



杨位钦 顾 岚 编著

时间序列分析与动态数据建模

(修订本)



北京理工大学出版社

时间序列分析 与动态数据建模

(修订本)

杨位钦 顾 岚 编著

北京理工大学出版社

内 容 简 介

本书首先讨论采样数据的检验和预处理,然后介绍平稳随机过程的基本特征及其模型。以线性ARMA模型为主,随后两个重点部分是时间域模型和频率域模型的估计方法,包括极大熵谱估计。在建模的基础上介绍了几种时间序列的预报方法。除单变量模型外,本书也讨论多变量的时域和频域模型及其估计,对趋势性、季节性、混合回归、疏系数、门限回归、双线性、指数等特定形式的模型也做了不同程度的讨论,最后一章是关于自适应AR模型的在线参数估计方法。

附录中提供了可在IBM-PC/XT微机上运行的FORTRAN程序及使用说明。本书可作为大学生、研究生的教材和有关科技人员的参考书。

时间序列分析与动态数据建模

(修订本)

杨位钦 顾 岚 编著

*

北京理工大学出版社出版

新华书店北京发行所发行 各地新华书店经售

国防科工委印刷厂印刷

*

787×1092毫米 32开本22.125印张 570千字

1988年12月第二版 1988年12月第一次印刷

ISBN7-81013-115-X/TP·11

印数: 1—3000册 定价: 5.40元

符 号 说 明

如果正文中没有另外的说明，下列符号的含义为

$\{\cdot\}$ ——集合。

$O[\cdot]$ ——与 $[\cdot]$ 内的量同阶。

$\equiv$ ——定义式或恒等式。

$\exp(\cdot)$ ——等于 $e^{(\cdot)}$ ，其中 e 为自然对数的底

$\chi^2_a(n)$ ， $\chi^2_{n;a}$ —— χ^2 变量，自由度为 n ，显著水平为 α 。

$F_a(v_1, v_2)$ ， $F_{v_1, v_2; \alpha}$ —— F 分布变量，自由度为 v_1 和 v_2 ，
显著水平为 α 。

$NID(\cdot, \cdot)$ ——独立正态分布， $(\cdot, \cdot)$ 中数值依次为均值与方差。

$\mathcal{N}(\cdot, \cdot)$ ——正态分布， $(\cdot, \cdot)$ 中数值依次为均值与方差

$E[\cdot]$ ， $\text{Var}[\cdot]$ ， $\text{Cov}[\cdot]$ ——分别为 $[\cdot]$ 中所指的量的期望、方差和协方差。

$S.E.[\cdot]$ —— $[\cdot]$ 中所指的量的标准差

$[\cdot]$ ——对某一量取整数值（舍弃非整数部分）

$*$ ——用作上标时表示取共轭复数。并列在两个函数中间时表示卷积运算。

$\det$ ——行列式。

$\hat{x}$ ——表示 x 的估计。

黑体（如 $\mathbf{x}$ ）表示该量为向量或矩阵。

第二版前言

自1986年第一次出版和1987年第二次印刷以来，这是本书的第二版。这次修订根据学科的发展和使用本书所提出的要求，增删了部分内容。较明显的是增加了第九章自适应模型及算法，这对实时建模和预报是比较实用的。另外加了一些习题。几乎每个章节都有不同程度的修改，但总的体系不变。计算机程序删掉一个，新增三个，并对部分程序作了修改。

我们希望这个修订版能够对读者的帮助效果更好些。

编者 1988年5月

前 言

对生产和科学研究等过程中的某一个变量或一组变量 $x(t)$ 进行观察测量,在一系列时刻 $t_1, t_2, \dots, t_n$ (t 为自变量且 $t_1 < t_2 < \dots < t_n$)得到的离散的有序数集合 $x(t_1), x(t_2), \dots, x(t_i), \dots, x(t_n)$ 称为离散数字时间序列(本书只讨论离散情况,简称为时间序列)。在实际问题中,自变量 t 除代表时间外,也可具有各种不同的物理意义,例如代表长度、温度、速度或其它单调递增地取值的物理量。

气象上的月降水量序列,水文上的河流流量序列,海洋上的海浪序列,天文上的太阳黑子序列,地球物理上的地震波序列,医学上的脑电波序列,机械系统的振动序列,雷达系统对活动目标的跟踪量测序列,化工反应系统的输入输出函数序列,自然界某种生物总数的消长序列,商业经济方面的产值、销售额的序列等等都是时间序列的具体例子。

用来分析各种相依有序的离散数据集合的整个方法称为时间序列分析。一般说来,时间序列很难用一个完全确定的函数或函数组给出来,它们大都具有统计规律性,可以通过概率分布函数或函数组对它们取值的规律性作统计描述,这种类型的时间序列称为随机的。本书主要讨论时间序列的统计分析和模型构造问题。

在生产、科研、社会生活和国防建设中,时间序列分析有着广泛的实际应用,其中主要包括:

1. 分析时间序列的统计规律性,推断产生时间序列的物理系统的性质,找出它的物理规律。

2. 根据对时间序列统计规律性的分析, 构造拟合它的最佳数学模型, 浓缩时间序列的信息, 简化对时间序列的表示。

3. 利用拟合的数学模型预报时间序列未来的可能取值, 给出预报结果的精度分析。

4. 根据拟合的数学模型模拟出时间序列, 用于分析和优化处理。

时间序列分析属于统计学科的分支, 近十多年来的发展很活跃, 原有的方法有很多新进展, 本书在介绍时序分析的基本原理和方法的时候注意具体应用, 为便于读者进行时序分析和建模的实践, 在附录中提供了能在 IBM-PC/XT 微机上使用的 FORT-RAN 程序, 书中的计算例子除某些引用的结果外都是按照这些程序进行处理的。应当说明, 虽然提供的程序具有一定的实用性, 但在编写上并非都是最优的。

本书的编写得到了北京工业学院张志方教授的热情鼓励和支持, 并承蒙中国科学院应用数学所陈兆国副研究员主审, 提出了宝贵的意见, 我们在此表示深切的谢意。

编者

一九八六年三月

目 录

前言

第一章 采样数据的检验和预处理	1
1.1 采样间隔和频率混叠	1
1.2 均值、方差和概率直方图	6
1.3 随机数据的正态性检验	13
1.4 随机数据的独立性检验	19
1.5 非平稳趋势的检验	20
1.6 剔点处理	23
1.7 提取趋势项	26
1.8 随机数据的周期性检验	31
第二章 平稳随机过程及其模型	35
2.1 平稳随机过程	35
2.2 自协方差和自相关函数	37
2.3 典型离散参数模型	43
1. 纯随机过程——白噪声 2. 一阶自回归过程 3. 二阶自回归过程 4. n 阶自回归过程 5. 滑动平均过程 6. 自回归滑动平均过程, 格林函数和逆函数	
2.4 平稳随机过程的频率域表示	67
1. 频谱分析的基本概念 2. 连续过程的功率谱和自协方差函数的关系 3. 离散过程的功率谱和自协方差函数的关系 4. 基本线性模型的功率谱 5. 平稳过程的谱展式	
第三章 时间域模型的估计	89
3.1 自协方差和自相关函数的估计	89
1. 两种估计方法 2. 关于协方差和自相关估计的Bartlett公式	
3.2 模型参数的相关矩估计	98
1. AR模型参数的矩估计 2. MA模型参数的矩估计 3. ARMA模型参数的矩估计	

3.3 模型参数的最小二乘估计 (LS估计)	107
1. 最小二乘估计 2. AR模型参数的LS估计 3. ARMA模型参数的LS估计	
3.4 模型参数的极大似然估计 (ML估计)	121
1. 极大似然估计 2. AR、MA和ARMA模型参数的极大似然估计	
3.5 模型阶数的确定	129
1. 残差方差图 2. 偏相关函数和模型 3. F检验用于模型定阶 4. FPE, AIC, BIC准则 5. 隅角法定阶	
3.6 时间序列建模的基本步骤	155
1. Box-Jenkins方法 2. Pandit-Wu方法的建模 3. 长自回归白噪声建模方法 4. 参数初始估计的逆函数法 附录: 隅角定理证明	
第四章 周期图与加窗谱估计	174
4.1 隐周期的估计	174
1. 周期图分析 2. 周期图的样本统计特性 3. 周期图的峰值检验 4. 非白噪声背景下的周期检验	
4.2 功率谱密度的周期图估计	190
1. 修正周期图与功率谱估计 2. 样本周期图的方差	
4.3 功率谱估计的两种基本方法	197
4.4 窗函数	208
1. 窗函数与谱的分辨力和泄漏 2. 几种常用的窗函数	
第五章 极大熵谱估计	226
5.1 谱熵和极大熵准则	226
1. 问题的提出 2. 高斯过程的熵和熵率 3. 功率谱和熵率的关系	
5.2 极大熵谱估计及BURG算法	232
5.3 极大熵谱估计的LS-LUD	241
第六章 时间序列的预报	251
6.1 平稳线性最小方差预报	251
6.2 最小方差预报的性质和方法	255
6.3 时间序列的新息实时预报	265
1. 新息预报的原理 2. 新息定理及实时预报	
第七章 多变量时间序列	273
7.1 多变量平稳过程的相关和谱特性	273

1. 双变量过程的相关特性 2. 双变量过程的谱特性 3. 双变量过程举例 4. 一般多变量过程	
7.2 具有线性关系的多变量过程谱分析	286
1. 严格线性关系的情况 2. 附加噪声时的线性关系	
7.3 互谱特性和相干函数的估计和应用	296
7.4 多变量过程的时间域模型	304
1. 多变量过程的矩阵差分方程表示 2. 多变量自回归模型的参数估计 3. 多变量自回归模型的FPE准则定阶 4. AR模型参数矩阵的计算方法与建模实例	
第八章 一些特定形式的模型	324
8.1 趋势性和季节性模型	324
1. ARIMA模型 2. 乘积型季节性模型 3. 组合模型	
8.2 混合回归模型及疏系数模型	348
1. 一般线性回归模型及参数的线性最小二乘估计 2. 混合回归模型 3. 疏系数模型 4. 实例分析	
8.3 门限回归(自回归)模型	375
1. 门限回归(自回归)模型的定义 2. 门限自回归模型的特性 3. 门限回归(自回归)模型的建立 4. 门限回归(自回归)模型的预报 5. 实例分析	
8.4 双线性模型和指数自回归模型	394
1. 双线性模型 2. 指数自回归模型	
第九章 自适应模型及估计的递推算法	400
9.1 自适应模型的特点和建模方法	400
9.2 最小均方误差准则和基于梯度估计的算法	402
1. 最小均方误差准则和Wiener最优滤波器 2. 最速下降法求解及其渐近性质 3. LMS算法与自适应AR模型	
9.3 Kalman滤波及其在自适应建模中的应用	417
1. Kalman滤波原理及状态估计 2. 基于一歩预报的Kalman滤波公式 3. Kalman滤波算法应用于自适应AR模型	
9.4 基于最小二乘准则的自适应递推算算法	432
1. 加权最小二乘准则及递推最小二乘(RLS)算法 2. RLS算法的收敛性 3. 线性预报和自适应AR模型的RLS算法	
9.5 递推最小二乘估计的格形算法结构	443

1. 预备知识 2. 最小二乘格形算法 (LSL算法)

习 题	462
附录一 时间序列分析与建模程序说明	474
1. 随机过程数据产生程序 (DAGENT)	474
2. 样本均值和样本方差的计算 (SMSV)	477
3. 直方图的计算 (PDFH)	478
4. 随机数据的独立性检验 (TESTI)	480
5. 随机数据中的隐含周期检验 (TESTP)	481
6. 多项式的提取 (POLYTR)	489
7. 逆序检验 (INORD)	490
8. 加窗最小二乘平稳化 (WINDLS)	494
9. 格林函数的计算 (GREENF)	496
10. 逆函数的计算 (INF)	497
11. 朱利准则 (JURYC)	498
12. 理论自相关函数的计算 (AUTOCR)	499
13. 偏相关函数的递推计算 (PARCR)	502
14. ARMA 谱的计算和 (ARMASP)	503
15. 样本自相关和互相关的快速计算 (CORELA)	505
16. 逆函数法估计模型参数的初值 (INGUS)	509
17. ARMA 模型参数的矩估计 (ARMAME)	511
18. 线性最小二乘法 (LSME)	513
19. Marquardt 阻尼最小二乘法 (MARQT)	514
20. 直接法的谱估计 (PMPSD)	520
21. 相关函数加窗谱估计 (CMPSD)	525
22. 极大熵谱估计的伯格算法 (MEBURG)	529
23. 预报的逆函数法 (INVERF)	533
24. 预报矢量法 (VECTF)	535
25. 多变量自回归模型的参数估计 (MVAR)	537
26. 混合回归模型及疏系数模型的估计 (SMLR)	542
27. 门限回归模型的估计 (TREG)	546
28. 双线性模型的估计 (BILIM)	548
29. 指数模型的估计 (EXPOM)	550

30. LMS算法自适应AR模型参数估计 (LMSAR).....	551
31. RLS算法自适应AR模型参数估计 (RLSAR)	553
32. LSL算法自适应AR模型参数估计 (LSLAR)	555
附录二 程序文本	557
附录三	671
1. 标准正态分布的累积分布函数表.....	671
2. χ^2 分布表.....	676
3. F 分布表.....	679
4. 调和分析中显著性费歇检验 (Fisher) 检验表.....	684
参考资料	690

第一章 采样数据的检验和预处理

在实际问题中,测量记录常常是时间的连续函数,为了数字计算处理上的方便,往往只按一定的时间间隔取值,怎样合理地选择采样间隔,使得采样所得的序列能如实地代表原来的时间序列,这是数据采集首先遇到的一个重要问题。

在得到大量的量测数据之后,还要进行初步的整理和必要的检验,以期去粗取精,为进一步深入分析提供较好的依据和参考。这些工作都属于预处理。

1.1 采样间隔和频率混叠

设记录到的连续时间 t 的函数为 $x(t)$ ($t_0 \leq t \leq t_0 + T$), 这里只讨论在等间隔上取值的时间序列

$$x_1, x_2, \dots, x_i, \dots, x_N \quad (1-1-1)$$

其中 $x = x(t_i) = x(t_0 + i\Delta)$ 。

在时间序列 $x(t)$ 的采样过程中,取不同的采样间隔 Δ 可以得到不同的数字时间序列。这种离散的序列抛弃了 $x(t)$ 在 $t_0 + (i-1)\Delta$ 和 $t_0 + i\Delta$ 之间的值,即这部分信息在采样后丧失了。显然, Δ 愈小,(1-1-1)所示序列和 $x(t)$ 的差别就愈小,信息损失就愈少。但随着 Δ 的减小,采样和统计处理的数据量 $N = T/\Delta$ 也就愈大,需要的计算机存储单元和计算时间(如建模所需的时间)也就增加了,所以采样间隔不能无限制地减小。

Shannon采样定理给出了理论的采样时间间隔,规定了能从采样数据复现原来信号所必须的最小采样频率(或最大采样间隔)。尽管由于很难事先确知时间序列的总体频谱特性,我们还

是可以根据经验、直观和实际的物理要求，参照采样定理，在不过分减小信息损失和不过分增加数据量之间作出合理的选择。

在反映时间序列的原有信息方面，有一个频率混叠现象是和 Δ 有直接关系的。问题的提出是：在给定的 Δ 下，离散序列 x_i 的频率分析能不能得出 $x(t)$ 原来具有的频率分量？

以对连续的谐波信号进行等间隔采样为例，可以说明频率混叠现象。图 1-1 中的实线为原来的连续信号，而虚线是根据采样点连接起来的。

从这个图可以归纳出离散化采样处理所产生的数字时间序列的某些性质：

(1) 如果实际信号频率小于或等于 $1/(2\Delta)$ Hz 即采样频率之半，则由采样点反映的频率和实际信号频率相同。(见图 1-1a, b, c, d)。

(2) 如果实际信号频率较高 (图 1-1e, f, g, h, i)，则根据采样点来“重建”原来信号时，就会得到比实际频率低的波形 (在 f 和 i 的情况下，甚至得到的不是正弦波而是直流信号)。

(3) 采样间隔 Δ 确定了频率分析中的最高分辨频率。直观地看，要确定出一个周期所需要的最少信息是三个采样点，即 2Δ 的时间区间，也就是说，以 Δ 为间隔采样时，能够检测出的最短周期为 2Δ 。或者说，能够分辨的最高频率是 $1/(2\Delta)$ 。这个极限频率称之为乃奎斯特 (Nyquist) 频率。它也等于采样频率 $f_s=1/\Delta$ 的一半。

上述性质可以从正弦信号的周期性得出，设

f_c = 连续信号的真实频率

f_d = 离散化信号的频率

则离散化的采样信号 $\cos 2\pi f_d t$ 可表为

$$\begin{aligned}\cos 2\pi f_c t \big|_{t=k\Delta} &= \cos 2\pi f_d k\Delta \\ &= \cos 2\pi (1/\Delta \pm f_d) k\Delta \\ &= \cos 2\pi (2/\Delta \pm f_d) k\Delta\end{aligned}$$

$$= \cos 2\pi (3/\Delta \pm f_d) k\Delta \dots\dots$$

$$(k=0, 1, 2, \dots)$$

$$(1-1-2)$$

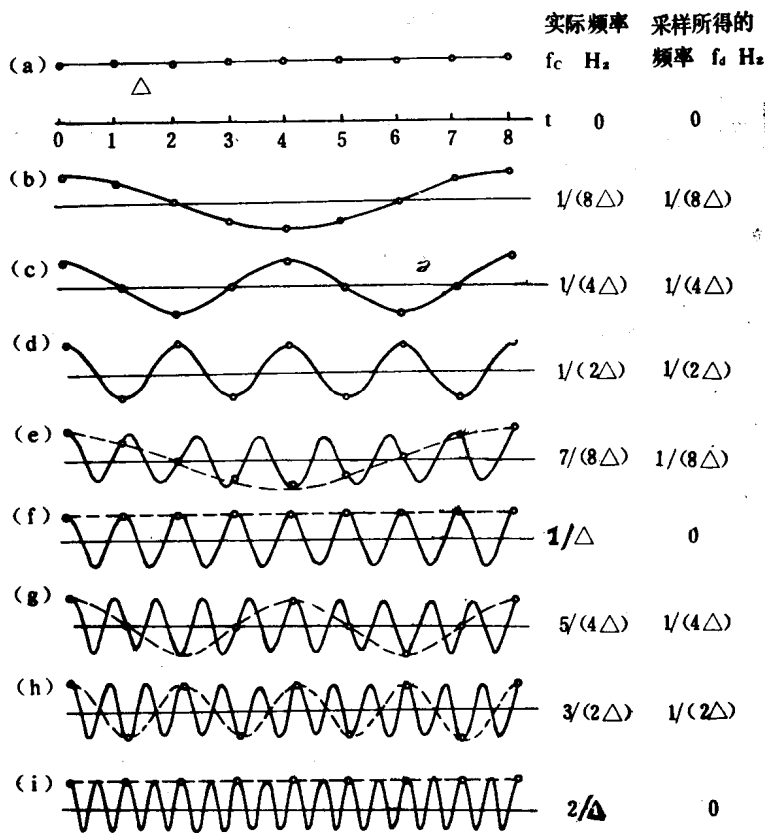


图1-1 正弦信号的频率混叠现象

可见，对于同样的采样信号频率(f_d)，可以有许许多多不同频率(f_c)的连续信号与之对应。例如若 $f_d = \frac{1}{8\Delta}$ ，则可有

$$f_c = \begin{cases} 1/\Delta \pm f_d = \begin{cases} 1/\Delta - 1/(8\Delta) = 7/(8\Delta) \\ 1/\Delta + 1/(8\Delta) = 9/(8\Delta) \end{cases} \\ 2/\Delta \pm f_d = \begin{cases} 2/\Delta - 1/(8\Delta) = 15/(8\Delta) \\ 2/\Delta + 1/(8\Delta) = 17/(8\Delta) \end{cases} \\ \dots\dots\dots \end{cases} \quad (1-1-3)$$

这意味着，如果采样得到的信号频率为 $f_d = 1/(8\Delta)$ ，我们就难于断定实际信号的频率究竟是 $7/(8\Delta)$ ，还是 $9/(8\Delta)$ ，还是 $15/(8\Delta)$ …，反过来说，如果实际信号具有 $f_c = 1/(8\Delta)$ ， $7/(8\Delta)$ 等多种频率成份，那么这些频率成份对于构成离散信号中频率为 $1/(8\Delta)$ 的分量都有贡献。即便实际信号并不存在 $1/(8\Delta)$ 的频率，采样信号仍可有该频率成分存在，因为连续信号中的 $7/(8\Delta)$ 和 $9/(8\Delta)$ ， $15/(8\Delta)$ …的频率成分也会在 $f_d = 1/(8\Delta)$ 的采样信号中反映出来。这种在采样数据中不希望有的效应就是“频率混叠”。

$$f_d \leftrightarrow f_c = n/\Delta \pm f_d \quad n=0, 1, 2, \dots \quad (1-1-4)$$

图1-2表示出 f_d 和 f_c 之间的多值对应关系。可以看出，各个 f_c 均对称于 $1/(2\Delta)$ ， $1/\Delta$ ， $3/(2\Delta)$ ， $2/\Delta$ ，…（即 f_{max} ， $2f_{max}$ ， $3f_{max}$ ，…），所以这些频率也称折叠频率。（图1-2a.b.c…和图1-1相对应）。

前已指出，靠无限减小 Δ （即增大折叠频率）是不现实的，它也会使出现频率混叠的范围扩大。通常是利用低通滤波处理的方法来解决频率混叠效应。假设低通滤波器是理想的（图1-3a），则在 $f > f_c$ 以外的频率成分被滤掉了，因此不会在 f_c 范围内引起混叠。不过，实际滤波器频率响应的模 M 是逐渐衰减的（图1-3b）， f_c 以外的信号并没有完全滤掉。通常截止频率 f_c 定义在 $M = 0.707$ 处，而真正被滤掉的频率成分在 $f_c > f_c$ 以外。设图1-3b低通滤波器的标称截止频率为5kHz，用它作为抗混叠滤波器，应当取多大的 Δ ？采样频率取为 $f_s = 2 \times 5 = 10\text{kHz}$ （每秒采10000个数据）行不行？由于5kHz附近高于截止频率的成分在滤波后依然存在，并反映到0—5kHz范围中来（如图1-3c），因此应当增大采样频

率, 例如取 $f_s = \frac{1}{\Delta} = 15\text{kHz}$ (即折叠频率为 7.5kHz , 如图1-3d)。

这时, 只有高于 10kHz 的频率成分才会反映到 $0-5\text{kHz}$ 范围内, 而因大于 10kHz 的成分经滤波器后已被充分衰减了, 因此频率混叠效应被大大削弱。

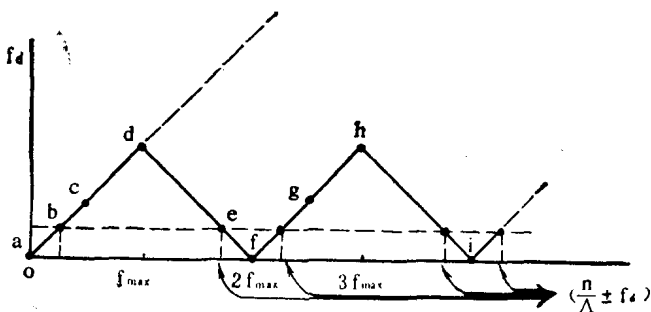


图1-2 连续信号频率与采样信号频率的关系

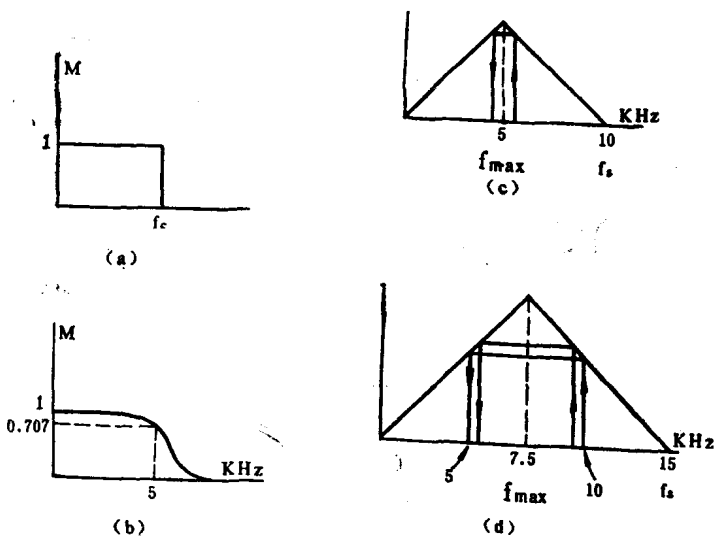


图1-3 滤波器的抗混叠作用