

一种非线性回归模型的线性解法^{*}

景奉水

(山东农业大学工程技术学院)

孙爱东

(山东农业大学食品科学系)

A LINEAR SOLUTION OF A NONLINEAR REGULATION MODEL

Jing Feng-shui

(Engineering and Technology College of Shandong Agri. Univ.)

Sun Ai-dong

(Food Science Department of Shandong Agri. Univ.)

Abstract

With three unfixed parameters (m, a, b) , the nonlinear regression model, $y = m - \exp(ax + b + \varepsilon)$, is considered traditionally to be out of solution by using linear conversion way. Depending on linear regression and iteration way, this paper presents a new way to solve this problem.

§ 1. 引言

一元线性回归模型

$$y = ax + b + \varepsilon, \quad (1)$$

其中 a 和 b 为不依赖于 x 的参数, 是最小二乘法容易处理的. 然而真正符合线性模型的毕竟很少, 决大多数还属于非线性情况. 有些非线性模型常被化成线性模型 (1) 处理, 如模型

$$y = \exp(ax + b + \varepsilon), \quad (2)$$

其中 a 和 b 为不依赖于 x 的参数. 大多数非线性模型, 却难以化成线性模型进行处理, 如模型

$$y = m - \exp(ax + b + \varepsilon), \quad (3)$$

* 1996年7月23日收到.

其中 m , a 和 b 均是独立于 x 的参数.

模型 (3) 是一种生长模型, 当 $a < 0$ 时, 它具有极限值 m , 代表了实际模型的生长极限. 生长模型在农业、生物、经济、化工等领域有着极为广泛的应用. 通常对该类模型的处理是先对原模型进行线性近似后, 再采用 Gauss-Newton 迭代法 (或其改进型) 解决. G-N 迭代法计算步骤较为繁杂, 对初值的依赖性大, 在迭代过程中可能出现较大的波动, 甚至会出现发散情况^[1].

§ 2. 线性处理方法

2.1 模型的线性化

模型 (3) 有三个待定参数 m , a 和 b , 不能象模型 (2) 那样直接转化成模型 (1) 的形式. 若将 a 和 b 划成一组参数, 将 m 独自划成一组参数并分别处理, 就可实现线性化. 固定 m , 视其为一常数, 则 (3) 式可变形为

$$m - y = \exp(ax + b + \varepsilon). \quad (4)$$

(4) 式两边取自然对数, 得

$$\ln(m - y) = ax + b + \varepsilon.$$

令

$$t = \ln(m - y), \quad (5)$$

代入上式可得

$$t = ax + b + \varepsilon, \quad (6)$$

(6) 式与模型 (1) 的形式完全一致. 由最小二乘法知^[2]: 使截差平方和

$$Q = \sum_{i=1}^n [t - (ax_i + b)]^2$$

达到最小值, 回归系数为

$$\begin{aligned} a &= \frac{S_{xt}}{S_{xx}}, \\ b &= \bar{t} - a\bar{x}, \end{aligned} \quad (7)$$

式中

$$\begin{aligned} \bar{x} &= \frac{1}{n} \sum_{i=1}^n x_i, & \bar{t} &= \frac{1}{n} \sum_{i=1}^n t_i, \\ S_{xx} &= \sum_{i=1}^n (x_i - \bar{x})^2, & S_{xt} &= \sum_{i=1}^n (x_i - \bar{x})(t_i - \bar{t}). \end{aligned}$$

样本相关系数

$$r_{xt} = \frac{S_{xt}}{\sqrt{S_{xx} \cdot S_{tt}}}, \quad (8)$$

式中

$$S_{tt} = \sum_{i=1}^n (t_i - \bar{t})^2.$$

从以上过程可知, 对于 m 的每一确定值, 都可由线性回归得到一组确定的 (a, b) . 因此 a 和 b 是 m 的函数.

设回归函数 $g(m, a, b) = m - \exp(ax + b)$, 由于 a, b 都为 m 的函数, 所以有

$$f(m) = g(m, a, b) = m - \exp(ax + b). \quad (9)$$

将 (9) 代入 (3), 得

$$y = f(m) + [\exp(ax + b) - \exp(ax + b + \varepsilon)] = f(m) + \varepsilon'. \quad (10)$$

设

$$S(m) = \sum_{i=1}^n [y_i - f_i(m)]^2, \quad (11)$$

$S(m)$ 即是回归函数与实验值在所有 x_i 处的偏差平方和.

根据最小二乘原理, m 值的选取应使得 $S(m)$ 最小. 设 $m = M$ 时, $S(m)$ 最小, 那么只要求出 M , 由 (4)-(6) 式, 非线性模型 (3) 式便转化成了线性模型 (6) 式.

2.2 M 的数值解法

若 $m = M$ 使得 $S(m)$ 最小, 必有

$$\left. \frac{dS(m)}{dm} \right|_{m=M} = 0, \quad (12)$$

$$\begin{aligned} \frac{dS(m)}{dm} &= \sum_{i=1}^n (-2)[y_i - f_i(m)] \frac{d(f_i(m))}{dm} \\ &= -2 \sum_{i=1}^n \left\{ [y_i - f_i(m)] \left[1 - \frac{\partial(\exp(ax_i + b))}{\partial a} \frac{da}{dm} - \frac{\partial(\exp(ax_i + b))}{\partial b} \frac{db}{dm} \right] \right\} \\ &= -2 \sum_{i=1}^n \left\{ [y_i - m + \exp(ax_i + b)] \left[1 - x_i \exp(ax_i + b) \frac{da}{dm} - \exp(ax_i + b) \frac{db}{dm} \right] \right\}. \end{aligned} \quad (13)$$

将 (5) 式代入 (7) 式, 对 m 求导可得

$$\begin{cases} \frac{da}{dm} = \frac{1}{S_{xx}} \sum_{i=1}^n (x_i - \bar{x}) \left(\frac{1}{m - y_i} - \frac{1}{n} \sum_{i=1}^n \frac{1}{m - y_i} \right), \\ \frac{db}{dm} = \frac{1}{n} \sum_{i=1}^n \frac{1}{m - y_i} - \frac{da}{dm} \bar{x}. \end{cases} \quad (14)$$

(13), (14) 式中 a, b, \bar{x}, S_{xx} 等的定义同 (7) 式.

直接由式 (12) 用解析法求 M 是不可能的, 为此我们采用取 m 系列值迭代逼近 M 的方法来获得近似值. 由式 (13), (14) 知, $S(m)$ 对 m 由二阶以上连续导数. $S(m)$ 在 $(M + \Delta m)$ 点的泰勒展开式为

$$S(M + \Delta m) - S(M) = \left. \frac{ds}{dm} \right|_{m=M+\Delta m} \cdot (-\Delta m) + \dots \approx \left. \frac{ds}{dm} \right|_{m=M+\Delta m} \cdot (-\Delta m). \quad (15)$$

上式说明, 当 $\Delta m \rightarrow 0$ 或 $\frac{dS}{dm} \rightarrow 0$ 时, 都使得 $S(M + \Delta m) \rightarrow S(M)$. 因此二者都可作为衡量迭代精度的标准. 本文在实例中采用的是 Δm . 由于 M 是未知的, 因此 Δm 的精确值无法确定, 但是 Δm 的范围可由以下步骤确定:

设精度要求为 $|\Delta m| < \delta, \delta > 0$, 当 $\frac{dS}{dm} < 0$ 时, 令 $m_1 = m$, 当 $\frac{dS}{dm} > 0$ 时, 令 $m_2 = m$. 若 $|m_2 - m_1| < \delta$, 则必有 $|\Delta m| = |m - M| < |m_2 - m_1| < \delta$ 成立.

2.3. 计算步骤

步 1 设置迭代精度 $\delta > 0$, 以及 m, m_1, m_2 和步长 step 的初始值;

步 2 由 (6), (7), (14) 式计算 $a, b, \frac{da}{dm}$ 和 $\frac{db}{dm}$;

步 3 由 (11) 式计算 $S(m)$, 由 (13) 式计算 $\frac{dS}{dm}$, 再根据 2.2 计算 m_1 和 m_2 ;

步 4 若 $|m_2 - m_1| < \delta$, 停止迭代, 转步 6;

步 5 计算 step, m , 转步 2;

步 6 计算 r_{xt} ;

步 7 输出 $m, a, b, S(m)$ 及 r_{xt} 等;

步 8 结束.

2.4 计算步骤中的几点问题

(1) 初始值的选择

由于 (3) 式是带有生长极限的模型, m 就代表了极限, 因此 m 的初值选用数据的最大值或再加一小正数即可 (在本文的实例中这一小正数采用的是初始步长值).

步长 step 的初始值不宜选得过大, 本文根据实际情况取初始步长为 1. m_1 和 m_2 的初时值均取零.

(2) 迭代过程中 m 和 step 的计算

每迭代一步, m 都要加上步长, 即 $m = m + \text{step}$. 在 $\frac{dS}{dm}$ 变号以前, step 的值保持不变; 一旦 $\frac{dS}{dm}$ 变号, step 的值应变小, 并变号 (本文取原来的负十分之一, 即 $\text{step} = -\text{step} \cdot 1/10$).

§ 3. 实 例

在一次月见草油萃取中, 所得数据如表 1 所示 [3]:

表 1 萃取数据

时间 (min.)	15	30	45	60	90	120
萃取率 (%)	18.5	36.4	43.0	54.1	61.0	63.8

由表 1 中的数据作散点图, 如图 1 所示. 根据散点图的走势, 可选取 (3) 式为回归模型. 取 $\delta = 0.1$, m 的初始值为 64.8, 表 1 数据的迭代结果如表 2 所示.

表 2 迭代结果

迭代次数	m	step	$ m_2 - m_1 $	$\frac{dS}{dm}$	$S(m)$
1	64.800	1.00	64.800	-129.471	66.787
2	65.800	1.00	65.800	-13.849	15.307
3	66.800	1.00	1.000	3.289	12.178
4	66.700	-0.10	0.900	2.505	11.888
5	66.600	-0.10	0.800	1.598	11.682
6	66.500	-0.10	0.700	0.545	11.574
7	66.400	-0.10	0.100	-0.676	11.578
8	66.410	0.01	0.090	-0.546	11.572

$$r_{rt} = 0.997, a = -0.028138, b = 4.28526, m = 66.410$$

将 a, b, m 值代入 (3) 式, 得 $y = 66.410 - \exp(-0.028138x + 4.28526)$. 该方程曲线如图 1 所示.

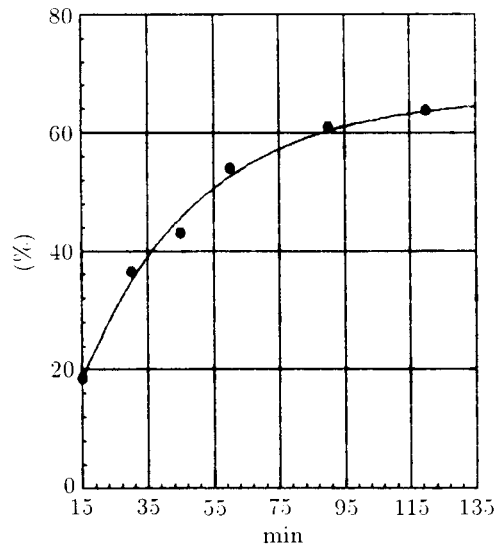


图 1 萃取数据的散点图和回归曲线

§ 4. 与 $G-N$ 解法的比较

根据 $G-N$ 法的要求, 将回归方程

$$y = g(m, a, b) = m - \exp(ax + b)$$

线性化处理, 其迭代公式为^[1]

$$\theta_{i+1} = \theta_i + [V^T(\theta_i)V(\theta_i)]^{-1}V^T(\theta_i)[Y - F(\theta_i)], \quad (16)$$

其中, θ_i 为迭代到第 i 步时的参数向量, 即 $\theta_i = (m_i, a_i, b_i)^T$, $V(\theta_i)$ 为 $n \times p$ 阶矩阵, n 代表采样数, p 为模型的参数数目. 其定义是:

$$V(\theta_i) = \begin{bmatrix} \left. \frac{\partial g}{\partial m_i} \right|_{x=x_1} & \left. \frac{\partial g}{\partial a_i} \right|_{x=x_1} & \left. \frac{\partial g}{\partial b_i} \right|_{x=x_1} \\ \left. \frac{\partial g}{\partial m_i} \right|_{x=x_2} & \left. \frac{\partial g}{\partial a_i} \right|_{x=x_2} & \left. \frac{\partial g}{\partial b_i} \right|_{x=x_2} \\ \vdots & \vdots & \vdots \\ \left. \frac{\partial g}{\partial m_i} \right|_{x=x_n} & \left. \frac{\partial g}{\partial a_i} \right|_{x=x_n} & \left. \frac{\partial g}{\partial b_i} \right|_{x=x_n} \end{bmatrix}. \quad (17)$$

Y 为参数向量, 即

$$Y = (y_1, y_2, \dots, y_n)^T.$$

$F(\theta_i)$ 为模型 (3) 在采样点所对应的函数值所构成的向量, 即

$$F(\theta_i) = [g(m_i, a_i, b_i)|_{x=x_1}, g(m_i, a_i, b_i)|_{x=x_2}, \dots, g(m_i, a_i, b_i)|_{x=x_n}]^T.$$

与前面介绍的方法比较, 除了计算量增大外, $G-N$ 法的突出问题是: 初值较难选择和对初值的依赖性. 一方面, $G-N$ 法需要 3 个初值: m_0, a_0 和 b_0 . 其中除 m_0 可根据采样值的最大值确定外, 根据模型 (3) 函数的特性, 只能判断出 a_0 须取一负数, 而 b_0 的取值范围是任意实数, 因此初值的选择质量难以保证; 另一方面, $G-N$ 法对初值的依赖性强, 初值稍加变化, 就会带来迭代过程的强烈波动, 甚至数值溢出等后果. 对于本文中的实例, 为了和前面介绍的方法比较, 我们将 m_0 取 64.8, b_0 取 1, 而对 a_0 分别取了 $-1, -0.01, -0.02$ 和 -0.03 四个值进行试算. 结果除 a_0 取 -0.02 迭代成功外, 其它都因计算机数值溢出而失败. 表 3 列出了 a_0 取 -0.02 和 -0.03 时的部分迭代结果.

表 3 G-N 法的部分结果

迭代 次数	$a_0 = -0.02$		$a_0 = -0.03$	
	$\theta_i = (m_i, a_i, b_i)$	$S_i = (Y - F(\theta_i)) ^2$	$\theta_i = (m_i, a_i, b_i)$	$S_i = (Y - F(\theta_i)) ^2$
1	65.1219, -0.1969, 25.6432	3223.0316	66.8254, 0.05058, 26.5019	3295.2330
2	56.3285, -0.1969, 24.6432	5.1153×10^{19}	29.5871, 0.05057, 25.5019	2.04938×10^{28}
3	56.3295, -0.1969, 23.6432	6.9229×10^{18}	29.2370, 0.05057, 24.5019,	2.77354×10^{27}
⋮	⋮	⋮	⋮	⋮
15	56.418, -0.1805, 9.4241	4716429.80	29.3874, 0.05057, 11.5019	1.04702×10^{17}
20	58.6047, -0.06359, 5.1389	8918.7964	29.3875, 0.05057, 7.5015	4.75367×10^{12}
23	66.9332, -0.02714, 4.2814	11.5026	29.3972, 0.05063, 4.49346	1.17868×10^{10}
24	66.9347, -0.02714, 4.281409	11.2562	29.4139, 0.05075, 3.47883	1.59605×10^9
25	66.9347, -0.02714, 4.281409	11.2562 达到精度 * 迭代结束	29.4594, 0.05106, 2.43894	2.16327 $\times 10^8$
28			30.7380, 0.06136, -1.8673	561237.81
29			计算溢出 **, 程序中断	

注: * 程序的精度 $|S_i - S_{i-1}| < 0.1$, ** 程序采用的是 C 语言的 8 字节双精度数.

§ 5. 结 论

本文应用线性回归和变步长迭代法解决了模型 $y = m - \exp(ax + b + \varepsilon)$ 的非线性回归问题. 新的方法比传统的 G-N 迭代法具有方法简洁、易选初值、收敛快而稳定性好等优点. 根据此方法, 作者用 C 语言开发了一回归程序, 成功地处理了文献 [3] 中的几十组实验数据.

本文的方法尽管针对一种模型提出, 但是除能转化成此模型的问题可使用外, 其它具有三个待定参数的回归模型问题也可借鉴之.

参 考 文 献

- [1] 韦博成, 近代非线性回归分析, 东南大学出版社, 1989.
- [2] 中国矿业学院数学教研室, 数学手册 (第二版), 科学出版社, 1990.
- [3] 孙爱东, CO₂ 超临界萃取技术提取天然材料中多种保健成分的研究, 山东农业大学硕士研究生论文, 1996.