CRM Master Research Projects

CENTRE DE RECERCA MATEMÀTICA

Bayesian Network Studies for Splicing Regulatory Elements

Author: Anna M. Kedzierska

Advisor: Marta Casanellas

Estudi de xarxes bayesianes per a elements de regulació de splicing.

Bayesian network studies for splicing regulatory elements.

Anna M. Kedzierska
supervisor: Marta Casanellas

Research report
Matematica Aplicada I
Universitat Politècnica de Catalunya

2008

Table of Contents

Introduction

The motivation to pursue the project and for its particular orientation is driven by the large-scale genomic data available. Though abundant, it exhibits noise and sparseness, thus sound inference techniques are needed to retrieve the underlying information. Statistics has proved fundamental in many real-world applications, including gene finding and phylogenetics [1]. With a plethora of biological data becoming available, it constitutes an integral part of the inference tools in Bioinformatics and Computational Biology.
Algebraic statistics describes statistical models via sets of polynomial equations. Therefore, through marrying these concepts with notions from algebraic geometry, it provides alternative tools for studying the statistical models. In particular, many of them can be described as real algebraic varieties ([16]). The use of algebraic geometry for statistical inference has demonstrated to be an efficient tool in phylogenetics (e.g. [3], [2], [4]).

We believe that algebraic models are capable of capturing the underlying network structure, while sidestepping the issue of model overparametrization. In this work we wish to focus on genomic data that exhibits certain interactions, e.g. transcription binding sites, splicing regulators, regulatory pathways. We think of our observations as random variables having certain underlying structure of interactions. Natural setting to study such models is provided by the directed graphical models commonly known as Bayesian Newtorks (Belief Networks). Graphical models based on directed graphs proved to be fundamental and efficient for representing joint probability distribution over many variables.

This refers to the situations, where either each or certain subsets of variables are conditionally independent of all but a few of the random variables considered. As a consequence, they are a commonly used tool on a wealth of applications, ranging from biomedical analyses to data mining and machine learning.
In Chapter 1, we introduce the notion of marginal and conditional independence. Next, we describe the architecture of interactions under study, directed Markov models. Details on their algebraic characterizations and an introductory study for a small network are given in Chapter 2. Finally in Chapter 4, we talk about hypotheses under study, future directions and possible applications to the real-world data

Chapter 1

Bayesian Networks

1.1 Conditional Independence

Let (Ω, Σ, P) be a probability space, where Ω is a sample space, Σ is the corresponding $\sigma-$algebra of subsets of Ω and P a probability measure.

Definition 1.1.1. A set of events $\{X_j \mid j \in J\} \in \Sigma$, are said to be *(mutually) independent* if and only if for all subsets of indices $I \subseteq J$ we have:

$$\forall_{I \subset J} P(\bigcap_{i \in I} X_i) = \prod_{i \in I} P(X_i)$$

Definition 1.1.2. For two disjoint sets $X_i, Y \in \Sigma$ such that $P(Y) > 0$, *conditional probability* of X_i given the values of Y is defined as follows:

$$P(X_i \mid Y) = \frac{P(X_i \cap Y)}{P(Y)}$$

Note that for any set Y occurring with a non-zero probability, $P(\cdot \mid Y)$ defines another probabilistic measure in Σ as any two sets having the same intersection with Y have also the same probability.

Definition 1.1.3. Pairwise disjoint subsets $X_i \in \Sigma$ are said to be *conditionally independent* given Y $(P(Y) > 0)$ if and only if:

$$P(\bigcap_{i \in I} X_i \mid Y) = \prod_{i \in I} P(X_i \mid Y)$$

In particular, let X_A,X_B,X_C be three pairwise disjoint subsets of Σ (e.g. $X_A = \{X_1, X_2\}$). X_A is conditionally independent of X_B given X_C (we write $X_A \perp X_B \mid X_C$) if:

$$X_A \perp X_B \mid X_C \iff P(X_A, X_B \mid X_C) = P(X_A \mid X_C) P(X_B \mid X_C).$$

Conditional independence takes its meaning under the joint probability distribution P, that is $X_A \perp X_B \mid X_C[P]$, however, is commonly skipped for notational convenience. Note that if $X_C = \emptyset$, then we get the independence of X_A and X_B. There are five basic conditional independence axioms:

1. symmetry $X_A \perp X_B \mid X_C \Longrightarrow X_B \perp X_A \mid X_C$
2. decomposition $X_A \perp (X_B \cup X_D) \mid X_C \Longrightarrow X_A \perp X_D \mid X_C$
3. weak union $X_A \perp (X_B \cup X_D) \mid X_C \Longrightarrow X_A \perp X_B \mid (X_C \cup X_D)$
4. contraction $X_A \perp X_B \mid (X_C \cup X_D) \wedge X_A \perp X_D \mid X_C \Longrightarrow X_A \perp (X_B \cup X_D) \mid X_C$
5. for strictly positive distributions $P > 0$, the intersection axiom holds $X_A \perp X_B \mid (X_C \cup X_D) \wedge X_A \perp X_C \mid (X_B \cup X_D) \Longrightarrow X_A \perp (X_B \cup X_C) \mid X_D$

The above list of statements is not exhaustive. The existence of additional ones derived independently or partially from the above axioms, proved useful in practical computations. More details can be found in [20], [12] or [15].

From here on, we restrict our focus to the case where $X = \{X_1, \ldots, X_n\}$ is a collection of discrete random variables. We assume each X_i takes values on a finite domain, $[d_i]$. The joint distribution of X, $P(X_1, \ldots, X_n)$, can be viewed as a $[d_1] \times \ldots \times [d_n]$ multidimensional table.

Definition 1.1.4. *A conditional independence* (CI) model M_{CI} is formally defined through a family of probability distributions satisfying a number of conditional independence restrictions:

$$M_{CI} = \{X_{A_1} \perp X_{B_1} \mid X_{C_1}, \ldots, X_{A_n} \perp X_{B_n} \mid X_{C_n}\} \tag{1.1.1}$$

where $\forall i \quad A_i, B_i, C_i$ are disjoint subsets of X

The CI statements $X_{A_i} \perp X_{B_i} \mid X_{C_i}$ reflect the dependencies between random variables. In this work, we focus on the CI statements derived from directed graphs. We investigate their description as algebraic models and in particular, as instances of algebraic varieties.

1.2 Conditional independence of BN

A particular class of conditional independence models are graphical models. The corresponding independence statements arise from the separating properties of the graph underlying the model. We will focus on the special case of the so-called directed graphical models, commonly referred to as Bayesian Networks (Belief Networks). For the reasons of being one of the fundamental tools in a wealth of applications, these models have been widely studied. Our introduction to the concepts and notation draws upon the one of [12].

Definition 1.2.1. *A directed acyclic graph* $G(V, E)$ (DAG) is a simple graph over a finite set of nodes V (vertices), where all the edges E are oriented (arcs) and there are no directed cycles.

The first component, V, corresponds to the random variables in X via a $1-1$ correspondence. Random variables defined over the nodes in a DAG, $X = \{X_1, \ldots, X_n\}$, have a joint distribution $p(x_1, \ldots, x_n) = p(X_1 = x_1, \ldots, X_n = x_n)$. Here, x_i represents a realization of a random variable X_i. The edges represent the (conditional) dependencies between the elements in V.

Two nodes are connected, that is $e(v_i, v_j) \in E$, whenever there exists an edge $v_i \rightarrow v_j$ for $v_i, v_j \in V$. We say that two nodes are neighbours or adjacent if either $e(v_i, v_j)$ or $e(v_j, v_i)$. By definition there are no directed cycles, being a

sequence of nodes $v_1, \ldots, v_l$, s.t. $\forall k = 1, \ldots l-1 \quad e(v_k, v_{k+1}) \in E$ and $e(v_1, v_l) \in E$. Acyclicity assures that following the arrows in the graph, it is impossible to return to the initial point. A parental set of v_i denoted by $pa(v_i)$ is a the set of such nodes v_j for which we have $e(v_j, v_i) \in E$. A path between two nodes, say v_i and v_j, is a set of vertices $\{v_1, \ldots, v_k\}$ s.t. $\forall l = 1, \ldots, k-1$ we have $e(v_l, v_{l+1}) \in E$. The set of ancestral nodes of v_i $(an(v_i))$ consists of all the nodes v_j such that there exists a path from v_j to v_i $(j \neq i)$. The set of descendants of v_i, $nd(v_i)$, are those v_j, for which exists a directed path from v_i to v_j. Lastly, the set of nondescendants of v_i is defined to be $nd(v_i) = V \setminus ((de(v_i) \cup pa(v_i))$.
Recursive factorization of a probability distribution according to a DAG is based on assigning a probability measure to each node, $p(X_i, pa(X_i))$ such that
$\sum_{i_j \in [d_i]} p(X_i = i_j \mid (i_1 \ldots i_n) \mid_{pa(X_i)}) = 1$ and

$$p(X) = \prod_{X_i \in X} p(X_i \mid pa(X_i)). \tag{1.2.1}$$

Note: for the sake of notation we denote by $(i_1 \ldots i_n) \mid_{pa(X_i)}$ the subset of indices corresponding to the parents of X_i. Unless stated otherwise, their realization is assumed fixed.
A probability distribution, p, *factorizes* over a DAG G if (1.2.1) holds. Alternatively, in line with the above interpretation that a particular variable is independent of the remaining variables given its parents, we say that the probability distribution p is Markov given G.

Definition 1.2.2. *A Bayesian Network* consists of an annotated directed acyclic graph, G, and a family of conditional probability distribution, β, for which the recursive factorization property, (1.2.1), holds.

From here onwards a Bayesian Network will be abbreviated as BN.
Note that β represents a set of parameters that quantify the network. The family of these conditional distributions is usually assumed fixed and chosen beforehand, thus the factorization depends solely on the underlying network structure. Specifically, if a probability distribution β is Markov given G, we say that it belongs to the model defined by G.

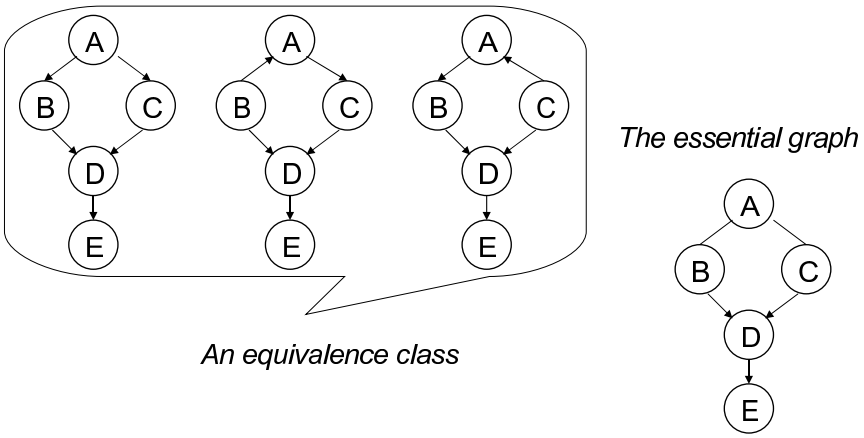

Figure 1.1: Illustration of the equivalence classes for a Bayesian Network on five nodes with the conditional independence model $M_{CI} = \{B \perp C \mid A, A \perp D \mid (B, C), (A, B, C) \perp E \mid D\}$; *Left:* the graphs constituting an equivalence class *Right:* the essential graph representing the three-element equivalence class defined by M_{CI}

As a simple example, consider BN in Figure 1.1 shows a few possible graph structure with four observed nodes. Usually in direct applications the vertices have a clear interpretation associated to them. For instance, one may consider a biomedical survey with binary outcomes described by the leftmost graph in Figure 1.1. Let us consider the effects of age the patients ($A = \{0 = child, 1 = adult\}$) examined for two types of diseases ($B, C = \{0 = presence, 1 = absence\}$) and being a smoker ($D = \{0 = no, 1 = yes\}$) on the general level of tiredness ($E = \{0 = low, 1 = high\}$). The factorization of the joint probability function takes the form: $p(A, B, C, D) = p(A)p(B \mid A)p(C \mid A)p(D \mid B, C)p(E \mid D)$.

1.2.0.1 Conditional independence and factorization theorem

There exists a range of independence relationships, the so-called Markov properties, that can be associated to a G. We will focus on the two predominant notions (e.g. [12] gives a detailed overview).
As a consequence of the factorization property (1.2.1), one can define:

Definition 1.2.3. *The directed local Markov property* with respect to a directed acyclic graph, G, states that any node is independent of its nondescendants given the values of its parents:

$$X_i \perp nd(X_i)) \mid pa(X_i) \quad \forall X_i \in V \tag{1.2.2}$$

Let us define a conditional independence model corresponding to (1.2.2) on G as

Definition 1.2.4. [*Local conditional independence model*]

$$M_{local} = \{X_i \perp nd(X_i) \mid pa(X_i) \quad \forall X_i \in V \quad \forall X_i \in V\} \tag{1.2.3}$$

Another notion of independence is provided by the global Markov property. It is based on the concept of $d-$separation according to G. A set X_C *d-separates* X_A and X_B if all the paths from X_A to X_B are blocked. We say that a path, π, between two nodes is *blocked* by elements in X_C if:

1. π contains a node in X_C, which is not a collider (serial or diverging)
2. π contains a node that is a collider, but it does not belong to X_C and neither does any of its descendants.

Definition 1.2.5. [Directed global Markov property]
Let p belong to a model defined by G on X. Then $X_A \perp X_B \mid X_C$ whenever X_A and X_B are d-separated by X_C in G.

The separation property defined by the global Markov independence statements is often denoted by $\perp_G$.
Analogously to the local CI model, *a global conditional independence model*, M_{global}, on G is defined through the sets of directed global CI statements:

Definition 1.2.6. [Global conditional independence model]

$$M_{global} = \{X_A \perp_G X_B \mid X_C\} \tag{1.2.4}$$

Note that its is not difficult to see that $M_{local} \subseteq M_{global}$. In general, the inclusion is strict (e.g. the equality holds when Gs are complete).

As we will see in later in the text, conditional independence statements correspond to polynomial equations on the joint distribution. Therefore, M_{local} and M_{global} will be interpreted as algebraic sets.

The following theorem proved in [12] (p.51) links the factorization property (probability theory) and the probabilistic independence (graph theory). It states that a family of distributions associated to a DAG can be described in two ways, which are equivalent.

Theorem 1.2.7. *[Factorization Theorem] Let X be a set of random variables with a joint distribution $p(X)$ and G a DAG over X. The following statements are equivalent:*

1. *$p(X)$ obeys the directed local Markov property*
2. *$p(X)$ obeys the directed global Markov property*
3. *$p(X)$ factors according to G*

Thus for every DAG, G, we have that a joint probability distribution p obeys the local Markov property with respect to G if and only if it obeys the global Markov property for G.

1.2.0.2 Network inference

As BN are probabilistic models, one can ask a number of questions ranging from the identification of the underlying structure of the graph to the parameter estimation. The latter, assumes the network to be known and asks about the Markovian transition matrices assigned to the arcs. Yet, there is a number of problematic issues including the choice of the probabilistic model β, optimal data size, dealing with missing values, overparametrization. Likewise, the number of possible graphs grows exponentially in the number of nodes. The Weissmtein's conjecture states that the number of DAGs with n labeled vertices is equal to the number of equivalence classes of the $n \times n$ $(0,1)$-matrices with positive real eigenvalues (http://www.research.att.com/~njas/sequences/A003087, [13]). If a_n is the number of DAGs with n labeled vertices, then

$$R_n = \sum_{k=1}^{n} (-1)^{k+1} \binom{n}{k} 2^{k(n-k)} a_{n-k},$$

where $n \geq 1, a_0 = 1$. Asymptotically a_n converges to $n! \frac{2^{\binom{n}{2}}}{Mp^n}$ as $n \to \infty$, where $p = 1.488(8)$ and $M = 0.474(74)$. In order to give an idea about the exponential growth of the search space through possible graphs, a few first counts for $n = 1, \ldots 5$ nodes are $(1, 3, 25, 543, 29281)$.

As the conditional independence statements vary depending on the model, one should be able to distinguish between the models based on the structure of M_{CI} (1.1.1). However, network identification is potentially challenged by the one-to-many correspondence between a statistical model represented and its associated DAGs. Even most successful algorithms for such model selection (e.g. PC-algorithm of [17]) assume (markovian) faithfulness, that is to say, that all the conditional statements belonging to the model can be read of the graph (have a perfect representation in the DAG).

However, it is known that as much as one can hope is to identify the network up to an equivalence class (see [15] for details):

Definition 1.2.8. Two directed acyclic graphs, G and G' are said to be *Markov equivalent* if M_{local} (M_{global}) agree on G and G'

Definition 1.2.9. *A skeleton* of a DAG is the set of arcs joining its vertices with the directions removed.

Theorem 1.2.10. *G and G' are equivalent if and only if*

1. *G and G' have the same skeletons*
2. *G and G' have the same unmarried parents (equivalently, the same $V-$structure, being its sets of colliders)*

The problem of graph equivalence can be eluded by the use of the so-called Partially Directed Acyclic Graph (PDAG). PDAGs, which bridge the properties of undirected and directed graphs, have received a great deal of attention thus far in the algebraic research.
This section makes mention of one the future directions of this research project concerning the translation of the PC-algorithm into the language of algebraic statistics (see section 3.3).

Chapter 2

Algebraic statistics for bayesian newtorks

2.1 Algebraic statistical models

This section brings together the ideas described earlier in the text. Bayesian networks can be described in two ways: implicitly by a set of conditional independence statements associated with the graph-based model, or parametrically through mapping the parameters onto the set of distributions (see Theorem 1.2.7). In the next sections we introduce both approaches from the algebraic point of view and state the Hammersley-Clifford theorem, which proves their equivalence. The focus of hereon will be placed on the algebraic description of the graphical models.
In order to introduce the algebraic description, we let $\Bbbk$ be an algebraically closed field and $\Bbbk^n$ be the affine n-space over $\Bbbk$. We call $\Bbbk[x_1 \dots x_n]$, a ring of polynomials in the indeterminants $x_1, \dots, x_n$ and coefficients in $\Bbbk$.

Definition 2.1.1. *An algebraic statistical model* is a parametric statistical model, where the probability distribution is a polynomial function in the parameters.

Namely, if X is a vector of discrete random variables with the cardinality of the space space m, a polynomial function is used to map the set of parameters

into a set of distributions.

$$\psi : \Bbbk^n \mapsto \Bbbk^m,$$

where $\forall x = x_1 \dots x_n \in \Bbbk^n$, $\psi(x_1 \dots x_n) = (g_1(x_1 \dots x_n), ..., g_m(x_1 \dots x_n))$ and $g_1, ..., g_m \in \Bbbk[x_1 \dots x_n]$. The description through one or more polynomial expressions arises from the factorization of the distribution according to the graph as defined in (1.2.1).

Definition 2.1.2. *An ideal I is a subset of $\Bbbk[x_1 \dots x_n]$ satisfying:*

1. $0 \in I$
2. *if $f, g \in I$, then $f + g \in I$,*
3. *if $f \in I$ and $h \in \Bbbk[x_1 \dots x_n]$, then $hf \in I$.*

We say that an ideal is *generated* by $f_1, \dots, f_n$ is defined if:

$$(f_1, \dots, f_n) = \{\sum_{i=1}^{n} a_i f_i \mid a_i \in \Bbbk[x_1 \dots x_n]\}$$

Definition 2.1.3. *An algebraic set $V \subset \Bbbk^n$* is the set of common zeroes of a collection of polynomials $S \subset \Bbbk[x_1 \dots x_n]$:

$$V = V(S) = \{x \in \Bbbk^n \mid f(x) = 0 \quad \forall f \in S\}$$

Definition 2.1.4. *Let $X \in \mathbb{R}^n$. The ideal of X is defined as:*

$$I(X) = \{f \in \Bbbk[x_1 \dots x_n] \mid f(x) = 0 \quad \forall x \in X\}$$

An algebraic set is a *variety* if and only if the polynomials defining it generate a prime ideal in the polynomial ring. In other words, it is an irreducible algebraic set.

As will be shown in Sections 2.1.1 and 2.1.2 that the independence models in (1.2.2) and (1.2.4)) can by viewed as solutions to sets of polynomial equation. On

the other hand, they lie in the closure of the image of a certain polynomial map (see Section 2.1.1). the DAG Hilbert's basis theorem states that every ideal in $\Bbbk[x_1 \dots x_n]$ is finitely generated, that is for every ideal I, there exists a finite set of polynomials $f_i \in \Bbbk[x_1 \dots x_n]$, s.t. $I = (f_1, \dots f_s)$. In particular, any algebraic set $V(S)$ is an algebraic set for a finite collection of polynomials $V(S) = V(< S >) = V(f_1, \dots, f_s)$. Therefore, so as to answer whether a probability distribution belongs to a model, it is always possible to choose a finite list of conditions to be checked.

As mentioned earlier in the text, graphical models can be described algebraically in two ways: parametrically and implicitly. Both approaches are based on the notion that those statistical models are the zero set of a list of polynomials. The characterization of the generating set of such polynomial equations is not a trivial task. However, as we will see, for a certain class of models the generators of the independence ideals are binomials.
Let $p_{i_1,\dots,i_n}$ be indeterminants denoting $p(X_i = i_1, \dots, X_n = i_n)$. We let *R[P]* to be the ring of polynomial functions generated by these unknowns in $\mathbb{R}^n$.
For an extensive introduction to the topic refer to [14] or [19] (for Markov fields). We briefly introduce both settings in their general form and next illustrate the ideas and motivation for this study by focusing on the specific case of the binary BN with three nodes.

2.1.1 Parametrization

Recall that $X = \{X_1, \ldots, X_n\}$ is a collection of n discrete random variables. Let us denote by $[d_i]$ the sample space for X_i. For notational convenience, let $|d_i| = \#[d_i]$ and $N = |d_1| \times |d_n|$ be the cardinality of $[d_i]$ and the product space of X, respectively. The joint distribution $p_i = p_{i_1 \ldots i_n} = P(X_k = i_k)_{k=1}^n, i = (i_k)_{k=1}^n$ defines a table of probabilities, $p = (p_i)$. Given the condition $\sum_{i=1}^N p_i = 1$, it becomes an element of the probability simplex:

$$\triangle_{N-1} = \{p \in \mathbb{R}^N : \quad \forall i \quad p_i \geq 0, \sum_{i=1}^{N} p_i = 1\} \subset \mathbb{R}^N_{\geq 0}$$

As mentioned in the previous sections, independence models constitute a subclass of the general log-linear models (also referred to as toric models) and thus have a neat representation in the parameter space of integer tables. For more details see e.g. [10]
We introduce q to be a set of model parameters:

$$q^i_{jk} := q^i_{i_j,(i_1 \ldots i_n)_{|pa(X_i)=k}} = p(X_i = i_j \mid pa(X_i) = k) \tag{2.1.1}$$

Note that k denotes here the values of the parental nodes of X_i. Formally, $k \subseteq (i_1 \ldots i_n) : \forall s \in k X_s \in pa(X_i)$. Let $R[Q]$ a polynomial ring spanned by these indeterminants and d be their cardinality, that is $d = |q| = \#[q^i_{jk}]$.
Define a matrix $\mathbf{A} = (a_{ij}) \in \mathbb{Z}^{d \times N}$ over non-negative integers with equal sums of column entries, that is $\forall j \quad \sum a_{ij} = const.$ We label the rows of $\mathbf{A}$ by the elements of $R[Q]$ and as a consequence, the columns are the elements of $R[P]$ in the monomial representation in $'$. Matrix $\mathbf{A}$ has an associated log-linear model consisting of a set of probability tables p such that:

$$\mathbb{M}_{\mathbf{A}} = \{p \in (\triangle_{N-1}) : \log p \in rowspan(\mathbf{A})\}.$$

The above is equivalent to stating that $\log p \in image(\mathbf{A}^T)$ if there exists $1 \leq k \leq d$ such that $p_i = \exp(A^T q_k), q_k \in \mathbb{R}^d$. This in turn provides a representation of the map from $\mathbb{R}^d$ to $\mathbb{R}^N$:

$$\psi : q = (q_i)_{i=1}^d \mapsto p = (p_i)_{i=1}^N \tag{2.1.2}$$

In general terms, the parameters q's are the so-called generators of the log-linear model and index the rows of the associated matrix $\mathbf{A}$. Note that the coordinates of ψ are polynomials in model parameters. This clearly follows from (2.1.1) and (1.2.1). Consequently we have:

$$\psi : p_{i_1 \ldots i_n} = \prod_{X_i \in X} q^i_{i_j,(i_1 \ldots, i_n|_{pa(X_i)})}$$

It follows from the definition of $\mathbf{A}$ that the monomials on the right-hand side of the above equation have equal degrees.
The image of ψ becomes the toric variety of the log-linear model $\mathbb{M}_{\mathbf{A}}$. Toric models have desirable properties from the algebraic standpoint, as their ideals are generated by binomials. It follows that the primary components of the toric ideal will also be generated by binomials. Indeed, it is a known result that varieties obtained through a monomial parametrization are defined by the binomial equations [18]. However, not every BN with its conditional independence structure is toric.

$image(\psi)$ is not in general an algebraic set, however, we can consider its closure, $\overline{image(\psi)}$ as the smaller algebraic set containing it. It can be thus checked that:

$$I(\overline{image(\psi)}) = I(image(\psi)) = \ker(\Psi),$$

where Ψ is the ring homomorphism

$$\Psi : \quad R[P] \mapsto R[Q], \quad p_{i_1,\ldots,i_n} \mapsto \prod_{X_i \in X} q^i_{i_j,(i_1 \ldots i_n|_{pa(X_j)})}$$

The above holds independently of the base field. *Hilbert-Nullstellensatz* (see e.g. [7]) establishes a $1-1$ correspondence between the radical ideals and varieties. This results suggests that oftentimes it is convenient to work over complex numbers field.

2.1.2 Implicit description

In Section 1.1 we described the construction of the conditional independence model, M_{CI} (see Definition (1.1.1)), defined by the set of restrictions on the subsets of random variables (X_A, X_B, X_C) in X. As before, we assume that random variables are discrete and X_i take values in $[d_i]$. In order to describe the independence statements $X_A \perp X_B \mid X_C$ implicitly, we introduce a set of indeterminants $p_{i_a i_b i_c} = P(X_A = i_a, X_B = i_b, X_C = i_c)$ in $R[P]$. Each of those expressions is obtained by marginalization of the joint probability $\mathbf{p}_i = p_{i_1 \dots i_n} = P(X_1 = i_1, \dots, X_n = i_n), \forall i_k \in [d_k]$. The *saturation* of an indeterminant $p_{i_a i_b i_c}$ with respect to the full space boils down to integrating out the remaining variables. Thus the saturated statements take the form:

$$p_{i_a i_b i_c} = p_{+i_a + \dots i_b + \dots i_c +}.$$

In the discrete case the statement $X_A \perp X_B \mid X_C$ boils down to rank-one restrictions placed on the corresponding $|d_C|$ (recall $|d_C| = \#[d_C]$) matrices of size $[d_A] \times [d_B]$. This, in turn, is equivalent to the vanishing condition on all the minors of the above matrices. Namely,

$$X_A \perp X_B \mid X_C \leftrightarrow p_{i_a i_b i_c} p_{j_a j_b i_c} - p_{i_a j_b i_c} p_{j_a i_b i_c} \forall i_a, j_a \in [d_A], i_b, j_b \in [d_B], i_c \in [d_C] \tag{2.1.3}$$

Note that the number of such equations equals $\binom{|d_A|}{2}\binom{|d_B|}{2} \times |d_C|)$.
The polynomials defined in (2.1.3) generate an ideal in $R[P]$ that we will call $I_{A \perp B|C}$. In turn, a collection of such statements defines an ideal I, being a sum of the ideals generated by each of them individually. As a consequence, we can define an independence ideal of the model given in (1.1.1):

$$I_{CI} = I_{A_1 \perp B_1 | C_!} + \dots + I_{A_n \perp B_n | C_n} \tag{2.1.4}$$

The common zeroes of the polynomial equations in I_{CI} is an algebraic set $V(I_{CI})$. Therefore, $V(I_{CI})$ is a subset of those tables $\mathbf{p} = (\mathbf{p}_i)_i$ in $\mathbb{R}^m$, which are the solution set of the polynomial equations defined by (2.1.4).

One approach to study the properties of such models is via the primary decomposition of I_{CI}. If I_{CI} is binomial, it is usually the case that the components of the decompositions are interpretable in terms of the conditional independence statements. It follows from (2.1.3) that when M_{CI} consists of the saturated independence statements, That is to say, if for all $X_A \perp X_B \mid X_C$ we have that $X_A \cup X_B \cup X_C = X$, I_{CI} will always be binomial. However, a straightforward interpretation of the elements in such decomposition is not possible in more complex cases.

2.1.3 Comparative study

In the previous sections, we defined the conditional independence models, (1.2.2) and (1.2.4), based on the notions of the local and global Markov properties associated to a G. *The local independence ideal*, I_{local}, is an independence ideal as defined in (2.1.4) generated by the local markovian statements contained in M_{local}. Similarly, I_{global} corresponds to the quadric (2.1.3) imposed by the global Markov conditional restrictions as defined in (1.2.4).

Definition 2.1.5. *The local/global independence variety* is defined as

$$V_{local} = V(I_{local}), \qquad V_{global} = V(I_{global})$$

There are usually more conditional independence statements in I_{global} than in I_{local}. Formally we have:

$$I_{local} \subseteq I_{global} \tag{2.1.5}$$

This implies the inverse relationship for the corresponding varieties:

$$V_{local} \supseteq V_{global} \tag{2.1.6}$$

Let us recall the parametric formulation of the independence model from section 2.1.1. The non-negative functions associated to the vertices of a DAG form the domain of the parametric map as defined in (2.1.2). In reference to the notation used previously, let us denote by q' the parameters as in (2.1.1) restricted to the probability simplex, that is to say:

$$q' = \{q : \sum_j q^i_{jk} = 1\}$$

Recall that this is equivalent to $\sum_{i_j \in [d_i]} \quad p(X_i = i_j \mid pa(X_i) = i_k) = 1$, where i_k is a particular realization of the parental set of states, $pa(X_i)$. We denote by ψ' the analog of the map (2.1.2):

$$\psi' : q' = (q'_i)_{i=1}^d \longrightarrow p = (p_i)_{i=1}^N \tag{2.1.7}$$

and by Ψ' the corresponding ring homomorphism. As a consequence of the above transformation, it can be easily checked that:

$$I_{local} \subseteq I_{global} \subseteq \ker(\Psi') \tag{2.1.8}$$

The following theorem states that algebraic varieties of directed graphical models given through explicit or parametric representation, coincide when restricted to the probability simplex:

Theorem 2.1.6 (Hammersley-Clifford theorem)**.**

$$V_{\geq 0}(I_{local} + < p-1 >) = V_{\geq 0}(I_{global} + < p-1 >) = image(\psi'_{\geq 0}) = \ker(\Psi'))$$

Chapter 3

Algebraic statistics for small bayesian newtorks

3.1 BN with three nodes

Figure 3.1: $M_1 = \{(X_1, X_2) \perp X_3\}$.

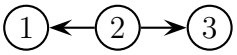

Figure 3.2: $M_2 = \{X_1 \perp X_3 \mid X_2\}$

Figure 3.3: $M_3 = \{X_1 \perp X_3\}$

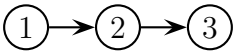

Figure 3.4: $M_4 = \{X_1 \perp X_3 \mid X_2\}$

Figure 3.5: $M_5 = \{X_1 \perp X_2 \perp X_3\}$

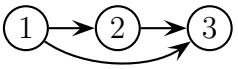

Figure 3.6: $M_6 = \emptyset$

Let G be a DAG on three binary nodes: $X = \{X_1,X_2,X_3\}$ and $\forall_{i=1:3} \quad [d_i] = \{0, 1\}$. We will use a computational algebra software SINGULAR [11] to study the relationship between I_{local}, I_{global} and $ker(\Psi)$ for binary Bayesian networks on three nodes.

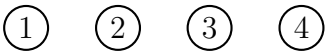

Figure 3.7: $M_1^4 = \{X_1 \perp X_2 \perp X_3 \perp X4\}$

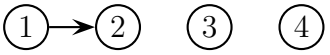

Figure 3.8: $M_2^4 = \{(X_1, X_2) \perp X_3 \perp X_4\}$.

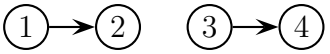

Figure 3.9: $M_3^4 = \{(X_1, X_2) \perp (X_3, X_4)\}$

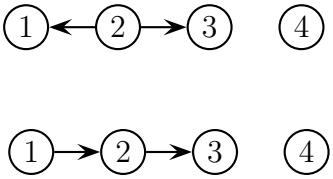

Figure 3.10: $M_4^4 = M_5^4 = \{X_1 \perp X_3 \mid X_2, (X_1, X_2, X_3) \perp X_4\}$

3.1.1 Parametric representation

We give the parametrized models as introduced in Section 2.1.1. From (1.2.1) we can factorize the joint distribution, $p := P(X = x) = (p_{ijk})_{ijk \in \{0,1\}}$, where $p_{i_1 i_2 i_3} = p(X_1 = i_1, X_2 = i_2, X_3 = i_3)$, according to G. Denoting by $q_{jk}^i := p(X_i = i_j \mid pa(X_i) = k)$ the parameters of the model, we get:

$$p_{ijk} = \prod_k q_{i_1 k}^1 \prod_k q_{i_2 k}^2 \prod_k q_{i_3 k}^3.$$

Note that for a DAG, the indeterminants q correspond to the parameters placed on the oriented edges.

For a BN on three binary nodes, a probability distribution takes values on the $N = 2^3 = 8$ dimensional binary domain:

$$p = (p_{000}, p_{001}, p_{010}, p_{011}, p_{100}, p_{101}, p_{110}, p_{111}).$$

It factorizes to G if it lies in the image of the map ψ as defined in (2.1.2). By the same token, a distribution belongs to a model if it lies in the kernel of the ring homomorphism $\Psi : R[Q] \mapsto R[P]$.
As a starting point we describe the definitions of the matrix **A** for the distinct graphs. Note that in each of the cases **A** has N columns consisting of the unit vectors $(e_1, e_2, e_3)^T$, where e_i is a unit vector in $\mathbb{R}^{|d_i|}$. Again, $d = |q|$.

Figure 3.1 The network corresponding can be described as $pa(X_1) = pa(X_3) = \emptyset, pa(X_2) = X_1,\ d = 8$
$q = (q^1_{0.}, q^1_{1.}, q^2_{00}, q^2_{10}, q^2_{01}, q^2_{11}, q^3_{0.}, q^3_{1.})$
Matrix **A** is shown in Tab. 3.1.

Figure 3.2 $pa(X_1) = pa(X_3) = \{X_2\}, pa(X_2) = \emptyset,\ d = 10$
$q = (q^1_{00}, q^1_{10}, q^1_{01}, q^1_{11}, q^2_{0.}, q^2_{1.},\ q^3_{00}, q^3_{10}, q^3_{01}, q^3_{11})$
The matrix of the transformation is depicted in Table 3.2.

Figure 3.3 $pa(X_1) = pa(X_3) = \emptyset, pa(X_2) = \{X_1, X_3\}, d = 12$
$q = (q^1_{0.}, q^1_{1.}, q^2_{0(00)}, q^2_{0(01)}, q^2_{1(10)}, q^2_{0(11)}, q^2_{1(00)}, q^2_{1(01)}, q^2_{1(10)}, q^2_{1(11)}, q^3_{0.}, q^3_{1.})$
Matrix A is shown in Tab. 3.3.

Figure 3.4 $pa(X_1) = \emptyset, pa(X_2) = \{X_1\}, pa(X_3) = \{X_2\}, d = 10$
$q = (q^0_{1.}, q^1_{1.}, q^2_{00}, q^2_{01}, q^2_{10}, q^2_{11}, q^3_{00}, q^3_{01}, q^3_{10}, q^3_{11})$
See Table 3.4 for exact description.

Figure 3.5 Subsequently, let us consider the marginal independence model, where the parent set is empty for every node. That is to say, $pa(X_1) = pa(X_2) = pa(X_3) = \emptyset$:

$$p_{ijk} = q^1_{i.} q^2_{j.} q^3_{k.}$$

Matrix **A** takes the form shown in Tab. 3.5.

Figure 3.6 $pa(X_1) = \emptyset, pa(X_2) = \{X_1\}, pa(X_3) = \{X_1, X_2\}, d = 14$
$t = (t_1, \ldots, t_{14}) = (q^1_{0.}, q^1_{1.}, q^2_{00}, q^2_{10}, q^2_{01}, q^2_{11}, q^3_{0(00)}, q^3_{0(01)}, q^3_{0(10)}, q^3_{0(11)}, q^3_{1(00)}, q^3_{1(01)}, q^3_{1(10)}, q^3_{1(11)})$
In this case, matrix **A** is depicted in Tab. 3.6

As a final case, we shift to the fully connected graph. There are no non-trivial conditional independence statements and the model is not identifiable.

3.1.2 Conditional independence statements

There are six isomorphic cases of the DAGs on three nodes (up to an order on the nodes). We consider all five non-trivial and a complete graph examples (see Figures 3.2– 3.6).
Note that networks 3.2 and 3.4 belong to the same equivalence class (see Section 1.2.0.2).
We will give explicit formulas for I_{local}, which for such small networks coincides with I_{global}. Recall that for binary random variables the generators of the independence ideal, I_{CI} (2.1.4), consist of certain 2×2 minors of the corresponding joint probability matrices (2.1.3).

Note: in SINGULAR binary variables are coded as $\{1, 2\}$, which correspond to our notation 0/1 notation. In addition, we label he model parameters by the vectorized indices $t = (t_1, t_2, \ldots, t_d)$.

1 The independence ideal for model M_1 is generated by the quadrics imposed by the following conditional independence restrictions:
$M_1 = \{X_2 \perp X_3 \mid X_1, X_1 \perp X_3, X_{1,2} \perp X_3\}$. The latter statement contains the two preceding ones, thus we have that $M_1 = \{X_{1,2} \perp X_3\}$. Therefore, the components of I_{M1} generated by two binomials derived from the rank restrictions of the matrices:

$$\begin{pmatrix} p_{.00} & p_{.01} \\ p_{.10} & p_{.11} \end{pmatrix}$$

where e.g. $p_{.00} \in \{p_{000}, p_{100}\}, I_{M_1^1} =< p_{000}p_{011} - p_{001}p_{010}, p_{100}p_{111} - p_{101}p_{110} >$ are contained in the ideal $I_{M_1^2}$ generated by the quadrics of

$$\begin{pmatrix} p_{0+0} & p_{0+1} \\ p_{1+0} & p_{1+1}, \end{pmatrix}$$

where e.g. $p_{0+0} = \sum_{i=0}^{1} p_{0i0}$.
Namely,

$$I_{M_1} = I_{M_1^1} + I_{M_1^2} =< (p_{000}+p_{010})(p_{101}+p_{111}) - (p_{100}+p_{110})(p_{001}+p_{011}) > .$$

Singular code for the direct parametrization through a polynomial map f_1 as defined by the toric model $\mathbf{A}$ =Tab 3.1:

```
ring p=0,(p(1..2)(1..2)(1..2)),dp;
ring t1=0,(t(1..8)),dp;
ideal t0=0;
map f1=p,t(1)*t(3)*t(7),t(1)*t(3)*t(8),t(1)*t(4)*t(7),
t(1)*t(4)*t(8),t(2)*t(5)*t(7),t(2)*t(5)*t(8),
t(2)*t(6)*t(7),t(2)*t(6)*t(8);
setring p;
ideal i1=preimage(t1,f1,t0);
```

leads to the following kernel of ring homomorphism Ψ_{f_1}:

```
_[1]=p(2)(1)(2)*p(2)(2)(1)-p(2)(1)(1)*p(2)(2)(2)
_[2]=p(1)(2)(2)*p(2)(2)(1)-p(1)(2)(1)*p(2)(2)(2)
_[3]=p(1)(1)(2)*p(2)(2)(1)-p(1)(1)(1)*p(2)(2)(2)
_[4]=p(1)(2)(2)*p(2)(1)(1)-p(1)(2)(1)*p(2)(1)(2)
_[5]=p(1)(1)(2)*p(2)(1)(1)-p(1)(1)(1)*p(2)(1)(2)
_[6]=p(1)(1)(2)*p(1)(2)(1)-p(1)(1)(1)*p(1)(2)(2)
```

We next confirm that indeed the inclusion of (2.1.8) holds:

```
matrix m1[2][2]=p(1)(1..2)(1..2);
matrix m2[2][2]=p(2)(1..2)(1..2);
matrix m11[2][2]=p(1..2)(1)(1..2);
matrix m21[2][2]=p(1..2)(2)(1..2);
ideal IM1=minor(m1,2),minor(m2,2),minor(m11+m21,2);
```

```
I[1]=-p(1)(1)(2)*p(1)(2)(1)+p(1)(1)(1)*p(1)(2)(2)
I[2]=-p(2)(1)(2)*p(2)(2)(1)+p(2)(1)(1)*p(2)(2)(2)
I[3]=-p(1)(1)(2)*p(2)(1)(1)-p(1)(2)(2)*p(2)(1)(1)
    + p(1)(1)(1)*p(2)(1)(2)+p(1)(2)(1)*p(2)(1)(2)
    - p(1)(1)(2)*p(2)(2)(1)-p(1)(2)(2)*p(2)(2)(1)
    + p(1)(1)(1)*p(2)(2)(2)+p(1)(2)(1)*p(2)(2)(2)

reduce(IM1,std(i1));
_[1]=0
_[2]=0
_[3]=0
```

Namely, $I_{M_1} \subseteq \ker(\Psi_{f_1})$.

2 The binomial independence ideal $I_{M_2} = I_{X_1 \perp X_3 | X_2}$ corresponds to all the 2×2 minors of the following matrices:
where e.g. $p_{0.0} \in \{p_{000}, p_{010}\}$. We get:

$$I_{M_2} =< p_{000}p_{101} - p_{001}p_{100}, p_{010}p_{111} - p_{011}p_{110} >$$

```
setring p;
matrix m21[2][2]=p(1..2)(1)(1..2);
matrix m22[2][2]=p(1..2)(2)(1..2);
ideal IM2=minor(m21,2),minor(m22,2);

ring t2=0,(t(1..10)),dp;
ideal t0=0;
map f2 = p,t(1)*t(5)*t(7),t(1)*t(5)*t(9),t(2)*t(6)*t(8),
t(2)*t(6)*t(10),t(3)*t(5)*t(7),t(3)*t(5)*t(9),
t(4)*t(6)*t(8),t(4)*t(6)*t(10);

setring p;
ideal i2=preimage(t2,f2,t0);
```

```
i2:=
_[1]=p(1)(2)(2)*p(2)(2)(1)-p(1)(2)(1)*p(2)(2)(2)
_[2]=p(1)(1)(2)*p(2)(1)(1)-p(1)(1)(1)*p(2)(1)(2)
```

As before, we check (2.1.8):

```
reduce(IM2,std(i2));
_[1]=0
_[2]=0
```

3 The marginal independence statement of **A** =Tab 3.3 in Figure 3.3, $X_1 \perp X_2$, equals the second term in I_{M_1}. That is, I_{M_3} is the ideal generated by the quadrics:

$$I_{M_3} =< (p_{000}+p_{010})(p_{101}+p_{111}) - (p_{100}+p_{110})(p_{001}+p_{011}) > .$$

```
setring p;
matrix m21[2][2]=p(1..2)(1)(1..2);
matrix m22[2][2]=p(1..2)(2)(1..2);
ideal IM3=minor(m21+m22,2);

ring t3=0,(t(1..12)),dp;
ideal t0=0;
map f3=p,t(1)*t(3)*t(11),t(1)*t(4)*t(12),t(1)*t(7)*t(11),
t(1)*t(8)*t(12),t(2)*t(5)*t(11),t(2)*t(6)*t(12),
t(2)*t(9)*t(11),t(2)*t(10)*t(12);
```

By restricting to: $t_1+t_2 = t_3+t_7 = t_4+t_8 = t_5+t_9 = t_6+t_{10} = t_{11}+t_{12}$, the preimage equals:

```
ideal k3=t(1)+t(2)-t(3)-t(7),t(3)+t(7)-t(4)-t(8),
t(4)+t(8)-t(5)-t(9),t(5)+t(9)-t(6)-t(10),
t(6)+t(10)-t(11)-t(12);
```

```
setring p;
ideal i3=preimage(t3,f3,k3);


i3[1]=
     p(1)(1)(2)*p(2)(1)(1)+p(1)(2)(2)*p(2)(1)(1)
    -p(1)(1)(1)*p(2)(1)(2)-p(1)(2)(1)*p(2)(1)(2)
    +p(1)(1)(2)*p(2)(2)(1)+p(1)(2)(2)*p(2)(2)(1)
    -p(1)(1)(1)*p(2)(2)(2)-p(1)(2)(1)*p(2)(2)(2);
```

and the (2.1.8) holds:

```
reduce(IM3,std(i3));
_[1]=0
```

4 For CI model of **A** =Tab 3.4 depicted in Figure 3.4, it holds that I_{M_4} and I_{M_2}. What follows, I_{M_4} is generated by the two binomials:

```
setring p;
matrix m21[2][2]=p(1..2)(1)(1..2);
matrix m22[2][2]=p(1..2)(2)(1..2);
ideal IM4=minor(m21,2),minor(m22,2);
IM4[1]=p(1)(2)(2)*p(2)(2)(1)-p(1)(2)(1)*p(2)(2)(2)
IM4[2]=p(1)(1)(2)*p(2)(1)(1)-p(1)(1)(1)*p(2)(1)(2)
```

Computing the kernel of Ψ_{f_4}:

```
ring t4=0,(t(1..10)),dp;
ideal t0=0;
map f4 = p,t(1)*t(3)*t(7),t(1)*t(3)*t(9),t(1)*t(4)*t(8),
t(1)*t(4)*t(10),t(2)*t(5)*t(7),t(2)*t(5)*t(9),
t(2)*t(6)*t(8),t(2)*t(6)*t(10);
setring p;
ideal i4=preimage(t4,f4,t0);
i4[1]=p(1)(2)(2)*p(2)(2)(1)-p(1)(2)(1)*p(2)(2)(2)
i4[2]=p(1)(1)(2)*p(2)(1)(1)-p(1)(1)(1)*p(2)(1)(2)
```

```
reduce(IM4,std(i4));
_[1]=0
_[2]=0
```

5 M_5 of marginal independence of the three random variables is described by $\mathbf{A}$ =Tab 3.5 and shown in Figure 3.5. Trivial SINGULAR calculations are given by:

```
setring p;
matrix m51[2][2]=p(1)(1..2)(1..2);
matrix m52[2][2]=p(2)(1..2)(1..2);
matrix m53[2][2]=p(1..2)(1)(1..2);
matrix m54[2][2]=p(1..2)(2)(1..2);
matrix m55[2][2]=p(1..2)(1..2)(1);
matrix m56[2][2]=p(1..2)(1..2)(2);
ideal IM5=minor(m51+m52,2),minor(m53+m54,2),minor(m55+m56,2);

ring t5=0,(t(1..6)),dp;
ideal t0=0;
map f5=p,t(1)*t(3)*t(5),t(1)*t(3)*t(6),t(1)*t(4)*t(5),
t(1)*t(4)*t(6),t(2)*t(3)*t(5),t(2)*t(3)*t(6),
t(2)*t(4)*t(5),t(2)*t(4)*t(6);
setring p;
ideal i5 = preimage(t5,f5,t0);
_[1]=p(2)(1)(2)*p(2)(2)(1)-p(2)(1)(1)*p(2)(2)(2)
_[2]=p(1)(2)(2)*p(2)(2)(1)-p(1)(2)(1)*p(2)(2)(2)
_[3]=p(1)(1)(2)*p(2)(2)(1)-p(1)(1)(1)*p(2)(2)(2)
_[4]=p(1)(2)(2)*p(2)(1)(2)-p(1)(1)(2)*p(2)(2)(2)
_[5]=p(1)(2)(1)*p(2)(1)(2)-p(1)(1)(1)*p(2)(2)(2)
_[6]=p(1)(2)(2)*p(2)(1)(1)-p(1)(1)(1)*p(2)(2)(2)
_[7]=p(1)(2)(1)*p(2)(1)(1)-p(1)(1)(1)*p(2)(2)(1)
_[8]=p(1)(1)(2)*p(2)(1)(1)-p(1)(1)(1)*p(2)(1)(2)
_[9]=p(1)(1)(2)*p(1)(2)(1)-p(1)(1)(1)*p(1)(2)(2)
```

and a confirming answer to (2.1.8).

```
reduce(IM5,std(i5));
_[1]=0
_[2]=0
_[3]=0
```

6 The final case of the complete graph of $\mathbf{A} = 3.6$ (see Figure 3.6) generates only trivial local independence statements:

$$M_6 = \{X_1 \perp \{X_2, X_3\}, X_2 \perp \{X_1, X_3\}, X_3 \perp \{X_1, X_2\}\}$$

Thus, $I_{M_6} = \emptyset$.

SINGULAR code for describing $\ker(\Psi)$ in this case is given below:

```
ring t6=0,(t(1..14)),dp;
ideal t0=0;
map f6=p,t(1)*t(3)*t(7),t(1)*t(3)*t(11),t(1)*t(4)*t(8),
t(1)*t(4)*t(12),t(2)*t(5)*t(9),t(2)*t(5)*t(13),
t(2)*t(7)*t(10),t(2)*t(7)*t(14);
setring p;
preimage(t6,f6,t0);
_[1]=0
```

3.2 BN with four nodes

Let us consider a few cases of the Bayesian Networks on four binary nodes: $X = \{X_1, X_2, X_3, X_4\}$ and $\forall i = 1:4 \quad [d_i] = \{0, 1\}$. For the computational limitations of the current parametric setting, we have to restrain the analyses to the DAGs with at most 12 parameters (see (2.1.1)). Henceforth, we consider 4 distinct equivalent classes (up to a permutation) of BN, as depicted in Figures 3.7−3.1. In the last equivalence class we include two networks representing

Table 3.1: M_1

$\boldsymbol{t}$	p_{000}	p_{001}	p_{010}	p_{011}	p_{100}	p_{101}	p_{110}	p_{111}
t_1	1	1	1	1	0	0	0	0
t_2	0	0	0	0	1	1	1	1
t_3	1	1	0	0	0	0	0	0
t_4	0	0	1	1	0	0	0	0
t_5	0	0	0	0	1	1	0	0
t_6	0	0	0	0	0	0	1	1
t_7	1	0	1	0	1	0	1	0
t_8	0	1	0	1	0	1	0	1

Table 3.2: M_2

$\boldsymbol{t}$	p_{000}	p_{001}	p_{010}	p_{011}	p_{100}	p_{101}	p_{110}	p_{111}
t_1	1	1	0	0	0	0	0	0
t_2	0	0	1	1	0	0	0	0
t_3	0	0	0	0	1	1	0	0
t_4	0	0	0	0	0	0	1	1
t_5	1	1	0	0	1	1	0	0
t_6	0	0	1	1	0	0	1	1
t_7	1	0	0	0	1	0	0	0
t_8	0	0	1	0	0	0	1	0
t_9	0	1	0	0	0	1	0	0
t_{10}	0	0	0	1	0	0	0	1

identical conditional independence modelas as in (1.1.1), defined nonetheless via different polynomial maps, (2.1.2).
This is an initial study strongly tied to the future directions of the project. Namely, as a further step, we wish to proved the Conjecture 3.3 (see Section 3.3). For its initial character, we provide model descriptions and SINGULAR commands. For comments and detailed descriptions and definitions, see Section 3.1.

The ring $R[P]$ is generated by 16 indeterminants:
$p = (p_{0000}, p_{0001}, p_{0010}, p_{0011}, p_{0100}, p_{0101}, p_{0110}, p_{0111}, p_{1000}, p_{1001}, p_{1010}, p_{1011}, p_{1100},$

Table 3.3: M_3

$\boldsymbol{t}$	p_{000}	p_{001}	p_{010}	p_{011}	p_{100}	p_{101}	p_{110}	p_{111}
t_1	1	1	1	1	0	0	0	0
t_2	0	0	0	0	1	1	1	1
t_3	1	0	0	0	0	0	0	0
t_4	0	1	0	0	0	0	0	0
t_5	0	0	0	0	1	0	0	0
t_6	0	0	0	0	0	1	0	0
t_7	0	0	1	0	0	0	0	0
t_8	0	0	0	1	0	0	0	0
t_9	0	0	0	0	0	0	1	0
t_{10}	0	0	0	0	0	0	0	1
t_{11}	1	0	1	0	1	0	1	0
t_{12}	0	1	0	1	0	1	0	1

Table 3.4: M_4

$\boldsymbol{t}$	p_{000}	p_{001}	p_{010}	p_{011}	p_{100}	p_{101}	p_{110}	p_{111}
t_1	1	1	1	1	0	0	0	0
t_2	0	0	0	0	1	1	1	1
t_3	1	1	0	0	0	0	0	0
t_4	0	0	1	1	0	0	0	0
t_5	0	0	0	0	1	1	0	0
t_6	0	0	0	0	0	0	1	1
t_7	1	0	0	0	1	0	0	0
t_8	0	0	1	0	0	0	1	0
t_9	0	1	0	0	0	1	0	0
t_{10}	0	0	0	1	0	0	0	1

Table 3.5: M_5 : model of marginal independence for three variable

$\boldsymbol{t}$	p_{000}	p_{001}	p_{010}	p_{011}	p_{100}	p_{101}	p_{110}	p_{111}
t_1	1	1	1	1	0	0	0	0
t_2	0	0	0	0	1	1	1	1
t_3	1	1	0	0	1	1	0	0
t_4	0	0	1	1	0	0	1	1
t_5	1	0	1	0	1	0	1	0
t_6	0	1	0	1	0	1	0	1

$p_{1101}, p_{1110}, p_{1111}$). Again, it factorizes according to the graph G if it lies in the image of the map ψ as defined in (2.1.2).
We use characteristic the 32003 for the base field in order to speed the computations and avoid running out of memory.

Table 3.6: M_6 : complete BN on three nodes

$\boldsymbol{t}$	p_{000}	p_{001}	p_{010}	p_{011}	p_{100}	p_{101}	p_{110}	p_{111}
t_1	1	1	1	1	0	0	0	0
t_2	0	0	0	0	1	1	1	1
t_3	1	1	0	0	0	0	0	0
t_4	0	0	1	1	0	0	0	0
t_5	0	0	0	0	1	1	0	0
t_6	0	0	0	0	0	0	1	1
t_7	1	0	0	0	0	0	0	0
t_8	0	0	1	0	0	0	0	0
t_9	0	0	0	0	1	0	0	0
t_{10}	0	0	0	0	0	0	1	0
t_{11}	0	1	0	0	0	0	0	0
t_{12}	0	0	0	1	0	0	0	0
t_{13}	0	0	0	0	0	1	0	0
t_{14}	0	0	0	0	0	0	0	1

Ad. Figure 3.7 $pa(X_1) = pa(X_2) = pa(X_3) = pa(X_4) = \emptyset$:

$$p_{ijkl} = q^1_{i.} q^2_{j.} q^3_{k.} q^4_{l.}$$

```
ring p4=32003,(p(1..2)(1..2)(1..2)(1..2)),dp;

%%%% X1 independent X2
matrix pom1[2][2]=p(1..2)(1..2)(1)(1);
matrix pom2[2][2]=p(1..2)(1..2)(1)(2);
matrix pom3[2][2]=p(1..2)(1..2)(2)(1);
matrix pom4[2][2]=p(1..2)(1..2)(2)(2);
ideal M11=minor(pom1+pom2+pom3+pom4,2);

%%%% X2 ind X3
matrix pom1[2][2]=p(1)(1..2)(1..2)(1);
matrix pom2[2][2]=p(1)(1..2)(1..2)(2);
matrix pom3[2][2]=p(2)(1..2)(1..2)(1);
matrix pom4[2][2]=p(2)(1..2)(1..2)(2);
ideal M12=minor(pom1+pom2+pom3+pom4,2);
%%%% X3 ind X4
matrix pom1[2][2]=p(1)(1)(1..2)(1..2);
matrix pom2[2][2]=p(1)(2)(1..2)(1..2);
matrix pom3[2][2]=p(2)(1)(1..2)(1..2);
matrix pom4[2][2]=p(2)(2)(1..2)(1..2);
ideal M13=minor(pom1+pom2+pom3+pom4,2);

ideal I41=M11,M12,M13;

ring t41=32003,(t(1..8)),dp;
ideal t0=0;
map f41=p4,
t(1)*t(3)*t(5)*t(7),t(1)*t(3)*t(5)*t(8),t(1)*t(3)*t(6)*t(7),
t(1)*t(3)*t(6)*t(8),t(1)*t(4)*t(5)*t(7),t(1)*t(4)*t(5)*t(8),
```

```
t(1)*t(4)*t(6)*t(7),t(1)*t(4)*t(6)*t(8),t(2)*t(3)*t(5)*t(7),
t(2)*t(3)*t(5)*t(8),t(2)*t(3)*t(6)*t(7),t(2)*t(3)*t(6)*t(8),
t(2)*t(4)*t(5)*t(7),t(2)*t(4)*t(5)*t(8),t(2)*t(4)*t(6)*t(7),
t(2)*t(4)*t(6)*t(8);
setring p4;
ideal k41=preimage(t41,f41,t0);

reduce(I41,std(k41));
_[1]=0
_[2]=0
_[3]=0
```

Ad. Figure 3.1 $M_2^4 = \{(X_1, X_2) \perp X_3 \perp X_4\}$
$pa(X_1) = P(X_3) = P(X_4) = \emptyset, pa(X_2) = X_1, d = 10,$
$q = (q^1_{0.}, q^1_{1.}, q^2_{00}, q^2_{01}, q^2_{10}, q^2_{11}, q^3_{0.}, q^3_{1.}, q^4_{0.}, q^4_{1.})$

```
%%% (X1,X2) ind X3

matrix pom1[2][4]=p(1..2)(1..2)(1)(1),p(1..2)(1..2)(2)(1);
matrix pom2[2][4]=p(1..2)(1..2)(1)(2),p(1..2)(1..2)(2)(2);

ideal M21=minor(pom1+pom2,2);

%%%% X3 ind X4

ideal I42=M13,M21;

ring t42=32003,(t(1..10)),dp;
ideal t0=0;
map f42=p4,
t(1)*t(3)*t(7)*t(9),t(1)*t(3)*t(7)*t(10),t(1)*t(3)*t(8)*t(9),
t(1)*t(3)*t(8)*t(10),t(1)*t(4)*t(7)*t(9),t(1)*t(4)*t(7)*t(10),
t(1)*t(4)*t(8)*t(9),t(1)*t(4)*t(8)*t(10),t(2)*t(5)*t(7)*t(9),
```

```
t(2)*t(5)*t(7)*t(10),t(2)*t(5)*t(8)*t(9),t(2)*t(5)*t(8)*t(10),
t(2)*t(6)*t(7)*t(9),t(2)*t(6)*t(7)*t(10),t(2)*t(6)*t(8)*t(9),
t(2)*t(6)*t(8)*t(10);
setring p4;
ideal k42=preimage(t42,f42,t0);

reduce(I42,std(k42));
_[1]=0
_[2]=0
_[3]=0
_[4]=0
_[5]=0
_[6]=0
_[7]=0
```

Ad. Figure 3.1 $M_3^4 = \{(X_1, X_2) \perp (X_3, X_4)\}$
$pa(X_1) = Pa(X_4) = \emptyset, pa(X_2) = \{X_1\}, pa(X_3) = \{X_2\}, d = 12$
$q = (q^1_{0.}, q^1_{1.}, q^2_{00}, q^2_{01}, q^2_{10}, q^2_{11}, q^3_{0.}, q^3_{1.}, q^4_{00}, q^4_{01}, q^4_{10}, q^4_{11})$

```
matrix M43[4][4];
M43[1,1..4]=p(1..2)(1..2)(1)(1);
M43[2,1..4]=p(1..2)(1..2)(2)(1);
M43[3,1..4]=p(1..2)(1..2)(2)(1);
M43[4,1..4]=p(1..2)(1..2)(2)(2);

ideal I43=minor(M43,2);

ring t43=32003,(t(1..12)),dp;
ideal t0=0;
map f43=p4,
t(1)*t(3)*t(7)*t(9),t(1)*t(3)*t(7)*t(10),t(1)*t(3)*t(8)*t(11),
t(1)*t(3)*t(8)*t(12),t(1)*t(4)*t(7)*t(9),t(1)*t(4)*t(7)*t(10),
```

```
t(1)*t(4)*t(8)*t(11),t(1)*t(4)*t(8)*t(12),t(2)*t(5)*t(7)*t(9),
t(2)*t(5)*t(7)*t(10),t(2)*t(5)*t(8)*t(11),t(2)*t(5)*t(8)*t(12),
t(2)*t(6)*t(7)*t(9),t(2)*t(6)*t(7)*t(10),t(2)*t(6)*t(8)*t(11),
t(2)*t(6)*t(8)*t(12);

setring p4;
ideal k43=preimage(t43,f43,t0);

reduce(I43,std(k43));
_[1]=0
...
_[30]=0
```

Fig. 3.1 $M_4^4: \quad pa(X_1) = pa(X_3) = \{X_2\}, pa(X_2) = \emptyset,\ d = 12$
$M_5^4: \quad pa(X_1) = pa(X_3) = \{X_2\}, pa(X_2) = \emptyset,\ d = 12$
$M_4^4 = M_5^4 = \{X_1 \perp X_3 \mid X_2, (X_1, X_2, X_3) \perp X_4\}$

- $M_4^4: \quad q = (q^1_{00}, q^1_{01}, q^1_{01}, q^1_{11}, q^2_{0.}, q^2_{1.}, q^3_{00}, q^3_{01}, q^3_{10}, q^3_{11}, q^4_{0.}, q^4_{1.})$

```
%%% X1 independent X3 given X2

matrix pom1[2][2]=p(1..2)(1)(1..2)(1),p(1..2)(1)(1..2)(1);
matrix pom2[2][2]=p(1..2)(1)(1..2)(2),p(1..2)(1)(1..2)(2);
matrix m441=pom1+pom2;

matrix pom1[2][2]=p(1..2)(2)(1..2)(1),p(1..2)(2)(1..2)(1);
matrix pom2[2][2]=p(1..2)(2)(1..2)(2),p(1..2)(2)(1..2)(2);
matrix M442=pom1+pom2;

ideal M44=minor(M441,2),minor(M442,2);

%%% (X1,X2,X3) independent X4
```

```
matrix M443[2][8]=p(1..2)(1..2)(1)(1),p(1..2)(1..2)(2)(1),
                  p(1..2)(1..2)(1)(2),p(1..2)(1..2)(2)(2);
ideal I44=minor(M443,2),M44;

ring t44 = 32003,(t(1..12)),dp;
ideal t0 = 0;
map f44 = p4,
t(1)*t(5)*t(7)*t(11),t(1)*t(5)*t(7)*t(12),
t(1)*t(5)*t(9)*t(11),t(1)*t(5)*t(9)*t(12),
t(2)*t(6)*t(8)*t(11),t(2)*t(6)*t(8)*t(12),
t(2)*t(6)*t(10)*t(11),t(2)*t(6)*t(10)*t(12),
t(3)*t(5)*t(7)*t(11),t(3)*t(5)*t(7)*t(12),
t(3)*t(5)*t(9)*t(11),t(3)*t(5)*t(9)*t(12),
t(4)*t(6)*t(8)*t(11),t(4)*t(6)*t(8)*t(12),
t(4)*t(6)*t(10)*t(11),t(4)*t(6)*t(10)*t(12);
setring p4;
ideal k44=preimage(t44,f44,t0);

reduce(I44,std(k44));
_[1]=0
...
_[30]=0
```

- $M_5^4: \quad q = (q^1_{0.}, q^1_{1.}, q^2_{00}, q^2_{01}, q^2_{01}, q^2_{11}, q^3_{00}, q^3_{01}, q^3_{10}, q^3_{11}, q^4_{0.}, q^4_{1.})$

```
ring p4=32003,(p(1..2)(1..2)(1..2)(1..2)),dp;
ring t45=32003,(t(1..12)),dp;
ideal t0=0;
map f45=p4,
t(1)*t(3)*t(7)*t(11),t(1)*t(3)*t(7)*t(12),
t(1)*t(3)*t(8)*t(11),t(1)*t(3)*t(8)*t(12),
t(1)*t(4)*t(9)*t(11),t(1)*t(4)*t(9)*t(12),
```

```
t(1)*t(4)*t(10)*t(11),t(1)*t(4)*t(10)*t(12),
t(2)*t(5)*t(7)*t(11),t(2)*t(5)*t(7)*t(12),
t(2)*t(5)*t(8)*t(11),t(2)*t(5)*t(8)*t(12),
t(2)*t(6)*t(9)*t(11),t(2)*t(6)*t(9)*t(12),
t(2)*t(6)*t(10)*t(11),t(2)*t(6)*t(10)*t(12);
setring p4;
ideal k45 = preimage(t45,f45,t0);

reduce(I44,std(k45));
_[1]=0
...
_[30]=0
```

Note that in the examples presented in this section, the inclusion (2.1.8) holds irrespective of the condition (2.1.7). That is to say, Ψ is defined as (2.1.2), which confirms the posed hypothesis (see 3.3).

3.3 Future directions

- Validate the conjecture of [9]: is it true that $I_{global} \equiv$ toric ideal $\ker(\psi)$ only on the probability simplex
- We have seen that $\Psi'(L) \subseteq V(I_{local})$, where

  $$\Psi' : L \subseteq \mathbb{R}^d \mapsto \mathbb{R}^N$$

  $L = \{\sum_j q^i_{jk} = 1, \forall k : (i_1 \dots i_n) \mid_{pa(X_i)} = k, i = 1, \dots, n\}$ and k is a realization of the parental set of a random variable X_i.
  However, in our computations in SINGULAR, in order to make the ideals homogeneous, we have not restricted precisely to these spaces (see 3.1.2), but to the projective version of them. Therefore we conjecture that:

  Conjecture 1. *Let* $H = \{\sum_j q^i_{jk_m} = q^l_{jk_l}, 1 \leq m < l \leq n$, *where* $k_s \in [k] := \left\{(i_1 \dots i_n) \mid_{pa(X_i)}\right\}$ *is the set of realizations of* $pa(X_i)\}$. *Now consider the*

map:

$$\mathbb{P} : H \subseteq \mathbb{P}^{d-1} \mapsto \mathbb{P}^{N-1}$$

$$[q] \mapsto [\Psi(q)]$$

Then

$$\mathbb{P}\Psi(H) \subseteq V(I_{local}) \subseteq \mathbb{P}^{N-1}.$$

In other words, the ring homomorphism $\Psi : R[P] \mapsto R[Q]$ *satisfies* $I_{local} \subseteq \mathbb{J}_H = (\sum_j q^1_{jk_0} - \sum_j q^l_{jk_l},$ *where* $k_0, k_l \in [k], \quad k_0 \neq k_l)$

- Despite a plethora of inference algorithms for Bayesian networks, their common drawbacks concern the limitations on the complexity of the allowed model: number of random variables and/or parameters.

  We will try to use the algebraic statistics approach to improve these algorithms. For that we will investigate the local set of generators of the ideal under study. If possible, we would like to provide a local complete intersection of our varieties at the points that are statistically meaningful. As this cannot be done by considering each particular Bayesian Network, it is interesting to find a recursive procedure in order to establish the local complete intersection of a BN on n nodes (BN_n) with a BN on $(n-1)$ nodes (BN_{n-1}).
  The goal of the inference algorithm of BNs on n nodes is, assuming that one has a distribution p corresponding to a certain BN on n nodes, infer the underlying network structure that corresponds to p. In terms of algebraic statistics this is equivalent to the following statement:
  assuming that $p \in \bigcup_{BN_n} image(\psi'_{BN})$, find the particular network BN^0 such that $p \in image(\psi_{BN^0})$ is defined inside $\bigcup_{BN_n} image(\psi'_{BN})$. The corresponding idea in the algebraic setting boils down to providing the generators (at least locally) of the ideal: $\frac{\ker(\Psi)_{BN^0}}{\bigcap_{BN_n} \ker(\Psi_{BN})}$

Chapter 4

Applications: case study of splicing regulators

4.1 Motivation

Modern science becomes increasingly cross-disciplinary and primary attention is given to developing methodologies of practical value. Our project was founded as a cooperation between the Universitat Politècnica de Catalunya and Centre de Regulació Genòmica (Centre of Genomic Regulation) in Barcelona. It is aimed to develop and ground mathematical methodologies that would prove useful in practical applications. In particular, the scientific focus of the Bioinformatics and Genomics group lies in the discovery and annotation of novel human genes and is one of a cornerstone teams of the ENCODE Project http://genome.ucsc.edu/ENCODE/.

We have performed the data analyses on human splicing factors, which we now hope to extend using algebraic models, as described in the previous sections.

In short, the biological hypothesis posed was that the conservation of sequences across genomes is often an indication of functionality. In particular, the notion of preservation influences the regulators involved in splicing. Detection of evolutionary constrained sequences across multiple genome alignments is therefore a widely used approach to identify functional regions in genomes, and a multitude

of methods have been developed to quantify the degree of sequence constraint.

Motif-based analysis is of great interest in genomics. It is being increasingly appreciated that the genomic sequence is intrinsically polysemic: the same DNA sequence often carries multiple meanings, that is, it is involved in different functions. The nucleotide sequence of the genome, therefore, is shaped by multiple contrasting evolutionary forces acting at different levels. Within protein coding regions, sequences may play a role in control of translation, translational efficiency, transcript stability, etc. (see [5]) and may therefore be subjected to additional selective forces not directly related to protein coding function. Methods have been recently developed to detect non protein coding selective pressure within protein coding regions (see [6]). Sequences involved in the definition of splice sites, and in the regulation of alternative splicing, the so-called exonic splicing enhancers (ESEs) and silencers (ESSs), may in this regard be particularly prevalent within protein coding regions. Evidence exists that these sequences are under specific selective pressure.

4.2 Data

We used reliable multiple nucleotide sequence alignments of coding exon sequences across six vertebrate species to infer the rate of evolutionary change at base pair resolution: five mammalian species, and used chicken, a bird, as a relative outgroup to infer the direction of the sequence changes. Within mammals, we used two primate species (human and macaque), two rodent species (mouse and rat), and an artyodactil (cow) as a relative outgroup. There exits the generally accepted tree topology relating these species.
Up to date the number of identified splicing-related regulatory subsequences comprises 78% of the total set of possible hexamers. Thus, given their ubiquitousness defining a pertinent set of motifs acting in splicing is a nontrivial task. We performed a crude pruning scheme on the set of 666 Hexamer Exonic Splicing Enhancers (ESEs) from [8] and obtained a set of 32 trusted pentamers.

In addition, we compiled a set of 886 "neutral" hexamers, that is of hexamers that to our knowledge have not yet been implicated in splicing regulation, in order to discriminate between the positive and negative sets.

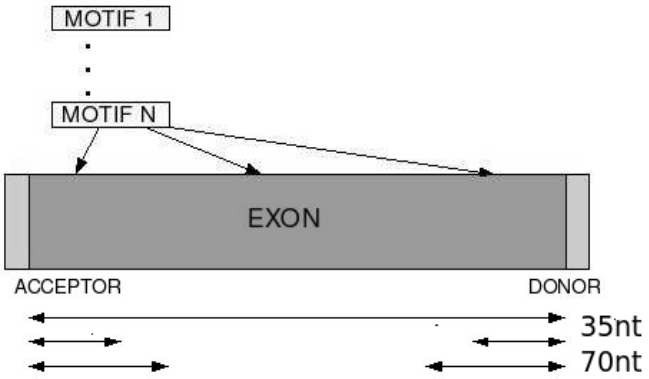

Figure 4.1: A schema representing the data extraction step

We placed the sets of trusted regulatory pentamers (32 sequences), and neutral hexamers (training, 531 sequences, and control, 355 sequences) on the human exon sequences. Only exact matches were considered. We then extracted the columns covered by these sequences from the orthologoues exon alignments. Alignment columns containing gaps were discarded (and the $3nt-$long boundaries on exon edges). For the schema of the mapping procedure see Figure 4.2.

We scored the conservation of the trusted 32 pentamers using two approaches and confirmed the intuitive claims as to their functionality. For instance, being the enhancing factors, we would expect them to support weak splice sites for recognition when splicing occurs. The manuscript containing the results is in its final phase of preparation.

4.3 Further applications

Nevertheless, our scoring procedures did not allow for the study of individual motifs. Also, we performed the analyses across distinct data sets in a pairwise

fashion. Even so, this initial study, relieved more complex cross-effects and implied the network-like structure of occurrences of the enhancers.
At the same time, we wish to avoid overparametrization, which is often the case for sparse biological data.
As the next step of the project, we wish to improve our results by employing the tools from algebraic statistics. We will consider the set of trusted ESEs and investigate their presence or absence in the human exons (and across many taxa). Thus created Bayesian network will have in principle 32 nodes with binary outcomes $\{E_1, \ldots, E_{32}\}$, where $E_i \in \{0 = \text{present}, 1 = \text{absent}\}$. We wish to investigate whether the occurrences of particular splicing factors is elevated in certain genomic regions (positioned in the vicinity of the exon junctions), predominant in certain type of data or co-regulated, that is accompanied by other functional motifs.

Splicing has been shown to be driven by a complex machinery, thus the need for precise and accurate inference methods. An alternative analysis will be performed with the use of transcription factor binding sites motifs.

Bibliography

[1] E. S. Allman and J. A. Rhodes. *Mathematical Models in Biology*. Cambridge University Press,, 2004.

[2] E. S. Allman and J. A. Rhodes. Molecular phylogenetics from an algebraic viewpoint. *Statistica Sinica*, pages 1299–1316, 2007.

[3] E. S. Allman and J. A. Rhodes. *Reconstructing Evolution: New Mathematical and Computational Advances*, chapter Phylogenetic Invariants. Oxford University Press, New York, 2007.

[4] M. Casanellas and J. Fernandez-Sanchez. Geometry of the kimura 3-parameter model. *arXiv:math/0702834*, 2007.

[5] J. V. Chamary, J. L. Parmley, and L. D. Hurst. Hearing silence: no royle, g. f. wanless, i. m. oggier, f. e. sloane, n. j. a. and wilf, h.n-neutral evolution at synonymous sites in mammals. *Nature Reviews Genetics*, 7:98–108, 2006.

[6] H. Chen and M. Blanchette. Detecting non-coding selective pressure in coding regions. *BMC Evol Biol*, 7 Suppl 1:S9, 2007.

[7] L. J. Cox, D. and D. O'Shea. *Ideals, Varieties, and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra*. Springer, 3rd edition, 2007.

[8] W. Fairbrother, R. Yeh, P. Sharp, and C. Burge. Predictive identification of exonic splicing enhancers in human genes. *Science*, 297:1007–1013, 2002.

[9] S. M. Garcia, L.D and B. Sturmfels. Algebraic geometry of bayesian networks. *J.of Symbolic Computation*, 2004.

[10] M. C. Geiger, D. and B. Sturmfels. On the toric algebra of graphical models. *The Annals of Statistics*, 2006.

[11] P. G. Greuel, G.M. and H. Schönemann. SINGULAR 3.0 - a computer algebra system for polynomial computations. In *Symbolic Computation and Automated Reasoning*.

[12] S. Lauritzen. *Graphical Models*. Oxford University Press, New York, 1996.

[13] B. D. McKay. Acyclic digraphs and eigenvalues of (0,1)-matrices. *J. Integer Sequences*, 2004.

[14] L. Pachter and B. Sturmfels, editors. *Algebraic Statistics for Computational Biology*. Cambridge University Press, New York, 2005,.

[15] J. Pearl. *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann Publishers Inc., San Francisco, 1988.

[16] R. E. Pistone, G. and H. Wynn. *Algebraic Statistics*. 2001. Monographs on Statistics and Applied Probability.

[17] G. C. Spirtes, P. and R. Scheines. Causation, prediction, search.

[18] B. Sturmfels. *Grobner Bases and Convex Polytopes*. University Lecture Series, 1997.

[19] B. Sturmfels. Solving systems of polynomial equations. May 2002.

[20] J. Whittaker. *Graphical Models in Applied Multivariate Statistics*. John Wiley and Sons, Chichester, 1978.

CENTRE DE RECERCA MATEMÀTICA

Campus de Bellaterra, Edifici C
08193 Bellaterra, Spain
Tel.: +34 93 581 10 81
Fax: +34 93 581 22 02
crm@crm.cat
www.crm.cat