RESEARCH ARTICLE

Lattice conditional independence models and Hibi ideals

Peter Caines[1] | **Fatemeh Mohammadi[2,3]** |
Eduardo Sáenz-de-Cabezón[4] | **Henry Wynn[5]**

[1]Department of Electrical and Computer
Engineering, McGill University,
Montreal, Canada

[2]Department of Mathematics, Algebra
and Geometry, Ghent University, Ghent,
Belgium

[3]Department of Mathematics and
Statistics, UiT – The Arctic University of
Norway, Tromsø, Norway

[4]Departamento de Matemáticas y
Computación, Universidad de La Rioja,
Logroño, Spain

[5]London School of Economics and
Political Science, London, UK

Correspondence
Fatemeh Mohammadi, Department of
Mathematics, Algebra and Geometry,
Ghent University, Ghent, Belgium and
Department of Mathematics and
Statistics, UiT – The Arctic University of
Norway, Tromsø, Norway.
Email: fatemeh.mohammadi@ugent.be

Abstract

Lattice conditional independence models [Andersson and Perlman, *Lattice models for conditional independence in a multivariate normal distribution*, Ann. Statist. **21** (1993), 1318–1358] are a class of models developed first for the Gaussian case in which a distributive lattice classifies all the conditional independence statements. The main result is that these models can equivalently be described via a transitive directed acyclic graph (TDAG) in which, as is normal for causal models, the conditional independence is in terms of conditioning on ancestors in the graph. We demonstrate that a parallel stream of research in algebra, the theory of Hibi ideals, not only maps directly to the lattice conditional independence models but gives a vehicle to generalise the theory from the linear Gaussian case. Given a distributive lattice (i) each conditional independence statement is associated with a Hibi relation defined on the lattice, (ii) the directed graph is given by chains in the lattice which correspond to chains of conditional independence, (iii) the elimination ideal of product terms in the chains gives the Hibi ideal and (iv) the TDAG can be recovered from a special bipartite graph constructed via the Alexander dual of the Hibi ideal. It is briefly demonstrated

that there are natural applications to statistical log-linear models, time series and Shannon information flow.

1 | INTRODUCTION

Lattice conditional independence (LCI) are a special type of statistical graphical models introduced by Andersson and Perlman [2] in the context of linear Gaussian models. The idea is that in the case of a distributive lattice of linear subspaces associated with the marginal models, all conditional independence statements could be classified via the intersections (meets) on the lattice. The theory comes with an important duality which links the LCI models isomorphically to a transitive-directed acyclic graph (TDAG) from [1], so that LCI could be seen as providing an important class of causal models. More precisely, each intersection on the lattice is mapped to a conditional independence statement conditioning on common ancestors in the TDAG.

The LCI models that arise from a TDAG are not in general graphical models based on directed acyclic graphs (DAGs) or undirected graphs (UGs). An important subclass of a graphical model is a decomposable graphical model, see [16], and it is possible to confuse these with the LCI models. There is an overlap between the two classes; see [1, 5]. An added complexity is that some DAG models give the same set of CIs as UGs; The DAGs equivalent to the same model are called *Markov equivalent*. The TDAG models have a unique set of CI statements but are not in general represented by UGs.

In this paper, we study LCI models through the lens of algebraic statistics, by connecting them to so-called Hibi rings that are both predicated on distributive (Boolean) lattices. Hibi rings and their properties have been extensively studied in commutative algebra; see, for example, [6, 9, 11]. The key point linking the two areas is the fact that for LCI models every member of the lattice is associated with a marginal distribution. This means that, initially, the algebraic relations are associated with margins. The existence of a connection between LCI models and Hibi ideals was suggested by Beerenwinkel, Eriksson and Sturmfels [3] within the context of Gröbner basis theory and maximum likelihood estimation. Here, we develop the algebraic connection in more detail, particularly in the relationship to the duality with the TDAG representation. The insight is that the TDAG representation that is related to conditional probability statements, suggests that there should be a representation of this duality in the algebraic context. This is indeed the case and, remarkably, the relation is that the Hibi (marginal) representation is the toric elimination ideal from the products of conditional probabilities given by the TDAG. Moreover, the TDAG itself can be read from a special bipartite graph arising from a particular Alexander duality.

In [4] it was suggested that LCI models have wide applicability, and in that paper, the application to time series was discussed. The work by [12] is a continuation of this approach in which a more formal link to Granger causality has been made. This, in particular, prompts the inclusion of time series as one of the main applications. Another application is to Shannon information where the LCI models lead to additive models for information.

The outline of the paper is as follows. In Section 1.1, we will give an elementary review of the LCI models with their connections to other types of statistical models, and in Section 1.2, we discuss the linear (Gaussian) case. In Section 2, we will describe the Hibi ideals. Section 3 will link all the concepts together. This includes an understanding of the effect of reversing the arrows in the TDAG. The final sections cover the two applications: time series and Shannon information.

1.1 | Elementary considerations

In teaching elementary statistics and probability theory, there is something of a conundrum in the expression of independence and CI. This is seen in as simple an example as two independent binary random variables $X_1, X_2 = \{0, 1\}$. Let

$$p_{ij} = \text{prob}\{(X_1 = i) \cap (X_2 = j)\} \ (i, j = 0, 1).$$

The issue is that there are two ways of expressing independence. Writing the margins as

$$p_{i+} = \sum_{j=0,1} p_{ij} \quad \text{and} \quad p_{+j} = \sum_{i=0,1} p_{ij},$$

the first way of expressing independence is

$$p_{ij} = p_{i+} p_{+j} \quad (i, j = 0, 1).$$

The second, and equivalent, condition (given all the $p_{ij} > 0$) is

$$p_{00} p_{11} - p_{10} p_{01} = 0.$$

In the algebraic context, this can be seen as a toric ideal condition in a certain polynomial ring, which is at the foundation of algebraic statistics for log-linear models. That is to say it arises as the elimination ideal from the power product representation

$$p_{ij} = s_i t_j,$$

or its log-linear equivalent

$$\log p_{ij} = \theta_i + \phi_j,$$

where $s_i = \exp(\theta_i)$ and $t_j = \exp(\phi_j)$.

With three binary random variables X_1, X_2, X_3, for the CI statement

$$X_1 \perp\!\!\!\perp X_2 \,|\, X_3,$$

we have either the marginal version

$$p_{ijk} p_{++k} = p_{i+k} p_{+jk} \quad (i, j, k = 0, 1), \tag{1.1}$$

or the toric ideal with one generator for each level of X_3

$$p_{000}p_{110} - p_{010}p_{100} = 0$$

$$p_{001}p_{111} - p_{011}p_{101} = 0.$$

The key idea in this paper is that the version (1.1) relies on a special set of margins indexed by subsets of indices $\{ijk, jk, ik, k\}$, which form a lattice generated by the 2-subsets $\{jk, ik\}$ under unions and intersections. We note, immediately, that there is a version of (1.1) for discrete random variables or continuous random variables and for larger collections of index sets. Let $[n] = \{1, \ldots, n\}$. For any subset of indices $I \subset [n]$, then, with slight abuse of notation we write p_I for the I-margin (summing or integrating over variables in $[n] \setminus I$).

Thus, let $I, J \subset [n]$. Then the CI statement

$$X_{I \setminus J} \perp\!\!\!\perp X_{J \setminus I} \mid X_{I \cap J}, \tag{1.2}$$

is represented by

$$p_{I \cup J} p_{I \cap J} = p_I p_J \quad \text{(given that } p_I > 0 \text{ for all } I\text{)}. \tag{1.3}$$

This is somewhat lazy notation: we should write this strictly as

$$p_{I \cup J}(x_{I \cup J}) p_{I \cap J}(x_{I \cap J}) = p_I(x_I) p_J(x_J), \tag{1.4}$$

for all $x = (x_1, \ldots, x_n)$ in the support of the distribution.

This shows the relationship with the sublattice generated by I and J in which the join operator is the union of sets $\cup$, and the meet operator is the intersection of sets $\cap$. More precisely, a term such as p_I is used both as terms in lattice theory and a (possibly marginal) probability distribution, in the continuous case a probability density function. We see clearly also how from an abstract point of view, (1.3) is a toric term indexed by the lattice.

One aim of the present paper is to show that if we have a Boolean (distributive) lattice formed from collection of sets $\mathbb{I}$, then there is a probability model associated with $\mathbb{I}$ such that *every* union/intersection formula of the type (1.3) gives a CI statement of the type (1.4). Moreover, we demonstrate the connection with the associated toric ideal of the lattice.

The toric ideal associated with a distributive lattice is called a Hibi ideal after the work of Hibi in [11]. Moreover, it does indeed hold that this defines a class of probability models which is legitimate to call lattice conditional independence (LCI) because exactly such models were introduced in the Gaussian (normal) case. The advantages of the algebraic approach using Hibi ideals are that (i) we can define LCI models for any type of random variables, discrete or continuous, (ii) entities such as the TDAG associated with the LCI model have a purely algebraic realisation and (iii) the ideal theory can be used to make statements about the probability models. One algebraic concept is that of the Alexander duality, which we will show has a precise relationship to the underlying TDAG.

Definition 1.1 (TDAG). A DAG $G = ([n], E)$ is *transitive* (TDAG) if $\{i \to j\} \in E$ and $\{j \to k\} \in E$ imply that $\{i \to k\} \in E$ (see Figures 2 and 3).

1.2 | Linear LCI models

Multivariate normal LCI models are defined in terms of collections of marginal distributions based on a distributive lattice of subsets of indices, as described above and will be made more formal later in the paper. The meet and join of the lattice will give intersections and unions of vector spaces associated with the margins.

Let us now discuss a single CI statement as in (1.2) for a (zero mean) multivariate normal random vector $X = (X_1, \dots, X_n)^T$. Let the full rank covariance matrix be Γ. First, we factor Γ in some standard way: $\Gamma = AA^T$, where A is a square matrix. We can then think of X as arising from what is sometimes called a moving average representation

$$X = AZ,$$

where Z is the zero mean multivariate normal random variable with the covariance matrix I_n (the $n \times n$ identity matrix). For index sets I, J and $K = I \cap J$ we can express, by selecting appropriate rows of A, the corresponding random vector as

$$X_I = A_I Z, \quad X_J = A_J Z, \quad X_K = A_K Z,$$

with covariance matrices, respectively:

$$\Gamma_I = A_I A_I^T, \quad \Gamma_J = A_J A_J^T, \quad \Gamma_K = A_K A_K^T.$$

(Note that A_I denotes the submatrix of A on the rows indexed by I). Using a standard formula known as Schur complement, the covariance matrix between X_I and X_J conditional on X_K is

$$A_I A_J^T - A_I A_K^T (A_K A_K^T)^{-1} A_K A_J^T. \tag{1.5}$$

For CI in the Gaussian case, it is necessary and sufficient that the matrix at (1.5) is zero.

Note that $P_K = A_K^T (A_K A_K^T)^{-1} A_K$ is the orthogonal projection operator onto the row space of A_K. Similarly, we can construct the orthogonal projections $P_I = A_I^T (A_I A_I^T)^{-1} A_I$ and $P_J = A_J^T (A_J A_J^T)^{-1} A_J$ for the row spaces of A_I and A_J, respectively. The CI condition can then be seen to be equivalent to

$$P_I (I_n - P_K) P_J = 0.$$

It is important to note that the condition is equivalent to the commutativity condition

$$P_I P_J = P_J P_I \ (= P_K),$$

which is necessary and sufficient for the CI.

If we extend the argument to all $X_I, I \in \mathbb{I}$, where $\mathbb{I}$ is a distributive sublattice, then we see that it is necessary and sufficient that all projections P_I, for all $I \in \mathbb{I}$ in the lattice commute. This implies that they are simultaneously diagonalisable with spectrum (eigenvalues) being 0 or 1. In fact the location of the values 1 in $[n]$ in the spectrum of each projector $I, J, K \in [n]$ corresponds precisely to the indicator of the corresponding index set: $P_I = \mathrm{diag}(q_1, \dots, q_n)$ where $q_i = 1$ for $i \in I$ and 0

otherwise. For example, for the sets $I = 13, J = 23$, and $K = 3$ we have the eigenvalues (spectrum) $\{1, 0, 1\}\{0, 1, 1\}$ and $\{0, 0, 1\}$. The sets where the spectrum is non-zero are a model for a subset lattice which is equivalent to the original lattice.

Important in the theory of LCI is that every distributive lattice, which we here associate with the collection of index sets $\mathbb{I}$, arises from a TDAG and vice versa: any distributive lattice defines a TDAG. Although we will study this in more detail in Section 3.1 (see also Definition 3.1), we can simply describe this 1-1 correspondence as follows. The corresponding TDAG, $G([n], E)$ has n vertices labelled by $1, \dots, n$, where each i corresponds to an X_i, $i = 1, \dots, n$ in our probability model, and $i \to j$ is in the edge set E if and only if i is in the ancestral set of j, meaning: $i \leqslant j$ in the partial ordering defined by the lattice.

2 | HIBI IDEALS AND LCI MODELS

In [11], Hibi introduced a class of algebras which nowadays are called Hibi rings. They are toric rings attached to finite posets, and may be viewed as natural generalisations of polynomial rings. We first recall the construction of Hibi ideals associated to distributive lattices. Recall that a poset in which every two elements have a 'meet' and a 'join' is called a lattice. A lattice $(\mathcal{L}, \vee, \wedge)$ is *distributive* if the following additional identities hold for all x, y and z in $\mathcal{L}$:

$$x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z) \quad \text{and} \quad x \wedge (y \vee z) = (x \vee y) \wedge (x \vee z).$$

Given a poset Q, a non-empty subset $I \subseteq Q$ is called an *order ideal* of Q if for any $i \in I$ and $j \in Q$, $j < i$ implies that $j \in I$. We denote $\mathcal{O}(Q)$ for the set of order ideals of Q. Note that if we define the union of sets $\cup$ as the join operator and the intersection of sets $\cap$ as the meet operator, then $(\mathcal{O}(Q), \cap, \cup)$ is naturally a distributive lattice.

Before defining Hibi ideals, we first fix our notation. Consider the poset Q on the elements $1, 2, \dots, n$ and the polynomial ring $R = \mathbb{C}[z_1, \dots, z_n, y_1, \dots, y_n]$ over the field of complex numbers on the variables z_i, y_i, where variables z_i, y_i are corresponding to the element i of Q. For every order ideal $I \subseteq Q$ we associate the monomials $u_I := \prod_{i \in I} z_i \prod_{i \notin I} y_i$ and $u_I' := \prod_{i \in I} z_i$. Then we define the following polynomial maps:

$$\varphi_Q : \mathbb{C}[p_I : I \in \mathcal{O}(Q)] \to \mathbb{C}[z_1, \dots, z_n, y_1, \dots, y_n] \quad \text{with} \quad p_I \mapsto u_I, \tag{2.1}$$

and

$$\varphi_Q' : \mathbb{C}[p_I : I \in \mathcal{O}(Q)] \to \mathbb{C}[z_1, \dots, z_n] \quad \text{with} \quad p_I \mapsto u_I'. \tag{2.2}$$

Definition 2.1 (Hibi ideal). The *Hibi ideal* of Q is defined as $L_Q = \text{Ker}(\varphi_Q) = \text{Ker}(\varphi_Q')$.

The equality above is implicit in Hibi's seminal paper [11]; see also [6, Theorem 1.1].

Before we explain LCI models in more detail, we recall the notion of join-irreducible elements of a lattice.

Definition 2.2. Let $(\mathcal{L}, \wedge, \vee)$ be a finite distributive lattice. An element $j \in \mathcal{L}$ is called *join-irreducible* if $j = \ell_1 \vee \ell_2$ for some $\ell_1, \ell_2 \in \mathcal{L}$ implies $j = \ell_1$ or $j = \ell_2$. Denote by $J(\mathcal{L})$ the set of join-irreducible elements in $\mathcal{L}$.
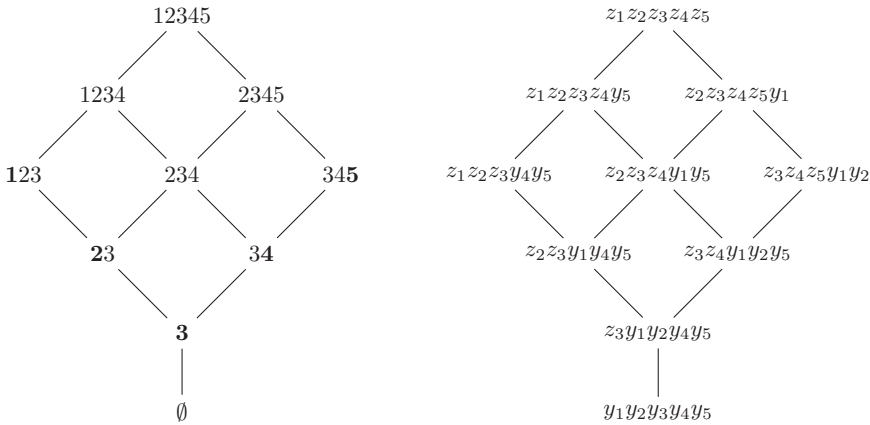
FIGURE 1 (Left) The lattice $\mathcal{L}$ of an LCI model. (Note that $\emptyset$ is always the unique minimum element of $\mathcal{L}$.) (Right) The monomials u_I associated to each element I of $\mathcal{L}$. Note that z_i corresponds to the lattice point which is the smallest subset containing i (marked in bold in $\mathcal{L}$)
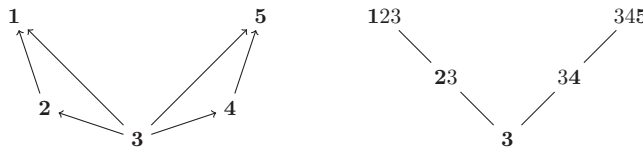

FIGURE 2 (Left) The TDAG of the above LCI model. (Right) The joint irreducible poset $J(\mathcal{L})$

Example 2.3. In examples, in order to simplify our notation we denote the set $\{i_1, i_2, \ldots, i_t\}$ by $i_1 i_2 \cdots i_t$, where $i_1 < i_2 < \cdots < i_t$. Figure 1 depicts the lattice generated by the sets 123, 234, 345 and the corresponding monomials u_I for each lattice element I. The non-empty join-irreducible elements of $\mathcal{L}$ are 3, 23, 34, 123 and 345. The corresponding join-irreducible post is depicted in Figure 2 (right). We have marked an element in bold, when it is the first time that it shows up in $\mathcal{L}$ (looking at $\mathcal{L}$ from bottom to top). The corresponding TDAG is shown in left that will be further explained in Section 3.

Remark 2.4. Note that, every element I of $J(\mathcal{L})$ is uniquely determined by an element i for which I is the smallest lattice element containing it. Hence, we may use the notation g_i for its corresponding monomial $u_I' = \prod_{j \leqslant i} z_j$. For example, Figure 1 shows that 123 is the smallest element of the lattice containing 1, so the monomial $z_1 z_2 z_3$ is denoted by g_1. Similarly, we have $g_2 = z_2 z_3$, $g_3 = z_3$, $g_4 = z_3 z_4$ and $g_5 = z_3 z_4 z_5$.

Note that the partial order on $\mathcal{L}$ induces a partial order on $J(\mathcal{L})$. Hence, taking the poset $J(\mathcal{L})$, we can define the Hibi ideal of $J(\mathcal{L})$ or simply the Hibi ideal of the lattice $\mathcal{L}$. More precisely, Birkhoff's fundamental structure Theorem [22, Theorem 3.4.1] guarantees that the lattices $(\mathcal{O}(J(\mathcal{L})), \cap, \cup)$ and $(\mathcal{L}, \cap, \cup)$ are isomorphic. One fundamental result in this context is that the Hibi ideal of $\mathcal{L}$ is generated by the following relations (see, for example, [10, Theorem 10.1.3]):

$$p_I p_J - p_{I \cap J} p_{I \cup J} \quad \text{with} \quad I \nsubseteq J \quad \text{and} \quad J \nsubseteq I \quad (I, J \in \mathcal{L}).$$

We can now interpret the terms $p_I, p_J, p_{I \cup J}, p_{I \cap J}$ as marginal probability distributions associated with their appropriate margins. The only bit of additional notation we need is the conditioning sign that we will develop in the next section. Thus, the above binomials in the Hibi ideal give us the CIs in (1.2).

3 | TRANSITIVE DIRECTED ACYCLIC GRAPHS

3.1 | The probabilistic interpretation of the join irreducible elements

We start by giving the formal definition of the TDAG associated to a distributive lattice.

Definition 3.1 (TDAG of a lattice). Let $\mathcal{L}$ be a distributed lattice whose elements are corresponding to a collection of subsets of $[m]$. Assume that $\mathcal{L}$ has n join-irreducible elements. The TDAG associated with $\mathcal{L}$ is a TDAG $G([n], E)$ where the vertex i represents a join-irreducible element of $\mathcal{L}$ containing i for the first time (ordering the join-irreducible elements from bottom to top) and a directed edge $e = \{i \to j\}$ belongs to E if and only if the join-irreducible elements i and j are comparable in the lattice and $i < j$.

Example 3.2. The associated TDAG of the lattice $\mathcal{L}$ in Figure 1 is depicted in Figure 2 (left).

Figures 1 and 2 show the relations between the LCI lattice, the set of monomials associated with the lattice and the corresponding TDAG of the lattice. Up to this point the development has been purely algebraic. Now we interpret the variables p_J, z_i and y_k in probabilistic terms.

We can give a formal meaning, or interpretation, of the z_i as

$$z_i = p_{\{j : j \leqslant i\} | \{j : j < i\}},$$

where $j \leqslant i$ is in the lattice sense. The notation means that we condition X_i on the 'ancestors' of X_i, with respect to the lattice. For instance, let us consider the example of the set 123 in the lattice $\mathcal{L}$ in Figure 1. This has an ascending chain of subsets leading to

$$3 \subset 23 \subset 123,$$

and a corresponding chain of monomials:

$$z_3, \ z_2 z_3, \ z_1 z_2 z_3.$$

The interpretation of the z_i is as the terms in the development of marginal p_{123} in terms of conditions. Thus,

$$p_{123} = p_3 p_{23 \mid 3} p_{123 \mid 23},$$

and we make the mapping

$$z_3 \leftrightarrow p_3, \ z_2 \leftrightarrow p_{23 \mid 3}, \ z_1 \leftrightarrow p_{123 \mid 23}.$$

Similarly, let us check p_{234}. Mapping to the z_i we have

$$z_2 z_3 z_4 \rightarrow \left(\frac{p_{23}}{p_3}\right) p_3 \left(\frac{p_{34}}{p_3}\right) = \frac{p_{23} p_{34}}{p_3} = p_{234}.$$

Note that the last equality derives from the equation $p_{23} p_{34} = p_3 p_{234}$ arising from the corresponding Hibi ideal. More generally, for every saturated chain of subsets

$$J_0 \subset J_1 \subset \cdots \subset J_r = J$$

we have that

$$p_J = p_{J_0} p_{J_1 \mid J_0} p_{J_2 \mid J_1} \cdots p_{J \mid J_{r-1}}$$
$$= z_{J_0} z_{J_1 \setminus J_0} z_{J_2 \setminus J_1} \cdots z_{J_r \setminus J_{r-1}}.$$

We should note that in our development we increment subsets with a single 'i', but in general the increment can be larger. In those cases, we can extend the definition of the index set, in an obvious way. We now state and prove the relationship between the marginal and conditional probabilities.

Theorem 3.3. *Consider a probability distribution with a set of positive marginals $p_J = \prod_{i \in J} z_i$ where the collection of sets J forms a distributive lattice satisfying the Hibi relations. Then*

(i) *the quantities z_i are conditional probabilities $z_i = p_{\{j : j \leqslant i\} \mid \{j : j < i\}}$;*
(ii) *the Hibi relations give all conditional independent statements associated with the lattice.*

Proof. The conditions satisfied by the Hibi ideal can be rearranged as

$$p_{I \cup J} = \frac{p_I p_J}{p_{I \cap J}}.$$

Now assume that we have a probability distribution with positive marginals supported on the lattice. Then the logarithm $\phi(I) = \log p_I$ induces a valuation

$$\phi(I \cup J) = \phi(I) + \phi(J) - \phi(I \cap J). \tag{3.1}$$

We then invoke the theory of valuations on a distributive lattice and claim that the conditional independence relations arise from Möbius inversion on the lattice, essentially inclusion–exclusion formula for $\phi(I)$. More precisely, considering the saturated chains of subsets

$$I \cap J = J_0 \subset J_1 \subset \cdots \subset J_r = J \quad \text{and} \quad I \cap J = I_0 \subset I_1 \subset \cdots \subset I_s = I,$$

Equation (3.1) translates to

$$p_{I \cup J} = \frac{p_I p_J}{p_{I \cap J}}$$
$$= \frac{p_I}{p_{I \cap J}} \frac{p_J}{p_{I \cap J}} p_{I \cap J}$$

$$= \left(\frac{p_I}{p_{I_{s-1}}} \cdots \frac{p_{I_1}}{p_{I \cap J}} \right) \left(\frac{p_J}{p_{J_{s-1}}} \cdots \frac{p_{J_1}}{p_{I \cap J}} \right) p_{I \cap J}$$

$$= \prod_{i=1}^{s} z_{I_i \setminus I_{i-1}} \prod_{j=1}^{r} z_{J_j \setminus J_{j-1}} z_{I \cap J}. \qquad \square$$

3.1.1 | The complementary lattice of $\mathcal{L}$ and the reverse TDAG

Given a lattice L on a collection of subsets $\mathbb{I}$ of $[n]$, we denote its complementary lattice with $\mathcal{L}^*$ whose elements are $[n] \setminus I$ (for $I \in \mathbb{I}$) and are reversely ordered comparing to $\mathcal{L}$. In exactly same way that we have defined the entities for $\mathcal{L}$, we can associate the dual entities q_J and the y_i for the complementary lattice $\mathcal{L}$. If we want to think of each arrow as a generalisation of time, we might say we condition on the 'descendants' of i. In the same way, the variable y_i has an interpretation as the conditional random variable for the distribution of random variables Y_i defined by the margins $q_J = p_{[n]}/p_{[n]\setminus J}$. Given the distribution defined by the q_J margins:

$$y_i = q_{\{j : j \geqslant i\} | \{j : j > i\}},$$

the question remains as what ideal do we obtain if we eliminate the y_i from the ideal

$$\left\langle q_J - \prod_{i \in J} y_i \; : \; J \in \mathcal{O}(Q) \right\rangle.$$

This has an interesting probabilistic interpretation. In Section 3.3, we will see that it relates to reversing the arrows in the associated TDAG of $\mathcal{L}$. We will also connect this graph operation with the Alexander duality of the corresponding ideal.

Just as we associate a marginal random variable X_I with every index set $I \in \mathbb{I}$, so we associate a random variable Y_J with every J in the complementary lattice $\mathcal{L}^*$ on the subsets $\mathbb{J}$. The random variable Y_J is interpreted as having the conditional distribution of X_J given $X_{[n]\setminus J}$. This has the distribution

$$q_J = \frac{p_{[n]}}{p_{[n]\setminus J}}. \tag{3.2}$$

The duality is clear: writing $I = [n] \setminus J$ we have

$$p_I q_J = p_{[n]}.$$

For the complementary lattice of Figure 1, it can be shown that Y_{12} and Y_{45} are independent, as

$$q_{12} q_{45} = \frac{p_{12345}}{p_{345}} \frac{p_{12345}}{p_{123}}$$

$$= \frac{p_{12345}^2}{p_{12345} p_3}$$

$$= \frac{p_{12345}}{p_3}$$

$$= q_{1245},$$

giving the factorisation for the required independence.

The Gaussian case reflects the duality well. If one considers the projection operator P_J associated with the Gaussian random variable X_J then the Gaussian random variable associated with the random variable $Y_{[n]\setminus I}$ has the orthogonal projector

$$Q_{[n]\setminus I} = I_n - P_I. \tag{3.3}$$

In our running example, we can see that

$$Q_{12}Q_{45} = (I - P_{345})(I - P_{123}) = I - P_{345} - P_{123} + P_3 = 0,$$

hence we recognise the conditional independence $Y_{12} \perp\!\!\!\perp Y_{45} \mid Y_3$.

3.2 | Log-linear models for margins

With a view to statistical modelling, we would like to point out that the z-representation leads to log-linear models, not only of the more familiar form for a single probability $p(x)$ but also for all the margins. If for a margin for the index set I we have the product form

$$p_I(x) = \prod_{i \in I} z_i(x)$$

and

$$\log p_I(x_I) = \sum_{i \in I} u_i(x_I),$$

where $u_i(x_I) = \log z_i(x_I)$, then this can be written in exponential form as

$$p_I(x) = \exp\left\{ \sum_{i \in I} u_i(x_I) \right\}.$$

Because of the dependence on x_I this should be referred to as a *local* model. In statistical modelling, one may start with a class of functions $\mathcal{U} = \{u(x_I)\}$, which defines the properties of our model and extend the class in a parametric way by introducing parameters $\theta_{i,I}$ and writing

$$\log p_I(x) = \exp\left\{ \sum_{i \in I} \theta_{i,I} u_i(x_I) \right\}.$$

Similar models can be written down for the q-margins. We do not discuss statistical analysis in this paper, but we refer to [18] for an overview in the Gaussian case. Maximum likelihood estimation and other statistical issues in the general case remain open, but we note the promising contribution from [3].

3.3 | Alexander duality and a special bipartite graph

We promised that the algebraic representation would yield certain results which may prove useful in the probability theory. Let $\mathcal{L}$ be a distributive lattice with join-irreducible poset $Q = J(\mathcal{L})$; see Definition 2.2. The significance of Hibi ideals is that the Alexander dual of the monomial ideal $M_Q = \langle u_I : I \in \mathcal{O}(Q) \rangle$ can be interpreted as the edge ideal of a special bipartite graph. To be precise, for a poset Q on $1, \dots, n$, if we define the bipartite graph G on the vertex set $\{z_1, \dots, z_n, y_1, \dots, y_n\}$ by saying that $\{z_i, y_j\}$ is an edge of G if and only if $i \leqslant j$ in Q, then the Alexander dual of the ideal M_Q is the edge ideal of G. It turned out that bipartite graphs obtained in this way are exactly the Cohen–Macaulay bipartite graphs [9, Lemma 3.1].

We first recall the notion of Alexander duality for squarefree monomial ideals which naturally arise in this context (see [19, Definition 5.20]).

Definition 3.4 (Alexander dual). Let $K[\mathbf{x}] = K[x_1, \dots, x_n]$ be a polynomial ring over a filed K in n variables. Let $M = \langle \mathbf{x}^{a_1}, \dots, \mathbf{x}^{a_r} \rangle$ be a squarefree monomial ideal in $K[\mathbf{x}]$, that is, it is generated by monomials $\mathbf{x}^{a_i}$ for $i = 1, \dots, r$, where every coordinate of $a_i = (a_{i,1}, \dots, a_{i,n})$ is either 0 or 1. The *Alexander dual* of M is the monomial ideal $M^* = \underline{m}^{a_1} \cap \cdots \cap \underline{m}^{a_r}$, where $\underline{m}^{a_i} = \langle x_j : a_{i,j} = 1 \rangle$.

Definition 3.5 (Edge ideal). Let G be a directed graph on the vertices $1, \dots, n$. Consider the polynomial ring $K[\mathbf{z}, \mathbf{y}] = K[z_1, \dots, z_n, y_1, \dots, y_n]$ on $2n$ variables, where z_i, y_i corresponds to the vertex i of G. The *edge ideal* of G is the ideal $I(G) = \langle z_i y_j : \{i \to j\} \text{ is an edge of } G \rangle$.

Example 3.6. Consider our running example depicted in Figures 1 and 2. The ideal M_Q is

$$M_Q = \langle y_1 y_2 y_3 y_4 y_5, z_3 y_1 y_2 y_4 y_5, z_2 z_3 y_1 y_4 y_5, z_3 z_4 y_1 y_2 y_5, z_1 z_2 z_3 y_4 y_5, z_2 z_3 z_4 y_1 y_5,$$

$$z_3 z_4 z_5 y_1 y_2, z_1 z_2 z_3 z_4 y_5, z_2 z_3 z_4 z_5 y_1, z_1 z_2 z_3 z_4 z_5 \rangle.$$

The Alexander dual of M_Q is generated by the degree 2 monomials as follows:

$$M_Q^* = \langle z_3 y_2, z_3 y_1, z_2 y_1, z_3 y_4, z_3 y_5, z_4 y_5, z_1 y_1, z_2 y_2, z_3 y_3, z_4 y_4, z_5 y_5 \rangle,$$

which is the edge ideal of the TDAG in the left hand of Figure 2, where we have removed the 'loops' corresponding to the monomials $z_i y_i$.

More formally, we have the following corollary [9, Lemma 3.1] adopting our notation.

Corollary 3.7. *Let $\mathcal{L}$ be a distribute lattice with the joint irreducible poset $Q = J(\mathcal{L})$. Then the Alexander dual of the ideal $M_Q = \langle u_I : I \in \mathcal{O}(Q) \rangle$ is the monomial ideal M_Q^* generated by $z_i y_j$, where $i \leqslant j$ in Q. In particular, M_Q^* is the edge ideal of the TDAG on the vertices $1, \dots, m$ with the edge set $\{i \to j : i \leqslant j \text{ in } Q\}$.*

This analysis gives a pleasing duality which we state in the following lemma.

Lemma 3.8. *Let $\mathcal{L}$ be a distribute lattice with the joint irreducible poset $Q = J(\mathcal{L})$. Let $I(i) = \{j : j \leqslant i\}$ and $J(j) = \{k : k \geqslant j\}$ for i, j in Q. Following the notation of §3.1.1, let p_I denote the probabilities associated to $\mathcal{L}$ with dual entities q_J associated to the complementary lattice of $\mathcal{L}$. Then, for every*

$i \leqslant j$ we have that

$$\frac{p_{I(j)}}{p_{I(i)}} = \frac{q_{J(j)}}{q_{J(i)}}.$$

Proof. Let $G = G([n], E)$ be the corresponding TDAG of $\mathcal{L}$. By Corollary 3.7, for every vertex i of G, $\{j \to i\}$ is an edge if j is in the set $I(i)$. Similarly, for the complementary lattice of $\mathcal{L}$, $\{k \to j\}$ is an edge in the corresponding TDAG if k is in the set $J(j)$. In other words, the corresponding TDAG is the reverse of G, that is, its edges are obtained by reversing the edges of G. Then, note that for any member $p_I q_J$ for disjoint sets I, J with $I \cup J = [n]$ we have

$$p_{I(i)} q_{J(j)} = p_{I(j)} q_{J(i)} = p_{[n]}.$$

Thus if $i \leqslant j$ we have

$$\frac{p_{I(j)}}{p_{I(i)}} = \frac{q_{J(j)}}{q_{J(i)}},$$

giving a symmetry in forward and backward conditional independence on the TDAG and the reverse TDAG. □

As we have seen above, the Alexander dual framework unifies the forward and backward TDAGs analogously to its role in the paths and cuts duality of algebraic approaches to system reliability [7, 20]. Moreover, it provides us with various computational toolkits on Macaulay2 [8] such as the 'Posets.m2' package.

4 | APPLICATIONS

4.1 | Time series

We now give a somewhat larger example showing how to apply the LCI principle to three univariate time series $\{X_{1t}, X_{2t}, X_{3t}\}$. We define a lattice conditional independence model which describes the conditional independence between the first time series up to time t and the third time series up to time t, given the second time up to time $t - 1$. Moreover, in the general case this holds for all $t = 1, 2, \dots$. For $t = 1$, we may simply assume that X_{11}, X_{21} and X_{31} are independent (for simplicity we have omitted cross-sectional conditional independence).
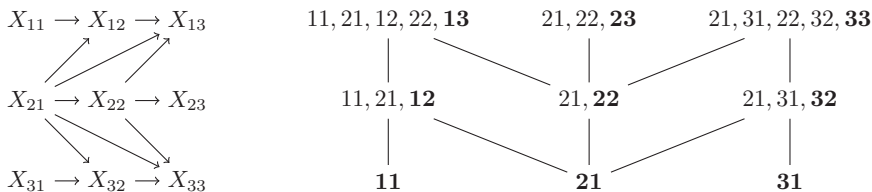

FIGURE 3 (Left) Time series for $t = 1, 2, 3$. To simplify the graph, the edges $X_{i1} \to X_{i3}$ (for $i = 1, 2, 3$) are not depicted. (Right) The join-irreducible poset o the corresponding lattice. The variable z_{ij} corresponds to the point which is the smallest subset containing ij (marked in bold)

Example 4.1. Figure 3 (left) shows the TDAG scheme for three time points. We have simplified the labelling in the middle to avoid the double suffix. The corresponding lattice $\mathcal{L}$ has 45 elements indexed by the following subsets (ordered by inclusions):

$\emptyset, \{11\}, \{21\}, \{11, 21\}, \{31\}, \{11, 31\}, \{21, 31\}, \{11, 21, 31\}, \{11, 21, 12\}, \{11, 21, 31, 12\}, \{21, 22\},$

$\{11, 21, 22\}, \{21, 31, 22\}, \{11, 21, 31, 22\}, \{11, 21, 12, 22\}, \{11, 21, 31, 12, 22\}, \{21, 31, 32\},$

$\{11, 21, 31, 32\}, \{11, 21, 31, 12, 32\}, \{21, 31, 22, 32\}, \{11, 21, 31, 22, 32\}, \{11, 21, 31, 12, 22, 32\},$

$\{11, 21, 12, 22, 13\}, \{11, 21, 31, 12, 22, 13\}, \{11, 21, 31, 12, 22, 32, 13\}, \{21, 22, 23\}, \{11, 21, 22, 23\},$

$\{21, 31, 22, 23\}, \{11, 21, 31, 22, 23\}, \{11, 21, 12, 22, 23\}, \{11, 21, 31, 12, 22, 23\}, \{21, 31, 22, 32, 23\},$

$\{11, 21, 31, 22, 32, 23\}, \{11, 21, 31, 12, 22, 32, 23\}, \{11, 21, 12, 22, 13, 23\}, \{11, 21, 31, 12, 22, 13, 23\},$

$\{11, 21, 31, 12, 22, 32, 13, 23\}, \{\{21, 31, 22, 32, 33\}, \{11, 21, 31, 22, 32, 33\},$

$\{11, 21, 31, 12, 22, 32, 33\}, 11, 21, 31, 12, 22, 32, 13, 33\}, \{21, 31, 22, 32, 23, 33\},$

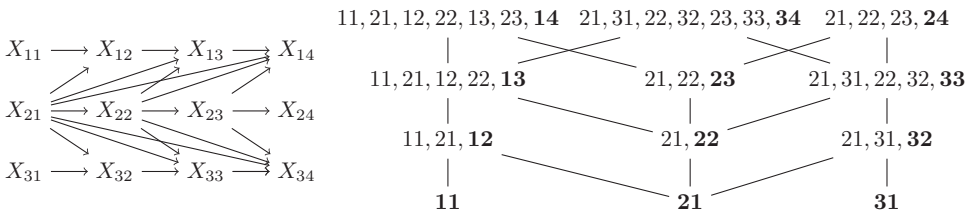$\{11, 21, 31, 22, 32, 23, 33\}, \{11, 21, 31, 12, 22, 32, 23, 33\}, \{11, 21, 31, 12, 22, 32, 13, 23, 33\}.$

FIGURE 4 (Left) Time series for $t = 1, 2, 3, 4$ while the curved edges $X_{ij} \to X_{ik}$ (for $i = 1, 2, 3$ and $j < k + 1$) are not depicted above. (Right) The join-irreducible poset of the corresponding lattice. Note that z_{ij} corresponds to the point which is the smallest subset containing ij (marked in bold)

Figure 4 (left) shows the corresponding poset $Q = J(\mathcal{L})$ of join-irreducible elements. To verify that the TDAG is recaptured using the Alexander dual, for Q in Figure 4 (left), consider the ideal

$$M_Q = \langle y_{11} y_{21} y_{31} y_{12} y_{22} y_{32} y_{13} y_{23} y_{33}, z_{11} y_{21} y_{31} y_{12} y_{22} y_{32} y_{13} y_{23} y_{33},$$

$$z_{21} z_{31} z_{22} z_{32} z_{23} z_{33} y_{11} y_{12} y_{13}, \ldots, z_{11} z_{21} z_{31} z_{12} z_{12} z_{32} z_{13} z_{23} z_{33} \rangle$$

generated by 45 monomials associated to the subsets above. Then the Alexander dual of M_Q is

$$M_Q^* = \langle z_{11} y_{12}, z_{21} y_{12}, z_{21} y_{22}, z_{21} y_{32}, z_{31} y_{32}, z_{11} y_{13}, z_{21} y_{13}, z_{12} y_{13}, z_{22} y_{13}, z_{21} y_{23}, z_{22} y_{23}, z_{21} y_{33},$$

$$z_{31} y_{33}, z_{32} y_{33}, z_{22} y_{33}, z_{11} y_{11}, z_{21} y_{21}, z_{31} y_{31}, z_{12} y_{12}, z_{22} y_{22}, z_{32} y_{32}, z_{13} y_{13}, z_{23} y_{23}, z_{33} y_{33} \rangle.$$

Now, following Corollary 3.7 we obtain the edge representation of the time series in Figure 3 (left) by removing the 'loops' corresponding to the monomials $z_{ij} y_{ij}$ in M_Q^*.

To preserve the conditional independence as time progresses, we need to show how to update the TDAG and the associated join-irreducible poset. If we add another time layer, then the new

join-irreducible poset has one more layer with the index sets

$$\{11, 12, 13, 14, 21, 22, 23\}, \{21, 22, 23, 24\}, \{21, 22, 23, 31, 32, 33, 34\}.$$

These are the generators which appear at the top of the previous poset for $t = 1, 2, 3$:

$$\{11, 12, 13, 21, 22\}, \{21, 22, 23\}, \{21, 22, 31, 32, 33\}.$$

It should be clear, then, that the rule for updating the lattice is just to update the ancestral sets at each time point. But this also has a rule in terms of the z_{ij}. More precisely, we have that

$$P_{\{11,12,13,14,21,22,23\}} = z_{23} z_{14} \cdot P_{\{11,12,13,21,22\}}$$

$$P_{\{21,22,23,24\}} = z_{24} \cdot P_{\{21,22,23\}}$$

$$P_{\{21,22,23,31,32,33,34\}} = z_{23} z_{34} \cdot P_{\{21,22,31,32,33\}}.$$

Thus the rule is to update the lattice generators with the conditional probabilities corresponding to the new arrows. In time series language, $z_{23}, z_{14}, z_{24}, z_{34}$ can be considered as the innovations corresponding to the conditional independence structure or, to put it more strongly, the innovations required to *preserve* the conditional independence. All new lattice-based conditional independence statements are obtained by expanding with subsets, inclusions and intersections in the usual way.

4.1.1 | Transfer functions

The z-products, which we have seen are strings of conditional probabilities, facilitate the updating of time series models, using transfer functions. We only give a sketch, here. We suggest that an appropriate approach is to use the generators g_i in the sense of Remark 2.4, where i uniquely determines the subset I which is the smallest join-irreducible element containing i). For example, as depicted in Figure 3 (right), the smallest set containing 13 is $\{11, 21, 12, 22, 13\}$. Let us denote the corresponding monomial with g_{13} as follows. Thus, the transfer functions can be based on updating the generators. These are products, or sets of products, but their logarithms give linear functions, creating a log-linear time series.

More precisely, in our example in Figure 3 the generators corresponding to the top-layer at time $t = 3$ are given by

$$g_{13} = z_{11} z_{21} z_{12} z_{22} z_{13},$$

$$g_{23} = z_{21} z_{22} z_{23},$$

$$g_{33} = z_{21} z_{31} z_{22} z_{32} z_{33}.$$

Taking the same LCI structure advancing to $t = 4$, as depicted in Figure 4, we multiple entrywise by a new vector, namely

$$u_{1t} = z_{14} z_{23},$$

$$u_{2t} = z_{24},$$

$$u_{3t} = z_{24} z_{34}.$$

We can express the updating symbolically, with obvious notation, by

$$g_{i,t+1} = g_{it} \circ u_{it},$$

for every $i = 1, \dots, 3$, where $\circ$ is the Schur (entrywise) product. Taking logarithms we write:

$$\log(g_{i,t+1}) = \log(g_{it}) + \log(u_{it}).$$

Now, using the shift operator S from time series we can write this as

$$S \cdot \log(g_{it}) = \log(g_{it}) + \log(u_{it}),$$

leading to

$$(S - 1)\log(g_{it}) = +\log(u_{it}),$$

and formally

$$\log(g_{it}) = (S - 1)^{-1} \log(u_{it}),$$
$$= T(1 - T)^{-1} \log(u_{it}),$$

Here $T = S^{-1}$ is the reverse (backward in time) shift operator. Then, by expanding

$$(1 - T)^{-1} = 1 + T + T^2 + \cdots,$$

and with suitable initial conditions, we can develop the process in terms of the 'innovations' u_{it}, noting that $T \log(u_{it}) = \log(u_{i,t-1})$.

4.2 | Information flow

The Shannon information associated with the margin I is

$$H(I) = \mathrm{E}\{\log(p_I(X_I))\},$$

where the expectation E is with respect to the full joint distribution (although we could compute the expectation with respect to the marginal random variable X_I). Taking logarithms and then expectation of the Hibi relations (assuming positivity of the probabilities) we have

$$p_I(X)p_J(X) = p_{I \cap J}(X)p_{I \cup J}(X),$$

and we see that H is a *valuation* on our distributive lattice: for all pairs $I, J \in \mathcal{L}$ we have

$$H(I \wedge J) + H(I \vee J) = H(I) + H(J).$$

The theory of valuations on distributive lattices is famously developed by Rota [15, 21]. It is important to note that, whereas $\phi(I) = \log \phi_I$ coming from a probability evaluation in Theorem 4.2

depends on the argument X_I, that is a local evaluation, the expectation in the evaluation of Shannon information removes the random variable X_I, leaving only the index I. Thus, we may write $H(I)$ as a numerical quantity attached directly to the margin I. For example, we can again use simple suffix notation. Thus, for simple conditional expectation we write:

$$H(123) = H(23) + H(34) - H(3).$$

Moreover, we have a TDAG representation of change in information. Returning to our running TDAG left branch $3 \subset 23 \subset 123$ from Figure 1, we have

$$H(123) = H(3) + \{H(23) - H(3)\} + \{H(123) - H(23)\}. \tag{4.1}$$

The three terms on the right hand side correspond respectively to the arrows in our TDAG: $3 \to 1, 3 \to 2, 2 \to 1$ (see Figure 2). We see that the TDAG describes a (cumulative) additive model for information. The transitive closure edge $3 \to 1$ simply contributes $H(3)$ to the total information at node 1.

The increments along each edge have an interpretation in terms of the Shannon result for disjoint index sets I, J:

$$H(I \cup J) = H(I) + E_I\{H(J \mid I)\}.$$

Thus, in our example we can write

$$H(123) = H(3) + E_3 H(2 \mid 3) + E_{23} H(1 \mid 23). \tag{4.2}$$

This interpretation of the TDAG as representing an information flow diagram, holds generally for LCI models, and is reminiscent of various types of compartmental flow models, for fluids, or perhaps a better analogy is for energy, recalling that Shannon entropy is the negative of Shannon information. Essentially, it arises from the underlying Markov structure.

We express this result as an informally stated theorem, without proof.

Theorem 4.2. *The TDAG of an LCI model provides an additive information model in which to each edge $(i \to j)$ is associated energy increment: $H(I) - H(J)$ where $I = J \cup \{i\}$.*

As we mentioned in the discussion above Theorem 4.2, one property of valuations on distributive lattices is the existence of inclusion–exclusion lemma. This allows us to express the information for any union of index sets in the lattice in terms of alternating sums related to intersections. Again from our running example using the index sets $\{123, 234, 345\}$ we have

$$H(12345) = H(123 \cup 234 \cup 345)$$

$$= H(123) + H(234) + H(345) - H(23) - H(3) - H(34) + H(3)$$

$$= H(123) + H(234) + H(345) - H(23) - H(34).$$

Note that the complexity of the first inclusion–exclusion identity is simplified. The simplification here is an example of a property called the running intersection property, which we now explain.

Definition 4.3. A sequence of index sets $I_1, \ldots, I_k$ has the *running intersection property* if for any triple I_i, I_j, I_k with $i < j < k$ we have $I_i \cap I_k \subset I_j$.

The result we have used is the following, which is known as Rota's inclusion–exclusion principle; see, for example, [13, 14].

Lemma 4.4. *For a sequence of index sets $\mathcal{I} = \{I_1, \ldots, I_k\}$ with the running intersection property and a valuation $\phi(I)$ on the distributed lattice generated by $\mathcal{I}$, we have the simplified inclusion–exclusion formula as*

$$\phi\left(\cup_{i=1}^k I_i\right) = \sum_{i=1}^k \phi(I_i) - \sum_{i,j=1,\ i<j}^k \phi(I_i \cap I_j).$$

It is worth noting that the fact that the generators of the LCI have the running intersection property shows that if we only consider the first-order intersections, then we have a decomposable graphical model [17]. We see this here in that the LCI gives many more inclusion–exclusion formulas than the single one above. For the LCI model every information associated with a margin, E_I is the sum of the information associated with each conditional probability given by a z_i (for $i \in I$).

5 | DISCUSSION

This paper is a contribution to the area of algebraic statistics but one which is somewhat different from the usual log-linear models' approach which leads to a toric representation for the raw probabilities. There *is* a toric representation, but for the marginal distributions. It is based on the simple observation that the distributive lattice definition of Hibi ideals is identically providing an exact mapping to a distributive lattice-based probabilistic model, called an LCI model. These were originally introduced for Gaussian models where the lattice arises from a particular selection of margins, but the formula is now quite general. The TDAG dual representation of the LCI model also has a purely algebraic representation in terms of a special bipartite graph arising from Alexander duality, a theory discovered independently of the probabilistic representation. The toric representation derives as the elimination ideal generated by certain products of variables z_i that can be themselves associated with conditional independences, and the $\log z_i$ give the terms in (i) appropriate log-linear models, (ii) time series models and (iii) models for Shannon information. We suggest that this algebraic underpinning of LCI models will help make them more widely applicable.

All surplus income from its publishing programme is used to support mathematicians and mathematics research in the form of research grants, conference grants, prizes, initiatives for early career researchers and the promotion of mathematics.

REFERENCES

1. S. A. Andersson, D. Madigan, M. D. Perlman, and C. M. Triggs, *On the relation between conditional independence models determined by finite distributive lattices and by directed acyclic graphs*, J. Statist. Plann. Inference **48** (1995), no. 1, 25–46.

2. S. A. Andersson and M. D. Perlman, *Lattice models for conditional independence in a multivariate normal distribution*, Ann. Statist. **21** (1993), 1318–1358.

3. N. Beerenwinkel, N. Eriksson, and B. Sturmfels, *Conjunctive Bayesian networks*, Bernoulli **13** (2007), 893–909.

4. P. E. Caines and H. P. Wynn, *An algebraic framework for Bayes nets of time series*, Modeling, estimation and control (A. Chiuso, S. Pinzoni, A. Ferrante, eds.), Springer, Berlin, 2007, pp. 45–57.

5. R. Castelo and A. Siebes, *A characterization of moral transitive acyclic directed graph Markov models as labeled trees*, J. Statist. Plann. Inference **115** (2003), no. 1, 235–259.

6. V. Ene, J. Herzog, and F. Mohammadi, *Monomial ideals and toric rings of Hibi type arising from a finite poset*, Eur. J. Combin. **32** (2011), no. 3, 404–421.

7. B. Giglio and H. P. Wynn, *Monomial ideals and the Scarf complex for coherent systems in reliability theory*, Ann. Statist. **32** (2004), no. 3, 1289–1311.

8. D. R. Grayson and M. E. Stillman, *Macaulay2, a software system for research in algebraic geometry.* https://faculty.math.illinois.edu/Macaulay2/.

9. J. Herzog and T. Hibi, *Distributive lattices, bipartite graphs and Alexander duality*, J. Algebraic Combin. **22** (2005), no. 3, 289–302.

10. J. Herzog and T. Hibi, *Monomial ideals*, Springer, Berlin, 2011, pp. 3–22.

11. T. Hibi, *Distributive lattices, affine semigroup rings and algebras with straightening laws*, Commutative algebra and combinatorics, Mathematical Society of Japan, Tokyo, 1987, pp. 93–109.

12. M. Jozsa, M. Petreczky, and M. K. Camlibel, *Causality and network graph in general bilinear state-space representations*, IEEE Trans. Automat. Control, IEEE, Piscataway, NJ, 2019, pp. 3623–3630.

13. D. A. Klain and G.-C. Rota, *Introduction to geometric probability*, Cambridge University Press, Cambridge, 1997.

14. K. H. Knuth, *Valuations on lattices and their application to information theory*, IEEE Int. Conf. Fuzzy Syst. IEEE, Piscataway, NJ, 2006, pp. 217–224.

15. J. P. Kung, G.-C. Rota, and C. H. Yan, *Combinatorics: the Rota way*, Cambridge University Press, Cambridge, 2009.

16. S. L. Lauritzen, *Graphical models*, vol. 17, Clarendon Press, Oxford, 1996.

17. S. L. Lauritzen and D. J. Spiegelhalter, *Local computations with probabilities on graphical structures and their application to expert systems*, J. R. Stat. Soc. **50** (1988), no. 2, 157–194.

18. H. Massam and E. Neher, *Estimation and testing for lattice conditional independence models on Euclidean Jordan algebras*, Ann. Statist. **26** (1998), 1051–1082.

19. E. Miller and B. Sturmfels, *Combinatorial commutative algebra*, vol. 227, Springer Science & Business Media, Berlin, 2004.

20. F. Mohammadi, E. Sáenz-de Cabezón, and H. P. Wynn, *The algebraic method in tree percolation*, SIAM J. Discrete Math. **30** (2016), no. 2, 1193–1212.

21. G.-C. Rota, *On the foundations of combinatorial theory I, Theory of Möbius functions*, Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete **2** (1964), no. 4, 340–368.

22. R. P. Stanley, *Enumerative Combinatorics*, vol. I, Wadsworth & Brooks/Cole, Monterey, CA, 1986.