Computational Information Geometry and Graphical Models

Frank Critchley¹ and Paul Marriott²

¹The Open University
²University 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 \((M, g, \nabla^\alpha)\) [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

(a) \(-1\)-geodesics in \(-1\)-simplex

(b) \(-1\)-geodesics in \(+1\)-simplex

(c) \(+1\)-geodesics in \(-1\)-simplex

(d) \(+1\)-geodesics in \(+1\)-simplex
Visualising IG: extended trinomial example

(a) $-1$-geodesics in $-1$-simplex
(b) $-1$-geodesics in $+1$-simplex
Visualising IG: extended trinomial example

(a) $-1$-geodesics in $-1$-simplex

(b) $-1$-geodesics in $+1$-simplex

(c) $+1$-geodesics in $-1$-simplex

(d) $+1$-geodesics in $+1$-simplex
Visualising IG: extended trinomial example

(a) \(-1\)-geodesics in \(-1\)-simplex

(b) \(-1\)-geodesics in \(+1\)-simplex

(c) \(+1\)-geodesics in \(-1\)-simplex

(d) \(+1\)-geodesics in \(+1\)-simplex
Visualising IG: extended trinomial example

(a) $-1$-geodesics in $-1$-simplex

(b) $-1$-geodesics in $+1$-simplex

(c) $+1$-geodesics in $-1$-simplex

(d) $+1$-geodesics in $+1$-simplex
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

\[ 1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 1 \]

- Models lie in 15-dimensional 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
- 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.
- This is due to total positivity.
- 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


References II


References III


References IV


References VI


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