

东南大学科技出版基金资助

非线性时间序列分析及其应用

王海燕 卢 山 著

科学出版社

北 京

内 容 简 介

本书以来自于确定性非线性系统的观测或实验时间序列为研究对象,在对问题的背景和意义进行分析的基础上,根据目前国内外关于单变量非线性时间序列分析的相关文献,总结了单变量非线性时间分析的基本流程,对单变量非线性时间序列分析的基本方法进行了详细综述。由于实际问题中常常可以获得多变量时间序列,本书把单变量非线性时间序列分析方法推广到多变量非线性时间序列的情形,着重研究了基于多变量时间序列的系统非线性性检验方法、多变量时间序列相空间重构方法和多变量非线性时间序列的预测方法等,最后把这些方法应用到证券市场的指数时间序列中。

本书自成体系,可作为系统工程、管理科学、金融工程、应用数学、生物医学工程、信号处理等专业高年级本科生、研究生和从事相关领域研究的科技工作者的参考书。

图书在版编目(CIP)数据

非线性时间序列分析及其应用 / 王海燕, 卢山著. —北京: 科学出版社, 2006

ISBN 7-03-018035-6

I. 非… II. ①王…②卢… III. 非线性-时间序列分析 IV. O211.61

中国版本图书馆 CIP 数据核字(2006)第 108009 号

责任编辑:何舒民 / 责任校对:赵 燕

责任印制:吕春珉 / 封面设计:耕者设计工作室

科 学 出 版 社 出 版

北京东黄城根北街 16 号

邮政编码:100717

<http://www.sciencep.com>

印刷

科学出版社发行 各地新华书店经销

*

2006 年 11 月第 一 版 开本:B5 (720× 1000)

2006 年 11 月第一次印刷 印张:11 1/2

印数:1—2 000 字数:230 000

定价:30.00 元

(如有印装质量问题,我社负责调换〈科印〉)

销售部电话 010-62136131 编辑部电话 010-62137026(BA03)

前 言

近年来,非线性科学得到了快速发展。实际问题中,面临大量复杂而无法直接建立解析数学模型的非线性系统,对这些非线性系统可以通过实验或观测的手段获得非线性时间序列,这些非线性时间序列中蕴含着丰富的系统动力学信息,如何提取这些信息并应用到实际问题中去解释、分析,甚至控制原来的复杂系统是非线性时间序列应用的一个重要方面。从 20 世纪 80 年代以来,非线性时间序列分析的相关理论已取得了许多重大成果,在实际复杂系统中也得到了有效的应用,为解决实际问题中像经济系统、生理系统、电力系统等这些用传统的确定性和随机性方法都无法很好刻画的复杂系统提供了一种全新的研究工具。

基于混沌理论的非线性时间序列分析目前在许多领域都引起广泛兴趣,其中的某些成果已在国内外得到应用,但目前国内系统介绍这方面的著作还很少,应用工作多是某一具体方法在实际问题中的应用,缺乏对理论和应用的系统总结。作者在给相关专业的博士生和硕士生上课时对非线性时间序列分析方法进行过介绍,引起了学生的广泛兴趣,所以在本书第二章对目前得到广泛应用的单变量非线性时间序列的基本流程进行了总结和归纳,对其分析方法进行了详细综述。同时作者结合近年来把单变量非线性时间序列分析方法推广到多变量非线性时间序列情形的研究实践,在本书第三至第五章详细介绍了基于多变量时间序列的系统非线性性检验方法、多变量时间序列相空间重构方法和多变量非线性时间序列的预测方法等,并在第六章把这些方法应用到证券市场的指数时间序列中,且进行了实证分析。本书介绍的非线性时间序列分析方法,也可应用到其他领域,例如,经济系统的汇率时间序列、生理系统的心(或脑)电图时间序列、电力系统电力负荷时间序列、水文预报的水沙(或流量)时间序列、机械系统故障诊断的振动信号检测时间序列等。

本书内容自成体系,学术上力求完整严谨,应用上力求通俗易懂,把复杂问题用简单的语言说清楚,使对这一领域感兴趣的读者通过对本书的阅读能很快进入这一领域并掌握非线性时间序列分析这一工具。对理论性较强的内容和相关的应用成果,本书标注了大量的参考文献,使不同学科领域的读者可以通过本书对非线性时间序列分析这一领域有一个较全面的了解。

本书在写作过程中参考或引用了许多学者的资料,作者已尽可能在文中作了标注,并在参考文献中列出,在此,谨对他们表示衷心的感谢。若某些引用的资料由于作者疏忽没有标注其出处,作者深表歉意。

本书由王海燕负责全书结构的策划和具体的撰写,卢山负责相关仿真分析的具体计算,研究生汤龙坤、朱梅、钱志强、赵巍、方芬等在仿真分析中也做了大量具体工作,借此机会向他们表示衷心的感谢。

本书得到了东南大学科技出版基金的资助,在此表示衷心的感谢。

由于非线性时间序列分析及其应用是理论和实践都很强的研究领域,尚有许多问题有待深入分析和研究,因此书中难免有不足之处,恳请读者批评指正。

王海燕

2006年6月于东南大学

目 录

第一章 绪论	1
1.1 时间序列的含义及分类	1
1.2 非线性时间序列的例子	2
1.3 研究非线性时间序列的意义	10
第二章 单变量非线性时间序列分析	12
2.1 非线性时间序列分析流程	12
2.2 相空间重构	16
2.2.1 延迟时间间隔的确定	17
2.2.2 嵌入维数的确定	19
2.3 几何不变量的计算	22
2.3.1 关联维数	22
2.3.2 Kolmogorov 熵和 Renyi 熵	24
2.3.3 Lyapunov 指数	25
2.4 观测时间序列平稳性的检验	28
2.5 基于观测时间序列的系统非线性性检验	30
2.5.1 零假设及替代时间序列的约束生成算法	31
2.5.2 判别统计量的选取	38
2.5.3 统计检验方法	39
2.6 基于观测时间序列的系统确定性检验	40
2.6.1 从非线性预测判断系统的确定性	41
2.6.2 利用递归图判断系统的确定性	43
2.7 观测时间序列噪声处理技术	45
2.7.1 噪声级别的估计	46
2.7.2 噪声的降低	47
第三章 基于多变量时间序列的系统非线性性检验	50
3.1 随机变量的线性冗余和广义冗余	50
3.2 时间序列广义冗余的计算	52
3.2.1 直方图及盒计数法	53
3.2.2 关联积分算法	53
3.3 系统非线性性的定性和定量检验	55
3.3.1 系统非线性性的定性检验	56

3.3.2	系统非线性性的定量检验·····	57
3.4	仿真模拟·····	59
第四章	多变量时间序列相空间重构 ·····	70
4.1	多变量时间序列相空间重构的流程·····	70
4.2	多变量时间序列中变量间的依赖关系·····	72
4.2.1	随机变量间统计依赖性的度量方法 ·····	73
4.2.2	观测时间序列统计依赖性的计算 ·····	76
4.2.3	应用举例·····	79
4.3	多变量时间序列相空间重构参数的确定·····	83
4.3.1	利用预测误差最小法确定嵌入维数 ·····	84
4.3.2	利用虚假最近邻点法确定嵌入维数 ·····	85
4.3.3	虚假最近邻点法确定嵌入维数算法的改进·····	86
4.3.4	嵌入维数算法的仿真计算·····	87
4.4	多变量时间序列重构相空间中几何不变量的计算·····	90
4.4.1	广义关联维数的计算 ·····	90
4.4.2	小数据量情况下最大 Lyapunov 指数的计算 ·····	94
4.5	多变量时间序列相空间重构中噪声的影响·····	98
第五章	多变量非线性时间序列预测方法 ·····	102
5.1	多变量非线性时间序列的局域预测法 ·····	102
5.1.1	局部平均预测法 ·····	103
5.1.2	局部线性预测法 ·····	103
5.1.3	局部多项式预测法 ·····	105
5.2	多变量非线性时间序列的全域预测法 ·····	106
5.2.1	多项式逼近预测法 ·····	107
5.2.2	神经网络预测法 ·····	107
5.2.3	径向基函数预测法 ·····	108
5.3	各种预测方法的预测效果对比分析 ·····	112
5.3.1	预测效果评价 ·····	112
5.3.2	仿真比较 ·····	113
5.4	基于正则化的多变量非线性时间序列预测方法 ·····	117
5.4.1	奇异值分解 ·····	117
5.4.2	最小二乘估计 ·····	118
5.4.3	正则化估计 ·····	120
5.4.4	基于正则化的局部线性和局部多项式预测的步骤 ·····	123
5.4.5	Lorenz 系统的仿真模拟 ·····	124

5.5	基于正则化的多变量非线性时间序列的自适应预测方法	129
5.5.1	基于正则化的自适应预测的步骤	129
5.5.2	Henon 映射的仿真检验	132
第六章	非线性时间序列分析法在证券市场中的应用	139
6.1	基于单变量时间序列的证券市场非线性性和确定性检验	139
6.1.1	样本数据及平稳化处理	139
6.1.2	证券市场的非线性性检验	141
6.1.3	证券市场的确定性检验	143
6.2	基于多变量时间序列的证券市场非线性性检验	145
6.2.1	样本数据及平稳化处理	145
6.2.2	证券市场的非线性性检验	147
6.3	上海证券市场单变量指数序列的预测研究	153
6.3.1	样本数据及相空间重构	154
6.3.2	基于正则化的自适应预测	156
6.4	上海证券市场多变量指数序列的预测研究	158
6.4.1	样本数据及相空间重构	159
6.4.2	局部多项式预测	161
6.4.3	基于正则化的局部线性和局部多项式预测	162
参考文献	165

第一章 绪 论

对实际问题中动力系统进行研究的一种常用方法是通过对系统本身的性质分析,用差分方程或微分方程建立系统的数学模型,然后通过对数学模型的分析来研究系统的演化规律。从数学上看,这是一个正问题,系统的演化规律能否正确刻画取决于对系统所做的假设和在此假设基础上所建立的数学模型的正确性。实际问题中有些系统很难通过这种方法来建立数学模型,如像经济系统、生理系统、社会系统等要建立精确的解析模型是非常困难的。但对这些系统通过实验或观测可以获得包含系统演化信息的时间序列,需要研究的是能否通过这些时间序列本身去获取所研究系统的相关信息以及如何通过这些时间序列去揭示所研究系统的演化行为。从数学上看,这是一个反问题,它对所研究系统本身的背景信息和知识要求较少,只是从获取的时间序列本身出发进行研究,因此研究起来会更困难,但对那些无法建立精确解析模型的实际系统提供了一种新的研究方法。时间序列分析法正是从这样的角度去研究复杂系统演化行为的一种方法。

1.1 时间序列的含义及分类

从纵向看,时间序列是指存在于自然科学或社会科学中的某一变量或指标的数值或观测值,按照其出现时间的先后次序,以相同的或不同的间隔时间排列的一组数值。它是某一现象或若干现象在不同时刻上的状态所形成的数据,反映的是现象以及现象之间关系的发展变化规律。从横向看,时间序列也可以是若干相关现象在某一时间点上所处的状态按一定顺序排序的一组数据,反映的是一定时间、地点条件下各相关现象之间存在的内在数值联系。因此,从系统的意义上看,时间序列就是某一系统在不同时间(地点、条件等)的响应。

时间序列是指所研究系统的历史行为的客观记录,因而它包含了系统结构特征及其运行规律,所以可通过对时间序列的研究来认识所研究系统的结构特征,揭示所研究系统的运行规律,进而用以预测、控制其未来行为,修正和重新设计系统使之按照希望的结构运行。

时间序列根据研究的出发点不同,可以有不同的分类。

1) 按研究系统复杂程度不同,可以分为线性时间序列和非线性时间序列。线性时间序列是指从线性系统通过观察或实验获取的时间序列。而非线性时间序列是指从非线性系统通过观察或实验获取的时间序列。线性时间序列比较简单,可以

通过线性回归等方法建立数学模型进行分析;非线性时间序列比较复杂,直接建立数学模型比较困难,有时甚至具有混沌特征或分形特征。

2) 按研究系统的确定性程度不同,可以分为随机时间序列和确定性时间序列。随机时间序列是指从随机系统通过观察或实验获取的时间序列,一般用统计分析方法进行研究。确定性时间序列是指从确定性系统通过观察或实验获取的时间序列,可以通过建立确定性数学模型进行研究,但当确定性系统具有混沌特征时,对这样的时间序列进行分析会非常复杂。

3) 按研究系统的观察变量多少不同,可以分为单变量时间序列和多变量时间序列。单变量时间序列是指从系统中只观察一个量所得到的时间序列。多变量时间序列是指从系统中同时观察多个变量所得到的多维时间序列。多变量时间序列中观察量之间可能是相关的,这样的多变量时间序列就有冗余。

本书研究的时间序列主要是指来自于非线性确定性系统的实测单变量或多变量时间序列。

1.2 非线性时间序列的例子

来自于非线性确定性系统的时间序列具有各种各样的特性,有些随时间的演化会趋向于一个稳定的状态,有些随时间的演化会周期性变化,还有些随时间的演化会表现出看起来是随机的行为。

1. 趋向于稳定状态的时间序列

例 1.1 考虑以下 Logistic 模型

$$x_{n+1} = R x_n (1 - x_n) \quad (1.1)$$

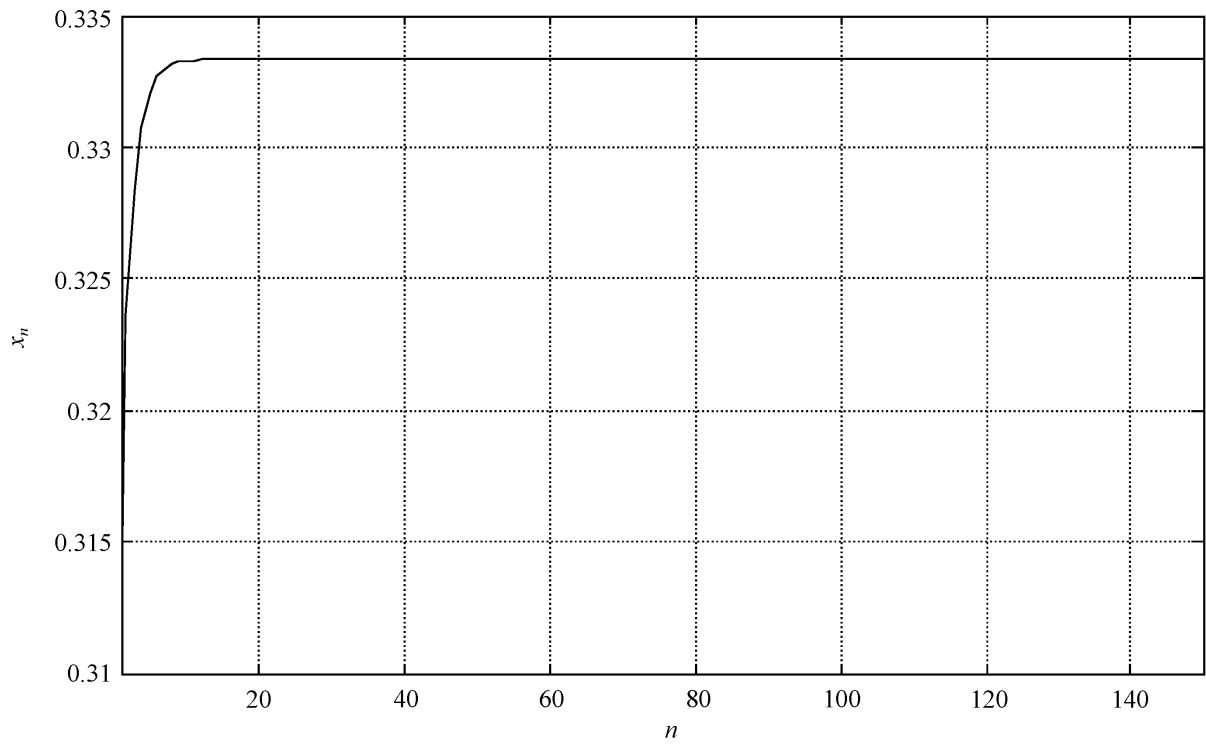
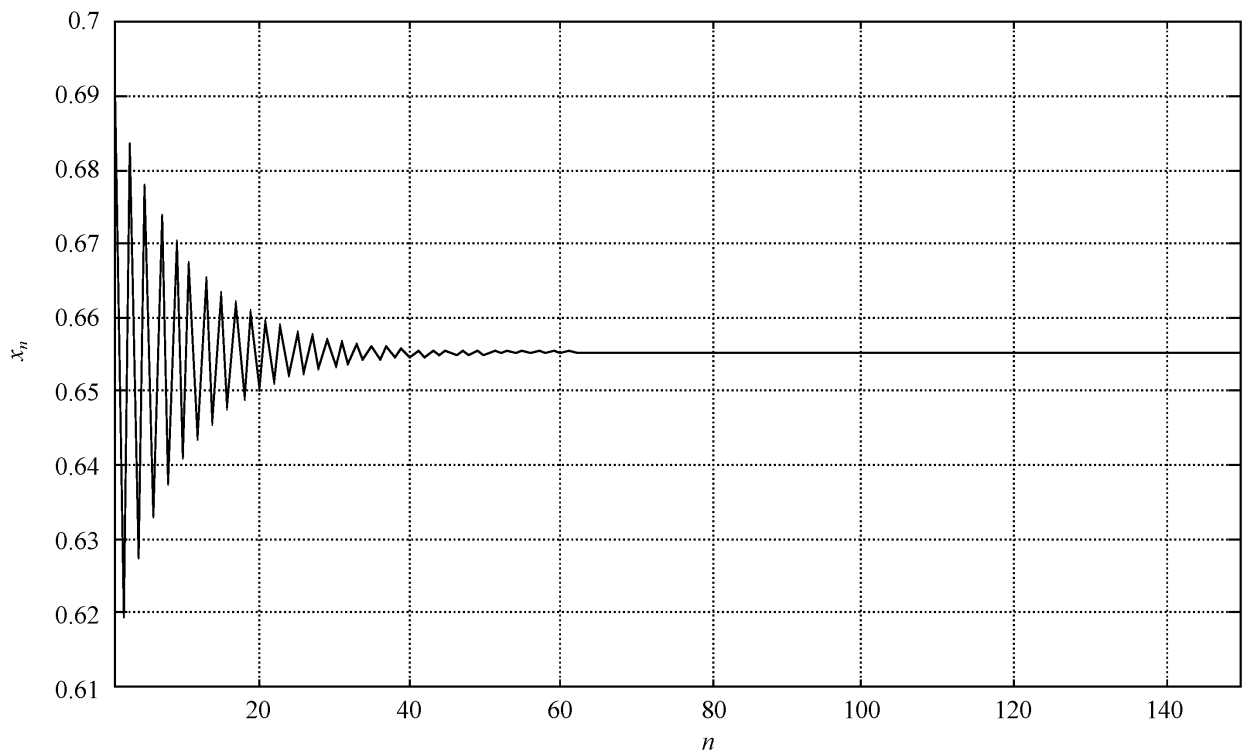
当 $R = 1.5$ 时,不管初始状态 x_0 在何处,随时间的演化,系统都将单调地趋向于 $\frac{1}{3}$,如图 1.1 所示($x_0 = 0.7$)。

当 $R = 2.9$ 时,不管初始状态 x_0 在何处,随时间的演化,系统都将交替地趋向于 $\frac{19}{29}$,如图 1.2 所示($x_0 = 0.7$)。

例 1.2 弹簧振子在策动力、弹性力和阻尼力的作用下做受迫振动,其振动方程为

$$\frac{d^2 x}{dt^2} + 2\beta \frac{dx}{dt} + \omega_0^2 x = f_0 \cos \alpha t \quad (1.2)$$

取物体所受的最大策动力 $f_0 = 100$,策动力的频率 $\alpha = 1$,振动物体的固有频率 $\omega = 0$,阻尼因子 $\beta = 0.1$ 。当初值 $x(0) = 5, x'(0) = 0$ 时,随时间的演化,系统将趋于稳定状态,如图 1.3 所示。

图 1.1 Logistic 模型中 $R=1.5$ 时的时间历程图图 1.2 Logistic 模型中 $R=2.9$ 时的时间历程图

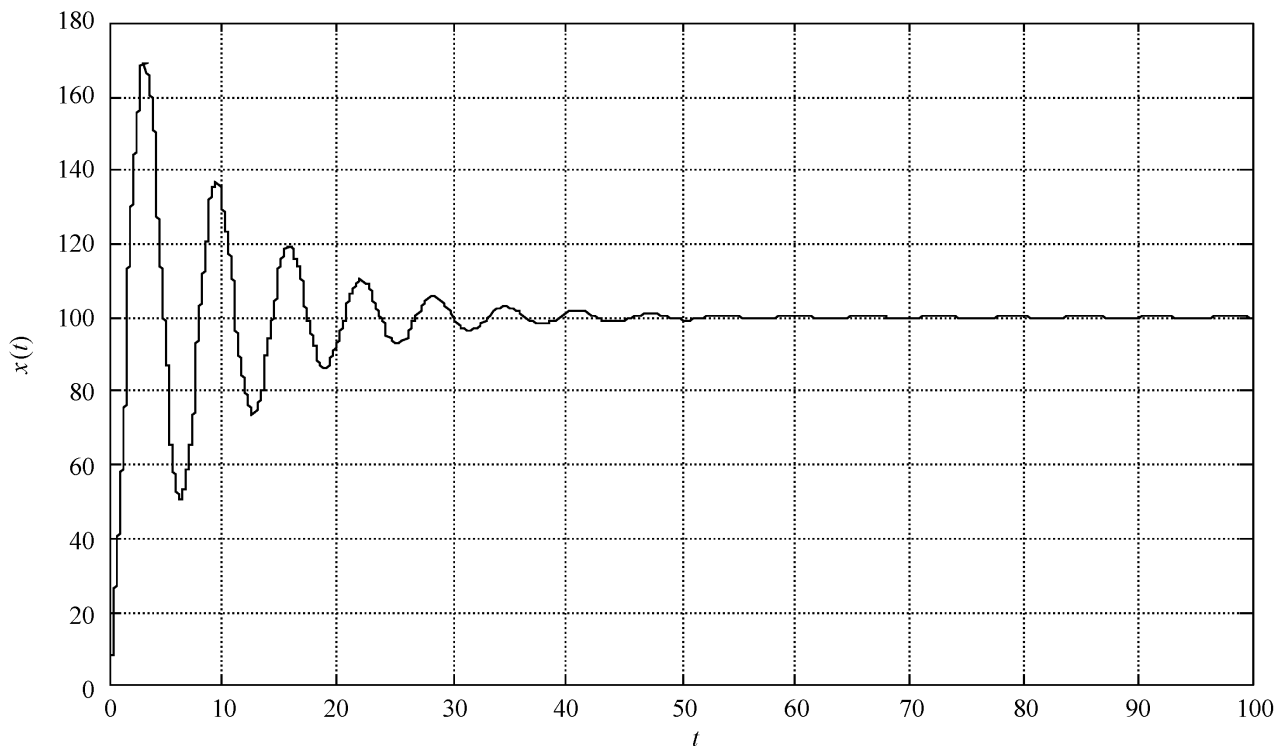


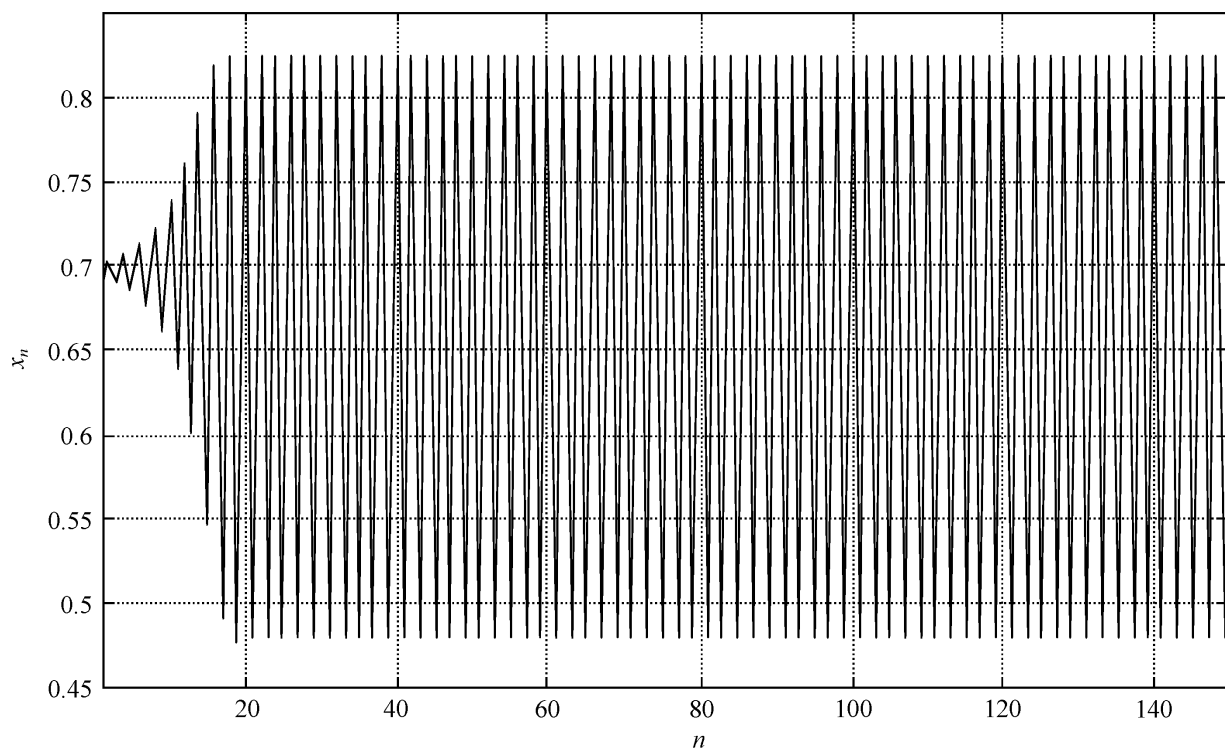
图 1.3 具有阻尼的振动方程的时间历程图

2. 趋向于周期状态的时间序列

例 1.3 考虑以下 Logistic 模型

$$x_{n+1} = 3.3x_n(1 - x_n) \quad (1.3)$$

不管初始状态 x_0 在何处, 随时间的演化, 系统都将在 0.48 和 0.82 两个状态之间周期性地变化, 如图 1.4 所示 ($x_0 = 0.7$)。

图 1.4 Logistic 模型中 $R = 3.3$ 时的时间历程图

例 1.4 弹簧振子在策动力、弹性力和阻尼力的作用下做受迫振动,其振动方程为

$$\frac{d^2x}{dt^2} + 2\beta \frac{dx}{dt} + \omega_0^2 x = f_0 \cos \alpha t \quad (1.4)$$

取物体所受的最大策动力 $f_0 = 100$, 策动力的频率 $\alpha = 1$, 振动物体的固有频率 $\omega = 1$, 阻尼因子 $\beta = 0.1$ 。当初值 $x(0) = 5, x'(0) = 0$ 时, 随时间的演化, 系统将趋于周期性变化状态, 如图 1.5 所示。

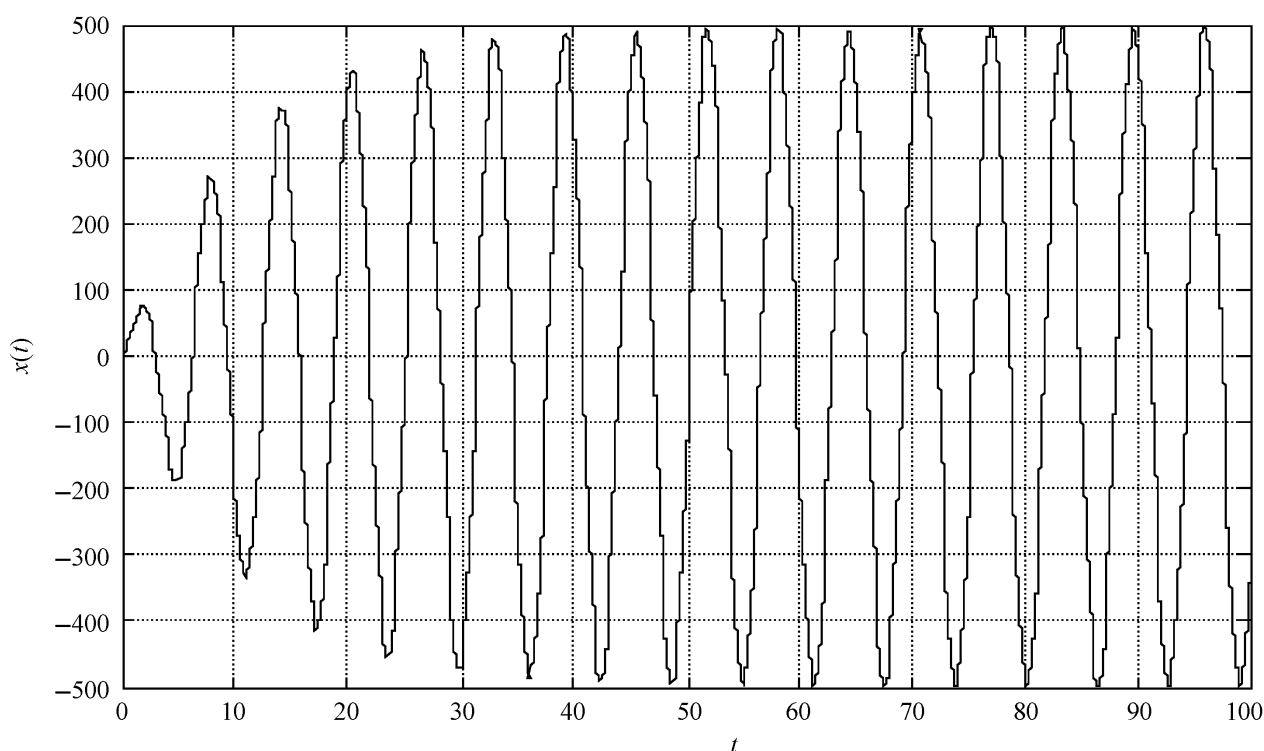


图 1.5 无阻尼的振动方程的时间历程图

3. 混沌时间序列

例 1.5 考虑以下 Logistic 模型

$$x_{n+1} = 4x_n(1 - x_n) \quad (1.5)$$

随时间的演化, 系统将出现不规则的振荡, 看起来好像是随机的, 如图 1.6 所示。

在 Logistic 模型 (1.5) 中, 初值的微小变化随着时间的演化会不断放大, 表 1.1 是 Logistic 模型 (1.5) 的三个非常接近的初值演化 50 次后的结果。可以看到刚开始时演化结果非常接近, 但随着演化次数的增加, 演化结果不断分离, 最后甚至变得毫无关系, 表明系统对初值具有非常敏感的依赖性, 也说明这样的系统只能进行短期预测, 要进行较长时间的预测会变得不正确。

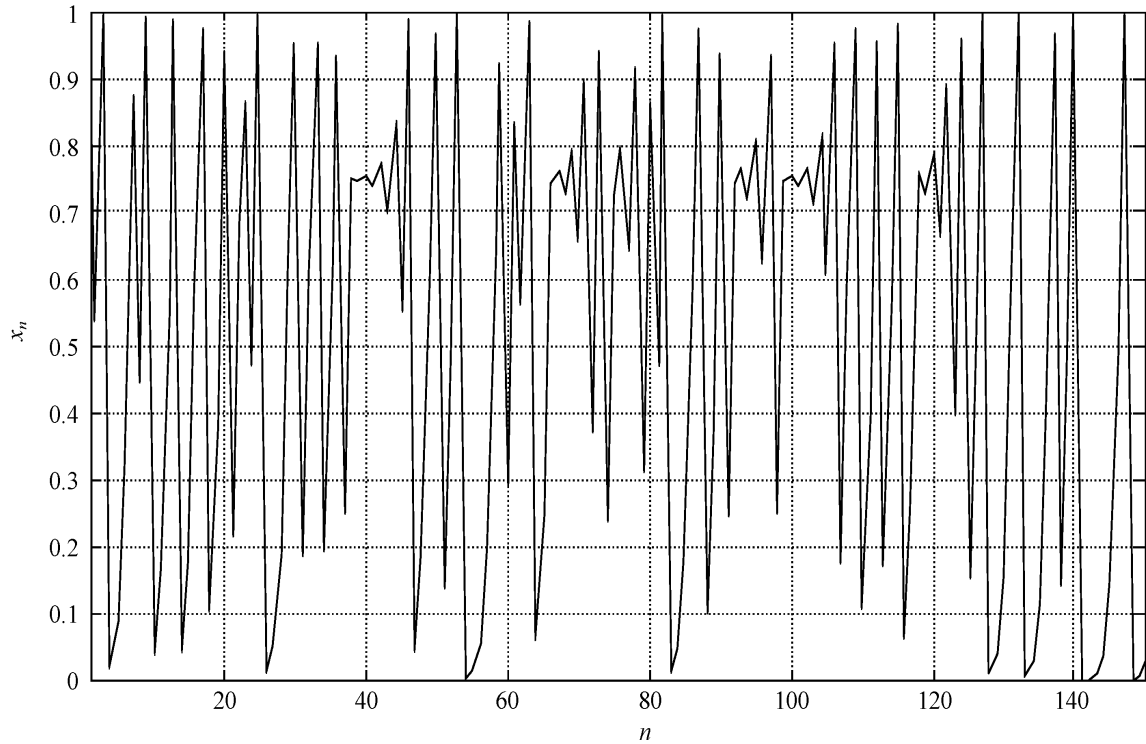


图 1.6 Logistic 模型中 $R=4$ 时的时间历程图

表 1.1 Logistic 模型 (1.5) 在不同初值下的演化结果

n	$x_n^{(1)}$	$x_n^{(2)}$	$x_n^{(3)}$
1	0.100 000 000 000 000 0	0.100 000 010 000 000 0	0.100 000 020 000 000 0
2	0.360 000 000 000 000 0	0.360 000 031 999 999 6	0.360 000 063 999 998 4
3	0.921 600 000 000 000 0	0.921 600 035 839 995 4	0.921 600 071 679 981 8
4	0.289 013 760 000 000 1	0.289 013 639 118 858 4	0.289 013 518 237 736 8
5	0.821 939 226 122 649 8	0.821 939 022 088 530 7	0.821 938 818 054 328 7
6	0.585 405 387 334 197 4	0.585 421 064 226 722 4	0.585 421 589 719 127 5
7	0.970 813 326 249 438 0	0.970 812 967 145 496 7	0.970 812 608 039 428 2
8	0.113 339 247 303 761 2	0.113 340 599 870 613 8	0.113 341 952 444 447 2
9	0.401 973 849 297 512 3	0.401 978 033 166 332 8	0.401 982 217 042 111 4
10	0.961 563 495 113 812 8	0.961 566 776 072 237 8	0.961 570 056 896 080 9
11	0.147 836 559 913 285 3	0.147 824 444 905 122 6	0.147 812 330 307 794 6
12	0.503 923 645 865 163 6	0.503 889 513 574 460 0	0.503 855 381 267 096 2
13	0.999 938 420 012 499 1	0.999 939 486 736 616 4	0.999 940 544 141 141 3
14	0.000 246 304 781 624 1	0.000 242 038 406 114 3	0.000 237 809 295 438 2
15	0.000 984 976 462 314 7	0.000 967 919 294 097 3	0.000 951 010 968 708 6
16	0.003 936 025 134 733 6	0.003 867 929 705 349 5	0.003 800 426 187 384 2
17	0.015 682 131 363 489 3	0.015 411 875 300 575 8	0.015 143 931 792 713 7
18	0.061 744 808 477 550 3	0.060 697 397 601 181 4	0.059 658 372 490 285 4

续表

n	$x_h^{(1)}$	$x_h^{(2)}$	$x_h^{(3)}$
19	0. 231 729 548 414 483 7	0. 228 052 894 102 501 9	0. 224 397 004 328 382 9
20	0. 712 123 859 224 412 5	0. 704 179 086 375 899 8	0. 696 171 955 107 322 4
21	0. 820 013 873 390 966 5	0. 833 243 602 746 811 4	0. 846 066 256 117 482 7
22	0. 590 364 483 349 241 7	0. 555 794 804 913 301 4	0. 520 952 585 507 315 6
23	0. 967 337 040 596 098 5	0. 987 547 758 978 746 6	0. 998 243 956 642 234 5
24	0. 126 384 361 947 522 4	0. 049 188 730 859 208 1	0. 007 011 838 677 964 5
25	0. 441 645 420 010 560 2	0. 187 076 798 462 674 1	0. 027 850 691 185 274 7
26	0. 986 378 971 977 024 3	0. 608 316 279 758 520 4	0. 108 300 120 743 108 5
27	0. 053 741 982 474 292 1	0. 953 070 334 157 095 7	0. 386 284 818 360 546 6
28	0. 203 415 127 176 099 9	0. 178 909 089 227 110 4	0. 948 275 429 858 824 5
29	0. 648 149 652 848 124 1	0. 587 602 508 076 145 0	0. 196 196 555 939 544 4
30	0. 912 206 721 443 921 2	0. 969 303 202 315 075 8	0. 630 813 869 508 022 5
31	0. 320 342 475 185 814 1	0. 119 018 017 187 260 0	0. 931 550 926 177 352 3
32	0. 870 892 695 110 560 7	0. 419 410 915 088 292 4	0. 255 055 192 461 877 9
33	0. 449 754 434 854 498 7	0. 974 021 597 572 374 4	0. 760 008 165 040 449 3
34	0. 989 901 532 732 836 8	0. 101 214 100 139 735 8	0. 729 583 016 449 193 9
35	0. 039 985 952 904 069 1	0. 363 879 224 290 557 4	0. 789 166 554 232 356 7
36	0. 153 548 305 897 690 6	0. 925 884 537 681 038 5	0. 665 530 815 653 542 0
37	0. 519 884 894 614 559 3	0. 274 489 442 256 832 4	0. 890 398 196 276 292 4
38	0. 998 418 363 864 671 5	0. 796 579 953 385 461 8	0. 390 356 993 376 870 0
39	0. 006 316 538 249 855 6	0. 648 161 324 999 509 2	0. 951 913 644 394 561 1
40	0. 025 106 558 377 574 7	0. 912 192 887 097 559 3	0. 183 096 232 040 104 9
41	0. 097 904 876 416 032 6	0. 320 388 095 304 714 9	0. 598 288 007 411 283 7
42	0. 353 278 046 359 975 8	0. 870 958 254 766 927 3	0. 961 357 870 396 477 7
43	0. 913 890 673 280 218 5	0. 449 559 892 881 101 7	0. 148 595 661 692 907 6
44	0. 314 778 042 286 589 9	0. 989 823 182 375 336 2	0. 506 059 964 075 818 1
45	0. 862 771 305 523 246 9	0. 040 293 000 030 792 6	0. 999 853 107 341 599 2
46	0. 473 587 919 555 836 3	0. 154 677 896 717 244 7	0. 000 587 484 323 791 0
47	0. 997 209 608 026 444 0	0. 523 010 579 937 496 3	0. 002 348 556 743 841 2
48	0. 011 130 422 744 759 7	0. 997 882 052 843 760 4	0. 009 372 164 100 248 7
49	0. 044 026 145 737 130 6	0. 008 453 845 824 331 8	0. 037 137 306 561 306 9
50	0. 168 351 376 914 654 3	0. 033 529 513 260 441 1	0. 143 032 508 090 713 6

例 1.6 考虑以下 Lorenz 系统

$$\begin{cases} \frac{dx}{dt} = \sigma(y - x) \\ \frac{dy}{dt} = x(r - z) - y \\ \frac{dz}{dt} = xy - bz \end{cases} \quad (1.6)$$

其中

$$\sigma = 10, r = 28, b = \frac{8}{3}$$

取初值 $x_0 = 15.34, y_0 = 13.68, z_0 = 37.91$, 利用四阶 Runge-Kute 法积分, 积分步长 $h = 0.04$, 得关于变量 x 的长为 8500 的时间序列。为了减少瞬态的影响, 去掉前 5000 个数据, 剩下的 3500 个数据的时间历程如图 1.7 所示。

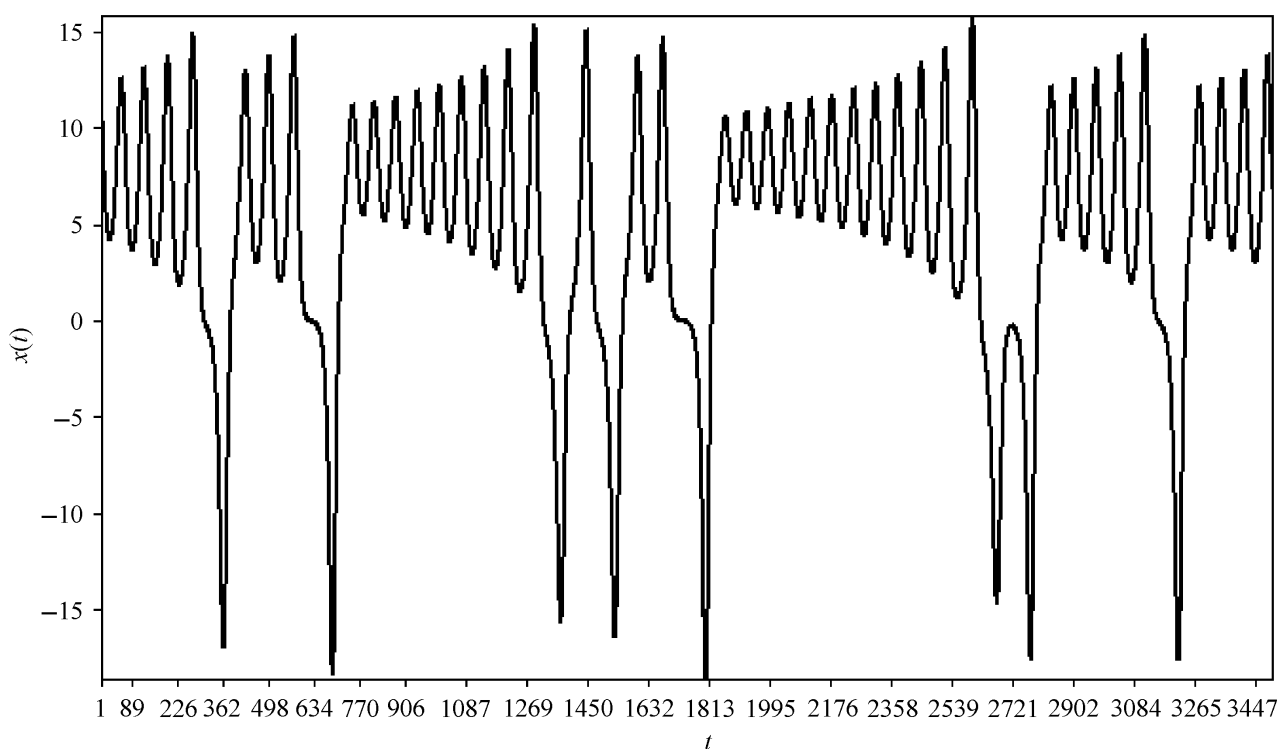


图 1.7 Lorenz 系统变量 x 的时间历程图

从图 1.7 中可以发现, 随时间的演化, 系统将出现不规则的振荡, 看起来好像是随机的, 实际上此时系统是混沌的。

4. 实际问题中的实测时间序列

实际问题中有大量的时间序列, 如天文中的太阳黑子时间序列、水文中的径流量时间序列、气象中的降雨量时间序列、生理系统中的心电图时间序列和经济系统中的股票指数时间序列等。

例 1.7 股票指数时间序列。

图 1.8 是从 1999 年 1 月 4 日至 2003 年 12 月 31 日的上海证券交易所综合股指时间序列,共 1196 个数据,可以看出股票指数时间序列表现为非常复杂的演化行为。

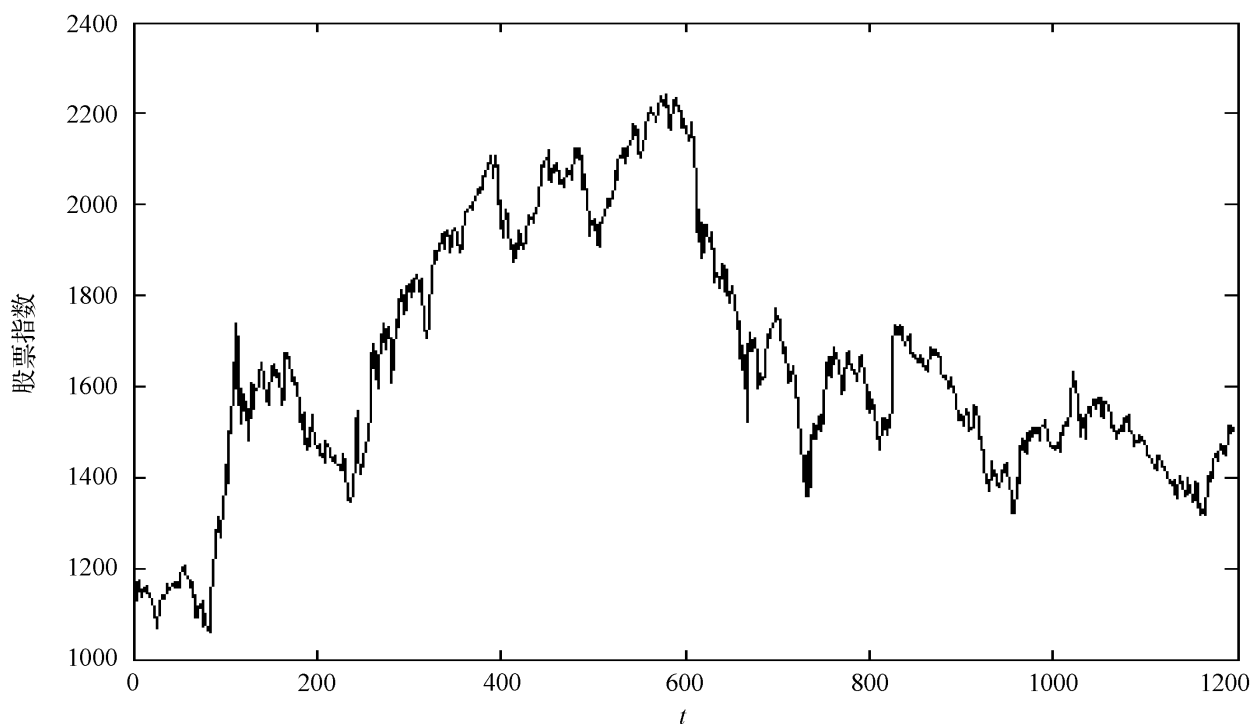


图 1.8 上海证券交易所综合股指时间序列历程图

例 1.8 太阳黑子数时间序列。

图 1.9 是 1700 年以来的太阳黑子数时间序列,可以看出太阳黑子时间序列也表现为非常复杂的演化行为。

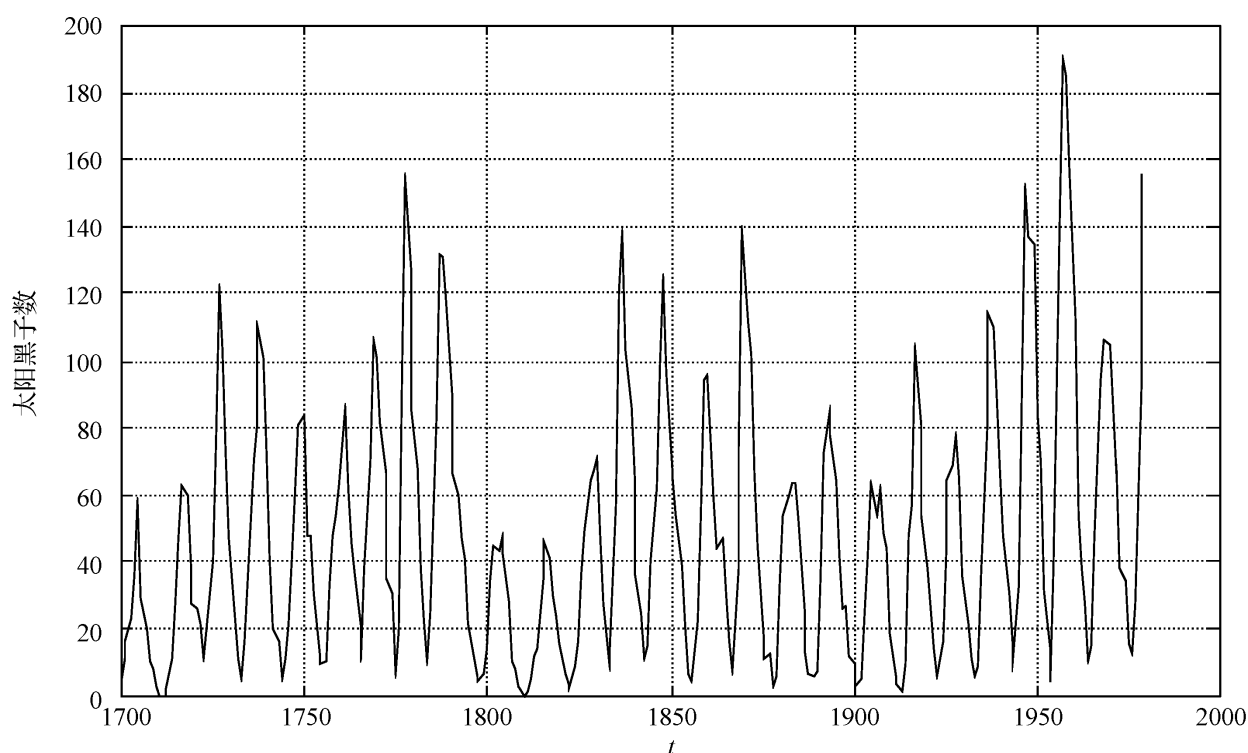


图 1.9 太阳黑子数时间序列历程图

从上面的时间序列例子可以看出,不同的时间序列具有不同的特性。例如,趋向于稳定状态、趋向于周期循环或出现看起来是毫无规律的混沌特征,具有相同特性的时间序列可能来自于用差分方程描述的非线性系统,也可能来自于由微分方程描述的系统。

实际问题中通过观测或实验手段获得的时间序列常常是比较复杂的,在不知道实际系统解析模型的情况下,分析观测时间序列的演变规律是掌握所研究系统动力学特性的重要手段。

1.3 研究非线性时间序列的意义

自 20 世纪 80 年代以来,有相当多的学者利用混沌、分形等非线性动力系统理论来研究实际问题中的复杂系统,并取得了不少研究成果。非线性动力系统理论是定量研究复杂系统的一种较好的方法,对于那些其演化过程不可预测但其不可预测性可以预测的复杂系统,即确定性系统内具有内在随机性的复杂系统,混沌理论成为一种有效的研究工具。利用非线性动力系统理论研究复杂系统需要建立非线性动力学模型,但实际问题中的复杂系统往往表现为多变量动态演化行为、多层次结构,由于获取的信息通常是不完备的,并具不确定性,且系统的结构参数和边界条件具有时变性和复杂性,因此往往难以直接建立起精确的解析形式的数学模型,常常依赖于通过观测或实验手段获取的时间序列进行分析。

如果通过观测或实验手段获得的时间序列是来自于线性系统,则时间序列可表示成不同幅度的正弦波的叠加,因此可用传统的线性方法和 Fourier 变换方法进行处理。但当获得的时间序列来自于非线性系统时,无论多么高阶次的线性模型都不可能对该非线性系统的行为做出恰当的解释;相反,一个低阶非线性模型或许能很好地刻画该动力系统,对这样的时间序列,非线性时间序列分析方法成为重要的研究工具。

经典的时间序列分析方法是把时间序列看成随机过程进行研究的,因而整个理论是以随机过程作为基础的。随着混沌现象的发现,人们发现在确定性系统内部也存在随机性,但这种复杂系统的本质特征往往不是随机因素而是非线性动力系统内的混沌因素造成的,对从这样的系统获取的时间序列,用随机过程方法分析显然是不合适的,而应当用混沌时间序列分析方法进行研究。

混沌运动与牛顿力学中运动的三种定常状态(静止、周期运动和准周期运动)不同,它是一种始终限于有限区域、轨道永不重复、性态复杂的运动,它具有对初值的极端敏感性、长期不可预测性及分形结构等特点,它的时间历程是非周期的或被认为是周期无穷大的、貌似随机的运动曲线。混沌运动不是随机运动,而是属于确定性运动的范畴,因为它是源于确定性系统,受到确定性激励,因而反映了确定性

系统的内在随机性。如果一个时间序列是由一个确定性系统通过观测或实验手段获得的,要考虑的是以下反问题:如何由时间序列来恢复并刻画原复杂系统?非线性时间序列分析就是研究如何通过复杂系统的观测或实验获得的时间序列来辨识和重构原复杂系统,分析原复杂系统的性质,刻画原复杂系统的特征,并对原复杂系统进行预测和控制。因此,研究非线性时间序列分析方法对解决实际问题中复杂系统的相关问题具有重要的意义。

第二章 单变量非线性时间序列分析

非线性时间序列分析方法的研究对象是未知数学模型或无法建立解析数学模型的复杂系统,研究过程中具有的信息是从复杂系统中通过观测或实验手段获得的单变量或多变量时间序列。研究目标就是寻求反映复杂系统本质特征的不变量,分析复杂系统的特征、内在变化规律及内部关系,最终解释、指导甚至控制实际问题中具体的复杂系统。

2.1 非线性时间序列分析流程

非线性时间序列分析可以看作是如图 2.1 所示的输入输出过程,输入的是没有任何背景知识的通过观测或实验手段获得的实际复杂系统的时间序列,输出的是所研究系统的某些特征。

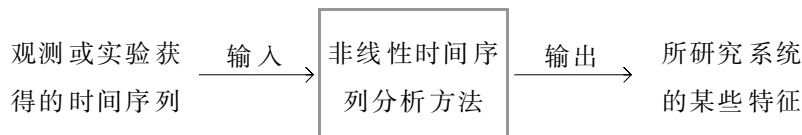


图 2.1 非线性时间序列分析的框架模型

为了具体说明如何从所研究的系统获得的非线性时间序列出发去研究系统的特征,下面通过两个例子来引出非线性时间序列分析的相空间重构思想。

例 2.1 考虑以下 Henon 映射

$$\begin{cases} x_{n+1} = 1 - 1.4x_n^2 + y_n \\ y_{n+1} = 0.3x_n \end{cases} \quad (2.1)$$

该系统虽然有两个状态变量 x_n, y_n , 但由于

$$x_{n+1} = 1 - 1.4x_n^2 + 0.3x_{n-1} \quad (2.2)$$

表明该系统实际上只与状态变量 x_n 的前两个时刻的状态有关,因此,如果未知该系统的数学模型,且只观测到状态变量 x_n 的有关信息,我们也可以从状态变量 x_n 来建立原系统的模型。对状态变量 x_n 进行相空间重构,即令

$$x_n = (x_n, x_{n-1}) \quad (2.3)$$

从而

$$x_{n+1} = G(x_n) = G(x_n, x_{n-1}) = 1 - 1.4x_n^2 + 0.3x_{n-1} \quad (2.4)$$

也就是说,只要能找到函数 G ,就可以从观测时间序列 $\{x_n\}$ 重构原来的系统。

例 2.2 考虑以下 Lorenz 系统

$$\begin{cases} \frac{dx}{dt} = \sigma(y - x) \\ \frac{dy}{dt} = x(r - z) - y \\ \frac{dz}{dt} = xy - bz \end{cases} \quad (2.5)$$

其中

$$\sigma = 10, r = 28, b = \frac{8}{3}$$

取初值 $x_0 = 15.34, y_0 = 13.68, z_0 = 37.91$, 利用四阶 Runge-Kute 法积分, 积分步长 $h = 0.04$, 得关于状态变量 x, y, z 的长为 2000 的时间序列, 从而可以画出如图 2.2 所示的 Lorenz 吸引子。另外, 如果我们不知道该系统的数学模型, 只观测到状态变量 x 的时间序列 x_n , 对状态变量 x_n 进行相空间重构, 令

$$x_n = (x_n, x_{n-4}, x_{n-8}) \quad (2.6)$$

从而可得如图 2.3 所示的 Lorenz 吸引子的重构图。由图 2.3 可以看出, 实际 Lorenz 吸引子与重构的 Lorenz 吸引子非常类似。

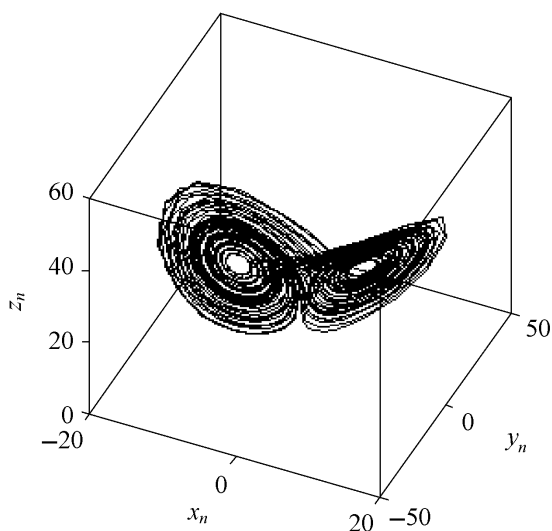


图 2.2 Lorenz 吸引子

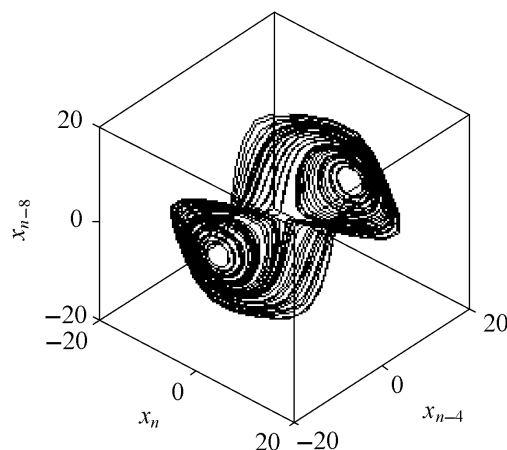


图 2.3 Lorenz 吸引子按 (x_n, x_{n-4}, x_{n-8}) 重构的相图

上面两个例子中的系统都具有混沌特征, 产生的时间序列是混沌时间序列, 例子中的相空间重构就是混沌时间序列分析的基础。相空间重构的基本原理是 F. Takens^[1] 和 R. Mane^[2] 的延迟嵌入定理, 它建立了观测信号系统时间波动和动力系统空间特征之间的桥梁。

单变量混沌时间序列分析的基本思想可以描述如下:

对未知数学模型的混沌动力系统, 通过观测或实验手段可以获得单变量时间

序列 $\{x_n\}$, 做以下相空间重构

$$x_n = (x_n, x_{n-\tau}, \dots, x_{n-(m-1)\tau}) \quad (2.7)$$

从而形成 m 维状态空间, 在重构的 m 维状态空间中可以建立数学模型

$$x_{n+1} = G(x_n) \quad (2.8)$$

F. Takens 和 R. Mane 证明了只要适当选取 m 和 τ , 原未知数学模型的混沌动力系统的几何特征与重构的 m 维状态空间的几何特征是等价的, 它们具有相同的拓扑结构, 这意味着原未知数学模型的混沌动力系统中的任何微分或拓扑不变量可以在重构的状态空间中计算, 并且可以通过在重构的 m 维状态空间中建立数学模型对原未知数学模型的动力系统进行预测, 进一步解释、分析、指导原未知数学模型的动力系统。这一过程可以用图 2.4 表示。

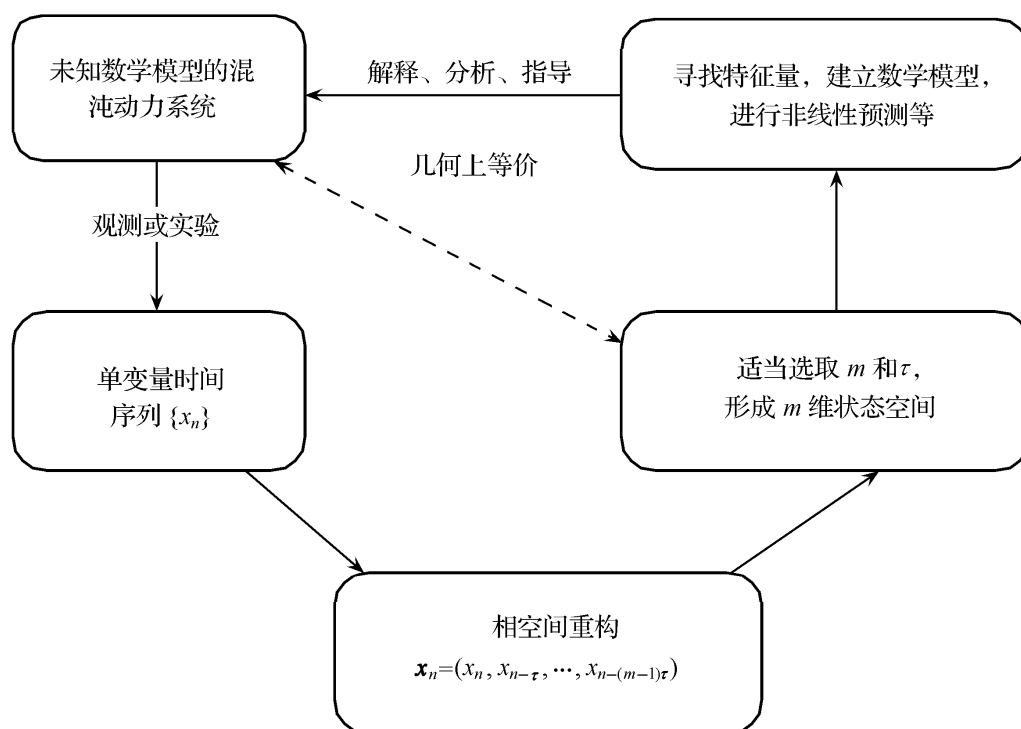


图 2.4 单变量混沌时间序列分析流程图

例 2.1 和例 2.2 中的时间序列就是系统的一个状态变量的观测值, 没有观测噪声, 而且已知原系统是混沌的, 但实际问题中复杂系统通过观测或实验手段获得的时间序列往往没有这么理想, 因此进行时间序列分析时还有许多工作需要做。

对实际问题中未知数学模型或无法建立解析数学模型的复杂系统, 通过观测或实验手段获得的时间序列可能是不平稳的, 可以通过适当的方法对时间序列进行平稳化处理, 因为混沌时间序列分析方法主要是针对平稳时间序列的。同时, 需要对通过观测或实验手段获得的时间序列进行非线性检验, 如果时间序列是来自于线性系统, 则可以通过线性方法对原系统进行建模; 如果时间序列来自于非线性系统, 需要继续检验其是来自于确定性系统还是来自于随机系统, 对随机时间序列

可以通过统计方法研究,对来自确定性系统的时间序列,还需要进一步判断其是否具有混沌特征,因为对混沌时间序列只能进行短期分析,其长期演化行为无法正确描述。另外,通过观测或实验手段获得的时间序列往往具有噪声,在进行分析处理前需要进行降噪处理。在上述分析的基础上,可以对时间序列进行相空间重构,并在重构的相空间中建立数学模型来分析原系统的演化行为。

因此,通过观测或实验手段获得的时间序列的分析流程可以用图 2.5 表示。

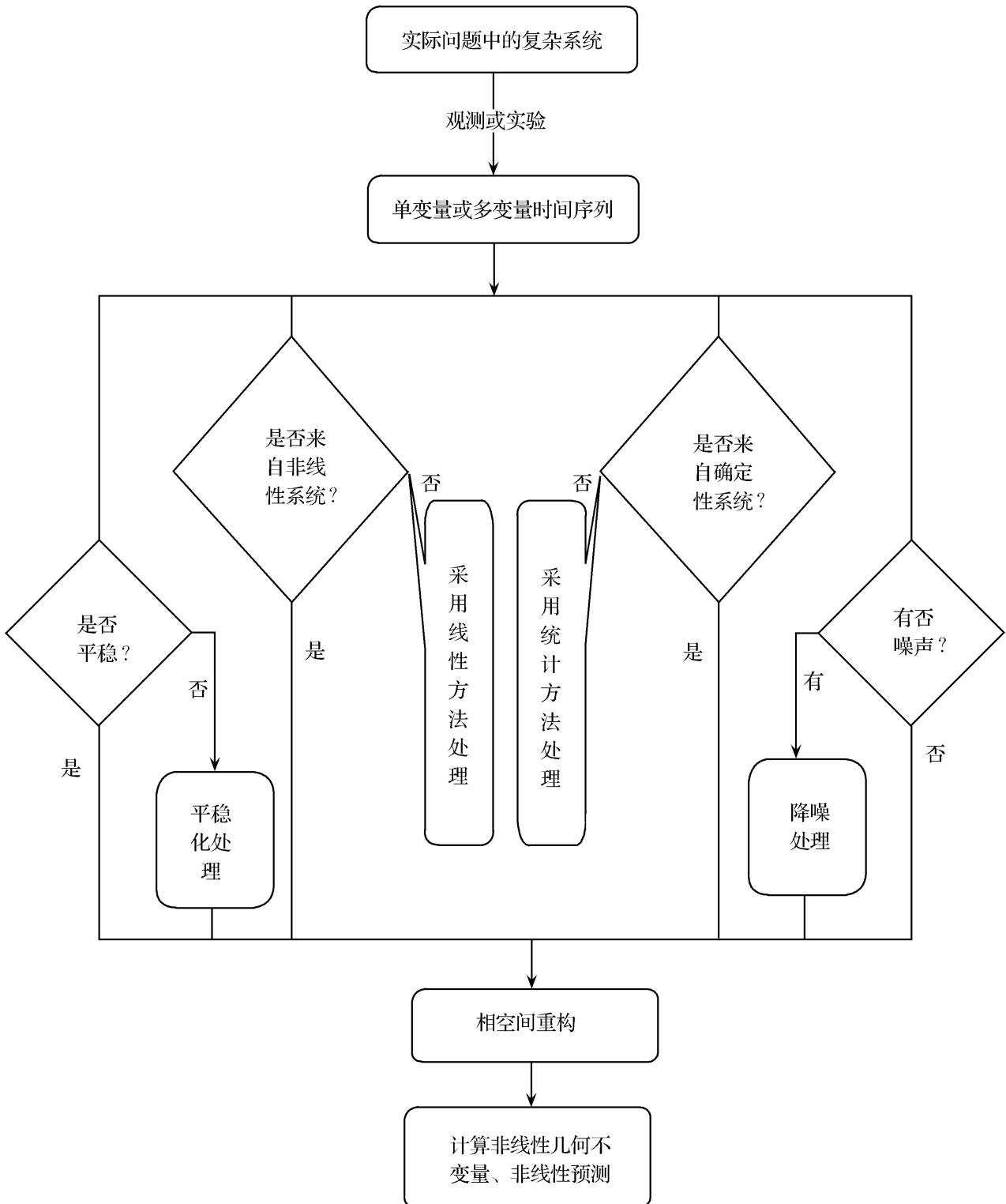


图 2.5 通过观测或实验手段获得时间序列的分析流程图

非线性时间序列分析的一个缺陷是从输出的结果即所研究系统的某些特征本身无法给出系统的实际意义解释,无法从这一过程获得系统的某些方面的改变对系统的影响,而通过建立系统数学模型研究系统行为的正向研究方法恰恰能做到这一点。但这并不妨碍对非线性时间序列分析方法的研究,因为在实际应用中经常是把正向的模型研究与反向的时间序列分析结合起来,根据实际系统的背景知识建立的正向模型可以给出系统的测量时间序列的特征,而通过对观测或实验手段获取的时间序列的分析可以检验和定量分析这些特征,同时非线性时间序列分析可以给出什么样的方程或模型适合所研究系统的相关信息,为建立系统动力学模型提供建议和参考。另外,也可以通过对已建立的系统动力学模型和非线性时间序列的预测正确程度的比较,分析已建立的系统动力学模型的精确性,为进一步修正系统动力学模型提供依据。

2.2 相空间重构

根据观测或实验手段获得的混沌时间序列去研究复杂系统这种方法的基础是相空间重构,它的基本思想是:由于混沌系统产生的轨迹经过一定时期的变化后,会最终做一种有规律的运动,产生一种规则的、有形的轨迹(混沌吸引子)。系统中任一分量的演化都是由与之相互作用的其他分量所决定的,因此这些相关分量的信息就隐含在任一分量的发展过程中,这样就可以从某一分量的时间序列中提取和恢复出动力系统原来的规律。为了重构一个等价的状态空间,只需考察一个分量,并将它在某些固定的时间延迟点上的观测量作为新维处理,即延迟值被看作是新的坐标,它们确定了某个多维状态空间中的一点。重复这一过程并测量相对于不同时间的各延迟量,就可以产生出许多这样的点,已经证明它可以将吸引子的许多性质保存下来,即用系统的一个观测量可以重构出原动力系统的模型,并可初步确定该系统的真实相空间的维数。由此可见,相空间重构就是从时间序列出发创建一个多维状态空间,它保持了原系统的许多几何不变量不变,这些几何不变量包括不动点的特征值、吸引子的分维数和轨线的 Lyapunov 指数等。相空间重构的依据是 F. Takens^[1]和 R. Mane^[2]的延迟嵌入定理。F. Takens 的论文对非数学专业的人来说很难读懂,R. Mane 的论文写得更抽象,但可通过下面的相空间重构过程来理解 F. Takens 和 R. Mane 的延迟嵌入定理。

设动力系统是由非线性差分方程

$$z_{n+1} = F(z_n) \quad (2.9)$$

表示的离散系统,或是由微分方程

$$\frac{dz(t)}{dt} = F(z(t)) \quad (2.10)$$

表示的连续系统。其中 z_n 或 $z(t)$ 是系统在时刻 n 或 t 的状态向量, $F(\cdot)$ 是向量值函数。时间序列 $\{x_n\}$ 是观测到的系统某一维输出, 即

$$x_n = h(z_n) + \omega_n \quad (2.11)$$

或

$$x_n = x(t_0 + n\Delta t) = h[z(t_0 + n\Delta t)] + \omega_n \quad (2.12)$$

式中, $h(\cdot)$ 为多元数量值函数; ω_n 为在观测或者测量过程中由于技术手段不完善或者精度不够引起的测量噪声。

根据 F. Takens 的定理^[1], 当 $\omega_n = 0$ 时, 观察到的时间序列 $\{x_n\}$ 以向量

$$x_n = (x_n, x_{n-\tau}, \dots, x_{n-(m-1)\tau}) \quad (2.13)$$

形成 m 维空间, 只要 $m \geq 2d + 1$, 动力系统的几何结构可以完全打开, 其中 d 是系统中吸引子的维数, τ 是正整数, 称为延迟时间间隔。条件 $m \geq 2d + 1$ 是动力系统重构的充分但不必要条件, 获得动力系统重构的整数 m 叫做嵌入维数。状态空间中 $x_n \rightarrow x_{n+1}$ 的演化反映了未知动力系统 $z_n \rightarrow z_{n+1}$ 或 $z(t) \rightarrow z(t+1)$ 的演化, 并且状态空间 R^m 中吸引子的几何特征与原动力系统的几何特征等价, 这意味着原动力系统中任何微分或拓扑不变量可以在重构的状态空间中计算。F. Takens 的定理是在无噪声的情况下考虑的, 即式(2.11)和式(2.12)中的 $\omega_n = 0$, 后来 T. Sauer 等把延迟嵌入定理推广到了具有噪声的情形^[3]。

对一组长为 N 的实测时间序列 $\{x_n\}_{n=1}^N$, 由式(2.13)可构造出 m 维状态向量

$$x_n = (x_n, x_{n-\tau}, \dots, x_{n-(m-1)\tau}) \in R^m, \quad n = N_0, N_0 + 1, \dots, N \quad (2.14)$$

其中 $N_0 = (m-1)\tau + 1$, τ 是延迟时间间隔。在 R^m 中以 L_2 或 L_∞ 范数定义 x_i 到 x_j 的距离, 即

$$\|x_i - x_j\|_2 = \sqrt{\sum_{l=0}^{m-1} [x_{i+l\tau} - x_{j+l\tau}]^2} \quad (2.15)$$

或

$$\|x_i - x_j\|_\infty = \max_{0 \leq l \leq m-1} |x_{i+l\tau} - x_{j+l\tau}| \quad (2.16)$$

以下在不特别说明的情况下, 记 R^m 中的范数为 $\|\cdot\|$, 但 L_∞ 范数在计算距离时实现较快。

为了能在重构的 R^m 空间中刻画原动力系统的性质, 需正确地确定延迟时间间隔 τ 和嵌入维数 m 。

2.2.1 延迟时间间隔的确定

由 F. Takens 的延迟嵌入定理知, 在时间序列无限长、无噪声、无限精确的情况下, 可以任意选择 τ , 但实测时间序列是有限长的, 且一般都有噪声污染和测量误差, 只能根据经验来选择 τ 。选择 τ 的基本思想是使 x_n 与 $x_{n+\tau}$ 具有某种程度的独

立但又不完全无关,以便它们能在重构的相空间中作为独立的坐标处理。如果 τ 太小,则 x_n 与 $x_{n+\tau}$ 的值充分靠近,以至于不能区分它们,从实际观点看不能提供两个独立的坐标,导致吸引子重构非常靠近相空间中的“对角线”,重构的相空间可能总是杂乱无规则的;如果 τ 太大, x_n 与 $x_{n+\tau}$ 可能会不相关,吸引子轨道会投影在两个完全不相关的方向上,不能反映相空间中轨线的真实演化规则。鉴于此,需要选择一个比较合适的延迟时间间隔 τ 。目前,可以使用的方法非常多,但从计算复杂性和使用的简便性等角度看,比较常用的方法主要有自相关函数法和平均互信息法。

1. 自相关函数法

自相关函数的基本思想是要考察观测量 x 和 $x_{n+\tau}$ 与平均观测量的差之间的线性相关性,即如果假设

$$x_{n+\tau} - \bar{x} = C_L(\tau)(x_n - \bar{x}) \quad (2.17)$$

其中

$$\bar{x} = \frac{1}{N} \sum_{n=1}^N x_n \quad (2.18)$$

则使

$$\sum_{n=1}^N [x_{n+\tau} - \bar{x} - C_L(\tau)(x_n - \bar{x})]^2 \quad (2.19)$$

最小的 $C_L(\tau)$ 为

$$C_L(\tau) = \frac{\frac{1}{N} \sum_{n=1}^N [x_{n+\tau} - \bar{x}][x_n - \bar{x}]}{\frac{1}{N} \sum_{n=1}^N [x_n - \bar{x}]^2} \quad (2.20)$$

称这样的 $C_L(\tau)$ 为线性自相关函数。取 $C_L(\tau)$ 第一次为零时的 τ 为延迟时间间隔,此时在平均意义下, x_n 和 $x_{n+\tau}$ 是线性无关的^[4]。

自相关函数法是比较简单的寻找时间延迟 τ 的一种方法,但这种方法只考虑到时间序列中线性关系,至于非线性关系并不清楚,所以并不适合所有情况。特别当自相关函数变化十分缓慢时,选择 τ 会非常困难。

2. 平均互信息法

不同于自相关函数法,平均互信息法将非线性关系也考虑在内,这种方法的根据是可从事件 b_j 在集 B 中发生的概率中得到关于 a_j 在集 A 发生概率的信息。文献[5, 6]给出了平均互信息的定义以及如何从时间序列计算平均互信息的方法。

设观测时间序列为 $\{x_n\}$,则在 n 和 $n+\tau$ 时刻观测量之间的平均互信息为

$$I(\tau) = \sum_{n=1}^N P[x_n, x_{n+\tau}] \log_2 \left[\frac{P[x_n, x_{n+\tau}]}{P(x_n)P(x_{n+\tau})} \right] \quad (2.21)$$

式中, $P(x_n), P(x_{n+\tau}), P(x_n, x_{n+\tau})$ 为概率。

概率 $P(x_n), P(x_{n+\tau})$ 可以通过计算时间序列的直方图获得, 联合概率 $P(x_n, x_{n+\tau})$ 可以通过计算时间序列的二维直方图获得, 利用计算机可以很方便地计算观测时间序列的平均互信息^[5, 6]。

文献[5]建议选择 $I(\tau)$ 的第一个局部最小时的 τ 为延迟时间间隔, 因为此时产生的冗余最小, 产生了最大的独立性。与自相关函数法相比, 平均互信息法考虑了非线性依赖性, 但仍有其局限性, 如有时可能无局部最小或对某些例子特别不合适^[7, 8]。

除了自相关函数法和平均互信息法外, 选择 τ 的方法还有重构展开法^[8]、高阶关联法^[9]、通过分析整体和局部混沌吸引子行为获得优化延迟时间的填充因子法^[10]等多种方法, 这些方法都有各自的特点, 但实际应用中用得较多的还是自相关函数法和平均互信息法。

2.2.2 嵌入维数的确定

关于嵌入维数, F. Takens^[1]、R. Mane^[2]、T. Sauer^[3]等先后从理论上证明了当 $m \geq 2d + 1$ 时可获得一个吸引子的嵌入, 其中 d 是吸引子的分形维数, 但这只是一个充分条件, 对观测时间序列选择 m 没有帮助。如果仅仅是计算关联维数, M. Z. Ding 等^[11]证明了对无噪声、无限长的时间序列, 只要取 m 为大于关联维数 d 的最小整数即可, 但对长度有限且具噪声的时间序列, m 要比 d 大得多。如果 m 选得太小, 则吸引子可能折叠以致在某些地方自相交, 这样在相交区域的一个小邻域内可能会包含来自吸引子不同部分的点; 如果 m 选得太大, 理论上是可以的, 但在实际应用中, 随着 m 的增加会大大增加吸引子几何不变量(如关联维数、Lyapunov 指数等)的计算工作量, 且噪声和舍入误差的影响会大大增加。

在实际应用中常有以下方法选取嵌入维数 m 。

1. 试算法

试算法是通过逐步增加计算过程中的嵌入维数, 观察什么时候某些几何不变量(例如, 关联积分、关联维数、Lyapunov 指数等)停止变化的方法。从理论上讲, 由于这些几何不变量是吸引子的几何性质, 当 m 大于最小嵌入维数时, 几何结构被完全打开, 因此这些不变量与嵌入维数无关, 取吸引子的几何不变量停止变化时的 m 为最小嵌入维数。这种方法的缺点是对数据要求较高(无噪声)、计算量大且比较主观。例如, P. Grassberger 通过增加嵌入维数 m , 计算关联积分 $C_N(r, m, \tau)$, 取当关联积分 $C_N(r, m, \tau)$ 不再变化时的 m 为嵌入维数^[12]。

2. 虚假邻点法

虚假邻点法建立在以下事实的基础上: 选择太小的嵌入维数将导致那些在原

相空间中离得比较远的点会在重构的相空间中靠近。其基本思想是当嵌入维数从 m 变到 $m + 1$ 时,考察轨线 x_n 的邻点中哪些是真实的邻点,哪些是虚假的邻点,当没有虚假邻点时,可以认为几何结构被完全打开^[13]。

设 x_n 的最近邻点为 $x_{\eta(n)}$,即

$$\begin{aligned} \|x_{\eta(n)} - x_n\|_2^{(m)} &= \min_{j=N_0, \dots, N, j \neq n} \|x_n - x_j\|_2 \\ &= \sqrt{\sum_{l=1}^{m-1} [x_{\eta(n)+l\tau} - x_{n+l\tau}]^2} \end{aligned} \quad (2.22)$$

当嵌入维数从 m 增加到 $m + 1$ 时,它们之间的距离变为

$$\begin{aligned} \|x_{\eta(n)} - x_n\|_2^{(m+1)} &= \sqrt{\sum_{l=0}^m [x_{\eta(n)+l\tau} - x_{n+l\tau}]^2} \\ &= \sqrt{[\|x_{\eta(n)} - x_n\|_2^{(m)}]^2 + [x_{\eta(n)+m\tau} - x_{n+m\tau}]^2} \end{aligned} \quad (2.23)$$

若 $\|x_{\eta(n)} - x_n\|_2^{(m+1)}$ 比 $\|x_{\eta(n)} - x_n\|_2^{(m)}$ 大很多,可以认为是由于高维吸引子中两个不相邻的点在投影到低维轨线上时变成相邻的两点造成的,因此这样的邻点是虚假的。具体地说,当

$$\sqrt{\frac{[\|x_{\eta(n)} - x_n\|_2^{(m+1)}]^2 - [\|x_{\eta(n)} - x_n\|_2^{(m)}]^2}{[\|x_{\eta(n)} - x_n\|_2^{(m+1)}]^2}} = \frac{|x_{\eta(n)+m\tau} - x_{n+m\tau}|}{\|x_{\eta(n)} - x_n\|_2^{(m+1)}} \geq R_{tol} \quad (2.24)$$

时, $x_{\eta(n)}$ 为 x_n 的虚假最近邻点,其中 R_{tol} 是某一个阈值,文献[13]建议 $R_{tol} \geq 10$,而文献[14]建议 R_{tol} 可取 15 左右。

观测时间序列通常具有噪声且长度有限,所以仅仅用上面的标准判别虚假最近邻点会不正确,为此, M. B. Kennel 等提出了增加以下标准,即当

$$\frac{\|x_{\eta(n)} - x_n\|_2^{(m+1)}}{\sqrt{\frac{1}{N} \sum_{k=1}^N \left[x_k - \frac{1}{N} \sum_{k=1}^N x_k \right]^2}} > A_{tol} \quad (2.25)$$

时, $x_{\eta(n)}$ 为 x_n 的虚假最近邻点,其中 A_{tol} 是阈值,文献[13]建议 $A_{tol} = 2$ 。

对式(2.14)中的每个 x_n ,按式(2.22)找出其最近邻点,并按照式(2.24)、式(2.25)的标准判别最近邻点是否是虚假的,从而可以计算出虚假最近邻点的比例。让 m 从 1 开始增加,计算每个 m 时的虚假最近邻点的比例,直到虚假最近邻点的比例小于 5% 或虚假最近邻点不再随着 m 的增加而减少时,可以认为吸引子几何结构完全打开,此时的 m 为嵌入维数。

从几何观点看,这是一种较好的方法,但在判断虚假邻点时阈值的不同选取会导致不同的结果,因此具有较大的主观性。从式(2.24)看,阈值应当由时间序列的方差确定,因此,不同的时间序列有不同的阈值,这意味着要给出一个合适和合理

的阈值是非常困难的甚至是不可能的,为此,文献[15]提出了改进的虚假邻点算法。

3. 改进的虚假邻点法

对式(2.14)中的 x_n , 类似于虚假最近邻点法的思想, 文献[15]定义

$$a(n, m) = \frac{\|x_{\eta(n)} - x_n\|_{\infty}^{(m+1)}}{\|x_{\eta(n)} - x_n\|_{\infty}^{(m)}} \quad (2.26)$$

这里用的是 L_{∞} 范数, 记所有 $a(n, m)$ 关于 n 的均值为

$$E(m) = \frac{1}{N - N_0 + 1} \sum_{n=N_0}^N a(n, m) \quad (2.27)$$

由于式(2.14)的相空间重构方式与文献[15]中有所差别, 式(2.14)是向后重构, 而文献[15]中的重构是向前重构, 所以式(2.27)与文献[15]有所不同, 但本质上是一样的。

$E(m)$ 只依赖于嵌入维数 m 和延迟时间间隔 τ , 为了研究嵌入维数从 m 变为 $m+1$ 时相空间的变化情况, 定义

$$E_1(m) = \frac{E(m+1)}{E(m)} \quad (2.28)$$

如果当 m 大于某个 m_0 时, $E_1(m)$ 停止变化, 则 m_0+1 就是重构相空间的最小嵌入维数^[15]。

对来自于随机系统的时间序列, 原则上, 随着 m 的增加, $E_1(m)$ 将永远不会达到饱和值, 但在实际计算中, 当 m 充分大时, 很难分辨 $E_1(m)$ 是在缓慢增加还是停止变化。事实上, 由于获得的观测时间序列是有限长的, 可能会出现虽然是随机时间序列, 但 $E_1(m)$ 会在某一 m 处停止变化。为了解决这一问题, 文献[15]定义了量

$$E^*(m) = \frac{1}{N - N_0 + 1} \sum_{n=N_0}^N |x_{\eta(n)-m\tau} - x_{n-m\tau}| \quad (2.29)$$

记

$$E_2(m) = \frac{E^*(m+1)}{E^*(m)} \quad (2.30)$$

对随机时间序列, 由于将来的值与过去的值独立, 对任何 m , $E_2(m)$ 将等于 1, 而对确定性系统的观测时间序列, $E_2(m)$ 显然与 m 有关, 对所有 m , $E_2(m)$ 不可能为常数。因此, 文献[15]建议可以通过同时计算 $E_1(m)$ 和 $E_2(m)$ 来确定时间序列相空间重构的最小嵌入维数, 同时也区分了时间序列是来自于确定性系统还是随机系统。

除了上述方法外, 确定嵌入维数的方法还有关联积分法^[16]、奇异值分解法^[17]以及上述各种的改进方法。

2.3 几何不变量的计算

对实际问题中具有混沌特征的复杂系统,通常可以通过一些几何不变量(例如,分形维数、Kolmogorov 熵和 Lyapunov 指数等)来定量刻画该系统的复杂程度,因此如何从观测或实验手段获得的时间序列去计算复杂系统的这些几何不变量是实际应用时所关心的问题。复杂系统的吸引子具有分数维数、正的 Kolmogorov 熵以及正的最大 Lyapunov 指数是判断系统是混沌的必要条件,但并不充分。分形维数的定义有很多种,从时间序列的角度看,关联维数是易于计算的一种分形维数。1983 年, P. Grassberger 和 I. Procaccia 提出了从时间序列中计算关联维数的 G-P 算法^[12, 18]以及计算 Kolmogorov 熵的公式^[19]。关联维数是系统复杂性程度的一种很好的度量。一般认为,大于关联维数的下一个整数是刻画系统所需的独立变量的个数,这为从时间序列恢复原复杂系统确定了一个框架。Lyapunov 指数度量了复杂系统的预测性,定量地刻画了初始靠近的状态空间轨线的指数发散率,从时间序列中计算最大 Lyapunov 指数的最早工作是在 1985 年由 A. Wolf 等提出的轨线法^[20]。

2.3.1 关联维数

相空间的最小嵌入维数 m 是系统复杂性的一个界,这种拓扑维数在一般情况下只是吸引子占据空间数量的一个非常粗糙的描述,为了给出更严格的描述,文献 [21~ 23] 引入了几种不同的维数概念,其中关联维数是最常用的一种维数度量,尤其对通过观测或实验获得的时间序列,可由 P. Grassberger 和 I. Procaccia 给出的 G-P 算法直接计算^[12, 18]。关联维数度量了所研究系统的复杂性,刻画了相空间中点的分布。在确定性系统中,关联维数就是生成相应复杂系统所必需的独立变量的个数,规则的确定性系统有整数关联维数,而混沌系统有非整数关联维数,大于此关联维数的下一个整数就是系统的独立变量的个数。但是某些关联的随机过程中也有非整数维数^[24],文献 [25] 证明了非整数关联维数不能成为混沌判断的充分条件,因为分形的 Brownian 运动,虽然不是混沌的,但也有非整数关联维数。

G-P 算法描述如下:在按式 (2.14) 重构的 m 维空间中, $\sum_{j \neq i} H(r - \|x_i - x_j\|)$ 表示除 x_i 本身外到 x_i 的距离小于 r 的 x_j 的点数,其中 $H(\cdot)$ 是 Heavside 函数,即当 $x \leq 0$ 时, $H(x) = 0$, 当 $x > 0$ 时, $H(x) = 1$ 。定义按上述方式重构的时间序列 $\{x_n\}_{n=1}^N$ 的关联积分为

$$C_N(r, \tau, m) = \frac{2}{(N - N_0)(N - N_0 + 1)} \sum_{i=N_0}^{N-1} \sum_{j=i+1}^N H(r - \|x_i - x_j\|) \quad (2.31)$$

$C_N(r, \tau, m)$ 描述了距离小于 r 的对点数的分布情况。如果在 r 的某一区间段内, 有

$$C_N(r, \tau, m) \propto r^d \quad (2.32)$$

则称 d 是关联维数, 这样定义的 d 就是近似刻画产生时间序列 $\{x_n\}_{n=1}^N$ 的复杂系统复杂程度的某种维数。对由观测或实验手段获得的时间序列, 使式(2.32)成立的 r 的某一区间段可通过对双对数 $\ln C_N(r, \tau, m) - \ln r$ 图观测的办法或通过文献[26]给出的方法寻找。

G-P 算法是计算关联维数的一种经典方法, 后来又有不少文献讨论了 G-P 算法的改进和其他计算关联维数的方法^[27~34]。

由观测或实验手段获得的时间序列对应混沌系统的关联维数就是双对数 $\ln C_N(r, \tau, m) - \ln r$ 图的线性部分的斜率, 由于连续采样点之间的相关性, 会造成“肩峰”效应, 为了消除这种“肩峰”效应, 文献[27]对关联积分作了以下修改, 即定义

$$C_N(r, \tau, m, w) = \frac{2}{(N - N_0 - w + 1)(N - N_0 - w + 2)} \sum_{i=N_0}^{N-w} \sum_{j=i+w}^N H[r - \|x_i - x_j\|] \quad (2.33)$$

式(2.33)中忽略了嵌入空间中非常靠近的对点关联积分的贡献, 只考虑满足 $|i - j| \geq w$ 的点, 其中 $w > \tau(2/N)^{2/m}$, 特别取 $w = \tau$ 即可。文献[27]指出 w 的严格选取并不重要。

在实际应用中, 关联维数的计算是非常费时的, 为了改善前面算法的计算效率, 文献[35]提出了一种修改算法。在关联积分中, 较短的距离起着更有意义的作用, 因此选取 r 的截断距离 r_0 , 构造宽度为 r_0 的盒子网格, 只计算在相同的或相邻的盒子中距离小于 r_0 的对点的距离, 这样可以把计算时间从 $O(N^2)$ 减到 $O(N \log_2 N)$ 。文献[35]给出了 r_0 的一个优化选择

$$r_0 = R(2/N)^{1/d} \quad (2.34)$$

式中, R 为时间序列的有效半径; N 为时间序列的长度; d 为关联维数的粗略估计。

当关联维数的值较大时, 所需的时间序列要求较长^[36], 而实际问题中, 观测或实验获得的时间序列一般都比较短, 为解决这一问题, 文献[30]利用极大似然法估计关联维数的方法就有对数据要求相对较少的优点, 其计算公式为

$$d = \left[\frac{\sum_{i>j} H[r_0 - \|x_i - x_j\|] \ln[r_0 / \|x_i - x_j\|]}{\sum_{i>j} H[r_0 - \|x_i - x_j\|]} \right]^{-1} \quad (2.35)$$

用式(2.35)估计 d 的近似方差为

$$\text{Var}(d) = \frac{2d^2}{p(N - N_0)(N - N_0 + 1)} [2q(N - 2) + 1] \quad (2.36)$$