

生物科学参考资料

第十七集

科学出版社

内 容 简 介

本集为生物数学专集,共收集了19篇文章。其中11篇较系统地介绍了生物数学及其分支学科,如数量分类学、数量遗传学、数量生理学、多变量统计分析、模糊数学、电子计算机及其在生物医学中的应用。其余8篇系室分析、因子分析中的均差最大旋转标准、多因素组群观测的回归模型、岭判别分析、逐步判别分析等数学模型或方法在生物医学中的应用实例,其中有的方法为我国科技工作者的研究成果。读者可通过本专集初步了解生物数学概貌,并开阔思路。

本专集可供生物、医药、农林、应用数学专业的研究人员及有关大专院校的师生参考。

生物科学参考资料

第十七集

责任编辑 马素卿

科学出版社出版

北京朝阳门内大街137号

中国科学院印刷厂印刷

新华书店北京发行所发行 各地新华书店经售

*

1983年2月第一版 开本:787×1092 1/16

1983年2月第一次印刷 印张:11

印数:0001—3,320 字数:252,000

统一书号:13031·2156

本社书号:2948·13—10

定价:1.75元

目 录

生物数学的发展.....	(1)
植物数量遗传学研究的成就和方向.....	(12)
科学的分类方法——数量分类学.....	(16)
数量生态学简介.....	(27)
模糊数学及其在生命科学中的应用.....	(42)
多变量统计分析在医学、生物学中的若干应用.....	(50)
电子计算机医学应用研究.....	(56)
计算机在生命科学中的一些应用.....	(65)
模拟: 植物生理学研究中的一个工具.....	(74)
预报作物产量的植物生理过程模拟模型的制定.....	(82)
室分析及其在医药学研究中的应用.....	(88)
药物体内作用的数学模型.....	(104)
无限制两代谢区的同一核素两种标记物参入的示踪动力学数学模型.....	(112)
因子分析中的均差最大旋转标准.....	(126)
多因素组群观测的回归模型及其在肿瘤流行病学中的应用.....	(136)
岭判别分析及其在预测疾病转归方面的应用.....	(149)
肺部球形病变计算机诊断——介绍一种离散变量逐步判别法.....	(155)
逐步判别分析在脉图诊断心血管疾病中应用的初步报告.....	(162)

生物数学的发展

杨纪珂

徐克学

(中国科技大学) (山西省晋东南地区药品检验所)

“数学的应用：在固体力学中是绝对的，在气体力学中是近似的，在液体力学中已经比较困难了；在物理学中多半是尝试性的和相对的；在化学中是最简单的一次方程式；在生物学中 $= 0$ 。”^[1] 这就是恩格斯在《自然辩证法》中的一段叙述。一个多世纪以后的今天，数学不仅广泛地应用于生物学，而且生物数学已作为一门独立学科耸立在科学的丛林之中，对生物学及其边缘学科的发展产生深远的影响。

二十世纪自然科学中的许多学科向生物学渗透，各种边缘学科相继兴起。因为数学是自然科学的基础，这一点使生物数学得天独厚。虽然生物数学起步较晚，它却比其它边缘学科更重要。马克思曾经指出，一种科学只有在成功地运用数学时，才算达到了真正完善的地步。德国一位生物统计学家说得更生动：“生物学 (Biology) 这个名词来源于希腊字 ($\betaίος$ 即生命)，这门科学由于应用了数学，获得第二次生命。”^[2] 应该再补上一句：数学，当它转向生物学，这门非生命科学将获得生命。

年青的生物数学正在突飞猛进地向前发展。要想展望它的未来，还是让我们先回顾一下它的过去吧。

数学与生物学互相渗透始于十九世纪末、二十世纪初，发展速度愈来愈快。请看下面对 4500 篇生物数学论文的统计表：^[3]

年 份	1870	1880	1890	1900	1910	1920	1930	1940	1950	1960
占百分比	0.4	0.7	3.7	6.3	4.4	11.4	29.6	12.5	31.0	

50 年代应用电子计算机技术后，生物数学发展得更快。请看统计表*：

年：	1950	1960	1970	1977
出版生物数学杂志种数：	5	11	23	42(1976)
平均每年召开有关生物数学学术会议次数：	0.2	1.7	5.1	

生物数学的发展经历了不同的历史时期。

1901 年皮尔逊创办“生物统计学” (Biometrika) 杂志，标志着数学开始向生物学渗透。生物数学的发展就是从生物统计学开始的。

1939 年“数学生物物理学通报” (Bulletin of Mathematical Biophysics) 杂志创办，创办人拉舍夫斯基 (N. Rashevsky) 把数学物理方法引进生物学，标志着生物数学发展的第二个里程碑。

* 根据中国科学院图书馆和中国科学技术资料情报研究所收藏图书资料，于 1978 年 11 月统计。

50年代电子计算机广泛应用,生物数学的发展又进入了一个新的阶段。由于生命现象的复杂性,生物数学需要进行大量运算,电子计算机的应用,使得许多生物数学问题的求解成为可能,为数学在生物学中的应用打开了新的局面,促使生物数学飞速发展。因此电子计算机是当代生物数学发展的基础。

当前,生物数学已经上升为一门独立的学科。74年联合国教科文组织编制的学科分类目录,已经将生物数学列为一门独立学科,归入生命科学类,与生物化学、生物物理鼎足而立。在这门学科之下,分支学科颇多。依使用的数学工具来划分,有生物统计、生物运筹、生物信息论、生物方程、生物控制论、系统分析、分类分析和生物几何学等。据在生物学中的应用来区别,生物数学包括数量遗传、数量生态、数量分类、数量生理和生物力学等。在生物数学领域内,许多分支如雨后春笋般地蓬勃兴起,新的分支还在不断涌现,表现异常活跃。

本文将着重从数学方法上讨论生物数学的发展,并对有关分支作简要介绍。

一、从生物物理到生物数学

对于生命现象的研究,人们最初把复杂的生命物质运动归结为简单的化学和物理学运动,利用比较成熟的物理学和化学知识总结生命现象的规律性。Thompson 在其巨著《论生长与形状》^[4]中,开始大量应用物理学与数学方法去研究生物,他的工作局限于对生物作静止的描绘,而 Lotka 则进了一步^[5,6],尝试用物理方法说明生命活动的规律。到了拉舍夫斯基,生物物理方法在生物学中的应用已更全面、更成熟。随着生物物理方法的日趋成熟,引用的数学工具愈多。生物物理与生物数学也就成为一对姐妹学科。

许多生物物理学的先驱者,把物理方法引进到生物学,却未料到获得的结果竟是生物数学的发展。基于这个事实,1924年 Lotka 的《物理生物学原理》(Elements of Physical Biology)^[5],在1956年再版缩写本时,不得不改名为《数学生物学原理》(Elements of Mathematical Biology)^[6]。1972年拉舍夫斯基也把他的杂志改名为《数学生物学通报》(Bulletin of Mathematical Biology)。生物物理学的发展导致生物数学的发展,这是30年代以后生物物理学发展的一种饶有兴趣的趋势。

这个事实重演了过去物理学、化学以及其它自然科学的发展导致数学发展的历史。这是不足为奇的,因为数学是一切自然科学的基础,数学本身就是从其它科学的实践中不断发展、丰富和抽象出来的。这种关系,物理学与数学之间表现得尤其明显。物理学的基本理论,从牛顿的力学原理开始,有热力学原理、电动力学原理、量子力学原理、爱因斯坦的相对论直到今天的基本粒子理论。物理学对客观世界认识得愈深刻,它所引用的数学就愈多。近代物理学的原理也开始向生物学渗透,出现了生物热力学、量子生物学。从物理学的观点对生命现象揭示得愈深刻,引用的数学也愈多,生物数学终于成为研究生物学不可缺少的工具。

数学物理方法在生物学中的应用产生了数学模型方法,成为研究生物学的重要手段。

举一个例子,著名的 Lotka-Volterra 方程,是描写捕食与被捕食生态关系的数学模型:

$$\frac{dN_1}{dt} = (a_1 - b_1 N_2) N_1$$

$$\frac{dN_2}{dt} = (-a_2 + b_2N_1)N_2,$$

其中变量 N_1 和 N_2 分别表示被捕食与捕食生物种群的大小; a_1 、 a_2 、 b_1 和 b_2 都是大于 0 的常数。我们用一组微分方程,说明种群变化应该遵从的规律。为了认识这种生态关系,只需对方程的解进行讨论。方程的解是:

$$a_2 \ln N_1 - b_2N_1 + a_1 \ln N_2 - b_1N_2 = C,$$

不同的常数 C 意味着不同的初始条件,可以画出它的曲线簇。从曲线的形状可以看到两个种群的消涨,按周期性此起彼伏变化的情况。还可以根据方程的解,算出种群变化的周期、振幅和平衡位置等等。从这个例子可以看到,一个生态学问题,可以借助数学方法作非常深刻、细致的分析。

能够进行这样分析的关键,在于是否能找到描述这种生物现象的数学模型。数学模型可以把一个生物现象抽象为数学问题。应用数学工具对数学模型进行分析,便能达到研究生命现象的目的,复杂的生命活动得到精确的模拟分析。

任何一个数学公式所表示的数学关系,只要能够反映客观世界物质运动的规律,都可以看作是数学模型。在非生命科学中,一种数学关系可以描述诸多的现象。譬如拉普拉斯方程 $\nabla^2 u = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} = 0$, 可以模拟热的传导、流体的流动、分子的扩散或者刻画电场的规律,一个简单的方程可以把许多物理现象囊括其中。在生物学中却非如此简单。就拿 Lotka-Volterra 方程所代表的生态关系来说,如果种群不是完全封闭的,考虑到有新的捕食者迁入,方程变为^[7]

$$\frac{dN_1}{dt} = a_1N_1 - b_1N_1N_2 + v$$

$$\frac{dN_2}{dt} = -a_2N_2 + b_2N_1N_2 \quad \circ$$

倘若计及种群内部存在自我干扰,方程又该写成^[8]

$$\frac{dN_1}{dt} = \left(a_1 - \frac{a_1}{\theta} N_1 - b_1N_2 \right) N_1$$

$$\frac{dN_2}{dt} = -(a_2 - b_2N_1)N_2 \quad \circ$$

如果是三个种群以上,当 N_2 是捕食种群, N_1 和 N_3 是被捕食种群,描述这种关系的数学模型是^[6]

$$\frac{dN_1}{dt} = a_1N_1 - bN_1N_2$$

$$\frac{dN_3}{dt} = a_3N_3 - bN_2N_3$$

$$\frac{dN_2}{dt} = KN_1N_2 + hN_2N_3 - d_2N_2 \quad \circ$$

不同的捕食与被捕食关系相互组合,将会引出更复杂的情形。

总之,与非生命科学不同,这里很难找到能概括许多现象的普遍模型。由于生命现象

的复杂性,针对不同的具体问题需要构造不同的数学关系,来模拟事物的变化。生物模型这个概念是适应生命现象的复杂性而产生的。

早期的数学模型几乎都以微分方程的形式出现。从 Lotka 的捕食与被捕食方程开始,各种方程模型被构造出来,对运动形式多种多样的生命现象进行定量分析。许多数学家被吸引到这个领域,做了大量的工作,在生物数学中发展为一个独特的数学方法——生物方程。因为生命现象常常以多元的复杂形式出现,偏微分方程在生物方程中占据重要的地位,许多偏微分方程的理论和数值解法,被引用于生物学的研究中。

生物方程作为数学模型,现在已经广泛应用于生态学、生理学、遗传学、人类学、心理学、医学和农业等方面。由于它对生命活动中的许多物理关系和生理过程,能够综合生物学、物理学、数学和化学的最新成果进行研究,从而把生物学中某些领域的研究推向前所未有的水平。但是微分方程的数学模型,在数学方法上并无实质性的创新,几乎全部沿袭物理学中的数学物理方法。

二、从非生命到有生命

机械的数学物理方法应用于生物学,很快就暴露出它的弱点。让我们的讨论先从一些具体例子入手。

物理学中热传导方程是

$$\nabla^2 u - \frac{\partial u}{\partial t} = f \quad。$$

哺乳动物体内热的扩散是一个十分复杂的问题,如果模仿物理学中所做过的事情,作一些简化假设,可以得到类似于热传导方程的模型^[9]

$$\nabla(k\nabla T) + m_b c_b (T_0 - T) + S_m = \rho c \frac{\partial T}{\partial t}。$$

这个模型已经把血液循环和新陈代谢活动产生的热源都考虑在内,但是模型仍然很不完善。至少我们可提出如下一系列问题:活组织不可能是物理学中具有均匀介质的物体;新陈代谢的水平与氧的供给状况有关,与进行代谢活动的各种化学物质以及酶的作用有关;除血液循环以外,组织之间也存在物质的交换;神经系统的控制等等。试看在冷的环境下,生物组织温度应该下降,但是在一定的条件下,冷对神经末梢的刺激可使微血管扩张,新陈代谢旺盛,组织内的温度反倒升高。在此,机械的数学物理方法就会完全失效。

数学转向生物学出现的困难还表现在另一方面,物理学中所阐述的原理,表现在自然界中常常是直接的、明确的,因而比较容易被人们认识;可是在生物学中,生命现象以大量的、重复的形式出现,生命活动的规律寓于大量的偶然事件之中,往往难于为人们发觉。

生物数学出现的困难不止这些,最使数学家伤脑筋的事乃是许多生命现象无法用数值来加以描绘或度量。这个带根本性的困难,后面我们还要深入讨论。

总而言之,生命现象是物质运动的高级形式,表现多种多样,它们以互相联系、互相制约、大量重复的形式表现出来;又在空间上和时间上构成错综复杂的有机整体。因此在使用数学方法上,与非生命科学应该有本质区别,如果因循守旧、机械地照搬过去的方法,出现许多困难是很自然的事。

数学从非生命科学转向生命科学,生物数学出现的基本困难可以归结为以下三种:

- (1) 被研究的特性常常受到随机性的干扰;
- (2) 事物的特性数量比较大;
- (3) 事物的特性表现复杂,出现了无法用数值(实数)表示的特性,即非实数特性。

这里所谓特性,含意较广,是指被研究的生物对象其本身所具有的属性、特征或与其它事物的联系。

生物数学就是在不断克服这些困难的过程中得到发展的。

微积分学的建立是数学发展历史上具有划时代意义的大事,在这以前,古典的初等数学,其数学方法上的特点是静止地描绘客观事物,并且分析事物之间静止的数量关系。微积分学的创立标志着数学方法从静止到运动的转变。

方法上的转变带来了数学本身的发展,这是过去发展的历史。当今数学从非生命转向有生命,出现了许多困难,要求数学方法上的根本变革,这一变革又孕育着数学将要朝着新的方向发展。因此是一件具有重要意义的事。

生物统计学的产生克服了第一种困难,为数学应用于生物学闯开了第一道关。生命现象以大量重复的形式表现出来,受到许多随机因素的干扰,如何从大量随机现象中找出生命活动的规律呢?统计方法把随机变量与实数中某一值建立联系,以这个值的变化来反映随机变量内在的规律性,达到对生命现象进行研究的目的。这是统计数学方法的思想实质。试看各种随机变量的数字特征、各种统计量和各种统计方法,无一不体现了这个思想。整个统计数学的基本概念——概率,它的定义^[10]是集合上的测度,也体现了这个思想。

生物统计是数学应用于生物学最早的一个分支,各种方法和理论比较完整。应生物学的需要,50年代多元统计理论又获得发展。统计方法在生物学中的应用,现在已普及到每个领域,成为从事生物学、农学和医学工作的常规手段。

生物统计使用的数学理论主要有概率论、集合论、线性代数等。由于统计与概率的理论比较完善,随机性困难基本上得到解决,在当前生物数学发展中,看来不是主要障碍,因此我们不进行更深入的讨论。

三、从归纳到综合

现在让我们讨论第二种困难,即特性多的困难。

非生命科学中克服多特性困难的方法有二:其一是突出主要特性,对次要特性忽略不计;其二是对一事物所表现的诸种特性进行分解,分别研究、“各个击破”。这种比较片面的、孤立的、机械的研究方法不能满足生命科学的需要,生物数学需要对许多特性在相互联系的水平上进行全面地研究,需要综合分析的数学方法。

生物数学方法向综合分析的方向转变,也反映在自然科学的发展过程中。早期的研究,人们把复杂的现象分割为不同的部分和属性,分别进行研究,从个别的研究探求普遍规律。这是归纳的方法。自然科学被人为地分割成许多学科分支,分化愈来愈细。“我们用世界上的一切归纳法都永远不能把归纳过程弄清楚。只有对这个过程的分析才能做到这一点。”^[11]二十世纪50年代以后,自然科学的研究方法开始从归纳向综合分析的方向转变,这个转变表现为许多学科互相渗透、方法互相利用。近代生物学还出现了如生态

学、分子生物学、医学生物学这样一些高度综合性的学科。

自然科学不同学科的彼此渗透,突出地表现在各门学科的数学化。因为数学是自然科学的基础,它能把不同学科重新综合于统一的整体。自然科学经历了“合久必分,分久必合”的过程,这个过程也深刻地反映在生物数学方法的转变上。

数学方法向综合分析的方向发展,从50年代以后愈来愈明显,统计数学产生了多元分析技术,广泛应用于生物学,近代又有系统分析,生物控制论和分类分析兴起,这些方法从不同的角度,对生命现象的多种特性进行综合分析。数学方法向综合分析方向每迈进一步,都促进数学向生物学渗透,使生物数学得到发展。

系统分析和生物控制论是最突出的能够进行综合分析的生物数学方法。这些方法并没有靠舍弃某些特性去克服多特性困难,而是把尽可能多的有关特性都包揽在内,采取具有数学理论根据的方法,把一个复杂的问题变成实际上可以解决的问题。

试看线性状态方程:

$$\frac{d}{dt} \mathbf{X} = \mathbf{F}\mathbf{X} + \mathbf{G}\mathbf{U},$$

方程所描绘的系统可以是多输入、多输出,态向量 $\mathbf{X}' = (x_1 \ x_2 \cdots x_n)$ 的维数,一般大到几十、上百,它显示生命活动异常复杂的状态。在此不仅没有把那些所谓次要的特性忽略,也没有孤立地看待每个因素,通过矩阵 $\mathbf{F}$ 和 $\mathbf{G}$ 把有关特性都联系在一起,在综合的水平上,进行全面地分析。

系统分析中分室网络的理论也是一个很好的综合分析方法。组合数学中的图论被成功地应用于此,错综复杂的联系有条不紊地表现在网络图形中。对描述生态学、生理学中物质、能量和其它种种信息的流通,这种研究方法卓有成效。

综合分析不仅体现在把输入和各种特性考虑在内,而且还把输出反过来对系统自身的影响也考虑进去。这就是控制论中反馈的思想,控制论的创始人维纳(N. Wiener),首先把无线电电子学中的这个概念介绍到生物学中来,成功地分析了一些神经生理现象^[11]。反馈的思想对生物数学很有意义,因为生命活动普遍存在着反馈现象。

系统分析和生物控制论,靠综合分析的手段去研究生物学问题,因此比机械的数学物理方法高出一着,能够从错综复杂的现象中探究生命活动内在规律性。

维纳的《控制论》一书问世(1948年),是控制论发展的起点;一般的系统理论首次由 Bertalanffy 总结(1930年)。这二个分支学科在数学方法上有许多共通之处,都是在40年代以后,从电气工程技术中发展起来的,并且一开始就与生物学建立密切的联系,生物学是其应用的一个主要方面。尤其是生态学、生理学、医学和农业,此外,还大量应用于商业经济等社会科学领域。

系统和控制理论使用的数学工具很广,主要有微分方程、概率论、博弈论、图论、信息论等。

生物数学方法的综合性,还表现在它能够把不同学科中共同存在的内容,集中起来进行综合研究,沿着横的方向,从共性中把不同学科联系起来,交织成一个整体。数学高度的抽象性本身就体现了这个思想,生物数学更体现了这个精神。例如分类分析方法所要解决的“物以类聚”的内容,在生物学中普遍存在,有动物分类、植物分类、微生物分类、群落分类、蛋白质分类、土壤分类、疾病分类……。分类分析方法是用数学工具对这些领域

的共同属性作综合研究。通过分类分析方法,把这些不同领域综合在一起,促进它们相互渗透、共同提高。又如生物信息论,通过信息量或熵这个共同属性去综合探讨生物学中许多领域。遗传学、进化论、分子生物学、神经生理学、心理学、生态学、生物分类学和生物热力学,在信息论的基础上获得了新的统一。

四、从实数到非实数

现在让我们讨论生物数学的第三种困难,即非实数困难。非实数特性在生物学中非常普遍,例如昆虫翅的特性分直翅、膜翅、鞘翅、鳞翅……;植物的花冠分蝶形、唇形、舌形……;生物体内组成蛋白质的各种氨基酸、组成核酸的各种核苷酸,这些都是非数值特性。生物学中的许多基本概念如亲缘关系、遗传性、变异性等等,作为研究特性时,也难以用数值进行描绘。还有一类可用图形表示的特性,如植物叶片的形状,昆虫翅的脉象、癌细胞的图象等也都属于非数值特性。

在生命世界中非数值特性比比皆是,变化多端。与前两种困难比较,对付它更为艰难,之所以难,因为它涉及整个数学的根基。数学,是一门研究空间形式与数量关系的科学。除个别抽象数学以外,绝大多数都确立在实数集合,或者由实数所扩展的集合上(如多维欧氏空间,复数集合)。它的数学方法体现在函数的概念中,即从实数集合到一个实数值的映象:

$$\text{实数集合} \xrightarrow{\text{映象}} \text{实数}$$

与实际应用有联系的数学包括基础数学、微积分、微分方程、概率论、统计数学……等,其数学方法的基本结构都如上所示。这些数学在科学发展的历史中,曾经扮演过非常重要的角色,然而当进入生命科学领域,它的数学基本结构不能容纳非实数特性。这个根本缺陷几乎成为长期以来数学与生物学之间不可逾越的鸿沟。二十世纪以前许多数学方法没有越过这条界线,现代生物数学中的个别方法开始跨越非实数性困难,未来的生物学将要与数学建立更密切的联系,因此,这是一件具有十分重要意义的事。

一些能够克服非实数困难的生物数学方法有生物信息论、分类分析、模糊数学等。这些数学分支起步较晚,理论与方法尚不十分完善,但是由于它们能够越过实数界线,一旦应用于生物学,很快就显示出蓬勃的生命力。

信息论中的信息量是这样确定的,如果 (X, S, μ) 是有限离散概率空间,将基本空间 $X = \{x_1, x_2, \dots, x_n\}$ 上的概率分布 $P_i = \mu(x_i)$ ($i = 1, 2, \dots, n$) 看作信源,信息量由下面的公式给出:

$$H(P_1, P_2, \dots, P_n) = - \sum_{i=1}^n P_i \log P_i,$$

这里,基本空间 X 并未限制在实数集合上,可以是普遍意义下的集合,因此信息量可以理解为普遍意义下的一集合函数。它可以越过实数,将一组特性集合与实数建立联系,克服生物数学中非实数困难。它的数学结构与实集合函数不同,表示如下:

$$\text{特性集合} \rightarrow \text{实数}.$$

概率的定义本来也是一个普遍意义下的集合函数,但是概率论和统计数学都建立在

随机变量的基础上,随机变量取值实数,这一点又把它的数学概念回复到实数集合上。因此信息论与概率统计在数学方法上不同,概率统计不能解决的生物学问题,信息论却可能予以解决。

信息量的定义,于40年代末由 Shannon 给出^[12],50年代苏联数学家 Хинчин^[13] 和 Фадеев^[14] 在一定的条件下,分别给出了它的唯一性证明。在这个基础上发展起来的信息理论,内容非常丰富,但是这些理论毕竟都是应电信技术的需要而建立起来,不能完全满足生物学的需要。

生物学中常常使用信息量去描述生命活动中的离散性,描述离散性的定义很多,Brillouin 的计算公式如下:

$$B_k(n; n_1, \dots, n_k) = \log_2 \left(\frac{n!}{n_1! \dots n_k!} \right)。$$

二年前英国数学家 Laxton^[15] 在六项假设条件下,证明了这个定义的唯一性。在此基础上,一种适合生物学需要的生物信息理论正在形成和发展。

生物信息论现已用于遗传学、进化论、分子生物学、神经生理学、心理学、生态学和生物分类学。生物学向微观发展,从分子水平揭示生命现象的本质,或者在探索人类大脑思维的奥秘中,生物信息论很可能还要发挥更大的作用。

分类分析所要解决的问题,就是俗话说的:“物以类聚”。这是研究自然科学的基本方法之一。分类分析的运算从一个庞大的特性数值矩阵开始,经过多个步骤,最后得到的分类结果常常用树系图表示,因此它的数学结构比较特殊,可以表示成:

特性集合 → 树系图。

它与一般的集合函数大不一样,从特性集合映象过去的不再是一个简单的数值,而是一个定量的隶属关系。这种关系包含着生物学家所需要了解的丰富内容,当然远非一个单纯的数值所能表达。

用数学方法解决分类问题比较困难,长期以来没有进展,直到电子计算技术发明以后,才有所突破。一经突破,发展之迅速令人怵目。最早 Tryon 将它应用于心理学^[16],50年代末美国生物统计学家 Sokal 和英国微生物学家 Sneath 应用于生物分类学^[17]。数学向生物分类渗透带来了新的科学方法,使这门古老的学科别开生面、焕然一新。两门学科的结合产生了边缘分支——数量分类学。

数学应用于生物分类,促使分类的数学理论和方法大大发展,借用了许多数学工具,包括概率、统计、图论、信息论、线性代数和模糊数学。由于它的数学理论和运算方法日臻完善,逐渐形成一种独特的生物数学方法,称之为分类分析。

应生物分类的不同需要,分类分析区分为表现性分析方法与分支性分析方法。表现性方法发展较早,通常称为聚类分析,现在已得到广泛应用。分支性分析方法于60年代开始确立,70年代得到突飞猛进的发展,现在已用于古生物学、分子遗传学和进化论。

今日的分类分析不仅仅为生物分类建立起演化的自然系统,而且能通过分类运算去分析复杂事物之间的因果关系,探索生命活动的客观规律。它的应用除生物分类外还有生态学、生物化学、生物地理、分子生物学和医学、甚至超出了生物科学,把它的触角伸展到更广的范围,如土壤学、气候学、考古学、环境科学和社会科学等领域。

模糊数学的产生基于对集合概念赋予新的理解。所谓模糊集,其实是集合概念的扩

充。任意一个集合 A ，都可以定义一个特征函数：

$$f_A(x) = \begin{cases} 1 & \text{当 } x \in A \\ 0 & \text{当 } x \notin A, \end{cases}$$

反之，任一个这种形式的函数 $f_A(x)$ ，也唯一地确定了一个集合 A 。如此一一对应的关系，我们简直可以把特征函数理解为集合 A ，二者是等价的。现在要问，如果特征函数的限制放宽，集合的概念又该怎样理解呢？譬如说，让函数的值可以取在区间 $[0, 1]$ 上任何一点，如此扩充了的集合概念，正是模糊集。集合概念的扩充，正好象过去代数学中开方运算的扩充可以导出无理数与复数；几何学中平行公理的改变引出非欧几何学。在数学发展的历史上，基本概念扩充屡见不鲜。

从一般集合概念扩充到模糊集合概念，对生命科学意义很大。在非生命科学中，事物的判断往往是明确的，非此即彼、界限分明。在这个意义下，集合的特征函数取值非 0 即 1。人们称这样的科学为“硬”科学，物理学是典型的硬科学。可是在生物学中，许多事物却是不十分明确的。在这个意义下，集合的特征函数应该能取到介于 0 与 1 之间的值。生物学的研究需要建立一种新的集合概念，来描述这种不明确的事物，应生物学等“软”科学的需要，模糊集合的概念应运而生，数学中产生了“软”的数学方法。

模糊集之间还可以定义并、交、余基本运算：

$$C = A \cup B \text{ (并)} \quad f_C(x) = \max\{f_A(x), f_B(x)\},$$

$$C = A \cap B \text{ (交)} \quad f_C(x) = \min\{f_A(x), f_B(x)\},$$

$$C = A' \text{ (余)} \quad f_C(x) = 1 - f_A(x)。$$

从模糊集的运算获得一个新的模糊集；对模糊集的分析也可以得到一个实数值，用以表示模糊集本身或它们相互之间的关系。因此模糊数学的数学方法，一方面体现在它的模糊概念中，另一方也可以用这样的结构来表达：

模糊集集合 $\rightarrow$ 模糊集或实数值。

模糊数学于 1965 年由 Zadeh 创立，在他的原始论文^[18]中，第一次提出了模糊集合的概念。因为集合是数学中最基本的概念，模糊集合的引进，动摇了整个数学的基础，构筑于集合概念之上整个数学大厦的面目都将为之一新，在短短的十多年中，产生了模糊聚类、模糊概率、模糊决策、模糊图象识别……。许多数学家还在寻找各种途径，把模糊理论与各数学分支建立联系。数学理论的模糊化，使得数学更能适合生物“软”科学的需要。模糊数学十分年青，正在发展中，是一门很有前途的生物数学方法。

五、从连续到不连续

为了克服非实数困难，生物数学方法又从连续向不连续发展。实数与连续性，在数学的基本概念中有着本质的联系。数学中连续性的严格定义来自极限的概念，数学分析的基本理论又告诉我们，极限的基本定理，即哥西（Cauchy）准则，与实数集合的完备性完全等价，连续性的概念是在实数集合上产生的。因此，生物数学的发展从实数到非实数与从连续到不连续，二者是一致的。实际上许多能克服非实数困难的生物数学方法都带有不连续性的特点。生物学中的数学模型，对离散模型的研究占据重要位置，信息论、矩阵论、图论和布尔代数等等不连续的数学理论在生物数学中得到广泛的应用。

自然科学中,不连续性早在本世纪初已被 Thompson 明确指出:“不连续的原理是我们所有学科种类中的固有性,不管是数学的、物理学的还是生物学的。”^[4]

物理学中分子学说的确立,说明物质是不连续的。1900 年德国物理学家普朗克提出能量量子的假设,说明能量分布的不连续性。也是在 1900 年,孟德尔的遗传理论再次被发现,又打破了生物学中物种进化连续性的观点。达尔文认为生物进化是连续地量的积累,不同物种之间都存在中间类型的过渡。从孟德尔的基因理论却引出相反的结论,在生物学中,基因相当于量子力学中的普朗克常数,一个基因的改变引起相应的生物突变,可以看作是生物进化的最小单位。著名量子物理学家薛定谔 (E. Schrödinger) 指出“突变实际上是由于基因分子中量子跃迁所引起的。”^[9]正是他,把这二个不连续的理论统一起来,开创了“量子生物学”。这个统一说明,不论是生命科学或非生命科学,从连续到不连续是普遍的原则。

自然辩证法告诉我们,物质的运动是不断地从量变到质变的过程。质变的飞跃意味着运动的不连续性。数学作为自然科学的基础,不仅能描述物质运动量的连续变化,也应该能描述质的不连续变化。因此从连续到不连续是数学发展的一种趋势。

数学的基本理论告诉我们,如果一个函数 $y = f(P)$, 在 P 点是可微的,则函数在 P 点是连续的。可微性蕴涵着连续性。因此,一切以微分概念为基础的数学理论都是连续的。以微分方程构造的数学模型,是生物数学中典型的连续性数学方法。这些方法只能适用于连续运动范围,对于出现不连续点或奇点(包括导函数不连续)的情形,将无能为力。而恰恰在这些连续性中断的关节点上,正是生物学常常需要研究的课题。

60 年代末,法国数学家 Thom^[20] 从拓扑学提出一种几何模型,能够描绘多维不连续现象,他的理论称为跃变论 (Catastrophe Theory)。Thom 指出,控制因素不超过 4 个,突变的类型只有 7 种。生物学、物理学和社会科学中许多不连续的现象都能够从中找到相应的跃变类型进行模拟。跃变理论弥补了传统数学物理方法的不足,对于处在飞跃的、临界状态的生命现象,可以用跃变的模型给予生动地描绘和解释。现在已成功地应用于生理学、生态学、心理学和组织胚胎学。在神经心理学的研究中,甚至用来指导医生对某些疾病的临床治疗。

数学领域中有许多分支,一直被看作是毫无应用价值的抽象数学,受到人们的冷遇。这些分支数学应用于生物学,在方法上能够越过不连续性困难,开始放出奇异的光彩。拓扑学是一个例子,下面再举一例。

在非实数特性中,有一类称为二元特性。所谓二元特性是指表现为二个对立状态的特性。例如动物是脊椎动物,还是无脊椎动物;有花植物是离瓣花冠,还是合瓣花冠;神经组织是处于兴奋状态,还是抑制状态;某些生理反应呈阳性,还是阴性等等。二元特性是一种不连续特性,在生命现象中屡见不鲜,连续性数学方法不能解决具有二元特性的生物学问题。

在此,我们讨论关于神经传递功能的二元特性问题,神经的兴奋由兴奋波沿着神经细胞的轴突传递,经过突触,在阀的控制下传给另一个神经细胞。兴奋波通过与否是一个不连续的二元特性。布尔代数是专门适于描述二元不连续特性的数学方法。1943 年麦卡洛克和皮茨在布尔代数的基础上,第一次给出描写神经传递现象的数学模型^[21]。经过多次改进,今日的神经网络数学模型已经能把学习和遗忘考虑在内。如果用电子计算机加

以实现,很象人的大脑,具有学习的功能。不连续的数学模型正在揭开人类大脑思维的奥秘。因而布尔代数博得了数学家、生理学家和心理学家们的垂青。

但是不要忘记,布尔代数是英国数学家乔治·布尔(George Boole)早在19世纪创立的,最初应用于数理逻辑,一个世纪以后的今天,布尔代数作为不连续的数学方法,在生物学和电子计算机技术中竟会放出如此绚丽的光彩!

我们已经回顾了生物数学发展的简短过程。生物学发展的需要,促使数学方法正在变革。在生物数学中,新的生长点正在酝酿形成,新的数学方法又正在不断涌现。如何评价当前生物数学所处的地位呢?未来生物数学发展的前景又将怎样呢?

数学转向生物学,从生物统计学开始,到数学方法上的根本变革,生物学把数学的发展引向新的境界。十七世纪牛顿与莱布尼茨从物理学研究中创造了微积分学,体现了数学方法从静止到运动的转变,这个转变导致现代数学的产生。当今数学从非生命到有生命,是一次更深刻的变革。数学方法开始向综合分析的方向发展,向克服非实数困难方向发展,向不连续方向发展。数学的发展在未来生物学中面临着新的转折点,准备着一次更大的飞跃。未来的生物数学将会出现比牛顿与莱布尼茨时代更为辉煌灿烂的前景!

数学,作为基础科学,与其它自然科学相结合,在量子力学、相对论等近代物理理论中把人类对非生命世界的认识推向新的阶段。在生命世界中,生物学与数学紧密相结合,在探索生命现象的奥秘中,将把人类对生命世界的认识提高到又一个崭新的阶段。未来的生物学由于引进了数学,将是定量的生物学,精密的生物学、创造生物与改造大自然的生物学!

参 考 文 献

- [1] 恩格斯:自然辩证法,第206,249页,人民出版社,1974.
- [2] Leuschner, D.: *Biometrische Zeitschrift*, 19(7), 575—580, 1977.
- [3] Chytil, M. K.: *Acta Biotheoretica*, 26(2), 137—150, 1977.
- [4] Thompson, D' A. W.: *On Growth and Form*, 1952.
- [5] Lotka, A. J.: *Elements of Physical Biology*, 1952.
- [6] Lotka, A. J.: *Element of Mathematical Biology*, 1956.
- [7] Walsh, G. R., *Bull. Math. Biol.*, 40(3), 319—333, 1978.
- [8] Bojadziev, G.: *Bull. Math. Biol.*, 40(3), 335—345, 1978.
- [9] Zhitomirskii, I. S. and Kagna, A. A., *Bull. Math. Biol.*, 40(4), 535—540, 1978.
- [10] Halmos, P. R.: *Measure Theory*, 187, 1950.
- [11] Wiener, N., 郝季仁译:控制论,科学出版社,1963.
- [12] Shannon, C. E.: *Bell System Technical Journal*, 27 379—423, 623—654, 1948.
- [13] Хинчин, А. Я.: *УМН*, XI, Выход, 1(67), 17—75, 1956.
- [14] Фадеев, Д. К.: *УМН*, XI, Выход, 1(67), 227—231, 1956.
- [15] Laxton, R. R.: *J. Theoretic Biology*, 70(1), 51—67, 1978.
- [16] Tryon, R. C. and Bailey, D. E.: *Cluster Analysis*, 1970.
- [17] Sokal, R. R. and Sneath, P. H. A.: *Principles of Numerical Taxonomy*, 1963.
- [18] Zadeh, L. A.: *Information and Control*, 8, 338—353, 1965.
- [19] Schrödinger, E.: *What is Life*, 1944.
- [20] Thom, R.: *Topology*, 8, 313—335, 1969.

植物数量遗传学研究的成就和方向

黄远樟 编译

(北京师范大学生物系、数学系数量遗传研究组)

美国内布拉斯加州林肯大学农学系的 Gardner, 在 1976 年国际数量遗传学论文集中发表了题为“植物数量遗传学研究的成就和方向”的文章, 作者回顾了植物数量遗传学研究的历史, 对植物数量遗传学研究的过去、现在和将来作了深刻的总结性分析和归纳, 很值得植物育种者和数量遗传工作者一读。现以该文为基础, 适当补充 1976 年以来植物数量遗传学研究的进展, 综述如下:

回顾植物数量遗传学的发展历史是很有意义的。Gardner 认为本世纪 50 年代和 60 年代的廿年可称之为植物数量遗传学和数量遗传学理论在植物育种方面应用的“黄金时代”。在 Fisher (1918)、Sewell Wright (1935) 和 Haldane (1924) 的基础上, 英国 Mather (1949) 等和美国北卡罗来纳州的 Comstock 及 Robinson 等继续深入研究和开展了这项工作。Cockerham (1954) 和 Kempthorne (1954) 因将上位性理论考虑到遗传模型里而作出了重要贡献, 进而加快了数量遗传学的深入研究。关于数量遗传模型, 虽然 Schnell (1961, 1963) 发展了连锁理论, 但在植物数量遗传学中则很少应用。Comstock 和 Robinson (1952) 在显性度的连锁效应方面的工作是十分重要的。

本世纪 70 年代, 在植物育种方面, 数量遗传学仍然起着重要作用, 但在许多部门内其重点已转移到植物育种中的生理和生化方面的研究。实践表明, 数量遗传学家与生理学家及生物化学家共同合作已是势在必行了。植物的许多生理过程和生化组成, 特别是谷物和饲料作物的品质性状, 如谷物的蛋白质含量及赖氨酸含量等生化指标性状均属数量性状, 都处于多基因控制之下, 最终还是要用数量遗传学的方法来研究。这样, 我们期待着生理学家和生物化学家去发展和完善测量技术, 以使得数量遗传研究中的大量个体或家系的测定成为可能。

数量遗传学的理论发展远远赶在植物育种应用的前面。众所周知, 植物是非常复杂的有机体, 植物在其一生中可对各种不同的环境变异作出各种不同的反应。在许多情况下, 我们却一直利用较为简单的遗传模型, 然而却又取得了很大的成就。

一、已取得的成就

仅就数量遗传学原理在玉米育种和植物育种中应用的一些进展简述如下:

(1) 交配设计与遗传模型及其相应的统计方法的分析和遗传解释已由 Comstock、Robinson (1948, 1952) 和 Griffing (1956) 等加以发展, 并由 Cockerham (1963) 在一篇很出色的论文中加以总结和详细说明。由 Kempthorne (1957)、Mather (1949)、Falconer (1960) 以至最近的 Mather 和 Jinks (1971) 所写的书对植物育种者来说是有价值的。

的。直到目前,国际上仍有许多植物育种工作者和数量遗传工作者广泛地使用 Hayman 和 Griffing 的双列杂交法来研究育种问题或数量遗传学问题。

(2) 大多数植物的遗传方差已被估算并用来作为植物育种的重要依据。在异花授粉作物如玉米中,其籽粒产量和其它性状中存在着明显的加性遗传方差,从而证明群内选择方案是有效的,而以往早期的玉米育种者则认为群内选择方案无效。在改良群体时,利用各种轮回选择方案而取得的成就也进一步证明这些加性遗传方差的估算是有效的。

(3) Hull 提出的作为玉米杂种优势主要原因的超显性假说已为北卡罗来纳州大学及内布拉斯加州大学的工作加以否定。在玉米,已证实了控制产量的基因所引起的明显的超显性连锁效应的存在 (Gardner, 1963)。

(4) 通过任何种类的轮回选择系统来预测植物遗传进度的理论基础已得到发展。其主要功绩归于 Comstock 和 Cockerham, 尽管他们的工作和想法是以 Lush 的工作为基础的,而 Lush 则又追溯到 Sewell Wright 和 Fisher 的早期工作。关于具体的选择系统,已有人提出和发表了许多预测遗传进度的公式。Empig、Gardner 和 Compton (1972) 总结了关于群内和群间选择系统的预测方程式,在选择不同的育种系统时,植物育种者广泛地使用了这些方法。

(5) 有关固定有利等位基因的概率理论已由 Kimura (木村资生)(1957)提出。该理论在轮回选择方案中提供了决定有效群体大小的最佳值。

(6) 选择指数的理论已得到发展,在选择标准中该理论提供了被研究性状的最优加权。然而,植物育种很少使用选择指数。

(7) 正是数量遗传学的研究,新的育种方案已得到发展,而旧的育种方案则得到修改和完善。Comstock、Robinson 和 Harvey (1949) 的交互轮回选择利用了所有种类的遗传变异。从商业育种观点来看,由于每代新品系和杂种的自然衍生,看来全同胞交互轮回选择 (Lonnquist 和 Williams, 1967; Hallauer, 1967) 更有前途。利用近交测验品种进行的全同胞交互轮回选择 (Eberhart、Debela 和 Hallauer, 1973; Eberhart 和 Russell, 1975) 提供了更明确的基因型鉴定,看起来是有前途的。由于使用格栅系统和其它环境控制 (Gardner, 1961), 改进的混合选择已在其它作物及玉米中 (Gardner, 1973) 获得成功。关于玉米穗行育种法, Lonnquist 的修改已经使得该系统成为最成功的系统之一 (Webel 和 Lonnquist, 1967)。

(8) 自花授粉作物的数量遗传研究,为了改良群体,已使轮回选择方法得到应用;并提供了更大范围的基因型变异,从中选出一些新品种或育种材料。在高粱,我们应用了遗传的雄性不育基因把正常的自花授粉作物转变为异花授粉作物,并利用轮回选择方案来改良这些群体。

(9) Sprague 等(1951)和 Comstock 等(1963)已阐明了基因型与环境的互相作用(作为表现型与它的基因型值的离差的一个分量)。Finlay (1964) 和 Eberhart、Russell (1966) 提出了基因型稳定性测定的方法。

(10) 本世纪 70 年代,一些遗传育种工作者如 Bhatt 和 Hussoim 等对合理选配亲本的问题进行了研究,提出合理选配亲本的重要原则之一就是杂交亲本的遗传差异要大。因此,如何测定亲本间遗传差异是很必要的。为此,育种工作者们运用了多元统计的方法对作物数量性状的遗传差异进行了分析,提出了数量性状遗传差异的一个度量——遗传距

离。再根据遗传距离的大小对亲本进行分类,为育种工作者在杂交育种工作中合理选配亲本提供依据。我国刘来福等^[1,2]曾以冬小麦为试验材料,用上述方法计算了50个冬小麦每两个亲本间的遗传距离,并根据遗传距离,对参加试验的冬小麦亲本进行分类。其研究结果与本地区育种实践在这面积累的经验相符。刘来福等人在另一试验中^[3]对利用完全双列杂交法所测出的各组合特殊配合力(换算成主成分的特殊配合力)与利用多元统计方法分析所测到的各组合的亲本遗传距离进行了比较分析,研究表明,利用多元统计方法所测定的遗传距离可作为杂交育种时正确选配亲本的一个依据。目前,聚类分析法已广泛应用于水稻、棉花、蔬菜等作物上,并已取得可喜的成果。

(11) 最后,在植物育种者对基本群体、亚群体和育种系统进行选择时,有关群体的数量遗传信息已经起着并将继续起着重要作用。

二、今后的研究方向

展望未来,数量遗传学依然在植物育种中起着重要作用,但还需要作进一步的改进和完善。

(1) 不同的轮回选择方案在决定最佳有效群体大小的方法中,必须作出改进。方案中有效群体太小了,以致使不需要的基因得到固定并限制了进一步的发展。

(2) 由于轮回选择方案,群体被结合或产生时,或当通过选择形成下一世代时,连锁和连锁不平衡效应要作进一步的研究。

(3) 对同时选择几个性状的方法需要作进一步的研究。不必要的相关响应常常发生在植物单个性状的选择上。对于有商业价值的群体,许多性状必须同时加以改良。

(4) 关于上位性研究的新途径及其在育种方案中的普遍使用的新方法均十分需要。在自交系间的杂种里所观察到的杂种优势中,以及在自花授粉栽培品种和无性繁殖植物中,上位性显得非常重要,而关于上位性我们则知道得很少。

(5) 植物的生理过程和植物的生化组成多半处在多基因控制之下,这将需要用数量遗传学方法来研究。今后,要完善测量方法,深入开展对生化指标如蛋白质含量、赖氨酸含量等的数量遗传学的研究。

(6) 如果细胞培养技术已发展到能容易地产生单倍体植物的地步,那么,数量遗传学家将面临一整套新问题。Griffing 预先处理了这个问题,并提出了关于单倍体的选择理论。

(7) 聚类分析方法已得到广泛的应用,但有些学者对此持怀疑态度,这迫使我们对该方法要作进一步研究,即解决下列问题:这种测定方法是否能客观地反映材料间真实的遗传差异。另外,既然能利用亲本间遗传距离大小预测 F_1 的表现,那么能否预测 F_2 以后各世代表现,这些将有待于作进一步的研究。

(8) 最后,在未来的年代里,我们将面临矿物能源不足的问题,于是化肥的使用、灌溉方式以及耕作制度等必将发生变化。这些变化将依次改变植物的环境和植物育种的目标。这样,目前处于最佳条件下的最有利的等位基因可能不再是最有利的等位基因了。

总之,数量遗传学在植物育种中的重要作用是不可忽视的。今后,仍要重视数量遗