

《光谱实验室》丛书之十一

数 理 统 计

在化学、光谱分析中的应用

〔苏〕纳利莫夫 原著

余 生等 编译



〔德〕高斯(1777-1855)

《光谱实验室》编辑部
数学是自然科学
的皇后。

——高斯

《光谱实验室》编辑部

《光谱实验室》丛书之十一

数理统计

在化学、光谱分析中的应用

[苏] 什利莫夫 原著

生等 编译

江苏工业学院图书馆
藏书章

1987

北京

内 容 提 要

本书根据“国际化学丛书”《The Application of Mathematical Statistics to Chemical Analysis》结合俄文原版编译而成。该书结合实际，由浅入深，通过具体例子比较全面和系统地阐述了数理统计在化学分析和光谱分析中的应用，是对分析工程师、技术人员、大学生、教师、研究生等的工作都有深远指导意义的基础读物。本书主要内容如下：1.数理统计所解决的问题；2.分析中误差的分类；3.随机变量及其特征；4.正态分布；5.泊松分布和二项分布；6.如何估计分析结果；7.方差分析；8.线性关系的统计；9.与统计实验设计有关的工作方法。书后还附录了十四个数理统计参数表、索引和有关数理统计的基本知识。

数理统计在化学、光谱分析中的应用

(学习班专用讲义)

[苏] 纳利莫夫 原著

余 生等 编译

编辑者：《光谱实验室》编辑部

(北京市学院南路76号)

发行者：冶金工业部钢铁研究总院

(北京市学院南路76号)

印刷者：北京市通县向阳印刷厂

(北京市通县小杨各庄)

字数：315000 1987年3月第1版 工本费：6.00元

编辑者的话

《数理统计在化学、光谱分析中的应用》一书是冶金部钢铁研究总院高级工程师余生等根据“国际化学丛书”翻译整理出来的。该书曾经科学出版社审查，同意出版，并排入了出版计划。后来，由于多种原因，该书的出版问题就拖了下来。粉碎“四人帮”后，我们去科学出版社联系，他们不愿管了。原因就是稍微“旧”了一点，因而出版价值不大。

我们认为，优秀著作是不能用它诞生的时间来否定或降低它在今天的价值的。文学作品是这样，自然科学著作也是这样。基础科学论著尤其如此。爱因斯坦的相对论是1905年发表的，至今已80年了，是不是过时了呢？看来，包括我们自己在内，现在活着许多人，还谈不上真正了解和懂得相对论，至于完满解释和应用就更难说了。可以预料，在目前和将来的相当长的一个历史阶段，相对论仍是照耀现代物理学发展的一盏明灯。

《数理统计在化学、光谱分析中的应用》一书，是一本优秀基础科学著作——论述严谨，层次分明，由浅入深，通过化学分析和光谱分析工作中的实例，论述了数理统计在此领域中的应用和重要价值。尽管最近几年来国内外都出版了一些数理统计在化学、光谱分析中应用的专著，但**未发现有哪本著作能象本书这样系统、深入和实用**，而且**目前仍被许多专家引用**。因此，本书对分析工程师、技术人员、实验员、教师和研究生的**目前和将来工作**，仍有着深远指导意义。我们认为，如果国内有较多的同行能比较彻底地看懂该书，我

国化学分析和光谱分析工作的科学化和正规化，将有较大的发展！

本书俄文原著的翻译、整理人员如下：第二、四、六章为余生（周开亿复核）；第一、三、五章为周开亿（余生复核，下同）；第七章为郑明祥；第八章为赵会元，第九章、附录和索引为余生。余生对初稿进行了全面审校，并结合英文译本加以订正。数理统计的基本知识由周开亿摘编。

本书的翻译和整理工作，一直是在姚元恺和姚影澄先生领导下进行的。

航空工业部 621 研究所孙亮庆等同志参加了本书最后审校工作。

本书责任编辑是周开亿。

由于数理统计所涉及的问题范围较广，需要学习的东西还有很多，我们自己下功夫也很不够，了解不多，所以错误之处在所难免，敬请读者发觉时，不吝指正。

《光谱实验室》编辑部

1986年12月

目 录

序言	(1)
1 数理统计所解决的问题	(4)
2 分析中误差的分类	(9)
3 随机变量及其特征	(22)
3.1 随机变量的分布	(22)
3.2 随机变量的平均值及方差	(25)
3.2.1 随机变量的平均值	(26)
3.2.2 方差及均方误差	(29)
3.3 根据日常分析中的数据计算方差	(36)
3.4 误差叠加定律	(39)
3.5 间接测量时的误差	(47)
3.5.1 吸收光谱分析中的误差	(48)
3.5.2 发射光谱分析中的误差	(50)
4 正态分布	(54)
4.1 正态分布函数	(54)
4.2 一些与正态分布有联系的特殊分布	(64)
4.2.1 t 分布	(65)
4.2.2 χ^2 分布	(73)
4.2.3 F 分布	(76)
4.2.4 r 分布	(80)
4.3 观测所得分布与正态分布相近似的判据	(81)
4.3.1 利用 χ^2 判据检验正态分布假设	(82)

4.3.2	利用 λ 判据检验正态分布假设	(88)
4.3.3	根据大量小子样检验正态性假设	(93)
4.3.4	图示法	(97)
4.4	分析工作中偏离正态分布的情况	(100)
5	泊松分布和二项分布	(111)
5.1	泊松分布	(111)
5.2	用泊松分布估计半定量分析结果的误差	(122)
5.3	二项分布	(128)
6	如何估计分析结果	(133)
6.1	用 t 检验法比较两组平均值	(133)
6.2	若干个方差的比较	(138)
6.3	测量结果均一性假设的检验	(141)
6.4	序贯分析	(150)
6.5	非参数统计法	(158)
6.5.1	两类数据间存在固定差异这一假设的检验	(158)
6.5.2	数据波动具有随机性这一假设的检验	(161)
6.5.3	以切贝雪夫不等式为基础的估计	(165)
7	方差分析	(169)
7.1	由单因素作用所引起的方差的测定	(169)
7.1.1	方差测定的概念和简单的例子	(169)
7.1.2	各组数据个数不等时的方差分析	(176)
7.1.3	应用方差分析的条件	(178)
7.1.4	应用方差分析研究方法误差的例子	(180)
7.2	多级分组的方差分析	(183)
7.2.1	二级分组的方差分析	(183)
7.2.2	三级分组数据的方差分析	(190)

7.3	复杂实验——多因素组群的方差分析…	(195)
7.3.1	有重复的两因素组群的方差分析 ……	(195)
7.3.2	无重复的三因素组群的方差分析 ……	(209)
7.4	根据拉丁方方法的实验设计 ……	(219)
7.5	方差分析的效力 ……	(221)
8	线性关系的统计 ……	(222)
8.1	最小二乘法 ……	(222)
8.1.1	校准曲线参数的测定 ……	(222)
8.1.2	非等精度测量的加权平均值 ……	(231)
8.2	回归分析 ……	(232)
8.2.1	线性假设的检验 ……	(232)
8.2.2	校准曲线参数和其理论预期值的比较 …	(236)
8.2.3	两条校准曲线的比较 ……	(247)
8.2.4	校准曲线平行移动这一假设的检验 ……	(253)
8.2.5	估计由校准曲线所求分析结果的置信界限 和误差 ……	(256)
8.3	相关分析 ……	(261)
8.4	应用不同统计分析方法的实验设计例子	(277)
9	与统计实验设计有关联的工作方法 ……	(286)
9.1	取样和实验条件的随机化；随机数字表 的应用 ……	(286)
9.2	重复测定次数的选定 ……	(290)
9.3	文件编制 ……	(300)
9.4	控制图 ……	(314)

附录

- 附表1 正态分布函数 $\Phi(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^u e^{-(u^2/2)} du$
数值表..... (319)
- 附表2 双倍标准拉普拉斯函数 $2\theta(u) =$
$$= \frac{2}{\sqrt{(2\pi)}} \int_0^u e^{(-u^2/2)} du$$
 数值表..... (321)
- 附表3 t 的数值表..... (322)
- 附表4 χ_p^2 的数值表..... (323)
- 附表5 概率 $P(\chi^2 > \chi_p^2)$ 的数值表..... (324)
- 附表6 F 的数值表..... (328)
- 附表7 相对偏离 $r = \frac{x_i - \bar{x}}{s\sqrt{(n-1)/n}}$ 的数值表 (332)
- 附表7A r_{\max} (或 r_{\min}) 的数值表..... (333)
- 附表8 相关系数 r_{xy} 的数值表..... (334)
- 附表9 不同 λ 时的概率 $P(\lambda)$ (335)
- 附表10 r 值出现的概率..... (335)
- 附表11 正负符号出现较少的次数 g (337)
- 附表12 连贯数 R 的上、下限 (339)
- 附表13 最大方差与 k 个方差总和的比值..... (340)
- 附表14 随机数字表 (342)
- 参考文献..... (346)
- 索引..... (358)
- 数理统计的基本知识 (366)

序 言

现代数理统计学，是十九世纪末和二十世纪初由于生物学工作者和经济学工作者的工作需要，在概率论基础上开始形成的。在近数十年来，数理统计作为一种研究方法，在农艺生物学、医学、机械制造、仪器制造、化学工业、冶金工业等科学技术领域中，已经开始得到充分应用。随后，这种统计研究方法，更得到突飞猛进的发展。不久以前，以概率论为基础的一种新学科——信息论——已经形成。信息论原来的任务是研究无线电技术中与信号传递有关的问题。在信息论基础上，控制论又开始发展。信息论在光学方面的应用，原先完全没有预料到。现在，在科学技术资料记录方面，信息论思想的应用，看来很有前途。由于核物理学的发展，概率论新的应用领域——核子计数统计学已经出现。

数理统计新的应用领域之一，就是与化学、光谱分析有关联的研究工作。在化学、光谱分析中应用数理统计方法的必要性，取决于一系列的因素。这里首先要指出的是，组分复杂新型合金和材料在生产中的运用，以及生产过程的不断加速，都迫使我们广泛采用化学、光谱分析方法；而这些方法所依据的过程，研究和了解得尚不够全面，不易严格控制，不易精确调节。因此迫切提出了寻找合理判别标准的问题，以比较由不同分析方法所得到的结果。

新的分析方法的研究及运用，要比其标准化快得多。这就不可避免会导致在每一个分析实验室内，经常碰到复杂的计量问题；不采用数理统计方法，要解决这些问题是不可能

的。分析工作者应当很好掌握数理统计方法，是很明显的事。

应用数理统计的每一个新领域，都要求有它自身专门的^一处理方式。在某一应用领域内所获得的统计研究的经验，不应机械地搬到另一个领域中去，即使它们之间似乎是很近似的。例如由解决计量问题和大地测量问题所研究得到误差理论，若不经形式上改变，就不可以生硬地搬到光谱分析的领域中来。因此，除了一般共同性的数理统计教科书外，对某一专门领域的工作人员来说，还需要专门的数理统计书籍。世界各国出版这类专门书籍已积累了很多经验；除已出版了相当多的一般性专供研究工作应用的数理统计方法的参考书籍外，也出版了一系列专为化学方面研究工作所用的参考书。

在本书中，企图把近来与化学、光谱分析有关的统计研究的工作，加以系统化和总结。我们推测在不久的将来，将会建立起以概率论为基础的、化学、光谱分析的普遍理论，正如曾成功地创立了测量方面的普遍理论——计量学——那样。在建立这种理论中，首先的和最困难的一步是要把涉及化学、光谱分析的问题，用数学语言表达出来。

本书的编写主要着重在实际应用，而以处方的形式写成。数理统计学的主要理论将不予论证，而是用化学、光谱分析中的具体例子来加以说明。理论性问题的讨论，仅以能理解所讨论的问题的计量方面是必需的为限。

学习本书内容，需具有高等院校的数学知识，并对概率论基础有一定认识。书中有关概率论方面各章，是为了提示读者，使能对概率论的主要理论予以回忆。

本书并不想对近代数理统计思想作全面的系统论述。

书中仅仅讨论与化学、光谱分析有关联的、在试验室已经应用、并在期刊上已经发表过的数理统计方法。

书中题材的排列顺序，是按照统计分析基本思想的发展先后为序，并用化学、光谱分析中的应用实例作说明。在这种排列方式中，为了解决同一形式的分析化学研究课题，不得不分几次前后反复讨论。然而我们认为这样来回重复是完全恰当的，因为在解决同一形式的光谱、化学分析课题时，既需要用到很简单的统计方法，也需要用到复杂的统计方法，而只有在讨论并了解了前面题材内容的基础上，才能够理解后面更复杂的统计方法。

1 数理统计所解决的问题

数理统计是一门以概率论为基础、利用实验结果来研究所探讨的现象的客观规律性的科学。

数理统计的这一定义具有普遍性，是因为在不同的科学技术领域中都有它的应用。数理统计在任一学科中的应用，总是以优先利用数理统计中某些特殊方面为主。在实验室工作中，特别是在化学、光谱分析工作中，数理统计是用概率论为基础的方法，主要对实验结果进行概括和分析。这可用下列事实来说明：在试验研究工作中，经常会遇到多种难以估计的因素的作用和这些因素的交互作用的情况。因而即使设置了一系列实验以后，一般还不能揭露出其中起作用的物理规律性。只有在不同条件、不同实验室、对不同对象的试验结果进行了比较之后，这种规律性才有可能搞清。但是，要进行这样的比较，只有用数理统计方法把实验结果整理和表达为便于保存、便于作进一步处理的紧凑形式后，才是可能的。所谓信息的概括(信息的紧缩、精简)，其中之一就是利用数理统计这一工具把有关分析方法的精确度的全部信息，表达为该分析方法误差的分布函数(或分布的规律性)，并以分布函数的参数来具体描述该分布函数。这些参数包括方差或均方误差、数学期望等。

在分析工作中，分析次数经常是不多的。这些少量观测值，可以看作是从某一假设的无限集合中随机抽得的子样。而这一假设的无限集合常称总体(或母体)。这就是实际观测值的数学模型。在这种情况下，从数学的观点看来，信息概

括的问题就是根据子样来求得某些量(随机变量的子样方差和算术平均值)。这些量是总体分布函数的参数(相当于方差、数学期望)的估计量。根据子样来估计确定总体的参数时,当然,会引入某种非肯定性成份;但该成份可用数理统计方法估算出来。在许多实验工作者之间,曾经广泛流传着一种完全不正确的看法,认为数理统计只适用于大量的数字资料。其实,现代数理统计,根据小子样,甚至在有些情况下只不过根据两个观测数据,就可能估计出总体的参数,并确定这些参数的置信界限。(当然,实验数据愈少,根据子样数值对总体参数估计就愈不够精确。)因此,数理统计一方面可以概括地代表实验结果;另一方面,在少数回次实验时,也能定量估计伴随着每一次实验的可疑成分(非肯定性成分)。

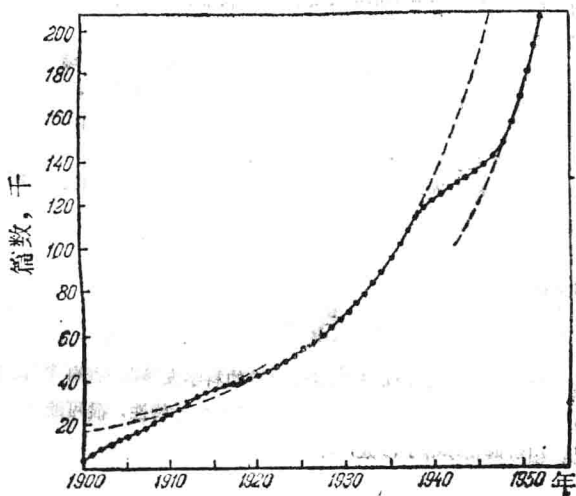


图1 1900年起在《物理文摘(Physics Abstracts)》上出版论文总数的增长
纵坐标表示累计总数(以千为单位);横坐标表示年份

由于在所有科技领域内发表的论文数目逐年增多，可靠方便的信息概括方法的拟定问题和标准化问题，就显得特别迫切。图1是研究报告篇数的增长曲线。这是根据1900年起发表在《物理文摘(*Physics Abstracts*)》上的材料绘制的。在差不多五十年的时期内，曲线呈指数形式^①。这种曲线的指数发展趋向，只在世界大战时才遭到破坏。每10—15年，出版论文总数就要增加一倍。出版论文数量的增大，不可避免地要使论文篇幅缩短。在近年来，某些杂志已经将论文篇幅缩短将近一半，看来今后还会更多地缩短。杂志编辑工作者任意削减论文篇幅，也会导致包含在实验资料中相当一部份信息的损失。因而，假如说，以往读者可能根据长篇的实

① 这种曲线的指数增长形式，从科学杂志的份数、从科研工作的人数以及研究工作的拨款数字，都可看出有这种特性。目前，杂志总数已超过十万，文摘杂志多于三百种。

在所有这些情况下，指数的常数值表明科学发展的指标为每10—15年内增加一倍。根据这曲线向过去历史外推，就得到纵坐标值等于1，那时为公元1700年，即牛顿时代，这一时代可以认为是近代科学发展的起点。因此，科学发展的指数增长特性，看来在近200—250年的时期内就已实现了。

科学发展的指数特性，可以从具有相当普遍性的、非常可能的假设推导出来。指数方程的表达式

$$y = ae^{kt}, \quad k > 0,$$

可以看作是微分方程

$$\frac{dy}{dt} = ky$$

的解。在这里，导数 $\frac{dy}{dt}$ 就是我们所感兴趣的科学发展指标的增长速度，也就是在单位时间内的增长数。因此，科学发展的指数特性，很可能是这样一种假设的结果：相对增长率等于常数，即

$$\frac{dy}{ydt} = \text{常数}.$$

不难看出，在10—15年内所有指标的加倍，相当于相对增长率为每年5—7%

验条件和处理方法的叙述，来估计实验结果的可靠程度，那末现在，这种可能性便消失了；而定量估计与每一实验有联系的非肯定性成分的必要性，就感到特别尖锐迫切了。

在各类分析方法规程中，以及在苏联国家标准（ГОСТ）中所介绍的有关分析精密度和准确度的报导资料；以及说明标准样品的资料（如说明标样均匀性、和确定标样中某成分平均值的可靠性的资料等等），同样都应看作是研究某一定问题时包含在大量实验工作中的信息的概括。

最后，日常分析结果也可以看作是信息的概括。这些信息是分析未知样品时若干次重复分析所得到的；在有些情况下，还可以是对标准样品进行核对分析时所得到的。

因此，可以肯定地说，借助于数理统计来概括和紧缩信息，应该是任何分析过程中的一个组成部分。只有当资料能表达为简洁紧凑的形式、所得数据的可靠性得以评价时，这种材料的分析，才能认为是完善的。

应用数理统计于信息概括的思想，是费歇尔(Fisher)提出的。大家知道，英美统计学校的创始人——费歇尔和皮尔逊(Pearson)是享有声誉的；在他们的著作中，常把信息概括转变为目的本身。这种对数理统计的使命之一的解释，当然不能认为正确；这种解释与辩证唯物论的观点是相矛盾的。后者认为，揭露客观规律性才是数理统计的主要使命。

信息的概括，事实上并不是目的本身，而是对客观存在的规律性的复杂认识过程的一个组成部分。让我们用下面的例子来说明这一观点：对发射光谱分析中的一个重要问题，如所谓“第三元素影响问题，曾有成百篇研究报告在不同程度上阐述过，并且每一篇报告利用有限的实验资料，对这一问题的某一特殊情况加以探讨。分别取这些报告中的任一

篇，都不包含足够的数 据，以致能够洞悉这一复杂现象的本质。但如果对所有这些论文的结果，借助于数理统计来加以概括总结，表达为某些便于比较、便于处理的标准形式，那末，在概括全部资料时，一方面可以更了解这一复杂现象的本质；另一方面，就可以为拟定分析方法作出重要的结论。

由此可见，如果我们利用数理统计来概括信息，那末使我们有可能把各个研究工作看作是一大堆实验工作的一部分；而把概括信息的过程本身看作是认识过程中的某一组成部分。这一概括的过程，目前之所以变得特别迫切重要，是因为需要进一步概括的、具有局部意义的各个单独研究项目愈来愈多。在这样的问题提法下，我们所感兴趣的数理统计，就可看作是控制论的使命。控制论首先涉及的是信息的概括和处理问题。

更为复杂的是我们感兴趣的数理统计的第二个使命——实验数据的分析。这两种数理统计的使命——实验数据的分析和简洁地表示所得到的实验结果——是彼此有机地联系着的。

统计分析的目的，一方面是使在实验所耗费的劳力为最小的情况下，能得到最大限度的信息；另一方面，是评价所得结果的可靠性。

当研究工作者利用有限的实验数据想要清楚地估计某一种因素所起的作用时，就必须着手搞统计分析。假如必须同时研究好几个因素所起的作用、以及这些因素相互之间的交互作用，则统计分析就更为复杂。由于统计方法的发展，在有些情况下，有可能彻底改变实验的工作方法。如果在过去研究工作者按古老办法安排一些实验，使得只是改变