



L-cumulants, *L*-cumulant embeddings and algebraic statistics

Piotr Zwiernik

Department of Statistics, University of California, Berkeley, CA 94720, USA

Abstract. Focusing on the discrete probabilistic setting we generalize the combinatorial definition of cumulants to *L*-cumulants. This generalization keeps all the desired properties of the classical cumulants like semi-invariance and vanishing for independent blocks of random variables. These properties make *L*-cumulants useful for the algebraic analysis of statistical models. We illustrate this for general Markov models and hidden Markov processes in the case when the hidden process is binary. The main motivation of this work is to understand cumulant-like coordinates in algebraic statistics and to give a more insightful explanation why tree cumulants give such an elegant description of binary hidden tree models. Moreover, we argue that *L*-cumulants can be used in the analysis of certain classical algebraic varieties.

2000 Mathematics Subject Classifications: Primary 62E99; Secondary 60C05, 46L54

Key Words and Phrases: Conditional independence models, discrete random variables, cumulants, free cumulants, Boolean cumulants, tree cumulants, central moments

1. Introduction

Although moments provide a convenient summary of properties of a probability distribution, it was observed that these properties can generally be described in a simpler way using cumulants (see for example [2, Section 2.4], [11, Chapter 2]). This is mainly because cumulants have the ability to capture symmetries and underlying independencies of a probability distribution. These striking features of cumulants make them an interesting object of statistical study both from a theoretical and practical point of view. In addition, as it was shown for example in [5, 16, 18], cumulants and moments can be used to analyze the geometry of statistical models.

Recently, in [27] we have suggested using a less standard system of coordinates which we called *tree cumulants*. This new coordinate system proved to be useful to analyze Bayesian networks on trees when some of the nodes are not observed. Various results on identifiability and geometry of these models have been obtained in [25, 26, 27], which encouraged us to study more general coordinate systems like that. In the present paper we propose a useful generalization of both cumulants and tree cumulants.

Email address: piotr.zwiernik@gmail.com (P. Zwiernik)

We work in a simple probabilistic setting. Let $X = (X_1, \dots, X_n)$ be a random vector such that each X_i takes $r_i \geq 2$ possible values, where each r_i is finite. The vector X takes values in a finite discrete set $\mathcal{X} = \prod_{i=1}^n \mathcal{X}_i \subseteq \mathbb{R}^n$ such that $|\mathcal{X}_i| = r_i$ for $i = 1, \dots, n$. Without loss of generality we set

$$\mathcal{X} = \{0, \dots, r_1 - 1\} \times \dots \times \{0, \dots, r_n - 1\}.$$

Any probability distribution of X can be written as a point $P = [p(x)] \in \mathbb{R}^{\mathcal{X}}$ such that $p(x) \geq 0$ for all $x \in \mathcal{X}$ and $\sum_{x \in \mathcal{X}} p(x) = 1$. The set of all such points is called the *probability simplex* and it is denoted by $\Delta_{\mathcal{X}}$.

For any function $f : \mathcal{X} \rightarrow \mathbb{R}$ the *expectation of $f(X)$* is given by

$$\mathbb{E}[f(X)] := \sum_{x \in \mathcal{X}} p(x)f(x).$$

Let $[n] := \{1, \dots, n\}$ and for any multiset $A = \{i_1, \dots, i_d\}$ of elements of $[n]$ let

$$X_A = (X_{i_1}, \dots, X_{i_d}).$$

In a similar way we define $x_A = (x_{i_1}, \dots, x_{i_d})$ and $\mathcal{X}_A = \mathcal{X}_{i_1} \times \dots \times \mathcal{X}_{i_d}$. For each such a multiset A we define the corresponding *moment*

$$\mu_A = \mathbb{E}[X_{i_1} \dots X_{i_d}]$$

and the *central moment*

$$\mu'_A = \mathbb{E}[(X_{i_1} - \mu_{i_1}) \dots (X_{i_d} - \mu_{i_d})].$$

Our convention is to write μ_A as $\mu_{i_1 \dots i_d}$, where $i_1 \leq \dots \leq i_d$. So for example if $A = \{1, 2, 4, 4, 4\}$, the corresponding moment is written as $\mu_{12444} = \mathbb{E}[X_1 X_2 X_4^3]$. The same convention applies to central moments. In particular, for every $i < j$, μ'_{ij} is the covariance between X_i and X_j .

To show how cumulants can be naturally generalized we first define them formally and then we discuss their basic properties. Cumulants are usually computed using the cumulant generating function, which is defined as the logarithm of the moment generating function. In this paper we use an alternative definition of cumulants using partitions (see for example [11, 15, 19]). We say that $\pi = B_1 | \dots | B_k$ is a *partition* (or a *set partition*) of $[n]$, if the *blocks* $B_i \neq \emptyset$ are disjoint sets whose union is $[n]$. A partition is called a *split* if it consists of two blocks. Let $\Pi([n])$ be the set of all set partitions of $[n]$. The *cumulant of the vector X* is defined as

$$k_{1 \dots n} = \sum_{\pi \in \Pi([n])} (-1)^{|\pi|-1} (|\pi| - 1)! \prod_{B \in \pi} \mu_B, \tag{1}$$

where the sum is over all set partitions of $[n]$, the product is over all blocks of a partition and $|\pi|$ denotes the number of blocks of π . For example, if $n = 3$ then there are five partitions in $\Pi([3])$: 123, 1|23, 2|13, 12|3 and 1|2|3 and (1) gives

$$k_{123} = \mu_{123} - \mu_1 \mu_{23} - \mu_2 \mu_{13} - \mu_{12} \mu_3 + 2\mu_1 \mu_2 \mu_3. \tag{2}$$

Equation (1) can be generalized for any multiset $A = \{i_1, \dots, i_d\}$ of elements of $[n]$ to obtain the cumulant of X_A . We use the bijection between A and $[d]$ and write

$$k_A = \sum_{\pi \in \Pi([d])} (-1)^{|\pi|-1} (|\pi| - 1)! \prod_{B \in \pi} \mu_{i_B}, \tag{3}$$

where $i_B = \{i_j : j \in B\}$. Hence for instance

$$k_{112} = \mu_{112} - 2\mu_1\mu_{12} - \mu_{11}\mu_2 + 2\mu_1^2\mu_2.$$

For each $x = (x_1, \dots, x_n) \in \mathcal{X}$ define a multiset $\mathcal{A}(x)$ as

$$\mathcal{A}(x) = \underbrace{\{1, \dots, 1\}}_{x_1 \text{ times}}, \dots, \underbrace{\{n, \dots, n\}}_{x_n \text{ times}} \tag{4}$$

and let $\mathcal{A}(\mathcal{X}) = \{\mathcal{A}(x) : x \in \mathcal{X}\}$. By the *moment aliasing principle* (see [12, Lemma 3]) there exists a polynomial isomorphism between $P = [p(x)]_{x \in \mathcal{X}}$ and two other systems of coordinates of $\mathbb{R}^{\mathcal{A}(\mathcal{X})} \simeq \mathbb{R}^{\mathcal{X}}$ given by moments $M = [\mu_{\mathcal{A}(x)}]_{x \in \mathcal{X}}$ and by cumulants $K = [k_{\mathcal{A}(x)}]_{x \in \mathcal{X}}$. In particular, every model $\mathcal{M} \subseteq \Delta_{\mathcal{X}}$, after a change of coordinates, can be equivalently expressed in terms of M or K .

In our discussion of cumulants the central concept is that of independence. Let $B \subseteq [n]$ and define the function $\mathbb{1}_{x_B}$ on \mathcal{X} by $\mathbb{1}_{x_B}(X) = 1$ if $X_B = x_B$ and $\mathbb{1}_{x_B}(X) = 0$ otherwise. By p_B denote the *marginal distribution of X_B* defined by

$$p_B(x_B) = \mathbb{E}[\mathbb{1}_{x_B}(X)] \quad \text{for every } x_B \in \mathcal{X}_B.$$

For any two disjoint subsets $I, J \subseteq [n]$ we say that X_I and X_J are *independent*, which we denote by $I \perp\!\!\!\perp J$ (or $X_I \perp\!\!\!\perp X_J$), if and only if

$$p_{I \cup J}(x_{I \cup J}) = p_I(x_I)p_J(x_J) \quad \text{for all } x \in \mathcal{X}.$$

The following formulation of independence in terms of moments will be helpful.

Lemma 1. *We have $I \perp\!\!\!\perp J$ for some disjoint sets $I, J \subseteq [n]$ if and only if*

$$\mu_{A \cup B} = \mu_A \mu_B \quad \text{for all nonempty } A \in \mathcal{A}(\mathcal{X}_I), B \in \mathcal{A}(\mathcal{X}_J),$$

where $\mathcal{A}(\mathcal{X}_I) = \{\mathcal{A}(x) : x \in \mathcal{X}_I\}$.

Proof. We use an alternative definition of independence (see [4, page 136]) which states that X_I and X_J are independent if and only if for any two L^2 -functions f, g we have

$$\mathbb{E}[f(X_I)g(X_J)] = \mathbb{E}[f(X_I)]\mathbb{E}[g(X_J)].$$

Now the 'if' direction of the lemma is immediate. The 'only if' direction uses the fact that the set of values of X is discrete and finite. In this case any function of X is a polynomial

function (can be represented as a polynomial in the entries of X), where the terms of these polynomials are $\prod_{i \in A} X_i$ for all $A \in \mathcal{A}(\mathcal{X})$. Thus, to check if $I \perp\!\!\!\perp J$, it remains to check if

$$\mathbb{E}[f(X_I)g(X_J)] = \mathbb{E}[f(X_I)]\mathbb{E}[g(X_J)]$$

for all polynomials f, g such that each f has only terms $\prod_{i \in A} X_i$ for all nonempty $A \in \mathcal{A}(\mathcal{X}_I)$ and g has only terms $\prod_{i \in B} X_i$ for all $B \in \mathcal{A}(\mathcal{X}_J)$. By expanding the terms of f and g it suffices to check that this property holds for each monomial, which is true by assumption. \square

Example 2. Let $m = 2$, $r_1 = 2$ and $r_2 = 3$. Then $\mathcal{X} = \{0, 1\} \times \{0, 1, 2\}$ and

$$\mathcal{A}(\mathcal{X}) = \{\emptyset, \{2\}, \{2, 2\}, \{1\}, \{1, 2\}, \{1, 2, 2\}\}.$$

Since $\mathcal{A}(\mathcal{X}_1) = \{\emptyset, \{1\}\}$ and $\mathcal{A}(\mathcal{X}_2) = \{\emptyset, \{2\}, \{2, 2\}\}$, by Lemma 1, we have $1 \perp\!\!\!\perp 2$ if and only if $\mu_{12} = \mu_1\mu_2$, $\mu_{122} = \mu_1\mu_{22}$, where $\mu_{122} = \mathbb{E}[X_1X_2^2]$.

Lemma 1 generalizes and we have $I_1 \perp\!\!\!\perp I_2 \perp\!\!\!\perp \dots \perp\!\!\!\perp I_r$ for some disjoint sets $I_1, \dots, I_r \subseteq [n]$ if and only if

$$\mu_{A_1 \dots A_r} = \prod_{i=1}^r \mu_{A_i}, \quad \text{for all } A_i \in \mathcal{A}(\mathcal{X}_{I_i}), i = 1, \dots, r, \tag{5}$$

where $A_1 \dots A_r$ is a shorthand notation for $A_1 \cup \dots \cup A_r$.

Cumulants satisfy the following four basic properties, which make them useful for statistical modelling.

- (P1) Whenever there exists a split of the set of indices $[n]$ of X into two block $A|B$ such that $A \perp\!\!\!\perp B$ then $k_{1\dots n} = 0$.
- (P2) For any $a \in \mathbb{R}^n$ define $\tilde{X} = X + a$ and for any multiset A by \tilde{k}_A denote the corresponding cumulant of \tilde{X}_A . Then $\tilde{k}_i = k_i + a_i$ for every $i = 1, \dots, n$, and $\tilde{k}_A = k_A$ whenever $|A| \geq 2$.
- (P3) Let $Q = [q_{ij}] \in \mathbb{R}^{m \times n}$, $X \in \mathbb{R}^n$ and let $\tilde{X} = QX \in \mathbb{R}^m$. Define \tilde{k}_A as the cumulant of \tilde{X}_A , where A is a multiset of elements of $[m]$. Let $K^{(d)} = [k_{i_1 \dots i_d}]$ denote the $(n \times \dots \times n)$ -tensor indexed by all multisets of elements of $[n]$ of size $d \geq 1$; and let $\tilde{K}^{(d)} = [\tilde{k}_{i_1 \dots i_d}]$ be the $(m \times \dots \times m)$ -tensor indexed by all multisets of elements of $[m]$. Then,

$$\tilde{K}^{(d)} = Q \cdot K^{(d)}, \quad \text{for every } d \geq 1$$

where for every multiset $\{i_1, \dots, i_d\}$ of elements of $[m]$:

$$(Q \cdot K^{(d)})_{i_1 \dots i_d} := \sum_{j_1=1}^n \dots \sum_{j_d=1}^n q_{i_1 j_1} \dots q_{i_d j_d} k_{j_1 \dots j_d}.$$

In other words cumulants under linear mappings transform as contravariant tensors.

(P4) For two random vectors X, Y of dimension n denote by $k_A(X)$, $k_A(Y)$ and $k_A(X+Y)$ the cumulants of X , Y and $X+Y$ respectively. If $X \perp\!\!\!\perp Y$ then $k_A(X+Y) = k_A(X) + k_A(Y)$ for every multiset A of elements of $[n]$.

In this paper we generalize cumulants by changing the set $\Pi([n])$ in (1) for other set partition lattices. The term $(-1)^{|\pi|}(|\pi|-1)!$ in each summand of (1) is replaced by another function of π which will be specified later. These generalized cumulants keep usually all properties (P1)-(P4) of classical cumulants. Also the Brillinger's conditional cumulants formula derived in [1] can be generalized under additional conditions.

Different forms of cumulants are known to researchers in non-commutative probability. For example *free cumulants* are used in the theory of random matrices [9, 20] and *Boolean cumulants* are applied to stochastic differential equations [10]. All those cumulants fall under our general definition. In Proposition 27 we show that central moments can be also represented as generalized cumulants. As an interesting implication we get a simple computationally efficient formula for central moments in terms of moments (see Lemma 28). The proof of this formula is straightforward.

As it has been already pointed out in [23], cumulants and cumulant-like quantities are also useful in algebraic geometry. The coordinate system given by cumulants has a number of useful properties. For example, the tangential variety $\text{Tan}((\mathbb{P}^1)^n)$, when expressed in binary cumulants, becomes toric. Also, the study of the secant variety $\text{Sec}((\mathbb{P}^1)^n)$ becomes easier when we change coordinates to binary tree cumulants. This happens because the induced parametrization in this new coordinate system becomes nearly monomial (see Section 3.3).

There are two main reasons why cumulants can be successfully used in algebraic geometry and in the geometric study in statistics. First, many interesting algebraic varieties coincide with some statistical models. Second, the whole machinery of cumulants is purely algebraic in the sense that nonnegativity of probabilities does not play any role. In fact the only condition which we impose on probabilities is that they sum to one. For that reason the same techniques can be applied to any complex tensor with coordinates summing to one. This observation links our work to the theory of umbral calculus [14].

This paper is organized as follows. In Section 2 we introduce some basic concepts of the theory of partially ordered sets. In Section 3 we define binary L -cumulants, which form a rather straightforward generalization of binary cumulants introduced in [23]. In Section 3.2 we present how binary L -cumulants may be used in algebraic geometry. This is then exemplified with a basic study of secant varieties in Section 3.3. The general definition of L -cumulants is provided in Section 4. In Section 5 we show that, under some mild conditions, all the basic properties (P1)-(P4) of classical cumulants hold also for L -cumulants. Moreover, in Section 5.3 we generalize the Brillinger's formula for cumulants in terms of conditional cumulants. In Section 6 we show how the results of this paper explain why tree cumulants work so well for tree models. We also provide a simple analysis of processes with an underlying hidden two-state Markov chain, which in particular gives a very simple parametrization of homogeneous binary hidden Markov models.

2. Basic combinatorics

In this section we introduce basic combinatorial concepts used later in the paper. For a more detailed treatment see [22]. Recall that $\pi = B_1 | \dots | B_k$ is a partition of $[n]$, if the *blocks* $B_i \neq \emptyset$ are disjoint sets whose union is $[n]$. Equivalently, a partition of $[n]$ corresponds to an equivalence relation \sim_π on $[n]$ where $i \sim_\pi j$ if i and j lie in the same block. Let now A be a multiset $A = \{i_1, \dots, i_d\}$ of elements of $[n]$. We define a partition π of A using a partition π of $[d]$ by $i_j \sim_\pi i_k$ if $j \sim_\pi k$ in $\Pi([d])$. The set of all partitions of A is denoted by $\Pi(A)$ and by definition it is isomorphic to $\Pi([d])$.

A *partially ordered set* \mathcal{P} (or *poset*) is a set together with an ordering \leq such that: $\pi \leq \pi$ for all $\pi \in \mathcal{P}$; if $\pi \leq \nu$ and $\nu \leq \pi$ then $\pi = \nu$; and if $\pi \leq \nu$ and $\nu \leq \delta$ then $\pi \leq \delta$ for all $\pi, \nu, \delta \in \mathcal{P}$. A *subset* of \mathcal{P} is any subset of \mathcal{P} with the same ordering. As an important example of a poset consider the set $\Pi([n])$ with the poset structure given by refinement ordering such that $\pi \leq \nu$ in $\Pi([n])$ if and only if every block of π is contained in a block of ν . For instance let $n = 5$, $\pi = 13|4|25$ and $\nu = 1235|4$ then $\pi \leq \nu$.

We say that \mathcal{P} has a $\hat{0}$ if there exists an element $\hat{0} \in \mathcal{P}$ such that $\pi \geq \hat{0}$ for all $\pi \in \mathcal{P}$. Similarly, \mathcal{P} has a $\hat{1}$ if there exists $\hat{1} \in \mathcal{P}$ such that $\pi \leq \hat{1}$ for all $\pi \in \mathcal{P}$. If π and ν belong to a poset \mathcal{P} , then an *upper bound* of π and ν is an element $\delta \in \mathcal{P}$ satisfying $\delta \geq \pi$ and $\delta \geq \nu$. A *least upper bound* of π and ν is an upper bound δ of π and ν such that every upper bound γ of π and ν satisfies $\gamma \geq \delta$. If a least upper bound of π and ν exists, then it is clearly unique and it is denoted by $\pi \vee \nu$. Dually one can define the *greatest lower bound* $\pi \wedge \nu$ when it exists. We call \vee the *join operator* and \wedge the *meet operator*.

A *lattice* is a poset L for which every pair of elements has a least upper bound and greatest lower bound. A *sublattice* of a lattice L is a nonempty subset of L which is a lattice with *the same* meet and join operations as L . Clearly all finite lattices have a $\hat{0}$ and $\hat{1}$. In particular $\Pi([n])$ forms a lattice where the n -block partition $1|2| \dots |n$ is the $\hat{0}$, and the one-block partition $[n]$ is the $\hat{1}$ of this lattice. A *meet semilattice* is a poset S for which every pair of elements has a least upper bound. A *meet subsemilattice* of S is a subset of S which forms a meet semilattice with the same meet operator as S . Dually we define a *join semilattice* and a *join subsemilattice*.

Definition 3. By a *partition lattice* of a set $[n]$ we mean any lattice L which forms a subset of $\Pi([n])$ and both the one block partition $[n]$ and the minimal partition $1|2| \dots |n$ lie in L .

Note that we do not require that a partition lattice forms a sublattice of $\Pi([n])$.

Definition 4. The following is a list of interesting set partition lattices.

- (1) A partition $\pi \in \Pi([n])$ is *non-crossing* if there is no quadruple of elements $i < j < k < l$ such that $i \sim_\pi k$, $j \sim_\pi l$ and $i \not\sim_\pi j$. The noncrossing partitions of $[n]$ form a lattice which we denote by $\text{NC}([n])$. This lattice is not a sublattice of $\Pi([n])$, however, it is a meet subsemilattice of $\Pi([n])$ because the meet operators coincide.

- (2) An *interval partition* of $[n]$ is a partition π of a form

$$1 \cdots i_1 | (i_1 + 1) \cdots i_2 | \cdots | (i_k + 1) \cdots n$$

for some $0 \leq k \leq n - 1$ and $1 \leq i_1 < \dots < i_k \leq n - 1$. The poset of all interval partitions is denoted by $\mathcal{I}([n])$. It forms a sublattice of $\Pi([n])$ isomorphic to the Boolean lattice of $[n - 1]$.

- (3) A partition $\pi \in \Pi([n])$ is called a *one-cluster partition* if it contains at most one block of size greater than one. In particular the one-block partition $[n]$ and the minimal partition $1|2|\dots|n$ are one-cluster partitions. The poset of all one-cluster partitions forms a lattice $\mathcal{C}([n])$, which is not a sublattice of $\Pi([n])$. It is isomorphic to the poset of all subsets of $[n]$ excluding singletons. It forms a meet subsemilattice of $\Pi([n])$.
- (4) Let $T = (V, E)$ be a fixed tree with set of nodes V , set of edges E and with n leaves labelled by $[n]$. Removing a subset of edges E' from E induces a forest. Restricting $[n]$ to the connected components of this forest gives a *tree partition* π induced by T . The set of all tree partitions induced by T is denoted by $\mathcal{T}^T([n])$ and it forms a lattice which is a meet subsemilattice of $\Pi([n])$. For an example of a tree and the induced lattice of partitions see Figure 1 (for $n = 4$) and Figure 2.

For every poset \mathcal{P} we define the *Möbius function* $\mathbf{m}_{\mathcal{P}} : \mathcal{P} \times \mathcal{P} \rightarrow \mathbb{R}$ by

$$\mathbf{m}_{\mathcal{P}}(\pi, \nu) = \begin{cases} 1, & \text{if } \pi = \nu, \\ -\sum_{\pi \leq \delta < \nu} \mathbf{m}_{\mathcal{P}}(\pi, \delta) & \text{if } \pi < \nu, \\ 0, & \text{otherwise.} \end{cases} \tag{6}$$

When there is no ambiguity we usually drop \mathcal{P} in the notation denoting the Möbius function on \mathcal{P} by \mathbf{m} . Note that directly from the definition in (6)

$$\sum_{\pi \leq \delta \leq \nu} \mathbf{m}_{\mathcal{P}}(\pi, \delta) = \begin{cases} 0 & \text{if } \pi < \nu \\ 1 & \text{if } \pi = \nu. \end{cases} \tag{7}$$

A special type of a subposet of \mathcal{P} is the *interval*

$$[\pi, \nu] = \{\delta \in \mathcal{P} : \pi \leq \delta \leq \nu\},$$

defined whenever $\pi \leq \nu$. The Möbius function on this subposet is naturally induced from the Möbius function on \mathcal{P} (see for example [13, Proposition 4]). For any two posets $\mathcal{P}_1, \mathcal{P}_2$ we define the poset $\mathcal{P}_1 \times \mathcal{P}_2$ as a set with the ordering $(\pi, \nu) \leq (\pi', \nu')$ if $\pi \leq \pi'$ and $\nu \leq \nu'$. The following result gives a convenient way of finding a Möbius function for posets constructed from other posets by taking products.

Proposition 5 (Proposition 3.1.2, [22]). *Let \mathcal{P}_1 and \mathcal{P}_2 be finite posets, and let $\mathcal{P}_1 \times \mathcal{P}_2$ be their direct product. If $(\pi, \nu) \leq (\pi', \nu')$ in $\mathcal{P}_1 \times \mathcal{P}_2$, then*

$$\mathbf{m}_{\mathcal{P}_1 \times \mathcal{P}_2}((\pi, \nu), (\pi', \nu')) = \mathbf{m}_{\mathcal{P}_1}(\pi, \nu) \mathbf{m}_{\mathcal{P}_2}(\pi', \nu').$$

The Möbius function is especially useful due to the following result.

Proposition 6 (Möbius inversion formula). *Let \mathcal{P} be a finite poset. Let $f, g : \mathcal{P} \rightarrow \mathbb{R}$. Then*

$$g(\pi) = \sum_{\nu \leq \pi} f(\nu), \quad \text{for all } \pi \in \mathcal{P},$$

if and only if

$$f(\pi) = \sum_{\nu \leq \pi} \mathbf{m}(\nu, \pi) g(\nu) \quad \text{for all } \pi \in \mathcal{P}.$$

For every lattice denote $\mathbf{m}(\pi) := \mathbf{m}(\pi, \hat{1})$. Later we will see that it is particularly important to identify values of $\mathbf{m}(\pi)$ for various partition lattices. For $\Pi([n])$ we have $\mathbf{m}(\pi) = (-1)^{|\pi|-1} (|\pi| - 1)!$. The lattice of interval partitions $\mathcal{I}([n])$ is isomorphic to the Boolean lattice of all subsets of $[n - 1]$ and hence $\mathbf{m}(\pi) = (-1)^{|\pi|-1}$. For the lattice of one-cluster partitions we have

$$\mathbf{m}(\pi) = \begin{cases} (-1)^{n-1} (n - 1) & \text{if } \pi = 1|2| \cdots |n, \text{ and} \\ (-1)^{|\pi|-1} & \text{otherwise.} \end{cases} \tag{8}$$

For the other cases in Definition 4 the Möbius function can be computed recursively.

3. Binary L -cumulants

In this section we discuss binary L -cumulants which generalize binary cumulants of [23]. Most of the technical results will be stated without proofs, which will then be given in a more general context in later sections.

3.1. Definition and basic facts

Assume that $\mathcal{X} = \{0, 1\}^n$, in which case $\mathcal{A}(\mathcal{X})$ is the set of all subsets of $[n]$. Let $L \subseteq \Pi([n])$ be a partition lattice of $[n]$. For every $I \subseteq [n]$ consider $L(I)$ as the subposet of $\Pi(I)$ obtained from L by constraining each partition to the subset I . The Möbius function on $L(I)$ is also denoted by \mathbf{m} unless it may lead to ambiguity in which case we write explicitly \mathbf{m}_I .

A *multiplicative function* on $L(I)$ is any function such that for every $\pi \in L(I)$

$$f(\pi) = \prod_{B \in \pi} f_B \quad \text{for some } f_B \in \mathbb{R}.$$

First consider the case when $L = \Pi([n])$. For every $I \subseteq [n]$ and $\nu \in \Pi(I)$ define

$$k(\nu) = \sum_{\pi \leq \nu} \mathbf{m}(\pi, \nu) \mu(\pi), \tag{9}$$

where $\mu(\pi) = \prod_{B \in \pi} \mu_B$ is a multiplicative function and the sum is taken over elements π of $\Pi(I)$ such that $\pi \leq \nu$. The one-block partition I is the unique maximal element of the lattice $\Pi(I)$. The Möbius function on $\Pi(I)$ satisfies $\mathbf{m}(\pi) := \mathbf{m}(\pi, I) = (-1)^{|\pi|-1} (|\pi| - 1)!$

for all $\pi \in \Pi(I)$. It follows by (3) that $k_I = k(I)$ and hence (9) evaluated at $\nu = I$ gives the definition of binary cumulants.

To get the inverse formula for moments in terms of cumulants we need the following result.

Lemma 7. *For every $\nu \in \Pi(I)$ we have $k(\nu) = \prod_{B \in \nu} k_B$, where $k(\nu)$ is defined by (9).*

Proof. Note that every interval $[\pi, \nu] \subseteq \Pi(I)$ is isomorphic to a product of intervals $\prod_{B \in \nu} [\pi(B), B] \subseteq \prod_{B \in \nu} \Pi(B)$, where $\pi(B)$ denotes π constrained to elements in $B \subseteq I$. By Proposition 5 a Möbius function on a product of posets is equal to the product of Möbius functions for each individual factor. Hence, (9) can be rewritten as

$$k(\nu) = \prod_{B \in \nu} \left(\sum_{\delta \in \Pi(B)} \mathfrak{m}_B(\delta) \mu(\delta) \right) = \prod_{B \in \nu} k_B,$$

which finishes the proof. □

The inverse formula for moments in terms of cumulants follows directly by Proposition 6 and Lemma 7. For every $I \subseteq [n]$ we have

$$\mu_I = \sum_{\pi \in \Pi(I)} k(\pi) = \sum_{\pi \in \Pi(I)} \prod_{B \in \pi} k_B. \tag{10}$$

We can directly generalize the definition of binary cumulants to *binary L-cumulants*. Let L be a partition lattice of $[n]$. Define binary L -cumulants by

$$\ell_I = \sum_{\pi \in L(I)} \mathfrak{m}(\pi) \prod_{B \in \pi} \mu_B \quad \text{for every } I \subseteq [n]. \tag{11}$$

By definition for every $I \subseteq [n]$ the maximal and minimal element of the lattice $L(I)$ coincide with the minimal and maximal element of $\Pi(I)$. In particular for every L we have $\ell_i = \mu_i$ for $i = 1, \dots, n$; and $\ell_{ij} = \mu_{ij} - \mu_i \mu_j$ for all $1 \leq i < j \leq n$. However, already when $n = 3$ not all L -cumulants coincide with cumulants.

Example 8. Let $n = 3$ and consider L -cumulants induced by the lattice of interval partitions. The lattice $\mathcal{I}([3])$ has four elements: 123 , $1|23$, $12|3$ and $1|2|3$ and $\mathfrak{m}(\pi) = (-1)^{|\pi|-1}$. Therefore, we have

$$\ell_{123} = \mu_{123} - \mu_1 \mu_{23} - \mu_{12} \mu_3 + \mu_1 \mu_2 \mu_3.$$

Compare this with the formula for k_{123} in (2) to note that not only the term $\mu_2 \mu_{13}$ is missing now in the formula for ℓ_{123} but also the coefficient of $\mu_1 \mu_2 \mu_3$ is 1 not 2.

Let $\pi \in \Pi([n])$ be a set partition into blocks B_1, \dots, B_r . Denote

$$\perp_{B \in \pi} X_B \quad := \quad X_{B_1} \perp \cdots \perp X_{B_r}.$$

By (5), $\perp_{B \in \pi} X_B$ if and only if

$$\mu_I = \mu(\pi(I)) \quad \text{for every } I \subseteq [n], \tag{12}$$

where $\pi(I)$ denotes π constrained to elements in B . So for example the full independence is given by the minimal partition $\pi = 1|2|\dots|n$ and $\mu_I = \prod_{i \in I} \mu_i$ for every $I \subseteq [n]$.

Below we list the basic facts about binary L -cumulants. They are proved in a more general setting in Section 5. The following result implies that (P1) holds for binary L -cumulants.

Proposition 9. *There exists a partition $\pi_0 \in L$ such that $\perp_{B \in \pi_0} X_B$ if and only if $\ell(\pi) = 0$ for all $\pi \not\leq \pi_0$, or equivalently, if $\ell_I = 0$ unless I is contained in one of the blocks of π_0 (equivalence follows from Theorem 24).*

Proof. The 'if' part of the proposition is given in a more general setting in Proposition 25. To prove the opposite implication use Theorem 24 to conclude that $\ell(\pi) = 0$ for all $\pi \not\leq \pi_0$ implies that $\mu_I = \mu(\pi_0(I))$ for all $I \subseteq [n]$ which by (12) implies $\perp_{B \in \pi_0} X_B$. \square

Example 10. Consider the situation of Example 8, where $n = 3$ and L -cumulants are defined by the lattice of interval partitions. If $X_1 \perp (X_2, X_3)$ then $\mu_{123} = \mu_1\mu_{23}$, $\mu_{12} = \mu_1\mu_2$ and $\mu_{13} = \mu_1\mu_3$. It follows that $\ell_{12} = \ell_{13} = \ell_{123} = 0$. On the other hand, the condition $X_2 \perp (X_1, X_3)$ does not imply that $\ell_{123} = 0$ because in this case

$$\ell_{123} = \mu_2\mu_{13} - \mu_1\mu_2\mu_3,$$

which is zero only when in addition $\mu_{13} = \mu_1\mu_3$ and hence when $X_1 \perp X_3$. Here there is no contradiction with Proposition 9 because $2|13 \notin \mathcal{I}(\{3\})$.

Under a minor additional condition the property (P2) also holds for binary L -cumulants.

Proposition 11. *Suppose that for every $i \in [n]$ the split $i|([n] \setminus i)$ lies in L . Let $\tilde{X} = X + a$, where $a \in \mathbb{R}^n$ and, for every $I \subseteq [n]$, by $\tilde{\ell}_I$ denote the corresponding L -cumulant of the subvector X_I . Then $\tilde{\ell}_i = \ell_i + a_i$ for all $i = 1, \dots, n$ and $\tilde{\ell}_I = \ell_I$ for any $I \subseteq [n]$ such that $|I| \geq 2$.*

Proof. This follows from Proposition 26. \square

Define *central binary L -cumulants* by replacing moments μ_B in (11) by central moments μ'_B . For every $I \subseteq [n]$ the corresponding central binary L -cumulant is denoted by ℓ'_I .

Lemma 12. *Under the assumptions of Proposition 11 we have $\ell'_I = \ell_I$ for every $I \subseteq [n]$ such that $|I| \geq 2$.*

Proof. Central binary L -cumulants of X can be alternatively defined as binary L -cumulants of \tilde{X} , where $\tilde{X}_i = X_i - \mathbb{E}X_i$. The lemma follows from Proposition 11. \square

In the next section we show how all these ideas can be applied in algebraic geometry.

3.2. Geometric applications

We consider algebraic varieties in either the real space $\mathbb{R}^{2^n} = \mathbb{R}^{2 \times \dots \times 2}$ or its complexification $\mathbb{C}^{2^n} = \mathbb{C}^{2 \times \dots \times 2}$, or projectivization $\mathbb{P}^{2^n-1} = \mathbb{P}(\mathbb{C}^{2 \times \dots \times 2})$. Each component \mathbb{C}^2 (or \mathbb{R}^2) has basis e_0, e_1 so that $e_{i_1} \otimes \dots \otimes e_{i_n}$ corresponds to $I \subseteq [n]$ for $i_j = 1$ if $j \in I$ and $i_j = 0$ otherwise. For example, if $n = 2$ and $\mu \in \mathbb{C}^{2 \times 2}$ then we write μ in our basis as

$$\mu = \mu_\emptyset e_0 \otimes e_0 + \mu_1 e_1 \otimes e_0 + \mu_2 e_0 \otimes e_1 + \mu_{12} e_1 \otimes e_1.$$

Formula (11) gives an isomorphism of the affine subspace $\mu_\emptyset = 1$ in \mathbb{R}^{2^n} (or \mathbb{C}^{2^n}), which forms a Zariski open subset of \mathbb{P}^{2^n-1} . The inverse map is computed in a more general case in (21).

We first show that some basic operations on the random vector X can encode interesting actions on the space of $2 \times \dots \times 2$ tensors. Define \tilde{X} such that $\tilde{X}_i = \lambda_i X_i$ for $\lambda_i \in \mathbb{C} \setminus \{0\}$ for $i = 1, \dots, n$. Multiplying each X_i by λ_i results in the change of moments from μ_I to $\tilde{\mu}_I = \prod_{i \in I} \lambda_i \mu_I$ and hence it corresponds to the action of the group D^n , where D a group of diagonal matrices of the form

$$\begin{bmatrix} 1 & 0 \\ 0 & \lambda \end{bmatrix} \quad \text{for } \lambda \in \mathbb{C} \setminus \{0\}.$$

Because L -cumulants are multilinear functions of the moments we conclude that this action is the same on the level of L -cumulants. We have $\tilde{\ell}_I = \prod_{i \in I} \lambda_i \ell_I$ for every $I \subseteq [n]$.

Suppose now that $\tilde{X} = X + b$, for $b = (b_1, \dots, b_n) \in \mathbb{C}^n$, and consider the group $U(2)^n$ where $U(2)$ is the unipotent group of 2×2 -matrices of the form

$$\begin{bmatrix} 1 & 0 \\ \lambda & 1 \end{bmatrix} \quad \text{for } \lambda \in \mathbb{C}.$$

Adding b to the vector X corresponds to the action of $U(2)^n$, with $\lambda_i = b_i$ for $i = 1, \dots, n$, on the space of moments. We illustrate this with an example that easily generalizes.

Example 13. Let $n = 2$ and denote by $\tilde{\mu} = [\tilde{\mu}_I]$ the moments of the vector $\tilde{X} = X + b$. We have $\tilde{\mu}_\emptyset = 1$, $\tilde{\mu}_i = \mathbb{E}(X_i + b_i) = \mu_i + b_i$ for $i = 1, 2$ and

$$\tilde{\mu}_{12} := \mathbb{E}[(X_1 + b_1)(X_2 + b_2)] = \mu_{12} + b_1 \mu_2 + \mu_1 b_2 + b_1 b_2.$$

Write $\mu = [\mu_I] \in \mathbb{C}^{2 \times 2}$:

$$\mu = e_0 \otimes e_0 + \mu_1 e_1 \otimes e_0 + \mu_2 e_0 \otimes e_1 + \mu_{12} e_1 \otimes e_1.$$

After applying the action of $U(2)^2$ with $\lambda_i = b_i$ for $i = 1, 2$ we obtain

$$\begin{aligned} \tilde{\mu} &= (e_0 + b_1 e_1) \otimes (e_0 + b_2 e_1) + \mu_1 e_1 \otimes (e_0 + b_2 e_1) + \mu_2 (e_0 + b_1 e_1) \otimes e_1 + \mu_{12} e_1 \otimes e_1 \\ &= e_0 \otimes e_0 + (\mu_1 + b_1) e_1 \otimes e_0 + (\mu_2 + b_2) e_0 \otimes e_1 + (\mu_{12} + b_1 \mu_2 + \mu_1 b_2 + b_1 b_2) e_1 \otimes e_1 \\ &= e_0 \otimes e_0 + \tilde{\mu}_1 e_1 \otimes e_0 + \tilde{\mu}_2 e_0 \otimes e_1 + \tilde{\mu}_{12} e_1 \otimes e_1, \end{aligned}$$

which confirms that translating X by $b \in \mathbb{R}^n$ corresponds to the action of $U(2)^n$ on μ .

For every $I \subseteq [n]$, denote by $\tilde{\ell}_I$ the L -cumulant of \tilde{X}_I . By Proposition 11, whenever every split $i|([n] \setminus i)$ lies in L , this complicated transformation of moments induced by $U(2)^n$ translates to a very simple transformation of cumulants. We have $\tilde{\ell}_i = \ell_i + b_i$ for $i \in [n]$ and $\tilde{\ell}_I = \ell_I$ for all $I \subseteq [n]$ such that $|I| \geq 2$ and hence all the higher order L -cumulants are invariant with respect to the action of $U(2)^n$ on the space of moments.

Changing values of the binary variables X_i from 0, 1 to b_i, a_i , means defining a new random vector \tilde{X} such that $\tilde{X}_i = (a_i - b_i)X_i + b_i$. We have just shown that changing values of the components of X corresponds to a natural action of the n -dimensional torus $(\mathbb{C}^*)^n$ with coordinates $a_i - b_i$ on the space $\mathbb{C}^{2^n - n - 1}$ whose coordinates are the higher order L -cumulants ℓ_I , $|I| \geq 2$. More specifically the L -cumulants of \tilde{X} , such that $\tilde{X}_i = (a_i - b_i)X_i + b_i$, are transformed by

$$\tilde{\ell}_I = \ell_I \cdot \prod_{i \in I} (a_i - b_i) \quad \text{for all } I \subseteq [n] \text{ and } |I| \geq 2$$

and $\tilde{\ell}_i = (a_i - b_i)\ell_i + b_i$ for $i = 1, \dots, n$. This leads to the following result.

Theorem 14. *Suppose that for every $i \in [n]$ the split $i|([n] \setminus i)$ lies in L . Then a subvariety of $\mathbb{C}^{2^n - 1}$ is invariant under changing values of components of X if and only if it is defined by \mathbb{Z}^n -homogeneous polynomials in ℓ_I with $|I| \geq 2$.*

Proof. See the proof of [23, Theorem 3.1]. □

Note that if a variety is invariant under the action of the *special linear group* $SL(2)^n$ then in particular it is invariant under $U(2)^n$.

Corollary 15. *Suppose that L is a partition lattice of $[n]$ such that for every $i \in [n]$ the split $i|([n] \setminus i)$ lies in L . Let V be a subvariety of the affine open subset given by $\mu_\emptyset = 1$ in the projective space $\mathbb{P}(\mathbb{C}^{2 \times \dots \times 2})$ and let \bar{V} denote its closure in that projective space. If \bar{V} is invariant under the action of $SL(2)^n$ then the ideal I_V that defines V is generated by \mathbb{Z}^n -homogeneous polynomials in the L -cumulants ℓ_I with $|I| \geq 2$.*

Another important reason why L -cumulants may be useful, apart from their invariance properties, is related to property (P1). Denote by $\text{Seg}((\mathbb{P}^1)^n)$ the *Segre variety*, which is an embedding of $(\mathbb{P}^1)^n$ into $\mathbb{P}^{2^n - 1}$. In statistics the Segre variety corresponds to the full independence model $X_1 \perp\!\!\!\perp \dots \perp\!\!\!\perp X_n$. In particular Proposition 9 implies that the image of $\text{Seg}((\mathbb{P}^1)^n)$ in the space given by L -cumulants is an affine subspace given by $\ell_I = 0$ for all $|I| \geq 2$ (see also [23, Remark 3.4]). Moreover, L -cumulants seem to be helpful also in the analysis of other algebraic varieties related to the Segre variety $\text{Seg}((\mathbb{P}^1)^n)$. For example the tangential variety $\text{Tan}((\mathbb{P}^1)^n)$ is toric when expressed in cumulants (see [23, Theorem 4.1]). In the following section we show how L -cumulants defined by a tree partition lattice can help to analyze the *secant variety* $\text{Sec}((\mathbb{P}^1)^n)$.

3.3. Binary tree cumulants for secant varieties

In [27] we defined tree cumulants, which gave a better understanding of certain statistical models related to trees. We write more on that in Section 6. In this section we show

how tree cumulants can be used to study secant varieties. Recall from Definition 4 that, for a fixed tree T with n -leaves, $\mathcal{T}^T([n])$ denotes the lattice of tree partitions of $[n]$ induced by T . Moreover, $\mathcal{T}^T(I)$ is the lattice of all tree partitions of I induced by $T(I)$, which is the smallest subtree of T containing all leaves in I . The tree cumulant of the subvector X_I for every $I \subseteq [n]$ is denoted by \mathfrak{t}_I . Tree cumulants are L -cumulants and hence defined by (11):

$$\mathfrak{t}_I = \sum_{\pi \in \mathcal{T}^T(I)} \mathfrak{m}(\pi) \prod_{B \in \pi} \mu_B, \quad \text{for all } I \subseteq [n]. \tag{13}$$

Remark 16. In [27, Section 3.2] binary tree cumulants were defined in terms of central moments by

$$\tilde{\mathfrak{t}}_I = \sum_{\pi \in \mathcal{T}^T(I)} \mathfrak{m}(\pi) \prod_{B \in \pi} \mu'_B \quad \text{for all } I \subseteq [n], |I| \geq 2,$$

and $\tilde{\mathfrak{t}}_i = \mu_i$ for $i \in [n]$. In particular $\tilde{\mathfrak{t}}_I$ for all $|I| \geq 2$ is just the corresponding central L -cumulant. Let $i \in [n]$ be one of the leaves. Removing the edge incident with i induces a split $i|([n] \setminus i)$ and hence the assumption of Proposition 11 holds and, by Lemma 12, it follows that $\tilde{\mathfrak{t}}_I = \mathfrak{t}_I$ for all $I \subseteq [n]$. In particular, both the definition in [27] and the one given in (13) are equivalent.

Let L be the lattice of tree partitions induced by the caterpillar tree in Figure 1. For example if $n = 4$ then the induced lattice is given in Figure 2. We first show how to compute L -cumulants $[\mathfrak{t}_I]$ without computing the Möbius function on the lattice L . By Remark 16 we can replace moments by central moments in the formula for \mathfrak{t}_I for all $I \subseteq [n]$ such that $|I| \geq 2$. This is very convenient because $\prod_{B \in \pi} \mu'_B$ is zero whenever π contains a singleton block. Note that the elements of L with no singleton blocks correspond to all interval partitions with no singleton blocks. If $n = 4$ then the elements of L with no singleton blocks are the two boldfaced elements in Figure 2. This gives that for all $I \subseteq [n]$ such that $|I| \geq 2$:

$$\mathfrak{t}_I = \sum_{\pi \in L(I)} \mathfrak{m}(\pi) \prod_{B \in \pi} \mu'_B = \sum_{\pi \in \mathcal{I}(I)} \mathfrak{m}(\pi) \prod_{B \in \pi} \mu'_B.$$

Both sums above are over all partitions in a poset of all interval partitions with no singleton blocks. Hence, both Möbius functions constrained to this poset need to coincide. The gain is that we already computed the Möbius function on the right-hand side explicitly obtaining $\mathfrak{m}(\pi) = (-1)^{|\pi|-1}$ (see the end of Section 2).

This allows us to write the map from moments $[\mu_I]$ to tree cumulants $[\mathfrak{t}_I]$ of the caterpillar tree as a composition of two maps: from moments to central moments and from central moments to tree cumulants induced by the caterpillar tree. We will show in the end of Section 5.1 that the first map can be written as

$$\mu'_I = \sum_{B \subseteq I} (-1)^{|I \setminus B|} \mu_B \prod_{i \in I \setminus B} \mu_i \quad \text{for all } I \subseteq [n], |I| \geq 2,$$

and we have just shown that the second map is given by $t_i = \mu_i$ for $i = 1, \dots, n$, and

$$t_I = \sum_{\pi \in \mathcal{I}(I)} (-1)^{|\pi|-1} \prod_{B \in \pi} \mu'_B \quad \text{for all } |I| \geq 2.$$

In particular, if $n = 4$ then $t_I = \mu'_I$ for all $2 \leq |I| \leq 3$ and

$$t_{1234} = \mu'_{1234} - \mu'_{12}\mu'_{34}.$$

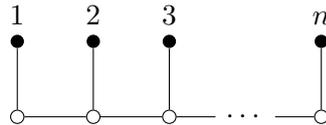


Figure 1: A caterpillar tree with n leaves/legs.

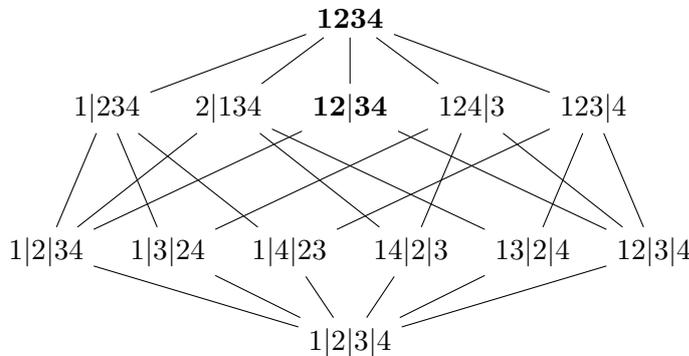


Figure 2: The Hasse diagram of the lattice of tree partitions induced by the tree in Figure 1 if $n = 4$.

We use this new coordinate system to study the secant variety $\text{Sec}((\mathbb{P}^1)^n)$. As an example consider the case when $n = 4$.

Example 17. The secant variety $\text{Sec}((\mathbb{P}^1)^4)$ is a projective variety in \mathbb{P}^{15} parametrized by 9 copies of \mathbb{P}^1 with coordinates (t_0, t) , (a_{0i}, a_i) and (b_{0i}, b_i) for $i = 1, 2, 3, 4$. The parametrization is given by

$$\mu_I = t_0 \prod_{i \in I^c} a_{0i} \prod_{i \in I} a_i + t \prod_{i \in I^c} b_{0i} \prod_{i \in I} b_i \quad \text{for all } I \subseteq [4],$$

where I^c denotes the complement of I in $\{1, 2, 3, 4\}$ and $\mu = [\mu_I]$ denotes the coordinates of the projective space \mathbb{P}^{15} . We want to describe the image of an open subset of the parameter space given by $a_{0i} = b_{0i} = 1$ for $i \in \{1, 2, 3, 4\}$ and $t_0 = 1 - t$. This image is described by

$$\mu_I = (1 - t) \prod_{i \in I} a_i + t \prod_{i \in I} b_i \tag{14}$$

and in particular $\mu_\emptyset = 1$.

Earlier in this section we explained how to compute $[t_I]$ from moments as a composition of two simple maps. From this we can also compute the induced parametrization directly. Here we will show an alternative way of proceeding for the secant variety $\text{Sec}((\mathbb{P}^1)^4)$ to present some other available techniques. First, use the parametrization of the secant in terms of classical cumulants. This parametrization was given in [23, Equations (18) and (19)], which implies that for every $i < j < k$

$$\begin{aligned} k_{ij} &= t(1-t)(b_i - a_i)(b_j - a_j) \\ k_{ijk} &= t(1-t)(1-2t)(b_i - a_i)(b_j - a_j)(b_k - a_k) \\ k_{1234} &= t(1-t)(6t^2 - 6t + 1) \prod_{i=1}^4 (b_i - a_i). \end{aligned} \tag{15}$$

Now we change coordinates from cumulants to binary tree cumulants $[t_I]$ using Proposition 21. In particular, as explained in Example 22, since $13|24$ and $14|23$ are the only partitions in $\Pi([4])$ which are not tree partitions of the caterpillar tree in Figure 1 for $n = 4$, this yields

$$t_{1234} = k_{1234} + k_{13}k_{24} + k_{14}k_{23} \tag{16}$$

and $t_I = k_I$ for all $I \subseteq [4]$ such that $|I| \leq 3$. From this it follows that for every $I \subseteq \{1, 2, 3, 4\}$ such that $|I| \geq 2$:

$$t_I = t(1-t)(1-2t)^{|I|-2} \prod_{i \in I} (b_i - a_i), \tag{17}$$

which for t_{1234} can be verified by direct computations. Now we can immediately check that

$$t_{I \cup J} t_{I' \cup J'} - t_{I \cup J'} t_{I' \cup J} = 0$$

holds on $\text{Sec}((\mathbb{P}^1)^4)$ for all distinct $I, I' \in \{\{i\}, \{j\}, \{i, j\}\}$ and $J, J' \in \{\{k\}, \{l\}, \{k, l\}\}$ and every split $ij|kl$ of $\{1, 2, 3, 4\}$. For example $12|34$ leads to a set of equations including $t_{13}t_{24} - t_{14}t_{23} = 0$ and $t_{1234}t_{13} - t_{123}t_{134} = 0$.

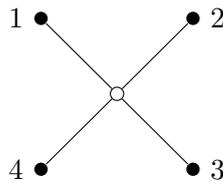


Figure 3: A 4-star tree.

This simple example can be generalized using the link between the secant varieties and certain statistical models (see [3, Section 4.1]). Define for any two disjoint $A, C \subseteq [n]$ the conditional probability of X_A given X_C as:

$$p_{A|C}(x_A|x_C) := \frac{p_{A \cup C}(x_A, x_C)}{p_C(x_C)} \quad \text{for all } x_C \in \mathcal{X}_C \text{ s.t. } p_C(x_C) \neq 0.$$

For any function f of X_A define the *conditional expectation of $f(X_A)$ given X_C* as a function of X_C given for any $x_C \in \mathcal{X}_C$

$$\mathbb{E}[f(X_A)|X_C = x_C] = \sum_{x_A \in \mathcal{X}_A} p_{A|C}(x_A|x_C)f(x_A).$$

We denote this conditional expectation by $\mathbb{E}[f(X_A)|X_C]$. If $f(X_A) = \prod_{i \in A} X_i$ then we simply write μ_A^C and $\mu_A^C(x_C) = \mathbb{E}[\prod_{i \in A} X_i | X_C = x_C]$. Note that μ_A^C is a random variable itself.

Similarly as in the case of Lemma 1 we can show that for disjoint $C, B_1, \dots, B_r \subseteq [n]$ the X_{B_i} 's are *jointly independent given X_C* if

$$\mu_{A_1 \dots A_r}^C = \prod_{i=1}^r \mu_{A_i}^C \quad \text{for all } A_i \subseteq B_i, i = 1, \dots, r,$$

In this case the marginal distribution of X_C satisfies

$$\mu_{A_1 \dots A_r} = \mathbb{E}[\mu_{A_1 \dots A_r}^C] = \sum_{x_C \in \mathcal{X}_C} p_C(x_C) \mu_{A_1 \dots A_r}^C(x_C). \tag{18}$$

For a statistician the parametrization in (14) corresponds to the parametrization of moments of the binary 4-star tree model (naive Bayes model) as given in Figure 3. The leaves of this tree correspond to a vector $X = (X_1, X_2, X_3, X_4)$ of binary *observed* variables and the inner node corresponds to a binary variable Y which is not observed. This model contains all possible moments of a binary vector X such that all components of X are jointly independent given Y . The parametrization in (14) is a special version of (18).

The fact that (14) can be rewritten in the easier form in (17) for any $n \geq 4$ follows from more general considerations in [27, Section 4]. We obtain the following procedure:

1. Consider any trivalent tree with n leaves, that is a tree such that each inner node has valency three.
2. Compute tree cumulants induced by this trivalent tree.
3. The induced parametrization of the n -star tree model in the coordinate system constructed in step 2 is (17), where now $I \subseteq [n]$ for $n \geq 4$. For more details check Section 6.

Of course, since we can pick any trivalent tree in step 1, the most natural choice is to pick the caterpillar tree. This is mainly because the computation of the corresponding tree cumulants is simple as it was presented earlier in this section. Now from the parametrization in (17) we easily verify that

$$t_{I \cup J} t_{I' \cup J'} - t_{I \cup J'} t_{I' \cup J} = 0 \tag{19}$$

holds on $\text{Sec}((\mathbb{P}^1)^n)$ for all non-empty subsets $I, I' \subseteq A$ and $J, J' \subseteq B$ where $A|B$ is a split of $[n]$.

Remark 18. It may seem that a more natural way to proceed in Example 17 was to construct tree cumulants induced directly by partitions of the 4-star tree in Figure 3. The tree partitions of the 4-star tree are equal to one-cluster partitions from Definition 4. By Proposition 27 this partition lattice induces central moments μ'_I . To compute the induced parametrization of the central moments note that $\mu'_I = k_I$ for all $2 \leq |I| \leq 3$. A direct check shows that

$$\mu'_{1234} = t(1-t)(3t^2 - 3t + 1) \prod_{i=1}^4 (b_i - a_i)$$

and we find that the relation between μ'_{1234} and other central moments is more complicated than in the case of tree cumulants induced by the caterpillar tree. In particular, the corresponding equations are no longer binomial like in (19).

4. The definition of L -cumulants

Let $A = \{i_1, \dots, i_d\}$ be a multiset. We define its *multisubset* $B \subseteq A$ as a multiset $B = \{i_j : j \in I\}$ for some $I \subseteq [d]$. For example if $A = \{1, 1, 2, 2\}$ then A has, among others, four multisubsets of the form $\{1, 2\}$. Let X be a finite discrete random vector with values in \mathcal{X} and let $\mathcal{A}(\mathcal{X})$ be the family of multisets associated to \mathcal{X} as given in (4). Consider any family $\mathbf{L} = (L(A))_{A \in \mathcal{A}(\mathcal{X})}$ of partition lattices such that $L(A)$ is a subposet of $\Pi(A)$ for every $A \in \mathcal{A}(\mathcal{X})$. Assume that the maximal and minimal elements of $L(A)$ coincide with the maximal and the minimal element of $\Pi(A)$ and denote them by A and $\hat{0}_A$ respectively. Moreover, for every $B \subseteq A$ the map $L(A) \rightarrow L(B)$ are surjections given by constraining partitions of A to B . Note that in particular, $L(A)$ need not be a sublattice of $\Pi(A)$ because the join and the meet operator of $L(A)$ and $\Pi(A)$ may differ.

The first two trivial examples of a family \mathbf{L} as above is $\mathbf{\Pi} = (\Pi(A))_{A \in \mathcal{A}(\mathcal{X})}$ and \mathbf{L} such that for every $A \in \mathcal{A}(\mathcal{X})$, $|A| \geq 2$, the lattice $L(A)$ is given by just two elements $\hat{0}_A$ and A . Other interesting examples are obtained from Definition 4 (excluding tree partitions), where $L(A)$ is assumed to be isomorphic to $L(|A|)$. The corresponding families of lattices are denoted by \mathbf{NC} (non-crossing), \mathbf{I} (interval) and \mathbf{C} (one-cluster). A definition of tree cumulants in this case requires construction of an \hat{A} -labelled tree, where \hat{A} is the maximal multiset A in $\mathcal{A}(\mathcal{X})$ corresponding to $x = (r_1 - 1, \dots, r_n - 1)$. This construction is not unique and for that reason we discuss tree cumulants only in very concrete examples.

By \mathbf{m}_A we denote the Möbius function on $L(A)$. The lattice will be always obvious from the context so we omit it in the notation. When A is also clear from the context we just write \mathbf{m} .

Definition 19 (L -cumulants). Let $X = (X_1, \dots, X_n)$ be a random vector. For any $A \in \mathcal{A}(\mathcal{X})$ and $\nu \in L(A)$ define

$$\ell(\nu) = \sum_{\pi \leq \nu} \mathbf{m}_A(\pi, \nu) \mu(\pi), \tag{20}$$

where $\mu(\pi) = \prod_{B \in \pi} \mu_B$. Then $\ell_A := \ell(A)$ is the L -cumulant of X_A .

If $\mathbf{L} = \mathbf{\Pi}$ then, because $\Pi(A) \simeq \Pi([A])$, we obtain the formula in (3) and hence this definition generalizes the classical cumulants. Other known L -cumulants were defined in the non-commutative probability literature. These are L -cumulants defined by \mathbf{NC} and \mathbf{I} , which are called *free cumulants* and *Boolean cumulants* respectively (see [20, 21]).

The map (20) is invertible with the inverse given by the Möbius inversion formula in Proposition 6. Thus for every $A \in \mathcal{A}(\mathcal{X})$

$$\mu_A = \mu(A) = \sum_{\pi \in L(A)} \ell(\pi). \tag{21}$$

Note that in general $\ell(\pi) \neq \prod_{B \in \pi} \ell_B$, as it was the case for cumulants. However, $\ell(\pi) = \prod_{B \in \pi} \ell_B$ whenever \mathbf{L} satisfies the following condition:

(C0) For every $A \in \mathcal{A}(\mathcal{X})$ and for any two partitions $\pi, \nu \in L(A)$ the interval $[\pi, \nu]$ is isomorphic to the product of intervals $\prod_{B \in \nu} [\pi(B), B] \subseteq \prod_{B \in \nu} L(B)$.

Condition (C0) is not very restrictive. In fact all partition lattices mentioned in Definition 4 satisfy this property. If (C0) holds, then, by Proposition 5, the Möbius function on $L(A)$ satisfies $\mathbf{m}_A(\pi, \nu) = \prod_{B \in \nu} \mathbf{m}_B(\pi(B), B)$. In particular (21) becomes

$$\mu_A = \sum_{\pi \in L(A)} \prod_{B \in \pi} \ell_B,$$

and the proof of this follows essentially the proof of Lemma 7.

Remark 20. By the moment aliasing there is a one-to-one correspondence between the probabilities $P = [p(x)]_{x \in \mathcal{X}}$ and moments $M = [\mu_A]_{A \in \mathcal{A}(\mathcal{X})}$ and hence also L -cumulants $\mathcal{L} = [\ell_A]_{A \in \mathcal{A}(\mathcal{X})}$.

Unlike in the case of cumulants, for general L -cumulants no generating function is known. It may be then useful to realize that L -cumulants can be expressed in terms of classical cumulants in a rather simple manner. The following result generalizes Theorem 4.1 in [8].

Proposition 21. *Let $L(A)$ be a lattice of set partitions of A in the family \mathbf{L} and let Π^* denote the set of elements $\pi \in \Pi(A)$ such that $[\pi, A] \cap L(A) = \{A\}$, where the interval $[\pi, A]$ is taken in $\Pi(A)$. We have*

$$\ell_A = \sum_{\pi \in \Pi^*} k(\pi) = \sum_{\pi \in \Pi^*} \prod_{B \in \pi} k_B.$$

Proof. In this proof $\delta \leq_{\Pi} \pi$ means that $\delta \leq \pi$ and $\delta \in \Pi(A)$. Similarly $\pi \geq_L \delta$ denotes $\pi \geq \delta$ and $\pi \in L(A)$. Expressing the L -cumulant in terms of moments and then the moments in terms of classical cumulants gives

$$\ell_A = \sum_{\pi \in L(A)} \mathbf{m}(\pi) \prod_{B \in \pi} \left(\sum_{\delta_B \in \Pi(B)} \prod_{C \in \delta_B} k_C \right) =$$

$$= \sum_{\pi \in L(A)} \mathbf{m}(\pi) \sum_{\delta \leq_{\Pi} \pi} \prod_{B \in \delta} k_B.$$

For every $\delta \in \Pi(A)$ let $\bar{\delta}$ denote the smallest element of $L(A)$ such that $\delta \leq_{\Pi} \bar{\delta}$. Then, by changing the order of summation, the above equation can be rewritten as

$$\ell_A = \sum_{\delta \in \Pi(A)} \prod_{B \in \delta} k_B \left(\sum_{\pi \geq_L \bar{\delta}} \mathbf{m}(\pi) \right).$$

By (7) the sum in brackets vanishes whenever $\bar{\delta} \neq A$. Therefore the whole expression is equal to $\sum_{\delta \in \Pi^*} \prod_{B \in \delta} k_B$. \square

Example 22. Let $n = 4$ and let L be the lattice of all set partitions in Figure 2. The only partitions of $\Pi([4])$ which are not in L are $13|24$ and $14|23$. Hence, they are also the only partitions satisfying the condition $[\pi, [4]] \cap L = [4]$. This, by Proposition 21, gives the formula for t_{1234} given in (16).

5. Basic properties of L -cumulants

In this section we show that L -cumulants satisfy properties similar to (P1)-(P4). The following lemma is central to most of the proofs of this section. It was first formulated by Weisner [24] in a special case and then generalized by Rota [13] for general lattices (see the corollary on page 351 therein).

Lemma 23. *Let L be a finite lattice with at least two elements, and let $\pi_0 \in L$ be such that $\pi_0 \neq \hat{1}$. Then for any $\delta \in L$*

$$\sum_{\pi: \pi \wedge \pi_0 = \delta} \mathbf{m}(\pi) = 0.$$

A special case of this result, when $\delta = \hat{0}$, is given in [22, Corollary 3.9.3]. It is a useful exercise to see that the proof given there generalizes to provide a proof of Lemma 23.

5.1. Independence and semi-invariance

To show that property (P1) holds for L -cumulants we first prove a more algebraic version of this result. This result is directly linked to the definition of independence formulated in terms of moments in (5).

Theorem 24. *Consider the L -cumulant of $X = (X_1, \dots, X_n)$ as in Definition 19. The following are equivalent:*

- (i) *There exists a partition $\pi_0 \in L$ such that $\pi_0 \neq [n]$ and for every $\pi \in L$ we have that $\mu(\pi) = \mu(\pi \wedge \pi_0)$,*

(ii) $\mu_I = \mu(\pi_0(I))$ for all $I \subseteq [n]$,

(iii) $\ell(\pi) = 0$ for all $\pi \not\leq \pi_0$,

(iv) $\ell_I = 0$ unless I is contained in a single block of π_0 .

Proof. The equivalence of (i) and (ii) follows from the fact that $\mu(\pi)$ is a multiplicative function of L . Hence (i) \Rightarrow (ii) follows by taking $\pi = [n]$ and then constraining to elements of I . The opposite implication follows by taking I to be blocks of π . We now prove (i) \Rightarrow (iii). Using Definition 19 we obtain

$$\begin{aligned} \ell(\nu) &= \sum_{\pi \leq \nu} \mathbf{m}(\pi, \nu) \mu(\pi) = \sum_{\pi \leq \nu} \mathbf{m}(\pi, \nu) \mu(\pi \wedge \pi_0) = \\ &= \sum_{\delta \leq \nu \wedge \pi_0} \left(\sum_{\pi \wedge \pi_0 = \delta} \mathbf{m}(\pi, \nu) \right) \mu(\delta), \end{aligned}$$

where the inner sum in the last expression is over all π in $[\hat{0}, \nu]$ such that $\pi \wedge \pi_0 = \delta$ (or $\pi \wedge (\pi_0 \wedge \nu) = \delta$). To show (iii), we are interested only in $\nu \not\leq \pi_0$ and hence we can assume that $\nu \neq \hat{0}$. The interval $[\hat{0}, \nu] \subseteq L$ is a lattice with at least two elements, and, whenever $\nu \not\leq \pi_0$, also $\pi_0 \wedge \nu \neq \nu$. Therefore, by Lemma 23 for all $\delta \leq \nu \wedge \pi_0$ the sum $\sum_{\pi \wedge \pi_0 = \delta} \mathbf{m}(\pi, \nu)$ vanishes. Hence $\ell(\nu) = 0$ unless $\nu \leq \pi_0$.

To show (iii) \Rightarrow (i) note that if $\ell(\delta) = 0$ for all $\delta \not\leq \pi_0$ then for every $\pi \in L$

$$\mu(\pi) = \sum_{\delta \leq \pi} \ell(\delta) = \sum_{\delta \leq \pi \wedge \pi_0} \ell(\delta) = \mu(\pi \wedge \pi_0).$$

To see that (iv) follows from (i) and (iii), apply (i) with $L(I)$ in place of L . If I is not contained in a block of π_0 then $\pi_0(I)$ is not the maximal element of $L(I)$ and by (i) this gives $\mu(\pi) = \mu(\pi \wedge \pi_0(I))$ for every $\pi \in L(I)$. Now $\ell_I = 0$ by (iii).

Finally we show that (iv) implies (ii) using induction with respect to $|I|$. If $I = \{i, j\}$ such that i and j lie in different blocks of π_0 then $\pi_0(I) = i|j$. Since $\ell_{ij} = \mu_{ij} - \mu_i \mu_j = 0$, (ii) holds if $|I| = 2$. Suppose now that (ii) holds for all $|I| < d$ and let now $I \subseteq [n]$ be such that $|I| = d$ and $\pi_0(I) \neq I$ (otherwise (ii) holds trivially). By (20) we have

$$\ell_I = \sum_{\pi \in L(I)} \mathbf{m}(\pi) \mu(\pi).$$

If $\pi < I$ then $\mu(\pi)$ is a product of some μ_B , where $|B| < d$ and hence by assumption $\mu(\pi) = \mu(\pi \wedge \pi_0(I))$. We can rewrite the above equation as

$$\ell_I = \mu_I - \mu(\pi_0(I)) + \sum_{\pi \in L(I)} \mathbf{m}(\pi) \mu(\pi \wedge \pi_0(I)). \tag{22}$$

The last summand can be rewritten as

$$\sum_{\delta \leq \pi_0(I)} \left[\sum_{\pi \wedge \pi_0(I) = \delta} \mathbf{m}(\pi) \right] \mu(\delta),$$

which is zero by Lemma 23 because $\pi_0(I) \neq I$. Therefore, (22) becomes $\ell_I = \mu_I - \mu(\pi_0(I))$. Since $\ell_I = 0$ by assumption, we obtain that (ii) holds for $|I| = d$ and hence it holds for all $I \subseteq [n]$. \square

This result gives an immediate corollary which generalizes property (P1) of the classical cumulants.

Proposition 25. *Suppose there exists a partition $\pi_0 \in L$ such that $\perp\!\!\!\perp_{B \in \pi_0} X_B$. Then $\ell(\pi) = 0$ for all $\pi \not\leq \pi_0$ or equivalently $\ell_A = 0$ unless all the elements of A are contained in a single block of π_0 .*

This proposition shows one of the important features of L -cumulants. For cumulants, by (P1), all marginal independencies imply that $k_{1\dots n} = 0$. In the case of L -cumulants only some of the independencies imply vanishing (see Example 10). Hence, this new coordinate system can be designed to better fit the model under consideration. This concept will be explained in more detail for tree cumulants in Section 6.

We formulate an additional condition on the family of lattices \mathbf{L} , which we require to hold only when this is explicitly stated.

(C1) For every $A \in \mathcal{A}(\mathcal{X})$ and every $i \in A$ the split $i|(A \setminus i)$ is in $L(A)$.

Among the partitions in Definition 4 only the lattice of interval partitions does not satisfy (C1).

Proposition 26 (Semi-invariance). *Let \mathbf{L} satisfy (C1) and $\tilde{X} = X + a$, where $a \in \mathbb{R}^n$ is any constant vector. Denote by $\tilde{\ell}_A$ the L -cumulant of \tilde{X}_A . Then $\tilde{\ell}_i = \ell_i + a_i$ for all $i = 1, \dots, n$ and $\tilde{\ell}_A = \ell_A$ for any multiset $A \in \mathcal{A}(\mathcal{X})$ such that $|A| \geq 2$.*

Proof. Without loss of generality assume $A = [n]$. Since $a = \sum a_i e_i$, where the e_i 's are the unit vectors in \mathbb{R}^n , it suffices to prove this result only in the case when a is such that a_1 is the only non-zero entry. In this case write $\tilde{X}_1 = X_1 + a_1$ as $X_1 - \mu_1 + (a_1 + \mu_1)$, where $\mu_1 = \mathbb{E}X_1$ and $a_1 + \mu_1 = \mathbb{E}\tilde{X}_1$. Hence, if the split $\pi_0 = 1|\{2, \dots, n\} \in L$ then for every $\pi \in L$,

$$\tilde{\mu}(\pi) = \mu(\pi) - \mu(\pi \wedge \pi_0) + \tilde{\mu}(\pi \wedge \pi_0).$$

It follows that

$$\begin{aligned} \tilde{\ell}_{1\dots n} &= \sum_{\pi \in L} \mathbf{m}(\pi) \mu(\pi) - \\ &\quad - \sum_{\pi \in L} \mathbf{m}(\pi) \mu(\pi \wedge \pi_0) + \sum_{\pi \in L} \mathbf{m}(\pi) \tilde{\mu}(\pi \wedge \pi_0). \end{aligned} \tag{23}$$

Since L is a lattice and $\pi_0 \neq [n]$, by Lemma 23 we have that $\sum_{\pi \wedge \pi_0 = \nu} \mathbf{m}(\pi) = 0$ for each $\nu \in L$ and hence the second and third summand in (23) are zero. The proof is completed because the first summand is exactly $\ell_{1\dots n}$. \square

The following result shows that the central moments are L -cumulants induced by the lattice of one-cluster partitions $\mathcal{C}([n])$.

Proposition 27. *Let X be a random vector with values in \mathcal{X} . Then the central moments μ'_A for $|A| \geq 2$ are equal to the corresponding L -cumulants induced by $\mathbf{C} = (\mathcal{C}(A))_{A \in \mathcal{A}(\mathcal{X})}$.*

Proof. Denote by \mathfrak{c} the L -cumulants induced by the family \mathbf{C} of one-cluster partition lattices. Let $A \in \mathcal{A}(\mathcal{X})$ be such that $|A| \geq 2$. Since every split of the form $i|(A \setminus i)$ is a one-cluster partition, by Proposition 26, we can write \mathfrak{c} in terms of the central moments

$$\mathfrak{c}_A = \sum_{\pi \in \mathcal{C}(A)} \mathfrak{m}(\pi) \prod_{B \in \pi} \mu'_A(B) \quad \text{for all } |A| \geq 2.$$

However, $\mu'_i = 0$ for every $i \in [n]$ and hence the only non-zero term of the above sum is where $\pi = A$, which proves that $\mathfrak{c}_A = \mu'_A$. \square

The correspondence between the lattice of one-cluster partitions and central moments gives also the following explicit, simple and computationally efficient formula for central moments in terms of moments.

Lemma 28. *Let X be a random vector with values in \mathcal{X} . For every $A \in \mathcal{A}(\mathcal{X})$ such that $|A| \geq 2$ we have:*

$$\mu'_A = \sum_{B \subseteq A} (-1)^{|A \setminus B|} \mu_B \prod_{i \in A \setminus B} \mu_i. \tag{24}$$

Proof. Use (8) and Proposition 27 to write

$$\mu'_A = \sum_{\emptyset < \pi \in \mathcal{C}(A)} (-1)^{|\pi|-1} \prod_{B \in \pi} \mu_B + (-1)^{|A|-1} (|A| - 1) \prod_{i \in A} \mu_i.$$

Let B_0 be the distinguished non-singleton block in each of the product $\prod_{B \in \pi} \mu_B$ above. Then $|\pi|-1 = |A \setminus B_0|$. Hence, every $\prod_{B \in \pi} \mu_B$ corresponds to some $\mu_{B_0} \prod_{i \in A \setminus B_0} \mu_i$ in (24) with the same coefficient. The remaining part is to check that the coefficient of $\prod_{i \in A} \mu_i$ is also the same, but this is an easy check. \square

Example 29. Let $A = \{1, 1, 2, 2\}$ and list all multisubsets of A as defined in the beginning of Section 4. We easily check that

$$\mu'_{1122} = \mu_{1122} - 2\mu_1\mu_{122} - 2\mu_2\mu_{112} + \mu_{11}\mu_2^2 + 4\mu_{12}\mu_1\mu_2 + \mu_1^2\mu_{22} - 3\mu_1^2\mu_2^2,$$

which can be verified also by hand.

5.2. Multilinear transformations

By property (P3) cumulants behave nicely under multilinear transformations. In this section, to study similar properties for general L -cumulants, we restrict to \mathbf{L} satisfying the following condition.

(C2) For every $A \in \mathcal{A}(\mathcal{X})$ the lattice $L(A)$ is isomorphic to $L([d])$, where $d = |A|$.

This property is satisfied by construction for $\mathbf{\Pi}$, \mathbf{I} , \mathbf{NC} , and \mathbf{C} . If (C2) holds then, for every d -tuple $(i_1, \dots, i_d) \in [n]^d$ we define $\mathcal{L}^{(d)}$ as a $n \times \dots \times n$ tensor of the form

$$\mathcal{L}_{i_1 \dots i_d}^{(d)} = \sum_{\pi \in L([d])} \mathfrak{m}(\pi) \prod_{B \in \pi} \tilde{\mu}_{i_B}. \tag{25}$$

Note that in general $\mathcal{L}_{i_1 \dots i_d}^{(d)}$ may differ from $\ell_{i_1 \dots i_d}$. For example if $L = \mathcal{I}([3])$ then $\ell_{213} = \ell_{123}$ because the definition of L -cumulants does not depend on the ordering of the elements in $[n]$. On the other hand, we have $\mathcal{L}_{123}^{(d)} \neq \mathcal{L}_{213}^{(d)}$ because

$$\mathcal{L}_{123}^{(d)} = \mu_{123} - \mu_1\mu_{23} - \mu_{12}\mu_3 + \mu_1\mu_2\mu_3$$

and

$$\mathcal{L}_{213}^{(d)} = \mu_{123} - \mu_2\mu_{13} - \mu_{12}\mu_3 + \mu_1\mu_2\mu_3.$$

The following proposition shows that the tensor $\mathcal{L}^{(d)}$, for any $d \geq 1$, under linear mappings transforms as a contravariant tensor.

Proposition 30. *Let $X = (X_1, \dots, X_n)$ be a random vector. Consider L -cumulants defined by \mathbf{L} satisfying (C2). Let $Q = [q_{ij}] \in \mathbb{R}^{m \times n}$ and $\tilde{X} = QX \in \mathbb{R}^m$. Define $[\tilde{\mu}_A]$, $[\tilde{\ell}_A]$ and $\tilde{\mathcal{L}}^{(d)}$ as counterparts of $[\mu_A]$, $[\ell_A]$ and $\mathcal{L}^{(d)}$ for \tilde{X} accordingly. Then for each $d \geq 1$,*

$$\tilde{\mathcal{L}}^{(d)} = Q \cdot \mathcal{L}^{(d)},$$

where $Q \cdot \mathcal{L}^{(d)}$ is the multilinear action on a d -dimensional tensor defined by

$$(Q \cdot \mathcal{L}^{(d)})_{i_1 \dots i_d} = \sum_{j_1=1}^n \dots \sum_{j_d=1}^n q_{i_1 j_1} \dots q_{i_d j_d} \mathcal{L}_{j_1 \dots j_d}^{(d)} \tag{26}$$

for each $d \geq 1$ and $i_1, \dots, i_d \in [m]$.

Proof. By (25) we have

$$(Q \cdot \mathcal{L}^{(d)})_{i_1 \dots i_d} = \sum_{j_1=1}^n \dots \sum_{j_d=1}^n q_{i_1 j_1} \dots q_{i_d j_d} \left(\sum_{\pi \in L([d])} \mathfrak{m}(\pi) \prod_{B \in \pi} \mu_{j_B} \right).$$

Write μ_{j_B} explicitly as $\mathbb{E} [\prod_{b \in B} X_{j_b}]$. Then, using (C2), after changing the ordering of products and summations we obtain

$$(Q \cdot \mathcal{L}^{(d)})_{i_1 \dots i_d} = \sum_{\pi \in L([d])} \mathfrak{m}(\pi) \prod_{B \in \pi} \mathbb{E} \left[\prod_{b \in B} \left(\sum_{j_b=1}^n q_{i_b j_b} X_{j_b} \right) \right].$$

Since $\tilde{X}_{i_b} = \sum_{j_b=1}^n q_{i_b j_b} X_{j_b}$ we obtain

$$(Q \cdot \mathcal{L}^{(d)})_{i_1 \dots i_d} = \sum_{\pi \in L([d])} \mathfrak{m}(\pi) \prod_{B \in \pi} \tilde{\mu}_{i_B} = (\tilde{\mathcal{L}}^{(d)})_{i_1 \dots i_d},$$

which finishes the proof. \square

Although for some \mathbf{L} the property (P3) may not hold, the homogeneity holds for all L -cumulants. Thus, if $\tilde{X} = (\lambda_1 X_1, \dots, \lambda_n X_n)$ for some $\lambda = (\lambda_1, \dots, \lambda_n) \in (\mathbb{R}^*)^n$ then $\tilde{\ell}_A = \prod_{i \in A} \lambda_i \ell_A$ for every $A \in \mathcal{A}(\mathcal{X})$.

5.3. Conditional L -cumulants

Suppose we are given the conditional cumulants of $X = (X_1, \dots, X_n)$ conditional on some random variable Y and we want to obtain the unconditional cumulants. This is a common problem with hidden variable models. On the level of moments this relationship is straightforward since

$$\mu_A = \mathbb{E}\left[\prod_{i \in A} X_i\right] = \mathbb{E}\left[\mathbb{E}\left[\prod_{i \in A} X_i | Y\right]\right]$$

for every multiset A . For cumulants, or more generally for L -cumulants, the situation is a bit more complicated.

For every multiset $A \in \mathcal{A}(\mathcal{X})$ denote by k_A^Y the *conditional cumulant of X_A given Y* , that is a cumulant computed as in Definition 19 but with moments replaced by conditional moments. Note that each k_A^Y is itself a random variable. For any $\pi \in \Pi(A)$, by \hat{k}_π denote the cumulant of the random vector $(k_B^Y)_{B \in \pi}$. It is known from [1] that for every $A \in \mathcal{A}(\mathcal{X})$:

$$k_A = \sum_{\pi \in L(A)} \hat{k}_\pi. \tag{27}$$

This in particular generalizes the well-known formula

$$\text{Cov}(X, Z) = \mathbb{E}[\text{Cov}(X, Z | Y)] + \text{Cov}(\mathbb{E}[X | Y], \mathbb{E}[Z | Y]).$$

In Theorem 31 we give a purely combinatorial proof of (27). For our purposes it is slightly more constructive than a similar proof of the same result in [19]. Also it immediately enables us to formulate this result for L -cumulants in the case when \mathbf{L} satisfies the following property.

(C3) For every $n \geq 0$ and each $\pi \in L$ the interval $[\pi, [n]] \subseteq L$ is isomorphic to $L([\pi])$.

This property is satisfied for $\mathbf{\Pi}$ (see [22, Example 3.10.4]). A sufficient condition for \mathbf{L} to satisfy (C3) is that for every $n \geq 0$ the lattice L forms a join subsemilattice of $\Pi([n])$. Therefore, \mathbf{I} as well as the lattice of tree partitions for sufficiently regular trees (for example caterpillars) both satisfy the property. Condition (C3) does not hold however for the lattice of one-cluster partitions, (general) tree partitions and non-crossing partitions.

For every multiset $A \in \mathcal{A}(\mathcal{X})$ denote by ℓ_A^Y the conditional L -cumulant of X_A given Y . For any $\pi \in L(A)$, by $\hat{\ell}_\pi$ denote the L -cumulant of the random vector $(\ell_B^Y)_{B \in \pi}$.

Theorem 31 (Brillinger’s formula for L -cumulants). *Let $X = (X_1, \dots, X_n)$ be a random vector and Y be a random variable. If \mathbf{L} satisfies (C3) then*

$$\ell_{1\dots n} = \sum_{\pi \in L} \hat{\ell}_\pi.$$

Proof. Since $\mu_B = \mathbb{E}\mu_B^Y$, by (21) we obtain the identity

$$\mu_B = \mathbb{E}\mu_B^Y = \sum_{\delta \in L(B)} \mathbb{E} \left[\prod_{C \in \delta} \ell_C^Y \right]. \tag{28}$$

Using (20) and replacing (28) for each μ_B we can write

$$\begin{aligned} \ell_{1\dots n} &= \sum_{\pi \in L} \mathbf{m}(\pi) \prod_{B \in \pi} \left(\sum_{\delta \in L(B)} \mathbb{E} \left[\prod_{C \in \delta} \ell_C^Y \right] \right) = \\ &= \sum_{\pi \in L} \mathbf{m}(\pi) \sum_{\delta \leq \pi} \prod_{B \in \pi} \mathbb{E} \left[\prod_{C \in \delta(B)} \ell_C^Y \right], \end{aligned}$$

where $\delta(B)$ denotes the partition $\delta \in L$ constrained to $B \in \pi$. We change the order of summation to obtain

$$\ell_{1\dots n} = \sum_{\delta \in L} \left(\sum_{\pi \geq \delta} \mathbf{m}(\pi) \prod_{B \in \pi} \mathbb{E} \left[\prod_{C \in \delta(B)} \ell_C^Y \right] \right). \tag{29}$$

For each $\delta = C_1 | \dots | C_r \in L$ denote the set of its blocks by $\mathbb{B}_\delta = \{C_1, \dots, C_r\}$. By (C3) the interval $[\delta, [n]]$ is isomorphic to $L(\mathbb{B}_\delta)$ which is isomorphic to $L([\delta])$ and hence the expression in brackets in (29) can be rewritten as

$$\sum_{\nu \in L(\mathbb{B}_\delta)} \mathbf{m}_{\mathbb{B}_\delta}(\nu, \mathbb{B}_\delta) \prod_{B \in \nu} \mathbb{E} \left[\prod_{C \in B} \ell_C^Y \right],$$

which by definition is equal to $\hat{\ell}_\delta$. □

If (C3) does not hold and we want to perform some efficient conditional computations, we can still use the classical Brillinger’s formula for cumulants and then translate them back to L -cumulants using Proposition 21. Moreover, for some special statistical models the following result may be useful. It works for all families \mathbf{L} .

Proposition 32. *Let $X = (X_1, \dots, X_n)$ be a random vector and Y a random variable. If $X_1 \perp\!\!\!\perp \dots \perp\!\!\!\perp X_n | Y$, then*

$$\ell_{1\dots n} = \hat{\ell}_{1|2|\dots|n},$$

where by definition $\hat{\ell}_{1|2|\dots|n}$ is the L -cumulant of the random vector $(\ell_1^Y, \dots, \ell_n^Y) = (\mu_1^Y, \dots, \mu_n^Y)$.

Proof. Since $X_1 \perp\!\!\!\perp \dots \perp\!\!\!\perp X_n | Y$, by Proposition 25, $\ell_C^Y = 0$ unless $|C| = 1$. Moreover, we have

$$\mu_B = \mathbb{E} \left[\prod_{i \in B} \mu_i^Y \right].$$

Using (20) and replacing the above identity for each μ_B we can write

$$\ell_{1\dots n} = \sum_{\pi \in L} \mathbf{m}(\pi) \prod_{B \in \pi} \mathbb{E} \left[\prod_{i \in B} \mu_i^Y \right].$$

But since $\ell_i^Y = \mu_i^Y$, the right hand side in the above equation is exactly the L -cumulant of the random vector $(\ell_1^Y, \dots, \ell_n^Y)$. \square

To see how this result may be relevant in geometry see Example 17.

6. Tree cumulants and hidden Markov processes

In this section we complement the discussion of tree cumulants and show how they can be used to analyze more general processes on trees.

6.1. Tree models

Let T^r be a rooted tree with vertex set V and edge set E , that is a tree with one distinguished node $r \in V$ called the root and all the edges directed away from r . Let $X = (X_v)_{v \in V}$ be a vector of binary random variables with values 0 and 1. Consider the Bayesian network for X represented by T^r . Each node v corresponds to a random variable X_v and the structure of T^r imposes some constraints on the joint distribution of X (see for example [7]). Define \mathcal{M}_T as the model obtained from this Bayesian network by taking the marginal distributions over the leaves of T^r . We call \mathcal{M}_T the *two-state general Markov model* (for example [17, Chapter 8]). We omit the rooting in the notation because the model does not depend on the rooting. In other words, for any alternative rooting the induced parametrization will lead to the same model.

The parametric formulation of the model is obtained by expressing the marginal distribution of X over the leaves of T^r in terms of the marginal distribution of the root r and conditional distributions of each $v \in V \setminus \{r\}$ given its parent in T^r denoted by $\text{pa}(v)$. Assume that T^r has n leaves and label them by elements of $[n]$. The distribution over the set of leaves satisfies

$$p(x_1, \dots, x_n) = \sum_{\mathcal{H}} p_r(x_r) \prod_{v \in V \setminus r} p_{v|\text{pa}(v)}(x_v | x_{\text{pa}(v)}), \quad (30)$$

where \mathcal{H} is the set of all $x \in \{0, 1\}^V$ such that the restriction to the leaves of T is equal to (x_1, \dots, x_n) . The model is given as the image of (30) in $\Delta_{\mathcal{X}}$, where each point corresponds to a different choice of values for conditional probabilities on the right hand side of this parametrization. If m denotes the number of inner nodes of T then this parametrization has 2^m terms. For large trees this is a big polynomial which complicates the geometric and algebraic analysis of these models.

The two-state general Markov model can be equivalently defined by a set of conditional independence statements. This follows from the general theory of graphical models (see [7, Section 3.2.2]). We say that two disjoint subsets A, B of the set of vertices V of T are *separated* by another subset C if every undirected path from a node in A to a node in B

necessarily crosses C . The set of all conditional independence statements which define the general Markov model are given by all $A \perp\!\!\!\perp B|C$ for all disjoint subsets $A, B, C \subseteq V$ such that C separates A and B . For example the 4-star tree model discussed in Section 3.3 is defined by $X_1 \perp\!\!\!\perp X_2 \perp\!\!\!\perp X_3 \perp\!\!\!\perp X_4|Y$ because the inner node separates all the leaves from each other.

Before we recall the main result of [27], let us give some intuition on why tree cumulants may be helpful in the study of tree models. Suppose that for some edge (u, v) in T^r we impose on the model \mathcal{M}_T that in addition $X_u \perp\!\!\!\perp X_v$. This corresponds to removing the edge (u, v) from T^r and considering the model of the induced forest. Let $A|B$ be the split of the set of leaves $[n]$ induced by removing the edge (u, v) . Then the independence statement $X_u \perp\!\!\!\perp X_v$ implies also that $X_A \perp\!\!\!\perp X_B$.

Example 33. Let T be the quartet tree in Figure 4 rooted in a . The independence $(X_1, X_2) \perp\!\!\!\perp (X_3, X_4)$ defines a valid submodel of the tree model for T . This submodel is defined by requesting $X_a \perp\!\!\!\perp X_b$ and hence it is given as the image of the subspace of the parameter space restricted to $p_{b|a}(1|0) = p_{b|a}(1|1)$.

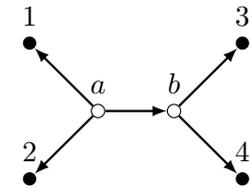


Figure 4: A quartet tree.

By Proposition 9 there exists a tree partition π_0 such that $\perp\!\!\!\perp_{B \in \pi_0} X_B$ if and only if $t_I = 0$ whenever $I \subseteq [n]$ is not completely contained in one of the blocks of π_0 . In Example 33, because $12|34$ is a valid tree partition, the marginal independence $(X_1, X_2) \perp\!\!\!\perp (X_3, X_4)$ holds if and only if $t_I = 0$ for all $I \subseteq \{1, 2, 3, 4\}$ such that I is not contained neither in $\{1, 2\}$ nor $\{3, 4\}$. Hence all $t_{13}, t_{14}, t_{23}, t_{24}, t_{134}, t_{234}, t_{123}, t_{124}$ and t_{1234} vanish whenever $p_{b|a}(1|0) = p_{b|a}(1|1)$. These kind of considerations help to understand why tree cumulants are helpful for describing the two-state general Markov models. They also help to intuitively understand the result in Theorem 34, which we now state formally.

Let $\eta_{uv} = p_{v|u}(1|1) - p_{v|u}(1|0)$ for each $(u, v) \in E$. As we have shown $\eta_{uv} = 0$ if and only if $X_u \perp\!\!\!\perp X_v$. Moreover, let $\bar{\mu}_v = 1 - 2\mu_v$ for $v \in V$.

Theorem 34 (Zwiernik, Smith [27]). *Let T be trivalent tree. Then the two-state general Markov model \mathcal{M}_T can be equivalently expressed in the space of tree cumulants by $l_i = \mu_i = \frac{1}{2}(1 - \bar{\mu}_i)$ for $i = 1, \dots, n$; and for all $|I| \geq 2$*

$$t_I = \frac{1}{4}(1 - \bar{\mu}_{r(I)}^2) \prod_{\deg(v)=3} \bar{\mu}_v \prod_{(u,v) \in E(I)} \eta_{uv},$$

where $V(I)$ and $E(I)$ denotes vertex and edge sets of the tree $T(I)$, $r(I)$ is the root of $T(I)$ and $\deg(v)$ denotes the valency of v in $T(I)$.

Example 35. Let T^r be a quartet tree in Figure 4. Then by Theorem 34 we have for example $t_{12} = \frac{1}{4}(1 - \bar{\mu}_a^2)\eta_{a1}\eta_{a2}$, $t_{13} = \frac{1}{4}(1 - \bar{\mu}_a^2)\eta_{a1}\eta_{ab}\eta_{b3}$, $t_{34} = \frac{1}{4}(1 - \bar{\mu}_b^2)\eta_{b3}\eta_{b4}$ and

$$t_{1234} = \frac{1}{4}(1 - \bar{\mu}_a^2)\bar{\mu}_a\bar{\mu}_b\eta_{1a}\eta_{2a}\eta_{ab}\eta_{b3}\eta_{b4}.$$

We also infer from this that

$$t_{I \cup J} t_{I' \cup J'} - t_{I \cup J'} t_{I' \cup J} = 0$$

for all $I, I' \in \{\{1\}, \{2\}, \{1, 2\}\}$ and $J, J' \in \{\{3\}, \{4\}, \{3, 4\}\}$.

Theorem 34 can be applied only for trivalent trees and hence it does not hold for n -star tree models discussed earlier (see also Remark 18). We can use however the fact that any non-trivalent tree model is a submodel of some model of a trivalent tree. Thus, if T is not trivalent then we take any trivalent tree T^* such that T can be obtained from T^* by edge contractions. Now the two-state general Markov model for T , when expressed in tree cumulants of T^* , is parametrized by $\ell_i = \mu_i$ for $i = 1, \dots, n$, and for all $|I| \geq 2$

$$t_I = \frac{1}{4}(1 - \bar{\mu}_{r(I)}^2) \prod_{v \in V(I) \setminus I} \bar{\mu}_v^{\deg(v)-2} \prod_{(u,v) \in E(I)} \eta_{uv}.$$

In the quartet tree of Example 35 we can contract the edge (a, b) to obtain the 4-star tree in Figure 3. This contraction corresponds to the subspace of the parameter space given by $\bar{\mu}_a = \bar{\mu}_b$ and $\eta_{ab} = 1$. This induces the parametrization of the secant variety given in Example 17. The same can be obtained for any n -star tree model with $n \geq 4$. For more details see [27].

6.2. Binary hidden Markov processes

We now show that tree cumulants can be useful also for other related statistical models. We consider models with an underlying two-state Markov chain which is not observed, where the observed variables are independent given this Markov chain. An example is given by the hidden Markov model or some simple cases of Markov switching models without autoregressive terms (see for example [6]). In this section we refer to all these models as *binary hidden Markov processes*.

Consider tree cumulants induced by the caterpillar tree and define the *normalized tree cumulants* as

$$\bar{t}_I = \prod_{i \in I} \frac{1}{\sqrt{k_{ii}}} t_I \quad \text{for all } I \subseteq [n],$$

which is always well defined if all the variables in the system are non-degenerate (a degenerate random variable takes only one value with nonzero probability). With this definition $\rho_{ij} := \bar{t}_{ij}$ is just the usual *correlation* between X_i and X_j , and $\gamma_i := \bar{t}_{iii}$ is the *skewness* of X_i .

In this section we deal with an observed vector $X = (X_1, \dots, X_n)$ and a hidden vector $H = (H_1, \dots, H_n)$. Since we need to consider mixed tree cumulants involving indices from both vectors, we introduce the following convention. Whenever an index involves i referring to H_i we write it as \underline{i} . Hence for example $k_{ij} = \text{Cov}(H_i, X_j)$, $k_{\underline{ii}} = \text{Var}(H_i)$, $k_{ii} = \text{Var}(X_i)$, $\gamma_{\underline{i}} = \mathbb{E}(H_i - \mathbb{E}H_i)^3 / \text{Var}(H_i)^{3/2}$, and $\mu'_{\underline{B}} = \mathbb{E}[\prod_{i \in B} (H_i - \mathbb{E}H_i)]$.

It is well known that for every random variable X , if Y is binary, then

$$\mathbb{E}(X|Y) = \mathbb{E}X + \text{Cov}(X, Y)(\text{Var}(Y))^{-1}(Y - \mathbb{E}Y), \tag{31}$$

where $\text{Cov}(X, Y)(\text{Var}(Y))^{-1}$ is the linear regression coefficient of X with respect to Y . The following proposition shows that the hidden Markov process has an elegant formulation and all its normalized tree cumulants are parametrized by correlations and skewnesses.

Proposition 36. *Let $X = (X_1, \dots, X_n)$ be a random vector and $H = (H_1, \dots, H_n)$ a binary random vector (both non-degenerate). Assume that $X_1 \perp\!\!\!\perp \dots \perp\!\!\!\perp X_n | H$ and the conditional distribution of X_i given H depends only on H_i for $i = 1, \dots, n$. Moreover, let H form a Markov chain. Then for every $I = \{i_1, \dots, i_d\}$ such that $1 \leq i_1 < \dots < i_d \leq n$ the corresponding normalized tree cumulant satisfies*

$$\bar{\mathfrak{t}}_I = \prod_{j=2}^{d-1} \gamma_{\underline{i_j}} \prod_{i=i_1}^{i_d-1} \rho_{\underline{i_{i+1}}} \prod_{i \in I} \rho_{\underline{i}}.$$

Proof. Before we prove the proposition we formulate the following result.

Lemma 37. *Suppose that $X = (X_1, \dots, X_n)$ is a binary random vector such that $i \perp\!\!\!\perp j \perp\!\!\!\perp C | r$ for some disjoint $i, j, r \in [n]$ and $C \subseteq [n]$. Let $\eta_{rA} = \mu'_{rA} k_{rr}^{-1}$ for every $A \subseteq [n]$ and $\tau_r = k_{rrr} k_{rr}^{-1}$. Then*

$$\mu'_{ijC} = \eta_{ri} \eta_{rj} k_{rr} \mu'_C + \eta_{ri} \eta_{rj} \mu'_{rC} \tau_r.$$

Proof. Let $U_A := \prod_{i \in A} (X_i - \mathbb{E}X_i)$ for every $A \subseteq [n]$. The conditional independence $i \perp\!\!\!\perp j \perp\!\!\!\perp C | r$ implies

$$\mathbb{E}[U_{ijC} | U_r] = \mathbb{E}[U_i | U_r] \mathbb{E}[U_j | U_r] \mathbb{E}[U_C | U_r].$$

Using (31) for the conditional expectations on the right hand side and then taking expectations on both sides yields

$$\mu'_{ijC} = \eta_{ri} \eta_{rj} k_{rr} \mu'_C + \eta_{ri} \eta_{rj} \eta_{rC} k_{rrr}.$$

Replace $\eta_{rC} = \mu'_{rC} k_{rr}^{-1}$ to obtain the formula in Lemma 37. □

To prove Proposition 36 first assume that $I = [n]$ and by L denote the lattice of tree partitions of the caterpillar tree with n leaves. We can divide the partitions in L into two groups:

1. partitions with 1 and 2 in two different blocks $1A$ and $2B$, and

2. partitions with 1 and 2 in a single block $12A$

By Remark 16 we can write

$$t_{1\dots n} = \sum_{\pi \in L} m(\pi) \prod_{B \in \pi} \mu'_B. \tag{32}$$

In the first group of partitions we always have either $A = \emptyset$ or $B = \emptyset$. Since $\mu'_1 = \mu'_2 = 0$, for every π in the first group the corresponding summand in (32) is zero. Let $\delta_0 = 12|3| \cdots |n$. The set of all partitions in the second group forms an interval $[\delta_0, [n]]$, which is isomorphic to the set of all tree partitions of the subtree T_2 of T with $n - 1$ leaves given by the hidden vertex $\underline{2}$ and the remaining leaves of T : $3, \dots, n$. This isomorphism is given by replacing each block $12A$ with a block $\underline{2}A$. Denote the lattice of all partitions of T_2 by L_2 . Since $[\delta_0, [n]] \simeq L_2$, the Möbius function on L restricted to this interval is equal to the Möbius function on L_2 .

For every $A \subseteq [n] \setminus \{1, 2\}$ we have that $X_1 \perp\!\!\!\perp X_2 \perp\!\!\!\perp X_A | H_2$ and hence, by Lemma 37

$$\mu'_{12A} = \eta_{21} \eta_{22} k_{22} \mu'_A + \eta_{21} \eta_{22} \mu'_{2A} \tau_2.$$

Therefore, (32) becomes

$$t_{1\dots n} = \sum_{\pi \in [\delta_0, [n]]} m(\pi) \prod_{B \in \pi} \mu'_B \cdot \eta_{21} \eta_{22} k_{22} \mu'_A + \eta_{21} \eta_{22} \tau_2 \sum_{\pi \in L_2} m_2(\pi) \prod_{B \in \pi} \mu'_B. \tag{33}$$

Let π_0 be a split $12|[n] \setminus \{1, 2\}$. For every $\pi \in [\delta_0, [n]]$ the partition $\pi \wedge \pi_0$ is the partition obtained from π by splitting the block $12A$ into two blocks 12 and A . With this notation the first summand in (33) can be rewritten as

$$\sum_{\nu \in [\delta_0, \pi_0]} \left[\sum_{\pi: \pi \wedge \pi_0 = \nu} m(\pi) \right] \prod_{B \in \nu} \mu'_B \cdot \eta_{21} \eta_{22} k_{22}.$$

Since the interval $[\delta_0, [n]]$ forms a lattice then by Lemma 23 the above expression is zero. Since $\sum_{\pi \in L_2} m(\pi) \prod_{B \in \pi} \mu'_B = t_{([n] \setminus \{1, 2\}) \cup \{a\}}$, then the second summand in (33) is

$$t_{1\dots n} = \eta_{21} \eta_{22} \tau_2 t_{([n] \setminus \{1, 2\}) \cup \{2\}}. \tag{34}$$

Using (31) we can also prove that $\eta_{21} = k_{11} \eta_{11} \eta_{12} k_{22}^{-1}$ (use the fact that $X_1 \perp\!\!\!\perp H_2 | H_1$). In the next step we can apply the same procedure as above to express $t_{([n] \setminus \{1, 2\}) \cup \{2\}}$ in (34) in terms of $k_{22}, \eta_{23}, \eta_{33}, \tau_3$ and $t_{([n] \setminus \{1, 2, 3\}) \cup \{3\}}$. We can do it recursively until we obtain

$$t_{1\dots n} = k_{11} \prod_{i=1}^n \eta_{ii} \prod_{i=1}^{n-1} \eta_{i, i+1} \prod_{i=2}^{n-1} \tau_i.$$

Divide both sides by $\sqrt{k_{11} \cdots k_{nn}}$. The main proposition follows for $I = [n]$ after some obvious algebraic rearrangements. In the general case we first use the formula for $[n]$ to conclude that for every $I = \{i_1, \dots, i_d\}$, where $i_1 < \dots < i_d$

$$\bar{t}_I = \prod_{j=2}^{d-1} \gamma_{i_j} \prod_{j=1}^{d-1} \rho_{i_j, i_{j+1}} \prod_{i \in I} \rho_{i, i}.$$

To prove the final formula, it remains to show that

$$\rho_{\underline{i}_j \underline{i}_{j+1}} = \prod_{i=i_j}^{i_{j+1}-1} \rho_{i \underline{i}_{i+1}},$$

which can be proved by induction using (31). \square

This proposition enables us to analyze the moment structure of hidden Markov processes.

Example 38 (Homogeneous binary hidden Markov model). Consider a homogeneous binary hidden Markov model. In this case $H = (H_i)_{i=1}^n$ forms a homogeneous two-state Markov chain which we assume to start from its stationary distribution. Moreover, the conditional distribution of X_i given H_i is the same for every $i = 1, \dots, n$. Under these assumptions the marginal distribution of H_1 is equal to the marginal distribution of H_i for every $i = 2, \dots, n$. Let γ be the skewness of H_1 , $\rho = \text{Corr}(H_1, H_2)$ be the one step correlation of the Markov chain H , and $b = \text{Corr}(H_1, X_1)$. By Proposition 36, for every $d \geq 2$ and $1 \leq i_1 < \dots < i_d \leq n$,

$$\bar{\mathfrak{t}}_{i_1 \dots i_d} = b^d \rho^{i_d - i_1} \gamma^{d-2}. \quad (35)$$

This in turn induces some constraints on the tree cumulants of the observed variables which may be useful to construct simple diagnostic tests for this class of models. For example it is easy to check that

$$\bar{\mathfrak{t}}_{i(i+2)} \bar{\mathfrak{t}}_{j(j+2)} = \bar{\mathfrak{t}}_{k(k+3)} \bar{\mathfrak{t}}_{l(l+1)} \quad \text{for every } i, j, k, l = 1, \dots, n$$

and that $\bar{\mathfrak{t}}_{ij} \bar{\mathfrak{t}}_{ik} \bar{\mathfrak{t}}_{jk} \geq 0$ for all $i < j < k$. The monomial parametrization in (35) enables us to obtain the equations for higher order tree cumulants.

ACKNOWLEDGEMENTS This research was conducted at Warwick University as part of the author's PhD thesis and then at TU Eindhoven, where the author was supported by Jan Draisma's Vidi grant from the Netherlands Organisation for Scientific Research (NWO). The author is grateful to Franz Lehner, Diane Maclagan, Kristian Ranestad, Jim Q. Smith, Bernd Sturmfels, the anonymous referees and the editors for helpful comments and discussions.

References

- [1] David R. Brillinger. The calculation of cumulants via conditioning. *Annals of the Institute of Statistical Mathematics*, 21(1):215–218, 1969.
- [2] Anthony C. Davison. *Statistical models*, volume 11 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 2003.

- [3] Mathias Drton, Bernd Sturmfels, and Seth Sullivant. *Lectures on Algebraic Statistics*. Oberwolfach Seminars Series. Birkhauser Verlag AG, 2009.
- [4] William Feller. *An Introduction to Probability Theory and Applications*, volume 2. John Wiley & Sons, New York, second edition, 1971.
- [5] Dan Geiger, David Heckerman, Henry King, and Christopher Meek. Stratified exponential families: graphical models and model selection. *Ann. Statist.*, 29(2):505–529, 2001.
- [6] James D. Hamilton. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica: Journal of the Econometric Society*, 57(2):357–384, 1989.
- [7] Steffen L. Lauritzen. *Graphical models*, volume 17 of *Oxford Statistical Science Series*. The Clarendon Press Oxford University Press, New York, 1996. Oxford Science Publications.
- [8] Franz Lehner. Free cumulants and enumeration of connected partitions. *European Journal of Combinatorics*, 23(8):1025–1031, 2002.
- [9] Franz Lehner. Cumulants in noncommutative probability theory I. Noncommutative exchangeability systems. *Mathematische Zeitschrift*, 248(1):67–100, 2004.
- [10] Kishiko Maruyama and Fumiaki Shibata. From two-state jump to Gaussian stochastic processes. *Physica A: Statistical and Theoretical Physics*, 149(3):447 – 471, 1988.
- [11] Peter McCullagh. *Tensor methods in statistics*. Monographs on Statistics and Applied Probability. Chapman & Hall, London, 1987.
- [12] Giovanni Pistone and Henry P. Wynn. Cumulant varieties. *Journal of Symbolic Computation*, 41(2):210–221, 2006.
- [13] Gian-Carlo Rota. On the foundations of combinatorial theory I. Theory of Möbius Functions. *Probability Theory and Related Fields*, 2(4):340–368, 1964.
- [14] Gian-Carlo Rota, D. Kahaner, and A. Odlyzko. On the foundations of combinatorial theory. VIII- Finite operator calculus(Umbral/finite operator/ calculus in combinatorial theory of special polynomial sequences as technique for expressing one polynomial set in terms of another). *Journal of Mathematical Analysis and Applications*, 42:684–760, 1973.
- [15] Gian-Carlo Rota and Jianhong Shen. On the combinatorics of cumulants. *J. Combin. Theory Ser. A*, 91(1-2):283–304, 2000. In memory of Gian-Carlo Rota.
- [16] Dmitry Rusakov and Dan Geiger. Asymptotic model selection for naive Bayesian networks. *J. Mach. Learn. Res.*, 6:1–35 (electronic), 2005.

- [17] Charles Semple and Mike Steel. *Phylogenetics*, volume 24 of *Oxford Lecture Series in Mathematics and its Applications*. Oxford University Press, Oxford, 2003.
- [18] Raffaella Settimi and Jim Q. Smith. Geometry, moments and conditional independence trees with hidden variables. *Ann. Statist.*, 28(4):1179–1205, 2000.
- [19] Terence P. Speed. Cumulants and partition lattices. *Austral. J. Statist.*, 25(2):378–388, 1983.
- [20] Roland Speicher. Free probability theory and non-crossing partitions. *Sém. Lothar. Combin.*, 39:Art. B39c, 38 pp. (electronic), 1997.
- [21] Roland Speicher and Reza Woroudi. Boolean convolution. In *Free probability theory (Waterloo, ON, 1995)*, volume 12 of *Fields Inst. Commun.*, pages 267–279. Amer. Math. Soc., Providence, RI, 1997.
- [22] Richard P. Stanley. *Enumerative combinatorics. Volume I*. Number 49 in Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2002.
- [23] Bernd Sturmfels and Piotr Zwiernik. Binary cumulant varieties. *to appear in Annals of Combinatorics*, March 2011.
- [24] Louis Weisner. Abstract theory of inversion of finite series. *Trans. Amer. Math. Soc.*, 38(3):474–484, 1935.
- [25] Piotr Zwiernik. An asymptotic behaviour of the marginal likelihood for general Markov models. *J. Mach. Learn. Res.*, 12:3283–3310, 2011.
- [26] Piotr Zwiernik and Jim Q. Smith. Implicit inequality constraints in a binary tree model. *Electron. J. Statist.*, 5:1276–1312, 2011.
- [27] Piotr Zwiernik and Jim Q. Smith. Tree-cumulants and the geometry of binary tree models. *Bernoulli*, 18(1):290–321, January 2012.