SPECIAL FEATURE

The Statistics of Rarity[1]

Ecologists deal with rarity in many guises—species can be rare, particular interactions may be uncommon, and catastrophic events that reshape landscapes are, by definition, infrequent. Although ecologists often seek out abundant species or events for their investigations out of convenience, rare species are often of special central concern to conservation biologists, reserve managers, and legislators, and historical legacies of rare events are pervasive in ecosystems. Statistical analysis and modeling of rare events is often necessary, but as MacKenzie et al. point out in their contribution to this Special Feature, ''rare species [or events] are simultaneously the species for which strong inferences about state variables and vital rates are most needed and the species for which such information is most difficult to obtain.'' The standard set of statistical tools used by the majority of ecologists are difficult or inappropriate to use when analyzing rare species or events, either because assumptions such as normality or homoscedasticity do not hold, or because the required sample sizes are impossibly large. The overall goal of this Special Feature is to present a cross-section of techniques for sampling, quantifying, and modeling rarity.

Methods for analyzing data on rare species and events come from a variety of disciplines and were originally designed for addressing specific questions. Most of these questions, including problems related to quality control in manufacturing, frequency of flooding, and econometrics, are unrelated to ecological questions, but the mathematical and statistical tools transcend disciplinary boundaries. As a result, most of the papers in this Special Feature represent collaborations between ecologists and statisticians and reflect the growing need for interdisciplinary cooperation to effectively address complex ecological questions.

The first step in a study of rarity is collecting the data. Sampling rare species is challenging precisely because they are hard to find. The first three papers discuss sampling methods for rare species. The papers by Edwards et al. and Philippi focus on increasing the efficiency of sampling when the focal taxon is rare. Edwards et al. present methods for stratifying the sampling effort by modeling occurrences of common species known to be associated with the rare species of interest. Their interest is simply in determining whether or not a species occurs at a particular site. In contrast, Philippi addresses the problem of designing a sampling regime to more accurately estimate the abundance of a rare species. Measures of abundance are required to determine the legal status of a species (e.g., secure, threatened, or endangered), and the adaptive cluster sampling described by Philippi is more efficient than simple random sampling for estimating the abundance of sessile species that tend to be clumped. MacKenzie et al. discuss sampling and modeling methods for estimating the occupancy of a given site or patch by a mobile species. Occupancy can be used as a surrogate for abundance, especially in capture–recapture studies of mobile animals. Of particular importance is improvement of the accuracy of occupancy estimates when the probability of actually observing the species of interest is less than 1. The generality of all three of these approaches is reflected in the diversity of example organisms: lichens (Edwards et al.), herbaceous plants (Philippi), and giant wetas, gaurs, and salamanders (MacKenzie et al.).

Deriving precise estimates of rare events and expressing the uncertainty surrounding these estimates is challenging because the study of rare events often involves small sample sizes that provide little confidence in the estimates. Just as the papers by Edwards et al. and MacKenzie et al. take advantage of auxiliary data to improve sampling efficiency and estimates of occupancy, the paper by Dixon et al. uses auxiliary data to increase the precision (decrease the variance or uncertainty) in estimates of the frequency of rare events. Dixon et al. apply simple Bayesian methods with informed prior probability distributions, stratified data, regression with continuous

[1] Reprints of this 85-page Special Feature are available for $12.75 each, either as pdf files or as hard copy. Prepayment is required. Order reprints from the Ecological Society of America, Attention: Reprint Department, 1707 H Street, N.W., Suite 400, Washington, DC 20006.

covariates, and aggregated data from additional samples to the question of how frequently a sit-and-wait predator, the carnivorous plant *Darlingtonia californica*, captures prey. Large-scale disturbances are also infrequent events, and Katz et al. illustrate techniques for distinguishing the ''signal'' of extreme events from the ''noise'' of temporal variability. These techniques derive from the large body of statistical work devoted to the study of extremes, which were introduced to ecologists by Steve Gaines and Mark Denny in their 1993 paper, ''The largest, smallest, highest, lowest, longest, and shortest: extremes in ecology'' (*Ecology* **74**:1677–1692). Gaines and Denny focused on ''light-tailed'' distributions, in which the frequency of extreme events decreases at a relatively rapid rate, but many uncommon disturbances, including fires, floods, and large hurricanes, have ''heavy-tailed'' distributions, in which the frequency of extreme events decreases at a relatively slow rate. In their analysis of sedimentation rates at Nicolay Bay, Canada over ~500 years and in Chesapeake Bay since 1800, Katz et al. illustrate the utility of the statistics of extremes and show that these indicators of hydrological, climatological, and anthropogenic disturbances are heavy-tailed. The implication of their results is that some apparently rare events are less ''rare'' than would appear on first glance.

Armed with reasonable samples and reasonably precise parameter estimates, it is possible to model the effects of biotic processes or environmental drivers on the rare species or event of interest. Katz et al. show that the Nicolay Bay sedimentation record reflects only hydrological perturbations, whereas the Chesapeake Bay record reflects both climatic and anthropogenic disturbances. In a return to rare species, Cunningham and Lindenmayer use generalized linear models with link functions designed to account for excess zeros in presence/absence or abundance data to determine predictors of occurrence of Leadbeater's possum in Australian eucalypt forests, and predictors of abundance of nesting frigatebirds and boobies in the Coral Sea.

The last two papers address issues of major importance to conservation planners and managers. Mao and Colwell discuss how to accurately estimate the species richness of a given assemblage when some (or many) of the species are rare. Two questions are of paramount interest. First, can rare species that go undetected in inventories be accounted for statistically? Second, when is an inventory complete? Their analysis, based on empirical data for breeding birds, seeds, and beetles, as well as from simulations, suggests that while it is relatively straightforward to estimate a lower bound for the number of species present in an assemblage, estimating the actual number of species present remains a challenging statistical problem. Managers cannot always wait for better data, however, and Doak et al. discuss how to build demographic models for rare species when few data are available. The counterintuitive conclusion from their simulations is that, for rare species, simple deterministic models of population growth and population viability analyses are preferable to data-hungry stochastic models, even though the former have known biases.

Many of the methods presented in these papers will be unfamiliar to ecologists raised on a steady diet of ANOVA, regression, and contingency tables. To encourage the use of these tools, the authors have published the statistical code and data sets accompanying their papers in *Ecological Archives*, or have made sophisticated packages freely available on their web sites. We hope that these methods will be used and that they will continue to evolve through productive collaborations between ecologists, statisticians, conservation planners, and managers.

—Aaron M. Ellison
Associate Editor-in-Chief
—Anurag A. Agrawal
Special Features Editor

Key words: adaptive cluster sampling; Bayesian inference; detection probability; extremes; general linear models; maximum likelihood; parameter estimation; Poisson distribution; population viability analysis; precision; rarity; sampling.