Model Selection Lab: Poisson Regression

Model Statement

We will use the continental U.S. bird richness data set for this lab. In a previous lab on MCMC we used a simple linear regression model with the log of the counts ($\log(y_i)$, for $i = 1, \ldots, n$) as the response variable and the state area as the predictor variable (i.e., covariate x_i).

For simplicity in the MCMC lab we transformed the counts using the log function and modeled this transformed response variable with a Gaussian distribution:

$$\log(y_i) \sim \mathrm{N}(\beta_0 + \beta_1 x_i, \sigma^2) \ . \tag{1}$$

Now suppose that we wish to model the bird counts (y_i) by state, directly. The support of y_i are the non-negative integers. Thus, a reasonable starting place for a data model for y_i is the Poisson distribution such that

$$y_i \sim \mathrm{Pois}(\lambda_i) \ . \tag{2}$$

Now we can link the "intensities" (λ_i) to the covariates (x_i) and regression coefficients ($\beta_0, \ldots, \beta_p$) using a log link function

$$\log(\lambda_i) = \beta_0 + \beta_1 x_{1,i} + \ldots + \beta_p x_{p,i} \ , \tag{3}$$

for a set of covariates ($x_{j,i}, j = 1, \ldots, p$). An important point here is that the λ_i are linked deterministically to the regression parameters β_j. Thus, we only need a prior for β_j (for $j = 1, \ldots, p$). It is common to see regression part of the model written as $\log(\lambda_i) = \beta_0 + \mathbf{x}_i'\boldsymbol{\beta}$ or $\log(\lambda_i) = \mathbf{x}_i'\boldsymbol{\beta}$, depending on whether the intercept is included in $\boldsymbol{\beta}$ (in the latter case, the first element of vector $\mathbf{x}_i$ is 1).

A reasonable prior for unconstrained regression coefficients is Gaussian (because the support for β_j includes all real numbers), thus we could use

$$\beta_j \sim \mathrm{N}(\mu_j, \sigma_j^2) \text{ for } j = 1, \ldots, p \ , \tag{4}$$

as priors. Note that it is common to specify the same prior mean and variance for all regression coefficients, but you don't have to.

Information Criteria

Recall that the deviance information criterion (DIC):

$$\text{DIC} = \hat{D} + 2p_D \ , \tag{5}$$

for $p_D = \bar{D} - \hat{D}$. These different forms of deviance can be computed using MCMC output from our model using

$$\hat{D} = -2\sum_{i=1}^{n} \log\left(\text{Pois}(y_i|\hat{\lambda}_i)\right) \ , \tag{6}$$

$$\bar{D} = -2\frac{\sum_{t=1}^{T}\sum_{i=1}^{n} \log\left(\text{Pois}(y_i|\exp(\beta_0^{(t)} + \beta_1^{(t)}x_{1,i} + \ldots + \beta_p^{(t)}x_{p,i}))\right)}{T} \ , \tag{7}$$

where $\hat{\lambda}_i$ is the posterior mean of λ and $\beta_j^{(t)}$ is the t^{th} MCMC sample (for $j = 1, \ldots, p$ and T total MCMC samples).

Similarly, the Watanabe-Akaike information criterion is

$$\text{WAIC} = -2\sum_{i=1}^{n} \text{lppd}_i + 2p_D \ , \tag{8}$$

where the 'lppd' stands for log posterior predictive density for y_i and can be calculated using MCMC as

$$\text{lppd}_i = \log\left(\frac{\sum_{t=1}^{T} \text{Pois}(y_i|\exp(\beta_0^{(t)} + \beta_1^{(t)}x_{1,i} + \ldots + \beta_p^{(t)}x_{p,i}))}{T}\right) \ , \tag{9}$$

and where Gelman et al. (2013) recommend calculating p_D as

$$p_D = \sum_{i=1}^{n}\left(\frac{\sum_{t=1}^{T}(\log(\text{Pois})_i^{(t)} - \sum_{t=1}^{T}\log(\text{Pois})_i^{(t)}/T)^2}{T}\right) \ , \tag{10}$$

where, $\log(\text{Pois})_i^{(t)} = \log\left(\text{Pois}(y_i|\exp(\beta_0^{(t)} + \beta_1^{(t)}x_{1,i} + \ldots + \beta_p^{(t)}x_{p,i}))\right)$.

The D_∞ criterion based on posterior predictive loss is defined as

$$D_\infty = \sum_{i=1}^{n}(y_i - E(\tilde{y}_i|\mathbf{y}))^2 + \sum_{i=1}^{n}\text{Var}(\tilde{y}_i|\mathbf{y}) \ . \tag{11}$$

To calculate $E(\tilde{y}_i|\mathbf{y})$ and $\text{Var}(\tilde{y}_i|\mathbf{y})$, first draw $\tilde{y}_i^{(t)} \sim \text{Pois}(y_i|\exp(\beta_0^{(t)} + \beta_1^{(t)}x_{1,i} + \ldots + \beta_p^{(t)}x_{p,i})$ on the t^{th} MCMC iteration for all $t = 1, \ldots, T$. Then $E(\tilde{y}_i|\mathbf{y})$ is the sample mean of the $\tilde{y}_i^{(t)}$ and $\text{Var}(\tilde{y}_i|\mathbf{y})$ is the sample variance over the T MCMC iterations.