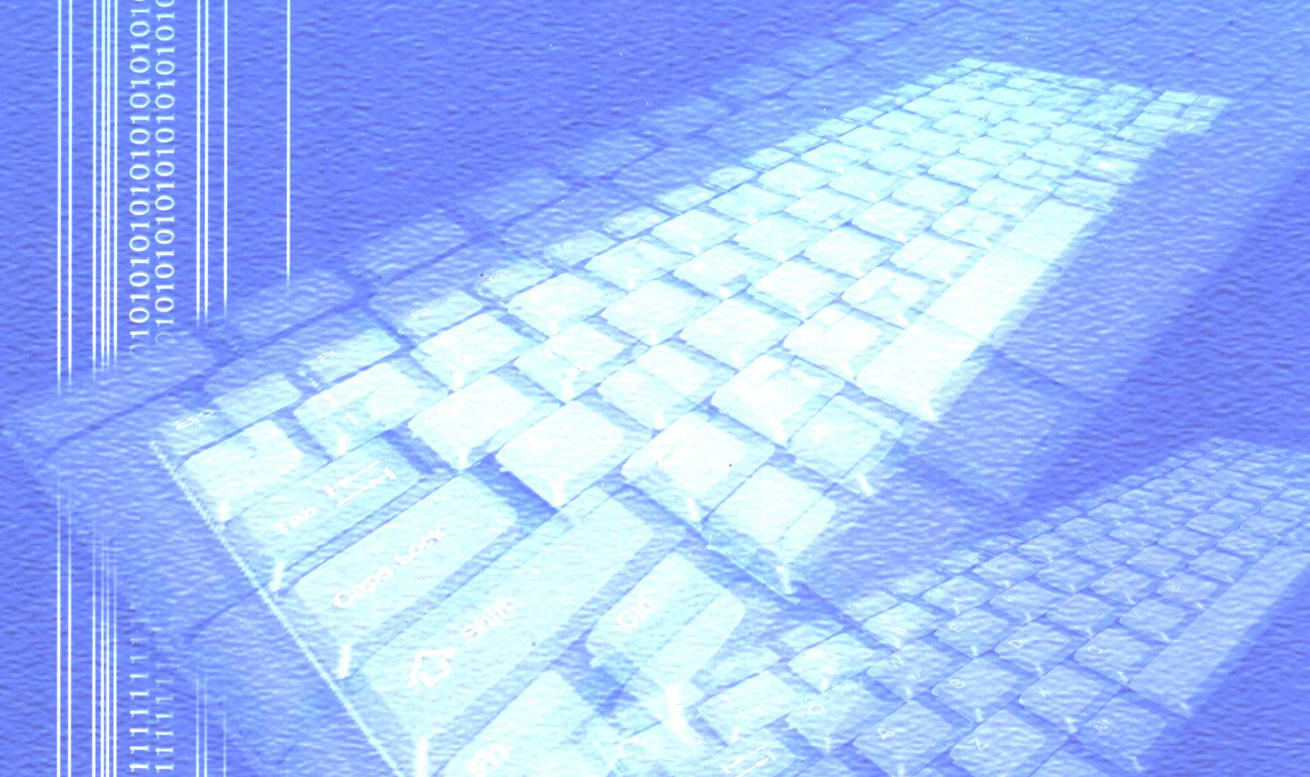
煤炭科学研究总院出版基金
煤炭科学研究总院西安分院 资助出版

地下水系统随机模拟与管理

DIXIASHUI XITONG SUIJI MONI YU GUANLI

虎维岳 李竞生 著



地质出版社

地下水系统随机模拟与管理

虎维岳　李竞生　著

地质出版社

·北　京·

内 容 提 要

本书较全面系统地介绍和论述了地下水系统随机模拟与管理的基本理论和应用技术。第1章介绍了随机问题的基本理论。第2章介绍了常见的随机问题分析与模拟方法。第3章介绍了随机规划理论。第4章介绍了地下水系统确定性模拟与管理的基本理论。第5章介绍了地下水系统随机模拟与管理的基本理论。第6章介绍了地下水系统随机模拟与管理模型的求解方法。第7章通过两个假设实例和一个实际工程介绍了地下水系统随机模拟与管理理论的应用。

图书在版编目（CIP）数据

地下水系统随机模拟与管理/虎维岳，李竞生著.-北京：地质出版社，2003.9
ISBN 7-116-03879-5

Ⅰ.地… Ⅱ.①虎… ②李… Ⅲ.地下水-水流模拟 Ⅳ.P641.2

中国版本图书馆 CIP 数据核字（2003）第060168号

责任编辑：赵俊磊 陈军中
责任校对：田建茹
出版发行：地质出版社
社址邮编：北京海淀区学院路31号，100083
电 话：（010）82324508（邮购部）
网 址：http：//www. gph. com. cn
电子邮箱：zbs@ gph. com. cn
传 真：（010）82310759
印 刷：北京中科印刷有限公司
开 本：787 mm×1092 mm 1/16
印 张：6.125
字 数：140千字
印 数：1—800册
版 次：2003年9月北京第一版·第一次印刷
定 价：20.00元
ISBN 7-116-03879-5/P·2391

前　言

地下水系统模拟与管理作为水文地质学研究的一个重要分支，其研究的对象、内容和技术方法都经历了一个萌芽、发展和完善的过程。这一变化过程都与其研究对象的复杂性和客观要求的准确性密切相关。由于人们对水资源在人类社会可持续发展中重要位置认识的提高以及地下水系统与人类社会活动、环境等因素的密切联系，使得对地下水系统的研究迅速地由对含水层系统行为规律的认识转向对这一系统的科学模拟、管理和调控。同时，由于地下水系统的复杂性、非线性、多尺度性、突变性和随机性等本质属性的存在，使得对地下水系统模拟与管理问题的研究变得十分复杂。

地下水系统模拟与管理问题就是在充分掌握和预知地下水系统的空间结构、埋藏条件以及各种影响和控制其行为规律的自然和人为因素的前提下，进而对其进行科学的、定量的模拟预测、管理和调控，使其向着人们所期望的目标发展和演化。地下水系统的行为规律又取决于与其相关联的外部环境和系统内部参数的时、空变化情况。这些外部环境因素和内部参数场则具有极强的随机性和不确定性。影响地下水系统的不确定性因素基本上可分为两大类：固有的不确定性和信息的不确定性。所谓固有的不确定性是指来自于一定自然性质和自然过程的不确定性（如大气降水等）。信息的不确定性主要是一种“噪声”，或者是对反映地下水系统性质的信息的不完全了解所带来的不确定性。这种不确定性是可以通过多种策略予以减少的。正是由于这些不确定性因素的存在，人们常常将对这类系统的计算机模拟形容为“垃圾输入，垃圾输出”。意思是指模型参数或输入数据的不确定性和误差将导致输出结果的不确定性和误差。前者的不确定性和误差又不可避免地来自于多个方面。首先，表现为模型的概化，由于概念模型是水文地质实体的概化，这一概化过程必将导致一些信息的丢失和忽视。其次，是含水层参数的不确定性和误差。由于描述含水层特征的主要参数（如渗透系数、储水系数、水动力弥散系数等）无法直接测量，而需要通过解译相关的勘探试验资料求取，且含水层一般都是非均质的，用有限的试验资料所求取的参数只能代表整个系统的局部参数值。再次，是输入数据的误差。在地下水系统模拟研究中，所需的多种输入数据（如边界性质和位置、初始条件、激发条件等）都不可避免地带有不确定性和误差。最后，是数值方法本身所具有的区域参数特征化误差、截断误差、数值波动误差等。由此可见，研究对象的客观随机性和不确定性决定了地下水系统模拟与管理模型的随机属性。

对于随机问题的研究主要出现于20世纪60年代中期，其主要原因就是当时摄动方法的兴起和概率论思想的普及。Song 和 Bogdanoff 关于无规则线性链索的研究[1]，采用了传递矩阵技术与摄动展开思想的结合，并给出了考虑随机参数的频率解答。Boyce 和 Goodwin（1964）研究了具有随机参数的弦和梁的特征值问题[2]。在他们的研究中考虑了参数的随机因素并使用了摄动逼近思想。到了20世纪70年代，许多学者将已有的研究成果与有限元技术相结合，形成了后来人们称之为随机有限元法（或概率有限元法）的基本思想，随

之以 Shinozuka 和 lenoe 的研究成果为代表[3~7]，人们开始采用随机模拟的方法对具有随机参数的系统进行研究，进而形成了解决随机问题的又一主导方法——随机模拟方法。到了20世纪80年代，以 Nakagiri 和 Hisada 为代表，对摄动有限元方法及其应用进行系列研究，其研究的热点是形函数不确定性问题[8~10]。随后，在泛函分析思想的框架下发展形成了次序正交分解方法[11~13]、反应量正交展开方法和混沌正交多项式等[14,15]分析和求解随机问题的理论和方法。

随着研究随机系统理论和方法的不断发展和完善，加之水文地质学所研究对象的固有随机性特点，使得随机水文地质学理论和方法也得到了迅速发展和应用，并形成了全分布分析和一阶、二阶矩分析。所谓全分布分析方法就是产生形成地下水系统中不确定性输入或参数的概率分布场并以此产生具体的地下水流概率分布场。一、二阶矩分析方法则是假设随机变量或函数的一、二阶矩完全可以刻画其随机分布特性并基于这一假设，只考虑水头或流量的均值与方差的分析方法。两种最主要的全分布技术就是衍生分布和蒙特卡罗（Monte-Carlo）模拟。衍生分布是一种为获得随机参数（函数）概率分布的分析技术，研究这一问题的代表人物有 Sagar 和 Kisiels T. 及 Eagleson's。蒙特卡罗方法则是一种应用更为广泛的方法，它主要是根据每组随机变量的分布特征产生一组随机分布的伪随机参数作为输入来模拟地下水系统的行为。并对模拟结果进行合成分析，形成水流或水头的估计分布特征。

Michael D 等提出了一阶和二阶矩序列分析，即利用一阶泰勒展开式分析随机变量的均值和方差、利用二阶泰勒展开式分析随机变量的均值、方差及协方差等特征指标。Sergio E.Serrano Soblev 空间探讨了求解地下水流问题的随机偏微分方程的理论和方法[17]。ShuGuang Li 等研究了求解地下水流问题的非平稳随机谱分析方法，并讨论了基于传统 Fourier 变换和数值模拟技术相结合的混合随机问题求解方法[18]。姚磊华博士在其博士学位论文中较系统地分析和讨论了地下水运动和溶质运移问题的随机数值模拟技术。对蒙特卡罗随机有限元法在地下水系统随机数值模拟中的应用问题进行了探讨，并根据蒙特卡罗法的特点，将 Neumann 展开式引入地下水系统的蒙特卡罗有限元模拟中，解决了矩阵求逆的效率问题，同时也讨论了摄动有限元法、Taylor 展开法等技术在随机地下水流模拟及溶质运移问题中的应用方法。

随着人们对地下水系统不确定性和随机性认识的不断深入以及地下水系统随机模拟技术的日趋完善，为随机地下水系统管理模型的建立和求姐奠定了良好的基础，使得随机地下水管理模型的建立和求解技术开始引起人们的广泛重视。地下水管理模型的广泛研究出现于20世纪70年代，其代表人物有 Chaudhry et al.（1974），Bredehoeft（1973）……Taylor 和 Luckey（1974），Bockstock 等（1977）[20~26]。其研究成果主要集中于简单的确定性水力管理模型和策略评价模型。在水力管理模型中形成至今仍被广泛应用的“嵌入法”模型和“响应矩阵法”模型。“嵌入法”和“响应矩阵法”两种方法各有优缺点，它们在具体应用中往往根据具体条件而加以选择。对于“嵌入法”，剖分后的整个地下水流方程都包含于规划管理模型中作为约束条件，使模拟模型作为管理模型的一部分进行求解。整个模拟区及各个模拟阶段的所有节点水头都将作为约束条件，其他条件作为附加约束条件加入到管理模型之中。该方法可以产生大量的关于含水层系统的信息量。许多地下水管理模型本身并不要求知道含水层的所有信息。因此，许多决策变量和约束条件无须加入到地下水系统

的规划与管理模型之中。为了减少计算工作量和提高模型效率，“嵌入法”较适合于小范围的稳定流管理问题。“响应矩阵法”是将水流模拟模型的一部分作为管理模型的约束条件。该方法只把管理模型所需要的信息量从模拟模型中耦合进来，因而其计算量小，方法简单，能有效地处理大规模的非稳定流含水层系统的管理问题。

进入 20 世纪 80 年代，地下水管理模型迅速地发展为动态规划、多级规划、多目标规划、整数规划等。Gorelick 等首先研究了地下水流与溶质运移问题相结合的非线性最优化问题[27]。1992 年，Chang 等应用有约束最优化控制算法形成了时变地下水质控制与净化策略。1994 年，Wang 和 Alfeld 以井位作为显式决策变量探讨了疏干孔群的优化设计问题[28]。

到了 20 世纪 90 年代后期，地下水管理模型的建模技术和求解方法有了质的飞跃，人们将研究工作的重点越来越多地集中于带有随机变量的随机地下水管理模型和具非线性约束与目标函数的非线性地下水管理模型的建模和求解技术。对于传统的地下水系统控制与管理问题，在求解过程中总希望目标函数和约束方程有明确的数学表示，并强烈地依赖于梯度求解过程。对于许多问题，求解梯度总是一件比较困难的事情，当以井位作为决策变量时，求导的函数在求解域内是非连续的，这一切都使得求导运算变得非常困难，甚至不可能。对于越来越多的非线性模型，随机模型和模糊模型，基于求导运算的优化算法很难保证其解收敛到全局最优点。因此，在地下水管理模型及求解技术方面人们正在借助于其他学科的研究成果寻求新的理论和方法。借助于遗传算法的研究成果，Dougherty 和 Marryott 采用模拟退火技术求解了地下水污染净化和控制的优化问题。他们在二维非均质各向异性含水层中利用模拟退火技术优化在预先指定的孔位中抽水强度的分配问题。Mc Kinney 和 Lin 利用人工遗传算法求解了稳态地下水管理模型的全局最优解[29]。Brian J. Ritzel 和 J. Wayland 将人工遗传算法与人工神经网络算法相结合求解了具有线性目标函数的含水层系统水质管理问题，并将该方法与基于梯度函数的传统 Neuton's 算法进行了比较[42]。

正是基于对水资源进行科学预测、模拟、管理和调配研究重要性的认识，并能够深刻把握地下水系统及其相关环境因素具有随机性和不确定性这一本质特点，本书较全面系统地介绍和论述了地下水系统随机模拟与管理的基本理论和应用技术。其中第 1 章介绍了随机问题的基本理论。第 2 章介绍了常见的随机问题分析与模拟方法。第 3 章介绍了随机规划理论。第 4 章介绍了地下水系统确定性模拟与管理的基本理论，该部分内容作为地下水系统随机模拟与管理的基础，本书引用了卢文喜、姚磊华等的研究成果。第 5 章介绍了地下水系统随机模拟与管理的基本理论。第 6 章介绍了地下水系统随机模拟与管理模型的求解方法。第 7 章通过两个假设实例和一个实际工程介绍了地下水系统随机模拟与管理理论的应用技术。

本书在编写过程中得到了姚磊华博士指教，张壮路、赵杨工程师、闫兰英高级工程师等为全书的制图、打印等做了大量的工作，在此一并表示诚挚的谢意。

限于作者的学识与水平，书中错漏之处难免，恳请读者指教。

作　者

2002 年于煤炭科学研究总院西安分院

第1章　随机问题的基本理论

对于一个客观的物理过程、物理现象，建立其行为规律的数学模型以进行定量分析和定量预测是现代工程分析与模拟的主要方法。无论是模拟问题、预测问题或是优化问题，数学模型的建立总是在一定的假定条件下来建立影响事物发生、发展和变化的某些因素之间的因果关系，进而通过一定的已知信息来获得事物发生、发展和变化的内在规律，以实现认识事物、预见事物和控制事物。

在工程设计和工程实践中，不管我们获取资料的手段如何，不管我们所获得的信息量有多大，总不能完全准确而毫无遗漏地反映事物的全部。对于地下水系统而言，由于其复杂性、时变性和不可完全揭示性，使得这种不确定性和信息的不完整性显得尤为突出。我们的预测和决策显然都是在不确定性条件下所进行，因而预测和决策的结果就不能完全置信。这种不确定性除了根据相似条件下的观测和模型实验的结果所推导出的数据具有某种程度的不完善以外，许多与工程行为本身密切相关的自然因素本身就具有随机的性质，当然是属于不确定性因素，这就决定了处理问题方法和观点是不确定性的，而不是确定性的。

影响事物因素的不确定性和处理问题方法的不完善性对生产、设计与决策带来的影响往往是不可忽视的。在工程设计、系统行为预测及工程决策过程中要对不确定性因素进行定量计算和评估就必须用到基于概率论的随机方法。同样，其设计与预测、决策的结果只能是概率意义上的不确定性结论而不是确定性结论，正像天气预报工作一样，预报者只能告诉我们未来天气降水的概率有多大，而不能简单地采用“是”或“非”来回答未来天气的降水情况。

1.1　水文地质中常见的不确定性问题

1.1.1　来自随机性的不确定性

在研究一个水文地质问题时，总会遇到大量的和我们所研究问题密切相关的随机性因素，即事物实际发生的结果(在一定程度上)具有不可预见性。事物的发生具有不可重复性，每次观测结果存在着一个范围，在这一范围内事物每次发生的结果不尽相同，如某地区每月的降水强度在不同年份中的变化情况(图1.1至图1.3)虽在一定的范围内遵循大致相同的变化规律，但不完全相同。

某河流流量及其水位标高的变化情况(图1.4至图1.5)。

1.1.2　来自量测误差的不确定性

一定的信息总是来自特定的量测仪器和量测方法，不同的量测仪器和量测方法往往会获得对同一物理量的不同量测结果。这种来自量测误差的不确定性在实际工程中屡见不鲜。如某矿井在放水试验时来自不同量测方法所获得的放水量结果(图1.6)。

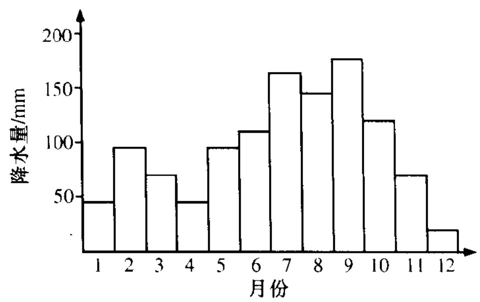


图 1.1　1990 年月平均降水量

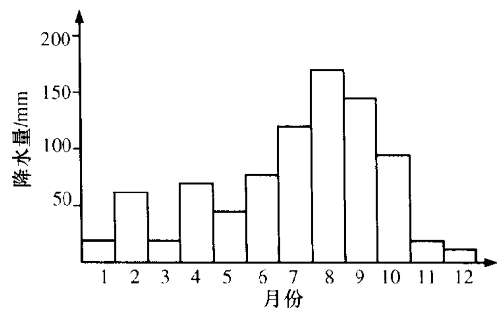


图 1.2　1992 年月平均降水量

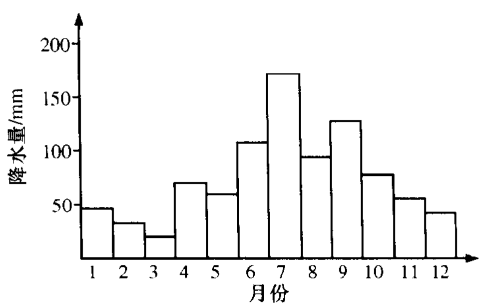


图 1.3　1995 年月平均降水量

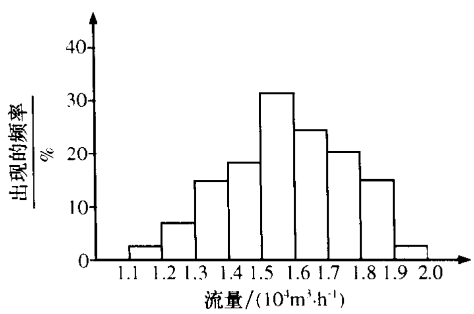


图 1.4　某河流 6 月份流量频率分布图

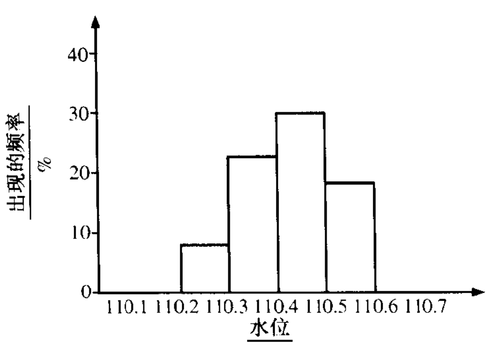


图 1.5　某河流 6 月份水位标高频率分布图

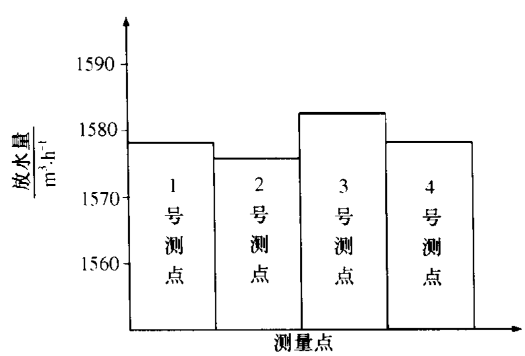


图 1.6　不同测点水量测量结果直方图

1.1.3　模型不完善性带来的估计误差不确定性

工程实际问题中的不确定性，并不完全由前述的不确定性引起，即观测中基本变量的可变性所引起，特别是像水文地质中常用的数学模拟、参数识别、预测预报、管理与决策之类的工程问题，往往使某些变量的值取决于具有不确定性的观测数据，这样误差就不可避免（尤其在数据有限的情况下），基于不同次的观测资料所获的参数结果自然会不相同，如表 1.1 所示某水源地利用不同年份观测资料所获的水文地质参数相差极大（计算方法完全相同）。

表 1.1 依不同年份资料所获的水文地质参数反演结果表

参数	参数分区号	1991 年资料	1992 年资料	1993 年资料	1994 年资料	1995 年资料	μ_{XY}	$D(\cdot)=\delta^2$
	1	463.093	486.954	507.046	487.259	460.473	480.965	1495.029
	2	53.915	56.693	59.032	56.729	53.610	55.996	20.264
	3	301.348	316.874	329.949	317.073	299.642	312.977	633.064
	4	374.518	393.815	410.064	394.062	372.399	388.972	977.818
	5	97.240	102.250	106.469	102.314	96.690	100.993	65.918
	6	46.213	48.594	50.599	48.625	45.952	47.997	14.888
	7	347.561	365.468	380.548	365.697	345.594	360.974	842.120
K_X	8	462.131	485.941	505.992	486.246	459.516	479.965	1488.819
	9	74.133	77.953	61.170	78.002	73.714	76.994	38.312
	10	253.209	266.255	277.241	266.422	251.776	262.981	446.962
	11	463.093	486.954	507.046	487.259	460.473	480.965	1395.029
	12	465.017	488.979	809.154	489.285	462.387	480.965	1507.487
	13	462.131	485.941	505.992	486.246	459.516	479.965	1488.819
	14	405.327	426.211	443.797	426.478	403.033	420.969	1145.313
	15	459.242	482.904	502.830	483.207	456.644	476.965	1470.267
	1	466.944	491.003	511.263	491.311	464.302	484.965	1519.998
	2	378.369	397.865	414.281	398.114	376.149	392.971	998.001
	3	336.045	352.308	366.844	352.528	333.149	347.975	782.560
	4	296.662	417.100	434.310	417.361	394.417	411.970	1096.868
	5	117.458	123.510	128.606	123.587	116.794	121.991	96.179
	6	48.139	50.619	52.708	50.651	47.866	49.996	16.155
	7	48.139	50.619	52.708	50.651	47.866	49.996	16.155
K_Y	8	69.320	72.891	75.899	72.937	68.927	71.995	33.498
	9	347.561	365.468	380.548	365.697	345.594	360.974	842.120
	10	464.056	487.966	508.100	488.272	461.430	481.965	1501.252
	11	401.176	422.162	439.581	422.426	399.204	416.970	1123.652
	12	165.619	488.979	509.154	489.285	462.387	482.965	1507.487
	13	460.205	483.917	503.884	484.220	457.601	477.965	1476.437
	14	462.131	485.941	505.992	486.246	459.516	479.965	1488.819
	15	98.203	103.263	107.523	103.327	97.647	101.993	67.229
	1	0.077	0.081	0.084	0.081	0.077	0.080	0.000
	2	0.077	0.081	0.084	0.081	0.077	0.080	0.000
	3	0.087	0.091	0.095	0.091	0.086	0.090	0.000
	4	0.087	0.091	0.095	0.233	0.086	0.090	0.000
	5	0.221	0.233	0.242	0.274	0.220	0.230	0.000
	6	0.260	0.273	0.285	0.284	0.258	0.270	0.000
	7	0.270	0.283	0.295	0.284	0.268	0.280	0.001
μ	8	0.270	0.283	0.295	0.294	0.268	0.280	0.001
	9	0.279	0.294	0.306	0.284	0.276	0.290	0.001
	10	0.270	0.283	0.295	0.284	0.268	0.280	0.001
	11	0.270	0.283	0.295	0.182	0.268	0.280	0.001
	12	0.173	0.182	0.190	0.284	0.172	0.180	0.001
	13	0.270	0.283	0.295	0.284	0.268	0.280	0.001
	14	0.270	0.283	0.295	0.284	0.268	0.280	0.001
	15	0.270	0.283	0.295	0.284	0.268	0.280	0.001

注：μ_{XY}为参数均值；$D(\cdot)=\delta^2$ 为参数方差。

同时，为了简化处理一些问题而建立的物理模型或数学模型（如公式、方程、算法、计算模拟程序等），甚至一些模型实验都是实际问题的某种理想化代表，因此总会存在某种程度的不准确性或信息不完善性。这也是构成问题不确定性的重要因素之一。诸如上述种种原因所造成的不确定性问题都需要采用基于概率论与随机理论的方法来处理。

1.2 概率与数理统计理论的基本概念

当讨论到不确定性问题时，总会涉及概率的概念，即某一事件相对于其他事件发生的可能性，也就是说某事件至少有一种以上发生的可能性，否则，问题将变成确定性问题。概率即是某一事件的发生相对于一切其他事件的发生的量的度量。因此，构成概率问题的先决条件是必须明确问题发生的所有可能性，即所谓可能性空间以及该空间的事件。

1.2.1 随机事件与样本空间

不确定性事件发生的所有可能性结果的集合构成了随机事件发生的样本空间，而样本空间中的每一个具体结果叫做该样本空间的随机事件。要深刻理解概率的概念，必须先知道频率的有关性质。一般地，设随机事件 A 在 n 次试验或观测中出现的次数为 n_A，则称

$$f_n(A) = n_A/n \tag{1.1}$$

为事件 A 在这 n 次试验或观测过程中出现的频率。事件 A 在多次观测中出现的频率虽为一个变数，但对多种物理现象的观测表明，当试验或观测的次数 n 逐渐增多时，$f_n(A)$在一个常数附近摆动，且逐渐稳定于这个常数，也就是说频率具有稳定性的性质。频率的稳定性性质对于我们认识随机现象的内在规律性，预测事物和控制事物具有重要意义。

对于样本空间 S 中的随机事件A，n 次试验中的频率具有下列性质。

(1) $0 \leqslant f_n(A) \leqslant 1$

(2) $f_n(S) = 1$

基于对频率概念的理解，假设 E 是一次随机试验，S 是试验的所有样本空间，对于试验的每个具体事件 A 赋予一个实数$P(A)$，则称 $P(A)$为事件 A 发生的概率，如果满足下列条件：

(1) $0 \leqslant P(A) \leqslant 1$

(2) $P(S) = 1$

(3) 对于两两不相容的事件 $A_k(k = 1,2,\cdots)$有：

$P(A_1 \cup A_2 \cup \cdots \cup A_n \cup \cdots) = P(A_1) + P(A_2) + P(A_3) + \cdots + P(A_n) + \cdots$

则称概率具有可列可加性。有关概率的运算法则参见文献[53]。

1.2.2 随机变量

为了全面研究随机事件和分析随机问题的内在规律性，揭示客观世界存在的不确定性或随机性问题的统计规律性，有必要了解随机变量的基本概念。

设 E 为随机试验，它的样本空间是 $S = \{e\}$。如果对于样本空间中的某个具体随机事件 $e \in S$ 有一个实数$X(e)$与之对应，这样，对于空间 S 中的每一个e 总有一个实值单值函数 $X(e)$，也就是产生了 S 与$X(e)$之间的函数对应关系，称 $X(e)$为随机变量。

设 X 为 $X(e)$所有可能取值的全体，则有下列示意图关系(图 1.7)：

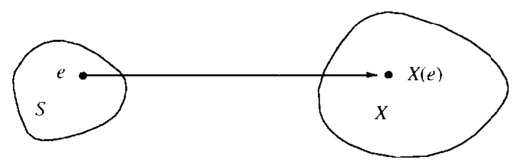


图 1.7

由于随机变量是随机事件的函数，随机事件的发生具有一定的概率。于是，随机变量的取值也有一定的概率，这一性质显示了随机变量与普通函数之间有着本质的差异，且普通函数是定义在实数轴上而随机变量则是定义在样本空间上的(样本空间元素不一定是实数)。

在样本空间 $S=\{e\}$上定义一个实值函数以便形成一个随机变量是分析随机问题常见的事情。如表 1.1 所示的水文地质参数就是一组随机变量，它是实现一次水文地质数据观测(一个随机事件)，根据一定的函数关系便可得到一组水文地质参数(随机变量)。随机变量的引入，主要是为了帮助我们利用数学分析的方法来分析和研究随机问题。

随机变量可分为离散型随机变量和连续型随机变量两种。所谓离散型随机变量是指其全部可能取到的值是有限多个或是可列无限多个。

一般地，设离散型随机变量 X 所有可能取的值为 $x_k(k=1,2,\cdots)$，X 取每个可能值的概率为：

$$P\{X=x_k\}=P_k \qquad k=1,2,\cdots \tag{1.2}$$

则 P_k 应满足下列两个条件：

(1) $P_k\geqslant 0 \qquad k=1,2,\cdots$

(2) $\sum\limits_{k=1}^{\infty}P_k=1$

式 $P\{X=x_k\}=P_k$ 称为离散型随机变量的概率分布或分布律，常见的离散型随机变量的概率分布有如下几种。

(1) (0-1)分布。对于一个随机事件可能发生的结果只有两种，即其样本空间只包含有两个元素 $S=\{e_1,e_2\}$，我们定义随机变量

$$X=x(t)=\begin{cases}1, & \text{当 } e=e_1\\ 0, & \text{当 } e=e_2\end{cases} \tag{1.3}$$

来描述和刻画这类随机问题，称其为(0-1)分布。

(2) 二项分布。设随机事件只有两种可能的结果，$S=\{e_1,e_2\}$，如事件 e_1 发生的概率为 p，则事件 e_2 发生的概率为 1-p，即有 $P\{x=e_1\}=p$

$$P\{x=e_2\}=1-p=q \tag{1.4}$$

如果将上述随机问题做 n 次贝努利试验，则事件 e_1 可能发生 $0,1,2,\cdots,n$ 次。通过计算不难发现事件 e_1 恰好发生 $k(0\leqslant k\leqslant n)$次的概率为：

$$P_n(k)=\binom{n}{k}p^k\cdot q^{n-k} \quad k=1,2,\cdots,n \tag{1.5}$$

注意到$\binom{n}{k}p^k\cdot q^{n-k}$刚好是二项式$(p+q)^n$ 的展开式中的第 $k+1$ 项，故我们称随机变量 X 服从参数 n,p 的二项分布，记为 $X\sim B(n,p)$。

(3) 泊松分布。设随机变量 X 所有可能取的值为 $0,1,2,\cdots$且取第 k 个值的概率为 $P\{x=k\}=\dfrac{\lambda^k\cdot \mathrm{e}^{-\lambda}}{k!}$，$k=0,1,2,\cdots$其中 $\lambda>0$ 是常数，则称 X 服从参数为λ 的泊松分布。记

为 $X \sim \pi(\lambda)$。 (1.6)

连续型随机变量及其概率密度：设有随机变量 X，它的分布函数为 $F(X)$，如存在有非负的函数 $f(x)$，使对于任意实数有：

$$F(X) = \int_{-x}^{x} f(t)\mathrm{d}t \tag{1.7}$$

则称 X 为连续型随机变量，$f(x)$称为 X 的概率密度函数。可简称为概率密度。$F(X)$称为 X 的分布函数。连续型随机变量的分布函数也是连续函数。

概率密度函数反映了样本空间中个别具体随机事件发生的相对概率大小，而随机变量的分布函数则反映了随机事件在某一特定的区域或时间域中出现的概率大小情况，概率密度函数 $f(x)$具有下列基本性质。

(1) $f(x) \geqslant 0$

(2) $\int_{-\infty}^{+\infty} f(x)\mathrm{d}x = 1$

(3) $P\{x_1 \leqslant X \leqslant x_2\} = F(x_2) - F(x_1) = \int_{x_1}^{x_2} f(x)\mathrm{d}x$

图 1.8 至图 1.11 反映了随机变量的概率密度函数与概率分布函数的基本意义。

几种常见的重要连续型随机变量分布有以下几种。

(1) 均匀分布。如果连续型随机变量 X 在某一特定区间(a,b)内取值，且其概率密度函数为：

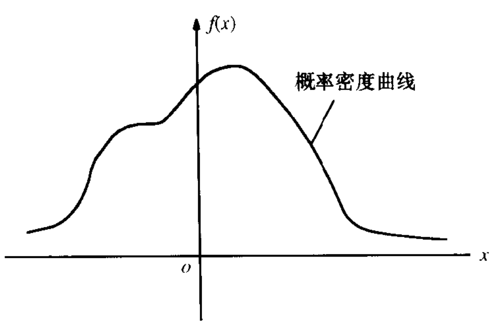


图 1.8

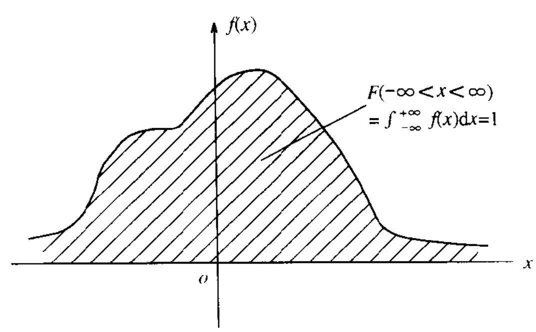


图 1.9

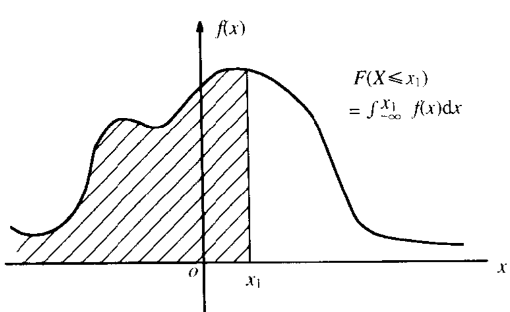


图 1.10

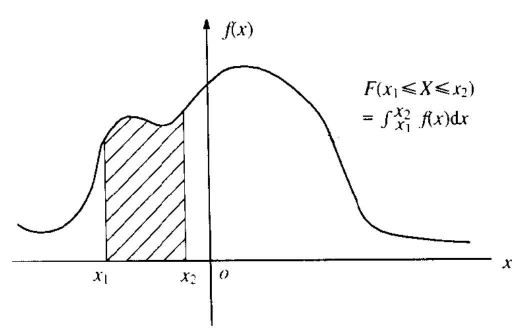


图 1.11

$$f(x)=\begin{cases}\dfrac{1}{b-a} & a<x<b\\ 0 & \text{其他}\end{cases} \tag{1.8}$$

则称 X 在 (a,b) 上服从均匀分布，其分布函数为：

$$F(x)=\begin{cases}0 & x<a\\ \dfrac{x-a}{b-a} & a\leqslant x<b\\ 1 & x\geqslant b\end{cases} \tag{1.9}$$

(2) 正态分布。如果连续型随机变量 X 的概率密度为：

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad -\infty<x<+\infty \tag{1.10}$$

式中：μ,σ——常数。X——服从参数为 μ,σ 的正态分布。具正态分布的随机变量的密度函数和分布函数典型示意图如图 1.12 与图 1.13。

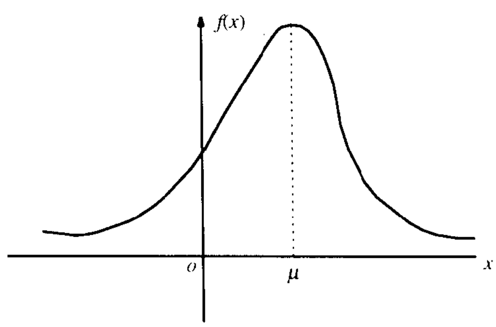


图 1.12

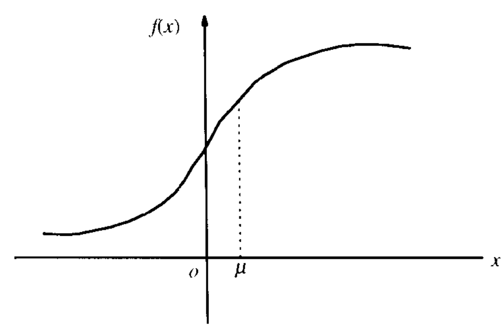


图 1.13

由式(1.10)与图 1.12 可知，μ 和 σ 是刻画正态分布随机变量的重要参数，μ 反映了随机变量在 $(-\infty,+\infty)$ 上出现的最大概率位置，而 σ 则反映了随机变量在 $(-\infty,+\infty)$ 上围绕以 μ 为中心的位置出现的集中程度，当 $\mu=0,\sigma=1$ 时，称 X 服从标准正态分布，其概率密度和分布函数可分别表示为：

$$\phi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}} \tag{1.11}$$

$$\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{t^2}{2}}dt \tag{1.12}$$

1.2.3 随机变量的数字特征

虽然一个随机变量的概率密度函数或分布函数能很好地描述和刻画随机变量的基本特征，但对于生产实践中所遇到的随机变量往往很难知道其具体的分布函数式，然而通过对随机变量的统计分析，会得到一些反映随机变量性质的重要的数字特征，如数学期望、方差、矩等。

若离散型随机变量 X 的分布律为：

$$P\{X=x_n\}=P_x \qquad k=1,2,\cdots$$

且 $\sum_{k=1}^{\infty} x_k P_k$ 绝对收敛，则称 $E(X) = \sum_{k=1}^{\infty} x_k P_k$ 为该随机变量的数学期望。

若 X 为连续型随机变量，其概率密度函数为 $f(x)$ 且积分

$$\int_{-\infty}^{\infty} Xf(x)\mathrm{d}x \text{ 绝对收敛，则称}$$

$$E(X) = \int_{-\infty}^{\infty} Xf(x)\mathrm{d}x \text{ 为随机变量 } X \text{ 的数学期望。}$$

由上述随机变量数学期望的定义可见，其物理意义相当于加权平均值。对于随机变量的函数的数学期望定义与随机变量的数学期望类同，随机变量的数学期望具有下列重要性质：

(1) 设 C 为常数，则 $E(C) = C$

(2) 设 X 为随机变量，C 为常数，则 $E(CX) = C \cdot E(X)$

(3) 设 X, Y 为任意两个随机变量，则 $E(X + Y) = E(X) + E(Y)$

(4) 设 X, Y 是两个相互独立的随机变量，则有：$E(X \cdot Y) = E(X) \cdot E(Y)$

随机变量的均值只反映了随机变量的平均水平，但对随机变量的每一个具体个体偏离平均水平的程度难以刻画，为了研究和分析随机变量偏离其均值的程度，需要引入随机变量方差的概念。

设 X 是一个随机变量，若 $E\{[X - E(X)]^2\}$ 存在，则称 $E\{[X - E(X)]^2\}$ 为 X 的方差，记为 $D(X)$ 或 $\mathrm{var}(X)$ 即：

$$D(X) = \mathrm{var}(X) = E\{[X - E(X)]^2\}$$

由上述公式不难看出方差实际上是平方差的概念，如果对方差开平方根，便可得到均方差或标准差，记为 $\sigma(X)$ 即：

$$\sigma(X) = \sqrt{D(X)}$$

关于随机变量方差的计算有下列重要公式：

$$D(X) = E(X^2) - [E(X)]^2 \tag{1.13}$$

随机变量的方差具有下列重要性质：

(1) 设 C 为常数，则 $D(C) = 0$

(2) 设 X 为一随机变量，C 为常数，则 $D(CX) = C^2 D(X)$

(3) 设 X, Y 为两个相互独立的随机变量，则有 $D(X + Y) = D(X) + D(Y)$

几种常见分布的随机变量的数字特征如表 1.2。

1.2.4 协方差与相关系数

前节介绍了一个随机变量的有关数字特征，但在实际工程中，往往是两个甚至两个以上的随机变量共存，且不同随机变量之间具有某种不同程度的关联性。为了研究不同随机变量之间的相互关系，需要了解协方差和相关系数的概念。

设 X, Y 为随机变量，则 X, Y 之间的协方差为：

$$\mathrm{cov}(X, Y) = E\{[X - E(X)] \cdot [Y - E(Y)]\} \tag{1.14}$$

而

$$\rho_{XY} = \frac{\mathrm{cov}(X, Y)}{\sqrt{D(X)} \cdot \sqrt{D(Y)}} \tag{1.15}$$

表 1.2

分布形式	分布律或概率密度	数学期望	方差
(0-1)分布	$P\{x=k\}=p^k\cdot q^{1-k}$, $k=0,1\quad 0<p<1\quad p-q=1$	p	$p\cdot q$
二项分布	$P_n(k)=\binom{n}{k}p^k\cdot q^{n-k}$, $k=1,2,\cdots,n\quad 0<p<1\quad p+q=1$	np	npq
泊松分布	$P\{x=k\}=\dfrac{\lambda^k\cdot e^{-\lambda}}{k!}$ $k=0,1,2,\cdots\lambda>0$	λ	λ
均匀分布	$f(x)=\begin{cases}\dfrac{1}{b-a} & a\leqslant x\leqslant b\\ 0 & 其他\end{cases}$	$\dfrac{a+b}{2}$	$\dfrac{(b-a)^2}{12}$
正态分布	$f(x)=\dfrac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ $-\infty<x<+\infty,\sigma>1$	μ	σ^2
威布尔分布	$f(x)=\begin{cases}\dfrac{\beta}{\eta}\left(\dfrac{x}{\eta}\right)^{\beta-1}\cdot e^{-\left(\frac{x}{\eta}\right)^{\beta}} & x>0\\ 0 & 其他\end{cases}$	$\eta\Gamma\left(\dfrac{1}{\beta}+1\right)$	$\eta^2\left\{\Gamma\left(\dfrac{1}{\beta}+1\right)-\left[\Gamma\left(\dfrac{1}{\beta}+1\right)\right]^2\right\}$

为 X,Y 的相关系数或标准协方差,协方差还有下列计算公式

$$\operatorname{cov}(X,Y)=E(X,Y)-E(X)\cdot E(Y) \tag{1.16}$$

协方差具有下列性质:

(1) $\operatorname{cov}(X,Y)=\operatorname{cov}(Y,X)$

(2) $\operatorname{cov}(aX,bY)=ab\cdot\operatorname{cov}(X,Y)$

(3) $\operatorname{cov}(X_1,X_2,Y)=\operatorname{cov}(X_1,Y)+\operatorname{cov}(X_2,Y)$

同样,对于随机变量 X,Y,若有 $E(X^k),k=1,2,\cdots$存在,则称其为 X 的 k 阶原点矩。若有 $E[(X-E(X^k))],k=1,2,\cdots$存在,则称其为 X 的 k 阶中心矩。若有 $E(X^k,Y^l),k,l=1,2,\cdots$存在,则称其为 X 和 Y 的 $k+l$ 阶中心混合矩。上述关于随机变量的矩的概念的引入,不难看出随机变量 X 的数学期望 $E(X)$就是 X 的一阶原点矩,而其方差就是二阶中心矩,协方差是随机变量的二阶中心混合矩。

二维随机变量的有关性质可以直接推广至 n 维随机变量,其中最常用的有 n 维随机变量的协方差阵:

设$(X_1,X_2,\cdots,X_n)$为 n 维随机变量,其两两变量间的二阶中心矩为:

$$C_{ij}=\operatorname{cov}(X_i,X_j)=E\{[X_i-E(X_i)][X_j-E(X_j)]$$

则称矩阵:

$$C = \begin{pmatrix} C_{11} & C_{12} & \cdots & C_{1n} \\ C_{21} & C_{22} & \cdots & C_{2n} \\ \vdots & & & \\ C_{n1} & C_{n2} & \cdots & C_{nn} \end{pmatrix}$$

为 n 维随机变量的协方差矩阵。由随机变量协方差的性质 $C_{ji}=C_{ij}$知，矩阵 C 为一个对称矩阵。

1.3 随机过程的基本概念

在客观世界中有些随机现象表示的是事物随机变化的过程，不能用随机变量或随机矢量来描述，而需要用一族无限多个随机变量来描绘，这就是随机过程。

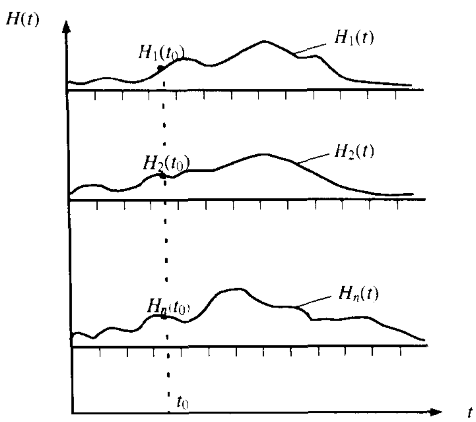


图 1.14

随机变量是指在同一条件下，事件每次发生的结果是随机的、不确定的，而随机过程是指在同样条件下，事物发生的某一过程是随机的、不可准确预知的。一个过程可能是由无限多个随机变量构成，而随机过程是由一族过程（随机出现的）构成的。如对某一个钻孔的水位进行连续观测，以 $H_0(t)$来表示水位，在第一个水文年观测到的水位曲线为 $H_1(t),\cdots$，在第 n 个水文年里观测到的水位为 $H_n(t)$，每个水文年里所得到的样本曲线都是随机的（图 1.14）。$\{H(t), t\in(0,\infty)\}$，怎样理解为由一族随机变量构成的呢？我们固定某一观测时间 t_0，考察 $H(t)$在每年 t_0 时刻的水位值 $H_1(t_0), H_2(t_0),\cdots,H_n(t_0)$，显然 $H(t_0)$是一个随机变量，而当 t 变化时，$H(t)$是一族随机变量。因此，$H(t)$是一个随机过程。

同样的道理，一个地区大气降水的过程，某条河流的流量或河水位变化过程都可看成是一个随机过程。由此可见，设$\{X(t), t\in T\}$为一随机过程，一次过程的观测可以视为随机过程的一个样本函数 $X_1(t)$，第 i 次过程的观测可视为随机过程的第 i 个样本函数 $X_i(t)$。n 次试验观测的结果可得样本函数 $X_1(t), X_2(t),\cdots,X_n(t)$。对于随机过程 $X(t)$，当 t 固定时，为一个随机变量，即随机过程在 t 时刻的状态。随机变量 $X(t), t\in T$（t 固定）的所有可能取的值构成一个实数集，称为随机过程的状态空间或值域，而每一个可能取的值称为一个状态。

随机过程可根据参数集 T 和状态空间的情况进行分类，一般地随机过程可分为下列四类：

(1) 离散参数、离散状态随机过程。

(2) 离散参数、连续状态随机过程。

(3) 连续参数、离散状态随机过程。

(4) 连续参数、连续状态随机过程。

1.3.1 有限维分布族

随机过程$\{X(t),t\in T\}$在每一时刻 t 的状态是一维随机变量,在任意两个时刻的状态是二维随机变量。随机过程的统计特征可用其不同固定时刻的不同维随机变量(矢量)的分布的全体来表示。

对任意固定的 $t\in T$,

$$F(x,t)=P\{X(t)\leqslant x\}$$

称为随机过程 $X(t)$在 t 时刻的一维分布函数。

对于任意两个固定的 $t_1,t_2\in T$

$$F(x_1,x_2,t_1,t_2)=P\{X(t_1)\leqslant x_1,X(t_2)\leqslant x_2\}$$

称为随机过程 $X(t)$的二维分布函数。

一般地,对于任意固定的 $t_1,t_2,\cdots,t_n\in T$,

$$F(x_1,x_2,\cdots,x_n,t_1,t_2,\cdots,t_n)=P\{X(t_1)\leqslant x_1,X(t_2)\leqslant x_2,\cdots,X(t_n)\leqslant x_n\}$$

称为随机过程 $X(t)$的 n 维分布函数。

随机过程 $X(t)$的一维分布函数,二维分布函数,…,n 维分布函数的全体称为随机过程的有限维分布函数族。

1.3.2 随机过程的数字特征

随机过程的数字特征是通过随机过程的有限维分布函数的数字特征来刻画,由于随机过程$\{X(t),t\in T\}$在每一个 $t\in T$ 的状态是一个随机变量,有其对应的数字特征。随 t 的不同取值,随机变量的数字特征是可以不同的,它的数学期望和方差是依赖于参数 t 的函数,我们称这一函数为随机过程的数字特征。设随机过程 $X(t)t\in T$ 的数学期望用 $m_X(t)$表示,则有:

$$m_X(t)=\mathrm{E}X(t)=\int_{-\infty}^{\infty}x\mathrm{d}F(x,t)\qquad t\in T\tag{1.17}$$

式中:$F(x,t)$——随机过程的一维分布函数。

若 $F(x,t)$为连续函数,则有:

$$m_X(t)=\mathrm{E}X(t)=\int_{-\infty}^{\infty}x\cdot f(x,t)\mathrm{d}x\qquad t\in T\tag{1.18}$$

式中:$f(x,t)$——一维分布密度函数。

如图 1.15 所示,当样本曲线数增加到一定数量后,$m_X(t)$基本为一条固定曲线,而样本曲线围绕其上下波动。

设随机过程 $X(t)t\in T$ 的方差用$D_X(t)$表示,则有:

$$D_X(t)=\mathrm{E}[X(t)-m_X(t)]^2\qquad t\in T\tag{1.19}$$

而 $\sigma_X(t)=\sqrt{D_X(t)}$称为随机过程的标准差。

随机过程的方差也是过程 t 的函数,它反映了每一个样本曲线对均值曲线 $m_X(t)$的偏离程度。

在分析实际工程问题时,随机过程的均方值具有物理意义,随机过程的均方值用$\Psi_X(t)$