This problem involves a Bayesian generalized linear model (GLM) to analyze a moth predation dataset. The example is modified from the Statistical Sleuth, and a summary of the problem is as follows. Near Liverpool, England, air pollution has caused the discoloration (blackening) of tree trunks, and the tree trunks become lighter as one travels from the city to the countryside, which experiences lower levels of pollution. One particular species of moth rests during the day on tree trunks (it is active at night), and their coloration acts as camouflage to protect them against predatory birds. In Liverpool, a high percentage of these moths are characterized by the dark morph, and the percentage of "typical" morphs (salt-and-pepper, "light") increases towards the countryside. Researchers used this gradient in tree trunk color to study the intensity of natural selection. They selected 7 sites between Liverpool and the countryside. They glued equal numbers of dead light and dead dark moths to tree trunks at each site. After 24 hours, they counted the number of moths of each morph that had been removed, presumably by predators.

We will construct a GLM—in particular, a logistic regression—to evaluate if the probability of removal differed between light and dark morphs, and if this difference depended on distance from Liverpool. We will also try to incorporate a site random effect, which may or may not be appropriate (food for thought). The dataset includes variables for Site (1, 2, …, 7), Morph (1 = light, 2 = dark), Dist (distance from Liverpool, km), Placed (number of moths originally glued to tree trunks), and Removed (number of moths removed from tree trunks). Let y_i, n_i, and x_i represent Removed, Placed, and Dist, respectively, and let m be the Morph index and s the Site index. We will begin with the following simple model that does not include site random effects:

$$y_i \sim Binomial\left(n_i, p_i\right)$$
$$\text{logit}(p_i) = \beta_{1,m(i)} + \beta_{2,m(i)} \cdot x_i$$
$$\beta_{1,m}, \beta_{2,m} \sim N\left(0, 0.00001\right) \quad for\ m = 1, 2$$

Are you able to identify the likelihood part, the process (or mean) model, and the prior part?

Activities

1. Program the above GLM (logistic regression) in R / JAGS.

2. Come-up with reasonable starting values for the parameters.

3. Run the above model with jags.model, using the initials from part 2. Using coda.samples, update the JAGS model for 5,000 iterations and monitor deviance and the β's. Evaluate convergence and mixing; compute posterior statistics. Upon running the model, consider the following questions:

    a. Does the probability of predation (removal) vary with distance from Liverpool?
    b. Does the probability of predation differ between morphs?
    c. Does the effect of distance on predation differ between morphs?
    d. How well does the model fit the data? How would you evaluate this?

4. Modify the above model to include a site random effect in the linear predictor (call this ε, where for site s, $\varepsilon_s \sim N(0, \tau_\varepsilon)$. Initialize the model with jags.model for 3 chains, with initials updated based on the model runs in part 3 above. Using coda.samples, run the model for 5,000 iterations and monitor deviance, the β's, the site random effects (ε_s), and their standard deviation (σ_ε). Quickly check for convergence so ensure that the burn-in isn't included in the calculation of posterior statistics and plots. Upon running the model, consider the following questions:

   a. Did the estimate of the β's change upon adding a site random effect?
   b. Did addition of the additive random effect created any noticeable identifiability problems?
   c. Does the probability of predation vary with distance from Liverpool?
   d. Does the probability of predation differ between morphs?
   e. Does the effect of distance on predation differ between morphs?
   f. How well does the model fit the data? Is fit improved by adding a site random effect?
   g. Why is adding a site random effect probably not appropriate here?