

Nonlinear Time
Series Analysis and
its
Application



非线性时间序列 分析及其应用

王海燕 卢山/著



科学出版社

www.sciencep.com

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>

双青印刷厂印刷

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<新欣>)

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

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

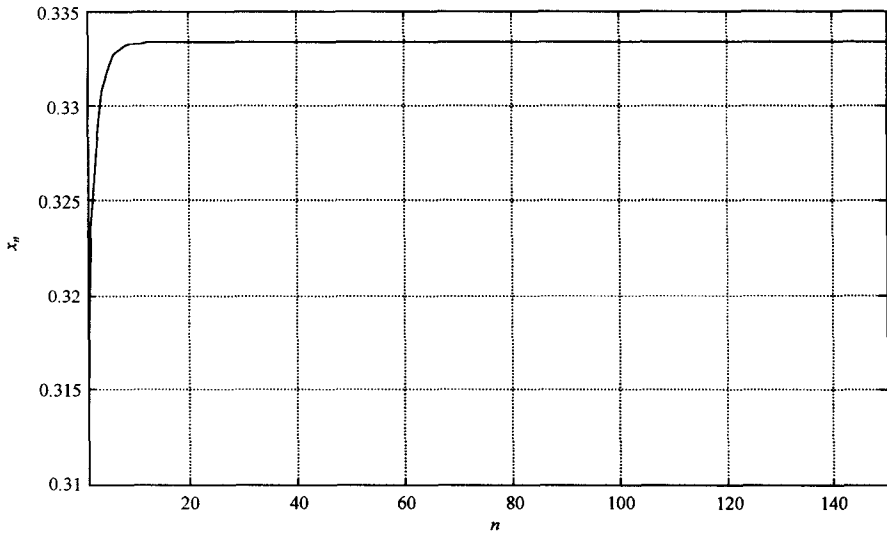
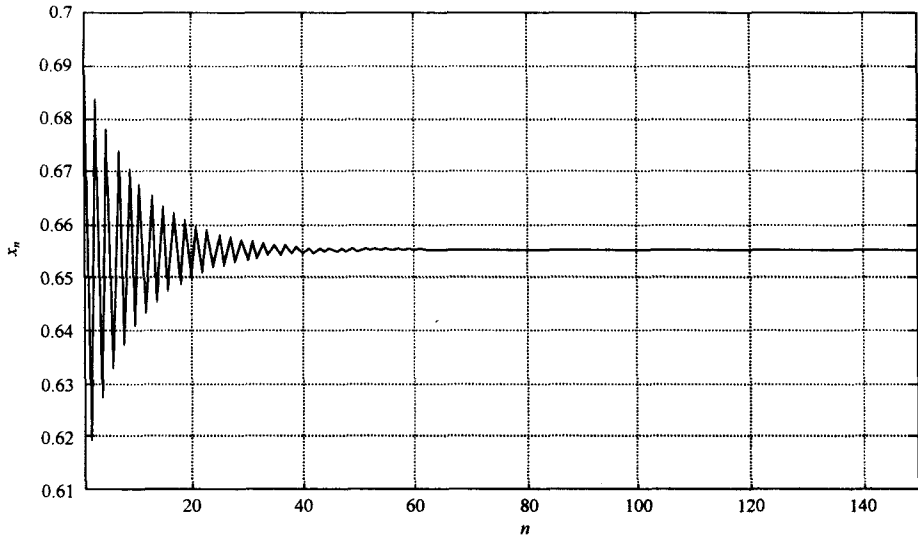
当 $R=1.5$ 时,不管初始状态 x_0 在何处,随时间的演化,系统都将单调地趋向于 $\frac{1}{3}$,如图 1.1 所示($x_0=0.7$)。

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

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

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 $\omega_0=0$,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

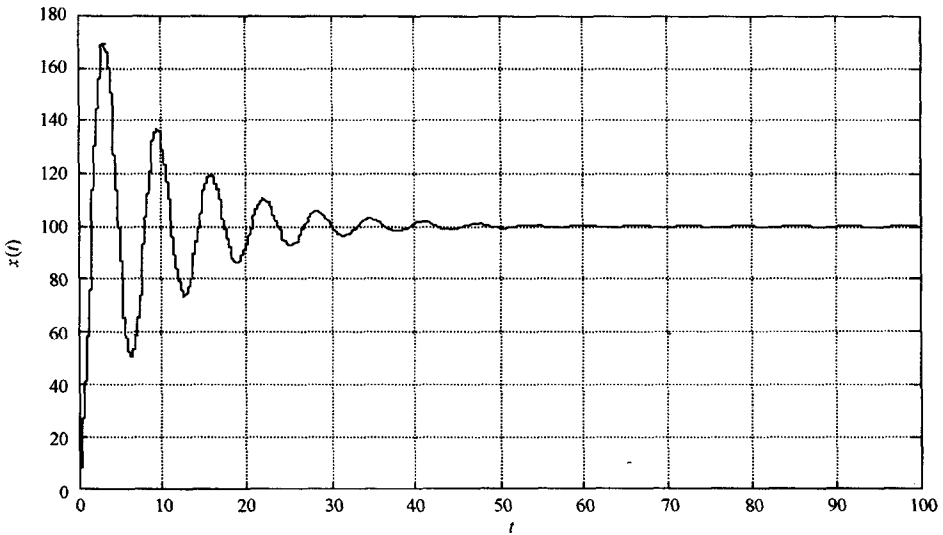


图 1.3 具有阻尼的振动方程的时间历程图

2. 趋向于周期状态的时间序列

例 1.3 考虑以下 Logistic 模型

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

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 ($x_0=0.7$)。

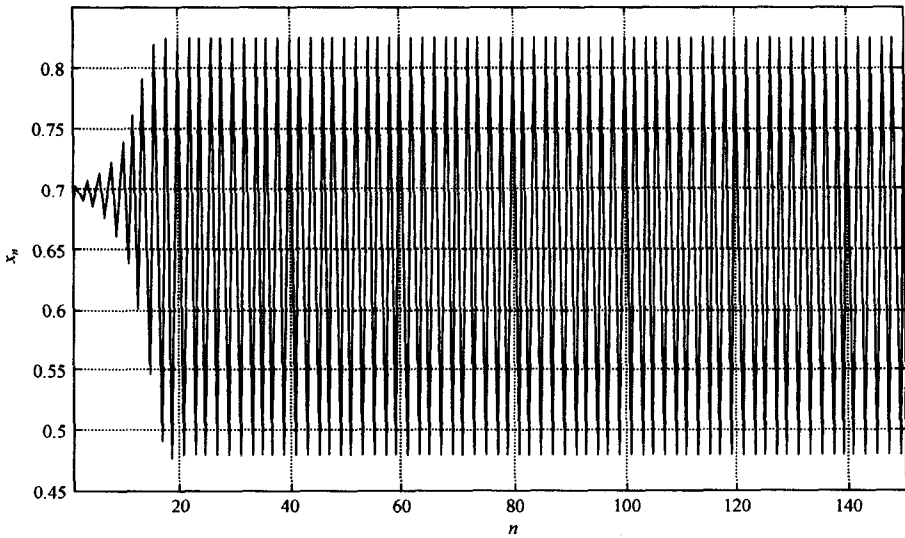


图 1.4 Logistic 模型中 $R=3.3$ 时的时间历程图

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

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

取物体所受的最大策动力 $f_0=100$,策动力的频率 $a=1$,振动物体的固有频率 $\omega_0=1$,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。

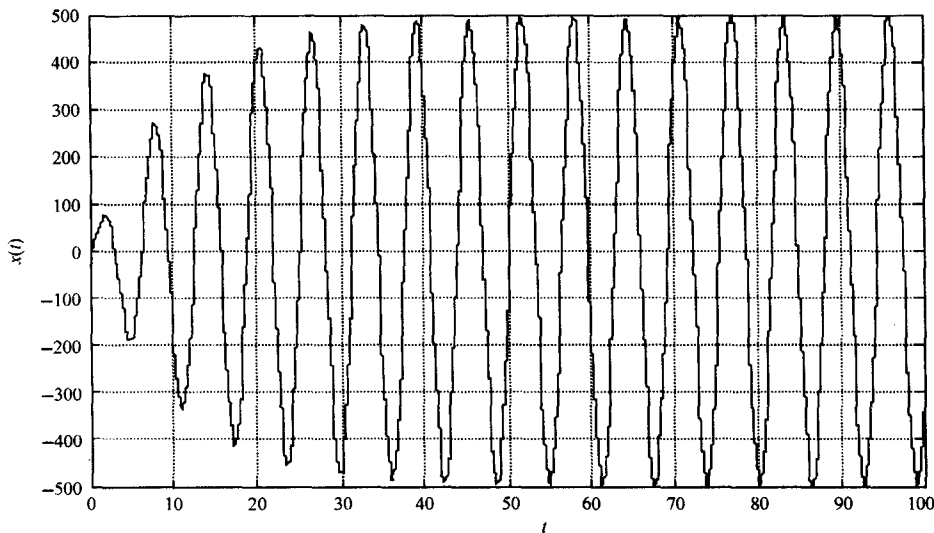


图 1.5 无阻尼的振动方程的时间历程图

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。

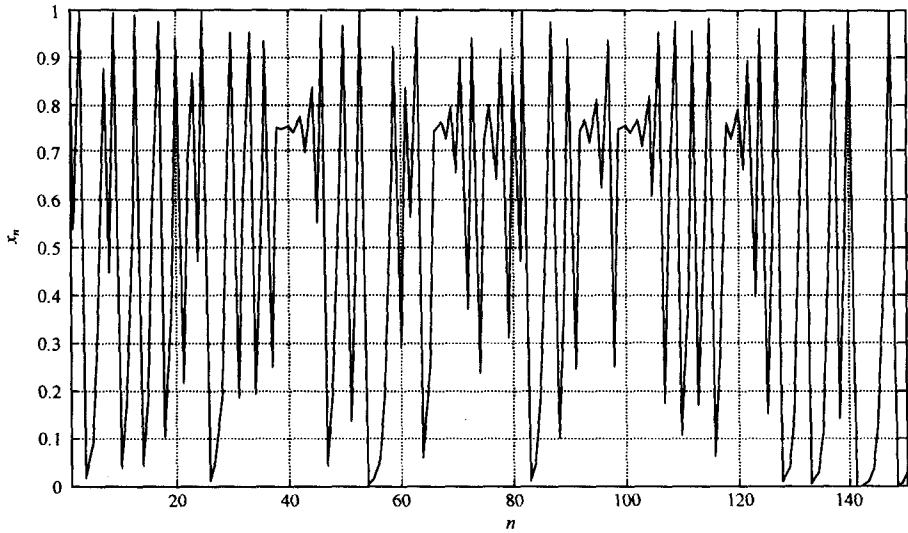


图 1.6 Logistic 模型中 $R=4$ 时的时间历程图

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	$x_n^{(1)}$	$x_n^{(2)}$	$x_n^{(3)}$
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。

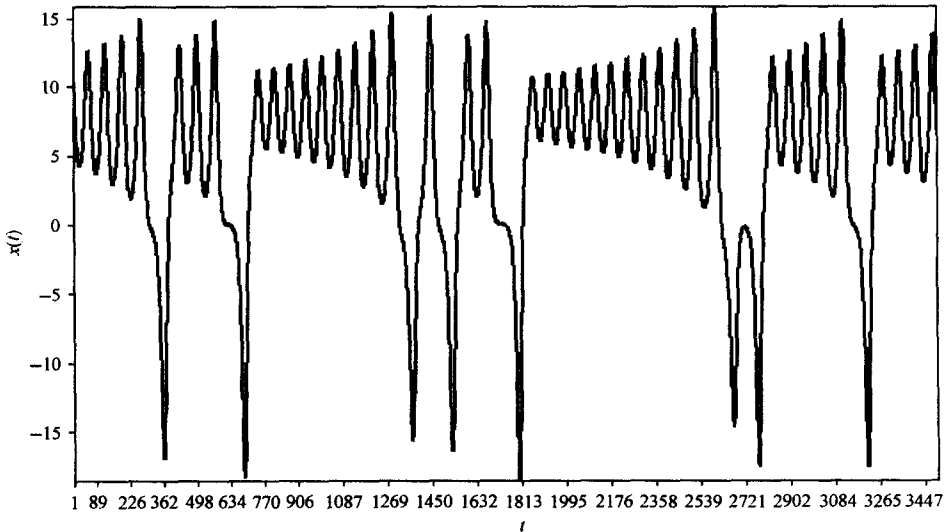


图 1.7 Lorenz 系统变量 x 的时间历程图

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。

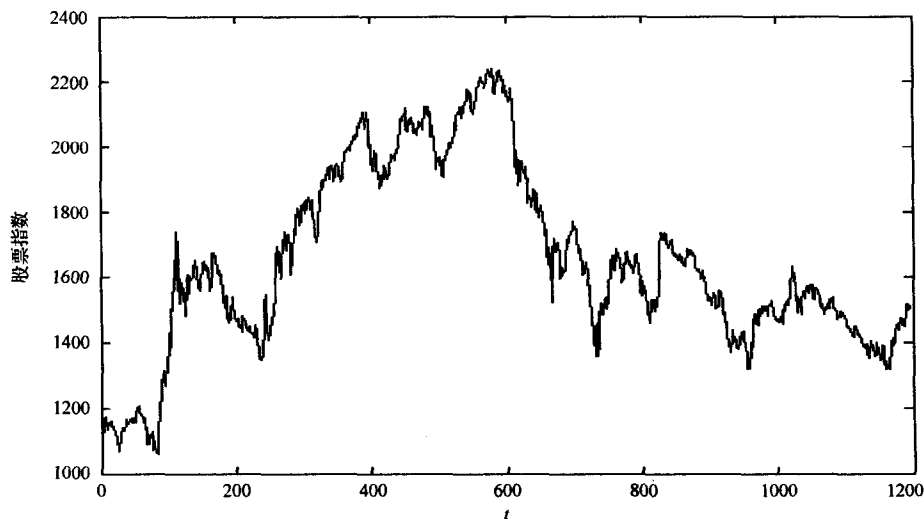


图 1.8 上海证券交易所综合股指时间序列历程图

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。

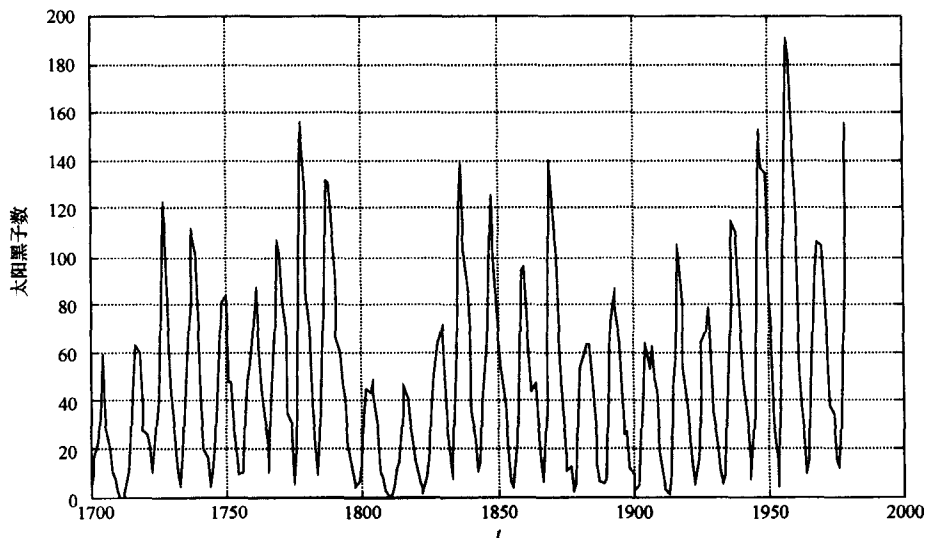


图 1.9 太阳黑子数时间序列历程图