ΣΤΑΤΙΣΤΙΚΗ ΦΥΣΙΚΗ
ΣΦ 2011
STATISTICAL PHYSICS

International Conference on
ΣΤΑΤΙΣΤΙΚΗ ΦΥΣΙΚΗ
Larnaca - Cyprus 11-15 July 2011

Algebraic Statistics and Information Geometry of Reversible Markov Chains

Giovanni Pistone and Maria Piera Rogantin

Collegio Carlo Alberto

UNIVERSITÀ DEGLI STUDI DI GENOVA

July 13 2011

Content

Part 1 Introduction to Algebraic Statistics: A-models, binomial ideals, Commutative Computer Algebra.

Part 2 Algebra of reversible Markov chains: Gröbner and Graver bases of the Kolmogorov's ideal.

Part 3 Algebra and geometry of Markov chains: work in progress.

PART ONE

A-model

- $\mathcal{X}$ is a finite sample space with reference measure μ.
- A is an integer matrix $A \in \mathbb{Z}_{\geq}^{m+1,\mathcal{X}}$.
- The elements of matrix A are $A_i(x)$, $i = 0 \ldots m, x \in \mathcal{X}$. We assume the row A_0 to be the constant 1.
- The x-column of A, say $A(x)$, is the multi-exponent of a monomial term denoted
$$t^{A(x)} = t_0 t_1^{A_1(x)} \cdots t_m^{A_m(x)}$$
- Matrix A defines a statistical model on $(\mathcal{X}, \mu)$ whose unnormalized probability densities are
$$q(x;t) = t^{A(x)}, \quad x \in \mathcal{X},$$
for all $t \in \mathbb{R}_{\geq}^{m+1}$ such that $q(\cdot, t)$ is not identically zero.
- The probability densities wrt μ in the A-model are
$$p(x;t) = q(x;t)/Z(t), \quad Z(t) = \sum_{x \in \mathcal{X}} q(x;t)\mu(x).$$
- If $t > 0$, $t = \log \theta$ and $q(x;\theta) = \exp(\theta \cdot A(x))$.

Example of A-model

- The simplest example is the Binomial(n, p):

$$\mathcal{X} = \{0, 1, 2, 3, \ldots, n\}, \quad \mu(x) = \binom{n}{x},$$

$$A = \begin{array}{cc} & \begin{array}{cccccc} 0 & 1 & 2 & 3 & \cdots & n \end{array} \\ \begin{array}{c} 0 \\ 1 \end{array} & \begin{bmatrix} 1 & 1 & 1 & 1 & \cdots & 1 \\ 0 & 1 & 2 & 3 & \cdots & n \end{bmatrix} \end{array}, \quad q(x; t_0, t_1) = t_0 t_1^x,$$

$$p(x; t) = \frac{t^x}{\sum_{x=0}^{n} t^x \binom{n}{x}} = \frac{t^x}{(1+t)^n}, \quad x \in \mathcal{X}, t \geq 0.$$

- Persi Diaconis and Bernd Sturmfels. Algebraic algorithms for sampling from conditional distributions. Ann. Statist., 26(1):363–397, 1998. ISSN 0090-5364;
- Giovanni Pistone, Eva Riccomagno, and Henry P. Wynn. Algebraic statistics, volume 89 of Monographs on Statistics and Applied Probability. Chapman & Hall/CRC, Boca Raton, FL, 2001. ISBN 1-58488-204-2. Computational commutative algebra in statistics;
- Dan Geiger, Christopher Meek, and Bernd Sturmfels. On the toric algebra of graphical models. Ann. Statist., 34(3):1463–1492, 2006. ISSN 0090-5364;
- Paolo Gibilisco, Eva Riccomagno, Maria Piera Rogantin, and Henry P. Wynn, editors. Algebraic and geometric methods in statistics. Cambridge University Press, Cambridge, 2010. ISBN 978-0-521-89619-1.

C-constrained A-model; identification

- In some applications the statistical model is further constrained by a matrix $C \in \mathbb{Z}^{k,n}$.

$$\begin{cases} q(x;t) & = & t^{A(x)}, \\ \sum_{x \in \mathcal{X}} C_i(x) q(x;t) & = & 0, \end{cases}$$

for $\quad x \in \mathcal{X}, t \in \mathbb{R}_{\geq}^{m+1}, i = 1 \dots k.$

- Assume $s, t \in \mathbb{R}_{>}^{m}$ and $p_s = p_t$. Denote by Z the normalizing constant. Then $p_t = p_s$ if, and only if,

$$Z(s) t^{A(x)} = Z(t) s^{A(x)}, \quad x \in \mathcal{X}$$

hence

$$\sum_{i=0}^{m} (\log t_i - \log s_i) A_i(x) = \log Z(t) - \log Z(s), \quad x \in \mathcal{X}.$$

The confounding condition is

$$\delta^T A = 1, \quad \delta_i = (\log t_i - \log s_i)/(\log Z(t) - \log Z(s)),$$

so that $\delta \in e_0 + \ker A$.

Toric ideals; closure of the A-model

- The ker of the ring homomorphism

$$k[q(x) \colon x \in \mathcal{X}] \ni q(x) \mapsto t^{A(x)} \in k[t_0, \dots, t_m]$$

is the toric ideal of A, $\mathrm{I}(A)$. It has a finite basis made of binomials of the form

$$\prod_{x \colon u(x)>0} q(x)^{u^+(x)} - \prod_{x \colon u(x)<0} q(x)^{u^-(x)}$$

with $u \in \mathbb{Z}^{\mathcal{X}}$, $Au = 0$.

- As $\sum_{x \in \mathcal{X}} u(x) = 0$, all the binomials are homogeneous polynomials so that all densities p_t in the A-model satisfy the same binomial equation.

Theorems

- The nonnegative part of the A-variety is the (weak) closure of the positive part of the A-model
- Let $\mathcal{H}$ be the Hilbert basis of $\mathrm{Span}\,(A_0, A_1, \dots) \cap Z_{\geq}^{\mathcal{X}}$. Let H be the matrix whose rows are the elements of $\mathcal{H}$ of minimal support. The H-model is equal to the nonnegative part of the A-variety

The binomial example

- The integer kernel of $A = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 & 5 \end{bmatrix}$ is $\mathbb{Q}$-generated by the rows of $K = \begin{bmatrix} 1 & -2 & 1 & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 & 0 \\ 0 & 0 & 1 & -2 & 1 & 0 \\ 0 & 0 & 0 & 1 & -2 & 1 \end{bmatrix}$ and the relevant binomials are

  $$q(0)q(2)-q(1)^2, q(1)q(3)-q(2)^2, q(2)q(4)-q(3)^2, q(3)q(5)-q(4)^2.$$

- The Hilbert basis of RowSpan A is

  $$\mathcal{H} = \begin{bmatrix} 0 & 1 & 2 & 3 & 4 & 5 \\ 5 & 4 & 3 & 2 & 1 & 0 \end{bmatrix} \quad \text{hence } q(x; t_1, t_2) = t_1^x t_2^{5-x}.$$

- Bernd Sturmfels. Gröbner bases and convex polytopes. American Mathematical Society, Providence, RI, 1996. ISBN 0-8218-0487-1;
- Dan Geiger, Christopher Meek, and Bernd Sturmfels. On the toric algebra of graphical models. Ann. Statist., 34(3):1463–1492, 2006. ISSN 0090-5364;
- Luigi Malagò and Giovanni Pistone. A note on the border of an exponential family. arXiv:1012.0637v1, 2010;
- J. Rauh, T. Kahle, and N. Ay. Support sets in exponential families and oriented matroid theory. Proceedings WUPES'09, invited for special issue of IJAR arXiv:0906.5462, 2009.

Example: 3 binary identical RVs, no-3-way-interaction

- $\mathcal{X} = \{+, -\}^3$. Matrix A is

	+++	-++	+-+	--+	++-	-+-	+--	---
I	1	1	1	1	1	1	1	1
1	0	1	0	1	0	1	0	1
2	0	0	1	1	0	0	1	1
3	0	0	0	0	1	1	1	1
12	0	1	1	0	0	1	1	0
13	0	1	0	1	1	0	1	0
23	0	0	1	1	1	1	0	0

- Constrain matrix C is

	+++	-++	+-+	--+	++-	-+-	+--	---
1=2	0	1	-1	0	0	1	-1	0
1=3	0	1	0	1	-1	0	-1	0

- The toric ideal $\mathrm{I}(A)$ is generated by

$$q(+++)q(--+)q(-+-)q(+--) \\ - q(-++)q(+-+)q(++-)q(---)$$

- $H = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \end{bmatrix}$ has 16 rows.
- The complete H-parameterization is quadratic.

Toric and Weyl

- Consider the design $\mathcal{D} \subset \mathbb{Z}_+^d$ with reference measure μ. Let $I(\mathcal{D})$ be its ideal of points. Consider the statistical model

$$q(x;t) = \prod_{i=1}^{d} t_i^{x_i}, \quad x \in \mathcal{D}, \quad t_j \geq 0, \quad j = 1, \ldots, d,$$

with normalizing constant

$$Z(t) = \sum_{x \in \mathcal{D}} t^x \mu(x)$$

It is the A-model with $A_i(x) = x_i$, $i = 1, \ldots, m$.

- In the Weyl algebra $\mathbb{C}\langle t_1 \ldots t_d, \partial_1 \ldots \partial_d \rangle$ define the operators

$$A(i,x) = t_i \partial_i - x_i = \partial_i t_i - (1 + x_i), \quad i = 1, \ldots, d, \quad x \in \mathcal{D},$$

where the second equality follows from the commutation relation $\partial_i t_i = 1 + t_i \partial_i$. For all $x \in \mathcal{D}$ we have

$$A(i,x) \bullet t^x = \partial_i \bullet (t_i t^x) - (1 + x_i) t^x = 0,$$

so that $t_i \partial_i \bullet t^x = x_i t^x$ and, by iteration, $(t_i \partial_i)^\alpha \bullet t^x = x_i^\alpha t^x$, $\alpha \in \mathbb{N}$.

- The operator $(t_i\partial_i)^\alpha$ applied to the polynomial $Z(t) \in \mathbb{C}[t_1, \dots, t_m]$ gives

$$(t_i\partial_i)^\alpha \bullet Z(t) = \sum_{x \in \mathcal{D}} (t_i\partial_i)^\alpha \bullet t^x = \sum_{x \in \mathcal{D}} x_i^\alpha t^x \quad \boxed{\mu(x) = 1}.$$

- Note the commutativity

$$(t_i\partial_i)(t_j\partial_j) = (t_j\partial_j)(t_i\partial_i),$$

hence

$$\prod_{i=1}^d (t_i\partial_i)^{\alpha_i} \bullet Z(t) = \sum_{x \in \mathcal{D}} \prod_{i=1}^d (t_i\partial_i)^{\alpha_i} \bullet t^x = \sum_{x \in \mathcal{D}} \left(\prod_{i=1}^d x_i^{\alpha_i} \right) t^x.$$

- By dividing by the normalizing constant we obtain he following expression for the moments:

$$Z(t)^{-1} \prod_{i=1}^d (t_i\partial_i)^{\alpha_i} \bullet Z(t) = Z(t)^{-1} \sum_{x \in \mathcal{D}} \prod_{i=1}^d (t_i\partial_i)^{\alpha_i} \bullet t^x = \mathrm{E}_t \left[X^\alpha \right].$$

From the ring homomorphism $A\colon \begin{cases} \mathbb{C}[x] & \to & \mathbb{C}\langle t_1 \dots t_d, \partial_1 \dots \partial_d \rangle \\ x_i & \mapsto & t_i \partial_i \end{cases}$

we have

$$A(f(x)) \bullet Z(t) = \sum_{x \in \mathcal{D}} f(x) t^x.$$

Theorem

1. *Let x^α, $\alpha \in M$, be a monomial basis for $\mathcal{D}$. Then $Z(t)$ satisfies the following system of $\#M = \#\mathcal{D}$ linear non-homogeneous differential equations:*
$$A(x^\alpha) \bullet Z(t) = \sum_{x \in \mathcal{D}} x^\alpha t^x, \quad \alpha \in M.$$

2. *Let $f_a(x)$ be the (reduced) indicator polynomial of $a \in \mathcal{D}$. Then $Z(t)$ satisfies the following system of $\#\mathcal{D}$ linear non-homogeneous differential equations:*
$$A(f_a(x)) \bullet Z(t) = t^a, \quad a \in \mathcal{D}$$

3. *Let $g(p_a \colon a \in \mathcal{D})$ be a polynomial in the toric ideal of the monomial homomorphism $p_a \mapsto t^a$. Then*
$$g\left(A(f_a(x)) \bullet Z(t) \colon a \in \mathcal{D}\right) = 0$$

PART TWO

Detailed balance

- A transition matrix $P_{v \to w}$, $v, w \in V$, satisfies the detailed balance conditions if $\kappa(v) > 0$, $v \in V$, and

$$\kappa(v) P_{v \to w} = \kappa(w) P_{w \to v}, \quad v, w \in V.$$

- It follows that $\pi(v) \propto \kappa(v)$ is an invariant probability and the Markov chain $(X_n)_{n=0,1,\ldots}$, has symmetric and stationary two-step joint distributions under π.
- Consider the simple (no-loop) directed graph $(V, \mathcal{A})$ s.t. $(v \to w) \in \mathcal{A}$ if, and only if, $v \neq w$ and $P_{v \to w} > 0$.
- The binomials

$$\kappa(v) P_{v \to w} - \kappa(w) P_{w \to v}, \quad (v \to w) \in \mathcal{A},$$

define a binomial ideal of the ring

$$\mathbb{Q}[k(v) \colon v \in V; P_{v \to w} \colon (v \to w) \in \mathcal{A}]$$

Reversibility on trajectories

- Let $\omega = v_0 \cdots v_n$ be a trajectory (path) in the connected graph $\mathcal{A}$ and let $r(\omega) = v_n \cdots v_0$ be the reversed trajectory.
- If the detailed balance holds, then the reversibility condition

$$\mathrm{P}(\omega) = \mathrm{P}(r\omega)$$

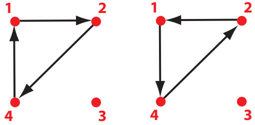
  holds for each trajectory ω.
- Write the detailed balance along the trajectory,

$$\begin{aligned}
\pi(v_0)P_{v_0 \to v_1} &= \pi(v_1)P_{v_1 \to v_0}, \\
\pi(v_1)P_{v_1 \to v_2} &= \pi(v_2)P_{v_2 \to v_1}, \\
&\vdots \\
\pi(v_{n-1})P_{v_{n-1} \to v_n} &= \pi(v_n)P_{v_n \to v_{n-1}},
\end{aligned}$$

  and clear $\pi(v_1) \cdots \pi(v_{n-1})$ in both sides of the product.
- Note the binomial structure.

Kolmogorov's condition

We denote by ω a closed trajectory, that is a trajectory on the graph such that the last state coincides with the first one, $\omega = v_0 v_1 \dots v_n v_0$, and by $r\omega$ the reversed trajectory $r\omega = v_0 v_n \dots v_1 v_0$



Theorem (Kolmogorov)

Let the Markov chain $(X_n)_{n=0,1,\dots}$ have transitions the connected graph $\mathcal{G}$.

- *If the process is reversible, for all closed trajectory*

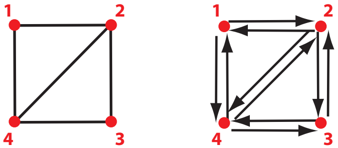
$$\boxed{P_{v_0 \to v_1} \cdots P_{v_n \to v_0} = P_{v_0 \to v_n} \cdots P_{v_1 \to v_0}}$$

- *If the equality is true for all closed trajectory, then the process is reversible.*

- Detailed balance, reversibility, Kolmogorov's condition are algebraic in nature and define binomial ideals. The Kolmogorov's condition does not involve the invariant probability π.
- P. Suomela. Invariant measures of time-reversible Markov chains. J. Appl. Probab., 16(1):226–229, 1979. ISSN 0021-9002.

Transition graph

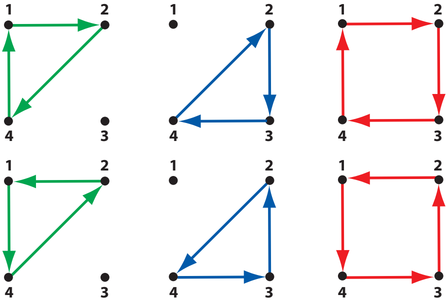
- From $\mathcal{G} = (V, \mathcal{E})$ an (undirected simple) graph, split each edge into two opposite arcs to get a connected directed graph (without loops) $\mathcal{O} = (V, \mathcal{A})$. The arc going from vertex v to vertex w is $(v \to w)$. The reversed arc is $r(v \to w) = (w \to v)$.



- A path or trajectory is a sequence of vertices $\omega = v_0 v_1 \cdots v_n$ with $(v_{k-1} \to v_k) \in \mathcal{A}$, $k = 1, \ldots, n$. The reversed path is $r\omega = v_n v_{n-1} \cdots v_0$. Equivalently, a path is a sequence of inter-connected arcs $\omega = a_1 \ldots a_n$, $a_k = (v_{k-1} \to v_k)$, and $r\omega = r(a_n) \ldots r(a_1)$.

Circuits, cycles

- A closed path $\omega = v_0 v_1 \cdots v_{n-1} v_0$ is any path going from an initial v_0 back to v_0; $r\omega = v_0 v_{n-1} \cdots v_1 v_0$ is the reversed closed path. If we do not distinguish any initial vertex, the equivalence class of closed paths is called a circuit.
- A closed path is elementary if it has no proper closed sub-path, i.e. if does not meet twice the same vertex except the initial one v_0. The circuit of an elementary closed path is a cycle.



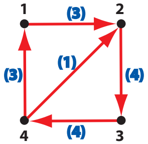
Kolmogorov's ideal

- With indeterminates $P = [P_{v \to w}]$, $(v \to w) \in \mathcal{A}$, form the ring $k[P_{v \to w} : (v \to w) \in \mathcal{A}]$. For a trajectory ω, define the monomial term

$$\omega = a_1 \cdots a_n \mapsto P^{\omega} = \prod_{k=1}^{n} P_{a_k} = \prod_{a \in \mathcal{A}} P_a^{N_a(\omega)},$$

with $N_a(\omega)$ the number of traversals of the arc a by the trajectory.

2→3 3→4 4→1 1→2
2→3 3→4 4→2
2→3 3→4 4→1 1→2
2→3 3→4 4→1 1→2



$P^3_{1\to2}\, P^4_{2\to3}\, P^4_{3\to4}\, P^3_{4\to1}\, P_{4\to2}$

Definition (K-ideal)

The Kolmogorov's ideal or K-ideal of the graph $\mathcal{G}$ is the ideal generated by the binomials $P^{\omega} - P^{r\omega}$, where ω is any circuit.

Bases of the K-ideal

- The binomials $P^{\omega} - P^{r\omega}$, where ω is any cycle, form a reduced universal Gröbner basis of the K-ideal.
- The K-ideal is generated by a finite set of binomials. A Gröbner basis is a special class of generating set of an ideal.
- Gröbner basis theory: David Cox, John Little, and Donal O'Shea. Ideals, varieties, and algorithms: An introduction to computational algebraic geometry and commutative algebra. Undergraduate Texts in Mathematics. Springer-Verlag, New York, second edition, 1997. ISBN 0-387-94680-2, Martin Kreuzer and Lorenzo Robbiano. Computational commutative algebra. 1. Springer-Verlag, Berlin, 2000. ISBN 3-540-67733-X.
- The theory is based on the existence of a monomial order, i.e. a total order on monomial term which is compatible with the product. Given such an order, the leading term $\mathrm{LT}(f)$ of the polynomial f is defined. A generating set is a Gröbner basis if the set of leading terms of the ideal is generated by the leading terms of monomials in the generating set.
- A Gröbner basis is reduced if the coefficient of the leading term of each element of the basis is 1 and no monomial in any element of the basis is in the ideal generated by the leading terms of the other element of the basis. The Gröbner basis property depend on the monomial order. However, a generating set is a universal Gröbner basis if it is a Gröbner basis for all monomial orders.

Gröbner basis

- The finite algorithm for computing a Gröbner basis depends on the definition of sygyzy. Given two polynomial f and g in the polynomial ring K, their sygyzy is the polynomial

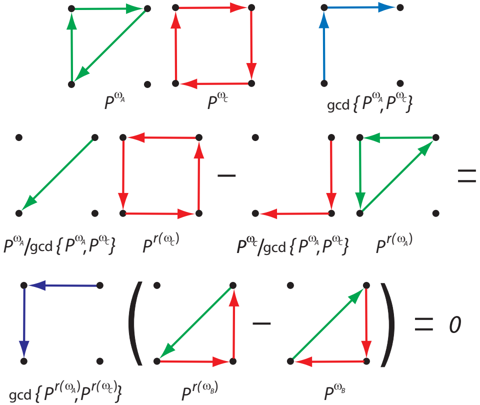
$$S(f,g) = \frac{\mathrm{LT}(g)}{\gcd(\mathrm{LT}(f), \mathrm{LT}(g))} f - \frac{\mathrm{LT}(f)}{\gcd(\mathrm{LT}(f), \mathrm{LT}(g))} g.$$

- A generating set of an ideal is a Gröbner basis if, and only if, it contains the sygyzy $S(f,g)$ whenever it contains f and g, see

- Chapter 6 in David Cox, John Little, and Donal O'Shea. Ideals, varieties, and algorithms: An introduction to computational algebraic geometry and commutative algebra. Undergraduate Texts in Mathematics. Springer-Verlag, New York, second edition, 1997. ISBN 0-387-94680-2, or
- Theorem 2.4.1 p. 111 of Martin Kreuzer and Lorenzo Robbiano. Computational commutative algebra. 1. Springer-Verlag, Berlin, 2000. ISBN 3-540-67733-X.
- Software for Gröbner bases: CoCoATeam. CoCoA: a system for doing Computations in Commutative Algebra. Available at cocoa.dima.unige.it, online.
- Software for Hilbert bases: 4ti2 team. 4ti2—a software package for algebraic, geometric and combinatorial problems on linear spaces. Available at www.4ti2.de.
- Giovanni Pistone and Maria Piera Rogantin. The algebra of reversible markov chains. arXiv:1007.4282v2 [math.ST], 2011

Proof (square with 1 diagonal)

Six cycles: $\omega_A = 1 \to 2 \to 4 \to 1$, $\omega_B = 2 \to 3 \to 4 \to 2$, $\omega_C = 1 \to 2 \to 3 \to 4 \to 1$, and the reversed arcs.



- In blue we have represented the common part of ω_A and ω_B. $t_i = P^{\omega_i}$, $r(t_i) = P^{r\omega_i}$, $i = A, B, C$.

Cycle space

- For each cycle ω define cycle vector

$$z_a(\omega) = \begin{cases} +1 & \text{if } a \text{ is an arc of } \omega, \\ -1 & \text{if } r(a) \text{ is an arc of } \omega, \\ 0 & \text{otherwise.} \end{cases} \qquad a \in \mathcal{A}.$$

	$(1 \to 2)$	$(2 \to 1)$	$(2 \to 3)$	$(3 \to 2)$	$(3 \to 4)$	$(4 \to 3)$	$(4 \to 1)$	$(1 \to 4)$	$(2 \to 4)$	$(4 \to 2)$
$z(\boldsymbol{\omega_A})$	1	-1	0	0	0	0	1	-1	1	-1
$z(\boldsymbol{\omega_B})$	0	0	1	-1	1	-1	0	0	-1	1
$z(\boldsymbol{\omega_C})$	1	-1	1	-1	1	-1	1	-1	0	0

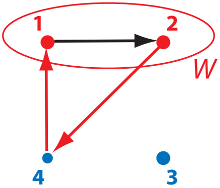
- We can find nonnegative integers $\lambda(\omega)$ such that $z(\bar{\omega}) = \sum_{\omega \in \mathcal{C}} \lambda(\omega) z(\omega)$, i.e. it belongs to the integer lattice generated by the cycle vectors.

- $Z(\mathcal{O})$ is the cycle space, i.e. the vector space generated by the cycle vectors.

Cocycle space of $\mathcal{O}$

- For each subset W of V, define cocycle vector

$$u_a(W) = \begin{cases} +1 & \text{if } a \text{ exits from } W, \\ -1 & \text{if } a \text{ enters into } W, \\ 0 & \text{otherwise.} \end{cases} \qquad a \in \mathcal{A}.$$



- The generated subspace of $k^{\mathcal{A}}$ is the cocycle space $U(\mathcal{O})$
- The cycle space and the cocycle space are orthogonal to each other. In fact for each cycle vector $z(\omega)$, cocycle vector $u(W)$, $z_a(\omega)u_a(W) = z_{r(a)}(\omega)u_{r(a)}(W)$, $a \in \mathcal{A}$, therefore

$$z(\omega)\cdot u(W) = 2\sum_{a\in\omega} u_a(W) = 2\left[\sum_{a\in\omega, u_a(W)=+1} 1 - \sum_{a\in\omega, u_a(W)=-1} 1\right] = 0.$$

- It is an orthogonal split the full vector space.

Toric ideals

- The matrix $U = [u_a(W)]_{W \subset V, a \in \mathcal{A}}$ whose rows are the cocycle vectors is the cocycle matrix.
- Consider the ring $k[P_a \colon a \in \mathcal{A}]$ and the Laurent ring $k(t_W \colon W \subset V)$, together with their homomorphism h defined by

$$h \colon P_a \longmapsto \prod_{W \subset V} t_W^{u_a(W)} = t^{u_a}.$$

- The kernel $I(U)$ of h is the toric ideal of U. It is a prime ideal and the binomials $P^{z^+} - P^{z^-}$, $z \in \mathbb{Z}^{\mathcal{A}}$, $Uz = 0$ are a generating set of $I(U)$ as a k-vector space.
- As for each cycle ω we have $Uz(\omega) = 0$, the cycle vector $z(\omega)$ belongs to $\ker_{\mathbb{Z}} U = \{z \in \mathbb{Z}^{\mathcal{A}} \colon Uz = 0\}$. Moreover, $P^{z^+(\omega)} = P^{\omega}$, $P^{z^-(\omega)} = P^{r\omega}$, therefore the K-ideal is contained in the toric ideal $I(U)$.

The K-ideal is toric

Theorem

The K-ideal is the toric ideal of the cocycle matrix.

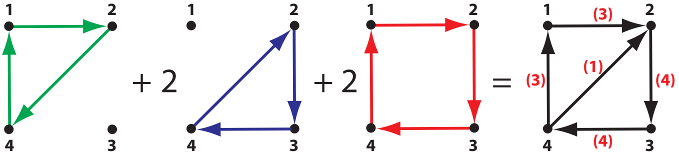
Definition (Graver basis)

$z(\omega_1)$ is conformal to $z(\omega_2)$, $z(\omega_1) \sqsubseteq z(\omega_2)$, if the component-wise product is non-negative and $|z(\omega_1)| \leq |z(\omega_2)|$ component-wise, i.e. $z_a(\omega_1)z_a(\omega_2) \geq 0$ and $|z_a(\omega_1)| \leq |z_a(\omega_2)|$ for all $a \in \mathcal{A}$. A Graver basis of $Z(\mathcal{O})$ is the set of the minimal elements with respect to the conformity partial order $\sqsubseteq$.

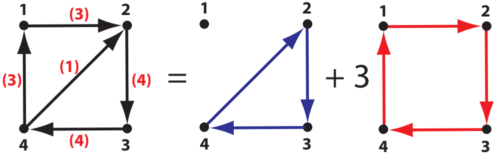
Theorem

1. *For each cycle vector $z \in Z(\mathcal{O})$, $z = \sum_{\omega \in \mathcal{C}} \lambda(\omega) z(\omega)$, there exist cycles $\omega_1, \ldots, \omega_n \in \mathcal{C}$ and positive integers $\alpha(\omega_1), \ldots, \alpha(\omega_n)$, such that $z^+ \geq z^+(\omega_i)$, $z^- \geq z^-(\omega_i)$, $i = 1, \ldots, n$ and $z = \sum_{i=1}^n \alpha(\omega_i) z(\omega_i)$.*
2. *The set $\{z(\omega) \colon \omega \in \mathcal{C}\}$ is a Graver basis of $\mathcal{Z}(\mathcal{O})$. The binomials of the cycles form a Graver basis of the K-ideal.*

Proof



$$z(\omega) = z(\omega_A) + 2z(\omega_B) + 2z(\omega_C) = (3,-3,4,-4,4,-4,0,0,-1,1)$$
$$z^+(\omega) = z^+(\omega_B) + 3z^+(\omega_C) = (3,\ 0\ ,4,\ 0\ ,4,\ 0\ ,0,0,\ 0\ ,1)$$



Positive K-ideal

- The strictly positive reversible transition probabilities on $\mathcal{O}$ are given by:

$$\begin{aligned} P_{v \to w} &= s(v,w) \prod_{S} t_S^{u_{v \to w}(S)} \\ &= s(v,w) \prod_{S:\ v \in S, w \notin S} t_S \prod_{S:\ w \in S, v \notin S} t_S^{-1}, \end{aligned}$$

  where $s(v,w) = s(w,v) > 0$, $t_S > 0$.
- The first set of parameters, $s(v,w)$, is a function of the edge.
- The second set of parameters, t_S, represent the deviation from symmetry. The second set of parameters is not identifiable because the rows of the U matrix are not linearly independent.
- The parametrization can be used to derive an explicit form of the invariant probability.

Parametric detailed balance

Theorem

Consider the strictly non-zero points on the K-variety.

1. *The symmetric parameters $s(e)$, $e \in \mathcal{E}$, are uniquely determined. The parameters t_S, $S \subset V$ are confounded by* $\ker U = \{U^t t = 0\}$.
2. *An identifiable parametrization is obtained by taking a subset of parameters corresponding to linearly independent rows, denoted by t_S, $S \subset \mathcal{S}$:*

$$P_{v \to w} = s(v,w) \prod_{S \subset \mathcal{S}:\ v \in S, w \notin S} t_S \prod_{S \subset \mathcal{S}:\ w \in S, v \notin S} t_S^{-1}$$

3. *The detailed balance equations, $\kappa(v) P_{v \to w} = \kappa(w) P_{w \to v}$, are verified if, and only if,*

$$\kappa(v) \propto \prod_{S:\ v \in S} t_S^{-2}$$

PART THREE

Markov Chains (MC)

- In a Markov chain with state space V, initial probability π_0 and stationary transitions $P_{u \to v}$, $u, v \in V$, the joint distribution up to time T on the sample space Ω_T is

$$P(\omega) = \prod_{v \in V} \pi_0(v)^{(X_0(\omega)=v)} \prod_{a \in \mathcal{A}} P_a^{N_a(\omega)}, \tag{MC}$$

  where $(V, \mathcal{A})$ is the directed graph defined by $u \to v \in \mathcal{A}$ if, and only if, $P_{u \to v} > 0$ and $N_a(\omega)$ is the number of times the arc a is contained in the path $\omega = a_1 \cdots a_T$.

- A MC is an instance of an A-model with $m = \#V + \#\mathcal{A}$ parameters and sample space $\Omega(\mathcal{A})$. We assume the MC to be connected and quasi-reversible, i.e. $a \in \mathcal{A}$ if, and only if, the reversed arc $r(a) \in \mathcal{A}$.

- The rows of A are

$$A_0(\omega) = 1, A_v(\omega) = (X_0(\omega) = v), A_a(\omega) = N_a(\omega)$$

  i.e the unnormalized density is

$$q(\omega; t) = t_0 \prod_{v \in V} t_v^{(X_0(\omega)=v)} \prod_{a \in \mathcal{A}} t_a^{N_a(\omega)} \tag{A}$$

A-model of a MC I

- The A-model is normalized by

$$Z(t) = t_0 \sum_{\omega \in \Omega(\mathcal{A})} \prod_{v \in V} t_v^{(X_0(\omega)=v)} \prod_{a \in \mathcal{A}} t_a^{N_a(\omega)}$$

  and it is a Markov proces with non-stationary transition probabilities.

- Define $a(v) = \sum_{w\colon (v \to w) \in \mathcal{A}} t_{v \to w}$; hence $P_{v \to w} = t_{v \to w}/a(v)$ is a transition probability supported by $\mathcal{A}$ and $\nu(v) = a(v)/\sum_v a(v)$ is a probability on V. Consider the change of parameters

$$\alpha\pi(v) = t_v, \quad \beta\nu(v)P_{v \to w} = t_{v \to w},$$

  to get the new parameterization of the unnormalized density

$$q(\omega; \pi, P) \propto \left(\prod_{v \in V} \nu(v)^{N_{v+}(\omega)} \right) \left(\prod_{v \in V} \pi(v)^{(X_0(\omega)=v)} \prod_{a \in \mathcal{A}} P_a^{N_a(\omega)} \right)$$

A-model of a MC II

- Note that for $\omega = v_0 v_1 \cdots v_T$

$$\prod_{v \in V} \nu(v)^{N_{v+}(\omega)} = \nu(v_0)\nu(v_1)\cdots\nu(v_T)$$

  which is constant if ν is constant.

- The (MC) model is derived from the (A) model by adding the parametric constrains

$$\sum_{w\colon\ v_1 \to w \in \mathcal{A}} t_{v_1 \to w} = \sum_{w\colon\ v_2 \to w \in \mathcal{A}} t_{v_2 \to w}, \quad v_1, v_2 \in V.$$

Discussion

- The information geometry of MCs and RMCs is best expressed in exponential form, i.e. with respect to the parameterization $\theta = \log t$. It is instrumental the non-parametric Information Geometry as described e.g. in
- Giovanni Pistone and Maria Piera Rogantin. The exponential statistical manifold: mean parameters, orthogonality and space transformations. Bernoulli, 5(4):721–760, August 1999. ISSN 1350-7265.
- The manifold of MCs and of RMCs in the corresponding *A*-model are curved exponential families whose tangent bundles are computed from the constrains.
- It should be of interest to find an orthogonal decomposition of a MC or *A*-model into a reversible component and the other orthogonal to the reversible manifold.
- The use of deformed exponentials does not exclude the algebraic theory as presented here; a toy example with the Kaniadakis's exponential was discussed in
- Giovanni Pistone. κ-exponential models from the geometrical viewpoint. The European Phisical Journal B Condensed Matter Physics, 71(1):29–37, July I 2009. ISSN 1434-6028. doi10.1140/epjb/e2009-00154-y. URL `http://dx.medra.org/10.1140/epjb/e2009-00154-y`.