

Bayesian statistics and modelling in ecology

Prof. Paul Blackwell

University of Sheffield, UK

February 2011

Bayesian statistics

Bayesian statistics: the systematic use of probability to express uncertainty and update beliefs in the light of data.

Bayes' Theorem tells us how to update our beliefs given data, but we must make our previous beliefs and assumptions explicit; "prior distributions".

Posterior prob. is proportional to prior prob. times likelihood

$$p(\theta|x) \propto p(\theta) \times p(x|\theta)$$

Here, x can be anything we observe; θ can be anything we are uncertain about: parameters; missing values; future values.

By expressing hierarchical relationships between unknowns, we can often 'borrow strength'.

After statistical analysis, we can make probability statements about parameters, functions of parameters, missing values, future values. . . individually and jointly.

Probability is the natural way to express uncertain predictions and assess the consequences of possible actions.

In fact Bayes' rule tells us how to make decisions! But we must specify our utility function.

Computation has been an obstacle to Bayesian methods, but there is an increasing range of algorithms—and software—to fix this.

MCMC (Markov chain Monte Carlo) has been vital.
Implementation in BUGS makes it accessible for many models.

MCMC is simulation-based, and results in a (usually very large!) sample from posterior distributions of interest.

Also ABC (Approximate Bayesian Computation), INLA (Integrated Nested Laplace Approximations).

Bayesian estimation for relative-abundance models

Duncan Gollcher, Bob O'Hara, Lorena Ruíz-Montoya, Luis Cayuela

Data: fruit-feeding butterflies in 4 areas of southern Mexico.

Assume Poisson distribution for the number of individuals for a species, with mean μ_i for species i .

Interested in the distribution of the μ_i s; this is often modelled as log-normal, but here a Gamma distribution is a better fit.

So

$$X_i \sim \text{Poisson}(\mu_i)$$

$$\mu_i \sim \text{Gamma}(\alpha, \beta)$$

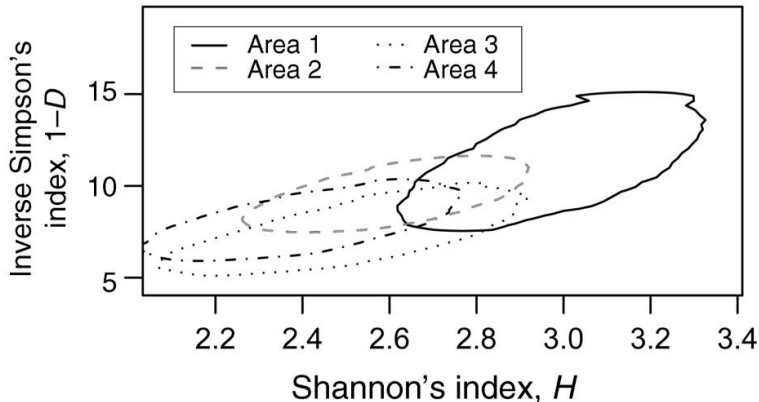
Parameters not so easy to interpret as in log-normal case, but can generate corresponding values of μ_i s and standard diversity indices: Shannon H and inverse Simpson $1/D$.

$$(\alpha, \beta) \rightarrow \mu \rightarrow X$$

↓

$$(H, D)$$

95% highest posterior density regions; Golicher et al (2006)
("1-D" should be "1/D")



Reconstructing animal territories

Blackwell (2001)—builds on joint work with David Macdonald, data from Hans Kruuk.

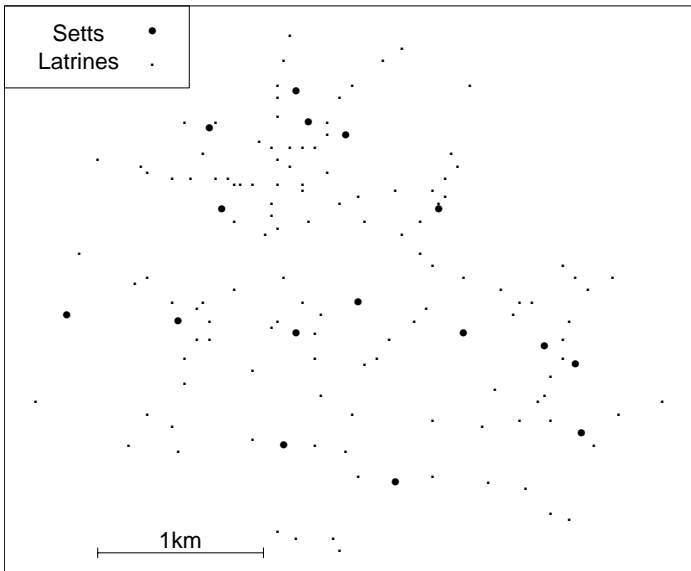
Looking at reconstructing a map of territories of badgers (*Meles meles*), based on information on latrines.

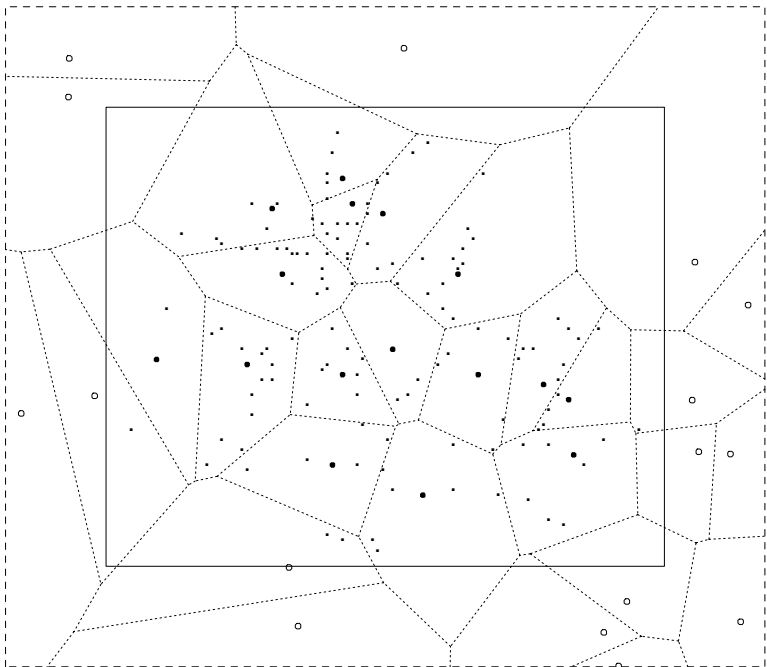
Represent map as a tessellation, defined by 'centres' of territories.

Model locations of latrines as a random point process, with intensity high near boundaries, low elsewhere.

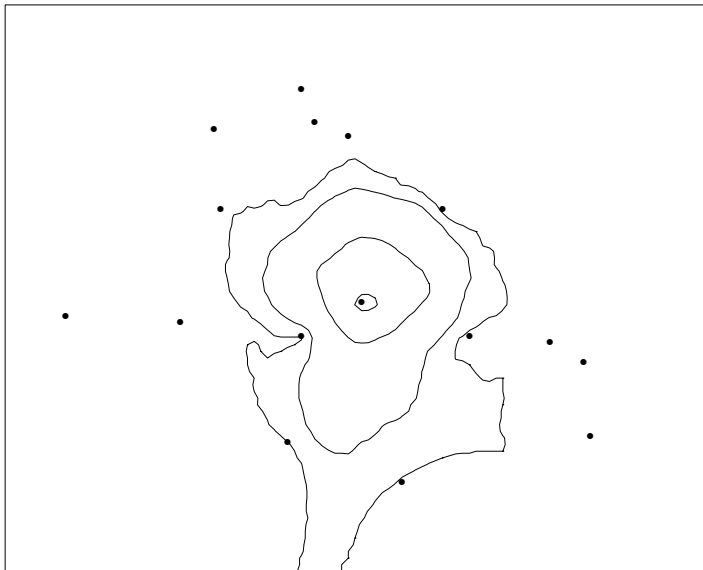
Can tackle edge effects by Geyer-Möller algorithm (add/drop completely unobserved centres).

Badger latrine data; Kruuk (1978), Blackwell (2001)

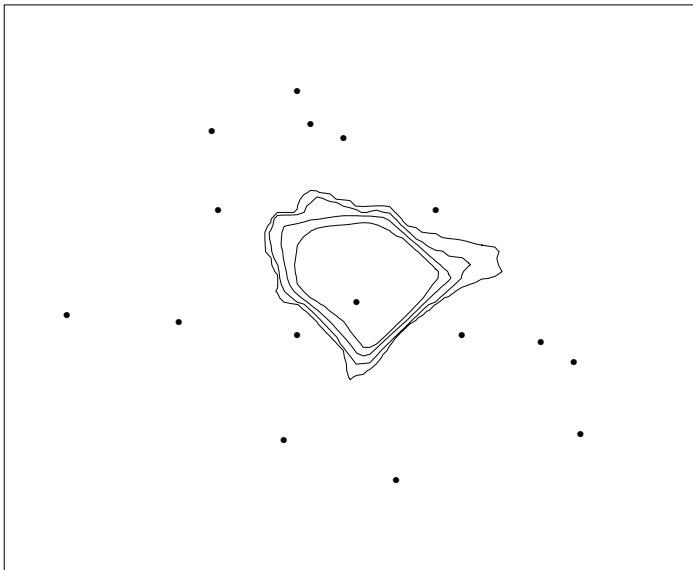




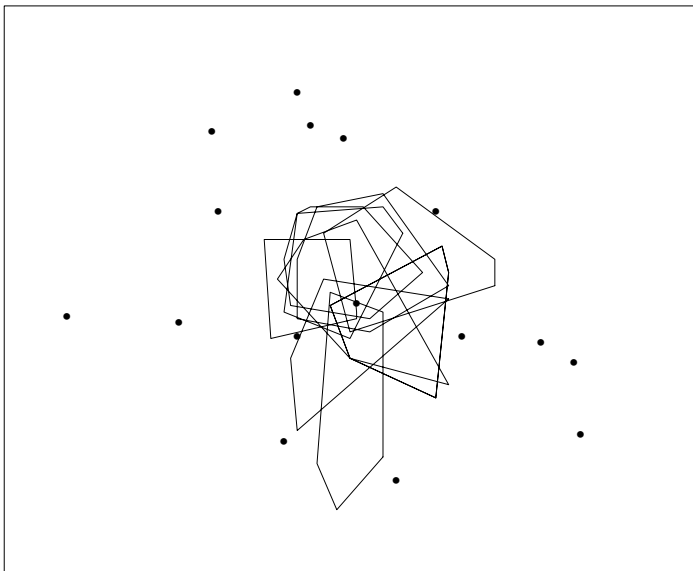
Prior probability contour plot



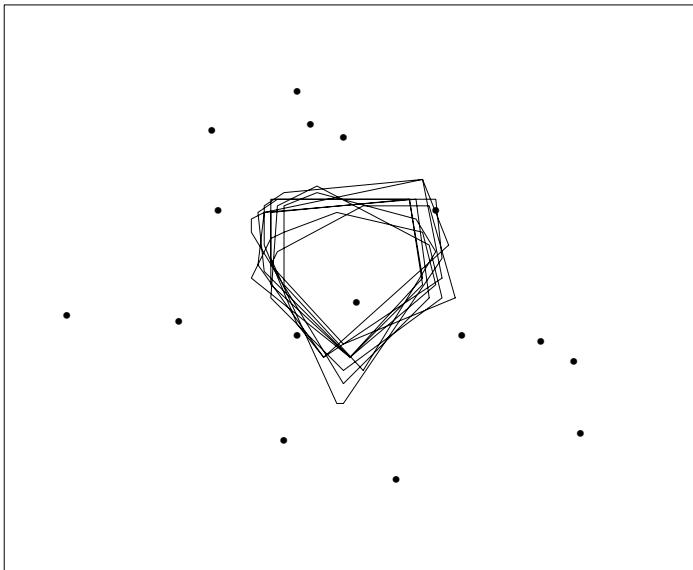
Posterior probability contour plot



Prior sample of boundaries



Posterior sample of boundaries



Changes in distribution of several invasive plant species

Gary Campbell, Ian Woodward, Angela Howard, PGB

Interested in a range of species, and how their dispersal is affected by the river network.

Data supplied by the Biological Records Centre: presence/absence of species in 10km cells, by year.

Modelling the cost to spread between grid cells depending on direction and landscape.

Presence (white)/absence (grey) of *Impatiens glandulifera*



$$Pr(X_{ij}^{(t+1)} = 1 | \{X_{rs}^{(t)}\}) = f\left(\min_{r,s: X_{rs}^{(t)}=1} \{c(r, s \rightarrow i, j)\}\right)$$

where $c(r, s \rightarrow i, j)$ is a cost function given by sum of costs of directed steps from one cell to the next.

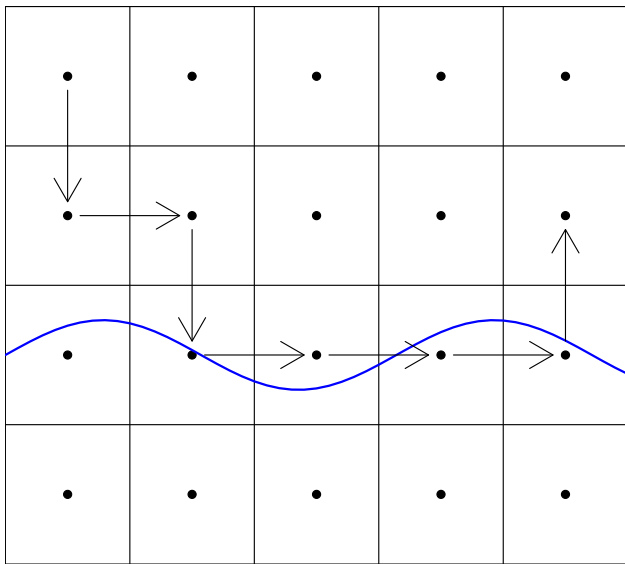
Cost of each step depends on parameters α (autochory speed) and β (hydrochory speed), and on local rivers in that cell.

Simulate seed movement on detailed landscapes to parametrise river effect i.e. how to summarise each cell.

Use Bayesian inference to estimate parameters.

We can then give e.g. realisations of future maps, probabilities in each cell.

Construction of cost surface



Cost surface for possible spread of *Impatiens glandulifera*



Bayesian radio-tracking

Keith Harris, Svetlana Tishkovskaya, PGB

Relationship between individual animal movement and habitat

Analysed radio-tracking data (David Macdonald) on wood mice (*Apodemus sylvaticus*) in and around farmland.

Movement follows an Ornstein-Uhlenbeck process (roughly, a random walk) with parameters/properties that depend on habitat (and/or on 'behaviour')—a switching diffusion model.

Analysis difficult because transitions between habitats unobserved.

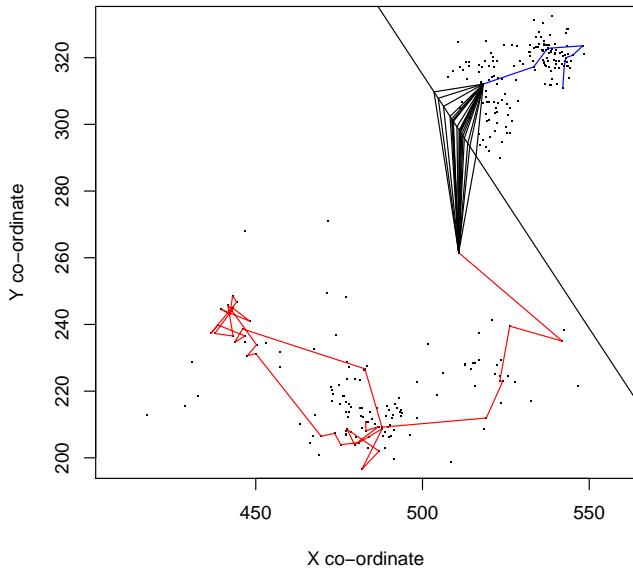
Changes in behaviour also unobserved.

Bayesian approach gives a natural way to handle missing values—crucially the times of boundary crossings—by ‘imputing’ values probabilistically.

General idea: data augmentation, very widely used.

Multiple imputation, to correctly represent uncertainty.

Trajectories of a wood mouse in pasture and woodland



Defining biodiversity

An aside. . . with nice connections!

Andrew Solow and Stephen Polasky

A measure of biodiversity with a probabilistic interpretation.

Motivated by the probability of finding a cure for a disease. Could be any species, but need to allow for similarity of species.

Say that any species is equally likely to provide a cure, but if one doesn't, then closely related species are unlikely to.

For a set of species S , define

$$V(S) = eF^{-1}e^T$$

where

$$F = (f(d_{ij})), \quad i, j \in S,$$

is a measure of relatedness and

$$e = (1, \dots, 1).$$

$V(S)$ is the effective number of species in S , in the sense that if the probability of a cure from any one species is p , and

$$Pr(\text{cure from } i | \text{cure from } j) = p + (1 - p)f(d_{ij}),$$

then the chance of a cure in S is the same as with $V(S)$ unrelated species.

There is a connection with the effective number of observations in a statistical experiment.

Also . . .

Simon Willerton, Tom Leinster

Generalise to the idea of ‘magnitude’—for example, the effective number of points in a metric space. There’s a connection with category theory . . .

Beta diversity across a seascape

Al Harborne, Pete Mumby, Kamila Żychaluk, John Hedley, PGB

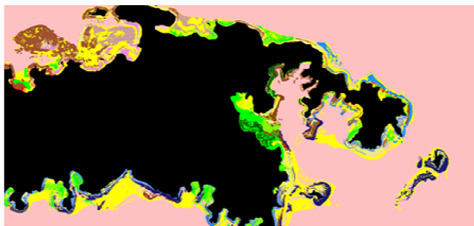
Looked at spatial patterns of beta diversity in a tropical marine (largely coral) ecosystem.

Measured diversity based on proportion and similarity of different benthic communities in a moving window around each point, of area 0.5 to 5.0 km².

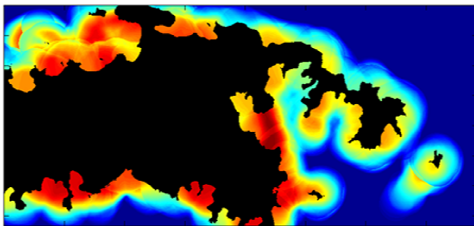
$$B_d = \log_{10} \sum_{i,j} (100D_{ij})^2 \times \frac{-\sum_{i=1}^H P_i \ln P_i}{\ln H}$$

where D_{ij} is a Bray-Curtis dissimilarity coefficient.

Marine data: (a) communities and (b) beta-diversity



(a)



(b)

Obviously strong spatial auto-correlation—can be allowed for.

Requires non-standard spatial representation—distance measured only within sea.

Want to 'explain' beta-diversity in terms of easily obtained variables.

Use depth and exposure—calculate wave exposure based on a mechanistic model, using fetch and depth.

Depth (and variability in depth) and exposure (and variability in exposure) accounted for 60% of variability in beta diversity, or more.

National Centre for Statistical Ecology (UK)

www.ncse.org.uk

International Statistical Ecology Conference (ISEC)
July 2012, near Oslo.