha \cdot \Pr(d|h) \cdot \Pr(s|d, h) \cdot \Pr(t|d). \end{aligned}$$

On the other hand, as D and H are conditionally independent given S and T , we have

$$\begin{aligned} \Pr(d|s, t, h) = \Pr(d|s, t) \\ = \alpha \cdot \Pr(d) \cdot \Pr(s|d) \cdot \Pr(t|d), \end{aligned}$$

The comparison of these two equations shows that in this case it is possible to build a model with only three variables, D , S and T and it is not necessary to explicitly represent H , even if the model is to be used at the hospital. Furthermore, the population probabilities $\Pr(d)$ and $\Pr(s|d)$ can be safely replaced by their corresponding hospital probabilities $\Pr(d|h)$ and $\Pr(s|d, h)$ — provided that both of them are replaced simultaneously. Also, the absence of a causal link $T \rightarrow H$ implies that $\Pr(t|d) = \Pr(t|d, h)$, i.e., the sensitivity and specificity of T are the same in the general population and at the hospital, which implies that the values obtained from a hospital database can be safely combined with those of the general population. Of course, if the result of the test may increase or decrease the probability that the patient is included in the database, our graphical model should contain a link $T \rightarrow H$, and the treatment of the problem would be the same as in the example in which there were two symptoms.

In summary, we have analyzed three examples. In the first one, it was not necessary to explicitly represent the fact that the patient is at the hospital once we know whether the patient presents with H . The model can be built by using either $\Pr(d)$ and $\Pr(s|d)$ or $\Pr(d|h)$ and $\Pr(s|d, h)$, but not a combination of them. In the second example, it was not possible to

build an accurate causal model based on the hospital data — it was necessary to use population data, which may be harder to obtain. In the third example, we assumed that the result of T did not lead to a selection bias. As a consequence, its sensitivity and specificity could be obtained from either the population or from the hospital database and safely combined with any other source.

In these examples, the variable that led to selection biases was the fact that the patient was admitted to the hospital. Other variable that might lead to a similar bias in some cases is the fact that the patient is alive. The conclusion we extract from this section is that, contrary to the usual practice in knowledge engineering, such variables may not be ignored. When building a Bayesian network, the causal graph must explicitly represent them. The causal Markov condition will help the knowledge engineer determine whether some of those variables can be removed from the graph, provided that the conditional probabilities of their ancestors are coherently chosen. In contrast, when a node is not an ancestor of any of those selection variables, its conditional probability is invariant and can be obtained from any source.

5 Discussion

Knowledge engineers quantifying probabilistic models usually combine various sources of information, such as existing textbook knowledge, available statistics, databases of cases, and expert judgments. In this paper we focused on dangers that this practice is facing. We have demonstrated that lack of attention to whether the sources are compatible and whether they can be combined can lead to erroneous behavior of the model. We studied the problem on three simple examples and offered guidelines for detecting when different sources of data can be safely combined. We based our explanation on the concept of probabilistic conditioning with respect to the underlying causal model of the domain. We have shown that data that seemingly capture the properties of the same population of individuals can imply different numerical properties of the population, depending on how they were collected.

Often we perceive such parameters as sensitivity and specificity as fairly population-independent. After all, sensitivity and specificity do not depend on the prevalence. We realize that different population characteristics, such as sex, race, diet, etc., may influence both sensitivity and specificity, but we forget about purely statistical phenomena such as conditioning. Please note that in our motivating example, the population was identical and we could attribute presence in the hospital to random variation. It was not the special

characteristics of the hospital patients that made them develop the symptom more or less likely than the general population.

An important conclusion that flows from our paper is that it is important to study the graphical part of the model under construction, reflect on the selection criteria that were applied in collecting each source of data, model them explicitly, and investigate whether the variables that the data is conditioned on can potentially bias the latter and affect the model.

Our motivating example was based on a medical data set, but the same argument can be made with respect to numbers obtained from human experts. Subjective probability judgments have been shown to rely on judgmental heuristics [Kahneman *et al.*, 1982] and they are very sensitive to prior experiences (in fact prior experiences are often all that probability judgments are based on). Humans have been shown to be able to match the probability of observed events with an amazing precision [Estes, 1976]. Physicians working in a hospital will tend to match the sensitivity and specificity of medical symptoms and tests that they observe in their practice. These are often determined by the circumstances, such as what brought the patients to the hospital or clinic in the first place. Physician experts will tend to at least adjust the parameters to what they observe in their practice. While their experience is valuable for building decision models for the particular clinics where they have worked, in general they cannot be readily used in other settings. Our recommendation is that the knowledge engineer warn the expert about the need for a careful choice of population for each probability estimate.

It seems to us that the incompatibility of various sources of information is more prevalent than most knowledge engineers realize. In every model, there are unmodeled factors, which, if conditioned upon in data collection, make data sets biased. It has been known that in some cases conditioning affects qualitative, structural properties of models, such as probabilistic independence, but to our knowledge no attention has been paid to its impact on model building through incompatibility of such seemingly robust local properties, as conditional probabilities. We hope that this paper will make knowledge engineers aware of the problem and prepare them better for dealing with it in practice.

Acknowledgments

The first author was supported in part by the National Science Foundation under Faculty Early Career Development (CAREER) Program, grant IRI-9624629, and by the Air Force Office of Scientific Research un-

der grants F49620-97-1-0225 and F49620-00-1-0112. The second author was supported by the Spanish CI-CYT, under grant TIC97-1135-C04. Our collaboration was enhanced by travel support from NATO Collaborative Linkage Grant number PST.CLG.976167.

References

- [Dawid, 1979] A. Philip Dawid. Conditional independence in statistical theory. *Journal of the Royal Statistical Society, Series B (Methodological)*, 41:1-31, 1979.
- [Díez *et al.*, 1997] F. J. Díez, J. Mira, E. Iturralde, and S. Zubillaga. DIAVAL, a Bayesian expert system for echocardiography. *Artificial Intelligence in Medicine*, 10:59-73, 1997.
- [Estes, 1976] W.K. Estes. The cognitive side of probability learning. *Psychological Review*, 83(1):37-64, January 1976.
- [Geiger *et al.*, 1990] Dan Geiger, Thomas S. Verma, and Judea Pearl. Identifying independence in Bayesian networks. *Networks*, 20(5):507-534, August 1990.
- [Henrion *et al.*, 1991] Max Henrion, John S. Breese, and Eric J. Horvitz. Decision Analysis and Expert Systems. *AI Magazine*, 12(4):64-91, Winter 1991.
- [Howard and Matheson, 1984] Ronald A. Howard and James E. Matheson. Influence diagrams. In Ronald A. Howard and James E. Matheson, editors, *The Principles and Applications of Decision Analysis*, pages 719-762. Strategic Decisions Group, Menlo Park, CA, 1984.
- [Kahneman *et al.*, 1982] Daniel Kahneman, Paul Slovic, and Amos Tversky, editors. *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge University Press, Cambridge, 1982.
- [Oniśko *et al.*, 1999] Agnieszka Oniśko, Marek J. Druzdzel, and Hanna Wasyluk. A Bayesian network model for diagnosis of liver disorders. In *Proceedings of the Eleventh Conference on Biocybernetics and Biomedical Engineering*, volume 2, pages 842-846, Warszawa, Poland, December 2-4 1999.
- [Pearl, 1988] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers, Inc., San Mateo, CA, 1988.
- [Spirtes *et al.*, 1993] Peter Spirtes, Clark Glymour, and Richard Scheines. *Causation, Prediction, and Search*. Springer Verlag, New York, 1993.