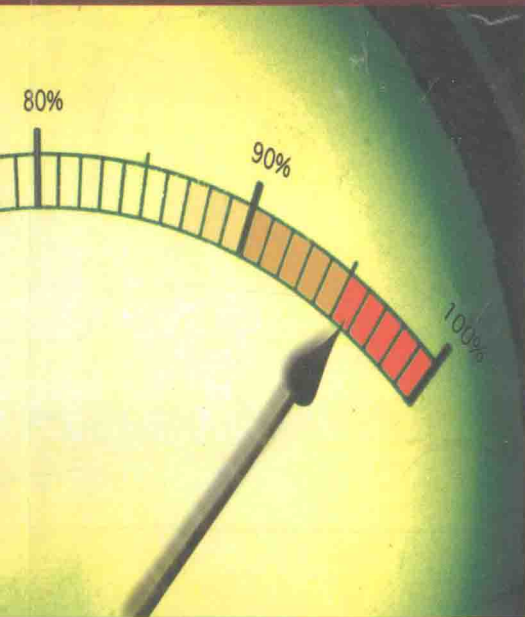


混沌与统计理念

储海林 编著



HUN DUN

YU TONG JI

中国物资出版社

混沌与统计理念

储海林 编著

中国物资出版社

图书在版编目(CIP)数据

混沌与统计理念/储海林编著. —北京:中国物资出版社,
2002.9

ISBN 7-5047-1902-1

I.混... II.储... III.统计学—研究 IV.C8

中国版本图书馆 CIP 数据核字(2002)第 072462 号

责任编辑:王 莉

封面设计:沃森广告

责任校对:王 莉

责任印制:方鹏远

中国物资出版社出版发行

网址:<http://www.clph.com.cn>

社址:北京市西城区月坛北街 25 号

电话:(010)68392746 邮编:100834

全国新华书店经销

山西财经大学印刷厂印刷

开本:850×1168mm 1/32 印张:16.5 字数:415 千字

2002 年 11 月第 1 版 2002 年 11 月第 1 次印刷

ISBN7-5047-1902-1/0·0030

印数:0001-2000 册

定价:26.00 元

(图书出现印装质量问题,本社负责调换)

前 言

本书的写作是一个学习、总结和提高的过程：它的格式不受某个固定的框架的所限，而是遵循思考的“流程”。书中引用许多他人的观点，只是为了说明本书的主题与宗旨——统计理念和混沌的关系。阅读本书不要求全部看懂，有些章节可略去不看，只看结论就行。之所以安排一些较深的数学内容，是想给那些需要进一步弄清混沌和统计关系的读者留下一个思考和探索的空间。值得一提的是，作者在研究（是 study about 而不是 study on）分形与非线性问题时，是基于直觉，是对“丛论”有一个“粗浅”的纲要性理解，这主要得益于王建华博士的“丛论”课程。特别是在看了霍金博士的《Nature of Space and Time》一书和翻译完了彼得斯先生的《分形市场分析》书后，对于从数据的角度来看混沌和动力系统这一类问题有了更实际和直觉的了解。对于“理论”、“实践”、“试验”这一类有可能带有深奥哲学意义的概念，从形而下的角度有了更深的理解。

本书各章节无明显的排序联系，它完全可以分章独立地去看，数学公式的计算仅仅是个形式，关键是要明了公式背后的形而下的背景。关于人文科学、诗歌与分形的联系，是受托姆教授、陈寅恪和黄仁宇先生的一些观点的启发而来，出发点没有问题，操作性的思路尚待进一步商榷。

再者，无论是分形与混沌还是统计学，都对非线性方法表现出了浓厚的兴趣。

很多数学大家的个人网页都说明了这一点。一些从事实际工作的人也进行了相当有益的探索，并取得了“分类”意义上的成就，

彼得斯(peters), 菲利普(philip), 常昆思(Kung - sik chan)和理查德(Richard)等人的工作及著作都反映了实际工作对分形和统计数据的需求。本书中无论是引用他人的观点, 还是我个人的观点, 作者本人都保留进一步探讨的余地。对于可能出现的错误, 文责自负。

学术概念的总结

(内容提要)

有一种理论认为历史在循环中移动。它像一个盘旋的楼梯间隔,当人类活动的过程充满一个间隔时,它就跃上一个新水平。人类文化转变的“钟摆”(盛衰),并不仅仅是同样事情一而再,再而三的重复。不管理论是否真实和实用,它是以一种隐喻的方式引起了我们的注意,聚焦于我们的注意时,它充当了一个隐喻。本书的主题代表和体现了这样一个循环:像恩格斯所说的螺旋式的上升,但不是鞅意义上的,如同在一条蜿蜒小路上疾驶的,简·爱和灰姑娘都曾乘坐过的欧洲中世纪的大马鞍子的驿车,这个驿车在小路所代表的曲线上的每一点都有可能沿切线方向出去,但是鞅没有,因为有一个像上帝一样的驭手使它不会出轨。至于说会不会在圆形环纹曲面的曲线上出轨,那就要看驭手的高明程度了。混沌给出了秩序的方式,秩序又依次给出混沌的新形式。在这个“钟摆”中,我们不是力求摧毁混沌,而是要驯服它。

在我们人类很久远的过去,自然被考虑为一个宽敞的创造物(曾像盘古,又像上帝的环境)。在自然世界中,模式的缺乏被归因于既有力量且又不可理解并统治它的神的一时的兴致。被统治的混沌及法则是不可想像的。

跨过几千年的区间,人类慢慢达到了一种理解,即自然有许多规则,这些规则能够被记录、分析、预测及开发。18世纪科学在消除蛮荒方面取得了巨大的成功,而自然法则的许多思想却留给发

现者。

规定在宇宙中每个质子运动的不可变更的法则要永远精确；科学家的任务是要解释那些法则对于任何有兴趣的特定现象的应用。混沌让位于有规律的世界。

然而，世界在进程中，它是我们视野中的宇宙。当今，甚至我们的钟表都不是由有规律的装置来组成——那么，我们的世界为什么一定如此呢？随着量子力学的出现，有规律的世界变成了宇宙的轮盘赌。基础的活动诸如放射性粒子的衰变，被认为是由机会而不是由法则来显示确定性的。不考虑量子力学蔚为壮观的成功，它的概率性特征（不仅仅是布朗形式的）没有唤起每一个人。爱因斯坦谈到量子力学，可是，他的哲学是抓住了经典力学全部过程的原因，在这里，量子不确定性是不可操作的。对于机会的掷骰子的隐喻的全面应用，确定性给机会留下空间了吗？

循环已经来到并充满旋转——但是在更高的水平。因为我们是由服从不可更变和准确性法则的系统发现者开始的，这个系统并不总是在可预测和规范的方式下运行。简单的法则不能产生简单的行为。确定性法则却能够产生呈现随机性的行为。秩序能够培育它自己那一类型的混沌。问题不是上帝是否掷骰子，而是上帝如何掷骰子。（霍金语）

这是一个戏剧性的发现。它的应用已经在我们的科学中引起全面影响。预测或重复实验的概念，一旦混沌眼睛（指类似猫头鹰眼的轨道形状）被人们理解，它就在新的方面被采用。我们所考虑的那些简单变成复杂的，以及扰乱性的新问题由于涉及到测试、可预测性、证明或理论的证明而无根据而堆积起来。

作为补偿，被考虑为复杂的事情将变得简单。呈现在结构和随机性中的现象可能实际上服从一个简单的法则。确定性混沌有它自己的法则而且导致新的实验技术。在自然界中不乏不规则性，而且它们当中的一些被证明是混沌数学的形而下神秘现象的

具体(或形而下的)表现形式。自然现象在这方面的说明已充盈于耳,而彼得斯等人在金融数据领域的研究为人文科学打开了新的研究方法领域之门。

来自于秩序的混沌

有序和无序,和谐与混沌之间的冲突必定代表人类整体理解力的深度的感觉。为什么会有如此之多的天地万物的神秘性以及如此之多的文化,是因为它们有共通性。在古希腊的宇宙学中,混沌即是宇宙的古老的空虚,而且是居住死者的底层社会。旧约中的神学曾说过:大地无形,且空,远离中心的表面笼罩着黑暗。我国的古代神话也有类似描述。很明显,这是脱离了知觉与观察的推理和理性。混沌最初是无形的质量,它来自于形成宇宙的造物者。秩序等同于善,而无序等同于罪恶。

某种与生俱来的冲动使得人类对自然界规则的理解进行抗争,去寻找宇宙的反复无常的复杂性背后的法则,使秩序脱离混沌。历法、日食、月食即为一例。他们同时还创造了一个万神殿以解释随机和无感觉世界中的反复无常,他们发现循环、形状、数字、数学。

过高的理性

数学是缘形而下世界的问题而开始的,并且以提供某些答案维持生计,但过程决不是恰好一个,通常的数学理念取自于它自己的生命。它好像存在于默默无闻之中,以纯粹的数学对象为自己的目的而被显示和讨论,直到它内部的秘密被剖析开,它形而下的特征(通常是几何学意义上的)因此被领悟。数学之所以是有效的,是因为它代表了人类大脑的基本的“语言方式”。我们之所以能够观察模型仅仅是数学的,是因为数学是我们的知觉的工具。数学之所以在组织形而下的实体中是有效的,是因为它基于形而下的实体而生灵感。它之所以成功,是因为它是宇宙的幻影(映射)。或许不存在实际的模型,存在的仅仅是那些我们微弱的存乎

一心的赋予,这些是哲学家的问题。实用主义的现实环绕于我们周围,存在于我们所看到的、所理解的东西中,而数学是最有效和最有价值的方法。牛顿的哲学在这方面是最好的体现,他的思想已被抽象为我们文化的直接的基础。

“自然有法则,我们能够发现它们”

简单,优雅,但难以理解。

秩序来自于混沌

过去的时尚或我们将来的视角?我们不能分辨,理论取代理论,范例推翻范例。只有一个事情保持不变:数学的相关性。自然的法则是数学。但是你用数学时一定要用脑子而不是用手。

混沌的几种解释

1. 名词:无形无序之物,存在于有序的宇宙之前。

2. 彻底的无序,绝对的混乱。

3. (数学上)发生在确定性系统的随机行为。这是两个更模糊的词——“随机”和“确定”。拉普拉斯的决定论对我们已经是很熟悉的了。“随机”意味着“偶然”、“任意”。为理解混沌现象,我们需要进一步讨论它们的意义,因为它现在的形式形成的定义是一个悖论。确定性行为是由精确的和不可破坏的法则所统治的;随机行为则相反,无法无序,由机会所统治。这样的话,混沌是完全由一法则所控制的无法则行为。

无所不适的方程

有规律世界的隐喻可以追溯到许久以前,我们很欣赏它是如何深深地扎根于我们的理念,在考察混沌以前,我们必须首先考察法则。

关于法则概念的演化,我们仅以几个概念性名词来说明它。

1. 具有悠久文化历史的民族在天文历法方面的数据记录,古希腊、中国、印度等。“在你不能清楚你的脚下之物时,你如何能分辨什么在天空运行”。

动态学仅仅起始于数学家们拽着他们的眼睛看宇宙,并且看得更近,更关键。

2. 宇宙学的革命(运动)

3. 希腊的齿轮

4. 中心的太阳

5. 摆的摇动

6. 重力与几何学

7. 宇宙(或世界)的体系。

当我们把上边七个问题放在一个范围内系统地考虑,他们就是动态学的开始,适用于他们的是方程或微分方程以及离散的差分方程。

“分析”一词,当今已被用来描述微积分的严格的行为形式:在理论的背后不是计算技巧。在18世纪计算的理论方面被实质性的扩大时,它得到了一个暗示:这个发展的主要设计师是欧拉,他们甚至把音乐与数学联系起来。分析、几何、演化、可选择的方向,已开始在人们的头脑中徘徊。

误差的法则

越来越多的聚众生事呈现出越来越多的无政府状态,而越来越完善政策(法规)的出台以统治和支配这种状态。这是非理性的无上法则。每当混沌元素的大样本在进行时,并在它们的大小的秩序支配之下,意外的和最美的法则形式证明将始终隐藏着,被控制的过程的顶端构成了一个不变比例的流动的曲线:假如它被归了类,假如它被发现,每个元素会发现一个适合它的恰当的位置。数学可以计算木星卫星的运动,可是不能计算暴风雪中雪花的形状。它能够描述肥皂泡的增长,却不能描述一棵树的生长。基于某种理由,数学家们相信,那么多的无序服从相同的基本法则;他们没有能力对任何结果运用那些法则的恰恰是事物的复杂性。

混沌(确定性混沌)处理长期演化——某些东西如何在一个长

时间变化。混沌时间序列看起来是无规律的。混沌的两个重要的、在实际应用上的题中应有之意是指:混沌条件下的预测是无价值的,而且复杂行为可能有简单的原因。在真实世界的的数据中辨认混沌是困难的,因为可适用的工具通常是在理想条件下显示的,而要在实践中完成这一任务是困难的。

混沌一般发生在确定性的、非线性的、动态的系统中,一些著名的非线性系统如今受到越来越多的评论和关注。动态系统能够在两个方式下演化:一个是在分离的情况下,有明显的原因(离散区间);另一个是连续的,当连续变化时,我们能够离散或连续的测度。依据离散观测值而来的方程,称作差分方程,而依据连续观测值的叫作微分方程。迭代是一个模拟正则,离散区间的数学方法。迭代也并入了在给定时间依赖它在以前时间的值的任何对象。混沌确是一个训练中的课题,在许多学科中,专家们正在研究它。

在自然界中,所提出的混沌的原因是在控制因素中的临界性增长,两个或更多过程的非线性迭代;以及环境噪声。基于混沌分析,数据能够揭示一个看起来不稳定的、古怪无常的时间序列实际上是确定性的,并能导致更精确的短期预测。它告诉我们对系统的长期预测是完全可行的,但要简要的建模,却未必理想。另一方面,混沌分析不能揭示基础性质的物理定律。

混沌理论(包括对初始条件的敏感性、长期的不可预测性,熵,李亚谱诺夫指数以及概率和其他方面)。

相空间或状态空间是一个抽象的数学空间,其坐标代表了特定化一个动态系统在一个特定时间的位相(或状态)的所需变量。子系统有三个或较少的变量,变量的观测值可以标绘在图上;在这种情况下,相空间是在图形的轴之间的想象性空间。相空间的两个类型是标准相空间和伪相空间。在图上(三个坐标或更少)的标

准相空间,每个轴代表一个关键变量(亦即温度、价格、重量)。标绘的点代表了系统在一个时间的状态,标绘的点的序列表明了系统如何随时间变化。在标准相空间上的标绘,相对而言,轴和坐标代表了同一个物理(形而下)特征的连续值。最普通的伪相空间标绘使用一个非常量时间区间(一个滞后)在两个连续测度之间,称作滞后空间标绘。滞后空间标绘的一个好的例子是一维映射的图形(映射是一个规则,它特定化了动态系统如何随时演化),一维映射是一个仅仅处理一个测度特征的函数。它的表现形式为 X_t 和 X_{t+1} 。在每一个相空间图形上的点的序列是相空间中的轨道,不仅仅是三维或再少一些,相空间和伪相空间的概念适用于任何数量的坐标和维数。在任何伪相空间的维数分析叫做嵌入维。在大多数混沌分析的类型中,嵌入其时间序列是一个基本步骤之一。

空间中的距离和直线

混沌理论在相空间中通常要求两点的直线距离。对于直线方程的参数,以及几个类型的值,不考虑维数和变量的个数,在两点之间的一个直线距离可以用一般的距离公式计算(彼得哥拉斯定理的重新排列)。对于任何变量的个数($x, y, z, \dots, w$)公式为:

$$L = ([X_2 - X_1]^2 + [Y_2 - Y_1]^2 + [Z_2 - Z_1]^2 + \dots + [W_2 - W_1]^2)^{\frac{1}{2}}$$

找到一个直线方程的参数有几个简便的方法,对于直线上任何两个点给定一个坐标。所有包括简短的、直接的使用坐标的代数步骤的方法和模型方程 $y = c + bx$ 。在两个已知点中间,任何选定的媒介性时间,线性内插是方便可行的:刚好建立了一个包含两个时间的变量值的比例。更精巧的方法是(指已知两点):假定两个点之间的弯曲关系(三次多项式);第三个内插的普通方法是得到两条直线的交点;第四个要求内插的通行方法是具有一个平面和一条直线的三维重点的估计,给定平面的方程和在平面每一侧

的两个离散测度,这个方法包括三个步骤:

1. 使用坐标值得到以时间定义的方程;
2. 替换那些定义为一个平面的方程以得到交的时间;
3. 插入交的时间为定义线段的方程,在交的时间得到 x 、 y 和 z 。

以上这些概念不仅适用于标准相空间而且适应于滞后相空间。

向量是一个有大小和方向的量,起始点在图形的原点的向量是坐标向量,并由它的终点坐标定义,相空间中的数据点也是一个坐标向量。向量的维数是它具有的元素(或坐标)的个数。结合式是两个向量的和,确定一个结果的两个方法是三角定律和平行四边形定律。由标量乘以向量产生另外一个向量。两个向量的点积(或内积或标量积)是它们的 x 坐标的积的和, y 、 z 轴的以此类推。

概率是一个极限相对频率。通常概率的估计是总体或样本事件的发生数和所有可能发生的数的比率。概率和频率分布是一个表明变量的值或各个类别所具有的总概率为 1 的表或图型。对于连续随机变量估计概率分布在数理统计是一个很大的问题。基于各种各样的原因,一般的直方图通常是不适合的。统计学已经提出了许多可能的程序,其中在混沌理论中有三个采用直方图是有用的,即核密度统计量。

对信息而言,概率论是封闭和可逆的。也就是,对于概率而言,估计一个数值,它的逆是数值性估计的基础。假如试验的结果和可能的类别是相等的(等概的),信息 I 是 $\log(1/P)$,它也等于状态或类别数的对数。一般而言,各个类别和结果的概率不是相等的,这样,我们必须加权每一类的信息。可能的序列或路径的数量戏剧性(几何性)地随时间而增加,这个增加,或者是可能状态的数量,或者是事件(时间)的数量的增加:路径 = 状态的事件次方,也就是 $r = S^c$

自相关

自相关或序列相关是一个反映变量随时间与自身的(正或负)相关程序。它测试两个值 X_t 和 X_{t+m} 之间的相互程序在时间序列之内。基于统计理论、预测和混沌分析,它的重要性是要发现原始数据是否是自相关。从数学上讲,对于给定一个滞后(自相关系数或序列相关系数),自相关就是对一个方差的自协方差(auto covariance)。

自协方差是一个对于特征程度的维数测试(在平方单位)。

时间序列数据的基础分析

检验时间序列数据的第一步是绘出数据,做一个自相关测试,并适用傅立叶分析。假如数据是自相关的,与混沌测试之前使数据脱离相关的可取性意见相左的是某些对于使数据脱离相关的理由的自相关数据:

- (1)可能给出一个错误的伪相空间标绘;
- (2)可能导致相关维和李亚普诺夫指数的错误估计;
- (3)对于许多统计分析是不适合的;
- (4)可能影响我们的预测方法的选择。

另一方面,不相关的数据在混沌性数据中探查确定性和确认非线性上会导致很大的问题。

去掉周期性和倾向性是在自相关系数中说明过度变异的两个主要方法。去掉周期性和倾向性是用滤波方法变换数据。滤波(线性滤波)是任何一种系统改变输入序列的数学技巧,诸如时间序列或幂谱。使它们成为一种新的输出序列,该序列有某种所要求的性质。有许多滤波类型:标准化是一个消除了周期性的滤波或者转换,它变换无论什么单位的数据为一个通用的或标准的标度(Scale);精典的简约是一个去掉周期和倾向性的滤波,它是一种将时间序列分为倾向季节性、循环性和噪声的传统的方法。随着标准化,计算是由季节数据所做的,这里主要的工具是移动平均

方法。

差分是另外一种去掉周期性和倾向性的选择方法,在通常的计算中,部分是因为简洁。逻辑斯蒂方程,是描述混沌许多主要特征的最流行的方法。它主要被用来说明控制参数的重要性。依赖于那个参数的值,迭代能够导致一个不动点(零或某个正数),周期加倍,混沌,以及我们还没有探查到的许多特征。控制参数类似某种调谐的旋钮,控制着系统的行为。当参数从零开始增加,对于混沌的完全的路径具有一个控制参数的临界值。通常涉及这一过程的术语是映射、区间、区间映射到自身、不可逆、可逆、轨道、框架和折叠。

吸引子是代表一类系统的各种可能稳定状态的相空间条件下的集合,轨道一般渐近的接近吸引子,而且至少大部分数学例子,不能全部得到一一映射的吸引子。

三种类型状态的混沌吸引子

点吸引子(不动点,在相空间中占零维);

周期性吸引子或“极限环”(两个或更多的值依序发生,并在相空间中占有二维);

环形图纹曲面(一类两个或更多极限环的连接表示,通常在相空间中采用三维周期的,在这里连接轨道几乎但不准确的重复自身)。

当控制参数增加时,混沌能够从各种路径出现。由于控制参数较小的指数性的增加,周期加倍是一个周期性的系统加倍,直到混沌形成。几个周期加倍的“普适性”法则描述了临界值,吸引子值的变化率,以及控制参数的变化率。普适性是指所有经历周期加倍的系统均可照此办理。

并非任何非线性方程能够导致混沌,例如,迭代一个简单的幂规则($y = ax^b$ 或指数规则 $y = ac^{bx}$)未必能产生混沌。对于产生混沌的非线性方程,必须提及的必要条件是在方程中的不可逆性,而

且至少有一个不稳定的不动点。参数的合适选择在一般的非线性方程中通常能够满足一个或更多的这样的条件并导致混沌。

混沌的特征通常如下：

1. 混沌来自于一个确定性过程；
2. 它仅仅在非线性系统里发生；
3. 对于大部分看起来是无组织和混乱的形式和运动尽管怀疑,但实际上,就随机性质而言(指高斯性质)通常能够通过所有统计上的检验；
4. 它发生于反馈系统——系统过去的事件影响今天的事件,今天的事件影响明天的事件；
5. 它可能来自相对简单的系统。对于离散的时间,混沌能够在只有一个变量的系统内发生；对于连续时间,它能够在三个变量之多的系统内发生；
6. 对于给定条件和控制参数,它是完全自产生的(Self-generated)。换句话说,其他(外部)变量的改变是不必要的；
7. 它不是数据不准确的结果,诸如样本误差或测试误差。 X_i (正确或错误)的任何特定值,只要控制参数在合适的范围之内,就能导致混沌；
8. 尽管是不连接的出现,混沌包括一个或更多的秩序或结构的类型；
9. 变量的范围具有有限的边界。在一个相空间中,边界限制吸引子于某个特定域；
10. 混沌行为的细节对于初始条件的改变是高度敏感的；
11. 长期行为的预测是无意义的,理由是无法达到对初始条件的敏感性和测度变量的无限精度；
12. 短期预测能够相对准确；
13. 关于初始条件的信息不可避免的损失。用数学家的行话来说是“不可逆的”。换句话说来说,我们无法确定一个混沌系统以

前的历史;

14. 无论在哪里,傅立叶谱是“宽”的(大部分不涉及噪声),它具有某种周期性的持续;

15. 相空间轨道可能有分形的性质;

16. 当控制参数系统性增加,初始的非混沌系统对于称作路(routes)的混沌服从几个选择的方案之一。

初始条件的敏感性依赖即是指一个在变量的初始值中,表面上无意义的差异可能随时间在产出上导致巨大的差异。测度误差、噪声或数据值的舍入可能导致在输入中的这种微小的误差。在这种意义上,对于初始条件的敏感性暗示着对测度误差和对噪声的敏感性。对初始条件的敏感性是混沌非常重要的特征,但它不是一个简单明了的指示器,随机数据能够表现它。对于初始条件的敏感性的实际结果是长期预测无意义,尽管系统可能是确定性的和由已知的法则控制。

不管它的名字和显现,混沌还是具有大量的秩序和规则性。例子就是窗口(在混沌统治内的),来自于窗口之内的混沌的各种各样的路径,是在相空间内的唯一的几何形状(混沌吸引子)和在吸引子上的较有名望的带状物,以及吸引子的分形结构。因为这些秩序化的特征,你可能看到在两个新显现出来的概念环境下注意到的混沌。一个是自组织——过程靠什么自组织系统,没有任何明显的外部影响,使自己看似无规则性或一致性的进入一个形式或结构。另一个是复杂性——在许多独立的动因中连续的平等交换所定义的动态行为,导致了一个随时间形成的较大和较复杂结构的可适用的自组织系统。

分形是一条线、曲面或一个形式在很大一个规律范围看起来相同的,换句话说,分形是自相似,而且规模不变。一个混沌吸引子通常是分形的,分形的两个主要类型是确定性分形和自然分形:确定性分形是由迭代等式或遵循特定的数值要求产生的,这样它