

# 田间害虫种群抽样技术中 指数法的参数估计

刘光祖 史美英<sup>✓</sup> 宋世德

(西北农业大学基础科学系, 陕西杨陵 712100)

**摘要** 针对 Tylor 指数法用广义 LS 法对其参数进行了估计, 并利用 BOX 百分偏差对估计偏差进行了度量, 从而可以有效的控制把估计精度不高的参数值用于田间害虫的抽样技术之中。

**关键词** 田间害虫, 指数法, 参数估计, 抽样技术

**中图分类号** O212.1, S433.89

在田间害虫种群的空间分布型和抽样技术研究中, 理论抽样数  $n$  和抽样模式  $T(n)$  中的参数一般是利用回归的方法通过试验资料来估计的<sup>[1~3]</sup>, 而参数的估计精度对  $n$  与  $T(n)$  的精度有着直接的影响。如何获得精度较高的参数值以提高抽样数  $n$  与抽样模式  $T(n)$  的精度, 正是本文所要讨论的问题。

## 1 理论抽样数与 Green 序贯抽样模式

$$\text{Tylor 指数法 } S^2 = a\bar{x}^b \quad (1)$$

$$\text{理论抽样数 } n = \left(\frac{t}{D}\right)^2 a\bar{x}^{b-1} \quad (2)$$

$$\text{Green 序贯抽样模式 } T(n) = \exp \left\{ \frac{\ln \frac{D^2}{a} + (b-1) \ln n}{b-2} \right\} \quad (3)$$

其中  $t$  为给定置信概率之下的  $t$  分布分位点值, 其他记号见文献[1, 2]。

## 2 参数 $a, b$ 的估计与估计偏差

### 2.1 $a, b$ 的广义 LS 估计

回归方程(1)是非线性回归方程, 在田间害虫种群的空间分布型及抽样技术研究中, 一般做法是将(1)式线性化, 按线性回归用 LS 法(最小二乘法)估计其参数的。大量的应用结果表明, 这种参数估计方法有其缺点, 线性化往往使 LS 法等方差的假定被破坏, 结果可能使参数的估计精度很差, 但用相关指数或 F 检验可能是显著或极显著的。如果将这种精度不高的参数估计值用于决定抽样数  $n$  与  $T(n)$ , 必然会影响抽样数与抽样模式的精度。为了获得较好的参数估计值以提高  $n$  与  $T(n)$  的精度, 现用广义 LS 法估计  $a, b$ 。

设回归方程

$$y = f(x, \theta) \quad (4)$$

式中,  $f$  为非线性函数,  $\theta = (\theta_1, \dots, \theta_p)'$  为一组参数。将(4)式线性化为

$$Y = A + BX \quad (5)$$

其中变换式为  $X = \varphi(x), Y = \psi(x, y)$ 。设  $(x_i, y_i) (i=1, 2, \dots, m)$  为  $m$  组观测值, 则  $A, B$  的估计值为<sup>[4]</sup>

$$\begin{cases} \hat{B} = \frac{\sum k_i X_i Y_i - (\sum k_i X_i)(\sum k_i Y_i) / \sum k_i}{\sum k_i X_i^2 - (\sum k_i X_i)^2 / \sum k_i} \\ \hat{A} = \frac{\sum k_i Y_i - \hat{B} \sum k_i X_i}{\sum k_i} \\ k_i = \left( \frac{dY_i}{dx_i} \right)^{-2} \end{cases} \quad (6)$$

由  $\hat{A}, \hat{B}$  可得  $\hat{\theta}$ 。

由(6)式, 可得(1)式中  $a, b$  的估计公式为

$$\begin{cases} \hat{b} = \frac{\sum s_i^4 \ln \bar{x}_i \ln s_i^2 - (\sum s_i^4 \ln \bar{x}_i)(\sum s_i^4 \ln s_i^2) / \sum s_i^4}{\sum s_i^4 (\ln \bar{x}_i)^2 - (\sum s_i^4 \ln \bar{x}_i)^2 / \sum s_i^4} \\ \hat{a} = \exp \left\{ \frac{\sum s_i^4 \ln s_i^2 - \hat{b} \sum s_i^4 \ln \bar{x}_i}{\sum s_i^4} \right\} \end{cases} \quad (7)$$

公式(7)的计算量虽然略大于用 LS 法估计  $a, b$  时的计算量, 但在很大程度上却能提高(1)式中  $a, b$  的估计精度。

## 2.2 $a, b$ 估计量的百分偏差

参数  $a, b$  的估计好坏, 可以通过计算  $\hat{a}, \hat{b}$  的百分偏差来度量, 其偏差计算公式为<sup>[5]</sup>

$$\text{Bias}(\hat{\theta}) = -\frac{\sigma^2}{2} (V'V)^{-1} V' \text{tr}[(V'V)^{-1} W] \quad (8)$$

$\hat{a}, \hat{b}$  的百分偏差计算公式为<sup>[5]</sup>

$$P(\hat{a}) = \frac{100 \text{Bias}(\hat{a})}{\hat{a}} \%, \quad P(\hat{b}) = \frac{100 \text{Bias}(\hat{b})}{\hat{b}} \% \quad (9)$$

若  $P(\hat{a}), P(\hat{b})$  均不超过 1%, 则认为估计值  $\hat{a}, \hat{b}$  的偏差较小, 把它们代入(2)式、(3)式, 可得到精度较高的  $n$  与  $T(n)$ 。

## 3 算 例

表 1 中的数据引自文献[2]。

表 1 黑绒金龟抽样资料

| 序 号       | 1       | 2        | 3       | 4        | 5       | 6       | 7       | 8       | 9       | 10      |
|-----------|---------|----------|---------|----------|---------|---------|---------|---------|---------|---------|
| 样点数       | 305     | 460      | 129     | 137      | 574     | 162     | 140     | 240     | 130     | 989     |
| $\bar{x}$ | 1.09836 | 3.30217  | 0.62791 | 4.55474  | 1.58711 | 1.59259 | 0.85741 | 1.46666 | 0.30769 | 0.31200 |
| $S^2$     | 2.86529 | 19.52287 | 1.13286 | 35.03537 | 5.02638 | 3.06869 | 2.03673 | 3.04491 | 0.53610 | 0.69860 |

### 3.1 $a, b$ 的广义 LS 估计

由公式(7)计算的  $\hat{a}, \hat{b}$  及残差平方和  $\sum_{i=1}^{10} (s_i^2 - \hat{s}_i^2)^2$  列于表 2 中。为了便于比较, 表 2 中还列出了由 LS 法估计的  $\hat{a}, \hat{b}$  值及残差平方和。

表2  $\hat{a}, \hat{b}$  值及残差平方和

|     | $\hat{a}$ | $\hat{b}$ | 回归方程                                   | $\sum (x_i^2 - \bar{x}_i^2)^2$ | 相关系数 $r$ |
|-----|-----------|-----------|----------------------------------------|--------------------------------|----------|
| GLS | 2.08676   | 1.86104   | $\hat{y}^2 = 2.08676\bar{x}^{1.86104}$ | 5.8560                         |          |
| LS  | 2.61398   | 1.45431   | $\hat{y}^2 = 2.61398\bar{x}^{1.45431}$ | 156.8245                       | 0.9733   |

由表2可见,用广义LS法估计的  $a, b$  值优于用LS法估计的  $a, b$  值。因为前者的残差平方和为5.856,而后者的残差平方和为156.8245,虽然相关系数(0.9733)接近于1,一般认为回归是极显著的,但实际上它不是最优回归方程。

### 3.2 $\hat{a}, \hat{b}$ 的百分偏差

对用广义LS法估计的  $\hat{a}=2.08676, \hat{b}=1.86104$ ,利用公式(8),(9)计算其估计偏差和百分偏差,得

$$\hat{\sigma}^2 = \frac{5.8560}{8} = 0.732, \quad \text{Bias}(\hat{\theta}) = (0.005728, 0.002364)'$$

$$P(\hat{a}) = \frac{100 \times 0.005728}{2.08676} \% = 0.3\%, \quad P(\hat{b}) = \frac{100 \times 0.002364}{1.86104} \% = 0.1\%$$

由于  $\hat{a}, \hat{b}$  的百分偏差均小于1%,所以其估计偏差较小,从而所得非线性回归方程是较优的回归方程。

对于用LS法估计的  $\hat{a}=2.61398, \hat{b}=1.45431$ ,计算它们的偏差与百分偏差,得

$$\hat{\sigma}^2 = 19.6031, \quad \text{Bias}(\hat{\theta}) = (0.16569, 0.07928)'$$

$$P(\hat{a}) = 6.3\%, \quad P(\hat{b}) = 5.5\%$$

这里  $\hat{a}, \hat{b}$  的百分偏差均超过了1%,它表明用LS法估计的  $\hat{a}, \hat{b}$  与真参数  $a, b$  的值差异太大,不适合将其选作回归方程(1)的估计值。

### 3.3 $n$ 与 $T(n)$ 的计算

将GLS估计值  $\hat{a}=2.08676, \hat{b}=1.86104$  代入(2),(3)式,则得黑绒金龟的理论抽样数和Green序贯抽样模式

$$n = 2.08676 \left( \frac{t}{D} \right)^2 \bar{x}^{0.13886} \quad (10)$$

$$T(n) = \exp \{ 5.2937 - 14.3926 \ln D - 6.1963 \ln n \} \quad (11)$$

由(10),(11)式可以计算出在给定  $D$  及置信概率(可得  $t$  值)之下的  $n$  与  $T(n)$  值。这里,给定  $D=0.3$ ,置信概率为65% ( $t \approx 1$ ),计算  $n$  与  $T(n)$  的值列于表3。表3中还列出了在相同的  $D, t$  值之下,由LS估计值  $\hat{a}=2.61398, \hat{b}=1.45431$  算得的  $n$  与  $T(n)$  值。

表3 两种不同参数值下黑绒金龟理论抽样数与序贯抽样比较

| $\bar{x}$          | 0.5 | 1  | 5  | 10  | 15  | 20  |
|--------------------|-----|----|----|-----|-----|-----|
| LS • Taylor $n$    | 42  | 29 | 12 | 8   | 7   | 6   |
| GLS • Taylor $n$   | 26  | 23 | 19 | 17  | 16  | 15  |
| LS • Green $T(n)$  | 6   | 24 | 80 | 160 | 231 | 345 |
| GLS • Green $T(n)$ | 21  | 29 | 60 | 85  | 95  | 108 |

从表3可以看出,理论抽样数  $n$  随黑绒金龟种群密度的增大而减小,抽样模式  $T(n)$  值随密度的增大而增大。两种不同参数估计方法所得的参数值差异,导致了  $n$  与  $T(n)$  值的很大差异。由LS法估计的  $\hat{a}, \hat{b}$  值算得的  $n$  值随平均密度的增大而迅速减小,而由广义

LS 法算得的  $n$  值随平均密度的增大在稳定减小;前者所得的  $T(n)$  值随平均密度增大而迅速增大,而后者所得的  $T(n)$  值随密度的增大而平稳增大,这种稳定的抽样结果符合抽样理论,也是实际问题研究者所希望得到的结果。

#### 4 小 结

1)公式(6)是一类非线性回归参数的估计公式,它适用于将函数(4)线性化为函数(5),所用变换式为  $x=\varphi(x)$ ,  $Y=\psi(x,y)$  的非线性回归的参数估计,用它估计这类非线性回归可提高回归精度。

2) $a, b$  的广义 LS 估计公式(7)是一种加权 LS 估计,当权重  $k_i=S_i^4(i=1,2,\dots,n)$  差异较大时,估计的效果更好。田间害虫种群空间分布型及抽样技术研究所展示的资料,其中大部分的权重差异较大,公式(7)适合估计(1)式的参数  $a, b$ 。

3)较高精度的参数估计值不仅能使理论抽样数和抽样模式精度提高,而且用它度量种群空间分布型的聚集度,其准确性也会随之提高。由公式(7)获得较为精确的参数估计值,由公式(9)可以度量所选的参数值是偏差较小者,这从方法上保证了  $n$  与  $T(n)$  具有较高的精度。

#### 参 考 文 献

- 1 刘绍友,侯有明,胡作栋等.渭北旱塬油菜苗期蚜虫空间分布型及抽样技术研究.干旱地区农业研究,1993,11(1):28~35
- 2 谢君成.黑城金龟田间分布型及抽样技术初步研究.干旱地区农业研究,1994,12(4):57~60
- 3 杨森,王贵生,王永祥等.小麦散黑穗病空间分布型及抽样技术.植物保护学报,1995,22(3):256~258
- 4 刘光祖,史美英.一类非线性回归模型参数的广义 LS 估计.工科数学,1996,12(2):127~131
- 5 刘光祖,吕新民,史美英等.非线性回归参数估计的百分偏差度量.西北农业大学学报,1996,24(4):79~84

### armetric Estimation of Exponent in Field Pest Population Sampling

Liu Guangzu Sh Meiying Song Shide

(Department of Basic Science, Northwestern Agricultural University, Yangling, Shaanxi 712100)

**Abstract** In this paper, parameters of Tylor exponent were estimated with the generalised LS, and the estimative deviation was demermined with Box percent deviation. Thus Could effectively control the application of the parameters without highly estimative accuracy to the field pest population samling techniques.

**Key words** field pest, exponent, parametric estimation, sampling technique