

## 12. Cokriging universal

Representamos por  $\{Z_i(x_{i,j}); i = 1, \dots, p; j = 1, \dots, n_i\}$  la muestra a partir de la cual pretendemos predecir  $Z_1(x_0)$ ,  $x_0 \in D$ . Al igual que en el caso univariante, dependiendo de las suposiciones sobre las funciones de tendencia  $m_1(x), \dots, m_p(x)$  se distinguen tres tipos de predicción lineal óptima multivariante:

1. Cokriging simple. Si las funciones de tendencia son conocidas.
2. Cokriging ordinario. Si las funciones de tendencia son desconocidas pero constantes.
3. Cokriging universal. Si las funciones de tendencia son desconocidas.

Evidentemente, el último caso es el más general y, por ello, el más utilizado. En el caso de cokriging universal se asume que cada tendencia  $m_i(x)$  puede expresarse como combinación lineal de funciones regresoras conocidas:

$$m_i(x) = \sum_{l=0}^{L_i} a_{i,l} f_i^l(x); i = 1, 2, \dots, p.$$

El predictor cokriging se define como el mejor predictor lineal insesgado (BLUP) calculado con todas las variables observadas:

$$Z_1^*(x_0) = \sum_{i=1}^p \sum_{j=1}^{n_i} \lambda_{i,j} Z_i(x_{i,j}).$$

1. Condición de insesgadez.

Para que el predictor sea insesgado son condiciones necesarias:

- $\sum_{j=1}^{n_1} \lambda_{1,j} f_1^l(x_{1,j}) = f_1^l(x_0); l = 0, 1, \dots, L_1; x_{1,j} \in S_1.$
- $\sum_{j=1}^{n_i} \lambda_{i,j} f_i^l(x_{i,j}) = 0; l = 0, 1, \dots, L_i; x_{i,j} \in S_i; i = 2, \dots, p.$

Si ambas se verifican, entonces:

$$E[Z_1^*(x_0)] = \sum_{l=0}^{L_1} a_{1,l} f_1^l(x_0) = E[Z_1(x_0)].$$

## 2. Condición de varianza mínima.

Al igual que en el caso unidimensional, se trata de un problema de minimización de varianza de predicción sujeto a las restricciones necesarias para asegurar la insesgadez del predictor. La función a minimizar se plantea utilizando  $L_1 + \dots + L_p + p$  multiplicadores de Lagange:

$$\begin{aligned} Var(Z_1^*(x_0) - Z_1(x_0)) &+ 2 \sum_{l=0}^{L_1} \mu_{1,l} \left( \sum_{j=1}^{n_1} \lambda_{1,j} f_1^l(x_{1,j}) - f_1^l(x_0) \right) \\ &+ 2 \sum_{i=2}^p \sum_{l=0}^{L_i} \mu_{i,l} \left( \sum_{j=1}^{n_i} \lambda_{i,j} f_i^l(x_{i,j}) \right). \end{aligned}$$

Mediante derivadas parciales de la expresión anterior se obtiene la expresión matricial de la cual se deduce el vector de pesos óptimo y la varianza de predicción.

A continuación se muestra el aspecto del sistema matricial a resolver en el caso de  $p = 2$  variables estacionarias de segundo orden y con tendencias  $m_1$  y  $m_2$  desconocidas pero constantes (cokriging ordinario).

$$\begin{pmatrix}
C_{1,1}(x_{1,1}, x_{1,1}) & \dots & C_{1,1}(x_{1,n_1}, x_{1,1}) & C_{2,1}(x_{2,1}, x_{2,1}) & \dots & C_{2,1}(x_{2,n_2}, x_{1,1}) & 1 & 0 \\
\vdots & C_{1,1}(x_{1,i}, x_{1,j}) & \vdots & \vdots & C_{2,1}(x_{2,i}, x_{1,j}) & \vdots & \vdots & \vdots \\
C_{1,1}(x_{1,1}, x_{1,n_1}) & \dots & C_{1,1}(x_{1,n_1}, x_{1,n_1}) & C_{2,1}(x_{2,1}, x_{1,n_1}) & \dots & C_{2,1}(x_{2,n_2}, x_{1,n_1}) & 1 & 0 \\
C_{1,2}(x_{1,1}, x_{2,1}) & \dots & C_{1,2}(x_{1,n_1}, x_{2,1}) & C_{2,2}(x_{2,1}, x_{2,1}) & \dots & C_{2,2}(x_{2,n_2}, x_{2,1}) & 0 & 1 \\
\vdots & C_{1,2}(x_{1,i}, x_{2,j}) & \vdots & \vdots & C_{2,2}(x_{2,i}, x_{2,j}) & \vdots & \vdots & \vdots \\
C_{1,2}(x_{1,1}, x_{2,n_2}) & \dots & C_{1,2}(x_{1,n_1}, x_{2,n_2}) & C_{2,2}(x_{2,1}, x_{2,n_2}) & \dots & C_{2,2}(x_{2,n_2}, x_{2,n_2}) & 0 & 1 \\
1 & \dots & 1 & 0 & \dots & 0 & 0 & 0 \\
0 & \dots & 0 & 1 & \dots & 1 & 0 & 0
\end{pmatrix}
=
\begin{pmatrix}
\lambda_{1,1} & C_{1,1}(x_0, x_{1,1}) \\
\vdots & \vdots \\
\lambda_{1,n_1} & C_{1,1}(x_0, x_{1,n_1}) \\
\lambda_{2,1} & C_{1,2}(x_0, x_{2,1}) \\
\vdots & \vdots \\
\lambda_{2,n_2} & C_{1,2}(x_0, x_{2,n_2}) \\
\mu_1 & 1 \\
\mu_2 & 0
\end{pmatrix}$$

Donde  $C_{r,s}(x_{r,i}, x_{s,j}) = \hat{C}_{r,s}(x_{s,j} - x_{r,i})$ .

## 12.1. Consideraciones prácticas

1. Cuando se trabaja con un proceso espacial multivariante deben calcularse y ajustarse  $p$  semivariogramas (o covariogramas)  $\gamma_{i,i}(h)$  y  $p(p-1)$  semivariogramas (o covariogramas) cruzados  $\gamma_{i,j}(h)$ . Los modelos teóricos escogidos para modelizar los semivariogramas experimentales deben verificar que la varianza de cualquier combinación lineal de estas variables sea siempre no negativa. Dicho de otra forma, debe garantizarse que la varianza de predicción sea siempre no negativa. Para solucionar este problema de elección de modelos suele recurrirse al modelo lineal de correogionalización, que está implementado en  $R$ , y que pasamos a describir:

Supongamos que existen  $q$  variables incorreladas dos a dos:  $Y_1(x), Y_2(x), \dots, Y_q(x)$ , de forma que las variables espaciales pueden expresarse como combinación lineal de ellas:

$$Z_i(x) = m_i(x) + \sum_{j=1}^q \alpha_{i,j} Y_j(x), \forall i = 1, 2, \dots, p.$$

De la expresión anterior, deducimos que un semivariograma (cruzado si  $i \neq j$  o no cruzado si  $i = j$ ) puede expresarse según:

$$\begin{aligned} \gamma_{i,j}(h) &= \alpha_{i,1}\alpha_{j,1}\gamma_1(h) + \alpha_{i,2}\alpha_{j,2}\gamma_2(h) + \dots + \alpha_{i,q}\alpha_{j,q}\gamma_q(h) \\ &= \beta_{i,j}^1\gamma_1(h) + \beta_{i,j}^2\gamma_2(h) + \dots + \beta_{i,j}^q\gamma_q(h), \end{aligned}$$

con  $\beta_{i,j}^k = \alpha_{i,k}\alpha_{j,k}$ .

Obsérvese que los semivariogramas  $\gamma_{i,j}$  son modelos anidados de los semivariogramas que modelizan la dependencia espacial de las variables  $Y_1(x), Y_2(x), \dots, Y_q(x)$ . Además los coeficientes  $\beta_{i,j}^k$  corresponden a los valores de pepita y meseta de las funciones  $\gamma_{i,j}$ .

Para asegurar que la varianza de cualquier combinación lineal de las variables  $Z_i(x)$  sea siempre no negativa debe verificarse que la matriz de coeficientes:

$$\begin{pmatrix} \beta_{1,1}^k & \dots & \beta_{1,p}^k \\ \vdots & \ddots & \vdots \\ \beta_{p,1}^k & \dots & \beta_{p,p}^k \end{pmatrix}$$

sea semidefinida positiva para todos los índices  $k = 1, 2, \dots, q$

Como se trata de una matriz simétrica la condición anterior es equivalente a comprobar que todos los determinantes de las submatrices cuadradas sean no negativos, es decir:

$$|\beta_{1,1}^k| \geq 0, \left| \begin{array}{cc} \beta_{1,1}^k & \beta_{1,2}^k \\ \beta_{2,1}^k & \beta_{2,2}^k \end{array} \right| \geq 0, \dots, \left| \begin{array}{ccc} \beta_{1,1}^k & \dots & \beta_{1,p}^k \\ \vdots & \ddots & \vdots \\ \beta_{p,1}^k & \dots & \beta_{p,p}^k \end{array} \right| \geq 0.$$

2. La función *variogram* de la librería *gstat* del programa *R* calcula covariogramas cruzados y semivariogramas cruzados.

Si se opta por el cálculo de semivariogramas cruzados, *R* utiliza la expresión de semivariograma cruzado en el caso de que las dos variables tengan el mismo número de observaciones y en las mismas coordenadas. En caso contrario calcula el pseudo-semivariograma cruzado. Esta opción se toma de forma automática, sin consultar al usuario.

3. La inclusión de variables secundarias siempre da lugar a predicciones para la variable principal con una varianza de predicción menor. Sin embargo el beneficio obtenido puede no ser significativo en muchos casos. En la práctica se ha comprobado que los sistemas de cokriging mejoran de forma significativa los resultados de los sistemas de kriging cuando hay un submuestreo de la variable principal en comparación con las secundarias,  $n_1 \ll n_i$ ;  $i \neq 1$ , y además las variables secundarias están fuertemente correladas con la principal.
4. En caso de que las variable sean incorreladas dos a dos, los sistemas de cokriging coinciden con los sistemas de kriging.

5. Debido a las restricciones impuestas sobre las variables secundarias:

$$\sum_{j=1}^{n_i} \lambda_{i,j} f_i^l(x_{i,j}) = 0; \quad l = 0, 1, \dots, L_i; \quad i = 2, \dots, p$$

puede demostrarse que los pesos de las variables secundarias son nulos,  $\lambda_{i,j} = 0$  para un  $i$  fijo, si el número de observaciones para esa variable  $i$ -ésima no es superior al número de funciones regresoras, es decir: si  $n_i \leq L_i$ .

6. Antes de calcular una predicción cokriging es aconsejable tipificar las variables, para evitar el efecto que produce trabajar con escalas diferentes.

**Ejercicio 5.** En la secuencia de aprendizaje encontraréis un script de *R*, llamado *Cokriging.r* que debéis ejecutar para contestar las siguientes cuestiones relativas al cokriging.

1. Comenta el coeficiente de correlación, ¿es significativo?
2. En el modelo lineal de correogionalización, identifica los coeficientes y los modelos que se corresponden con lo visto en teoría:

$$\gamma_{i,j}(h) = \beta_{i,j}^1 \gamma_1(h) + \beta_{i,j}^2 \gamma_2(h) + \cdots + \beta_{i,j}^q \gamma_q(h).$$

3. Comprobación de que se verifican las condiciones necesarias para asegurar que las varianzas de predicción serán siempre no negativas: Para asegurar que la varianza de cualquier combinación lineal de las variables  $Z_i(x)$  sea siempre no negativa debe verificarse que ciertas matrices sean semidefinidas positivas. Identifica las matrices en este ejemplo.
4. Ajuste de semivariogramas cruzados experimentales a modelos teóricos: A la vista de los errores cometidos en los ajustes, ¿qué semivariograma ha sido ajustado con menor error? ¿Y cuál con mayor error?
5. Cálculo de predicciones para la variable principal: Viendo las opciones utilizadas en el fichero de ejemplo, ¿qué tipo de cokriging se está utilizando? ¿Las variables son estacionarias de segundo orden o intrínsecamente estacionarias?
6. Representación gráfica de los valores y de las varianzas de predicción: La gráfica representa los valores obtenidos utilizando variables tipificadas pero (sin necesidad de programar en *R*, sólo razonando con la teoría) puedes dar una idea de cómo

crees que serían los semivariogramas teóricos y las varianzas de predicción para las variables no tipificadas?

**Observaciones** para el ejercicio 5.

1. Debéis entregarlo, como muy tarde, el 30 de abril.
2. Espero un fichero de texto con las contestaciones, no tenéis que programar nada.