

数据建模及预报

SHUJU JIANMO YUBAO

张树侠 吴简彤 编著

哈尔滨工程大学出版社

数据建模及预报

张树侠 吴简彤 编著

哈尔滨工程大学出版社

内容简介

近代各项科学实验与工程实践,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得所期望的结果,正是编写本书的目的所在。

本书共九章,首先简要介绍了数据的分类与描述;接着较全面地讲述了自回归滑动平均(ARMA)模型的特性、建模与预报;并结合应用对陀螺漂移控制和 GPS/SINS 组合导航系统的数据处理进行了较详细的描述;最后对基于神经网络的信号处理和小波分析基本原理作了介绍。

本书可作为导航、制导与控制、测控技术与仪器、控制理论与工程等专业高年级学生和相关专业研究生的教材和参考书。对从事上述专业的工程技术人员也有一定的参考价值。

数据建模及预报

Shuju Jianmo ji Yubao

张树侠 吴简彤 编著

责任编辑 李 英

*

哈尔滨工程大学出版社出版发行

哈尔滨市南通街145号 哈工程大学11号楼

发行部电话(0451)2519328 邮编:150001

新华书店经销

东北农业大学印刷厂印刷

*

开本 787×1092 1/16 印张 10.75 字数 260 千字

1999年4月第1版 1999年4月第1次印刷

印数:1~2000册

ISBN 7-81007-908-5

TP·80 定价:13.00元

出版说明

根据国务院发(1978)23号文件批转试行的《关于高等学校教材编审出版若干问题的暂行规定》,中国船舶工业总公司负责全国高等学校船舶类专业规划教材编审、出版的组织工作。

为做好教材编审组织工作,中国船舶工业总公司相应地成立了“船舶工程”、“船舶动力”两个教材委员会和“船电自动化”、“惯性导航及仪器”、“水声电子工程”、“液压”、“水中兵器”五个教材小组,聘请了有关院校的教授、专家 50 余人参加工作。船舶类专业教材委员会(小组)是有关船舶类专业教材建设研究、指导、规划和评审方面的专家组织,主要任务是协助船舶总公司做好高等学校船舶类专业教材的编审工作,为教材质量审查把关。

经过前四轮教材建设,共出版教材 300 余种,建立了较完善的规章制度,扩大了出版渠道,在教材的编审依据、计划体制、出版体制等方面实行了有成效的改革,这些为“九五”期间船舶类专业教材建设奠定了良好基础。根据国家教委对“九五”期间高校教材建设要“抓好重点教材,全面提高质量,继续增加品种,整体优化配套,深化管理体制和运行机制的改革,加强组织领导”的要求,船舶总公司于 1996 年又制定了“全国高等学校船舶类专业教材(九五)选题规划”。列入规划的选题共 133 种,其中部委级重点选题 49 种,一般选题 84 种。

“九五”教材规划是在我国发展社会主义市场经济条件下第一个教材规划,为适应社会主义市场经济外部环境,“九五”船舶类专业教材建设实行指导性计划体制。即在指导性教材计划指导下,教材编审出版由主编学校负责组织实施,教材委员会(小组)进行质量审查,教材编审室组织协调。

“九五”期间要突出抓好重点教材,全面提高教材质量,为此教材建设引入竞争机制,通过教材委员会(小组)评审、择优确定主编,实行主编负责制。教材质量审查实行主审、复审制,聘请主编校以外的专家审稿,最后教材委员会(小组)复审,复审合格后由有关教材委员会(小组)发出版推荐证书,出版社方可出版。全国高校船舶类专业规划教材,就是通过严密的编审程序和高标准、严要求的审稿工作来保证教材质量。

为完成“九五”教材规划,主编学校应充分发挥主导作用。规划教材的立项是由学校申报,立项后由主编校组织实施,教材出版后由学校组织选用,学校是教材编写与教材选用的行为主体,教材计划的执行主要取决于主编校工作情况。希望有关高校切实负起责任,各有关方面积极配合,为完成“九五”船舶类专业教材规划、为编写出版更多的精品教材而努力。

由于水平和经验局限,教材的编审出版工作和教材本身还会有很多缺点和不足,希望各有关高校、同行专家和广大读者提出宝贵意见,以便改进提高。

中国船舶工业总公司教材编审室
一九九七年四月

前 言

在信息社会中,人们的时间倾向是未来,特别要注意未来的结构变化,这是人类的预测、控制能力达到一定水平的标志。“面向未来”概括了这个时代的特征。

工程、物理、生物学或社会科学等领域的实际工作者和研究人员经常要对一系列的观察数据进行处理和分析。一般说来,试验、观察的结果总是受当时当地条件的限制,得到的结果总具有一定的随机性,同样的事件,这次试验结果就可能与下一次的试验结果不一样,这是由于有某些未知的原因及噪声的影响,使数据中含有不需要的成分所致。如何对试验结果进行去伪存真的分析处理,并依据对这些数据的定量分析得到一个数学模型,且用它来简洁而又全面地表征所研究的系统,就是本书的目的。这个模型可以用来分析系统并且预测它在变化着的条件下的行为,根据分析所得到的信息还可以进一步对系统中某些因素和变量适当改变以达到在某种意义下的最优性能。

在系统分析中,传统的方法是根据系统的结构,应用某些基本原理(例如牛顿定律、能量守恒定律、基尔霍夫定律等)建立起系统的微分方程;如果系统比较复杂,则应用实验的方法(例如频率响应实验等)进行分析。而在时间序列分析中,统计工作者或经济工作者对所获得的数据则是以某种数学模型进行拟合,以使用这种模型进行预报和预测,所以时间序列分析和系统分析是两个不同的分析方法。

但实际上它们是相互联系着的。本教材的目的是将时间序列分析和系统分析结合在一起,向系统分析工作者提供一种新的工具,这种新的工具已用于系统辨识、图象分析、物理模拟、控制以及工程设计。特别是对于预报最有用处,这也正是时间序列发展的本来目的。

书中所讨论的内容是根据依次排列的观察数据序列进行系统的建模、分析、预报和控制。产生数据的系统可能是容易得到的某种具体的形式,例如一个化学过程或机械震动系统,也可能是难以具体化的抽象形式,例如飞机客票销售量的变化情况。书中所研究的方法对于上述两种情况都可适用。这种方法可以只根据数据建立用随机差分方程或微分方程形式表示的数学模型。差分方程可用来表示抽象系统,它可用于预测和控制,而微分方程则可用于对产生数据的系统进行物理解释,以及用来表征、分析它的品质(如稳定性等),并且也可以将它和别的系统进行比较,并通过改变有关组成部分对系统进行修改。若再加上从数据以外所获得的关于系统的定性或定量的知识,则微分方程可以用于改善系统的设计。

书中特别注重一阶和二阶系统,并加以详细的讨论。一阶系统使读者能够容易掌握基本概念而不会过多地陷入数学问题,二阶系统因为具有一些其他特点和实际重要性而值得仔细讨论。对于更高阶的系统则给定了有关的表达式,但省略了对它们的推导。

本书是编著者在近几年来讲授《数据建模及预报》课程的讲义基础上,并结合当前的科研实践而编写成的,其中1~5章由张树侠同志编写,6~9章由吴简彤同志编写。哈尔滨工业大学冯幼田教授主审了全书,对书稿提出了许多宝贵意见和建议。在本教材的立

项与编写过程中得到了中国船舶工业总公司教编室主任李堃的指导和鼓励。在此,对他们的辛勤劳动和热情帮助表示衷心的感谢。

本教材引用了国内外许多学者的论著和科研成果,在此谨致谢意。

限于编著者的水平和能力,本书难免还存在错误和不妥之处,殷切希望读者指正与批评。

编著者

1998年9月于哈尔滨

目 录

第 1 章 数据的分类和描述	1
1.1 确定性数据的分类	1
1.2 随机数据的分类	4
1.3 随机数据的基本特性	6
1.4 随机数据的联合特性	13
第 2 章 自回归滑动平均 (ARMA) 模型	18
2.1 线性回归模型	18
2.2 一阶自回归模型	22
2.3 自回归滑动平均模型 $ARMA(2,1)$	27
2.4 $ARMA(n, n-1)$ 模型和建模方案 (策略)	32
第 3 章 ARMA 模型的特性	36
3.1 格林函数和稳定性	36
3.2 逆函数和可逆性	45
3.3 自协方差函数	49
3.4 偏自相关函数和自频谱	57
第 4 章 建 模	60
4.1 建模的系统方法	60
4.2 关于自回归部分的阶数增量问题	65
4.3 估 计	66
4.4 模型的适用性的检验	72
4.5 建模步骤及举例	74
第 5 章 预 报	80
5.1 预报可以看作为正交投影	80
5.2 以条件期望作预报	82
5.3 预报的校正	90
5.4 指数平滑	90
第 6 章 随机漂移的控制	96
6.1 问题的提出	96
6.2 顺馈控制系统设计	97
6.3 反馈控制系统设计	99
6.4 顺馈—反馈联合控制系统	101
第 7 章 GPS/SINS 组合导航系统中的数据处理	103
7.1 概述	103

7.2	GPS 全球定位系统	103
7.3	GPS 误差信号的分析与建模	106
7.4	捷联式惯性导航系统原理及误差分析	110
7.5	GPS/SINS 组合导航数据处理	114
第 8 章	基于神经网络的信号检测与估计	122
8.1	大脑神经元	122
8.2	大脑的人工神经网络模型	124
8.3	基本神经网络模型	130
8.4	神经网络信号检测	140
8.5	神经网络滤波	144
8.6	神经网络谱估计	148
第 9 章	小波分析基本原理	150
9.1	引 言	150
9.2	小波变换	151
9.3	数字信号的二进小波变换	155
附录	158
参考文献	161

第 1 章 数据的分类和描述

表示物理现象和社会经济现象的任何观察数据,都可以广泛地分为确定性和非确定性的两类。能够用明确的数学关系式描述称为确定性数据,例如简单弹性质量系统,如图 1.1 所示。

假定刚体离开平衡位置的距离为 X ,松手的时间记为 $t=0$,根据力学的基本定律或重复观察,可以建立如下关系:

$$x(t) = X \cos\sqrt{\frac{K}{m}}t \quad (t \geq 0) \tag{1.1}$$

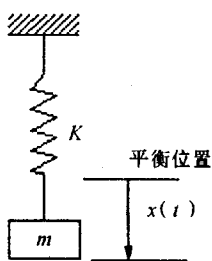


图 1.1 简单弹性质量系统

方程式(1.1)确定了刚体在以后任意瞬时的精确位置。因此表示刚体运动的物理数据是确定性的。但是实践中有许多物理现象产生的数据是不确定性的。例如,起伏海面的波高,噪声发生器的输出电信号等,都不能用明确的数学关系式来描述,无法预测未来时刻的精确值,这些数据在性质上是随机的,不能用精确的方程式,而只能用概率术语和统计平均来描述。

从分析的观点来研究,若一个产生数据的实验,在一定误差范围内,能够重复许多次得到相同的结果,则一般可以认为这些数据是确定性的,若不能设计一种实验,重复进行并产生相同的结果,则一般认为这些数据是随机性的。

1.1 确定性数据的分类

图 1.2 表示了确定性数据的这些不同类别。

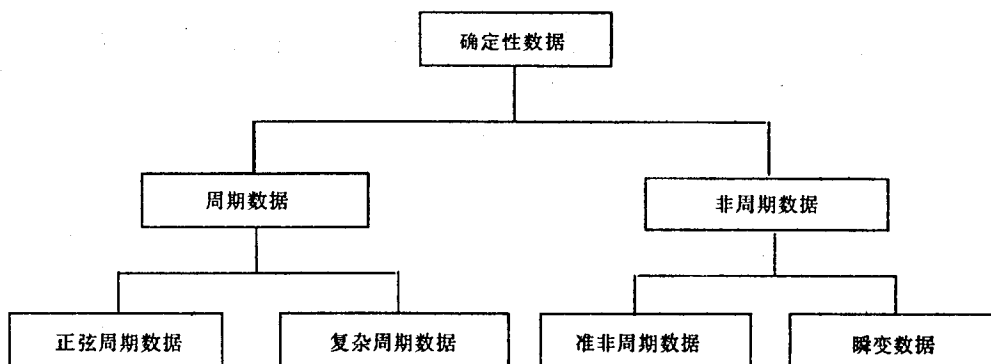


图 1.2 确定性数据分类

1.1.1 正弦周期数据

正弦周期数据可以用下列形式的时变函数表示：

$$x(t) = X \cdot \sin(2\pi f_p t + \theta_0) \quad (1.2)$$

式中： X ：振幅； f_p ：频率； θ_0 ：相对于时间原点的初始相角。当 $\theta_0 = 0$ 时，(1.2) 式又可以画成随时间变化的记录曲线图和振幅频率图(频谱)，见图 1.3。

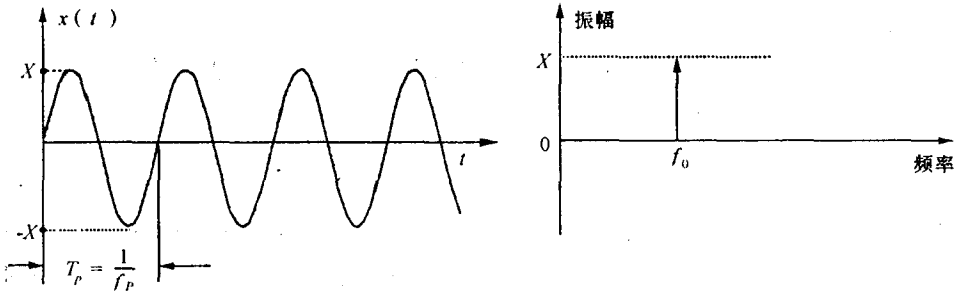


图 1.3 正弦数据的时间历程与谱

正弦数据完成一个循环所需要的时间，称为周期 T_p ，单位时间内的循环数称为频率 f_p ，频率与周期的关系是 $T_p = \frac{1}{f_p}$ 。图 1.3 的频谱由一个具体频率 f_p 上的振幅 X 构成，这样的谱称为离散谱或线谱。

1.1.2 复杂周期数据

复杂周期数据用周期性的时变函数表示如下：

$$x(t) = x(t + nT_p) \quad n = 1, 2, 3, \dots \quad (1.3)$$

这种时变函数在一般区间上具有基波的整数倍数的波形。复杂周期数据可以用富里叶级数(Fourier)表示，即：

$$x(t) = X_0 + \sum_{n=1}^{\infty} X_n \cos(2\pi n f_1 t - \theta_n) \quad (1.4)$$

$$\text{式中：} X_0 = \frac{a_0}{2}, f_1 = \frac{1}{T_p}, X_n = \sqrt{a_n^2 + b_n^2}, n = 1, 2, 3, \dots$$

$$\theta_n = \tan^{-1}\left(\frac{b_n}{a_n}\right), n = 1, 2, 3, \dots$$

$$a_n = \frac{2}{T_p} \int_0^{T_p} x(t) \cdot \cos(2\pi n f_1 t) dt, n = 1, 2, 3, \dots$$

$$b_n = \frac{2}{T_p} \int_0^{T_p} x(t) \cdot \sin(2\pi n f_1 t) dt, n = 1, 2, 3, \dots$$

式(1.4)表明了复杂周期数据由一个静态分量 X_0 和无限个被称为谐振的正弦分量

组成。谐振分量的频率都是基频 f_1 的整数倍数。其幅频特性可用图(1.4)所示的离散谱来表征。

产生复杂周期数据的物理现象远远多于产生简单正弦数据的物理现象。事实上,归类为正弦的数据往往是一个复杂数据的近似表示。例如多缸复式发动机的振动响应,通常会表现出宏观的谐振分量。

1.1.3 准周期数据

如上所述,周期数据一般可以简化成一系列的正弦波,反过来两个和几个频率成正比例的正弦波加起来,将组成一个周期数据。但是,任意频率的两个或几个正弦波之和,一般不是周期的。可用如下形式的时变函数表示:

$$x(t) = \sum_{n=1}^{\infty} X_n \sin(2\pi f_n t + \theta_n) \quad (1.5)$$

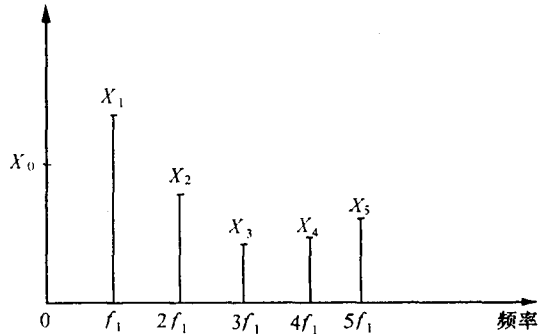


图 1.4 复杂周期数据的谱

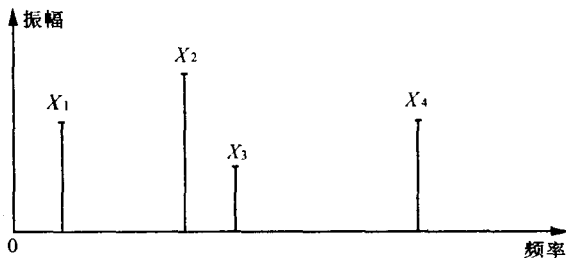


图 1.5 准周期数据的谱

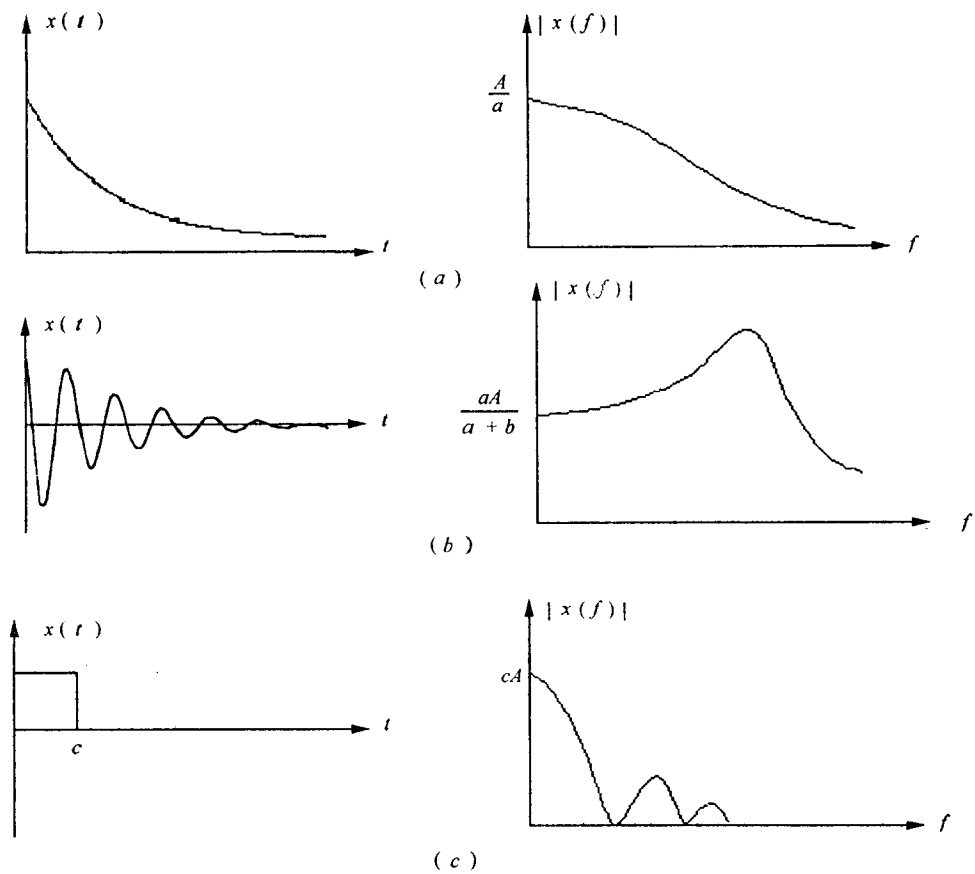
式(1.5)表示一种准周期数据,这是一种非周期数据。实践中,当两个或几个无关联的周期性现象混合作用时,常会产生准周期性的数据。多组螺旋桨推进飞机在发动机不同步时的振动响应,就是一个很好的例子。周期数据的一个重要性质是:若相角 $= 0$ 则式(1.5)可以用离散谱来表征。如图 1.5 所示。这与处理复杂周期性数据一样,其差别只在于分量的各频率不再是有理数关系。

1.1.4 瞬变非周期数据

前面所讨论的准周期以外的非周期数据都属于瞬变数据。与周期和准周期数据不同,瞬变数据有一个重要的特征,就是不能用离散谱表示,在大多数情况下,瞬变数据可以用富里叶积分表示成连续谱:

$$X(f) = \int_{-\infty}^{\infty} x(t) e^{-2\pi f t} dt \quad (1.6)$$

富里叶谱 $X(f)$ 一般是一个复数 $X(f) = |X(f)| e^{j\theta(f)}$, 其中: $|X(f)|$ 是 $X(f)$ 的模, $\theta(f)$ 是幅角。图 1.6 表示三个瞬变数据时间变化曲线和对应的富里叶谱。



$$\begin{aligned}
 (a) \quad x(t) &= \begin{cases} Ae^{-at} & t \geq 0 \\ 0 & t < 0 \end{cases} \\
 (b) \quad x(t) &= \begin{cases} Ae^{-at} \cos bt & t \geq 0 \\ 0 & t < 0 \end{cases} \\
 (c) \quad x(t) &= \begin{cases} A & 0 \leq t \leq c \\ 0 & c < t < 0 \end{cases}
 \end{aligned}$$

图 1.6 瞬变数据及谱

1.2 随机数据的分类

如前所述,表示随机现象的数据不能用精确的数学关系式描述,因为这种现象的每次观察都是不一样的。表示随机现象的单个时间历程,称为样本函数(在有限时间上观察,称为样本记录);随机现象可能产生的全部样本函数的集合,称为随机过程。其分类如图 1.7 所示。

1.2.1 平稳随机过程

如果把一个物理现象看做一个随机过程,这个现象在任何时刻的特性可以用随机过

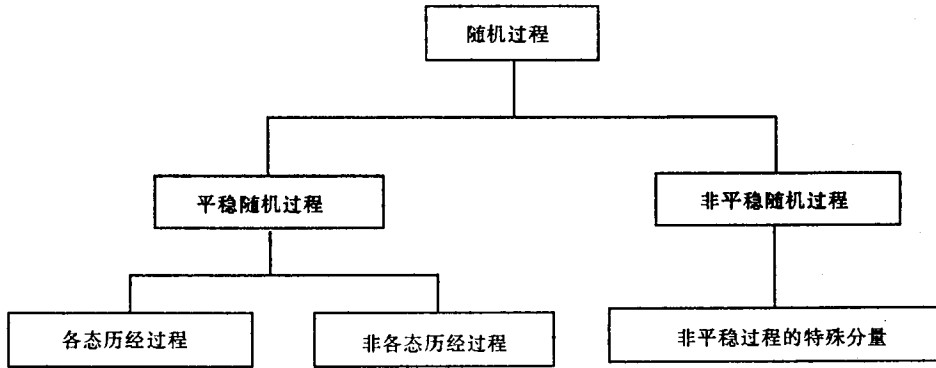


图 1.7 随机数据分类

程样本函数集合 $\{x(t)\}$ (也称总体) 的平均值来描述, 其某一时刻 t_1 的均值(一阶矩)为:

$$\mu_x(t_1) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=1}^N x_k(t_1) \quad (1.7)$$

而随机过程不同时刻之值的相关性(联合矩, 亦称为自相关函数) $R_x(t_1, t_1 + \tau)$ 为:

$$R_x(t_1, t_1 + \tau) = \lim_{N \rightarrow \infty} \sum_{k=1}^N x_k(t_1) \cdot x_k(t_1 + \tau) \quad (1.8)$$

其中最后假定求和时, 各样本函数都有等可能性。一般情况, 若 $\mu_x(t_1)$ 和 $R_x(t_1, t_1 + \tau)$ 都随时间 t_1 的改变而改变, 则称随机过程 $\{x(t)\}$ 是非平稳的。在特殊情况下, 若 $\mu_x(t_1)$ 和 $R_x(t_1, t_1 + \tau)$ 不随时间 t_1 变化, 则称 $\{x(t)\}$ 为弱平稳或广义平稳, 则此时 $\mu_x(t_1)$ 为常数, $\mu_x(t_1) = \mu_x$, $R_x(t_1, t_1 + \tau) = R_x(\tau)$ 。对于高阶矩和联合矩都不随时间变化则称 $\{x(t)\}$ 为强平稳或严格平稳的, 在许多实际应用中, 过程通常具有弱平稳, 往往也可以假设其为强平稳的。

1.2.2 各态历经过程

前面讨论了如何用具体时刻的总体平均来确定随机过程的特征。但是, 在大多数情况下, 也可以用总体中某样本函数的时间来确定平稳随机过程的特性。例如取随机过程中相当于 k 的样本函数, 它的均值和自相关函数分别为:

$$\mu_x(k) = \lim_{N \rightarrow \infty} \frac{1}{T} \int_0^T x_k(t) dt \quad (1.9)$$

$$R_x(\tau, k) = \lim_{N \rightarrow \infty} \frac{1}{T} \int_0^T x_k(t) x_k(t + \tau) dt \quad (1.10)$$

若随机过程 $\{x(t)\}$ 是平稳的, 而且用不同样本函数计算式(1.9)、(1.10)中的 $\mu_x(k)$ 和 $R_x(\tau, k)$ 的结果都一样, 则称此随机过程为各态历经的, 对于各态历经的随机过程, 按时间平均的值和自相关函数(以及所有其他的按时间平均的量)等于相应的随机过程总体平均值, 即 $\mu_x(k) = \mu_x$, $R_x(\tau, k) = R_x(\tau)$ 。注意, 只有平稳过程才能是各态历经的。

各态历经随机过程的所有特征可以用单个样本函数上的时间平均来描述, 因此, 各态历经过程显然是随机过程中很重要的一类。实践中表示平稳现象的随机数据一般都是各

态历经的。因此在大多数情况下,都可以用单个观察到的时间历经来测定平稳随机现象的特性。

1.2.3 非平稳随机过程

若不满足本节 1.2.1 中所叙述的平稳性要求的随机过程,称为非平稳随机过程。除非另有限制,非平稳过程的特性一般是随时间而变化的,只能用组成过程的样本函数的总体瞬时平均来确定,而在实践中,不易得到足够数量的样本记录来精确地测量总体平均性质。这一事实,妨碍了非平稳随机数据实用测量和分析技术的发展。在许多情况下,实际物理现象产生的非平稳数据,可以进一步分成特殊类别,以便简化测量和分析方法。例如,有些随机数据可以用这样的非平稳随机过程 $\{y(t)\}$ 来描述,它们每部样本函数 $y(t) = A(t)x(t)$,其中 $x(t)$ 是来自平稳随机过程 $\{y(t)\}$ 的样本函数, $A(t)$ 是一个确定乘法因子。这就是说,这类数据可以用具有共同确定性时间趋势的样本函数组成的非平稳随机过程来表示。

1.3 随机数据的基本特性

习惯上用四种主要的统计函数表示随机数据的基本特性。它们是:①均方值;②概率密度函数;③自相关函数;④功率谱密度函数。均方值提供数据强度方向的基本描述,概率密度函数提供数据在振幅域内的有关特征,自相关函数和功率谱密度函数分别在时域和频率上提供类似的信息。对于平稳数据,功率谱密度函数不会比自相关函数提供更多的信息,因为两者是互为富里叶变换,但是,两者所给出的信息形式是不同的,在具体应用时,往往用某种形式而不希望用另一种形式。下面仅对平稳随机数据的这些不同特性作概括的定义。假定数据是各态历经的,从而可以用单个样本记录时间来确定这些特性。

1.3.1 均方值(均值和方差)

任何随机数据的一般强度,可以用基本的均方值来描述,它是时间历程平方值的简单平均值。样本时间历程记录 $x(t)$ 的均方值可表示为:

$$\psi_x^2 = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T x^2(t) dt \quad (1.11)$$

均方值的平方根称为均方根值。

我们常常希望将数据看做静态分量(即不随时间变化的)和动态分量(即波动分量)之和。静态分量可以用均值来描述,此均值是所有的简单平均,用公式来表示,均值 μ_x 为:

$$\mu_x = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T x(t) dt \quad (1.12)$$

动态分量可以用方差来描述,方差是关于均值的简单均方值,用公式来表示,方差 σ_x^2 为:

$$\sigma_x^2 = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T (x(t) - \mu_x)^2 dt \quad (1.13)$$

方差的正平方根称为标准差。展开式(1.13),可知方差等于均方差减去均值的平方,即:

$$\sigma_x^2 = \psi_x^2 - \mu_x^2 \quad (1.14)$$

1.3.2 概率密度函数

随机数据的概率密度函数表示瞬时数据值落在某指定范围内的概率。考虑图 1.8 所示的样本时间历程记录 $x(t)$, $x(t)$ 值在 x 和 $x + \Delta x$ 范围内的概率可取 T_x/T 之比得到。这里 T_x 是在观察时间 T 内, $x(t)$ 落在 $(x, x + \Delta x)$ 范围内的总时间。当 T 趋于无穷时,

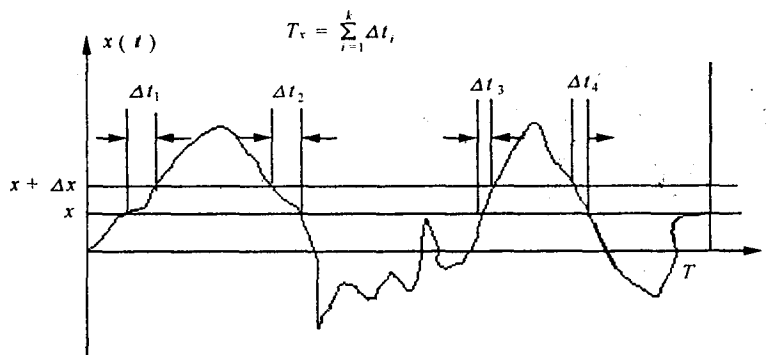


图 1.8 概率测量

此时比值将趋于正确概率值,用公式表示,即:

$$\text{prob}[x < x(t) \leq x + \Delta x] = \lim_{T \rightarrow \infty} \frac{T_x}{T} \quad (1.15)$$

$$\text{prob}[x < x(t) \leq x + \Delta x] \approx p(x)\Delta x \quad (1.16)$$

更精确一些,有:

$$p(x) = \lim_{\Delta x \rightarrow 0} \frac{\text{prob}[x < x(t) < x + \Delta x]}{\Delta x} = \lim_{\Delta x \rightarrow 0} \frac{1}{\Delta x} \left(\lim_{T \rightarrow \infty} \frac{T_x}{T} \right) \quad (1.17)$$

概率密度函数 $p(x)$ 恒为非负实值函数。

瞬时值 $x(t)$ 小于或等于某值的概率定义为 $P(x)$, 它等于概率密度函数从 $-\infty$ 到 x 的积分。函数 $P(x)$ 称为概率分布函数或累积概率分布函数, 不应与概率密度函数混淆。明确一些讲,

$$P(x) = \text{prob}[x(t) < x] = \int_{-\infty}^x p(x) dx \quad (1.18)$$

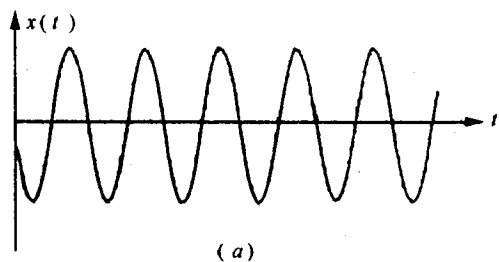
分布函数 $P(x)$ 的值应在 0 和 1 之间, 因为 $x(t)$ 小于 $-\infty$ 的概率显然为 0, 而 $x(t)$ 小于 ∞ 的概率显然为 1。 $x(t)$ 落在任何区域 (x_1, x_2) 内的概率为:

$$P(x_2) - P(x_1) = \text{prob}(x_1 < x(t) \leq x_2) = \int_{x_1}^{x_2} p(x) dx \quad (1.19)$$

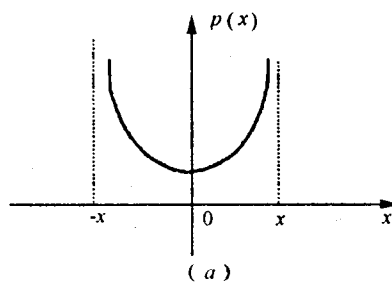
用概率密度函数 $p(x)$ 表示 $x(t)$ 的均值为:

$$\mu_x = \int_{-\infty}^{+\infty} xp(x) dx \quad (1.20)$$

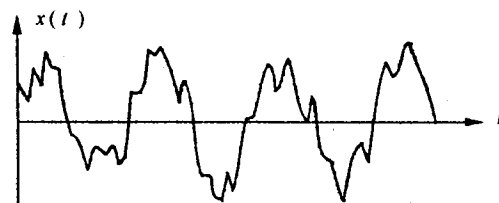
对于很小的 Δx , 可定义概率密度函数 $p(x)$ 如下:



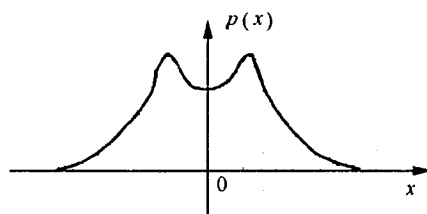
(a)



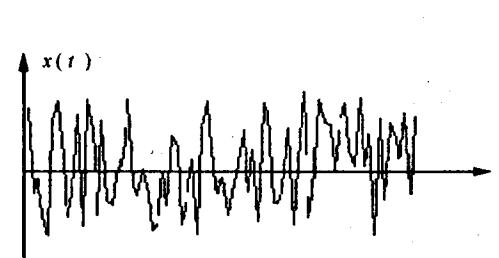
(a)



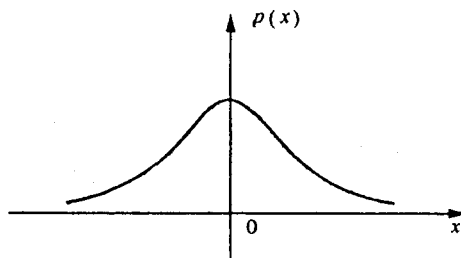
(b)



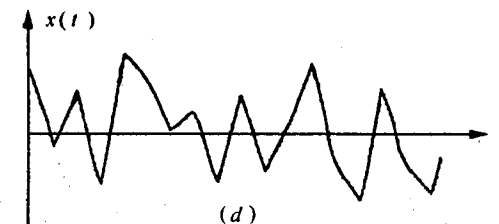
(b)



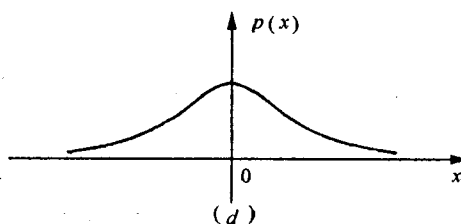
(c)



(c)



(d)



(d)

图 1.9 四个具体时间历程

图 1.10 概率密度函数

这就是说,均值 μ_x 是在所有 x 值上的加权线性和,类似地,均方值为:

$$\psi_x^2 = \int_{-\infty}^{+\infty} x^2 p(x) dx \quad (1.21)$$

它是 $x^2(t)$ 在所有 x 值上加权线性和。

为了帮助理解概率密度函数的实际意义,考虑实践中可能存在的四种样本历程记录:(a)正弦波;(b)正弦波加随机噪声;(c)窄带随机噪声;(d)宽带随机噪声。它们的典型时间历程图如图 1.9 所示。为了方便起见,对所有记录均假设均值为零,即 $\mu_x = 0$ 。

注意,正弦波通常是作为确定性数据,因为它完全可以用方程:

$x(t) = X \cdot \sin(2\pi f_0 t + \theta_k)$ 来描述。但是,正弦波也可以当做随机过程 $\{x(t)\} = \{X \cdot \sin(2\pi f_0 t + \theta_k)\}$ 的样本函数,其中每个样本函数 $x_k(t)$ 的初始相角 θ_k 是一个随机变量,这里就这样来理解正弦波,以使用概率术语来描述它。

图 1.10 给出了这四个例子的概率密度函数 $[p(x)$ 对 $x]$ 。图 1.10(a)画的是正弦波的盆状概率密度,它由下式定义:

$$p(x) = \begin{cases} (\pi \sqrt{X^2 - x^2})^{-1} & |x| < X \\ 0 & |x| \geq X \end{cases} \quad (1.22)$$

图 1.10(c)和(d)所表示的铃状概率密度图,分别代表了窄带,宽带随机数据的典型。这些概率密度取如下的经典高斯形式:

$$p(x) = (\sigma_x \sqrt{2\pi})^{-1} \cdot e^{-\frac{x^2}{2\sigma_x^2}} \quad (1.23)$$

图 1.10(b)给出了正弦波加随机噪声的概率密度,它具有上述两种情况的主要特点。图 1.10 的四个例子示范了从正弦宽带随机噪声各概率密度曲线的明确变化趋势。

实际观察到的数据的概率密度函数,主要用来描述数据瞬时值的概率。从图 1.10 也可以看出,概率密度函数也可能用来区分正弦数据和随机数据。而且各种非线性物理效应也会在概率密度函数中有所表现。

1.3.3 自相关函数

随机数据的自相关函数描述一个时刻的数据值与另一个时刻数据数据值之间的依赖关系。考虑图 1.11 所表示的样本时间历程记录。估计在 t 时刻和 $t + \tau$ 时刻上的自相关性时,可以在观察时间 T 上对两个值的乘积作平均得到。 T 趋于无穷时,平均乘积的极限将接近一个正确的自相关函数。用公式来表示,就是:

$$R_x(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T x(t) \cdot x(t + \tau) dt \quad (1.24)$$

量 $R_x(\tau)$ 恒为实值偶函数,可正可负,且在 $\tau = 0$ 时有最大值。用数学公式表示,就是:

$$R_x(-\tau) = R_x(\tau) \quad (1.25)$$

$$R_x(0) \geq R_x(\tau) \text{ (所有的 } \tau) \quad (1.26)$$

$x(t)$ 的均值可以用自相关函数表示(有包括诸如正弦波的特例)为:

$$\mu_x = \sqrt{R_x(\infty)} \quad (1.27)$$

也就是说: $x(t)$ 的均值等于时间位移变得很大时的自相关函数的正平方根。类似地, $x(t)$ 的均方值:

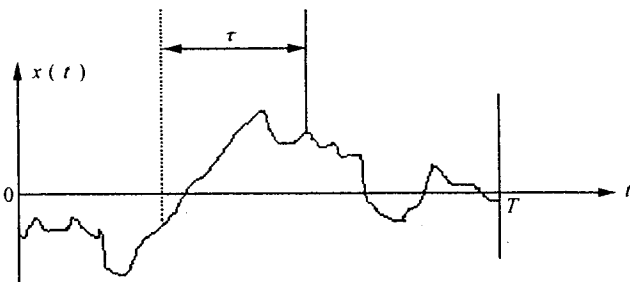


图 1.11 自相关测量