**Introduction**

The species abundance distribution (SAD) has been a major focus of both theoretical and empirical traditions in community ecology, largely because it is one of the very few ecological phenomena to conform to a consistent pattern approaching a law (Lawton, McGill, …). Regardless of the biological specifics of an assemblage, it will almost certainly be dominated by a few very abundant species, with relatively many rare species rounding out the abundance distribution (). Tempted by this remarkably general pattern, community ecologists have developed an array of theories to explain the underlying biological processes, as well as methods to divine the internal state of a system based on subtle variations in the SAD (). Ironically, this same generality of form has confounded our efforts to extract biological information from the SAD. There is precious little variation in form among empirical SADs, even in systems subject to experimental manipulations (Supp, ...). Most theories produce similar predictions and perform comparably well when compared to observed distributions, making it difficult to use the SAD to evaluate competing theories (McGill).

Recently, multiple lines of reasoning suggest an explanation and a remedy for this puzzle. The SAD may be subject to a statistical constraint that encourages it to follow a hollow curve, even in the absence of any biological mechanism, and warps our perception of the important aspects of the pattern. Frank () used a statistical mechanics framework to show that a log-series – one of the most common mathematical approximations for empirical SADs – emerges as a statistical by-product of [RMD forgets the precise process and needs to revisit Frank’s stuff]. Harte and associates () used the principle of entropy maximization to find the most-likely form for numerous ecological distributions, given sparse information about broad community properties and minimal assumptions regarding ecological mechanism. They also found that a log-series emerges as the most-likely form for the SAD (). This log-series is an excellent fit to empirical SADs (White et al 2012). Finally, Locey and White (2012) put forward an entirely non-biological derivation for the statistical constraint on the SAD. They leveraged the fact that there are a finite number of ways to divide *N* individuals among *S* species, and used combinatorics to characterize the range of *possible* forms for SADs, or feasible set, for realistic values of S and N. They showed that practically all feasible sets have strong central tendencies towards hollow curves ().

If, as seems apparent, SADs are statistically driven to approximate hollow curves before biology even enters the picture, it is no surprise that we have struggled to interpret the hollow curve in biological terms. But there is hope for a way forward: if we recognize and account for the statistical constraint, we may be able to isolate subtler but more biologically meaningful variation in SADs. Specifically, we can describe SADs not in terms of their absolute shape, but in terms of their shape *relative to what we would expect simply due to statistical constraint* (). If empirical SADs consistently deviate from their distribution of likely forms, these deviations may signal biological processes operating over and above the statistical constraint (). If they instead tend to track their statistically-most-likely forms, there may be little information about biological process to extract from the SAD alone – either because no processes consistently regulate the SAD, or because so many processes operate in various directions that the emergent distribution is indistinguishable from randomness ()

The first step is to characterize how empirical SADs generally compare to their statistical baselines and quantify any consistent deviations. Locey and White ()’s initial efforts along these lines showed that empirical SADs are often more skewed and less even than the central tendencies of their feasible sets. Here, we extend their approach to more specifically quantify the extent and ways in which empirical SADs deviate from (or conform to) their feasible sets. We test whether these deviations correspond to statistical properties, such as overall S and N and the range of variation present in the feasible set, that may further affect detectable patterns. We also test whether deviations vary between communities of different taxonomic [and other?] groups.

Although new conceptual frameworks for interpreting deviations will undoubtedly develop as this approach matures, we can already test one such framework. Borrowing reasoning from statistical physics, Newman and Harte () proposed that systems at “steady state” should more closely resemble random expectations than systems that have recently undergone manipulation or disturbance. At steady state, numerous countervailing processes are in equilibrium, and may cancel each other out at the scale of the SAD. Disturbances may temporarily overwhelm these processes and drive the system into a relatively nonrandom state. Although Supp et al () found that manipulation affected the shape of the SAD only when S and N also changed, their approach focused only on the *overall* shape of the SAD and not changes in the shape of the SAD relative to random expectation. We re-analyze the manipulated animal community data from Supp () to evaluate whether, in keeping with Newman and Harte’s proposition, manipulations induce increased deviations between observed SADs and their statistical constraints.

**Results**

1. Datasets x S and N space
   1. The datasets occupy mostly different, but somewhat overlapping, S and N space.
2. Effects of S and N on the characteristics of the feasible set
   1. The size of the feasible set increases dramatically with S, N, and N/S.
   2. Mean, median, standard deviation, and range of both skew and evenness vary systematically with S, N, and N/S, and consequentially each other and the size of the feasible set.
   3. This variation can appear counterintuitive when we make comparisons over large ranges of S and N. For example, a very large community with one superabundant species (say, one species with 10000 individuals and a thousand singletons) can score as more even than a less species-rich community with more evenly distributed abundances.
   4. The variation is much more intuitive *within a feasible set*; that is, when comparing vectors all with the same S and N. Therefore we advocate focusing on the rank score, and not the raw metrics, for these purposes.
3. Sampling
   1. For small communities, we can exhaustively sample the feasible set; for most, even 4000 samples is a miniscule (think exp(-20)) slice of the number of elements.
   2. The samples appear to be uniformly distributed: increasing the number of samples increases the resolution but not the shape of summary distributions.
4. Nonsingletons percentile scores
   1. There is considerable variation in percentile score.
      1. Some, but by no means all, of this corresponds to gradients in the state variables and feasible set characteristics. Because the datasets occupy different regions of S and N space, we cannot confidently differentiate between variation driven by S and N and variation driven by biological differences between the datasets.
   2. Disproportionately often, real SADs are very skewed and uneven compared to their feasible sets. This holds across almost all S and N, with a few pronounced exceptions.
      1. Communities with small feasible sets, which we were able to exhaustively sample, have *less* remarkable SADs than communities with larger feasible sets.
      2. For large communities where N approaches S, we see unusually *even* – as opposed to uneven – observed SADs.
5. Sensitivity to singletons
   1. Supplementing with presumed missed species almost always increases deviation, but rarely with a large magnitude. So if anything, we’re underestimating weirdness.
   2. The number or proportion of species added (relative to the original S), or the starting percentile value, do not predict the shift in percentile value from adding singletons.
6. Manipulations
   1. Manipulated communities do not differ in any consistent direction from control communities in their deviations.

**Results II**

Effects of S and N on the feasible set

Consistent with Locey and White (2012), the state variables S and N have a profound effect on the number and characteristics of forms represented in the feasible set.

The size of the feasible set itself increases dramatically with both S and N, and with the ratio of N:S. Communities with very low S or N, or a low ratio of N:S, necessarily have relatively few unique elements in the feasible set. This can be problematic for comparing observed distributions to the feasible set. In the limit of only a few possible forms for the SAD, we cannot say that one general form dominates and there is little to be learned from confirming that the observed distribution is one of these few.

The mean, standard deviation, and range of both evenness and skewness of samples from the feasible set vary systematically over the range of S and N represented in our datasets. While the most unusual feasible sets are for values of S and N that also generate relatively few (<2000) samples, there is still appreciable variation in large feasible sets. In particular, communities with a relatively low N:S ratio – even if the actual values for both S and N are large – have the most unusual feasible sets. These feasible sets have high values and low variability for evenness, and high values and *high* variability for skewness, compared to the other well-sampled S and N combinations.

The datasets differ markedly in their values of S and N. All of the low-N/S communities mentioned above are from the Gentry dataset. However, except for this extreme subset, the datasets do not differ systematically in the characteristics of their feasible sets.

Observed vs. sampled SADs

Across all communities, the percentile ranks for observed values of both skewness and evenness are disproportionately concentrated at the extremes. Observed communities rank very high in skewness, and very low in evenness, dramatically more than would be expected at random. At random, we would expect 1% of samples to fall into each percentile bin. For all percentiles >75, the % of skewness samples exceeds 1%, and for all percentiles <20, the % of evenness samples exceeds 1%. Values in the 100th and 0th percentile for skewness and evenness (i.e. entirely outside the values represented in samples from the feasible set) were 5 and 10 times more prevalent than random, respectively. However, a considerable proportion of observed SADs fall well within the distribution of skewness and evenness values represented by their feasible sets. Although it occurs much more often than it should by chance, it is by no means a rule that observed SADs deviate strongly from their feasible sets.

These qualitative results hold when broken out by each dataset. The BBS and FIA datasets have the *lowest* proportion of highly skewed and uneven observed SADs. The Gentry dataset deviates from the general pattern in that it has a slight overrepresentation of highly *even* and *non-skewed* observed SADs. These correspond almost perfectly to communities with an exceptionally low N:S ratio, and whose feasible sets deviate the most from the rest of S and N space (see above). Specifically, these feasible sets are more even than most; these observed SADs are therefore unusually even elements from unusually even, and generally exceptional, pools.

Except at extreme values of N/S, percentile rank does not vary systematically with N/S. For very low N/S, skewness tends to be low and evenness high. This is entirely driven by the Gentry sites discussed above. At very high N/S, there are virtually no intermediate percentile ranks. These feasible set are the largest, and it is not surprising that 2500 samples does not capture the full range of variation, and the extreme values signal the edge of our ability to measure variation in the deviation. Other feasible set characteristics (mean, sd, and range of skewness and evenness) do not have strong relationships with percentile rank.

Effects of rarefaction

SADs with species supplemented via rarefaction tended to have, if anything, *more* extreme values of skewness and evenness than the raw SADs. This is true for all datasets, with somewhat more noise in the BBS and FIA datasets than the others. These datasets have the most samples, and the highest proportion of non-extreme percentile values. There is no obvious relationship between the percentile rank of the raw SAD and the change induced due to rarefaction, nor between the number or proportion of species added via rarefaction and the resulting change.

Effects of manipulation

RMD is still vetting the MACD data wrangling. It **looked like** no consistent effect of treatment, but I want to be sure I’m joining the treatment-control sites correctly.

There is no significant relationship between season or treatment and percentile rank for the Portal plants.

**Discussion takeaways**

There are difficult-to-unravel constraints on the available forms of the feasible set for some S and N. This is a cautionary signal that we should really only compare SADs with comparable S & N **or** compare them strictly in terms of how they compare to their respective FS; the baseline shifts. Except in the extremes, the percentile rank appears to vary unrelated to characteristics of the feasible set.

Observed SADs often differ markedly from their expected forms given the feasible set. Specifically, they tend to be dramatically less even and more skewed than their feasible sets. We should look for ways to predict and understand what drives some communities to deviate so strongly. However, we should also recognize that the hollow curve pattern is often a red herring once put in context of the feasible set.

One route forward may be to trace deviation over time within communities. Consistent or variable? Precipitating factors?

The amount of variation in both characteristics of the feasible set and the position of observed distributions relative to the feasible set means we should be able to use the deviation to compare predictive performance of different theories.

We did not detect an effect of disturbance on the shape of the SAD relative to its feasible set. This is inconsistent with the idea put forth by Newman and Harte, that disturbance should amplify deviations between the actual state of a system and its statistical baseline. We may need a larger database, or a more systematic taxonomy of the types and impacts of disturbance to detect such a signal. Or, these deviations may be unrelated to disturbance history.

If the species abundance distribution is any indication, statistical constraints may have an underappreciated impact on other community abundance patterns. Characterizing those constraints for other macroecological distributions may present more of a challenge, but is a critical piece of context for identifying and interpreting biological signal within subtle and potentially deceptive statistical noise.