Journal of Machine Learning Research 12 (2011) 3283-3310 Submitted 12/10; Revised 9/11; Published 11/11

An Asymptotic Behaviour of the Marginal Likelihood for General Markov Models

Piotr Zwiernik PIOTR.ZWIERNIK@GMAIL.COM
Institute for Pure and Applied Mathematics
460 Portola Plaza
Box 957121
Los Angeles, CA 90095-7121

Editor: Kenji Fukumizu

Abstract

The standard Bayesian Information Criterion (BIC) is derived under regularity conditions which are not always satisfied in the case of graphical models with hidden variables. In this paper we derive the BIC for the binary graphical tree models where all the inner nodes of a tree represent binary hidden variables. This provides an extension of a similar formula given by Rusakov and Geiger for naive Bayes models. The main tool used in this paper is the connection between the growth behavior of marginal likelihood integrals and the real log-canonical threshold.

Keywords: BIC, marginal likelihood, singular models, tree models, Bayesian networks, real log-canonical threshold

1. Introduction

A key step in the Bayesian learning of graphical models is to compute the *marginal likelihood* of the data, which is the *likelihood function* averaged over the parameters with respect to the prior distribution. Given a fully observed system, the theory of graphical models provides a simple way to obtain the marginal likelihood. This was explained for example by Cooper and Herskovits (1992) and Heckerman et al. (1995). However, when some of the variables in the system are *hidden* (never observed), the exact determination of the marginal likelihood is typically intractable (for example Chickering and Heckerman, 1997). This motivates the search for efficient techniques to approximate the marginal likelihood. In this paper we focus on the large sample behavior of the marginal likelihood called the *Bayes Information Criterion* (BIC).

To present basic results on the BIC we need to introduce some notation. Let X be a random variable with values in $[m] := \{1,\ldots,m\}$. Its distribution $q = (q_1,\ldots,q_m)$ can be identified with a point in the *probability simplex* $\Delta_{m-1} = \{x \in \mathbb{R}^m : \sum_i x_i = 1, x_i \geq 0\} \subseteq \mathbb{R}^m$. Consider a map $p : \Theta \to \Delta_{m-1}$ and let $\mathcal{M} = p(\Theta)$ be a parametric discrete model for X with the parameter space Θ and parametrization p. Let $X^{(N)} = X^1,\ldots,X^N$ be a random sample from the distribution $q \in \Delta_{m-1}$. By Z_N we denote the marginal likelihood and by $L(\theta; X^{(N)}, \mathcal{M}) = \mathbb{P}(X^{(N)}|\mathcal{M}, \theta)$ the likelihood function. Thus

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

where θ denotes the model parameters and $\varphi(\theta)$ is a prior distribution on Θ. The *stochastic complexity* is defined by

$$F_N = -\log Z_N$$

and the entropy function by

$$S = -\sum_{i=1}^{m} q_i \log q_i.$$

In statistical theory to obtain the BIC we usually require that the asymptotic limit of the likelihood function, as $N \to \infty$, is maximized over a unique point in the interior of the parameter space where the Jacobian matrix of the parametrization is full rank. For the class of problems for which this assumption is satisfied Schwarz (1978) and Haughton (1988) showed that, as $N \to \infty$,

$$\mathbb{E}F_N = NS + \frac{d}{2}\log N + O(1),$$

where $d = \dim\Theta$ (Watanabe, 2009, Corollary 1.15 and Section 6). The same formula works if the limit of the likelihood is maximized over a finite number of points in the interior of Θ. Geometrically, for large sample sizes function Z_N concentrates around the maxima. This enables us to apply the Laplace approximation locally in the neighborhood of each maximum.

It can be proved (see Proposition 5) that the above formula can be generalized for the case when the set, over which the limit of the likelihood is maximized, forms a sufficiently regular compact subset of the parameter space. Denote this subset by $\widehat{\Theta}$. Then, as $N \to \infty$,

$$\mathbb{E}F_N = NS + \frac{d-d'}{2}\log N + O(1), \tag{1}$$

where $d' = \dim\widehat{\Theta}$. Note that in our case $\widehat{\Theta}$ is a set of zeros of a real analytic function. Therefore, it will be always a semi-analytic set, that is given by $\{g_1(\theta) \geq 0, \ldots, g_r(\theta) \geq 0\}$, where g_i are all analytic functions. It follows that the dimension is well defined (Bierstone and Milman, 1988, Remark 2.12).

In the case of models with hidden variables the locus of the points maximizing the limit of the likelihood may not be sufficiently regular. In this case the likelihood will have a different asymptotic behavior around different points and relatively more mass of the marginal likelihood integral will be related to neighborhoods of singular points (see Figure 1). For these points we cannot use the Laplace approximation. Nevertheless, the computation of the BIC is still possible using results of Watanabe (2009) and some earlier works of Arnold, Varchenko and collaborators (Arnold et al., 1988). This formula will differ from the standard BIC. First, the coefficient of $\log N$ can be different from $\frac{d-d'}{2}$ in (1). Second, we sometimes encounter an additional $\log\log N$ term affecting the asymptotics (see Theorem 4).

Let again q be the true data generating distribution and $\mathcal{M} = p(\Theta)$ a discrete parametric statistical model with the parameter space Θ. Let $\varphi : \Theta \to \mathbb{R}$ be a prior distribution. Throughout the paper we always assume:

(A1) The prior distribution φ is strictly positive, bounded and smooth on Θ.

(A2) There exists $\theta \in \Theta$ such that $p(\theta) = q$ and q lies in the interior of the probability simplex.

(A3) The set $\Theta \subseteq \mathbb{R}^d$ is a compact and *semianalytic set* of dimension d.

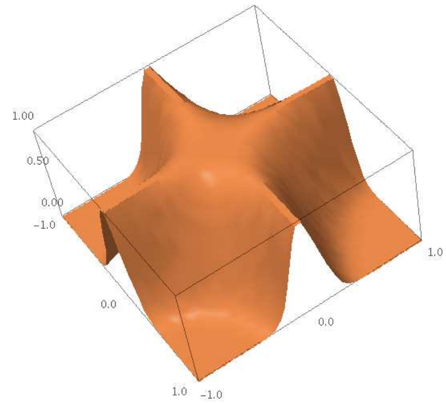


Figure 1: The case when the likelihood is maximized over a singular subset of $\Theta = [-1,1]^2$ given by $\theta_1\theta_2 = 0$.

In this paper we consider an important class of parametric models with large number of hidden variables, called *general Markov models*, assuming for simplicity that all the random variables in the system are binary. This model class is extensively used in phylogenetics (Semple and Steel, 2003, Chapter 8) and in the analysis of causal systems (see Pearl and Tarsi, 1986). We begin with a quick informal introduction to general Markov models which is then formalized in Section 3.1. Let $T = (V, E)$ be an undirected tree with the vertex set V and the edge set E. Let T^r denote T *rooted in* r, that is a tree with one distinguished vertex r and all the edges directed away from r. Consider the *Markov process* $Y = (Y_v)_{v\in V}$ on T^r, which by definition is the Bayesian network on T^r. Then, the *general Markov model* is a family of marginal distributions over the subvector of Y corresponding to the leaves of T^r. It is well known that this model class does not depend on the rooting. Therefore, we denote this model class, omitting the rooting, by $\mathcal{M}_T$.

For a tree T with n leaves we denote the subvector of Y corresponding to the leaves of T by $X = (X_1, \ldots, X_n)$ with some arbitrary numbering of leaves. The subvector of Y corresponding to the inner nodes is denoted by H. By construction the general Markov model is a statistical model for X. Let $q \in \mathcal{M}_T$ be the true distribution and $\widehat{\Sigma} = [\hat{\mu}_{ij}]$ be the covariance matrix of X. A surprising fact proved in this paper is that the zeros in $\widehat{\Sigma}$, or equivalently, marginal independencies between components of X, completely determine the asymptotics for the marginal likelihood.

We say that two nodes u, v of T are *separated* by another node w, if w lies on the unique path between u and v. Let l_2 denote the number of *inner* nodes v of T such that for each triple i, j, k of leaves separated in T by v we have $\hat{\mu}_{ij}\hat{\mu}_{ik}\hat{\mu}_{jk} = 0$ but there exist leaves i, j separated by v such that $\hat{\mu}_{ij} \neq 0$. In terms of the conditional independence defining the general Markov model, an inner node v contributes to l_2 if for every three leaves i, j, k such that $X_i \perp\!\!\!\perp X_j \perp\!\!\!\perp X_k | H_v$ at least two are marginally independent but there exist two leaves i, j such that $X_i \perp\!\!\!\perp X_j | H_v$ but not $X_i \perp\!\!\!\perp X_j$. In addition, we say that an inner node v is *degenerate (or q-degenerate)* if for any two leaves i, j separated by v we have $\hat{\mu}_{ij} = 0$. In other words v is degenerate if for every i, j such that $X_i \perp\!\!\!\perp X_j | H_v$ also $X_i \perp\!\!\!\perp X_j$. All other nodes are called *nondegenerate*.

We denote by n_e the number of edges of T and by n_v the number of its nodes. The following result is a special instance of (Watanabe, 2009, the Main Formula II, p. 34):

Theorem 1 *Let $T^r = (V,E)$ be a tree rooted in r and $X^{(N)}$ be a random sample from q. With assumptions (A1), (A2) and (A3), if there are no q-degenerate nodes then, as $N \to \infty$,*

$$\mathbb{E}F_N = NS + \frac{n_v + n_e - 2l_2}{2} \log N + O(1).$$

Note, in particular, that the above formula is independent of the rooting.

Example 1 *Let $n = 4$ and assume that data are generated from the Bayesian network given by the quartet tree in Figure 2. If q is such that $\widehat{\Sigma}$ has no zeros then $l_2 = 0$ and the coefficient of $\log N$ is*

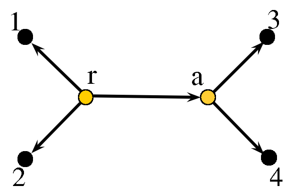


Figure 2: A quartet tree rooted in r.

$\frac{11}{2}$. This corresponds to the classical BIC since the dimension of the parameter space is 11. If the true distribution $q \in \mathcal{M}_T$ satisfies in addition the marginal independence condition $X_1 \perp\!\!\!\perp (X_2, X_3, X_4)$ then $\hat{\mu}_{1i} = 0$ for $i = 2, 3, 4$ and r contributes to l_2. We depict this situation on the left hand side in Figure 3. Here the dashed edge means that for every pair i, j of leaves separated by this edge $\hat{\mu}_{ij} = 0$ and an inner node contributes to l_2 if its valency, in the forest with the dashed edges removed, is 2. In this case $l_2 = 1$ and the coefficient of $\log N$ is $\frac{9}{2}$. If, in addition, q satisfies $X_1 \perp\!\!\!\perp X_3 \perp\!\!\!\perp (X_2, X_4)$ then $l_2 = 2$ and hence the coefficient is $\frac{7}{2}$. The corresponding graph is depicted in the middle of Figure 3.

Example 2 (Naive Bayes model) *Consider a star tree with one inner node and n leaves. If there are no degenerate nodes this corresponds to q being either a regular point or a type 1 singularity as defined by Rusakov and Geiger (2005). If $l_2 = 0$ then q is a regular point and the coefficient of $\log N$ is equal to $\frac{2n+1}{2}$. If $l_2 = 1$ then q is a type 1 singularity and the coefficient is equal to $\frac{2n-1}{2}$. This corresponds exactly to (Rusakov and Geiger, 2005, Theorem 4). If the inner node is degenerate this corresponds to the type 2 singularity which does not satisfy assumptions of Theorem 1.*

If q is such that there are degenerate nodes the computation of the BIC is much harder because the likelihood in this case maximizes over a singular subset of the parameter space. The case of star

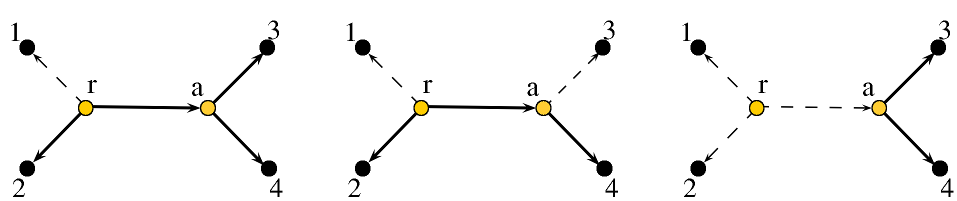


Figure 3: Three graphs representing submodels of the quartet tree model with some additional marginal independencies. In the third case the root is degenerate.

trees was investigated by Rusakov and Geiger (2005). In this paper we obtain a closed form formula for the BIC in the case of *trivalent trees*, whose all inner nodes have valency three. This is provided in Theorem 2 which together with Theorem 1 are the main results of this paper. The importance of trivalent trees follows mainly from the fact that any other tree model is a submodel of a model for a trivalent tree. They also form a natural class for models of evolution in biology.

If T is trivalent then for every inner node $v \in V$ there exist $A, B, C \subseteq [n]$ such that $A \cup B \cup C = [n]$ and A, B, C are separated by v. By the defining conditional independence conditions we have that $X_A \perp\!\!\!\perp X_B \perp\!\!\!\perp X_C | Y_v$, where $X_A = (X_i)_{i \in A}$. In this case we call v degenerate if q is such that $X_A \perp\!\!\!\perp X_B \perp\!\!\!\perp X_C$. Let l_0 denote the number of degenerate nodes.

Theorem 2 *Let $T^r = (V, E)$ be a rooted trivalent tree with $n \geq 3$ leaves and root r. With assumptions (A1), (A2) and (A3) if r is degenerate but all its neighbors are not, then, as $N \to \infty$,*

$$\mathbb{E}F_N = NS + \left(\frac{n_v + n_e - 2l_2}{2} - \frac{5l_0 + 1}{4} \right) \log N + O(1).$$

In all other cases, as $N \to \infty$,

$$\mathbb{E}F_N = NS + \left(\frac{n_v + n_e - 2l_2}{2} - \frac{5l_0}{4} \right) \log N - c \log\log N + O(1),$$

where c is a nonnegative integer. Moreover $c = 0$ always if either both r is nondegenerate or if r and all its neighbors are degenerate.

The coefficients of $\log N$ above are given in this special form to show the correction term with respect to the coefficient in the smooth case in Theorem 1.

Example 3 *Consider again the quartet model from Example 1. If $\widehat{\Sigma}$ has no zeros then $l_2 = 0$, $l_0 = 0$ and we get the same formula as previously with coefficient $\frac{11}{2}$. Now assume that q is such that in addition the marginal independence $X_1 \perp\!\!\!\perp X_2 \perp\!\!\!\perp (X_3, X_4)$ holds. The situation is depicted on the right hand side in Figure 3. The edge (r, a) is dashed since for any two leaves separated by this edge the corresponding covariance is zero. In this case $l_2 = 1$, $l_0 = 1$, the root is degenerate but all its neighbors are not and hence, by Theorem 2, the coefficient of $\log N$ is 3 and $c = 0$. Consider finally the case when all off-diagonal elements of $\widehat{\Sigma}$ are zero. In this case $l_2 = 0$ and $l_0 = 2$ and hence the coefficient of $\log N$ is also 3. However, later in Example 5 and Remark 29 we will see that c may be strictly greater than zero in this case.*

Following Rusakov and Geiger (2005), the main method of the proof is to change the coordinates of the model so that the induced parameterization becomes simple. This gives us a much better insight into the model structure which is described by Zwiernik and Smith (2011b) and Zwiernik and Smith (2011a). Since the BIC is invariant with respect to these changes, the reparameterized problem still gives the solution to the original question. Our main analytical tool is the real log-canonical threshold (see for example Watanabe, 2009). This is an important geometric invariant which in certain cases can be computed in a relatively simple way using discrete geometry. The relevance of this invariant to the BIC is given by Theorem 4. We remark that the techniques developed in this paper can be applied to obtain the BIC also in the case of non-trivalent trees.

The paper is organized as follows. In Section 2, following Watanabe (2009), we provide the theory of asymptotic expansion of marginal likelihood integrals. This theory enables us to analyze the asymptotic behavior of the marginal likelihood without the standard regularity assumptions. In Section 3 we define Bayesian networks on rooted trees. We also obtain a basic result on the BIC in the case when the observed likelihood is maximised over a sufficiently smooth subset of the parameter space. This gives a simple proof of Theorem 1. The proof of Theorem 2 is more technical and hence divided into three main steps split between Sections 4, 5 and 6. Finally, in Section 7, we combine all these results.

2. Asymptotics of Marginal Likelihood Integrals

In this section we introduce the real log-canonical threshold and link it to the asymptotic behavior of marginal likelihood integrals. We present how this enables us to obtain the BIC in the case of a general class of statistical models, which is mostly based on previous results of Sumio Watanabe.

2.1 The Real Log-canonical Threshold

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 $\Theta \subset \mathbb{R}^d$ satisfying (A3), let $\mathcal{A}_\Theta(\mathbb{R}^d)$ be the ring of real functions analytic at each point $\theta_0 \in \Theta$. If $f \in \mathcal{A}_\Theta(\mathbb{R}^d)$, then for every $\theta_0 \in \Theta$, f can be locally represented as power series centered at θ_0. Denote by $\mathcal{A}^{\geq}_\Theta(\mathbb{R}^d)$ the 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 our notation writing $\mathcal{A}_{\theta_0}$, $\mathcal{A}_\Theta$ and $\mathcal{A}^{\geq}_\Theta$.

Definition 3 (The real log-canonical threshold) *Given a compact semianalytic set $\Theta \subseteq \mathbb{R}^d$ such that* $\dim \Theta = d$*, a real analytic function $f \in \mathcal{A}^{\geq}_\Theta(\mathbb{R}^d)$ and a smooth positive function $\varphi : \mathbb{R}^d \to \mathbb{R}$, consider the zeta function defined as*

$$\zeta(z) = \int_\Theta f(\theta)^{-z}\varphi(\theta)\mathrm{d}\theta. \tag{2}$$

By Theorem 2.4 of Watanabe (2009) this function is extended to a meromorphic function on the entire complex line and its poles are real and positive. The real log-canonical threshold of f denoted by $\mathrm{rlct}_\Theta(f;\varphi)$ *is the smallest pole of $\zeta(z)$. By* $\mathrm{mult}_\Theta(f;\varphi)$ *we denote the multiplicity of this pole. 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 $\mathcal{M} = p(\Theta) \subseteq \Delta_{m-1}$ be a parametric discrete model and $q \in \Delta_{m-1}$ be a probability distribution. With $\mathcal{M}$ and q fixed the Kullback-Leibler distance $K : \Theta \to \mathbb{R}$ is defined by

$$K(\theta) = \sum_{i=1}^{m} q_i \log \frac{q_i}{p_i(\theta)}. \tag{3}$$

It is well known that $K(\theta) \geq 0$ on Θ and $K(\theta) = 0$ if and only if $p(\theta) = q$. If q is the true data generating distribution then assumption (A2) means that $\widehat{\Theta} = \{\theta : K(\theta) = 0\}$ is non-empty.

The following theorem gives the motivation to study the real log-canonical threshold in the statistical context.

Theorem 4 (Watanabe) *Let $\mathcal{M}$ be a parametric discrete statistical model, q the true data generating distribution and K the corresponding Kullback-Leibler distance. With assumptions (A1), (A2) and (A3), as $N \to \infty$,*

$$\mathbb{E}F_N = NS + \mathrm{rlct}_\Theta(K;\varphi)\log N + (\mathrm{mult}_\Theta(K;\varphi)-1)\log\log N + O(1).$$

To compute the real log-canonical threshold we split the integral in (2) into a sum of finitely many integrals over small neighbourhoods Θ_0 of some points $\theta_0 \in \Theta$ for which we have efficient tools of computation. We can always do this using the partition of unity since Θ is compact. For each of the local integrals we use Hironaka's theorem to reduce it to a locally monomial case. The details are presented by Watanabe (2009).

Let $\theta_0 \in \Theta$ and let W_0 be any sufficiently small open ball around θ_0 in $\mathbb{R}^d$. Then, by Theorem 2.4 of Watanabe (2009), $\mathrm{RLCT}_{W_0}(f;\varphi)$ does not depend on the choice of W_0 and hence it is denoted by $\mathrm{RLCT}_{\theta_0}(f;\varphi)$. If $f(\theta_0) \neq 0$ then $\mathrm{RLCT}_{W_0}(f;\varphi) = (\infty, d)$ and hence we can constrain only to points θ_0 such that $f(\theta_0) = 0$. In our context this means that we consider only points in the q-fiber $\widehat{\Theta}$.

The local computations give the answer to the global question because, by (Lin, 2011, Proposition 2.5), the set of pairs $\mathrm{RLCT}_{\Theta_0}(f;\varphi)$ for $\theta_0 \in \Theta$ has a minimum and

$$\mathrm{RLCT}_\Theta(f;\varphi) = \min_{\theta_0 \in \Theta} \mathrm{RLCT}_{\Theta_0}(f;\varphi), \tag{4}$$

where $\Theta_0 = W_0 \cap \Theta$ is the neighbourhood of θ_0 in Θ. For each $\theta_0 \in \Theta$ to compute $\mathrm{RLCT}_{\Theta_0}(f;\varphi)$ we consider two cases. If θ_0 lies in the interior of Θ then we can assume $\Theta_0 = W_0$ and hence $\mathrm{RLCT}_{\Theta_0}(f;\varphi) = \mathrm{RLCT}_{\theta_0}(f;\varphi)$. If $\theta_0 \in \mathrm{bd}(\Theta)$, where $\mathrm{bd}(\Theta)$ denotes the set of boundary points of Θ, the computations may change significantly because the real log-canonical threshold depends on the boundary conditions (cf. Example 2.7 of Lin, 2011). Nevertheless, it can be showed that at least if there exists an open subset $U \subseteq \mathbb{R}^d$ such that $U \supset \Theta_0$ and $f \in \mathcal{A}_U^{\geq}(\mathbb{R}^d)$ then

$$\mathrm{RLCT}_{\Theta_0}(f) \geq \mathrm{RLCT}_{\theta_0}(f). \tag{5}$$

Because in this case

$$\int_{W_0} (f(\theta))^{-z} \mathrm{d}\theta = \int_{\Theta_0} (f(\theta))^{-z} \mathrm{d}\theta + \int_{W_0 \setminus \Theta_0} (f(\theta))^{-z} \mathrm{d}\theta$$

which implies that $\mathrm{RLCT}_{\theta_0}(f) = \min\{\mathrm{RLCT}_{\Theta_0}(f), \mathrm{RLCT}_{W_0\setminus\Theta_0}(f)\}$.

Finally, whenever $\widehat{\Theta} \neq \emptyset$ we have

$$\mathrm{RLCT}_\Theta(K) = \min_{\theta_0 \in \widehat{\Theta}} \mathrm{RLCT}_{\Theta_0}(K). \tag{6}$$

The following result enables us to obtain the BIC in the smooth case.

Proposition 5 *Let $\mathcal{M}$ be a parametric statistical model with parametrization p, and q be the true data generating distribution. Let $K \in \mathcal{A}_\Theta^{\geq}(\mathbb{R}^d)$ be the Kullback-Leibler distance defined in (3). Given (A1), (A2) and (A3) assume that there exists a smooth manifold $M \subseteq \mathbb{R}^d$ satisfying $\widehat{\Theta} = M \cap \Theta$. Then, as $N \to \infty$,*

$$\mathbb{E}F_N = NS + \frac{d-d'}{2}\log N + O(1),$$

where $d' = \dim \widehat{\Theta}$.

Proof By assumption (A1) there exist two constants $c, C > 0$ such that $c < \varphi(\theta) < C$ on Θ. Therefore

$$c\int_{\Theta}(K(\theta))^{-z}\mathrm{d}\theta < \zeta(z) < C\int_{\Theta}(K(\theta))^{-z}\mathrm{d}\theta$$

and it follows that $\mathrm{RLCT}_{\Theta}(K;\varphi) = \mathrm{RLCT}_{\Theta}(K)$. By Theorem 4 it suffices to prove the following lemma which generalises Proposition 3.3 of Saito (2007).

Lemma 6 *Let $\Theta \subset \mathbb{R}^d$ be a compact semianalytic set and $f \in \mathcal{A}_{\Theta}^{\geq}(\mathbb{R}^d)$. If there exists a smooth manifold $M \subseteq \mathbb{R}^d$ such that $\widehat{\Theta} = M \cap \Theta$ and $\theta_0 \in \widehat{\Theta}$ then* $\mathrm{RLCT}_{\theta_0}(f) = \mathrm{RLCT}_{\Theta}(f) = (\frac{d-d'}{2}, 1)$ *where $d' = \dim\widehat{\Theta}$.*

To prove this recall that the real log-canonical threshold $\mathrm{RLCT}_{\theta_0}(f)$ does not depend on the choice of a sufficiently small neighborhood W_0 of θ_0. Since $\widehat{\Theta} = M \cap \Theta$ and M is a smooth manifold it follows that for each point θ_0 of $\widehat{\Theta}$ there exists an open neighborhood W_0 of θ_0 in $\mathbb{R}^d$ with local coordinates $w_1, \dots, w_d$ centered at θ_0 such that the local equation of $\widehat{\Theta}$ is $w_1^2 + \cdots + w_c^2 = 0$, where $c = d - d'$. A single blow-up π at the origin satisfies all the conditions of Hironaka's Theorem since in the new coordinates over one of the charts $f(\pi(u)) = u_1^2 a(u)$ where $a(u)$ is nowhere vanishing and $\pi'(u) = u_1^{c-1}$. For other charts the situation is the same and hence $\mathrm{RLCT}_{\theta_0}(f) = (c/2, 1)$. Since by (4) $\mathrm{RLCT}_{\Theta}(f) = \min_{\theta_0 \in \Theta} \mathrm{RLCT}_{W_0 \cap \Theta}(f)$ it suffices to show that if θ_0 is a boundary point of Θ then $\mathrm{RLCT}_{W_0\cap\Theta}(f) \geq (c/2, 1)$. But this follows from (5) and the fact that $\mathrm{RLCT}_{\theta_0}(f) = (c/2, 1)$ as θ_0 is a smooth point of M. The lemma is hence proved. ∎

3. General Markov Models

In this section we formally define the general Markov model $\mathcal{M}_T$ and give in Theorem 1 the asymptotic expansion of the marginal likelihood when q and $\mathcal{M}_T$ satisfy conditions of Proposition 5.

3.1 Definition of the Model Class

All random variables considered in this paper are assumed to be binary with the value either 0 or 1. Let $T^r = (V, E)$ be a rooted tree. For any directed edge $e = (k, l) \in E$ we say that k and l are *adjacent* and k is a *parent* of l and we denote it by $k = \mathrm{pa}(l)$. For every $\beta \in \{0,1\}^V$ let $p_\beta = \mathbb{P}(\bigcap_{v\in V}\{Y_v = \beta_v\})$. The *Markov process* on T^r is a sequence $Y = (Y_v)_{v\in V}$ of binary random variables such that for each $\beta = (\beta_v)_{v\in V} \in \{0,1\}^V$

$$p_\beta(\theta) = \theta^{(r)}_{\beta_r} \prod_{v \in V\setminus r} \theta^{(v)}_{\beta_v|\beta_{\mathrm{pa}(v)}}, \tag{7}$$

where $\theta^{(v)}_{\beta_v|\beta_{\mathrm{pa}(v)}} = \mathbb{P}(Y_v = \beta_v | Y_{\mathrm{pa}(v)} = \beta_{\mathrm{pa}(v)})$ and $\theta^{(r)}_{\beta_r} = \mathbb{P}(Y_r = \beta_r)$. In a more standard statistical language these models are just fully observed Bayesian networks on rooted trees. Recall that $n_e = |E|$ and $n_v = |V|$. 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^r defined by (7) 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 $\Theta_T = [0,1]^{2n_e+1}$.

The general Markov model on T^r is induced from the Markov process on T^r by assuming that all the inner nodes represent hidden random variables. Hence we consider induced marginal probability distributions over the leaves of T^r. The set of leaves is denoted by L. We assume that T^r has n leaves and hence we can associate L with the set $[n]$ given some arbitrary numbering of the leaves. Let $Y = (X,H)$ where $X = (X_1,\dots,X_n)$ denotes the variables represented by the leaves of T^r and H denotes the vector of variables represented by inner nodes, that is $X = (Y_v)_{v\in L}$ and $H = (Y_v)_{v\in V\setminus L}$. We define the general Markov model $\mathcal{M}_T$ to be the model in the probability simplex Δ_{2^n-1} obtained by summing out in (7) over all possible values of the inner nodes. By definition $\mathcal{M}_T$ is the image of the map $p : \Theta_T \to \Delta_{2^n-1}$ given by

$$p_\alpha(\theta) = \sum_{\mathcal{H}} \theta^{(r)}_{\beta_r} \prod_{v\in V\setminus r} \theta^{(v)}_{\beta_v|\beta_{\mathrm{pa}(v)}} \qquad \text{for any } \alpha \in \{0,1\}^L,$$

where $\mathcal{H}$ is the set of all vectors $\beta = (\beta_v)_{v\in V}$ such that $(\beta_v)_{v\in L} = \alpha$. Because the model class does not depend on the rooting we usually omit the root r in the notation. For a more detailed treatment see (Semple and Steel, 2003, Chapter 8).

3.2 The BIC in the Smooth Case

For $q \in \mathcal{M}_T$ let $\widehat{\Sigma} = [\hat{\mu}_{ij}] \in \mathbb{R}^{n\times n}$ be the corresponding covariance matrix of the random vector represented by the leaves of T. We define the q-fiber as

$$\widehat{\Theta}_T \quad = \quad \{\theta \in \Theta_T : p(\theta) = q\} \quad = \quad \{\theta \in \Theta_T : K(\theta) = 0\}.$$

The geometry of $\widehat{\Theta}_T$ is directly related to the real log-canonical threshold of the Kullback-Leibler distance. We now show that this geometry is determined by zeros in $\widehat{\Sigma}$. For this we need to introduce some new concepts. We say that that an edge $e \in E$ is *isolated relative to* q if $\hat{\mu}_{ij} = 0$ for all $i, j \in [n]$ such that $e \in E(ij)$, where $E(ij)$ denotes the set of edges in the path joining i and j. By $\widehat{E} \subseteq E$ we denote the set of all edges of T which are isolated relative to q. By $\widehat{T} = (V, E\setminus\widehat{E})$ we denote the forest obtained from T by removing edges in $\widehat{E}$.

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

By construction all the inner nodes of T have either degree zero in $\widehat{T}$ or the degree is strictly greater than one. We say that a node $v \in V$ is *non-degenerate with respect to* q 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* q. Note that this coincides with the definition of a degenerate node given in the introduction. Moreover, the isolated edges in Examples 1 and 3 correspond precisely to the dashed edges in Figure 3. The set of all nodes which are degenerate with respect to q is denoted by $\widehat{V}$.

Proposition 7 (Zwiernik and Smith, 2011b) *Let T be a tree with n leaves. Let $q \in \mathcal{M}_T$ and let $\widehat{T}$ be defined as above. If each of the inner nodes of T has degree at least two in $\widehat{T}$ then $\widehat{\Theta}_T$ is a manifold with corners and* $\dim\widehat{\Theta}_T = 2l_2$, *where l_2 is the number of nodes which have degree two in $\widehat{T}$.*

In this way we can compute the asymptotic behavior of the marginal likelihood in the case when assumptions of Proposition 7 are satisfied.

Proposition 8 *Let T be a tree and $\mathcal{M}_T$ be the corresponding general Markov model. Let $q \in \mathcal{M}_T$ be the real distribution generating the data such that each inner node of T has degree at least two in $\widehat{T}$. Then*

$$\mathrm{RLCT}_{\Theta_T}(K) = \left(\frac{n_v + n_e - 2l_2}{2}, 1 \right).$$

Proof Since every inner node of T has degree at least two in $\widehat{T}$ then by Proposition 7 there exists a smooth manifold $M \subseteq \mathbb{R}^{n_v + n_e}$ such that $\widehat{\Theta}_T = M \cap \Theta_T$ and $\dim \widehat{\Theta} = 2l_2$. The result follows from Proposition 5 and the fact that $\dim \Theta_T = 2n_e + 1 = n_v + n_e$. ∎

By Theorem 4, Proposition 8 implies Theorem 1 since l_2 in its statement is exactly the number of inner nodes v such that the degree of v in $\widehat{T}$ is two.

Remark 9 *Theorem 1 is still true if (A1) is replaced by the assumption that the prior distribution is bounded on Θ_T and there exists an open subset of Θ_T with a non-empty intersection with $\widehat{\Theta}_T$ where the prior is strictly positive. In particular we can use conjugate Beta priors $\theta^{(v)}_{1|i} \sim \mathrm{Beta}(\alpha^{(v)}_i, \beta^{(v)}_i)$ as long as $\alpha^{(v)}_i, \beta^{(v)}_i \geq 1$.*

4. The Ideal-theoretic Approach

In this section we define the real log-canonical threshold of an ideal. Theorem 11 translates the problem of finding the real log-canonical threshold of the Kullback-Leibler distance into algebra. We then analyse general Markov models from this perspective. In Theorem 14 we apply a useful change of coordinates which enables us to work out the real log-canonical threshold in the singular case.

4.1 The Real Log-canonical Threshold of an Ideal

Let $f_1, \ldots, f_r \in \mathcal{A}_\Theta$ then the *ideal generated* by $f_1, \ldots, f_r$ is a subset of $\mathcal{A}_\Theta$ denoted by

$$\langle f_1, \ldots, f_r \rangle = \{ f \in \mathcal{A}_\Theta : \ f(\theta) = \sum_{i=1}^{r} h_i(\theta) f_i(\theta), h_i \in \mathcal{A}_\Theta \}.$$

Following Lin (2011) we generalize the notion of the real log-canonical thresholds to the ideal $I = \langle f_1, \ldots, f_r \rangle$. This mirrors the analytic definition of the log-canonical threshold of an ideal (see for example Lazarsfeld, 2004, Section 9.3.D). By definition

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

where $f(\theta) = f_1^2(\theta) + \cdots + f_r^2(\theta)$. By (Lin, 2011, Proposition 4.5) the real log-canonical threshold does not depend on the choice of generators of I.

The following important proposition enables us to use the full power of the ideal-theoretic approach.

Proposition 10 *Let $f, g \in \mathcal{A}_\Theta(\mathbb{R}^d)$ and let I be an ideal in $\mathcal{A}_\Theta(\mathbb{R}^d)$. Then*

i *Let* $\rho : \Omega \to \Theta$ *be a proper real analytic isomorphism and* $\rho^* I = \{f \circ \rho : f \in I\}$ *be the pullback of* I *to* $\mathcal{A}_\Omega$. *Then,*

$$\mathrm{RLCT}_\Theta(I;\varphi) = \mathrm{RLCT}_\Omega(\rho^* I; (\varphi \circ \rho)|\rho'|),$$

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

ii *If* φ *is positive and bounded on* Θ *then*

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

iii *If there exist constants* $c, c' > 0$ *such that* $c\, g(\theta) \le f(\theta) \le c' g(\theta)$ *for every* $\theta \in \Theta$ *then* $\mathrm{RLCT}_\Theta(f) = \mathrm{RLCT}_\Theta(g)$.

iv *Let* $I = \langle f_1, \ldots, f_r \rangle$ *and* $J = \langle g_1, \ldots, g_r \rangle$ *where* $g_i = u_i f_i$ *for* $i = 1, \ldots, r$ *and there exist positive constants* c, C *such that* $c < u_i(\theta) < C$ *for all* $\theta \in \Theta$ *and for all* $i = 1, \ldots, r$. *Then* $\mathrm{RLCT}_\Theta(I) = \mathrm{RLCT}_\Theta(J)$.

The ideal-theoretic approach proves to be useful in a fairly general statistical context:

Theorem 11 (Lin, 2011) *Let* $p = (p_1, \ldots, p_m) : \Theta \to \Delta_{m-1}$ *be a polynomial mapping and* $\mathcal{M} = p(\Theta)$ *be the statistical model of* X *with values in* $[m]$. *For a given point* $q \in \mathcal{M}$ *define*

$$\overline{\mathscr{I}} = \langle p_1(\theta) - q_1, \ldots, p_m(\theta) - q_m \rangle \subset \mathcal{A}_\Theta. \tag{8}$$

Let q *denote the true data generating distribution and* $K(\theta)$ *be the corresponding Kullback-Leibler distance defined in (3). Moreover, let* φ *the prior distribution on* Θ *satisfying (A1). Then*

$$\mathrm{RLCT}_\Theta(K;\varphi) = \mathrm{RLCT}_\Theta(\overline{\mathscr{I}};\varphi) = \mathrm{RLCT}_\Theta(\overline{\mathscr{I}}), \tag{9}$$

where the second equation in (9) follows from Proposition 10 ii.

We now perform the change of coordinates $f_{\theta\omega} : \Theta_T \to \Omega_T$, $f_{p\kappa} : \Delta_{2^n-1} \to \mathcal{K}_T$ discussed in detail by Zwiernik and Smith (2011b). We have the following diagram, where the top row is the original parametrization and where the induced parameterisation ψ_T is given in the bottom row.

$$\begin{array}{ccc} \Theta_T & \xrightarrow{\;p\;} & \Delta_{2^n-1} \\ f_{\omega\theta}\uparrow\;\downarrow f_{\theta\omega} & & f_{\kappa p}\uparrow\;\downarrow f_{p\kappa} \\ \Omega_T & \overset{\psi_T}{\dashrightarrow} & \mathcal{K}_T \end{array}$$

Here $f_{\theta\omega}$ and $f_{p\kappa}$ are polynomial isomorphisms, and hence, by Proposition 10 (i), in our computations of the real-log canonical threshold, we can alternatively constrain to the bottom row of the diagram. We denote the coordinates of $\mathcal{K}_T$ by $\kappa = (\kappa_I)$ for $I \subseteq [n]$, $I \neq \emptyset$. The coordinates of Ω_T are denoted by $\omega = ((s_v), (\eta_{uv}))$ for all $v \in V$ and $(u,v) \in E$. Both ω and κ have a statistical meaning as described by Zwiernik and Smith (2011b). However, in this work we use them in a purely algebraic manner. We just note that the coordinates of $\mathcal{K}_T$ are certain functions of the moments. In particular, $\kappa_i = \mathbb{E} X_i$ for $i = 1, \ldots, n$ and each κ_{ij} corresponds to the covariance between X_i and X_j. Interpretation of other coordinates is more complicated.

Simple linear constraints defining Θ_T become only slightly more complicated when expressed in the new parameters. The choice of parameter values is not free anymore in the sense that constraints for each of the parameters involve other parameters. The new parameter space Ω_T is given by $s_r \in [-1,1]$ and for each $(u,v) \in E$ (cf. Equation (19) in Zwiernik and Smith, 2011b)

$$\begin{array}{c} -(1+s_v) \leq (1-s_u)\eta_{uv} \leq (1-s_v) \\ -(1-s_v) \leq (1+s_u)\eta_{uv} \leq (1+s_v). \end{array} \tag{10}$$

In the new coordinate system the situation is more tractable because ψ_T has a simpler structure.

Proposition 12 (Zwiernik and Smith, 2011b) *Let $T^r = (V,E)$ be a rooted trivalent tree with n leaves. Then for each $i = 1,\dots,n$ one has $\kappa_i(\omega) = \frac{1}{2}(1-s_i)$ and*

$$\kappa_I(\omega) = \frac{1}{4}(1-s_{r(I)}^2) \prod_{v \in V(I)\setminus I} s_v^{\deg v - 2} \prod_{(u,v)\in E(I)} \eta_{uv} \quad \text{for all } |I| \geq 2,$$

where the degree of $v \in V(I)$ is considered in $T(I) = (V(I),E(I))$, which is the smallest subtree of T containing I.

Let $\mathscr{I}$ denote the pullback of the ideal $\overline{\mathscr{I}} \subseteq \mathcal{A}_{\Theta_T}$ to the ideal in $\mathcal{A}_{\Omega_T}$ induced by $f_{\theta\omega}$. Thus $\mathscr{I} = f_{\omega\theta}^* \overline{\mathscr{I}} = \{f \circ f_{\omega\theta} : f \in \overline{\mathscr{I}}\}$. The ideal describes $\widehat{\Omega}_T = f_{\theta\omega}(\widehat{\Theta}_T)$ as a subset of Ω_T. Let $[n]_{\geq k}$ denote all subsets of $[n]$ with at least k elements. Then the pullback of $\overline{\mathscr{I}}$ satisfies

$$\mathscr{I} = \langle \kappa_1 - \hat{\kappa}_1, \dots, \kappa_n - \hat{\kappa}_n \rangle + \left(\sum_{I \in [n]_{\geq 2}} \langle \kappa_I(\omega) - \hat{\kappa}_I \rangle \right), \tag{11}$$

where $\hat{\kappa}_I$ are the corresponding coordinates of $f_{p\kappa}(q)$. Here the sum of ideals results in another ideal with the generating set which is the sum of generating sets of the summands.

For local computations we use the following reduction.

Proposition 13 (Lin, 2011) *Let $I \subseteq \mathcal{A}_{x_0}(\mathbb{R}^m)$, $J \subseteq \mathcal{A}_{y_0}(\mathbb{R}^n)$ be two ideals. If $\mathrm{RLCT}_{x_0}(I) = (\lambda_x, m_x)$ and $\mathrm{RLCT}_{y_0}(J) = (\lambda_y, m_y)$ then*

$$\mathrm{RLCT}_{(x_0,y_0)}(I+J) = (\lambda_x + \lambda_y, m_x + m_y - 1).$$

Theorem 14 *Let T^r be a rooted tree with n leaves and $q \in \mathcal{M}_T$. Let $\overline{\mathscr{I}}$ be the ideal defined by (8) and $\mathscr{I}$ the ideal defined by (11). Then*

$$\mathrm{RLCT}_{\Theta_T}(\overline{\mathscr{I}}) = \mathrm{RLCT}_{\Omega_T}(\mathscr{I}) = \min_{\omega_0 \in \widehat{\Omega}_T} \mathrm{RLCT}_{\Omega_0}(\mathscr{I}),$$

where Ω_0 is a sufficiently small neighborhood of ω_0 in Ω_T. Moreover, let $\mathscr{J} = \sum_{I \in [n]_{\geq 2}} \langle \kappa_I(\omega) - \hat{\kappa}_I \rangle$. Then, for every $\omega_0 \in \widehat{\Omega}_T$

$$\mathrm{RLCT}_{\omega_0}(\mathscr{I}) = \left(\frac{n}{2}, 0\right) + \mathrm{RLCT}_{\omega_0}(\mathscr{J}). \tag{12}$$

Proof Since $f_{\omega\theta}$ is an isomorphism with a constant Jacobian then the first part of the theorem follows from Proposition 10 (i). Let now W be an ε-box around $\omega_0 = ((s_v^0),(\eta_{uv}^0))$. If T is rooted in an inner leaf then by Proposition 12 the ideal $\mathscr{J}$ does not depend on $s_1,\dots,s_n$. Since for every $i = 1,\dots,n$ the expression $\kappa_i - \hat{\kappa}_i$ depends only on s_i then

$$\mathrm{RLCT}_{\omega_0}(\langle \kappa_1 - \hat{\kappa}_1,\dots,\kappa_n - \hat{\kappa}_n\rangle) = (\frac{n}{2},1),$$

which can be easily checked (see for example Proposition 3.3 of Saito, 2007). Equation (12) follows from Proposition 13.

Now assume that T is rooted in one of the leaves. In this case both $\langle \kappa_1 - \hat{\kappa}_1,\dots,\kappa_n - \hat{\kappa}_n\rangle$ and $\mathscr{J}$ depend on s_r because $\kappa_I(\omega) = (1-s_r^2)\, f_I(\omega)$ for some monomial $f_I(\omega)$ whenever $r \in I$. Therefore, we cannot use Proposition 13 directly. However, by assumption (A2), q lies in the interior of the probability simplex and hence $\hat{\kappa}_i \in (0,1)$ for $i = 1,\dots,n$ which is equivalent to $s_i^0 \in (-1,1)$. Therefore, for each ω_0 one can find two positive constants c, C such that $c \le 1 - s_r^2 \le C$ in W. By Proposition 10 (iv) the real log canonical threshold of $\mathscr{J}$ in W is equal to the real log-canonical threshold of a an ideal with generators induced from the generators of $\mathscr{J}$ by replacing each $1 - s_r^2$ by 1. Now again (12) follows from Proposition 13. ■

5. The Main Reduction Step

Recall that $\hat{\kappa}_{ij} = \mathrm{Cov}(X_i, X_j)$. In this section we prove a technical result which enables us to reduce the computations of $\mathrm{RLCT}_{\omega_0}(\mathscr{J})$ to two simpler cases. First, when q is such that $\hat{\kappa}_{ij} \neq 0$ for all $i, j \in [n]$. Second, when q is such that $\hat{\kappa}_{ij} = 0$ for all $i, j \in [n]$. Moreover, the second case is reduced to computations for monomial ideals which are amenable to various combinatorial techniques.

Let T be a trivalent tree with $n \ge 3$ leaves and let $q \in \mathcal{M}_T$. If all the equivalence classes in $[\widehat{E}]$ are singletones or $[\widehat{E}]$ is empty, which is equivalent to every inner node being of degree at least two in $\widehat{T}$, then Theorem 1 gives us the asymptotic behavior of the marginal likelihood. Thus, let assume that there is at least one class in $[\widehat{E}]$ which is not a singleton. Let $T_1,\dots,T_k$ denote trees representing the equivalence classes in $[\widehat{E}]$ and let $S_1,\dots,S_m$ denote trees induced by the connected components of $E \setminus \widehat{E}$. Let $L_1,\dots L_k$ denote the sets of leaves of $T_1,\dots,T_k$. For each S_i $i = 1,\dots,m$ by Remark 5.2 (iv) of Zwiernik and Smith (2011b) its set of leaves denoted by $[n_i]$ is a subset of $[n]$. For each S_i the number of nodes, edges and nodes of degree 2 in $\widehat{T}$ is denoted by n_v^i, n_e^i and l_2^i respectively. We illustrate this notation in Figure 4 where the dashed edges represent edges in $\widehat{E}$. Simpler examples are given in Figure 3.

Lemma 15 *Let $T = (V,E)$ be a trivalent rooted tree with $n \ge 4$ leaves and let $q \in \mathcal{M}_T$. Let $\mathscr{J} = \sum_{I \in [n]_{\ge 2}} \langle \kappa_I(\omega) - \hat{\kappa}_I\rangle$ as in Theorem 14. If $\omega_0 \in \widehat{\Omega}_T$ then*

$$\mathrm{RLCT}_{\omega_0}(\mathscr{J}) = \sum_{i=1}^{m} \mathrm{RLCT}_{\omega_0}(\mathscr{J}(S_i)) + \sum_{i=1}^{k} \mathrm{RLCT}_{\omega_0}(\mathscr{J}(T_i)) + (0, 1-m-k), \qquad (13)$$

where $\mathscr{J}(S_i) = \sum_{I \in [n_i]_{\ge 2}} \langle \kappa_I(\omega) - \hat{\kappa}_I\rangle$ for $i = 1,\dots,m$ and $\mathscr{J}(T_i) = \sum_{w,w' \in L_i} \langle \kappa_{ww'}(\omega)\rangle$ for $i = 1,\dots,k$.

Proof We first show that $\sum_{I:\hat{\kappa}_I = 0} \langle \kappa_I(\omega)\rangle = \sum_{i,j:\hat{\kappa}_{ij}=0} \langle \kappa_{ij}(\omega)\rangle$. The inclusion "⊇" is clear. We now show "⊆". First note that for every $I \in [n]_{\ge 2}$ if $\hat{\kappa}_I = 0$ then either $\eta_e^0 = 0$ for an edge $e \in E(I)$ or

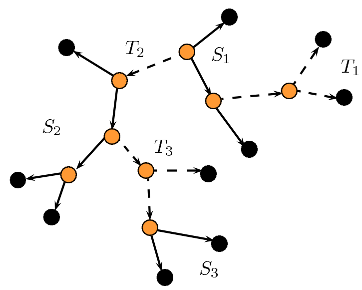


Figure 4: An example of a forest $\widehat{T}$ induced by a point q.

$s^2_{r(I)} = 1$. There exist $i, j \in I$ such that such that $\hat{\kappa}_{ij} = 0$ and the $r(ij) = r(I)$. It follows by Proposition 12 that $\kappa_I(\omega) = \kappa_{ij}(\omega) f(\omega)$ for a polynomial $f(\omega)$ and therefore the inclusion "$\subseteq$" is also true. This implies

$$\mathscr{I} = \sum_{I:\hat{\kappa}_I \neq 0} \langle \kappa_I(\omega) - \hat{\kappa}_I \rangle + \sum_{I:\hat{\kappa}_I = 0} \langle \kappa_I(\omega) \rangle = \sum_{i=1}^{m} \mathscr{I}(S_i) + \sum_{i,j:\hat{\kappa}_{ij}=0} \langle \kappa_{ij}(\omega) \rangle.$$

Hence, to proof the lemma, it suffices to show that for every $\omega_0 \in \widehat{\Omega}_T$

$$\mathrm{RLCT}_{\omega_0}(\sum_{i=1}^{m} \mathscr{I}(S_i) + \sum_{i,j:\hat{\kappa}_{ij}=0} \langle \kappa_{ij}(\omega) \rangle) \tag{14}$$

is equal to the right hand side of (13).

If $e \in E \setminus \widehat{E}$ then by definition there exist $i, j \in [n]$ such that $\hat{\kappa}_{ij} \neq 0$ and $e \in E(ij)$. Since, by Proposition 12, $\hat{\kappa}_{ij} = \eta_e^0 f(\omega_0)$ for a polynomial f then in particular $\eta_e^0 \neq 0$. It follows that for a sufficiently small ε for each $E' \subseteq E \setminus \widehat{E}$ one can find positive constants $c(\varepsilon), C(\varepsilon)$ such that $c(\varepsilon) \leq \prod_{e \in E'} \eta_e \leq C(\varepsilon)$ holds in the ε-box around ω_0. Similarly if $v \notin \widehat{V}$ (cf. Section 3.2) then there exist positive constants $d(\varepsilon), D(\varepsilon)$ such that $d(\varepsilon) \leq (1 - s_v^2) \leq D(\varepsilon)$ in the ε-box around ω_0. It follows by Proposition 10 (iv) that in computations of the real log-canonical threshold in (14) we can replace each $\kappa_{ij}(\omega)$ by

$$(1 - s^2_{r(ij)})^{\delta_{r(ij)}} \prod_{e \in E(ij) \cap \widehat{E}} \eta_e \tag{15}$$

where $\delta_{r(ij)} = 1$ if $r(ij) \in \widehat{V}$ and $\delta_{r(ij)} = 0$ otherwise. Thus, in (14) we can replace the ideal $\sum_{i,j:\hat{\kappa}_{ij}=0} \langle \kappa_{ij}(\omega) \rangle$ by the ideal $\mathscr{I}_1 = \sum_{i,j:\hat{\kappa}_{ij}=0} \langle (1 - s^2_{r(ij)})^{\delta_{r(ij)}} \prod_{e \in E(ij) \cap \widehat{E}} \eta_e \rangle$. However, if we define

$$\mathscr{I}_2 = \sum_{i=1}^{k} \sum_{w,w' \in L_i} \langle (1 - s^2_{r(ww')})^{\delta_{r(ww')}} \prod_{e \in E(ww')} \eta_e \rangle \tag{16}$$

then it can be checked that $\mathscr{I}_1 = \mathscr{I}_2$. To show that $\mathscr{I}_2 \subseteq \mathscr{I}_1$, fix $j = 1, \ldots, k$ and $w, w' \in L_j$, and show that the corresponding generator of $\mathscr{I}_2$ lies in $\mathscr{I}_1$. Note that by construction each of w, w' either has degree two in $\widehat{T}$ or is a leaf of T. Hence, by the definition of $\widehat{E}$, there exist $i, j \in [n]$ such that $E(ij) \cap \widehat{E} = E(ww')$. It follows that each generator in (16) is also in the set of generators of $\mathscr{I}_1$ and hence $\mathscr{I}_2 \subseteq \mathscr{I}_1$. To show the opposite inclusion, note that, if $E(ij)$ intersects with more than one component $T_1, \ldots, T_k$ then the corresponding generator in (15) is a product of some generators in (16) and hence it lies in $\mathscr{I}_2$.

Since the generators of every $\mathscr{J}(S_i)$ for $i=1,\ldots,m$ and every

$$\sum_{w,w'\in L_j}\langle(1-s^2_{r(ww')})^{\delta_{r(ww')}}\prod_{e\in E(ww')}\eta_e\rangle$$

for $j=1,\ldots,k$ involve disjoint sets of variables then by Proposition 13 the term in (14) is equal to

$$\sum_{i=1}^{m}\mathrm{RLCT}_{\omega_0}(\mathscr{J}(S_i))+\sum_{i=1}^{k}\mathrm{RLCT}_{\omega_0}\left(\sum_{w,w'\in L_j}\langle(1-s^2_{r(ww')})^{\delta_{r(ww')}}\prod_{e\in E(ww')}\eta_e\rangle\right)+(0,1-m-k).$$

Again by Proposition 10 (iv) for each $i=1,\ldots,k$

$$\mathrm{RLCT}_{\omega_0}\left(\sum_{w,w'\in L_i}\langle(1-s^2_{r(ww')})^{\delta_{r(ww')}}\prod_{e\in E(ww')}\eta_e\rangle\right)=\mathrm{RLCT}_{\omega_0}(\mathscr{J}(T_i))$$

which finishes the proof. ∎

We note that, by Proposition 8 and the formula in (12) for each S_i:

$$\mathrm{RLCT}(\mathscr{J}(S_i))+\frac{n_i}{2}=\frac{n^i_v+n^i_e-2l^i_2}{2}. \tag{17}$$

6. The Case of Zero Covariances

In this subsection we assume that $q\in\mathcal{M}_T$ is such that $\hat{\kappa}_{ij}=0$ for all $i,j\in[n]$. This implies the full joint marginal independence $X_1\perp\!\!\!\perp\cdots\perp\!\!\!\perp X_n$. The aim is to prove the following proposition.

Proposition 16 *Let T be a trivalent tree with $n\neq 3$ leaves rooted in $r\in V$. Let $q\in\mathcal{M}_T$ be such that $\hat{\kappa}_{ij}=0$ for all $i,j\in[n]$. Let $\mathscr{J}$ be the ideal defined in Theorem 14. Then*

$$\min_{\omega_0\in\widehat{\Omega}_T}\mathrm{RLCT}_{\omega_0}(\mathscr{J})=\left(\frac{n}{4},m\right),$$

where $m=1$ if either r is a leaf of T or r together with all its neighbors are all inner nodes of T. In all other cases we cannot obtain an explicit upper bound for m and hence $m\geq 1$.

The strategy of the proof of Proposition 16 is as follows. First, in Section 6.1, we show that the local computations can be restricted to a special subset of $\widehat{\Omega}_T$ over which $\mathscr{J}$ can be replaced by a monomial ideal. Then, in Section 6.2, we present a method to compute the real log-canonical threshold of a monomial ideal. We use this method in Section 6.3.

6.1 The Deepest Singularity

Note that, by Lemma 15, $\mathrm{RLCT}_{\omega_0}(\mathscr{J})=\mathrm{RLCT}_{\omega_0}(\sum_{i,j\in[n]}\langle\kappa_{ij}(\omega)\rangle)$ so without loss of generality we will assume in this section that $\mathscr{J}=\sum_{i,j\in[n]}\langle\kappa_{ij}(\omega)\rangle$. Moreover, for each $v\in V$ these ideals depend on s_v only through the value of s_v^2. It follows that the computations can be reduced only to points satisfying $s_v\geq 0$ for all inner nodes v of T. Henceforth, in this section, we always assume this is the case. We define the *deepest singularity* of $\widehat{\Omega}_T$ as

$$\widehat{\Omega}_{\mathrm{deep}}:=\{\omega\in\widehat{\Omega}_T:\eta_e=0\text{ for all }e\in\widehat{E},\ s_v=1\text{ for all }v\in\widehat{V}\}.$$

We note that, since $\hat{\kappa}_{ij} = 0$ for all $i, j \in [n]$, then $\widehat{E} = E$ and $\widehat{V}$ is equal to the set of all inner nodes of T. It follows that $\widehat{\Omega}_{\mathrm{deep}}$ is an affine subspace constrained to Ω_T with all coordinates either 0 or 1.

Proposition 17 *Let T be a tree with n leaves. Let $q \in \mathcal{M}_T$ such that $\hat{\kappa}_{ij} = 0$ for all $i, j \in [n]$. Then*

$$\min_{\omega_0 \in \widehat{\Omega}_T} \mathrm{RLCT}_{\omega_0}(\mathscr{J}) = \min_{\omega_0 \in \widehat{\Omega}_{\mathrm{deep}}} \mathrm{RLCT}_{\omega_0}(\mathscr{J}).$$

Proof We first show that $\widehat{\Omega}_T$ is a union of affine subspaces constrained to Ω_T with a common intersection given by $\widehat{\Omega}_{\mathrm{deep}}$. Let $V_0 \subseteq \widehat{V}$ and $E_0 \subseteq \widehat{E}$ and

$$\Omega_{(V_0,E_0)} = \{\omega \in \widehat{\Omega}_T : \ s_v = 1 \text{ for all } v \in V_0, \ \eta_{uv} = 0 \text{ for all } (u,v) \in E_0\}.$$

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

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

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

Each $\Omega_{(V_0,E_0)}$ is an affine subspace in $\mathbb{R}^{|V|+|E|}$, denoted by $M_{(V_0,E_0)}$, constrained to Ω_T. Let $\mathcal{S}$ denote the intersection lattice of all $M_{(V_0,E_0)}$ for (V_0, E_0) minimal with ordering denoted by $\leq$. For each $i \in \mathcal{S}$ let $M^{(i)}$ denote the corresponding intersection and define

$$S_i = M^{(i)} \setminus \bigcup_{j<i} M^{(j)}.$$

In this way we obtain an *$\mathcal{S}$-induced decomposition* of $\mathbb{R}^{|V|+|E|}$ (cf. Section 3.1 in Goresky and MacPherson, 1988).

By (Lazarsfeld, 2004, Example 9.3.17) the function $\omega \mapsto \mathrm{rlct}_\omega(\mathscr{J})$ is lower semicontinuous (the argument used there works over the real numbers). This means that for every $\omega_0 \in \Omega_T$ and $\varepsilon > 0$ there exists a neighborhood U of ω_0 such that $\mathrm{rlct}_{\omega_0}(\mathscr{J}) \leq \mathrm{rlct}_\omega(\mathscr{J}) + \varepsilon$ for all $\omega \in U$. Since the set of values of the real log-canonical threshold is discrete this means that for every $\omega_0 \in \widehat{\Omega}_T$ and any sufficiently small neighborhood W_0 of ω_0, one has $\mathrm{rlct}_{\omega_0}(\mathscr{J}) \leq \mathrm{rlct}_\omega(\mathscr{J})$ for all $\omega \in W_0$. Moreover, $\mathrm{rlct}(\mathscr{J})$ is constant on each S_i. Since for any neighborhood W_0 of $\omega_0 \in \widehat{\Omega}_{\mathrm{deep}}$ we have $W_0 \cap S_i \neq \emptyset$ for all $i \in \mathcal{S}$ then necessarily the minimum of the real log-canonical threshold is attained for a point in the deepest singularity. ■

Proposition 17 shows that in the singular case we can restrict our analysis to the neighborhood of $\widehat{\Omega}_{\mathrm{deep}}$. Often however, we also consider points in a bigger set

$$\widehat{\Omega}_0 = \{\omega \in \widehat{\Omega}_T : \eta_{uv} = 0 \text{ for all } (u,v) \in \widehat{E}\}.$$

Note that $\widehat{\Omega}_{\mathrm{deep}}$ lies on the boundary of Ω_T (cf. (10)) but $\widehat{\Omega}_0$ also contains internal points of Ω_T which will be crucial for some of the arguments later.

We now formulate another technical lemma which enables us to reduce computations to the monomial case.

Lemma 18 *Assume that $q \in \mathcal{M}_T$ is such that $\hat{\kappa}_{ij} = 0$ for all $i, j \in [n]$. Let $\mathscr{J}(\omega_0)$ be the ideal $\mathscr{J}$ translated to the origin. Then for every $\omega_0 \in \widehat{\Omega}_0$*

$$\mathrm{RLCT}_0(\mathscr{J}(\omega_0)) = \mathrm{RLCT}_0(\mathscr{J}'), \tag{18}$$

where $\mathscr{J}'$ is a monomial ideal such that each $\kappa_{ij}(\omega+\omega_0)$ in the set of generators of $\mathscr{J}(\omega_0)$ is replaced either by

$$\begin{array}{ll} s_{r(ij)} \prod_{(u,v)\in E(ij)} \eta_{uv} & \text{if } s^0_{r(ij)} = 1, \text{ or by} \\ \prod_{(u,v)\in E(ij)} \eta_{uv} & \text{if } s^0_{r(ij)} \neq 1. \end{array}$$

Proof Let $i, j \in [n]$ and assume that $\omega_0 = ((s^0_v), (\eta^0_e)) \in \widehat{\Omega}_0$ so that $\eta^0_e = 0$ for all $e \in E$. Then, by Proposition 12:

$$\kappa_{ij}(\omega+\omega_0) = \frac{1}{4}(1-(s_{r(ij)}+s^0_{r(ij)})^2) \prod_{e\in E(ij)} \eta_e. \tag{19}$$

If $s^0_{r(ij)} \neq 1$ for a sufficiently small $\varepsilon > 0$ there exist positive constants $c(\varepsilon)$, $C(\varepsilon)$ such that $c(\varepsilon) < 1-(s_{r(ij)}+s^0_{r(ij)})^2 < C(\varepsilon)$ for $s_{r(ij)} \in (-\varepsilon, \varepsilon)$. Therefore, by Proposition 10 (iv), we can replace this term in (19) with 1. If $s^0_{r(ij)} = 1$ rewrite $1-(1+s^2_{r(ij)})^2$ as $-s_{r(ij)}(2+s_{r(ij)})$. For a sufficiently small ε we can find two positive constants $c(\varepsilon), C(\varepsilon)$ such that $c < 2+s_{r(ij)} < C$ whenever $s_{r(ij)} \in (-\varepsilon, \varepsilon)$. Again, by Proposition 10 (iv), we can replace $2+s_{r(ij)}$ with 1. This proves Equation (18). ■

Since $\mathscr{J}'$ is a monomial ideal then, by (Lin, 2011, Proposition 4.11) and Theorem 20 below, we can compute $\mathrm{RLCT}_0(\mathscr{J}')$ using the method of Newton diagrams. We present this method in the following subsection.

6.2 Newton Diagram Method

Given an analytic function $f \in \mathcal{A}_0(\mathbb{R}^d)$ we pick local coordinates $x = (x_1, \ldots, x_d)$ in a neighborhood of the origin. This allows us to represent f as a 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

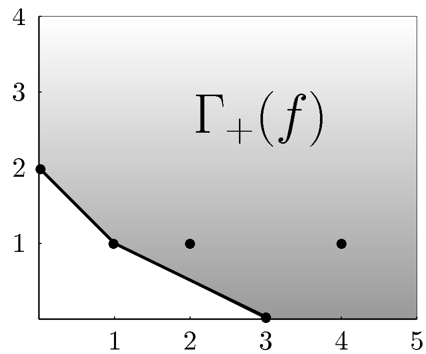
$$\{\alpha+\alpha' : c_\alpha \neq 0, \alpha' \in \mathbb{R}^d_{\geq 0}\}.$$

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

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

If γ is a subset of $\Gamma_+(f)$ then we define $f_\gamma(x) = \sum_{\alpha\in\gamma\cap\mathbb{N}^d} c_\alpha x^\alpha$. The *principal part* of f is, by definition, the sum of all terms of f supported on all compact faces of $\Gamma_+(f)$.

Example 4 *Let $f(x,y) = x^3+2xy+6x^2y+3x^4y+y^2$. Then the Newton diagram looks as follows:*



where the dots correspond to the terms of f. There are only two bounded facets of $\Gamma_+(f)$ and the principal part of f is equal to x^3+xy+y^2.

Definition 19 *The principal part of the power series f with real coefficients is $\mathbb{R}$-nondegenerate if for all compact faces γ of $\Gamma_+(f)$*

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

From the geometric point of view this condition means that the singular locus of the hypersurface defined by $f_\gamma(x)=0$ lies outside of $(\mathbb{R}^*)^n$ for all compact faces γ of $\Gamma_+(f)$.

The following theorem shows that, if the principal part of f is $\mathbb{R}$-nondegenerate and $f\in\mathcal{A}_\Theta^{\geq}$, the computations are greatly facilitated. An example of an application of these methods in statistical analysis can be found in Yamazaki and Watanabe (2004).

Theorem 20 (Arnold et al., 1988) *Let $f\in\mathcal{A}_0^{\geq}(\mathbb{R}^d)$ and $f(0)=0$. If the principal part of f is $\mathbb{R}$-nondegenerate then $\mathrm{RLCT}_0(f)=(\frac{1}{t},c)$ where t is the smallest number such that the vector $(t,\ldots,t)$ hits the polyhedron $\Gamma_+(f)$ and c is the codimension of the face it hits.*

For a proof see Theorem 4.8, Lin (2011).

Let now $f\in\mathcal{A}_{\theta_0}^{\geq}$ such that $f(\theta_0)=0$. We can then center f at θ_0 obtaining a function in $\mathcal{A}_0^{\geq}$. Then we can use Theorem 20 to compute $\mathrm{RLCT}_{\theta_0}(f)$.

Remark 21 *Note that this theorem in general will not give us $\mathrm{RLCT}_{\Theta_0}(f)$ if 0 is a boundary point of Θ in which case we also need to resolve the defining inequalities. For a discussion see (Arnold et al., 1988, Section 8.3.4) and Example 2.7 in Lin (2011).*

6.3 Proof of Proposition 16

Let $n\geq 4$. For each $\omega_0\in\widehat{\Omega}_0$, let $\delta=\delta(\omega_0)\in\{0,1\}^V$ denote the indicator vector satisfying $\delta_v=1$ if $v\in V$ is such that $s_v^0=1$ and $\delta_v=0$ otherwise. In particular $\delta_i=0$ for all $i=1,\ldots,n$ because the leaves, by (A2), are assumed to be non-degenerate. Let $\mathcal{V}_\delta=\mathbb{R}^{n_e+|\delta|}=\mathbb{R}^{|\delta|}\times\mathbb{R}^{n_e}$, where $|\delta|=\sum_v\delta_v$, be the real space with variables representing the edges $(x_e)_{e\in E}$ and nodes (y_v) for all v such that $\delta_v=1$. With some arbitrary numbering of the nodes and edges we order the variables as follows: $y_1\prec\cdots\prec y_{|\delta|}\prec x_{e_1}\prec\cdots\prec x_{e_{n_e}}$. In Lemma 18, for each $\omega_0\in\widehat{\Omega}_0$, we reduced our computations to the analysis of $\mathrm{RLCT}_0(\mathcal{J}')$ where $\mathcal{J}'$ has a simple monomial form. Let Q_δ be a polynomial function on Ω_T defined as a sum of squares of generators of $\mathscr{J}'$. In particular $\mathrm{RLCT}_0(\mathscr{J}') = \mathrm{RLCT}_0(Q_\delta)$. The exponents of terms of the polynomial $Q_\delta(\omega)$ are vectors in $\{0,2\}^{n_e+|\delta|}$. We have that

$$Q_\delta(\omega) = \sum_{i\neq j\in[n]} s_{r(ij)}^{2\delta_{r(ij)}} \prod_{(u,v)\in E(ij)} \eta_{uv}^2. \tag{20}$$

The convex hull, in $\mathcal{V}_\delta$, of the exponents of the terms in Q_δ is called the *Newton polytope* of Q_δ and denoted $\Gamma(Q_\delta)$. We now investigate this polytope which is needed to understand the polyhedron $\Gamma_+(Q_\delta)$, which is needed to use Theorem 20. Since each term of Q_δ corresponds to a path between two leaves then the construction of the Newton polytope $\Gamma(Q_\delta) \subset \mathcal{V}_\delta$ gives a direct relationship between paths in T and the points generating the polytope. Convex combinations of points corresponding to paths give rise to points in the polytope. Let $E_0 \subseteq E$ be the subset of edges of T such that one of the ends is in the set of leaves of T. We call these edges *terminal*. Note that each point generating $\Gamma(Q_\delta)$ satisfies $\sum_{e\in E_0} x_e = 4$. This follows from the fact that each of these points corresponds to a path between two leaves in T and every such a path need to cross exactly two terminal edges. Consequently each point of $\Gamma(Q_\delta)$ needs to satisfy this equation as well. The induced facet of the Newton polyhedron $\Gamma_+(Q_\delta)$ is given as

$$F_0 = \{(\mathbf{y},\mathbf{x}) \in \Gamma_+(Q_\delta) : \sum_{e\in E_0} x_e = 4\} \tag{21}$$

and each point of $\Gamma_+(Q_\delta)$ satisfies $\sum_{e\in E_0} x_e \geq 4$.

The following lemma proves one part of Proposition 16.

Lemma 22 (The real log-canonical threshold of $\mathscr{J}$) *Under assumptions of Proposition 16 we have that* $\mathrm{rclt}_0(\mathscr{J}') = \frac{n}{4}$.

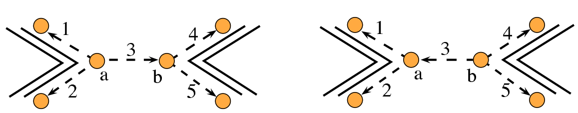
Proof If $n = 2$ then, since $s_1^0, s_2^0 \neq 1$, by Lemma 18 we have that

$$\mathrm{RLCT}_{\omega_0}(\mathscr{J}') = \mathrm{RLCT}_0(\eta_{12}^2) = \left(\frac{1}{2}, 1\right).$$

Therefore Proposition 16 holds in this case. Now assume that $n \geq 4$. By Theorem 20 we have to show that $t = \frac{4}{n}$ is the smallest t such that the vector $(t,\ldots,t)$ hits $\Gamma_+(Q_\delta)$. To show that $\frac{4}{n}\mathbf{1} \in \Gamma_+(Q_\delta)$ we construct a point $q \in \Gamma(Q_\delta)$ such that $q \leq \frac{4}{n}\mathbf{1}$ coordinatewise. The point is constructed as follows.

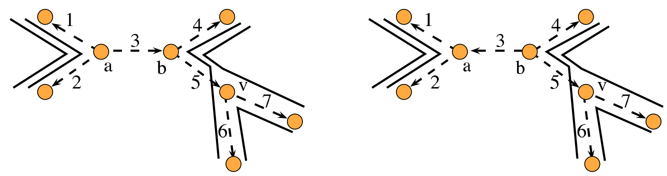
Construction 23 *Let $T = (V,E)$ be a trivalent tree with $n \geq 4$ leaves, rooted in r. We present two constructions of networks of paths between the leaves of T.*

The first construction is for the case when the root is degenerate, $\delta_r = 1$. In this case T is necessarily 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 corresponds to a point in $\Gamma(Q_\delta)$. We order the coordinates of $\mathcal{V}_\delta = \mathbb{R}^{5+|\delta|}$ by $y_a \prec y_b \prec x_1 \prec \cdots \prec x_5$ where y_a, y_b are included only if $\delta_a, \delta_b = 1$. For example the point corresponding to the path involving edges e_1 and e_2 is $(2,0;2,2,0,0,0)$. 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. 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. Now we extend the network by adding e_6 to one of the paths terminating in v and e_7 to the other. Next 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 and hence 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 $\Gamma(Q_\delta)$ denoted by $q_1,\ldots,q_n$. Let $q = \frac{1}{n}\sum_{i=1}^n q_i$ then $q \in \Gamma(Q_\delta)$ is given by $x_{ab} = 0$, $x_e = \frac{4}{n}$ for all $e \in E \setminus (a,b)$. The other coordinates by construction satisfy $y_a = \frac{4}{n}$, $y_b = \frac{4}{n}$ if $\delta_b = 1$, and $y_v = \frac{2}{n}$ for all $v \in V \setminus \{a,b\}$ such that $\delta_v = 1$.

If $\delta_r = 0$ 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. This makes eight paths and each edge is covered exactly four times. The coordinates of the point representing the barycenter of all paths in the network satisfy $x_e = 1$ for all $e \in E$ and $y_v = \frac{1}{2}$ for all v such that $\delta_v = 1$. 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 $q \in \Gamma(Q_\delta)$ with coordinates given by $y_v = \frac{2}{n}$ for all $v \in V$ such that $\delta_v = 1$ and $x_e = \frac{4}{n}$ for $e \in E$. (This finishes the construction.)

The point $\frac{4}{n}\mathbf{1}$ lies in $\Gamma_+(Q_\delta)$, which follows from Construction 23 and the fact that the constructed point $q \in \Gamma(Q_\delta)$ satisfies $q \leq \frac{4}{n}\mathbf{1}$. Moreover, for any $s < \frac{4}{n}$ the point $s(1,\ldots,1)$ does not satisfy $\sum_{e\in E_0} x_e \geq 4$ and hence it cannot be in $\Gamma_+(Q_\delta)$. It follows that $\frac{4}{n}$ is the smallest t such that $t\mathbf{1} \in \Gamma_+(Q_\delta)$ and therefore $\mathrm{rlct}_0(\mathscr{J}') = \frac{n}{4}$. Note that the result does not depend on δ.

■

To compute the multiplicity of the real log-canonical threshold of Q_δ we have to get a better understanding of the polyhedron $\Gamma_+(Q_\delta)$. According to Theorem 20 we need to find the codimension of the face of $\Gamma_+(Q_\delta)$ hit by the vector $\frac{4}{n}\mathbf{1}$. First we find the hyperplane representation of the Newton polytope $\Gamma(Q_\delta)$ reducing the problem to a simpler but equivalent one.

Definition 24 (A pair-edge incidence polytope) *Let $T = (V,E)$ be a trivalent tree with $n \geq 4$ leaves. We define a polytope $P_n \subset \mathbb{R}^{n_e}$, where $n_e = 2n-3$, as the convex combination of points $(q_{ij})_{i,j\in[n]}$ where k-th coordinate of q_{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 in (Mihaescu and Pachter, 2008, Definition 1).*

The reason to study the pair-edge incidence polytope is that its structure can be handled easily and it can be shown to be affinely equivalent to $\Gamma(Q_\delta)$. The latter is immediate if $\delta = (0,\dots,0)$ since $Q_{\mathbf{0}} = 2P_n$. For an arbitrary δ fix a rooting r of T and define a linear map $f_r : \mathbb{R}^{n_e} \to \mathbb{R}^{|\delta|}$ as follows. For each $v \in V \setminus r$ such that $\delta_v = 1$ set

$$y_v = \frac{1}{2}(x_{v\mathrm{ch}_1(v)} + x_{v\mathrm{ch}_2(v)} - x_{\mathrm{pa}(v)v}),$$

where $\mathrm{ch}_1(v)$, $\mathrm{ch}_2(v)$ denotes the two children of v. If $\delta_r = 1$ then set

$$y_r = \frac{1}{2}(x_{r\mathrm{ch}_1(r)} + x_{r\mathrm{ch}_2(r)} + x_{r\mathrm{ch}_3(r)}).$$

The map $(\mathrm{id} \times f_r) : \mathbb{R}^{n_e} \to \mathbb{R}^{n_e} \times \mathbb{R}^{|\delta|}$ satisfies $(\mathrm{id} \times f_r)(2P_n) = \Gamma(Q_\delta)$ because, 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, that is if the path crosses both children of v.

Lemma 25 *Let $P_n \subset \mathbb{R}^{n_e}$ be the pair-edge incidence polytope for a trivalent tree with n leaves where $n \geq 4$. Then $\dim(P_n) = n_e - 1 = 2n - 4$. The unique equation defining the affine span of P_n is $\sum_{e\in E_0} x_e = 2$. 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*

$$x_{e_1(v)} + x_{e_2(v)} - x_{e_3(v)} \geq 0, \quad x_{e_2(v)} + x_{e_3(v)} - x_{e_1(v)} \geq 0, \quad \textit{and } x_{e_3(v)} + x_{e_1(v)} - x_{e_2(v)} \geq 0 \quad \textit{for all } v \in V. \tag{22}$$

Proof Let M_n be the pair-edge incidence matrix, that is a $\binom{n}{2} \times n_e$ matrix with rows corresponding to the points defining P_n. By (Mihaescu and Pachter, 2008, Lemma 1) the matrix has full rank and hence P_n has codimension one in $\mathbb{R}^{n_e}$. 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. This can be checked directly for $n = 4$ using POLYMAKE Gawrilow and Joswig (2005). Assume this is true for all $k < n$. By Q_n we will denote the polyhedron defined by the equation $\sum_{e\in E_0} x_e = 2$ and $3(n-2)$ inequalities given by (22). We want to show that $P_n = Q_n$. It is obvious that $P_n \subseteq Q_n$ since all points generating P_n satisfy the equation and the inequalities. We show that the opposite inclusion also holds.

Consider any cherry $\{e_1, e_2\} \subset E$ in the tree given by two leaves, which we denote by 1, 2, and the separating inner node a. Define a projection $\pi : \mathbb{R}^{n_e} \to \mathbb{R}^{n_e - 2}$ on the coordinates related to all the edges apart from the two in the cherry. We now show that $\pi(Q_n) = \widehat{Q}_{n-1}$, where $\widehat{P} = \mathrm{conv}\{0, P\}$ is a cone with the base given by P. The projection $\pi(Q_n)$ 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

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

Denote the edge incident with e_1, e_2 by e_3 and the related coordinates of x by $x_{e_1}, x_{e_2}, x_{e_3}$. The three inequalities involving x_{e_1} and x_{e_2} do not affect the projection since they imply that

$$\max\{x_{e_1} - x_{e_2}, x_{e_2} - x_{e_1}\} \leq x_{e_3} \leq x_{e_1} + x_{e_2}$$

and hence in particular if $x_{e_1} = x_{e_2}$ the constraint becomes $[0, 2x_{e_1}]$. Consequently the set given by $x_{e_1} + x_{e_2} - x_{e_3} \geq 0$, $x_{e_1} + x_{e_3} - x_{e_2} \geq 0$, $x_{e_2} + x_{e_3} - x_{e_1} \geq 0$ projects down to $\mathbb{R}_{\geq 0}$. However, since $\widehat{Q}_{n-1}$ is contained in the nonnegative orthant, there are no additional constraints on x_{e_3}. Inequalities in (22) 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 $\pi(x)$ for $x \in Q_n$ is a convex combination of the points generating P_{n-1} and zero, that is $\pi(x) = \sum c_{ij} p_{ij}$ where the sum is over all $i \neq j \in \{a, 3, \ldots, n\}$ and $c_{ij} \geq 0$, $\sum c_{ij} \leq 1$. Next, we lift this combination back to Q_n, and show, that any such a lift has to lie in P_n. This would imply that in particular $x \in P_n$. Let y denote a lift of $\pi(x)$ to Q_n. We have

$$y = \sum c_{ij} r_{ij} + \left(1 - \sum c_{ij}\right) r_0,$$

where r_{ij} is a lift of $\pi(p_{ij})$ and r_0 is a lift of the origin. It suffices to show that each r_{ij} and r_0 necessarily lie in P_n.

Consider the following three cases. First, if $p_{ij} \in P_{n-1}$ is such that $x_{e_3} = 0$. Since $P_{n-1} = Q_{n-1}$ and Q_{n-1} satisfy the equation $\sum_{e \in E_0 \setminus \{e_1, e_2\}} x_e + x_{e_3} = 2$, sum of all the other coordinates related to the terminal edges of the smaller tree is 2. Hence, if we lift $\pi(p_{ij})$ to Q_n, then $x_{e_3} = 0$ and

$$x_{e_1} + x_{e_2} \geq 0, \quad x_{e_1} - x_{e_2} \geq 0, \quad x_{e_2} - x_{e_1} \geq 0$$

by plugging $x_{e_3} = 0$ into the three inequalities for the node a. But since $r_{ij} \in Q_n$ 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 $x_{e_1} + x_{e_2} = 0$ and hence $x_{e_1} = x_{e_2} = 0$. Consequently, r_{ij} is a vertex of P_n corresponding to the path between i and j. Second, if p_{ij} is a vertex of P_{n-1} such that $x_{e_3} = 1$, then the sum of all the other coordinates of p_{ij} related to the terminal edges of the smaller tree is 1. Because the lift lies in Q_n we have $x_{e_1} + x_{e_2} = 1$. The additional inequalities give that $x_{e_1}, x_{e_2} \geq 0$. Hence in this case r_{ij} is a convex combination of two points in P_n corresponding to paths terminating in either of the nodes 1 or 2. Finally, we can easily check that zero lifts uniquely to a point in P_n corresponding to the path $E(12)$ joining the leaves 1 and 2. Indeed, from the equation defining Q_n we have $x_{e_1} + x_{e_2} = 2$ and from the inequalities since $x_{e_3} = 0$ we have $x_{e_1} = x_{e_2} = 1$. Therefore every lift y of $\pi(x)$ to Q_n can be written as a convex combination of points generating P_n and hence $y \in P_n$. Consequently $x \in P_n$ and hence $Q_n \subseteq P_n$. ■

Lemma 25 shows that P_n has an extremely simple structure. The inequalities give a polyhedral cone and the equation cuts out the polytope P_n as a slice of this cone. The result gives us also the representation of $\Gamma(Q_\delta)$ in terms of the defining equations and inequalities.

Proposition 26 (Structure of $\Gamma(Q_\delta)$) *Polytope $\Gamma(Q_\delta) \subset \mathcal{V}_\delta$ is given as an intersection of the sets defined by the inequalities in (22) together with $|\delta| + 1$ equations given by*

$$\begin{array}{ll} 2y_v = x_{v\mathrm{ch}_1(v)} + x_{v\mathrm{ch}_2(v)} - x_{\mathrm{pa}(v)v} & \text{for all } v \neq r \text{ such that } \delta_v = 1, \\ 2y_r = x_{r\mathrm{ch}_1(r)} + x_{r\mathrm{ch}_2(r)} + x_{r\mathrm{ch}_3(r)} & \text{if } \delta_r = 1, \text{ and} \\ \sum_{e \in E_0} x_e = 4. & \end{array} \tag{23}$$

From this we can partially understand the structure of $\Gamma_+(Q_\delta)$. First note that $\Gamma_+(f) = \Gamma(f) + \mathbb{R}^d_{\geq 0}$, where the plus denotes the Minkowski sum. The *Minkowski sum* of two polyhedra is by definition

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

Lemma 27 *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 $c \geq 0$.*

Now we are ready to compute multiplicities of the real log-canonical threshold $\mathrm{RLCT}_0(Q_\delta)$ at least in certain cases. This completes the proof of Proposition 16.

Lemma 28 (Computing multiplicities) *Let T be a trivalent tree with $n \geq 4$ leaves, rooted in r. Let $q \in \mathcal{M}_T$ be such that $\hat{\kappa}_{ij} = 0$ for all $i, j \in [n]$ and $\omega_0 \in \widehat{\Omega}_0$. Let $\delta = \delta(\omega_0)$ be such that $\delta_v = 1$ if $s_v^0 = 1$ and it is zero otherwise. Define $Q_\delta(\omega)$ as in (20). If either: (i) $\delta_r = 0$ or (ii) $\delta_r = 1$ and $\delta_v = 1$ for all $(r, v) \in E$ then* $\mathrm{mult}_0(Q_\delta) = 1$.

Proof 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 $\Gamma_+(Q_\delta)$ is decomposed as a face of the standard cone $\mathbb{R}^{n_e+|\delta|}_{\geq 0} \subset \mathcal{V}_\delta$ plus a face of $\Gamma(Q_\delta)$. We say that a face of $\Gamma(Q_\delta)$ induces a facet of $\Gamma_+(Q_\delta)$ if there exists a face of the standard cone $\mathbb{R}^{n_e+|\delta|}_{\geq 0}$ such that the Minkowski sum of these two faces gives a facet of $\Gamma_+(Q_\delta)$. Since the dimension $\Gamma(Q_\delta)$ is lower than the dimension of the resulting polyhedron it turns out that one face of $\Gamma(Q_\delta)$ can induce more than one facet of $\Gamma_+(Q_\delta)$. In particular $\Gamma(Q_\delta)$ itself induces more than one facet where one of them is F_0 given by (21).

Every facet of $\Gamma_+(Q_\delta)$ containing the point $\frac{4}{n}\mathbf{1}$, after normalizing the coefficients to sum to n, that is $\sum_v \alpha_v + \sum_e \beta_e = n$, is of the form

$$\sum_v \alpha_v y_v + \sum_e \beta_e x_e \geq 4, \tag{24}$$

where by Lemma 27 we can assume that $\alpha_v, \beta_e \geq 0$. Our approach can be summarized as follows. Using Construction 23 we provide coordinates of a point $p \in \Gamma(Q_\delta)$ such that $\frac{4}{n}\mathbf{1}$ lies on the boundary of $p + \mathbb{R}^{n_e+|\delta|}_{\geq 0}$. Then $\frac{4}{n}\mathbf{1}$ can only lie on faces of $\Gamma_+(Q_\delta)$ induced by faces of $\Gamma(Q_\delta)$ containing p. To show that the multiplicity is exactly 1 we need to show that $\frac{4}{n}\mathbf{1}$ lies in the interior of F_0.

First, assume that $\delta_r = 0$ which corresponds to the case when the root r represents a non-degenerate random variable. Consider the point $p \in \Gamma(Q_\delta)$ induced by the network of $2n$ paths given in the second part of Construction 23. Since $x_e = \frac{4}{n}$ for all $e \in E$ then from the description of $\Gamma(Q_\delta)$ in Lemma 26 we can check that all defining inequalities are strict for this point. Therefore p lies in the interior of $\Gamma(Q_\delta)$ and the only facets of $\Gamma_+(Q_\delta)$ containing p are these induced by $\Gamma(Q_\delta)$ itself. The equation defining a facet induced by $\Gamma(Q_\delta)$ has to be obtained as a combination of the defining equations: $\sum_{e \in E_0} x_e = 4$ and $|\delta|$ equations

$$2y_v - x_{v\mathrm{ch}_1(v)} - x_{v\mathrm{ch}_2(v)} + x_{\mathrm{pa}(v)v} = 0 \tag{25}$$

for all $v \in V$ such that $\delta_v = 1$. We check possible combinations such that the form of the induced inequality in (24) is attained. The first inequality, defining F_0, is already of this form (cf. (21)). The sum of all the 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 (24) is 4) a non-negative (since the coefficients in front of y_v need to be non-negative) combination of equations in (25). However, since the sum of the coefficients in (25) is $+1$, this contradicts the assumption that the sum of coefficients in the defining inequality is n. Consequently, if $\delta_r = 0$ the codimension of the face hit by $\frac{4}{n}\mathbf{1}$ is 1 and hence by Theorem 20 we have that $\mathrm{mult}_0(Q_\delta) = 1$.

Second, if $\delta_r = 1$ and $\delta_v = 1$ for all children of r in T then since all the nodes adjacent to r (denote them by a, b, c) are inner we have three different ways of conducting the construction of the n-path network in Construction 23 (by omitting each of the incident edges). Hence we get three different points and their barycenter satisfies $x_{ra} = x_{rb} = x_{rc} = \frac{8}{3n}$ and $x_e = \frac{4}{n}$ for all the other edges; $y_r = \frac{4}{n}$, $y_a = y_b = y_c = \frac{8}{3n}$ and $y_v = \frac{2}{n}$ for all the other inner nodes. Denote this point by p and note that $p \leq \frac{4}{n}\mathbf{1}$. By the facet description of $\Gamma(Q_\delta)$ derived in Proposition 26 we can check that this point cannot lie in any of the facets defining $\Gamma(Q_\delta)$ and hence it is an interior point of the polytope. As in the first case it means that the facets of $\Gamma_+(Q_\delta)$ containing p are induced by $\Gamma(Q_\delta)$. By Proposition 26 the affine span is given by (23). Since the sum of coefficients in the equation involving y_r 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 (25) each with coefficient $t_v \geq 0$ and then add the equation in (23) involving y_r 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 $t_a - \sum_{v \neq r} t_v = -\sum_{v \neq r,a} t_v$. Since it has to be 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 $t_v = 0$ for all inner nodes v. Consequently the only possible facet of $\Gamma_+(Q_\delta)$ containing $\frac{4}{n}\mathbf{1}$ is F_0 and hence again $\mathrm{mult}_0(Q_\delta) = 1$. ■

The following example shows that in certain cases $\mathrm{mult}_0(Q_\delta)$ can be strictly greater than 1.

Example 5 *Consider the quartet tree model with q such that $\hat{\kappa}_{ij} = 0$ for all $i, j = 1,2,3,4$. In this case $\Gamma(Q_\delta) \subseteq \mathbb{R}^7$ has six vertices $(2,0;2,2,0,0,0)$, $(2,0;2,0,2,2,0)$, $(2,0;2,0,2,0,2)$, $(2,0;2,0,2,2,0)$, $(2,0;0,2,2,0,2)$ and $(0,2;0,0,0,2,2)$. The facet description of the Newton polyhedron $\Gamma_+(Q_\delta)$ can be easily computed using* POLYMAKE *Gawrilow and Joswig (2005). From this description it is easily checked that the point $(1,1;1,1,1,1,1)$ lies on two facets of $\Gamma_+(Q_\delta)$. It follows that the codimension of the face hit by this vector is two, or equivalently,* $\mathrm{mult}_0(Q_\delta) = 2$.

7. Proof of Theorem 2

In this section we complete the proof of Theorem 2 using results from the previous sections. We split it into three steps.

7.1 Step 1

To analyze the asymptotic behavior of the stochastic complexity F_N, by Theorem 4, equivalently we can compute $\mathrm{RLCT}_{\Theta_T}(K;\varphi)$, where K is the Kullback-Leibler distance defined in (3) and φ is the prior distribution satisfying (A1). By Theorem 11 and Theorem 14 this real log-canonical threshold is equal to $\mathrm{RLCT}_{\Omega_T}(\mathscr{I})$, where $\mathscr{I}$ is the ideal defined by (11).

7.2 Step 2

We compute separately $\mathrm{RLCT}_{\Omega_T}(\mathscr{I})$ in the case when $n=3$. If T is rooted in the inner node the expansion for $\mathbb{E}F_N$ follows from Theorem 4 in Rusakov and Geiger (2005). Thus if $\widehat{E}=E$, which in Rusakov and Geiger (2005) corresponds to the type 2 singularity, then

$$\mathbb{E}F_N = NS + 2\log N + O(1) \quad \text{or} \qquad \mathrm{RLCT}_{\Omega_T}(\mathscr{I}) = (2,1). \tag{26}$$

Since all the neighbours of the root are leaves and hence, by (A2), they are non-degenerate we need only to make sure that the first equation in Theorem 2 gives (26). This follows from the fact that $l_2=0$ and $l_0=1$, where l_i $i=0,1,2,3$ defined in the introduction is the number of *inner* nodes of T whose degree in $\widehat{T}$ is i. In the case when $|\widehat{E}|=1$ (type 1 singularity) we have

$$\mathbb{E}F_N = NS + \frac{5}{2}\log N + O(1) \quad \text{or} \qquad \mathrm{RLCT}_{\Omega_T}(\mathscr{I}) = \left(\frac{5}{2},1\right).$$

The second equation in Theorem 2 holds since $l_2=1$, $l_0=0$ and $c=0$. If $\widehat{E}=\emptyset$ we have

$$\mathbb{E}F_N = NS + \frac{7}{2}\log N + O(1) \quad \text{or} \qquad \mathrm{RLCT}_{\Omega_T}(\mathscr{I}) = \left(\frac{7}{2},1\right),$$

which again is true since $l_2=0$, $l_0=0$ and $c=0$.

Now assume that T is rooted in a leaf, say 1. If there exists $i,j=1,2,3$ such that $\hat{\kappa}_{ij}\neq 0$ (or equivalently $|\widehat{E}|\leq 1$) then $\widehat{V}=\emptyset$ and by Proposition 8

$$\mathbb{E}F_N = NS + \frac{7-2l_2}{2} + O(1) \quad \text{or} \qquad \mathrm{RLCT}_{\Omega_T}(\mathscr{I}) = \left(\frac{7-2l_2}{2},1\right).$$

If $\widehat{E}=E$ then $\widehat{V}\neq\emptyset$ and by Theorem 14 for every $\omega_0\in\widehat{\Omega}_T$

$$\mathrm{RLCT}_{\omega_0}(\mathscr{I}) = \left(\frac{3}{2},0\right) + \mathrm{RLCT}_{\omega_0}(\mathscr{J}).$$

Moreover, by Lemma 18, for every $\omega_0\in\widehat{\Omega}_0$

$$\mathrm{RLCT}_{\omega_0}(\mathscr{J}) = \mathrm{RLCT}_0(\langle \eta_{1h}\eta_{h2}, \eta_{1h}\eta_{h3}, s_h^{\delta_h}\eta_{h2}\eta_{h3}\rangle),$$

where $\delta_h=1$ if $s_h^0=1$ and $\delta_h=1$ otherwise. It can be checked directly by using the Newton diagram method and Theorem 20 that $\mathrm{RLCT}_{\omega_0}(\mathscr{J})=(\frac{3}{4},1)$ both if $\delta_h=0$ and $\delta_h=1$ and hence $\mathrm{RLCT}_{\omega_0}(\mathscr{I})=(\frac{9}{4},1)$. Since the points in $\widehat{\Omega}_0$ such that $s_h^0\neq 1$ lie in the interior of Ω_T then for these points $\mathrm{RLCT}_{\omega_0}(\mathscr{I})=\mathrm{RLCT}_{\Omega_0}(\mathscr{I})$ where Ω_0 is a neighborhood of ω_0 in Ω_T. Hence, by (6), we have that

$$\mathrm{RLCT}_{\Omega_T}(\mathscr{I}) = \min_{\omega_0\in\widehat{\Omega}_T}\mathrm{RLCT}_{\Omega_0}(\mathscr{I}) \leq \min_{\omega_0\in\widehat{\Omega}_0}\mathrm{RLCT}_{\Omega_0}(\mathscr{I}) = \left(\frac{9}{4},1\right).$$

On the other hand, by (5) and then Proposition 17, we obtain the following inequalities

$$\mathrm{RLCT}_{\Omega_T}(\mathscr{I}) \geq \min_{\omega_0\in\widehat{\Omega}_T}\mathrm{RLCT}_{\omega_0}(\mathscr{I}) \geq \min_{\omega_0\in\widehat{\Omega}_{\mathrm{deep}}}\mathrm{RLCT}_{\omega_0}(\mathscr{I}) = \left(\frac{9}{4},1\right).$$

It follows that

$$\mathbb{E}F_N = NS + \frac{9}{4}\log N + O(1) \quad \text{or} \qquad \mathrm{RLCT}_{\Omega_T}(\mathscr{I}) = \left(\frac{9}{4},1\right),$$

which gives the the second equation in Theorem 2 since in this case $l_2=c=0$ and $l_0=1$.

7.3 Step 3, Case 1

Assume now that $n \geq 4$ and $r \notin \widehat{V}$. In this case, using notation from Section 5, every T_i for $i = 1, \ldots, k$ is rooted in one of its leaves. Hence $\mathrm{RLCT}_{\omega_0}(\mathscr{J}(T_i)) = (\frac{|L_i|}{4}, 1)$ for every $i = 1, \ldots k$. If $|L_i| \neq 3$ this follows from Proposition 16. If $|L_i| = 3$ it follows from Case 2 above. By Lemma 15 and Equation (17), for every $\omega_0 \in \widehat{\Omega}_0$ we have that

$$\mathrm{rlct}_{\omega_0}(\mathscr{I}) = \frac{n}{2} + \sum_{i=1}^{m} \frac{n_v^i + n_e^i - n_i - 2l_2^i}{2} + \sum_{i=1}^{k} \frac{|L_i|}{4},$$

where n_v^i, n_e^i, l_2^i are respectively the number of vertices, edges and and degree two nodes in $\widehat{T}$ of S_i; and L_i is the set of leaves of T_i. Let m_i denote the number of nodes of $\widehat{T}$ whose degree is i. Note that $m_2 = l_2$ but m_0 does not necessarily equal l_0. We now use three simple formulas: $\sum_i n_v^i = m_1 + m_2 + m_3$ (that is only degree zero nodes of $\widehat{T}$ do not lie in the S_i's), $\sum_i n_e^i = |E \setminus \widehat{E}|$ (that is $E \setminus \widehat{E}$ is the set of all edges of all the S_i's) and $\sum_i |L_i| = m_2 + n - m_1$ (that is the leaves of all the T_i's are precisely the degree two nodes of $\widehat{T}$ and these leaves of T which have degree zero in $\widehat{T}$). Moreover, for any graph with the vertex set V and the edge set E, $\sum_{v \in V} \deg(v) = 2n_e$ (see Semple and Steel, 2003, Corollary 1.2.2). Therefore, with the formula applied for the forest $\widehat{T}$, we have $m_1 + 2m_2 + 3m_3 = 2|E \setminus \widehat{E}|$. Using these four formulas together we show that $\mathrm{rlct}_{\omega_0}(\mathscr{I}) = \frac{1}{4}(3n + m_2 + 5m_3)$. The final formula for the coefficient follows from the fact that $l_2 = m_2$ and $l_0 = n_v - n - m_2 - m_3$. Moreover, since $\delta_r = 0$ for all $\omega_0 \in \widehat{\Omega}_0$ then, by Lemma 28, $\mathrm{mult}_0(\mathscr{J}(T_i)) = 1$ for every $\omega_0 \in \widehat{\Omega}_0$. Therefore,

$$\mathrm{RLCT}_{\omega_0}(\mathscr{I}) = \left(\frac{n_v + n_e - 2l_2}{2} - \frac{5l_0}{4}, 1 \right). \tag{27}$$

Now we show that $\mathrm{RLCT}_{\Omega_T}(\mathscr{I})$ also has the same form. Let ω_2 be a point in $\widehat{\Omega}_0$ such that $s_v \neq 1$ for all $v \in V$ and let $\omega_1 \in \widehat{\Omega}_{\mathrm{deep}}$. Equation (27) is true both if $\omega_0 = \omega_1$ and $\omega_0 = \omega_2$ and hence $\mathrm{RLCT}_{\omega_1}(\mathscr{I}) = \mathrm{RLCT}_{\omega_2}(\mathscr{I})$. However, since ω_2 is an inner point of Ω_T, it follows from the definition of $\mathrm{RLCT}_{\Omega_T}(\mathscr{I})$ as the minimum over all points in Ω_T, that

$$\mathrm{RLCT}_{\Omega_T}(\mathscr{I}) \leq \mathrm{RLCT}_0(\mathscr{I}_{\omega_2}).$$

On the other hand by (5) and Proposition 17

$$\mathrm{RLCT}_0(\mathscr{I}_{\omega_1}) = \min_{\omega_0 \in \widehat{\Omega}_T} \mathrm{RLCT}_0(\mathscr{I}_{\omega_0}) \leq \min_{\omega_0 \in \widehat{\Omega}_T} \mathrm{RLCT}_{\Omega_0}(\mathscr{I}_{\omega_0}) = \mathrm{RLCT}_{\Omega_T}(\mathscr{I}).$$

Therefore, if $r \notin \widehat{V}$, then in fact $\mathrm{RLCT}_{\Omega_T}(\mathscr{I}) = (\lambda, 1)$, where λ is the coefficient in (27), and

$$\mathbb{E}F_N = NS + \lambda \log N + O(1).$$

The main formula in Theorem 2 is proved in this case because $c = 0$.

7.4 Step 3, Case 2

Let now $n \geq 4$ and $r \in \widehat{V}$. Let $1 \leq j \leq k$ be such that r is an inner node of T_j and $\omega_0 \in \widehat{\Omega}_0$. For all $i \neq j$, T_i is rooted in one of its leaves. Therefore, by Lemma 22, Lemma 28 and Step 2 above for all $i \neq j$ we have that $\mathrm{RLCT}_{\omega_0}(\mathscr{J}(T_i)) = (|L_i|/4, 1)$. It remains to compute $\mathrm{RLCT}_{\omega_0}(\mathscr{J}(T_j))$. If $|L_j| = 3$ then $\mathrm{RLCT}_{\omega_0}(\mathscr{J}(T_j)) = (1/2, 1) = ((|L_j| - 1)/4, 1)$ by the Step 2 above, (cf. (26)). In this case the computations are the same as in Step 3, Case 1 but with a difference of $\frac{1}{4}$ in the real log-canonical threshold. We obtain

$$\mathbb{E}F_N = NS + \left(\frac{n_v + n_e - 2l_2}{2} - \frac{5l_0 + 1}{4}\right)\log N + O(1).$$

However, if $|L_j| \geq 4$ then, by Lemma 22, $\mathrm{rlct}_0(\mathscr{J}(T_j)) = |L_j|/4$ and hence as in Step 3, Case 1 we have $\sum_{i=1}^{k} \mathrm{rlct}_0(\mathscr{J}_{\omega_0}(T_i)) = \frac{1}{4}(n - m_1 + m_2)$. Therefore $\mathrm{rlct}_{\Omega_T}(\mathscr{I}) = \lambda$. We compute the multiplicity by considering different subcases. If all the neighbours of r are degenerate then for all points $\omega_0 \in \widehat{\Omega}_{\mathrm{deep}}$ we have that $\delta_r = 1$ and $\delta_v = 1$ for all neighbours v or r. It follows from Lemma 28 that $\mathrm{mult}_{\omega_0}(\mathscr{J}(T_j)) = 1$ and hence $\mathrm{mult}_{\Omega_T}(\mathscr{I}) = 1$. Therefore,

$$\mathbb{E}F_N = NS + \frac{1}{4}(3n + l_2 + 5l_3)\log N + O(1).$$

Otherwise we do not have explicit bounds on the multiplicity. Since $\mathrm{mult}_{\Omega_T}(\mathscr{I}) \geq 1$ then

$$\mathbb{E}F_N = NS + \frac{1}{4}(3n + l_2 + 5l_3)\log N - (m-1)\log\log N + O(1),$$

where $m \geq 1$. This finishes the proof of Theorem 2. □

Remark 29 *Example 5 showed that* $\mathrm{mult}_{\omega_0}(\mathscr{I})$ *may be strictly greater than* 1 *for some boundary points of* Ω_T. *The analysis of how it affects the computation of* $\mathrm{mult}_{\Omega_T}(\mathscr{I})$ *is highly complicated as it involves resolution of the boundary constraints. Typically we are just able to provide upper bounds. For example, since in Example 5 we have* $\mathrm{mult}_{\Omega_0}(\mathscr{I}) = \mathrm{mult}_{\omega_0}(\mathscr{I})$ *then, by (5),* $\mathrm{mult}_{\Omega_0}(\mathscr{I}) \leq \mathrm{mult}_{\omega_0}(\mathscr{I}) = 2$.

Acknowledgments

I am especially grateful to Shaowei Lin for a number of illuminating discussions and introducing me to the theory of log-canonical thresholds. I also want to thank Diane Maclagan and the referees for helpful comments.

References

Vladimir I. Arnold, Sabir M. Guseĭn-Zade, and Aleksandr N. Varchenko. *Singularities of Differentiable Maps*, volume II. Birkhäuser, 1988.

Edward Bierstone and Pierre D. Milman. Semianalytic and subanalytic sets. *Publications Mathématiques de l'IHÉS*, 67(1):5–42, 1988.

David Maxwell Chickering and David Heckerman. Efficient approximations for the marginal likelihood of Bayesian networks with hidden variables. *Machine Learning*, 29(2):181–212, 1997.

Gregory F. Cooper and Edward Herskovits. A Bayesian method for the induction of probabilistic networks from data. *Machine Learning*, 9(4):309–347, 1992.

Ewgenij Gawrilow and Michael Joswig. Geometric reasoning with polymake. `arXiv:math.CO/0507273`, 2005.

Mark Goresky and Robert MacPherson. *Stratified Morse Theory*, volume 14 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]*. Springer-Verlag, Berlin, 1988. ISBN 3-540-17300-5.

Dominique Haughton. On the choice of a model to fit data from an exponential family. *Ann. Statist*, 16(1):342–355, 1988.

David Heckerman, Dan Geiger, and David Maxwell Chickering. Learning Bayesian networks: The combination of knowledge and statistical data. *Machine Learning*, 20(3):197–243, 1995.

Robert Lazarsfeld. *Positivity in Algebraic Geometry*. A Series of Modern Surveys in Mathematics. Springer Verlag, 2004.

Shaowei Lin. Asymptotic Approximation of Marginal Likelihood Integrals. `arXiv:1003.5338`, November 2011. submitted.

Radu Mihaescu and Lior Pachter. Combinatorics of least-squares trees. *Proceedings of the National Academy of Sciences of the United States of America*, 105(36):13206, 2008.

Judea Pearl and Michael Tarsi. Structuring causal trees. *J. Complexity*, 2(1):60–77, 1986. ISSN 0885-064X. Complexity of approximately solved problems (Morningside Heights, N.Y., 1985).

Dmitry Rusakov and Dan Geiger. Asymptotic model selection for naive Bayesian networks. *J. Mach. Learn. Res.*, 6:1–35 (electronic), 2005. ISSN 1532-4435.

Morihiko Saito. On real log canonical thresholds. `arXiv:0707.2308`, 2007.

Gideon Schwarz. Estimating the dimension of a model. *Annals of Statistics*, 6(2):461–464, 1978.

Charles Semple and Mike Steel. *Phylogenetics*, volume 24 of *Oxford Lecture Series in Mathematics and its Applications*. Oxford University Press, Oxford, 2003. ISBN 0-19-850942-1.

Sumio Watanabe. *Algebraic Geometry and Statistical Learning Theory*. Number 25 in Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press, 2009. ISBN-13: 9780521864671.

Keisuke Yamazaki and Sumio Watanabe. Newton diagram and stochastic complexity in mixture of binomial distributions. In *Algorithmic Learning Theory*, volume 3244 of *Lecture Notes in Comput. Sci.*, pages 350–364. Springer, Berlin, 2004.

Piotr Zwiernik and Jim Q. Smith. Implicit inequality constraints in a binary tree model. *Electron. J. Statist.*, 5:1276–1312, 2011a. ISSN 1935-7524. doi: 10.1214/11-EJS640.

Piotr Zwiernik and Jim Q. Smith. Tree-cumulants and the geometry of binary tree models. *to appear in Bernoulli*, 2011b.