**Figure legends**

Figure 1. Distribution of communities from each dataset in terms of total abundance (N) and species richness (S). Communities range from few species and individuals (lower left corner) to speciose communities with many individuals (upper right). However, datasets are not evenly distributed across this state space due to differences in their sampling intensity, design, and underlying biology (e.g. productivity, regional richness, taxonomic group, or other factors that influence the capacity of a community to support large numbers of species and individuals). In particular, note that the FIA dataset comprises very small communities, and communities from the Gentry dataset are extreme in both their high species richness and low average abundance.

Figure 2. Large feasible sets may allow better detection of deviations from the statistical baseline by generating more specific, narrowly-defined baselines. We illustrate this phenomenon using 3 hypothetical communities: a small community (*S* = 4, *N* = 34; top row), an intermediate community (*S* = 13, *N* = 315; middle row), and a large community (*S*= 44, *N* = 13360; bottom row). The large communiity has approximately 6.59e+70 possible SADs in its feasible set, while the intermediate community has 1.001e+12 and the small community has only 297. For every SAD sampled from the feasible set (left column), we calculate the skewness (color scale) or other summary metrics (not shown). The distributions of these values (right column) constitute the statistical baseline. We define a “breadth index” as the ratio of the range encompassed in the two-tailed 95% density interval (distance between red lines, right), compared to the full range of values for the statistic (distance between the maximum and minimum values). As *S* and *N* increase, the size of the feasible set increases, resulting in a narrower statistical baseline (smaller breadth index) – thus enabling easier detection of deviations that may be the result of ecological processes affecting the SAD.

Figure 3. Many ecological communities are highly unusual compared to their statistical baselines. Percentile ranks are calculated by comparing each community to its sampled feasible set, with very high or very low percentile ranks reflecting extreme values relative to statistical baselines. The vertical red lines mark the 95th percentile for the dissimilarity to the central tendency, and the 2.5th and 97.5th percentiles for all other metrics. Species abundance distributions that are sampled at random from the feasible set will produce percentile ranks that are roughly uniformly distributed from 0 to 100, with approximately 5% of values above the 95th percentile or 2.5% of values above and below the 2.5th and 97.5th percentiles, respectively. In contrast, most datasets have more communities that are highly skewed or uneven than would be expected by chance. The percentile values shown are the mean of the percentile scores defined as the proportion of comparison values <=, and <, the focal value. In calculating the proportion of sites with extreme values, the <= designation gives an appropriately conservative estimate of the proportion of high values, but overestimates the proportion of very low values, and the reverse occurs for the < designation.

Figure 4. The variability of the feasible set, defined as either the mean dissimilarity of elements of the feasible set to the central tendency of the feasible set, or via a breadth index (see Figure 1), decreases as the number of possible SADs in the feasible set becomes very large. Highly variable feasible sets constitute broad, poorly-defined statistical baselines that may impede our ability to confidently detect deviations between observations and what is expected given the baseline. Small feasible sets, which occur for small combinations of S and N, are often highly variable. The majority of these small, highly variable feasible sets occur for communities in the FIA and Mammal Community databases. Although the Gentry dataset also contains communities with small feasible sets, these communities also have a very low ratio of N to S, meaning their entire feasible sets may be constrained to be more self-similar than small feasible sets in general (see Dissimilarity to central tendency). There is, however, substantial additional variation in the dissimilarity and breadth indices not accounted for by the size of the feasible set or the ratio of N to S.

Figure 5. Summaries of how resampling to adjust for under-detection of rare species (green) and subsampling (blue) change the proportion of extreme values observed for each metric and dataset. The horizontal black lines mark the approximate proportions of extreme values that would be expected at random: 5% for dissimilarity to the central tendency, and 2.5% for all other metrics. Adjusting for rare species consistently increases the proportion of extreme values relative to the raw SADs, while subsampling often decreases it but does not eliminate or change the direction of the effect. The exception is for the FIA dataset, which does not show strong deviations for either raw or resampled SADs. Shown are the effects and directions observed for most datasets; for complete results of resampling, including the opposite direction effects, see A7.

Figure 6. Very small communities (e.g. those with fewer than 2000 possible SADs in the feasible set; upper rows) exhibit more variable, broadly-defined statistical baselines (top) and less consistently extreme observed values relative to their feasible sets (bottom). 2000 possible SADs is used as a cutoff because it allows for a comparison using a substantial number of communities from the FIA and two other datasets. Of these datasets, the FIA is the most dominated by very small communities (68% of FIA sites have fewer than 2000 possible SADs, compared to 34% for the Mammal Community and 7% for the Miscellaneous Abundance databases). Results shown are for skewness; for complete results see Appendix A10.