How to be happy when your data are SAD

Andrew J. Rominger, John Harte

April 26, 2016

1 Introduction

1. SAD long standing in ecology
2. Any useful thoery of biodiversity, in addition to whatever other patterns it may predict, should make meaningful and accurate predictions of the SAD.
3. Here I
   (a) briefly summarize the most relevant SAD theories, laying to rest several irrelevant theories
   (b) present a set of best statistical practices for SAD testing, clarifying misconceptions such as the veil-line
   (c) explore the future of SAD theories, including the exciting universality and utility of the negative binomial distribution

1.1 A SAD history

1. Review of theories predicting SADs:
   (a) broken stick
   (b) negative binomial
   (c) log series
   (d) log normal
   (e) zipf
   (f) neutral theory
   (g) mechanismless theory
2. mathematical connections between theories

1.2 Bad SAD theories

1. The log-normal is not valid and not useful
2. The Gambin model is also not valid nor well supported

1.3 SAD theories for the future

Return to Fisher's formulation and Preston's ergodic hypothesis

2 SAD subsamples and mixtures

1. SADs are sampled by indiviudal not by species.
2. The veil line arrises by inappropriately sampling by species.
3. SADs have a discrete spatio-temporal extent. This is arbitrary but important: The log normal shape arrises by mixing samples (i.e. multiple years of moth trapping, combining distinct spatial samples)
4. How SADs scale is another issue and an interesting topic on inquery. The fact that there is scale depenence in SADs doesn't negate the fact that sampling them must be viewed as discrete in space, time and taxonomy.

3 Best and worst statistical practices for comparing theoretical and emperical SADs

Do not bin data. Deal with subsampling appropriately

3.1 Comparing theory to data

1. connection between PMF, CDF, RAD
2. comparison of goodness of fit methods
3. a note on hierarchical models and how commonly used normal hyperdistribuitons with log and logit link functions might nearly always be wrong

4 Biological and statistical limits to interpreting SADs

1. How small a sample is too small?
2. Can we distinguish log normal hyperdistributions from Gamma hyperdistributions?
3. Recent mechanismless theory says SADs might just be statistical outcome, so how much should we really infer from them?