Bayesian Regression

Bayesian Modeling for Socio-Environmental Data

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

August 2017

Outline

- Understand Bayesian inference using familiar examples.
- Appreciate one-to-one relationship between math and JAGS code.
- Learn some programming tricks.

Notation

$$\underbrace{g\left(\boldsymbol{\theta}, x\right)}_{\text{deterministic model}}$$

Embedding the deterministic model

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

$$\left[\boldsymbol{\theta}, \sigma^2\right] \propto \left[y \mid \boldsymbol{\theta}, \sigma^2\right]\left[\boldsymbol{\theta}\right]\left[\sigma^2\right]$$

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

$$\left[\boldsymbol{\theta}, \sigma^2\right] \propto \left[y \mid g\left(\boldsymbol{\theta}, x\right), \sigma^2\right]\left[\boldsymbol{\theta}\right]\left[\sigma^2\right]$$
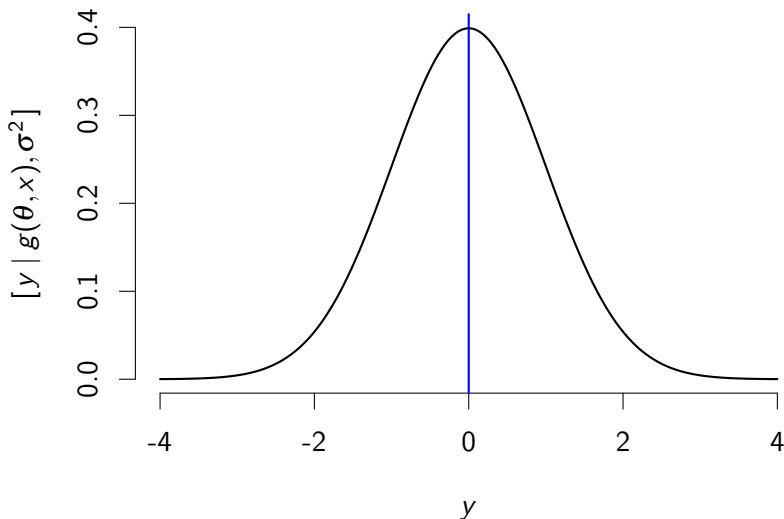
The general set up for a simple Bayesian regression model

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

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

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

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	lognormal or 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

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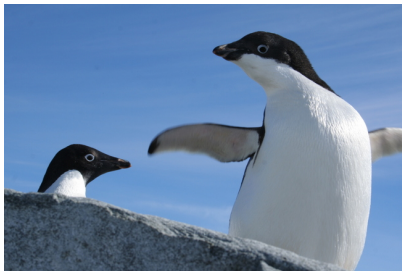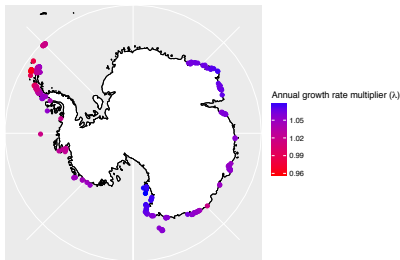

Photo c/o Heather J. Lynch

Normal data, continuous and real valued

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

```
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

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

```
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])
```

Bernoulli, data 0 or 1 (aka logistic)

$$
\begin{aligned}
[\beta_0, \beta_1 \mid \mathbf{y}] &\propto \prod_{i=1}^{N} \mathrm{Bernoulli}\big(y_i \mid g(\beta_0, \beta_1, x_i)\big) \\
&\quad \times \mathrm{normal}\big(\beta_0 \mid 0, 3\big) \, \mathrm{normal}\big(\beta_1 \mid 0, 6\big) \\
g(\beta_0, \beta_1, x_i) &= \frac{e^{\beta_0 + \beta_1 x_i}}{e^{\beta_0 + \beta_1 x_i} + 1}
\end{aligned}
$$

```
b0 ~ dnorm(0, .368)
b1 ~ dnorm(0, .368)
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)

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

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)

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

```
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)

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

```
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)

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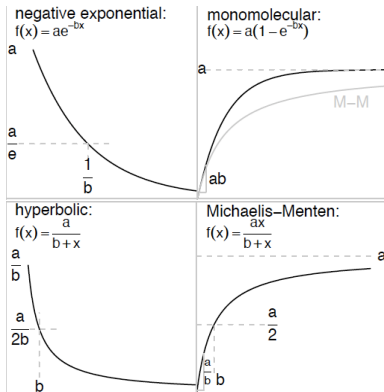g(\beta_0, \beta_1, x_i) &= \beta_0 + \beta_1 x_i
\end{aligned}
$$

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

negative exponential:
$f(x) = ae^{-bx}$

monomolecular:
$f(x) = a(1 - e^{-bx})$

hyperbolic:
$f(x) = \frac{a}{b+x}$

Michaelis–Menten:
$f(x) = \frac{ax}{b+x}$

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.
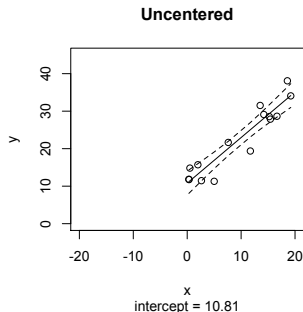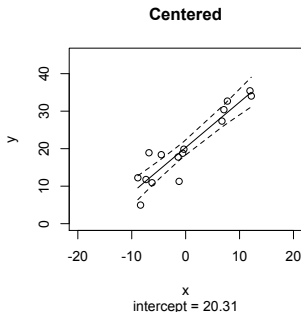
Centering predictor data

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

```
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

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

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

```
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?)