Web-based Supplementary Materials for "Biodiversity Analysis Using Rank Abundance Distributions" by S. D. Foster and P. K. Dunstan).

January 14, 2009

Web Appendix A : Computation for Modified Dirichlet-Multinomial Model

We maximise the log-likelihood using numerical optimisation. The beta-binomial distribution for parameters N, θ_j and θ is

$$\Pr\left(n_j; N, \theta_j, \theta\right) = \binom{N}{n_j} \frac{B(n_j + \theta_j, N - n_j + \theta)}{B(\theta_j, \theta)} \tag{1}$$

where $B(\cdot,\cdot)$ is the beta function (Abramowitz & Stegun, 1964) and can be expressed as

$$B(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}.$$

The log-likelihood for all of the T sites is

$$\ell_{\boldsymbol{n}}(\nu, \boldsymbol{\tau}_{\boldsymbol{n}}, \theta; \{\boldsymbol{n}\}) = \sum_{i=1}^{T} \left[\sum_{k=1}^{S} \log\left\{ \Pr\left(n_k | \boldsymbol{n}^{(k-1)} \right) \right\} \right]$$

where all the probability distribution functions are beta-binomial and their parameters are defined by their arguments. For clarity we have removed notation that shows the variable n_k and its distribution vary with site.

Calculation of the log-likelihood requires special consideration as large values of θ and θ_j will dominate the contributions from n_j and N, and are likely to cause machine round-off errors. This is a significant problem when optimising using numerical methods.

The machine error problem can be alleviated using a number of methods. The first is to explicitly express the beta-binomial distribution in terms of Stirling's approximation to the log-gamma function (Abramowitz & Stegun, 1964, page 257). This approximation, of some predefined order, is frequently used when evaluating the log-gamma function. We use a first-order Stirling's approximation and note that there are many simplifications which make the approximation used to calculate (1) more robust to machine error. The actual approximation that we use for a single site's likelihood contribution is

$$\begin{aligned}\log\left(\Pr\left(n_j; N, \theta_j, \theta\right)\right) \approx \log\binom{N}{n_j} &+ n_j \log\left(n_j + \theta_j\right) + \left(\theta_j - \frac{1}{2}\right)\log\left(1 + \frac{n_j}{\theta_j}\right) \\ &+ (N - n_j)\log\left(N - n_j + \theta\right) + \left(\theta - \frac{1}{2}\right)\log\left(1 + \frac{N - n_j}{\theta}\right) \\ &- N\log\left(N + \theta_j + \theta\right) - \left(\theta_j + \theta - \frac{1}{2}\right)\log\left(1 + \frac{N}{\theta_j + \theta}\right).\end{aligned}$$

The second and more effective method for increasing numerical stability involves altering the log-likelihood function to a limiting form when these unstable conditions are likely to occur. Observe that when the over-dispersion factor for the beta-binomial approaches 1 the beta-binomial distribution approaches the binomial distribution. This implies that an approximation can be obtained by using the binomial distribution when the over-dispersion factor is 1 to within predefined tolerances. The condition for defining the values of the over-dispersion factor is simply

$$\frac{\left(N - \sum_{k=1}^{j-1} n_k\right) - 1}{1 + \phi_j \left(\theta_j + \theta - \sum_{k=1}^{j} \theta_k\right)} < \epsilon$$

where ϵ is some small real number. We chose $\epsilon = 10e^{-10}$.

Numerical optimisation of any function can be problematic and this situation is no exception. It has been our experience that standardising covariates and using orthogonal contrasts for polynomial terms can alleviate the problem. When estimates are required on the original covariate scale then back-transformations are required.

Web Table 1: Estimates for the Multinomial Model

Table 1: Parameter estimates for the multinomial model for $\boldsymbol{n}_i|S_i, N_i$. Explanatory variables were chosen by the model selection for the modified Dirichlet-multinomial model.

Model Term[1]	Polynomial Order	Estimate	Standard Error
mean	-	0.817	0.678
temperature	1	-0.105	0.043
	2	0.005	0.003
	3	-7.902×10^{-5}	6.702×10^{-6}
scaled abundance	1	5.570	0.552
	2	-8.947	1.103
	3	5.914	0.903
scaled richness	1	-7.206	1.189
	2	6.628	6.212
salinity SD*	1	2.715	1.167
	2	-5.797	4.088
oxygen	1	0.082	0.024
scaled abundance : oxygen	1:1	-0.160	0.075

[1]All terms are coefficients of $\log(j)$. *Standard deviation.

Web Table 2: Estimates for Models of Classic Evenness Metrics

Table 2:

Model	Polynomial Term	Number of Parameters[1,2]	AIC[2]	Estimate[4]	Standard Error[3]
Simpson's D					
mean	-	2	-93.326	0.425	0.402
scaled richness	1	5	-139.696	62.455	10.835
	2			-619.866	124.239
	3			2288.877	501.625
scaled abundance	1	7	-159.935	-5.864	0.931
	2			4.439	1.178
length of tow	1	10	-210.6977	0.001	0.001
	2			-4.861×10^{-7}	4.362×10^{-7}
	3			1.020×10^{-10}	8.583×10^{-11}
scaled richness : length of tow	1 : 1	11	-217.279	0.014	0.005
Pielou's J					
mean	-	2	-511.909	0.843	0.037
scaled abundance	1	3	-544.576	-2.041	0.312
scaled richness	1	4	-590.035	4.762	1.125
scaled abundance	2	5	-624.837	3.726	1.021
	3			-2.426	0.928
oxygen SD*	1	6	-627.705	0.233	0.096
scaled richness	2	7	-627.909	-9.079	5.178
scaled richness : oxygen SD*	1 : 1	8	-628.000	-3.713	2.043
depth	1	10	-628.071	-1.13×10^{-4}	7.98×10^{-5}
	2			1.06×10^{-7}	5.97×10^{-8}

[1]Including residual variance. [2]From model including terms higher in the table. [3]From selected model. *Standard deviation.

References

Abramowitz, M., & Stegun, I. A. 1964. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. ninth Dover printing, tenth GPO printing edn. New York: Dover.