

# Metropolis: Logistic regression

## Logistic regression model

Younger male sparrows may or may not nest during a mating season, perhaps depending on their physical characteristics. Researchers have recorded the nesting success of 43 young male sparrows of the same age, as well as their wingspan.

We'll view  $Y_i$  as the binary outcome that sparrow  $i$  successfully nests. So  $Y_i$  is naturally modeled as Bernoulli. We will try to understand the variability in nesting success using its wingspan  $x_i$ . We'll fit a **logistic regression model**, which is the following *generalized linear* model:

$$Y_i | p_i \stackrel{\text{ind}}{\sim} \text{Bern}(p_i) \quad i = 1, \dots, n$$
$$\log \left( \frac{p_i}{1 - p_i} \right) = \beta_0 + \beta_1 x_i$$

Note:  $\mathbb{E}[Y_i | x_i, \beta_0, \beta_1] = p_i$ .

The log-odds  $\log(\frac{p_i}{1-p_i})$  is known as the “logit link”. Note that this is a “generalized linear” model because some function of the conditional mean is linear in the predictors, as opposed to the mean itself.

Note that in the model above,  $\beta_0$  and  $\beta_1$  are the unknown parameters. Once we know  $\beta_0$  and  $\beta_1$ ,  $p_i$  is deterministic:

$$\log \left( \frac{p_i}{1 - p_i} \right) = \beta_0 + \beta_1 x_i$$
$$\frac{p_i}{1 - p_i} = e^{\beta_0 + \beta_1 x_i}$$
$$p_i = (1 - p_i) e^{\beta_0 + \beta_1 x_i}$$
$$p_i (1 + e^{\beta_0 + \beta_1 x_i}) = e^{\beta_0 + \beta_1 x_i}$$
$$p_i = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_i)}}$$

## Priors and likelihood

We might take the following independent priors for the two coefficients:

$$\beta_0 \sim N(\mu_0, \sigma_0^2)$$

$$\beta_1 \sim N(\mu_1, \sigma_1^2)$$

Now that we have the sampling model and priors, we should try to get the posterior to understand the effect of wingspan on probability of successfully nesting. In order to do so, we will use the Metropolis Algorithm to obtain samples from  $f(\beta_0, \beta_1 | \mathbf{y})$ .

Let's try to write the likelihood and then the log-likelihood to later use in the acceptance ratio step. To make things clearer, note that  $f(y_i | x_i, p_i) \equiv f(y_i | x_i, \beta_0, \beta_1)$ .

$$f(\mathbf{y} | \mathbf{x}, \beta_0, \beta_1) = f(\mathbf{y} | \mathbf{x}, \mathbf{p}) = \prod_{i=1}^n f(y_i | x_i, p_i) = \prod_{i=1}^n p_i^{y_i} (1 - p_i)^{1-y_i}$$

So the log-likelihood is:

$$\log f(\mathbf{y} | \mathbf{x}, \beta_0, \beta_1) = \log \prod_{i=1}^n p_i^{y_i} (1 - p_i)^{1-y_i} = \sum_{i=1}^n [y_i \log p_i + (1 - y_i) \log(1 - p_i)]$$

## Metropolis for 2D (or > 1D) case

To go from the 1D case to multivariate case for the Metropolis Algorithm, we will be like a Gibbs sampler where we iterate through Metropolis updates for each parameter, conditioning on the current values of the remaining parameters. That is, we will alternate between:

1. Doing a Metropolis update/step for  $\beta_0$ , holding  $\beta_1$  fixed. That is, proposing  $\beta_0^*$  given the current value of  $\beta_0$ .
2. Doing a Metropolis update/step for  $\beta_1$ , holding  $\beta_0$  fixed.

We might call this “component-wise” Metropolis.

Note that in our model under the choice of independent priors, the acceptance ratio is quite easy to calculate! For example, if we are currently at the  $s$ -th iteration and proposing  $\beta_0^*$ , we want to evaluate the following acceptance ratio (which we will log in the code):

$$\begin{aligned} r &= \frac{f(\beta_0^*, \beta_1^{(s-1)} | \mathbf{y})}{f(\beta_0^{(s-1)}, \beta_1^{(s-1)} | \mathbf{y})} = \frac{f(\mathbf{y} | \beta_0^*, \beta_1^{(s-1)}) f(\beta_0^*) f(\beta_1^{(s-1)})}{f(\mathbf{y})} \times \frac{f(\mathbf{y})}{f(\mathbf{y} | \beta_0^{(s-1)}, \beta_1^{(s-1)}) f(\beta_0^{(s-1)}) f(\beta_1^{(s-1)})} \\ &= \frac{f(\mathbf{y} | \beta_0^*, \beta_1^{(s-1)})}{f(\mathbf{y} | \beta_0^{(s-1)}, \beta_1^{(s-1)})} \times \frac{f(\beta_0^*)}{f(\beta_0^{(s-1)})} \end{aligned}$$

## Code

```
library(tidyverse)
sparrow <- read_csv("../handouts/sparrows.csv")
# when fitting logistic regression, we like to standardize predictors
x <- scale(sparrow$wingspan)
y <- sparrow$nest

# set prior values
mu0 <- 0
mu1 <- 0
sigma0 <- 5
sigma1 <- 5

# initialize
beta0 <- 0
beta1 <- 0

# get p based on initial values of beta0, beta1
p <- 1 / (1 + exp(-(beta0 + beta1 * x)))
prop_sd0 <- 1 # proposal sd for beta0
prop_sd1 <- 1 # proposal sd for beta1
S <- 5000
BETA <- matrix(NA, nrow = S, ncol = 2)
beta_acc <- matrix(0, nrow = S, ncol = 2) # col1 = beta0 accept, col2 = beta1 accept
set.seed(412)
for(s in 1:S){
  # propose beta0
  beta0_prop <- rnorm(1, beta0, prop_sd0)
  # evaluate p at proposed value of beta0, holding beta1 where it currently is
  p_prop <- 1/(1 + exp(-(beta0_prop + beta1 * x)))
  # get acceptance ratio (log scale)
  log_r <- sum(y*log(p_prop) + (1-y)*log(1-p_prop)) -
    (sum(y*log(p) + (1-y)*log(1-p))) +
    dnorm(beta0_prop, mu0, sigma0, log = T) -
    dnorm(beta0, mu0, sigma0, log = T)
  # decide
  if(log(runif(1)) < log_r){
    beta0 <- beta0_prop
    p <- p_prop # update p as well!
    beta_acc[s,1] <- 1
  }
}
```

```

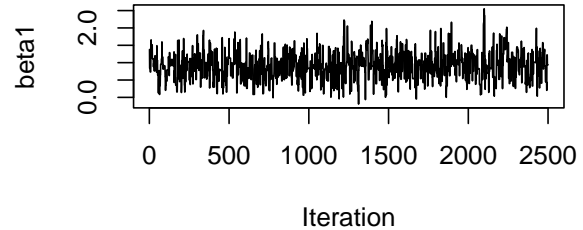
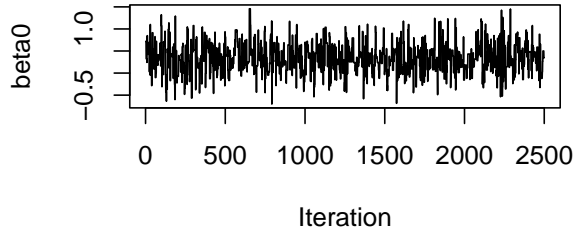
# do it all again, but now for beta1
beta1_prop <- rnorm(1, beta1, prop_sd1)
p_prop <- 1/(1 + exp(- (beta0 + beta1_prop * x)))
log_r <- sum(y*log(p_prop) + (1-y)*log(1-p_prop)) -
  (sum(y*log(p) + (1-y)*log(1-p))) +
  dnorm(beta1_prop, mu1, sigma1, log = T) -
  dnorm(beta1, mu1, sigma1, log = T)
if(log(runif(1)) < log_r){
  beta1 <- beta1_prop
  p <- p_prop
  beta_acc[s,2] <- 1
}

# store
BETA[s,] <- c(beta0, beta1)
}
BETA <- BETA[-c(1:(S/2)),]

```

Diagnose acceptance probabilities for  $\beta_0$  and  $\beta_1$  proposals:

```
[1] 0.3886 0.4290
```



## Posterior inference

As before, we can approximate posterior quantities:

	Posterior mean	2.5%	97.5%
$\beta_0$	0.340	-0.326	1.049
$\beta_1$	0.901	0.134	1.718

Lastly, as an example of a **derived quantity**, I can obtain (the posterior distribution of) the probability of nesting success for a bird with a wingspan of 14cm. That is, I can obtain the

distribution of  $\Pr(Y_i = 1|x_i = 14, \beta_0, \beta_1)$  using my MCMC samples and the formula for  $p_i$ :

$$p_i = \frac{1}{1 + e^{-(\beta_0 + \beta_1 \times 14^*)}}$$

\* Note, it's actually the standardized version of 14cm we want given our previous work.

```
xbar <- mean(sparrow$wingspan)
s_x <- sd(sparrow$wingspan)
xi_scaled <- (14 - xbar)/s_x # standardize wingspan of 14cm
prob_i <- (1 + exp(-(BETA[,1] + BETA[,2] * xi_scaled)))^(-1)
```

