

INTRODUCCIÓN

En los temas anteriores se han introducido las definiciones de la función de covarianza y el variograma, así como algunos resultados y modelos relacionados. En este tema se abordan los problemas de estimación e interpolación/predicción. El estudio se centrará en el caso espacial. Concretamente, se introduce la definición de variograma empírico, a partir del cual se tratará de ajustar algún modelo teórico apropiado, y la técnica de interpolación/predicción denominada krigeado (o *kriging*). En la sección siguiente se revisan y amplían conocimientos sobre modelos de variogramas espaciales.

ALGUNOS MODELOS DE VARIOGRAMA ESPACIAL VÁLIDOS

Se distinguen dos familias de funciones simples que representan de forma adecuada las características de los variogramas. Una representa las variaciones no acotadas y la otra las acotadas. Las presentaremos en su forma isotrópica, es decir, dependiendo sólo de la distancia y no de la dirección.

Variación aleatoria no acotada

La idea de varianza infinita puede parecer poco realista, en un sentido estricto, desde el punto de vista físico. Sin embargo, algunas situaciones en las que se muestrean regiones pequeñas (en el espacio y/o tiempo) sugieren que si se amplía la región, la variación tiende a ser cada vez mayor.

El modelo más simple para la variación no acotada son las funciones potenciales:

$$\gamma(h) = wh^\alpha \quad \text{para } 0 < \alpha < 2, \quad (1)$$

donde w describe la intensidad de variación y α describe la curvatura.

Una forma de mirar estas funciones no acotadas es considerar el movimiento browniano unidimensional. Supongamos una partícula cuya velocidad en la posición $x + h$ depende de la velocidad en la posición previa x y puede ser representada como

$$Z(x + h) = \beta Z(x) + \epsilon,$$

donde ϵ es una variable aleatoria gaussiana independiente y β es un parámetro. Si $\beta = 1$, su variograma es entonces

$$2\gamma(h) = E[\{Z(x+h) - Z(x)\}^2] = |h|^k.$$

Variación aleatoria acotada

Estos modelos son más comunes en la práctica que los no acotados.

El modelo más simple es el lineal acotado, que se forma mediante la unión de dos líneas: la primera se incrementa con pendiente constante y la otra es constante a partir de un punto dado

$$\gamma(h) = \begin{cases} C \left(\frac{h}{a} \right) & , \quad h \leq a \\ C & , \quad h > a \end{cases}$$

donde a es el rango y C es la varianza. Este modelo no puede usarse para dimensiones superiores a 1.

Otros modelos muy utilizados son:

Esférico

$$\gamma(h) = \begin{cases} C \left(\frac{3h}{2a} - \frac{1}{2} \left(\frac{h}{a} \right)^3 \right) & , \quad 0 \leq h < a \\ C & , \quad a \leq h \end{cases}$$

Exponencial

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

Gaussiano

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

Potencial

$$\gamma(h) = \alpha h^\beta, \quad \text{con } 0 < \beta < 2$$

Efecto pepita puro

$$\gamma(h) = \begin{cases} C, & 0 < h \\ 0, & h = 0 \end{cases}$$

Los modelos anteriores son todos isotrópicos, es decir, la estructura de dependencia espacial sea función sólo de h , la distancia. Puede que, además, dependa de la dirección, en cuyo caso tendremos un modelo anisotrópico. En algunas circunstancias, la anisotropía se puede transformar a isotropía mediante un cambio de coordenadas, girando un ángulo determinado y reescalando los ejes (esencialmente, se trata de transformar una elipse en un círculo). En este caso se habla de anisotropía geométrica.

ESTIMACIÓN DEL VARIOGRAMA

El variograma se suele estimar mediante el variograma empírico. Éste se basa en calcular la media de las diferencias al cuadrado entre las observaciones asociadas a coordenadas que distan entre sí (dos a dos y, en el caso más general, teniendo en cuenta la direccionalidad) un vector \mathbf{h} :

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2n(\mathbf{h})} \sum_{i=1}^{n(\mathbf{h})} [Z(\mathbf{s}_i + \mathbf{h}) - Z(\mathbf{s}_i)]^2$$

con $n(\mathbf{h})$ = número de pares de variables con distancia vectorial \mathbf{h} .

En la práctica, dado que usualmente las observaciones están irregularmente espaciadas, se consideran ‘entornos’ de distancias definidos por rangos de tolerancia en longitud y dirección.

Una vez estimado el variograma empírico, se suelen dibujar estos puntos para ajustar el modelo teórico que se suponga que tiene el proceso mediante un ajuste de mínimos cuadrados.

KRIGEADO

Una de las técnicas de predicción más aplicadas en el contexto espacial, encaminadas a la predicción, es el kriging. Esta técnica consiste en construir el predictor lineal con menor error cuadrático medio de predicción, que generalmente depende de las propiedades de segundo orden del proceso $X(\cdot)$. El nombre de kriging se debe a que el precursor de esta técnica fue el ingeniero de minas D.G. Krige, quien en 1950 desarrolló métodos empíricos para la predicción espacial. Sin embargo, la formulación de este tipo de predictores ha sido realizada por otros autores en años posteriores. Según las hipótesis que se consideren sobre el proceso $X(\cdot)$, la formulación y la expresión del predictor será distinta. Por ello, distinguimos diversos tipos de kriging. Sin embargo, la situación común a todos ellos que se intenta resolver es la siguiente:

Sea $X(s)$ un campo espacial, con $s \in D$, siendo D el dominio de definición del campo. Supongamos que disponemos de datos observados en las localizaciones s_1, \dots, s_n ,

$$X(s_1), \dots, X(s_n),$$

y que estamos interesados en predecir $X(s_0)$, $s_0 \in D$. Denotemos por $\hat{X}(s_0)$ al predictor obtenido a partir de los datos observados. El error cometido al predecir $X(s_0)$ mediante $\hat{X}(s_0)$ es

$$X(s_0) - \hat{X}(s_0),$$

y el error cuadrático medio de predicción está dado por

$$E[X(s_0) - \hat{X}(s_0)]^2, \tag{2}$$

donde la esperanza es considerada con respecto a la distribución conjunta de $X(\mathbf{s}_0)$ y $\mathbf{X}_S = (X(\mathbf{s}_1), \dots, X(\mathbf{s}_n))'$. El *mejor* predictor se define como aquél que minimiza el error cuadrático medio de predicción dado por la expresión (2) y se obtiene como la esperanza de $X(\mathbf{s}_0)$ condicionada a \mathbf{X}_S . En la práctica es difícil obtener dicha esperanza sólo a partir de los n datos observados, ya que en cierta medida ello presupone el conocimiento de una distribución $(n + 1)$ -dimensional. Por ello, se suele descomponer el proceso como suma de una tendencia determinística $\mu(\mathbf{s})$ y un proceso aleatorio de media cero $\epsilon(\mathbf{s})$, según la ecuación

$$X(\mathbf{s}) = \mu(\mathbf{s}) + \epsilon(\mathbf{s}), \quad \mathbf{s} \in D.$$

A continuación presentamos los tipos de kriging más usuales, que dependen de las hipótesis que realicemos sobre el proceso y de las restricciones impuestas sobre el predictor.

Kriging simple

Se supone que tanto la función de tendencia $\mu(\mathbf{s})$ como la función de covarianza de $\epsilon(\mathbf{s})$ son conocidas. Se considera también que el predictor adopta la expresión

$$\hat{X}(\mathbf{s}_0) = \sum_{i=1}^n \lambda_i X(\mathbf{s}_i) + k,$$

donde k y $\lambda_1, \dots, \lambda_n$ son parámetros desconocidos que se calcularán imponiendo que el error cuadrático medio sea mínimo. Los parámetros λ_i se denominan *coeficientes de ponderación* o

coeficientes de krigeado y las ecuaciones que resultan de imponer dicha condición se denominan *ecuaciones de krigeado*.

El error cuadrático medio en este caso viene dado por

$$E \left[X(\mathbf{s}_0) - \sum_{i=1}^n \lambda_i X(\mathbf{s}_i) - k \right]^2 = \text{Var} \left[X(\mathbf{s}_0) - \sum_{i=1}^n \lambda_i X(\mathbf{s}_i) \right] + \left[\mu(\mathbf{s}_0) - \sum_{i=1}^n \lambda_i \mu(\mathbf{s}_i) - k \right]^2.$$

Bajo estas hipótesis, el predictor óptimo viene determinado por

$$k = \mu(\mathbf{s}_0) - \sum_{i=1}^n \lambda_i \mu(\mathbf{s}_i),$$

$$(\lambda_1, \dots, \lambda_n) = \Sigma_{X(\mathbf{s}_0), \mathbf{X}_S} \Sigma_{\mathbf{X}_S}^{-1},$$

siendo $\Sigma_{\mathbf{X}_S}$ la matriz de covarianzas de \mathbf{X}_S y $\Sigma_{X(\mathbf{s}_0), \mathbf{X}_S}$ el vector de covarianzas entre $X(\mathbf{s}_0)$ y \mathbf{X}_S . El predictor adopta, pues, la expresión

$$\hat{X}(\mathbf{s}_0) = \mu(\mathbf{s}_0) + \Sigma_{X(\mathbf{s}_0), \mathbf{X}_S} \Sigma_{\mathbf{X}_S}^{-1} (\mathbf{X}_S - \mu_S),$$

siendo $\mu_S = (\mu(\mathbf{s}_1), \dots, \mu(\mathbf{s}_n))'$. El predictor obtenido de esta forma es insesgado y el error cuadrático medio viene dado por

$$\sigma_k^2(\mathbf{s}_0) = \Sigma_{X(\mathbf{s}_0)} - \sum_{i=1}^n \lambda_i \Sigma_{X(\mathbf{s}_0), X(\mathbf{s}_i)}.$$

En el caso en que $\epsilon(\mathbf{s})$ sea un proceso gaussiano, el predictor lineal $\hat{X}(\mathbf{s}_0)$ coincide con $E[X(\mathbf{s}_0)|\mathbf{X}_S]$ y es, por tanto, el mejor entre todos los predictores. Es claro que este método constituye un método de interpolación, dado que $\hat{X}(\mathbf{s}_i) = X(\mathbf{s}_i)$.

En la práctica, la media determinística $\mu(\mathbf{s})$ y la estructura de covarianza no son conocidas; en ese caso, el predictor obtenido en el krigeado simple no es útil.

Krigeado ordinario

Se supone ahora que $\mu(\mathbf{s}) = \mu$, donde μ es desconocido, y se exige que el predictor sea lineal, insesgado y con varianza mínima, dado por la expresión

$$\hat{X}(\mathbf{s}_0) = \sum_{i=1}^n \lambda_i X(\mathbf{s}_i), \quad \sum_{i=1}^n \lambda_i = 1.$$

Esta última condición sobre los coeficientes λ_i garantiza la insesgadez del predictor, puesto que

$$E[\hat{X}(\mathbf{s}_0)] = \sum_{i=1}^n \lambda_i \mu = \mu.$$

Suponemos además que $\epsilon(\mathbf{s})$ es un proceso intrínsecamente estacionario, es decir,

$$\text{Var} [\epsilon(\mathbf{s} + \mathbf{h}) - \epsilon(\mathbf{s})] = 2\gamma(\mathbf{h}), \quad \forall \mathbf{s}, \mathbf{s} + \mathbf{h} \in D,$$

donde la función $\gamma(\cdot)$ no depende de \mathbf{s} . A la función $2\gamma(\cdot)$ se le denomina *variograma*. Bajo estas condiciones, el predictor lineal óptimo minimizará

$$E \left[X(\mathbf{s}_0) - \sum_{i=1}^n \lambda_i X(\mathbf{s}_i) \right]^2 - 2\eta \left(\sum_{i=1}^n \lambda_i - 1 \right),$$

donde η es el multiplicador de Lagrange. Si desarrollamos el cuadrado que aparece en la esperanza y aplicamos la condición de insesgadez, $\sum_{i=1}^n \lambda_i = 1$, obtenemos

$$\left[X(\mathbf{s}_0) - \sum_{i=1}^n \lambda_i X(\mathbf{s}_i) \right]^2 = -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j (X(\mathbf{s}_i) - X(\mathbf{s}_j))^2 + \sum_{i=1}^n \lambda_i (X(\mathbf{s}_0) - X(\mathbf{s}_i))^2, \quad (3)$$

con lo que la función a minimizar vendrá dada por

$$-\sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \gamma(\mathbf{s}_i - \mathbf{s}_j) + 2 \sum_{i=1}^n \lambda_i \gamma(\mathbf{s}_0 - \mathbf{s}_i) - 2\eta \left(\sum_{i=1}^n \lambda_i - 1 \right).$$

Si diferenciamos esta ecuación con respecto a los coeficientes λ_i y al parámetro η e igualamos a cero, obtenemos las ecuaciones

$$-\sum_{j=1}^n \lambda_j \gamma(\mathbf{s}_i - \mathbf{s}_j) + \gamma(\mathbf{s}_0 - \mathbf{s}_i) - \eta = 0, \quad i = 1, \dots, n,$$

$$\sum_{i=1}^n \lambda_i = 1.$$

Los coeficientes $\lambda = (\lambda_1, \dots, \lambda_n)'$ y η están dados por

$$\lambda' = \left(\Gamma_{X(s_0), \mathbf{X}_S} + \mathbf{1} \frac{(1 - \mathbf{1}' \Gamma_{\mathbf{X}_S}^{-1} \Gamma_{X(s_0), \mathbf{X}_S})}{\mathbf{1}' \Gamma_{\mathbf{X}_S}^{-1} \mathbf{1}} \right)' \Gamma_{\mathbf{X}_S}^{-1},$$

$$\eta = - \frac{1 - \mathbf{1}' \Gamma_{\mathbf{X}_S}^{-1} \Gamma_{X(s_0), \mathbf{X}_S}}{\mathbf{1}' \Gamma_{\mathbf{X}_S}^{-1} \mathbf{1}},$$

siendo $\Gamma_{X(s_0), \mathbf{X}_S} = (\gamma(s_0 - s_1), \dots, \gamma(s_0 - s_n))$ y $\Gamma_{\mathbf{X}_S}$ la matriz de dimensión $n \times n$ con elemento (i, j) -ésimo dado por $\gamma(s_i - s_j)$, y $\mathbf{1} = (1, \dots, 1)'$.

En términos de la función de covarianzas, la expresión anterior para los coeficientes λ_i y para el parámetro η puede deducirse siguiendo los mismos pasos. En este sentido, únicamente habrá que hacer notar que

$$\left[X(s_0) - \sum_{i=1}^n \lambda_i X(s_i) \right]^2 = (X(s_0) - \mu)^2 + \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j (X(s_i) - \mu)(X(s_j) - \mu)$$

$$- 2 \sum_{i=1}^n \lambda_i (X(s_0) - \mu)(X(s_i) - \mu),$$

con lo que los coeficientes de kriging y el multiplicador de Lagrange vendrán dados por

$$\lambda' = \left(\Sigma_{X(s_0), \mathbf{X}_S} + \mathbf{1} \frac{(1 - \mathbf{1}' \Sigma_{\mathbf{X}_S}^{-1} \Sigma_{X(s_0), \mathbf{X}_S})}{\mathbf{1}' \Sigma_{\mathbf{X}_S}^{-1} \mathbf{1}} \right)' \Sigma_{\mathbf{X}_S}^{-1},$$

$$\eta = - \frac{1 - \mathbf{1}' \Sigma_{\mathbf{X}_S}^{-1} \Sigma_{X(s_0), \mathbf{X}_S}}{\mathbf{1}' \Sigma_{\mathbf{X}_S}^{-1} \mathbf{1}},$$

donde $\Sigma_{X(s_0), \mathbf{X}_S} = (C(\mathbf{s}_0 - \mathbf{s}_1), \dots, C(\mathbf{s}_0 - \mathbf{s}_n))$ y $\Sigma_{\mathbf{X}_S}$ es la matriz $n \times n$ cuyo elemento (i, j) -ésimo es $C(\mathbf{s}_i - \mathbf{s}_j)$.

El error cuadrático medio para la predicción, algunas veces denominado *varianza del kriging*, viene dado por

$$\begin{aligned} \sigma_k^2(\mathbf{s}_0) &= \sum_{i=1}^n \lambda_i \gamma(\mathbf{s}_0 - \mathbf{s}_i) + \eta \\ &= \Gamma'_{X(s_0), \mathbf{X}_S} \Gamma_{\mathbf{X}_S}^{-1} \Gamma_{X(s_0), \mathbf{X}_S} - \frac{(\mathbf{1}' \Gamma_{\mathbf{X}_S}^{-1} \Gamma_{X(s_0), \mathbf{X}_S} - 1)^2}{\mathbf{1}' \Gamma_{\mathbf{X}_S}^{-1} \mathbf{1}} \\ &= \sum_{i=1}^n \lambda_i \gamma(\mathbf{s}_0 - \mathbf{s}_i) - \frac{(1 - \mathbf{1}' \Gamma_{\mathbf{X}_S}^{-1} \Gamma'_{X(s_0), \mathbf{X}_S})^2}{\mathbf{1}' \Gamma_{\mathbf{X}_S}^{-1} \mathbf{1}} \\ &= 2 \sum_{i=1}^n \lambda_i \gamma(\mathbf{s}_0 - \mathbf{s}_i) - \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \gamma(\mathbf{s}_i - \mathbf{s}_j). \end{aligned}$$

Krigeado universal

A veces es más realista suponer que la media es una combinación lineal de funciones conocidas. Es decir, se supone que la función de tendencia es desconocida y que puede expresarse mediante

$$\mu(\mathbf{s}) = \sum_{j=0}^m f_j(\mathbf{s})\beta_j,$$

donde las funciones $f_0(\mathbf{s}), \dots, f_m(\mathbf{s})$ son conocidas y los coeficientes β_0, \dots, β_m son desconocidos. En la práctica, $\mu(\mathbf{s})$ suele representarse mediante polinomios, por ejemplo, mononios del tipo $s_1^r s_2^s$, o funciones que dependan de alguna característica ligada a las localizaciones, por ejemplo altura. Como en el caso anterior, se supone que $\epsilon(\mathbf{s})$ es un proceso intrínsecamente estacionario. Obviamente, para las n localizaciones donde se observa el proceso podemos escribir de forma matricial

$$\mathbf{X}_S = \mathbf{F}_S \beta + \epsilon,$$

donde \mathbf{F}_S es una matriz $n \times (m + 1)$ cuyo elemento (i, j) -ésimo viene dado por $f_{j-1}(\mathbf{s}_i)$.

Se impone que el predictor sea lineal, insesgado y con mínima varianza. El predictor óptimo adopta la expresión

$$\hat{X}(\mathbf{s}_0) = \sum_{i=1}^n \lambda_i X(\mathbf{s}_i),$$

con la restricción de insesgadez

$$(\lambda_1, \dots, \lambda_n) \mathbf{F}_S = (f_0(\mathbf{s}_0), \dots, f_m(\mathbf{s}_0)).$$

En este caso, la función a minimizar es

$$E \left[X(\mathbf{s}_0) - \sum_{i=1}^n \lambda_i X(\mathbf{s}_i) \right]^2 - 2 \sum_{j=0}^m \eta_j \left[\sum_{i=1}^n \lambda_i f_j(\mathbf{s}_i) - f_j(\mathbf{s}_0) \right],$$

donde η_j son los multiplicadores de Lagrange. Usualmente se considera $f_0(\mathbf{s}) \equiv 1$, con lo que una de las restricciones sobre los coeficientes λ_i es que $\sum_{i=1}^n \lambda_i = 1$. Así, haciendo el mismo desarrollo que en (3), la función a minimizar se reduce a

$$- \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \gamma(\mathbf{s}_i - \mathbf{s}_j) + 2 \sum_{i=1}^n \lambda_i \gamma(\mathbf{s}_0 - \mathbf{s}_i) - 2 \sum_{j=0}^m \eta_j \left[\sum_{i=1}^n \lambda_i f_j(\mathbf{s}_i) - f_j(\mathbf{s}_0) \right].$$

Usando la misma notación que para el krigado ordinario, la expresión de los estimadores de los parámetros se obtiene resolviendo un sistema de ecuaciones similar, obtenido al derivar la expresión anterior con respecto a los coeficientes λ_i y los parámetros η_j e igualar estas ecuaciones a cero. Los estimadores así obtenidos responden a la expresión

$$\lambda' = \left\{ \Gamma'_{X(\mathbf{s}_0), \mathbf{X}_S} + \mathbf{F}_S \left(\mathbf{F}_S' \Gamma_{\mathbf{X}_S}^{-1} \mathbf{F}_S \right)^{-1} \left((f_0(\mathbf{s}_0), \dots, f_m(\mathbf{s}_0)) - \mathbf{F}_S' \Gamma_{\mathbf{X}_S}^{-1} \mathbf{F}_S \right) \right\}' \Gamma_{\mathbf{X}_S}^{-1},$$

$$\eta = \left((f_0(\mathbf{s}_0), \dots, f_m(\mathbf{s}_0)) - \mathbf{F}_S' \Gamma_{\mathbf{X}_S}^{-1} \Gamma_{X(\mathbf{s}_0), \mathbf{X}_S} \right)' \left(\mathbf{F}_S' \Gamma_{\mathbf{X}_S}^{-1} \mathbf{F}_S \right)^{-1}.$$

La expresión de estos coeficientes en función de las covarianzas es análoga a la anterior, reemplazando $\Gamma_{X(\mathbf{s}_0), \mathbf{X}_S}$ por $\Sigma_{X(\mathbf{s}_0), \mathbf{X}_S}$ y $\Gamma_{\mathbf{X}_S}$ por $\Sigma_{\mathbf{X}_S}$, con el mismo significado que en el apartado anterior.

El error cuadrático medio está dado por

$$\begin{aligned}
\sigma_k^2(\mathbf{s}_0) &= \sum_{i=1}^n \lambda_i \gamma(\mathbf{s}_0 - \mathbf{s}_i) + \sum_{j=0}^m \eta_j f_j(\mathbf{s}_0) \\
&= \Gamma'_{X(\mathbf{s}_0), \mathbf{X}_S} \Gamma_{\mathbf{X}_S}^{-1} \Gamma_{X(\mathbf{s}_0), \mathbf{X}_S} - \left((f_0(\mathbf{s}_0), \dots, f_m(\mathbf{s}_0)) - \mathbf{F}'_S \Gamma_{\mathbf{X}_S}^{-1} \Gamma_{X(\mathbf{s}_0), \mathbf{X}_S} \right)' \\
&\quad (\mathbf{F}'_S \Gamma_{\mathbf{X}_S}^{-1} \mathbf{F}_S)^{-1} \left((f_0(\mathbf{s}_0), \dots, f_m(\mathbf{s}_0)) - \mathbf{F}'_S \Gamma_{\mathbf{X}_S}^{-1} \Gamma_{X(\mathbf{s}_0), \mathbf{X}_S} \right) \\
&= 2 \sum_{i=1}^n \lambda_i \gamma(\mathbf{s}_0 - \mathbf{s}_i) - \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \gamma(\mathbf{s}_i - \mathbf{s}_j).
\end{aligned}$$

Observemos que el kriging ordinario es un caso particular del kriging universal en el que $m = 0$ y $f_0(\mathbf{s}) \equiv 1$.

Cokriging

En los apartados anteriores se ha considerado que el interés era la predicción de una única variable aleatoria en un punto \mathbf{s}_0 de interés. Una extensión natural es la consideración conjunta de varias variables. Ésta es la filosofía del cokriging, en el que habrá que utilizar tanto la información espacial de cada variable como la relación que existe entre las diferentes variables consideradas. Esto conllevará un número de cálculos sensiblemente mayor, con lo que el tiempo de cálculo aumentará. Se presentará también el problema de estimar modelos multivariantes para la función de covarianzas o el semivariograma entre las variables.

Denotemos por X^1, \dots, X^p a las p variables de interés que se supone han sido observadas en n localizaciones s_1, \dots, s_n .

Algunas situaciones donde el cokrigado puede resolver problemas de predicción son los siguientes:

- Todas las variables a excepción de X^1 han sido muestreadas en s_0 y estamos interesados en predecir X^1 en s_0 , con lo que el problema de predicción se reduce a encontrar los valores de los coeficientes λ_{ij} en la siguiente expresión:

$$\hat{X}^1(s_0) = \sum_{i=2}^p \lambda_{i1} X^i(s_0) + \sum_{i=1}^p \sum_{j=1}^n \lambda_{ij} X^i(s_j),$$

imponiendo que la predicción sea insesgada y que la varianza del error de predicción sea mínima.

- Estimación de los errores de medida: Para una variable aleatoria de interés $X(s)$ no observable directamente cuyos valores medidos son $Y(s)$, se supone que

$$Y(s) = X(s) + \epsilon(s),$$

donde los errores de medición $\epsilon(s)$ son independientes de $X(s)$, con media cero y no correlacionados espacialmente. Claramente las variables X e Y son dependientes y se puede plantear el problema de estimar $X(s)$ a partir de $Y(s)$.

- Estimación conjunta de las p variables unidimensionales en una nueva localización s_0 .

Denotamos la covarianza cruzada como

$$C_{ij}(\mathbf{s}, \mathbf{s} + \mathbf{h}) = \text{Cov}(X^i(\mathbf{s}), X^j(\mathbf{s} + \mathbf{h})),$$

y el semivariograma cruzado como

$$\gamma_{ij}(\mathbf{s}, \mathbf{s} + \mathbf{h}) = \frac{1}{2} E \left[(X^i(\mathbf{s} + \mathbf{h}) - X^i(\mathbf{s})) (X^j(\mathbf{s} + \mathbf{h}) - X^j(\mathbf{s})) \right].$$

Esta función verifica que

$$\begin{aligned} \gamma_{ij}(\mathbf{s}, \mathbf{s} + \mathbf{h}) &= \gamma_{ji}(\mathbf{s}, \mathbf{s} + \mathbf{h}), \\ \gamma_{ij}(\mathbf{s}, \mathbf{s} + \mathbf{h}) &= \gamma_{ij}(\mathbf{s} + \mathbf{h}, \mathbf{s}), \end{aligned}$$

propiedades que, en general, no verifica la función de covarianza.

A continuación presentamos el cokrigado ordinario. Para un estudio más detallado sobre esta técnica ver, por ejemplo, Cressie (1991), Wackernagel (1998), Olea (1999), Webster y Oliver (2001), entre otros. En este caso se suponen además las siguientes hipótesis:

$$\begin{aligned} E[X^i(\mathbf{s})] &= \mu^i, \quad i = 1, \dots, p, \\ \text{Cov}[X^i(\mathbf{s}), X^j(\mathbf{s} + \mathbf{h})] &= C_{ij}(\mathbf{s}, \mathbf{s} + \mathbf{h}) \equiv C_{ij}(\mathbf{s} - (\mathbf{s} + \mathbf{h})) \equiv C_{ij}(-\mathbf{h}) \equiv C_{ij}(\mathbf{h}). \end{aligned}$$

Es decir, las medias son constantes para cada variable (y desconocidas) y las variables aleatorias son conjuntamente estacionarias de segundo orden, dependiendo la covarianza de la distancia entre los puntos donde se aplica, y, obviamente, de las variables sobre las que se realiza.

La situación que consideramos de interés es la última, es decir, queremos predecir todas las variables en una nueva localización mediante una combinación lineal de las observaciones. Así pues, deberemos determinar los siguientes valores de los parámetros λ_{ij}^k :

$$\hat{X}^i(\mathbf{s}_0) = \sum_{k=1}^n \sum_{j=1}^p \lambda_{ij}^k X^j(\mathbf{s}_k), \quad i = 1, \dots, p.$$

Las ecuaciones anteriores se pueden representar en forma matricial de la siguiente forma:

$$\begin{pmatrix} \hat{X}^1(\mathbf{s}_0) \\ \vdots \\ \hat{X}^p(\mathbf{s}_0) \end{pmatrix} = \sum_{k=1}^n \begin{pmatrix} \lambda_{11}^k & \dots & \lambda_{1p}^k \\ \vdots & \ddots & \vdots \\ \lambda_{p1}^k & \dots & \lambda_{pp}^k \end{pmatrix} \begin{pmatrix} X^1(\mathbf{s}_k) \\ \vdots \\ X^p(\mathbf{s}_k) \end{pmatrix} = \sum_{k=1}^n \Lambda_k \begin{pmatrix} X^1(\mathbf{s}_k) \\ \vdots \\ X^p(\mathbf{s}_k) \end{pmatrix},$$

donde Λ_k se identifica como la matriz de la ecuación central.

En este caso obtenemos p errores de predicción $X^i(\mathbf{s}_0) - \hat{X}^i(\mathbf{s}_0)$. La condición de insesgadez establece que

$$E \left[X^i(\mathbf{s}_0) - \hat{X}^i(\mathbf{s}_0) \right] = 0.$$

Debido a que las medias son constantes para cualquier localización, esta restricción se puede expresar como

$$\sum_{k=1}^n \Lambda_k = \mathbf{I}_p,$$

donde \mathbf{I}_p es la matriz identidad de orden p . Estas condiciones también se pueden expresar en forma no matricial como

$$\sum_{k=1}^n \lambda_{ij}^k = \delta_{ij},$$

siendo δ_{ij} la delta de Krönecker.

Como criterio de optimalidad vamos a adoptar que la suma de las varianzas de los errores de predicción cometidos en cada variable sea mínima, es decir,

$$\sum_{j=1}^p \text{Var} \left[\hat{X}^j(\mathbf{s}_0) - X^j(\mathbf{s}_0) \right].$$

Si usamos notación matricial $\mathbf{X}(\mathbf{s}) = (X^1(\mathbf{s}), \dots, X^p(\mathbf{s}))'$ el criterio anterior se puede escribir como

$$E \left[(\hat{\mathbf{X}}(\mathbf{s}_0) - \mathbf{X}(\mathbf{s}_0))' (\hat{\mathbf{X}}(\mathbf{s}_0) - \mathbf{X}(\mathbf{s}_0)) \right].$$

Dado que la cantidad anterior es escalar, coincide con su traza y, aplicando propiedades de ésta, obtenemos que es igual a

$$\begin{aligned}
& \text{tr} \left[E \left[(\hat{\mathbf{X}}(\mathbf{s}_0) - \mathbf{X}(\mathbf{s}_0))(\hat{\mathbf{X}}(\mathbf{s}_0) - \mathbf{X}(\mathbf{s}_0))' \right] \right] \\
&= \text{tr} \left[\text{Cov} \left(\hat{\mathbf{X}}(\mathbf{s}_0) - \mathbf{X}(\mathbf{s}_0), \hat{\mathbf{X}}(\mathbf{s}_0) - \mathbf{X}(\mathbf{s}_0) \right) \right] \\
&= \text{tr} \left[E \left[\hat{\mathbf{X}}(\mathbf{s}_0) \hat{\mathbf{X}}(\mathbf{s}_0)' \right] + E \left[\mathbf{X}(\mathbf{s}_0) \mathbf{X}(\mathbf{s}_0)' \right] - E \left[\mathbf{X}(\mathbf{s}_0) \hat{\mathbf{X}}(\mathbf{s}_0)' \right] - E \left[\hat{\mathbf{X}}(\mathbf{s}_0) \mathbf{X}(\mathbf{s}_0)' \right] \right] \\
&= \text{tr} \left[\sum_{l=1}^n \sum_{k=1}^n \Lambda_l E \left[\mathbf{X}(\mathbf{s}_l) \mathbf{X}(\mathbf{s}_k)' \right] \Lambda_k' \right] - 2 \text{tr} \left[\sum_{k=1}^n \Lambda_k E \left[\mathbf{X}(\mathbf{s}_k) \mathbf{X}(\mathbf{s}_0)' \right] \right] \\
&\quad + \text{tr} \left[E \left[\mathbf{X}(\mathbf{s}_0) \mathbf{X}(\mathbf{s}_0)' \right] \right].
\end{aligned}$$

Definiendo la matriz $\mathbf{C}(\mathbf{h}) = (C_{ij}(\mathbf{h}))$, y teniendo en cuenta que $\sum_{k=1}^n \Lambda_k = \mathbf{I}_p$, tenemos que la expresión anterior puede escribirse como

$$\text{tr} \left[\sum_{l=1}^n \sum_{k=1}^n \Lambda_l \mathbf{C}(\mathbf{s}_l - \mathbf{s}_k) \Lambda_k' \right] - 2 \text{tr} \left[\sum_{k=1}^n \Lambda_k \mathbf{C}(\mathbf{s}_k - \mathbf{s}_0) \right] + \text{tr} [\mathbf{C}(\mathbf{s}_0)].$$

Si hacemos mínima esta función y además tenemos en cuenta las restricciones sobre las matrices Λ_k , el sistema a resolver es de la forma

$$\begin{pmatrix} \mathbf{C}(\mathbf{s}_1 - \mathbf{s}_1) & \dots & \mathbf{C}(\mathbf{s}_1 - \mathbf{s}_n) & \mathbf{I}_p \\ \vdots & \ddots & \vdots & \vdots \\ \mathbf{C}(\mathbf{s}_n - \mathbf{s}_1) & \dots & \mathbf{C}(\mathbf{s}_n - \mathbf{s}_n) & \mathbf{I}_p \\ \mathbf{I}_p & \dots & \mathbf{I}_p & \mathbf{0} \end{pmatrix} \begin{pmatrix} \Lambda_1 \\ \vdots \\ \Lambda_n \\ \eta \end{pmatrix} = \begin{pmatrix} \mathbf{C}(\mathbf{s}_1 - \mathbf{s}_0) \\ \vdots \\ \mathbf{C}(\mathbf{s}_n - \mathbf{s}_0) \\ \mathbf{I}_p \end{pmatrix},$$

donde η son los multiplicadores de Lagrange.

REFERENCIAS

Se indican algunos textos de interés en relación con los contenidos de este tema.

- N. Cressie (1991). *Statistics for Spatial Data*. Wiley.
- H. Wackernagel (1998). *Multivariate Geostatistics*. Springer.
- R. Olea (1999). *Geostatistics for Engineers and Earth Scientists*. Kluwer.
- R. Webster y M.A. Oliver (2001). *Geostatistics for Environmental Scientists*. Wiley.