Bernoulli **18**(1), 2012, 290–321
DOI: 10.3150/10-BEJ338

Tree cumulants and the geometry of binary tree models

PIOTR ZWIERNIK* and JIM Q. SMITH**

Department of Statistics, University of Warwick, Coventry CV4 7AL, UK.
E-mail: *piotr.zwiernik@gmail.com; **j.q.smith@warwick.ac.uk

In this paper we investigate undirected discrete graphical tree models when all the variables in the system are binary, where leaves represent the observable variables and where all the inner nodes are unobserved. A novel approach based on the theory of partially ordered sets allows us to obtain a convenient parametrization of this model class. The construction of the proposed coordinate system mirrors the combinatorial definition of cumulants. A simple product-like form of the resulting parametrization gives insight into identifiability issues associated with this model class. In particular, we provide necessary and sufficient conditions for such a model to be identified up to the switching of labels of the inner nodes. When these conditions hold, we give explicit formulas for the parameters of the model. Whenever the model fails to be identified, we use the new parametrization to describe the geometry of the unidentified parameter space. We illustrate these results using a simple example.

Keywords: binary data; central moments; conditional independence; cumulants; general Markov models; graphical models on trees; hidden data; identifiability; Möbius function

1. Introduction

Discrete graphical models have become a very popular tool in the statistical analysis of multivariate problems (see, e.g., [7,19]). When all the variables in the system are observed, they exhibit a useful modularity. In particular, it is possible to estimate all the conditional probabilities that parametrize such models, maximum likelihood estimates are simple sample proportions and a conjugate Bayesian analysis is straightforward. However, if the values of some of the variables are unobserved, then the resulting model for the observed variables often becomes very complex, making inference much more difficult.

The complicated structure of models with hidden variables usually leads to difficulties in establishing the identifiability of their parameters (see, e.g., [1]). In this paper, we show how algebraic and combinatorial techniques can help. We focus on graphical models where the underlying graph is a tree and all the inner nodes represent hidden variables. In the computational biology literature, these models are called the general Markov models (see, e.g., [14]), tree models or tree decomposable distributions (cf. [10]). Building on results of Chang [4], in this paper we analyze issues associated with identifiability of such a tree model when all its variables are binary, paying particular attention to the geometry of the unidentified space. In particular, we obtain necessary and sufficient conditions for this model to be locally identified, which gives a stronger version of Theorem 4.1 in [4]. When these conditions are satisfied, we also obtain exact formulae for its parameters in terms of the marginal distribution over the observed variables.

1350-7265

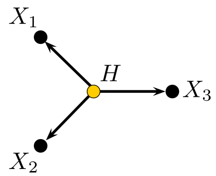


Figure 1. The tripod tree model.

Our strategy is to define a new parametrization of this model class. The new coordinate system is based on moments rather than conditional probabilities. This helps us to exploit various invariance properties of tree models, which, in turn, enables us to express the dependence structure implied by the tree more elegantly. Furthermore, because the parametrization is based on well-understood moments, the implied dependence structure becomes more transparent.

The motivation of this methodology sprung from the study of the tripod tree model, which is the simplest naive Bayes model. The model is a graphical model given in Figure 1, where the black nodes represent three observed variables, X_1, X_2, X_3, and the white node indicates a hidden variable H that remains *hidden*; that is, its values are never directly observed. We assume all the variables in the system have values in $\{0, 1\}$. For $\alpha = (\alpha_1, \alpha_2, \alpha_3) \in \{0, 1\}^3$ let $p_\alpha = \mathbb{P}(X_1 = \alpha_1, X_2 = \alpha_2, X_3 = \alpha_3)$. This model would usually be parametrized using conditional probabilities. In this case we would write

$$p_\alpha = \sum_{i=0}^{1} \theta_i^{(h)} \theta_{\alpha_1|i}^{(1)} \theta_{\alpha_2|i}^{(2)} \theta_{\alpha_3|i}^{(3)}, \tag{1}$$

where $\theta_i^{(h)} = \mathbb{P}(H = i)$ and $\theta_{\alpha_j|i}^{(j)} = \mathbb{P}(X_j = \alpha_j | H = i)$. It can be seen that there are seven free parameters needed to specify p_α, namely: $\theta_1^{(h)}$ together with $\theta_{1|i}^{(j)}$ for $i = 0, 1$ and $j = 1, 2, 3$.

However, the definition of this model given in (1) becomes more transparent when expressed in terms of moments. It is easy to check that there is a one-to-one correspondence between the probabilities p_α for $\alpha \in \{0, 1\}^3$ and the four central moments $\mu_{ij} = \mathbb{E}(X_i - \lambda_i)(X_j - \lambda_j)$ for $i, j = 1, 2, 3$ and $\mu_{123} = \mathbb{E}(X_1 - \lambda_1)(X_2 - \lambda_2)(X_3 - \lambda_3)$ supplemented by the three means $\lambda_i = \mathbb{E}X_i$ for $i = 1, 2, 3$ (cf. Appendix A.1).

Let $\bar{\mu}_h = 1 - 2\theta_1^{(h)}$, $\bar{\mu}_i = 1 - 2\lambda_i$ and $\eta_{h,i} = \theta_{1|1}^{(i)} - \theta_{1|0}^{(i)}$ for $i = 1, 2, 3$. We can now write an explicit isomorphism between the original seven parameters $(\theta_1^{(h)}, (\theta_{1|0}^{(i)}, \theta_{1|1}^{(i)}))$ and new parameters $(\bar{\mu}_h, (\bar{\mu}_i), (\eta_{h,i}))$ for $i = 1, 2, 3$. Thus, in [15], it is shown that in the new coordinate system, together with the new parameters, the model class is equivalently given by

$$\begin{aligned} \lambda_i &= \tfrac{1}{2}(1 - \bar{\mu}_i) \qquad \text{for } i = 1, 2, 3, \\ \mu_{ij} &= \tfrac{1}{4}(1 - \bar{\mu}_h^2)\eta_{h,i}\eta_{h,j} \qquad \text{for all } i \neq j \in \{1, 2, 3\} \quad \text{and} \\ \mu_{123} &= \tfrac{1}{4}(1 - \bar{\mu}_h^2)\bar{\mu}_h \eta_{h,1}\eta_{h,2}\eta_{h,3}. \end{aligned} \tag{2}$$

The product-like form of this parametrization enables us to see various interesting constraints on the observed nodes. For example, by multiplying formulae for μ_{12}, μ_{13} and μ_{23} in (2) together we can see that $\mu_{12}\mu_{13}\mu_{23} \geq 0$ must hold. It also allows us to find explicit formulae for the parameters of the model in terms of the marginal distribution on the set of observed variables. For example, when $\mu_{12}\mu_{13}\mu_{23} \neq 0$ by substituting (2) for all the observed moments, we see that

$$\bar{\mu}_h^2 = \frac{\mu_{123}^2}{\mu_{123}^2 + 4\mu_{12}\mu_{13}\mu_{23}}, \qquad \eta_{h,i}^2 = \frac{\mu_{123}^2 + 4\mu_{12}\mu_{13}\mu_{23}}{\mu_{jk}^2} \qquad \text{for } i = 1, 2, 3. \tag{3}$$

Now a similar parametrization is known for general naive Bayesian models; see the Appendix in [6]. The new parametrization for this model class was used in [13] to approximate a marginal likelihood where the sample size was large, in [6] to understand the local geometry of the model class and in [2] to provide the full description of these models in terms of the defining equations and inequalities.

Naive Bayesian models are a particular example of general Markov models. The class of tree models is somewhat more complicated than the naive Bayesian models and needs new tools to examine its geometry. In this paper, we investigate the moment structures induced by tree models using the theory of partially ordered sets and Möbius functions. Similar methods were used in the combinatorial theory of cumulants (see [12,17]) for a poset of all partitions of a finite set. To our knowledge, this paper is the first to use more general posets in statistical analysis, although a similar approach can be found in the theory of free probability (see, e.g., [18]).

The paper is organized as follows. In Section 2 we define and analyze the moment structures of the class of models under consideration. In Section 3 we define tree-cumulants, which form a new coordinate system for this model class. In Section 4 we reparametrize the model and show that the induced parametrization on the observed margin has an elegant product-like form. We apply this reparametrization in Section 5, analyzing the local geometry of the tree models and the geometry of the subsets of the parameter space that give the same set of marginal distributions on the set of observed variables. In Section 6 we illustrate this method using a simple general Markov model given by a tree with two hidden nodes.

2. Independence models on trees

In this section, we introduce models defined by global Markov properties on trees.

2.1. Preliminaries on trees

A *graph* G is an ordered pair (V, E) consisting of a non-empty set V of *nodes* (or *vertices*) and a set E of *edges*, each of which is an element of $V \times V$. An edge $(u, v) \in E$ is *directed* if the pair (u, v) is ordered and we represent the edge by an arrow from u to v. If (u, v) is not an ordered pair, then we say that (u, v) is an *undirected edge*. Graphs with only (un)directed edges are called (un)directed. If $e = (u, v)$ is an edge of a graph G, then u and v are called *adjacent* and e is said to be *incident with* u and v. If $v \in V$, the *degree* of v is denoted by $\deg(v)$, and is the number of edges incident with v. A *path* in a graph G is a sequence of nodes $(v_1, v_2, \dots, v_k)$ such that, for all $i = 1, \dots, k-1$, v_i and v_{i+1} are adjacent. If, in addition, $v_1 = v_k$, then the path is called a *cycle*. A graph is *connected* if each pair of nodes in G can be joined by a path.

A (*directed*) *tree* $T = (V, E)$ is a connected (*directed*) graph with no cycles. A node of T of degree one is called a *leaf*. A node of T that is not a leaf is called an *inner node*. An edge e of T is *inner* if both nodes incident with e are inner nodes. A connected subgraph of T is a *subtree* of T. A *rooted tree*, T^r, is a directed tree that has one distinguished node called the *root*, denoted by the letter r, and edges that are directed away from r. Let T^r be a rooted tree. For every node v of T^r we let $\mathrm{pa}(v)$ denote the set of nodes u such that $(u, v) \in E$. If v is the root, then $\mathrm{pa}(v) = \varnothing$. Otherwise $\mathrm{pa}(v)$ is a singleton.

For any $W \subseteq V$ we define $T(W)$ as the minimal subtree of T whose set of nodes contains W. We say that $T(W)$ is the subtree of T *spanned on* W. Henceforth, denote the edge set of $T(W)$ by $E(W)$ and its set of nodes by $V(W)$. If T is rooted, then let $r(W)$ denote the unique node v of $T(W)$ such that $\mathrm{pa}(v) \cap V(W)$ is the empty set.

Let $T = (V, E)$ be a tree where $e = (u, v)$ denotes one of its edges. Then *contracting* e results in another tree, denoted by T/e, with the edge e removed and its incident nodes u and v identified. Similarly, for any $E' \subseteq E$ we denote the tree obtained from T by contracting all edges in E' by T/E'. If $v \in V$ such that $\deg v = 2$, then to *suppress* v we simply contract one of the edges incident with v. The resulting tree is denoted by T/v.

2.2. Models defined by global Markov properties

In this paper, we always assume that random variables are binary, taking either value 0 or 1. The vector Y has as its components all variables in the graphical model, that is, both hidden and observed variables. Denote the subvector of Y of observed variables by X and the subvector of hidden variables by H.

Let $T = (V, E)$ be an undirected tree. For any three disjoint subsets $A, B, C \subseteq V$ we say that C *separates* A and B in T, denoted by $A \perp_T B|C$, if each path from a node in A to a node B passes through a node in C. For any $A \subseteq V$ let Y_A denote the subvector of $Y = (Y_v)_{v \in V}$ with elements indexed by A, that is, $Y_A = (Y_v)_{v \in A}$. We are interested in statistical models for Y defined by global Markov properties (GMP) on T. By definition (see, e.g., [7], Section 3.2.1), these models are specified through the set of conditional independence statements of the form:

$$\{Y_A \perp\!\!\!\perp Y_B | Y_C\colon \text{ for all } A, B, C \subset V \text{ s.t. } A \perp_T B|C\}. \tag{4}$$

Let $\widetilde{\mathcal{M}}_T$ denote the space of probability distributions of (X, H) satisfying the global Markov properties on T. We now let $\mathcal{M}_T$ denote the space of marginal probability distribution on X induced from distributions over (X, H), which are in $\widetilde{\mathcal{M}}_T$.

2.3. Models for rooted trees

We next present the parametric formulation of the models presented in the previous section. A Markov process on a rooted tree T^r is a collection of random variables, $\{Y_v\colon v \in V\}$, such that for each $\alpha = (\alpha_v)_{v\in V} \in \{0,1\}^V$

$$p_\alpha(\theta) = \prod_{v\in V} \theta^{(v)}_{\alpha_v|\alpha_{\mathrm{pa}(v)}}, \tag{5}$$

where $\mathrm{pa}(r)$ is the empty set, $\theta = (\theta^{(v)}_{\alpha_v|\alpha_{\mathrm{pa}(v)}})$ and

$$\theta^{(v)}_{\alpha_v|\alpha_{\mathrm{pa}(v)}} = \mathbb{P}\big(Y_v = \alpha_v | Y_{\mathrm{pa}(v)} = \alpha_{\mathrm{pa}(v)}\big).$$

Since $\theta^{(r)}_0 + \theta^{(r)}_1 = 1$ and $\theta^{(v)}_{0|i} + \theta^{(v)}_{1|i} = 1$ for all $v \in V \setminus \{r\}$ and $i = 0, 1$, the set of parameters consists of exactly $2|E| + 1$ free parameters: we have two parameters, $\theta^{(v)}_{1|0}$, $\theta^{(v)}_{1|1}$, for each edge $(u, v) \in E$ and one parameter, $\theta^{(r)}_1$, for the root. We denote the parameter space by $\Theta_T = [0,1]^{2|E|+1}$.

Suppose that T^r has n leaves representing a binary random vector, $X = (X_1, \dots, X_n)$, and let

$$\Delta_{2^n-1} = \Big\{p \in \mathbb{R}^{2^n}\colon \sum_\beta p_\beta = 1, p_\beta \geq 0\Big\} \tag{6}$$

with indices β ranging over $\{0,1\}^n$ be the probability simplex of all possible distributions of X. Equation (5) induces a polynomial map, $f_T : \Theta_T \to \Delta_{2^n-1}$, obtained by marginalization over all the inner nodes of T, giving the marginal mass function $p_\beta(\theta)$ as

$$p_\beta(\theta) = \sum_{\mathcal{H}} \prod_{v\in V} \theta^{(v)}_{\alpha_v|\alpha_{\mathrm{pa}(v)}}. \tag{7}$$

Here, $\mathcal{H}$ denotes the set of all $\alpha \in \{0,1\}^V$ such that the restriction to the leaves of T is equal to β. The image of this map is, by definition, the general Markov model on T^r (cf. [14], Section 8.3, [10]).

Standard theory in graphical models tells us that the Markov process on T^r is equal to $\widetilde{\mathcal{M}}_T$ and, consequently, that the general Markov on T^r model is equal to $\mathcal{M}_T$. Indeed, since T^r is a perfect directed graph (see Section 2.1.3 in [7]), by [7], Theorem 3.28, the Markov properties are equivalent to the factorization with respect to the undirected version of T^r, which is just T. Since T is decomposable, by [7], Proposition 3.19, the factorization according to T is equivalent to the global Markov properties on T.

In this paper, we often focus on *trivalent trees*, that is, trees such that every inner node has degree three. This is an important subclass because, by the well-known lemma below (see, e.g., [10], Section 2), the nodes of valency two in a given tree add nothing to the model class $\mathcal{M}_T$.

Lemma 2.1. *Let T be a tree. Let $v \in V$ be a node of degree two and let $T' = T/v$ be the tree obtained from T by suppressing v. Then $P \in \mathcal{M}_T$ if and only if $P \in \mathcal{M}_{T'}$.*

Corollary 2.2. *Let T be a tree and let i, j, k be any three leaves of T. The marginal model on (X_i, X_j, X_k) induced from $\mathcal{M}_T$ and denoted by $\mathcal{M}_{T(ijk)}$ is equivalent to the tripod tree model where the tripod tree is given in Figure 1.*

In addition, the model corresponding to any tree is a submodel of a model corresponding to a trivalent tree. To show this, we need the following definition.

Definition 2.3. *Let T be any tree. A trivalent expansion of T, denoted by T^*, is any tree $T^* = (V^*, E^*)$ whose each inner node has degree at most three and there exists a set of inner nodes $E' \subseteq E^*$ such that $T = T^*/E'$.*

Lemma 2.4. *Let T be a tree and $T^* = (V^*, E^*)$ its trivalent expansion with $E' \subseteq E^*$ such that $T = T^*/E'$. Then $\mathcal{M}_T \subseteq \mathcal{M}_{T^*}$.*

Proof. Let p be a point in $\mathcal{M}_T$. Then $p = f_T(\theta)$ for some $\theta \in \Theta_T$. Identifying edges of T^* and T in the obvious way, we can write $E^* = E' \cup E$. Define $\theta^* \in \Theta_{T^*}$ as follows. For all $\alpha_u, \alpha_v \in \{0, 1\}$

$$\begin{aligned} \theta^{*(v)}_{\alpha_v|\alpha_u} &= \theta^{(v)}_{\alpha_v|\alpha_u} \qquad \text{for every } (u, v) \in E, \\ \theta^{*(v)}_{\alpha_v|\alpha_u} &= \delta_{\alpha_u\alpha_v} \qquad \text{for every } (u, v) \in E', \end{aligned} \tag{8}$$

where δ_{ij} denotes the Kronecker's delta. It is now simple to check that $f_{T^*}(\theta^*) = p$. It follows that $p \in \mathcal{M}_{T^*}$. $\square$

For these reasons, we can usually safely restrict our attention to trivalent trees.

2.4. Moments and conditional independence

Let $X = (X_1, \dots, X_n)$ be a random vector and for each $\beta = (\beta_1, \dots, \beta_n) \in \mathbb{N}^n$ denote $X^\beta = \prod_i X_i^{\beta_i}$. We shall denote $\mathbb{E}X^\beta$ by λ_β and $\mathbb{E}U^\beta$ by μ_β, where $U_i = X_i - \mathbb{E}X_i$. When $\beta \in \{0, 1\}^n$, it is often convenient to use an alternate notation. Thus, for subsets $I \subseteq [n] := \{1, 2, \dots, n\}$, we let $\lambda_I = \mathbb{E}(\prod_{i\in I} X_i)$, $\mu_I = \mathbb{E}(\prod_{i\in I} U_i)$. Note that λ_{e_i}, where e_i is the standard basis vector in $\mathbb{R}^n$, can also be denoted by λ_i for $i = 1, \dots, n$.

The model $\mathcal{M}_T$ in the previous section is given in terms of the probabilities as the image of the map in (7). We find it convenient to change these coordinates. Let $[n]_{\geq 2}$ denote all subsets of $[n]$ with at least two elements. Denote by $\mathcal{C}_n$ the set of values of all the means $\lambda_1, \dots, \lambda_n$ together with central moments μ_I such that $I \in [n]_{\geq 2}$ for all possible probabilities in Δ_{2^n-1}. There exists a polynomial isomorphism, $f_{p\mu} : \Delta_{2^n-1} \to \mathcal{C}_n$, with the inverse denoted by $f_{\mu p}$ (for details see Appendix A.1). Consequently, we can express any distribution in the general Markov model in terms of its central moments and means.

For any two sets A, B let AB denote $A \cup B$. If $X_A \perp\!\!\!\perp X_B$, then $\mu_{IJ} = \mu_I\mu_J$ for all non-empty $I \subseteq A$, $J \subseteq B$. However, when all variables are binary, we also have a converse result. Thus, if for all non-empty $I \subseteq A$, $J \subseteq B$ we have that $\mu_{IJ} = \mu_I\mu_J$, then $X_A \perp\!\!\!\perp X_B$. Indeed, the independence expressed in terms of moments (see, e.g., Feller [5], page 136) gives

$$X_A \perp\!\!\!\perp X_B \iff \operatorname{Cov}(f(X_A), g(X_B)) = 0 \qquad \text{for all } f \in L^2(\mathcal{X}_A), g \in L^2(\mathcal{X}_B). \tag{9}$$

Since our variables are binary, all the functions of X_A and X_B are just polynomials with square-free monomials. Equivalently, every function of X_A or X_B can be written as a polynomial with square-free monomials in U_A or U_B, respectively. For instance, because $X_1, X_2 \in \{0, 1\}$,

$$X_1^{10} X_2^3 = X_1 X_2 = (U_1 + \lambda_1)(U_2 + \lambda_2) = U_1 U_2 + \lambda_2 U_1 + \lambda_1 U_2 + \lambda_1 \lambda_2.$$

Since the covariance is a bilinear form, Settimi and Smith [16] concluded that the independence can be checked only on these monomials and (9) can be rewritten as

$$X_A \perp\!\!\!\perp X_B \iff \operatorname{Cov}(U_A^\alpha, U_B^\beta) = 0 \qquad \text{for all } \alpha \in \{0, 1\}^{|A|}, \beta \in \{0, 1\}^{\beta}. \tag{10}$$

However, $\operatorname{Cov}(U_A^\alpha, U_B^\beta) = 0$ holds for each non-zero $\alpha \in \{0, 1\}^{|A|}$ and $\beta \in \{0, 1\}^{|B|}$ if and only if $\mu_{IJ} = \mu_I \mu_J$ for each $I \subseteq A$, $J \subseteq B$.

We can generalize the result above. For a random variable H_a let $\lambda_a = \mathbb{E} H_a$ and $U_a = H_a - \lambda_a$. For each $I \subseteq [n]$ let $U_I = \prod_{i \in I} U_i$ and

$$\eta_{a,I} = \mathbb{E}(U_I U_a) / \operatorname{Var}(H_a). \tag{11}$$

Note that under this notation $\operatorname{Var}(H_a) = \lambda_a (1 - \lambda_a)$.

Proposition 2.5. *Let H_a be a non-degenerate random variable. With the notation above, we have $X_A \perp\!\!\!\perp X_B | H_a$ if and only if for all non-empty $I \subseteq A$, $J \subseteq B$*

$$\begin{aligned} \mu_{IJ} &= \mu_I \mu_J + \lambda_a (1 - \lambda_a) \eta_{a,I} \eta_{a,J}, \\ \eta_{a,IJ} &= \mu_I \eta_{a,J} + \eta_{a,I} \mu_J + (1 - 2\lambda_a) \eta_{a,I} \eta_{a,J}. \end{aligned} \tag{12}$$

Proof. The definition of independence given in (10) induces a condition for $X_A \perp\!\!\!\perp X_B | H_a$. Thus, for each $I \subseteq A$, $J \subseteq B$ we have

$$\operatorname{Cov}(U_I, U_J | H_a = 0) = \operatorname{Cov}(U_I, U_J | H_a = 1) = 0, \tag{13}$$

so, in particular,

$$\begin{aligned} \lambda_a \operatorname{Cov}(U_I, U_J | H_a = 1) + (1 - \lambda_a) \operatorname{Cov}(U_I, U_J | H_a = 0) &= 0, \\ \operatorname{Cov}(U_I, U_J | H_a = 0) - \operatorname{Cov}(U_I, U_J | H_a = 1) &= 0. \end{aligned} \tag{14}$$

Moreover, for any $I \subseteq [n]$, one has $\mathbb{E}(U_I | H_a) = \mu_I + \eta_{a,I} U_a$, and hence

$$\operatorname{Cov}(U_I, U_J | H_a) = \mu_{IJ} - \mu_I \mu_J + (\eta_{a,IJ} - \eta_{a,I} \mu_J - \mu_I \eta_{a,J}) U_a - \eta_{a,I} \eta_{a,J} U_a^2. \tag{15}$$

Equation (12) now follows from substituting (15) into (14). $\square$

3. Tree posets and tree cumulants

In this section, we use the theory of partially ordered sets to propose a further change of coordinates. In the new coordinate system it is possible to parametrize the marginal model $\mathcal{M}_T$ in a product form (see Proposition 4.1) in contrast to the complicated polynomial mapping given in (7).

3.1. The poset of edge partitions

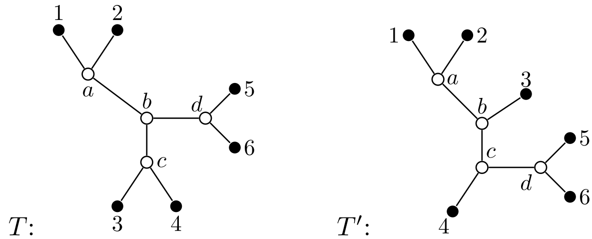
Let $T = (V, E)$ be a tree with n leaves. We identify the set of leaves of T with the set $[n]$. For any $e \in E$ we let $T \setminus e$ denote the *forest* obtained from T by removing e, that is, the subgraph of T given as a collection of disjoint trees with the set of nodes given by V and the set of edges given by $E \setminus e$. Similarly, for any $E' \subseteq E$, we let $T \setminus E'$ denote the forest obtained by removing all the edges in E'. An *edge split* is a partition of the set of leaves, $[n]$, of T into two non-empty sets induced by removing an edge e from E and restricting $[n]$ to the connected components of $T \setminus e$. By an *edge partition*, we mean any partition $B_1|B_2|\cdots|B_k$ of the set of leaves induced by considering connected components of $T \setminus E'$ for some $E' \subseteq E$. Call each subset B_i in this partition a *block*.

Henceforth let Π_T denote the poset of all edge partitions of the set of leaves induced by edges of T. The ordering is induced from the ordering of the poset of all partitions of the set of leaves (see [20], Example 3.1.1.d). Thus, for two partitions, $\pi = B_1|\cdots|B_k$ and $\nu = C_1|\cdots|C_l$, we write $\pi \leq \nu$ if every block of π is contained in one of the blocks of ν. To make this more explicit, define the following equivalence relation on the subsets of E. For $E_1, E_2 \subseteq E$ we say $E_1 \sim E_2$ if and only if removing E_1 induces the same partition of the set of leaves $[n]$ as removing E_2. For example, in Figure 1 the partition, $1|2|3$, can be obtained either by removing any two edges or by removing all them. However, the only way to obtain the partition, $12|3$, is by removing the edge incident with the third leaf.

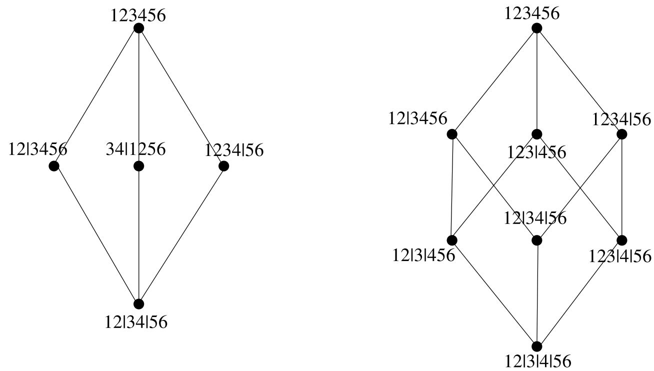
Let $\overline{E}_\pi$ denote the element of the equivalence class of subsets of E inducing the partition π, which is maximal with respect to inclusion. Suppose that $\pi \in \Pi_T$ is obtained by removing edges in the subset of the set of edges E_π and $\nu \in \Pi_T$ is obtained by removing edges in E_ν. Write $\pi \leq \nu$ if and only if $\overline{E}_\pi \supseteq \overline{E}_\nu$ and call π a *subpartition* of ν.

An *interval*, $[\pi, \nu]$, for π and ν in Π_T, is the set of all elements δ such that $\pi \leq \delta \leq \nu$. The poset Π_T forms a lattice (cf. [20], Section 3.3). To show this, we define $\pi \vee \nu \in \Pi_T$ ($\pi \wedge \nu \in \Pi_T$) as an element in Π_T obtained by removing $\overline{E}_\pi \cap \overline{E}_\nu$ ($\overline{E}_\pi \cup \overline{E}_\nu$). We have $\pi \vee \nu \geq \pi$, $\pi \vee \nu \geq \nu$ ($\pi \wedge \nu \leq \pi$, $\pi \wedge \nu \leq \nu$) and, if there exists another $\delta \in \Pi_T$ such that $\delta \geq \pi$, $\delta \geq \nu$ ($\delta \leq \pi$, $\delta \leq \nu$), then $\delta \geq \pi \vee \nu$ ($\delta \leq \pi \wedge \nu$). The element $\pi \vee \nu$ ($\pi \wedge \nu$) is called the *join* (the *meet*) of π and ν. The poset Π_T has a unique minimal element, $1|2|\cdots|n$, induced by removing all edges in E and the maximal one with no edges removed, which is equal to a single block, $[n]$. The maximal and minimal element of a lattice will be denoted by $\hat{1}$ and $\hat{0}$, respectively.

The number of elements in these posets is typically large. However, the key concepts can be presented using a simpler poset. Let $\widetilde{\Pi}_T$ denote a subposet of Π_T containing partitions obtained by removing only inner edges and consider, for example, the two different trivalent trees T and T', both with six leaves, given below



Their associated posets, $\widetilde{\Pi}_T$ and $\widetilde{\Pi}_{T'}$, are, respectively,



So, for example, 12|34|56 is an edge partition in $\widetilde{\Pi}_T$ and is a subpartition of any other edge partition $\nu \in \widetilde{\Pi}_T$. It can be obtained by removing either any two inner edges from (a, b), (b, c) and (b, d), or all of them. Since, for $\pi = 12|34|56$, there are no subpartitions of π, it follows that π is the minimal element of $\widetilde{\Pi}_T$. In $\widetilde{\Pi}_{T'}$, there is only one way to obtain this partition. Namely, by removing (a, b) and (c, d). However, note that this partition is not minimal in $\widetilde{\Pi}_{T'}$ because $12|3|4|56 < \pi$.

For any poset Π a *Möbius function* $\mathfrak{m}_\Pi : \Pi \times \Pi \to \mathbb{R}$ is defined by $\mathfrak{m}_\Pi(\pi, \pi) = 1$ for every $\pi \in \Pi$, $\mathfrak{m}_\Pi(\pi, \nu) = -\sum_{\pi \le \delta < \nu} \mathfrak{m}_\Pi(\pi, \delta)$ for $\pi < \nu$ in Π and is zero otherwise (cf. [20], Section 3.7). Recall that for any $W \subset V$, $T(W)$ denotes the subtree of T spanned on W (see Section 2.1). We denote $\mathfrak{m}_{\Pi_{T(W)}} := \mathfrak{m}_W$ and $\mathfrak{m}_{\Pi_T} := \mathfrak{m}$, and let $\hat{0}_W$ and $\hat{1}_W$ denote the minimal and the maximal element of $\Pi_{T(W)}$, respectively. For any partition $\pi \in \Pi_T$ the interval $[\hat{0}, \pi]$ has a natural structure of a product of posets for blocks of π, namely $\prod_{B \in \pi} \Pi_{T(B)}$, where the product is over all blocks B of π. By Proposition 3.8.2 in [20], the Möbius function on the product of posets $\prod_{B \in \pi} \Pi_{T(B)}$ can be written as a product of Möbius functions for each of the posets $\Pi_{T(B)}$. Thus, for $\nu \le \pi$ in Π_T

$$\mathfrak{m}(\nu, \pi) = \prod_{B \in \pi} \mathfrak{m}_B(\nu_B, \hat{1}_B), \tag{16}$$

where $\nu_B \in \Pi_{T(B)}$ is the restriction of $\nu \in \Pi_T$ to the block containing only elements from $B \subset [n]$ (it is well defined since $\nu \le \pi$) and $\pi_B = \hat{1}_B$ for each B.

In the next section, we will use the Möbius function of the poset of tree partitions to derive a useful change of coordinates on $\mathcal{M}_T$.

3.2. An induced change of coordinates

Assume that each inner node of T has degree at most three and consider a map, $f_{\mu\kappa}:\mathbb{R}^n \times \mathbb{R}^{2^n} \to \mathbb{R}^n \times \mathbb{R}^{2^n}$, where the coordinates in the domain are denoted by $\lambda_1, \ldots, \lambda_n$ and μ_I for $I \subseteq [n]$ and the coordinates in the image are denoted by $\lambda_1, \ldots, \lambda_n$ and κ_I for $I \subseteq [n]$. The map is defined as the identity on the first n coordinates corresponding to the means and

$$\kappa_I = \sum_{\pi \in \Pi_{T(I)}} \mathfrak{m}_I(\pi, \hat{1}_I) \prod_{B \in \pi} \mu_B \qquad \text{for all } I \subseteq [n]. \tag{17}$$

It is easy to prove that the Jacobian of $f_{\mu\kappa}$ is equal to 1, so, in particular, this is constant. To see this, order the variables in such a way that the first n coordinates both in $\mathcal{K}_T$ and $\mathcal{C}_n$ are $\lambda_1, \ldots, \lambda_n$ and let κ_I precede κ_J (μ_I precede μ_J) as long as $I \subset J$. The Jacobian matrix of $f_{\mu\kappa}$ is then lower triangular with each of its diagonal entries equal to 1. It follows that the modulus of its determinant is always 1.

The map, $f_{\mu\kappa}$, is a regular polynomial map with a regular polynomial inverse $f_{\kappa\mu}$. Therefore, it gives a change of coordinates from the central moments with means to a coordinate system given by $\lambda_1, \ldots, \lambda_n$ and κ_I for $I \subseteq [n]$. Its inverse map is given by

$$\mu_I = \sum_{\pi \in \Pi_{T(I)}} \prod_{B \in \pi} \kappa_B \qquad \text{for all } I \in [n]_{\geq 2}. \tag{18}$$

To show (18), define two functions on $\Pi_{T(I)}$: $\alpha(\pi) = \prod_{B\in\pi} \mu_B$ and $\beta(\pi) = \prod_{B\in\pi} \kappa_B$. For each $\pi \in \Pi_{T(I)}$, by (17),

$$\begin{aligned}\beta(\pi) &= \prod_{B\in\pi} \kappa_B = \prod_{B\in\pi} \Bigg(\sum_{\nu_B \in \Pi_{T(B)}} \mathfrak{m}_B(\nu_B, \hat{1}_B) \prod_{C \in \nu_B} \mu_C \Bigg) \\ &= \sum_{\nu \leq \pi} \prod_{B \in \pi} \mathfrak{m}_B(\nu_B, \hat{1}_B) \prod_{C \in \nu} \mu_C,\end{aligned}$$

where ν is an element of $\Pi_{T(I)}$ such that its restriction to each of the blocks $B \in \pi$ is equal to ν_B. By the product formula in (16), we have $\prod_{B\in\pi} \mathfrak{m}_B(\nu_B, \hat{1}_B) = \mathfrak{m}_I(\nu, \pi)$. Therefore, $\beta(\pi) = \sum_{\nu\leq\pi} \mathfrak{m}_I(\nu, \pi)\alpha(\nu)$ for all $\pi \in \Pi_{T(I)}$. Equation (18) now follows on applying the Möbius inversion formula in Proposition 3.7.1 in [20].

Denote $\mathcal{K}_T = f_{\mu\kappa}(\mathcal{C}_n)$. Since $\mathcal{K}_T$ is contained in a subset of $\mathbb{R}^n \times \mathbb{R}^{2^n}$ given by $\kappa_\varnothing = \kappa_1 = \cdots = \kappa_n = 0$, a system of coordinates on $\mathcal{K}_T$ is given by λ_i for $i = 1, \ldots, n$ and κ_I for $I \in [n]_{\geq 2}$. This system of coordinates is called *tree cumulants*. The name is justified by (17) because one of the definitions of classical cumulants is the following. Let $\Pi(I)$ denote the set of all partitions of

$I = \{i_1, \ldots, i_k\} \in [n]_{\geq 2}$ (see [20], Example 3.1.1.d). Then, for all $k > 1$

$$\mathrm{Cum}(X_{i_1}, \ldots, X_{i_k}) = \sum_{\pi \in \Pi(I)} \mathfrak{m}_{\Pi(I)}(\pi, \hat{1}_I) \prod_{B \in \pi} \mu_B, \tag{19}$$

where the product is over all blocks of π. Moreover, for every $\pi \in \Pi(I)$

$$\mathfrak{m}_{\Pi(I)}(\pi, \hat{1}_I) = (-1)^{|\pi|-1}(|\pi| - 1)!,$$

where $|\pi|$ denotes the number of blocks in π. Note that the usual definition of cumulants uses non-central moments instead of central moments in (19). It can be shown that both definitions are equivalent for all cumulants of order greater than one because the classical cumulants are translation invariant. The definition in (19) is thus essentially the same as (17) but with a different defining poset (cf. [12,17]).

Using a basic result in the theory of lattices, Lemma 3.2 shows that certain features of classical cumulants are also shared by tree cumulants (cf. Section 2.1 of [8]).

Lemma 3.1 (Corollary in [11], Section 5). *Let L be a finite lattice and let $\pi_0 \neq \hat{1}$ in L. Then, for any ν in L*

$$\sum_{\pi \wedge \pi_0 = \nu} \mathfrak{m}(\pi, \hat{1}) = 0.$$

Lemma 3.2. *Let T be a tree with n leaves. Whenever there exists an edge split $C_1|C_2 \in \Pi_T$ of the set of leaves $[n]$ such that $X_{C_1} \perp\!\!\!\perp X_{C_2}$, then $\kappa_{1\cdots n} = 0$.*

Proof. Let π_0 be the split $C_1|C_2$ such that $X_{C_1} \perp\!\!\!\perp X_{C_2}$. It follows that $\mu_{1\cdots n}$ is equal to $\mu_{C_1}\mu_{C_2}$. More generally, for any $I \in [n]_{\geq 2}$,

$$\mu_I = \mu_{C_1 \cap I}\mu_{C_2 \cap I}.$$

Consequently, for any partition $\pi \in \Pi_T$

$$\prod_{B \in \pi} \mu_B = \prod_{B \in \pi \wedge \pi_0} \mu_B. \tag{20}$$

Using (17) and (20), we obtain

$$\kappa_{1\cdots n} = \sum_{\pi \in \Pi_T} \mathfrak{m}(\pi, \hat{1}) \prod_{B \in \pi} \mu_B = \sum_{\pi \in \Pi_T} \mathfrak{m}(\pi, \hat{1}) \prod_{B \in \pi \wedge \pi_0} \mu_B.$$

Since $\pi \wedge \pi_0 \leq \pi_0$, by grouping all partitions $\pi \in \Pi_T$ giving the same partition, after taking the meet with π_0, we can rewrite the sum as

$$\kappa_{1\cdots n} = \sum_{\pi \in \Pi_T} \mathfrak{m}(\pi, \hat{1}) \prod_{B \in \pi \wedge \pi_0} \mu_B = \sum_{\nu \leq \pi_0} \left(\sum_{\pi \wedge \pi_0 = \nu} \mathfrak{m}(\pi, \hat{1}) \right) \prod_{B \in \pi \wedge \pi_0} \mu_B.$$

However, this is zero since by Lemma 3.1 each of $\sum_{\pi \wedge \pi_0 = \nu} \mathfrak{m}(\pi, \hat{1})$ is zero. □

4. The induced parametrization

We now define a new parameter space, Ω_T, with $|V|+|E|$ parameters denoted by $\eta_{u,v}$ for all $(u,v)\in E$ and $\bar{\mu}_v$ for all $v\in V$. The map between the two parameter spaces is given by

$$\begin{aligned} \eta_{u,v} &= \theta^{(v)}_{1|1}-\theta^{(v)}_{1|0} \qquad \text{for all } (u,v)\in E \quad \text{and} \\ \bar{\mu}_v &= 1-2\lambda_v \qquad \text{for each } v\in V, \end{aligned} \tag{21}$$

where λ_v is a polynomial in the original parameters in Θ_T. The details are given in Appendix A.2, where the inverse map is given by (36). It follows that the change of parameters between Θ_T and Ω_T is a polynomial isomorphism.

It can be checked that if $\mathrm{Var}(Y_u)>0$, then $\eta_{u,v}=\mathbb{E}(U_uU_v)/\mathrm{Var}(Y_u)$ is the regression coefficient of Y_v on Y_u. Therefore, $\eta_{u,v}$, defined above, coincides with the definition of $\eta_{u,v}$ in (11). If $\mathrm{Var}(Y_u)=0$, then the formula in (11) is not well defined; however, (21) always is.

Proposition 4.1 below motivates the whole section and demonstrates why our new coordinate system is particularly useful. Henceforth let $\mathcal{M}^{\kappa}_T=(f_{\mu\kappa}\circ f_{p\mu})(\mathcal{M}_T)\subseteq\mathcal{K}_T$.

Proposition 4.1. *Let $T=(V,E)$ be a rooted tree with n leaves such that each inner node has degree at most three. Then $\mathcal{M}^{\kappa}_T$ is given as the image of $\psi_T:\Omega_T\to\mathcal{K}_T$. Here ψ_T is defined by $\lambda_i=\frac{1}{2}(1-\bar{\mu}_i)$ for $i=1,\dots,n$ and*

$$\kappa_I=\frac{1}{4}\big(1-\bar{\mu}^2_{r(I)}\big)\prod_{v\in V(I)\setminus I}\bar{\mu}_v^{\deg(v)-2}\prod_{(u,v)\in E(I)}\eta_{u,v} \qquad \text{for each } I\in[n]_{\geq 2}, \tag{22}$$

where the degree is taken in $T(I)=(V(I),E(I))$ and $r(I)$ denotes the root of $T(I)$ (cf. Section 2.1).

The proof is given in Appendix B.

By Lemma 2.4 we can obtain the parametrization of $\mathcal{M}_T$ for any non-trivalent tree $T=(V,E)$ using a parametrization for its trivalent expansion $T^*=(V^*,E^*)$. Let E' be the subset of inner nodes of E^* given in Definition 2.3, so that $T^*/E'=T$. Let $\{V^*\}$ denote the equivalence classes of subsets of V^* such that $v\sim v'$ if and only if v becomes identified with v' in T in the process of contracting E' in T^*. There exists a natural identification of V with $\{V^*\}$. Let $\{v\}$ denote the equivalence class of $v\in V^*$ or the corresponding node in T. In particular, since E' is a set of inner edges, the class $\{i\}$ of every leaf $i\in[n]$ can be naturally identified with i and hence $\{V^*\setminus[n]\}=\{V^*\}\setminus[n]$.

Lemma 4.2. *Let T be any tree and T^* be its trivalent expansion. If κ^*_I for $I\in[n]_{\geq 2}$ are tree cumulants of T^*, then $\mathcal{M}^{\kappa}_T$ is given in $\mathcal{K}_{T^*}$ as the image of a map that is the identity on the coordinates corresponding to $\bar{\mu}_i$ for $i=1,\dots,n$ and, for each $I\in[n]_{\geq 2}$,*

$$\kappa^*_I=\frac{1}{4}\big(1-\bar{\mu}^2_{r(I)}\big)\prod_{v\in V(I)\setminus I}\bar{\mu}_v^{\deg(v)-2}\prod_{(u,v)\in E(I)}\eta_{u,v}, \tag{23}$$

where $T(I) = (V(I), E(I))$ is the subtree of T spanned on I.

Proof. By Lemma 2.4 and equation (8), $\mathcal{M}_T \subseteq \mathcal{M}_{T^*}$ is the image $f_{T^*}(\Theta_T)$, where Θ_T is the subset of Θ_{T^*} given by setting $\theta^{*(v)}_{\alpha_v|\alpha_u} = \delta_{\alpha_u\alpha_v}$ for every edge $(u, v) \in E'$ and $\theta^{*(v)}_{\alpha_v|\alpha_u} = \theta^{(v)}_{\alpha_v|\alpha_u}$ otherwise. In the new parameters, Ω_T is isomorphic to the subset of Ω_{T^*} given by

$$\begin{aligned} \eta^*_{u,v} &= \eta_{u,v} \qquad \text{for all } (u,v) \notin E', \\ \eta^*_{u,v} &= 1 \qquad \text{for all } (u,v) \in E' \quad \text{and} \\ \bar{\mu}^*_v &= \bar{\mu}_{\{v\}} \qquad \text{for all } v \in V^*. \end{aligned} \tag{24}$$

Denote the root of T^* by r^*. We show (23) for $I = [n]$. The general case can be proved with an obvious change in notation. By Proposition 4.1, the model $\mathcal{M}_{T^*}$ is parametrized by

$$\kappa^*_{1\cdots n} = \frac{1}{4}(1 - \bar{\mu}^{*2}_{r^*}) \prod_{v \in V^* \setminus [n]} \bar{\mu}^{*\deg(v)-2}_v \prod_{(u,v) \in E^*} \eta^*_{u,v}. \tag{25}$$

Since $E^* = E \cup E'$ by applying (24), $\prod_{(u,v)\in E^*} \eta^*_{u,v}$ becomes $\prod_{(u,v)\in E} \eta_{u,v}$, where we have identified E with $E^* \setminus E'$. For every $w \in V^*$, whenever $\deg\{w\} \geq 3$, we have that $\deg\{w\} = |\{w\}| + 2$. Therefore, if $\deg\{w\} \geq 3$, then the degree of each $v \in \{w\}$ in T^* equals 3. Hence

$$\sum_{v \in \{w\}} (\deg v - 2) = \sum_{v \in \{w\}} 1 = |\{w\}| = \deg\{w\} - 2.$$

It follows that, after applying (24), $\prod_{v\in\{w\}} \bar{\mu}^{*\deg v - 2}_v$ becomes $\bar{\mu}^{\deg\{w\}-2}_{\{w\}}$. The last statement is also true if $\deg\{w\} = 2$. For, in this case, $\deg w = 2$ in T^* and w is the only element in $\{w\}$. Moreover, E' is necessarily contained in the set of inner edges of T^*. It follows that $\prod_{v\in V^*\setminus[n]} \bar{\mu}^{*\deg(v)-2}_v$ in (25) becomes

$$\prod_{\{w\} \in \{V^*\} \setminus [n]} \bar{\mu}^{\deg(\{w\})-2}_{\{w\}} = \prod_{v \in V \setminus [n]} \bar{\mu}^{\deg(v)-2}_v.$$

In addition, $\{r^*\}$ becomes the root of T denoted by r. Therefore, (25) becomes

$$\kappa^*_{1\cdots n} = \frac{1}{4}(1 - \bar{\mu}^2_r) \prod_{v \in V \setminus [n]} \bar{\mu}^{\deg(v)-2}_v \prod_{(u,v) \in E} \eta_{u,v},$$

which is exactly (23) for $I = [n]$. □

Remark 4.3. For every $v \in V$ the variance $\mathrm{Var}(Y_v)$ is zero if and only if $\bar{\mu}^2_v = 1$. Hence, in the case when $\bar{\mu}^2_v < 1$, the variable Y_v is non-degenerate. In phylogenetics it is usually assumed that $\bar{\mu}^2_r < 1$ for the root r of T and $\eta_{u,v} \neq 0$ for all $(u, v) \in E$ (cf. Conditions (M1) and (M2) in Section 8.2, [14]). It is shown in Section 8.2 in [14] that (M1) and (M2) imply the weaker condition $\bar{\mu}^2_v < 1$ for all $v \in V$. Over the subset of Ω_T on which this weaker condition holds, we can apply another smooth transformation on both the parameter and model space. This leads to a further simplification of the parametrization in (22) presented in Appendix A.3.

5. Singularities and the geometry of unidentified subspaces

The identifiability of general Markov models can be addressed here geometrically. For any $q \in \mathcal{M}_T$ the preimage $\widehat{\Theta}_T := f_T^{-1}(q)$, that is, the set of parameter values that is consistent with the known probability model q, is called the *q-fiber*. In this section, we analyze the geometry of these fibers, determining when they are finite and thus when the model is locally identifiable. We will also be interested in when the fibers are smooth subsets of Θ_T and when they are singular. We use methods similar to the ones presented in a different context by Moulton and Steel in [9], Section 6. The results in this section generalize similar results for the naive Bayes models (cf. [6], Theorem 7).

First we analyze the geometric description of Ω_T. This gives a set of implicit inequalities constraining each q-fiber. Simple linear constraints defining Θ_T become only slightly more complicated when expressed in the new parameters. The choice of parameter values is not free anymore in the sense that the constraining equations for each of the parameters involve the values of other parameters. By (36), Ω_T is given by $\bar{\mu}_r \in [-1, 1]$ and for each $(u, v) \in E$

$$\begin{aligned} -(1+\bar{\mu}_v) \le (1-\bar{\mu}_u)\eta_{u,v} \le (1-\bar{\mu}_v), \\ -(1-\bar{\mu}_v) \le (1+\bar{\mu}_u)\eta_{u,v} \le (1+\bar{\mu}_v). \end{aligned} \tag{26}$$

For $\hat{p} \in \mathcal{M}_T$ let $\widehat{\Sigma} = [\hat{\mu}_{ij}] \in \mathbb{R}^{n\times n}$ be the covariance matrix of the observed variables labelled by the leaves of T computed with respect to $\hat{p}$. We show that the geometry of the $\hat{p}$-fiber, denoted by $\widehat{\Theta}_T$, is determined by zeros in $\widehat{\Sigma}$. Let $\hat{\lambda}_i$ be the expected value of X_i. Then, for every point in the $\hat{p}$-fiber, we have $\bar{\mu}_i = \hat{\mu}_i = 1 - 2\hat{\lambda}_i$ for all $i = 1, \dots, n$. Without loss we always assume that $\hat{\lambda}_i(1-\hat{\lambda}_i) \neq 0$ (or, equivalently, that $\hat{\mu}_i^2 \neq 1$) for all $i = 1, \dots, n$.

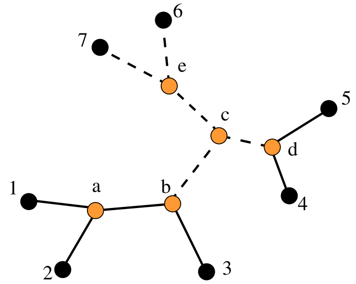
It is easier to analyze the geometry of $\hat{p}$-fibers in Ω_T. Therefore transform $\widehat{\Theta}$ to Ω_T using the mapping $f_{\theta\omega}$. The image of this map, denoted by $\widehat{\Omega}_T$, is isomorphic to $\widehat{\Theta}_T$. Let $\hat{\kappa}_{ij}$ denote the corresponding second-order tree cumulants in the point $f_{p\kappa}(\hat{p})$. Since $\kappa_{ij} = \mu_{ij}$ for all $i, j \in [n]$, from (22) for any $\omega_0 = ((\bar{\mu}_v^0), (\eta_{u,v}^0)) \in \widehat{\Omega}_T$ we have that

$$\hat{\mu}_{ij} = \mu_{ij}(\omega_0) = \frac{1}{4}\left(1 - \left(\bar{\mu}^0_{r(ij)}\right)^2\right) \prod_{(u,v)\in E(ij)} \eta^0_{u,v}. \tag{27}$$

We say that that an edge, $e \in E$, is *isolated relative to* $\hat{p}$ if $\hat{\mu}_{ij} = 0$ for all $i, j \in [n]$ such that $e \in E(ij)$. We denote the set of all edges of T that are isolated relative to $\hat{p}$ by $\widehat{E} \subseteq E$. We define the $\hat{p}$-forest $\widehat{T}$ as the forest obtained from T by removing edges in $\widehat{E}$ so that $\widehat{T} = T \setminus \widehat{E}$. Hence, the set of vertices of $\widehat{T}$ is equal to the set of vertices of T and the set of edges is equal to $E \setminus \widehat{E}$.

We illustrate this construction in the example below. Let T be the tree given in Figure 2 and assume that the covariance matrix contains zeros given in the provided 7×7 matrix, where the asterisks mean any non-zero values such that the matrix is positive semidefinite. It can be checked that $\widehat{E} = \{(b, c), (c, d), (c, e), (e, 6), (e, 7)\}$ and these edges are depicted as dashed lines. The forest, $\widehat{T}$, is obtained by removing the edges in $\widehat{E}$.

We now define relations on $\widehat{E}$ and $E \setminus \widehat{E}$. For two edges, e, e', with either $\{e, e'\} \subset \widehat{E}$ or $\{e, e'\} \subset E \setminus \widehat{E}$, write $e \sim e'$ if either $e = e'$ or e and e' are adjacent and all the edges that are



$$\widehat{\Sigma} = \begin{bmatrix} * & * & * & 0 & 0 & 0 & 0 \\ & * & * & 0 & 0 & 0 & 0 \\ & & * & 0 & 0 & 0 & 0 \\ & & & * & * & 0 & 0 \\ & & & & * & 0 & 0 \\ & & & & & * & 0 \\ & & & & & & * \end{bmatrix}$$

Figure 2. An example of a tree and a sample covariance matrix. The dashed lines depict the edges isolated with respect to $\hat{p}$.

incident with both e and e' are isolated relative to $\hat{p}$. We now construct the transitive closure of $\sim$ restricted to pairs of edges in $\widehat{E}$ to form an equivalence relation on $\widehat{E}$. Consider a graph with nodes representing elements of $\widehat{E}$ and put an edge between e, e' whenever $e \sim e'$. Then the equivalence classes correspond to connected components of this graph. In the same way, we take the transitive closure of $\sim$ restricted to the pairs of edges in $E \setminus \widehat{E}$ to form an equivalence relation in $E \setminus \widehat{E}$. We will let $[\widehat{E}]$ and $[E \setminus \widehat{E}]$ denote the set of equivalence classes of $\widehat{E}$ and $E \setminus \widehat{E}$, respectively. For the tree from the example above, $[\widehat{E}]$ is one element given by a subtree of T spanned on $\{b, d, 6, 7\}$ and

$$[E \setminus \widehat{E}] = \{\{(1, a)\}, \{(2, a)\}, \{(a, b), (b, 3)\}, \{(d, 4), (d, 5)\}\}.$$

By construction, all the inner nodes of T have either degree zero in $\widehat{T}$ or the degree is strictly greater than one. The following lemma shows that whenever the degree of an inner node in $\widehat{T}$ is not zero, the node represents a non-degenerate random variable.

Lemma 5.1. *Let $\hat{p} \in \mathcal{M}_T$. If $v \in V$ is an inner node of T such that* $\deg(v) \geq 2$ *in the $\hat{p}$-forest $\widehat{T}$, then the variable H_v cannot be degenerate.*

Proof. By construction, if $\deg(v) \geq 2$ in $\widehat{T}$, then there exists $i, j \in [n]$ such that $\hat{\mu}_{ij} \neq 0$ and v lies on the path between i and j. Suppose that H_v is degenerate. Then the global Markov properties in (4) imply that $X_i \perp\!\!\!\perp X_j$. But then $\hat{\mu}_{ij} = 0$ and we obtain the contradiction. □

We now list some basic statements, partly based on Lemma 6.4 in [9], which follow directly definitions above.

Remark 5.2. Let $T = (V, E)$ be a tree with n leaves, let $\mathcal{M}_T$ be the corresponding general Markov model and suppose that $\hat{p} \in \mathcal{M}_T$.

(i) The edges in any equivalence class of $[\widehat{E}]$ form a connected subgraph of T. If T is trivalent, then this subgraph is either a single edge or a trivalent tree.

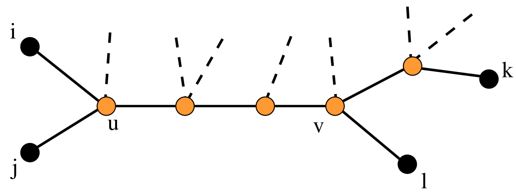
(ii) If each inner node of T has degree at least two in $\widehat{T}$, then all the equivalence classes in $[\widehat{E}]$ are just single edges. If each inner node has degree at least three in $\widehat{T}$, then all equivalence classes in $[E \setminus \widehat{E}]$ are single edges.

(iii) The edges in any equivalence class in $[E \setminus \widehat{E}]$ can be ordered so that they form a path in T.

(iv) Every connected component of $\widehat{T}$ is either a single node or a tree with its set of leaves contained in $[n]$.

Lemma 5.3. *Let $E(uv) \subset E$ be any path as in Remark 5.2*(iii), *which is an element of* $[E \setminus \widehat{E}]$. *Then the quantities μ_{uv}^2 and $\eta_{u,v}^2$ are constant on $\widehat{\Omega}_T$ and non-zero. It is possible to determine their values from $\hat{p}$.*

Proof. First note that the degree of each inner node on the path between u and v in $\widehat{T}$ must be exactly two. Moreover, the degree of both u and v in $\widehat{T}$ must be at least three unless u or v is a leaf. Consider the case when both u and v are inner nodes of T. In this case, these nodes have degrees at least three in $\widehat{T}$ and we can find four leaves i, j, k, l such that u separates i from j in $\widehat{T}$, v separates k and l and $\{u, v\}$ separates $\{i, j\}$ from $\{k, l\}$ as in the graph below.



Furthermore, by construction, $\hat{\mu}_{ij}, \hat{\mu}_{kl}, \hat{\mu}_{ik}, \hat{\mu}_{jl}$ are all non-zero. Consider the marginal models for $T(ijk)$ and $T(ikl)$. By Corollary 2.2, these are equivalent to models associated with tripod trees as in Figure 1. Hence, from (3) we have that

$$\bar{\mu}_u^2 = \frac{\hat{\mu}_{ijk}^2}{\hat{\mu}_{ijk}^2 + 4\hat{\mu}_{ij}\hat{\mu}_{ik}\hat{\mu}_{jk}}, \qquad \bar{\mu}_v^2 = \frac{\hat{\mu}_{ikl}^2}{\hat{\mu}_{ikl}^2 + 4\hat{\mu}_{ik}\hat{\mu}_{il}\hat{\mu}_{kl}}. \tag{28}$$

These equations are well defined since $\hat{\mu}_{ij}\hat{\mu}_{ik}\hat{\mu}_{jk} > 0$ and $\hat{\mu}_{ik}\hat{\mu}_{il}\hat{\mu}_{kl} > 0$. Consider the quantity $\frac{\hat{\mu}_{ik}\hat{\mu}_{jl}}{\hat{\mu}_{ij}\hat{\mu}_{kl}}$ and substitute (27) for each of the terms. A simple rearrangement now gives that

$$\frac{\hat{\mu}_{ik}\hat{\mu}_{jl}}{\hat{\mu}_{ij}\hat{\mu}_{kl}} = \frac{1-\bar{\mu}_u^2}{1-\bar{\mu}_v^2}\eta_{u,v}^2(\omega),$$

where $\eta_{u,v}(\omega) = \frac{1-\bar{\mu}_{r(uv)}^2}{1-\bar{\mu}_u^2}\prod_{(w,w')\in E(uv)}\eta_{w,w'}$. Therefore, substituting for $\bar{\mu}_u^2$, $\bar{\mu}_v^2$ using (28) implies that $\eta_{u,v}^2$ is constant on $\widehat{\Omega}_T$ and non-zero. Its value can be determined as a function of $\hat{p}$. Also the value of μ_{uv}^2 is constant since $\mu_{uv}^2 = \frac{1}{16}(1-\bar{\mu}_u^2)^2\eta_{u,v}^2$.

If either u or v is a leaf of T, then the argument is very similar. Thus, if u is a leaf, then consider any two leaves i, j of T such that v separates u, i, j in $\widehat{T}$. In particular, as in (28),

$$\bar{\mu}_v^2 = \frac{\hat{\mu}_{uij}^2}{\hat{\mu}_{uij}^2 + 4\hat{\mu}_{ui}\hat{\mu}_{uj}\hat{\mu}_{ij}}.$$

Moreover, $\eta_{u,v}(\omega)$ must be determined, since from (27)

$$\frac{\hat{\mu}_{ui}\hat{\mu}_{uj}}{\hat{\mu}_{ij}} = \frac{1}{4}(1-\bar{\mu}_v^2)\eta_{u,v}^2(\omega),$$

from which it follows that $\eta_{u,v}^2$ has to be constant on the $\hat{p}$-fiber. $\square$

The following theorem shows that the geometry of the $\hat{p}$-fiber $\widehat{\Omega}_T$ is determined by the zeros of the covariance matrix $\widehat{\Sigma}$.

Theorem 5.4 (The geometry of the $\hat{p}$-fiber – the smooth case). *Let $\hat{p} \in \mathcal{M}_T$. If each of the inner nodes of T has degree at least three in the $\hat{p}$-forest $\widehat{T}$, then the $\hat{p}$-fiber is a finite set of points of cardinality $2^{|V|-n}$. If each of the inner nodes of T has degree at least two in $\widehat{T}$, then the $\hat{p}$-fiber is diffeomorphic to a disjoint union of polyhedra. In particular, it is a manifold with corners. Its dimension is $2l_2$, where l_2 is the number of degree-2 nodes in $\widehat{T}$.*

The proof is given in Appendix C.

If T is trivalent, then the $\hat{p}$-fiber is finite if and only if for all $i, j \in [n]$ $\mu_{ij} \neq 0$. The proof of Theorem 5.4 provides explicit formulae for the parameters in this case when the $\hat{p}$-fiber is a finite number of points.

Corollary 5.5. *Let T be a tree such that each inner node has degree at least three and let $\hat{p} \in \mathcal{M}_T$. Consider the $\hat{p}$-forest $\widehat{T}$. If every inner node of T has degree at least three in $\widehat{T}$, then, by Remark 5.2*(ii), *both $[\widehat{E}]$ and $[E \setminus \widehat{E}]$ consist of singletons. In this case, every point in the $\hat{p}$-fiber satisfies*

$$\begin{aligned} \bar{\mu}_i &= \hat{\mu}_i \qquad \text{for all } i = 1, \dots, n, \\ \eta_{u,v} &= 0 \qquad \text{for all } (u,v) \in [\widehat{E}]. \end{aligned} \tag{29}$$

Moreover, for any inner node v of T, if $i, j, k \in [n]$ are any three leaves separated by v in T such that $\hat{\mu}_{ij}\hat{\mu}_{ik}\hat{\mu}_{jk} \neq 0$, then

$$\bar{\mu}_v^2 = \frac{\hat{\mu}_{ijk}^2}{\hat{\mu}_{ijk}^2 + 4\hat{\mu}_{ij}\hat{\mu}_{ik}\hat{\mu}_{jk}}$$

for any terminal edge $(v,i) \in E \setminus \widehat{E}$, where v is an inner node and $i \in [n]$ is a leaf of T. Let j, k be any two leaves such that v separates i, j, k and $\hat{\mu}_{jk} \neq 0$. Then

$$\eta_{v,i}^2 = \frac{\hat{\mu}_{ijk}^2 + 4\hat{\mu}_{ij}\hat{\mu}_{ik}\hat{\mu}_{jk}}{\hat{\mu}_{jk}^2}.$$

Moreover, for any inner edge $(u,v) \in E \setminus \widehat{E}$ let $i, j, k, l \in [n]$ be any four leaves of T such that u separates i and j in $\widehat{T}$, v separates j and k in $\widehat{T}$ and (u,v) separates $\{i, j\}$ from $\{k, l\}$ in $\widehat{T}$.

Then

$$\eta_{u,v}^2 = \frac{\hat{\mu}_{il}^2}{\hat{\mu}_{ij}^2} \frac{\hat{\mu}_{ijk}^2 + 4\hat{\mu}_{ij}\hat{\mu}_{ik}\hat{\mu}_{jk}}{\hat{\mu}_{ikl}^2 + 4\hat{\mu}_{ik}\hat{\mu}_{il}\hat{\mu}_{kl}}.$$

Remark 5.6. The choice of signs of the $\bar{\mu}_v$ and $\eta_{u,v}$ in Corollary 5.5 is not completely free and has to be consistent with signs of tree cumulants via (22) (see Appendix D).

The singular case when there is at least one degree-zero inner node is more complicated. We begin with an example.

Example 5.7. Let $T = (V, E)$ be the tripod tree rooted in the inner node as in Figure 1 and let $\hat{p} \in \mathcal{M}_T$. The degree of h in the $\hat{p}$-forest $\widehat{T}$ is less than two if and only if $\hat{\mu}_{ij} = 0$ for all $i \neq j = 1, 2, 3$. In this situation, $\widehat{E} = E$ and the $\hat{p}$-fiber $\widehat{\Omega}_T$ is given as a subset of Ω_T by equations for the sample means $\bar{\mu}_i = \hat{\mu}_i$ for $i = 1, 2, 3$ together with the three additional equations

$$(1 - \bar{\mu}_h^2)\eta_{h,1}\eta_{h,2} = 0, \qquad (1 - \bar{\mu}_h^2)\eta_{h,1}\eta_{h,3} = 0, \qquad (1 - \bar{\mu}_h^2)\eta_{h,2}\eta_{h,3} = 0.$$

Geometrically, in the subspace given by $\bar{\mu}_i = \hat{\mu}_i$ for $i = 1, 2, 3$, this is a union of two three-dimensional hyperplanes $\{\bar{\mu}_h = \pm 1\}$ and three planes given by $\{\eta_{h,1} = \eta_{h,2} = 0\}$, $\{\eta_{h,1} = \eta_{h,3} = 0\}$ and $\{\eta_{h,2} = \eta_{h,3} = 0\}$ subject to the additional inequality constraints defining Ω_T and given by (26). In particular, it is not a regular set since it has self-intersection points given by $1 - \bar{\mu}_h^2 = \eta_{h,1} = \eta_{h,2} = \eta_{h,3} = 0$.

This geometry is mirrored in the general case. We first need two definitions. We say that a node $v \in V$ is *non-degenerate (with respect to $\hat{p}$)* if either v is a leaf of T or $\deg v \geq 2$ in $\widehat{T}$. Otherwise, we say that the node is *degenerate with respect to $\hat{p}$*. The set of all nodes that are degenerate with respect to $\hat{p}$ is denoted by $\widehat{V}$. By Lemma 5.1, for all $v \in V \setminus \widehat{V}$, $\mathrm{Var}(Y_v) \neq 0$, where the variance is computed with respect to $\hat{p}$. Hence v is non-degenerate if and only if Y_v is a non-degenerate random variable.

We define the *deepest singularity* of $\widehat{\Omega}_T$ as

$$\widehat{\Omega}_{\text{deep}} := \{\omega \in \widehat{\Omega}_T \colon\ \eta_{u,v} = 0, \bar{\mu}_v^2 = 1 \text{ for all } (u, v) \in \widehat{E}, v \in \widehat{V}\}. \tag{30}$$

Theorem 5.8 (The geometry of the $\hat{p}$-fiber – the singular case). *If $\widehat{V}$ is non-empty, then the $\hat{p}$-fiber is a singular variety given as a union of intersecting smooth manifolds in $\mathbb{R}^{|V|+|E|}$ restricted to Ω_T. Their common intersection locus restricted to Ω_T is given by $\widehat{\Omega}_{\text{deep}}$, which lies on the boundary of Ω_T.*

The proof is given in Appendix C.

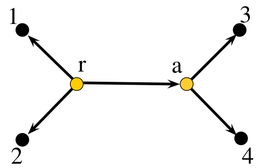


Figure 3. The quartet tree.

6. Example: The quartet tree model

In this section, we study the first non-trivial example: the quartet tree model given by the tree in Figure 3. The model is parametrized as in (7) by the root distribution and conditional probabilities attached to each of the edges. We set the values of the parameters to $\theta_1^{(r)} = 0.8$, $\theta_{1|0}^{(1)} = 0.8$, $\theta_{1|1}^{(1)} = 0.3$, $\theta_{1|0}^{(2)} = 0.7$, $\theta_{1|1}^{(2)} = 0.3$, $\theta_{1|0}^{(a)} = 0.8$, $\theta_{1|1}^{(a)} = 0.3$, $\theta_{1|0}^{(3)} = 0.7$, $\theta_{1|1}^{(3)} = 0.3$, $\theta_{1|0}^{(4)} = 0.7$, $\theta_{1|1}^{(4)} = 0.3$. Using (7) we can then calculate the corresponding probabilities over the observed nodes that are given in the third column in the table below. The change of coordinates $f_{p\lambda}$ presented in Appendix A.1 and $f_{\mu\kappa}$ in Section 3.2 gives the corresponding non-central moments and tree cumulants that are shown in Table 1. Formula (21) enables us to calculate the values for the new parameters as: $\eta_{r,1} = 0.5$, $\eta_{r,2} = 0.4$, $\eta_{r,a} = 0.5$, $\eta_{a,3} = 0.4$, $\eta_{a,4} = 0.4$ and $\bar{\mu}_1 = -0.4$, $\bar{\mu}_2 = -0.24$, $\bar{\mu}_3 = -0.16$, $\bar{\mu}_4 = -0.16$, $\bar{\mu}_r = -0.6$, $\bar{\mu}_a = -0.4$. It is easy to verify that (22)

Table 1. Moments and tree cumulants for a probability assignment in $\mathcal{M}_T$

α	I	p_α	λ_I	κ_I
0000	$\varnothing$	0.0444	1.0000	0
0001	4	0.0307	0.5800	0
0010	3	0.0307	0.5800	0
0011	34	0.0403	0.3700	0.0336
0100	2	0.0346	0.6200	0
0101	24	0.0323	0.3724	0.0128
0110	23	0.0323	0.3724	0.0128
0111	234	0.0547	0.2422	−0.0020
1000	1	0.0482	0.7000	0
1001	14	0.0491	0.4220	0.0160
1010	13	0.0491	0.4220	0.0160
1011	134	0.0875	0.2750	−0.0026
1100	12	0.0828	0.4660	0.0320
1101	124	0.0979	0.2853	−0.0038
1110	123	0.0979	0.2853	−0.0038
1111	1234	0.1875	0.1875	0.0006

holds in this example. For instance,

$$\kappa_{1234} = \frac{1}{4}(1 - \bar{\mu}_r^2)\bar{\mu}_r \bar{\mu}_a \eta_{r,1}\eta_{r,2}\eta_{r,a}\eta_{a,3}\eta_{a,4} = 0.0006,$$

which equates with the value in the table. In general, higher-order tree cumulants tend to be very small.

If we have only tree cumulants $K \in \mathcal{M}_T^{\kappa}$, we can still identify the parameters of the model up to the label switching on the inner nodes using Corollary 5.5. Recall that if $|I| \le 3$, then $\kappa_I = \mu_I$ so, for example,

$$\bar{\mu}_r^2 = \frac{\mu_{123}^2}{\mu_{123}^2 + 4\mu_{12}\mu_{13}\mu_{23}} = 0.36,$$

$$\eta_{r,1}^2 = \frac{\mu_{123}^2 + 4\mu_{12}\mu_{13}\mu_{23}}{\mu_{23}^2} = 0.25,$$

$$\eta_{r,a}^2 = \frac{\mu_{14}^2}{\mu_{12}^2} \frac{\mu_{123}^2 + 4\mu_{12}\mu_{13}\mu_{23}}{\mu_{134}^2 + 4\mu_{13}\mu_{14}\mu_{34}} = 0.25.$$

Note that the entries in Table 1 can be computed in several different ways. However, by Corollary 5.5 this does not matter. For instance, to compute $\bar{\mu}_r$ we picked $1, 2, 3$ as three leaves separated by r. If, instead of $1, 2, 3,$ we used $1, 2, 4,$ the answer would be the same since

$$\bar{\mu}_r^2 = \frac{\mu_{124}^2}{\mu_{124}^2 + 4\mu_{12}\mu_{14}\mu_{24}} = 0.36.$$

Finally, in Appendix D we show that in this case we have exactly four possible distinct choices for combinations of signs of these parameters. The first one is the original one with all $\eta_{u,v} > 0$, which we denote by ω:

$$\begin{array}{lllll} \eta_{r,1} = 0.5, & \eta_{r,2} = 0.4, & \eta_{r,a} = 0.5, & \eta_{a,3} = 0.4, & \eta_{a,4} = 0.4, \\ \bar{\mu}_r = -0.6, & \bar{\mu}_a = -0.4, & & & \end{array}$$

where we omit $\bar{\mu}_i$ for $i = 1, 2, 3, 4$ since these are constant for all points in $\widehat{\Omega}_T$. We obtain three remaining points by using local sign switching as defined in Appendix D, which are $(\eta_{r,1}, \eta_{r,2}, \eta_{r,a}, \eta_{a,3}, \eta_{a,4}, \bar{\mu}_r, \bar{\mu}_a) = (-0.5, -0.4, -0.5, 0.4, 0.4, 0.6, -0.4)$ or $(0.5, 0.4, -0.5, -0.4, -0.4, -0.6, 0.4)$ or $(-0.5, -0.4, 0.5, -0.4, -0.4, -0.6, -0.4)$.

7. Discussion

The reparametrization of Bayesian tree models with hidden variables given herein has illuminated the structure of these tree models and has enabled us to establish some identifiability results. However, the applicability of the new coordinate system reaches far beyond understanding identifiability. Some additional results will be presented in forthcoming papers where we generalize both results of [2] and [15], obtaining the full semi-algebraic description of this model class, and results of [13], on the asymptotic approximation of the marginal likelihood integrals.

The results given here can be extended in a straightforward way to the case when all hidden variables are binary but all leaf variables are arbitrary. It is less clear how the methods extend to tree models for arbitrary finite discrete random variables, or more generally, to other discrete graphical models. However, the extension to Gaussian models on trees appears to be straightforward.

The definition of tree cumulants in (17) can be generalized using other posets than Π_T. This opens many interesting possibilities to investigate more general coordinate systems for binary models. They all share certain useful properties of classical cumulants. In particular, Lemma 3.2 is true if the poset of tree partitions is replaced by any other lattice of partitions. We will report on this result in a forthcoming paper.

Appendix A: Change of coordinates

A.1. From probabilities to central moments

Let Δ_{2^n-1} be the set of all possible probability distributions of a binary vector $X=(X_1,\dots,X_n)$ as defined in (6). Let $\mathcal{C}_n$ be the set of all possible central moments μ_I for $I\in[n]_{\geq 2}$ and means $\lambda_1,\dots,\lambda_n$. In this section, we show that there exists a polynomial isomorphism between Δ_{2^n-1} and $\mathcal{C}_n$.

First, perform a change of coordinates from the raw probabilities $p=[p_\alpha]$ to the non-central moments $\lambda=[\lambda_\alpha]$ for $\alpha=(\alpha_1,\dots,\alpha_n)\in\{0,1\}^n$. This is a linear map $f_{p\lambda}:\mathbb{R}^{2^n}\to\mathbb{R}^{2^n}$, where $\lambda=f_{p\lambda}(p)$ is defined as follows:

$$\lambda_\alpha=\sum_{\alpha\leq\beta\leq\mathbf{1}} p_\beta \qquad \text{for any } \alpha\in\{0,1\}^n, \tag{31}$$

where $\mathbf{1}$ denotes the vector of ones and the sum is over all binary vectors β such that $\alpha\leq\beta\leq\mathbf{1}$ in the sense that $\alpha_i\leq\beta_i\leq 1$ for all $i=1,\dots,n$. In particular, $\lambda_{\mathbf{0}}=1$ for all probability distributions. Therefore, the image $\mathcal{L}_n=f_{p\lambda}(\Delta_{2^n-1})$ is contained in the hyperplane defined by $\lambda_{\mathbf{0}}=1$. The map, $f_{p\lambda}:\Delta_{2^n-1}\to\mathcal{L}_n$, is invertible and hence we can obtain coordinates on $\mathcal{L}_n$ given by λ_α for all $\alpha\in\{0,1\}^n$ such that $\alpha\neq\mathbf{0}$. The inverse of $f_{p\lambda}$ is the map, $f_{\lambda p}=f_{p\lambda}^{-1}:\mathcal{L}_n\to\Delta_{2^n-1}$, and is given by

$$p_\alpha=\sum_{\alpha\leq\beta\leq\mathbf{1}}(-1)^{|\beta-\alpha|}\lambda_\beta \qquad \text{for } \alpha=(\alpha_1,\dots,\alpha_n)\in\{0,1\}^n. \tag{32}$$

The linearity of the expectation implies that the central moments can be expressed in terms of non-central moments. In particular,

$$\mu_\alpha=\sum_{\mathbf{0}\leq\beta\leq\alpha}(-1)^{|\beta|}\lambda_{\alpha-\beta}\prod_{i=1}^{n}\lambda_{e_i}^{\beta_i} \qquad \text{for } \alpha\in\{0,1\}^n, \tag{33}$$

where $|\beta| = \sum_i \beta_i$. Using these equations, we can transform variables from the non-central moments $[\lambda_\alpha]$ to another set of variables given by all the means $\lambda_{e_1}, \ldots, \lambda_{e_n}$, where $e_1, \ldots, e_n$ are standard basis vectors in $\mathbb{R}^n$, and central moments $[\mu_\alpha]$ for $\alpha \in \{0, 1\}^n$. The polynomial mapping $f_{\lambda\mu} : \mathbb{R}^{2^n} \to \mathbb{R}^n \times \mathbb{R}^{2^n}$ is the identity on the first n variables corresponding to the means $\lambda_{e_1}, \ldots, \lambda_{e_n}$ and is defined by (33) on the remaining variables. The image of $f_{\lambda\mu}$ is contained in the subspace $\mathcal{H} \subset \mathbb{R}^n \times \mathbb{R}^{2^n}$ given by $\mu_{e_1} = \cdots = \mu_{e_n} = 0$. It is easy to show (see, e.g., equation (5), [3]) that the inverse of $f_{\lambda\mu} : \mathbb{R}^{2^n} \to \mathcal{H}$ is given as $f_{\mu\lambda} = f_{\lambda\mu}^{-1} : \mathcal{H} \to \mathbb{R}^{2^n}$ defined by

$$\lambda_\alpha = \sum_{\mathbf{0} \leq \beta \leq \alpha} \mu_{\alpha-\beta} \prod_{i=1}^{n} \lambda_{e_i}^{\beta_i} \qquad \text{for } \alpha \in \{0, 1\}^n. \tag{34}$$

Let $\mathcal{C}_n$ denote $f_{\lambda\mu}(\mathcal{L}_n)$. Then $\mathcal{C}_n$ is contained in $\mathcal{H}$ and $\mu_{\mathbf{0}} = 1$. We have, therefore, obtained coordinates of $\mathcal{C}_n$ given by $\lambda_{e_1}, \ldots, \lambda_{e_n}$ together with μ_α for all $\alpha \in \{0, 1\}^n$ such that $|\alpha| \geq 2$.

A.2. A reparametrization for general Markov models

Let $T = (V, E)$ be a rooted tree with n leaves and root r. Note that for a tree $1 + 2|E| = |V| + |E|$ so the number of free parameters in (5) and (7) is $|V| + |E|$. We define a polynomial map $f_{\theta\omega} : \mathbb{R}^{|V|+|E|} \to \mathbb{R}^{|V|+|E|}$ from the original set of parameters of Θ_T given by the root distribution and the conditional probabilities for each of the edges to a set of parameters given as follows:

$$\begin{aligned} \eta_{u,v} &= \theta^{(v)}_{1|1} - \theta^{(v)}_{1|0} \qquad \text{for each } (u, v) \in E \quad \text{and} \\ \bar{\mu}_v &= 1 - 2\lambda_v \qquad \text{for each } v \in V, \end{aligned} \tag{35}$$

where $\lambda_v = \mathbb{E}Y_v$ is a polynomial in the original parameters θ of degree depending on the path from the root to v. Let $(r, v_1, \ldots, v_k, v)$ be a directed path in T. Then

$$\lambda_v = \sum_{\alpha \in \{0,1\}^{k+1}} \theta^{(v)}_{1|\alpha_k} \theta^{(v_k)}_{\alpha_k|\alpha_{k-1}} \cdots \theta^{(r)}_{\alpha_r}.$$

Let $\Omega_T = f_{\theta\omega}(\Theta_T)$. The inverse map $f_{\omega\theta} : \Omega_T \to \Theta_T$ has the following form. For each edge $(u, v) \in E$ we have

$$\begin{aligned} \theta^{(v)}_{1|0} &= \frac{1 - \bar{\mu}_v}{2} - \eta_{u,v} \frac{1 - \bar{\mu}_u}{2}, \\ \theta^{(v)}_{1|1} &= \frac{1 - \bar{\mu}_v}{2} + \eta_{u,v} \frac{1 + \bar{\mu}_u}{2} \end{aligned} \tag{36}$$

and $\theta^{(r)}_1 = \frac{1 - \bar{\mu}_r}{2}$.

A.3. The non-degenerate case

In this section, we derive the submodel of $\mathcal{M}_T^\kappa = \psi_T(\Omega_T)$, defined as the image of ψ_T constrained to the subset Ω_T^0 of Ω_T given by $\bar{\mu}_v^2 < 1$ for all $v \in V$. We define a smooth transformation on Ω_T^0 that enables us to change coordinates from $((\bar{\mu}_v), (\eta_{u,v}))$ to $((\bar{\rho}_v), (\rho_{uv}))$, where

$$\bar{\rho}_v = \frac{2\bar{\mu}_v}{\sqrt{1-\bar{\mu}_v^2}}, \qquad \rho_{uv} = \sqrt{\frac{1-\bar{\mu}_u^2}{1-\bar{\mu}_v^2}}\,\eta_{u,v}. \tag{37}$$

It is easily checked that this map is invertible since

$$\bar{\mu}_v = \frac{\bar{\rho}_v}{\sqrt{4+\bar{\rho}_v^2}}, \qquad \eta_{u,v} = \sqrt{\frac{4+\bar{\rho}_u^2}{4+\bar{\rho}_v^2}}\,\rho_{uv}. \tag{38}$$

The inequality constraints defining Ω_T^0 are given by (26) and the fact that $\bar{\mu}_v \in (-1,1)$ for all $v \in V$. To express this in terms of the new coordinates, let t_v be defined by

$$t_v = \sqrt{1+\left(\frac{\bar{\rho}_v}{2}\right)^2} + \frac{\bar{\rho}_v}{2} \in (0,\infty). \tag{39}$$

Then (26) becomes

$$\begin{aligned} -t_u t_v \le \rho_{uv} \le \frac{t_u}{t_v}, \\ -\frac{1}{t_u t_v} \le \rho_{uv} \le \frac{t_v}{t_u}. \end{aligned} \tag{40}$$

Transform the tree cumulants to a new coordinate system given by $\bar{\rho}_1, \ldots, \bar{\rho}_n$ and

$$\rho_I = \frac{2^{|I|}\kappa_I}{\prod_{i\in I}\sqrt{1-\bar{\mu}_i^2}} \qquad \text{for all } I \in [n]_{\ge 2}, \tag{41}$$

so that ρ_{ij} is the correlation between X_i and X_j. The change of coordinates on Ω_T^0 and $\mathcal{K}_T$ induces a new parametrization of $\mathcal{M}_T^0$. The parametrization is given by the identity on the first n coordinates corresponding to $\bar{\rho}_i$ for $i = 1, \ldots, n$ and

$$\rho_I = \prod_{v\in V(I)\setminus I} \bar{\rho}_v^{\deg(v)-2} \prod_{(u,v)\in E(I)} \rho_{uv} \qquad \text{for all } I \in [n]_{\ge 2}. \tag{42}$$

In particular, each ρ_I has an attractive monomial form. To prove (42), simply substitute (38) and (41) into (22) to obtain

$$\rho_I \prod_{i\in I} \frac{1}{\sqrt{4+\rho_i^2}} = \frac{1}{4+\rho_{r(I)}^2} \prod_{v\in V(I)\setminus I} \left(\frac{\bar{\rho}_v}{\sqrt{4+\bar{\rho}_v^2}}\right)^{\deg v-2} \prod_{(u,v)\in E(I)} \sqrt{\frac{4+\bar{\rho}_u^2}{4+\bar{\rho}_v^2}}\,\rho_{uv}$$

or, equivalently,

$$\rho_I = \prod_{v\in V(I)\setminus I} \bar{\rho}_v^{\deg v-2} \prod_{(u,v)\in E(I)} \rho_{uv}$$
$$\times \frac{1}{4+\rho_{r(I)}^2} \prod_{v\in V(I)} \left(\frac{1}{\sqrt{4+\bar{\rho}_v^2}}\right)^{\deg v-2} \prod_{v\in V(I)} \sqrt{\frac{4+\bar{\rho}^2_{\mathrm{pa}(v)}}{4+\bar{\rho}_v^2}}.$$

Next, we show that the term in the second line of the equation above is equal to one. This follows from the fact that every $v \in V(I)$ apart from the root is a parent of exactly $\deg(v)-1$ nodes and has one parent; the root has no parents and is a parent of $\deg(r(I))$ nodes.

Appendix B: Proof of Proposition 4.1

It suffices to prove (22) for $I=[n]$ because the general result for $I \subset [n]$ obviously follows by restriction to the subtree $T(I)$ since each inner node of $T(I)$ has degree at most three. The proof proceeds by induction with respect to the number of leaves of T. First, we show that the result is true for $n=2$. Since by definition $\kappa_{12}=\mu_{12}$ we need to prove that

$$\mu_{12} = \frac{1}{4}(1-\bar{\mu}_r^2) \prod_{(u,v)\in E} \eta_{u,v}, \tag{43}$$

where r is the root of T. If any of the nodes of V represents a degenerate random variable, then the global Markov properties in (4) imply that $X_1 \perp\!\!\!\perp X_2$. In this case, the left-hand side of (43) is zero. However, as we show next, one of the factors on the right-hand side of (43) must vanish as well. We prove this by contradiction. Suppose that both $\bar{\mu}_r^2 \neq 1$ and $\eta_{u,v} \neq 0$ for all $(u,v)\in E$. By Remark 4.3, this implies that all the nodes of T represent non-degenerate random variables, which leads to contradiction.

So assume now that every random variable in the system is non-degenerate. From (12), by taking $I=\{1\}$, $J=\{2\}$, we have

$$\mu_{12} = \tfrac{1}{4}(1-\bar{\mu}_r^2)\eta_{r,1}\eta_{r,2}$$

so it suffices to show that

$$\begin{aligned}(1-\bar{\mu}_r^2)\eta_{r,1} &= (1-\bar{\mu}_r^2) \prod_{(u,v)\in E(r1)} \eta_{u,v} \quad \text{and}\\ (1-\bar{\mu}_r^2)\eta_{r,2} &= (1-\bar{\mu}_r^2) \prod_{(u,v)\in E(r2)} \eta_{u,v}.\end{aligned} \tag{44}$$

If $r=1$ or r is a parent of 1, then the first equation in (44) is trivially satisfied. Assume that the length of the path between r and 1 is greater than one. Let $(r, h_m, h_{m-1}, \ldots, h_1, 1)$ be the directed path $E(r1)$ joining r with 1. Then, because $Y_r \perp\!\!\!\perp Y_1 | Y_{h_1}$, by (12) we have that

$$\tfrac{1}{4}(1-\bar{\mu}_r^2)\eta_{r,1} = \mu_{r1} = \tfrac{1}{4}(1-\bar{\mu}_{h_1}^2)\eta_{h_1,r}\eta_{h_1,1}. \tag{45}$$

Similarly, because $Y_r \perp\!\!\!\perp Y_{h_k} | Y_{h_{k+1}}$ for each $k = 1, \ldots, m-1$, then again by (12)

$$\tfrac{1}{4}(1 - \bar{\mu}_{h_k}^2)\eta_{h_k,r} = \tfrac{1}{4}(1 - \bar{\mu}_{h_{k+1}}^2)\eta_{h_{k+1},r}\eta_{h_{k+1},h_k}.$$

Substituting this expression for all subsequent $k = 1, \ldots, m-1$ into (45) we can now conclude that

$$\tfrac{1}{4}(1 - \bar{\mu}_r^2)\eta_{r,1} = \tfrac{1}{4}(1 - \bar{\mu}_{h_m}^2)\eta_{h_m,r}\eta_{h_m,h_{m-1}} \cdots \eta_{h_2,h_1}\eta_{h_1,1}. \tag{46}$$

But since $\frac{1}{4}(1 - \bar{\mu}_{h_m}^2)\eta_{h_m,r} = \mu_{rh_m} = \frac{1}{4}(1 - \bar{\mu}_r^2)\eta_{r,h_m}$, equation (46) implies that

$$(1 - \bar{\mu}_r^2)\eta_{r,1} = (1 - \bar{\mu}_r^2) \prod_{(u,v)\in E(r1)} \eta_{u,v}. \tag{47}$$

The second equation in (44) is proved simply by changing the index from 1 to 2 above.

Now assume the proposition is true for all $k \le n-1$ and let T be a tree with n leaves. If one of the inner nodes of T is degenerate, then by the global Markov properties in (4) there exists an edge split $C_1|C_2$ of the set of leaves such that $X_{C_1} \perp\!\!\!\perp X_{C_2}$. The left-hand side is zero by Lemma 3.2. Again, by Remark 4.3, if both $\bar{\mu}_r^2 \neq 1$ and $\eta_{u,v} \neq 0$ for all $(u, v) \in E$, then $\bar{\mu}_v^2 \neq 1$ for all $v \in V$. Hence, on the right-hand side of equation (43), either $\bar{\mu}_r^2 = 1$ or one of the $\eta_{u,v}$ vanishes. Consequently, (43) is satisfied.

We assume now that all the inner nodes of T represent non-degenerate random variables. As $n \ge 3$, we can always find two leaves separated from all the other leaves by an inner node. We shall call such a pair an *extended cherry*. Denote the leaves by $1, 2$ and the inner node by a. Let $A = \{3, \ldots, n\}$ and let $T(aA)$ be the minimal subtree of T spanned $a \cup A$. Note that the global Markov properties in (4) give that, for each $C \subseteq A$, we have $(X_1, X_2) \perp\!\!\!\perp X_C | H_a$. Using (12), we can conclude that

$$\mu_{12C} = \mu_{12}\mu_C + \tfrac{1}{4}(1 - \bar{\mu}_a^2)\eta_{a,12}\eta_{a,C} = \mu_{12}\mu_C + \eta_{a,12}\mu_{aC}. \tag{48}$$

Let $e \in E$ be the edge incident with a separating 1 and 2 from all other leaves, that is, such that e induces the split $\nu = 12|\hat{1}_A$. For each $\pi \in \Pi_T$, if π is induced by removing $E_\pi \subset E$, then $\pi \wedge \nu$ is induced by removing $E_\pi \cup e$. Let $\rho = 12|\hat{0}_A \in \Pi_T$. Since $\{1, 2\}$ forms an extended cherry and all the inner nodes of T have degree at most three, it follows that a necessarily has degree three in T and is a leaf of $T(aA)$. The *trimming map* with respect to $\{1, 2\}$ is the map $[\rho, \hat{1}] \to \Pi_{T(aA)}$ such that $\pi \mapsto \widetilde{\pi}$ is defined by changing the block $12C$ in $\pi \in [\rho, \hat{1}]$ to aC. Note that the trimming map constitutes an isomorphism of posets between $[\rho, \hat{1}]$ and $\Pi_{T(aA)}$.

It follows from the definition of tree cumulants in (17) that

$$\kappa_{1\cdots n} = \sum_{\pi\in[\rho,\hat{1}]} \mathfrak{m}(\pi, \hat{1}) \prod_{B\in\pi} \mu_B + \sum_{\pi\notin[\rho,\hat{1}]} \mathfrak{m}(\pi, \hat{1}) \prod_{B\in\pi} \mu_B. \tag{49}$$

The second summand in (49) is zero since every $\pi \in \Pi_T$ such that $\pi \notin [\rho, \hat{1}]$ necessarily contains either 1 or 2 as one of the blocks and $\mu_1 = \mu_2 = 0$. Applying (48) to each μ_{12C} for each $\pi \in$

$[\rho, \hat{1}]$, we obtain

$$\prod_{B\in\pi} \mu_B = \prod_{B\in\pi\wedge\nu} \mu_B + \eta_{a,12} \prod_{B\in\widetilde{\pi}} \mu_B$$

and hence

$$\kappa_{1\cdots n} = \sum_{\pi\in[\rho,\hat{1}]} \mathfrak{m}(\pi, \hat{1}) \prod_{B\in\pi\wedge\nu} \mu_B + \eta_{a,12} \sum_{\pi\in[\rho,\hat{1}]} \mathfrak{m}(\pi, \hat{1}) \prod_{B\in\widetilde{\pi}} \mu_B. \tag{50}$$

The first summand in (50) can be rewritten as

$$\sum_{\delta\in[\rho,\nu]} \left[\left(\sum_{\pi\wedge\nu=\delta} \mathfrak{m}(\pi, \hat{1}) \right) \prod_{B\in\delta} \mu_B \right]. \tag{51}$$

However, from Lemma 3.1, since $\nu \neq \hat{1}$, for each δ the sum $\sum_{\pi\wedge\nu=\delta} \mathfrak{m}(\pi, \hat{1})$ in (51) is zero. It follows that

$$\kappa_{1\cdots n} = \eta_{a,12} \sum_{\pi\in[\rho,\hat{1}]} \mathfrak{m}(\pi, \hat{1}) \prod_{B\in\widetilde{\pi}} \mu_B.$$

By Proposition 4 in [11], the Möbius function of $[\rho, \hat{1}]$ is equal to the restriction of the Möbius function on Π_T to the interval $[\rho, \hat{1}]$. The trimming map constitutes an isomorphism between $[\rho, \hat{1}]$ and $\Pi_T(aA)$. Consequently, the Möbius function on $[\rho, \hat{1}]$ is equal to the Möbius function on $\Pi_{T(aA)}$. It follows that

$$\begin{aligned}\kappa_{1\cdots n} &= \eta_{a,12} \left(\sum_{\pi\in[\rho,\hat{1}]} \mathfrak{m}(\pi, \hat{1}) \prod_{B\in\widetilde{\pi}} \mu_B \right) \\ &= \eta_{a,12} \left(\sum_{\pi\in\Pi_{T(aA)}} \mathfrak{m}_{aA}(\pi, \hat{1}_{aA}) \prod_{B\in\pi} \mu_B \right) = \eta_{a,12}\kappa_{aA}.\end{aligned}$$

Since $X_1 \perp\!\!\!\perp X_2 | H_a$, by the second equation in Proposition 2.5, $\eta_{a,12} = \bar{\mu}_a \eta_{a,1} \eta_{a,2}$. Since $|aA| = n-1$, by using the induction assumption

$$\kappa_{aA} = \frac{1}{4}\left(1 - \bar{\mu}^2_{r(aA)}\right) \prod_{v\in V(aA)\setminus aA} \bar{\mu}_v^{\deg(v)-2} \prod_{(u,v)\in E(aA)} \eta_{u,v},$$

where the degree is taken in $T(aA)$. We have two possible scenarios: either $r(aA) \neq a$ or $r(aA) = a$. In the first case, $r(a1) = r(a2) = a$ and by (47)

$$\eta_{a,1}\eta_{a,2} = \prod_{(u,v)\in E(12)} \eta_{u,v}$$

and hence

$$\kappa_{1\cdots n} = \left(\bar{\mu}_a \prod_{(u,v)\in E(12)} \eta_{u,v} \right) \kappa_{aA}. \tag{52}$$

In the second case, either $r(a1)=a$ and $r(a2)=r$ or $r(a1)=r$ and $r(a2)=a$ and so

$$\eta_{a,1}\eta_{a,2}=\frac{1-\bar{\mu}_r^2}{1-\bar{\mu}_a^2}\bar{\mu}_a\prod_{(u,v)\in E(12)}\eta_{u,v}.$$

Hence,

$$\kappa_{1\cdots n}=\left(\frac{1-\bar{\mu}_r^2}{1-\bar{\mu}_a^2}\prod_{(u,v)\in E(12)}\eta_{u,v}\right)\kappa_{aA}. \tag{53}$$

The degree of a in T is three and the degree of all the other inner nodes of $T(12)$ is two. Moreover, $E=E(aA)\cup E(12)$ and $V\setminus[n]=(V(aA)\setminus aA)\cup(V(12)\setminus\{1,2\})$. It follows that both (52) and (53) satisfy (22).

Appendix C: Proofs of the theorems

Proof of Theorem 5.4. If each inner node of T has degree at least three in $\widehat{T}$, then for each inner node u it is possible to find $i,j,k\in[n]$ separated by u in $\widehat{T}$. So $\hat{\mu}_{ij}\hat{\mu}_{ik}\hat{\mu}_{jk}\neq 0$. Thus, by (28), we can determine all values $\bar{\mu}_u^2=\hat{\mu}_u^2\neq 1$. Since, by Remark 5.2(ii), all the equivalence classes in $[E\setminus\widehat{E}]$ are just single edges, we can identify all $\eta_{u,v}^2=\hat{\eta}_{u,v}^2\neq 0$ for all $(u,v)\in E\setminus\widehat{E}$ by Lemma 5.3.

We now show that, because all equivalence classes in $[\widehat{E}]$ are singletons, $\eta_{w,w'}=0$ for every $(w,w')\in\widehat{E}$. By construction, for each $(w,w')\in\widehat{E}$, either both w and w' have degrees at least three in $\widehat{T}$ or one of them is a leaf and the other has degree at least three in $\widehat{T}$. Therefore, there exist $i,j\in[n]$ such that $E(ij)\cap\widehat{E}=\{(w,w')\}$ by the construction of $\widehat{E}$. We have that $\hat{\mu}_{ij}=0$. However, $\eta_{u,v}=\hat{\eta}_{u,v}\neq 0$ for all $(u,v)\in E\setminus\widehat{E}$. Because $\bar{\mu}_{r(ij)}^2=\hat{\mu}_{r(ij)}^2\neq 1$, it follows by (27) that $\eta_{w,w'}=0$. Therefore, the values of all the parameters are fixed up to signs and in this case $\widehat{\Omega}_T$ is finite. The proof that there are exactly $2^{|V|-n}$ points in this fiber is provided in Appendix D.

To prove the second statement of Theorem 5.4, first note that, since every inner node of T has degree at least two in $\widehat{T}$, it follows by Lemma 5.1 that for each $v\in V$, $\bar{\mu}_v^2<1$. This implies that the $\hat{p}$-fiber lies in $\Omega_T^0\subset\Omega_T$ as defined in Appendix A.3. We can apply a smooth transformation over this subset to a second space $\Omega_T'\subseteq\mathbb{R}^{|V|+|E|}$ whose coordinates are given by $\bar{\rho}_v$ for $v\in V$ and ρ_{uv} for $(u,v)\in E$. The map is defined by (37) and is invertible with the inverse defined in (38).

To investigate the geometry of the $\hat{p}$-fiber in Ω_T', first list all the defining constraints. For all $i=1,\ldots,n$ we have that $\bar{\mu}_i=\hat{\mu}_i$ because $\hat{p}$ determines the sample means of the observed nodes. Hence the value of $\bar{\rho}_i$ is determined as well. Write $\bar{\rho}_i=\hat{\rho}_i$ for all $i=1,\ldots,n$, where $\hat{\rho}_i$ is the image of $\hat{\mu}_i$ under (37). For each inner node v whose degree in $\widehat{T}$ is at least three, we can find $i,j,k\in[n]$ separated in $\widehat{T}$ by v. The value of $\bar{\mu}_v^2$ is determined by (28), which is well defined because $\hat{\mu}_{ij}\hat{\mu}_{ik}\hat{\mu}_{jk}>0$. Therefore, the value of $\bar{\rho}_v^2$, for each v whose degree in $\widehat{T}$ is at least three, is fixed $\bar{\rho}_v^2=\hat{\rho}_v^2$, where $\hat{\rho}_v^2=\frac{4\hat{\mu}_v^2}{1-\hat{\mu}_v^2}$ by (37).

Next, we show that for every $(u, v) \in \widehat{E}$ we must have that $\rho_{uv} = 0$. This follows by essentially the same argument as in the first part of the proof. Because the degrees of both u and v are at least two, there exist $i, j \in [n]$ such that $E(ij) \cap \widehat{E} = \{(u, v)\}$. In particular, $\hat{\mu}_{ij} = 0$ and so by (27) $\eta_{u,v} = 0$. Moreover, for any path $E(kl)$ in $[E \setminus \widehat{E}]$ the value of ρ_{kl}^2 is constant by Lemma 5.3. So write $\rho_{kl} = \hat{\rho}_{kl}$. By (42), we have that

$$\hat{\rho}_{kl} = \prod_{(u,v)\in E(kl)} \rho_{uv}. \tag{54}$$

Finally, for any degree-two node v the parameter $\bar{\rho}_v$ can take any real value and each ρ_{uv} is constrained to satisfy (40). This completes the list of constraints defining the image of the $\hat{p}$-fiber in Ω'_T.

We now show that this image is diffeomorphic to a union of polyhedra. Let $\rho = ((\bar{\rho}_v), (\rho_{uv}))$ be any point in the transformed $\hat{p}$-fiber. Then ρ lies in a linear subspace $\mathcal{L}$ of $\mathbb{R}^{|V|+|E|}$ given by $\rho_{uv} = 0$ for all $(u, v) \in \widehat{E}$. Since $\rho_{uv} \neq 0$ for all $(u, v) \in E \setminus \widehat{E}$, we can define the following further smooth change of coordinates on $\mathcal{L}$. Let $s : E \to \{-1, 0, 1\}$ be any possible sign assignment for (ρ_{uv}) such that $s(u, v) = \operatorname{sgn}(\rho_{uv})$ and $\operatorname{sgn}(\rho_{ij}) = \prod_{(u,v)\in E(ij)} s(u, v)$ for all $i, j \in [n]$ (cf. Appendix D). Then s induces an open orthant $\mathbb{R}_s^{|E\setminus\widehat{E}|}$ defined by $s(u, v)\rho_{uv} > 0$ for all $(u, v) \in E \setminus \widehat{E}$. Moreover, the disjoint union of $\mathcal{U}_s = \mathbb{R}^{|V|} \times \mathbb{R}_s^{|E\setminus\widehat{E}|} \subset \mathcal{L}$, for all possible sign assignments s, covers the $\hat{p}$-fiber, that is, each point of the $\hat{p}$-fiber lies in one of the $\mathcal{U}_s$. Note also that on each $\mathcal{U}_s$ the sign of $\bar{\rho}_v$ for all nodes of the degree at least three is fixed. This follows from the fact that by (42)

$$\rho_{ijk} = \bar{\rho}_v \prod_{(u,w)\in E(ijk)} \rho_{uw},$$

for any three leaves $i, j, k \in [n]$ separated by v in $\widehat{T}$. Since on each $\mathcal{U}_s$ the signs of ρ_{uw} for all $(u, w) \in E(ijk)$ are fixed, the sign of $\bar{\rho}_v$ also has to be fixed to match the sign of ρ_{ijk}. We write $\bar{\rho}_v = \hat{\rho}_v^s$ on $\mathcal{U}_s$.

On each $\mathcal{U}_s$ define a map to the space $\mathbb{R}^{|V|+|E\setminus\widehat{E}|}$ with coordinates given by ν_{uv} for $(u, v) \in E \setminus \widehat{E}$ and z_v for $v \in V$. The map is a diffeomorphism defined as follows. We set

$$\nu_{uv} = \log(s(u, v)\rho_{uv}) \qquad \text{for all } (u, v) \in E \setminus \widehat{E}.$$

Next, for every $v \in V$ we substitute $\bar{\rho}_v$ for t_v as defined in (39). This is an invertible transformation because

$$\bar{\rho}_v = \frac{t_v^2 - 1}{t_v},$$

which is well defined since $t_v > 0$ for all $v \in V$. We then simply substitute t_v for $z_v = \log t_v$.

In this new coordinate system, the $\hat{p}$-fiber restricted to $\mathcal{U}_s$ is a union of polyhedra. The defining constraints are as follows. First,

$$\begin{aligned} z_i &= \hat{z}_i && \text{for all leaves } i = 1, \ldots, n, \\ z_v &= \hat{z}_v^s && \text{for all } v \text{ with degree at least three in } \widehat{T}. \end{aligned} \tag{55}$$

Here, $\hat{z}_i, \hat{z}_v^s$ are real numbers obtained as images of $\hat{\rho}_i$, $\hat{\rho}_v^s$, respectively. Moreover, for each $E(kl) \in [E \setminus \widehat{E}]$

$$\sum_{(u,v)\in E(kl)} \nu_{uv} = \log|\hat{\rho}_{kl}| \tag{56}$$

subject to additional inequality constraints

$$\begin{aligned} &\nu_{uv} \le \min\{z_u - z_v, z_v - z_u\} \qquad \text{if } s(u,v) = 1, \\ &\nu_{uv} \le \min\{z_u + z_v, -z_u - z_v\} \qquad \text{if } s(u,v) = -1, \text{ for each } (u,v) \in E \setminus \widehat{E} \text{ and} \\ &z_v > 0 \qquad \text{for the inner nodes of degree 2.} \end{aligned} \tag{57}$$

These inequalities follow from (40). Since all these constraints are linear, they define a polyhedron in $\mathbb{R}^{|V|+|E\setminus\widehat{E}|}$. Therefore the $\hat{p}$-fiber is a disjoint union of subsets each of which is diffeomorphic to a polyhedron.

To show the dimension of each polyhedron is equal to $2l_2$, we must ensure that the dimension of the smallest affine subspace containing this polyhedron is $2l_2$. Since $z_v > 0$ for all $v \in V$ it is easily checked that the inequalities in (57) do not induce any equality. Therefore, the description of the affine span is obtained from the description of the polyhedron (given by (55)–(57)) by suppressing all inequalities in (57). The dimension of the ambient space is $|V| + |E \setminus \widehat{E}|$; the codimension is given by the number of equations in (55) and (56). Hence the codimension is equal to $|V| - l_2 + |[E \setminus \widehat{E}]|$. For each $E(kl) \in [E \setminus \widehat{E}]$ one has that $|E(kl)| - 1$ is equal to the number of degree-two nodes in $E(kl)$. By summing over all $E(kl)$ it follows that $|E \setminus \widehat{E}| - |[E \setminus \widehat{E}]| = l_2$. Therefore, the dimension of the polyhedron is given by

$$(|V| + |E \setminus \widehat{E}|) - (|V| - l_2 + |[E \setminus \widehat{E}]|) = 2l_2.$$

Since the dimension of the affine span of a polyhedron is equal to its dimension, the dimension is equal to $2l_2$ as required. □

Proof of Theorem 5.8. Let $V_0 \subseteq \widehat{V}$ and $E_0 \subseteq \widehat{E}$ and

$$\Omega_{(V_0,E_0)} = \{\omega \in \Omega_T\colon\ \bar{\mu}_v^2 = 1 \text{ for all } v \in V_0, \eta_{u,v} = 0 \text{ for all } (u,v) \in E_0\}. \tag{58}$$

We say that (V_0, E_0) is *minimal for* $\widehat{\Sigma}$ if for every point ω in $\Omega_{(V_0,E_0)}$ and for every $i, j \in [n]$ such that $\hat{\mu}_{ij} = 0$ we have that $\mu_{ij}(\omega) = 0$ and furthermore that (V_0, E_0) is minimal with such a property (with respect to inclusion on both coordinates).

To illustrate the motivation behind this definition, consider the tripod tree singular case in Example 5.7. If T is rooted in the inner node, we have four minimal subsets of $2^{\widehat{V}} \times 2^{\widehat{E}}$: $(\{h\}, \varnothing)$, $(\varnothing, \{(h,1), (h,2)\})$, $(\varnothing, \{(h,1), (h,3)\})$ and $(\varnothing, \{(h,2), (h,3)\})$.

We now show that the $\hat{p}$-fiber satisfies

$$\widehat{\Omega}_T = \bigcup_{(V_0,E_0)\ \text{min.}} \Omega_{(V_0,E_0)} \cap \widehat{\Omega}_T. \tag{59}$$

The first inclusion "$\subseteq$" follows from the fact that if $\omega \in \widehat{\Omega}_T$, then $\mu_{ij}(\omega) = \hat{\mu}_{ij}$ for all $i, j \in [n]$. In particular, $\mu_{ij}(\omega) = 0$ whenever $\hat{\mu}_{ij} = 0$. Therefore, $\omega \in \Omega_{(V_0,E_0)} \cap \widehat{\Omega}_T$ for (V_0, E_0) minimal. The second inclusion is obvious.

For each minimal (V_0, E_0) the set $\Omega_{(V_0,E_0)} \cap \widehat{\Omega}_T$ is a union of disjoint manifolds in $\mathbb{R}^{|V|+|E|}$ constrained to Ω_T. To show this, consider first all the connected components $T_i = (V_i, E_i)$ for $i = 1, \dots, k$ of $\widehat{T}$ except isolated inner nodes of $\widehat{T}$. By Remark 5.2(iv), all these components are trees with a set of leaves contained in $[n]$. The projection of the parameter space Ω_T to the parameters for the marginal model $\mathcal{M}^{\kappa}_{T_i}$ is denoted by Ω_i. It is therefore a projection of Ω_T on $\bar{\mu}_v$ for $v \in V_i$ and $\eta_{u,v}$ for $(u, v) \in E_i$. By Theorem 5.4, each component T_i induces a manifold with corners in Ω_i, denoted by $\widehat{\Omega}_i$. Hence there exists a manifold M_i in $\mathbb{R}^{|V_i|+|E_i|}$ such that $\widehat{\Omega}_i = M_i \cap \Omega_i$. The constraints on the remaining coordinates are given by: $\bar{\mu}_v^2 = 1$ for all $v \in V_0$ and $\eta_{u,v} = 0$ for $(u, v) \in E_0$. These algebraic equations define a union $M_{(V_0,E_0)}$ of affine subspaces in $\mathbb{R}^{|\widehat{V}|+|\widehat{E}|}$ with coordinates given by $\bar{\mu}_v$ for $v \in \widehat{V}$ and $\eta_{u,v}$ for $(u, v) \in \widehat{E}$.

For each (V_0, E_0), consider the union of manifolds $M \subset \mathbb{R}^{|V|+|E|}$ given as the Cartesian product of $M_{(V_0,E_0)}$ and M_i for $i = 1, \dots, k$. The restriction of M to Ω_T is exactly $\Omega_{(V_0,E_0)} \cap \widehat{\Omega}_T$. Now we have that

$$\bigcap_{(V_0,E_0)\,\text{min.}} \bigl(M_{(V_0,E_0)} \times M_1 \times \cdots \times M_k\bigr) = \Biggl(\bigcap_{(V_0,E_0)\,\text{min.}} M_{(V_0,E_0)}\Biggr) \times M_1 \times \cdots \times M_k. \tag{60}$$

However, $\bigcap_{(V_0,E_0)\,\text{min.}} M_{(V_0,E_0)}$ is equal to

$$\bigl\{\omega \in \mathbb{R}^{|V|+|E|} \colon \bar{\mu}_v^2 = 1 \text{ for all } v \in \widehat{V}, \eta_{u,v} = 0 \text{ for all } (u, v) \in \widehat{E}\bigr\},$$

where, after the restriction to Ω_T, the intersection in (60) is equal to the deepest singularity. □

Appendix D: Sign patterns for parameters

Let $\hat{p} \in \mathcal{M}_T$ such that each inner node of T has degree at least three in the corresponding forest $\widehat{T}$. By the proof of Theorem 5.4, there is a finite number of points $\theta \in \Theta_T$ such that $f_T(\theta) = \hat{p}$. By definition, this set of points is denoted by $\widehat{\Theta}_T$. Corollary 5.5 gives the formulae for the parameters modulo signs, which suggests that $|\widehat{\Theta}_T| = 2^{|V|+|E|}$. However, not all sign choices are possible. Let m be the number of inner nodes of T. We will show that the number of possible choices of signs is, in fact, equal to 2^m, that is, $|\widehat{\Theta}_T| = 2^m$. We also show how to obtain all the points in $\widehat{\Theta}_T$ given one of them. This construction becomes especially simple when expressed in the new parameters defined by (35).

Let θ be a point in $\widehat{\Theta}_T$ ($\widehat{\Theta}_T$ is finite and non-empty) and let $\omega = f_{\theta\omega}(\theta)$. We assign signs to each edge of T using the map $s : E \to \{-1, 0, 1\}$ such that for every $(u, v) \in E$, $s(u, v) = \operatorname{sgn}(\eta_{u,v})$, where $\eta_{u,v}$ are parameters in ω. Let h be an inner node of T. On $\widehat{\Omega}_T$ we define the operation of local sign switching δ_h such that $\delta_h(\omega) = \omega'$ where $\eta'_{u,v} = -\eta_{u,v}$ if one of the ends of (u, v) is in h and $\eta'_{u,v} = \eta_{u,v}$ otherwise; $\bar{\mu}'_h = -\bar{\mu}_h$ and $\bar{\mu}'_v = \bar{\mu}_v$ for all $v \neq h$. We have that

$\bar{\mu}'_i = \bar{\mu}_i$ and hence $\lambda'_i = \lambda_i$ for all leaves $i = 1, \dots, n$. Let now $I \in [n]_{\geq 2}$. Then, from (22),

$$\kappa_I(\omega') = \frac{1}{4}\left(1 - \bar{\mu}^2_{r(I)}\right) \prod_{v \in V(I)\setminus I} (\bar{\mu}'_v)^{\deg(v)-2} \prod_{(u,v)\in E(I)} \eta'_{u,v}.$$

We have two cases: either h lies in $V(I)$ or not. In the first case,

$$\kappa_I(\omega') = (-1)^{\deg(h)-2}(-1)^{\deg(h)}\kappa_I(\omega) = \kappa_I(\omega).$$

In the second case, $\omega' = \omega$ and hence trivially $\kappa_I(\omega') = \kappa_I(\omega)$. It follows that $\omega' \in \widehat{\Omega}_T$ and therefore the operator $\delta_h : \widehat{\Omega}_T \to \widehat{\Omega}_T$ is well defined. The local sign switchings form a group $\mathcal{G}$ that is isomorphic to the multiplicative group Z_2^m. By composing distinct local switchings we obtain 2^m different points in $\widehat{\Omega}_T$. Hence the orbit of ω in $\widehat{\Omega}_T$ has exactly 2^m elements.

It remains to show that there are no other orbits of $\mathcal{G}$ in $\widehat{\Omega}_T$. Let $\omega \in \widehat{\Omega}_T$ and let ω' be a point in Ω_T such that $(\eta'_{u,v})^2 = \eta_{u,v}{}^2$ for all $(u,v) \in E$ and $(\bar{\mu}'_v)^2 = \bar{\mu}_v^2$ for all inner nodes v of T, which is a necessary condition for ω' to be in $\widehat{\Omega}_T$. Assume that ω' is not in the orbit of ω. We will show below that this implies that ω' cannot lie in the $\hat{p}$-fiber. It will then follow that the orbit of ω constitutes the whole $\widehat{\Omega}_T$ and hence $|\widehat{\Omega}_T| = 2^m$.

We proceed by contradiction. Thus, let $\omega' \in \widehat{\Omega}_T$ and we want to show that $\omega' = \delta(\omega)$ for some $\delta \in \mathcal{G}$. Since ω can be replaced by any other point in its orbit, we can assume that $\mathrm{sgn}(\bar{\mu}_v) = \mathrm{sgn}(\bar{\mu}'_v)$ for all $v \in V$. Since $\omega, \omega' \in \widehat{\Omega}_T$, for every $i, j, k \in [n]$ by (22) applied for κ_{ij} and κ_{ijk}, respectively, we have that

$$\prod_{(u,v)\in E(ij)} s(u,v) = \prod_{(u,v)\in E(ij)} s'(u,v), \qquad \prod_{(u,v)\in E(ijk)} s(u,v) = \prod_{(u,v)\in E(ijk)} s'(u,v).$$

It follows that $\prod_{(u,v)\in E(vi)} s(u,v) = \prod_{(u,v)\in E(vi)} s'(u,v)$ for each inner node v and leaf i. It immediately implies that $s(u,v) = s'(u,v)$ for all $(u,v) \in E$ and hence $\omega = \omega'$. In this way we have shown that ω' is in the orbit of ω under $\mathcal{G}$.

Acknowledgement

We are very grateful to a referee whose extensive comments enabled us to substantially improve this paper.

References

[1] Allman, E.S., Matias, C. and Rhodes, J.A. (2009). Identifiability of parameters in latent structure models with many observed variables. *Ann. Statist.* **37** 3099–3132. MR2549554

[2] Auvray, V., Geurts, P. and Wehenkel, L. (2006). A semi-algebraic description of discrete naive Bayes models with two hidden classes. In *Proc. Ninth International Symposium on Artificial Intelligence and Mathematics, Fort Lauderdale, Florida*. Available at http://anytime.cs.umass.edu/aimath06/.

[3] Balakrishnan, N., Johnson, N.L. and Kotz, S. (1998). A note on relationships between moments, central moments and cumulants from multivariate distributions. *Statist*. *Probab*. *Lett*. **39** 49–54. MR1649335
[4] Chang, J.T. (1996). Full reconstruction of Markov models on evolutionary trees: Identifiability and consistency. *Math*. *Biosci*. **137** 51–73. MR1410044
[5] Feller, W. (1971). *An Introduction to Probability Theory and Its Applications*. *Vol*. *II*, 2nd ed. New York: Wiley. MR0270403
[6] Geiger, D., Heckerman, D., King, H. and Meek, C. (2001). Stratified exponential families: Graphical models and model selection. *Ann*. *Statist*. **29** 505–529. MR1863967
[7] Lauritzen, S.L. (1996). *Graphical Models*. *Oxford Statistical Science Series* **17**. Oxford: Clarendon Press. MR1419991
[8] McCullagh, P. (1987). *Tensor Methods in Statistics*. London: Chapman & Hall. MR0907286
[9] Moulton, V. and Steel, M. (2004). Peeling phylogenetic 'oranges'. *Adv*. *in Appl*. *Math*. **33** 710–727. MR2095862
[10] Pearl, J. and Tarsi, M. (1986). Structuring causal trees. *J*. *Complexity* **2** 60–77. MR0925434
[11] Rota, G.C. (1964). On the foundations of combinatorial theory. I. Theory of Möbius functions. *Probab*. *Theory Related Fields* **2** 340–368. MR0174487
[12] Rota, G.C. and Shen, J. (2000). On the combinatorics of cumulants. *J*. *Combin*. *Theory Ser*. *A* **91** 283–304. MR1779783
[13] Rusakov, D. and Geiger, D. (2005). Asymptotic model selection for naive Bayesian networks. *J*. *Mach*. *Learn*. *Res*. **6** 1–35 (electronic). MR2249813
[14] Semple, C. and Steel, M. (2003). *Phylogenetics*. *Oxford Lecture Series in Mathematics and Its Applications* **24**. Oxford: Oxford Univ. Press. MR2060009
[15] Settimi, R. and Smith, J.Q. (1998). On the geometry of Bayesian graphical models with hidden variables. In *UAI* (G.F. Cooper and M. Serafín, eds.) 472–479. San Francisco: Morgan Kaufmann.
[16] Settimi, R. and Smith, J.Q. (2000). Geometry, moments and conditional independence trees with hidden variables. *Ann*. *Statist*. **28** 1179–1205. MR1811324
[17] Speed, T.P. (1983). Cumulants and partition lattices. *Austral*. *J*. *Statist*. **25** 378–388. MR0725217
[18] Speicher, R. (1994). Multiplicative functions on the lattice of noncrossing partitions and free convolution. *Math*. *Ann*. **298** 611–628. MR1268597
[19] Spiegelhalter, D.J., Dawid, A.P., Lauritzen, S.L. and Cowell, R.G. (1993). Bayesian analysis in expert systems. *Statist*. *Sci*. **8** 219–283. With comments and a rejoinder by the authors. MR1243594
[20] Stanley, R.P. (2002). *Enumerative Combinatorics*. *Volume I*. *Cambridge Studies in Advanced Mathematics* **49**. Cambridge: Cambridge Univ. Press.

Received May 2010 and revised October 2010