

Applied Time Series Analysis with R

应用时间序列分析

R软件陪同

- 具有浓郁统计学味道的时间序列分析教材，没有复杂的数学证明过程。
- 通过案例阐述相关的概念和方法，涉及各领域真实数据的分析，是系统学习时间序列分析方法及其应用的首选之作。

吴喜之 刘苗 编著



机械工业出版社
China Machine Press

· 华章应用统计系列 ·

应用时间序列分析

R软件陪同

Applied Time Series Analysis with R

吴喜之 刘苗 编著



机械工业出版社
China Machine Press

图书在版编目 (CIP) 数据

应用时间序列分析: R 软件陪同 / 吴喜之, 刘苗编著. —北京: 机械工业出版社, 2014.7
(华章应用统计系列)

ISBN 978-7-111-46816-5

I. 应… II. ①吴… ②刘… III. 时间序列分析 – 高等学校 – 教材 IV. O211.61

中国版本图书馆 CIP 数据核字 (2014) 第 108158 号

本书通过案例讲述有关的概念和方法, 不仅介绍了 ARMA 模型、状态空间模型、Kalman 滤波、单位根检验和 GARCH 模型等一元时间序列方法, 还介绍了很多最新的多元时间序列方法, 如线性协整、门限协整、VAR 模型、Granger 因果检验、神经网络模型、可加 AR 模型和谱估计等. 书中强调对真实的时间序列数据进行分析, 全程使用 R 软件分析了各个科学领域的实际数据, 还分析了金融和经济数据的例子.

本书通俗易懂, 理论与应用并重, 可作为高等院校统计学和经济管理等专业“时间序列分析”相关课程的教材, 对金融和互联网等领域的相关从业者也极具参考价值.



出版发行: 机械工业出版社 (北京市西城区百万庄大街 22 号 邮政编码: 100037)

责任编辑: 明永玲

印 刷: 北京市荣盛彩色印刷有限公司

版 次: 2014 年 7 月第 1 版第 1 次印刷

开 本: 186mm × 240mm 1/16

印 张: 11.75

书 号: ISBN 978-7-111-46816-5

定 价: 39.00 元

凡购本书, 如有缺页、倒页、脱页, 由本社发行部调换

客服热线: (010) 88378991 88361066

投稿热线: (010) 88379604

购书热线: (010) 68326294 88379649 68995259

读者信箱: hzjsj@hzbook.com

版权所有·侵权必究

封底无防伪标均为盗版

本书法律顾问: 北京大成律师事务所 韩光 / 邹晓东

前言

同类教材的特点

首先,一些教材偏重于数学理论和推导.作者多为数学出身,他们习惯于数学的严格性和导出精确而又漂亮的数学结论.这些书适用于那些愿意为时间序列的数学理论研究做出贡献的读者.

其次,国内教材中一元时间序列往往占绝大部分篇幅,而且包含在各种数学假定下的各种定理和结果.这是因为一元时间序列的数学描述确实很漂亮,很多结果都能够以比较简洁的数学语言表达出来.而多元时间序列则很不一样,在一元情况下很漂亮的结果,在多元情况下就完全不同了.在数学上,复杂的表达是不被人们所喜爱的,因此,多元时间序列很难在数学味道很浓的教科书中展开.

很多教材对于真实时间序列的数据分析强调得不够.那些数学味道强的书,由于其主要目的不是分析实际数据,而且实际数据往往很难满足书中的数学假定,过多地讨论实际应用并不是这些书的重点.

另外有一些教材的确强调应用,而且作者很多也不是数学出身,书中也列举了一些数学假定和结论,但是往往没有花篇幅去完善和系统化,更没有用简明扼要的语言去做解释,使得无论是数学还是非数学出身的读者都不能很好地理解所用模型背后的机理.

在涉及统计软件使用方面,数学味道强的书完全不用任何软件是可以理解的,但很多偏重于应用的教科书只介绍昂贵的“傻瓜式”商业软件就不值得提倡了,因为介绍昂贵商业软件的教材客观上鼓励了使用盗版软件.商业软件不透明,代码保密,而且没有体现最新的成果,完全不能满足实际工作者的需要.

本书的特点

本书的目标读者是非数学专业出身的各类人员,可以是本科生或者研究生,也可以是在校教师或者实际工作者.我们力图用简单通俗的语言阐述有关的基本概念和计算,并尽量通过案例来讲述各种时间序列方法,使得非数学背景的读者可以较容易地理解.同时,我们也把有关的数学结构用简单完整的方式阐述,以供读者参考.

本书以应用为导向,除了介绍常用的一元时间序列方法,也尽可能多地介绍最新的多元时间序列方法.书中不仅包括各个科学领域的实际数据,还包括金融和经济数据的例子.本书例题用到的实际数据都可以从网上下载,或者在华章公司的网站(www.hzbook.com)获取.

本书全程使用免费、公开、透明的开源编程软件R^①, 而且提供全部代码. R软件是世界上使用者最多的数据分析软件, 有着非常强大的统计界的支持, 发展很快, 每天都有许多新的程序包被加入, 到2014年5月初, R的统计程序包数量已经达到5500多个, 而2009年年底只有不到1000个. 新的统计方法大都以R程序包的形式首先展现在世人面前, 这是任何商业软件所望尘莫及的. 我们希望读者能够尽快地学会如何使用R软件解决读者自己的实际问题或分析数据.

欢迎读者提出宝贵意见, 以使得我们在本书再版时予以改进.

作者
2014年5月

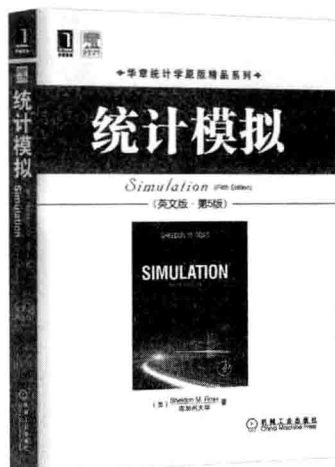
① R的第二版和第三版的授权在网站<http://www.R-project.org/Licenses/>中.

推荐阅读



随机过程 (原书第2版)

作者: (美) Sheldon M. Ross 译者: 龚光鲁 ISBN: 978-7-111-43029-2 定价: 79.00元



统计模拟 (英文版·第5版)

作者: (美) Sheldon M. Ross ISBN: 978-7-111-42045-3 定价: 59.00元



例解回归分析 (原书第5版)

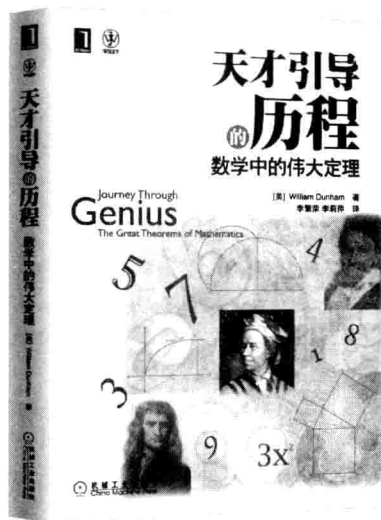
作者: (美) Samprit Chatterjee 等 译者: 郑忠国 等 ISBN: 978-7-111-43156-5 定价: 69.00元



金融数据分析导论: 基于R语言

作者: (美) Ruey S. Tsay 译者: 李洪成 等 ISBN: 978-7-111-43506-8 定价: 69.00元

推荐阅读



天才引导的历程：数学中的伟大定理

作者：William Dunham ISBN: 978-7-111-40329-6 定价：45.00元

“Dunham的这本书如此特别，是我以前从未遇到过的……娓娓道来的一个个推理精巧与颇具洞察力的个案，引人入胜。”

——当代美国最著名的科普作家、科幻小说家、文学评论家Isaac Asimov

“把Dunham的这本书推荐给所有热爱探索、思想活跃的人们，不管他们感兴趣的是艺术还是科学，阅读本书都是一次重要的文化体验。”

——英国著名数学教育家暨《上帝掷骰子吗？》的作者Ian Stewart

“如果有一本书让你读了会对数学产生毕生爱好的话，那肯定就是Dunham的这本书。”

——Amazon读者评论

“这门几乎每个人都觉得沉闷、无聊、呆板的学科，在Dunham的笔下充满生机与活力……我是拥有计算机科学学位的外行，但是我喜欢这本书……Dunham巧妙地将数学中的伟大定理编织成数学史，使得本书容易理解，而且我敢说，事实上很有趣！本书是一颗珍宝，每一个爱好数学的人都不能与它失之交臂。”

——Amazon读者评论

这是二十多年来一直畅销不衰的数学大众读物！长期雄踞亚马逊数学大众读物畅销榜榜首！

本书将两千多年的数学发展历程融为十二章内容，每章都包含了三个基本组成部分，即历史背景、人物传记以及在这些“数学杰作”中所表现出的创造性。作者精心挑选了一些杰出的数学家及其所创造的伟大定理，如欧几里得、阿基米德、牛顿和欧拉。这一个个伟大的定理，不仅串起了历史的年轮，更是串起了数学这门学科所涵盖的各个深邃而不乏实用性的领域。当然，这不是一本典型的数学教材，而是一本大众读物，它会让热爱数学的人体会到绝处逢生的喜悦，让讨厌数学的人从此爱上数学。

目 录

前言

第1章 引言	1
1.1 时间序列的特点	1
1.2 时间序列例子	1
1.3 R软件入门	5
1.3.1 简介	5
1.3.2 动手	7
1.4 本书的内容	7
1.5 习题	8
第2章 一元时间序列的基本概念和模型	10
2.1 时间序列的平稳性及相关性度量	10
2.1.1 平稳、自协方差函数和自相关函数	10
2.1.2 差分算子和后移算子	12
2.2 白噪声	13
2.3 随机游走	14
2.4 趋势平稳过程	15
2.5 一般线性模型	15
2.6 MA模型	16
2.7 AR模型	19
2.8 ARMA模型	23
2.9 ARIMA模型	24
2.10 季节模型	26
2.11 习题	26

第3章 一元时间序列数据的拟合及预测	29
3.1 一些估计和预测方法的基本数学原理	29
3.1.1 ARMA模型的最大似然估计	29
3.1.2 ARMA模型的矩估计方法	30
3.1.3 预测的基本目的	30
3.1.4 简单指数平滑	30
3.1.5 Holt-Winters滤波预测方法	32
3.1.6 ARMA模型预测的基本数学原理	33
3.2 一元时间序列数据实例分析	36
3.2.1 差分、平滑和时间序列的分解	36
3.2.2 ARMA模型和ARIMA模型	49
3.2.3 例1.2中Auckland降水序列的综合分析	60
3.3 习题	67
第4章 状态空间模型和Kalman滤波简介*	69
4.1 动机	69
4.2 结构时间序列模型	69
4.2.1 局部水平模型	70
4.2.2 局部线性趋势模型	70
4.2.3 季节效应	70
4.3 一般状态空间模型	71
4.3.1 随时间变化系数的回归	71
4.3.2 ARMA模型的状态空间模型形式	71
4.3.3 结构时间序列的一般状态空间模型表示	74
4.4 Kalman滤波	75
4.5 状态空间数据例子	76
4.5.1 一元局部水平模型例子	76
4.5.2 二元局部水平模型Kalman滤波例子	78
4.5.3 包含季节因素的局部水平多元模型Kalman 滤波例子	80

第5章 单位根检验	83
5.1 单整和单位根	83
5.2 单位根检验	84
5.2.1 DF检验、ADF检验以及PP检验	85
5.2.2 KPSS检验	88
第6章 长期记忆过程: ARFIMA模型*	90
6.1 介于 $I(0)$ 及 $I(1)$ 之间的长期记忆序列	90
6.2 ARFIMA过程	91
6.3 ARFIMA模型拟合例3.4尼罗河流量数据	93
第7章 GARCH模型	95
7.1 时间序列的波动	95
7.2 模型的描述	98
7.3 数据的拟合	100
7.3.1 例7.1数据的拟合	100
7.3.2 例7.2数据的拟合	102
7.4 GARCH模型的延伸	104
7.4.1 一组GARCH模型	104
7.4.2 FGARCH模型族	106
7.4.3 ARFIMA-GARCH模型族拟合例7.2数据	107
第8章 多元时间序列的基本概念和模型	112
8.1 平稳性	112
8.2 交叉协方差矩阵和相关矩阵	112
8.3 一般线性模型	113
8.4 VARMA模型	113
8.5 协整模型和Granger因果检验	114
8.5.1 VECM和协整	114
8.5.2 协整检验	115

8.5.3	Granger因果检验	117
第9章	多元时间序列数据的拟合及预测	119
9.1	例9.1数据的协整检验和Granger因果检验	120
9.1.1	Johansen协整检验	120
9.1.2	Engle-Granger协整检验	120
9.1.3	Pillips-Ouliaris协整检验	121
9.1.4	例9.1数据的Granger因果检验	122
9.2	用VAR、VARX及状态空间模型拟合例9.1数据	122
9.2.1	用VAR拟合及预测例9.1数据	122
9.2.2	用VARX模型拟合及预测例9.1数据	125
9.2.3	用状态空间模型拟合及预测例9.1 数据	127
9.3	习题	129
第10章	非线性时间序列	130
10.1	非线性例子	130
10.2	线性AR模型	133
10.3	自门限自回归模型	133
10.3.1	一个门限参数的模型	134
10.3.2	两个门限参数的模型	135
10.3.3	Hansen检验	137
10.4	Logistic 平滑过渡自回归模型	138
10.5	神经网络模型	140
10.6	可加AR模型	141
10.7	模型的比较	142
10.8	门限协整	142
10.8.1	向量误差修正模型	142
10.8.2	向量误差修正模型的估计	143
10.8.3	向量误差修正模型的检验	145

第11章 谱分析简介*	147
11.1 周期性时间序列	147
11.2 谱密度	150
11.3 谱分布函数	151
11.4 自相关母函数和谱密度	152
11.5 时不变线性滤波器	155
11.6 谱估计	158
11.6.1 通过样本自协方差函数估计谱密度	158
11.6.2 通过周期图估计谱密度	158
11.6.3 非参数谱密度估计	160
11.6.4 参数谱密度估计	162
附录 使用R软件练习	164
参考文献	172

第1章 引言

1.1 时间序列的特点

本书研究的时间序列主要是根据时间先后对同样的对象按等间隔时间收集的数据,比如每日的平均气温、每天的销售额、每月的降水量、每秒钟容器的压力等等。虽然这些序列所描述的对象本身(比如容器压力)可能是连续的,但由于观测值并不连续,这种时间序列数据称为离散(discrete)时间序列。当然也有连续(continuous)时间序列,它是不间断地取值得到的时间序列,比如地震波和一些状态控制(比如月球车是否苏醒的用1-0表示的状态)的连续观测值,但对象不一定是连续的(比如0-1状态)。和用于简单最小二乘回归的横截面数据不同,时间序列的观测值有可能是相关的,比如今天的物价和昨天的物价相关,本月的CPI和上个月的CPI相关等等。

因为待研究的时间序列变量个数的不同,时间序列分为一元时间序列和多元时间序列。在使用时间序列进行预测时往往使用一个(一元)或一组(多元)时间序列的历史数据来预测其未来的值,这在一元时间序列的应用中尤其明显。在多元时间序列中也有用其他时间序列作为输入来预测另一些时间序列的情况。这些特点使得时间序列更注重序列本身的自相关,并且利用过去时间序列的模式来预测未来。因此,研究时间序列本身的性质就很重要了。对于没有规律的时间序列,或者受到其他未知的或不可观测变量影响的时间序列,人们基本上无能为力。而且,为了有足够的关于序列模式的信息,时间序列分析通常需要较多的历史数据。只有那些比较“规范”的而且信息量较大的时间序列才有可能被建模并且用于预测。此外,一些时间序列模型有很强的数学假定,但这些最多根据经验的假定永远无法用数据验证。如果不注意到这些局限性,则会导致滥用和误导。

什么是较“规范”的序列呢?比方说,对于一元时间序列来说,就是在进行一系列差分变换之后可以转变为“平稳”的序列,“平稳”是时间序列分析中的一个重要概念。一般地,平稳的序列可以用可操作的数学模型来描述和处理;而对于多元时间序列,则希望各个序列的线性组合是平稳的,这样人们可以用数学方式来建模。这里所用的数学模型通常都是线性模型,或者是部分线性模型。

一般来说,具有规律的降水、气压、气温、湿度等序列是比较规范的时间序列。在与人类社会活动相关的序列中,食品和日用品的销售等是比较有规律的,一些宏观经济数据也是如此。但诸如股票、期货等贸易及金融数据就很难有较明确的规律性,这些时序的变动依赖于许多其他因素,而且与产生数据的市场的成熟程度有关。

下面介绍一些时间序列的例子。

1.2 时间序列例子

例 1.1 美国年度经济数据(sf.csv). 该数据来自Fisher (1994)^①。该数据的变量为DATE(相应于年末)、PRICEIND(1900—1938年道琼斯工业及交通运输指数,拼接到1939—1985年的纽约证交所综合指数)、PREMIUM(每年美国股票溢价)、USVEL2(组

^① 可以从网页<http://qed.econ.queensu.ca/jae/1994-v9.S/>下载。

约证交所成交率)和GROWTH(美国工业生产增长指数). 图1.1为该序列的点图. 该图是用下面语句生成的:

```
w=read.csv("sf.csv")
w=ts(w[,-1],start=c(1899,1),freq=1)
plot(w,main="",type="o",pch=16)
```

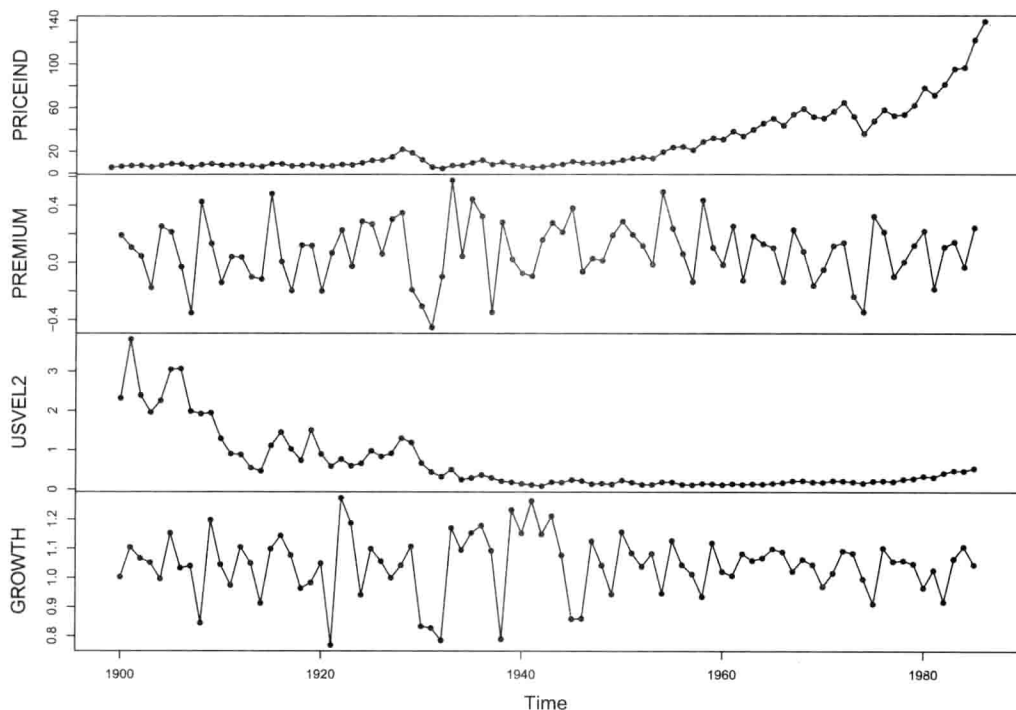


图 1.1 例1.1美国年度经济数据

这是一组经济时间序列, Fisher (1994)研究了这些序列之间的关系. PRICEIND和USVEL2序列随着时间变化表现出上升和下降的长期趋势. PREMIUM和GROWTH序列的波动较大, 且呈现一定的周期性. 在宏观经济分析领域, 这些时间序列数据扮演了很重要的角色. 通过对这些序列的预测和分析可以更好地为国家经济的管理服务. 本书的很多方法主要是用于经济金融领域的时间序列.

例 1.2 2000年1月到2012年10月新西兰5个城市(Auckland, Christchurch, Dunedin, Hamilton, Wellington)的月度总降水数据(NZRainfall.csv). 该数据可从网上下载^①. 图1.2为该序列的点图. 该图是用下面语句生成的:

^① 网址<http://new.censusatschool.org.nz/resource/time-series-data-sets-2013/>.

```
w=read.csv("NZRainfall.csv")
x=ts(w[,-1],start=c(2000,1),freq=12)
plot(x,ylab="Monthly rainfall (mm)",type="o",pch=16,nc=1,main="")
title("New Zealand Rainfall")
```

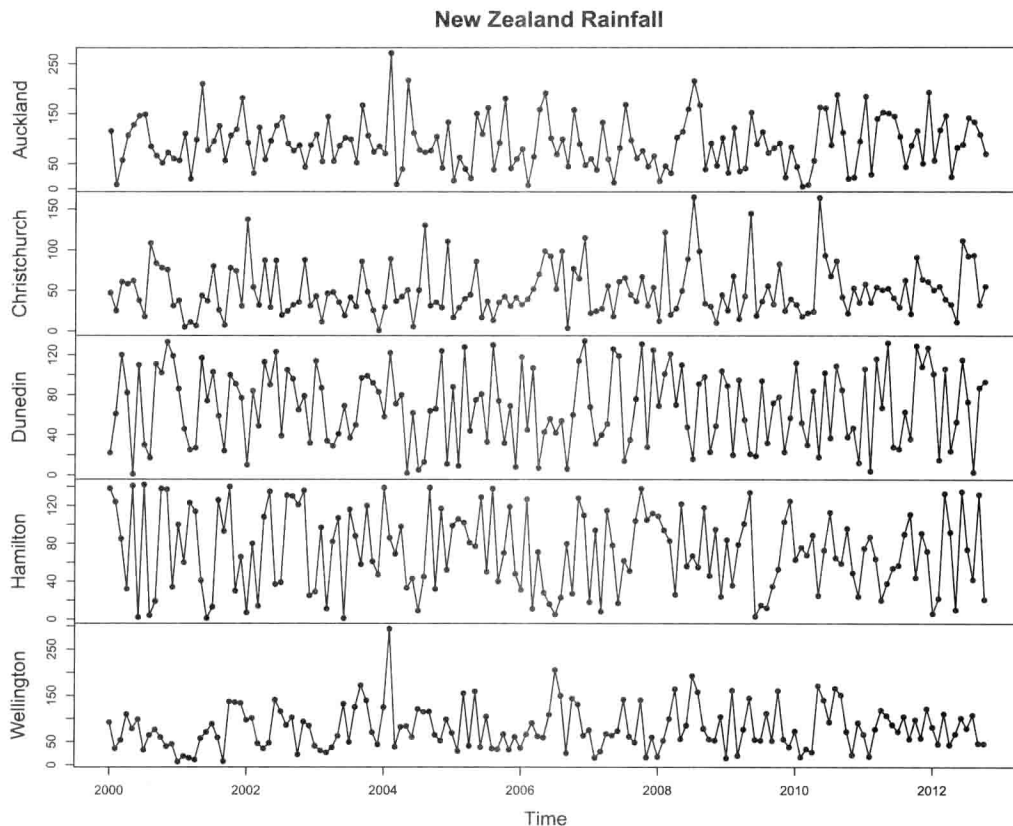


图 1.2 例1.2: 新西兰5个城市2000年1月到2012年10月的月度总降水量

这是一组降水数据,与例1.1比较,这组数据有比较强的周期性,可以看出降水量随月度变化呈现出较有规律的波动.从平均降水量上看,Dunedin和Hamilton的月度降水量比较相似,主要集中在0mm及120mm之间.Auckland和Wellington降水量的峰值相对较高.总体来说,月度降水量属于比较规律的自然现象,所以这几个序列并没有明显的总体上升或下降趋势变化.诸如降水和气温等气象数据是时间序列中比较有规律的一族.

例 1.3 1855—2011年苏格兰的结婚和离婚数据(marriages-divorces-1855-2011.csv). 该数据可从网下载^①.图1.3为该序列的点图.该图是用下面语句生成的:

^① 网址<http://www.gro-scotland.gov.uk/statistics/at-a-glance/time-series-datasets.html>.


```
w=read.csv("marriages-divorces-1855-2011.csv")
x=ts(w[,-1],start=c(1855),freq=1)
plot(x,ylab="Number of Marriages and Divorces",main="",
type="o",pch=16,plot.type="single",lty=1:2)
title("Marriage and Divorce in Scotland")
legend("topleft",c("Marriage","Divorce"),lty=1:2)
```



图 1.3 例1.3中1855—2012年苏格兰的结婚和离婚数据

这是一组1855—2011年苏格兰的结婚和离婚数据，没有周期性，但有上升及下降等趋势。这种社会学的时间序列例子很多，但并不见得容易用数学语言来描述。

前面的几个例子都是关于离散时间序列的，下面的描述性例子是关于连续时间序列的。这种连续时间序列不是本书研究的目标。此外，在本节的时间序列点图中，凡是离散时间序列的点图都包含了观测值点及它们的连线，以后的时间序列图一般都不一定标出观测值的点，但一定标出观测点之间的连线(因此看上去是连续的)。

例 1.4 状态为0和1的连续监控时间序列。图1.4为该序列的点图。该序列是模拟出来的。

