The 12th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU), Utrecht, The Netherlands, 2013

Detecting marginal and conditional independencies between events and learning their causal structure.

Jan Lemeire[1,4], Stijn Meganck[1,3], Albrecht Zimmermann[2], and Thomas Dhollander[3]

[1] ETRO Department, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium
{jan.lemeire, stijn.meganck}@vub.ac.be
[2] Katholieke Universiteit Leuven, Celestijnenlaan 200A, 3001 Leuven, Belgium
albrecht.zimmermann@cs.kuleuven.be
[3] D square N.V., Kempische Steenweg 297 bus 3, B-3500 Hasselt, Belgium
thomas.dhollander@dsquare.be
[4] iMinds, Department of Future Media and Imaging, Gaston Crommenlaan 8 (box 102), B-9050 Ghent Belgium

Abstract. Consider data given as a sequence of events, where each event has a timestamp and is of a specific type. We introduce a test for detecting marginal independence between events of two given types and for conditional independence when conditioned on one type. The independence test is based on comparing the delays between two successive events of the given types with the delays that would occur in the independent situation. We define a Causal Event Model (CEM) for modeling the event-generating mechanisms. The model is based on the assumption that events are either spontaneous or caused by others and that the causal mechanisms depend on the event type. The causal structure is defined by a directed graph which may contain cycles. Based on the independence test, an algorithm is designed to uncover the causal structure. The results show many similarities with Bayesian network theory, except that the order of events has to be taken into account. Experiments on simulated data show the accuracy of the test and the correctness of the learning algorithm when assumed that the spontaneous events are generated by a Poisson process.

1 Introduction

In this paper we consider the following problem. The data is a sequence of events $\mathcal{E} = \langle (E_1, t_1), (E_2, t_2), \ldots \rangle$ where E_i represents an event type and t_i, the time of occurrence (also called timestamp) of the ith event, is a real value $\in [0, T]$, with T the end time of the sequence. E_i take values from a finite set of event types, the event domain $\mathcal{D}$. Fig. 1 shows an example event sequence with $\mathcal{D} = \{A, B, C, D\}$. When there is no confusion possible we denote events (E_i, t_i) with lower case e_i. Event types are denoted with upper case and sets with boldface letters. The question is to infer (1) independencies and (2) causal relations between the events.

If events of type A can cause events of type B, which we write as $A \rightarrow B$, then sequences $\langle t_{A_1}, t_{A_2} \ldots t_{A_k} \rangle$ and $\langle t_{B_1}, t_{B_2} \ldots t_{B_l} \rangle$ are correlated, where t_{A_i} and t_{B_j} are the timestamps of the A and B events respectively. We want a test to identify such correlation. We also want a test to identify conditional independencies. For causal model

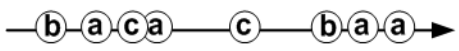

Fig. 1. Example event sequence.

$A \rightarrow C \rightarrow B$, A is independent from B when conditioned on C, which we write as $A \perp\!\!\!\perp B|C$.

The problem has extensively been studied for series of continuous-valued dynamic variables, see for instance Granger causality [1]. Methods for analyzing sequences of events, on the other hand, have been studied in the data mining community. The main technique is episode mining where an *episode* is defined as an ordered tuple of events. Occurrences of episodes are counted and highly-frequent episodes are considered as relevant.

The independence test we propose here is based on the information given by the intervals between successive events of a given episode. The intervals measured from the data will be compared with the intervals in the case in which the events would be generated independently. If both interval arrays appear as being generated from different distributions, the events are correlated.

Our approach for learning the causal structure is similar to the approach as used in causal model theory in which a causal model is represented by a Bayesian network [2, 3]. In Bayesian network theory, conditional independencies are defined over the joint probability distribution and a link is drawn between causality and dependencies through the causal Markov condition. The conditional independencies following from the causal structure can then be used to learn the causal structure from data.

In the next section we define a Causal Event Model for reflecting the event-generating mechanisms. We show that it is more general than current settings. In Section 3 we define marginal and conditional independence between events. Section 4 draws the link between causation and correlation in our framework. Section 5 defines the conditional independence tests. Section 6 gives a causal structure learning algorithm and Section 7 provides experiments with simulated data.

2 Causal Event Model

The model for the event-generating mechanisms is based on the following assumptions. (a) Events have exogenous causes (called *spontaneous events*) or are caused by other (*effect events*). (b) The causal mechanisms depend on the type of event. This does not mean that event c literally causes event e. It is possible that the mechanism responsible for generating event c (e.g. when a variable passes a certain threshold) affects another mechanism which triggers event e. In such case, the event related to the cause can happen after the effect event. Here, (c) we will assume that cause events happen before their effects. (d) The causal mechanism only generates one event (or none) of a specific type.

The effect event counterfactually depends on the cause events; if one of the causes would not have happened, the effect event would not have happened. The event sequence $\mathcal{E}$ can then be split up into two sequences: the spontaneous events $\mathcal{E}_s$ and the

effect events $\mathcal{E}_e$. Instantiations of events belong to either $\mathcal{E}_s$ or $\mathcal{E}_e$, however events of a certain type can occur in both. Each non-spontaneous event has one or more causes: an effect event is linked to one or more events. $\forall e_i \in \mathcal{E}_e, \exists \boldsymbol{c} \subset \mathcal{E} : \boldsymbol{c} \rightarrow e_i$. This is indicated in Fig. 2. We call it the *Causal Event Sequence Graph* (CESG). It constitutes a Directed Acyclic Graph (DAG).

Fig. 2. Example event sequence with the causal relations between the events.

On the other hand, a graph describing the mechanisms responsible for generating the effect events should not be cyclic. We only assumed that the mechanism depend on the event types, in the sense that events from a certain type are responsible for generating events from another type in the future. So, $A \rightarrow B$ means that some events of type B are caused by events of type A. If an A event, say e_i, causes a B event, say e_j, then t_j depends on t_i. This is represented by $P(\triangle^* t_B|A)$, a probability distribution over $\triangle^* t_B$ which is defined as the time interval between cause and effect, $t_j - t_i$ in the case of e_i causing e_j. The asterisk denotes that it is an interval between causally-connected events. The probability distribution can often be described by a Weibull distribution. It should be noted that the sum $P_{total} = \int_t P(\triangle^* t_B = t|A)dt$ can be smaller than 1, indicating that A in some cases does not generate B. By defining P over the time difference, time invariance is incorporated into our system.

The causal structure can be represented by a directed graph which can be cyclic, and can have bidirected edges or loops. Fig. 3 shows the causal structure responsible for the event sequence of Fig. 2. The parameterization is that for each node X and for all parents Pa of X, there is $P(\triangle^* t_X|Pa)$ which specifies a distribution of the time delay. These distributions represent the generation of X by Pa. This is shown in Fig. 4. If X has multiple parents, they can all independently generate X or the generation of X happens by a mutual occurrence of multiple parent events. $Pa_1 \ldots Pa_k \rightarrow X$ is described by $P(\triangle^* t_X|Pa_1, \triangle t_{Pa_2}, \ldots, \triangle t_{Pa_k})$. The distribution gives the time to X after the occurrence of Pa_1 and occurrence of Pa_i $(i = 2 \ldots k)$ with a time difference of $\triangle t_{Pa_i}$.

The directed graph together we the parameterization we call a *Causal Event Model* (CEM). The CEM can be considered as a generic template to produce the CESG, which is often called the 'rollout'.

2.1 Related work

Temporal Nodes Bayesian Networks (TNBNs) [4] are a special kind of Bayesian network which are parameterized considering delays (relative times). When an initial (spontaneous) event occurs, its occurrence gives the reference time. The nodes represent variables. Events occur when variables pass a certain threshold. This is indeed often the case, but we do not want to make any assumption about the 'meaning' of the events and use event variables as nodes.

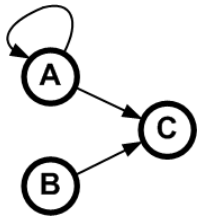

Fig. 3. Causal structure used for the experiments.

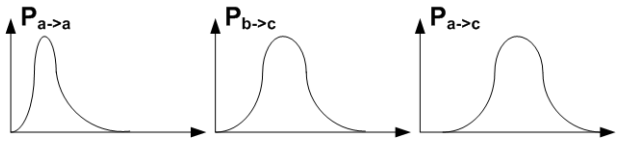

Fig. 4. Parameterization for some families of the causal structure of Fig. 3.

Networks of Probabilistic Events in Discrete Time (NPEDTs) [5] are also defined over event variables with a parameterization similar to ours. NPEDTs are, however, more restrictive: each event can happen at most once and no self-references or cycles are allowed in the graph. By this restriction, NPEDTs are genuine Bayesian networks, while a CEM is not.

[6] uses a dynamic Bayesian network to model the relations between the event variables. The limitation of a dynamic Bayesian network is that you need to draw a link between a node and another node in the future. This fixes the time interval between cause and effect. [6] use the dynamic Bayesian network in combination with episode mining, but because of this limitation, they limit themselves to fixed-delay episodes. In our case we make no assumption about the time interval between the cause and effect event. We even allow continuous time intervals.

Finally, it must be noted that all 3 models here discussed discretize the time.

3 Independence relationships

We define marginal and conditional independence on the distributions over the time intervals between successive events. This is motivated by the following. Events generate new events that will happen in the future. The causal mechanism defines the time interval between cause C and effect E, which we denoted by $\triangle^* t_E$. The knowledge of an event happening at time t contains information on the occurrence of the causally-related future events. We therefore consider the time to the first future occurrence of an event of a specific type. We denote this time delay as $\triangle^1 t_E$.

3.1 Interval to first occurrence ($\triangle^1$)

We denote by $P_{ab}(\triangle^1 t_B|A)$ the probability that the first event of type B after a random time t happens at time $t + \triangle^1 t_B$ given that an event of type A occurred at time t. The

subscript ab indicates that event b must happen after a. The definitions for independency will be based on this distribution instead of causal distributions $\triangle^* t$, because it is measurable from data. Before that, two important consequences of the model have to be discussed.

It must be noted that if $A \rightarrow B$, this does not necessarily mean that $P(\triangle^* t_B|A)$ and $P_{ab}(\triangle^1 t_B|A)$ are the same, since the first event B after event A might be a spontaneous event or caused by other events, and accidently occurring right after A. The relation between $P(\triangle^* t_B|A)$ and $P_{ab}(\triangle^1 t_B|A)$ is calculated as follows. The probability for having the *first* occurrence of a B event at relative time t is the probability that B occurs at time t multiplied with the probability that no B occurred before that. This is expressed by the following equation. With $P(\triangle t_B = t|A)$ we denote the probability that a B event happens at time t after an A event.

$$P(\triangle^1 t_B = t|A) = P(\triangle t_B = t|A).(1 - P(\triangle^1 t_B < t|A)) \quad (1)$$

$$= (\sum_i P_i(\triangle^* t_B = t|A)).(1 - \int_{t'=0}^{t'=t} P(\triangle^1 t_B = t'|A)dt') \quad (2)$$

The first probability of the right hand side of Eq. 1 is determined by all possible direct causes of B (denoted by index i), the second is an integral adding all previous probabilities. It results in a recursive formula, given by Eq. 2. If the probability $P(\triangle t_B = t|A)$ is a constant, the result is an exponential distribution.

Next, assume the causal model is $A \rightarrow B$ and A is spontaneously generated by a Poisson process with rate λ. The first event B after an event A can then be (1) the event caused by that A or (2) a B event caused by another A event. For the latter holds that $P_2(\triangle t_B = t|A) = P(t_B = t) = \lambda$ since it is unrelated to A. $P(\triangle^1 t_B = t|A)$ is a combination of both given by above equation. The resulting distribution mainly depends on which distribution 'comes first'. The distribution with most of its weight for small delays greatly determines $P(\triangle^1 t_B = t|A)$.

3.2 Marginal independence

Marginal independence is defined as follows:

$$A \underset{ab}{\perp\!\!\!\perp} B \Leftrightarrow P_{ab}(\triangle^1 t_B|A) = P(\triangle^1 t_B) \quad (3)$$

where $P(\triangle^1 t_B)$ is the probability that the first event of type B after time t happens at time $t + \triangle^1 t_B$ given a random time t.

An important difference with statistical independence defined over a joint probability distribution is that the order should be taken into account: $A \underset{ab}{\perp\!\!\!\perp} B$ means that knowledge about an A event has no information on the next B event, while $\underset{ba}{\perp\!\!\!\perp}$ is about information of an B event over the next A event. Hence:

$$A \underset{ab}{\perp\!\!\!\perp} B \not\Leftrightarrow A \underset{ba}{\perp\!\!\!\perp} B. \quad (4)$$

While it can be shown that symmetry holds for a given order:

$$A \underset{ab}{\perp\!\!\!\perp} B \Leftrightarrow B \underset{ab}{\perp\!\!\!\perp} A \quad (5)$$

A special case is autocorrelation. $P(\triangle^1 t_B|B)$ is the probability that the first event of type B after a random time t happens at time $t + \triangle^1 t_B$ given that another event of type B occurred at time t. To check for autocorrelation we check whether $P(\triangle^1 t_B|B) = P(\triangle^1 t_B)$. We denote B autocorrelated as $\overset{\frown}{B}$.

3.3 Conditional independence

Conditional independence is also defined for a specific event ordering over its arguments. The order is described by an episode.

$$\begin{aligned} A \underset{ep(A,B,\boldsymbol{S})}{\perp\!\!\!\perp} B|\boldsymbol{S} \Leftrightarrow \\ P_{e=ep(A,B,\boldsymbol{S})}(\triangle^1 t_B|s^*, \triangle^1(\boldsymbol{S} \setminus s^*), \triangle^1 t_A) \\ = P_{e\setminus A}(\triangle^1 t_B|s^*, \triangle^1(\boldsymbol{S} \setminus s^*)) \end{aligned} \tag{6}$$

with $ep(A, B, \boldsymbol{S})$ an episode over A, B and $\boldsymbol{S}$, and s^* the first element of $\boldsymbol{S}$ in the episode, all $\triangle$s are defined with respect of s^*. $e \setminus A$ denotes the episode e from which A is removed. Note that the distributions do not depend on the choice of s^* among $\boldsymbol{S}$; it only sets the reference time.

4 Causation implies correlation and vice versa

In this section we draw the relation between causation and correlation as defined in the previous sections. The relation is grounded by the assumption that causally unrelated events are independent.

This assumption is also expressed by Reichenbach's principle: if A and B are correlated, then either A causes B, either B causes A or either there is a common cause of A and B.

In the following we will also assume that there are no unknown (latent) common causes. Together with the independence assumption this implies that there is no correlation if no causal relation in the model. Except that the spontaneous events from a specific type will be autocorrelated when their occurrence is not random.

4.1 Correlation and the causal event sequence graph.

Consider $\mathcal{I}(cem)$ the conditional independencies of a causal event model, consider $CEM(G)$ all causal event models compatible with graph G. We are interested in the conditional independencies that hold for all $CEMs$ compatible with G (the intersection): $\mathcal{I}(G) = \cap_{CEM(G)}\mathcal{I}(cem)$. These independencies follow from the causal structure, independent from the parameterization. Specific parameterizations may lead to additional independencies.

The following theorem proves that the conditional independence statements from $\mathcal{I}(G)$ can be extracted from the Causal Event Sequence Graphs (CESG) compatible with G by d-separation. We recall the definition. $\boldsymbol{X}$ and $\boldsymbol{Y}$ are d-separated by $\boldsymbol{Z}$ if every path between $\boldsymbol{X}$ and $\boldsymbol{Y}$ is blocked by $\boldsymbol{Z}$. An (undirected) path is said to be blocked by $\boldsymbol{Z}$

if it contains a collider $\rightarrow \cdot \leftarrow$ whose descendants are not in $\boldsymbol{Z}$ or a non-collider $\rightarrow \cdot \rightarrow$ or $\leftarrow \cdot \rightarrow$ or $\leftarrow \cdot \leftarrow$ that is in $\boldsymbol{Z}$ [2].

Theorem 1. *$A \underset{ep(A,B,\boldsymbol{S})}{\perp\!\!\!\perp} B|\boldsymbol{S}$ is not in $\mathcal{I}(G)$ if and only if there is a subset of nodes in a Causal Event Sequence Graph compatible with G which forms an occurrence of the episode $ep(A, B, \boldsymbol{S})$ in the event sequence such that $a \not\perp b|\boldsymbol{s}$.*

Proof. If there is an active path between a and b in the sequence graph, we prove that t_a and t_b are bounded by the causal delay distributions of the network. Then there exists at least one parameterization which bounds the occurrence of b to the time of occurrence of a, such that the independence does not hold. For any triple $x \rightarrow y \rightarrow z$ along the path, t_z is bounded by t_y and also by t_x. The same applies for $x \leftarrow y \leftarrow z$. For any triple $x \leftarrow y \rightarrow z$ along the path, both t_z and t_x are bounded by t_y which makes them also depend on each other. For any triple $x \rightarrow y \leftarrow z$, t_y is bounded by t_x and t_z but this does not imply that t_x and t_z are dependent unless y or one of its descendants is conditioned on. In that case, t_y is known and together with t_x this gives information about t_z. Combining these bounds proves that an active path implies a conditional dependency. If, on the other hand, there is no path, then a and b events are assumed to be independent. If there is a path, but blocked by an event, say c, then t_c constrains t_b, but t_a does not further bounds t_c.

4.2 Correlation and the causal event model.

d-separation is not readily usable to identify conditional independencies from the Causal Event Model. A related criterion will be established here.

Definition 1 **(d-separation in CEM).** *A path p between two nodes A and B is said to be blocked by a set $\boldsymbol{S} = \{\boldsymbol{S}_1, \boldsymbol{S}_2, \boldsymbol{S}_3\}$, with $\boldsymbol{S}_1, \boldsymbol{S}_3 \subset \boldsymbol{E}$ and $\boldsymbol{S}_2 \subset \boldsymbol{E} \backslash \{A, B\}$, corresponding to an ordered episode $(\boldsymbol{s}_1, a, \boldsymbol{s}_2, b, \boldsymbol{s}_3)$ if:*

- *on p there is a fork $X \leftarrow Y \rightarrow Z$ and $Y \in S_1$*
- *on p there is a chain $X \rightarrow Y \rightarrow Z$ and $Y \in S_2$*

and

- *on p there is no collider $X \rightarrow Y \leftarrow Z$ for which either Y or any of its descendants $\in S_3$*

When all paths between A and B are blocked by $\boldsymbol{S} = \{\boldsymbol{S}_1, \boldsymbol{S}_2, \boldsymbol{S}_3\}$ we say that A is d-separated from B given $\boldsymbol{S}$ denoted as $A \underset{\boldsymbol{s}_1 a \boldsymbol{s}_2 b \boldsymbol{s}_3}{\perp} B|\boldsymbol{S}_1\boldsymbol{S}_2\boldsymbol{S}_3$, otherwise we call them d-connected denoted as $A \underset{\boldsymbol{s}_1 a \boldsymbol{s}_2 b \boldsymbol{s}_3}{\not\perp} B|\boldsymbol{S}_1\boldsymbol{S}_2\boldsymbol{S}_3$.

Theorem 2. *$A \underset{\boldsymbol{s}_1 a \boldsymbol{s}_2 b \boldsymbol{s}_3}{\perp\!\!\!\perp} B|\boldsymbol{S}_1\boldsymbol{S}_2\boldsymbol{S}_3 \Leftrightarrow A \underset{\boldsymbol{s}_1 a \boldsymbol{s}_2 b \boldsymbol{s}_3}{\perp} B|\boldsymbol{S}_1\boldsymbol{S}_2\boldsymbol{S}_3$, with $\boldsymbol{S}_1, \boldsymbol{S}_2, \boldsymbol{S}_3 \subset \boldsymbol{E}$.*

Proof. $\Leftarrow$
Assume $A \underset{\mathbf{s}_1 a \mathbf{s}_2 b \mathbf{s}_3}{\perp} B|\mathbf{S}_1\mathbf{S}_2\mathbf{S}_3$. Conditioning on events in $\mathbf{S}_3$ cannot block a path between A and B in the CEM. For each $e \in \mathbf{S}_1$, in the corresponding CESG e will appear before

Algorithm 1 Marginal independence test for $A \underset{ab}{\perp\!\!\!\perp} B$.

Given: Set of possible event types $\mathcal{D}$ and event sequence $\mathcal{E} = \langle (E_1, t_1), \ldots, (E_n, t_n) \rangle$

1. Count the number of occurrences of B in $S = n$.
2. Generate a new sequence S' with the same A events as in S, and add n events of type B with random timestamp $\in [0, T]$. If A and B are the same (self-correlation test), sequence S' should only contain the randomly generated events.
3. For both sequences S and S', generate the sequence of intervals I and I' between each occurrence of A and the first occurrence of B after that of A.
4. If the Kolmogorov-Smirnov test applied on I and I' returns 'equal', the test returns true (meaning independence).

a and b. If in the CESG there is a causally directed path from e to both a and b then conditioning on e closes the path $a \leftarrow \ldots \leftarrow e \rightarrow \ldots \rightarrow b$. If there is no directed path from e to both a and b in the CESG then a is trivially d-separated from b. Similar observations can be made for $e \in \mathbf{S}_2$, where there either is a causally directed path $a \rightarrow \ldots \rightarrow e \rightarrow \ldots \rightarrow b$ or a is again trivially d-separated from b in the corresponding CESG. Therefore $\Leftarrow$ follows from Theorem 1.

$\Rightarrow$

Assume $A \underset{\mathbf{s}_1 a \mathbf{s}_2 b \mathbf{s}_3}{\perp\!\!\!\perp} B|\mathbf{S}_1\mathbf{S}_2\mathbf{S}_3$ and $A \underset{\mathbf{s}_1 a \mathbf{s}_2 b \mathbf{s}_3}{\not\perp\!\!\!\perp} B|\mathbf{S}_1\mathbf{S}_2\mathbf{S}_3$. This implies that $A \underset{\mathbf{s}_1 a \mathbf{s}_2 b \mathbf{s}_3}{\perp} B|\mathbf{S}_1\mathbf{S}_2\mathbf{S}_3$ in the corresponding CESG (Theorem 1). Conditioning on events in $\mathbf{S}_1$ and $\mathbf{S}_2$ cannot d-connect A and B in the CEM, so $A \underset{ab\mathbf{s}_3}{\not\perp\!\!\!\perp} B|\mathbf{S}_3$. This means that there is a $E \in \mathbf{S}_3$ such that there is a causally directed path from both A and B to E, or an edge $A \rightarrow B$ in the CEM. This however is contradicted by the lack of such paths in the corresponding CESG.

5 Independence test

The goal is to define a test which identifies $.\underset{ep}{\perp\!\!\!\perp}.|.$ such that in the generic case: $\perp \Leftrightarrow \underset{ep}{\perp\!\!\!\perp}$.

For testing $X \underset{\cdot\cdot}{\perp\!\!\!\perp} Y|\mathbf{Z}$, we have to compare the distribution $P_{ep(Y,\mathbf{Z})}(\triangle^1 t_Y|\mathbf{Z})$ reflecting the independence situation, with the actual distribution $P_{ep(X,Y,\mathbf{Z})}(\triangle^1 t_Y|\mathbf{Z}, X)$ estimated from the data. We will use the Kolmogorov-Smirnov test which works on the data directly. The test identifies whether two samples are drawn from the same distribution, without making any assumption about the distribution of data. The exact significance probability is calculated using the method of [7].

Note that all tests have linear complexity with respect to the sequence size.

5.1 Marginal independence

The algorithm is described by Alg. 1. The algorithm measures the distribution over $\triangle^1$, which is different from that of $\triangle^*$ but as discussed in Sec. 3.1, $\triangle^*$ comes close to $\triangle^1$ if the average delta is smaller than that of the other causes.

Algorithm 2 Conditional independence test for $A \underset{ep(abc)}{\perp\!\!\!\perp} B|C$.

Given: Set of possible event types $\mathcal{D}$ and event sequence $\mathcal{E} = \langle (E_1, t_1), \ldots, (E_n, t_n) \rangle$

1. For each occurrence of episode $ep(abc)$, add interval $t_b - t_a$ to sequence I, add interval $t_c - t_a$ to sequence I_1 and $t_b - t_c$ to I_2.
2. Now shuffle sequence I_1 randomly such that the order of the elements gets completely different from that of I_2.
3. Construct sequence I' by adding the elements of I_1 to those of I_2 (interval i of I_1 is summed with interval i of I_2).
4. If the Kolmogorov-Smirnov test applied on I and I' returns 'equal', the test returns true (meaning conditional independence).

5.2 Conditional independence

Next, we give an algorithm to test for independence when conditioned on one variable. The test is based on randomization of the intervals with respect to the time of occurrence of the conditioning variable. This creates the reference distribution. Alg. 2 describes the test procedure. If the occurrence of an A event is irrelevant for the occurrence of B when C is known, the time interval between A and B is irrelevant with respect of C. The distribution for the null hypothesis is then constructed by randomizing (swapping) the intervals between A and B.

6 Causal structure learning

Here we present a modified version of the PC algorithm to detect the causal structure called EPC: Algorithm 3. We define a complete directed CEM as a complete graph with all bi-directed edges and self-references for each variable. The description of the algorithm simplifies as each bi-directed edge $A \leftrightarrow B$ is considered as two edges $A \rightarrow B$ and $A \leftarrow B$. Since we can directly make a difference between these edges by looking at ab and ba episodes (through the asymmetry, see Section 2), we do not have to add an orientation phase or end up with a class of equivalent models under the given independencies such as the PC algorithm.

All CIs discovered in the data are following from the causal structure.

Theorem 3. *Under faithfulness, the EPC algorithm returns the correct CEM given an oracle for the independence tests.*

Proof. By faithfulness, no adjacent nodes can become independent when conditioned on any subset of other nodes. Non-adjacent nodes are either marginally independent or become independent when conditioned on one of the nodes along each path.

7 Experiments and Evaluation

In this section we experimentally analyze the accuracy of the independence test and the learning algorithm. It is then compared with the results obtained by episode mining.

Algorithm 3 EPC

Given: Set of possible event types $\mathcal{D}$ and event sequence $\mathcal{E} = \langle (E_1, t_1), \ldots, (E_n, t_n) \rangle$

1. Initialization with complete directed CEM G over $\mathcal{D}$
2. For each edge $A \to B$ in G (including $\overset{\frown}{A}$, i.e. B=A),
   (Consider each bi-directed edge $A \leftrightarrow B$ as two edges $A \to B$ and $A \leftarrow B$)
   $\forall \mathbf{S}_1 \subset \mathcal{D}$ and $\forall \mathbf{S}_2 \subset \mathcal{D} \backslash \{B\}$:
   If $A \underset{\mathbf{s}_1 a \mathbf{s}_2 b}{\perp\!\!\!\perp} B | \mathbf{S}_1, \mathbf{S}_2$, remove $A \to B$ from G

7.1 Influence of causal delay and sample size

As discussed in Sec. 3.1, the delay between cause and effect plays an important role in identifying $\triangle^*$ from $\triangle^1$. We experimentally studied this with data generated from model $X \to Y \to Z$ with the following parameterization. X is generated by a Poisson process with rate 0.01 (meaning that on average every 100 time units an event occurs). The parameterization of both causal relations, $X \to Y$ and $Y \to Z$, is given by a Gaussian distribution with given mean and the standard deviation is set to the square root of the mean (a mean of 100 thus gives the same average delay as that of the Poisson processes). There is a probability of 0.2 that no effect event is generated. Table 1 shows the minimum episode occurrences necessary to correctly identify the given dependencies in 10 experiments. A dash means that the minimum count exceeded 4000.

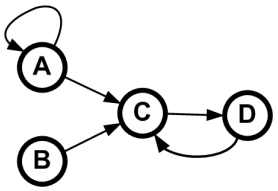

Fig. 5. Causal structure of the example event sequence.

Secondly, data was generated from the model of Fig. 5. The parameterization was set as follow. Spontaneous events are generated for A, B and D by a Poisson process with rate 0.01. The parameterization of each causal relation (each edge), $P(\triangle^* E|C)$, is given by a Gaussian distribution with given mean and the standard deviation set to be the square root of the mean. For all causal relations there is a probability of 0.2 that no effect event is generated, except for $A \to A$ and $D \to C$ for which the probability of no effect is set to 0.8 to avoid cascading effects due to the cycles. The results are also shown in Table 1.

The results show clearly that the detection of dependencies is accurate for small sample sizes for simple models but becomes harder when multiple causes are into play, such as for detecting the dependency between A and D. The self-correlation of A is also hard to detect since the causal relation is only fired with low probability (0.2). Finally, it

must be noted that the test rarely makes errors on detecting conditional independencies. For model $X \to Y \to Z$ and over all experiments, testing $X \underset{xyz}{\perp\!\!\!\perp} Z|Y$ gave an accuracy of 99.3%. The same accuracy was obtained when testing $A \underset{acd}{\perp\!\!\!\perp} D|C$ in the second model.

Table 1. *Number of episodes necessary to correctly identify the following (in)dependencies that hold for the given models with varying Gaussian mean. The lowest row shows the minimal sequence size to correctly learn model $X \to Y \to Z$.*

mean	20	40	60	80	100	120
$X \underset{xy}{\not\perp\!\!\!\perp} Y$	19	20	58	59	56	138
$X \underset{xz}{\not\perp\!\!\!\perp} Z$	59	135	137	297	303	1242
$X \underset{xyz}{\not\perp\!\!\!\perp} Y\|Z$	185	186	185	89	90	87
$A \underset{ac}{\not\perp\!\!\!\perp} C$	25	75	378	790	3181	-
$A \underset{ad}{\not\perp\!\!\!\perp} D$	53	941	2341	-	-	-
$A \underset{abc}{\not\perp\!\!\!\perp} B\|C$	127	159	126	129	128	121
$A \underset{acd}{\not\perp\!\!\!\perp} C\|D$	583	565	267	260	263	257
$A \underset{aa}{\not\perp\!\!\!\perp} A$	375	1223	3695	-	-	-
learning	3068	686	685	677	668	690

7.2 Causal structure Learning

The lowest row of Table 1 shows the minimal sequence size (number of events) to correctly learn model $X \to Y \to Z$ with the parameterization specified in the previous section. It shows that only a relatively small sample size is needed to learn simple models. The high sample size needed to learn the model with mean 20 is needed since the delays come close to being deterministic (small standard deviation) which results in violations of faithfulness.

The accuracy of the learning algorithm depends on the correctness of the independence test and the validity of faithfulness. This was confirmed by our experiments with randomly-generated models and different sample sizes. The following causes of failure were detected:

1. Large causal delays and small sample sizes increase the number of test errors, as discussed in the previous section. When for a single event multiple causes come into play, it's harder to detect the dependencies.
2. Exact violations of faithfulness or near-to-unfaithful situations. For example, when the causal delay is nearly deterministic. These cases are similar to the problems in learning causally-interpreted Bayesian networks. See for instance [8] for discussion of the problems and modifications of the PC algorithm to handle such violations.
3. Finally, our conditional independence test is limited to one conditioning variable. This means that if two variables are related by two difference causal paths, they

are dependent and will not become independent when conditioned on only one variable. It introduces a false positive edge.

8 Conclusions

We created a procedure with linear complexity for testing marginal and conditional independence between events in event sequences. The test is accurate in detecting dependencies coming from causal relations if the average interval between cause and effects is smaller than that of spontaneous events or other causes. We defined a very general model, the Causal Event Model (CEM), to describe the underlying event-generating mechanisms. As opposed to other event models, it is not a Bayesian network since it allows cycles. Based on the conditional independencies an algorithm could be constructed to learn the correct causal structure under faithfulness and causal sufficiency.

9 Acknowledgements

This research was partly funded by the Prognostics for Optimal Maintenance (POM) project (grant nr. 100031; `www.pom-sbo.org`) which is financially supported by the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen).

References

1. Granger, C.W.J.: Investigating causal relations by econometric models and cross-spectral methods. Econometrica **37**(3) (1969) 424–438
2. Pearl, J.: Causality. Models, Reasoning, and Inference. Cambridge University Press (2000)
3. Spirtes, P., Glymour, C., Scheines, R.: Causation, Prediction, and Search. 2nd edn. Springer Verlag (1993)
4. Arroyo-Figueroa, G., Sucar, L.E.: Temporal bayesian network of events for diagnosis and prediction in dynamic domains. Applied Intelligence **23** (October 2005) 77–86
5. Galán, S.F., Díez, F.J.: Networks of probabilistic events in discrete time. Int. J. Approx. Reasoning **30**(3) (2002) 181–202
6. Patnaik, D., Laxman, S., Ramakrishnan, N.: Inferring dynamic bayesian networks using frequent episode mining. CoRR (2009)
7. Nikiforov, A.: Algorithm as 288: Exact smirnov two-sample tests for arbitrary distributions. Applied Statistics **43** (1994) 265–284
8. Lemeire, J., Meganck, S., Cartella, F., Liu, T.: Conservative independence-based causal structure learning in absence of adjacency faithfulness. Int. J. Approx. Reasoning **53**(9) (2012) 1305–1325