

ALGORITMO EFICIENTE PARA PROCESOS DE ESTIMACION GEOESTADISTICOS

EFFICIENT ALGORITHM FOR GEOSTATISTICS ESTIMATION PROCESS

RAFAEL TRUJILLO CODORNIU
MANUEL LORES VIDAL
MIRELIS RASUA LOPEZ

E-mail: rtrujillo@ismm.edu.cu
 Instituto Superior Minero Metalúrgico de Moa

RESUMEN: En el presente artículo se expone un algoritmo que permite disminuir de manera sustancial el número de operaciones usualmente requerido para realizar la estimación del valor de una variable regionalizada, a partir de los valores conocidos de la misma en un conjunto de puntos del espacio o del plano. El algoritmo puede usarse para acelerar la estimación por Kriging, o por el método de las funciones de autocorrelación permitiendo, por tanto, mayor rapidez en la generación de mapas y otros cálculos asociados.

Palabras clave: Geoestadística, Kriging, estimación.

ABSTRACT: In this article is exposed an algorithm that allows to diminish, substantially, the number of operations that are usually required to carry out the estimation of the value of a variable, starting from the known values of the same, in a group of points of the space or the plane. The algorithm can be used to accelerate estimation by Kriging, or by the autocorrelation functions method, and permit, therefore, greater speed in the generation of maps and other associated calculations.

Key words: Geostatistics, Kriging, estimation.

INTRODUCCIÓN

La modelación matemática de un yacimiento es, en esencia, la reconstrucción del cuadro geológico tridimensional del mismo, a partir de datos aislados obtenidos a través del muestreo. Este proceso combina la imaginación con las formulaciones matemáticas idóneas, de modo que con la información disponible (que generalmente es insuficiente debido al costo de su obtención) pueda lograrse un modelo adecuado.

La estimación es una de las partes más importantes del proceso de modelación. La misma implica la aplicación de una serie de algoritmos matemáticos para interpolar o extrapolar un atributo medido sobre un conjunto de puntos, distribuidos espacialmente de forma irregular o muy dispersa, a un nodo de interés. Las técnicas de estimación varían desde la simple triangulación y la aproximación polinomial, hasta algoritmos mucho más complejos, como los que se abordan en la Geoestadística. Los métodos geoestadísticos ofrecen (Deraisme y Fouquet, 1996), como regla, mejores resultados, ya que parten del concepto intuitivo de que mientras más cerca estén 2 puntos dentro del bloque modelado, menor será la variación del atributo medido entre ellos. Es decir, que siempre existe una determinada correlación estadística entre el valor del atributo en un punto y los valores de ese mismo atributo en puntos "vecinos" o "ceranos".¹ Esta suposición, además de ser a todas luces muy lógica, ha sido confirmada por la práctica, y por ello la difusión de los métodos geoestadísticos ha crecido hasta el punto

de convertirse en el sistema predominante actualmente, ya que el uso de los ordenadores permite soslayar lo complejo y laborioso que resulta su aplicación.

Es muy común que a partir de un número finito y relativamente pequeño de valores, se tenga que estimar el atributo en un número grande de puntos. Esto se puede apreciar con claridad cuando se desean obtener mapas o estimar recursos para cuyo cálculo preciso necesitamos de redes muy finas. En estos casos la laboriosidad de los métodos geoestadísticos hacen esta tarea compleja (desde el punto de vista del número de operaciones requerida y del tiempo promedio de su ejecución) y muchos optan (ver, por ejemplo: Legrá y Trujillo, 1997) por utilizar en las etapas finales del refinamiento de la red otros métodos de estimación (de precisión más dudosa, pero de más rápido cálculo). En el presente trabajo exponemos un algoritmo que permite simplificar sustancialmente el proceso de estimación geoestadístico.

DESARROLLO DEL ALGORITMO

Sean $P_1 = P_1(x_1, y_1, z_1)$; $P_2 = P_2(x_2, y_2, z_2)$; ... $P_n = P_n(x_n, y_n, z_n)$ n puntos del Espacio R^3 para los cuales es conocido el valor del atributo v . Es decir, son conocidos $v_1 = v(P_1)$; $v_2 = v(P_2)$; ... $v_n = v(P_n)$. Supongamos que la estimación del valor del

¹ En la geoestadística puede formalizarse matemáticamente el concepto de punto "cerano", pero esto no es necesario para la comprensión del cuadro general que se expone.

atributo v en el punto $P = P(x,y,z)$ se efectúa a través de la media ponderada:

$$v = v(P) = \alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_n v_n = \sum_{k=1}^n \alpha_k v_k \quad (1)$$

donde los ponderadores $\alpha_1, \alpha_2 \dots \alpha_n$ se obtienen como solución del Sistema:

$$\sum_{i=1}^n A_{ki} \alpha_i = b_k; \quad 1 \leq k \leq n \quad (2)$$

Asumiremos que los coeficientes A_{ki} de la matriz del sistema de ecuaciones (2) no dependen de $P = P(x,y,z)$, a diferencia de las coordenadas del vector libre $b_k = b_k(P)$, que sí dependen explícitamente de P . Consideraremos, además, que la matriz A es simétrica (es decir, satisface la condición $A_{ik} = A_{ki}$) y no singular (o sea, con determinante desigual a cero), con lo cual se garantiza que la solución del Sistema (2) existe y es única. Como veremos más adelante, estas suposiciones son válidas en el caso típico de las estimaciones por Kriging o por el método de las funciones de autocorrelación (Bernal, 1994).

El problema consiste en que para estimar el valor del atributo v en cada punto P es necesario resolver un Sistema de Ecuaciones de $n \times n$. Es bien conocido que la solución de un Sistema de Ecuaciones Lineales por los métodos más eficientes, necesita un número de operaciones proporcional al cubo del orden del Sistema, es decir, $O(n^3)$.² Si se necesitan estimar N puntos, se tendrían que realizar $O(Nn^3)$ operaciones.

Para disminuir el número de operaciones que se deben realizar y posibilitar la estimación de todos los puntos en un plazo razonable, se emplean varias técnicas. En primer lugar, se limita el orden n del Sistema de Ecuaciones. Dicho en otros términos, se considera que influyen en el valor del punto P sólo los valores de los puntos más "cercanos". Para esto se especifica un umbral de correlación γ , a partir del mismo, un elipsoide de correlación con centro en P . Es posible demostrar que si un punto se encuentra fuera del elipsoide de correlación, la influencia de su valor en el valor de P no es (de acuerdo al umbral seleccionado) estadísticamente significativa. Si después de esta técnica el número de puntos que se debe considerar en la estimación sigue siendo alto, algunos sistemas hacen una eliminación posterior (ya no tan justificable desde el punto de vista estadístico) hasta que obtengan un valor de n manejable (generalmente entre 8 y 24).

Es fácil ver, además, que como la matriz $A = A_{ik}$ no depende del punto que se va a estimar y es invertible, puede realizarse un efectivo ahorro de operaciones si previamente se invierte la matriz, y el sistema (2) se resuelve por la fórmula:

$$\bar{\alpha} = A^{-1} \bar{b}; \quad \bar{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_n); \quad \bar{b} = (b_1, b_2, \dots, b_n) \quad (3)$$

² La expresión $O(n^3)$ puede interpretarse como una función $f(n)$ tal, que $\lim_{n \rightarrow \infty} \frac{f(n)}{n^3} = C \neq 0$.

Es decir, es una función que crece a la misma velocidad que n^3 .

En efecto, la matriz A tendría que invertirse sólo una vez y luego se emplearía este resultado para las N estimaciones necesarias. Como el producto de una matriz por un vector necesita $O(n^2)$ operaciones, es posible apreciar que por este método se necesitarían $O(Nn^2)$ operaciones.

Pasemos a la descripción del algoritmo que se propone. La fórmula de estimación (1) puede escribirse como el producto escalar de dos vectores:

$$v = v(P) = \alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_n v_n = (\bar{\alpha}, \bar{v}) \quad (4)$$

donde

$$\bar{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_n) \quad \text{y} \quad \bar{v} = (v_1, v_2, \dots, v_n)$$

Si se sustituye en (4) la expresión (3), obtenemos:

$$v = v(P) = (\bar{\alpha}, \bar{v}) = (A^{-1} \bar{b}, \bar{v}) \quad (5)$$

Como es conocido, el producto escalar de vectores reales satisface la siguiente relación de dualidad:

$$(B\bar{x}, \bar{y}) = (\bar{x}, B'\bar{y}) \quad \bar{x}, \bar{y} \in \mathfrak{R}^n \quad (6)$$

donde B es una matriz arbitraria de $n \times n$ y B' la matriz transpuesta a B . Si B es simétrica, se cumple que $B = B'$ y $B^{-1} = (B^{-1})'$, por tanto, la combinación de (5) y (6) conduce a:

$$v = v(P) = (\bar{\alpha}, \bar{v}) = (A^{-1} \bar{b}, \bar{v}) = (\bar{b}, (A^{-1})' \bar{v}) = (\bar{b}, A^{-1} \bar{v}) \quad (7)$$

Sea $\bar{c} = A^{-1} \bar{v}$. Evidentemente el vector \bar{c} no depende de P y puede ser calculado una sola vez en todo el proceso resolviendo el sistema:

$$\sum_{i=1}^n A_{ki} \alpha_i = b_k; \quad 1 \leq k \leq n \quad (8)$$

luego de lo cual para cada punto P donde se desee efectuar la estimación tendríamos:

$$v = v(P) = (\bar{b}, \bar{c}) = (\bar{b}(P), \bar{c}) \quad (9)$$

Note que con esta fórmula de estimación, una vez calculado el vector \bar{b} sólo se debe ejecutar el producto escalar y por ello la estimación de un punto requiere de $O(n)$ operaciones y la de N puntos, de $O(Nn)$ operaciones.

Veamos ahora la factibilidad de este algoritmo para los diferentes tipos de Kriging. Como es conocido (ver, por ejemplo, Chica Olmo, 1988; Legrá y Guardiola, 1999; Legrá et al., 1999), el Kriging es un método de estimación lineal insesgado, óptimo en el sentido que minimiza la varianza de estimación; se han estudiado varios casos, entre los que distinguiremos tres:

1. La Esperanza Matemática de la variable v es nula o conocida a priori y además la covarianza de v es una función de la distancia euclidiana entre dos puntos. En este caso los coeficientes de ponderación α_i se calculan resolviendo el sistema

$$\sum_{i=1}^n A_{ki} \alpha_i = b_k; \quad 1 \leq k \leq n$$

donde A_{ki} es la covarianza entre los puntos P_i y P_k , y b_k es la covarianza entre los puntos P y P_k .

La estimación se efectúa por la fórmula:

$$v = v(P) = M_v + \sum_{k=1}^n \alpha_k (v_k - M_v)$$

Como la covarianza depende de la distancia euclidiana y $d(P_i, P_k) = d(P_k, P_i)$, la matriz A es simétrica y no depende de P, por lo que se cumplen los postulados planteados al inicio para la aplicación del algoritmo. En efecto

$$v = v(P) = M_v + (\bar{\alpha}, \bar{v}') = M_v + (A^{-1} \bar{b}, \bar{v}') = M_v + (\bar{b}, A^{-1} \bar{v}') = M_v + (\bar{b}, \bar{c}')$$

donde:

$$\bar{v}' = (v_1 - M_v, v_2 - M_v, \dots, v_n - M_v) \text{ y } \bar{c}' = A^{-1} \bar{v}'$$

Puede apreciarse que determinando al inicio de los cálculos el vector \bar{c}' , garantizamos que el resto de las estimaciones se efectúen calculando un simple producto escalar (por supuesto, una vez determinado el vector \bar{b}).

2. La Esperanza Matemática de v es constante, pero desconocida $E(v(P)) = m$, y la covarianza es una función de la distancia euclidiana entre dos puntos. En este caso los coeficientes de ponderación α_i se calculan resolviendo el sistema:

$$\begin{cases} \sum_{i=1}^n A_{ki} \alpha_i = b_k + \mu \\ \sum_{i=1}^n \alpha_i = 1 \end{cases} \quad (10)$$

y la estimación se hace por la fórmula (1), donde A_{ki} es la covarianza entre los puntos P_i y P_k , y b_k es la covarianza entre los puntos P y P_k . Observe que en este caso el sistema de ecuaciones lineales tiene orden $n + 1$, ya que aparece una incógnita auxiliar, que es el multiplicador de Lagrange μ . Si se efectúa un cambio de variable $\mu = -\mu$, se puede escribir el sistema (10) en la forma:

$$\begin{cases} \sum_{i=1}^n A_{ki} \alpha_i + \mu = b_k \\ \sum_{i=1}^n \alpha_i = 1 \end{cases} \quad (11)$$

cuya matriz es evidentemente:

$$\tilde{A} = \begin{pmatrix} A_{11} & A_{12} & \dots & A_{1n} & 1 \\ A_{21} & A_{22} & \dots & A_{2n} & 1 \\ \vdots & \vdots & \dots & \vdots & \vdots \\ A_{n1} & A_{n2} & \dots & A_{nn} & 1 \\ 1 & 1 & \dots & 1 & 0 \end{pmatrix}$$

la cual es simétrica en virtud de la simetría de A. Luego el algoritmo expuesto puede aplicarse si utilizan las siguientes notaciones:

$$\begin{aligned} \bar{v}' &= (v_1, v_2, \dots, v_n, 0) \in \mathfrak{R}^{n+1} \\ \bar{b}' &= (b_1, b_2, \dots, b_n, 1) \in \mathfrak{R}^{n+1} \\ \bar{\alpha}' &= (\alpha_1, \alpha_2, \dots, \alpha_n, \mu) \in \mathfrak{R}^{n+1} \end{aligned}$$

En efecto:

$$v = v(P) = \sum_{k=1}^n \alpha_k v_k = (\bar{\alpha}, \bar{v}') = (\bar{\alpha}', \bar{v}') = (\tilde{A}^{-1} \bar{b}', \bar{v}') = (\bar{b}', \tilde{A}^{-1} \bar{v}') = (\bar{b}', \bar{c}')$$

donde el vector $\bar{c}' \in \mathfrak{R}^{n+1}$ se calcula resolviendo el Sistema: $\tilde{A} \bar{c}' = \bar{v}'$.

3. La variable v es una variable estrictamente intrínseca, sin desviación, dotada de un variograma g , pero no de covarianza. En este caso los coeficientes de ponderación α_i se calculan resolviendo el sistema:

$$\begin{cases} \sum_{i=1}^n A_{ki} \alpha_i + \mu = b_k \\ \sum_{i=1}^n \alpha_i = 1 \end{cases}$$

y la estimación se hace por la fórmula (1), donde:

$$A_{ki} = \gamma(d(P_k, P_i)); \quad b_k = \gamma(d(P_k, P))$$

Es fácil apreciar que este caso puede resolverse de manera análoga al caso 2 sin necesidad de efectuar cambios de variable.

CONCLUSIONES

El algoritmo propuesto es sencillo y permite disminuir el número de operaciones necesarias para efectuar estimaciones múltiples, partiendo de la posibilidad de concentrar en el vector \bar{c}' toda la información que depende de la distribución espacial de los puntos P_i y dejando en el vector $\bar{b} = \bar{b}(P)$ la información que depende específicamente de la posición del punto que se va a estimar. La disminución de la cantidad de operaciones permite trabajar con un número mayor de puntos y utilizar los métodos geoestadísticos, incluso en las últimas etapas de refinamiento de la red.

BIBLIOGRAFÍA

BERNAL, H., S.: "La formación de la calidad de la mena en la planificación operativa de los trabajos mineros en la cantera Moa", *Minería y Geología*, 1 (11) :47-49, 1994.

CHICA OLMO, MARIO: *Análisis Geoestadístico en el estudio de la explotación de los Recursos Minerales*, Universidad de Granada, España, 1988.

DERAISME, J. Y CH. DE FOUQUET: "The geostatistical approach for reserves", *Mining Magazine*, 174 (5) :309-313, May, 1996.

LEGRÁ, A. A. Y R. A. TRUJILLO: "Modelación de mediciones geólogo-mineras sobre una red rectangular mediante spline cúbico natural", *Minería y Geología*, 2 (14) : 15-16, 1997.

LEGRÁ, A.A. Y R.L. GUARDIOLA: "Contribución a la Práctica del Análisis Variográfico y de la Estimación por Kriging", *Minería y Geología*, 2 (16) :83-93, 1999.

LEGRÁ, A. A.; O.R. SILVA Y O. BELETE: "Modelación de una Superficie Topográfica a partir de la Relación entre el Kriging y la Interpolación Lineal en R^m ", *Minería y Geología*, 1 (16) :58-61, 1999.