

Zoey Diggory
Biol 862 Annotated Bibliography
Species Richness Estimators and Diversity Indices

This is truly an annotated bibliography: the richness estimator and diversity index literature is huge! Below is a small selection of the articles I found to be either the most helpful in understanding the subject matter, the best illustrations of the estimation methods, or the most interesting to read.

Bunge, J. and M. Fitzpatrick. 1993. Estimating the number of species: a review. Journal of the American Statistical Association 88(421): 364-373.

As the title implies, this is a review of different types of species estimators. Estimators are sorted into two primary schools – sampling-theoretic methods and data-analytic methods. Data-analytic methods include curve-fitting and lognormal estimators. Sampling-theoretic methods are sorted into various parametric and non-parametric methods. Performance of the various estimators is compared. In their opinion nearly all of the estimators reviewed have drawbacks that are significant enough to raise reasonable doubts about their conclusions. Their provisional choice of an estimator is a non-parametric Chao estimator. Finally, recommendations for future research in estimators are made. Because this is from a statistical journal it is a little tough to get through, but it is one of the most frequently cited articles in the species richness estimator literature and describes the underlying theory and assumptions of various schools of estimators well.

Chao, A. and S. Lee. 1992. Estimating the number of classes via sample coverage. Journal of the American Statistical Association 87(417): 210-217.

This is just one of several articles by Chao and her colleagues that describe the derivation of several non-parametric estimators that are based on capture-recapture data. This article describes a non-parametric estimator that uses the idea of sample coverage. It is referred to as “Chao2” in later review articles and includes a variance calculation. The estimator requires only incidence data and is based on the ratio of the number of species observed in only one sample and those observed in only two samples. The authors find that the estimator is biased to underestimating values. The article is hard to get through, but is cited extensively in the richness estimator literature. This estimator is described succinctly in a variety of review articles that are easier to read. Other Chao estimators are covered in the following articles: Chao1 - requires abundance data, relies on ratio between singletons and doubletons (Chao 1984); ACE – Abundance Coverage Estimator is based on ratio between rare (<10 individuals) and common (>10 individuals) species (Chao et al. 1993); ICE – Incidence Coverage Estimator is based on ratio of species observed in only one sample and those observed in only two samples (Chao et al. 1993).

Colwell, R.K. and J.A. Coddington. 1994. Estimating terrestrial biodiversity through extrapolation. Philosophical Transactions of the Royal Society of London 345: 101-118.

This is, arguably, the holy grail of richness estimator literature. In contrast to the Chao and Nichols capture-recapture approach to richness estimation (i.e., estimators are needed because species have different detection probabilities), Colwell and Coddington promote the use of estimators based on the overwhelming biodiversity of the planet – we can't possibly detect all species because there are so many. This article provides a good introduction to their argument and reviews several approaches to estimating richness: extrapolating species accumulation curves, fitting parametric models of relative abundance, and non-parametric methods. They discuss the performance of these

estimators using several data sets from tropical regions. They also discuss, in detail, measures of complementarity and the use of ratios (e.g., taxon ratios) to estimate richness. This article discusses many estimation methods broadly, so it is a good introduction to the subject, but not detailed enough for one to thoroughly understand the subject.

Deiller, A., J.N. Walter, and M. Tremolieres. 2001. Effects of flood interruption on species richness, diversity and floristic composition of woody regeneration in the upper Rhine alluvial hardwood forest. *Regulated Rivers: Research and Management* 17: 393-405.

This article describes a study on the Upper Rhine River to compare the vegetation composition on floodplains that have been isolated from flooding for varying numbers of years. The study used a first-order jackknife estimator to calculate species richness from seed rain, seed bank, and extant vegetation data. Shannon-Weaver diversity and equitability indices were used to compare species richness and abundance across sites. The study found that while flooded sites had the lowest species diversity, Shannon-Weaver values were higher because abundances were more even. Implications of flood restoration are then discussed. This study exemplifies several of the pitfalls of using richness estimators and diversity indices and serves as a good case study after one has learned about the derivation and limitation of richness estimators and diversity indices. Also, this study is a rare example of estimators and indices in practice; most of the literature is made up of comparative studies.

Gotelli, N.J. and R.K. Colwell. 2001. Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters* 4: 379-391.

This is a detailed look into two schools of richness estimation and provides a good compliment to articles about parametric and non-parametric estimators. The use of taxon sampling curves (e.g., accumulation and rarefaction) to extrapolate the number of species and compare community assemblages is reviewed. The differences between individual- and sample-based curves are discussed. The use of category-subcategory ratios (e.g., species-per-individual and species-per-genus) is also discussed. The limitations of these estimation methods are described well as well as their varying performance under different sampling schemes. If nothing else, this article helps illustrate that there are a seemingly endless number of methods that one can use to calculate potentially meaningless estimates of species richness.

Longino, J.T., J.T. Coddington, and R.K. Colwell. 2002. The ant fauna of a tropical rain forest: estimating species richness three different ways. *Ecology* 83(3): 689-702.

This study is a great introduction to the comparative performance of several richness estimators. As such, it is frequently cited in the richness estimator literature. The investigators have achieved a thorough inventory of the ant fauna in a tropical rain forest, and compare the performance of several estimators using varying stopping points in their species accumulation curve. This study seems to have gotten fairly close to observing the “true” number of ant species, and the calculated estimates are compared to this value (most studies have only a weak basis for the true number of species – species-area curves, expert opinion, museum collections, etc.). Three estimation methods are compared: fitting data to a lognormal distribution, the asymptote of the Michaelis-Menton equation, and Chao’s Incidence-based Coverage Estimator. The performance of all three estimators is disappointing, as they were all demonstrated to be strongly dependent on sample size (i.e., estimators only performed well when almost all species had been observed) and fairly unstable (i.e., estimates were not predictable). This article should be required reading for anyone interested in richness estimators.

Magurran, A.E. 2004. Measuring biological diversity. Blackwell Science Ltd., Oxford, UK.

This book is a terrific introduction to the study of the commonness and rarity of species. The book covers species abundance models (i.e., abundance distribution), sampling techniques, richness estimators, and diversity indices. Because the book is so recent, its discussions are quite up-to-date and her discussions are very inclusive. Explanation and comparisons of estimators are clear and concise, particularly after reading earlier chapters. This is a great review, and easy read, after getting acquainted with estimators in more detail following a review of the journal literature.

Mouillot, D. and A. Lepre tre. 1999. A comparison of species diversity estimators. Resource Population Ecology 41: 203-215.

Nearly all discussions of diversity indices, including this study, begin with the sentence: "Although having been much criticized, diversity indices are still widely used...". As the title implies, the performance of four diversity indices are compared: Shannon, Simpson, Camargo evenness, and Pielou regularity indices. The study uses simulated data and ANOVA to compare the bias, coefficient of variation, and root-mean-squared error of each index. They conclude that the Camargo index has the highest associated error, the Shannon and Pielou indices are the most biased, and that the Simpson index performs best. This is in keeping with other comparative studies, which consistently find that the Simpson index performs well (for a diversity index that is). This is a good introduction to the limitation of indices, with a strong mathematical basis.

Nichols, J.D., T. Boulinier, J.E. Hines, K.H. Pollock, and J.R. Sauer. 1998. Inference methods for spatial variation in species richness and community composition when not all species are detected. Conservation Biology 12(6): 1390-1398.

Along with Royle & Nichols (2003), this study is a good introduction to one school of richness estimators that is based on capture-recapture data (Chao non-parametric estimators are also of this school). The article outlines a study design and data analyses that can be applied to presence/absence data collected at different sampling occasions or locations. Like several of the Chao estimators, the analyses use the proportion of shared species between sites and the proportion of species found in only one location to estimate relative species richness that can be compared between study sites. Their estimation methods are illustrated using data from the North American Breeding Bird Survey. Limitations of the estimation method are not described well, and because the Nichols et al. approach has not yet made its way into many comparative studies, it is hard for the reader to evaluate the usefulness of the method presented. That said, because of the similarity in underlying data and ratios, one might assume that the Nichols method is comparable to Chao estimators.

O'Hara, R.B. 2005. Species richness estimators: how many species can dance on the head of a pin? Journal of Animal Ecology 74: 375-386.

Along with Longino et al. 2002, this should be required reading for those interesting in species richness estimators. This comparative study provides a great discussion of the underlying mathematics of several estimators, so that the basis for, and limitations of, the compared estimates are clear. Two non-parametric (Chao1, ACE), four rarefaction curves, and two parametric (negative binomial, Poisson) estimators are described and compared using simulated data. The only estimation method that gave a good estimate was the correct parametric estimate for the simulated data. Accumulation curves and non-parametric methods all underestimated the number of species. The study concludes that unless the actual number of species is known (and then estimation is unnecessary), or

the form of the underlying sampling distribution is known, the estimates cannot be relied upon. Ouch.

Petersen, F.T. and R. Meier. 2003. Testing species-richness estimation methods on single-sample collection data using the Danish Diptera. *Biodiversity and Conservation* 12: 667-686.

A comparison of species richness estimators in a worst-case, but most-likely, scenario. The study goal was to estimate total species richness of Diptera in Denmark using only museum collections. The performance of ACE, Chao1, Lognormal, and Poisson lognormal estimators was compared. None of the estimators performed well – they all greatly underestimated species richness – but the estimate of the true number of species (with which estimator performance was being compared) was pretty rough, based on a species-area curve and expert opinion. The study illustrates the fact that the performance of estimators is generally only as good as the data being analyzed and the investigators conclude that abundance-based estimators should only be used on almost complete samples of randomly distributed organisms. This is a good contrast to the Longino et al. (2002) study, which was a best-case scenario in regard to the “true” number of species.