

用麦夸方法最优拟合逻辑斯谛曲线*

王莽莽 李典谟

(中国科学院动物研究所)

摘 要

一般对非线性逻辑斯谛生长曲线的拟合是采用首先对原方程线性化以后用线性最小二乘的方法。此方法不是最优的。本文提出用麦夸方法对曲线进行拟合,并且比较了 Gause、Andrewartha、May、Pearl、Krebs、万昌秀等人提出的方法与计算结果。麦夸方法对生物实验及生态学中诸多非线性曲线的参数估计具有普遍的意义。

当种群在一个有限空间增长时,随着种群密度的上升,对有限空间资源和其他的生活必需条件的竞争必将加剧,从而影响种群的生殖率和存活率,以至降低种群的实际增长率,最后种群停止增长,有时甚至下降。因此,我们可设想在一个环境条件下,有一个允许的最大种群值 K 。通常称 K 为环境容纳量(carrying capacity)。Pearl等(1920)提出了逻辑斯谛方程(logistic equation)描述这种有限空间中的种群生长。

$$\frac{dN}{dt} = rN \left(\frac{K - N}{K} \right)$$

其中 K 为环境容纳量。 N 为 t 时刻的种群数量; r 为种群内禀增长率,方程的积分式为:

$$N = \frac{K}{1 + e^{a - rt}}$$

其中 a 为积分常数。

一、逻辑斯谛曲线拟合存在的主要问题

曲线拟合是生态学工作中的一个重要研究课题。一般的问题可叙述成下列形式:假设通过实验或观测得到 n 对数据值:

$$(x_i, y_i) \quad (i = 1, 2, 3, \dots, n)$$

而又假设 y 关于自变量 x 有下面函数关系:

$$y = f(x; K, a, r)$$

其中 K, a, r 是待定的参数。

曲线拟合即是要寻找参数 K, a, r 使函数所描述的曲线与每个数据的残差平方和 R

$$R = \sum_{i=1}^n [y_i - f(x_i, K, a, r)]^2$$

* 本文曾得到兰仲雄先生指导,特此致谢。

最小。于是 K 、 a 、 r 必须满足下述方程组:

$$\begin{cases} \frac{\partial R}{\partial K} = 0 \\ \frac{\partial R}{\partial a} = 0 \\ \frac{\partial R}{\partial r} = 0 \end{cases}$$

如果函数 f 是线性的, 这就是线性方程的最小二乘解。如果函数 f 是非线性的, 问题就略复杂一点。对于逻辑斯谛曲线, 函数有

$$N = f(t; K, a, r) = \frac{K}{1 + e^{a - rt}} \quad (1)$$

形式。一般过去用的方法都是把方程 (1) 用取对数的办法变成线性方程, 然后用线性方程去拟合参数。即把逻辑斯谛曲线化为:

$$\ln\left(\frac{K - N}{N}\right) = a - rt$$

其中 $K > N$ 。令 $y = \ln\left(\frac{K - N}{N}\right)$, 于是非线性拟合问题化为下述线性拟合问题: 求 a 、 r 使 $\hat{y} =$

$a - rt$ 和经验数据 $y = \ln\left(\frac{K - N}{N}\right)$ 的残差平方和最小。

这样拟合的结果, 实际上不是最优的。概括来说有下列两个问题:

(1) 为了要计算 $y = \ln\left(\frac{K - N}{N}\right)$ 的值, 必须先估出 K 值。文献上记载的估计 K 值的方法大体上有: 目测法、三点法、平均值法和枚举选优法 (参看万昌秀等, 1983)。前三种方法对 K 值的估计都带有强烈的主观色彩, 很难把 K 值统一在唯一的水平上。枚举选优法虽然对 K 值选取有客观的比较, 但实际上它还是用模拟的方法选优。它的缺点在于: ① 这种模拟法选优往往工作量很大; ② 这种方法选的“优”往往不是真正的最优参数, 如果最优参数值不在给定的范围内, 这种方法就不能找到最优的参数值。上面提到的方法无论哪一种在估计 a 、 r 时都必须先估 K 值, 使得 a 、 r 的估计都依赖于 K 的精度。并且对 $\frac{K - N}{N} = 0$ 的点需要舍去, 使得本来有限的的数据由于舍去部分而失去一些有用的信息。

(2) 用上面所述方法求得的参数, 只是其变量 y 的残差最小二乘解。一般说来已不是原变量 N 的残差最小二乘解 (参看 Glass, 1969; Zar, 1968; 冯康, 1979)。因此, 使用枚举选优法找到了最适的 K 值, 但用这个 K 值算得的 a 与 r 只是变换后的线性方程的最小二乘解, 并不是原逻辑斯谛方程的最小二乘解。

由上述原因, 我们提出了用麦夸方法 (Marquardt's Method) 来估计 K 、 a 、 r 的值。

二、麦夸方法的基本原理

为了表达方便,不妨把逻辑斯谛曲线写成下述形式

$$N = f(t; b_1, b_2, b_3) = \frac{b_1}{1 + e^{b_2 - b_3 t}}$$

为估计这些 b_i ($i = 1, 2, 3$), 先给一个初值 $b_i^{(0)}$, 并记初值与真值 b_i 之差为 Δ_i

$$b_i = b_i^{(0)} + \Delta_i \quad (i = 1, 2, 3)$$

于是估计未知真值 b_i 的问题就化为确定修正值 Δ_i 。把函数 f 在 $b_i^{(0)}$ 附近作台劳 (Taylor) 展开, 并略去 Δ_i 的二次及二次以上的项得:

$$f(t_k, b_1, b_2, b_3) = f_{k0} + \frac{\partial f_{k0}}{\partial b_1} \Delta_1 + \frac{\partial f_{k0}}{\partial b_2} \Delta_2 + \frac{\partial f_{k0}}{\partial b_3} \Delta_3$$

其中: $f_{k0} = f(t_k, b_1^{(0)}, b_2^{(0)}, b_3^{(0)})$

$$\left. \begin{aligned} \frac{\partial f_{k0}}{\partial b_i} &= \frac{\partial f(t_k, b_1, b_2, b_3)}{\partial b_i} \\ &\quad \begin{cases} t = t_k \\ b_1 = b_1^{(0)} \\ b_2 = b_2^{(0)} \\ b_3 = b_3^{(0)} \end{cases} \end{aligned} \right\}$$

($i = 1, 2, 3$)

当 $b_i^{(0)}$ 给定后, 它们是 t 的函数, 可算出。

那么, 残差平方和

$$R = \sum_{k=1}^n [N_k - f(t_k; b_1, b_2, b_3)]^2 \approx \sum_{k=1}^n [N_k - (f_{k0} + \frac{\partial f_{k0}}{\partial b_1} \Delta_1 + \frac{\partial f_{k0}}{\partial b_2} \Delta_2 + \frac{\partial f_{k0}}{\partial b_3} \Delta_3)]^2$$

按照 $\frac{\partial R}{\partial b_i} = 0$ ($i = 1, 2, 3$) 的条件, 可以得到 Δ_i 要满足的方程组:

$$\begin{cases} a_{11} \Delta_1 + a_{12} \Delta_2 + a_{13} \Delta_3 = a_{1N} \\ a_{21} \Delta_1 + a_{22} \Delta_2 + a_{23} \Delta_3 = a_{2N} \\ a_{31} \Delta_1 + a_{32} \Delta_2 + a_{33} \Delta_3 = a_{3N} \end{cases} \quad (2)$$

其中

$$a_{ij} = \sum_{k=1}^n \left(\frac{\partial f_{k0}}{\partial b_i} \right) \left(\frac{\partial f_{k0}}{\partial b_j} \right) \quad (i, j = 1, 2, 3) \quad (3)$$

$$a_{iN} = \sum_{k=1}^n \frac{\partial f_{k0}}{\partial b_i} (N_k - f_{k0})$$

当观测值(t_k, N_k)给定后, 并给出近似值 $b_i^{(0)}$ 后, 系数 a_i 及 a_{iN} 可按(3)式算出, 因而由方程组(2)解出 Δ_i 进而得到 b_i 值。当 $|\Delta_i|$ 值较大时, 可令当前的 b_i 值代替原来的近似值 $b_i^{(0)}$, 重复计算方程(3)得出 a_i 及 a_{iN} , 并解方程组(2)得出新的 Δ_i (进而得 b_i)。这种过程可以反复迭代直到 $|\Delta_i|$ 的值小到预先给定的值为止, 这时最后得到的 b_i 即为所求。这种算法为高斯-牛顿法。

这种方法的难点在于: 如果初值 $b_i^{(0)}$ 选得不好, 迭代过程可能发散, 即迭代后的 b_i 有可能离真值越来越远。Marquardt (1963) 修改了这方法, 提出在方程组系数矩阵的主对角线上加一个量 d , 并证明了: ①这个量 d 不影响原方程的局部最小; ②这个量 d 改变了方程组解表示的向量的长度, 当 $d \rightarrow \infty$ 时, 解向量长度趋于零; ③当这个量 d 趋于无穷时, 解向量的方向与残差平方和 R 在参数空间中的最速下降方向一致。这三点保证了方程组的解在寻找残差平方和最小值时, 方向准确, 步长适度。同时也放宽了对初值误差的要求。关于详细的麦夸算法, 有兴趣者可参阅Marquardt (1963, 1966)。

三、拟合的实施及结果

上述思想可以通过计算机程序完成。为使程序具有普及性, 我们用BASIC语言编制了此算法的程序并在 TRS-80, IBM-PC/XT 微处理机上通过运行。在程序运行时只需输入数据个数与 K 、 a 、 r 的初值。观测数据可以存在磁盘中, 其文件名应为MAR/DAT。当 K 、 a 、 r 收敛后, 打印机输出它们的值及对应的残差平方和(用 G 表示)。若输出结果不理想—— G 值过高, 则要改变 K 、 a 、 r 的输入初值并重新运算。这种情况常见于对逻辑斯谛方程中各参数的生物学意义不了解, 使初值过大或各参数比例不合适。

例如, 对 Carlson (1913) 培养酵母菌的实验数据, 用麦夸方法计算出的参数值与其他人计算的结果相比较见表1。这个结果所给的初值是 $K = 1000$ 、 $a = 6$ 、 $r = 2$ 、运算收敛后, 其参数所算出的残差平方和是最小的。

表 1 酵母菌实验数据计算结果的比较
table.1 Comparison of the values of K, a, r from saccharomycete of Carlson
(1913) experiment data using different methods

结 果 \ 计 算 者	Pearl	Krebs	笔者
	1927	1978	1984
K	665	665	662
a	4.1896	4.2	4.2758
r	0.5355	0.53	0.5475
残差平方和 G	249	392	192

对 Gause (1934) 的草履虫实验数据, 用各种不同的方法拟合的结果相比较见表2。

表 2 草履虫实验数据计算结果的比较

table 2 comparison of the values of K, a, r from Gause(1934) experiment data using different methods

计算结果	计算者及所用方法		Gause	Andrewartha May		万昌秀等	笔者
			1934 目测法	1954 平均值法	1976 目测法	1981 枚举选优法	1984 麦夸法
P. Aurelia 每 0.5 毫升的数据	一环 (I)	K		450		443	448
		a		5.041		5.0683	3.8171
		r		1.022		1.0451	0.8048
	残差平方和			13,133		12,489	8,640
	半环 (II)	K		240		241	244
		a		5.101		5.2021	4.2512
r			1.460		1.5284	1.2378	
残差平方和			2,811		2,692	2,367	
P. Caudatum 每 0.5 毫升的数据	一环 (III)	K	137	140	130	130	131
		a	4.2121	3.894	4.1589	4.0101	3.713
		r	1.0772	1.015	1.016	1.1414	1.0597
	残差平方和		2,868	3,118	3,338	2,245	2,222
	半环 (IV)	K	64	65	65	59	60
		a	3.4340	3.185	3.450	3.2086	2.819
r		0.7944	0.855	0.855	0.9379	0.821	
残差平方和		1,318	1,233	1,281	971	956	

从表 1 与表 2 中可看出所有的方法中, 唯有麦夸方法寻找出的参数其残差平方和最小。

四、讨 论

麦夸方法用于逻辑斯谛曲线的拟合同以前通常用的拟合方法比较有以下优点:

(1) 动态的估计 K 、 a 、 r 的值。避免了过去在估计过程中 K 、 a 、 r 精度上的相互依赖。使曲线参数有一个客观的标准。

(2) 这种方法估计参数的过程是求原非线性方程的残差平方和最小, 这样比起迄今为止生态学中常用的变换法拟合逻辑斯谛曲线 (即寻求变换后线性方程的残差平方和最小) 的方法来说, 确实要拟合得更好, 这才能算是原非线性曲线的最优参数估计值。计算结果证实了这一点。

(3) 原则上说, 这种方法不仅适合于逻辑斯谛曲线的拟合, 而且适合于拟合其他的非线性曲线, 因此在生物学中有广泛的应用前途。

(4) 在将此方法编成程序后, 在计算机上实现拟合十分简便, 快速。

此方法比高斯-牛顿方法对初值的要求虽然放宽了一些, 但若初值选择不好有时仍会陷入残差平方和的局部最小点, 而不能自拔。有两个途径可解决这个难点, 一个途径是可多选取几个初值点, 如果它们都收敛到同一点, 则这一点可能就是全局最小, 另一个途径是运用随机搜索的方法。

参 考 文 献

- 万昌秀等 1983 逻辑斯谛曲线的一种拟合方法. 生态学报 3 (3): 288—296.
冯康 1978 数值计算方法. 第152—160页. 国防工业出版社.
Andrewartha, H. G. and L. C. Birch 1954 The Distribution and abundance of Animals, The university of Chicago Press, P. 348—354.
Gause, G. G. 1934 The Struggle for Existence, Reprinted 1964 by Hafner, New York.
Glass, N. R. 1967 A technique for fitting nonlinear models to biological data. *Ecology* 48: 1010—1013.
Marquardt, D. W. 1963 An algorithm for least-squares estimation of nonlinear parameters. *J. Soc. Indust. Appl. Math.* 11 (2): 431—441.
Marquardt, D. W. 1966 Least-squares estimation of nonlinear parameters, IBM SHARE Library Distribution No. 309401, MLIN 2 August 1966.
Pearl, R. and L. J. Reed 1920 On the rate of growth of the population of the United States since 1790 and its mathematical representation. *Proc. Nat. Acad. Sc.* 6: 275—288.
Pearl, R. 1925 The Biology of population growth, Reprinted 1930 P. 10.
Zar, J. H. 1968 Calculation and miscalculation of the allometric equation as a model in biological data. *Bioscience* 18: 1118—1120.

FITTING LOGISITIC CURVE BY MARQUARDT'S ALGORITHM

Wang Mangmang Li Dianmo

(Institute of Zoology, Academia Sinica)

To transform the nonlinear equation to provide a mathematically equivalent linear equation is a common method for fitting logistic equation in the past. But it will not yield the minimum residuals for the untransformed equation. In this paper, Marquardt's algorithm is used for determining the values of parameters in logistic equation that will give the best least-squares fit to a set of data. The comparison of present method with the methods, used by Gause, Andrewartha, May, and Wan, are given in the paper. The result shows that Marquardt's method, which gives least residuals, fits to the data best.