Dipartimento di Statistica
"Giuseppe Parenti"

WORKING PAPER 2004/08

Dipartimento di Statistica "G. Parenti" – Viale Morgagni 59 – 50134 Firenze - www.ds.unifi.it

Chain Graphs
for Multilevel Models

Anna Gottard, Carla Rampichini

Università degli Studi
di Firenze

Statistics for experimental and technological research

Chain Graphs for Multilevel Models

Anna Gottard and Carla Rampichini

Department of Statistics "G. Parenti", University of Florence, Italy.

Summary. The present work proposes a possible solution to extend graphical models for correlated data. Particularly, the paper focused on hierarchical data structures, considering two-level random intercept models. The proposed solution allows to use the existing chain graphs theory in a straightforward way. After a brief introduction to multilevel models and a description of the conditional independencies derived from the model, the paper defines chain graphs for multilevel models. Some examples illustrates and an application to real data show the usefulness of the proposed models.

Keywords: chain graphs, conditional independence, hierarchical data, multilevel models, two-level random intercept model

1. Introduction

Probabilistic independence is a very important way to look at a statistical model. Graphical models are a key technique in dealing with this topic. In recent years, the literature on graphical models has grown considerably, particularly at a theoretical level. Their use in applied statistics, however, is lagged behind, they mostly are considered a theoretical topic. The purpose of this article is to show how the theory of conditional independence and graphical models can be successfully employed in the analysis of complex data sets, such as hierarchical data structures, which are largely present in applications. The class of models considered in this paper is that of multilevel models (Snijders and Bosker, 1999). These models are useful tools to treat clustered correlated data, particularly if one is interested in relations among variables at different levels in a hierarchical system. In case of hierarchically clustered data, observations belonging to the same cluster are not independent, while observations of different clusters are independent. In multilevel models, in order to deal with this kind of dependence, a latent variable is introduced in the model and the independence among observations is assumed conditionally on it.

The present work proposes an extension of chain graphs to represent multilevel models. As a first step only the two-level random intercept model is analyzed.

Other authors proposed alternative representations: Johnson and Hoeting (2003) proposed a two component graphical chain model to represent a random effects model with categorical variables measured on many random sites, while Buntine (1994) uses plates to represent multilevel data structures in a Bayesian framework.

Section 2 introduces multilevel models and describes the conditional independencies derived from the model. Section 3 provides some basics to understand the the semantics used in this work and define chain graph for multilevel models. Section 4 illustrates some examples of multilevel graph models. Section 5 presents an application to real data, while Section 6 concludes giving some direction for future research.

2. Conditional independence in multilevel models

Many kinds of data have a hierarchical, nested, or clustered structure: for example, repeated measures on the same subjects over time, students in schools, patients in hospitals and so on. Statistical units belonging to the same cluster have or tend to have similar observable and unobservable characteristics. This implies that the group and its members both influence and are influenced by the group membership. Ignoring this relationship leads to overlooking the importance of group effects, and may also render invalid many of the traditional statistical analysis techniques used for studying data relationships. In multilevel models a random variable is inserted to take into account such kinds of dependency among observations.

Consider the random vector $\mathbf{Y} = (Y_{11}, \ldots, Y_{n_1 1}, \ldots, Y_{n_J J})$ and the matrix of random vectors $\mathbf{Z} = (Z_1, \cdots, Z_K)$ on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. In a multilevel framework we assume that

$$Y_{ij} \not\perp\!\!\!\perp Y_{i'j} \mid \mathbf{Z}, \quad \forall i \neq i', \quad i, i' = 1, 2, \ldots, n_j$$

given that i, i' are two statistical units belonging to the same group j, while

$$Y_{ij} \perp\!\!\!\perp Y_{i'j'} \mid \mathbf{Z} \qquad \text{and} \qquad Y_{ij} \perp\!\!\!\perp Y_{i'j'}, \quad \forall j \neq j', \quad j, j' = 1, 2, \ldots, J$$

since j, j' are two different groups.

If Y is a continuous response variable, assuming without loss of generality only one explanatory variable Z, the basic linear two-level random intercept model is specified in the following way (Goldstein, 2003):

$$Y_{ij} = \alpha + \beta \mathbf{Z}_{ij} + \tau \mathcal{U}_j + \varepsilon_{ij}, \tag{1}$$

with $i = 1, 2, \ldots, n_j$ statistical units for the j-th group ($j = 1, 2, \ldots, J$). In (1), α is the intercept; $\mathbf{Z}_{ij}$ is the matrix of the explanatory variables and β the corresponding vector of fixed coefficients; τ is the square root of the second level variance; the random variables ε_{ij} and $\mathcal{U}_j$ are the disturbances, respectively at the first (individual) and second (group) levels, under the hypotheses: (i) $E(\varepsilon_{ij}) = 0$ and $Var(\varepsilon_{ij}) = \sigma^2$, (ii) $\mathcal{U}_j \stackrel{iid}{\sim} N(0, 1)$, (iii) the ε_{ij}'s and $\mathcal{U}_j$'s are mutually independent, (iv) $\mathcal{U}_j \perp\!\!\!\perp \mathbf{Z}$.

According to model (1) and the (i)-(iv) assumptions the following statements hold:

- $Y_{ij} \perp\!\!\!\perp Y_{i'j} \mid \mathcal{U}_j, \mathbf{Z}$;
- $\tau = 0 \Rightarrow f(\mathbf{Y}_j \mid \mathcal{U}_j, \mathbf{Z}) = f(\mathbf{Y}_j \mid \mathbf{Z})$, that is $\mathbf{Y}_j \perp\!\!\!\perp \mathcal{U}_j \mid \mathbf{Z}$.
- $f(\mathcal{U}_j \mid \mathbf{Z}) = f(\mathcal{U}_j)$

Therefore the joint probability distribution for each group j, $j = 1, 2, \ldots, J$, can be factorized as:

$$\begin{aligned} f(\mathbf{Y}_j, \mathcal{U}_j, \mathbf{Z}) &= f(\mathbf{Y}_j \mid \mathcal{U}_j, \mathbf{Z}) f(\mathcal{U}_j) f(\mathbf{Z}) = f(\mathbf{Y}_j \mid \mathcal{U}_j, \mathbf{Z}) f(\mathcal{U}_j) f(\mathbf{Z}) \\ &= \left[\prod_{i=1}^{n_j} f(y_{ij} \mid \mathcal{U}_j, \mathbf{Z}) \right] f(\mathcal{U}_j) f(\mathbf{Z}) \end{aligned} \tag{2}$$

where $\mathbf{Y}_j = \{y_{1j}, \ldots, y_{n_j j}\}'$.

If the joint density is absolutely continuous and strictly positive, for the conditional independencies properties and given the assumption $\mathcal{U}_j \perp\!\!\!\perp \mathbf{Z}$, if $\mathbf{Y}_j \perp\!\!\!\perp \mathcal{U}_j \mid \mathbf{Z}$ then $\mathcal{U}_j \perp\!\!\!\perp (Y, \mathbf{Z})$ and the model (1) reduces to a standard regression model. Nevertheless, the assumption $\mathcal{U}_j \perp\!\!\!\perp \mathbf{Z}$ does not imply that $\mathcal{U}_j \perp\!\!\!\perp \mathbf{Z} \mid \mathbf{Y}_j$, unless $\mathbf{Y}_j \perp\!\!\!\perp \mathcal{U}_j$ or $\mathbf{Y}_j \perp\!\!\!\perp \mathbf{Z}$.

It is well known that under the Gaussian distribution, conditional independence is equivalent to the assumption that a certain element of the concentration matrix is equal to zero and marginal independence among variables is equivalent to the assumption that the corresponding element of the covariance matrix is zero (Whittaker, 1990). Therefore, if $(\mathbf{Y}, \mathcal{U}, Z) \sim MN(\mu, \mathbf{\Sigma})$, it is useful to look at the covariance and concentration matrices derived from model (1).

For each group j, $\mathbf{V}_j = \{\mathbf{Y}_j, \mathcal{U}_j, \mathbf{Z}\}$ and $\mathbf{V}_j \sim MN(E(\mathbf{V}_j), \Sigma_{\mathbf{V}_j})$, where $E(\mathbf{V}_j) = (\mu_{\mathbf{Y}_j}, 0, \mu_{\mathbf{Z}})$, with $\mathbf{\Sigma} = \mathbf{I}_{J\times J} \otimes \mathbf{\Sigma}_{\mathbf{V}_j}$.

Following Cox and Wermuth (1996), given that $\mathcal{U}_j \perp\!\!\!\perp \mathbf{Z}$, we can write the covariance matrix of $\mathbf{V}_j$ as:

$$\mathbf{\Sigma}_{V_j} = \begin{pmatrix} \Sigma_{\mathbf{Y}_j\mathbf{Y}_j} & \Sigma_{\mathbf{Y}_jU} & \Sigma_{\mathbf{Y}_j\mathbf{Z}} \\ \Sigma_{U\mathbf{Y}_j} & \Sigma_{UU} & \Sigma_{U\mathbf{Z}} \\ \Sigma_{\mathbf{Z}\mathbf{Y}_j} & \Sigma_{\mathbf{Z}U} & \Sigma_{\mathbf{Z}\mathbf{Z}} \end{pmatrix} = \begin{pmatrix} \Sigma_{\mathbf{Y}_j\mathbf{Y}_j} & \Sigma_{\mathbf{Y}_jU} & \Sigma_{\mathbf{Y}_j\mathbf{Z}} \\ \Sigma_{U\mathbf{Y}_j} & \Sigma_{UU} & \mathbf{0} \\ \Sigma_{\mathbf{Z}\mathbf{Y}_j} & \mathbf{0} & \Sigma_{\mathbf{Z}\mathbf{Z}} \end{pmatrix}$$

Note that, since $Y_{ij} \perp\!\!\!\perp Y_{i'j} \mid (\mathcal{U}_j, \mathbf{Z})$, the element $\Sigma^{\mathbf{Y}_j\mathbf{Y}_j}$ of the concentration matrix $\Sigma^{-1}_{V_j}$ is a diagonal matrix. Moreover, $\mathcal{U}_j \perp\!\!\!\perp \mathbf{Z} \Leftrightarrow \Sigma_{U\mathbf{Z}} = \mathbf{0}$, in Σ_{V_j}, but in general $\mathcal{U}_j \not\!\perp\!\!\!\perp \mathbf{Z} \mid \mathbf{Y}_j$,so the element $\Sigma^{U\mathbf{Z}}$ of the concentration matrix might not be a null matrix.

The normal vector $\mathbf{Y}_j$ and $\mathcal{U}_j$ are conditionally independent given $\mathbf{Z}$ if and only if either (Whittaker, 1990):

i. cov$(\mathbf{Y}_j, \mathcal{U}_j \mid \mathbf{Z}) = 0$ or $\Sigma_{\mathbf{Y}_jU|\mathbf{Z}} = \Sigma_{\mathbf{Y}_jU} - \Sigma_{\mathbf{Y}_j\mathbf{Z}}\Sigma^{-1}_{\mathbf{Z}\mathbf{Z}}\Sigma_{U\mathbf{Z}} = 0$;
ii. or the block of the concentration matrix $\Sigma^{\mathbf{Y}_jU} = \mathbf{0}$.

Under model (1), assuming without loss generality $n_j = 2$ so that $\mathbf{Y}'_j = (Y_{1j}, Y_{2j})$ and only one explanatory variable Z, the Normal distribution hypothesis for $\mathbf{Y}_j$ corresponds to a normal distribution for the ε_{ij}. As a consequence, conditional on $\mathcal{U}_j$ and Z, the response Y_{ij} on each observation is normally distributed with $\mathbb{E}[Y_{ij} \mid \mathcal{U}_j = u_j, Z = z_{ij}] = \alpha + \beta z_{ij} + \tau u_j$, var$(Y_{ij}|\mathcal{U}_j = u_j, Z = z_{ij}) = \sigma^2$ and cov$(Y_{ij}, Y_{i'j}|\mathcal{U}_j = u_j, Z) = 0$, while marginally with respect $\mathcal{U}_j$ var$(Y_{ij} \mid Z = z_{ij}) = \sigma^2 + \tau^2$ and cov$(Y_{ij}, Y_{i'j'} \mid Z = z_{ij})$ is equal to zero if $j \neq j'$ and equal to τ^2 if $j = j'$.

Thus, the covariance matrix $\mathbf{\Sigma}_j$ of the observed variables can be written in terms of the multilevel model parameters:

$$\mathbf{\Sigma}_j = \begin{pmatrix} \sigma^2 + \tau^2 + \beta^2\sigma^2_Z & \tau^2 + \beta^2\sigma^2_Z & \beta\sigma^2_Z \\ \tau^2 + \beta^2\sigma^2_Z & \sigma^2 + \tau^2 + \beta^2\sigma^2_Z & \beta\sigma^2_Z \\ \beta\sigma^2_Z & \beta\sigma^2_Z & \sigma^2_Z \end{pmatrix} \tag{3}$$

while the observed concentration matrix Σ^{-1}_j has the form:

$$\mathbf{\Sigma}^{-1}_j = \frac{1}{\mid \Sigma_j \mid} \begin{pmatrix} \cdot & \tau^2 & -\beta\sigma^2_Z\sigma^2 \\ \tau^2 & \cdot & -\beta\sigma^2_Z\sigma^2 \\ -\beta\sigma^2_Z\sigma^2 & -\beta\sigma^2_Z\sigma^2 & \cdot \end{pmatrix} \tag{4}$$

Note that $\sigma^{Y_{1j}Y_{2j}} = \tau^2$, so if $\tau = 0$ then $\sigma^{Y_{1j},Y_{2j}} = 0$, so $Y_{1j} \perp\!\!\!\perp Y_{2j} \mid Z$, that is the Y_{ij} are independent and identical distributed variables. Moreover, it can be seen that $\Sigma_{\mathbf{Y}_jZ} - \Sigma_{\mathbf{Y}_jU}\Sigma^{-1}_{UU}\Sigma_{ZU} = \beta\sigma^2_Z$. Recalling that the normal vector $\mathbf{Y}_j$ and Z are conditionally independent given $\mathcal{U}_j$ if and only if either (Whittaker, 1990):

i. cov$(\mathbf{Y}_j, Z \mid \mathcal{U}_j) = 0$ or $\Sigma_{\mathbf{Y}_jZ|U} = \Sigma_{\mathbf{Y}_jZ} - \Sigma_{\mathbf{Y}_jU}\Sigma^{-1}_{UU}\Sigma_{ZU} = 0$;
ii. or the block of the concentration matrix $\Sigma^{\mathbf{Y}_jZ} = \mathbf{0}$,

it is manifest that if $\beta = 0$ then $(Y_{1j}, Y_{2j}) \perp\!\!\!\perp Z \mid \mathcal{U}$, and this implies that $Y_{1j} \perp\!\!\!\perp Z \mid \mathcal{U}$ and $Y_{2j} \perp\!\!\!\perp Z \mid \mathcal{U}$.

3. Chain graphs for multilevel models

There are many types of graphs in literature, with different semantics. This Section provides some basics to understand the the semantic used in this work and define chain graph multilevel models. For an extended presentation of chain graphs theoretical concepts refer to Lauritzen (1996).

A graph $\mathcal{G}$ is a pair (V, E), where V is a nonempty set of *nodes*, representing random variables, and E is a subset of the set $V \times V$ of ordered pair of distinct nodes, called *edges*. The assumption that E consists of pair of distinct nodes, implies that there are no self-loops. We say that $u, v \in V$ are *neighbors*, if both (v, u) and $(u, v) \in E$; these edges are called undirected (—) and indicate a sort of symmetric association. Edges with either (v, u) or $(u, v) \in E$ are called directed edges or arrows ($\longrightarrow$), representing a sort of asymmetric relation. When $u \longrightarrow v$, u is called *parent* of v and v is called *child* of u. A lack of connection between two nodes represents a sort of conditional independence, according to Markov properties. A *chain graph* is a graph admitting both undirected and directed edges without partially directed cycles. In a chain graph the set V can be partitioned into an ordered sequence of nonempty subsets, called *blocks*, forming the so-called *dependence chain* $V = \mathcal{C}_1 \cup \ldots \cup \mathcal{C}_r$, with $\mathcal{C}_i \cap \mathcal{C}_j = \emptyset$, $\forall i \neq j$, $i, j = 1, 2, \ldots, r$. Nodes in a same block can be joined only by undirected edges, while nodes in different blocks can be connected only by arrows, the direction being the one specified by the ordering of the blocks. Given a probability measure P over $\mathcal{X} = \mathcal{X}_V$, a product probability space for all the random variables in V, it can be said that P obeys to the *block-recursive Markov property* with respect to a chain graph $\mathcal{G}$ if, for each non–adjacent pair $u, v \in V$, $u \perp\!\!\!\perp v \mid \tilde{V}_{uv} \backslash (u, v)$, where $\tilde{V}_{uv}$ is the largest subset of V containing all random variables in the same block or in a previous block of u and v. Hereafter, in this paper, the so-called Lauritzen-Wermuth-Frydenberg (LWF) Markov properties (Lauritzen and Wermuth, 1989; Frydenberg, 1990) are adopted.

Chain graph models concentrate their attention on the relationships among variables, given that individuals are regarded as independent. This is not true when the response variables are correlated into groups of statistical units as in multilevel models. This calls for a new kind of chain graphs, where individual responses are represented by nodes in the graph. We start giving some definitions of the new object strictly necessary to the construction of this new kind of chain graph.

DEFINITION 3.1. An *individual node* is a node of a graph $\mathcal{G}$ representing a random variable of a specific statistical unit.

Given that an individual node represent also a random variable, it will be impossible to distinguish it by an aggregate node but its label: for example, for a variable Y, Y will be the label of the aggregate node, while Y_i will be the label for the individual node.

DEFINITION 3.2. *An* individual graph *is a graph $\mathcal{G} = (V, E)$ with V containing individual nodes.*

Note that it is advisable to include individual nodes in a graph if the conditional independence structure among the individual random variables represented can be modified by the inclusion or the omission of one or more random variables. For example, let Y_i and Y_j be two individual nodes for the statistical units i and j respectively, and Z an explanatory random variable. If, as in simple random samples, $Y_i \perp\!\!\!\perp Y_j$ for construction, it is not expedient to represent both Y_i and Y_j in the graph. On the contrary, if $Y_i \not\!\perp\!\!\!\perp Y_j$ for construction then, the introduction of individual nodes is to be preferred.

DEFINITION 3.3. *A* grouping latent node *is a latent node representing an unobserved random variable $\mathcal{U}$ such that*

$$Y_{ij} \perp\!\!\!\perp Y_{i'j} \mid \mathcal{U}_j, \mathbf{Z} \qquad \textit{and} \qquad Y_{ij} \not\!\perp\!\!\!\perp Y_{i'j} \mid \mathbf{Z}$$

for $Y_{ik}, Y_{jk} \in ch(\mathcal{U})$.

Note that the variable $\mathcal{U}_j$ represents the kind of unobserved factor mentioned in Section 2. In the graph, the grouping latent node will be drawn as $\oslash$, as it belongs to the class of latent nodes described in Cox and Wermuth (1996), such that marginalizing over this kind of node leads to a correlation between its children nodes.

DEFINITION 3.4. *A* random–intercetpt multilevel graph $\mathcal{G} = (V, E)$ *is an individual graph fulfilling the following conditions:*

(a) with respect to the variable nature, the set of nodes V is partitioned into three subsets: the set $\mathcal{I}$, of the individual nodes, with cardinality $|\mathcal{I}| \geq 2$, the set Ψ containing grouping latent nodes, and the set Λ of all the other random variables;

(b) $\mathcal{G}$ is a chain graph, with V partitioned into an ordered sequence of *blocks*, $V = \mathcal{C}_1 \cup \ldots \cup \mathcal{C}_r$, such that

  (b.1) if $v \in \mathcal{I}$ then $v \in \mathcal{C}_r$;

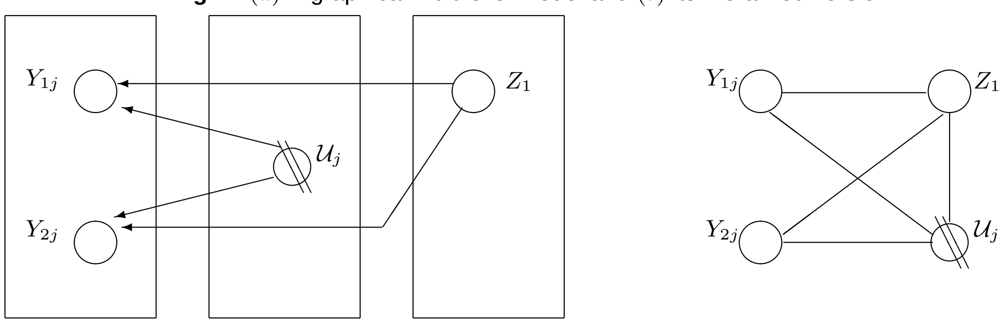
  (b.2) if $v \in \Psi$ then $v \in \mathcal{C}_{r-1}$;

  (b.3) if $v \in \Lambda$ then $v \in \bigcup_{i=1}^{r-2} \mathcal{C}_i$;

(c) $E = (E^I, E^\Upsilon)$. E^Υ contains the pairs (u, v), where $u, v \in \Upsilon = \Lambda \cup \Psi$, E^I contains the pairs (u, i), where $u \in \Upsilon, i \in \mathcal{I}$: if the edge (u, i) exists, then also the edge (u, i') is an element of E^I, for each $i' \in \mathcal{I}$. Therefore the set E^I can be represented as:
$E^I = \big\{\{(u, i), \forall i \in \mathcal{I}\}, \cdots, \{(v, i), \forall i \in \mathcal{I}\}\big\}$.

Note that each element of E^I is a set with the same cardinality of $\mathcal{I}$, that is: $|\{(u, i), \forall i \in \mathcal{I}\}| = |\mathcal{I}|$. As defined in point (b.2) Definition 3.4, in a multilevel graph a grouping latent node is inserted in the second-last block, because $\mathcal{U}_j$ represents the residual heterogeneity in the responses after considering all the observable explanatory variables. Moreover, the point (c) of Definition 3.4 states that each Z affects every Y_{ij} in the same manner, so the conditional independencies involving Y_{ij} are the same for all the i, j. The principal advantage of this formulation is that usual Markov properties of block recursive graphical models and factorization criterion (Lauritzen, 1996) apply also in this case.

Fig. 1. (a) A graphical multilevel model and (b) its moralized version.



(a) (b)

To assess the global Markov property, the moralized version of a multilevel graph has to be drawn and this can be done following the usual rules for chain graphs. As an example, in Figure 1 a graphical two-level random intercept model and its moralized version are represented. Note that $\mathcal{U} \perp\!\!\!\perp Z_1$, like is manifest from part (a) of Figure 1, while from part (b) of Figure 1 it can be read that $\mathcal{U} \not\perp\!\!\!\perp Z_1 \mid (Y_{1j}, Y_{2j})$.

It is worth to note that if $\mathcal{U}_j \notin pa(Y_{ij}, Y_{i'j})$, then the individual graph is not necessary and can be misleading. In this case:

i. the individual nodes set $\mathcal{I}$ collapses into a single node and the multiple arrows originating from the Z to the individual nodes are collapsed in a single arrow (see example 4.3);

ii. the collapsed graph is Markov equivalent to the corresponding individual graph.

It is important to remark that the collapsed graph is still a chain graph, but it is no more an individual graph.

4. Examples

The following examples help to illustrate the graphs relative to multilevel models and their interpretation.

4.1. Two level null random-intercept model

Consider the two level null random-intercept model

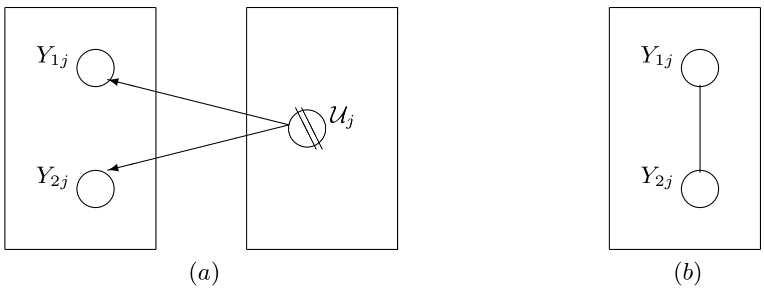
$$y_{ij} = \alpha_0 + \tau u_j + \varepsilon_{ij} \tag{5}$$

where y_{ij} is the response variable of the i-th individual of the j-th cluster, $i = 1, 2, \ldots, n_j$, $j = 1, 2, \ldots, J$, α_0 is the common mean, u_j are i.i.d random variables representing the j-th cluster deviation from the mean, while ε_{ij} are i.i.d. individual residuals. The graph of this model is reported in Figure 2. According to the graph, the joint probability distribution can be factorized as:

$$f(\mathbf{Y}_j, \mathcal{U}_j) = f(\mathbf{Y}_j \mid \mathcal{U}_j) f(\mathcal{U}_j) = \left[\prod_{i=1}^{n_j} f(y_{ij} \mid \mathcal{U}_j) \right] f(\mathcal{U}_j) \tag{6}$$

where $\mathbf{Y}_j = \{y_{1j}, \ldots, y_{n_j j}\}'$.

Fig. 2. A graphical multilevel null model: (a) conditioning on $\mathcal{U}_j$ and (b) marginalizing over it.



Looking at part (b) of Figure 2, it can be seen that marginalizing with respect to the grouping latent node $\mathcal{U}_j$ leads to a connection between its children nodes.

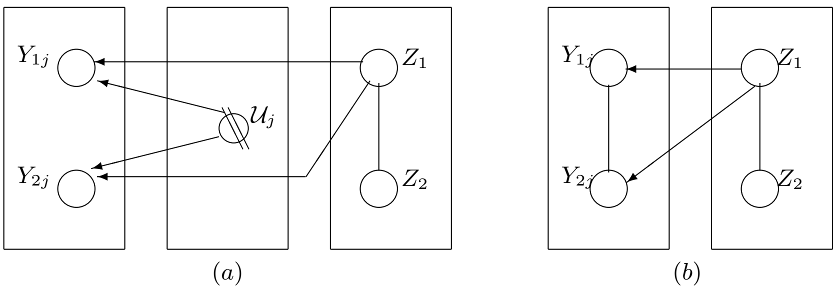
4.2. Two-level random-intercept model with covariates

The following model (7) corresponds to the classical two-level random intercept model with two explanatory variables:

$$y_{ij} = \alpha_0 + \beta_1 z_{1ij} + \beta_2 z_{2ij} + \tau u_j + \varepsilon_{ij} \tag{7}$$

Its factorization is that of equation (2), where $\mathbf{Z} = (Z_1, Z_2)'$. The chain graph corresponding to model (7) is represented in Figure 3, whenever β_2 is not significantly different from zero.

Fig. 3. A graphical multilevel model: (a) conditioning on grouping and (b) marginalizing over it.



It is worth to note that the LRT is not a valid test to compare two nested random-intercept models with a different number of explanatory variables because the latent variable $\mathcal{U}$ might include also the effect of the omitted explanatory variables. For instance, the null model of Figure 2 cannot be compared with LRT to the model with explanatory variables of Figure 3. In this case, a Wald test has to be preferred.

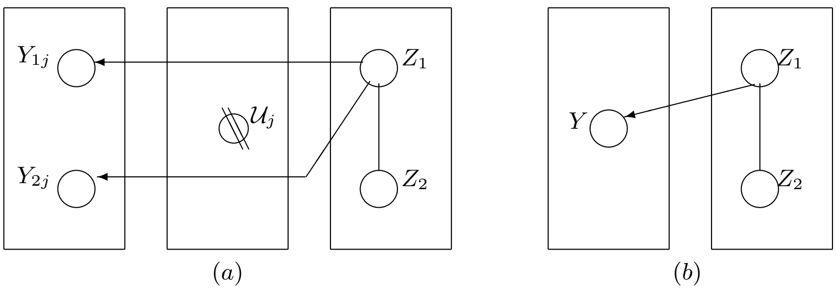
4.3. An ordinary regression model with covariates

The following is an example of a misleading individual graph. Consider the the model

$$y_{ij} = \alpha_0 + \beta_1 z_{1ij} + \beta_{21} z_{2ij} + \varepsilon_{ij} \tag{8}$$

The corresponding graph, reported in Figure 4, states that $Y_{1j} \perp\!\!\!\perp Y_{2j} \mid Z_1$, but actually it is $Y_{1j} \perp\!\!\!\perp Y_{2j}$. Moreover, the node $\mathcal{U}_j$ is a singleton and can be deleted from the graph without changing the conditional independencies set.

Fig. 4. An example of misleading graphical multilevel model (a) and its correct collapsed version (b).



Moreover, from the graph of Figure 4 it can be seen that $Y \perp\!\!\!\perp \mathcal{U} \mid Z$ and also $Y \perp\!\!\!\perp \mathcal{U}$, because $\mathcal{U} \perp\!\!\!\perp Z$.

Table 1. Graduates by job position (2000).

Code	*Job position*	*Frequency*	*Percent*
0	no job	249	8.02
1	temporary job	1458	46.97
2	stable job	1397	45.01

5. Application

The proposed models are used to analyze some of the data gathered by a telephone survey conducted, about two years after the degree, on the 2000's graduates of the University of Florence. The main goal is to determine factors influencing graduates job position. Moreover, we want to evaluate the degree programmes on the basis of the probability of finding a stable occupation of graduates. The pure response variable Y, *job position at the date of the interview*, is a polytomous variable (Table 5†), taking on value '0' if the graduate doesn't have a job, '1' if he/she has a temporary job and '2' if he/she has a stable job. Many observed and unobserved variables may influence the probability of finding a temporary or stable job. Variables can be ordered into a sequence of blocks, according to subject-matter knowledge or time. The unexplained variability at the course programme level is represented by a latent node. The variables definition and the blocks ordering is reported in Table 2. According to the sequence of blocks, the joint density function can be factorized as:

$$f(\mathbf{Y}_j, \mathcal{U}_j, \mathbf{Z}) = f(\mathbf{Y}_j, | \, \mathcal{U}_j, \mathbf{Z}) f(\mathcal{U}_j) f(\mathbf{Z})$$

where $f(\mathbf{Z}) = \prod_{m=1}^{6} f(\mathbf{Z}_m \mid \mathbf{Z}_{m-1}, \ldots, \mathbf{Z}_1)$, with $\mathbf{Z}_m$ denoting the vector of variables in the m-th block.

The estimation of a chain graph model is a difficult task, requiring many steps. The estimation strategy adopted in this work is based on fitting univariate appropriate regression models according to the dependent variable scale and to the recursive nature of the chain graph (Cox and Wermuth, 1996). To take account of the hierarchical structure of the data and of the multinomial nature of the pure response variable, a suitable graphical multilevel model for polytomous response (Skrondal and Rabe-Hesketh, 2003) is developed and fitted by means of maximum likelihood with adaptive Gaussian quadrature.

The dependence structure of the data is quite complex. Only 5 of the 15 explanatory variables have a direct effect on the job position, but only one (X_{51} = *professional training*) among them has no effect at all on the pure response variable. For example, the parents educational level has only an indirect effect on the pure response through the high school type and the high school rank. The introduction of a latent node, representing the course programme effect, is substantial: the likelihood ratio test comparing the models with and without the latent grouping effect is significant. The proportion of residual variance explained by the grouping is about 8.25%. To give an idea of usefulness of the proposed approach, Figure 5 represents a part of the resulting chain graph, where all undirected edges and only some arrows are drawn to highlight all direct and some indirect effects on the pure response. It is important to stress that graphical models are not a causal models: particularly, in this application, the analysis is conditional on (*a*) the choice to enroll at the university after the high school and (*b*) to not drop out during the university career.

†Only graduates having looked for a job have been considered.

Table 2. Variables definition and blocks ordering

Z	*Name*	*Description*	*Block*
11	father educational level	1-5	1 (pure explanatory)
12	mother educational level	1-5	
13	male	1 if male	
21	high school type	4 types	2 (intermediates)
22	high school rank	36-60	
23	regular career	1 if regular	
31	enrollment year	1(<1990)-5(>1995)	3 (intermediates)
32	short degree	1 if short degree	
33	school	11 schools	
41	examinations rank average	18-31	4 (intermediates)
42	duration index	0.91-4.86	
51	training	1 if training done	
52	honors	1 if graduate with honors	
53	age at the degree	22-51	5 (intermediates)
61	degree-interview distance	12-33 months	6 (intermediate)
u_j	course programme	56 course programmes	(latent node)
y_{ij}	job position	0 no, 1 temporary, 2 stable	7 (pure response)

6. Concluding remarks

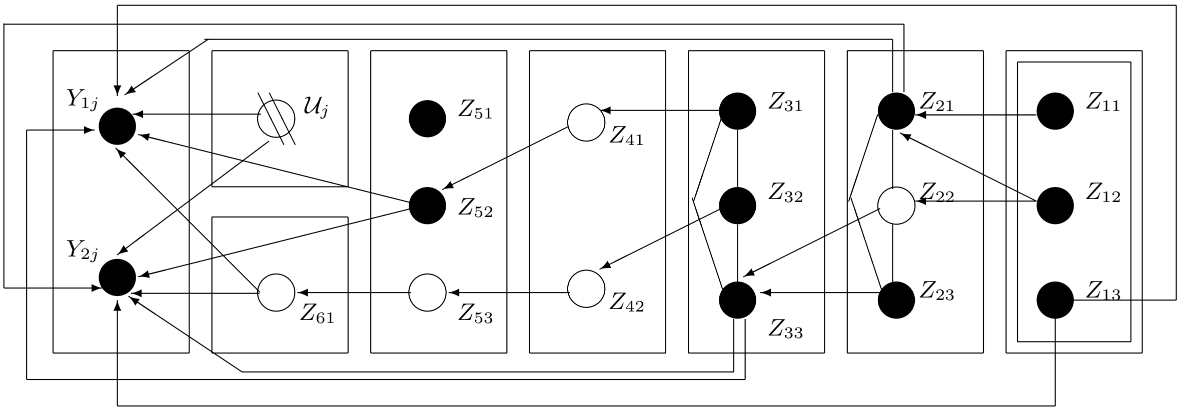
The work proposes a possible solution to extend graphical models for correlated data. Particularly, the paper focused on hierarchical data structures, considering two-level random intercept models. The proposed solution allows to use the existing chain graphs theory in a straightforward way.

This definition of a chain graph for multilevel models combines the potentialities of graph models with that of multilevel models. Moreover, it forces to make clear the conditional independence hypotheses underlying the multilevel model specification. Furthermore, the introduction of individual response nodes highlight that in multilevel models the unit of analysis is the cluster, while the response is multivariate. The estimation procedure used in the application could be improved further, relying on the graph properties.

The simply random-intercept model can be extended in many ways, in order to taking into account a more complex dependence structure of the responses, including random slopes, and a more complex hierarchical structure, allowing more that two-level of hierarchy, or cross-classified structures.

Acknowledgements Part of this work was presented at the *Celebrating Statistics: International conference in honour of Sir David Cox on the occasion of his 80th birthday*, Neuchâtel, Switzerland, 14-18 july 2004. We are grateful to Sir D.R. Cox for his constructive and valuable comments and suggestions. Financial support from MIUR research project COFIN2003 *Learning on graphical models and causal inference with special regard to socio-economic, genetic and forensic themes* (project no.2003138887-004) is gratefully acknowledged.

Fig. 5. Part of the resulting chain graph



References

Buntine, W. L. (1994). Operations for learning with graphical models. *Journal of Artificial Intelligence Research 2*, 159–225.

Cox, D. and N. Wermuth (1996). *Multivariate Dependencies. Models, analysis and interpretation*. London: Chapman and Hall.

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

Goldstein, H. (2003). *Multilevel Statistical Models 3rd edition*. London: Arnold.

Johnson, D. and J. A. Hoeting (2003). Random effects graphical models for multiple site sampling. *submitted to The Annals of Statistics*.

Lauritzen, S. (1996). *Graphical Models*. Oxford: Oxford Science Publications.

Lauritzen, S. and N. Wermuth (1989). Graphical models for the association between variables, some of which are qualitative and some quantitative. *Ann. Statist. 17*, 31–57.

Skrondal, A. and S. Rabe-Hesketh (2003). Multilevel logistic regression for polytomous data and rankings. *Psychometrika 68*, 267–287.

Snijders, T. and R. Bosker (1999). *Multilevel Analysis. An introduction to basic and advanced multilevel modelling*. London: Sage.

Whittaker, J. (1990). *Graphical Models in Applied Multivariate Statistics*. Wiley: New York.

Copyright © 2004
Anna Gottard, Carla Rampichini