

中国人口发展的R/S分析

陈 嵘 王 放 李后强 艾南山

引 言

近年来,人口系统的非线性特性研究越来越受到学术界的重视。这是因为,人们在认识和分析人口系统时面临许多经典人口学理论所难以解决的问题,迫切要求新方法和新手段的引入。在非线性和科学研究中,非线性人口学被列为重点项目。四川大学人口研究所和物理系已承担1992年国家青年社会科学基金资助的“混沌与人口系统的内禀随机性”研究项目,对人口系统非线性特性作了一些尝试性的探索^①。本文是系列研究的一部分。下面拟就中国人口系统的时间过程中所显示出的分形特征进行剖析。

分形理论由法国著名数学家曼德尔布罗特(B.B.Mandelbrot)在70年代中期创立。它是近年在国际上兴起的非线性科学中的前沿数学工具,主要研究和揭示复杂的自然和社会现象中所隐藏着的有规律性、层次性和标度不变性,是一门横跨自然科学、社会科学和思维科学的新学科,被称之为继突变论、协同学、耗散结构和混沌学后的又一探索复杂性对象的新方法。目前已在许多领域取得辉煌成果。人们普遍认为,自从分形科学诞生以后,就再也不能用老眼光看世界了。分形概念和原理可升华为一种方法论——分形论。它为人们通过部分推及整体、从有限中认识无限提供了新的指导。国际国内已召开多次关于分形理论与社会学、哲学问题的研讨会,一股分形热潮正席卷整个科学界。但是,国内人口学的分形研究却只是刚刚开始。

所谓“分形”,指英文“Fractal”,原意是“破碎的、极不规则的”,实际是指其组成部分以某种方式与整体相似的形(Shape)^②,或者是指在很宽的尺度范围内,无特征尺度却有自相似性或自仿射性的一种现象^③。因此,分形的外表结构极为复杂和无规,但内部却是有规可寻的。如人口出生率、死亡率随时间的变化等等,都是分形理论的研究对象。描述分形特征的定量参数是分维(fractal dimension)。它除具有通常的欧几里德维数(点0维、直线一维、平面二维、立体三维……)的内涵外,还包含着刻画客体复杂程度和不规则性的深刻意义,将经典维数进行了连续化拓展,即可以取分数维,如2.4维、1.5维等。分维是研究分形所必需的最少信息量。分维 D_q 的类型在理论上有无穷多种($q=-\infty, \dots, -1, 0, 1, \dots, +\infty$),每种分维都有其描述对象和自身的含义。本文仅用到“盒子维数”,定义为:用大小为 ε 的盒子去覆盖分形曲线,覆盖完时所必需的最少盒子数 $N(\varepsilon)$,则盒子维数(也称覆盖分维、局域分维) D_B 为:

$$D_B = -\lim_{\varepsilon \rightarrow 0} \frac{\lg N(\varepsilon)}{\lg \varepsilon}$$

就人口发展过程而言,可将观测数据(例如总人口数、出生率、死亡率等)按年代顺序

① 参见王放、何承金、艾南山、李后强:《分形人口学初探》,载《大自然探索》,1991年2期。

② Feder J., Fractals, Plenum, 1988。

③ 郝柏林:《分形和分维》,载《科学杂志》,1986年第1期。

组成一时间序列。以时间为横坐标,观测数据为纵坐标,各点相连成曲折线段。对这种非光滑(不可微分)的分形曲线,用经典数学处理显得很困难,而用分形理论则比较容易。本文用分形理论中的R/S分析方法^①求出解放以来中国人口发展过程曲线的分维值,由此找出其变化规律,并预测将来的变化趋势。

一、人口发展过程的分形特征

人口系统是一个复杂的开放系统,内部有众多的交互作用因素。对人口系统的发展过程和未来的演化趋势可用分形理论来阐述。图1显示解放后40年中国人口的出生率、死亡率变化情况。从图1可见,人口的发展过程具有分形的特征,无论是出生率、死亡率还是自然增长率及总人口数,都沿着这种分形曲线作波动式演化。从几何角度看,这种演化呈现出的波动规律通过仿射坐标变换具有不变性,即通过某些方向的拉伸和压缩,不同区段的波形与整体相似(自仿射性)。这便是在历史长河这条时间轴上人口发展过程所呈现出的典型分形特征^②。

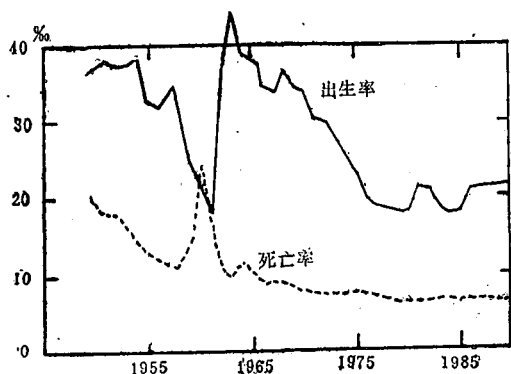


图1 中国人口的出生率、死亡率变化(1949~1990)

另一方面,由于人口系统的开放性,其发展过程时刻都受到外界偶然因素和内在随机作用的干扰,故人口过程不仅受确定性规律的支配,同时也受偶然性、随机性因素的影响,但其统计学规律却是十分明显的。在人口预测中,宋健等学者的方法是:首先作出一系列假设,然后定量地建立起决定人口发展过程的主要因素(即出生、死亡、迁移等)之间的平衡关系,最后得到描述人口发展过程的数学方程式。本文所采用的R/S分析方法与之相比,不同之处在于,基于人口发展同时具备确定性和随机性的特点,不对未来环境作出假设,从某一变化过程的历史状态出发,预测该过程在与过去相同的环境条件下,未来的自然发展状态。基本方法是考虑一时间序列 $B(t)$,定义如下量:

$$\text{差值序列} \quad \xi(t) = B(t) - B(t-1) \quad t=1, 2, 3, \dots$$

$$\text{均值序列} \quad \langle \xi \rangle_\tau = \frac{1}{\tau} \sum_{t=1}^{\tau} \xi(t) \quad t=1, 2, 3, \dots$$

$$\text{累积离差} \quad X(t, \tau) = \sum_{\mu=1}^t \{\xi(\mu) - \langle \xi \rangle_\tau\} \quad t=1, 2, \dots, \tau$$

$$\text{极差} \quad R(\tau) = \max_{1 \leq t \leq \tau} X(t, \tau) - \min_{1 \leq t \leq \tau} X(t, \tau) \quad \tau=1, 2, \dots$$

$$\text{标准偏差} \quad S(\tau) = \left\{ \frac{1}{\tau} \sum_{t=1}^{\tau} [\xi(t) - \langle \xi \rangle_\tau]^2 \right\}^{\frac{1}{2}} \quad \tau=1, 2, \dots$$

研究发现,有如下标度关系^③:

$$\frac{R(\tau)}{S(\tau)} = (a \tau)^H \quad (1)$$

① 黄登仕、李后强:《分形几何学: R/S分析与分式布朗运动》,载《自然杂志》1990年第8期。

② 李后强、程光钰:《分形与分维》,四川教育出版社,1990年版。

③ Hurst H.E. et al., Long-Term Storage: An Experimental Study, Constable, 1965.

其中 a 为常数, H 称作赫斯特指数或 H 指数, 它与盒子维数的关系为^①,

$$D_B = 2 - H \quad (2)$$

理论证明^②了关联函数

$$C(t) = \frac{\langle -\Delta B(-t)\Delta B(t) \rangle}{\langle \Delta B(t)^2 \rangle} = 2^{3-2D_B} - 1 \quad (3)$$

$\langle \dots \rangle$ 表示求平均, $\Delta B(t)$ 是未来的增量, $\Delta B(-t)$ 是过去的增量。关联函数 $C(t)$ 反映了事物发展的未来状态与过去历史的相关特性。可见, 当 $H=0.5$ 即 $D_B=1.5$ 时, $C(t)=0$ 。过去 $\Delta B(-t)$ 与未来 $\Delta B(t)$ 无关; $H>0.5$ 即 $D_B<1.5$ 时, $C(t)>0$, 则呈正相关。即过去一段时间内的增长趋势意味着未来相同时间间隔内也有一个增长趋势, 反之亦然。并且 H 偏离 0.5 越远, 这种相关性越明显。当 $H<0.5$ 时, $C(t)<0$, “过去”与“未来”有负相关特性。因此, 只要求出一定时间范围内某一人口过程的 H 指数, 便可由(2)式得出盒子维数 D_B , 再由(3)式分析过去的发展与未来趋势之间的相关特性。 H 指数可根据(1)式拟合求得。

二、中国人口的未来发展趋势

本文用 R/S 法分析影响我国人口发展的两个重要因素: 出生率、死亡率以及二者的直接作用结果——总人口数和自然增长率的发展趋势。所采用的分析数据是从1949~1990年这四项指标各年的变化数据(见表1)。

表1 中国历年总人口数、出生率、死亡率、自然增长率(1949~1990)

年份	总人口数 (万人)	出生率 (‰)	死亡率 (‰)	自然增长率 (‰)	年份	总人口数 (万人)	出生率 (‰)	死亡率 (‰)	自然增长率 (‰)
1949	54 167	36.00	20.00	16.00	1970	82 992	33.43	7.60	25.83
1950	55 196	37.00	18.00	19.00	1971	85 229	30.65	7.32	23.33
1951	56 300	37.80	17.80	20.00	1972	871 77	29.77	7.61	22.16
1952	57 482	37.00	17.00	20.00	1973	89 211	27.93	7.04	20.89
1953	58 796	37.00	14.00	23.00	1974	90 859	24.82	7.34	17.48
1954	60 266	37.97	13.18	24.79	1975	92 420	23.01	7.32	15.69
1955	61 465	32.60	12.28	20.32	1976	93 717	19.91	7.25	12.66
1956	62 828	31.90	11.40	20.50	1977	94 974	18.93	6.87	12.06
1957	64 653	34.03	10.80	23.23	1978	96 259	18.25	6.25	12.00
1958	65 994	29.22	11.98	17.24	1979	97 542	17.82	6.21	11.61
1959	67 207	24.78	14.59	10.19	1980	98 705	18.21	6.34	11.87
1960	66 207	20.86	25.43	-4.75	1981	100 072	20.91	6.36	14.55
1961	65 859	18.02	14.24	3.78	1982	101 590	21.09	6.60	14.49
1962	67 295	37.01	10.02	26.99	1983	102 764	18.62	7.08	11.54
1963	69 172	43.37	10.04	33.33	1984	103 876	17.50	6.69	10.81
1964	70 499	39.14	11.50	27.64	1985	105 044	17.80	6.57	11.23
1965	72 538	37.88	9.50	28.38	1986	106 529	20.77	6.69	14.08
1966	74 542	35.05	8.83	26.22	1987	108 073	21.04	6.65	14.39
1967	76 368	33.96	8.43	25.53	1988	109 614	20.78	6.58	14.20
1968	78 534	35.59	8.21	27.38	1989 ^①	111 191	20.83	6.50	14.33
1969	80 671	34.11	8.03	26.08	1990 ^②	113 368	20.98	6.28	14.70

资料来源: ①中国社会科学院人口研究所:《中国人口年鉴》(1990), 经济管理出版社, 1991。

②《中华人民共和国国家统计局关于1990年人口普查主要数据的公报》第1号, 《人民日报》1990. 10. 30。

③其余数据取自国家统计局编《中国统计年鉴》1989, 中国统计出版社, 1990。

① 同第28页③; ② 同第28页①及第27页②。

将以上四个分析指标作为 $B(t)$ ，计算出H指数，记为 $\frac{y_2}{y_1} H(y_1' - y_2')$ 。 y_1 、 y_2 是选取数据的起始、终止年份， $y_1' - y_2'$ 表示该起止年内所拟合的区间。为便于直观分析，定义并绘制H指数曲线图，简称H图形。横坐标为年份、纵坐标为H值。在某一起止年内，例如1949~1990年，将 ${}_{99}^{}H(1949 \sim 1951)$ （该点的横坐标为1951年，以下类似）、 ${}_{99}^{}H(1949 \sim 1952)$ 、…、 ${}_{99}^{}H(1949 \sim 1990)$ 共40个点连成一条曲线，起点为(1949, 0)点，然后，将拟合区间后移一年，即 ${}_{99}^{}H(1950 \sim 1952)$ 、 ${}_{99}^{}H(1950 \sim 1953)$ 、…、 ${}_{99}^{}H(1950 \sim 1990)$ 共59个点连线。以此类推则得1949~1990年间H图形。

(一) 总人口数。作1949~1990年中国总人口数的 $l_n \tau - l_n R/S$ （见图2），得到拟合直线，斜率即为 ${}_{99}^{}H(1949 \sim 1990) = 0.74$ ，具有较为明显的正相关。为了得到详细情况，可根据数据作出中国总人口数的H图形，其中包含40条H指数曲线，这些曲线按其走向趋势大致分为三组：1949~1958年、1959~1961年、1962~1990年，同时在1966年左右有弯曲。这三组H曲线相对集中，故选出

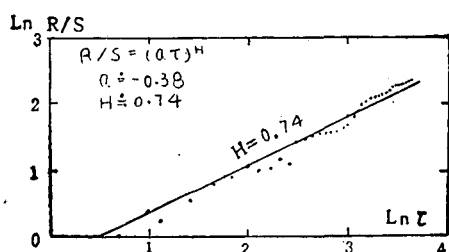


图2 1949~1990年间中国总人口数的 $\ln \tau$ 和 $\ln R/S$ 关系

${}_{99}^{}H(1959 \sim 1990) = 0.71$ $D_B = 1.29$
作出1959~1990年的H图形，发现其中H值均大于0.5， D_B 则小于1.5，而且还有1974年以来的突变。选出

$${}_{74}^{}H(1974 \sim 1990) = 0.62 \quad D_B = 1.38$$

通过对总人口数H指数变化的分析，我们可以看到，40余年来，中国人口数量是逐步增加的，未来40年里必将有与过去相同的变化，即人口继续增加。而且由于H值偏离0.5较远，维数接近1，故这种正相关较明显。这正是所谓的“滞后效应”。虽然中国从1974年开始开展计划生育，1980年提倡一对夫妇只生一个孩子，使得人口增长率大为下降，但由于人口的惯性增长、生育率控制的不稳定性以及中国的人口基数大，中国人口在未来四五十年间仍将继续增长。从1986~1990年，中国人口的自然增长率一直在14%以上，1990年达14.7%。即使是在未来的40年中，人口增长率能控制在年平均10%左右，低于解放后40年中的最低水平（1959~1961年除外）即1984年的水平，预计到2030年，中国的总人口仍将达16.88亿。宋健等在1979年底曾进行过中国人口的百年预测^①：如果妇女的总和生育率保持在2.3，则到1990年中国的总人口数可达11.19亿（与第四次人口普查数据相差不是很大），而到2030年可达16.88亿，与从1990年开始到2030年年平均递增10%所达到的人口数也相差不大。由此看来，今后中国妇女的总和生育率要想降到2.3以下并非易事。

解放40年来，中国的经济形势变化很大，特别是人口政策的重大改变，导致人口发展过程产生一些突变。这些突变也能通过H指数得到反映。1949~1958年间，新中国的国民经济迅速恢复和发展，城乡人民的物质生活水平和身体健康水平都有一定程度的提高，形成了人口迅速增长的局面。这期间的H值约为0.5， D_B 约为1.5。这表明解放初期的10年，中国人口发展在纯自然状态下运行。由于没有采取任何人口控制措施，使得人口数量持续高速增长。

1951~1961年3年间，H值陡然上升且突破1的界限。这是一种反常现象。这期间由于

① 宋健、于景元：《人口控制论》，科学出版社，1985年。

1958年后的“大跃进”和“反右倾”的错误，加上严重的自然灾害，使得出生率大幅度下降，而死亡率则迅速上升，造成解放以来绝无仅有的总人口下降。

由H值反映的1962、1966、1974年等的突变，原因十分明显：1962年以后人口出现补偿性回升。到1966年左右，由于“文化大革命”造成生育上的无政府主义状态，直到1974年H值开始下降，但仍保持远大于0.5的势头。1973年8月国务院成立计划生育领导小组，在全国范围内大力开展计划生育工作，人口增长速度放慢，但以后每年的净增人数仍相当可观。这是由于约等于1的分维使这种增长过程具有强大的惯性。分段考虑的结果与整体分析是一致的，从一个侧面反映了人口发展的分形特征。

(二) 出生率。计算得到： $H(1949\sim1990)=0.58$ ， $D_H=1.42$ ，说明过去的发展与未来趋势之间有较弱的正相关。因过去出生率总的变化是下降的，所以未来总的发展趋势也将是下降的。随着经济的发展和城镇化水平的提高，同时继续执行计划生育政策，必将对出生率有一定程度的控制作用。此外，由于人口年龄结构的变化，年轻人在总人口中的比例逐渐减少，也将导致出生率的下降。但是，出生率的H值并不明显大于0.5，因此，这种“过去”与“未来”的正相关是很微弱的。图3是1949~1990年间的出生率H图形的简化(从众多的H曲线中抽取几条具有代表性的曲线)，的确可见H值偏离0.5不远，而且在1979年前后有明显的弯曲，故将分析的起点移至1978年，有：

$$H(1978\sim1990)=0.52 \quad D_H=1.48$$

$$H(1979\sim1990)=0.56 \quad D_H=1.44$$

两值平均约为0.54，很接近于无相关性的随机过程。这不仅说明了整个90年代中国的出生率将会有起伏，而且不是“持续稳定”地下降，同时也显示80年代虽然中国强有力地实施了计划生育政策，使出生率大大下降，但出生率仍然波动较大，在80年代初和后半期明显高于中期(参见图1，表1)。在图3中，从1979年到1986年前后，H曲线弯成“V”型，且在0.5之下(“V”字的最低点在1981年)。因此，受其负相关影响的年限就将是1987年到1994年前后，也就是说在这7年左右的时间内出生率将有回升。这与实际情况十分吻合。因为在这期间中国面临着第三次生育高潮。通过对出生率H值的分析，我们看到，中国未来控制出生率的任务十分艰巨，无论是使出生率继续下降还是保持现有水平都不是轻而易举的事。

(三) 死亡率。图4a为1949~1990年中国死亡率H指数曲线的简化图。该图中1959年以后H曲线十分集中，表明其置信度高，约为0.4；在图4b中，1959~1990年间的H值介于0.5~0.6之间。平均H值应在小于0.5的方向，逼近0.5。说明死亡率在未来30年与前30年有较弱的负相关特性，同时存在波动起伏，即在2020年前中国死亡率水平将在波动中伴有缓慢上升的趋势。这种变化趋势表明中国现阶段的死亡率是过去与今后一定时期内的最低水平。但是，70年代以后由于出生率和生育率降低，人口年龄结构从年轻型向中老年型过渡，老年人口比重相对提高，而老年人口的死亡率较高，这就势必造成总人口的死亡率上升。因此，随着中国人口老龄化的逐步到来，中国人口死亡率会有一定幅度的上升。

(四) 自然增长率。计算表明，自然增长率的H曲线变化与出生率H曲线变化十分相似。这是因为自然增长率为出生率与死亡率之差，而死亡率的H曲线在0.5附近变化平缓。故与

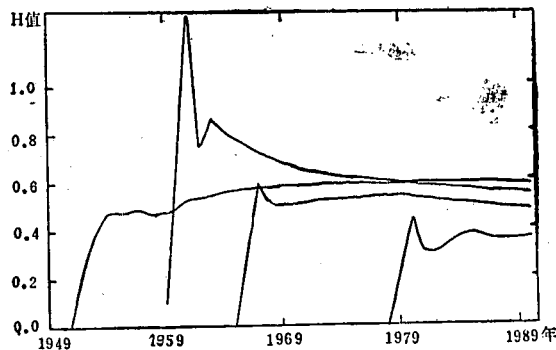


图3 中国出生率H指数曲线简化图(1949~1990)

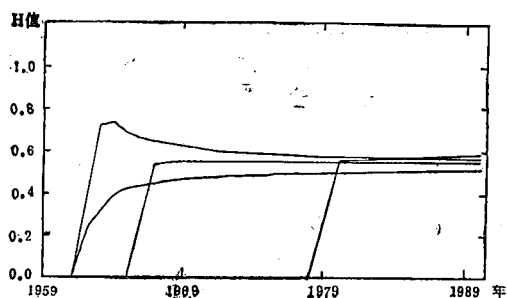
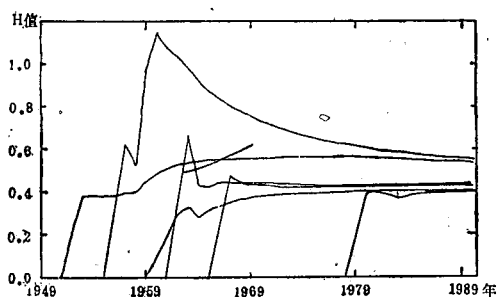


图4a 中国死亡率H指数曲线简化图(1949~1990) 图4b 中国死亡率H指数曲线简化图(1959~1990)

出生率对比, 仍选取几个主要H指数:

$$\frac{1}{4}H(1949\sim1990)=0.59 \quad D_B=1.41$$

$$\frac{7}{8}H(1978\sim1990)=0.57 \quad D_B=1.43$$

$$\frac{7}{8}H(1979\sim1990)=0.61 \quad D_B=1.39$$

这些值与出生率相应值较接近, 只是后两值偏大些, 因此可得结论, 自然增长率在未来一定时期内呈下降趋势(比出生率下降较明显), 但也伴有波动性的回升, 现阶段从1986年左右开始的回升将持续到1994年前后。

三 结 语

本文引入分形理论中的R/S分析方法, 对中国1949~1990年间的总人口数、出生率及死亡率和自然增长率这4个序列的分形特征进行了研究, 得到赫斯特指数(H)和分维值(D_B), 进而分析了这些序列的发展特性并预测了未来中国人口的发展趋势。总的来看, 在未来40年中, 中国的出生率和自然增长率将有不太明显的下降趋势, 但伴有起伏; 死亡率则有一定幅度的上升, 总人口数将持续增长到下个世纪中期。这些发展趋势意味着中国人口发展逐步走向稳定化。但要彻底实现中国人口稳固的转变还需要50~60年的时间。

通过以上分析, 我们还看到, 中国人口的增长速度能降低到现今水平已属不易, 要巩固和发展已有成果, 需要我们乃至几代人作进一步的努力。分析结果同时表明, 分形与R/S分析方法有巨大的优势和潜力, 只要合理运用于人口学中, 有可能成为探讨人口系统发展特性及规律的常规手段。

(本文责任编辑: 宋黎明)

(作者工作单位: 四川大学人口研究所、物理系)