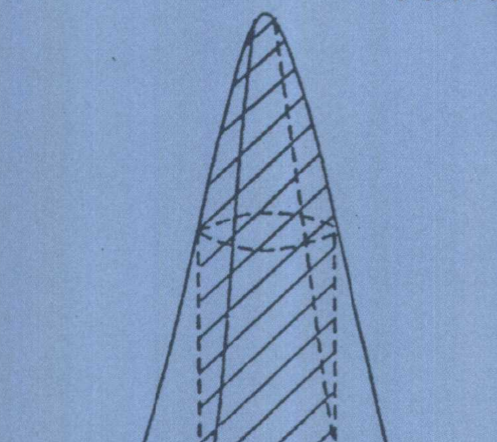
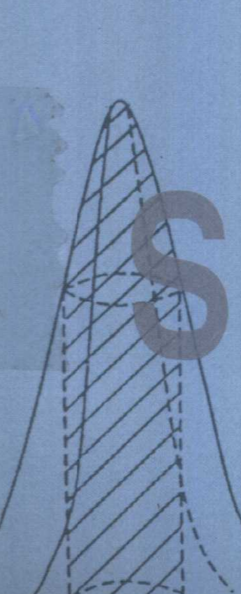
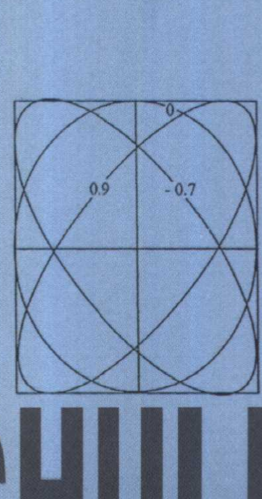
实验的数据处理

■ 李耀清/编著

中国科学技术大学出版社







SHIYAN DE SHUJU CHULI

实验的数据处理

李耀清　编著

中国科学技术大学出版社
2003・合肥

内容简介

介绍物理实验中常用的数据处理方法和工具，包括概率数理统计和数字信号处理两大部分。主要内容有：概率统计基础、误差理论、参数估计（包括贝叶斯估计）、假设检验、最小二乘法与曲线拟合、序列卷积与离散富里叶变换、快速富里叶变换、实验谱的去卷积等，阐述重点放在物理概念的理解和实际应用方面，书中附有大量的实例，便于自学。其内容适合作为实验物理类本科生和研究生相关课程的教材。

图书在版编目(CIP)数据

实验的数据处理/李耀清编著. —合肥：中国科学技术大学出版社，2003.9

ISBN 7-312-01589-1

Ⅰ. 实…　Ⅱ. 李…　Ⅲ. 物理-实验数据-数据处理-高等学校-教材　Ⅳ. O4—33

中国版本图书馆 CIP 数据核字(2003)第 053692 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，邮编：230026，发行电话：0551-3602905，3602906）

合肥义兴印务有限责任公司印刷

全国新华书店经销

开本：787mm×1092mm　1/16　　印张：14　　字数：360 千

2003 年 9 月第 1 版　　2003 年 9 月第 1 次印刷

印数：1—3000 册

ISBN 7-312-01589-1/O·273　　定价：21.00 元

序

本书是在我多年讲授的《实验的数据处理》课讲义基础上整理而成的。书中内容大致根据我所在的核与粒子物理专业教学和科研实践的需要选取，适当兼顾其它物理专业，经不断的补充、完善，便形成今天的样子。

在各种高科技飞速发展的今天，实验物理类学生面临着需要学习和掌握的知识越来越繁杂的局面。如何将以往传统的教学内容浓缩筛选，并补充进新的知识，以便在有限的学时内完成预期的教学计划，这是当今每一个高校教师都面对的课题。笔者也是遵循这一思路来选取和安排本书内容的。

在以往的同类教材中，《实验的数据处理》内容大多在概率与数理统计的范围内选取。随着当今数字计算技术的广泛普及应用，数字信号处理在物理实验的数据处理中占据越来越重要的地位，已成为当今物理实验工作者必须掌握的基本技术中的一部分。因此，在浓缩数理统计部分内容的基础上，笔者将数字信号处理中自认为是必须了解和掌握的重要内容纳入了本教材。此外，根据我所在专业的需要，书中还收集了实验谱去卷积方面的有关内容。内容广而杂，算是本书的一个特色吧。

作为实验物理的一门专业基础课程，本教材更注重于物理概念的阐述和实际的应用方面。因此书中尽量避免繁琐的数学推导和证明，使学生的注意力更多集中于基本概念的理解和方法的实际应用上。

特别要感谢李惕碚教授，他的《实验的数学处理》一书给我提供了一个很好的参照系。同样要感谢清华大学与南京大学的同行，他们的同类教材给我提供了部分实例。

本书可作为物理类研究生及高年级本科生相关课程的教材。

笔者水平有限，书中错误在所难免，欢迎各位批评指正。

作　　者

2003 年 9 月于合肥

前　言

做过物理实验的人都有这样的经验，实验观测值具有一定的偶然性。例如，放射性计数的测量，每次测量其计数率一般都不相同。当放射源的强度衰减在测量期间可忽略不计时，计数率一般都在某一固定值上下波动。即使测量一个物体的重量或长度，不同的人或不同次测量其结果也会不同，依然是在某一固定值上下波动。所谓实验的数据处理就是从带有偶然性的实验观测值出发，用数学方法导出规律性结论的过程。偶然性，又叫随机性。物理实验的观测结果具有随机性的原因在于：

1. 测量的偶然误差

由于实验技术水平的限制，总会存在着某些观测者尚不能完全控制的偶然因素，例如温度、湿度、气流等，造成测量的偶然误差。此外，观测者本身感官分辨本领的限制也是偶然误差的一个来源 。如测量金属丝的直径、杆的长度，以及读指针仪表时，眼的歪斜会造成测量结果的偶然误差。偶然误差的存在使测量数据带有随机性。

2. 物理现象本身的随机性

按照量子力学测不准原理，对处于同一个态的微观粒子，测量同一个可观测的物理量时，即使不存在任何测量误差，各次测量结果也不会相同，除非粒子处于这个可观测量的本征态。例如，同一种基本粒子的寿命，其实测值分布在相当长的时间范围内，服从指数分布。又如，高能物理实验中产生的不稳定粒子的质量也有一个分布，具体到某一个粒子，其质量是不确定的。这类测量结果的离散程度不是由测量的偶然误差造成，而是由物理现象本身固有的随机性决定的。

正是由于上述原因，使得物理过程的规律性往往被表面的偶然性所掩盖，必须运用适当的数学工具才能由实验观测数据导出正确的结论来。

例如，用闪烁计数器和火花室组成的荷电粒子探测器寻找宇宙线中的分数荷电粒子（夸克）。仪器记录$\frac{1}{3}$e粒子的效率$P'=0.55$，实验条件$\Omega ST=1.9\times10^{9}$ sr. cm^2. sec，其中，Ω(sr)是仪器的有效立体角，S(cm^2)是探测器的工作面积，T(sec)是工作时间。如果在实验的全部工作时间里，没有记录到$\frac{1}{3}$e粒子，问：宇宙线中有没有$\frac{1}{3}$e粒子？如果有，流强是多少？

显然，这个问题不能简单地回答没有，也不能用一个准确的数字来描述流强的大小。随着本课程的深入，这个问题将会得到较好的解决。从这个例子可以看出，实验数据处理的内容和重要性已大大超出了经典的误差处理范畴。

除了经典的误差分析外，实验数据的处理还包括以下方面的内容：

1. 如果我们已知理论上存在某个物理规律 $y=f(x;c)$，c 为已知的参数，要用实验来验证这一规律。由于种种原因，实验测得的物理量观测值 y^* 不可能同理论值 y 完全吻合。如何根据实验观测结果对理论上的物理规律作出正确判断呢？这类问题属于假设检验问题。

2. 如果物理公式 $y=f(x;c)$ 中的参数 c 并不知道，要从实验观测值 y^* 出发，运用数理统计方法来确定参数 c 的数值及其置信范围。这类问题属于参数估计。

3. 预先根本不知道物理规律，要从实验数据中寻找出物理规律，或者寻找某个经验的函数表达式来近似物理规律，这类问题属于曲线拟合的范围。

随着计算机技术的飞速发展，数字信号处理技术在实验数据的处理中已得到了广泛应用。例如，谱的去卷积，投影的图像重建等领域都用到了数字信号处理技术。可以说掌握基本的数字信号处理方法已成为对每一个实验物理工作者的基本要求。因此，包括快速富里叶变换在内的数字信号处理的一些重要内容也纳入了本书。

此外，谱的去卷积也是当今实验物理工作者经常面临的技术问题，书中也作了适当的介绍。

作　　者

2003 年 7 月于合肥

目　录

第1章　随机变量及常用分布

从前言的介绍我们已知道物理实验结果具有随机性。在处理这种具有随机性的测量数据时，必须应用随机量数学——定量研究随机现象规律性的科学。否则，就得不出合理的结论，也无法知道结论的可靠程度。

随机量数学是一门正在迅速发展的学科，它包括概率论、数理统计和随机过程理论，在工农业生产、科学研究及经济领域中正在得到越来越广泛的应用。本章的内容是介绍概率数理统计的基础知识。

1.1　随机事件的概率

1.1.1　随机事件及其概率

在一定的试验条件下，现象 A 可能发生，也可能不发生，并且只有发生或不发生这两种可能性。这是偶然现象中比较常见的一种形态。我们把发生了现象 A 的事件叫做随机事件 A，简称事件 A。

例如：掷硬币试验，可以把正面朝上的现象定为事件 A。如果掷多次，会发现正面朝上的次数总是在试验总次数的二分之一左右摆动。

如果在即定条件下进行试验，总共试验 N 次，其中现象 A 发生了 N_A 次，则称比值 N_A/N 为事件 A 的频率。重复进行多组这样的试验，会发现事件 A 的频率总是在某个值的上下摆动，并且随着每组试验次数 N 的增多，频率上下摆动的平均幅度趋于减小。因此，在统计意义上，事件 A 的频率存在一极限值，叫做事件 A 的概率，记作 $P_r(A)$。

$$P_r(A) \equiv \lim_{N\to\infty} \frac{N_A}{N} \tag{1.1}$$

例如，一袋中装有三个白球，七个黑球，则在袋中摸到白球的概率就应是 0.3。

根据随机事件的概率定义可知，$0 \leqslant P_r(A) \leqslant 1$。

若试验中事件 A 从不出现，则 $P_r(A)=0$。

若每次试验事件 A 必定出现，则 $P_r(A)=1$。

事件之间总是有联系的，下面介绍事件之间的几种关系。

1）事件的和 $A+B$

若事件 A 与事件 B 是两个不同的随机事件，A 和 B 的和事件是指 A 与 B 至少有一个发生的事件。如果用两个圆分别表示事件 A 和 B 的集合，如图 1.1 所示，画斜线的区域就代表事件 $A+B$ 的集合。显然，只有事件 A 发生，或者只有事件 B 发生，或者事件 A 和事件 B 同时发生的事件都属于事件 $A+B$。

例如，有四张扑克牌，方块、红心、黑桃和梅花各一张。每次摸两张牌，其中含有方块的情况定为事件 A，含有红心的情况定为事件 B。事件 $A+B$ 就是摸到的两张牌中至少

有一张是方块或红心的事件。

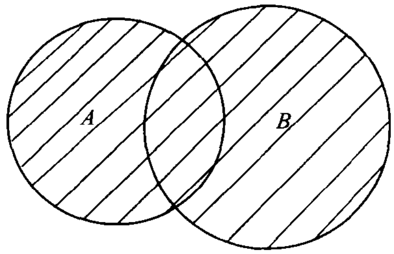


图 1.1　事件的和

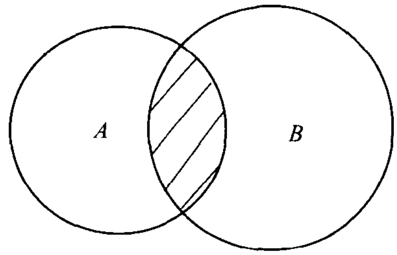


图 1.2　事件的积

2)事件的积 AB

定义事件 A 与事件 B 的积事件 AB 为 A 和 B 同时发生的事件。图 1.2 中两个事件重叠的区域就是积事件 AB 的区域。例如,前例中,同时摸到方块和红心的事件就是积事件 AB。

和事件和积事件的定义可以推广到多个事件的情况。

和事件 $A_1+A_2+\cdots A_N$ 是指 $A_1,A_2,\cdots,A_N$ 中至少有一个发生的事件。

积事件 $A_1A_2\cdots A_N$ 是指 $A_1,A_2,\cdots,A_N$ 同时发生的事件。

3)互斥事件

如果事件 A 和事件 B 不可能在一次事件中同时发生,即

$$P_r(AB)=0$$

则称事件 A 与事件 B 是互斥事件。图 1.3 中的事件 A 与事件 B 就是互斥事件。

例如,掷硬币,出现正面与出现反面这两个事件就是互斥事件。

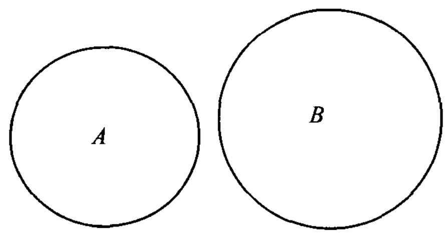


图 1.3　互斥事件

4)逆事件

对事件 A 而言,不是 A 的其它事件都称为 A 的逆事件,记作 $\overline{A}$。

1.1.2　随机事件的概率公式

一、和事件的概率公式

从图 1.1 可以看出,一般情况下,和事件 $A+B$ 的概率不等于事件 A 的概率与事件 B 的概率之和 $P_r(A)+P_r(B)$,因为两个圆重叠的部分,即 A 和 B 同时发生的事件的概率计算了两次,所以 A 和 B 的和事件的概率公式应当是

$$P_r(A+B)=P_r(A)+P_r(B)-P_r(AB) \tag{1.2}$$

如果 A 和 B 是互斥事件,即 $P_r(AB)=0$,则有

$$P_r(A+B)=P_r(A)+P_r(B) \tag{1.3}$$

显然,对于 N 个两两互斥的随机事件 $A_1,A_2,\cdots,A_N$,其和事件的概率为

$$P_r\left(\sum_{i=1}^{N}A_i\right)=\sum_{i=1}^{N}P_r(A_i) \tag{1.4}$$

例 1.1 有两个摇号机，每个摇号机每次可摇出 0,1,…,9 等十个数字中的一个数。摇两次号，第一次在其中的一个机上摇，第二次在另一个机上摇。用 A 表示第一次摇到数字 6 的事件，B 表示第二次摇到数字 6 的事件，则至少摇到一个数字 6 的概率是

$$\begin{aligned}P_r(A+B) &= P_r(A)+P_r(B)-P_r(AB)\\ &= 1/10+1/10-1/10\times 1/10 = 19/100\end{aligned}$$

二、条件概率

B 发生的条件下 A 发生的概率，叫做 A 对于 B 的条件概率，记作 $P_r(A\backslash B)$。

如果在 N 次试验中，B 发生 N_B 次，A 和 B 同时发生 N_{AB} 次，按定义

$$P_r(A\backslash B) = \lim_{N\to\infty}\frac{N_{AB}}{N_B} = \lim_{N\to\infty}\frac{N_{AB}/N}{N_B/N} = \frac{P_r(AB)}{P_r(B)} \tag{1.5}$$

例如，从一副扑克牌中连抽两次牌，用 A 表示第一次抽到方块，B 表示第二次抽到方块。则在第一次抽到方块的条件下第二次抽到方块的概率可写作 $P_r(B\backslash A)$。

三、积事件概率公式

由(1.5)式可得积事件概率公式

$$P_r(AB) = P_r(B)P_r(A\backslash B) \tag{1.6}$$

同理推得

$$P_r(AB) = P_r(A)P_r(B\backslash A) \tag{1.6'}$$

利用(1.6)式，容易导出多个事件之积的概率公式为

$$\begin{aligned}&P_r(A_1A_2\cdots A_{N-1}A_N) = P_r(A_1A_2\cdots A_{N-1})P_r(A_N\backslash A_1A_2\cdots A_{N-1})\\ &= P_r(A_1A_2\cdots A_{N-2})P_r(A_{N-1}\backslash A_1A_2\cdots A_{N-2})P_r(A_N\backslash A_1A_2\cdots A_{N-1})\cdots\cdots\\ &= P_r(A_1)P_r(A_2\backslash A_1)P_r(A_3\backslash A_1A_2)\cdots P_r(A_N\backslash A_1A_2\cdots A_{N-1})\end{aligned} \tag{1.7}$$

例 1.2 从一副扑克牌中连抽两张都是方块的概率。用 A 表示第一张抽到方块，B 表示第二张抽到方块，A 和 B 同时出现的概率为

$$P_r(AB) = P_r(A)P_r(B\backslash A) = 1/4\times 12/51 = 1/17$$

对于事件 A 和事件 B，如果 A 事件发生的概率不受 B 事件是否发生的影响，即

$$P_r(A\backslash B) = P_r(A) \tag{1.8}$$

则称事件 A 独立于事件 B。事件独立的意义在于，事件 A 在总体中所占的比例与事件 A 在事件 B 中所占的比例是相等的。

例如，从一副扑克牌中抽牌，抽到方块为事件 B，抽到 K 为事件 A，则事件 A 对事件 B 是独立的。因为

$$P_r(A\backslash B) = P_r(A) = 1/13$$

独立是相互的，若事件 A 独立于事件 B，必有事件 B 独立于事件 A，即

$$P_r(B\backslash A) = P_r(B) \tag{1.8'}$$

对于独立事件 A 和 B，积事件概率公式(1.6)可写成

$$P_r(AB) = P_r(B)P_r(A) \tag{1.9}$$

反之，如果事件 A 和 B 的概率满足式(1.8)、(1.8′)及(1.9)中的任意一个，则事件 A 和 B 就是互相独立的事件。

例 1.3 图 1.4 为一继电器网络,继电器导通的事件用 E_i 表示($i=1,2,3$)。每一个继电器导通的概率为 α。求 A,B 两端点之间导通的概率 $P_r(E)$。

根据题意,$P_r(E_1)=P_r(E_2)=P_r(E_3)=\alpha$

$$\begin{aligned}P_r(E) &= P_r[E_1+(E_2E_3)]\\ &= P_r(E_1)+P_r(E_2E_3)-P_r(E_1)P_r(E_2E_3)\\ &= P_r(E_1)+P_r(E_3)P_r(E_2)-P_r(E_1)P_r(E_2)P_r(E_3)\\ &= \alpha+\alpha^2-\alpha^3\end{aligned}$$

本题求解中视各继电器导通的事件是独立的。

四、全概率公式

用符号$\{A_i\}$表示某一组事件(A_1,A_2,…,)。如果任意一次试验的结果都至少发生$\{A_i\}$中的一个事件,则称$\{A_i\}$是一个事件的完备集。显然,对于事件的完备集$\{A_i\}$,有

$$P_r(\sum_i A_i)=1$$

式中,$\sum\limits_i$ 表示对$\{A_i\}$中所有的事件求和。

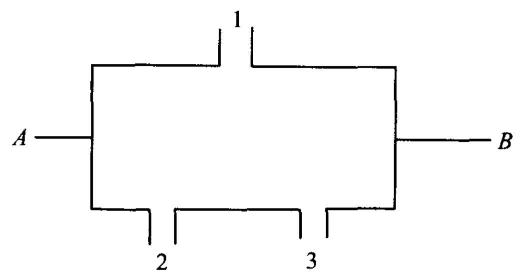


图 1.4 继电器网络

如果$\{A_i\}$是一个互斥事件的完备集,由式(1.4)可将上式变为

$$\sum_i P_r(A_i)=1$$

而对任一随机事件 B 有下列全概率公式:

$$P_r(B)=\sum_i P_r(B\backslash A_i)P_r(A_i) \tag{1.10}$$

全概率公式的证明如下:由于$\{A_i\}$是事件的完备集,B 总是伴随$\{A_i\}$中的事件同时发生,又由于$\{A_i\}$中的事件是两两互斥的,B 又只能伴随$\{A_i\}$中的一个同时发生。所以,事件 B 可以表示成下列互斥的积事件之和:

$$B=A_1B+A_2B+\cdots=\sum_i A_iB$$

利用公式(1.4)和(1.6)可得

$$P_r(B)=\sum_i P_r(A_iB)=\sum_i P_r(B\backslash A_i)P_r(A_i)$$

例如,一副扑克牌,方块、红心、梅花和黑桃构成互斥完备集$\{A_i\}$,设抽到 K 为随机事件 B,则抽到 K 的概率 $P_r(K)$为

$$P_r(K)=\sum_i P_r(K\backslash A_i)P_r(A_i)=\sum_{i=1}^{4}(\frac{1}{13}\times\frac{1}{4})=\frac{1}{13}$$

五、贝叶斯定理

由式(1.6)和式(1.6′)可导出

$$P_r(A\backslash B)P_r(B)=P_r(AB)=P_r(B\backslash A)P_r(A)$$

因此有

$$P_r(A\backslash B) = \frac{P_r(B\backslash A)P_r(A)}{P_r(B)} \tag{1.11}$$

式(1.11)就是关于条件概率的贝叶斯定理。贝叶斯定理告诉我们，事件 A 对于事件 B 的条件概率，可以由事件 A 和事件 B 的概率以及 B 对于 A 的条件概率算出。事实上，根据概率和条件概率的定义也可以直接导出贝叶斯定理。

设 N 为总的试验次数，N_{AB} 为 A 和 B 同时发生的次数，N_A 为 A 发生的次数，N_B 为 B 发生的次数，有

$$\frac{P_r(B\backslash A)P_r(A)}{P_r(B)} = \lim_{N\to\infty} \frac{\dfrac{N_{AB}N_A}{N_A N}}{\dfrac{N_B}{N}} = \lim_{N\to\infty}\frac{N_{AB}}{N_B} = P_r(A\backslash B)$$

如果 $A = \sum\limits_i A_i$，而 $\{A_i\}$ 是一个互斥事件的完备集，利用全概率公式(1.10)，贝叶斯公式又可写成

$$P_r(A\backslash B) = \frac{P_r(B\backslash A)P_r(A)}{\sum\limits_i P_r(B\backslash A_i)P_r(A_i)} \tag{1.11$'$}$$

例 1.4 分别用 B_1、B_2 和 B_3 表示三人，每人持二钱币。B_1 有两个金币，B_2 有一金币和一银币，B_3 有两个银币。随机从一人的手中拿走一枚钱币，假设拿走的是一枚金币。试问这人手中余下一枚也是金币的概率是几何?

设 A 表示所拿走的一枚是金币的事件，实际上就是要计算在事件 A 发生的条件下，被拿走金币者是 B_1 的条件概率 $P_r(B_1\backslash A)$。

利用(1.11′)式有

$$P_r(B_1\backslash A) = \frac{P_r(A\backslash B_1)P_r(B_1)}{\sum\limits_i P_r(A\backslash B_i)P_r(B_i)}$$

根据题意有

$$P_r(A\backslash B_1) = 1, P_r(A\backslash B_2) = 1/2, P_r(A\backslash B_3) = 0, P_r(B_i) = 1/3 (i = 1,2,3)$$

因此有

$$P_r(B_1\backslash A) = \frac{1\times 1/3}{1\times 1/3 + 1/2\times 1/3 + 0\times 1/3} = 2/3$$

贝叶斯定理在统计推断理论中占有重要的地位。实验数据处理的任务常常是根据观测结果来推断某个物理量的真值。更确切的说，是推断在出现该观测结果的条件下物理量取各种可能值的概率。适当地应用贝叶斯定理，可以把问题转化为：在物理量取各种可能值的条件下，求出出现特定观测结果的概率，而这个概率通常可以计算出来。第三章介绍参数估计的贝叶斯方法时，将更详细地讨论贝叶斯定理在统计推断问题上的应用。

1.2 随机变量及其概率分布

1.2.1 随机变量和随机样本

考虑一个简单的摇号试验，摇出的号数是一随机事件，它取 1,2,…,9,0 之中的一个