Mixed continuous and binary graphs: The relation between factor and Ising models

Lourens Waldorp

University of Amsterdam, Weesperplein 4, 1018 XA, the Netherlands

Abstract

Undirected graphical models have many applications in such areas as machine learning, image processing, and, recently, psychology. Here we are interested in a typical application from the social sciences: latent variable modeling. In this context it is often of interest to know what the graph on the binary variables looks like when the latent (continuous) variables are unobserved (are marginalized over). Assuming that the continuous variables are Gaussian and the binary nodes have logistic probability conditioned on the continuous variable, we show that the binary nodes adjacent to the continuous nodes, are completely connected whenever the adjacent continuous nodes are connected. Next, we show that the mean and variance parameters of the continuous nodes affect the cliques of the original graph, and that the variance and covariance parameters affect combinations of the cliques. As a special case we obtain the Ising model when the binary nodes only have edges to continuous nodes. We illustrate these results with examples and provide a description of how these results can be applied to distinguishing between a latent variable model and an Ising model.

Key words: undirected graph, mixed graphical model, Ising model, factor model

1 Introduction

Graphical models are popular in many applications such as machine learning, image processing, social science, and recently psychology. In psychology a popular model is the so-called latent variable model (Bollen, 1989). For example, to explain the symptoms of depression (observed variables) a single unobserved (latent) variable is used to explain the dependence structure among the symptoms (Cramer et al., 2010; Borsboom et al., 2011; Schmittmann et al., 2013). If a latent variable is a trait, then such a model is sometimes referred to as an item response model. As another example, consider image analysis with a

Email address: `waldorp@uva.nl` (Lourens Waldorp).

12 March 2015

slice of the brain obtained from a functional magnetic resonance image (fMRI). Here the objective is to know how brain regions communicate to obtain complex behavior. Then, knowing what kind of underlying structure could have created this image could help with dimension reduction. This dimension reduction makes it feasible to construct networks of functionally active brain regions.

Here we investigate the effect of eliminating continuous variables (marginalizing) in an undirected graph with mixed binary and continuous variables. Our main objective is to discover how the parameters of Gaussian variables will affect the graphical structure (nodes and their interactions) of binary variables with logistic probability. The Gaussian variables correspond to unobserved variables and the binary variables to observed variables.

Previous work on mixed (hybrid) graphs of discrete and continuous variables includes the conditional Gaussian (Lauritzen, 1996). A limitation of these networks is that a Gaussian node cannot be a parent of a discrete node. This issue is taken up in several papers, where the joint distribution of Gaussian and discrete variables (with logistic probability function) are approximated by either a quadratic lower bound (Gaussian shape, variational approach) (Murphy, 1999; Lerner et al., 2001) or a mixture of truncated exponentials (Cobb and Shenoy, 2004). In contrast to these advances, we focus here on the exact joint probability and the marginal distribution where only the continuous (Gaussian) variables are integrated out. Our main interest is in how and where in the network of discrete variables the Gaussian parameters affect the resulting marginal distribution. This is more in line with the exact results by Lauritzen (1996), but then on undirected graphs where continuous variables can be parents of discrete variables.

We first use a result by Castillo et al. (1998) where it is shown that only the nodes adjacent to the continuous variables are affected by the marginalization. We then show that marginalization over normally distributed variables which are combined linearly, results in graphs where the mean and variance parameters affect the cliques of the binary variables and the variances and covariance parameters affect the combinations of cliques of the binary nodes. As a special case we show that when all binary nodes are only connected to continuous variables, then the marginal distribution over the binary nodes is an Ising model. We also suggest some hypothesis test to distinguish between modeling with or without a latent variable.

This paper is organized as follows. We first provide some background to undirected graphical models. Then we show that the marginal distribution of the binary nodes is affected only in the nodes adjacent to the continuous nodes (boundary). We use this result to show that if we assume a multivariate normal distribution for the continuous variables and a conditional Bernoulli distribution for the binary variables, then the mean and variance parameters of the multivariate Gaussian are associated with the cliques of the original graph and that the variance and covariance parameters are associated with the combinations of cliques. To illustrate the theory we provide several examples.

Finally, we show how these results can be applied to models in the social sciences.

2 Undirected graphical models

An undirected graphical model or Markov random field is a set of probability distributions representing the structure of some graph G. There are two equivalent ways of defining a Markov random field: (i) in terms of Markov properties and (ii) in terms of the factorization property.

Let $G = (V, E)$ be an undirected graph, where V is the set of nodes $\{1, 2, \dots, p\}$ and $E = V \times V$ is the set of edges $\{(s,t) : s,t \in V\}$, with size $|E| = m$. A subset of nodes Q is a cutset or separator set of the graph if removing Q results in two (or more) components. For instance, Q is a cutset if any path between any two nodes $s \in A$ and $t \in B$ must go through some $q \in Q$. A clique is is a subset of nodes in $C \subset V$ such that all nodes in C are connected, that is, for any $s, t \in C$ it holds that $(s,t) \in E$. A maximal clique is a clique such that including any other node in V will not be a clique.

For an undirected graph G, we associate with each vertex $s \in V$ a random variable $X_s \in \mathcal{X}$ for discrete and $Y_s \in \mathcal{Y}$ for continuous variables. For any subset $A \subset V$ of nodes in we define a configuration $x_A = \{x_s : s \in A\}$. The nodes in V can be associated with both discrete variables X_s with $s \in D \subseteq V$ and continuous variables Y_s with $s \in U \subseteq V$, where $D \cup U = V$ an $D \cap U = \varnothing$. A configuration x_C for a clique C is $\{x_s, s \in C\}$, where C can contain both discrete and continuous variables. An edge set restricted to the edges among a subset $D \subseteq V$ is denoted by E_D. For subsets of nodes A, B, and W, we denote by $X_A \perp\!\!\!\perp X_B \mid X_W$ that X_A is conditionally independent of X_B given X_W.

Definition 2 (Markov property) A random vector X is Markov with respect to G if $X_A \perp\!\!\!\perp X_B \mid X_W$ whenever W is a cutset that yields two disjoint subsets A and B.

For each clique C in the set of all cliques $\mathcal{C}$ of graph G a compatibility function $\psi_C : \mathcal{X}^{|C \cap D|} \cup \mathcal{Y}^{|C \cap U|} \to \mathbb{R}_+$ maps the states of the nodes in clique C to the positive reals. When normalized, the product of the compatibility functions defines the distribution.

Definition 3 (Factorization property) The distribution of the random vector Z factorizes according to graph G if it can be represented by a product of compatibility functions of the cliques

$$p(x) = \prod_{C \in \mathcal{C}} \psi_C(x_C) \tag{1}$$

For strictly positive distributions the Hammersly-Clifford theorem says that the Markov and factorization properties are equivalent (Cowell et al., 1999; Lauritzen, 1996).

Definition 4 (Exponential family) Given a random vector X with values in $\mathcal{X}^p$, let $\phi(x) = (\phi_\alpha : \alpha \in \mathcal{I})$ be a vector of functions (potential functions or sufficient statistics), with $\phi_\alpha : \mathcal{X}^p \to \mathbb{R}$ and $\mathcal{I}$ an index set. Associated to this vector of sufficient statistics is a vector $\eta(\theta)$ of (canonical) parameters. A distribution is exponential family if it has the density function

$$p(x) = \exp\left[\eta(\theta)'\phi(x) - A(\theta)\right] \tag{2}$$

where $A(\theta)$ is known as the log partition or cumulant function and is defined by

$$A(\theta) = \log \int_{\mathcal{X}^p} \exp\left[\eta(\theta)'\phi(x)\right] \nu(dx) \tag{3}$$

The cumulant function ensures that the probability sums to one and can also be used to determine the moments of the distribution $p(x)$ (Koller and Friedman, 2009; Wainwright and Jordan, 2008; Bickel and Doksum, 2007; Brown, 1986). An exponential family distribution is minimal if the set of sufficient statistics $(\phi_\alpha : \alpha \in \mathcal{I})$ are linearly independent.

Example 5 (Ising model) Let $X \in \mathcal{X}^{|D|} = \{0,1\}^{|D|}$ be a binary random vector associated with the graph $G = (D, E)$. If we assume that each X_s is distributed as Bernoulli then we have an exponential family for X. The Ising model is known from statistical physics to model the magnetic field (see e.g., Kindermann et al., 1980; Cipra, 1987; Kolaczyk, 2009). The Ising model considers cliques of sizes one and two nodes only, so the interactions are at most pairwise. Let θ be the parameter vector containing the $|D| + |E|$ parameters. The distribution can be written as

$$p_\theta(x) = \exp\left[\sum_{s\in V}\theta_s x_s + \sum_{(s,t)\in E}\theta_{st}x_s x_t - A(\theta)\right] \tag{4}$$

where

$$A(\theta) := \log \sum_{x\in\{0,1\}^{|D|}} \exp\left[\sum_{s\in V}\theta_s x_s + \sum_{(s,t)\in E}\theta_{st}x_s x_t\right]$$

is the log normalization constant. It is immediate that the Ising model is exponential family with sufficient statistics $\phi(x) = (x_s, s \in D; x_s x_t, (s,t) \in E)$. It is also minimal since the functions in $\phi(x)$ are linearly independent, i.e. $\langle u, \phi(x)\rangle$ is not a constant a.e. for any nonzero $u \in \mathbb{R}^{|D|+|E|}$.

Example 6 (Gaussian random field) Let $Y \in \mathcal{Y}^{|U|} = \mathbb{R}^{|U|}$ be a continuous random vector associated with the graph $G = (U, E)$. If we assume a multivariate Gaussian (normal) distribution for Y with mean μ and covariance Σ, then the usual form of the distribution is

$$p_\theta(y) = c_\theta \exp\left[-\frac{1}{2}(y-\mu)'\Sigma^{-1}(y-\mu)\right]$$

It can be shown that the Markov property results in zeros of the precision (inverse covariance) matrix (Lauritzen, 1996). If $\Gamma_{st} = 0$ in the precision matrix $\Gamma = \Sigma^{-1}$, then s and t are conditionally independent given all other variables and $(s,t) \notin E$. The distribution is again exponential family with the form

$$p_\theta(y) = \exp\left[\sum_{s\in V}(\mu'\Gamma_s y_s + \Gamma_{ss}y_s^2) - \frac{1}{2}\sum_{(s,t)\in E}\Gamma_{st}y_s y_t - A(\theta)\right] \tag{5}$$

where the parameters $\theta = (\mu, \Gamma)$ are collected in a single vector and Γ_s is the sth column of Γ and the log normalization function is

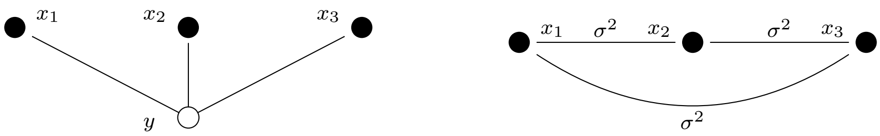
$$\begin{aligned} A(\theta) = \log \int_{\mathbb{R}^p} & \exp\left[\sum_{s\in V}(\mu'\Gamma_s y_s + \Gamma_{ss}y_s^2) + \sum_{(s,t)\in E}\Gamma_{st}y_s y_t\right] dy \\ & \times \exp\left[-\frac{1}{2}\sum_{(s,t)\in E}\Gamma_{st}\mu_s\mu_t\right] \end{aligned}$$

The terms of cross products follow from the Markov property for Gaussian random fields, that $\Gamma_{st} = 0$ whenever $Y_s \perp\!\!\!\perp Y_t \mid Y_{V\setminus\{t,s\}}$. The sufficient statistics for the Gaussian random field are $\phi(x) = (y_s, y_s^2, s \in V; y_s y_t, (s,t) \in E)$.

3 Marginal distribution of the binary variables

We use the factorization of compatibility functions for the cliques to determine the joint distribution of a mixed graph with binary and continuous variables. From this distribution we can obtain the marginal distribution. We first derive a general result from Castillo et al. (1998) that shows that the marginal is changed only at the nodes adjacent to the continuous variables. Then when we additionally assume that the continuous variables are multivariate normal and the binary variables Bernoulli with logistic function, then the resulting marginal is exponential family, and can for some graphs be interpreted as an Ising model.

To show how the continuous variables in the graph affect the marginal distribution we use the factorization of Definition 3 . The compatibility functions are probabilities without normalization. Upon integration over the continuous variables in U we normalize the obtained function to obtain a probability distribution. We first give an example to motivate our main result.



(a) (b)

Fig. 1. Graphs of Example 7 for the distribution $p(x, y)$ (a) and for the marginal $p(x)$ (b). In (a) is the one-factor model with nodes in the boundary ∂G, where the continuous variable separates all nodes x_s with $s \in D$. The marginal $p(x)$ in (b) is seen to follow the distribution from Lemma 10 such that the remaining variables form a clique $C = \{x_1, x_2, x_3\}$.

Example 7 Consider the graph $G = (V, E)$ in Fig. 1. Let $X \in \{0, 1\}^3$ be a binary random vector and $Y \in \mathbb{R}$ is the continuous variable. We identify the set D with the random variables $\{X_1, X_2, X_3\}$ and the set U with $\{Y\}$ for convenience. This graph is known in psychometrics as the one-factor model with three indicators (Bollen, 1989) or an item response model. The continuous variable separates the binary variables, known as local independence. The joint distribution $p_\theta(x, y)$ can be obtained from the factorization property of Definition 3, where the compatibility functions of the cliques are multiplied. The cliques are $C_1 = \{x_1, y\}$, $C_2 = \{x_2, y\}$, and $C_3 = \{x_3, y\}$. Let μ be the mean of the variable Y and σ^2 its variance. For the binary variables we assume the logistic function for the probability of x_s conditional on y, i.e.,

$$p_\theta(x_s \mid y) = \frac{\exp[x_s y]}{1 + \exp[y]} \tag{6}$$

Using the unnormalized probabilities as compatibility functions, the factorization property of Definition 3 gives

$$\begin{aligned} p_\theta(x, y) &\propto \psi_{C_1}(x_1, y)\psi_{C_2}(x_2, y)\psi_{C_3}(x_3, y) \\ &= \exp\left[\sum_{s \in D} x_s y - \frac{1}{2}(y - \mu)^2/\sigma^2\right] \\ &= \exp\left[\left(\mu/\sigma^2 + \sum_{s \in D} x_s\right) y - \frac{1}{2}y^2/\sigma^2 - \frac{1}{2}\mu^2/\sigma^2\right] \end{aligned}$$

Let $\tilde{E}$ denote the set of edges among the nodes in D after marginalizing. Integrating with respect to y using the Gaussian integral with linear component gives the marginal

$$\begin{aligned} p_\theta(x) &= \exp\left[\frac{1}{2}\left(\mu/\sigma^2 + \sum_{s\in D} x_s\right)^2 \sigma^2 - \frac{1}{2}\mu^2/\sigma^2 - A(\theta)\right] \\ &= \exp\left[(\mu + \sigma^2/2)\sum_{s\in D} x_s + \sigma^2 \sum_{(s,t)\in\tilde{E}} x_s x_t - A(\theta)\right] \end{aligned}$$

where $A(\theta)$ is the normalizing constant, and $\tilde{E} = \{(1,2),(2,3),(1,3)\}$. The resulting graph can be seen in Fig. 1. This immediately shows that we are dealing with a specific case of the Ising model, seen in (4).

This example shows that the Gaussian integral including a linear component gives the pairwise interactions of the Ising model. This happens because in the graph corresponding to the joint distribution there were no edges between the binary nodes, and so the only cliques are those including the continuous variable. The effect of marginalizing in completely continuous graphs was given by Castillo et al. (1998). They showed that only the nodes that are adjacent to the nodes over which is marginalized (boundary) are affected.

Definition 4 (Boundary and closure) Given a graph $G = (V, E)$ and a subset $A \in V$, the boundary ∂A of A is the set of nodes that are not in A but have neighbors in A, i.e., $\partial A = \{s \notin A : (t,s) \in E, t \in A\}$. The closure of $A \subseteq V$, denoted $\bar{A}$, is the set A together with the boundary ∂A.

Castillo et al. (1998) also showed that the marginal graph has cliques whenever the nodes in the marginal are connected by adjecent nodes that are connected and subsequently marginalized over. Denote by $s \sim t$ that there is a path between s and t, that is, that there are edges (s, v_1), (v_1, v_2), …, (v_k, t) that connect s to t. Note that a path defines an equivalence relation.

Definition 8 (Connectivity components) A connectivity component τ_s for s is defined as the set of nodes in the graph $G = (V, E)$ that are connected to s, that is $\tau_s = \{t \in V\backslash\{s\} : t \sim s\}$. Let $\mathcal{T}$ be the set of connectivity components in G. For each connectivity component $\tau \in \mathcal{T}$ we define the completed edge set $\tilde{E}_\tau = \{(s,t) : s,t \in \tau\}$ as the edge set with an edge (s,t) for every $s, t \in \tau$ in the same connectivity component. It follows that $\tilde{E}_{\mathcal{T}} = \cup_{\tau\in\mathcal{T}}\tilde{E}_\tau$.

We will be concerned with the marginal of nodes in D. We therefore need to know which nodes in D are connected to some nodes in U of the same connectivity component. We denote by $\mathcal{T}_U$ the connectivity components restricted to U.

Example 9 Consider the graph in Figure 2 where we have nodes in D, (x_1, x_2, x_3) and nodes in U, (y_1, y_2). All nodes are in the same connectiv-

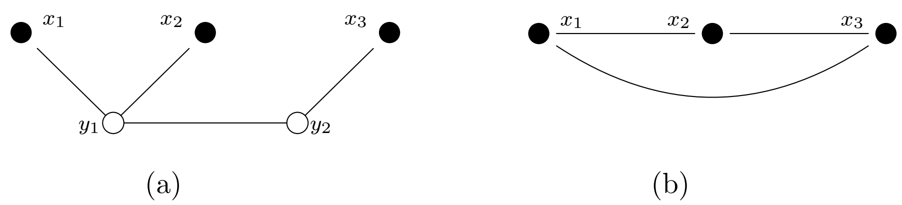


Fig. 2. Graphs of Example 9 for the distribution $p(x, y)$ (a) and for the marginal $p(x)$ (b).

ity component. And so $\mathcal{T} = \{\tau\} = D \cup U$. It follows that the two nodes in U are also in the same connectivity component since there is a (direct) path between y_1 and y_2. Hence $\mathcal{T}_U = U$. The completed edge set of D is $\{(x_1, x_2), (x_1, x_3), (x_2, x_3)\}$, which is the completed edge set of the boundary of the connectivity component in U, i.e., $\tilde{E}_{\partial\tau(U)}$.

We derive the result by Castillo et al. (1998) here in Lemma 1, and show that their result can be applied when marginalizing over continuous nodes in a mixed graph with binary and continuous nodes. All connectivity components are completely in U.

Lemma 10 Let $G = (V, E)$ be an undirected graph with discrete variables $(X_s : s \in D)$ and continuous variables $(Y_s : s \in U)$ and joint distribution $p_\theta(x, y)$, which is positive almost everywhere. If $p_\theta(x, y)$ factorizes according to G, then the marginal distribution $p_\theta(x)$ factorizes according to the graph $\tilde{G}_D = (D, \tilde{E}_D)$, where $\tilde{E}_D = E_D \cup \tilde{E}_{\partial\mathcal{T}_U}$ and the completed edge set for the boundary of the connectivity components is $\tilde{E}_{\partial\mathcal{T}_U} = \cup_{\tau\in\mathcal{T}_U}\{(s, t) : s, t \in \partial\tau\}$.

The proof is in the appendix. Lemma 10 shows that for any connectivity component τ in U the nodes of the binary variables in D will be fully connected if they were connected to that component in U. Considering Example 2 again, Lemma 10 shows that all nodes in D are connected in the marginal, as in Figure 2(b).

We use this result to show how the parameters of the Gaussian variables end up in the marginalized distribution, which in some cases is the Ising model. The assumption of multivariate normality of the continuous variables in U with mean μ and covariance Σ, immediately gives another characterization of the completed edge set $\tilde{E}_{\partial\mathcal{T}_U}$. In fact, any path can be given as a function of the concentration parameters (inverse covariance), and so, any non-zero correlation indicates a path between nodes. It follows that the completed edge set can be described as a clique for all nodes in the same connectivity component.

Lemma 11 Let the nodes $Y \in \mathbb{R}^{|U|}$ be multivariate normally distributed with mean μ and covariance Σ. Then the set of edges from the connectivity components $\tilde{E}_{\mathcal{T}_U}$ that are in U can be characterized as $\tilde{E}_{\mathcal{T}_U} = \{(s, t) : \Sigma_{st} \neq$

$0, s, t \in U\}$.

The proof is in the appendix. Lemma 11 corresponds to the intuition that when two nodes are connected, the correlation (covariance) is nonzero.

We furthermore assume that the discrete variables are conditionally distributed as Bernoulli given the nodes in U, and

$$p(x_s \mid y_{\partial s}) = \frac{\exp(x_s \sum_{t \in \partial s} y_t)}{1 + \exp(\sum_{t \in \partial s} y_t)} \tag{7}$$

This is a common assumption in latent variable modeling (e.g., Bollen, 1989; Holland, 1990) and is in general popular. For our main result we use this logistic function without normalization; we normalize after integrating out the Gaussian variables, as in Example 3. Let $\partial G = \partial U \cup \partial D$ be the union of both boundaries.

The use of the latent variable model led us to combine the binary and continuous variables such that the binary variables in a single clique can be multiplied and that the continuous variables in the boundary ∂D are combined linearly.

Definition 12 (Clique function) The clique function $c_C(x_C, y_C) : \mathcal{X}^{|C \cap D|} \times \mathcal{Y}^{|C \cap \partial D|} \to \mathbb{R}$ maps the states of the discrete and continuous variables adjacent to the discrete variables in clique $C \subseteq \bar{D}$ to the reals by

$$c_C(x_C, y_C) = \prod_{s \in C \cap D} x_s \sum_{t \in C \cap \partial D} y_t$$

The vector of these functions for all cliques is denoted by $c(x, y)$.

This definition ensures that the continuous variables are combined linearly and that the event that all nodes in the clique are 1 if $\{x_s = 1 : \forall s \in C\}$ gives $c_C(x_C) = 1$ and zero otherwise. This gives us the sufficinet statistics $\phi(x)$ for the cliques in the binary variables (see, e.g., Wainwright and Jordan, 2008). Note that in this definition the cliques are either in D or in the set of nodes adjacent to D, the boundary ∂D. This definition corresponds closely to the one for the conditional Gaussian distribution by Lauritzen (1996), except here the graph is undirected. Associated with the clique function is a selection matrix.

Definition 13 Let $K_{\partial G}$ be a selection matrix for nodes in the graph G, such that $c(x)' K'_{\partial G} y$ is the sum of cliques in the boundary set ∂G that satisfies the clique function in Definition 12. Specifically, the vector $(K_D)_s = 1$ if there is a clique $C \subseteq D$ indexed by s in $c(x)$ that is completely in D and 0 otherwise. And $(K_{\partial G})_{st} = 1$ if $C \cap \partial D \neq \varnothing$ and $C \cap \partial U \neq \varnothing$ for this clique indexed by t in the vector $c(x)$ and variable y_s of $s \in C \subseteq U$, and 0 otherwise.

Definition 13 is a convenient way to combine the binary and continuous variables that satisfies the clique function in Definition 12. This allows us to use the multivariate Gaussian integral to obtain the main result. Using Definitions 12 and 13 and the factorization property of Definition 2 we can write the compatibility function $\psi(x, y)$ of both continuous and binary variables as a product of cliques $C \in \mathcal{C}$ of binary nodes in D, of Gaussian nodes in U, and of mixed binary and Gaussian nodes

$$\begin{aligned}\psi(x,y) &= \prod_{C\subseteq D}\psi(x_C)\prod_{C\subseteq D\cup U}\psi(x_C,y_C)\prod_{C\subseteq U}\psi(y_C)\\ &= \exp\left[\sum_{C\subseteq D}c_C(x_C)+\sum_{C\subseteq\partial G}c_C(x_C,y_C)-\frac{1}{2}(y-\mu)'\Sigma^{-1}(y-\mu)\right]\\ &= \exp\left[c(x)'K_D'+(c(x)'K_{\partial G}'+\mu'\Sigma^{-1})y-\frac{1}{2}y'\Sigma^{-1}y-\frac{1}{2}\mu'\Sigma^{-1}\mu\right] \quad (8)\end{aligned}$$

where the last expression is the canonical form used in Lauritzen (1996) for the joint distribution of discrete and Gaussian variables. In Murphy (1999) and Lerner et al. (2001) a variational approach was used as an approximation to the joint probability. In general this is required since then marginalization over both continuous and discrete variables is closed. Another approach to this issue is taken up in Cobb and Shenoy (2004) where a mixture of truncated exponentials is used that also results in closure upon marginalization over both continuous and discrete variables. Here we focus on marginalization over continuous variables only since we aim to investigate the effect of (hidden) Gaussian variables on a discrete network.

Our main result is that the marginal distribution of the binary variables is exponential family, given the assumptions of Gaussian continuous variables. Furthermore, it shows that the mean and variance parameters of the continuous variables in U affect the cliques of the boundary of U, and the covariance parameters affect the interactions between the cliques in the boundary of U that are in the same connectivity component. In some cases the marginal distribution is identical to the Ising model (see Corollary 15).

Theorem 14 Let $G = (V, E)$ be an undirected graph with $V = \{U, D\}$ with continuous variables $Y \in \mathbb{R}^{|U|}$ associated with nodes $s \in U$ and binary variables $X \in \{0, 1\}^{|D|}$ associated with $s \in D$. Let the distribution of Y be multivariate normal with mean μ and covariance Σ, and the distribution of X_s, for $s \in D\backslash\partial U$ is Bernoulli, and for $s \in \partial U$ (nodes adjacent to nodes in U), conditionally distributed as Bernoulli given $y_{\partial s}$, with probability of success the logistic function. A set C represents a set of nodes that form a clique. If $p_\theta(x, y)$ factorizes according to G, then the marginal distribution $p_\theta(x)$ and factorizes over $\tilde{G} = (V, \tilde{E}_D)$, where where $\tilde{E}_D = E_D \cup \tilde{E}_{\partial\mathcal{T}_U}$. This marginal is exponential family over $\{0,1\}^{|D|}$ with the form

$$p_\theta(x) = \exp\Bigg[\sum_{C\subseteq D} c_C(x_C) + \sum_{C\in\partial\mathcal{C}}\sum_{s\in C\cap\partial D}(\mu_s + \Sigma_{ss}/2)c_C(x_C) \\ + \sum_{C_1\neq C_2\in\partial\mathcal{C}}\sum_{s,t\in\tau(C_1\cup C_2)}\Sigma_{st}c_{C_1}(x_{C_1})c_{C_2}(x_{C_2}) - A(\theta)\Bigg] \tag{9}$$

where $A(\theta)$ is the log-partition function

$$A(\theta) = \log\Bigg\{[(2\pi)^{|U|}|\Sigma|]^{1/2} \\ \sum_{x\in\{0,1\}^{|D|}}\exp\Bigg[\sum_{C\subseteq D} c_C(x_C) + \sum_{C\in\partial\mathcal{C}}\sum_{s\in C\cap\partial D}(\mu_s + \Sigma_{ss}/2)c_C(x_C) \\ + \sum_{C_1\neq C_2\in\partial\mathcal{C}}\sum_{s,t\in\tau(C_1\cup C_2)}\Sigma_{st}c_{C_1}(x_{C_1})c_{C_2}(x_{C_2})\Bigg]\Bigg\} \tag{10}$$

which ensures that the sum over all states is 1.

The proof is in Section 4. It can be seen that the nodes that are not adjacent to U are unaffected by marginalization, as in Lemma 1. Additionally, the means and variances of the continuous nodes that are adjacent to nodes in D are seen to affect the binary nodes in the boundary ∂U. And finally, the variances and covariances will affect the edges in the marginal graph $\tilde{G}$ when two nodes are in the same connectivity component τ.

Using Definition 13 for the selection matrices of binary and continuous nodes in cliques, we can rewrite the result in equation (9) as

$$p_\theta(x) = \exp\left[K_D c(x) + \mu' K_{\partial G}c(x) + \frac{1}{2}c(x)' K'_{\partial G}\Sigma K_{\partial G}c(x) - A(\theta)\right] \tag{11}$$

This corresponds to equation (6.2) in Lauritzen (p. 159, 1996). Note that the affine transformations with K_D and $K_{\partial G}$ are not necessarily one-to-one, and so the representation is in general not minimal (Brown, 1986).

One of the interesting consequences of Theorem 14 is that for graphs that only have binary nodes in the boundary of the continuous variables in U and have no edges between them in E_D, the marginal distribution is an Ising model.

Corollary 15 (Ising model) Let G be a graph with binary and continuous nodes such that the graph consists only of the closure of U, that is $\bar{U} = U \cup \partial U$. Assume that the continuous nodes are Gaussian and the discrete nodes are conditional Bernoulli as in Theorem 14. If $E_D = \varnothing$, there are no edges among nodes in D, then the marginal distribution is the Ising model. Furthermore, the mean and variance parameters are associated with the binary variables

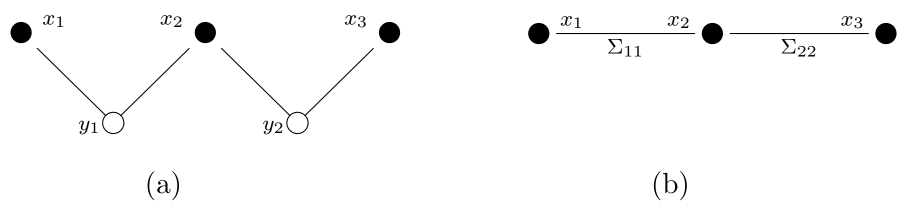


Fig. 2. Graphs of Example 16 for the distribution $p(x, y)$ (a) and for the marginal $p(x)$ (b).

and the covariance parameters are associated with the interactions, i.e.,

$$p_\theta(x) = \exp\left[\sum_{s\in D}\sum_{t\in\partial s}(\mu_t + \Sigma_{tt}/2)x_s + \sum_{u,v\in\partial\tau}\sum_{s,t\in\tau(\partial uv)}\Sigma_{st}x_u x_v - A(\theta)\right] \quad (12)$$

with sufficient statistic $\phi(x) = (x_s, s \in D; x_s x_t, (s,t) \in \tilde{E}_D)'$ and parameters $\theta = (\mu_t + \Sigma_{tt}/2, t \in \partial D \subseteq U; \Sigma_{st}, (s,t) \in E_{\mathcal{T}_U})'$.

In Example 7 we have already seen that Corollary 15 applies. There the marginal graph was a clique (verifying Lemma 1) and there were at most pairwise interactions with the variance of Y as the parameter. This example will be considered later again in the section on applications.

3.1 Examples

We now consider some examples to illustrate the main result in the previous section.

Example 16 Consider the graph in Fig. 2 which is referred to as a two-factor model with cross-loadings (on x_2). Here we have four cliques $C_1 = \{x_1, y_1\}$, $C_2 = \{x_2, y_1\}$, $C_3 = \{x_2, y_2\}$, and $C_4 = \{x_3, y_2\}$. Let $x = (x_1, x_2, x_3)$ and $y = (y_1, y_2)$. The matrix $K_D = 0$ and c is the identity function here, and

$$K_{\partial G} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}$$

Then

$$p_\theta(x, y) \propto \exp\left[(x_1 + x_2)y_1 + (x_2 + x_3)y_2 - \frac{1}{2}(y - \mu)'\Sigma^{-1}(y - \mu)\right]$$

where $\Sigma = \mathrm{diag}(\Sigma_{11}, \Sigma_{22})$. Integrating with respect to (y_1, y_2) and some algebra

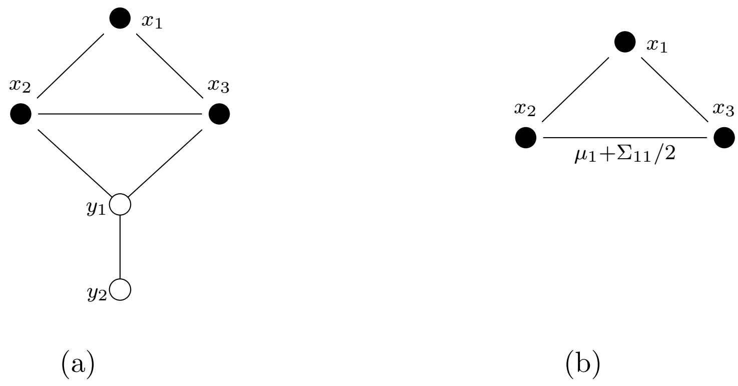


Fig. 3. Graphs of Example 17 for the distribution $p(x, y)$ (a) and for the marginal $p(x)$ (b).

gives

$$\begin{aligned} p_\theta(x) &\propto \exp[(\mu_1 + \Sigma_{11}/2)x_1 + [\mu_1 + \mu_2 + (\Sigma_{11} + \Sigma_{22})/2]x_2 + (\mu_2 + \Sigma_{22}/2)x_3 \\ &\quad + \Sigma_{11}x_1x_2 + \Sigma_{22}x_2x_3] \\ &= \exp\left[\sum_{s\in D} \theta_s x_s + \sum_{(s,t)\in\tilde{E}} \theta_{st} x_s x_t\right] \end{aligned}$$

This again verifies Lemma 10, which specifies that $\tilde{E}_D = \{(x_1, x_2), (x_2, x_3)\}$, and so $(x_1, x_3) \notin \tilde{E}_D$. This is because there are two connectivity components $\tau_1 = \{y_1\}$ and $\tau_2 = \{y_2\}$ in U. Lemma 10 then specifies that the boundaries $\partial\tau_1$ and $\partial\tau_2$ when $1 \nsim 2$, with $1, 2 \in U$, do not form a clique if they were not a clique in the original graph with E. The marginal also has the form of an Ising model, thus verifying Corollary 15.

Example 17 Consider the graph in Fig. 3. The cliques are $C_1 = \{x_1, x_2, x_3\}$, $C_2 = \{x_2, x_3, y_1\}$, and $C_3 = \{y_1, y_2\}$. Then the clique function for nodes s in $D \cup \partial D$ are $c_{C_1}(x_{C_1}) = x_1x_2x_3$ and $c_{C_2}(x_{C_2}, y_{C_2}) = x_2x_3y_1$. Then

$$p_\theta(x, y) \propto \exp\left[x_1x_2x_3 + x_2x_3y_1 - \frac{1}{2}(y-\mu)'\Sigma^{-1}(y-\mu)\right]$$

Let $c(x) = (x_1x_2x_3, x_2x_3)$ and $y = (y_1, y_2)$, and choose $K_D = (1, 0)$ and

$$K'_{\partial G} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$$

Then the marginal is

$$p_\theta(x) = \exp\left[x_1x_2x_3 + (\mu_1 + \Sigma_{11}/2)\, x_2x_3 - A(\theta)\right]$$

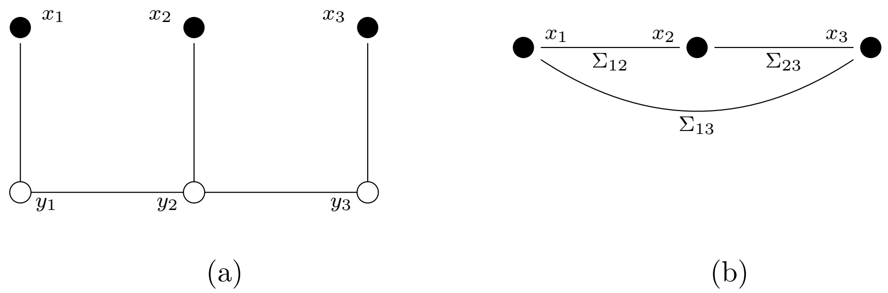


Fig. 4. Graphs of Example 18, (a) for the distribution $p(x, y)$ and (b) for the marginal $p(x)$.

Here we see that there is no conditional independence as expected but that the mean and variance of the Gaussian variable determine the interaction between the nodes adjacent to y_1, which is a result of the marginalization.

Example 18 Consider the graph in Fig. 4. This is a possible representation of a model (hidden Markov model) where mood changes over time continuously according to Y and what can be measured is a binary variable X. The cliques are $C_1 = \{x_1, y_1\}$, $C_2 = \{y_1, y_2\}$, $C_3 = \{x_2, y_2\}$, $C_4 = \{y_2, y_3\}$, and $C_5 = \{x_3, y_3\}$. Then the clique functions for nodes s in $D \cup \partial D$ are $c_{C_s}(x_{C_s}, y_{C_s}) = x_s y_s$. Then

$$p_\theta(x, y) \propto \exp\left[x_1 y_1 + x_2 y_2 + x_3 y_3 - \frac{1}{2}(y - \mu)'\Sigma^{-1}(y - \mu)\right]$$

We know from Lemma 10 that the resulting graph will be completely connected (a single clique) because all nodes in ∂D are in the same connectivity component. In this particular case where each node s in D is connected to a single node t in U, we can use a single index set for both nodes in D and U, say, $\mathcal{I} = \{1, 2, 3\}$. Then the marginal is

$$p_\theta(x) = \exp\left[\sum_{s\in\mathcal{I}}(\mu_s + \Sigma_{ss}/2)x_s + \sum_{s<t\in\mathcal{I}} \Sigma_{st} x_s x_t - A(\theta)\right]$$

From this last equation it is clear that this is the Ising model (Corollary 15) and that this is minimal exponential. This is because we have that the vector of sufficient statistics $\phi(x) = (x_s, s \in D; x_s x_t, (s,t) \in \tilde{E}_D)$ of dimension $|D| + |\tilde{E}_D| = 6$ are linearly independent. The vector of parameters is $\eta = (\mu_s + \Sigma_{ss}/2, s \in D; \Sigma_{st}, (s,t) \in \tilde{E}_D)$ with dimension $|U| + |\tilde{E}_D| = 6$.

4 Proof of Theorem 14

We can focus on the nodes adjacent to U (the boundary of U) since by Lemma 10 we know that the nodes in D that are not adjacent to U will not be directly affected by marginalizing over y. The compatibility functions for each clique that have no nodes in U, $C \cap U = \varnothing$, is $\psi_C = \exp[c_C(x_C)]$. By Definition 12 we have that the nodes $s \in D$ in the boundary ∂U depend only on the nodes in ∂s, and so these variables in D are distributed as $X_s \mid y_{\partial s} \sim \text{Bernoulli}(p)$, with p given in (7). It then follows from Definition 6 that for any clique $C \subseteq \partial G$ such that $C \cap \partial D \neq \varnothing$ and $C \cap \partial U \neq \varnothing$ $(C \in \partial \mathcal{C})$, the compatibility function is

$$\psi_C(x_C, y_C) = \exp\left[\prod_{s \in C \cap D} x_s \sum_{t \in C \cap \partial D} y_t\right] \tag{13}$$

We additionally assumed that the continuous vector $Y \in \mathbb{R}^u$ is multivariate normal with parameters (μ, Σ). The compatibility function for the vector y is then

$$\psi_U(y) = \exp\left[-\frac{1}{2}(y-\mu)'\Sigma^{-1}(y-\mu)\right] \tag{14}$$

where the cliques can be identified by the patterns of zeros in Σ^{-1}. Let $c(x)$ be the vector of cliques $C \subseteq D$. By assumption we can use Definition 2 and multiply the ψ_C for all cliques. If we use the selection matrices K_D and $K_{\partial G}$ we have the function for all cliques

$$\begin{aligned}
\psi(x, y) &= \exp\left[\sum_{C \subseteq D} c_C(x_C) + \sum_{C \subseteq \partial G} c_C(x_C, y_C) - \frac{1}{2}(y-\mu)'\Sigma^{-1}(y-\mu)\right] \\
&= \exp\left[c(x)'K_D' + (c(x)'K_{\partial G}' + \mu'\Sigma^{-1})y - \frac{1}{2}y'\Sigma^{-1}y - \frac{1}{2}\mu'\Sigma^{-1}\mu\right]
\end{aligned}$$

Integrating with respect to y using the multivariate Gaussian integral gives

$$\psi(x) = a(\Sigma)\exp\left[K_D c(x) + \mu' K_{\partial G} c(x) + \frac{1}{2}c(x)'K_{\partial G}'\Sigma K_{\partial G} c(x)\right] \tag{15}$$

where $a(\Sigma) = [(2\pi)^{|U|}|\Sigma|]^{-1/2}$. The matrix $K_{\partial G}$ selects cliques with nodes in both boundaries ∂D and ∂U. Hence, for any node $s \in U$ in clique $C \subseteq \partial G$ we have $\mu_s c_C(x_C)$. And by Lemma 11 we have that for any two nodes s and t in U that are not connected, i.e., $s \nsim t$ and so are in different connectivity components, the covariance $\Sigma_{st} = 0$. Hence, for any $s \neq t \in \tau \cap (C_1 \cup C_2) \subseteq \partial G$ we find $\Sigma_{st} c_{C_1}(x_{C_1}) c_{C_2}(x_{C_2})$. Note that for any $C \subseteq D$, $c_C(x_C)^2 = c_C(x_C)$ because all variables are binary, which explains why the variances (diagonal of Σ) end up in the cliques of the nodes. We conclude that the marginal factorizes according to $\tilde{G}$ with edge set $\tilde{E}_D$, where the cliques in the boundary

∂U are determined by connectivity components in U. From this exposition and equation (15) we can write the marginal as equation (9). This proves the form of the marginal.

To show that the marginal is exponential family, let the vector of sufficient statistics be $\phi(x) = (c_C(x_C), C \in D, c_{C_1}(x_{C_1})c_{C_2}(x_{C_2}), C_1, C_2 \in D)$. The dimension of $\phi(x)$ is $k + q$, where k is the number of cliques in D and q is the number of combinations of cliques in D. Let the vector of parameters be $\theta = (\mu_s, s \in U; \Sigma_{st}, s, t \in E_{\mathcal{T}_U})$ with dimension $|U| + |E_{\mathcal{T}_U}|$. The reparameterization for exponential family is $\eta : \mathbb{R}^{|U|+|E_{\mathcal{T}_U}|} \to \mathbb{R}^{k+q}$, which is what $K_{\partial G}$ does. We can rewrite the exponential in (15) as the inner product with terms of single cliques $\langle K_D + \mu' K_{\partial G}, c(x)\rangle$ and combinations of cliques $\langle K'_{\partial G}\Sigma K_{\partial G}, c(x)c(x)'\rangle$, where the last is the matrix inner product. Then the parameter vector is $\eta(\theta) = (K_D + \mu' K_{\partial G}, K'_{\partial G}\Sigma K_{\partial G})$ which is a linear transformation of the Gaussian parameters. This representation complies with Definition 3 of exponential family. This proves the theorem. □

5 Application to factor analysis and Ising model

One application is the so-called factor model, a model that (usually) has continuous latent variables explain the observed correlation pattern in the data. Examples of factor models are given in Fig. 1 and 2. One of the issues in modeling is whether the factor model is appropriate; perhaps an Ising model would be better. Here we suggest a way to answer this question by a test on the parameters of the Ising models.

In a factor model, the probability $p(x_s \mid y)$ of a correct ($X_s = 1$) or incorrect ($X_s = 0$) response on x_s is modeled by a linear combination of continuous (latent) variables, called factors using the logistic function in (7). For a one-factor model the log-odds of the two possible probabilities (logit) reveals the linear structure

$$\text{logit}(x_s \mid y) = \log \frac{p(x_s \mid y)}{1 - p(x_s \mid y)} = \theta_0 + \theta_1 x_s y$$

Then the logit of a correct response on x_s is the linear function $\theta_0 + \theta_1 y$ and the logit of an incorrect response is θ_0. In the psychometrics literature this model with θ_0 fixed over all $s \in D$ is known as the Rasch model (de Boeck and Wilson, 2004). This model can easily be extended to more than one factor by a linear combination of the continuous variables. This idea corresponds to our Definition 12 of the clique function to combine binary and continuous variables. Definition 12 is more general in that it incorporates cliques that have more than one binary variable. Estimates of the parameters can be obtained in the generalized linear models framework (McCullogh and Searle, 2001).

An important assumption in these models is that the binary variables are independent given the continuous variable(s), sometimes known as local independence (Bollen, 1989). This assumption can easily be tested using the ideas presented here. If there is a one-factor model (a single continuous variable) underlying the observed configuration x, then the marginal distribution $p(x)$ is complete (fully connected) and each edge has the same parameter. This is formally stated in Corollary 19.

Corollary 19 Let $G = (V, E)$ be a graph where each node $s \in D$ has one edge that connects it to the only node $t \in U$ (see Fig. 1), with y_t distributed as $N(\mu, \sigma^2)$. If the assumptions in Theorem 14 hold, then

(a) the marginal graph $\tilde{G}$ is complete
(b) the marginal distribution $p(x)$ is an Ising model
(c) all parameters for the nodes in D are $\mu + \sigma^2/2$
(d) all parameters for pairwise interactions in $\tilde{E}_D$ are σ^2.

This corollary implies that in a test for a one-factor model one simply has to consider the equivalence of the interaction paramters θ_{st}, for all $(s,t) \in \tilde{E}_D$. This assumption is therefore relatively easy to check. It is possible to estimate the parameters θ in the Ising model using a generalized linear model, as described in Bühlmann and van de Geer (2011), and then test the hypothesis that $H_0 : \theta = \alpha 1$, where 1 is the vector $(1, 1, \ldots, 1)$. If it is true, then a one-factor model will do equally well as an Ising model where all nodes are connected. If the hypothesis is rejected, then an Ising model is a better description than a factor model.

This example was relatively simple because it was assumed that the edges between the nodes in D and the single node in U were all identical. A related result assuming different parameters for each of the edges, say, $\alpha = (\alpha_{y1}, \alpha_{y2}, \alpha_{y3})$, as shown in Fig. 5, is easily seen to result in the graph in the right panel. This is easily seen from Corollary 19. We know from the clique function in Definition 6 that the cliques over the boundary of D have an additional parameter such that for $s \in C \cap D$, $c_C(x_C, y_C) = x_s \alpha_{ys} y$. From the basic rules for linear combinations of random normal variables we then have the marginal

$$p(x) = \exp\left[\sum_{s \in D}(\alpha_{ys}\mu + \alpha_{ys}^2\sigma^2/2)x_s + \sum_{(s,t)\in\tilde{E}_D} \alpha_{ys}\alpha_{yt}\sigma^2 x_s x_t - A(\theta, \alpha)\right] \quad (16)$$

which gives the desired result.

We can now test for equality of coefficients $\alpha_{sy} = \alpha_{ty}$ for any (s,t), even without knowledge of the variance of the latent variable or the parameters in α. Let $\gamma_{st} = \alpha_{ys}\alpha_{yt}\sigma^2$. Then under $H_0 : \alpha_{ys} = \alpha, \forall s \in D$, we have $\gamma_{st} = \gamma_{su}$ is equivalent to $\alpha_{yt} = \alpha_{yu}$. Hence we can use the estimates of γ to test for equality of coefficients in α, which indicates similar connection strengths of the latent variable to the binary, observed variables.

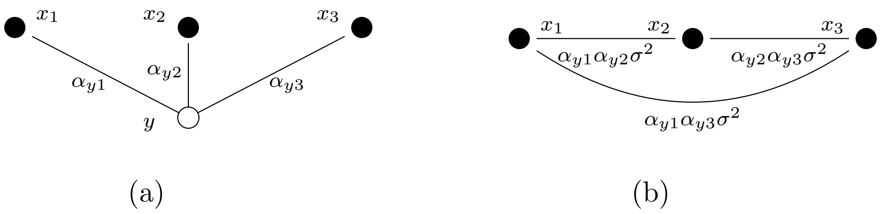


Fig. 5. Graphs of a factor model (a), including different coefficients between the continuous node y and the binary nodes x, and for the marginal $p(x)$ in (b).

6 Discussion

We have shown that for a graph with mixed Gaussian and Bernoulli variables the effect of marginalization can be described in terms of the means, variances, and covariances of the Gaussian variables in the marginal network of binary nodes. Specifically, the means and variances end up as parameters of the nodes in the marginal graph, and the variances and covariances as parameters of the edges.

One interesting consequence is that a common model in the social sciences, the latent variable model, reverts to an Ising model when the continuous variable(s) is (are) marginalized over. The latent variable is a convenient construct and from the theory presented here it can be tested whether an underlying structure, like a single latent variable, is plausible.

Appendix

Proof of Lemma 10 The marginal $p(x)$ is obtained by integrating over $(y_s : s \in U)$, i.e., only the nodes of continuous Gaussian variables in U. By assumption the factorization for $p(x, y)$ holds. We therefore need only consider integration over the compatibility functions for cliques that have nodes in U. That is

$$\int \prod_{C \in \mathcal{C}} \psi_C(x_C, y_C) dy = \prod_{C \cap U = \varnothing} \psi_C(x_C) \int \prod_{C \cap U \neq \varnothing} \psi_C(x_C, y_C) dy$$

No clique C in $\mathcal{C}$ can have nodes that are in two different connectivity components τ_s and τ_t of $\mathcal{T}_U$ when $t \nsim s$. Hence, for any clique C we need to integrate only over the nodes that are in the same connectivity component τ. Hence

$$\int \prod_{C \cap U \neq \varnothing} \psi_C(x_C, y_C) dy = \prod_{\tau \in \mathcal{T}_U} \int \prod_{C \cap \tau \neq \varnothing} \psi_C(x_C, y_C) dy_\tau$$

where y_τ refers to the variables of nodes in the connectivity component τ. Since integration over all y_s for $s \in U$ is performed, we have that the integral $\int \prod_{C \cap \tau \neq \varnothing} \psi_C(x_C, y_C) dy_\tau$ is a function of nodes in D that are in the boundary of $\tau \in U$ only for each C. This is because the only nodes in D that are in $C \cap \tau$ are adjacent to U; otherwise the clique C would contain only nodes of D and would be in the first part of the marginal distribution. Hence, the compatibility function becomes $\psi_{\partial\tau}(\partial\tau)$, a function of the boundary of the connectivity component τ. Therefore,

$$\prod_{\tau \in \mathcal{T}_U} \int \prod_{C \cap \tau \neq \varnothing} \psi_C(x_C, y_C) dy_\tau = \prod_{\tau \in \mathcal{T}_U} \psi_{\partial\tau}(\partial\tau)$$

For any connectivity component τ the integral term $\psi_{\partial\tau}(\partial\tau)$ forms a clique on the boundary $\partial\tau$ and need not be factorized further. This shows that the marginal $p(x)$ factorizes according to the graph $\tilde{G}_D = (V, \tilde{E}_D)$, where $\tilde{E}_D = E_D \cup \tilde{E}_{\partial\mathcal{T}_U}$ and $\tilde{E}_{\partial\mathcal{T}_U} = \cup_{\tau \in \mathcal{T}_U} \{(s,t) : s,t \in \partial\tau\}$ is the completed edge set with all edges in the boundaries $\partial\tau$ for all connectivity components $\tau \in \mathcal{T}_U$. □

Proof of Lemma 11 We use a theorem by Jones and West (2005) that characterizes a covariance Σ_{st} as a product of elements of the concentration matrix $\Gamma = \Sigma^{-1}$. Let $\Gamma_{\setminus P}$ be the concentration matrix with rows and columns deleted in the set of nodes in P (a path). Denote for a path between s and t of length k, $P = \{(s, v_1), (v_1, v_2), \dots, (v_k, t)\}$

$$d_P = (-1)^{k+1} \gamma_{sv_1} \gamma_{v_1 v_2} \cdots \gamma_{v_k t} |\Gamma_{\setminus P}|$$

Then Theorem 1 of Jones and West (2005) states that the covariance between s and t is

$$\Sigma_{st} = \frac{1}{|\Gamma|} \sum_{P \in \mathcal{P}} d_P$$

If s and t are not in the same connectivity component then in each path P there will be a $\gamma_{v_i v_j} = 0$. And so for each P, $d_P = 0$. Hence, the covariance $\Sigma_{st} = 0$ if and only if there is no path between s and t. □

Proof of Corollary 15 By Theorem 14 we know that $C \subseteq D = \varnothing$ and so the first term in the exponential function is zero. The cliques in both boundaries ∂D and ∂U have by assumption no edges in D, and so each clique in this graph has at most one node in D. Hence, for any $s \in C$ we have $c_C(x_C, y_C) = x_s \sum_{t \in C \cap \partial D} y_t$. From this we see that the mean and variances of the y_t in the boundary ∂D are multiplied with the nodes in D of the same clique, which explains the second term in the exponential. Since each clique function contains only a single binary node, it is also obvious that interactions in terms of squares between the cliques in the closure of D will result in pairs of variables $x_s x_t$. This explains the third term. □

Proof of Corollary 19 (a) Since there is only a single continuous variable in U, there is only one connectivity component. From Lemma 10 we immediately have that the marginal graph $\tilde{G}$ is complete. (b) It is easy to see that this is a special case of Corollary 15, and so the marginal is an Ising model. (c) Since there is only one node in U, Corollary 15 implies that each node in D has parameter $\mu + \sigma^2/2$. (d) Again from Corollary 15 it follows that each interaction term $x_s x_t$ has parameter σ^2. $\square$

References

Bickel, P., Doksum, K., 2007. Mathematical Statistics: Basic ideas and selected topics. Pearson Prentice Hall.

Bollen, K., 1989. Structural equations with latent variables. New York: John Wiley and Sons.

Borsboom, D., Cramer, A., Schmittmann, V., Epskamp, S., Waldorp, L., 2011. The small world of psychopathology. PLoS ONE 6 (11), e27407.

Brown, L., 1986. Fundamentals of statistical exponential families. Institute of Mathematical Statistics.

Bühlmann, P., van de Geer, S., 2011. Statistics for High-Dimensional Data: Methods, Theory and Applications. Springer.

Castillo, E., Ferrandiz, J., Sanmartin, P., 1998. Marginalizing in undirected graphs and hypergraphs. In: Proceedings of the Fourteenth conference on Uncertainty in Artificial Intelligence. Morgan Kaufmann Publishers, pp. 69–78.

Cipra, B. A., 1987. An introduction to the ising model. The American Mathematical Monthly 94 (10), 937–959.

Cobb, B., Shenoy, P., 2004. Hybrid influence diagrams using misxtures of truncated exponentials. In: Chickering, M., Halpern, J. (Eds.), Proceedings of the Conference on Uncertainty in Artificial Intelligence. pp. 85–93.

Cowell, R., Dawid, A., Lauritzen, S., Spiegelhalter, D., 1999. Probabilistic networks and expert systems. Springer.

Cramer, A. O. J., Waldorp, L. J., van der Maas, H. L. J., Borsboom, D., 2010. Comorbidity: A network perspective. Behavioral and Brain Sciences 33 (2-3), 137–150.

de Boeck, P., Wilson, M., 2004. Explanatory item response models: A generalized linear and nonlinear approach. Springer.

Holland, P. W., 1990. The dutch identity: A new tool for the study of item response models. Psychometrika 55 (1), 5–18.

Jones, B., West, M., 2005. Covariance decomposition in undirected gaussian graphical models. Biometrika 92 (4), 79–786.

Kindermann, R., Snell, J. L., et al., 1980. Markov random fields and their applications. Vol. 1. American Mathematical Society Providence, RI.

Kolaczyk, E. D., 2009. Statistical analysis of network data: Methods and models. Springer, New York, NY.

Koller, D., Friedman, N., 2009. Probabilistic graphical models: Principles and techniques. MIT Press.

Lauritzen, S., 1996. Graphical Models. Oxford University Press.

Lerner, U., Segal, E., Koller, D., 2001. Exact inference in networks with discrete children of continuous parents. In: Proceedings of the 17th Conference on Uncertainty in Artificial Intelligence. pp. 319–328.

McCullogh, C., Searle, S., 2001. Generalized, linear, and mixed models. New York: John Wiley & Sons.

Murphy, K., 1999. A variational approach for bayesian networks with discrete and continuous latent variables. In: Proceedings of the Conference on Uncertainty in Artificial Intelligence. pp. 467–475.

Schmittmann, V. D., Cramer, A. O., Waldorp, L. J., Epskamp, S., Kievit, R. A., Borsboom, D., 2013. Deconstructing the construct: A network perspective on psychological phenomena. New Ideas in Psychology 31 (1), 43–53.

Wainwright, M. J., Jordan, M. I., 2008. Graphical models, exponential families, and variational inference. Foundations and Trends in Machine Learning 1 (1-2), 1–305.