Bayesian Regression

Bayesian Modeling for Socio-Environmental Data

Christian Che-Castaldo

August 2016

Outline

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

Notation

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

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. . .

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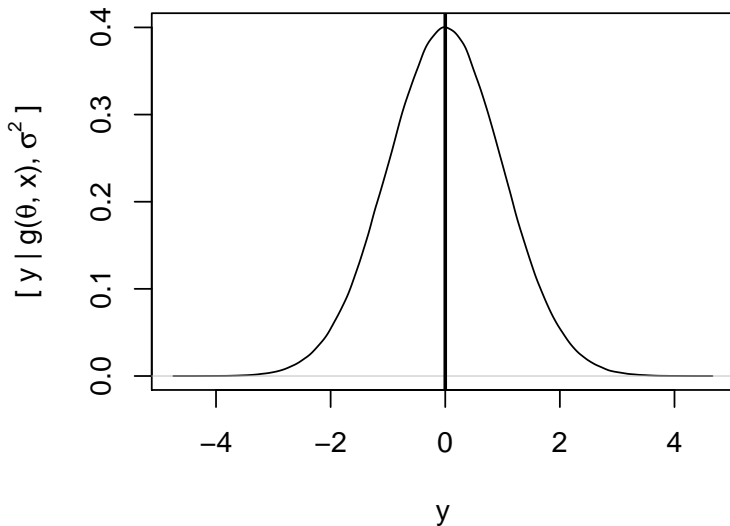
The general set up for a simple Bayesian regression model

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

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

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

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	$\mu = e^{\beta_0 + \beta_1 x}$	$\log\left(\mu\right) = \beta_0 + \beta_1 x$
0 or 1	Bernoulli	$\mu = \dfrac{\exp\left(\beta_0 + \beta_1 x\right)}{\exp\left(\beta_0 + \beta_1 x\right) + 1}$	$\mathrm{logit}\left(\mu\right) = \log\left(\frac{\mu}{1-\mu}\right) = \beta_0 + \beta_1 x$
$0 - 1$	beta	$\mu = \dfrac{\exp\left(\beta_0 + \beta_1 x\right)}{\exp\left(\beta_0 + \beta_1 x\right) + 1}$	$\mathrm{logit}\left(\mu\right) = \log\left(\frac{\mu}{1-\mu}\right) = \beta_0 + \beta_1 x$
continuous, strictly positive	lognormal or gamma	$\mu = e^{\beta_0 + \beta_1 x}$	$\log\left(\mu\right) = \beta_0 + \beta_1 x$

Lots of flexibility as a modeler

Continent-wide Adélie penguin population dynamics

$$z_{s,y} \sim \text{lognormal}(z_{s,y} \mid g(\beta_1, \beta_2, \beta_3, \beta_4, \epsilon_y, z_{s,y-1}), \sigma^{process})$$

$$g(\theta) = \log(z_{s,y-1}e^{\beta_1 + \beta_2\text{wsic}_{s,y} + \beta_3\text{ssic}_{s,y} + \beta_4\text{krill}_{s,y} + \epsilon_y})$$
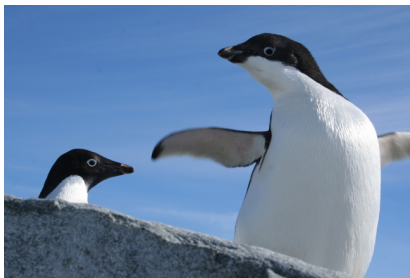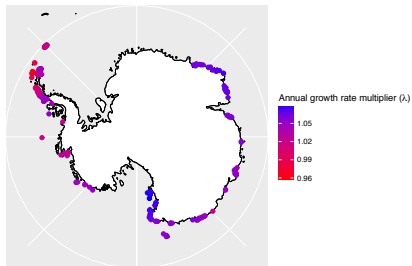

Annual growth rate multiplier (λ)
- 1.05
- 1.02
- 0.99
- 0.96

Photo c/o Heather J. Lynch

Normal data, continuous and real valued

$$
\begin{aligned}
\left[\beta_0, \beta_1, \sigma \mid \mathbf{y}\right] &\propto \prod_{i=1}^{N} \text{normal}\left(y_i \mid g\left(\beta_0, \beta_1, x_i\right), \sigma^2\right) \times \\
&\quad \text{normal}\left(\beta_0 \mid 0, .001\right) \text{normal}\left(\beta_1 \mid 0, .001\right) \times \\
&\quad \text{uniform}\left(\sigma \mid 0, 100\right) \\
g\left(\beta_0, \beta_1, x_i\right) &= \beta_0 + \beta_1 x_i
\end{aligned}
$$

```
b0 ~ dnorm(0, .0001)
b1 ~ dnorm(0, .0001)
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

$$
\begin{aligned}
\left[\beta_0, \beta_1 \mid \mathbf{y}\right] &\propto \prod_{i=1}^{N} \mathrm{Poisson}\left(y_i \mid g\left(\beta_0, \beta_1, x_i\right)\right) \times \\
&\quad \mathrm{normal}\left(\beta_0 \mid 0, .001\right) \mathrm{normal}\left(\beta_1 \mid 0, .001\right) \\
g\left(\beta_0, \beta_1, x_i\right) &= e^{\beta_0 + \beta_1 x_i}
\end{aligned}
$$

```
b0 ~ dnorm(0, .0001)
b1 ~ dnorm(0, .0001)
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)

$$[\beta_0, \beta_1 \mid \mathbf{y}] \propto \prod_{i=1}^{N} \mathrm{Bernoulli}\big(y_i \mid g\big(\beta_0, \beta_1, x_i\big)\big) \times$$

$$\mathrm{normal}\big(\beta_0 \mid 0, .001\big)\,\mathrm{normal}\big(\beta_1 \mid 0, .001\big)$$

$$g\big(\beta_0, \beta_1, x_i\big) = \frac{e^{\beta_0 + \beta_1 x_i}}{e^{\beta_0 + \beta_1 x_i} + 1}$$

```
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])
```
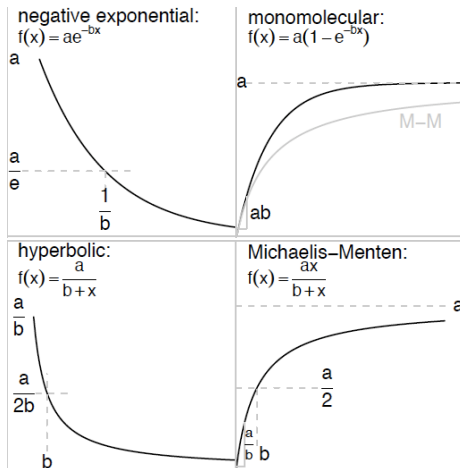
lognormal, data continuous and > 0

$$
\begin{aligned}
\left[\beta_0, \beta_1, \sigma \mid \mathbf{y}\right] &\propto \prod_{i=1}^{N} \text{lognormal}\left(y_i \mid g\left(\beta_0, \beta_1, x_i\right), \sigma^2\right) \times \\
&\quad \text{normal}\left(\beta_0 \mid 0, .001\right) \text{normal}\left(\beta_1 \mid 0, .001\right) \times \\
&\quad \text{uniform}\left(\sigma \mid 0, 100\right) \\
g\left(\beta_0, \beta_1, x_i\right) &= e^{\beta_0 + \beta_1 x_i}
\end{aligned}
$$

Talk about the interpretation of σ and the bounding trick

```
b0 ~ dnorm(0, .001)
b1 ~ dnorm(0, .001)
sigma ~ dunif(0, 100)
tau <- 1/sigma^2
for(i in 1:length(y)){
  mu[i] <- log(b0 + b1 * x[i])
    y[i] ~ dlnorm(max(.000001, mu[i]), tau)
}
```

Nonlinear regression

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

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

M–M

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

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

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.
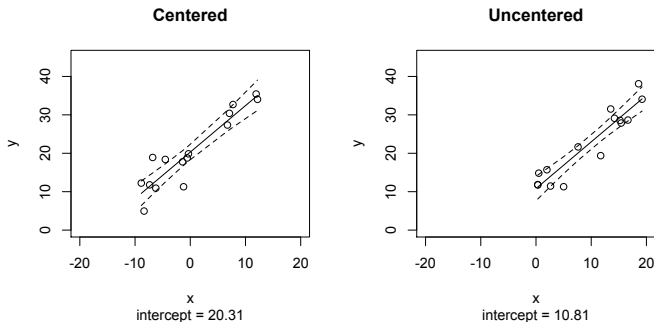- To make the intercept more easily interpretable.

Centering predictor data

$$
\begin{aligned}
\left[\beta_0, \beta_1, \sigma \mid \mathbf{y}\right] &\propto \prod_{i=1}^{N} \text{normal}\big(y_i \mid g\big(\beta_0, \beta_1, x_i, \bar{x}\big), \sigma^2\big) \times \\
&\quad \text{normal}\big(\beta_0 \mid 0, .001\big) \, \text{normal}\big(\beta_1 \mid 0, .001\big) \times \\
&\quad \text{uniform}\big(\sigma \mid 0, 100\big) \\
g\big(\beta_0, \beta_1, x_i\big) &= \beta_0 + \beta_1\big(x_i - \bar{x}\big)
\end{aligned}
$$

```
b0 ~ dnorm(0, .0001)
b1 ~ dnorm(0, .0001)
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

$$
\begin{aligned}
\left[\beta_0, \beta_1, \sigma \mid \mathbf{y}\right] \quad \propto \quad & \prod_{i=1}^{N} \text{normal}\big(y_i \mid g\big(\beta_0, \beta_1, x_i, \bar{x}, \sigma_x\big), \sigma^2\big) \times \\
& \text{normal}\big(\beta_0 \mid 0, .001\big) \, \text{normal}\big(\beta_1 \mid 0, .001\big) \times \\
& \text{uniform}\big(\sigma \mid 0, 100\big) \\
g\big(\beta_0, \beta_1, x_i\big) \quad = \quad & \beta_0 + \beta_1\Big(\frac{x_i - \bar{x}}{\sigma_x}\Big)
\end{aligned}
$$

```
b0 ~ dnorm(0, .0001)
b1 ~ dnorm(0, .0001)
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

$$
\begin{aligned}
y_i &= \beta_0 + \beta_1 \left(\frac{x_i - \bar{x}}{\sigma_x} \right) \\
y_i &= \beta_0 + \frac{\beta_1}{\sigma_x} - \frac{\beta_1 \bar{x}}{\sigma_x} \\
B_0 &= \beta_0 - \frac{\beta_1 \bar{x}}{\sigma_x} \\
B_1 &= \frac{\beta_1}{\sigma_x}
\end{aligned}
$$

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