# 多样性指数的应用*

李冠国

（山东海洋学院）

近十多年来，在生态学文献中出现了许多关于“多样性指数（Diversity Index）”的应用和评论的文章，其主要目的是，要将生态系或生物群落组成结构方面的这一特性与其他的一些特性，如生产力、稳定性、环境类型等联系起来，或是要利用这一特性的变化来判断外来干扰（如污染等）的影响。但是，这些目的究竟得到了或者可能得到多大程度的满足，仍然是一个很难确定的问题。这次会议上，有人对所采用的两种不同的多样性指数的真实意义提出了一些疑问。这里，我想就这一问题谈一点看法。

生态系或生物群落中的组成结构，即其中有哪些种以及各个种的数量有多大的比重，与生态系中的环境因素以及物质能量转换密切相关。不同区域或不同生态系的生物组成结构是不相同的，同一区域或同一生态系的生物组成结构在不同时间也有变化。当遭受到大规模偶然事件或人为干预时，原来的生物组成结构也要受到影响。因此，生物组成结构的调查分析对于生态系的研究及其应用都有重要意义。

以往这种调查分析是用列表的方法进行的。其主要内容有：优势种的种类和优势的大小，种的数目和各个种个体数目的分配情况，特有种的有无，等等。不过，这样的分析工作比较困难，其一，它要求研究人员有较丰富的经验；其二，由于缺乏客观的定量指标，难于相互进行比较，或者容易产生意见分歧。近来发展的多样性指数的分析方法，却补足了这方面的缺点。

多样性指数是根据生物组成结构中种的数目和各个种的个体数目的分配特点而设计的一种数值指标；种数越多或各个种的个体数分配越均匀，多样性就越大。一种好的多样性指数

* 本文系1978年3月，“中国科学院环境污染和生态学学术会议”上的专题发言。

主义，成为一名光荣的共产主义战士。他严于解剖自己，严于对待亲近的学生和后辈，在关键时刻，又总是按党的政策实事求是说明情况，热情加以保护。

在海洋学界，海洋资料的保密问题长期沒有得到认真的解决。由于受极左思潮的影响，个别当权者主张一律保密，从而限制了海洋科研成果的必要交流；这些人还以“政治条件”不合格为理由限制某些海洋科学工作者的工作。1956年6月科学院召开学部大会期间，童弟周同志向竺可桢同志提到海洋所一位多年从事海浪研究的同志因家庭问题不得不改行时，他表示不解，认为这不符合周总理一再强调的重在政治表现的政策。他要童弟周和赵九章同志转告那位同志，不要背包袱，要安心工作。

1960年前后，科学院海洋所有几位科学工作者发表了几篇关于潮流大面、预报方法和海洋底质特征等文章，其中有的在发表前曾经有关部门的领导审查同意，但其后被人以严重洩密告到中央有关部门，并逐步升级，使作者或研究所的领导都感到压力很大。竺可桢同志告诉我们，关于科技资料的保密问题，聂副总理最近有指示；海洋研究成果的保密是个老问题，至今无章可循，可以参考别人的经验，订出具体的保密条例来共同遵守。不要什么都保密了事，妨碍学术交流。他还支持我们据实申辩，

不仅应当能够反映出生物群落中生物组成结构在这方面的特性，而且应当能被用来对不同区域、不同生态系或同一区域同一生态系不同时间的生物组成结构进行比较。它最好能在一定范围内不受样本量大小的影响，但对生物组成结构本身的变化则比较敏感。当然，一个好的多样性指数，不仅要有比较严密的数学基础、同时还应当有一定的生物学依据。

现有的各式各样的多样性指数很多，可以将它们分为两类作一初步评论。这两种类型恰好也包括了这次会议上，有关报告中所采用的两种指数。

## (一)

一种类型的多样性指数是，简单地从实用目的出发，并不考虑或强调有关的生物学理论基础，而主要是采用一些经典的数学方法求得一个比较客观的数值指标。这里先举出 Simpson指数作为一例加以说明。为了便于比较，我们将 Simpson (1949) 所用的符号和叙述稍加修改。

假设在一个包含无限个体的生物群落中，包括有 S 个种，各个种的个体数在总体中所占的比重分别为$P_1$，$P_2$，$P_3$，……$P_s$。如果定义$\lambda=\Sigma P^2$，则$\lambda$可以反映该群落中个体分配的集中程度。$\lambda$可以是从 1/S 到 1 的任何数值，数值越大，说明分配越集中。若群落中全部个体都属于一个种，$\lambda$ 的值为 1；若群落中各个种所占的比重相等，$\lambda$的值为1/S。$\lambda$ 值原来的意义可以被看作为，从这一群落中随机取出的两个个体恰好是属于同一个种的概率。这是众所熟知的一个简单的古典概型问题。

Simpson 接着将这一概念应用到从这样的群落所采的随机样本上。设样本中有 N 个个体，分配到各个种的个体数分别为 $n_1$，$n_2$，$n_3$，…… $n_S$（$\Sigma n=N$）。Simpson 认为，以 $l=\frac{\Sigma n(n-1)}{N(N-1)}$，则$l$是$\lambda$的一个无偏估计值。不难看出，$l$ 原来的意义是，从样本中随机取出的两个个体恰好是属于同一个种的概率。$l$ 的数值大小范围为 0到1。实际上，只有当 n全不为 0 或 1 时，$l$才可能是$\lambda$的一个无偏估计。

$\lambda$或$l$是测定群落生物组成结构的集中情况的，数值大小与多样性大小恰相反。Williams (1964) 将 $l$ 的公式改为 $\frac{N(N-1)}{\Sigma n(n-1)}$，并称它为一种“多样性指数”，数值范围从 1 到∞。这一数值的原来意义，平均来说，需要从样本中随机取多少次成对的个体才能得到属于同一个种的一对个体。

虽然Simpson 指数数值的大小决定于样本所包含的种的数目，也决定于各个种个体数的

---

不要束缚手脚。后经我们实事求是地进行答辩，终于作出不算泄密的结论。随后，竺可桢又亲自批准科学出版社为《海洋科学集刊》出版保密版，以便使那些确实需要保密的成果资料得到保密，而又能在应有范围内充分交流。

竺可桢同志十分注意争取海外科学家回祖国参加社会主义建设的工作。他对毛汉礼同志回国的支持和回国后在工作、生活、思想各方面的关怀和严格要求就是个突出的例子。

十年浩劫期间，海洋研究所是个重灾户，党委书记孙自平同志、副所长张玺教授等都被迫害致死，大批知识分子、干部以种种莫须有的罪名被非法关押审查，其中对几个科学家以“里通外国的特务分子”罪名被进行残酷批斗。为了弄到那几位科学家的所谓“罪证”，有人专程去竺可桢同志处外调。竺可桢同志当时行动也受到约束，但仍据实予以驳斥，他的这些铿锵有力的声音，在“动乱”中传来，使我们倍感温暖，至今回忆起来，仍是对我们的亲切鼓励和鞭策！

粉碎“四人帮”，迎来了科学的春天。竺可桢同志毕生关怀的我国海洋科学事业正在蓬勃发展。我们缅怀竺可桢同志，努力学习竺可桢同志坚韧不拔勇于攀登科学高峰的革命精神，为祖国的海洋科学研究和“四化”事业，作出微薄的贡献。

分配情况，但是它完全没有联系任何有关的生物学规律。它的数值不能完全排除样本量大小的影响，其受个体数较多的种的影响大，个体数稀少的种的作用则被轻视。但在一般自然生物群落中，个体数少的种的数目是比较多的。

Simpson 指数不能很好地被用于对不同样本进行比较。Fager（1972）曾经采用 SI*=1.0－SI（SI为Simpson原来的$l$），以 SI* 作为新的Simpson 指数，数值范围为 0 到 1。为了便于在不同样本间进行比较，Fager 先将从各样本所计算出的指数加以标准化，即用样本的 S 和 N 先计算出可能的最大值和最小值（假定分配最均匀和最集中情况下的两个理论值），再求出标准 SI* =（实测值－最小值）/（最大值－最小值）。不过，这样标准化以后的结果，对于样本间的比较工作，并没有多少实质性改进。

在这一类型的多样性指数中，还应当举出Shannon指数H′。这一指数在很多生态学调查（包括污染调查）中被采用，本次会议报告中也有采用这一指数的。

Shannon 指数的计算公式可以是：

$$H' = -\sum_{1}^{S} P_i \log P_i .$$

其中，S是整个生物群落中所包含的种的数目；$P_i$是第$i$ 种的个体数在全部群落总个体数中所占的比重，也就是从群落中随机抽取一个个体，这一个体属于第 $i$ 种的概率。为什么要用这样的公式来计算多样性指数呢？下面先简单地作一说明。

Shannon 指数原来是在信息论中测度信息量的。测度信息量的一个很好的方法是测度信息所消除的不肯定性。这就联系到了应用有限概率格式的问题。例如，我们要选购一件衣服，可供选择的样式有n种，都同样令人满意。那么，选中某一种样式的概率为 1/n，而选中哪一种样式的不肯定性是与这一概率大小有关系的。如果有 4 种样式，n=4，选中某一种样式的概率为1/n=1/4；如果有6种样式，n=6，选中某一种样式的概率为1/n=1/6。后一情况下的不肯定性比前一情况下要大些。显然，不肯定性是由概率决定的，可以设不肯定性=f(1/n)。

如果我们要购买的衣服还有m 种颜色供挑选，各种颜色都同样令人满意，那么，选定某一种颜色的概率为1/m，不肯定性=f(1/m)；如果我们把衣服的样式和颜色一同考虑，选中某一样式的某一颜色的衣服的概率就应当是1/n×1/m=1/nm，不肯定性=f(1/nm)。不肯定性是用来测度信息量的，信息量应当是有可加性的，因此我们要求这些函数能满足：

$$f(1/nm) = f(1/n) + f(1/m) .$$

这一要求可以通过采用对数函数来满足，即

$$\log(1/nm) = \log(1/n) + \log(1/m) ,$$

由于信息量应当是正数，而概率却是≤1的，所以在这些对数函数前面都加上负号，用－log P来测度信息量。

不难看出，Shannon 指数就是按照这样的方式建立起来的。为了按照各个种的概率加权求统计平均值，对每一个－log $P_i$都先乘以$P_i$再求和，于是得到：

$$H' = -\sum_{1}^{S} P_i \log P_i .$$

有人坚持主张这里的对数应当以 2 为底。实际上，作为多样性指数，没有必要那样做。采用自然对数不仅查表方便，而且在一些推导中有用，这里就不讨论了。以 2 为底的对数可以给出用二进位制编码来标记各个独立事件所需要的位数（信息位）。显然，这一性质在多样性指数的应用中没有意义。

Shannon 指数的公式，原来适用于群落中总个体数极大，实际上可以被当作为无限的情况。种的数目已知，$P_i$通过取容积很大的样本来估计。但实际工作中，样本量大小往往是不够的，也不能保证采到全部的种。因此，从样本中估计出的这一指数值是受样本量大小的影响的，而且总是有偏低的倾向。关于这方面的问题，Pielou (1966a, 1966b, 1975) 曾有评论可供参考。

总起来看，Shannon指数所依据的是，群落生物组成结构多样性大的，所计算出来的不肯定性也大，指数的设计也是以古典概型为基础的，缺乏直接的生物学意义。前面所列举的Simpson指数的主要缺点，在这一指数的应用中同样存在，应当引起注意。最近，Manzi等（1977）用$J'=H'/H_{max}$来反映群落结构的均匀性（$H_{max}=\log S$）；用Gleason指数来反映种的丰富程度，二者配合使用。这样的做法可能是比较好的。

## （二）

另一种类型的多样性指数是，根据大量实际观察中所看到的生物群落或生态类群中的种的组成和个体数的分配规律来设计的。我们认为，这一类型的多样性指数是比较好的，特别是Fisher的$\alpha$指数很有意义。仍然以N代表样本中个体总数，S代表种数，则$\alpha$与N和S有下列关系：

$$S=\alpha \ln(1+N/\alpha),$$

$$N=\alpha(e^{S/\alpha}-1).$$

$\alpha$指数的意义需要从它的导出过程来认识。

$\alpha$指数是Fisher与Corbet和Williams（1943），在联合发表的一组论文中提出的。当时他们所讨论的是，对一个含有很多种的混合种群所采的随机样本中，有一定个体数的种的出现频率问题。Corbet和Williams提出的实例表现出，在自然界中普遍存在的一种现象，即在一类生物所组成的混合种群中，优势种（个体数众多）只有少数，常见种（个体数中等）和稀有种（个体数很少的）的种数较多。过去曾以为这样的一种分布近似地服从于一个调和级数，

$$n_1,\ \frac{n_1}{2},\ \frac{n_1}{3},\ \cdots\cdots.$$

其中，$n_1$为样本中以1个个体出现的种数，其余各项依次为以2，3，……等个个体出现的种数。Fisher等的工作指出，用一个对数数列可以最完善地反映实际分布情况，

$$n_1,\ \frac{n_1}{2}x,\ \frac{n_1}{3}x^2,\ \cdots\cdots.$$

其中，$x$为小于1的一个常数（参看Williams，1944，1964）。Fisher的$\alpha$指数$=n_1/x$。

这里只对Fisher的推导过程提出其中的两个要点：

1．在生物学取样中，有一个大家所熟知的规律，即如果从均匀材料中（如一种浮游植物的培养稀释液中）依次取独立的、等量的样本，则各样本中的个体数是一个随机变量，它们服从泊松分布，

$$P_{(n)}=e^{-m}\frac{m^n}{n!}(n=0,1,2,\cdots\cdots)$$

其中，$n$代表样本中所观察到的个体数；$P_{(n)}$是样本中观察到$n$个个体的概率；$m$是泊松分布的唯一参数，即个体数的期望值，或者说是$n$的平均值，它与样本量大小和种群密度成正比。如果取样对象包含许多不同种，但各个种群密度相等，则从这一混合种群中取一个样本所得的各个种的个体数也服从于这一概率分布。

如果被取样的材料是不均匀的，或所取样本量大小不等，我们就必然得到一个由对应于一些不同$m$值的分布所构成的混合分布。同样地，如果被取样的对象包含密度不等的许多种群，则在一个样本中各个种出现的个体数也遵从这样的混合分布。这些情况正是我们调查分析群落生物组成结构工作中所面临的情况。Fisher推导出在混合种群的样本中具有$n$个个体的种的概率：

$$P_{(n)}=\frac{(K+n-1)!}{(K-1)!\ n!}\ \frac{P^n}{(1+P)^{k+n}},$$

$$n=0,1,2\cdots\cdots.$$

由于这一概率分布与负二项展开式有联系，因此被称为负二项分布。参数P与样本量大小成正比；K与不同种个体数的数学期望$m$的方差成反比，是表征被取样生物群固有特性的一个参量。

2．由于在自然界中一个群落或生态类群中的优势种很少，常见种和稀有种较多，因此生物组成结构总是比较分散而多样的。这样的生物学背景，在数学上就表现为K非常接近于

0。另外，在取样时，0频率无法观测到；因为从样本中无法知道沒有观察到的种的数目。根据这两种实际情况，可以令前式中的$K=0$，引进一个$\alpha$代替其中的$1/(K-1)!$，用$x$代替$P/(1+P)$，就可以得到有$n$个个体的种的期望数的表达式：

$$\frac{\alpha}{n}x^n,\ n=1,2,\ \cdots\cdots.$$

注意此式中$n$不能为0。

从这一表达式出发，经过简单的级数求和运算，便得到前面所给的$N$，$S$和$\alpha$的关系式。

## (三)

从以上讨论中可以看出，Fisher的$\alpha$指数利用了自然界中生物数量分布的一个普遍规律。这一规律早在1924和1925年就由Gleason以另一种形式报道过。Gleason在研究植物种数与样方大小的关系时，发现其中有对数函数关系，Margalef (1958) 根据这一关系设计了一种多样性指数。

$$d=(S-1)/lnN.$$

Margalef利用这一指数分析海洋浮游植物时序演替和空间非均匀性，有很好的结果。该指数应当被称作Gleason指数，但也有人称它为Margalef指数。这一指数在这次会议的报告中曾被采用。不难看出，Gleason指数与Fisher的$\alpha$指数有密切联系，当$N$和$S$值较大时，前者就是后者的近似值。

关于这两种指数所依据的分布规律，Williams(1964) 还收集了许多例证。这里需要指出一点，从已有的一些文献报道和我们的实践经验来看，$\alpha$指数特别适合于对生物群落中某几个生态类群进行生物组成结构的分析。如果从生态小生境 (niche) 的理论来探讨这一情况，可能会引导出很有意义的结果。

Fisher的$\alpha$指数不仅有上述的生物学意义，在统计上还具有一个很突出的优点，即它是一个无偏估计，可以被看作为不受样本量大小的影响。正是这一优点使我们有可能做到定量地判断不同样本所代表的生物群落或混合种群的生物组成结构的异同（参看Williams,1944)。例如：(1)如果几个样本是取自同一生物群的，平均说来，不仅它们会有同一$\alpha$指数，而且当两个或更多的样本合并起来后，其样本仍然有同一$\alpha$指数值；(2)不同生物群可能凑巧有相近似的$\alpha$指数值。但是，当它们的样本被合并起来后，其样本$\alpha$指数值比原来样本中任一个$\alpha$指数值都大些；(3)两个样本合并后的$\alpha$指数值最大可达原来两个样本$\alpha$指数值的和（两个样本中沒有共同种)，一般则处于这一最大值与原有$\alpha$指数值的中间数值，取决于样本所代表的生物群的生物组成结构的异同情况。关于$\alpha$指数在这一方面的用途还需要进一步地探讨和发展。前面提到过，Margalef (1958) 曾经用Gleason指数近似地分析海洋浮游植物的时序演替和空间非均匀性，是一很好的开端。

因为Fisher的$\alpha$指数值也是依据样本来计算的，在应用$\alpha$指数时，我们不仅计算出它的数值，而且要考虑到它的抽样误差。为了解决这一问题，Fisher在同一篇论文中曾导出了$\alpha$的方差表达式；

$$Var(\alpha)=\frac{\alpha^3\{(N+\alpha)^2 ln\frac{2N+\alpha}{N+\alpha}-\alpha N\}}{(SN+S\alpha-N\alpha)^2}$$

以上我们对Fisher的$\alpha$指数进行了比较详细的讨论，因为我们认为它是一个比较好的多样性指数。虽然$\alpha$和它的方差的计算比较麻烦，但在当今电子计算机十分普及的形势下，这一点已经不再有任何困难了。不过，必须强调指出，多样性指数的应用，象其他数学模式的应用一样，可以有助于我们的工作，但不能以之代替生物的活动规律。因此，在应用时仍然必须进行一般的生态学检验，才能得出比较可靠的结果。另外，目前生物数学方面有很大的发展，有很多新途径，如数值分类法、因子分析等，有潜在的实用价值，值得我们进一步地广泛探讨。