

第十章 生态学家在系统分析中的任务

(记 录)

阳含熙1980.9.24.

系统分析科学研究,从历史上讲是一种新的研究方式,把系统看作一个整体,它的结构,功能以及功能结构间的关系。首先要把整体定义下来,是怎样一个单位,即境界问题。

整体研究的重点是主要部分的相互关系之间的研究。这是一种综合的多学科的研究。作为生态系统研究来讲牵涉到的主要部分是生物,与其周围环境之间的相互关系的研究。对于整体的各个部分的相互关系采取多学科,多兵种的途径,建立一个数学的模型。到底怎样建立数学模型,今天下午来谈谈这个问题。生态学家作为主要组织者必须要在专业知识之外还要具备一定的数学知识,才能参与建立数学模型,并了解这个数学模型能否达到研究目的。而这点国内外都认为是一项困难的工作。今天我向大家汇报一下自己的初步看法。

生态学家从广义上讲生物学家,对数学知识的应用时间是很长的。数理统计研究最先是生物科学开始的。统计学科发展的历史可分为几个阶段:

1890年—1940年大约五十年时间是统计学科的形成阶段。这个时期统计学研究主要在英国与生物学科有关的研究机构进行的。最著名的两个人,一个是这五十年中的前二十五年,大约1890年到1915年以前的阶段,卡尔·皮尔生(KARL·PEARSON)提出了最有名的“相关系数”的公式,现在人人都知道了。从性质上决定两个实体或性质相互间的关系,而且通过这个公式可以找出两种实体或性质上的关系。并且通过“相关系数”的研究从一种实体(性质)来估量另一种实体(性质),这就是所谓“回归”(REGRESSION)的研究。它最大的功用是可以从一个容易观测的实体(性质)来估量另外一个难以观测或研究的实体(性质)。这个发现对当时来讲是一个很大的促进。

在后二十五年有一个最重要的代表人物是阿·爱·费雪(R·A·FISHER),工作地点在英国最古老有名的ROTHAMSTED农业实验研究站,当初实验站成立已好几十年了,积累了大量的数据,无法处理和进一步解释,就找了这样一个数学家,建立了数理统计这个学科,包括现在大家都知道的几个部分,其中一个就是取样的原理和知识。有随机取样,机械取样等等,现在这些老的方法还在应用,原理是正确的。另一方面,他们建立了许多实验设计方法即:随机区组、拉丁方、裂区实验等等。五十年代开始出现的上边的方法更进步的方法,即析因试验。这一系列的许许多多的实验方法。第三部分是他们提出了一系列分析数据的方法,其中最完整的是“方差分析”(老的翻译名称叫“变量分析”)。以上这些部分组合起来之后,对当时的农业科学研究,包括栽培

试验,肥料试验,育种试验起了很大推动作用,也取得了许多成果。许多新兴的推广品种都是在以上一整套试验设计数据分析方法中取得的,包括中国的小麦2905,水稻的中农28,以及解放后西北的碧蚂小麦,这一系列都是他们创造的方法取得的成果。

数理统计的发展这两个阶段,在今天看起来有局限性,在那些方面呢?我们试图来分析一下:

1.统计科学规定的从取样到分析的总过程中的许多方法,有一些基本的理论假设是很不容易满足的。比如用得最多最普遍的“方差分析”假定了必须具备以下五个条件:

<1>数据要随机取得,这在大规模的实验室,小面积农作物按测定设计易作到,在大面积的资源调查中经常碰到一些问题,不易在野外随机取得数据。<2>必须假定数据是独立的不相互干扰的;如肥料实验,用四种肥料看对作物的效应,用四种肥料就有四种处理,如果把一种处理都排在一块地块上,把另外一种处理放在另外一块地块上,这显然就犯了错误。因为,地块与地块之间有差异性,这个差异就要干扰到你所研究的处理。因为你想找出不同肥料的效应,而地块本身的差异性就影响了结果。因而这个数据得出来是不可靠的。有一个英国的生态学家举过一个例子,他是在1936年写这篇文章的。他说象现在这样调查一个植物对土壤PH的反应的方法是不可靠的。如果你找到一个植物,在植物下边挖它的土壤,测这土壤的PH值,就说这个植物最适于在哪种PH值中生长。如果这样做,就变成电杆也能找出它的最宜PH值。错误在那里呢?首先,没有先找到地块之中PH值的频率分布,如果知道了地块的原来PH值的频率分布,而后找出植物生活生长的土壤PH值频率分布,这两个分布互相比较之后才能找出植物最适于生活在哪种PH值。我很同意杰弗斯的意见,同志们反映意见时讲,不要再讲取样了,不要再讲实验设计了,杰弗斯讲:“那不行!包括美国在内,在这方面犯的错误很多,很多,我一定要讲。”这是他的一种看法,同志们可有保留意见。不过,我是同意杰弗斯的意见的。因为,好的数学方法对坏的数据起不到任何作用,唯一作用是给了它一个漂亮形式而已,而这种结果实际上是“欺骗”。任何高级数学方法不可能增加数据本身任何性质,而对数据本身的设计取样有错误的话,得出来的结论肯定是错误的。不能因为数学方法得出任何改进。<2>所有的数据必须是独立的,用我们数学语言来讲,同一类数据它的方差应该是零的,如果不是同一类数据它们的协方差应该是零。这一条做起来是不容易的。这是第二个条件。<3>方差分析数据要求方差是齐性的(HOMOGENEITY of VARIANCE)这一点是不容易的。有许多数据,方差是不一样的,方差的齐性,数据处理专门有一些方法,叫转换(TRANSFORMATION),方差差异比较大,经过数据处理改进它的缺点,如果不能达到这个目的,就不能用这个方法。有许多检验的方法,最简单的方法是检验它的变异系数,这是粗略的检验方法。<4>数据要求正态分布,数据表现出一个钟形曲线,如果数据不是这种情况,要用转换方式法。<5>数据必须是可加性的。什么叫可加性,数据是同一个测量单位测定的,就叫可公度性(COMMEASURABILITY)。数据是可公度的,而且又是独立的,这些数据就叫具有可加性,必须满足这么几个条件,否则方差分析得出结果有偏差。而这些条件有时是不能都满足的。

2.早期的数理统计方法的计算工具是台式计算机。这样的台式计算机与电子计算机无法比拟,计算量是有限的,金陵大学一百台以上台式计算机天天计算数据,每年用做

全国农作物情报报告。今天，只用一台电子计算机就行了。任何科学的发展必须先改进工具，在方法上，工具上都要有所突破之后，随后才是理论上的突破。

3. 早期的试验设计面积小，每一小区一个平方米到几个平方米，面积都很狭小，这样小的面积取得的结果，要在大面积上来应用是有限制的。

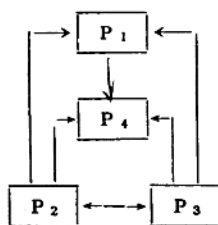
4. 时间过长。一个试验，从做实验到生产中推广应用，中间经过这样几个阶段：田间实验，区域实验，常常十几年以上时间。

5. 试验因子有限，如超过四个因子，拉丁方实验就要有许多排列，有很多困难。尽管早期数理统计方面取得了许许多多成果，但不是令人十分满意的。解放后，有许多人把数理统计抛弃掉，因数理统计有局限性。这些不但反应在数理统计上，也反映在许多其它数学方法上。如早期的聚类分析的八种方法，有许多现在不再用了。又如主分量方法，在1930年代这方法提出来应用在心理学上和人类学上，计算非常费时。现在，我们用电子计算机通常就只需半小时左右，许许多多计算方法，没有先进工具就没办法推广。尽管数理统计学首先是从生物学发展起来的，而且应用在生物科学中，有一个时期生物统计学BIOMETRY与统计学STATISTICS几乎是同义语，现在分开了。但从四十年代之后，数理统计大量的应用在物理学，化学上。特别是物理学在这方面应用的越来越多，因为它不象生命现象那样复杂，六五年后出现了一个时代，有人把统计学的这一阶段定为1965—1990年。电子计算机最早于1946年发明，用在生物学上是1960年代，大量应用是1965年以后，现在是稳定性发展。许多国家，为了适应这种情况，把六十年代之前毕业的大学生重新回炉进行数学训练，学习电子计算机。英国朋友讲生物学家对数学感到非常困难，在一些国际会议上，数学家坐在一起，生态学家或生物学家坐在一起，双方找不出一个共同语言。但是，这种情况逐步有改进，生态学家要应用数学这个工具，自己也要走一段路程，走三分之一或一半以上的路程，了解一些数学方法的特点，它的理论根据和应用范围。

在生物学上常用到的曲线方程就有下面几种，我们还可以举出一些线性代数应用的例子，也并不十分复杂。

<1> 生态系统

这里：



$$\begin{pmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

<2> 曲线方程的拟合

$$(0, 1), (1, 3), (2, 4), (3, 4); M = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix}$$

$$M'M = \begin{pmatrix} 4 & 6 \\ 6 & 14 \end{pmatrix}$$

$$(M'M)^{-1} = \frac{1}{10} \begin{pmatrix} 7 & -3 \\ -3 & 2 \end{pmatrix}$$

$$V = (M'M)^{-1}M'Y = \frac{1}{10} \begin{pmatrix} 7 & -3 \\ -3 & 2 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \end{pmatrix} \begin{pmatrix} 1 \\ 3 \\ 4 \\ 4 \end{pmatrix}$$

$$= \begin{pmatrix} 1.5 \\ 1 \end{pmatrix}$$

$$Y = 1.5 + x$$

<3> 森林演替的马尔柯夫过程:

转移矩阵 杨树 桦树

$$P = \begin{pmatrix} .8 & .3 \\ .2 & .7 \end{pmatrix} \quad \begin{array}{l} \text{活80\%} \\ \text{死20\%} \end{array}$$

初始向量

$$X^0 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad X^1 = PX^0 = \begin{pmatrix} .8 & .3 \\ .2 & .7 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} .3 \\ .7 \end{pmatrix}$$

$$X^2 = PX^1 = \begin{pmatrix} .8 & .3 \\ .2 & .7 \end{pmatrix} \begin{pmatrix} .3 \\ .7 \end{pmatrix} = \begin{pmatrix} .45 \\ .55 \end{pmatrix}$$

$$X^3 = PX^2 = \begin{pmatrix} .8 & .3 \\ .2 & .7 \end{pmatrix} \begin{pmatrix} .45 \\ .55 \end{pmatrix} = \begin{pmatrix} .525 \\ .475 \end{pmatrix}$$

$$X^4 = \begin{pmatrix} .563 \\ .438 \end{pmatrix}$$

$$X^5 = \begin{pmatrix} .581 \\ .419 \end{pmatrix}$$

$$X^6 = \begin{pmatrix} .591 \\ .409 \end{pmatrix}$$

$$X^7 = \begin{pmatrix} .595 \\ .405 \end{pmatrix}$$

$$X^8 = \begin{pmatrix} .598 \\ .402 \end{pmatrix}$$

$$X^9 = \begin{pmatrix} .599 \\ .401 \end{pmatrix}$$

$$X^{10} = \begin{pmatrix} .599 \\ .401 \end{pmatrix}$$

$$X^{11} = \begin{pmatrix} .600 \\ .400 \end{pmatrix}$$

< 4 > 生态学常用的曲线 (曲线⑤略)

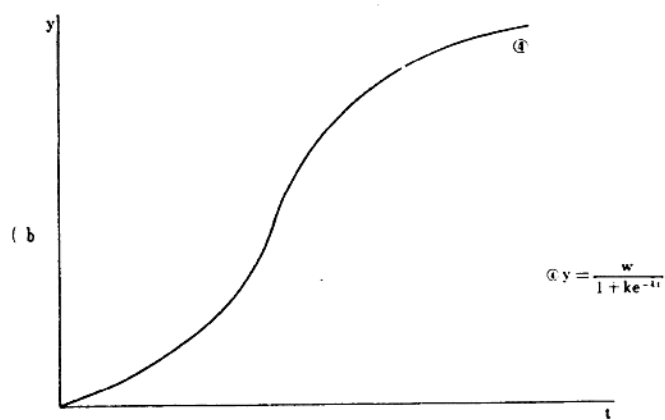
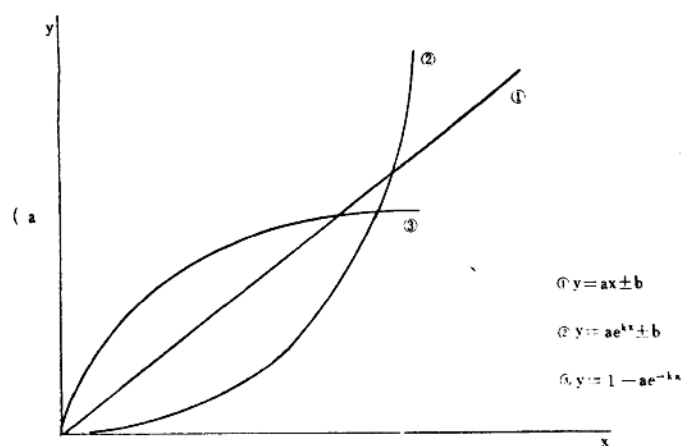


图10.1

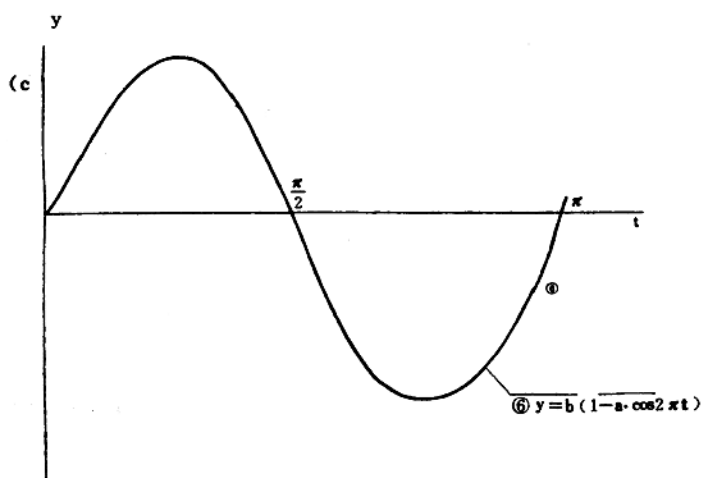


图10.2