

生物群落多样性的测度方法

VI :与多样性测度有关的统计问题*

刘灿然 马克平

(中国科学院植物研究所,北京 100093)

吕延华

(河南教育学院,郑州 450003)

康永亮

(河南省禹州市农业局,河南 462500)

摘要 物种多样性测度是(群落的)总体参数,它们常常是未知的,需要通过抽样将它们估计出来,因此,必须了解估计量的抽样性质。本文对与一些多样性测度的均值、方差的估计和假设检验以及大、小样本分布等有关的问题作了综述。可以看出大多数多样性测度的抽样性质还不清楚,有些甚至根本就没有研究过。Pielou的合并样方法和刀切法是两个比较通用的方法,用它们可以解决其中的一些问题。但相比之下,刀切法更实用。

关键词 群落,物种多样性,多样性测度,抽样,估计,刀切法

Measurement of biotic community diversity VI: the statistical aspects of diversity measures/LIU Can-Ran¹⁾, MA Ke-Ping¹⁾, Lü Yan-Hua²⁾, KANG Yong-Liang³⁾

Abstract Species diversity is a community parameter. It's usually unknown, and must be estimated with a sample or samples from the community. Therefore the sampling properties of the estimators must be examined in order to accurately estimate species diversity of a community. The estimation of means and variances, hypothesis testing, and large or small sample distributions for some diversity measures are reviewed in this paper. However, the sampling properties of many diversity measures remain unclear, and some of them have not been studied at all. It is concluded that Pielou's pooled quadrat method and jackknife method are two better methods to solve some of the problems, and jackknife method is more practical in application than the other.

Key words community, species diversity, diversity measure, sampling, estimation, jackknife method

Author's address 1)Institute of Botany, Chinese Academy of Sciences, Beijing 100093

2)Henan Education College, Zhengzhou 450003

3)Yuzhou Bureau of Agriculture, Henan 462500

1 引言

物种多样性测度是(群落的)总体参数^[1],在实际中,它们的值是未知的,需要用从群落中抽取的样本估计出来^[2]。由于估计量本身的偏差以及抽样误差的存在,估计值往往偏离真实值。因此,只有对这些多样性测度的抽样性质(包括大小样本分布、均值和方差等)有充分的了解,才能对群落的多样性作出合理的估计,并对不同群落的多样性作出正确的评价。

2 物种丰富度

物种丰富度是一个群落中的物种数目,它是基于物种的存在与否,而不是物种的相对多度

的一个多样性测度^[3]。

2.1 观察的物种数目

设 S 为一个包含大量个体的群落中实际的物种数目, p_i 为第 i 个种的相对多度, 并设一个包含 n 个个体的样本中出现的物种数目为 s , 它具有均值^[4]

$$E(s) = S - \sum_{i=1}^S (1 - p_i)^n$$

和方差^[3]

$$\text{Var}(s) = \sum_{i=1}^S (1 - p_i)^n [1 - (1 - p_i)^n] + 2 \sum_{i>j} [(1 - p_i - p_j)^n - (1 - p_i)^n (1 - p_j)^n]$$

只有当 $p_i n > 1$ 时, 种 i 才可能在大小为 n 的样本中出现。因此, 在一个多样性很高的群落中, 观察的物种数目 s 可能严重地低估真实的物种数目 S 。因为即使抽取相当大的样本, 稀有种也常常会不在其中出现^[3]。

关于群落物种数目的估计方法, 笔者已做过评述^[5], Colwell 和 Coddington(1994)^[6]也介绍了几种用不同大小的样本通过外推估计物种数目的方法。估计群落中真实物种数目的所有外推方法都包含有或隐含有关于稀有种多度分布形式的假设, 因此, 它们在一个未知的程度上存在着误差^[3]。

2.2 稀疏多样性测度(Rarefaction diversity measurement)

Sanders^[7]在关于洋底生物多样性研究中使用了 $E(S_n)$ (在包含 N 个个体、 S 个种的一个集合中, 随机抽取包含 n 个个体的一个样本, 在其中期望出现的物种数目), 提出了“稀疏”(Rarefaction)这个术语, 但他的方法高估了期望出现的物种数目^[8-11]。Hurlbert^[8]和 Simberloff^[10]给出了这个过高估计的一般理由, 他们指出真正的 $E(S_n)$ 是由超几何分布确定的。后者还写了一个简单的计算机程序来模拟随机抽样过程, 从而估计真正的 $E(S_n)$ 和这个参数的置信区间^[11]。Hurlbert^[8]给出了

$$E(S_n) = S - (C_N^n)^{-1} \sum_{i=1}^S C_{N-N_i}^n, \quad ,$$

Heck^[11]又给出了

$$\text{Var}(S_n) = (C_N^n)^{-1} \left[\sum_{i=1}^S C_{N-N_i}^n \left(1 - \frac{C_{N-N_i}^n}{C_N^n}\right) + 2 \sum_{i<j} \left(C_{N-N_i-N_j}^n - \frac{C_{N-N_i}^n C_{N-N_j}^n}{C_N^n}\right) \right], \quad ,$$

其中 N_i 是未稀疏的样本中种 i 的个体数。

如果 n 比 N 小得多, 以致随机地抽取 n 个个体可以近似地看作有放回的抽样, 或者如果有生物学上的理由认为相继地抽取个体并不影响原始样本中的种-个体分布 (species-individuals distribution), 那么就可以应用多项式分布, 而不用超几何分布, 并且

$$E(S_n) = S - \sum_{i=1}^S (1 - N_i/N)^n$$

$$\text{Var}(S_n) = \sum_{i=1}^S (1 - N_i/N)^n [1 - (1 - N_i/N)^n] + 2 \sum_{i<j} [(1 - N_i/N - N_j/N)^n - (1 - N_i/N)^n (1 - N_j/N)^n] \quad .$$

上述 $E(S_n)$ 被 Peet^[12]和 Hurlbert^[8]称为物种丰富度测度 (species richness measure), 被 Sanders^[7]称为多样性测度 (diversity measure), Smith 和 Grassle^[2]则称其为期望物种测度 (expected species measure) 以及期望物种多样性 (expected species diversity)。对于给定的个体指

数(individual index) n 期望物种测度满足 Pielou^[13] 关于多样性测度的描述, 即依赖于样本中的物种数目和样本中的个体在各物种间分布的均匀程度。最好将期望物种测度看作一族多样性测度, 因为对不同的个体指数 n , 它们对稀有种具有不同的敏感性^[2]。期望物种测度将 Simpson 多样性指数加以扩展, 在保留了 Simpson 指数好的抽样性质的同时, 又使稀有种具有更大的敏感性。对于大的 n , 期望物种测度是文献中见到的唯一的一个既对稀有种敏感又无偏的测度^[2]。

Smith 和 Grassle^[2]对 Hurlbert^[8]的期望物种测度给出了一个更准确的定义。

假设有一个包含 S 个种的有限总体, 其中种 i 有 N_i 个个体, $N = (N_1, N_2, \dots, N_S)$ 代表整个有限总体, 令随机变量 s 表示具有 n 个个体的样本中的物种数, 则对于给定的有限总体 N , 在无放回抽样抽取的一个包含 n 个个体的随机样本中期望出现的物种数目为:

$$E(s/N) = \sum_{i=1}^S [1 - C_{N-N_i}^n / C_N^n] \quad (1)$$

其中, N 是有限总体中的总个体数, $N = \sum_{i=1}^S N_i$ 。

假设有一个多项式总体, 其中 p_i 表示总体中种 i 的个体数所占的比例, $p = (p_1, p_2, \dots, p_S)$ 是描述该总体的一个概率向量, 多项式抽样模型假设总体是无限的, 并且对抽取的每一个个体的识别不依赖于抽取的其它个体。换句话说, 就是在种内不存在聚集性。于是, 在具有 n 个个体的一个样本中期望出现的物种数目为:

$$S(n) = E(s/p) = \sum_{i=1}^S [1 - (1 - p_i)^n]$$

当个体指数 $n=2$ 时, $S(n)$ 与 Simpson 多样性指数 ($SI = 1 - \sum_{i=1}^S p_i^2$) 相联系, 即 $S(2) = 1 + SI$ 。对于小的 n , $S(n)$ 强调更富有的种; 对于较大的 n , 该测度对较不富有的种也有反应。当然, 对于大的 n , $S(n)$ 趋近于在多项式总体中出现的总物种数。

与其它多样性指数相比, Hurlbert 的 $S(n)$ 和 Simpson 的 SI 的优势表现在它们都有最小方差无偏估计量。假设有来自多项式总体 p 的包含 N 个个体的一个样本, 其中种 i 有 N_i 个个体。令 $N = (N_1, N_2, \dots, N_S)$ 表示描述该随机样本的向量, 通过该随机样本 N , 我们希望估计从该总体中抽取的一个包含 n 个个体的样本中期望出现的物种数目 $S(n)$, 其最小方差无偏估计量为:

$$\hat{S}(n) = E(s/N)$$

其中, $E(s/N)$ 由 (1) 式给出。

Smith 和 Grassle^[2]还给出了 $\hat{S}(n)$ 的方差的一个无偏估计量, 用它可以构造 $S(n)$ 的一个近似的置信区间; 他们还讨论了在非同质总体中相应的估计问题。

3 多样性指数

3.1 Shannon 指数

Shannon 多样性指数假设个体是从一个“无限大”的总体中随机抽取的^[13,14], 它还假设总体中的所有种都在样本中出现, 它可由下式计算

$$H' = - \sum_{i=1}^S p_i \log p_i$$

其中 p_i 是种 i 的个体数在总体中所占的比例。

假设有一个包含 N 个个体的随机样本, 其中种 i 的个体数为 n_i , 则 p_i 有极大似然估计量^[14,15]

$$\hat{p}_i = n_i/N \quad .$$

从而, Shannon 指数可以用下式来估计

$$\hat{H}' = - \sum_{i=1}^S (n_i/N) \log(n_i/N) \quad .$$

Hutcheson^[16]计算了 \hat{H}' 的期望和方差,

$$E(\hat{H}') = - \sum p_i \ln p_i - \frac{S-1}{N} + \frac{1 - \sum p_i^{-1}}{12N^2} + \frac{\sum (p_i^{-1} - p_i^{-2})}{12N^3} + \dots \quad (2)$$

$$\text{Var}(\hat{H}') = \frac{\sum p_i (\ln p_i)^2 - (\sum p_i \ln p_i)^2}{N} + \frac{S-1}{2N^2} + \dots \quad (3)$$

对于大多数应用来说, \hat{H}' 的偏差很小^[12] (2)式中从第3项开始的余项和以及(3)式中从第2项开始的余项和实际上都很小。一个更重要的误差来源就是总体中的物种没有在样本中全部出现^[12,14], 这个误差随物种数在样本中所占比例的减小而增加^[14]。

Hutcheson^[16]根据(2)、(3)两式, 提出了一个 t 检验方法, 来检验两个群落的多样性指数之间差异的显著性^[12,14], 所用的统计量为:

$$t = \frac{\hat{H}'_1 - \hat{H}'_2}{[\text{Var}(\hat{H}'_1) + \text{Var}(\hat{H}'_2)]^{1/2}} \quad (4)$$

其中 \hat{H}'_1 和 \hat{H}'_2 分别是用样本1和样本2对群落1和群落2所作的多样性估计, $\text{Var}(\hat{H}'_1)$ 和 $\text{Var}(\hat{H}'_2)$ 分别是相应的方差估计。与(4)相应的自由度为:

$$df = \frac{[\text{Var}(\hat{H}'_1) + \text{Var}(\hat{H}'_2)]^2}{[\text{Var}(\hat{H}'_1)]^2/N_1 + [\text{Var}(\hat{H}'_2)]^2/N_2}$$

其中 N_1 和 N_2 分别是样本1和样本2中的总个体数。

然而, 由于(3)式方差的计算还需要总体中物种数的知识, 这就大大地降低了这种方法在生态学中的实用性。从(2)式可以看出 \hat{H}' 的偏差也依赖于真实的物种数目 S , 而它在一般情况下是未知的。因此, Shannon 指数的无偏估计量不存在。对 H' 准确的估计要求抽取大量的个体, 使得 $2N$ 比总体中真实的物种数目 S 大得多^[3]。

实际上, Bowman 等已经广泛地评述了 Shannon 多样性指数的小样本性质。他们指出, 平均来说从样本中得到的估计值将低估真正的总体值。也就是说该指数是有偏的, 其偏差依赖于样本大小。当样本很大时, 偏差趋近于0。然而, 由于该偏差也依赖于真正的总体多样性, 故在偏差与样本大小之间不存在简单的关系。从两个不同的总体(即群落)中抽取的样本估计的多样性之间的差别可能来自三个方面: 真正的总体多样性之间的差别, 估计量的依赖于样本大小的偏差, 以及随机抽样误差。并且要把这三种影响分离出来也是很困难的, 尤其是当样本很小以及偏差很大的时候^[2]。

Taylor 指出, 如果对一些样本计算了 Shannon 指数, 那么这些指数本身将是正态分布的。这个性质使得有可能应用参数统计(包括强有力的方差分析)方法来比较这组样本所在的各种总体的多样性。这是比较不同生境中的多样性的一个有用的方法, 尤其是在有一些重复的时候^[14]。

Gill 和 Joanes^[17]假设总的物种数目 S 是已知的, 并且相对较小。他们利用 Bayes 方法对 Shannon 多样性指数提出了两组估计量

$$h_k = - \sum_{i=1}^S [(n_i + k)/(n + kS)] \log[(n_i + k)/(n + kS)], \quad k > 0$$

$$\bar{h}_k = \sum_{i=1}^S \frac{n_i + k}{n + kS} [\psi(n + kS + 1) - \psi(n_i + k + 1)], \quad k \geq 0$$

其中 $\psi(x)$ 为 digamma 函数 n_i 是包含 S 个种的随机样本中种 i 的个体数 $n = \sum_{i=1}^S n_i$ k 为一个常数。

他们还推导出了 h_k 和 \bar{h}_k 的期望和方差(但由于方差的式子太长而在原文中没有给出):

$$E(h_k) = \log(n + kS) - \frac{1}{n + kS} \sum_{r=0}^n C_n^r(k+r) \log(k+r) \sum_{i=1}^S p_i^r (1-p_i)^{n-r}$$

$$E(\bar{h}_k) = \psi(n + kS + 1) - \frac{1}{n + kS} \sum_{r=0}^n C_n^r(k+r) \psi(k+r+1) \sum_{i=1}^S p_i^r (1-p_i)^{n-r}$$

其中 p_i 是总体中种 i 所占的比例($i = 1, 2, \dots, S$)。如果期望总体中各个种的分布相当均匀,则应该取较大的 k ;反之,就取较小的 k 。 $k = 1$ 对应于总体中的物种分布符合 MacArthur 的“断棍”模型(broken stick model);如果认为总体中的物种分布比“断棍”模型还均匀的话就取 $k > 1$;反之,取 $0 \leq k < 1$ 。

Tong^[11]对相当广泛的一类多样性测度研究了它们的大样本性质。设 $\hat{p}_1, \hat{p}_2, \dots, \hat{p}_s$ 包含 N 个个体的群落中随机有放回地抽取的包含 n 个个体的一个随机样本中的样本相对多度向量 $\mathbf{p} = (p_1, p_2, \dots, p_s)$ 为总体(群落)相对多度向量 $f(\mathbf{p})$ 为群落的一个多样性指数 $f(\hat{\mathbf{p}})$,为其估计量。如果 $f(\hat{\mathbf{p}})$ 可微,使得

$$\Phi_i(\mathbf{p}) = \frac{\partial}{\partial p_i} f(\hat{\mathbf{p}}) \Big|_{\hat{\mathbf{p}} = \mathbf{p}}$$

存在 $j = 1, 2, \dots, S$ 。并且如果

$$\tau^2 = \tau^2(\mathbf{p}) = \sum_{i=1}^S p_i \Phi_i^2(\mathbf{p}) - \left[\sum_{i=1}^S p_i \Phi_i(\mathbf{p}) \right]^2 > 0$$

则当 $n \rightarrow \infty$ 时,

$$n^{1/2} [f(\hat{\mathbf{p}}) - f(\mathbf{p})] \rightarrow N(0, \tau^2)$$

从而 $f(\hat{\mathbf{p}})$ 的极限分布是正态分布 $N(f(\mathbf{p}), \tau^2)$ 。

Shannon 多样性指数 $H' = - \sum_{i=1}^S p_i \log p_i$ 是 Tong^[11]的结果的一个特例,对

$$\hat{H}' = \sum_{i=1}^S \hat{p}_i \log \hat{p}_i \quad \text{有,}$$

$$n^{1/2} (\hat{H}' - H') \rightarrow N\left(0, \sum_{i=1}^S p_i (\log p_i)^2 - (H')^2\right)。$$

从而 \hat{H}' 的极限分布是 $N\left(H', \mu \left[\sum_{i=1}^S p_i (\log p_i)^2 - (H')^2 \right]\right)$ 。

Kempton 和 Wedderburn^[18]假设总体的种-多度分布为 Gamma 分布,推导出了 Shannon 多样性指数 H' 及其变形 $\exp(H')$ 的大样本均值和方差。May^[19]也讨论了 H' 及其它几个指数在假设总体的种-多度分布为对数正态分布和断棍分布时的抽样性质。

上述的讨论都是基于个体计数的。然而,在应用中,大多数生态学家都认为一个种对一个群落的重要性应该基于一个连续量(如一个种的生物量或生产力)的相对数量^[8,20,21](这些量笼统地称为“重要值”)。实际中,这些重要值有时就代替了多样性指数中的 n_i ,而指数中的 N 则被所有种的重要值之和所取代。这样一来,方差就是未知的,而将用个体数计算指数的方差的公式拿来直接套用是不恰当的。这样的估计依赖于重要值测量的单位,而测度公式本身是不变的^[21]。

Lyons^[21]对基于固定权重的相当广泛的一类多样性指数推导出了渐近分布和方差。假设一个总体包含 S 个种, 种 i 的相对多度为 $p_i, i = 1, 2, \dots, S, \sum_{i=1}^S p_i = 1$; w_i 为赋予种 i 的每个个体的固定权重或重要值。再假设样本中种 i 的个体数为 $n_i, i = 1, 2, \dots, S$, 并且 $N = \sum_{i=1}^S N_i$, 则 $w_i n_i / \sum w_i n_i$ 就代表样本中种 i 的相对重要性, 通过适当地选择函数 Q 和 $f_i (i = 1, 2, \dots, S)$ 可以将相当广泛的一类多样性指数表示为:

$$Q[\sum f_i(w_i n_i / \sum w_i n_i)]$$

于是, 总体多样性可以表示为: $\mathcal{T}(p_1, p_2, \dots, p_s) = Q[\sum f_i(w_i p_i / W_p)]$,

对应的样本统计量可以表示为: $\mathcal{T}(\hat{p}_1, \hat{p}_2, \dots, \hat{p}_s) = Q[\sum f_i(w_i \hat{p}_i / W_N)]$

其中, $W_N = \sum w_i \hat{p}_i, W_p = \sum w_i p_i$ 。

Lyons^[21]指出, $\mathcal{T}(\hat{p}_1, \hat{p}_2, \dots, \hat{p}_s)$ 渐近地服从均值为 $\mathcal{T}(p_1, p_2, \dots, p_s)$, 方差为:

$$\frac{1}{NW_p^2} Q[\sum f_i(w_i p_i / W_p)]^2 \sum w_i^2 p_i [f'_i(w_i p_i / W_p) - \sum (w_i p_i / W_p) f'_i(w_i p_i / W_p)]^2$$

Shannon 多样性指数是 Lyons^[21]所考虑的一类多样性指数中的一个特例。于是,

$$\hat{H}' = \mathcal{T}(\hat{p}_1, \hat{p}_2, \dots, \hat{p}_s) = -\sum (w_i \hat{p}_i / W_N) \log(w_i \hat{p}_i / W_N)$$

$$H' = \mathcal{T}(p_1, p_2, \dots, p_s) = -\sum (w_i p_i / W_p) \log(w_i p_i / W_p)$$

对于很大的 N , 有 $\text{Var}(\hat{H}') \approx \frac{1}{NW_p^2} \sum w_i^2 p_i [\log(w_i p_i / W_p) + H']^2$

如果每个种的权重都相同, 则

$$\text{Var}(\hat{H}') \approx \frac{1}{N} [\sum p_i \log^2 p_i - (H')^2]$$

$\exp(H')$ 也是上述一般情况的一个特例。

$$\exp(\hat{H}') = \exp[-\sum (w_i \hat{p}_i / W_N) \log(w_i \hat{p}_i / W_N)]$$

$$\exp(H') = \exp[-\sum (w_i p_i / W_p) \log(w_i p_i / W_p)]$$

并且,

$$\text{Var}[\exp(\hat{H}')] \approx \exp(2H') \text{Var}(\hat{H}')$$

上述公式中的方差可以用 \hat{p}_i 代替 p_i 估计出来。

用 \hat{t} 表示对多样性指数的估计, $\hat{\sigma}^2$ 表示对方差的估计, 则对上述任意一个加权的多样性指数都可以用统计量

$$Z = (\hat{t}_1 - \hat{t}_2) [\hat{\sigma}_1^2 + \hat{\sigma}_2^2]^{1/2}$$

对用大小分别为 N_1 和 N_2 的两个独立的样本估计的多样性进行差异显著性检验, 并且这个统计量 Z 是渐近地服从正态分布的。还有一个优点就是这个统计量对用来确定权重的测量单位是不变的^[21]。

上述的讨论都假设随机抽样是对个体进行的。然而, 集团抽样常常比对个体的简单随机抽样要容易得多。例如, 用样方对植物抽样、用网或收集器对动物抽样更容易^[22]。因此, 上述各种方法对这种情况就不再适用。一个更实际的方法就是利用刀切法, 它不要求关于个体的随机抽样, 并且可以采用重要值等作为多度指标, 而不必用个体数。Zahl^[23]、Routledge^[22]等已将该方法应用于 Shannon 指数 H' 及其变型 $\exp(H')$ 的研究^[14, 24]。笔者已将该方法在生物多样性测度研究中的应用做过详细的介绍^[25]。

3.2 Simpson 指数

Simpson 多样性指数可表示为 $D = 1 - \sum_{i=1}^S p_i^2$

其中 S 为群落中的物种数, p_i 为种 i 的个体数所占的比例, $i = 1, 2, \dots, S$ 。这一公式是由 Gini 最早提出的, 由 Berger 和 Parker^[26] 建议使用的^[11]。它是 Simpson 的优势度指数的补^[27], 即 Simpson 多样性指数与 Simpson 优势度指数之和为 1。

设有包含 s 个种、 N 个个体的一个随机样本, 种 i 的个体数为 n_i , $i = 1, 2, \dots, s$ 。则 D 的最小方差无偏估计量为^[15, 28]:

$$\hat{D} = 1 - \sum_{i=1}^s \frac{n_i(n_i - 1)}{N(N - 1)}。$$

其方差为^[3, 28]

$$\text{Var}(\hat{D}) \approx \frac{4}{N} \left[\sum_{i=1}^s p_i^3 - \left(\sum_{i=1}^s p_i^2 \right)^2 \right] \quad (5)$$

D 的另一个估计量是^[3]

$$\hat{D}_1 = 1 - \sum_{i=1}^s (\hat{p}_i)^2 = 1 - \sum_{i=1}^s \left(\frac{n_i}{N} \right)^2,$$

其均值为:

$$E(\hat{D}_1) = \left(1 - \frac{1}{N} \right) D$$

其大样本方差同(5)式。因此, D 还有一个无偏估计量

$$\hat{D}_2 = \frac{N}{N-1} \hat{D}_1 = \frac{N}{N-1} \left[1 - \sum_{i=1}^s \left(\frac{n_i}{N} \right)^2 \right]。$$

Tong^[11]关于一类广泛的多样性指数得出的结论对一般的 Simpson 指数

$H_\alpha(\mathbf{p}) = \frac{1}{1-\alpha} \log \sum_{i=1}^s p_i^\alpha$ ($\alpha > 0$) 也成立, 即对 $\alpha > 0, \alpha \neq 1$, 有

$$n^{1/2} [H_\alpha(\hat{\mathbf{p}}) - H_\alpha(\mathbf{p})] \rightarrow N \{ 0, [\alpha / (1-\alpha)]^2 \left(\sum_{i=1}^s p_i^{2\alpha-1} \right) / \left(\sum_{i=1}^s p_i^\alpha \right)^2 - 1 \},$$

即 $H_\alpha(\hat{\mathbf{p}})$ 渐近地服从正态分布

$$N \{ H_\alpha(\mathbf{p}), [\alpha / (1-\alpha)]^2 \left(\sum_{i=1}^s p_i^{2\alpha-1} \right) / \left(\sum_{i=1}^s p_i^\alpha \right)^2 - 1 \}。$$

Kempton 和 Wedderburn^[18]还研究了 Simpson 指数在种-多度分布为 Gamma 分布时的抽样性质, 并给出了其均值和方差。

Lyons 和 Hutcheson 提出了比较来自两个独立样本的多样性的方法, 但整个检验只有在对个体计数的情况下才有效^[21]。Lyons^[21]关于一类广泛的用重要值加权的多样性指数的结果对加权的 Simpson 指数 $D = 1 - \sum_{i=1}^s (w_i p_i / W_p)^2$ 也成立, 且 D 的估计量为

$\hat{D} = 1 - \sum_{i=1}^s (w_i \hat{p}_i / W_N)^2$, 其方差为:

$$\text{Var}(\hat{D}) \approx \frac{4}{N W_p^4} \sum_{i=1}^s w_i^2 p_i [w_i p_i - W_p (1 - D)]^2 \quad (6)$$

其中, 各参数及变量的意义同前, 当权重都相同时(6)式就变为(5)式。对用两个独立的样本得出的多样性指数进行比较可照 3.1 节中的方法同样进行。

Routledge^[22]以及 Heltshel 和 Forrester^[29]已将刀切法用于 Simpson 多样性指数的研究, 其最大优点就是允许样方抽样和使用连续多度指标而不是个体计数。

3.3 Brillouin 指数

Margalef 最先用 Brillouin 指数

$$H = \frac{1}{N} \log \frac{N!}{N_1! N_2! \dots N_s!}$$

来测量生态多样性^[15]。Pielou^[13,15,30]的合并样方方法(pooled quadrat method)提供了该指数的估计和抽样方差估计的方法^[14],其作法如下:

抽取一系列随机设置的样方,并按随机顺序将它们合并,再用 Brillouin 指数对合并后的样方计算多样性 H_k ,其中 k 是合并的样方数。将 H_k 对 k 作图。设该曲线变平时的样方数为 t ,曲线变平后的部分用来估计总体多样性 H_{pop} ,再计算

$$h_k = \frac{M_k H_k - M_{k-1} H_{k-1}}{M_k - M_{k-1}} \quad k = t, t+1, \dots, m$$

其中 M_k 是第 k 个合并后的样方中的个体数, m 是总的样方数。

尽管这种决定曲线开始变平时的点 t 是主观的,但只要 $\{h_k | k = t, t+1, \dots, m\}$ 不存在序列相关,就认为 t 已经正确选择了^[13,14]。于是,就可以用

$$\hat{H} = \frac{1}{n-t+1} \sum_{k=t}^m h_k$$

来估计总体的多样性^[14],并将其记为 \bar{h} ,则此估计的抽样方差为:

$$\text{Var}(\hat{H}) = \frac{1}{m(m-1)} \left(\sum_{k=t}^m h_k^2 - m \bar{h}^2 \right)$$

由于 m 个样方随机合并时,共有 $m!$ 种可能的结果。因此,不同的合并顺序得到的结果是不同的。Lloyd 等^[3,13]建议改变合并顺序,对得到的几个不同的结果取其中值作为总体多样性的最终估计^[13,14,32],Heyer 和 Berven^[33]也采用了类似的处理方法^[32]。Heltshel 和 Forrester^[32]认为 Pielou 的合并样方方法的一种自然的扩展可能是对不同的合并顺序,取其结果的平均值作为总体多样性的最终估计。

Heltshel 和 Bitz 以及 Lyons 认为 Pielou 的合并样方方法可以应用于其它多样性指数^[14,21]。这时,多度指标就可以选用诸如“重要值”之类的其它指标,而不必是个体数了。在个体计数的情况下, Lyons 和 Hutcheson 对 Shannon 指数和 Simpson 指数推导出了相继值 $\{h_k | k = t, t+1, \dots, m\}$ 的协方差^[21,34]。

刀切法也可应用于 Brillouin 指数的估计。Heltshel 和 Forrester^[29]通过模拟研究了样方抽样时,用刀切法估计 Brillouin 指数的情况。认为小样方趋于给出该指数较小的偏差和较小的方差估计。Heltshel 和 Bitz 利用 Pielou 的合并样方方法和刀切法通过模拟比较了 4 个多样性指数的行为^[32]。Heltshel 和 Forrester^[32]还证明了在一定的条件下, Pielou 的合并样方估计量等价于 Zahl^[23]的刀切估计量,并认为刀切法是估计多样性指数的一个更实用的方法。

4 均匀度指数

均匀度是指群落中各个物种的多度分布的均匀程度。常用的均匀度指数是通过三种方法构造的:

$$(1) \quad E = \frac{D - D_{\min}}{D_{\max} - D_{\min}}$$

$$(2) \quad E = \frac{D}{D_{\max}}$$

$$(3) E = \frac{D_{max} - D}{D_{max} - D_{min}}$$

这样得到的均匀度指数大都依赖于物种数目,要计算这些指数,必须先知道群落中的物种数 S ,这是在大多数生态学应用中不可能确定的。人们常常用样本中观察到的物种数 s 去代替群落中的总物种数 S 来计算均匀度指数。这样做存在两个严重的问题:一是 s 总是低估 S ,从而估计的均匀度值总是偏高;其二是样本大小的差别或者 s 的简单随机变异都会严重地影响所得的结果^[12]。

抽样性质比较清楚的均匀度指数为数很少,Troussellier 和 Legendre^[24]提出的“功能均匀度指数”(‘functional evenness’ index) E 是其中之一:

$$E = \frac{1}{C \log 0.5} \sum_{i=1}^C [p_i \log p_i + (1 - p_i) \log(1 - p_i)], 0 < p_i < 1 \quad (7)$$

其中, C 是双重特征数(the number of binary characters $\chi = 27$)。他们指出由于(7)式中的分母是常数而不是变量,因此,对 Shannon 多样性指数 H' 适合的方法对 E 也适合,这包括刀切法,通过它们可以得到 E 的标准误差。

另外,对移动数均匀度指数的抽样性质也进行过研究。移动数是 Fager^[9]提出用来度量样本的物种多样性的。其含义是将一个样本的频度分布(即将各个种按多度从大到小的顺序排列的分布)变换成一个均匀分布需要移动的步数(即所需要的工作量 W)。为了对一个样本得到 W 的唯一的值,需要先将频度按从大到小的顺序排列,令 $n_{(1)} \geq n_{(2)} \geq \dots \geq n_{(s)}$ 表示顺序值。

设样本中观察到的物种数为 s ,则 W 可用下式计算

$$W = \frac{N(s+1)}{2} - \frac{r(s-r)}{2} - \sum_{i=1}^s in_{(i)}$$

其中, $N = \sum_{i=1}^s n_i$, r 是 N 除以 s 后所得的余数。

Fager^[9]指出, W 是多样性的一个相反的测度,因为对一个完全均匀的分布它达到最小值。而对一个完全偏斜的分布它达到最大值。

$$M_s = \frac{(N-s)(s-1)}{2} - \frac{r(s-r)}{2}$$

因此, Fager 定义样本的移动数均匀度指数为:

$$T_s = 1 - \frac{W}{M_s} = 1 - \frac{N}{M_s} \left[\frac{s+1}{2} - \frac{r(s-r)}{2N} - \frac{1}{N} \sum_{i=1}^s in_{(i)} \right]$$

Lyons 和 Hutcheson^[35]定义总体的移动数均匀度指数为:

$$T = \frac{2}{S-1} \left[\sum_{i=1}^S ip_{(i)} - 1 \right]$$

其中, S 为总体中所包含的物种数, $p_{(i)}$ 是第 i 个最富有的种的相对多度($i=1, 2, \dots, S$)。实际上, T_s 可以看作是总体的移动数均匀度指数的估计量。他们认为在物种频度(即个体数)分布渐近地服从正态分布这个一般的假设之下,移动数均匀度指数 T_s 的大样本分布是近似正态的;对完全均匀的总体,可以用 beta 分布来拟合 T_s 的(小样本)分布;当总体中的物种数未知时, T_s 用来自非均匀总体的小样本所作的估计偏高。

5 结语

从前面的讨论可以看到,并不是每个多样性测度的抽样性质都很清楚,还有大量的测度其

抽样性质并不清楚,甚至根本没有研究过,如大多数均匀度指数。Smith 和 Wilson^[27]列出了十几个均匀度指数,却很少涉及其抽样性质。即使在研究过的测度中,也大多是大样本性质清楚了,而小样本性质不清楚。因此,这方面还有大量的工作要做。在现有的方法中,Pielou 的合并样方方法和刀切法是两个比较通用的方法,对大多数测度都应该适用,同时,它们也不要求对个体的随机抽样,即可以使用样方抽样。但相比之下,刀切法是一个更实用的方法^[32]。

参 考 文 献

- 1 Tong Y L. Some distribution properties of the sample species-diversity indices and their applications. *Biometrics*, 1983, **39**: 999 ~ 1008
- 2 Smith W, Grassle J F. Sampling properties of a family of diversity measures. *Biometrics*, 1977, **33**: 283 ~ 292
- 3 Lande R. Statistics and partitioning of species diversity, and similarity among multiple communities. *Oikos*, 1996, **76**(1): 5 ~ 13
- 4 Grassle J F, Smith W. A similarity measure sensitive to the contribution of rare species and its use in investigation of variation in marine benthic communities. *Oecologia*, 1976, **25**: 13 ~ 22
- 5 刘灿然, 马克平. 生物群落多样性的测度方法 V 生物群落物种数目的估计方法. *生态学报*, 1997, **17**(6): 601 ~ 610
- 6 Colwell R K, Coddington J A. Estimating terrestrial biodiversity through extrapolation. *Philos. Trans. R. Soc. Lond. B.*, 1994, **345**: 101 ~ 118
- 7 Sanders H L. Marine benthic diversity: a comparative study. *American Naturalist*, 1968, **102**: 243 ~ 282
- 8 Hulbert S H. The nonconcept of species diversity: a critique and alternative parameters. *Ecology*, 1971, **52**: 577 ~ 586
- 9 Fager E W. Diversity: a sampling study. *American Naturalist*, 1972, **106**: 293 ~ 310
- 10 Simberloff D. Properties of rarefaction diversity measurements. *American Naturalist*, 1972, **106**: 414 ~ 415
- 11 Heck K L, Belle G van, Simberloff D. Explicit calculation of rarefaction diversity measurement and the determination of sufficient sample size. *Ecology*, 1975, **56**: 1459 ~ 1461
- 12 Peet R K. The measurement of species diversity. *Annual Review of Ecology and Systematics*, 1974, **5**: 285 ~ 307
- 13 Pielou E C. *Ecological diversity*. New York: John Wiley & Sons, 1975
- 14 Magurran A E. *Ecological diversity and its measurement*. Princeton: Princeton Univ. Press, 1988
- 15 Pielou E C. 数学生态学(第二版)(卢泽愚译). 北京: 科学出版社, 1988
- 16 Hutcheson K. A test for comparing diversities based on Shannon formula. *Journal of Theoretical Biology*, 1970, **29**: 151 ~ 154
- 17 Gill C A, Joanes D N. Bayesian estimation of Shannon's index of diversity. *Biometrika*, 1979, **66**(1): 81 ~ 85
- 18 Kempton R A, Wedderburn R W M. A comparison of three measures of species diversity. *Biometrics*, 1978, **34**: 25 ~ 37
- 19 May R M. Patterns of species abundance and diversity. In: Cody M L, J M Diamond (eds.). *Ecology and Evolution of Communities*, Cambridge: Harvard University Press, 1975, 81 ~ 120
- 20 Whittaker R H. Dominance and diversity of land plant communities. *Science*, 1965, **147**: 250 ~ 260
- 21 Lyons N I. Comparing diversity indices based on counts weighted by biomass or other importance values. *American Naturalist*, 1981, **118**: 438 ~ 442
- 22 Routledge R D. Bias in estimating diversity of large uncensured communities. *Ecology*, 1980, **61**: 276 ~ 281
- 23 Zahl S. Jackknifing an index of diversity. *Ecology*, 1977, **58**: 907 ~ 913
- 24 Troussellier M, Legendre P. A functional evenness index for microbial ecology. *Microbial Ecology*, 1981, **7**: 283 ~ 296
- 25 刘灿然, 马克平, 周文能, 张新生. 生物群落多样性的测度方法 IV 刀切法和自助法在生物多样性测度研究中的应用. *生物多样性*, 1997, **5**(1): 61 ~ 66
- 26 Berger W H, Parker F L. Diversity of planktonic Foraminifera in deep sea sediments. *Science*, 1970, **168**: 1345 ~ 1347
- 27 Smith B, Wilson J B. A consumer's guide to evenness indices. *Oikos*, 1996, **76**: 70 ~ 82
- 28 Engen S. Comments on two different approaches to the analysis of species frequency data. *Biometrics*, 1977, **33**: 205 ~ 213
- 29 Heltshel J F, Forrester N E. Statistical evaluation of the jackknife estimate of diversity when using quadrat samples. *Ecology*, **66**(1): 107 ~ 111

- 30 Pielou E C. The measurement of diversity in different types of biological collections. *Journal of Theoretical Biology* , 1966 , **13** : 131 ~ 144
- 31 Lloyd M , Inger R F , King F W. On the diversity of reptile and amphibian species in a Bornean rain forest. *American Naturalist* , 1968 , **102** : 497 ~ 515
- 32 Heltshe J F , Forrester N E. Estimating diversity using quadrat sampling. *Biometrics* , 1983 , **39** : 1073 ~ 1076
- 33 Heyer W R , Berven K A. Species diversities of herpetofaunal samples from similar microhabitats at two tropical sites. *Ecology* , 1973 , **54** : 642 ~ 645
- 34 Lyons N I , Hutcheson K. A stopping rule based on Simpson's index of diversity. *Journal of Theoretical Biology* , 1979 , **77** : 497 ~ 503
- 35 Lyons N I , Hutcheson K. Distributional properties of the number of moves index of diversity. *Biometrics* , 1988 , **44** : 131 ~ 140

欢迎为《生物多样性》撰稿

欢迎订阅一九九九年期刊

《生物多样性》期刊由中国科学院生物多样性委员会主办,中国科学院植物研究所、动物研究所和微生物研究所共同承办,是中国第一个全国性、学术性和专门性的生物多样性公开出版物。于1993年10月创办。国内外公开发行。

刊登主要内容分为三类:一类,开展生物多样性基础和应用研究的学术论文、简报、技术方法和综述;二类,国家和地方的重要或重大生态工程对生物多样性保护、可持续利用的理论分析及评估;三类,介绍中国在保护和持续利用生物多样性方面所制定、在地区或世界范围内有较大影响的政策、法规等。

本刊是从事生物多样性研究和教学的科技人员、大专院校师生、科技情报信息中心和图书馆必备的文献资料;关注生物多样性保护及持续利用的国家政府有关部委的决策者及实施者、经济界人士和自然保护区管理人员阅读本刊也将受益匪浅。

《生物多样性》现为季刊,每季度中月21日出版,16开本、80页,每期定价9.50元,全年4期共38.00元。《生物多样性》期刊由科学出版社出版,自1999年起,进入邮局发行网,邮发代号为82-858,欢迎广大读者到当地邮局订阅,如漏订,可在编辑部补订,补订时需加包挂寄费,全年共43.00元。

特告:1999年起《生物多样性》将恢复每年增刊1期英文版本(与中文版本的内容不同)。本刊欢迎广大作者撰写英文稿件,需订英文刊的读者可直接与编辑部联系。

编辑部备有自创刊以来各年度的各期刊物,新订户可以向编辑部索取订单予以补购。同时还备有少量精装合订本:1993(中、英文版本)与1994年(中英文版本)合订为一册(Vol. 1~2),每册92.50元;1995年(中、英文版本)合订本(Vol. 3),每册59.00元。需邮购者请加寄刊价18%的包挂寄费。汇款收到即寄刊物和报销凭证。

编辑部地址:100093 北京香山南辛村20号中国科学院植物所院内,电话:010-62591431-6137。银行汇款:100037 北京三里河路36号,中国工商银行北京分行南礼士路支行百万庄分理处;户头:《生物多样性》编辑部,帐号:014-144587-41

生物多样性研究丛书

1 生物多样性研究丛书

- (1)《中国动植物遗传多样性》(胡志昂/张亚平主编) 35.00元/本
- (2)《保护生物学》(蒋志刚/马克平主编) 38.00元/本
- (3)《中国森林多样性及其地理分异》(陈灵芝主编) 35.00元/本
- (4)《生物多样性研究的原理与方法》(中国科学院生物多样性委员会) 27.50元/本

2 生物多样性译丛

- (1)《生物多样性译丛(三)》 41.00元/本
- (2)《生物多样性公约指南》(又名《生物多样性译丛(四)》) 30.00元/本

以上书籍对于从事生物多样性科学、生物学、生态学及环境保护研究的科技人员和高等院校师生以及自然保护区管理工作具有重要的参考价值。

对以上6本书有兴趣的读者可与《生物多样性》编辑部联系。