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-

zen and Speed, 1980; Edwards and Kreiner, 1983; Wermuth and Lauritzen, 1983). In this discussion I will, however, use the term more generally to refer to statistical models that host some conditional independence properties. Hierarchical models (Lindley and Smith, 1972; Morris, 1987) are a class of statistical models that immediately come to my mind when discussing graphical models. Inherent in hierarchical models is the notion of conditional independence across observations at one stage and across parameters at another stage. Consider for example a two-stage normal hierarchical model used to combine information across experiments. The observed data will consist of a summary measure from each experiment, y_i, and an associated measure of precision, V_i. In a random effects model, it is assumed that for $i = 1, 2, \ldots, k$ studies

$$y_i|\theta_i \overset{\text{indep.}}{\sim} \mathrm{N}(\theta_i, V_i), \tag{1}$$

$$\theta_i|\mu,\tau^2 \overset{\text{indep.}}{\sim} \mathrm{N}(\mu,\tau^2), \tag{2}$$

where θ_i represents the underlying study effect for the ith experiment and μ and τ^2 are the hyperparameters of the mixing distribution governing the generation of each underlying study effect. The directed graph corresponding to this model will have k separate nodes for each summary measure, k separate nodes for each underlying study effect and a node for each of the hyperparameters. Unlike the CHILD network discussed by Spiegelhalter and colleagues (Figure 2 in their article) and the examples considered in Cox and Wermuth's paper, only a subset of the nodes in the graph representing this hierarchical model will ever be observed. Substantially more complicated hierarchical models, those with more stages and more dependency structure such as the multiprocess models of Harrison and Stevens (1976), can be represented graphically.

The value of displaying the qualitative structure of statistical models has been vastly underutilized by statisticians but appreciated in other branches of science. In the medical arena, we frequently encounter graphical representations of decision models, namely *decision trees*. In its simplest form, the decision tree is a singly connected graph in which some nodes represent risk factors such as age and gender, some nodes represent complications and symptoms and some nodes represent decisions. For example, researchers may be interested in investigating whether older patients who are suspected of having an acute myocardial infarction will benefit from thrombolytic therapy. A decision-analytic model is then built using information from the experts (cardiologists) and from the results of clinical trials (e.g., the rate of incapacitating complications from thrombolytic therapy for older patients). Some statisticians are investigating methods of quantifying uncertainty in medical decision analysis (Katz and Hui, 1989) because, typically, statistical error is not incorporated in most decision analyses. Clearly, the expert system methodology could play a substantial role in this effort—propagation of the uncertainty attached to the decision tree inputs is naturally accommodated within the graphical framework.

More recently we have witnessed in the statistical literature the use of graphical representations to understand the dependency structure in order to perform the "correct" computations. For example, Bernardinelli and Montomoli (1992) use a graphical representation of a hierarchical model of relative risk mortality to display the qualitative structure of the data but also to indicate which conditional distributions must be specified to calculate the joint distribution. Gilks et al. (1993) construct a graphical model for modeling precursors of cervical cancer in an application of Gibbs sampling in medicine for a similar specification purpose.

Finally, as Professor Spiegelhalter and his colleagues have indicated, the graphical model can be used to perform efficient probability calculations in high dimensional problems. The main goal is to have queries regarding certain sets of variables answered quickly. This is achieved through local computations performed through an algorithm designed to capitalize on the dependency structure embedded in the statistical model. In the expert system setting, the computational efficiency of the propagation algorithm is obvious. However, it has been shown that even in standard models, computation within a graphical framework can be beneficial. Normand and Tritchler (1992) discuss the use of a graphical model as the computational device for updating parameter estimates in a hierarchical model and show that the graphical model characterizes the hierarchical model and its computations in a unified way.

3. DATA PROPAGATION IN GRAPHS

Because one of the central roles for the expert system is that of updating the system once evidence has been realized, I will recast for the reader the essence of how this is achieved. The task at hand is the following: information is observed and consequently, the joint distribution needs to be updated in light of this new information. Essentially the problem becomes one of conditioning and a brute force approach is clearly undesirable in high dimensional problems. It is worth recalling that there were over a billion possible configurations in the CHILD network. Propagation refers to the transmission of hereditary features to or through offspring and this is the "divide and conquer" strategy employed in graphical models: the joint set of random variables is divided into subgroups, a source subgroup is identified, a marginal is taken in the source subgroup

and then that marginal multiplies a function on a destination subgroup. A marginal is then taken in the destination subgroup and that marginal multiplies a function on its destination subgroup and so on and so on. Professor Spiegelhalter and his colleagues refer to the subgroups as belief universes and equate these universes to the cliques of the relevant undirected graph. A clique is a set of random variables such that no further factorization of the probability function characterizing the distribution is possible; that is, there are no further independence constraints on the elements in the clique. Ideally the state space of the cliques should be small otherwise the efficiency gains through the use of the propagation algorithm will be lost. The propagation algorithm described in the article by Professor Spiegelhalter and colleagues is based on the junction tree. The junction tree may be thought of as a singly connected graph in which each node consists of sets of random variables (the *cliques*). In the case of multinomial random variables, any node in the junction tree may be used as a root for propagation. The steps necessary to transform the original directed graph into the junction tree (referred to as compilation by Professor Spiegelhalter and his colleagues) are many and sometimes nontrivial.

4. LIMITATIONS OF GRAPHICAL MODELS

The (potential) limitations of graphical models that I envision are related only to one of my three motivating reasons for using graphical models, and these have to do with efficient computation. First, Professor Spiegelhalter and his colleagues have indicated the importance of the size of the state spaces of the cliques obtained after triangulation in measuring the computational benefits of a graphical model approach. In preserving all the induced dependence relationships in a model through the "moralizing" procedure, the dimensionalities of the cliques are increased. These dimensionalities are further increased after triangulation. It is not immediately clear in which statistical models the computational advantages of a graphical model approach will be realized. Further research into identifying classes of statistical models that could benefit from the computational efficiency of graphical models needs to be undertaken.

Second, the junction tree algorithm for propagation in graphical models works well in models in which the random variables arise from a multinomial distribution. There does exist an algorithm that mimics the junction tree algorithm in models for which some of the variables are multinomial and some are Gaussian. However in these latter graphs (mixed graphs), only means and variances are propagated. Moreover, there is an additional requirement in the compilation process for marked graphs, that of strong decomposability, that further increases the dimensionalities of the cliques. In addition, for graphs that host other distributions, Monte Carlo methods have to replace exact methods.

Third, I am not satisfied with how well one can assess model fit in graphical models. I have a difficult time assessing model adequacy in a logistic regression model with more than five covariates! I will not equate my model-checking capabilities to those of Professor Spiegelhalter and his colleagues but surely model assessment involving the number of variables typified in the expert system domain requires a tremendous amount of skill. In the CHILD network, how can one assess whether age at presentation is related quadratically to disease or whether age at presentation is related quadratically to lung disease but only linearly to the remaining five diseases? How important is correct specification of the functional form of the model variables and how important are "missing links" in predicting the state of a particular configuration? The node monitors proposed by Professor Spiegelhalter and colleagues are admittedly using a prequential approach but I hope research will extend to model diagnostic methods using other endpoints.

Graphical models will play an increasingly important role in the structuring of statistical analyses for complex problems. These models enhance communication among researchers, thereby facilitating scientific modeling, and provide a unifying approach to computation. Research into automating algorithms for distributions other than the multinomial and the Gaussian distributions should be explored. More examples of graphical models need to identified and analyzed, and the effects of model misspecification on prediction quantified. In closing, I thank the authors for presenting their valuable ideas.

ACKNOWLEDGMENT

This research was partially supported by AHCPR Grant RO1 HS07118-01.