Statistics at the University of Edinburgh

Research Group Overview

- Statistics lies at the heart of Data Science.
- Broad interests in applied and theoretical areas.
- 8 Statistics members based in the School of Mathematics (including 2 arriving in December) with a range of research interests.
- Lead innovative statistical research by developing novel methodological and theoretical techniques and apply modern tools to different fields.

Research Areas and Applications

- **Computational Statistics:** Efficient Bayesian and classical model-fitting approaches; new computational algorithms.
- **Statistical Modelling and Methods:** clustering; dimension reduction; extremes; graphical modelling; hidden Markov models; hierarchical models; high-dimensional modelling; inverse problems; missing data; state-space models.
- **Applications:** astrophysics; biostatistics; ecology; environment; finance; genomics; medicine.

$$\pi(\theta|x) = \frac{f(x|\theta)p(\theta)}{f(x)}$$

Prof Ruth King
School of Mathematics

(Statistics)

$$f(\boldsymbol{x}|N, \boldsymbol{\theta}, \boldsymbol{\epsilon}_{1:n}) \propto \prod_{i=1}^{n} f(\boldsymbol{x}_i|\boldsymbol{\theta}, \boldsymbol{\epsilon}_i) \times \frac{N!}{(N-n)!}(1-p^*)^{N-n}$$

Who am I?

- I am an applied statistician – motivated by real data and associated questions of interest.
- To answer the questions we need to:
  - develop sensible and realistic models that describe the main underlying characteristics of the data;
  - fit these models to data in a tractable manner; and
  - interpret the data accordingly.

Research interests

- Bayesian inference;
- Missing data (a very common problem!);
- Hidden (semi-)Markov models (these are really cool!);
- State-space models;
- Incorporating individual heterogeneity;
- Integrated models;
- Applications to ecology and epidemiology.

Ecology

- The field of ecology is very data rich (think of all the times your biology colleagues are on fieldwork!).
- However, there is often as associated lack of skills to fully analyse the collected data.
- New statistical models and associated model-fitting tools are rapidly being developed to provide robust analyses of the available data.

Case study 1

- Capture-recapture studies are often collected on wildlife populations when abundance and/or survival probabilities are of interest.
- These involve observers going into the field at a series of capture events.
- At each capture event all observed individuals are uniquely identified (possibly by attaching a ring/tag), recorded and released.

Data

- The data correspond to the encounter history of each individual observed within the study.
- An example encounter history:

1 0 0 0 1 1 0 2

- Additional individual-level information may also be recorded e.g. weight, “state”.

Questions of interest

- How can we incorporate previous states (e.g. breeding/non-breeding) in to current survival probabilities when individuals move between states?
- Do individuals exhibit "memory" with regard to their changes in state (typically assume change in state only depends on current state)? And if so how can we also include additional covariate information?
- Given the above models, how can we efficiently fit the models to real data – particularly as datasets increase in size?

Case study 2

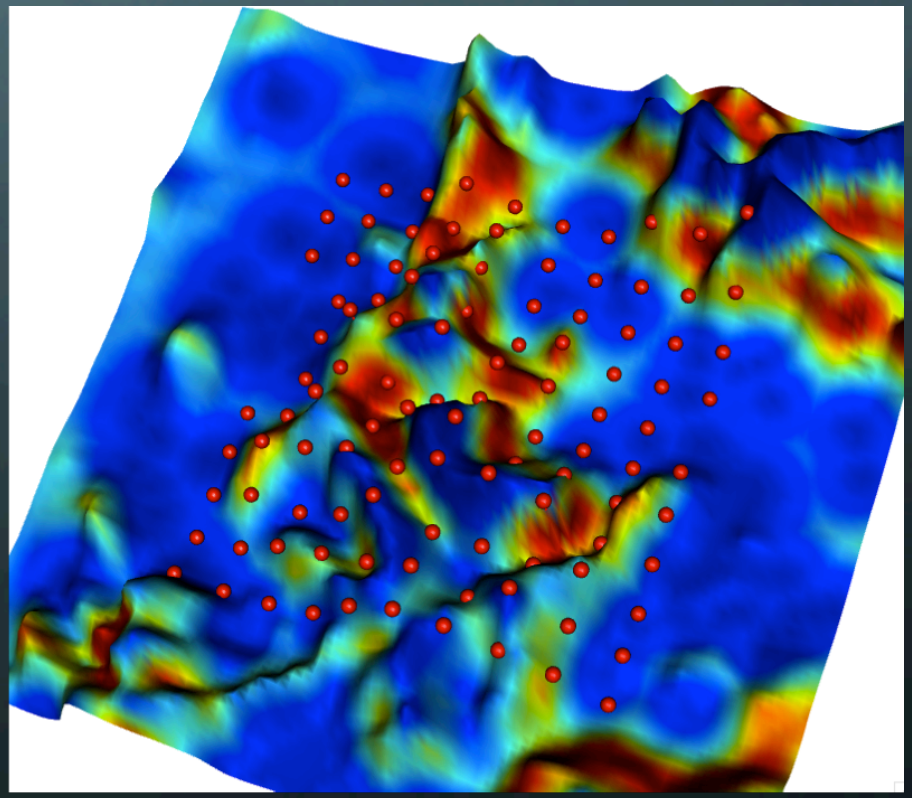
- For wildlife management/conservation data are often collected on individual animals.
- These often take the form of trapping or observing animals via a series of “traps”:

Spatially explicit capture-recapture

- The capture probability of individuals differ over traps – dependent on the home-range of the individual.
- Traditionally, the capture probability is simply specified as a function of the fixed (unobserved) home-range centre of the individual.
- This ignores the known locations of the individuals from observed trappings/sightings, so that:
  - Information is thrown away (poor practice);
  - It implicitly allows animals to “teleport”.

Output

- The output of the statistical analysis is in the form of an estimated density for the population (in terms of home range centres).



Questions of interest

- How can we incorporate information relating to known observed locations of individuals (i.e. when trapped) into the analysis?
- What happens if the animals under study are territorial/social (i.e. non independent individuals)?
- How can we incorporate “moving” home range centres?
- How do we incorporate inhomogeneous landscapes?
- How can we fit the models efficiently?

Thankyou!

- If you want to know more – come and ask!