

线性规划法在植物生态学数学模型 参数估计中的应用初探

李凤民 苗云 张振万

(西北植物研究所, 陕西杨陵 712100)

提 要 本文提出了一种植物生态学数学模型参数估计的新方法——线性规划法, 并结合最小二乘法对模型参数的估计进行了实例分析。认为某种程度上前者更优于后者, 是一种值得进一步研究的方法。

关键词 植物生态学模型; 线性规划法; 最小二乘法; 参数估计

植物生态学研究日益从定性走向定量, 构造的数学模型多种多样。其中, 数据的规律性分析和曲线拟合占有重要位置。数学模型中参数的估计目前普遍采用最小二乘法。本文试图提出另一种参数估计的方法——线性规划法, 以求更合理更全面地对数据进行分析处理。本文在讨论中兼与最小二乘法进行对比分析。

一、线性规划的基本思想及模型的建立

1. 线性规划法的基本思想及与最小二乘法的比较

设: 欲建立的植物生态学数学模型可用下面的函数式表示:

$$y = f(a_h, x_j) \quad h=1, 2, \dots, u, \quad j=1, 2, \dots, m \quad (u \geq m)$$

其中, a_h 为模型参数, x_j 为模型变量。

另设: 有一组实测值为

$$(x_{ij}, y_i) \quad i=1, 2, \dots, N \quad j=1, 2, \dots, M$$

线性规划法^[1]的基本思想是确定一组 a_h 值, 使得

$$Q(\hat{a}_h) = \sum_{i=1}^N |y_i - f(\hat{a}_h, x_{ij})| \quad (1)$$

达到最小。

最小二乘法^[2]的基本思想是确定一组 a_h 值, 使得

$$Q(\hat{a}_h) = \sum_{i=1}^N [y_i - f(\hat{a}_h, x_{ij})]^2 \quad (2)$$

达到最小。

可见, 二者的根本差别在于前者要求模型计算值 $\hat{y}_i$ 与实测值 y_i 之差的绝对值之和

最小。后者则要求二者之差的平方和最小。

2. 线性规划法参数估计的过程

设: $e_i = y_i - \hat{y}_i$, 则有

$$f(\alpha_h, x_{ij}) + e_i = y_i \quad (3)$$

其中, x_{ij} 和 y_i 为实测值。 α_h 和 e_i 为欲求未知量, 二者可以为正或为负。线性规划模型要求约束方程中欲求变量全部为负。故作如下变换:

$$\alpha_h = \beta_{1h} - \beta_{2h}, e_i = \eta_i - \rho_i \quad (4)$$

其中, $\beta_{1h}, \beta_{2h}, \eta_i, \rho_i \geq 0$, η_i 为正偏差, ρ_i 为负偏差。

因此, (3) 式即可变为

$$f(\beta_{1h} - \beta_{2h}, x_{ij}) + \eta_i - \rho_i = y_i \quad (5)$$

(1) 式变为

$$\text{Min } Q(\beta_{1h} - \beta_{2h}) = \sum_{i=1}^N (\eta_i - \rho_i) \quad (6)$$

至此, 便构成了估计 α_h 的线性规划模型为

$$\text{目标函数 } \text{Min } Q(\beta_{1h} - \beta_{2h}) = \sum_{i=1}^N (\eta_i - \rho_i)$$

$$\text{约束条件 } f(\beta_{1h} - \beta_{2h}, x_{ij}) + \eta_i - \rho_i = y_i$$

$$\text{非负条件 } \beta_{1h}, \beta_{2h}, \eta_i, \rho_i \geq 0$$

该模型可用求解线性规划问题的最一般方法——单纯形法求解。特别由于本模型中 η_i 的系数为 1, 约束方程系数自身可构成单位阵, 使计算过程更为简单。

当由此模型求出 $\beta_{1h}, \beta_{2h}, \eta_i$ 和 ρ_i 后, 再经 (4) 式变换出欲估计的原模型参数 α_h , 以及实测值与估计值之间的差值 e_i 。

二、应用实例分析

例1. 多重线性回归方程参数的估计

对笔者1988年5月16日在宁夏盐池测定的人工草地苜蓿蒸腾强度与环境因子的资料, 利用线性规划法和最小二乘法进行对比分析。设蒸腾强度(失水 g/小时/鲜重 g)与温度(℃)、相对湿度、风速(m/s)具有如下关系:

$$y = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_3$$

原始资料和计算结果列于表1。

由线性规划法求得的回归方程为

$$\hat{y} = 0.5086x_1 + 0.1090x_2 - 0.3258x_3 - 10.6935 \quad R = 0.9496$$

由最小二乘法得到的回归方程为

$$\hat{y} = 0.4312x_1 + 0.08851x_2 - 0.3381x_3 - 8.6080 \quad R = 0.9660$$

对两个相关系数的差异作显著性检验得:

$$t(26) = 0.512 < t_{0.25}(26) = 0.684$$

即两个方程相关系数无显著差异。

两种方法出现的残差极值都在 No.10, 此时正值下午4时(北京夏令时), 苜蓿也出现了植物蒸腾普遍存在的“午休”现象(y_{10} 显著低于相邻两个值, 而环境因子却无

异常)。可见,本例中两种方法检出的异常点(本文暂以残差极值点作为异常点)是一致的、合理的。

例2. 二次抛物线回归方程的参数估计

用上例中前13个蒸腾数据与时间作二次抛物线拟合,原始数据及计算结果列于表2。其中,时间都已减去基数7(蒸腾从早晨7点开始测定),然后换算成小时计算。设蒸腾强度与时间有如下关系:

$$y = a_0 + a_1x + a_2x^2$$

由线性规划法所得的方程为:

$$\hat{y} = 0.1221 + 0.3624x + 0.02014x^2 \quad R = 0.9367$$

由最小二乘法所得方程为:

$$\hat{y} = 0.07056 + 0.3970x - 0.02431x^2 \quad R = 0.9481$$

对两个相关系数差异作显著性检验得:

$$t(20) = 0.2299 < t_{0.25}(20) = 0.687$$

即两个方程与原值的相关性是一致的。

在异常点检验方面,线性规划法检出两个异常点,分别为 No.10 和 No.13。最小二乘法检出一个异常点为 No.10。从上例中可清楚地看到, No.10 作为异常点显然是合理的。对于 No.13, 从 y_i 值中也可以看到, $y_{13} = 1.082$, 与 $y_{12} = 1.654$ 相差竟达 0.572, 是所有 y_i 与相邻值之差中最大的。显然,它作为异常点无疑也是合适的(至于 No.13 如何成为异常点,以及上例中 No.14—16 数据的变化是由植物的环境因子和植物的蒸腾特点决定的,本文不予讨论)。可见,在这一模型中线性规划法比最小二乘法更优越一些。

例3. 指数曲线方程的参数估计

一天中,植物蒸腾耗水量(每克干物质所损失的水量)与时间的关系可近似地用一条指数曲线描述,其表达式如下:

$$y = ax^b$$

试估计参数 a、b。资料及计算结果列于表3。

由线性规划法所得方程为:

$$\hat{y} = 5.6998x^{1.1989} \quad R = 0.9898$$

由最小二乘法所得方程为:

$$\hat{y} = 8.3864x^{0.9846} \quad R = 0.9911$$

对两个相关系数的差异作显著性检验得:

$$t(26) = 0.179 < t_{0.25}(26) = 0.684$$

即:二者无显著差异。

从残差来看,线性规划法的 $\sum |e_i|$ 和 $|\sum e_i|$ 都比最小二乘法的相应值小得多,说明对前者来说,实测值能较好地分布在拟合曲线的两侧,使实测数据得到较好的逼近,且拟合精度也好于最小二乘法。由于残差出现上述特点,由残差检出的异常点也出现类似的偏差。线性规划法检出的异常点为 No.16, 其它点的残差都显著小于该点的残差(绝对值)。最小二乘法检出的异常点为 No.13, 而残差值与之接近的点还有 No.11,

表 1 蒸腾强度与温度、湿度和风速的关系

Table 1 The relation of transpiration intensity and temperature, humidity as well as wind rate

编号 No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	合计Sum	$\Sigma e_i $
蒸腾强度 Transpiration intensity y_i	0.179	0.459	0.776	0.987	1.175	1.334	1.709	1.832	1.732	1.372	1.876	1.654	1.082	0.683	0.16	0.14	17.150	
温度 Temperature x_1	9.8	14.5	18.2	19.6	20.5	21.5	22.0	23.0	23.5	23.8	24.0	23.2	22.5	20.2	18.5	17.4		
湿度 Humidity x_2	60	43.0	31.5	21.0	18.0	16.0	15.0	12.5	11.2	9.5	10.0	10.0	15.0	20.5	27.0	33.5		
风速 Wind rate x_3	2.0	3.3	4.0	2.6	2.0	2.0	2.0	2.6	2.0	2.3	2.6	1.6	4.0	3.3	4.6	4.6		
线性规划法 Linear program method ϵ_i	0.179	0.293	0.693	0.717	1.094	1.394	1.479	1.520	1.828	1.698	1.756	1.675	1.082	0.740	0.160	0.309	16.557	
最小二乘法 Minimum square method ϵ_i	0	0.166	0.082	0.270	0.080	0	0.230	0.312	-0.096	-0.326*	0.120	-0.021	0	-0.057	0	-0.109	0.593	1.929
异常点残差 *Scrap value of the abnormal point.	0.252	0.334	0.675	0.822	1.191	1.412	1.528	1.536	1.839	1.717	1.746	1.739	1.068	0.800	0.203	0.304	17.150	2.004

表 2 蒸腾强度与时间的关系

Table 2 The relation between transpiration intensity and time

编号 No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	合计Sum	$\Sigma e_i $
蒸腾强度 Transpiration intensity y_i	0.179	0.459	0.776	0.987	1.175	1.334	1.709	1.832	1.732	1.372	1.876	1.654	1.082					
时间 Time x_i	0.0833	0.9833	2.0000	3.2333	4.1667	5.0333	5.9667	7.1333	8.0000	8.9667	9.9333	11.2000	12.0667					
线性规划法 Linear program method ϵ_i	0	0.010	-0.096	-0.107	-0.102	0.142	0.150	0	-0.380*	0.142	0	-0.480*	-0.694	1.636				
最小二乘法 Minimum square method ϵ_i	0.103	0.437	0.767	1.100	1.302	1.453	1.573	1.665	1.690	1.675	1.614	1.466	1.320					
异常点残差 *Scrap value of the abnormal point.	0.022	0.009	-0.113	-0.127	-0.119	0.136	0.167	0.042	-0.303*	0.262	0.188	-0.238	0	1.793				

表 3 蒸腾耗水量与时间的关系

Table 3 The relation between water transpiration and time

编号 No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	合计Sum	$\Sigma e_i $
蒸腾耗水量 Water transpiration y_i	1.348	4.687	10.304	17.574	26.281	35.909	48.343	61.735	74.232	84.423	97.917	109.965	117.809	122.783	123.936	124.464		
时间 Time x_i	0.0833	0.9833	2.0000	3.2333	4.1667	5.0333	5.9667	7.1333	8.0000	8.9667	9.9333	11.2000	12.0667	13.0000	13.8833	14.6000		
线性规划法 Linear program method ϵ_i	0.291	5.586	13.067	23.220	31.455	39.438	48.343	59.864	68.670	78.716	88.978	102.723	112.308	122.783	132.834	141.082		
最小二乘法 Minimum square method ϵ_i	1.057	-0.899	-2.763	-5.647	-5.174	-3.529	0	1.871	6.562	5.707	8.939	7.242	5.501	0	-8.898	-16.618*	7.648	79.306
异常点残差 *Scrap value of the dubious abnormal point.	0.726	8.249	16.595	26.630	34.184	41.174	48.681	58.040	64.977	72.701	80.412	90.498	97.389	104.802	111.809	117.490		
最小二乘法 Minimum square method ϵ_i	0.622	-3.562	-6.291	-9.056	-7.903	-5.265	-0.338	3.695	9.255	11.722	17.505*	19.467*	20.420*	17.981	12.127	6.974	87.353	152.183

No.19, No.14等。这就使得由此检出的异常点的真实性发生怀疑。所以,由此例的结果也不难看出,线性规划法更优于最小二乘法。

三、讨 论

1. 从多重线性回归看,最小二乘法所得的残差的代数和 $\sum e_i$ 为零(把二次抛物线方程中的 x^2 项视为实测值,则该模型便视为二重线性回归方程,故其结果与例1一致)。此为其性质之一。在平面上它表现为回归直线通过散点图的几何中心 $(\bar{x}, \bar{y})$, 而线性规划法不具备此性质,故其 $\sum e_i$ 不为零。

2. 由于出发点的不同,线性规划法所得的 $\sum |e_i|$ 值一般要小于最小二乘法,且由于二者在回归预测的相关系数上差异极不显著,所以,可以认为:在建立回归预测模型时,一般地讲,线性规划法优于最小二乘法。

3. 建立数学模型,根据残差出现的特点检验分析实测数据的规律性是数据分析的一个重要手段。线性规划法的分析效果与最小二乘法相当,或更优。可以单独选用线性规划法,或与最小二乘法结合使用。

4. 同最小二乘法一样,由于线性规划法要求约束方程全部为线性,故对欲建立的非线性模型要先划成线性形式。然后才能求解。

5. 相对地讲,线性规划法的计算要比最小二乘法更复杂一些。但随着微型计算机的日益普及,计算将不再成为问题。

6. 同最小二乘法相比,线性规划法估计模型参数的理论还很不成熟,还缺乏一套较为系统完备的理论基础,这是需要进一步研究的。

参 考 文 献

- 1 伊格尼齐奥著,闵仲求等译。单目标和多目标系统线性规划,上海:同济大学出版社,1986:188—192
- 2 中国科学院计算中心概率统计组。概率统计计算,北京:科学出版社,1979:105—143

A PRELIMINARY STUDY ON THE APPLICATION OF THE LINEAR PROGRAM METHOD TO THE ESTIMATION OF PARAMETERS IN PLANT ECOLOGICAL MODEL

Li Fengmin, Miao Yun, Zhang Zhenwan

(Northwest Institute of Botany, Shaanxi 712100)

Abstract In this paper, the authors suggest using linear program method for the estimation of parameters in plant ecological models. The results from the linear program method and the minimum square method were compared and analysed with three examples. It is obvious that the former is superior than the latter in general. So linear program method is consciously worthy for further studying and applying in ecological investigation.

Key words Plant ecological model, Linear program method, Minimum square method, Estimation of parameter