A note on the border of an exponential family

Luigi Malagò and Giovanni Pistone

Abstract Limits of densities belonging to an exponential family appear in many applications, e.g. Gibbs models in Statistical Physics, relaxed combinatorial optimization, coding theory, critical likelihood computations, Bayes priors with singular support, random generation of factorial designs. We discuss the problem from the methodological point of view in the case of a finite state space. We prove two characterizations of the limit distributions, both based on a suitable description of the marginal polytope (convex hull of canonical statistics' values). First, the set of limit densities is equal to the set of conditional densities given a face of the marginal polytope. Second, in the lattice case there exists a parametric presentation, in monomial form, of the closure of the statistical model.

Key words: Algebraic Statistics, Convex Support, Extended Exponential Family, Statistical Modeling.

1 Background

We consider the *exponential family* defined by the family of densities

$$p(x;\theta) = \exp\left(\sum_{j=1}^{m} \theta_j T_j(x) - \psi(\theta)\right), \quad \theta \in \mathbb{R}^m, \tag{1}$$

on a finite state space $(\mathscr{X}, \mu)$ with $n = \#\mathscr{X}$ points and reference measure μ. Many monographs have been devoted to the study of this important class of statistical

Luigi Malagò
DEI Politecnico di Milano, Italy. e-mail: malago@elet.polimi.it

Giovanni Pistone
Collegio Carlo Alberto, Moncalieri, Italy. e-mail: giovanni.pistone@gmail.com

models, e.g. [1, 2, 11]. In this section we have collected facts from this theory and its algebraic version in order to introduce to our result discussed in Section 2.

Different exponential families could represent the same statistical model. Consider the orthogonal decomposition $\mathrm{Span}\,(1, T_1, \ldots, T_m) = 1 \oplus V \subset L^2(\mathscr{X}, \mu)$. In fact, $V \subset L_0^2(\mathscr{X}, \mu)$. For each density p in the exponential family (1) there exists a unique $v \in V$ such that $p(x) = \mathrm{e}^{v - K_\mu(v)}$, see [14, 3].

The *canonical statistics*

$$T = (T_1, \ldots, T_m) \colon \mathscr{X} \to \mathscr{Y} = T(\mathscr{X}) \subset \mathbb{R}^m$$

map the statistical model (1) to the *canonical exponential family*

$$p(y; \theta) = \exp\left(\sum_{j=1}^m \theta_j y_j - \psi(\theta) \right), \quad \theta \in \mathbb{R}^m, \tag{2}$$

where the new state space is $(\mathscr{Y}, \nu)$, with $\nu = \mu \circ T^{-1}$. In Equation (2), the canonical statistics are coordinate projections $y \mapsto y_j$, $j = 1, \ldots, m$.

1.1 Monomial and implicit presentations

Other useful parameterization of the exponential family (1) are available, in particular the mean parameterization which shall be discussed in Section 1.3. In this paper we focus on a less known parameterization, i.e. the *monomial parameterization*, which is obtained from (1) by introducing the exponentials $t_j = \mathrm{e}^{\theta_j}$ of each canonical parameter θ_j, $j = 1, \ldots, m$,

$$p(x; t) \propto \prod_{j=1}^m t_j^{T_j(x)}, \quad t \in \mathbb{R}_>^m. \tag{3}$$

This presentation is especially useful in the lattice case, i.e. when the canonical statistics are integer valued. This is the case which has been studied with the methods of Algebraic Statistics, see e.g. [13, Sec. 6.9], [9].

While Equations (1) and (3) are equivalent for positive densities, an interesting phenomenon appears if the conditions $t_j > 0$ are relaxed to $t_j \geq 0$. In such a case, (3) makes sense and an extension of the original model is obtained, see [15, 16]. For example, assume we let just one of the t_j's, say t_1, to be zero. It follows that the corresponding unnormalized density is zero if $T_1(x) \neq 0$ and is positive for $T_1(x) = 0$, giving rise to densities with support $\{T_1 = 0\}$ which form a new exponential family. Thus, the exponential family (1) is extended to include exponential families with defective support. Unfortunately, such extension depends on the canonical statistics used to describe the statistical model as an exponential family. For example, if the chosen canonical statistics are never zero, no such extension is possible.

Statistical models of type (1) admit an *implicit representation*, see [13, 12]. Let $1 \oplus V = \mathrm{Span}\,(T_0 = 1, T_1, \ldots, T_m)$ be the linear space generated by the canonical statistics together with the constant 1, and let $w_1, \ldots, w_l$ be a linear basis of the orthogonal space $(1 \oplus V)^\perp$, i.e., $1, T_1, \ldots, T_m, w_1, \ldots, w_l$ is a linear basis of $L^2(\mathscr{X}, \mu)$ and

$$\sum_{x \in \mathscr{X}} w_i(x) T_j(x) \mu(x) = 0, \quad i = 1, \ldots, l, \quad j = 0, \ldots, m.$$

If we introduce the $(m+1) \times n$ matrix $A = [T_j(x)\mu(x)]$, $j = 0, \ldots, m$, $x \in \mathscr{X}$, $T_0 = 1$, then $\mathrm{Span}\,(w_1, \ldots, w_l) = \ker A$. The case where A is integer valued is discussed in [9]. The general case is discussed in [17].

Since $\log p(\cdot; \theta)$ is an affine function of the canonical statistics T_j's, a density p belongs to the exponential model (1) if and only if p is a positive density of $(\mathscr{X}, \mu)$ and

$$\sum_{x \in \mathscr{X}} w(x) \mu(x) \log p(x) = 0, \quad w \in \mathrm{Span}\,(w_1, \ldots, w_l). \tag{4}$$

More precisely, if $p = p(\cdot; \theta)$ in (1) for a θ, then (4) holds true; vice versa, if $\sum_{x \in \mathscr{X}} w(x)\mu(x) \log p(x) = 0$ holds true for $w = w_i$, $i = 1, \ldots, l$, then $p = p(\cdot; \theta)$ for some θ.

Equation (4) is equivalent to the following equation

$$\prod_{x \in \mathscr{X}} p(x)^{w(x)} = 1, \quad w/\mu \in \mathrm{Span}\,(w_1, \ldots, w_l), \tag{5}$$

or, clearing the denominators,

$$\prod_{x:\, w(x)>0} p(x)^{w^+(x)} = \prod_{x:\, w(x)<0} p(x)^{w^-(x)}, \quad w/\mu \in \mathrm{Span}\,(w_1, \ldots, w_l), \tag{6}$$

where $w = w^+ - w^-$ and $w^+, w^- \geq 0$. Equation (6) makes sense outside the exponential model, i.e. if we assume $p(x) \geq 0$. Assume $\mathscr{X}_0 = \mathrm{Supp}\, p$ is strictly contained in $\mathscr{X}$ and satisfies Equation (5). Therefore, p belongs to the exponential model associated to the space V_0, with $1 \oplus V_0 = \mathrm{Span}\,\left(w_{1|\mathscr{X}_0}, \ldots, w_{l|\mathscr{X}_0}\right)^\perp \subset L^2\left(\mathscr{X}_0, \mu_{|\mathscr{X}_0}\right)$.

1.2 Toric statistical models

From now on we assume that the $m \times \mathscr{X}$ matrix $A = [T_j(x)]_{j=1,\ldots,d; x \in \mathscr{X}}$, is nonnegative integer valued. The nonnegativity assumption does not restrict the class of model we consider. We define

$$\mathscr{L}^\star(A) = \left\{ y \in \mathbb{Z}^{\mathscr{X}} : y \neq 0, Ay = 0 \right\}$$

be the *lattice* of A. We denote by $A(x)$, $x \in \mathscr{X}$, the columns of A. The model (3) is written $p(x; t) = t^{A(x)} = t_1^{A_1(x)} \cdots t^{A_m(x)}$ and it is called *A-model*.

Consider the homomorphism τ from the polynomial ring $\mathbb{Q}[q(x)\colon x \in \mathscr{X}]$ into the polynomial ring $\mathbb{Q}[t_j\colon j = 1,\ldots,d]$ defined by

$$\tau\colon q(x) \mapsto \prod_{j=1}^{d} t_j^{T_j(x)} = t^{A(x)}, \quad x \in \mathscr{X}.$$

The kernel of τ is a polynomial ideal $\mathrm{Ideal}(A)$, which is called the *toric ideal* of A. It is proved in [21] that $\mathrm{Ideal}(A)$ is generated as a vector space by the binomials

$$\prod_{x \in \mathscr{X}} q(x)^{w_+(x)} - \prod_{x \in \mathscr{X}} q(x)^{w_-(x)}, \quad k \in \mathscr{L}^{\star}(A)$$

and it is generated as an ideal by a finite subset of such binomials, i.e. the binomials where k is an element of the Graver basis of $\mathscr{L}^{\star}(A)$. Note that the binomials are homogeneous if, and only if, $1 \in \mathrm{Span}(A)$.

Assume now that $t_1,\ldots,t_d$ take nonnegative and not all zero real values and consider the parameterization

$$q(x) = t^{A(x)}, \quad x \in \mathscr{X}, \quad t \in Q = \mathbb{R}_+^d \setminus \{0\}.$$

Note that $t^{A(x)} = \prod_{j\colon A_j(x)\neq 0} t_j^{A_j(x)}$. Each $q(x)$ is nonnegative and strily positive if $t_1,\ldots,t_d > 0$. Let I be a subset of indices, $I \subset \{1,\ldots,d\}$ such that $t_j = 0$ for all $j \in I$. Then $q(x) \neq 0$ for all $x \in \mathscr{X}$ such that $A_j(x) = 0$, $j \in I$.

There exists at least one $x \in \mathscr{X}$ where $q(x) \neq 0$ if, and only if, each column of A contains at least one zero. In such a case, we have defined a parameterization of unnormalized probabilities q with parameters in the vertex-less quadrant:

$$Q \ni t \mapsto p(x;t) = \frac{t^{A(x)}}{\sum_{x \in \mathscr{X}} t^{A(x)}}$$

Let us study the confounding induced by such a parameterization on strictly positive parameters. If

$$\frac{s^{A(x)}}{\sum_{x \in \mathscr{X}} s^{A(x)}} = \frac{t^{A(x)}}{\sum_{x \in \mathscr{X}} t^{A(x)}}, \quad x \in \mathscr{X},$$

then the unnormalized probabilities are proportional and

$$\prod_{j=1}^{d} \left(\frac{s_j}{t_j}\right)^{A_j(x)} = \mathrm{constant}, x \in \mathscr{X}.$$

or

$$\sum_{j=1}^{d} (\log s_j - \log t_j) A_j(x) = \mathrm{constant}.$$

If, and only if, $1 \in \operatorname{Rank}(A)$, there exists vectors $c = (c_1, \ldots, c_d)$ such that $A(x)c =$ constant and $\log s_j - \log t_j = c_j$, $j = 1, \ldots, d$ or $s_j = \mathrm{e}^{c_j} t_j$. Confounding is reduced to the confounding of uniform probability.

1.3 Trace, closure, marginal polytope

In the present section we discuss two general methods under which the reduction of the support appears, namely the trace operation and the limit operation. For each event $\mathscr{S} \subset \mathscr{X}$, the *trace* on $\mathscr{S}$ of the exponential family in (1) is the exponential family defined on $(\mathscr{S}, \mu_{|S})$ by conditioning on $\mathscr{S}$.

We denote by $\mathscr{M}_>$ the convex set of strictly positive densities and by $\mathscr{M}_\geq$ the convex set of densities. Both sets are endowed with the weak topology, i.e., if p_n, $n = 1, 2, \ldots$, and p are densities, then $\lim_{n\to\infty} p_n = p$ means $\lim_{n\to\infty} p_n(x) = p(x)$ for all $x \in \mathscr{X}$. In general, the exponential model (1) is not closed in the weak topology. The *extended exponential family* is the closure in the weak topology of an exponential family (1). An extended exponential family according to this definition is a set of densities. A proper parameterization of the extended family requires the use of the expectation parameters and the identification of their range.

Definition 1. The *convex support*, cf. e.g [4, 2, 7], or *marginal polytope*, see [23], and also [10], of the exponential family (1) is the convex hull of $\mathscr{Y} = T(\mathscr{X})$,

$$\operatorname{co}(\operatorname{im} T) = \left\{ \eta \in \mathbb{R}^m, \eta = \sum_{j=1}^m \lambda_j t_j : \lambda_j \geq 0, \sum_{j=1}^m \lambda_j = 1 \right\}.$$

The previous set-up covers the behavior of the exponential family and its parameterization with the expectation parameters in the interior of the marginal polytope, see [2]. The discussion of the parameterization of the extended family requires the notion of exposed subset.

Definition 2. 1. A *face* of the marginal polytope M is a subset $F \subset M$ such that there exists an affine mapping $A \colon \mathbb{R}^m \ni t \mapsto A(t) \in \mathbb{R}$ which is zero on F and strictly positive on $M \setminus F$.
2. A subset $S \subset \mathscr{X}$ is *exposed for the exponential family* (2) if $S = T^{-1}(F)$ and F is a face of the marginal polytope.

The following theorem is a minor improvement of known results.

Theorem 1. *Let* θ_n, $n = 1, 2, \ldots$, *be a sequence of parameters in Equation* (1) *such that for some* $q \in \mathscr{M}_\geq$ *we have* $\lim_{n\to\infty} p(x; \theta_n) = q(x)$, *i.e.,* q *belongs to the extended exponential model.*

1. *If the support of* q *is full,* $\{q > 0\} = \mathscr{X}$, *then* q *belongs to the exponential family* (1) *for some parameter value* $\theta = \lim_{n\to\infty} \theta_n$.

2. *If the support of q is defective, then the sequence θ_n is not convergent, $\operatorname{Supp} q = \{q > 0\}$ is an exposed subset of $\mathscr{X}$, and q belongs to the trace of the exponential family on the reduced support.*

Proof. Let $\mathscr{X}_0 = \{x \in \mathscr{X} : q(x) > 0\}$, $\mathscr{X}_1 = \{x \in \mathscr{X} : q(x) = 0\}$. For each $x \in \mathscr{X}_0$, we have $\lim_{n \to \infty} \log p(x; \theta_n) = \log q(x)$ by continuity; for each $x \in \mathscr{X}_1$, we have $\lim_{n \to \infty} \log p(x; \theta_n) = -\infty$. From (4) we get

$$\sum_{x \in \mathscr{X}_0} \log q(x) k(x) \mu(x) + \lim_{n \to \infty} \sum_{x \in \mathscr{X}_1} \log p(x; \theta_n) k(x) \mu(x) = 0, \qquad (7)$$

with $k \in \operatorname{Span}(w_1, \ldots, w_l)$.

1. If the set $\mathscr{X}_1$ is empty, then q belongs to the exponential model because Equation (7) reduces to (4). The convergence $\lim_{n \to \infty} \eta_n = \lim_{n \to \infty} \operatorname{E}_{\theta_n}[T] = \operatorname{E}_q[T] = \eta$ in M° implies the convergence of the θ parameters (mod the identifiability constraints).
2. If the set $\mathscr{X}_1$ is not empty, the second term of the LHS of (7) has to be finite, so that no linear combination of the w_i's can be definite in sign. Otherwise, the limit would diverge. In other words, the problem

$$k : \mathscr{X}_1 \ni x \mapsto \sum_{i=1}^{l} \lambda_i w_i(x) \geq 0 \text{ and } k \neq 0 \text{ for at least one } x \qquad (8)$$

is not satisfiable. By the Theorem of the alternative, see e.g. [19, Ch. 15], the non satisfiability of (13) is equivalent to the existence of a positive solution $u^{(1)}(x) > 0$, $x \in \mathscr{X}_1$, to the problem

$$\sum_{x \in \mathscr{X}_1} u^{(1)}(x) k(x) \mu(x) = 0, \quad k \in \operatorname{Span}(w_1, \ldots, w_l).$$

The random variable

$$u(x) = \begin{cases} 0 & \text{if } x \in \mathscr{X}_0, \\ u^{(1)}(x) & \text{if } x \in \mathscr{X}_1, \end{cases}$$

is orthogonal to all w_i's, so that there exist $a_0, a_1, \ldots, a_m$ such that

$$u(x) = a_0 + \sum_{j=1}^{m} a_j T_j(x). \qquad (9)$$

The conclusion on the support now follows from (9). In fact, for each $t \in \mathscr{Y}$ such that $T^{-1}(t) \in \mathscr{X}_1$ the linear function $a_0 + \sum_j a_j t_j$ is positive, while for each t such that $T^{-1}(t) \in \mathscr{X}_0$ takes value zero, so that the points in $\mathscr{X}_1$ are the points of an exposed set of the face of M identified by (9).

Finally, on the support of q, $\log q$ is a linear combination of the T_j's being a limit in the linear space generated by those functions. $\square$

Theorem 2. *If q belongs to the trace of the exponential family* (1) *with respect to an exposed subset S, then q belongs to the extended exponential model.*

Proof. We generate sequences that admit as limit a generic density in the trace model by considering a one-dimensional (Gibbs) sub-model. Let F be the face of the marginal polytope such that $S = T^{-1}(F)$ and let A be an affine function such that $A(\eta) = 0$ for $\eta \in F$ and $A(\eta) > 0$ for $\eta \in M \setminus F$. We can chose A such that $A \circ T$ belongs to the space generated by $1, T_1, \ldots, T_m$, i.e. $A \circ T = \alpha_0 + \sum_{j=1}^m \alpha_j T_j$. We can take $\alpha_0 = 0$ if $1 \in \operatorname{Span}(T_j \colon j = 1, \ldots, m)$.

Let $\bar{\theta}$ be a value of the canonical parameter such that

$$
q(x) = \begin{cases} \dfrac{\exp\left(\sum_{j=1}^m \bar{\theta}_j T_j(x)\right)}{\sum_{x \in S} \exp\left(\sum_{j=1}^m \bar{\theta}_j T_j(x)\right) \mu(x)} & \text{if } x \in S, \\ 0 & \text{if } x \in \mathscr{X} \setminus S. \end{cases}
$$

For $\beta \in \mathbb{R}$,

$$
\beta A + \sum_{j=1}^m \bar{\theta}_j T_j = \sum_{j=1}^m (\beta \alpha_j + \bar{\theta}_j) T_j + \beta \alpha_0,
$$

so that the one-dimensional statistical model

$$
p_\beta = \exp\left(\beta(A - \alpha_0) + \sum_{j=1}^m \bar{\theta}_j T_j - \psi(\beta \alpha + \bar{\theta})\right), \quad \beta \in \mathbb{R},
$$

is a sub-model of (1). The family of densities

$$
\frac{p_\beta}{p_0} = \exp\left(\beta(A - \alpha_0) - \left(\psi(\beta \alpha + \bar{\theta}) - \psi(\bar{\theta})\right)\right)
$$

is a one-dimensional exponential family whose canonical statistics $A - \alpha_0$ reaches its minimum value $-\alpha_0$ on S. Therefore, if $\beta_n \to -\infty$, $n \to \infty$, its limit is the uniform distribution on S and, consequently, p_{β_n} is convergent to q. $\quad\square$

2 Extended families

In this section we assume the exponential family (1) to be of lattice type, i.e. we assume that the $m \times n$ matrix $A = [T_j(x)\mu(x)]$, $j = 1, \ldots, m$ and $x \in \mathscr{X}$, is non-negative integer valued. Hence, the exponential family can be written as in Equation (3) and takes the monomial parametric form

$$
p(x; \zeta) \propto \prod_{j \colon A_j(x) > 0} \zeta_j^{A_j(x)}, \quad \zeta_j \geq 0, \quad j = 1, \ldots, m. \tag{10}
$$

In [9] the statistical model (10) is called the *A-model*, see also [8]. If all ζ_j's are positive, then (10) is the exponential family with a different parameterization. If we

let one, or more, of the ζ_j's to be zero, either the monomials in (10) are zero for all $x \in \mathscr{X}$, in which case no probability is defined, or the monomials are non-zero for some x, giving rise to a statistical model with restricted support, see the discussion in [9].

Each integer vector k such that $Ak = 0$, i.e. $k \in \ker_{\mathbb{Z}} A$, splits into its positive and negative part, $k = k^+ - k^-$, and we have

$$\prod_{x:\, k^+(x)>0} p(x,\zeta)^{k^+(x)} = \prod_{x:\, k^-(x)>0} p(x;\zeta)^{k^-(x)}, \quad k \in \ker_{\mathbb{Z}} A. \qquad (11)$$

The statistical model defined by the infinite system of binomial equations (11) is called the *toric model* of A, as defined in [13]. Again, if all the probabilities in (11) are positive, then the toric model is just the exponential family. If some probabilities are zero, then the toric model implies the A-model. In fact, substitution of (10) into (11) leads to an algebraic identity, without any restriction on the parameters ζ_j.

The existence of a finite generating set for Equation (11) is discussed in details in [9], see also [8]. Moreover, in [9] it is proved that each probability in the extended exponential family satisfies (11). We shall obtain a related result in a different way.

Consider a second $l \times n$ matrix B with the same integer ker as A. The exponential model would be the same, but the border cases of the A-model could be different then the border cases of the B-model. The problem of finding a suitable maximal monomial model was considered first in [16] and it is fully discussed in [17]. Rapallo's method has been applied in [6] to the Bayesian analysis of tables with structural zeros. Here, we show that all of the extended exponential family is actually parameterized by this maximal monomial model. For a related approach see also [18].

The maximality of the monomial model is defined as follows. Consider the model matrix $A \in \mathbb{Z}^{m \times n}$. If the constant vector $\mathbf{1}$ does not belong to the row space switch to the matrix $[\mathbf{1}A]$ which defines the same exponential model. Let the column span of the orthogonal matrix $K = [w_1 \cdots w_l] \in \mathbb{Z}^{n \times l}$ be $\ker_{\mathbb{Q}} A$. The integer matrix K can be computed by a symbolic algebra software, such as [5, 22]. Numeric software might be unsuitable because it will normally produce floating point unit vectors, not integer vectors.

Consider all possible rows of a non-negative matrix equivalent to A, i.e. producing the same statistical model when all the parameters are strictly positive:

$$\mathscr{B} = \left\{ b \in \mathrm{Span}_{\mathbb{Q}}(A) : b \neq 0, b \in \mathbb{Z}_+^n \right\} = \left\{ b \in \mathbb{Z}_+^n : b \neq 0, b^T K = 0 \right\}.$$

The set $\mathscr{B}$ is closed for the sum of vectors. It is proved in [20] that a unique inclusion-minimal generating set, called *Hilbert basis*, exists. The Hilbert basis can be computed by symbolic software [5, 22]. It is a $\mathbb{Q}$-generating set but it is usually much larger than a lattice basis.

The following theorem was stated first il [16] without a complete proof, see also the discussion in [17].

Theorem 3. *Let us consider the set $\mathscr{B}$ of non-negative and non-zero integer vectors that are orthogonal to $\ker_{\mathbb{Z}} A$ and let $b_1, \ldots, b_l$ be its unique Hilbert basis. Define a $l \times n$ matrix B whose rows are the elements of the Hilbert basis. Hence, the extended exponential family is fully parametrized by the B-model with non-negative parameters i.e. each one of the maximal exposed subsets of the A-model is obtained by letting one of the ζ_j's to be zero.*

Proof. The constant vector belongs to $\mathscr{B}$, therefore $1, b_1, \ldots, b_l$ is a $\mathbb{Q}$-vector generating set, possibly non-minimal. In fact, any rational basis of $\mathrm{Span}_{\mathbb{Q}}(A)$ becomes an integer basis by multiplication by a suitable integer; the integer basis is transformed to an integer positive basis by adding, where needed, a constant integer vector; each of the vector obtained in such a way belong to $\mathscr{B}$.

The sets $S_j = \left\{ x \in \mathscr{X} : b_j(x) = 0 \right\}$, $j = 1, \ldots, l$ are non empty. In fact, if $m_j = \min_x b_j(x) > 0$, as $b_j(x) \neq 0$ for some x, the vector $b_j - m_j \mathbf{1}$ belongs to $\mathscr{B}$, and therefore can be represented as

$$ b_j(x) - m_j = \sum_{i=1}^{l} n_i b_i(x), \quad x \in \mathscr{X}. $$

If $n_j = 0$, the basis is not minimal. If $n_j \geq 1$, subtracting $b_j(x)$ from both sides, we get, by inspection of the signs of the two sides, that $m_j = 0$.

Each of the S_j's is an exposed set of the exponential family. In fact, each element of the Hilbert basis belongs to the row $\mathbb{Q} - \mathrm{Span}$ of the original matrix A, so that

$$ b_j(x) = \beta_{0j} + \sum_{i=1}^{m} \beta_{ij} a_i(x), \quad j = 1, \ldots, l, $$

where a_i is the i-th row of A. The definition of exposed set is easily checked.

Vice-versa, let $\mathscr{S}$ be an exposed set, i.e.

$$ b(x) = \beta_0 + \sum_{i=1}^{m} \beta_i a_i(x), $$

with $\mathscr{S} = \{x \colon b(x) = 0\}$ and $b(x) > 0$ for each $x \notin \mathscr{S}$. As A has integer entries, the coefficients $\beta_0, \beta_1, \ldots, \beta_l$ can be chosen to have integer values, therefore $b \in \mathscr{B}$ and it is a sum of elements of the Hilbert basis,

$$ b(x) = \sum_{j=1}^{l} \alpha_j b_j(x), \quad \alpha_j \in \mathbb{Z}_+, \quad j = 1, \ldots, l. $$

Therefore, $S = \cap_{j \colon \alpha_j \neq 0} S_j$. $\square$

Remark 1. The additive representation of b for maximal exposed sets contains only one term. However, the Hilbert basis might contain an element b_j such that its zero set S_j is the intersection of other S_j's. In such a case, such a b_j could be dropped from the B-model without loosing any part of the extended family.

3 Examples

3.1 4-cycle

The 4-cycle is the exponential family

$$\exp\left(\theta_D D + \theta_C C + \theta_B B + \theta_A A + \theta_{BA} BA + \theta_{CB} CB + \theta_{DC} DC + \theta_{AD} AD - \psi(\theta)\right)$$

where A, B, C, D are the coordinates of $\mathscr{X} = \{\pm 1\}^4$. The model matrix A and the B matrix are shown in the following edited R output:

$\mathscr{X}$	I	D	C	B	A	BA	CB	DC	DA	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	b_{10}	b_{11}	b_{12}	b_{13}	b_{14}	b_{15}	b_{16}	b_{17}	b_{18}	b_{19}	b_{20}	b_{21}	b_{22}	b_{23}	b_{24}
++++	1	1	1	1	1	1	1	1	1	0	1	1	1	0	1	0	0	0	0	0	0	0	0	0	1	1	0	0	1	0	0	1	0
+++-	1	1	1	1	-1	-1	1	1	-1	0	1	1	0	0	0	0	0	0	1	1	1	0	0	0	0	1	0	0	1	0	1	0	0
++-+	1	1	1	-1	1	1	-1	1	1	0	1	0	1	0	0	0	1	0	1	0	0	0	0	0	0	1	1	0	0	0	1	1	0
++--	1	1	1	-1	-1	1	-1	1	-1	0	1	0	0	0	0	0	0	0	1	1	0	0	1	0	0	1	1	1	0	0	0	1	0
+-++	1	-1	1	1	1	1	1	-1	-1	1	0	0	1	0	1	1	1	0	0	0	0	1	0	1	0	1	0	0	0	0	0	1	0
+-+-	1	-1	1	1	-1	-1	1	-1	1	1	0	0	0	0	0	1	1	0	1	1	1	0	1	0	0	0	0	0	0	1	0	0	1
+--+	1	-1	1	-1	1	-1	-1	-1	1	1	0	0	1	0	0	0	1	0	0	0	0	1	0	0	1	0	0	0	1	0	1	0	1
+---	1	-1	1	-1	-1	1	-1	-1	-1	1	0	0	0	0	0	0	0	0	1	1	0	1	1	1	0	1	1	1	1	1	0	1	0
-+++	1	-1	1	1	1	1	1	-1	-1	0	0	1	0	1	1	0	0	0	0	0	0	0	1	0	1	1	0	0	0	1	1	0	0
-++-	1	-1	1	1	-1	-1	1	-1	1	0	0	1	0	0	0	0	0	1	0	1	0	0	0	1	0	0	0	0	1	1	0	0	0
-+-+	1	-1	1	-1	1	-1	-1	-1	-1	0	0	0	1	0	0	1	0	1	0	0	0	0	1	0	0	0	1	0	1	0	1	1	0	1
-+--	1	-1	1	-1	-1	1	-1	-1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	1	1	1	0	1	0	1	0
--++	1	-1	-1	1	1	1	-1	1	-1	0	0	0	1	1	1	0	0	1	0	0	0	1	1	0	1	0	0	0	0	1	0	0	1
--+-	1	-1	-1	1	-1	-1	-1	1	1	0	0	0	0	0	0	1	1	1	1	0	1	0	0	1	0	0	0	1	0	0	1	0	0
---+	1	-1	-1	-1	1	-1	1	1	-1	0	0	0	0	1	0	1	0	0	0	0	1	0	1	1	0	0	0	0	1	0	1	0	1
----	1	-1	-1	-1	-1	1	1	1	1	0	0	0	0	0	0	0	0	1	0	0	0	1	0	1	1	1	1	0	1	0	0	1	0

In this exemple all the b_j vectors are binary vectors, which implies they are all indispensable. The vectors $F_j = 1 - b_j$ are the indicator functions of the S_j sets. The polynomial representation is (after multiplication by $1/16$):

θ	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8	F_9	F_{10}	F_{11}	F_{12}	F_{13}	F_{14}	F_{15}	F_{16}	F_{17}	F_{18}	F_{19}	F_{20}	F_{21}	F_{22}	F_{23}	F_{24}
I	12	12	12	12	12	12	12	8	12	8	12	12	12	8	12	8	8	12	12	8	12	8	8	12
D	-4	-4	0	-4	4	0	0	0	4	0	-4	0	0	0	4	0	0	0	0	0	4	0	0	0
C	4	-4	-4	0	0	0	4	0	0	4	0	4	0	0	0	-4	0	-4	0	0	0	0	0	0
B	0	0	-4	0	0	-4	-4	0	0	0	0	-4	4	0	0	0	4	4	0	0	0	0	0	4
A	0	0	0	-4	-4	-4	0	0	4	4	0	0	0	0	4	0	0	0	4	0	0	0	0	-4
BA	0	0	0	0	-4	0	4	0	4	0	4	0	-4	0	-4	-4	-4	0	4	0	4	-4	4	
CB	0	0	-4	0	0	0	4	4	0	4	0	0	-4	0	-4	0	4	-4	0	-4	0	-4	4	0
DC	4	-4	0	0	0	0	4	0	-4	0	0	0	4	-4	-4	0	0	-4	4	4	-4	4	-4	0
DA	0	0	0	-4	4	0	0	-4	-4	4	4	0	0	4	0	-4	4	0	0	-4	0	4	-4	0

e.g.

$$F_1 = \frac{3}{4} - \frac{1}{4}D + \frac{1}{4}C + \frac{1}{4}DC$$

that is $DC = D - C$ on S_1.

The Gröbner basis of each ideal $\langle A^2 - 1, B^2 - 1, C^2 - 1, D^2 - 1, F_j - 1 \rangle$ reveals in a different way the aliasing induced on each facet.

Next: polynomial representation of the model.

3.2 Markov chain

Let X_t, $t = 0, 1, \ldots, n$ be a Markov chain with stationary transitions on the binary state space $\{0, 1\}$. Let us denote by $t_x = P(X_0 = x)$, $x = 0, 1$, the initial probability and with $t_{xy} = P(X_1 = y | X_0 = x)$, $x, y = 0, 1$, the transition probabilities. For each

trajectory $\omega \in \mathscr{X} = \{0,1\}^{n+1}$ the probability of the trajectory is

$$p(\omega) = t_0^{(1-X_0(\omega))} t_1^{X_0} \prod_{x,y=0}^{1} t_{xy}^{N_{xy}(\omega)}, \tag{12}$$

where $N_{xy}(\omega)$ is the number of transitions from x to y appearing in the trajectory ω.

This in an instance of the toric model of the $\mathscr{X} \times 6$ matrix whose rows are

$$\left[(1 - X_0) \ X_0 \ N_{00} \ N_{01} \ N_{10} \ N_{11}\right].$$

Let us compute the confounding, i.e. find the vectors $c \in \mathbb{R}^6$ such that

$$c_0(1 - X_0(\omega)) + c_1 X_0(\omega)$$
$$+ c_{00} N_{00}(\omega) + c_{01} N_{01}(\omega) + C_{10} N_{10}(\omega) + c_{11} N_{11}(\omega) = \alpha, \quad \omega \in \mathscr{X}.$$

for some α. Note the following equalities:

$$N_{00} = \sum_{t=1}^{n} (1 - X_{t-1})(1 - X_t) = n - X_0 - 2\sum_{t=1}^{n-1} X_t - X_n + \sum_{t=1}^{n} X_{t-1} X_t,$$

$$N_{01} = \sum_{t=1}^{n} (1 - X_{t-1}) X_t = \sum_{t=1}^{n-1} X_t + X_n - \sum_{t=1}^{n} X_{t-1} X_t,$$

$$N_{10} = \sum_{t=1}^{n} X_{t-1}(1 - X_t) = X_0 + \sum_{t=1}^{n-1} X_t - \sum_{t=1}^{n} X_{t-1} X_t,$$

$$N_{11} = \sum_{t=1}^{n} X_{t-1} X_t.$$

Expanding the equation for c and observing that the vectors 1, X_0, $\sum_{t=1}^{n-1} X_t$, X_n, $\sum_{t=1}^{n} X_{t-1} X_t$ are linearly independent, we obtain, equating to zero the coefficient of each vector, that

$$c_0 + n c_{00} = \alpha$$
$$c_1 - c_0 - c_{00} + c_{10} = 0$$
$$-2 c_{00} + c_{01} + c_{10} = 0$$
$$-c_{00} + c_{01} = 0$$
$$c_{00} - c_{01} - c_{10} + c_{11} = 0$$

The solution of the previous system is

$$c_0 = c_1, \quad c_{00} = c_{01} = c_{10} = c_{11}.$$

It follows that an identifiable parameterization of the exponential model from (12) is

$$p(\omega) = t_0^{(1-X_0(\omega))} t_1^{X_0} \prod_{x,y=0}^{1} t_{xy}^{N_{xy}(\omega)}, \quad t_0 + t_1 = 1, \quad \sum_{xy} t_{xy} = 2, \qquad (13)$$

while the Markov chain model is the submodel

$$p(\omega) = t_0^{(1-X_0(\omega))} t_1^{X_0} \prod_{x,y=0}^{1} t_{xy}^{N_{xy}(\omega)}, \quad t_0 + t_1 = 1, \quad t_{00} + t_{01} = 1, \quad t_{10} + t_{11} = 1. \quad (14)$$

The orthogonal space of the model matrix is generated by the vector $k = (n, n, 1, 1, 1, 1)$

References

1. Barndorff-Nielsen, O.E.: Information and Exponential Families in Statistical Theory. John Wiley & Sons, New York (1978)
2. Brown, L.D.: Fundamentals of statistical exponential families with applications in statistical decision theory. Institute of Mathematical Statistics Lecture Notes—Monograph Series, 9. Institute of Mathematical Statistics, Hayward, CA (1986)
3. Cena, A., Pistone, G.: Exponential statistical manifold. Ann. Inst. Statist. Math. **59**(1), 27–56 (2007)
4. Čencov, N.N.: Statistical decision rules and optimal inference, *Translations of Mathematical Monographs*, vol. 53. American Mathematical Society, Providence, R.I. (1982). Translation from the Russian edited by Lev J. Leifman
5. CoCoATeam: CoCoA: a system for doing Computations in Commutative Algebra. Available at cocoa.dima.unige.it (online)
6. Consonni, G., Pistone, G.: Algebraic Bayesian analysis of contingency tables with possibly zero-probability cells. Statist. Sinica **17**(4), 1355–1370 (2007)
7. Csiszár, I., Matúš, F.: Closures of exponential families. Ann. Probab. **33**(2), 582–600 (2005)
8. Drton, M., Sturmfels, B., Sullivant, S.: Lectures on Algebraic Statistics. No. 39 in Oberwolfach Seminars. Birkhäuser (2009)
9. Geiger, D., Meek, C., Sturmfels, B.: On the toric algebra of graphical models. Ann. Statist. **34**(3), 1463–1492 (2006)
10. Kahle, T.: Neighborliness of marginal polytopes (2010). Accepted in *Contributions to Algebra and Geometry,* arXiv:0809.0786
11. Letac, G.: Lectures on natural exponential families and their variance functions, *Monografías de Matemática [Mathematical Monographs]*, vol. 50. Instituto de Matemática Pura e Aplicada (IMPA), Rio de Janeiro (1992)
12. Pistone, G.: Algebraic varieties vs. differentiable manifolds in statistical models. In: P. Gibilisco, E. Riccomagno, M. Rogantin, H.P. Wynn (eds.) Algebraic and Geometric Methods in Statistics, chap. 21, pp. 339–363. Cambridge University Press (2009)

13. Pistone, G., Riccomagno, E., Wynn, H.P.: Algebraic statistics. Computational commutative algebra in statistics, *Monographs on Statistics and Applied Probability*, vol. 89. Chapman & Hall/CRC, Boca Raton, FL (2001)
14. Pistone, G., Sempi, C.: An infinite-dimensional geometric structure on the space of all the probability measures equivalent to a given one. Ann. Statist. **23**(5), 1543–1561 (1995)
15. Rapallo, F.: Log-linear models and toric ideals. Ph.D. thesis, Dottorato in Matematica Università Genova, Università di Torino, Politecnico di Torino (2003)
16. Rapallo, F.: Toric statistical models: parametric and binomial representations. Ann. Inst. Statist. Math. **59**(4), 727–740 (2007). DOI 10.1007/s10463-006-0079-z
17. Rauh, J., Kahle, T., Ay, N.: Support sets in exponential families and oriented matroid theory (2009). Proceedings WUPES'09, invited for special issue of IJAR arXiv:0906.5462
18. Rinaldo, A., Fienberg, S.E., Zhou, Y.: On the geometry of discrete exponential families with application to exponential random graph models. Electronic Journal of Statistics **3**, 446–484 (2009)
19. Roman, S.: Advanced linear algebra, *Graduate Texts in Mathematics*, vol. 135, second edn. Springer, New York (2005)
20. Schrijver, A.: Theory of linear and integer programming. Wiley-Interscience Series in Discrete Mathematics. John Wiley & Sons Ltd., Chichester (1986)
21. Sturmfels, B.: Gröbner bases and convex polytopes. American Mathematical Society, Providence, RI (1996)
22. 4ti2 team: 4ti2—a software package for algebraic, geometric and combinatorial problems on linear spaces. Available at www.4ti2.de
23. Wainwright, M.J., Jordan, M.I.: Graphical models, exponential families, and variational inference. Foundations and Trends in Machine Learning **1**(1-2), 1–305 (2008). DOI 10.1561/2200000001