Bayesian Regression

Models for Socio-Environmental Data

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

June 07, 2019

SESYNC

Outline

- Be able to write proper Bayesian regression models for different types of data.
- Appreciate one-to-one relationship between math and JAGS code.
- Be able to interpret coefficients of general linear models.

Where are we?

Deterministic models

| general linear |
| nonlinear |
| differential equations |
| difference equations |
| auto-regressive |
| occupancy |
| state-transition |
| integral-projection |

univariate and
multivariate

Types of data

| real numbers |
| non-negative real numbers |
| counts |
| 0 to 1 |
| 0 or 1 |
| counts in categories |
| proportions in categories |
| ordinal categories |

A great follow-up

This book should be in your library:

The general Bayesian set-up

Recall that the posterior distribution of the unobserved quantities conditional on the observed ones is proportional to their joint distribution:

$$[\theta|y] \propto [\theta, y].$$

The joint distribution can be factored into a likelihood and priors for simple Bayesian models:

$$\left[\theta, \sigma^2\right] = \left[y \mid \theta, \sigma^2\right]\left[\theta\right]\left[\sigma^2\right]$$

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

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

Simple Bayesian regression models

As always, we start with a deterministic model:

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

where β is a vector of regression coefficients and $\mathbf{x}_i$ is a vector or predictor variables. We use likelihood to connect the predictions of our model to data:

$$\underbrace{[y_i \mid \mu_i, \sigma^2]}_{\text{stochastic model}}$$

$$[\beta, \sigma^2 \mid \mathbf{y}] \propto \prod_{i=1}^{n} \left[y_i \mid g\left(\beta, x_i\right), \sigma^2\right] \left[\theta\right] \left[\sigma^2\right]$$

We choose appropriate deterministic functions (linear or non-linear) and appropriate probability distributions to compose specific models.
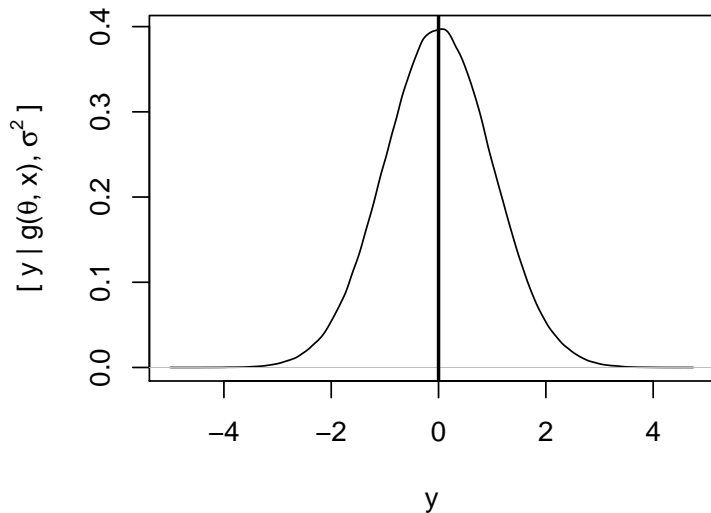
Identical notation

$$y_i = g(\beta, x_i) + \varepsilon_i$$
$$\varepsilon_i \sim \text{normal}(0, \sigma^2)$$

is the same as:

$$y_i \sim \text{normal}\big(g(\beta, x_i), \sigma^2\big),$$

but the second notation is much more flexible because it doesn't require additive errors.

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	$\mu = \frac{\exp(\beta_0 + \beta_1 x)}{\exp(\beta_0 + \beta_1 x) + 1}$	$\mathrm{logit}(\mu) = \log\left(\frac{\mu}{1-\mu}\right) = \beta_0 + \beta_1 x$
$[0, 1]$	beta	$\mu = \frac{\exp(\beta_0 + \beta_1 x)}{\exp(\beta_0 + \beta_1 x) + 1}$	$\mathrm{logit}(\mu) = \log\left(\frac{\mu}{1-\mu}\right) = \beta_0 + \beta_1 x$
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

Bayesian regression models are building blocks for more complex models.
For example, the continent-wide Adélie penguin population dynamics:

$$z_{s,y} \sim \text{lognormal}\big(z_{s,y} \mid g(\beta_1, \beta_2, z_{s,y-1}), \sigma^2\big)$$
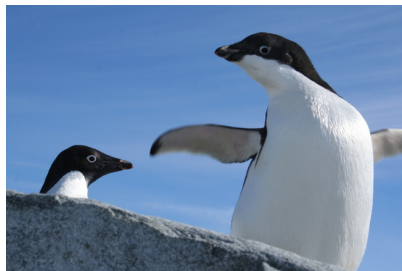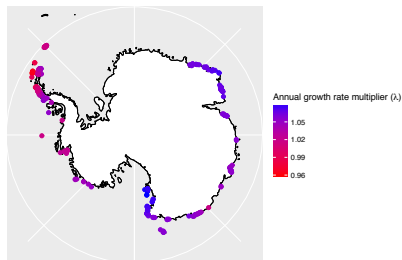$$g(\theta) = \log\big(z_{s,y-1} e^{\beta_1 + \beta_2 \text{wsic}_{s,y} + \beta_3 \text{ssic}_{s,y}}\big)$$

Photo c/o Heather J. Lynch

Normal data, continuous and real valued

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

```
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)
}
```

Counts, discrete and non-negative

You have collected some count data ($y = 12, 17, 1, 0, 31, \ldots, 25$) at n sites, along with a covariate (x) which you believe is likely to affect these counts. Write a model regressing y on x.

- Choose a specific deterministic and stochastic model.
- Write out the DAG, posterior distribution, and joint distribution for your model.
- Interpret the coefficients of your model.

Poisson, discrete and non-negative

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

```
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 non-negative

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

Exponentiate β_0 and β_1 and report counts or multiplicative changes in counts.
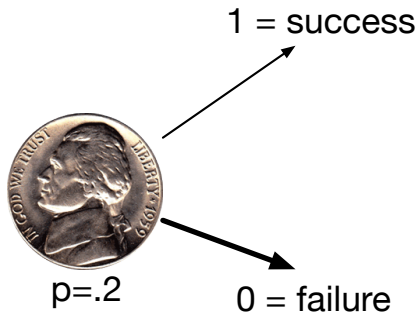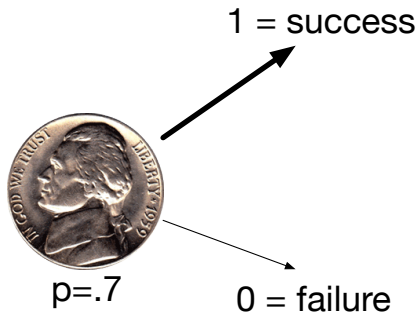
$e^{\beta_1} = \frac{e^{\beta_0 + \beta_1 x}}{e^{\beta_0}}$ which is the multiplicative change in the mean count per unit change in x.

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."

Binary data, 0 or 1 (aka logistic)

1 = success

p

0 = failure

Binary data, 0 or 1 (aka logistic)

1 = success
0 = failure
p=.7

1 = success
0 = failure
p=.2
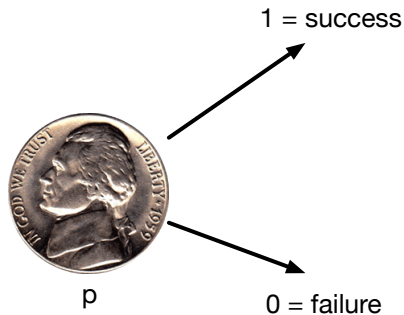
Binary data, 0 or 1 (aka logistic)

What happens when we want to make p a function of a continuous predictor?

- probability: p $[0,1]$
- odds: $\frac{p}{1-p}$ $[0,\infty)$
- log odds: $\log(\text{odds})$ $(-\infty,\infty)$

Moving between probabilty and log odds

- logit$() = \log\left(\frac{p}{1-p}\right)$: input is probability, output is log odds
- inverse logit$() = \frac{e^{\log(\text{odds})}}{e^{\log(\text{odds})}+1}$: input is log(odds), output is probability

Binary data, 0 or 1 (aka logistic)

inverse logit mapping: input is log(odds), output is probability

log(odds)
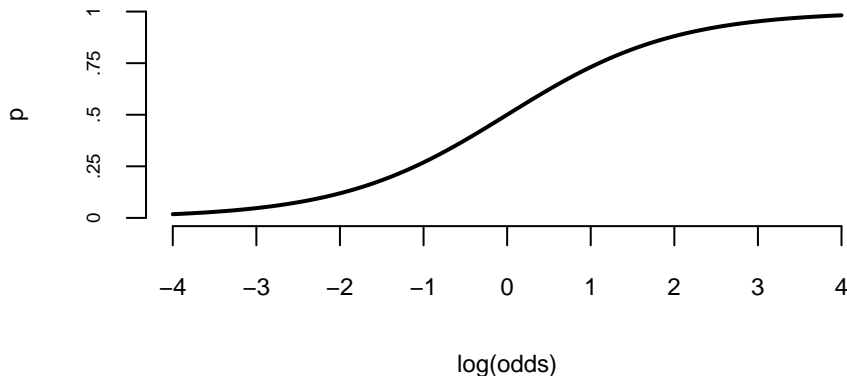
Binary data, 0 or 1 (aka logistic)

You have collected some binary data ($y = 1, 0, 0, 1, 1, 0, 1, \ldots, 1$) at n sites, along with a covariate (x) which you believe is likely to affect these counts. Write a model regressing y on x.

- Choose a specific deterministic and stochastic model.
- Write out the DAG, posterior distribution, and joint distribution for your model.
- Interpret the coefficients of your model.

Bernoulli, 0 or 1 (aka logistic)

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

```
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] ~ dbern(p[i])
```
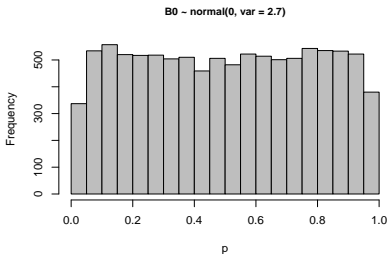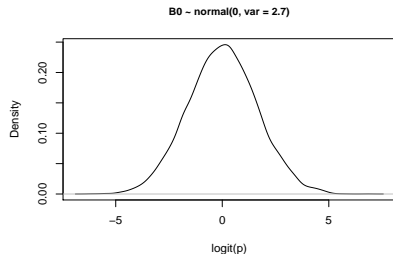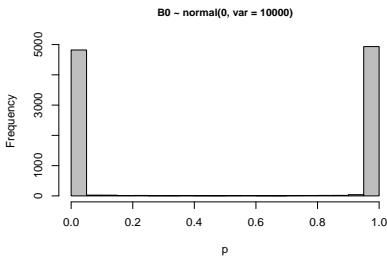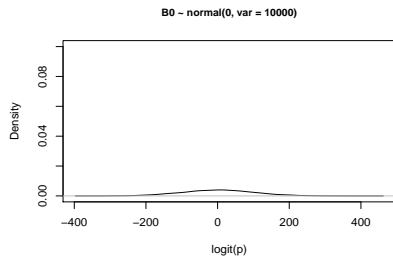
Interpreting logit coefficients

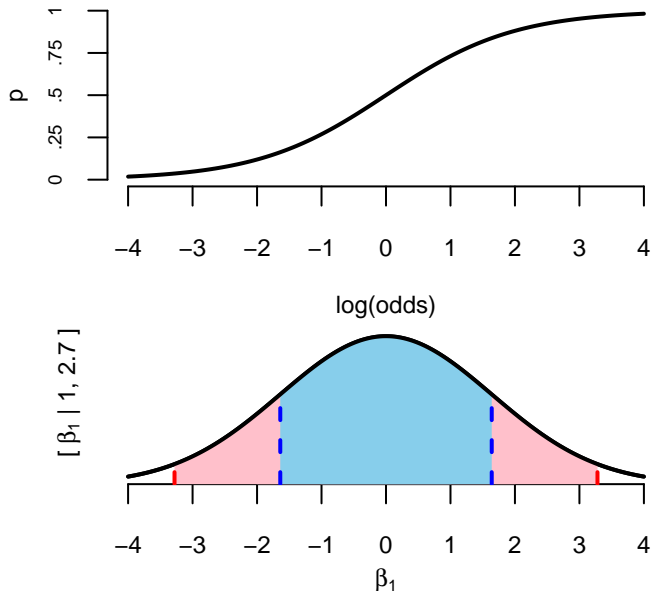Exponentiate β_0 and β_1 and report odds and odds ratios.

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

For example: "The odds of detecting weevils in upland willow stems were 3.2 (95% CI: 2.3, 4.8) times greater than detecting them in riparian willow stems."

Choosing reasonable flat priors on logit intercept

Choosing reasonable flat priors on logit effects

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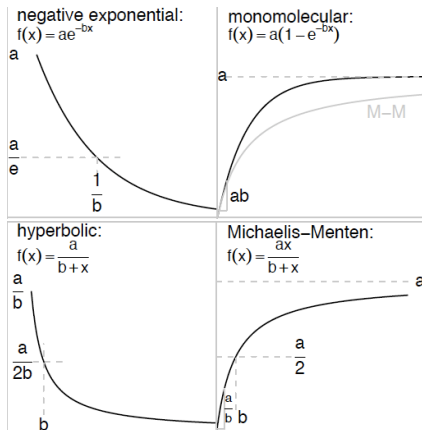
lognormal, data continuous and > 0 (not log link)

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

```
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)
}
```

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.