Graphical Models for Marked Point Processes based on Local Independence

VANESSA DIDELEZ

University College London

Abstract

A new class of graphical models capturing the dependence structure of events that occur in time is proposed. The graphs represent so–called local independencies, 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 that the corresponding local independence graphs differ considerably from classical graphical models. Hence a new notion of graph separation, 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 dependencies as well as computational simplifications are discussed.

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, getting married, starting a family, durations of unemployment, illness etc. 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, change of medication etc. (cf. 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), hierarchical Bayesian modelling (cf. BUGS project `http://www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml`), causal reasoning (Pearl, 2000; Spirtes et al., 2000), as well as sociological, bio–medical, and econometric applications. For overviews and many different applications see for instance the monographs by Whittaker (1990), Cox and Wermuth (1996), Edwards (2000).

While the 'classical' graphical models are concerned with representing conditional independence structures among random variables, variations have been proposed to deal with feedback systems (Sprites, 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 getting on its way. 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) uses graphs to represent Granger–causality, which is a dynamic concept of dependence. For continuous time one approach, called Dynamic Bayesian Networks, is to discretise time and provide directed acyclic graphs that encode the independence structure for the transitions from t to $t+1$ (Dean and Kanazawa, 1989). The approach proposed by Nodelman et al. (2002, 2003) comes closest to the kind of graphs that we will consider. In their Continuous Time Bayesian Networks they represent a multistate Markov process with nodes corresponding to subprocesses and edges corresponding to dependencies 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 e.g. in Aalen et al. (1980). A generalisation to processes with a Doob–Meyer decomposition can be found in Aalen (1987) who focusses 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' used by Allard et al. (2001) is a different one.

We first set out the necessary notations 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 generalisation to a version that allows to condition on the past of other processes and hence describes dynamic dependencies for multivariate processes.

2

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 independencies that are preserved after marginalising over some of the processes. In section 3.3, it is shown how the likelihood of a process with given local independence graph factorises and implications are discussed. The potential of local independence graphs is discussed in section 4 and proofs are given in the appendix.

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 some 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 $N_k(t)$* associated with an MPP are then given by

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

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 dependencies 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 involved. Hence we denote the internal filtration of a marked point process by $\mathcal{F}_t = \sigma\{(T_s, E_s) \mid T_s \leq t, E_s \in \mathcal{E}\}$ which is equal to $\sigma\{(N_1(s), \ldots, N_K(s)) \mid s \leq t\}$, while for $A \subset \{1, \ldots, K\}$ we define the filtrations of a subprocess as $\mathcal{F}_t^A = \sigma\{\mathbf{N}_A(s) \mid s \leq t\}$; in particular $\mathcal{F}_t^k$ is the internal filtration of an individual counting process N_k.

Under rather general assumptions (cf. Fleming and Harrington, 1991, p. 61), a Doob–Meyer decomposition of $N_k(t)$ into a compensator and a martingale exists. Both these processes depend on the considered filtration which is here taken to be the internal filtration

of the *whole* MPP Y. We will 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. $\Lambda_k(t) = \int_0^t \lambda_k(s)ds$. Heuristically we have (Andersen et al., 1993, p. 52)

$$\lambda_k(t)dt = E(N_k(dt) \mid \mathcal{F}_{t-}). \tag{1}$$

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, $\lambda_k(t)dt$ is our best prediction of the immediately following behaviour of N_k. Note that for the theory developed in this paper the setting can slightly be generalised, not requiring absolute continuity of compensators (cf. 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 based on the past of the *whole* MPP, and the $\mathcal{F}_t^A$–intensities based on the past of the subprocess on marks in A. The latter can be computed using the innovation theorem (Brémaud, 1981, pp. 83), and a way of doing so, 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 2.1 *No jumps at the same time*
The $\mathcal{F}_t$–martingales $N_k - \int \lambda_k(s)ds$ 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 2.1 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 dependencies 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 multi–state processes can be represented as marked point process 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 (cf. Aalen, 1987).

Definition 2.2 *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 w.r.t $\mathcal{F}_t^1$ for all $t \in \mathcal{T}$. Otherwise we speak of local dependence.

The process N_1 being locally independent of N_2 is symbolised by $N_2 \nrightarrow N_1$. Interchangeably we will sometimes say that e_1 is locally independent of e_2, or $e_2 \nrightarrow e_1$.

The essence of the above 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 to the full one $\mathcal{F}_t$. This implies that we do not lose any essential information by ignoring how often and when event e_2 has occurred before t. One could say that if $N_2 \nrightarrow N_1$ then the presence of N_1 is conditionally independent of the past of N_2 given the past of N_1, or heuristically[1]

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

where $A \perp\!\!\!\perp B \mid C$ means 'A is conditionally independent of B given C' (cf. Dawid, 1979). For general processes, this is a stronger property than local independence but it holds for marked point processes with Assumption 2.1 as their distributions are determined by the intensities. Note that (2) does not imply that for $u > 0$: $N_1(t+u) \perp\!\!\!\perp \mathcal{F}_{t-}^2 \mid \mathcal{F}_{t-}^1$ — hence the name *local* independence. Also, $N_1(t) \perp\!\!\!\perp N_2(t)$ will only hold if the two processes are mutually locally independent of each other. Without Assumption 2.1 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) model two events in the life of an individual women: occurrence of a particular skin disease and onset of menopause. Their analysis reveals that the intensity for developing this skin disease is greater once the menopause has started than before. In contrast, and as one 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

---

[1]Note that (2) is an informal way of saying that $N_1(dt)$ 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 (3), (12) or (15), should be interpreted correspondingly.

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 events. This requires conditioning on the past of other processes as follows.

Definition 2.3 *Local independence (multivariate).*
Let $\mathbf{N} = (N_1, \ldots, N_K)$ be a multivariate counting processes associated with an MPP. Let further A, B, 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 $\mathbf{N}_A \nrightarrow \mathbf{N}_B \mid \mathbf{N}_C$ or briefly $A \nrightarrow B \mid C$. Otherwise, $\mathbf{N}_B$ is *locally dependent* on $\mathbf{N}_A$ given $\mathbf{N}_C$, i.e. $A \to B \mid C$. 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. In analogy to (2) the above definition of multivariate local independence is with Assumption 2.1 equivalent to

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

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 programmes exist to assist the elderly through regular home visits by a nurse. This is meant to reduce unnecessary hospitalisations 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 hospitalisation are monitored. In addition, it is plausible that the underlying health status of the elderly person may also affect the rate of hospitalisation 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 hospitalisation directly, i.e. whether survival is locally independent of the visits process given the hospitalisation 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 will 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, p. 122).

In order to see the relation with local independence, we briefly review Granger non–causality (Granger, 1969). Let $X_V = \{X_V(t) \mid t \in \mathbf{Z}\}$ 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 $\overline{X}_A(t) = \{X_A(s) \mid s \leq t\}$. Then, for disjoint subsets $A, B \subset V$ we say that X_A is strongly Granger-non causal for X_B if

$$X_B(t) \perp\!\!\!\perp \overline{X}_A(t-1) \mid \overline{X}_{V \setminus A}(t-1),$$

for all $t \in Z$. The interpretation is similar as 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 one of all other components $C = V \setminus (A \cup B)$, in analogy to (3). Also note that the above does not imply that $X_B(t+u) \perp\!\!\!\perp \overline{X}_A(t-1) \mid \overline{X}_{V \setminus A}(t-1)$ for $u > 0$, again analogous to local independence. Eichler (1999, 2000) investigates a graphical representation and rules to determine when the condition $\overline{X}_{V \setminus A}(t-1)$ can be reduced to proper subsets $\overline{X}_C(t-1)$, $C \subset V \setminus A$.

Finally, let us indicate how the definition of local independence can be generalised 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 zero once death has occurred. However, the dependence is 'trivial' and not of much interest. Let T be a $\mathcal{F}_t$–stopping time and let $\mathbf{N}^T = (N_1^T, \ldots, N_K^T)$ be the multivariate counting process stopped at T. Then the intensities of N_k^T are given by λ_k^T, $k \in V$, and local independence can be formalised as follows.

Definition 2.4 *Local independence for stopped processes.*
Let $\mathbf{N}^T = (N_1^T, \ldots, N_K^T)$ be a multivariate counting processes associated to an MPP and stopped at time T. Then we say that $A \not\to B \mid C$ if there exist $\mathcal{F}_t^{B \cup C}$–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 $\lambda_k^T(t) = \tilde{\lambda}_k(t)\mathbf{1}\{t \leq T\}$, $k \in B$.

The local independencies in a stopped process have to be interpreted as being valid as long as $t \leq T$.

3 Local independence graphs

We first give the definition of local independence graphs and then investigate what can be read off 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 symbol for local dependence as in the following small example.

Example 1 ctd. *Skin disease.*
The local independence graph for the relation between menopause and skin disease is very simple, cf. Figure 1. Notice that even with this simple example there is no way of expressing the local independence using a classical graph based on conditional independence for the two times $T_1 =$ 'time of occurrence of skin disease' and $T_2 =$ 'time of occurrence of menopause' as these are simply dependent.
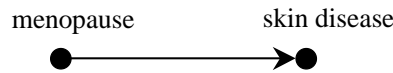

Figure 1: Local independence graph for skin disease example.

For general local independence structures, we will have directed graphs that may have more than one directed edge between a pair of vertices, in 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 edges. The graph is said to be *directed* if $E \subset \{(j, k) | j, k \in V, j \neq k\}$. Later we will 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, $j \longrightarrow k$. If $(j, k) \in E$ and $(k, j) \in E$ this is shown by stacked arrows, $j \overrightarrow{\longleftarrow} k$.

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

Definition 3.1 *Local independence graph*
Let $\mathbf{N}_V = (N_1, \ldots, N_K)$ be a multivariate counting process associated to 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

$$\text{for all } j, k \in V: \quad (j, k) \notin E \quad \Rightarrow \quad \{j\} \nrightarrow \{k\} | V \backslash \{j, k\}. \tag{4}$$
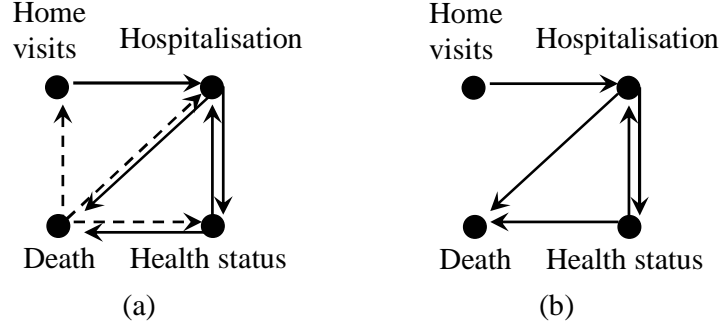

Figure 2: Home visits example; (a) for the whole process; (b) for the stopped process.

Example 2 ctd. *Home visits*

The graph in Figure 2(a) is for the whole process, while (b) shows the local independencies 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 'Hospitalisation' given 'Health status' as well as of 'Health status' given 'Hospitalisation' (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 hospitalisation and health history, and that the health process is also locally independent of the visits given hospitalisation 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 hospitalisation.

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 (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 independencies. The *local dynamic Markov property* 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 (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 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 notations. If $a \longrightarrow b$ then a is called a *parent* of b and b is a *child* of a (if $a \mathrel{\substack{\longrightarrow \\ \longleftarrow}} b$ 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 3.2 *Local dynamic Markov property*
Let $G = (V, E)$ be a directed graph. For an MPP Y the property

$$\text{for all } k \in V: \quad V\backslash\mathrm{cl}(k) \not\to \{k\} \mid \mathrm{pa}(k), \tag{5}$$

is called the *local dynamic Markov property w.r.t. 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 a.s. identical which is however prevented by the orthogonality assumption 2.1. As shown in Appendix A.3, the exact condition for the property (5) to follow from (4) is that

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

where we define $\mathcal{F}^\emptyset = \{\emptyset, \Omega\}$. Property (6) is called 'conditional measurable separability' (Florens et al., 1990) and formalises the intuitive notion that the components of $\mathbf{N}$ are

'different' enough to ensure that common events are necessarily due to common components.

Example 2 ctd. *Home visits*

Let us consider the question whether the four processes are sufficiently different to ensure (6). If the health process is measured in a way such that it is determined by the number and duration of past hospitalisation, not taking any other information into account, this assumption might be violated. However, it makes sense and we will assume for this example that the 'Health status' reflects more aspects of a person's health than just past hospitalisations. Then it seems plausible that (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, hospitalisation 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 symbolised 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, 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 moralisation will also be applied to an *induced subgraph G_A*, $A \subset V$, 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 obtained by deleting all directed edges of G starting in B.

Definition 3.3 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, C is given in Appendix A.2).

Note that except for the fact that we delete edges starting in B, which makes δ–separation asymmetric, the definition parallels the one 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' due 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 3.4 *Global dynamic Markov property*
Let $\mathbf{N}_V = (N_1, \dots, N_K)$ be a multivariate counting process associated to an MPP Y and $G = (V, E)$ a directed graph. The property that

$$\text{for all disjoint } A, B, C \subset V: \quad C \;\delta\text{–separates } A \text{ from } B \text{ in } G \quad \Rightarrow \quad A \not\rightarrow B \mid C. \quad (7)$$

is called the *global dynamic Markov property w.r.t. 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, (7) is only meaningful if it can be linked to the definition of local independence graphs addressed next.

Theorem 3.5 *Equivalence of dynamic Markov properties*
Let Y be a marked point process and $G = (V, E)$ a directed graph. Under the assumption of (6) and further regularity conditions (cf. Appendix A.3), the pairwise, local and global dynamic Markov properties, i.e. (4), (5) and (7) are equivalent.

The proof is given in Appendix A.3.

Example 2 ctd. *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 Figure 2 we can infer that survival is still locally independent of 'Home visits' given only the hospitalisation but ignoring the health process. Graphically this means we have to check whether the node 'Hospitalisation' alone separates 'Home visits' from 'Death'. As can be seen from the corresponding moral graph (for the stopped process) in Figure 3 this is not

12

the case. Hence, even though the home visits are assumed to be determined externally in Figure 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 hospitalisation with preceding home visit predicts survival differently from a hospitalisation without preceding home visit — the former might mean that the health was especially bad and hence hospitalisation was necessary, while the latter allows minor health problems that could have been treated by a nurse who was not available.
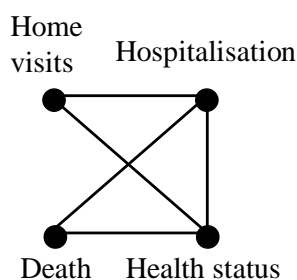

Figure 3: Moral graph for home visits example.

Notice that 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 Figure 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 hospitalisations will typically give misleading results due to the conditional association between 'Home visits' and 'Death' given 'Hospitalisation'.

Example 3. *Chemo therapy 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 chemo therapy. The tumour size 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 chemo therapy may be delayed or discontinued if the patient shows a toxic reaction following the treatment but also if the
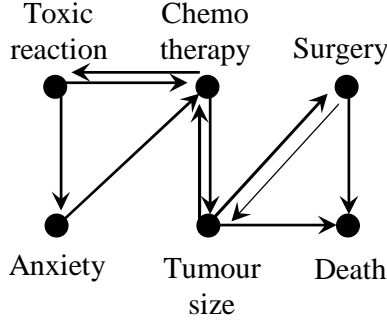
Figure 4: Local independence graph for chemo therapy example (stopped process).

patient requests a delay, possibly due to an increased state of anxiety. Except for tumour size all processes count one type of event that can occur once or more often. Tumour size is measured categorically depending on the number and palpable size of lesions and can be regarded as a multistate process. Figure 4 shows a hypothetical local independence structure. For instance it assumes that survival locally depends on the tumour size and whether surgery has taken place, but once this information is 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 will assume that all patients are 'healthy' except for the breast cancer so that this violation is excluded.

Figure 5 shows the different moral graphs constructed from Figure 4 to investigate δ–separations. Graph (a) shows δ–separation from the node 'Death' allowing us to read off, for instance, that 'Chemo therapy' is not δ–separated from 'Death' by 'Tumour size' alone reflecting that chemo therapy predicts survival if surgery history is ignored. This is plausible because knowing that for example a decrease in tumor size was preceded by a treatment cycle is informative for surgery, making it less likely than without preceding chemo therapy; 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 the graph (a) that 'Death' is locally independent of 'Anxiety' given either the set {'Surgery', 'Tumour size'} or {'Chemo therapy', 'Tumour size'}, the latter implying that once we know the chemo therapy history in addition to the development of the tumour size 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 has to be part of the separating set. Figure 5(b) shows that 'Anxiety' is δ–separated from 'Surgery' by any set that includes 'Tumour size' and similarly for
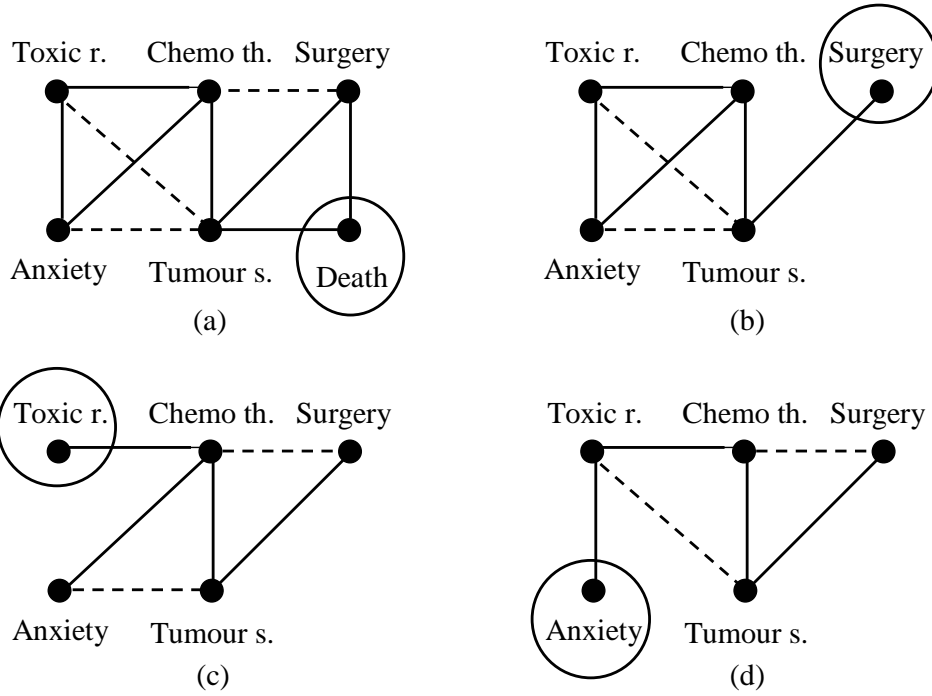
Figure 5: Different moral graphs for chemo therapy example. Dotted edges have been added due to a 'common child' in Figure 4 and circles indicate that arrows out of these nodes have been deleted before moralising.

(c) that it is δ–separated from 'Toxic reaction' by any set that includes 'Chemo therapy' — in these two cases δ–separation does not tell us more than the local dynamic Markov property (5). From graph (d) we see that 'Anxiety' itself is locally independent of 'Chemo therapy' and 'Tumour size' given 'Toxic reaction' and of 'Surgery' given either 'Toxic reaction' or the set {'Chemo therapy', 'Tumour size'}.

3.3 Likelihood factorisation and implications

In order to discuss properties and implications for the likelihood for graphical MPPs we will 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 realisation of the *history process* $H_t = \{(T_s, E_s) | T_s \leq t\}$. As for filtrations, H_{t-} denotes the *strict pre–t history process*. Additionally, H_t^A, $A \subset \{1, \ldots, K\}$, defined as

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

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 $e_s = e_u$ can be a history, i.e. a realization of H_t. Note that (up to completion by null sets) the different filtrations can be regarded as being generated by the history processes, i.e. $\mathcal{F}_t^A = \sigma\{H_t^A\}$, $A \subset \{1, \ldots, K\}$.

15

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

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

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

$$\begin{aligned}
L(t|H_t) &= \prod_{k=1}^{K} \prod_{T_s \leq t} \lambda_k(T_s)^{\mathbf{1}\{E_s = e_k\}} \cdot \exp\left(-\int_0^t \sum_{k=1}^{K} \lambda_k(s)\, ds\right) \\
&= \prod_{k=1}^{K} \left(\prod_{T_{s(k)} \leq t} \lambda_k(T_{s(k)}) \cdot \exp\left(-\int_0^t \lambda_k(s)\, ds\right)\right),
\end{aligned}$$

where $T_{s(k)}$ with $E_{s(k)} = e_k$ are the occurrence times of mark e_k. The inner product of the above 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 $\mathrm{cl}(k)$ 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 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, symbolised by $H_t^{\mathrm{cl}(k)}$.

It follows that under (6) the likelihood factorises as

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

which parallels the factorisation 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 case.

Example 2 ctd. *Home visits*
From Figure 2(b) we obtain the following factorisation

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

for $t \leq$ time of death, where vi, ho, hs, d stand for 'visits', 'hospitalisation', 'health status' and 'death' respectively.

Two consequences of the above factorisation regarding the relation of local and conditional independence are given next.

Theorem 3.6 *Conditional independencies*

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

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

The proof is given in appendix A.4. The graph separation that A, B, C have to satisfy for (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 $k \in A$ (resp. B) the $\mathcal{F}_t^{A \cup B \cup C}$–intensity is $\mathcal{F}_t^{A \cup C}$ (resp. $\mathcal{F}_t^{B \cup C}$)– measurable; otherwise there would be edges linking A and B in the graph and they could not be separated. A property similar to (11) has been noted by Schweder (1970, Theorems 3 and 4) for Markov processes. With a similar argument we can reformulate (2) and (3): for any $B \subset V$ we have

$$\mathbf{N}_B(t) \perp\!\!\!\perp \mathcal{F}_{t-}^{V \setminus \text{cl}(B)} \mid \mathcal{F}_{t-}^{\text{cl}(B)}, \tag{12}$$

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

Example 3 ctd. *Chemo therapy cycles*

Let $A =$'Surgery', $B =$'Toxic reaction' and $C = \{$'Chemo therapy', 'Tumour size'$\}$. Then $(G_{\text{An}(A \cup B \cup C)})^m$ is the same as Figure 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 (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 occured given we know the tumour size development up to t and when chemo therapy has been administered. Note that as mentioned above 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 'Chemo therapy' is $\mathcal{F}_t^{B \cup C}$–measurable; the latter can be seen, using δ–separation, by checking that 'Chemo therapy' is locally independent from $A =$'Surgery' given $\{$'Chemo therapy', 'Toxic reaction' $\}$ (the relevant moral graph happens to be the same as Figure 5(b)).

3.4 Extensions

Local independence graphs can easily be extended to include time–fixed covariates, such as sex, age and socioeconomic background of patients etc. The filtration at the start, $\mathcal{F}_0$, then has to 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, cf. 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 independencies.

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 dependencies. 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 1st child', 'giving birth to 2nd child' etc. are considered, it might be relevant whether a woman is married or not when considering the event of 'giving birth to 1st child' but it might not be relevant anymore when considering 'giving birth to 2nd 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 discretise 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 dependencies 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 in order to avoid having to represent logical and uninteresting dependencies. More generally one might want to relax in Definition 2.3 the requirement 'for all $t \in \mathcal{T}$' and instead consider suitably defined intervals based on stopping times. For example, it might be the case that the independence structure is very different between the time of finishing education and starting the first job than 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, off the graph without the need to derive explicit formulae for these intensities, and similarly we can read off relations among subprocesses such as properties (11) and (12). This facilitates reasoning about complex dependencies, 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, but for the following few comments. Local independence graphs represent (in)dependencies in $E(N_k(dt)|\mathcal{F}_{t-})$, where conditioning is on *having observed $\mathcal{F}_{t-}$* and, as we saw in Example 2, it makes a difference to what dependencies there are whether we condition on $\mathcal{F}_{t-}$ or different subsets (or even extensions) thereof. Causal inference is about predicting $N_k(dt)$ 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 dependencies — 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 in order to justify the use of the term 'causality' for what is now known as Granger-causality. E.g. if, in Figure 2(b), we are satisfied that by including 'Health' all relevant processes have been taken into account, then we could say that homevisits 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 confer 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 (i) inference when a graphical structure is given, e.g. from

background knowledge, but we still want to quantify the strength of the dependencies, and (ii) 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–, semi– 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) provide a first attempt, restricted to Markov processes, at exploiting the graphical structure to find the graph itself when it is not postulated based on background knowledge. Clearly, generalisations would be desirable.

A   Appendix

The appendix is targeted at proving Theorem 3.5, 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; Pearl and Paz, 1987; Pearl, 1988; Dawid, 1998) which have been generalised to the asymmetric case by Didelez (2006).

A.1   Properties of local independence

Proposition A.1 *Properties of local independence*
The following properties hold for local independence:

(i) *left redundancy:* for all $A, B \subset V$: $A \not\rightarrow B | A$,

(ii) *left decomposition:* for all $A, B, C \subset V$ and $D \subset A$: if $A \not\rightarrow B | C$ then $D \not\rightarrow B | C$,

(iii) *left weak union:* for all $A, B, C \subset V$ and $D \subset A$: if $A \not\rightarrow B | C$ then $A \not\rightarrow B | (C \cup D)$
and
right weak union: for all $A, B, C \subset V$ and $D \subset B$: if $A \not\rightarrow B | C$ then $A \not\rightarrow B | (C \cup D)$,

(iv) *left contraction:*
for all $A, B, C, D \subset V$: if $A \not\rightarrow B | C$ and $D \not\rightarrow B | (A \cup C)$ then $(A \cup D) \not\rightarrow B | C$,

(v) *right intersection:*
for all $A, B, C \subset V$: if $A \not\rightarrow B | C$ and $A \not\rightarrow C | B$ then $A \not\rightarrow (B \cup C) | (B \cap C)$.

Proof:

(i) 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.

(ii) 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 that the same must hold for the $\mathcal{F}_t^{B\cup C\cup D}$–intensities $\lambda_k(t)$, $k \in B$, for $D \subset A$.

(iii) 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.

(iv) 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.

(v) The property of right intersection can be checked by noting that in the definition of local independence the filtration w.r.t. which the intensity process should be measurable is always generated at least by the process itself. $\qquad\qquad\square$

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

$$A \not\rightarrow B \mid C \quad \Leftrightarrow \quad A\backslash C \not\rightarrow B \mid C. \tag{13}$$

It is also always true that $A \not\rightarrow B \mid C \Rightarrow A \not\rightarrow B\backslash C \mid C$, 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, cf. Lauritzen, 1996).

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

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

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 (6) we get 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 (13), left decomposition and left intersection we have that for disjoint

$A, B, C, D \subset V$

$$A \not\rightarrow B \mid (C \cup D) \text{ and } C \not\rightarrow B \mid (A \cup D) \Rightarrow (A \cup C) \not\rightarrow B \mid D \qquad (14)$$

This follows from Corollary 4.3 of Didelez (2006).

The last property that I want to consider, the 'right' counterpart of left decomposition 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 A.3 *Conditions for right decomposition of local independence*
Consider a marked point process and assume that the cumulative counting process $\sum N_k$ is non–explosive and that intensites exist. Let $A, B, C \subset V$, $D \subset B$, with $(B \cap A) \setminus (C \cup D) = \emptyset$. The following property, called *right decomposition,*

$$A \not\rightarrow B \mid C \;\Rightarrow\; A \not\rightarrow D \mid C$$

holds under the conditions that

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

and

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

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

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

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

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

as well as

$$\mathcal{F}_t^{A^*} \perp\!\!\!\perp \mathcal{F}_t^{B^*} \mid \mathcal{F}_t^{C \cup D}. \qquad (16)$$

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 and interpretation (1) we have

$$
\begin{aligned}
\tilde{\lambda}_D(t)dt &= E(N_D(dt) \mid \mathcal{F}_{t-}^{A \cup C \cup D}) = E(N_D(dt) \mid \mathcal{F}_{t-}^{A^* \cup C^* \cup D}) \\
&= E(E(N_D(dt) \mid \mathcal{F}_{t-}^{A^* \cup B^* \cup C^* \cup D}) \mid \mathcal{F}_{t-}^{A^* \cup C^* \cup D}) \\
&= E(E(N_D(dt) \mid \mathcal{F}_{t-}^{B^* \cup C^* \cup D}) \mid \mathcal{F}_{t-}^{A^* \cup C^* \cup D}) \qquad \text{using (15)} \\
&= E(N_D(dt) \mid \mathcal{F}_{t-}^{C^* \cup D}) \qquad\qquad \text{using (16)} \\
&= E(N_D(dt) \mid \mathcal{F}_{t-}^{C \cup D}),
\end{aligned}
$$

as desired.

A.2 Properties of δ–separation

For a general investigation of the properties of δ–separation we need to complete Definition 3.3 by the case that A, B and C are not disjoint: we then define that C δ–separates A from B if $C\backslash B$ δ–separates $A\backslash(B \cup C)$ 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 $(G_{\mathrm{An}(A \cup B)}^B)^m$.

It can be shown (Didelez, 2006) that δ–separation satisfies the same properties as local independence given above in Proposition A.1 if we replace $A \not\rightarrow B|C$ by "C δ–separates A from B" which we will write as $A \,\mathrm{IR}_\delta\, B|C$. In particular it satisfies left redundancy, left decomposition, let and right weak union, 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 analogous to those in Proposition A.3. In particular we have the following special case of right decomposition

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

which is Lemma 4.11 in Didelez (2006).

In addition, we want to show how δ–separation can be read off a local independence graph in a different but equivalent way to Definition 3.3. 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 3.5 are easier to show 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 (i) directed edges of the trail do not meet head–to–head at γ and $\gamma \in C$, or (ii) directed edges of the trail meet head–to–head at γ and γ as well as all its descendants are no elements of C. Otherwise the trail is called *active.*

Proposition A.4 *Trail condition for δ–separation*
Let $G = (V, E)$ be a directed graph and A, B, 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, 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, pp.22; cf. also Didelez, 2006).

A.3 Proof of Theorem 3.5

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

Assume that (4) holds and that C δ–separates A from B in the local independence graph. We have to show that $A \not\rightarrow B \mid C$, i.e. the $\mathcal{F}_t^{A \cup B \cup C}$–intenities $\lambda_k(t)$, $k \in B$, 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 (7) trivially holds. If $|C| < |V| - 2$ then either A or B consist 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.

$$\{\alpha\} \perp\!\!\!\perp_\delta B \mid C \cup (A \backslash \{\alpha\})$$

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

$$A \backslash \{\alpha\} \, \mathrm{IR}_\delta B \mid (C \cup \{\alpha\}).$$

Therefore, we have by the induction hypothesis that

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

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

If $|B| > 1$ we can show by a similar reasoning, applying (17) to $\{\beta\} \in B$, that $A \not\to B \mid C$.

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 $\mathrm{An}(A \cup B \cup C)$, 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 (cf. Proposition A.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 get

$$\{\gamma\} \, \mathrm{IR}_\delta B \mid (A \cup C).$$

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

$$A \, \mathrm{IR}_\delta B \mid (C \cup \{\gamma\}).$$

It follows with the induction hypothesis that

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

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

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

$$
\begin{aligned}
&(i) \quad \text{either } \{\gamma\} \perp\!\!\!\perp_g B \mid (A \cup C) \text{ in } (\tilde{G}^B)^m \\
&(ii) \quad \text{or } A \perp\!\!\!\perp_g \{\gamma\} \mid (B \cup C) \text{ in } (\tilde{G}^B)^m.
\end{aligned}
$$

(i) In the first case $\{\gamma\} \, \mathrm{IR}_\delta B \mid (A \cup C)$ and it follows from left contraction that

$$(A \cup \{\gamma\}) \, \mathrm{IR}_\delta B \mid C.$$

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

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

25

Left intersection according to (14) and left decomposition for local independence yields $A \not\rightarrow B \mid C$.

(ii) The second case is the most complicated and the proof makes now use of right decomposition for local independence under the conditions given in Proposition A.3. First, we have from (ii) that $A \perp\!\!\!\perp_g \{\gamma\} \mid B \cup C$ in $(G_{\mathrm{An}(A \cup B \cup C)})^m$ 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

$$A \, \mathrm{IR}_\delta \{\gamma\} \mid (B \cup C).$$

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

$$A \not\rightarrow B \mid (C \cup \{\gamma\}) \text{ and } A \not\rightarrow \{\gamma\} \mid (B \cup C).$$

With right intersection for local independence we get $A \not\rightarrow (B \cup \{\gamma\}) \mid C$. In addition, $\{\gamma\} \not\rightarrow A \mid (B \cup C)$ by the same arguments as given above for $A \not\rightarrow \{\gamma\} \mid (B \cup C)$. In order to apply Proposition A.3 we still have to show that for all $k \in C$ either $A \, \mathrm{IR}_\delta \{k\} \mid (C \cup B \cup \{\gamma\})$ or $\{\gamma\} \, \mathrm{IR}_\delta \{k\} \mid (C \cup B \cup A)$ which by the induction hypothesis implies the corresponding local independencies. To see this, assume that there exists a vertex $k \in C$ for which neither holds. With the trail condition we then have that in $G_{\mathrm{An}(A \cup B \cup C)}$ there exists an allowed trail from A and γ, respectively, 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 $(C \cup B \cup \{\gamma\}) \backslash \{k\}$ respective $(C \cup B \cup A) \backslash \{k\}$. 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 $G_{\mathrm{An}(A \cup B \cup C)}$. This in turn contradicts the separation of A and γ by $B \cup C$ in $G^B_{\mathrm{An}(A \cup B \cup C)}$ because the edges starting in B cannot contribute to this trail. Consequently we can apply right decomposition and get the desired result.

A.4   Proof of Theorem 3.6

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

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

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

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

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 $(G_{\mathrm{An}(A \cup B \cup C)})^m$. The above thus corresponds to the factorisation property of undirected graphs which in turn implies the global Markov property for undirected graphs (Lauritzen, 1996, p. 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. *Scandinavian Actuarial Journal*, 177 – 190.

Aalen, O. O., Ø. Borgan, N. Keiding, and J. Thormann (1980) Interaction between Life History Events: Nonparametric Analysis of Prospective and Retrospective Data in the Presence of Censoring. *Scandinavian Journal of Statistics 7*, 161 – 171.

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

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

Arjas, E., P. Haara, and I. Norros (1992) Filtering the Histories of a Partially Observed Marked Point Process. *Stochastic Processes and Applications 40*, 225 – 250.

Arjas, E., and J. Parner (2004) Causal Reasoning from Longitudinal Data. *Scandinavian Journal of Statistics 31*, 171 – 201.

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

Comte, F., E. Renault (1996) Long memory continuous time models. *Journal of Econometrics 73*, 101 – 149.

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

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

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

Dahlhaus, R. (2000) Graphical Interaction Models for Multivariate Time Series. *Metrika 51*, 157 – 172.

Dawid, A. P. (1979) Conditional Independence in Statistical Theory. *Journal of the Royal Statistical Society, Series B 41*, 1 – 31.

Dawid, A. P. (1998) Conditional Independence. In S. Kotz, C. B. Read, and D. L. Banks (Eds.), *Encyclopedia of Statistical Sciences, Update Volume 2*, pp. 146 – 155. Wiley–Interscience, New York.

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

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

Didelez, V. (2006) Asymmetric separation for local independence graphs. In R. Dechter and T. Richardson (Eds.), *Proceedings of the 22nd Annual Conference on Uncertainty and Artificial Intelligence*, pp. 130 – 137. UAI Press, Oregon.

Didelez, V. (2007) Graphical models for composable finite Markov processes. To appear in *Scandinavian Journal of Statistics*.

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

Eerola, M. (1994) *Probabilistic Causality in Longitudinal Studies*, Volume 92 of *Lecture Notes in Statistics.* Springer, Berlin.

Eichler, M. (1999) *Graphical Models in Time Series Analysis.* Ph. D. thesis, University of Heidelberg.

Eichler, M. (2000) Causality Graphs for Stationary Time Series. Technical report, University of Heidelberg.

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

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

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

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

Granger, C. W. J. (1969) Investigating Causal Relations by Econometric Models and Cross–spectral Methods. *Econometrica 37*, 424 – 438.

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

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

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

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

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

Lok, J. J (2001) *Statistical modelling of causal effects in time.* Ph. D. thesis, Free University of Amsterdam.

Nodelman, U., Shelton, C. R. and D. Koller (2002) Continuous Time Bayesian Networks. Proceedings of the 18th Annual Conference on Uncertainty in Artificial Intelligence, 378 – 387. Morgan Kaufman, San Francisco.

Nodelman, U., Shelton, C. R. and D. Koller (2003) Learning Continuous Time Bayesian Networks. Proceedings of the 19th Annual Conference on Uncertainty in Artificial Intelligence, 451 – 458. Morgan Kaufman, San Francisco.

Pearl, J. (1988) *Probabilistic Reasoning in Intelligent Systems.* Morgan Kaufman, San Mateo.

Pearl, J. (2000) *Causality. Models, Reasoning, and Inference.* Cambridge University Press, Cambridge.

Pearl, J. and A. Paz (1987) Graphoids: A Graph–Based Logic for Reasoning about Relevancy Relations. In B. Du Boulay, D. Hogg, and L. Steel (Eds.), *Advances in Artificial Intelligence – II*, pp. 357 – 363.

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. *Mathematical Modelling 7*, 1393 – 1512.

Robins J. (1997) Causal inference from complex longitudinal data. In M. Berkane, *Latent Variable Modeling and Applications to Causality.* Lecture Notes in Statistics 120. Springer, New York.

Schweder, T. (1970) Composable Markov Processes. *Journal of Applied Probability 7*, 400 – 410.

Spirtes, P. (1995) Directed cyclic graphical representations of feedback models. In *Proceedings of the 11th Annual Conference on Uncertainty and Artificial Intelligence*, pp. 491 – 498. Morgan Kaufman, San Francisco, California.

Spirtes, P., C. Glymour, and R. Scheines (2000) *Causation, Prediction, and Search* (2nd ed.) MIT Press, Cambridge, Massachusetts.

Vass, M., Avlund, K., Hendriksen, C., Andersen C. K. and N. Keiding (2002) Preventive home visits to older people in Denmark: methodology of a randomized controlled study. *Aging Clin. Exp. Res 14*, 509 – 515.

Vass, M., Avlund, K., Kvist, K., Hendriksen, C., Andersen C. K. and N. Keiding (2004) Structured home visits to older people. Are they only of benefit for women? A randomised controlled trial. *Scand. J. Prim. Health Care 22*, 106 – 111.

Verma, T. and J. Pearl (1990) Causal Networks: Semantics and Expressiveness. In R. D. Shachter, T. S. Levitt, L. N. Kanal, and J. F. Lemmer (Eds.), *Uncertainty and Artificial Intelligence*, pp. 69 – 76. North Holland, Amsterdam.

Whittaker, J. (1990) *Graphical Models in Applied Multivariate Statistics.* Wiley, Chichester.