quires the graph to be chordal for there to be equivalence, whereas Theorem 1+ puts no requirements on it. Theorem 2* requires the hypergraph to be acyclic for there to be equivalence, whereas Theorem 2+ requires only that it be conformal. Theorem 3* requires the set of conditional independencies to have a conflict-free cover for there to be equivalence, whereas Theorem 3+ puts no requirements on it (actually, the closure with respect to strictly positive distributions of a set of conditional independencies is always graph-generated).

As far as I know, Theorems 1*, 2* and 3* are new, although, by now, they are probably not unexpected.

Parallel developments in the two fields have occurred in the past, with neither aware of the other, apparently. For example, Vorob'ev's (1962) results on extending consistent marginal distributions parallel similar results for the extension of consistent databases (Beeri et al., 1983). And Beeri and Kifer's (1986a, 1986b, 1987) work on fixing sets of multivalued dependencies that have intersection anomalies parallels Dawid's (1979b) method for fixing up sets of conditional independencies.

3. MODELS AND DATA

Two simple but important points, each mentioned in both papers and neither having to do directly with graph theory, deserve to be emphasized. First, both papers take the position that a model represents the substantive knowledge that an expert brings to the problem prior to seeing specifically relevant data. One practical consequence of such a position is that statisticians cannot work in a vacuum; rather, they must interact and communicate effectively with domain specialists. And, on a more philosophical note, this position highlights the fact that a scientifically meaningful model for the data is as much a subjective prior assessment of the relative likelihood of possible values as is a scientifically meaningful model for the parameters of such a model. Second, SDLC stress and CW mention that observed data allow us not only to estimate parameters in the model but also to monitor and, if need be, to critique the model. It is refreshing to see frequentists concerned about representing expert knowledge and Bayesians worried about model criticism.

4. SOME QUESTIONS FOR THE AUTHORS

Can you have discrete variables in chain graphs with dashed edges? Can you explain why the diagnostic ability of the Bayesian network was not as good as that of the CART-like algorithm? From Table 6, it appears that for 110 cases (of 168) the Bayesian network assigned the correct diagnosis the highest probability; what were the ranks of the correct diagnoses for the other 58 cases? Has anyone created Bayesian networks with both discrete and continuous variables? Of course, with mixed models the number of parameters in each distribution will not stay fixed after updating. Has anyone considered creating a "Bayesian chip" that could be used to create truly parallel "Bayesian machines"?

Reading and thinking about these papers has been a real pleasure.

Comment: What's Next?

David Madigan

These papers represent two of the many different graphical modeling camps that have emerged from a flurry of activity in the past decade. The paper by Cox and Wermuth falls within the statistical graphical modeling camp and provides a useful generalization of that body of work. There is, of course, a price to be paid for this generality, namely that the interpretation of the graphs is more complex. I cannot resist complementing the authors on the remarkable feat of finding an example for each of the different graphical models they propose.

The paper by Spiegelhalter, Dawid, Lauritzen and Cowell falls within the probabilistic expert system camp. This is a tour de force by researchers responsible for much of the astonishing progress in this area. Ten years ago, probabilistic models were shunned by the artificial intelligence community. That they are now widely accepted and used is due in large measure to the insights and efforts of the authors, along with other pioneers such as Judea Pearl and Peter Cheeseman.

I will confine my remaining comments to the Spiegelhalter et al. paper and explore some open questions that I believe will rapidly become important, now that

David Madigan is Assistant Professor, Department of Statistics, GN-22, University of Washington, Seattle, Washington 98195.

many basic technical issues are being successfully solved.

WHAT CAN YOU DO WITH A GRAPHICAL MODEL?

My primary concern is with the apparent mismatch between the informal, qualitative character of human reasoning and the rigorous, formal, quantitative approach of graphical models (Henrion, Breese and Horvitz, 1991). Knowledge-based system builders now have access to knowledge representation tools of considerable expressive power and flexibility (e.g., Skuce, 1991) while the poor graphical modeler has to make do with nodes, links and probability distributions. These concerns are practically motivated. At the University of Washington we are constructing an intelligent tutoring system (ITS) for basic statistics. At the heart of any ITS is an explicit model of the student's knowledge. Acknowledging the inherent uncertainty, we use a Bayesian graphical model for this purpose. However, a second ITS component concerns instructional strategy—the procedural knowledge of experienced teachers. Graphical models fail dismally to represent this knowledge, yet a simple rule-based system does a reasonable job. In a project at the Fred Hutchinson Cancer Research Center in Seattle, we are constructing a knowledge-based system to assist nurses who handle telephone calls from bone marrow transplant patients and their physicians (Bradshaw et al., 1993). Graphical models can calculate the probabilities of various complications, but cannot represent the heuristic knowledge of experienced nurses as they manage the call. In general, the range of potential applications for graphical models is considerably smaller than for knowledge-based systems.

There may be a way out of this dilemma: a number of authors have suggested combining conventional knowledge-based systems with probabilistic models. The key to the success of such hybrid systems is that each component contributes to the portion of the process that it does best: the knowledge-based components guide the interaction by using rough rules-of-thumb that can help to quickly scope, categorize, gather information about, structure and interpret important aspects of the problem; the probabilistic components rely on carefully crafted assessments of uncertainty to provide specific answers about particular situations in a rigorous manner (Bradshaw et al., 1993; Szolovits and Pauker, 1978). Control rests with the knowledge-based component, which calls the probabilistic component as required.

Closely related to this is the emerging area of "knowledge-based model construction" (KBMC). The effective application of belief network tools requires a relatively high level of modeling sophistication, and model construction has proven to be a serious bottleneck. These tools contain some of the algorithms of probabilistic modeling, but cannot embody the experience and intuition of the skilled modeler. KBMC seeks to combine probabilistic modeling tools (including belief networks and influence diagrams) with a knowledge-based system that helps domain experts without extensive training in probabilistic modeling to build, evaluate and refine probabilistic models (Breese, 1989; Goldman and Breese, 1992; Holtzman, 1989). For complex problem domains, sharing and re-use of model components is vital: the knowledge base could dynamically assemble a probabilistic model, tailored to the problem at hand, from model fragments (Almond, Bradshaw and Madigan, 1993). Notable applications of KBMC technology include the Boeing Company's DDUCKS tool, a knowledge-based influence diagram workbench (Bradshaw et al., 1991) and the text understanding application of Goldman and Charniak (1992).

In short, it seems likely that in the future, graphical models will not exist as stand-alone applications, but rather will be embedded in larger systems, encompassing a variety of knowledge bases, databases and models.

MODEL UNCERTAINTY

An alternative to KBMC is to automatically induce models from existing databases. This is discussed by the authors in subsection 5.4. They begin by stating that "An approach that takes model comparison to its full consequence is to induce the network directly from data ... ignoring the prior structural and quantitative information available." Why does the "full consequence" involve the absence of prior information? One of the great advantages of the Bayesian graphical model approach is that prior knowledge, both structural and quantitative, can *realistically* be elicited and incorporated into both model selection and subsequent inference (Madigan and York, 1993). Indeed, with even a modest number of nodes, the graphical model space is vast, and there is a concern that in the absence of *some* prior knowledge, model selection procedures may fail (Draper, 1993).

Historically, model selection procedures have focused on finding the single "best" model. However, this ignores model uncertainty, leading to poorly calibrated predictions: it will often be seen in retrospect that one's uncertainty bands were not wide enough (Draper, 1993). A Bayesian solution to this problem involves averaging over all plausible models when making inferences about quantities of interest (see, for example, Raftery, 1988, and Kass and Raftery, 1993). Indeed Hodges (1987) comments that "what is clear is that when the time comes for betting on what the future holds, one's uncertainty about that future should be fully represented, and model [averaging] is the only tool around." In many applications, however, because of the size of the model space and awkward integrals, this averaging will not be a practical proposition, and approximations are required. Draper (1993) describes "model expansion": averaging over all plausible models in the neighborhood of a "good" model. Madigan and Raftery (1991) describe an approach for Bayesian graphical models that involves seeking out the most plausible models and averaging over them. Raftery (1993) applies this to structural equation models. Madigan and York (1993) suggest a Markov Chain Monte Carlo approach that provides a workable approximation to the complete solution. These methods can also be applied to incomplete data (Madigan and Kong, in preparation). The point is that with Bayesian graphical models, correctly accounting for model uncertainty is entirely possible.

Model averaging in the context of expert systems raises special problems: displaying multiple models requires careful software design; enhanced explanation facilities are required; software for model prior elicitation is needed. The issue of compatible priors in alternative models, addressed by the authors in Section 8, is of considerable importance. While the procedure suggested seems reasonable, a more general framework is required. Certainly, when precisely specified probabilities are involved, the procedure should be used with extreme caution.

INTERCAMP COMMUNICATION

Other (independence) graphical modeling camps are to be found within decision analysis, philosophy of science and statistics. Several different camps are located in computer science. To date, these camps have communicated remarkably effectively with each other, fostering rapid progress. The challenge we face is to maintain the communication. The gulf between the two papers here demonstrates both the diversity of the progress and the extent of the challenge.

ACKNOWLEDGMENTS

I am indebted to Russell Almond, Adrian Raftery, Jeremy York and especially Jeff Bradshaw and David Draper for helpful discussions. This work was supported in part by a grant from the NSF.

Comment

Sharon-Lise Normand

1. INTRODUCTION

The authors of these two highly complementary articles are to be congratulated on their timely contributions to the readership of *Statistical Science* and to statisticians in general. The article by Spiegelhalter and colleagues provides a comprehensive review of the most recent *statistical* developments in expert systems, guiding us through a complete analysis in the expert system domain. Cox and Wermuth present a pointed discussion on the interpretation and graphical representation of linear dependencies for continuous valued random variables. In this discussion I will expand upon the range of applications of graphical models and emphasize some specific areas discussed by the authors. Specifically, my comments will address (1) the role of graphical models in statistical inference, (2) data propagation in graphs and (3) limitations of graphical models.

Sharon-Lise Normand is Assistant Professor of Biostatistics, Department of Health Care Policy, Harvard Medical School, 25 Shattuck Street, Parcel B, 1st Floor, Boston, Massachusetts 02115.

2. THE ROLE OF GRAPHICAL MODELS

Graphical models can play an important role in structuring statistical analyses, in performing complicated computations and in communicating results. Thus the motivation for creating a graphical representation of a statistical model is threefold: (1) the graph provides an effective vehicle for communication among researchers, (2) the graph displays a knowledge map of the dependency structure posited in the model and finally (3) the graph can be transformed into a static secondary structure that can be used for efficient probability calculations. Professor Spiegelhalter and his colleagues touch on all three reasons with emphasis placed on calculating probabilities while Professors Cox and Wermuth stress the value of the graph as a knowledge map. It is particularly important to note that one may choose to exploit any or all three reasons for using a graphical model.

The term *graphical model* has a very precise definition in the contingency table literature (Darroch, Laurit-