

几种常用的多阶段增长曲线的拟合方法

袁志发 颀科让

(西北农业大学基础课部)

摘 要

由于外界条件的变化及内部生长机理的控制等原因,很多生长现象具有阶段性。本文就阶梯型、折线型和多阶段Logistic及Misherlish等形式的阶段增长模型进行了拟合方法的研究。

关键词 多阶段增长曲线;阶梯型增长模型;折线型增长模型;多阶段Logistic增长模型;多阶段Misherlish增长模型

由于外界条件的变化及内部增长机理的控制等原因,很多增长现象都具有阶段性。本文的目的在于给出几种常用的不同形式的多阶段增长曲线的拟合方法。

一、阶梯型增长曲线的拟合

阶梯型增长曲线的方程为:

$$f(t) = \begin{cases} A_1, a_0 \leq t < a_1; \\ A_2, a_1 \leq t < a_2; \\ \dots\dots\dots \\ A_K, a_{K-1} \leq t \leq a_K; \end{cases} \quad (1)$$

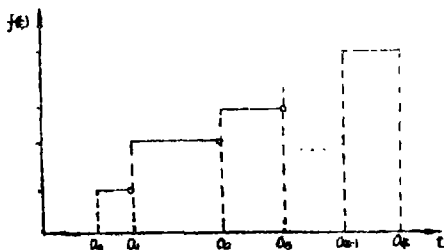


图1

其图象如图1所示。

假设我们所观察到的阶梯增长序列为

$$X_1, X_2, \dots, X_n \quad (2)$$

$f(t)$ 的拟合步骤如下面所述:

用方差分析划分阶段

1. 对序列(2)进行粗略分析,大致分为K个子序列:

本文于1985年1月24日收到。

$$\begin{aligned}
 &X_1, X_2, \dots, X_{n_1}, \\
 &X_{n_1+1}, X_{n_1+2}, \dots, X_{n_1+n_2}, \\
 &\dots \\
 &X_{n_1+n_2+\dots+n_{k-1}+1}, X_{n_1+n_2+\dots+n_{k-1}+2}, \dots, \\
 &X_{n_1+n_2+\dots+n_{k-1}+n_k} = X_n
 \end{aligned} \quad (3)$$

序列(2)的总和为 $T = \sum_{t=1}^n X_t$, $R = \sum_{t=1}^n X_t^2$, $C = T^2/n$,

各子序列的和分别为 $T_1, T_2, \dots, T_K$, $Q = \sum_{i=1}^K T_i^2/n_i$ 。则

$$\begin{cases} MS_K = \frac{1}{K-1} (Q - C) \\ MSe = \frac{1}{n-K} (R - Q) \end{cases} \quad (4)$$

MS_K 称为阶段间均方, MSe 称为误差均方。若方差比

$$\frac{MS_K}{MSe} > F_{\alpha}(k-1, n-k) \quad (5)$$

则可认为在水平 $K\alpha$ 之下, 序列(2)分为 K 个阶段是显著的。

2. 找出相临两阶段的最合适的分点。由于 C 和 R 与 K 无关, 故第 i 阶段和第 $i+1$ 阶段最合适的分介点应是 $\frac{T_i^2}{n_i} + \frac{T_{i+1}^2}{n_{i+1}}$ 最大。利用这个原则可逐阶段找出最合适的分介点。

3. 设最合适的分介点为 $X_{n_1}, X_{n_1+n_2}, \dots, X_{n_1+n_2+\dots+n_{k-1}}$ 则 $\hat{a}_1 = n_1$, $\hat{a}_2 = n_1 + n_2, \dots, \hat{a}_{k-1} = n_1 + n_2 + \dots + n_{k-1}$, $\hat{A}_1 = \frac{T_1}{n_1} = \bar{X}_1, \hat{A}_2 = T_2/n_2 = \bar{X}_2, \dots, \hat{A}_K = T_K/n_K = \bar{X}_K$, 误差 $\hat{\sigma}^2 = MSe$ 。

[例1] 利用表1数据, 拟合阶梯函数 $f(t)$ 。

表 1

	0	1	2	3	4	5	6	7	8	9
1900										18.4
1910	18.4	18.5	17.7	17.9	19.5	17.8	18.0	19.1	18.2	19.6
1920	19.5	18.9	18.8	18.5	19.6	20.2	20.1	18.9	19.0	18.9
1930	19.7	18.7	19.5	19.7	19.5	19.2	19.1	19.2	18.8	19.5
1940	18.7	19.0	19.2	19.5	18.8	18.6	19.7	18.8	19.4	19.6
1950	19.3	19.2	19.6	18.5	18.0	19.3	18.8	17.7	19.1	19.1
1960	18.4	19.2	19.0	19.5	18.8	18.8	18.8	19.6	18.8	17.7
1970	19.4	18.3	17.9	18.9	18.9	19.8	18.2	19.5		

假设我们观察到的折线增长序列为

$$\begin{array}{c|c} t & t_1, t_2, \dots, t_n \\ \hline X & X_1, X_2, \dots, X_n \end{array} \quad (7)$$

$f(t)$ 的拟合步骤如下所述。

1、由 x_1 起顺序增多观察点求相关系数 r_1 , 直到 $(1 - r_1^2) / (n_1 - 2) = S_{r_1}^2$ 最小为止 (n_1 是用的观察点数), 则得到了第一段折线, 其方程为前 n_1 个点所拟合的直线方程。

由第 $(n_1 + 1)$ 个点开始顺序增点数求相关系数 r_2 , 直到 $(1 - r_2^2) / (n_2 - 2) = S_{r_2}^2$ 最小为止 (n_2 是 r_2 所用的点数), 则得到了第二段折线, 其方程为从第 $n_1 + 1$ 个点开始的 n_2 个观察点所拟合的直线方程。仿此可得到 $r_3, r_4, \dots, r_k$ 及其所对应的直线方程。由 r_i 可检验第 i 段折线的显著性

$$S_{r_i} = \sqrt{(1 - r_i^2) / (n_i - 2)}.$$

2、设 $Z_i = \ln \left(\frac{1+r}{1-r} \right)^{1/2}$, 将 r_i 及 r_{i+1} 代入求得 Z_i 与 Z_{i+1} , 则可用 t 分布检验 r_i 及 r_{i+1} 的差异显著性, 自由度为 $(n_1 + n_2 - 6)$,

$$t = (Z_i - Z_{i+1}) / \sqrt{\frac{1}{n_i - 3} + \frac{2}{n_{i+1} - 3}}.$$

3、由 $a_i + b_i t = a_{i+1} + b_{i+1} t$ 可解出 $\hat{c}_i$ 。

〔例2〕某地花生油价格如表4所示, 试拟合折线型增长函数。

表4 某地花生油批发价格

月 年	1	2	3	4	5	6	7	8	9	10	11	12
1953							525	510	465	460	515	560
1954	615	605	595	570	530	520	520	480	440	480	540	570
1955	700	740	720	715	710	690	660	640	635	780	800	835
1956	875	900	895	840	810	765	750	685	680	770	785	800

①第一段折线的拟合

用1953年7, 8, 9三个月数据得 $r_1 = -0.92201$, $(1 - r_1^2) / (3 - 2) = 0.14988$;

用1953年7, 8, 9, 10四个月数据得 $r_1 = -0.74753$, $(1 - r_1^2) / (4 - 2) = 0.22859$;

用1953年7, 8, 9, 10, 11五个月数据得 $r_1 = -0.23801$, $(1 - r_1^2) / (5 - 2) = 0.31445$,

故第一段折线只能由7, 8, 9三个月数据拟合, 其方程为:

$$X = 923.3 - 55t, \quad r_1 = -0.92201$$

②第二段折线的拟合

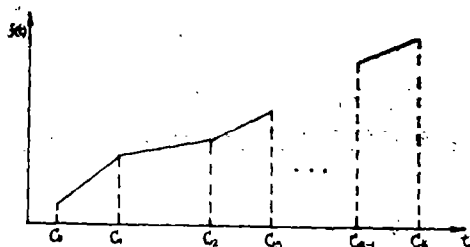


图2

用1953年10, 11, 12及1954年1月四个月数据拟合, $r_2 = 0.9992$,

$$(1 - r_2^2) / (4 - 2) = 0.00077.$$

再加上1954年2月数据, 得 $r_2 = 0.95521$, $(1 - r_2^2) / (5 - 2) = 0.02919$ 。故第二段折线只能用前面四个数据拟合, 其方程为:

$$X = -49 + 51t, \quad r_2 = 0.9992^{**}$$

③第三段折线的拟合

用1954年2—9月8个数据得 $r_3 = -0.97866$, $(1 - r_3^2) / (8 - 2) = 0.00704$; 用1954年2—10月9个数据得 $r_3 = -0.94973$, $(1 - r_3^2) / (9 - 2) = 0.01400$; 故第二段折线只能用1954年2—9月数据拟合, 其方程为:

$$X = 926.25 - 22.5t, \quad r_3 = -0.97866^{**}$$

④第四段折线的拟合

用1954年10—12月3个数据得 $r_4 = 0.98198$, $(1 - r_4^2) / (3 - 2) = 0.03571$; 用1954年10—12月及1955年1月4个数据得 $r_4 = 0.95917$, $(1 - r_4^2) / (4 - 2) = 0.04$; 再加上1955年2月数据得 $r_4 = 0.97824$, $(1 - r_4^2) / (5 - 2) = 0.01434$; 再加上1955年3月数据得 $r_4 = 0.94849$, $(1 - r_4^2) / (6 - 2) = 0.02509$ 。故第四段折线只能用5个数据来拟合, 其方程为:

$$X = -1026 + 68t, \quad r_4 = 0.97824^{**}$$

⑤第五段折线的拟合

用1955年3—8月6个数据得 $r_5 = -0.95855$, $(1 - r_5^2) / (6 - 2) = 0.0203$; 再加上9月数据得 $r_5 = -0.9713$, $(1 - r_5^2) / (7 - 2) = 0.01128$; 再加上10月份数据得 $r_5 = -0.13279$, $(1 - r_5^2) / (8 - 2) = 0.16373$ 。故第五段折线只能用3—9月的7个数据拟合, 其方程为:

$$X = 1168.9 - 16.25t$$

$$r_5 = -0.97139^{**}$$

⑥第六段折线的拟合

用1955年10—12月及1956年1—2月5个数据得 $r_6 = 0.99463$, $(1 - r_6^2) / (5 - 2) = 0.00357$; 再加上3月份数据得 $r_6 = 0.96901$, $(1 - r_6^2) / (6 - 2) = 0.01525$ 。故第六段折线用前5个数据拟合, 其方程为:

$$X = -296 + 31.5t, \quad r_6 = 0.99463^{**}$$

⑦第七段折线的拟合

用1956年3—8月6个数据得 $r_7 = -0.98977$, $(1 - r_7^2) / (6 - 2) = 0.00509$; 加上9月份数据得 $r_7 = -0.98791$, $(1 - r_7^2) / (7 - 2) = 0.00481$; 再加上10月份数据得 $r_7 = -0.82029$, $(1 - r_7^2) / (8 - 2) = 0.05452$ 。故第七段折线应用1956年3—9月

的七个数据拟合,其方程为

$$X = 2297.5 - 36.25t \quad r_7 = -0.98791. **$$

⑧第八段折线的拟合

用1956年10—12月三个数据得 $r_8 = 1$, $(1 - r_8)/(3 - 2) = 0$ 。故第八条折线方程为:

$$X = 80 + 15t, \quad r_8 = 1 **$$

综上所述,所拟合的折线型增长曲线为:

$$f(t) = \begin{cases} 923.3 - 55t, & 7 \leq t \leq 9, \quad (1953 \text{ 年 } 7-9 \text{ 月}) \\ -19 + 51t, & 10 \leq t \leq 13, \quad (1953 \text{ 年 } 10 \text{ 月} - 1954 \text{ 年 } 1 \text{ 月}) \\ 926.25 - 22.5t, & 14 \leq t \leq 21, \quad (1954 \text{ 年 } 2-9 \text{ 月}) \\ -1026 + 68t, & 22 \leq t \leq 26, \quad (1954 \text{ 年 } 10 \text{ 月} - 1955 \text{ 年 } 2 \text{ 月}) \\ 1168.9 - 16.25t, & 27 \leq t \leq 33, \quad (1955 \text{ 年 } 3 \text{ 月} - 9 \text{ 月}) \\ -296 + 31.5t, & 34 \leq t \leq 38, \quad (1955 \text{ 年 } 10 \text{ 月} - 1956 \text{ 年 } 2 \text{ 月}) \\ 2297.5 - 36.25t, & 39 \leq t \leq 45, \quad (1956 \text{ 年 } 3-9 \text{ 月}) \\ 80 + 15t, & 46 \leq t \leq 48, \quad (1956 \text{ 年 } 10-12 \text{ 月}) \end{cases}$$

相临两阶段折线的异同(回归系数)可通过它们的相关系数的比较获得结论。相临两阶段的转折时间的估计为(单位:月): $\hat{C}_1 = 9.17, \hat{C}_2 = 13.26, \hat{C}_3 = 21.57, \hat{C}_4 = 26.05, \hat{C}_5 = 30.68$ (估计值未在33到34之间。说明花生价格有暴涨), $\hat{C}_6 = 38.28, \hat{C}_7 = 43.26$ (有暴涨)。各阶段的回归系数说明每月花生油价格平均上升或下降多少。

三、多阶段Logistic增长曲线与Misherlich增长曲线的拟合

Logistic有限增长曲线的方程为

$$y = \frac{A}{1 + ae^{-bt}} \quad (a > 0)$$

由于Logistic曲线的上限为A,故两阶段的Logistic曲线之间有“突变”增长发生,这就产生了两阶段之间“突变”点的估计和两段Logistic曲线的光滑衔接问题。下面我们以两段Logistic曲线为例说明之(多阶段可仿此进行)。

两阶段Logistic增长曲线的方程式为

$$f(t) = \begin{cases} \frac{A_1}{1 + a_1 e^{-b_1 t}}, & 0 \leq t \leq T_1; \\ H(t), & T_1 \leq t \leq T_2; \\ \frac{A_2}{1 + a_2 e^{-b_2 t}}, & t > T_2 \end{cases} \quad (8)$$

$f(t)$ 的图象如图3所示。

设我们所观察到的两阶段Logistic增长序列为:

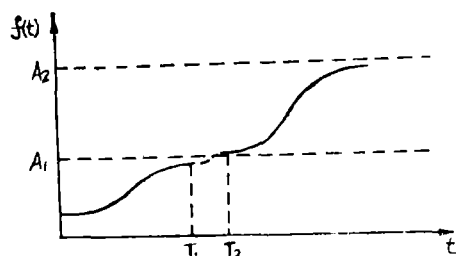


图 3

$$\frac{t}{X} \left| \begin{matrix} t_1, t_2, \dots, t_n \\ X_1, X_2, \dots, X_n \end{matrix} \right. \quad (9)$$

$f(t)$ 的拟合步骤如下所述。

1. 拟合 $\hat{X} = \frac{A_1}{1 + a_1 e^{-b_1 t}}$, 确定 T_1 和 T_2 ,

用 0.618 法优选 A_1 后线性化可估算出使

$$\sum \left(X_i - \frac{A_1}{1 + a_1 e^{-b_1 t_i}} \right)^2,$$

最小的 $\hat{A}_1$ 、 $\hat{a}_1$ 和 $\hat{b}_1$, 若最后一个拟合点为 (t_k, X_k) , 则 $\hat{T}_1 = t_k$, 而 $\hat{T}_2 = t_{k+1}$ 。

2. 用 (t_{k+1}, X_{k+1}) , (t_{k+2}, X_{k+2}) , $\dots$, (t_n, X_n) 拟合

$$\hat{X}' = \frac{A_2}{1 + a_2 e^{-b_2 t}}$$

估计出 $\hat{A}_2$ 、 $\hat{a}_2$ 和 $\hat{b}_2$ 。

3. 选择“突变”增长函数。由于

$$\frac{A_1}{1 + a_1 e^{-b_1 t}}$$

的上限为 A_1 , 即 $X = A_1$ 为其水平渐近线, 故从 T_1 开始, 应有一“突变”增长函数 $H(t)$ 存在, 使增长突破 A_1 而转向第二阶段的增长。设增长是连续的, 我们不妨要求 $H(t)$ 在接点 $(T_1, X(T_1))$ 及 $(T_2, X(T_2))$ 处满足如下条件:

$$\begin{cases} H(T_1) = \hat{X}(T_1) = m_1 \\ H(T_2) = \hat{X}(T_2) = m_2 \\ H'(T_1) = \hat{X}'(T_1) = m_3 \\ H'(T_2) = \hat{X}'(T_2) = m_4 \end{cases} \quad (10)$$

由于三次多项式亦是“S”型曲线, 故我们选取 $H(t) = at^3 + bt^2 + ct + d$ 。由 (10) 可得

$$\begin{cases} aT_1^3 + bT_1^2 + cT_1 + d = \frac{A_1}{1 + a_1 e^{-b_1 T_1}} = m_1 \\ aT_2^3 + bT_2^2 + cT_2 + d = \frac{A_2}{1 + a_2 e^{-b_2 T_2}} = m_2 \\ 3aT_1^2 + 2bT_1 + c = \frac{a_1 b_1 A_1 e^{-b_1 T_1}}{(1 + a_1 e^{-b_1 T_1})^2} = m_3 \\ 3aT_2^2 + 2bT_2 + c = \frac{a_2 b_2 A_2 e^{-b_2 T_2}}{(1 + a_2 e^{-b_2 T_2})^2} = m_4 \end{cases} \quad (11)$$

$$\text{令 } T = \begin{bmatrix} 1 & T_1 & T_1^2 & T_1^3 \\ 1 & T_2 & T_2^2 & T_2^3 \\ 0 & 1 & 2T_1 & 3T_1^2 \\ 0 & 1 & 2T_2 & 3T_2^2 \end{bmatrix} \quad B = \begin{bmatrix} d \\ c \\ b \\ a \end{bmatrix} \quad M = \begin{bmatrix} m_1 \\ m_2 \\ m_3 \\ m_4 \end{bmatrix}$$

则式 (11) 可写成

$$TB = M \quad (12)$$

由于 $|T| = -(T_2 - T_1)^4 \neq 0$, 故 T^{-1} 存在。因而

$$B = T^{-1}M \quad (13)$$

其中

$$T^{-1} = (T_2 - T_1)^{-2} \begin{bmatrix} T_2^2(T_2 - 3T_1) & T_1^2(3T_2 - T_1) & -T_1T_2^2(T_2 - T_1) & -T_1^2T_2(T_2 - T_1) \\ 6T_1T_2 & -6T_1T_2 & T_2(T_2 - T_1)(T_2 + 2T_1) & T_1(T_2 - T_1)(T_1 + 2T_2) \\ -3(T_2 + T_1) & 3(T_2 + T_1) & -(T_2 - T_1)(T_1 + 2T_2) & -(T_2 - T_1)(2T_1 + T_2) \\ 2 & -2 & (T_2 - T_1) & (T_2 - T_1) \end{bmatrix}$$

即

$$H(t) = [1, t, t^2, t^3] T^{-1}M \quad (14)$$

显然, 还可利用 $H(t)$ 在 $[T_1, T_2]$ 上的性质讨论一些符合对象增长机理的分析, 如什么时刻突破了 A_1 等。

Mitsherlich 增长曲线的方程为 $y = A + ae^{-bt}$ ($a < 0$), 上限为 A 。两阶段的 Mitsherlich 增长曲线的方程为:

$$f(t) = \begin{cases} A_1 + a_1 e^{-b_1 t} & 0 \leq t \leq T_1 \\ H(t) & T_1 \leq t \leq T_2 \\ A_2 + a_2 e^{-b_2 t} & t \geq T_2 \end{cases} \quad (15)$$

其图象如图 4 所示。

其拟合步骤与两阶段 Logistic 增长曲线相

同, $H(t)$ 亦为三次多项式, 矩阵 T 相同,

B 相同, 矩阵 M 中的元素含义不同, 其中

$$m_1 = A_1 + a_1 e^{-b_1 T_1}, \quad m_2 = A_2 + a_2 e^{-b_2 T_2},$$

$$m_3 = -a_1 b_1 e^{-b_1 T_1}, \quad m_4 = -a_2 b_2 e^{-b_2 T_2}.$$

$$H(t) = [1, t, t^2, t^3] T^{-1}M$$

【例 3】设有一两阶段 Logistic 增长曲线为

$$f(t) = \begin{cases} \frac{3}{1 + 2e^{-0.5t}} & 0 \leq t \leq 10; \\ H(t) & 10 \leq t \leq 11; \\ \frac{4.4}{1 + 4e^{-0.2t}} & t \geq 11; \end{cases}$$

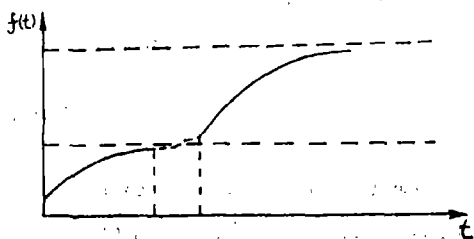


图 4

其中, $A_1 = 3$, $a_1 = 2$, $b_1 = 0.5$; $A_2 = 4.4$, $a_2 = 4$, $b_2 = 0.2$,

$$H(t) = at^3 + bt^2 + ct + d; T_1 = 10, T_2 = 11.$$

计算结果:

$$T^{-1} = \begin{bmatrix} -2299 & 2300 & -1210 & -1100 \\ 660 & -660 & 341 & 320 \\ -63 & 63 & -32 & -31 \\ 2 & -2 & 1 & 1 \end{bmatrix}$$

$$M' = [2.96011 \quad 3.04875 \quad 0.01968 \quad 0.18726]$$

$$a = 0.02966, b = -0.8505, c = 8.13168, d = -22.9669$$

$$H(t) = 0.02966t^3 - 0.8505t^2 + 8.13168t - 22.9669, [10, 11]$$

当 $t = 10.68$ 时, $H(t) = 3.0007$, 即 $t = 10.68$ 为两阶段间的实变点。

参 考 文 献

- [1] 杨纪柯、齐南翔、陈霖:《生物数学概论》, 科学出版社, 1982年。
- [2] 丁士晟:《多元分析方法及其应用》, 吉林人民出版社, 1981年。
- [3] 袁志发、刘光祖、肖俊璋: 0.618法在两种常用生长曲线拟合中的应用, 《西北农学院学报》, 1984年, 第3期。

Several Kinds of Methods of Fitness of Multiple Growth Curves

Yuan Zhifa

Xie Kerang

(Department of Basic Course, the Northwestern Agricultural University)

Abstract

Owing to the variations in external conditions and control of internal growth mechanism, many growth phenomena may have the phaso characteristics. This paper analyses the multiple growth models of echelon patterns, analytical models and multiple Logistic and Misherlish by using the methods of fitness.

key words: multiple growth curve, growth model of echelon pattern, growth model of analytical model, growth model of multiple Logistic, growth model of multiple Misherlish