Quad-trees, efficient networks, inferred fibres

2011 W-CAS Afternoon

Wilfrid S. Kendall
w.s.kendall@warwick.ac.uk

Department of Statistics, University of Warwick

1st December 2011

Introduction

This talk gives brief sketches of three snippets of research.

1. Quad-trees (Kendall and Wilson 2003): image analysis in depth;
2. Efficient networks (Aldous and WSK 2008): how to build fast networks that connect efficiently;
3. Inferred fibres (Hill, Kendall, and Thönnes 2011): guessing curves given an associated point pattern.

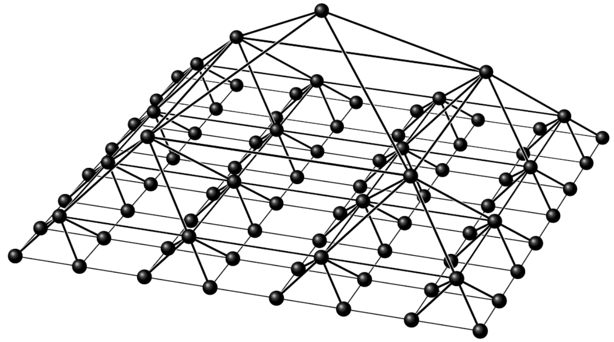
   Common theme: Using probability to build useful models.

Ising images

Consider modelling a binary image using an Ising model.

1. Do this by envisioning an ideal image as a finite cartesian lattice, with bond strengths J_1 expressing the thought that neighbouring pixels are likely to be similar (a "local Bayesian prior"). The actual observed image is a duplicate lattice in which neighbouring pixels are un-related to each other, but relate to corresponding ideal pixels by bonds of strength K.
2. We can simulate this using the heat bath algorithm. However we wish to simulate from the Ising model *conditioned* on the observed noisy image. Because the heat bath algorithm is reversible, we simply fix observed pixels!
3. The heat-bath algorithm can be viewed as a collection of correlated but very simple reflected random walks.

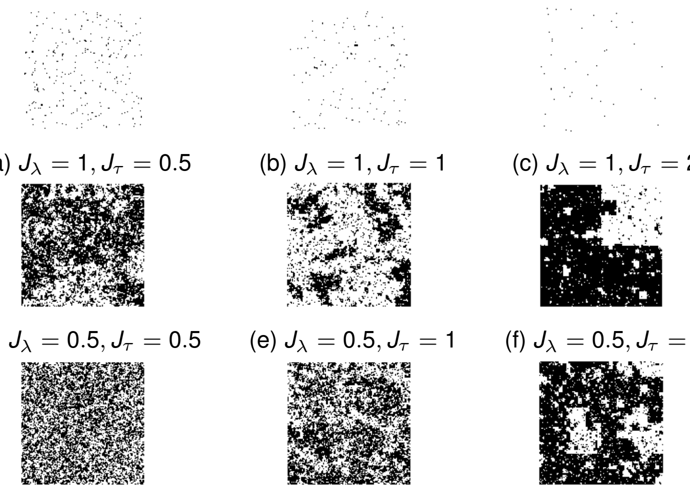
Multiresolution (I)



Kendall and Wilson (2003): Ising model built on a *quadtree*; different parent-child and horizontal neighbour connection strengths (J_τ, J_λ).
Question: in which range of parameters is the model suitable for image analysis?

Multiresolution (II)

http://www.dcs.warwick.ac.uk/~rgw/sira/sim.html



(a) $J_\lambda = 1, J_\tau = 0.5$ (b) $J_\lambda = 1, J_\tau = 1$ (c) $J_\lambda = 1, J_\tau = 2$

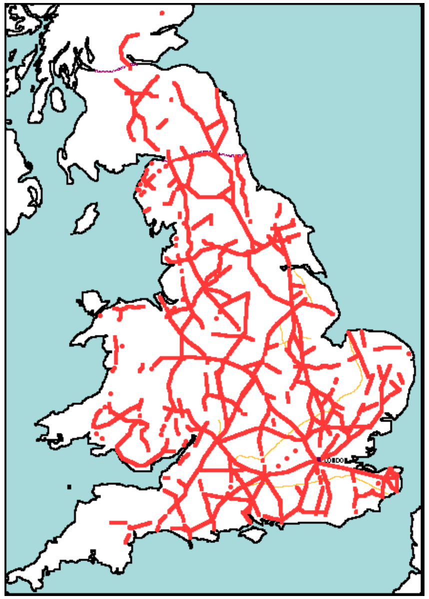
(d) $J_\lambda = 0.5, J_\tau = 0.5$ (e) $J_\lambda = 0.5, J_\tau = 1$ (f) $J_\lambda = 0.5, J_\tau = 2$

(g) $J_\lambda = 0.25, J_\tau = 0.5$ (h) $J_\lambda = 0.25, J_\tau = 1$ (i) $J_\lambda = 0.25, J_\tau = 2$

Best region is around the bottom right-hand corner.

Introduction | Quad-trees | Efficient networks | Inferred fibres | Conclusion | References

An ancient optimization problem



A Roman Emperor's dilemma:

PRO: Roads are needed to move legions quickly around the country;

CON: Roads are expensive to build and maintain;

Pro optimo quod faciendum est?

Introduction | Quad-trees | Efficient networks | Inferred fibres | Conclusion | References

A problem in frustrated optimization

Consider N cities $x^{(N)} = \{x_1, \ldots, x_N\}$ in square side $\sqrt{N}$.
Assess road network $G = G(x^{(N)})$ connecting cities by:

- network total road length $\text{len}(G)$ (minimized by Steiner minimum tree $\text{ST}(x^{(N)})$); versus
- average network distance between two random cities,

$$\text{average}(G) = \frac{1}{N(N-1)} \sum\sum_{i \neq j} \text{dist}_G(x_i, x_j),$$

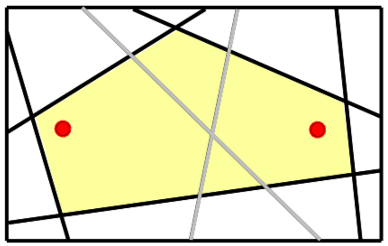
(minimized by laying tarmac for complete graph).

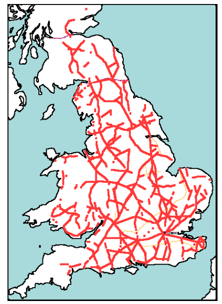
- Perhaps the average ratio would be a good measure of performance?

$$\frac{1}{N(N-1)} \sum\sum_{i \neq j} \frac{\text{dist}_G(x_i, x_j)}{\|x_i - x_j\|}$$

Introduction | Quad-trees | Efficient networks | Inferred fibres | Conclusion | References

Answer to first question (I)





- Augment Steiner tree by a low-intensity invariant Poisson line process Π_1.
- Unit intensity is $\frac{1}{2}\,\mathrm{d}\,r\,\mathrm{d}\,\theta$: we will use this and scale.
- Pick two cities x and y at distance $n = \sqrt{N}$ units apart.
  Remove lines separating the two cities;
  focus on cell $\mathcal{C}_{x,y}$ containing the two cities.

Asymptotics

Theorem

Careful asymptotics for $n \to \infty$ show that

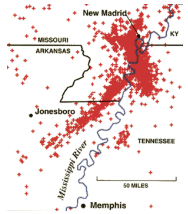
$$\mathbb{E}\left[\tfrac{1}{2}\,\text{len}\,\partial\mathcal{C}_{x,y}\right] \quad = \quad n + \tfrac{1}{4}\iint_{\mathbb{R}^2}(\alpha - \sin\alpha)\exp\left(-\tfrac{1}{2}(\eta - n)\right)\mathrm{d}\,z \quad \approx \quad n + \frac{4}{3}\left(\log n + \gamma + \frac{5}{3}\right)$$

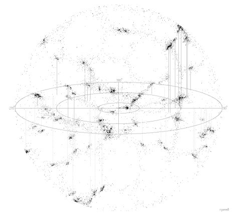
where $\gamma = 0.57721\ldots$ is the Euler-Mascheroni constant.

Thus a unit-intensity invariant Poisson line process is within $O(\log n)$ of providing connections which are as efficient as Euclidean connections. Sparse versions allow us to modify Steiner tree networks at infinitesimal cost to be only logarithmically worse than complete Euclidean graphs!

Three typical application contexts:

Inferring fibre structure from associated point sets:

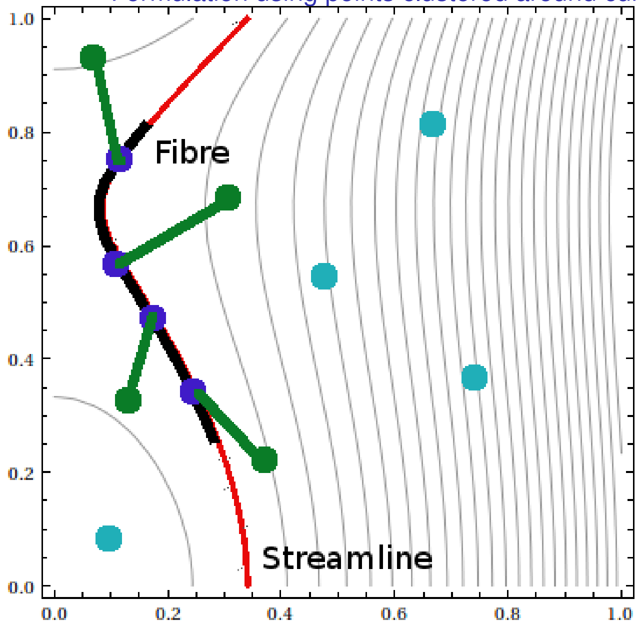




1. Fingerprint sweat pores
   Extracted from fingerprint a002-5 from NIST Special database 30 (Watson 2001).
2. Earthquake epicentres
   Epicentres in New Madrid region,
   taken from CERI (Center for Earthquake Research and Information).
3. Universe within 500 Mly
   Image: Richard Powell (`atlasoftheuniverse.com/nearsc.html`:
   Creative Commons Attribution-ShareAlike 2.5 License).

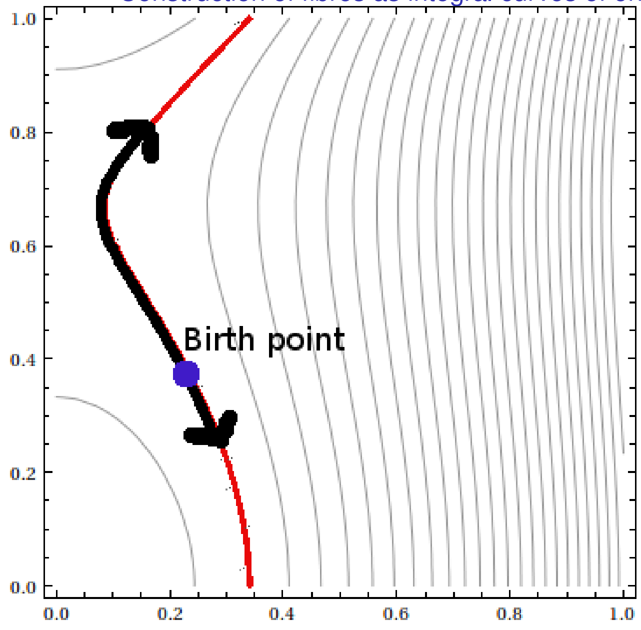
Our statistical model (I)

Formulation using points clustered around curvilinear fibres



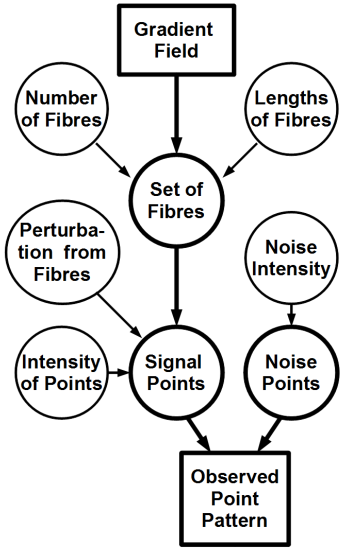
Our statistical model (II)

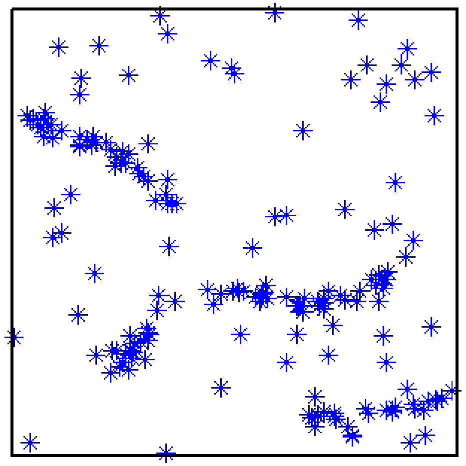
Construction of fibres as integral curves of orientation field



Our statistical model (III)

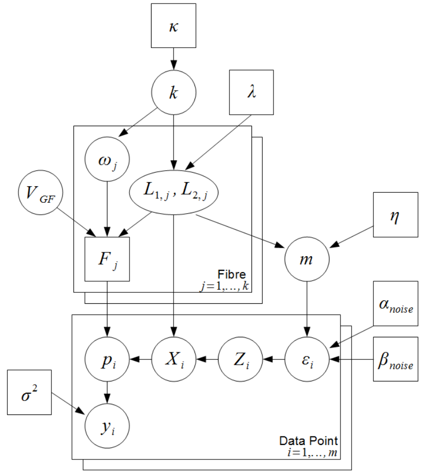
Building up a (simplified) DAG





Our statistical model (IV)

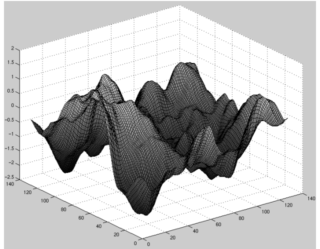
DAG of full model



Use MCMC for statistical analysis (variant of “heat-bath”)

Orientation Field

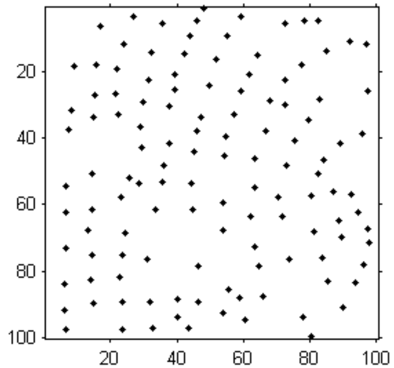
- Calculating an appropriate orientation field is key.
- Possible approaches include using random field theory, *eg* extending a Gaussian field . . .

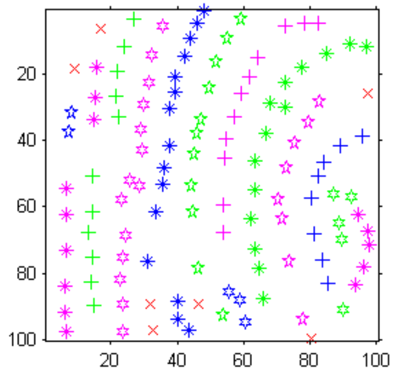


- . . . but the configuration space of orientation fields is huge.
- Use Empirical Bayes to evade resulting problems.

Fingerprints

Estimate of clustering of signal points





Advantage of statistically principled approach:
we can get a quantitative handle on random variation.

Conclusion

Bespoke probability models for interesting situations.

Aldous, D. J. and WSK (2008, March).
Short-length routes in low-cost networks via Poisson line patterns.
Advances in Applied Probability 40(1), 1–21.

Hill, B. J., W. S. Kendall, and E. Thönnes (2011).
Fibre-generated Point Processes and Fields of Orientations.
arXiv(1109.0701), 31.

Kendall, W. S. and R. G. Wilson (2003, March).
Ising models and multiresolution quad-trees.
Advances in Applied Probability 35(1), 96–122.

Stoyan, D., WSK, and J. Mecke (1995).
Stochastic geometry and its applications (Second ed.).
Chichester: John Wiley & Sons.

Watson, C. (2001).
Dual Resolution Images from Paired Fingerprint Cards.