

43-47

拟合三参数对数正态分布时第三参数 c 的优选方法探讨

刘向瑞 刘刚

(中国地质大学·武汉·430074)

p: 62.5.1

A

在地质数据中,许多数据常呈正偏或负偏分布,但并不服从对数或反对数正态分布。这时可考虑它们是否服从三参数对数或反对数正态分布。本文给出一种估计第三参数 c 的方法,它可快速、客观、较优地确定此参数,弥补以往凭经验多次选择、反复试算等缺陷,为检验地质数据的统计分布型式,掌握分布规律,从而用于统计预测等提供了方便。

关键词 正偏 负偏 众数 三参数对数(反对数)正态分布

地质数据
优选法



图 1 数据图

对许多地质实测数据(记为 x),常希望通过统计分析来寻找其概率分布型式,然后再利用这种规律进行推断或预测。对于其直方图呈对称分布的数据,可以直接检验其是否服从正态分布。但实际地质数据是很复杂的,常可见到分布呈正偏(高峰偏左)或负偏(高峰偏右)的现象,其中又有一些不服从通常的对数或反对数正态分布。此时可考虑试用三参数对数或反对数正态分布来拟合。在这种三参数对数或反对数正态分布中,首先要确定第三参数 c ,然后才可估计出其它两个参数(均值和方差)。因此,参数 c 的估计起着重要的作用。 c 估计得好,数据 $\ln(x-c)$ 或 $\ln(c-x)$ 才可能转化为基本对称分布的型式。下面给出一种估计 c 的方法及其数学原理和应用实例。

1 三参数对数或反对数正态分布

设 X 为一随机变量,其观测数据为 x 。若 $\ln(x-c)$ (c 是某常数, $x > c$) 服从正态分布 $N(\mu, \sigma^2)$, 则表明 x 服从三参数 (μ, σ^2, c) 对数正态分布,其概率密度为

$$f(x) = \begin{cases} \frac{1}{\sqrt{2\pi\sigma(x-c)}} \exp\left\{-\frac{1}{2\sigma^2}[\ln(x-c) - \mu]^2\right\}, & x > c \\ 0, & x \leq c \end{cases}$$

若 $\ln(c-x)$ ($c > x$) 服从正态分布 $N(\mu, \sigma^2)$, 则表明 x 服从三参数 (μ, σ^2, c) 反对数正态分布,其概率密度为

$$f(x) = \begin{cases} \frac{1}{\sqrt{2\pi\sigma(c-x)}} \exp\left\{-\frac{1}{2\sigma^2}[\ln(c-x) - \mu]^2\right\}, & x < c \\ 0, & x \geq c \end{cases}$$

在拟合三参数对数或反对数正态分布时,首先要根据样本数据 x 合理地估计出第三参数 c 来,然后由数据 $\ln(x-c)$ 或 $\ln(c-x)$ 估计出 μ 和 σ^2 。最后,再检验 $\ln(x-c)$ 或 $\ln(c-x)$ 是否服从正态分布。

本文 1995 年 4 月收到,于地 2 编辑。

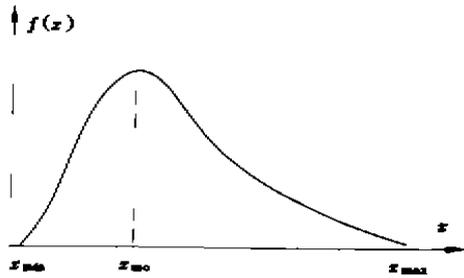


图 1

2 优选参数 c 的数学原理和计算公式

设数据 x 呈正偏分布, 其最小值为 x_{\min} , 最大值为 x_{\max} , 众数为 x_{mo} , 其频率密度曲线 $f(x)$ 如图 1 所示。如果 $c(x > c)$ 选得不当, 则 $\ln(x-c)$ 的频率密度 $f_1[\ln(x-c)]$ 可能还呈正偏, 或转为负偏, 如图 2(a, b) 所示。笔者认为, 当 c 为最优选择时, 应使 $\ln(x-c)(x > c)$ 呈近似对称分布, 这就基本上消除了正偏或负偏, 使峰值居于中央。设 $t = \ln(x-c)$, 此时 t 的频率密度曲线

$f_1(t)$ 应呈近似对称状, 见图 2(c)。这时 $\ln(x-c)$ 才最有可能服从正态分布。在这种峰值居中的条件下显然有下式成立

$$\ln(x_{mo} - c) = \frac{1}{2} [\ln(x_{\min} - c) + \ln(x_{\max} - c)] \quad (1)$$

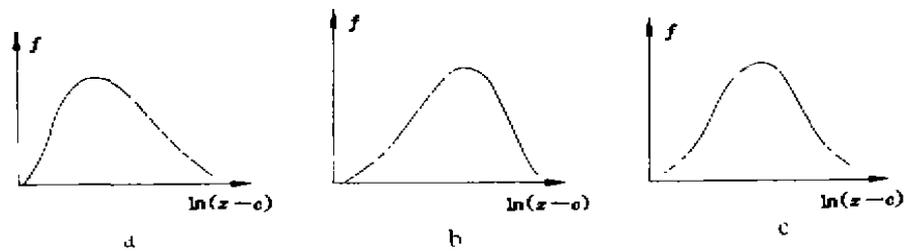


图 2

这是由于 x_{mo} 是数据 x 的众数, 故 $(x_{mo} - c)$ 必是 $(x - c)$ 的众数。又因 $\ln(x - c)$ 是单调增函数, 故 $\ln(x_{mo} - c)$ 必是 $\ln(x - c)$ 的众数, 就是说 $f_1[\ln(x_{mo} - c)]$ 是 $f_1(t)$ 的峰值。因为正态分布的均值、中位数和众数三者是合一的, 故若 $f_1(t)$ 是正态概率密度, 其众数就是其中位数, 即众数必定在最小数和最大数的中央, 即必有(1)式成立。

根据(1)式即可推导出求参数 c 的优选公式(2)。将(1)式变形可得

$$\begin{aligned} \ln(x_{mo} - c) &= \ln[(x_{\min} - c)(x_{\max} - c)]^{\frac{1}{2}}, \\ (x_{mo} - c) &= [(x_{\min} - c)(x_{\max} - c)]^{\frac{1}{2}}, \\ (x_{mo} - c)^2 &= (x_{\min} - c)(x_{\max} - c), \\ x_{mo}^2 - 2cx_{mo} &= x_{\min}x_{\max} - (x_{\min} + x_{\max})c, \text{ 故得} \\ c &= \frac{x_{mo}^2 - x_{\min}x_{\max}}{2x_{mo} - x_{\min} - x_{\max}}, \quad (2x_{mo} - x_{\min} - x_{\max} \neq 0) \end{aligned} \quad (2)$$

还可证明, 由(2)式确定的 c 具有以下性质:

(1) 当 $2x_{mo} - x_{\min} - x_{\max} < 0$ 时有 $c < x_{\min}$, 且

$$\ln(x_{mo} - c) = \frac{1}{2} [\ln(x_{\min} - c) + \ln(x_{\max} - c)]$$

此时, 因有 $x_{mo} < \frac{1}{2}(x_{\min} + x_{\max})$, 表明高峰偏左, 是属正偏, 应检验是否服从三参数对数正态分

布。

(2) 当 $2x_{m_0} - x_{min} - x_{max} > 0$ 时有 $c > x_{max}$, 且

$$\ln(c - x_{m_0}) = \frac{1}{2} [\ln(c - x_{min}) + \ln(c - x_{max})] \quad (3)$$

此时, 因有 $x_{m_0} > \frac{1}{2}(x_{min} + x_{max})$, 表明高峰偏右, 是属负偏, 应检验是否服从三参数反对数正态分布。

综上所述, 由(2)式确定的 c 确实可使 $\ln(x-c)$ 或 $\ln(c-x)$ 的峰值居中, 呈近似对称分布。至于是否服从正态分布, 还须通过拟合优度检验。当检验通过时, 才可确信原始数据是服从三参数对数或反对数正态分布的。

3 众数 x_{m_0} 与最大数 x_{max} 的选取方法

由(2)式可知, c 的确定完全依赖于 x_{m_0} , x_{min} 和 x_{max} 的选取。由于 x 往往是正值数据, 故 x_{min} 的下限为零, 变化范围不大, 可就取数据中最小的数, 或比它略小较为整齐的数。但是 x_{m_0} 和 x_{max} 的变化则无限制了, 它们选取的好坏对 c 值的影响很大。因此, 有必要研究如何合理地确定 x_{m_0} 与 x_{max} 。根据经验, 可以从数据 x 的直方图出发, 采用几何方法求得 x_{m_0} 与 x_{max} 计算公式。设数据 x 的方图如图 3 所示。

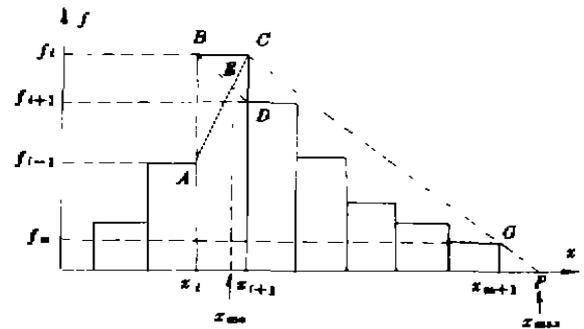


图 3

(1) x_{m_0} 的确定

选取频率最大(记为 f_i)的区间 $[x_i, x_{i+1}]$ 上的直方形与左、右相邻的两个直方形。设 AC 与 BD 交于 E 点, 则 E 点的横坐标即为 x_{m_0} 。用公式写出即为

$$x_{m_0} = x_i + \frac{AB \cdot BC}{AB + CD} = x_i + \frac{(f_i - f_{i-1})(x_{i-1} - x_i)}{2f_i - f_{i-1} - f_{i+1}} \quad (4)$$

式中, f_{i-1} 和 f_{i+1} 分别是 f_i 左、右两个组的频率。

(2) x_{max} 的确定

由于数据有随机性, 一般要取比实测数据中最大的数据再大一些的数作为 x_{max} , 但大到什么程度则未有规定。可以借助直方图(图 3)来确定 x_{max} , 即将直方图中最大频率的直方形的右上角顶与最右端一个直方形的右上角顶相连接, 即连 CG , 并延长之使与 x 轴交于 P 点, P 点的横坐标即可取为 x_{max} 。用公式写出即为

$$x_{max} = x_{n+1} + \frac{f_m(x_{m-1} - x_{i+1})}{f_i - f_m} \quad (5)$$

式中, x_{n+1} 是最右端一个组的上限; f_m 是最右端一个组的频率。

4 实例及效果

例 1 有一批数据 x (109 个), 其直方图(图 4(a))呈正偏。取自自然对数 $\ln x$ 后, 其直方图(图 4(b))又略呈负偏。用上法先估算出 $x_{min} = 2.5$, $x_{m_0} = 130.839$, $x_{max} = 36.895$, $c = -17.365$ 。

从 $c < x_{\min}$ 可知原数据分布是正偏的。于是,取 $\ln(x-c)$,其直方图(图 4(c))基本对称。经用 χ^2 拟合优度检验, $\chi^2 = 1.1459 < \chi_{0.10}^2(5) = 9.236$, 接受 $\ln(x-c)$ 为正态分布的假设。可以认为原始数据 x 服从 $\mu = 4.069, \sigma^2 = 0.134, c = -17.365$ 的三参数对数正态分布。

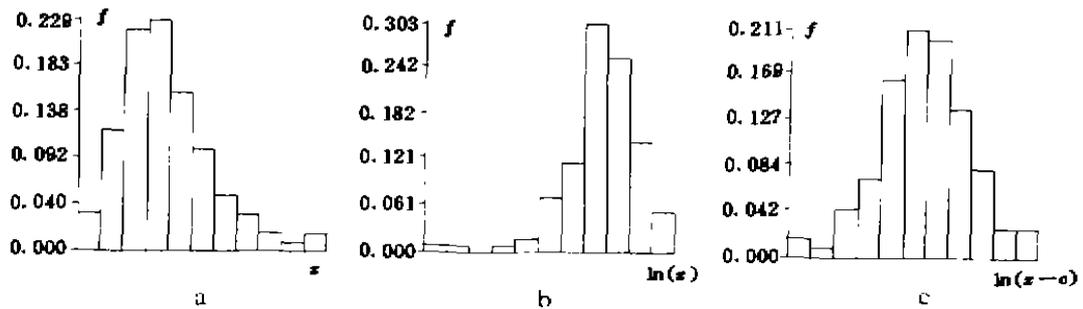


图 4

例 2 有一批数据 x (109 个),其直方图(图 5(a))呈负偏。取 $a = x_{\max} = 37$,再取 $\ln(37-x)$,其直方图(图 5(b))仍略呈负偏。用上法先估算出 $x_{\min} = 2.00, x_{\max} = 37.00, x_{\text{mo}} = 24.5, c = 52.626$ 。从 $c > x_{\max}$ 可知原数据分布是负偏的。于是,取 $\ln(c-x)$,其直方图(图 5(c))基本对称。经用 χ^2 拟合优度检验,得 $\chi^2 = 1.0067 < \chi_{0.10}^2(5) = 6.064$, 接受 $\ln(c-x)$ 为正态分布的假设。故可认为原始数据 x 服从 $\mu = 3.375, \sigma^2 = 0.044, c = 52.626$ 的三参数反对数正态分布。

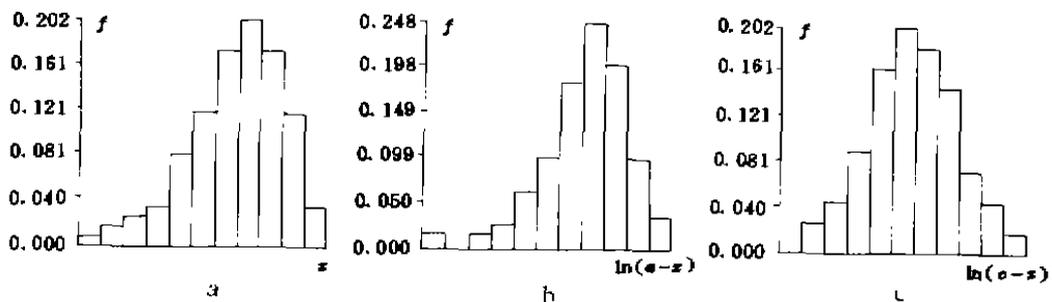


图 5

笔者已编出计算机程序,上机后,不仅很快就算出 c 值来,而且也可以打印出前后形成的直方图供比较和参考。经验证明,由于第三参数 c 可起到调节分布偏度的作用。故经 c 的调节,分布变得更加对称,因而也更接近正态分布。若经本法优选的 c 调节后仍不服从正态分布,则必是在分布的峰度上与正态分布差别较大。

此法对直方图呈单峰且各组频率起伏不大,连续性较好的数据效果较佳,否则效果差些。另外,此法的结果在一定程度上还有赖于作直方图方法的合理性。此法可广泛用于各类地质、物化探、遥感数据的处理。

参考文献

程光华, 蒋耀松, 张一斌. 概率统计. 北京: 地质出版社, 1980: 176~180.

Discussion on Optimal Estimation Method of the Third Parameter c When Fitting a Lognormal Distribution with Three Parameters

Liu Dianrui Liu Gang

Many geological data often show positive (or negative) skew distributions but do not follow lognormal (or inverse lognormal) distributions. In this case it is possible to consider whether they follow the lognormal (or inverse lognormal) distribution with three parameters. This paper proposes a new method for the optimal estimation of the third parameter c of these distributions. This method can estimate the third parameter c rapidly, objectively and optimally, so it is helpful to the statistical distribution testing and the statistical prediction of mineral resources.

Key words: positive skew, negative skew, mode, lognormal (inverse lognormal) distribution with three parameters

X 射线荧光分析钼原矿技术在钼矿山投入应用

栾川钼矿虽然储量大、品位高,但品位分布极不均匀,给矿山采矿作业带来了一定的难度,经济效益也受到了影响。所以在加强地质工作的同时,还采用新技术——X 射线荧光仪测试手段,圈定贫富矿段,快速正确的指导生产,取得了良好的经济效益。

1 X 射线荧光仪分析方法

X 射线荧光分析技术是一种核地球物理勘探方法,它用轻便的 X 射线荧光仪,在探槽、坑道、采场等山地工程的岩墙表面或矿体露头上,直接测定矿石在放射性同位素源照射下产生的特征 X 射线强度,并据其测量结果来确定矿石品位,具有快速、准确等特点。

2 X 射线荧光分析技术在生产中的应用

通过与成都地质学院合作,开发了 X 射线荧光分析钼品位的技术。该方法选用 30mm 作为仪器的源样距,按 $8 \times 8\text{cm}^2$ 测网进行测定,取一个地质原样的全部测点的“特散比”平均值作为参量,用“特散比”法计算钼品位。1992~1993 年,在室内测定粉样 2000 多个,节约化学分析费用 4 万余元(化学分析样每个 20 元左右,X 射线荧光分析仅 1 元左

右)。1994 年,在井下坑壁上测定钼品位 500 多个,节约经费 2 万元(刻槽取样及化学分析每个样 40 元左右)。目前对块样的研究测定也有了很大进展,这对出矿管理,及时提供出矿品位,起到了积极的作用。

近年来,此方法正式投入了生产,据统计,在所属各矿井采矿场和掘进工程中,应用 X 射线荧光分析技术测定钼品位 3000 多个,正确地指挥了生产,起到了先锋作用。例如在 1530 中段一坑道壁上,进行了 X 射线荧光分析测定后,确定了该处钼品位很低,无开采价值,马上停止了掘进;在 1801 工区 1450 中段采场中,进行了 X 射线荧光分析测定后,准确圈定出矿石与废石的界线,正确指挥了出矿作业,提高了矿石品位,降低了矿石贫化率。

总之,应用 X 射线荧光分析测定技术,不仅具有及时、准确、无损、见效快等优点,并且能节省大量的人力、物力,对指导生产,增加经济效益,起到了先锋作用,并对减少粉尘,改善劳动条件,更是其他方法难以比拟的,我们相信,在不久的将来,一定能代替传统的刻槽取样和化学分析方法,使矿山经济效益得到更大的提高。

(栾川栾川铝业公司 杨森林)