

Applications of Algebraic Geometry  
IMA Annual Program Year Workshop  
**Applications in Biology, Dynamics, and Statistics**  
March 5-9, 2007

**Information Geometry (IG) and Algebraic Statistics (AS)**

Giovanni Pistone (DIMAT Politecnico di Torino)

<http://staff.polito.it/giovanni.pistone/>  
giovanni.pistone@polito.it

## Example: new cancer incidence and gender

- D. Geiger, C. Meek, and B. Sturmfels. On the toric algebra of graphical models. *Ann. Statist.*, 34:1463–1492, 2006.
- F. Rapallo. Toric statistical models: Parametric and binomial representations. *AISM*, 2007. DOI 10.1007/s10463-006-0079-z
- G. Consonni and G. Pistone. Bayesian analysis of contingency tables with possibly zero-probability cells. Technical Report 0703123, arXiv, 2007

The following table reports data for different types of cancer separated by gender for Alaska in year 1989 (Smirnoff 2003).

| <i>Type of cancer</i> | Female | Male | Total |
|-----------------------|--------|------|-------|
| Lung                  | 38     | 90   | 128   |
| Melanoma              | 15     | 15   | 30    |
| Ovarian               | 18     | *    | 18    |
| Prostate              | *      | 111  | 111   |
| Stomach               | 0      | 5    | 5     |
| Total                 | 71     | 221  | 292   |

Clearly cells  $(3,2)$  and  $(4,1)$  are structural zeros, while we regard the zero count corresponding to the combination  $(\text{Stomach}, \text{Female})=(5,1)$  as a possibly zero-probability cell.

## Quasi-independence

- A typical assumption that is of interest in this case is that of *quasi-independence* ( $QI$ ), corresponding to the standard independence assumption for all sub-tables, excluding those having a structural zero.
- For this hypothesis, Simonoff (2003) finds a  $p$ -value between 2% and 3%, depending on the method employed. The data thus seem to provide significant evidence against the  $QI$ -model, although this evidence is not very strong.
- Let  $I = \{1, 2, 3, 4, 5\}$ ,  $J = \{1, 2\}$  denote the set of levels for the rows and columns respectively, and consider the two-way table with cells in the set  $A = I \times J \setminus \{(3, 2), (4, 1)\}$ , i.e. with cells  $(3, 2)$  and  $(4, 1)$  missing.
- Under the  $QI$ -model the un-normalized cell probabilities  $q_{ij}$  are given by

$$q_{ij} = \rho_i \psi_j, \quad (i, j) \in A \quad QI - \text{model}$$

- If  $q_{ij} > 0$ ,  $(i, j) \in A$ ,

$$\log q_{i,j} = \alpha_i + \beta_j, \quad (i, j) \in A$$

with  $\alpha_i = \log \rho_i, \beta_j = \log \psi_j \in \mathbb{R}$ .

## Exponential and extended exponential model

- The design matrix  $M$  and an orthogonal matrix  $K$  are:

$$M = \begin{array}{c} \alpha_1 \quad \alpha_2 \quad \alpha_3 \quad \alpha_4 \quad \alpha_5 \quad \beta_1 \quad \beta_2 \\ \begin{array}{l} 11 \\ 21 \\ 31 \\ 51 \\ 12 \\ 22 \\ 42 \\ 52 \end{array} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 \end{bmatrix}, \end{array} \quad K = \begin{array}{c} k_1 \quad k_2 \\ \begin{array}{l} 11 \\ 21 \\ 31 \\ 51 \\ 12 \\ 22 \\ 42 \\ 52 \end{array} \begin{bmatrix} 1 & 0 \\ -1 & -1 \\ 0 & 0 \\ 0 & 1 \\ -1 & 0 \\ 1 & 1 \\ 0 & 0 \\ 0 & -1 \end{bmatrix}$$

- If  $q_{ij} > 0$ ,  $(i, j) \in A$ , the  $QI$ -model is equivalent to implicit binomial model

$$\begin{cases} q_{11}q_{22} - q_{21}q_{12} = 0 \\ q_{51}q_{22} - q_{21}q_{52} = 0. \end{cases}$$

The above equations are the standard conditions for independence in the two  $2 \times 2$  tables with rows  $\{1, 2\}$ , respectively  $\{2, 5\}$ . This is equivalent to the independence of the sub-table  $\{1, 2, 5\} \times \{1, 2\}$

## Maximal model

- The maximal design matrix  $M_{\max}$  and the maximal model in monomial form are computable with Computer Algebra Software.

$$M_{\max} = \begin{array}{c} 11 \\ 21 \\ 31 \\ 51 \\ 12 \\ 22 \\ 42 \\ 52 \end{array} \begin{array}{cccccccc} \zeta_1 & \zeta_2 & \zeta_3 & \zeta_4 & \zeta_5 & \zeta_6 & \zeta_7 \\ \left[ \begin{array}{cccccccc} 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 \end{array} \right] \end{array} \left\{ \begin{array}{l} q_{11} = \zeta_5 \zeta_7 \\ q_{21} = \zeta_3 \zeta_7 \\ q_{31} = \zeta_1 \\ q_{51} = \zeta_4 \zeta_7 \\ q_{12} = \zeta_5 \zeta_6 \\ q_{22} = \zeta_3 \zeta_6 \\ q_{42} = \zeta_2 \\ q_{52} = \zeta_4 \zeta_6 \end{array} \right.$$

- CoCoATeam. CoCoA: a system for doing Computations in Commutative Algebra. Available at <http://cocoa.dima.unige.it>.
- 4ti2 team. 4ti2 – a software package for algebraic, geometric and combinatorial problems on linear spaces. Available at [www.4ti2.de](http://www.4ti2.de).
- The number of **instances** (also called feasible sets) for the  $QI$ -model is 87.

## Bayesian analysis

- The model which imposes no restriction on the cell probabilities, save the zero-probability cells  $(3, 2)$  and  $(4, 1)$ , is called the Structural Zero model  $SZ$ . The number of  $SZ$ -instances is equal to  $2^8 - 1 = 255$ .
- Only two of the above  $SZ$ -instances are logically consistent with the observed data: the one giving a positive probability to all eight free cells; and the one giving zero-probability to cell  $(5, 1)$  only. We label these instances  $SZ_0$  and  $SZ_1$ .
- There exists only one logically consistent instance of the quasi-independence model, i.e. that having all positive cell-probabilities (except for the two cells corresponding to structural zeros), which we label  $QI_0$ .
- The models  $SZ_0$  and  $SZ_1$  are nested, both in the sense of the maximal parameterization and of the supports (faces), so that their a-priori Dirichlet distributions on the parameters should be related.  $ZI$  could be parameterized by  $QI_0$  and an orthogonal component.
- The two models  $SZ_0$  and  $SZ_1$  are different exponential models and an a-priori on their union disintegrates into a discrete a-priori for the model and a conditional Dirichlet given the model.

## Statistical models as differentiable manifolds

- Statistical models have a very rich mathematical structure. We can approach this from many points of view: Commutative Algebra, Convex Analysis, Differential Geometry.
- The Differential Geometry approach is the oldest: statistical models are Riemannian manifolds according: C. R. Rao. Information and accuracy attainable in the estimation of statistical parameters. *Bulletin of Calcutta Mathematical Society*, 37:81–89, 1945.
- More recently, it has been found that there are other manifold structures of interest, called by Amari  $\alpha$ -Bundles and  $\alpha$ -Connections.
- Most of the constructions in the literature are restricted to parametric statistical models. However, there are applications where infinite dimensional statistical models appear, e.g. in Mathematical Finance the set of martingale measures on the Wiener space.
- A general coordinate-free construction is important at the conceptual level, because in such a framework the “big picture” comes out more clearly, as in S. Lang. *Differential and Riemannian manifolds*, volume 160 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, third edition, 1995.

## Exponential statistical manifolds

The theory of **exponential statistical manifolds modeled on Orlicz spaces**, has been developed in a joint work with C. Sempì, M.-P. Rogantin, P. Gibilisco, A. Cena, D. Imparato, B. Trivellato (1993-2007).

- G. Pistone and C. Sempì. An infinite-dimensional geometric structure on the space of all the probability measures equivalent to a given one. *Ann. Statist.*, 23(5):1543–1561, October 1995;
- G. Pistone and M. P. Rogantin. The exponential statistical manifold: mean parameters, orthogonality and space transformations. *Bernoulli*, 5(4):721–760, August 1999;
- P. Gibilisco and G. Pistone. Connections on non-parametric statistical manifolds by Orlicz space geometry. *IDAQP*, 1(2):325–347, 1998;
- A. Cena and G. Pistone. Exponential statistical manifold. *AIMS*, 59:27–56, 2007. DOI 10.1007/s10463-006-0096-y.

**The main focus is in finding a rigorous functional setting for the IG as described in**

- S. Amari and H. Nagaoka. *Methods of information geometry*. American Mathematical Society, Providence, RI, 2000. Translated from the 1993 Japanese original by Daishi Harada.



## Statistical varieties

**In the case of a finite state space or in the case of special distributions, IG has an interesting interface with AS.** This interface is both conceptually interesting and computationally useful.

- P. Diaconis and B. Sturmfels. Algebraic algorithms for sampling from conditional distributions. *Ann. Statist.*, 26(1):363–397, 1998.
- G. Pistone and H. P. Wynn. Finitely generated cumulants. *Statist. Sinica*, 9(4):1029–1052, October 1999.
- G. Pistone, E. Riccomagno, and H. P. Wynn. *Algebraic Statistics: Computational Commutative Algebra in Statistics*. Chapman&Hall, 2001.
- D. Geiger, C. Meek, and B. Sturmfels. On the toric algebra of graphical models. *Ann. Statist.*, 34:1463–1492, 2006.

**The key notion is that of exponential statistical model which has long been known to have special features, both analytic and algebraic.**

The study of general infinite dimensional exponential models suggests how to avoid unnatural parameterization of the finite state space models. In turn, the finite state space case suggests how to deal with limit cases.

# Outline

## Exponential statistical model

- Basic construction.
- Exponential and mixture connections.
- Models, divergence, tangent bundle.

## Finite state space or special distributions

## Extended exponential model

## Tables with zero cells

## IG as a special Banach manifold

Given a general probability space  $(X, \mathcal{X}, \mu)$ ,

- $\mathcal{M}_>$  is the set of all densities which are positive  $\mu$ -a.s.
- $\mathcal{M}_>$  is thought to be the “maximal” regular statistical model.
- We want to assign a manifold structure to this maximal model in such a way that each specific statistical model could be considered as a “**sub-manifold**” of  $\mathcal{M}_>$ . We will see that the notion of sub-manifold to be used is not obvious.
- The model space for the manifold, locally at each  $p \in \mathcal{M}_>$ , is a subspace of **centered random variables** of a suitable **Orlicz space**. Orlicz spaces are constructed in analogy with Lebesgue spaces, by imposing conditions of the form  $E_p[\Phi(u)] = \int \Phi(u)p d\mu < +\infty$ , for a suitable function  $\Phi$ . If  $\Phi(x) = |x|^a$ , then the Lebesgue  $L^a$  spaces are obtained.

A reference on Orlicz spaces is:

- M. M. Rao and Z. D. Ren. *Applications of Orlicz spaces*, volume 250 of *Monographs and Textbooks in Pure and Applied Mathematics*. Marcel Dekker Inc., New York, 2002.

## Orlicz spaces

- The Jung function  $\Phi(x) = \cosh x - 1$  is used instead of the equivalent and more commonly used  $e^{|x|} - |x| - 1$ .
- $\Psi$  denotes its conjugate Jung function or the equivalent  $(1 + |y|) \log(1 + |y|) - |y|$ .
- A random variable  $u$  belongs to the vector space  $L^\Phi(p)$  if for some  $\alpha > 0$   $E_p[\Phi(\alpha u)] < +\infty$ .
- The closed unit ball of  $L^\Phi(p)$  consists of all  $u$ 's such that  $E_p[\Phi(u)] \leq 1$ .
- The open unit ball  $B(0, 1)$  consists of those  $u$ 's such that  $\alpha u$  is in the closed unit ball for some  $\alpha > 1$ .
- The Banach space  $L^\Phi(p)$  is not separable, unless the sample space is finite. In this sense it is an un-natural choice.
- However,  **$L^\Phi(p)$  is natural for statistics because for each  $u \in L^\Phi(p)$  the Laplace transform of  $u$  is well defined at 0 and the one-dimensional exponential model  $p(\theta) \propto e^{\theta u}$  is well defined around 0 (and viceversa).**
- The space  $L^\Psi(p)$  is separable and it is the pre-dual of  $L^\Phi(p)$ , with pairing  $E_p[uv]$ . For  $1 < a < +\infty$ ,  $L^\Phi(p) \subset L^a(p) \subset L^\Psi(p)$ .

## Moment functional

- For each  $p \in \mathcal{M}_>$ , the moment functional is  $M_p(u) = \mathbb{E}_p [e^u]$ .
- $M_p(0) = 1$ ; otherwise, for each  $u \neq 0$ ,  $M_p(u) > 1$ .
- $M_p$  is convex and lower semi-continuous, and its proper domain

$$\text{dom}(M_p) = \{u \in L^\Phi(p \cdot \mu) : M_p(u) < \infty\}$$

is a convex set which contains the open unit ball  $B_p(0, 1)$  of  $L^\Phi(p)$ .

**Th**  $M_p$  is infinitely Gâteaux-differentiable in the interior of its proper domain, the  $n$ th-derivative at  $u \in \text{dom}(M_p)$  in the direction  $v \in L^\Phi(p)$  being

$$\left. \frac{d^n}{dt^n} M_p(u + tv) \right|_{t=0} = \mathbb{E}_p [v^n e^u];$$

**Th**  $M_p$  is bounded, infinitely Fréchet-differentiable and analytic on the open unit ball of  $L^\Phi(p)$ , the  $n$ th-derivative at  $u \in B(0, 1)$  evaluated in  $(v_1, \dots, v_n) \in L^\Phi(p) \times \dots \times L^\Phi(p)$  is

$$D^n M_p(u)(v_1, \dots, v_n) = \mathbb{E}_p [v_1 \cdots v_n e^u].$$

In particular,

$$D^n M_p(0)(v_1, \dots, v_n) = \mathbb{E}_p [v_1, \dots, v_n].$$

## Connected component

- We associate to each density  $p$  a space of  $p$ -centered random variables: scores, estimating functions . . . . It is technically crucial to discuss how the relevant spaces depend on the density  $p$ .
- Given  $p, q \in \mathcal{M}_>$ , the exponential model  $p(\theta) \propto p^{1-\theta}q^\theta$ ,  $0 \leq \theta \leq 1$  connects the two given densities as end points of a curve. **This curve need not to be continuous.** So we ask for more.

**D** We say that  $p, q \in \mathcal{M}_>$  are **connected by an open exponential arc** if there exist  $r \in \mathcal{M}_>$ ,  $u \in L_0^\Phi(r)$  and an open interval  $I$  that contains 0, and such that  $p(t) \propto e^{tu} \cdot r$ ,  $t \in I$ , is an exponential model containing both  $p$  and  $q$ .

**Th** **Let  $p$  and  $q$  be densities connected by an open exponential arc. Then the Banach spaces  $L^\Phi(p)$  and  $L^\Phi(q)$  are equal as vector spaces and their norms are equivalent.**

**Th** For all  $q$  that are connected to  $p$  by an open exponential arc, the Orlicz space of centered random variables at  $q$  is

$$L_0^\Phi(q) = \left\{ u \in L^\Phi(p) \mid \mathbb{E}_p \left[ \frac{q}{p} u \right] = 0 \right\}$$

then  $\mathbb{E}_p [{}^*k u] = 0$ ,  ${}^*k = \left( \frac{q}{p} - 1 \right)$ ,  $u \in L^\Phi(p)$ .

## Cumulant functional

- For each  $p \in \mathcal{M}_>$  and the  $u$  in the set  $\mathcal{S}_p$  of  $p$ -centered random variables in the interior of the proper domain of  $M_p$ , the cumulant functional is  $K_p(u) = \log M_p(u)$ .
- $K_p$  is infinitely Gâteaux-differentiable.
- $K_p$  is bounded, infinitely Fréchet-differentiable and analytic on the open unit ball of  $L_0^\Phi(p)$ .
- The mapping

$$e_p : \begin{cases} \mathcal{S}_p \rightarrow \mathcal{M}_> \\ u \mapsto e^{u-K_p(u)} p \end{cases}$$

is a parameterization of a subset of  $\mathcal{M}_>$ .

- If  $e_p(\mathcal{S}_p) = \mathcal{E}(p)$ , the corresponding **chart** is

$$s_p : \mathcal{E}(p) \ni q \mapsto \log \left( \frac{q}{p} \right) - \mathbb{E}_p \left[ \log \left( \frac{q}{p} \right) \right] \in \mathcal{S}_p.$$

- **If  $q = e_p(u)$ , then  $K_p(u) = \mathbb{E}_p \left[ \log \left( \frac{p}{q} \right) \right]$  and**

$$\begin{aligned} DK_p(u)(v) &= \mathbb{E}_q[v] \\ D^2K_p(u)(v, w) &= \mathbb{E}_q[vw] \end{aligned}$$

## Exponential manifold

**D** For every  $p \in \mathcal{M}$ , the **maximal exponential model at  $p$**  is defined to be the family of densities

$$\mathcal{E}(p) := \left\{ e^{u - K_p(u)} p : u \in \text{dom}^\circ K_p \right\} \subseteq \mathcal{M}_>.$$

**Th** **The maximal exponential model at  $p$ ,  $\mathcal{E}(p)$  is equal to the set of all densities  $q \in \mathcal{M}_>$  connected to  $p$  by an open exponential arc.**

**Th** **The collection of charts  $\{(\mathcal{E}_p, s_p) : p \in \mathcal{M}_>\}$  is an affine  $\mathcal{C}^\infty$  atlas on  $\mathcal{M}_>$ .** The transition maps are

$$s_{p_2} \circ e_{p_1} : u \mapsto u + \log \left( \frac{p_1}{p_2} \right) - E_{p_2} \left[ u + \log \left( \frac{p_1}{p_2} \right) \right]$$

- The chart domains are either equal or disjoint because they correspond to the connected components (in the sense of open arcs).
- $\mathcal{E}(p)$  is convex. reference measure.
- The derivative of the transition map  $s_{p_2} \circ e_{p_1}$  is

$$L_0^\phi(p_1) \ni u \mapsto u - E_{p_2}[u] \in L_0^\phi(p_2)$$

which is an isomorphism, because  $p_1$  and  $p_2$  are connected by an open exponential arc.



## Divergence

- For each non-negative density  $q$  and each  $p \in \mathcal{M}_>$ , the divergence is defined by  $D(q||p) = \mathbb{E}_q [\log q/p]$ .
- Restricting to the manifold  $\mathcal{M}_> \times \mathcal{M}_>$  and then to the proper domain, we can restrict to  $\mathcal{E}(p) \times \mathcal{E}(p)$  and compute the representative in that chart:

$$D(q_1||q_2) = \mathbb{E}_{q_1} [u_1 - u_2] - K_p(u_1) + K_p(u_2)$$

Then, the divergence, as a real function on  $\mathcal{E}(p) \times \mathcal{E}(p)$ , is infinitely Gâteaux-differentiable.

- Its partial derivatives are

$$D_1 D(q_1||q_2) \cdot v = \text{Cov}_{q_1} [u_1 - u_2, v]$$

$$D_2 D(q_1||q_2) \cdot v = \mathbb{E}_{q_2} [v] - \mathbb{E}_{q_1} [v]$$

- In particular we can write

$$D(q||p) = \mathbb{E}_p \left[ \left( \frac{q}{p} - 1 \right) u \right] - K_p(u)$$

which shows that  $D(q||p)$ , as a function of  $\left( \frac{q}{p} - 1 \right)$ , is the conjugate of the convex function  $K_p$ . Equivalently,

$$D(q||p) + D(p||q) = \mathbb{E}_p \left[ \left( \frac{q}{p} - 1 \right) u \right] = \mathbb{E}_q [u]$$

## The exponential geometry

The previous theory is intended to capture the essence and to generalize the idea of **curved exponential model** as defined by

- B. Efron. Defining the curvature of a statistical problem (with applications to second-order efficiency). *The Annals of Statistics*, 3:1189–1242, 1975. (with discussion)
- B. Efron. The geometry of exponential families. *Ann. Statist.*, 6(2):362–376, 1978;
- A. P. Dawid. Discussion of a paper by Bradley Efron. *The Annals of Statistics*, 3:1231–1234, 1975
- A. P. Dawid. Further comments on a paper by Bradley Efron. *The Annals of Statistics*, 5:1249, 1977

From the work of S.-I. Amari, we know that there is a second geometry on probabilities, whose geodesics are mixtures. This structure is a **connection** on a special vector bundle of the  $\mathcal{M}_>$ -manifold. A related manifold on the set  $\mathcal{M}_1$  of *normalized* random variables is defined by charts of the form

$$q \mapsto \frac{q}{p} - 1$$

Locally, the  $q$ 's will have finite KL-divergence  $D(q||p)$ .

## Mixture manifold 1

- We enlarge  $\mathcal{M}$  > considering the sets

$$\mathcal{M}_{\geq} = \left\{ p \in L^1(\mu) : p \geq 0, \int p d\mu = 1 \right\}$$

$$\mathcal{M}_1 = \left\{ p \in L^1(\mu) : \int p d\mu = 1 \right\}.$$

- For each  $p \in \mathcal{M}_{\geq}$ , we consider the subset of  $L_0^{\Psi}(p)$  defined by:

$${}^*\mathcal{E}(p) = \left\{ q \in \mathcal{P} : \frac{q}{p} \in L^{\Psi}(p) \right\}$$

On such a domain, we define the charts

$$\eta_p : \begin{cases} {}^*\mathcal{E}(p) \rightarrow L_0^{\Psi}(p) \\ q \mapsto \frac{q}{p} - 1 \end{cases}$$

and the associated parameterizations

$$L_0^{\Psi}(p) \ni u \mapsto (u + 1)p \in {}^*\mathcal{U}_p.$$

The collection of sets  $\{{}^*\mathcal{E}(p)\}_{p \in \mathcal{M}_1}$  is a covering of  $\mathcal{M}_1$ :

- If  $p \in \mathcal{M}$ , then  $\mathcal{E}(p) \subset {}^*\mathcal{E}(p)$ .
- If  $p_1, p_2 \in \mathcal{E}(p)$ , then  ${}^*\mathcal{E}(p_1) = {}^*\mathcal{E}(p_2)$ .

## Mixture manifold 2

**Th** The set of charts

$$\{(*\mathcal{E}(q), \eta_q) : q \in \mathcal{E}(p)\}$$

is an affine  $\mathcal{C}^\infty$ -atlas on  $*\mathcal{E}(p)$ , so it defines a manifold modeled on the Banach space  $L_0^\Psi(p)$ .

- For each pair  $p_1, p_2 \in \mathcal{E}(p)$  the transition map is

$$\eta_{p_2} \circ \eta_{p_1}^{-1} : \begin{cases} L_0^\psi p_1 \rightarrow L_0^\psi p_2 \\ u \mapsto u \frac{p_1}{p_2} + \frac{p_1}{p_2} - 1 \end{cases}$$

**Th** For each  $q \in \mathcal{M}_\geq$ , the divergence  $D(q||p)$  finite if and only if  $q \in *\mathcal{E}(p)$ :

$$D(q||p) = \mathbb{E}_p \left[ \frac{q}{p} \log \left( \frac{q}{p} \right) \right] < \infty \iff q \in *\mathcal{E}(p)$$

**Th** For each density  $p \in \mathcal{M}_\mu$ , the inclusion  $j : \mathcal{E}(p) \hookrightarrow *\mathcal{E}(p)$  is of class  $\mathcal{C}^\infty$ .

## Exponential models

- Let  $V$  be a subspace of  $L_0^\phi$ . We call exponential model based on  $V$  the “flat manifold”

$$\mathcal{E}_V(p) = \left\{ e^{u - K_p(u)} p \mid u \in V \cap \mathcal{S}_p \right\}$$

- Let

$$V^\perp = \left\{ k \in L_0^\Psi(p) \mid E_p[ku] = 0, u \in V \right\}$$

be the orthogonal space of  $V$ . If  $V$  is closed, then

$$q \in \mathcal{E}_V(p) \iff E_p \left[ k \log \frac{q}{p} \right] = 0, k \in V^\perp$$

- Note that  $V + V^\perp$  is not a splitting of the space  $L_0^\phi(p)$ . In fact, the proper notion of statistical model appears to be different from what is usually termed a sub-manifold, because, in general, there is no orthogonal splitting of subspaces in  $L^\Phi(p)$ . The proper splitting space consists of the orthogonal space of the space of the model in the pre-dual space  $L_0^\Psi(p)$ .
- If we take  $k = \frac{q}{p} - 1$ ,  $q \in^* E(p)$ , then the orthogonality becomes a special case of the Csiszar pitagorean theorem

$$D(q||p) + D(p||q) = 0$$

## Tangent space

- Let  $p(\theta)$ ,  $\theta \in I$  be a one-dimensional statistical model. If  $p(0) = p$  and  $p(\theta) \in \mathcal{E}(p)$ , then  $p(\theta) = e^{p(\theta) - K_p(p(\theta))} p$ . The tangent vector at  $\theta$  is computed in the exponential chart centered at  $p$  as

$$T_{p(\theta)}p(\cdot) = \frac{d}{d\theta}u(\theta) = \frac{d}{d\theta} \log \left( \frac{p(\theta)}{p} \right) + E_{p(\theta)} \left[ \frac{d}{d\theta}u(\theta) \right]$$

and, at  $\theta = 0$ ,

$$T_p p(\cdot) = \left. \frac{d}{d\theta}u(\theta) \right|_{\theta=0} = \left. \frac{d}{d\theta} \log \left( \frac{p(\theta)}{p} \right) \right|_{\theta=0}$$

- The exponential model

$$e^{\theta u'(0) - K_p(u'(0))} p, \theta \in I$$

is the **tangent exponential model**.

- The statistical model

$$\{q \in {}^* \mathcal{E}(p) \mid E_q[u'(0)] = 0\}$$

is the **orthogonal mixture model**.

## Classical exponential models

- A classical exponential model is

$$p(x; \theta) = \exp \left( \sum_{i=1}^d \theta_i T_i(x) - \Psi(\theta) \right), \quad \theta \in \Theta$$

where  $\Psi = \widehat{T^* \mu}$  is the Laplace transform of the distribution of  $T$  and  $\Theta$  is its domain.

- If  $H = \text{Span}(1, T_1, \dots, T_d)$ , the membership of  $p_\theta$  in the model is equivalent to

$$\log p(x; \theta) = \sum_{i=1}^d \theta_i T_i(x) - \Psi(\theta) \in H$$

- The coordinate in the exponential manifold is

$$s(p_\theta) = \sum_{i=1}^d \theta_i \left( T_i(x) - \frac{\partial}{\partial \theta_i} \Psi(\theta) \right)$$

and the cumulant function is

$$K(u) = \Psi(\theta) - \sum_{i=1}^d \frac{\partial}{\partial \theta_i} \Psi(\theta)$$

## Sufficiency

- The supporting probability space can be reduced by sufficiency. In fact, if  $\mathcal{T} = \sigma(T_1, \dots, T_d)$  is the sufficient  $\sigma$ -algebra, it suffices to the smaller probability space  $(T(X), \mathcal{T}, \mu|_{\mathcal{T}})$ .
- Assume the all the  $T_i$ 's take a finite number of values. Then, the space of measurable functions which are functions of  $T$  is the algebra generated by  $T_1, \dots, T_d$ . If a smaller number of generators can be found, this will lead to a reduction of the dimension of the model space.
- The computation of the sufficient algebra, of a generating set of this algebra, and of a linear basis, are all relevant, because all statistic based on the likelihood will be a member of this algebra.
- More specifically, assume that the original sample space is finite and that all values of the  $T_i$  are rational. Then, we can map each probability of the original model to a finite subset  $\mathcal{D}$  of  $\mathbb{Q}^d$  to obtain the equivalent exponential model

$$p(t; \theta) = \exp \left( \sum_{i=1}^d \theta_i t_i - \Psi(\theta) \right) T^* \mu(t), \quad t \in \mathcal{D}$$



## Design theory

- In algebraic Design Theory, the points of a finite state space (or treatment space) are labelled with points in  $\mathbb{Q}^d$ . In turn, the sample space is described as a 0-dimensional algebraic variety of an ideal  $I$  of the ring  $R = \mathbb{Q}[x_1 \dots x_d]$ . The algebra of real functions on the sample space is described by the quotient ring  $R/I$  and has a hierarchical linear basis of monomial. Other labels have proved to be useful, in particular the  $n$ -roots of the unity.
- This same approach should lead to a useful algebraic presentation of the sufficient algebra of exponential models on a finite sample space.
- G. Pistone, E. Riccomagno, and H. P. Wynn. *Algebraic Statistics: Computational Commutative Algebra in Statistics*. Chapman&Hall, 2001
- G. Pistone and M. Rogantin. Indicator function and complex coding for mixed fractional factorial designs (revised 2006). Technical Report 17, Dipartimento di Matematica, Politecnico di Torino, 2006
- H Maruri-Aguilar, R Notari and E Riccomagno. “On the description and identifiability analysis of mixture designs” (Accepted for publication in *Statistica Sinica* 2007)

## Example: Exponential versus toric

- $\mathcal{X}$  is a finite sample space and  $p$  is a given probability density. In particular, a multi-way contingency table identified by a collection of factors  $X = \{X_1, \dots, X_F\}$ . If  $\mathcal{I}_f$  denotes the set of levels for the factor  $X_f$ ,  $f = 1, \dots, F$ , the state space is a product space, i.e.  $\mathcal{X} = \times_{f=1}^F \mathcal{I}_f$ .
- A log-linear model assumes that  $p(x) > 0$  and that  $\log p(x)$  belongs to a linear subspace  $H$  of  $\mathbb{R}^{\mathcal{X}}$ . If  $H$  is spanned by  $\{T_1, \dots, T_s\}$ , where the  $T_j$ 's are integer valued functions, we can write the log-linear model as

$$\log p(x) = \sum_{j=1}^s (\log \zeta_j) T_j(x),$$

- The unnormalized probability  $q$  satisfies the same equation

$$q(x) = \zeta_1^{T_1(x)} \dots \zeta_s^{T_s(x)}, \quad \zeta_j \geq 0, \quad j = 0, \dots, s,$$

where the parameters  $\zeta_1, \dots, \zeta_s$  are subject to non-negativity constraint only.

## Example: orthogonality

- A third expression of the same model can be derived by elimination of the indeterminates  $\zeta_1, \dots, \zeta_s$  in the monomial parameterization equation. In fact, if  $M = [T_1(x) \ \cdots \ T_s(x)]_{x \in \mathcal{Q}}$  is the design matrix, the orthogonal space of its range can be generated by integer valued vectors with zero sum  $K = [k_1 \ \cdots \ k_r]$ , then

$$\prod_x q(x)^{k_j(x)} = \prod_x \left( \zeta_1^{T_1(x)} \cdots \zeta_s^{T_s(x)} \right)^{k_j(x)} = \zeta_1^{T_1(x) \cdot k_j(x)} \cdots \zeta_s^{T_s(x) \cdot k_j(x)} = 1,$$

- As the sum of the elements of each  $k_j$ ,  $j = 1, \dots, r$ , is zero, the sum of the elements of both the positive part  $k_j^+$  and the negative part  $k_j^-$  are equal, so that we could write

$$\prod_x q(x)^{k_j(x)^+} - \prod_x q(x)^{k_j(x)^-} = 0, \quad j = 1, \dots, r.$$

- It follows that the toric model implies a set of  $r$  binomial and homogeneous equations in the un-normalized probabilities  $q(x)$ ,  $x \in \mathcal{Q}$ .
  - Given a finite state log-linear model and all its limit points, a specific set of configurations of zero-probability cells arises. This set cannot be recovered by setting to zero some parameters in a generic toric parametric representation.

## Example: instances

- However, there exists a “maximal” parametric toric representation that gives a full parameterization of the extended model.
  1. All toric models compatible with the implicit binomial model are characterized by a string of  $T$ 's exponents, which is a non-negative integer vector orthogonal to the vectors  $[k_1 \dots k_r]$ .
  2. The lattice of non-negative integer vectors  $t \in \mathbb{N}_+^{\mathbb{Q}}$  such that the condition  $t \cdot k_j = 0$  holds for each  $j = 1, \dots, r$ , has a finite number of generators that can be computed with symbolic software.
  3. If the generators are  $S_1, \dots, S_u$ , then the “maximal” toric model is

$$q(x) = \zeta_1^{S_1(x)} \dots \zeta_u^{S_u(x)} \quad x \in \mathbb{Q}.$$

- The support of the resulting probability will be the set  $\mathbb{Q}_1 = \{x \in \mathbb{Q} : S_1(x) = 0\}$ . On such a restricted support, the model will be again toric:

$$q(x) = \zeta_2^{S_2(x)} \dots \zeta_u^{S_u(x)} \quad x \in \mathbb{Q}_1.$$

or exponential if all the other parameters  $\zeta_2, \dots, \zeta_u$  are assumed to be strictly positive.

- **In this sense, we say that each toric model is a union of exponential models with different supports. Each one of these models is called an instance of the model.**