Checkerboards And Missing Species Combinations: Are Ecological Communities Assembled By Chance?

Nicholas J. Gotelli

24 January 2016

How are communities of plants and animals organized in nature? This is a challenging question, because there are too many species in most communities to study this problem experimentally.

For example, even with only 3 species (A,B, and C), there are 8 different "communities" or distinct species combinations that can be formed, based on the presence or absence of each species in the community: 0,A,B,C,AB,BC,AC, and ABC. Each combination represents a different possible community (including the 0 possibility of no species being present).With S species, there are 2S possible species combinations that can be formed, which is far too many for a wieldy experiment.

So ecologists have often taken a different tack. They use so-called "natural experiments" in which they let nature do the work and generate the replicates. Islands in an archipelago are often treated as natural replicates. The patterns of species association among a set of islands may provide important clues for how species interact.

The data for such an analysis can be represented as a binary matrix in which each row is a species, each column is an island (or site), and the matrix entries represent the presence (1) or absence (0) of a particular species on a particular island. For a matrix M with i = 1 to S species and j = 1 to R replicate islands, Mij indicates the presence or absence of species i on island j.

It is too much work to identify all of the different species of vertebrates, invertebrates, plant, and microbes on a set of islands, or even a single island. Instead, ecologists have often restricted their analysis to a single taxonomic group of organisms that have similar body sizes and similar life histories. Such collections (or "ecological guilds") of species may often be competing for shared food or habitat resources.

For example, here is a matrix illustrating the occurrence of the 27 species of finches (Subfamily Fringillidae) on the 25 major islands of the Greater and Lesser Antilles. Most of these species feed on seeds, which might be a shared, limiting resource that prevents all species from living together on the same island.

#### Figure 1 here

Figure 1. A binary presence-absence matrix. Each row represents a different species of island bird (finches in the Subfamily Fringillidae), and each column represents a different island. Ecologists have often used these kinds of data matrices to infer whether species interactions are present.

In 1975, the ecologist Jared Diamond (best known for his 199x Pulitzer-Prize-winning book *Guns, Germs, and Steel*), asserted that these kinds of matrices contained the statistical signature of strong species interactions, which he boldly outlined as a series of community "assembly rules".

In one of his assembly rules, Diamond proposed that some pairs of species were such strong competitors that they never occured together on the same island, forming a "checkerboard distribution". For example, from the matrix above, we notice that the species in rows 3 and 6 (x and y) form a checkerboard pair and never co-occur:

#### Figure 2 here

If we treat each island as a statistical replicate, these data can be analyzed as a 2 x 2 contingency table. Fisher's exact test can be used to ask whether the occurrence of species X is independent of that of species Y.

However, Diamond was not interested so much in the pattern for a single species in the assemblage. Rather, he argued that these checkerboard pairs could frequently be found in a community, and that they indicated competition for limited resources was important. For the West Indies finch matrix, there are 91 pairs of species that form perfect checkerboards.

Diamond's second assembly rule was based on the observation that some species combinations were never observed in nature. His inference was that these missing combinations were "forbidden" and again reflected the signature of interspecific competition. For the West Indies finch matrix, there are only x unique species combinations represented on the islands that are sampled.

But how many checkerboard pairs and how many species combinations would be expected if species interactions were *not* important? This is the null hypothesis that was not tested explicitly in Diamond's paper.

Is 91 checkerboards more than we would expect by chance? With 27 species, there are x unique pairs that can be formed, so perhaps finding 91/x pairs as checkerboards is not so extreme. Similarly, with 27 species, there are 227 possible species combinations that could be formed. But with a sample of only 25 islands, the maximum possible value is only 25, which would happen if every island had a unique mixture of species. In this matrix, there are only 23 combinations that are observed. Is 23 combinations on 25 islands sampled unusually small, given that x combinations are theoretically possible?

In a provocative response to Diamond (1975), Connor and Simberloff (1979) introduced a Monte Carlo analysis (usually referred to as a "null model analysis" in this literature) to try and answer that question.

Connor and Simberloff asked, What would this matrix look like if each species colonized the different islands independently and species interactions were not important? Answering that question precisely would require a detailed ecological model with species-specific parameters for colonization, extinction, and persistence. There are many such models in the theoretical ecology literature, but getting the data to estimate all of those parameters is another story.

Connor and Simberloff argued that a randomization or reshuffling of the existing data matrix might mimic the properties of a null or non-interactive community, without getting bogged down in the problem of estimating colonzation and extinction parameters.

Following standard Monte Carlo procedure, Connor and Simberloff randomized the observed presence-absence matrix, and calculated the number of checkerboard pairs and the number of species combinations that were found in this null matrix. Repeating this process 1000 times yields a distribution of the number of checkerboard pairs for this null hypothesis. Using a standard frequentist approach, we can then compare with the observed data and calculate how extreme or improbable the data are, given this null hypothesis as .

Exactly the same procedure could be used to evaluate the number of forbidden species combinations that are found in the matrix.

But exactly how should the observed data be reshuffled to mimic a colonization extinction model in which species occurrences on islands are independent of one another? The devil is in the details, and the development of different algorithms for randomizing a presence-absence matrix and quantifying the patterns has remained an active research frontier for over 35 years.

The simplest initial approach is a kind of "isotropic" null hypothesis in which we simply reshuffle with equal probability all of the matrix elements. This null model preserves the dimensions of the matrix (the number of rows and columsn) and it preserves, the "fill" of the matrix, that is the proportion of presences (0.xx for the West Indices finch matrix). But there are no other constraints on the occurrence of species on islands.

Here are the results for the 1000 randomizations of the West Indies finch matrix, with separate analyses for the number of checkerboard pairs and the number of species combinations. By this test, we would conclude that there are fewer checkerboard pairs than expected by chance, and also fewer observed combinations.

#### Figure 3 here

However, this randomization implies that all species have an equal probability of occurrence, and all sites have an equal chance of receiving species. Both assumptions seem unrealistic: some islands (especially large ones) are more likely to accumulate species than others, and some species are likely to be more common than others. These differences among species and among islands could occur even in the absence of any species interactions.

Two alternative algorithms incorporate more realistic assumptions. First, we could assume that different islands accumulate a random sample of species. The number of species on an island in each null assemblage is the same as in the original matrix. In otherwords, the marginal column sums of the matrix are maintained in each simulation. This model can be simulated by randomizing the elements within each column of the matrix.

Alternatively, we can think of the different specis as "sampling" the islands. In this model, the commonness and rarity of each species are preserved, the islands are treated as equiprobable, and the algorithm preserves the row totals of the original matrix for each randomization.

For these two models, the results are.....

Connor and Simberloff chose to impose both constraints, preserving simultaneously the row and column totals of the matrix to generated the expected number of checkerboards and observed species combinations, conditional on these marginal constraints. Figure x shows one of these simulated matrices, which is visually much more similar to the original matrix than random matrices produced by the other algorithm. When tested against this pattern....

Connor and Simberloff's original conclusion (based on a different data set and some different metrics) was that the evidence for Diamond's assembly rules was weak, and that observed patterns of species co-occurrence could be better explained by "chance" than by competition.