Evaluating Causal effects using Chain Event Graphs

Peter Thwaites and Jim Smith
University of Warwick Statistics Department
Coventry, United Kingdom

Abstract

The Chain Event Graph (CEG) is a coloured mixed graph used for the representation of finite discrete distributions. It can be derived from an Event Tree (ET) together with a set of equivalence statements relating to the probabilistic structure of the ET. CEGs are especially useful for representing and analysing asymmetric processes, and collections of implied conditional independence statements over a variety of functions can be read from their topology. The CEG is also a valuable framework for expressing causal hypotheses, and manipulated-probability expressions analogous to that given by Pearl in his Back Door Theorem can be derived. The expression we derive here is valid for a far larger set of interventions than can be analysed using Bayesian Networks (BNs), and also for models which have insufficient symmetry to be described adequately by a Bayesian Network.

1 Introduction

Bayesian Networks are good graphical representations for many discrete joint probability distributions. However, many asymmetric models (by which we mean models with non-symmetric sample space structures) cannot be fully described by a BN. Such processes arise frequently in, for example, biological regulation, risk analysis and Bayesian policy analysis.

In eliciting these models it is usually sensible to start with an Event Tree (Shafer, 1996), which is essentially a description of how the process unfolds rather than how the system might appear to an observer. Working with an ET can be quite cumbersome, but they do reflect any model asymmetry, both in model development and in model sample space structure.

The Chain Event Graph (Riccomagno & Smith, 2005; Smith & Anderson, 2006; Thwaites & Smith, 2006a) is a graphical structure designed for analysis of asymmetric systems. It retains the advantages of the ET, whilst typically having far fewer edges and vertices. Moreover, the CEG can be read for a rich collection of conditional independence properties of the model. Unlike Jaeger's very useful Probabilistic Decision Graph (2002) this includes all the properties that can be read from the equivalent BN if the CEG represents a symmetric model and far more if the model is asymmetric.

In the next section we show how CEGs can be constructed. We then describe how we can use CEGs to analyse the effects of Causal manipulation.

Bayesian Networks are often extended to apply also to a control space. When it is valid to make this extension the BN is called causal. Although there is debate (Pearl, 2000; Lauritzen, 2001; Dawid, 2002) about terminology, it is certainly the case that BNs are useful for analysing (in Pearl's notation) the effects of manipulations of the form $Do\ X = x_0$ in symmetric models, where X is a variable to be manipulated and x_0 the setting that this variable is to be manipulated to. This type of intervention, which might be termed atomic, is actually a rather coarse manipulation since we would need to extend the space to make predictions of effects when X is manipulated to **any** value. Although there is a case for only considering such manipulations when a model is very symmetric, it is too coarse to capture many of the manipulations we might want to consider in asymmetric environments. We use this paper to show how CEGs can be used to analyse a far more refined *singular* manipulation in models which may have insufficient symmetry to be described adequately by a Bayesian Network.

2 CEG construction

We can produce a CEG from an Event Tree which we believe represents the model (see for example Figure 1). This ET is just a graphical description of how the process unfolds, and the set of atoms of the Event Space (or path sigma algebra) of the tree is simply the set of root to leaf paths within the tree. Any random variables defined on the tree are measureable with respect to this path sigma algebra.

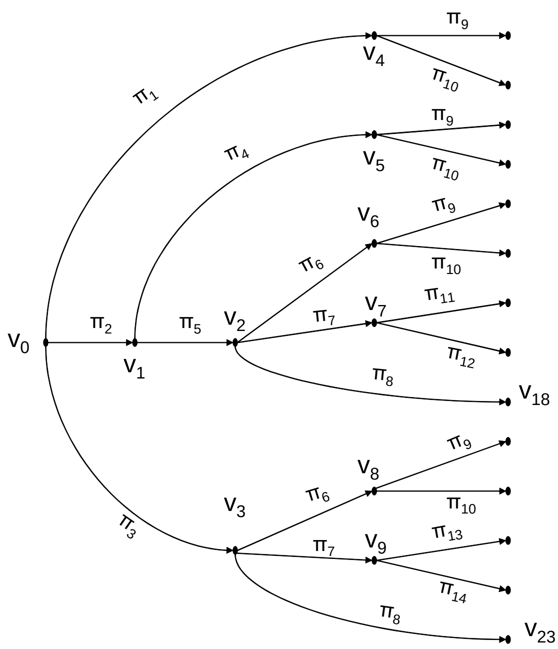

Figure 1. ET for machine example.

Example 1.
A machine in a production line utilises two replaceable components A and B. Faults in these components do not automatically cause the machine to fail, but do affect the quality of the product, so the machine incorporates an automated monitoring system, which is completely reliable for finding faults in A, but which can detect a fault in B when it is functioning correctly.

In any monitoring cycle, component A is checked first, and there are three initial possibilities: A, B checked and no faults found (π_1 on the ET in Figure 1); A checked, fault found, machine switched off (π_2); A checked, no fault found, B checked, fault found, machine switched off (π_3).

If A is found faulty it is replaced and the machine switched back on (vertex v_1), and B is then checked. B is then either found not faulty (π_4), or faulty and the machine switched off (π_5).

If B is found faulty by the monitoring system, then it is removed and checked (vertices v_2 and v_3). There are then three possibilities, whose probabilities are independent of whether or not component A has been replaced: B is not in fact faulty, the machine is reset and restarted (π_6); B is faulty, is successfully replaced and the machine restarted (π_7); B is faulty, is replaced unsuccessfully and the machine is left off until the engineer can see it (π_8).

At the time of any monitoring cycle, the quality of the product produced (π_{10}) is unaffected by the replacement of A unless B is also replaced. It is however dependent on the *effectiveness* of B which depends on its *age*, but also, if it is a **new** component, on the *age* of A; so:
$\pi(\textit{good product} \mid \textit{A and B replaced}) = \pi_{12}$
$> \pi(\textit{good product} \mid \textit{only B replaced}) = \pi_{14}$
$> \pi(\textit{good product} \mid \textit{B not replaced}) = \pi_{10}$

An ET for this set-up is given in Figure 1 and a derived CEG in Figure 2. Note that:

- The subtrees rooted in the vertices v_4, v_5, v_6 and v_8 of the ET are identical (both in physical structure and in probability distribution), so these vertices have been conjoined into the vertex (or *position*) w_4 in the CEG.
- The subtrees rooted in v_2 and v_3 are not identical (as $\pi_{11} \neq \pi_{13}$, $\pi_{12} \neq \pi_{14}$), but the edges leaving v_2 and v_3 carry identical probabilities. The equivalent *positions* in the CEG w_2 and w_3 have been joined by an undirected edge.
- All leaf-vertices of the ET have been conjoined into one sink-vertex in the CEG, labelled w_∞.

To complete the transformation, note that $v_i \to w_i$ for $0 \leq i \leq 3$, $v_7 \to w_5$ and $v_9 \to w_6$.

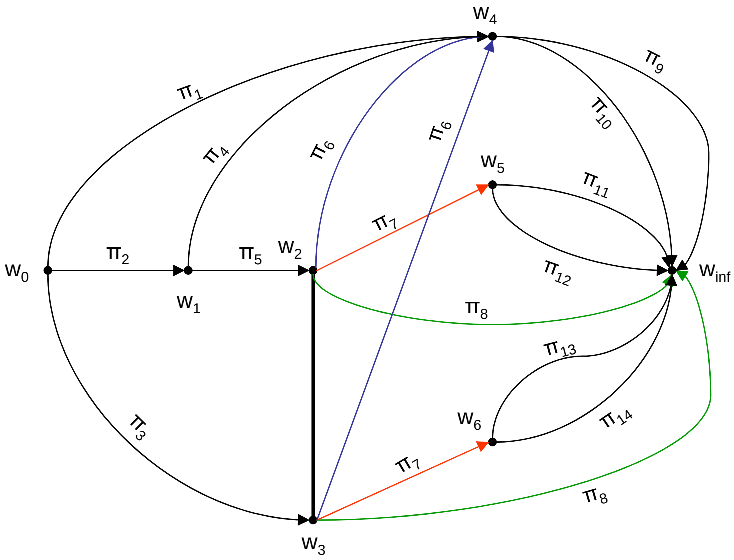

Figure 2. CEG for machine example.

A formal description of the process is as follows: Consider the ET $T = (V(T), E(T))$ where each element of $E(T)$ has an associated edge probability. Let $S(T) \subset V(T)$ be the set of non-leaf vertices of the ET.
Let $v_i \prec v_j$ indicate that there is a path joining vertices v_i and v_j in the ET, and that v_i precedes v_j on this path.
Let $\mathbb{X}(v)$ be the sample space of $X(v)$, the random variable associated with the vertex v ($\mathbb{X}(v)$ can be thought of as the set of edges leaving v, so in our example, $\mathbb{X}(v_1) = \{B\ found\ not\ faulty, B\ found\ faulty\}$).

For any $v \in S(T)$, $v_l \in V(T) \backslash S(T)$ such that $v \prec v_l$:

- Label $v = {v_\lambda}^0$
- Let ${v_\lambda}^{i+1}$ be the vertex such that ${v_\lambda}^i \prec {v_\lambda}^{i+1} \prec v_l$ for which there is no vertex v' such that ${v_\lambda}^i \prec v' \prec {v_\lambda}^{i+1}$ for $i \geq 0$
- Label $v_l = {v_\lambda}^m$, where the path λ consists of m edges of the form $e({v_\lambda}^i, {v_\lambda}^{i+1})$

Definition 1. For any $v_1, v_2 \in S(T)$, v_1 and v_2 are termed *equivalent*, *iff* there is a bijection ψ which maps the set of paths (and component edges)
$\Lambda_1 = \{\lambda_1(v_1, v_{l_1}) \mid v_{l_1} \in V(T) \backslash S(T)\}$ onto $\Lambda_2 = \{\lambda_2(v_2, v_{l_2}) \mid v_{l_2} \in V(T) \backslash S(T)\}$ in such a way that:

(a) $\psi(e({v_{\lambda_1}}^i, {v_{\lambda_1}}^{i+1})) = e(\psi({v_{\lambda_1}}^i), \psi({v_{\lambda_1}}^{i+1})) = e({v_{\lambda_2}}^i, {v_{\lambda_2}}^{i+1})$ for $0 \leq i \leq m(\lambda)$
(b) $\pi({v_{\lambda_1}}^{i+1} \mid {v_{\lambda_1}}^i) = \pi({v_{\lambda_2}}^{i+1} \mid {v_{\lambda_2}}^i)$
where ${v_{\lambda_1}}^{i+1}$ and ${v_{\lambda_2}}^{i+1}$ label the same **value** on the sample spaces $\mathbb{X}(v^i_{\lambda_1})$ and $\mathbb{X}(v^i_{\lambda_2})$ for $i \geq 0$.

The set of equivalence classes induced by the bijection ψ is denoted $K(T)$, and the elements of $K(T)$ are called *positions*.

Definition 2. For any $v_1, v_2 \in S(T)$, v_1 and v_2 are termed *stage-equivalent*, *iff* there is a bijection ϕ which maps the set of edges
$E_1 = \{e_1(v_1, {v_1}') \mid {v_1}' \in \mathbb{X}(v_1)\}$ onto $E_2 = \{e_2(v_2, {v_2}') \mid {v_2}' \in \mathbb{X}(v_2)\}$ in such a way that:
$\pi({v_1}' \mid v_1) = \pi(\phi({v_1}') \mid \phi(v_1)) = \pi({v_2}' \mid v_2)$
where ${v_1}'$ and ${v_2}'$ label the same **value** on the sample spaces $\mathbb{X}(v_1)$ and $\mathbb{X}(v_2)$.

The set of equivalence classes induced by the bijection ϕ is denoted $L(T)$, and the elements of $L(T)$ are called *stages*.

A CEG $C(T)$ of our model is constructed as follows:

(1) $V(C(T)) = K(T) \cup \{w_\infty\}$
(2) Each $w, w' \in K(T)$ will correspond to a set of $v, v' \in S(T)$. If, for such v, v', $\exists$ a directed edge $e(v, v') \in E(T)$, then $\exists$ a directed edge $e(w, w') \in E(C(T))$
(3) If $\exists$ an edge $e(v, v_l) \in E(T)$ *st* $v \in S(T)$

and $v_l \in V(T) \backslash S(T)$, then $\exists$ a directed edge $e(w, w_\infty) \in E(C(T))$

(4) If two vertices $v_1, v_2 \in S(T)$ are stage-equivalent, then $\exists$ an undirected edge $e(w_1, w_2) \in E(C(T))$

(5) If w_1 and w_2 are in the same stage (ie: if v_1, v_2 are stage-equivalent in $E(T)$), and if $\pi(v_1' \mid v_1) = \pi(v_2' \mid v_2)$ then the edges $e(w_1, w_1')$ and $e(w_2, w_2')$ have the same *label* or *colour* in $E(C(T))$.

Note that in our example, w_2 and w_3 are in the same stage and that $\pi(v_6|v_2) = \pi(v_8|v_3)$, $\pi(v_7|v_2) = \pi(v_9|v_3)$, $\pi(v_{18}|v_2) = \pi(v_{23}|v_3)$, so the edges $e(w_2, w_4)$ and $e(w_3, w_4)$ are coloured the same, as are the edges $e(w_2, w_5)$ and $e(w_3, w_6)$ and as are $e(w_2, w_\infty)$ and $e(w_3, w_\infty)$.

More detail on CEG construction can be found in Smith & Anderson (2006), as can a detailed description of how CEGs are read. We conclude section 2 of this paper by looking at two ideas that will be used extensively in the next section.

Firstly, when we say that a position w in our CEG or the set of edges leaving w have an associated *variable*, we are not referring to the measurement-variables of a BN-representation of the problem, each of which must take a value for any atomic event, but to a more flexible construct defined through stage-equivalence in the underlying tree. The exit-edges of a position are simply the collection of possible immediate outcomes in the next step of the process given the history up to that position. A *setting* (or *value* or *level*) is then simply a possible realisation of a variable in this collection.

Secondly, we use these edge probabilities to define the probabilities of composite events in the path sigma field of our CEG:

Definition 3. For two positions w, w' with $w \prec w'$, let $\pi_\lambda(w' \mid w)$ be the probability associated with the path $\lambda(w, w')$. Note that this will be a product of edge probabilities.

$$\text{Define} \quad \pi(w' \mid w) \triangleq \sum_{\lambda \in \Lambda} \pi_\lambda(w' \mid w)$$

where Λ is the set of all paths from w to w'.

Note that the combination rules for path probabilities on CEGs (directly analogous to those for trees) give us that for any 3 positions w_1, w_2, w_3, with $w_1 \prec w_2 \prec w_3$, we have that $\pi(w_3 \mid w_1, w_2) = \pi(w_3 \mid w_2)$; that is the probability that we pass through position w_3 given that we have passed through positions w_1 and w_2 is simply the probability that we pass through position w_3 given that we have passed through position w_2.

3 Manipulations of CEGs

The simplest types of intervention are of the form *Do* $X = x_0$ for some variable X and setting x_0, and these are really the only interventions that can be satisfactorily analysed using BNs. In this paper we consider a much more general intervention where not only the setting of the manipulated variable, but the variable itself may be different depending on the settings of other variables within the system.

We can model such interventions by the production of a *manipulated* CEG $\hat{C}$ in parallel with our *idle* CEG C. In the intervention considered here every path in our CEG is manipulated by having **one** component edge given a probability of 1 or 0. All edges with zero probabilities and branches stemming from such edges are removed (or *pruned*) from $\hat{C}$ (note that in this paper all edges on any CEG will have non-zero probabilities). We will call such an intervention a *singular* manipulation, and denote it *Do Int.*

Definition 4. A subset W_X of positions of C qualifies as a singular manipulation set if:

(1) all root-to-sink paths in C pass through exactly one position in $pa(W_X)$, where $w \in pa(W_X)$ if $w \prec w'$ for some $w' \in W_X$ and there exists an edge $e(w, w')$

(2) each position in $pa(W_X)$ has exactly one child in W_X, by which we mean that for $w \in pa(W_X)$, there exists exactly one $w' \in W_X$ such that there exists an edge $e(w, w')$

A singular manipulation is then an intervention such that:

(a) for each $w \in pa(W_X)$ and corresponding $w' \in W_X$, $\hat{\pi}(w' \mid w) = 1$

(b) for any $w \in pa(W_X)$ and $w' \notin W_X$ such that $w \prec w'$ and there exists an edge $e(w, w')$, then $\hat{\pi}(w' \mid w) = 0$, and this edge is removed (or *pruned*) in $\hat{C}$

(c) for any $w \notin pa(W_X)$ and w' such that $w \prec w'$ and there exists an edge $e(w, w')$, then $\hat{\pi}(w' \mid w) = \pi(w' \mid w)$

where $\hat{\pi}$ is a probability in our manipulated CEG $\hat{C}$.

Let $W_X = \{w_j\}$, $pa(W_X) = \{w_j^i\}$. Each position in $pa(W_X)$ has exactly one child in W_X so elements of $pa(W_X)$ can be intially identified by their child in W_X (ie by the index j). But a position in W_X could have more than one parent in $pa(W_X)$, so we distinguish these parents by a second index i. For each pair (w_j^i, w_j) let X_j^i be the variable associated with the edge $e(w_j^i, w_j)$ and x_j^i be the setting of this variable on this edge.

If we also consider a response variable Y downstream from the set of positions W_X, then we can show (using for example Pearl's definition of Do) that:

$$\pi(y \mid Do\ Int) = \sum_{i,j} \Big[\pi(w_j^i \mid w_0)\ \pi(y \mid w_j) \Big] \qquad (3.1)$$

Pearl's own Back Door expression (below) is a simplification of the general manipulated-probability expression used with BNs.

$$\pi(y \mid Do\ x_0) = \sum_z \pi(y \mid z, x_0)\ \pi(z) \qquad (3.2)$$

Z here is a subset of the measurement-variables of the BN which obey certain conditions. If Z is chosen carefully then we can calculate $\pi(y \mid Do\ x_0)$ without conditioning on the full set of measurement-variables.

In this paper we use the topology of the CEG to produce an analogous expression to (3.2) for our more general *singular* manipulation, by using a set of positions W_Z downstream from the intervention which can stand-in for the set of positions W_X used in expression (3.1). As with Pearl's expression, the use of such a set W_Z will reduce the complexity of the general expression (3.1) as well as possibly allowing us to sidestep identifiability problems associated with it.

Following Pearl, we have two conditions, which if satisfied, are sufficient for W_Z to be considered a Back Door blocking set. We give the first here, and the second following a few further definitions.

(A) For all $w_j \in W_X$, every $w_j - w_\infty$ path in C must pass through exactly one position $w_k \in W_Z$

The obvious notation for use with CEGs is a path-based one. However most practitioners will be more familiar with expressions such as (3.2), so we here develop a few ideas to allow us to express our causal expression in a similar fashion. The first step in this process is to note that any position w in a CEG has a **unique** set $q(w)$ associated with it, where:

- $Q(w)$ is the **minimum** set of variables, by specifying the settings (or *values* or *levels*) of which, we can describe the union of all $w_0 - w$ paths
- $q(w)$ are the settings of $Q(w)$ which fully describe the union of all $w_0 - w$ paths

Formally this means that:

$$q(w) = \bigcup_{\lambda \in \Lambda_w} q(\lambda)$$

where $q(\lambda)$ are the settings on the $w_0 - w$ path λ, and Λ_w is the set of all $w_0 - w$ paths.

Letting $Z(w)$ be the set of variables encountered on edges upstream of w, $X(w)$ be the set of variables encountered on edges downstream of w, and $R(w) = Z(w) \backslash Q(w)$, we note that the conditional independence statement encoded by the position w is of the form:

$$X(w) \amalg R(w) \mid q(w)$$

In the CEG in Figure 2 for example, $X(w_4) \amalg R(w_4) \mid q(w_4)$ tells us that product quality is independent of the monitoring system responses, given that B is not replaced.

Note that the definition of $q(w)$ means that the variable-settings within it might not always correspond to simple values of the variables within $Q(w)$. None-the-less, we have found that $q(w)$ is typically simpler than each individual $q(\lambda)$.

We now use the ideas outlined above to expand the expression (3.1). As the position w_k is uniquely defined by $q(w_k)$, we can write, without ambiguity $\pi(y \mid w_k) = \pi(y \mid q(w_k))$. Using condition (A) we get:

$$\begin{aligned} &\pi(y \mid Do\ Int) \qquad (3.3)\\ &= \sum_{i,j} \Big[\pi(w_j^i \mid w_0) \sum_k \pi(w_k, y \mid w_j) \Big] \\ &= \sum_{i,j,k} \pi(w_j^i \mid w_0)\ \pi(y \mid w_j, w_k)\ \pi(w_k \mid w_j) \\ &= \sum_{i,j,k} \pi(w_j^i \mid w_0)\ \pi(y \mid w_k)\ \pi(w_k \mid w_j) \\ &= \sum_k \Big[\sum_{i,j} \pi(w_j^i|w_0)\pi(w_k|w_j) \Big] \pi(y|q(w_k)) \end{aligned}$$

The equivalence of $\pi(y \mid w_j, w_k)$ and $\pi(y \mid w_k)$ is a consequence of the equivalence of $\pi(w_3 \mid w_1, w_2)$ and $\pi(w_3 \mid w_2)$ noted at the end of section 2, and proved in Thwaites & Smith (2006b).

We now need a number of technical definitions before we can introduce our second condition and proceed to our causal expression. Examples illustrating these definitions can be found in Thwaites & Smith (2006b).

Now, the position w_k can also be fully described by the union of disjoint events, each of which is (by conditions (1) and (A)) a $w_0 - w_j^i - w_k$ path for some $w_j^i \in pa(W_X)$. These events divide into 2 distinct sets:

(1) $w_0 - w_j^i - w_j - w_k$ paths

(2) paths that do **not** utilise the x_j^i edge when leaving w_j^i (formally $w_0 - w_j^i - w' - w_k$ paths where there exists an edge $e(w_j^i, w')$, but $w' \notin W_X$).

We can combine the events in set (1) into *composite* or *C-paths* so that each *C-path* passes through exactly one w_j^i and can be uniquely characterised by a pair $q_j^i(w_k) = (x_j^i, z_j^i(w_k))$, where z_j^i is defined as follows:

- $Z_j^i(w_k)$ is the **minimum** set of variables, by specifying the settings of which, we can describe the union of all $w_0 - w_j^i - w_j - w_k$ paths (with X_j^i excluded from this set)
- $z_j^i(w_k)$ are the settings of $Z_j^i(w_k)$ which (with the addition of $X_j^i = x_j^i$) fully describe the union of all $w_0 - w_j^i - w_j - w_k$ paths

Definition 5. We express this formally as: Let $q_j^i(w_k) = \bigcup_{\lambda \in \Lambda} q(\lambda)$, where $q(\lambda)$ are the settings on the $w_0 - w_j^i - w_j - w_k$ path λ, and Λ is the set of all $w_0 - w_j^i - w_j - w_k$ paths in C. Let $Q_j^i(w_k)$ be the set of variables present in $q_j^i(w_k)$.
Define $Z_j^i(w_k)$ as $Q_j^i(w_k) \backslash X_j^i$. Let $z_j^i(w_k)$ be the settings of $Z_j^i(w_k)$ compatible with $q_j^i(w_k)$.

Our $X_j^i, x_j^i, Z_j^i(w_k), z_j^i(w_k)$ are directly analogous to Pearl's X, x, Z and z in expression (3.2), and fulfil similar roles in our final causal expression.

The following rather technical definitions are **only** required for an understanding of the proof. Definition 7 deals with the idea of a *descendant* which is very similar to the analogous idea in BNs, and is needed for condition (B).

Definition 6.
Define $q(w_j^i)$ analogously with the definition of $q(w)$. Let $Q(w_j^i)$ be the set of variables present in $q(w_j^i)$.
Define $Q_j(w_k)$ as $Z_j^i(w_k) \backslash Q(w_j^i)$. Note that $Q(w_j^i) \subset Z_j^i(w_k)$. Let $q_j(w_k)$ be the settings of $Q_j(w_k)$ compatible with $z_j^i(w_k)$ (or $q_j^i(w_k)$).

We can therefore write:

$$z_j^i(w_k) = (q(w_j^i), q_j(w_k))$$
$$q_j^i(w_k) = (x_j^i, z_j^i(w_k)) = (q(w_j^i), x_j^i, q_j(w_k))$$

We can also combine the events in set (2) into *C-paths*, each of which can be uniquely characterised by $r_j^i(w_k) = \bigcup_{\mu \in M} q(\mu)$, where $q(\mu)$ are the settings on the $w_0 - w_j^i - w' - w_k$ path μ, and M is the set of all $w_0 - w_j^i - w' - w_k$ paths in C. We can therefore write:

$$q(w_k) = \Big[\bigcup_{i,j} q_j^i(w_k) \Big] \bigcup \Big[\bigcup_{i,j} r_j^i(w_k) \Big]$$

Definition 7. Consider variables $A, D, \{B_m\}$ defined on our CEG C. Then D is a descendant of A in C if there exists a sequence of (not necessarily adjacent) edges $e_1, \ldots e_n$ forming part of a $w_0 - w_\infty$ path in C where the edges $e_1, \ldots e_n$ are labelled respectively $b_1 \mid (a, \ldots)$, $b_2 \mid (b_1, \ldots), \ldots b_{n-1} \mid (b_{n-2}, \ldots)$, $d \mid (b_{n-1}, \ldots)$, or if there exists an edge forming part of a $w_0 - w_\infty$ path in C labelled $d \mid (a, \ldots)$; where $a, b_1, b_2, \ldots b_{n-1}, d$ are settings of $A, B_1, B_2, \ldots B_{n-1}, D$.

We are now in a position to state our 2nd condition.

(B) In the sub-CEG C_j^i with w_j^i as root-node, $Q_j(w_k)$ must contain no descendants of X_j^i for all i, j, for each position w_k

Checking that condition (B) is fulfilled is actually straightforward on a CEG, especially since we will know which manipulations we intend to investigate, and can usually construct our CEG so as to make $Q_j(w_k)$ as *small* as possible for all values of j, k.

We can now replace expression (3.3) by a Back Door expression for singular manipulations:

Proposition 1.

$$\pi(y \mid Do\ Int) = \sum_k \Big[\sum_{i,j} \pi(z_j^i(w_k))\Big]\ \pi(y \mid q(w_k)) \tag{3.4}$$

Proof.
Consider

$$\begin{aligned}
&\pi(w_k \mid w_j) = \pi(w_k \mid w_j^i, w_j) \\
&= \pi(q(w_k) \mid q(w_j^i), x_j^i) \\
&= \pi\Big(\Big[\bigcup_{m,n} q_n^m(w_k)\Big] \cup \Big[\bigcup_{m,n} r_n^m(w_k)\Big] | q(w_j^i), x_j^i\Big) \\
&= \sum_{m,n} \pi(q_n^m(w_k) \mid q(w_j^i), x_j^i) \\
&\quad + \sum_{m,n} \pi(r_n^m(w_k) \mid q(w_j^i), x_j^i)
\end{aligned}$$

since disjoint.

$$\begin{aligned}
&= \pi(q_j^i(w_k) | q(w_j^i), x_j^i) + \pi(r_j^i(w_k) | q(w_j^i), x_j^i) \\
&= \pi(q_j^i(w_k) \mid q(w_j^i), x_j^i) \\
&= \pi(q(w_j^i), x_j^i, q_j(w_k) \mid q(w_j^i), x_j^i) \\
&= \pi(q_j(w_k) \mid q(w_j^i), x_j^i)
\end{aligned}$$

But this is simply the probability that $Q_j(w_k) = q_j(w_k)$ given that $X_j^i = x_j^i$ in the sub-CEG C_j^i.
Condition (B) implies that $X_j^i \amalg Q_j(w_k)$ in C_j^i since X_j^i has no parents in this CEG. So we get:

$$\pi(w_k \mid w_j) = \pi(q_j(w_k) \mid q(w_j^i))$$

Substituting this into expression (3.3), we get:

$$\begin{aligned}
&\pi(y \mid Do\ Int) \\
&= \sum_k \Big[\sum_{i,j} \pi(w_j^i \mid w_0)\ \pi(q_j(w_k) \mid q(w_j^i))\Big] \\
&\qquad \times \pi(y \mid q(w_k)) \\
&= \sum_k \Big[\sum_{i,j} \pi(q(w_j^i))\ \pi(q_j(w_k) \mid q(w_j^i))\Big] \\
&\qquad \times \pi(y \mid q(w_k)) \\
&= \sum_k \Big[\sum_{i,j} \pi(z_j^i(w_k))\Big]\ \pi(y \mid q(w_k)) \qquad \square
\end{aligned}$$

It is possible to show that the expression $\pi(y \mid q(w_k))$ can be replaced by a probability conditioned on a **single** $w_0 - w_k$ path, and moreover that even on that path Y may well be independent of some of the variables encountered given the path-settings of the others — for details see Thwaites & Smith (2006b).

We also noted earlier that $q(w) = \bigcup q(\lambda)$ is typically simpler than each individual $q(\lambda)$. In most instances $\sum_{i,j} \pi(z_j^i(w_k))$ will be the probability of a union of disjoint events which will also typically be simpler than an individual $z_j^i(w_k)$. We can deduce that calculating $\sum_{i,j} \pi(z_j^i(w_k))$ is unlikely to be a complex task.

We conclude this section by demonstrating how expression (3.4) is related to Pearl's Back Door expression (3.2):
Consider the intervention $Do\ X = x_0$, and let $X_j^i = X$ and $x_j^i = x_0$ for all i, j. Combine all our $w_0 - w_j^i - w_j - w_k$ *C-paths*, and write $\bigcup_{i,j} q_j^i(w_k) = (x_0, z(w_k))$. Rephrase conditions (2) and (B) as:

(2) each position in $pa(W_X)$ has exactly one of its outward edges in C labelled x_0, and this edge joins the position in $pa(W_X)$ to a position in W_X

(B) $Z(w_k)$ must contain no descendants of X (where $Z(W_k)$ is defined from $z(w_k)$ in the obvious manner)

Then with a little work, we can replace expression (3.4) by:

$$\pi(y \mid Do\ Int)$$
$$= \sum_k \pi(z(w_k))\ \pi(y \mid x_0, z(w_k))$$

If $Z(w_k)$ contains the same variables for all k, and $z(w_k)$ runs through all settings of $Z(w_k)$ as we run through all w_k, then this expression reduces to Pearl's expression (3.2).

4 Causal Analysis on CEGs and BNs

The principal advantage that CEGs have over BNs when it comes to Causal analysis is their flexibility. In a BN the kind of manipulations we can consider are severely restricted (see for example section 2.6 of Lauritzen (2001) where he comments on Shafer (1996)), whereas using a CEG we can tackle not only the conventional $Do\ X = x_0$ manipulations of symmetric models, but also the analysis of interventions on asymmetric models and manipulations where both the manipulated-variable and the manipulated-variable value can differ for different settings of other variables. It is also the case that our blocking sets are sets of values or settings, and do not need to correspond to any fixed subset of the original problem random variables.

For simplicity of exposition in this paper we have not fully exploited the potential flexibility of the CEG, considering only formulae associated with a blocking-set W_Z of *positions* downstream of W_X. We can also consider sets of *stages* upstream of W_X, and combinations of the two. Also, we have only discussed one particular fairly coarse example of an intervention. There are often circumstances where some paths in our CEG are not manipulated at all, for example in a treatment regime where only patients with certain combinations of symptoms (ie at certain positions or stages) are treated. There are also non-singular interventions where (for instance) a manipulation, rather than forcing a path to follow one specific edge at some vertex, instead provides a probability distribution for the outgoing edges of that vertex.

So not only are CEGs ideal representations of asymmetric discrete models, retaining the convenience of a tree-form for describing how a process unfolds but also expressing a rich collection of the model's conditional independence properties, but their event-based causal analysis has distinct advantages over the variable-based analysis one performs on Bayesian Networks.

References

A.P. Dawid. 2002. Influence Diagrams for Causal Modelling and Inference. *International Statistical Review 70.*

M. Jaeger. 2002. Probabilistic Decision Graphs — Combining Verification and AI techniques for Probabilistic Inference. In *PGM '02 Proceedings of the 1st European Workshop on Probabilistic Graphical Models,* pages 81-88.

S.L. Lauritzen. 2001. Causal Inference from Graphical Models. In O.E. Barndorff-Nielsen et al (Eds) *Complex Stochastic Systems.* Chapman and Hall / CRC.

J. Pearl. 2000. *Causality: models, reasoning and inference.* Cambridge University Press.

E. Riccomagno and J.Q. Smith. 2005. The Causal Manipulation and Bayesian Estimation of Chain Event Graphs. CRiSM Research Report no. 05-16, University of Warwick.

G. Shafer. 1996. *The Art of Causal Conjecture.* MIT Press.

J.Q. Smith and P.E. Anderson. 2006. Conditional independence and Chain Event Graphs. Accepted, subject to revision, by *Journal of Artificial Intelligence.*

P.A. Thwaites and J.Q. Smith. 2006a. Non-symmetric models, Chain Event Graphs and propagation. In *IPMU '06 Proceedings of the 11th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems.*

P.A. Thwaites and J.Q. Smith. 2006b. Singular manipulations on Chain Event Graphs. Warwick Research Report, University of Warwick.