Toward Learning Graphical and Causal Process Models

Christopher Meek
Microsoft Research
One Microsoft Way
Redmond, WA 98052
meek@microsoft.com

Abstract

We describe an approach to learning causal models that leverages temporal information. We posit the existence of a graphical description of a causal process that generates observations through time. We explore assumptions connecting the graphical description with the statistical process and what one can infer about the causal structure of the process under these assumptions.

1 Introduction

Data that measure the temporal dynamics of systems is pervasive. The goal of this paper is to describe an approach to the development of a sound approach to causal inference for dynamic systems. One of the popular extant approaches is Granger causality (Granger 1969) which fails to be sound in the presence of latent variables. Granger causality is typically applied in discrete-time continuous valued time-series. Roughly speaking, in a multivariate time series X a set of variables are the Granger-causes of X_j if the historical values of this set of variables (including X_j) are necessary and sufficient for optimal prediction. Unfortunately a variable deemed a Granger-cause can arise due to either a latent common cause or as a result of a direct causal relationship and thus the approach cannot be used to determine causal relationships if one does not exclude the possibility of latent variables.

In this paper, we explore how one can leverage the assumption that causes must precede effects to inform causal conclusions drawn from observations of a temporal statistical process. The approach taken here is similar to the approach developed by Verma and Pearl (1990) and Spirtes, Glymour, and Scheines (2001) for atemporal causal discovery. One key ingredient in our approach is a new asymmetric graphical separation criterion for directed (possibly cyclic) graphs called δ*separation which plays an analogous role as d-separation in the work of Verma and Pearl (1990) and Spirtes, Glymour, and Scheines (2001). Another key ingredient is the process independence statement that plays an analogous role to the independence statement. Conceptually, we assume that we can test whether a process independence statements about observable quantities holds by observing the process and that these observation provide insight into the causal structure governing the process. In particular, we posit the existence of a graphical description of a causal process and make assumptions that connect δ*separation with observable process independence statements. We explore what can be inferred about the causal structure of the process under various observability assumptions. While the ultimate goal is to create a sound and complete method for causal inference for observations from a stochastic dynamic system, this paper represents some initial steps towards this ultimate goal. In particular, the results in Section 3.2 can be viewed sufficient conditions for Granger causality and, in Section 3.3, we present sufficient conditions under which we can make sound inferences about causal relationships including the existence of causal relationships and the existence and non-existence of latent common causal relationships.

As presented in Section 3, our causal discovery algorithm assumes the existence of an oracle for process independence statements. Our approach of abstracting away the details of how one connects process independence statements with particular statistical processes allows us to simultaneously make progress on the causal discovery problem for multiple distinct statistical processes such as marked point processes, Gaussian processes and dynamic Bayesian networks. In Section 4, we discuss two particular statistical processes and their associated process independence statements. In Section 5, we discuss some related work and open research questions.

2 Graphical Separation

We use $\mathcal{G} = \langle \mathcal{L}, \mathcal{E} \rangle$ to denote a directed graph where $\mathcal{L}$ is a set of vertices and $\mathcal{E} \subseteq \langle \mathcal{L} \times \mathcal{L} \rangle$ is a set of edges represented as ordered pairs. We write $a \to b$ if $\langle a, b \rangle \in \mathcal{E}$ and say that a is a parent of b and b is the child of a. Note that, in addition to allowing cycles, we also allow that a vertex can be its own parent and child (i.e., a self-edge $a \to a$). We use the shorthand $a \leftrightarrow b$ to indicate that $a \to b$ and $b \to a$.

A *path* in $\mathcal{G}$ is a sequence $\langle l_1, \ldots, l_n \rangle$ where there is an edge between successive pairs of vertices in $\mathcal{G}$. The length of a path $p = \langle l_1, \ldots, l_n \rangle$ is $|p| = n$ and a path p is termed a *trivial path* if $|p| = 1$. A vertex l_i on path $p = \langle l_1, \ldots, l_n \rangle$ is a *collider* on p if $l_{i-1} \to l_i$ and $l_i \leftarrow l_{i+1}$ and a *non-collider* otherwise. A *directed path* in graph $\mathcal{G} = \langle \mathcal{L}, \mathcal{E} \rangle$ is a sequence of vertices $\langle l_1, \ldots, l_n \rangle$ such that $\langle l_i, l_{i+1} \rangle \in \mathcal{E}$. The *source* of a directed path is the first vertex in the path. We denote the set of ancestors for a set A by $An(A)$. The ancestor relation is reflexive and thus $A \subseteq An(A)$.

We define a graph separation criterion called d*separation for directed graphs which is an extension of d-separation (Pearl 1988). An extension of d-separation is required as a pure vertex separation criterion like d-separation cannot separate a vertex from itself which is required to appropriately handle self-edges in directed graphs. A path p d*connects vertices a and b given the set of vertices C in graph $\mathcal{G}$ if every collider on p is in $An(C)$ and every non-collider on p is not in C. For sets of vertices $A, B, C \subseteq \mathcal{L}$ where $A \cap C = \emptyset$ we say that B is d*separated from A by C in graph $\mathcal{G}$ if and only if there does not exist a non-trivial d*connecting path between some $a \in A$ and some $b \in B$ given C in $\mathcal{G}$.

There are two key differences from Pearl's d-separation that allow us to appropriately handle cyclic directed graphs. First, we restrict d*separation statements to sets in which $A \cap C$ is the empty set but allow the sets A and B to overlap. Second, d*connecting paths must be non-trivial. These modifications enable us to use d*separation statements to distinguish between graphs in which there is a self-edge ($a \to a$) and one in which there is not.

We use directed graphs to represent temporal statistical processes. We associate the vertices $\mathcal{L}$ with a set of possible observation types (i.e., things that can happen). The edges denote potential dependencies between observations and the absence of a directed edge from observation type a to observation type b indicates that the process that generates observations of type b does not directly depend on the history of observations of type a. Analogous to the use of d-separation for directed acyclic graphs, we would like a graphical separation criterion for directed graphs to answer questions about how past observations influence future observations. Due in part to the fact that a directed graph does not explicitly encode temporal information we cannot simply apply d*separation on the directed graph. Instead, we define δ*separation which extends the graphical δ-separation of Didelez (2008) to handle self-edges. For sets $A, B, C \subset \mathcal{L}$ where $A \cap C = \emptyset$ we say that that B is *δ*separated* from A given C (or simply $\delta(A, C, B)$) in $\mathcal{G}$ if an only if B is d-separated from A given C in the B-historical dependency graph $\mathcal{G}^B$ where $\mathcal{G}^B = \langle \mathcal{L}, \mathcal{E}^B \rangle$ and $\mathcal{E}^B = \mathcal{E} \backslash \{\langle b, a \rangle \in \mathcal{E} | b \in B, a \neq b\}$. Note that δ*separation is not symmetric in the first and third arguments due to the use of the graph $\mathcal{G}^B$.

3 Learning the Structure of a Causal Process

Our aim is to connect statistical processes with causal graphs and to learn the causal graph governing a system of observed events. We assume that there is a statistical process governing what and when events happen. We denote a statistical process for a set of observation types $\mathcal{L}$ by $\mathcal{P}_{\mathcal{L}}$. We also assume that we can observe the process to determine the whether process independence statements hold. We will write $PI(A, C, B)$ to indicate that the process associated with observations of type B does not depend on the history of observations of type A given the history of observations of type C in a given process $\mathcal{P}_{\mathcal{L}}$ (where $A \cap C = \emptyset$). We write $\neg PI(A, C, B)$ if this is not the case. We call such statements *process independence statements*. We note that process independence statements need not correspond to statistical independence statements and, as with δ*separation, there is no expectation that such process independence statements ought to be symmetric. In this section, we assume the existence of a process independence oracle for the relevant statistical process. In Section 4, we discuss particular statistical processes and the problem of testing process independence statements for those processes.

A process $\mathcal{P}_{\mathcal{L}}$ satisfies the *Causal Factorization Assumption* with respect to a causal process graph $\mathcal{G} = \langle \mathcal{L}, \mathcal{E} \rangle$ if and only if for all $A, B, C \subset \mathcal{L}$ where $A \cap B = \emptyset$ it is the case that $\delta(A, B, C) \Rightarrow PI(A, B, C)$

A process $\mathcal{P}_{\mathcal{L}}$ satisfies the *Causal Dependence Assumption* with respect to a causal process graph $\mathcal{G} = \langle \mathcal{L}, \mathcal{E} \rangle$ if and only if for all $A, B, C \subset \mathcal{L}$ where $A \cap B = \emptyset$ it is the case that $PI(A, B, C) \Rightarrow \delta(A, B, C)$

The Causal Analysis (CA) Algorithm (Algorithm 1) uses a process independence oracle to construct a directed graph. We use $\pi_l^{\mathcal{G}}$ to denote the parents of l in graph $\mathcal{G}$ and $|B|$ to denote the cardinality of the set B. The basic idea is to use process independence statements to remove edges from an initially complete graph. This algorithm is analogous to the PC Algorithm of Spirtes, Glymour and Scheines (2001) but does not have an orientation phase.

Note that the output of the CA algorithm is a directed graph and that any edges presented do not necessarily indicate a causal relationship. In the remainder of this section we explore the interpretation of the output of the CA algorithm under various assumptions. Recall that $a \leftrightarrow b$ simply indicates that $a \rightarrow b$ and $b \rightarrow a$ and not the existence of a latent common cause.

```
Input: A set of events L and a process P_L
Output: A directed graph G
Let G = ⟨L, E⟩ be a complete directed graph.;
foreach l ∈ L do
    Let n = 0;
    foreach l' ∈ π_l^G do
        foreach B ⊆ π_l^G \ {l'} where |B| = n do
            if PI(l', B, l) holds in P_L then
                E = E \ ⟨l', l⟩
            end
        end
        Let n = n + 1;
    end
end
Return G = ⟨L, E⟩;
```

Algorithm 1: The Causal Analysis (CA) Algorithm

Theorem 1 (Complete Observations)**.** *If $\mathcal{P}_{\mathcal{L}}$ satisfies both the causal dependence and factorization assumptions with respect to $\mathcal{G}$ then algorithm $CA(\mathcal{L}, \mathcal{P}_{\mathcal{L}})$ returns $\mathcal{G}' = \mathcal{G}$.*

Lemma 1. *If $\mathcal{P}_{\mathcal{L}}$ satisfies the causal dependence assumption for $\mathcal{G} = \langle \mathcal{L}, \mathcal{E} \rangle$ and algorithm $CA(\mathcal{L}, \mathcal{P}_{\mathcal{L}})$ returns $\mathcal{G}' = \langle \mathcal{L}, \mathcal{E}' \rangle$ then if $l' \rightarrow l \in \mathcal{E}$ then $l' \rightarrow l \in \mathcal{E}'$.*

Lemma 2. *If $\mathcal{P}_{\mathcal{L}}$ satisfies both the causal dependence and factorization assumptions for $\mathcal{G} = \langle \mathcal{L}, \mathcal{E} \rangle$ and algorithm $CA(\mathcal{L}, \mathcal{P}_{\mathcal{L}})$ returns $\mathcal{G}' = \langle \mathcal{L}, \mathcal{E}' \rangle$ then if $l' \rightarrow l \notin \mathcal{E}$ then $l' \rightarrow l \notin \mathcal{E}'$.*

Proof of Theorem 1: The theorem follows from Lemmas 1 and 2.

3.1 Absence of a direct causal relationship

Next we consider the case in which some of the event types in the system are not observed. We let $\mathcal{O} \subseteq \mathcal{L}$ be the set of observed event types. In this case we will assume that the causal factorization and dependence assumptions hold for a process $\mathcal{P}_{\mathcal{L}}$ and some causal process graph $\mathcal{G}$. Our causal factorization and dependence assumptions allow us to focus on δ*separation in $\mathcal{G}$ by assuming that the observed process independence statements accurately reflect the δ*separation statements about $\mathcal{G}$ for the observed observation types. In order to understand and interpret the output of the CA algorithm we need to understand the conditions that lead to edges in the final output. We begin by defining the concept of vertex blockability relative to a set of observed event types.

We say that a vertex a is b-unblockable relative to $\mathcal{O}$ in $\mathcal{G}$ if and only if for all $C \subseteq \mathcal{O} \setminus \{a, b\}$ $\neg\delta(a, C, b)$ is true of $\mathcal{G}$. Otherwise the vertex is said to be b-blockable relative to $\mathcal{O}$. Note that if $b \rightarrow b$ then if $b \in \mathcal{O}$ b is b-unblockable relative to $\mathcal{O}$.

We say that l is a *direct cause* of l' relative to $\mathcal{O}$ for causal process graph $\mathcal{G}$ if and only if there exists a directed path $\langle l_1, \ldots, l_n \rangle$ where $l_1 = l$ and $l_n = l'$ and $l_i \notin \mathcal{O}$ for $(1 < i < n)$. We call the path in the definition of direct cause a *witnessing path* that l is a direct cause of l'. We let D_b denote the set of observed direct causes of the variable b relative to $\mathcal{O}$, that is, members of $\mathcal{O}$ that are direct causes of b relative to $\mathcal{O}$.

Example 1. Let $\mathcal{E} = \{a \rightarrow c, c \rightarrow b\}$, $\mathcal{L} = \{a, b, c\}$ and $\mathcal{O} = \{a, b\}$. The vertex a is b-unblockable relative to $\mathcal{O}$ for $\mathcal{G} = \langle \mathcal{L}, \mathcal{E} \rangle$ but the vertex b is a-blockable relative to $\mathcal{O}$. In this example, a is a direct cause of b relative to $\mathcal{O}$ in graph $\mathcal{G}$ and $a \rightarrow c \rightarrow b$ is a witnessing path for this fact.

Lemma 3. *If l' is a direct cause of l relative to $\mathcal{O}$ in $\mathcal{G}$ then l' is l-unblockable relative to $\mathcal{O}$ in $\mathcal{G}$.*

The following lemma allows us to make causal inferences using the causal analysis algorithm about the absence of a direct causal relationship.

Lemma 4. *If $\mathcal{P}_{\mathcal{L}}$ satisfies the causal dependence assumption with respect to $\mathcal{G}$ then, in the graph $\mathcal{G}'$ output by CA($\mathcal{O}$,$\mathcal{P}_{\mathcal{L}}$), the set of parents for each event type include all of its direct causes relative to $\mathcal{O}$.*

In particular, if the algorithm finds that an event type a is not a parent of event type b then a is not a direct cause of b.

3.2 Causal sufficiency

In the section, we restrict the type of unobserved event types which enables us to make strong inferences about the causal structure of a process. In particular we assume causal sufficiency which is essentially an assumption that there are no latent confounding processes.

A set of event types $\mathcal{O} \subset \mathcal{L}$ is *causally sufficient* with respect to a graph $\mathcal{G} = \langle \mathcal{L}, \mathcal{E} \rangle$ if and only if every common cause of $l, l' \in \mathcal{O}$ is in the set of event types

$\mathcal{O}$.

A directed graph $\mathcal{G}' = \langle \mathcal{O}, \mathcal{E}' \rangle$ is *causally correct* with respect to a graph $\mathcal{G} = \langle \mathcal{L}, \mathcal{E} \rangle$ if for every edge $\langle a, b \rangle \in \mathcal{E}'$ a is a direct cause of b with respect to $\mathcal{O}$ in $\mathcal{G}$.

Theorem 2 (Causal Sufficiency)**.** *If $\mathcal{P}_\mathcal{L}$ satisfies both the causal dependence and factorization assumptions for $\mathcal{G} = \langle \mathcal{L}, \mathcal{E} \rangle$ and $\mathcal{O} \subseteq \mathcal{L}$ is causally sufficient with respect to $\mathcal{G}$ then the graph $\mathcal{G}'$ returned by algorithm $CA(\mathcal{O}, \mathcal{P}_\mathcal{L})$ is causally correct with respect to $\mathcal{G}$ and $\mathcal{O}$.*

Lemma 5. *If $\mathcal{P}_\mathcal{L}$ satisfies the causal dependence and factorization assumptions with respect to $\mathcal{G}$ and $\mathcal{O}$ is causally sufficient for $\mathcal{G}$ then the output of the CA algorithm removes the edge $a \to b$ if a is not a direct cause of b relative to $\mathcal{O}$.*

3.3 Causal insufficiency

We have shown that the CA algorithm can provide causally accurate information under the assumptions of causal sufficiency, causal factorization and causal dependence. In this section we consider removing the assumption of causal sufficiency.

Example 2. Let $\mathcal{E} = \{a \leftarrow c, c \to b\}$, $\mathcal{L} = \{a, b, c\}$ and $\mathcal{O} = \{a, b\}$. The observed event types $\mathcal{O}$ are not causally sufficient for the graph $\mathcal{G} = \langle \mathcal{L}, \mathcal{E} \rangle$. In addition, the CA algorithm fails to provide output that is causally correct. In particular, the CA algorithm yields the graph in which $a \to b$ and $b \to a$ despite the fact that neither is a a cause of b in $\mathcal{G}$ nor is b a cause of a.

Our aim is to graphically characterize vertex separability. We do so using the idea of an inducing path in a directed graph that was introduced for directed acyclic graphs by Verma and Pearl (1990). For a pair of vertices a, b, we define $A_{ab} = An(\{a\}) \cup An(\{b\}) \setminus \{a, b\}$. A path p between $\langle a, b \rangle$ is an *inducing path* relative to $\mathcal{O}$ if and only if (1) every vertex on $p \in \mathcal{O}$ is a collider on p and (2) Every collider on p is in A_{ab}. An inducing path $p = \langle l_1 = a, \ldots, l_n = b \rangle$ from a to b is into b if $l_{n-1} \to l_n$. An inducing path $p = \langle l_1 = a, \ldots, l_n = b \rangle$ from a to b is out of a if $l_1 \to l_2$.

Lemma 6. *For a directed graph $\mathcal{G}$ the following three statements are equivalent:*

(a) *A vertex a is b-unblockable relative to $\mathcal{O}$ in graph $\mathcal{G}$*

(b) *There is an inducing path between a and b relative to $\mathcal{O}$ in graph $\mathcal{G}^b$. Note this inducing path must be into b.*

(c) $\neg\delta(a, \mathcal{O} \cap A_{ab}, b)$ *in $\mathcal{G}$.*

We say that a is a *cause* of b in $\mathcal{G}$ and if there is a directed path from a to b in $\mathcal{G}$.

We aim to find common features of all graphs that are consistent with the observed pattern of process independence statements. Latent processes, however, can mask the causal nature of the observed pattern of dependencies.

For a pair of vertices a, b and graph $\mathcal{G}$ we say that there is a *potential indirect inducing path* into b relative to $\mathcal{O}$ if and only if (1) there is a vertex $c_1 \in \mathcal{O} \setminus \{a, b\}$ such that $a \to b$ in $\mathcal{G}$ and (2) there is a sequence of vertices $c_1, \ldots, c_n \subseteq \mathcal{O} \setminus \{a, c\}$ such that $c_i \leftrightarrow c_{i+1}$ and $c_n \leftrightarrow b$ in $\mathcal{G}$.

Lemma 7. *For any set of observed variable $\mathcal{O}$, if a graph has an inducing path between observed variables a, b into b containing another observed variable then the output of the CA algorithm will contain a potential indirect inducing path into b.*

Theorem 3 (Sufficient Cause)**.** *If $\mathcal{P}_\mathcal{L}$ satisfies both the causal dependence and factorization assumptions for $\mathcal{G} = \langle \mathcal{L}, \mathcal{E} \rangle$ then if CA produces $\mathcal{G}'$ with vertices $\mathcal{O} \subseteq \mathcal{L}$ for which the subgraph over $\{a, b\}$ is $a \to b$ and $\mathcal{G}'$ contains no potential inducing path between a, b into b then a is a cause of b in $\mathcal{G}$.*

Lemma 8. *If $\mathcal{P}_\mathcal{L}$ satisfies both the causal dependence and factorization assumptions for $\mathcal{G} = \langle \mathcal{L}, \mathcal{E} \rangle$ and CA produces $\mathcal{G}'$ with vertices $\mathcal{O} \subseteq \mathcal{L}$ for which the subgraph over $\{a, b, c\}$ is $a \leftrightarrow b \leftrightarrow c$ then*

- *if $PI(a, \emptyset, c)$ and $PI(c, \emptyset, a)$ then there is a latent common causes of a, b and a (possibly distinct) latent cause of b, c and b is not a direct cause of c and b is not a direct cause of a.*
- *if $PI(a, b, c)$ then there is no latent common causes of b, c, b is a cause of c in $\mathcal{G}$.*

4 Statistical Processes and Process Independence

Our approach to causal discovery through the observation of a dynamic process is applicable to different temporal statistical processes. The key connection required is a connection between process independence statements and the observations from a particular statistical process. In this section we consider two distinct statistical processes and discuss process independence for these processes.

4.1 Dynamic Bayesian Networks

Dynamic Bayesian networks (DBNs) are a popular discrete-time model that can capture temporal dynamics of a statistical process. A DBN is a statistical model of an infinite set of variables indexed by time. A variable X_i^t denotes the i^{th} variable at time t. We use $X = X_1, \ldots, X_n$ to denote the set of *variable types* in the DBN, that is, a variable with an unspecified time component and X^t to denote the set of variables at time t. The DBN specifies the evolution of X^t as a stochastic function of the value of previous variables X^{t-i} $(i > 0)$. In particular, the variable X_i^t is a stochastic function of the value of its parents in a graph. The causal process graph associated with a causal DBN is a graph over the variable types of the DBN X where there is an edge $X_i \to X_j$ if there exists a t, i such that there is an edge $X_i^{t-i} \to X_j$ in the DBN. Thus, the parent relationship of the causal process graph captures the dependence of a variable type on the history of other variable types. Furthermore, process independence statements $PI(X_i, C, X_j)$ correspond to a set of independence statements of the form $I(X_i^1, \ldots, X_i^{t-1}, X_C^1, \ldots, X_C^{t-1}, X_j^t)$. Without further assumptions, testing process independence would be unfeasible but if we focus on stationary processes with finite temporal dependency we can potentially test process independence statements.

4.2 Graphical Event Models

In this section, we define Conditional Intensity Models and Graphical Event Models (GEMs) and connect these models with previous work on the class of Piecewise-Constant Conditional Intensity Models and Poisson Networks. We assume that events of different types are distinguished by labels l drawn from a finite alphabet $\mathcal{L}$. An event is then composed of a non-negative time-stamp t and a label l. A *history* is an event sequence $h = \{(t_i, l_i)\}_{i=1}^n$ where $0 < t_1 < \cdots < t_n$, and our data is a specific history denoted by $\mathcal{D}$. Given data $\mathcal{D}$, we define the *history at time t* as $h(t, \mathcal{D}) = \{(t_i, l_i) \mid (t_i, l_i) \in \mathcal{D}, t_i \leq t\}$. We suppress $\mathcal{D}$ from $h(t, \mathcal{D})$ when clear from context and write $h_i = h(t_{i-1})$. By convention $t_0 = 0$. We define the *ending time* $t(h)$ of a history h as the time of the last event in h: $t(h) = \max_{(t,l)\in h} t$ so that $t(h_i) = t_{i-1}$.

A *Conditional Intensity Model* (CIM) is a set of non-negative *conditional intensity functions* indexed by label $\{\lambda_l(t|h;\theta)\}_{l\in\mathcal{L}}$. The data likelihood for this model is

$$p(\mathcal{D}|\theta) = \prod_{l\in\mathcal{L}} \prod_{i=1}^{n} \lambda_l(t_i|h_i,\theta)^{\mathbf{1}_l(l_i)} e^{-\Lambda_l(t_i|h_i;\theta)} \qquad (1)$$

where $\Lambda_l(t|h;\theta) = \int_{-\infty}^{t} \lambda_l(\tau|h;\theta)d\tau$ and the function $\mathbf{1}_l(l')$ is one if $l' = l$ and zero otherwise. The conditional intensities are assumed to satisfy $\lambda_l(t|h;\theta) = 0$ for $t \leq t(h)$ to ensure that $t_i > t_{i-1} = t(h_i)$. These modeling assumptions are quite weak. In fact, any distribution for $\mathcal{D}$ in which the timestamps are continuous random variables can be written in this form. For more details see [1, 2]. Despite the fact that the modeling assumptions are weak, these models offer a powerful approach for decomposing the dependencies of different event types on the past. In particular, this per label conditional specification allows one to model detailed label-specific dependence on past events.

Next we define a graphical conditional intensity model that we call a graphical event model (GEM). A filtered history for $A \subseteq \mathcal{L}$ as $[h]_A = \{(t_i, l_i)|(t_i, l_i) \in h \wedge l_i \in A\}$. A GEM is a pair $< \mathcal{G}, \theta >$, where $\mathcal{G} =< \mathcal{L}, \mathcal{E} >$ is a directed graph over a set of event types and edges in $\mathcal{E}$ represent potential dependencies among event types. The parameters $\theta = \{\theta_l\}_{l\in\mathcal{L}}$ parameterize the intensity functions for each event type. In particular, $\lambda_l(t|h_t, \theta_l) = \lambda_l(t|[h_t]_{\pi_l}, \theta_l)$ where π_l is the set of parents for l in $\mathcal{G}$. As in the case of the DBN, a process independence statement correspond to testing a dependence of an event type on set of event histories. One potential approach to testing a process independence $PI(a, C, b)$ is to estimate/learn an intensity function for b using the event histories for $\{a\} \cup C$ and see if the intensity model depends on the event history for a. The work by Gunawardana et al (2011) on learning piecewise continuous intensity models is a good starting point for this approach.

5 Discussion

One of the goals for the research direction described in this paper is the development a sound approach to causal inference for dynamic systems. One of the popular extant approaches is that of Granger causality which fails on this account. This approach is typically applied in a discrete-time continuous valued time-series and, thus, can be viewed as a dynamic Bayesian network. Roughly speaking, in a multivariate time series X a set of variables are the Granger-causes of X_j if the historical values of this set of variables (including X_j) are necessary and sufficient for optimal prediction. Unfortunately this approach does not appropriately handle latent common causes. In particular, for both of the scenarios described in Lemma 8 it is the case that each of the variables is a Granger cause of its neighbors while this relationships need not be causal as the lemma demonstrates. In fact, it is easy to construct stochastic processes with latent factors which demonstrate that the inferential approach to Granger causality is not sound with respect to causal relations.

There has been much work related to causal discovery and the estimation of causal effects in time-series. As discussed above, the work on Granger causality (Granger 1969) is the most well known. The short-comings of this approach are also well known (e.g., Eichler 2007) and there has been some work in trying to address these known short comings. For instance, Eichler (2007) proposes a similar approach to the approach described here but differs in that it allows for the possibility of "simultaneous correlation" which requires the use of an alternative definition of separation. In addition, while providing definitions of cause and spurious cause, sufficient conditions for the identification of causal relationships are not presented. The work of Entner and Hoyer (2010) considers the problem of causal discovery from time series data using limited dependence vector autoregressive models and the FCI algorithm that uses conditional independence tests to identify the structure. Our approach of using δ*separation is inspired by the work of Didelez (2008) who defined δ-separation and shows the connection between that graphical separation criterion and local independence of marked point processes. Our extension to δ*separation allows for the appropriate treatment of self-edges which are essential in any self-excitatory or self-inhibitory dynamic process. Another more loosely connected work is that of Eichler and Didelez (2007) that considers the estimation of causal effects based on an intervention in a time-series.

While the results described in this paper offer hope for developing a methodologically sound approach to causal inference for dynamic systems, there is much work that needs to be done. Here are some of the open research questions.

- Non-parametric tests for process independence for various type of temporal statistical processes

- Soundness and completeness results for δ*separation analogous to those provided by Pearl (1988), Meek (1995) and Spirtes et al (2001) for d-separation. Note that Didelez (2008) has shown the soundness of δ-separation for a family of marked point processes related to GEMs.

- A representation for equivalence classes of causal graphs with respect to δ*separation in the case of causal insufficiency ($\mathcal{O} \subset \mathcal{L}$) analogous to those developed by Verma and Pearl (1990) and Spirtes et al (2001) that captures the common casual aspects of the set of graphs in the equivalence class.

Acknowledgments

Thanks to Asela Gunawardana and two anonymous reviewers for their comments on an earlier draft of this paper.

References

[1] D. J. Daley and D. Vere-Jones. *An Introduction to the Theory of Point Processes: Elementary Theory and Methods*, volume I. Springer, second edition, 2002.

[2] Vanessa Didelez. Graphical models for marked point processes based on local independence. *JRSS-B*, 70(1):245–264, 2008.

[3] Michael Eichler. Granger causality and path diagrams for multivariate time series. *Journal of Econometrics*, 137:334–353, 2007.

[4] Michael Eichler and Vanessa Didelez. Causal reasoning in graphical time series models. In *Uncertainty in Artificial Intelligence*, pages 109–116, 2007.

[5] Doris Entner and Patrik O. Hoyer. On causal discovery from time series data using FCI. In *Probabilistic Graphical Models*, pages 121–128, 2010.

[6] C.W.J. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37:424–438, 1969.

[7] Asela Gunawardana, Christopher Meek, and Puyang Xu. A model for temporal dependencies in event streams. In *Advances in Neural Information Processing Systems*, 2011.

[8] C. Meek. Strong completeness and faithfulness in Bayesian networks. In *Proceedings of Eleventh Conference on Uncertainty in Artificial Intelligence,* Montreal, QU, pages 411–418. Morgan Kaufmann, August 1995.

[9] P. Spirtes, C. Glymour, and R. Scheines. *Causation, Prediction, and Search, Second Edition.* MIT Press, Cambridge, MA, second edition, 2001.

[10] T. Verma and J. Pearl. Equivalence and synthesis of causal models. In *Proceedings of Sixth Conference on Uncertainty in Artificial Intelligence,* Boston, MA, pages 220–227. Morgan Kaufmann, July 1990.