



Análisis de Señales en Geofísica

15° Clase Geoestadística

Prof. Ricardo C. Rebollo



Facultad de Ciencias Astronómicas y Geofísicas,
Universidad Nacional de La Plata, Argentina





Introducción:

La geoestadística, a diferencia de la estadística clásica, nos ofrece una forma de describir la continuidad espacial de los datos, que es una característica fundamental de innumerables procesos naturales, proveyéndonos de métodos de adaptación de las técnicas clásicas de regresión para incluir en ellas la información disponible de la continuidad espacial. Esta información es incorporada en los métodos de estimación mediante alguna de las siguientes funciones: covarianza, variograma o correlograma.

En las ciencias de la Tierra es habitual trabajar con variables definidas en todos los puntos de una determinada región, es decir, variables definidas en un dominio continuo Ω . Sin embargo, habitualmente sólo disponemos de observaciones de estas variables en algunos puntos aislados de la región.

El problema que se nos plantea, es cómo estimar estas cantidades en aquellos puntos donde no fueron observadas. Las variables con las que tratamos son variables determinísticas, y son función de las coordenadas espaciales que definen su ubicación, motivo por el cual lo deseable sería proponer un modelo determinístico que nos permita predecir los valores en los puntos no observados.



Introducción:

Por lo general el proceso físico por el cual se generaron estas cantidades es sumamente complejo y no disponemos de la información, ni de los conocimientos necesarios, para caracterizar un modelo determinístico apropiado.

La incertidumbre de lo que ocurre en los puntos no observados es extremadamente elevada. Por esta razón se hace necesario un enfoque geoestadístico del método de estimación basado en modelos probabilísticos que reconozcan la existencia de estas incertidumbres inevitables.

En un modelo probabilístico la variable determinística en cuestión es considerada una variable aleatoria $V(x)$ y los datos disponibles $v(x_i)$ son considerados realizaciones puntuales de un proceso aleatorio continuo. En ningún momento debemos perder de vista que este modelo está en conflicto con la realidad. El proceso físico determinístico generador de nuestra variable, es extremadamente complicado y nuestro entendimiento del mismo es tan pobre que dada su complejidad se nos presenta en apariencia con un comportamiento aleatorio, pero esto no significa que el proceso sea aleatorio, sólo significa que somos ignorantes.



Covarianza, Variograma y Correlograma:

La información de la continuidad espacial de nuestros datos se puede definir por medio de alguna de las siguientes funciones:

➤ *Covarianza:*

$$Cov(V(x_i), V(x_j)) = E\{(V(x_i) - E\{V(x_i)\})(V(x_j) - E\{V(x_j)\})\}$$

➤ Variograma o Semivariograma:

$$\gamma(V(x_i), V(x_j)) = \frac{1}{2} E\{(V(x_i) - V(x_j))^2\}$$

➤ Correlograma:

$$\rho(V(x_i), V(x_j)) = \frac{Cov(V(x_i), V(x_j))}{\sqrt{Var(V(x_i))Var(V(x_j))}}$$

La varianza de $V(x)$ está dada por:

$$Var(V(x_i)) = Cov(V(x_i), V(x_i)) = E\{(V(x_i) - E\{V(x_i)\})^2\} = \sigma_{V(x_i)}^2$$



Kriging Ordinario:

El método de estimación de kriging ordinario es el mejor estimador lineal no sesgado. Es lineal porque el valor estimado es obtenido como un promedio pesado de los datos disponibles. Es no sesgado porque la estimación del error medio residual m_R es cero. Es el mejor, u óptimo en el sentido de los cuadrados mínimos, porque minimiza la varianza σ_R^2 de los errores de estimación del modelo probabilístico:

$$\hat{v}(x_0) = \sum_{i=1}^N w_i v(x_i) \quad \textit{estimador lineal}$$
$$m_R = 0 \quad \textit{estimador no sesgado}$$
$$\sigma_R^2 \textit{ mínima} \quad \textit{estimador óptimo}$$

Utilizaremos como modelo de nuestros datos una función aleatoria $V(x)$. Los N datos observados serán considerados como una realización $v(x_i)$ de la variable aleatoria $V(x_i)$ para $i = 1, N$. Esta función aleatoria nos permitirá expresar el error, su valor medio y su varianza, como promedios estadísticos.



Kriging Ordinario:

En cada punto x_0 de la región de trabajo donde no tenemos un valor observado $v(x_0)$, estimaremos el valor desconocido mediante un promedio pesado $\hat{v}(x_j)$ de los datos disponibles.

Si definimos el error $r(x_j)$ de un determinado valor estimado en el punto x_j como la diferencia entre el valor estimado y el valor no observado en la misma ubicación, tenemos:

$$r(x_j) = \hat{v}(x_j) - v(x_j)$$

Entonces el valor medio del error de K estimadores estará dado por:

$$m_r = \frac{1}{K} \sum_{j=1}^K r(x_j) = \frac{1}{K} \sum_{j=1}^K (\hat{v}(x_j) - v(x_j))$$

Desafortunadamente, no podemos hacer demasiado con esta ecuación ya que involucra cantidades que no conocemos, como son los valores verdaderos $v(x_1), v(x_2), \dots, v(x_K)$, en los puntos de estimación.

La solución probabilística de este problema consiste en la conceptualización de los valores desconocidos como realizaciones de un proceso aleatorio, y resolver el problema para el modelo conceptual.



Kriging Ordinario:

Si el modelo probabilístico es estacionario de segundo orden las variables aleatorias tendrán el mismo valor medio y la misma varianza en todos los puntos x de la región. El valor esperado de la variable aleatoria estará dado por $E\{V(x)\} = m_V$. Todo par de variables aleatorias estacionarias tendrá una única función de distribución de probabilidades conjunta que dependerá exclusivamente de la separación entre las dos variables y no de sus posiciones absolutas. La covarianza entre dos variables aleatorias separadas por una distancia h , estará dada por $C(h)$:

$$C(h) = Cov(V(x), V(x+h)) = E\{[V(x) - E\{V(x)\}][V(x+h) - E\{V(x+h)\}]\}$$

$$C(h) = E\{[V(x) - m_V][V(x+h) - m_V]\} = E\{V(x)V(x+h)\} - m_V^2$$

Cada valor de este modelo probabilístico es visto como una realización de una variable aleatoria. Los mismos valores observados son considerados realizaciones de variables aleatorias, como también lo son los valores exactos en los puntos no observados. Los valores estimados en esos puntos también serán realizaciones de variables aleatorias, puesto que son un promedio pesado de las variables aleatorias consideradas en los puntos observados:

$$\hat{V}(x_0) = \sum_{i=1}^N w_i V(x_i)$$



Kriging Ordinario:

Análogamente, el error de estimación $R(x_0)$, definido como la diferencia entre la variable aleatoria asociada al valor estimado y la variable aleatoria de la cual el valor real es una realización, es también una variable aleatoria:

$$R(x_0) = \hat{V}(x_0) - V(x_0)$$

Sustituyendo en esta última ecuación la ecuación (1), podemos expresar $R(x_0)$ en función de las $N + 1$ variables aleatorias involucradas en el problema planteado:

$$R(x_0) = \sum_{i=1}^N w_i V(x_i) - V(x_0)$$
$$m_R = E\{R(x_0)\} = E\left\{\sum_{i=1}^N w_i V(x_i) - V(x_0)\right\} = \sum_{i=1}^N w_i E\{V(x_i)\} - E\{V(x_0)\}$$

Como asumimos que la función aleatoria $V(x)$ es estacionaria, su esperanza matemática $E\{V(x)\}$ es la misma en todos los puntos de la región, entonces podemos escribir:

$$m_R = E\{R(x_0)\} = \sum_{i=1}^N w_i E\{V\} - E\{V\} = E\{V\} \left(\sum_{i=1}^N w_i - 1 \right)$$



Kriging Ordinario:

Un estimador no sesgado es aquel para el cual la esperanza matemática del error de estimación es igual a cero.

El valor de la esperanza matemática del error m_R en cualquier punto de la región es habitualmente referido como el sesgo (*bias*) del valor estimado. Como queremos que nuestro estimador sea no sesgado (*unbiased*), al igualar esta esperanza matemática a cero:

$$m_R = E\{R(x_0)\} = E\{V\} \left(\sum_{i=1}^N w_i - 1 \right) = 0$$

Obtenemos:

$$\sum_{i=1}^N w_i - 1 = 0$$

Es decir que al imponer la condición de que la suma de los pesos sea igual a 1, estamos obteniendo un estimador no sesgado de nuestra variable.



La varianza del error del modelo probabilístico:

Obtengamos una expresión para la varianza del error del modelo probabilístico:

$$\sigma_R^2 = \text{Var}\{R(x_0)\} = \text{Var}\left\{\left(\hat{V}(x_0) - V(x_0)\right)\right\}$$

$$\sigma_R^2 = E\left\{\left[\left(\hat{V}(x_0) - V(x_0)\right) - E\{\hat{V}(x_0) - V(x_0)\}\right]^2\right\} =$$

$$= \text{Cov}\{\hat{V}(x_0), \hat{V}(x_0)\} - \text{Cov}\{\hat{V}(x_0), V(x_0)\} - \text{Cov}\{V(x_0), \hat{V}(x_0)\} + \text{Cov}\{V(x_0), V(x_0)\}$$

$$\sigma_R^2 = \text{Var}\{\hat{V}(x_0)\} - 2 \text{Cov}\{\hat{V}(x_0), V(x_0)\} + \text{Var}\{V(x_0)\}$$

La varianza de una combinación lineal de variables aleatorias está dada por:

$$\text{Var}\{\hat{V}(x_0)\} = \text{Var}\left\{\sum_{i=1}^N w_i V(x_i)\right\} = \sum_{i=1}^N \sum_{j=1}^N w_i w_j \text{Cov}\{V(x_i), V(x_j)\}$$

Para simplificar la notación no escribiremos más el vector posición x con su subíndice entre paréntesis sino solamente el subíndice.



La varianza del error del modelo probabilístico:

$$\sigma_R^2 = Var\{\hat{V}_0\} - 2 Cov\{\hat{V}_0, V_0\} + Var\{V_0\}$$

Como asumimos que nuestras variables aleatorias son estacionarias, entonces todas tendrán el mismo valor medio y la misma varianza:

$$Var\{V_0\} = Var(V_i) = \sigma_V^2$$

Además podemos escribir:

$$\begin{aligned} Cov\{\hat{V}_0, V_0\} &= Cov\left\{\left(\sum_{i=1}^N w_i V_i\right), V_0\right\} = E\left\{\sum_{i=1}^N w_i V_i V_0\right\} - E\left\{\sum_{i=1}^N w_i V_i\right\} E\{V_0\} = \\ &= \sum_{i=1}^N w_i E\{V_i V_0\} - \sum_{i=1}^N w_i E\{V_i\} E\{V_0\} = \sum_{i=1}^N w_i Cov\{V_i, V_0\} = \sum_{i=1}^N w_i C_{i0} \end{aligned}$$

Combinando estas expresiones obtenemos la siguiente expresión para la varianza del error:

$$\sigma_R^2 = \sigma_V^2 + \sum_{i=1}^N \sum_{j=1}^N w_i w_j C_{ij} - 2 \sum_{i=1}^N w_i C_{i0}$$



El parámetro de Lagrange:

Para imponer la condición de que el estimador sea no sesgado vamos a introducir en la expresión de σ_R^2 , un término nulo escalado por una nueva incógnita μ conocida como multiplicador o parámetro de Lagrange:

$$\sigma_R^2 = \sigma_V^2 + \sum_{i=1}^N \sum_{j=1}^N w_i w_j C_{ij} - 2 \sum_{i=1}^N w_i C_{i0} + 2\mu \left(\sum_{i=1}^N w_i - 1 \right)$$

Este último término no modifica el valor de la varianza del error ya que es nulo, pero impone a la solución buscada la condición de no sesgada.

Ahora debemos minimizar la varianza del error derivándola respecto de los N pesos w_i y respecto del parámetro de Lagrange μ , obteniendo un sistema de $N + 1$ ecuaciones lineales con $N + 1$ incógnitas, los N pesos w_i y el parámetro μ .

$$\frac{\partial \sigma_R^2}{\partial \mu} = \frac{\partial \left(2\mu \left(\sum_{i=1}^N w_i - 1 \right) \right)}{\partial \mu} = 2 \sum_{i=1}^N w_i - 2 = 0$$
$$\sum_{i=1}^N w_i = 1$$

Es decir que al derivar la varianza del error respecto del parámetro μ obtenemos la condición de solución no sesgada.



Minimización de la varianza del error:

El primer término de la varianza del error σ_R^2 es constante e igual a la varianza de la variable aleatoria, por lo tanto sus derivada respecto de los pesos y del parámetro μ son nulas.

Veamos en detalle la diferenciación de la varianza del error σ_R^2 respecto de w_1 , las derivadas respecto de los pesos restantes pueden ser calculadas de manera análoga. Calculemos la derivada de a un término a la vez; comencemos con el segundo término:

$$\frac{\partial(\sum_{i=1}^N \sum_{j=1}^N w_i w_j C_{ij})}{\partial w_1} = \frac{\partial(w_1^2 C_{11} + 2w_1 \sum_{j=2}^N w_j C_{1j})}{\partial w_1} = 2w_1 C_{11} + 2 \sum_{j=2}^N w_j C_{1j} = 2 \sum_{j=1}^N w_j C_{1j}$$

En la sumatoria del tercer término, sólo uno de sus términos involucra al peso w_1 :

$$2 \frac{\partial(\sum_{i=1}^N w_i C_{i0})}{\partial w_1} = 2 \frac{\partial(w_1 C_{10})}{\partial w_1} = 2C_{10}$$

Finalmente en la sumatoria del último término, encontramos análogamente que un único término de esta sumatoria involucra al peso w_1 :

$$2 \frac{\partial(\mu(\sum_{i=1}^N w_i - 1))}{\partial w_1} = 2 \frac{\partial(\mu w_1)}{\partial w_1} = 2\mu$$



Minimización de la varianza del error:

Juntando las derivadas de los diferentes términos podemos escribir la derivada de la varianza del error respecto de w_1 , del siguiente modo:

$$\frac{\partial \sigma_R^2}{\partial w_1} = 2 \sum_{j=1}^N w_j C_{1j} - 2C_{10} + 2\mu$$

Igualando esta derivada a cero obtenemos la siguiente ecuación:

$$\sum_{j=1}^N w_j C_{1j} + \mu = C_{10}$$

Calculando de manera análoga las derivadas de σ_R^2 respecto de los otros pesos y considerando su derivada respecto del multiplicador de Lagrange, obtenemos el siguiente sistema de $N + 1$ ecuaciones con $N + 1$ incógnitas:

$$\left\{ \begin{array}{l} \sum_{j=1}^N w_j C_{ij} + \mu = C_{i0} \quad \forall i = 1, N \\ \sum_{i=1}^N w_i = 1 \end{array} \right.$$

Este sistema de ecuaciones es habitualmente conocido como sistema de ecuaciones de kriging ordinario y nos permite obtener los pesos que minimizan σ_R^2 .



Sistema de Ecuaciones de Kriging Ordinario:

El sistema de ecuaciones de kriging ordinario puede ser escrito en notación matricial del siguiente modo:

$$\begin{bmatrix} C_{11} & \cdots & C_{1N} & 1 \\ \vdots & \ddots & \vdots & \vdots \\ C_{N1} & \cdots & C_{NN} & 1 \\ 1 & \cdots & 1 & 0 \end{bmatrix} \begin{bmatrix} w_1 \\ \vdots \\ w_N \\ \mu \end{bmatrix} = \begin{bmatrix} C_{10} \\ \vdots \\ C_{N0} \\ 1 \end{bmatrix}$$

$\mathbf{C} \mathbf{W} = \mathbf{D}$

La solución estará dada por:

$$\mathbf{W} = \mathbf{C}^{-1} \mathbf{D}$$

Para resolver este sistema y obtener los N pesos w_i y el parámetro de Lagrange incluidos en el vector columna \mathbf{W} , debemos definir los $N^2 + N$ valores de covarianza involucrados en el sistema de ecuaciones incluidos en la matriz \mathbf{C} y en el vector columna \mathbf{D} .

Los valores de la covarianza son obtenidos definiendo una función parametrizada $C(h)$, donde $h = |x_j - x_i|$ es un escalar si la continuidad espacial es isotrópica, o $h = x_j - x_i$ es un vector si la continuidad espacial es anisotrópica.

Los pesos que nos da la solución de este sistema, nos permitirán estimar un valor no sesgado de la variable en cuestión en el punto genérico x_0 como una suma pesada de los valores conocidos. Kriging ordinario es un interpolador exacto si x_0 coincide con un punto observado, entonces el valor estimado será igual al valor observado.



Varianza de Kriging Ordinario:

La varianza σ_R^2 del error del valor estimado será mínima y es posible calcularla. Con este propósito tomemos del sistema de ecuaciones de kriging ordinario de la diapositiva 14 las siguientes ecuaciones:

$$\sum_{j=1}^N w_j C_{ij} + \mu = C_{i0} \quad \forall i = 1, N$$

Multipliquemos cada una de estas N ecuaciones por el peso w_i :

$$w_i \left(\sum_{j=1}^N w_j C_{ij} + \mu \right) = w_i C_{i0} \quad \forall i = 1, N$$

A continuación sumemos estas N ecuaciones:

$$\sum_{i=1}^N w_i \sum_{j=1}^N w_j C_{ij} + \sum_{i=1}^N w_i \mu = \sum_{i=1}^N w_i C_{i0}$$
$$\sum_{i=1}^N w_i \sum_{j=1}^N w_j C_{ij} = \sum_{i=1}^N w_i C_{i0} - \mu \sum_{i=1}^N w_i$$

Como los pesos suman 1, el último término es simplemente μ .



Varianza de Kriging Ordinario:

$$\sum_{i=1}^N \sum_{j=1}^N w_i w_j C_{ij} = \sum_{i=1}^N w_i C_{i0} - \mu$$

Sustituyendo esta ecuación en la expresión de la varianza del error σ_R^2 de la diapositiva 11 ,obtenemos:

$$\begin{aligned} \sigma_R^2 &= \sigma_V^2 + \sum_{i=1}^N w_i C_{i0} - \mu - 2 \sum_{i=1}^N w_i C_{i0} \\ \sigma_R^2 &= \sigma_V^2 - \left(\sum_{i=1}^N w_i C_{i0} + \mu \right) = \sigma_{KO}^2 \end{aligned}$$

O en forma matricial:

$$\sigma_{KO}^2 = \sigma_V^2 - \mathbf{W} \cdot \mathbf{D}$$

Esta varianza minimizada del error es habitualmente llamada varianza de kriging ordinario, utilizando la notación σ_{KO}^2 . No debemos olvidar que esta varianza fue calculada a partir de un modelo probabilístico y no se corresponde con la verdadera varianza del error σ_r^2 que no nos es posible calcular.



Relación entre la covarianza C_{ij} y el variograma γ_{ij} :

Cuando derivamos la expresión para la varianza del error, asumimos que nuestro modelo probabilístico era estacionario de segundo orden. Esta hipótesis nos permite establecer la siguiente relación entre la covarianza y el correlograma:

$$\begin{aligned}\gamma_{ij} &= \frac{1}{2} E \left\{ (V_i - V_j)^2 \right\} = \frac{1}{2} E \{ V_i^2 \} + \frac{1}{2} E \{ V_j^2 \} - E \{ V_i V_j \} \\ \gamma_{ij} &= E \{ V^2 \} - E \{ V_i V_j \} = E \{ V^2 \} - m_V^2 - (E \{ V_i V_j \} - m_V^2) \\ \gamma_{ij} &= \sigma_V^2 - C_{ij} \quad \Rightarrow \quad C(h) = \sigma_V^2 - \gamma(h) = \gamma(\infty) - \gamma(h)\end{aligned}$$

La relación entre la covarianza y el correlograma será:

$$\rho_{ij} = \frac{C_{ij}}{\sqrt{C_{ii}C_{jj}}} = \frac{C_{ij}}{\sigma_V^2}$$

Tengamos presente que estas relaciones son válidas para el modelo probabilístico con valor medio y varianza constantes. Esto no implica que las mismas relaciones existan entre la covarianza, el variograma y el correlograma de nuestros datos reales. Estas hipótesis son las que nos permitieron expresar las ecuaciones de kriging ordinario en función del variograma o del correlograma. La práctica común en geoestadística es parametrizar el variograma $\gamma(h)$, y luego por razones de eficiencia computacional resolver el sistema utilizando la covarianza $C(h)$.



Interpretación intuitiva de Kriging Ordinario:

Es esencial entender intuitivamente el rol que cumplen la matriz \mathbf{C} y el vector columna \mathbf{D} en el cálculo de los pesos $\mathbf{W} = \mathbf{C}^{-1}\mathbf{D}$ que intervendrán en la obtención del valor estimado:

$$\hat{v}(x_0) = \sum_{i=1}^N w_i v(x_i)$$

Comprender el rol de \mathbf{C} y \mathbf{D} nos permitirá hacer los ajustes necesarios en los parámetros del variograma para mejorar la estimación.

Haciendo una comparación de kriging ordinario con el método de estimación por medio de la suma pesada con las inversas de las distancias, podemos decir que el vector \mathbf{D} contiene información análoga a las inversas de las distancias, sólo que en vez de pensar en distancias geométricas son distancias estadísticas o variabilidad espacial de nuestro observable. La matriz \mathbf{C} en aquel caso sólo sería la matriz identidad \mathbf{I} multiplicada por la constante de normalización que garantiza que la suma de los pesos sea igual a la unidad.

En kriging ordinario, la matriz \mathbf{C} contiene las inversas de las distancias estadísticas entre los puntos con valores observados. Lo que distingue a kriging ordinario no es el uso de distancias estadísticas, sino que multiplicar a \mathbf{D} por \mathbf{C}^{-1} además de normalizar, está compensando a la solución por la redundancia de los datos observados producida por la manera en que los datos están agrupados y distribuidos espacialmente entorno al punto de estimación (*declustering*).



Parámetros del variograma:

Veamos como los diferentes parámetros del variograma afectan a los pesos de kriging ordinario. Aun cuando los datos estén regularmente grillados y tengan un buen comportamiento, ajustar un variograma parametrizado con el variograma de la muestra o variograma experimental, involucra importantes decisiones del interprete. El variograma experimental no provee información para distancias cortas menores que la distancia mínima existente entre los datos observados. Dependiendo de lo exhaustivo que sean los datos observados tampoco nos proveerá información para distancias intermedias. El comportamiento del variograma en un entorno próximo al origen, incluyendo el llamado efecto pepita (*nugget effect*), no puede ser determinado con el variograma experimental, sin embargo este comportamiento tiene una inmensa influencia en el resultado de la estimación.

Dos de los parámetros que debemos definir del variograma son el umbral (*sill*) y el rango (*rank*). El umbral es el valor máximo que alcanza el variograma y es igual a la varianza de la variable observada σ_V^2 . El rango a es la distancia a la cual el variograma alcanza el valor del umbral o el 95% del valor del umbral si el variograma tiende asintóticamente a dicho valor.

Parámetros del variograma:

Los tres modelos de variogramas habitualmente utilizados son:

➤ Modelo esférico:

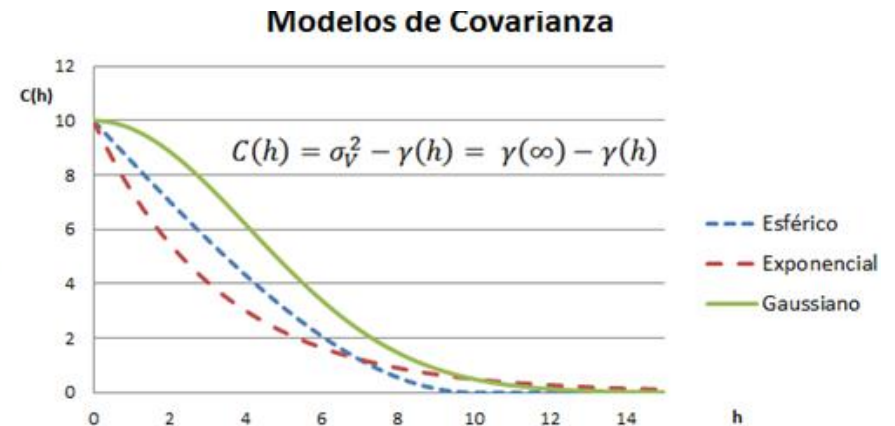
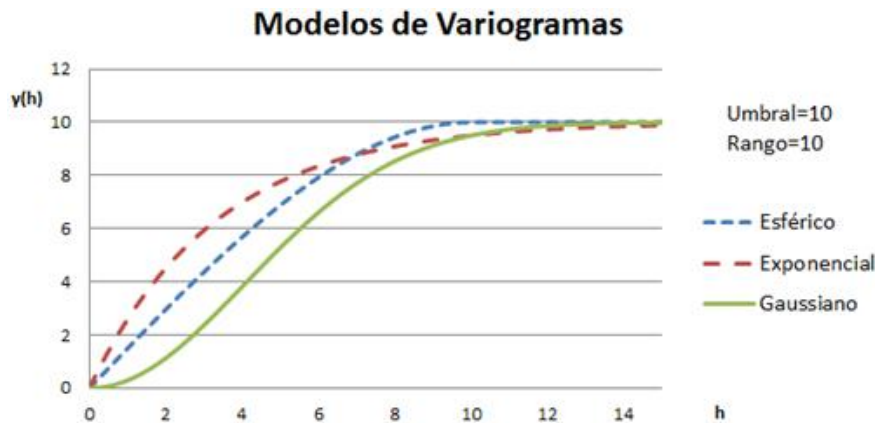
$$\gamma(h) = \begin{cases} \sigma_V^2 \left(1.5 \left(\frac{|h|}{a} \right) - 0.5 \left(\frac{|h|}{a} \right)^3 \right) & \text{si } |h| \leq a \\ \sigma_V^2 & \text{si } |h| > a \end{cases}$$

➤ Modelo exponencial:

$$\gamma(h) = \sigma_V^2 \left(1 - e^{-3 \frac{|h|}{a}} \right)$$

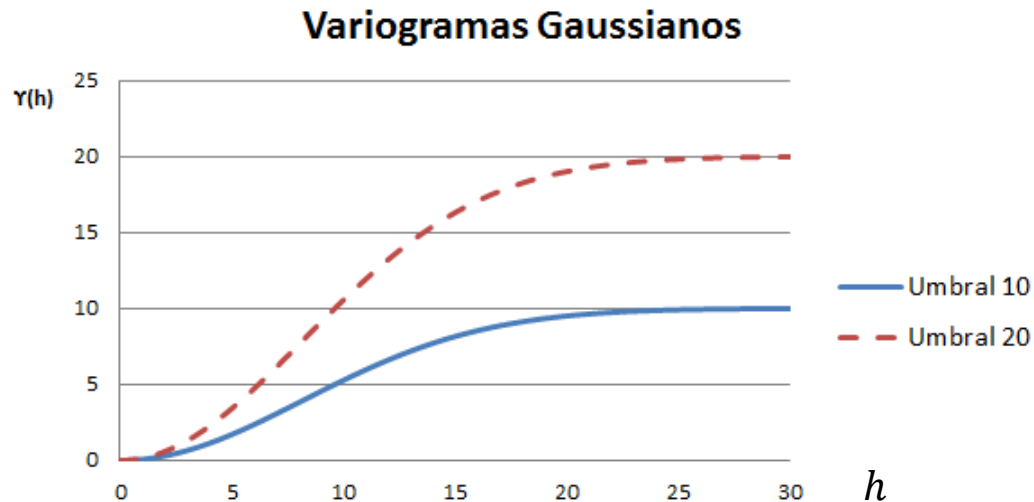
➤ Modelo gaussiano:

$$\gamma(h) = \sigma_V^2 \left(1 - e^{-3 \left(\frac{h}{a} \right)^2} \right)$$



Efecto de un factor de escala:

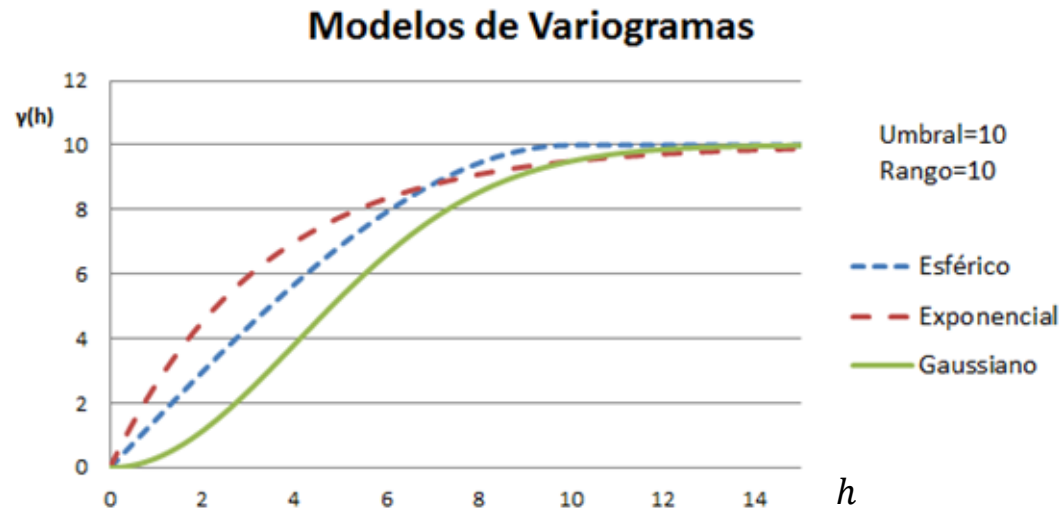
La siguiente figura muestra dos variogramas gaussianos que difieren solamente en un factor de escala, lo que produce un cambio en el valor del umbral:



Multiplicar los variogramas por un factor de escala no tiene incidencia en el cálculo de los pesos de kriging ordinario debido a que los pesos están normalizados, sin embargo sí tiene efecto en el cálculo de la varianza σ_{K0}^2 de kriging ordinario. Mientras que el valor estimado no cambia, la estimación de la varianza es modificada por el mismo factor que fue usado para escalar el variograma.

Efecto de la forma del variograma:

La siguiente figura muestra tres variogramas que poseen el mismo umbral y el mismo rango, pero difieren en la forma:



El variograma gaussiano varía más suavemente que los otros en un entorno del origen, con lo que le da más peso a los puntos más cercanos al punto de estimación, lo cual implica una mayor continuidad espacial de la solución, es decir los valores estimados variarán más suavemente con un variograma gaussiano que con uno exponencial o uno esférico.



Efecto pantalla:

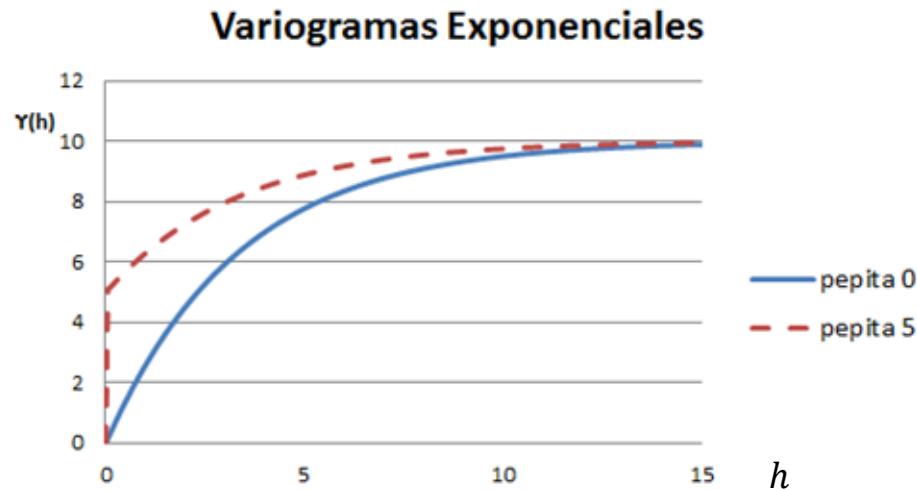
Cuando mirando desde el punto de estimación, vemos un punto observado escondido detrás de otro punto observado, es probable, especialmente con el variograma gaussiano, que el punto más alejado reciba un peso negativo, este efecto es denominado efecto pantalla (*screen effect*). Este efecto es parte de lo que la multiplicación de D por C^{-1} produce. El grado con el cual una observación escondida detrás de otra pierde su influencia en la estimación depende del patrón de continuidad espacial definido por la forma del variograma. El uso de variogramas que varíen suavemente cerca del origen producirá que el efecto pantalla sea mucho más pronunciado, frecuentemente asignando pesos negativos a los puntos observados escondidos.

La ventaja de métodos de estimación que asignen pesos menores que cero o mayores que uno, respetando la condición de que la suma de los pesos sea igual a la unidad (condición de estimador no sesgado) es que pueden conducir a valores estimados mayores al mayor valor observado o menores que el menor valor observado. Es razonable pensar que los valores verdaderos que estamos estimando puedan sobrepasar los límites de los valores observados.

La desventaja es que existe la posibilidad de estimar valores negativos cuando la variable estimada es necesariamente positiva. En estos casos está perfectamente justificado asignarles a estos valores negativos estimados el valor cero.

Efecto pepita:

La siguiente figura muestra dos variogramas que difieren solamente en que el de línea roja a trazos posee un salto en el origen del 50% del valor del umbral, este salto es denominado por los ingenieros en minas efecto pepita:



$$\gamma_1(h) = 10\left(1 - e^{-3\frac{|h|}{10}}\right)$$

$$\gamma_2(h) = \begin{cases} 0 & \text{si } h = 0 \\ 5 + 5\left(1 - e^{-3\frac{|h|}{10}}\right) & \text{si } h > 0 \end{cases}$$

Los pesos calculados utilizando $\gamma_2(h)$ son más similares entre ellos que los calculamos utilizando $\gamma_1(h)$. Dándole un valor mayor a este parámetro logramos que el valor estimado sea más similar a un simple promedio de los datos disponibles. Además produce que sea mayor la varianza σ_{K0}^2 de kriging ordinario.



Efecto pepita puro:

Si el variograma representa un efecto pepita puro, tendrá la siguiente expresión:

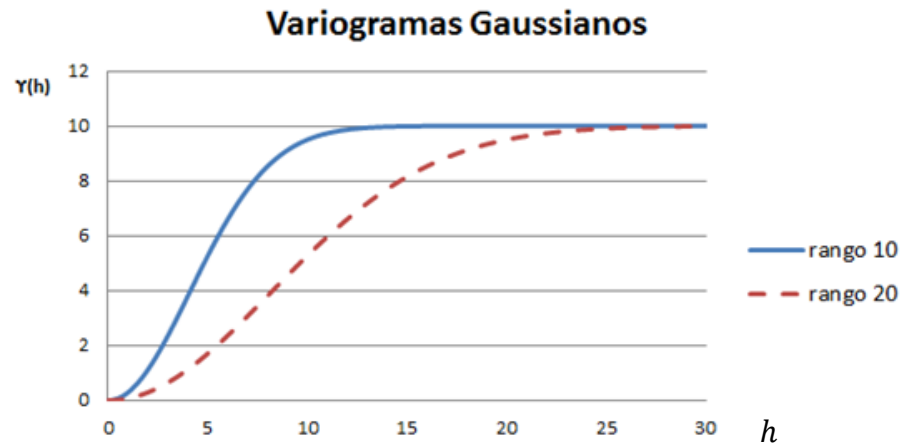
$$\gamma(h) = \begin{cases} 0 & \text{si } h = 0 \\ C_0 & \text{si } h > 0 \end{cases}$$

Este variograma no tiene en cuenta la redundancia, ni la distribución geométrica, ni el agrupamiento de los valores observados. En términos de distancias estadísticas todos los puntos observados están a la misma distancia del punto estimado. Con este modelo probabilístico todos los pesos tomarán el valor $1/N$.

El efecto pepita puro implica una ausencia total de correlación espacial de nuestros valores observados, los valores de nuestros datos no guardan ninguna similitud ni siquiera con los valores vecinos más próximos. Es importante tener presente que la información contenida en la correlación cruzada entre dos variables difiere de la información contenida en la covarianza cruzada únicamente en un factor de normalización.

Efecto del rango:

El rango es habitualmente el parámetro al que se le asigna mayor importancia. La siguiente figura muestra dos variogramas que difieren solamente en sus rangos, $\gamma_2(h)$ posee un rango dos veces mayor que el valor de $\gamma_1(h)$:



$$\gamma_1(h) = 10 \left(1 - e^{-3\left(\frac{h}{10}\right)^2} \right) \quad \gamma_2(h) = 10 \left(1 - e^{-3\left(\frac{h}{20}\right)^2} \right) = \gamma_1(h/2)$$

Este cambio tiene un efecto relativamente menor en la variación de los pesos calculados. A pesar de estos pequeños cambios en los pesos, se producen cambios apreciables en los valores estimados. La varianza σ_{KO}^2 calculada con el variograma de mayor rango es menor, ya que en términos de distancias estadísticas los puntos observados se encuentran dos veces más cerca unos de otros y del punto de estimación. Si el rango es muy pequeño, el resultado sería muy parecido al obtenido con un modelo de efecto pepita puro.



Efecto de la anisotropía:

En todos los ejemplos de variogramas que vimos hasta ahora, en ninguno de ellos tuvimos en cuenta la presencia de anisotropía. Únicamente tuvimos en cuenta la distancia $h_{ij} = |x_j - x_i|$ entre los puntos para cuantificar la continuidad espacial, es decir, consideramos que la continuidad espacial era la misma en todas las direcciones. Sin embargo, en muchos datos es posible observar una mayor continuidad espacial en determinada dirección. En ese caso además de considerar la variación de la continuidad espacial con la distancia también debemos hacerlo con el acimut, lo cual implica considerar que h es una magnitud vectorial o bien que el variograma es función de dos variables: la distancia y el acimut.

En el caso planimétrico bidimensional, la forma habitual de tener en cuenta la anisotropía es considerar dos variogramas para direcciones perpendiculares asociados a las direcciones de mayor y menor continuidad espacial, y combinarlos en una única función que depende de la distancia y del acimut.

Con un modelo anisotrópico se le asignará mayor peso a los puntos observados que definan un acimut con el punto estimado igual al acimut de mayor continuidad espacial del variograma, y menor peso a los puntos observados ubicados en una dirección perpendicular, es decir en la dirección de menor continuidad espacial. Lo habitual es variar el rango del variograma respecto a direcciones perpendiculares, el cociente entre el rango mayor y el menor es denominado razón de anisotropía.



Efecto de la anisotropía:

Observe que si evalúa un variograma $\gamma_a(h)$ con rango a , a una distancia h , se obtiene el mismo resultado que si evalúa un variograma $\gamma_1(h/a)$ con rango 1, a una distancia h/a . Es decir que si ambos variogramas poseen iguales parámetros a excepción del rango, podemos escribir:

$$\gamma_1(h/a) = \gamma_a(h)$$

Entonces, para el caso de dos direcciones planimétricas, si a_x es el rango en la dirección \hat{e}_x , y a_y es el rango en la dirección \hat{e}_y , el variograma $\gamma(\mathbf{h})$ que tiene en cuenta la anisotropía, se puede obtener de la siguiente manera:

$$\gamma(\mathbf{h}) = \gamma(h_x, h_y) = \gamma_1\left(\sqrt{(h_x/a_x)^2 + (h_y/a_y)^2}\right)$$

Claro que tenemos que orientar nuestros ejes en las direcciones de máxima y mínima continuidad o aplicar una rotación de coordenadas.

Información cualitativa disponible, como por ejemplo: la interpretación geológica de un depósito de minerales, la presencia de vientos que prevalecen de determinada dirección en estudios de contaminación atmosférica, direcciones preferenciales del flujo en un reservorio, orientaciones principales de los esfuerzos tectónicos, etc., puede ser incorporada en estos métodos de estimación por medio de un modelo de variograma que considere la presencia de anisotropía.



Validación cruzada:

En una prueba de validación cruzada se estima el valor en un punto donde existe una observación, sin utilizar en el proceso de estimación el valor observado en ese punto, sólo utilizando los valores observados restantes. Podemos verlo como un experimento en el cual simulamos nunca haber observado ese valor. Una vez estimado el valor podemos compararlo con el valor observado no incluido en el cálculo.

Este procedimiento se puede repetir para todos los puntos observados, y la comparación de los valores estimados y los valores observados puede darnos una idea de cuán bueno es nuestro método de estimación.

Aunque el concepto de validación cruzada es una manera ingeniosa para comparar valores estimados con valores verdaderos, en la práctica hay que considerar con cautela las conclusiones alcanzadas porque pueden ser engañosas.



Restricciones al modelo de variograma:

Si deseamos que el sistema de ecuaciones de kriging ordinario tenga una y sólo una solución estable, debemos asegurarnos que la siguiente matriz \mathbf{K} , de orden $N + 1$, sea definida positiva:

$$\mathbf{K} = \begin{bmatrix} C_{00} & C_{01} & \cdots & C_{0N} \\ C_{10} & C_{11} & \cdots & C_{1N} \\ \vdots & \vdots & \ddots & \vdots \\ C_{N0} & C_{N1} & \cdots & C_{NN} \end{bmatrix}$$

Es decir que para todo vector columna \mathbf{x} no nulo de dimensión $N + 1$, se cumpla la siguiente condición:

$$\mathbf{x}^T \mathbf{K} \mathbf{x} = \sum_{i=0}^N \sum_{j=0}^N x_i x_j K_{ij} > 0$$

Una forma de satisfacer esta condición es utilizando modelos de variogramas que es sabido que son definidos positivos. Aunque esta condición puede parecerse muy restrictiva, podemos combinar aquellas funciones que sabemos que son definidas positivas para obtener nuevas funciones que sean definidas positivas. Las expresiones de los modelos de variogramas vistos: esférico, exponencial y gaussiano, son funciones definidas positivas. Los correlogramas experimentales, calculados a partir de los valores observados, en los cuales habitualmente hay que interpolar para rellenar zonas sin información, raramente satisfacen esta condición.



Cokriging:

Hasta ahora hemos estimado el valor de nuestra variable en determinada posición en función de los valores observados $v(x_i)$ de la misma variable en posiciones vecinas. Sin embargo, es posible que no solo dispongamos de valores observados de nuestra variable primaria $v(x_i)$ sino que además dispongamos de valores observados de otras variables secundarias que estén *cross-correlacionadas* (*cross-correlated*) con la variable primaria, y que por lo tanto contengan información de gran utilidad para la estimación de la variable primaria. Parece razonable que utilicemos la información contenida en la covarianza cruzada entre la variable primaria y las variables secundarias para reducir aún más la varianza del error de estimación. Presentaremos los formalismos matemáticos del método de cokriging que hace uso de la información contenida en la covarianza cruzada para el caso de una única variable secundaria $u(x_j)$. El método puede ser generalizado sencillamente para una cantidad arbitraria de variables secundarias que posean un nivel importante de correlación cruzada con la variable primaria.

La utilización de una variable secundaria usualmente mejora la estimación de la variable primaria cuando disponemos de un muestreo más exhaustivo de la variable secundaria, si ambas variables fueron observadas en los mismos puntos entonces cokriging no producirá ninguna mejora en el resultado de la estimación.



Cokriging para una única variable secundaria:

El valor estimado por el método de cokriging para la variable primaria en el punto x_0 será una combinación lineal de los N valores observados de la variable primaria $v(x_i)$ y de los M valores observados de la variable secundaria $u(x_j)$:

$$\hat{v}(x_0) = \sum_{i=1}^N w_i v(x_i) + \sum_{j=1}^M p_j u(x_j)$$

Donde w_i y p_j son los pesos de cokriging aplicados a las variables observadas $v(x_i)$ y $u(x_j)$ respectivamente para obtener el valor estimado.

Análogamente a lo realizado para derivar el método de kriging ordinario, haciendo la hipótesis de que las variables involucradas son variables aleatorias estacionarias de segundo orden, podemos derivar el método de cokriging:

$$R(x_0) = \hat{V}(x_0) - V(x_0)$$
$$R(x_0) = \sum_{i=1}^N w_i V(x_i) + \sum_{j=1}^M p_j U(x_j) - V(x_0)$$



La varianza del error de Cokriging:

Obtengamos una expresión para la varianza del error del modelo probabilístico de Cokriging:

$$\begin{aligned}\sigma_R^2 &= \text{Var}\{R(x_0)\} = \text{Var}\left\{\left(\hat{V}(x_0) - V(x_0)\right)\right\} \\ &= E\left\{\left[\left(\hat{V}(x_0) - V(x_0)\right) - E\{\hat{V}(x_0) - V(x_0)\}\right]^2\right\} \\ \sigma_R^2 &= \sum_{i=1}^N \sum_{j=1}^N w_i w_j \text{Cov}\{V_i, V_j\} + \sum_{i=1}^M \sum_{j=1}^M p_i p_j \text{Cov}\{U_i, U_j\} + 2 \sum_{i=1}^N \sum_{j=1}^M w_i p_j \text{Cov}\{V_i, U_j\} - \\ &\quad - 2 \sum_{i=1}^N w_i \text{Cov}\{V_i, V_0\} - 2 \sum_{j=1}^M p_j \text{Cov}\{U_j, V_0\} + \text{Cov}\{V_0, V_0\}\end{aligned}$$

Donde $\text{Cov}\{V_i, U_j\}$ es la covarianza cruzada entre las variables aleatorias V_i y U_j .

Los pesos deben satisfacer dos condiciones, la primera es que deben ser tales que el valor estimado sea no sesgado, es decir $m_R = 0$, y la segunda es que deben ser tales que la varianza del error de cokriging $\sigma_R^2 = \sigma_{CK}^2$ sea mínima.



Estimador Cokriging no sesgado:

Primero veamos como satisfacer la condición de estimador no sesgado:

$$R(x_0) = \hat{V}(x_0) - V(x_0)$$

$$m_R = E\{R(x_0)\} = E\left\{\sum_{i=1}^N w_i V(x_i) + \sum_{j=1}^M p_j U(x_j) - V(x_0)\right\}$$

$$m_R = E\{R(x_0)\} = \sum_{i=1}^N w_i E\{V(x_i)\} + \sum_{j=1}^M p_j E\{U(x_j)\} - E\{V(x_0)\}$$

Como asumimos que las funciones aleatorias $V(x)$ y $U(x)$ son estacionarias de segundo orden, sus esperanzas matemáticas $m_V = E\{V(x)\}$ y $m_U = E\{U(x)\}$ son las mismas en todos los puntos de la región, entonces podemos escribir:

$$m_R = E\{R(x_0)\} = E\{V\} \sum_{i=1}^N w_i + E\{U\} \sum_{j=1}^M p_j - E\{V\}$$

$$m_R = m_V \left(\sum_{i=1}^N w_i - 1 \right) + m_U \sum_{j=1}^M p_j$$



Estimador Cokriging no sesgado:

El valor de la esperanza matemática del error m_R es referido como el sesgo del valor estimado. Como queremos que nuestro estimador sea no sesgado, igualamos el error medio residual de nuestro modelo probabilístico a cero:

$$m_R = m_V \left(\sum_{i=1}^N w_i - 1 \right) + m_U \sum_{j=1}^M p_j = 0$$

De esta ecuación podemos deducir que la manera de satisfacer la condición de estimador no sesgado es haciendo que la suma de los pesos w_i sea igual a 1 y que la suma de los pesos p_j sea igual a 0:

$$\sum_{i=1}^N w_i = 1$$

$$\sum_{j=1}^M p_j = 0$$



Multiplicadores de Lagrange:

Busquemos los pesos que minimicen la varianza del error y que además que satisfagan a las dos condiciones anteriores necesarias para que el estimador sea no sesgado. Para alcanzar el objetivo de minimizar una función con dos constricciones haremos uso del método de los multiplicadores de Lagrange. Para implementar el método simplemente igualamos cada condición a cero las multiplicamos por un multiplicador de Lagrange diferente para cada una y se las sumamos a la expresión de la varianza del error, lo cual nos conduce a la siguiente expresión:

$$\sigma_R^2 = Var\{R(x_0)\} + 2\mu_V \left(\sum_{i=1}^N w_i - 1 \right) + 2\mu_U \left(\sum_{j=1}^M p_j \right)$$

Donde μ_V y μ_U son los multiplicadores de Lagrange. Observe que estos dos términos agregados son nulos, por lo tanto no modifican a la varianza del error pero si le imponen las condiciones deseadas al cálculo de los pesos.



Minimización de la varianza del error:

Para minimizar σ_R^2 debemos calcular sus derivadas parciales respecto de los $N + M$ pesos y respecto de los dos multiplicadores de Lagrange:

$$\frac{\partial \sigma_R^2}{\partial w_j} = 2 \sum_{i=1}^N w_i \text{Cov}\{V_i, V_j\} + 2 \sum_{i=1}^M p_i \text{Cov}\{U_i, V_j\} - 2 \text{Cov}\{V_0, V_j\} + 2\mu_V \quad \text{para } j = 1, N$$

$$\frac{\partial \sigma_R^2}{\partial p_j} = 2 \sum_{i=1}^N w_i \text{Cov}\{V_i, U_j\} + 2 \sum_{i=1}^M p_i \text{Cov}\{U_i, U_j\} - 2 \text{Cov}\{V_0, U_j\} + 2\mu_U \quad \text{para } j = 1, M$$

$$\frac{\partial \sigma_R^2}{\partial \mu_V} = 2 \left(\sum_{i=1}^N w_i - 1 \right)$$

$$\frac{\partial \sigma_R^2}{\partial \mu_U} = 2 \sum_{i=1}^M p_i$$



Sistema de Cokriging:

Finalmente el sistema de cokriging es obtenido igualando cada una de estas $N + M + 2$ ecuaciones a 0. Reordenando los términos individuales obtenemos:

$$\left\{ \begin{array}{l} \sum_{i=1}^N w_i \text{Cov}\{V_i, V_j\} + \sum_{i=1}^M p_i \text{Cov}\{U_i, V_j\} + \mu_V = \text{Cov}\{V_0, V_j\} \quad \text{para } j = 1, N \\ \sum_{i=1}^N w_i \text{Cov}\{V_i, U_j\} + \sum_{i=1}^M p_i \text{Cov}\{U_i, U_j\} + \mu_U = \text{Cov}\{V_0, U_j\} \quad \text{para } j = 1, M \\ \sum_{i=1}^N w_i = 1 \\ \sum_{i=1}^M p_i = 0 \end{array} \right.$$

Sistema de Cokriging:

Expresado matricialmente nos queda:

$$\begin{bmatrix} \mathbf{C}_{VV} & \mathbf{C}_{UV} & \mathbf{1} & \mathbf{0} \\ \mathbf{C}_{VU} & \mathbf{C}_{UU} & \mathbf{0} & \mathbf{1} \\ \mathbf{I}_N & \mathbf{0}_M & \mathbf{0} & \mathbf{0} \\ \mathbf{0}_N & \mathbf{I}_M & \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{W} \\ \mathbf{P} \\ \mu_V \\ \mu_U \end{bmatrix} = \begin{bmatrix} \mathbf{C}_{V_0V} \\ \mathbf{C}_{V_0U} \\ 1 \\ 0 \end{bmatrix}$$

donde:

$$\mathbf{C}_{VV} = \begin{bmatrix} C_{VV_{11}} & \cdots & C_{VV_{1N}} \\ \vdots & \ddots & \vdots \\ C_{VV_{N1}} & \cdots & C_{VV_{NN}} \end{bmatrix} \quad \mathbf{C}_{UU} = \begin{bmatrix} C_{UU_{11}} & \cdots & C_{UU_{1M}} \\ \vdots & \ddots & \vdots \\ C_{UU_{M1}} & \cdots & C_{UU_{MM}} \end{bmatrix}$$

$$\mathbf{C}_{UV} = \begin{bmatrix} C_{UV_{11}} & \cdots & C_{UV_{1N}} \\ \vdots & \ddots & \vdots \\ C_{UV_{M1}} & \cdots & C_{UV_{MN}} \end{bmatrix} \quad \mathbf{C}_{VU} = \begin{bmatrix} C_{VU_{11}} & \cdots & C_{VU_{1M}} \\ \vdots & \ddots & \vdots \\ C_{VU_{N1}} & \cdots & C_{VU_{NM}} \end{bmatrix}$$

$$\mathbf{W} = [w_1 \quad \cdots \quad w_N]^T \quad \mathbf{P} = [p_1 \quad \cdots \quad p_M]^T$$

$$\mathbf{C}_{V_0V} = [C_{VV_{01}} \quad \cdots \quad C_{VV_{0N}}]^T \quad \mathbf{C}_{V_0U} = [C_{VU_{01}} \quad \cdots \quad C_{VU_{0M}}]^T$$

\mathbf{I}_N es una matriz identidad de orden N , \mathbf{I}_M es una matriz identidad de orden M , $\mathbf{0}_N$ es una matriz cuadrada nula de orden N y $\mathbf{0}_M$ es una matriz cuadrada nula de orden M .



Varianza de Cokriging:

La expresión para obtener la varianza minimizada del error σ_R^2 , puede ser simplificada haciendo las sustituciones correspondientes, utilizando las expresiones de los multiplicadores de Lagrange. La versión simplificada nos queda:

$$\sigma_R^2 = Cov\{V_0, V_0\} - \mu_V - \sum_{i=1}^N w_i Cov\{V_i, V_0\} - \sum_{j=1}^M p_j Cov\{U_j, V_0\}$$

$$\sigma_R^2 = \sigma_{CK}^2 = \sigma_V^2 - \mu_V - \sum_{i=1}^N w_i Cov\{V_i, V_0\} - \sum_{j=1}^M p_j Cov\{U_j, V_0\}$$



Varianza de Cokriging:

El sistema de cokriging puede ser escrito en función del variograma asumiendo que la covarianza cruzada es simétrica, es decir:

$$Cov\{V_i, U_j\} = Cov\{U_j, V_i\}$$

Aunque la covarianza cruzada puede no ser simétrica, la práctica más común es modelarla como una función simétrica. Entonces la continuidad espacial definida en función del variograma puede ser expresada en función de la covarianza utilizando la siguiente expresión:

$$C_{VU}(h) = \gamma_{VU}(\infty) - \gamma_{VU}(h)$$

Para garantizar que la solución del sistema de cokriging exista y sea única, tanto los variogramas como los variogramas cruzados deben ser definidos positivos. Para los variogramas cruzados se utilizan habitualmente las mismas funciones parametrizadas vistas para los variogramas, donde el umbral estará dado por la covarianza cruzada entre las variables primaria y secundaria en un mismo punto, es decir para $h = 0$.



Bibliografía:

- Wackernagel, H. (2003). Multivariate Geostatistics. Springer.
- Davis, J. (1986). Statistics and Data Analysis in Geology. John Wiley & Sons.
- Isaaks, E. & Srivastava, M. (1989). An Introduction to Applied Geostatistics. Oxford.
- Brand, S. (1970). Statistical and Computational Methods in Data Analysis. North Holland.
- [Apunte de Geoestadística de la Cátedra.](#)