
Causal Reasoning for Events in Continuous Time: A Decision–Theoretic Approach

Vanessa Didelez
School of Mathematics
University of Bristol
Vanessa.Didelez@bristol.ac.uk

Abstract

The dynamics of events occurring in continuous time can be modelled using marked point processes, or multi-state processes. Here, we review and extend the work of Røysland et al. (2015) on causal reasoning with local independence graphs for marked point processes in the context of survival analysis. We relate the results to the decision-theoretic approach of Dawid & Didelez (2010) using influence diagrams, and present additional identifying conditions.

1 INTRODUCTION

Dynamic dependence structures among the occurrence of different types of events in continuous time can be represented by local independence graphs as developed by Didelez (2006, 2007, 2008). In related work, Røysland (2011, 2012) showed how causal inference based on inverse probability weighting (IPW), well known for longitudinal data (Robins et al., 2000), can be extended to the continuous-time situation using a martingale approach. Røysland et al. (2015) combine these and give graphical rules for the identifiability of the effect of interventions, which in the context of events in time take the form of changes to the intensities of specific processes, e.g. a treatment process.

As we discuss here, the approach of Røysland et al. (2015) can be seen as the time-continuous version of Dawid & Didelez (2010), who develop a decision theoretic approach for sequential decisions in longitudinal settings and use a graphical representation with influence diagrams that include decision nodes. This provides an explicit representation of the target of inference as well as allowing us to use simple graphical rules to check identifiability.

2 LOCAL INDEPENDENCE GRAPHS

The notion of dynamic dependence on which we focus here can be stated as follows. For stochastic processes $X(t), Y(t), Z(t)$ we say informally that $X(t)$ is locally independent of $Y(t)$ given $Z(t)$ if the present of $X(t)$ is independent of the past of $Y(t)$ given the past of both $X(t), Z(t)$. Slightly more formally we can write this as

$$X(t) \perp\!\!\!\perp \mathcal{F}_{t^-}^Y \mid \mathcal{F}_{t^-}^{X,Z}$$

where \mathcal{F}_t^k are filtrations generated by $X_k(t)$, i.e. the sets of information becoming available over time. Note that this is an asymmetric type of independence as discussed in detail in Didelez (2006).

Marked Point Processes

More formally we consider a marked point process (MPP) to describe the occurrence of different types of events \mathcal{E} ; this can be represented by a set of counting processes $\{N_j(t)\}$ for each type of event $j \in \mathcal{E}$. It may often be too detailed to model the dependence structure between all possible types of events; e.g. the event ‘stop treatment’ can necessarily only happen after the event ‘start treatment’ and the two events are therefore trivially dependent. Instead of a MPP one can therefore group certain events together to obtain a multi-state process with several components $\mathbf{Y}_V(t) = \mathbf{Y}(t) = (Y_1(t), \dots, Y_K(t))$, $V = 1, \dots, K$, where e.g. $Y_k(t)$ describes the treatment process with states ‘on / off treatment’. Note that the components $Y_k(t)$ need to be such that none of them systematically change state at the same time, i.e. $\mathbf{Y}(t)$ is *composable* (see Didelez, 2007). Further each $Y_k(t)$ can be described by a set of counting processes, one for each change of state, so that the whole $\mathbf{Y}(t)$ is itself an MPP. In the following we will not clearly distinguish between a component $Y_k(t)$ of a composable multi-state process, or a counting process $N_j(t)$ for an individual event.

Under mild regularity conditions, the Doob–Meyer Theorem tells us that each counting process can be decomposed:

$$Y_k(t) = \underbrace{\Lambda_k(t)}_{\text{predictable}} + \underbrace{M_k(t)}_{\text{martingale}},$$

where $\Lambda_k(t)$ is predictable based on the history \mathcal{F}_{t-}^V of whole \mathbf{Y}_V and $M_k(t)$ is an \mathcal{F}_t^V -martingale. We will assume that the \mathcal{F}_t^V -intensity processes $\lambda_k(t)$ exist and have the following interpretation:

$$\Lambda_k(t) = \int_0^t \lambda_k(s) ds, \quad \lambda_k(t) dt = E(N_k(dt) | \mathcal{F}_{t-}^V).$$

Local Independence

From the above we see that $\lambda_k(t)$ fully describes the dependence of a process' infinitesimal short-term expectation on the past. Any independencies must therefore be reflected in the structure of the intensity; if we find, for instance, that $\lambda_k(t)$ remains unchanged regardless of whether an event of type $j \neq k$ has occurred in the past, then we say there is a local independence.

Indeed, the formal definition is that Y_k is *locally independent* of Y_j given $\mathbf{Y}_{V \setminus \{j,k\}}$ if $\lambda_k(t)$ is $\mathcal{F}_t^{V \setminus \{j\}}$ -measurable, i.e. the intensity process remains the same when information on the past of Y_j is omitted. We write this as $Y_j \not\rightarrow Y_k | \mathbf{Y}_{V \setminus \{j,k\}}$. Note that $\mathcal{F}_t^{V \setminus \{j\}}$ always contains the past of the component Y_k itself. Meek's (2014) approach allows for cases where $\lambda_k(t)$ is $\mathcal{F}_t^{V \setminus \{k\}}$ -measurable.

Graphs and δ -Separation

The *local independence graph* $G = (V, E)$ of a multi-state process $Y_V(t) = (Y_1(t), \dots, Y_K(t))$ (or an MPP) is given such that the absence of a directed edge indicates a local independence, i.e.

$$(j, k) \notin E \Rightarrow Y_j \not\rightarrow Y_k | \mathbf{Y}_{V \setminus \{j,k\}}.$$

The resulting graphs are directed, can have two directed edges between any two vertices, and can have cycles. Note that $\text{pa}(k) \cap \text{ch}(k) \neq \emptyset$ is possible, and similar for ancestors and descendants etc.

Under regularity conditions, the definition implies that the intensity process λ_k for Y_k is $\mathcal{F}^{\text{cl}(k)}$ -measurable (Didelez, 2008), where $\text{cl}(k)$ is the closure (i.e. the set of parents and k itself).

As for conditional independence graphs, certain separations on a local independence graph imply further local independencies. However, a different notion of separation is required, δ -separation: define G^B as the graph obtained after deleting all edges emanating from

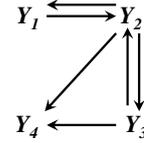


Figure 1: A local independence graph.

nodes in set B ; then we say that C δ -separates A from B in the local independence graph G if it separates A and B in the undirected graph $(G_{\text{An}(A \cup B \cup C)}^B)^m$ obtained by moralising the subgraph of G^B on the ancestral set $\text{An}(A \cup B \cup C)$. Note that δ -separation is asymmetric, i.e. δ -separating A from B is not the same as B from A . Meek (2014) introduces self-edges so to be able to distinguish the case where a process is locally independent of itself or not, and generalises the above to δ^* -separation.

A key result of Didelez (2008) is that, under mild regularity conditions, we have for subsets $A, B, C \subset V$:

$$\text{if } C \text{ } \delta\text{-separates } A \text{ from } B \text{ then } Y_A \not\rightarrow Y_B | Y_C.$$

The above is not obvious as the \mathcal{F}_t^V -intensity and the $\mathcal{F}_t^{A \cup B \cup C}$ -intensity of a process can be very different.

Example I: The graph in Figure 1 encodes for instance that $Y_1 \not\rightarrow Y_4 | (Y_2, Y_3)$. Using δ -separation we can verify that this is not preserved without Y_3 , i.e. it is not the case that $Y_1 \not\rightarrow Y_4 | (Y_2)$. This is because of the ‘selection effect’: knowing something about the past of $Y_2(t)$ makes the past of $Y_1(t)$ informative for past of $Y_3(t)$ and therefore predictive of $Y_4(t)$.

3 CAUSAL VALIDITY

So far we described a notion, and graphical representation, of dynamic (in)dependence based on how the present of a subprocess depends or not on the past of other processes; in other words, a notion of time-lagged (in)dependence. As it is based on the intensity process it can be considered as characterised by infinitesimal short-term predictions, which is very much parallel to so-called ‘Granger-causality’ (Granger, 1969). However, much of the causal inference literature formalises causality in terms of (sometimes hypothetical) interventions. For instance a DAG is termed causal if the set of variables \mathbf{X}_V is sufficiently ‘rich’ so that an intervention that changes how a variable X_k is generated corresponds to replacing $p(x_k | x_{\text{pa}(k)})$ with a different $\tilde{p}(x_k)$ in the factorisation

$$p(\mathbf{x}_V) = \prod_{i \in V} p(x_i | \mathbf{x}_{\text{pa}(i)}).$$

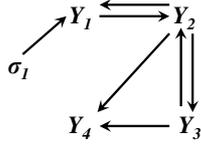


Figure 2: An augmented local independence graph with intervention indicator σ_1 .

Røysland et al. (2015) extend this notion of intervention to local independence graphs by assuming that the intervention replaces the intensity process λ_k of Y_k by a different one $\tilde{\lambda}_k$, which will typically be measurable with respect to a smaller subset of processes, e.g. those relevant to and observable by the decision maker.

Remember that for a given local independence graph G , each intensity process λ_k is $\mathcal{F}^{\text{cl}(k)}$ -measurable. Røysland et al. (2015) then define this graph to be *causally valid* for an intervention in Y_k if this corresponds to replacing λ_k by $\tilde{\lambda}_k$ while *all other intensities* λ_j , $j \neq k$, remain the same under the intervention.

Intervention Indicator

In analogy to the influence diagrams of Dawid (2002, 2012), it can be helpful to indicate graphically that an intervention modifying the intensity of Y_k is being considered, by adding an intervention node σ_k . For the basic set-up chosen here, σ_k would itself not be a process and simply take values in $\{o, e\}$ to indicate the original system with intensity λ_k when $\sigma_k = o$, or the intervened system with intensity $\tilde{\lambda}_k$ when $\sigma_k = e$. The absence of any edges involving σ_k other than $\sigma_k \rightarrow Y_k$ then represents the causal validity assumption, in analogy to *extended stability* of Dawid & Didelez (2010).

Example I (ctd.): The graph in Figure 2 is augmented with the intervention node σ_1 to indicate that Y_1 is subject to possibly different intensities in the two different regimes. The absence of edges between σ_1 and other nodes indicates that their observational $\mathcal{F}_t^{\text{cl}(k)}$ -intensities remain the same under intervention.

Re-Weighting

Similar to the case of longitudinal data, it turns out that inference about the dynamics between events under the intervened system can be obtained by re-weighting. Specifically the weights are given as $W(t) :=$

$$\prod_{s \leq t} \left(\frac{\tilde{\lambda}_k(s)}{\lambda_k(s)} \right)^{\Delta N_k(s)} \exp \left(\int_0^t \lambda_k(s) - \tilde{\lambda}_k(s) ds \right).$$

For these to be well-defined, in particular for $\tilde{P} \ll P$, we need $W(t)$ to be uniformly integrable which can be interpreted as $\lambda_k(t), \tilde{\lambda}_k(t)$ not being ‘too different’, e.g. $W(t)$ could be uniformly bounded. In fact, if $\Lambda_k(t)$ is assumed absolutely continuous such that $\lambda_k(t)$ exists, then it is e.g. not possible to re-weight with an intervention that has discrete jumps of $N_k(t)$ at fixed time points. Note that this can be regarded as correspondent of the ‘positivity’ condition typically made in many causal inference contexts.

Censoring and Re-Weighting

In the context of survival or duration data it is almost inevitable to have censoring (e.g. due to the end of the study). Censoring in itself can be regarded as an event and modelled with a counting process that jumps when the observation is censored. This then allows us to express assumptions about the censoring in terms of its intensity process. A common assumption is *independent censoring* which can be stated as the relevant process (e.g. survival) being locally independent of the censoring process, possibly conditional on other observed processes. The most obvious violation of this assumption occurs when there are unobserved common causes for censoring and survival.

Moreover, censoring can be linked to the above ideas of intervention and re-weighting in the following sense. The target of inference is typically a population where no censoring occurs (e.g. future patients) or where censoring is entirely random and stochastically independent of other processes. Hence we can say that the target is to replace the censoring intensity by a different intensity that does not depend on the past. When this is possible given the observed processes therefore depends among others on whether the local independence graph on all events including censoring is causally valid wrt. the censoring process. Røysland et al. (2015) discuss this further and give an example where censoring is independent, but based on a local independence graph that is not causally valid and hence leading to incorrect inference. For the remainder of the paper here we do not further consider censoring.

4 IDENTIFICATION

In the following we assume that the index set of processes is $V = V_0 \cup X \cup L \cup U$ where V_0 are observable processes of interest (‘outcome’ processes), X (or counting process N_X) is the process in which we want to intervene changing its intensity, L is a set of observable processes in which we are not interested, and U is a set of unobservable processes.

Definition 1:

Let G be the local independence graph for processes $V = V_0 \cup X \cup L \cup U$; assume causal validity wrt. X . Consider an intervention in X that changes its observational \mathcal{F}^V -intensity λ_X to a \mathcal{F}^{V_0} -intensity $\tilde{\lambda}_X$. We say that the effect of such an intervention on V_0 is *identified* by L if the \mathcal{F}^{V_0} -intensities for every counting process $N \in V_0$ under the intervention exist and are given by re-weighting with the above weights $W(t)$.

Røysland et al. (2015) show the following sufficient condition for identification:

Proposition 2:

In the situation of Definition 1, if $U \not\rightarrow X \mid (V_0 \cup L)$, then the effect on V_0 of intervening in X is identified by L .

Example I (ctd.): In Figure 2, assume we are interested in the effect of an intervention in $X = Y_1$ on $V_0 = Y_4$ and let $L = Y_2$ and $U = Y_3$. Then we see that Proposition 2 is satisfied, meaning that re-weighting will allow us to compute aspects of the possibly modified behaviour of Y_4 under an intervention that changes the intensity process of Y_1 , where the weights require no observation of Y_3 .

The condition of Proposition 2 is the point process analogue of *sequential randomisation* in Dawid & Didelez (2010); it is in fact satisfied iff $U \cap \text{pa}(X) = \emptyset$. In other words, it formalises the notion that given the past of observed processes, $X(t)$ is at any time t independent of the past of unobserved processes. Dawid & Didelez (2010) show that this implies ‘simple stability’ which in turn is a sufficient identifying condition for sequential interventions in their longitudinal (time-discrete) setting. Here, we define the time-continuous marked point process analogue as follows.

Definition 3:

With the preconditions of Definition 1, and the augmented local independence graph G^σ with intervention node σ_X , we define that *simple stability* holds if

$$\sigma_X \not\rightarrow (L \cup V_0) \mid X.$$

We conjecture that identification can in fact be obtained under the wider assumption of simple stability.

Conjecture 4:

Assume the preconditions of Definition 1, and the augmented local independence graph G^σ (i.e. causal validity wrt. X).

If simple stability holds, then the effect on V_0 of intervening in X is identified by L .

Corollary 5:

The condition of Proposition 2 implies simple stability.

We can now formulate a result corresponding to Dawid

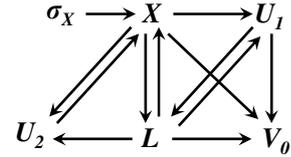


Figure 3: An augmented local independence satisfying simple stability.

& Didelez’ (2010) notion of ‘sequential irrelevance’; this condition allows unobserved processes in U to affect the treatment process X as long as they are ‘irrelevant’ to the other processes of interest.

Corollary 6:

Assume the preconditions of Definition 1, and the augmented local independence graph G^σ (i.e. causal validity wrt. X). Then $U \not\rightarrow (V_0 \cup L) \mid X$ implies simple stability.

Both, Corollary 5 and 6 are sufficient but not necessary for simple stability as the following example demonstrates.

Example II: The graph in Figure 3 shows a situation where $U = (U_1, U_2)$ satisfies neither Proposition 2 nor Corollary 6. However, simple stability is satisfied. Note that U_1 alone fulfills Corollary 6 and U_2 alone Proposition 2. All these would be destroyed by an edge between U_1 and U_2 .

5 DISCUSSION

More generality? In the time-discrete case, more general conditions for causal effect identification can and have been given than those analogous to simple stability. Specific to sequential decisions in longitudinal data these are for example addressed in Pearl & Robins (1995), Robins (1997), Dawid & Didelez (2010; section 8). It appears not straightforward to generalise these to the time-continuous situation with local independence graphs considered here, as it assumes stationarity of the dependence structure, while such more general criteria are typically relevant when the structure changes over time. However, it is possible to generalise local independence graphs to some extent in order to take non-stationarity of (in)dependencies into account, e.g. some independencies might hold before a certain event has happened and others afterwards leading to a sequence of graphs that are valid in intervals defined by stopping times (Didelez, 2008).

Why an intervention indicator? The decision theoretic approach to causality makes it formally and graphically explicit that an intervention in a particular node is being considered and what assumptions are involved

(Dawid, 2012). This allows greater clarity, e.g. regarding the target of inference; but in our case it also allows to formulate conditions for identification that do not need to refer to or characterise unobservable processes U . The flip side is that one might miss an intuition for what kinds of U violate the conditions, which may impede justifying the assumption of simple stability. Here, we have linked the results to the notions of sequential randomisation / irrelevance of U which provide some intuition.

Causal Search? We assumed that the local independence graph is given and that subject matter knowledge justifies causal validity wrt. certain events or processes. Meek (2014) addresses learning the graph. Under a completeness assumption this is in principle (i.e. given an oracle test for local independence) straightforward as there are no issues of Markov-equivalence due to the asymmetry of local independence in time, i.e. all edges can easily be oriented. Meek (2014) further gives results for cases of unobserved processes, e.g. causal insufficiency. However, the main practical problem in any real application will be a suitable test for local independence. In low-dimensional settings with few events, this can be done almost non-parametrically e.g. by testing equality of survival-curves; but in higher dimensions this becomes prohibitive. One could make simplifying assumptions, such as assuming a Markov process; in this context it is important to be aware that if $\mathbf{Y}_V(t)$ is Markov, then a subprocess $\mathbf{Y}_A(t)$, $A \subset V$ is typically not.

APPENDIX

Proof of Conjecture 4: see Røysland & Didelez (2015).

Proof of Corollary 5:

Remember that in the augmented local independence graph G^σ , assuming causal validity wrt. X , there is only a single edge involving σ_X pointing into X . Further, the condition of Proposition 2 is satisfied iff $U \cap \text{pa}(X) = \emptyset$ in G . The graphical check of δ -separation for simple stability involves removing all outgoing edges from $V_0 \cup L$; in the resulting graph before moralisation, there are no edges into X except the one from σ_X . Hence, in the moral graph, σ_X only has an edge with X and Definition 3 is satisfied.

Proof of Corollary 6:

As above, in the augmented local independence graph G^σ there is only a single edge involving σ_X pointing into X . The graphical check of δ -separation for simple stability, furthermore, involves removing all outgoing edges out of $V_0 \cup L$ and with the condition of Corollary 6 this means that there are no edges

between U and $V_0 \cup L$ at all. Hence, even if there are moral edges between σ_X and U these do not lead to paths between $V_0 \cup L$ and σ_X in the relevant moral graph and Definition 3 is satisfied.

Acknowledgement

Financial support from the Leverhulme Trust (RF-2011-320) is gratefully acknowledged.

References

- Dawid, A.P. (2002). Influence diagrams for causal modelling and inference. *International Statistical Review*, 70:161-189.
- Dawid, A.P. (2012). The Decision-Theoretic Approach to Causal Inference. Chapter 4 in: *Causality – Statistical Perspectives and Applications* (eds. C.Berzuini, A.P.Dawid, L.Bernardinelli), Wiley.
- Dawid, A.P., Didelez, V. (2010). Identifying the consequences of dynamic treatment strategies: A decision theoretic overview. *Statistics Surveys*, 4:184-231.
- Didelez, V. (2006). Asymmetric separation for local independence graphs. In *Proc. of 22nd UAI Conference*, 130-137. AUAI Press.
- Didelez, V. (2007). Graphical models for composable finite Markov processes. *Scandinavian Journal of Statistics*, 34:169-185.
- Didelez, V. (2008). Graphical models for marked point processes based on local independence. *JRSSB*, 70(1):245-264.
- Granger, C.W.J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* 37:424-438.
- Meek, C. (2014). Toward learning graphical and causal process models. In *Proc. of 31th UAI Conference Causality Workshop*, 43-48. AUAI Press.
- Pearl, J., Robins, J. (1995). Probabilistic evaluation of sequential plans from causal models with hidden variables. In *Proc of 11th UAI Conference*, 444-453. Morgan Kaufmann Publishers, San Francisco.
- Robins, J.M. (1997). Causal inference from complex longitudinal data. In: *Latent Variable Modeling and Applications to Causality*, (ed. M. Berkane). *Lecture Notes in Statistics* 120:69-117. Springer, New York.

- Robins, J.M., Hernan, M.A., Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology*, 11:550-560.
- Røysland, K. (2011). A martingale approach to continuous time marginal structural models. *Bernoulli*, 17(3):895-915.
- Røysland, K. (2012). Counterfactual analyses with graphical models based on local independence. *Annals of Statistics*, 40(4):2162-2194.
- Røysland, K., Didelez, V., Nygard, M., Lange, T., Aalen, O.O. (2015). Causal reasoning in survival analysis: re-weighting and local independence graphs. *Submitted*.
- Røysland, K., Didelez, V. (2015). General criteria for identification of causal effects between events in continuous time. *In preparation*.