GRAPHICAL MODELS, WITH POTENTIAL AS/GS SYNERGIES

Drawing on Algebraic and Geometric Statistics respectively, two recent research projects have been seeking, in their own contexts, to

expose the implicit/hidden inferential impact of modelling assumptions:

(1) Among others, Jim Smith and Piotr Zwiernik have been exploring the inferential effects of varying the conditional independence statements - or, rather, *assumptions* - together comprising a Graphical Model

(2) Frank Critchley, Paul Marriott, Paul Vos and Karim Anaya-Izquierdo have been exploring the inferential effects of varying modelling assumptions, within a Computational Information Geometry framework.

The question naturally arises as to how best to identify and develop potential synergies between these projects, briefly reviewed below, possibilities including:

(a) use the extended multinomial model framework of (2) to, potentially, provide distributions for diagnostic test statistics being developed in (1)

(b) exploring the potential for developing the approximate cut tool of (2) to examine (approximate) factorisations on the likelihood implicit in graphical models.

1. Graphical Models

Graphical models form a family of statistical models which code various conditional independence (CI) statements in a compact way (see [7]). A graphical model is given by a random vector $\mathbf{X}$ and a graph $G = (V, E)$. The nodes of G represent random variables in the vector and edges give some information about independence. There are various different types of graphical models. In the most classical undirected graph model a missing edge between $i, j \in V$ corresponds to the conditional independence statement $X_i \perp\!\!\!\perp X_j | \mathbf{X} \setminus \{X_i, X_j\}$. Recently, perhaps the most well studied class of graphical models is the Bayesian Network (BN). These have been central to the development of probabilisitic expert systems. They have also been the framework - propagated especially by Judea Pearl - that at the moment best explains complex causal structures in high dimensional problems.

Once graphical modellers turned their minds to inferential issues, where the parameters of statistical graphical models needed to be estimated, it was noted that - at least for finite discrete problems, or parameterised families of graphical models - functions of the corresponding likelihoods could often be defined by particular families of rational functions. In particular, when all the variables in the system are observed, graphical models exhibit a useful modularity and the corresponding geometry was therefore straightforward and very elegant. In particular, if the defining graph was decomposable, it was possible to estimate all the conditional probabilities that parametrise such models, maximum likelihood estimates being simple sample proportions, while the conjugate Bayesian analysis was straightforward (see e.g. [13]). This was reflected in the beauty of the underlying geometry of these model classes.

However, in many of the current applications, it is common for many variables in a statistical model not to be observable directly, but only implicitly. In the vernacular, such graphical models are said to have some of their nodes *hidden* or *systematically missing*. In this case, the associated geometry of graphical models is much richer and more interesting. Indeed, the graph itself is not nearly such a powerful summary. For example, unlike the full sampling case, the number of levels in some of the variables in the problem (here, the hidden ones) had an absolutely critical impact on the geometry of the model class; and hence, for example, on the extent and nature of unidentifiability of the system; and hence on the properties of different inferential techniques. Such information is never represented in graphical models.Typically, for example,

- there are no closed form formulae for the maximum likelihood estimators,
- the likelihood function has many optima, and
- models are not always identified.

In addition, many of the techniques developed for regular models do not apply (see [15, Section 1.2]). Naïve techniques based only on the graph therefore then start to fail, and a good understanding of the underlying geometry is absolutely critical to a proper appreciation of learning about the model class from data (see [3], [5]).

There are various issues which may be addressed using computational algebraic geometry. Recently, some progress has been made in the analysis of identifiability of graphical models with hidden variables. A large part of this work is based on resurrecting an old theorem of Joseph Kruskal [6] (see [1] for the most general treatment). This work gives very general conditions to assure that a model is generically identifiable. In certain cases, a more detailed analysis of identifiability is available (see [14], [17]).

Some effort has also been put into understanding the semi-algebraic structure of graphical models with hidden variables ([10], [11], [9], [16]). In statistical terms this would complement the identifiability issue. It appears that a key point is to embed the model being analysed in the space of its moments, rather than of raw probabilities (see also [2], [4]). In a recent paper, [17], we showed that there exists an embedding of tree models with hidden variables having very nice properties. The new coordinates, called tree cumulants, are based on moments and their definition resembles the combinatorial definition of cumulants. The definition of cumulants involves the poset of set partitions of a finite set ([8], [12]), whereas the definition of tree cumulants uses the poset of partitions of a finite set (of leaves) induced by removing edges of a tree.

2. Computational Information Geometry for Statistical Science

Computational Information Geometry for Statistical Science provides an operational vehicle for, in particular:

(a) sensitivity analysis and

(b) opening up - and extending - practically important, powerful results from classical information geometry to the mainstream user.

It does this by, respectively:

(a) working in ‘a universal space of all possible models’, allowing the geometry to determine the limits of empirical knowledge, following the principle of ‘learn what you can; explore what you can’t’

(b) removing the requirement of having first to become fluent in the language of differential geometry, which had locked away these results behind notational and conceptual bars.

Computational Information Geometry (CIG) has been shown through a broad range of important examples to be both operational and practically useful. The associated general theory develops classical information geometry so as to apply to new mathematical structures. Finiteness is essential in building an operational computational framework. Given certain choices about it, computational information geometry encompasses and extends both (i) classical information geometry and (ii) the convex geometry behind Lindsay's (1995) approach to inference in mixture models, opening up new approaches to the hard inferential problems found in the latter context.

Illustrative examples include mixture models, curved exponential families and examples, such as the generalised extreme value distributions, falling outside the standard regularity conditions of classical information geometry. Work in progress is targeting applications of CIG to generalised linear models, the workhorse of much of applied statistics.

Future developments include developing links to non- and semi-parametric inference. In particular, to empirical likelihood, estimating equation, quasi-likelihood, bootstrap and maximum entropy methods.

Further details are given in three attached talks (CIG1.pdf - CIG3.pdf).

References

[1] E. S. Allman, C. Matias, and J. A. Rhodes, *Identifiability of parameters in latent structure models with many observed variables*, Ann. Statist, 37 (2009), pp. 3099–3132.

[2] N. Beerenwinkel, N. Eriksson, and B. Sturmfels, *Conjunctive Bayesian networks*, Bernoulli, 13 (2007), pp. 893–909.

[3] D. M. Chickering and D. Heckerman, *Efficient approximations for the marginal likelihood of Bayesian networks with hidden variables*, Machine Learning, 29 (1997), pp. 181–212.

[4] M. Drton and T. Richardson, *Binary models for marginal independence*, Journal of the Royal Statistical Society: Series B (Statistical Methodology), 70 (2008), pp. 287–309.

[5] D. Geiger, D. Heckerman, H. King, and C. Meek, *Stratified exponential families: graphical models and model selection*, Ann. Statist., 29 (2001), pp. 505–529.

[6] J. B. Kruskal, *Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics*, Linear Algebra and Appl., 18 (1977), pp. 95–138.

[7] S. L. Lauritzen, *Graphical models*, vol. 17 of Oxford Statistical Science Series, The Clarendon Press Oxford University Press, New York, 1996. Oxford Science Publications.

[8] P. McCullagh, *Tensor methods in statistics*, Monographs on Statistics and Applied Probability, Chapman & Hall, London, 1987.

[9] D. Mond, J. Q. Smith, and D. van Straten, *Stochastic factorizations, sandwiched simplices and the topology of the space of explanations*, R. Soc. Lond. Proc. Ser. A Math. Phys. Eng. Sci., 459 (2003), pp. 2821–2845.

[10] R. Settimi and J. Q. Smith, *On the Geometry of Bayesian Graphical Models with Hidden Variables*, in UAI, G. F. Cooper and S. Moral, eds., Morgan Kaufmann, 1998, pp. 472–479.

[11] R. Settimi and J. Q. Smith, *Geometry, moments and conditional independence trees with hidden variables*, Ann. Statist., 28 (2000), pp. 1179–1205.

[12] T. P. Speed, *Cumulants and partition lattices*, Austral. J. Statist, 25 (1983), pp. 378–388.

[13] D. J. Spiegelhalter, A. P. Dawid, S. L. Lauritzen, and R. G. Cowell, *Bayesian analysis in expert systems*, Statist. Sci., 8 (1993), pp. 219–283. With comments and a rejoinder by the authors.

[14] E. Stanghellini and B. Vantaggi, *On the identification of discrete graphical models with hidden nodes*, Arxiv preprint arXiv:1009.4279, (2010).

[15] S. Watanabe, *Algebraic Geometry and Statistical Learning Theory*, no. 25 in Cambridge Monographs on Applied and Computational Mathematics, Cambridge University Press, 2009. ISBN-13: 9780521864671.

[16] P. Zwiernik and J. Q. Smith, *The semialgebraic description of tree models for binary data*. arXiv:0904.1980, May 2010.

[17] P. Zwiernik and J. Q. Smith, *Tree-cumulants and the geometry of binary tree models*. arXiv:1004.4360, 2010. to appear in Bernoulli.