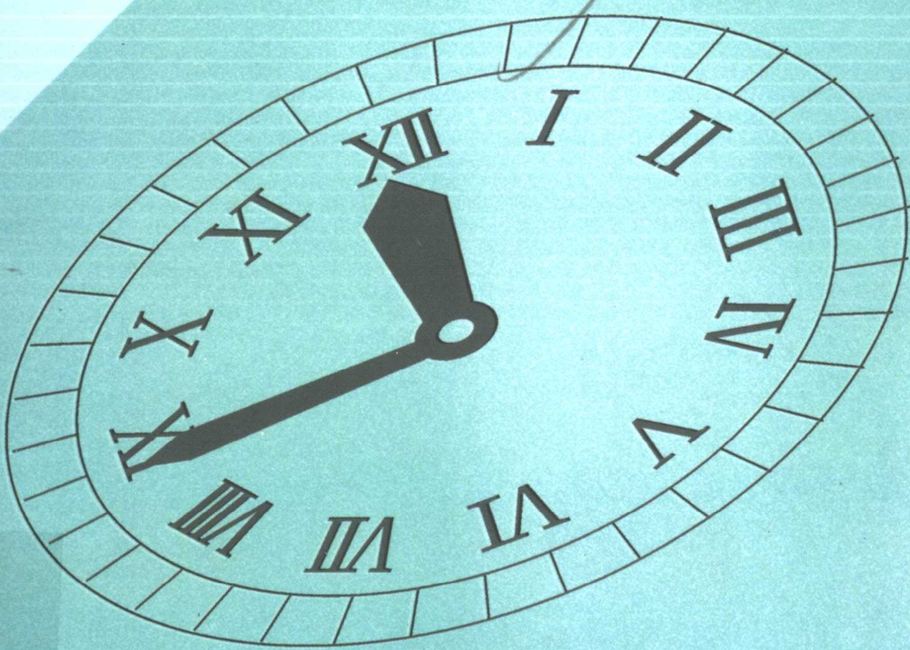
时间序列分析与综合

吴怀宇 编著

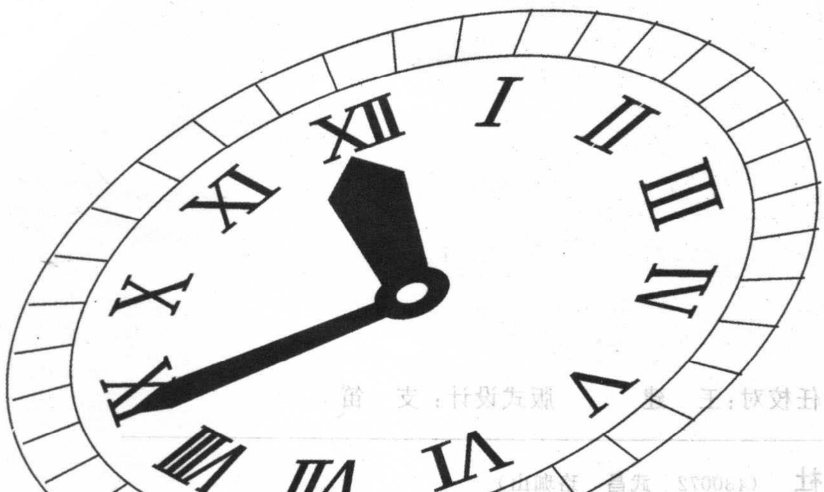
全国优秀出版社
武汉大学出版社



时间序列分析与综合

吴怀宇 编著

全国优秀出版社
武汉大学出版社



图书在版编目(CIP)数据

时间序列分析与综合/吴怀宇编著．—武汉：武汉大学出版社，2004．12
ISBN 7-307-04462-5

Ⅰ．时…　Ⅱ．吴…　Ⅲ．时间序列分析　Ⅳ．O211．61

中国版本图书馆 CIP 数据核字(2004)第 077449 号

责任编辑：李汉保　　责任校对：王　建　　版式设计：支　笛

出版发行：**武汉大学出版社**　（430072　武昌　珞珈山）
（电子邮件：wdp4@whu.edu.cn　网址：www.wdp.whu.edu.cn）
印刷：武汉中远印务有限公司
开本：787×1092　1/16　印张：13.5　字数：322 千字　插页：1
版次：2004 年 12 月第 1 版　2004 年 12 月第 1 次印刷
ISBN 7-307-04462-5/O·315　定价：20.00 元

内容简介

本书系统地介绍了时间序列分析的基本理论、基本方法、工程应用以及最新研究进展。全书共分7章，主要内容包括：时间序列分析的数学基础；时间序列模型及其特征函数；时间序列模型辨识与参数估计；时间序列模型结构识别与定阶方法；时间序列分析在带钢张力信号建模与分析中的应用；时间序列分析在飞行姿态建模与分析中的应用；时间序列分析的最新研究进展，包括时间序列与小波变换、时间序列与混沌理论、时间序列与分形理论、时间序列与数据挖掘等。

本书可作为高等院校自动化、电子信息、计算机、仪器仪表、机械、化工、航空航天等相关专业的高年级本科生、硕士生的教材，也可供从事动态数据分析和处理的相关科技工作者参考。

ABSTRACT

In this book, the basic theories, methods, engineering applications and its some recent developments in time series analysis are introduced systematically. The book consists of seven main parts: mathematic foundation of time series analysis; time series models and their characteristic functions; identification for time series models and their parametric estimation; model structure selection criteria; applications of time series analysis in engineering (i); applications of time series analysis in engineering (ii); some recent developments in time series analysis which includes time series and wavelet transformation, time series and chaos theory, time series and fractal theory, time series and data mining.

This book will be useful as a reference for both senior undergraduate and postgraduate levels in automation, electrical engineering, computer engineering, instrument engineering, mechanical engineering, chemical engineering, aeroastro engineering, etc. Also, it can be served as a reference for researchers in dynamic data analysis and processing area.

序

时间序列分析是数理统计中的一个重要分支,用随机过程理论和数理统计方法研究随机数据序列的规律。时间序列分析在经济领域中的研究和应用一直很活跃,并扩展到社会、气象、水利、交通、信息、农业和工业等领域。George E. P. Box 和 Gwilym M. Jenkins (1979)合著的《时间序列分析:预测和控制》(Time Series Analysis - Forecasting and Control)一书曾经引起广泛的重视,其后国内外的许多学者在多个领域进行了大量的研究,包括理论和应用,出版了许多专著,例如 J. Hamilton(1994)的《Time Series Analysis》是一本有影响易读的教科书,涉及一些经济计量学的内容。我国杨叔子等著(1991)的《时间序列分析的工程应用》,获得了第一届国家图书奖。此外,还有专门的期刊 Journal of Time Series Analysis、国际会议和专题讨论。2003 年诺贝尔经济学奖授予美国经济学家罗伯特·恩格尔(Robert F. Engle)和英国经济学家克莱夫·格兰杰(Clive W. J. Granger),表彰他们在分析经济时间序列方法(methods of analyzing economic time series)方面所作出的贡献,特别是他们提出的根据"时间变化变易率(time-varying volatility, ARCH 模型)"和根据"同趋势"(common trends, cointegration)的分析方法。这又一次说明了时间序列的研究和应用仍在发展中。

许多时间序列分析的工作也常常出现在系统辨识(System Identification)的文章和书刊中。在有的控制系统的著作中也把时间序列分析作为一个重要内容。特别是现代时间序列分析中出现了许多新的理论和方法,继维纳(Wiener)滤波和卡尔曼(Kalman)滤波理论和方法之后,现代时间序列分析已经成为从事经济、社会和工程等重要领域的科技工作者的重要工具。现代时间序列分析也包含以前的和现在的许多研究成果,例如卡尔曼滤波、小波分析等。

值得注意的是,时间序列分析不但在经济和社会等领域有应用价值,而且在工程领域也得到了十分广泛的应用。例如,生物工程中的 DNA 序列分析和生物医学信号序列分析、电子随机信号时序建模、机械故障诊断中的振动和噪声信号时序分析、工业自动控制过程时序建模与预报、精细化工过程的时序分析等。目前,国内外许多高等院校已将"时间序列分析"设为管理工程、电子信息、计算机和自动化、机电工程等专业的本科高年级学生和研究生的一门重要课程。前面已经提到,涉及经济和社会的时间序列分析的著作很多,工程和自然科学的读者希望有适合他们阅读的书籍,特别是成功的应用范例。

本书作者多年从事时间序列分析理论与应用方面的教学和科研工作,主持和完成了多项与时间序列分析相关的研究课题,阅读了国内外的大量著作,对本领域的研究方向有深入了解。《时间序列分析与综合》一书有两个特色:一是注重理论与实践相结合,突出时间序列分析的工程应用,综合了作者近年来的教学心得和相关的科研成果,这些应用成果进一步拓宽了时间序列分析的应用范围;二是吸收和归纳了近年来国内外在时间序列分析方面的最新研究成果。本书也有涉及工程应用的章节,特别是时间序列与其他学科交叉、融合的创新

方法，主要包括时间序列与小波分析、时间序列与混沌理论、时间序列与分形理论、时间序列与数据挖掘等。选题新颖，内容简明，注重基础，面向应用。

本书是为高等院校高年级学生或研究生的教学用书而著，也可供致力于时间序列分析与应用研究的科技工作者阅读和参考。

周兆英教授
2004年10月于北京清华园

前　言

在自然科学、社会科学及工程技术的许多领域,普遍存在着按时间顺序发生的具有概率特征的各种随机现象,人们通过观测把这些现象记录下来便成为可供分析的随机数据。所谓时间序列通常就是指这种有序的随机数据。这些数据有时本身就是离散数据,有时是随机的连续信号的采样值。实际上,时间序列就是离散的随机过程。例如,太阳黑子数目的变化、地震波的变化、雷达系统跟踪误差的变化、人脑电波的变化、市场物价的变化、机械振动信号的变化、自动化炼钢过程中钢水含碳量的变化等都属于随机过程。由于各自的物理背景不同,它们包含的信息和呈现的规律是千变万化、错综复杂的。人们正是希望通过分析这些数据序列,达到认识事物、掌握事物规律的目的。然而,如果仅从数据本身作一些简单处理(如简单的统计分析)是远远不能满足要求的。时间序列分析提供了一套具有科学依据的动态数据处理方法,该方法的主要手段是对各种类型的数据,采用相应的数学模型去近似描述。通过对模型的分析研究,便可更本质地了解数据的内在结构和复杂特性,从而达到预测其发展趋势并进行必要的控制的目的。

近年来,随着计算技术和信号处理技术的迅速发展,一方面,时间序列分析的理论和方法更趋完善,特别是在参数估计算法、模型结构识别与定阶方法以及与智能计算技术的融合等方面都取得了丰硕的研究成果。另一方面,时间序列分析的应用范围日益广泛,并且应用成果都处在一个较高水平层面上。例如,在控制工程领域,过程控制系统、运动控制系统的时间序列建模与预测;在 Internet 技术中,网络流量的时间序列模型分析;在数据库理论研究中,数据挖掘的时间序列方法;在电子信息领域,随机信号的时间序列建模与分析;在生物工程领域,DNA 序列分析与计算;在生物医学工程领域,生物肌电信号的时间序列分析;在机械故障诊断研究中,无损检测信号的时间序列分析;在精细化工控制中运用时间序列谱分析技术,等等。

本书主要涵盖三个方面的内容:一是时间序列分析的基本理论和计算方法;二是时间序列分析的工程应用;三是时间序列分析的最新研究进展。全书共分 7 章。第 1 章介绍了时间序列分析的数学基础;第 2 章介绍了时间序列模型及其特征函数;第 3 章介绍了时间序列模型辨识与参数估计;第 4 章介绍了时间序列模型结构识别与定阶方法;第 5 章介绍了时间序列分析在带钢张力信号建模与分析中的应用;第 6 章介绍了时间序列分析在飞行姿态建模与分析中的应用;第 7 章介绍了近几年国内外时间序列分析的最新研究进展。

本书的特色之一是注重理论与实践相结合,突出时间序列分析的工程应用,综合了作者近年来的教学心得和承担完成的与时间序列分析相关的科研成果,这些应用成果进一步弥补和拓宽了时间序列分析的应用范围。本书的特色之二是吸收和归纳了近年来国内外学者在时间序列分析方面的最新研究成果。特别是时间序列与其他学科领域交叉、融合的新进展和新方法,主要包括时间序列与小波分析、时间序列与混沌理论、时间序列与分形理论、时

间序列与数据挖掘等，全书取材新颖，内容丰富。

在本书的编写过程中，作者得到了香港城市大学、香港理工大学、清华大学、武汉大学、湖北省科技厅、湖北省教育厅、武汉钢铁(集团)公司、武汉科技大学等单位的领导和同行专家的鼓励和全力支持。研究生李俊、李腾飞、赵新、艾峥、陈媛、陈乔礼等同学参与了本书的多次校对工作。在此，作者谨向所有在出版本书过程中给予帮助和支持的领导、专家、学者、同事、研究生们表示衷心的感谢！

此外，在本书的编写过程中，作者借鉴和参考了近几年国内外大量的相关研究专著、教材和论文，在此，作者谨向本书参考文献中所列出的文献的编著者们表示衷心的谢意！

作者希望本书能在运用时间序列分析的基本理论和方法解决工程实际问题方面，对读者有所帮助和启示。尽管如此，由于作者水平所限，书中的错误或不妥之处在所难免，恳请读者批评指正。

最后，作者还要特别感谢武汉大学出版社的各位领导对于本书的出版所给予的大力支持，也衷心感谢各位编辑为本书的出版所付出的大量心血。

本书出版得到国家教育部重点研究基金(205098)、湖北省自然科学基金(2004ABA003)、湖北省教育厅重大研究基金(Z200511001)和武汉科技大学博士基金等项目的资助。

吴怀宇

2004 年 8 月

目　录

Contents

第 1 章　时间序列分析的数学基础

§1.1　随机过程的概念与基本类型

1.1.1　随机过程的基本概念

初等概率论研究的主要对象是一个或有限个随机变量或随机向量，尽管我们有时也讨论随机变量序列，但通常要假定序列之间是相互独立的．随着科学技术的发展，我们必须对某些随机现象的变化过程进行研究，这就需要考虑无穷多个随机变量，并且解决问题的出发点不是随机变量的 N 个独立样本，而是无穷多个随机变量的一次具体观测．因此，我们必须用一族随机变量才能刻画这种随机现象的全部统计特性．这样的随机变量族通常称为随机过程．

例 1.1.1　在气象预报中，若以 X_t 表示某地区第 t 次统计所得到的该天最高气温，则 X_t 是随机变量．为了预报该地区未来的气温，我们必须研究随机过程 $\{X_t, t=0,1,\cdots\}$ 的统计特性．

例 1.1.2　在通信工程中，电话交换台在时间段 $[0,t]$ 内接到的呼唤次数是与 t 有关的随机变量 $X(t)$，对于固定的 t，$X(t)$ 是一个取非负整数的随机变量，则 $\{X(t), t\in[0,\infty)\}$ 是随机过程．

例 1.1.3　在描述生物群体的演变过程中，若以 X_t 表示在时刻 t 群体的个数，则对每一个 t，X_t 是一个随机变量．假设我们从 $t=0$ 开始每隔 24 小时对群体的个数观测一次，则 $\{X_t, t=0,1,\cdots\}$ 是随机过程．

例 1.1.4　在海浪分析中，需要观测某固定点处海平面的垂直振动．设 $X(t)$ 表示在时刻 t 处的海平面相对于平均海平面的高度，则 $X(t)$ 是随机变量，而 $\{X(t), t\in[0,\infty)\}$ 是随机过程．

这里有必要先介绍随机过程分析中的几个常用名词．随机试验所有可能结果组成的集合称为这个试验的样本空间或基本事件空间，记为 Ω．Ω 中的元素 e 称为样本点或基本事件，Ω 的子集 A 称为事件，样本空间 Ω 称为必然事件，空集 $\varnothing$ 称为不可能事件．$\mathscr{F}$ 是 Ω 的某些子集组成的集合族．P 是 $(\Omega,\mathscr{F})$ 上的概率．

定义 1.1.1　设 $(\Omega,\mathscr{F},P)$ 是概率空间．T 是给定的参数集，若对每个 $t\in T$，有一个随机变量 $X(t,e)$ 与之对应，则称随机变量族 $\{X(t,e), t\in T\}$ 是 $(\Omega,\mathscr{F},P)$ 上的随机过程，简记为随机过程 $\{X(t), t\in T\}$．

随机过程 $\{X(t), t\in T\}$ 可以解释为一个物理系统．$X(t)$ 表示系统在时刻 t 所处的状态．$X(t)$ 的所有可能状态所构成的集合称为状态空间，记为 I．必须指出，参数 t 既可以指时间，又可以指别的．当 t 是向量时，则随机过程又可称为随机场．此外，从数学的观点来说，随机过程 $\{X(t,e), t\in T\}$ 是定义在 $T\times\Omega$ 上的二元函数．对固定的 t，$X(t,e)$ 是 $(\Omega,\mathscr{F},P)$

上的随机变量.对固定的 e, $X(t,e)$是定义在 T 上的普通函数,称为随机过程$\{X(t,e), t\in T\}$的一个样本函数,样本函数的全体又称为样本函数空间.

根据参数 T 及状态空间 I 是可列集或非可列集,通常把随机过程分为以下四种类型:

(1) T 和 I 都是可列的;

(2) T 非可列, I 可列;

(3) T 可列, I 非可列;

(4) T 和 I 都非可列.

参数集 T 可列的随机过程又称为随机序列或时间序列,一般用$\{X_t, t=0,\pm1,\pm2,\cdots\}$表示,如例 1.1.1 和例 1.1.3.状态空间 I 可列的随机过程又称为可列过程,如例 1.1.2 和例 1.1.3.例 1.1.4 对应于(4)类随机过程.

随机过程的分类,除上述按参数集 T 与状态空间 I 是否可列外,还可以进一步根据 X_t 之间的概率关系进行分类,如独立增量过程、马尔可夫过程以及平稳过程等.

1.1.2 随机过程的数字特征

研究随机现象,主要是研究它的统计特性.由概率论知,有限个随机变量的统计特性完全由它们的联合分布函数所确定.由于随机过程可视为一族或无穷多个随机变量,由概率论可知,采用无穷维分布函数的方法来描述其统计特性是行不通的,可行的办法是采用有限维分布函数族来刻画随机过程的统计特性.

定义 1.1.2 设 $X_T=\{X(t), t\in T\}$是随机过程,对任意 $n\geqslant 1$ 和 $t_1,t_2,\cdots,t_n\in T$,随机向量$(X(t_1),X(t_2),\cdots,X(t_n))$的联合分布函数为

$$F_{t_1,\cdots,t_n}(x_1,x_2,\cdots,x_n)=P\{X(t_1)\leqslant x_1,\cdots,X(t_n)\leqslant x_n\}. \tag{1.1.1}$$

这些分布函数的全体

$$F=\{F_{t_1,\cdots,t_n}(x_1,x_2,\cdots,x_n), t_1,t_2,\cdots,t_n\in T, n\geqslant 1\} \tag{1.1.2}$$

称为 $X_T=\{X_t, t\in T\}$的有限维分布函数族.

可以证明,随机过程 $X_T=\{X(t), t\in T\}$的有限维分布函数族 F 具有如下性质:

(1) 对称性.对于$\{t_1,t_2,\cdots,t_n\}$的任意排列$\{t_{i_1},t_{i_2},\cdots,t_{i_n}\}$,有

$$F_{t_1,\cdots,t_n}(x_1,\cdots,x_n)=F_{t_{i_1},\cdots,t_{i_n}}(x_{i_1},\cdots,x_{i_n}). \tag{1.1.3}$$

(2) 相容性.当 $m<n$ 时,有

$$F_{t_1,\cdots,t_m}(x_1,x_2,\cdots,x_m)=F_{t_1,\cdots,t_m,\cdots,t_n}(x_1,x_2,\cdots,x_m,\infty,\cdots,\infty). \tag{1.1.4}$$

反之,对给定的满足对称性和相容性条件的分布函数族 F,是否一定存在一个以 F 作为有限维分布函数族的随机过程呢?这就是随机过程的存在性定理要回答的问题.

定理 1.1.1 柯尔莫哥洛夫(Kolmogorov)存在定理 设已给参数集 T 及满足对称性和相容性条件的分布函数族 F,则必存在概率空间$(\Omega,\mathscr{F},P)$及定义在其上的随机过程$\{X(t), t\in T\}$, $X(t)$的有限维分布函数族是 F.

柯尔莫哥洛夫定理说明,随机过程的有限维分布函数族是随机过程概率特征的完整描述.由于随机变量的分布函数和特征函数的一一对应关系,随机过程的概率特征也可以通过随机过程的有限维特征函数族

$$\Phi = \{g_{t_1,\cdots,t_n}(\theta_1,\theta_2,\cdots,\theta_n), t_1,t_2,\cdots,t_n \in T, n \geqslant 1\} \tag{1.1.5}$$

来完整地描述,其中

$$g_{t_1,\cdots,t_n}(\theta_1,\theta_2,\cdots,\theta_n) = E\left[\exp\left(\mathrm{i}\sum_{j=1}^{n}\theta_j X(t_j)\right)\right]. \tag{1.1.6}$$

在实际问题中,要知道随机过程的全部有限维分布函数族是不可能的,一般是利用随机过程的某些统计特征来取代 F.随机过程常用的统计特征定义如下.

定义 1.1.3　设 $X_T=\{X(t),t\in T\}$是随机过程,如果对任意 $t\in T$,$E[X(t)]$存在,则称函数

$$m_X(t) = E[X(t)], t \in T \tag{1.1.7}$$

为 X_T 的均值函数.

对任意 $t\in T$,若 $E[X(t)]^2$ 存在,则称 X_T 为二阶矩过程;若 $E[X(t)]$存在,则称 X_T 为一阶矩过程,称

$$B_X(s,t) = E[(X(s) - m_X(s))(X(t) - m_X(t))], \quad s,t \in T \tag{1.1.8}$$

为 X_T 的协方差函数.称

$$D_X(t) = B_X(t,t) = E[X(t) - m_X(t)]^2, \quad t \in T \tag{1.1.9}$$

为 X_T 的方差函数.称

$$R_X(s,t) = E[X(s)X(t)], \quad s,t \in T \tag{1.1.10}$$

为 X_T 的相关函数.

由许瓦兹(Schwartz)不等式知,二阶矩过程的协方差函数和相关函数一定存在,且满足下列关系

$$B_X(s,t) = R_X(s,t) - m_X(s)m_X(t).$$

特别地,当 $m_X(t)=0$ 时,$B_X(s,t)=R_X(s,t)$.

均值函数 $m_X(t)$是随机过程$\{X(t),t\in T\}$在时刻 t 的平均值,方差函数 $D_X(t)$是随机过程在时刻 t 对均值 $m_X(t)$的偏离程度,而协方差函数 $B_X(s,t)$和相关函数 $R_X(s,t)$则反映随机过程$\{X(t),t\in T\}$在时刻 s 和 t 时的线性相关程度.

例 1.1.5　设随机过程

$$X(t) = Y\cos(\theta t) + Z\sin(\theta t), \ t > 0,$$

其中,Y 与 Z 是相互独立的随机变量,且 $E(Y)=E(Z)=0$,$D(Y)=D(Z)=\sigma^2$,求$\{X(t),t>0\}$的均值函数 $m_X(t)$和协方差函数 $B_X(s,t)$.

解　由数学期望的性质知

$$\begin{aligned} E[X(t)] &= E[Y\cos(\theta t) + Z\sin(\theta t)] \\ &= \cos(\theta t)E(Y) + \sin(\theta t)E(Z) = 0. \end{aligned}$$

由于 Y 与 Z 相互独立,故

$$\begin{aligned} R_X(s,t) &= E[X(s)X(t)] = E[Y\cos(\theta s) + Z\sin(\theta s)][Y\cos(\theta t) + Z\sin(\theta t)] \\ &= \cos(\theta s)\cos(\theta t)E(Y^2) + \sin(\theta s)\sin(\theta t)E(Z^2) = \sigma^2\cos[(t-s)\theta]. \end{aligned}$$

例1.1.6　设随机过程

$$X(t) = Y + Zt, \ t > 0,$$

其中 Y 与 Z 是相互独立的$N(0,1)$随机变量,求$\{X(t),t>0\}$的一维和二维概率密度函数.