

样品观测值实质上都是三维数据。所以应试用本法进行三元回归分析,以图了解金属分布的全部型式。

化探异常延拓方法不依赖于经验评价指标,强调异常自身垂直变异规律,认为成矿元素含量向深部明显呈双指数规律变富是矿致异常的根本特征,是区分工业矿化与分散矿化的定性标志,所圈出的趋势剖面中的成矿浓度带对于指导探矿无疑是一种有力的暗示。该法使用条件简单,在圈定整个异常形态的基础上,只要做1~3条精测横剖面,对岩石样品进行定量分析即可,适用于原生异常检查评价。但由于原生晕形成机制相当复杂,形成理论尚不完善充分,所有矿致异常金属含量变化肯定服从一定规律,但是否完全遵循双指数函数规律不敢断言,因此,此法尚需与其它方法配合验证,目前尚处初步试验阶段,需在实践中不断充实完善。

主要参考文献

- [1] 秦皇岛冶金地质进修学院编:《现代成矿理论及勘查地球化学汇编》,第一、二、三、四集,1981
- [2] H. E. 霍克斯等(谢学锦译),《矿产勘查的地球化学》,地质科学院物探研究所印,1974
- [3] 地质部情报研究所编译,《化探资料选编》一、二、三、四集,1974~1981
- [4] 吴锡生,《化探讲座》,宁夏地质局物探队科技情报组印,1980
- [5] 云南地质局物探大队编,《云南物化探》,1977年第5期
- [6] 福建地质局科技情报室编,《数学地质》(一)、(二)
- [7] 中国科学院地质研究所编著,《数学地质引论》,地质出版社,1977
- [8] 小乔治S. 科克等,《地质数据统计分析》上、下册,科学出版社,1978
- [9] 孙作为等,《物理化学》,地质出版社,1979

元素背景值计算的几个问题

——与李清同志商榷

蒋 志

李清同志在《元素背景值计算问题探讨》^[1]一文中,论述了柯氏检验、偏度与峰度检验在处理化探数据中的应用,无疑地具有实际意义,这里不再赘述。不过,就笔者所见,在该文的后半部份,还有值得商榷的地方。下面是笔者的一些浅见,提出来供讨论。

1. 元素含量所遵循的分布问题 该文指出了“同一批数据既符合算术正态分布,又符合对数正态分布时”^[1]的现象。原则上,随机数据所遵循的分布必须是图形一致或类似的分布。而正态分布与对数正态分布的图形在大多数场合下并不相似,只有在对数均值 m 较大、对数标准差 s 较小时的对数正态分布才趋近正态分布。此时,随机变量 x_i 的对数可表示为:

$$\begin{aligned}\ln x_i &= \ln(e^m + \Delta_i) \\ &= m + \ln(1 + \Delta_i e^{-m}) \\ &\approx m + \Delta_i e^{-m}\end{aligned}\quad (1)$$

式(1)的最后结果是因 m 较大、 s 较小时,只有 Δ_i 值较小时才有实际意义,而展开近似得到。从式(1)左端看,是随机变量 $\ln x_i$ 服从正态分布,即 x_i 服从对数正态分布;从式(1)的右端看,是变量 $\Delta_i = x_i - e^m$ 服从正态分布,即 x_i 服从正态分布。在理论上,式(1)所示的是 x_i 表面上服从对数正态分布,实质上服从正态分布,并非是 x_i 既服从正态分布,又服从对数正态分布。在实际上,这样的数据通常按正态分布处理,而不按对数正态分布处理。

2. 算术平均值和几何平均值问题 该文指出,元素含量平均值“一般按算术计算偏大,按对数计算偏小”,并解释为“由于算术平均值受边界含量影响较大,尤其样品数较少时,个别异常含量点的加入可使算术平均值明显偏高”^[1]。由于算术平均值 $\bar{x}$ 和几何平均值 $\bar{x}_g$ 的表示式分别为:

$$\bar{x} = (x_1 + x_2 + \cdots + x_n) / n \quad (2)$$

$$\bar{x}_g = \sqrt[n]{x_1 x_2 \cdots x_n} \quad (3)$$

而哥西 (Cauchy) 不等式^[2]的表示式可写为:

$$(x_1 + x_2 + \cdots + x_n)/n \geq \sqrt[n]{x_1 x_2 \cdots x_n} \quad (4)$$

式 (4) 在 $x_i \geq 0$ 的条件下成立, 并且, 只有当 x_i 全相等时, 式 (4) 的等号才成立。

因此, 算术平均值大于几何平均值是已证明了的数学结论, 而并非边界含量影响或个别异常含量加入的影响。

特别是在该文所用数据的情况下, x_i 不仅大于或等于零, 不全等, 而且服从一定的概率分布。其算术平均值和几何平均值可以直接由所遵循的理论分布函数求出。例如, 该文图 1 给出了第 17 号地层 Mn 元素含量分布数据, 对该数据重新处理后, 得到 Mn 含量 x 的如下理论分布函数 (图 1)。

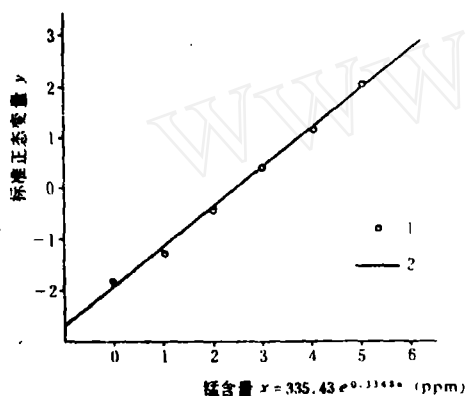


图 1 第 17 号地层锰含量分布

1—统计值^[1]; 2—式 (5) 计算值

$$F(x) = \Phi\left(\frac{\ln x - m}{s}\right) \quad (5)$$

式 (5) 中, $m = 6.646$, $s = 0.425$, $\Phi(y)$ 为如下概率积分

$$\Phi(y) = \int_{-\infty}^y \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \quad (6)$$

据式 (5), 随机变量 x 的算术平均值 $\bar{x}$ 、方差 σ^2 和几何平均值 $\bar{x}_g$ 可分别写为^[3]:

$$\bar{x} = e^{m + s^2/2} \quad (7)$$

$$\sigma^2 = e^{2m + s^2} (e^{s^2} - 1) \quad (8)$$

$$\bar{x}_g = e^m \quad (9)$$

应用前述 m 、 s 数据和式 (7)、(9), 可求得第 17 号地层中 Mn 元素含量的算术平均值和几何平均值如表 1。

第 17 号地层 Mn 含量平均值 表 1

计算方法	算术平均	几何平均	备 注
实际计算值	843	769	文献 [1] 中表 1
公式计算值	842	770	本文式 (7)、(9)

从表 1 看, 理论函数计算值与实际数据计算值, 完全一致, 进一步表明算术平均值和几何平均值的不同是由式 (6)、(7) 不同决定的, 并非边界品位和个别异常含量的影响。

3. 异常下限 T 的计算问题 该文提出了对数计算时的异常下限 T 公式 (为讨论方便, 本文改为自然对数) 如下:

$$\ln T = \ln \bar{x} + y \ln \sigma \quad (10)$$

并取 $y = 1.6$, 问题是式 (10) 中 $\bar{x}$ 和 σ 的意义不明确, 对 $\bar{x}$ 、 σ 同时取对数以及对 y 的取值也值得讨论。

如果式 (10) 中 $\bar{x}$ 是算术平均值, σ 是标准差, 则据式 (7)、(8)、(10) 可得:

$$\ln T = m + \frac{s^2}{2} + y \left[m + \frac{s^2}{2} + \frac{1}{2} \ln(e^{s^2} - 1) \right] \quad (11)$$

如果式 (10) 中 $\bar{x}$ 是几何平均值 $\bar{x}_g$, σ 是相对于 $\bar{x}_g$ 的“标准差” σ_g (这是极不严格的概念), 显然 σ_g 可写为:

$$\sigma_g = \sqrt{\sigma^2 + (\bar{x} - \bar{x}_g)^2} \quad (12)$$

由式 (8)、(9)、(10) 和 (12) 可得:

$$\ln T = m + y \left[m + \frac{1}{2} \ln(e^{2s^2} - 2e^{s^2/2} + 1) \right] \quad (13)$$

式 (11) 和 (13) 无论如何也不能从概率论上加以解释。

如果式 (10) 中 $\bar{x}$ 是几何平均值 $\bar{x}_g$, σ 定义为:

$$\sigma = e^s \quad (14)$$

由式 (9)、(10) 和 (14) 可得:

$$\ln T = m + y s \quad (15)$$

式 (15) 是可以从概率论上解释的, 但式 (14) 则不能称为“标准差”, 概率论中也没有形如式 (14) 的定义公式。

当然,也可用

$$T = \bar{x} + 2\sigma \quad (16)$$

和 $T = \bar{x}_s + 2\sigma_s \quad (17)$

计算异常下限。

用式 (11)、(13)、(15)、(16) 和 (17) 以及后面的式 (20) 到第17号地层Mn元素的含量数据上,得T计算值于表2。

表 2

公 式	$m = 6.646,$ $s = 0.425$	T 计算值 (ppm)
(11)	$y = 1.6$	1.106×10^7
(13)	$y = 1.6$	1.041×10^7
(15)	$y = 1.6$	1.519×10^7
(16)		1.591×10^7
(17)		1.533×10^7
(20)		1.800×10^7

从表2看,式(15)可能是该文公式的原意,但式(10)“标准差”的对数值含义不明。如果进一步引用该文求 y 的方程,即

$$\bar{x} + 2\sigma = e^{m+ys} \quad (18)$$

据式(7)、(8)和(18)可求得 y 是 s 的函数(图2)。

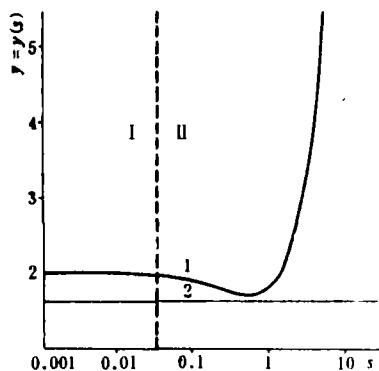


图 2 $y \sim s$ 的函数关系

I—对数正态趋近正态区; II—对数正态区;

1—式(19)计算值; 2— $y = 1.6$ [1]

$$y = \frac{s}{2} + \frac{1}{s} \ln(1 + 2\sqrt{e^{s^2} - 1}) \quad (19)$$

由图2和式(19)看,当 $s \rightarrow 0$ 时, $y \rightarrow 2$ 。这表明当对数正态分布趋近正态分布时, y 值为2,与该文所谓“模糊”区的 $y = 1.6$ 不同;当

$s \rightarrow \infty$ 时, $y \rightarrow \infty$,表明 y 不是在很小的范围内变化,而是在无穷区间上变化; y 的最小值约为1.7,大于1.6。应当强调的是,通常确定异常下界用 $\bar{x} + 2\sigma$ 是在正态随机变量的情况下。因此,对数正态变量的异常下界用 $\bar{x} + 2\sigma$ 估计并不合适,而应该用

$$\ln T = m + 2s \quad (20)$$

当然,式(20)中的系数2并非纯理论结果,而是通过理论与实际比较约定的。如果把2改为1.6,如式(15)那样,只要在经验上可行,也无不可。但这不能也不必用理论证明,只须直接根据经验即可。

综上所述,要强调的是:

第一,在实际处理元素背景含量所遵循的分布时,对单一分布来说,如果能用正态分布表示,就尽量用正态分布表示;对叠加分布来说,叠加的项数应尽量少,例如,通常为二项,至多三项即可。

第二,应用何种平均值,要视元素背景值分布所遵从的概率分布类型而定。如果遵从正态分布,则应该用其最可几值——算术平均值;如果遵从对数正态分布,则应该用其最可几值——几何平均值。因为最可几值被观测到的可能性最大,做为平均背景值较合适。

第三,异常含量下限的计算,如果元素含量是单一分布,正态数据时以 $\bar{x} + y\sigma$ 估计,对数正态数据时以 $m + ys$ 估计。 y 的数值可根据实际情况取3或2或小于2;如果是二项叠加分布,并大致可估计到第二项分布为异常分布,其所占的权与实际相当,则可根据统计决策规则[4]计算出异常下限值。

感谢梁厚锋工程师审阅手稿,并提出宝贵意见。

参考文献

- [1] 李清,地质与勘探,1983,12,42~47
- [2] Hardy, G. H., 不等式,科学出版社,1965,15~16
- [3] 王梓坤,概率论基础及其应用,科学出版社,1976,64~65
- [4] Helstrom, C. W., 信号检测的统计理论,上海科学技术出版社,1965,54~65