理论物理教程之七

计 算 物 理 学

马文淦 编著

中国科学技术大学出版社
2001·合肥

需要全本请访问 www.ertongbook.com

内 容 简 介

本书系统而详尽地讲述了计算物理领域涉及的基本概念、数学基础与方法；充分注意到了计算物理课程的特点，强调了计算方法与物理学科之间的关系。书中不仅以大篇幅讲述了在传统物理课题中常用的数值计算方法，如偏微分方程的数值求解方法、神经元网络方法以及计算机模拟方法中的随机模拟方法——蒙特卡洛方法和确定性模拟方法——分子动力学方法，而且细致地讲述了计算机符号处理系统及其在理论物理中的应用。这是综合讲授计算物理方法的教材中有其特色的地方。书中还提供了计算物理方法在理论和实验物理领域中应用的实例。

本书内容丰富，体系完整，适合于作为高等学校物理类高年级大学生和研究生的教学用书，也可以作为其他相关专业的师生及科研工作者的参考书。

图书在版编目（CIP）数据

计算物理学 / 马文淦编著. —合肥：中国科学技术大学出版社，2001.4
ISBN 7–312–01214–0

Ⅰ. **计…** Ⅱ. **马…** Ⅲ. **物理学–数值计算–计算方法** Ⅳ. O411

中国版本图书馆 CIP **数据核字（** 2000）**第** 36013 **号**

中国科学技术大学出版社出版发行

（ 安徽省合肥市金寨路 96 **号，**230026）

中国科学技术大学印刷厂印刷

全国新华书店经销

开本：787×1092/16 **印张：** 10.5 **字数：** 256 **千**

2001 **年** 4 **月第** 1 **版** 2001 **年** 4 **月第** 1 **次印刷**

印数：1—1500 **册** **定价：**13.00 **元**

（ 凡图书出现印装质量问题，请向承印厂要求调换）

前　言

计算物理学作为新兴的学科分支是物理学、数学在过去百余年来取得巨大成就的基础上，伴随着计算机科学近几十年来突飞猛进的发展而逐步发展起来的。计算物理早已与实验物理和理论物理形成三足鼎立之势，甚至可以说它将成为现代物理大厦的“栋梁”。今天的年轻一代物理工作者，无论是从事基础理论或实验研究，还是从事应用基础或工程研究的，都必须学习和掌握计算物理的概念和方法。

从 1980 年代中期开始，中国科学技术大学近代物理系就已经开设了《计算物理学》课程。本书大部分内容取自过去十多年来笔者在该系给高年级学生讲授的《计算物理学》必修课教材。该课程也作为近代物理系研究生的选修课。该课程虽然主要针对粒子物理和原子核物理、等离子体物理等专业的学生，但对诸如物理化学、材料科学、工程物理等专业的本科学生和研究生也有很大的吸引力。计算物理学是一门边缘学科，它的讲授涉及到物理学、数学和计算机科学的知识。因此，本教材是针对已经具备物理学基础知识、并具有一定数学水平和一般的计算机编程能力的读者而编写的。

计算物理学所包含的内容是相当广泛的。本书的内容力求较为全面，但由于讲授课程的时间和本人实践经验的限制，本书中仅特别选择了近代物理学中应用比较广泛，但又不是本科学生很容易掌握的一部分方法和技术作为教材内容。本书介绍了计算物理学中两大类计算：一类可以称做是计算机数值计算方法（第二、三、四、五、六、九章），另一类则可以称为计算机符号计算（第七、八章）。在计算机数值计算方法中介绍了偏微分方程的数值求解方法（第四章的有限差分法和第五章的有限元素法）和计算机模拟方法。在计算机模拟的内容中又包含了蒙特卡洛模拟方法及其应用（第二、三章）和确定性模拟方法（第六章的分子动力学方法）。在第九章中介绍的神经元网络方法实际上还是一种数值计算方法，它是近年来在粒子物理研究中用得较为成功的方法。

由于作者本人水平所限，本书在选择内容的合理性，叙述的科学性方面可能还有值得斟酌的地方，错漏之处也在所难免。希望读者们批评指正。

作者在编写本书的过程中得到过舒伯尔（F.Schoeberl）教授、韩良博士和张杰博士的帮助，在此表示深切的谢意。

马 文 淦

2000 年春于中国科大

目　　次

前言 I

第一章　引言 1

1.1 **计算物理的起源和发展** 1
1.2 **计算物理学在物理学研究中的应用** 2

第二章　蒙特卡洛方法 5

2.1 **蒙特卡洛方法的基础知识** 5
2.2 **随机数与伪随机数** 9
2.3 **任意分布的伪随机变量的抽样** 14
2.4 **蒙特卡洛计算中减少方差的技巧** 15

第三章　蒙特卡洛方法的若干应用 40

3. 1 **蒙特卡洛方法在定积分计算中的应用** 40
3.2 **事例产生器** 44
3.3 **高能物理实验中蒙特卡洛方法的应用** 46
3.4 **随机游动及应用** 50
3.5 **在量子力学中的蒙特卡洛方法** 54
3.6 **在统计力学中的蒙特卡洛方法** 60

第四章 有限差分方法 64

4.1 **引言** 64
4.2 **有限差分法和偏微分方程** 66
4.3 **有限差分方程组的迭代解法** 70
4.4 **求解泊松方程的直接法** 75

第五章 有限元素方法 79

5.1 **有限元素方法的基本思想** 79
5.2 **二维场的有限元素法** 82
5.3 **有限元素法与有限差分法的比较** 90

第六章　分子动力学方法 91

6.1 **引言** 91
6.2 **分子运动方程的数值求解** 92
6.3 **分子动力学模拟的基本步骤** 94

6.4 平衡态分子动力学模拟 98
第七章 计算机符号处理 103
7.1 引言 103
7.2 通用符号处理系统的特点和功能 105
7.3 Mathematica 语言编程 108
第八章 Mathematica 在理论物理中的应用举例 112
8.1 粒子在中心力场中的运动问题 112
8.2 求非相对论性薛定格方程本征能量限 119
第九章 神经元网络方法及其应用举例 143
9.1 神经元网络法 143
9.2 高能物理中的神经元网络应用举例 147
附录 149
附录 A 贝斯理论 149
附录 B 一些常用分布密度函数的抽样 149
附录 C 求解常微分方程的近似方法 151
附录 D 三角形型函数积分式的证明 153
附录 E Mathematica 函数和指令 154

需要全本请访问 www.ertongbook.com

第一章　引　　言

计算物理学是英文“ Computational Physics ”的中译文。通常人们也把它等同于计算机物理学(Computer Physics)。它是一门新兴的边缘科学，是物理学、数学、计算机科学三者相结合的产物。计算物理学也是物理学的一个分支，它与理论物理、实验物理有密切联系，但又保持着自己相对的独立性。如果要给计算物理学做一个定义的话，我们可以采用下面这个有代表性的概括：计算物理学是以计算机及计算机技术为工具和手段，运用计算数学的方法，解决复杂物理问题的一门应用科学。计算物理学已经给复杂体系的物理规律、物理性质的研究提供了重要手段，对物理学的发展起着极大的推动作用。

计算物理学作为一门新兴的学科，它是怎样发展起来的？它与理论物理、实验物理有什么区别和联系？计算物理学在物理学中的应用情况如何？这就是本章要介绍的内容。

1.1　计算物理的起源和发展

19 世纪中叶以前，可以说物理学还基本上是一门基于实验的科学。1862 年麦克斯韦(Maxwell)将电磁规律总结为麦克斯韦方程，进而在理论上预言了电磁波的存在。这使人们看到了物理理论思维的巨大威力。从此理论物理开始成为一门独立的物理学分支。到了 20 世纪初，物理学理论经历了两次重大的突破，相继诞生了量子力学和相对论。理论物理开始成为一门成熟的学科。传统意义上的物理学便具有了理论物理和实验物理（ 应用物理包括在内）两大支柱。物理学便成为实验物理和理论物理密切结合的学科。正是物理学这样的“ 理论与实践相结合 ”的研究方式，才大大促进了该学科的发展，并引发了 20 世纪科学技术的重大革命。这个革命对人类的社会生活产生了重大影响。其中一个重要的方面就是电子计算机的发明和应用。

物理学研究与计算机和计算机技术紧密结合起始于 20 世纪 40 年代。当时正值第二次世界大战时期，美国在研制核武器的工作中，要求准确地计算出与热核爆炸有关的一切，迫切需要解决在瞬时发生的最复杂的物理过程的数值计算问题。然而这是采用传统的解析方法求解或手工数值计算根本办不到的。这样，计算机在物理学研究中的应用成为不可避免的了，计算物理学因此得以产生。第二次世界大战之后，计算机技术的迅速发展又为计算物理学的发展打下了坚实的基础，大大增强了人们从事科学研究的能力，促进了各个学科之间的交叉渗透，使计算物理学得以蓬勃发展。

理论物理是从一系列的基本物理原理出发（ 例如：质量守恒、动量守恒、角动量守恒、电荷守恒、万有引力规律、静电作用规律以及电磁感应规律等），列出数学方程，再用传统的数学分析方法求出显示的解析解。通过这些解析解所得到的结论与实验观测结果进行对比分析，从而解释已知的实验现象并预测未来的发展。实验物理是以实验和观测为基本手段来揭示新的物理现象，奠定理论物理对物理现象作进一步研究的基础，从而为发现新的理论提供依据，或者检验理论物理推论的正确性及应用范围。计算物理则是计算机科学、数学和物理

学三者间新兴的交叉学科或边缘学科。计算物理学研究的主要内容是如何应用高速计算机作为工具，去解决物理学研究中极其复杂的问题。例如：在高能物理实验中，由于实验技术的发展和测量精度的提高，实验规模越来越大，实验数据量惊人地增加，被测实验数据在单位时间内的产额非常高，因而单靠人力和通常的电子仪器已无法完成实验设备的管理和实验数据的处理工作。又如电子反常磁矩修正的计算。对四阶修正的手工解析计算已经相当繁杂，而对六阶修正的计算已经包含了 72 个费曼图，手工解析运算已不可能完成。类似这样的复杂系统的控制和大量繁杂的计算工作，计算机的应用就成为不可避免的了。计算物理学对解决复杂物理问题的巨大能力，使它成为物理学的第三支柱，并在物理学研究中占有重要的位置。

计算物理学与理论物理和实验物理有着密切的联系。计算物理学的研究内容涉及到物理学的各个领域。一方面，计算物理学所依据的理论原理和数学方程是由理论物理提供的，其结论还需要理论物理来分析检验；另一方面，计算物理学所依赖的数据是由实验物理提供的，其结果还要由实验来检验。对实验物理而言，计算物理学可以帮助解决实验数据的分析、控制实验设备、自动化数据获取以及模拟实验过程等问题。对理论物理而言，计算物理学可以为理论物理研究提供计算数据，为理论计算提供进行复杂的数值和解析运算的方法和手段。总之，计算物理学是与理论物理、实验物理互相联系、互相依赖、相辅相成的，它为理论物理研究开辟了一个新的途径，也对实验物理研究的发展起了巨大的推动作用。

1.2 计算物理学在物理学研究中的应用

自 20 世纪 40 年代以来，由于人们受到在原子弹设计中使用计算机而取得了巨大成功的启示，计算物理的方法和技巧也迅速地从核物理向其他学科领域渗透，从军事研究转向基础科学研究，从而大大丰富了计算物理学的内容。在 60 年代以前，计算机还主要用在物理问题的数值计算和模拟。而到 60 年代以后计算机又进一步深入到实验室控制和数据获取自动化和理论解析运算自动化方面。1962 年，在低能物理实验中就开始了计算机与实验的联机工作。1964 年，在高能物理实验中开始采用计算机高速可靠地采集和处理数据信息，以满足粒子物理实验对高事例率、大数据量处理和大型仪器设备控制的要求。当今在我们物理学研究中计算机的应用已经是无处不在了。计算机在物理学中的应用可以大致分为四类：计算机数值分析、计算机符号处理、计算机模拟和计算机实时控制。

通常在物理研究中，我们从已知的物理规律出发得到描写物理过程的抽象数学公式后，最后或许要做数值分析以便与实验结果对照或作为实验的参考数据。如果对一个简单的数学公式进行数值求解，也许我们还可以用纸和笔，手工就计算出数值结果来。但是对更复杂系统的数学处理，我们就不得不在计算机上用计算机特有的数值计算方法来计算了。在这种工作方式下，计算机成了物理学研究的数值分析的工具。

计算机在物理学中应用的另一个重要方面是利用计算机的符号处理系统进行解析计算、公式的推导和高精度的数值计算。这在理论物理研究领域的意义就特别重大。当前在天体物理、核和粒子物理研究中已经广泛采用计算机符号处理系统来做复杂的公式推导和高精度计算，还发展出许多用于各个领域研究的计算机符号计算程序包。

借助于计算机的符号和数值计算程序，我们可以方便地解析和数值地计算各种复杂的数学物理问题,诸如多重不定积分和定积分、大型数字或者符号矩阵的计算、求解复杂的微分方程等等。随着计算机技术的高速发展，今天我们已经能够在个人微机上做复杂的符号和数值

计算了。计算机的确在物理学的计算中起到了十分重要的作用。

我们还要指出：计算机的数值计算功能对物理学研究的用途决不仅仅是可以得到数值结果，更为重要的是，它为物理学家提供了“ 计算机模拟实验 ”这个新的研究手段。例如统计物理中有个自回避随机迁移问题，它是在随机游动中加上了一个限制，即以后的游走步子不能穿过以前各步所走过的路径。这样的问题就不再像一般的随机迁移问题那样可以用通常的微分方程来描写系统的统计行为。对这类物理问题的研究，计算机模拟实验就几乎是唯一的研究方法。即使对于一些有解析方程描述的问题，由于系统的复杂性，往往用计算机模拟比做数值计算更为方便。计算机模拟实验基本上不受实验条件、时间和空间的限制，这就使它具有极大的灵活性和随意性。也就是说，只要建立起理论模型，我们就能进行计算机模拟实验，即使这样的实验现象在自然界可能是不存在的，或者该实验在时间和空间上都是在实验室无法进行的。因此，通过计算机模拟实验会给物理学家带来新的物理概念，发现新的物理现象。当前计算机模拟已经成为继理论和实验研究方法外，物理学研究的第三种手段。

物理实验中的计算机控制也是十分重要的。现在几乎所有的大型实验中，它的大多数实验设备都通过接口与控制计算机相连接，并结合在线数据获取和分析程序对实验装置的整个实验进程做实时控制，使物理实验可以在没有人在场的情况下自己监测设备的正常运行，自动采集和分析实验数据。

一般来说，计算机在物理实验中的应用大致可以分为两个部分，即计算机的在线分析和离线分析。在实验装置运行过程中由计算机实现数据获取和数据分析就称为实验的在线分析。以粒子物理实验为例，在线分析的任务包括四个方面：

（ 1）控制系统运行。根据物理实验对物理事例的选择要求和对在线系统构成部分的管理需要，设计一定的程序逻辑，采用计算机实现对整个在线系统运行的控制。

（ 2）采集实验数据。将探测器记录到的事例信息，加速器运行中的束流状态及某些仪器设备的工作状态信息，采集送入计算机；计算机又以规定的格式将这些实验数据记入到计算机的外部存贮设备（ 磁带、磁盘或光盘）中。

（ 3）监视仪器状态。计算机定时或不定时地监视探测器工作状态的情况，加速器束流的流强变化等。一旦出现不正常情况，计算机将送出状态信息，通知值班人员，或自动作出预先规定好的处理。

（ 4）数据在线分析。在实验进行期间，对在线系统获取的数据信息，由计算机按各种方式进行取样分析。数据分析的范围是由在线系统的分析能力决定的。在一个能力较强的系统中，数据分析还包括按一定的物理要求对事例进行判别与选择，实现粒子作用事例的图像重建。这些分析的目的是为了观察仪器设备安排和事例选择方案的实施情况，以便在实验运行期间研究和发现问题，改善实验设计。

粒子物理实验的离线分析是将实验数据送到计算中心做进一步的浓缩、过滤和理论分析工作。粒子物理的离线分析还包括对物理过程的理论模拟、探测器模拟、本底分析、理论和实验事例的分析对照等。粒子物理的离线分析又可以划分为两部分工作。一个是事例模拟；另一个是物理分析。事例模拟也就是 计算机实验 ”，它包括对所研究过程及可能形成该过程本底的背景过程的模拟。这个模拟过程是从理论模拟产生出终态产物的各物理参数（ 包括能量、动量、方向、粒子种类等）开始，再通过探测器模拟，得到格式与实验数据记录相同的模拟数据。探测器模拟包含了终态粒子通过实验装置时，在各个探测器上留下的能量和时间的数字化信息。物理分析工作主要包括事例的径迹重建、各类事例的筛选和物理参数的计算分析。分析的数据对象既包括实验数据，也包括模拟数据。上面介绍的计算机在粒子物理研

究中的应用，就是属于通常称做“ 计算高能物理学”(Computational High Energy Physics)**的学科领域。**

计算机在物理学研究中还有其他许多用途，比如，用于语言文字处理、通过计算机网络进行信息或科学数据的交流传递、计算机辅助教学等等。这里我们不再赘述。

计算物理学是计算机在自然科学的应用中发展较早的学科之一。虽然它的研究对象是物理学，但是它的研究方法可以推广到其他的自然科学领域，甚至包括社会科学、思维科学、决策和管理科学等社会科学领域。计算物理学研究的一些特点和优点，甚至它的一些研究成果都可以去支持这些领域的研究工作。毫无疑问，计算物理学的发展将对自然科学和社会科学领域的计算机应用研究起着极大的推动作用。

第二章　蒙特卡洛方法

计算机模拟实验在物理学研究中占着越来越重要的地位。从计算机模拟采用的方法来看，它大致可以分为两种类型。一种类型为随机模拟方法或统计试验方法，又称蒙特卡洛(Monte Carlo)方法。它是通过不断产生随机数序列来模拟过程。自然界中有的过程本身就是随机的过程，物理现象中如粒子的衰变过程、粒子在介质中的输运过程……等等。当然蒙特卡洛方法也可以借助概率模型来解决不直接具有随机性的确定性问题。另一类为确定性模拟方法。它是通过数值求解一个个粒子的运动方程来模拟整个系统的行为。在统计物理中称为分子动力学（ Molecular Dynamics）方法。关于分子动力学方法我们将在第六章中介绍。此外，近年来还发展了神经元网络方法和原胞自动机方法。我们将在第九章介绍神经元网络方法及其应用举例。

蒙特卡洛方法的提出可以追溯到 19 世纪末期，但是实际上直到 20 世纪 40 年代以后，随着电子计算机的发展，该方法才得到迅速的发展和应用。在第二次世界大战中，蒙特卡洛方法首先被美国的科学家应用于原子弹的研制中。目前这一方法已经广泛运用到物理学的许多领域。甚至像系统工程、科学管理、生物遗传、社会科学等学科领域也采用了这种研究方法。这些都充分表现出这种方法完全区别于其他的方法，具有独特功能和优越性。

2.1　蒙特卡洛方法的基础知识

一、基本思想

所谓蒙特卡洛方法，就是根据待求随机问题的变化规律，根据物理现象本身的统计规律，或者人为地构造出一个合适的概率模型，依照该模型进行大量的统计实验，使它的某些统计参量正好是待求问题的解。下面我们举两个最简单的例子来说明上面解释的内涵。

尽管现在人们都认为：在当今的研究工作中离开了电子计算机，很难想像蒙特卡洛方法计算还能够进行。但是实际上远在计算机出现以前，蒙特卡洛方法就已经被仔细研究过了。著名的巴夫昂(Buffon)投针实验就是巴夫昂在 1777 年提出的求π的近似值的方法。该试验方案是：在平坦桌面上划一组相距为 s 的平行线，向此桌面随意地投掷长度$l=s$的细针，那么从针与平行线相交的概率就可以得到π的数值。

该试验方案的原理是基于数学统计理论所得到的结论，即此试验中细针与平行线相交的概率为$2/\pi$。这个数学统计理论的结果可以简单地计算如下：设针与平行线的垂直方向的夹角为α，那么针在平行线垂直方向的投影长度为$l\cdot|\cos\alpha|$。对于确定的α夹角，细针与平行线相交的概率为投影长度与平行线间距之比，即$\frac{l\cdot|\cos\alpha|}{s}=|\cos\alpha|$。由于$\alpha$是在$[0,\pi]$间均匀分布的，所以$|\cos\alpha|$的平均值为

$$\frac{1}{\pi}\int_0^{\pi}|\cos\alpha|\mathrm{d}\alpha=\frac{2}{\pi} \tag{2.1.1}$$

假如在 N 次投针中，有 M 次和平行线相交。当 N 充分大时，相交的频率M/N就近似为细

针与平行线的相交的概率。因此结合公式(2.1.1)，我们得到

$$\pi \approx \frac{2N}{M} \tag{2.1.2}$$

然而，这种投针法的试验结果，效率和精度都很差。我们现在来计算一下经过 n 次投针后得到的π值精度。设 p 为细针与平行线相交的概率($p = 2/\pi$)，则针与平行线相交的次数应满足二项式分布，其期望值为np，方差应为$np(1-p)$。因而$2/\pi$值的方差为 $p(1-p)/n$，标准误差为$\sqrt{\frac{p(1-p)}{n}}$。将 $p = 2/\pi$ 的标准误差改写为π 的标准误差$2.37/\sqrt{n}$ (这里我们必须先知道π值来计算，但也可以通过试验数据来估计π值)。这意味着试验所得的π值的不确定性的范围如下：

对 100 次投针为，0.2374

对 10,000 次投针为，0.0237

对 1,000,000 次投针为，0.0024

显然这种方法比用其他方法计算π值所引起的不确定范围要大得多。实际上不应当采用这种方法来计算π值。这里我们只是将它作为蒙特卡洛方法在表面上与随机过程无关的领域中应用的一个典型实例。

作为第二个例子，我们考虑一个简单的定积分计算

$$I = \int_0^1 f(x)\mathrm{d}x \tag{2.1.3}$$

假定被积函数 $y = f(x)$ 在积分范围的值是在区间 $0 \le y \le 1$，如图(2.1.1)所示。这时我们可以随机地向正方形内投点，最后统计落在曲线下的点数 M，当总的掷点数 N 充分大时，M/N就近似等于积分值 I。

根据这两个例子，我们可以将蒙特卡洛方法的基本思想总结如下：当问题可以抽象为某个确定的数学问题时，应当首先建立一个恰当的概率模型，即确定某个随机事件 A 或随机变量 X(如上面例子中的投针实验，细针与平行线相交的事件；求定积分中的随机变量 f)，使得待求的解等于随机事件出现的概率或随机变量的数学期望值。然后进行模拟实验，即重复多次地模拟随机事件 A 或随机变量 X。最后对随机实验结果进行统计平均，求出 A 出现的频数或 X 的平均值作为问题的近似解。这种方法也叫做间接模拟。

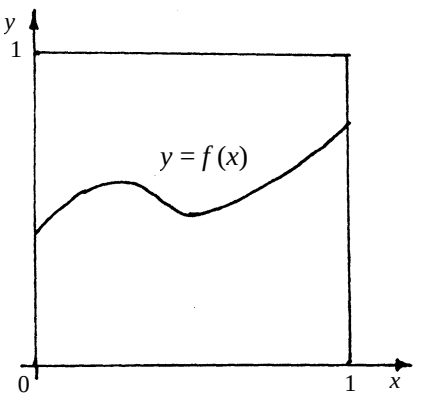


图 2.1.1 定积分计算示意图

对求解问题本身就具有概率和统计性的情况，例如中子在介质中的传播，核衰变过程等，我们可以使用直接模拟法。该方法是按照实际问题所遵循的概率统计规律，用电子计算机进行直接的抽样试验，对结果的处理和间接法相同。直接模拟法最充分体现出蒙特卡洛方法无可比拟的特殊性和优越性，因而得到广泛的应用。该方法也就是通常所说的“ 计算机实验 ”。

二、随机变量和随机变量的分布

随机变量是一个可以取不止一个值的变量(通常在连续区间取值)，并且人们可能无法事先预言它取的某一特定值。虽然这种变量的值无法预言，但其分布是可能了解的。假定我们研究连续的随机变量，由随机变量的分布可以得到它取某给定值的概率，即

$$g(u)\mathrm{d}u = P[u < u' < u + \mathrm{d}u]. \tag{2.1.4}$$

需要全本请访问 www.ertongbook.com

$g(u)$称为 u 的概率分布密度函数，它表示随机变量u'取u到$u+\mathrm{d}u$之间值的概率。物理学家们常常用概率密度函数来表达u'的分布。但是，数学上有时采用分布函数更为方便。分布函数定义为：

$$G(u)=\int_{-\infty}^{u} g(x)\,\mathrm{d}x \tag{2.1.5}$$

则

$$g(u)=\mathrm{d}G(u)/\mathrm{d}u \tag{2.1.6}$$

注意：$G(u)$是一个在$[0,1]$区间取值的单调递增函数。通常$g(u)$是归一化的分布密度函数，因而该函数对所有的u值范围的积分值应当为1。

三、随机变量的独立性

假如我们考虑两个随机变量u'和v'的分布，则必须引进这两个变量的联合分布密度函数$h(u,v)$，此时带来的数学问题就更为复杂。但是在$h(u,v)=p(u)\cdot q(v)$这种特殊情况下，u'和v'是彼此独立的随机变量。对于两个以上的变量来说，随机变量独立性的概念就更复杂了。此时仅考虑两个变量之间的独立性是不够的。事实上，所有变量有可能两两间是相互独立的，而在三个变量，甚至更多变量的组合之间却是相关的。我们举如下例子来说明：如果r和s是两个均匀分布在$[0,1]$区间的相互独立的随机变量，由此我们可以构造三个新的变量

$$\begin{aligned} x &= r \\ y &= s \\ z &= (r+s)\bmod 1 \end{aligned} \tag{2.1.7}$$

此时x,y,z也都是均匀分布在区间$[0,1]$的随机变量，并且所有的$(x,y),(y,z)$和(x,z)组合都是独立的(括号内任一个变量值的选取并不对括号中另一个变量的取值有影响)。但是(x,y,z)中，任意两个变量的值可以确定出第三个变量的值。此时它们之间存在明显的相关性。

四、期望值、方差和协方差

一个函数$f(u')$的数学期望值定义为该函数的平均值

$$E\{f\}=\int f(u)\mathrm{d}G(u)=\int f(u)g(u)\mathrm{d}u \tag{2.1.8}$$

上式中，$G(u)$ $G(u)$是独立变量u'的分布函数。通常u'是在$[a,b]$区间均匀分布的随机变量，即$\mathrm{d}G=\mathrm{d}u/(b-a)$。这时的期望值可以写为

$$E\{f\}=\frac{1}{b-a}\int_a^b f(u)\,\mathrm{d}u \tag{2.1.9}$$

类似地，可以定义变量u'的期望值为u的平均值

$$E\{u'\}=\int u\mathrm{d}G(u)=\int ug(u)\mathrm{d}u \tag{2.1.10}$$

一个函数或变量的方差是可以用下式来定义的

$$V\{f\}=E\{(f-E\{f\})^2\}=\int[f-E\{f\}]^2\mathrm{d}G \tag{2.1.11}$$

注意：在上式中计算 f 的期望值时，需要做一次积分，而求方差时还需做一次积分。

方差的平方根叫做标准误差。由于标准误差与其真值有相同的量纲，因而它比方差更具有物理意义。但是求标准误差时的平方根运算在数学处理时很不方便。标准误差也很容易解释为平方值的均方根误差。如果将求期望值和求方差的运算作为算符，我们可以证明出这些算符作用在随机变量的线形组合式上的一些简单规则。假如 x 和 y 是随机变量，c 是一个常数，则

$$E\{cx+y\}=cE\{x\}+E\{y\} \tag{2.1.12}$$

$$V\{cx+y\}=c^2V\{x\}+V\{y\}+2cE\{(y-E\{y\})(x-E\{x\})\} \tag{2.1.13}$$

因而期望值算符是一个线性算符，而方差算符是非线性算符。公式(2.1.13)右边最后一项称为 x 和 y 间的协方差。如果 x 和 y 是独立随机变量，则 x 和 y 间的协方差为零。通常协方差为正值时，我们称 x 和 y 是正关联；反之，我们称 x 和 y 是负关联。不过我们也要注意：(1)即使 x 与 y 的协方差为零，我们也不能肯定 x 和 y 是否是独立变量。(2)尽管方差算符是非线性的，但如果 x 和 y 是独立变量，则

$$V\{x+y\}=V\{x\}+V\{y\} \tag{2.1.14}$$

五、大数法则和中心极限定理

概率论中的大数法则和中心极限定理是蒙特卡洛方法的基础。大数定理反映了大量随机数之和的性质。如果函数 h 在 $[a,b]$ 区间，以均匀的概率分布密度随机地取 n 个数 u_i，对每个 u_i 计算出函数值 $h(u_i)$。大数法则告诉我们这些函数值之和除以 n 所得的值将收敛于函数 h 的期望值，即

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}h(u_i)=\frac{1}{b-a}\int_a^b h(u)\,\mathrm{d}u \tag{2.1.15}$$

公式(2.1.15)的左边正是公式右边积分的蒙特卡洛估计值。根据这个法则，我们在抽取足够多的随机样本后，计算得到的积分的蒙特卡洛估计值将收敛于该积分的正确结果。若要对收敛的程度进行研究，并作出各种误差估计，则要用到中心极限定理。中心极限定理可以近似地告诉我们：在有足够大，但又有限的 n 值的情况下，蒙特卡洛估计值是如何分布的。该定理指出：无论单个随机变量的分布如何，许多独立随机变量之和总是满足正则分布(即高斯分布)。高斯分布可以由给定的期望值 μ 和方差 σ^2 完全确定下来。通常用 $N(\mu,\sigma^2)$ 来表示高斯分布密度函数

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left[-(x-\mu)^2/2\sigma^2\right] \tag{2.1.16}$$

如果公式(2.1.15)右边积分的期望值为 $\bar{I}$，公式左边用蒙特卡洛估计的值为 I_n，标准误差为 σ，则当 n 充分大时，对任意的 $\lambda(\lambda>0)$，有

$$P\left\{\left|\frac{I_n-\bar{I}}{\sigma/\sqrt{n}}\right|<\lambda\right\}=\frac{1}{\sqrt{2\pi}}\int_{-\lambda}^{\lambda}\mathrm{e}^{-t^2/2}\,\mathrm{d}t=1-\alpha \tag{2.1.17}$$

这说明：该积分的期望值与蒙特卡洛估计值之差在范围

$$\left|I_n-\bar{I}\right|<\lambda\frac{\sigma}{\sqrt{n}} \tag{2.1.18}$$

内的概率为 $1-\alpha$，α 称为显著水平，$1-\alpha$ 称为置信水平。σ 为蒙特卡洛估计值的标准误差，$\sigma^2=V\{f\}/n$。α 与 λ 的关系可以由公式(2.1.17)求得。也有专门的数学用表可查。例如取置信水平 $1-\alpha=99\%$，可以查得 $\lambda=3$。这可以解释为：不等式 $\left|I_n-\bar{I}\right|<3\frac{\sigma}{\sqrt{n}}$ 成立的概率为99%。

同样$1-\alpha=95\%$时，$\lambda=2$，其解释与上面一例相似。

从上面的分析看到，蒙特卡洛方法的误差与σ^2和n有关(见公式(2.1.18))。为了减小误差，就应当选取最优的随机变量，使其方差最小。对同一个问题，往往会有多个可供选择的随机变量，这时就应当择优而用之。在方差固定时，增加模拟次数可以有效地减小误差。如试验次数增加 100 倍，精度提高 10 倍。当然这样做就增加了计算机计算机时，提高了费用。所以在考虑蒙特卡洛方法的精确度时，不能只是简单地减少方差和增加模拟次数，还要同时兼顾计算费用，即机时耗费。通常以方差和费用的乘积作为衡量方法优劣的标准。

蒙特卡洛方法精度的概念也不是通常意义下收敛于真值，而是在某一置信度，或者说某一概率意义下收敛于真值。也就是说，精度是带有随机性的，我们只能知道有多大的可能性具有某一精度，而不能认准一定具有某一精度。

2.2 随机数与伪随机数

原则上，一个随机数仅仅是指随机变量所取的某一个特定的值。但是在蒙特卡洛方法的研究中“随机”一词包含了各种其他不同的含义。人们常常采用已经确定好的数列来作蒙特卡洛研究。这些数列从统计意义上讲并不是随机的，但却具有与真正的随机数序列相似的某些特性。这些数列可以分为三种不同的类型：真随机数列，准随机数列和伪随机数列。需要指出的是：在实际应用中，有两个完全独立的术语概念很容易引起混淆。它们是数列的随机特性和它的分布。实际上，一个完美的随机数序列可能具有某种分布(例如：均匀分布、高斯分布等)，但是具有某种分布的数列却可能完全不是随机的。

一、真随机数

真随机数数列是不可预计的，因而也不可能重复产生两个相同的真随机数数列。真随机数只能用某些随机物理过程来产生。例如：放射性衰变、电子设备的热噪音、宇宙射线的触发时间等等。如果采用随机物理过程来产生蒙特卡洛计算用的随机数，理论上不存在什么问题。但在实际应用时，要作出速度很快(例如每秒产生上百个浮点数)，而又准确的随机数物理过程产生器是非常困难的。有时甚至还要做较多的计算工作。

弗里吉雷欧(Frigerio)等人在 1975 年至 1978 年做过下面所述的工作。他们用一个α粒子放射源和一个高分辨率的计数器做成的装置，在 20 ms 时间内平均记录了 24.315 个α粒子。当计数为偶数时，便在磁带上记录二进制的“1”。他们还仔细地对奇数计数的几率并不精确等于 1/2 所引起的偏差进行了修正。这个装置每小时可以产生大约 6000 个 31 比特(bits)的真随机数。这些数被存储在磁带上，并通过了一系列的“随机数”检验用于蒙特卡洛计算当中。

这里我们对消除偏差的技巧做些介绍。利用上面介绍的装置得到的“0”或者“1”的真随机数序列中，0 和 1 出现的几率 $P(0)$和 $P(1)$可能并不精确等于 1/2。我们从原始的真随机数序列出发，将序列中的二进制数依次成对组合；如果这组中的两个数相同，则舍去这两个数；如果这组中的两个数不相同，则保留第二个二进制数而丢弃第一个数。这样构成的一个新序列可以保证：在原始序列中的数是相互独立的情况下，“0”和“1”出现的概率相等。这一点可以从如下的计算中看出：“0”出现在新序列中的概率为$P'(0)=P(1)P(0)$。这是因为新序列中的“0”只能在原始序列中“1”后面跟着“0”时才出现。同样“1”在新序列中出现的概率$P'(1)=P(0)P(1)$。因而无论 $P(0)$和 $P(1)$等于什么值，$P'(0)$ 和 $P'(1)$ 都相等。由于在构成新

序列时，舍去了一组数的几率为 $P^2(0)+P^2(1)$ ，因而 $P'(0)+P'(1)$ 不等于 1，而小于或等于 1/2。在这种方法中，对两个数不相同的一组数至少要丢掉一个二进制数。很明显，它的产生效率为 $P(0)P(1)=P(1-P)$，其中 P 为 $P(0)$或 $P(1)$。其产生效率的最大值为 25 %。

我们再回顾一下前面曾叙述过的巴夫昂投针实验来说明在真随机数产生器中由于物理偏差所引起的问题。首先，在投针实验中平行线间间距必须保证为一个常数值，并在所要求的误差范围内与针长相等。如果我们仅要求 π 值的一至二位有效数字，这个要求是不难满足的，但是如果要求更多位的有效数字，这就比较困难了。第二，正确地判断临界状态下的针与平行线的相交也非易事。第三，我们还必须保证针的投掷位置和角度的分布是均匀分布的。为保证角度分布的均匀性，可以在投针的时候，让针迅速旋转，并采用非常平的、摩擦系数是各向同性的桌面。此外，针位置的分布决不是均匀分布的，而是在投掷目标点周围服从高斯分布。在实际应用中，我们必须由实验来决定这一分布宽度，并且要对它引起的偏差做类似于前面所述的由弗里吉雷欧等人所做的复杂修正。

二、准随机数

准随机数序列并不具有随机性质，仅仅是它用来处理问题时能够得到正确结果。准随机数概念是来自如下的事实：对伪随机数来说，要实现其严格数学意义上的随机性，在理论上是不可能的，在实际应用中也没有这个必要。关键是要保证“随机”数数列具有能产生出所需要的结果的必要特性。例如，在多重积分和大多数模拟研究中，多维空间的每个点或模拟事例被认为是相互独立的，而这些点或事例的顺序则似乎并不重要。因而我们可以在大多数运算中，放心地置随机性的概念于不顾。同样，我们也可以不考虑对某些分布均匀性的涨落程度。事实上在许多情况下，超均匀的分布比真随机数的均匀分布更合乎实际需要。

从严格的意义上来讲，若放弃了所有随机性的要求，采用不具有“随机”性特性的数列的方法，我们已经不能再将它纳入蒙特卡洛计算的范畴了。但是如果将蒙特卡洛方法的概念扩大到包括准随机数序列，这样可能更恰当一些。因为准蒙特卡洛方法仍然保留了蒙特卡洛方法的一些基本的特性。例如，它可以用于非常高维空间中的计算；利用它计算多重积分时与积分重数无关，甚至对非常高维数的积分计算，其计算量增加也很少；它对函数连续性要求很强等等。事实上，准蒙特卡洛方法是将蒙特卡洛方法处理问题的维数，向高维扩展的方法。由此可见准蒙特卡洛方法的理论与真蒙特卡洛的理论很接近，而与求积分的理论差别很大。

三、伪随机数

在实际应用中的随机数通常都是通过某些数学公式计算而产生的伪随机数。这样的伪随机数从数学意义上讲已经一点不是随机的了。但是，只要伪随机数能够通过随机数的一系列的统计检验，我们就可以把它当作真随机数而放心地使用。这样我们就可以很经济地、重复地产生出随机数。这里我们需要了解满足物理问题的计算机模拟需要的伪随机数的标准是什么。理论上，我们要求伪随机数产生器具备以下特征：良好的统计分布特性、高效率的伪随机数产生、伪随机数产生的循环周期长和伪随机数可以重复产生。其中满足良好的统计性质是最重要的。然而实际使用的伪随机数产生程序还没有一个是十全十美的，因此我们要求产生出的伪随机数应当能通过尽可能多的统计检验，以便人们放心地使用。我们在本章以下内容中将主要介绍伪随机数的产生和应用。这里我们首先讨论一下在实际应用中，如何产生和检验伪随机数。

1. 伪随机数的产生方法

伪随机数产生器产生的实际上是伪随机数序列。最基本的产生器是均匀分布的伪随机数产生器。最早的伪随机数产生器可能是冯·诺曼的平方取中法。该方法首先给出一个 $2r$ 位的

数，取它的中间的 r 位数码作为第一个伪随机数；然后将第一个伪随机数平方构成一个新的 $2r$ 位数，再取中间的 r 位数作为第二个伪随机数……。如此循环便得到一个伪随机数序列。类似上述方法，利用十进制公式表示 $2r$ 位数 x_n 的递推公式。

$$\begin{aligned} x_{n+1} &= \left[10^{-r} x_n^2\right] \left(\mathrm{mod} 10^{2r}\right) \\ \xi_n &= x_n / 10^{2r} \end{aligned} \tag{2.2.1}$$

这样得到的 $\{\xi_i\}$ 伪随机数序列是分布在[0，1]上的。相应的二进制递推公式为(x_n 为 $2r$ 位二进制数)：

$$\begin{aligned} x_{n+1} &= \left[2^{-r} x_n^2\right] \left(\mathrm{mod} 2^{2r}\right) \\ \xi_n &= x_n / 2^{2r} \end{aligned} \tag{2.2.2}$$

上面公式中 $[x]$ 表示对 x 取整。运算 $A = B(\mathrm{mod}\ M)$ 表示 A 等于 B 被 M 整除后的余数。如果选择初始数 x_0 适当，这种方法可以得到似乎是随机的一长串数。但是这种方法不是很好，现在已很少使用。这主要是因为该方法产生的数列具有周期性，有些数(如零)甚至会紧接着重复出现。

实际使用的伪随机数产生器常常比平方取中法简单。如今比较流行，并用得最多的是同余产生器。我们通过如下的线形同余关系式来产生数列。

$$\begin{aligned} x_{n+1} &= (a x_n + c)(\mathrm{mod}\, m) \\ \xi_n &= x_n / m \end{aligned} \tag{2.2.3}$$

其中 x_0 称为种子，改变它的值就得到基本序列的不同区段。a, c, x_0, m 为大于零的整数，分别叫做乘子、增量、初值和模。使用时需要仔细地挑选模数 m 和乘子 a，使得产生出的伪随机数的循环周期要尽可能长。$c \neq 0$ 时能实现最大的周期，但是得到的伪随机数的特性不好。$c \neq 0$ 的这类情况称为混合同余发生器。通常选取 x_0 为任意非负整数，乘子 a 和增量 c 取如下形式

$$a = 4q + 1 \quad , \quad c = 2p + 1 \tag{2.2.4}$$

p 和 q 为正整数。这两种方法中的 p, q, x_0, m 值的选择一般是通过定性分析和计算机试验来选择，以使得到的伪随机数列具有足够长的周期，而且独立性和均匀性都能通过一系列的检验。

$c = 0$ 的情况叫做乘同余法，由于减少了一个加法，因而可以使产生伪随机数的速度快些。这种方法产生的伪随机数递推公式为

$$\begin{aligned} x_{n+1} &= a x_n (\mathrm{mod}\, m) \\ \xi_n &= x_n / m \end{aligned} \tag{2.2.5}$$

x_0, a, m 也为正整数，并分别叫做初值、乘子和模。

还有许多其他的产生伪随机数的方法，例如混沌法伪随机数产生、反馈移位寄存器（RNG）等，这里我们不再赘述。

2. 伪随机数的统计检验

前面已经提到，伪随机数特性好坏是通过各种统计检验来确定的，这些检验包括均匀性检验、独立性检验、组合规律检验、无连贯性检验、参数检验等等[1, 2]。其中最基本的是它的均匀性和独立性的好坏检验。所谓均匀性是指在[0，1]区域内等长度区间的随机数分布的个数应相等。独立性是按先后顺序出现的若干个随机数中，每一个数的出现都和它前后的各个数无关。下面就介绍这两种检验。需要指出的是：一个好的伪随机数序列除了能通过这两种主要的统计检验外，还需要能通过别的多种检验。能通过的检验越多，则该产生器就越优良可靠。

（1）均匀性检验——频率检验。均匀性检验的方法很多。这里介绍χ^2方法。设有在区间[0，1]上的伪随机数序列为$\{\xi_1,\xi_2,...,\xi_N\}$。如果该伪随机数是均匀分布的，则将[0，1]区间分成k个相等的子区间后，落在每个子区间的伪随机数个数N_i应当近似为N/k。此数也称频数。它的统计误差$\sigma_i=\sqrt{N_i}=\sqrt{N/k}$。统计量$\chi^2$按定义应为

$$\chi^2=\sum_{i=1}^{k}\frac{(N_i-N/k)^2}{N/k}=\frac{k}{N}\sum_{i-1}^{k}\left(N_i-N/k\right)^2 \tag{2.2.6}$$

χ^2在此问题中应服从$\chi^2(k-1)$的分布。据此可以假定一个显著性水平值来进行检验。我们可以从χ^2表查得$(k-1)$个自由度的显著水平为α时的t_α值。如果由(2.2.6)式计算出来的χ^2小于t_α，则认为在α的显著水平下，原伪随机数在[0，1]区间是均匀分布的假定是正确的。如果计算得到的χ^2大于t_α，则认为在α的显著水平下，伪随机数不满足均匀性的要求。通常取显著水平为0.01或0.05。为了反映均匀性分布的特性，k的取值不宜太小，但也不能太大。一般选取的k值，要能使每个子区间有若干个伪随机数时就比较合适。

（2）独立性检验——无重复联列检验。这里我们也只介绍独立性检验的一种比较简单的方法。如果把[0，1]上的伪随机数序列$\{\xi_1,\xi_2,...,\xi_{2N}\}$分成两列：

$$\xi_1,\xi_3,...\xi_{2i-1},...,\xi_{2N-1}$$
$$\xi_2,\xi_4,..\xi_{2i},...,\xi_{2N}$$

第一列作为随机变量x的取值，第二列作为随机变量y的取值。在x–y平面内的单位正方形域$[0\le x\le 1，0\le y\le 1]$上，分别以平行于坐标轴的平行线，将正方形域分成$k\times k$个相同面积的小正方形网格。落在每个网格内的随机数的频数n_{ij}应当近似等于$2N/k^2$。由此可以算出χ^2为

$$\chi^2=\sum_{i,j=1}^{k}\frac{k^2}{2N}\left(n_{ij}-\frac{2N}{k^2}\right)^2 \tag{2.2.7}$$

χ^2应满足$\chi^2\left((k-1)^2\right)$的分布。据此可以采用均匀性检验的$\chi^2$方法，假定出显著性水平来进行检验。我们也可以把伪随机数序列分为三列、四列……，用与上面所述相似的方法进行多维独立性检验。

3. 独立于计算机机型的伪随机数产生器

上面曾介绍过的伪随机数产生器是与计算机所能容纳的整数位数有关。在实际应用中，我们有时希望使用能够在各种型号的计算机上工作、并产生相同伪随机数序列的产生器。这种产生器的实现基于如下的思想：如果要产生[0,1]区间的伪随机浮点数，可以选择精度最低的计算机作为标准精度。而对字长较长的计算机，我们将用较低的数位人为置零的方法，即在高精度的计算机上进行较低精度的运算。一般来说，这样的伪随机数产生器无论从伪随机数的重复周期和产生伪随机数的速度都不算理想，但它却可以在大多数较大的计算机上工作。 这里我们以 CERN 程序库中的伪随机数产生子程序 RN32 为例。该程序选择 IBM 计算机的 32 位字长作为最小精度。缺省的起始整数为 65539，也可以输入“ 种子 ”作为起始整数。将起始整数(或前一个整数)乘以 69069，结果只保留较低的 31 位数，这个 31 位整数又作为下一个伪随机数的“ 种子 ”。浮点伪随机数是通过如下操作得到的：将“ 种子 ”的最后 8 位数置零，以保证浮点整数的表示；再将此结果乘以2^{-31}就得到伪随机浮点数。不同计算机上的 RN32 子程序的 FORTRAN 语法及浮点表示有稍许不同。下面便是在 RN32 子程序的源程序的 CDC 及 IBM 版本。这些伪随机数产生器产生的前几个数近似为：R1=0.10791504…，R2=0.58747506…。

```
      FUNCTION RN32(IDUMMY)
C       CDC VERSION    F.JAMES,   1978
```


需要全本请访问 www.ertongbook.com