

25-28

时空域中的指示克立格理论研究

肖斌

赵鹏大

侯景儒

(北京大学·北京·100871) (中国地质大学·北京·100083) (北京科技大学·北京·100083)

对时空域内的指示克立格进行了较全面的研究,主要研究内容有:(1)时空域内的指示克立格(STIK)变换;(2)指示二阶距;(3)单元指示克立格(STIK):①简单指示克立格(STSIK),②普通指示克立格(STOIK),③中值指示克立格(STMIK);(4)多元指示克立格:①指示协同克立格(STCOIK),②主分量指示克立格(STPCIK)。

关键词 时空域 简单指示克立格 普通指示克立格 中值指示克立格 指示协同克立格 主分量指示克立格

地质、物化探数据的分布常出现一个长尾巴,这种分布特征影响了数据的整体分布规律及数据处理与估计的精度。在实际应用中,被研究的母体取样有限且分布不均;特异值的存在使变异函数不稳定;数据的分布假设无法检验等一系列的问题使得参数地质统计学的应用受到一定限制。A Journel 发展了无需对数据分布作任何假设的非参数地质统计学,提出了一些非参数地质统计学方法。

1 时空域中的指示克立格(STIK)转换

时空域中的指示克立格(STIK)的研究是在原始数据进行指示转换的基础上进行的,对不同类型的变量有不同的转换方法。

设在每一时刻 $\tau(\tau=1, \dots, T)$, 在 N_τ 个空间位置 $\alpha(\alpha=1, \dots, N_\tau)$ 上,对某一区域化变量 Z 得到一组观测值,如金矿床中的金的品位值,将这一组观测值记为:

$$\{Z[(\tau, x)_\alpha]; \tau = 1, \dots, T; \alpha = 1, \dots, N_\tau\}$$

x 表示空间中的一个支撑; τ 表示时间。

若区域化变量是连续分布的,要对该区域化变量 z 在研究邻域内进行指示转换,应先确定阈值,即临界值($z^l, l=1, \dots, L$, l 表示第 l 个临界值)。指示转换的常用表达式为:

$$E\{I[(\tau, x)_\alpha; z^l]\} = 1 \times \text{Prob}\{Z[(\tau, x)_\alpha] \leq z^l\} + 0 \times \text{Prob}\{Z[(\tau, x)_\alpha] > z^l\} = \text{Prob}\{Z[(\tau, x)_\alpha] \leq z^l\} = F(z^l)$$

式中 $F(z^l)$ 表示品位分布函数在边界品位 z^l 处的品位值。

当 $I[(\tau, x)_\alpha; z^l]$ 和 $I[(\tau+t, x+h)_\alpha; z^l]$ 为被矢量 (t, h) 分离的二随机变量时,可定义如下二阶距^[2,4]:

$$I[(\tau, x)_\alpha; z^l] = \begin{cases} 1, & \text{若 } (\tau, x)_\alpha \text{ 点处 } Z[(\tau, x)_\alpha] \leq z^l \\ 0, & \text{若 } (\tau, x)_\alpha \text{ 点处 } Z[(\tau, x)_\alpha] > z^l \end{cases} \quad (1)$$

上面是 $[0, 1]$ 二分表示法,也可以将其进行三分 $[-1, 0, 1]$ 。通过(1)式可将原始数据转换成以 0 和 1 表示、按原有分布特征分布的一组指示转换值 $I[(\tau, x)_\alpha; z^l]$ 。

如果区域化变量 $Z[(\tau, x)_\alpha]$ 是连续性变量,那正确地选择临界值 z^l 就显得相当重要:若临界值太多,推断和计算就显得烦琐,花费也就昂贵;临界值太少,则某些有贡献的细节就会丢失。

若该区域化变量是离散分布的,则应用指示克立格就显得方便。如果区域化变量 $Z[(\tau, x)_\alpha]$ 是一个二分变量,如在岩性分析时,则指示值可这样得到:在时空位置 $(\tau, x)_\alpha$ 处,某一特定岩性出现或占优势时其值设为 1,否则为 0。

2 指示二阶距

原始数据 $Z[(\tau, x)_\alpha]$ 服从随机分布,在原始数据基础上转换的指示数据 $I[(\tau, x)_\alpha; z^l]$ 也为随机分布。在已知边界品位 z^l 的条件下,随机函数 $I[(\tau, x)_\alpha; z^l]$ 服从 $[0, 1]$ 二项分布。二分指示随机函数 $I[(\tau, x)_\alpha; z^l]$ 的平稳期望是函数 $Z[(\tau, x)_\alpha]$ 的累加分布函数(cdf)本身:

(1)非中心化互协方差:

$$K_I(t, h; z^l) = \text{Prob}\{Z[(\tau+t, x+h)_\alpha] \leq z^l\} \text{ 和 } \text{Prob}\{Z[(\tau, x)_\alpha] \leq z^l\} \quad (3)$$

(2)非中心化协方差:

$$K_I(t, h; z^l, z^l) = \text{Prob}\{Z[(\tau+t, x+h)_\alpha] \leq z^l\} \text{ 和 } \text{Prob}\{Z[(\tau, x)_\alpha] \leq z^l\} \quad (4)$$

(3) 中心化协方差:

$$C_I(t, h; z^l) = K_I(t, h; z^l) - F^2(z^l) \quad (5)$$

(4) 方差:

$$\text{Var}[I(\tau, x; z^l)] = C_I(0, 0; z^l) = F(z^l) - F^2(z^l) \quad (6)$$

(5) 半变异函数:

$$\gamma_I(t, h; z^l) = 0.5 \times E\{[I(\tau + t, x + h; z^l) - I(\tau, x; z^l)]^2\} =$$

$$C_I(0, 0; z^l) - C_I(t, h; z^l) = F(z^l) - K_I(t, h; z^l) \quad (7)$$

(6) 指示半变异函数与指示协方差的关系:

$$C_I(t, h; z^l) = C_I(0, 0; z^l) - \gamma_I(t, h; z^l) \quad (8)$$

3 单元指示克立格(STIK)

对连续变量进行指示克立格的目的并不仅仅是要得到指示转换值 $I[(\tau, x)_a; z^l]$ 而是需要估计当 $I[(\tau, x)_a; z^l]$ 为某值或处于某个区间时的概率。 $I[(\tau, x)_a; z^l]$ 可看成原始数据在临界值 z^l 处的条件累加分布函数(ccdf), 指示克立格为 ccdf 在临界值 z^l 处提供了最小方差估计:

对 $I(W; z^l)$ 进行估计的 STIK 估计值为:

$$I^*(W; z^l)_{\text{STIK}} = \{ \text{Prob}\{Z[(\tau, x)_a] \leq z^l\} \}_{\text{STIK}} = \sum_{\tau=1}^T \sum_{\alpha=1}^{N_\tau} \lambda_{\alpha\tau}(z^l) \times I[(\tau, x)_a; z^l] + [1 - \sum_{\alpha=1}^{N_\tau} \lambda_{\alpha\tau}(z^l) \times F(z^l)] \quad (11)$$

上面 $\lambda_{\alpha\tau}(z^l)$ 表示对应于临界值 z^l 的简单克立格权值, 这些权值可由 STIK 方程组求得:

$$\sum_{\tau=1}^T \sum_{\alpha=1}^{N_\tau} \lambda_{\alpha\tau}(z^l) \times C_I(W_{\alpha\tau}, W_{\beta\tau}; z^l) = C_I(W, W_{\beta\tau}; z^l) \quad (12)$$

式中: $C_I(W_{\alpha\tau}, W_{\beta\tau}; z^l) = \text{Cov}\{I[(\tau, x)_a; z^l], I[(\tau + t, x + h)_\beta; z^l]\}$ 是在临界值 z^l 处的指示协方差; W 为时刻 τ 处的某待估域; $\tau = 1, \dots, T$; $\alpha, \beta = 1, \dots, N_\tau$; 它有 $\sum_{\tau=1}^T N_\tau$ 个未知数, 共有 $\sum_{\tau=1}^T N_\tau$ 个方程, 故在正定条件下方程组有确定解。

克立格方差 σ_k^2 是在 σ_E^2 最小的基础上得到,

$$\sigma_k^2 = C_I(W, W) - \sum_{\tau=1}^T \sum_{\alpha=1}^{N_\tau} \lambda_{\alpha\tau}(z^l) \times C_I(W, W_{\alpha\tau}; z^l) \quad (13)$$

如果有 L 个临界值 z^l , 则简单克立格需要 L 个指示协方差 $C_I(t, h; z^l)$ 和 L 个 cdf 函数值 $F(z^l)$, 需要解 L 次 STIK 方程组。如果原始数据有丛集化现象, 则样品的 cdf 值应去丛集化。

3.2 普通指示克立格(STOIK)

当指示值的平稳期望未知时, 要估计时空域中某待估点 W 的指示值, 则应考虑普通指示克立格。

时空域中, 在时刻 τ 处某待估域 W 的实际指示值为 $I(W; z^l)$, 相应的估计值为 $I^*(W; z^l)$:

$$I^*(W; z^l) = \sum_{\tau=1}^T \sum_{\alpha=1}^{N_\tau} \lambda_{\alpha\tau}(z^l) \times I[(\tau, x)_a; z^l] \quad (14)$$

要使估计无偏, 则 $E[I(W; z^l) - I^*(W; z^l)] = 0$

在二阶平稳条件下, 无偏条件为:

$$\sum_{\tau=1}^T \sum_{\alpha=1}^{N_\tau} \lambda_{\alpha\tau}(z^l) = 1 \quad (15)$$

$$I^*[(\tau, x)_a; z^l] = E\{I[(\tau, x)_a; z^l] | (n)\}^* =$$

$$\text{Prob}^*\{Z[(\tau, x)_a] \leq z^l | (n)\} \quad (9)$$

这里 n 表示在时空支撑 $(\tau, x)_a$ 的邻域内可利用的满足条件的数据数。

指示克立格估计在 L 个临界值 $[z^l, l = 1, \dots, L]$ 处重复进行。根据 L 个指示克立格估计而建立的 ccdf 函数表示待估计值 $Z[(\tau, x)_a]$ 的不确定性概率分布模型。下面分别介绍时空域中几种单元指示克立格技术。

3.1 简单指示克立格(STSIK)

当 $E[I[(\tau, x)_a; z^l]] = m$, m 为常数且已知时, 可以用时空域中的简单克立格法来研究。时空域中, 在时刻 τ 处某待估域 W 的实际指示值为 $I(W; z^l)$ (也可写成 $I[(\tau, x)_a; z^l]$, x 为时空域中的待估域 W 在空间域中的一个支撑, 后同):

$$I(W; z^l) = 1 / (T \times V) \int \int_{\tau} I[(\tau, x)_a; z^l] d\tau dx \quad (10)$$

在时刻 τ 处对 $I(W; z^l)$ 进行 STOIK 估计的估计方差计算式 σ_E^2 为:

$$\sigma_E^2 = C_I(W, W) - 2 \times \sum_{\tau=1}^T \sum_{\alpha=1}^{N_\tau} \lambda_{\alpha\tau}(z^l) \times C_I(W, W_{\alpha\tau}; z^l) + \sum_{\tau=1}^T \sum_{\alpha=1}^{N_\tau} \sum_{\beta=1}^{N_\tau} \lambda_{\alpha\tau}(z^l) \times \lambda_{\beta\tau}(z^l) \times C_I(W_{\alpha\tau}, W_{\beta\tau}; z^l) \quad (16)$$

上式中系数 $\lambda_{\alpha\tau}(z^l)$ 可以在估计方差最小和无偏条件下, 由下面 STOIK 方程组求得:

$$\sum_{\tau=1}^T \sum_{\alpha=1}^{N_\tau} \lambda_{\alpha\tau}(z^l) \times C_I(W_{\alpha\tau}, W_{\beta\tau}; z^l) - \mu = C_I(W, W_{\beta\tau}; z^l) \sum_{\tau=1}^T \sum_{\alpha=1}^{N_\tau} \lambda_{\alpha\tau}(z^l) = 1 \quad (17)$$

式中: μ 对应于无偏条件 $\sum_{\tau=1}^T \sum_{\alpha=1}^{N_\tau} \lambda_{\alpha\tau}(z^l) = 1$ 的拉格朗日乘子。

克立格方差 σ_k^2 是在 σ_E^2 最小的基础上得到,

$$\sigma_k^2 = C_I(W, W) - \sum_{\tau=1}^T \sum_{\alpha=1}^{N_\tau} \lambda_{\alpha\tau}(z^l) \times C_I(W, W_{\alpha\tau}; z^l) + \mu \quad (18)$$

3.3 中值指示克立格(STMIK)

在指示克立格中, 若选用 L 个指示值 z^l , 那么就

有 L 个相互不同的指示协方差 $C_l(t, h; z^l)$ 。然而有时样品指示协方差或变异函数之间相互成比例, 这时相应的连续随机函数模型 $Z(\tau, x)$ 就是所谓的“镶嵌”(mosaic)模型, 因而:

$$\rho_z(t, h) = \rho_l(t, h; z^l) = \rho_l(t, h; z^l, z^l), \forall z^l, z^l \quad (19)$$

上式中, $\rho_z(t, h)$ 和 $\rho_l(t, h; z^l, z^l)$ 是连续区域化变量的随机函数 $Z(\tau, x)$ 跟它的指示转换函数之间的相关函数和互相关函数。

在这种情况下, 单个的相关函数更好估计, 可直接从区域化变量的相关函数求得, 也可从中央临界值 $[z^l, z^l = M]$ 处样品指示相关函数求得, 此时, $F(M) = 0.5$ 。事实上, 根据定义, 在中央临界值处, 所有特异值的指示数据以 1 和 0 均匀分布。

在模型(19)下的指示克立格称为中值指示克立格, 它只需要一个中值指示变异函数就可满足 L 个临界值的需要。此外, 对所有的临界值, 指示数据的构型相同时, 一个 STIK 方程组就可解出用于所有临界值的权值。如在 STSIK 中:

$$I^*(W; z^l)_{\text{STSIK}} = \sum_{\tau=1}^T \sum_{\alpha=1}^{N_\tau} \lambda_{\alpha\tau}(z^l) I[(\tau, x)_{\alpha}; z^l] + \left[1 - \sum_{\alpha=1}^{N_\tau} \lambda_{\alpha\tau}(z^l) \times F(z^l)\right] \quad (20)$$

其中 $\lambda_{\alpha\tau}(z^l)$ 是估计权值, 它们对所有的临界值 z^l 都相同, 可由一个 STSIK 方程组求出:

$$\sum_{\tau=1}^T \sum_{\alpha=1}^{N_\tau} \lambda_{\alpha\tau}(z^l) \times C_l(W_{\alpha\tau}, W_{\beta\tau}; z^l) = C_l(W, W_{\beta\tau}; z^l) \quad (21)$$

协方差 $C_l(W_{\alpha\tau}, W_{\beta\tau}; z^l)$ 可从区域化变量 z 的样品协方差得到模拟, 更好的办法是根据样品中值指示协方差来进行模拟。

4 多元指示克立格

在一定的研究领域里, 实际研究对象中的各区域化变量大都具有或多或少的联系, 这些区域化变量之间表现出一种协同行为, 绝对不相关的几个区域化变量是很少见的。如在油藏描述中, 剩余油的分布就与孔隙度、渗透率及饱和度等都有关; 在金矿床中某点的金品位常与和金矿床共生或伴生的其它元素品位(如铜品位)密切相关。现代地质统计学考虑到了各区域化变量之间的相互关系, 在时空域中的协同区域化理论的基础上, 充分利用已有的各种信息, 使得研究结果更合理、更稳健、更精确。下面简单地介绍两种多元指示克立格技术。

4.1 指示协同克立格(STCOIK)

设有 K 个区域化变量, 在时刻 τ 处某待估点 W

有 K 个指示值 $[I_k(W; z_k), k = 1, \dots, K]$, 对主变量 k_0 的指示值 STCOIK 估计表达式为:

$$I_{k_0}[W; z_{k_0}]_{\text{STCOIK}} = \sum_{k=1}^K \sum_{\tau=1}^T \sum_{\alpha=1}^{N_{\tau k}} \lambda_{\alpha\tau k} \times I_k[(\tau, x)_{\alpha k}; z_k] \quad (22)$$

相应的协同克立格方程组需要 k^2 个直接、互指示协方差矩阵

$$C_l[(t, h)_{\alpha k}; z_k, z_{k'}] = C_{\alpha k} \{I_k[(\tau, x)_{\alpha k}; z_k], I_{k'}[(\tau + t, x + h)_{\alpha k'}; z_{k'}]\} \quad (23)$$

根据主变量与次级变量的不同作用, 可以选用不同的协同克立格技术来进行各种估计: 如主要考虑主变量的作用时用普通协同克立格; 当主变量与次级变量同样重要时, 则可选用标准协同克立格。

对于 K 较大时, 直接推断和联合模拟 K 个协方差函数就有点不现实。一个解决方案是要求有一个先验二元分布模型; 另一个办法是对那些不具相互关系的指示变量值进行线性转换。

采用先验二元分布模型; 相当于在参考一些或全部指示协方差模型的基础上模拟实际数据。用得最多的二元分布模型是对原始数据进行正态转换后的二元高斯模型: $Z(\tau, x) \rightarrow Y(\tau, x) = \varphi(Z(\tau, x))$ 。通过一些关系式, 所有的指示(互)协方差都可从经正态转换后的 $Y(\cdot)$ 协方差来决定。

4.2 主分量指示克立格(STPCIK)

对时空域中相距 (t_0, h_0) 的指示协方差矩阵 $\{C_l[(t_0, h_0)_{\alpha k}; z_k, z_{k'}], k = 1, \dots, K\}$ 利用克立格求解其主分量而形成 STPCIK, 通常时空域中的空间距离 h_0 和时间距离 t_0 取得很小或为 0。

K 个主分量 $\{Y_k[(\tau, x)_{\alpha k}], k = 1, \dots, K\}$ 是原来的 K 个指示变量的线性组合:

$$Y_k[(\tau, x)_{\alpha k}] = \sum_{k'=1}^K \lambda_{\alpha k, k'} \times I_{k'}[(\tau, x)_{\alpha k'}; z_{k'}] \quad \text{即} \quad Y = A^T \times I \quad (24)$$

其中 $K \times K$ 阶矩阵 $A^T = [\alpha k, k']$ 是正交矩阵 A 的转置矩阵, 正交矩阵 A 是从协方差矩阵进行谱系分解而得到:

$$C_l[(t_0, h_0)_{\alpha k}; z_k, z_{k'}] = A \Lambda A^T \quad (25)$$

Λ 是具 K 个指示协方差矩阵特征的对角矩阵。

通过这种构造, 以时空矢量 (t, h_0) 分隔的 K 个主分量 $Y_k(\tau, x)$ 相互正交:

$$C_{0V}\{Y_k[(\tau, x)_{\alpha k}], Y_{k'}[(\tau + t, x + h)_{\alpha k'}]\} \approx 0, \forall t, h \quad \forall k \neq k' \quad (26)$$

通过协同克立格方程组来求得 K 个主分量 $Y_k(\tau, x)$

的值,这大大减少了用单元克立格来进行估计和推断的花费。

此外,当区域化变量呈连续分布时,将 K 个主分量按时空相关性以降序排列,高阶主分量与纯块金效应行为有关:

$$C_{OV}\{Y_k[(\tau, x)_{ark}], Y_k[(\tau + t, x + h)_{ark}]\} \approx 0, \quad h \neq 0, \quad k \text{ 接近 } K \quad (27)$$

因而,对高阶主分量 $Y_k(\tau, x)$ 进行克立格估计实际就是在滑动研究邻域内对低阶主分量 $Y_k[(\tau, x)_{ark}]$ 的简单平均。那么,原来协同克立格方程组中的 K 个指示值 $I_k[(\tau, x)_{ark}; z_k]$ 就可以减少到按降序排列的前 k' 个, $k' < K$ 。

从指示克立格求出的主分量经线性转换就可以得到相应的对指示值的大致无偏的协同克立格方程组估计。在时刻 τ 外,某待估点 W 的主分量指示克立格估值为:

$$I_k[W; z_k]_{STPCIK}^* = \sum_{k'=1}^k \lambda_{ak, k'} \times Y_k'[(\tau, x)_{ark'}] \\ \text{即 } I^* = A \times Y^* \quad (28)$$

指示克立格主要用在条件模拟中产生条件概率。指示克立格本身并不能保证结果的概率估计、连续变量的 *ccdf* 函数、分类变量的离散概率、序次

相关等方面的正确性,它只能提供这种处理的概率分布。

参考文献

- 1 肖斌,赵鹏大,侯景儒,等.时空多元协同克立格的理论研究.物探化探计算技术,1998,20(1):1~8
- 2 侯景儒.指示克立格理论与方法.地质与勘探,1990,25(3):28~36
- 3 侯景儒,黄竞先,吴雨沛.非参数及多元地质统计学的理论分析及其应用.北京:冶金工业出版社,1992.77~156
- 4 赵鹏大,李紫金,胡旺亮,等.矿床统计预测.北京:地质出版社,1994.47~93
- 5 吴雨沛,侯景儒.非参数地质统计学及其在矿业中的应用.有色金属,1990(2):41~45;1990(3):39~44;1990(4):41~46;1990(5):45~58;1990(6):40~42
- 6 周蒂,陈汉宗.稳健统计学与地球化学数据的统计分析.地球科学,1991,16(3):273~280
- 7 Deutsch CV, Journel AG. GSLIB. Oxford: Oxford University Press, 1992.73~84
- 8 Jacques Rivoinard. Disjunctive kriging and non-linear geostatistics. Oxford: Clarendon Press, 1994.27~67
- 9 Journel AG. Nonparametric estimation of spatial distributions. Math Geol, 1983, 15(3):445~468
- 10 Carr JR, Bailey RE. Use of indicator variograms for an enhanced spatial analysis. Math Geol, 17(8):797~812

THE THEORY OF INDICATOR KRIGING STUDY IN TEMPORAL - SPATIAL DOMAIN

Xiao Bin, Zhao Pengda, Hou Jingru

Indicator Kriging comprehensively in temporal - spatial domain were discussed. It is included: (1) the transformation of indicator Kriging in temporal - spatial domain; (2) second order moment of Indicator Kriging; (3) univariate indicator Kriging system: ① Simple Indicator Kriging (STSIK); ② Ordinary Indicator Kriging (STOIK); ③ Median Indicator Kriging (STMIK); (4) Multi-variate Indicator Kriging: ① Indicator Cokriging (STCOIK); ② Indicator Principal Component Kriging (STPCIK).

Key words temporal - spatial domain, simple indicator kriging, ordinary indicator kriging, median indicator kriging, indicator cokriging, indicator principal component kriging

第一作者简介:

肖斌男, 1969年生。1993年毕业于华东地质学院地质系找矿勘探专业, 1996年在该校获铀矿勘查专业硕士学位, 1999年在中国地质大学(北京)获博士学位, 现在北京大学从事博士后研究。主要从事三维GIS研究和地质统计学的理论与应用工作。

通讯地址: 北京市海淀区 北京大学地质学系 邮政编码: 100871

