ASYMPTOTIC MODEL SELECTION AND IDENTIFIABILITY OF DIRECTED TREE MODELS WITH HIDDEN VARIABLES

PIOTR ZWIERNIK

Abstract. The standard Bayesian Information Criterion (BIC) is derived under some regularity conditions which are not always satisfied by the graphical models with hidden variables. In this paper we derive the BIC score for Bayesian networks in the case when the data is binary and the underlying graph is a rooted tree and all the inner nodes represent hidden variables. This provides a direct generalization of a similar formula given by Rusakov and Geiger in [10]. Geometric results obtained in this paper are complementary to the results in the previous paper [18] extending our understanding of this class of models. The main tool used in this paper is the connection between asymptotic approximation of Laplace integrals and the real log-canonical threshold.

1. Introduction

Let $X^{(N)} = X^1, \dots, X^N$ be a random sample of random variables with values in a finite discrete space $[k] := \{1, \dots, k\}$. We assume that the true model for the data is $\mathcal{M}_0$. In this paper we focus on the Bayesian approach to the model selection. Hence, given a finite set of possible parametric models we choose a model $\mathcal{M}$ with likelihood function $f(\theta; X^{(N)}, \mathcal{M}) = \mathbb{P}(X^{(N)}|\mathcal{M}, \theta)$ according to the maximum a posteriori probability given the observed data:

$$\mathbb{P}(\mathcal{M}|X^{(N)}) \propto g(\mathcal{M})\mathbb{P}(X^{(N)}|\mathcal{M}) = g(\mathcal{M}) \int_{\Theta} f(\theta; X^{(N)}, \mathcal{M})\varphi(\theta|\mathcal{M})\mathrm{d}\theta,$$

where $\theta \in \Theta$ denotes the model parameters, $\Theta \subseteq \mathbb{R}^d$ is the parameter space, $g(\mathcal{M})$ is a prior distribution on the set of considered models, and $\varphi(\theta|\mathcal{M})$ is a prior distribution on Θ given $\mathcal{M}$. We focus on the model selection using large sample approximations for $\log \mathbb{P}(\mathcal{M}|X^{(N)})$ called the BIC score.

We assume here that $g(\mathcal{M})$ is uniform on the space of all the considered models. This reduces the problem to maximizing the marginal likelihood $\mathbb{P}(X^{(N)}|\mathcal{M})$. From now on we fix $\mathcal{M}$ and hence we frequently omit it in the notation. The model is assumed to be an image of a real analytic map $p : \Theta \to \Delta_{k-1}$, where Δ_{k-1} is the probability simplex $\{x \in \mathbb{R}^k : x_i \geq 0, \sum x_i = 1\}$. We assume that $\Theta \subset \mathbb{R}^d$ is a compact and *semianalytic set*, i.e. $\Theta = \{x \in \mathbb{R}^d : g_1(x) \geq 0, \dots, g_l(x) \geq 0\}$, where g_i are real analytic functions. The prior distribution $\varphi : \Theta \to \mathbb{R}$ is assumed to be strictly positive, bounded and smooth on Θ.

Assuming the observations in $X^{(N)}$ are independent we can write $Z(N) := \log \mathbb{P}(X^{(N)}|\mathcal{M})$ as a function of sample proportions $\hat{p}_i^{(N)} = N_i/N \in \Delta_{k-1}$ where N_i is the number of times that pattern $i \in [k]$ was observed in the data $X^{(N)}$. Let

Key words and phrases. Bayesian model selection, BIC, marginal likelihood, singular models, tree models, Bayesian networks, hidden data, asymptotic approximation of integrals, Laplace integrals, real canonical threshold, Newton diagram method, pair-edge incidence polytope.

$\ell(p(\theta);X) = \log f(\theta;X)$ be the log-likelihood for a single observation. Then the log-likelihood of the data can be rewritten as

$$\ell_N(p(\theta)) = \sum_{i=1}^{k} N_i \log p_i(\theta) = N\ell(p(\theta);\hat{p}^{(N)}). \tag{1}$$

In this paper following [10] we always assume that there exists N_0 such that $\hat{p}^{(N)} = \hat{p} \in \mathcal{M}$ for all $N > N_0$ and $\hat{p}$ has positive entries. With this assumption the maximum likelihood estimates are given as $\widehat{\Theta} := p^{-1}(\hat{p}) \subset \Theta$. This follows from the properties of the multinomial distribution. For given $\hat{p}$ define the normalized log-likelihood as a function $f : \Theta \to \mathbb{R}$

$$f(\theta) = f(p(\theta);\hat{p}) = \ell(\hat{p};\hat{p}) - \ell(p(\theta);\hat{p}) \geq 0 \tag{2}$$

and denote

$$I(N) := \int_{\Theta} \exp\{-Nf(\theta))\}\,\varphi(\theta)\mathrm{d}\theta. \tag{3}$$

Then the logarithm of the marginal likelihood can be written as $Z(N) = \hat{\ell}_N + \log I(N)$, where $\hat{\ell}_N = \ell_N(\hat{p})$.

It follows from [1], [7], [16] that under the above assumptions the geometry of $\widehat{\Theta}$ contains all the data we need to obtain the asymptotic approximation for $Z(N)$. This leads to the concept of the real log-canonical threshold of $\widehat{\Theta}$, which will be presented in the next section. In the standard setting one assumes that $\widehat{\Theta}$ is a single point. In this case the posterior distribution is asymptotically normal and to approximate $Z(N)$ for large N we use the Laplace approximation which gives us the BIC score

$$Z(N) = \hat{\ell}_N - \frac{d}{2}\log N + O(1), \tag{4}$$

where $d = \dim \Theta$ and $O(1)$ is a standard notation which means that the omitted term of the approximation is bounded by a constant. This evaluation was first performed by Schwarz [12] for Linear Exponential models and then by Haughton [4] for Curved Exponential models under some additional technical assumptions.

In this paper we investigate the asymptotic approximation for $Z(N)$ for directed graphical models induced by trees such that all the variables in the system are binary and in addition we do not observe the inner nodes. These models in general do not allow us to use the standard asymptotic approximation as in (4) since the MLE in this case is never unique. This involves the advanced analytical tools for approximating general Laplace integrals introduced in this context by Watanabe [15] which link to some earlier results of Varchenko (see e.g. [1]). An algebraic treatment of the real log-canonical threshold presented in [7] and in this paper may simplify some asymptotic approximation techniques used in statistics.

For a given tree T and some data $X^{(N)}$ denote the set of possible MLEs by $\widehat{\Theta}_T \subset \Theta_T$, where Θ_T is the parameter space for the model (for details see Section 3). A surprising fact proved in this paper is that the second-order moments between the observable variables in the system completely determine the geometry of $\widehat{\Theta}_T$ and hence also the asymptotics. In fact the only thing we need to know is the set of sample covariances that vanish. An especially nice formula for the approximation of the marginal likelihood is given in the case of trivalent trees, i.e. the trees such that each inner node has degree three (see Theorem 25). For example if all the covariances between the leaves are nonzero then $\widehat{\Theta}_T$ is a finite set of points and

$$Z(N) = \hat{\ell}_N - \frac{4n-5}{2}\log N + O(1),$$

where n is the number of leaves of T. Since in this case $\dim \Theta_T = 4n-5$ this is exactly the formula in Equation (4). However, in general this formula does not apply. First, the coefficient of $\log N$ can be different than $\dim \Theta$. Second, one can obtain an additional $\log\log N$ term affecting the asymptotics. For example for $n \geq 4$ if all the sample covariances vanish and the tree is rooted in a leaf (for relevant definitions see Section 3) then

$$Z(N) = \hat{\ell}_N - \frac{3n}{4}\log N + O(1).$$

If the tree is rooted in an inner node then

$$Z(N) = \hat{\ell}_N - \frac{3n}{4}\log N + O(\log\log N),$$

where the coefficient in front of $\log\log N$ depends on some additional conditions.

The paper is organized as follows. In Section 2 we provide the theory of asymptotic approximation of the marginal likelihood integrals. This theory allows to approximate marginal likelihood without standard regularity assumptions. We link these concepts with the real log-canonical threshold which allows us to use simple algebraic arguments. In Section 3 we define Bayesian networks on rooted trees and provide a useful parametrization of these models in terms of the tree-cumulants introduced in a previous paper [18]. In Section 4 we analyze the geometry of MLE sets using some combinatorial insight. In Section 5 we present some further results on asymptotics and links with the real log-canonical threshold. In Theorem 24 we provide the approximation formula in the case when this set is a smooth subset of the parameter space. In Theorem 25 we state the main result of this paper which gives formulas for the BIC score for Bayesian networks on trivalent rooted trees with hidden variables. The rest of the paper is devoted to the proof of this result. We first reduce the problem using techniques which mimic the ones proposed by Rusakov and Geiger [10]. Then we finish the proof using some polyhedral geometry and the method of Newton diagrams. The paper is concluded with a short discussion.

Acknowledgements

I am especially grateful to Shaowei Lin for a number of helpful comments and for introducing me to the theory of log-canonical thresholds. I also want to thank Diane Maclagan and Jim Q. Smith for helpful discussions at the early stage of this work. A part of this project was done during my stay at UC Berkeley in October 2009. Special thanks go to American Study and Student Exchange Committee (ASSEC) and the Department of Statistics at the University of Warwick for funding this visit.

2. Asymptotics of marginal likelihood integrals

Given $\theta_0 \in \mathbb{R}^d$, let $\mathcal{A}_{\theta_0}(\mathbb{R}^d)$ be the ring of real-valued functions $f : \mathbb{R}^d \to \mathbb{R}$ that are analytic at θ_0. Given a subset $\Theta \subset \mathbb{R}^d$, let $\mathcal{A}_\Theta(\mathbb{R}^d)$ be the ring of real functions analytic at each point $\theta_0 \in \Theta$. In $\mathcal{A}_\Theta(\mathbb{R}^d)$ each function can be locally represented as a power series centered at θ_0. By $\mathcal{A}_\Theta^+(\mathbb{R}^d)$ we denote a subset of

$\mathcal{A}_\Theta(\mathbb{R}^d)$ consisting of all non-negative functions. Usually the ambient space is clear from the context and in this case we omit it in the notation writing $\mathcal{A}_{\theta_0}$ and so on.

Given a compact subset Θ of $\mathbb{R}^d$, a real analytic function $f \in \mathcal{A}_\Theta^+$ and a smooth positive function $\varphi : \mathbb{R}^d \to \mathbb{R}$, consider the *zeta function* defined as

$$\zeta(z) = \int_\Theta f(\theta)^{-z}\varphi(\theta)\mathrm{d}\theta.$$

This function is extended to a meromorphic function in z on the entire complex line. If $\zeta(z)$ has a smallest pole, we denote the pole and its multiplicity by $\mathrm{rlct}_\Theta(f;\varphi)$ and $\mathrm{mult}_\Theta(f;\varphi)$. By convention if $\zeta(z)$ has no poles then $\mathrm{rlct}_\Theta(f;\varphi) = \infty$ and $\mathrm{mult}_\Theta(f;\varphi) = d$. If $\varphi(\theta) \equiv 1$ then we omit φ in the notation writing $\mathrm{rclt}_\Theta(f)$ and $\mathrm{mult}_\Theta(f)$. Define $\mathrm{RLCT}_\Theta(f;\varphi)$ to be the pair $(\mathrm{rlct}_\Theta(f;\varphi), \mathrm{mult}_\Theta(f;\varphi))$, and we order these pairs so that $(r_1, m_1) > (r_2, m_2)$ if $r_1 > r_2$, or $r_1 = r_2$ and $m_1 < m_2$.

Let $\theta_0 \in \Theta$ and let W_0 be any sufficiently small neighbourhood of θ_0 in $\mathbb{R}^d$. Then then by the proof of [7, Lemma 2.4] $\mathrm{RLCT}_{W_0}(f;\varphi)$ does not depend on the choice of W_0 and it is denoted by $\mathrm{RLCT}_{\theta_0}(f;\varphi)$. If θ_0 lies in the interior of Θ then $W_0 \cap \Theta = W_0$ and $\mathrm{RLCT}_{W_0\cap\Theta}(f;\varphi) = \mathrm{RLCT}_{\theta_0}(f;\varphi)$. However, if θ_0 is a boundary point of Θ that this usually does not hold and one can show that (c.f. [7])

$$\mathrm{rlct}_{W_0\cap\Theta}(f;\varphi) \geq \mathrm{rlct}_{\theta_0}(f;\varphi), \quad \mathrm{mult}_{W_0\cap\Theta}(f;\varphi) \leq \mathrm{mult}_{\theta_0}(f;\varphi). \tag{5}$$

By [7, Proposition 2.5] the set of pairs $\mathrm{RLCT}_{W_0\cap\Theta}(f;\varphi)$ for $\theta_0 \in \Theta$ has a minimum and

$$\mathrm{RLCT}_\Theta(f;\varphi) = \min_{\theta_0\in\Theta} \mathrm{RLCT}_{W_0\cap\Theta}(f;\varphi). \tag{6}$$

In particular if f is the normalized log-likelihood defined in (2) then $\mathrm{RLCT}_\Theta(f;\varphi) = \mathrm{RLCT}_{\widehat{\Theta}}(f;\varphi)$ since for all $\theta \in \Theta \setminus \widehat{\Theta}$ we have $\mathrm{RLCT}_\theta(f;\varphi) = (\infty, d)$.

Section 7.2 in [1] and independently Section 2.4 and Section 6.2 in [16] relate computation of the asymptotic expansion for $I(N)$ in (3) to computation the poles and their multiplicities of $\zeta(z)$ (c.f. [11, Corollary 2]). The following result is based on the results in [15]. To prove it we can alternatively use some techniques from [1].

Theorem 1 ([16], §6)**.** *Let Θ be a compact semianalytic subset of $\mathbb{R}^d$ and $f \in \mathcal{A}_\Theta^+$. Let $I(N)$ be defined as in (3). Then as $N \to \infty$*

$$\log I(N) = -\mathrm{rlct}_\Theta(f;\varphi)\log N + (\mathrm{mult}_\Theta(f;\varphi) - 1)\log\log N + O(1).$$

A guide to the proof. One first shows the theorem in the case when both f and φ are monomials which follows from [1, Theorem 7.3]. The general case follows by applications of the Hironaka's theorem on the resolution of singularities. For detail see also the proof of [7, Theorem 1.1]. □

Remark 2. The important connection between the theory of real log-canonical threshold and the resolution of singularities has been omitted in this paper. For details see [16]. In particular [1, Lemma 7.3] shows that $\mathrm{rltc}_\Theta(f;\phi)$ is always a rational number and $\mathrm{mult}_\Theta(f;\phi)$ is an integer.

Remark 3. Note that there is a substantial difference between the real log-canonical threshold and the log-canonical threshold which is an important invariant used in algebraic geometry (see e.g. [6, Section 9.3.B]). Let $f \in \mathbb{R}[x_1, \ldots, x_d]$ be a polynomial with real coefficients. By $f_\mathbb{C}$ we denote its compactification, i.e. the same polynomial but as a function on $\mathbb{C}^d$. Saito [11] showed that $\mathrm{rlct}(f) \geq \mathrm{lct}(f_\mathbb{C})$.

As an example let $f(x,y,z)=x^2+y^2+z^2$. By Kollar [5, Example 8.15] we have $\mathrm{lct}_0(f_{\mathbb{C}})=1$ and one can easily show that over the real numbers a single blow-up at the origin (see e.g. [16, Section 3.5]) allows to compute the poles of $\zeta(z)$ (c.f. Proposition 3.3 in [11]) giving $\mathrm{rlct}_0(f)=3/2$.

Let $\theta_0\in\Theta$ and $f_1,\ldots,f_r\in\mathcal{A}_{\theta_0}$ then the *ideal generated* by $\{f_1,\ldots,f_r\}$ is by definition $\{f\in\mathcal{A}_{\theta_0}:\ f(\theta)=\sum_{i=1}^r h_i(\theta)f_i(\theta), h_i\in\mathcal{A}_{\theta_0}\}$. We denote it by $\langle f_1,\ldots,f_r\rangle$. Following [7] we generalize the notion of the log-canonical thresholds to an ideal $I=\langle f_1,\ldots,f_r\rangle$. By definition

$$\mathrm{RLCT}_{\theta_0}(I;\varphi)=\mathrm{RLCT}_{\theta_0}(\langle f_1,\ldots,f_r\rangle;\varphi):=\mathrm{RLCT}_{\theta_0}(f;\varphi),$$

where $f(\theta)=f_1^2(\theta)+\cdots+f_r^2(\theta)$. Below we list some basic properties of the real log-canonical threshold (see also [7, Section 4]). Most of them mimic analogous properties of the log-canonical threshold (c.f. [6, Section 9.3.B]). In particular the next result shows that the real log-canonical threshold of an ideal is well defined.

Lemma 4 (Proposition 4.5, [7])**.** *Given $x_0\in\mathbb{R}^d$ let $f_1,\ldots,f_r,g_1,\ldots,g_s\in\mathcal{A}_{x_0}$. If $\{f_1,\ldots,f_r\}$ and $\{g_1,\ldots,g_s\}$ generate the same ideal $I\subset\mathcal{A}_{x_0}$, then*

$$\mathrm{RLCT}_{x_0}(\langle f_1,\ldots,f_r\rangle;\varphi)=\mathrm{RLCT}_{x_0}(\langle g_1,\ldots,g_s\rangle;\varphi).$$

Lemma 5 (Proposition 4.7, [7])**.** *Let $\rho:\mathbb{R}^d\to\Theta$ be a proper real analytic isomorphism and let $f\in\mathcal{A}_{x_0}$. Denote $y_0=\rho^{-1}(x_0)$. Then,*

$$\mathrm{RLCT}_{x_0}(f;\varphi)=\mathrm{RLCT}_{y_0}(f\circ\rho;(\varphi\circ\rho)|\rho'|),$$

where $|\rho'|$ denotes the Jacobian of ρ.

The following three results follow easily from the interpretation of the real log-canonical threshold and its multiplicity as coefficients in the asymptotic approximation of $I(N)$.

Lemma 6. *If φ is positive and bounded on Θ then*

$$\mathrm{RLCT}_\Theta(f;\varphi)=\mathrm{RLCT}_\Theta(f).$$

Lemma 7. *Let $I_x\in\mathcal{A}_0(\mathbb{R}^m)$, $I_y\in\mathcal{A}_0(\mathbb{R}^n)$ be two ideal generated by $f_i(x)$ for $i\in[r]$ and $g_j(y)$ for $j\in[s]$ respectively. Let $I_x+I_y\in\mathcal{A}_0(\mathbb{R}^{m+n})$ denote the ideal generated by all f_i and g_j. If $\mathrm{RLCT}_0(I_x)=(\lambda_x,m_x)$ and $\mathrm{RLCT}_0(I_y)=(\lambda_y,m_y)$ then $\mathrm{RLCT}_0(I_x+I_y)=(\lambda_x+\lambda_y,m_x+m_y-1)$.*

Lemma 8. *Let $f,g\in\mathcal{A}_\Theta$. If there exist constants $c,c'>0$ such that $c\,g(x)\le f(x)\le c'g(x)$ for every $x\in\Theta$ then $\mathrm{RLCT}_\Theta(f)=\mathrm{RLCT}_\Theta(g)$.*

2.1. **Netwon diagram method.** In certain situations there exists a nice combinatorial way to compute the real log-canonical threshold. We base this section on [1, Chapter 8]. An example of an application of these methods in statistical analysis can be found in [17].

Let f be formal power series in $x_1,\ldots,x_d$ such that $f(x)=\sum_\alpha c_\alpha x^\alpha$. The exponents of terms of the polynomial f are vectors in $\mathbb{N}^d$. The *Newton polyhedron* of f denoted by $\Gamma_+(f)$ is the convex hull of the subset

$$\{\alpha+\alpha':\ c_\alpha\neq 0,\alpha'\in\mathbb{R}^d_{\ge 0}\}.$$

A subset $\gamma\subset\Gamma_+(f)$ is a *face* of $\Gamma_+(f)$ if there exists $\beta\in\mathbb{R}^d$ such that

$$\gamma=\{\alpha\in\Gamma_+(f):\ \langle\alpha,\beta\rangle\le\langle\alpha',\beta\rangle\ \text{ for all }\alpha'\in\Gamma_+(f)\}.$$

If γ is a subset of $\Gamma_+(f)$ then we define $f_\gamma(x) = \sum_{\alpha\in\gamma} c_\alpha x^\alpha$. The *principal part* of f is by definition the sum of all f_γ, where γ goes through the set of all compact faces of $\Gamma_+(f)$.

If f is a polynomial then the convex hull of the exponents of the terms in the sum is called the *Newton polytope* and denoted $\Gamma(f)$. In this case $\Gamma_+(f) = \Gamma(f) + \mathbb{R}^d_{\geq 0}$, where the plus denotes the Minkowski sum, where the *Minkowski sum* of two polyhedra is by definition

$$\Gamma_1 + \Gamma_2 = \{x + y \in \mathbb{R}^d : x \in \Gamma_1, y \in \Gamma_2\}.$$

Definition 9. The principal part of the power series f with real coefficients is $\mathbb{R}$-*nondegenerate* if for all compact faces γ of $\Gamma_+(f)$

$$\left\{x \in \mathbb{R}^n : \frac{\partial f_\gamma}{\partial x_1}(x) = \cdots = \frac{\partial f_\gamma}{\partial x_n}(x) = 0\right\} \subseteq \{\omega \in \mathbb{R}^n : x_1 \cdots x_n = 0\}. \tag{7}$$

We say that $I = \langle f_1, \ldots, f_r\rangle$ is $\mathbb{R}$-nondegenerate if $f = \sum f_i^2$ is $\mathbb{R}$-nondegenerate.

From the general theory (c.f. [1]) we know that if the principal part of f is $\mathbb{R}$-nondegenerate and $f \in \mathcal{A}^+_\Theta$ it greatly facilitates the computations in Theorem 1.

Theorem 10. *Let $\theta_0 \in \Theta \subset \mathbb{R}^d$, $f \in \mathcal{A}^+_{\theta_0}$ and $f(\theta_0) = 0$. If the principal part of f is $\mathbb{R}$-nondegenerate then $\mathrm{RLCT}_{\theta_0}(f) = (\frac{1}{t}, c)$ where t is the smallest number such that $t\,(1, \ldots, 1)$ hits $\Gamma_+(f)$ and c is the codimension of the face it hits.*

A guide to the proof. This result uses the toric resolution of singularities. One can apply Theorem 7.6 and Theorem 8.6 in [1]. □

Remark 11. Note that this theorem in general will not give us $\mathrm{RLCT}_{W_0\cap\Theta}$ if θ_0 is a boundary point of Θ. For a discussion see [1, Section 8.3.4] and [7].

3. Independence tree models

In this section we present the statistical models induced by trees. We introduce the marginal likelihood integral for the model. The asymptotic approximation of this integral, which is the main aim of this paper, involves techniques presented in the previous section.

3.1. **General Markov models.** All random variables considered in this paper are assumed to be binary with values in $\{0, 1\}$. Let $T = (V, E)$ be a tree with the vertex set V and the set of edges E. Denote $n_e = |E|$ and $n_v = |V|$. We assume that T is a *rooted tree* with root $r \in V$, i.e. a directed tree with one distinguished vertex r and all the edges directed away from r. For any $e = (k, l) \in E$ we say that k and l are *adjacent* and k is a *parent* of l and we denote it by $k = \mathrm{pa}(l)$. Let $p_\beta = \mathbb{P}(\bigcap_{v\in V}\{Y_v = \beta_v\})$. A *Markov process* on a rooted tree T is a sequence $Y = (Y_v)_{v\in V}$ of random variables such that for each $\beta = (\beta_v)_{v\in V} \in \{0, 1\}^{n_v}$

$$p_\beta(\theta) = \prod_{v\in V} \theta^{(v)}_{\beta_v|\beta_{\mathrm{pa}(v)}}, \tag{8}$$

where $\mathrm{pa}(r) = \emptyset$, $\theta = (\theta^{(v)}_{\beta_v|\beta_{\mathrm{pa}(v)}})$ and $\theta^{(v)}_{\beta_v|\beta_{\mathrm{pa}(v)}} = \mathbb{P}(Y_v = \beta_v | Y_{\mathrm{pa}(v)} = \beta_{\mathrm{pa}(v)})$. In a more standard statistical language these models are just fully observed Bayesian networks on rooted trees.

The models analyzed in this paper are induced from Markov processes on trees by assuming that all the inner nodes represent random variables which are not observed directly. Let $Y = (X, H)$ where $X = (X_1, \ldots, X_n)$ denotes the variables represented by the leaves of T and H denotes the vector of variables represented by inner nodes. By $[n]$ we denote the set of leaves of T. Since $\theta^{(v)}_{0|i} + \theta^{(v)}_{1|i} = 1$ for all $v \in V$ and $i = 0, 1$ then the Markov process on T defined by Equation (8) has exactly $2n_e + 1$ free parameters in the vector θ: one for the root distribution $\theta^{(r)}_1$ and two for each edge $(u, v) \in E$ given by $\theta^{(v)}_{1|0}$ and $\theta^{(v)}_{1|1}$. The parameter space is $\Theta_T = [0, 1]^{2n_e+1}$. Let $\mathcal{M}_T$ be a model in Δ_{2^n-1} obtained by summing out in (8) all possible values of the inner nodes which induces a map $p : \Theta_T \to \Delta_{2^n-1}$ parametrizing $\mathcal{M}_T$ which coordinate-wise for any $\alpha \in \{0, 1\}^n$ is given by

$$p_\alpha(\theta) = \sum_{\mathcal{H}} \prod_{v \in V} \theta^{(v)}_{\beta_v|\beta_{\mathrm{pa}(v)}}, \tag{9}$$

where $\mathcal{H}$ is the set of all vectors $\beta = (\beta_v)_{v \in V}$ such that $\beta_i = \alpha_i$ for all $i \in [n]$. We call it a *general Markov model* on T.

Let $X^{(N)} = (X^1, \ldots, X^N)$ denote observations of the random vector representing the leaves of T and let (N_α) for $\alpha \in \{0, 1\}^n$ be the sufficient statistic given by sample counts. As the introduction we write $\hat{p} = [\hat{p}_\alpha]$, where $\hat{p}_\alpha = N_\alpha/N$, for sample proportions assuming $\hat{p} \in \mathcal{M}_T$. Define the normalized likelihood function $f : \Theta_T \to \mathbb{R}$ as in (2). Then $f \in \mathcal{A}^+_{\Theta_T}$. Since by assumption $\varphi(\theta)$ is positive and bounded on Θ_T then by Theorem 1 and Lemma 6 to obtain the asymptotic approximation for $I(N)$ we need to compute $\mathrm{RLCT}_{\Theta_T}(f) = \mathrm{RLCT}_{\widehat{\Theta}_T}(f)$ (c.f. Equation 6). Let $\theta_0 \in \widehat{\Theta}_T$ and let

$$I_{\theta_0} = \bigcap_{\alpha \in \{0,1\}^n} \langle p_\alpha(\theta) - \hat{p}_\alpha \rangle \subset \mathcal{A}_{\theta_0} \tag{10}$$

be the ideal defining $\widehat{\Theta}_T$ locally near θ_0. By [7, Theorem 1.2] we have

$$\mathrm{RLCT}_{\theta_0}(f) = \mathrm{RLCT}_{\theta_0}(I_{\theta_0}). \tag{11}$$

3.2. **A reparametrization of the model.** In a previous paper [18] we derived a formula for a useful alternative coordinate system for $\mathcal{M}_T$. It allows us to find another more convenient basis for I_{θ_0}. To introduce the result we define a partially ordered set (poset) Π_T (for details see [18, Section 4]) of all partitions of the set of leaves $[n]$ obtained by removing some inner edges of T and considering connected components of the resulting forest. The order on this poset is defined as follows. For any $E_1 \subseteq E$ let $\overline{E}_1$ denote the maximal with respect to inclusion subset of E inducing the same partition of $[n]$ as E_1. Now if partition π_1 is obtained by removing a subset E_1 of the set of edges and π_2 is obtained by removing E_2 then $\pi_1 \preceq \pi_2$ if and only if $\overline{E}_1 \subseteq \overline{E}_2$. In other words $\pi_1 \preceq \pi_2$ if and only if π_2 is a refinement of π_1, i.e. can be obtained from π_1 after removing some additional inner edges of T.

For any poset Π we define its Möbius function $m : \Pi \times \Pi \to \mathbb{R}$ as a uniquely defined function satisfying (c.f. [14, Chapter 3])

$$m(\pi, \pi) = 1 \text{ for every } \pi \in \Pi \text{ and } m(\pi_1, \pi_2) = -\sum_{\pi_1 \preceq \pi \prec \pi_2} m_I(\pi_1, \pi).$$

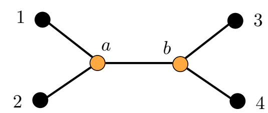
For any $I \subseteq [n]$ we define $T(I)$ as the minimal subtree of T spanned by I. By m_I we denote the Möbius function of $\Pi_{T(I)}$.

The change of the basis for I_{θ_0} is performed as follows. First we express the raw probabilities $[p_\alpha(\theta)]$ for $\alpha \in \{0,1\}^n$ in terms of a new system of polynomials given by all the means $\lambda_i(\theta) = \mathbb{E}X_i$ supplemented with all the central moments $\mu_I(\theta) = \mathbb{E}\left(\prod_{i\in I}(X_i - \lambda_i)\right)$ for $I \in [n]_{\geq 2}$ (for details see [18]), where $[n]_{\geq k}$ denotes all subsets of $[n]$ with at least k elements. Finally, make a further change of the basis such that the first n polynomials are $s_i(\theta) = 1 - 2\lambda_i(\theta) \in [-1,1]$ and the remaining ones are the tree cumulants defined as

$$\kappa_I(\theta) = \sum_{\pi \in \Pi_{\overline{T}(I)}} m_I(0_I, \pi) \prod_{B\in\pi} \mu_B(\theta) \quad \text{for all } I \subseteq [n], \tag{12}$$

where 0_I denotes I as an element of $\Pi_{T(I)}$, i.e. the trivial partition with no inner edges deleted and $\overline{T}$ is tree such that T can be obtained from $\overline{T}$ by edge contractions and each inner node of $\overline{T}$ has degree at most three. The new basis for I_{θ_0} is expressed in terms of $(s_i(\theta))_{i\in[n]}$ and $(\kappa_I(\theta))_{I\in[n]_{\geq 2}}$. In this paper the exact form of the map in Equation (12) is not important. We just use the fact that this map is a regular polynomial map with a constant Jacobian (equal to one here). Nevertheless we provide a simple example presenting the basic idea behind the change of basis.

Example 12. Let T be the quartet tree below



There is only one inner edge. For example to obtain κ_{1234} using Equation (12) set $I = \{1,2,3,4\}$ and note that $\Pi_{T(I)}$ has only two elements: the trivial partition $0 = (1234)$ and the partition obtained by removing the inner edge $(12)(34)$. By the definition of the Möbius function $m_I(0,0) = 1$ and $m_I(0,(12)(34)) = -m_I(0,0) = -1$ and hence

$$\kappa_{1234} = \mu_{1234} - \mu_{12}\mu_{34},$$

where the central moments can be easily computed from the probability distribution. To compute κ_{123} note that $\Pi_T(123)$ consists of two elements: $0 = (123)$ and $(12)(3)$. Then since $\mu_3 = 0$ we have

$$\kappa_{123} = \mu_{123} - \mu_{12}\mu_3 = \mu_{123}.$$

In [18] it is also shown that polynomial maps defined above have inverse maps which are also regular polynomial maps and hence they induce an isomorphism between I_{θ_0} and another ideal with the basis expressed in terms of tree cumulants.

Corollary 13. *Let $\hat{p} \in \mathcal{M}_T$, $\theta_0 \in \widehat{\Theta}_T$ and $I_{\theta_0} \subset \mathcal{A}_{\theta_0}$ be as defined in (10). Then*

$$I_{\theta_0} = \left(\bigcup_{i=1}^{n} \langle s_i(\theta) - \hat{s}_i\rangle\right) + \left(\bigcup_{I\in[n]_{\geq 2}} \langle \kappa_I(\theta) - \hat{\kappa}_I\rangle\right),$$

where $\hat{s}_i = s_i(\theta)$ and $\hat{\kappa}_I = \kappa_I(\theta)$ for any $\theta \in \widehat{\Theta}_T$.

The next step is to change the coordinates to obtain a more transparent representation of $\widehat{\Theta}_T$. Define the following set of $n_v + n_e$ parameters. For every directed edge $(u,v)\in E$

$$\begin{aligned}\eta_{uv} &= \theta^{(v)}_{1|1}-\theta^{(v)}_{1|0}\in[-1,1] \text{ and}\\ s_v &= 1-2\lambda_v\in[-1,1]\quad \text{for each } v\in V,\end{aligned}\tag{13}$$

where $\lambda_v=\mathbb{E}Y_v$ is a polynomial in the original parameters θ of degree depending on the distance of v from the root r. If $r,v_1,\ldots,v_k,v$ is a directed path in T then

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

We denote the new parameter space by $\Omega_T\subset[-1,1]^{n_e+n_v}$ and the coordinates by $\omega=((s_v),(\eta_e))$ for $v\in V, e\in E$. Note that s_i for $i\in[n]$ are both in the system of tree cumulants and in the set of parameters. Since T is a tree than one can easily show that $2n_e+1=n_v+n_e$ and hence $\dim\Omega_T=\dim\Theta_T$. The change of parameters defined above is denoted by $f_{\theta\omega}:\Theta_T\to\Omega_T$. It is a regular polynomial map with a regular inverse (see [18, Section 3.2]). We write $N(I)$ for the set of inner nodes of $T(I)$ and $E(I)$ for its set of edges. The parametrization of $\mathcal{M}_T$ in the system of tree cumulants is given by the following proposition.

Proposition 14 (Proposition 13 in [18])**.** *Let $T=(V,E)$ be a rooted tree with n leaves. Then for each $i\in[n]$ one has $s_i(\omega)=s_i$. Moreover, for any $I\in[n]_{\geq 2}$ let $r(I)$ be the unique root of $T(I)$. Then*

$$\kappa_I(\omega)=\frac{1}{4}(1-s^2_{r(I)})\prod_{v\in N(I)}s_v^{\deg v-2}\prod_{(u,v)\in E(I)}\eta_{uv}\quad \textit{for all } I\in[n]_{\geq 2},\tag{14}$$

where the degree of $v\in N(I)$ is considered in the subtree $T(I)$.

For any two variables Y_u,Y_v define $\eta_{u,v}=\mathrm{Cov}(Y_u,Y_v)/\mathrm{VarY_u}$ where $\mathrm{Var}(Y_u)=\frac{1}{4}(1-s_u^2)$. One can show that if $(u,v)\in E$ then η_{uv} defined above concides with $\eta_{u,v}$ and we have $(1-s_u^2)\eta_{u,v}=(1-s_v^2)\eta_{v,u}$. The geometric structure of Ω_T is more complicated than the structure of Θ_T. The new parameter space looks as follows. For any choice of the values for $(s_v)_{v\in V}$, where $s_v\in[-1,1]$, we obtain the following constraints on the remaining parameters (c.f. [18, Equation (14)])

$$\begin{aligned}-\min\{(1+s_u)(1+s_v),(1-s_u)(1-s_v)\}&\leq(1-s_u^2)\eta_{u,v}\leq\\ &\leq\min\{(1+s_u)(1-s_v),(1-s_u)(1+s_v)\}.\end{aligned}\tag{15}$$

Recall that we assume that for N large enough $\hat{p}^{(N)}=\hat{p}\in\mathcal{M}_T$ and then $p^{-1}(\hat{p})=\widehat{\Theta}_T$ is the MLE. Since $\omega(\theta)$ is an isomorphism with a constant Jacobian then by Lemma 5 and Lemma 6 we have $\mathrm{RLCT}_{\theta_0}(f(\theta))=\mathrm{RLCT}_{\omega_0}(f(\omega))$, where $\omega_0=\omega(\theta_0)$. Consequently one has $\mathrm{RLCT}_{\widehat{\Theta}_T}(f(\theta))=\mathrm{RLCT}_{\widehat{\Omega}_T}(f(\theta(\omega)))$, where $\widehat{\Omega}_T=\omega(\widehat{\Theta}_T)$. We call $\widehat{\Omega}_T$ the *ML fiber* of $\hat{p}$.

By Corollary 13 to compute $\mathrm{RLCT}_{\omega_0}(f)$ it suffices to compute the real log-canonical threshold at ω_0 for the ideal in $\mathcal{A}_{\omega_0}$ generated by $s_i(\omega)-\hat{s}_i=s_i-s_i^0$ for all $i\in[n]$ and $\kappa_I(\omega)-\hat{\kappa}_I$ for all $I\in[n]_{\geq 2}$. If T is rooted in an inner node then $\kappa_I(\omega)$ does not depend on s_i for any $i\in[n]$ (c.f. Equation (14)) and hence by Lemma 7 one can compute separately $\mathrm{RLCT}_{\omega_0}(\langle s_i-s_i^0\rangle)=(1/2,1)$ for all $i\in[n]$ and

$$(\lambda,m)=\mathrm{RLCT}_{\omega_0}\left(\sum_{I\in[n]_{\geq 2}}(\kappa_I(\omega)-\hat{\kappa}_I)^2\right).\tag{16}$$

Then one has

$$\mathrm{RLCT}_{\omega_0}(I_{\omega_0}) = (n/2 + \lambda, m) \tag{17}$$

We assumed above that T is rooted in an inner node. However, this reduction remains valid if T is rooted in one of the leafs - denote it by r. Indeed, by our assumption $\hat{p}$ has only positive entries and then all the leaves represent non-degenerate random variables implying $s_i^0 \in (-1, 1)$ for all $i \in [n]$. It follows that $(1 - s_r^2)$ is a unit in $\mathcal{A}_{\omega_0}$ and for all $I \in [n]_{\geq 2}$ the ideal $\langle \kappa_I(\omega) - \hat{\kappa}_I \rangle \subset \mathcal{A}_{\omega_0}$ does not depend on s_r which again allows us to use Lemma 7.

For a fixed $\hat{p}$ we denote by $J_{\omega_0} \subset \mathcal{A}_{\omega_0}$ the ideal generated by $\kappa_I(\omega) - \hat{\kappa}_I$ for all $I \in [n]_{\geq 2}$. By (16) and (17) the real canonical threshold of this ideal allows us to compute the real log-canonical threshold of I_{ω_0}. In the rest of the paper we focus on computing the real log-canonical threshold of J_{ω_0} for different points $\omega_0 \in \widehat{\Omega}_T$. This by Equation (11) allows us to compute $\mathrm{RLCT}_{\theta_0}(f)$ for points $\theta_0 \in \widehat{\Theta}_T$ and consequently $\mathrm{RLCT}_{\Theta}(f)$.

4. Singularities and the geometry of ML fibers

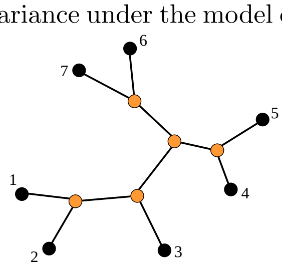
The geometry of the ML fiber drives the asymptotics of $I(N)$. In this section we analyze this using some ideas similar to the ones presented in a different context by Moulton and Steel in [9, Section 6]. This generalizes similar results for the star trees (c.f. [3, Theorem 7]).

For $\hat{p} \in \mathcal{M}_T$ let $\hat{\Sigma} = [\hat{\kappa}_{ij}] \in \mathbb{R}^{n \times n}$ be the matrix of all pairwise sample covariances between the leaves of T. We show that the geometry of ML fibers is determined by zeros in $\hat{\Sigma}$. From Equation (14) for any $\omega_0 = (s_v^0, \eta_e^0) \in \widehat{\Omega}_T$ we have

$$\hat{\kappa}_{ij} = \kappa_{ij}(\omega_0) = \frac{1}{4}\left(1 - (s^0_{r(ij)})^2\right) \prod_{(u,v) \in E(ij)} \eta^0_{uv}. \tag{18}$$

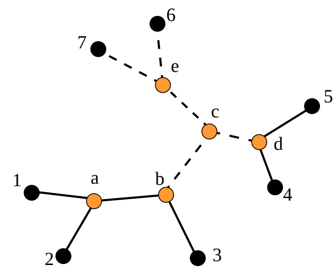
We say that that an edge $e \in E$ is *isolated relative to* $\hat{p}$ if $\hat{\kappa}_{ij} = 0$ for all $i, j \in [n]$ such that $e \in E(ij)$. By $\widehat{E} \subseteq E$ we denote the set of all edges of T which are isolated relative to $\hat{p}$. By $\widehat{T} = (V, E \setminus \widehat{E})$ we denote the forest obtained from T by removing edges in $\widehat{E}$.

Example 15. Let T be a tree given below and assume that the estimator of the sample covariance under the model contains zeros given in the following 7×7-matrix



$$\hat{\mu} = \begin{bmatrix} * & * & * & 0 & 0 & 0 & 0 \\ & * & * & 0 & 0 & 0 & 0 \\ & & * & 0 & 0 & 0 & 0 \\ & & & * & * & 0 & 0 \\ & & & & * & 0 & 0 \\ & & & & & * & 0 \\ & & & & & & * \end{bmatrix}$$

where the asterisks mean any non-zero values making the matrix positive semi-definite. One can check that $\widehat{E} = \{(b, c), (c, d), (c, e), (e, 6), (e, 7)\}$ and if we depict the edges isolated relative to $\hat{p}$ by dashed lines then $\widehat{T}$ looks as follows



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 incident with both e and e' are isolated relative to $\hat{p}$. Let us now take the transitive closure of $\sim$ restricted to pairs of edges in $\widehat{E}$ to form an equivalence relation on $\widehat{E}$. Similarly, 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.

Example 16. Consider the tree from the example above. Here $[\widehat{E}]$ is one element given by a subtree of T spanned by $\{b,d,6,7\}$ and

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

By the 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 then the node represents a non-degenerate random variable.

Lemma 17. *If $v\in V$ is an inner node of T such that $\deg(v)\geq 2$ in $\widehat{T}$ then there exists $\epsilon>0$ such that $(s_v^0)^2\leq 1-\epsilon$ for every $\omega_0\in\widehat{\Omega}_T$. In particular variable H_v cannot be degenerate.*

Proof. By definition of $\widehat{T}$ one can always find two leaves $i,j\in[n]$ such that $\hat{\kappa}_{ij}\neq 0$ and v lies on the path $\mathcal{P}_T(i,j)$. Since $X_i \perp\!\!\!\perp X_j|H_v$ then one can easily show (c.f. [18, Equation (19)]) that

$$\hat{\kappa}_{ij}=\frac{1}{4}(1-(s_v^0)^2)\eta_{v,i}(\omega_0)\eta_{v,j}(\omega_0)$$

and $|\eta_{v,i}(\omega_0)|,|\eta_{v,j}(\omega_0)|\leq 1$. We obtain the result taking $\epsilon=4|\hat{\kappa}_{ij}|$. $\square$

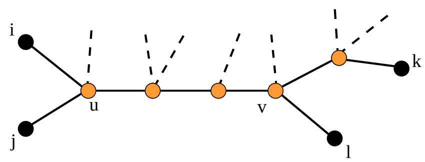
We now formulate a lemma which is partly based on Lemma 6.4 in [9].

Lemma 18. *Let $T=(V,E)$ be a tree with n leaves, and suppose $\hat{p}\in\mathcal{M}_T$.*

- **(i):** *The edges in any equivalence class of $[\widehat{E}]$ form a connected subgraph of T. Moreover, 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 $[E\setminus\widehat{E}]$ form a path in T. For any such path $\mathcal{P}_T(u,v)$ $\kappa_{uv}(\omega)$ is constant on $\widehat{\Omega}_T$ and one can identify its value up to the sign from the sample proportions $\hat{p}$.*

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

Proof. The only not obvious statement is the second part of (iii). First note that degree of each inner node of $\mathcal{P}_T(u,v)$ in $\widehat{T}$ must be at least two. Moreover, degree of v in $\widehat{T}$ must be at least three unless v is a leaf and the same for u. Consequently, by Lemma 17 all the nodes in the path represent non-degenerate random variables and one can prove the statement modifying proof of Proposition 18 in [18]. For example if both u and v are inner nodes of T then they 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 T, v separates k and l and $\{u,v\}$ separates $\{i,j\}$ from $\{k,l\}$ like on the graph below.



Moreover by the construction we can require that $\hat{\kappa}_{ij}, \hat{\kappa}_{kl}, \hat{\kappa}_{ik}, \hat{\kappa}_{jl}$ are all non-zero. One can show using [18, Equation (8)] that no matter where the root is we have $\kappa_{ij} = (1-s_u^2)\eta_{u,i}\eta_{u,j}$, $\kappa_{kl} = (1-s_v^2)\eta_{v,k}\eta_{v,l}$, $\kappa_{ik} = (1-s_u^2)\eta_{u,i}\eta_{u,v}\eta_{v,k}$ and $\kappa_{jl} = (1-s_u^2)\eta_{u,j}\eta_{u,v}\eta_{v,l}$. This implies that in $\widehat{\Omega}_T$ one has

$$\frac{\hat{\kappa}_{ik}\hat{\kappa}_{jl}}{\hat{\kappa}_{ij}\hat{\kappa}_{kl}} = \frac{1-s_u^2}{1-s_v^2}\eta_{u,v}^2(\omega) = \rho_{uv}^2(\omega),$$

where ρ_{uv} is the correlation between H_u, H_v. In particular ρ_{uv}^2 is constant on $\widehat{\Omega}_T$ and its value is identified by $\hat{\mu}$. Since both u and v have degree at least three in $\widehat{T}$ then by the proof of [18, Proposition 18] the values of s_u^2 and s_v^2 are fixed and can be identified from $\hat{p}$ and hence since $\rho_{uv}^2(\omega)$ is constant the same applies to $\kappa_{uv}^2(\omega)$. If either u or v is a leaf of T the argument is very similar. $\square$

The following proposition shows that the geometry of the ML fiber $\widehat{\Omega}_T$ is determined by zeros in the sample covariance matrix $\hat{\Sigma}$.

Proposition 19 (The geometry of the ML fiber - the smooth case)**.** *Let $\hat{p} \in \mathcal{M}_T$ and let $\widehat{T}$ be defined as above. If each of the inner nodes of T has degree at least three in $\widehat{T}$ then the ML fiber is a finite set of points. If each of the inner nodes of T has degree at least two in $\widehat{T}$ then all the points of the ML fiber are smooth.*

Proof. If each inner node of T has degree at least three in $\widehat{T}$ then for each h one can find $i,j,k \in [n]$ separated from each other by h such that $\hat{\kappa}_{ij}\hat{\kappa}_{ik}\hat{\kappa}_{jk} \neq 0$ and then by the proof of Proposition 18 in [18] we can identify all the parameters of the model up to the choice of labels of the inner nodes.

For the second statement: Consider any equivalence class in $[\widehat{E}]$. By Lemma 18 (ii) the equivalence classes are just single edges and for each such an edge e one can find two leaves i,j such that the path between i and j crosses $\widehat{E}$ only through e. However one easily checks that if an edge e' satisfies $e' \in E \setminus \widehat{E}$ then necessarily $\eta_{e'} \neq 0$ in $\widehat{\Omega}_T$. Moreover, each inner node on the path between i and j is nondegenerate and therefore $\hat{\kappa}_{ij} = 0$ if and only if $\eta_e^0 = 0$ which means that the value of this parameter is fixed in $\widehat{\Omega}_T$. Moreover, for any inner node h of T which has degree at least three in $\widehat{T}$ and for every leaf $i \in [n]$ one can identify s_h and s_i and their values are fixed in $\widehat{\Omega}_T$ again following the lines of the proof of Proposition 18 in [18]. By Lemma 18 (iii) all the equivalence classes in $[E \setminus \widehat{E}]$ are paths and for any such path $\mathcal{P}_T(k,l)$ one can identify $\hat{\kappa}_{kl} \neq 0$ up to the sign. This gives us an equation

$$\frac{1}{4}(1 - s_{r(kl)}^2) \prod_{e \in \mathcal{P}_T(k,l)} \eta_e = \hat{\kappa}_{kl} \neq 0$$

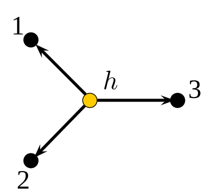
defining a regular subset in a subspace of Ω_T with coordinates given by $s_{r(kl)}$ and η_e for all edges e of $\mathcal{P}_T(k,l)$. If $r(kl)$ is a leaf or $\deg(r(kl)) \geq 3$ in $\widehat{T}$ then $s_{r(kl)}$ has a fixed value on $\widehat{\Omega}_T$ by the preceding statement.

It remains to show that there are no other constraints on η_e for $e \in E(uv)$. Note however that if for some $I \subseteq [n]$ such that $T(I)$ has no vertices in $\widehat{E}$ then $T(I)$ either contains all the edges of $E(uv)$ or non of them. In the second case the values of η_e do not matter since $\kappa_I = 0$ anyway. In the first case the values of η_e matter only through the value of κ_{uv}.

All these algebraic constraints involve different parameters and hence $\widehat{\Omega}_T$ is given as a subset of Ω_T which is a product of the regular subsets given above. In particular $\widehat{\Omega}_T$ is a regular subset of Ω_T. □

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

Example 20. Let $T = (V, E)$ be the tripod tree rooted in the inner node.



The degree of h in $\widehat{T}$ is less than two if and only if $\hat{\kappa}_{ij} = 0$ for all $i \neq j = 1,2,3$. In this situation $\widehat{E} = E$ and the ML fiber $\widehat{\Omega}_T$ is given as a subset of Ω_T by three equations

$$(1 - s_h^2)\eta_{h1}\eta_{h2} = 0,\ (1 - s_h^2)\eta_{h1}\eta_{h3} = 0,\ (1 - s_h^2)\eta_{h2}\eta_{h3} = 0$$

and $s_i = 1 - 2\hat{\lambda}_i$ for $i = 1,2,3$, where $\hat{\lambda}_i$ denotes the sample mean. Geometrically this is a sum of two hyperplanes $s_h^2 = 1$ and three planes given by $\eta_{h1} = \eta_{h2} = 0$, $\eta_{h1} = \eta_{h3} = 0$ and $\eta_{h2} = \eta_{h3} = 0$ plus additional inequality constraints given by Equation (15). In particular it is not a regular set since it has self-intersection points.

The situation changes if T is rooted in one of the leaves, say in 1. In this case $\widehat{\Omega}_T$ is given as a sum of the three planes $\{\eta_{1h} = \eta_{h2} = 0\}$, $\{\eta_{1h} = \eta_{h3} = 0\}$, $\{\eta_{h2} = \eta_{h3} = 0\}$ plus the two planes given by $\{s_h^2 = 1, \eta_{1h} = 0\}$. Note however that in both situations (the two different rootings) the sets are equal as the subsets of Ω_T since by (15) $s_h^2 = 1$ implies $\eta_{1h} = 0$.

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 which are degenerate with respect to $\hat{p}$ is denoted by $\widehat{V}$. By

Lemma 17 the set $V \setminus \widehat{V}$ is the set of all the nodes representing random variables which cannot be degenerate given $\hat{p}$. We define the *deepest singularity* of $\widehat{\Omega}_T$ as

$$\widehat{\Omega}_{\text{deep}} := \{\omega \in \widehat{\Omega}_T : \eta_e = 0 \ \forall e \in \widehat{E},\ \ s_v^2 = 1 \ \forall v \in \widehat{V}\}. \tag{19}$$

A priori it is not obvious that $\widehat{\Omega}_{\text{deep}}$ is not an empty set. We check carefully that there are always points in Ω_T satisfying the constraints of $\widehat{\Omega}_{\text{deep}}$. Let $(u, v) \in E$, by Equation (15) there are three cases to consider. If u and v satisfy $s_u^2, s_v^2 \neq 1$ and $(u, v) \in \widehat{E}$ then $\eta_{uv} = 0$ satisfy the inequality (15) for any value $s_u, s_v \in (-1, 1)$. The case when $u \in \widehat{V}$ is trivial. Finally if $u \notin \widehat{V}$ and $v \in \widehat{V}$ then the inequality is satisfied if and only if $\eta_{uv} = 0$ but it still in Ω_T. Hence $\widehat{\Omega}_{\text{deep}}$ is non-empty. Note however that unless $\widehat{V}$ is empty $\widehat{\Omega}_{\text{deep}}$ will always lie on the boundary of Ω.

Proposition 21 (The geometry of the ML fiber - the singular case)**.** *If $\widehat{V}$ is non-empty then the ML fiber is a collection of smooth varieties. Their common intersection locus is given by $\widehat{\Omega}_{\text{deep}}$.*

Proof. First assume that all the inner node of T are in $\widehat{V}$. In particular for all $i, j \in [n]$ we have $\hat{\kappa}_{ij} = 0$. Let $A \times B \subseteq \widehat{V} \times \widehat{E}$. We say that $A \times B$ is *minimal for* $\widehat{\Sigma}$ if for every point ω in

$$\Omega_{A\times B} = \{\omega \in \Omega_T : s_v^2 = 1 \text{ for all } v \in A, \eta_e = 0 \text{ for all } e \in B\}$$

and for every $i, j \in [n]$ we have $\kappa_{ij}(\omega) = 0$ and $A \times B$ is minimal with such a property (with respect to inclusion). Note that the ML fiber in this case is given as the sum of all $\Omega_{A\times B}$ for all $A \times B$ minimal for $\widehat{\Sigma}$ constrained to Ω_T. In particular every $\Omega_{A\times B}$ is an affine subspace constrained to Ω_T so it is smooth.

The general case can be shown in a similar manner. We use the regular case to show that each of the components is a smooth variety. □

The next result allows us to restrict our analysis to the neighborhood of $\widehat{\Omega}_{\text{deep}}$.

Lemma 22. *Let $\omega_0 \in \widehat{\Omega}_{\text{deep}}$ then every open neighbourhood of ω_0 contains all the irreducible components of $\widehat{\Omega}_T$ and their intersection. Therefore we have*

$$\min_{\omega_0 \in \Omega_T} \text{RLCT}_{\theta_0}(f) = \text{RLCT}_{\widehat{\Omega}_{\text{deep}}}(f).$$

Proof. The first statement follows from the proof of the proposition above. Let $\mathcal{V}_1, \ldots, \mathcal{V}_r$ be all the irreducible components of $\widehat{\Omega}_T$ from the proposition above. To prove the second part note first that from the first part we can construct the intersection lattice whose elements are subsets $I \subseteq [r]$ such that $\mathcal{V}_I = \bigcap_{i\in I} \mathcal{V}_i$. In particular $[r]$ as an element of this lattice corresponds to $\widehat{\Omega}_{\text{deep}}$ and each $i \in [r]$ corresponds to $\mathcal{V}_i$. Define $\mathcal{U}_I = \mathcal{V}_I \setminus \bigcup_{J \supsetneq I} \mathcal{V}_J$. Since each $\mathcal{V}_I$ corresponds to vanishing some of the variables then the function $\omega \mapsto \text{rlct}_\omega(f)$ is constant on each of $\mathcal{U}_I$.

By [6, Exercise 9.3.17] the function $\omega \mapsto \text{rlct}_\omega(f)$ is lower semicontinuous (the argument used there works over the real numbers). Since the set of values of the real log-canonical threshold is discrete it means that for $\omega_0 \in \widehat{\Omega}_T$ and any sufficiently small neighbourhood W_0 of ω_0 one has $\text{rlct}_{\omega_0}(f) \leq \text{rlct}_\omega(f)$ for all $\omega \in W_0$. Since for any neighborhood W_0 of $\omega_0 \in \widehat{\Omega}_{\text{deep}}$ we have $W_0 \cap \mathcal{U}_I \neq \emptyset$ for all $I \subseteq [r]$ then necessarily the minimum of the real log-canonical threshold is attained for a point from the deepest singularity. □

5. Asymptotics for phylogenetic tree models

Before we state the main theorem of the paper we provide a direct corollary of Proposition 19. The following useful result generalizes Proposition 3.3 in [11].

Lemma 23. *Let $\Theta \subset \mathbb{R}^d$ and $f \in \mathcal{A}^+_\Theta$. If $f^{-1}(0)$ is a smooth subset of an open set $U \supset \Theta$ and $x_0 \in f^{-1}(0)$ then $\mathrm{RLCT}_{x_0}(f) = \mathrm{RLCT}_\Theta(f) = (c/2, 1)$ where $c = \operatorname{codim}(f^{-1}(0))$.*

Proof. Recall that the real log-canonical threshold does not depend on the choice of a neighbourhood of x_0. Since $f^{-1}(0)$ is smooth then there exists an open neighbourhood of x_0 with local equations $u_1, \dots, u_d$ such that the local equation of $f^{-1}(0)$ is $u_1^2 + \cdots + u_c^2 = 0$. A single blow-up π at the origin does the job since in the new coordinates $f(\pi(\tilde{x})) = \tilde{x}_1^2 u(\tilde{x})$ where $u(\tilde{x})$ is nowhere vanishing and $\pi'(\tilde{x}) = \tilde{x}_1^{c-1}$. Hence by [7, Proposition 2.1] $\mathrm{RLCT}_{x_0}(f) = (c/2, 1)$. Since by (6) $\mathrm{RLCT}_\Theta(f) = \min_{x_0 \in \Theta} \mathrm{RLCT}_{W_0 \cap \Theta}(f)$ then it suffices to show that if x_0 is a boundary point of Θ then $\mathrm{RLCT}_{W_0 \cap \Theta} \geq (c/2, 1)$ but this follows from (5) and the fact that $\mathrm{RLCT}_{x_0}(f) = (c/2, 1)$ as θ_0 is a smooth point of $f^{-1}(0)$ in U. □

The result gives us a way to compute the asymptotic approximation for the marginal likelihood in a non-singular case.

Theorem 24. *Let $\hat{p} \in \mathcal{M}_T$ be such that each inner node of T has degree at least two in $\widehat{T}$. Then*

$$Z(N) = \hat{\ell}_N - \frac{1}{2}(n_v + n_e - 2l_2)\log(N) + O(1),$$

where l_2 is the number of the inner nodes of T with degree two in $\widehat{T}$.

Proof. Since every inner node of T has degree at least two in $\widehat{T}$ then by Proposition 19 the ML fiber $\widehat{\Omega}_T$ is a smooth subset of Ω_T and one can see that the constraints defining it give a smooth subset in any open subset in $\mathbb{R}^{n_v+n_e}$ containing Ω. Hence by Lemma 23 it suffices to show that $\operatorname{codim}(\widehat{\Omega}_T) = n_v + n_e - 2l_2$. This can be seen directly from the proof of Proposition 19 by listing all the constraints on $\widehat{\Omega}_T$.

First, for all $e \in \widehat{E}$ we have $\eta_e = 0$ in $\widehat{\Omega}_T$. If v is either a leaf of T or its degree in $\widehat{T}$ is at least three then s_v is fixed in $\widehat{\Omega}_T$ (this gives $n_v - l_2$ further constraints). Moreover, by Lemma 18 (iii) for all paths $\mathcal{P}_T(u, v)$ in $[E \setminus \widehat{E}]$ the value of $\kappa_{uv}(\omega)$ is nonzero and fixed on $\widehat{\Omega}_T$ where

$$\kappa_{uv}(\omega) = \frac{1}{4}(1 - s_{r(uv)}^2) \prod_{e \in E(uv)} \eta_e,$$

which gives an equation involving η_e for e in the path between u and v and $s_{r(uv)}$ unless $r(uv) \in \{u, v\}$ since in this case the value of $s_{r(uv)}$ is fixed on $\widehat{\Omega}_T$ ($r(uv)$ has either the degree at least three in $\widehat{T}$ or it is a leaf). Consequently, for each path $\mathcal{P}_T(u, v) \subset E$ one has $e_{uv} - l_{uv} = 1$ where e_{uv} is the number of edges and l_{uv} is the number of degree two nodes in $\mathcal{P}_T(u, v)$. Since all the paths have disjoint sets of edges and their terminal nodes are either leaves or nodes of degree greater or equal to three in $\widehat{T}$ then the number of paths is equal to $\sum 1 = \sum(e_{uv} - l_{uv}) = |E \setminus \widehat{E}| - l_2$ where the sum is over all the paths in $[E \setminus \widehat{E}]$. Since there are no other constraints on $\widehat{\Omega}_T$ it follows that the codimension of $\widehat{\Omega}_T$ is $|\widehat{E}| + (n_v - l_2) + (|E \setminus \widehat{E}| - l_2) = n_v + n_e - 2l_2$. □

It is a known fact that without lost of generality we can assume that all the inner nodes of T have degree at least three. It follows from the fact that $p \in \mathcal{M}_T$ if and only if $p \in \mathcal{M}_{T'}$ where T' is a tree with degree two nodes and T is obtained from T' by contracting all these nodes, i.e. replacing each pair of edges (u, v), (v, w) where $\deg v = 2$ by an edge (u, w) and removing the node v. Moreover, for every tree T the model $\mathcal{M}_T$ can be realized as a submodel of $\mathcal{M}_{\overline{T}}$, for some trivalent tree $\overline{T}$, with some simple constraints on the parameter space. For example the 4-star model with four leaves and one inner node is a submodel of the quartet tree model (the trivalent tree with four leaves). To see this write the parametrization for the quartet tree model (c.f. Proposition 14) denoting the inner nodes by a, b and set $\eta_{a,b} = 1$ and $s_a = s_b$.

For these reasons the trivalent case seems the most interesting and since it is also easier to analyze for the singular case we consider only this situation. In the remaining part of the paper we prove the following theorem which is our main result.

Theorem 25. *Let $T = (V, E)$ be a trivalent tree with $n \geq 3$ leaves. Let $\mathcal{M}_T \subseteq \Delta_{2^n-1}$ be the general Markov model on T parametrized as in Equation (9). Let $I(N)$ be defined by (3) for $\hat{p} \in \mathcal{M}_T$ and φ which is bounded and strictly positive. If the root is degenerate but all its neighbors are not then as $N \to \infty$*

$$Z(N) = \hat{\ell}_N - \frac{1}{4}(3n + l_2 + 5l_3 - 1)\log N + O(1). \tag{20}$$

where l_i is the number of nodes in T of degree i in $\widehat{T}$ for $i = 2, 3$. If either the root is non-degenerate or it is degenerate and all its neighbours are degenerate as well then as $N \to \infty$

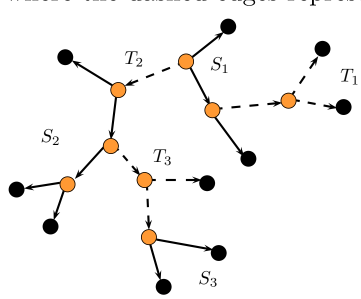
$$Z(N) = \hat{\ell}_N - \frac{1}{4}(3n + l_2 + 5l_3)\log N + O(1). \tag{21}$$

In all other cases

$$Z(N) = \hat{\ell}_N - \frac{1}{4}(3n + l_2 + 5l_3)\log N + O(\log\log N). \tag{22}$$

The proof of this result is provided in the end of the paper.

First we show that we can divide the problem into simpler subproblems. Note that in this part we do not have to assume that the tree is trivalent. By $T_1, \ldots, T_k$ denote trees representing the equivalence classes in $[\widehat{E}]$ and by $S_1, \ldots, S_m$ denote trees induced by the connected components of $E \setminus \widehat{E}$. By $L_1, \ldots L_k$ we denote the sets of leaves of $T_1, \ldots, T_k$. For each S_i $i = 1, \ldots, m$ by Lemma 18 (iv) its set of leaves denoted by $[n_i]$ is a subset of $[n]$. To have a concrete example we may consider the graph below where the dashed edges represent edges in $\widehat{E}$.



The following technical lemma allows us to restrict the analysis to each of the T_i for $i = 1, \ldots, k$ and S_i for $i = 1, \ldots, m$ separately.

Lemma 26. *Let $T = (V, E)$ be a rooted tree with n leaves and let $\hat{p} \in \mathcal{M}_T$. Let $\omega_0 \in \widehat{\Omega}_{\text{deep}}$ then without a loss of generality assume that $s_v = 1$ for all $v \in \widehat{V}$ and denote $t_v = 1 - s_v$. Let $J_{\omega_0} \subset \mathcal{A}_{\omega_0}$ as defined in the end of Section 3. Then*

$$J_{\omega_0} = \bigcup_{i=1}^{m} I_i + \bigcup_{i=1}^{k} J_i,$$

where $I_i = \bigcup_{I \in [n_i]_{\geq 2}} \langle \kappa_I(\omega) - \hat{\kappa}_I \rangle$ for $i \in [m]$ and $J_i = \bigcup_{k,l \in L_i} \left\langle t_{r(kl)} \prod_{e \in E(kl)} \eta_e \right\rangle$ if T_i is rooted in an inner node; and

$$J_i = \bigcup_{l \in L_i} \left\langle \prod_{e \in E(rl)} \eta_e \right\rangle + \bigcup_{k,l \in L_i \setminus r} \left\langle t_{r(kl)} \prod_{e \in E(rl)} \eta_e \right\rangle$$

if it is rooted in one of the leaves denoted by r.

Proof. First we show that if $i, j \in [n]$ such that $\kappa_{ij}(\omega_0) = 0$ then $\kappa_I(\omega_0) = 0$ for all I such that $i, j \in I$. Indeed, since $\omega_0 \in \widehat{\Omega}_{\text{deep}}$ then $\kappa_{ij} = 0$ if and only if the path between i and j contains an edge in $\widehat{E}$. But this holds if and only if $\kappa_I = 0$ for all I such that $i, j \in I$. From this it follows that $\langle \kappa_I(\omega) \rangle \subseteq \langle \kappa_{ij}(\omega) \rangle$ in $\mathcal{A}_{\omega_0}$ whenever $i, j \in I$ and consequently

$$\bigcup_{I \in [n]_{\geq 2}} \langle \kappa_I(\omega) - \hat{\kappa}_I \rangle = \left(\bigcup_{i=1}^{m} I_i \right) + \left(\bigcup_{k,l : \hat{\kappa}_{kl} = 0} \langle \kappa_{kl}(\omega) \rangle \right).$$

From Proposition 14 we have

$$\kappa_{kl}(\omega) = \frac{1}{4}(1 - s_{r(kl)}^2) \prod_{e \notin \widehat{E}_{kl}} \eta_e \prod_{e \in \widehat{E}_{kl}} \eta_e. \tag{23}$$

If $r(kl) \notin \widehat{V}$ then $(1 - s_{r(kl)}^2) \prod_{e \notin \widehat{E}_{kl}} \eta_e$ is a unit in $\mathcal{A}_{\omega_0}$ and hence from (23) $\langle \kappa_{kl}(\omega) \rangle = \langle \prod_{e \in \widehat{E}_{kl}} \eta_e \rangle$. Similarly if $r(kl) \in \widehat{V}$ then by our assumption $s^0_{r(kl)} = 1$ and hence $1 + s_{r(kl)}$ is a unit in $\mathcal{A}_{\omega_0}$ and then $\langle \kappa_{kl}(\omega) \rangle = \langle t_{r(kl)} \prod_{e \in \widehat{E}_{kl}} \eta_e \rangle$. Finally one can easily check that in both cases $\langle \kappa_{kl}(\omega) \rangle \subseteq \bigcup J_i$. The opposite inclusion is left as an exercise. □

Note that all $m + k$ parts of J_{ω_0} in Lemma 26 depend on different parameters and by Proposition 4.6 in [7] one can compute the real log-canonical threshold for each of the parts separately. The computation for the first m parts is simple since for each S_i for $i \in [m]$ we can use Theorem 24. In the next section we focus on computation the real log-canonical threshold of J_i for $i = 1, \ldots, k$.

6. The case of trivalent trees

In what follows we continue the analysis from the previous section restricting ourselves exclusively to trivalent trees. In this case we can obtain direct formulas for the BIC score. Note that if T is trivalent with n leaves then the number of inner nodes is $n - 2$ and the number of the edges is $2n - 3$. Moreover, by Lemma 18 (i) all the connected components of $[\widehat{E}]$ are either single edges or trivalent trees. One can easily check that for $i = 1, \dots, k$ if T_i is a single edge e and then $\mathrm{rlct}(J_i) = 1/2$, where $J_i = \langle \eta_e \rangle$. Hence in what follows we assume that $|L_i| > 2$.

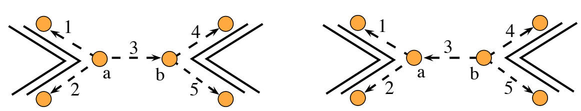
We fix $1 \le i \le k$ and analyse contribution to the marginal likelihood related to T_i. Since for each $i = 1, \dots, k$ the ideal J_i defined in Lemma 26 is monomial then by [7, Corollary 5.11] it is $\mathbb{R}$-nondegenerate and hence by Theorem 10 we can use the Newton diagram method to compute the real log-canonical threshold. Let $\omega_0 \in \widehat{\Omega}_{\mathrm{deep}}$. Without loss of generality we assume again that $s_h = 1$ for all inner nodes h and we change the variables $t_h \mapsto 1 - s_h$. By $Q_i(\omega)$ we denote the generator of $J_i \subset \mathcal{A}_{\omega_0}$ given as a sum of squares of the generating monomials as in Lemma 26. From the proof of Proposition 21 it follows that the variety of J_i is just a sum of affine subspaces intersecting in $\widehat{\Omega}_{\mathrm{deep}}$.

Let $\mathbb{R}^{c+d}$ be a real space with variables representing the edges $(x_e)_{e \in E_i}$ and the inner nodes $(y_v)_{v \in N_i}$ of $T_i = (V_i, E_i)$ where $N_i = V_i \setminus L_i$, $d = |E_i|$ and $c = |N_i|$. We order the variables as follows: $y_1 \prec \cdots \prec y_c \prec x_1 \prec \cdots \prec x_d$. The exponents of terms of the polynomial $Q_i(\omega)$ are vectors in $\{0, 2\}^{c+d}$.

Now for $\omega_0 \in \widehat{\Omega}_{\mathrm{deep}}$ we compute $\mathrm{RLCT}_{\omega_0}(Q_i)$ using the method of Newton diagrams. This part involves a considerable amount of polyhedral geometry. To simplify the notation we fix $i = 1, \dots, k$ such that T_i is not a single edge and denote T_i by T and Q_i by Q. Let $c = |L_i| - 2$ be the number of the inner nodes of T and let $d = 2|L_i| - 3$ be the number of edges of T. The construction of the Newton polytope $\Gamma(Q) \subset \mathbb{R}^{c+d}$ gives a direct relationship between paths in T and points generating the polytope. Combinations of paths give rise to points in the polytope. The following construction is particularly useful.

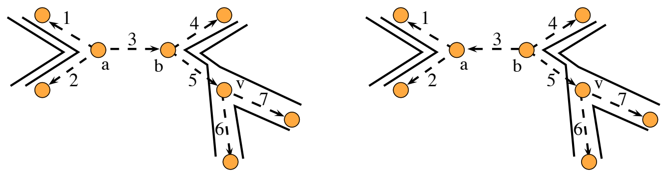
Construction 27. Let $T = (V, E)$ be a trivalent rooted tree with $n \ge 4$ leaves. We present two constructions of networks of paths between the leaves of T.

The first construction is conducted for T rooted in an inner node. If $n = 4$ then the network consists of the two paths within cherries counted with multiplicity two.



Each of the paths correspond to a point in $\Gamma(Q)$. We order the coordinates of $\mathbb{R}^7$ by $y_a \prec y_b \prec x_1 \prec \cdots \prec x_5$. For example the point corresponding to the path involving edges e_1 and e_2 is $(2, 0; 2, 2, 0, 0, 0)$ and this does not depend on whether T is rooted in a or b. The barycenter of the points corresponding to all the four paths in the network is $(1, 1; 1, 1, 0, 1, 1)$ both if T is rooted in a or b.

If $n > 4$ then we build the network recursively. Assume that T is rooted in an inner node a and pick an inner edge (a, b). Label the edges incident with a and b as for the quartet tree above and consider the subtree given by the quartet tree. Draw four paths as on the picture above. We build up a network recursively. Let v be any leaf of the quartet subtree which is not a leaf of T and label the two additional edges incident with v by e_6 and e_7. Then we extend the network by adding e_6 to one of the paths terminating in v and e_7 to the other. Moreover we add an additional path involving only e_6 and e_7 like on the picture below. By construction v is the root of the additional path. We extend the network cherry by cherry until it covers all terminal edges.



Note that we have made some choices building up the network the construction is not unique. However, each of the inner nodes is always a root of at least one and at most two paths. Moreover each edge is covered at most twice and each terminating edge is covered exactly two times. We have n paths in the network, all representing points of $\Gamma(Q)$ denoted by $p_1, \ldots, p_n$. Let $p = \frac{1}{n}\sum_{i=1}^{n} p_i$ then $p \in \Gamma(Q)$ is given by $x_{ab} = 0$, $x_e = 4/n$ for all $e \in E \setminus (a,b)$, $y_a = y_b = 4/n$ and $y_v = 2/n$ for all $v \in N \setminus \{a,b\}$. Hence p depends on the construction above only through the choice of the omitted edge not covered by any path.

If T is rooted in a leaf then we proceed as follows. For $n = 4$ consider a network of all the possible paths all counted with multiplicity one apart from the cherry paths (paths of length two) counted with multiplicity two. It makes eight paths and each edge is covered exactly four times. With the order of the coordinates as above the coordinates of the point representing the barrycenter of all paths in the network are $(1/2, 1/2; 1, 1, 1, 1, 1)$. This construction generalizes recursively in a similar way as the one for T rooted in an inner node. We always have $2n$ paths and each edge is covered exactly four times. The network induces a point $p \in \Gamma(Q)$ such that $y_h = 2/n$ for all $h \in N$ and $x_e = 4/n$ for $e \in E$.

Let E_0 denote the set of terminal edges of T, i.e. the edges incident with a leaf. The hyperspace given by $\sum_{e \in E_0} x_e = 4$ contains $\Gamma(Q)$ which follows from the fact that each path in T necessarily crosses exactly two terminal edges and hence each point generating $\Gamma(Q)$ satisfies the equation. Hence $\sum_{e \in E_0} x_e \geq 4$ defines a facet of $\Gamma_+(Q)$ both if T is rooted in a leaf or in an inner node. We denote this facet by F_0.

Lemma 28 (Computing the real log-canonical threshold)**.** *Let T be a trivalent tree with $n \geq 4$ leaves. Then the vector $t(1, \ldots, 1)$ hits F_0 for $t = 4/n$ and it is the smallest number such that $t\mathbf{1} \in \Gamma_+(Q)$.*

Proof. The fact that $4/n\,\mathbf{1} \in \Gamma_+(Q)$ follows from Lemma 27 and the fact that the constructed point $p \in \Gamma(Q)$ satisfies $p \leq \frac{4}{n}\mathbf{1}$ both if T is rooted in an inner node or in a leaf and hence $4/n\mathbf{1} \in p + \mathbb{R}^{c+d}_{\geq 0}$. The result then follows since for any $s < t$ the vector $s(1, \ldots, 1)$ does not satisfy $\sum_{e \in E_0} x_e \geq 4$ and hence it cannot be in $\Gamma_+(Q)$. $\square$

Since $\omega_0 \in \widehat{\Omega}_{\text{deep}}$ then $\text{rlct}_{\omega_0}(Q) = \text{rlct}_{\Omega}(Q)$. By [7, Corollary 5.11] this gives $\text{rlct}_{\omega_0}(Q) = n/4$. Note that $\widehat{\Omega}_{\text{deep}}$ lies on the boundary of Ω. However it is not a problem since any point in $\widehat{\Omega}_T$ such that $\eta_e = 0$ for all $e \in \widehat{E}$ (and we can find such points in the interior of Ω) gives the same value of rlct. To compute the multiplicity we have to get a better understanding of the polyhedron $\Gamma_+(Q)$. First we find the hyperplane representation of the Newton polytope $\Gamma(Q)$ reducing the problem to a simpler but equivalent one.

Definition 29 (A pair-edge incidence polytope)**.** Let $T = (V, E)$ be a trivalent tree with $n \geq 4$ leaves. We define a polytope $P_n \subset \mathbb{R}^d$, where $d = n_e = 2n - 3$, as a convex combination of points $(p_{ij})_{\{i,j\}\in[n]_2}$ where k-th coordinate of p_{ij} is one if the k-th edge is in the path between i and j and there is zero otherwise. We call P_n a pair-edge incidence polytope by analogy to the pair-edge incidence matrix defined by Mihaescu and Pachter [8, Definition 1].

The reason to study this polytope is that its structure can be handled easily and it can be shown to be affinely equivalent to $\Gamma(Q)$. To see this fix a rooting r of T and define a linear map $f_r : \mathbb{R}^d \to \mathbb{R}^c$ as follows: for each $v \in V \setminus r$ set $y_v = 1/2(x_{v\mathrm{ch}_1(v)} + x_{v\mathrm{ch}_2(v)} - x_{\mathrm{pa}(v)v})$, where $\mathrm{ch}_1(v)$, $\mathrm{ch}_2(v)$ denotes the two children of v, and $y_r = 1/2(x_{r\mathrm{ch}_1(r)} + x_{r\mathrm{ch}_2(r)} + x_{r\mathrm{ch}_3(r)})$. Then one can check that for a map $(\mathrm{id} \times f_r) : \mathbb{R}^d \to \mathbb{R}^d \times \mathbb{R}^c$ one has $(\mathrm{id} \times f_r)(2P_n) = \Gamma(Q)$ which follows from the fact that for each point $y_r = 2$ if and only if the path crosses r and for any other node $y_v = 2$ if and only if the path crosses v and v is the root of the path, i.e. if the path crosses both children of v.

Lemma 30. *Let $P_n \subset \mathbb{R}^d$ be a pair-edge incidence polytope for a trivalent tree with n leaves where $n \geq 4$. Then $\dim(P_n) = d - 1 = 2n - 4$. Hence the codimension of P_n is one and the affine subspace defining P_n is given by $\sum_{e\in E_0} x_e = 2$. For each inner node $v \in V$ let $e_1(v)$, $e_2(v)$, $e_3(v)$ denote the three adjacent edges. Then exactly $3(n-2)$ facets define P_n and they are given by*

$$x_{e_1(v)}+x_{e_2(v)}-x_{e_3(v)} \geq 0, x_{e_2(v)}+x_{e_3(v)}-x_{e_1(v)} \geq 0, x_{e_3(v)}+x_{e_1(v)}-x_{e_2(v)} \geq 0 \tag{24}$$

for all $v \in V$.

Proof. Let M_n be the pair-edge incidence matrix, i.e. a $\binom{n}{2} \times d$ matrix with rows corresponding to the points defining P_n. By Lemma 1 in [8] the matrix has full rank and hence P_n has codimension one in $\mathbb{R}^d$. Moreover since each path necessarily crosses two terminal edges then each point generating P_n satisfies the equation $\sum_{e\in E_0} x_e = 2$ and hence this is the equation defining the affine subspace containing P_n.

Now we show that the inequalities give a valid facet description for P_n. It is a direct check for $n = 4$ (e.g. using Polymake [2]). Assume it is true for all $k < n$. By Q_n we will define the polytope defined by the unique equation and $3(n-2)$ inequalities. It is obvious that $P_n \subseteq Q_n$ since all points generating P_n satisfy the equation and the inequalities. We will show the opposite inclusion.

Consider any cherry $\{e_1, e_2\} \subset E$ in the tree given by leaves denoted by 1, 2 and the separating inner node a. Define a projection $\pi : \mathbb{R}^d \to \mathbb{R}^{d-2}$ on the coordinates related to all the edges apart from the two in the cherry. We have $\pi(Q_n) = \widehat{Q}_{n-1}$, where $\widehat{P} = \mathrm{conv}\{0, P\}$. Indeed, $\pi(Q_n)$ is described by all the triples of inequalities for all the inner nodes apart from the one incident with the cherry and the defining equation becomes an inequality

$$\sum_{e\in E_0\setminus\{e_1,e_2\}} x_e \leq 2.$$

Moreover, inequalities in Equation (24) define a polyhedral cone and the equation $\sum_{e\in E_0\setminus\{e_1,e_2\}} x_e = t$ for $t \geq 0$ cuts out a bounded slice of the cone which is equal to $t \cdot P_{n-1}$. The sum of all these for $t \in [0, 2]$ is exactly $\widehat{Q}_{n-1}$. Since $\widehat{Q}_{n-1} = \widehat{P}_{n-1}$ by induction then each $\pi(x)$ is a convex combination of the points generating P_{n-1} and zero, i.e. $\pi(x) = \sum c_{ij} p_{ij}$ where the sum is over all $i \neq j \in \{a, 3, \ldots, n\}$ and

$\sum c_{ij} \leq 1$. Next we lift this combination to Q_n. We have

$$x = \sum c_{ij}\pi^{-1}(p_{ij}) + \left(1 - \sum c_{ij}\right) \pi^{-1}(0)$$

and we want to show that $Q_n \subseteq P_n$. But to show this it suffices to show that any lift of $p_{ij} \in P_{n-1}$ and zero to Q_n will necessary imply that $x \in P_n$

Denote the edge incident with e_1, e_2 by e_3 and the related coordinates of x by x_1, x_2, x_3. Consider the following three cases. First, if $p_{ij} \in P_{n-1}$ is such that $x_3 = 0$ then sum of all the other coordinates related to the terminal edges is two since $P_{n-1} = Q_{n-1}$ and Q_{n-1} satisfy the equation $\sum_{e\in E_0\setminus\{e_1,e_2\}} x_e = 2$. Hence if we lift p_{ij} to Q_n then $x_3 = 0$ and

$$x_1 + x_2 \geq 0, \quad x_1 - x_2 \geq 0, \quad x_2 - x_1 \geq 0$$

by plugging $x_3 = 0$ into the three inequalities for the node a. But since since $\pi^{-1}(p_{ij})$ must also satisfy the equation $\sum_{e\in E_0} x_e = 2$ and since we already have $\sum_{e\in E_0\setminus\{e_1,e_2\}} x_e = 2$ then $x_1 + x_2 = 0$ and hence $x_1 = x_2 = 0$. Consequently, p_{ij} lifts to a vertex of P_n. Second, if p_{ij} is a vertex of P_{n-1} such that $x_3 = 1$ then the sum of all the other coordinates of p_{ij} related to the terminal edges is one and hence since the lift is in Q_n we have $x_1 + x_2 = 1$. The additional inequalities give that $x_1, x_2 \geq 0$. Hence in this case p_{ij} lifts to a convex combination of two points in P_n - one corresponding to a path finishing in one of the edges and the other in the other. Finally, one can easily check that zero lifts uniquely to a point in P_n corresponding to the path $\mathcal{P}_T(1,2)$. Indeed, from the equation defining Q_n we have $x_1 + x_2 = 2$ and from the inequalities since $x_3 = 0$ we have $x_1 = x_2 = 1$. Consequently, x can be written as a convex combination of points generating P_n and hence $x \in P_n$. Consequently $Q_n \subseteq P_n$. $\square$

The lemma shows that P_n has an extremely simple structure. The inequalities give a polyhedral cone and the equation cuts out its slice of height two. The result gives us also the H-representation of $\Gamma(Q)$.

Proposition 31 (Structure of $\Gamma(Q)$)**.** *Polytope $\Gamma(Q) \subset \mathbb{R}^{c+d}$ is given as an intersection of the sets defined by the facets in Equation (24) plus $n-1$ equations and if T is rooted in one of the leaves then*

$$\sum_{e\in E_0} x_e = 4 \qquad \text{and} \tag{25}$$
$$2y_v = x_{v\mathrm{ch}_1(v)} + x_{v\mathrm{ch}_2(v)} - x_{\mathrm{pa}(v)v} \quad \text{for all } v \in N.$$

If T is rooted in one of the inner nodes then we have

$$\sum_{e\in E_0} x_e = 4 \qquad \text{and} \tag{26}$$
$$2y_r = x_{r\mathrm{ch}_1(r)} + x_{r\mathrm{ch}_2(r)} + x_{r\mathrm{ch}_3(r)} \qquad \text{and}$$
$$2y_v = x_{v\mathrm{ch}_1(v)} + x_{v\mathrm{ch}_2(v)} - x_{\mathrm{pa}(v)v} \quad \text{for all } v \in N.$$

From this we can partially understand the structure of $\Gamma_+(Q)$. First we state the following basic fact.

Lemma 32. *Let $\Gamma \subset \mathbb{R}^n_{\geq 0}$ be a polytope and let Γ_+ be the Minkowski sum of Γ and the standard cone $\mathbb{R}^n_{\geq 0}$. Then all the facets of Γ_+ are of the form $\sum_i a_i x_i \geq c$ where $a_i \geq 0$ and $c \geq 0$.*

Proof. It suffices to show that all faces of Γ with the supporting hyperplane given by $\sum_i a_i x_i \geq c$ where $c \geq 0$ and some of $a_i < 0$ have a nonempty intersection with the interior of Γ_+. Let F be a face like that and let after renumbering $a_1, \ldots, a_k > 0$, $a_{k+1}, \ldots, a_l < 0$ and $a_{l+1} = \ldots = a_n = 0$. Let $q = (q_1, \ldots, q_n)$ be a point in the interior of F. Then we can decrease coordinates $x_1, \ldots, x_k$ by $\epsilon/a_1, \ldots, \epsilon/a_k$ and coordinates $x_{k+1}, \ldots, x_l$ by $-\epsilon/a_{k+1}, \ldots, -\epsilon/a_l$ respectively. Moreover, we decrease $x_{l+1}, \ldots, x_n$ by some small δ. If $\epsilon > 0$ and $\delta > 0$ are sufficiently small then we still stay in the F reaching a point q'. Since the cone $q' + \mathbb{R}^n_{\geq 0}$ is contained in $\Gamma + \mathbb{R}^n_{\geq 0}$ and q lies in its interior then the lemma is proved. □

Now we are ready to compute multiplicities of the real log-canonical threshold $\mathrm{RLCT}_{\omega_0}(Q)$ which by [7, Corollary 5.11] are given as the codimension of the face hit by the vector of ones (codimension of the largest face containing the point $4/n\mathbf{1}$). Note that since $\widehat{\Omega}_{\text{deep}}$ lies on the boundary of Ω then by Corollary 11 we get only inequalities.

Lemma 33 (Computing multiplicities)**.** *Let T be a trivalent tree with $n \geq 4$ leaves and rooted in r. If either r is one of the leaves or r is an inner node with no adjacent leaves then $w = 1$. Otherwise $w \geq 1$.*

Proof. The proof is moved to the appendix. □

Now we return to the general case of a tree T for which $\widehat{T}$ decomposes into T_i and S_i. We are ready to proof the main theorem of the paper.

Proof of Theorem 25. The case when $n = 3$ and T is rooted in the inner node follows from the main theorem in [10]. The case when $n = 3$ and T is rooted in a leaf is left as an exercise. Assume that $n \geq 4$. To compute the asymptotic approximation for (3) we need to compute $\mathrm{RLCT}_\Theta(f;\varphi)$. By Lemma 6 this is equivalent to computing $\mathrm{RLCT}_\Theta(f)$. We can constrain to the deepest singularity and then by Lemma 4 instead of f we can take the ideal in Lemma 26. Since its various components depend on different parameters we can analyze the bits for T_i and S_i separately. Let $\omega_0 \in \widehat{\Omega}_{\text{deep}}$. Since $s_i(\omega + \omega_0) - s_i(\omega_0) = s_i$ then we have

$$\mathrm{RLCT}_{\omega_0}\left(\sum_{i=1}^n (s_i(\omega+\omega_0) - s_i(\omega_0))^2\right) = \left(\frac{n}{2}, 1\right). \tag{27}$$

Let I_i for each $i = 1, \ldots, m$ de defined as in Lemma 26. By the regular case resolved in Theorem 24, for each S_i we have

$$\mathrm{RLCT}_{\omega_0}(I_i) = \left(\frac{n_v^i + n_e^i - 2l_2^i - n_i}{2}, 1\right).$$

We subtracted n_i in the above formula since it has been already counted in (27). Note that $n_v^i = n_i + l_2^i + l_3^i$ and $\sum_{i=1}^m n_e^i = |E \setminus \widehat{E}|$. Hence

$$\sum_{i=1}^m \left(n_v^i + n_e^i - 2l_2^i - n_i\right) = \sum_{i=1}^m \left(l_3^i - l_2^i + n_e^i\right) = l_3 - l_2 + |E \setminus \widehat{E}|.$$

By Lemma 7 one has

$$\mathrm{RLCT}_{\omega_0}\left(\sum_{i=1}^m \sum_{I \in [n_i]_{\geq 2}} (\kappa_I(\omega+\omega_0) - \kappa_I(\omega_0))^2\right) = \left(\frac{l_3 - l_2 + |E \setminus \widehat{E}|}{2}, 1\right). \tag{28}$$

Let r be the root of T and assume that $r \notin \widehat{V}$. In this case all T_i for $i = 1, \ldots, k$ are rooted in one of the leaves. Let Ω_0 be any sufficiently small ball around $\omega_0 \in \widehat{\Omega}_{\text{deep}}$. Since $\widehat{\Omega}_{\text{deep}}$ lies on the boundary Ω_T then by Lemma 28 together with Equation (5) we have $\text{rlct}_{\Omega_0 \cap \Omega_T}(Q_i) \geq \text{rlct}_{\omega_0}(Q_i) = (|L_i|/4, 1)$. However in this case we can easily provide a point on $\widehat{\Omega}_T$ in the interior of Ω with the same real log-canonical threshold as $\text{rlct}_{\omega_0}(Q_i)$ and hence in fact $\text{RLCT}_{\Omega_T}(Q_i) = (|L_i|/4, 1)$. This follows from the fact that both Construction 27 and Lemma 28 remain valid even if all for all the inner nodes $s_h^2 \neq 1$.

Note that since T is a trivalent tree then $\sum_{i=1}^{k} |L_i| = n - l_1 + l_2$ and hence

$$\text{RLCT}_{\omega_0}\left(\sum_{i=1}^{k} Q_i\right) = \left(\frac{n - l_1 + l_2}{4}, 1\right). \tag{29}$$

To write the real log-canonical thresholds in (27), (28) and (29) in a convenient form note that for any graph with the vertex set V and the edge set E one satisfies $\sum_{v \in V} \deg(v) = 2n_e$ (see e.g. Corollary 1.2.2 in [13]). In particular for $\widehat{T}$ we have $l_1 + 2l_2 + 3l_3 = 2|E \setminus \widehat{E}|$. Consequently the coefficient $3n - l_1 - l_2 + 2l_3 + 2|E \setminus \widehat{E}|$ obtained by summing the real log-canonical thresholds above can be rewritten as $3n + l_2 + 5l_3$. Hence, if $r \notin \widehat{V}$ then

$$Z(N) = \hat{\ell}_N - \frac{1}{4}(3n + l_2 + 5l_3) \log N + O(1).$$

Let $r \in \widehat{V}$ and let j be such that r is an inner node of T_j. The case when $r \in \widehat{V}$ and is slightly more complicated. There are two reasons for that. The first one is that the case for each T_i the case $|L_i| = 3$ leads to different real log-canonical thresholds depending on the position of the root. If $|L_i| = 3$ and T_j is rooted in one of the leaves then as above $\text{RLCT}_{\omega_0}(Q_i) = (3/4, 1)$. However, if T_i is rooted in the inner node (and hence $j = i$) then $\text{RLCT}_{\omega_0}(Q_i) = (1/2, 1)$ (c.f. [10]). Note that at most one of all the T_i can be rooted in the inner node. If all the neighbors of r represent non-degenerate random variables then r is the inner node of T_j and $|L_j| = 3$. In this case the real log-canonical threshold has to be slightly modified but we essentially repeat the computations above obtaining

$$Z(N) = \hat{\ell}_N - \frac{1}{4}(3n + l_2 + 5l_3 - 1) \log N + O(1).$$

In all other cases we do not have to modify the first coefficient of the asymptotic approximation but we may need to change the second one according to Lemma 33. There exists one T_i such that r is its inner node. By Lemma 33 if all the neighbours of r are degenerate then $\text{mult}_\Theta(f) = 1$ and hence

$$Z(N) = \hat{\ell}_N - \frac{1}{4}(3n + l_2 + 5l_3) \log N + O(1).$$

Otherwise we have $\text{mult}_\Theta(f) \geq 1$ and

$$Z(N) = \hat{\ell}_N - \frac{1}{4}(3n + l_2 + 5l_3) \log N + O(\log \log N).$$

$\square$

7. Discussion

In this paper we obtained the asymptotic approximation for the marginal likelihood for the directed tree models when all the variables in the system are binary and all the inner nodes of the tree represent hidden variables. We provided nice combinatorial, algebraic and geometric insight into this problem.

The results were derived under some additional assumptions. The positivity assumption saying that $\hat{p}_\alpha > 0$ for all $\alpha \in \{0,1\}$ seems plausible for real data sets. The assumption that $\hat{p} \in \mathcal{M}_T$ seems restrictive but it was beyond the scope of this paper to relax it. The most controversial is the assumption that the prior density for the parameters is positive everywhere on the parameter space. In fact in the Bayesian analysis one usually uses some popular family of distributions which do not satisfy this assumption. We believe that careful analysis similar to the one presented here allows to understand more general situations.

One could argue that the condition for the sample covariances to vanish has measure zero with respect to the distribution of $X^{(N)}$ under the model $\mathcal{M}$ and therefore it can be neglected. However, as we have shown in [18] the models under consideration have usually complicated geometry and they have small dimension relative to the ambient space. Hence, although $p^{(N)}$ will usually not lie in $\mathcal{M}_T$ its hypothetical limit $\hat{p}$ (given the model is true) may lie on a locus of the model corresponding to some degeneracies with a positive probability. Example of this is given in the discussion in [3].

Appendix A. Proofs

Proof of Lemma 33. A standard result for Minkowski sums says that each face of a Minkowski sum of two polyhedra can be decomposed as a sum of two faces of the summands and this decomposition is unique. Each facet of $\Gamma_+(Q)$ is decomposed as a face of $\mathbb{R}^{c+d}_{\geq 0}$ plus a face of $\Gamma(Q)$. We say that a face of $\Gamma(Q)$ induces a facet of $\Gamma_+(Q)$ if there exists a face of the standard cone $\mathbb{R}^{c+d}_{\geq 0}$ such that the sum of these two faces gives a facet of $\Gamma_+(Q)$. However, since the dimension $\Gamma(Q)$ is lower than the dimension of the resulting polyhedron it turns out that one face of $\Gamma(Q)$ can induce more than one facet of $\Gamma_+(Q)$. In particular $\Gamma(Q)$ itself induces more than one facet and one of them is F_0 given by $\sum_{e\in E_0} x_e \geq 4$.

Every facet of $\Gamma_+(Q)$ containing $4/n\mathbf{1}$ after normalizing the coefficients to sum to n, i.e. $\sum_v \alpha_v + \sum_e \beta_e = n$, is of the form

$$\sum_v \alpha_v y_v + \sum_e \beta_e x_e \geq 4. \tag{30}$$

Our approach can be summarized as follows. Using Construction 27 we provide coordinates of a point $q \in \Gamma(Q)$ such that $4/n\mathbf{1}$ lies on the boundary of $q + \mathbb{R}^{c+d}_{\geq 0}$. Then $4/n\mathbf{1}$ can only lie on faces of $\Gamma_+(Q)$ induced by faces of $\Gamma(Q)$ containing q.

First, assume that T is rooted in one of the leaves. Consider a point $p \in \Gamma(Q)$ induced by a network constructed in the second part of Lemma 27. From the description of $\Gamma(Q)$ in Lemma 31 we can check that p lies in the interior of $\Gamma(Q)$ since all facet defining inequalities are strict for this point. Hence the only facets of $\Gamma_+(Q)$ containing p are these induced by $\Gamma(Q)$ itself. How many of these facets can be represented in the form of (30). We can find this representation by checking combinations of the defining equations: $\sum_{e\in E_0} x_e = 4$ and

$$2y_v - x_{v\mathrm{ch}_1(v)} - x_{v\mathrm{ch}_2(v)} + x_{\mathrm{pa}(v)v} = 0 \tag{31}$$

for all inner nodes v. The first one defining F_0 is already of this form (the sum of coefficients is n since there are n terminal edges). Any other facet has to be obtained by adding to the first equation (since the right hand side in (30) is 4) a non-negative (since the coefficients in front of y_v need to be non-negative) combination of equations in (31). However, since the sum of the coefficients in (31) is $+1$ this contradicts the assumption that the sum of coefficients in the defining inequality is n. Consequently, if T is rooted in one of the leaves then the codimension of the face hit by $4/n\mathbf{1}$ is one and hence $w = 1$.

Second, if the root is an inner node with no adjacent leaves we use a similar argument. Since all the nodes adjacent to r (denote them by a, b, c) are inner we have three different ways of conducting the construction in Lemma 27 (by omitting each of the incident edges). Hence we get three different points and their barrycenter satisfies $x_{ra} = x_{rb} = x_{rc} = 8/3n$ and $x_e = 4/n$ for all the other edges; $y_r = 4/n$, $y_a = y_b = y_c = 8/3n$ and $y_v = 2/n$ for all the other inner nodes. Denote this point by q. By the facet description of $\Gamma(Q)$ derived in Lemma 31 we can check that this point cannot lie in any of the facets defining $\Gamma(Q)$ and hence it is an interior point of the polytope. As in the first case it means that the facets of $\Gamma_+(Q)$ containing q are induced by $\Gamma(Q)$. Here the affine span is given by the equation defining F_0, the equations (31) for all inner edges v apart from the root and in addition for the root we have

$$2y_r - x_{ra} - x_{rb} - x_{rc} = 0. \tag{32}$$

Since the sum of coefficients in the above equation is negative we cannot use the same argument as in the first case. Instead we add to $\sum_{e\in E_0} x_e = 4$ a non-negative combination of equations in (31) each with coefficient $t_v \geq 0$ and then we add (32) with coefficient $\sum_{v\neq r} t_v$. The sum of coefficients in the resulting equation will be n by construction. The coefficient of x_{ra} is $t_a - \sum_{v\neq r} t_v$. Since it has to non-negative it follows that $t_v = 0$ for all v apart from a. However, by checking the coefficient of x_{rb} one deduces that in fact $t_v = 0$ for all inner nodes v. Consequently the only possible facet of $\Gamma_+(Q)$ containing $4/n\mathbf{1}$ is F_0 and hence $w = 1$.

Third, if there are exactly two inner nodes adjacent to r, say a and b, then in a similar fashion one constructs q as a barrycenter of two points obtained by omitting one of the two edges. The point defined in this way is given by $y_r = 4/n$, $y_a = y_b = 3/n$ and $y_v = 2/n$ for all other inner nodes, $x_{ra} = x_{rb} = 2/n$ and $x_e = 4/n$ for all other edges. Denote the only leaf adjacent to r by 1. By Lemma 31 the point lies on exactly one of the facets of $\Gamma(Q)$, namely the one given by $x_{ra} + x_{rb} - x_{r1} \geq 0$. Using the same as in the previous paragraph we show that the only facet of $\Gamma_+(Q)$ induced by $\Gamma(Q)$ and containing $4/n\mathbf{1}$ is F_0. We analyze the facets induced by the inequality above. We have

$$\sum_{e\in E_0} x_e + \sum_v t_v \left(2y_v - x_{v\mathrm{ch}_1(v)} - x_{v\mathrm{ch}_2(v)} + x_{\mathrm{pa}(v)v}\right) +$$
$$+ s\left(x_{ra} + x_{rb} - x_{r1}\right) + \big(s + \sum_v t_v\big)\left(2y_r - x_{ra} - x_{rb} - x_{r1}\right) \geq 4$$

where we assume that $s > 0$. The sum of all the coefficients again by construction is equal to n. We check the coefficients of x_{ra} and x_{rb} to deduce that $t_v = 0$ for all

v. We obtain

$$2sy_r + \sum_{e\in E_0\setminus(r,1)} x_e + (1-2s)x_{r1} \geq 4$$

and hence $s \in [0, 1/2]$, where zero just gives F_0 and $s = 1/2$ gives a new facet $y_r + \sum_{e\in E_0\setminus(r,1)} x_e \geq 4$. Consequently $4/n\mathbf{1}$ lies in two facets and then necessarily $w = 2$.

In the last case, when there are exactly two leaves adjacent to r. The point p defined in Lemma 27 can be constructed only in one way. It satisfies $y_r = y_a = 4/n$, $y_v = 2/n$ for the rest of inner nodes, $x_{ra} = 0$ and $x_e = 4/n$ for the rest of the edges. It can be checked that the point satisfies four inequalities in Equation (24) as equalities namely

$$\begin{aligned} x_{r1} + x_{ra} - x_{r2} \geq 0, \quad x_{r2} + x_{ra} - x_{r1} \geq 0, \\ x_{ab} + x_{ra} - x_{ac} \geq 0, \quad x_{ac} + x_{ra} - x_{ab} \geq 0, \end{aligned} \tag{33}$$

where b, c are the nodes adjacent to a. Now we consider three cases: both b and c are inner, exactly one of them is a leaf, and both are leaves (T is a quartet tree). In all the cases we easily show there is at least one facet apart from F_0 containing $4/n\mathbf{1}$. It follows that $w \geq 2$.

□

References

[1] V. Arnold, S. Gusein-Zade, and A. Varchenko, *Singularities of differentiable maps*, vol. II, Birkhäuser, 1988.
[2] E. Gawrilow and M. Joswig, *Geometric reasoning with polymake*. arXiv:math.CO/0507273, 2005.
[3] D. Geiger, D. Heckerman, H. King, and C. Meek, *Stratified exponential families: graphical models and model selection*, Ann. Statist., 29 (2001), pp. 505–529.
[4] D. Haughton, *On the choice of a model to fit data from an exponential family*, Ann. Statist, 16 (1988), pp. 342–355.
[5] J. Kollár, *Singularities of pairs*, in Proceedings of Symposia in Pure Mathematics, vol. 62, 1997, pp. 221–288.
[6] R. Lazarsfeld, *Positivity in algebraic geometry*, A Series of Modern Surveys in Mathematics, Springer Verlag, 2004.
[7] S. Lin, *Asymptotic approximation of marginal likelihood integrals*. in preparation, 2010.
[8] R. Mihaescu and L. Pachter, *Combinatorics of least-squares trees*, Proceedings of the National Academy of Sciences of the United States of America, 105 (2008), p. 13206.
[9] V. Moulton and M. Steel, *Peeling phylogenetic 'oranges'*, Advances in Applied Mathematics, 33 (2004), pp. 710–727.
[10] D. Rusakov and D. Geiger, *Asymptotic model selection for naive Bayesian networks*, J. Mach. Learn. Res., 6 (2005), pp. 1–35 (electronic).
[11] M. Saito, *On real log canonical thresholds*, Arxiv preprint arXiv:0707.2308 v, 1 (2007).
[12] G. Schwarz, *Estimating the dimension of a model*, Annals of Statistics, 6 (1978), pp. 461–464.
[13] C. Semple and M. Steel, *Phylogenetics*, vol. 24 of Oxford Lecture Series in Mathematics and its Applications, Oxford University Press, Oxford, 2003.
[14] R. P. Stanley, *Enumerative combinatorics. Volume I*, no. 49 in Cambridge Studies in Advanced Mathematics, Cambridge University Press, 2002.
[15] S. Watanabe, *Algebraic Analysis for Nonidentifiable Learning Machines*, Neural Computation, 13 (2001), pp. 899–933.
[16] S. Watanabe, *Algebraic Geometry and Statistical Learning Theory*, no. 25 in Cambridge Monographs on Applied and Computational Mathematics, Cambridge University Press, 2009. ISBN-13: 9780521864671.

[17] K. Yamazaki and S. Watanabe, *Newton diagram and stochastic complexity in mixture of binomial distributions*, in Algorithmic learning theory, vol. 3244 of Lecture Notes in Comput. Sci., Springer, Berlin, 2004, pp. 350–364.

[18] P. Zwiernik and J. Q. Smith, *The geometry of conditional independence tree models with hidden variables*. Arxiv preprint arXiv:0904.1980, January 2010.

Piotr Zwiernik, University of Warwick, Department of Statistics, CV7AL, Coventry, UK.

E-mail address: p.w.zwiernik@warwick.ac.uk