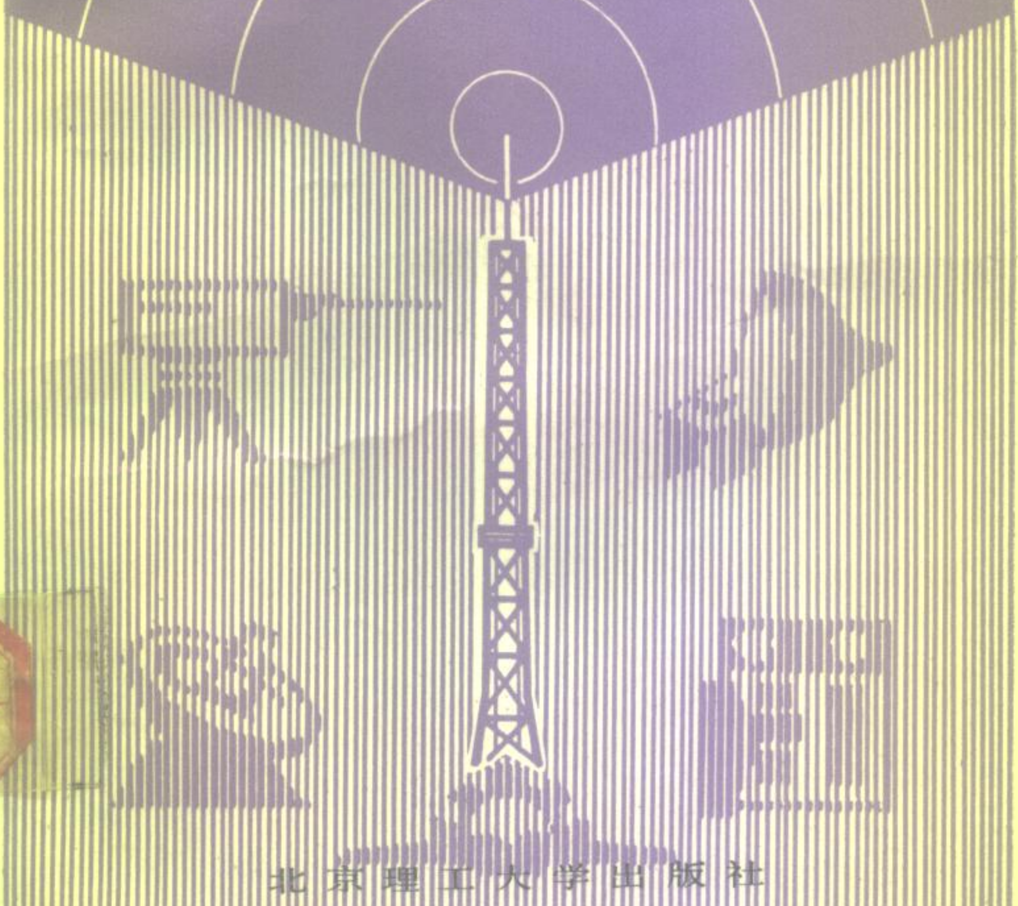
高等学校教材

动态数据的统计分析

甘仞初　编著



北京理工大学出版社

动态数据的统计分析

甘仞初　编著

北京理工大学出版社

内 容 简 介

本书系原电子工业部统编教材。阐述依据观测数据对动态过程建模和谱分析的理论、方法及其应用。主要内容有：动态数据的统计分析基础、平稳序列和连续过程的线性动态模型、模型的参数估计、结构辨识和建模方法，动态过程的预报，谱分析基础和谱估计方法。本书面向工程和管理等应用领域，系统性强、内容丰富、实例较多。可供工科和管理类各专业研究生、高年级大学生40—60学时课程教学之用，亦可供有关科技人员、管理人员自学、参考用。

动态数据的统计分析

甘仞初

*

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

*

850×1168毫米　32开本　12.75 印张　330 千字

1991年5月第一版　　1991年5月第一次印刷

ISBN 7-81013-400-0/C・22

印数：1—2800册　　　　定价：4.15元

出版说明

根据国务院关于高等学校教材工作分工的规定，我部承担了全国高等学校、中等专业学校工科电子类专业教材的编审、出版的组织工作。由于各有关院校及参与编审工作的广大教师共同努力，有关出版社的紧密配合，从1978年至1985年，已编审、出版了两轮教材，正在陆续供给高等学校和中等专业学校教学使用。

为了使工科电子类专业教材能更好地适应“三个面向”的需要，贯彻“努力提高教材质量，逐步实现教材多样化，增加不同品种、不同层次、不同学术观点、不同风格、不同改革试验的教材”的精神，我部所属的七个高等学校教材编审委员会和两个中等专业学校教材编审委员会，在总结前两轮教材工作的基础上，结合教育形势的发展和教学改革的需要，制订了1986～1990年的“七五”(第三轮)教材编审出版规划。列入规划的教材、实验教材、教学参考书等近400种选题。这批教材的评选推荐和编写工作由各编委会直接组织进行。

这批教材的书稿，是从通过教学实践、师生反映较好的讲义中经院校推荐，由编审委员会(小组)评选择优产生出来的。广大编审者、各编审委员会和有关出版社为保证教材的出版和提高教材的质量，作出了不懈的努力。

限于水平和经验，这批教材的编审、出版工作还会有缺点和不足之处，希望使用教材的单位，广大教师和同学积极提出批评建议，共同为不断提高工科电子类专业教材的质量而努力。

电子工业部教材办公室

序　言

本教材系按原电子工业部的工科电子类专业教材1986—1990年编审出版规划，由工科电子类教材编审委员会自动控制编审小组征稿，推荐出版，责任编委是吴沧浦教授。

本教材由北方交通大学黄琇坤教授主审。

随着科学技术的进步和社会经济的发展，在工程技术、经济管理、自然科学和社会科学的许多领域，经验的、定性的传统方法正逐渐被科学的、定量的现代方法所替代，人们日益重视对各种现象的定量观测及有关数据的收集与分析。电子计算机的迅速发展和广泛应用，为处理与分析大量数据提供了日臻完善的手段。

自然的和人为的各种现象中，依时间变化且不同时间的状况相互关联的现象极为普遍。这类现象称为动态过程。产生动态过程的系统称为动态系统。动态数据就是对动态过程进行观测所记录的依时间变化的数据。统计学中把按时间顺序排列的数据称为时间序列；工程中把依时间变化的随机现象的记录称随机信号或随机振动数据；数学中把某个随机过程的观测数据称为该过程的实现或样本函数。从数据处理和系统分析的意义上来说，这些都统称为动态数据。国民经济中逐年工、农业总产值和国民收入的记录数据，工、商企业中产量、质量、能源消耗、价格以及销售额的逐日、逐月或逐年记录数据，机械振动、语音、噪声等的记录数据，飞行器的运动状态的观测数据，自然现象中的气候、气象、水文、地质随时间变化的各种观测记录以及医学中的脑电、心电记录数据等等，都是实际生活中动态数据的例子。有些现象的主要特征虽不随时间明显变化，但依其他参量变化，如机械制

此为试读，需要完整PDF请访问：www.ertongbook.com

造中的表面粗糙度数据，地理中的地貌数据，均依空间位置变化。在研究其变化规律时，也按动态数据进行分析。因此，动态数据这一概念概括了极为广泛的自然和社会生活中的现象。

动态数据有两个重要特性。其一是时序性。几乎所有随时间变化的现象在不同时间的状态都相互关联，当前的状态是过去状态某种程度的继续，因而动态过程的观测数据，均依观测时间的顺序排列。改变观测数据的时间顺序，就会影响对所观测的现象进行正确的分析。一般说来，两不同时刻的间隔愈小，同一现象两状态之间的关联愈密切。例如物价的波动、销售额的增减，化工过程中温度、浓度、粘度的变化，气候的变迁，物体的移动等等，都是如此。所谓相关性、记忆性、惯性或动态性质，都是在不同场合下对这种相互关联特性的描述，反映了所研究的事物发展变化的规律性。其二是随机性。几乎所有实际系统都处在复杂的环境中。在依时间变化的各类现象中，绝大多数为其未来状态不能预先确知的过程。数学中定义的随机过程就属于这类过程。有些动态数据反映的现象并不完全符合数学上严格定义的随机过程的条件，但用基于随机过程理论的一些方法分析、研究这些现象，仍可得到有效的结果。因此，本书中提到的动态数据，均作为随机过程的观测数据进行研究。

动态数据的统计分析，就是应用系统的、统计的方法，依据观测数据研究随机现象依时间变化的规律及其应用。动态数据统计分析的理论和方法可分为时域分析和频域（谱）分析两部分。对动态数据进行分析的目的，从应用的角度来看，大体可归纳为以下几个方面：

1． 根据观测数据推断所研究的动态过程的统计特性，建立动态过程或动态系统的数学模型，为系统分析、设计和控制提供特性数据；

2． 预报动态过程或系统的未来状态；

3． 根据观测数据反映的动态过程统计特性的差异，对未知

系统的类型进行识别，或者对已知系统特性的变迁与可能出现的故障进行诊断；

4. 根据观测数据，推断有关系统或过程之间的固有联系；

5. 利用按观测数据建立的数学模型，在一定条件下复现某些随机现象，对这种现象或有关系统进行仿真研究。

1982年以来，作者为北京理工大学自动控制、系统工程、管理工程、机械工程、电子工程、化学工程和飞行器工程等专业的研究生及高年级本科生讲授“动态数据的统计分析”课程。本书是在该课程讲稿的基础上并吸取了作者在美国威斯康星大学及回国后的部分研究工作编写的。作为教材，由于教学时数的限制并考虑到讲授对象均来自工程与管理各应用领域，故本书着重讨论时域分析中比较实用的线性模型理论与方法，并且集中研究一维问题。此外，还简要介绍了谱分析的基础知识。在讨论中，本书在重视理论推导的严密性的同时，尽量避免涉及过深的数学背景。具备我国大学工科本科数、理基础（包括概率论与数理统计的基本知识）的学生、教师以及其他研究工作者和实际工作者均可阅读此书。

全书内容共分8章。由于我国工程与管理方面大多数研究生与大学本科专业教学计划中不设“随机过程理论与应用”课程，本书第一章扼要介绍了随机过程，特别是平稳随机过程理论以作为研究后续章节的预备知识。这一章还讨论了利用样本函数估计随机过程一、二阶数字特征的方法。第二章对平稳序列的线性模型——自回归滑动平均($ARMA$)模型和具有输入项的自回归滑动平均($ARMAX$)模型进行了分析。这是时域线性模型理论的基本内容。在一些应用领域，连续过程的连续模型仍然是分析、设计和控制所用的一类模型。而连续过程离散化又给数据采集、处理和建模带来很大方便，为此，第三章研究连续线性模型的离散化及连续模型与离散模型的参数关系和相互转换问题。非平稳过程是实际生活中极为广泛的一大类随机过程。非平稳序列的线

性模型及其应用在第四章讨论。依据观测数据建立所研究的过程或系统的线性动态模型，是动态数据时域分析的核心问题。依据观测数据估计模型参数的方法在第五章中讨论。这一章介绍了模型参数估计的线性最小二乘法及其推广算法、非线性最小二乘算法以及最大似然法。第六章研究模型结构辨识问题，还介绍了建立*ARMA*模型的波克斯(G.E.P.Box)—詹金斯(G.M.Jenkins)方法和潘迪特(S.M.Pandit)—吴(S.M.Wu)方法。作为平稳和非平稳线性模型应用的一个方面，第七章讨论了利用线性模型进行预报的方法。第八章是平稳过程的谱分析，介绍了连续和离散傅里叶变换以及谱分析的基本概念，讨论了经典谱估计方法和现代谱估计中的*ARMA*谱估计方法。

如前所述，动态数据这一概念概括了极为广泛的自然和社会生活现象，因而动态数据统计分析的理论和方法具有十分广阔的应用范围。“动态数据的统计分析”这门课程系面向工程、管理各应用领域，因此本书所举各例只是为了介绍方法的应用，而不可能对某特定专业的问题进行具体、详细的分析。各例所用观测数据的来源，凡不是作者收集的，均在相应的附表中作了说明。

“动态数据的统计分析”课程教学课内时数为60学时，其中课堂讲授45～50学时，课程大型作业所用课内时数为10～15学时。本书是按课程教学要求编写的。为了巩固课堂所学知识，培养学生应用所学理论与方法来分析、解决实际问题的能力，大型作业要求每个学生结合自己所学专业的一个实际问题，进行动态数据收集、相关分析、建模、模型分析和谱分析等工作，并写出详细的、有自己见解的分析报告。北京理工大学管理信息系统实验室提供上述工作所需用的计算机软件及与本书内容有关的其他计算机程序。学生还可自选小课题，利用上述软件进行研究。

编写本书时，作者参阅了大量国内外文献。其中凡直接引用的在正文中都作了注明并列在书末的参考文献目录中，未直接引用但对作者编写本书有启发、参照作用或可供读者阅读本书时参

考的，也选择其中一部分列入参考文献目录。

本书责任编委北京理工大学吴沧浦教授对本书的编写给予了热情的鼓励、支持与帮助，本书主审、北京交通大学黄绣坤教授在百忙中详细审阅了书稿并提出了许多宝贵意见，作者谨表示衷心的感谢。

目　录

此为试读，需要完整PDF请访问：www.ertongbook.com

第一章　随机过程的统计分析基础

§1.1　随机过程的概念

自然界和社会生活中，事物的变化和发展过程，呈现出错综复杂的情况。有些过程可以用确定性的函数描述，这类过程称为确定性过程。确定性过程的一个显著特征是其变化和发展可以预先确知。但是，人们所遇到的大量现象是不能预先确知其变化和发展状况的，这些现象不可能用确定性的函数来描述。对于观测者来说，即使有可能在相同的条件下进行重复观测或实验，所得结果也不尽相同。客观事物所表现的这种不确定性，称为随机性。具有随机性的现象，称为随机现象。

在概率论中，对随机现象进行观测或试验，称为随机试验，简称试验。随机变量是概率论中对随机试验结果的定量描述。但随机变量的概念没有包含变化和发展的因素。许多随机现象的试验或观测结果不仅与试验条件有关，还随时间或其他参量而发生变化。对这类随机试验的结果的定量描述，就是随机过程(Stochastic Processes)。

随机过程的定义：设某随机试验的结果，对于给定的参数集T中每一个元素t，均对应于一随机变量$X(t)$，则称依赖于t的一族随机变量$\{X(t),\ t\in T\}$为一随机过程。其中t为时间或其他参数。

依上述定义的随机过程$X(t)$，对于某个固定的参数$t=t_1$，$X(t)$的取值$X(t_1)$称为此过程在t_1的状态。状态可以是多维的，即为向量；也可以是一维的，即为标量。随机过程的状态可以是离散型的，即$X(t)$对于固定的t的可能取值只是实数轴上有限个

或可列个点。这类随机过程称为离散状态的随机过程。如某车间每天加工的零件中废品的个数就是这类过程的一个例子。与此对应的是状态连续的随机过程，即状态在实数轴的某区域连续取值的随机过程。这类过程的一个典型例子就是机械系统的随机振动。参数集T可以是离散型的，即T中的元素t在实数轴上的有限个或可列个点上取值，或者是连续型的，即t在整个实数轴或某个区间连续取值。对应于上述两种参数集的随机过程分别称为离散参数的随机过程和连续参数的随机过程。一个商店每日销售额是离散参数随机过程的例子，而供电电网某点的电压波动则属于连续参数的随机过程。本书中我们只讨论状态连续的情况，因此以后我们将离散参数和连续参数的随机过程分别简称为离散随机过程和连续随机过程，有时又称离散随机过程为随机序列。

随机过程的一族观测数据称为该过程的一个样本函数或实现。图1.1所示曲线是某旅行车行驶中随机振动测试数据，这是连续随机过程样本函数的一个例子（数据见书末附表$A1$，表中是经采样的数据）。离散随机过程样本函数的例子如图1.2所示。这是太阳黑子年平均数的部分记录（数据见附表$A2$，图中只示出了表中部分数据）。

图1.1 某旅行车随机振动数据

为了便于数据处理，连续随机过程的样本函数常要通过采样加以离散化。因此，动态数据统计分析的主要对象是离散过程。

随机过程理论是动态数据统计分析的理论基础。虽然一些领

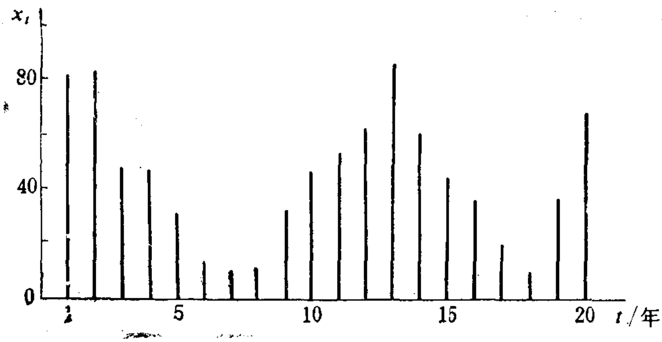


图1.2　太阳黑子年平均数部分数据

域中所要处理的许多实际过程如某些经济过程、地震波、语言等，未必能概括于数学上严格定义的随机过程之中，但基于随机过程理论的一些分析方法在处理这些实际问题中仍可获得很大成功。因此，在本书中，动态数据或时间序列均作为随机过程的样本函数来分析、处理。

§1.2　随机过程的有限维概率分布函数和数字特征

一、有限维概率分布函数和概率密度函数

由概率论可知，一个随机变量统计规律的最完善描述是它的概率分布函数或概率密度函数。多个随机变量的统计关系的最完善描述是它们的联合概率分布函数或联合概率密度函数。随机过程是依时间(或其他参数)变化的一族随机变量，因此，随机过程任意时刻的状态的概率分布函数或概率密度函数，以及各不同时刻的状态的联合概率分布函数或联合概率密度函数就被用来描述随机过程的统计规律。

1. 一维概率分布函数和概率密度函数

设$X(t)$为一随机过程，T为其参数集，对于任一固定的参数$t\in T$，$X(t)$是一个随机变量，它的分布函数一般与t有关，记为

$$F_{X(t)}(x,t)=P\{X(t)<x\} \tag{1.2.1}$$

称$F_{X(t)}(x,t)$为$X(t)$的一维概率分布函数，有时简记为$F(x,t)$或$F(x_t)$。

如果存在一函数$f_{X(t)}(x,t)$使

$$F_{X(t)}(x,t)=\int_{-\infty}^{x} f_{X(t)}(z,t)\,\mathrm{d}z \tag{1.2.2}$$

成立，则称$f_{X(t)}(x,t)$为$X(t)$的一维概率密度函数，简称一维概率密度，有时简记为$f(x,t)$或$f(x_t)$。

一维概率分布函数或一维概率密度函数描述随机过程各个孤立时刻的状态的统计特性，但不能反映各不同时刻状态之间的统计关系。

2. 二维概率分布函数和概率密度函数

对于任意两时刻$t_1,t_2\in T$，随机过程$X(t)$的状态$X(t_1)$、$X(t_2)$的联合概率分布函数

$$F_{X(t)}(x_1,x_2,t_1,t_2)=P\{X(t_1)<x_1,\ X(t_2)<x_2\} \tag{1.2.3}$$

称为$X(t)$的二维概率分布函数，有时简记为$F(x_1,x_2,t_1,t_2)$或$F(x_{t_1},x_{t_2})$。

若存在一函数$f_{X(t)}(x_1,x_2,t_1,t_2)$使

$$F_{X(t)}(x_1,x_2,t_1,t_2)=\int_{-\infty}^{x_1}\int_{-\infty}^{x_2} f_{X(t)}(z_1,z_2,t_1,t_2)\,\mathrm{d}z_1\mathrm{d}z_2 \tag{1.2.4}$$

成立，则称$f_{X(t)}(x_1,x_2,t_1,t_2)$为$X(t)$的二维概率密度函数，有时简记为$f(x_1,x_2,t_1,t_2)$或$f(x_{t1},x_{t2})$。

3. n维概率分布函数和概率密度函数

对于时间t的任意n个数值$t_1,t_2,\cdots,t_n\in T$，$X(t)$的n个状态$X(t_1)$、$X(t_2)$、$\cdots$，$X(t_n)$的联合概率分布函数

$$F_{X(t)}(x_1,x_2,\cdots,x_n;t_1,t_2,\cdots,t_n)=$$

$$=P\{X(t_1)<x_1, X(t_2)<x_2, \cdots, X(t_n)<x_n\} \tag{1.2.5}$$

称为随机过程$X(t)$的n维概率分布函数，有时简记为$F(x_1, x_2, \cdots, x_n; t_1, t_2, \cdots, t_n)$或$F(x_{t_1}, x_{t_2}, \cdots, x_{t_n})$。

与一维、二维情形一样，若有函数$f_{X(t)}(x_1, x_2, \cdots x_n; t_1, t_2, \cdots, t_n)$使

$$\begin{aligned} &F_{X(t)}(x_1, x_2, \cdots, x_n; t_1, t_2, \cdots, t_n) \\ &=\int_{-\infty}^{x_1}\int_{-\infty}^{x_2}\cdots\int_{-\infty}^{x_n} f_{X(t)}(z_1, z_2, \cdots, z_n; t_1, t_2, \\ &\cdots, t_n)\mathrm{d}z_1\mathrm{d}z_2\cdots\mathrm{d}z_n \end{aligned} \tag{1.2.6}$$

成立，则称$f_{X(t)}(x_1, x_2, \cdots, x_n; t_1, t_2, \cdots, t_n)$为随机过程$X(t)$的$n$维概率密度函数，有时简记为$f(x_1, x_2, \cdots, x_n; t_1, t_2, \cdots, t_n)$或$f(x_{t_1}, x_{t_2}, \cdots, x_{t_n})$。

二维分布特性概率分布函数或概率密度函数描述随机过程两任意时刻的状态的统计特性和两时刻状态间的统计关系，n维分布特性概率分布函数或概率密度函数描述随机过程任意n个时刻的状态的统计特性和它们之间的统计关系。由概率分布函数的性质可知，对于任意正整数$m<n$，一随机过程的m维分布特性完全取决于n维分布特性。显然，n愈大，对随机过程的统计规律的描述愈完善。若对于任一有限数n，某随机过程的n维分布特性已知，则完全掌握了此随机过程的统计规律。有限维概率分布函数或概率密度函数也是随机过程统计特性其他描述形式的基础。

4. $(n+m)$维联合分布

设$X(t)$和$Y(t')$为两随机过程，它们的参数集分别为T和T'。对于参数(时间)t的任意n个数值$t_1, t_2, \cdots, t_n \in T$和$t'$的任意$m$个数值$t'_1, t'_2, \cdots, t'_m \in T'$，$X(t)$的$n$个状态$X(t_1)$、$X(t_2)$、$\cdots$、$X(t_n)$和$Y(t')$的$m$个状态$Y(t'_1), Y(t'_2), \cdots, Y(t'_m)$的联合概率分布函数

$$\begin{aligned} &F_{X(t),Y(t')}(x_1, \cdots, x_n; t_1, \cdots, t_n; y_1, \cdots, y_m; t'_1, \cdots, t'_m) \\ &=P[X(t_1)<x_1, \cdots, X(t_n)<x_n; Y(t'_1)<y_1, \cdots, Y(t'_m) \\ &<y_m] \end{aligned} \tag{1.2.7}$$

此为试读，需要完整PDF请访问：www.ertongbook.com