

Random effect multinomial model with Dirichlet prior

A. J. Rominger

21 April 2017

The model

Given the number of sequencing reads assigned to each of S species, we want to estimate the number of individuals for each species that went into the sequencing run, call that vector x . We know the total number of reads N_{reads} , the number of species S , the total number of individuals N that were sequenced, and the vector of number of reads per species x_{read} . We assume that stochastic evolutionary processes have led to some species having a greater propensity to be sequenced. These stochastic processes we summarize in a random effect ν . We further assume that within orders these random effects are constant, thus each species within the same order gets the same random effect.

Thus our model for the number of reads is:

$$x_{reads} \sim \text{multinom} \left(N_{reads}, \frac{x\nu}{\sum_i x_i \nu_i} \right)$$

Thus the vector of probabilities for the multinomial distribution is proportional to the unknown abundance times the unknown random effect ν . Given that ν is constant within orders, varies across orders, and must be positive, we model it as

$$\log(\nu_{order=j}) \sim \text{norm}(0, \sigma_\nu^2)$$

That is, for order j the log of the random effect is distributed normally with mean 0 and variance σ_ν^2 .