

Programme and Abstracts for WOGAS2

Programme

Tues 6 April

- 11:00 REGISTRATION OPENS (coffee available)
- 12:00 BUFFET LUNCH
- 12:50-13:00 Opening Remarks
- 13:00-13:10 Frank Critchley
Emerging geometries for statistical science: I. A very short introduction
- 13:10-14:05 Peter Jupp
A review of geometric statistics
- 14:05-15:00 Henry Wynn
A review of algebraic statistics
- 15:00-15:30 TEA
- 15:30-16:15 John Copas
Model uncertainty: minimax confidence limits from models that fit the data
- 16:15-17:00 Paul Marriott
Computational information geometry: Geometry of model choice
- 17:00-17:30 MORE TEA!
- 17:30-18:00 Giovanni Pistone
Algebra of reversible Markov chains
- 18:00-18:30 Elena Stanghellini
On the identification of discrete graphical models with hidden nodes
- 18:30-19:45 POSTER RECEPTION (beginning with 1-minute summaries)
John Aston Maurizio Filippone Helene Gehrman
Luigi Malagò Hugo Maruri-Aguilar Maria Sofia Massa
Johannes Rauh Eva Riccomagno Germain Van Bever
- 20:00 DINNER

Wed 7 April

- 09:30-10:15 Piotr Zwiernik
Posets, Möbius functions and tree-cumulants:
- 10:15-11:00 Hélène Massam
The geometry of hierarchical discrete loglinear models
- 11:00-11:30 COFFEE
- 11:30-12:15 Karim Anaya-Izquierdo
Computational information geometry: Model sensitivity and approximate cuts
- 12:15-13:00 Nihat Ay
Information geometry of polytopes - general theory and applications to stochastic matrices
- 13:00-14:30 LUNCH
- 14:30-15:15 Ian Dinwoodie
Syzygies for Metropolis proposal kernels
- 15:15-16:00 Nanny Wermuth
Special features of probability distributions: which are detectable by algebraic methods?
- 16:00-16:30 TEA
- 16:30-16:40 Frank Critchley
Emerging geometries for statistical science: II. Future directions
- 16:40-17:00 Roundtable Discussion
- 17:00 WORKSHOP CLOSES

Abstracts

TALKS

Frank Critchley *Open University*

Emerging Geometries for Statistical Science:I. A Very Short Introduction

A very short introduction to geometric and algebraic methods in statistics is offered, as background to the twin review talks which follow.

Peter Jupp *University of St. Andrews*

A review of geometric statistics

Henry Wynn *LSE*

A review of algebraic statistics

John Copas *University of Warwick*

Model uncertainty: minimax confidence limits from models that fit the data.

A goodness-of-fit statistic can be thought of as a measure of the distance between the data and the projection of the data onto the model. A goodness-of-fit test therefore identifies a neighbourhood of models which fit the data. Each model gives confidence limits for a parameter of interest. This talk will look at the minimax of these upper confidence limits — max over the neighbourhood of models but min over the choice of goodness-of-fit test. Similarly, maximin for the lower confidence limit. This suggests some challenging questions about the use of models and model diagnostics.

Paul Marriott *University of Waterloo*

Computational Information Geometry: Geometry of Model Choice

Our project uses computational information geometry to develop diagnostic tools to help understand sensitivity to model choice by building an appropriate perturbation space for a given inference problem. Focused on generalised linear models, the long-term aim is to engage with statistical practice via appropriate R software encapsulating the underlying geometry, whose details the user may ignore.

This talk exploits the universality of the extended multinomial model to examine the interaction between prior considerations (scientific and otherwise), model choice and final inference. A range

of examples is used for illustration. There are strong points of contact with the work of Copas and Eguchi.

EPSRC support under grant EP/E017878/1 is gratefully acknowledged.

(joint work with Karim Anaya-Izquierdo, Frank Critchley and Paul Vos)

Giovanni Pistone *Politecnico di Torino*

Algebra of Reversible Markov Chains

A Markov Chain with stationary un-normalized probability π and transition matrix P_{xy} , $x, y \in V$, is reversible if the detailed balance conditions $\pi(x)P_{xy} = \pi(y)P_{yx}$, $x \neq y$, are satisfied. In turn, the detailed balance conditions generate a toric ideal of the ring $\mathbb{Q}[\pi(x), P_{xy}]$, so that reversible Markov Chains are the non-negative part of a toric variety. Elimination of the indeterminate π produces again a toric ideal, whose binomial generators are the Kolmogorov's equations on the closed paths of the undirected graph of the transition matrix P . We discuss the parameterization of this model and the effect of imposing a toric model on the π 's.

(Joint work with Maria Piera Rogantin)

Elena Stanghellini *Università degli Studi di Perugia*

On the identification of discrete graphical models with hidden nodes

Conditions are presented for local identifiability of discrete undirected graphical models with a binary hidden node. These models can be obtained by extending the latent class model to allow for conditional associations between the observed variables. The conditions are based on the topology of the undirected graph and rely on the faithfulness assumption. For models that are not locally identified we derive the expression of the parameter (sub)space where the identifiability breaks down.

Piotr Zwiernik *University of Warwick*

Posets, Möbius functions and tree-cumulants:

It has been noted by several authors that in the case of multivariate models cumulants often form a convenient system of coordinates. We investigate Bayesian networks on rooted trees where all variables in the system are binary and the inner nodes represent hidden variables. We show that in this case we can construct a more flexible change of coordinates. This change depends on classical results in the theory of partially ordered sets and it mirrors the combinatorial definition of cumulants. The new coordinates give us a good understanding of the structure of the models under consideration. The parameterization of the model becomes monomial which in particular allows us to obtain the equations and inequalities defining the model. There is also another direct application which I want to present briefly during this talk. The nice structure of the parameterization allows us to understand the identifiability issues for this class of models: the formulae for the estimators in the case when the model is identified and the structure of the MLE fibers in the case when it is not. This insight gives us also some understanding of the asymptotics for this class of models.

Hélène Massam *York University*

The geometry of hierarchical discrete loglinear models

Given a multi-way contingency table of counts, a standard problem is that of model selection in the class of hierarchical loglinear models. In Bayesian analysis, this is done through the Bayes factor which is the ratio of the posterior probabilities of two models given the data. When the prior distribution on the class of models is chosen to be uniform and the prior distribution on the loglinear parameters is the Diaconis-Ylvisaker conjugate prior, this problem is reduced to computing the ratio of normalizing constants for the prior and posterior distributions on the loglinear parameters. As a prior on the loglinear parameters, we use the prior defined by Massam, Liu and Dobra (2009) which comprises two hyperparameters. One of these hyperparameters is the equivalent sample size α which represents the total number of counts of a fictive contingency table. In practice, it is very interesting to take this equivalent sample size as small as possible as this has the effect of "regularizing" the selected models. In studying the behaviour of the Bayes factor as α tends to 0, we are led to study the geometry of the convex hull of the support of the multinomial hierarchical model. The faces of this convex hull are of particular importance. We show that as α tends to 0, the Bayes factor is equivalent to a power of α which we compute explicitly. This power depends on the position of the data point with respect to the faces of the convex hulls of the support of each one of the two multinomial hierarchical models.

Karim Anaya-Izquierdo *Open University*

Computational Information Geometry: Model Sensitivity and Approximate Cuts

Our project uses computational information geometry to develop diagnostic tools to help understand sensitivity to model choice by building an appropriate perturbation space for a given inference problem. Focused on generalised linear models, the long-term aim is to engage with statistical practice via appropriate R software encapsulating the underlying geometry, whose details the user may ignore.

This talk focuses on the role played by the concept of an approximate cut. Amongst other features, the perturbation space thereby constructed allows us to expose the differences in inferences resulting from different models, albeit that they may be equally well supported by the data. A running example illustrates the development.

EPSRC support under grant EP/E017878/1 is gratefully acknowledged.

(joint work with Frank Critchley, Paul Marriott and Paul Vos)

Nihat Ay *Max Planck Institute*

Information geometry of polytopes - general theory and applications to stochastic matrices

In my talk I review basic information-geometric concepts and extend them to polytopes. In particular, I discuss a notion of exponential families in polytopes and present results for the problem of maximizing the entropy distance from such an exponential family. Based on the general theory

I consider as example the polytope of stochastic matrices and the maximization of a generalized version of multi-information. A direct application of the above-mentioned results to the context of stochastic matrices gives an upper bound on the conditional entropy of a stochastic matrix with locally maximal multi-information.

Ian Dinwoodie *Duke University*

Syzygies for Metropolis Proposal Kernels

When the Metropolis algorithm is used to sample from a constrained set of points by targeting a concentrated exponential distribution, the base chain can sometimes be improved by computing generators of a syzygy module in advance.

Nanny Wermuth *Chalmers/University of Gothenburg*

Special features of probability distributions: which are detectable by algebraic methods?

Examples of special features of families of probability distributions are given and their consequences discussed. These include families without the intersection property, without the composition and incomplete families. The question is which of these special features can be detected with the help of algebraic methods.

Frank Critchley *Open University*

Emerging Geometries for Statistical Science: II. Future Directions

A few pointers are offered to promote discussion of future directions in geometric and algebraic methods in statistics.

POSTERS

John Aston *University of Warwick*

Geometry of Mutual Information in Three-Way Contingency Tables

Likelihood ratio (LR) tests for association and for interaction are examined for three-way contingency tables, in particular, the widely used $2 \times 2 \times K$ table. Mutual information identities are used to characterize the geometry of the information decomposition and the logical relationship between the omnibus LR test for conditional independence across K strata and its two independent components, the LR tests for interaction and for uniform association. The latter two tests are logically connected to formulating a natural two-step test for conditional independence. It is suggested to use the proposed two-step test with reduced nominal levels in contrast to the Breslow-Day test and the Cochran-Mantel-Haenszel test. This yields efficient interval estimation for both the interaction parameter and the common odds ratio, compared to using the Mantel-Haenszel estimate. This allows the development of power analysis for testing general hypotheses of varied interactions, using an invariant Pythagorean law of relative entropy.

Maurizio Filippone *University of Glasgow*

Inference for Gaussian Process Emulation of Nonlinear Differential Equation Models

Computer models aim at building a description of a system with some degree of accuracy. Combining observations from the system with results from simulations through computer codes in artificial scenarios, can provide a way to gain some understanding of the process underlying the system. We study the approach proposed in (Kennedy and O'Hagan, 2001), where the output of the real system is modeled as a combination of the output of the computer code and a term accounting for the model inadequacy, with both terms modeled using Gaussian Processes. Additionally, a simple i.i.d. noise model is considered to account for the noise in the measuring process. We apply this idea to computer codes based on nonlinear differential equation models, that are known to have complex posterior distributions over the input parameters. Advanced inference techniques are necessary to estimate the uncertainty in the model parameters along with the model inadequacy. We compare a variety of sampling techniques to obtain posterior distributions over the parameters of the model. In particular, we study the application of the recently proposed Riemannian Manifold Hamiltonian Monte Carlo by (Girolami et al, 2009) to this problem.

Helene Gehrman *University of Oxford*

Topology of the Set of Vertex and Edge Colored Graphs Representing Graphical Gaussian Models with Symmetry Constraints

Højsgaard and Lauritzen (2008) introduced three types of graphical Gaussian models with symmetry constraints on their parameters: equality between specified elements of the concentration matrix (RCO), equality between specified partial correlations (RCOR) and restrictions that are generated by permutation symmetry (RCOP). The models can be represented by vertex and edge colored graphs, where parameters that are associated with equally colored vertices or edges are restricted

to being identical. The set of vertex and edge colored graphs on a fixed vertex set V denoted by \mathcal{C}_V is the central object of our study. We show that \mathcal{C}_V is a lattice with respect to model inclusion of the represented models and give its meet and join operations. We further study subsets of \mathcal{C}_V which are of statistical interest: (i) Π_V , the set of graphs representing RCOP models, (ii) B_V , the set of graphs which yield the same model restrictions independently of whether they determine an RCON or RCOR model, (iii) P_V , the set of graphs yielding equality of maximum likelihood and least squares estimators for the model means, and (iv) $R_V = B_V \cap P_V$. We show that each of the four sets is a lattice and give their meet and join operations. Among other results on the set structures we show that each graph not lying in one of the four subsets above has a supremum inside the set. We have implemented the computation of the supremum for Π_V and B_V in a computer program and are in the process of doing so for P_V and R_V .

Luigi Malagò *Politecnico di Milano*

Towards the Geometry of Stochastic Relaxation: Optimization of Pseudo-Boolean Functions by Estimation of Covariances

We want to find the minimum of a pseudo-Boolean function, that is, a real-valued function defined over a vector of binary variables. These functions are also known in the literature as set functions. We introduce the notion of stochastic relaxation, i.e., we look for the minima of a pseudo-Boolean function by minimizing its expected value over a set of probability densities. By doing this, the parameters of the statistical model become the new variables of the optimization problem. The new continuous optimization problem also appears in the literature in the context of stochastic optimization. We choose statistical models that belong to the exponential family, and we justify this choice with results related to the characterization of its tangent space and topological closure. We also provide a characterization of the stationary points of the relaxed problem, together with a study of the minimizing sequences with reduced support. Finally, we show how stochastic relaxation can be interpreted as a unifying framework for the analysis of classical techniques in linear programming and stochastic methods in optimization. In particular we describe a stochastic algorithm where the gradient of the expected value of the pseudo-Boolean function is estimated by empirical covariances between the function and the sufficient statistics of the exponential models.

Hugo Maruri-Aguilar *LSE*

Holes and aberration

Algebraic techniques have been used to identify polynomial models in experimental design. A feature of models identified with such techniques is that they are of low weighted degree. This lead to a generalisation of the idea of aberration, i.e. algebraic models are of low aberration. when designs are fractions of 2^k factorial designs, we can further describe models from an elementary topological viewpoint. That is, models are simplicial complexes and the quantity and quality of holes of such complexes is described by Betti numbers. In summary, our poster presents two forms of describing models: by degree and by connectivity. Some examples are included.

Maria Sofia Massa *Università degli Studi di Padova*

Towards algebraic combinations of Gaussian graphical models

Massa and Lauritzen (2010) introduce a general framework for combining families of distributions and then specialize it for the case of Gaussian graphical models. Firstly, I will review some of the laws useful for combining two Gaussian graphical models and the distinction between graphical and non-graphical combinations. Graphical combinations are themselves Gaussian graphical models and therefore they can be described by means of an undirected graph G and a multivariate normal distribution satisfying the global Markov property with respect to G . This does not hold for combinations of non-graphical type where the law of the combination does not represent a Gaussian graphical model. In this context, I will present some examples of graphical and non-graphical combinations. Then, I will propose an approach for non-graphical combinations based on performing combinations of algebraic type. Here, the intent is to investigate all constraints induced on the common variables and describe them by means of algebraic techniques.

Johannes Rauh *Max Planck Institute*

Support Sets in Exponential Families and Oriented Matroid Theory

My poster presents results from a joint work with Nihat Ay and Thomas Kahle (preprint available at arXiv:0906.5462). We study how results from algebraic statistics generalize to the case of non-algebraic exponential families on finite sets. Here an exponential family is called algebraic if it has an integer-valued matrix of sufficient statistics. In this case the exponential family is the intersection of an algebraic variety with the probability simplex, which makes available the powerful tools of computational commutative algebra. While most relevant examples of exponential families are algebraic, it turns out that ignoring the algebraic properties yields another viewpoint which better captures the continuity aspects of exponential families. A lot of properties can be deduced from an oriented matroid naturally associated to the exponential family: The closure of a discrete exponential family is described by a finite set of equations corresponding to the circuits of this matroid. These equations are similar to the equations used in algebraic statistics, although they need not be polynomial in the general case. This description allows for a combinatorial study of the possible support sets in the closure of an exponential family. In particular, if two exponential families induce the same oriented matroid, then their closures have the same support sets. Finally we find a surjective (but not injective) parametrization of the closure of the exponential family by adding more parameters. These parameters also have an interpretation in the matroid picture. The parametrization generalizes the monomial parametrization used in algebraic statistics in the case of algebraic exponential families.

Eva Riccomagno *Università degli Studi di Genova*

Hermite polynomial aliasing in Gaussian quadrature

A representation of Hermite polynomials of degree $2n + 1$, as sum of an element in the polynomial ideal generated by the roots of the Hermite polynomial of degree n and of a remainder, suggests a folding of multivariate polynomials over a finite set of points. From this, the expectation of some polynomial combinations of random variables normally distributed is computed. This is related to

quadrature formulas and has strong links with designs of experiments.

(joint work with G. Pistone)

Germain Van Bever *Université Libre de Bruxelles*

Shape depth

The concept of depth introduced in Tukey (1975) is today regarded as one of the most powerful method for multivariate data analysis. Depth was strictly restricted to the location setup until the extension to linear regression models by Rousseuw and Hubert (1999), then the extension to any parametric model by Mizera (2002). In view of the success of location depth, it is puzzling that the properties of general parametric depth remain virtually unexplored, but maybe for the robustness properties of the resulting maximum depth estimators. As a first effort to study the nature of depth in general parametric models, and in particular its interplay with semiparametric inference, we consider elliptical models indexed by a shape matrix and a radial density. The shape matrix here is essentially a normalized version (with respect to a scale functional) of the covariance matrix, and is the natural parameter of interest in various problems, such as, e.g., in principal component analysis. We study the concept of shape depth associated with the general definition of Mizera (2002). We mainly focus on equivariance properties and on the independence of shape depth with respect to the scale functional used.

(joint work with Davy Paindaveine)
