Comment: Graphical Models, Causality and Intervention

Judea Pearl

I am grateful for the opportunity to respond to these two excellent papers. Although graphical models are intuitively compelling for conceptualizing statistical associations, the scientific community generally views such models with hesitancy and suspicion. The two papers before us demonstrate the use of graphs – specifically, directed acyclic graphs (DAGs) – as a mathematical tool of great versatility and thus promise to make graphical languages more common in statistical analysis. In fact, I find my own views in such close agreement with those of the authors that any attempt on my part to comment directly on their work would amount to sheer repetition. Instead, as the editor suggested, I would like to provide a personal perspective on current and future developments in the areas of graphical and causal modeling. A complementary account of the evolution of belief networks is given in Pearl (1993a).

I will focus on the connection between graphical models and the notion of causality in statistical analysis. This connection has been treated very cautiously in the papers before us. In Lauritzen and Spiegelhalter (1988), the graphs were called "causal networks," for which the authors were criticized; they have agreed to refrain from using the word "causal." In the current paper, Spiegelhalter et al. deemphasize the causal interpretation of the arcs in favor of the "irrelevance" interpretation. I think this retreat is regrettable for two reasons: first, causal associations are the primary source of judgments about irrelevance, and, second, rejecting the causal interpretation of arcs prevents us from using graphical models for making legitimate predictions about the effect of actions. Such predictions are indispensable in applications such as treatment management and policy analysis. I would like to supplement the discussion with an account of how causal models and graphical models are related.

It is generally accepted that, because they provide information about the dynamics of the system under study, causal models, regardless of how they are discovered or tested, are more useful than associational models. In other words, whereas the joint distribution tells us how probable events are and how probabilities would change with subsequent observations, the causal model also tells us how these probabilities would change as a result of external interventions in the system. For this reason, causal models (or "structural models" as they are often called) have been the target of relentless scientific pursuit and, at the same time, the center of much controversy and speculation. What I would like to discuss in this commentary is how complex information about external interventions can be organized and represented graphically and, conversely, how the graphical representation can be used to facilitate quantitative predictions of the effects of interventions.

The basic idea goes back to Simon (1977) and is stated succinctly in his foreword to Glymour et al. (1987): "The advantage of representing the system by structural equations that describe the direct causal mechanisms is that if we obtain some knowledge that one or more of these mechanisms has been altered, we can use the remaining equations to predict the consequences – the new equilibrium." Here, by "mechanism" Simon means any stable relationship between two or more variables that remains invariant to external influences until it falls directly under such influences.

This mechanism-based model was adapted in Pearl and Verma (1991) for defining probabilistic causal theories; each child-parent family in a DAG Γ represents a deterministic function $X_i = f_i(\mathbf{pa}_i, \varepsilon_i)$, where $\mathbf{pa}_i$ are the parents of variable X_i in Γ, and ε_i, $0 < i < n$, are mutually independent, arbitrarily distributed random disturbances. Characterizing each child-parent relationship as a deterministic function, instead of the usual conditional probability $P(x_i \mid \mathbf{pa}_i)$, imposes equivalent independence constraints on the resulting distributions and leads to the same recursive decomposition

$$(1) \qquad P(x_1, \ldots, x_n) = \prod_i P(x_i \mid \mathbf{pa}_i)$$

that appears in Eq. (1) of Spiegelhalter et al.'s article. However, the functional characterization also specifies how the resulting distribution would change in response to external interventions, since, by convention, each function is presumed to remain constant unless specifically altered. This formulation is merely a nonlinear generalization of the usual structural equation models, where function constancy (or stability) is implicitly

Judea Pearl is Professor of Computer Science and Director of the Cognitive Systems Laboratory, University of California Los Angeles, 405 Hilgard Avenue, Los Angeles, California 90024.