**Intro (to results/discussion section)**

We detect appreciable differences between empirical SADs and their feasible sets, with a range of variation that may be a promising source of new information and statistical leverage. We observe these deviations more consistently for large communities than for smaller ones – in particular, the very small communities that comprise most of the FIA dataset - and we suggest that the range of variability of the feasible sets of small communities limits our ability to detect deviations. Adjusting for rarefaction increases the strength of this signal.

**Empirical SADs compared to their baselines**

For four of the five datasets we analyzed – BBS, Gentry, MCDB, and MIsc. Abund – empirical SADs are highly skewed and highly uneven relative to their feasible sets much more frequently than would be expected by chance. Combined across these four datasets, 16% of observed SADs are more skewed than 95% of their feasible sets, and 31% are less even than 95% of their feasible sets. By chance we would expect only 5% of observed distributions to fall in these extremes. Adjusting for rarefaction increases the strength of this signal: for these datasets, 18% of adjusted SADs are more skewed than 95% of their feasible sets, and 38% are less even.

This may be evidence that the shape of the SAD is not entirely a statistical artefact – that there are indeed biological processes that generate a particular, highly uneven, form for many empirical SADs. So far, we have struggled to understand or even identify these processes because we have been looking at the wrong *aspects* of this form. We can use the deviations between empirical SADs and their feasible sets as a new source of leverage for fitting models and evaluating theories. A model that deviates from the feasible set *consistent with observed distributions* will be much more convincing than one that simply predicts the central tendency of the feasible set, even though both models will predict plausible-seeming hollow curves. A logical starting point will be to test predictions from established theories (e.g. neutral theory, METE, ???) and common functional approximations (logseries, exponential, and log normal) for the SAD to evaluate which ones make accurate predictions regarding *deviations*.

While there is an overall signal of high unevenness in these four datasets, there is also considerable heterogeneity in how empirical SADs compare to their feasible sets. Some this variation may be statistical, driven by special cases regarding S and N. For example, certain Gentry communities have very low average abundances, which forces all elements of the feasible set to be fairly even; coincidentally, these are the only group of communities for which empirical SADs are unusually *even* relative to their feasible sets. However, we also see considerable variation between communities with similar values of S and N. More focused comparisons between communities may show whether there are identifiable differences between these communities that systematically cause some to deviate and some not – differences to which we would not otherwise have been attuned.

We detect considerably less pronounced deviations for communities from the FIA communities. For FIA, percentile scores are near uniformly-distributed for skewness (5% of observations are more skewed than 95% of the feasible set), and much noisier than any of the other datasets for evenness (11.5% of observations are less even). We suspect that this is primarily an effect of small community size: the FIA communities are the smallest in our database. There is not an obvious difference between FIA communities and comparably-sized communities from other datasets in the distribution of percentile values – although note that the vast majority of communities in this size range are FIA.

**Small-community effects**

Small community size may affect our ability to distinguish between deviations and randomness via its effect on the variability of forms represented in the feasible set. We found that the feasible sets for small communities have broader distributions of evenness and especially skewness than those for large communities. For communities of the sizes represented in the FIA dataset, the 95% interval of skewness values often encompasses more than 80% of the entire range of values; for larger communities, the 95% interval spans closer to 60% of the full range. This is consistent with concepts from statistical mechanics. Large communities have many components that can be arranged in many ways, and most of these arrangements cluster around a relatively specific highly-likely state. If an observation differs even a small amount from this most-likely state, it is readily detectable as highly unlikely to have occurred by chance. In contrast, small communities have relatively few possible arrangements and relatively broad distributions of likely shapes. Observations may deviate from the most-likely form, but only the most extreme deviations will be highly *unlikely* given the breadth of the corresponding probability distribution.

These small-community considerations appear to be relevant for ranges of S and N that are quite common in ecology. The FIA communities range in size from x to y species and x to y individuals. These are by no means hard thresholds, but they do indicate a general range of values below which we have relatively little power to distinguish deviation from randomness. Unless we can develop more sensitive methods for identifying deviations even in these small communities, we stand to learn the most by focusing on SADs from relatively large communities. In the meantime, sampling the range of forms represented in the feasible set provides context to identify when the breadth of variation in the feasible set is so great it diminishes our power to detect deviations.

**Defining the statistical baseline**

There are multiple plausible approaches to defining the statistical baseline for the SAD, of which we have taken only one. Specifically, we have taken the feasible set defined as the set of possible *unique, unordered* divisions of N individuals into S species. Other assumptions – regarding the distinguishability of individuals and species, for example – may be equally valid and may generate substantively different statistical baselines. In the absence of a firm logical argument for one set of assumptions over another, we suggest performing analyses similar to the one presented here using alternative assumptions. This is a common problem, or a fatal problem?????? NOT SURE.