3.1 Markov Chain example

We will continue with the example of the Markov Chain described in our last lecture. Recall that we had an integer matrix $A \in \mathbb{Z}^{d \times m}$ and we indexed its columns by the probabilities $p_{ijkl} := \theta_{ij} \cdot \theta_{jk} \cdot \theta_{kl}$ of our Markov Chain model. Each row corresponded to a parameter $\theta_{ij}$.

For example, $p_{0000} = \theta_{00}^3$, $p_{0001} = \theta_{00}^2 \theta_{01}$. Our matrix $A$ is:

\[ \begin{pmatrix}
\theta_{00} & \theta_{01} & \theta_{10} & \theta_{11} \\
0 & 1 & 1 & 2 & 1 & 1 & 0 & 1 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 1 & 2 & 0 & 0 & 0 & 1 & 1 & 1 & 2 & 3
\end{pmatrix} \]

As we remarked during last lecture, some columns of matrix $A$ are equal, for example $p_{0010} = p_{0100} = p_{1001}$, or equivalently, $p_{0010} - p_{0000} = 0$, $p_{0010} - p_{1001} = 0$. However, this type of relation is not the only one that holds. In the following, we describe a minimal set of relations that characterize the columns of $A$. We will describe three types of relations: linear ones and two different types of relations of higher order.

- Linear relations

\[ p_{0010} = p_{0100} = p_{1001} ; \quad p_{0110} = p_{1011} = p_{1101} . \]

- Four quadratic relations:

\[ p_{0011}^2 = p_{0001} p_{0111} ; \quad p_{1100}^2 = p_{1000} p_{1110} ; \quad p_{1001}^2 = p_{0001} p_{1010} ; \quad p_{1101}^2 = p_{1010} p_{1111} . \] (\*)

- Nine other relations, which are meaningful:

\[ p_{0111} p_{1010} = p_{0101} p_{1110} ; \quad p_{0111} p_{1000} = p_{0001} p_{1110} ; \quad p_{0101} p_{1000} = p_{0001} p_{1100} ; \]
\[ p_{0111} p_{1010}^2 = p_{0101} p_{1111}^2 ; \quad p_{0111} p_{1110} = p_{0101} p_{1111}^2 ; \quad p_{0001} p_{1000}^2 = p_{0000} p_{1010} ; \]
\[ p_{0000}^2 p_{0101} = p_{0001}^2 p_{1000} ; \quad p_{0000}^3 p_{0110} = p_{1000}^3 p_{1111} ; \quad p_{0000}^3 p_{0111} = p_{0001}^2 p_{1111} . \]

Note that the relations come from the independence model. On the other hand, the number of relations comes from $9 = 2 \cdot 2 + 2 + 3 \neq 3 \cdot 3$, which we will explain later. We will show that these relations can be understood geometrically.
3.2 Geometry

In the present section we will introduce some geometric concepts. Geometry will be an important tool for Computational Biology, in particular for the problem of alignments: geometry will show us how to proceed.

**Definition 3.1.** Given a set of points $\Gamma := \{P_1, \ldots, P_m\} \subseteq \mathbb{R}^d$, we define the convex hull of the set $\Gamma$ as the set of points:

$$\text{ConvHull}(\Gamma) := \left\{ \sum_{i=1}^{m} \lambda_i P_i \mid P_i \in \Gamma, \lambda_i \geq 0, \sum_{i=1}^{r} \lambda_i = 1 \right\},$$

i.e. the set of convex linear combinations of the points in $\Gamma$.

Note that in our case, $\Gamma := \{p_{ijkl} \mid (ijkl) \in \{0,1\}^4\}$ is the set of indices of the columns of $A$.

The advantage we have in our example is that, although the points lie in $\mathbb{R}^4$, since the sum of the columns of $A$ is constant (i.e. the points in $\Gamma$), the points of $\text{ConvHull}(\Gamma)$ lie on a hypersurface, namely $x_1 + x_2 + x_3 + x_4 = 3$ in $\mathbb{R}^4$, so they describe a 3-dimensional variety. This set is in 1-1 correspondence with a subset of $\mathbb{R}^3$. This can be achieved via the projection map $\pi : \mathbb{R}^4 \rightarrow \mathbb{R}^3$ sending a point to its first three coordinates. Note that $\pi$ is injective over the set $\text{ConvHull}(\Gamma)$.

So using this projection map, we are able to draw the convex hull of the points $p_{ijkl}$, which is what we call a polytope. Figure 3.2 describes the polytope associated to the Markov

![Figure 3.1. Polytope of the Markov chain model for $n = 4$](image-url)
chain model. Note that it has three pairs of parallel lines, which are indicated in our figure. It consists of eight vertices, two triangles, two pentagons and two rectangles.

From the figure, we see that the eight vertices are labeled by \((ijkl)\) (since each one corresponds to the point \(p_{ijkl}\)). Note that we had 12 distinct points \(p_{ijk}\) (because we had four repeated columns) but our polytope has 8 vertices. The remaining four points lie inside the polytope. In fact, they lie on four edges of the polytope.

Let us show where the missing points are. They correspond to the points \(p_{0011}, p_{1100}, p_{1001}\) and \(p_{1101}\). The answer to this is given by the four quadratic relations \((*)\). For example, we will see where we can find \(p_{0011}\) in our polytope. We know that

\[
p_{0011}^2 = p_{0001} \cdot p_{0111},
\]

which comes from the expression \((\theta_{00}\theta_{01}\theta_{11})^2 = (\theta_{00}^2\theta_{01})(\theta_{01}\theta_{11}^2)\). If we look at the corresponding exponent vectors, this says that

\[
p_{0011}^2 \downarrow \downarrow \downarrow 2 \cdot (1,1,0,1) = (2,1,0,0) + (0,1,0,2)
\]

or equivalently,

\[
(1,1,0,1) = \frac{1}{2} \cdot ((2,1,0,0) + (0,1,0,2))
\]

i.e. the point \(p_{0011}\) is the midpoint of the segment between \(p_{0001}\) and \(p_{0111}\). This corresponds to the fact that the equation \(p_{0011}^2 = p_{0001} \cdot p_{0111}\) corresponds to the real relation \(p_{0011} = \sqrt{p_{0001} \cdot p_{0111}}\), since all probabilities are positive real numbers. Figure 3.2 shows the “missing points” in red.

In this way, we see that the geometry determines the relations among the points \(p_{ijkl}\). In fact, a minimal set of relations will be the ones we need to describe the polytope. For this, we need only to be able to describe the boundary of the polytope by means of relations among the vertices, and to give relations for describing the “missing points”: the four repeated columns plus the four points in the interior of the polytope. This gives the first four linear relations plus the four quadratic relations \((*)\). The relations describing the boundary of the polytope will give the remaining nine equations.

In our case, the boundary of the polytope consists of two pentagons, two triangles and two rectangles. To describe each rectangle we need to say that the corresponding four vertices lie in a plane: this gives one relation per rectangle. For each pentagon, we can choose two relations describing it. We simply pick two collections of four of its vertices and we give a linear relation saying that the four points lie in the same plane. For the triangle there is nothing to say since three points always determine a triangle unless they are collinear, which is not the case. To finish, we need to add three more relations, which say that we have three pairs of parallel edges. Therefore, we need \(2 \cdot 2 + 2 \cdot 1 + 3 = 9\) relations.

**Remark.** In fact, what we wrote down is a minimal set of relations, i.e. a minimal set (with respect to inclusion) of relations that describes completely our polytope and the sixteen
points $p_{ijkl}$ of our model. This *minimal* set will be closely related to what will be a *reduced* Gröbner basis.

### 3.3 Meaning of the polytope

In the following we will describe the inference meaning of the polytope, i.e. what it is useful for.

Suppose we fix our parameters $\theta_{00}, \theta_{01}, \theta_{10}, \theta_{11}$. A basic question we could ask is the following. What is the sequence (Markov Chain) with highest probability? One solution for this would be to compute all of the $p_{ijkl}$ and then choose the biggest number(s) among them. But we want to derive a faster way of doing this.

This example is the simplest case of what it is called maximum “a posteriori” inference. It is closely related to the polytope construction.

Since we are working with the polytope, i.e. convex linear combinations of points, it is natural to consider $\log(p_{ijkl})$ instead of $p_{ijkl}$. In this case we will have new equations:

$$\log(p_{ijkl}) = \log(\theta_{ij} \theta_{jk} \theta_{kl}) = \log(\theta_{ij}) + \log(\theta_{jk}) + \log(\theta_{kl}) .$$

For example:

$$\log(p_{0000}) = \log(\theta_{00}^3) = 3 \cdot \log(\theta_{00}) := 3s_{00} ;$$
$$\log(p_{0001}) = \log(\theta_{00}^2 \theta_{01}) = 2 \cdot \log(\theta_{00}) + \log(\theta_{01}) := 2 \cdot s_{00} + s_{01} ;$$
$$\log(p_{1111}) = \log(\theta_{11}^3) = 3 \cdot \log(\theta_{11}) := 3s_{11} ;$$
where we’ve introduced a change of coordinates, namely $s_{ij} := \log(\theta_{ij})$.

From this discussion we can derive an equivalent question to the highest probability one. Namely, given the vectors $(u_{0000}, \ldots, u_{ijkl}, \ldots, u_{1111})$, find one that maximizes $U \cdot S^t$, where

$$U = \begin{pmatrix} u_{0000} \\ \vdots \\ u_{1111} \end{pmatrix} \in \mathbb{R}^{m \times 4} \text{ and } S = (s_{00}, s_{01}, s_{10}, s_{11}) \in \mathbb{R}^4.$$

For example, $\log(p_{0000})$ corresponds to $(3, 0, 0, 0) \cdot S = 3s_{00}$. In general, each coordinate $(ijkl)$ of $U \cdot S^t$ will correspond to $\log(p_{ijkl})$.

The key fact in this problem is that the maximal points of $U \cdot S$ will define a face of our polytope. Moreover, for generic choices, the answer will be a vertex of the polytope. The genericity conditions will be given by avoiding the relations that describes our model. For example, if we allow our parameters to satisfy $p_{0000} = p_{1111}$, i.e. $\theta_{00}^3 = \theta_{11}^3$ (or equivalently $3s_{00} = 3s_{11}$) then the solution will lie on an edge of the polytope.

### 3.4 Back to polytopes

We will now give some few basic terminology concerning the theory of polytopes.

**Definition 3.2.** Given $P = \text{ConvHull}(\Gamma)$ a polytope described as the convex hull of a finite set of points $\Gamma \subset \mathbb{R}^d$ we define the dimension of $P$ as $\dim P = \dim(\Gamma)$, i.e. the dimension of the affine space spanned by the set $\Gamma$.

**Definition 3.3.** Given a polytope $P$ and a vector $w \in \mathbb{R}^d$ of $S$ ($w$: weight, $S$: score), the set of points where the linear function $f_w : P \rightarrow \mathbb{R}$, given by $x \mapsto w \cdot x$, attains its minimum is called a face of $P$. Namely,

$$\text{face}_w(P) = \{ x \in P : x \cdot w \leq y \cdot w \ \forall \ y \in P \}.$$

Note that, alternatively, we could define a face as the set of points where the same function attains its maximum. This choice would be in agreement with our question about highest probabilities. However, to agree with the textbook reference of our course we will use the previous definition of a face by means of minimizing the functional $x \cdot w$. Moreover, note that by choosing $-w$ instead of $w$ we can answer the maximum probability problem.

**Definition 3.4.** There are distinguished names for some faces, according to their dimensions:

- a vertex is a face of dimension 0;

---

1 to be defined in the next section
2 “Algebraic Statistics for Computational Biology”
Figure 3.3. The unitary square in $\mathbb{R}^2$ as a polyhedron

- an edge is a face of dimension 1;
- a ridge is a face of dimension $\dim P - 2$;
- a facet is a face of dimension $\dim P - 1$.

For example, in the case of our polytope (Figure 3.2), an edge is also a ridge, since $\dim P = 3$.

Definition 3.5. The f-vector of a polytope $P$ is a vector $(f_0, f_1, \ldots)$ summarizing the number of faces of each dimension in the polytope.

Example. For $P$ our model polytope (Figure 3.2) we have $f$-vector($P$) = $(8, 12, 6)$, since $P$ has 8 vertices, 12 edges and 6 facets.

We now make an important remark. There are two ways of describing a polytope: as the convex hull of a finite set of points or as the intersection of half-spaces. Each half-space will correspond to a facet of the polytope. Figure 3.3 shows the case of a square in $\mathbb{R}^2$ as the intersection of four half-spaces. The grey-scale describes each partial intersections of two, three or four half-spaces.

A polytope will be a special kind of polyhedron.

Definition 3.6. Given a matrix $A \in \mathbb{R}^{m \times d}$ and a vector $b \in \mathbb{R}^d$ we define a polyhedron $P$ as the set of points

$$P = \{ x \in \mathbb{R}^d : Ax^t \geq b \}.$$ 

This expression of $P$ by inequalities will be called the “H-representation” of a polytope, by contrast with the “V-representation” by means of ConvHull($\Gamma$).

Remark. Note that from the definition of a polyhedron we see that $P$ might be unbounded. For example, this will be the case if we pick one half-space in $\mathbb{R}^d$.

The following result characterizes the H-representation of a polytope:

Theorem 3.7. (Minkowski-Weyl Theorem)

The bounded polyhedra are precisely the polytopes (i.e. V-representation).
Remark. An interesting question that arises is how to go back and forth between the H-representation and the V-representation. During this course we will try to give an answer to this issue.

Remark. If we let $w^* = \min \{ x \cdot w : w \in P \}$, then face$_w(P) = \{ x \in P : x \omega \leq w^* \}$. Since $P$ is bounded, face$_w(P)$ is also bounded. Therefore by Theorem 3.7, the faces of a polytope are polytopes as well.

3.5 Equations for describing a statistical model

In this section, we intend to give an answer to the following question: how do we find the polynomial relations that describe a model? The punchline for this will be the use of Gröbner bases. But before this we need to say what is a Gröbner basis. For the moment, we will say only that Groebner bases are generalizations of three different algorithms:

- Gaussian elimination,
- Euclidean algorithm,
- Integer programming.

Before giving the precise definition, we need to introduce some algebraic concepts that will allow us to explain in an algebraic way our construction of the statistical model $F$.

In the first place we need two polynomial rings (over the rational numbers):

$$R := \mathbb{Q}[p_{ijkl} : ijk] = \mathbb{Q}[p_{0000}, \ldots, p_{1111}],$$

$$S := \mathbb{Q}[a_{00}, a_{01}, a_{10}, a_{11}].$$

(Note that we’ve replaced the parameters $\theta_{ij}$ by the indeterminates $a_{ij}$. There are various reasons for doing this. The first one is that the notation $a_{ij}$ is more common in the literature. The second one is that the polynomial rings will be often represented in computer programs, and in this context it’s better to use $a$’s instead of $\theta$’s.)

What are polynomial rings? For example, the elements of $R$ are formal sums of expressions of the form $b_{\underline{n}} \cdot p_{0000}^{n_{0000}} \cdot p_{0001}^{n_{0001}} \cdot \ldots \cdot p_{1111}^{n_{1111}}$, where $\underline{n} = (n_{0000}, \ldots, n_{1111}) \in \mathbb{N}_0^m$ and $b_{\underline{n}} \in \mathbb{Q}$. We can add and multiply polynomials in the natural way and these two operations have nice properties. These properties make $S$ and $R$ into rings.

Recall from last lecture that the statistical model corresponds to a function $F : \Theta \subseteq \mathbb{R}^d \to P \subseteq \mathbb{R}^m$. How do we interpret this function by means of the rings $S$ and $R$? The key fact is that $F$ gives a map of rings in the opposite direction, namely

$$f : R \to S \quad p_{ijkl} \mapsto a_{ij} a_{jk} a_{kl}.$$ 

Namely, given a polynomial $\varphi(p_{0000}, \ldots, p_{1111}) \in R$, we have $f(\varphi(p_{0000}, \ldots, p_{ijkl}, \ldots, p_{1111})) = \varphi(a_{00} a_{00} a_{00}, \ldots, a_{ij} a_{jk} a_{kl}, \ldots, a_{11} a_{11} a_{11}).$
Our task will be the following: find the elements of $R$ that get mapped to 0 (i.e., polynomials in the variables $p_{ijkl}$ that vanish under the function $f$). This elements will provide us with the relations we are looking for. Thus, we need to characterize:

$$\ker(f) := \{ g \in R : f(g) = 0 \}.$$  

The set $\ker(f)$ (called kernel of the map $f$) has some good properties. Namely, if $g \in \ker(f)$ and $h \in R$ then $h \cdot g \mapsto f(h) \cdot f(g) = f(h) \cdot 0 = 0$, so $h \cdot g \in \ker(f)$. In addition, if $g_1, g_2 \in \ker(f)$ then $g_1 + g_2 \mapsto f(g_1) + f(g_2) = 0 + 0 = 0$, so $g_1 + g_2 \in \ker(f)$. These two properties will characterize what is called an ideal of the ring $R$.

**Definition 3.8.** Let $\mathcal{F} \subset \mathbb{Q}[p_{000}, \ldots, p_{111}] = R$, then the ideal generated by $\mathcal{F}$ is the set

$$\langle \mathcal{F} \rangle = \left\{ \sum_{i \in I, \text{finite}} h_i f_i \mid h_i \in \mathbb{Q}[p_{000}, \ldots, p_{111}], f_i \in \mathcal{F} \right\}.$$  

An ideal is any set of the form $\langle \mathcal{F} \rangle$ for some subset $\mathcal{F}$.

Note that the set $\mathcal{F}$ need not be finite. In the ring of polynomials, Hilbert's Basis Theorem will tell us that we can find a finite subset of $\mathcal{F}$ generating the same ideal. This key fact will enable us to perform computations with ideals.

**Theorem 3.9.** *(Hilbert's Basis Theorem)*  
Every infinite set of polynomials $\mathcal{F} \subset \mathbb{Q}[p]$ has a finite subset $\mathcal{F}' \subset \mathcal{F}$ s.t. $\langle \mathcal{F} \rangle = \langle \mathcal{F}' \rangle$. In other words: every ideal in $\mathbb{Q}[p]$ is finitely generated.

**Remark.** The previous result gives us the answer to our question. We will have a finite set of generators of $\ker(f)$, that is a finite number of relations among the $p_{ijkl}$'s that will describe our model.

Gröbner basis will enter the scenario as a “nice” finite set of generators of the ideal $\ker(f)$. They will have nice properties that will enable us to do computations for answering several questions regarding our statistical model $F$.

### 3.6 Homework

**Convex Hull for $n = 5$.** Find the convex hull corresponding to the Markov Chain with $n = 5$ (i.e. $m = 2^5$), i.e. describe the convex hull by means of equations.

**Extra credit** Do the same problem but in the case of $n = 20$. The key point for this is that the geometry of the polytopes for different $n$'s gets fixed if $n \gg 0$. Thus, although we have $2^{20}$ sequences, the polytope will have few vertices.