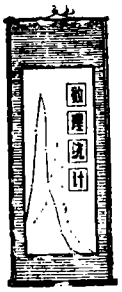

数理统计在化探中的应用简介(二)

地　文

4.概率分布类型

在数理统计理论中有一条非常重要的定律，叫做中心极限定律。它用严格的数学步骤证明，如果变量ζ是由许多互相独立的偶然因素相加的结果，而这些因素对ζ的影响没有一个是非常显著的，即设：

$\zeta = \alpha + \beta + \cdots\cdots + \varepsilon + \cdots\cdots$，而α、β、ε均为起伏不很显著的变量，则ζ服从正态分布规律。正态分布的一些基本特点已在上期中作了介绍。例如上期例子中对同一个均匀试料作多次重复的化学分析，其结果必定服从正态分布规律。因为每次分析结果中都包含着称样、熔样、稀释、分液、试剂体积、反应条件、读数等等因素的影响，使测定结果这一变量接近中心极限定律的要求。但光谱半定量分析不满足这一条件，因为谱线黑度与含量是对数关系，黑度测读的误差对含量的影响显著地大于其他因素。

如我们考察取自同一岩体不同地点的样品中某一微迹元素的含量，或某一单矿物中某一微迹元素的含量，可以发现这些含量之间的变化，除受到取样、加工、分析等偶然因素的影响外，还往往包含着某种起显著作用的因素的影响。例如质量作用定律、晶体化学规律等。由于这些基本规律的作用，使岩体不同部位的微迹元素含量发生某种依赖关系，而不是互相独立的，因而不满足中心极限定律的要求，使这种含量的概率分布呈现复杂的情况。目前在理论上还没有研究清楚，但有一种比较有代表性的变量在地球化学研究中有很大意义，这种变量由许多相互独立的偶然因素以倍数的方式作用的结果所决定。例如次生晕中观察到的金属含量，在其形成过程中发生的分散、淋滤、淋积、残余富集和生物作用等，这些因素的影响总是以其原有含量为基础增加或减少若干倍。又如对同一期热液活动形成的原生晕中金属的含量也可作类似的分析。这类变量可表示为：　$\eta = \alpha \cdot \beta \cdot \cdots\cdots \cdot \varepsilon \cdot \cdots\cdots$

两边取对数得：$\log\eta = \log\alpha + \log\beta + \cdots\cdots + \log\varepsilon + \cdots\cdots$

令　$\log\eta = \eta'$，$\log\alpha = \alpha'$，$\log\beta = \beta'$，……等。

则：　$\eta' = \alpha' + \beta' + \cdots\cdots + \varepsilon' + \cdots\cdots$。

如果α′，β′，ε′等也不包括显著过大的影响，则我们又回到了正态分布的情况。这就是说，根据中心极限定律可以推知变量η的对数值$\log\eta$服从正态分布。这种分布称为对数正态分布。它也有两个基本参数，其一是平常所说的几何平均值，其二是对数方根差。这类分布在地球化学资料的统计分析中占有很重要的地位。在50—60年代，地球化学界曾对岩石和矿物中元素含量的分布型式进行了热烈的讨论。大体上可以归纳出如下一些看法：

1.一定的地球化学指标在一定的地质－地球化学过程之后（或简单地说在一定的地质单元之内），总会形成一个统计学上的母体，并对应地有某种理论上的概率分布。例如铜的含量，在没有矿化的地段是一种分布(背景分布)，在矿化地段是一种分布（异常分布），在含矿岩体中是一种分布，在无矿岩体中又是一种分布；在花岗岩体中是一种分布，在灰岩中又是一种分布，等等。这些分布的差别，可以是型式的不同，也可以是同型式而参数不同。

2.岩石、矿石中的常量元素大多数服从正态分布，而微量元素大多数服从对数正态分布。在少数情况下当某种微量元素近似均匀地分散在多种矿物中时该元素呈正态分

© 1994-2009 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

布，当某元素主要集中于一种矿物中时，呈对数正态分布。

3.一单地球化学作用形成的浓度分布，通常服从对数正态分布，而多次作用叠加的结果，就向正态分布接近。如新鲜花岗岩中亲铜元素呈对数正态分布，而在经多次表生作用结果形成的土壤中则呈正态分布。

4.除正态分布和对数正态分布以外，还可能存在其他分布，如β分布等，因不常用，这里不讨论。

5.复合母体的分解

当一套子样包含着取自两个以上不同母体的个体时，概率分布就复杂化了。这是化探工作中最常遇到的情况。这里着重介绍一种用概率格纸的分解方法。

分解由两个母体A和B形成的复合母体时，要考虑如下三种情况：

1.母体A、B的分布型式。这里有三种可能的组合：两个都是正态母体；两个都是对数正态母体；一个是正态母体，另一个是对数正态母体。

2.母体A和B的分布参数的差别。如果参数相差不显著，则无法用统计方法分解。

3.子样中来自母体A的样品数n_A与来自母体B的样品数n_B在样品总数$n=n_A+n_B$中所占的比例。设$P=\frac{n_A}{n}$，$q=\frac{n_B}{n}=1-p$

可由下式计算出复合分布：

$$F_{A+B}(X)=PF_A(X)+(1-P)F_B(X)。$$

当A、B的分布参数已知时，给定一个P值，就可以根据给定的$F_A(X)$，$F_B(X)$算出一条$F_{A+B}(X)$曲线。取一系列P值，即可得到一张列线图，图5、图6即为两个例子。再改变A的参数即可得到一套列线图。当我们根据实测数据在对数概率格纸上作出累积分布曲线后，就可由列线图中找到一条接近的曲线。如图5中虚线为实测曲线，由列线图中可读出A、B的分布参数及它们所占的比例，然后就可以根据地质地球化学条件对A、B分别说明其地质意义。

当A、B的分布型式相同而参数相差较大时，复合分布曲线的高、低含量端均呈直

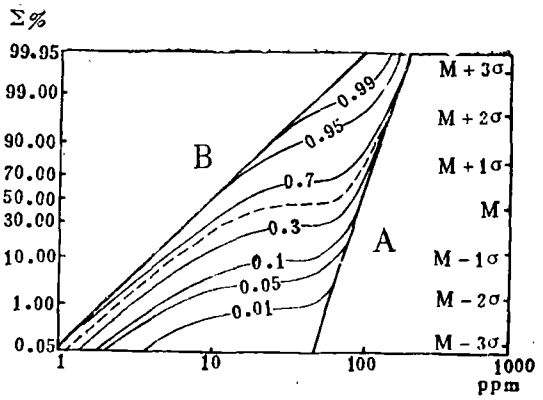

图5

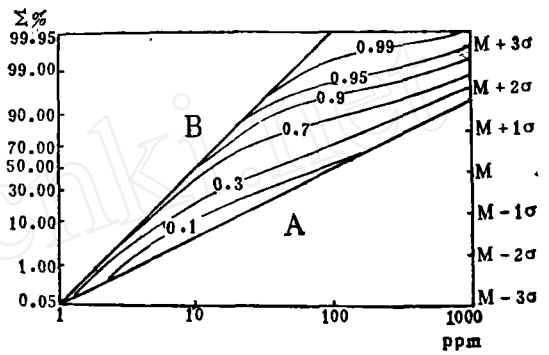

图6

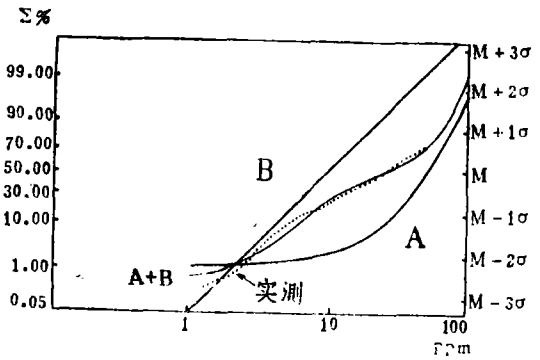

图7

线，延长这两条直线就可以大致地求出A和B。但是实际上A、B的分布型式可以不同，参数变化可以无穷，所以要作出一套完整的列线图是很困难的。因此建议采用经验选择法，即根据实测曲线的形状假设A、B及混合比P，算出复合曲线与之对比，根据二者的接近程度修改A、B参数及P值，如此反复几次即可。例如图7中的虚线为1250个腐植质样品灰份含量实测累积分布曲线，这个

© 1994-2009 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

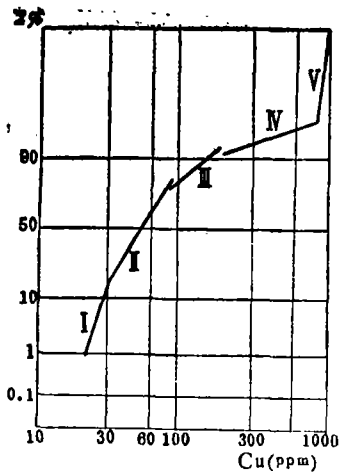

图8

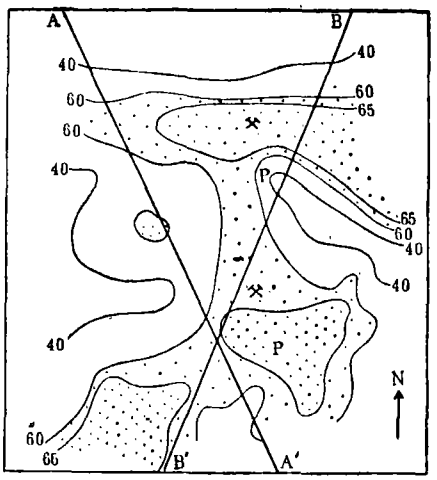

图9 父－矿山，P－勘探区，下同。

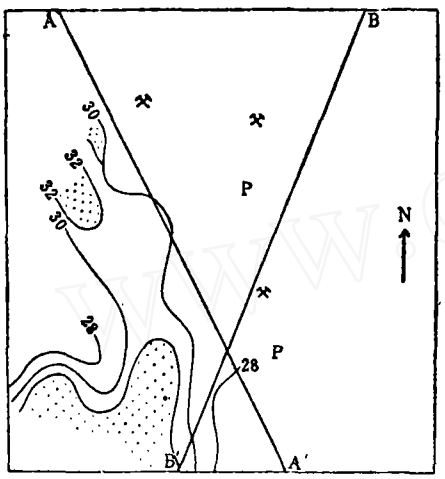

图10

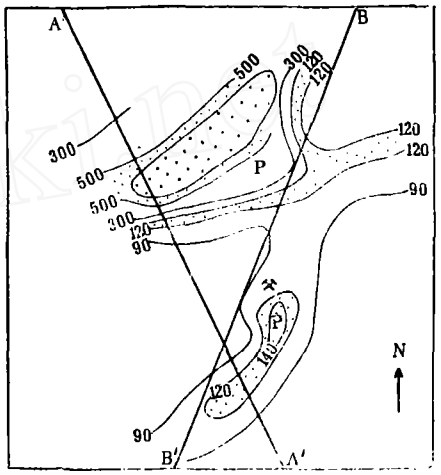

图11

分布和A、B的混合分布（P＝0.5）曲线很符合。其中A为正态分布，B为对数正态分布。一般认为正态分布反映有机成因的灰份，对数正态分布反映无机成因的灰份。它们在样品中各占50%。

6.概率分布的概念在资料整理中的应用

化探原始数据的一般处理方法是求出一个异常下限，据此将数据分成背景和异常两部分。这样就不能充分利用分析成果，特别是小比例尺大面积工作中，不能反映测区所包含的各种地质情况和自然环境。如能利用化探数据中所包含的多重统计分布，对每一种可分辨的统计分布单独作图，就可以从一批原始资料中得出若干张图，说明更多的地质问题。下面以美国亚利桑那州某地区的铜分散流测量为例说明这种整理方法。

第一步，把所有分析数据按含量递增的顺序排列。在对数概率格纸上作出累积频率分布图，并分出若干不同斜率的直线段，每一直线代表单一频率分布。如图8，可分出5个直线段。第二步，对每一直线段范围内的样品分别作出等量线图。例如对属于最低直线段（第一段）的数据单独作图，可得低Cu背景图（图9）。这一地区产金而没有铜矿。Cu的分布与岩性有关。根据第二段直线作出的高铜背景图（图10），它反映了当地的断裂构造。根据第三、四、五段直线作出铜异常图（图11）。近东西向的低异常是高铜岩石引起的，测区北部的最高异常与已知铜矿采区一致，北东—西南向的中等异常也与几个铜矿区有关。

© 1994-2009 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net