Mixture models, zero inflation, occupancy

Models for Socio-Environmental Data

Chris Che-Castaldo, Mary B. Collins, and N. Thompson Hobbs

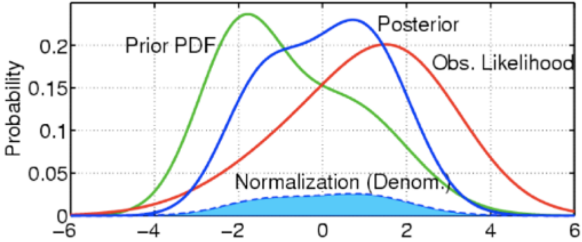
August 22, 2017

SESYNC

Flow of ideas

- Mixture models in general
- Zero-inflation as a useful example of mixture models
- Occupancy as an example of zero-inflation

Remember these bumpy distributions?



Introduction

Mixture Distribution:

- $[\mathbf{y}|\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, p] = p[\mathbf{y}|\boldsymbol{\theta}_1]_1 + (1-p)[\mathbf{y}|\boldsymbol{\theta}_2]_2$
  - $0 \leq p \leq 1$.
  - $[\mathbf{y}|\theta_1]_1$ and $[\mathbf{y}|\theta_2]_2$ integrate to 1.

K-Mixture Distribution:

- $[\mathbf{y}|\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_K, \mathbf{p}] = \sum_{k=1}^{K} p_k[\mathbf{y}|\boldsymbol{\theta}_k]_k$
  - $p_k \geq 0$ for all k.
  - $\sum_{k=1}^{K} p_k = 1$.
  - all $[\mathbf{y}|\theta_k]_k$ integrate to 1.

Introduction

Mixture Distribution:

- $[\mathbf{y}|\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, p] = p[\mathbf{y}|\boldsymbol{\theta}_1]_1 + (1-p)[\mathbf{y}|\boldsymbol{\theta}_2]_2$
  - $0 \leq p \leq 1$.
  - $[\mathbf{y}|\boldsymbol{\theta}_1]_1$ and $[\mathbf{y}|\boldsymbol{\theta}_2]_2$ integrate to 1.

K-Mixture Distribution:

- $[\mathbf{y}|\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_K, \mathbf{p}] = \sum_{k=1}^{K} p_k[\mathbf{y}|\boldsymbol{\theta}_k]_k$
  - $p_k \geq 0$ for all k.
  - $\sum_{k=1}^{K} p_k = 1$.
  - all $[\mathbf{y}|\boldsymbol{\theta}_k]_k$ integrate to 1.

Introduction

Mixture Distribution:

- $[\mathbf{y}|\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, p] = p[\mathbf{y}|\boldsymbol{\theta}_1]_1 + (1-p)[\mathbf{y}|\boldsymbol{\theta}_2]_2$
  - $0 \le p \le 1$.
  - $[\mathbf{y}|\boldsymbol{\theta}_1]_1$ and $[\mathbf{y}|\boldsymbol{\theta}_2]_2$ integrate to 1.

K-Mixture Distribution:

- $[\mathbf{y}|\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_K, \mathbf{p}] = \sum_{k=1}^{K} p_k [\mathbf{y}|\boldsymbol{\theta}_k]_k$
  - $p_k \ge 0$ for all k.
  - $\sum_{k=1}^{K} p_k = 1$.
  - all $[\mathbf{y}|\boldsymbol{\theta}_k]_k$ integrate to 1.

Introduction

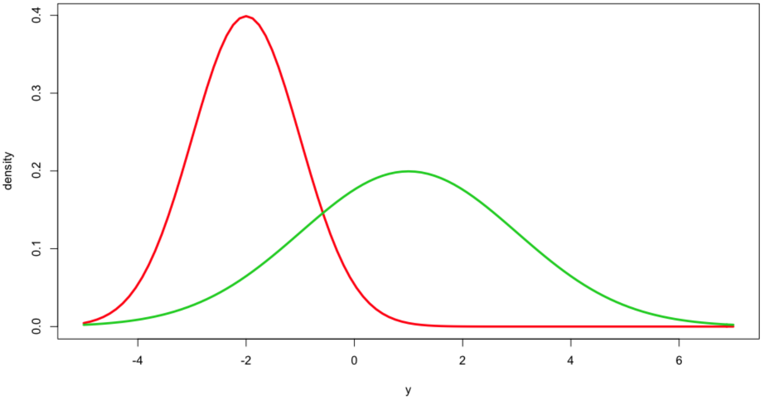
Mixture Distribution:

- $[\mathbf{y}|\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, p] = p[\mathbf{y}|\boldsymbol{\theta}_1]_1 + (1-p)[\mathbf{y}|\boldsymbol{\theta}_2]_2$
  - $0 \leq p \leq 1$.
  - $[\mathbf{y}|\boldsymbol{\theta}_1]_1$ and $[\mathbf{y}|\boldsymbol{\theta}_2]_2$ integrate to 1.

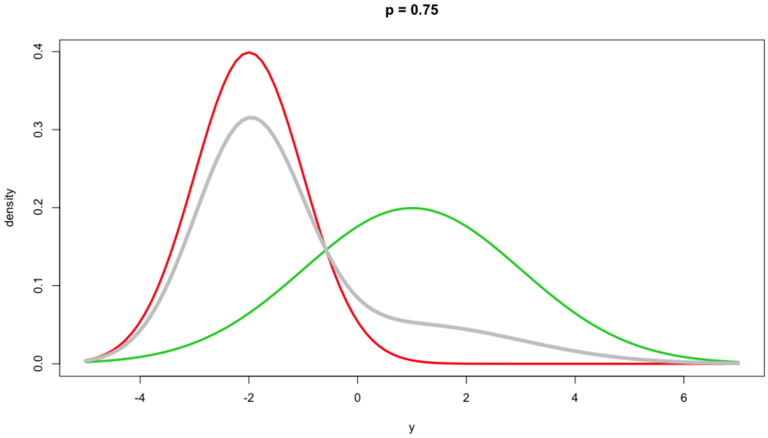
K-Mixture Distribution:

- $[\mathbf{y}|\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_K, \mathbf{p}] = \sum_{k=1}^{K} p_k[\mathbf{y}|\boldsymbol{\theta}_k]_k$
  - $p_k \geq 0$ for all k.
  - $\sum_{k=1}^{K} p_k = 1$.
  - all $[\mathbf{y}|\boldsymbol{\theta}_k]_k$ integrate to 1.

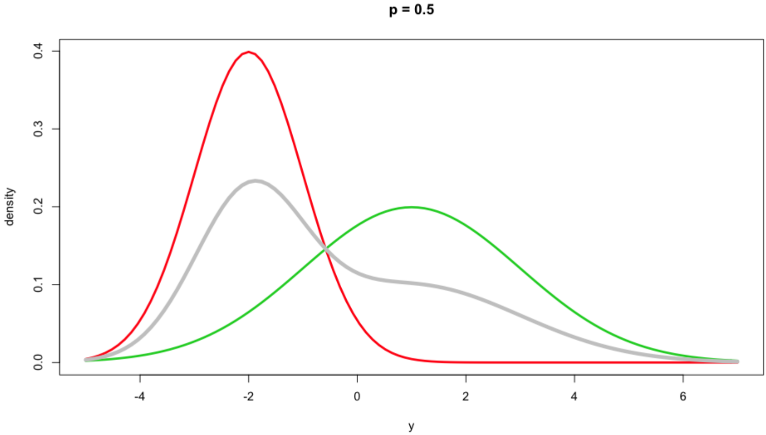
Introduction



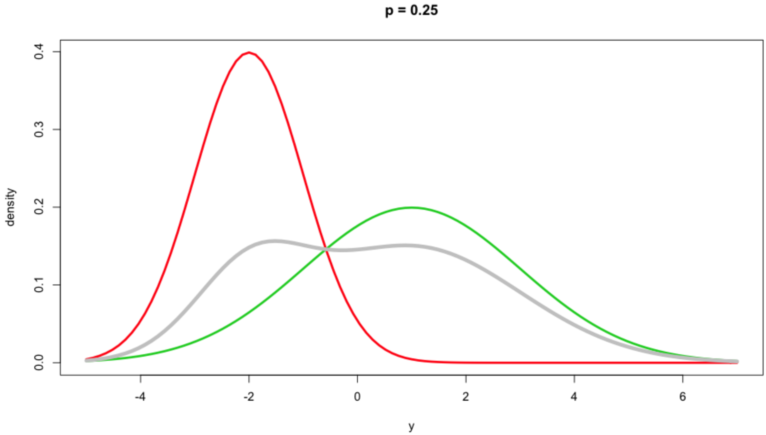
Introduction



Introduction



Introduction



Example: Darwin's Finches Model (Hendry et al. 2006):

$$y_i \sim p \cdot \text{normal}(\mu_1, \sigma^2) + (1 - p) \cdot \text{normal}(\mu_2, \sigma^2)$$

- $i = 1, \ldots, n$
- $\mu_1 \neq \mu_2$
- $0 < p < 1$

Example: Darwin's Finches

- Use latent (auxiliary) variables to make the mixture model hierarchical:

$$y_i \sim \begin{cases} \text{normal}(\mu_1, \sigma^2) & \text{if } z_i = 1 \\ \text{normal}(\mu_2, \sigma^2) & \text{if } z_i = 0 \end{cases}$$

where,

$$z_i \sim \text{Bernoulli}(p)$$

and,

$$\begin{aligned} p &\sim \text{beta}(\alpha, \beta) \\ \mu_1 &\sim \text{normal}(0, 100) \\ \mu_2 &\sim \text{normal}(0, 100) \\ \sigma^2 &\sim \text{inverse gamma}(.01, .01) \end{aligned}$$

Directed acyclic graph

Draw it.

Implementation Posterior

$$[\mu_1, \mu_2, \sigma^2, \mathbf{z}, p|\mathbf{y}] \propto \prod_{i=1}^{n} [y_i|\mu_1, \sigma^2]^{z_i} [y_i|\mu_2, \sigma^2]^{1-z_i} [z_i|p][p][\mu_1][\mu_2][\sigma^2]$$

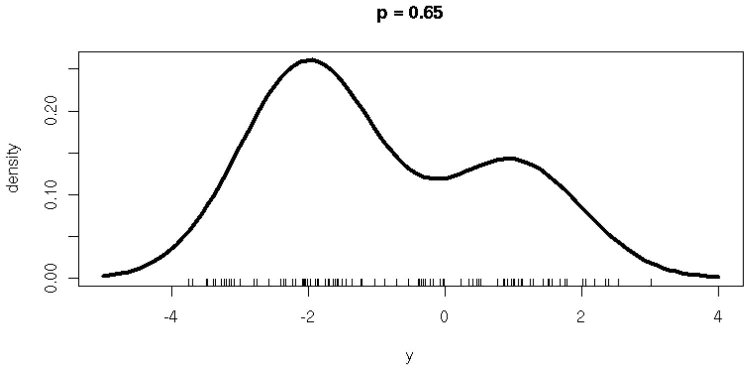
Full conditionals:

$$\begin{aligned}
[\mu_1|\cdot] &\propto \left(\prod_{i=1}^{n} [y_i|\mu_1, \sigma^2]^{z_i} \right) [\mu_1] \\
[\mu_2|\cdot] &\propto \left(\prod_{i=1}^{n} [y_i|\mu_2, \sigma^2]^{1-z_i} \right) [\mu_2] \\
[\sigma^2|\cdot] &\propto \left(\prod_{i=1}^{n} [y_i|\mu_2, \sigma^2]^{z_i} [y_i|\mu_2, \sigma^2]^{1-z_i} \right) [\sigma^2] \\
[z_i|\cdot] &\propto [y_i|\mu_1, \sigma^2]^{z_i} [y_i|\mu_2, \sigma^2]^{1-z_i} [z_i|p] \\
[p|\cdot] &\propto \left(\prod_{i=1}^{n} [z_i|p] \right) [p]
\end{aligned} \tag{1}$$

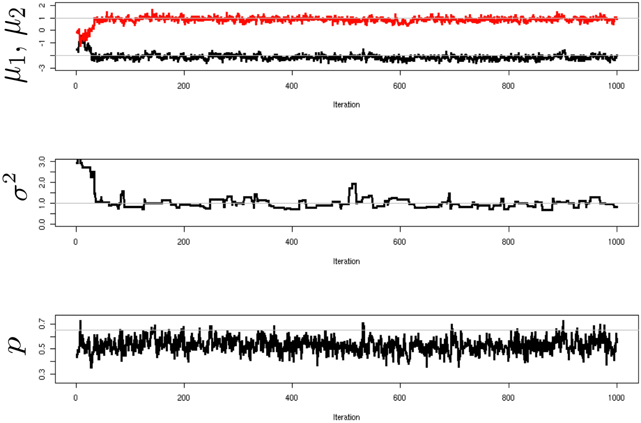
Implementation of MCMC:

1. Choose initial values for parameters. Use a Gibbs update for all unobserved quantities:
2. Sample from $[\mu_1|\cdot]$
3. Sample from $[\mu_2|\cdot]$.
4. Sample from $[\sigma^2|\cdot]$.
5. Sample from $[z_i|\cdot]$ for $i = 1, \ldots, n$.
6. Sample from $[p|\cdot]$.
7. Repeat 2–6 until convergence, then as many MCMC samples as desired.

Data analysis finch model



MCMC output

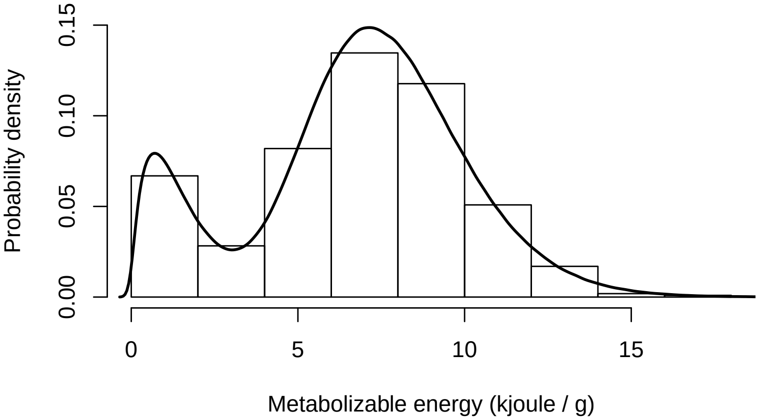


Some JAGS coding tricks

```
model{
 #inits for sigma[2]  > sigma[1], m[2] > mu[1]
  sigma[1] ~ dgamma(4/100, 2/100)
  sigma[2] ~ dgamma(25/200,5/200) T(sigma[1],)
  mu[1] ~ dgamma(4/100,2/100)
  mu[2] ~ dgamma(49/500, 7/500) T(mu[1],)
  p ~ dunif(0,1)
  alpha[1] <- mu[1]^2 / sigma[1]^2
  beta[1] <- mu[1] / sigma[1]^2
  alpha[2] <- mu[2]^2 / sigma[2]^2
  beta[2] <- mu[2] / sigma[2]^2

  for(i in 1:length(y)){
    z[i] ~ dbern(p)
    alpha.mix[i] <- z[i] * alpha[1] + (1-z[i]) * alpha[2]
    beta.mix[i] <- z[i] * beta[1] + (1-z[i]) * beta[2]
    y[i] ~ dgamma(alpha.mix[i],beta.mix[i])
    y.new[i] ~ dgamma(alpha.mix[i],beta.mix[i])
    } # end of i
  #posterior predictive checks should go here
} #end of model
```

Model fit



0.00
0.05
0.10
0.15
Probability density
0
5
10
15
Metabolizable energy (kjoule / g)

Martin et al. 2005

Ecology Letters, (2005) 8: 1235–1246 doi: 10.1111/j.1461-0248.2005.00826.x

REVIEWS AND SYNTHESES

Zero tolerance ecology: improving ecological inference by modelling the source of zero observations

Tara G. Martin,[1]* Brendan A. Wintle,[2] Jonathan R. Rhodes,[3] Petra M. Kuhnert,[4] Scott A. Field,[5] Samantha J. Low-Choy,[6] Andrew J. Tyre[7†] and Hugh P. Possingham[1]

Abstract

A common feature of ecological data sets is their tendency to contain many zero values. Statistical inference based on such data are likely to be inefficient or wrong unless careful thought is given to how these zeros arose and how best to model them. In this paper, we propose a framework for understanding how zero-inflated data sets originate and deciding how best to model them. We define and classify the different kinds of zeros that occur in ecological data and describe how they arise: either from 'true zero' or 'false zero' observations. After reviewing recent developments in modelling zero-inflated data sets, we use practical examples to demonstrate how failing to account for the source of zero inflation can reduce our ability to detect relationships in ecological data and at worst lead to incorrect inference. The adoption of methods that explicitly model the sources of zero observations will sharpen insights and improve the robustness of ecological analyses.

Keywords

Bayesian inference, detectability, excess zeros, false negative, mixture model, observation error, sampling error, zero-inflated binomial, zero-inflated Poisson, zero inflation.

Ecology Letters (2005) 8: 1235–1246

Zero-inflation as mixture model

Imagine that you sampled many plots along a coastline, counting the number of species of mussels within each plot. In essence there are two sources of zeros. Some zeros arise because the plot was placed areas that are not mussel habitat, while other zeros occur in plots placed in mussel habitat but that contain no mussels as a result of sampling variation. The Poisson distribution offers a logical choice for modeling the distribution of counts in mussel habitat, but it cannot portray the zeros that arise because plots were placed in areas where mussels never live.

Zero-inflation as mixture model

$$y_i \sim \begin{cases} 0 & w_i = 1 \\ \text{Poisson}(\lambda) & w_i = 0 \end{cases}$$

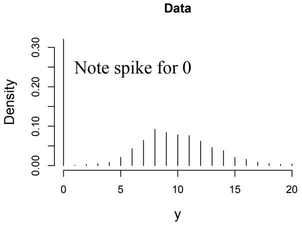
$$\begin{aligned} y_i &\sim \text{Poisson}\left(y_i|\lambda(1-w_i)\right) \cdot \text{Bernoulli}\left(w_i|\phi\right) \text{beta}\left(\phi|1,1\right) \\ &\times \text{gamma}(\lambda|.01,.01), \end{aligned}$$

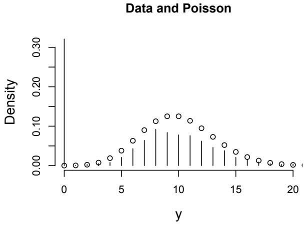
where λ is the average number of mussels per plot; ϕ is the probability that the plot is outside of mussel habitat, and w_i is an indicator variable for non-habitat.

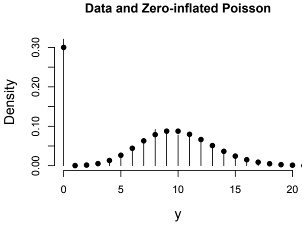
Bayesian network (DAG)

Draw it.

Zero-inflation as mixture model







Zero-inflation example

Zero-inflated models

Poisson:

$$[\lambda, \alpha, \beta, \mathbf{w}, \phi | \mathbf{y}] \propto \prod_{i=1}^{n} \text{Poisson}(y_i | \lambda(1 - w_i)\text{Bernoulli}(w_i | \phi)$$
$$\times \text{beta}(\phi | \alpha, \beta)[\lambda][\alpha][\beta]$$

Binomial:

$$[p, \alpha, \beta, \mathbf{w}, \phi | \mathbf{y}] \propto \prod_{j=1}^{J} \text{binomial}(y_i | n_i, p(1 - w_i)\text{Bernoulli}(w_i | \phi)$$
$$\times \text{beta}(\phi | \alpha, \beta)[p][\alpha][\beta]$$

In both cases ϕ is the probability of a zero that is not accounted for by sampling variation alone. Negative binomial and multinomial random variables can also be modeled this way.

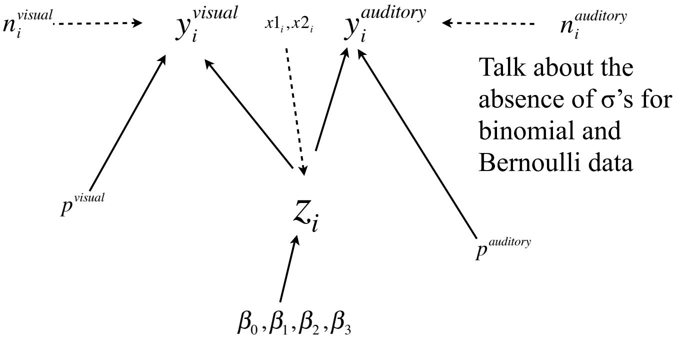
Occupancy models as application of zero inflation

- We want to understand controls on presence or absence of individuals, traits, etc. with discrete categories.
- We have a problem interpreting zeros.
  - The individual is truly absent.
  - The individual is present but undetected.
- We want to use a model to explain spatial or temporal variation in presence or absence.
- We need to model uncertainty arising from two sources:
  - The failure of the model to portray the process
  - The error in our observations arising because we fail to perfectly observe presence or absence

Exercise: Courtesy of McCarthy 2007: Box 5.9

- Kristen Parris studied controls on the distribution of tree frogs in the riparian zone of streams on the east coast of sub-tropical Australia.
- Multiple surveys were conducted at 64 sites using 2 observation methods, visual searches at night and auditory searches for responses to taped calls.
- We assume the presence /absence of frogs at a site does not change during the multiple surveys. (Frogs don't fly.) This is known as the *closure* assumption
- Presence / absence of frogs was modeled as a function of 1) stream size (measured as annual volume of rainfall in the catchment above the site) 2) the presence or absence of palms at the site (an indicator of mesic or xeric conditions) and 3) the interaction between palms and stream size.
- Data are the total number of times frogs were detected using each method for each site and the number of surveys for each site.

Bayesian network (DAG)



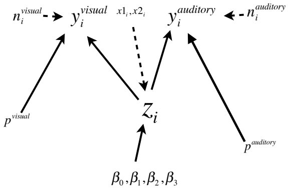
Posterior and joint distributions:

$$[\boldsymbol{z}, \boldsymbol{\beta}, p^{auditory}, p^{visual} | \mathbf{y}^{auditory}, \mathbf{y}^{visual}] \propto$$

$$\prod_{i=1}^{64} \text{binomal}\left(y_i^{auditory} \mid n_i^{auditory}, z_i \cdot p^{auditory}\right) \text{binomial}\left(y_i^{visual} \mid n_i^{visual}, z_i \cdot p^{visual}\right) \times$$

$$\text{Bernoulli}\left(z_i \mid \text{invlogit}\left(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2\right)\right) \times$$

$$\prod_{j=0}^{3} \text{normal}\left(\beta_j \mid 0, .0001\right) \text{uniform}(p^{visual} \mid 0,1) \text{uniform}(p^{auditory} \mid 0,1)$$

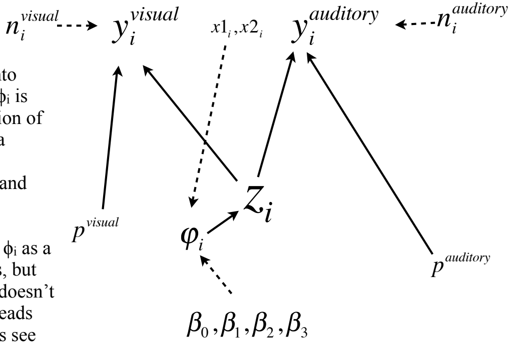


A common error:

Important

Here is where you can run into trouble. (I did.) Notice that ϕ_i is simply a calculation, a function of the β's and the x's. It is not a stochastic quantity.

The relationship between z_i and the β's and the x's form the stochastic quantity. You are certainly free to estimate the ϕ_i as a function of random variables, but putting them in the diagram doesn't make sense in terms of the heads and tails of arrows helping us see the conditioning. If you want them in the diagram, use a different type of arrow indicating a calculated quantity, as I did here.



Code

```
model
{
  b0 ~ dnorm(0, 1.0E-6)  # uninformative priors for the variables
  b[1] ~ dnorm(0, 1.0E-6)
  b[2] ~ dnorm(0, 1.0E-6)
  b[3] ~ dnorm(0, 1.0E-6)
  p.visual ~ dunif(0, 1)  # detection probabilities when the species is present
  p.auditory ~ dunif(0, 1)
  mLnCV <- mean(LnCV[])  # average catchment volume
  for (i in 1:64)  # for each of the 64 sites
  {
    phi[i] <- ilogit(b0 + b[1]*(LnCV[i] - mLnCV) + b[2]*palms[i] + b[3]*(LnCV[i] -
mLnCV)*palms[i])  # probability of presence using centered data
 z[i] ~ dbern(phi[i])  # actual, latent presence at the site, 0 or 1
    y.visual[i] ~ dbin(p.visual*z[i], n.visual[i])  # number eye detections with
    y.auditory[i] ~ dbin(p.auditory*z[i], n.auditory[i])  # number of ear detec-
tions
  }
  # predicted relationships--derived quantities; note that predic-
tions are done using centered data. There is no back-transform.
  for (i in 1:20)
  {
    LVol[i] <- 2 + 3*i/20  # covers the range of stream sizes
    logit(predpalms[i]) <- b0 + (b[1] + b[3])*(LVol[i] - mLnCV) + b[2]
    logit(prednopalms[i]) <- b0 + b[1]*(LVol[i] - mLnCV)
  }
}
```

Implementation tricks

Refer to code above

- Initialize all latent quantities i.e., all z[i] = 1.The phi[i] don't need to be initialized because they are derived quantities.
- You may need to bound phi[i], i.e.:

```
phi[i] <- max(min(ilogit(a + b[1]*(LnCV[i] -
mLnCV) + b[2]*palms[i] + b[3]*(LnCV[i] -
mLnCV))*palms[i],.0000001),.999999)
```