ıve. Now I
ch to these
ng. But the
Friedman's

COMMENT†

of "A diagrammatic approach to evidence," by R.D. Friedman

Glenn Shafer*

In his articles for this symposium, Richard Friedman introduces a method of diagramming probability models. He then uses this method to develop probabilistic explications of legal ideas along lines that were pioneered by Richard Lempert.

My comments will be directed towards two main goals. First, I will try to nail down the technical meaning of Friedman's diagrams. Second, I will criticize the general enterprise that Lempert initiated and that Friedman is carrying on. In my view, this enterprise is flawed by a misunderstanding of the relation between objective and subjective probability models. Finally, I will briefly comment on a secondary issue: the relative strength of Bayesian and frequency ideas in the statistical community.

What Do the Diagrams Mean?

Friedman's articles are a model of exposition. They will be accessible to many readers who have little mathematical training and little experience with probability. Readers who are familiar with the mathematical literature on probability (or who want to learn more from this literature) will be frustrated, however, by Friedman's failure to relate his diagrams to standard ideas and nomenclature and his failure to state the technical meaning of his diagrams in concise mathematical language. I will try to remedy these failures here.

What, exactly, does one of Friedman's diagrams tell us about joint probability distributions? It appears that the diagrams are supposed to communicate conditional independence assumptions.

The events are arranged in columns. Each column consists of mutually exclusive events, one of which is known to be true; this means that a column represents a random variable. The diagrams that I understand seem to be expanded versions of simpler diagrams relating these random variables. In Figures A and B below, I draw these simpler diagrams for Friedman's Figures 10 and 15, respectively. These simpler diagrams are trees, and the nodes in the trees are random variables. The same three random variables appear in both trees: C for clouds, R for rain, and P for the picnic. The variable C has two possible values, cloudy and not cloudy. The variable R

* Professor of Statistics, University of Kansas School of Business.

Boston University Law Review 66 1986

has two possible values, rain and no rain. And the variable P has three possible values, timely, delayed, and cancelled (I have not included nodes corresponding to the nodes labeled O in Friedman's diagrams.).

We can express Friedman's conditional independence assumptions very simply in terms of Figures A and B: conditional on a value of one of the random variables in the tree, other random variables that are separated when this random variable is removed are independent.

C→R→P

Figure A

C ↗ R
C ↘ P

Figure B

Probabilists call a tree of random variables having this property a Markov tree.[1] When the tree is actually a chain, as it is in Figures A and B, it is called a Markov chain.

This interpretation of Friedman's diagrams affords no significance to the direction of the arrows, because the definition of a Markov tree does not depend on directions being associated with the links between the random variables. It is because of this that Friedman is able to reverse the directions of the links so freely. A Markov tree does have the following property, however. If we designate one of the random variables as the initial variable, and assign each link the direction outward from this initial variable, then the entire joint probability distribution is determined by the distribution for the initial variable and the probability transition matrices in the directions of the arrows. This property partially accounts for the significance these arrows have for Friedman.

To get from a Markov tree back to Friedman's picture, we must expand each random variable into a column of nodes representing the possible values of the random variable. An arrow from variable X to variable Y in the Markov tree then becomes a collection of arrows, one from each value x for X to each value y for Y such that Pr[Y = y/X = x] = 0. This kind of elaboration is quite common in the case of a Markov chain with arrows pointing in one direction along the chain, as in Figure 10. I cannot say, however, that I have ever seen a diagram like Figure 15 before. Friedman's Figure 18 remains within the general framework I have just sketched: it is special only in that one variable in the chain is a refinement of the preceding one.

Figures 16 and 19 seem to correspond to networks that are not trees. Although they can perhaps be related to the idea of a Markov field, which is

---

[1] *See, e.g.*, Darroch, Lauritzen, & Speed, *Markov Fields and Log-Linear Interaction Models for Contingency Tables*, 8 ANNALS STATISTICS 522-525 (1980).

more general than the idea of a Markov tree, I do not fully understand Friedman's intentions in these figures. (Is Figure 16 supposed to convey the information that $P(G_3|NC) = 0.5$?)

Causal Models

Friedman's conditional independence assumptions come from causal models. Figure 10 is based on a model that says that whether the picnic is delayed or cancelled will be determined in the afternoon, and that this decision will be directly influenced only by whether it is raining in the afternoon.

This emphasis on causal models suggests that Friedman is using an objective interpretation of probability. If causal independence assumptions translate directly into probabilistic independence assumptions, then the probabilities must be probabilities in the causal model.

The causal aspect of Friedman's models gives further meaning to the directions of the arrows in his diagrams. Though he sometimes reverses the directions of the arrows, he always seems to begin with models in which the arrows indicate the direction of causation.

Markov trees that have this causal interpretation have recently been studied[2] and generalized[3] by Judea Pearl, in the Computer Science Department at UCLA. Pearl's generalization involves trees of random variables in which more than one arrow may point to the same random variable. Trees of this general kind have also been extensively studied by biologists, sociologists, and economists, under the rubric "path analysis."[4] These scientists have been primarily concerned, however, with models for statistical data, and have usually considered only continuous random variables. Friedman, like Pearl, is concerned with conceptual rather than statistical modeling, and emphasizes discrete rather than continuous variables.

A Spurious Subjectivity

Friedman's causal models are attractive, but they are obviously unrealistic for legal applications. The problem, as Friedman himself explains, is that

---

[2] *See* Pearl, *Reverend Bayes on Inference Engines: A Distributed Hierarchical Approach*, in PROCEEDINGS OF THE SECOND NATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE, AMERICAN ASSOCIATION FOR ARTIFICIAL INTELLIGENCE 133-36 (1982).

[3] *See* Kim & Pearl, *A Computational Model for Combined Causal and Diagnostic Reasoning in Inference Systems*, in PROCEEDINGS OF THE EIGHTH INTERNATIONAL JOINT CONFERENCE ON ARTIFICIAL INTELLIGENCE 190-93 (1983); *see generally* J. Pearl, *Fusion, Propagation and Structuring in Bayesian Networks*, 29 ARTIFICIAL INTELLIGENCE 241-48 (1986).

[4] *See e.g.*, Wright, *The Method of Path Coefficients*, 5 ANNALS MATHEMATICAL STATISTICS 161-215 (1934).

the objective probabilities called for by these models are seldom available in the context of adjudicative factfinding.

Friedman tries to rescue the relevance of his work from this unfortunate fact by appealing to the personalistic interpretation of probability, according to which rationality and consistency demand that a person have subjective values for these probabilities. There is no reason, however, to expect events that are causally independent to be independent with respect to a subjective probability distribution. If we believed that C and P were independent given R (as indicated by Figure A), then we might adopt as our subjective probability distribution an average of several objective distributions for which C and P are independent given R. Yet we should not expect C and P to be independent given R with respect to this average.

A numerical example may make the point clearer. Suppose we know that C and P are objectively independent given R. Suppose we also know the objective conditional probabilities for R given C:

Pr[rain / cloudy] = 0.8, Pr[rain / not cloudy] = 0.1,
Pr[no rain / cloudy] = 0.2, Pr[no rain / not cloudy] = 0.9.

But suppose we do not know the objective probabilities for C or the objective conditional probabilities for P given R. We know instead that there are two impossibilities: Either Pr[cloudy] = .03, and

Pr[timely / rain] = 0, Pr[timely / no rain] = 0.8,
Pr[delayed / rain] = 0.2, Pr[delayed / no rain] = 0.1,
Pr[cancelled / rain] = 0.8, Pr[cancelled / no rain] = 0.1.

Or else Pr[cloudy] = 0.7, and

Pr[timely / rain] = 0, Pr[timely / no rain] = 0.8,
Pr[delayed / rain] = 0.8, Pr[delayed / no rain] = 0.1,
Pr[cancelled / rain] = 0.2, Pr[cancelled / no rain] = 0.1.

Suppose we give both possibilities the subjective probability 0.5. This determines a subjective joint probability distribution for C, R, and P. But relative to this subjective distribution, C and P are not independent given R. We find, for example, that

Pr[delayed / rain & cloudy] = 0.62,

while

Pr[delayed / rain & not cloudy] = 0.38.

Even given that it rains in the afternoon, whether the picnic is delayed is not independent of whether it was cloudy in the morning. This means that Figure A is not valid for the subjective distribution.

The lesson I draw from this example is that Friedman's diagrams do not survive his transition from objective to subjective probability.

Explaining Relevance

This lesson is pertinent to Lempert's Bayesian account of relevance.[5] According to this account, an item of evidence is relevant if its likelihood ratio differs substantially from one. More precisely, evidence E is relevant to hypothesis H given prior evidence F if the likelihood ratio

Pr[E/F & H]/Pr[E/F & not H]

differs substantially from one. This is equivalent to saying that E and H are substantially dependent given F. In later elaborations of his views, Lempert has emphasized that the probabilities in this ratio are subjective and may vary from person to person, and so E is relevant if the ratio is substantially different from one for at least one reasonable person.

In the example I just gave, the subjective likelihood ratio is 0.62/0.38, or 1.63, which is substantially different from one, even though we have a causal model in which E and H are independent given F. Is the particular subjective probability distribution I used unreasonable? It is as reasonable as any. There are no particular grounds on which to justify it, but neither are there any grounds on which to justify any particular alternative.

Yet the lack of justification for this subjective distribution surely makes it inappropriate to claim that the relevance of E has been established. It seems to me that Lempert's Bayesian account of relevance is defective because it takes too seriously the range of reasonable but unfounded subjective distributions.

A simple and more sensible non-Bayesian account can be given of the significance of a causal model in which E and H are independent given F but the objective probabilities are left unspecified. This account says simply that the causal model does not establish the relevance of E to H. This is a purely negative result. It leaves open the possibility that some other argument might establish the relevance of E to H.

Who Is In The Majority?

In a footnote, Friedman suggests that the perception that Bayesians are in a minority among statisticians is outdated. For those interested in this question, here are some judgments offered by two distinguished Bayesian statisticians, David Blackwell and Morris DeGroot, in a recent conversation published in the inaugural issue of the new journal, *Statistical Science*.

Blackwell: Sort of a steady 5-10% of all work in statistical inference is done from a Bayesian point of view.

DeGroot: I see the Bayesian approach growing, but it certainly is not sweeping the field by any means.[6]

---

[5] *See* Lempert, *Modeling Relevance*, 75 MICH. L. REV. 1021 (1977).

[6] DeGroot, *A Conversation with David Blackwell*, 1 STATISTICAL SCI. 40, 48 (1986).