Finding happiness in your SAD

Intro

- SAD is fundamental to ecological theory
- As such we should have robust methods to test SAD predictions
- Despite decades of work on SAD, there is little agreement about best practices
- Here we lay out guidance to be happy with your SAD
  - SAD models are mathematically related to one another and recognizing that can help us identify what we're really testing
  - Common visualization methods are mathematical transformations of one another, but some are better than others
    - use CDF or Rank
    - Binning data does not help identify nor illustrate the shape of the SAD; this has been said many times before, can we all finally agree to stop?
  - Sampling is a real consideration but the veil line is overly simplistic
    - sub-sampling largely preservers parametric forms of SADs
    - sub-sampling produces predictable changes in parameter values of SADs
    - small sample sizes make it hard to discriminate between models
  - AIC is great for model comparison (widely agreed on) but is not great for goodness of fit
  - We explore a range of goodness of fit metrics and decide on the following guidelines
    - hypothesis tests to reject a model based on exact tests are good
    - sample size can have a strong influence on goodness of fit metrics so to compare goodness of fit across samples of different size care is needed; we propose XYZ method for this use case
- We apply these best practices to several iconic SAD datasets, uncovering several new finds
  - something about which models fit best (not new, but sets the stage)
  - something about whether those best fitting models are actually well-fit (new)
  - something about sufficiency of data to discriminate amoung models (new)
  - something about scaling of parameters across space or something (new)

Mathematical relationships between SAD forms

- We first use known results from the applied mathematics literature to draw connections between parametric forms of SAD models
- *Maybe: We then use KL divergences to understand how identifiable two major classes of SAD models are*
  - *what is KL divergence*
  - *how does sample size come into play*
- **Figure** map of SADs and (maybe, maybe not at all: map of plognorm and tnegbiom parameter spaces)

Visualization practices

- We use known results from probability theory to re-iterate relationships between SAD plots
- We use known results from applied math to argue for the use of certain plots over others
- We use simulation to demonstrate a few pitfalls of binning
- **Figure** map of plot types
- **Figure** binning and modes

Exploring the effects of sampling on SAD inference

- Defining SAD sampling: it's over individuals not species
- Simulation experiment
  - *Supp Figure* how does sample size influence parameter estimates
  - **Figure** how does sample size influence model competition

Developing the best goodness of fit metrics

- The metrics to be considered and why
  - classic metrics (e.g. KS)
  - MSE on RAD
  - "exact test" metric
  - order statistic metric
- Simulation set-up
  - simulate under known model
  - fit that model or a different model
  - calculate goodness of fit
  - do under different sample sizes
- *Supp Figure* scaling of goodness of fit metrics and sample size
- **Figure** ability of goodness of fit metrics to accurately capture divergence from known SAD models

Studying SADs in the wild

- **Figure** spatial structure of SAD in BBS (or something)
- **Figure** data sufficiency

Discussion

- Summarize key take-homes
- address areas for continued improvement (both in data generation and analysis)
- Make a final point about the continued relevance of SADs