Hierarchical Models of Species Abundance and Occurrence

J. Andrew Royle
USGS Patuxent Wildlife Research Center

MBI Ohio State University
April 4, 2006

Introduction

Ecology: *The study of variation in abundance and occurrence*

Inference problems:

- Spatial distribution (range) of a species
- Population size/abundance/density
- Landscape factors that influence abundance/occurrence
- Environmental factors that influence . . .
- Population, metapopulation or community dynamics: interactions with other species, density dependence, spatial interactions, etc..

“I think that population ecology would be better off if all statisticians were taken out and shot!”

– C.J. Krebs

Observational Studies in Ecology

- In ecology, we have an acute inability to observe the state variable of interest.

- Data are not only "noisy", but also biased: **Imperfect detection** (or non-detection bias)

  Observations: $y(s) \sim \text{Binomial}(N(s), p)$
  - $N(s)$ = (local) population size
  - both p, $N(s)$ unknown

Imperfect Detection – Detectability

This issue has attracted considerable attention in the development of methods for sampling populations, and in developing models and inference procedures from observational data.

It might be important as a component of the observation model because functions of N may not be invariant to sampling bias.

e.g., Hypotheses about $N(s)$, or the metapopulation distribution $g(N;\theta)$:

- $E[N(s)]$
- $\Pr(N(s) > 0)$

Prevailing Views: The Classical View

Observation/sampling driven (classical view): Modeling and inference are focused on the detection process (observation model) part of the problem.

- Partial likelihood (conditional estimators):

$$[data|\text{“captured”}, p][\text{“captured”}|N]$$

- $N(s)$ is a derived parameter:

$$N(s) = c(\hat{p}) \times n(s)$$

The Classical View

This view has been fostered to a large extent by the advent of software that perform certain, limited types of analyses, converting sample data to estimates of N. To achieve broader objectives, the practitioner then will have to use the results in a plug-in second procedure.

- Software driven
- Complex models for p
- $N(s)$ is a derived parameter
- ad hoc, multi-stage procedures (statistics on statistics)

Prevailing Views: Process-Driven view

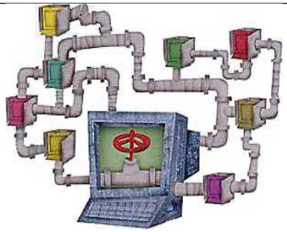
Process driven view: Modeling/estimation and inference are focused on the process component of the model

Characterized by:

- Complex models for process
- Disregard sampling processes
- Interpretation as if $p = 1$

Philosophical Ordination

Rube Goldberg



- software driven
- complex models for p
- "statistics on statistics"

$\Longleftarrow$? ? ? $\Longrightarrow$

Head-in-sand

- Ignore sampling
- Complex models for process

The Conceptual Middle: Hierarchical Models

Most ecological inference problems and sampling situations yield naturally to a hierarchical construction, distinct models for:

- $[data|process, parameters]$
- $[process|parameters]$

Why is this the conceptual middle?

Rube-Goldberg: $[data|parameters]$

Head-in-Sand: $[process|parameters]$

Hierarchical Models

Hierarchical models are concerned with 3 quantities:

1. *Observations*, $y(s,t)$ – "data"

- *observation model* $f(y|z)$

2. *State variable*, $z(s,t)$: outcome of ecological *process* of interest

- *process model* $g(z|\theta)$

3. *Parameters*, θ: govern the state process

- *prior distribution*: $[\theta]$

The Joy of Hierarchical Models

Factorization of the joint probability distribution:

$$[y, z, \theta] = f(y|z)g(z|\theta)[\theta]$$

Why is this useful? In practice, it can be very difficult to conceptualize or describe $[y, z, \theta]$ directly. But, each of $f(y|z)$, $g(z|\theta)$, $[\theta]$ might be very simple and easy to construct.

The Swiss Breeding Bird Survey

All About the Swiss BBS:

- Annual, conducted during the breeding season
- **Sample units**: 300 1 km^2 quadrats
- Each unit sampled 3 times.
- Covariates: elevation, forest cover influence abundance
- Covariates: quadratic day-of-year effect on detection probability
- Variation in route length among quadrats

Data: counts $\mathbf{x}_i = (x_{i1}, x_{i2}, x_{i3})$ for quadrat i

Hierarchical Model for Swiss Bird Data

1. Process model – N_i = abundance of the *quadrat*. $N_i \sim \text{Poisson}(\lambda_i)$.

$$\log(\lambda_i) = \alpha_0 + \alpha_1 \text{Forest}_i + \alpha_2 \text{Elev}_i + \alpha_3 \text{Elev}_i^2.$$

2. Spatial coverage bias – $M_i \leq N_i$ is the population of quadrat i that is *exposed* to sampling by a route of length L_i:

$$M_i \sim \text{Binomial}(N_i, p(L_i))$$

$p(L_i) \longrightarrow 1$ as $L_i \rightarrow \infty$ (as L_i increases, exposed population $\longrightarrow$ quadrat population):

$$p(L_i) = \exp\{-\alpha_4 / L_i\}$$

3. Observation model:

$$x_{ij} \sim \text{Binomial}(M_i, \theta_j)$$

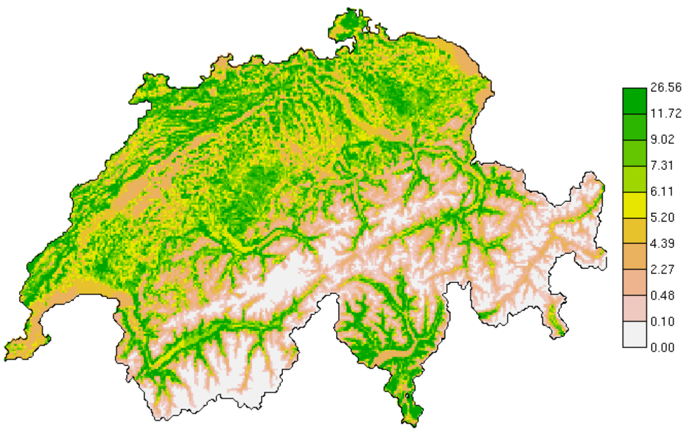
Detection probability:

$$logit(\theta_j) = \beta_0 + \beta_1\texttt{day} + \beta_2\texttt{day}^2$$

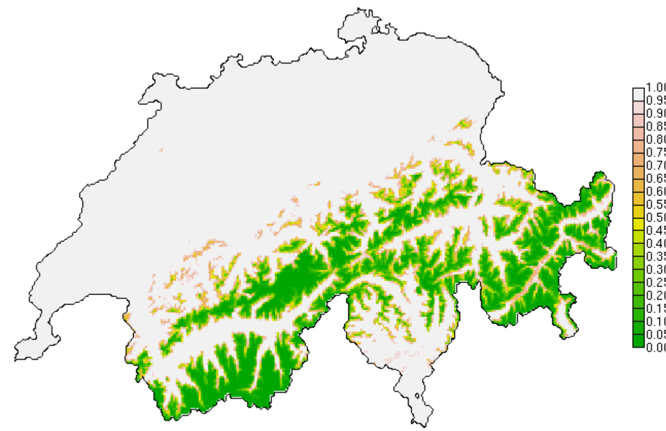
Pretty Pictures

Eurasian Jay

Abundance (territories/km^2)



$Pr(N(s) > 0)$



Summary

- Ecology is a science largely concerned with understanding spatial and temporal variation in abundance and occurrence.
- Most studies are observational.
- Hierarchical models provide a conceptually clear, flexible, rigorous framework for modeling and inference
- In practice, it can be very difficult to conceptualize or describe probability models directly. But each of $[data|process, parameters]$, $[process|parameters]$, $[parameters]$ are frequently simple and easy to construct.