中国原子能科学研究院科学技术丛书

数据的分析和处理

赵志祥　编著

中国原子能出版社

中国原子能科学研究院科学技术丛书

数据的分析和处理

赵志祥　编著

中国原子能出版社

图书在版编目(CIP)数据

数据的分析和处理 / 赵志祥编著. —北京:中国原子能出版社,2012.12

ISBN 978-7-5022-5770-5

Ⅰ.①数… Ⅱ.①赵… Ⅲ.①统计数据—统计分析(数学)②数据处理 Ⅳ.①O212.1②TP274

中国版本图书馆 CIP 数据核字(2012)第 272329 号

数据的分析和处理

出版发行	中国原子能出版社(北京市海淀区阜成路 43 号 100048)
责任编辑	黄厚坤
责任校对	冯莲凤
责任印制	潘玉玲
印　　刷	北京盛通印刷股份有限公司
经　　销	全国新华书店
开　　本	787mm×1092mm 1/16
印　　张	24.5
字　　数	466 千字
版　　次	2012 年 12 月第 1 版 2012 年 12 月第 1 次印刷
书　　号	ISBN 978-7-5022-5770-5
印　　数	1—2500 **定　价 88.00 元**

网址:http://www.aep.com.cn **E-mail:atomep123@126.com**

发行电话:010-68452845

《中国原子能科学研究院科学技术丛书》

总　序

中国原子能科学研究院创建于1950年，是我国核科学技术的发祥地和先导性、基础性、前瞻性的综合性核科学技术研究基地。

在党中央和上级部门的关怀和指导下，中国原子能科学研究院为我国的国防建设、国民经济建设和核科学技术的发展做出了重要贡献，造就了7位“两弹一星”功勋科学家和60多位两院院士，培养了大批科技人才，在核物理、核化学与放射化学、反应堆工程技术、加速器工程技术、同位素技术、核电子学与核探测技术、辐射防护、放射性计量等学科形成了自己的特色和优势，并拥有核科学与技术和物理学两个一级学科硕士、博士学位授予权。

为了系统地总结原子能院在核科学技术相关优势学科积累的知识和经验，吸收和借鉴国内外核科学技术最新成果，促进我国核科技事业的发展，我院决定组织出版《中国原子能科学研究院科学技术丛书》，并选定王淦昌、肖伦、丁大钊、王乃彦、阮可强等院士编著的《惯性约束核聚变》、《放射性同位素技术》、《中子物理学——原理、方法与应用》、《新兴的强激光》、《核临界安全》5本专著首批出版，今后还将组织撰写更多的学术专著纳入本丛书系列。

谨以此套丛书献给为我国核科技事业献身的人们！

《中国原子能科学研究院科学技术丛书》出版委员会

2005年9月1日

前　言

在一切做定量测量和判断的科学研究及其应用的领域，都要认真考虑有关量值的误差及这些误差在量值传递过程中的性质和行为。典型的例子是在核科学技术及其应用的领域，在那里，不仅数据本身，而且数据的误差都受到了越来越多的重视。数据和数据的误差紧密地与核工程的经济性和安全性等重大问题联系起来。为了确保核工程与核产品的安全性，在设计上必须要留出余量，而为了获得更好的经济性，这个余量又不能留得太多。在核工业的许多工程领域，关键数据的误差哪怕减少一点，都会带来巨大的经济效益。在安全性和经济性之间的这种博弈，使得对数据的误差分析和处理过程的科学研究工作具有重大的现实意义。

从20世纪70年代以来，由于核工程设计能力和水平的提高，微观核数据的误差往往居于工程设计计算量误差的主导地位。这一事实不仅推动了核数据工作的发展，而且导致了世界范围的核数据专门队伍的形成，促进了核数据处理和误差分析研究工作的开展，而大型计算机的发展和快速计算能力的提高使得上述进展成为可能。

作者从20世纪70年代末期开始进入了核数据处理和误差分析这一领域，并在这一领域持续工作了三十多年。因而，本书在传统的章节中在下列方面相当多地融入了作者多年的研究成果和观点：

关于误差的分类　数据的误差可以分成偶然误差和关联误差两大类，它们在量值传递过程中的性质有着明显的差异。对于数据评价和使用者来说，还可能遇到一种不包括在报道误差中的疏失误差，它多半也是一种关联误差，只是由于难以在一个测量系统上发现而被疏失了。

宏观测量系统和微观测量系统　在实验数据的数学处理中，三种不同性质的误差在一定前提下可以认为都具有偶然误差的随机性质。微观测量系统和宏观测量系统的引进有助于描述和理解三类误差性质。在同一微观系统上对同一物理量做重复测量，测量的关联误差不因测量

次数的增加而减少，在同一微观系统上对某一物理曲线做测量时，测量曲线的关联误差不会因测量点的加密而减少，在同一微观系统上甚至不能发现可能存在的关联误差。而每一微观系统中的关联误差在宏观系统中具有偶然误差的性质。基于这一点，尽可能地设计独立于其他微观系统的实验可以使得关联误差随机化。

协方差矩阵 20世纪70年代中期以前，关联误差在进行数据处理和误差分析时常常被忽略。这不仅反映在有关的专著中，也反映在一些实际使用的数据处理程序中。这之后，情况发生了变化。在工程设计中，不考虑数据的关联误差便不能正确地估计计算量的误差。这就要求在数据处理的全部过程中考虑全部的误差信息—协方差矩阵，同时要求实验工作者做完整的误差报道，即报道包括各种关联误差的协方差矩阵。

协方差矩阵的引入对过去常用的数据处理和误差分析方法的适用性提出了挑战。本书建立了适用于用协方差矩阵表示的关联误差存在时的数据处理方法和公式体系。

当处理带协方差矩阵的数据的合并与拟合时，将会遇到一个高阶矩阵的求逆问题，计算上费时很多，而且容易出现病态矩阵问题，本书给出了一些在适当近似下解决高阶矩阵求逆的方法和技巧。

数据空间和导出量空间 在数据空间和导出量空间应用最小二乘法得到的结果是等价的，但是迭代的方式不同。按照人们的传统习惯在导出量空间处理数据非常容易犯错误，因此要特别小心。作者强烈建议尽可能在数据空间处理数据。

导出量处理中的PPP问题 数据处理的对象常常是导出量，其为直接测量量的函数。当考虑关联误差时，使用现成的误差传递公式和最小二乘拟合公式处理导出量的合并及拟合时，常常发生极其荒谬的现象。这一由美国ORNL实验室的R. Peelle所遇到并提请国际核数据界注意的问题，即所谓PPP(Peelle's Pertinent Puzzle)问题，困扰了国际核数据界十余年。PPP总是在导出量空间发生，是不正确地理解和使用误差传递公式和最小二乘法造成的。作者对这一问题的解决，使相关数据的处理实际成为可能，受到了国际核数据界的重视。

贝叶斯方法 对贝叶斯方法的实际应用，历史上一直存在争论，相

当长的一个时期贝叶斯方法受到“冷落”，一些统计学家甚至主张完全放弃它。近三十年来，贝叶斯方法有了很大的进展，由于群论方法的引进，采用贝叶斯假设即无差别原理确定验前分布所引起的悖论均被解决。在包括核数据等在内的许多领域，贝叶斯方法得到了越来越多的应用。本书介绍了现代贝叶斯统计推断的基本概念和应用实例。

模型理论计算值的误差 在本书中讨论这一问题是因为在某种意义上它可以等价为一个实验数据的数学处理问题。和曲线拟合一样，需要解决的是如何由实验数据的协方差矩阵估计参数的协方差矩阵，不同的是，还必须考虑到因理论模型的缺陷带来的不确定性对理论预言值的影响。本书基于作者的工作经验对这一问题进行了初步的探讨。

以上可以认为是体现本书特色的一些方面。

为了使读者易于理解本书所述的概念和结论，作者给出了一些数据测量和评价的例子。这些实例大多是实际问题的简化。本书的大部分章节曾作为教材由作者在核工业研究生院讲授过。我们希望本书成为数据测量、评价和使用者掌握数据处理方法的手册，它应当也是其他领域科技工作者的一本有价值的参考书。

由于水平有限，本书难免存在缺点和错误，恳望读者不吝指正。应当指出的是，本书是按照实验测量和数据评价的观点和要求写成的。有些问题的表述在数学家看来也许是不严格的。比如当写出一个积分时，并不去说明它是否收敛，当对一个函数求导时，也并不去证明这个函数的导数一定存在等。作者希望本书能给读者一个清晰的思路和实验测量、数据评价工作者一个能够理解的结论。

作　者

2011 年 10 月

目　录

第1章　随机事件和随机变量

1.1　随机事件及其概率

1.1.1　随机事件

随机事件是指统计意义下的偶然事件，它在一次试验中可能发生，也可能不发生。我们无法肯定地回答一个随机事件在一次试验中是否能够发生。例如，我们无法肯定地回答“某一放射性原子核在时间 t 内是否衰变”或“某一测量值是否高于真值”这样一些问题。在这两个例子中，“某一放射性原子核在时间 t 内衰变”或“某一测量值高于真值”都是随机事件。随机事件并不是毫无规律可循，所谓“随机”是指一次试验而言。如果多次重复试验，随机事件就会显示出很好的统计规律性。有时把随机事件简称为事件。

1.1.2　随机事件概率的定义

定量地描述随机事件 A 发生的可能性大小的量称为随机事件 A 的概率，记为 $Pr(A)$。

在概率论发展的历史上，曾经有几种定义概率和计算概率的方法。

古典概率

设某一随机试验中，可能发生的全体基本事件总数是有限的，且具有相等的可能性，则对于任意事件 A，若基本事件总数为 n，事件 A 包含的基本事件数目为 k，则事件 A 发生的概率由下式计算

$$Pr(A)=\frac{k}{n} \tag{1.1.1}$$

上式称为**古典概率**。

例 1.1　短袜问题

2 只黄色短袜和 8 只白色短袜混放在抽屉中，随机地抽出三只，得到一双黄色短袜的概率有多大？

基本事件总数：$n=C_{10}^{3}=\dfrac{10\times 9\times 8}{1\times 2\times 3}=120$

抽出 2 只黄色短袜加任何一只白色短袜的基本事件数目：