Beyond the hollow curve

* Species abundance distributions almost always follow a classic “hollow curve” pattern. There are a few abundant, and many rare, species, regardless of the system or taxon in question.
* Puzzlingly, most theories also predict hollow curves. This has made it difficult to use the species abundance distribution to evaluate competing theories.
* Recent work (Locey and White; Frank; Blonder?) has identified an underlying statistical constraint on the species abundance distribution.
* Locey and White found that most possible forms for the species abundance distribution are hollow curves. That is, in the absence of any strong force (or in the presence of numerous countervailing forces), we should in fact expect the species abundance distribution to follow some form of hollow curve based simply on the mathematical realities of partitioning N individuals across S categories.
* If this is the case, focusing on the “hollow curve” designation overlooks variation in the SAD that may hold the key to distinguishing between forms that are highly likely to emerge at random and those that are less likely. Even unlikely forms may be generally “hollow curves”; the important variation lies in the more nuanced aspects.
* Exposing variation in SADs in this regard may be the key to evaluating theoretical predictions for SADs and/or evaluating the possibility that, in aggregate, ecological communities behave essentially randomly.
* Previously, exploring the full range of SADS within the feasible set and comparing them to observed distributions has been prohibitively computationally intensive. Here, we use a new algorithm for rapid unbiased sampling of the feasible set space, and a compilation of abundance datasets for X communities, to (1) intensively sample the range of possible forms for a vast collection of SADs and (2) evaluate whether ecological communities consistently deviate from their statistical baselines or simply track their most-likely forms.

Phenomenological vs stat mechanics framing...

* Phenomenological:
  + SADs are almost always hollow curves. This makes them hard to tell apart.
  + We can sample the feasible set to map the range of variation in possible forms, then compare a distribution of interest to this corpus
  + This allows us to differentiate between distributions that simply track their baseline and those that are unusual
* Statistical mechanics:
  + Either because there is no force operating on a system, or because numerous forces cancel each other out, it might appear random
  + Distinguishing between highly-likely-at-random and unlikely observations may help us identify the forces operating in ecological communities that shape drive SADs away from a random baseline.

Methods

* Sampling the feasible set
  + Define the problem
    - The feasible set is defined as all possible unique unordered vectors of *S* integers that sum to *N*.
    - This is a form of constrained integer partitioning, which is a challenging problem that the number theorists are deeply into but have not solved concisely.
    - As S and N increase, the number of possible forms expands rapidly. Rather than enumerate all possible forms, we sample the distribution. It is therefore crucial that we sample the distribution at random.
  + Explain algorithm
    - Gnomes:
      * Allocate 1 individual to all species.
      * Sequentially redistribute remaining individuals and species according to the possible ways.
      * (Hao wrote an actual robust version of this with 0 gnomes).
* Describing variation in distributions
  + We use skewness and Simpson’s diversity to summarize SADs.
  + Illustrate how Simpson’s and skewness map on to a FS
* Comparison to observed SADs
  + The raw value of skewness or evenness for a community is not informative, because the range of probable values varies depending on S and N. Therefore, we focus on where an observed distribution’s value falls relative to the distribution of values of the samples from its feasible set.
* Compiling SADs
  + We use the same compilation of community abundance data as several previous studies of the SAD (Baldridge, Locey and White, Xiao).
    - Gentry
    - FIA plots
    - BBS – 2009
    - Mammal community database
    - Miscellaneous abundance database
  + Gentry, FIA, and BBS all include a single year of data. MCDB and Misc Abund contain varying temporal scales.

Results

* SADs are generally hollow curves (heatmaps)
* That said, empirical SADs are consistently more skewed and less even than most of their feasible sets. Depending on the dataset.
* This does not appear to map on to any obvious axes of S, N, S:N.

Discussion

* Although many empirical SADs do appear to track their statistical constraints, in all datasets, highly skewed and uneven vectors are overrepresented.
* Understanding the correlates and causes of the variation in the unlikeliness of the SAD will be a major new horizon.
* This method reveals a new and promising way to extract whatever information there is to be found in the SAD and leverage that to understand what makes some ecosystems deviate from a random baseline.

Over decades of ecologists’ theorizing, the species abundance distribution has proven a tempting but surprisingly slippery source of information about the processes structuring ecological communities. Almost all communities have a large proportion of rare species, leading to classic “hollow curve” SADs. Generations of ecologists have attempted to use this remarkably general pattern to evaluate theory and make inference about communities’ state and structuring forces. These efforts have been frustrated, at least in part, by the same generality that makes the SAD so intriguing: most theories also predict hollow curves, and the range of variation in the form of empirical SADs is small.

Recent work by Locey and White (2013) offers an explanation for, and potential way out of, this predicament. They demonstrated that hollow curves dominate the range of *possible* forms, or “feasible set”, for the SAD. Simply due to the mathematical constraints imposed by partitioning a particular number of individuals into a particular number of species, the most-likely form for the SAD is a hollow curve. The hollow curve itself is not necessarily a symptom of any additional process, and efforts to reproduce or explain the hollow curve may not yield compelling insights into *ecology*.

However, we may be able to use the feasible set, or range of possible forms, for the SAD as a baseline against which to evaluate real SADs. If an empirical SAD is unusually even or skewed compared to its feasible set, this might indicate that biological forces have driven that community away from its most-likely, random, state. If such deviations are consistent across communities, we may be able to use the *deviations* as the biological signature around which to develop and test theory. On the other hand, if empirical SADs are unremarkable compared to their feasible sets, we may not be able to detect a biological signal in the SAD itself.

Here, we compare Xnumber of empirical SADs to their feasible sets. Consistent with Locey and White, we find that most feasible sets are overwhelmingly dominated by hollow curves. Within this general “hollow curve” form, we document appreciable variation in shape as captured by both skewness and Simpson’s evenness. We show that empirical SADs vary in their position relative to their feasible sets, and that exceptionally skewed and uneven SADs are disproportionately common in ecological communities. We suggest refocusing research on the SAD to identify the correlates and develop explanations for these common, but not ubiquitous, deviations.