

Triangular systems for symmetric binary variables

Nanny Wermuth*

Mathematical Statistics, Chalmers/Gothenburg University, Sweden
e-mail: wermuth@chalmers.se

Giovanni M. Marchetti†

Department of Statistics, University of Florence, Italy
e-mail: giovanni.marchetti@ds.unifi.it

and

D.R. Cox

Nuffield College, Oxford, UK
e-mail: david.cox@nuffield.ox.ac.uk

Abstract: We introduce and study distributions of sets of binary variables that are symmetric, that is each has equally probable levels. The joint distribution of these special types of binary variables, if generated by a recursive process of linear main effects is essentially parametrized in terms of marginal correlations. This contrasts with the log-linear formulation of joint probabilities in which parameters measure conditional associations given all remaining variables. The new formulation permits useful comparisons of different types of graphical Markov models and leads to a close approximation of Gaussian orthant probabilities.

AMS 2000 subject classifications: Primary 62E10; secondary 62H17, 62H20.

Keywords and phrases: Graphical Markov models, linear in probability models, log-linear models, recursive generating processes.

Received June 2009.

Contents

1	Introduction	933
2	A linear triangular system for symmetric binary variables	934
	2.1 Definition and background	934
	2.2 Some basic properties	937
3	A logit triangular system for symmetric binary variables	939
4	Orthant probabilities	940
5	Types of graphical Markov models	942

*Supported in part by the Swedish Research Society via the Gothenburg Stochastic Center and by the Swedish Strategic Fund via the Gothenburg Mathematical Modelling Center.

†Supported in part by MIUR, Rome, under the project PRIN 2007 XECZ7L002.

5.1	Markov chains, covariance and concentration graph models . . .	943
5.2	A concentration chain	944
5.3	A covariance chain	945
5.4	Seemingly unrelated regression and multivariate regression chains	946
5.5	Mutual conditional independence of A_1, A_2, A_3, A_4 given A_5 . . .	948
5.6	Mutual independence of A_2, A_3, A_4, A_5 , common response A_1 . .	948
A	Proof of equations (2.3), (2.9)	949
B	Effect parameter expansions	951
C	A Gaussian and a symmetric binary correlation matrix	952
	Acknowledgements	953
	References	953

1. Introduction

The explicit study of dependences and independences among a set of random variables is easiest for the multivariate Gaussian distribution since there are only linear relations and no interactive effects. Its key features include preservation of functional form under marginalization and conditioning and the use of triangular systems in only main effects to model recursive data generating processes. The latter in particular are strongly linked to a matrix formulation, itself in turn connected with the representation of independences by graphs; see for instance [Marchetti and Wermuth \(2009\)](#).

In the present paper, we give similar results for binary variables that are symmetric. Such binary variables may but need not result by median-dichotomizing continuous variables. We denote the two equally likely levels by -1 and 1 , so that all variables are standardized to have marginally zero mean and unit variance. Then we give linear representations of probabilities involving parameters of which some can be interpreted as covariances and equivalently as correlations and others as linear regression coefficients. A strong parallel with Gaussian theory is established and used to compare different types of graphical Markov models, including seemingly unrelated regression in binary variables.

Two related approaches in logarithms of probabilities are studied briefly. For three variables, but not for more, the linear in probability representation is equivalent to a joint log linear model and to a logit triangular version. It is shown also that when all pairs of variables are equally correlated, the linear representation is remarkably close to that formed by median dichotomy of a latent multivariate Gaussian distribution.

Whenever measurement tools are questionnaires, median-dichotomized variables are widely used. Especially for visual analogue scales, the resulting symmetric binary variables often lead to appropriate data summaries. Symmetric binary variables may also result by design, for instance, when in medical studies controls are matched to an equal number of cases. But, since linear triangular systems of main effects for symmetric binary variables appear to have not been studied before, they are so far also absent in applications of graphical Markov models, see for instance [Green, Hjort and Richardson \(2003\)](#), and of discrete variable models; see e.g. [Bergsma, Croon and Hagenaars \(2009\)](#).

2. A linear triangular system for symmetric binary variables

2.1. Definition and background

We consider p binary variables A_s , $s = 1, \dots, p$. Variable A_s has two equally probable levels a_s , coded as 1 for success and -1 for failure. For up to four variables, we sometimes denote the variables by $A = A_1, B = A_2, C = A_3, D = A_4$ and the levels, respectively, by i, j, k, l .

Joint and conditional probabilities are written in a condensed form, for instance for variables A, B, C

$$\pi_{ijk} = \Pr(A = i, B = j, C = k), \quad \pi_{i|jk} = \pi_{ijk} / \sum_i \pi_{ijk}.$$

Sometimes, the notation is supplemented by superscripts indicating which variables correspond to the given levels, such as in

$$\pi_{111}^{ABC} = \Pr(A = 1, B = 1, C = 1), \quad \pi_{111}^{BCD} = \Pr(B = 1 | C = 1, D = 1).$$

The linear triangular system of exclusively main effects in four variables is

$$\begin{aligned} \pi_{i|jkl}^{A_1|A_2A_3A_4} &= \frac{1}{2} \{1 + i(\eta_{12}j + \eta_{13}k + \eta_{14}l)\} \\ \pi_{j|kl}^{A_2|A_3A_4} &= \frac{1}{2} \{1 + j(\eta_{23}k + \eta_{24}l)\} \\ \pi_{k|l}^{A_3|A_4} &= \frac{1}{2} \{1 + k(\eta_{34}l)\} \\ \pi_l^{A_4} &= \frac{1}{2}, \end{aligned} \tag{2.1}$$

where, with ρ_{st} denoting the correlation coefficient of A_s, A_t for $s < t$, with \mathbf{P} the correlation matrix and with $r(s) = \{s + 1, \dots, p\}$,

$$\begin{aligned} \eta_{34} &= \rho_{34} \\ (\eta_{23} \ \eta_{24}) &= (\rho_{23} \ \rho_{24}) \mathbf{P}_{r(2),r(2)}^{-1} \\ (\eta_{12} \ \eta_{13} \ \eta_{14}) &= (\rho_{12} \ \rho_{13} \ \rho_{14}) \mathbf{P}_{r(1),r(1)}^{-1}. \end{aligned} \tag{2.2}$$

This form of the η 's as linear regression coefficients in a conventional sense generalizes directly to p variables and stems from the close connection for binary variables between probabilities and expectations. For example, on multiplying the first equation in (2.1) by i , using $i^2 = 1$, and summing over i

$$E(A_1 | A_2 = j, A_3 = k, A_4 = l) = \eta_{12}j + \eta_{13}k + \eta_{14}l.$$

The joint distribution of the four variables has from (2.1) the form

$$\pi_{ijkl}^{A_1A_2A_3A_4} = \frac{1}{16} (1 + \rho_{12}ij + \rho_{13}ik + \rho_{14}il + \rho_{23}jk + \rho_{24}jl + \rho_{34}kl + \eta^{1234}ijkl), \tag{2.3}$$

with

$$\eta^{1234} = (\eta_{12} \ \eta_{13} \ \eta_{14})(\rho_{34} \ \rho_{24} \ \rho_{23})^T. \tag{2.4}$$

After multiplying (2.3) for instance by ij and summing over i, j, k, l , the correlation coefficient of A_1, A_2 results as

$$E(A_1A_2) = (\pi_{11}^{A_1A_2} + \pi_{-1-1}^{A_1A_2}) - (\pi_{-11}^{A_1A_2} + \pi_{1-1}^{A_1A_2}) = \rho_{12}. \tag{2.5}$$

In Appendix A, equations (2.3) and (2.4) are derived from (2.1) by multiplying the conditional probabilities in a way that generalizes to p variables. Given the joint probabilities, the parameters of the joint distribution may be computed by an effect expansion of π^{A_1, \dots, A_p} , a vector of the joint probabilities; see Appendix B.

Joint distributions generated via linear triangular systems of main effects for symmetric binary variables relate for $p \geq 4$ to the Bahadur expansion (1961) of general densities as given by Streitberg (1990), who proved existence and uniqueness without reference to any stepwise generating processes.

Here, we see that these parameters relate also directly to the triangular decomposition of the concentration matrix, which is for variables standardized to have zero means and unit variances, the inverse of the correlation matrix \mathbf{P} . This triangular decomposition is denoted here by $(\mathbf{H}, \mathbf{\Delta}^{-1})$ and gives

$$\mathbf{P}^{-1} = \mathbf{H}^T \mathbf{\Delta}^{-1} \mathbf{H},$$

where \mathbf{H} is a unit upper-triangular matrix, which has ones along the diagonal and zero entries below the diagonal, and $\mathbf{\Delta}$ is a diagonal matrix with linear regression variances along the diagonal. The decomposition exists for every positive definite \mathbf{P} and is unique for a given complete ordering $(1, \dots, p)$.

In particular, for the linear triangular systems treated here

$$\mathbf{H}_{st} = -\eta_{st}, \quad \mathbf{H}_{s,r(s)} = -\mathbf{P}_{s,r(s)} \mathbf{P}_{r(s),r(s)}^{-1} \text{ for } s < t \in \{1, \dots, p\}, \tag{2.6}$$

and for four variables, the diagonal elements of $\mathbf{\Delta}$ are $\delta_{44} = 1, \delta_{33} = 1 - \rho_{34}^2,$

$$\delta_{22} = 1 - (\eta_{23} \ \eta_{24})(\rho_{23} \ \rho_{34})^T, \quad \delta_{11} = 1 - (\eta_{12} \ \eta_{13} \ \eta_{14})(\rho_{12} \ \rho_{13}, \ \rho_{14})^T.$$

For $p > 2$, the δ_{ss} are not the conditional variances of the binary variables, $\text{var}(A_s|A_{r(s)})$, but their expected values with respect to $A_{r(s)}$.

For p symmetric binary variables, the triangular system in linear main effect parameters defines the joint distribution as a product of the univariate conditional probabilities, where $\pi_1^{A_p} = \frac{1}{2}$ and

$$\pi_{a_s|a_{r(s)}}^{A_s|A_{r(s)}} = \frac{1}{2}(1 + a_s\{a_{s+1}\eta_{s,s+1} + \dots + a_p\eta_{sp}\}), \tag{2.7}$$

with $(\eta_{s,s+1}, \dots, \eta_{sp}) = -\mathbf{H}_{s,r(s)}$ as in (2.6) and

$$E(A_s|A_{r(s)}) = a_{s+1}\eta_{s,s+1} + \dots + a_p\eta_{sp}$$

$$\text{var}(A_s|A_{r(s)}) = 1 - (a_{s+1}\eta_{s,s+1} + \dots + a_p\eta_{sp})^2.$$

The main effects, η_{st} for $t > s$, and the conditional expectations are fully determined by the marginal correlations via the triangular decomposition of \mathbf{P}^{-1} .

The triangular decomposition of a concentration matrix may be obtained as a byproduct when applying the operator of partial inversion (Wermuth, Wiedenbeck and Cox, 2006) repeatedly to the covariance matrix or, for standardized variables, to the correlation matrix. If (a, b) denotes any ordered split of $\{1, 2, \dots, p\}$, and e.g. $\mathbf{P}_{ab} = (\mathbf{P})_{a,b}$, then

$$\text{inv}_b \mathbf{P} = \begin{pmatrix} \mathbf{P}_{aa|b} & \mathbf{P}_{ab} \mathbf{P}_{bb}^{-1} \\ \sim & \mathbf{P}_{bb}^{-1} \end{pmatrix}, \quad \mathbf{P}_{aa|b} = \mathbf{P}_{ab} - \mathbf{P}_{ab} \mathbf{P}_{bb}^{-1} \mathbf{P}_{ba},$$

where \sim denotes entries which are symmetric up to the sign.

Linear triangular systems for continuous responses have been studied in econometrics under the name of linear recursive equations with uncorrelated residuals; see Goldberger (1964) and in genetics as path analysis models; see Wright (1934). In that context, the regression coefficients have been called least-squares regression coefficients in the population; see Cram er (1946), p. 302. To cover path analysis and similar models for general types of densities, the name triangular system was introduced; see Wermuth and Cox (2004).

For general types of binary variables, the models most closely analogous to path analysis are triangular systems of logit regressions; see Goodman (1973) and here Section 3. The vanishing of a logit regression coefficient indicates always conditional independence given the remaining directly explanatory variables; see also Fienberg (2007). It is in the special case of symmetric binary variables considered here, that the vanishing of a main effect in probabilities in (2.7) coincides with the vanishing of a corresponding logit regression coefficient.

The symmetry of the binary variables leads to missing odd-order terms in the effect expansion of π^{A_1, \dots, A_p} . The presence of only even-order moments is also a feature of joint Gaussian distributions so that Gaussian probabilities, such as all variables being jointly positive, are well approximated by the corresponding probabilities for the binary variables; see Section 4.

By the relation of the linear main effects to the triangular decomposition of the correlation matrix, independence constraints on the joint probability distribution generated by a linear triangular system (2.7) in symmetric binary variables have similar implications as in a joint Gaussian distributions. For instance for three disjoint subsets a, b, c of $\{1, \dots, p\}$, the conditional independence $A_a \perp\!\!\!\perp A_b | A_c$, often written compactly as $a \perp\!\!\!\perp b | c$, imposes the following constraint on the correlation matrix. The submatrix \mathbf{P}_{ab} is replaced by \mathbf{P}_{ab}^* , where

$$\mathbf{P}_{ab}^* = \mathbf{P}_{ac} \mathbf{P}_{cc}^{-1} \mathbf{P}_{cb}, \tag{2.8}$$

all other entries are left unchanged. This is a unique modification of \mathbf{P} which maximizes the determinant; see Dempster (1972). The effect of this modification of a correlation matrix is that $\mathbf{P}_{ab|c}^* = 0$. For p variables having a multivariate Gaussian distribution, this is known to be equivalent to $a \perp\!\!\!\perp b | c$. For joint distributions of four or more symmetric binary variables generated as in (2.7), it is only a necessary condition for $a \perp\!\!\!\perp b | c$; see Appendix C.

In the linear triangular systems (2.7), mutual conditional independence is captured by zero parameters of this stepwise generating process while marginal

mutual independence is reflected in sets of zero parameters of the joint distribution. For $p = 4$, for instance $A_1 \perp\!\!\!\perp A_2 \perp\!\!\!\perp A_3 | A_4$ holds if $0 = \eta_{12} = \eta_{13} = \eta_{23}$ and $A_2 \perp\!\!\!\perp A_3 \perp\!\!\!\perp A_4$ holds if $0 = \rho_{23} = \rho_{24} = \rho_{34}$. Likelihood ratio tests apply just as for general categorical variables, but maximum-likelihood estimation is often simpler than for general binary variables. These features makes the family attractive for illustrating basic distinctions and similarities between various types of graphical Markov models, as studied for discrete variables by [Wermuth \(1976\)](#), [Darroch, Lauritzen and Speed \(1980\)](#), [Wermuth and Cox \(2004\)](#), [Drton \(2009\)](#), [Marchetti and Lupporelli \(2009\)](#); see Section 5.

2.2. Some basic properties

[Streitberg \(1999\)](#) stresses that estimation in constrained Bahadur expansions is undeveloped and that a truncated Bahadur expansion may be negative. The closed form (2.4) of the four-factor interaction term makes the first statement transparent for joint distributions generated by (2.7).

A linear four-factor effect can be zero if and only if some of the marginal or conditional linear regression coefficients are zero as well. For instance, when $\eta_{12} = 0$ in (2.1), and all other parameters in this process generating the joint distribution are nonzero, then the linear four-factor effect does not vanish unless $0 = \rho_{23} = \rho_{24}$ or $\eta_{14}\rho_{23} = -\eta_{13}\rho_{24}$.

Marginal independences may be implied by independence constraints on (2.7), or result by parametric cancellations, which are very special parametric constellations, for instance $0 = \rho_{23}$ if $\eta_{23} = -\eta_{24}\rho_{34}$; see Appendix A. For examples of parametric cancellations in joint Gaussian distributions see [Wermuth and Cox \(1998\)](#).

An important feature of (2.7) is that the symmetry in all margins, $\frac{1}{2} = \pi_1 = \pi_{-1}$, carries over to the joint probabilities, so that for instance for $p = 4$

$$\pi_{1111} = \pi_{-1-1-1-1}, \quad \pi_{-1111} = \pi_{1-1-1-1}, \dots, \quad \pi_{-1-1-11} = \pi_{111-1}.$$

That is, the probabilities of any given pattern of level combinations and that of the reversed pattern are the same.

For $p \geq 4$, every marginal trivariate distribution of (2.7) is of the same form as the joint probability π_{jkl}^{BCD} in (2.1), that is

$$\pi_{jkl}^{BCD} = \frac{1}{8}(1 + \rho_{23}jk + \rho_{24}jl + \rho_{34}kl). \tag{2.9}$$

Such distributions have in general four distinct probabilities, since

$$\pi_{111} = \pi_{-1-1-1}, \quad \pi_{-111} = \pi_{1-1-1}, \quad \pi_{1-11} = \pi_{-11-1}, \quad \pi_{-1-11} = \pi_{11-1}.$$

Sets of marginal correlations that give valid trivariate distributions (2.9) satisfy

$$\begin{aligned} \rho_{23} + \rho_{24} + \rho_{34} &\geq -1, & \rho_{23} - \rho_{24} - \rho_{34} &\geq -1, \\ -\rho_{23} + \rho_{24} - \rho_{34} &\geq -1, & -\rho_{23} - \rho_{24} + \rho_{34} &\geq -1 \end{aligned}$$

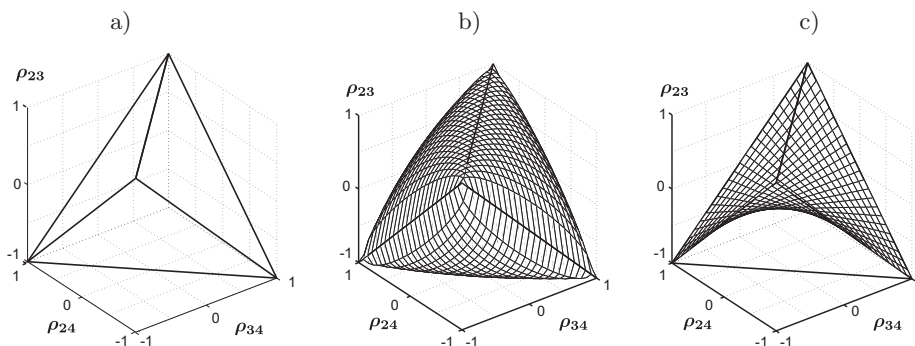


FIG 1. a) The simplex of sets of correlations for a valid density of B, C, D ; b) curved boundaries for correlations satisfying just positive-definiteness of \mathbf{P} ; c) surface of points excluded when requesting absence of parametric cancellation for pair (C, D) .

so that they belong to the simplex of Figure 1a), defined by the convex hull of the points $V_1 = (1, -1, -1), V_2 = (1, 1, 1), V_3 = (-1, -1, 1), V_4 = (-1, 1, -1)$. Within the curved boundaries of Figure 1b) are sets of correlations constrained only by the positive definiteness of the 3×3 correlation matrix. The maximal Euclidean distance of these boundaries to the simplex is $1/\sqrt{12}$. Figure 1c) shows the surface obtained by avoiding parametric cancellation for pair (C, D) in the distribution of B, C, D generated by (2.1) with nonvanishing $\eta_{23}, \eta_{24}, \rho_{34}$. That is, when $\rho_{34} = \rho_{23}\rho_{24}$ were to be explicitly excluded.

With only linear two-factor effects in (2.9), there is in this special trivariate family of symmetric binary variables no additive interaction as defined for three-dimensional contingency tables by Lazarsfeld (1961), or Lancaster (1969); see also Streitberg (1990).

In fact, it also has no multiplicative interaction. To see this, we look at odds for success, say level 1 of variable A_2 , compared to failure given its two explanatory variables A_3, A_4 , denoted by $\text{ods}_{A_2|kl} = \pi_{1|kl}^{A_2 A_3 A_4} / \pi_{-1|kl}^{A_2 A_3 A_4}$. There are two odds reversals in (2.9) with

$$\text{ods}_{A_2|11} = 1/\text{ods}_{A_2|-1,-1}, \quad \text{ods}_{A_2|-11} = 1/\text{ods}_{A_2|1-1}.$$

Therefore, the conditional odds ratios of success to failure, abbreviated here by odr, coincide at the two levels of the third variable, where

$$\text{odr}(A_2 A_3 | A_4 = l) = \frac{\pi_{11l} \pi_{-1-1l}}{\pi_{-11l} \pi_{1-1l}} = \text{ods}_{A_2|1l} / \text{ods}_{A_2|-1l}.$$

Equal odds-ratios at both levels of A_4 are equivalent to a zero three-factor effect in $\log \pi_{jkl}^{A_2 A_3 A_4}$; the 2^3 table is said to be without a multiplicative interaction. If these two odds-ratios are in addition equal to one then $2 \perp\!\!\!\perp 3 | 4$

Multiplicative interaction was defined for three-dimensional contingency tables by Bartlett (1935) and extended later for measuring associations in general log-linear models; see for instance Fienberg (2007).

For general trivariate log-linear models, expected counts have been defined as $m_{ijk} = n\pi_{ijk}$, where $n = \sum n_{ijk}$ is the total number of observed counts. For such log-linear models without a multiplicative interaction, maximum-likelihood estimates are obtained by equating the three marginal two-way tables of m_{ijk} to the three marginal two-way tables of n_{ijk} ; see Birch (1963). An iterative algorithm is needed to solve the equations. One such algorithm is iterative proportional fitting, for which Darroch and Ratcliff (1972) proved convergence.

By contrast, for symmetric binary variables, closed form estimates are available. The set of minimal sufficient statistics for model (2.9) consists of observed sums of counts, sums for two successes and two failures, and sums for the mixture of success and failure

$$n_{11}^{A_s A_t} + n_{-1-1}^{A_s A_t}, \quad n_{-11}^{A_s A_t} + n_{1-1}^{A_s A_t}.$$

These lead to closed form maximum-likelihood estimates of ρ_{st} , in terms of differences of the cross sums

$$\hat{\rho}_{st} = \{(n_{11}^{A_s A_t} + n_{-1-1}^{A_s A_t}) - (n_{-11}^{A_s A_t} + n_{1-1}^{A_s A_t})\}/n, \tag{2.10}$$

and to closed form estimates of the marginal odds-ratios by the one-to-one relations

$$\text{odr}(A_s A_t) = \{(1 + \rho_{st})/(1 - \rho_{st})\}^2, \quad \rho_{st} = \tanh\{\frac{1}{4} \log \text{odr}(A_s A_t)\}. \tag{2.11}$$

For $p > 3$, the method of moments yields with (2.10) and (2.5) all elements of an unconstrained correlation matrix of the binary variables. In this paper we shall not discuss inference for models with independence constraints. Instead, we describe and compare examples of models in different types of model classes.

Equal odds-ratios for any variable pair, at the fixed level combinations of all remaining variables, is equivalent for this variable pair having zero parameters of all orders higher than two in the log-linear model for A_1, \dots, A_p . Log-linear models with this property satisfied for all variable pairs have been studied and named binary quadratic exponential distributions by Cox and Wermuth (1994).

Such joint distributions are in general determined by sets of full two-way marginal tables and, for the subclass of symmetric binary variables considered here, by the sets of cross sums defining also the correlation coefficients; see (2.5). Nevertheless for $p > 3$, these distributions are in general different from those generated by a linear triangular system (2.7) as well as from those generated by a logit triangular system; see the next section.

3. A logit triangular system for symmetric binary variables

A logit is the logarithm of the odds, for instance,

$$\text{logit } \pi_{1|l}^{A_3|A_4} = \log \text{ods}_{A_3|l}, \quad \text{logit } \pi_{1|kl}^{A_2|A_3 A_4} = \log \text{ods}_{A_2|kl}.$$

The logit triangular system of main effects in four symmetric binary variables is

$$\begin{aligned}
 \text{logit } \pi_{1|jkl}^{A_1|A_2A_3A_4} &= \theta_{12}j + \theta_{13}k + \theta_{14}l \\
 \text{logit } \pi_{1|kl}^{A_2|A_3A_4} &= \theta_{23}k + \theta_{24}l \\
 \text{logit } \pi_{1|l}^{A_3|A_4} &= \theta_{34}l \\
 \text{logit } \pi_1^{A_4} &= 0,
 \end{aligned}
 \tag{3.1}$$

where

$$\begin{aligned}
 \exp(\theta_{34}) &= \frac{1}{2}\text{odr}(A_3A_4), \quad \exp(\theta_{2s}) = \frac{1}{2}\text{odr}(A_2A_s|A_t = a_t) \text{ for } s \neq t \in r(2), \\
 \exp(\theta_{1s}) &= \frac{1}{2}\text{odr}(A_1A_s|A_t = a_t, A_u = a_u) \text{ for } s \neq t \neq u \in r(1).
 \end{aligned}$$

After exponentiating, each given logit equation can be solved for the corresponding probabilities of success and of failure.

For symmetric binary variables, the trivariate marginal distributions coincide for the logit triangular system and for the linear triangular system of only main effects, since they have no additive and no multiplicative interaction; see the discussion of (2.9). For $p > 3$, the joint distributions generated by (3.1) will typically contain higher than two-factor effects of even order which differ from those generated by (2.1).

It can however be shown that for up to four variables, families of joint distributions can be obtained with (3.1) which are quadratic exponential and lead to an equal correlation matrix \mathbf{P} . The following vector of probabilities, gives an example for $\rho = 0.83\bar{3}$ and $p = 4$,

$$(\pi^{A_1A_2A_3A_4})^T = (81 \ 3 \ 3 \ 1 \ 3 \ 1 \ 1 \ 3 \ 3 \ 1 \ 1 \ 3 \ 1 \ 3 \ 3 \ 81)/192.$$

These distributions introduce six non-vanishing two-factor effects and one four-factor effect in the linear expansion of $\pi^{A_1 \dots A_4}$. For instance, the four-factor terms obtained for $\pi_{i|jkl}^{A_1A_2A_3A_4}$ from the linear triangular system, (2.1) as given in (4.3) below, and from the logit triangular system (3.1) are for equal correlation $\rho = 0.55\bar{5}, 0.83\bar{3}, 0.963$, respectively, 0.44, 0.78, 0.95. and 0.41, 0.75, 0.94. Thus, the joint quadrivariate distributions differ, but it would take large numbers of observations to discriminate between them in any given set of data.

In this paper, we explore properties of the linear triangular system (2.7). In the next section, we give its relation to a standardized joint Gaussian distribution having equal correlations.

4. Orthant probabilities

Attempts to approximate the Gaussian multivariate integral have a long tradition; see for instance [Schläfli \(1858\)](#), [Sheppard \(1898\)](#), [Moran \(1956\)](#), [McFadden](#)

(1955, 1956), Cheng (1969). For a more recent overview see Johnson, Kotz and Balakrishnan (1995). An orthant probability is the probability that all variables are simultaneously positive.

Sheppard (1898) derived the bivariate orthant probability in closed form, McFadden (1955) extended the result to trivariate distributions and Cheng (1969) used the di-logarithm for the quadrivariate distribution. From (2.7), bivariate and trivariate probability distributions for symmetric binary variables having equal correlation ρ are

$$\pi_{a_s, a_t}^{A_s, A_t} = \frac{1}{4}(1 + \rho a_s a_t), \quad \pi_{ijk} = \frac{1}{8}\{1 + \rho i(j+k) + \rho jk\}, \quad (4.1)$$

and from Sheppard’s result, the correlation coefficient ρ^* in an underlying Gaussian distribution satisfies $\rho^* = \sin(\frac{1}{2}\pi \rho)$, where π is Archimedes’ constant. In this context, each symmetric binary variable results by dichotomizing a corresponding standardized Gaussian variable at its median.

For equally correlated binary variables, the effect parameters η in (2.7) reduce to

$$\eta_{st} = \rho / (1 + \{d - 1\}\rho), \text{ for all } t \in r(s),$$

where d denotes the number of explanatory variables, i.e. is the dimension of $r(s)$. For $p = 4, 6, 8$, the even order interaction terms in the joint probabilities generated by the linear triangular system of only main effects are, respectively,

$$3 \eta_{12} \eta_{34}, \quad 15 \eta_{12} \eta_{34} \eta_{56}, \quad 105 \eta_{12} \eta_{34} \eta_{56} \eta_{78}, \quad (4.2)$$

where indices of the η ’s do not overlap.

Table 1 compares the binary orthant probabilities for a few selected values of odds being in favor of success with orthant probabilities of standardized and equally correlated Gaussian variables, as obtained with Matlab. Displayed are on top, the selected value of the odds, $\text{ods}_{A_{a_s}|1}$, the bivariate orthant probability, π_{11} , the correlations of the binary variables, ρ , and the corresponding correlation ρ^* of the Gaussian variables. For each p , there are two rows of orthant probabilities, the first row contains values given by Matlab for the Gaussian distribution and the second row contains values given by (4.4) for the linear triangular system. For $p = 4$, the former coincide with Cheng’s evaluations in terms of di-logarithms up to the 6th decimal place. As it turned out, the two types of orthant probabilities agree at least up to the third decimal place, also for all other values of $\rho > 0$ not shown in the Table 1.

The p ’th order interaction introduced for even p , and denoted by int_p , may be expressed directly in terms of ρ as

$$\text{int}_p = \rho^{p/2} \prod_{h=0}^{\frac{1}{2}(p-2)} (1 + 2h) / (1 + 2h\rho). \quad (4.3)$$

TABLE 1
 Orthant probabilities for standardized, equally correlated Gaussian variables and for symmetric binary variables generated by a triangular system of main effects

ods $A_{a_s 1}$	101/99	21/19	11/7	3	8	19	39
π_{11}	0.2525	0.2625	0.3056	0.3750	0.4444	0.4750	0.4875
ρ for Lintri	0.0100	0.0500	0.2222	0.5000	0.7777	0.9000	0.9500
ρ^* for Gauss	0.0157	0.0785	0.3420	0.7071	0.9397	0.9877	0.9969
$p = 4$, Gauss	0.0663	0.0817	0.1522	0.2734	0.3986	0.4542	0.4771
Lintri	0.0663	0.0817	0.1522	0.2736	0.3987	0.4544	0.4772
$p = 5$, Gauss	0.0344	0.0479	0.1167	0.2461	0.3853	0.4481	0.4740
Lintri	0.0344	0.0479	0.1167	0.2463	0.3858	0.4485	0.4742
$p = 6$, Gauss	0.0180	0.0290	0.0927	0.2256	0.3749	0.4432	0.4715
Lintri	0.0180	0.0289	0.0926	0.2259	0.3757	0.4439	0.4720
$p = 7$, Gauss	0.0095	0.0180	0.0756	0.2095	0.3663	0.4392	0.4695
Lintri	0.0095	0.0179	0.0756	0.2100	0.3676	0.4402	0.4701
$p = 8$, Gauss	0.0051	0.0114	0.0630	0.1964	0.3592	0.4358	0.4677
Lintri	0.0051	0.0114	0.0629	0.1970	0.3609	0.4371	0.4685
$p = 9$, Gauss	0.0027	0.0074	0.0534	0.1855	0.3530	0.4328	0.4662
Lintri	0.0027	0.0074	0.0533	0.1862	0.3551	0.4343	0.4671
$p = 10$, Gauss	0.0015	0.0049	0.0459	0.1762	0.3475	0.4301	0.4648
Lintri	0.0015	0.0049	0.0457	0.1771	0.3500	0.4319	0.4659

Similarly, the corresponding orthant probabilities for $p \geq 3$ are

$$\pi_{1 \dots 1}^{A_1 \dots A_p} = \begin{cases} 2^{-p} \prod_{h=\frac{1}{2}(p-2)}^{p-2} \{1 + (1 + 2h)\rho\} / \prod_{h=0}^{\frac{1}{2}(p-2)} (1 + 2h\rho) & \text{if } p \text{ is even,} \\ 2^{-p} \prod_{h=\frac{1}{2}(p-1)}^{p-2} \{1 + (1 + 2h)\rho\} / \prod_{h=0}^{\frac{1}{2}(p-3)} (1 + 2h\rho) & \text{if } p \text{ is odd,} \end{cases} \quad (4.4)$$

or, more explicitly, for instance for $p = 4, 5$, respectively,

$$\pi_{1111} = 2^{-4}(1 + 3\rho)(1 + 5\rho)/(1 + 2\rho), \quad \pi_{11111} = 2^{-5}(1 + 5\rho)(1 + 7\rho)/(1 + 2\rho).$$

5. Types of graphical Markov models

We now discuss distributions for symmetric binary variables generated by triangular systems (2.7) with independence constraints. The resulting graphs consist of nodes and of edges that couple pairs of nodes. Edges may be arrows, dashed lines or full lines. A path is a sequence of nodes for which the consecutive nodes are distinct and are coupled by an edge. Every path of three or more nodes has some inner nodes. These are nodes of the path that differ from the path endpoints. A path of at least two arrows is direction-preserving if every inner

node has an incoming and an outgoing arrow. A directed graph is acyclic if it is impossible to return to the same node following any direction-preserving path.

We assume in this section that edges present in the graph correspond to positive correlations and that a positive probability is associated with each level combination.

5.1. Markov chains, covariance and concentration graph models

A Markov chain is a special type of a triangular system represented by a directed graph consisting of a single direction-preserving path which is a special case of a directed acyclic graph. For four variables, such a path is

$$1 \leftarrow 2 \leftarrow 3 \leftarrow 4.$$

In the linear triangular system (2.1), the constraints are then $0 = \eta_{13} = \eta_{14} = \eta_{24}$, or, equivalently, $1 \perp\!\!\!\perp 34|2$ and $2 \perp\!\!\!\perp 4|3$. Every quadrivariate joint density with the above Markov chain structure factorizes. The factorization is, written in a condensed notation in terms of the nodes in the graph,

$$f_{1234} = f_{1|2} f_{2|3} f_{3|4} f_4.$$

This chain is represented in the correlation matrix of the symmetric binary variables by a particular pattern of correlations, which is the same pattern as for the corresponding Gaussian Markov chain. The diagonal matrix $\mathbf{\Delta}$ is defined by $\delta_{ss} = 1 - \rho_{s,s+1}^2$ for $s = 1, \dots, 4$. The matrix of regression coefficients \mathbf{H} and the correlation matrix $\mathbf{P} = \mathbf{H}^{-1} \mathbf{\Delta} \mathbf{H}^{-T}$ are

$$\mathbf{H} = \begin{pmatrix} 1 & -\rho_{12} & 0 & 0 \\ & 1 & -\rho_{23} & 0 \\ & & 1 & -\rho_{34} \\ \mathbf{0} & & & 1 \end{pmatrix}, \quad \mathbf{P} = \begin{pmatrix} 1 & \rho_{12} & \rho_{12} \rho_{23} & \rho_{12} \rho_{23} \rho_{34} \\ \cdot & 1 & \rho_{23} & \rho_{23} \rho_{34} \\ \cdot & \cdot & 1 & \rho_{34} \\ \cdot & \cdot & \cdot & 1 \end{pmatrix}.$$

where the dots indicate symmetric entries. Thus, the correlations corresponding to edges present in the generating directed graph match the simple correlations in \mathbf{P} ; the three other correlations for pairs (1,3), (1,4), (2,4) are products of correlations along the path and therefore called sometimes merely induced or spurious correlations.

The zero constraints on the regression coefficients turn here into more complex constraints on the correlation matrix. The Markov chain for (A_1, A_2, A_3, A_4) has as conditioning set $r(s)$ for variable A_s , $s = 1, 2, 3$, while the correlation matrix contains the bivariate association for each variable pair. For each missing arrow, the bivariate association results by ignoring some variables in the conditioning sets defined implicitly by the generating Markov chain. Therefore, the more complex structure in \mathbf{P} is sometimes said to represent a case of under-conditioning. Each of the induced correlations may be computed, for population values as well as for maximum-likelihood estimates, by using Wright's rules

for tracing paths whenever the directed graph is decomposable, i.e. contains no sink-oriented V-configuration; [Wermuth \(1980\)](#). Such a V-configuration is a subgraph of three nodes, \circ , having two arrows as edges and both arrows point to the common neighbor: $\circ \rightarrow \circ \leftarrow \circ$; see also [Section 5.6](#).

In a covariance graph, that is an undirected graph with edges shown by dashed lines, each edge corresponds to a marginal, pairwise association. Thus, for instance a missing edge for pair (1,3) means $1 \perp\!\!\!\perp 3$. For symmetric binary variables generated by [\(2.1\)](#), a covariance graph model is specified by zero constraints on marginal correlations. In general, some additional latent variables are needed to generate a covariance chain model recursively with univariate conditional distributions having exclusively zero parameter constraints; see [Pearl and Wermuth \(1994\)](#), [Holland and Rosenbaum \(1986\)](#).

For the above Markov chain, none of the marginal pairwise associations vanishes, so that the covariance graph induced by the generating directed graph for the given four variables is complete, i.e. it has no missing edges. In such cases, it cannot be recognized from the two graphs alone that the two model specifications captured by the matrices $(\mathbf{H}, \Delta^{-1})$ and \mathbf{P} are Markov equivalent, i.e. that they define the same independence structure.

By contrast, in a concentration graph, that is an undirected graph with edges shown by full lines, each edge corresponds to a conditional association given all of the remaining variables. Thus, for instance a missing edge for pair (1,3) in a concentration graph of four nodes means $1 \perp\!\!\!\perp 3|24$. For discrete variables, these conditional associations and independences are reflected in the parameters of log-linear models for the joint probabilities, that is in the effect expansion of $\log \pi^{A_1 \dots A_p}$; see [Appendix B](#). There is conditional independence of a variable pair if the log-linear two-factor effect and all higher-order effects of this pair vanish and this happens if and only if the conditional odds-ratios of this pair equal one at all level combinations of the remaining variables. For a Markov chain, the induced concentration graph has the same edge set as the Markov chain graph; see the following section.

For the symmetric binary variables discussed here, zero partial correlation coefficients given all remaining variables are necessary for the corresponding conditional independence statement to hold. But starting with four variables, zero partial correlations are no longer sufficient, see [Appendix C](#). Partial correlations given all remaining may be computed from \mathbf{P}^{-1} with elements ρ^{st} , see for instance [Cox and Wermuth \(1996\)](#), p. 69, as

$$\rho_{st|\{1,\dots,p\}\setminus\{st\}} = -\rho^{st} / \sqrt{\rho^{ss}\rho^{tt}}.$$

5.2. A concentration chain

For the Markov chain given in the previous section, the order of the variables was specified implicitly as (1,2,3,4) and all independences implied by the directed graph can be derived from the two defining statements $1 \perp\!\!\!\perp 34|2$ and $2 \perp\!\!\!\perp 4|3$. For instance, the statements $1 \perp\!\!\!\perp 3|24$ and $1 \perp\!\!\!\perp 4|23$ and $2 \perp\!\!\!\perp 4|13$ are implied. These

three independences happen to define a concentration graph in four nodes and edges present for pairs (1,2), (2,3) and (3,4), that is the undirected graph

$$1 \text{ --- } 2 \text{ --- } 3 \text{ --- } 4.$$

Conversely, given the defining independences of this concentration graph, all independence statements implied by the Markov chain graph of Section 5.1 can be derived, so that the two graphs are Markov equivalent.

For the symmetric binary variables, each edge present in the above concentration graph corresponds to a two-factor term in the effect expansion of $\log \pi^{A_1 A_2 A_3 A_4}$, where

$$\pi_{ijkl}^{A_1 A_2 A_3 A_4} = \frac{1}{16} \{ (1 + \rho_{12} ij)(1 + \rho_{23} jk)(1 + \rho_{34} kl) \}.$$

Then $\log f_{1234}$ is the sum of terms involving at most two variables. Therefore, the joint log-linear model has no higher-order than two-factor effects and no two-factor effects for all pairs of missing edges in the generating directed graph; see also (2.11).

In general, the characterizing feature for Markov equivalence of a directed acyclic graph and a concentration graph is that the generating directed graph is decomposable; see Frydenberg (1990). Thus, from a decomposable directed graph the corresponding induced concentration graph is obtained by replacing each arrow by a full line. In such a case, the concentration graph induced by the generating directed acyclic graph has an unchanged edge set and the two graphs are then said to have the same skeleton.

5.3. A covariance chain

A covariance chain for the ordered pairs (1, 2), (2, 3), (3, 4) has covariance graph

$$1 \text{ --- } 2 \text{ --- } 3 \text{ --- } 4,$$

where each edge present corresponds to the bivariate association of the corresponding variable pair. This is another undirected graph with the same skeleton as those discussed in the previous two graphs, but it represents a quite different independence structure.

For the linear triangular system (2.1), the correlation matrix \mathbf{P} to this graph and the matrix \mathbf{H} of the triangular decomposition of \mathbf{P}^{-1} are

$$\mathbf{P} = \begin{pmatrix} 1 & \rho_{12} & 0 & 0 \\ \cdot & 1 & \rho_{23} & 0 \\ \cdot & \cdot & 1 & \rho_{34} \\ \cdot & \cdot & \cdot & 1 \end{pmatrix}, \quad \mathbf{H} = \begin{pmatrix} 1 & -\eta_{12} & \eta_{12} \eta_{23} & -\eta_{12} \eta_{23} \eta_{34} \\ & 1 & -\eta_{23} & \eta_{23} \eta_{34} \\ & & 1 & -\eta_{34} \\ \mathbf{0} & & & 1 \end{pmatrix}.$$

Here, the dependences that do not correspond to edges present in the covariance graph, are products of the coefficients corresponding to the non-vanishing marginal correlations. Thus, simple zero constraints on the marginal correlations in \mathbf{P} , correspond to more complex nonlinear constraints in \mathbf{H} . Therefore,

the more complex structure in \mathbf{H} which may be regarded as a case of over-conditioning.

The directed graph in the induced by the covariance chain is complete, since none of the induced conditional dependences of A_s on $A_{r(s)}$ vanishes. Therefore, there is also no simplifying factorization of the joint density and there is no direct representation of the structure in the log-linear formulation of $\pi_{ijkl}^{A_1 A_2 A_3 A_4}$ that is in the effect expansion of $\log \pi^{A_1 A_2 A_3 A_4}$. The two model specifications captured by the matrices $(\mathbf{H}, \Delta^{-1})$ and \mathbf{P} are again Markov equivalent but this is not reflected in the corresponding two graphs alone, that is in the complete directed graph corresponding to \mathbf{H} , and in the covariance graph corresponding to \mathbf{P} , which is here a chain of dashed lines.

For a joint Gaussian distribution, maximum-likelihood estimation for a chain of covariances and for a chain of concentrations are quite different, even though both are defined without reference to an order of the variables; see [Wermuth, Cox and Marchetti \(2006\)](#) or [Cox \(2006\)](#), pp. 120-124. For binary variables, estimation under a covariance graph model maybe is less complex than in the Gaussian case; see [Drton \(2009\)](#). This holds also for symmetric binary variables when the multivariate regression is equivalent to a covariance chain; see the next section.

5.4. Seemingly unrelated regression and multivariate regression chains

A seemingly unrelated regression model, [Zellner \(1962\)](#), that is Markov equivalent to the covariance chain in the previous section has the graph

$$1 \longrightarrow 2 \text{ --- } 3 \longleftarrow 4,$$

where variables A_1, A_4 are independent explanatory variables for the joint responses A_2, A_3 . With a set of arrows pointing to a set of nodes that may be connected by undirected edges, one speaks sometimes of a joint response graph.

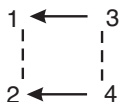
The graph implies in particular that A_1, A_4 are both generated before A_2, A_3 . We therefore reordered the correlation matrix accordingly, i.e. as (2, 3, 1, 4), to give \mathbf{P}' . This matrix and the changes that the result after partial inversion with respect to the explanatory variables are

$$\mathbf{P}' = \begin{pmatrix} 1 & \rho_{23} & 0 & \rho_{12} \\ \cdot & 1 & \rho_{34} & 0 \\ \cdot & \cdot & 1 & 0 \\ \cdot & \cdot & \cdot & 1 \end{pmatrix}, \quad \text{inv}_{3,4} \mathbf{P}' = \begin{pmatrix} 1 - \rho_{12}^2 & \rho_{23} & 0 & \rho_{12} \\ \cdot & 1 - \rho_{34}^2 & \rho_{34} & 0 \\ \sim & \sim & 1 & 0 \\ \sim & \sim & \cdot & 1 \end{pmatrix}.$$

By using the more explicit Yule-Cochran notation, where for instance $\beta_{1|2.3}$ denotes the coefficient of variable A_2 in linear regression of variable A_1 on A_2, A_3 , we obtain from equations (A.1), (A.2) and (A.4) in Appendix A that the nonzero two-factor terms in $\pi^{A_2 A_3 | A_1 A_4}$ reduce to $\beta_{2|1.3} = \rho_{12}$, $\beta_{3|4.1} = \rho_{34}$, $\eta_{23} + \eta_{24}\rho_{13} + \eta_{21}\rho_{34} = \rho_{23}$ and $\eta_{21}\eta_{34} + \eta_{24}\eta_{12} = \rho_{12}\rho_{34}$ to give

$$\pi_{jk|il}^{A_2 A_3 | A_1 A_4} = \frac{1}{4}(1 + \rho_{12}ij + \rho_{23}jk + \rho_{34}jl + \rho_{12}\rho_{34}ijkl).$$

The next, more complex seemingly unrelated regression model has associated explanatory variables and a joint response graph reflecting $1 \perp\!\!\!\perp 4|3$ and $2 \perp\!\!\!\perp 3|4$



In the model for this graph, the conditional distribution of the responses (1, 2) given (3, 4) has in general the same form as for independent explanatory variables, the case described first in this subsection.

Whenever a conditional distribution is complemented by a marginal one in such a way that its parameters do not depend on those of the margin, then the joint density gives independently varying factors, say $f_{12|34}$, f_{34} , where

$$f_{1234} = f_{12|34}f_{34},$$

and for maximum-likelihood estimation, each of the two parts of the likelihood can then be maximized separately. This is typically achieved, whenever the only types of constraint are independence statements captured by missing edges in graphs. By contrast, this is for instance not possible, when a parametric cancellation, such as illustrated in Figure 1, were to be explicitly excluded or when regression coefficients in different equations had to take on equal values.

The conditional probabilities are here

$$\pi_{ij|kl}^{A_1A_2|A_3A_4} = \frac{1}{4}(1 + \rho_{12}ij + \rho_{13}ik + \rho_{24}jl + \rho_{13}\rho_{24}ijkl),$$

so that the joint distribution is given by

$$\pi_{ijkl}^{A_1A_2A_3A_4} = \frac{1}{16}\pi_{ij|kl}^{A_1A_2|A_3A_4}(1 + \rho_{34}kl).$$

There is only a small step from this last example to the formulation of multivariate regressions chains and the corresponding graphs; see [Wermuth and Cox \(2004\)](#). Let for example $\{1, \dots, p\} = (a, b, c, d)$ be an ordered partition of $\{1, \dots, p\}$ and suppose that

$$f = f_{a|bcd}f_{b|cd}f_{c|d}f_d.$$

There results a factorisation corresponding to the joint or single responses within the chain components a, b, c, d . Each response has as potential explanatory variables all variables in the past but not those within the same chain component nor any variable in the future. Among several responses on equal standing, i.e. variables within the same chain component, the pairwise associations are modelled given the past.

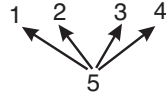
The corresponding chain graph has within each chain component a covariance graph of dashed lines and represents between chain components regressions given the past by arrows. For general discrete variables, [Marchetti and Lupparelli \(2009\)](#) show how such a multivariate regression chain is formulated as a multivariate logistic model of [Glonck and McCullagh \(1995\)](#). They also prove equivalence of its independence structure to the one obtained with a more complex formulation that depends only on the graph; see for instance [Drton \(2009\)](#).

5.5. Mutual conditional independence of A_1, A_2, A_3, A_4 given A_5

For mutually independent discrete variables given a latent variable, it is known that no constraints are implied on an observed contingency table provided that the latent variable has a sufficiently large number of levels; see [Holland and Rosenbaum \(1986\)](#). However, if all variables including the latent variable are binary then the correlation matrix of the observed variables satisfies the tetrad conditions, just like in the Gaussian case; see e.g. equation (21) of [Wermuth \(1976\)](#).

But, unless the joint probabilities even of symmetric binary variables may be generated via a triangular system of only main effects (2.7), a partial correlation coefficient of two binary variables given a third may be zero even if the three variables are strongly associated; see [Wermuth \(1998\)](#) for an example from psychological research.

We take here four symmetric binary variables to satisfy mutual conditional independence given A_5 and to have positive correlations with A_5 . The decomposable directed generating graph is



where each V-configuration of the graph implies conditional independence of the uncoupled variable pair given A_5 .

The matrix \mathbf{H} and the induced correlation matrix, $\mathbf{P} = \mathbf{H}^{-1} \mathbf{\Delta} \mathbf{H}^{-T}$, are

$$\mathbf{P} = \begin{pmatrix} 1 & \rho_{15}\rho_{25} & \rho_{15}\rho_{35} & \rho_{15}\rho_{45} & \rho_{15} \\ \cdot & 1 & \rho_{25}\rho_{35} & \rho_{25}\rho_{45} & \rho_{25} \\ \cdot & \cdot & 1 & \rho_{35}\rho_{45} & \rho_{35} \\ \cdot & \cdot & \cdot & 1 & \rho_{45} \\ \cdot & \cdot & \cdot & \cdot & 1 \end{pmatrix}, \quad \mathbf{H} = \begin{pmatrix} 1 & 0 & 0 & 0 & -\rho_{15} \\ & 1 & 0 & 0 & -\rho_{25} \\ & & 1 & 0 & -\rho_{35} \\ & & & 1 & -\rho_{45} \\ \mathbf{0} & & & & 1 \end{pmatrix},$$

and $\delta_{ss} = 1 - \rho_{s5}^2$ for $s = 1, \dots, 4$.

For instance for the submatrix $\mathbf{P}_{\{1,2\},\{3,4\}}$, the tetrad conditions can then be stated as

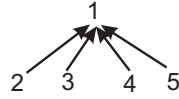
$$\rho_{13}/\rho_{23} = \rho_{14}/\rho_{24}.$$

This holds since every marginal correlation of the first four variables is induced by their correlations with A_5 to satisfy $\rho_{st|5} = 0$, that is

$$\rho_{st} = \rho_{s5}\rho_{t5} \text{ for } s \neq t \in \{1, 2, 3, 4\}.$$

5.6. Mutual independence of A_2, A_3, A_4, A_5 , common response A_1

The final example treats mutual marginal independence of four variables. With the four variables having a common response, the non-decomposable generating graph is



The joint distribution in symmetric binary variables, generated by (2.7), is in this case defined by an identity matrix for $\mathbf{H}_{r(1),r(1)}$, $\delta_{ss} = 1$ for $s = 2, \dots, 5$,

$$\mathbf{H}_{1,r(1)} = (-\rho_{12} \quad -\rho_{13} \quad -\rho_{14} \quad -\rho_{15}), \quad \delta_{11} = 1 - \sum_{t=2}^{t=5} \rho_{1t}^2.$$

If data to this model were analyzed by a log-linear model that is by an effect expansion of $\log \pi^{A_1 \dots A_5}$, then one would see all variables being associated. The reason is that for this model, the induced partial correlation for a pair $A_s A_t$ of the explanatory variables given all remaining variables is a multiple of $\rho_{s1} \rho_{t1}$.

Ignoring the information about which variables are responses and which independent explanatory variables by treating all variables on equal standing, is another instance of over-conditioning, that is here of inducing associations by enlarging the conditioning sets, $r(s)$ of A_s , for $s = 2, 3, 4, 5$ given in the generating process, to include the common response A_1 .

Thus, as we have seen, independences may lead to simple zero constraints in one formulation but may appear as more complex constraints in parameter equivalent but different model formulations. It is therefore in general rarely useful to restrict model fitting and data analysis to one particular class of graphical Markov models unless there is strong prior knowledge. Also, Markov equivalent graphs may be most helpful in suggesting possible alternative interpretations of a given model, but Markov equivalence of two models need not be reflected in the corresponding graphs. Thus by using only one class graphs to learn about models, one may overlook some important Markov equivalent models.

The outstanding features of the described linear triangular systems are the simplicity of the parametrization of different types of independence models for few variables, the close approximation of the orthant probabilities of equally correlated Gaussian variables by the corresponding binary orthant probabilities and the complete reversal of odds for success in two-way margins.

Because of the symmetry in the joint probability distributions it may be difficult to distinguish for a given small set of data among different types of independence structure. Moreover the permissible range of association parameters reduces when the number of variables in the generating process is increasing. It remains to be seen for which substantive questions these special binary distributions may become attractive. Possibly, complete reversals in odds for success of some features or events can be expected when in medical case-control studies, particular genes are present for cases and absent for controls.

Appendix A: Proof of equations (2.3), (2.9)

By noting that $a_s^2 = 1$, we have from (2.1)

$$\pi_{jkl}^{A_2 A_3 A_4} = \pi_{j|kl}^{A_2|A_3 A_4} \pi_{kl}^{A_3 A_4} = \frac{1}{8}(1 + \eta_{23}jk + \eta_{24}jl)(1 + \rho_{34}kl)$$

$$= \frac{1}{8}\{1 + (\eta_{23} + \eta_{24}\rho_{34})jk + (\eta_{24} + \eta_{23}\rho_{34})jl + \rho_{34}kl\}.$$

and for instance, from a matrix multiplied by its inverse giving the identity matrix,

$$\eta_{24} + \eta_{23}\rho_{34} = (\eta_{23} \ \eta_{24})(\rho_{34} \ 1)^T = (\rho_{23} \ \rho_{24}) \begin{pmatrix} 1 & \rho_{34} \\ \cdot & 1 \end{pmatrix}^{-1} (\rho_{34} \ 1)^T = \rho_{24}.$$

With $\eta_{23} + \eta_{23}\eta_{34} = \rho_{23}$ obtained similarly, equation (2.9) follows.

The above argument proves also Cochran’s recursion relation for regression coefficients, Cochran (1938). With the coefficients written in the more explicit Yule-Cochran notation, where for instance $\beta_{2|3.4}$ denotes the coefficient of variable X_2 in a linear least-squares regression of variable X_2 on X_3, X_4 , one obtains

$$\beta_{2|4} = \beta_{2|4.3} + \beta_{2|3.4}\beta_{3|4}, \tag{A.1}$$

here applied to regression coefficients of binary variables, standardized to have mean zero and unit variance. See Cox and Wermuth (2003), Ma, Xie and Geng (2006), Cox (2007), Xie, Ma and Geng (2008) for interpretations and consequences of (A.1).

From (2.1) and (A.1), giving $\eta_{23} = \beta_{2|3.4}$, $\eta_{24} = \beta_{2|4.3}$,

$$\eta_{14} + \eta_{12}\eta_{24} = \beta_{1|4.3}, \quad \eta_{13} + \eta_{12}\eta_{23} = \beta_{1|3.4},$$

and with $\pi_{ij|kl}^{A_1A_2|A_3A_4} = \pi_{i|jkl}^{A_1|A_2A_3A_4}\pi_{j|kl}^{A_2|A_3A_4}$, one obtains

$$\begin{aligned} \pi_{ij|kl}^{A_1A_2|A_3A_4} &= \frac{1}{4}\{1 + \beta_{1|3.4}ik + \beta_{1|4.3}il + \beta_{2|3.4}jk + \beta_{2|4.3}jl \\ &+ (\eta_{12} + \eta_{13}\eta_{23} + \eta_{14}\eta_{24})ij + (\eta_{13}\eta_{24} + \eta_{14}\eta_{23})ijkl\}. \end{aligned} \tag{A.2}$$

From

$$\begin{pmatrix} 1 & \rho_{23} & \rho_{24} \\ \cdot & 1 & \rho_{34} \\ \cdot & \cdot & 1 \end{pmatrix}^{-1} (1 \ \rho_{23} \ \rho_{24})^T = (1 \ 0 \ 0)^T, \tag{A.3}$$

one gets an extension of Cochran’s formula as

$$\eta_{12} + \eta_{13}\rho_{23} + \eta_{14}\rho_{24} = (\eta_{12} \ \eta_{13} \ \eta_{14})(1 \ \rho_{23} \ \rho_{24})^T = \rho_{12}. \tag{A.4}$$

Then, from (A.2) with $\pi_{ijkl} = \pi_{ij|kl}^{A_1A_2|A_3A_4}\{\frac{1}{4}(1 + \rho_{34}kl)\}$ we have

$$\begin{aligned} \pi_{ijkl} &= \frac{1}{16}[1 + \rho_{12}ij + (\beta_{1|3.4} + \beta_{1|4.3}\rho_{34})ik + (\beta_{1|4.3} + \beta_{1|3.4}\rho_{34})il \\ &+ (\beta_{2|3.4} + \beta_{2|4.3}\rho_{34})jk + (\beta_{2|4.3} + \beta_{2|3.4}\rho_{34})jl + \rho_{34}kl \\ &+ \{\eta_{12}\rho_{34} + \eta_{13}(\beta_{2|4.3} + \beta_{2|3.4}\rho_{34}) + \eta_{14}(\beta_{2|4.3} + \beta_{2|3.4}\rho_{34})\}ijkl], \end{aligned}$$

and hence, after repeatedly applying Cochran’s recursion relation (A.1), the claimed form of (2.3).

It is equation (A.3) which extends directly to higher dimensions. This leads to further generalisations of Cochran’s relation, in addition to (A.4) and hence

to direct expressions of a marginal correlation in terms weighted sums of η 's and ρ 's. Products of parameters without overlapping indices are not related in this way and induce higher-order interactions, such as given in (4.2) for the equal correlation case. In general, for instance the six-factor interaction contains terms such as

$$\beta_{1|2.3456}\beta_{3|4.56}\beta_{5|6}, \quad \beta_{1|3.2456}\beta_{2|4.56}\beta_{5|6}.$$

Such expressions look complex, but they are readily computed via the effect parameter expansion of the joint probabilities that are given by the linear triangular system.

Appendix B: Effect parameter expansions

Effect parameter expansions for binary variables are given by Yates' algorithm (Yates, 1937), which has been extended to general discrete variables by Good (1958) using contrast matrices.

Every contrast matrix is the inverse of a corresponding design matrix, \mathbf{D}_p , that can for a 2^p table be obtained in terms of a 2×2 matrix \mathbf{D}_1 via Kronecker products, where here for effect coding

$$\mathbf{D}_1 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \quad \mathbf{D}_1^{-1} = \frac{1}{2}\mathbf{D}_1, \quad \mathbf{D}_2 = \mathbf{D}_1 \otimes \mathbf{D}_1, \quad \mathbf{D}_3 = \mathbf{D}_1 \otimes \mathbf{D}_2, \quad \dots$$

Thus, since the inverse of a matrix of Kronecker products is the Kronecker product of the inverses

$$\mathbf{C}_p = \mathbf{D}_p^{-1} = 2^{-p} \mathbf{D}_p.$$

When writing the 2^p probabilities in a column vector π^{A_1, \dots, A_p} such that the levels of the first variable change fastest, then the levels of the second variable next and the levels of the p 'th variable last, then the effect parameters that result with $\mathbf{C}_p \pi^{A_1, \dots, A_p}$ are in a lexicographical order in which the overall effect λ_- is followed by the main effect $\lambda_1^{A_1}$ of A_1 , then $(\lambda_1^{A_2}, \lambda_{11}^{A_1 A_2})$ is added to $(\lambda_-, \lambda_1^{A_1})$, then $(\lambda_1^{A_3}, \lambda_{11}^{A_1 A_3}, \lambda_{11}^{A_2 A_3}, \lambda_{111}^{A_1 A_2 A_3})$ follows the first four terms and so on, adding at the next step, the additional variable in the superscripts and a level 1 in the subscripts of each of the previous terms.

Thus, for a vector of joint probabilities with structure given by (2.9), the linear effect parameter expansion is

$$(\mathbf{C}_3 \pi_{ijk}^{A_1 A_2 A_3})^T = \frac{1}{8}(1 \ 0 \ 0 \ \rho_{12} \ 0 \ \rho_{13} \ \rho_{23} \ 0)$$

and for the three correlations being nonzero, precisely the entries in position 2,3,5 and 8 are also zero in $(\mathbf{C}_3 \log \pi_{ijk}^{A_1 A_2 A_3})^T$, the log-linear effect expansion.

For any constant c , the linear effect terms of $c \pi_{ijk}^{A_1 A_2 A_3}$ are multiplied by c , while the effect parameters of $\log c \pi_{ijk}^{A_1 A_2 A_3}$ remain unchanged except that $\log c$ is added to the overall effect. Thus, for comparisons of linear and log-linear parameters it is sometimes convenient to divide the probability vector by one

of its probabilities to obtain ones in some entries. For instance, for model (2.9) with equal correlations, we have with $\alpha = \pi_{111}/\pi_{-111}$ and $\beta = \log \alpha$

$$(\pi^{A_1 A_2 A_3})^T / \pi_{-111} = (\alpha \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ \alpha),$$

$$\log\{(\pi^{A_1 A_2 A_3})^T / \pi_{-111}\} = (\beta \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ \beta).$$

so that the log-linear effect expansion becomes $\frac{1}{4}(\beta \ 0 \ 0 \ \beta \ 0 \ \beta \ \beta \ 0)$.

Appendix C: A Gaussian and a symmetric binary correlation matrix

Here, an example of a correlation matrix is provided, in which precisely one independence, $1 \perp\!\!\!\perp 2|34$, holds whenever it represents the correlation structure of a joint Gaussian distribution since $\rho_{12|34} = 0$ so that $\rho_{12}^* = 0.617\bar{3}$ is given by (2.8) with $a = \{1\}, b = \{2\}, c = \{34\}$. But as we shall see, when the same matrix represents the correlation structure of a linear triangular system of symmetric binary variables, it depends on the order in which the variables are generated, whether $1 \perp\!\!\!\perp 2|34$ holds or not.

The following display on the left, shows in the order (1,2,3,4), marginal correlations in the lower half, and partial correlations given the two remaining variables in the upper half; on the right in the upper half, the upper part of the matrix **H** for generating the joint density the order (1,2,3,4) and in the lower half, the upper part of the matrix **H'** for the order (4,3,2,1) but rearranged so that indices s, t correspond to t, s in the upper half:

$$\begin{pmatrix} 1.0000 & \mathbf{0} & 0.2285 & 0.4180 \\ 0.617\bar{3} & 1.00 & 0.2489 & 0.4451 \\ 0.7000 & 0.72 & 1.0000 & 0.3900 \\ 0.7600 & 0.78 & 0.8000 & 1.0000 \end{pmatrix} \begin{pmatrix} 1.0000 & \mathbf{0} & -0.2556 & -0.5556 \\ -0.617\bar{3} & 1.0000 & -0.2667 & -0.5667 \\ -0.4129 & -0.4651 & 1.0000 & -0.8000 \\ -0.3145 & -0.3496 & -0.3282 & 1.0000 \end{pmatrix}$$

The two joint distributions of the binary variables differ substantially for the two orders. For (1,2,3,4), the conditional probability vectors $\pi^{A_3|A_4}$ are generated first, then $\pi^{A_2|A_3A_4}$ and last $\pi^{A_1|A_2A_3A_4}$ while for (4,3,2,1), $\pi^{A_2|A_1}$ comes first, then $\pi^{A_3|A_2A_1}$ and $\pi^{A_4|A_3A_2A_1}$ last. With the following index vector, the joint probabilities obtained for (4,3,2,1) are arranged to imply the correlation matrix as it has been used for (1,2,3,4), i.e. the same values in the same order

$$(1, 9, 5, 13, 3, 11, 7, 15, 2, 10, 6, 14, 4, 12, 8, 16).$$

We get as the first case for the variable order (1,2,3,4)

$1000\pi_{ijk1}$		levels $jk1$ of A_2, A_3, A_4			
levels of A_1		111	-111	1-11	-1-11
$i = 1$		373.54	33.96	21.12	11.38
$i = -1$		38.96	3.54	11.38	6.12
$\text{odr}(A_1 A_2 A_3 = k, A_4 = 1)$		1		1	

and as the second case for the variable order (4,3,2,1), with the probabilities rearranged to be directly comparable to those obtained for (1,2,3,4),

1000 π_{ijk1}				
levels of A_1	levels $jk1$ of A_2, A_3, A_4			
	111	-111	1 - 11	-1 - 11
$i = 1$	378.19	29.31	16.48	16.02
$i = -1$	34.31	8.19	16.02	1.48
odr($A_1 A_2 A_3 = k, A_4 = 1$)	3.081		0.095	

Because of the symmetry of the joint distributions, it is enough to list i, j, k at level 1 of A_4 . Since in the second case the conditional odds-ratios $\text{odr}(A_1 A_2 | A_3 = k, A_4 = 1)$ are quite different, there is a substantial four-factor effect in the log-linear expansion. It is only in the first case that $1 \perp\!\!\!\perp 2 | 34$ holds.

In both cases, non-vanishing four-factor terms result in the linear effect expansion of the probabilities. This illustrates in particular that there may be a conditional independence in the presence of a linear four-factor effect. The second case shows with $\rho_{12|34} = 0$, that there may be a linear independence without a corresponding conditional independence statement being satisfied.

Thus, in spite of the strong similarities to standardized, equally correlated Gaussian distributions, some important properties of Gaussian distributions do not carry over to the described joint distributions of symmetric binary variables.

Acknowledgements

Computations were carried out with Matlab R2009a. We thank the referees for their careful reading and constructive comments.

References

- BAHADUR, R.R. (1961). A representation of the distribution of joint responses to n dichotomous items. In: *Studies in item analysis and prediction*. H. Solomon (ed.) 158–176. Stanford University Press. [MR0121893](#)
- BARTLETT, M.S. (1935). Contingency table interactions. *Suppl. J. Roy. Statist. Soc.*, **2**, 248–252.
- BERGSMAN, W., CROON, M. & HAGENAARS, J.A. (2009) *Marginal Models*. Springer, New York.
- BIRCH, M.W. (1963). Maximum likelihood in three-way contingency tables. *J. Roy. Statist. Soc. B* **25**, 220–233. [MR0168065](#)
- CHENG, M.C. (1969). The orthant probabilities of four Gaussian variates. *Ann. Math. Statist.* **40**, 152–161. [MR0235596](#)
- COCHRAN, W.G. (1938). The omission or addition of an independent variate in multiple linear regression. *Suppl. J. Roy. Statist. Soc.* **5**, 171–176.
- COX, D.R. (2006). *Principles of Statistical Inference*. Cambridge University Press.

- COX, D.R. (2007). On a generalization of a result of W.G. Cochran. *Biometrika*, **94**, 400–410. [MR2410022](#)
- COX, D.R. & WERMUTH, N. (1994). A note on the quadratic exponential binary distribution. *Biometrika*, **81**, 403–408. [MR1294901](#)
- COX, D.R. & WERMUTH, N. (1996). *Multivariate Dependencices – Models, Analysis and Interpretation*. London: Chapman & Hall. [MR1456990](#)
- COX, D.R. & WERMUTH, N. (2003). A general condition for avoiding effect reversal after marginalization. *J. Roy. Statist. Soc. B*, **65**, 937–941. [MR2017879](#)
- CRAMÈR, H. (1946). *Mathematical methods of statistics*. Princeton University Press. [MR1816288](#)
- DARROCH, J.N. & RATCLIFF, D. (1972) Generalized iterative scaling for log-linear models. *Ann. Math. Statist.*, **43**, 1470–1480. [MR0345337](#)
- DARROCH, J.N., LAURITZEN, S.L. & SPEED, T.P. (1980). Markov fields and log-linear models for contingency tables. *Ann. Statist.* **8**, 522–539. [MR0568718](#)
- DEMPSTER, A. P. (1972). Covariance selection. *Biometrics* **28**, 157–175.
- DRTON, M. (2009). Discrete chain graph models. *Bernoulli*. to appear.
- FIENBERG, S. (2007). The analysis of cross-classified categorical data. Springer (2007), New York.
- FRYDENBERG, M. (1990). The chain graph Markov property. *Scan. J. Statist.* **17**, 333–353. [MR1096723](#)
- GLONEK, G.J.N. & MCCULLAGH, P. (1995). Multivariate logistic models. *J. Roy. Statist. Soc. B*, **57**, 533–546.
- GOLDBERGER, A.S. (1964). *Econometric Theory*. Wiley, New York. [MR0158754](#)
- GOOD, I. J. (1958). The interaction algorithm and practical Fourier analysis. *J. Roy. Statist. Soc. B* **20**, 361–372. [MR0126937](#)
- GOODMAN, L.A. (1973). The analysis of multidimensional contingency tables when some variables are posterior to others: a modified path analysis approach. *Biometrika*, **60**, 179–192. [MR0415839](#)
- GREEN, P.J., HJORT, N. & RICHARDSON, S. (eds.) (2003). *Models for highly structured stochastic systems*. Oxford University Press.
- HOLLAND, P.W. & ROSENBAUM, P. (1986). Conditional association and unidimensionality in monotone latent variable models. *Ann. Statist.*, **14**, 1523–1543. [MR0868316](#)
- JOHNSON, N.L., KOTZ, S. & BALAKRISHNAN, N. (1995). Distributions in statistics. Continuous univariate distributions, Vol. 2, 2nd edn., Wiley, New York.
- LANCASTER, H.O. (1969). *The Chi-Squared Distribution*. Wiley, New York. [MR0253452](#)
- LAZARSFELD, P.F. (1961). The algebra of dichotomous systems. In: Studies in item analysis and prediction. H. Solomon (ed.) 111-157. Stanford University Press. [MR0121892](#)
- MA, Z.M., XIE, X.C. & GENG, Z. (2006). Collapsibility of distribution dependence. *J. Roy. Statist. Soc. B*, **68**, 127–133. [MR2212578](#)
- MARCHETTI G.M. & LUPPARELLI, M. (2009). Chain graph models of multivariate regression type for categorical data. Submitted.

- MARCHETTI, G.M. & WERMUTH, N. (2009). Matrix representations and independencies in directed acyclic graphs. *Ann. Statist.*, **37**, 961–978. [MR2502657](#)
- MCFADDEN, J.A. (1955). Urn models of correlation and a comparison with the multivariate normal integral. *Ann. Math. Statist.* **26**, 478–489. [MR0070863](#)
- MCFADDEN, J.A. (1956). An approximation for the symmetric quadrivariate normal integral. *Biometrika* **43**, 206–207. [MR0077051](#)
- MORAN, P.A.P. (1956). The numerical evaluation of a class of integrals. *Proc. Cambridge Philos. Soc., Statist.* **52**, 230–233. [MR0076450](#)
- PEARL, J. & WERMUTH, N. (1994). When can association graphs admit a causal interpretation? In: *Models and Data, Artificial Intelligence and Statistics IV*. 205–214. P. Cheeseman and W. Oldford (eds.). New York: Springer.
- SCHLÄFLI, L. (1858). On the multiple integral. *Quart. J. Math.*, **2**, 269–301, **3** 54–68, 97–108.
- SHEPPARD, W.F. (1898). On the geometrical treatment of the 'normal curve' of statistics, with special reference to correlation and to the theory of error. *Proc. Roy. Soc. London*, **62**, 171–173.
- STREITBERG, B. (1990). Lancaster interactions revisited. *Ann. Statist.*, **18**, 1878–1885. [MR1074442](#)
- STREITBERG, B. (1999). Exploring interactions in high-dimensional tables: a bootstrap alternative to log-linear models. *Ann. Statist.*, **27**, 405–413. [MR1701118](#)
- WERMUTH, N. (1976). Analogies between multiplicative models for contingency tables and covariance selection. *Biometrics*, **32**, 95–108. [MR0403088](#)
- WERMUTH, N. (1980) Linear recursive equations, covariance selection, and path analysis. *J. Amer. Statist. Assoc.* **75** 963–972. [MR0600984](#)
- WERMUTH, N. (1998). Pairwise independence. *Encyclopedia of Biostatistics*. P. Armitage and T. Colton (eds). New York: Wiley, 3244–3246.
- WERMUTH, N. & COX, D.R. (1998). On association models defined over independence graphs. *Bernoulli*, **4**, 477–495. [MR1679794](#)
- WERMUTH, N. & COX, D.R. (2004). Joint response graphs and separation induced by triangular systems. *J. Roy. Statist. Soc. B*, **66**, 687–717. [MR2088296](#)
- WERMUTH, N., COX, D.R. & MARCHETTI, G.M. (2006). Covariance chains. *Bernoulli*, **12**, 841–862. [MR2265345](#)
- WERMUTH, N., WIEDENBECK, M. & COX, D.R. (2006). Partial inversion for linear systems and partial closure of independence graphs. *BIT, Numerical Mathematics*, **46**, 883–901. [MR2285213](#)
- WRIGHT, S. (1934). The method of path coefficients. *Ann. Math. Statist.*, **5**, 161–215.
- XIE, X.C., MA, Z.M. & GENG, Z. (2008). Some association measures and their collapsibility. *Stat. Sinica*. 1165–1183. [MR2440406](#)
- YATES, F. (1937). *The Design and Analysis of Factorial Experiments*. Harpenden: Imperial Bureau of Soil Science.
- ZELLNER, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *J. Amer. Statist. Assoc.* **57**, 348–368. [MR0139235](#)