

金矿品位的对数正态分布

杨尔煦



金矿中特高品位样品的处理问题,多年来在金矿储量计算中是一个有争论的问题。有的矿区规定,凡是大于平均品位12~15倍的特高品位要删去,有的则按格拉布斯公式进行取舍。这些处理方法的共同点就是要把特高品位删去,这主要是由于采用算术平均值来计算矿块品位所引起的。一个特高品位的影响要相当于20~30个一般品位的数据。如果不删去,将使平均品位增加1~2倍。一般认为这是一种偶然现象,不能代表矿体的真实情况,就将它删去了。简单的推理看来好象是正确的,因此这种处理方法也就沿用了很多年。

当采用算术平均值时,特高品位的处理问题是无法解决的,只有删去了之。其实,对这个问题再深入一步,就要研究采用算术平均值的办法是否正确。由概率论得知,采用算术平均值必须有一个先决条件,即所观测的量必需符合正态分布(高斯分布)。反之,若不管所观测的量是什么分布,统统采用算术平均值是不妥的。例如,一些铁矿品位的分布大体上符合正态分布,可采用算术平均值。而金矿品位或厘米·克/吨值却大大偏离正态分布,如玲珑金矿52号脉的厘米·克/吨分布(图1a),并不符合正态分布,采用算术平均值计算金矿矿块品位是不恰当的。

能否使用算术平均值来计算金矿储量的问题,在南非金矿地质学家中争论得最早^[1]。在本世纪20~40年代中,人们发现算术平均值往往比实际开采的结果要高,但说不清原因何在。当时也有几位地质学家写文章,反对放弃算术平均值的方法。由于找不到一种代替算术平均值的新方法,因此,在20~40年代南非金矿仍然普遍使用算术平均值法。直到1947年西歇尔(Sichel)发表了金矿品位分布的类型接近于对数正态分布

以后,就开始采用几何平均值推出平均值的估值来代替算术平均值。克立格(Krige)1960年发表了南非24个大金矿的分布,不管金或是铀都符合三参数对数正态分布。从这以后,地质统计方法也得到了迅速的发展,并创立了地质统计学^[2]。几何平均值不存在特高品位删去的问题,当只有一个母体时,所出现的特高品位也是整个统计母体中的一员,当然应该参与计算。而且,若某一金矿中的特高品位出现较多时,正好说明这个金矿可能是个富矿。如果在金矿勘探初期,打了10个钻孔,其中有一个钻孔出现了特高品位,这就意味着这个金矿是大有希望的,因此特高品位在评价金矿时是一个重要标志。

现在已经是80年代了,我国即将出版的有关岩金和砂金矿地质勘探规范仍然规定采用算术平均值法,并提出对特高品位要进行处理。这二点显然是值得商榷的。我们应该吸收南非金矿的经验,改变采用算术平均值的作法,采用几何平均值推出平均值估值来估算矿块品位。本文的目的也是想介绍和应用南非金矿的经验,并与金矿地质工作者们共同研究这个问题,使我们对金矿的评价和勘探工作在认识上来一个改革。

一、金矿对数正态分布的一例

这里介绍的是玲珑金矿52号脉的一个中段,在中段下面有14个钻孔。在中段的沿脉内每隔4米有一个横切含金石英脉的水平刻槽样,共54个样品。为了使钻孔样的分析结果和坑道样的分析结果可以相比,对钻孔样品都计算了它们到水平面上的矿体投影厚度,这样14个钻孔有14个样品。然后求出每个样品的品位和厚度的乘积 Z ,单位用厘米·克/吨表示。用68个子样来研究一个母体,显然是少了一点,但已经可以看出金矿的特征了。

图1a所示为用普通数作横坐标的直方图,表明52号脉的厘米·克/吨分布形态是极不对称的,在右侧有一个拖长的尾巴,说明高品位和特

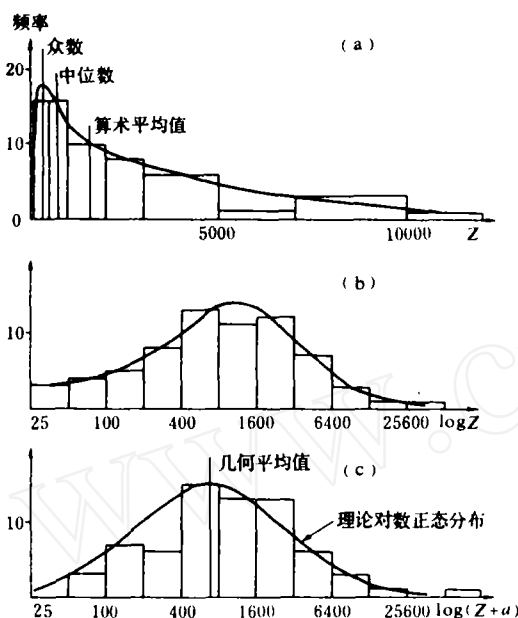


图1 52号脉品位和厚度乘积Z的分布曲线

高品位所占的量不可忽视,必需在计算中统一考虑。由于52号脉分布的这种极不对称性,使得分布的众数,即频率最大的值,位于中位数及算术平均值的左边,是三者中最小的数值。其次是中位数,即累计频率为50%的数值。在对数正态分布中,中位数就是几何平均值。最大的是算术平均值,位于最右边。这就是对数正态分布曲线的特征,它使众数、中位数和算术平均值三者分离了。这和正态分布时的情况大不一样,在正态分布时,其众数、中位数和算术平均值都是重合的。

当我们用对数方法来绘制分布曲线时,同样的68个数据得出了图1 b的曲线。这个曲线初看起来大致是对称的,但仔细看时,对数分布曲线右边陡而左边缓,形成了负歪斜的分布曲线。

为了纠正这个负歪斜的形态,在Z值后面加一个常数a,横坐标变成了 $\log(Z+a)$, (a的求法在后面再讨论)。做了这样一个修改以后,分布曲线就变为比较对称了(图1 c),与理论正态分布曲线基本拟合。这样就形成了1951年克立格发表的三参数对数正态分布曲线。数学表达式如下:

$$f(Z) = [\sigma\sqrt{2\pi}]^{-1} \cdot \exp\left[-\frac{1}{2}\sigma^{-2}(\log_e(Z+a)-\xi)^2\right]$$

上式 $f(Z)$ 为分布的密度函数,

ξ 为 $\log_e(Z+a)$ 的平均值,即 $(Z+a)$ 的几何平均值,

σ^2 为 $\log_e(Z+a)$ 的方差,即 $(Z+a)$ 的对数方差,a为第三参数,它因矿体而变化。

用这个理论的三参数对数正态分布曲线来拟合52号脉,得到的三个参数值如下:

$$\xi = \log_e 700 = 6.5511$$

$$a = 50 \text{ 厘米} \cdot \text{克/吨}$$

$$\sigma^2 = 1.7$$

这三个参数可以比较完整而又简单地来表征一个矿体的基本特点。在这里要注意,我们所统计的68个子样是无选择地把这条含金石英脉的所有样品都放进去,既有够边界品位的,也有不够边界品位的数据,因此这里的 ξ 是所有 $(Z+a)$ 值的几何平均值, σ^2 表示 $(Z+a)$ 值的变化程度。a是随不同矿体而变化的,也表示矿体的一个特征,但其物理意义尚不清楚。用这三个参数来表示一个矿体的特征,要比目前规范中常用的算术平均品位、品位变化系数和厚度变化系数更科学一些。

二、对数正态分布的检验

是否属于对数正态分布,可以用对数概率纸来检验。把 $(Z+a)$ 和累计频率对应的值展在对数概率纸上,当属于对数正态分布时就成一条直线。如图2所示,最下面一条就是52号脉原始的Z值和累计频率的曲线。该曲线的中部基本上是一条直线,但其下部偏离了直线的位置。这时要加第三参数a,这里选择了二个数字50和100, $(Z+a)$ 值曲线的尾部马上向上升高。当选择适当的第三参数a就可以使曲线的下部和直线吻合。这就是求参数a的图解法。

图中直线AB是与52号脉拟合最佳的对数正态分布,根据AB直线求出52号脉的其他参数。AB直线和50%累计频率的交点处的 $(Z+a)$ 值,就是 $(Z+a)$ 值的中位数,即几何平均值 e^ξ 为700厘米·克/吨。

通过对数概率纸的中心点C,作一平行于AB的直线CD。这CD直线右上角所表示的刻度就是该分布的对数方差 σ^2 (以e为底)。一个分

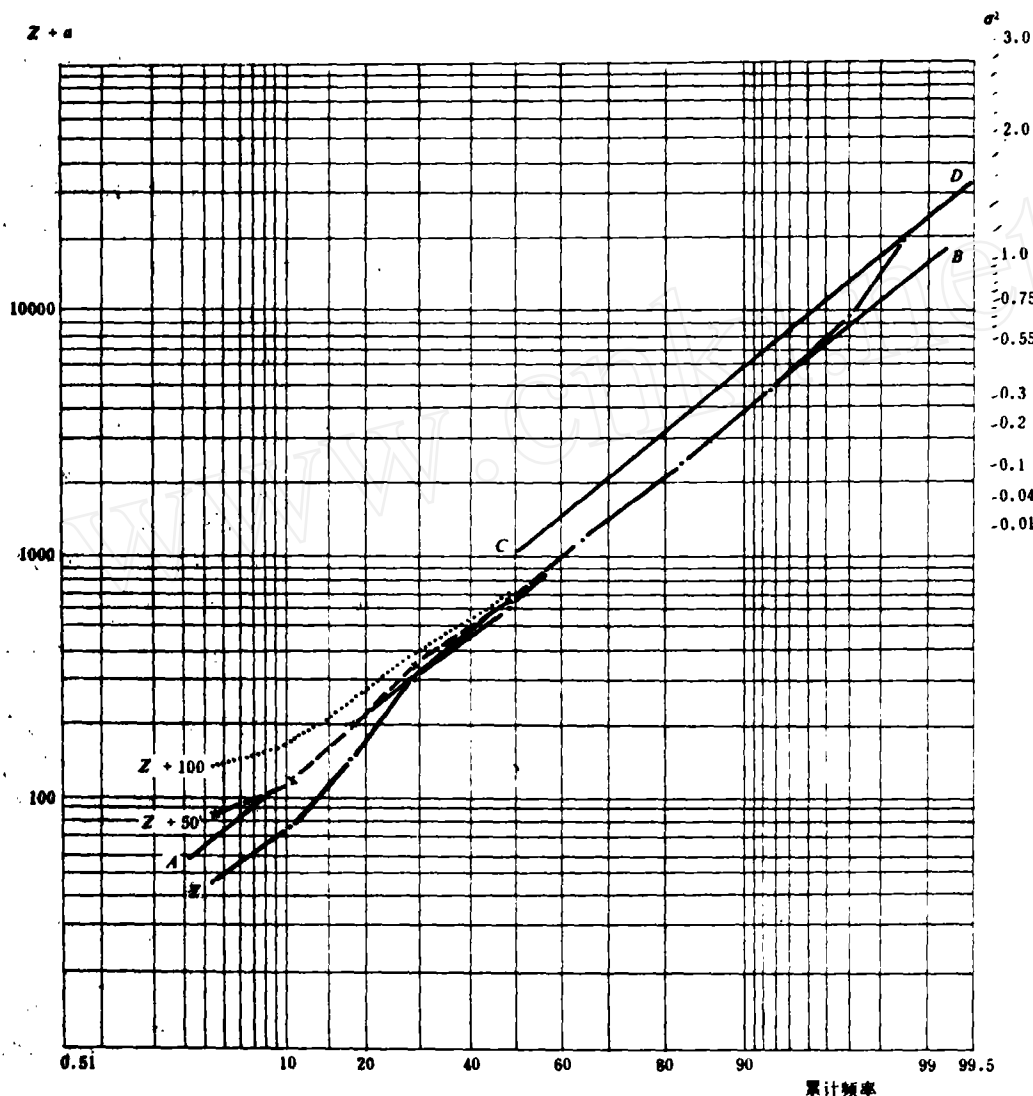


图2 对数正态分布检验图

布的对数方差 σ^2 决定于在对数概率纸上所绘直线的斜率。那么52号脉的 σ^2 为 1.7，符合一般金矿的 σ^2 在 0.1 ~ 3.0 之间的情况。

对数正态分布的检验方法简单易行。关键在于所取的样品是否存在偏差，取样的方法是否正确。例如在用钢粒钻探时，取出的矿芯，尤其是对硬脆的含金石英脉，往往是不完整的。矿芯往往顺着小的裂隙面破碎了，而金的矿物正好在这些面上磨成粉末。这种情况下，钢粒钻探即使符合国家规定的矿芯采取率（在75%以上），也会使钻探所获得的品位值比真正品位偏小。这种取样方法本身引起的偏差，任何统计方法也难予弥

补。因此，为了使所取样品可靠，在金矿区的勘探中要强调必须采用金刚石钻探，以保证取样的质量。此外，用岩芯劈开机劈分岩芯，对金矿来讲也会带来偏差。最好用金刚砂切片机切开。这些类似的问题对铁矿来说可能无所谓，但对金矿来讲，采样、加工和化验的每一步都要进行认真的分析，使所有的原始数据 Z 都是可靠的。

在取样方法正确的前提下，地质统计学还要求每个样品的规格都是一样的，分布是均匀的，这样三个参数才能是正确的。

前面所举的52号脉，严格来讲是不合要求的。中段里取样是沿脉每隔4米一个槽，刻槽的规格

是 10×3 厘米², 钻孔中的取样面积大体上是15厘米², 相距100米左右。显然坑内的54个样品代表一种类型, 钻孔中的14个样品又代表另一种类型。这二种样品大小不相等, 分布也不均匀, 混合起来显然会引起偏差, 再加上样品的数量太少, 用这样的样品虽然也获得了一条直线, 但它不一定能代表母体的特征, 即矿体固有的特征。

真正的分布曲线, 要从矿山的开拓工程中才能求出。这要在沿脉、穿脉和上山中按一定的间距进行系统刻槽取样或其他方法的取样。在一个矿山中通过多年的工作积累才能求出分布曲线, 求出各条矿脉的参数, 从中找出规律。

在勘探阶段一般来讲是很难取得可靠的分布, 因为它的工程很有限, 这时可利用已知矿体的规律来进行推断。另一种情况是没有已知矿体而是处于勘探阶段, 往往只打了一些钻孔和少量的坑道。这些样品的数量较少, 用这些少量样品来推断母体是一件困难的事, 这要采用统计学中的小子样推断理论来解决。

三、对数正态分布中的一些数学关系

对数正态分布的函数表达式(1)实际上是从正态分布函数套过来的。只是用 $\log_e(Z+a)$ 代替了 Z 。根据密度函数 $f(Z)$ 可知

$$\begin{aligned} & \int_{-\infty}^{\infty} f(Z) d\log_e(Z+a) \\ &= \int_0^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \frac{1}{(Z+a)} \exp\left[-\frac{1}{2}\sigma^{-2}(\log_e(Z+a)-\xi)^2\right] dZ \\ &= \int_0^{\infty} F(Z) dZ \\ &= 1 \end{aligned} \quad (2)$$

为了简化, 改变一下变数。

$$\text{设 } W = \frac{1}{\sigma}(\log_e(Z+a) - \xi) \quad (3)$$

$$dW = \frac{1}{\sigma(Z+a)} dZ$$

$$Z = \exp(W\sigma + \xi) - a$$

将(3)式代入(2)式得

2. 总体的算术平均值 $\bar{Z}$

$$\begin{aligned} \bar{Z} &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} Z \exp\left(-\frac{W^2}{2}\right) dW \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} (e^{W\sigma + \xi} - a) \exp\left(-\frac{W^2}{2}\right) dW \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{W^2}{2} + W\sigma + \xi\right] dW - a \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(W-\sigma)^2}{2} + \frac{1}{2}\sigma^2 + \xi\right] dW - a \\ &= \exp\left(\frac{1}{2}\sigma^2 + \xi\right) - a \end{aligned} \quad (6)$$

3. 总体的众数 Z_{mo}

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}W^2\right) dW = 1 \quad (4)$$

(4) 式就是正态分布函数。

现在利用(2)式可以求出对数正态分布的各种特征值。

1. 总体的几何平均值 Z_g :

$$\because \log_e(Z_g + a) = \xi$$

$$\therefore Z_g = e^{\xi} - a \quad (5)$$

求 $F(Z) = f(Z) \cdot \frac{1}{Z+a}$ 的极大值, 即 $F'(Z) = 0$ 。

$$F'(Z) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\sigma^{-2}(\log_e(Z+a) - \xi)^2\right] \cdot \left\{ \frac{-(\log_e(Z+a) - \xi)}{(Z+a)^2\sigma^2} - \frac{1}{(Z+a)^2} \right\} = 0$$

$$\therefore \log_e(Z_{m0} + a) - \xi = -\sigma^2$$

$$Z_{m0} = \exp(\xi - \sigma^2) - a \quad (7)$$

4. $\log_e(Z+a)$ 的方差 σ^2

5. $Z+a$ 的方差 s^2

$$\begin{aligned} s^2 &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \left[(Z+a - \exp(\xi + \frac{\sigma^2}{2}))^2 \right] e^{-\frac{W^2}{2}} dW \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{2\xi} (e^{W\sigma} - e^{\frac{\sigma^2}{2}})^2 e^{-\frac{W^2}{2}} dW \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{2\xi} (e^{2W\sigma} - 2e^{W\sigma + \frac{1}{2}\sigma^2} + e^{\sigma^2}) e^{-\frac{W^2}{2}} dW \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{2\xi} \exp\left[-\frac{1}{2}(W - 2\sigma)^2 + 2\sigma^2\right] dW - \\ &\quad - 2 \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{2\xi} \exp\left[-\frac{1}{2}(W - \sigma)^2 + \sigma^2\right] dW + \\ &\quad + e^{2\xi + \sigma^2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{W^2}{2}\right) dW \\ &= e^{2\xi + 2\sigma^2} - e^{2\xi + \sigma^2} \\ &= e^{2\xi + \sigma^2} (e^{\sigma^2} - 1) \\ &= (\bar{Z} + a)^2 (e^{\sigma^2} - 1) \end{aligned} \quad (8)$$

6. 边界品位 Z_i 以上 Z 的平均值估值 Z_i^*

$$\begin{aligned} Z_i^* &= \frac{\int_{W_i}^{\infty} \frac{1}{\sqrt{2\pi}} Z \exp\left(-\frac{W^2}{2}\right) dW}{\int_{W_i}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{W^2}{2}\right) dW} \\ &= \frac{\int_{W_i}^{\infty} \frac{1}{\sqrt{2\pi}} (e^{W\sigma + \xi} - a) \exp\left(-\frac{W^2}{2}\right) dW}{\int_{W_i}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{W^2}{2}\right) dW} \\ &= e^{\xi + \frac{1}{2}\sigma^2} \cdot \frac{\int_{W_i}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}W^2\right) dW}{\int_{W_i}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}W^2\right) dW} - a \end{aligned} \quad (9)$$

其中 $W_i = \frac{1}{\sigma} [\log_e(Z_i + a) - \xi]$

7. 边界品位 Z_i 以上 $\log_e(Z+a)$ 的平均值 $[\log_e(Z+a)]_i^*$ 和几何平均值估值 $Z_{i,g}^*$

$$\begin{aligned}
[\log_e(Z+a)]_i^* &= \frac{\int_{w_i}^{\infty} \frac{1}{\sqrt{2\pi}} \log_e(Z+a) \exp\left(-\frac{W^2}{2}\right) dW}{\int_{w_i}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{W^2}{2}\right) dW} \\
&= \frac{\int_{w_i}^{\infty} \frac{1}{\sqrt{2\pi}} (W\sigma + \xi) \exp\left(-\frac{W^2}{2}\right) dW}{\int_{w_i}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{W^2}{2}\right) dW} \\
&= \frac{\int_{w_i}^{\infty} \frac{\sigma}{\sqrt{2\pi}} \exp\left(-\frac{W^2}{2}\right) d\left(\frac{W^2}{2}\right)}{\int_{w_i}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{W^2}{2}\right) dW} + \xi \\
&= \frac{\frac{\sigma}{\sqrt{2\pi}} \exp\left(-\frac{W_i^2}{2}\right)}{\int_{w_i}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{W^2}{2}\right) dW} + \xi \\
&= \xi^*
\end{aligned}$$

$$\therefore Z_{ig}^* = e^{\xi^*} - a \quad (10)$$

根据52号脉的参数 $\xi = 6.5511$, $\sigma^2 = 1.7$, $a = 50$ 可以求出该脉的其他数据:

1. 总体的几何平均值 $Z_g = e^{\xi} - a$ 为 650 厘米·克/吨。
2. 总体的众数 $Z_{mo} = e^{\xi - \sigma^2} - a$ 为 78 厘米·克/吨。
3. 总体的算术平均值 $\bar{Z} = e^{\xi + \frac{1}{2}\sigma^2} - a$ 为 1587 厘米·克/吨。
4. 按边界品位 2 克/吨, 几何平均厚度 1.00 米, 计算 $Z_i = 200$, $W_i = -0.7897$
5. 边界品位以上的平均值估值 Z_i^* 为 1996 厘米·克/吨, 以几何平均值厚度除, 边界品位以上的平均品位估值为 19.96 克/吨。
6. 边界品位以上的平均 $\log_e(Z+a)$ 为 7.4559, 求出边界品位以上的几何平均值估值 Z_{ig}^* 为 1087 厘米·克/吨, 几何平均品位为 10.87 克/吨。

52号脉地质报告书中所述算术平均品位为 21.57 克/吨, 算术平均厚度为 1.81 米, 和上面用拟合曲线所推算出的平均值估值 19.96 克/吨相

仿。但报告书中算术平均厚度为 1.81 米和几何平均厚度 1.00 米相差很多, 这也是由于有一个特大厚度 15.19 米所影响。根据以上计算, 都使用几何平均值估值时, 品位是 10.87 克/吨, 厚度为 1.00 米。说明算术平均值还是偏高了, 几何平均值又偏低了, 南非的经验是采用几何平均值推出的平均值估值, 即 19.96 克/吨。52号脉的真实情况要等以后开采中加以证实。这里只是当作一个例子算一下。

以上的例子, 由于样品数量少, 分布不均匀, 样品规格也不一致, 因此在计算中不可避免的带来很多误差。但金矿对数正态分布的规律基本上是对的, 这就是处理金矿数据的出发点, 即利用几何平均值的稳定性来推算平均值的估值。

参 考 文 献

- [1] D. G. Krige, 用于矿产评价的对数正态地质统计学, 南非矿冶研究院丛书, 1978
- [2] A. G. 儒尔奈耳, CH. J. 尤日布雷茨, 矿业地质统计学, 冶金工业出版社, 1982