Why Become a Bayesian

FW8051 Statistics for Ecologists

Department of Fisheries, Wildlife and Conservation Biology

Why Bayes?

Some areas where Bayesian methods are a natural fit:

- Mixed effects models
- Generalized additive models (smoothing splines)
- Mixture distributions
- Models with conditionally independent (sometimes latent) random variables

Random effects Models

What makes a model a random effects model?

Some of the parameters are assumed to be drawn from a statistical distribution

- Bayesian methods are a natural fit (all parameters are assigned a distribution)
- Just adding a hierarchical structure (so, some parameters come from a *common* distribution)
- Use posterior distribution to estimate cluster-specific parameters and their uncertainty (nothing new!)

Frequentist methods (formally think of parameters as fixed, not random):

- Use Empirical Bayes to “predict” random parameters (BLUPs)
- Not clear how to account for uncertainty in these predictions

Smoothing splines

$$PSS = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \lambda \int f''(x)dx$$

- Penalties can be justified as priors (for amount of smoothness)
- Frequentist methods appeal to Bayesian philosophy when constructing uncertainty intervals

Conditionally Independent Variables

Models with conditionally independent (sometimes latent) random variables

- Population time series with measurement error
- Mark-recapture
- Occupancy and N-mixture models
- Integrated population models

State Space Models

State-space models (e.g., population time series data):[1]

$$N_{t+1} = N_t exp(a + bN_t + \epsilon_t) \text{ (equation for changes in state)}$$

$$\epsilon_t \sim N(0, \sigma_p^2) \text{ captures "process variation"}$$

We count animals, leading to $C_t \neq N_t$, but our interest is in the latent variable, N_t (and a, b).

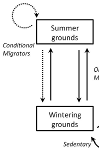
Observation model connects C_t to N_t:

- $log(C_t) \sim N(log(N_t), \sigma_o^2)$ (observation process)
- or $C_t \sim Binomial(N_t, p)$ (observation process)

[1] Mech, D. and J. Fieberg. 2015. Growth Rates and Variances of Unexploited Wolf Populations in Dynamic Equilibria. Wildlife Society Bulletin 39: 41–48. doi: 10.1002/wsb.511.

How does winter severity influence the probability that white-tailed deer will migrate to their winter range?[2]

- Obligate migrators migrate every year
- Facultative migrators migrate when winters are severe



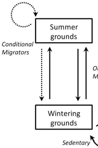
[2] Fieberg, J. and Conn, P. 2014. A hidden Markov model to identify and adjust for selection bias: An example involving mixed migration strategies. Ecology and Evolution 4(10):1903-1912.

$$y_{ij} = \begin{cases} 1 & \text{if deer i migrates in year j} \\ 0 & \text{otherwise} \end{cases}$$

- Pr(migrate | obligate migrator) = 1
- Pr(migrate in year j | facultative migrator) = θ_j

$\theta_j = \texttt{plogis}(\beta_0 + \beta_1 WSI_j)$

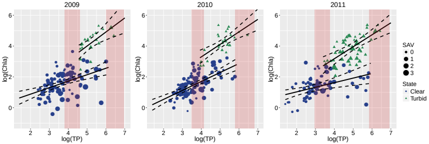
$$z_i = \begin{cases} 1 & \text{if deer i is an obligate migrator} \\ 0 & \text{if deer i is a facultative migrator} \end{cases}$$



z_i is partially observed (= 0 if did not migrate in $\geq$ 1 year, NA otherwise)

$Y_{ij} \sim Bernouli(\Theta_{ij}); \Theta_{ij} = z_i + (1 - z_i)\theta_{ij}$

Mixture Distributions

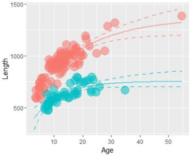


Relationship between Chla and Phosphorus depends on lake state (a latent variable)

$$chla_i = \alpha_0 + \tau S_i + b_0(1 - S_i)log(P_i) + b_1 S_i log(P_i) + \epsilon_i$$

$$S_i \sim Bern(p_i) = \begin{cases} 1 & \text{if the lake is turbid} \\ 0 & \text{if the lake is clear} \end{cases}$$

Mixture Distributions: Cubera Snapper



z_i = 1 if morph type I and 0 if morph type II; $z_i \sim dbern(p)$

Kery and Schaub's book: Bayesian Population Analysis using WinBugs

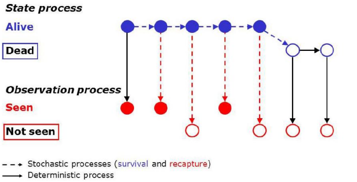


Fig. 7-1: Example of the state and observation process of a marked individual over time for the CJS model. The sequence of true states in this individual is z = [1,1,1,1,1,0,0], and the observed capture-history is y = [1,1,0,1,0,0,0].

Maximum Likelihood versus Bayesian Methods

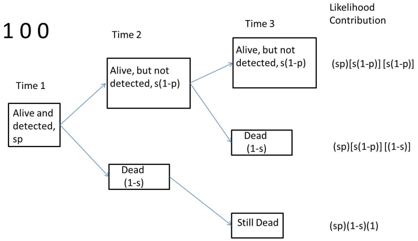
Likelihood formulation requires thinking of all the ways to get a capture-history ("alive, but not seen" or "dead so can't be detected")

- Think back to probability trees on your earlier homework

Bayesian methods are easier (I think), because we can break the problem down into conditionally independent events, modeled as Bernoulli random variables:

- Survival, given status (alive/dead) at previous time point
- Detection given current status (alive/dead)

Capture history 1 0 0



(sp)[s(1-p)] [s(1-p)] + (sp)[s(1-p)] [(1-s)] + (sp)(1-s)(1)

Bayesian formulation: 1 0 0

Survival:

- $z_{i,t}$ = 1 if individual i is alive at time t (0 if dead).
- $z_{i,t}|z_{i,t-1} \sim Bernoulli(sz_{i,t-1})$ (probability = 0 if already dead)

Detection:

- $y_{i,t}$ = 1 if individual i is seen at time t (0 otherwise)
- $y_{i,t} \sim Bernoulli(pz_{i,t})$ (probability = 0 if dead)

JAGS: 1 0 0

```
for (i in 1:nindividuals){
    # Survive first time period to be detected?
    z[i,1] ~ dbern(s)

   # Detected given alive?
    y[i,1] ~ dbern(z[i,1]*p)

  for(t in 2:3){
   # Survive (if alive previous time point z[i-1]=1)
    z[i,t] ~ dbern(s*z[t-1])

   # Detected given alive
    y[i,t] ~ dbern(z[i,t]*p)
  }
}
```

$z_{i,t}$ are only partially observed:

- If seen at time t, know $z_{i,t}$ = 1 (for all time points prior to t)
- If not seen at t, $z_{i,t}$ = NA.

JAGS will estimate/impute the missing $z_{i,t}$!

Occupancy Models

In presence-absence surveys, 0's can arise in two ways:

- The species is absent from the sample unit
- The species is present but not detected.

The processes affecting presence-absence may differ from those that influence detection | presence.

- $logit(\psi_i) = \alpha_0 + \alpha_1 x$ where ψ_i = probability species is present
- $logit(p_i) = \beta_0 + \beta_1 z$ where p_i = probability of detecting a species given it is present

Two options for model fitting:

- Use a probability tree to determine all the ways to observe particular capture histories, then use maximum likelihood (e.g., `optim`)
- Consider partially observed variables and conditional independencies within a Bayesian formulation.

Occupancy Models: Maximum Likelihood

Consider a single site and survey:

Define a random variable, Y_i as:

$$Y_i = \begin{cases} 1 & \text{if the species is observed at site i} \\ 0 & \text{if the species is not observed at site i} \end{cases}$$

- ψ = probability the site is occupied
- p = probability the species is detected, given the site is occupied

What is $P(Y = 1)$?

$P(Y = 1)$ = P(observed | present)P(present) = ψp

What is $P(Y = 0)$?

$P(Y = 0)$ = P(Not observed | present)P(present) + P(not observed | not present)P(not present) = $(1 - p)\psi + 1 - \psi = 1 - \psi p$

Likelihood: single visit

$$L = \prod_i (\psi p)^{Y_i}(1 - \psi p)^{(1-Y_i)}$$

- Likelihood contribution of a 1 is ψp
- Likelihood contribution of a 0 is $1 - \psi p$

Note: with single visits (and, with constant detection and occupancy probabilities), we cannot estimate both ψ and p

With multiple visits we can...

Alternative way

$$z_i = \begin{cases} 1 & \text{if the species is present at site i} \\ 0 & \text{if the species is absent from site i} \end{cases}$$

$z_i \sim Bernouli(\psi)$ (ψ = probability a site is occupied)

$$Y_i = \begin{cases} 1 & \text{if the species is observed at site i} \\ 0 & \text{if the species is not observed at site i} \end{cases}$$

$y_i \sim Bernouli(pz_i)$ (probability detected, given occupied = p)

Again, z_i are only partially observed:

- If see a species, know the site is occupied ($z_i = 1$)
- If do not see a species, z_i = NA (JAGS will impute/estimate)

JAGS

```
for(i in 1:nsites){
  # Occupied Y/N
  z[i] ~ dbern(psi)

  # Detected at a site (depends on psi and p)
  y[i] ~ dbern(z[i]*p)
}
```

But, we still need multiple visits to identify ψ and p.

Multiple Visits: Occupancy model

Assumptions:

- Each visit is independent (observers do not ‘remember’ where they previous found animals and adjust their search effort)
- The population is closed (presence-absence does not change during the course of the survey)
- Usually assume no “false-detections”, $P(Y_{ij} = 1 | z_i = 0) = 0$

These are all big assumptions!

There is also a cost to visiting sites multiple times

Multiple visits: start with 2

4 Possibilities:

- (0,0)
- (0,1)
- (1,0)
- (1,1)

Write down the probability of each capture history.

- ψ = probability the site is occupied
- p = probability the species is detected, given the site is occupied

(0,0) = $\psi(1-p)(1-p) + (1-\psi)$

(0,1) or (1,0) = $\psi p(1-p)$

(1,1) = ψp^2

Multiple visits: start with 2

Assume we have

- n_{00} of these (0,0)
- n_{01} of either (0,1) or (1,0)
- n_{11} of (1,1)

$$L \propto [\psi(1-p)(1-p) + (1-\psi)]^{n_{00}}[\psi p(1-p)]^{n_{01}}[\psi p^2]^{n_{11}}$$

Could use `optim` to solve or p and ψ

(see also Excel spreadsheet Todd Arnold put together under **Handouts**, Occupancy tab)

JAGS

```
for(i in 1:nsites){
  # Occupied Y/N
  z[i] ~ dbern(psi[i])
}
for(j in 1:nvisits){
  # Detected at a site (depends on psi and p)
  y[i,j] ~ dbern(z[i]*p)
}
```

Willow Tit Example

In-class Exercise using data from:

Royle, J.A. and R. M. Dorazio. 2008. Hierarchical Modeling and inference in Ecology. Academic Press, New York, NY.

Occupancy models

Can model the influence of covariates on:

- Abundance: $logit(\psi) = \beta_0 + \beta_1 * elevation$
- Detection: $logit(p) = \gamma_0 + \gamma_1 * wind.speed$

These models are considerably more complex than logistic regression models.

They also require repeated visits to sites.

When is the effort worth it?

Logistic Regression

When fitting logistic regression models to data that are impacted by perfect detection, we are essentially modeling:

$$logit(\psi p) = \beta_0 + x\beta_1$$

Consequences (Kery 238):

- Estimators of occupancy probabilities will be biased low whenever $p < 1$
- Estimators of βs, describing covariate relationships with occupancy rates, will be biased toward 0 whenever $p < 1$
- Variables that explain variability in detection probabilities may show up as important predictors of *apparent presence*

The other side of the coing: Brian McGill's first blog[3]

Welsh AH, Lindenmayer DB, Donnelly CF (2013) Fitting and Interpreting Occupancy Models. PLoS ONE 8: e52015.

- Maximum likelihood can be *hard*
    - Many cases with multiple solutions; answer you get may depend on starting values you supply to the solver
    - Solutions may converge to the boundary (where ψ and/or p were either 0 or 1; theoretically impossible)
    - Problems are worse when data were sparse
- Detection may depend on abundance, so correcting for detection may give you a worse index of abundance.

[3] https://dynamicecology.wordpress.com/2013/01/11/is-using-detection-probabilities-a-case-of-statistical-machismo/

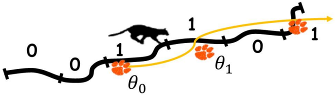
Follow-up[4]

Guillera-Arroita G, Lahoz-Monfort JJ, MacKenzie DI, Wintle BA, McCarthy MA (2014) Ignoring imperfect detection in biological surveys is dangerous: a response to ‘Fitting and Interpreting Occupancy Models’. PLoS ONE 9: e99571.

- If ψ is low ($<$ 50%) and *p* is also low, ignoring detection probabilities gives a **better** estimate of ψ (in terms of mean-squared error).
- Accounting for detection works best when ψ is high ($>$ 50%) and *p* is low.
- When *p* is high (> 50% and definitely when > 80%), both methods work equally well regardless of ψ

[4] https://dynamicecology.wordpress.com/2014/09/15/detection-probabilities-statistical-machismo-and-estimator-theory/

Occupancy and Correlated Data



- π = probability species is present to the left of the first segment
- θ_0 = probability species is present given it was not present in the previous replicate
- θ_1 = probability species is present given it was present in the previous replicate

Hines et al. (2010), Hines et al. (2014), Aing et al. (2011)

Capture history: 0101

From Hines et al. 2014:

$$\Pr(h_i = 0101) = \psi[([\pi\theta' + (1-\pi)\theta](1-p)\theta' + [\pi(1-\theta') + (1-\pi)(1-\theta)]\theta)p \times [(1-\theta')\theta + \theta'(1-p)\theta']p] \quad \text{eqn 1}$$

Previously non-sampled site was occupied or not, first sample occupied or not. . . .

Capture history: 00

$$\begin{aligned}\Pr(h_i = 00) = (1-\psi) + \psi\{&([\pi\theta' + (1-\pi)\theta](1-p)\theta' + [\pi(1-\theta')\\ &+ (1-\pi)(1-\theta)]\theta)(1-p) + [\pi\theta' + (1-\pi)\theta](1-p)(1-\theta')\\ &+ [\pi(1-\theta') + (1-\pi)(1-\theta)](1-\theta)\}\end{aligned}$$

Bayesian

Conditional Bernoulli random variables:

- $z_i \sim dbern(\psi_i)$ Site occupied ($z_i = 1$) or not ($z_i = 0$)
- $f_i \sim dbern(z_i \pi)$ segment 0 (not sampled) within site i occupied “locally” or not.
- $y_{i1} \sim dbern(z_i f_i \theta' + z_i(1 - f_i)\theta)$ first sampled segment at site i occupied locally or not.
- $y_{ij} \sim dbern(z_i y_{ij-1} \theta' + z_i(1 - y_{ij-1})\theta)$ segment j at site i occupied locally or not
- $o_{ij} \sim dbern(y_{ij} p)$ observed on segment j at site i.

```
# Likelihoood
for(i in 1:nsites){ # M = number of sites
  # Occupied
  z[i] ~ dbern(psi[i])       # Occupied Y/N
  logit(psi[i]) <- b0 + b1*AGR[i] + b2*SeaSTM[i]

  # Site "0" (non-surveyed site) local presence
  first.obs[i]~dbern(z[i]*pi)

  # For site 1
  # y = probability of local presence | site presence (z=1)
  loc.psi[i,1]<-z[i]*theta[1+first.obs[i]]
  y[i,1] ~ dbern(loc.psi[i,1])  # local presence site 1

  # Detection | local abundance
  logit(p[i])<-g0+g1*FNS3km[i]
  obs[i,1]~dbern(p[i]*y[i,1])

  for(j in 2:nsurveys){ # number of visits
  # y = probability of local presence | site presence (z=1)
  loc.psi[i,j]<-z[i]*theta[1+y[i,j-1]]
  y[i,j] ~ dbern(loc.psi[i,j])  # local presence site 1

  # Detection | local abundance
   obs[i,j]~dbern(p[i]*y[i,j])

  }
}
```