华章数学译丛

HZ BOOKS
华章教育

49

Time Series Analysis with Applications in R

(Second Edition)

时间序列分析及应用

R语言

（原书第2版）

（美） Jonathan D. Cryer
Kung-Sik Chan 著

潘红宇 等译

机械工业出版社
China Machine Press

华 学 译 丛
49

Time Series Analysis with Applications in R (Second Edition)

时间序列分析及应用

R语言（原书第2版）

阜阳师范学院图书馆 藏书

（美） Jonathan D. Cryer
Kung-Sik Chan 著

潘红宇 等译

机械工业出版社
China Machine Press

本书以易于理解的方式讲述了时间序列模型及其应用，主要内容包括：趋势、平稳时间序列模型、非平稳时间序列模型、模型识别、参数估计、模型诊断、预测、季节模型、时间序列回归模型、异方差时间序列模型、谱分析入门、谱估计、门限模型. 对所有的思想和方法，都用真实数据集和模拟数据集进行了说明.

本书可作为高等院校统计、经济、商科、工程及定量社会科学等专业学生的教材或教学参考书，同时也可供相关技术人员使用.

Translation from the English language edition: *Time Series Analysis with Applications in R, Second Edition* (ISBN 978-0-387-75958-6) by Jonathan D. Cryer and Kung-Sik Chan.

Copyright © 2008 Springer Science+Business Media, LLC.

Springer is a part of Springer Science+Business Media.

All Rights Reserved.

本书中文简体字版由 Springer Science+Business Media 授权机械工业出版社独家出版. 未经出版者书面许可，不得以任何方式复制或抄袭本书内容.

封底无防伪标均为盗版

版权所有，侵权必究

本书法律顾问　北京市展达律师事务所

本书版权登记号：图字：01-2010-6353

图书在版编目（CIP）数据

时间序列分析及应用：R 语言（原书第 2 版）/（美）克莱尔（Cryer, J. D.）等著；潘红宇等译. —北京：机械工业出版社，2011.1

（华章数学译丛）

书名原文：Time Series Analysis with Applications in R, Second Edition

ISBN 978-7-111-32572-7

Ⅰ. 时…　Ⅱ. ①克…　②潘…　Ⅲ. 时间序列分析　Ⅳ. O211.61

中国版本图书馆 CIP 数据核字（2010）第 231026 号

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：迟振春

北京市荣盛彩色印刷有限公司印刷

2011 年 1 月第 1 版第 1 次印刷

186mm×240mm · 22.5 印张

标准书号：ISBN 978-7-111-32572-7

定价：48.00 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

购书热线：（010）68326294；88379649；68995259

投稿热线：（010）88379604

读者信箱：hzjsj@hzbook.com

译 者 序

时间序列的教材版本众多，其中有的教材侧重理论的讲述，读者需要具备较深厚的数学基础，主要阅读对象是统计类专业的学生；有的教材则注重模型的应用，理论和技术细节不是重点，主要面向经济类专业的学生. 而由 Jonathan D. Cryer 和 Kung-Sik Chan 所著的本书则均衡地介绍了时间序列的理论与应用，使之能满足更多专业方向的学生和研究者的需求. 在理论方面，本书给出一般性理论描述的同时，注重通过各种简单特例演绎具体的推导过程，因而清晰地阐释了有关结论，方便读者对理论的理解. 在一些章节后面，还通过附录的方式，补充给出了正文里有关结论的数学推导和相关基本概念，为熟悉数学理论的读者提供了深入理解各类方法的素材. 在应用方面，本书提供了基于模拟数据和取材广泛的真实数据的丰富例证，通过基于模拟数据的例子使读者深刻认识时间序列的基本性质，而通过基于真实数据的例子使读者体会模型的实际应用效果.

本书的另一个特色是，各章都提供了实现所有实证结果的 R 程序，并在附录Ⅰ里对 R 语言给出了详细介绍. 作为一款免费的软件，R 语言可提供广泛的统计和作图技术，并且具有高度扩展性，已为统计学家和计量经济学家广泛使用. 另外，很多最新的理论方法的 R 实现程序还可以很方便地从网络上查找到. 通过本教材的学习，读者能够快速掌握 R 软件的使用方法，利用既有的程序达成研究目的.

本书适合一学期的教学安排，内容介绍深入浅出，概念严谨准确，是一本不错的入门教材.

本书的翻译工作凝聚了多人的劳动，具体分工如下：潘红宇翻译了第 1、2、7、13、14、15 章，王玲玲翻译了第 3 章和第 9 章，李瑶帆翻译了第 4 章和第 5 章，梁丽英翻译了第 6 章和第 8 章，关晨翻译了第 10 章和第 11 章，闵敏翻译了第 12 章，秦智鹏翻译了第 16 章. 最后，由潘红宇对全书进行了统稿.

由于译者水平有限，译稿中的疏漏在所难免，敬请读者批评指正.

译 者

2010 年 11 月

前　言

自 1970 年 George E. P. Box 和 Gwilym M. Jenkins 的奠基性著作《时间序列分析：预报与控制》(Time Series Analysis: Forecasting and Control) 问世以来（该书已在 Gregory C. Reinsel 的加盟下，于 1994 年出版了第 3 版），时间序列分析的理论和实践都有了飞速发展，也涌现了一大批时间序列方面的书籍，遗憾的是有些书欠缺实际应用的论述，而另一些书则理论基础薄弱. 本书力图同时介绍时间序列的理论与应用，采用自然的方式将两方面内容有机地结合起来，使之能够为范围更广的学生和实践工作者所理解和接受.

本书可作为一学期的课程教材，供统计、经济、商科、工程及定量社会科学等专业学生使用，读者需要具备从基本应用统计到多元线性回归等多方面的基础知识. 只要读者具备求解类似平方和最小化等问题深度的微积分知识，即可阅读本书. 而要深入理解本书部分理论，则需具备基于微积分的统计引论基础. 书中附录回顾了有关期望、方差、协方差、相关等概念，并简述了条件期望的性质以及最小均方误差预测等内容. 全书采用来自不同专业领域的实际时间序列数据来阐述方法论. 本书还包括部分高级内容，教师可酌情选讲.

书中所有的图和数值输出均使用 R 软件得到，该软件可从 www. r-project. org 上的 "The R Project for Statistical Computing" 得到. 为简明起见，我们对部分数值输出进行了处理. 作为一款免费软件，R 软件的源代码格式符合自由软件基金会 GNU 通用公共许可证规定，可以在 UNIX 平台及 Windows、MacOS 等类似操作系统上运行.

R 是一种可用于统计计算和作图的编程语言及环境，可提供广泛的统计（例如，时间序列分析、线性及非线性建模、经典的统计检验）及作图技术，并且具有高度的扩展性. 在附录 I "R 入门" 中，采用与本书内容相配的方式对 R 软件进行了介绍. 本书作者之一（Kung-Sik Chan）制作了大量可用于本书的新增或增强的 R 函数，列于附录 I 的最后，并可从 R 计划网站 www. r-project. org 上的 TSA 程序包中找到. 我们还为每一章建立了 R 命令脚本文件，可从本书网站 www. stat. uiowa. edu/～kchan/TSA. htm 下载. 本书中，每一个图表下都给出了 R 代码；习题所需数据集合均有相关的文件名，例如洛杉矶降雨量的数据文件命名为 larain. 而如果读者使用的是 TSA 程序包，则该数据集合是程序包的一部分，可通过 R 命令 data(larain) 取得.

本书所有数据集合以 ASCII 码文件的形式放在网站上，并在文件第一行标注了各自名称. 书中很多图及计算结果，使用 SAS©、Splus©、Statgraphics©、SCA©、EViews©、RATS©、Ox© 及其他软件也可以得到.

本书是 1986 年 PWS-Kent Publishing (Duxbury Press) 出版的 Jonathan Cryer 所著《时间序列分析》第 2 版，新版在补充大量最新材料、数据集合和习题的同时，仍然包括所有已为读者熟悉的原版内容. 其中既有与原版内容融为一体的若干新论题，如涉及单位根检验、扩展自相关函数、ARIMA 模型子集以及自助法等内容，也有时间序列回归模型、异方差时间序列模型、谱分析和门限模型等全新的章节. 与基本内容相比，新章节内容的难度水平有某种程度的提高，但我们确信本书所采用的讨论方式易为读者接受，一定会对广泛的读者群学习相关内容有所帮助. 虽然涉及非线性时间序列模型的第 15 章（门限模型）位于本书的最后，但相关

试读结束：需要全本请在线购买：　www.ertongbook.com

内容也可以根据教学需要提前讲授，例如可以在第 12 章之后讲授. 同样，讨论谱分析的第 13、14 两章内容也可以在第 10 章之后学习.

感谢 Springer 出版社“统计学丛书”的责任编辑 John Kimmel，他在本书长时间的写作过程中给予了作者持续的关注与指导. 伦敦经济学院的汤家豪教授、中国台北中央研究院的蔡恒修教授、西北大学的 Noelle Samia 教授、中国香港大学的李伟强教授和吴启宏教授、奥斯陆大学的 Nils Christian Stenseth 教授等热心研读过书稿部分章节，Jun Yan 教授曾以本版初稿为教材在艾奥瓦大学某班授课，感谢他们对本书提出的宝贵的建设性意见. 感谢 Samuel Hao 帮助整理习题解答和附录 I 两部分内容. 这里还要对在不同阶段匿名审阅并帮助改进书稿的审阅者一并致谢. 最后，作者之一（Jonathan D. Cryer）也借此机会，向在写作新版第一稿时提供了墨西哥城圣地亚哥俱乐部 Casa de Artes 舒适写作环境的 Dan、Marian 和 Gene 表达诚挚的谢意.

Jonathan D. Cryer
Kung-Sik Chan
2008 年 1 月于艾奥瓦州艾奥瓦城

目　录

第1章 引 论

通过一系列时间点上的观测来获取数据是司空见惯的活动. 在商业上，我们会观测周利率、日股票闭盘价、月价格指数、年销售量等. 在气象上，我们会观测每天的最高温度和最低温度、年降水与干旱指数、每小时的风速等. 在农业上，我们会记录每年作物和牲畜产量、土壤侵蚀、出口销售等方面的数字. 在生物科学上，我们会观测每毫秒心电活动的状况. 在生态学上，我们会记录动物种群数量的变动情况. 实际上，需要研究时间序列的领域是难以罗列的. 时间序列分析的目的一般有两个方面：一是认识产生观测序列的随机机制，即建立数据生成模型；二是基于序列的历史数据，也许还要考虑其他相关序列或因素，对序列未来的可能取值给出预测或预报.

本章将从广泛的应用领域中，介绍一些时间序列的实例. 时间序列及其模型的一个独特的性质是，通常我们不能假定观测值独立取自同一总体（例如，取自均值不同的总体），时间序列分析的要点是研究具有相关性质的模型.

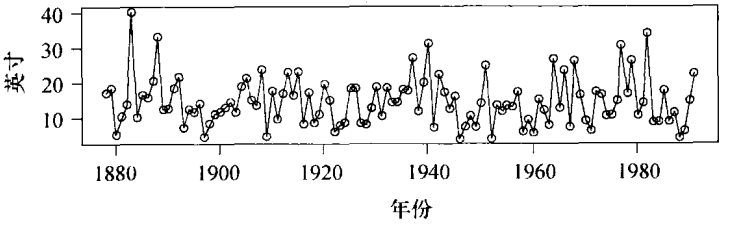
1.1 时间序列举例

本节介绍几个时间序列的例子，后续章节将进一步对其讨论.

洛杉矶年降水量

图表 1-1 是加利福尼亚州洛杉矶地区 100 多年来的年降水量时间序列图. 从图中可以看出，降水量在这些年有显著的差异——有的年份降水量低，有的年份降水量高，其他年份介于两者之间. 对洛杉矶来说，1883 年无疑是湿度特别大的一年，而 1983 年则相当干燥. 为了分析和建模需要，我们关心的是相邻年份的降水量是否存在某种关联. 若是，则可能依据当年的降水量数据预测来年的降水量. 我们可以画出相邻年份降水量的散点图，通过图形来研究这个问题.

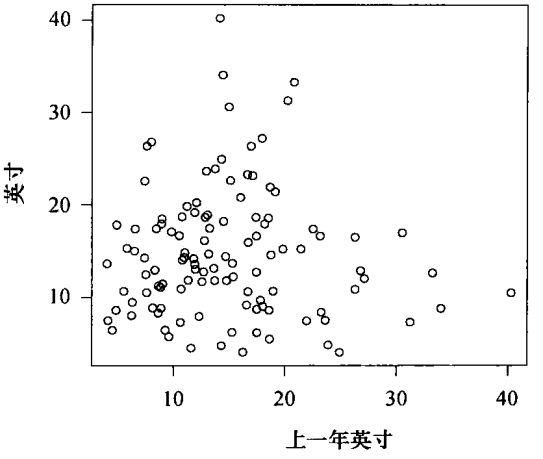
图表 1-1 洛杉矶年降水量时间序列图




```
> library(TSA)
> win.graph(width=4.875, height=2.5,pointsize=8)
> data(larain); plot(larain,ylab='Inches',xlab='Year',type='o')
```

图表 1-2 是由此绘出的降水量散点图. 例如，右下角的点显示降水量非常大的 1883 年·40 英寸的降水量，其后 1884 年降水量中等（约 12 英寸）. 图中靠近顶部的点表明 40 英寸降水量的年份，其上一年降水量比较典型，大约 15 英寸.

图表 1-2 洛杉矶当年降水量与去年降水量散点图



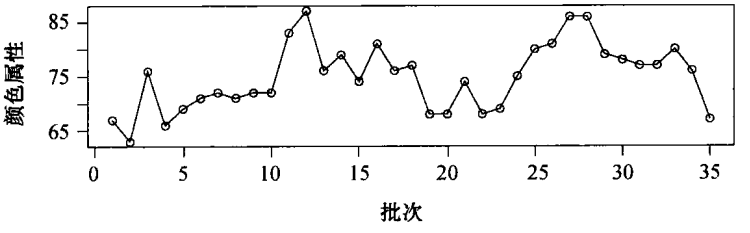

```
> win.graph(width=3,height=3,pointsize=8)
> plot(y=larain,x=zlag(larain),ylab='Inches',
   xlab='Previous Year Inches')
```

该图给人的主要印象是当年降水量与去年降水量几乎没有什么联系，既无“趋势”，也没有一般倾向. 上一年与当年降水量的相关性非常小，从预测和建模的角度，这样的时间序列没什么研究意义.

化工过程

第二个例子是来自某化工过程的时间序列. 这里变量度量的是过程中连续批次颜色的属性. 图表 1-3 是颜色值的时间序列图. 相邻时刻的颜色值差别不大，似乎相互之间存在关联.

图表 1-3 某化工过程中颜色属性的时间序列图



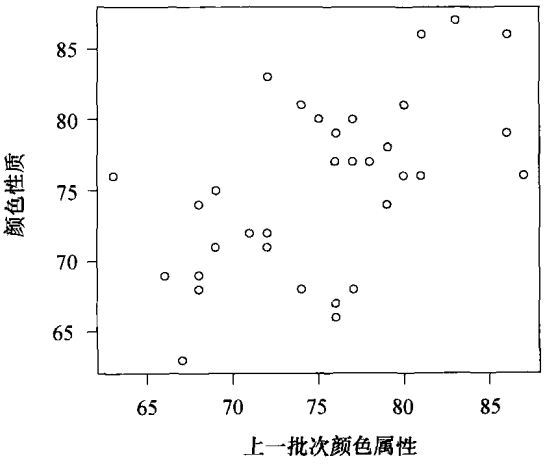

```
> win.graph(width=4.875, height=2.5,pointsize=8)
> data(color)
> plot(color,ylab='Color Property',xlab='Batch',type='o')
```

像第一个例子那样，制作一个相邻数据的散点图更能说明问题.

图表 1-4 是相邻颜色值的散点图. 该图显示了一个稍微向上的趋势——数值较小后面的批

次趋于较小的值，中等值后面的批次趋于中等值，数值较大后面的批次趋于较大的值．该趋势明显但并不非常强烈，例如，散点图的相关系数约为 0.6.

图表 1-4　当前颜色值与前期颜色值的散点图

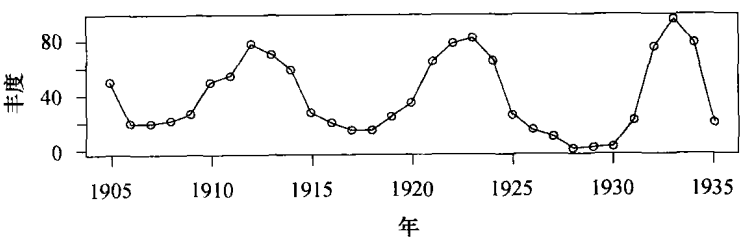



```
> win.graph(width=3,height=3,pointsize=8)
> plot(y=color,x=zlag(color),ylab='Color Property',
   xlab='Previous Batch Color Property')
```

加拿大野兔年丰度

第三个例子是加拿大野兔的年丰度．图表 1-5 给出了大约 30 年里该丰度的时间序列图．此例中，相邻数据联系非常密切，年度数据间没有大的变化．从图表 1-6 当年与上一年数量的散点图上，可以看到相邻数据间存在明显的相关性．类似上例，图形显示出一个向上的趋势——较小数值次年趋于较小的数值，中等数值趋于中等数值，较大的数值趋于较大的数值．

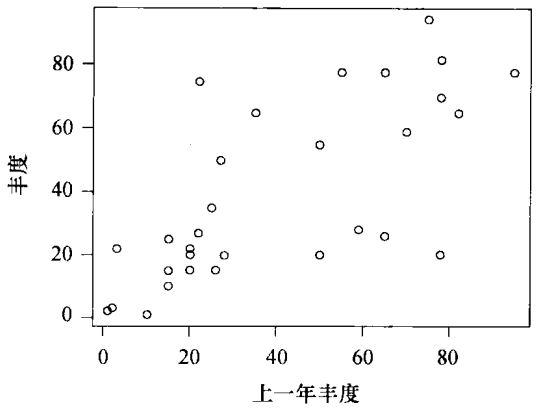
图表 1-5　加拿大野兔丰度




```
> win.graph(width=4.875, height=2.5,pointsize=8)
> data(hare); plot(hare,ylab='Abundance',xlab='Year',type='o')
```

试读结束：需要全本请在线购买：　www.ertongbook.com

图表 1-6 当年与上一年野兔丰度散点图

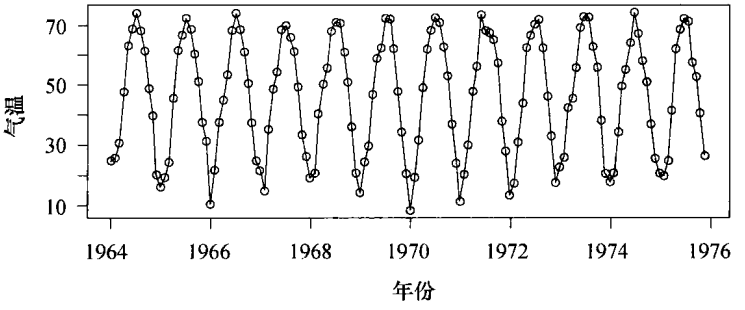



```
> win.graph(width=3, height=3,pointsize=8)
> plot(y=hare,x=zlag(hare),ylab='Abundance',
   xlab='Previous Year Abundance')
```

艾奥瓦州迪比克市月平均气温

艾奥瓦州迪比克市若干年里月平均气温（华氏度）的记录见图表 1-7.

图表 1-7 艾奥瓦州迪比克市月平均气温




```
> win.graph(width=4.875, height=2.5,pointsize=8)
> data(tempdub); plot(tempdub,ylab='Temperature',type='o')
```

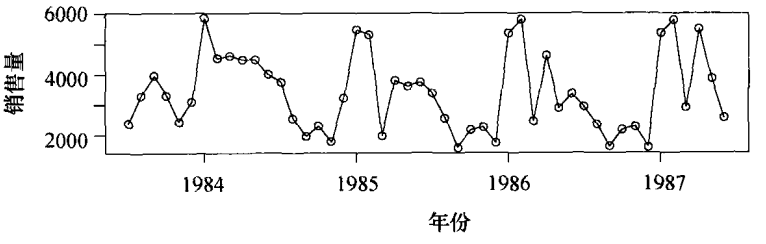
该时间序列显示了一种规则的被称为**季节性**的模式. 当分布在 12 个月的观测值以某种方式联系起来后，月度季节性就会显现出来. 比如，每年 1 月、2 月相当寒冷，温度近似，但是与温暖的 6 月、7 月、8 月的温度差别很大. 不同年份 1 月的温度值有差异，同样，不同年份 6 月的数值也有变化. 适用于这类序列的模型，必须反映出数据间的差异及其相似性. 此例中数据产生季节性的原因很好理解——北半球面向太阳的倾角随季节变化所致.

滤油器月销售量

本章最后一个例子是出售给经销商的滤油器月销售量. 该滤油器由 John Deere 生产，是

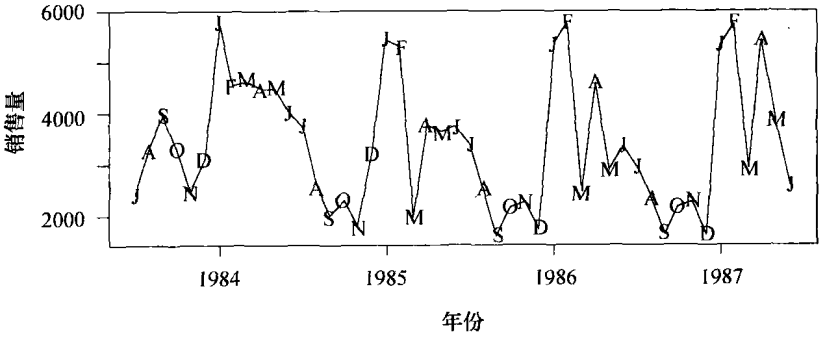
用于建筑设备的特制滤油器. 向作者之一提供数据时，经理说，“没有理由认为销售量存在季节性.”假如各年 1 月与 1 月的数据间存在关联趋势，2 月与 2 月的数据间存在关联趋势，等等，那么季节性就存在. 图表 1-8 没有显示出明显的季节性. 图表 1-9 与图表 1-8 相同，但是增加了有意义的符号对数据点进行标识. 图中所有 1 月的数值标以相同字母 J，2 月的数值标以 F，3 月的标以 M，等等[⊖]. 借助这些点所标记的符号，很容易看出所有 1 月、2 月等冬季月份的销量往往很高，而 9 月、10 月、11 月等月份的销量一般很低. 数据的季节性在这个处理过的时间序列图上是显而易见的.

图表 1-8　滤油器月销售量




```
> data(oilfilters); plot(oilfilters,type='o',ylab='Sales')
```

图表 1-9　以特殊符号绘制的滤油器月销售量



J=1月（和6月、7月）
F=2月, M=3月（和5月）, 以此类推

```
> plot(oilfilters,type='l',ylab='Sales')
> points(y=oilfilters,x=time(oilfilters),
    pch=as.vector(season(oilfilters)))
```

总之，我们的目的是，强调恰当和有益于发现特定模式的绘图方法，有利于找到符合时间

⊖ 看图时，读者仍然需要区分 1 月、6 月和 7 月（译者注：January，June，July），同样要区分 3 月与 5 月（译者注：March，May），4 月与 8 月（译者注：April，August），但通过对比相邻点上的符号很容易区分它们.

序列数据的合适模型. 在后续的章节中，我们将探讨一些在时间序列模型里引入季节性的不同方式.

1.2 建模策略

给时间序列寻找合适的模型并非易事. 下面介绍 Box and Jenkins(1976) 书中推崇的多步建模策略，该过程包括三个可反复使用的主要步骤：

1. **模型识别**（或称作辨识）
2. **模型拟合**
3. **模型诊断**

在模型识别阶段，从时间序列模型类中选择适合观测数据的模型. 这一步我们可以观察该序列的时间序列图，从数据出发计算一些不同的统计量，也可以利用任何生成该序列的背景知识，比如生物学、商业或生态学方面的知识等. 需要强调的是目前选取的模型是待考的，将在以后的分析过程中予以修正.

模型选取遵从**简约原则**，即应在能充分表示时间序列的前提下模型所含参数个数最少，正像 Parzen(1982，68 页) 中引用的爱因斯坦的名言所说，“凡事皆宜尽力简化，只要不失之草率.[⊖]”

模型所含的一个或多个参数需由观测序列给出估计，模型拟合就是要找到给定模型之未知参数的最优估计，我们采用的估计优化准则是最小二乘法则和极大似然法则.

模型诊断关注于完成设定、估计步骤之后所确定模型的质量评估问题. 模型对数据的拟合度有多好？适用模型的前提能否合理地得到满足？经诊断没有不足之处，则建模过程就结束了，所确定的模型也就可以应用了，例如可用于预测序列未来的取值等. 否则，针对找到的不足选取其他模型，即重新开始模型设定步骤. 采用这种方法，通过反复进行上述三个步骤，我们最终能够找到一个理想的、可接受的模型.

由于建模的每一步都要进行繁杂的计算，所以实际中可以依赖已有的统计软件进行这些计算，并绘出图形.

1.3 历史上的时间序列图

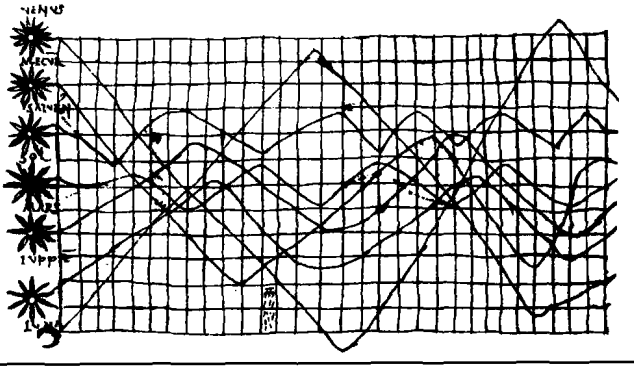
根据 Tufte(1983，28 页) 的说法，“时间序列图是图形设计最常用的形式. 其一个维度沿着秒、分、时、日、周、月、年，乃至千年等规则的时间节律延伸，时间标度的自然顺序赋予了这种设计以解释的力量和效率，这一点在其他图形设计上了无痕迹.”

图表 1-10 重现了有可能是目前已知的最古老的时间序列图，这个显示行星轨道倾角[⊜]的图可以上溯到公元 10 世纪（或 11 世纪），Tufte 对此的评价是：“这似乎是数据作图历史上一个神秘而孤立的杰作，因为目前所知下一个时间序列图的出现是 800 多年以后的事了.”

⊖ 爱因斯坦的原文是 “everything should be made as simple as possible，but not simpler”. ——译者注

⊜ 见 Tufte(1983，28 页).

图表 1-10　一个 10 世纪的时间序列图



1.4　本书概述

第 2 章介绍均值、方差、相关函数等基本概念，以重要的平稳性概念结束. 第 3 章探讨趋势分析的内容，并针对常用的有确定性趋势的模型，例如有线性时间趋势及季节性均值的模型，讨论了估计和验证的方法.

自第 4 章起，开始介绍平稳时间序列的参数模型，即所谓的自回归滑动平均（ARMA）模型（也称为 Box-Jenkins 模型）. 此类模型扩展后，可以涵盖某些随机非平稳的情况——ARIMA 模型，相关论述见第 5 章.

第 6、7、8 这三章介绍 ARIMA 模型核心的建模策略，包括初步设定模型的技术（见第 6 章），使用最小二乘法和极大似然法有效估计模型参数（见第 7 章），以及确定模型对数据如何更好拟合的方法（见第 8 章）.

第 9 章深入阐述 ARIMA 模型的最小均方误差预测理论与方法. 第 10 章把第 4 章到第 9 章介绍的思路进一步扩展，以应用于随机季节模型分析上. 后续各章探讨选定的若干高等主题.

习题

1.1　应用软件绘出与图表 1-2 一样的时间序列图，数据在名为 **larain** 的文件中[⊖].

1.2　绘出与图表 1-3 一样的时间序列图，数据文件名是 **color**.

1.3　模拟一个长度 48，完全随机的独立正态分布过程，并绘出时间序列图. 看看是否显示出“随机性”？使用不同的模拟样本，多次重复本练习.

1.4　模拟一个长度 48，完全随机、2 个自由度的独立χ^2分布过程，并绘出时间序列图. 看看是否显示出“随机性”和非正态性？使用不同的模拟样本，多次重复本练习.

1.5　模拟一个长度 48，完全随机、5 个自由度的独立 t 分布过程，并绘出时间序列图. 看看是否显示出“随机性”和非正态性？使用不同的模拟样本，多次重复本练习.

1.6　绘出与图表 1-9 一样的时间序列图，并对图上迪比克市温度序列给出月度标志，相关数据在名为 **tempdub** 的文件中.

⊖ 如果读者已从 www. r-project. org 下载并安装了 R 程序包 TSA，则 **larain** 数据可通过 R 命令 **data(larain)** 取得. 在本书主页 www. stat. uiowa. edu/～kchan/TSA. htm 上也有相关数据的 ASCII 文件.

第2章 基本概念

本章介绍时间序列模型理论中的基本概念. 特别地，我们介绍了随机过程、均值、协方差函数、平稳过程和自相关函数等概念.

2.1 时间序列与随机过程

随机变量序列 $\{Y_t: t=0, \pm1, \pm2, \pm3, \cdots\}$ 称为一个**随机过程**，并以之作为观测时间序列的模型. 已知该过程完整的概率结构是由所有 Y 的有限联合分布构成的分布族决定的. 幸运的是，联合分布中的大部分信息可以通过均值、方差和协方差加以描述，我们无需直接处理这些多元分布. 因此，我们将把注意力集中在对一阶和二阶矩的研究上. (如果 Y 的联合分布是多元正态分布，则所有的联合分布都可以由一阶和二阶矩完全确定.)

2.2 均值、方差和协方差

对随机过程 $\{Y_t: t=0, \pm1, \pm2, \pm3, \cdots\}$，**均值函数**定义如下：

$$\mu_t = E(Y_t), t = 0, \pm1, \pm2, \cdots \tag{2.2.1}$$

即 μ_t 恰是过程在 t 时刻的期望值. 一般地，不同时刻 μ_t 可取不同的值.

自协方差函数 $\gamma_{t,s}$ 定义如下：

$$\gamma_{t,s} = \mathrm{Cov}(Y_t, Y_s), t, s = 0, \pm1, \pm2, \cdots \tag{2.2.2}$$

其中 $\mathrm{Cov}(Y_t, Y_s)=E[(Y_t-\mu_t)(Y_s-\mu_s)]=E(Y_tY_s)-\mu_t\mu_s$.

自相关函数 $\rho_{t,s}$ 由下式给出：

$$\rho_{t,s} = \mathrm{Corr}(Y_t, Y_s), t, s = 0, \pm1, \pm2, \cdots \tag{2.2.3}$$

其中

$$\mathrm{Corr}(Y_t, Y_s) = \frac{\mathrm{Cov}(Y_t, Y_s)}{\sqrt{\mathrm{Var}(Y_t)\mathrm{Var}(Y_s)}} = \frac{\gamma_{t,s}}{\sqrt{\gamma_{t,t}\gamma_{s,s}}} \tag{2.2.4}$$

在附录 A 中，我们回顾了期望、方差、协方差和相关的基本性质.

回忆下述结论，协方差和相关系数都是随机变量间（线性）相关关系的度量，而某种程度上无量纲的相关系数更容易理解，那么从已知的结果及前述定义，可得如下的重要性质：

$$\left.\begin{array}{ll} \gamma_{t,t} = \mathrm{Var}(Y_t) & \rho_{t,t} = 1 \\ \gamma_{t,s} = \gamma_{s,t} & \rho_{t,s} = \rho_{s,t} \\ |\gamma_{t,s}| \leqslant \sqrt{\gamma_{t,t}\gamma_{s,s}} & |\rho_{t,s}| \leqslant 1 \end{array}\right\} \tag{2.2.5}$$

$\rho_{t,s}$ 的值接近 ±1 时，说明（线性）相关程度强，而接近 0 时，则说明（线性）相关程度弱. 若 $\rho_{t,s}=0$，则称 Y_t 和 Y_s 不相关.

在研究不同时间序列模型协方差的性质时，反复用到如下结论：如果 $c_1, c_2, \cdots, c_m$ 和 $d_1, d_2, \cdots, d_n$ 表示常数，$t_1, t_2, \cdots, t_m$ 和 $s_1, s_2, \cdots, s_n$ 表示时点，则有：

$$\mathrm{Cov}\left[\sum_{i=1}^{m} c_i Y_{t_i}, \sum_{j=1}^{n} d_j Y_{s_j}\right] = \sum_{i=1}^{m}\sum_{j=1}^{n} c_i d_j \mathrm{Cov}(Y_{t_i}, Y_{s_j}) \tag{2.2.6}$$

虽然方程（2.2.6）证明繁琐，但仅仅是直接应用了期望的线性性质. 作为它的一个特例，可得如下为人熟知的结果：

$$\mathrm{Var}\left[\sum_{i=1}^{n} c_i Y_{t_i}\right] = \sum_{i=1}^{n} c_i^2 \mathrm{Var}(Y_{t_i}) + 2\sum_{i=2}^{n}\sum_{j=1}^{i-1} c_i c_j \mathrm{Cov}(Y_{t_i}, Y_{t_j}) \tag{2.2.7}$$

随机游动

令 e_1，e_2，…为均值为 0，方差是 σ_e^2 的独立同分布的随机变量序列，观测时间序列 $\{Y_t: t=1, 2, \cdots\}$ 构造如下：

$$\left.\begin{aligned} Y_1 &= e_1 \\ Y_2 &= e_1 + e_2 \\ &\vdots \\ Y_t &= e_1 + e_2 + \cdots + e_t \end{aligned}\right\} \tag{2.2.8}$$

也可写成：

$$Y_t = Y_{t-1} + e_t \tag{2.2.9}$$

其“初始条件”$Y_1 = e_1$. 如果把 e 解释为沿数轴（前向或后向）游动的“步长”的大小，那么 Y_t 就是在时刻 t，“漫步者”到达的位置. 根据方程（2.2.8），得到均值函数：

$$\mu_t = E(Y_t) = E(e_1 + e_2 + \cdots + e_t) = E(e_1) + E(e_2) + \cdots + E(e_t) = 0 + 0 + \cdots + 0$$

因而，对所有的 t，

$$\mu_t = 0 \tag{2.2.10}$$

还可以得到：

$$\mathrm{Var}(Y_t) = \mathrm{Var}(e_1 + e_2 + \cdots + e_t) = \mathrm{Var}(e_1) + \mathrm{Var}(e_2) + \cdots + \mathrm{Var}(e_t) = \sigma_e^2 + \sigma_e^2 + \cdots + \sigma_e^2$$

故有

$$\mathrm{Var}(Y_t) = t\sigma_e^2 \tag{2.2.11}$$

注意随机过程的方差随时间线性地增长.

为了了解协方差函数，假设 $1 \leqslant t \leqslant s$，那么可以得到：

$$\gamma_{t,s} = \mathrm{Cov}(Y_t, Y_s) = \mathrm{Cov}(e_1 + e_2 + \cdots + e_t, e_1 + e_2 + \cdots + e_t + e_{t+1} + \cdots + e_s)$$

由方程（2.2.6），我们有：

$$\gamma_{t,s} = \sum_{i=1}^{s}\sum_{j=1}^{t} \mathrm{Cov}(e_i, e_j)$$

但是，除了 $i=j$ 以外，协方差都为 0. 当 $i=j$ 时，协方差等于 $\mathrm{Var}(e_i)=\sigma_e^2$. 这样的项恰有 t 个，因此 $\gamma_{t,s}=t\sigma_e^2$.

因为 $\gamma_{t,s}=\gamma_{s,t}$，故在所有时点 t 和 s 上，可以确定自协方差函数，记为：

$$\gamma_{t,s} = t\sigma_e^2, \quad 1 \leqslant t \leqslant s \tag{2.2.12}$$

易得随机游动的自相关函数为：

$$\rho_{t,s} = \frac{\gamma_{t,s}}{\sqrt{\gamma_{t,t}\gamma_{s,s}}} = \sqrt{\frac{t}{s}}, \quad 1 \leqslant t \leqslant s \tag{2.2.13}$$

下面的数值有助于对随机游动行为的理解.

试读结束：需要全本请在线购买： www.ertongbook.com