Binary models of marginal independence: a comparison of different approaches

Monia Lupparelli and Luca La Rocca

1 Background and notation

Graphical models of marginal independence, introduced by Cox & Wermuth [1] with the name of covariance graph models, are also known in the literature as bidirected graph models following Richardson [6]. These models are of special interest because they belong to the families of ancestral and summary graphs; see Wermuth & Sadeghi [8] for a review. Let $\mathscr{G} = (V, E)$ be a bidirected graph defined by a finite set $V = \{1, \dots, p\}$ of nodes and a symmetric set of edges $E \subseteq V \times V$ drawn as bidirected. Under the *pairwise Markov property*, for a random vector $X_V = (X_v)_{v \in V}$, a missing edge between a pair of nodes $(u, v) \notin E$ corresponds to the marginal independence $X_u \perp\!\!\!\perp X_v$. The set of all independencies encoded by $\mathscr{G}$ can be derived using the *connected set Markov property*: given any disconnected set $D \subseteq V$ of nodes in $\mathscr{G}$, the vectors associated to its connected components $X_{C_1}, \dots, X_{C_r}$ are mutually independent; see Fig. 1 for an illustration and Richardson [6] for technical details.

In this paper we consider a vector X_V of binary variables following a cross-classified Bernoulli distribution on $\{0,1\}^p$ with joint probability parameter π. As π varies over the simplex, this defines an exponential family whose mean parameter (also called Möbius parameter) is the vector $\mu = (\mu_A)_{A \subseteq V}$, where $\mu_A = P(X_A = 1)$. These probabilities are obtained as linear functions of the joint probabilities, i.e., $\mu = Z\pi$, where Z is a square matrix and is the inverse of the Möbius matrix M such that $\pi = M\mu$; see Drton & Richardson [2] for technical details.

Choosing a parameterization for graphical models of marginal independence is a crucial task mainly for two reasons: a) pairwise independencies do not imply higher

Monia Lupparelli
University of Bologna, Department of Statistical Sciences, Via delle Belle Arti 41, 40126 Bologna, Italy, e-mail: monia.lupparelli@unibo.it

Luca La Rocca
University of Modena and Reggio Emilia, Department of Communications and Economics, Viale Allegri 9, 42121 Reggio Emilia, Italy, e-mail: luca.larocca@unimore.it

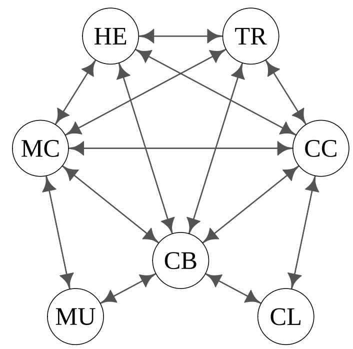

Fig. 1 Bidirected graph for seven variables. Given the disconnected sets $\{CL, TR, HE, MU, MC\}$ and $\{MU, TR, HE, CL, CC\}$, the independence model under the connected set Markov property is $CL \perp\!\!\!\perp \{TR, HE, MU, MC\}$ and $MU \perp\!\!\!\perp \{TR, HE, CL, CC\}$.

order independencies; b) it should be possible to implement additional constraints for the selection of parsimonious submodels. More in detail, desiderable properties of a convenient parameterization include: i) parameter interpretability, ii) upward compatibility (invariance of parameters with respect to marginalization), iii) closed-form representation of the joint probabilities, iv) availability of efficient procedures for maximum likelihood estimation.

Drton & Richardson [2] defined the independencies of a binary graphical model of marginal independence by means of multiplicative constraints on the mean parameter, i.e., letting $\mu_D = \mu_{C_1} \times \ldots \times \mu_{C_r}$ for every disconnected set $D \subseteq V$ of the graph. For instance, the graph in Fig. 1 implies $\mu_{\{MU,CL,CC\}} = \mu_{\{MU\}} \times \mu_{\{CL,CC\}}$. This parameterization satisfies the upward compatibility property, allows one to write the likelihood function in closed form, and is amenable to efficient maximum likelihood estimation via the *iterative conditional fitting* algorithm of Drton & Richardson [2]. However, the mean parameter does not have an immediate interpretation in terms of dependence/independence and, since non-linear constraints are used, parsimonious submodels cannot be easily specified.

Successively, Lupparelli *et al.* [5] considered the multivariate logistic parameter $\eta = (\eta_A)_{A \subseteq V}$ of Glonek & McCullagh [4], defined by the link function $\eta = C \log L\pi$ for suitable rectangular matrices C and L, and specified a bidirected graph model by means of the linear constraints $\eta_D = 0$, where D varies in the disconnected subsets of V. For instance, the graph in Fig. 1 implies $\eta_{\{MU,CL,CC\}} = 0$. This parameterization satisfies the upward compatibility property, its parameters are interpretable measures of association and, since the independence model corresponds to zero interactions, parsimonious models can be easily specified by setting further interactions to zero. Nevertheless, the likelihood function cannot be written in closed form as a function of η, because the inverse mapping $\eta \mapsto \pi$ is not analytically available, and classical constrained likelihood maximization, although straightforward, is hampered by the large dimension of the rectangular matrices C and L.

More recently, Roverato *et al.* [7] studied a parameterization based on a log-linear expansion of the parameter μ. The next section shows how this parameterization can be used to define parsimonious bidirected graph submodels.

2 The log-mean linear parameterization

The LML (log-mean linear) parameter $\gamma = (\gamma_A)_{A \subseteq V}$ is defined by the mapping

$$\gamma = M^\top \log \mu \tag{1}$$

and its elements are measures of association in marginal distributions, e.g., the first, second and third order interactions are $\gamma_{\{v\}} = \log \mu_{\{v\}}$, $\gamma_{\{u,v\}} = \log \frac{\mu_{\{u,v\}}}{\mu_{\{u\}}\mu_{\{v\}}}$, $\gamma_{\{u,v,w\}} = \log \frac{\mu_{\{u,v,w\}}\mu_{\{u\}}\mu_{\{v\}}\mu_{\{w\}}}{\mu_{\{u,v\}}\mu_{\{u,w\}}\mu_{\{v,w\}}}$, with $u,v,w \in V$. Hence, the LML parameterization satisfies the upward compatibility, and it allows to write the joint probabilities in closed form, because the mapping $\gamma \mapsto \pi$ is analytically computed as $\pi = M \exp(Z^\top \gamma)$. Although the link function defining the LML parameter is of the form of the link function defining the multivariate logistic transform, in the latter the contrast and marginalization matrices corresponding to M and Z are rectangular of size $t \times 2^p$ with $t \gg 2^p$. Therefore, classical constrained likelihood maximization is more efficient when constraints are expressed through γ.

Roverato *et al.* [7] proved that, given two disjoint subsets of variables $A, B \subseteq V$, the marginal independence $X_A \perp\!\!\!\perp X_B$ holds if and only if $\gamma_{a \cup b} = 0$ for every $a \subseteq A$ and $b \subseteq B$ with $a, b \neq \emptyset$. Then, a binary bidirected graph model can be defined by letting $\gamma_D = 0$ for every disconnected set $D \subseteq V$ of the graph. For instance, the graph in Fig. 1 implies $\gamma_{\{MU,CL,CC\}} = 0$. Additional zero constraints on higher order interactions can be introduced to obtain a parsimonious submodel, like with the multivariate logistic parameter η, but a more interpretable simplification option is also available: given three disjoint subsets of variables $A, B, C \subseteq V$, the *code-specific* subpopulation independence $X_A \perp\!\!\!\perp X_B | \{X_C = 1\}$ can be specified by a set of linear constraints on γ; see Cor. 8 in Roverato *et al.* [7].

One should observe that distinct codings (labelling of variable values as 0 and 1) will result in distinct parsimonious submodels. When no coding has a special status, Roverato *et al.* [7] suggest to adopt the *maximal count coding*, which labels variable values so that the cell 1_V of the table contains the largest count. This coding allows to test code-specific independencies in conditional distributions with more observations, which results in increased inferential power. Roverato *et al.* [7] also consider a medical application where a special coding is of interest.

3 An illustrative example

We illustrate how the LML parameterization can be used to select bidirected graph submodels for a data set originally analyzed by Drton & Richardson [2]. These authors used bidirected graph models to examine seven questions relating to trust and social institutions, taken from the U.S. General Social Survey between 1975 and 1994. The seven binary variables collected over 13,486 individuals are: TR (Can most people be trusted?), HE (Do you think most people are usually helpful?), *MU* and *MC* (Are you a member of a labour union / church?), *CL* and *CC* and *CB* (Do you have confidence in law / religious congregations / business?). Using the mean parameterization, and a backward stepwise selection procedure, Drton & Richardson [2] found that the bidirected graph model in Fig. 1 achieves a good fit with deviance $\chi^2_{(26)} = 32.67$ ($p = 0.172$, $AIC = -19.3$). The great number of parameters suggests that a more parsimonious model could fit the data equally well, but the mean parameterization does not help with this kind of simplification.

We analyze the same data set using the LML parameterization to specify the model in Fig. 1 and adopting a procedure based on two steps for the selection of a parsimonious submodel. As a first step, we search the data for code-specific independencies. Leaving the choice of an optimal procedure for this search as a topic for future research, we focus on a simple heuristic strategy that could be especially useful in presence of large tables: we focus on the cliques of the graph in Fig. 1, because they have a good potential for simplification (not having been simplified at all in the model). For each clique Cl_i, $i = 1,2,3$, we search for code-specific subpopulation independencies in the partial distribution $X_{Cl_i}|\{X_{V \setminus Cl_i} = 1\}$, and find that the variables $\{MU, MC, CB\}$ are mutually independent in the subpopulation defined by $\{TR = 1, HE = 1, CL = 1, CC = 1\}$. Including this code-specific independence in the model, we achieve a fit with deviance $\chi^2_{(30)} = 38.51$ ($p = 0.137$, $AIC = -21.5$). As a second step, we further simplify the model by testing the significance (based on Wald's test) of the remaining higher order LML interactions. We finally obtain a bidirected graph submodel with deviance $\chi^2_{(92)} = 99.50$ ($p = 0.279$, $AIC = -84.5$).

The same data set has been also analyzed by Evans & Richardson [3] using the multivariate logistic parameterization. These authors find a parsimonious submodel by vanishing all 4-5-6 and 7-way interactions, together with those 3-way interactions that are not significant in their selection procedure. They obtain a bidirected graph submodel with deviance $\chi^2_{(102)} = 111.48$ ($p = 0.245$, $AIC = -92.5$).

References

1. Cox, D.R., Wermuth, N.: Linear dependencies represented by chain graphs. Statist. Sci. **30**, 145–157 (1993)
2. Drton, M., Richardson, T.S.: Binary models for marginal independence. J. R. Stat. Soc. Ser. B Stat. Methodol. **70**, 287–309 (2008)
3. Evans, R.J., Richardson, T.S.: Marginal log-linear parameters for graphical Markov models. Cornell University arXiv:1105.6075v2[stat.ME] (2012)
4. Glonek, G.J.N., McCullagh, P.: Multivariate logistic models. J. R. Stat. Soc. Ser. B Methodol. **57**, 533–546 (1995)
5. Lupparelli, M., Marchetti, G.M., Bergsma, W.P.: Parameterizations and fitting of bi-directed graph models to categorical data. Scand. J. Stat. **36**, 559–576 (2009)
6. Richardson, T.S.: Markov property for acyclic directed mixed graphs. Scand. J. Stat. **30**, 145–157 (2003)
7. Roverato, A., Lupparelli, M., La Rocca, L.: Log-mean linear models for binary data. Cornell University arXiv:1109.6239v2[stat.ME] (2012)
8. Wermuth, N., Sadeghi, K.: Sequences of regressions and their independences. TEST (2012) doi: 10.1007/s11749-012-0290-6