1. Background
   1. SADs have a consistent form, which has tempted us to look for meaning
   2. Theories all predict hollow curves, and this makes meaning slippery
   3. A strong possible explanation = idea of a statistical baseline
      1. Our intuition of what a surprising distribution looks like needs to take into account what it would look like at random. In the case of the SAD, Locey and White () *[and White et al 2012?]* demonstrated that the feasible set of possible forms of the SAD is overwhelmingly dominated by hollow curves, meaning it is not at all surprising that the majority of both observed and theoretically-derived SADs conform to this qualitative pattern.
      2. The idea pervades macroecology *(but is there a more specific formulation of it?)* that no or many countervailing processes might leave systems appearing random, but that strong process might systematically drive systems to deviate from random
   4. Given this logic
      1. We need to use the statistical baseline to properly contextualize observed SADs.
      2. We need to quantify, and trace drivers of variation in, the deviation between observed SADs and randomness
      3. We should explore whether disturbances (introduced via experimental manipulation) induce increases in these deviations
   5. Here we
      1. Compare the shape of X observed SADs to extensive samples from their respective feasible sets
      2. Establish that observed SADs are far from representative of their feasible set, but that the degree of deviation varies considerably from community to community
      3. Test whether experimental manipulation induces a predictable shift in the position of a community’s SAD relative to its statistical baseline
2. Methods
   1. Data
      1. To look for generalities in the position of observed SADs relative to randomness, we revisit two large databases of community abundance data used in past work on the SAD: White et al (2012) and additional datasets found in Baldridge et al (2014). These include the Gentry and FIA forest plots, a subset of the Breeding Bird Survey, the Mammal Community Database, and the Miscellaneous Abundance Database.
      2. To test whether disturbance affects the SAD relative to its constraint, we revisit two databases previously used to address this question without taking into account the statistical constraint. We compare control and manipulated communities in the winter and summer plant communities from the Portal Project, and control and treatment communities from the multi-taxa Manipulated Animal Community Database.
      3. Communities with very low richness or abundance have extremely small feasible sets. We therefore exclude communities with fewer than 2 species, or an average abundance of 1.
      4. It is computationally intensive (although technically possible) to sample the feasible set for communities with very high richness and abundance. We therefore exclude communities with more than 25k (?) individuals or more than 8k species? *[This is not a simple limit – what matters is the combined S \* N – so may need to revisit this language to be precise]*
      5. Incomplete sampling is always a concern with empirical abundance data. To account for the possible impact of unobserved species on our results, we conducted all analyses on both the raw data and on communities supplemented via rarefaction. For every SAD, we estimated the true number of species present. We assumed that unobserved species are likely to be rare, and adjusted for their presumed omission by adding one individual each for each additional species estimated to be present.
   2. Sampling the feasible set
      1. [algorithm] implemented in diazrenata/feasiblesads
   3. Comparing observed SADs to the feasible set
      1. We calculated two summary statistics to describe the shape of abundance distributions, skewness and Simpson’s evenness.
      2. For each focal SAD, we drew 2500 samples from the feasible set of an abundance distribution with the appropriate S and N.
      3. Some combinations of S and N have a feasible set that is much smaller than 2500. It may not be appropriate to draw conclusions based on such a small feasible space. If 2500 draws from the feasible set did not yield 2000 or more unique distributions, we did not include that community in subsequent analyses.
      4. We constructed the distribution of skewness and evenness values of the sampled SADs, and calculated the percentile rank of the focal SAD’s statistics relative to these distributions. This allows us to compare the *position* of SADs from communities with different S and N and therefore different feasible sets.
      5. If SADs simply track the feasible set, we would expect observed percentile values to be uniformly distributed in aggregate.
   4. Correlates of percentile position
      1. Beta regression to test whether average abundance, mean/sd of FS statistic predict percentile position?
      2. For now, try just the R2 of a lm()
   5. Effect of disturbance on percentile position
      1. Just a t-test, treatment v control? Portal broken out by season.
3. Results
   1. Overall position of observed SADs relative to FS
   2. Effect of rarefaction on qualitative results (more extreme)
   3. Correlates of observational SAD positions
   4. Effects of manipulation on SAD position
4. Discussion

[ QUESTIONS ]

* Do we have to do this all as a maximal model or OK to break out by dataset
* How much of this do we really have to test statistically vs just show the plots?