**EcoSimR Presentation**

EcoSimR is a Null Model Analysis package for Ecological Data. For co-occurrence null models, size ratio null models, and niche overlap. Today, focusing on co-occurrence null models.

* Null model: “ A pattern-generating model that is based on randomization of ecological data or random sampling from a known or imagined distribution [which] is designed with respect to some ecological or evolutionary process of interest”
* Gotelli & Graves 1996

**History/Emergence**

Diamond (1975) proposed an interpretation of species distribution based on interspecific competition which became a point of great contention in the following years:

1. Some combinations of species will never be found in nature because they represent “forbidden combinations” which cannot coexist due to interspecific interaction
2. Particular pairs of species may never co-occur (checkerboard pairs)

Connor & Simberloff (1979) challenged Diamond’s assembly rules by arguing that only by disproving a null hypothesis that species are distributed randomly would one be able to demonstrate that competition is responsible for species distribution. Introduced class of null models to ecological analysis.

The presence absence matrix is commonly used for ecological and biogeographical analysis.

* Rows represent species
* Columns are sites or samples
* Entries are the presence (1) or absence (0) of species in a site

|  |  |  |  |
| --- | --- | --- | --- |
| Species | Site A | Site B | Site C |
| Species 1 | 0 | 1 | 0 |
| Species 2 | 1 | 0 | 0 |
| Species 3 | 0 | 1 | 1 |

Null Model Analysis Using Presence-Absence Matrices (Gotelli 2000; following steps of classical statistical randomization tests, Eddington 1987)

1. **Form hypothesis -** Condense theoretical model predictions into simple hypothesis about the structure of a presence absence matrix
2. **Define index X-** Define index X that describes co-occurrence pattern in PA matrix
3. **Measure X obs -** Measure X obs for the PA matrix
4. **Randomize with null model and record X sim -** Randomize the PA matrix according to a null model and record X sim for the randomized matrix
5. **Repeat and generate frequency histogram of Xsim -** Repeat step 4 many times (usually 1000) to generate a frequency histogram of Xsim
6. **Compare Xobs to Xsim -** Interpret Xobs in comparison to Xsim using classical statistical inference
7. **Evaluate hypothesis -** Use results to evaluate original hypothesis

Co-occurrence Indices

Quantify the average degree of co-occurrence (segregated – random-aggregated) for all possible pairs of species

Count the number of checkerboard species pairs and the number of unique species combinations in the matrix

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **CHECKER** | **C Score** | **V ratio** | **COMBO** |
| Description | The number of species pairs forming perfect checkerboard distributions | Checkerboard score- the average pairwise species segregation – based on the number of checkerboard units | Variance ratio – the ratio of the variance of the column sums to the sum of the column variances | Number of unique species combinations |
| Calculation | Scan matrix rows |  |  |  |
|  |  | Higher the C Score, the less co-occurrence, on average, between all the species pairs.  Large C score – more segregated  Small C score – more aggregated | Close to 1 indicates random  Larger than 1 – aggregated  Smaller than 1-segregated   1. All sites have same number of species |  |
| Source |  |  | Schulter (1984) |  |

Null Model Algorithms – EcoSimR has 3 basic strategies to randomize a PA matrix based on marginal row and column totals of the matrix

* Equiprobable - Margin totals allowed to vary randomly and equiprobably
* Proportional – Probabilites proportional to margin totals in the original matrix
* Fixed-Fixed – marginal totals fixed and identical to the totals of the original matrix
* 3 constraints^2 dimensions = 9 simple null models

|  |  |  |  |
| --- | --- | --- | --- |
| **Constraint** | **Columns equiprobable** | **Columns proportional** | **Column sums fixed** |
| **Rows equiprobable** | SIM1 | SIM6 | SIM3 |
| **Rows proportional** | SIM7 | SIM8 | SIM5 |
| **Row sums fixed** | SIM2 | SIM4 | SIM9 |

Sim10: supply external weighting data for rows and columns

Gotelli (2000) found that SIM2, SIM4, and SIM9 have the lowest probability of Type 1 errors, because species occurrence frequencies are maintained. This occurs because species co-occurrence tests are very sensitive to variation in species occurrence frequencies, and therefore preserving the row totals as constraints in the null model is advisable.

SIM2 – appropriate for analyzing sample lists: lists of species collected with equal sampling effort

SIM9- appropriate for analyzing island lists of species – extensive lists collected with thorough sampling effort.