Specification searches using MAG models

Thomas Richardson, Heiko Bailer, Moulinath Banerjee
University of Washington, Department of Statistics
Seattle, WA 98195-4322, USA
$\{$*tsr, heiko, mouli*$\}$*@stat.washington.edu*

The class of maximal ancestral graph (MAG) models can be used to model the Markov structure among the observed variables in a data generating process. A general data generating process may be described by a directed acyclic graph (DAG) under marginalization, representing unobserved variables, and conditioning, corresponding to selection effects. In the Gaussian case, MAG models can be parametrized via a recursive set of linear equation with correlated errors. In this paper we define the class of maximal ancestral graphs and give the associated parametrization.

Keywords: Graphical Gaussian models; latent variables; selection variables; maximal ancestral graphs; MAG models; curved exponential family.

1. Introduction

Specification searches for directed acyclic graph (DAG) models with latent (unmeasured) variables face several problems: first the number of possible models is infinite, and second, calculating scores for latent variable (LV) models is generally much slower than calculating scores for models without LVs. In addition, general LV models have a number of other features which make them hard to work with: LV models may be underidentified, leading to multimodal or flat likelihood surfaces; further, as described by Geiger and Meek (1998), LV models are stratified exponential families, rather than curved exponential families (like DAGs without LVs), and consequently the results which guarantee the asymptotic consistency of scores such as BIC do not apply; finally, LV models do not have a well-defined dimension.

To address some of these problems Spirtes, Richardson and Meek (1997) have introduced a class of graphical Gaussian models, called MAG models, which do not include latent variables, but do impose the independence constraints given by latent variable models over the observed variables, and *only* these constraints. Thus for any given LV model there is a MAG model to which it is Markov equivalent. However, since LV models often impose non-independence constraints, the corresponding MAG model will parametrize a superset of the distributions parametrized by the LV model. In contrast to latent variable models, MAG models are always statistically identifiable, and have a well-defined dimension, forming curved exponential families.

2. Ancestral Mixed Graphs

In this paper we consider *mixed* graphs in which three types of edge $\{-, \rightarrow, \leftrightarrow\}$ may occur. We naturally extend the d-separation criterion (see Pearl, 1988) to mixed graphs as follows: a pair of consecutive edges meeting at a vertex z on a path form a *collider* if both edges have an arrowhead at z, i.e. $\rightarrow z \leftarrow$, $\leftrightarrow z \leftrightarrow$, $\leftrightarrow z \leftarrow$, $\rightarrow z \leftrightarrow$. Two consecutive edges which do not form a collider are said to form a *non-collider*. A vertex a is said to be an *ancestor* of a vertex b if **either** there is a directed path $a \rightarrow \cdots \rightarrow b$ on which every edge is of the form '$\rightarrow$', and has the same orientation, **or** $a = b$. $an(x)$ is the set of vertices which are ancestors of x.

A path between vertices x and y in a mixed graph is said to be *m-connecting given a set* Z if (i) every non-collider on the path is not in Z, and (ii) every collider on the path is an ancestor of Z. If there is no path m-connecting x and y given Z, then x and y are said to be *m-separated* given Z. Sets X and Y are said to be *m-separated* given Z, if for every pair x, y, with $x \in X$ and $y \in Y$, x and y are m-separated given Z.

A mixed graph $\mathcal{M}$, is *ancestral* if the following conditions hold in $\mathcal{M}$:

(i) $\mathcal{M}$ is acyclic: if $a \neq b$, and $a \in \text{an}(b)$ then $b \notin \text{an}(a)$;

(ii) If there is an edge $a \leftrightarrow b$ in $\mathcal{M}$, then $a \notin \text{an}(b)$ and $b \notin \text{an}(a)$.

(iii) If there is an edge $a - b$ in $\mathcal{M}$ then there is no edge $c \rightarrow a$ or $c \leftrightarrow a$ in $\mathcal{M}$.

It is not hard to show that at most one edge (of any type) may occur between each pair of vertices in an ancestral mixed graph. It is always possible to partition the vertices of an ancestral mixed graph into two sets Ω, Δ such that the induced subgraph on Ω is completely undirected, and the induced subgraph on Δ contains no undirected edges; further, if there is an edge connecting a vertex $u \in \Omega$ to a vertex $d \in \Delta$ then it is oriented as $u \rightarrow d$. (The induced subgraphs on Ω and Δ need not be connected.) For the MAG in Figure 1 (e), such a partition is $\Omega = \{a, b, c\}$, $\Delta = \{d, e, f\}$. This decomposition is significant because it implies a corresponding factorization of the joint density:

$$\text{P}(\Omega, \Delta) = \text{P}(\Omega) \cdot \text{P}(\Delta \mid \Omega)$$

Thus we may parametrize the two pieces separately. The model for Ω is given by an undirected graphical model. It remains to parametrize $\text{P}(\Delta \mid \Omega)$ which corresponds to a an acyclic mixed graph containing no undirected edges; graphs of this type are termed *directed* acyclic mixed graphs.

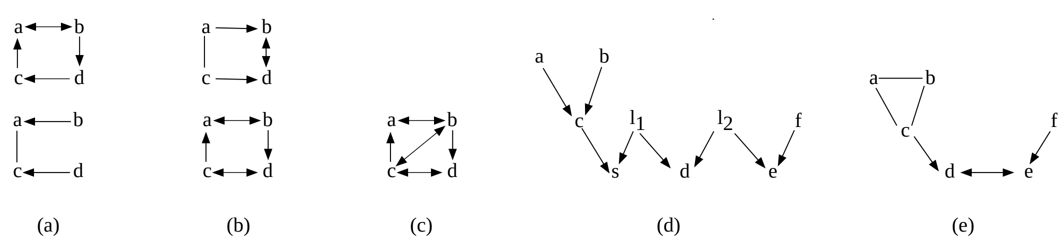

Figure 1. ***(a) Two mixed graphs that are not ancestral; (b) two maximal ancestral mixed graphs; (c) an ancestral mixed graph that is not maximal; (d) A DAG $\mathcal{G}$ with vertex set $V = \{a, b, c, d, e, f, l_1, l_2, s\}$; (e) the MAG $\mathcal{M}$ corresponding to $\mathcal{G}$ under the partition $O = \{a, b, c, d, e, f\}$, $S = \{s\}$, and $L = \{l_1, l_2\}$.***

An ancestral mixed graph is said to be a *maximal* ancestral mixed graph (MAG) if it satisfies the following condition:

(iv) If there is no edge between a and b, then there is some subset Z of the other vertices such that a and b are m-separated given Z.

Figure 1(c) shows a simple example of an ancestral mixed graph which is *not* maximal: there is no edge between a and d in the graph, and yet a and d are m-connected by all four subsets of the other variables $\{b, c\}$. The pairwise Markov properties for DAGs and undirected graphs (see Lauritzen, 1996), imply that every graph in either of these classes is maximal. A non-maximal ancestral mixed graph can be converted into a MAG by adding bi-directed edges ($\leftrightarrow$), between every pair of vertices for which no m-separating set exists. The resulting maximal ancestral graph will represent the same set of m-separation as held in the non-maximal graph.

Maximal ancestral mixed graphs lead to a natural parametrization of the associated Markov model in the Gaussian case: A Gaussian MAG *model* is a set of multivariate Gaussian distributions P, such that for all disjoint sets X, Y and Z, (where Z may be empty), if X is m-separated from Y given Z in the MAG, then $X \perp\!\!\!\perp Y \mid Z$ in P. A Gaussian MAG model (with means fixed at zero) can be parametrized as follows: The undirected component $P(\Omega)$ is parametrized via a covariance selection model (Dempster, 1974). The directed component, $P(\Delta \mid \Omega)$, is parametrized as follows:

(i) Associate with each y_j in Δ a linear equation, expressing y_j as a linear function of its parents plus an error term:

$$y_j = \sum_{x_i \in \mathrm{pa}(y_j)} \alpha_{ij} x_i + \epsilon_{y_j}$$

where $\mathrm{pa}(y_j)$ is the set of parents of y_j.

(ii) Specify a multivariate Gaussian distribution over $(\epsilon_y)_{y \in \Delta}$ (with mean zero) satisfying the condition that if there is not a double-headed edge $x \leftrightarrow y$ in the graph, then $\mathrm{Cov}(\epsilon_x, \epsilon_y) = 0$, but otherwise unrestricted.

Note that the MAG model parametrizes *all* Gaussian distributions in which the conditional independence relations corresponding to the m-separation relations hold. For this to hold for this parametrization it is necessary that the ancestral mixed graph be maximal. (See Spirtes *et al.*, 1997.)

MAGs are closely related to the summary graphs in Cox and Wermuth (1996), though there are key differences. In particular, in a summary graph it is possible to have a pair of vertices x and y that are not adjacent and yet there is no subset of the remaining variables which make x and y conditionally independent; it is also possible to have more than one edge between a pair of vertices in a summary graph.

3. Directed Acyclic Graphs with Latent and Selection Variables

Cox and Wermuth (1996) and Spirtes *et al.* (1997) consider a DAG $\mathcal{G}$ with vertex set V, partitioned into observed (O), latent (L), and selection (S) subsets. The interpretation is that $\mathcal{G}$ represents a causal, or data-generating mechanism; O represents the subset of the variables that are observed; S represents a set of variables which, due to the nature of the mechanism selecting the sample, are conditioned on in the subpopulation from which the sample is drawn; the variables in L are not observed and for this reason are called *latent*.

Spirtes *et al.* show that given such a DAG $\mathcal{G}$, with vertex set V, partitioned into (O, S, L) there is a corresponding MAG $\mathcal{M}$ with vertex set O, such that for disjoint sets $X, Y, Z \subseteq O$, X and Y are m-separated given $Z \cup S$ in $\mathcal{G}$, if and only if X and Y are m-separated given Z in $\mathcal{M}$. Thus the MAG captures the independencies holding among the observed variables in the selected subpopulation. See Figure 1(d),(e) for an example.

A vertex x is said to be *anterior* to y if there is a path from x to y on which every edge $\langle u, v \rangle$ is either undirected, $u - v$, or directed, $u \to v$, with orientation from x to y; $\mathrm{ant}_{\mathcal{M}}(x)$ is the set of vertices anterior to x in $\mathcal{M}$, and similarly $\mathrm{ant}_{\mathcal{M}}(X) = \{y \mid y \in \mathrm{ant}_{\mathcal{M}}(x) \text{ for some } x \in X\}$. Finally note that for a DAG $\mathcal{G}$, x is anterior to y if and only if x is an ancestor of y.

The algorithm for creating a MAG $\mathcal{M}$ from a DAG $\mathcal{G}$ with vertex set partitioned into O, S, L, requires three steps:

(i) Form an undirected graph $\mathcal{M}$ with vertex set O in which there is an edge $x - y$ if and only if for every subset $Z \subseteq O$, x and y are m-connected given $Z \cup S$ in $\mathcal{G}$.

(ii) If there is an edge $x - y$ in $\mathcal{M}$, and $x \notin \text{ant}_{\mathcal{G}}(\{y\} \cup S)$, and $y \notin \text{ant}_{\mathcal{G}}(\{x\} \cup S)$ then replace $x - y$ with $x \leftrightarrow y$.

(iii) If there is an edge $x - y$ in $\mathcal{M}$, and $x \in \text{ant}_{\mathcal{G}}(\{y\} \cup S)$, but $y \notin \text{ant}_{\mathcal{G}}(\{x\} \cup S)$ then replace $x - y$ with $x \rightarrow y$.

The algorithm can in fact be applied directly to an ancestral graph (not simply a DAG) whose vertices are partitioned O, S, L, to generate a MAG encoding the m-separation relations holding between vertices in O given S in an ancestral graph $\mathcal{G}$. This is why steps (ii) and (iii) above are defined via anterior sets rather than ancestral sets (for DAGs these are equivalent).

Step (i), together with the fact stated above, that $\mathcal{M}$ captures all m-separation relations holding between disjoint subsets X, Y of O given a third subset Z union S, ensures that the resulting graph $\mathcal{M}$ is maximal. It is simple to check that the orientation rules given in steps (ii) and (iii) result in a mixed graph that is ancestral.

Note that if there is an edge $x - y$ in $\mathcal{M}$, after steps (ii) and (iii), then both x and y are in $\text{ant}_{\mathcal{G}}(S)$. Further since in a DAG $\mathcal{G}$ there are no undirected edges $\text{ant}_{\mathcal{G}}(X) = \text{an}_{\mathcal{G}}(X)$, the set of ancestors of X in $\mathcal{G}$, so both x and y are ancestors of vertices in S (in $\mathcal{G}$).

REFERENCES

Cox, D.R. and Wermuth, N. (1996). *Multivariate Dependencies.* Chapman and Hall.

Dempster, A.P. (1972). Covariance Selection. *Biometrics*, **28**, 157-175.

Geiger, D. and Meek, C. (1998) Graphical Models and Exponential Families. *Proceedings of the 14th Conference on Uncertainty and Artificial Intelligence*, (G. Cooper and S.Moral eds.) Morgan Kaufmann: San Francisco, CA, 156-165.

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

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

Spirtes, P., Richardson, T. and Meek, C. (1997). The dimensionality of mixed ancestral graphs. Technical Report, CMU-PHIL-83, Philosophy Department, CMU.

RÉSUMÉ

La famille de graphes ancestraux maximaux (GAM) peut être utilisée pour modeler la structure Markovienne des variables observées dans les processus générateurs de données. Un processus générateur de données général peut être décrit à l'aide d'un graphe acyclique directionnel (GAD) qui est marginalisé et conditionné. Dans le cas Gaussien, les GAMs peuvent être paramétrisés via un ensemble recurrent d'équations linéaires avec erreurs corrélées. Dans ce papier, nous définissons la famille de graphes ancestraux maximaux ainsi que leur paramétrisation. Finalement, nous présentons un algorithm qui transforme un GAD, qui est marginalisé et conditionné, en un GAM.