Semantics of Causal DAG Models

Directed acyclic graphs (DAGs) are commonly used to represent causal models.

I will compare and contrast the semantics of DAGs representing Spirtes et al. model to that of DAGs representing the non-parametric structural equation (NPSEM) model of Pearl (1995) and the ønest fully randomized causally interpreted structured tree graph (FRCISTG) model of Robins (1986).

This discussion will be more philosophical than other contributions to this volume for the following reason: the major controversies in this øeld are often focused upon the causal rather than the statistical interpretation of various analytic procedures

For example, the ønest FRCISTG and NPSEM models both assume the existence of counterfactual variables, Dawid denies their existence, and the Spirtes et al (1993) model is ȷagnosticȷ.

1 Causal models and their DAG representation

We are given a DAG G with vertex set of random variables $V = (V_1, \ldots, V_M)$ with density $f_V(v)$ ordered so that V_j is not a descendant of V_m for $m > j$.

We shall use the following notational conventions.

For any random variable Z, we let a calligraphic $\mathcal{Z}$ denote the support (i.e., the set of possible realizations z) of Z. For any $z_0, \ldots, z_m$, deøne $\overline{z}_m = (z_1, \ldots, z_m)$.

By convention $\overline{z}_0 \equiv z_0 \equiv 0$.

Let X denote any subset of V and let x be a realization of X.

Both the ønest FRCISTG and NPSEM causal models assume the existence of the counterfactual random variable $V_m(x)$ encoding the value the variable V_m would have if, possibly contrary to fact, X were set to x, where $V_m(x)$ is assumed to be be well deøned in the sense that there is reasonable agreement as to the hypothetical intervention (i.e., closest possible world) which sets X to x (Robins and Greenland, 2000).

Finest FRCISTG Causal Model: A ønest FRCISTG model assumes

(i) all one-step ahead counterfactuals $V_m(\overline{v}_{m-1})$ exist,

(ii) $V_m(\overline{v}_{m-1}) \equiv V_m(pa_m)$ is a function of $\overline{v}_{m-1}$ only through the the values pa_m of $V_m's$ parents on G,

(iii) both the observed variables V_m and the counterfactuals $V_m(x)$ for any $X \subset V$ are obtained recursively from the $V_m(\overline{v}_{m-1})$ e.g. $V_3 = V_3\{V_1, V_2(V_1)\}$ and $V_3(v_1) = V_3\{v_1, V_2(v_1)\}$,

(iv)

$$\{V_{m+1}(\overline{v}_m), \ldots, V_M(\overline{v}_{M-1})\} \coprod V_m \mid \overline{V}_{m-1} = \overline{v}_{m-1}, \text{for all } m \text{ and all } \overline{v}_{M-1} \in \overline{\mathcal{V}}_{M-1} \quad (1)$$

where $\overline{v}_k$ is a subvector of $\overline{v}_{M-1}$ for $k < M-1$.

NPSEM Causal Model:A NPSEM assumes that there exists mutually independent random variables U_m and deterministic unknown functions f_m such that the counterfactual

$$V_m\left(\overline{v}_{m-1}\right) \equiv V_m\left(pa_m\right)$$

is given by

$$f_m\left(pa_m, U_m\right)$$

and both the observed variables V_m and the counterfactuals $V_m(x)$ for any $X \subset V$ are obtained recursively from the $V_m\left(\overline{v}_{m-1}\right)$ as above.

The relationship between NPSEMs and ønest FR CISTGs is given in the following.

Lemma 1: A NPSEM can be equivalently characterized by (i) - (iv) under the deønition of a FR CISTG except with Eq. (1) replaced by

$$\left\{V_{m+1}\left(\overline{v}_m\right), \ldots, V_M\left(\overline{v}_{M-1}\right)\right\} \coprod V_m\left(\overline{v}^{**}_{m-1}\right) \mid \overline{V}_{m-1} = \overline{v}^{*}_{m-1} \text{ for all } m, \text{ all } \overline{v}_{M-1} \in \overline{\mathcal{V}}_{M-1}, \text{and all } \overline{v}^{**}_{m-1}, \overline{v}^{*}_{m-1} \quad (2)$$

Thus a NPSEM is a ønest FRCISTG but the converse is false, because a FR-CISTG assumes independence of $\left\{V_{m+1}\left(\overline{v}_m\right), \ldots, V_M\left(\overline{v}_{M-1}\right)\right\}$ and $V_m\left(\overline{v}^{**}_{m-1}\right)$ given $V_{m-1} = \overline{v}^{*}_{m-1}$ only when $\overline{v}^{**}_{m-1} = \overline{v}^{*}_{m-1} = \overline{v}_{m-1}$.

Remark: In my 1995 Biometrika comment on Pearl (1995), I incorrectly claimed in my Lemma 1 that a NPSEM and a ønest FR CISTG were equivalent.

Butch Tsiatis pointed out to me that I had failed to note that an NPSEM satisøed the stronger assumption of Eq. (2).

The proof of the (corrected) Lemma 1 proceeds as in my Biometrika comment on Pearl (1995) where the following Lemma was also proved.

Lemma 2: If a DAG G represents a FRCISTG, then the density $f_V(V)$ of the observables V satisøes the Markov factorization

$$f_V(v) = \prod_{j=1}^{M} f\left(v_j \mid pa_j\right) \quad . \quad (3)$$

Before deøning the agnostic causal model of Spirtes et. al. (1993) we need to discuss intervention distributions and the g-computation algorithm functional..

Intervention distributions on FR CISTGs: Suppose we are given a set of variables $X = \{X_1, ..., X_k\} \subset V$ and an intervention $DAG\ G^{\tilde{}} = G(X)$ that agrees with $DAG\ G$ except the parents $PA^{\tilde{}}_m$ of $X_m \in X$ may diœer from the parents of X_m on G.

A non-random $G^{\tilde{}} -$ speciøc treatment regime $g^{\tilde{}}$ is a collection of functions $g^{\tilde{}} = \left\{g^{\tilde{}}_1, \ldots, g^{\tilde{}}_k; g^{\tilde{}}_m : \mathcal{PA}^{\tilde{}}_m \to \mathcal{X}_m\right\}$ that gives the value $g^{\tilde{}}_m\left(pa^{\tilde{}}_m\right)$ that we will set X_m to when $PA^{\tilde{}}_m$ takes the value $pa^{\tilde{}}_m$.

When for each m, X_m has no parents on $G^{\tilde{}}$, so that $g_m^{\tilde{}}\left(pa_m^{\tilde{}}\right)$ is a constant, say x_m^*, we say regime $g^{\tilde{}}$ is non-dynamic and write $g^{\tilde{}} = x^* = \{x_1^*, ..., x_k^*\}$. Otherwise, $g^{\tilde{}}$ is dynamic.

The counterfactual random variable $V_j\left(g^{\tilde{}}\right)$ associated with regime $g^{\tilde{}}$ is recursively deøned (i) to be the one step ahead counterfactual $V_j\left(\overline{v}_{j-1}\right)$ evaluated at $\overline{v}_{j-1} = \overline{V}_{j-1}\left(g^{\tilde{}}\right)$ when $V_j \in V \backslash X$

and (ii) to be $g_m^{\tilde{}}\left(pa_m^{\tilde{}}\right)$ with $pa_m^{\tilde{}}$ equal to the counterfactul $PA_m^{\tilde{}}\left(g^{\tilde{}}\right)$ when $V_j = X_m \in X$.

Lemma 3 (Robins, 1986): If DAG G represents a FRCISTG, then for any set of variables $X \subset V$, and associated intervention $DAG\ G^{\tilde{}} = G(X)$, and any treatment regime $g^{\tilde{}}$, the (intervention) density

$$f_{V(g^{\tilde{}})}(v)$$

of the counterfactual $V\left(g^{\tilde{}}\right)$ is a functional of the density $f_V(v)$ of V and thus is non-parametrically identiøed from data V.. This functional, which I have referred to as the g-computation algorithm functional or density (hereafter g-functional or density), is the density $f_{g^{\tilde{}}}(v)$ obtained by recursively modifying the product on the right-hand side of (3) as follows:

for $v_j = x_m$ with $X_m \in X$ remove the term $f(v_j \mid pa_j)$ from the product and set $v_j = x_m$ to the value $g_m^{\tilde{}}\left(pa_m^{\tilde{}}\right)$ elsewhere in (3).

Agnostic Causal Model: The agnostic causal model simply assumes that the joint distribution of V factors as in (3) and that the joint density of V under the regime $g^{\tilde{}}$ on a graph $G^{\tilde{}}$ is given by the g-functional $f_{g^{\tilde{}}}(v)$.

The agnostic causal model is a simpliøed version of the causal DAG models of Sprites et al. (1993) and Pearl (1993). Although this model assumes that density of V under the intervention $g^{\tilde{}}$ is well deøned, the model makes no reference to counterfactual variables and is agnostic as to their existence. In his paper, Phil Dawid embraces a restricted version of the agnostic model in which only a subset of the variables in V can be manipulated (i.e., set) which he refers to as the decision variables. This model is closely related to the causal model discussed by Heckerman and Shachter (1995). The randomized causally-interpreted structured tree graph (RCISTG) of Robins (1987) likewise restricts the set of variables in V that can be manipulated. The relationship of a RCISTG model to an FRCISTG model is analogous to that of Dawid's restricted agnostic model to the agnostic model.

In his paper Dawid was interested in the marginal intervention distribution $f_{g^{\tilde{}}}(y) = \int \cdots \int f_{g^{\tilde{}}}(v)\, d\mu(y^c)$ of a subset Y of the variables in $V = (Y, Y^c)$, say, and further that data are obtained only on some subset V^* of the variables in V with $Y \subset V^*$. In this case one wishes to know whether the intervention distribution $f_{g^{\tilde{}}}(y)$ of Y is identiøed from (i.e., is a functional of) the marginal distribution $f_{V^*}(v^*)$ of V^* and, if not, to set bounds on $f_{g^{\tilde{}}}(y)$ as discussed by Dawid. SuŒcient conditions for identiøcation have been derived by Galles and Pearl (1995) for univariate (i.e., time-independent) interventions, Pearl and Robins (1995) for non-dynamic regimes, and Robins (1997) for dynamic regimes.

We refer the reader to the above references for additional discussion.