

Anne Chao, Nicholas J. Gotelli, T. C. Hsieh, Elizabeth L. Sander, K. H. Ma, Robert K. Colwell, and Aaron M. Ellison. 2013. Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies. *Ecology Monographs*.

Appendix A: A binomial product model can incorporate spatial aggregation for quadrat sampling

Quadrat sampling has been widely used to estimate abundance of plants and other sessile organisms. Our model for incidence data (Eqs. 2a and 2b in the main text), when applied to quadrat sampling, can incorporate spatial aggregation or clustering.

We first show how a binomial product can be obtained from quadrat sampling. Suppose that the region under investigation is divided into T^* disjoint quadrats of the same area and a sample of T quadrats are randomly selected. Then each selected quadrat is surveyed and the presence or non-detection of any species for each of these T quadrats is recorded. We assume that T is small relative to T^* , so that sampling with or without replacement differ little.

Let M_i be the *unknown* number of quadrats occupied by the i th species, $i = 1, 2, \dots, S$. Assume that, in each sampling unit, the conditional probability of recording species i in any selected quadrat (given that species i is present) is $0 < \alpha_i \leq 1$. That is, any selected sampling unit need not be completely censused. For this reason, our model also can be applied to sampling mobile animals. The model assumes that, out of these M_i quadrats, species i can only be recorded in U_i quadrats. Here, U_i is also *unknown* and $M_i \geq U_i \geq 1$. (For any species with $U_i = 0$, there is no chance to detect this species in any sample, so it should be excluded in the estimating target.) Thus, U_i is a truncated binomial distribution with probability

$$P(U_i = u) = \binom{M_i}{u} \alpha_i^u (1 - \alpha_i)^{M_i - u} / \{1 - (1 - \alpha_i)^{M_i}\}, \quad u = 1, 2, \dots, M_i. \quad (\text{A.1})$$

In the other $T^* - U_i$ quadrats, either species i is absent or it is present but cannot be detected. Here we may assume any type of distribution for the probability α_i (e.g., α_i may be a constant, or α_i follows a uniform or beta distribution).

As discussed in the main text, we can form a species-by-quadrat incidence matrix (W_{ij}) with S rows and T columns. The value of element W_{ij} of this matrix is 1 if species i is recorded in the j th quadrat, and 0 otherwise. The row sum of the incidence matrix $Y_i = \sum_{j=1}^T W_{ij}$ denotes the incidence-based frequency of species i , $i = 1, 2, \dots, S$. The sample frequencies (Y_1, Y_2, \dots, Y_S) given U_i , $i = 1, 2, \dots, S$, follow a binomial product distribution:

$$P(Y_i = y_i, i = 1, 2, \dots, S | U_i, i = 1, 2, \dots, S) = \prod_{i=1}^S \binom{T}{y_i} \left(\frac{U_i}{T^*} \right)^{y_i} \left(1 - \frac{U_i}{T^*} \right)^{T-y_i}, \quad 1 \leq u_i \leq M_i. \quad (\text{A.2})$$

Combining (A.1) and (A.2), Chao and Chiu (2013) showed that the sample frequencies (Y_1, Y_2, \dots, Y_S) follow approximately a binomial product distribution:

$$P(Y_i = y_i, i = 1, 2, \dots, S) = \prod_{i=1}^S \binom{T}{y_i} \left(\frac{M_i \alpha_i}{T^*} \right)^{y_i} \left(1 - \frac{M_i \alpha_i}{T^*} \right)^{T-y_i}. \quad (\text{A.3})$$

This is a special case of the binomial product model with T sampling units and species incidence probability π_i for species i ; see Eq. 2b in the main text. Here, the probability π_i for species i in any quadrat is a product of rate M_i / T^* and α_i . Because there are no restrictions on the variables (M_1, M_2, \dots, M_S) , this model is valid even if species are spatially aggregated or clustered.

The theory we prove here implies that the proposed binomial product model can be applied to any species distribution patterns in a region. Colwell et al. (2004, their Table 2) provided an illustrative example to demonstrate that two contrasting patterns (one region with association between species and the other with dissociation between species) give the same sample-based rarefaction curves as long as the sample frequencies (Y_1, Y_2, \dots, Y_S) for the two regions are identical.

LITERATURE CITED

- Chao, A., and C.-H. Chiu. 2013. Estimation of species richness and shared species richness. To appear as an article in *Handbook of Methods and Applications of Statistics in the Atmospheric and Earth Sciences*. Wiley, New York, USA.
- Colwell, R. K., C. X. Mao, and J. Chang. 2004. Interpolating, extrapolating, and comparing incidence-based species accumulation curves. *Ecology* 85:2717-2727.