

7
高等学校教学参考书

概率图在矿床勘探中的应用

〔加拿大〕A. J. 平克莱 著

地 质 出 版 社

139911

圖書分類號碼

概率圖在可靠數據中的應用

作者: 李國華, 李國華, 李國華

中國科學院
數學研究所
北京

圖書分類號碼

高等学校教学参考书

概率图在矿床勘探中的应用

〔加拿大〕A.J.辛克莱 著

赵鹏大 等译

任建新 校

地质出版社

内 容 提 要

加拿大 A. J. 辛克莱著的《概率图在矿床勘探中的应用》一书包括八章。第一章为绪言，第二、三章阐述正态及对数正态单一总体在各种概率图上的图形特征，四、五两章着重叙述两个总体混合的双峰概率曲线及其筛分过程，第六、七两章进一步探讨了二个以上总体混合的多峰分布概率图形及其筛分，同时阐述了化探数据背景值的估计问题，第八章讨论了几个附加的问题。

本书简明扼要、实用性强，可作高等地质院校应用地球化学专业的教学参考书，并供数学地质工作者及广大地质勘探工作者阅读。

Applications of Probability Graphs in Mineral Exploration

By Alastair J. Sinclair

Dept. of Geological Sciences

University of B. C.

Vancouver, B. C.

概率图在矿床勘探中的应用

〔加拿大〕 A. J. 辛克莱 著

赵鹏大 等译 任建新 校

责任编辑 赫祥安

地质部教育司教材室编辑

地质出版社出版

(北京西四)

地质印刷厂印刷

(北京安德路47号)

新华书店北京发行所发行·各地新华书店经售

开本: $850 \times 1168^{1/32}$ 印张: 3 字数: 74,000

1981年7月北京第一版·1981年7月北京第一次印刷

印数1—3,910册·定价0.48元

统一书号: 15038·教113

译者的话

辛克莱著的这本小册子《概率图在矿床勘探中的应用》是一本带普及性的统计分析专著。说它带有普及性，是因为书中介绍的主要方法——地质勘探数据的概率图解法，可以毫无困难地在广大地质工作者中普及推广。一切和数据打交道的地质工作者，不论他们从事的学科范围如何，都可以有效地应用书中介绍的方法去研究数据。在这当中，并不需要特殊的计算工具和设备，也不需要太多的数学准备或基础。然而方法的应用却能获得很多重要信息，提供有助于解决实际问题的线索。因此，可以认为概率图解法是解决某些地质勘探问题的一种多快好省的辅助工具。

尽管本书具有这种普及性，但它所谈及的问题却是统计分析中带有根本性的基础理论问题，即数据的概率分布模型问题。众所周知，数理统计的许多基本理论和方法是以数据所服从的分布律为基础的。如许多单元和多元统计分析方法以数据服从正态分布为前提，如果数据偏离正态分布，则需进行相应变换使其呈正态分布，否则只能是一种近似或转而采用其它非参数性统计方法。另外，数据的概率分布模型或统计分布特征是研究对象成因特点的象征。就各种地质体而言，不同成因的地质体其数据的概率模型也各异。由于地质作用的长期性和复杂性，地质体往往经历了多个成因阶段或具有复杂的成因过程。因此，我们对各种地质体取样观测所得数据，往往代表多种成因或多次作用迭加的综合结果。表现在统计分布特征上，它们一般为由多总体混合而形成的多峰复合型分布。

可见，为了正确地进行统计分析工作，首先应查明所研究的数据集是单一成因总体抑或为多成因混合总体。如系前者，则应进一步查明其分布律；如系后者，则应设法将原始数据筛分成不

同的成分总体，然后再对各成分总体分别进行统计分析和地质解释，这对涉及成因问题的研究尤其重要。

数据的统计分布模型不仅具有重要的理论意义，而且有很大的实际意义，如化探中异常下限值的确定，编制各种等值线图时合理等值线值的选择，各种地质现象发生的概率估计等。因此，对做为概率统计分析基础的数据分布律的研究是一项十分重要的工作。

近年来，我国数学地质工作蓬勃开展，可以预料，随着电子计算技术的普及还将有更大的发展。当前，各种多元统计方法已相当广泛地应用于各项地质研究。但值得注意的是，在一些工作中，由于忽视了数据基本分布特征的研究，把本来属于混合总体的数据，不做任何筛分而笼统地进行简单处理；或者把不同空间部位具有不同混合程度的数据等同对待，不加区别地做为一个整体进行处理。凡此种种，势必影响统计分析效果，甚至可能导致错误的结论。

本书详细介绍了用概率图解法查明数据分布律的方法，特别着重阐明了双峰和多峰混合分布的筛分方法和原理。作者从已知各成分总体混合比例的理想条件入手，研究并建立了各种混合分布的概率图模式，其中也包括单一总体的截尾分布模式。作者强调在确定混合分布并筛分不同成分总体的基础上，对数据进行合理的有效分组；这就有可能研究各组数据的空间分布和它们的地质背景，分别估计各成分总体的统计参数并做进一步的统计处理。作者引用一系列实例说明正态和对数正态混合总体的各种模式，并阐述其筛分过程，这些都是有参考意义的。

如前所述，本书所涉及的是数学地质的一个基础理论问题，然而作者主要注意力集中于方法的阐述。有关混合分布的成因理论并未做深入探讨，对影响统计分布性质的各种因素未作必要的分析，至于各种类型混合分布的地质意义也缺乏详细探讨，这些不足之处有待广大地质工作者在实际应用中加以补充和提高。

本书由武汉地质学院勘探教研室数学地质组同志译出，参加

翻译的有胡旺亮、李紫金、冉宗培、陈丙芳、邹海卿及赵鹏大等同志。译稿蒙武汉地质学院勘探教研室任建新同志校阅，特致谢意。

赵鹏大

1978.8.21

序

累积概率图在许多不同领域里种类繁多的问题中都能应用。本书的主要对象是矿床勘探工作者，但所论述的方法当能得到更为广泛的应用。文章是以这样的思路组织的：介绍某个问题时，最好的办法是先有一个清楚的理论探讨，然后继之以实例。作者自始至终都在尽量利用他亲自熟悉的实例，但愿这样的局限性不致于妨碍所述方法之更一般性的使用。

在过去几年里，作者益发专心地研究了许多类型矿床勘探数据的统计分析，并逐渐写成了这本小册子。有三本参考书对作者观点的形成很有帮助，这三本书的作者是：坦南特 (Tennant) 和怀特 (White) (1959)、莱皮特尔 (Lepeltier) (1969)、博尔维肯 (Bolviken) (1971)，特致谢意。本书的某些章节曾多次在校内和产业部门作为讲座的基础使用过，非常感谢这些讲座的许多参加者的建设性意见。由加拿大全国科学研究委员会和动力、矿山资源部资助的若干研究项目对本文引用的概念和例子也有所帮助。通过研究合同而从动力、矿山及资源部取得经费资助使原稿的准备成为可能。

勘探地球化学家协会的计算机应用委员会成员加勒特 (R. G. Garrett)、米斯奇 (A. T. Miesch)、罗斯 (A. W. Rose) 和霍斯内尔 (R. F. Horsnail)，以及我的同事弗莱彻 (W. K. Fletcher) 和蒙哥马利 (J. H. Montgomery) 等对原稿提出了十分有益的建设性意见。勿容置疑，所有出错和遗漏的责任都归于作者。

分析化学家本茨恩 (L. F. A. Bentzen) 和奥尔 (J. F. W. Orr) 分别提供了技术性帮助，科克伦 (D. R. Cochrane) 先生善意地提供了用到的某些地球物理实例。插图由迈尔斯 (M. Waskett-Myers) 绘制，各阶段的底稿打字是由马伦 (J. Mullen)

夫人和海伍德 (C. Heywood) 夫人完成的。我对 所有 这些人的
贡献表示感谢。

A. J. 辛克莱

1976年2月19日于温哥华

目 录

第 I 章	绪言	(1)
第 II 章	累积概率纸和单一总体	(8)
第 III 章	单一总体的实例	(19)
第 IV 章	两个假设总体的概率图	(26)
第 V 章	双峰概率曲线图的实例	(41)
第 VI 章	三个或更多个总体联合的概率图	(54)
第 VII 章	多峰数据的有效分组—— 地球化学数据异常下限的估计	(61)
第 VIII 章	附加的论题	(71)
参考文献		(84)

第 I 章 绪 言

I-0: 概述

概率图的应用仅仅需要对简单的统计概念有一个一般性的了解,我们假定读者对这种概念是熟悉的。诸如算术平均数、方差、标准差、正态密度分布等这些基本术语将不详细论述,必要时可参考大量入门统计书藉中的任何一本。

概率纸的介绍推后至第 II 章。第 I 章里的几个标题考虑的或是叙述概率图本身,或是叙述本书中概率图所涉及的特殊应用。素材的安排试图有助于说明从大家所熟悉的直方图到概率图的简单进展。

I-1: 矿床勘探数据

矿床勘探过程中测得的许多变量都是连续的或近于连续的。例如,用作 Cu 分析的 700 个土样可能得出 Cu 的数值范围是 10—100ppm,此变量(土壤中的 Cu)在这个界线内是连续的,因为至少在理论上任何中间值都可能被某一样品所实现。当然,事实上绝不会有数值例如恰好被测得为 927.341。与连续变量相反的是所谓离散变量,即只取特定值的那些变量,例如,矿石光片中矿物种数是个离散变量,在一块给定的矿石光片中矿物种数可以是 1、2、3 等等,而不会是 1.372 种矿物。在本文中我们关心的是那些连续的或近于连续的变量。

对勘探过程中所得连续定量变量的分析方法有很多,我们只对那些与讨论概率图有关的方法才予以考虑。最常见的一种方法是对表列数据或根据经验圈定的数据进行主观地质分析,这种数据解释方法越发正规地辅以各种复杂程度的统计分析。即便是编制一张简单的直方图,都可看作是一种初步统计步骤。无论数据的

统计研究如何复杂，人们都可以确信，统计研究必须包括计算平均数和标准差，可能还包括数据的某种图示法。随着数据的增加，平均数的计算以及特别是标准差的计算，如果用手算方法，就会变得更加困难。而在野外取得这些参数的估计量的实际困难，似乎是不可克服的。幸好，正如我们将看到的那样，累积概率纸的使用在一定程度上克服了这些困难。

一般说来，一组数值代表总体的一个统计样本。从规则网格点上获得的500个磁法测量数值就代表对数值的一次抽样，如果网格起始于另外某个点，则将得到有某些差异的另外一组500个值。可能获得无限多个这样的统计样本。我们的那个样本（500个数值）只不过是其中的一个现实。这里要注意“Sample”这个词在统计学和地质学中用法上的区别，单个的土壤 Sample（土样）在地质上是一个样品，而在统计学上则它仅仅是统计样本的一个元素。

I-2: 直方图

直方图是表示数量信息的一种常用方法。图 I-1 所示三个直方图是矿床勘探数据（此处为矿石品位）总体中常见的几种变化形式，它们分别表示负偏倚[图 I-1(a)]、对称[图 I-1(b)]和正偏倚密度分布[图 I-1(c)]。直方图作为数据的一种直观表示法有以下几个明显的优点：（1）样本的整个数据范围一目了然；（2）能很容易地看出众数；（3）能很快估计出最大丰度值的范围；（4）数据密度分布的全貌也是清楚的。在某些情况下，直方图对区分背景和异常值之间的异常下限是有用的。为此目的，直方图得到了相当广泛的应用。直方图的一个附带优点是，数据的初步分组提供了一种根据分组数据方法计算平均数和方差的比较简便的形式。

为了达到预定的目的，数据必需具备适当的质量——即它们一定是有代表性的（无偏的），同时，用于获得“数字”的测量技术必须有足够的精度。最优数据收集法可以建立在大面积地球化

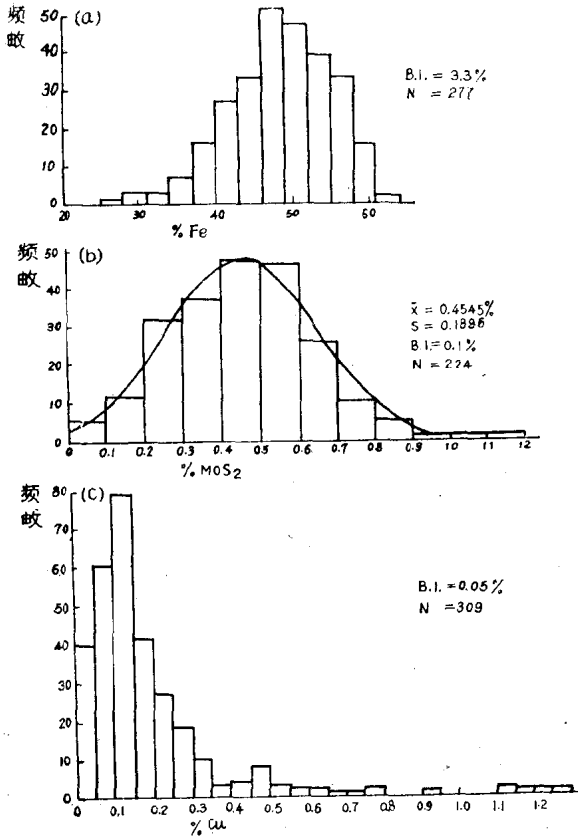


图 I—1 实际的矿石品位密度分布直方图三例。

(a) 负偏倚；(b) 对称；(c) 正偏倚。

B.I. 为组距，N 为样本大小。对对称分布拟合了一条与实际数据有相同平均数和标准差的正态曲线（辛克莱，1972）。

学和地球物理测量中常见的定向调查的基础上。

关于直方图的制作，有几点是值得指出的：首先是组距的选择，肖 (Shaw) (1964) 认为，组距最好选择在数据标准差的四分之一和二分之一之间。如果组距太大，分布的真实形式会被掩盖，组距太小，则所得直方图缺口太多。

编制直方图的第二个问题是第一组从何处开始。第一组选择

在何处开始，并不是一件像法规一样严格的事情，但将其步骤标准化，并使有一或二个分组对于平均值呈对称排列，则似乎是合理的。

如果要与其它直方图做对比，那么直方图应该用百分率作纵座标（频率），否则，对两个或更多个直方图（每个直方图所依据的数值个数显著不同）进行有效的直观对比是难办的或不可能的。一般说来，编制直方图时最好包括将下述内容列制成表：

（1）标题，（2） N ——样本大小，（3）组距，（4）数据的平均数和标准差。

I-3: 连续密度分布

对于大样本来说，随着直方图组距的缩小，显然越来越容易通过各组的顶部勾画出一条光滑的连续曲线。因而，对许多连续变量的频率分布，用一条被称为密度分布的光滑数学曲线来逼近是有可能的〔见图 I-1(b)〕。可以设想，为了估计实际数据的全部可能分布，需要很多这样的数学模型，而且虽然这一点在理论上可能是正确的，但幸而实际中并非如此。自然界大多数变量所呈现的频率分布形状能被为数不多的数学模型所近似。事实上，在本书中我们将限于考虑正态分布和对数正态分布这两种特定形式，在本章的后一部分将设法证明这种考虑是正确的。大家应当记住，还有许多离散的和连续的密度分布模型正在使用，其中包括二项分布、泊松分布和 γ 分布。为了参照这些分布，读者得查阅典型的统计书籍。

高斯正态分布最先是作为测量误差理论提出的。譬如，我们也许想测验土样化学分析方法的复验性，为此可以把一个拌匀的样品分成10份分样，其中每一分样都用同样的方法分析。由于分析过程中存在随机误差，所测得的10个数值未必完全相同，测定值围绕着平均数的分布即服从所谓正态或高斯密度分布，其密度分布由下式给出：

$$Y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(x-\mu)^2/\sigma^2}$$

式中 μ 是算术平均数, x 是任意测定值, σ^2 是总体方差。其图形表示乃是如图 I-1(b) 所示的钟状曲线。

I-4: 累积分布

准备用作标准直方图的数据, 还可以表示为累积直方图, 办法是把任意一组的频率加上它前面所有组的总频率。频率累加既可以从数值范围的高值端开始, 也可以从低值端开始。这种表示法在沉积学中常见于用来分析沉积物的粒度, 频率是由粗粒级向细粒级累加。一个例子示于图 I-2(a) 中。在图上, 标准直方图和累积直方图的关系是显而易见的。将该累积直方图及一条光滑累积曲线重新作在图 I-2(b) 上, 但座标已倒换。在这个假拟的例子中, 数据是由高值向低值累积的。

这种数据表示方法已经满足了在概率图纸上作图的要求 (参

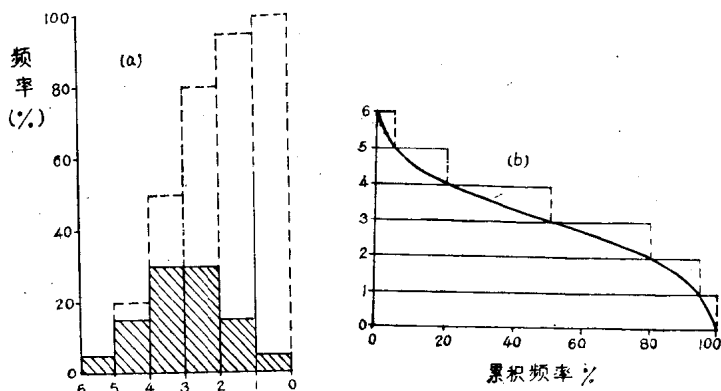


图 I-2

(a) 一个假拟数据集的直方图和累积直方图。注意数据是从高值向低值累积的。(b) 用相反的座标重画 (a) 的累积直方图, 并配有一条连续曲线 (累积曲线), 这种直方图的图形与本书中所用的概率图的关系很明显。注意, 通过适当地改变刻度, 累积曲线可以转变为直线。

见 II-1 节)。累积直方图的概念对理解概率图是易懂而又重要的。

I-5: 对数正态分布

对数正态分布的最简单的概念,可看作是一组数据对数值(对任意底)的正态分布。对数正态分布已由艾奇逊(Aitchison)和布朗(Brown)(1957)作过详细阐述。如同正态分布是加法误差的理论基础一样,对数正态分布可作为乘法误差的理论基础。

大量文献谈到原始数据的密度分布形态或形式(如阿伦斯 Ahrens, 1954; 肖 Shaw, 1961; 贝克尔 Becker, 和黑兹恩 Hazen, 1961; 拉吉昂诺夫 Rodionov, 1961)。很多作者认为,许多地学变量具有非常接近对数正态律的密度分布。例如,下列变量通常表现出很强的对数正态性:(1)地球化学中的微量元素(肖 1961);(2)许多地球物理变量(斯利克特 Slichter, 1955);(3)矿床的品位和储量(辛克莱, 1974, b);(4)沉积物粒度数据(哈里斯 Harris, 1958);(5)水藏的容水量(黑兹恩, 1914);(6)油藏的规模(麦克罗森 McCrossan, 1969)等等。另外,地质工作者日常处理的某些变量虽呈正态分布,但变量的本性就体现了一种对数变换,如 pH 测量值和沉积粒度数据(单位为 ϕ)都是常见的例子。

在强调文献中普遍承认大部分地学变量呈对数正态分布,以及强调矿床勘探中日常观测的许多变量都非常近似对数正态分布的同时,还必须要注意到对这样一种简单的解释仍存在着某些含混不清和复杂性。例如,并没有明确规定的或者不言而喻的基本“对数正态”律,对数正态模型仅仅是对现实的充分逼近而言。必须看到,实际数据在所配密度分布的尾部常常偏离被配的经验模型。很难准确地确定一个给定的模型在何处不再适用于实际数据。但可以毫不怀疑,比如说,一个对数正态模型肯定不能扩展到建立这个模型的数据范围以外的数据上去。另外,除了对数正态分布之外,还会碰到其它各种分布(戈维特 Govett, 1975),

人们必须时时注意到这种可能性。

使一对数正态分布模型与许多实际数据相符的最严重问题，多半是与多峰分布有关。多峰分布中各成分总体彼此没有明显重迭时，每个成分总体可以单个地检验其对数正态性。这种多峰对数正态分布的发生频率表明，预期重迭总体也近似于对数正态模型是合乎逻辑的。当然，后者的解释实际上已经解决得很好（例如蒙哥马利 Montgomery 等，1975；萨格和辛克莱，1974）。本书通篇有例子检验混合对数正态总体的存在。

避开烦琐哲学的争论，作者认为在地学中通常遇到的许多变量，特别是那些在矿床勘探工作中经常测定的变量非常接近对数正态，或接近于两个或多个对数正态总体的混合。很多其他变量则近似于正态密度分布。从实际观点出发，这就意味着许多数据通常可以在标准对数概率图上或算术概率图上加以表示和说明。