1
Comments on Graphical Models for Stochastic Processes

Vanessa Didelez
University College London

1 Introduction

Graphical models have proven to be a valuable concept in various fields of multivariate data analysis as underpinned by several contributions to this book. However, the pure definition of conditional independence graphs is not satisfactory for the representation of two closely related dependence concepts: causal and dynamic dependence — both sharing for instance the property of being asymmetric, as opposed to conditional dependence. In graphical models directed edges symbolize specific sets of conditional independence hypotheses. Thus, without additional assumptions they neither imply nor assume causal relations or a temporal ordering of the variables. We may even construct examples where due to latent variables or to a selection process the directions of the edges do not coincide with the flow of time or causality (Lauritzen and Richardson, 2001). Therefore, extensions and modifications of conditional independence graphs have been developed. The main approaches applying graphical models for an adequate representation of causal relations are due to Spirtes et al. (1993) and Pearl (1995, 2000). Recently, graphical models have also been modified to cope with dynamic dependencies among stochastic processes (Dahlhaus, 2000; Eichler, 2000; Didelez, 2001).

In the following, I want to focus on two aspects: First, the different notions of causality occurring in the indicated literature are compared. Second, the specific problems arising in the continuous–time situation are addressed.

2 Granger causality versus intervention causality

The causality graphs introduced by Eichler (2000) are based on dynamic dependencies among the components of multivariate time series which are defined in terms of Granger causality (Granger, 1969):

Granger causality: *Let* $\{\Omega_t\}$ *be all the relevant information in the universe. Then, a subprocess* $\{X_t\} \subset \{\Omega_t\}$ *is causal for another process* $\{Y_t\}$ *if the optimal prediction w.r.t.* $\{Y_t\}$ *is less precise when based on* $\{\Omega_t\}\backslash\{X_t\}$ *instead of* $\{\Omega_t\}$.

In other words, only a process which is *directly* relevant for an optimal prediction is regarded as a cause. Obviously, the assumption of no unmeasured confounders, and thus the identifiability of Granger causes, is highly questionable in any real data situation (cf. the example by Dahlhaus and Eichler in this volume where CO and NO are correlated because both are emitted by cars). In any given setting with finite information and without background knowledge on the causal structure, the analysis is instead mostly restricted to associations between lagged variables. Granger causality nevertheless provides a very intuitive concept of dynamic association which can be generalized to a large class of stochastic processes as addressed below.

In contrast, the causality concept prevailing in the literature on graphical models or Bayesian networks is usually concerned with the effects of *interventions* as opposed to the pure observation of a condition. For stochastic processes an intervention could mean that the values of a process at each point in time are set to values that have been fixed in advance. We could for instance think of persons suffering from a chronic disease who have to follow a specific scheme when taking the prescribed drugs over an indefinite period of time. However, the definition of a dynamic intervention should also allow for interventions taking place only at selected points in time instead of all the time, and for interventions to be conditional on the past (cf. Pearl, 1994). An approach to the analysis of such sequential plans using graphs has been proposed by Pearl and Robins (1995). Without going into the technical details, let me formulate an alternative definition of dynamic causality as follows.

Intervention causality: *A process $\{X_t\}$ is causal for $\{Y_t\}$ if an intervention w.r.t. the former affects the prediction of the latter.*

Although not explicit in this definition, the identifiability of some causal effect might require that specific covariate processes have been measured (Pearl and Robins, 1995; Parner and Arjas, 1999). But in contrast to Granger causality these do not necessarily include 'all the information in the universe'. Additionally, in the above definition it does not matter whether the effect of an intervention is due to a direct influence from $\{X_t\}$ to $\{Y_t\}$ or due to an indirect pathway. Thus, any Granger cause (assuming no unobserved confounding process) is also a cause w.r.t intervention but not vice versa.

Following the tradition of intervention graphs or influence diagrams (cf. the contribution of A.P. Dawid to this book), dynamic graphical models could be extended to take interventions into account for instance by adding intervention nodes. As for the static situation, this should simplify the derivation of appropriate formulae for causal effects of interventions as well as the deduction of conditions for their identifiability. Note that the effect of an intervention can be regarded as a causal parameter (Pearl, 2000, p. 39) which is to be estimated from observational data. This does not necessarily require the actual feasibility

of an intervention.

3 The continuous–time situation

As noted by R. Dahlhaus and M. Eichler in their contribution to this book, a drawback of the different graphical models for time series is their sensitivity to the choice of intervals between the measurements. Larger intervals correspond to marginalizing over the time in between. One the one hand, this will typically create additional correlations due to common causes or mediating events occurring in the meantime. On the other hand, marginalization might also hide genuine short–term correlations. One is even led to suppose that the undirected edges in the TSC or causality graphs indicating instantaneous causality (being less obvious but also relevant in the PC graphs) are only due either to such meantime effects or to whole unobserved processes. Not only does this challenge the interpretation in terms of causality — in the sense of Granger or of interventions — but additionally, the alternative Markov properties which apply to TSC graphs do not agree with the foregoing explanation of instantaneous causality. It has been shown that general AMP chain graphs cannot be generated by marginalizing directed acyclic graphs representing the underlying data generating process (Richardson, 1998). Yet, it is not clear how else instantaneous causality could arise.

In addition to the foregoing problems regarding the interpretation of graphical models for time series, it is difficult to follow the advice of choosing the intervals small enough in order to prevent spurious correlation when the underlying process is continuous in time. In this situation one should instead consider a generalization of the idea of dynamic dependencies inherent in Granger causality to the continuous–time situation. The basic approach goes back to Schweder (1970) who introduced the notion of *local independence* for Markov processes. Aalen (1987) extended this to stochastic processes that admit a Doob–Meyer decomposition and Florens and Fougère (1996) showed the analogy to Granger causality. Local independence reads as follows.

Local independence: *A process* $\{Y_t\}$ *is said to be locally independent of* $\{X_t\}$ *given some further information* $\{Z_t\}$ *if the predictable part, i.e. the compensator* Λ_t^Y *of* $\{Y_t\}$ *remains the same regardless of whether it is conditional on the past* $F_{t^-}^{XYZ}$ *of all three processes or only on the past* $F_{t^-}^{YZ}$ *of* $\{Y_t\}$ *and* $\{Z_t\}$.

Aalen (1987) additionally assumed that the innovations of the involved processes are uncorrelated because a notion of independence between processes would make no sense if they were fed by the same innovations. This assumption should in practice be carefully checked since it may again be violated by some unmeasured common causal process. Note that the above definition does not refer to 'all information in the universe'. Consequently, it does not claim to define causal relations.

A graphical representation of local independence structures is straightfor-

ward by representing the components of a multivariate process as vertices and drawing directed edges for any local dependence whilst omitting the edge when there is a local independence. The resulting local independence graphs have been analyzed for the special case of counting processes in Didelez (2001). The global Markov properties in these graphs can be obtained by the application of a newly developed separation criterion called *δ–separation* which takes into account that local independence is asymmetric and that the graphs may be cyclic. The augmentation of the graphs with intervention nodes for the computation and identification of causal effects seems to be a promising approach.

While it is often assumed that point processes can be observed in continuous time by simply registering the time of each event, this rarely seems possible for other types of stochastic processes which vary steadily in time. As an example one might consider the medical time series typically arising in intensive care online monitoring which are often recorded in intervals of a few seconds. A graphical representation of the dependence structure would have to take into account that the time between consecutive measurements is always unobserved. Thus, a class of graphical models generated by marginalizing suitable continuous–time models over the meantime of the intervals is called for. This should be subject to future research.

References

Didelez, V. (2001). *Graphical Models for Event History Analysis based on Local Independence.* Ph.D. thesis. Logos, Berlin.

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

Lauritzen, S.L. and Richardson, T.S. (2001). Chain graph models and their causal interpretation. Technical report, Aalborg University, Denmark.

Parner, J. and Arjas, E. (1999). Causal reasoning from longitudinal data. Technical report, University of Helsinki, Finland.

Pearl, J. (1994). A probabilistic calculus of actions. In R. Lopez de Mantaras and D. Poole (Eds.), *Uncertainty in Artificial Intelligence*, Volume 10, 454–462. Morgan Kaufmann, San Mateo.

Pearl, J. and Robins, J. (1995). Probabilistic evaluation of sequential plans from causal models with hidden variables. In P. Besnard and S. Hanks (Eds.), *Uncertainty in Artificial Intelligence*, Volume 11, 444–453. Morgan Kaufmann, San Francisco.

Richardson (1998). Chain graphs and symmetric associations. In M. Jordan (Ed.), *Learning in Graphical Models*, 231-260. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Spirtes, P., Glymour, C. and Scheines, R. (1993). *Causation, Prediction and Search.* Volume 81 of *Lecture Notes in Statistics.* Springer, New York.