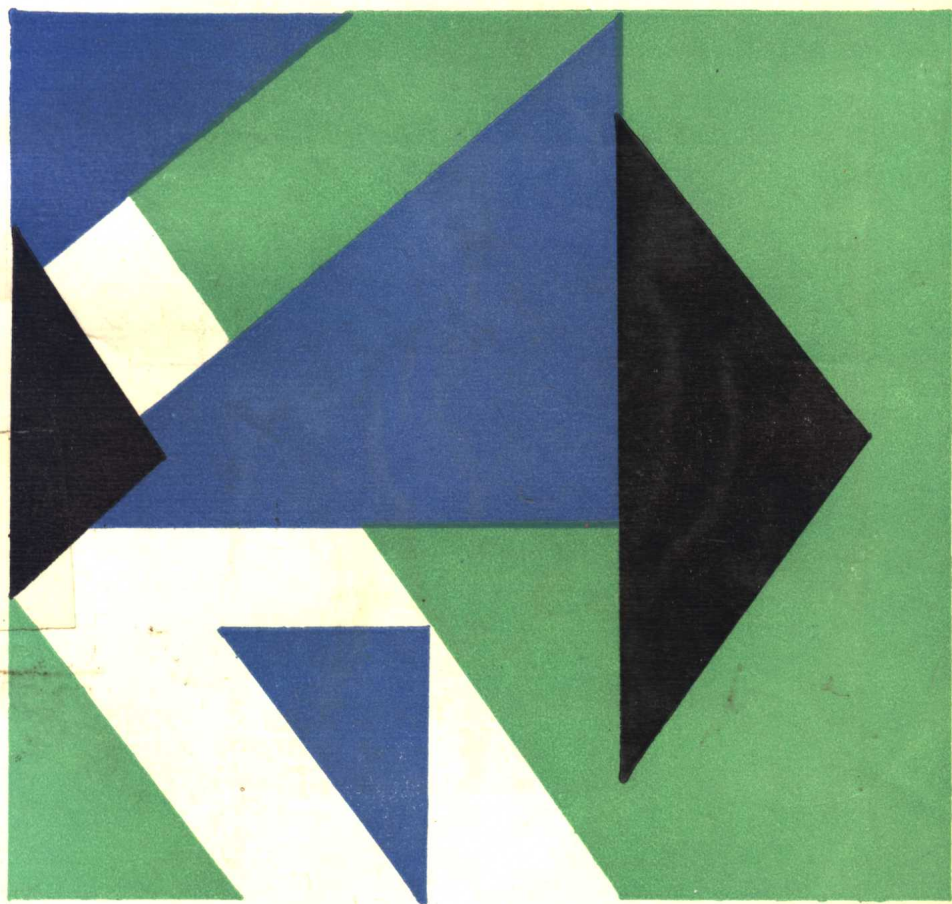


高等学校试用教材

时间序列分析

常学将 陈 敏 王明生 编著

高等教育出版社



高等学校试用教材

时间序列分析

常学将 陈 敏 王明生 编著

高等教育出版社

(京) 112 号

内 容 提 要

本书以讲授实用的时间序列分析的方法为主,也介绍了必要的理论知识。内容分为时间序列与随机过程、线性过程、ARMA 模型的建模、平稳时间序列的预报、非平稳时间序列的 ARIMA 模型、非线性模型与混合回归模型、谱分析、多维时间序列介绍、估计理论等九章,可供高等院校数学、应用数学、数理统计专业本科生与工科、财经类等专业的研究生作教材使用。

高等学校试用教材

时间序列分析

常学将 陈 敏 王明生 编著

*

高等教育出版社出版

新华书店总店科技发行所发行

中国科学院印刷厂印装

◆

开本 850×1168 1/32 印张 17 字数 440 100

1993 年 11 月第 1 版 1993 年 11 月第 1 次印刷

印数 0001—2 215

ISBN 7-04-004434-X/O·1249

定价 6.45 元

前 言

近 20 年来,时间序列分析在理论与应用两方面都得到蓬勃发展,成为概率统计学中一个内容十分丰富的重要分支.它在工程技术、气象、水文、地震、生物医学、经济管理以及军事科学等许多领域都得到广泛应用.这些应用研究的共同特点是处理动态数据,进行分析、预测与控制.时间序列分析着重研究具有随机性的动态数据,其早期工作是借助于富氏分析,从频率的角度揭示数据的规律性,即所谓的频域方法.由于快速富氏变换的产生,使得频域方法长期以来经久不衰.此外,借助于依时间推移变量之间的相关结构所提供的信息进行研究,即所谓时域分析方法,也是时间序列分析研究的有力工具.既然时间序列分析主要是处理随机性动态数据,那么概率统计知识是不可缺少的基础.目前,时间序列分析这个分支已趋于成熟,并奠定了自己的理论基础,特别是渐近理论保证了实用性方法的可靠性.因此,许多高校已把时间序列分析作为选修课开设.为了满足教学的需要,常学将于 1982 年编写了“时间序列分析讲义”初稿,此后和几位作者在教学中多次进行修改扩充,在山西大学数学系、山西师范大学数学系和山西财经学院计统系作为本科或研究生教材试用过多遍.最后作者根据 1991 年 12 月高等学校理科数学、力学教材编审委员会概率论与数理统计编审组提出的内容和要求,再次修改扩充,形成这本书.

本书的主要对象是大专院校数学、应用数学和数理统计专业本科生和工科、财经类有关专业的研究生.本书的宗旨是以讲授实用的典型方法为主,而又不回避必要的理论介绍.初学者在读完本书大部分章节后,若能独立进行数据处理和分析,又在道理上知其所以然,本书的目的就算达到了.为了便于教学和科技人员

参考,我们把方法与理论证明分别相对地作了集中。当然,如果将本书的理论较深的章节充分讲授或自学,将使读者能更深刻地理解时间序列的理论背景,也可作为其理论与应用进一步研究的导引。

本书共分九章:第一、二章介绍平稳序列和线性模型的基础理论,作为以后各章的理论基础和概率论的背景材料;第三章讲授 ARMA 模型参数估计与建模方法,是时序方法的代表章节,而频域方法的代表章节则放在第七章,讲述谱估计的种种方法;时间序列应用的重要目的之一是预报,放在第四章;非平稳序列、非线性模型和多维平稳序列等则安排在第五、六、八章讲授;渐近理论作为前述各章给出的各种实用性方法的理论保证,安排在本书最后一章,由于它侧重于理论推理,初学者只需了解其有关结论。某些对时间序列分析研究有兴趣的读者如能详细阅读第九章,必将是有益的。全书仅需读者熟悉数学分析、线性代数和概率统计知识即可顺利阅读。

本书吸收了某些能使本科生易于接受的最新成果,更新了某些旧的方法,丰富了一些基本理论,例如 ARMA 模型参数估计的 H-R[33] 和王一陈[34]方法,黄大威方法[35,36],MA 模型参数的逆自相关函数估计方法,渐近理论中安鸿志等[24,68]和黄大威[4]的一些深刻结果。此外,书中还包括了作者的部分研究结果,比如带干扰噪声 AR 过程的参数估计[38]和逆自相关函数估计的渐近性质[53,60]。当然,以上种种只是初步介绍,其详细理论可参阅有关文献。

本书由常学将、陈敏、王明生编写(其中王明生编第一、四、五章,陈敏编第二、三、六、九章,第七、八章由常学将编),由陈敏统一整理修改,常学将最后定稿。编写过程中参考或使用了国内学者、同行的有关著作和论文,已在参考文献中一一列出,在此一并表示感谢。

1992年8月在山西大学召开了本书的审稿会,审稿会邀请邓

集贤教授、韩志刚教授、朱燕堂教授、吴诗咏教授、范金城教授和高尚华副编审参加，由邓集贤教授主审。与会专家认真审阅了全书，对全书的选材及细节提出了十分有益的修改意见，在此基础上我们又对全书作了修改。在本书出版之际，我们向各位专家、教授深表谢忱。由于我们水平所限，书中难免存在缺点和错误，恳请读者批评指正。

编 者

1993.1

DAAZ1 / 04

目 录

第一章 时间序列与随机过程	1
§ 1.1 什么是时间序列	1
§ 1.2 随机过程的基本概念	9
§ 1.3 宽平稳过程	17
§ 1.4 平稳序列的谱表示	26
§ 1.5 平稳序列的遍历性	35
§ 1.6 线性最小方差估计	44
§ 1.7 线性差分方程	47
习题.....	51
第二章 线性过程	54
§ 2.1 噪声过程与线性过程	54
§ 2.2 MA(q) 模型	65
§ 2.3 AR(p) 模型	73
§ 2.4 ARMA(p, q) 模型	89
§ 2.5 线性过程的逆自相关函数及其性质	103
习题.....	110
第三章 ARMA 模型的建模	113
§ 3.1 线性平稳时间序列自协方差函数和自相关函数的估计.....	113
§ 3.2 模型的初步识别	119
§ 3.3 AR 模型的参数估计	125
§ 3.4 AR(p)模型的定阶.....	143
§ 3.5 MA 模型的参数估计和定阶	162
§ 3.6 ARMA 模型的参数估计和定阶	167
§ 3.7 带有干扰噪声的 AR 模型的建模	191
§ 3.8 疏系数 AR 模型的建模	196

习题	201
第四章 平稳时间序列的预报	204
§ 4.1 最优线性预报	204
§ 4.2 线性过程的预报	208
§ 4.3 ARMA 序列的平稳预报	212
§ 4.4 ARMA 序列的新息预报	225
§ 4.5 谱已知的预报问题	237
习题	248
第五章 非平稳时间序列的 ARIMA 模型	251
§ 5.1 ARIMA 模型	251
§ 5.2 季节性 ARIMA 模型	257
§ 5.3 ARIMA 模型建模	259
§ 5.4 ARIMA 模型的预报	272
§ 5.5 长记忆过程	278
习题	291
第六章 非线性模型与混合回归模型	293
§ 6.1 引言	293
§ 6.2 非线性模型的特点及其检验	295
§ 6.3 门限自回归模型	298
§ 6.4 双线性模型	310
§ 6.5 混合回归模型	317
习题	324
第七章 谱分析	327
§ 7.1 谱估计的周期图方法	327
§ 7.2 最大熵谱估计与其它谱密度估计方法	360
§ 7.3 离散谱的检测与估计	386
§ 7.4 周期采样与随机采样	407
习题	416
第八章 多维时间序列介绍	421
§ 8.1 多维线性平稳序列	421
§ 8.2 多维 AR(p) 模型的建模	428

§ 8.3 多维平稳序列的谱估计	434
§ 8.4 多维时间序列的主成分分析	439
§ 8.5 多维时间序列的典型分析	446
习题	451
第九章 时间序列分析的估计理论	453
§ 9.1 预备知识	453
§ 9.2 自协方差函数和自相关函数估计的相合性	462
§ 9.3 样本自协方差函数和样本自相关函数各阶矩的渐近 性质	467
§ 9.4 样本自协方差函数和样本逆自协方差函数的渐近分 布	471
§ 9.5 样本自协方差函数和样本逆自协方差函数的一致收 敛速度	483
§ 9.6 AR(p) 模型参数估计的渐近性质	486
§ 9.7 关于 AR 模型定阶方法的强相合性	495
§ 9.8 ARMA 模型极大似然估计的渐近性质	505
§ 9.9 其它估计量的渐近性质	514
习题	520
附录 判别高阶多项式的根是否在单位圆外的 Jury 准则	524
附表	525
参考文献	532

第一章 时间序列与随机过程

§ 1.1 什么是时间序列

所谓时间序列就是按时间次序排列的观测值集合。按照研究的现象或问题的不同，可以得到各种时间序列。例如经济学家观察某种物价指数波动，气象学家研究某地旬降雨量，电气工程师研究电子接收机的内部噪声等都会观测到依某种度量单位测量的一串数据，其自然顺序就是按出现的时间先后排列次序而得到的时间序列。在实际中获得的时间序列大致有两种类型，一是依连续时间进行观测或记录，一是依离散时间进行观测或记录。我们以后主要涉及离散情况，对依连续时间变化的现象，由于计算机数据处理的需要，要进行离散“采样”以得到观测值，比如对连续观测的噪声信号依等时间间隔取值；或是按一定的时间周期来获得观测值，比如某地旬降雨量。这里观测值的取得是随着研究者的兴趣所在而有各种不同的取法，离散时刻的取法也是随研究者的需要而定。比如，研究旬降雨量，可以每 10 天取一个累积值，而研究噪声信号，则未必需要等时间间隔取观测值。

把依时间变化的观测值记为时间 t 的函数 x_t ，时间的变化范围可以是有限集 $T = \{t_1, \dots, t_n\}$ ，可以是可列集 $T = \{t_1, t_2, \dots, t_n, \dots\}$ ，也可以是在某实数区间 $T = \{t: a \leq t \leq b\}$ ($-\infty < a < b \leq \infty$) 连续取值，但对这种情况，需要进行离散采样，通常按某种原则在有限个时刻 t_1, t_2, \dots, t_n 处取得相应的时间序列 $x_{t_1}, x_{t_2}, \dots, x_{t_n}$ 。最简单的采样方法是等时间间隔采样，称为等距采样，即令 $t_i = ih, h > 0, i = 1, 2, \dots, n$ ，在时刻 t_i 处取观测值 $x_{t_i}, i = 1, 2, \dots, n$ 。但需要指出的是，在采样过程中，由于我们

只是在特定的时刻 t_1, t_2, \dots, t_n 处取观测值, 这自然会有一部分信息在采样过程中损失掉. 显然, h 愈小, 损失的信息就愈小, 但却加大了计算机存储单元和运算工作量. 还有所谓随机采样等, 这些内容将在第七章中讨论.

如果时间序列 $\{x_t\}$ 是某确定性函数在 t 时刻的观测值, 比如 $x_t = \cos(\alpha t + \beta)$, α 与 β 是已知常数, 则称 x_t 为确定性时间序列. 如果时间序列 $\{x_t\}$ 不但是时间 t 的函数, 而且还是服从某种概率分布的随机序列, 则称 $\{x_t\}$ 是随机性时间序列, 这是本书研究的对象. 对于随机性时间序列 $\{x_t\}$, 在 x_t 未抽出之前 (未到时刻 t), 只知道 x_t 以某种分布取各种值, 不能断言必取某值, 而抽出之后 (即已到达时刻 t), 则取得该随机变量 x_t 的一个样本值. 因此, 我们所获得的时间序列 $\{x_t\}$ 实质上是一族随机变量的样本. 故严格地讲, 应定义时间序列 $\{x_t\}$ 是某个随机过程的一个样本 (或现实). 这样, 对时间序列的理论研究是以对应的随机过程的研究为基础的. 由于以后的内容只涉及参数空间为离散的随机过程, 故主要讨论随机序列的统计分析. 当讨论时间序列 $\{x_t\}$ 的概率结构时, 常使用随机序列这一术语, 而讨论模型参数的统计推断时, 则使用术语时间序列, 以示 $\{x_t\}$ 是随机序列的样本.

下面我们来举一些时间序列分析中的典型例子, 它们都是来自于实际问题.

例 1.1.1 天文上的太阳黑子活动. 太阳黑子的活动直接影响着地球气候的变化, 引起旱、涝灾害. 太阳黑子聚集到一定程度会产生耀斑, 释放出的大量能量所产生的电磁辐射和粒子辐射传播到地球可造成无线电通讯突然中断, 其高能粒子还威胁着宇航飞行器及宇航人员的安全. 图 1.1.1 是沃尔夫 (Wolf) 年平均太阳黑子数 (1749—1924) 的数据图, 确切数据见本书附表 1.

历史上最早用时间序列研究这一列数据的是尤尔 (Yule), 他于 1927 年提出了 AR 模型, 用这一模型, 根据沃尔夫太阳黑子数的统计分析来描述太阳黑子活动情况. 50 年代以来, 这方面的

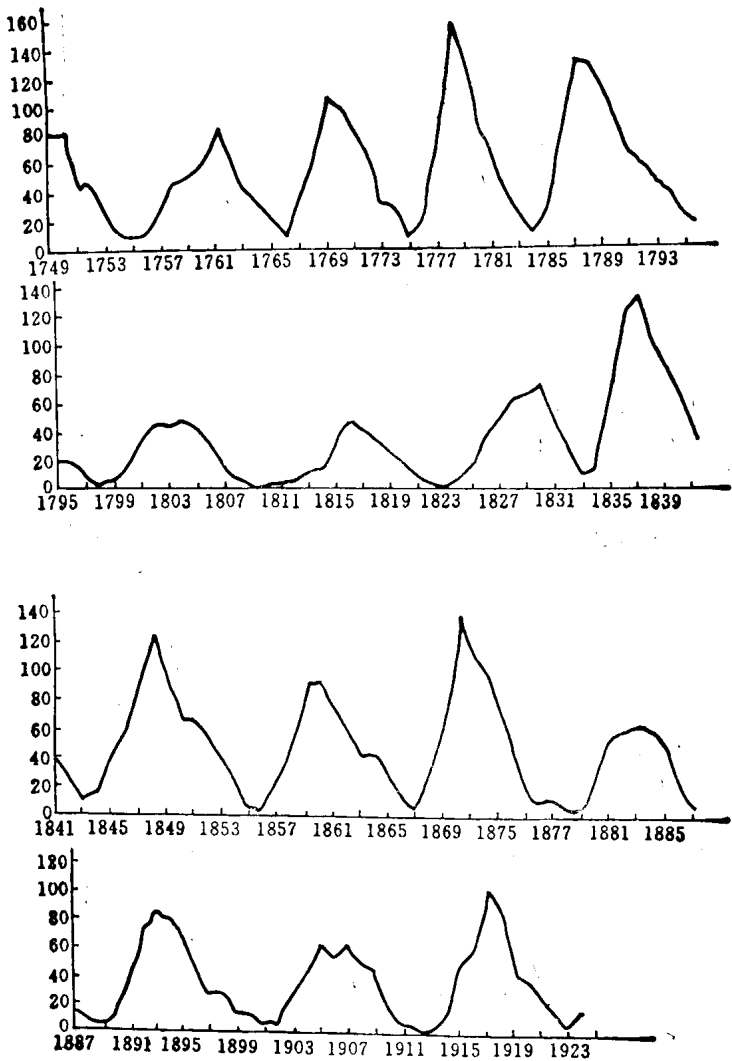


图 1.1.1 沃尔夫的太阳黑子数

有关研究文献急增,其中以鲍克斯与詹金斯 (Box and Jenkins)、汤家豪的工作较有影响。

例 1.1.2 加拿大马堪泽河流域山猫捕获量。图 1.1.2 是这种山猫每年捕捉量的 114 个数据 (1821—1943 年), 图中所画的是以 10 为底取对数后的数据值。确切数据见附表 2。像太阳黑子数据一样, 这组数据也是时间序列分析中有名的数据。各家新创的方法都以能更精确地预报这组数据作为方法优越性的凭证。

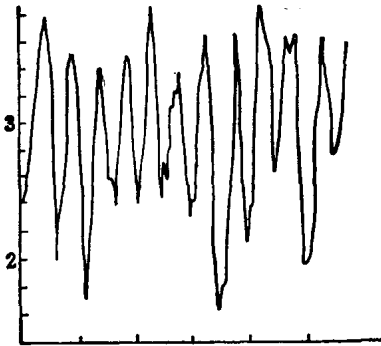


图 1.1.2 加拿大山猫年捕获量

种山猫每年捕捉量的 114 个数据 (1821—1943 年), 图中所画的是以 10 为底取对数后的数据值。确切数据见附表 2。像太阳黑子数据一样, 这组数据也是时间序列分析中有名的数据。各家新创的方法都以能更精确地预报这组数据

数据作为方法优越性的凭证。

例 1.1.3 国际航线旅客客票数据。图 1.1.3 给出某国际航空公司 1949—1960 年间客票月总数 (单位: 千张) 的时间序列曲线。直观上看, 每年有一次大的峰值和一次小的峰值, 并且逐年不断增长。

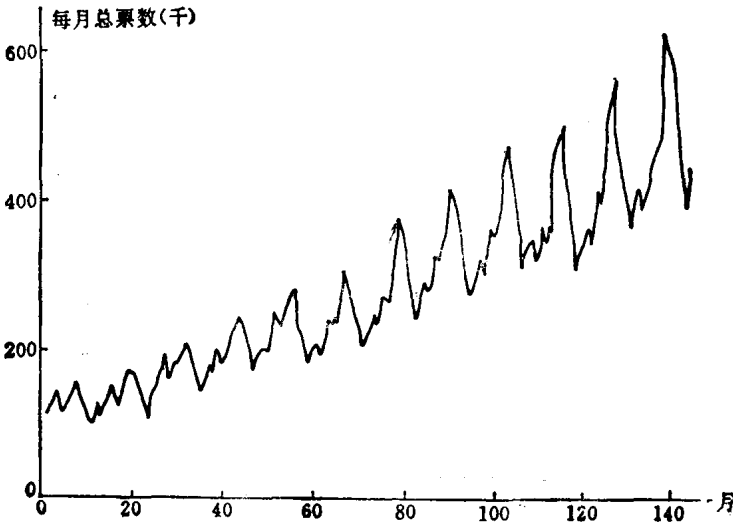


图 1.1.3 国际航线客票(单位: 千张)

加。

图 1.1.4 是我国铁路客流量的统计曲线，记录了 1971—1981 年客票月总数。从铁路客流量的时间序列曲线上可见，每年都有一次较大的峰值，大约是在 1、2 月份，也就是每年的春节前后有一次最大的峰值。时间序列分析的任务就是根据以往的数据找出变化规律，预报将来的客流量。

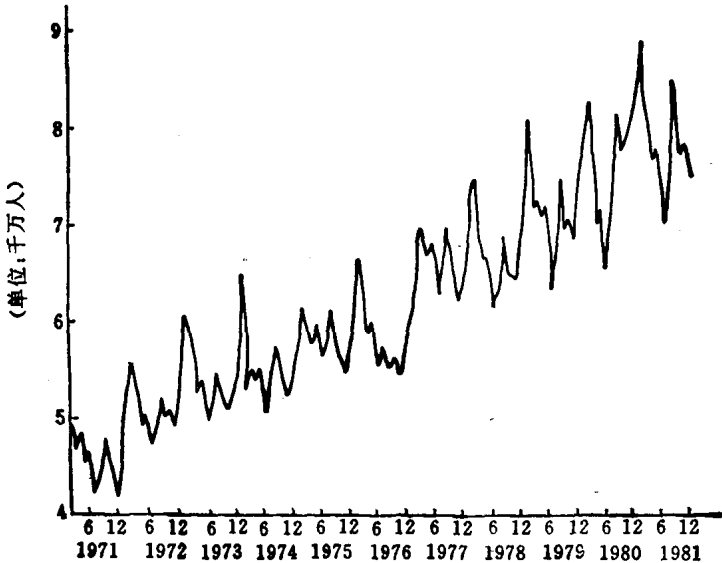


图 1.1.4 铁路客流量统计曲线(单位:千万人)

例 1.1.4 图 1.1.5 和图 1.1.6 给出了美国从 1965 年 7 月—1970 年 4 月的每月零售价格指数 (CPI) 和批发价格指数 (WPI) 的统计曲线。数据的曲线清楚地显示出一种趋势。时间序列分析的任务就是对原始数据进行细致的分析,找出其蕴含的趋势(确定性趋势或季节性趋势),对数据建立适当的时间序列模型。

例 1.1.5 图 1.1.7 是一条脑电图数据曲线。脑电图可用来判断人体的很多特征,比如根据脑电图可以判断一个人属正常智力还是愚型。脑电图是一条随时间连续变化的曲线,可按前面所述

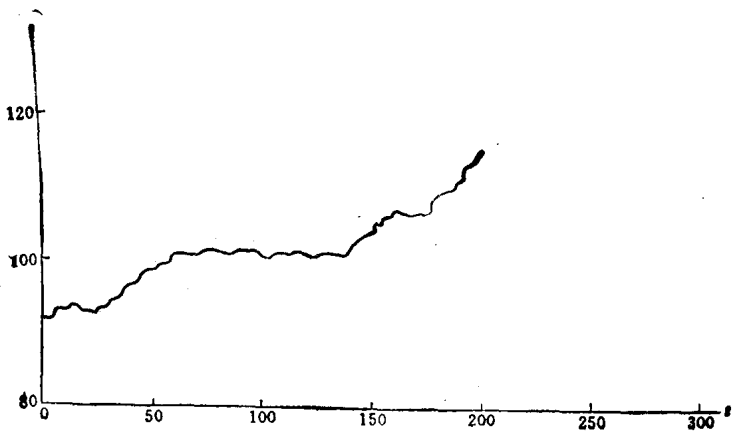


图 1.1.5 美国的每月零售价格指数 (CPI)

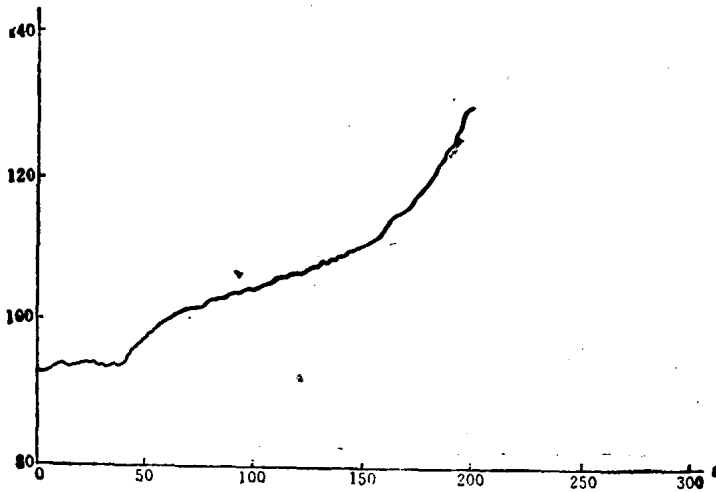


图 1.1.6 美国的每月批发价格指数 (WPI)

的离散化采样方法获得时间序列, 进而通过分析所得时间序列来推断人体的健康状况. 关于这方面的研究可参见[46].

例 1.1.6 图 1.1.8 是某地按小时的平均用电负荷曲线, 以 x_k

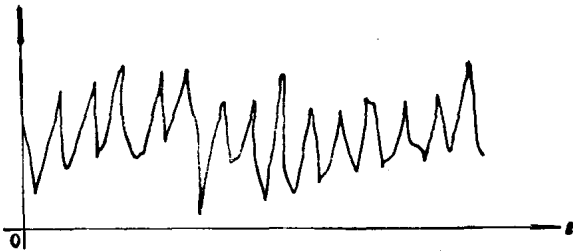


图 1.1.7 脑电图曲线

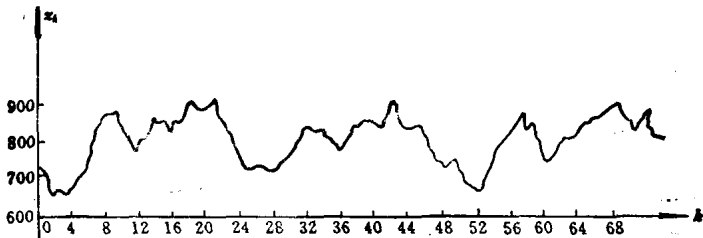


图 1.1.8 某地按小时平均用电负荷曲线

表示第 k 小时某地平均用电负荷量。利用现在和过去时刻的用电时间序列,可以预报未来某时刻的用电量,有关讨论可见[47]。

例 1.1.7 图 1.1.9 是我国地震局沈阳台记录的主应力方向变化曲线,每旬一个数据。从图中可见,在 1975 年初的海城地震和 1976 年夏的唐山地震时,主应力方向都有突变现象。这就启发我们,可根据数据分析主应力方向的变化,为预报地震提供一种方法。

例 1.1.8 图 1.1.10 是地球地极运动变化曲线。现代天文测量技术发现,地球围绕地轴自转的速率不是一个常数,地极(地轴在南北极和地表面的交点)并不固定在一点上,而是在一定的范围内摆动。如果把两个相差 90° 的经度方向作为 x 轴和 y 轴,则地极的位置可由在这两个数轴上的坐标来确定。每隔一定时间测一

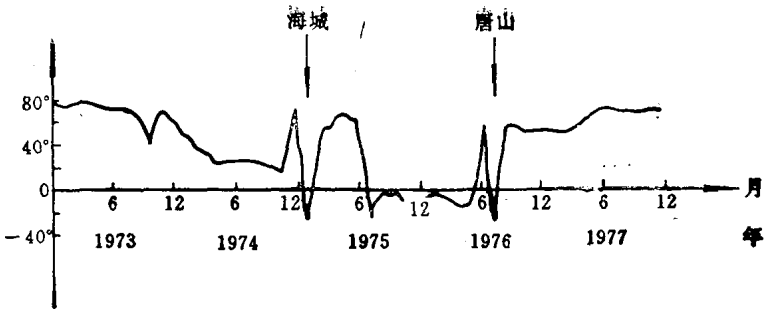


图 1.1.9 沈阳站记录的主应力方向

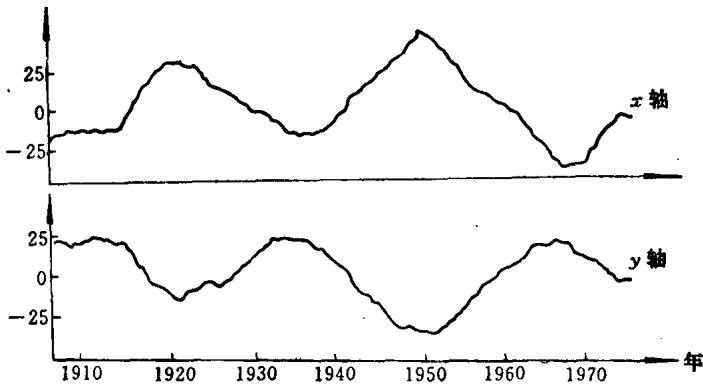


图 1.1.10 地极运动变化曲线

次,便得到两条不同坐标的时间序列,研究这种时间序列的变化规律对飞船的定位跟踪、大地测量及地震监测都有重要的意义.我国天文工作者在这方面作了很多工作,比如[48].

从上面的应用实例可见,时间序列分析在自然现象的研究、工业、医学、生物以及经济管理等领域都有应用,不过这并非包括了所有的应用方面.时间序列分析的目的在于处理和分析数据,以揭示客观现象的本质和内在规律性,达到有效地对客观现象及其变化规律进行预报和控制.在时间序列的统计分析中,常用所谓“时域分析”方法和“频域分析”方法去处理数据.前者是从时间结构