

逻辑斯谛曲线的一种拟合方法

万昌秀 梁中宇

(四川大学生物系)

一、前 言

逻辑斯谛方程 (Logistic equation) 指出了有限空间和种群增长的基本规律, 是研究种群动态的重要工具。长期以来, 人们运用逻辑斯谛方程处理实验数据, 得出了种种曲线拟合的方法。然而, 这些方法都存在着一定的缺点, 使逻辑斯谛方程的应用受到某些限制。随着种群生态学的不断发展, 对实验数据的处理要求更精确、更迅速, 因此进一步探讨曲线拟合的方法就显得日渐重要。本文试图寻求一条通往最佳拟合的途径, 借助于电子计算机取得了较为满意的结果。

本文所用的数据 (见表1) 引自Gause (1934) 的草履虫实验。该实验曾被Andrewartha和Birch (1954) 称赞为“最细致、最精确”的实验 (见原文第9章), 并且多次被人们

表 1

t (天)	<i>P. Aurelia</i> 每0.5毫升中的个体数		<i>P. Caudatum</i> 每0.5毫升中的个体数	
	一环(I)	半环(II)	一环(III)	半环(IV)
0	2	2	2	2
1	6	3	6	5
2	24	29	31	22
3	75	92	46	16
4	182	173	76	39
5	264	210	115	52
6	318	210	118	54
7	373	240	140	47
8	396	—	125	50
9	443	—	137	76
10	454	249	162	69
11	420	219	124	51
12	438	255	135	57
13	492	252	133	70
14	468	270	110	53
15	400	240	113	59
16	472	249	127	57

引用过。为了叙述方便,表中的四组数据分别以<I>、<II>、<III>、<IV>代之。“一环”、“半环”是每次在培养基中所加的细菌食物量。

二、逻辑斯谛方程及其拟合概况

逻辑斯谛方程来源于人口研究,在马尔萨斯模型(Malthus model)不断修正的基础上,人口学家提出了如下的种群增长的微分方程: $\frac{dN}{dt} = rN\left(1 - \frac{N}{K}\right)$, 其积分式 $N = \frac{K}{1 + e^{-a-rt}}$ 就是著名的逻辑斯谛方程。式中 K 是种群生长的上渐近值, 又叫环境容纳量 (carrying capacity); r 为种群的内在增长率 (intrinsic growth rate); N 表示种群密度, 通常用单位体积个体数目表示; t 代表时间; a 为积分常数。 K 、 a 、 r 是三个待定参数。所谓逻辑斯谛曲线拟合的关键就是确定这三个参数的值。

现有资料所报道的拟合方法概括起来大体上可以分为目测法、三点法和平均值法三类。

目测法的要求是预先拟出几个 K 的可能值, 然后将这些 K 值分别代入 $\ln \frac{K-N}{N}$ 中, 并以此对 t 作图 (注意, 对逻辑斯谛方程的积分式两边取对数, 整理后即得 $\ln \frac{K-N}{N} = a - rt$),

“如果图上各点连线接近一条直线, 说明该 K 值适中; 若向上弯曲, 则 K 值偏高; 向下弯曲, K 值偏低” [参看 Плохинский' (1961), 306页]。郭祖超在 1965 年也曾用过此法估计 K 值^②。

这种方法的粗糙, 不可避免地带来结果的偏差。将 Gause 的数据 <III> 按此法作图, 发现所有的曲线都向上弯曲。为了清楚起见, 我们仅选三条曲线 (达到平衡后观察值的上界、下界和二者之间任一值) 为代表, 其余曲线都在这个范围内, 而且相似。从图上无法判断哪一条更接近直线 (参看图 1)。

Pearl 和 Reed (1920) 曾经利用公式

$$K = \frac{2P_1P_2P_3 - P_1^2(P_2 + P_3)}{P_1P_3 - P_2^2}$$

求 K 值。其中 P_1 、 P_2 、 P_3 分别代表三个等距离的横坐标所对应的纵坐标值, 因此称为三点法。

这个方法似乎十分简便, 但在数轴上取三个等远的点却有若干种方法, 如何取又无严格的规定, 令人无从着手。例如数据 <III> 仅 17 个观察值, 取法就有 81 种之多。取点不同, K 值各异, 无从分辨其优劣。

Andrewartha 和 Birch (1954) 对实验数据较完整的情况提出: 用达到平衡后所得观察值的平均数作为 K 的初始值。简称平均值法。但开始达到平衡的位置难于确定。如对数据 <III>, 从图上描点情况来看, 当 $t=5$ 、6、7、8 时均可看作开始达到平衡 (参看图 6)。

以上三种方法都不易确定 K 值, 而且对于所定 K 值好否也仅以主观判断而定。

K 值确定后, 将逻辑斯谛方程两边取对数可得直线方程 $\ln \frac{K-N}{N} = a - rt$ 。利用直线回归的最小二乘法代入 N 和 t 的值, 即可求出 a 和 r 。但是, a 和 r 除直接依赖于 K 外, 还与观察点

的选取个数有关。Плохинский (1961) 采取的是将计算中出现的第一个 $N \geq K$ 的点和它后面的点全部舍去, 只取前面 $N < K$ 的点, 即随着 K 的取值大小不同, 被取观察点个数也就多少不等。而 Andrewartha 和 Birch (1954) 则是“通过图形上升部分的数据确定, 但如果拟合不够好, 可以通过改变 K 或从计算 a 和 r 中除去某些点来改进。”(摘自第 9 章第 22 节, 笔者译) 取法不同, a 和 r 也不同。

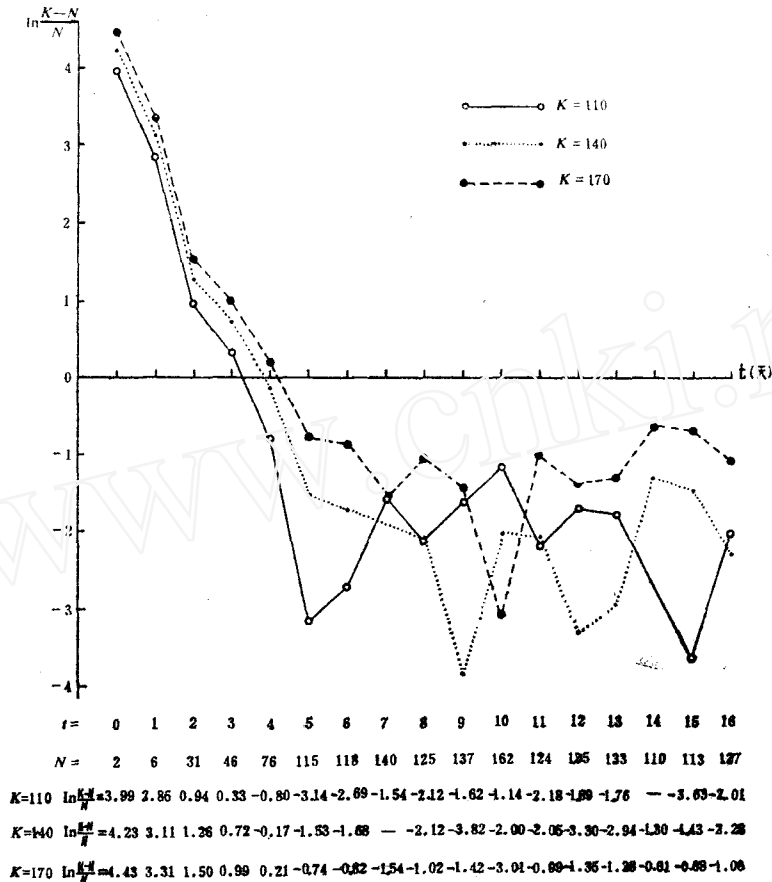


图 1 不同 K 值情况下 $\ln \frac{K-N}{N}$ 对 t 的作图

综上所述, 目前的拟合方法很不理想。由于方法带有浓厚的主观色彩, 其结果必然差异较大、难以确定。对同一组数据, 不同的人会得到不同的结果 (参看表 3)。因此为了使逻辑斯谛方程的应用更准确可靠, 为了使曲线拟合更接近客观, 进一步改进确定参数 K 、 a 、 r 的方法是完全必要的。

三、枚举选优法

在科研或生产中, 力求多快好省是我们努力的目标。寻找实现这一目标的最佳条件, 获得有关问题的最优答案, 这一切在数学上归结为“最优化问题”。解决最优化问题的方法叫做优选法。

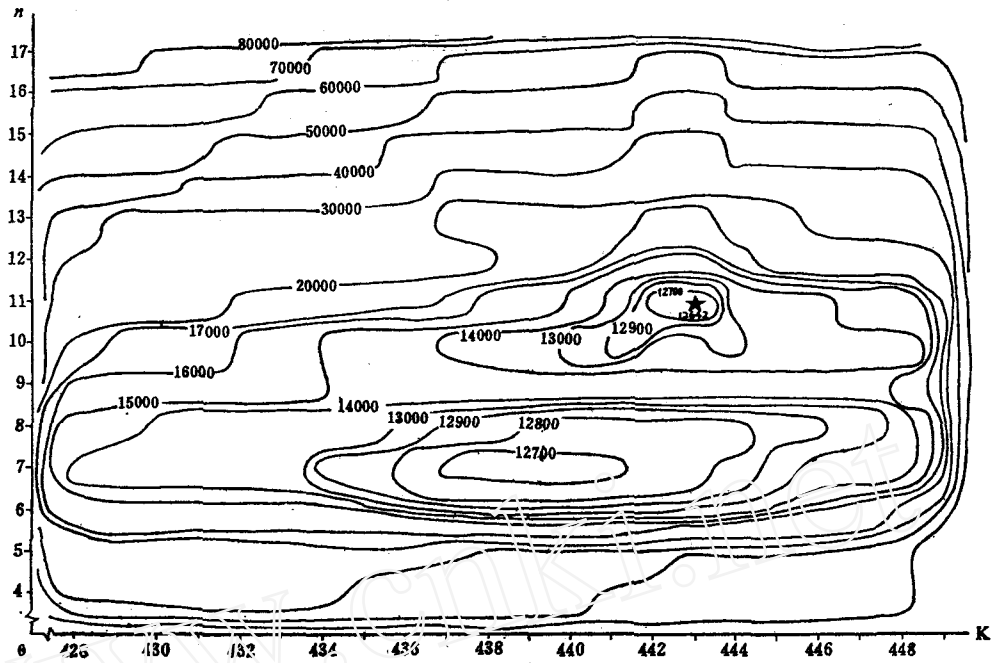


图 2 目标函数等高线图Gause数据(I)

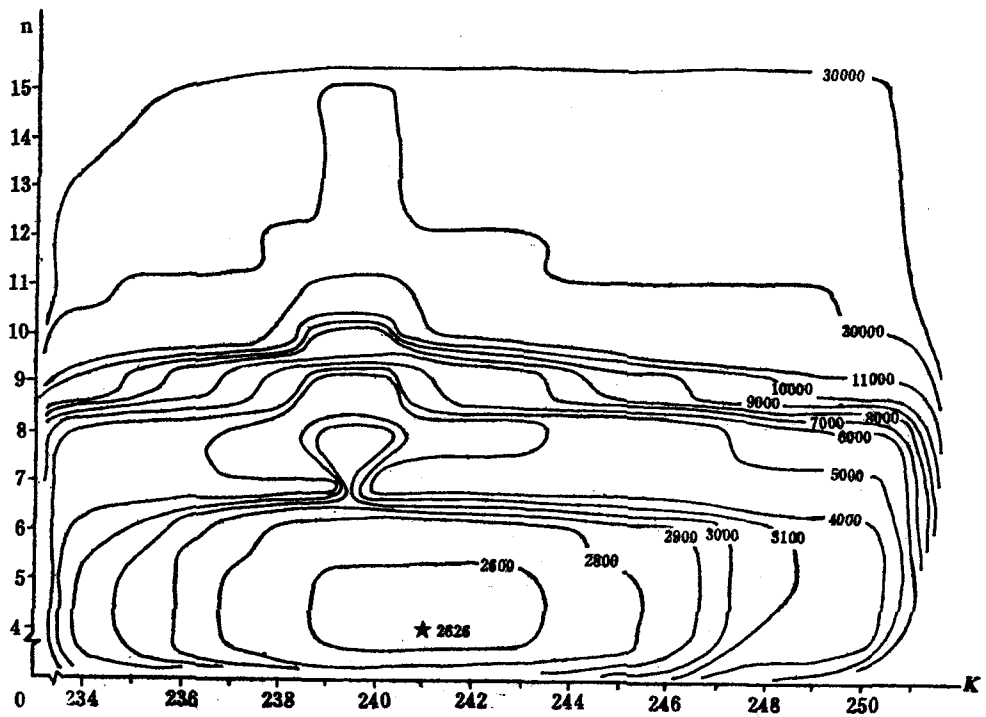


图 3 目标函数等高线图Gause数据(II)

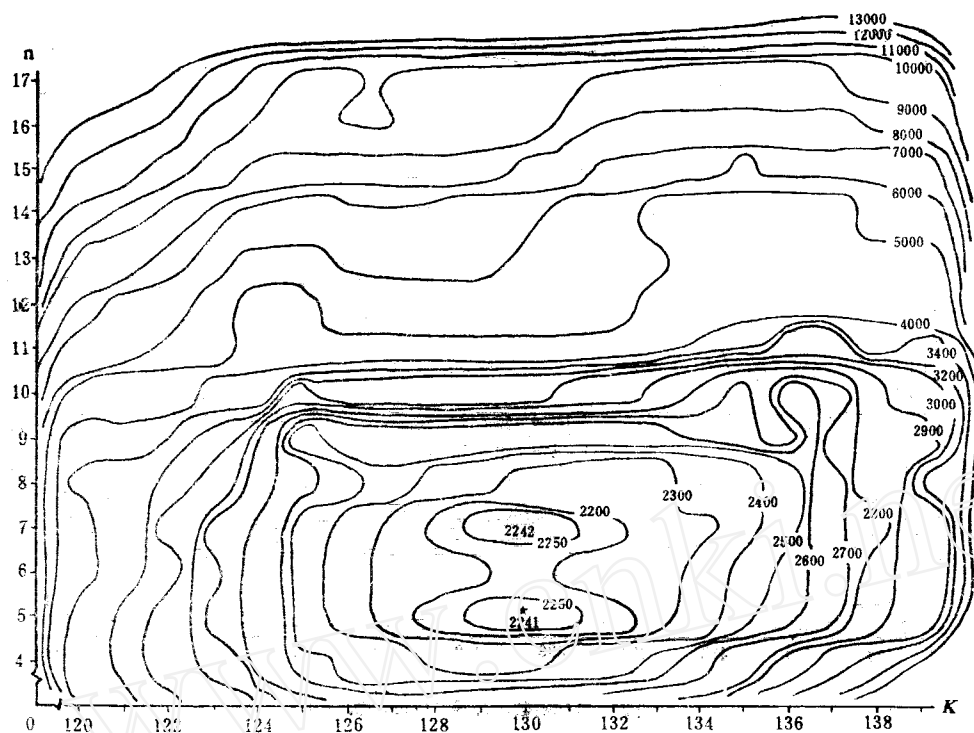


图4 目标函数等高线图Gause数据(Ⅲ)

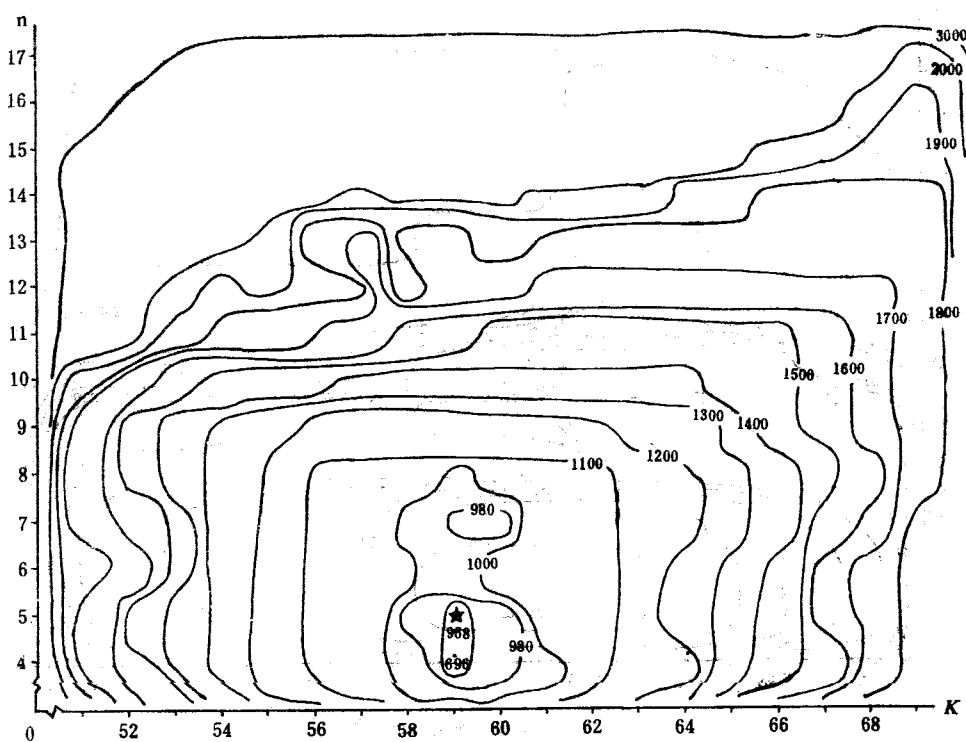


图5 目标函数等高线图Gause数据(Ⅳ)

在逻辑斯谛曲线拟合中, 我们的目标是理论曲线与观察点最大限度地接近。这个目标可以用观察值 N 和求得的理论值 $\hat{N}$ 的离差平方和 $\sum(N - \hat{N})^2$ 来表示。 $\sum(N - \hat{N})^2$ 愈小, 二者愈接近, 则拟合效果愈好。因此, 在优选中可以把 $\sum(N - \hat{N})^2$ 作为评定优劣的标准, 称为目标函数。而影响优劣的因素则是参数 K 和取点个数 n 。故此目标函数属于双因素目标函数。

令 $Z = f(x, y)$ 代表双因素目标函数, 则它表示空间的一个曲面。 x 、 y 、 z 即前面谈及的 K 、 n 、 $\sum(N - \hat{N})^2$ 。对于曲面上的每一个点 (x, y, z) , 都有一个高度 Z (相对于给定的坐标系而言) 与之相应。把曲面上具有相同高度的点联起来, 将得到数条曲线, 每条曲线上的点都有各自相同的高度, 因此称为等高线。

为了寻求一条通往 K 和 n 的最优解的途径, 将表 1 所列的四组数据的目标函数在可能的范围内逐个求出, 然后画出各自的等高线图。(参看图 2—5)

从图上可以看出, 四组数据其目标函数的峰形就有三种情况: 单峰 (图 3); 双峰且在竖直或水平方向重迭 (图 4、5); 双峰不重迭 (图 2)。对于前两种情况, 直接用“优选法”(中国科学院数学研究所, 1978) 一书中介绍的多因素坐标轮换法可以很快找出最佳点。本文目的是针对普遍情况, 故对此暂不赘述。对于第三种情况, 用这种方法是行不通的。从画等高线所得的启示, 用枚举法选优即可找到各种峰形的最佳点, 其理由如下:

(1) 有确定的目标函数: 以观察值和理论值的离差平方和 $\sum(N - \hat{N})^2$ 最小者为最优。故有明确的判断优劣的标准。

(2) 存在最佳配合点: 尽管由于各种原因目标函数的图形很复杂, 但总的趋势存在着一个或多个峰。每个峰尖应是该峰所在区域内的最好值, 即与周围各点比较峰尖的值最小。因此, 只要有峰存在, 就有峰尖即好点, 比较好点就可以找到最佳点。

(3) K 和 n 可枚举: 进行优选的因素 K 、 n 皆为离散变量。 K 通常用种群包含的个体数表示。小至酵母细胞、草履虫, 大至人口都以个数计。 n 是所取观察点个数。因此 K 和 n 都是有限个正整数, 是可以一一列举的。

(4) 保证优选的措施: 为了确保优选的顺利进行, 必须避免在 $\ln \frac{K - N}{N}$ 中出现真数 $\frac{K - N}{N} \leq 0$ 的情况发生。对于 $\frac{K - N}{N} < 0$, 用 $\frac{1}{2} \ln \left(\frac{K - N}{N} \right)^2$ 代替 $\ln \frac{K - N}{N}$ 即可避免; 对于 $\frac{K - N}{N} = 0$, 由于该情况出现的机会很少, 因此计算时可舍去该点, 而在编制电子计算机程

序时, 为了使计算不至于中断, 可以加上条件语句: “如果 $K = N$, 那末 $K := K + 0.5$ ” (实际程序中的书写格式是: $\text{eIe } KK[JI] = NN[I] \quad \text{eTe } KK[JI] := KK[JI] + 0.5;$)。这样处理对整个运算结果没有影响。

(5) 计算机的应用: 在电子计算机普遍使用的今天, 利用计算机进行枚举选优可在很短的时间内对大量数据迅速选出最佳结果 (参看表 2)。

总的来说, 枚举选优是可行的。与多因素优选中的坐标轮换法比较, 对一般的种群实验或观测数据进行逻辑斯谛曲线拟合, 该法更为简单可行, 且稳妥可靠。

四、拟合步骤及其结果

枚举选优法的步骤如下 (以数据<Ⅲ>为例):

1. 确定选优的范围

K值范围: 将数据在直角坐标系上描点 (如图6), 找出达到平衡后的最大值162和最小值110, 并分别作为K的上下界 [110, 162]¹⁾,

n的范围: 取4个—17个观察数据, 即取4个(0—3), 5个(0—4), ……17个(0—16)点²⁾。

2. 由K计算a、r

每取一个K值, 对应的n可以有14种取法 ($n=4, 5, \dots, 17$)。

例如当 $K=134$, n 取 (0—3) 时, 将 $t=0, 1, 2, 3$ 和 $N=2, 6, 31, 46$ 代入直线方程 $\frac{1}{2} \ln \left(\frac{K-N}{N} \right)^2 = a - rt$, 利用直线回归公式求出 $a=4.1492$, $r=1.2479$ 。

3. 求出理论值 $\hat{N}$

将 $K=134$, $a=4.1492$, $r=1.2479$ 代入 $\hat{N} = \frac{K}{1+e^{a-rt}}$ 中, 令 $t=0, 1, 2, \dots, 16$ 可得出的一组理论值。

4. 计算离差平方和S

将实验数据N和相应的理论值 $\hat{N}$ 代入公式得 $S = \sum_{i=0}^{16} (N - \hat{N})^2 = 2735$ 。当 $K=134$ 时, 由于n的取法有14种, 故可求出14对a、r值, 也就可以求出14个S值。以此类推, 当K取110—162时就有 $53 \times 14 = 742$ 个S值。

5. 选出最优解

逐个比较所有的S, 最后选得的结果是: 当 $K=130$, $n=5$ 时 $S=2244$ 为最小。故此时的 $K=130$, $a=4.0107$, $r=1.1414$ 为所求的最优解。

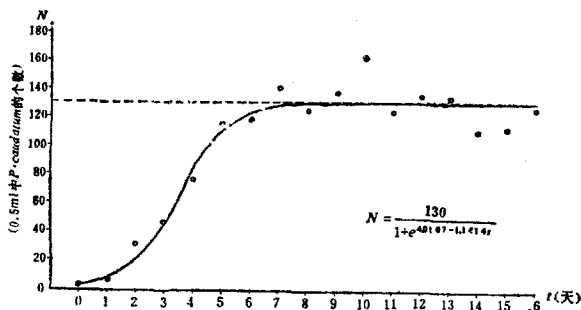


图6 对数据<Ⅲ>配合的逻辑斯谛曲线

1) 对于确定K的上下界值, 其范围大点小点关系不甚大。拟合经验可知, 最佳K值是在达到平衡后的观察点平均值附近。因此以上取法不会漏掉K的最佳值。

2) 由平面几何学可知, 确定一条直线至少两点, 确定一段弧至少三点, 因此要确定一条S形曲线不能少于四点。

6. 写出拟合方程, 绘出拟合曲线配合数据<Ⅰ>的逻辑斯谛方程是

$$N = \frac{130}{1 + e^{4.0107 - 1.1414t}}$$

枚举选优时比较的次数多、计算量大, 即使范围缩小也有几十次甚至上百次, 但使用电子计算机也可以迅速得出结果 (如表 2)。

表 2 Gause 实验数据计算情况和拟合结果。

数 据	K 值范围	n 的取法 (种)	比较次数 (次)	上机时间	拟合结果和取观察值的个数
I	400—492	14	1302	1'30"	$N = \frac{443}{1 + e^{5.0683 - 1.0451t}}$ $n = 11(0-10)$
Ⅱ	210—270	12	732	50"	$N = \frac{241}{1 + e^{3.1021 - 1.1284t}}$ $n = 4(0-3)$
Ⅲ	110—162	14	742	51"	$N = \frac{130}{1 + e^{4.0101 - 1.1414t}}$ $n = 5(0-4)$
Ⅳ	47—76	14	420	29"	$N = \frac{59}{1 + e^{3.2086 - 0.9379t}}$ $n = 5(0-4)$

* 表 2 的结果是利用 ALGOL-60 算法语言编制的程序, 使用的成都机械局设计院计算站的 TQ-16 数字电子计算机, 运算速度 12 万次/秒。该表结果也在 DJS-21 型上用 ALGOL-121 语言复核过。

五、讨 论

本文提出的枚举选优法仅仅是初步的。与常用的拟合方法比较有如下优点:

- (1) 给出了判断拟合好否的客观标准;
- (2) 适用于一般情况, 即无论目标函数是单峰还是多峰;
- (3) 由于逐个比较后选出的最佳点, 因而结果更准确, 拟合效果更理想;
- (4) 由于运用电子计算机, 使计算过程更迅速。

但该法的不足之处是, 由于进行大量的比较, 计算量很大, 因此不宜于一般的计算。还有待于从更迅速地找出最优点方面改进, 使之应用更方便。

尽管本文的重点是关于拟合方法的探讨, 但是我们对观察数值进行拟合的目的还在于找出种群的内禀增长率 r , 因此关于逻辑斯谛方程中的 r 值问题, 此处还要略加讨论。

逻辑斯谛方程首先假定 r 是不受环境约束与密度无关的, 对于同一个种群来说是恒定的。但是从表 3 来看, 数据<Ⅲ>、<Ⅳ>是同一种群 *Paramecium caudatum* 在其它条件一样仅食物更新率不同的情况下得到的数据。<Ⅲ>所加食物比<Ⅳ>每次多一倍 (<Ⅰ>与<Ⅱ>情况类似)。May (1976) 对该数据的拟合结果 (参看表 3) 得出了“ K 值与资源更新率成正比, 而不管更新率究竟是高还是低, r 值均相同”的结论。但从表 3 可知, 无论用哪一种拟合方法求得的两种情况下的 r 值都相差甚远, 谈不上“相同”。如果说前人拟合方法不理想, 使 r 值

的结果不确定,那末用枚举选优法所得的确定的结果就应该符合“ r 值均相同”的结论。但实际上相差甚远,因此这种假设是不能成立的。笔者认为,此 r 值的不恒定或者是 Gause 的实验精度未能达到模型要求的条件;或者是逻辑斯谛模型本身有待改进(或许是逻辑斯谛方程的假设在实验条件下不易满足)。欲得到近似恒定的 r 值,还可以从模型本身或实验设计上进一步考虑和探讨。

表 3

数 据	结 果	目 测 法 Gause (1934)	平 均 值 法 Andrewartha (1954)	目 测 法 May (1976)	枚举选优法 电子计算机 (1981)
<Ⅲ>*	K	137	140	130	130
	r	1.0772(或1.2444)	1.015	1.016	1.1427
<Ⅳ>	K	64	65	65	59
	r	0.7944	0.855	0.855	0.8926

* Gause 的<Ⅲ>因缺 r 值,根据 $K=137$ 分别取 $n=8$ 和 $n=4$ 可求得 $r=1.0772$ 和 1.2444 。Kostitzin (1939)的资料仅有拟合曲线而无数据。

参 考 文 献

- 中国科学院数学研究所运筹室优选小组 1978 优选法。45, 126, 144页。
 郭祖超 1965 医用数理统计方法。452—458页。
 Andrewartha, H. G. and L. C. Birch 1954 The Distribution and abundance of Animals. Chicago Illinois. 347—355.
 Gause, G. F. 1934 The struggle for existence. Baltimore 90—113, 145 p.
 Kostitzin, V. A. 1939 Mathematical Biology. London Harrap. (据摘录, 未见原文)
 May, R. M. 1976. Theoretical Ecology. Oxford. 孙儒泳译 理论生态学, 科学出版社, 1980。99—100页。
 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. Washington.
 Плохинский, Н. А. 1961. Биометрия. Новосибирск. 306—308.

A METHOD FITTING THE LOGISTIC CURVE

WAN CHANGXIU LIANG ZHONGYU

(University of Sichuan)

The common methods to fit the Logistic Curve have some disadvantages that the obtained result is indeterminate and variant, and there is no objective standard to judge it. This paper provides an optimization method of enumeration, which is well reflected in our determination of the parameters K , a and r . The procedure is that we first utilize the digital computer to determine the parameters K , a and r , and then compare one after another by the objective standard of search as a result of our obtaining the more exact values of these parameters within a very short time. The result shows that this method fit the Logistic Curve generally.