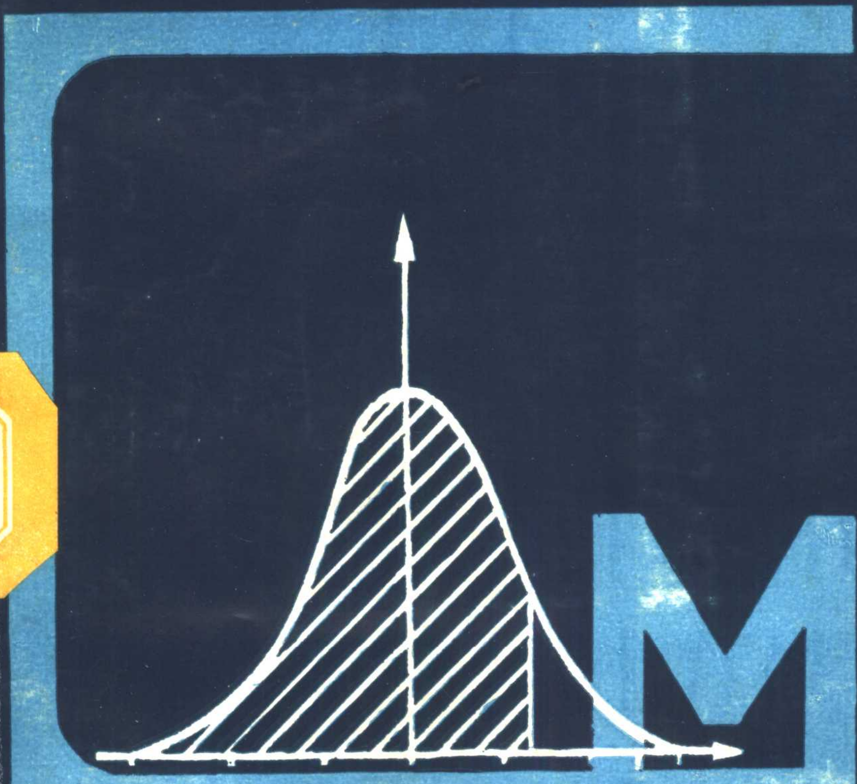
贵州人民出版社

分析化学中常用数理统计方法

焦云飞 编著



分析化学中
常用数理统计方法

焦云飞　编著

贵州人民出版社

本书受贵州省科技图书出版基金委员会资助

责任编辑
封面设计 施福根

分析化学中常用数理统计方法
焦云飞 编著
贵州人民出版社出版发行
(贵阳市延安中路9号)
贵州新华印刷厂印刷 贵州省新华书店经销
787×1092毫米 32开本 8.25印张 170千字
1990年1月第1版 1990年1月第1次印刷
印数1——1,500
ISBN 7-221-01365-9
O·06 定价：2.70元

前　言

数理统计学是数学领域里新兴的一门数学分支，随着这一分支在其它各类学科中的广泛应用，逐渐成为各学科研究工作中的得力工具。分析化学也不例外，近年来亦大量地使用数理统计方法去处理实验中获得的各种数据，综合各种信息，判断实验结果的好坏，评价实验方法的优劣，指导实验方案的设计与选择，等等。可以肯定，由于分析化学本身的不断发展，数理统计方法将会在分析化学的各个领域中得到更加广泛的应用。

为了更好地普及推广数理统计方法在分析化学中的应用，帮助广大分析工作者学习掌握这门知识，提高运用这一数学工具的能力，本书通俗地介绍了数理统计学中的一些基本概念，基础知识，为力求理论联系实际，还介绍了这些基础知识在分析化学工作中的应用实例，以便读者仿用；并配有一定数量的练习题。可供分析工作者和高等学校分析化学专业的学生阅读使用。

这里特别感谢：东北师范大学化学系吴立民教授对本书进行了审阅；贵州省理化测试中心朱昭阳同志、中国科学院贵阳地球化学研究所张光宇同志提供了部分资料；贵州师范大学化学系杨长府同志对本书各章节的修改做了大量的工作，提供了部分资料，并对全书进行了最后校阅；贵州师范

大学化学系袁惠民同志也参加了本书的校阅。

由于自己的水平有限，书中难免存在错误和不当之处，敬请广大读者批评指正。

编著者

1989年6月

目　录

第一章　数理统计学中的一些基础知识

§ 1-1　数理统计学简介

数理统计学是数学的一门分支。概括地说，数理统计学是一门关于数据资料的收集、整理、分析和推断的科学。随着概率论的发展，更多地应用概率论对各类统计资料进行深入的分析研究，已成为事实。数理统计的中心任务，就是从分析子样所遵循的规律入手，来判断母体所遵循的规律。

数理统计的方法及考虑的问题，不同于一般的资料统计，它侧重于应用随机现象本身的规律性来考虑资料的收集、整理和分析，从而找出相应的随机变量的分布规律和它的数字特征来。那么什么是随机现象呢？简言之，就是此类现象在个别的试验中表现出不确定性；而在大量重复试验中，又表现出具有统计规律性。显而易见，随着大量的重复性试验的不断进行，且对随机现象进行足够多次的观察之后，被研究的随机现象的规律便能清楚地呈现出来。但实际中所能允许的观察次数永远只能是有限的，有时甚至是少数的。数理统计学的任务就是要利用这些有限的信息，进行统计推断，尽可能作出精确、可靠的结论。

在数理统计里，不是对所研究的全部对象进行观察，而

是抽出其中的一部分进行观察，并通过这些观察所获得的数据，对所研究的全体进行推断。由于推断是根据部分数据作出的，不可能包含研究对象的全部特征，因此，所获得的结论，必然会包含着不定的因素。概率就是这种不定因素的量度。

数据统计学研究的内容随着科学技术和生产的不断发展而逐步扩大，概括地说可分为两类：一类是试验的设计和研究，即研究如何更合理、更有效地获得观察资料的方法；一类是统计推断，即研究如何利用一定的资料对我们所关心的问题作出尽可能精确、可靠的结论。本书拟应用数理统计规律对分析化学中测得的数据进行处理、分析、推断，从而找出规律性的东西来。

§ 1-2 频率与概率

一、必然事件和不可能事件

在自然界中有一些现象，我们完全可以预先知道在一定条件下，它们必然会发生。例如“人总是要死”，“从手上抛出的石块必然要落到地上”等等。因此，在一定条件下，必然发生的某一事件，我们称它为必然事件。

反之，在一定条件下，必然不发生的事件称为不可能事件。例如“人能长生不老”，“没有外力作用，水能从低处向高处流”等等。

二、随机事件及随机试验

在生产实际与科学实验中，经常会遇到这样的事件，在

一定条件下，此事件可能发生，亦可能不发生。例如“投掷一枚均匀的硬币，出现正面向上”，“在一批混有次品的产品中任意抽取一件恰好是次品”等等。这种在一定条件下可能发生也可能不发生的事件，称为随机事件。

为了探索随机事件的规律性，常常需要对随机事件进行观察，这种观察总是在一定条件下进行的，我们把每次观察看做是一个实验，观察的结果就是实验结果。如果在一定条件下实验结果不一样，例如投掷一枚硬币，可能出现正面向上，也可能出现反面向上，即在相同条件下，实验的结果不一样，而每一个可能的实验结果，都有一定的出现机会，我们称这样的试验为随机试验。

某一随机事件的发生与否在大量重复试验中具有某种规律性，揭示和研究这种规律性就是概率论要研究的问题。

三、频率与概率

任一随机试验存在许多可能的结果，观测者常常希望知道其中某一结果出现的可能性有多大。同时还希望将此结果出现的可能性的大小用一个数来表达。

现假设E为任一随机试验，A为其中任一事件（相当于一种可能出现的结果），在完全相同的条件下，把E独立地重复进行n次试验，观察在n次试验中A出现的次数。则E的试验次数和A出现的次数可用下式表示

$$f(A)=\frac{m}{n}$$

式中m是事件A在n次试验中出现的次数，称为频数；$f(A)$是事件A在n次试验中出现的比值，称为频率。频率有时也可用百分数来表示。例如：在固定试验条件下，做某合金钢

中镍含量的测定，重复进行了50次，分析结果如表1-1所列。

表1-1　　某合金钢中镍含量测定

测得值（A_i）	7.32	7.33	7.34	7.35	7.36	7.37	7.38	7.39	7.40	7.41
频　数（m）	1	3	6	9	11	10	6	2	1	1
频　数 $f(A)$	0.02	0.06	0.12	0.18	0.22	0.20	0.12	0.04	0.02	0.02

在E试验条件不变的情况下，对E进行多次重复试验，则A出现的频率便会在一常数P附近摆动，从总的趋势来看，试验次数n越大，$f(A)$的摆动幅度越小，因此在统计意义上就认为随机事件A出现的频率$f(A)$，在$n\to\infty$时存在着极限P，P是常数，是用来刻划随机事件A出现的可能性大小的，称为A出现的概率，记为$P(A)=P$。

历史上有不少人作过成千上万次掷钱币的试验，表1-2列出了部分人的实验记录。

表1-2

实　验　者	投掷次数n	出现“正面朝上”的次数m	频　　率
De Morgn	2048	1061	0.518
Buffon	4040	2048	0.5069
Pearson	12000	6019	0.5016
Pearson	24000	12012	0.5005

不难看出，投掷次数越多，频率越接近0.5，因此，0.5就是“正面朝上”出现的概率。

从概率的统计意义中，我们可以看出概率的几个性质：

（1）因为我们把概率作为频率$\frac{m}{n}$（这里n很大，$0\leqslant m\leqslant n$）的稳定值，所以任何事件A的概率$P(A)$总是介于0与1之间，即：

$$0\leqslant P\leqslant 1$$

（2）必然事件u的概率等于1，则：

$$P(u)=1$$

（3）不可能事件V的概率等于零，则；

$$P(V)=0$$

从上述频率与概率的概念可以看出，它们之间有着紧密的联系，可以证明，在相当广泛的条件下，有$\lim\limits_{n\to\infty} f(A)=P(A)$，即随机试验的频率在其概率值的附近摆动，并随着试验次数的增多而趋近其概率值。这是随机试验存在着内部规律性的表现，据此我们才可以在n充分大时，取其频率作为概率的近似值。在许多实际问题中，当概率不容易知道时，我们往往就是这样做的。

这里必须指出，频率和概率是两个不同的概念：频率是对事件出现后的分布的一种计量，概率则是对事件发生的可能性的一种计量。

四、概率的运算法则

（一）概率加法法则

如果事件A与事件B（或更多的事件）互不相容，即事件A与B不能同时都发生，则至少发生其中任一事件A或B的概率为全部事件的概率之和：

$$P(A+B)=P(A)+P(B)$$

上述表达了概率可进行加和的重要特征，从概率的定义看，也不难证明公式的正确性。设在 n 次重复试验中，事件 A 发生了 m_A 次，事件 B 发生了 m_B 次，由于 A 与 B 互不相容，故事件 $A+B$ 发生了 m_A+m_B 次，而事件 A 的 $P(A)\approx\frac{m_A}{n}$，事件 B 的 $P(B)\approx\frac{m_B}{n}$，且 $\frac{m_A}{n}+\frac{m_B}{n}=\frac{m_A+m_B}{n}$，所以 $\frac{m_A+m_B}{n}\approx P(A)+P(B)$；但是，$\frac{m_A+m_B}{n}$ 就是事件 $A+B$ 发生的频率 $f(A+B)$，当 n 足够地大时，$\frac{m_A+m_B}{n}$ 与 $(A+B)$ 就很接近，因此就可得出 $P(A+B)=P(A)+P(B)$。下面通过具体例子说明公式的应用。

例如：某工厂的产品分为一级品、二级品、三级品三档，在正常生产的条件下，出现二级品的概率是 7 %，出现三级品的概率是 3 %，其余都是一级品，求出现非一级品的概率。

假如一只一只地抽验 100 只产品，根据概率的频率含义，大约将抽到 7 只二级品，显然，一只产品不可能同时既是二级品，又是三级品，这样，在 100 只产品中大约还会抽到 3 只三级品。因此在 100 只产品中，大约就有10只是非一级品，根据这样的分析，抽到非一级品的概率应该是10%。

解：记抽验一只产品是二级品为事件 A，记抽验一只产品是三级品为事件 B，抽验一只产品是非一级品就是事件 $A+B$，由于事件 A 和 B 是互不相容事件，所以：

$$P(A+B)=P(A)+P(B)=7\%+3\%=10\%$$

应用上述运算公式要特别注意必须有“事件A与B互不相容”这一条件，例如上例中的“一只产品不可能同时既是二级品又是三级品”这一事实，否则就容易导致错误。同时，依据上述运算公式不难推广出有n个不相容事件的情形。即有n个事件A_1、A_2、A_3、……A_n为互不相容，则：

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

（二）概率乘法法则

事件A和事件B同时出现的概率，等于事件A的概率乘以事件B在给定A的条件下的概率；同样也等于事件B的概率乘以事件A在给定B的条件下的概率，即：

$$P(A\cdot B)=P(A)\cdot P(B|A)=P(B)\cdot P(A|B)$$

概率相乘原则要求事件相互独立，所谓“独立”就是A、B两事件互相不影响。非独立事件之间不能用乘法法则进行计算。公式中的$P(B|A)$称为事件A发生的条件下事件B也发生的条件概率。

例题：一盒子中装有10支晶体管，4支是次品，6支是正品。在其中取2次，每次任取1支，作不放回抽样，问两次都拿到正品管子的概率是多大？

解：设事件A为“第一次拿到的是正品管子”；事件B为“第二次拿到的也是正品管子”。依题意可得：

$$P(A)=\frac{6}{10}，则\ P(B|A)=\frac{5}{9}，$$

$$\therefore\quad P(AB)=P(A)\cdot P(B|A)=\frac{6}{10}\times\frac{5}{9}=\frac{1}{3}。$$

即两次都拿到正品管子的概率为$\frac{1}{3}$。

从上例看出，如果A和B均为小概率事件，则它们同时

发生的概率要小得多。

显然，当事件A_1、A_2、A_3……A_n相互独立时有：

$$P(A_1A_2A_3\cdots\cdots A_n)=P(A_1)P(A_2)P(A_3)\cdots\cdots P(A_n)$$

怎样判断一些事件是否相互独立呢？很多情况下，根据对事件本质的分析就可以知道，不一定都要进行复杂的计算。

§ 1-3 随机变量及其分布密度

随机变量是不能事先准确地预言其结果的，在相同条件下，可以进行重复试验，试验的结果可以用数字来表示，因此，随机试验的结果是数量化的变量，我们称为随机变量。例如，掷抛均匀的硬币出现“正面”用实数1表示，出现“反面”用实数2表示等。它具有两个特征：①在一组观测值中随机变量取哪一个值是无法确定的；②在许多次观测值中，随机变量是遵照一定规律而出现的。用什么来描述这个特征呢？这就需要引入分布密度的概念。

由于任何一个随机变量都有它的统计分布规律，而这种统计分布规律又因随机试验的重复次数的不断增加，最终表现为随机变量的取值在一定范围内具有波动性，我们称这种波动性为随机变量的分布密度。

例如我们研究某一大学的电话总机呼叫次数的规律，在上午8点以前，呼叫的电话次数较少，8点以后，大家都上班了，呼叫次数急剧增加，过了一段时间，呼叫次数渐渐减少了。到了12点以后，大家下班了，呼叫次数减少到最低

数。下午两点以后，又都上班了，呼叫次数又急剧增加，过一段时间以后，呼叫次数又逐渐减少了。假使6点下班，6点以后，呼叫次数又减少到最低数。由此可看出，此总机呼叫次数不是均匀分布的。在单位时间内，时而多时而少，是有波浪起伏的。单位时间内呼叫次数的分布密度，可以通过长期统计出分布规律的。我们可以用某一个 $f(x)$ 函数来表示这种规律。因此，我们说随机变量的分布密度是具有一定规律的。

在实际问题中广泛存在着随机变量。比如，在工业生产中，随机地取一件产品，问它的质量指标（强度、硬度、光洁度。粘合力、纤度等等）是多少？这些指标就可以看作是一个随机变量，它们都有自己的分布规律。

我们用 x 表示随机变量。要想弄清一个随机变量的内部规律，就必须知道分布密度，至少也要知道它的数学特征（主要的是数学期望及方差）。当我们研究的随机变量不多或知之甚少的时候，确定一个随机变量的分布密度是很困难的。为了研究方便，常把随机变量分为两种：一种是离散型的随机变量，另一种是连续型的随机变量。

（一）离散型随机变量

这种类型的随机变量 x，可能取的值能够一一列举出来，在数轴上只能取有限个或一联串孤立值。例如：一位士兵进行实弹射击，每次射中的环数是一个随机变量，显然这个随机变量可取值为0、1、2、……10，但每次射击前随机变量取什么值是不确定的。然而，在大量射击中此随机变量的各个可能值的概率是可以确定的。下表是某士兵在大量射击中的结果：