Appendices

1. Simlulated community structures

I randomly assigned each species pair an interaction magnitude, β_{ij}, drawn from an exponential distribution with rate parameter 1. I also randomly assigned three quarters of the interaction coefficients to be negative, and the remaining quarter of the coefficients to be positive.

For the simulated landscapes where the abiotic environment was constant across locations, each species' α coefficient was drawn from a normal distribution with mean -1 and standard deviation 1. For the remaining landscapes, two environmental variables, x_1 and x_2 were sampled from independent standard normals with different values for each location. Each species' response to these two environmental variables was a linear function of these two environmental variables, with coefficients drawn from normal distributions with mean 0 and standard deviation 2. In this way, species' α coefficients (and thus their occurrence probabilities) depended on external environmental factors.

Once the "true" coefficients had been generated for each site on the landscape, I generated possible landscapes using Gibbs sampling. In each round of Gibbs sampling, I cycled through all the species, randomly updating each one's presence/absence vector in response to its conditional occurrence probability:

$$p(y_i) = \text{logistic}(\alpha_i + \sum_j \beta_{ij} y_j),$$

where the logistic function is $\frac{1}{1+e^{-x}}$ (Murphy 2012). After 1000 rounds of sampling, I continued this procedure until I obtained a landscape matrix where all of the species occurred at least once. I treated the resulting matrices as "observed" data for analysis.

2. Log-likelihood gradient

The *energy* of a Markov network is defined as

$$E(y; \alpha, \beta) = -\sum_i \alpha_i y_i - \sum_{i \neq j} \beta_{ij} y_i y_j.$$

The energy function can be used to define the log-likelihood of a given y vector as

$$\log \mathcal{L}(y; \alpha, \beta) = -E(y; \alpha, \beta) - \log(Z(\alpha, \beta)).$$

Here, $Z(\alpha, \beta)$ is the *partition function*, a scaling factor defined as

$$Z(\alpha, \beta) = \sum_{y \in Y} e^{-E(y;\alpha,\beta)},$$

where Y is the set of all possible y vectors.

The partial derivative of the log-likelihood with respect to α_i is

$$\frac{\partial}{\partial \alpha_i} \log \mathcal{L}(y; \alpha, \beta) = y_i - p(y_i; \alpha, \beta).$$

The first term, y_i, is zero if species i is absent in the observed assemblage and one if it's present. The latter term, $p(y_i; \alpha, \beta)$, describes the expected probability of observing species i under the current values of α and β. It comes from the derivative of the partition function, as derived in (learning Boltzmann machines, Murphy 2012, etc.). Following the gradient of α_i adjusts the expected probability of observing species i until it matches the observed value and the two terms in the gradient cancel one another out.

The partial derivative of the log-likelihood with respect to β_{ij} can be derived similarly as

$$\frac{\partial}{\partial \alpha_i} \log \mathcal{L}(y; \alpha, \beta) = y_i y_j - p(y_i y_j; \alpha, \beta).$$

Following this gradient adjusts the expected probability of co-occurrence between species i and species j until this value matches the observed co-occurrence frequency and the two terms in the gradient cancel one another out.

References

Murphy, K. P. 2012. Machine Learning: A Probabilistic Perspective. The MIT Press.