Comment

Joe R. Hill

The authors of these two papers are among the most active nodes in an ever growing hypergraph of interesting papers on statistical applications of graph theory. It is an honor to discuss these two new hyperedges.

My discussion is divided into four parts. Section 1 discusses statistical applications of graph theory. Section 2 briefly describes ways of leveraging parallels between probability and database theory. Section 3 highlights two important points made in each of the papers. Finally, Section 4 asks some specific questions.

1. STATISTICAL APPLICATIONS OF GRAPH THEORY

Graph theory has a lot to offer statisticians. Consequently, graph theory is quickly becoming an integral part of modern statistics. Graphs, both directed and undirected, and hypergraphs can be used to (a) represent qualitative multivariate relationships, (b) specify and visualize multivariate statistical models, (c) determine statistical properties of multivariate models and (d) develop computationally efficient algorithms for dealing with large multivariate models. The first two of these contribute to effective communication between applications experts and statisticians. The third helps statisticians develop appropriate statistical theory. The fourth makes computing feasible for more complicated problems.

Graphical models provide a flexible paradigm for describing multivariate statistical models. They can have *discrete* variables (as in Bayesian networks, graphical and recursive loglinear models for contingency tables, and influence diagrams for applied decision analysis), or *continuous* variables (as in covariance selection and structural equation models). Conditional Gaussian models (Lauritzen and Wermuth, 1989; Wermuth and Lauritzen, 1990) provide a framework for having both kinds of variables in a single graphical model. Graphical models can have *directed* edges (as in Bayesian networks, influence diagrams and regression models) or *undirected edges* (as in graphical and decomposable loglinear models, covariance selection models and Markov random field models for image restoration). Chain graphs provide a framework for having both kinds of edges in a single graphical model.

In their paper, Cox and Wermuth (CW) introduce, for multivariate normal models, the concept of *dashed* edges as a way to represent constraints on covariance matrices (i.e., to represent marginal independencies), complementing the use of *full* edges to represent constraints on concentration matrices (i.e., to represent conditional independencies). They illustrate the use of the new enriched class of models with a number of empirical examples.

Spiegelhalter, Dawid, Lauritzen and Cowell (SDLC) give a status report on their ongoing development of Bayesian networks for expert systems. They have carefully combined a number of methods. They elicit Bayesian graphical models from medical experts. They use graphical ideas to convert the model into a computationally efficient form. They apply Bayesian estimation techniques to "learn" probability parameters as additional data are observed, and they use significance testing methods to monitor and critique the model.

SDLC provide an effective method for eliciting the qualitative, the probabilistic and the initial quantitative aspects of an expert-defined model. The key to their method is to use a directed acyclic graph to represent the qualitative relationships between variables. Nearly everything else follows from this graph.

This graph determines a recursive factorization of the joint distribution with, for each variable, a factor that is the conditional distribution of that variable, given its parents. This representation of the joint distribution has two advantages. First, the number of probabilities that the expert has to specify is considerably less than for a general joint distribution that does not encode the implied conditional independencies as efficiently. Second, these probabilities are "easy" for an expert to specify for three reasons: (a) the expert has to think about the distribution of only one variable at a time, (b) each distribution is conditioned on the parents of the variable, which are the variables that directly influence it and (c) the conditioning events can be thought of as fixed scenarios. In short, it is easy for an expert to think about the probability distribution of a single "effect" given its immediate "causes." This second advantage contrasts sharply with the problems associated with directly specifying an overall joint distribution. In that case, the expert would not be able to think conditionally but would have to think in multiple dimensions simultaneously and would typically have to specify many very small probabilities.

Once the model has been specified, it is converted to a junction tree representation for efficient computation. This conversion is carried out in a series of steps

Joe R. Hill is R & D Manager, EDS Research, 5951 Jefferson St. NE, Albuquerque, New Mexico 87109.