

生物群落多样性的测度方法

Ⅲ 与物种-多度分布模型有关的统计问题

刘灿然 马克平

(中国科学院植物研究所, 北京 100044)

周文能

(中国农业科学院, 北京 100081)

在过去的几十年中,生态学家们对物种相对多度分布进行了广泛的研究,采用的方法主要有两种,即物种重要性顺序-多度表(ranked-abundance list)和物种-多度分布(species-abundance distribution),并且提出了很多数学模型^[1,2]。第一种方法主要涉及曲线拟合问题,Wilson 对此有过系统的评述^[3];对于第二种方法,笔者曾对几个主要模型及其生态学含义作过介绍^[2]。本文拟对与物种-多度分布模型有关的统计问题进行讨论。

1 对数正态分布

Preston 1948 年第一次将对数正态分布应用于物种-多度分布的研究。他发现对于充分大的包含很多种的个体集合,当个体数在对数尺度上分组时,物种-多度分布可以很好地与正态分布相吻合^[4]。对数正态分布的概率密度函数为:

$$g(\lambda) = (\lambda\sigma\sqrt{2\pi})^{-1}\exp\{-[\log_2(\lambda/m)]^2/(2\sigma^2)\}, 0 < \lambda < \infty \quad (1.1)$$

令 $X = \log_2\lambda, \alpha = \log_2m$, 则(1.1)就变为:

$$P(X) = (\sigma\sqrt{2\pi})^{-1}\exp[-(X-\alpha)^2/(2\sigma^2)], -\infty < X < \infty \quad (1.2)$$

这就是正态分布的概率密度函数。其中, λ 为物种的个体数, σ 为 λ 经过对数变换后的标准差, α 为其平均数,即

$$\sigma^2 = \text{Var}(X) = \text{Var}(\log_2\lambda), \alpha = E(X) = E(\log_2\lambda)$$

他首先提出拟合对数正态分布,采用称为“倍程”(octaves)的方法对个体数 λ 的值进行分组。即取 $\lambda=1, 2, 4, 8, \dots$ 作为组的边界,落在边界上的一个种,例如它包含 2^x 个个体,则认为种的一半分布在倍程 $(2^{x-1}, 2^x)$ 内,而另一半分布在倍程 $(2^x, 2^{x+1})$ 内^[5],将个体数作以2为底的对数变换后,组的边界变为 $0, 1, 2, 3, 4, 5, \dots$ 。于是,原来的第 X 个倍程 $(2^{X-1}, 2^X)$ 就变成了现在的 $(X-1, X)$,倍程的中点现在就是 $X-1/2$ 。其中, $X=1, 2, 3, \dots$ 。

如果分布是非截断的,则有关的参数估计和拟合问题在普通的数理统计教科书中都有介绍。但是在物种-多度分布研究中,这些分布几乎总是左边截断的,它反映了这样一个事实,即一些稀有种在调查时漏掉了。从而,总种数 S^* 是未知的,需要估计出来^[4]。Preston 称对数正态分布曲线左边的截断为“隐线”(veil line)^[5],它是出现在个体数为1的地方^[4]。

Hald 1949 年发表了数表,借此可以得到截断对数正态分布参数的极大似然估计。为了同样的

目的, Cohen 1959 年和 1961 年发表了简化的数表, Patrick 等 1954 年还介绍过一种图形方法^[4]。Gauch 和 Chase 提出了一个变化参数方法(a variation of the parameters algorithm)^[6], 用它可以得到参数的最小二乘估计。这种方法是假设模型包含了一个对数正态型的项再加上一个随机误差, 因此, 是一个回归方法^[4]。Slocomb 等也提出了对截断对数正态分布中的参数进行极大似然估计的方法^[4]。他指出, 截断对(1.2)式的影响就是用一个新的常数 A 去代替其中的常数 $(\sigma\sqrt{2\pi})^{-1}$, 并将该函数的定义域限制为 $X > 0$ 。于是, 概率密度函数(1.2)就变为

$$P(X) = A \exp[-(X - \alpha)^2 / (2\sigma^2)], X > 0 \quad (1.3)$$

α 和 σ 的意义同上, X 是所考虑的某“倍程”到原点(即截断点)的距离, 常数 A 是 α 和 σ^2 的函数。

在实际处理数据时, 是以“隐线”作为 Y 轴, 它在 X 轴上的位置作为原点, 于是 X 的第一个值为 0.5, 这就是第一个倍程的中点, 并设 $h(X)$ 为第 $X+0.5$ 个倍程中种的频数(其中, $X=0.5, 1.5, 2.5, 3.5, \dots$), 边界点上的频数如前进行分裂。设 n 为观察的总种数与单一种(即被一个个体代表的种)数目一半之差, 则 $n = \sum h(X)$ 。由(1.3), 可知

$$A \int_0^{\infty} \exp[-(X - \alpha)^2 / (2\sigma^2)] dX = 1$$

于是

$$A^{-1} = \int_0^{\infty} \exp[-(X - \alpha)^2 / (2\sigma^2)] dX = \sigma \int_{-\alpha/\sigma}^{\infty} \exp(-y^2/2) dy \quad (1.4)$$

因此, 似然函数的对数可以写成

$$L = n \ln A - \left(\frac{1}{2\sigma^2}\right) \sum h(X)(X - \alpha)^2 \quad (1.5)$$

简单起见, 令 $b = \alpha/\sigma$, 则 L 关于 b 和 σ 的极大化是关于 α 和 σ 的极大化。因 σ 是非截断正态分布的标准差, 它必须是非负的, 因此, 必须对其估计施加非负限制。于是需要极大化的函数实际上就是

$$L = n \ln A - \left(\frac{1}{2\sigma^2}\right) \sum h(X)(X - b\sigma)^2 \quad (1.6)$$

其中,

$$A^{-1} = \sigma \int_{-b}^{\infty} \exp(-y^2/2) dy \quad (1.7)$$

限制条件为

$$\sigma \geq 0 \quad (1.8)$$

求 L 关于 b 和 σ 的一阶偏导数, 并令其分别等于 0, 就得到似然方程

$$\sum h(X)X - bn\sigma - n\sigma^2 A \exp(-b^2/2) = 0 \quad (1.9)$$

和

$$\sum h(X)X^2 - n\sigma^2 - b\sigma \sum h(X)X = 0 \quad (1.10)$$

于是, 用 Newton-Raphson 方法, 根据(1.8)的限制条件, 解方程(1.7)、(1.9)和(1.10)就可得到参数 A, b, σ (进而求出 α) 的极大似然估计。Gauch 和 Chase^[6]关于参数初值的选取方法对这些方程的求解也会有帮助。

Pielou 也给出了拟合截断对数正态分布的方法^[7], 并被 Magurran^[8]采纳。此过程包括将观察变量 r (个体数) 转换为对数(即令 $X = \log_{10} r$)。然后, 对 X 值拟合正态分布。因为是将 r 作为连续变量对待, 所以, 要用区间 $(0, \frac{1}{2}]$, $(\frac{1}{2}, 1 \frac{1}{2}]$, $(1 \frac{1}{2}, 2 \frac{1}{2}]$, \dots , $(r - \frac{1}{2}, r + \frac{1}{2}]$, \dots 来代替离散值 $0, 1, 2, \dots, r, \dots$ 。由于值 $r=0$ 从原始(离散)数据中漏掉了, 从而区间 $(0, 1/2)$ 也从数据的连续表示中漏掉了。因此要拟合的正态分布是在 $X_0 = \log_{10} 0.5 = -0.30103$ 处(左边)截断的。拟合过程如下:

(1) 令 $X_i = \log_{10} r_i$, $i = 1, 2, \dots, S$;

(2) 得到 X 的观察均值与方差, 即 $\bar{X} = \frac{1}{S} \sum_{i=1}^S X_i$, $\sigma^2 = \frac{1}{S} \sum_{i=1}^S (X_i - \bar{X})^2$, 其中, S 是观察的种数;

(3) 计算 $v = \sigma^2 / (\bar{X} - X_0)^2$, 其中 $X_0 = -0.30103$;

(4) 从 cohen(1961)的表 1 或 Magurran^[8]的附录 3 中得到对应于(3)中 v 的“辅助估计函数”(auxiliary estimation function) $\hat{\theta}$;

(5) 得到 X 的均值与和方差的估计 $\hat{\mu}_X$ 和 \hat{V}_X : $\hat{\mu}_X = \bar{X} - \hat{\theta}(\bar{X} - X_0)$, $\hat{V}_X = \sigma^2 + \hat{\theta}(\bar{X} - X_0)^2$;

(6) 得到对应于截断点 X_0 的标准化的正态变量 $Z_0 = (X_0 - \hat{\mu}_X) / \sqrt{\hat{V}_X}$

(7) 从正态分布表中查 $P_0 = P\{Z \leq Z_0\}$;

(8) 得到群落中总种数的估计: $\hat{S}^* = \frac{S}{1 - P_0}$;

(9) 计算期望频数;

(10) 对观察频数和期望频数进行适当的分组, 借助 χ^2 分布进行分布的适合性检验, 由于在整个过程中利用数据估计了两个参数 μ_X 和 V_X , 所以 χ^2 分布的自由度应取为分组数-3。

另外, Cohen(1961)还给出了 $\hat{\mu}_X$ 和 $\sqrt{\hat{V}_X}$ 的抽样方差估计

$$\hat{\text{Var}}(\hat{\mu}_X) = \frac{\mu_{11} \hat{V}_X}{S}, \quad \hat{\text{Var}}(\sqrt{\hat{V}_X}) = \frac{\mu_{22} \hat{V}_X}{S}$$

其中, μ_{11} 和 μ_{22} 是从 Cohen(1961)的表 3 中得到的对应于 $\eta (= Z)$ 的两个值^[7]。

由 Pielou^[7]中的式(3.17)可以推出, 对数正态分布的变量 r 和正态分布的变量 $X = \log_{10} r$ 的均值与方差之间的关系(1.12)式, 从而可以得到分布(1.2)中的两个参数 α 和 σ 的估计。

$$\mu_r = 10^{\mu_X + V_X(\ln 10)/2}, \quad V_r = [10^{(\ln 10)V_X} - 1]10^{2\mu_X + V_X(\ln 10)} \quad (1.12)$$

2 Poisson—对数正态分布

以前用对数正态分布拟合物种-多度分布数据时(如 Preston(1948)), 都是将观察分布作为截断的对数正态分布对待, 这种处理忽视了 Poisson 抽样可变性的影响。于是, Grundy(1951)提出用 Poisson—对数正态分布来拟合物种-多度分布数据^[9]。

假设观察的物种-多度分布是具有 Poisson 参数 λ 的复合 Poisson 变量, 而 λ 服从对数正态分布, 即假设样本中第 j 种的个体数是均值为 λ_j 的一个 Poisson 变量, 而 $\{\lambda_j \mid j = 1, 2, \dots, S^*\}$

可以认为是来自对数正态分布的 S^* 个独立观察。对数正态分布的密度函数为:

$$f(\lambda) = \frac{1}{\lambda \sqrt{2\pi V}} \exp\left[-\frac{1}{2V}(\ln \lambda - M)^2\right] \quad (2.1)$$

这等价于假设 $\ln \lambda$ 是正态分布的。于是有

$$E(\ln \lambda) = M, \quad \text{Var}(\ln \lambda) = V \quad (2.2)$$

因此, 样本中一个种可以由 r 个个体代表的概率为 Poisson—对数正态分布^[5]

$$\begin{aligned} P_r &= \frac{1}{r! \sqrt{2\pi V}} \int_0^\infty e^{-\lambda} \lambda^r \exp\left[-\frac{1}{2V}(\ln \lambda - M)^2\right] d\lambda \\ &= \frac{1}{r! \sqrt{2\pi V}} \int_0^\infty \exp\left[-\lambda + r \ln \lambda - \frac{1}{2V}(\ln \lambda - M)^2\right] \frac{d\lambda}{\lambda} \end{aligned} \quad (2.3)$$

其中 $r = 0, 1, 2, \dots$ 。Grundy(1951)提出, 令 $\ln \lambda = y$, 得到(2.3)的另一种形式

$$P_r = \frac{1}{r! \sqrt{2\pi V}} \int_{-\infty}^{\infty} \exp[-e^{-y} + ry - (y - M)^2 / (2V)] dy \quad (2.4)$$

$$= \frac{\exp(rM + r^2V/2)}{r! \sqrt{2\pi V}} \int_{-\infty}^{\infty} \exp[-e^{-y} - (y - M - rV)^2 / (2V)] dy \quad (2.5)$$

$r=0, 1, 2, \dots$ 。这个积分没有简明的表达式,但是 Bulmer^[9]推出了一个近似公式,对于较大的 r ,

$$P_r \approx \frac{1}{r \sqrt{2\pi V}} \exp[-(\ln r - M)^2 / (2V)] \{1 + [(\ln r - M)^2 / V + \ln r - M - 1] / (2rV)\} \quad (2.6)$$

对于实际中可能遇到的 M 和 V 值,当 $r \geq 10$ 时,与数值积分得到的结果相比,这个近似值的误差小于 10^{-3} 。Bulmer 还进一步给出了分布(2.4)中参数 M 和 V 的极大似然估计方法。如果数据是来自截断 Poisson-对数正态分布的大小为 S 的一个样本, n_r 为由样本中 r 个个体代表的种数,则其对数似然方程为:

$$L = \sum_{r=1}^{\infty} n_r \ln[P_r / (1 - P_0)] = \sum_{r=1}^{\infty} n_r \ln P_r - S \ln(1 - P_0) \quad (2.7)$$

令 $\theta_1 = M, \theta_2 = V$, 则

$$\frac{\partial L}{\partial \theta_i} = \sum_{r=1}^{\infty} \frac{n_r}{P_r} \frac{\partial P_r}{\partial \theta_i} + \frac{S}{1 - P_0} \frac{\partial P_0}{\partial \theta_i} \quad (2.8)$$

再微分一次,并取期望值,就得到信息矩阵的 (i, j) 元素

$$-E\left(\frac{\partial^2 L}{\partial \theta_i \partial \theta_j}\right) = \frac{S}{1 - P_0} \left[\sum_{r=1}^{\infty} \frac{1}{P_r} \frac{\partial P_r}{\partial \theta_i} \frac{\partial P_r}{\partial \theta_j} - (1 - P_0)^{-1} \frac{\partial P_0}{\partial \theta_i} \frac{\partial P_0}{\partial \theta_j} \right] \quad (2.9)$$

为了计算 $\frac{\partial P_r}{\partial \theta_i}$, 应用(2.4)中 P_r 的表达式,通过分部积分,得到

$$\frac{\partial P_r}{\partial M} = rP_r - (r + 1)P_{r+1} \quad (2.10)$$

$$\frac{\partial P_r}{\partial V} = \frac{1}{2} \frac{\partial^2 P_r}{\partial M^2} = \frac{1}{2} [r^2 P_r - (r + 1)(2r + 1)P_{r+1} + (r + 1)(r + 2)P_{r+2}] \quad (2.11)$$

(2.10)和(2.11)两式用于式(2.8)的计算。于是,解似然方程

$$\begin{cases} \frac{\partial L}{\partial M} = \sum_{r=1}^{\infty} \frac{n_r}{P_r} \frac{\partial P_r}{\partial M} + \frac{S}{1 - P_0} \frac{\partial P_0}{\partial M} = 0 \\ \frac{\partial L}{\partial V} = \sum_{r=1}^{\infty} \frac{n_r}{P_r} \frac{\partial P_r}{\partial V} + \frac{S}{1 - P_0} \frac{\partial P_0}{\partial V} = 0 \end{cases}$$

即可得到参数 M 和 V 的极大似然估计。

式(2.9)中的无限和,对前一部分(如, r 从 1 到 R)可以直接计算。余项($r > R$)可以通过下式求得:

$$P_r \approx \frac{1}{r \sqrt{2\pi V}} \exp[-(\ln r - M)^2 / (2V)]$$

于是, $\frac{\partial \ln P_r}{\partial \theta_i}$ 是 $\ln r - M$ 的线性或二次函数。我们再注意到

$$\sum_{r=R+1}^{\infty} \frac{1}{P_r} \frac{\partial P_r}{\partial \theta_i} \frac{\partial P_r}{\partial \theta_j} = \sum_{r=R+1}^{\infty} P_r \frac{\partial \ln P_r}{\partial \theta_i} \frac{\partial \ln P_r}{\partial \theta_j} \quad (2.12)$$

通过对应于 $(R + 1/2, \infty)$ 上的积分来近似从 $R + 1$ 到 ∞ 的和,并做变换 $y = \ln r - M$, 方程(2.12)可以用截断正态分布的矩近似地表达出来。

在截断分布的情况下,总的种数(包括漏查的)可以用下式来估计: $\hat{S}^* = S/(1 - \hat{P}_0)$
这个估计的抽样方差近似地表达为:

$$\text{Var}(\hat{S}^*) = \frac{\text{Var}(S)}{(1 - P_0)^2} + S^2 \text{Var}\left(\frac{1}{1 - P_0}\right) = \frac{\text{Var}(S)}{(1 - P_0)^2} + \frac{S^2}{(1 - P_0)^4} \text{Var}(\hat{P}_0)$$

其中, $\text{Var}(S) = S^* P_0 (1 - P_0) \approx S P_0$

$$\begin{aligned} \text{Var}(\hat{P}_0) &= \left(\frac{\partial P_0}{\partial M}\right)^2 \text{Var}(\hat{M}) + 2 \frac{\partial P_0}{\partial M} \frac{\partial P_0}{\partial V} \text{Cov}(\hat{M}, \hat{V}) + \left(\frac{\partial P_0}{\partial V}\right)^2 \text{Var}(\hat{V}) \\ &= P_1^2 \text{Var}(\hat{M}) + P_1(P_1 - 2P_2) \text{Cov}(\hat{M}, \hat{V}) + \frac{1}{4}(P_1 - 2P_2)^2 \text{Var}(\hat{V}) \end{aligned}$$

其中, $\text{Var}(\hat{M})$ 、 $\text{Var}(\hat{V})$ 、 $\text{Cov}(\hat{M}, \hat{V})$ 可以用信息矩阵的元素(2.9)计算出来。

Kempton 等^[10]提出了另一种进行参数的极大似然估计的方法。设 S 是总体中的总种数,能被样本中的 r 个个体代表的期望种数为 $\{E_r\}$, $r = 1, 2, \dots$ 。其中, $E_r = S^* P_r$; 对应的观察种数为 $\{n_r\}$, $r = 1, 2, \dots$ 。他的极大似然估计方法表示数据对模型的“支持”,它是模型参数的函数,其核心是选取使该“支持”达到最大的数值集作为参数估计。假设 $\{n_r\}$ 是具有均值 $\{E_r\}$ 的独立 Poisson 变量,这在样本占总体的比例很小时应该是切合实际的。函数

$$F(\theta) = \sum_{r=1}^{\infty} \{n_r \ln \left[\frac{E_r(\theta)}{n_r} \right] - [E_r(\theta) - n_r]\}$$

表达了数据 $\{n_r\}$ 对于参数集 θ 的“支持”。为了得到极大似然估计 $\hat{\theta}$, 这个函数一定要达到最大。因为观察的物种-多度分布通常具有很大的范围,考虑分组观察及它们对应的“支持”函数更可行。

这种估计方法也提供了一个模型对观察值的适合性检验方法。当模型正确时,如果任一组中的期望频数都不是太小(如 < 1), 则 $-2F(\hat{\theta})$ 近似服从 χ^2 分布, 自由度 $df =$ 分组数 - 模型中被估计的参数个数。除此之外, 还可用通常的 χ^2 适合性检验, 判断拟合效果, 其 χ^2 分布的自由度 $df =$ 分组数 - 3。

3 对数级数分布

Fisher^[11]首次将对数级数分布引入生态学中物种-多度分布关系的研究^[8,12]。这个分布可以表示为^[13,14]:

$$P_r = \theta \frac{X^r}{r}, \quad 0 < X < 1, r = 1, 2, \dots \quad (3.1)$$

这是一个种被样本中的 r 个个体代表的概率。其中, $\theta = -1/\ln(1 - X)$ 。

若 S 为群落中的总种数, 令 $\alpha = S^* \theta$, 则

$$E_r = \alpha \frac{X^r}{r}, \quad r = 1, 2, \dots \quad (3.2)$$

表示被样本中的 r 个个体代表的种数的期望值, 它对应的观察值为 n_r 。

设 S 为样本中的总种数, N 为样本中的总个体数, 则由(3.2), 有

$$E(S) = \sum_{r=1}^{\infty} E_r = \sum_{r=1}^{\infty} \alpha \frac{X^r}{r} = \alpha [-\ln(1 - X)] \quad (3.3)$$

$$E(N) = \sum_{r=1}^{\infty} r E_r = \sum_{r=1}^{\infty} r \alpha \frac{X^r}{r} = \frac{\alpha X}{1 - X} \quad (3.4)$$

以观察值 S 和 N 分别代替 $E(S)$ 和 $E(N)$, 则上述两式变为

$$S = \alpha [-\ln(1 - X)] \quad (3.3')$$

$$N = \frac{\alpha X}{1 - X} \quad (3.4')$$

进而得到

$$S/N = (1 - X)[- \ln(1 - X)]/X \quad (3.5)$$

利用迭代法解(3.5),即得到参数 X 的极大似然估计 \hat{X} ,再将 \hat{X} 代入(3.4'),即得到 α 的估计值

$$\hat{\alpha} = \frac{N(1 - \hat{X})}{\hat{X}} \quad (3.6)$$

Anscombe(1950)给出的大样本下 $\hat{\alpha}$ 的抽样方差为^[15]:

$$\text{Var}(\hat{\alpha}) = \frac{\alpha}{-\ln(1 - X)} \quad (3.7)$$

Fisher 为了方便解方程(3.3')和(3.4'),构造了一个数表,他还给出了 $\hat{\alpha}$ 的抽样方差^[11]:

$$\text{Var}(\hat{\alpha}) = \alpha^2[(N + \alpha)^2 \ln \frac{2N + \alpha}{N + \alpha} - 2N]/(SN + S\alpha - N\alpha)^2 \quad (3.8)$$

为了解方程(3.5),Williamson 和 Bretherton(1964)以及 Patil 和 Wani(1965)也都构造了数表,并且对于很大的 S ,后者还给出了 \hat{X} 的偏差(它是负的)和方差。其中的一个方差公式是:

$$\text{Var}(\hat{X}|S) = X(1 - X)^2/[S\theta(1 - \theta X)] + O(S^{-2}) \quad (3.9)$$

其中, $\theta = \frac{1}{-\ln(1 - X)}$ 。他们还给出了 $O(S^{-2})$ 项的表达式,Bowman 等做了进一步的讨论^[16]。

Anscombe(1950)给出另一个似然方程

$$S = \hat{\alpha} \ln(1 + \frac{N}{\alpha}) \quad (3.10)$$

解之,可以得到参数 α 的极大似然估计 $\hat{\alpha}$,这个估计量的大样本方差为^[17]:

$$\text{Var}(\hat{\alpha}) = \alpha/[\ln(1 + \frac{N}{\alpha}) + \alpha/N - 1]$$

Kempton 等^[10]也用了该方法,并介绍了一种对几次的调查数据求取一个共同的 α 的方法。

Patil^[13]基于大小为 S 的具有频数分布 $\{n_r | r = 1, 2, \dots, \sum n_r = S\}$ 的一个随机样本,考察了对数级数分布(3.1)的参数 X 的极大似然估计,它等同于一阶矩估计量。他给出的似然方程为:

$$\bar{r} = \mu(\hat{X}) \quad (3.11)$$

其中, $\bar{r} = \sum r n_r / S$, $\mu = \mu(X) = \theta X / (1 - X) = \frac{X}{1 - X} \frac{-1}{\ln(1 - X)}$

解(3.11),得到参数 X 的极大似然估计 \hat{X} ,其渐近方差为: $\text{Var}(\hat{X}) = X^2 / (\mu_2 S)$

其中, μ_2 是分布(3.1)的二阶中心矩,即方差,有 $\mu_2 = \mu(\frac{1}{1 - X} - \mu)$ 。同时,他还给出了另外三个估计量:

(1) 一阶矩第一室估计(First-moment first-cell estimate) X^*

$$X^* = 1 - \frac{n_1}{N^{(1)}}$$

$$V(X^*) = \frac{1}{S\mu^2} [a_{11} - 2(1 - X)a_{12} + (1 - X)^2 a_{22}]$$

(2) 二阶矩估计 \tilde{X} :

$$\tilde{X} = 1 - \frac{N^{(1)}}{N^{(2)}}$$

$$V(\tilde{X}) = \frac{1}{Sm_2^2} [b_{11} - 2(1 - X)b_{12} + (1 - X)^2 b_{22}]$$

(3) 比估计(Ratio estimate) X' :
$$X' = \frac{1}{S} \sum_{r=2}^{\infty} \left(\frac{r}{r-1}\right) n_r$$

$$\sigma^2(X') = \frac{1}{S} \left[\sum_{r=2}^{\infty} \left(\frac{r}{r-1}\right)^2 P_r - X^2 \right]$$

其中, $N^{(i)} = \sum r^i n_r, i = 1, 2; a_{11} = P_1(1 - P_1), a_{12} = P_1(1 - \mu), a_{22} = \mu_2; b_{11} = \mu_2, b_{12} = m_2 - \mu m_2,$
 $b_{22} = m_4 - m_2^2, \mu$ 和 μ_2 同前, $m_2 = \frac{\mu}{1 - X}, m_4 = \mu \frac{(1 + X)^2 + 2X}{(1 - X)^3}; V$ 表示渐近方差, σ^2 表示精确的方差。他还进一步比较了三个估计量的渐近有效性和渐近偏差, 其中 X^* 是有效的, X' 是 X 的唯一的线性无偏估计量。总起来看, 由于 X^* 简单、偏差较小、有效性较高, 而被认为是对数级数分布的参数 X 的较好的估计量。他还找到了 X^* 的渐近方差 $\text{Var}(X^* | S) \approx 2(1 - X)^2 / (S\theta)$ (当 $S \rightarrow \infty$)。当 $X \rightarrow 0$ 时, 这个方差趋近于(3.9)给出的极大似然估计量 \hat{X} 的方差; 当 $X \rightarrow 1$ 时, 也不会超过那个方差的二倍。可以用 X^* 来作为迭代求解(3.5)时对 X 的初始近似^[16]。

Birch^[14]提出了求解似然方程(3.11)的一个算法。

令
$$X = \frac{\zeta - 1}{\zeta}, \text{ 其中 } \zeta = \frac{1}{1 - X} \quad (3.12)$$

于是, (3.11)变为

$$\zeta - 1 - \bar{r} \ln \zeta = 0 \quad (3.13)$$

而(3.13)可以用 Newton 迭代法来解。如果 ζ_i 是第 i 次近似, 则第 $i+1$ 次近似为

$$\zeta_{i+1} = \frac{1 + \bar{r}(\ln \zeta_i - 1)}{1 - \bar{r}/\zeta_i}, \quad i = 1, 2, \dots$$

他推荐使用下式作为对 ζ 的第一次近似

$$\zeta_1 = 1 + [k(\bar{r} - 1) + 2] \ln \bar{r}$$

只要满足 $1 \leq k \leq 1 + \frac{2}{3}$ 就行, 简单的是令 $k = 1$, 但仅当 \bar{r} 特别大的时候这个 k 才是正确的。他对某些 \bar{r} 值给出了建议的 k 值。当 \bar{r} 接近 1 时, 使用下列 Taylor 级数来给出 ζ_1 的值可能更准确:

$$\zeta_1 = 1 + 2(\bar{r} - 1) + \frac{2}{3}(\bar{r} - 1)^2 - \frac{2}{9}(\bar{r} - 1)^3 + \frac{14}{135}(\bar{r} - 1)^4 - \dots$$

Engen^[18]比较了估计参数 α 的四种方法, 只有第三种方法得到的估计量的方差, 当 $X > 0.982$ 时, 小于 Fisher 的估计量的方差。

Patil 等(1966)讨论了参数 X 的最小方差无偏估计量, 并给出了方差估计^[16]。Wani 等^[12]进一步比较了对数级数分布参数的最小方差无偏估计量(MVUE)和极大似然估计量(ML)。指出由于涉及到的方程没有显式解, 所以两种方法都需要数表, 在这些估计量之间没有简单的准则可寻。ML 估计量是有偏的, 而 MVUE 在某些情况下是不合理的。对于大的 X 值, ML 估计量中的偏差则较小, 并且随着样本含量的增加, 这个偏差保持下降。他们还通过 Clopper-Pearson 系统提出了对数级数的参数 X 的区间估计。

Anscombe(1950)指出 $\hat{\alpha}$ 的分布而是渐近正态的, 也给出了 $\hat{\alpha}$ 的抽样方差的一个估计 $\text{Var}(\hat{\alpha}) = \frac{\hat{\alpha}}{\ln X(1 - X)}$, 因此, 可以据此确定 α 的置信区间^[7]。

Kempton 等^[10]提出的利用“支持”函数(support function)进行参数极大似然估计的方法在这里也适用, 并且可以进行适合性检验。与 Fisher 等(1943)提出的极大似然估计方法(解方程 3.5)相比, 他们更喜欢这种“支持”函数方法。

Lo 等^[19]则讨论了不变多度分布中参数的极大似然估计问题,并将其结果应用于对数级数分布。Watterson^[16]指出,由 Karlin, McGregor 和 Ewens 提出的三个模型可以导出 Fisher 对数级数分布的 5 种变形。他对各种变形的分布和参数的统计性质做了详细的讨论。

关于分布的适合性检验,最常用的方法还是 χ^2 检验。在计算分组的理论频数时,需要计算 $\sum_a^b \alpha X^r/r$, 当 $b - a$ 较大时,用 Biliss(1965)给出的公式

$$\sum_a^b E^r \approx \alpha \left\{ \sum_a^b \frac{1}{r} - (b - a + 1)(1 - X) + \frac{b^{(2)} - (a - 1)^{(2)}}{2(2!)}(1 - X)^2 - \frac{b^{(3)} - (a - 1)^{(3)}}{3(3!)}(1 - X)^3 + \dots + (-1)^m \frac{b^{(m)} + (a - 1)^{(m)}}{m(m!)}(1 - X)^m \right\}$$

计算是很方便的。其中 $b^{(m)} = b(b - 1)(b - 2)\dots(b - m + 1)$ 。当 $m = 3$ 或 4 时,就可以达到较高的精度^[7]。由于对数级数分布实质上只包含一个参数(另一个可以用它表示出来)^[7],故检验时 χ^2 分布的自由度取为分组数 - 2。

另外, Fisher^[11]还提出了一个对数级数分布的适合性检验方法。但是,此方法受到了 Holgate^[20]的质疑,他指出 $k = 0$ 不能作为零假设。

4 负二项分布

Brain(1953)首次将负二项分布引入物种-多度分布关系的研究^[18]。Pielou^[7]对此分布有过详细的描述。假设群落中总共有 S^* 个种,样本中第 j 个种的个体数是参数为 λ_j 的 Poisson 变量,不同的 λ_j 值构成了带有两个参数的 Γ -分布(即 Pearson III 型分布)的 S^* 个独立随机变量,即 λ 的分布密度函数为

$$f(\lambda) = \frac{p^{-k} \lambda^{k-1} \exp(-\lambda/p)}{\Gamma(k)}, 0 < \lambda < \infty \quad (4.1)$$

其中,正常数 p 和 k 是 Γ -分布的两个参数

于是,任何一个种可以被样本中的 r 个个体代表的概率为

$$q_r = \int_0^{\infty} \frac{\lambda^r e^{-\lambda}}{r!} f(\lambda) d\lambda = \frac{\Gamma(k+r)}{r! \Gamma(k)} \left(\frac{p}{1+p}\right)^r \left(\frac{1}{1+p}\right)^k, r = 0, 1, 2, \dots \quad (4.2)$$

这是负二项分布的通项。这意味着一个种包含有 0 个个体的概率为: $q_0 = (1+p)^{-k} > 0$ 但是“空”(empty)种是观察不到的,因此,一个观察到的种可以被样本中 r 个个体代表的概率为

$$q'_r = \frac{q_r}{1 - q_0} = \frac{\Gamma(k+r)}{r! \Gamma(k)} \left(\frac{p}{1+p}\right)^r \frac{1}{(1+p)^k - 1}, r = 1, 2, \dots \quad (4.3)$$

这就是 0-截断的负二项分布的通式。此分布的均值和方差为

$$E(r) = \frac{kp}{1 - (1+p)^{-k}}, \text{Var}(r) = (1+p+kp)E(r) - [E(r)]^2 \quad (4.4)$$

于是,分布中的两个参数 k 和 p 可以通过矩估计而得到,具体如下:

设样本中包含 N 个个体,分属于 S 个种,有 n_r 个种包含 r 个个体。于是, $\sum n_r = S$, $\sum r n_r = N$, 而 $\bar{r} = N/S$ 就是要拟合的分布的样本均值,样本二阶原点矩为 $m'_2 = \sum r^2 n_r / \sum n_r$ 。

令

$$\begin{cases} \bar{r} = E(r) & (4.5) \\ m'_2 = \text{Var}(r) + [E(r)]^2 & (4.6) \end{cases}$$

即

$$\left\{ \begin{array}{l} \frac{N}{S} = \frac{kp}{1 - (1+p)^{-k}} \end{array} \right. \quad (4.5')$$

$$\left\{ \begin{array}{l} \sum r^2 n_r / \sum n_r = \frac{kp(1+p+kp)}{1 - (1+p)^{-k}} \end{array} \right. \quad (4.6')$$

利用迭代方法解这两个方程,即可以得到参数 k 和 p 的矩估计 \hat{k} 和 \hat{p} 。

利用该截断分布还可以估计总种数 S^* ,只需让“观察的”与期望的“空”种的频数相等,即

$$S^* - S = S^*(1+p)^{-k} \quad (4.7)$$

将 k 和 p 代入式(4.7),即得到 S^* 的估计量

$$\hat{S}^* = \frac{S}{1 - (1+\hat{p})^{-\hat{k}}} \quad (4.8)$$

但遗憾的是,到目前为止, \hat{p} 、 \hat{k} 和 \hat{S}^* 的抽样方差还不知道^[7]。

此外,Pielou^[7]还介绍了负二项分布(4.3)的拟合问题。首先要计算 n_r 的期望值,即

$$E(n_r) = S^* q'_r, \quad r = 1, 2, \dots \quad (4.9)$$

在(4.3)中,令 $r=1$,得

$$q'_1 = \hat{k} \left(\frac{\hat{p}}{1+\hat{p}} \right) \frac{1}{(1+\hat{p})^{\hat{k}} - 1} \quad (4.10)$$

然后用递推关系

$$q'_{r+1} = \left(\frac{k+r}{1+r} \right) \left(\frac{p}{1+p} \right) q'_r, \quad r = 2, 3, \dots \quad (4.11)$$

对频数进行对数分组比较理想,传统的分组方法还是 Preston(1948)的“倍程”分组方法。为了计算各组的期望频数,最好的方法(Brain, 1953)是由组边界值 r , 得到 q'_r 的值,于是,以 $r=i$ 和 $r=j$ ($i < j$) 为边界的组的近似期望频数为 $S^*(j-i)(q'_i + q'_j)/2$ 。然后进行 χ^2 适合性检验,自由度为分组数-3。

5 负二项分布和对数级数分布的扩展

5.1 负二项分布的扩展

Engen^[18]对负二项分布模型做了扩展,把参数 k 的范围从 $(0, \infty)$ 扩大到 $(-1, \infty)$ 。假设 p_i 是随机抽出的一个个体属于第 i 种的概率, $i = 1, 2, \dots, S^*$ 。其中 S^* 可能是无穷大, $\sum_{i=1}^{S^*} p_i = 1$ 。扩展的负二项分布模型的基本假设是经过尺度调整后的多度(scaled abundance)在 a 和 b 之间的种数近似地为^[21]

$$\int_a^b \frac{\alpha^{k+1}}{\Gamma(k+1)} p^{k-1} e^{-\alpha p} dp \quad (5.1)$$

其中, $k > -1, \alpha > 0$ 是群落的结构特征参数,这里“尺度调整”的意思是经过调整使多度之和为1。既然种频率函数是对所有正的 p 定义的。所以必须限制参数空间,使得 $p > 1$ 的种数近似为0。

如果观察的个体数是具有均值 νP_i 的 Poisson 变量,能够用样本中 r 个个体代表的种数的期望值 E_r 近似地为

$$\alpha \omega^k \frac{\Gamma(k+r)}{\Gamma(k+1)r!} (1-\omega)^r, r = 1, 2, \dots \quad (5.2)$$

其中 $\omega = \alpha / (\alpha + \nu)$ 。

Engen^[18]评价了扩展的负二项分布模型的三种估计方法,即虚拟极大似然方法(pseudo maximum likelihood method)、虚拟矩方法(pseudo moment method)和逆矩方法(inverse moment method),然后指出,虚拟矩方法是最有效的。

$$\text{令 } S = \sum n_r, \quad N = \sum r n_r, \quad R = \frac{1}{N} \sum r(r-1)n_r, \quad m = \frac{N}{S}$$

则估计方程为

$$m e^{-R\omega \ln \omega / (1-\omega)} = 1 - \omega(R - m + 1) \quad (5.3)$$

和

$$k = R\omega / (1 - \omega) - 1 \quad (5.4)$$

这两个方程有平凡解: $\omega = \omega_1 = \frac{1}{R+1}, \quad k = 0$

可以证明 $1 - e^{-R} < \frac{R}{m}$ 是(5.3)和(5.4)存在唯一的非平凡解的充分条件,并且

$$\frac{R}{m} \neq \ln(R+1) \Rightarrow \omega \neq \omega_1, \quad k \neq 0 \quad (5.5)$$

根据(5.5),利用迭代法,从 ω_1 开始可以很容易地求出(5.3)的解 $\hat{\omega}$,再代入(5.4)即可求出 \hat{k} 。

由(5.2)可得

$$E(N) = a(1 - \omega) / \omega \quad (5.6)$$

在上式中以 N 代替 $E(N)$,并将 $\hat{\omega}$ 代入可以求出 \hat{a} 。

Engen^[18]指出,似乎在很多情况下,该分布都可很好地拟合。当 n_1 值太大而不能用对数级数描述时,扩展的负二项分布是 Preston(1948)的对数正态模型的一种替代,而对数级数分布只不过是扩展的负二项分布在 $k=0$ 时的特殊情况。如按 Cohen(1968)的推导,MacArthur 著名的“分割线段”(broken stick)模型则是 $k=1$ 时的特例。对 k 的估计自动地将群落结构纳入了著名的物种-多度分布模型之中。Engen 给出的标准差使我们通过简单的计算就可以检验对对数级数或分割线段模型的偏离^[22]。

5.2 β -分布模型与扩展的对数级数分布

对数级数模型已成功地用于描述很多种数据的频率分布,但是,样本分布常常比对数级数更偏斜。因此,为了更好地拟合这些分布,有必要对对数级数加以扩展^[21]。

5.2.1 β -分布模型

Fisher 在推导对数级数分布时,假设各个种的多度服从 Γ -分布(即 Pearson III 型分布)。对这一假设的一个自然推广就是假设总体中种的多度服从 II 类 β -分布:

$$\begin{aligned} f(x) &= \int_0^\infty \frac{a^k x^{k-1}}{\Gamma(k)} e^{-ax} \frac{b^{-q} a^{q-1}}{\Gamma(q)} e^{-a/b} da \\ &= \frac{1}{b(k,q)} \frac{b^k x^{k-1}}{(1+bx)^{k+q}}, \quad x \geq 0 \end{aligned} \quad (5.7)$$

如果样本是从总体中随机抽取的,则由样本中 r 个个体代表的期望种数为

$$E_r = \frac{S^*}{r!} \frac{b^k}{B(k,q)} \int_0^\infty \frac{x^{r+k-1} e^{-x}}{(1+bx)^{k+q}} dx, \quad r = 0, 1, \dots \quad (5.8)$$

这就是 β -分布模型。其中 S^* 为群落中的总种数, k, b, q 为参数。

与(5.8)对应的渐近级数形式为:

$$E_r = \frac{S^*}{q^k B(k,q)} \sum_{i=0}^{\infty} \frac{\Gamma(k+r+i)}{i! r!} a_i \left(1 + \frac{k}{q}\right)^i c^r \left(1 + c + \frac{k}{q}\right)^{-(k+r+i)} \quad (5.9)$$

其中 $c = \frac{1}{qb}$, a_i 由下列递推关系给出: $a_{i+1} = \frac{i}{k+q}(a_{i-1} - a_i)$

其中 $a_0 = 1, a_1 = 0$ 。当 $q \rightarrow \infty$ 时, $q^k B(k, q) \rightarrow \Gamma(k)$, (5.9) 就变成通常的负二项分布。

5.2.2 扩展的对数级数分布

在(5.8)中令 $k \rightarrow 0$, 就给出扩展的对数级数, 即由样本中的 r 个个体代表的期望种数为

$$E_r = \frac{\alpha}{r!} \int_0^{\infty} \frac{x^{r-1} e^{-x}}{(1+bx)^q} dx, \quad r = 1, 2, \dots \quad (5.10)$$

在样本中期望出现的种数和个体数分别为:

当 $u = \frac{1}{q} < 1$ 时

$$E(S) = \alpha \int_0^{\infty} \frac{1 - e^{-x}}{x(1+bx)^q} dx \quad (5.11)$$

$$E(N) = \frac{\alpha}{b(q-1)} = \frac{1}{1-u} \cdot \frac{\alpha X}{1-X}, \quad (5.12)$$

其中, $X = \frac{1}{1+bq}$

当 $u \geq 1$ 时, 期望的种数总是有限的, 期望的个体数却变成无限了。与(5.10)对应的级数形式为:

$$E_r = \frac{\alpha X^r}{r} \sum_{i=0}^{\infty} \frac{(i+r-1)!}{i!(r-1)!} a_i (1-X)^i \quad (5.13)$$

其中, a_i 由下列递推关系给出: $a_{i+1} = \frac{i}{q}(a_{i-1} - a_i)$, $a_0 = 1, a_1 = 0$

当 $b\sqrt{r} < 0.1$ 时, 与(5.10)的数值积分相比, (5.13)给出的近似值的误差小于 10^{-5} 。对于这个范围之外的小的 r , 将通过对(5.10)的数值积分来计算 E_r 。

5.2.3 参数估计

设样本中由 r 个个体代表的种数为 n_r , 还假设 $\{n_r\}$ 为具有均值 $\{E_r\}$ 的独立 Poisson 变量, 其中, E_r 由(5.8)或(5.10)给出。则对数似然函数为:

$$L(\theta) = \sum_{i=0}^{\infty} \{n_r \ln E_r(\theta) - E_r(\theta) - \ln n_r!\}, \quad (5.14)$$

参数 θ 的极大似然估计可以通过下列方程计算:

$$\frac{\partial L}{\partial \theta} = \sum_{r=1}^{\infty} \left(\frac{n_r}{E_r} - 1 \right) \frac{\partial E_r}{\partial \theta} = 0 \quad (5.15)$$

但这种方法的有效性却受到了质疑(Rao, 1971), 对于扩展的对数级数分布, 极大似然方程可通过

$$S = E(S) = \alpha \int_0^{\infty} \frac{1 - e^{-x}}{x(1+bx)^q} dx, N = E_1 + \sum_{r=1}^{\infty} (r+1)n_r \frac{E_{r+1}}{E_r} \quad (5.16)$$

$$\sum_{r=1}^{\infty} \left[\left(\frac{n_r}{E_r} - 1 \right) \int_0^{\infty} \frac{x^{r-1} e^{-x} \ln(1+bx)}{r!(1+bx)^q} dx \right] = 0 \quad (5.17)$$

从(5.14)得到。不需通过解(5.15), 而直接通过最优化程序, 寻求使(5.14)达到最大的参数。搜寻过程是在一个二维面上进行的, 受到 $\sum E_r = \sum n_r$ (由 5.16 的第一式给出) 的约束。

如果 $u = \frac{1}{q}$, 则 u 的初值可取为 0, X 的初值可通过先拟合 Fisher 的对数级数得到。对于扩展的对数级数来说, 参数的矩估计比极大似然估计简单得多, 这只需使观察和期望的种数及个体数相等[(5.11)和(5.12)中的 $E(S)$ 和 $E(N)$ 换成 S 和 N], 再加上方程

$$\sum r(r-1)n_r = \frac{\alpha X^2}{(1-X)^2(1-u)(1-2u)}, \quad u < 1/2$$

解这三个方程就可得到参数的矩估计。然而,这个方法的效率很低,除非 u 很小,当 $u \geq 1/2$ 时,该方法完全失效。

适合性检验还是用 χ^2 检验,对于扩展的对数级数分布,自由度取为分组数-3。

6 有待解决的几个问题

在物种-多度分布模型的研究中,涉及到的相当一部分统计问题已经得到解决,并成功地应用于一些生态现象的解释。然而,由于模型的复杂性,致使其中一些问题至今尚未解决;同时,由于生态学研究的要求,又提出了一些新的问题,Engen^[23]、Hughes^[24]、Pielou^[7]等对此都作过讨论,归纳起来主要有以下5个方面。

6.1 抽样假设的一般改进

以前,对物种-多度分布模型的处理几乎都是假设,对于给定的 $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_{s^*})$ (总体中 S^* 个种的多度分布), X_1, X_2, \dots, X_{s^*} (样本中的多度)是具有均值 $\lambda_1, \lambda_2, \dots, \lambda_{s^*}$ 的独立 Poisson 变量。这个假设过于简单,因为在生物抽样中,常常发现方差大于均值,因此,有必要研究更真实的抽样假设的作用。

6.2 多度的其它定义

“多度”(abundance)这一概念在生态学文献中不是唯一定义的。统计学家倾向于将一个种的多度定义为个体的密度(每单位的平均个体数)。然而没有理由不去用种的生产力(如生物量)的连续度量表示多度,因为对很多植物,计数个体是不可能的^[25]。如果将多度定义为生物量密度,则基础观测数据就不是 $n = (n_1, n_2, \dots, n_i, \dots)$ (其中 n_i 是个体数等于 i 的种数),而且参数估计过程就会更复杂。

6.3 参数估计

一些分布中参数的估计还不完善。例如,截断负二项分布和对数正态分布中,群落的总种数 S^* 的估计量的抽样方差还不知道,也不能进行区间估计,从而对估计精度的判断以及不同群落之间的比较都是不可能的^[7]。

6.4 适合性检验

如果 X_i (样本中种 i 的多度)是独立同分布的,则 n_1, n_2, \dots 的确是一个频数分布(a histogram)。于是,传统的适合性检验(如 χ^2 检验)就可以应用。Engen^[26]定义了相当大一类模型(重新排序的对称模型),其中 \mathbf{n} 是一个频数分布。除 Kempton 和 Taylor(1978)以外,还没有其他人检验过这些模型对数据的适合程度,也就没有理由认为传统的适合性检验一般来说可以应用。一些作者也将顺序观测多度(orderd observed abundances) $X^{(1)} \geq X^{(2)} \geq \dots \geq X^{(s)}$ 看作一个频数分布,而应用传统的适合性检验,这肯定是不恰当的。

Engen(1979)和 Routledge(1980)则认为,在用适合性检验(如 χ^2 检验)去评价样本数据对对数模型的适合性时,这些方法具有相当低的分辨力,尤其是当分组数很少的时候。并且,适合性检验常常对几个模型都不拒绝,究竟哪个模型是最恰当的,至今尚未定论^[24]。

6.5 多度模型与相似性

比较两个群落时,当然要比较它们的结构分布,这是以来自两个群落的样本(如 $\mathbf{n}^{(1)}$ 和 $\mathbf{n}^{(2)}$)为基础进行估计的。尽管 $\mathbf{n}^{(1)}$ 和 $\mathbf{n}^{(2)}$ 对单个群落 1 和 2 来说是唯一的可观测的变量,但是,如果我们同时观测两个群落,则它们就不再是唯一可观测的变量。定义 n_{ij} 似乎更自然,它是来自群落 1 中的

i 个个体和来自群落 2 中的 j 个个体代表的种数。现在的问题是,是否可能定义一些同时随机描述两个群落的“二维”多度模型以适用于对 n_{ij} 的研究。

参 考 文 献

- 1 May R, Patterns of species abundance and diversity. In: M T Cody, J M Diamond (eds.), *Ecology and Evolution of Community*, Cambridge MA: Harvard University Press, 1975, 81~120
- 2 马克平, 生物群落多样性的测度方法. 钱迎倩, 马克平主编, 生物多样性研究的原理与方法. 北京: 中国科学技术出版社, 1994, 141~165
- 3 Wilson J B, Methods for fitting dominance/diversity curves. *Journal of Vegetation Science*, 1991, 2: 35~46
- 4 Slocomb J, B Stauffer, K L Dickson, On fitting the truncated lognormal distribution to species-abundance data using maximum likelihood estimation. *Ecology*, 1977, 58: 693~696
- 5 Pielou E C, 1985. 卢泽愚译, 数学生态学(第二版). 北京: 科学出版社, 1988
- 6 Gauch H G Jr, G B Chase, Fitting the Gaussian curve to ecological data. *Ecology*, 1974, 55: 1377~1381
- 7 Pielou E C, *Ecological diversity*. New York: John Wiley & Sons INC., 1975
- 8 Magurran A E, *Ecological diversity and its measurement*. New Jersey: Princeton University Press, 1988
- 9 Bulmer M G, On fitting the Poisson lognormal distribution to species abundance data. *Biometrics*, 1974, 30: 101~110
- 10 Kempton R A, L R Taylor, Log-series and log-normal parameters as diversity discriminants for the Lepidoptera. *J. Anim. Ecol.*, 1974, 43: 381~399
- 11 Fisher R A, A S Corbet, C B Williams, The relation between the number of species and the number of individuals in a random samples of an animal population. *J. Anim. Ecol.*, 1943, 12: 42~58
- 12 Wani J K, H P Lo, Clopper-Pearson system of confidence intervals for the logarithmic series distribution. *Biometrics*, 1975, 31: 771~775
- 13 Patil G P, Some methods of estimation for the logarithmic distribution. *Biometrics*, 1962, 18: 68~75
- 14 Birch M W, An Algorithm for the logarithmic series Distribution, *Biometrics*, 1963, 19: 651~652
- 15 Taylor L R, R A Kempton, I P Woiwod, Diversity statistics and the log-series model. *J. Anim. Ecol.* 1976, 45: 255~272.
- 16 Watterson G A, Models for the logarithmic species abundance distributions. *Theor. Pop. Biol.*, 1974, 6: 217~250
- 17 Kempton R A, R W M Wedderburn, A comparison of three measures of species diversity. *Biometrics*, 1978, 34: 25~37
- 18 Engen S, On species frequency models. *Biometrika*, 1974, 61(2): 263~270
- 19 Lo H P, J K Wani, Maximum likelihood estimation of the parameters of the invariant abundance distributions. *Biometrics*, 1983, 39: 977~986
- 20 Holgate P, Species frequency distributions. *Biometrika*, 1969, 56(3): 651~660
- 21 Kempton R A, A generalized form of Fisher's logarithmic Series. *Biometrika*, 1975, 62(1): 29~38
- 22 Engen S, The coverage of a random sample from a biological community. *Biometrics*, 1975, 31: 201~208
- 23 Engen S, Stochastic abundance models in ecology. *Biometrics*, 1979, 35: 331~338
- 24 Hughes R G, Theories and models of species abundance. *Am. Nat.* 1986, 128(6): 879~899
- 25 Kempton R A, The structure of species abundance and measurement of diversity. *Biometrics*, 1979, 35: 307~321
- 26 Engen S, Comments on two different approaches to the analysis of species frequency data, *Biometrics*, 1977, 33: 205~213