（京）新登字 039 号

图书在版编目（CIP）数据

分析化学手册．第十分册，化学计量学/梁逸曾，俞汝勤主编．—2 版．—北京：化学工业出版社，2000.12
ISBN 7-5025-3111-4

Ⅰ．分… Ⅱ．①梁… ②俞… Ⅲ．①分析化学-手册 ②化学计量学-手册 Ⅳ．065-62

中国版本图书馆 CIP 数据核字(2000)第 86044 号

分 析 化 学 手 册
（第二版）
第 十 分 册
化 学 计 量 学
梁逸曾 俞汝勤 主编
责任编辑：任惠敏
责任校对：洪雅姝
封面设计：于 兵
*
化学工业出版社出版发行
（北京市朝阳区惠新里 3 号 邮政编码 100029）
发行电话：(010) 64918013
http://www.cip.com.cn
*
新华书店北京发行所经销
北京市燕山印刷厂印刷
北京市燕山印刷厂装订
开本 787×1092 毫米 1/16 印张 47¾ 字数 1190 千字
2000 年 12 月第 2 版 2000 年 12 月北京第 1 次印刷
印 数：1—3000
ISBN 7-5025-3111-4/TQ·1343
定 价：110.00 元

第二版前言

分析化学是人们获得物质化学组成和结构信息的科学。由于多学科的交叉渗透，现代分析化学已发展成为一个庞大的学科体系，建立起了比较成熟的多种分析方法，包括色谱分析、电化学分析、光谱分析、波谱分析、质谱分析、化学分析、热分析、放射分析、生化分析等。它一方面在科学研究中起着至关重要的作用，极大地推动着其他学科的发展；另一方面还直接服务于国民经济和生产建设的需要。同时，当代科学技术和人类生产活动的飞速发展也向分析化学学科提出了严峻的挑战，并带来了前所未有的发展机会。

我国的分析化学学科在新中国建立以来，特别是改革开放以后，取得了长足的发展。到目前为止，在全国范围内已形成了一支以中国科学院和高等院校及各部委研究所为核心的分析化学科研队伍，和一个涉及生物、环境、材料、临床、医药、地质、冶金、石化、宇航、商检、法医、侦破和考古等领域的庞大分析检验队伍，共同构成了我国分析化学学科研究发展的源泉和推广应用的基地。在多年的发展过程中，无论是分析化学的基础理论，还是实际应用方面，都已形成了丰富的知识和经验的积累，需要进一步的总结和推广。

《分析化学手册》是一部比较全面的反映现代分析技术，供化学工作者使用的专业工具套书。手册第一版自1979年出版以来，在读者中形成了一定的影响，已成为许多分析化验室的必备图书。但由于受组稿时的历史条件所限，加上近20年来是世界和我国的科学技术，包括分析化学学科飞速发展的时期，原手册第一版在内容和编排上已不能全面反映当前我国分析化学的发展现状。因此，根据广大读者的要求，我们组织了这套《分析化学手册》的修订工作。

在第一版原有6个分册的基础上，这次经扩充和修订为以下10个分册：

第一分册　基础知识与安全知识
第二分册　化学分析
第三分册　光谱分析
第四分册　电分析化学
第五分册　气相色谱分析
第六分册　液相色谱分析
第七分册　核磁共振波谱分析
第八分册　热分析
第九分册　质谱分析
第十分册　化学计量学

其中第一分册为基础内容，收集了分析工作中常用的基础数据、分析实验室

的安全知识及分析数据的常规处理、计算机应用的基础知识。第十分册所涉及的化学计量学是近些年来发展非常迅速的化学学科的一个分支，与分析化学有着特殊密切的关系，它应用数学和统计学的方法，并引入计算机科学的发展成果，其研究对象几乎涉及分析化学的所有过程，对于设计或选择最优的分析方法，解析大量的化学分析数据以最大限度地获取化学信息等具有普遍的指导意义，因此修订时增加这一部分内容。其他各分册均是按分析方法及所采用的主要仪器类型来划分，大体包括两方面的内容：基础原理、基础数据部分和实际应用部分。

本次修订，在内容上我们着重收录了基础性的理论和发展较为成熟的方法及应用，注意推陈出新，更新有关数据，增补各自领域近些年的新发展新成果，特别是计算机应用、多种分析手段联用技术的发展，以及分析技术应用于生命科学等的内容。

在编排方式上，进一步突出了手册的可查性。各册均编排主题词索引，与目录相互补充。手册中所涉及的名词术语统一采用国家自然科学名词审定委员会发布的标准，计量单位参照国家标准《GB 3100～3102—93·量和单位》的有关规定贯彻执行。其他凡有国家标准的也一律采用相关最新标准。

第二版的重编修订工作得到了我国分析化学界的大力支持，包括 11 位中国科学院院士在内的近 30 位知名专家、学者应邀担任了手册修订的编委会成员，全套书的修订出版凝聚着他们大量的心血和期望，在此谨向他们，以及在编写过程中曾给予我们热情支持与帮助的有关院校、科研单位及厂矿企业的专家和同行们，致以衷心的感谢。同时我们也真诚地期待着广大读者的热情关注和批评指正。

《分析化学手册》编委会

1996 年 6 月

第一篇　化学计量学方法

第一章　现代分析化学与分析信息理论和方法

第一节　现代分析化学——化学量测与化学信息的新兴学科

分析化学学科正经历着巨大的变革[1~4]。由于近年来物理学和电子学的发展，各种新型分析仪器相继问世，昔日的以化学分析为主的经典分析化学已发展成为一门包括众多仪器分析（色谱分析，电化学分析，光化学分析，波谱分析，质谱分析，热分析，放射分析，表面分析，等）为主的现代分析化学。正因为分析手段的不断扩展，广大分析化学家们感到以“溶液平衡”为基础的经典分析化学已不能代表现代分析化学学科发展的全貌，致使 Leihaisky 提出的“不管你喜欢不喜欢，化学正在走出分析化学”的论点广为流传。分析化学界出现了“化学正在走出分析化学”，分析化学应重名为“分析物理”，“分析科学”的议论。1985 年 11 月和1989 年 10 月在维也纳分别召开了第一次和第二次“国际分析化学的哲学和历史会议”，探讨分析化学的某些基本哲学问题。这说明分析化学学科正处在一个急剧分化的发展时期。美国《分析化学》的主编 Murray 在题为“化学量测科学”一文中指出：“用拓展的眼光来看待今天的分析化学是有益和有帮助的，它的发展已使之成为一门创造和应用新概念、新原理和仪器的策略来测量化学体系及其组分的学科，简言之，分析化学已成为一门化学量测科学。”[5]如果我们遵循着“分析化学是一门化学量测科学”的思路，就可以发现，分析化学学科当今的变革不是“化学正在走出分析化学”，而是“物理和新仪器正在走进分析化学”，它使分析化学家手中拥有更多的化学量测工具和手段，为分析化学家解决各学科发展所需解决的复杂的分析难题提供了更有力的武器。如何更有效地使用和发展这些新的分析手段和工具，怎样有效地从这些新的化学量测工具和手段中获取化学家们所需的有关的化学组成和结构信息、以及其他各种有用的化学信息，当是目前分析化学家急需解决的一个新问题。

首先，让我们来考察一下化学量测的全过程。化学量测的全过程似应从选择分析方法和采样开始，经化学量测的试验设计、量测过程的正确控制和优化、分析仪器所得信号的预处理、各类分析仪器数据定性定量分析，再到分析结果的统计推断、分析过程机理研究、所得纯组分波谱图的结构解析等，直至有用决策信息的提取。由此看来，化学量测过程是一个很复杂而且内涵极其丰富的过程，它每一步的有效完成实际都包括了相当丰富的内容，需要有很多关于化学、数学和物理的基础知识。如果说经典的分析化学主要是以“溶液平衡”为基础的话，那么，现代分析化学则是一门包括如何有效地进行各种化学试样的处理（包括分离等化学基础）、各种有关分析仪器所需的物理知识（物理基础）和化学知识（如各种波谱解析）、以及怎样进行最优采样、实验设计或选择最优化学量测方法，并通过解析化学量测数据以最大限度地获取化学及其相关信息（数学基础）的一门综合性极强的化学分支学科。

进行化学量测的基本目的是获取有关物质系统的化学成分与结构方面定性与定量的信息。凯塞尔（Kaiser）在有关分析方法基础的专著中[6]，界定分析化学学科的内涵为取得所研

究的物质组成的知识的有计划的信息过程。卡特曼（Kateman）等[7]从三个方面阐述分析化学的任务：利用已有的分析方法，提供关于物质化学成分的信息——日常例行分析工作；研究利用不同学科的原理、方法取得有关物质系统的相关化学信息的过程——分析化学的科学研究工作；研究利用现有分析方法取得关于物质系统的信息的策略——分析实验的组织工作。柯瓦尔斯基（Kowalski）在一篇题为“分析化学作为信息科学”的论文中[8]，指出分析化学发展史正经历极为重要的时期。这个时期对分析化学学科的重要性，是与现代科学与社会发生的一系列重大变化相关联的。这种重大变化首先源于计算机科学与信息科学的发展。该文援引美国科学基金会的报告，认为信息已成为美国社会的极重要组成部分，约半数的劳动力从事与信息相关的工作，挣取一半以上的劳动收入。在这样一个信息具有如此重要地位的社会中，分析化学起什么作用？该文作者认为，分析化学现在是，而且过去也一直是一门信息科学。在化学的各个分支学科中，分析化学担负的任务与其他分支学科的不同处，就在于分析化学的研究对象，不是直接提供某种具体的例如无机和有机材料，而是提供与这些材料的化学成分和结构相关的信息，研究获取这些信息的最优方法与策略。当然，分析化学工作者是与其他化学工作者分工而又合作，共同去完成生产与科研向化学提出的使命的。不但分析化学与无机化学、有机化学等传统的化学分支学科关系如此，分析化学与一些新兴的边缘学科如环境化学的关系亦是如此。例如，在IUPAC30届学术大会上关于“环境的挑战”的学术讨论中，出现这样的学术报告题目：“分析数据的生物学意义——所有环境课题都源于分析化学家！”[9]。将分析化学认作通过化学量测取得化学信息的科学，并不是说分析化学发展到今天才具有这种性质，也并不完全是由于信息对当代社会的重要性，人们才有意强调这一事实。然而，在分析化学得到飞速发展的今天，重新认识分析化学作为通过化学量测来提供化学信息的科学这一性质，反映了分析化学的新发展，而且，这一新发展可能还是质的飞跃。那就是，分析化学工作者已不仅是单纯的分析数据的提供者，而是解决实际问题有用的化学信息的提供者。

值得提出的是，近年来随着计算机，特别是微型计算机大量进入化学特别是分析化学实验室，一门新的化学分支学科——化学计量学应运而生，为有效进行化学量测和提供化学信息开辟了新通路，为分析化学提供了新机遇。化学计量学是一门交叉学科，它应用数学、统计学与计算机科学的工具和手段及其最新成果来设计或选择最优化学量测方法，并通过解析化学量测数据以最大限度地获取化学及其相关信息，自然，它首先就在分析化学中得到了普遍认同。自20世纪70年代中期诞生以来，在20世纪80年代得到长足发展，至今已逐步日趋成熟。图1-1将化学计量学与分析化学作为一门化学量测和化学信息学科之间的关系以图示方法表示出来了。

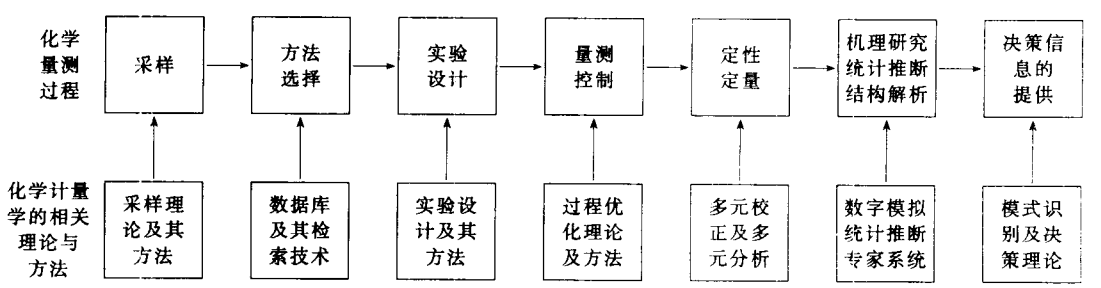

图1-1　化学量测流程与化学计量学方法的关系示意图

从此图我们可以看出，对于化学量测的每一步，化学计量学都有相应的理论和方法学，研

究如何来使化学量测和数据解析过程变得更有效，实质上，化学计量学就是一门关于化学量测的理论基础和方法学的化学分支学科[10]。分析化学与化学计量学有着如此密切的关系，就不难理解为什么美国化学会的分析化学杂志——“Analytical Chemistry”要将化学计量学作为一个分支领域，从1980年起就每两年对它的发展进行一次总结性综述了。我国分析化学家高鸿先生在十几年前就预言过：“数学在分析化学中的应用日益重要，如果说20世纪60年代是分析化学与电子学结合的时代，70年代是分析化学与电子计算机结合的年代，80年代很可能是分析化学与统计学和应用数学紧密结合的时代”[11]。从这一角度来看，化学计量学已经成为分析化学学科的必要数学基础。我们将化学计量学中从分析化学的信息理论基础、化学采样理论及方法、化学量测过程的试验设计与优化、化学量测数据的多元校正和多元分辨的定性定量解析，一直到分析过程机理研究、数值模拟和计算机波谱解析、有用决策信息的提取，其中包括化学模式识别、化学构效关系研究直至人工智能与化学专家系统的所有内容汇成一书，来作为分析化学手册的第十分册，其初衷也在于此。

分析化学作为一门化学量测和化学信息科学，对其量测过程有效性及效率的估计和评价就显得十分重要了。分析信息理论从信息理论的角度来研究化学量测过程。如果我们将通讯处理信息的过程与化学分析中的化学量测过程来进行比较，就可容易发现这两个过程十分相似（参见图1-2）[12]。

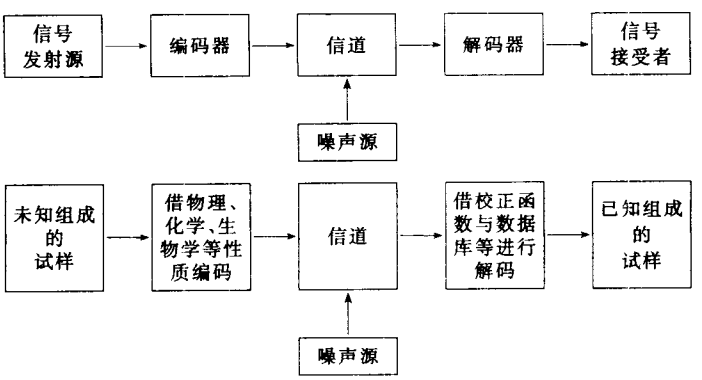

图1-2　通讯信息处理过程与化学量测过程的比较

进行化学量测的目的就在于消除或减少被量测的化学系统在化学成分、结构及其他相关信息的“不确定度”。分析信息理论和方法可提供相应的概念和方法来定量表征化学量测系统的“不确定度”及化学量测过程中体系之“不确定度”的消除或减少的定量度量（或称为化学量测过程的信息量之获取）。它不但可以对不同的分析过程，如分离、定性鉴定、定量测定等进行信息理论分析，还可对现代分析化学中各类复杂分析仪器的提供信息之能力进行合理评价，为分析化学量测提供了一套选择不同分析过程或不同分析仪器的理论基础和具体定量方法。本章将系统地介绍这些概念和方法。

第二节　熵——化学量测的“不确定度”的定量度量

一、分析试验与不确定度

分析工作的目的是取得有关未知试样的化学成分与结构的相关信息。所以，在进行分析测试之前，必存在某种不确定性或称“不确定度”。设有一份试样，其中可能含有 k 种离子中

的某一种，定性分析的任务就是确定这一离子确切是何种离子。在分析测试之前，存在有 k 种可能性，或称分析实验有 k 个可能的结局。如将上述情况用数学式表示，并设上述 k 种可能结局为 a_1，a_2，…，a_k；发生这 k 种可能结局的概率为 p_1，p_2，…，p_k，则有

$$A=\begin{matrix}a_1,\ a_2,\ \cdots,\ a_k\\ p_1,\ p_2,\ \cdots,\ p_k\end{matrix}$$

在此 A 表示发生上述定性分析实验的一个事件。如果存在有两个不同定性分析实验的事件，分别记为 A_1 和 A_2，并由以下的形式表示出的两个数表表示出，即

$$A_1=\begin{matrix}a_1,a_2,\cdots,a_k\\ 1,\ 0,\cdots,0\end{matrix};\qquad A_2=\begin{matrix}a_1,\ a_2,\ \cdots,\ a_k\\ 1/k,1/k,\cdots,1/k\end{matrix}$$

很明显，因为对于事件 A_1，它发生的可能性实际上只有一种，即只可能发生 a_1，而发生 a_2，…，a_k 的概率都等于零，所以，此事件的“不确定度”实际上不存在，如果存在这样一种“不确定度”的定量度量标准，对于事件 A_1，其值应为零；然而，对于事件 A_2，它发生的可能性就有 k 种，即发生 a_1，a_2，…，a_k 的概率都相等，故此事件的“不确定度”很大，如存在一种“不确定度”的度量标准，那么对于事件 A_2，其值应该很大或至少要大于零。以下我们将要讨论的熵的概念，就是这样“不确定度”的一种定量度量。

二、不确定度与仙农熵

对于一个随机事件的数表，仙农熵的定义如下

$$H=-\sum_{i=1}^{k}p_i\log p_i \tag{1-1}$$

仙农给出了上述定义并称 H 为熵。lg 一般表示以 e 为底，此时所得熵的单位为奈特（nat）；如果以 10 为底，其单位为的特（dit）；而以 2 为底时，其单位则为比特（bit）。从式（1-1）可知，如对前述的两个实例，用式（1-1）计算可得

$$H(A_1)=-1\log 1=0$$

$$\begin{aligned}H(A_2)&=-\sum(1/k)\log(1/k)\\&=-\lg(1/k)=\log(k)\end{aligned}$$

从上述结果明显可知，事件 A_2 的不确定度大于事件 A_1 的不确定度。

三、仙农熵的性质

（1）非负性，即

$$H(p_1,p_2,\cdots,p_k)\geqslant 0$$

这是因为 $p_i(i=1,2,\cdots,k)$ 为概率，故有 $1\geqslant p_i\geqslant 0$，所以 $\lg(p_i)$ 不可能取正值，即 $(-\sum p_i\log p_i)$ 才取正值。

（2）对于确定性事件，其熵为零。即

$$H(1,0,\cdots,0)=0$$

所谓确定性事件，即为该事件只有一种可能结果，如用数学语言表达则为，对于发生某一可能结果的概率为 1，而发生其他可能结果的概率都为零。前面讨论的事件 A_1 就是确定性事件的一个例子。注意到在此实际上是引入了一个人为的假设，即

$$\lim_{p\to 0}(-p\log p)=0 \tag{1-2}$$

式中，p 为一表示概率的实数。

引入这一假设是因为 log (0) 在数学上为一无意义的数。可是，如从极限的角度来看，引入 (1-2) 式在数学上是完全合理的。如假设 $p=1/e^n$，此时

$$-\log p=n$$

而

$$-p\log p=n/e^n\to 0 \tag{1-3}$$

随着 n 的增大，n/e^n 将很快接近于零。实因此时 p 是一个比 $(-\log p)$ 更高阶的无穷小。所以上式实际上是用极限的概念来避免对 log(0)的直接计算。有了式 (1-3) 的定义，所以有

$$H(1,0,\cdots,0)=-1\log(1)-\sum 0\log 0=-1\log(1)=0$$

(3) 对于等概率结果的事件，其熵最大。即

$$H(p_1,p_2,\cdots,p_k)\leqslant H(1/k,1/k,\cdots,1/k)=\log(k) \tag{1-4}$$

所谓等概率结果的事件，即该事件有 k 种可能结果，而且，发生这 k 种结果的概率都相等，即发生某一可能结果的概率都为$(1/k)$。由此性质可知，等概率结果的事件的不确定度最大。

四、条件熵与可疑度

今考察两个随机试验 A 与 B，其可能结果为

$$A:\ A_1,\ A_2,\ \cdots,\ A_n$$

$$B:\ B_1,\ B_2,\ \cdots,\ B_m$$

则两事件同时发生时的熵为

$$H_{n,m}(AB)=-\sum_{i=1}^{n}\sum_{j=1}^{m}p(A_iB_j)\log[p(A_iB_j)] \tag{1-5}$$

条件熵的定义：如果试验 A 出现结果 A_k 时，试验 B 在此条件下的熵为

$$H_m(B|A_k)=-\sum p(B_i|A_k)\log[p(B_i|A_k)]$$

相对应也可定义试验 B 出现结果 B_k 时，试验 A 在此条件下的熵

$$H_n(A|B_k)=-\sum p(A_i|B_k)\log[p(A_i|B_k)]$$

更进一步，还可定义在进行试验 A 的前提下试验 B 的条件熵

$$\begin{aligned}H(B|A)&=\sum p(A_k)H_m(B|A_k)\\&=\sum p(A_k)[-\sum p(B_i|A_k)\log[p(B_i|A_k)]\\&=-\sum\sum p(A_k)p(B_i|A_k)\log[p(B_i|A_k)]\end{aligned}$$

或在进行试验 B 的前提下试验 A 的条件熵

$$H(A|B)=-\sum\sum p(B_k)p(A_i|B_k)\log[p(A_i|B_k)] \tag{1-6}$$

条件熵的性质：

① $H(A|B)=H_{n,m}(AB)-H_m(B)$

$H(B|A)=H_{n,m}(AB)-H_n(A)$

此性质可由积事件(即同时发生的事件)的概率公式(参阅第十一章)直接导出。

② $H(A|B)\leqslant H_n(A)$

$H(B|A)\leqslant H_m(B)$

此性质的证明可参阅文献 [12]。

③ 如果随机试验 A 与 B 相互独立，此时 $p(A|B)=p(A)$(参阅第十一章)，亦即分析实验 B 与待解决的分析课题 (试验 A) 毫不相干，当然做试验 B 将得不到试验 A 的任何信息，也无从减少关于试验 A 的“不确定度”。此时，有

$$H_{n,m}(AB)=H_n(A)+H_m(B) \tag{1-7}$$

五、可疑度、互信息与散度

可疑度的定义：从条件熵的定义可知，$H(A|B)$实际表述了在进行了试验 B 以后的试验 A 的“不确定度”，所以，$H(A|B)$ 又称为可疑度。这样定义的可疑度，实际上可以说是反映了分析仪器或方法提供信息的能力与解决给定分析课题的需要之间的差距的定量度量。

互信息的定义：

$$I(A;B)=H_n(A)-H_n(A|B)=H_m(B)-H_m(B|A)=I(B;A) \tag{1-8}$$

式中，$I(A;B)$ 称为 B 关于 A 的互信息，或 $I(B;A)$称为 A 关于 B 的互信息。同时，式(1-6) 也可称为信息守恒定律或信息平衡原理。

散度或卡尔贝克（Kullback）信息量的定义：

设有 p、q 两个概率分布，

验前分布 $p:(p_1,p_2,\cdots,p_n),\sum p_i=1$

验后分布 $q:(q_1,q_2,\cdots,q_n),\sum q_i=1$

当分布 p 被分布 q 取代时，散度或卡尔贝克信息量定义为

$$I(q\ //\ p)=\sum q_i\log(q_i/p_i) \tag{1-9}$$

类似地可给出连续变量的散度或卡尔贝克信息量的定义：

$$I(q\ //\ p)=\int q(x)\log(q(x)/p(x))\mathrm{d}x \tag{1-10}$$

第三节　定性分析的信息理论和方法

定性分析作为分析工作的重要组成部分，提供的是关于物质成分、结构特征方面的化学信息，回答的是“是什么?”这一问题。在这一节我们将就有关定性鉴定方法的信息定量评价方法、色谱及层析方法实验调优的信息理论和方法、质谱及红外光谱的编码与检索的信息理论和方法等方面来分别加以介绍。

一、不同定性分析鉴定方法的信息量估价

在定性分析鉴定方法的信息量估价中，一般可分为两种不同的方法，一种是针对一具体定性实验而言，以实验前的结果“不确定度”与实验后的结果“不确定度”之差别来估价此定性实验的信息量，亦即用前节所讨论的实验前与实验后的熵之差来估价实验的信息量，这样的信息量估价可包括以下几种情况：

（一）结构定性分析的信息量

关于结构分析的结果之信息量计算十分简单。结构分析过程可用图 1-3 所示的简单框图表出。

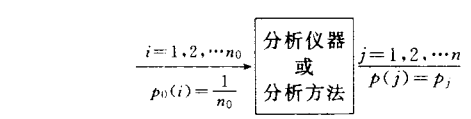

图 1-3　结构分析过程单框示意图

作为分析仪器或方法的输入，可以是 n_0 个等概率的可能化学结构。而分析仪器或分析方法的输出，可能给出 n 种尚不能分辨的结构。每种结构的验后概率一般是不相等的。结构分

析提供的信息量可按下式来进行计算。

$$I(A//B)=H(A)-H(B)$$

在此，A 表示试验前可能存在不同化学结构的事件；B 表示进行了分析仪器或分析方法试验后可能存在不同化学结构的事件；而 $I(A//B)$ 则表示在进行了分析仪器或分析方法的试验 B 后所获得的信息量。很明显，进行了分析仪器或分析方法的试验 B 后所获得的信息量应为 A、B 两事件的“不确定度”即熵之差[13]。

分析前：

$$p_0(i)=1/n_0 \qquad (i=1,2,\cdots,n_0)$$
$$H(p_0)=\ln(n_0)$$

分析后：

$$p(j)=p_j \qquad (j=1,2,\cdots,n)$$
$$H(p)=-\sum_{j=1}^{n}p_j\ln(p_j)$$
$$I(A//B)=H(A)-H(B)$$
$$=\ln(n_0)+\sum p_j\ln(p_j)$$

如假定实验后各可能结构均为等概率，则

$$I(A//B)=\ln(n_0/n) \tag{1-11}$$

注意到此时 $I(A//B)$ 与卡尔贝克（Kullback）信息量定义完全一致了。

（二）定性化学反应分析的信息量

设有一纯溶液，它可能是下述离子之一的试液：Ag^+，Pb^{2+}，Al^{3+}，Zn^{2+}，Na^+ 或 K^+。今用经典定性分析方法进行分离鉴定。设加入试剂盐酸，如何估算正反应（发生沉淀）及无反应时的信息量[14]？

正反应（发生沉淀）时，因可与盐酸产生白色沉淀的在这 6 种离子中只可能是 Ag^+ 和 Pb^{2+}，所以，经此反应后，可能离子的范围从 6 种变成了 2 种，根据式（1-11），可得其化学反应过程所得的信息量为

$$I(A//B)=H(A)-H(B)=\ln(6/2)=\ln 3$$

无反应时，因不与盐酸产生白色沉淀的在这 6 种离子中可能为 Al^{3+}，Zn^{2+}，Na^+ 或 K^+ 中任何一种，所以，经此反应后，可能离子的范围从 6 种变成了 4 种，根据式（1-11），其化学反应过程所得信息量为

$$I(A//B)=\ln(6/4)=\ln(1.5)$$

（三）测定物理常数鉴定有机化合物

测定熔点、沸点、折射率、密度等物理性质常用于有机化合物的鉴定。今试以纯物质的熔点测定为例，来考察某次物理常数测定所获得的信息量。设所测物质在分析之前就已知应是属于在温度 100°C 至 200°C 的一种物质，又已知在此温度范围之内可能存在 200 种有机化合物，如进而设它们将等概率分布于此温度区间，则在量测之前的不确定度，即熵为

$$H(A)=\ln(200)$$

经量测后知，其熔点为(100±1)°C，从已知在此温度范围之内可能存在 200 种有机化合物且它们将等概率分布于此温度区间，则测定后的熵为

$$H(B)=\ln[(200/100)\times 2]=\ln(4)$$

此实验的信息量为

$$H(A)-H(B)=\ln(200/4)=\ln(50)$$

二、仪器定性分析的信息量

如前所述，在定性分析的信息量估价中，一般可分为两种不同的方法，前一节已经讨论了针对某一具体定性实验信息量估价方法，亦即用实验前与实验后的熵之差来估价实验的信息量，在这一节里，将讨论另一种专门针对仪器定性分析的信息量估价方法，这类方法之主要思路是直接估价仪器信号的熵，亦即该仪器可获得的信息量。化学计量学方法就是通过采用化学实验的方法努力增大仪器的分析能给出的信息量或是采用编码的方式使得仪器分析的结果能尽量多地给出信息，以达到用信息量来作为目标函数，从而提高仪器定性分析的效率。

（一）薄层及纸上层析分离与定性鉴定

薄层色谱与纸色谱分离一般以 R_f 值表述，不同的化学物质具有不同的 R_f 值。如将薄层色谱分成 m 个间距，对落入每个间距的 R_f 值的个数 n_k 计数，则薄层及纸上层析分离与定性鉴定的信息量可由下式给出[15]：

$$I=-\sum_{k=1}^{m}(n_k/n)\ln(n_k/n) \tag{1-12}$$

此式是较早将信息理论引入薄板层析的马萨特（Massart）[15]所用的式子，它的物理意义是很直观的。如 n 个化合物均落在某个间距中，则此时 I 为零，未获得如何分离与信息。I 值只有在 n 个化合物均匀分布在 m 个间距时最大。此信息量可用作寻找最佳展开试剂的目标函数。

（二）色谱分离鉴定的信息量

在色谱分析中，一般是利用保留指数来进行定性分析，与薄层色谱和纸色谱的不同之处是，每一个化合物都由一个色谱峰来表示，存在着色谱峰相互重叠的问题。所以，色谱分离鉴定的信息量比起薄层色谱和纸色谱的信息量计算多了一项。仿薄层色谱的处理方法，将保留时间进行区间离散化，即划分为等长的 (Δy) m 段，统计一定数量的化合物（总数 n）的保留指数落入不同段的频数，则有

$$I=-\sum_{k=1}^{m}(n_k/n)\ln(n_k/n)+\ln(\Delta y)-\ln\sqrt{2\pi e\sigma_e^2} \tag{1-13}$$

式中，$\ln(\Delta y)$ 为一个常数，所以，对于不同的色谱柱，此项没有差别，一般可采用将 Δy 取为 1 而去掉。式中的第三项 $\ln\sqrt{2\pi e\sigma_e^2}$ 实际是来自前一节所定义的可疑度，在此表述了在进行了色谱分离以后的化合物靠保留指数来定性时的“不确定度”，也就是说是色谱峰的熵。其中 σ_e 为色谱峰的标准差。注意到在此我们是假设了每一种物质的色谱峰的标准差都是一样的。因为由式（1-6）定义的可疑度为

$$H(A|B)=-\sum\sum p(B_k)p(A_i|B_k)\log[p(A_i|B_k)]$$

在此，$p(B_k)$ 表示第 k 种化合物出现的概率，在此假设为等概率，即 $(1/n)$，$p(A_i|B_k)$ 为第 k 种化合物在色谱仪中的信号，一般假设为标准差为 σ_e 高斯色谱峰。对此积分所得结果就是 $-\ln\sqrt{2\pi e\sigma_e^2}$。详细讨论可参见文献 [12]。由式（1-12）定义的信息量标准，可用于选择不同流动相或色谱柱。

（三）质谱定性鉴定的信息原理

质谱是本世纪出现的分析方法。质谱仪的功能是产生带电离子，包括母离子和原分子的离子碎片，并按离子的质荷比（m/z）对化合物进行区分，“质谱”是不同离子数目的记录，每种离子的相对数目对每种化合物（包括同分异构体）将是特征的。质谱仪能提供关于有机化合物结构和固态试样元素分析的大量信息。所谓“化合物的质谱”，包含大量的离子碎片，且这些碎片离子的相对丰度时常超过母离子。分子碎片化的独特性有助于化合物的鉴别工作。质谱鉴定可用不同的方法进行。一种是解析法，即研究与假设破碎模式，并反过来从纯化合物质谱中的碎片离子来构思分子的结构。另一种办法是检索，即不管质谱图的含义，从已有的数据库中检索。本小节主要讨论信息理论在质谱检索中的应用。

利用质谱进行定性鉴定，我们先对其讯号能提供的信息量作一粗略估计。设用一低分辨率质谱仪，其质荷比区间仅 200 原子质量单位，如用信息论的观点就是 200 个信道。又设每一质量数位置我们仅区别有峰（其编码为 1）与无峰（其编码为 0），使用这种 0-1 编码时，则每个信道的最大可提供的信息量为 1bit（ld2，在此 ld 表示以 2 为底的对数），如果做到每个信道都相互不相关，这一低分辨率的质谱仪理论上可提供 200bit 的信息量，即大约能分辨 2^{200}种不同化合物，这个数目当然远远大于目前已知的有机化合物的个数。

现试从统计角度来探讨一下将谱图编码后，互相重复的可能性。设有 N 个谱图，按 0-1 编码，每个谱图可认为含 n 个“信道”，即原子离子质量单位数。在研究谱图 $\boldsymbol{x}$ 与 $\boldsymbol{y}$ 的重复问题时，可将每一谱图认作一个向量：

$$\boldsymbol{x}=(x_1,x_2,\cdots,x_n)$$

$$\boldsymbol{y}=(y_1,y_2,\cdots,y_n)$$

向量中的每一元素 x_i 或 y_i 的取值只有 0 或 1，即有峰时为 1，无峰时为 0。定义函数 $F(x_i,y_i)$，将编码重复的情况及不重复的情况可分别表述为

$$F(x_i,y_i)=0 \qquad \text{当 } x_i=y_i$$

$$F(x_i,y_i)=1 \qquad \text{当 } x_i\neq y_i$$

再定义函数 D，以表述所有信道中编码不重复数之和，

$$D=\sum_{i=1}^{n}F(x_i,y_i)$$

可以证明函数 D 的期望值 D 在信道 i（在质谱中即为某一质荷比数）出现 1 的概率，及信道 i 出现 0 的概率皆为 0.5 时可达最大[12]。

实际上，如从信息理论来考虑，同样可以得出上述结论。对于质谱的 0-1 编码，每个信道 i 的熵值可由式（1-14）表出：

$$H_i=-\sum_{k=1}^{2}p_{ik}\mathrm{ld}p_{ik}=1 \tag{1-14}$$

式中，p_{ik}（$k=1$，2）表示信道 i 在 N 张质谱图中出现 1（$k=1$ 时）或 0（$k=2$ 时）的概率，显然，根据等概率事件熵最大原理，只有当 $p_{ik}=0.5(k=1,2)$时，H_i 为最大。当质谱的 n 个信道都相互不相关时，此时对 N 个质谱的 0-1 编码的总信息量为

$$H=\sum_{i=1}^{n}H_i=\sum_{i=1}^{n}\left(-\sum_{k=1}^{2}p_{ik}\lg p_{ik}\right)=n$$

但实际编码由于各信道之间存在相关性，上述所得实际是 0-1 编码可以得到的极大值，将其记为 $H_{\max}$。定义实际可得信息量 H_n 与最大可得编码信息量 $H_{\max}$之比为编码效率(code efficien-

cy,CE),

$$CE\%=(H_n/H_{max})\times 100\%=(H_n/n)\times 100\%$$

在 20 世纪 70 年代初期关于信息原理用于质谱编码检索的早期论文中[16,17],研究了如何进行编码能提供最大的信息,而又尽可能节省计算机内存与缩短检索时间等问题。

在 0-1 编码的问题中,需确定在何种情况下应定义信道 i 具有峰即编码为 1。为此需确定一临界水平值(TL 值),强度大于(或等于)此值认作有峰,否则认作无峰。显然,TL 值如选择过低,则噪声信号亦可能被认作分析信号,在极端情况下,每一信道均出现 1,信息量将降至零;如 TL 值取得过高,则许多微弱强度的分析信号均被认作为零,亦将损失信息量,以基峰的 0.01%至 0.3%强度作 TL 值时,信息量是恒定的,当 TL 值上升到约 1%以上时,信息量下降。因取很低的 TL 值易受噪声的影响,折衷的方案是取 1%的 TL 值。当然,对不同的质量数取不同的 TL 值,或通过调整 TL 值使相应质量数的信息量 H_i 达到最佳值。

TL 值对信息量有一定影响,但借调整 TL 值增大信息量是有限的。p_{ik}值基本上是质谱峰的自然分布概率所决定的。有的质量数位置出现峰的概率很小,有的 p_{ik}值则较大。减少对计算机内存要求及增大编码效率的办法之一,是舍弃 p_{ik}差别很大的质量数位置,这对信息量影响不大,但能减少对内存的要求。更好的办法是在考虑到信道之间相关性的基础上,将某些信道合并,以尽量减少信道而使信息量减少不多。如已知质量数为 86 的 $p(86_1)$(出现有峰的概率)为 0.236,而质量数为 87 的 $p(87_1)$为 0.246,如将二信道合并,在此两信道信号不相关的条件下,可得 $p_{合,1}$为 0.48。然而,实际上并非如此。对数据库进行分析可知,如在质量数为 86 位置出现峰,则有 56%的情况在质量数为 87 处亦有峰,即条件概率 $p(87_1|86_1)=0.56$,可见两质量数位置(此两信道)是相关的。下就以此例来说明二信道合并时信息量的计算。注意到两信道合并后,所谓有峰(编码为 1)的情况包括以下三种情况:① 在质量数为 86 位置出现峰,且在质量数为 87 处亦有峰;② 在质量数为 86 位置出现峰,但在质量数为 87 处无峰;③ 在质量数为 86 位置无峰,但在质量数为 87 处有峰。所以,出现编码为 1 的概率应为

$$\begin{aligned}
p_{合,1}&=p(87_1,86_0)+p(87_0,86_1)+p(87_1,86_1)\\
&=p(87_1)p(86_0|87_1)+p(86_1)p(87_0|86_1)+p(86_1)p(87_1|86_1)\\
&=p(87_1)[1-p(86_1|87_1)]+p(86_1)[1-p(87_1|86_1)]+0.246\times 0.56\\
&=p(87_1)\{1-[p(86_1)p(87_1|86_1)]/p(87_1)\}+0.236[1-0.56]+0.246\times 0.56\\
&=p(87_1)-[p(86_1)p(87_1|86_1)]+0.236[1-0.56]+0.246\times 0.56\\
&=0.246+0.236-0.236\times 0.56\\
&=0.35<0.48
\end{aligned}$$

可见,上例合并二峰位置得不到出现 1 的概率 0.48。对质谱峰的研究表明,二质量数相差 1(1H)、2(2H)、13(CH)、14(CH_2)、15(CH_3)等数值时,相关关系最为显著。相关关系本身在编码中可考虑作为压缩维数以节约内存和缩短检索时间的依据。显然,将相关的质量数位置合并损失有用信息较少。

Wangen 等[17]对有关文献编辑的质谱图集[18]中 6652 条低分辨质谱图进行编码试验。这些质谱系在不同实验室用不同的电子轰击离子化仪器(30～100eV)测得,涉及约 5000 种不同的化合物,即不少化合物在此图集中存在不同实验室测得的互相重复的质谱。借此可校验编码的使用效果。作者比较了 0.1%、1.0%、5.0%三种 TL 值用于编码,约 5000 种化合物摄制的 6652 条质谱经编码后,相互重复的组数见表 1-1。由表可见,TL 值由 0.1 升至 1.0%,相似异构体或相似化合物给出重复的质谱编码的情况并无显著变化,1.0%是较佳的 TL 值。

表 1-1 编码质谱的相互重复组数

TL 值/%	重复组数	涉及谱图数	结构重复的化合物特征		
			同一化合物	相似异构体	相似结构及其他
0.1	167	370	112	42	14
1.0	209	451	156	38	15
5.0	376	836	335	98	43

表 1-1 涉及 352 个质量位置。按上述合并质量数位置的办法，将原 352 个质量数合并至 80 个，并取 TL 值为 1%，在表 1-1 的基础上增加的重复编码数见表 1-2，如利用相关关系进一步合并一些质量数位置，使信道数压缩至 48，进一步增加的重复编码质谱图数目亦列于表 1-2。从表 1-2 可以看出，经压缩维数后信道的编码效率 CE 值增高。80 信道的最大熵值为 80bit，实际提供的信息量为 76.5bit，故编码效率为 95.6%，较使用 352 维信道时，虽绝对信息量由 131.7bit 降至 76.5bit，编码效率则较原 37.4%提高甚多，即能更有效地利用有限的计算机内存。

表 1-2 压缩编码维数对检索的影响

维　数		80	48
较维数为 352 时增加的重复质谱图数		51	230
增加的重复谱图涉及的化合物特征	同一化合物	22	76
	类似异构体	13	115
	相近结构及其他	16	60
新增重复编码所涉及的谱图数		115	555
总信息量/bit		76.5	45.8
编码效率(CE)		95.6	95.4
与 352 维编码比较减少的熵值/bit		55.2	85.9

上述讨论的 0-1 编码是最简单的编码。实际上，在某一质量数位置有峰时，峰的强度还包含有化学信息。设每一信道的峰强度可分成 m 个阶梯或水平，在编码时应将这部分信息亦加以考虑。令 p_k 为 i 信道的峰落入 k 水平的概率，则在各信道是相互独立的情况下，有

$$H_n = \sum_{i=1}^{n} H_i$$

其中

$$H_i = -\sum_{k=1}^{m} p_{ik} \lg p_{ik}$$

（四）红外光谱定性鉴定的信息原理

上节讨论的关于质谱编码的信息原理，对各种电磁波谱分析原则上均是适用的。各种电磁波谱中，红外光谱是较常用的有机化合物波谱鉴定手段。本节将以红外光谱为例进行简要讨论，其他电磁波谱分析可举一反三类推，如将前节中的质量数位置改为红外光谱的波数或波长小区间，可对红外吸收光谱进行编码，凡吸收等于或超过某一约定的临界水平 TL 值者均编码为 1，否则为 0，即得 0-1 编码，如再将吸收峰带强度划分为若干水平，则还可进一步得到更多的化学信息。

有关文献[19]按上述方法对 ASTM 红外光谱索引（ASTM Infrared Spectral Index）中的红

外光谱数据库进行编码，所有文件包括了约 102000 条红外光谱（按 Wy-andotte-ASTM 法编码[20]是将 2.0～15.9μm 波长区域划分为宽度为 0.1μm 的 140 波长段，在每一波长段，如吸收强度超出选定的 TL 值，则编码为 1，否则为 0）。由 ASTM 文件编码 96900 条红外光谱（WL-1）。如无误差及相关关系的影响存在，WL-1 文件的信息量在选取 100 个峰时约 50bit。实际由于误差及相关关系的影响，只能达到 20bit 左右。当然，20bit 信息量仍足以区别 10^8 个化合物，故其信息功能仍是很强的。红外光谱与质谱有不同处，因波长区段或波数区段的划分与质量数相比，有任意性。为了说明波长（或波数）区段划分的宽度对信息量的影响，取 WL-1 文件的一个子集 WLS-1（含 5100 条光谱）进行扩充谱带“窗”的试验。按 0.1μm“窗”宽度编码后，对含有峰（编码为 1）的波长段作如下处理：如某一波长区段含有峰（1），则将其相邻的二波长段亦赋予“1”，即得“窗”宽为 0.3μm 的编码（记为 WLS-1-3）。进而将 WLS-1-3 文件再按相同办法扩“窗”，可得“窗”宽为 0.5μm 文件（WLS-1-5）。这一扩窗步骤示于表 1-3，对原按 0.1μm 波长区段编码的文件进行上述处理后，对信息量将发生影响。这种影响与三个因素有关：① 扩“窗”后由于许多波长区段的“1”增加，即概率 p_1 值增加，信息量一般上升。当然，p_1 值增至超过 0.5 以后信息量又下降。这个因素可称为概率效应。② 误差效应。误差将使信息量降低。随着“窗”扩大这一影响的相对值将减弱。③ 相关效应。扩“窗”肯定导致相关关系的增加。例如 j 位置有峰，$j-1$ 与 $j+1$ 亦赋值“1”。这导致信息量降低。表 1-4 例举将 WLS-1 的窗扩至 0.3μm（WLS-1-3）及 0.5μm（WLS-1-5）信息量较 WLS-1 发生的变化。“+”代表信息量增加，“−”代表减少。对 WLS-1 本身而言，亦存在误差与相关关系对信息量的影响，在表中亦列出以资比较。总之，将光谱编码的“窗”适当扩充，例如由 0.1μm 扩至 0.3μm，对增加信息量是有利的。

表 1-3　红外光谱编码的扩“窗”步骤示例

文　件	“窗”宽/μm	峰　位　置						
		$j-2$	$j-1$	j	$j+1$	$j+2$	$j+3$	$j+4$
WLS-1	0.1	0	0	1	0	1	0	0
WLS-1-3	0.3	0	1	1	1	1	1	0
WLS-1-5	0.5	1	1	1	1	1	1	1

表 1-4　红外光谱编码信息量（bit）与“窗”宽的关系

影　响　因　素	WLS-1	WLS-1-3	WLS-1-5
概率效应（+）	—	22.6	8.3
误差效应（−）	26.5	12.3	6.1
相关效应（−）	6.1	20.7	14.0
总效应		+8.0	+0.4

对于 WL-1 文件，当取全部 140 信道编码时，其编码效率为 15%。择优选出 93 信道的编码效率增至 23%。如无误差存在，140 信道的效率当为 41%，93 信道当为 51%。可见，误差效应对信息量有重大影响。扩“窗”能提高编码效率。例如，对优选的信道，当 WLS-1 的编码效率为 24%时，扩至 WLS-1-3 为 41%，WLS-1-5 为 46%。

在记录用于编码检索的红外光谱时，应固定实验操作条件，例如先将整个光谱区的基线调至透射率（T）＝（95±2）%，继将最强的谱带调整至 T＝（5±2）%（改变吸收池厚度），不时校正波长表度，对所用化合物进行严格提纯精制，峰强度取最高峰处的透射率与基线的差值：峰强度低于一定 TL 值时编码为 0，否则为 1。TL 值较高时，出现的“0”增加，使有

用的可供选用的信道数下降。文献[21]采用了3%、5%、10%三种TL进行比较，不同TL值选出的有用信道数及信息量列于表1-5和表1-6。表1-5是将波数（wave number，WN）由4000cm^{-1}按每25cm^{-1}为一段均分为140段（称WN-码），表1-6是将波长（wave length，WL）2.0μm起按每0.1μm波长段分为140段（称WL-码）。误差项的影响与TL值有关，在将最强带调整至5%时，若$T=1\%$的误差，则在近TL值处造成的误差当TL为3%时$T=0.25\%$，当TL为10%时$T=0.7\%$，作较坏估计，取TL值为3%时，由此造成的信息量下降为0.03bit。另一项可能的误差是记录峰值的实验误差，对WL-码，在长波长区这一误差将较显著，故WN-码的可靠性较高。光谱区间划分按WL-码在（3500～2500）cm^{-1}区域峰数目较少，这是由于在这一区域将两个或两个以上的峰编码为一个"1"，用WN-码则可免此弊，故WL-码的信息量一般很低。对较高波数区，则可能出现相反情况，但考虑到相关关系等因素，总的说来WN-码的信息量较大，故选择WN-码一般效果较佳。

表1-5 TL值对WN-码信息量（bit）的影响

数据集①	$TL=3\%$		$TL=5\%$		$TL=10\%$	
	信息量	信道数	信息量	信道数	信息量	信道数
CHS-WN	15.1	26	13.4	22	7.5	14
CHU-WN	31.5	47	29.7	42	24.9	35
CH-WN	32.8	58	18.8	19	25.0	47
ALC-WN	28.7	45	28.7	44	28.5	45
ETH-WN	17.5	21	15.0	17	22.2	34
CARB-WN	15.5	28	15.4	26	14.9	25
A-WN	41.3	73	39.9	65	36.5	61

① CHS——饱和烃；CHU——不饱和烃；CH——饱和烃及不饱和烃；ALC——醇；ETH——醚；CARB——醛/酮；A——以上全部谱图。

表1-6 TL值对WL-码信息量（bit）的影响

数据集	$TL=3\%$		$TL=5\%$		$TL=10\%$	
	信息量	信道数	信息量	信道数	信息量	信道数
CHS-WL	12.2	25	9.6	18	5.8	18
CHU-WL	26.3	15	25.5	50	19.9	34
CH-WL	28.2	57	26.1	50	16.9	32
ALC WL	24.8	48	24.5	48	24.1	48
ETH-WL	21.2	35	20.5	36	20.1	36
CARB-WL	—	—	13.9	21	14.1	23
A-WL	44.0	99	42.6	93	37.1	88

Heite等[22]研究了以相关系数作指标将峰位置按数值分类法分组，并以信息量作指标由每组选出熵值最大的"信道"，由此可得到相互较独立的信息量较大的红外峰位置，在此基础上再进一步重新编码。对5100个化合物组成的样本（ASTM红外光谱索引），在由140个光谱分段选出40个编码时，97.7%的编码是不重复的。Bink等[23]用主成分分析技术研究红外光谱的结构相关模式，用多元线性回归分析法分类，用可疑度作为分类时的优化指标。Ritter等[24]研究了互信息与最大似然分类器分类功能的关系。

第四节　定量分析的信息理论和方法

关于定量分析的信息理论基础，Eckschlager等曾经进行过较系统的研究，在"Collection Czechoslov. Chem. Communicationsz"杂志上发表过系列论文，并总结于其专著[25]中。在本节

中，将叙及这些研究的主要成果。

一、定量测定的信息量[25,26]

在进行定量测定之前，对被分析试样的成分的浓度范围往往并非完全一无所知。一般可假设待测组分 x 的含量在 $[x_1, x_2]$ 区间内，服从均匀分布，故其验前概率分布为

$$p_0 = 1/(x_2 - x_1)$$

在关于试样成分浓度范围确实一无所知的情况，则 $x_1 = 0\%$，$x_2 = 100\%$，上述假设仍成立。

在完成定量分析之后，分析结果一般服从正态分布，即 $x \sim N(\mu, \sigma^2)$。此处 μ 为试样中待测组分含量的真值，σ^2 为总体方差。则验后分布 $p(x)$ 是正态概率密度函数。用散度或卡尔贝克（Kullback）信息量的定义：

$$I(p \,//\, p_0) = \int_{x_1}^{x_2} p(x)\ln[p(x)/p_0(x)]\mathrm{d}x \tag{1-15}$$

为积分方便均取自然对数，故其单位为奈特（Nat）。对式（1-15）积分得

$$I(p \,//\, p_0) = \ln[(x_2 - x_1)/(\sigma\sqrt{2\pi e})] \tag{1-16}$$

实际上，在完成定量分析之后，分析结果一般用均值 x 表示，则此时可用学生分布来代替正态分布，当测定次数为 n，则分析结果的置信区间为

$$x \pm t_{\alpha,\varphi} \times (s/\sqrt{n})$$

在此，均分差 s（为 σ 的样本估计）可由下式求得

$$s = \sqrt{1/(n-1)\sum(x_i - x)^2}$$

$t_{\alpha,\varphi}$ 为学生分布的临界值，α 为置信率，$\varphi = n - 1$ 为自由度。近似地，我们可认为分析测定后的后验分布为

$$p'(x) = 1/[2t_{\alpha,\varphi}(s/\sqrt{n})]$$

由此求得的信息量为

$$I(p' \,//\, p_0) = \ln[(x_2 - x_1)\sqrt{n}]/(2st_{\alpha,\varphi}) \tag{1-17}$$

从上式可以看出，由学生分布求出的信息量是一个测定次数的函数，一般说来，测定次数的增加且精密度高（测定次数增加时，s 不增大）可增加定量分析的信量。

如果分析结果存在系统误差 δ，即

$$\delta = |x - \mu|$$

在这种情况下，信息增益由下式给出

$$I(p_2 \,//\, p_0) = I(p \,//\, p_0) - I(p_2 \,//\, p)$$

在此，$p_0(x)$ 为待测分析物的验前分布，$p(x)$ 的意义同前，是服从 $N(\mu, \sigma^2)$ 的正态概率密度函数，而 $p_2(x)$ 则是服从 $N(\mu + \delta, \sigma^2)$ 的正态概率密度函数，在这里 $I(p_2 \,//\, p)$ 代表了以 p_2 代替 p 后引起的信息量的降低。经运算可得

$$I(p_2 \,//\, p_0) = \ln[(x_2 - x_1)/(\sigma\sqrt{2\pi e})] - [(\delta/\sigma)^2/2] \tag{1-18}$$

显然，当 $\delta = 0$ 时，仍得式（1-10）计算的 $I(p \,//\, p_0)$ 值。由此可见，分析方法的准确度影响定量分析的信息量。

二、提高分析精密度与准确度的信息量

前节已述及分析方法的精密度与准确度对信息量的影响。我们再进一步考察，当用一精