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

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

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

Giovanni Pistone
DIMAT Politecnico di Torino and Collegio Carlo Alberto, Moncalieri, Italy. e-mail: `giovanni.pistone@gmail.com`

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 models, e.g. [1, 3, 2].

The *canonical statistics*

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

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

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

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

Another useful parameterization is the *monomial parameterization*, which is obtained from (1) by introducing the exponentials $\zeta_j=\mathrm{e}^{\theta_j}$ of each canonical parameter θ_j, $j=1,\dots,m$,

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

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. [11, Sec. 6.9], [8].

While Equations (1) and (3) are equivalent for positive densities, an interesting phenomenon appears if the conditions $\zeta_j>0$ are relaxed to $\zeta_j\geq 0$. In such a case, (3) makes sense and an extension of the original model is obtained, see [12, 13]. Assume we let just one of the ζ_j's, say ζ_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. The exponential family (1) is extended with exponential families of defective support and the extension actually depends on the canonical statistics used to describe the exponential statistical model. For example, if the canonical statistics are never zero, no such extension is possible.

Statistical models of type (1) admit an *implicit representation*, see [11, 10]. Let $1\oplus V=\operatorname{Span}(T_0=1,T_1,\dots,T_m)$ be the linear space generated by the canonical statistics together with the constant 1, and let $k_1,\dots,k_l$ be a linear basis of the orthogonal space $(1\oplus V)^\perp$, i.e., $1,T_1,\dots,T_m,k_1,\dots,k_l$ is a linear basis of $L^2(\mathscr{X},\mu)$ and

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

If we introduce the $(m+1)\times n$ matrix $A=[T_j(x)\mu(x)]$, $j=0,\dots,m$, $x\in\mathscr{X}$, $T_0=1$, then $\operatorname{Span}(k_1,\dots,k_l)=\ker A$. The case where A is integer valued is discussed in [8]. The general case is discussed in [14].

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

$$\sum_{x\in\mathscr{X}} k(x)\mu(x)\log p(x) = 0, \quad k \in \operatorname{Span}(k_1,\dots,k_l). \tag{4}$$

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

Equation (4) is equivalent to the following equation

$$\prod_{x\in\mathscr{X}} p(x)^{h(x)} = 1, \quad \frac{h(x)}{\mu(x)} \in \operatorname{Span}(k_1,\dots,k_l), \tag{5}$$

or, clearing the denominators,

$$\prod_{x\colon h(x)>0} p(x)^{h^+(x)} = \prod_{x\colon h(x)<0} p(x)^{h^-(x)}, \quad \frac{h(x)}{\mu(x)} \in \operatorname{Span}(k_1,\dots,k_l), \tag{6}$$

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

2 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, $n = 1,2,\dots$, 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 [3, 2, 6], or *marginal polytope*, see [19], and also [9], 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

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

with $k \in \operatorname{Span}(k_1,\ldots,k_l)$.

1. If the set $\mathscr{X}_1$ is empty, then q belongs to the exponential model because Equation (7) reduces to (4). The convergence $\lim_{n\to\infty} \eta_n = \lim_{n\to\infty} \mathrm{E}_{\theta_n}[T] = \mathrm{E}_q[T] = \eta$ 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 k_i's can be definite in sign. Otherwise, the limit would diverge. In other words, the problem

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

is not satisfiable. By the Theorem of the alternative, see e.g. [16, Ch. 15], the non satisfiability of (8) is equivalent to the existence of a positive solution $u^{(1)}(x) > 0$, $x \in \mathscr{X}_1$, to the problem

$$\sum_{x\in\mathscr{X}_1} u^{(1)}(x)k(x)\mu(x) = 0, \quad k \in \operatorname{Span}(k_1,\ldots,k_l)\,.$$

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 k_i's, so that there exist $a_0, a_1, \ldots, a_m$ such that

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

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. □

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. □

3 Toric 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 $A = [T_j(x)\mu(x)]$, $j = 1,\dots,m$ 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

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

In [8] the statistical model (10) is called the *A-model*, see also [7]. 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 [8].

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

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

The statistical model defined by the infinite system of binomial equations (11) is called the *toric model* of A, as defined in [11]. 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 [8], see also [7]. Moreover, in [8] it is proved that each probability in the extended exponential family satisfies (11). We shall obtain a related result in a different way.

One could 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 could be different. The problem of finding a suitable maximal monomial model is considered in [13]. This approach has been applied in [5] 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 [15].

The maximality of the monomial model is defined as follows. Consider the matrix $A \in \mathbb{Z}^{m\times n}$ and assume that the constant vectors belong to the row space. Let the column span of the matrix $K = [k_1 \cdots k_l] \in \mathbb{Z}^{n\times l}$ be $\ker_{\mathbb{Q}} A$. The integer matrix K can be computed by a symbolic algebra software, such as [4, 18]. Consider all possible rows of a matrix equivalent to A,

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

The set $\mathscr{B}$ is closed for the sum of vectors. It is proved in [17] that an inclusion-minimal generating set, called Hilbert basis, exists. The Hilbert basis can be computed by symbolic software and it is usually larger than a linear basis.

Theorem 3. *Let us consider the set 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 the* $l \times n$ *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. Each one of the maximal exposed subsets of the* A*-model, i.e.* $S = T^{-1}(F)$ *and* F *is a facet, is obtained by letting one of the* ζ_j*'s to be zero.*

Proof. The constant vector belongs to $\mathscr{B}$, so that $1, b_1, \ldots, b_l$ is a generating set, possibly non-minimal. The sets $S_j = \left\{ x \in \mathscr{X} : b_j(x) = 0 \right\}$, $j = 1, \ldots, l$ are not empty, unless b_j is constant. In fact, if $m_j = \min_x b_j(x) > 0$ and $b_j(x) \neq 0$ for some x, the vector $b_j - m_j$ 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 $n_j \neq 0$, 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 non empty S_j is an exposed set of the exponential family. In fact, each element of the Hilbert basis belongs to the row $\mathbb{Q}$-space 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 S be an exposed set, i.e.

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

with $S = \{x \colon b(x) = 0\}$ and $b(x) > 0$ for each $x \notin S$. 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,

$$b(x) = \sum_{j=1}^{l} \alpha_j b_j(x).$$

Therefore, $S = \cap_{j \colon \alpha_j \neq 0} S_j$. The additive representation of b for maximal exposed sets contains only one term. $\square$

Acknowledgements Part of this work was prepared by L. Malagò during a visit to the group led by Nihat Ay at Max Planck Institute for Mathematics in the Sciences. L. Malagò thanks his supervisor M. Matteucci for support. G. Pistone acknowledges the support of DIMAT Politecnico di Torino, SAMSI NC, and Collegio Carlo Alberto Moncalieri. The authors thank T. Kahle for suggestions and comments on an early draft and S. Sullivan for suggestions about Theorem 3.

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. 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
4. CoCoATeam: CoCoA: a system for doing Computations in Commutative Algebra. Available at cocoa.dima.unige.it
5. Consonni, G., Pistone, G.: Algebraic Bayesian analysis of contingency tables with possibly zero-probability cells. Statist. Sinica **17**(4), 1355–1370 (2007)
6. Csiszár, I., Matúš, F.: Closures of exponential families. Ann. Probab. **33**(2), 582–600 (2005)
7. Drton, M., Sturmfels, B., Sullivant, S.: Lectures on Algebraic Statistics. No. 39 in Oberwolfach Seminars. Birkhäuser (2009)
8. Geiger, D., Meek, C., Sturmfels, B.: On the toric algebra of graphical models. Ann. Statist. **34**(3), 1463–1492 (2006)
9. Kahle, T.: Neighborliness of marginal polytopes (2010). Accepted in *Contributions to Algebra and Geometry,* arXiv:0809.0786
10. 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)
11. 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)
12. Rapallo, F.: Log-linear models and toric ideals. Ph.D. thesis, Dottorato in Matematica Università Genova, Università di Torino, Politecnico di Torino (2003)
13. 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
14. 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
15. 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)
16. Roman, S.: Advanced linear algebra, *Graduate Texts in Mathematics*, vol. 135, second edn. Springer, New York (2005)
17. Schrijver, A.: Theory of linear and integer programming. Wiley-Interscience Series in Discrete Mathematics. John Wiley & Sons Ltd., Chichester (1986)
18. 4ti2 team: 4ti2—a software package for algebraic, geometric and combinatorial problems on linear spaces. Available at www.4ti2.de
19. 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