Information Geometry and its Applications III

August 02 - 06, 2010
Max-Planck-Institut für Mathematik in den Naturwissenschaften, Leipzig

Finite generation of cumulants

Giovanni Pistone

Collegio Carlo Alberto, Moncalieri, Italy

August 6, 2010

Joint work with Henry P. Wynn, LSE London England

Summary

Algebraic Statistics

Elimination in polynomial ideals

Finite generation

Toric statistical models

Reversible Markov chains

Toric and differential ideals

- Markov bases of toric ideals Persi Diaconis and Bernd Sturmfels. Algebraic algorithms for sampling from conditional distributions. *Ann. Statist.*, 26(1): 363–397, 1998. ISSN 0090-5364 preprint 1993
- Gröbner bases in Desing of Experiments Giovanni Pistone and Henry P. Wynn. Generalised confounding with Gröbner bases. *Biometrika*, 83(3): 653–666, March 1996. ISSN 0006-3444
- The name of the game Giovanni Pistone, Eva Riccomagno, and Henry P. Wynn. *Algebraic statistics. Computational commutative algebra in statistics*, volume 89 of *Monographs on Statistics and Applied Probability*. Chapman & Hall/CRC, Boca Raton, FL, 2001. ISBN 1-58488-204-2

Ideals, bases

- $R = k[x_1, \ldots, x_d]$ is the ring of polynomials in the inderrminates $x_1, \ldots, x_d$ with coefficients in field k.
- Polynomials $f_1, \ldots, f_m$ generate the ideal

$$\langle f_1, \ldots, f_m \rangle = \left\{ \sum_{j=1}^{m} g_j f_j : g_j \in R \right\}$$

- Every ideal has many finite generating set or bases
- A monomial order is a type of total order on monomials which is compatible with product. Given a monomial order it is possible to write every polynomial in decreasing order and to identify its leading term.
- The elimination ideal is the ideal

$$\langle f_1, \ldots, f_m \rangle \cap k[x_1, \ldots, x_l], \quad l \leq d$$

- CCA Martin Kreuzer and Lorenzo Robbiano. *Computational commutative algebra. 1*. Springer-Verlag, Berlin, 2000. ISBN 3-540-67733-X

```
Use R::= Q[x[1..4],t,y[1..2]], Lex; -- ring
Eqs:=[x[1]-(1-x[2])*x[2], -- first bernoulli
      x[3]-(1-x[4])*x[4], -- second bernoulli
      y[1]-x[2]-x[4], -- sum of k'
      y[2]-x[1]-x[3], -- sum of k''
      t-x[2]+x[4]]; -- parameter
I:=Ideal(Eqs);
GBasis(I); -- Groebner basis
Elim(x,I); -- Elimination ideal

[-x[2] - x[4] + y[1],
 x[3] + x[4]^2 - x[4],
 x[1] + x[2]^2 - x[2],
 -2x[4] - t + y[1],
 -1/2t^2 - 1/2y[1]^2 + y[1] - y[2]]
-------------------------------
Ideal(-1/2t^2 - 1/2y[1]^2 + y[1] - y[2])
```


- CoCoATeam. CoCoA: a system for doing Computations in Commutative Algebra. Available at cocoa.dima.unige.it, online. L. Robbiano team leader.

Multivariate cumulant

Definition (Moment and cumulant generating function)

- X is a random vector in $\mathbb{R}^m$.
- For $\theta \in \mathbb{R}^m$, $\theta \cdot X = \sum_{i=1}^m \theta_i X_i$ is the scalar product.
- D_X is the *interior* of the convex set

$$\left\{\theta \in \mathbb{R}^m : E[e^{\theta \cdot X}] < +\infty\right\}.$$

- If $D_X \neq \emptyset$, then the *moment (generating) function* M_X and *cumulant (generating) function* K_X of X are the functions defined for each $t \in D_X$ by the equations

$$M_X(\theta) = E[e^{\theta \cdot X}],$$
$$K_X(\theta) = \log M_X(\theta).$$

Monomial and moment aliasing

- Let D be a finite set of points in $\mathbb{R}^m$, $I(D)$ the design ideal in $\mathbb{R}[x_1, \ldots, x_m]$, $g_1, \ldots, g_k$ a polynomial basis of $I(D)$, x^α, $\alpha \in L$, a linear monomial basis of $\mathbb{R}[x_1, \ldots, x_m]/I(D)$. Given a Gröbner basis, the monomials that are not divider by a leading term for such a (linear) basis.
- This is the usual setting of the algebric theory of Design of Experiments. Each equation $g(x) = 0$, $g \in I(D)$, is an aliasing relation between terms.
- Let

$$H(x) = \exp\left(\sum_{i=1}^{n} s_i x_i\right) = \sum_{\alpha \in L} b_\alpha(s) x^\alpha .$$

Therefore $M_X(s) = \sum_{\beta \geq 0} \frac{s^\beta \mu_\beta}{\beta!} = \sum_{\alpha \in L} b_\alpha(s) \mu_\alpha$,

$$\mu_\beta = \sum_{\alpha \in L} b_{\alpha,\beta} \mu_\alpha , \quad b_{\alpha,\beta} = D_\beta b_\alpha(s)|_{s=0}$$

- The monomial basis is computed by CoCoA
- the coefficients $b_\alpha(s)$ are obtained by interpolation

Cumulant aliasing

For a discrete distribution and monomial order τ every cumulant $\mu_\beta, \beta \geq 0$ is expressible as a linear function of the moments $\mu_\alpha,\ \alpha \in L$, whose coefficients depend only the support and choice of monomial ordering, not the $p(x)$.

Theorem (Cumulants aliasing)

For a discrete distribution and monomial order τ every cumulant $\kappa_\beta, \beta \geq 0$ is expressible as a polynomial function of the cumulant $\kappa_\alpha,\ \alpha \in L$, whose form is only dependent of the support and monomial ordering.

- Giovanni Pistone and Henry P. Wynn. Cumulant varieties. *Journal of Symbolic Computation*, 41(2):210–221, 2006. ISSN 0747-7171

Finite generation

Definition

The cumulants of X are called *finitely generated* if there exist polynomials

$$F_{hk}(\eta_i : i = 1, \ldots, m; \gamma_{ij} : i \leq j = 1, \ldots, m) \quad , \quad h \leq k = 1, \ldots, m \quad ,$$

such that the corresponding system of equations can be uniquely solved for $\gamma = (\gamma_{ij})_{1 \leq h \leq k \leq m}$ as a function of $\eta = (\eta_i)_{1 \leq i \leq m}$, around the point

$$\eta_0 = K'_X(0) \quad , \quad \gamma_0 = K''_X(0) \quad ,$$

and the equations

$$F_{hk}(K'_X(t), K''_X(t)) = 0 \quad , \quad h \leq k = 1, \ldots, m \quad ,$$

hold in a neighborhood of 0. The polynomials $F = (F_{hk})_{h \leq k = 1, \ldots, m}$ are called *generating polynomials of* X.

CoCoA

```
Use R::= Q[x[1..4],t,y[1..2]], Lex; -- ring
Eqs:=[x[1]-(1-x[2])*x[2], -- first bernoulli
      x[3]-(1-x[4])*x[4], -- second bernoulli
      y[1]-x[2]-x[4], -- sum of k'
      y[2]-x[1]-x[3], -- sum of k''
      t-x[2]+x[4]]; -- parameter
I:=Ideal(Eqs);
GBasis(I);
Elim(x,I);

[-x[2] - x[4] + y[1],
 x[3] + x[4]^2 - x[4],
 x[1] + x[2]^2 - x[2],
 -2x[4] - t + y[1],
 -1/2t^2 - 1/2y[1]^2 + y[1] - y[2]]
-------------------------------
Ideal(-1/2t^2 - 1/2y[1]^2 + y[1] - y[2])
```


- CoCoATeam. CoCoA: a system for doing Computations in Commutative Algebra. Available at cocoa.dima.unige.it, online

Variance function: Morris

- The following table is adapted from [Morris, 1982, Table 1], where all the distributions such as the variance function is a quadratic polynomial in the mean are studied.
- In our terms, the variance $K''(\theta)$ and the mean $K'(\theta)$ are related by a *generating polynomial* of degree 2.

Distribution	Parameters	Generating polynomial
Normal $\mathrm{N}(\mu, \sigma^2)$	$\mu,\ \sigma^2$	$K'' - \sigma^2$
Poisson $\mathrm{P}(\lambda)$	λ	$K'' - K'$
Gamma $\Gamma(\alpha, \lambda)$	$\alpha,\ \lambda$	$\alpha K'' - (K')^2$
Binomial $\mathrm{Bin}(n, p)$	$n,\ p$	$nK'' - K'(n - K')$
Negative Binomial $\mathrm{NegBin}(r, p)$	$r,\ p$	$rK'' - K'(r + K')$
Generalised Hyperbolic Secant	$r,\ \lambda = \tan t$	$rK'' - (K')^2 - r^2$

Finite generation; example

- The generating polynomial uniquely defines the corresponding distribution. E.g. the differential equation for $\eta(\theta) = K'(\theta)$ in the GHS case is
$$r\eta'(\theta) = \eta(\theta)^2 + r^2, \quad \eta(0) = 0$$
The unique solution is
$$\eta(\theta) = r \tan t$$
so that
$$K(\theta) = r \int_0^t \tan \tau \, d\tau = r \log \sec t.$$
- All cumulants are polynomials in the mean parameter. E.g. for the GHS distribution
$$r^n K^{(n)}(\theta) = f_n(K'(\theta)), n = 2, 3, \ldots$$
where
$$f_{n+1}(\eta) = f_n'(\eta)(\eta^2 + r^2).$$

Finite generation

- The *Laplace (double exponential) density* with parameter 1 has cumulant function $K(t) = -\log(1-t^2)$. Then the first and second derivatives are

$$K'(t) = \frac{2t}{1-t^2},$$
$$K''(t) = 2\frac{1+t^2}{(1-t^2)^2}.$$

The generating polynomial is

$$(K'')^2 - 2(1+(K')^2)K'' + (K')^2 + (K')^4.$$

- The uniform density on $\{0,1,2\}$ has generating polynomial

$$3(K')^4 + 2K' - 2K'' + 11(K')^2 - 12K'K'' - 12(K')^3 + 6(K')^2(K'') + 3(K'')^2$$

Finite generation

Theorem

The FGC property is stable for

1. *joining independent components, in particular sampling;*
2. *invertible linear transformations;*
3. *convolutions of the same distribution.*

Theorem

- *Every discrete distribution supported on an equally spaced set of reals has the FGC property.*
- *Every finite mixture of exponential random variables has the FGC property.*
- *Let $p_X(x)$ be the density function of a random variable with the FGC property. Then if Y is a random variable with density $g(y)p_X(y)$ where $g(y)$ is polynomial then Y also has the FGC property.*

Finite generation: discussion

- For U[0, 1] the MGF is $M(\theta) = \frac{e^\theta - 1}{\theta}$.
- This involves θ and e^θ. We set $z = \frac{1}{e^\theta - 1}$ and $t = \frac{1}{\theta}$, so that $z' = -(1+z)z \; t' = -t^2$ and

$$
\begin{aligned}
K' &= 1 + z - t \\
K'' &= -z - z^2 + t^2 \\
K''' &= z + 3z^2 + 2z^3 - 2t^3
\end{aligned}
$$

- Algebraic elimination of t and z gives

$$
\begin{aligned}
&(K')^6 - 5(K')^5 - 3(K')^4K'' + 17/2(K')^4 + 2(K')^3K'' - 4(K')^3K''' \\
&- 6(K')^3 + 3(K')^2(K'')^2 + (K')^2K'' + 6(K')^2K''' + 3/2(K')^2 \\
&- 5K'(K'')^2 - 3K'K''' - (K'')^3 + 5/2(K'')^2 - 1/2K'' + 1/2K'''
\end{aligned}
$$

Toric ideals

- Let be given an integer model matrix X with rows $x \in \mathcal{D}$ and d columns.
- Consider the ring $k[y_x \colon x \in \mathcal{D}]$ and the Laurent ring $k(t_1, \ldots, t_d)$, together with their homomorphism A defined by

$$A \colon y_x \longmapsto \prod_{j=1}^{d} t_j^{A_{x,j}} = t^{A(x)},$$

- The kernel $I(A)$ of h is called the *toric ideal* of A,

$$I(A) = \left\{ f \in k[y_x \colon x \in \mathcal{D}] : f\left(t^{A(x)} \colon x \in \mathcal{D}\right) = 0 \right\}.$$

- The toric ideal $I(A)$ is a prime ideal and the binomials

$$P^{z^+} - P^{z^-}, \quad z \in \mathbb{Z}^{\mathcal{D}}, \quad A^T z = 0,$$

  are a generating set of $I(A)$ as a k-vector space.
- In particular, Hilbert says that a finite generating set of the ideal is formed by selecting a finite subset of such binomials.

- Bernd Sturmfels. *Gröbner bases and convex polytopes*. American Mathematical Society, Providence, RI, 1996. ISBN 0-8218-0487-1

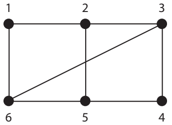
Toric ideals in statistics

- For the 2×2 independence model parameterized as $p_{x_1,x_2} = t_0 t_1^{x_1} t_2^{x_2}$, one computes the invariant:

$$A = \begin{matrix} & 1 & x_1 & x_2 \\ ++ & +1 & +1 & +1 \\ +- & +1 & +1 & -1 \\ -+ & +1 & -1 & +1 \\ -- & +1 & -1 & -1 \end{matrix}, \quad z = \begin{bmatrix} +1 \\ -1 \\ -1 \\ +1 \end{bmatrix}, \quad p_{++}p_{--} - p_{+-}p_{-+} \in I(A)$$

- Mathias Drton, Bernd Sturmfels, and Seth Sullivant. *Lectures on Algebraic Statistics*. Number 39 in Oberwolfach Seminars. Birkhäuser, 2009. ISBN 978-3-7643-8904-8
- Viceversa, one could go from the invariants to the parameterization.
- Giovanni Pistone and Maria Piera Rogantin. Algebra of revesible Markov chains. arXiv:1007.4282v1, 2010

CoCoA elimination




```
Use S::=Q[t,k[1..6],p[1..6,1..6]];
Set Indentation;
NI:=6; M:=[];
Define Lista(L,NI);
   For I:=1 To NI Do
       For J:=1 To I-1 Do
           Append(L,k[I]p[I,J]-k[J]p[J,I]); EndFor;
   EndFor; Return L; EndDefine;
N:=Lista(M,NI);
LL:=t*Product([k[I]|I In 1..NI])-1; Append(N,LL);
P0:=[p[1,3],p[1,4],p[1,5],p[2,4],p[2,6], p[3,1],p[3,5],
p[4,1],p[4,2],p[4,6],p[5,1],p[5,3],p[6,2],p[6,4]];
N:=Concat(N,P0);
E:=Elim(k,Ideal(N)); GB:=ReducedGBasis(E);  GB;
```

CoCoA output

```
GB;
[
 p[1,3],  p[1,4],  p[1,5],  p[2,4],  p[2,6],  p[3,1],  p[3,5],
 p[4,1],  p[4,2],  p[4,6],  p[5,1],  p[5,3],  p[6,2],  p[6,4],

  p[2,3]p[3,4]p[4,5]p[5,2] - p[2,5]p[3,2]p[4,3]p[5,4],
  p[1,2]p[2,3]p[3,6]p[6,1] - p[1,6]p[2,1]p[3,2]p[6,3],
  p[1,2]p[2,5]p[5,6]p[6,1] - p[1,6]p[2,1]p[5,2]p[6,5],
  p[2,5]p[3,2]p[5,6]p[6,3] - p[2,3]p[3,6]p[5,2]p[6,5],
  p[3,4]p[4,5]p[5,6]p[6,3] - p[3,6]p[4,3]p[5,4]p[6,5],
  p[1,2]p[2,5]p[3,6]p[4,3]p[5,4]p[6,1] -
        p[1,6]p[2,1]p[3,4]p[4,5]p[5,2]p[6,3],
  p[1,2]p[2,3]p[3,4]p[4,5]p[5,6]p[6,1] -
        p[1,6]p[2,1]p[3,2]p[4,3]p[5,4]p[6,5]]
```

Toric model and Weyl

- Consider the design (sample space) $\mathcal{D} \subset \mathbb{Z}_+^d$ with reference measure μ, e.g. $\mu = 1$.
- The design ideal is

$$I(\mathcal{D}) = \{f \in \mathbb{Q}[x_1, \dots, x_d] : f(x) = 0, x \in \mathcal{D}\}.$$

- Consider the toric statistical model

$$p(x;t) \propto \prod_{j=1}^{d} t_j^{x_j}, \quad x \in \mathcal{D}, \quad t_j \geq 0, \quad j = 1, \dots, d,$$

- The normalizing constant (partition funtion) is

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

- There exists a polynomial $p(t,x) \in \mathbb{Q}[t,x]$ such that $p(t,x) = t^x$, $x \in \mathcal{D}$.

Weyl differential algebra

- The Weyl algebra is the ring of differential operators $\mathbb{C}\langle t_1 \dots t_d, \partial_1 \dots \partial_d \rangle$ where everything commutes but

$$\partial_i t_i - t_i \partial_i = 1$$

- Define the operators

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

  where the second equality follows from the commutation relation.

- 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 $Z(t) \in \mathbb{C}[t_1, \dots, t_d]$ gives

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

- S. C. Coutinho. *A primer of algebraic D-modules*, volume 33 of *London Mathematical Society Student Texts*. Cambridge University Press, Cambridge, 1995. ISBN 0-521-55119-6; 0-521-55908-1. doi10.1017/CBO9780511623653. URL http://dx.doi.org/10.1017/CBO9780511623653

- Note the commutativity

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

hence we have an action of multivatiate monomials:

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

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

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

- By consider the ring homomorphism

$$\begin{array}{rccc} A\colon & \mathbb{C}[x] & \to & \mathbb{C}\langle t_1 \ldots t_d, \partial_1 \ldots \partial_d\rangle \\ & x_i & \mapsto & t_i\partial_i \end{array}$$

We have

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

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-omogeneous differential equations:*

$$A(x^\alpha) \bullet Z(t) = \sum_{x \in \mathcal{D}} x^\alpha t^x \mu(x), \quad \alpha \in M.$$

2. *Let $f_a(x)$ be the (reduced) indicator polynomomial of $a \in \mathcal{D}$. Then $Z(t)$ satisfies the following system of $\#\mathcal{D}$ linear non-omogeneous differential equations:*

$$A(f_a(x)) \bullet Z(t) = \mu(a) t^a, \quad a \in \mathcal{D}$$

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*

$$g\left(\mu(x)^{-1} A(f_a(x)) \bullet Z(t) \colon a \in \mathcal{D}\right) = 0$$

4. *Also for cumulants.*

- G. Pistone, H. Wynn, in progress

IGAIA 3 was made for us by

- Nihat Ay Max Planck Institute for Mathematics in the Sciences Information Theory of Cognitive Systems Group Germany
- Paolo Gibilisco Università degli Studi di Roma Tor Vergata Facoltà di Economia Italy
- František Matúš Academy of Sciences of the Czech Republic Institute of Information Theory and Automation Czech Republic
- Antje Vandenberg Max Planck Institute for Mathematics in the Sciences

We appreciate! Thanks!