数字时间序列分析

[美] R. K. 奥特内斯 L. 伊诺克森 著

王子仁 马忠安 译

国防工业出版社

数字时间序列分析

〔美〕R.K.奥特内斯 L.伊诺克森 著
王子仁 马忠安 译
汪 浩 王振钧 等校

国防工业出版社

内 容 简 介

全书共十二章，主要介绍了数据的预处理，递归数字滤波，快速傅里叶变换，以及随机量的各种主要函数的算法。第一章复习了数学的预备概念；第二章专述数据的预处理；第三章递归数字滤波；第四章快速傅里叶变换；第五章至第八章介绍了相关函数、功率谱密度和互谱密度的各种算法；第九章为传递函数和凝聚函数的算法；第十章为概率密度函数的算法；第十一章介绍一些非平稳数据的分析方法；第十二章介绍试验与实例。每章末均附有一定数量的习题。本书可供航天、航空、自动化、无线电通信、地震学、海洋学、核过程等领域的科技人员和高等院校师生参考。

DIGITAL TIME SERIES ANALYSIS
R. K. Otnes and L. Enochson
John Wiley and Sons, Inc. 1972

*

数字时间序列分析

〔美〕 R. K. 奥特内斯
L. 伊诺克森 著

王子仁 马忠安 译
汪 浩 王振钧 等校

*

国防工业出版社 出版
新华书店北京发行所发行 各地新华书店经售
国防工业出版社印刷厂印装

*

850×1168 1/32 印张 11 3/8 290 千字
1982年6月第一版 1985年10月第二次印刷 印数：5,201—6,400册
统一书号：15034·2325 定价：2.00元

译　序

近代各项科学试验，特别是大型试验，往往包含着大量的试验数据，如何从这些大量的数据中迅速地获取所期望的结果，实为当务之急。本书系统地介绍了关于平稳随机函数如何进行数据处理问题，着重叙述如何把模拟量形式的时间历程转换成离散形式的数字数据，然后利用所介绍的各种数学方法，在电子计算机上运算，以迅速获取数据结果。本书形式上是数学教材，但又非纯数学式的推导，而是着重于数学方法在实际中的应用。本书介绍用电子计算机进行数字式数据处理的算法，必将有力地推动数据处理工作的前进步伐，同时亦将有利于四个现代化的实现。

本书第一章至第四章及附录A由马忠安同志翻译，第五章至第十二章及附录B由王子仁同志翻译；参加本书校订的有汪浩、王振钧、陈桂彬、邹丛青、刘德铭、李仲哲等同志；在初译期间，胡同庆、张忠翰同志也给予了不少帮助。

由于原书中错误或漏印之处较多，虽经译者、校者尽力校核改正，但肯定仍有不少疏漏未发现。并且由于我们水平有限，加以时间仓促，错误仍然难免，希望读者批评指正。

译　者

原　　序

本书系由作者们在时间序列分析方面，特别是在试验数据分析方面的经验积累而成的。书中所提供的内容业已在实际中加以应用，因而具有经过实践考验的优点。

本书是为了满足两种需要而编写的。首先，它可以作为诸如振动、声学、医学、地震学、通信和海洋学等方面从事数字数据分析的科技工作者的工具书和参考书。

其次，本书中的许多内容也适合作为工程和有关领域中具有大学研究生水平的学生学习时间序列分析的教材。为此，第一章所介绍的内容是复习一下为了学习本书所必需的一些知识。

（下略）

R. K. 奥特内斯

L. 伊诺克森

目　录

第一章 预备概念

1.1 引 言

本章的目的是复习一下本书所用的数学知识，并着重介绍某些数学结论或方法，这对于理解本书内容是必不可少的。其要点如下：

连续型傅里叶变换；

离散型傅里叶变换；

这两种变换间的差别；

矩形函数的傅里叶变换；

平均值、方差和功率的概念；

χ^2分布、正态分布和白噪声。

了解了上述这些概念，对学习本书以后各章将有很大帮助。因此，如果读者不熟悉这些概念，建议读一读这一章，并仔细地研究所举的例题和习题。

本书的主要目的不是讨论数学上的细节。因而，在本章和以后各章中，对许多数学证明仅作概略的叙述或者完全省略掉。如果读者需要了解较完整的数学细节，可以参考汉南（Hannan, 1960 年）、郭和凯泽（Kuo and Kaiser, 1966 年）、或帕仁(Parzen, 1962 年）等人的著作。

数据处理包括通常称之为时间历程的记录数据的处理。一般来说，这种时间历程就是测量仪器的输出，它被记录在某些介质诸如图纸、穿孔纸带或磁带等上面。例如，一种特定股票的每日收盘价格组成的数据记录，就是一个时间历程。按照通常对于时间历程的数学定义来说，股票的价格为因变量，而记载价格的时

间则为自变量。在本书中，最常见的自变量是时间和频率，可能还有许多其他的自变量如位置或速度坐标等，但它们在实用中较少被采用。因此，在数据记录中，不应该认为将时间作为自变量是一成不变的。本书所讨论的方法完全是一般性的，例如可以用距离或者深度去取代时间，这只要将各项作相应的变动就行了。

在以后的讨论中，并不需要对函数作精确的定义。可以把所讨论的绝大多数函数认为仅仅是一种过程，即由自变量的每一个值所决定的因变量也是一个单一值。当然也有例外的情况，如在某些条件下，函数呈现出具有多个因变量或多个自变量的情况。

本书所讨论的时间历程由于它们来源不同而受到若干限制，主要有下列四种：

1. 记录长度是有限的；
2. 数据的范围是有限的；
3. 数据是采样的；
4. 数据是离散的（数字化到有限的精确度）。

例1.1 用一台记录式的数字电压表去监控某个实验。在该实验中，传感器将被观测的物理量转换成电压，该电压输入到电压表中。数字电压表是以 1 伏为一级进行记录的，记录的范围为 ±100 伏，每秒采样一次，而实验取 100 次采样，即记录长度限于 100 秒。可见，高于 100 伏或者低于 −100 伏的数值就记录不上；出现于 1 秒采样间隔之间的电压也观测不到；同时小于 1 伏的波动也将被略掉。

用这样的系统观测所得的数值带有一定的局限性。这些局限性在后续各节频率域的分析中将继续起作用。

时间函数以下述方式表示：

$x(t)$ = 连续函数

及 x_i = 已采样的函数

通常采样间隔是均匀的，以 Δt 表示。于是，我们可写为：

$$x_i = x(i\Delta t) \tag{1-1}$$

函数 x 可以是整个测试过程中的任一被测物理量，又可以是数据处理过程中某一给定点上被测定的任一时间历程。如果必须同时对两个时间历程进行研究的话，那么第一个时间历程将以 x 表示，第二个以 y 表示。

注意，在讨论时间历程时，我们应用下述惯例：单个字母 x 一般表示函数或过程，而符号 $x(t)$ 表示函数 x 在时间 t 上的特定数值。

1.2 均　值

当所研究的时间历程是随机过程时，会出现很多问题。为了避免反复不断地讨论这些问题，现作如下假定：除非特殊说明外，当所讨论的时间历程是随机过程时，认为它们均是平稳的和各态历经的。这个专题将推迟到第十章中去讨论，在那一章中，这些术语的定义将更为确切些。

在上述假设下，函数的时间均值（平均值）等同于该函数的总体均值，并可以用时间均值来代替总体均值。连续数据的时间均值 $\bar{x}$ 定义为

$$\bar{x}=\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}x(t)dt \tag{1-2}$$

当然，可供利用的仅仅是记录的一个有限部分。假定在时间 $t=a$ 时记录开始，而当 $t=b$ 时记录结束，则这段数据上的时间均值为

$$\bar{x}=\frac{1}{b-a}\int_{a}^{b}x(t)dt \tag{1-3}$$

如果数据是采样的，则（1-3）式取下面形式：

$$\bar{x}=\frac{1}{N}\sum_{i=k}^{N-1+k}x_i \tag{1-4}$$

其中，$k\Delta t=a$，$(N-1+k)\Delta t=b$。

为方便起见，通常取 k 为零，所以（1-4）式可以写为

$$\bar{x}=\frac{1}{N}\sum_{i=0}^{N-1}x_i \tag{1-5}$$

在最后三个 $\bar{x}$ 的定义式中，可以添加脚标以表明平均值是在间隔（a，b）内取得的。不过，$\bar{x}$ 总是在给定点上所讨论的全部记录（指被用到的全部记录）的时间平均值。对于不符合这种规定的情况，当它们出现时应另作注明。

1.3 方 差

在 1.2 节中所讨论的假设条件下，时间历程的方差为 s_x^2，这里

$$s_x^2=\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}[x(t)-\bar{x}]^2dt \tag{1-6}$$

对于有限记录长度，上式变为

$$s_x^2=\frac{1}{t_2-t_1}\int_{t_1}^{t_2}[x(t)-\bar{x}]^2dt \tag{1-7}$$

x 的标准差是方差 s_x^2 的正平方根值。用(1-7)式来计算 s_x 的仪表被称为真值均方根($r.m.s$)表。还有其它一些类型的仪表，它们仅仅对于某些特定的周期函数(通常为正弦波)，给出 $r.m.s$ 值。

对于采样数据，s_x^2 的定义为

$$s_x^2=\frac{1}{N-1}\sum_{i=0}^{N-1}(x_i-\bar{x})^2 \tag{1-8}$$

在连续情况下，x 的均方值 ψ_x^2 为

$$\psi_x^2=\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}x^2(t)dt \tag{1-9}$$

如同方差一样，根据需要上式也可以应用于有限长度的记录以及离散型式的记录。

例1.2 假设x为一正弦波

$$x(t)=A\sin(2\pi f_c t+\phi) \qquad -\infty<t<\infty \tag{1-10}$$

那么

$$s_x^2=\lim_{T\to\infty}\frac{1}{T}\int_{-T/2}^{T/2}A^2\sin^2(2\pi f_c t+\phi)dt$$

$$=A^2\lim_{T\to\infty}\frac{1}{T}\int_{-T/2}^{T/2}\sin^2(2\pi f_c t+\phi)dt$$

$$=A^2\lim_{T\to\infty}\frac{1}{T}\left[\frac{T}{2}+\{\text{正弦项}\leqslant 1\}\right]=\frac{A^2}{2} \tag{1-11}$$

注意，对于这类函数而言，ψ_x^2● 和 s_x^2 是相等的，因为平均值为零。

1.4 正态分布和χ^2分布

设变量x的平均值为μ，方差为σ^2。如果x的概率密度函数$\phi(x)$为

$$\phi(x)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \tag{1-12}$$

那么x就是**正态分布**或者说x具有**高斯分布**。对于这样的高斯变量，其概率分布函数$\Phi(x)$被定义为

$$\Phi(x)=\int_{-\infty}^{x}\phi(\xi)d\xi \tag{1-13}$$

换句话说：

$$\Phi(x_0)=[-\infty<x\leqslant x_0\text{ 的概率}]$$
$$\Phi(x_1)-\Phi(x_0)=[x_0<x\leqslant x_1\text{ 的概率}]\ (\text{设}x_1>x_0)$$
$$1-\Phi(x_0)=[x_0<x<\infty\text{的概率}] \tag{1-14}$$

几个常用的Φ值（当$\mu=0$及$\sigma=1$）为

$$\frac{1}{\sqrt{2\pi}}\int_{-1}^{1}\exp\left(-\frac{x^2}{2}\right)dx=0.682$$

● 原文误为Φ_x^2。——校者

$$\frac{1}{\sqrt{2\pi}}\int_{-2}^{2}\exp\left(-\frac{x^2}{2}\right)dx=0.954$$

$$\frac{1}{\sqrt{2\pi}}\int_{-3}^{3}\exp\left(-\frac{x^2}{2}\right)dx=0.997^{●} \tag{1-15}$$

正态分布是一种理论上的概念，而实际上真正正态数据通常是不可能的，这里主要的问题是一个范围问题。大多数的数据函数具有有限的范围，而正态分布数据必须具有无限的范围。

例1.3 设一根2cm长的棒由一组人员来测量，每个人员都使用同一把尺子。假定所得的测量值呈正态分布，测量的真实平均值为2cm，方差为0.25cm²●，测量值小于或等于−0.5cm的概率为

$$\frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{-0.5}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]dx$$

以 $y=(x-\mu)/\sigma$ 取代得

$$=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\frac{-0.5-2}{0.5}}\exp\left(-\frac{y^2}{2}\right)dy$$

$$=p(-5.0)\approx 0.0000001 \tag{1-16}$$

因而，棒长度的负测量值具有有限的概率（尽管很小），这当然是很荒谬的。但是，完全可以证明正态分布是正确的，因为它不管怎样与真实分布都是极其近似的。

作正态性假设（即数据为正态分布）还有许多理由，主要如下：

1. 由于人们对于正态分布具有丰富的知识，这就使人们易于对某一给定的状态建立模型，从而对统计学进行讨论。

2. 中心极限定理（克拉默Cramer，1946年）指出，如果将足够多的作任意分布的变量相加起来，其和总是趋于正态分布的。

● 原文误为0.97。——校者

● 原文误为0.25cm。——校者

由于滤波数据（参看第三章）等效于许多观察值的相加，所以滤波后的数据就趋于正态分布。

正态分布函数的一些重要性质是：

1．正态变量 x 的密度函数和分布函数完全由平均值 μ_x 和方差 σ_x^2 来确定。

2．x 的各阶矩为

$$E[x^n]=\begin{cases}\mu & n=1\\ \mu^2+\sigma^2 & n=2\\ \mu(\mu^2+3\sigma^2) & n=3\\ \cdots\cdots & \end{cases}\tag{1-17}$$

各阶中心矩为

$$E[(x-\mu)^n]=\begin{cases}0 & n=1,3,\cdots\\ 1\cdot 3\cdots(n-1)\sigma^n & n=2,4,\cdots\end{cases}\tag{1-18}$$

如果 $\mu=0$，可得出一个非常重要的四阶矩关系式

$$E[x_a x_b x_c x_d]=E[x_a x_b]E[x_c x_d]+E[x_a x_c]E[x_b x_d]+E[x_a x_d]E[x_b x_c]\tag{1-19}$$

χ^2 分布

设 x_1，x_2，…，x_n 为独立的平均值为零，方差为 1 的高斯变量。我们将 χ_n^2 定义为

$$\chi_n^2=\sum_{i=1}^{n} x_i^2\tag{1-20}$$

这时 χ_n^2 被称为具有 n 个自由度（$d.f.$）的 χ^2 变量，且具有 χ^2 分布。χ_n^2 的密度函数为

$$p(\chi_n^2)=[2^{n/2}\Gamma(n/2)]^{-1}(\chi_n^2)^{\frac{n}{2}-1}\exp[-\chi^2_{n/2}]$$

$$\chi^2\geqslant 0\tag{1-21}$$

● 原文误为 $i-1$。——译者

● 原文误为 χ_n。——校者

● 原文漏指数 $\frac{n}{2}-1$。——校者

式中$\Gamma(a)$为伽玛(GAMMA)函数。

在时间序列数据分析中，χ^2分布主要用于讨论样本方差和功率谱密度的变动度。如果$\{x_i\}$的平均值为零，而$\{x_i\}$的N个样本值被用来计算样本方差s^2，那么真实方差值在两个界值之间

$$B_1 \leqslant \sigma^2 \leqslant B_2 \tag{1-22}$$

的概率为$(1-\alpha)$，其中B_1和B_2的定义为

$$B_1 = \frac{ns^2}{\chi^2_{n;\alpha/2}}$$

$$B_2 = \frac{ns^2}{\chi^2_{n;1-\alpha/2}} \qquad n = N-1 \tag{1-23}$$●

注意，B_1和B_2是s^2、α和χ^2● 的函数，区间(B_1, B_2)被称为**置信区间**；并且也可把它叫做$[(1-\alpha)100]\%$**置信区间**。

作为例子●，假设由正态分布随机变量中取出$N=31$个独立的观测值，并期望样本方差具有90%的置信区间。再假设采样平均值$\bar{x}=58.61$，样本方差$s^2=33.43$，那么理论方差σ^2的界值可由下式算出：

$$\frac{30s^2}{43.77} < \sigma^2 < \frac{30s^2}{18.49} \text{的概率} = 0.90$$

式中

$\chi^2_{30,0.05}=43.77$，即为使$p(\chi^2_{30})=0.05$的χ^2_{30}。

$\chi^2_{30,0.95}$● $=18.49$，即为使$p(\chi^2_{30})=0.95$的χ^2_{30}。

简化一下，上述界值为

$$22.91 \leqslant \sigma^2 \leqslant 54.24$$●

可见，平均起来每十次试验真实方差就要越出该区间一次。

● 原式中B_1、B_2的公式颠倒了。——校者

● 原文指数“2”漏掉。——校者

● 取自布莱克曼和图基的著作(1958年)。

● 原文误作95，另将“，”误写为“；”。——校者

● 原文误作54.22，且漏刊“≤”号。——校者

1.5 傅里叶变换

当谈到傅里叶变换时，应该记住它的四种主要形式，而每种形式又有它自己的变形。通常，不同的作者使用不同的乘法常数就会出现不同的形式，所以这些形式间的差别很小。另一方面，这四种基本形式间的差别，又是由于所应用的范围和定义域不同而造成的。也就是说，在这四种形式中，自变量和因变量的定义方式是不同的。表 1.1 示出这四种主要形式及它们的应用范围。

虽然进行时间序列数据分析时主要使用形式Ⅱ和Ⅲ，但还是首先讨论一下形式Ⅰ，因为它是人们最熟悉的了。在振动分析、通讯理论以及许多其他的物理应用领域中，形式Ⅰ是一种最通用的形式。

$$X(f)=\lim_{T\to\infty}\int_{-T}^{T} x(t)e^{-j2\pi ft}dt \tag{1-24}$$

这里，如果 x 的平方在勒贝格意义❶下是可积分的话，就可以证明 X 是存在的。通常，这个限制远比实际应用中所遇到的要宽得多，也就是说，这个限制仅是充分的而不是必要的。有两个值得注意的例外情况是正弦函数和余弦函数，这两个函数的平方和积分是无界的。然而，也将看到它们的傅里叶变换仍是可能存在的。还可证明存在这样一个函数 x_1，即

$$x_1(t)=\lim_{F\to\infty}\int_{-F}^{F} X(f)e^{j2\pi ft}df \tag{1-25}$$

这个函数 x_1 与 x 十分相象。事实上下述方程式

$$x(t)-x_1(t)=0 \tag{1-26}$$

对于绝大多数 t 值均成立，用术语说即（1-26）式几乎处处为真。于是，式（1-25）所定义的运算几乎等同于式（1-24）所给定的相反步骤。因为它们之间的差别甚微，在大多数实际情况中甚至

❶ 细节可参考维纳（Wiener）1930年的著作。——原注