Triangular systems for symmetric binary variables

Nanny Wermuth, Mathematical Statistics, Chalmers/Gothenburg University, Sweden
Giovanni M. Marchetti, Department of Statistics, University of Florence, Italy
D.R. Cox, Nuffield College, Oxford, 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.*
Keywords: Graphical Markov models, linear in probability models, log-linear models, odds-ratios, partial correlation, recursive equations.

1 Introduction

The explicit study of dependencies and independencies 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 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 independencies 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 levels by -1 and 1, so that all variables are standardized to have marginally zero mean and unit variance. This means, in particular, that linear representations of probabilities involve parameters of which some can be interpreted as correlation coefficients and others as regression coefficients. A strong parallel with Gaussian theory is established and used to compare typical cases of 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 log linear or 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.

2 A linear triangular system for symmetric binary variables

2.1 Definition and background

We consider p binary variables A_s, $s = 1, \ldots, p$. 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

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

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

$$\pi_{111}^{ABC} = \Pr(A = 1, B = 1, C = 1), \qquad \pi_{1|11}^{B|CD} = \Pr(B = 1|C = 1, D = 1).$$

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

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

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

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

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 (1) by i and summing

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

The joint distribution of the four variables has from (1) the form

$$\pi^{A_1A_2A_3A_4}_{ijkl} = \tfrac{1}{16}(1 + \rho_{12}ij + \rho_{13}ik + \rho_{14}il + \rho_{23}jk + \rho_{24}jl + \rho_{34}kl + \eta^{1234}ijkl), \tag{3}$$

with

$$\eta^{1234} = (\eta_{12} \quad \eta_{13} \quad \eta_{14})(\rho_{34} \quad \rho_{24} \quad \rho_{23})^{\mathrm{T}}. \tag{4}$$

From (3), the correlation coefficient of A_s, A_t results with

$$E(A_sA_t) = (\pi^{A_sA_t}_{11} + \pi^{A_sA_t}_{-1-1}) - (\pi^{A_sA_t}_{-11} + \pi^{A_sA_t}_{1-1}) = \rho_{st}. \tag{5}$$

In Appendix A, equations (3) and (4) are derived from (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 $\pi^{A_1,\ldots,A_p}$, the vector of the joint probabilities; see Appendix B.

The parameters of the linear triangular system of p symmetric binary variables relate also directly to the triangular decomposition of a concentration matrix, which is, when the variables are standardized to have zero means and unit variances, the inverse of the correlation matrix $\mathbf{P}$. This triangular decomposition is $(\mathbf{A}, \mathbf{\Delta}^{-1})$ and gives $\mathbf{P}^{-1} = \mathbf{A}^{\mathrm{T}}\mathbf{\Delta}^{-1}\mathbf{A}$, where $\mathbf{A}$ 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 $(1, \ldots, p)$.

In particular, for the linear triangular system treated here

$$\mathbf{A}_{st} = -\eta_{st}, \quad \mathbf{A}_{s,r(s)} = -\mathbf{P}_{s,r(s)}\mathbf{P}^{-1}_{r(s),r(s)} \text{ for } s < t \in \{1, \dots, p\}, \tag{6}$$

and for four variables

$$\delta_{11} = 1 - (\eta_{12} \;\; \eta_{13} \;\; \eta_{14})(\rho_{12} \;\; \rho_{13}, \;\; \rho_{14})^{\mathrm{T}}, \; \delta_{22} = 1 - (\eta_{23} \;\; \eta_{24})(\rho_{23} \;\; \rho_{34})^{\mathrm{T}}, \; \delta_{33} = 1 - \rho^2_{34},$$

and $\delta_{44} = 1$. For $p > 2$, the δ_{ss} are not the conditional variances of the binary variables, $\mathrm{var}(A_s|A_{r(s)})$, but their expected values with respect to $A_{r(i)}$.

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 $\pi^{A_p}_1 = \frac{1}{2}$ and

$$\pi^{A_s|A_{r(s)}}_{a_s|a_{r(s)}} = \tfrac{1}{2}(1 + a_s\{a_{s+1}\eta_{s,s+1} + \cdots + a_p\eta_{sp}\}), \tag{7}$$

with $(\eta_{s,s+1}, \dots, \eta_{sp}) = -\mathbf{A}_{s,r(s)}$ as in (6) and

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

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

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

Linear triangular systems for continuous responses have been studied in econometrics under the name of linear recursive equations with uncorrelated residuals, and in genetics as path analysis models. There, the regression coefficients have been called least-squares regression coefficients in the population; see Cramér (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).

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

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

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

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 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 in (7) main effects in probabilities coincide with linear regression coefficients.

Joint distributions generated via linear triangular systems 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, such as (7). The symmetry of the special binary variables considered here, leads to missing odd-order terms in the effect expansion of $\pi^{A_1,\ldots,A_p}$. The presence of only even-order effects makes the joint binary distributions similar to joint Gaussian distributions and leads to a good approximation of the probabilities that all Gaussian variables are jointly positive, as shown for equal correlation in 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 (7) in symmetric binary variables have similar implications as in a joint Gaussian distributions. For instance for three disjoint subsets a, b, c of $\{1, \ldots, p\}$, the conditional independence $A_a \perp\!\!\!\perp A_b | A_c$, often written compactly as $a \perp\!\!\!\perp b | c$, is imposed on the correlation matrix when replacing $\mathbf{P}_{ab}$ by $\mathbf{P}^*_{ab}$, where

$$\mathbf{P}^*_{ab} = \mathbf{P}_{ac}\mathbf{P}^{-1}_{cc}\mathbf{P}_{cb}, \tag{8}$$

and leaving all other entries 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 $\mathbf{P}^*_{ab|c} = 0$ which is known to be equivalent to $a \perp\!\!\!\perp b | c$, for variables $X_1, \ldots, X_p$ having a multivariate Gaussian distribution.

In the linear triangular systems (7), mutual conditional independence shows in zero parameters of this stepwise generating process in univariate conditional distributions, such as for (1) with $A_1 \perp\!\!\!\perp A_2 \perp\!\!\!\perp A_3 | A_4$ in $0 = \eta_{12} = \eta_{13} = \eta_{23}$, while marginal mutual independence shows in zero parameters of the joint distribution, such as for (3) with $A_2 \perp\!\!\!\perp A_3 \perp\!\!\!\perp A_4$ in $0 = \rho_{23} = \rho_{24} = \rho_{34}$. This makes the family attractive for illustrating in Section 4 basic distinctions and similarities between various types of graphical

Markov models, as studied for discrete variables by Darroch, Lauritzen and Speed (1980), Wermuth and Cox (2004), Drton (2009), Marchetti and Lupparelli (2009).

2.2 Some basic properties

Streitberg (1999) stresses that estimation in constrained Bahadur expansions is still undeveloped and that a truncated Bahadur expansion need not give a valid density, i.e. be nonnegative. Here, the closed form (4) of the four-factor interaction term makes the last statement transparent. 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 (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 (7), or result for instance 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 (7) is that the symmetry in all margins, $\frac{1}{2} = \pi_1 = \pi_{-1}$, carries over to the joint probabilities, so that

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

which says that the probability of any given pattern of level combinations coincides with the probability of the reversed pattern.

For $p \geq 4$, every marginal trivariate distribution of (7) is of the same form as π^{BCD}_{jkl} in (1) that is

$$\pi^{BCD}_{jkl} = \tfrac{1}{8}(1 + \rho_{23}jk + \rho_{24}jl + \rho_{34}kl). \qquad (9)$$

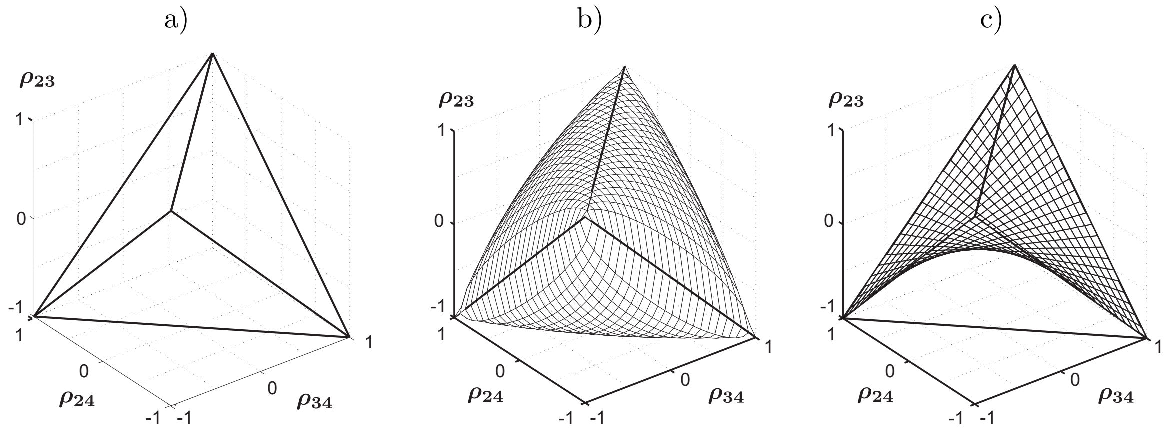
Such distributions have in general four distinct probabilities since

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

Sets of marginal correlations that give valid trivariate distributions (9) satisfy

$$\rho_{23} + \rho_{24} + \rho_{34} \geq -1, \quad \rho_{23} - \rho_{24} - \rho_{34} \geq -1, \quad -\rho_{23} + \rho_{24} - \rho_{34} \geq -1,$$

Figure 1: a) the simplex for 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 distance of these boundaries to the simplex is 1/12. Figure 1c) shows the surface excluded by requesting absence of one instance of parametric cancellation in the distribution of B, C, D generated by (1) with nonvanishing $\eta_{23}, \eta_{24}, \rho_{34}$, that is excluding $\rho_{23} = \rho_{24}\rho_{34}$.

With only linear two-factor effects in (9), there is in this special trivariate family of symmetric binary variables by definition no additive interaction as obtained for three-dimensional contingency tables by Lazarsfeld (1961), or Lancaster (1969); see also Streitberg (1990). In fact, it also has no log-linear 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 $\text{ods}_{A_2|jk} = \pi_{1|jk}^{A_2|A_3A_4} / \pi_{-1|jk}^{A_2|A_3A_4}$. There are two odds reversals in (9) with

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

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

$$\text{odr}(A_2A_3|A_4 = l) = \frac{\pi_{11l}\pi_{-1-1l}}{\pi_{-11l}\pi_{1-1l}} = \text{ods}_{A_2|1l}/\text{ods}_{A_2|-1l}.$$

Consequently, there is no multiplicative interaction, as defined for three-dimensional contingency tables by Bartlett (1935) and as extended later for measuring associations in general log-linear models; see for instance Fienberg (2007).

For general trivariate log-linear models without a three-factor effect, the maximum-likelihood estimates of the expected counts $m_{ijk} = n\pi_{ijk}$, where $n = \sum n_{ijk}$ is the total number of observed counts; 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, such as iterative proportional fitting, for which Darroch and Ratcliff (1972) proved convergence.

For symmetric binary variables, closed form estimates are available instead, since the set of minimal sufficient statistics for model (9) consists of the three observed sums of counts each for two successes and two failures, and for the mixture of success and failure

$$n_{11}^{A_sA_t} + n_{-1-1}^{A_sA_t}, \qquad n_{-11}^{A_sA_t} + n_{1-1}^{A_sA_t}.$$

These lead to closed form maximum-likelihood estimates in terms of differences of the cross sums

$$\hat{\rho}_{st} = \{(n_{11}^{A_sA_t} + n_{-1-1}^{A_sA_t}) - (n_{-11}^{A_sA_t} + n_{1-1}^{A_sA_t})\}/n, \tag{10}$$

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

$$\mathrm{odr}(A_sA_t) = \{(1+\rho_{st})/(1-\rho_{st})\}^2, \qquad \rho_{st} = \tanh\{\tfrac{1}{4}\log \mathrm{odr}(A_sA_t)\}. \tag{11}$$

For $p > 3$, the method of moments yields with (10) and (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 but compare different types of models.

Equal odds-ratios for any variable pair, at the fixed level combinations of all remaining variables, is equivalent for this variable pair to have zero effect parameters of all orders higher than two in the log-linear model for $A_1, \dots, A_p$, that is in the effect expansion of $\log \pi^{A_1 \dots A_p}$. Log-linear models with this property satisfied for all variable pairs have been studied under the name of binary quadratic exponential distributions by Cox and Wermuth (1994). Such joint distributions are, as (7), fully determined by given sets of two-way marginal tables. Nevertheless for $p \geq 3$, they are in general distinct from those generated by the linear triangular system, as well as from those generated by a logit triangular system in main effects; see the next section.

3 A logit triangular system for symmetric binary variables

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

$$\mathrm{logit}\, \pi_{1|l}^{A_3|A_4} = \log \mathrm{ods}_{A_3|l}, \quad \mathrm{logit}\, \pi_{1|kl}^{A_2|A_3A_4} = \log \mathrm{ods}_{A_2|kl}.$$

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

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

where

$$\exp(\theta_{34}) = \tfrac{1}{2}\mathrm{odr}(A_3A_4), \quad \exp(\theta_{2s}) = \tfrac{1}{2}\mathrm{odr}(A_2A_s|A_t = a_t) \text{ for } s \neq t \in r(2),$$

$$\exp(\theta_{1s}) = \tfrac{1}{2}\mathrm{odr}(A_1A_s|A_t = a_t, A_u = a_u) \text{ for } s \neq t \neq u \in r(1).$$

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 (9). For $p > 3$, the joint distributions generated by (12) will typically contain higher than two-factor effects which differ from those generated by (1) for $\log \pi_{ijkl}$.

However it can be shown that for up to four variables, families of joint distributions can be obtained with (12) which are quadratic exponential and lead to equal correlations for all variable pairs. The following vector of probabilities, gives an example for $\rho = 0.83\bar{3}$ and $p = 4$,

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

These distributions introduce non-vanishing two-and four-factor effects in the linear expansion of $\pi^{A_1 \dots A_4}$. For instance, for equal correlations $\rho = 0.55\bar{5},\ 0.83\bar{3},\ 0.963$, the four-factor terms obtained for $\pi_{ijkl}^{A_1A_2A_3A_4}$ from (1) as given in (15) below and from (12) are, 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 a given set of data. Here, we continue to explore properties of the linear triangular system (7), in the next section 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 (1948), McFadden (1956), Cheng (1969). For a more recent overview see Johnson, Kotz and Balakrishnan (1996). An orthant probability is the probability that all variables are simultaneously positive.

Sheppard (1898) derived the bivariate orthant probability in closed form, McFadden (1956) extended the result to trivariate distributions and Cheng (1969) used the di-logarithm for the quadrivariate distribution. From (7), bivariate and trivariate probability distributions for symmetric binary variables having equal correlation ρ are

$$\pi^{A_s,A_t}_{a_s,a_t} = \tfrac{1}{4}(1 + \rho\, a_s a_t), \qquad \pi_{ijk} = \tfrac{1}{8}\{1 + \rho i\,(j+k) + jk\}, \tag{13}$$

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 (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 $r(i)$. 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,

$$3\,\eta_{12}\,\eta_{34}, \qquad 15\,\eta_{12}\,\eta_{34}\,\eta_{56}, \qquad 105\,\eta_{12}\,\eta_{34}\,\eta_{56}\,\eta_{78}. \tag{14}$$

We may express the p'th order interaction introduced for even p, and denoted by int_p, directly in terms of ρ as

$$\text{int}_p = \rho^{p/2} \textstyle\prod_{h=0}^{\frac{1}{2}(p-2)} (1+2h)/(1+2h\rho). \tag{15}$$

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

$$\pi_{1\ldots1}^{A_1\ldots A_p} = \begin{cases} 2^{-p}\prod_{h=\frac{1}{2}(p-2)}^{p-2}\{1+(1+2h)\rho\}/\prod_{h=0}^{\frac{1}{2}(p-2)}(1+2h\rho) & \text{if } p \text{ is even,} \\ 2^{-p}\prod_{h=\frac{1}{2}(p-1)}^{p-2}\{1+(1+2h)\rho\}/\prod_{h=0}^{\frac{1}{2}(p-3)}(1+2h\rho) & \text{if } p \text{ is odd,} \end{cases} \tag{16}$$

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

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

The following table compares the orthant probabilities (16) for a few selected values

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

$\text{ods}_{A_{a_s\|1}}$	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	$0.222\bar{2}$	0.5000	$0.777\bar{7}$	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

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 second row contains values given by (16) for the linear triangular system, and the first row contains values given by Matlab for the Gaussian distribution. For $p = 4$, the latter 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 here.

5 Types of graphical Markov models

5.1 A Markov chain and covariance and concentration graph models

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

$$1 \prec\!\!- 2 \prec\!\!- 3 \prec\!\!- 4.$$

In the linear triangular system (1), the constraints are then $0 = \eta_{13} = \eta_{14} = \eta_{24}$, or, equivalently, $1 \perp\!\!\!\perp 34|2$ and $2 \perp\!\!\!\perp 4|3$.

This chain is also reflected in the correlation matrix of the binary variables by a particular pattern of correlations, just like the pattern in correlations for the same Gaussian Markov chain. The correlation matrix $\mathbf{P}$ and the matrix of regression coefficients $\mathbf{A}$ are

$$\mathbf{P} = \begin{pmatrix} 1 & \rho_{12} & \rho_{12}\,\rho_{23} & \rho_{12}\,\rho_{23}\,\rho_{34} \\ . & 1 & \rho_{23} & \rho_{23}\,\rho_{34} \\ . & . & 1 & \rho_{34} \\ . & . & . & 1 \end{pmatrix}, \ \mathbf{A} = \begin{pmatrix} 1 & -\rho_{12} & 0 & 0 \\ . & 1 & -\rho_{23} & 0 \\ . & . & 1 & -\rho_{34} \\ . & . & . & 1 \end{pmatrix}, \ \delta_{ss} = 1 - \rho_{s,s+1}^2.$$

This shows that the correlations corresponding to edges present in the graph match the simple correlations computed from the joint distribution in (1), all others are products of correlations along the path.

Thus, the zero constraints on the regression coefficients turn into more complex constraints on the correlation matrix. One may regard the latter as an instance of under-conditioning, that is of inducing associations by ignoring some variables in conditioning sets $r(s)$ for variable A_s in the generating process that are important intermediate variables. Such variables form a direction-preserving path to A_s in the generating directed graph.

In a covariance graph, an undirected graph with edges shown by dashed lines, an edge present corresponds to a marginal, pairwise association. For symmetric binary variables, a covariance graph model is specified by independence constraints on the the marginal correlations of (3). In the example of the above Markov chain, none of marginal pairwise associations vanishes so that the covariance graph induced by the generating directed graph has no missing edges, i.e. it is complete. It can therefore not be recognized from the two graphs that both model specifications are Markov equivalent, i.e. define the same independence structure for the given type of distribution, and that they are here defined by the same set of three distinct correlations.

By contrast, in a concentration graph, an undirected graph with edges shown by full lines, an edge present corresponds to a conditional association given all of the remaining variables. For discrete variables, these are reflected in the parameters of log-linear models for the joint probabilities, that is in the effect expansion of $\log \pi^{A_1 \ldots A_p}$; see Appendix B. There is conditional independence of a variable pair if the 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 the symmetric binary variables discussed here, zero partial correlation coefficients given all remaining variables are necessary for the corresponding conditional independence statement to hold that is for the joint probabilities to factorize accordingly. These partial correlations may be computed from $\mathbf{P}^{-1}$ with elements ρ^{st}, see for instance Cox and Wermuth (1996) p. 69, as

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

For a Markov chain, there is such a factorization and the induced concentration graph has the same edge set as the Markov chain graph; see the following section.

5.2 A concentration chain

In the concentration graph induced by a Markov chain, there are no additional edges since the concentration graph represents in this case the same independence structure as its generating graph, even though defined differently. The characterizing feature for Markov equivalence of these two types of graph is in general that the generating directed graph is decomposable. This is reflected in the directed graph by the lack of any sink-oriented V-configurations, that is by the lack of any subgraph induced by three nodes, $\circ$, having two edges and arrows pointing to the common neighbor: $\circ\!\!\longrightarrow\!\circ\longleftarrow\!\!\circ$; see also Section 5.6.

Thus, from a decomposable generating graph the induced concentration graph is obtained by replacing each arrow by a full line. In the example of Section 4.1,

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

is the induced concentration graph and each edge present corresponds here to a two-factor term in the effect expansion of $\log \pi^{A_1A_2A_3A_4}$, where

$$\pi_{ijkl}^{A_1A_2A_3A_4} = \tfrac{1}{16}\{(1+\rho_{12}ij)(1+\rho_{23}jk)(1+\rho_{34}kl)\}.$$

Any quadrivariate joint density with the above Markov chain structure factorizes. This is written in a condensed notation of the nodes in the graphs, as

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

Since then $\log f_{1234}$ is the sum of terms involving at most two variables, also a 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 concentration graph; see also (11).

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 corresponds to the marginal association of the corresponding variable pair. In spite of the similarity of this type of undirected graph to the Markov chain of the previous section, maximum-likelihood-estimation under this undirected graph model is quite different. For an exposition in the case of Gaussian distributions; see Wermuth, Cox and Marchetti (2006).

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

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

that is 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 $\mathbf{P}$, correspond to nonlinear constraint in $\mathbf{A}$. In this example, the sets of parameters of the two models are in a one-to-one relation, i.e. they are parameter equivalent and hence also Markov equivalent.

The corresponding directed graph 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 also no direct representation of the structure in the log-linear formulation of $\pi_{ijkl}^{A_1A_2A_3A_4}$ that is in the effect expansion of $\log \pi^{A_1A_2A_3A_4}$. But, there is a simple equivalent multivariate regression chain graph model given in the following section.

5.4 Seemingly unrelated regression and multivariate regression chains

A seemingly unrelated regression model, Zellner (962), that is Markov equivalent to the covariance chain in the previous section has the joint response graph

$$1 \longrightarrow\!\!\!\succ 2 - - - 3 \prec\!\!\!\longleftarrow 4,$$

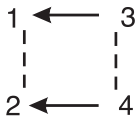
where variables A_1, A_4 are independent explanatory variables for the joint responses A_2, A_3. This implies in particular that A_1, A_4 are both generated before A_2, A_3. The accordingly ordered correlation matrix and this matrix after partial inversion with respect to the explanatory variables are

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

By using the more explicit Yule-Cochran notation, where for instance $\beta_{1|2.3}$ denotes the coefficient of variable A_2 in a linear regression of variable A_1 on A_2, A_3, we have from equations (17), (18) and (20) in Appendix A given the three zero correlations that the nonzero two-factor terms in $\pi^{A_2A_3|A_1A_4}$ 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

$$\pi_{jk|il}^{A_2A_3|A_1A_4} = \tfrac{1}{4}(1 + \rho_{12}ij + \rho_{23}jk + \rho_{34}jl + \rho_{12}\rho_{34}ijkl).$$

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



For the model to this graph, the conditional distribution of the responses has in general the same form as for independent explanatory variables, the case described first in this section. If a conditional distribution is added to a marginal one so that its parameters do not depend on those of the margin, then the joint density gives independently varying factors $f_{12|34}$, f_{34}, where

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

The conditional probabilities are here

$$\pi_{ij|kl}^{A_1A_2|A_3A_3} = \tfrac{1}{4}(1 + \rho_{12}ij + \rho_{13}ik + \rho_{24}jl + \rho_{13}\rho_{24}ijkl),$$

so that the joint distribution is given by

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

and for estimation, each part can be maximized separately.

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

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

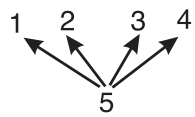
give a factorisation corresponding to the joint or single responses within the chain components a, b, c, d, then each response has as potential explanatory variables the variables in the past but not those within the same chain component nor any variable of the future. Among several responses on equal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

The corresponding chain graph has within each chain component a covariance graph of dashed lines and it 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 Glonek and McCullagh (1995) and 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 if only the latent variable has a sufficiently large number of levels; see Holland and Rosenbaum (1984). But, as we shall show, if all variables including the latent variable are binary and symmetric, then the correlation matrix of the observed variables satisfies the tetrad conditions, just like in the Gaussian case.

As an example we take four variables to be conditionally independent given A_5 so that we have the following decomposable generating graph



The induced correlation matrix $\mathbf{P} = \mathbf{A}^{-1}\mathbf{\Delta}\mathbf{A}^{-\mathrm{T}}$ is

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

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

For instance for the submatrix $\mathbf{P}_{\{1,2\},\{3,4\}}$, the tetrad conditions state that

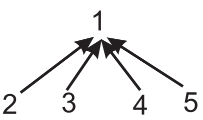
$$\rho_{13}/\rho_{23} = \rho_{14}/\rho_{24},$$

which 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 and a 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 (7), is in this case defined by an identity matrix for $\mathbf{A}_{r(1),r(1)}$ and

$$\mathbf{A}_{1,r(1)} = (-\rho_{12} \quad -\rho_{13} \quad -\rho_{14} \quad -\rho_{15}), \quad \delta_{11} - \textstyle\sum_{t=2}^{t=5} \rho_{1t}^2, \quad \delta_{ss} = 1 \text{ for } s = 2, \ldots, 5.$$

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

Thus, ignoring the information about response and independent explanatory variables by treating all variables on equal standing is an instance of over-conditioning, that is here of inducing associations by enlarging the conditioning sets $r(s)$ of A_s in the generating process to include common responses of A_s and A_t, for $t > s$.

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, different formulations. It is therefore in general rarely useful to restrict model fitting and data analysis to only one class of graphical Markov models. Also, Markov equivalent graphs may be most helpful in suggesting possible alternative interpretations of a given model, but parameter equivalence of two models need not be reflected in the corresponding graphs. Thus by using only graphs, one may overlook some Markov equivalent models.

Appendix A, proof of equations (3), (9)

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

$$\pi_{jkl}^{A_2A_3A_4} = \pi_{j|kl}^{A_2|A_3A_4}\pi_{kl}^{A_3A_4} = \tfrac{1}{8}(1+\eta_{23}jk+\eta_{24}jl)(1+\rho_{34}kl)$$
$$= \tfrac{1}{8}\{1+(\eta_{23}+\eta_{24}\rho_{34})jk+(\eta_{24}+\eta_{23}\rho_{34})jl+\rho_{34}kl\}.$$

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

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

With $\eta_{23}+\eta_{23}\eta_{34} = \rho_{23}$ obtained similarly, equation (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_3 in a linear least-squares regression of variable X_2 on X_3, X_4, one obtains

$$\beta_{2|4} = \beta_{2|4.3}+\beta_{2|3.4}\beta_{3|4}, \tag{17}$$

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 (17).

From (1) and (17), 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

$$\begin{aligned}\pi_{ij|kl}^{A_1A_2|A_3A_4} &= \tfrac{1}{4}\{1+\beta_{1|3.4}ik+\beta_{1|4.3}il+\beta_{2|3.4}jk+\beta_{2|4.3}jl \\ &+ (\eta_{12}+\eta_{13}\eta_{23}+\eta_{14}\eta_{24})ij+(\eta_{13}\eta_{24}+\eta_{14}\eta_{23})ijkl\}.\end{aligned} \tag{18}$$

From

$$\begin{pmatrix} 1 & \rho_{23} & \rho_{24} \\ . & 1 & \rho_{34} \\ . & . & 1 \end{pmatrix}^{-1} (1 \quad \rho_{23} \quad \rho_{24})^{\mathrm{T}} = (1 \quad 0 \quad 0)^{\mathrm{T}}, \tag{19}$$

one gets an extension of Cochran's formula as

$$\eta_{12} + \eta_{13}\rho_{23} + \eta_{14}\rho_{24} = (\eta_{12} \quad \eta_{13} \quad \eta_{14})(1 \quad \rho_{23} \quad \rho_{24})^{\mathrm{T}} = \rho_{12}, . \tag{20}$$

Then, from (18) with $\pi_{ijkl} = \pi_{ij|kl}^{A_1A_2|A_3A_4}\{\frac{1}{4}(1 + \rho_{34}kl)\}$ we have

$$\pi_{ijkl} = \tfrac{1}{16}[1 + \rho_{12}ij + (\beta_{1|3.4} + \beta_{1|4.3}\rho_{34})ik + (\beta_{1|4.3} + \beta_{1|3.4}\rho_{34})il + (\beta_{2|3.4} + \beta_{2|4.3}\rho_{34})jk$$

$$+(\beta_{2|4.3}+\beta_{2|3.4}\rho_{34})jl+\rho_{34}kl+\{\eta_{12}\rho_{34}+\eta_{13}(\beta_{2|4.3}+\beta_{2|3.4}\rho_{34})+\eta_{14}(\beta_{2|4.3}+\beta_{2|3.4}\rho_{34})\}ijkl],$$

and hence, after repeatedly applying Cochran's recursion relation (17), the claimed form of (3).

It is equation (19) which extends directly to higher dimensions. This leads to further generalisations of Cochran's relation, in additon to (20) 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 (14) for the equal correlation case. In general, for instance the six-factor interaction contains terms such as

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

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, $\mathcal{D}_p$, that can for a 2^p table be obtained in terms of a 2×2 matrix $\mathcal{D}_1$ via Kronecker products, where here for effect coding

$$\boldsymbol{\mathcal{D}_1} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \quad \boldsymbol{\mathcal{D}_1^{-1}} = \tfrac{1}{2}\boldsymbol{\mathcal{D}_1}, \quad \boldsymbol{\mathcal{D}_2} = \boldsymbol{\mathcal{D}_1} \otimes \boldsymbol{\mathcal{D}_1}, \quad \boldsymbol{\mathcal{D}_3} = \boldsymbol{\mathcal{D}_1} \otimes \boldsymbol{\mathcal{D}_2}, \; \ldots .$$

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

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

When writing the 2^p probabilities in a column vector $\pi^{A_1,\dots,A_p}$ 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 $\mathcal{C}_p\,\pi^{A_1,\dots,A_p}$ 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 $(\lambda_1^{A_2},\ \ \lambda_{11}^{A_1A_2})$ is added to $(\lambda_-,\ \ \lambda_1^{A_1})$, then $(\lambda_1^{A_3},\ \ \lambda_{11}^{A_1A_3}\ \ \lambda_{11}^{A_2A_3}, \lambda_{111}^{A_1A_2A_3})$ 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 (9), the linear effect parameter expansion is

$$(\mathcal{C}_{\mathbf{3}}\,\pi_{ijkl}^{A_1A_2A_3A_4})^{\mathrm{T}} = \tfrac{1}{8}(1\quad 0\quad 0\quad \rho_{12}\quad 0\quad \rho_{13}\quad \rho_{23}\quad 0)$$

and for the three correlations being nonzero, precisely the entries in position 2,3,5 and 8 are also zero in the log-linear effect expansion that is in $(\mathcal{C}_{\mathbf{3}}\ \log\pi_{ijkl}^{A_1A_2A_3A_4})^{\mathrm{T}}$.

For any constant c, the linear effect terms of $c\,\pi_{ijkl}^{A_1A_2A_3A_4}$ are multiplied by c, while the effect parameters of $\log c\,\pi_{ijkl}^{A_1A_2A_3A_4}$ remain unchanged except that $\log c$ is added to the overall efffect. Thus, for comparisons of linear and log-linear parameters it is sometimes convenient to divide a probability vector by one of its probabilities to obtain ones in some entries. For instance, for model (9) with equal correlations, we have with $\alpha = \pi_{111}/\pi_{-111}$ and $\beta = \log\alpha$

$$(\pi^{A_1,\dots,A_3})^{\mathrm{T}}/\pi_{-111} = (\alpha\ 1\ 1\ 1\ 1\ 1\ 1\ \alpha), \quad \log(\pi^{A_1,\dots,A_3})^{\mathrm{T}}/\pi_{-111} = (\beta\ 0\ 0\ 0\ 0\ 0\ 0\ \beta).$$

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

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.) 158-76. Stanford University Press.

Bartlett, M.S. (1935). Contingency table interactions. Suppl. J. Roy. Statist. Soc., **2**, 248-252.

Birch, M.W. (1963). Maximum likelihood in three-way contingency tables. *J. Roy. Statist. Soc. B* **25**, 220–233.

Cheng, M.C. (1969). The orthant probabilities of four Gaussian variates. *Ann. Math. Statist.* **40**, 152–161.

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. (2007). On a generalization of a result of W.G. Cochran. *Biometrika*, **94**, 400-410.

Cox, D.R. & Wermuth, N. (1994). A note on the quadratic exponential binary distribution. *Biometrika*, **81**, 403-408.

Cox, D.R. & Wermuth, N. (1996). *Multivariate Dependencies – Models,Analysis and Interpretation.* London: Chapman & Hall.

Cox, D.R. & Wermuth, N. (2003). A general condition for avoiding effect reversal after marginalization. *J. Roy. Statist. Soc. B*, **65**, 937-941.

Cramér, H. (1946). *Mathematical methods of statistics.* Princeton University Press.

Darroch, J.N. & Ratcliff, D. (1972) Generalized iterative scaling for log-linear models. *Ann. Math. Statist.* , **43**, 1470-1480.

Darroch, J.N., Lauritzen, S.L. & Speed, T.P. (1980). Markov fields and log-linear models for contingency tables. *Ann. Statist.* **8**, 522–539.

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

Drton, M. (2009). Discrete chain graph models. *Bernoulli.* to appear.

Good, I. J. (1958). The interaction algorithm and practical Fourier analysis. *J. Roy. Statist. Soc. B* **20,** 361-72.

Fienberg, S. (2007). The analysis of cross-classified categorical data. Springer (2007), New York.

Glonek, G.J.N. & McCullagh, P. (1995). Multivariate logistic models. *J. Roy. Statist. Soc. B*, **57**, 533-546.

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.

Holland, P.W. & Rosenbaum, P. (1986). Conditional association and unidimensionality in monotone latent variable models. *Ann. Statist.*, **14**, 1523–1543.

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.

Lazarsfeld, P.F. (1961). The algebra of dichotomous systems. In: Studies in item analysis and prediction. H. Solomon (ed.) 111-157. Stanford University Press.

Ma, Z.M., Xie, X.C. & Geng, Z. (2006). Collapsibility of distribution dependence. *J. Roy. Statist. Soc. B*, **68**, 127-133.

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.

McFadden, J.A. (1955). An approximation for the symmetric quadrivariate normal integral. *Biometrika* **43**, 206-207.

Moran, P.A.P. (1956). The numerical evaluation of a class of integrals. *Proc. Cambridge Philos. Soc., Statist.* **52**, 230–233.

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 especial 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.

Streitberg, B. (1999). Exploring interactions in high-dimensional tables: a bootstrap alternative to log-linear models. *Ann. Statist.*, **27**, 405-413.

Wermuth, N. & Cox, D.R. (1998). On association models defined over independence graphs. *Bernoulli*, **4**, 477-495.

Wermuth, N. & Cox, D.R. (2004). Joint response graphs and separation induced by triangular systems. *J. Roy. Statist. Soc. B*, **66**, 687-717.

Wermuth, N., Cox, D.R. & Marchetti, G.M. (2006). Covariance chains. *Bernoulli*, **12**, 841-862.

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.

Xie, X.C., Ma, Z.M. & Geng, Z. (2008). Some association measures and their collapsibility. *Stat. Sinica.* 1165-1183.

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.