Bayesian methods for ecological and environmental modelling

Trainers:

Lindsay Banin, David Cameron,
Pete Henrys & Peter Levy

UK Centre for
Ecology & Hydrology

Hierarchical modelling

Part 2: Non-normal distributions

Lindsay Banin

UK Centre for Ecology & Hydrology

What we will cover in Session 5b

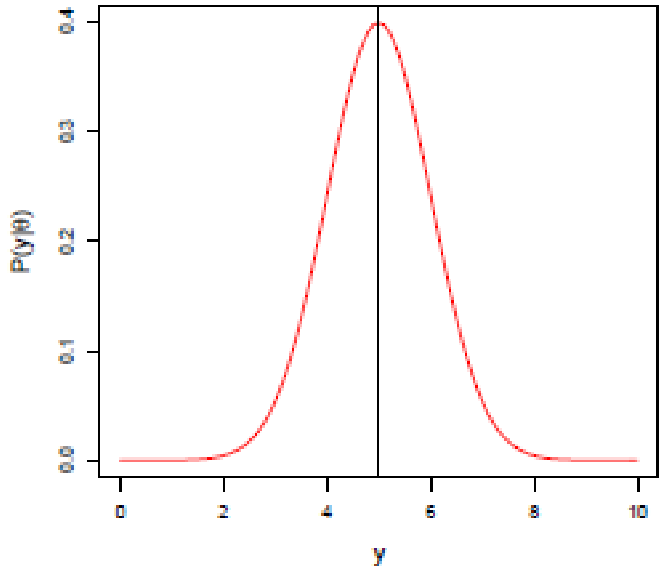
- We will consider data which take probability distributions other than Normal, and why that matters in Bayesian analyses
- Implement examples of Bayesian generalised linear (mixed-effects) models using rstanarm
- Re-cap of the day and opportunity for questions

Probability distributions

The values of random variables are governed by a probability distribution

Probability model:

$y_i \sim f\left(\mu_i, \sigma\right)$, μ_i and σ^2 are parameters of the distribution $f()$



From Tom Hobbs, Colorado State University

Probability distributions

A toolbox of *f()*'s for ecological data (and later, parameters, latent states)

- <u>Discrete</u>
  - Poisson
  - binomial
  - negative binomial
  - Bernoulli
  - multinomial
- <u>Continuous</u>
  - normal
  - multivariate normal
  - lognormal
  - uniform
  - beta
  - gamma
  - Dirichlet

From Tom Hobbs, Colorado State University

Probability distributions

A toolbox of *f()*'s for ecological data (and later, parameters, latent states)

- Discrete
  - Poisson
  - binomial
  - negative binomial
  - Bernoulli
  - multinomial
- Continuous
  - normal
  - multivariate normal
  - lognormal
  - uniform
  - beta
  - gamma
  - Dirichlet

Count data appears frequently in ecological datasets (e.g. abundance of individuals) – modelled as a Poisson or negative binomial distribution

From Tom Hobbs, Colorado State University

Probability distributions

A toolbox of *f()*'s for ecological data (and later, parameters, latent states)

- Discrete
  - Poisson
  - binomial
  - negative binomial
  - Bernoulli
  - multinomial
- Continuous
  - normal
  - multivariate normal
  - lognormal
  - uniform
  - beta
  - gamma
  - Dirichlet

Count data appears frequently in ecological datasets (e.g. abundance of individuals) – modelled as a Poisson or negative binomial distribution

Yes/no trials & outcomes data – e.g. survival

From Tom Hobbs, Colorado State University

Probability distributions

A toolbox of *f()*'s for ecological data (and later, parameters, latent states)

- Discrete
  - Poisson
  - binomial
  - negative binomial
  - Bernoulli
  - multinomial
- Continuous
  - normal
  - multivariate normal
  - lognormal
  - uniform
  - beta
  - gamma
  - Dirichlet

Count data appears frequently in ecological datasets (e.g. abundance of individuals) – modelled as a Poisson or negative binomial distribution

Yes/no trials & outcomes data – e.g. survival

Continuous percentage data

Strictly positive (shape & scale describe skew) – e.g. biomass, rainfall

From Tom Hobbs, Colorado State University

Generalised linear (mixed effects) models in rstanarm

- In GLMs **link functions** relate the expected values in the response to the linear predictors in the model

Following is a table of several exponential-family distributions in common use and the data they are typically used for, along with the canonical link functions and their inverses (sometimes referred to as the mean function, as done here).

Common distributions with typical uses and canonical link functions

Distribution	Support of distribution	Typical uses	Link name	Link function, $\mathbf{X}\boldsymbol{\beta} = g(\mu)$	Mean function
Normal	real: $(-\infty, +\infty)$	Linear-response data	Identity	$\mathbf{X}\boldsymbol{\beta} = \mu$	$\mu = \mathbf{X}\boldsymbol{\beta}$
Exponential	real: $(0, +\infty)$	Exponential-response data, scale parameters	Negative inverse	$\mathbf{X}\boldsymbol{\beta} = -\mu^{-1}$	$\mu = -(\mathbf{X}\boldsymbol{\beta})^{-1}$
Gamma					
Inverse Gaussian	real: $(0, +\infty)$		Inverse squared	$\mathbf{X}\boldsymbol{\beta} = \mu^{-2}$	$\mu = (\mathbf{X}\boldsymbol{\beta})^{-1/2}$
Poisson	integer: $0, 1, 2, \ldots$	count of occurrences in fixed amount of time/space	Log	$\mathbf{X}\boldsymbol{\beta} = \ln(\mu)$	$\mu = \exp(\mathbf{X}\boldsymbol{\beta})$
Bernoulli	integer: $\{0, 1\}$	outcome of single yes/no occurrence		$\mathbf{X}\boldsymbol{\beta} = \ln\left(\frac{\mu}{1-\mu}\right)$	
Binomial	integer: $0, 1, \ldots, N$	count of # of "yes" occurrences out of N yes/no occurrences		$\mathbf{X}\boldsymbol{\beta} = \ln\left(\frac{\mu}{n-\mu}\right)$	

https://en.wikipedia.org/wiki/Generalized_linear_model

Practical 5b

- Returning to **rstanarm** package…
- Example using count data (GLM with Poisson and negative binomial distributions)
- Example using binary outcome data (GLM with binomial distribution)

- See examples in Hobbs & Hooten for non-Gaussian hierarchical models in JAGS

What we covered in this session

- Non-Gaussian data types and distributions – common in ecology!
- We have extended implementation of linear (mixed effects) models to *generalised* linear mixed effects models using rstanarm
- We have considered the use of conjugate priors for when we do not take the GLM approach to the analysis

Conclusion
What we covered today:

- Four-step modelling framework
- Model evaluation steps
- The motivation for using hierarchical models – a flexible approach for complex models
- Implementing and evaluating Normal and generalised mixed effects models in rstanarm
- Writing models and implementing Bayesian analyses in JAGS

Any Questions?

This concludes

Hierarchical modelling Part 3

UK Centre for Ecology & Hydrology

Your Feedback

What Went Well?

Even Better If!

Tomorrow, Thursday

Same start time 9:30 am

We will cover

6a Measurement and uncertainty

6b Process-based modelling – part 1

7a Process-based modelling – part 2

7b Machine learning / Miscellaneous

Bayesian methods for ecological and environmental modelling

Trainers:

Lindsay Banin, David Cameron,
Pete Henrys & Peter Levy

UK Centre for
Ecology & Hydrology