J. R. Statist. Soc. B (2008)
70, *Part* 1, *pp.* 245–264

Graphical models for marked point processes based on local independence

Vanessa Didelez

University of Bristol, UK

[Received May 2006. Revised September 2007]

Summary. A new class of graphical models capturing the dependence structure of events that occur in time is proposed. The graphs represent so-called local independences, meaning that the intensities of certain types of events are independent of some (but not necessarilly all) events in the past. This dynamic concept of independence is asymmetric, similar to Granger non-causality, so the corresponding local independence graphs differ considerably from classical graphical models. Hence a new notion of graph separation, which is called δ-separation, is introduced and implications for the underlying model as well as for likelihood inference are explored. Benefits regarding facilitation of reasoning about and understanding of dynamic dependences as well as computational simplifications are discussed.

Keywords: Conditional independence; Counting processes; Event history analysis; Granger causality; Graphoid; Multistate models

1. Introduction

Marked point processes are commonly used to model event history data, a term originating from sociology where it is often of interest to investigate the dynamics behind events such as finishing college, finding a job, marrying, starting a family, durations of unemployment and illness. But comparable data situations also occur in other contexts, e.g. in survival analyses with intermediate events such as the onset of a side effect or a change of medication (see Keiding (1999)). Longitudinal studies and the careful analysis of the underlying processes are crucial for gaining insight into the driving forces of inherently dynamic systems, but having to deal with the multidimensionality as well as with the dynamic nature of these systems makes this a very complex undertaking.

Graphical models deal with complex data structures that arise whenever the interrelationship of variables in a multivariate setting is investigated. Over the last two decades, they have proven to be a valuable tool for probabilistic modelling and multivariate data analysis in such different fields as expert systems and artificial intelligence (Pearl, 1988; Cowell *et al.*, 1999; Jordan, 1999) and hierarchical Bayesian modelling (see the BUGS project at `http://www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml`), causal reasoning (Pearl, 2000; Spirtes *et al.*, 2000), as well as sociological, biomedical and econometric applications. For overviews and many different applications see for instance Whittaker (1990), Cox and Wermuth (1996) and Edwards (2000).

Whereas the 'classical' graphical models are concerned with representing conditional independence structures among random variables, variations have been proposed to deal with feedback

Address for correspondence: Vanessa Didelez, Department of Mathematics, University of Bristol, University Walk, Bristol, BS8 1TW, UK.
E-mail: vanessa.didelez@bristol.ac.uk

 1369–7412/08/70245

systems (Spirtes, 1995; Koster, 1996) but are still based on cross-sectional data. The application of graphical models to truly time-dependent data, such as event histories or time series, is only slowly beginning to progress. Dahlhaus (2000) has proposed graphical models for multivariate time series, but his approach does not capture how the present or future of the system depends on or is affected by the past. Instead, Eichler (1999, 2000) used graphs to represent Granger causality, which is a dynamic concept of dependence. For continuous time one approach, called dynamic Bayesian networks, is to discretize time and to provide directed acyclic graphs (DAGs) that encode the independence structure for the transitions from t to $t+1$ (Dean and Kanazawa, 1989). The approach that was proposed by Nodelman *et al.* (2002, 2003) comes closest to the kind of graphs that we shall consider. In their continuous time Bayesian networks they represent a multistate Markov process with nodes corresponding to subprocesses and edges corresponding to dependences of transition rates on states of other subprocesses.

In this paper, we propose and investigate the properties of graphs that represent so-called 'local independence' structures in event history data. The basic idea of local independence is that, once we know about specific past events, the intensity of a considered future event is independent of other past events. It has been developed by Schweder (1970) for the case of Markov processes and applied for example in Aalen *et al.* (1980). A generalization to processes with a Doob–Meyer decomposition can be found in Aalen (1987) who focused on the bivariate case, i.e. local dependence between two processes. Here we extend this approach to more than two processes. The analogy of the bivariate case to Granger non-causality has been pointed out by Florens and Fougère (1996); see also Comte and Renault (1996). Note that the notion of 'local independence' that was used by Allard *et al.* (2001) is a different one.

We first set out the necessary notation and assumptions for marked point processes in Section 2.1 followed by the formal definition of local independence in Section 2.2, the emphasis being on the generalization to a version that allows us to condition on the past of other processes and hence describes dynamic dependences for multivariate processes. Section 3.1 defines graphs that are appropriate to represent local (in)dependence. The main results are given in Section 3.2. The properties of local independence graphs are investigated. In particular we prove that a new notion of graph separation, called δ-separation, can inform us about independences that are preserved after marginalizing over some of the processes. In Section 3.3, it is shown how the likelihood of a process with given local independence graph factorizes and implications are discussed. The potential of local independence graphs is discussed in Section 4 and proofs are given in Appendix A.

2. Local independence for marked point processes

Marked point processes are briefly reviewed in Section 2.1, using the notation of Andersen *et al.* (1993). In Section 2.2 the concept of local independence is explained in detail.

2.1. Marked point processes and counting processes

Let $\mathcal{E}=\{e_1,\ldots,e_K\}$, $K<\infty$, denote the (finite) *mark space,* i.e. the set containing all *types* of events of interest for one observational unit, and $\mathcal{T}$ the *time space* in which the observations take place. We assume that time is measured continuously so that we have $\mathcal{T}=[0,\tau]$ or $\mathcal{T}=[0,\tau)$ where $\tau<\infty$. The *marked point process (MPP) Y* consists of *events* given by pairs of variables (T_s, E_s), $s=1,2,\ldots,$ on a probability space $(\Omega,\mathcal{F},P)$ where $T_s\in\mathcal{T}$, $0<T_1<T_2\ldots$ are the times of occurrences of the respective types of events $E_s\in\mathcal{E}$. Assume that the MPP is non-explosive, i.e. only a finite number of events occurs in the time span $\mathcal{T}$. The *mark-specific counting processes* $N_k(t)$ that are associated with an MPP are then given by

$$N_k(t) = \sum_{T_s \leqslant t} \mathbf{1}\{E_s = e_k\}, \qquad k = 1, \ldots, K.$$

We write $\mathbf{N} = (N_1, \ldots, N_K)$ for the multivariate counting process, and $\mathbf{N}_A$, $A \subset \{1, \ldots, K\}$, for the vector $(N_k)_{k \in A}$, calling $\mathbf{N}_A$ a *subprocess*, with $\mathbf{N}_V = \mathbf{N}$.

To investigate dependences of the present on the past it will be important to have some notation for the history of some subset or all of the processes that are involved. Hence we denote the internal filtration of a marked point process by $\mathcal{F}_t = \sigma\{(T_s, E_s) | T_s \leqslant t, E_s \in \mathcal{E}\}$ which is equal to $\sigma\{(N_1(s), \ldots, N_K(s)) | s \leqslant t\}$, whereas for $A \subset \{1, \ldots, K\}$ we define the filtrations of a subprocess as $\mathcal{F}_t^A = \sigma\{\mathbf{N}_A(s) | s \leqslant t\}$; in particular $\mathcal{F}_t^k$ is the internal filtration of an individual counting process N_k.

Under quite general assumptions (see Fleming and Harrington (1991), page 61), a Doob–Meyer decomposition of $N_k(t)$ into a compensator and a martingale exists. Both these processes depend on the filtration considered, which is here taken to be the internal filtration of the *whole* MPP Y. We shall assume throughout that all the $\mathcal{F}_t$-compensators Λ_k are absolutely continuous and predictable so that intensity processes $\lambda_k(t)$ exist, which are taken to be predictable versions of the derivatives of the compensators, i.e. $\Lambda_k(t) = \int_0^t \lambda_k(s)\, \mathrm{d}s$. Heuristically we have (Andersen *et al.* (1993), page 52)

$$\lambda_k(t)\, \mathrm{d}t = E\{N_k(\mathrm{d}t) | \mathcal{F}_{t^-}\}. \tag{1}$$

More formally this means that the differences $N_k - \Lambda_k$ are $\mathcal{F}_t$-martingales.

An interpretation of property (1) is that, given the information on the history of the whole MPP up to just before time t, $\lambda_k(t)\, \mathrm{d}t$ is our best prediction of the immediately following behaviour of N_k. For the theory that is developed in this paper the setting can slightly be generalized, not requiring absolute continuity of compensators (see Didelez (2000)), as might be relevant when certain types of events can only occur at fixed times.

It will be important to distinguish between the $\mathcal{F}_t$-intensity that is based on the past of the *whole* MPP, and the $\mathcal{F}_t^A$-intensities that are based on the past of the subprocess on marks in A. The latter can be computed by using the innovation theorem (Brémaud (1981), page 83), and a way of doing so, which is especially relevant to our setting, is given in Arjas *et al.* (1992).

The following is a standard assumption in counting process theory but we want to highlight it as it plays a particularly important role for local independence graphs.

Assumption 1 (no jumps at the same time). The $\mathcal{F}_t$-martingales $N_k - \int \lambda_k(s)\, \mathrm{d}s$ are assumed to be orthogonal for $k \in \{1, \ldots, K\}$, meaning that none of $N_1, \ldots, N_K$ jump at the same time. This is implied by the above assumption that all compensators are absolutely continuous if in addition no two counting processes N_j and N_k are counting the same type of event.

Assumption 1 might be violated, e.g. when investigating the survival times of couples and there is a small but non-zero chance that they die at the same time, in a car accident for instance. The reason for imposing this assumption is that we want to explain dependences between events by the past not by common innovations. If one wants to allow events to occur at the same time, then such a simultaneous occurrence defines a new mark in the mark space.

Note that general multistate processes can be represented as marked point processes with every transition between two states being a mark. This is explored in more detail in Didelez (2007) for Markov processes.

2.2. Local independence

The bivariate case is defined as follows (see Aalen (1987)).

Definition 1 (local independence (bivariate)). Let Y be an MPP with $\mathcal{E}=\{e_1, e_2\}$ and N_1 and N_2 the associated counting processes on $(\Omega, \mathcal{F}, P)$. Then, N_1 is said to be *locally independent* of N_2 over $\mathcal{T}$ if $\lambda_1(t)$ is measurable with respect to $\mathcal{F}_t^1$ for all $t \in \mathcal{T}$. Otherwise we speak of local dependence.

The process N_1 being locally independent of N_2 is symbolized by $N_2 \not\to N_1$. Interchangeably we shall sometimes say that e_1 is locally independent of e_2, or $e_2 \not\to e_1$.

The essence of this definition is that the intensity $\lambda_1(t)$, i.e. our 'short-term' prediction of N_1, remains the same under the reduced filtration $\mathcal{F}_t^1$ as compared with the full filtration $\mathcal{F}_t$. This implies that we do not lose any essential information by ignoring how often and when event e_2 has occurred before t. We could say that if $N_2 \not\to N_1$ then the presence of N_1 is conditionally independent of the past of N_2 given the past of N_1, or heuristically

$$N_1(t) \perp\!\!\!\perp \mathcal{F}_{t^-}^2 | \mathcal{F}_{t^-}^1, \tag{2}$$

where $A \perp\!\!\!\perp B|C$ means that 'A is conditionally independent of B given C' (see Dawid (1979)). (Expression (2) is an informal way of saying that $N_1(\mathrm{d}t)$ is conditionally independent of $\{T_s | T_s < t, E_s = 2, s = 1, 2, \ldots\}$ given $\{T_s | T_s < t, E_s = 1, s = 1, 2, \ldots\}$. This and similar statements later, like expressions (3), (12) or (15), should be interpreted correspondingly.) For general processes, this is a stronger property than local independence but it holds for marked point processes with assumption 1, as their distributions are determined by the intensities. Expression (2) does not imply that, for $u > 0$, $N_1(t+u) \perp\!\!\!\perp \mathcal{F}_{t^-}^2 | \mathcal{F}_{t^-}^1$—hence the name *local* independence. Also, $N_1(t) \perp\!\!\!\perp N_2(t)$ will hold only if the two processes are mutually locally independent of each other. Without assumption 1 the $\mathcal{F}_t^1$-measurability of $\lambda_1(t)$ would, for instance, trivially be true if $e_1 = e_2$ but, in such a case, we would not want to speak of independence of e_1 and e_2.

Example 1: skin disease—in a study with women of a certain age, Aalen *et al.* (1980) modelled two events in the life of an individual woman: the occurrence of a particular skin disease and onset of menopause. Their analysis revealed that the intensity for developing this skin disease is greater once the menopause has started than before. In contrast, and as we would expect, the intensity for onset of menopause does not depend on whether the person has earlier developed this skin disease. We can therefore say that menopause is locally independent of this skin disease but not vice versa. Note that, in whatever way the onset of skin disease and menopause are measured, it is assumed that they do not start systematically at exactly the same time, corresponding to the above 'no jumps at the same time' assumption.

Let us now turn to the case of more than two types of event. This requires conditioning on the past of other processes as follows.

Definition 2 (local independence (multivariate)). Let $\mathbf{N} = (N_1, \ldots, N_K)$ be a multivariate counting process that is associated with an MPP. Let further A, B and C be disjoint subsets of $\{1, \ldots, K\}$. We then say that a *subprocess* $\mathbf{N}_B$ *is locally independent of* $\mathbf{N}_A$ *given* $\mathbf{N}_C$ over $\mathcal{T}$ if all $\mathcal{F}_t^{A\cup B\cup C}$-intensities λ_k, $k \in B$, are measurable with respect to $\mathcal{F}_t^{B\cup C}$ for all $t \in \mathcal{T}$. This is denoted by $\mathbf{N}_A \not\to \mathbf{N}_B | \mathbf{N}_C$ or in short $A \not\to B|C$. Otherwise, $\mathbf{N}_B$ is *locally dependent* on $\mathbf{N}_A$ given $\mathbf{N}_C$, i.e. $A \to B|C$. If $C = \emptyset$ then B is *marginally* locally (in)dependent of A.

Conditioning on a subset C thus means that we retain the information about whether and when events with marks in C have occurred in the past when considering the intensities for marks in B. If these intensities are independent of the information on whether and when events with marks in A have occurred we have conditional local independence. By analogy with expression (2) this definition of multivariate local independence is with assumption 1 equivalent to

$$N_B(t) \perp\!\!\!\perp \mathcal{F}_{t^-}^A | \mathcal{F}_{t^-}^{B\cup C} \qquad \forall t \in \mathcal{T}. \tag{3}$$

Example 2: home visits—this example is not taken from the literature but is inspired by real studies (e.g. Vass *et al.* (2002, 2004)). In some countries there are programmes to assist the elderly through regular home visits by a nurse. This is meant to reduce unnecessary hospitalizations while increasing the quality of life for the person. It is hoped that such a programme increases the survival time. The times of the visits as well as the times and durations of hospitalization are monitored. In addition, it is plausible that the underlying health status of the elderly person may also affect the rate of hospitalization and predict survival. This interplay of events for an individual elderly person can be represented as an MPP if 'health status' is regarded as a multistate process. Assume that the timing of the home visits is determined externally, e.g. by the availability of nurses which has nothing to do with the patient's development, i.e. the visits are assumed locally independent of all the remaining processes. It might then be of interest to investigate whether the visits affect only the rate of hospitalization directly, i.e. whether survival is locally independent of the visits process given the hospitalization and health history or even given only a subset thereof.

As can easily be checked, local (in)dependence needs to be neither symmetric, reflexive nor transitive. However, since in most practical situations a subprocess depends at least on its own past we shall assume throughout that local dependence is reflexive. An example for a subprocess that depends only on the history of a different subprocess and not on its own history is given in Cox and Isham (1980), page 122.

To see the relation with local independence, we briefly review Granger non-causality (Granger, 1969). Let $X_V = \{X_V(t) | t \in \mathbf{Z}\}$ with $X_V(t) = (X_1(t), \ldots, X_K(t))$ be a multivariate time series, where $V = \{1, \ldots, K\}$ is the index set. For any $A \subset V$ we define $X_A = \{X_A(t)\}$ as the multivariate subprocess with components X_a, $a \in A$. Further let $\bar{X}_A(t) = \{X_A(s) | s \leqslant t\}$. Then, for disjoint subsets $A, B \subset V$, we say that X_A is strongly Granger non-causal for X_B if

$$X_B(t) \perp\!\!\!\perp \bar{X}_A(t-1) | \bar{X}_{V \backslash A}(t-1),$$

for all $t \in Z$. The interpretation is similar to that for local independence, i.e. the present value of X_B is independent of the past of X_A given its own past and the past of all other components $C = V \backslash (A \cup B)$, by analogy with expression (3). Also note that the above does not imply that

$$X_B(t+u) \perp\!\!\!\perp \bar{X}_A(t-1) | \bar{X}_{V \backslash A}(t-1)$$

for $u > 0$, which is again analogous to local independence. Eichler (1999, 2000) investigated a graphical representation and rules to determine when the condition $\bar{X}_{V \backslash A}(t-1)$ can be reduced to proper subsets $\bar{X}_C(t-1)$, $C \subset V \backslash A$.

Finally, let us indicate how the definition of local independence can be generalized to stopped processes. This is relevant when there are absorbing states such as death. In that case all other events will be locally dependent on this one because all intensities are 0 once death has occurred. However, the dependence is 'trivial' and not of much interest. Let T be an $\mathcal{F}_t$-stopping time and let $\mathbf{N}^T = (N_1^T, \ldots, N_K^T)$ be the multivariate counting process that is stopped at T. Then the intensities of N_k^T are given by λ_k^T, $k \in V$, and local independence can be formalized as follows.

Definition 3 (local independence for stopped processes). Let $\mathbf{N}^T = (N_1^T, \ldots, N_K^T)$ be a multivariate counting process that is associated with an MPP and stopped at time T. Then we say that $A \not\rightarrow B | C$ if there are $\mathcal{F}_t^{B \cup C}$-measurable processes $\tilde{\lambda}_k$, $k \in B$, such that the $\mathcal{F}_t^{A \cup B \cup C}$-intensities of $\mathbf{N}_B^T$ are given by $\lambda_k^T(t) = \tilde{\lambda}_k(t)\, \mathbf{1}\{t \leqslant T\}$, $k \in B$.

The local independences in a stopped process must be interpreted as being valid as long as $t \leqslant T$.

3. Local independence graphs

We first give the definition of local independence graphs and then investigate what can be read from these graphs.

3.1. Definition of local independence graphs

An obvious way of representing the local independence structure of an MPP by a graph is to depict the marks as vertices and to use an arrow as a symbol for local dependence as in the following small example.

Example 1 (continued): skin disease—the local independence graph for the relation between menopause and skin disease is very simple: Fig. 1. Even with this simple example there is no way of expressing the local independence by using a classical graph based on conditional independence for the two times T_1, the 'time of occurrence of skin disease', and T_2, the 'time of occurrence of menopause', as these are simply dependent.

For general local independence structures, we shall have directed graphs that may have more than one directed edge between a pair of vertices, in the case of mutual local dependence, and that may have cycles. More formally, a *graph* is a pair $G=(V,E)$, where $V=\{1,\ldots,K\}$ is a finite set of vertices and E is a set of edges. The graph is said to be *directed* if $E\subset\{(j,k)|j,k\in V, j\neq k\}$. Later we shall also need the notion of an *undirected* graph where $E\subset\{\{j,k\}|j,k\in V, j\neq k\}$. Undirected edges $\{j,k\}$ are depicted by lines, j—k, and directed edges (j,k) by arrows, $j\rightarrow k$. If $(j,k)\in E$ and $(k,j)\in E$ this is shown by stacked arrows, $j\rightleftarrows k$.

The following property (4) is called the *pairwise dynamic Markov property*, where we say *dynamic* to emphasize the difference from graphs that are based on conditional independence.

Definition 4 (local independence graph). Let $\mathbf{N}_V=(N_1,\ldots,N_K)$ be a multivariate counting process that is associated with an MPP Y with mark space $\mathcal{E}=\{e_1,\ldots,e_K\}$. Let further $G=(V,E)$ be a directed graph, $V=\{1,\ldots,K\}$. Then, G is called a *local independence graph* of Y if, for all $j,k\in V$,

$$(j,k)\notin E\Rightarrow\{j\}\not\rightarrow\{k\}|V\backslash\{j,k\}. \qquad (4)$$

Example 2 (continued): home visits—the graph in Fig. 2(a) is for the whole process, whereas Fig. 2(b) shows the local independences for the stopped process (stopping when death occurs). There are no arrows into Home visits representing that the rate of visits is locally independent of Hospitalization given Health status as well as of Health status given Hospitalization (while the person is still alive), reflecting that the visits are determined externally. The latter local independence might be violated if the nurses, on their own account, increase the frequency of their visits when they notice that the person's health is deteriorating. The graph further represents that survival is locally independent of the visits given hospitalization and health history, and that the health process is also locally independent of the visits given hospitalization history (while the person is still alive, obviously). These absent edges could reflect the null hypothesis when investigating whether the visits affect survival in other ways than through changing the rate of hospitalization.

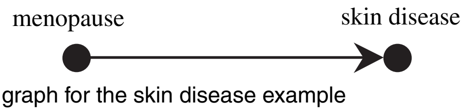

Fig. 1. Local independence graph for the skin disease example

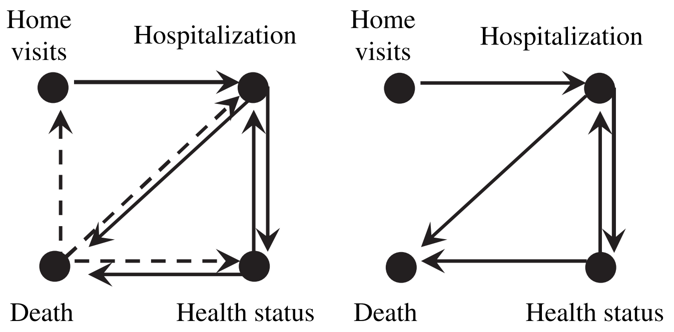

Fig. 2. Home visits example (a) for the whole process and (b) for the stopped process

3.2. Dynamic Markov properties

The local independence graphs as defined above allow (under mild assumptions) more properties to be read off, concerning the dependence structure, than just those given by expression (4). We may in particular query the graph with the aim of dimension reduction, i.e. with questions about which other processes can be ignored while investigating certain local independences. The *local dynamic Markov property* that is addressed in Section 3.2.1 tells us about the immediately relevant information when considering a single mark e_k and corresponding N_k. Further, the *global dynamic Markov property* in Section 3.2.2 gives graphical rules to identify when the *separating* set itself can be reduced, i.e. when in expression (4) we do not need to condition on *all* $V\backslash\{j,k\}$ but just on a true subset. For this we need the notion of δ-separation that is also introduced in Section 3.2.2.

Some more graph notation will be required. A *path* between two nodes is defined in the obvious way (the formal definition is given in Appendix A.2): we distinguish between *undirected paths* for undirected graphs, *directed paths*, preserving the direction of edges, for directed graphs and *trails* for connections in directed graphs that do not preserve the direction. For directed graphs we further require the following almost self-explanatory notation. If $a \rightarrow b$ then a is called a *parent* of b and b is a *child* of a (if $a \rightleftarrows b$ then a is both, a child and a parent of b); $\mathrm{pa}(A)$ denotes the set of all parents of nodes in $A \subset V$ without A itself, and $\mathrm{ch}(A)$ analogously the set of children of A. The set $\mathrm{cl}(A) = \mathrm{pa}(A) \cup A$ is called the *closure* of A. If there is a directed path from a to b then a is an *ancestor* of b and b is a descendant of a; the corresponding set notation is $\mathrm{an}(A)$ and $\mathrm{de}(A)$ (always excluding A itself). Consequently, $\mathrm{nd}(A) = V\backslash(\mathrm{de}(A) \cup A)$ are the *non-descendants* of A. If $\mathrm{pa}(A) = \emptyset$, then A is called *ancestral*. In general, $\mathrm{An}(A)$ is the *smallest ancestral set* containing A, given by $A \cup \mathrm{an}(A)$.

3.2.1. *Local dynamic Markov property*

Definition 5 (local dynamic Markov property). Let $G = (V, E)$ be a directed graph. For an MPP Y the property, for all $k \in V$,

$$V\backslash\mathrm{cl}(k) \not\rightarrow \{k\} | \mathrm{pa}(k) \tag{5}$$

is called the *local dynamic Markov property with respect to G*.

In other words, property (5) says that every $\mathcal{F}_t$-intensity λ_k is $\mathcal{F}_t^{\mathrm{cl}(k)}$ measurable, which clearly implies that for any ancestral set A the intensity λ_A is $\mathcal{F}_t^A$ measurable. This property could for instance be violated if two components in $\mathrm{pa}(k)$ were almost surely identical, which is, however, prevented by the orthogonality assumption 1. As shown in Appendix A.3, the exact condition for property (5) to follow from property (4) is that

$$\mathcal{F}_t^A \cap \mathcal{F}_t^B = \mathcal{F}_t^{A\cap B} \qquad \forall A, B \subset V, \quad \forall t \in \mathcal{T}, \tag{6}$$

where we define $\mathcal{F}^{\emptyset} = \{\emptyset, \Omega\}$. Property (6) is called 'conditional measurable separability' (Florens *et al.*, 1990) and formalizes the intuitive notion that the components of $\mathbf{N}$ are sufficiently 'different' to ensure that common events are necessarily due to common components.

Example 2 (continued): home visits—let us consider the question whether the four processes are sufficiently different to ensure property (6). If the health process is measured in a way such that it is determined by the number and duration of past hospitalizations, not taking any other information into account, this assumption might be violated. However, it makes sense and we shall assume for this example that the Health status reflects more aspects of a person's health than just past hospitalizations. Then it seems plausible that property (6) is satisfied as the other processes are clearly capturing different information anyway. Consequently we can use the local dynamic Markov property to read off that the visits process is locally independent of both, hospitalization and health status (while the person is still alive).

3.2.2. δ-separation and the global dynamic Markov property

In undirected graphs we say that subsets $A, B \subset V$ are separated by $C \subset V$ if any path between elements in A and elements in B is intersected by C. This is symbolized by $A \perp\!\!\!\perp_g B|C$. In classical graphical models every such separation induces conditional independence between A and B given C regardless of whether (A, B, C) is a partition of V or not. This can obviously lead to considerable dimension reduction if C is chosen minimally and the graph is sparse. To obtain a similar result for local independence graphs we require a suitable notion of separation called δ-separation, which is introduced below after some more graph notation.

The *moral* graph G^m is given by inserting undirected edges between any two vertices that have a common child (if they are not already joined) and then making all edges undirected (two directed edges between a pair of nodes are replaced by one undirected edge). This procedure of moralization will also be applied to an *induced subgraph* G_A, $A \subset V$, which is defined as (A, E_A) with E_A the subset of E containing only edges between pairs of nodes in A. Finally, for $B \subset V$, let G^B denote the graph that is obtained by deleting all directed edges of G starting in B.

Definition 6 (δ-separation). Let $G = (V, E)$ be a directed graph. Then, we say for pairwise disjoint subsets $A, B, C \subset V$ that C *δ-separates A from B in G* if $A \perp\!\!\!\perp_g B|C$ in the undirected graph $(G^B_{\mathrm{An}(A\cup B\cup C)})^m$ (the case of non-disjoint A, B and C is given in Appendix A.2).

Except for the fact that we delete edges starting in B, which makes δ-separation asymmetric, the definition parallels the definition for DAGs. This initial edge deletion can heuristically be explained by the fact that we want to separate the present of B from the past of A and hence we disregard the 'future' of B which is where the edges out of B point to; for the same reason only the ancestral set $\mathrm{An}(A \cup B \cup C)$ is considered. As for DAGs the insertion of moral edges is necessary whenever we condition on a common 'child' owing to a 'selection effect' by which two marginally independent variables (or processes) that affect a third variable (or process) become dependent when conditioning on this third variable. Further properties of δ-separation are discussed in Didelez (2006).

Definition 7 (global dynamic Markov property). Let $\mathbf{N}_V = (N_1, \ldots, N_K)$ be a multivariate counting process that is associated with an MPP Y and $G = (V, E)$ a directed graph. The property that, for all disjoint $A, B, C \subset V$,

$$C\ \delta\text{-separates } A \text{ from } B \text{ in } G \Rightarrow A \not\rightarrow B|C \qquad (7)$$

is called the *global dynamic Markov property with respect to G.*

The significance of the global Markov property is that it provides a way to verify whether a subset $C \subset V\backslash(A \cup B)$ is given such that $A \not\rightarrow B|C$, i.e. local independence is preserved even *when ignoring* information on the past of processes in $V\backslash(A \cup B \cup C)$. Of course, property (7) is only meaningful if it can be linked to the definition of local independence graphs that is addressed next.

Theorem 1 (equivalence of dynamic Markov properties). Let Y be a marked point process and $G=(V, E)$ a directed graph. Under the assumption of property (6) and further regularity conditions (see Appendix A.3), the pairwise, local and global dynamic Markov properties, i.e. expressions (4), (5) and (7), are equivalent.

The proof is given in Appendix A.3.

Example 2 (continued): home visits—the underlying health status of an elderly person may be difficult to measure accurately in practice. Let us therefore investigate the local independence structure when ignoring this underlying process altogether; in particular consider the question of whether from Fig. 2 we can infer that survival is still locally independent of Home visits given only the hospitalization but ignoring the health process. Graphically this means we must check whether the node Hospitalization alone separates Home visits from Death. As can be seen from the corresponding moral graph (for the stopped process) in Fig. 3 this is not so. Hence, even though the home visits are assumed to be determined externally in Fig. 2 and do not affect survival directly, ignoring the underlying health process may lead to a 'spurious' local dependence of survival on the home visits. The reason is that, for instance, a history of hospitalization with a preceding home visit predicts survival differently from a hospitalization without a preceding home visit—the former might mean that the health was especially bad and hence hospitalization was necessary, whereas the latter allows minor health problems that could have been treated by a nurse who was not available.

Intuitively it is clear that if the intensity of the visits depended on the underlying health status, i.e. if there was a directed edge from Health to Visits, we could talk of confounding. Hence it is rather surprising that even when the frequency of the visits is controlled externally we may find a spurious dependence. For the discrete time case Robins (1986, 1997) has demonstrated that nevertheless in a situation like Fig. 2 we can draw causal conclusions even when no information on the underlying health process is available. However, standard methods that just model the intensity for survival with time varying covariates for the times of previous home visits and hospitalizations will typically give misleading results due to the conditional association between Home visits and Death given Hospitalization.

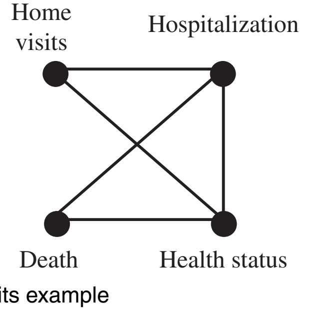

Fig. 3. Moral graph for the home visits example

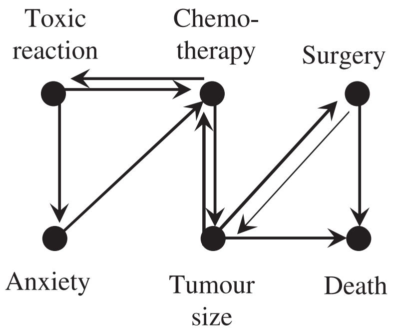

Fig. 4. Local independence graph for the chemotherapy example (a stopped process)

Example 3 (chemotherapy cycles)—to see a more complex example of a local independence graph consider a hypothetical study (which is inspired by real studies) where early stage breast cancer patients are observed over a period of several months during which they receive at least one but usually more cycles of chemotherapy. The size of the tumour is monitored through palpation. The doctors will consider removing the tumour by surgery if the size does not decrease and surgery is almost certain if it increases. Furthermore, the chemotherapy may be delayed or discontinued if the patient shows a toxic reaction following the treatment but also if the patient requests a delay, possibly due to an increased state of anxiety. Except for the size of the tumour all processes count one type of event that can occur once or more often. The size of the tumour is measured categorically depending on the number and palpable size of lesions and can be regarded as a multistate process. Fig. 4 shows a hypothetical local independence structure. For instance it assumes that survival locally depends on the size of the tumour and whether surgery has taken place, but once this information has been given none of the other processes are relevant for the intensity of death. Note that this particular assumption could plausibly be violated because toxic reactions and anxiety may reflect other health problems, but for simplicity we shall assume that all patients are 'healthy' except for the breast cancer so that this violation is excluded.

Fig. 5 shows the different moral graphs that are constructed from Fig. 4 to investigate δ-separations. Fig. 5(a) shows δ-separation from the node Death allowing us to read off, for instance, that Chemotherapy is not δ separated from Death by Tumour size alone, reflecting that chemotherapy predicts survival if surgery history is ignored. This is plausible because knowing that for example a decrease in the size of a tumour was preceded by a treatment cycle is informative for surgery, making it less likely than without preceding chemotherapy; and whether surgery has taken place, in turn, predicts the survival chances. As Anxiety is problematic to observe and measure we may further be interested in the question of when it can be ignored. We see from Fig. 5(a) that Death is locally independent of Anxiety given either the set {Surgery, Tumour size} or {Chemotherapy, Tumour size}, the latter implying that once we know the chemotherapy history in addition to the development of the size of the tumour then anxiety will not inform us any further about the intensity for death regardless of whether surgery and toxic reaction history is known or not. But note that even though Anxiety does not affect Tumour size directly the latter must be part of the separating set. Fig. 5(b) shows that Anxiety is δ separated from Surgery by any set that includes Tumour size and similarly for Fig. 5(c) that it is δ separated from Toxic reaction by any set that includes Chemotherapy—in these two cases δ-separation does not tell us more than the local dynamic Markov property (5). From Fig. 5(d) we see that Anxiety itself is locally independent of Chemotherapy and Tumour size given Toxic reaction and of Surgery given either Toxic reaction or the set {Chemotherapy, Tumour size}.

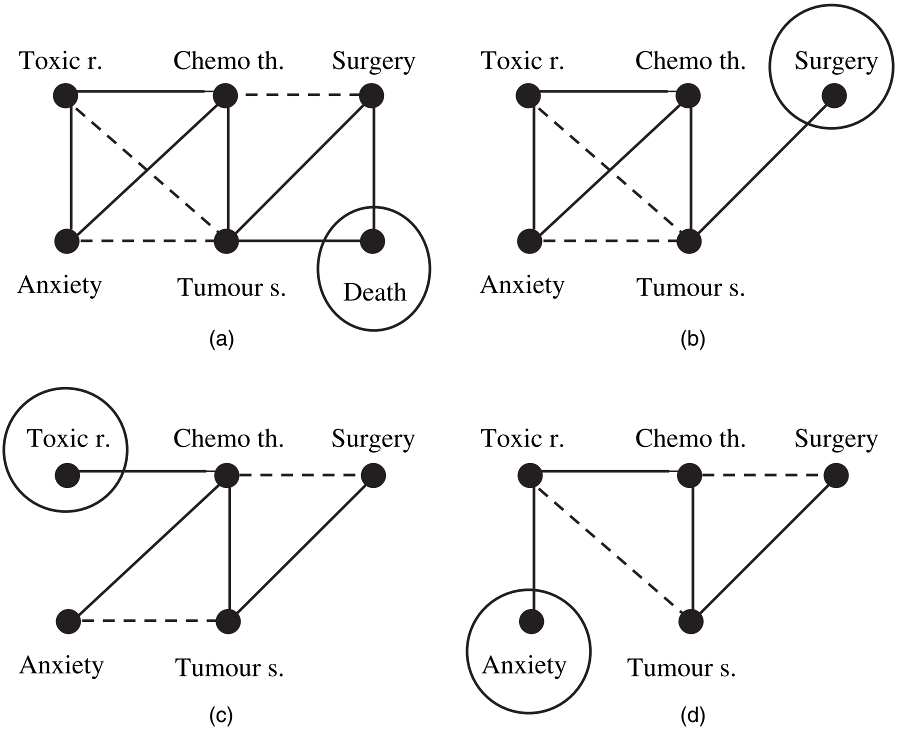

Fig. 5. Different moral graphs for the chemotherapy example: broken edges have been added as a result of a 'common child' in Fig. 4, and circles indicate that arrows out of these nodes have been deleted before moralizing

3.3. Likelihood factorization and implications

To discuss properties and implications for the likelihood for graphical MPPs we shall regard the data consisting of times and types of events $(t_1, e_1), (t_2, e_2), \ldots, (t_n, e_n)$ as a realization of the *history process* $H_t = \{(T_s, E_s) | T_s \leqslant t\}$. As for filtrations, H_{t^-} denotes the *strict pre-t history process*. Additionally, H_t^A, $A \subset \{1, \ldots, K\}$, which is defined as

$$H_t^A = \{(T_s, E_s) | T_s \leqslant t \text{ and } \exists k \in A : E_s = e_k, s = 1, 2 \ldots\},$$

denotes *the history process restricted to the marks in A*. Any set of marked points for which it holds that $t_s = t_u$, $s \neq u$, implies that $e_s = e_u$ can be a history, i.e. a realization of H_t. Note that (up to completion by null sets) the various filtrations can be regarded as being generated by the history processes, i.e. $\mathcal{F}_t^A = \sigma\{H_t^A\}$, $A \subset \{1, \ldots, K\}$.

Before deriving the likelihood for a given local independence graph, we recall it for the general case. On the basis of the mark-specific intensity processes $\lambda_k(t)$ the corresponding *crude* intensity process is given by $\lambda(t) = \Sigma_{k=1}^K \lambda_k(t)$. This is the intensity process of the cumulative counting process $\Sigma_k N_k$. The likelihood process $L(t|H_t)$ is then given as

$$L(t|H_t) = \prod_{T_s \leqslant t} \lambda_{E_s}(T_s) \exp\left\{ -\int_0^t \lambda(s)\, \mathrm{d}s \right\}. \tag{8}$$

To see how the likelihood is affected by G being a local independence graph of Y, we first rewrite equation (8) as

$$L(t|H_t) = \prod_{k=1}^{K} \prod_{T_s \leqslant t} \lambda_k(T_s)^{\mathbf{1}\{E_s=e_k\}} \exp\left\{ -\int_0^t \sum_{k=1}^{K} \lambda_k(s)\,\mathrm{d}s \right\}$$
$$= \prod_{k=1}^{K} \left[\prod_{T_{s(k)} \leqslant t} \lambda_k(T_{s(k)}) \exp\left\{ -\int_0^t \lambda_k(s)\,\mathrm{d}s \right\} \right],$$

where $T_{s(k)}$ with $E_{s(k)} = e_k$ are the occurrence times of mark e_k. The inner product of this equation can be regarded as the mark-specific likelihood and is denoted by $L_k(t|H_t)$. Now, by the definition of a local independence graph and the equivalence of the pairwise and local dynamic Markov properties under condition (6) we have that $\lambda_k(s)$ is $\mathcal{F}_t^{\mathrm{cl}(k)}$ measurable, where cl(k) is the closure of node k. Hence, it follows that

$$L_k(t|H_t) = L_k(t|H_t^{\mathrm{cl}(k)}), \tag{9}$$

i.e. the mark-specific likelihood L_k that is based on the whole past remains the same if the available information is restricted to how often and when those marks that are parents of e_k in the graph and e_k itself have occurred in the past, which is symbolized by $H_t^{\mathrm{cl}(k)}$.

It follows that under condition (6) the likelihood factorizes as

$$L(t|H_t) = \prod_{k \in V} L_k(t|H_t^{\mathrm{cl}(k)}), \tag{10}$$

which parallels the factorization for DAGs where the joint density is decomposed into the univariate conditional distributions given the parents. Here, we replace the parents by the closure because we also need to condition on the past of a component itself, which is not required in the static cases.

Example 2 (continued): home visits—from Fig. 2(b) we obtain the factorization

$$L(t|H_t) = L_{\mathrm{vi}}(t|H_t^{\{\mathrm{vi}\}})\; L_{\mathrm{ho}}(t|H_t^{\{\mathrm{vi,ho,hs}\}})\; L_{\mathrm{hs}}(t|H_t^{\{\mathrm{ho,hs}\}})\; L_{\mathrm{d}}(t|H_t^{\{\mathrm{d,ho,hs}\}})$$

for $t \leqslant$ time of death, where vi, ho, hs and d stand for visits, hospitalization, health status and death respectively.

Two consequences of this factorization regarding the relationship of local and conditional independence are given next.

Theorem 2 (conditional independences). For an MPP with local independence graph G and disjoint $A, B, C \subset V$, such that C separates A and B, i.e. $A \perp\!\!\!\perp_g B|C$, in $(G_{\mathrm{An}(A \cup B \cup C)})^m$, we have

$$\mathcal{F}_t^A \perp\!\!\!\perp \mathcal{F}_t^B | \mathcal{F}_t^C \qquad \forall t \in \mathcal{T}. \tag{11}$$

The proof is given in Appendix A.4. The graph separation that A, B and C must satisfy for condition (11) implies that for each $k \in C$ the $\mathcal{F}_t^{A \cup B \cup C}$-intensity λ_k is either $\mathcal{F}_t^{A \cup C}$ or $\mathcal{F}_t^{B \cup C}$ measurable; otherwise C could not separate A and B in the moral graph. Also, of course, we have that for each $k \in A$ and $k \in B$ the $\mathcal{F}_t^{A \cup B \cup C}$-intensity is respectively $\mathcal{F}_t^{A \cup C}$ and $\mathcal{F}_t^{B \cup C}$ measurable; otherwise there would be edges linking A and B in the graph and they could not be separated. A property similar to condition (11) has been noted by Schweder (1970), theorems 3 and 4, for Markov processes. With a similar argument we can reformulate expressions (2) and (3): for any $B \subset V$ we have

$$\mathbf{N}_B(t) \perp\!\!\!\perp \mathcal{F}_{t^-}^{V \backslash \mathrm{cl}(B)} | \mathcal{F}_{t^-}^{\mathrm{cl}(B)}, \tag{12}$$

i.e. the present of $\mathbf{N}_B$ is independent of the past of $\mathbf{N}_{V \backslash \mathrm{cl}(B)}$ given the past of $\mathbf{N}_{\mathrm{cl}(B)}$.

Example 3 (continued): chemotherapy cycles—let $A =$ Surgery, $B =$ Toxic reaction and $C =$ {Chemotherapy, Tumour size}. Then $(G_{\mathrm{An(A \cup B \cup C)}})^m$ is the same as Fig. 5(a), where the node

Death could be omitted as we are conditioning on the patient being alive anyway, and indeed A and B are separated by C. With condition (11) we can infer that at any time t (before death) the whole surgery history, i.e. whether and when surgery has taken place before t, is independent of whether and when toxic reactions have occurred given that we know the tumour size development up to t and when chemotherapy has been administered. As mentioned earlier we have for the nodes in C that the $\mathcal{F}_t^{A\cup B\cup C}$-intensity for Tumour size is $\mathcal{F}_t^{A\cup C}$ and the one for Chemotherapy is $\mathcal{F}_t^{B\cup C}$ measurable; the latter can be seen, by using δ-separation, by checking that Chemotherapy is locally independent from $A=$Surgery given {Chemotherapy, Toxic reaction} (the relevant moral graph happens to be the same as Fig. 5(b)).

3.4. Extensions

Local independence graphs can easily be extended to include time-fixed covariates, such as sex, age and socio-economic background of patients. The filtration at the start, $\mathcal{F}_0$, then must be enlarged to include the information on these variables. They can be represented by additional nodes in the graph with the restrictions that the subgraph on the non-dynamic nodes must be a DAG (or chain graph; see Gottard (2002)) and no directed edges are allowed to point from processes to time-fixed covariates. A process being locally independent of a time-fixed covariate means that the intensity does not depend on this particular covariate given all the other covariates and information on the past of all processes. δ-separation can still be applied to find further local independences.

The nodes in a local independence graph do not necessarily have to stand for only one mark (or the associated counting process); marks can be combined into one node as has been done in examples 2 and 3 with Health status and Tumour size. This might be of interest when there are logical dependences. For example if a particular illness is considered then the events 'falling ill' and 'recovering' from this illness are trivially locally dependent. If one node is used to represent a collection of marks corresponding to a multivariate subprocess of the whole multivariate counting process then an arrow into this node will mean that the intensity of at least one (but not necessarily all) of these marks depends on the origin of the arrow. However, some interesting information could be lost. For instance, if events such as 'giving birth to first child', 'giving birth to second child' etc. are considered, it might be relevant whether a woman is married or not when considering the event of giving birth to first child but it might not be relevant anymore when considering giving birth to second child.

In many data situations the mark space is not finite, e.g. when measuring the magnitude of electrical impulse, the amount of income in a new job or the dosage of a drug. One could then discretize the mark space, e.g. in 'finding a well-paid job' and 'finding a badly paid job'. However, it must be suspected that too many of these types of events will generate too many logical dependences that are of no interest and will make the graphs crammed.

As we have seen in some of the examples it is sometimes sensible to consider stopped processes to avoid having to represent logical and uninteresting dependences. More generally we might want to relax in definition 2 the requirement 'for all $t\in\mathcal{T}$' and instead consider suitably defined intervals that are based on stopping times. For example, it might be that the independence structure between the time of finishing education and starting the first job is very different from that before or after that. This deserves further investigation.

4. Discussion and conclusions

The main point of graphical models is that they allow certain algebraic manipulations to be replaced by graphical ones. In the case of local independence graphs, we can read properties of intensity processes with respect to different, in particular reduced, filtrations from the graph without the need to derive explicit formulae for these intensities, and similarly we can read off relationships between subprocesses such as properties (11) and (12). This facilitates reasoning about complex dependences, especially in the face of unobservable information, and simplifies calculations by reducing dimensionality.

Clearly, it is tempting to interpret local independence graphs causally. However, we regard causal inference as a topic of its own and it is not the aim of this paper to go into much detail in this respect, except for the following few comments. Local independence graphs represent (in)dependences in $E\{N_k(\mathrm{d}t)|\mathcal{F}_{t^-}\}$, where conditioning is on *having observed* $\mathcal{F}_{t^-}$ and, as we saw in example 2, it makes a difference to what dependences there are whether we condition on $\mathcal{F}_{t^-}$ or different subsets (or even extensions) thereof. Causal inference is about predicting $N_k(\mathrm{d}t)$ after *intervening* in $\mathcal{F}_{t^-}$, e.g. by modulating the times of the home visits to be once a week in the home visits example. It is well known that conditioning on observation is not the same as conditioning on intervention ('seeing' and 'doing' in Pearl (2000)). Hence, without further assumptions, the arrows in local independence graphs do not necessarily represent causal dependences—the intensity of an event being dependent on whether another event has been observed before does not imply causation in the same way as correlation does not imply causation. Such further assumptions could be that all 'relevant' events (or processes) have been taken into account, like originally proposed by Granger to justify the use of the term 'causality' for what is now known as Granger causality. For example if, in Fig. 2(b), we are satisfied that by including Health all relevant processes have been taken into account, then we could say that home visits are indirectly causal for Death. Obviously, in this particular example, there are many other relevant processes, like the occurrence of illnesses or death of the partner, that might be relevant. However, the literature on (non-dynamic) graphical models and causality has shown that causal inference is possible under weaker assumptions. Analogous results based on local independence graphs would require more prerequisites than we have given in this paper. Hence this is a topic for further research. For non-graphical approaches to causal reasoning in a continuous time event history setting see Eerola (1994), Lok (2001), Arjas and Parner (2004) and Fosen *et al.* (2004).

Another issue is the question of statistical inference for local independence graphs. This can be subdivided into

(a) inference when a graphical structure is given, e.g. from background knowledge, but we still want to quantify the strength of the dependences, and
(b) finding the graph from data if nothing about the local independence structure is known beforehand, which can be regarded as a particular kind of model selection or search task.

The former has partly been addressed in Section 3.3, where more specific results will depend on the actual modelling assumptions about the intensity processes which in turn will depend on the particular application. Estimation and testing within the class of Markov processes is tackled in Didelez (2007). More generally, local independence graphs can be combined with non-parametric, semiparametric or parametric methods but more research is required to investigate how the graphical representation of the local independence structure can simplify inference in particular settings. As to model search, Nodelman *et al.* (2003) provided a first attempt, which was restricted to Markov processes, at exploiting the graphical structure to find the graph itself when it is not postulated on the basis of background knowledge. Clearly, generalizations would be desirable.

Acknowledgements

I thank Iris Pigeot for her persistent encouragement and advice, and Niels Keiding for valuable comments and suggestions. Financial support by the German Research Foundation (Sonderforschungsbereich 386) is also gratefully acknowledged. The referees' comments have greatly helped to improve the paper.

Appendix A

The appendix is targeted at proving theorem 1 but to do so we first give some more results on the properties of local independence and δ-separation which will be used in that proof. These are explored along the lines of the graphoid axioms (Dawid, 1979, 1998; Pearl and Paz, 1987; Pearl, 1988) which have been generalized to the asymmetric case by Didelez (2006).

A.1. Properties of local independence

Proposition 1 (properties of local independence). The following properties hold for local independence:

(a) for all $A, B \subset V$: $A \not\to B|A$ (*left redundancy*);
(b) for all $A, B, C \subset V$ and $D \subset A$, if $A \not\to B|C$ then $D \not\to B|C$ (*left decomposition*);
(c) for all $A, B, C \subset V$ and $D \subset A$, if $A \not\to B|C$ then $A \not\to B|(C \cup D)$ (*left weak union*) and for all $A, B, C \subset V$ and $D \subset B$, if $A \not\to B|C$ then $A \not\to B|(C \cup D)$ (*right weak union*);
(d) for all $A, B, C, D \subset V$, if $A \not\to B|C$ and $D \not\to B|(A \cup C)$ then $(A \cup D) \not\to B|C$ (*left contraction*);
(e) for all $A, B, C \subset V$, if $A \not\to B|C$ and $A \not\to C|B$ then $A \not\to (B \cup C)|(B \cap C)$ (*right intersection*).

Proof.

(a) Left redundancy holds since obviously the $\mathcal{F}_t^{A\cup B}$-intensities of $\mathbf{N}_B$ are $\mathcal{F}_t^{A\cup B}$ measurable, i.e., if the past of $\mathbf{N}_A$ is known, then the past of $\mathbf{N}_A$ is of course irrelevant.
(b) Left decomposition holds since the $\mathcal{F}_t^{A\cup B\cup C}$-intensities $\lambda_k(t)$, $k \in B$, are $\mathcal{F}_t^{B\cup C}$ measurable by assumption so the same must hold for the $\mathcal{F}_t^{B\cup C\cup D}$-intensities $\lambda_k(t)$, $k \in B$, for $D \subset A$.
(c) Left and right weak union also trivially hold since adding information on the past of components that are already uninformative (left) or included (right) does not change the intensity.
(d) Left contraction holds since we have that the $\mathcal{F}_t^{A\cup B\cup C\cup D}$-intensities λ_k, $k \in B$, are by assumption $\mathcal{F}_t^{A\cup B\cup C}$ measurable and these are again by assumption $\mathcal{F}_t^{B\cup C}$ measurable.
(e) The property of right intersection can be checked by noting that in the definition of local independence the filtration with respect to which the intensity process should be measurable is always generated at least by the process itself. □

Note that left redundancy, left decomposition and left contraction imply that

$$A \not\to B|C \Leftrightarrow A\backslash C \not\to B|C. \tag{13}$$

It is also always true that $A \not\to B|C \Rightarrow A \not\to B\backslash C|C$, but we do not have equivalence here.

The following property will be important for the equivalence of pairwise, local and global dynamic Markov properties (just like in the well-known case of undirected conditional independence graphs; see Lauritzen (1996)).

Proposition 2 (left intersection for local independence). Under the assumption of property (6) local independence satisfies the following property which is called *left intersection*: for all $A, B, C \subset V$,

$$\text{if } A \not\to B|C \text{ and } C \not\to B|A \text{ then } (A \cup C) \not\to B|(A \cap C).$$

Proof. Left intersection assumes that the $\mathcal{F}_t^{A\cup B\cup C}$-intensities $\lambda_k(t)$, $k \in B$, are $\mathcal{F}_t^{B\cup C}$ as well as $\mathcal{F}_t^{A\cup B}$ measurable. With condition (6) we obtain that they are $\mathcal{F}_t^{B\cup(A\cap C)}$ measurable, which yields the desired result. □

The following can be regarded as an alternative version of the above property of left intersection. With expression (13), left decomposition and left intersection we have that, for disjoint $A, B, C, D \subset V$,

$$A \not\rightarrow B|(C \cup D) \text{ and } C \not\rightarrow B|(A \cup D) \Rightarrow (A \cup C) \not\rightarrow B|D. \tag{14}$$

This follows from corollary 4.3 of Didelez (2006). □

The last property that we consider, the 'right' counterpart of left decomposition that was given above, makes a statement about the irrelevance of a process $\mathbf{N}_A$ after discarding part of the possibly *relevant* information $\mathbf{N}_{B \setminus D}$. If the irrelevance of $\mathbf{N}_A$ is due to knowing the past of $\mathbf{N}_{B \setminus D}$ then it will not necessarily be irrelevant anymore if the latter is discarded.

Proposition 3 (conditions for right decomposition of local independence). Consider a marked point process and assume that the cumulative counting process ΣN_k is non-explosive and that intensities exist. Let $A, B, C \subset V$, $D \subset B$, with $(B \cap A) \setminus (C \cup D) = \emptyset$. The property, which is called *right decomposition*,

$$A \not\rightarrow B|C \Rightarrow A \not\rightarrow D|C$$

holds under the conditions that

$$B \not\rightarrow A \setminus (C \cup D)|(C \cup D)$$

and

$$A \not\rightarrow \{k\}|C \cup B \text{ or } B \not\rightarrow \{k\}|(C \cup D \cup A)$$

for all $k \in C \setminus D$.

Proof. In this proof we proceed somewhat informally for simplicity. The formal proof is based on the results of Arjas *et al.* (1992) and is given in Didelez (2000), page 72.

Redefine $A^* = A \setminus (C \cup D)$, $B^* = B \setminus D$ and $C^* = C \setminus D$. Then $A^* \cap B^* = \emptyset$ and, with conditions (12) and (11), it can be shown that the assumptions of the present proposition imply that

$$N_D(t) \perp\!\!\!\perp \mathcal{F}_{t^-}^{A^*} | \mathcal{F}_{t^-}^{B^* \cup C^* \cup D} \tag{15}$$

as well as

$$\mathcal{F}_t^{A^*} \perp\!\!\!\perp \mathcal{F}_t^{B^*} | \mathcal{F}_t^{C \cup D}. \tag{16}$$

We want to show that the $\mathcal{F}_t^{A \cup C \cup D}$-intensity $\tilde{\lambda}_D(t)$ of $\mathbf{N}_D(t)$ is $\mathcal{F}_t^{C \cup D}$ measurable. With the above expressions and interpretation (1) we have

$$\begin{aligned}
\tilde{\lambda}_D(t)\,\mathrm{d}t &= E\{N_D(\mathrm{d}t)|\mathcal{F}_{t^-}^{A \cup C \cup D}\} = E\{N_D(\mathrm{d}t)|\mathcal{F}_{t^-}^{A^* \cup C^* \cup D}\} \\
&= E[E\{N_D(\mathrm{d}t)|\mathcal{F}_{t^-}^{A^* \cup B^* \cup C^* \cup D}\}|\mathcal{F}_{t^-}^{A^* \cup C^* \cup D}] \\
&= E[E\{N_D(\mathrm{d}t)|\mathcal{F}_{t^-}^{B^* \cup C^* \cup D}\}|\mathcal{F}_{t^-}^{A^* \cup C^* \cup D}] \qquad \text{by using expression (15)} \\
&= E\{N_D(\mathrm{d}t)|\mathcal{F}_{t^-}^{C^* \cup D}\} \qquad \text{by using expression (16)} \\
&= E\{N_D(\mathrm{d}t)|\mathcal{F}_{t^-}^{C \cup D}\},
\end{aligned}$$

as desired.

A.2. Properties of δ-separation

For a general investigation of the properties of δ-separation we need to complete definition 6 by the case that A, B and C are not disjoint: we then define that C δ-separates A from B if $C \setminus B$ δ-separates $A \setminus (B \cup C)$ from B. We further define that the empty set is always δ separated from B. Additionally, we define that the empty set δ-separates A from B if A and B are unconnected in $(G^B_{\mathrm{An}(A \cup B)})^m$.

It can be shown (Didelez, 2006) that δ-separation satisfies the same properties as local independence given above in proposition 1 if we replace $A \not\rightarrow B|C$ by 'C δ-separates A from B' which we shall write as $A\textsc{ir}_\delta B|C$. In particular it satisfies left redundancy, left decomposition, left and right weak union, and left and right contraction as well as left and right intersection without requiring further asumptions. The property of right decomposition holds for δ-separation under conditions that are analogous to those in proposition 3. In particular we have the following special case of right decomposition:

$$A \text{IR}_\delta B|C, D \subset B \Rightarrow A \text{IR}_\delta D|(C \cup B)\backslash D \tag{17}$$

which is lemma 4.11 in Didelez (2006).

In addition, we want to show how δ-separation can be read from a local independence graph in a different but equivalent way to definition 6. We mention this, firstly, because it will be more familiar to readers who use d-separation for DAGs (Pearl, 1988; Verma and Pearl, 1990) and, secondly, because some parts of the proof of theorem 2 are easier to show by using this alternative way of checking δ-separation.

First, the different notions of paths and trails need to be made more stringent. Consider a directed or undirected graph $G=(V,E)$. An ordered $(n+1)$-tuple $(j_0,\ldots,j_n)$ of distinct vertices is called an *undirected path* from j_0 to j_n if $\{j_{i-1},j_i\}\in E$ and a *directed path* if $(j_{i-1},j_i)\in E$ for all $i=1,\ldots,n$. A (directed) path of length n with $j_0=j_n$ is called a *(directed) cycle*. A subgraph $\pi=(V',E')$ of G with $V'=\{j_0,\ldots,j_n\}$ and $E'=\{e_1,\ldots,e_n\}\subset E$ is called a *trail* between j_0 and j_n if $e_i=(j_{i-1},j_i)$, or $e_i=(j_i,j_{i-1})$ or $e_i=\{j_i,j_{i-1}\}$ for all $i=1,\ldots,n$. Further, for a directed graph we say that a trail between j and k is *blocked by* C if it contains a vertex γ such that either

(a) directed edges of the trail do not meet head to head at γ and $\gamma\in C$ or
(b) directed edges of the trail meet head to head at γ and γ as well as all its descendants are not elements of C.

Otherwise the trail is called *active*.

Proposition 4 (trail condition for δ-separation). Let $G=(V,E)$ be a directed graph and A, B and C pairwise disjoint subsets of V. Define that any *allowed trail from A to B* contains no edge of the form $(b,k), b\in B, k\in V\backslash B$. For disjoint subsets A, B and C of V, we have that C δ-separates A from B if and only if all allowed trails from A to B are blocked by C.

The proof is given in Didelez (2000), page 22; see also Didelez (2006).

A.3. Proof of theorem 1

It is easily checked that condition (7) implies condition (5), which implies condition (4). First, pa(k) always δ-separates $V\backslash(\text{pa}(k)\cup\{k\})$ from $\{k\}$ in G; hence condition (5) is just a special case of condition (7). Also, it is easy to see that the equivalence of the pairwise and local dynamic Markov properties immediately follows from left intersection assuming condition (6), left weak union and left decomposition. Thus, the following proof considers situations where A, B and C do not form a partition of V or $\text{pa}(B)\not\subset C$. The structure of the proof corresponds to the proof that was given by Lauritzen (1996), page 34, for the equivalence of the Markov properties in undirected conditional independence graphs. Owing to the asymmetry of local independence, however, this version is more involved.

Assume that condition (4) holds and that C δ-separates A from B in the local independence graph. We must show that $A\not\rightarrow B|C$, i.e. the $\mathcal{F}_t^{A\cup B\cup C}$-intensities $\lambda_k(t)$, $k\in B$, are $\mathcal{F}_t^{B\cup C}$ measurable. The proof is via backward induction on the number $|C|$ of vertices in the separating set. If $|C|=|V|-2$ then both A and B consist of only one element and condition (7) trivially holds. If $|C|<|V|-2$ then either A or B consists of more than one element.

Let us first consider the case that A,B,C is a partition of V and none of them is empty. If $|A|>1$ let $\alpha\in A$. Then, by left weak union and left decomposition of δ-separation we have that $C\cup(A\backslash\{\alpha\})$ δ-separates $\{\alpha\}$ from B, i.e.

$$\{\alpha\}\text{IR}_\delta B|C\cup(A\backslash\{\alpha\})$$

and $C\cup\{\alpha\}$ δ-separates $A\backslash\{\alpha\}$ from B in G, i.e.

$$A\backslash\{\alpha\}\text{IR}_\delta B|(C\cup\{\alpha\}).$$

Therefore, we have by the induction hypothesis that

$$\{\alpha\}\not\rightarrow B|C\cup(A\backslash\{\alpha\}) \text{ and } A\backslash\{\alpha\}\not\rightarrow B|(C\cup\{\alpha\}).$$

From this it follows with the modified version of left intersection as given in expression (14) (which can be applied because of the assumption that condition (6) holds) that $A\not\rightarrow B|C$ as desired.

If $|B|>1$ we can show by a similar reasoning, applying expression (17) to $\{\beta\}\in B$, that $A\not\rightarrow B|C$. Let us now consider the case that $A,B,C\subset V$ are disjoint but no partition of V. First, we assume that they are a partition of $\text{An}(A\cup B\cup C)$, i.e. that $A\cup B\cup C$ is an ancestral set. Let $\gamma\in V\backslash(A\cup B\cup C)$, i.e. γ is not an ancestor of $A \cup B \cup C$. Thus, every allowed trail (see proposition 4) from γ to B is blocked by $A \cup C$ since any such trail includes an edge (k, b) for some $b \in B$ where no edges meet head to head in k and $k \in A \cup C$. Therefore, we obtain

$$\{\gamma\}\text{IR}_\delta B|(A \cup C).$$

Application of left contraction, weak union and decomposition for δ-separation yields

$$A\text{IR}_\delta B|(C \cup \{\gamma\}).$$

It follows with the induction hypothesis that

$$A \not\to B|(C \cup \{\gamma\}) \text{ as well as } \{\gamma\} \not\to B|(A \cup C).$$

With left intersection as given by expression (14) and left decomposition for local independence we obtain the desired result.

Finally, let A, B and C be disjoint subsets of V and $A \cup B \cup C$ not necessarily an ancestral set. Choose $\gamma \in \text{an}(A \cup B \cup C)$ and define $\tilde{G}^B = G^B_{\text{An}(A \cup B \cup C)}$. Since $A \perp\!\!\!\perp_g B|C$ in $(\tilde{G}^B)^m$ we know from the properties of ordinary graph separation that

(a) either $\{\gamma\} \perp\!\!\!\perp_g B|(A \cup C)$ in $(\tilde{G}^B)^m$
(b) or $A \perp\!\!\!\perp_g \{\gamma\}|(B \cup C)$ in $(\tilde{G}^B)^m$.

In case (a) $\{\gamma\}\text{IR}_\delta B|(A \cup C)$ and it follows from left contraction that

$$(A \cup \{\gamma\})\text{IR}_\delta B|C.$$

Application of left weak union and left decomposition yields $A\text{IR}_\delta B|(C \cup \{\gamma\})$. With the induction hypothesis we therefore obtain

$$A \not\to B|(C \cup \{\gamma\}) \text{ and } \{\gamma\} \not\to B|(A \cup C).$$

Left intersection according to expression (14) and left decomposition for local independence yield $A \not\to B|C$.

Case (b) is the more complicated and the proof makes use now of right decomposition for local independence under the conditions that are given in proposition 3. First, we have from (b) that $A \perp\!\!\!\perp_g \{\gamma\}|B \cup C$ in $(G_{\text{An}(A \cup B \cup C)})^m$ since the additional edges starting in B can only yield additional paths between A and γ that must be intersected by B. Since deleting further edges out of γ does not create new paths, it holds that

$$A\text{IR}_\delta \{\gamma\}|(B \cup C).$$

With $A\text{IR}_\delta B|C$, application of right contraction for δ-separation yields $A\text{IR}_\delta (B \cup \{\gamma\})|C$. Now, we can apply property (17) to obtain $A\text{IR}_\delta B|(C \cup \{\gamma\})$ from where it follows with the induction hypothesis that

$$A \not\to B|(C \cup \{\gamma\}) \text{ and } A \not\to \{\gamma\}|(B \cup C).$$

With right intersection for local independence we obtain $A \not\to (B \cup \{\gamma\})|C$. In addition, $\{\gamma\} \not\to A|(B \cup C)$ by the same arguments as given above for $A \not\to \{\gamma\}|(B \cup C)$. To apply proposition 3 we still have to show that for all $k \in C$ either $A\text{IR}_\delta \{k\}|(C \cup B \cup \{\gamma\})$ or $\{\gamma\}\text{IR}_\delta \{k\}|(C \cup B \cup A)$, which by the induction hypothesis implies the corresponding local independences. To see this, assume that there is a vertex $k \in C$ for which neither holds. With the trail condition we then have that in $G_{\text{An}(A \cup B \cup C)}$ there is an allowed trail from A and γ to k such that every vertex where edges do not meet head to head are not in $(C \cup B \cup \{\gamma\}) \backslash \{k\}$ and $(C \cup B \cup A) \backslash \{k\}$ respectively, and every vertex where edges meet head to head or some of their descendants are in $(C \cup B \cup \{\gamma\}) \backslash \{k\}$ and $(C \cup B \cup A) \backslash \{k\}$ respectively. This would yield a path between A and γ which is not blocked by $C \cup B$ (note that k is a head-to-head node on this trail) in $G_{\text{An}(A \cup B \cup C)}$. This in turn contradicts the separation of A and γ by $B \cup C$ in $G^B_{\text{An}(A \cup B \cup C)}$ because the edges starting in B cannot contribute to this trail. Consequently we can apply right decomposition and obtain the desired result.

A.4. Proof of theorem 2

The factorization (10) implies that the marginal likelihood for the marked point process discarding events that are not in $\text{An}(A \cup B \cup C)$, i.e. $\{(T_s, E_s)|s = 1, 2, \ldots; E_s \in \mathcal{E}_{\text{An}(A \cup B \cup C)}\}$, is given by

$$L(t|H_t^{\text{An}(A \cup B \cup C)}) = \prod_{k \in \text{An}(A \cup B \cup C)} L_k(t|H_t^{\text{cl}(k)})$$

as none of the intensities of N_k, $k \in \mathrm{An}(A \cup B \cup C)$, depend on $V \setminus \mathrm{An}(A \cup B \cup C)$. Hence, the likelihood may be written as a product over factors that depend only on $\mathrm{cl}(k)$, $k \in \mathrm{An}(A \cup B \cup C)$. Let $\mathcal{C} = \{\mathrm{cl}(k) | k \in \mathrm{An}(A \cup B \cup C)\}$ be the set containing all such sets and let $g_c(t|\cdot)$, $c \in \mathcal{C}$, be these factors. Then, we have

$$L(t|H_t^{\mathrm{An}(A\cup B\cup C)}) = \prod_{c \in \mathcal{C}} g_c(t|H_t^c).$$

Further, by rearranging the factors the sets in $\mathcal{C}$ can be taken to be the 'cliques', i.e. the maximal fully connected sets of nodes, of the graph $(G_{\mathrm{An}(A\cup B\cup C)})^m$. This thus corresponds to the factorization property of undirected graphs which in turn implies the global Markov property for undirected graphs (Lauritzen (1996), page 35). This means that when we have a separation, like C separating A and B, in this graph $(G_{\mathrm{An}(A\cup B\cup C)})^m$ the corresponding conditional independence (11) holds, which completes the proof.

References

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

Aalen, O. O., Borgan, Ø., Keiding, N. and 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.

Allard, D., Brix, A. and Chadoeuf, J. (2001) Testing local independence between two point processes. *Biometrics*, **57**, 508–517.

Andersen, P. K., Borgan, Ø., Gill, R. D. and Keiding, N. (1993) *Statistical Models based on Counting Processes*. New York: Springer.

Arjas, E., Haara, P. and Norros, I. (1992) Filtering the histories of a partially observed marked point process. *Stoch. Processes Appl.*, **40**, 225–250.

Arjas, E.and Parner, J. (2004) Causal reasoning from longitudinal data. *Scand. J. Statist.*, **31**, 171–201.

Brémaud, P. (1981) *Point Processes and Queues: Martingale Dynamics*. New York: Springer.

Comte, F. and Renault, E. (1996) Long memory continuous time models. *J. Econometr.*, **73**, 101–149.

Cowell, R. G., Dawid, A. P., Lauritzen, S. L. and Spiegelhalter, D. J. (1999) *Probabilistic Networks and Expert Systems*. New York: Springer.

Cox, D. R. and Isham, V. S. (1980) *Point Processes*. London: Chapman and Hall.

Cox, D. R. and Wermuth, N. (1996) *Multivariate Dependencies—Models, Analysis and Interpretation*. London: Chapman and Hall.

Dahlhaus, R. (2000) Graphical interaction models for multivariate time series. *Metrika*, **51**, 157–172.

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

Dawid, A. P. (1998) Conditional independence. In *Encyclopedia of Statistical Sciences*, update vol. 2 (eds S. Kotz, C. B. Read and D. L. Banks), pp. 146–155. New York: Wiley-Interscience.

Dean, T. and Kanazawa, K. (1989) A model for reasoning about persistence and causation. *Computnl Intell.*, **5**, 142–150.

Didelez, V. (2000) *Graphical Models for Event History Data based on Local Independence*. Berlin: Logos.

Didelez, V. (2006) Asymmetric separation for local independence graphs. In *Proc. 22nd A. Conf. Uncertainty in Artificial Intelligence* (eds R. Dechter and T. Richardson), pp. 130–137. Arlington: AUAI.

Didelez, V. (2007) Graphical models for composable finite Markov processes. *Scand. J. Statist.*, **34**, 169–185.

Edwards, D. (2000) *Introduction to Graphical Modelling*, 2nd edn. New York: Springer.

Eerola, M. (1994) Probabilistic causality in longitudinal studies. *Lect. Notes Statist.*, **92**.

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.

Fleming, T. R. and Harrington, D. P. (1991) *Counting Processes and Survival Analysis*. New York: Wiley.

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

Florens, J. P., Mouchart, M. and Rolin, J. M. (1990) *Elements of Bayesian Statistics*. New York: Dekker.

Fosen, J., Borgan, O., Weedon-Fekjoer, H. and Aalen, O. O. (2004) Path analysis for survival data with recurrent events. *Technical Report*. University of Oslo, Oslo.

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.

Jordan, M. (1999) *Learning in Graphical Models*. Cambridge: MIT Press.

Keiding, N. (1999) Event history analysis and inference from observational epidemiology. *Statist. Med.*, **18**, 2353–2363.

Koster, J. T. A. (1996) Markov properties of non-recursive causal graphs. *Ann. Statist.*, **24**, 2148–2177.

Lauritzen, S. L. (1996) *Graphical Models*. Oxford: Clarendon.

Lok, J. J. (2001) Statistical modelling of causal effects in time. *PhD Thesis*. Free University of Amsterdam, Amsterdam.
Nodelman, U., Shelton, C. R. and Koller, D. (2002) Continuous time Bayesian networks. In *Proc. 18th A. Conf. Uncertainty in Artificial Intelligence*, pp. 378–387. San Francisco: Morgan Kaufmann.
Nodelman, U., Shelton, C. R. and Koller, D. (2003) Learning continuous time Bayesian networks. In *Proc. 19th A. Conf. Uncertainty in Artificial Intelligence*, pp. 451–458. San Francisco: Morgan Kaufmann.
Pearl, J. (1988) *Probabilistic Reasoning in Intelligent Systems*. San Mateo: Morgan Kaufmann.
Pearl, J. (2000) *Causality: Models, Reasoning, and Inference*. Cambridge: Cambridge University Press.
Pearl, J. and 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 and L. Steel), pp. 357–363. Amsterdam: North-Holland.
Robins, J. (1986) A new approach to causal inference in mortality studies with sustained exposure period—applications to control of the healthy workers survivor effect. *Math. Modllng*, **7**, 1393–1512.
Robins, J. (1997) Causal inference from complex longitudinal data. *Lect. Notes Statist.*, **120**, 69–117.
Schweder, T. (1970) Composable Markov processes. *J. Appl. Probab.*, **7**, 400–410.
Spirtes, P. (1995) Directed cyclic graphical representations of feedback models. In *Proc. 11th A. Conf. Uncertainty in Artificial Intelligence*, pp. 491–498. San Francisco: Morgan Kaufmann.
Spirtes, P., Glymour, C. and Scheines, R. (2000) *Causation, Prediction, and Search*, 2nd edn. Cambridge: MIT Press.
Vass, M., Avlund, K., Hendriksen, C., Andersen, C. K. and Keiding, N. (2002) Preventive home visits to older people in Denmark: methodology of a randomized controlled study. *Agng Clin. Exptl Res.*, **14**, 509–515.
Vass, M., Avlund, K., Kvist, K., Hendriksen, C., Andersen, C. K. and Keiding, N. (2004) Structured home visits to older people: are they only of benefit for women?: a randomised controlled trial. *Scand. J. Prim. Hlth Care*, **22**, 106–111.
Verma, T. and Pearl, J. (1990) Causal networks: semantics and expressiveness. In *Uncertainty and Artificial Intelligence* (eds R. D. Shachter, T. S. Levitt, L. N. Kanal and J. F. Lemmer), pp. 69–76. Amsterdam: North-Holland.
Whittaker, J. (1990) *Graphical Models in Applied Multivariate Statistics*. Chichester: Wiley.