

Graphical Models for Composable Finite Markov Processes

VANESSA DIDELEZ

Department of Statistical Science, University College London

ABSTRACT. Composable Markov processes were introduced by Schweder (1970) in order to capture the idea that a process can be composed of different components where some of these only depend on a subset of the other components. Here we propose a graphical representation of this kind of dependence which has been called ‘local dependence’. It is shown that the graph allows to read off further independencies characterizing the underlying Markov process. Also, some standard methods for inference are adapted to exploit the graphical representation, e.g. for testing local independence.

Key words: conditional independence, counting processes, Granger causality, local independence, multistate models

1. Introduction

Composable (finite) Markov processes (CFMPs) were introduced by Schweder (1970). The idea is that often a process can be composed of different components according to its state space, e.g. in an illness–death model there would be a component describing the illness state (healthy/ill) and another one for the death state (alive/dead). We would only talk of different components if they cannot change state at the same time, e.g. a transition from being healthy and alive to being ill and dead at the same time can trivially be excluded. In more complex situations certain independence assumptions among such components might be made and CFMPs allow to model these. In particular, we are interested in so-called local independence, where the transition intensity of one component does not depend on the state of another specific component. Schweder (1970) investigated a number of properties of these processes, addressing, e.g. the question when local independence implies stochastic independence of whole subprocesses.

In the present article, we extend Schweder’s results and put them into the framework of graphical models. Directed (possibly cyclic) graphs are used to represent the local (in)dependence structure. It is shown that these graphs allow to read off further independencies characterizing the Markov process based on graphical separations and collapsibility. Well-known methods of inference for Markov processes can easily be adapted to deal with CFMPs taking the graphical structure into account.

Some previous work on modifying classical graphical models for dynamic data situations – time series or more generally stochastic processes – is closely related to our approach. The graphs proposed in the present paper are a special case of local independence graphs developed by Didelez (2000) for general stochastic processes with Doob–Meyer decomposition, which is due to the one-to-one relation between Markov processes and marked point processes (Didelez, 2005). However, for the case of CFMPs it is more natural to represent a whole component of the process as a node in the graph instead of every individual mark of a marked point process because every transition constitutes a mark and hence graphs could become very complex and contain many trivial dependencies. Gottard (2002) gives an example and application of how Schweder’s local independence and graphical chain models can

be combined, also based on marked point processes. Recently, graphs for Markov processes have been used in artificial intelligence (Nodelman *et al.*, 2002, 2003), where they are called continuous time Bayesian networks, and methods have been developed that enable fast computations on the graphs so as to answer queries about the distribution of the Markov processes. Also, there is a close connection between local independence and Granger-non-causality for time series; essentially, local independence is the continuous time version of the latter (Florens & Fougère, 1996). Hence, our local independence graphs can be regarded as the continuous time version of the ones put forward by Eichler (1999, 2000) for representing dependencies in multivariate time-series based on Granger-causality.

The outline of the paper is as follows. In section 2, we revise the definition of CFMPs. Section 3 gives the definition of local independence for CFMPs and derives some further conditional independence properties for these processes that are implied by local independence. The main topic, local independence graphs, is introduced in section 4. First, we give some graph terminology as required by allowing for cyclic graphs. Secondly, the local independence graphs are defined in terms of graphical Markov properties which we call ‘dynamic’ to distinguish them from the Markov properties of ‘classical’ graphs (e.g. Lauritzen, 1996, ch. 3). Implications of these graphs are discussed in section 4.3 and methods of inference are addressed in section 4.4. Section 5 discusses our proposal and points out possible extensions as well as topics for further research.

2. Composable Markov processes

Throughout we consider a continuous time (finite) first-order Markov process $Y = \{Y(t) | t \in \mathcal{T}\}$, where \mathcal{S} is the finite set of states of Y and $\mathcal{T} = [0, \tau)$ is the time interval of interest. The distribution of such a Markov process Y is characterized by its *transition intensities* $\alpha_{qr}(t)$, $t \in \mathcal{T}$, with

$$\alpha_{qr}(t) = \lim_{h \downarrow 0} \frac{1}{h} P(Y(t+h) = r | Y(t) = q), \quad q \neq r \in \mathcal{S}.$$

These transition intensities are assumed to exist, i.e. $\alpha_{qr}(t) < \infty$ for all $q \neq r$, and to be continuous and bounded functions of t on any closed interval in \mathcal{T} .

We make use of the following relation between Markov processes and counting processes (cf. Andersen *et al.*, 1993, p. 94). A finite state Markov process with state space \mathcal{S} and with the above transition intensities can equivalently be represented by its *associated multivariate counting process* given by $\mathbf{N} = (N_{qr} | q, r \in \mathcal{S}, q \neq r)$ with

$$N_{qr}(t) = \sum_{s \leq t} \mathbf{1}\{Y(s^-) = q \text{ and } Y(s) = r\}, \quad q \neq r,$$

where $Y(s^-) = \lim_{h \downarrow 0} Y(s - h)$. With the internal filtration $\mathcal{F}_t = \sigma\{Y(s) | s \leq t\}$, $t \in \mathcal{T}$, its \mathcal{F}_t -compensator has components

$$\Lambda_{qr}(t) = \int_0^t Z_q(s) \alpha_{qr}(s) ds,$$

where $Z_q(t) = \mathbf{1}\{Y(t^-) = q\}$. The compensator exists if the involved transition intensities $\alpha_{qr}(t)$ exist as assumed here. The corresponding intensity processes are given by

$$\lambda_{qr}(t) = Z_q(t) \alpha_{qr}(t), \quad q, r \in \mathcal{S}, \quad q \neq r. \quad (1)$$

Later, we will also use the notation $\alpha(t; (q, r))$, $\lambda(t; (q, r))$, etc. because the subscripts become more complicated.

In order to model that state changes of a Markov process Y are due to changes in different components, Schweder (1970) suggests the notion of composability. Note that in the following the process will be denoted by boldface letters when it is regarded as consisting of different components and $\bigotimes_{i=1}^N A_i$ denotes the Cartesian product of sets A_1, \dots, A_N .

Definition 1

Assume that for $K \geq 2$ there are spaces $\mathcal{S}_k, k \in V = \{1, \dots, K\}$, $|\mathcal{S}_k| \geq 2$, with a one-to-one mapping f of \mathcal{S} onto $\bigotimes_{k \in V} \mathcal{S}_k$ so that elements $\mathbf{y} \in \mathcal{S}$ can be identified with elements $(y_1, \dots, y_K) \in \bigotimes_{k \in V} \mathcal{S}_k$. Then, a finite Markov process \mathbf{Y} is a composable process with components Y_1, \dots, Y_K given by $f(\mathbf{Y}(t)) = (Y_1(t), \dots, Y_K(t))$ if for all $A \subset V$, $|A| \geq 2$,

$$\lim_{h \downarrow 0} \frac{1}{h} P \left(\bigcap_{k \in A} \{Y_k(t+h) \neq y_k\} \middle| \bigcap_{k \in A} \{Y_k(t) = y_k\} \right) = 0 \quad (2)$$

for all $y_k \in \mathcal{S}_k, k \in V$, and $t \in \mathcal{T}$. We then write $\mathbf{Y} \sim (Y_1, \dots, Y_K)$ and call this a composable finite Markov process (CFMP).

Property (2) implies that the probability for changes in more than one component within a short period of length h is of magnitude $o(h)$. Therefore, any change of state can be represented as a change in only one of the components which justifies saying that the process is *composed* of different components. Note that the composition is not necessarily unique. If for example $\mathbf{Y} \sim (Y_1, \dots, Y_K)$ then $\mathbf{Y} \sim (\mathbf{Y}_A, \mathbf{Y}_B)$ with $A \subset V$ and $B = V \setminus A$, where \mathbf{Y}_A denotes the subprocess on the state space $\mathcal{S}_A = \bigotimes_{k \in A} \mathcal{S}_k$ and \mathbf{Y}_B analogously.

It follows immediately for any $\mathbf{y}, \mathbf{y}' \in \bigotimes_{k \in V} \mathcal{S}_k$ with $\mathbf{y} \neq \mathbf{y}'$ that the transition intensities $\alpha_{\mathbf{y}\mathbf{y}'}(t)$ of a CFMP, denoted as $\alpha(t; (\mathbf{y}, \mathbf{y}'))$, equal zero if \mathbf{y} and \mathbf{y}' differ on more than one component. More formally, they are given by

$$\alpha(t; (\mathbf{y}, \mathbf{y}')) = \begin{cases} \alpha_k(t; (y_k, y'_k)), & y_k \neq y'_k \text{ and } \mathbf{y}_{-k} = \mathbf{y}'_{-k} \\ 0, & \text{else,} \end{cases} \quad (3)$$

where $\mathbf{y}_{-k} = \mathbf{y}_{V \setminus \{k\}}$, and

$$\alpha_k(t; (\mathbf{y}, y'_k)) = \lim_{h \downarrow 0} \frac{1}{h} P(Y_k(t+h) = y'_k | \mathbf{Y}(t) = \mathbf{y}).$$

Example 1. The idea of a composable finite state Markov process has been exploited in an analysis by Aalen *et al.* (1980). The aim was to investigate the dependence between two events: (i) the occurrence of a specific kind of chronic skin disease and (ii) the beginning of menopause. A third event was potentially relevant, (iii) death (in fact, a fourth event was considered, ‘being sampled’, but we will ignore that for the moment). The states can hence be defined as 0 for initial state, M for menopause having started, S for skin disease having occurred, SM for both, and D_0 as well as D_S for death before or after occurrence of the skin disease. The standard graphical representation of such a situation is via a transition graph as given in Fig. 1, where the μ s denotes the transition intensities to death and the α s the transition intensities between the skin disease and menopause states.

We can see that it is not possible that the events skin disease, menopause or death occur at the same time as there is no transition from 0 to SM . This allows us to regard the whole process as being composed of three components: Y_1 , disease process; Y_2 , menopause process; and Y_3 , death process. All these have two states, 0 (event has not yet occurred) and 1 (event has occurred). Hence $\mathcal{S} = \mathcal{S}_1 \times \mathcal{S}_2 \times \mathcal{S}_3$ with $\mathcal{S}_i = \{0, 1\}$ such that (2) is satisfied and the only non-zero transition intensities are $\alpha_1(t; ((0, 0, 0), 1))$ and $\alpha_1(t; ((0, 1, 0), 1))$ corresponding

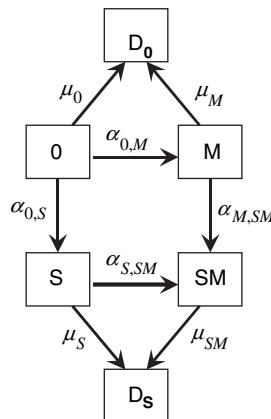


Fig. 1. Transition graph for example 1.

to $\alpha_{0,S}(t)$ and $\alpha_{M,SM}(t)$ for Y_1 , $\alpha_2(t;((0,0,0),1))$ and $\alpha_2(t;((1,0,0),1))$ corresponding to $\alpha_{0,M}(t)$ and $\alpha_{S,SM}(t)$ for Y_2 as well as $\alpha_3(t;((y_1,y_2,0),1))$ corresponding to $\mu_0(t), \dots, \mu_{SM}(t)$ for Y_3 , $(y_1, y_2) \in \mathcal{S}_1 \times \mathcal{S}_2$.

While a transition graph clearly shows what states there are and which transitions are possible, in particular which ones are zero so that absorbing states are easily identified, it does not show independencies between events that correspond to equalities between transition intensities. For instance, if $\alpha_2(t;((0,0,0),1)) = \alpha_2(t;((1,0,0),1))$ then we can say that the occurrence of the event menopause is independent of prior occurrence of skin disease. It is this kind of independencies that we want to focus on and provide an alternative graphically representation for.

3. Local independence

In this section, we introduce and discuss the property of local independence for composable Markov processes. In section 3.2 further properties and implications regarding independencies in these processes are derived.

3.1. Definition

The above property (3) of the transition intensities of a CFMP $\mathbf{Y} \sim (Y_1, \dots, Y_K)$ implies that the dependence structure of its components is determined by the quantities $\alpha_k(t;(\mathbf{y}, y'_k)), \mathbf{y} \in \mathcal{S}, y'_k \in \mathcal{S}_k, k \in V$, alone. As

$$h \cdot \alpha_k(t;(\mathbf{y}, y'_k)) \approx P(Y_k(t+h) = y'_k | \mathbf{Y}(t) = \mathbf{y}) \quad (4)$$

we can say that if $\alpha_k(t;(\mathbf{y}, y'_k))$ is independent of some components of the origin state \mathbf{y} then this is also the case for the probability $P(Y_k(t+h) = y'_k | \mathbf{Y}(t) = \mathbf{y})$ of an instantaneous change in Y_k , and one would intuitively speak of local independence, ‘local’ because it only holds for infinitesimal h . This motivates the following definition.

Definition 2

For a CFMP \mathbf{Y} we say that a component Y_k is locally independent of $Y_j, j \neq k$, given $\mathbf{Y}_{V \setminus \{j, k\}}$ if and only if $\alpha_k(t;(\mathbf{y}, y'_k))$ is constant in the j th component y_j of the first argument for any

$\mathbf{y}_{-j} \in \mathcal{S}_{-j}$ and $y'_k \in \mathcal{S}_k$, $y'_k \neq y_k$. Local independence is symbolized by $Y_j \not\rightarrow Y_k | \mathbf{Y}_{V \setminus \{j, k\}}$ (or shorter $j \not\rightarrow k | V \setminus \{j, k\}$).

Note that the property $j \not\rightarrow k | V \setminus \{j, k\}$ is not a marginal property of the two components Y_j and Y_k as it may depend on the other components \mathbf{y}_{-k} in $\alpha_k(t; (\mathbf{y}, y'_k))$. For example, if the information contained in \mathbf{y}_{-k} is altered by ignoring some components then local independence of Y_k on Y_j might not be preserved.

The case of local independence between subprocesses, i.e. between more than two individual components, is a straightforward generalization of definition 2. For a partition A, B, C of V , we say that $\mathbf{Y}_B \not\rightarrow \mathbf{Y}_A | \mathbf{Y}_C$ if and only if for all $k \in A$ the transition intensities $\alpha_k(t; (\mathbf{y}, y'_k))$ are constant in all the components $y_j, j \in B$, of the first argument for any $\mathbf{y}_{AUC} \in \mathcal{S}_{AUC}$ and $y'_k \in \mathcal{S}_k$, $y'_k \neq y_k$.

Example 1 continued. Based on background knowledge, Aalen *et al.* (1980) assumed that occurrence of death, Y_3 , is locally independent of skin disease and menopause, Y_1 and Y_2 , i.e. there was no differential mortality. Note that if a different type of disease were considered this might not be plausible, but in this particular case it is. It was further found, as one result of the analysis, that menopause is locally independent of skin disease given the person is still alive, i.e. given Y_3 . In contrast, the intensity for developing the skin disease increased after menopause had started given Y_3 . Formally, this means that $\alpha_1(t; ((0, 0, 0), 1)) < \alpha_1(t; ((0, 1, 0), 1))$ while $\alpha_2(t; ((0, 0, 0), 1)) = \alpha_2(t; ((1, 0, 0), 1))$. Note that Y_1 and Y_2 must be locally dependent on Y_3 because once death has occurred all other transitions are zero. In order to avoid such trivial dependencies we may instead consider the stopped process, where we stop at the time when death occurs (more details on this are addressed in Didelez, 2000, p.107; see also Didelez, 2005). A graphical representation using arrows for local dependence is given in Fig. 2. This will be formalized in section 4 below.

Note that arbitrary subprocesses of a CFMP are not necessarily Markov processes. We therefore do not formally consider the situation $\mathbf{Y}_B \not\rightarrow \mathbf{Y}_A | \mathbf{Y}_C$ when A, B, C are not a partition (but see proposition 5 and corollary 2 further down). In this context it should be mentioned that a generalization of local independence to non-Markov processes can loosely be defined as ‘the present of \mathbf{Y}_A is independent of the past of \mathbf{Y}_B given the past of $\mathbf{Y}_{V \setminus B}$ ’. In time-series analysis, the corresponding concept is known as Granger non-causality (Granger, 1969). This intuitive notion of general local independence can be made more formal as a property of the transition-specific compensators (Didelez, 2005; see also Aalen, 1987). We say that a process \mathbf{Y}_A is locally independent of another process \mathbf{Y}_B given \mathbf{Y}_C when its \mathcal{F}_t^{ABC} compensator (or intensity process, if it exists) is measurable with respect to \mathcal{F}_t^{AC} , where \mathcal{F}_t^{ABC} and \mathcal{F}_t^{AC} denotes the internal filtration generated by the three/two subprocesses respectively.

That this is the case for CFMPs can be seen as follows. The associated multivariate counting process for a single component $Y_j, j = 1, \dots, K$, is given by the set of counting processes, each counting a change of state where the destination state differs from the origin in the j th component, i.e. $\mathbf{N}_j(t) = (N_j(t; (y_j, y'_j)) | y_j, y'_j \in \mathcal{S}_j, y_j \neq y'_j)$ with



Fig. 2. Local independence graph for example 1, original process (left) and stopped process (right).

$$N(t; (y_j, y'_j)) = \sum_{s \leq t} \mathbf{1}\{Y_j(s^-) = y_j, Y_j(s) = y'_j\}.$$

Let \mathcal{F}_t be the filtration generated by the whole process \mathbf{Y} . Then, the \mathcal{F}_t -intensities of \mathbf{N}_j are given by $\lambda_j(t) = (\lambda(t; (y_j, y'_j)))|_{y_j, y'_j \in \mathcal{S}_j, y_j \neq y'_j}$ with

$$\lambda(t; (y_j, y'_j)) = \sum_{\mathbf{y}_{-j} \in \mathcal{S}_{-j}} Z_{\mathbf{y}}(t) \alpha_j(t; (\mathbf{y}, y'_j)), \quad j = 1, \dots, K.$$

From this and recalling (1) as well as $Z_{\mathbf{y}}(t) = \mathbf{1}\{\mathbf{Y}(t^-) = \mathbf{y}\}$, it is obvious that the following statements are equivalent: (i) $\alpha_j(t; (\mathbf{y}, y'_j))$ is constant in y_k for all $\mathbf{y}_{-k} \in \mathcal{S}_{-k}$ and all $y'_j \in \mathcal{S}_j, y'_j \neq y_j$; and (ii) $\lambda_j(t)$ is \mathcal{F}_t -measurable for all $t \in \mathcal{T}$.

3.2. Properties of local independence

It can easily be seen that local independence is not symmetric, i.e. if \mathbf{Y}_A is locally independent of \mathbf{Y}_B then \mathbf{Y}_B is not necessarily locally independent of \mathbf{Y}_A . Similarly it is not transitive; if \mathbf{Y}_A is locally independent of \mathbf{Y}_B and \mathbf{Y}_B is locally independent of \mathbf{Y}_C then it is not necessary that \mathbf{Y}_A is locally independent of \mathbf{Y}_C . We will assume that local dependence is reflexive, i.e. \mathbf{Y}_A is always locally dependent of \mathbf{Y}_A .

As detailed in Didelez (2006) we can regard local independence as an asymmetric irrelevance relation comparable with (symmetric) graphoids (Dawid, 1979; Pearl & Paz, 1987). In the following, however, we only require two particular properties of graphoids stated in the next proposition.

Proposition 1

Local independence for CFMPs satisfies the following properties for a partition (A, B, C, D) of V .

1. $A \not\rightarrow B|(C \cup D)$ and $C \not\rightarrow B|(A \cup D)$ then $(A \cup C) \not\rightarrow B|D$ (left intersection)
2. $A \not\rightarrow B|(C \cup D)$ and $A \not\rightarrow C|(B \cup D)$ then $A \not\rightarrow (B \cup C)|D$ (right intersection)

Proof.

1. We know that all the transition intensities $\alpha_k(t; (\mathbf{y}, y'_k))$, $k \in B$, are constant in all the components $y_j, j \in A$, of the first argument for any $\mathbf{y}_{B \cup C \cup D} \in \mathcal{S}_{B \cup C \cup D}$ and are also constant in all the components $y_j, j \in C$, of the first argument for any $\mathbf{y}_{A \cup B \cup D} \in \mathcal{S}_{A \cup B \cup D}$. Hence they are constant in all the components $y_j, j \in (A \cup C)$, of the first argument for any $\mathbf{y}_{B \cup D} \in \mathcal{S}_{B \cup D}$.
2. is obvious from the definition of local independence.

Note that left intersection relies on property (2), e.g. if Y_k is locally dependent of all other components except Y_j and Y_l then

$$\alpha_k(t; (\mathbf{y}, y'_k)) = \alpha_k(t; (\mathbf{y}_{-j}, y'_k)) = \alpha_k(t; (\mathbf{y}_{-l}, y'_k)).$$

The case $\alpha_k(t; (\mathbf{y}_{V \setminus \{j, l\}}, y'_k)) \neq \alpha_k(t; (\mathbf{y}, y'_k))$ can only occur if Y_j and Y_l carry the same information, but this contradicts (2).

Example 1 Continued. In this example, we have that $Y_2 \not\rightarrow Y_3|Y_1$ and $Y_1 \not\rightarrow Y_3|Y_2$ hence by left intersection we obtain $\mathbf{Y}_{\{1, 2\}} \not\rightarrow Y_3$, or in words, the intensity for death is independent of

whether and when the skin disease and/or the menopause have occurred in the past. Further $Y_1 \not\rightarrow Y_2 | Y_3$ yields $Y_1 \not\rightarrow Y_{\{2,3\}}$ by right intersection. As this example is fairly simple, we give a more complex one next.

Example 2. Consider, as a hypothetical example, a CFMP with four components $\mathbf{Y} = (Y_1, Y_2, Y_3, Y_4)$, where Y_1 describes the employment status of women with states $y_1 = 0$ = ‘not employed’, $y_1 = 1$ = ‘employed’, Y_2 the fertility with states y_2 = ‘number of children’ (choosing a maximum m so as to make the state space finite), Y_3 the martial status with states $y_3 = 0$ = ‘single’, $y_3 = 1$ = ‘married’, $y_3 = 2$ = ‘divorced’ and Y_4 indicating whether the person still lives with her parents or not with states $y_4 = 0$ = ‘not at home’ and $y_4 = 1$ = ‘at home’. Note that a transition graph would have to represent $2 \times m \times 3 \times 2$ states, even though there are only four components. One might hypothesize the following local independencies: given the information about whether the woman is employed and married, knowing that she lives at home or not adds no new information w.r.t. the intensity of having a child, i.e. $Y_4 \not\rightarrow Y_2 | \mathbf{Y}_{\{1,3\}}$. Further, knowing the number of children, the employment status as well as the information about where the women lives adds no information w.r.t. the intensity of getting married or divorced, i.e. $Y_1 \not\rightarrow Y_3 | \mathbf{Y}_{\{2,4\}}$ and $Y_4 \not\rightarrow Y_3 | \mathbf{Y}_{\{1,2\}}$. (These assumptions are represented in Fig. 7 later, when local independence graphs have formally been introduced.) With the first local independence it holds that

$$\alpha_2(t; ((y_1, x, y_3, 0), x+1)) = \alpha_2(t; ((y_1, x, y_3, 1), x+1)),$$

for $y_1 = 0, 1, 2$, $y_3 = 0, 1$, and for all $x \in \{0, \dots, m\}$. From the second pair of local independencies we have

$$\alpha_3(t; ((0, x, 0, 0), 1)) = \alpha_3(t; ((1, x, 0, 0), 1)) = \alpha_3(t; ((0, x, 0, 1), 1)) = \alpha_3(t; ((1, x, 0, 1), 1))$$

for all $x \in \{0, \dots, m\}$, and analogous equalities hold for transitions between the states ‘married’ and ‘divorced’. Thus we can write $\alpha_3(t; ((y_1, x, 0, y_4), 1)) = \alpha_3(t; ((x, 0), 1))$, etc. This confirms the property of left intersection, by which $\mathbf{Y}_{\{1,4\}} \not\rightarrow Y_3 | Y_2$. With regard to the corresponding formulation via the intensity processes for all possible transitions from y_3 to y'_3 in component Y_3 , which are $0 \rightarrow 1$, $1 \rightarrow 2$ and $2 \rightarrow 1$, we have that

$$\lambda_3(t; (y_3, y'_3)) = \sum_x \alpha_3(t; ((x, y_3), y'_3)) \mathbf{1}\{Y_2(t^-) = x, Y_3(t^-) = y_3\}$$

which is obviously measurable w.r.t. $\mathcal{F}_t^{(23)}$ again confirming left intersection.

3.3. Conditional independence in CFMPs

Schweder (1970) has shown that local independence is equivalent to a specific conditional independence within a partition $(\mathbf{Y}_A, \mathbf{Y}_B)$ of \mathbf{Y} . If \mathbf{Y}_A is locally independent of \mathbf{Y}_B , i.e. $B \not\rightarrow A$, then (and only then) the following conditional independencies hold

$$\mathbf{Y}_A(t+h) \perp\!\!\!\perp \mathbf{Y}_B(t) | \mathbf{Y}_A(t) \quad \forall h > 0; \quad t, t+h \in \mathcal{T}, \quad (5)$$

where $X \perp\!\!\!\perp Y | Z$ means X is conditionally independent of Y given Z (Dawid, 1979). Hence the infinitesimal independence in (4) holds for any time lag h in this case. Further, the sub-process $\mathbf{Y}_A(t)$ is again a Markov process with transition intensities $\alpha_k(t; (y_A, y'_k))$, $k \in A$. Property (5) suggest the heuristic interpretation of the local independence $B \not\rightarrow A$ that the future of \mathbf{Y}_A is independent of the present of \mathbf{Y}_B given its own present value. It also implies that the independence structure of \mathbf{Y}_A and \mathbf{Y}_B may be depicted in a graphical chain model

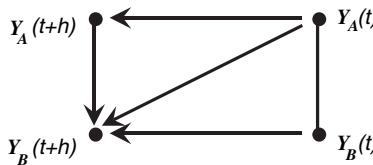


Fig. 3. Graphical chain model representing the local independence $B \not\rightarrow A$.

(Wermuth & Lauritzen, 1990; Frydenberg, 1990) as depicted in Fig. 3. In this graph, the only conditional independence is the one given in (5) as represented by a missing directed edge from $Y_B(t)$ to $Y_A(t+h)$. However, it is not straightforward to generalize (5) to more complex local independence structures. In the above situation, no other conditional independencies hold if \mathbf{Y}_B locally depends on \mathbf{Y}_A . In particular, it does not hold that $\mathbf{Y}_B(t+h) \perp\!\!\!\perp \mathbf{Y}_A(t+h) | \mathbf{Y}(t)$ due to marginalizing over $\mathbf{Y}(s)$, $t < s < t+h$. Hence, in the graphical chain model of Fig. 3 we must include a directed edge from $Y_A(t+h)$ to $Y_B(t+h)$.

Example 1 continued. As explained earlier it follows from left intersection that in equation (5) and Fig. 3 we can choose $B = \{1, 2\}$ and $A = \{3\}$ yielding $Y_3(t+h) \perp\!\!\!\perp \mathbf{Y}_{\{1, 2\}}(t) | Y_3(t)$ meaning that the state of the alive/dead process at time $t+h$ is independent of the states of menopause and skin disease at time t given its own state at time t . Alternatively, we can choose $B = \{1\}$ and $A = \{2, 3\}$ to obtain $\mathbf{Y}_{\{2, 3\}}(t+h) \perp\!\!\!\perp Y_1(t) | \mathbf{Y}_{\{2, 3\}}(t)$.

Next, property (5) is generalized to the case where (A, B) is not a partition of V .

Proposition 2

Let $\mathbf{Y} \sim (Y_1, \dots, Y_K)$ be a CFMP and $A, B \subset V$ with $A \cap B = \emptyset$ and $A \cup B = V$. Define $C = V \setminus (A \cup B)$. Assume that $B \not\rightarrow A | C$.

1. If $C \not\rightarrow A | B$ then \mathbf{Y}_A is a Markov process and property (5) still holds.
2. If $B \not\rightarrow C | A$ then $\mathbf{Y}_{A \cup C}$ is a Markov process and we have the following conditional independence:

$$\mathbf{Y}_A(t+h) \perp\!\!\!\perp \mathbf{Y}_B(t) | \mathbf{Y}_{A \cup C}(t) \quad \forall h > 0; t, t+h \in \mathcal{T}. \quad (6)$$

Proof. The first part follows as $(B \not\rightarrow A | C) \wedge (C \not\rightarrow A | B) \Rightarrow (B \cup C \not\rightarrow A)$ (left intersection), and property (5) yields $\mathbf{Y}_A(t+h) \perp\!\!\!\perp \mathbf{Y}_{B \cup C}(t) | \mathbf{Y}_A(t)$, where $\mathbf{Y}_C(t)$ can be dropped due to the properties of conditional independence (Dawid, 1979). In order to show the second part, note that $B \not\rightarrow A | C$ and $B \not\rightarrow C | A$ implies $B \not\rightarrow A \cup C$ by right intersection. Therefore, we obtain again with property (5)

$$\mathbf{Y}_{A \cup C}(t+h) \perp\!\!\!\perp \mathbf{Y}_B(t) | \mathbf{Y}_{A \cup C}(t), \quad \forall h > 0; t, t+h \in \mathcal{T},$$

which implies the desired result.

In the above proposition, the remaining case is the situation where \mathbf{Y}_A locally depends on \mathbf{Y}_C given \mathbf{Y}_B , and \mathbf{Y}_C locally depends on \mathbf{Y}_B given \mathbf{Y}_A . Then, no conditional independence statement similar to (5) is possible without further restrictions due to possible intermediate transitions in $\mathbf{Y}_C(s)$, $t < s < t+h$, which may induce a dependence between $\mathbf{Y}_A(t+h)$ and $\mathbf{Y}_B(t)$.

Example 2 continued. From the local independencies $Y_4 \not\rightarrow Y_2 | \mathbf{Y}_{\{1, 2\}}$ and $\mathbf{Y}_{\{1, 4\}} \not\rightarrow Y_3 | Y_2$ we cannot infer any independencies using proposition 2. If, in addition, we assumed $Y_4 \not\rightarrow Y_1 | \mathbf{Y}_{\{2, 3\}}$, i.e. the employment status is locally independent of whether the person still lives at home or not, then $Y_4 \not\rightarrow \mathbf{Y}_{\{1, 2, 3\}}$, where $\mathbf{Y}_{\{1, 2, 3\}}$ would still be a Markov process, and using part 2 of proposition 2 we obtain

$$Y_k(t+h) \perp\!\!\!\perp Y_4(t) | \mathbf{Y}_{\{1, 2, 3\}}(t),$$

for $k=1, 2, 3$, where the conditioning set cannot be reduced any further.

The above results have consequences for the situation where a continuous time CFMP can only be observed in discrete time, say at t_1, t_2, \dots, t_n . Proposition 2 implies that the conditional independence structure of $\mathbf{Y}(t_1), \dots, \mathbf{Y}(t_n)$ cannot be captured by a graphical chain model. At least, the characteristic independencies could not be read off a conditional independence graph since there would typically be several dependencies on past components as well as relations among the components at a given discrete point in time due to marginalizing over the time in between measurements.

The results can also be linked to theorem 1 of Hoem (1969) in the following way. Assume that by some intervention we are able to set some of the transition intensities regarding changes in the subprocess \mathbf{Y}_B to zero, e.g. if a cure for a certain disease is found then death from this cause can be eliminated. Hoem (1969) calls the resulting process a partial process. A consequence of proposition 2 is that in this partial process the transition intensities regarding changes in \mathbf{Y}_A remain the same as in the original one if A, B and C satisfy its conditions.

4. Local independence graphs

Graphical models for the representation of conditional independence structure have become more and more popular in the last decades (e.g. Whittaker, 1990; Cox & Wermuth, 1996; Lauritzen, 1996; Edwards, 2000) as they offer many advantages regarding the interpretation and communication of models as well as computational feasibility of statistical and probabilistic inference. The idea of graphical models will now also be exploited to depict the local independence structure of a CFMP.

4.1. Graph terminology

A *directed graph* is a pair $G=(V, E)$, where $V=\{1, \dots, K\}$ is a finite set of vertices representing the components of a CFMP and E is a set of directed edges $(j, k) \in V \times V$, $j \neq k$. In the visualization of a graph directed edges (j, k) are represented by arrows, $j \rightarrow k$. In contrast to the usual conditional independence graphs, we allow that $(j, k) \in E$ and $(k, j) \in E$ which is shown using two stacked arrows, $j \rightleftarrows k$. An undirected edge is given by $\{j, k\}$ and represented as $j-k$. An ordered $(n+1)$ -tuple (j_0, \dots, j_n) of distinct vertices (except possibly $j_0=j_n$) is called a directed path of length n from j_0 to j_n if $(j_{i-1}, j_i) \in E$ for all $i=1, \dots, n$. For $j_0=j_n$ we speak of a *directed cycle*. Such directed cycles are also allowed in the graphs we will consider, as opposed to the well-known directed acyclic graphs (DAGs). For $A \subset V$ the *induced subgraph* G_A is defined as (A, E_A) with $E_A=E \cap (A \times A)$. Furthermore, the *parents* of $A \subset V$ are defined as the set $\text{pa}(A)=\{k \in V \setminus A \mid \exists j \in A : (k, j) \in E\}$, i.e. as the set of those vertices of $V \setminus A$ pointing to at least one vertex in A . A vertex where a directed edge points to is called a *child*. Thus, the set $\text{ch}(A)=\{k \in V \setminus A \mid \exists j \in A : (j, k) \in E\}$ is called the *children* of A . The union of $\text{pa}(A)$ and A itself is referred to as closure of A or $\text{cl}(A)$. If we consider paths instead of single directed edges we can define the *ancestors* of A as $\text{an}(A)=\{k \in V \setminus A \mid \exists j \in A$

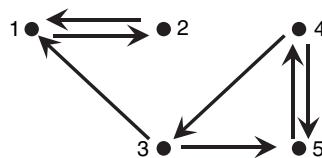


Fig. 4. Directed cyclic graph illustrating the terminology.

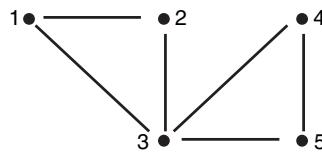
Fig. 5. Subgraph $G_{\{1, 2, 4\}}$ of graph in Fig. 4.

Fig. 6. Moral graph of graph in Fig. 4.

with a path from k to $j\}$, where in case of $\text{pa}(A)=\emptyset$, A is called *ancestral*. Analogously the *descendants* are $\text{de}(A)=\{k \in V \setminus A \mid \exists j \in A \text{ with a path from } j \text{ to } k\}$ and the *non-descendants* are $\text{nd}(A)=V \setminus (A \cup \text{de}(A))$. For general $A \subset V$, $\text{An}(A)$ is the *smallest ancestral* set containing A , given as $A \cup \text{an}(A)$. For cyclic graphs we can construct the so-called *moral graph* G^m in the same way as for DAGs: connect all parents of a common child in G by an undirected edge (if there is no directed edge) and then make G undirected, i.e. replace all directed edges by undirected ones; formally, $G^m=(V, E^m)$ with $E^m=\{(j, k) \mid (j, k) \in E \text{ or } (k, j) \in E \text{ or } \exists i \in V \text{ with } (j, i) \in E \wedge (k, i) \in E\}$. In these undirected graphs we speak of C separating A and B if every path between A and B is intersected by C .

Example 3. The graph in Fig. 4 is an example of a directed graph that contains some cycles. Here, $V=\{1, \dots, 5\}$ and $E=\{(1, 2), (2, 1), (3, 1), (3, 5), (4, 3), (4, 5), (5, 4)\}$. Hence, for instance, we have that $(4, 3, 1, 2)$ forms a directed path from node 4 to 2, while $(3, 5, 4, 3)$ forms a cycle. Further, for example, $\text{pa}(1)=\{2, 3\}$ while $\text{ch}(1)=\{2\}$ and $\text{cl}(1)=\{1, 2, 3\}$. Ancestors are, for instance, the sets $\text{an}(1)=\{2, 3, 4, 5\}$ or $\text{an}(3)=\{4, 5\}$. The only ancestral set apart from (trivially) V itself is $\{3, 4, 5\}$. An example of an induced subgraph G_A with $A=\{1, 2, 4\}$ is shown in Fig. 5. The moral graph G^m is given in Fig. 6. This has an additional edge between nodes 2 and 3 due to their common child node 1. For instance, from G^m we see that node 1 is separated from node 4 by 3, or that nodes 2 and 3 are not separated due to the new edge.

4.2. Dynamic graphical Markov properties

Combining now the notion of graphs and the concept of local independence, a local independence graph of a CFMP is defined as follows.

Definition 3

The local independence graph of a CFMP $\mathbf{Y}_V \sim (Y_1, \dots, Y_K)$ is defined as a directed graph $G = (V, E)$, where for all $j, k \in V$

$$(j, k) \notin E \Rightarrow \{j\} \not\rightarrow \{k\} \mid V \setminus \{j, k\}. \quad (7)$$

Property (7) is called the *pairwise dynamic Markov property* (Didelez, 2000 p. 88). For our purposes we will exploit the equivalence of the above pairwise property to the following.

Proposition 3

Let $\mathbf{Y} \sim (Y_1, \dots, Y_K)$ be a CFMP and $G = (V, E)$ a directed graph. Then property (7) is equivalent to

$$\forall k \in V: V \setminus \text{cl}(k) \not\rightarrow \{k\} \mid \text{pa}(k). \quad (8)$$

Proof. That (7) implies (8) follows from the left intersection property. Now we show that (8) implies (7). From definition 2 it follows with (8) that $\alpha_k(t; (\mathbf{y}, y'_k))$ is a constant function of any $y_j, j \in V \setminus \{\text{cl}(k)\}$ so that $\alpha_k(t; (\mathbf{y}, y'_k)) = \alpha_k(t; (\mathbf{y}_{\text{cl}(k)}, y'_k))$ for all $\mathbf{y}_{\text{cl}(k)} \in \mathcal{S}_{\text{cl}(k)}$ and $y'_k \in \mathcal{S}_k$ with $y'_k \neq y_k, k \in V$. As $V \setminus \{\text{cl}(k)\} = \{j \in V \mid (j, k) \notin E, k \neq j\}$, (7) follows.

The above proposition implies that we can from now on write $\alpha_k(t; (\mathbf{y}, y'_k)) = \alpha_k(t; (\mathbf{y}_{\text{cl}(k)}, y'_k))$. Property (8) may be interpreted such that at any time t the relevant information for determining the intensity for a transition in the k th component is completely contained in the states of the components $\mathbf{Y}_{\text{cl}(k)}$.

Further, the above means that for every component Y_k of a CFMP \mathbf{Y} we can specify conditional intensity matrices (Nodelman *et al.*, 2002) given the states of the components $\mathbf{Y}_{\text{pa}(k)}$ that are parents of Y_k in the local independence graph G . Let a conditional intensity matrix be denoted by $Q_{k|\text{pa}(k)}(t; \mathbf{y}_{\text{pa}(k)})$ then the (q, r) th entry is given by $\gamma_{qr}(t; \mathbf{y}_{\text{pa}(k)}) = \alpha_k(t; (\mathbf{y}_{\text{cl}(k)}, r))$ where in $\mathbf{y}_{\text{cl}(k)}$, $y_k = q$ and $q, r \in \mathcal{S}_k$. The transition behaviour of Y_k is hence characterised by a set of these matrices, one for each value $\mathbf{y}_{\text{pa}(k)} \in \mathcal{S}_{\text{pa}(k)}$.

Example 2 continued. Let us again consider the example of a CFMP $\mathbf{Y} = (Y_1, \dots, Y_4)$, where Y_1 describes the employment status of women, Y_2 the fertility, Y_3 the marital status, and Y_4 indicates whether the person still lives with her parents or not. The pairwise local independencies $Y_4 \not\rightarrow Y_2 \mid \mathbf{Y}_{\{1,3\}}$ and $Y_1 \not\rightarrow Y_3 \mid \mathbf{Y}_{\{2,4\}}$ as well as $Y_4 \not\rightarrow Y_3 \mid \mathbf{Y}_{\{1,2\}}$ result in the local independence graph shown in Fig. 7. By the local dynamic Markov property we have that $\mathbf{Y}_{\{1,4\}} \not\rightarrow Y_3 \mid Y_2$ as $\text{pa}(3) = \{2\}$. As $\text{pa}(1) = \{2, 3, 4\}$, the conditional intensity matrices for Y_1 , for instance, are given by the 2×2 matrices $Q_{1|2,3,4}(t; \mathbf{y}_{\{1,2,3\}})$, and there are $6m$ of them. In contrast, the conditional intensity matrices $Q_{3|2}(t; y_2)$ for Y_3 are 3×3 matrices and there are only m of them, one for each state of Y_2 , as this is the only graph parent of Y_3 .

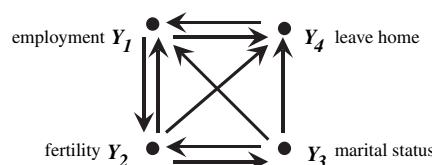


Fig. 7. Local independence graph for the CFMP of example 2.

4.3. Further properties of local independence graphs

For classical conditional independence graphs, collapsibility is an important property which allows reduction of complexity in different regards. Collapsibility of local independence graphs for CFMPs is considered next.

Proposition 4

Let $\mathbf{Y} \sim (Y_1, \dots, Y_K)$ be a CFMP with local independence graph G . If $\text{pa}(A) = \emptyset$, i.e. if A is ancestral, then \mathbf{Y}_A is a CFMP and its local independence graph is given by the induced subgraph $G_A = (A, E_A)$.

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

$$\alpha_A(t; (\mathbf{y}_A, \mathbf{y}')) = \begin{cases} \alpha_k(t; (\mathbf{y}_A, y'_k)), & \text{if } y_k \neq y'_k \text{ and } \mathbf{y}_{A \setminus \{k\}} = \mathbf{y}'_{A \setminus \{k\}} \\ 0, & \text{otherwise.} \end{cases}$$

This proves the claim.

We now restate proposition 2 in graphical terms.

Corollary 1

Let $\mathbf{Y} \sim (Y_1, \dots, Y_K)$ be a CFMP with local independence graph G , and $A, B \subset V$ with $A \cap B = \emptyset$ and $A \cup B \neq V$. Define $C = V \setminus (A \cup B)$. Then

1. if $\text{pa}(A) = \emptyset$ property (5) holds;
2. if $\text{pa}(A \cup C) = \emptyset$ property (6) holds.

Proof. The first condition implies $(B \cup C) \not\rightarrow A$ and the second $B \not\rightarrow (A \cup C)$ from where the proof of proposition 2 applies.

Situation 1. of corollary 1 is depicted in graph Fig. 8A (but note that each of A , B or C could be a collection of more than one vertex). Here the subprocess \mathbf{Y}_A is ‘exogenous’, i.e. there is no ‘input’ from the other processes. Hence it is plausible that property (5) applies. In graph (B) we have situation 2. from the corollary. In this case the subprocess $\mathbf{Y}_{A \cup C}$ is ancestral and hence a Markov process even when we ignore \mathbf{Y}_B . Again it is plausible that (6) holds. Note that both of these arguments can be applied to the situation of example 1, depicted in Fig. 2, with $A = \{3\}$, $B = \{1\}$ and $C = \{2\}$. In example C of Fig. 8 we can also say that $\mathbf{Y}_A(t+h) \perp\!\!\!\perp \mathbf{Y}_B(t) | \mathbf{Y}_A(t)$ without conditioning on $\mathbf{Y}_C(t)$ as past values of the subprocess \mathbf{Y}_C are not informative about \mathbf{Y}_B . More formally this is a consequence of proposition 5 below.

So far we have discussed the implications of a local independence structure for (in)dependencies between the future of some components and the past of others. The following results

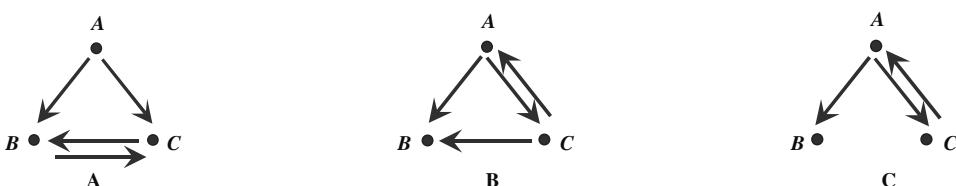


Fig. 8. Graphs illustrating corollary 1.

state which independencies between the whole processes can be read off the local independence graphs. We will denote by $\bar{\mathbf{Y}}_A(t)$ the whole (sub)process up to t , $\{\mathbf{Y}_A(s)|s \leq t\}$.

Proposition 5

Let $\mathbf{Y} \sim (Y_1, \dots, Y_K)$ be a CFMP with local independence graph G . If A, B, C are disjoint subsets of V such that C separates A and B in $G_{\text{An}(A \cup B \cup C)}^m$ then

$$\bar{\mathbf{Y}}_A(t) \perp\!\!\!\perp \bar{\mathbf{Y}}_B(t) | \bar{\mathbf{Y}}_C(t) \quad \forall t \in \mathcal{T}.$$

Proof. This follows from the corresponding property for general counting processes shown in Didelez (2000, p. 100) or Didelez (2005).

Note that in the above either $\mathbf{Y}_{A \cup C}$ or $\mathbf{Y}_{B \cup C}$ must be a Markov process because A has no parents in B and vice versa, and C cannot have parents in both, A and B , otherwise moralization would create an undirected edge between these nodes which would not be intersected by C . Hence either $\text{pa}(A \cup C) = \emptyset$ or $\text{pa}(B \cup C) = \emptyset$. This is, for instance, the situation in Fig. 8C with roles of A and C inverted, so that here $\bar{\mathbf{Y}}_B(t) \perp\!\!\!\perp \bar{\mathbf{Y}}_C(t) | \bar{\mathbf{Y}}_A(t)$.

Example 1 continued. Figure 9 shows the moral graphs relevant for example 1 (cf. the two local independence graphs in Fig. 2). The graph on the left is for the original process and shows that at any time t the histories of the three processes are mutually dependent. The second graph shows the moral graph of the stopped process and can be interpreted as follows: $\bar{\mathbf{Y}}_1(t)$ and $\bar{\mathbf{Y}}_2(t)$ are mutually dependent but independent of $\bar{\mathbf{Y}}_3(t)$ as long as t is smaller than the stopping time when death occurs.

A consequence of the above is stated next and corresponds to theorem 3.9 of Nodelman *et al.* (2002), resembling the local Markov property of DAGs (Lauritzen, 1996, ch. 3).

Corollary 2

Let $\mathbf{Y} \sim (Y_1, \dots, Y_K)$ be a CFMP with local independence graph G . Then

$$\bar{\mathbf{Y}}_A(t) \perp\!\!\!\perp \bar{\mathbf{Y}}_{\text{nd}(A) \setminus \text{pa}(A)}(t) | \bar{\mathbf{Y}}_{\text{pa}(A)}(t) \quad \forall t \in \mathcal{T}.$$

Proof. This is easily seen from proposition 5 by noting that A and $\text{nd}(A)$ are always separated by $\text{pa}(A)$ in $G_{\text{An}(A, \text{nd}(A))}^m$.

In order to illustrate the above proposition we need a more complex example than those used so far.

Example 4. The example in Fig. 10 is taken from Nodelman *et al.* (2002). It depicts how a drug takes effect as mediated by how fast it reaches the blood stream, which could be affected by how recently the person has eaten, and further its effect on joint pain or drowsiness. Most of the components of this process can be formulated as having two states, e.g. Y_2 feeling joint



Fig. 9. Moral graph for example 1, original process (left) and stopped process (right).

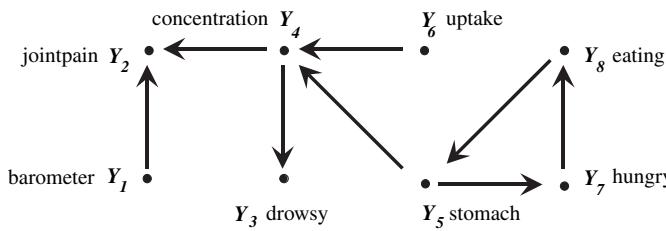


Fig. 10. Local independence graph for drug effect example.

pain has states 'yes' or 'no' at any given time t . Only atmospheric pressure, Y_1 , and concentration of the drug in the blood, Y_4 , may have more states, such as 'low', 'middle' and 'high'. The local independence assumptions here are that whether or not a person feels joint pain (Y_2) is locally independent of any other information once we know the level of concentration of the drug in the blood (Y_4) and the atmospheric pressure (Y_1). Further, whether a person is drowsy or not (Y_3) is assumed to be locally independent of any other information once we know the level of concentration of the drug in the blood, while this concentration is itself locally independent of the rest once it is known whether the drug has been taken in the past (Y_6) and whether the person has a full stomach (Y_5). Being hungry (Y_7), eating (Y_8) and having a full stomach (Y_5) form an obvious cycle.

We can apply proposition 4 in order to identify all subprocesses that are Markov processes; apart from V itself these are all ancestral sets, $\{5, 7, 8\}$, $\{6\}$, $\{4, 5, \dots, 8\}$, $\{3, 4, \dots, 8\}$ and $\{1\}$. From proposition 5 and the corresponding moral graph $G_{\text{An}(3, 4, 5, 7, 8)}^m$ in Fig. 11 we can further infer that

$$\bar{Y}_{\{3, 4\}}(t) \perp\!\!\!\perp \bar{Y}_{\{7, 8\}}(t) \mid \bar{Y}_5(t) \quad \forall t \in \mathcal{T},$$

meaning that at any given time t the past history of concentration and drowsiness is independent of the past history of being hungry and eating given we know the history of full stomach. From corollary 2 we see for example with $\text{pa}(4) = \{5, 6\}$ and $\text{nd}(4) \setminus \text{pa}(4) = \{1, 7, 8\}$

$$\bar{Y}_4(t) \perp\!\!\!\perp \bar{Y}_{\{1, 7, 8\}}(t) \mid \bar{Y}_{\{5, 6\}}(t) \quad \forall t \in \mathcal{T}.$$

4.4. Inference for local independence graphs

The likelihood of a composable Markov process can be derived from the likelihood of counting processes (Andersen *et al.*, 1993, p. 95) exploiting the relation described in section 2. With the local Markov property (8), the likelihood can be factorized according to the graphical structure such that the k th factor of the likelihood only depends on the states $\mathbf{y}_{\text{cl}(k)}$ taken by those components that are in the closure of the k th component (Didelez, 2005). More formally, let \mathbf{Y}_V be a CFMP and G the associated local independence graph. With

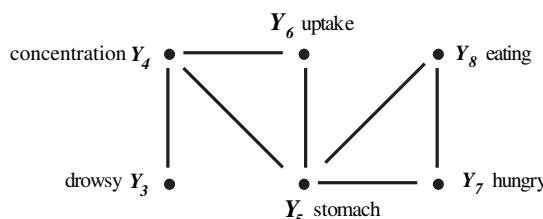


Fig. 11. Moral graph for drug effect example.

$\bar{\mathbf{Y}}_V(t) = \{\mathbf{Y}_V(s) | 0 \leq s \leq t\}$, the likelihood $L(t|\bar{\mathbf{Y}}_V(t))$ then reads as

$$L(t|\bar{\mathbf{Y}}_V(t)) = \prod_{k \in V} L_k(t|\bar{\mathbf{Y}}_{\text{cl}(k)}(t)),$$

where

$$L_k(t|\bar{\mathbf{Y}}_{\text{cl}(k)}(t)) = \prod_{y_k \neq y'_k} \prod_{T_{s(\mathbf{y}_{\text{cl}(k)}, y'_k)} \leq t} \lambda_k(T_{s(\mathbf{y}_{\text{cl}(k)}, y'_k)}) \cdot \exp\left(-\int_0^t \lambda_k(s) ds\right) \quad (9)$$

and where $T_{s(\mathbf{y}_{\text{cl}(k)}, y'_k)}$ are the times of occurrences of a transition in the k th component from y_k to y'_k given that the remaining components have the values $\mathbf{y}_{\text{cl}(k) \setminus \{k\}}$. The intensities $\lambda_k(T_{s(\mathbf{y}_{\text{cl}(k)}, y'_k)})$ are defined analogously to (1), where the transition from q to r is now the one from $\mathbf{y}_{\text{cl}(k)}$ to y'_k but for ease of notation the index $(\mathbf{y}_{\text{cl}(k)}, y'_k)$ is replaced by k .

When aiming at inference about the local independence structure the main hypothesis is the one on pairwise local independence of, say, Y_k from Y_j given the remaining components. Assume that $G = (V, E)$ is the local independence graph of a CFMP \mathbf{Y}_V containing the edge $(j, k) \in E$. We want to test whether the latter can be omitted. This yields the following null hypothesis: for all $\mathbf{y}_{\text{cl}(k)}, \tilde{\mathbf{y}}_{\text{cl}(k)} \in \mathcal{S}_{\text{cl}(k)}$ where $\mathbf{y}_{\text{cl}(k)}$ and $\tilde{\mathbf{y}}_{\text{cl}(k)}$ only differ in the j th component, and for all $y'_k \neq y_k$ it holds that

$$\alpha_k(t; (\mathbf{y}_{\text{cl}(k)}, y'_k)) = \alpha_k(t; (\tilde{\mathbf{y}}_{\text{cl}(k)}, y'_k)) \quad \forall t \in \mathcal{T}. \quad (10)$$

A likelihood ratio (LR) test of this hypothesis can be carried out based on the log-likelihoods \hat{l}_k and \hat{l}_k^0 (Andersen *et al.*, 1993, p. 420), derived from maximizing (9) under the model G and under the null model, i.e. the graph G without the edge (j, k) respectively. Similarly, we may construct more general LR tests for testing a graph $G^0 = (V, E^0)$ versus a graph $G^1 = (V, E^1)$, $E^0 \subset E^1$, i.e. they may differ in more than one edge but are nested. Then, the LR test is based on comparing the log-likelihoods for those $k \in V$ where $j \neq k$ exists with $(j, k) \in E^1$ and $(j, k) \notin E^0$. In the stationary case, explicit formulae for the maximum likelihood estimators and hence the LR test statistic can be derived (Didelez, 2000, p. 131); for a Bayesian approach to inference in homogeneous graphical Markov processes (see Nodelman *et al.*, 2003). In the non-stationary case, a parametric assumption about the time dependence is required for likelihood-based inference.

Alternatively, if no parametric assumptions are made for the intensity processes non-parametric inference can be based on the cumulative hazards

$$A_k(t; (\mathbf{y}_{\text{cl}(k)}, y'_k)) = \int_0^t \alpha(s; (\mathbf{y}_{\text{cl}(k)}, y'_k)) ds.$$

These can be estimated by the Nelson–Aalen estimator. If we then want to test for omission of the edge (j, k) the null hypothesis is similar to (10) but replacing intensities by cumulative hazards. To carry out the tests the methods of Andersen *et al.* (1993, p. 345) can be used, i.e. essentially we compare the estimators of the above cumulative hazard for each $y_j \in \mathcal{S}_j$ with the one where the origin state $\mathbf{y}_{\text{cl}(k)}$ is replaced by $\mathbf{y}_{\text{cl}(k) \setminus \{j\}}$, and this has to be performed for each value of $\mathbf{y}_{\text{cl}(k) \setminus \{j\}}$.

Example 4 continued. Consider the situation of Fig. 10 but with an additional arrow from Y_5 (full stomach) to Y_3 (drowsiness). Assume we want to test whether this directed edge can be omitted, i.e. whether drowsiness is locally independent of full stomach given the concentration of the drug in the blood. The assumption of the remaining local independencies means that we do not need to take $\mathbf{Y}_{\{1, 2, 6, 7, 8\}}$ into account at all. Under a parametric model we would test whether $\alpha_3(t; (\mathbf{y}_{\{3, 4, 5\}}, y'_3))$ is equal to $\alpha_3(t; (\mathbf{y}_{\{3, 4\}}, y'_3))$ for all t and for each

value of y_5 and each value of $(y_{\{3,4\}}, y'_3)$. This means we investigate whether for each given level of Y_4 (concentration of drug in blood) the transition intensities for becoming drowsy is the same for those with full stomach and those with empty stomach. And this has to be repeated for the transition intensities for becoming awake.

5. Discussion

We have proposed a graphical way of representing local independencies within a composable Markov process, i.e. a process that can be thought of as consisting of several components. These graphs provide a simple and intuitive way of reading off local independencies and further properties of the Markov process, such as which subprocesses are again Markov or which subprocesses are independent of each other possibly conditional on some third subprocess. Hence, local independence graphs facilitate and help to structure the communication, reasoning and analysis in complex dynamic data situations.

It is possible to include time-constant variables in these graphs, representing them by nodes that cannot be children of any of the nodes that represent processes in order to reflect that time-constant variables are fixed at $t=0$, before any of the processes start. Also, a generalization to non-Markov processes is possible (Didelez, 2005, 2006), but requires a more general definition of local independence. It is then possible to define a suitable notion of graph separation for cyclic graphs that allows to read off local independencies between subprocesses without requiring to stay in the class of Markov processes.

One of the main benefits of graphical models is that they allow to simplify computations by exploiting the graphical structure. This is particularly relevant for expert system which are based on DAGs representing conditional independencies (Cowell *et al.*, 1999). Such expert systems require a fully specified probability model and are then used to compute fast answers to ‘queries’ such as ‘if we observe that a patient has a specific side effect what is the probability that he will recover within a day?’. In the case of local independence graphs such queries can in principle be answered exactly based on the matrix of transition intensities for the whole process but computations quickly become very complex with increasing number of components. As discussed by Nodelman *et al.* (2002) an acceleration of such computations exploiting the graphical structure is not straightforward even in the homogeneous case. Instead, they propose a much faster approximate method which is based on the cliques of the graph. A different type of query would be to ask what the local (in)dependence structure is when we condition on certain events happening in the future, e.g. because we have only sampled the ‘cases’ so that the data represent a process of which we know that by a certain time the transition into the state ‘ill’ has taken place. Hoem (1969) deals with this question and calls such processes ‘purged’ processes. Again, it is in principle possible to calculate the transition intensities of a purged process exactly, but approximations based on the graphical structure may be much faster. Also, it would be desirable to have additional graphical rules that can tell us which local independencies of the original process are preserved in the purged process.

Another situation where acceleration of computation is called for is when the graph is unknown and has to be found from data. This seems to be simpler for local independence graphs than for DAGs as they are allowed to be cyclic, which means that there are no restrictions on the parent sets, and there is no problem of Markov equivalence of different graphs, i.e. different local independence graphs always represent different probabilistic models. For a Bayesian approach to model search in homogeneous Markov processes see Nodelman *et al.* (2003). Clearly, more research into the fitting of local independence graphs for specific applications as well as into fast computations and model search is still needed.

Acknowledgements

I thank Iris Pigeot for her persistent encouragement and advice. Financial support by the German Research Foundation (SFB 386) is also gratefully acknowledged. This paper has been finalized while I was a member of an international research group studying Statistical Analysis of Complex Event History Data at the Centre for Advanced Study at the Norwegian Academy of Science and Letters in Oslo during the academic year 2005/2006. Helpful comments from the editor and a referee are greatly appreciated.

References

Aalen, O. O. (1987). Dynamic modelling and causality. *Scand. Actuar. J.* 177–190.

Aalen, O. O., Borgan, Ø., Keiding, N. & Thormann, J. (1980). Interaction between life history events: nonparametric analysis of prospective and retrospective data in the presence of censoring. *Scand. J. Statist.* **7**, 161–171.

Andersen, P. K., Borgan, Ø., Gill, R. D. & Keiding, N. (1993). *Statistical models based on counting processes*. Springer, New York.

Cowell, R. G., Dawid, A. P., Lauritzen, S. L. & Spiegelhalter, D. J. (1999). *Probabilistic networks and expert systems*. Springer, New York.

Cox, D. R. & Wermuth, N. (1996). *Multivariate dependencies – models, analysis and interpretation*. Chapman and Hall, London.

Dawid, A. P. (1979). Conditional independence in statistical theory. *J. Roy. Statist. Soc. Ser. B* **41**, 1–31.

Didelez, V. (2000). *Graphical models for event history data based on local independence*. Logos, Berlin.

Didelez, V. (2005). *Graphical models for marked point processes based on local independence*. Technical Report 258, University College London, London.

Didelez, V. (2006). Asymmetric separation for local independence graphs. *Proceedings of the 22nd Annual Conference on Uncertainty in Artificial Intelligence*, 130–137. AUAI Press, Corvallis, Dregon.

Edwards, D. (2000). *Introduction to graphical modelling*, 2nd edn. Springer, New York.

Eichler, M. (1999). *Graphical models in time series analysis*. PhD thesis, University of Heidelberg, Heidelberg.

Eichler, M. (2000). *Causality graphs for stationary time series*. Technical report, University of Heidelberg, Heidelberg.

Florens, J. P. & Fougère, D. (1996). Noncausality in continuous time. *Econometrica* **64**, 1195–1212.

Frydenberg, M. (1990). The chain graph Markov property. *Scand. J. Statist.* **17**, 333–353.

Gottard, A. (2002). *Graphical duration models*. Technical Report 2002/07, University of Florence, Florence.

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

Hoem, J. M. (1969). Purged and partial Markov chains. *Skand. Aktuar. Tidskr.* 147–155.

Lauritzen, S. L. (1996). *Graphical models*. Clarendon Press, Oxford.

Nodelman, U., Shelton, C. R. & Koller, D. (2002). Continuous time Bayesian networks. In *Proceedings of the 18th Annual Conference on Uncertainty in Artificial Intelligence*, 378–387. Morgan Kaufman, San Francisco, CA. Available at: <http://www.auai.org/> (accessed July 18, 2006).

Nodelman, U., Shelton, C. R. & Koller, D. (2003). Learning continuous time Bayesian networks. In *Proceedings of the 19th Annual Conference on Uncertainty in Artificial Intelligence*, 451–458. Morgan Kaufman, San Francisco, CA. Available at: <http://www.auai.org/> (accessed July 18, 2006).

Pearl, J. & Paz, A. (1987). Graphoids: a graph-based logic for reasoning about relevancy relations. In *Advances in Artificial Intelligence–II*, (eds B. Du Boulay, D. Hogg, & L. Steel) 357–363. North Holland, Amsterdam.

Schweder, T. (1970). Composable Markov processes. *J. Appl. Probab.* **7**, 400–410.

Wermuth, N. & Lauritzen, S. L. (1990). On substantive research hypotheses, conditional independence graphs and graphical chain models (with discussion). *J. Roy. Statist. Soc. Ser. B* **52**, 21–72.

Whittaker, J. (1990). *Graphical models in applied multivariate statistics*. Wiley, Chichester.

Received October 2005, in final form June 2006

Vanessa Didelez, Department of Statistical Science, University College London, WC1E6BT London, UK.
E-mail: vanessa@stats.ucl.ac.uk