

数 学 地 质

论 文 汇 编

成 都 地 质 学 院

一九八一年一月

目 录

一、地质变量的分布及其应用·····	杨 龙 王方戎 朱章森 (1)
二、关于矿床条件模拟基本的理论和程序设计的初步探讨 ·····	王柏钧 周晓茹 (10)
三、亲合度与自组迭代分析·····	王柏钧 胡远来 (27)
四、聚类分析的自动作图程序·····	胡远来 (37)
五、攀枝花——西昌地区基性超基性岩体含矿性数学地质研究 ·····	成都地质学院数学地质研究室 川西攀枝花地质综合地质队 (41)
六、用数学地质方法判定南岭花岗岩含矿性·····	朱章森 姜若维 叶韵琴 (55)
七、数学地质在酒西盆地北部单斜带找油中的应用 ·····	成都地质学院 陈立官 程光瑛 李鸿智 玉门石油管理局研究院勘探室 (62) 张印斋
八、关于酸化压裂层段的定量预测 ·····	成都地质学院 程光瑛 陈立官 (70) 四川石油管理局井下工艺所 熊广燕 王国光
九、关于地层对比的数学模式·····	胡远来 王柏钧 徐仲平 (80)
十、东南沿海花岗岩地质年代的判别分析·····	朱章森 卢玉蓉 叶韵琴 (89)
十一、地下河流域岩溶水位的趋势分析·····	卫 迦 (97)
十二、遂南气田香溪群二段储油物性的多元统计分析 ·····	成都地质学院 陈立官 (102) 四川石油管理局川中矿区 陈立平
十三、现场破裂压力的多元逐步回归 ·····	成都地质学院采收率研究室 (108) 四川石油管理局井下工艺所

地质变量的分布及其应用

杨龙 王方戎 朱章森

数学地质作为一门把地质模型用概率模型来描述、检验和解释的科学,已为传统地质问题,如各种地质分类、对比、推断、预测、成因等的研究,提供了一套新的工具和严格的思维方法,颇受人们的重视,同时也取得了不少的进展。但作为数学地质的基础工作,地质变量的特征、分布及其变化性的研究,在国内除少数人员比较重视外,大多数人都有忽视的倾向。他们忽略了产生数学模型的条件,不了解,比如“地质统计学”的产生与发展就是以研究地质变量的空间相关性开始,通过它的分布的讨论,而使线性与非线性的理论趋于完善的。如不及时扭转这一现象,将会有碍于对现有数学地质模型的有效应用和进一步发展。本文仅从地质变量的分布出发,谈谈如何更好地对现有数学地质模型的有效应用,及其有关的正态性检验和转换的一些常用而有效的方法。

一、地质变量的分布

由于地质成因的复杂性和多样性,地质变量对于它的每一个个体,总是会表现出变异或差异的特点,所以人们总是把它看成随机变量。随机变量的所有观测值的数学模型是由它的理论分布来描述的,知道了地质变量的理论分布,就能提出相应的统计方法,对此作出数学处理。所以在应用一些现有的统计方法时,千万不能忘记了这些地质变量的理论分布,否则就可能产生模型的不适应性。

地质变量的理论分布很多,如玻璃沙中的石英分析和铬铁矿床中的铬元素分析都可能是负偏斜分布,而金元素化验又可能是对数正态分布。在基性超基性岩体中,主要造岩氧化物如 SiO_2 、 Al_2O_3 、 MnO 、 MgO 、 CaO 、 NiO 的分布可能是正态分布。而组成基性岩的次要成分,如 TiO_2 、 Na_2O 、 K_2O 和成矿元素 Cr_2O_3 又可能是对数正态分布等等。但是,在我们所要研究的地质变量中,它们的理论分布往往预先并不知道,或者不容易知道。这样,一般就借助于该地质变量观察值的经验分布来确定了。应当指出,用此办法是很难唯一地确定出其理论分布的。正如要从对称、钟形的经验频率曲线(如图一)区分出它所对应的理论频率分布是正态分布 $f(X) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2}$ 或柯西分布 $f(X) = K/[K^2 + (X - K_0)^2]/\pi$ (如图二)一样困难。好在多数地质变量的理论

分布都是正态或接近于正态的，而且正态分布在数学上便于处理，有一套较为成熟的处理方法。特别是很多统计方法的线性模型都是在正态的假定下推出来的。在此假设下，如果地质变量不服从正态分布，就会直接给模型计算的效果带来影响。因此，在用此种模型时，我们总是先假定我们所研究的地质变量的理论分布是正态的，然后通过它们的观察所的经验分布的检验，看是否与理论分布的一致性作为此假定成立与否的依据。如果原观察值不是正态的，我们还可以通过转换函数把每个观察值变换成服从正态分布的观察值，使与模型的假设条件相符合。这里对如何正确地判断观察数据是否服从正态分布就显得特别重要了。

二、正态性检验

正态性检验尽管很重要，但方法并不十分完备。比如三元以上的正态性检验，就还没有一个一般的方法，就是二元的检验也极为复杂，只有单变量的正态性检验，方法才比较多，而且也较成熟。这里分别介绍在电子计算机上已经实现的分别适用于大样本和小样本的两种单变量的简便方法。

1. 拟合检验与累积频数检验

分布函数的拟合检验和累积频数检验都是适用于大样本和连续分布函数的检验。

设 ξ 的分布为一正态分布函数 $F(X)$ ，且参数 μ 、 σ 已知，对 ξ 进行 N 次独立观测，得到 ξ 的 N 个观测值 $\xi_1, \xi_2, \dots, \xi_N$ 组成一观测子样。现将通过观测子样的经验分布与理论正态分布的差异是否显著，来判断 ξ 是否为正态分布，其方法有二。

(1) 拟合检验

以比较经验频率和理论频率的差异是否显著为标准的检验法，称为分布函数的拟合检验。其方法的一般步骤如下：

1) 将 ξ 的观测数据的取值分为 K 个区间，求出每个区间上的频率，并将区间 $[a_{k-1}, a_k]$ 上的频率记为 f_k ，($k = 1, 2, \dots, K$)。

2) 根据正态分布 $N(\mu, \sigma^2)$ 的理论频公式

$$P_k = \int_{(a_{k-1} - \mu) / \sigma}^{(a_k - \mu) / \sigma} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

计算出 ξ 在区间 $[a_{k-1}, a_k]$ 上取值的理论频率 p_k ($k = 1, 2, \dots, K$)。

3) 构造统计量——经验分布与正态分布之差异度：

$$\chi^2_q = N \sum_{k=1}^K \frac{(f_k - p_k)^2}{p_k}$$

渐近服从自由度为 $k-1$ 的 χ^2 分布。当 N 充分大时，它可以通过 χ^2 分布的概率值 $P(\chi^2 \geq \chi^2_q)$ 的大小，看出经验分布与正态分布之差异是否显著，从而达到肯定或否定原正态假设之目的。标准量 P_q 应从问题的具体情况选择，一般采用 $P_q = 0.03 \sim 0.05$ 之间。

(2) 累积频率检验

以比较经验累积频率与理论累积频率的差异是否显著为标准的检验方法，称为分布函数的累积频率检验。其方法的一般步骤如下：

1) 将 ξ 的观测数据的取值区间 (a, b) 分为 K 个小区间，并求出经验累积频率

$$S_k(x) = \begin{cases} 0 & (-\infty, a] \\ \frac{k}{\sum_{i=1}^K f_i} & (a_{k-1}, a_k] \\ 1 & (b, +\infty) \end{cases}$$

其中 f_i 为小区间 $(a_{i-1}, a_i]$ 上的频率。

2) 根据正态分布 $N(\mu, \sigma^2)$ 的理论分布公式

$$F(X) = \int_{-\infty}^{(X-\mu)/\sigma} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

计算出 $F(a_k)$ ， $(k=1, 2, \dots, K)$ 。

3) 构造统计量——理论分布与经验累积频率之间的最大偏差

$$\lambda = \sqrt{N} \max_{1 \leq k \leq K} |F(a_k) - S_k(a_k)|$$

$$= \sqrt{N} \max_{1 \leq k \leq K} |F(a_k) - \frac{k}{\sum_{i=1}^K f_i}|$$

当 N 充分大时，它渐近服从柯尔莫哥洛夫——斯米尔诺夫分布。这样就可以通过 λ 的概率值 $Q(\lambda)$ 的大小，判断出经验分布与理论分布之差异是否显著，从而达到肯定或否定原正态假设之目的，显著性水平 α ，一般取为 $0.3 \sim 0.5$ ，如取 $\alpha = 0.5$ ，由柯尔莫哥洛夫分布表，可查得临界值 $\lambda_\alpha = 1.36$ ，当

$$\lambda > 1.36$$

时，可以拒绝接受 ξ 的正态性假设。

正态分布的参数 μ 和 σ^2 未知时，可以用它们的估值，即用样本的均值 $\bar{X}$ 和样本方差 S^2 代替。这时的统计量 χ^2 ，由于引入了统计量 $\bar{X}$ 和 S^2 ， χ^2 分布的自由度减少两个，渐

近服从自由度为 $K-3$ 的 χ^2 分布。对于累积频率检验，若 $F(X)$ 引入了样本统计量 $\bar{X}$ 、 S^2 ，统计检验的标准有所下降，使用时应加以小心。

2. 夏皮罗——威尔克检定法

前面的正态性检验法都是对大样本而言特别有效，在样本容量较少，如 $n \leq 50$ 时，S·S·夏皮罗及 M·B 威尔克严格用正态分布的顺序统计量的方差——协方差数据，算出一正态性检验的系数表（见*），其检验方法也是首先假设 ξ 服从正态分布。具体步骤如下：

1) 按观测子样数据的大小，排成

$$X_1 < X_2 < \dots < X_n$$

2) 查夏皮罗——威尔克的 α_{in} 系数表，查出对应于 n 值的诸 α_{in} 值。

3) 计算统计量

$$W = \frac{\left\{ \sum_{i=1}^n \alpha_{in} (X_{(n-i+1)} - X_{(i)})^2 \right\}}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

其中分子的 $\sum_{i=1}^n$ ，当 n 为偶数时为 $\sum_{i=1}^{n/2}$ ，当 n 为奇数时为 $\sum_{i=1}^{(n+1)/2}$ 。

4) 给出危险率 α （信度），查夏皮罗——威尔克的 $W(n, \alpha)$ 表（见*）中相应的 $W(n, \alpha)$ 值，如果

$$W \leq W(n, \alpha)$$

则在信度为 α 的意义下否定原正态性的假设。

三、转换函数

变换是观测值的一个函数，它定义一个新的观测值

$$Z = f(X)$$

这里 X 是原观测值， Z 是新观测值， f 是一变换，它可以是一线性函数，也可以是一非线性函数。函数的线性性质也决定变换的线性性质。线性变换不改变分布的形状，只改变尺度；非线性变换则不但可以改变尺度，而且能改变分布的形状。因此人们常利用非线性变换的这一特点，将非正态分布的观测数据变为正态分布的数据。这样的非线性函数，通常称为（正态性）转换函数。在对一些非正态分布的地质观测数据作正态性转换时，还能达到稳定方差和获得可加性的目的。如金的化验值 X ，可表为

$$X = \mu \cdot e^{\varepsilon}$$

其中 μ 为 X 的期望值, ε 为均值为零, 方差恒定的观测误差, 且服从正态分布 $N(0, D\varepsilon)$ 。从上式可看出金化验值的方差是依赖于 μ 的函数, 它反映了在低品位矿石中, 金化验值的品位变化性小, 在高品位矿石中品位变化大, 是一非恒定方差的模型。由于上式可写为

$$X = f(\varepsilon),$$

且 f 是一个非线性函数, 因此 X 的分布为非正态的, 且 X 对 μ 和 e^{ε} 不具可加性, 即 μ 和 e^{ε} 为乘积模型。现若对它作转换, 只须取对数, 即有

$$Y = \ln X = \ln \mu + \varepsilon$$

此时 Y 不但服从正态分布, 而且具有恒定方差 $D\varepsilon$, 对 $\ln \mu$ 和 ε 还有可加性, 这样就更有利于作不同线性模型的处理。当然, 这不是说对任一非正态的分布都能通过正态转化而达到上述三种目的, 应主要根据所选用的处理方法的要求, 达到某一或二目的即可。比如, 在非平稳时间序列分析中, 主要关心的就不一定是正态化, 而是需要化为具有恒定方差的可加模型。

在变换观测值使其服从正态分布时, 转换函数当然和原观测值的分布有关, 比如二项分布, 可用角变按 (或称反正弦变按) $\sin^{-1} \sqrt{X}$, Poisson 分布可通过平方根变换, 使其正态或近似正态化。离散的 Poisson 分布和二项分布, 在地质变量中虽然很少见, 但接近于它们的连续分布, 如偏度中等的仍可用平方根变换, 弱偏的也可用角度变换。其中弱负偏 (或左偏) 可用反正弦变换 $\sin^{-1} (\sqrt{X/10n})$ 、弱正偏 (或右偏) 可用反余弦变换 $\cos^{-1} (\sqrt{X/10n})$, n 取 X_i 中最大值的整数位数。对地质数据应用得最广的还是适用于正偏度大的对数变换, 它除了能把对数正态分布的数据变为正态外, 还可获得可加性和方差齐性。对数变换法虽然能用于最大范围的正数, 但确不能用于零或接近于零的值, 这对地质数据而言是很难免的。因此, 在取对数之前应先把各数加 1。对数的底, 原则上任何底数都可以, 不过在考虑到要对小于某数的观测值作平方根变换, 这时就可以取某数为底的对数。比如以 10 为底, 此时, $\log_{10}(X+1)$ 的性质对在 10 以内的数值来说, 同平方根变换一样, 对在 10 以上的数值来说, 与 $\log_{10}(X)$ 相差不多。

以上变换都是以观测数据的分布为已知的几类特殊分布为依据的。但是, 地质观测数据的分布是多样的, 且难于知道或根本就不知道, 在这种情况下我们只能直接从观测数据的经验分布出发, 来寻找一个近似的正态转换函数。这里介绍一种用埃尔米特正交多项式表示的近似转换函数。设

$$Y = \Phi^{-1}(X)$$

或

$$X = \Phi(Y)$$

其中 X 代表非正态分布的观测数据, Y 代表经 Φ^{-1} 变换后为标准正态观测数据。注意到

关系式 $X = \Phi(Y)$ 包含 X 的分布 $F(X)$ 经 $\Phi^{-1}(X)$ 转换为 Y 的标准正态分布 $G(Y)$ 函数, 或标准正态分布函数 $G(Y)$ 经 $\Phi(Y)$ 转换为 $F(X)$ 分布, 即

$$F(X) \xleftrightarrow{\Phi^{-1}} G(Y)$$

为此, 我们可用 X 的经验分布 $F^*(X)$ 代替 $F(X)$, 则可通过 $F(X_i) = G(Y_i)$ 把 X 与标准正态变量 Y 之间的近似对应关系 $X = \Phi^*(Y)$ 建立起来。建立这种近似的转换函数, 一般采用图解法 (当然也可由程序在计算机上实现)。如非洲南部花岗岩中钾是一非正态分布, 它的近似转换就可通过钾的经验分布函数 $F^*(X)$ (如图三) 与标准正态的分布曲线 (如图四) $G(Y)$, 通过 $F(X_i) = G(Y_i)$ 的图解法, 找出 X 与 Y 之间的近似对应关系 (如图五)。即转换函数 $X = \Phi^*(Y)$ 。然后把 $\Phi(Y)$ 展为 K 阶埃尔米特多项式, 即

$$\Phi(Y) = \sum_{k=0}^K \varphi_k \eta_k(Y) = \sum_{k=0}^K \frac{\psi_k}{k!} H_k(Y)$$

其中 $H_k(Y)$ 是由

$$(e^{-y^2}) H_k(Y) = \frac{d^k}{dy^k} (e^{-y^2})$$

定义的非标准化的埃尔米特正交多项式, ψ_k 由积分

$$\psi_k = \int_{-\infty}^{\infty} \Phi(Y) H_k(Y) e^{-y^2} dy$$

所确定。

计算 ψ_k 的问题, 可化为一个积分数字计算问题, 因而有不同的解法。但一般都采用高斯——埃尔米特积分法, 即

$$\psi_k^* = \sum_{i=1}^m W_i H_k(Y_i) \Phi(Y_i)$$

其中 ψ_k^* 为 ψ_k 的估计量, m 为高斯法的点数, m 必须大于 ψ_k 中的埃尔米特多项式中的最高阶 K , W_i 为高斯积分中的数系数, 可查表, 也可通过下面公式直接计算。

$$W_i = \frac{k! \sqrt{\pi} 2^{K+1}}{[H_{K+1}(Y_i)]^2}$$

Y_i 为横坐标, $\Phi(Y_i)$ 为 $\Phi(Y)$ 在 Y_i 处的值。

埃尔米特多项式 $H_k(Y)$ 的最高阶 K , 应决定于 $\Phi(Y)$ 的形状, 衡量的标准是

X与 $\Phi(Y)$ 是否有相同的一、二阶距。即

$$M = E\{X\} = E\{\varphi(Y)\} = \varphi_0 = \psi_0.$$

$$\sigma^2 = E\{X^2\} = E\{\varphi^2(Y)\} = E\left\{\sum_{k=0}^K \varphi_k \eta_k(Y)\right\}^2$$

$$= \sum_{k=0}^K \varphi_k^2 = \sum_{k=0}^K \psi_k^2 / k!$$

如果我们取 $\varphi_0 = M$ ，就可以通过二阶距的接近程度

$$d(K) = \left| \sigma^2 - \sum_{k=0}^K \frac{\psi_k^2}{k!} \right|$$

来确定。即使 $d(K) < \delta$ （允许误差）时的K。到此，用埃尔米特多项式表示的转换函数，就完全确定了。

在地质统计学中，析取克里格方程组，就是利用上述——正交的埃尔米特多项式所表示的——转换函数建立起来的。从而为非线性估计开创了新的途径。不过在统计分析中，使用转换函数应特别慎重。因为某些偏倚分布，可能反映着元素在其空间上，有着规律性的变化，此时指示样品可能来自几个总体，分布可能各不相同。在这种情况下，数据转换是不妥当的，应该从成因分析入手，筛分不同总体，除非可以证明，偏倚分布不具空间意义，才能用转换函数的方法，对数据进行处理。

四、后记

地质变量的分布特征的研究，不仅对现有数学地质模型的有效应用有好处，而且某些地质变量的分布特征本身就可用来对区域成矿进行预测。

A·B·维斯捷利乌斯（1960）指出：化学元素在地壳中分布可分为两类：

1）单一地化作用用的产物，其分布是正态的。

2）多次地化作用的产物，其分布是偏斜的，而正向偏斜程度还可能指示矿化富集的程度。

1968年N·A·马拉霍夫运用数理统计方法，研究苏联乌拉尔超基性岩的含矿(Cr)性后，他指出，不含矿(Cr)超基性岩中， Cr_2O_3 呈正态分布，而含矿超基性岩中， Cr_2O_3 呈对数正态分布。

1976年黑龙江省地质局地科所铬矿组，对含矿(铬)与非铬超基性岩中，铬尖晶石

分布取得同样的结论，他们认为与铬矿化有关的铬尖晶石在含矿岩相中的分布往往是由某种特定的地质因素所决定（如岩体产状、岩相和构造条件），造成 Cr_2O_3 在岩体中分布不均匀性。

1974年夏林圻，对我国西北地区一些超基性岩含铬性进行了统计研究，仍然发现不含铬岩体 Cr_2O_3 为正态分布，而七个含矿（铬）的岩体， Cr_2O_3 分布曲线皆成明显的正偏。且反映分布曲线偏斜程度的三阶中心距 μ_3 的数值大小，还能近似的反映出矿床规模的正消长规律

矿化岩体	$\mu_3 < 0.01$
小型矿床岩体	$0.01 < \mu_3 < 0.03$
中型矿床岩体	$0.03 < \mu_3 < 0.05$
大型矿床岩体	$\mu_3 > 0.05$

据此，他指出青海L_N与E岩体 μ_3 数值很高，而目前查明之矿床与之不相称，应进一步工作，以尽可能找到更大的矿床的预测。

1980年孙念义、罗决诚用航空放射性测量 γ 总量的资料，对华南花岗岩作统计研究后指出，无矿的H岩体属正态分布，含矿岩体，一般放射性元素分布有绝对和相对不均匀性，分布曲线有呈较大的正偏或负偏的特点。

这就是说，放射性测量 γ 和地质变量 Cr_2O_3 的分布特征与含矿性具有一样的规律，并可据此规律进行区域性的成矿预测，至于其它地质变量是否也有同样的规律呢？这就不能简单地肯定和否定了，还需要细微的分析和深入的研究，不过至少还会有某些地质变量也会具有如上之规律的。据统计，在一般情况下，铜、铝、锌、钼矿多数是服从对数正态分布的，而镁、磷矿则多数是倒对数正态分布的。如西非和南美洲的一些淋滤型铁矿，由于淋滤作用，使 SiO_2 从氧化铁中流失而富集成的氧化铁矿，其中Fe的分布就为倒对数正态分布如图六。这种分布与德兴铜矿铜的分布不一样，即分布曲线的尾部拖长的方向正好相反。

人们在研究这个世界上潜在的巨大铁矿之一的西非Nimla山脉中的Guinea淋滤型矿床，发现上部地表经风化淋滤作用而富集形成的氧化铁矿，其Fe的品位分布曲线，十分吻合于均值为66.6%、方差为2.3的标准倒对数正态分布曲线，而深部原生带中的Fe的品位分布曲线，则呈近似的正态分布形状。为什么是近似呢？这可能是原生带混杂有过度带的样品（风化淋滤较深的那一部分的样品）。这就是说，若一个矿床矿石类型比较均一，则其实验分布曲线与标准分布曲线十分吻合，反之，若矿床中矿石类型不是均一的，则实验分布曲线也很不规则。因此对于不规则的实验分布曲线，首先就应该根据地质条件，弄清不均一的原因，然后才根据研究的目的，具体分别地对不同类型的矿化带进行研究。

参考文献

地质统计方法

1. 地质数据统计分析 [美] 小乔治·S·科克 著
理查德·F·林克
2. 概率统计计算 中国科学院计算中心概率统计组
- *3. 误差分析方法 何国伟 著
4. 数值计算方法 冯 康 等编
5. Mining Geostatistics A.G. Journel
6. 试用数理统计方法预测超基性岩性的含矿(铬)性
西北地质科学研究所 夏林圻
7. 运用数学地质方法研究超基性岩的化学成分及含矿性
黑龙江省地质科学研究所超基性岩组
8. 中南区产铀花岗岩区域放射场特征 孙念义 罗决诚
9. 地质勘探中的统计分析 武汉地质学院勘探教研室

关于矿床条件模拟的基本理论和程序设计的初步探讨

成都地质学院 王柏钧* 周晓笈

一九七三年以来,在法国地质统计学家G·Matheron提出的转动带法(The“Turning Bands” Method)的基础上, A·Journel 等人发展了地质统计学中的一种新方法——矿床条件模拟。这种模拟的矿床是在最大限度地从有关矿床的真实资料中提取有用信息的基础上,结合克里金法(Kriging)作出的一种网度很密的数学模型。它与克里金法有着密切的联系:在矿床条件模拟中必须使用克里金;但二者所建立起来的数学模型却各有其不同的特点和用途。克里金法的主要用途是估计矿床的储量,圈定矿体,衡量估计质量的标准是无偏性以及使估计方差最小。但是这样建立起来的数学模型往往并不能重现有关区域化变量(例如矿石品位)的空间变化,因为使估计方差最小就包含着对真实品位的分布加以园滑化。因此,在涉及到必须了解有关区域化变量在空间中的波动变化的问题中(例如在矿山设计中研究出矿品位的波动和配矿问题),使用克里金法模型是不恰当的。而对矿床进行条件模拟,虽然在每一点的模拟值并不是可能的最优估计值,(可以证明,条件化模拟值对真实值的估计方差正好是克里金方差的两倍)但它却能更好地重现真实值在空间中的波动变化,具有与真实值相同的主要统计分布特征(均值、协方差或变异函数,以及直方图),并在真实资料点上与真实值相同。因此,近年来矿床条件模拟正在矿床勘探和矿山设计工作中得到日益广泛的应用。例如:(1)研究比较各种估计矿量的方法;(2)研究为达到预定的目的所要求的勘探网度;(3)研究在选别开采,确定原地资源的可回采矿量的时候,有效资料数量对其影响如何;(4)研究最小开采单元的尺寸,富矿与贫矿块段在空间分布对开采的影响,回采矿石的贫化等问题;(5)研究配矿方法;(6)研究生产规模与边界品位,等等。

条件模拟的基本理论

条件模拟的对象,可以是和矿床有关的各种区域化变量,如矿石品位,覆盖层厚度等等。区域化变量可以解释为某一个随机函数 $Z(X)$ 的一次实现,这一类的随机函数

* 王柏钧为指导教师

可以用它的分布函数与协方差或变异函数模型来加以描述。条件模拟即是要从随机函数 $Z(X)$ 的这一次实现中引出这一类随机函数的其他实现 $Z_s(X)$ ，并只保留那些在取样点上与实际资料值相符的实现 $Z_{sc}(X)$ ，即 $Z_{sc}(Xa) = Z(Xa) \forall Xa \in \text{矿床资料组 I}$ 。这些实现 $Z_{sc}(X)$ 被称为区域化变量 $Z(X)$ 的条件模拟。

条件模拟的基本步骤有两个，即：

- (i) 用转动带法进行非条件模拟；
- (ii) 用克里金法对 (i) 所得到的非条件模拟值进行条件化。

1. 在一维空间的非条件模拟

转动带法采用了先在一维空间，即沿一条直线模拟一个随机函数 $Y(u)$ ，其协方差为 $C^{(1)}(S)$ ，然后经过转动得出具有所要求的协方差 $C(\gamma)$ 的三维随机函数的实现。

在地质统计学中，通常应用的三维空间模型，如球状模型或指数模型，其相应的一维空间的协方差 $C^{(1)}(S)$ 可以用函数 $f(S)$ 与其变换 $\check{f}(S) = f(-S)$ 的褶积来表示：

$$C^{(1)}(S) = f(S) * \check{f}(S) = \int_{-\infty}^{+\infty} f(u) f(u+S) du \quad (I.1-1)$$

因此，一维空间的模拟可以用具有权函数 $f(u)$ 的滑动平均法这种简单的方法来进行。

考虑一个弱平稳（二阶距平稳）的随机量 $T(dr)$ ，使得

$$E\{T(dr)\} = 0,$$

$$E\{T(d\gamma) \cdot T(d\gamma')\} = \begin{cases} 0, & \text{若 } \gamma \neq \gamma' \\ \sigma^2 d\gamma, & \text{若 } \gamma = \gamma' \end{cases} \quad (I.1-2)$$

每一个随机量 $T(d\gamma)$ 被权函数 $f(u)$ 正则化之后，其值即确定了下面的一维空间的随机函数 $Y(u)$ ：

$$Y(u) = \int_{-\infty}^{+\infty} f(u+\gamma) T(d\gamma) \quad (I.1-3)$$

$Y(u)$ 为一平稳随机函数。以 $C^{(1)}(S)$ 表其协方差，由 (I.1-2) 式可得

$$\begin{aligned} C^{(1)}(S) &= E\{Y(u) Y(u+S)\} \\ &= E\left\{\int_{-\infty}^{+\infty} f(u+\gamma) f(u+\gamma'+S) T(d\gamma) T(d\gamma')\right\} \end{aligned}$$

$$= \int_{-\infty}^{+\infty} f(u+\gamma) f(u+\gamma+S) \sigma^2 d\gamma$$

$$= \sigma^2 f(S) * \check{f}(S) \quad (I.1-4)$$

将 (I.1-4) 式与 (I.1-1) 式对比, 可以看出, 前者只多了一个 σ^2 , 如果我们利用一个校正系数来对它加以调整, 我们就得到了所要求的一维空间平稳随机函数 $Y(u)$, 其协方差为: $C^{(1)}(S) = f(S) * \check{f}(S)$ 。

前面的随机量 $T(d\gamma)$, 实际上表示了一种纯随机波动, 它的离散形式就是一系列具有相同的分布函数的相互独立的随机变量 T_i , 十分类似于所谓“白噪声”。为了方便, 我们把它取作均匀分布的随机变数列, 其期望为 0, 其特定的方差 $E\{T^2\} = \sigma^2$ 。(参见 [2]) 然后, 使这一系列随机变量 T_i 的实现沿一条直线以相等的间隔 b 分布, 即令 $t_{i-k}, \dots, t_i, \dots, t_{i+k}$ 为在点 $i-k, \dots, i, \dots, i+k$ 上的实现值, 用函数 $f(u)$ 加权以正则化, 取滑动平均 y_i , 在 i 点定义 y_i 为

$$y_i = \sum_{k=-\infty}^{+\infty} t_{i+k} f(kb) \quad (I.1-5)$$

在实际计算中, 我们必须使当 $|k|$ 增大时, $f(kb)$ 很快减小到 0, 这样才能使得对离散点的求和范围不必取从 $-\infty$ 直到 $+\infty$ 。

这些 y_i 值可以看成是一个平稳的一维空间随机函数 $Y_i = Y(u)$ 的一个实现, 这个随机函数的期望为 0, 协方差为

$$C_d^{(1)}(S) = E\{Y_i Y_{i+s}\}$$

$$\left\{ \sum_{k=-\infty}^{+\infty} T_{i+k} f(kb) \sum_{k'=-\infty}^{+\infty} T_{i+s+k'} f(k'b) \right\}$$

令变量 $L = S + k'$, 代入上式, 即得

$$C_d^{(1)}(S) = E \left\{ \sum_{k=-\infty}^{+\infty} \sum_{l=-\infty}^{+\infty} T_{i+k} T_{i+l} f(kb) f(l-b) \right\}$$

当 $k \neq l$ 时, T_{i+k} 与 T_{i+l} 相互独立, 且方差 $E\{T^2\} = \sigma^2$, 故上式简化为

$$C_d^{(1)}(S) = \sigma^2 \sum_{k=-\infty}^{+\infty} f(kb) f(kb-bS) \quad (I.1-6)$$

这一离散型的协方差 $C_d^{(1)}(S)$ 即是公式 (I. 1 - 1) 中以积分定义的一维协方差 $C^{(1)}(S)$ 的近似值。

偏差 $C^{(1)}(S) - C_d^{(1)}(S)$ 是由于离散近似造成的, 在实际计算中只要通过人工计算参数 $\sigma^2 = E\{T^2\}$ 来保证使 $C^{(1)}(0) = C_d^{(1)}(0)$, 就可以求得校正系数, 从而校正这一偏差。

2. 在三维空间的非条件模拟——转动带法

以变量 $Z_S(X)$ 与 $Z_O(X)$ 表示在三维空间中的区域化变量, $X \in R^3$ 。考虑 R^3 中的一条直线 D_1 , 以及定义在直线 D_1 上的一个一维随机函数 $Y(X_{D_1})$, 这一随机函数是弱平稳的, 其期望 $E\{Y(X_{D_1})\} = 0$, 其一维协方差为 $C^{(1)}(h_{D_1})$ 。令 X_{D_1} 为任意一点 X 在这条直线 D_1 上的射影, 考虑三维空间中的随机函数 $Z_1(X)$, 其定义为 $Z_1(X) = Y(X_{D_1})$, $\forall X \in R^3$ 。这一随机函数是弱平稳的, 其期望为 0, 三维空间的协方差等于

$$E\{Z_1(X)Z_1(X+h)\} = E\{Y(X_{D_1})Y(X_{D_1}+h_{D_1})\} = C^{(1)}(h_{D_1})$$

式中 h_{D_1} 为矢量 h 在直线 D_1 上的射影。

在具体做法上, 为了产生一个实现 $Z_1(X)$, 我们对以平面 $\{X_{D_1} = \text{常数}\}$ 为中心的薄片 (或条带) 上的所有点都给予直线 D_1 上点 X_{D_1} 处的模拟值 $y(X_{D_1})$, 上述平面是在点 X_{D_1} 处与直线 D_1 垂直的, 如图 1 所示。这一薄片在直线 D_1 上的模拟值两侧是等厚的, 其厚度正相当于我们在第 1 节中所取的直线上的间隔 b 。然后考虑 N 条直线 $D_1, D_2, \dots, D_n$, 它们的方向分别相当于在单位球中均匀分布的单位矢量 $k_1, k_2, \dots, k_n$ 的方向。在每一条直线 D_i 上分别产生随机函数 $Y(X_{D_i})$ 的一个实现 $y(X_{D_i}), y(X_{D_i})$ 是与 $Y(X_{D_i})$ 同构的, 即它们具有相同的期望值和相同的二阶矩——协方差或变异函数。这个随机函数 $\{Y(X_{D_i}), i = 1, \dots, N\}$ 是相互独立的。用上述办法, 对应于每一个一维的实现 $y(X_{D_i})$, 都可以得到一个三维的实现 $Z_i(X) = y(X_{D_i}), \forall X \in R^3$ 。最后赋予每一点 X 的值则是 N 条直线上的 N 个值的和。

$$Z_s(X) = \frac{1}{\sqrt{N}} \sum_{i=1}^N Z_i(X)$$

由于N条直线能够通过三维空间的转动来互相产生，定义在它们上面的薄片（或条带）也是如此，这种方法因而被称为转动带法。

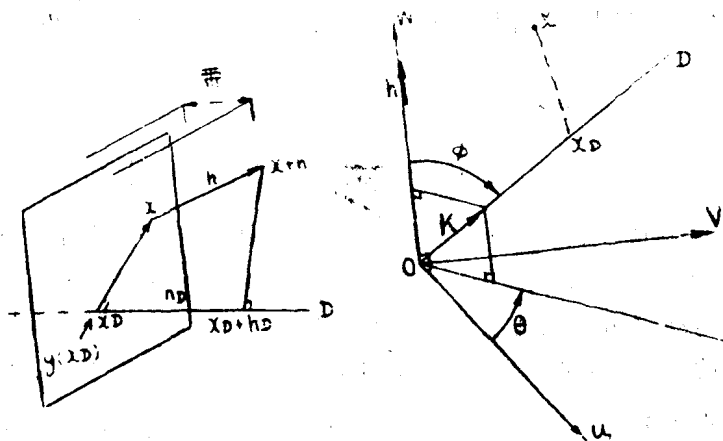
最终得到的实现 $Z_s(X)$ 是三维随机函数 $Z_s(X) = Z(u, v, w)$ 的一个实现。这个随机函数是弱平稳的，期望为0，其协方差为

$$C(h) = E \left\{ Z_s(X) Z_s(X+h) \right\} = \frac{1}{N} \sum_{i=1}^N C^{(1)}(h_{D_i})$$

当 $N \rightarrow \infty$ 时，这一协方差将趋向于各向同性的协方差

$$C(r) = \frac{1}{2\pi} \int_{\frac{1}{2} \text{ 单位球面}} C^{(1)}(\langle h, k \rangle) dk \quad (I.2-1)$$

这里 $\langle h, k \rangle = |h| \cos \psi$ ，是矢量 h 在轴 k 上的射影， $r = |h| = \sqrt{h_u^2 + h_v^2 + h_w^2}$ ，为矢量 h 的模，参看图1。



积分 (I.2-1) 可用球坐标表为

$$C(r) = \frac{1}{2\pi} \int_0^{2\pi} d\phi \int_0^{\pi} C^{(1)}(r \cos \psi) \sin \psi d\psi$$

$$= \frac{1}{\gamma} \int_0^{\gamma} C^{(1)}(S) ds$$

这里利用了变换 $S = \gamma \cos \psi$ 。当 h 与 k 的方向一致时, $\psi = 0$, 于是 $S = \gamma$ 。

在实际情况下, 三维协方差 $C(\gamma)$ 是给定的, 这样, 在 N 条直线中的每一条上要模拟的一维协方差 $C^{(1)}(S)$ 都可用下面的微分来给出

$$C^{(1)}(S) = -\frac{\partial}{\partial S} C(S) \quad (I. 2 - 2)$$

这样, 在直线 D_i 上要模拟的随机函数 $Y(u_i)$ 的一维协方差 $C^{(1)}(S)$ 总是可以确定的, 从而可以用第 1 节中所述的方法得到直线 D_i 上对随机函数 $Y(u_i)$ 的非条件模拟。把这些直线与其相应的条带在三维空间 R^3 中转动, 则可以求得具有指定的各向同性协方差 $C(\gamma)$ 的三维空间的模拟 $Z_S(u, v, w)$ 。

3. 常用协方差模型

地质统计学中通常采用的三维空间协方差模型, 有球状模型, 指数模型等。可用公式 (I. 2 - 2) 求出其相应的一维空间协方差 $C^{(1)}(S)$ 。可以证明, 它可以用

$$C^{(1)}(S) = f(S) * f(S) \quad \text{的形式来表示。}$$

(1) 球状模型

球状模型可用公式表示如下:

$$C(\gamma) = \begin{cases} K \left[1 - \frac{3}{2} \cdot \frac{\gamma}{a} + \frac{1}{2} \cdot \left(\frac{\gamma}{a} \right)^3 \right] & \forall \gamma \in [0, a] \\ 0 & \forall \gamma > a \end{cases} \quad (I. 3 - 1)$$

式中 a 为变程 (range), 常数 $K = C(0)$ 是三维空间随机函数 $Z(X)$ 的先验方差, 也就是相应半变异函数 $\gamma(h)$ 的基台值 (sill)。

利用公式 (I. 2 - 2) 可以求得对应的一维空间协方差:

$$C^{(1)}(S) = \begin{cases} K \left[1 - 3 \cdot \frac{s}{a} + 2 \cdot \left(\frac{s}{a} \right)^3 \right] & \forall S \in [0, a] \\ 0 & \forall s > a \end{cases} \quad (I. 3 - 2)$$

这个一维空间协方差用褶积 $C^{(1)}(S) = f * f$ 的形式表示时, 其权函数为: