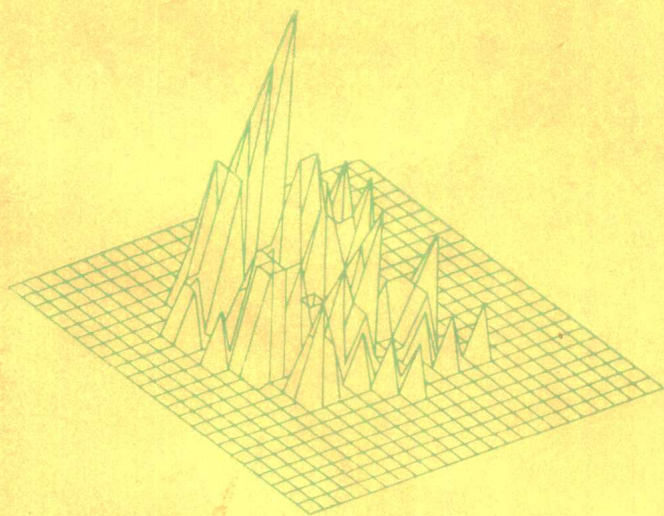


时间序列及其谱分析

陈兆国 编著



科学出版社

时间序列及其谱分析

陈兆国 编著

科学出版社

1988

内 容 简 介

本书是一本高级科普读物,主要介绍时间序列分析的方法和概念,使读者对这一学科中的基本思想有所了解,并且能掌握一些处理动态数据的最新方法,初步具有对实际数据进行分析、建模、预报、分类、检测的能力。

本书可供从事数据分析和处理的科研人员或工程技术人员阅读,也可供应用数学与概率统计专业的师生参考。

时间序列及其谱分析

陈兆国 编著

责任编辑 毕 颖

科 学 出 版 社 出 版

北京朝阳门内大街137号

中国科学院印刷厂印刷

新华书店北京发行所发行 各地新华书店经售

*

1988年7月第一版 开本:787×1092 1/32

1988年7月第一次印刷 印张:12 1/8

印数:0001—3,766 字数:275,000

ISBN 7-03-000423-X/O · 117

定 价: 3.15 元

序

虽然有关时间序列分析方面的著作在国外已非罕见，在国内也出版了一些，但阅读本书，仍会在不少方面有使人耳目一新之感。这本书反映了国内外时间序列分析的研究和应用的许多新进展，以相当的篇幅介绍了一些重要的新方法和有关理论，对于一些传统的或其它在国内已经较常用的有效方法，也作了简明扼要的叙述，可使读者对于这一领域的历史和现状有一完整的概念。

本书首先列出大量国内在时间序列分析应用中比较成功的实例，这对于各行业读者来说一定会有很大的吸引力。这些实例大部分来自全国第一届时间序列分析会议。我曾与本书作者和其他同志共同主持该次会议，深感国内这方面的成就，应当予以总结、宣扬，以激励心志。作者此举，也使我如愿以偿。本书以介绍实例为开端，接着深入浅出地阐述了时间序列分析的各种基本概念，然后介绍各类从数据得到统计规律的具体方法——建模与谱分析，并穿插以实例。这种处理途径，使直观、概念与方法浑然一体，是写高级科普图书的一种尝试。

本书作者从事时间序列分析的理论和研究方法研究多年，曾随国际著名的时间序列分析权威汉恩 (Hannan) 教授工作两年，受其深刻影响。所以，这本书虽然是以介绍方法为主，但读者在某些章节中可以领略一些汉恩学派的学术特点、研究和处理问题的思路，以及作者本人的造诣。作者近年来从事平稳序列线性建模及谱分析理论与方法的研究，他的这些工

作成就在书中也有所反映。

本书不仅对希望用时间序列分析方法解决实际问题的读者是一本实用教材，对有志于从事时间序列分析理论研究的青年读者也是一个有益的向导。沿着书中指明的某些重要文献前进，可掌握现代时间序列分析的基本理论和某些发展趋势。

愿此书能成为学习和使用时间序列分析理论与方法的读者的良师益友。

张志方

1986年7月于北京工业学院

前 言

目前,时间序列分析在经济、管理、天文、气象、水文、海洋、地球物理、生物、医学、机械、化工、纺织、电子、冶金、农林等很多科研和生产部门都得到了广泛的应用,并取得了很好的效果。本书旨在对时间序列分析中最基本的概念和方法作一比较系统的介绍,并展示某些具体的应用实例。

本书以说明基本概念和提供方法为主,并力图把它们建立在严格的概率论、数理统计和其他基础数学的理论上。但由于本书的性质和篇幅关系,对于其中的大部分定理和命题只着重于解释它们的背景和意义,严格的数学证明可查阅所引的参考文献。本书也提供了少数其他文献中没有或不易找到的某些定理的简捷证明。

本书收集了近几年来国内外在时间序列建模和谱分析方面比较实用的最新方法,也介绍了传统的方法,并且尽量避免和国内已出版的同类书籍重复,但为了本书体系的完整性,有时不得不做一些简要阐述。对于一些较老的且较繁琐的方法(例如,著名的鲍克思-詹金斯方法及各种非线性 ARMA 建模法)都不再涉及。

本书的理论和方法是自封的,如不追究定理和命题证明,读者可不查阅文献。但为了使有兴趣的读者进一步了解定理的证明,了解时序分析的某些历史、现状和发展,为学习更多的方法与理论提供一些建议,本书引用了一些参考文献。

本书试图采用由浅入深的叙述层次,在第一章第一节中给出时间序列的直观概念、一般内容、使用方法和许多实例,

第二至四节对平稳序列谱的概念、过滤的概念、各种序列的特点和性质作了比较通俗的介绍。第五节给出了预报方法。这几节基本上都属于平稳过程本身理论的范畴。第六节阐述了后两章所必需的统计学中的基本概念和两个最基本的统计量：样本协方差函数与周期图。该节在统计分析方面是承上启下的主要纽带。

第二章和第三章是本书的主要部分，分别介绍时间序列的时域分析和频域分析。第二章研究数据的分析方法和如何用它们建模，先从一维线性模型的参数估计和定阶开始，再向多维、非平稳、非线性等几个方面推广。第三章以一维平稳序列为主，前两节讨论各种谱密度的估计方法；第三节研究隐蔽频率的发现与估计方法；最后一节介绍多维序列频率成分统计分析中一些最基本的概念。通过第二、三章的学习，读者可以掌握用数据进行时序建模和谱分析的一些主要方法，并加以正确运用。

阅读本书不需要概率统计和高等数学的深奥理论，但读者最好能对概率统计中一些最基本的概念先有比较正确和深刻的理解，并且熟悉线性代数和矩阵论中的基本概念和运算规则。对于与实际工作已有很多接触的工程技术人员、科研人员、有关专业的教师及应用概率统计专业的学生，接受本书中的概念和方法是不困难的。

作者衷心感谢中国科学院应用数学研究所方开泰研究员对本书出版的关心和支持，衷心感谢北京工业学院张志方教授和王建新同志审阅本书并提出极为有价值的修改意见。

由于作者水平有限，错误、遗漏和其他不足之处在所难免，恳请读者批评指正。

作 者

1986 年于北京

目 录

序

前言

第一章 绪论与基本概念	1
第一节 时间序列分析的内容与应用	1
一、什么是时间序列	1
二、时间序列的概率定义、时域分析和参数模型	7
三、频域分析——谱及其估计与检验	14
四、应用	21
第二节 平稳过程及其谱	27
一、平稳过程的概率定义	27
二、平稳过程的功率谱及其协方差函数谱表示	30
三、平稳序列协方差函数谱表示、采样定理	43
四、平稳过程和平稳序列本身的谱表示	48
第三节 平稳序列的过滤与多维平稳序列	53
一、线性滤波器	53
二、平稳序列通过线性滤波器的变化	62
三、互协方差、互谱和多维平稳序列	64
第四节 线性序列、ARMA 序列	71
一、一般线性序列	71
二、线性序列	74
三、ARMA 序列	76
四、有理谱密度和 ARMA 序列的平稳可逆性	85
第五节 平稳序列的预报	90
一、最佳线性预报	90
二、平稳序列的预报	93
三、一般线性序列的预报	98

四、ARMA 序列的预报	102
第六节 统计方法与基本统计量	111
一、参数估计	111
二、估计量的评定	117
三、序列的样本均值和样本协方差、遍历性	122
四、有限傅里叶变换与周期图	129
第二章 时间序列模型	135
第一节 ARMA 模型的参数估计	135
一、AR 模型的参数估计	135
二、ARMA 模型的参数估计	146
第二节 ARMA 模型的定阶	156
一、自相关与偏相关定阶法	156
二、FPE、AIC、BIC 及用于 AR 模型定阶	161
三、ARMA 模型线性定阶法	169
第三节 多维 ARMA 模型、子集模型与模型检验	176
一、多维 ARMA 模型概述	176
二、多维 AR 模型与维特尔递推	179
三、子集模型——疏系数模型	183
四、模型残差的检验	191
第四节 回归与差分	196
一、对均值函数回归的模型	197
二、只对趋势差分的 ARIMA 模型	208
三、含有按季节差分的 ARIMA 模型	215
第五节 非线性模型	222
一、非线性现象与非线性模型	222
二、非线性模型的特点与非线性数据的处理	228
三、TAR 模型	237
四、非线性数据研究范例	241
第三章 谱分析	253
第一节 引言、谱密度估计的窗方法	253
一、引言	253
二、谱窗估计的定义	256
三、谱窗估计的渐近性质与估计精度	264

四、带宽、分辨率与窗的选择.....	272
五、数据加窗	282
第二节 极大熵与其他谱密度估计法	286
一、自回归逼近与极大熵原理	286
二、尤尔-沃克法与伯格法.....	292
三、最小二乘估计与马布尔递推	299
四、特征值方法及其他	311
第三节 谱跳跃的检验与估计	318
一、基于周期图坐标分布的隐蔽频率检验	319
二、隐蔽频率的精细估计	327
三、用周期图的样本性质寻找隐蔽频率	334
第四节 多维序列谱分析	354
一、多维序列的频率成分、谱密度的估计.....	354
二、复回归系数、全相干.....	357
三、主成分分析与典型相关分析	365
附表.....	373
参考文献.....	376

第一章 绪论与基本概念

第一节 时间序列分析的内容与应用

一、什么是时间序列

一个时间序列就是随时间变化的一列数。常记为 $\{x(t)\}$ 或 $\{y(t)\}$ 等。当用图形表示时,可把数据点用线段连接成曲线,这列数或曲线的特点是:第一,数据或数据点的位置依赖时刻,或者说数据取值随时间变化。第二,在每一时刻取什么值或点在什么位置有很大的随机性,不可能完全准确地用历史值预报。第三,前后时刻(不一定是相邻时刻)的数值或数据点在什么位置有一定的相关性。第四,有时从整体来看,有某种增长或下降的趋势,或出现周现性变化的现象。下面我们来看几个例子,请读者注意这些例子的上述特点。

例 1.1 图 1.1 是沃尔夫 (Wolf) 年平均太阳黑子数部分数据图,确切数字见本书附表 1。

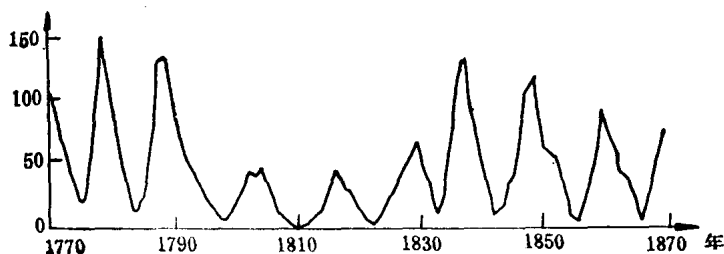


图 1.1 年平均太阳黑子数

最早用时间序列研究这一列数据的是尤尔 (Yule)。从五十年代以来，有关的文献急剧增加，其中莫兰 (Moran)、鲍克思与詹金斯 (Box and Jenkins^[1])、赤池 (Akaike)，特别是近年来汤家豪的工作^[2]都是较有影响的。由于太阳黑子的活动从长期来讲直接影响地球气候的变化，产生旱灾、涝灾，影响农牧业和野生动植物的生态及人类生活环境；从短期来讲，太阳黑子聚集到一定程度会产生耀斑，这时释放出的大量能量所产生的电磁辐射和粒子辐射可传播到地球，对电离层和磁层的突然骚扰会使无线电通讯中断，其高能粒子还威胁着宇航飞行器及人员的安全。那么，能否利用已有的数据从看来是不易捉摸的变化中找出一些规律，对未来的太阳黑子数作长期、中期以至短期的预报呢？这个问题将在后面回答。

例 1.2 图 1.2 是地极运动的变化曲线。

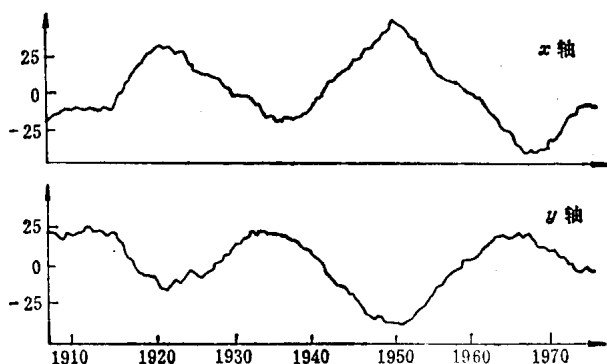


图 1.2 地极运动长周期摆动

我们知道，地球围绕着地轴自转。现代的天文测量技术发现，地球自转速率不是一个常数，而地极（地轴在南北极和地球表面的交点）并不固定在一一点上，地极大约在一个篮球场那样大的范围内摆动，如果把两个相差 90° 的经度方向作为 x 轴和 y 轴，那么地极的位置可由在这两个数轴上的坐标决

定。每隔一定时间测一次便得到两条不同坐标的时间序列。由于包括地极移动在内的地球自转参数对飞船的定位跟踪、大地测量及地震监测等都具有重要的意义，所以研究这个时间序列的变化规律是天文工作者的一大课题。

例 1.3 图 1.3 是我国地震局沈阳台站记录的主应力方向变化曲线，每旬一个数据，用折线相连而成。

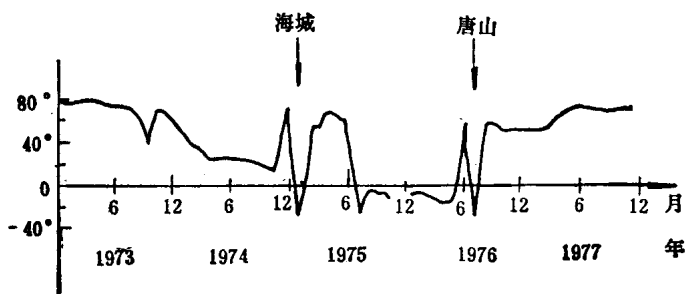


图 1.3 沈阳站主应力方向

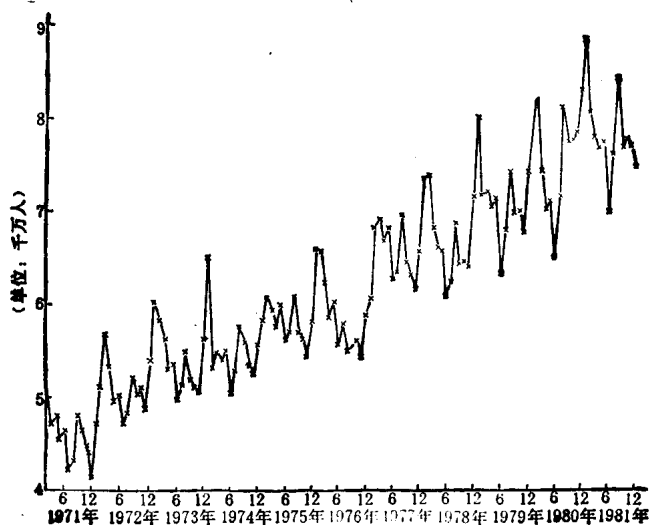


图 1.4 全国铁路客流量统计曲线

从图中可以看到, 在 1975 年初的海城地震和 1976 年夏的唐山地震时, 主应力方向都有一个突然变化的现象。但这是事后分析, 能否根据数据分析地震的前兆而给出某种预报的方法呢? 这个问题也将在后面回答。

例 1.4 图 1.4 是全国铁路客流量的统计曲线, 每月一个数据, 构成一条时间序列的曲线。直观上, 每年在春节前后有一个高峰, 并且逐年不断增加。我们的任务是根据以往的数据找出一些变化规律, 用以预报未来的月客流量。

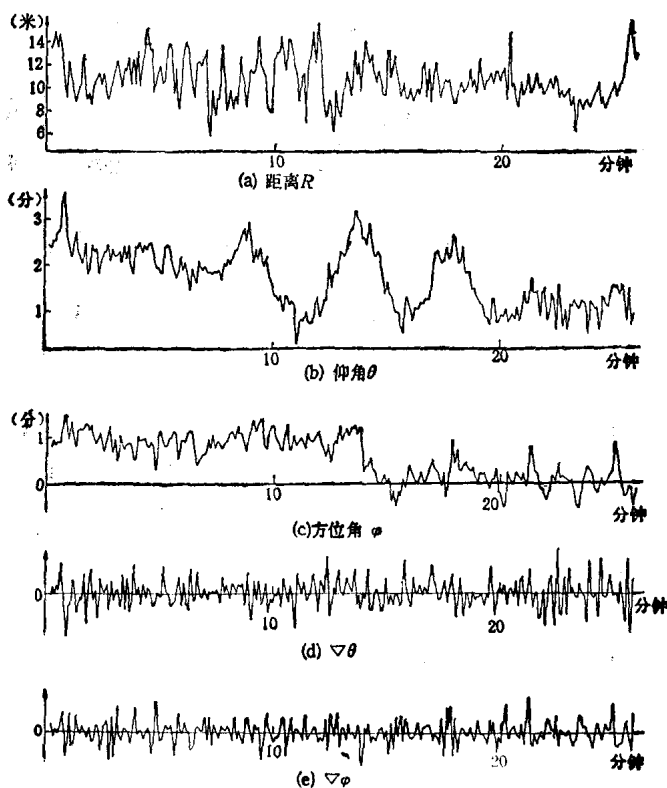


图 1.5 雷达测量误差曲线及其差分曲线

例 1.5 一部跟踪雷达的质量,最终由对跟踪目标的测量误差来决定。从某时刻(设为 0)开始,以等时间间隔进行采样,得第 t 时刻目标的位置数据为 $(\hat{R}(t), \hat{\theta}(t), \hat{\varphi}(t))$, 其中 R 为距离, θ 为仰角, φ 为方位角。如果在用雷达进行测量的同时,又用比雷达更精密的仪器,如经纬仪,也对目标进行测量,第 t 时刻数据为 $(R(t), \theta(t), \varphi(t))$, 可以把它们作为真实数值。那么测量误差可由 $(\Delta R(t), \Delta \theta(t), \Delta \varphi(t))$ 来表示, 其中, $\Delta R(t) = \hat{R}(t) - R(t)$ 。余者类似。图 1.5 的 (a), (b), (c), 分别是某次测量得到的 $\Delta R(t), \Delta \theta(t), \Delta \varphi(t)$ 的数据曲线, 它们都是时间序列。从图中可以直觉到这三个时间序列的前后数据依赖关系规律不太一样。对每个序列找出一些变化的规律,不仅可以鉴定雷达的质量,还可以在使用时预报跟踪误差,从而予以适当补偿。

例 1.6 图 1.6 所示的是加拿大马塔泽河流域山猫每年捕获量的 114 个数据序列(数据见附表 2), 由于数据变化范围太大, 图中画的是以 10 为底取对数后的数据值。由于这种山猫皮是名贵的皮毛, 故每年捕获量都有记录。从捕获量中可以看出当年这种山猫群体的总

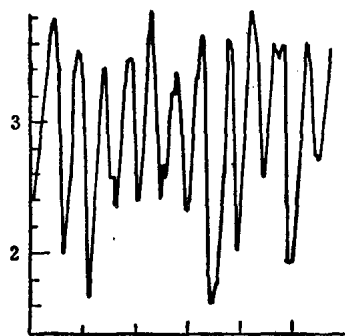


图 1.6 加拿大山猫每年捕获量(对数)

数。象太阳黑子一样,这也是时间序列分析中一组有名的数据。各家新创的方法都以能更精确地预报这条数据来作为方法优越的凭证。虽然在保护野生动物的今日,用它作预报无太大价值,但用它仍可找出某些生态学的规律。

例 1.7 图 1.7 是一条脑电图数据曲线。脑电图(EEG)

对临床具有重要意义,因而被广泛采用。用它可以判断人体很多特征,如健康或有某种疾病,正常智力或属愚型等等。按一定的时间间隔采样,可得时间序列,分析这种时间序列的特性可以推定人体的生理指标。



图 1.7 一条脑电图曲线

在大部分情况下,数据是随时间变化的,但有时也可以是随空间变化的。例如钻井时岩石的硬度是随深度变化的。但我们在讨论数量关系时,不去计较这些物理背景,仍称之为“时间序列”。所有的时间序列都可用 $\{x(t)\}$ 来表示, $x(t)$ 表示第 t 个数据。很多时间序列是天然的序列,如例 1.1、例 1.4、例 1.6,它是每年(或每月)一个年平均数(或月平均数)或年总数。但也有的时间序列从本质上看是连续变化过程中的一些观测值,如前面例 1.5 中提到的“采样”,也可称之为对连续过程的离散化。由于数字电子计算机的发展,处理离散的数据要比连续曲线更为精确、方便。但这里有个问题,就是采样间隔要多密才能不失(或基本不失)原来连续曲线所提供的全部信息?这是所谓的采样定理,我们将在下一节中介绍。

有时,在同一时刻不只对应一个量,而必须同时考虑几个有关系的因素,才能达到要研究的目的。如例 1.5 中,如果把雷达对 R 、 θ 、 φ 三个量的测量误差单独进行分析,可能达不到预期的精度,因现时刻 R 的误差不仅影响下时刻 R 的误差,而且也可能影响下一时刻 θ 和 φ 的误差,余者类似。所以我们常常令 $(\Delta R(t), \Delta \theta(t), \Delta \varphi(t))' \equiv x(t)$ 。对这三个误差一起讨论相互之间的关系和前后的影响,就构成了多维时间序

列。多维时间序列不可能用一条曲线表示，通常只能依靠数学公式表达。对于一个出现多个因素的实际问题，是应当作为一个多维序列来讨论，还是分成几个一维序列单独考虑，要看有没有必要和结果是否满意来决定。从本质上来说，取决于各因素之间的相互影响是否大。

二、时间序列的概率定义、时域分析和参数模型

前面我们已经提到，某时间序列 $\{x(t)\}$ 的特点可以用不同时刻上取值相互依赖的情况为标志。下面从两个极端情况入手对时间序列给出更科学的定义。

第一种，象下面的时间序列

$$x(t) = \beta_1 \cos \omega t + \beta_2 \sin \omega t \quad (1.1)$$

$$x(t) = \beta_1 + e^{\beta_2 t} \quad (1.2)$$

$$x(t) = \beta_0 + \beta_1 t \quad (1.3)$$

其中 β_i, ω 可以是常数，也可以是随机变量（即纯决定过程）。无论哪种情况，对于这样的序列，只要知道两个或三个时刻的 $x(t)$ 的值，便可定出 β_i 及 ω 的值，于是整个序列都成为已知的。从预报的观点来看，没有任何不确定性。

第二种序列，在任意两个不同时刻，比如， t 和 s ， $x(t)$ 和 $x(s)$ 都是随机变量且不相干，或更苛刻些，相互独立。这样的序列，有没有 t 时刻以前的信息，对于预报 $x(t)$ 都是一样的。因为 $x(t)$ 如何变化和 $x(s)$ 已经取什么值是没有（或基本上没有）关系的。其中最常用的是白噪声序列（为什么叫“白噪声”将在下一段里解释）。称 $\{\varepsilon(t)\}$ 是白噪声序列，是指对于每个 t ， $\varepsilon(t)$ 是一个随机变量，而且均值相同（通常可设为0，否则都减去这均值），方差相同，设为 σ^2 ，在不同时刻上互不相干。用公式表示，即