

# Hybrid symbolic-numerical approaches: a discussion

H.P. Wynn, London School of Economics, UK

We are concerned with the relationship between the area known as algebraic statistics (AS) and the use of differential geometry in statistics, captured by the term information geometry (IG). From the point of view of the WOGAS series this relationship may be thought of as central, but it is a largely unexplored frontier and technically quite hard.

Algebraic statistics can be defined as the use of computational commutative algebra in statistics, in particular polynomial rings and algebraic varieties, together with associated operations: quotients, factorisations, elimination and so on [2]. We are familiar with polynomials from polynomial regression and the expression for independence and conditional independence in algebraic form. Many algebraic operations can be carried out using powerful computer algebra packages: CoCoa [1], Singular, Macauley 2, 4ti2, GAP and more. The talk by Hugo Maruri-Aguillar shows the power of algebraic methods. Information geometry has its own entities: manifolds, curvature, flows, tensors, connections etc.

Both areas have a great deal of overlap in the areas of statistics : estimation, testing etc. But there are some important differences. Algebraic statistics has made most headway in intrinsically discrete models such as Poisson/multinomial log- linear models (exponential family), experimental design and mixture/latent models, although there is some newer work on Gaussian covariance models. Information geometry is fundamentally concerned with asymptotic statistics, where curvature of manifolds in parameter space and the other quantities mentioned above affect asymptotic efficiency. For a comprehensive account of algebraic statistics see [5] and the recent collections [4] [3]. The aim of [3] was to make some link between AS and IG; see the overview article by Giovanni Pistone. Some key references on IG are given in Paul Marriott's discussion document at this workshop.

One natural interface is where the manifolds in the IG framework are algebraic varieties. In that case there is some hope that the algebra will give insight and the algebraic packages can be used to compute some of the differential geometry entities. This works most naturally for the standard exponential family and shows some promise for curved exponential family, provided that the non-linear constraints can be expressed algebraically in the mean,  $\eta$ , or the natural parameter,  $\theta$ . An example of a computational issue is that the maximum likelihood estimators may be algebraic in that they are solutions of algebraic equations, but numeric in that these equations may not have a closed form. This feature may then be inherited by any quantity requiring a plug-in MLE, such as a bias correction (see the talk by Kei Kobayashi).

Having said that algebraic statistics comes into its own for discrete problem one can expect that approximations to integrals and differentials that arise in information geometry using some kind of discretization will allow algebraic methods to be imported. This discretization is at the heart of numerical analysis so that a research programme

to investigate how AS interfaces with IG via discretization, might review where algebraic geometry and traditional numerical analysis are already close. One hopeful area is quadrature. Careful placing of quadrature points allows exact integration of polynomial function with respect to a given measure, eg Gaussian. The “sparse grids” used in stochastic finite element methods (SFEM), for both integration and interpolation, are interestingly close to factorial designs. The talk by Mark Richardson discusses (polynomial and therefore analytic) Chebychev approximations to analytic functions. An area where polynomial algebras and (linear) ordinary and partial differential equations link together is differential algebras (Weyl algebras, D-modules). These are just beginning to be used in statistics.

To discuss effective hybrid methods may be premature, but the ingredients are there. We can ask a few speculative questions:

1. Can the more advanced areas of algebra make a contribution eg fast algorithms for minimal free resolutions, Betti numbers?
2. Should more use be made of numerical methods which start with algebra, such as fast numerical methods to solve polynomial equations?
3. What is the relation between experimental design for observation and interpolation points for quadrature?
4. Can we capture stochastics algebraically eg by using methods from SFEM?
5. Can we combine symbolic differential algebra with approximation methods eg via chebfun
6. Should we extend the range of kernels from polynomial: eg to Fourier and Haar?

## References

- [1] CoCoATeam: CoCoALib: a C++ library for doing Computations in Commutative Algebra, Available at <http://cocoa.dima.unige.it>
- [2] Cox D., Little, J. and OShea, D. Ideals, varieties, and algorithms: an introduction to computational algebraic geometry and commutative algebra, 3rd ed., Undergraduate Texts in Mathematics, Springer, New York, 2007.
- [3] Algebraic and Geometric Methods in Statistics. (Volume in honour of G. Pistone). E. Riccomagno, M Rigontin, H. P. Wynn (eds). Springer Verlag. 2008.
- [4] Algebraic methods in statistics and probability II. M Viana, H. P. Wynn (eds). American Mathematical Society, Contemporary Mathematics, 287, 2010.
- [5] Drton, M., Sturmfels, B. and Sullivant, S. Lectures on algebraic statistics. Springer. 2009.