

Bayesian Regression

Bayesian Modeling for Socio-Environmental Data

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

May 2018



Outline

- Understand Bayesian inference using familiar examples.
- Appreciate one-to-one relationship between math and JAGS code.
- Empowering you to properly interpret your parameters.
- Learn some programming tricks.
- Bayesian R^2 .

Notation

$$\underbrace{g(\boldsymbol{\theta}, x)}$$

deterministic model

Embedding the deterministic model

The joint distribution in a simple Bayesian model is the product of the likelihood and the priors.

$$[\theta, \sigma^2] \propto [y \mid \theta, \sigma^2] [\theta] [\sigma^2]$$

A deterministic model of an ecological process is embedded in the likelihood like this...

$$[\theta, \sigma^2] \propto [y \mid g(\theta, x), \sigma^2] [\theta] [\sigma^2]$$

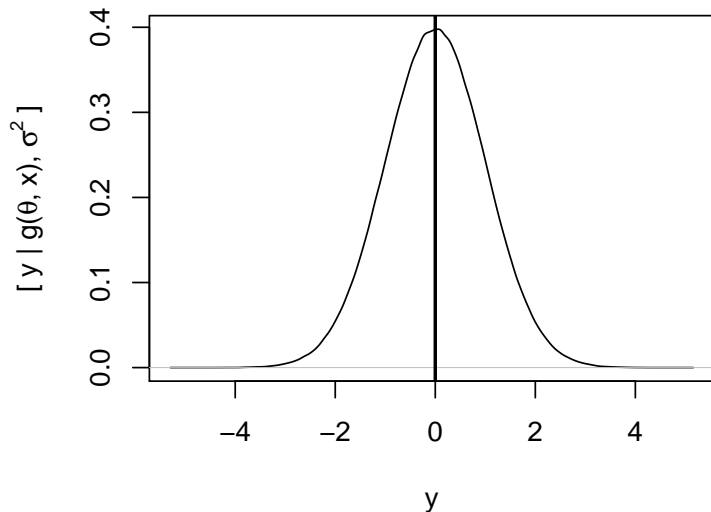
The general set up for a simple Bayesian regression model

$$\mu_i = \underbrace{g(\boldsymbol{\theta}, x_i)}_{\text{deterministic model}}$$

$$y_i \sim \underbrace{f(\mu_i, \sigma^2)}_{\text{stochastic model}}$$

$$[\boldsymbol{\theta}, \sigma^2] \propto [y \mid g(\boldsymbol{\theta}), \sigma^2] [\boldsymbol{\theta}] [\sigma^2]$$

A graphical representation



You don't have to be normal!

Data (y-values)	Distribution	"Mean" function	Link
continuous, real valued	normal	$\mu = \beta_0 + \beta_1 x$	NA
discrete, strictly positive	Poisson	$\mu = e^{\beta_0 + \beta_1 x}$	$\log(\mu) = \beta_0 + \beta_1 x$
0 or 1	Bernoulli	$\mu = \frac{\exp(\beta_0 + \beta_1 x)}{\exp(\beta_0 + \beta_1 x) + 1}$	$\text{logit}(\mu) = \log\left(\frac{\mu}{1-\mu}\right) = \beta_0 + \beta_1 x$
[0, 1]	beta	$\mu = \frac{\exp(\beta_0 + \beta_1 x)}{\exp(\beta_0 + \beta_1 x) + 1}$	$\text{logit}(\mu) = \log\left(\frac{\mu}{1-\mu}\right) = \beta_0 + \beta_1 x$
continuous, strictly positive, variance \uparrow as a f(mean)	lognormal	$\mu = e^{\beta_0 + \beta_1 x}$	$\log(\mu) = \beta_0 + \beta_1 x$
continuous, strictly positive, constant variance	gamma	$\mu = e^{\beta_0 + \beta_1 x}$	$\log(\mu) = \beta_0 + \beta_1 x$

Lots of flexibility as a modeler

Continent-wide Adélie penguin population dynamics

$$z_{s,y} \sim \text{lognormal}(z_{s,y} \mid g(\beta_1, \beta_2, z_{s,y-1}), \sigma_p)$$
$$g(\theta) = \log(z_{s,y-1} e^{\beta_1 + \beta_2 \text{wsic}_{s,y} + \beta_3 \text{ssic}_{s,y}})$$

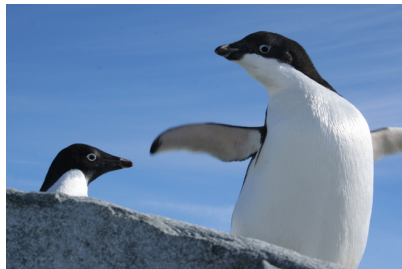
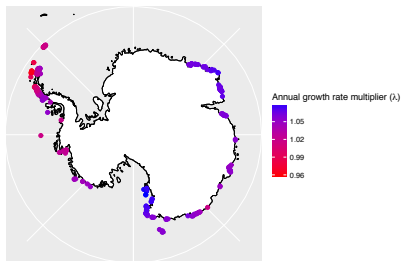


Photo c/o Heather J. Lynch

Normal data, continuous and real valued

$$\begin{aligned} [\beta_0, \beta_1, \sigma \mid \mathbf{y}] &\propto \prod_{i=1}^N \text{normal}(y_i \mid g(\beta_0, \beta_1, x_i), \sigma^2) \\ &\quad \times \text{normal}(\beta_0 \mid 0, 1000) \text{normal}(\beta_1 \mid 0, 1000) \\ &\quad \times \text{uniform}(\sigma \mid 0, 100) \\ g(\beta_0, \beta_1, x_i) &= \beta_0 + \beta_1 x_i \end{aligned}$$

```
b0 ~ dnorm(0, .001)
b1 ~ dnorm(0, .001)
sigma ~ dunif(0, 100)
tau <- 1/sigma^2
for (i in 1:length(y)){
  mu[i] <- b0 + b1 * x[i]
  y[i] ~ dnorm(mu[i], tau)
}
```

Poisson, discrete and positive

$$\begin{aligned} [\beta_0, \beta_1 \mid \mathbf{y}] &\propto \prod_{i=1}^N \text{Poisson}(y_i \mid g(\beta_0, \beta_1, x_i)) \\ &\quad \times \text{normal}(\beta_0 \mid 0, 1000) \text{normal}(\beta_1 \mid 0, 1000) \\ g(\beta_0, \beta_1, x_i) &= e^{\beta_0 + \beta_1 x_i} \end{aligned}$$

```
b0 ~ dnorm(0, .001)
b1 ~ dnorm(0, .001)
for(i in 1:length(y)){
  log(mu[i]) <- b0 + b1 * x[i]
  y[i] ~ dpois(mu[i])
}
```

or

```
mu[i] <- exp(b0 + b1 * x[i])
y[i] ~ dpois(mu[i])
```

Poisson, discrete and positive

$$\begin{aligned} [\beta_0, \beta_1 \mid \mathbf{y}] &\propto \prod_{i=1}^N \text{Poisson}(y_i \mid g(\beta_0, \beta_1, x_i)) \\ &\quad \times \text{normal}(\beta_0 \mid 0, 1000) \text{normal}(\beta_1 \mid 0, 1000) \\ g(\beta_0, \beta_1, x_i) &= e^{\beta_0 + \beta_1 x_i} \end{aligned}$$

Exponentiate β_0 and β_1 and report counts or multiplicative changes in counts.

$e^{\beta_1} = \frac{e^{\beta_0 + \beta_1 x}}{e^{\beta_0}}$ which is the multiplicative change in the mean count per change in x .

For example: “Mean western toad juvenile abundance is reduced by a factor of 5.1 (95% CI: 3.4, 10.8) per unit change in UV-B radiation.”

Bernoulli, data 0 or 1 (aka logistic)

$$\begin{aligned} [\beta_0, \beta_1 \mid \mathbf{y}] &\propto \prod_{i=1}^N \text{Bernoulli}(y_i \mid g(\beta_0, \beta_1, x_i)) \\ &\quad \times \text{normal}(\beta_0 \mid 0, 2.7) \text{normal}(\beta_1 \mid 0, 2.7) \\ g(\beta_0, \beta_1, x_i) &= \frac{e^{\beta_0 + \beta_1 x_i}}{e^{\beta_0 + \beta_1 x_i} + 1} \end{aligned}$$

```
b0 ~ dnorm(0, .37)
b1 ~ dnorm(0, .37)
for(i in 1:length(y)){
  logit(p[i]) <- b0 + b1 * x[i]
  y[i] ~ dbern(p[i])
}
```

or

```
p[i] <- inv.logit(b0 + b1 * x[i])
y[i] ~ dbin(p[i])
```

lognormal, data continuous and > 0 (log link)

$$\begin{aligned} [\beta_0, \beta_1, \sigma \mid \mathbf{y}] &\propto \prod_{i=1}^N \text{lognormal}(y_i \mid \log(g(\beta_0, \beta_1, x_i)), \sigma^2) \\ &\quad \times \text{normal}(\beta_0 \mid 0, 1000) \text{normal}(\beta_1 \mid 0, 1000) \\ &\quad \times \text{uniform}(\sigma \mid 0, 5) \\ g(\beta_0, \beta_1, x_i) &= e^{\beta_0 + \beta_1 x_i} \end{aligned}$$

Talk about the interpretation of σ .

```
b0 ~ dnorm(0, .001)
b1 ~ dnorm(0, .001)
sigma ~ dunif(0, 5)
tau <- 1/sigma^2
for(i in 1:length(y)){
  mu[i] <- exp(b0 + b1 * x[i])
  y[i] ~ dlnorm(log(mu[i]), tau)
}
```

lognormal, data continuous and > 0 (not log link)

$$\begin{aligned} [\beta_0, \beta_1, \sigma \mid \mathbf{y}] &\propto \prod_{i=2}^N \text{lognormal}(y_i \mid \log(g(\beta_0, \beta_1, y_{i-1})), \sigma^2) \\ &\quad \times \text{normal}(\beta_0 \mid 0, 1000) \text{normal}(\beta_1 \mid 0, 1000) \\ &\quad \times \text{uniform}(\sigma \mid 0, 5) \text{uniform}(y_1 \mid 1, 1E6) \end{aligned}$$

$$g(\beta_0, \beta_1, y_{i-1}) = y_{i-1} e^{\beta_0 + \beta_1 y_{i-1}}$$

```
b0 ~ dnorm(0, .001)
b1 ~ dnorm(0, .001)
sigma ~ dunif(0, 5); tau <- 1/sigma^2
y[1] ~ dunif(1, 1E6)
for(i in 2:length(y)){
  mu[i] <- y[i-1] * exp(b0 + b1 * y[i-1])
  y[i] ~ dlnorm(log(mu[i]), tau)
}
```

lognormal, data continuous and > 0 (max function)

$$\begin{aligned} [\beta_0, \beta_1, \sigma \mid \mathbf{y}] &\propto \prod_{i=1}^N \text{lognormal}(y_i \mid \log(g(\beta_0, \beta_1, x_i)), \sigma^2) \\ &\quad \times \text{normal}(\beta_0 \mid 0, 1000) \text{normal}(\beta_1 \mid 0, 1000) \\ &\quad \times \text{uniform}(\sigma \mid 0, 5) \\ g(\beta_0, \beta_1, x_i) &= \max(\beta_0 + \beta_1 x_i, 0.0001) \end{aligned}$$

```
b0 ~ dnorm(0, .001)
b1 ~ dnorm(0, .001)
sigma ~ dunif(0, 5)
tau <- 1/sigma^2
for(i in 1:length(y)){
  mu[i] <- b0 + b1 * x[i]
  y[i] ~ dlnorm(log(max(mu[i], .000001)), tau)
}
```

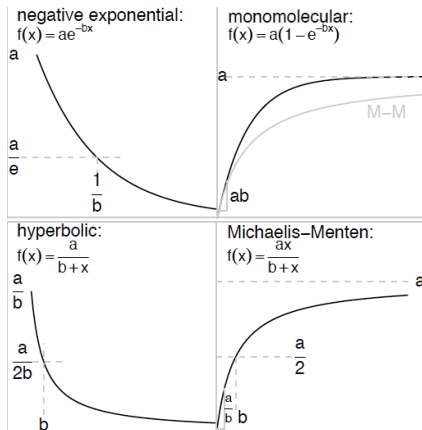
lognormal, data continuous and > 0 (priors)

$$\begin{aligned} [\beta_0, \beta_1, \sigma \mid \mathbf{y}] &\propto \prod_{i=1}^N \text{lognormal}(y_i \mid \log(g(\beta_0, \beta_1, x_i)), \sigma^2) \\ &\quad \times \text{uniform}(\beta_0 \mid .001, 1000) \text{uniform}(\beta_1 \mid .001, 1000) \\ &\quad \times \text{uniform}(\sigma \mid 0, 5) \\ g(\beta_0, \beta_1, x_i) &= \beta_0 + \beta_1 x_i \end{aligned}$$

Is this really a good idea?

```
b0 ~ dunif(.001, 1000)
b1 ~ dunif(.001, 1000)
sigma ~ dunif(0, 5)
tau <- 1/sigma^2
for(i in 1:length(y)){
  mu[i] <- b0 + b1 * x[i]
  y[i] ~ dlnorm(log(mu[i]), tau)
}
```


Nonlinear regression



Figures c/o Bolker, B. 2008. *Ecological Models and Data in R*. Princeton University Press, Princeton, NJ. USA.

Centering and standardizing

The remainder of the slides apply to all of the general linear models, but I will use a simple linear for normally distributed data as an example.

Centering predictor data

$$y_i = \beta_0 + \beta_1(x_i - \bar{x})$$

Why complicate things...?

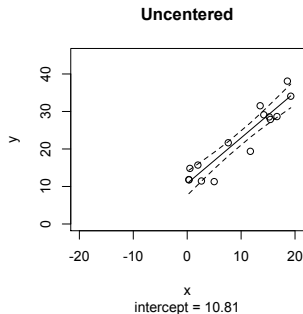
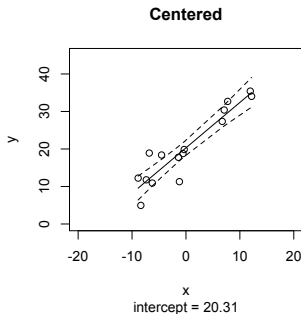
- To reduce autocorrelation in MCMC chain and speed convergence.
- To make the intercept more easily interpretable.

Centering predictor data

$$\begin{aligned} [\beta_0, \beta_1, \sigma \mid \mathbf{y}] &\propto \prod_{i=1}^N \text{normal}(y_i \mid g(\beta_0, \beta_1, x_i, \bar{x}), \sigma^2) \\ &\quad \times \text{normal}(\beta_0 \mid 0, 1000) \text{normal}(\beta_1 \mid 0, 1000) \\ &\quad \times \text{uniform}(\sigma \mid 0, 100) \\ g(\beta_0, \beta_1, x_i) &= \beta_0 + \beta_1(x_i - \bar{x}) \end{aligned}$$

```
b0 ~ dnorm(0, .001)
b1 ~ dnorm(0, .001)
sigma ~ dunif(0, 100)
tau <- 1/sigma^2
xBar <- mean(x)
for (i in 1:length(y)){
  mu[i] <- b0 + b1 * (x[i] - xBar)
  y[i] ~ dnorm(mu[i], tau)
}
b0_UC <- b0 - b1 * xBar
```

Recovering uncentered parameters



$$B_0 = \beta_0 - \beta_1 \bar{x}$$

$$B_1 = \beta_1$$

- For this to work properly, all of the coefficients in the model must be *added*.
- Slopes will not be the same if there is an interaction term or quadratic. In these cases, back transforming is not simple.

Standardizing predictor data

$$y_i = \beta_0 + \beta_1 \left(\frac{x_i - \bar{x}}{\sigma_x} \right)$$

Why complicate things...?

- To reduce autocorrelation in MCMC chain and speed convergence.
- To make the intercept more easily interpretable.
- To make parameters more easily comparable.

Standardizing predictor data

$$\begin{aligned} [\beta_0, \beta_1, \sigma \mid \mathbf{y}] &\propto \prod_{i=1}^N \text{normal}(y_i \mid g(\beta_0, \beta_1, x_i, \bar{x}, \sigma_x), \sigma^2) \\ &\quad \times \text{normal}(\beta_0 \mid 0, 1000) \text{normal}(\beta_1 \mid 0, 1000) \\ &\quad \times \text{uniform}(\sigma \mid 0, 100) \\ g(\beta_0, \beta_1, x_i) &= \beta_0 + \beta_1 \left(\frac{x_i - \bar{x}}{\sigma_x} \right) \end{aligned}$$

```
b0 ~ dnorm(0, .001)
b1 ~ dnorm(0, .001)
sigma ~ dunif(0, 100)
tau <- 1/sigma^2
xBar <- mean(x)
xSD <- sd(x)
for (i in 1:length(y)){
  mu[i] <- b0 + b1 * ((x[i] - xBar)/xSD)
  y[i] ~ dnorm(mu[i], tau)
}
```

Recovering unstandardized parameters

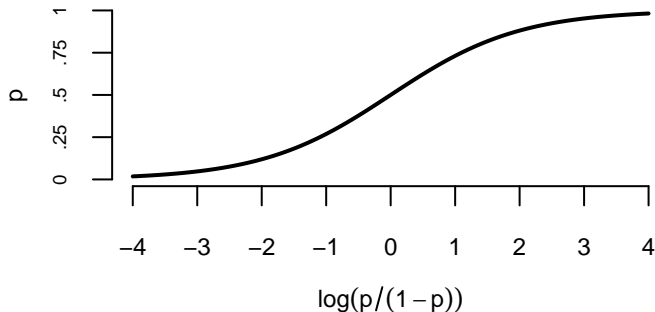
$$B_0 = \beta_0 - \frac{\beta_1 \bar{x}}{\sigma_x}$$
$$B_1 = \frac{\beta_1}{\sigma_x}$$

- This only works if there are not squared values or interactions.
- Generally, I back-transform predictions not parameters. (How?)

Bernoulli, data 0 or 1 (aka logistic)

$$\begin{aligned} [\beta_0, \beta_1 \mid \mathbf{y}] &\propto \prod_{i=1}^N \text{Bernoulli}(y_i \mid p_i) \\ &\quad \times \text{normal}(\beta_0 \mid 0, 2.7) \text{normal}(\beta_1 \mid 0, 2.7) \\ p_i &= \frac{e^{\beta_0 + \beta_1 x_i}}{e^{\beta_0 + \beta_1 x_i} + 1} \\ \text{logit}(p_i) &= \log\left(\frac{p_i}{1 - p_i}\right) = \beta_0 + \beta_1 x_i \end{aligned}$$

Log odds vs probability

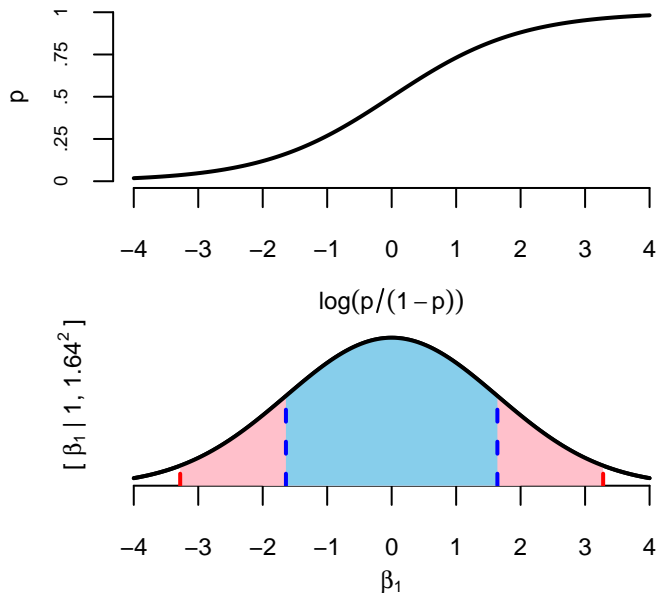


Exponentiate β_0 and β_1 and report odds and odds ratios.

$e^{\beta_1} = \frac{e^{\beta_0 + \beta_1 x}}{e^{\beta_0}}$ which is an odds ratio, i.e. the change in the odds.

For example: “The median odds of detecting weevil larvae in upland willow stems were 3.2 (95% CI: 2.3, 4.8) times greater than in riparian willow stems for stems of all sizes surveyed.”

Choosing reasonable flat priors on logit effects



Bayesian R^2

You can summarize a model's fit by computing its R^2 , or the proportion of variance in the response explained by the model predictors.

$$R^2 = 1 - \frac{\sum_{i=1}^n \epsilon_i^2}{\sum_{i=1}^n y_i^2}$$

where $\sum_{i=1}^n x_i = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$.

```
for(i in 1:length(y)){  
  log(mu[i]) <- b0 + b1 * x[i]  
  y[i] ~ dpois(mu[i])  
  e.resid[i] <- (y[i] - mu[i])^2  
  e.y[i] <- (y[i] - mean(y))^2  
}  
R2 <- 1 - sum(e.resid[])/sum(e.y[])
```