**Figure legends**

Figure 1. 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* = 7, *N* = 71; 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 60,289 and the small community has only 297. For every SAD sampled from the feasible set (left column), we calculate the skewness (color scale) and evenness (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 one-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 2. Many ecological communities are more skewed (left) or uneven (right) than 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 line marks the 95th percentile for skewness and the 5th percentile for evenness. Species abundance distributions that are sampled at random from the feasible set will produce percentile ranks that are uniformly distributed from 0 to 100, with approximately 5% of values above or below the 95th and 5th percentiles, respectively. In contrast, most datasets have more communities that are highly skewed or uneven than would be expected by chance.

Figure 3. Breadth indices of skewness (left) and evenness (right) indicate varying ability to detect the deviations between observations and the statistical baseline. The breadth index (see Figure 1) quantifies how narrowly-defined the statistical baseline is; high values indicate broad, poorly-defined statistical baselines that may impede our ability to confidently detect deviations between observations and what is expected given the baseline. Most datasets contain a mixture of communities with broad and narrow statistical baselines, but some – particularly the skewness baseline for the Forest Inventory and Analysis – have consistently high breadth indices across all of their communities, suggesting that skewness is not an effective metric for distinguishing empirical observations from the feasible set. In general, the breadth index for evenness (right panels) indicates more narrow statistical baselines than those for skewness.

Figure 4. Small communities exhibit consistently broad statistical baselines (top), and consistently weak evidence of deviations for observed SADs (bottom), regardless of the originating dataset. For a subset of 371 communities from the Forest Inventory and Analysis with communities in other datasets with matching *S* and *N*, we generate distributions of breadth indices for skewness and evenness (top) and compute corresponding percentile ranks for the observed SADs (bottom). Visually, there is no difference between FIA (left panels) and other datasets (right panels), when they are matched in *S* and *N*. This is confirmed by Kolmogorov-Smirnov tests for the breadth indices (for evenness, *D* = 0.04 and *p* = 0.91; for skewness, *D* = 0.03 and *p* > 0.99) and percentile ranks (for evenness, *D* = 0.04 and *p* = 0.87; for skewness, *D* = 0.07 and *p* = 0.37).