

1 **Methods**

2 **The multinomial null model**

3 Let $N_{i,j}$ be the number of fish of species i at site j . If the species are
4 distributed randomly across sites, then the distribution of of species at site
5 j should follow a multinomial distribution such that

$$\mathbf{N}_j \sim \text{Multinomial}(a, \mathbf{p}) \quad (1)$$

6 where a is the total number of fish counted, i_{max} is the maximum number of
7 species that can be observed and \mathbf{N}_j and \mathbf{p} are vectors of length i_{max} . Here,
8 p_i is the probability of observing species i , and \mathbf{p} is the vector of probabilities.

9 In our analysis, we use the maximum likelihood estimator for p_i , which
 10 is simply the proportion of the total number of fish that are from species i
 11 (Fig. 1)

$$\hat{p}_i = \frac{\sum_j N_{i,j}}{\sum_{i,j} N_{i,j}} \quad (2)$$

12 To determine whether \mathbf{N}_j follows this distribution, we compare the diver-
 13 sity of sites in \mathbf{N} to the diversity of random samples from the multinomial
 14 distribution defined by Equation 1. Let $\boldsymbol{\chi}(a)$ be a sample of a fish from a
 15 multinomial distribution with probability $\hat{\mathbf{p}}$,

$$\boldsymbol{\chi}(a) \sim \text{Multinomial}(a, \hat{\mathbf{p}}) \quad (3)$$

16 To calculate the diversity of each sample, we first determine which species
 17 are present or absent in the sample and then we sum across species.

$$S_{\boldsymbol{\chi}}(a) = \sum_i \tilde{\chi}_i(a) = \begin{cases} 1 & \text{if } \chi_i(a) > 0 \\ 0 & \text{if } \chi_i(a) = 0 \end{cases} \quad (4)$$

18 where S , is the species richness or diversity. Note that $S_{\boldsymbol{\chi}}$ is a random
 19 variable because it is a function of the random variable $\boldsymbol{\chi}$ given in Equation
 20 3. Similarly, let us define the species richness or diversity of our data

$$S_j(a) = \sum_i \tilde{N}_{i,j}(a) = \begin{cases} 1 & \text{if } N_{i,j}(a) > 0 \\ 0 & \text{if } N_{i,j}(a) = 0 \end{cases} \quad (5)$$

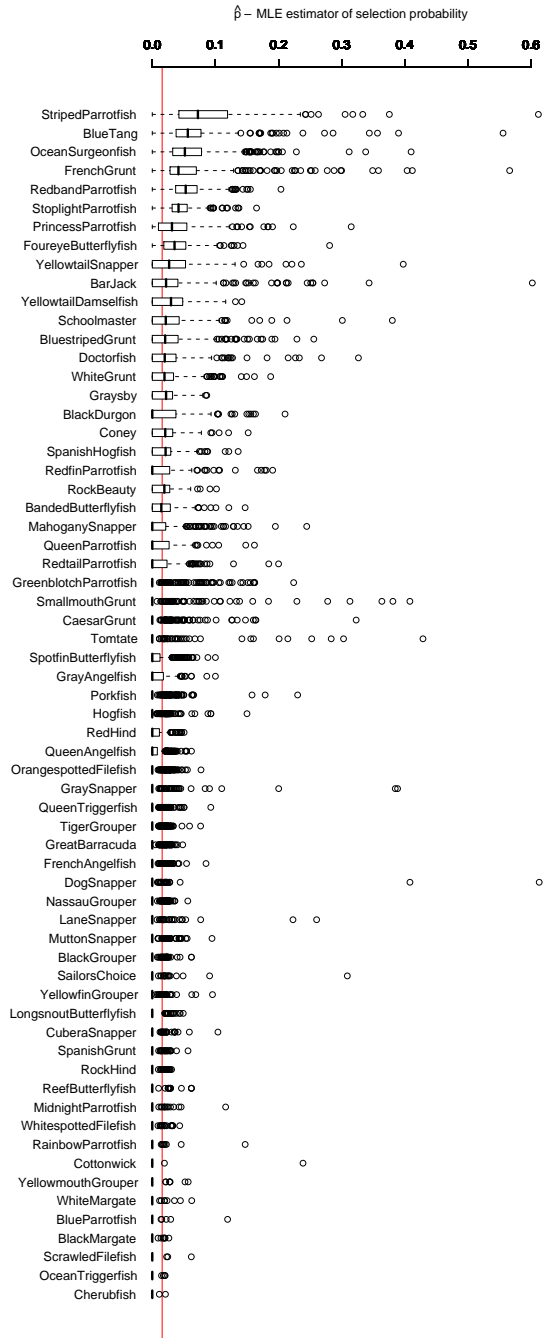


Figure 1: Boxplots of the proportion of species i found at site j .

21 Here, S_j is the diversity observed at site j and it is not a random variable.
22 In Figure 2, we have plotted the distribution of diversities from our ran-
23 dom samples, $S_\chi(a)$, and we compare this to the diversity from our data,
24 $S_j(a)$. A cursory inspection reveals that the data does not follow the random
25 distribution.

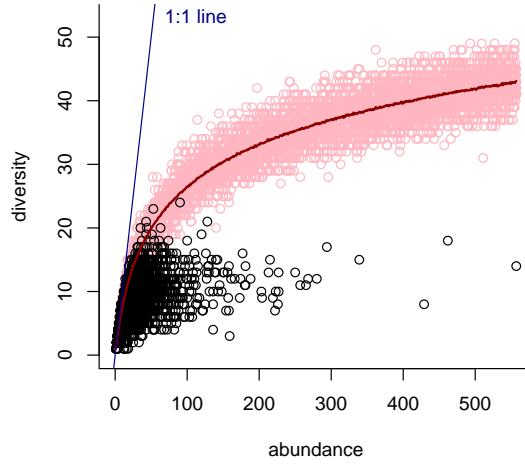


Figure 2: Diversity as a function of abundance. The pink points are random draws from the distribution of S_χ (see Equation 3). The dark red line is the mean of the distribution. The black dots are S_j .

26 The residuals analysis

27 We believe that the deviation from the null model can be explained by com-
 28 munity assemblages that are predictable by habitat. To do this analysis, we
 29 calculate the data residuals for diversity with respect to abundance.

$$\delta(a) = S_j(a) - \bar{S}_j(a) \quad (6)$$

30 where $\bar{S}_j(a)$ is the expected number of species observed at abundance a and
 31 δ is the number of species more or less than average observed at a given site.

32 We then regress our habitat variates on δ . Regression analysis goes here.

$$\delta \sim a + \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_h x_h + \epsilon \quad (7)$$

33 Partitioning the data

34 We partition the data based on the regression analysis. Maybe instead of
35 regression we should be doing ordination or clustering or something.

36 The summed multinomial distribution

37 Based on the partitions found in the previous section, let $N_{i,j,k}$ be the number
38 of fish of species i , at site j from partition k . Similarly, let \mathbf{N}_k be a matrix
39 of dimension $i_{max} \times j_{max}$ and containing all the fish of partition k .

40 Let's define the multinomial distributions for each partition. Let $\hat{\mathbf{p}}_k$ be
41 the probability of drawing species within partition k . As before, this is based
42 on the proportion of fish of species i found within partition k

$$\hat{p}_{i,k} = \frac{\sum_j N_{i,j,k}}{\sum_{i,j} N_{i,j,k}} \quad (8)$$

43 $\hat{\mathbf{p}}_k$ has length equal to the number of species observed in partition k . We next
44 find a sample from the multinomial distribution associated with partition k

$$\chi_k(a) \sim \text{Multinomial}(a, \hat{\mathbf{p}}_k) \quad (9)$$

45 we next calculate the diversity of this sample according to Equation 10

$$S_{\chi}(a, k) = \sum_i \tilde{\chi}_{i,k}(a) = \begin{cases} 1 & \text{if } \chi_{i,k}(a) > 0 \\ 0 & \text{if } \chi_{i,k}(a) = 0 \end{cases} \quad (10)$$

46 and we sum across partitions to get the summed distribution

$$S_K(a) = \sum_k S_{\chi}(a, k) \quad (11)$$

47 where S_K is the summed multinomial distribution. We can now compare our
48 data, $S_j(a)$, to this new distribution (see Figure ??).