Capítulo 4

Kriging con Tendencia

Kriging con Tendencia es una extensión de Kriging Ordinario donde la media desconocida no es constante y se asume que es una combinación lineal de funciones base, fijas, definidas para todo punto del dominio.

En lo que sigue se considera que la muestra $A = \{(z_1, x_1), ...(z_j, x_j), ...(z_n, x_n)\}$, C, w, L etc, tienen idéntico significado que el que tienen en Kriging Ordinario.

4.1. Hipótesis

Se presupone que el proceso en cada punto del dominio es la suma de un modelo determinístico de Tendencia (regresión) más un proceso estacionario de segundo orden de media 0. Más precisamente:

$$z(x_j) = \sum_{k=1}^{p} f_k(x_j)\beta_k + e(x_j) \tag{4.1}$$

donde $e(x_j)$ es estacionario de segundo orden con $E(e(x_j)) = 0$. Luego, la aleatoriedad de z_j está dada por el último término y las p funciones, f, representan aportes a la Tendencia existente en cada punto del campo; este aporte es determinístico. Luego,

$$E(z_j) = \sum_{k=1}^{p} f_k(x_j)\beta_k = m(x_j)$$

donde m es la Tendencia (trend) en el punto x_j.

Como la tendencia en ecu. 4.1 es determinística la estructura de covarianza de z es la de e. Luego, $C = Cov(Z) = Cov(E)$ con $E = (e_1, ...e_j, ...e_n)^T$. Se asume, además, que la Tendencia se capta con un número pequeño de funciones $p \ll n$.

4.2. Propósitos de Kriging con Tendencia

Dada la muestra A y un punto arbitrario x_0, donde se desconoce z_0:

1. construir un predictor **lineal insesgado** $\hat{z_0}$ de z_0 de modo de **minimizar la varianza del error** (**BLUP**), donde se entiende por *error* la variable aleatoria $Error = z_0 - \hat{z_0}$, diferencia entre z_0 y el predictor $\hat{z_0}$; y **simultáneamente**:

2. estimar el vector β de los coeficientes de la combinación lineal que definen la Tendencia m.

Para simplificar la notación las componentes de la Tendencia se incluyen en una matriz $F:\ nxp$:

$$F = \begin{bmatrix} f_1(x_1) & f_2(x_1) & ... & f_p(x_1) \\ f_1(x_2) & ... & ... & ... \\ ... & ... & ... & ... \\ f_1(x_n) & ... & ... & f_p(x_n) \end{bmatrix}$$

donde cada fila está formada por los valores de las componentes de la Tendencia en un punto x_j de la muestra.

4.3. Estimación de la media del proceso

Una primera aproximación al problema de estimación de la media $m(x_0)$ es su formulación como un problema de regresión generalizado:

$$Z = F\beta + E \quad Var(E) = C \quad E(E) = 0 \tag{4.2}$$

Sea F_0 el vector de los componentes $f_k(x_0)$ de Tendencia en x_0.

$$F_0^T = [f_1(x_0), f_2(x_0), ..., f_p(x_0)]$$

Si la estructura de covarianza C es conocida la ecu. 4.2 expresa un modelo lineal generalizado cuya solución es:

$$\hat{\beta} = inv\,(F^T\,inv(C)\,F)\,F^T inv(C)\,Z \tag{4.3}$$

Luego, el valor esperado de la media del proceso en un punto x_0 arbitrario es:

$$m(x_0) = F_0^T\hat{\beta} = F_0^T\,inv\,(F^T\,inv(C)\,F)\,F^T inv(C)\,Z$$

Lo que sigue es la solución del problema de regresión generalizado que puede omitirse.

NOTA: Solución del modelo lineal generalizado 4.2

▼Si el modelo es $Z = F\beta + E \quad Var(E) = C \quad E(E) = 0$, sea una matriz A raíz cuadrada de la inversa, $A = inv(C)^{-\frac{1}{2}}$ que existe por ser C definida positiva; entonces si se pre-multiplica por A: $AZ = AF\beta + AE \quad con\, Var(AE) = I$. Se tiene así un modelo lineal clásico $Y = X\beta + \epsilon$ con $AZ = Y$ $AF = X$ $AE = \epsilon$ cuya solución clásica es:

$$\hat{\beta} = inv(X^T X)\, X^T Y$$

o su equivalente:

$$\widehat{\beta} = (F^T A^T AF)^{-1}\, (F^T A^T)AZ = (F^T inv(C)F)^{-1}\, F^T inv(C)Z$$

▲

4.4. Predicción de z_0 y varianza del error

La proposición que sigue muestra que el predictor óptimo tiene una estructura simular al de Kriging Ordinario; en la ecu. 4.6 el primer término tiene los mismos pesos $w^T inv(C)$ que pre-multiplican el desvío de los valores de la muestra, Z, respecto del modelo de Tendencia $F\hat{\beta}$ y el segundo término es simplemente el valor de la Tendencia en x_0. En cuanto a la varianza, es la misma de Kriging Simple más un término positivo asociado a la incertidumbre en la determinación de la Tendencia.

4.4.1. Proposición

Los pesos $\hat{\alpha}$ del predictor lineal óptimo de z_0 , $\alpha^T Z$, están dados por:

$$\widehat{\alpha} = inv(C)\, \left(I - FhF^T inv(C)\right) w + inv(C)\, FhF_0 \tag{4.4}$$

con

$$h = \left(F^T inv(C)\, F\right)^{-1} \tag{4.5}$$

El predictor óptimo:

$$\widehat{\alpha}^T \mathrm{Z} = w^T inv(C)\, \left(Z - F\hat{\beta}\right) + F_0^T \hat{\beta} \tag{4.6}$$

donde $\hat{\beta}$ es la expresión ecu. 4.3 que representa la Tendencia en x_0.
La varianza del error asociada:

$$Var(Error) = \sigma_0^2 - w^T inv(C)\, w + h\left\|F_0 - F^T inv(C)\, w\right\|^2 \tag{4.7}$$

Demostración

▼

1. Que sea insesgado **para todo** β, $E(\hat{z_0}) = E(z_0) \Leftrightarrow \alpha^T E(Z) = \alpha^T F\beta = F_0^T \beta$; luego, para que sea válido para todo β se debe cumplir:

$$\alpha^T F = F_0^T$$

Si $Error = z_0 - \alpha^T Z$ se desea minimizar la varianza del Error:

$$Min\, Var(Error) = Min\, Var(z_0 - \alpha^T Z)$$

La primera condición es un sistema de p ecuaciones en α que tendrá múltiples soluciones en la medida que $n > p$.

El proceso de minimización tendrá que satisfacer estas ecuaciones así que su formulación por Lagrange incluye un vector multiplicador λ, de dimensión p (un multiplicador por cada restricción) :

$$min\, H(\alpha, \lambda) = Var(z_0 - \alpha^T Z) + 2(\alpha^T F - F_0^T)\lambda$$

desarrollando el primer término

$$min\, H(\alpha, \lambda) = Var(z_0) + \alpha^T C\alpha - 2\alpha^T w + 2(\alpha^T F - F_0^T)\lambda$$

Anulando su gradiente respecto de α queda:

$$C\alpha + F\lambda = w$$

y respecto de λ queda :

$$F^T \alpha = F_0$$

Ambas ecuaciones conforman un sistema de ecuaciones de $n + p$ incógnitas $\begin{pmatrix} \alpha \\ \lambda \end{pmatrix}$ que se puede expresar matricialmente:

$$\begin{bmatrix} C & F \\ F^T & 0 \end{bmatrix} \begin{pmatrix} \alpha \\ \lambda \end{pmatrix} = \begin{pmatrix} w \\ F_0 \end{pmatrix}$$

luego la solución es:

$$\begin{pmatrix} \alpha \\ \lambda \end{pmatrix} = \begin{bmatrix} C & F \\ F^T & 0 \end{bmatrix}^{-1} \begin{pmatrix} w \\ F_0 \end{pmatrix}$$

y aplicando la fórmula de inversión de una matriz particionada de ese tipo, queda:

$$\begin{pmatrix} \alpha \\ \lambda \end{pmatrix} = \begin{bmatrix} inv(C)\left(I - FhF^T inv(C)\right) & inv(C)\, Fh \\ hF^T inv(C) & -h \end{bmatrix} \begin{pmatrix} w \\ F_0 \end{pmatrix} \tag{4.8}$$

$$\widehat{\alpha} = inv(C)\ \left(I - FhF^T inv(C)\right) w + inv(C)\, FhF_0$$

$$\widehat{\lambda} = h\left(F^T inv(C)\, w - F_0\right) \tag{4.9}$$

De donde el predictor óptimo de z_0 es:

$$\widehat{\alpha}^T \mathrm{Z} = w^T inv(C)\ \left(I - FhF^T inv(C)\right) Z + F_0^T hF^T\, inv(C)\, Z$$

y sustituyendo la estimación del parámetro ecu. 4.3 $\hat{\beta} = hF^T inv(C)\, Z$

$$\widehat{\alpha}^T \mathrm{Z} = w^T inv(C)\ \left(Z - F\hat{\beta}\right) + F_0^T \hat{\beta}$$

el resultado buscado. En cuanto a la varianza del error, sustituyendo el valor de la estimación de α en

$$Var(z_0 - \alpha^T Z) = Var(z_0) + \alpha^T C\alpha - 2\alpha^T w$$

queda expresada como:

$$Var(Error) = \sigma_0^2 - w^T inv(C)\, w + \left(F_0^T - w^T inv(C)\, F\right) h \left(F_0 - F^T inv(C)\, w\right)$$

o su expresión equivalente dada en la ecu. 4.7.▲

Corolario

Si x_0 está alejado de la muestra, $w = 0$, entonces se predice con la Tendencia $F_0^T \hat{\beta}$ y la varianza del error es mayor o igual a la del proceso $Var(Error) \geq \sigma_0^2$.