A model of biological and phylogenetic effects on mammalian co-occurrence

Peter D Smits

Committee on Evolutionary Biology, University of Chicago

February 2, 2015

1 Methods

1.1 Species information

1.1.1 Informal phylogeny

1.2 Biogeographic network

1.3 Co-occurrence model

Poisson distribution with inverse-scale λ.

$$\text{Poisson}(n_i|\lambda_i) = \frac{1}{n_i!}\lambda^{n_i}\exp(-\lambda_i) \tag{1}$$

$$\lambda_i = \exp(\beta^T\mathbf{X_i} + h_i) \tag{2}$$

Negative binomial distribution parameterized as mean μ and overdispersion ϕ.

$$\text{Negative binomial}(y_i|\mu_i,\phi) = \binom{y_i + \phi - 1}{y_i}\left(\frac{\mu_i}{\mu_i + \phi}\right)^{y_i}\left(\frac{\phi}{\mu_i + \phi}\right)^{\phi} \tag{3}$$

$$\mu_i = \exp(\beta^T\mathbf{X_i} + h_i) \tag{4}$$

$$\phi \sim \text{halfCauchy}(2.5) \tag{5}$$

$\mathbf{X}$ is defined as an $n \times K$ matrix where n is the number of observations and K is the number of covariates of interest. As described above, the covariates of interest are the dietary and locomotor categories of a species and its body mass. While body mass is a continuous covariate, dietary and locomotor categories are index variables that need to be transformed into multiple binary covariates or indicator variables. To do this, for each of these variables were each transformed
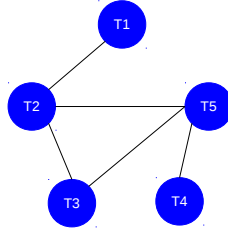
1

(a)
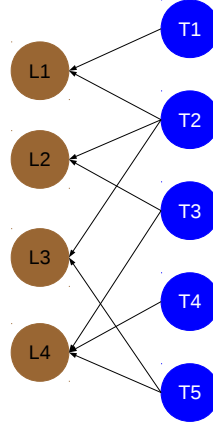
(b)

Figure 1: Bipartite and one mode graphs.

into $n \times (k-1)$ matrices where $k-1$ is the number of categories of the index variable (3 and 4, respectively). Only $k-1$ indicator variables are necessary as the intercept takes on the remaining category. Finally, a column of 1-s in included in the matrix $\mathbf{X}$ whose corresponding β coefficient is the intercept. In total, K equals 7.

For the parameterizations of the means for both the Poisson (Eq. 2) and negative binomial models (Eq. 4), a unique coefficient β is assigned to each of the covariates and are given a diffuse, weakly informative prior ($\beta_k \sim \mathcal{N}(0, 10)$).

h, or phylogenetic effect, is defined as a random multivariate normally distributed

variable whose covariance matrix known up to a constant σ_p (Eq. 6).

$$h \sim \text{Multivariate } \mathcal{N}(0, \boldsymbol{\Sigma}) \qquad (6)$$
$$\boldsymbol{\Sigma} = \sigma_p^2 \mathbf{V}_{phy}$$
$$\sigma_p \sim \text{halfCauchy}(2.5).$$

The overdispersion parameter ϕ of the negative binomial distribution was given weakly informative half-Cauchy prior distribution (Eq. 5).

Graphical summaries of the Poisson and negative binomial models, along with all values for the prior distributions, are presented in Figures 2 and 3, respectively.
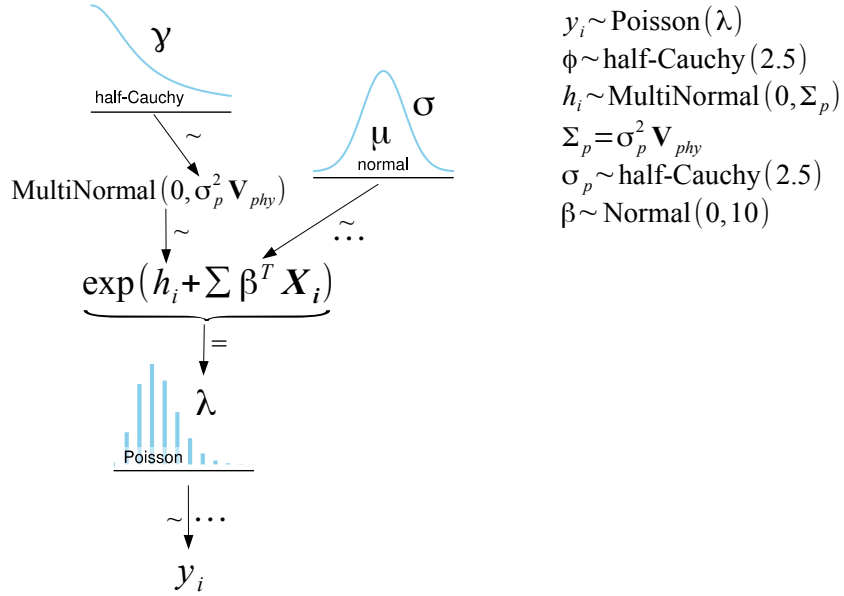

Figure 2: Poisson model

1.4 Posterior predictive checks

A models utility in making predictive and descriptive statements is dependent on its fit to the data and the questions of interest.

1.5 Predictive power

$$y_i \sim \text{NegBinom}(\mu, \phi)$$
$$\phi \sim \text{half-Cauchy}(2.5)$$
$$h_i \sim \text{MultiNormal}(0, \Sigma_p)$$
$$\Sigma_p = \sigma_p^2 \mathbf{V}_{phy}$$
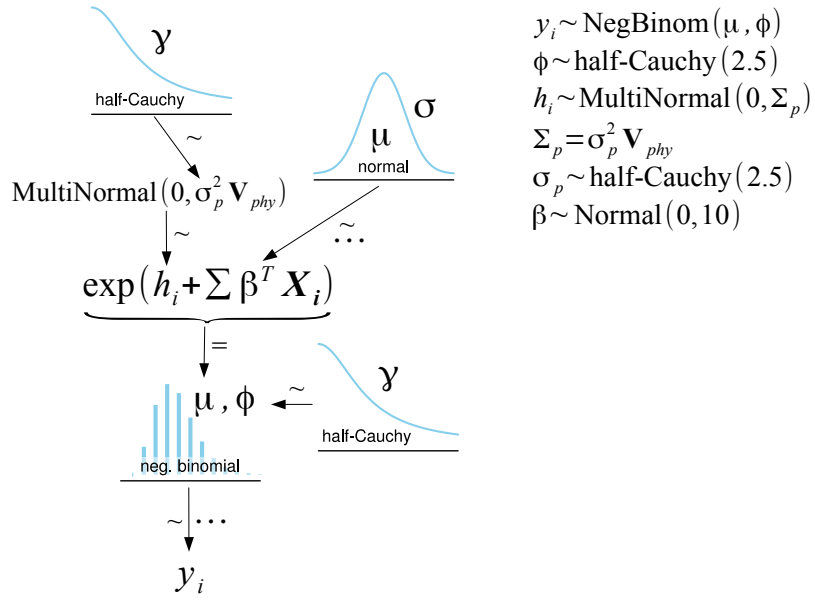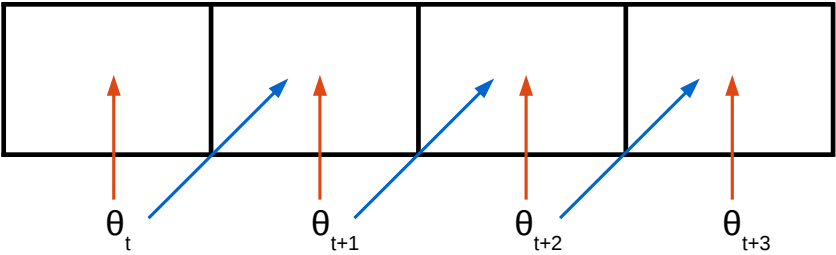$$\sigma_p \sim \text{half-Cauchy}(2.5)$$
$$\beta \sim \text{Normal}(0, 10)$$

Figure 3: Negative binomial model

Figure 4: CAPTION