Computational Information Geometry and Graphical Models

Frank Critchley[1] and Paul Marriott[2]

[1]The Open University

[2]University of Waterloo

Workshop on Graphical Models:
Mathematics, Statistics and Computer Science
April 16-18, 2012

Overview

- Information Geometry - our embedding approach
- Focus on computation (CIG)
- Boundaries - essential for graphical models
- MCMC - geometric approach
- Curvature and mixture models
- Simplicial asymptotics - lack of uniformity

Acknowledgements

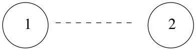
- Paul Vos
- Karim Anaya-Izquierdo
- Mark Girolami
- Simon Byrne
- EPSRC
- NSERC

Information Geometry

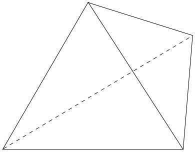
- Information geometry (M, g, ∇^{α}) [20] intrinsic
- We work by embedding in ‘space of all possible models’
- Models can be simplicial rather than manifolds: non-constant dimension
- Operational means finite dimensional
- Computational Information Geometry on extended multinomials

Discrete models

- Basic model have set of binary random variables



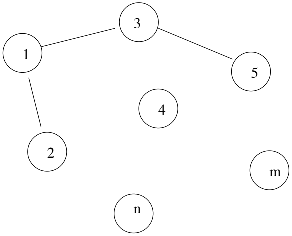
- Look at all possible joint distributions: simplex



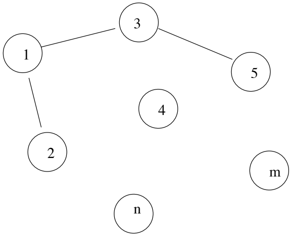
- Models are sub-families of simplex

Extended Multinomial

- Look at discrete graphical models

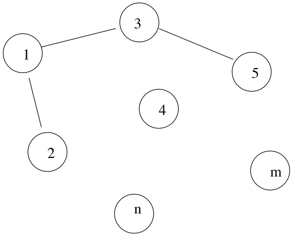


Extended Multinomial



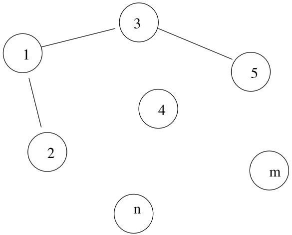
- Look at discrete graphical models
- Space of distributions simplicial

Extended Multinomial



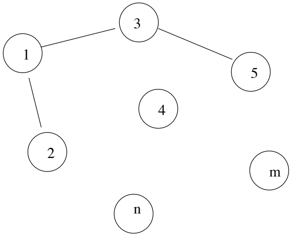
- Look at discrete graphical models
- Space of distributions simplicial
- Boundaries where probabilities are zero

Extended Multinomial



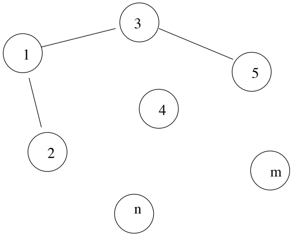
- Look at discrete graphical models
- Space of distributions simplicial
- Boundaries where probabilities are zero
- Information geometry of extended multinomial models

Extended Multinomial



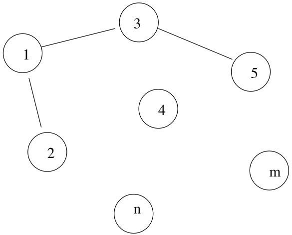
- Look at discrete graphical models
- Space of distributions simplicial
- Boundaries where probabilities are zero
- Information geometry of extended multinomial models
- applications to graphical models and elsewhere

Extended Multinomial



- Look at discrete graphical models
- Space of distributions simplicial
- Boundaries where probabilities are zero
- Information geometry of extended multinomial models
- applications to graphical models and elsewhere
- proxy for space of all models

Extended Multinomial

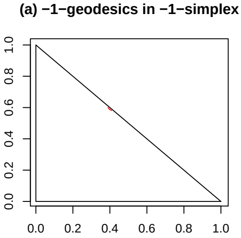


- Look at discrete graphical models
- Space of distributions simplicial
- Boundaries where probabilities are zero
- Information geometry of extended multinomial models
- applications to graphical models and elsewhere
- proxy for space of all models
- IG all explicit

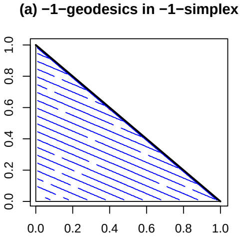
Information Geometry

- How to connect two probability density or mass functions $f(x)$ and $g(x)$ in some space of models?
  - -1: $\rho f(x) + (1-\rho)g(x)$
  - +1: $\frac{f(x)^{\rho} g(x)^{1-\rho}}{C(\rho)}$
- Two different affine structures used simultaneously
  - -1: Mixture affine geometry on unit measures
  - +1: Exponential affine geometry on positive measures
- Fisher Information's roles
  - measures angles and lengths
  - maps between $+1$ and -1 representations of tangent vectors, [3], [4], [19]

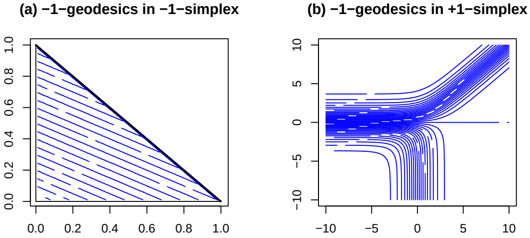
Visualising IG: extended trinomial example



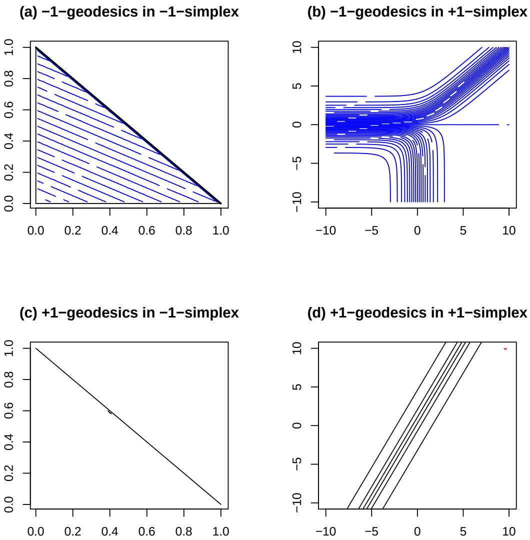
Visualising IG: extended trinomial example



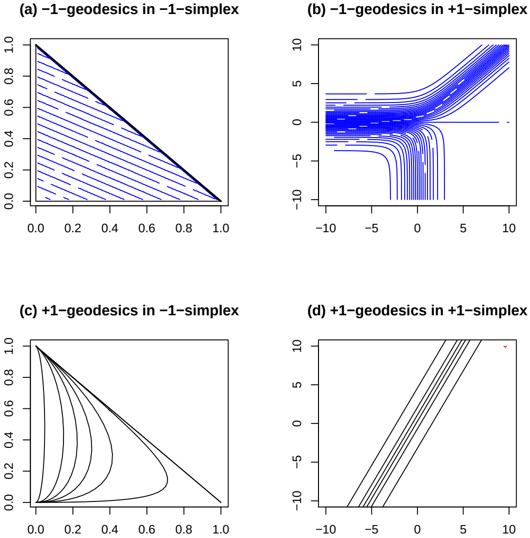
Visualising IG: extended trinomial example



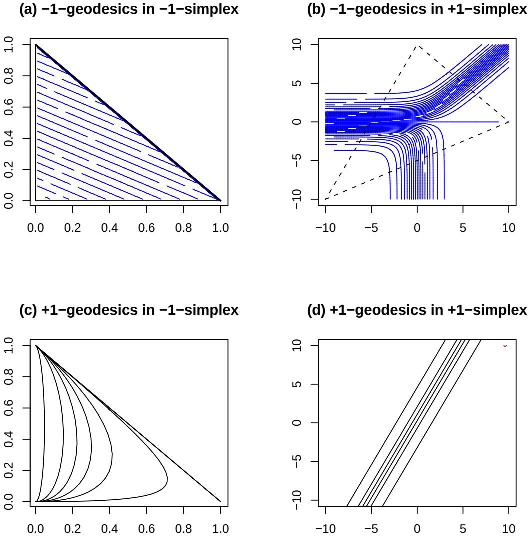
Visualising IG: extended trinomial example



Visualising IG: extended trinomial example

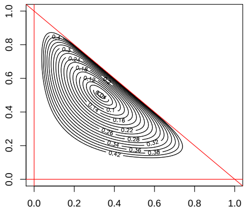


Visualising IG: extended trinomial example



Riemannian Geometry

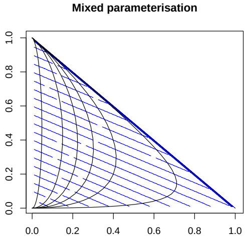
- The 0-geometry defined by the Fisher information metric
- Look at 0-geodesic spheres in simplex



- These are smoothly attached to boundary: *c.f.* ± 1-geodesics
- Can use this smooth structure in MCMC

Duality

- There exists a mixed parameterisation [6] as solution of differential equation



- -1-geodesics Fisher orthogonal to $+1$-geodesics
- Limit of mixed parameters give extended exponential family
- Key to structural theorem [3] and idea of inferential cuts

Geometry of likelihood

For sparse extended multinomial models:

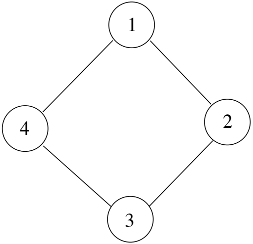
- Quadratic approximations to log-likelihood fail globally
- Many -1-flat directions
- MCMC
- Asymptotics

Examples: Graphs, Networks and IG

- Full exponential families [17]
- Very high-dimensional models [21] - MCMC is one tool here
- Curved exponential families - curvature can be very high
- Closure of exponential families- boundaries

Graphical models: FEF

- Consider the example from [14] of the cyclic graph of order 4 with binary values at each node



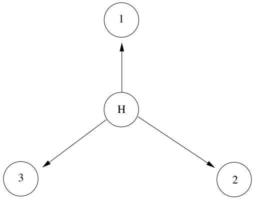
- Models lie in 15-dimensonal simplex, but with constraints imposed by conditional independence
- Constraints linear in $+1$-affine parameters

$$\eta_i + \eta_j = \eta_k + \eta_l$$

- So get 7-dimensional full exponential family

DAG with hidden variables

- In multinomials independence is expressible as a finite set of polynomial equalities
- Add hidden variables



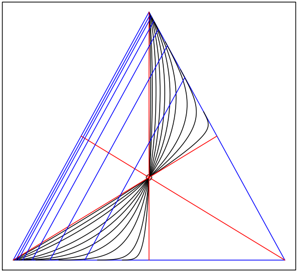
- Example lies in 7 dimensional simplex- mixes over a 3 dimensional CEF
- The model space is not a manifold but a variety- union of different dimensional manifolds- extended exponential family

Attaching to the boundaries

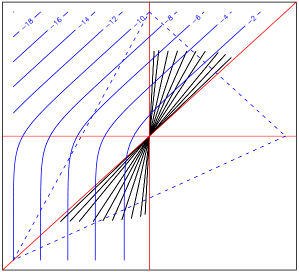
- Models are low-dimensional families within high dimensional simplexes
- Need to understand how models are attached to boundaries
- Extended exponential families: see also [30], [9]
- Need to be able to compute limit points in computationally efficient way
- Use linear programming and convex geometry techniques
- The following plot is generic

Attaching to the boundaries

Minus 1 plot



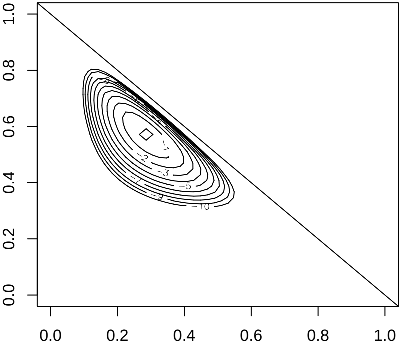
Plus 1 plot



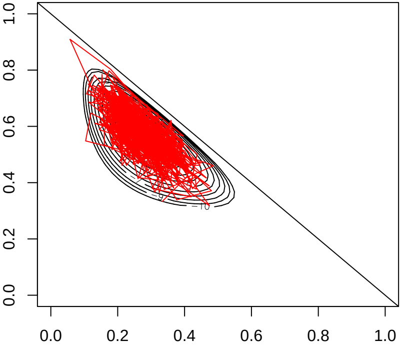
Geometric MCMC

- When doing MCMC on models the boundaries matter
- Riemann manifold Metropolis adjusted Langevin algorithm, [13] uses Fisher metric structure.
- Use a Metropolis-Hastings where the random-walk proposals have variance determined by Fisher metric
- Have seen how the 0-geometry smoothly attaches the boundaries
- Other geometric approaches are under active development - cuts and mixed parameterisations

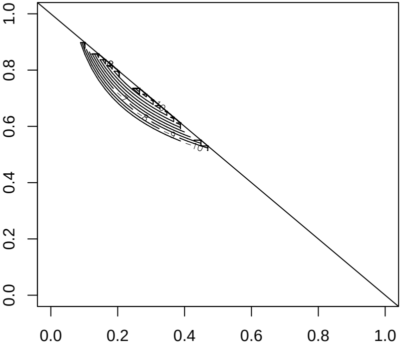
Geometric MCMC



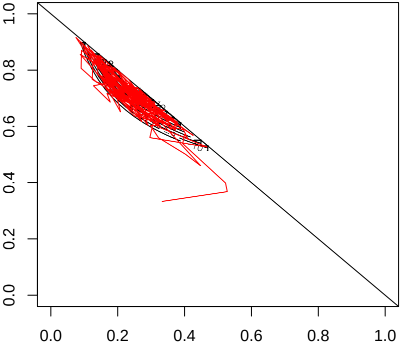
Geometric MCMC



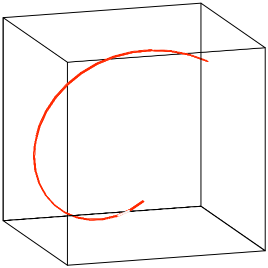
Geometric MCMC



Geometric MCMC

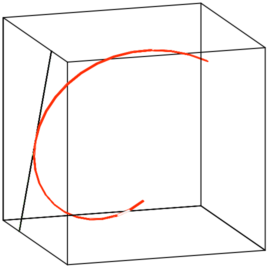


Embedding curvature and affine approximation



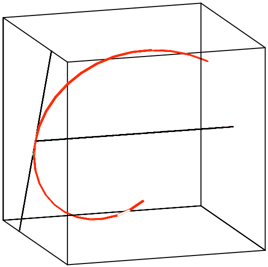
- Curvature(s) key part(s) of differential geometry
- Tangent space gives best linear approximation
- Tangent and curvature gives best two dimensional affine embedding space

Embedding curvature and affine approximation



- Curvature(s) key part(s) of differential geometry
- Tangent space gives best linear approximation
- Tangent and curvature gives best two dimensional affine embedding space

Embedding curvature and affine approximation



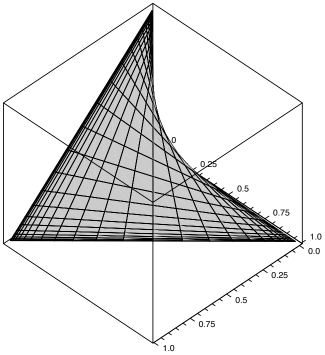
- Curvature(s) key part(s) of differential geometry
- Tangent space gives best linear approximation
- Tangent and curvature gives best two dimensional affine embedding space

Curvature and Dimension Reduction

- Dimension reduction (best approximating subspaces) via tangent and curvature
- Different affine geometries give different dimension reduction
- Low dimensional $+1$-affine spaces give approximate sufficient statistics [28]
- Low dimensional -1 approximations give limits to identification and computation in mixture models [27], [2]

Example: tripod and bipod

- Use -1-affine approximations
- tri and bi pod example are ruled surfaces: exploit this for computing 2-hull



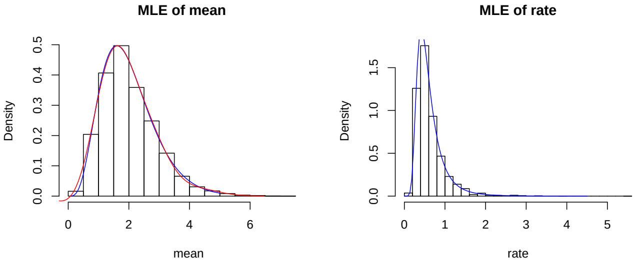
- IG gives ways to explore convex hull efficiently

Mixing paradox

- There is a paradoxical aspect to mixing in high dimensional extended multinomial model
- The convex hull of any open interval of a one-dimensional exponential family is of full dimension $\cdots$
- This is due to total positivity
- $\cdots$ But, there exist low dimensional approximations to this convex hull based on curvature, [2]

Asymptotic expansions

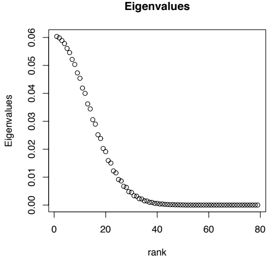
- Strong links between IG and higher order asymptotic expansions [7]
- Can apply Edgeworth, saddlepoint or Laplace expansions [32]



- Flexible, tractable given IG, invariance properties clear [3]

Asymptotic expansions

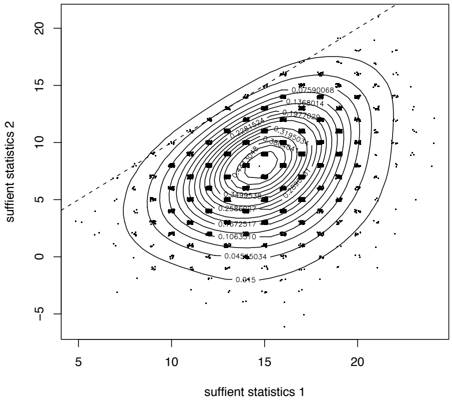
- High dimensional calculus through tensor analysis, McCullagh [29]
- Many terms need to be computed in high dimensional problems
- Singularity of Fisher information matters



- Fisher information can be singular (or infinite) [24]

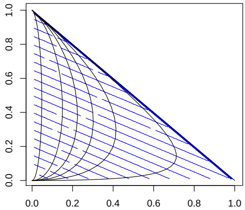
Asymptotic expansions

- Works in high dimensional examples
- Edgeworth: boundaries and discretisation effects
- Laplace expansion can be problematic in high-dimension- spectrum of FI



Asymptotic expansions

- First order Normal approximations require controlling statistical curvature
- As near boundary statistical curvature can be unbounded - most asymptotic formula not uniform across simplex



- Higher order corrections define the area where first order formula can be used
- Note centre of very high-dimensional simplex also problematic- get discretisation effects

Summary

- Information Geometry - our embedding approach
- Focus on computation (CIG)
- Boundaries - essential for graphical models
- MCMC - geometric approach
- Curvature and mixture models
- Simplicial asymptotics - lack of uniformity
- CIG useful generally in statistical modelling

References I

[1] Anaya-Izquierdo, K., Critchley, F, Marriott P. & Vos P. (2010), Towards information geometry on the space of all distributions, *Preprint*

[2] Anaya-Izquierdo, K and Marriott, P. (2007) Local mixtures of Exponential families, *Bernoulli* Vol. 13, No. 3, 623-640.

[3] Amari, S.-I. (1985). *Differential-Geometrical Methods in Statistics*. Lecture Notes in Statistics, No. 28, New York: Springer.

[4] Amari, S.-I. and Nagaoka, H. (2000). *Methods of Information Geometry*. Providence, Rhode Island: American Mathematical Society.

[5] Barndorff-Nielsen, O., (1978) *Information and exponential families in statistical theory*, London: John Wiley & Sons

[6] Barndorff-Nielsen, O. E. and Blaesild, P. (1983). *Exponential models with affine dual foliations*. Annals of Statistics, 11(3):753–769.

[7] Barndorff-Nielsen, O.E. and Cox, D.R, (1994), *Inference and Asymptotics*, Chapman & Hall:London

References II

[8] Brown, L. D. (1986). *Fundamentals of statistical exponential families: with applications in statistical decision theory*, Institute of Mathematical Statistics

[9] Csiszar, I. and Matus, F., (2005). Closures of exponential families, *The Annals of Probability*, 33(2):582–600

[10] Efron, B. (1975). Defining the curvature of a statistical problem (with applications to second order efficiency), *The Annals of Statistics*, 3(6):1189–1242

[11] Fukumizu, K. (2005). Infinite dimensional exponential families by reproducing kernel hilbert spaces. *Proceedings of the 2nd International Symposium on Information Geometry and its Applications*,p324-333.

[12] Gibilisco, P. and Pistone, G. (1998). Connections on non-parametric statistical manifolds by orlicz space geometry. *Infinite Dimensional Analysis, Quantum Probability and Related Topics,* 1(2):325-347.

References III

[13] Girolami M. and Calderhead, B. (2011), Riemann manifold Langevin and Hamiltonian Monte Carlo methods, J. R. Statist. Soc. B 73, Part 2, pp. 1-37

[14] D. Geiger, D.Heckerman, H.King and C. Meek (2001) Stratified Exponential Families: Graphical Models and Model Selection, *Annals of Statistics*, Vol. 29, No. 2, pp 505-529

[15] Grunwald P.D. and Dawid A.P. (2004) Game theory, maximum entropy, minimum discrepancy and robust Bayesian decision theory, *Annals of Statistics*, 32, 4 1267-1433

[16] Hand, D.J., Daly, F., Lunn, A.D., McConway, K.J. and Ostrowski, E. (1994), A handbook of small data sets , Chapman & Hall, London.

[17] D. Hunter, (2007) Curved exponential family models for social networks, *Social Networks*, 29 216–230.

[18] Hunter D, Handcock M. (2006) Inference in curved exponential family models for networks *J. of Computational and Graphical Statistics* 15, 565-583

References IV

[19] Kass, R. E. and Vos, P. W. (1997). *Geometrical Foundations of Asymptotic Inference*. New York: Wiley.

[20] Lauritzen, S. L. (1987). *Statistical manifolds*, in Differential geometry in statistical inference: IMS

[21] Lazega, E., Pattison, P.E., (1999) Multiplexity, generalized exchange and cooperation in organizations: a case study. Social Networks 21, 67–90

[22] L. L. Kupper L.L., and Haseman J.K., (1978), The Use of a Correlated Binomial Model for the Analysis of Certain Toxicological Experiments, *Biometrics*, Vol. 34, No. 1 (Mar., 1978), pp. 69-76

[23] Mary L. Lesperance and John D. Kalbfleisch (1992) An Algorithm for Computing the Nonparametric MLE of a Mixing Distribution*JASA* Vol. 87, No. 417 (Mar., 1992), pp. 120-126

[24] Li P., Chen J., & Marriott P., (2009) Non-finite Fisher information and homogeneity: the EM approach, *Biometrika* 96, 2 pp 411-426.

References V

[25] Lindsay, B.G. (1995). *Mixture models: Theory, Geometry, and Applications*, Hayward CA: Institute of Mathematical Sciences.

[26] Marriott P and West S, (2002), On the Geometry of Censored Models, *Calcutta Statistical Association Bulletin* 52, pp 235-250.

[27] Marriott, P (2002), On the local geometry of Mixture Models, *Biometrika*, 89, 1, pp 77-89

[28] Marriott, P., & Vos, P. (2004), On The Global Geometry of Parametric Models and Information Recovery, *Bernoulli*, **10** (2), 1-11

[29] McCullagh P. (1987) *Tensor Methods in Statistics*, Chapman and Hall, London.

[30] Rinaldo, Alessandro and Fienberg Stephen (2009), *On the geometry of discrete exponential families with applications to exponential random graph models*, EJS, 3, pp. 446-484

[31] Snijders, T. A. B. (2002), Markov Chain Monte Carlo Estimation of Exponential Random Graph Models, *Journal of Social Structure*, 3.

References VI

[32] Small, C.G. (2010) *Expansions and asymptotics for statistics*, Chapman and Hall

[33] Strauss, D., and Ikeda, M. (1990), Pseudolikelihood Estimation for Social Networks,*Journal of the American Statistical Association* , 85, 204–212.

[34] Shun Z and McCullagh P. (1995) Laplace approximations of high dimensional integrals JRSS B