



Didelez:

Local independence graphs for composable Markov processes

Sonderforschungsbereich 386, Paper 158 (1999)

Online unter: <http://epub.ub.uni-muenchen.de/>

Projektpartner



Local independence graphs for composable Markov processes

Vanessa Didelez
University of Munich, Institute of Statistics
Ludwigstrasse 33, D-80539 Munich, didelez@stat.uni-muenchen.de

ABSTRACT:

The concept of local independence is used to define local independence graphs representing the dynamic dependence structure of several continuous time processes which jointly form a so-called composable Markov process. Specific properties of this new class of graphs are discussed such as the role of separating sets. Further insight is gained by considering possible extensions to the discrete time situation. It is shown that the latter case can be reduced to classical graphical interaction models.

1 Introduction

This paper proposes a modelling technique for multistate Markov processes which uses graphs to visualize the underlying dependence structure. In order to speak of dependences the notion of compossibility is introduced by formalizing the assumption that the whole process consists of different components. This has first been defined by Schweder (1970) and applied to event history analysis for example by Aalen et al. (1980). In general terms, local independence means that a component of the process is independent of the past of another component given its own past and possibly the past of the remaining components. In the context of event history analysis this is exactly what is meant by saying that an event does not depend on the prior occurrence of another event, i.e. the intensity is not altered by the past event. A graphical representation should capture all such independences and possibly further properties of the underlying statistical model.

Classical graphical models as treated for instance by Lauritzen (1996) are based on conditional stochastical independence relations among the involved variables and it is the conditional independence structure that is represented by the graph. In contrast, local independence graphs are based on local independence relations among the components of a composable stochastic process, i.e. the vertices represent no random variables but stochastic processes or events which are registered by suitable processes. The emphasis of the present paper lies on the discussion of the graphical Markov properties for such local independence graphs and on possible extensions to the discrete time situation.

The outline of the paper is as follows. First, the concept of local independence for composable Markov processes is defined and discussed following Schweder (1970). Based on this, local independence graphs are defined in Section 3. In this context, we present the implications of separation and give conditions for the equivalence of different independence properties suggested by the graph in analogy to the pairwise, local, and global Markov properties in classical graphs. In Section 4 we treat the discrete time situation and indicate two possible approaches which are derived from the concept of local independence for the continuous time situation. As will be seen, the first approach appears somewhat unsuited for practical purposes but instead indicates the generalization of the concept of local independence graphs to non-Markov processes. The second approach assumes that a continuous time Markov process underlies the discrete one and derives its properties for a given local independence structure of the underlying process. Finally, we discuss the concept of local independence graphs with respect to estimation, inclusion of covariate information, and causality aspects.

2 Local independence in Markov processes

In this and the following section we consider Markov processes $Y = \{Y(t)|t \in \mathcal{T}\}$ with finite state space \mathcal{S} and continuous time scale, i.e. $\mathcal{T} = [0, \tau]$. The transition intensities are given by

$$\alpha_t(y; y') = \lim_{h \downarrow 0} \frac{1}{h} P(Y(t+h) = y' | Y(t) = y), \quad y \neq y' \in \mathcal{S}.$$

We assume that the transition intensities exist, i.e.

$$\alpha_t(y; y') < \infty \quad \forall y \neq y', \tag{1}$$

and that they are continuous and bounded functions of t on any closed interval in \mathcal{T} . If the process is regarded as consisting of different components it will be denoted by bold face letters as in the following definition.

Definition 2.1 *Composable finite Markov process, CFMP (Schweder, 1970)*

Let $\mathbf{Y} = \{\mathbf{Y}(t)|t \in \mathcal{T}\}$ be a Markov process with finite state space \mathcal{S} and transition intensities that hold (1). Let further $V = \{1, \dots, K\}$, $K \geq 2$, and assume that there are K spaces \mathcal{S}_j , $j \in V$, with $|\mathcal{S}_j| \geq 2$, and that there exists a one-to-one mapping f of \mathcal{S} onto $\bigotimes_{j \in V} \mathcal{S}_j$ so that elements $y \in \mathcal{S}$ can be identified with elements $(y_1, \dots, y_K) \in \bigotimes_{j \in V} \mathcal{S}_j$. Then, \mathbf{Y} is a *composable finite Markov process (CFMP)* with components Y_1, \dots, Y_K such that $f(\mathbf{Y}(t)) = (Y_1(t), \dots, Y_K(t))$ if $\forall A \subseteq V$, $|A| \geq 2$,

$$\lim_{h \downarrow 0} \frac{1}{h} P \left(\bigcap_{j \in A} \{Y_j(t+h) \neq y_j\} \mid \bigcap_{j \in A} \{Y_j(t) = y_j\} \right) = 0 \tag{2}$$

$\forall y_j \in \mathcal{S}_j$, $j \in V$, and $t \in \mathcal{T}$. Composability is denoted by $\mathbf{Y} \sim (Y_1, \dots, Y_K)$. //

The definition states that for a composable process the probability that more than one component changes in a short period of length h is of magnitude $o(h)$. This formalizes the intuitive claim that the different components should not describe the same or similar phenomenons or events. Consequently, in continuous time they should have probability zero of changing states at the same time. Note that the composition is not necessarily unique. If for example $\mathbf{Y} \sim (Y_1, \dots, Y_K)$ then $\mathbf{Y} \sim (\mathbf{Y}_A, \mathbf{Y}_{V \setminus A})$. In the following we treat $f(\mathbf{y}) = (y_1, \dots, y_K)$ as if $\mathbf{y} = (y_1, \dots, y_K)$.

Definition 2.1 implies that the transition intensities of a CFMP have the following properties.

Corollary 2.2 *Transition intensities for CFMP*

Let $\mathbf{Y} \sim (Y_1, \dots, Y_K)$ be a CFMP.

1. The intensity $\alpha_t(\mathbf{y}; \mathbf{y}')$, $t \in \mathcal{T}$, for any $\mathbf{y} \neq \mathbf{y}'$ is given by

$$\alpha_t(\mathbf{y}; \mathbf{y}') = \begin{cases} \alpha_t^j(\mathbf{y}; y'_j), & y_j \neq y'_j \wedge \mathbf{y}_{-j} = \mathbf{y}'_{-j}, \quad j \in V \\ 0, & \text{otherwise} \end{cases}$$

with $\mathbf{y}_{-j} = (y_1, \dots, y_{j-1}, y_{j+1}, \dots, y_K)$, i.e. the intensity equals 0 if \mathbf{y} and \mathbf{y}' differ on more than one component, where

$$\alpha_t^j(\mathbf{y}; y'_j) = \lim_{h \downarrow 0} \frac{1}{h} P(Y_j(t+h) = y'_j | \mathbf{Y}(t) = \mathbf{y}).$$

2. The total intensity $\bar{\alpha}_t(\mathbf{y})$, $t \in \mathcal{T}$, is given as a continuous and bounded function of t with

$$\bar{\alpha}_t(\mathbf{y}) = \lim_{h \downarrow 0} \frac{1}{h} (1 - P(\mathbf{Y}(t+h) = \mathbf{y} | \mathbf{Y}(t) = \mathbf{y})) = - \sum_{j \in V} \sum_{y'_j \neq y_j} \alpha_t^j(\mathbf{y}; y'_j).$$

//

The dependence structure of the components Y_1, \dots, Y_K is thus determined by the quantities $\alpha_t^j(\mathbf{y}; y'_j)$, $\mathbf{y} \in \mathcal{S}$, $y'_j \in \mathcal{S}_j$, $j \in V$. Since

$$h \cdot \alpha_t^j(\mathbf{y}; y'_j) \approx P(Y_j(t+h) = y'_j | \mathbf{Y}(t) = \mathbf{y})$$

we can say, that if $\alpha_t^j(\mathbf{y}; y'_j)$ is independent of some components of the first argument \mathbf{y} this also holds for the probability $P(Y_j(t+h) = y'_j | \mathbf{Y}(t) = \mathbf{y})$ of an instantaneous change in Y_j , and one would intuitively speak of local independence. This is formalized in the following definition.

Definition 2.3 *Local independence in a CFMP (Schweder, 1970)*

Let $\mathbf{Y} \sim (Y_1, \dots, Y_K)$ be a CFMP. Then, Y_j is *locally independent* of Y_k , $k \neq j$, if $\forall \mathbf{y}_{-k} \in \mathcal{S}_{-k}$ and $y'_j \in \mathcal{S}_j$, $y'_j \neq y_j$, $\alpha_t^j(\mathbf{y}; y'_j)$ is a constant function of y_k . This is denoted by $Y_j \perp^l Y_k$. Otherwise, Y_j is locally dependent on Y_k and we write $Y_j \not\perp^l Y_k$.

For $A \subset V$ we have that $\mathbf{Y} \sim (\mathbf{Y}_A, \mathbf{Y}_{V \setminus A})$ so that the vector-valued local independence $\mathbf{Y}_A \perp^l \mathbf{Y}_B$ is defined by $\alpha_t^A(\mathbf{y}; \mathbf{y}'_A)$ being a constant function of \mathbf{y}_B in the first argument $\forall \mathbf{y}_{-B} \in \mathcal{S}_{-B}$ and $\mathbf{y}'_A \in \mathcal{S}_A$, $\mathbf{y}'_A \neq \mathbf{y}_A$, $B \subseteq V \setminus A$. //

A more general definition which is not restricted to Markov processes would postulate that the presence of Y_j is independent of the past of Y_k given past of Y_{-k} . Similar concepts can be found in time series analysis (Granger, 1969) or in the framework of general stochastic processes considered in Mykland (1986).

Remark 2.4 *Local independence: not marginal*

Note that the property $Y_j \perp \!\!\! \perp Y_k$ is not a marginal property of the two involved components since it may depend on the other components \mathbf{y}_{-k} in $\alpha_t^j(\mathbf{y}; y'_j)$. If for example the information contained in \mathbf{y}_{-k} is altered by discarding some of these components then local independence of Y_j on Y_k might not be preserved. In general, the reduced process $\mathbf{Y}'(t) \sim (Y_j(t), Y_k(t))$ is neither a Markov process nor does $Y_j \perp \!\!\! \perp Y_k$ have to hold with respect to \mathbf{Y}' . //

As can easily be checked, local independence is neither necessarily symmetric nor reflexive nor transitive. Thus, we make the following assumption.

Assumption 2.5 *Reflexivity*

Since in most practical situations a process depends at least on its own past we only consider stochastic processes where local dependence is reflexive. //

Let us present two short examples within the framework of event history analysis where local independence could be of special interest. The first one describes the situation of several non-recurrent events. Aalen et al. (1980) consider a data example where the interesting events are menopause, subdivided into induced and natural menopause, and occurrence of a certain skin disease (pustulosis palmo-plantaris). This is modelled as a Markov process with states 0, i.e. no event has occurred so far, states (M) , (I) , or (D) i.e. menopause, induced menopause, or disease, respectively, has occurred but none of the other events, and states (MD) or (ID) if both, menopause and disease have occurred. It seems reasonable to assume that the indicator processes $Y_I(t), Y_M(t), Y_D(t)$ form a CFMP (one could of course doubt the Markov assumption). It is clear that $Y_M \perp \!\!\! \perp Y_I$ and $Y_I \perp \!\!\! \perp Y_M$ since if one of the events M or I has occurred then the intensity for the other one equals zero. Interestingly the analysis of the author shows that $Y_M \perp \!\!\! \perp Y_D$ whereas $Y_D \perp \!\!\! \perp Y_M$ and a corresponding result for induced menopause. More generally, consider events A_1, \dots, A_K then $Y_k(t) = \mathbf{1}(A_k \text{ has occurred before or at time } t)$ form the components of a CFMP if no two or more events may occur systematically at the same time. The state space is given by all combinations $\{0, 1\}^K$ of the events having occurred or not. The local independence structure indicates those events the occurrences of which can be discarded when assessing the intensity for a specific event, i.e. the dependence structure between past and present events.

A special case of event history analysis is survival analysis. Here, the occurrence of a specific event marks the transition into an absorbing state. As discussed for instance by Andersen (1986), survival analysis with time dependent covariates may appropriately be modelled using multistate Markov processes. Assume that the process $Y(t) = \mathbf{1}(\text{survival at least until time } t)$ describes the survival status and the covariate processes X_1, \dots, X_K the occurrence of intermediate events, as for instance onset of a side effect

of a medicamentation. If it can be ruled out that any of the corresponding counting processes jump at the same time we again have that (Y, X_1, \dots, X_K) form a CFMP. Trivially we have that $X_k \not\perp^l Y$ holds for $k = 1, \dots, K$, since after death no further transitions are possible, i.e. every intensity for a change from state $(y = 1, x_1, \dots, x_K)$ into any other is zero. Schweder (1970) proposes in this case to condition on being alive in order to assess the interaction between the covariates. A statistical test of the hypothesis that a specific covariate X_k has no influence on the survival is in this setting equivalent to assessing that $Y \perp^l X_k$. The presented approach additionally allows for the modelling of intermediate events. It is for example possible that $Y \perp^l X_k$ but $Y \not\perp^l X_j$ and $X_j \not\perp^l X_k$ which implies that X_k has an indirect effect on survival.

3 Local independence graphs

Local independence graphs are defined to represent the local independence structure of a CFMP. This calls for a new kind of graph which allows for cycles and bidirected edges since reciprocal local dependence is possible.

Definition 3.1 *Directed reciprocal graph / subgraph*

A *directed reciprocal graph* is a pair $G = (V, E)$, where $V = \{1, \dots, K\}$ is a finite set of vertices and E is a set of directed edges, i.e. $E \subseteq E^1(V) = \{(j, k) | j, k \in V, j \neq k\}$. If $A \subset V$ the induced *subgraph* G_A is defined as (A, E_A) with $E_A = E \cap E^1(A)$. //

In the visualization of the graph directed edges (j, k) are represented by arrows, $j \rightarrow k$. In contrast to the convention, the case $(j, k) \in E$ and $(k, j) \in E$ is shown as double headed arrow, $j \longleftrightarrow k$. Two vertices j and k are elements of a cycle if there exists a (directed) path from j to k as well as from k to j . Note that the usual graph terminology (cf. Koster, 1996, who also considers non-recursive graphs) can still be applied to directed reciprocal graphs.

The following definition provides the link between the above defined graphs and local independence structures.

Definition 3.2 *Local independence graph*

Let $\mathbf{Y} \sim (Y_1, \dots, Y_K)$ be a CFMP and $G = (V, E)$ a directed reciprocal graph. The distribution of \mathbf{Y} is *graphical with respect to G* if

$$Y_j \perp^l Y_k \quad \forall (k, j) \notin E, k \neq j. \quad (3)$$

The graph G is then called *local independence graph of \mathbf{Y}* . //

Property (3) is the analogue of the pairwise Markov property of a conditional independence graph. It is therefore called *pairwise dynamic Markov property (DP)*. As stated in Assumption 2.5, $Y_i \not\perp^l Y_i \forall i \in V$, but this is not shown by a special symbol in the graph.

Example: Consider a CFMP $\mathbf{Y} \sim (Y_1, Y_2, Y_3)$. The graph in Figure 1 is the local independence graph of \mathbf{Y} if $Y_1 \perp^l Y_3$, $Y_2 \perp^l Y_3$, $Y_3 \perp^l Y_1$. The only cycle component is given by $\{1, 2\}$. //

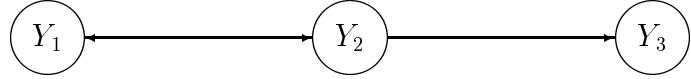


Figure 1

The visualization of the local independence structure is all the more of interest as additional properties can be read off the graph. These could be further Markov properties as addressed in the following.

Theorem 3.3 *Local dynamic Markov property*

Let $\mathbf{Y} \sim (Y_1, \dots, Y_K)$ be a CFMP and $G = (V, E)$ a directed reciprocal graph. Then, property (DP) is equivalent to

$$Y_j \perp^l Y_{V \setminus (\text{pa}(j) \cup \{j\})},$$

which is called the *local dynamic Markov property (DL)*. //

Proof:

First, we show that (DL) implies (DP). From Definition 2.3 it follows with (DL) that $\alpha_t^j(\mathbf{y}; y'_j)$ is a constant function of any y_k with $k \in V \setminus \{\text{pa}(j) \cup \{j\}\}$ so that $\alpha_t^j(\mathbf{y}; y'_j) = \alpha_t^j(\mathbf{y}_{\text{pa}(j) \cup \{j\}}; y'_j) \forall \mathbf{y}_{\text{pa}(j) \cup \{j\}} \in \mathcal{S}_{\text{pa}(j) \cup \{j\}}$ and $y'_j \in \mathcal{S}_j$ with $y'_j \neq y_j$, $j \in V$. Since $V \setminus \{\text{pa}(j) \cup \{j\}\} = \{k \in V \mid (k, j) \notin E, k \neq j\}$, it follows immediately that (DP) holds. Now we show that (DP) implies (DL). By (DP) we have for any $j \in V$: $\alpha_t^j(\mathbf{y}; y'_j) = \alpha_t^j(\mathbf{y}_{-k}; y'_j) \forall k \in V$ with $(k, j) \notin E$. Assume that $\text{pa}(j) = \{k, l\}$ then $\alpha_t^j(\mathbf{y}; y'_j) = \alpha_t^j(\mathbf{y}_{-k}; y'_j) = \alpha_t^j(\mathbf{y}_{-l}; y'_j)$. The case that $\alpha_t^j(\mathbf{y}_{-l, k}; y'_j) \neq \alpha_t^j(\mathbf{y}; y'_j)$ can only occur if Y_l and Y_k carry the same information, but this contradicts the assumption of composability. The same argument can be applied to the general situation of $\text{pa}(j) \subset V$. \square

The above proof of (DP) implying (DL) relies on the composability of the process and on the fact that \mathbf{Y} is a Markov process. Therefore, Theorem 3.3 does not necessarily hold for more general processes. However, for the transition intensities of Markov processes we may now write $\alpha_t^j(\mathbf{y}; y_j) = \alpha_t^j(\mathbf{y}_{\text{cl}(j)}; y_j)$, where $\text{cl}(j) = \text{pa}(j) \cup \{j\}$ is the closure of j .

The analogue of the global Markov property in conditional independence graphs is concerned with the role of separating sets and subprocesses of the original process. Since we restrict our considerations to Markov processes we use a result of Schweder (1970) stating which subprocesses of a CFMP \mathbf{Y} are still Markov processes in order to show the following lemma.

Lemma 3.4 *Local independence graph for subprocess*

Let \mathbf{Y}_A , $A \subset V$, be a subprocess of a CFMP $\mathbf{Y} \sim (Y_1, \dots, Y_K)$ with local independence graph G and assume $\text{pa}(A) = \emptyset$, i.e. A is an anterior set. The local independence graph of \mathbf{Y}_A is given as the subgraph $G_A = (A, E_A)$ of G . //

Proof:

It follows from Schweder (1970, Theorem 2) that for $\text{pa}(A) = \emptyset$ \mathbf{Y}_A is a Markov process with transition intensities

$$\alpha_t^A(\mathbf{y}_A; y'_j) = \begin{cases} \alpha_t^j(\mathbf{y}_A; y'_j), & y_j \neq y'_j \wedge \mathbf{y}_{A \setminus \{j\}} = \mathbf{y}'_{A \setminus \{j\}}, j \in A \\ 0, & \text{else} \end{cases}, \quad t \in \mathcal{T}.$$

This immediately yields the proof. \square

From Lemma 3.4 it follows that within the class of CFMP any set $A \subset V$ with $\text{pa}(A) = \emptyset$ is collapsible, i.e. the marginal distribution has the same local dependence structure as the corresponding subgraph. In contrast, an arbitrary subprocess \mathbf{Y}_A with $\text{pa}(A) \neq \emptyset$ is not necessarily a Markov process, i.e. the class of CFMPs is not closed under marginalization (see Remark 2.4) and the dependence structure may differ from the corresponding subgraph. The following definition of separating sets takes that into account.

Definition 3.5 δ^* -separation

Let $G = (V, E)$ be a directed reciprocal graph. For three disjoint subsets $A, B, C \subset V$ we say that C δ^* -separates A from B if

- (a) each path from B to A has elements in C ,
- (b) and $\text{pa}(A \cup B \cup C) = \emptyset$.

//

Thus, δ^* -separation guarantees that the subprocess given by ignoring the components in $V \setminus (A \cup B \cup C)$, i.e. $\mathbf{Y}_{A \cup B \cup C}$, is a Markov process. Note that the definition is not symmetric in A and B . For general CFMPs we get the following result.

Theorem 3.6 δ^* -global dynamic Markov property

Let $\mathbf{Y} \sim (Y_1, \dots, Y_K)$ be a CFMP and $G = (V, E)$ a directed reciprocal graph. Then property (DP) is equivalent to the δ^* -global dynamic Markov property (DG^*) which is defined as follows: \forall disjoint sets $A, B, C \subset V$ with C δ^* -separating A from B it holds that

$$Y_A \perp^l Y_B \text{ with respect to the local independence graph of } \mathbf{Y}_{A \cup B \cup C}. \quad (4)$$

//

Proof:

The implication $(DP) \Rightarrow (DG^*)$ is an immediate consequence of Lemma (3.4).

The fact that (DG^*) implies (DP) can be seen by showing that $(DG^*) \Rightarrow (DL)$. For any $j \in V$ we have that $\text{pa}(j)$ δ^* -separates $\{j\}$ from $V \setminus \text{cl}(j)$ with $D = \emptyset$ and $(DL) \Rightarrow (DP)$ follows from Theorem 3.3. \square

Theorems 3.3 and 3.6 thus yield the equivalence of properties (DP), (DL) and (DG^*) . However, the δ^* -separation is not as exhaustive as one would wish because of condition (b) which is mainly due to the fact that the class of CFMPs is not closed under

marginalization. Results for more general separations can be obtained by generalizing the concept of local independence to non-Markov processes. For the continuous time situation this requires some deeper results on the behaviour of subprocesses of CFMPs which are not considered here. The discrete time situation is addressed in the following section.

4 Discrete time models

Until now we have restricted our considerations to continuous time Markov processes. There are different possible approaches to the general discrete time situation in event history analysis. If several processes with finite state spaces are considered, there is at first the possibility to model the dependence structure via classical graphical chain models for discrete variables where each discrete time point represents a chain component. In case that there are many time points, many processes, and/or a high dimensional state space, this would result in high dimensional loglinear models with the eventual problem of non identifiability or the need of a very large sample size. Thus, some restrictive assumptions such as the one of p -th order Markov processes or a repetition of the same dependence structure between any p chain components are called for. Classical chain graphs have for instance been applied by Klein et al. (1995) or in a more general framework not restricted to discrete variables by Lynggaard & Walther (1993). Statistical software packages such as DIGRAM (described for instance in Klein et al., 1995) provide the necessary computational support.

In this section we discuss two extensions to the discrete time situation that are more closely related to the ideas of the foregoing sections. On the one hand, the assumptions of a composable Markov process may be maintained where the condition of no more than one jump at a time translates to the discrete time case as stochastically independent innovations. This seems to be a sensible assumption only if the space between two sequential points in time is not too large in relation to the considered events. We deal with this case in the first subsection. On the other hand, one may assume that a continuous time Markov process with a local independence structure modelled as above underlies the discrete process. Then, the innovations may no longer be independent. In the second subsection we present a proposition that deals with this situation.

4.1 Composable finite Markov chains

Let $Y = \{Y(t) | t \in \mathcal{T}\}$ be a Markov chain with state space \mathcal{S} and discrete time scale, i.e. $\mathcal{T} = \{t_0, t_1, t_2, \dots\}$, where $0 = t_0 < t_1 < t_2 < \dots$. The single-step transition probabilities are given by

$$\alpha_i(y; y') = P(Y(t_i) = y' | Y(t_{i-1}) = y), \quad y, y' \in \mathcal{S}, i = 1, 2, \dots$$

Like in the preceding section we assume that the process consists of different components.

Definition 4.1 *Composable finite Markov chain, CFMC*

Let $\mathbf{Y} = \{\mathbf{Y}(t) | t \in \mathcal{T}\}$ be a Markov chain with $\mathbf{Y} \sim (Y_1, \dots, Y_K)$ as in Definition 2.1, where condition (2) is modified to

$$\alpha_i(\mathbf{y}; \mathbf{y}') = \prod_{j=1}^K \alpha_i^j(\mathbf{y}; y'_j), \quad \forall \mathbf{y}, \mathbf{y}' \in \mathcal{S}, i = 1, 2, \dots, \quad (5)$$

where $\alpha_i^j(\mathbf{y}; y'_j) = P(Y_j(t_i) = y'_j | \mathbf{Y}(t_{i-1}) = \mathbf{y})$. The process \mathbf{Y} is then called *composable finite state Markov chain, CFMC*. //

Remark 4.2 *Independent innovations*

The above definition implies by (5) that for any $i = 1, 2, \dots$ all components $Y_j(t_i)$, $j = 1, \dots, K$, are stochastically independent of one another given $\mathbf{Y}(t_{i-1})$. To put it differently, any dependences in the marginal distribution of $Y_1(t_i), \dots, Y_K(t_i)$ should vanish by conditioning on the preceding time point. This is in analogy to (2) because of the underlying idea that the different processes should not be fed by the same innovations. If there were dependences within the vector $\mathbf{Y}(t_i)$ given $\mathbf{Y}(t_{i-1})$ this would mean that the latter is not the 'only cause' of the former, but that there is a 'common development' during the time period t_{i-1} to t_i . Consider the example of different measures concerning the physical status of a person. If these are measured daily one could under certain circumstances say that the development of the different measures is independent from one day to another given the preceding day. But if they are observed weekly it should reasonably be assumed that there is a common development during a week. Obviously, assumption (5) will not be appropriate in many practical situations, since it is often sensible to assume that there are common causes during t_{i-1} and t_i . //

In complete analogy to the preceding section we define for a CFMC $\mathbf{Y} \sim (Y_1, \dots, Y_K)$ that

$$Y_j \perp^l Y_k \Leftrightarrow \alpha_i^j(\mathbf{y}; y'_j) \text{ constant function of } y_k \quad \forall \mathbf{y}_{-k} \in \mathcal{S}_{-k}, y'_j \in \mathcal{S}_j.$$

If $G = (V, E)$ is the corresponding local independence graph of \mathbf{Y} as in Definition 3.2 it is straightforward from (5) that

$$\alpha_i(\mathbf{y}; \mathbf{y}') = \prod_{j=1}^K \alpha_i^j(\mathbf{y}_{\text{cl}(j)}; y'_j), \quad \forall \mathbf{y}, \mathbf{y}' \in \mathcal{S}, i = 1, 2, \dots, \quad (6)$$

where $\text{cl}(j) = \text{pa}(j) \cup \{j\}$ is the closure of j . It follows from (6) that the independence structure of \mathbf{Y} can equivalently be described by a classical DAG \tilde{G} with directed edges from $Y_j(t_{i-1})$ to $Y_k(t_i)$, $i = 1, 2, \dots$, if $(j, k) \in E$ or $j = k$, $j, k \in V$ (for the definition of a DAG cf. Lauritzen, 1996). All independence properties of the distribution can be read off this DAG. An exception is the starting distribution, i.e. the distribution of $\mathbf{Y}(t_0)$ which is not necessarily the independence distribution, so that one possibly has to condition on $\mathbf{Y}(t_0)$.

It is obvious that a result analogous to Lemma 3.4 and to Theorems 3.3, 3.6 holds for CFMCs, too. In contrast to Theorem 3.6 we now consider the behaviour of subprocesses \mathbf{Y}_A , where $\text{pa}(A) \neq \emptyset$, when marginalizing over $\mathbf{Y}_{V \setminus A}$. As noted above, such a subprocess is not necessarily a Markov process but we will see that the notion of local independence can still be applied. Instead of the single step transition probabilities, one has to consider the transition probabilities conditional on the whole past of the process. The following proposition shows that the subprocess \mathbf{Y}_A depends on this past via $\text{pa}(A)$ because this is a subset of the set $V \setminus A$ with respect to which we marginalize and via $\text{pa}(V \setminus A)$, i.e. the parents of the set with respect to which we marginalize. Only for subsets $A^0 \subset A$ with $\text{pa}(A^0) \cap V \setminus A = \emptyset$, \mathbf{Y}_{A^0} is still a Markov process.

Proposition 4.3 *Subprocess of a CFMC*

Let $\mathbf{Y} \sim (Y_1, \dots, Y_K)$ be a CFMC with local independence graph G and consider a subprocess \mathbf{Y}_A , $A \subset V$. Let further $A = A^0 \cup A^1$ for disjoint sets A^0, A^1 , where $\text{pa}(A^0) \cap \bar{A} = \emptyset$, and $A' = \text{pa}(\bar{A}) \cap A$ with $\bar{A} = V \setminus A$. Assume for the starting distribution that it factorizes according to $P(\mathbf{Y}(t_0)) = P(\mathbf{Y}_{A^0}(t_0))P(\mathbf{Y}_{A^1}(t_0))P(\mathbf{Y}_{\bar{A}}(t_0))$. Then it holds for the transition probabilities of the subprocess \mathbf{Y}_A that

$$\begin{aligned} P(\mathbf{Y}_A(t_i) = \mathbf{y}_A^i | \mathbf{Y}_A(t_{i-1}) = \mathbf{y}_A^{i-1}, \dots, \mathbf{Y}_A(t_0) = \mathbf{y}_A^0) \\ = P(\mathbf{Y}_{A^0}(t_i) = \mathbf{y}_{A^0}^i | \mathbf{Y}_{\text{cl}(A^0)}(t_{i-1}) = \mathbf{y}_{\text{cl}(A^0)}^{i-1}) \\ \cdot P(\mathbf{Y}_{A^1}(t_i) = \mathbf{y}_{A^1}^i | \mathbf{Y}_A(t_{i-1}) = \mathbf{y}_A^{i-1}, \dots, \mathbf{Y}_A(t_0) = \mathbf{y}_A^0) \quad (7) \end{aligned}$$

with

$$\begin{aligned} P(\mathbf{Y}_{A^1}(t_i) = \mathbf{y}_{A^1}^i | \mathbf{Y}_A(t_{i-1}) = \mathbf{y}_A^{i-1}, \dots, \mathbf{Y}_A(t_0) = \mathbf{y}_A^0) = \\ \frac{\sum_{\bar{y}_{\bar{A}}} \left[\prod_{j=1}^i P(\mathbf{Y}_{A^1}(t_j) = \mathbf{y}_{A^1}^j | \mathbf{Y}_{\text{cl}(A^1)}(t_{j-1}) = \mathbf{y}_{\text{cl}(A^1)}^{j-1}) \omega_j(\mathbf{y}_{\bar{A}}^{j-1}, \mathbf{y}_{\text{cl}(\bar{A})}^{j-2}) \right]}{\sum_{\bar{y}_{\bar{A}}} \left[\prod_{j=1}^{i-1} P(\mathbf{Y}_{A^1}(t_j) = \mathbf{y}_{A^1}^j | \mathbf{Y}_{\text{cl}(A^1)}(t_{j-1}) = \mathbf{y}_{\text{cl}(A^1)}^{j-1}) \omega_j(\mathbf{y}_{\bar{A}}^{j-1}, \mathbf{y}_{\text{cl}(\bar{A})}^{j-2}) \right]} \quad (8) \\ \text{with } \omega_j(\mathbf{y}_{\bar{A}}^{j-1}, \mathbf{y}_{\text{cl}(\bar{A})}^{j-2}) = \begin{cases} P(\mathbf{Y}_{\bar{A}}(t_0) = \mathbf{y}_{\bar{A}}^0), & j = 1 \\ P(\mathbf{Y}_{\bar{A}}(t_{j-1}) = \mathbf{y}_{\bar{A}}^{j-1} | \mathbf{Y}_{\text{cl}(\bar{A})}(t_{j-2}) = \mathbf{y}_{\text{cl}(\bar{A})}^{j-2}), & j > 1 \end{cases} \end{aligned}$$

where $\sum_{\bar{y}_{\bar{A}}}$ means summarizing over all $(\mathbf{y}_{\bar{A}}^0, \dots, \mathbf{y}_{\bar{A}}^i) \in \bigotimes_{j=0}^i \mathcal{S}_{\bar{A}}$. //

Since the proof is technical and of no substantial interest it is referred to the appendix.

Plausibly, the subprocess \mathbf{Y}_{A^0} still has the Markov property since it is not touched by $\mathbf{Y}_{\bar{A}}$ whereas the remaining \mathbf{Y}_{A^1} depends through $\text{pa}(A^1) \cap \bar{A}$ on the whole past of the subprocess. The second factor of (7) may even depend on the past of \mathbf{Y}_{A^0} if $\text{pa}(\bar{A}) \cap A^0 \neq \emptyset$ as can be seen from the weights in (8).

Example: In order to illustrate the Proposition 4.3 consider again the first example of a local independence graph with vertices $\{(1, 2), (2, 1), (2, 3)\}$ (Figure 1). The subprocess Y_{12} is still a Markov process since $3 \notin \text{pa}(1, 2)$. In contrast, this is not the case if the vertices are given by $\{(1, 2), (2, 3), (3, 1)\}$ (Figure 2 (a)). Here we have for $A = \{1, 2\}$ that $A^0 = \{2\}$, $A^1 = \{1\}$, and $A' = \{2\}$. It follows from Proposition 4.3

that the local independence $Y_1 \perp^l Y_2$ which holds with respect to \mathbf{Y}_{123} is not preserved when marginalizing over Y_3 because of the vertices (2, 3) and (3, 1).

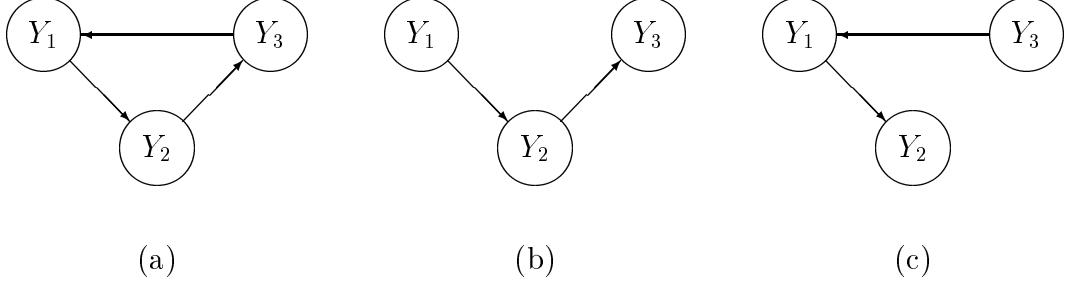


Figure 2

If the edge (3, 1) was absent as in Figure 2 (b) \mathbf{Y}_{12} would be a Markov process and $Y_1 \perp^l Y_2$ with respect to the reduced subprocess \mathbf{Y}_{12} . If, in contrast, the edge (2, 3) was absent (Figure 2 (c)), \mathbf{Y}_{12} would be no Markov process but $P(Y_1(t_i) = y_1(t_i) | \text{history of } \mathbf{Y}_{12}) = P(Y_1(t_i) = y_1(t_i) | \text{history of } Y_1)$, i.e. the transition probability is independent of $Y_2(t_{i-1}), \dots, Y_2(t_0)$. //

The latter consideration gives rise to the following more general definition of local independence.

Remark 4.4 *General local independence for CFMC*

In the situation of Proposition 4.3 we speak of local independence within \mathbf{Y}_A in the following sense: Any Y_j , $j \in A^1$, is locally independent of Y_k , $k \in A$, if the second factor in (7), i.e. (8), is independent of $Y_k(t_i), \dots, Y_k(t_0)$. //

Lemma 4.5 *General local independence within a subprocess of a CFMC*

With the assumptions as in Proposition 4.3 and with Remark 4.4 we have

(a) $\forall j \in A^1$:

$$Y_j \perp^l Y_k \quad \forall k \in A \setminus (\text{cl}(A^1) \cup A'); \quad (9)$$

(b) $\forall j \in A^0$: $Y_j \perp^l Y_k \quad \forall k \in A \setminus \text{cl}(j)$. //

Proof:

To see (a) consider (8): if $k \notin \text{cl}(A^1)$ then $P(\mathbf{Y}_{A^1}(t_j) | \mathbf{Y}_{\text{cl}(A^1)}(t_{j-1}))$ is independent of k and if additionally $k \notin A'$ then $\omega_j(\mathbf{y}_A^{j-1}, \mathbf{y}_{\text{cl}(A)}^{j-2})$ is also independent of k . Thus the second factor of (7) is independent of k .

Part (b) is obvious from the first factor in (7) which gives the probability for a change in \mathbf{Y}_{A^0} . \square

The above lemma is not exhaustive since Y_j in (9) may in special cases be locally independent on even more components. Consequently, the local independence graph of a general subprocess cannot be given explicitly as in Lemma 3.4 since it depends on the structure of (8). However, we can see that a sufficient condition for $\mathbf{Y}_{A^1} \perp \!\!\! \perp \mathbf{Y}_B$, $B \subset A^0$, is that $\text{pa}(A^1) \cap B = \emptyset$ and $B \cap A' = \emptyset$ hold.

The foregoing considerations lead to a more general definition of separation in local independence graphs and to restate the global Markov property as well as Theorem 3.6 for CFMCs.

Definition 4.6 δ -separation

Let $G = (V, E)$ be a directed reciprocal graph. For three disjoint subsets $A, B, C \subset V$ and $D = V \setminus (A \cup B \cup C)$ we say that C δ -separates A from B if

- (a) each path from B to A has elements in C ,
- (b) and either $\text{pa}(A) \cap D = \emptyset$ or $\text{pa}(B) \cap D = \emptyset$.

//

The δ -global dynamic Markov property (DG) is defined in analogy to (4) by replacing the δ^* -separation by the δ -separation. The necessity of condition (b) in the above definition can be understood by the following argument: If the subset A as well as the subset B have parents in D then the marginalization with respect to D could induce dependences between the former which are not captured by C . This phenomenon is very similar to classical graphical models.

With Definition 4.6 we now have the following result.

Theorem 4.7 Equivalence of Markov properties for CFMC

Let $\mathbf{Y} \sim (Y_1, \dots, Y_K)$ be a CFMC and G a directed reciprocal graph. Then it holds that $(\text{DP}) \Leftrightarrow (\text{DL}) \Leftrightarrow (\text{DG})$. //

Proof:

The first equivalence can be shown as in Theorem 3.3 and $(\text{DG}) \Rightarrow (\text{DP})$ as in Theorem 3.6.

To see $(\text{DP}) \Rightarrow (\text{DG})$ consider first the situation that $\text{pa}(A) \cap D = \emptyset$. Then A takes the role of A^0 in Proposition 4.3. Since $B \cap \text{cl}(A) = \emptyset$ because of C separating A from B Lemma 4.5 (b) yields the proof.

If $\text{pa}(A) \cap D \neq \emptyset$ we have $\text{pa}(B) \cap D = \emptyset$ and $\text{pa}(D) \cap B = \emptyset$ because of the definition of δ -separation. Thus, the set A takes the role of A^1 in Proposition 4.3 and since $B \cap \text{cl}(A) = \emptyset$ as well as $\text{pa}(D) \cap B = \emptyset$ Lemma 4.5 (a) yields the proof. \square

We have seen for the case of CFMCs that Definition 2.3 can be generalized. Aalen (1987) proposes an even more general definition with respect to counting processes. Within that framework more general results concerning the global Markov property as in Theorem 3.6 can be obtained. This, however, requires some deeper results on the behaviour of subprocesses of CFMPs, which are not considered here.

4.2 Discretizing a CFMP

In this section we relax assumption (5) which is crucial to the results of the foregoing section. Instead, we consider discrete time processes that result from CFMPs which are only observed in discrete time, i.e. we have to cope with the fact that the CFMP ‘continues’ between the points in time where it can be observed. The following proposition indicates the conditional independences within such a discretized process. The aim is to ‘translate’ the local independence graph into an appropriate chain graph similar to the independence structure of a CFMC being represented by a DAG (cf. equation (6)).

Proposition 4.8 *Generated chain model*

Let $\mathbf{Y} \sim (Y_1, \dots, Y_K)$ be a CFMP with local independence graph G , and let $t_i \in \mathcal{T}$ with $t_0 = 0$, $t_i < t_{i+1}$, $i = 0, 1, 2, \dots$, be discrete points in time. We will describe the distribution of the discretized process $\mathbf{Y}' = \{\mathbf{Y}'(i) | i = 1, 2, \dots\}$ with $\mathbf{Y}'(i) = \mathbf{Y}(t_i)$. It holds for all $i = 1, 2, \dots$ that

- (a) $P(\mathbf{Y}'(i) = \mathbf{y}_i | \mathbf{Y}'(i-k) = \mathbf{y}_{i-k}, \dots, \mathbf{Y}'(1) = \mathbf{y}_1) = P(\mathbf{Y}'(i) = \mathbf{y}_i | \mathbf{Y}'(i-k) = \mathbf{y}_{i-k})$, for $k = 1, \dots, i-1$;
- (b) for all $j \in V$:

$$Y'_j(i) \perp \mathbf{Y}'_{V \setminus \text{an}_G(j)}(i-k) | \mathbf{Y}'_{\text{an}_G(j)}(i-k) \quad \forall k = 1, \dots, i-1; \quad (10)$$

- (c) the distribution of $\mathbf{Y}'(i) | \mathbf{Y}'(i-k)$ is G' -Markov (in the classical sense) $\forall k = 1, \dots, i-1$, where $G' = (V, E')$ is the classical chain graph with edges $E' = \{(j, k) \in V \times V | \exists \text{ directed path from } j \text{ to } k \text{ in } G \text{ or } \text{an}_G(j) \cap \text{an}_G(k) \neq \emptyset, j \neq k\}$. //

Proof:

Part (a) follows from the Markov property of the underlying CFMP.

Part (b) follows from $\mathbf{Y} \sim (\mathbf{Y}_{V \setminus \text{an}_G(j)}, \mathbf{Y}_{\text{an}_G(j)}) \forall j \in V$, where $\mathbf{Y}_{\text{an}_G(j)} \perp \mathbf{Y}_{V \setminus \text{an}_G(j)}^l$ so that $\mathbf{Y}_j \perp \mathbf{Y}_{V \setminus \text{an}_G(j)} \forall j \in V$, and by application of Theorem 1 of Schweder (1970).

To see (c) we have to show that $\mathbf{Y}'_A(i) \perp \mathbf{Y}'_B(i) | \mathbf{Y}'_C(i) \mathbf{Y}'(i-k)$ whenever C separates A and B in the moral graph of the smallest anterior subgraph of G' containing A , B , and C . By the construction of the graph G' we have that it holds for the separating set with respect to G that (i) no element in C has ancestors in A as well as in B ; (ii) neither A has ancestors in B nor vice versa; (iii) A and B have no common ancestors. Thus, C can be partitioned into sets C_A and C_B with $\text{an}_G(C_A) \cap B = \emptyset$ and $\text{an}_G(B) \cap C_A = \emptyset$, $\text{an}_G(C_B) \cap A = \emptyset$ and $\text{an}_G(A) \cap C_B = \emptyset$, and $\mathbf{Y}_{C_A} \perp \mathbf{Y}_{C_B}$. It follows that $\mathbf{Y}_{A \cup C_A}$ and $\mathbf{Y}_{B \cup C_B}$ are stochastically independent processes. This yields the proposition. \square

The second part of the above theorem states that $Y_j(t+h)$, $h > 0$, is conditionally independent of those $Y_k(t)$ where there is no directed path from k to j in the local independence graph G given the components where there is such a path. In (10) $\text{an}_G(j)$ cannot be replaced by $\text{cl}_G(j)$, i.e. by those components of which Y_j is locally dependent, since $\mathbf{Y}_{\text{cl}_G(j)} \not\perp \mathbf{Y}_{V \setminus \text{cl}_G(j)}$ if there exists $l \in \text{cl}_G(j)$ and $k \in V \setminus \text{cl}_G(j)$ with $(k, l) \in E$. It follows that independences are preserved only if they are not conveyed by intermediates when marginalizing over the time between t and $t+h$.

The proof of the third part makes clear that the conditional independences within the distribution of $\mathbf{Y}'(i)|\mathbf{Y}'(i-k)$ are at the same time marginal independences. This can be explained by the fact that marginalization over the time between t_{i-k} and t_i implies marginalizing over all 'common causes' during this time which are given by common ancestors. Independences are only preserved between those processes who have no common ancestors and these are already marginally independent. It can be supposed that in practical situations there will usually be more (conditional) independences since local independence is restrictively defined by assuming that the independence holds for all $t \in \mathcal{T}$. The property of the transition intensities described in Definition 2.3 may hold for a subset $T \subset \mathcal{T}$ yielding more independences than those which can be read off the local independence graph.

Example: To illustrate the above theorem consider a CFMP $\mathbf{Y} \sim (Y_1, Y_2, Y_3, Y_4)$ with local independence graph $G = (V, E)$, $E = \{(1, 2), (3, 2), (3, 4), (4, 3)\}$ (Figure 3).

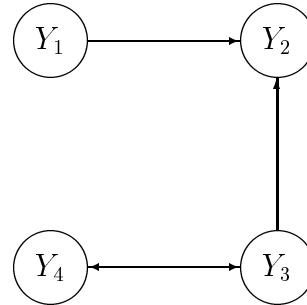


Figure 3

By Lemma 3.4 we know that each of the subprocesses Y_1 , \mathbf{Y}_{34} , and \mathbf{Y}_{134} are Markov processes. By part (b) of the above theorem we further have that for two discrete points in time t and $t + h$, $h > 0$: $Y_1(t+h) \perp\!\!\!\perp Y_{234}(t)|Y_1(t)$, $Y_3(t+h) \perp\!\!\!\perp Y_{12}(t)|Y_{34}(t)$, and $Y_4(t+h) \perp\!\!\!\perp Y_{12}(t)|Y_{34}(t)$.

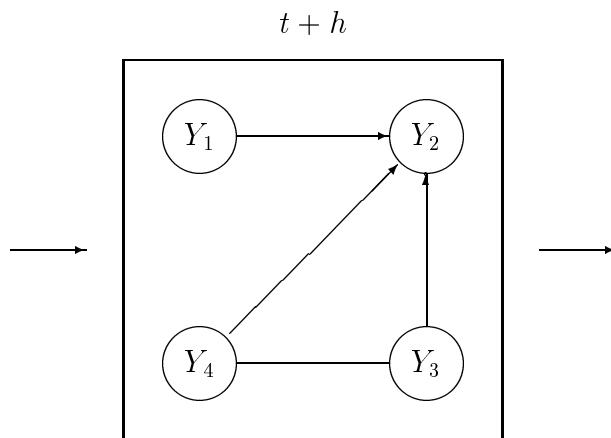


Figure 4

From part (c) it follows that the conditional distribution of $\mathbf{Y}(t+h)|\mathbf{Y}(t)$ holds the classical Markov properties of the graph given in Figure 4, above. This illustrates for example that $Y_1(t+h) \perp Y_3(t+h)|Y_4(t+h)\mathbf{Y}(t)$ as well as $Y_1(t+h) \perp Y_3(t+h)|\mathbf{Y}(t)$. Combining Theorems 2 and 3 from Schweder (1970) we additionally get that Y_1 and Y_{34} are marginally independent Markov processes so that $Y_1 \perp Y_3$ and $Y_1 \perp Y_4$. //

5 Discussion

A further important question related to the proposed local independence graphs concerns estimation and test procedures. Since we restrict our considerations to Markov processes, standard results on estimating and testing within this class can be applied (cf. Andersen et al., 1993). The restrictions given by the graphical structure have to be taken into account by modelling the transition intensities $\alpha_t^j(\mathbf{y}_{\text{cl}(j)}; y^j)$ as functions only of $\mathbf{y}_{\text{cl}(j)}$. Two general approaches are available: the nonparametric and the (semi-)parametric. The former mainly relies on the Nelson–Aalen estimator for the integrated transition intensities. Within the nonparametric framework k –sample tests of equality of transition intensities are available if no a priori assumptions on the local dependence structure are possible. In the (semi-)parametric approach $\alpha_t^j(\mathbf{y}_{\text{cl}(j)}; y^j)$ is modelled as a specific function of $\mathbf{y}_{\text{cl}(j)}$ but we would like to restrict ourselves from going into details here.

The modelling and analyzing of local independence structures can be useful in several fields of application. In survival analysis, for instance, typically not only the survival status but also different time varying covariates describing health status or onset of side effects are observed (cf. Klein et al. 1995). It should be pointed out that time–constant as well as time–varying covariates can be included in local independence graphs. Firstly, the former are equivalent to processes that do not change in time, i.e. they are a priori locally independent of all varying processes. Secondly, if a time constant covariate has a significant influence on one of the processes, say the one describing the survival status, then the latter is obviously locally dependent on this covariate. To put it differently: Time constant covariates can be represented in the local independence graphs by special symbols, where directed edges from processes to constant covariates are forbidden. Additionally, one could analyze the dependence structure among the time–constant covariates using classical graphical modelling techniques.

Time varying covariates can be included in the analysis by assuming appropriate multistate Markov processes. The composability is then given in a very natural way. The distinction of exogenous and endogenous covariate processes translates to the local independence framework as follows. Exogenous covariate processes are locally independent of any of the remaining processes, i.e. directed edges pointing at them are forbidden, endogenous covariate processes may possibly be locally dependent of any of the remaining ones.

Finally we would like to mention the affinity of local dependence to a certain notion of causality as described by Aalen (1987). In a system of processes represented by

$\mathbf{Y} \sim (Y_1, \dots, Y_K)$ any Y_j with $Y_k \not\perp^l Y_j$, $j \neq k$, can be regarded as a cause of Y_k in the sense that it cannot be replaced by any other subprocess in the system whereas Y_j with $Y_k \perp^l Y_j$, $j \neq k$, is no cause of Y_k since the development of Y_k does not depend on Y_j given $\mathbf{Y}_{\text{cl}(k)}$. Thus, $\mathbf{Y}_{\text{cl}(k)}$ can be interpreted as minimal causal set for Y_k . Note, that this always refers to the specific system of processes taken into consideration. Discarding a subset of \mathbf{Y} may change the local dependence structure and thus the causal structure as shown in the context of δ -separation as well as adding new information, i.e. when considering $\mathbf{Y}' \sim (Y_1, \dots, Y_K, Y_{K+1}, \dots, Y_{K+L})$ instead of \mathbf{Y} . The concept of local independence seems to be a sensible starting point for establishing, analyzing, and visualizing complex causal relations and could therefore be a rewarding topic of future research.

Appendix

Proof of Proposition 4.3:

For the sake of simplicity we write $P(\mathbf{Y}(t_j) = \mathbf{y}^j | \mathbf{Y}(t_{j-1}) = \mathbf{y}^{j-1}) = P(\mathbf{Y}(t_j) | \mathbf{Y}(t_{j-1}))$. Since

$$P(\mathbf{Y}_A(t_i) | \mathbf{Y}_A(t_{i-1}), \dots, \mathbf{Y}_A(t_0)) = \frac{P(\mathbf{Y}_A(t_i), \mathbf{Y}_A(t_{i-1}), \dots, \mathbf{Y}_A(t_0))}{P(\mathbf{Y}_A(t_{i-1}), \dots, \mathbf{Y}_A(t_0))} \quad (11)$$

consider first

$$\begin{aligned} P(\mathbf{Y}_A(t_i), \mathbf{Y}_A(t_{i-1}), \dots, \mathbf{Y}_A(t_0)) &= \sum_{\vec{y}_{\bar{A}}} P(\mathbf{Y}(t_i), \mathbf{Y}(t_{i-1}), \dots, \mathbf{Y}(t_0)) \\ &= \sum_{\vec{y}_{\bar{A}}} P(\mathbf{Y}(t_0)) \prod_{j=1}^i P(\mathbf{Y}(t_j) | \mathbf{Y}(t_{j-1})). \end{aligned} \quad (12)$$

Since

$P(\mathbf{Y}(t_j) | \mathbf{Y}(t_{j-1})) = P(\mathbf{Y}_{A^0}(t_j) | \mathbf{Y}_{\text{cl}(A^0)}(t_{j-1})) P(\mathbf{Y}_{A^1}(t_j) | \mathbf{Y}_{\text{cl}(A^1)}(t_{j-1})) P(\mathbf{Y}_{\bar{A}}(t_j) | \mathbf{Y}_{\text{cl}(\bar{A})}(t_{j-1}))$, where $\text{cl}(A^0) \cap \bar{A} = \emptyset$, we get that (12) equals

$$\begin{aligned} &[P(\mathbf{Y}_{A^0}(t_0)) \prod_{j=1}^i P(\mathbf{Y}_{A^0}(t_j) | \mathbf{Y}_{\text{cl}(A^0)}(t_{j-1}))] P(\mathbf{Y}_{A^1}(t_0)) \\ &\sum_{\vec{y}_{\bar{A}}} P(\mathbf{Y}_{\bar{A}}(t_0)) \prod_{j=1}^i P(\mathbf{Y}_{A^1}(t_j) | \mathbf{Y}_{\text{cl}(A^1)}(t_{j-1})) P(\mathbf{Y}_{\bar{A}}(t_j) | \mathbf{Y}_{\text{cl}(\bar{A})}(t_{j-1})). \end{aligned} \quad (13)$$

Further, it follows from

$$\sum_{\vec{y}_{\bar{A}}^i} P(\mathbf{Y}_{A^1}(t_i) | \mathbf{Y}_{\text{cl}(A^1)}(t_{i-1})) P(\mathbf{Y}_{\bar{A}}(t_i) | \mathbf{Y}_{\text{cl}(\bar{A})}(t_{i-1})) = P(\mathbf{Y}_{A^1}(t_i) | \mathbf{Y}_{\text{cl}(A^1)}(t_{i-1}))$$

that (13) equals $\sum_{\vec{y}_{\bar{A}}} \prod_{j=1}^i P(\mathbf{Y}_{A^1}(t_j) | \mathbf{Y}_{\text{cl}(A^1)}(t_{j-1})) P(\mathbf{Y}_{\bar{A}}(t_{j-1}) | \mathbf{Y}_{\text{cl}(\bar{A})}(t_{j-2}))$ with $P(\mathbf{Y}_{\bar{A}}(t_{j-1}) | \mathbf{Y}_{\text{cl}(\bar{A})}(t_{j-2})) = P(\mathbf{Y}_{\bar{A}}(t_0))$ for $j = 1$.

The same argument applied to the denominator of (11) yields the desired result. \square

References

AALEN, O.O., 1987: Dynamic modeling and causality. *Scand. Actuar. J.*, 177-190.

AALEN, O.O., BORGAN, Ø., KEIDING, N., THORMANN, J., 1980: Interaction between life history events — nonparametric analysis for prospective and retrospective data in the presence of censoring, *Scand. J. Stat.*, 7, 161-171.

ANDERSEN, P.K., 1986: Time-dependent covariates and Markov processes. In: Moolgavkar, S.H. and Prentice, R.L., (eds.), *Modern statistical methods in chronic disease epidemiology*, 82-103. Wiley, New York.

ANDERSEN, P.K., BORGAN, Ø., GILL, R.D., KEIDING, N., 1993: *Statistical models based on counting processes*. Springer, New York.

GRANGER, C.W.J., 1969: Investigating causal relations by econometric models and cross spectral methods. *Econometrica*, 37, 424-438.

KLEIN, P.J., KEIDING, N., KREINER, S., 1995: Graphical models for panel studies, illustrated on data from the Framingham heart study. *Stat. in Med.*, 14, 1265-1290.

KOSTER, J.T.A., 1996: Markov properties of nonrecursive causal models. *Ann. of Statist.*, 25, 2148-2177.

LAURITZEN, S.L., 1996: *Graphical models*. Oxford University Press.

LYNGGAARD, H., WALTHER, K.H., 1993: Dynamic modelling with mixed graphical association models. Master's Thesis, Aalborg University, Denmark.

MYKLAND, P., 1986: Statistical causality. Technical report No. 14, Department of Mathematics, University of Bergen, Norway.

SCHWEDER, T., 1970: Composable Markov processes. *J. Appl. Prob.*, 7, 400-410.

**This work has been supported by the Deutsche Forschungsgemeinschaft,
Sonderforschungsbereich 386.**