

用 Pearson III型曲线拟合菜 蚜种群空间分布型的初步研究*

沈佐锐 管致和

(北京农业大学植保系)

P. Dagnelie J. Deltour

(比利时 Gembloux 农学院)

摘 要

为寻求某些研究 r -对策昆虫种群在其生境空间的概率分布的方法, 我们用 Pearson III型曲线 (简称 P-III 分布) 拟合了菜蚜种群的 9 批抽样数据, 它们是 1981 年在菜蚜种群发生期的不同日期抽取于北京郊区的秋白菜田间。

我们可以通过 MTLK 方案将 P-III 分布用于拟合 9 批样本中经过初步判断具有可拟性的 3 批数据:

M——初值估计采用矩法, 即以矩法估计 P-III 分布的三个基本参数: 平均数 Ex , 离势系数 Cv 和偏差系数 Cs , 做为下面将要进行的适线法之初始值。

T——P-III 分布的累积概率值 p 是由数表给出, 具体做法是由给定的 Cs 值和抽样数据 x 的标准化变量 Φ 值在数表中查得 p 值。

L——以适线法实现良好的拟合, 该法是在计算机上通过一项最优化技术, 即在 Cv - Cs 直角坐标系中产生一组正三角形 (简称 GRT 法), 以自动得到拟合的良好结果。

K——对拟合优度采用柯尔莫哥洛夫检验。

拟合结果, 3 批样本中有两批可理论地拟合。该 P-III 分布的概率密度函数为

$$f(x) = \frac{\beta^a}{\Gamma(a)} (x-b)^{a-1} * \exp(-\beta * (x-b)), \quad (b \leq x < \infty)$$

其中 $b = Ex * (1 - 2 * Cv / Cs)$, $a = 4 / Cs^2$, $\beta = 2 / (Ex * Cv * Cs)$, 经过拟合后, b 、 a 和 β 的取值便更加适合了。

本文得出的结论是:

1. 连续型概率分布可以被用来拟合某些 r -对策昆虫种群的田间抽样数据, 这是与传统的方法不同的。这里所谓传统的方法是指用于拟合 k -对策昆虫种群和某些 r -对策昆虫种群的田间抽样数据的方法, 例如采用 Poisson 分布、负二项分布、Neyman A 型分布等离散型分布。

2. 借助于 Pearson 曲线系统有可能描述蚜虫种群的空间分布型动态。

3. 做为初步的研究, 有些问题尚有待于进一步解决, 如现有的 GRT 法实际上是在拟合优度的非均匀梯度场中实行的。

* 论文修改过程中得到林昌善教授、丁岩钦副研究员、陈维博和徐汝梅等同志的指点和鼓励, 谨此一并致谢。

一般认为, 昆虫种群空间格局(spatial pattern)的研究方法分 3 种: 频率拟合法, 聚集指标测度法及凭借 Taylor 公式和 Iwao 公式的分析法(丁岩钦, 1980; 徐汝梅等, 1980)。后两种方法避免了第一种方法计算复杂烦琐, 有时无解或多解等问题, 显出了一定的优越性。然而, 计算机技术的长足发展, 特别是微型机的日益普及, 使我们有可能把频率拟合法从困境中解脱出来。了解了种群在其生境空间的概率分布模式毕竟有某些独到的意义, 诸如有利于进行随机过程和蒙特卡洛方法的研究, 及预报模型的检验等等。

以往关于昆虫种群空间概率分布的研究, 在方法上基本是利用有限的几种离散型理论分布, 如 Poisson 分布、二项分布、负二项分布、Neyman 分布等, 做为昆虫种群空间分布的模型。现在凭借计算机, 是否可以引入其它理论分布, 特别是把连续型分布用于 r -对策昆虫种群的研究, 这就是本文探讨的课题。

我们把菜蚜种群做为研究材料。蚜虫是典型的 r -对策昆虫(R.M. 梅, 1976)。其“机会主义对策”的特点是在综合了兼性多型、世代重迭和有翅蚜迁飞等现象的基础上形成的。其种群数量和空间分布的特点是: ①变化快; ②在其发生期中可达到相当大的数量; ③样本中个体数变量的离散程度高。显然, 在这种情况下, 就不宜采用离散性模型, 也不宜采用变通性不强的仅具一、两个参数的连续型概率分布来拟合。面对比较复杂的观测变量分布, Kendall 和 Stuart (1977) 指出, 在数学上描述它们显然需要某种变通性较强的体系 (elastic systems)。本文即探讨他们所介绍的三种主要途径之一: 利用 Pearson 曲线系统。

一、菜蚜种群的调查数据及其初步分析

本研究的基本数据来自 1981 年 9 月至 11 月在北京郊区东北旺农业科学实验站秋白菜地里的调查。该地长 38.7 米, 宽 21 米, 东靠马车道, 南接菠菜地 (9 月 7 日直播, 收获期比白菜晚), 北、西两面是玉米地 (10 月初收割)。白菜品种为小核桃纹。7 月 3 日苗床播种。8 月 27 日移栽入菜地。行向南北; 行距 56 厘米; 株距 49 厘米。田间管理由实验站负责。整个生长期中不施农药, 其他管理措施均按常规进行。

调查方法: 经缓苗、定苗后, 调查自 9 月 10 日起至 11 月 5 日收获止。原则上每周调查一次, 每次调查时间掌握在 10:30—15:30 之间(管致和, 1962)。采用集团抽样, 在随机取定的调查行内以血球计数器记录每株白菜上的蚜量 (头/株)。

京郊秋白菜上的蚜群系由桃蚜 (*Myzus persicae*) 和萝卜蚜 (*Lipaphis erysimi*) 组成。每种蚜虫都有翅型和无翅型, 每种翅型又可区别出成蚜和若蚜 (暂不更细地鉴定若蚜的龄期) (管致和, 1962)。这样, 种群中就有 8 个基本组份。数据可供我们研究该种群的结构动态。但本文仅从蚜虫种群整体着手, 研究其空间分布。

共取得 9 个日期的样本, 各日期的样本量由于取样时间的约束和种群密度的影响而不等, 但都在 180 头/株以上。这些变量组成的样本可视为来自 9 个概率分布总体, 把它们串起来研究便可得到对菜蚜种群田间概率分布动态的了解。为了粗略地描述此动态, 可先用矩法估计这些分布的形状参数:

E_x ——样本均数, 即蚜群密度估计量;

C_v ——离势系数, 或称变异系数;

C_s ——偏差系数;

C_e ——峰度系数。

结果列于表 1。

表 1 各样本的分布参数 (1981)
table 1 distribution parameters of the samples

样 本	抽 样 量	Ex	C_v	C_s	C_e	$I\delta$
9.10	521	3.3474	0.8413	1.4381	2.7681	1.5370
9.14	523	19.669	0.8008	1.9263	6.9962	1.5546
9.17	475	67.712	0.7920	3.1856	15.837	1.7669
9.24	183	343.65	1.230	7.0424	66.839	2.4488
10.2	186	596.22	0.7313	1.3156	1.7029	1.5362
10.10	188	574.25	0.8574	2.9921	16.233	1.7489
10.22	264	390.83	1.037	3.6990	19.596	2.5396
10.31	227	378.11	0.8920	2.6852	8.3987	1.7984
11.5	218	353.24	0.8276	2.0190	4.3953	1.6803

该表中,

- Ex 值反映了这些分布的重心是按 Sd 曲线变化的。
- C_v 值是这些分布对于 Ex 的相对离散程度的测度。
- C_s 值均为正, 说明这些分布都是正偏。
- C_e 值均大于零, 说明这些分布的峰都比正态分布的峰要宽。

表中还给出了 Morisita 指数 $I\delta$ 值, 可见各日期样本的 $I\delta$ 值都大于 1, 说明这些分布都是聚集的。

然而, 上述处理并没有分析出来它们具体服从何种概率分布, 我们将用 Pearson 曲线系进一步探讨。

二、用 P -Ⅲ 分布拟合菜蚜种群田间抽样数据的方法和结果

利用 Pearson 曲线系的矩比进行初步判断 (Dagnelie, 1973; Johnson 等, 1963; Pearson 等, 1972), 9 月 14 日和 10 月 2 日的样本数据有可能是 Pearson Ⅲ 型分布 (简称 P -Ⅲ 分布)。

P -Ⅲ 分布具有概率密度函数

$$f(x) = \frac{\beta^a}{\Gamma(a)} (x-b)^{a-1} * \exp(-\beta(x-b)), (b \leq x < \infty) \quad (2.1)$$

其中 $a, \beta > 0$, 两者连同 b 称为原始参数。这是三参数分布。

我们用适线法拟合该分布, 该法具有 4 个要素: M, T, L 和 K 。

(1) 要素 M

即用矩法 (method of moment) 计算样本的数字特征:

$$Ex_0 = \bar{x} = \frac{1}{N} \sum x_i \quad (2.2)$$

$$Cv_0 = \frac{\sigma}{\bar{x}} = \sqrt{\frac{1}{N-1} \sum (Ki - 1)^2} \quad (2.3)$$

$$Cs_0 = \frac{\mu_3}{\sigma^3} = \frac{\sum (Ki - 1)^3}{(N-3)Cv^3} \quad (2.4)$$

式中 $Ki = \frac{x_i}{\bar{x}}$, 下标 0 表示它们是用矩法估计的。

Ex , Cv 和 Cs 通常称为基本参数, 它们与原始参数的关系为

$$b = Ex \left(1 - \frac{2Cv}{Cs} \right) \quad (2.5)$$

$$a = \frac{4}{Cs^2} \quad (2.6)$$

$$\beta = \frac{2}{Ex \cdot Cv \cdot Cs} \quad (2.7)$$

(2) 要素 T

即用数表(numerical table)给出 P-III 分布的函数值。华东水利学院(1980)提供了 P-III 分布 Φ 值表, 由表中 Cs 值和标准化变量 Φ , 可查得相应的超过累积概率 pi 。

$$\Phi_i = \frac{x_i - \bar{x}}{\sigma} = \frac{x_i - Ex}{Ex \cdot Cv} \quad (2.8)$$

(3) 要素 L

即用适线法(line-fitted method)对 Cv 和 Cs 做更准确的估计。这是因为用矩法估计时, 高阶矩参数误差较大。

首先将样本数据由大至小排列:

$$X = (x_1, x_2, \dots, x_N)。$$

样本分布函数定义为

$$S_N(x_i) = \frac{X \text{ 中大于或等于 } x_i \text{ 的数据个数}}{N} \quad (2.9)$$

然后在概率纸上做出 $S_N(X)$ 曲线。如图 1 所示。这就是将用适线法拟合的目标曲线。其数学形式为

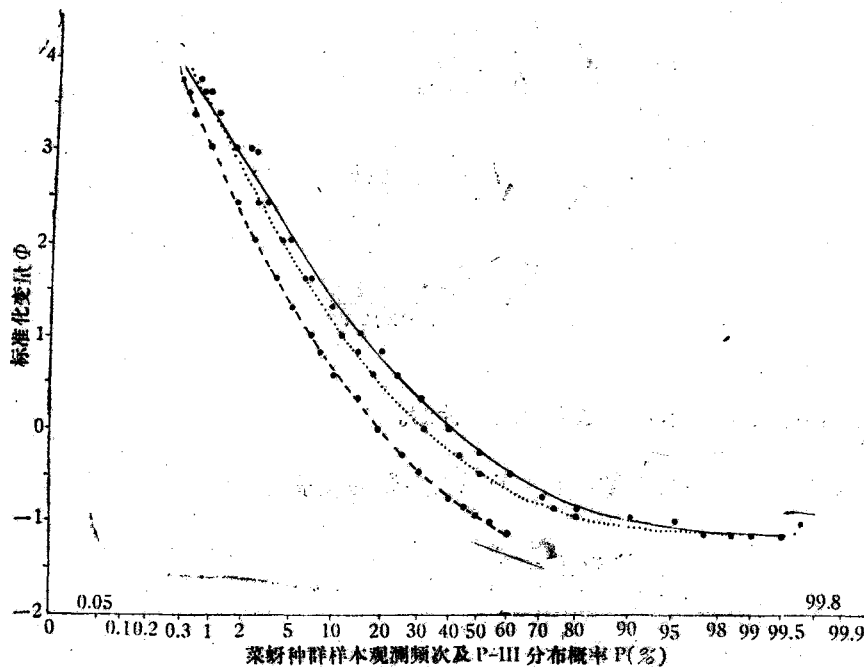


图1 观测曲线与P-Ⅲ分布曲线的比较(10月2日样本)
fig.1 the observed curve and fitted curves by the methods of moment
and of line-fitted, for sample(10.2)

— 观测频次累积曲线
---- 由矩法给出的P-Ⅲ分布曲线
..... 由适线法给出的P-Ⅲ分布曲线

$$S_N(X) = \begin{cases} 0, & \text{若 } X > x_1; \\ \frac{i}{N}, & \text{若 } x_i \geq X > x_{i+1}; (i=1, 2, \dots, N-1) \\ 1, & \text{若 } X \leq x_N \end{cases}$$

它是随机变量分布函数 $p(x)$ 的近似形式。当 $N \rightarrow \infty$, 有 $S_N(x) \rightarrow p(x) \rightarrow 1$ 。

现以 (Ex_0, Cv_0, Cs_0) 为初值进行适线。即在式(2.8)中代入 Ex_0 和 Cv_0 , 计算出 Φ_i , 再于P-Ⅲ分布 Φ 值表中 $Cs = Cs_0$ 一行查得相应的 P_i 值。查表过程中往往需要一两次内插, 可采用拉格朗日插值法。

用这样得到的 (Φ, p) 点的集合在概率纸上做图, 即得到相应的理论分布曲线。一开始它往往不与目标曲线拟合, 故需不断调整 Cv 和 Cs , 直至拟合良好。

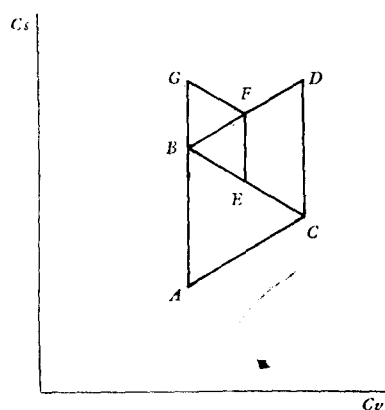
为使这样的最优化过程在计算机上自动实现, 本文介绍一种正三角形最优化搜索法(简称GRT法)。图2给出了其说明。

(4) 要素K

即用Kolmogrov检验来判定适线的良好性。

其要点是计算

$$D_N = \text{MAX} \{ |S_N(x_i) - p(\Phi_i)| \} \\ 1 \leq i \leq N$$

图 2 以 GRT 法寻找最优 (Cv, Cs) 点的过程示意fig. 2 process of seeking optimum of Cv - Cs pairs by the GRT method

A 为用矩法求得的初始点 (Cv_0, Cs_0) 。

$B = A \times CL$, CL 为一经验系数, 或称搜索跨度。

C 是以 A、B 为顶点的正三角形的第三个顶点, 由于由矩法估计的 Cv_0 和 Cs_0 往往偏低, 所以 B、C 通常取在 Cv 、 Cs 增大的方向。

D 是点 B、C 的中点, 且 B、C 优于 A 时所选定的。

E 和 F: 比较 B、C、D 三点, 若 B 优于 D, 则 E 取在边 $\overline{BC}$ 的中点, F 取在 $\overline{BD}$ 的中点。

G 是 B 和 F 优于 E 时所确定的点, 也是正三角形的一个顶点。如此类推, 直至结束。结束的判据有两个: ①符合 kolmogorov 检验; ②达到计算所要求的精度, 这里规定最后两点的距离等于或小于 0.01。

然后同

$$D_\alpha = \frac{1.63}{\sqrt{N}}$$

比较, D_α 是当 $N > 50$ 时该检验在显著水平 $\alpha = 0.01$ 时的临界值。若 $D_N < D_\alpha$, 则在 $\alpha = 0.01$ 显著水平上没有理由说明拟合得不好。

根据以上 4 要素编制的程序是用 TRS-80 I 型机配用的 Level II BASIC 语言写成的。利用该程序, 我们对 9 个日期的样本逐一进行了拟合。表 2 为拟合结果, 证实了初步判断。表中 Exg 、 Cvg 和 Csg 是由适线法得到的 P-III 分布基本参数的估计量。

表 2 P-III 分布拟合数据的结果 (1981)

table 2 results of the fit of P-III distribution to data

样 本	Ex	Cv	Cs	Exg	Cvg	Csg	解的存在
9.10	3.3474	0.8413	1.4381	—	—	—	无
9.14	19.6692	0.8008	1.9263	19.6692	0.8008	1.9725	有
9.17	67.7116	0.7920	3.1856	—	—	—	无
9.24	343.653	1.230	7.0425	—	—	—	无
10.2	596.221	0.7313	1.3156	596.221	0.7313	1.8181	有
10.10	574.245	0.8574	2.9921	—	—	—	无
10.22	390.833	1.037	3.6990	—	—	—	无
10.31	378.110	0.8920	2.6582	—	—	—	无
11.5	353.344	0.8276	2.0190	—	—	—	无

由式(2.5)~(2.7), 可得到9月14日和10月2日样本的P-Ⅲ分布的原始参数。

图1把矩法和适线法给出的P-Ⅲ分布曲线同实际观测数据曲线做了对照。

三、问题和讨论

1. 9月14日和10月2日样本的拟合过程说明, 尽管我们采用的最优化方法是同时调整 Cv 和 Cs 的, 但在最终的拟合值中 Cv 又返回到 Cv_0 , 即所需调整的仅仅是 Cs_0 。这一点证实, 从样本数据用矩法计算得到的高阶矩参数, 误差往往较大, 以致值得怀疑它做为总体参数估计量时的准确度。这个准确度是可以通过增加抽样量得到提高的。对此可采用 Monte Carlo 方法进一步研究(丛树铮等, 1980; Solomon等, 1978)。就P-Ⅲ分布, 我们所做Monte Carlo 试验的结果表明, 在本研究中所用的样本容量, 对于用矩法算得的平均数 $\bar{X}$ 和变异系数 Cv 的准确度已可满足要求; 但对于偏差系数 Cs , 则达到准确度要求所需的样本容量起码应在600以上。那么, 对于容量显著小于600的样本, 用文中介绍的MTLK法拟合的结果, Cs 值的修正确实更大些(见表2)。

2. 本文介绍的GRT法尚需改进。在具体拟合过程中我们发现, 该法存在不稳定性问题。具体表现是, 在图2的 $Cv-Cs$ 平面上寻优的过程中相邻两点, 甚至距离很近的两点, 其拟合优度(goodness of fit)有时相差甚大。 $Cv-Cs$ 平面绝不是均匀的“优度梯度场”(Shen, 1983)。

3. 本文的结果仅表明, 菜蚜种群发展的某些阶段, 其空间分布可能服从P-Ⅲ分布; Shen (1983)的研究表明, 在另一些发展阶段, 则可能服从Pearson曲线系中其它类型的分布。换言之, 利用Pearson曲线系有可能描述菜蚜种群在其发生期中空间分布的变化动态。

参 考 文 献

- 丁岩钦 1980 昆虫种群数学生态学原理及应用。科学出版社, 第113—124页。
- 丛树铮、谭维炎等 1980 水文频率计算中参数估计方法的统计试验研究。水利学报(3): 1—15。
- 徐汝梅、李兆华 1980 温室白粉虱成虫空间分布型的研究。昆虫学报23(3): 265—275。
- 管致和 1962 京郊菜蚜发生规律的研究初报。植物保护学报1(2): 23—32。
- 华东水利学院主编 1980 水文学的概率统计基础。水利出版社。第8.2.3和8.3.3节及附录I表4。
- R.M.梅(孙儒泳等译) 1976 理论生态学。第3章。科学出版社。
- Dagnelie, P. 1973 Théorie et méthodes statistiques. Vol.1. Les presses agronomiques de Gembloux, A.S.B. L. Gembloux, Belgique, p.378.
- Johnson, N.L. et al. 1963 Table of percentage points of Pearson curves, for given $\sqrt{\beta_1}$ and β_2 , expressed in standard measure. *Biometrika* 50(3 and 4): 459—462.
- Kendall, M. and A. Stuart 1977 The Advanced Theory of Statistics. 1: Distribution Theory, 4th ed. London, Griffin, Sec. 6.1.
- Pearson, E.S. and H.O. Hartley 1972 Biometrika Tables for Statisticians, V.2. Cambridge, pp. 75—77.
- Shen, Z.R. 1983 Methodological discussion about analysis of aphid population dynamics on Chinese cabbage in the first stage of research. Scientific report of stagiaire, Faculté des Sciences Agronomiques de l'Etat, 5800 Gembloux, Belgique. (unpublished).
- Solomon, H. and M.A. Stephens 1978 Approximations to density functions using Pearson curves. *J. Amer. Statis. Assoc.* 73(361): 153—160.

PRELIMINARY APPLICATION OF PEASON CURVE SYSTEM TO STUDY ON PROBABILITY DISTRIBUTIONS OF A CABBAGE APHID POPULATION IN ITS INHABITED SPACE

Shen Zuorui Guan Zhihe
(Dept. of Plant Protection, Beijing
Agricultural Univ., Beijing, China)

P. Dagnelie J. Deltour
(Faculté des Sciences Agronomiques
de l'Etat, 5800 Gembloux, Belgique)

To find some methods for studying probability distributions of r -strategic insect populations in their inhabited spaces, the type III of peason curve system (the P-III distribution) is used to fit for 9 samples obtained at different dates during a mixed population of *Myzus persicae* Sulzer and *Lipaphis erysimi* Kalt-enbach grew on Chinese cabbage in the suburbs of Beijing in autumn, 1981.

The P-III distribution may be fitted to the data of three samples with pre-judged fittability among the 9 samples, in terms of the MTLK scheme;

M—the moment method to estimate the three fundamental parameters of the P-III distribution, the mean Ex , the coefficient of variation Cv and the coefficient of skewness Cs , as first fitted values;

T—the numerical table giving P-values, cumulative probabilities of the P-III distribution according to given Cs and Φ -values, standardized variables of the sample data, x 's.

L—the line-fitted method to realize a good fit which is reached automatically on a computer, with an optimization technique of generating a set of regular triangles in the coordinate system of Cv and Cs (the GRT method, abbr.).

K—the Kolmogorov test determining the goodness of fit.

As results, two of the 3 samples are fitted theoretically. And with $b = Ex * (1 - 2 * Cv / Cs)$, $a = 4 / Cs^2$ and $\beta = 2 / (Ex * Cv * Cs)$, the probability density function of P-III distribution is valued more properly:

$$f(x) = \frac{\beta^a}{\Gamma(a)} (x-b)^{a-1} \exp[-\beta(x-b)], \quad (b \leq x < \infty),$$

The report comes to the conclusion that:

1. The continuous probability distribution can be used in fit of some field sample data of r -strategic insect populations as contrasted with conventional methods in fit of that of K -strategic insect populations and some r -strategic ones by discrete distributions, as such the Poisson, the negative binomial and the A-type of Neymans', etc.

2. It is possible to describe the spatial distribution pattern dynamics of the aphid populations, to certain extent at least, with the aid of the Pearson curve system.

3. As a preliminary research, there are quite a few problems to be solved further, for example, the GRT method actually running in a non-uniform gradient field of goodness of fit.