Capítulo 3

Kriging Ordinario

Kriging Ordinario es similar a Kriging Simple y sólo se diferencia en el hecho que la media del proceso, $E(z_j) = \mu$, es constante pero desconocida y deberá ser estimada. La estimación de μ también se construye a partir de una combinación lineal de la muestra; además, como $\hat{\mu}$ es un estimador se determinará su varianza para poder hacer inferencia acerca del verdadero valor de μ.

Se sugiere ver previamente los Capítulos I y II (Kriging Simple) ya que la notación y los resultados allí presentados se usan en este capítulo.

3.1. Objetivos de Kriging Ordinario

Dada la muestra $A = \{(z_1, x_1), ...(z_j, x_j), ...(z_n, x_n)\}$ del campo aleatorio en el espacio R^p , es decir: $z_1, ...z_j, ...z_n$ las variables aleatorias definidas en los puntos $x_1, ...x_j, ...x_n$ del campo en consideración, y asumiendo como hipótesis que:

- El proceso es estacionario de segundo orden
- Su estructura de covarianza es conocida
- La media del proceso, $E(z_j) = \mu$, es desconocida

Kriging Ordinario persigue los siguientes objetivos:

1. Construir un estimador **lineal insesgado** $\hat{\mu}$ de μ de **varianza mínima (BLUE),** y determinar dicha varianza con la finalidad de hacer inferencia sobre el verdadero valor de la media μ**.**
2. Dada la muestra A y punto arbitrario x_0, donde se desconoce z_0, 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 $\hat{z}_0$. Además, determinar dicha varianza.

Nótese que tanto $\hat{z_0}$ como *error* son predictores, es decir variables aleatorias!! Sólo cuando se toma una muestra y $z_1, ...z_j, ...z_n$ son valores muestrales específicos de una realización, Kriging Simple genera un mapa de predicción de la realización y otro de la varianza del error. Los resultados que siguen corresponden a propiedades de las variables aleatorias, es decir antes del proceso de tomar valores especificos de $z_1, ...z_j, ...z_n$.

3.2. Estimación de la media μ del proceso y su varianza

3.2.1. Estimación de la media μ

La siguiente proposición establece que el estimador óptimo de la media μ del proceso depende de la posición relativa de los puntos de la muestra, información contenida en la matriz C de covarianza.

Proposición

El estimador lineal insesgado óptimo $\hat{\mu}$ es un promedio ponderado de los valores de la muestra:

$$\widehat{\mu} = \beta^T Z \qquad con \quad \overline{\beta} = \frac{inv(C)\, L}{L^T inv(C)\, L} \tag{3.1}$$

Siendo la varianza mínima del error

$$Var(\widehat{\mu}) = Var(\hat{\beta}^T Z) = \frac{1}{L^T inv(C)\, L} \tag{3.2}$$

Demostración:

▼Se construye un estimador lineal de μ, $(\beta^T Z)$, que sea insesgado y de varianza mínima. Si es insesgado se cumple:
$E(\beta^T Z) = \beta^T E(Z) = \beta^T \mu L = \mu$, luego

$$\beta^T L = 1 \tag{3.3}$$

La suma de los β_j debe ser 1. Es decir, un promedio ponderado de Z.

La varianza del estimador es: $Var(\beta^T Z) = \beta^T C \beta$
Su optimización es un problema de minimización con la restricción ecu.3.3, que se resuelve mediante multiplicadores de Lagrange agregando un parámetro 2δ por la restricción

$$min\, H(\beta, \lambda) = \beta^T C \beta + 2\delta\, (\beta^T L - 1)$$

donde por comodidad el multiplicador se expresa como 2δ. Si anulamos los gradientes respecto de las dos variables de la función $H(\beta, \delta)$ a minimizar :

$$\bigtriangledown_{\beta} = 2C\beta + 2\delta L = 0$$

$$\bigtriangledown_{\delta} = 2(\beta^T L - 1) = 0$$

de la primera:

$$\beta = -\delta inv(C)L$$

y multiplicando a la izquierda por L^Ty usando la segunda igualdad:

$$1 = -\lambda\delta Linv(C)L \qquad \delta = \frac{-1}{L^T inv(C)L}$$

entonces:

$$\overline{\beta} = \frac{inv(C)L}{L^T inv(C)L}$$

En cuanto a su varianza

$$Var(\bar{\beta}^T Z) = \bar{\beta}^T C\beta = \frac{L^T inv(C)}{L^T inv(C)L} C \frac{inv(C)L}{L^T inv(C)L} = \frac{1}{L^T inv(C)\, L} \blacktriangle$$

3.2.2. Corolarios

3.2.2.1. $\overline{\beta}$ es fijo, independiente del muestreo y dado sólo por la estructura de covarianza, $Cov(Z)$.

Como se ha supuesto que la matriz de covarianza depende de la posición de los x_j, y para este conjunto $\{x_1, ...x_j, ...x_n\}$ la matriz de covarianza $Cov(Z)$ es conocida, el estimador $\overline{\beta}$ no depende de la muestra Z, es decir es un vector constante.

3.2.2.2. $\widehat{\mu}$ es aleatorio ya que es un promedio ponderado de los valores de la muestra Z

Es un promedio ponderado ya que $L^T\overline{\beta} = \frac{L^T Cov(Z)^{-1} L}{L^T Cov(Z)^{-1} L} = 1$

3.2.2.3. Los pesos de $\overline{\beta}$ dependen de la posición relativa entre los puntos x y de los parámetros del modelo de covarianza

3.2.2.4. $Var(\widehat{\mu})$ **es conocida previa al muestreo y es proporcional a la varianza del proceso.**

3.2.2.5. Si las observaciones de la muestra están muy alejadas, $C = \sigma^2 I_n$**, entonces** $Var(\widehat{\mu}) = \frac{\sigma^2}{n}$ **y** $\widehat{\mu} = m = \frac{L^T Z}{n}$ **coinciden con los clásicos de la media muestral.**

3.3. Predicción de z_0 y varianza del error

La proposición que sigue resume las características del predictor (BLUP) de Kriging Ordinario:

primero, que la predicción óptima depende de la estructura de covarianza de la muestra, C, de la covarianza entre los puntos de la muestra y el punto donde se desea predecir , w y de la media estimada del proceso $\hat{\mu}$; y segundo, que la varianza del error depende de C, w, de la varianza de la media $Var(\widehat{\mu})$ y de la varianza del proceso, σ^2.

Proposición

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

$$\hat{\alpha} = \bar{\gamma} + Var(\hat{\mu})\, coef\, inv(C)\, L \tag{3.4}$$

o su equivalente sólo en función de C *y* w:

$$\hat{\alpha} = inv(C)w + \frac{1}{L^T inv(C)\, L}(1 - w^T inv(C)L) inv(C)L$$

el predictor por:

$$\widehat{z_0} = \widehat{\alpha}^T \mathrm{Z} = \widehat{\mu} + w^T inv(C)\,(Z - L\widehat{\mu}) \tag{3.5}$$

y respecto del error cometido, la varianza del error está dada por:

$$Var(Error_0) = \sigma^2 - w^T inv(C) w + Var(\hat{\mu})(1 - w^T inv(C)L)^2 \tag{3.6}$$

Demostración:

▼

Si el predictor es insesgado se cumple que $E(\alpha^T Z) = E(z_0) = \mu \Leftrightarrow \alpha^T E(Z) = \alpha^T \mu L = \mu \Leftrightarrow \alpha^T L = 1$ es decir un promedio ponderado.

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

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

De nuevo es un problema de minimización con restricciones, cuyo lagrangiano es:

$$min\, H(\alpha,\lambda) = Var(z_0 - \alpha^T Z) + 2\lambda(\alpha^T L - 1)$$

El primer término es:

$$Var(z_0) + Var(\alpha^T Z) - 2cov(z_0, \alpha Z) = Var(z_0) + \alpha^T C\alpha - 2\alpha^T w \tag{3.7}$$

Luego se minimiza:

$$min\, H(\alpha,\lambda) = Var(z_0) + \alpha^T C\alpha - 2\alpha^T w + 2\lambda(\alpha^T L - 1)$$

Si anulamos los gradientes de la función $H(\alpha,\lambda)$respecto de α y λ se tiene :
$\bigtriangledown_\alpha = 2C\alpha - 2w + 2\lambda L = 0$
$\bigtriangledown_\lambda = 2(\alpha^T L - 1) = 0$
quedando el sistema con dos ecuaciones:

$$C\alpha + \lambda L \quad = \quad w \tag{3.8}$$

$$L^T\alpha \quad = \quad 1 \tag{3.9}$$

Este sistema se puede expresar matricialmente:

$$\begin{bmatrix} C & L \\ L^T & 0 \end{bmatrix} \begin{pmatrix} \alpha \\ \lambda \end{pmatrix} = \begin{pmatrix} w \\ 1 \end{pmatrix} \tag{3.10}$$

Para simplificar los cálculos que siguen se notará $h = \left(L^T inv(C)\, L\right)^{-1} = Var(\hat{\mu})$.
El sistema 3.10 es:

$$\left\{ \begin{array}{c} C\alpha + \lambda L = w \\ L^T\alpha = 1 \end{array} \right\}$$

luego de la primera ecuación $\alpha + \lambda\, inv(C)\, L = inv(C)\, w$, que multiplicando por L^T y aplicando la segunda se tiene:

$$1 + \lambda L^T\, inv(C)\, L = L^T inv(C)\, w$$

usando 2.4 y despejando:

$$\lambda = h(L^T\hat{\gamma} - 1) \tag{3.11}$$

$$\hat{\alpha} = \bar{\gamma} - \lambda\, inv(C)\, L = \bar{\gamma} + h\, coef\, inv(C)\, L \tag{3.12}$$

o su equivalente en función de $\hat{\beta}$ ecu, 3.1

$$\hat{\alpha} = \bar{\gamma} + h\, coef\, inv(C)\, L = \hat{\beta} + \bar{\gamma} - L^T\bar{\gamma}\hat{\beta} \tag{3.13}$$

Luego, el predictor óptimo de z_0 es:

$$\widehat{\alpha}^T Z = \hat{\beta}Z + \bar{\gamma}^T(Z - L\hat{\beta}^T Z)$$

y sustituyendo la media del proceso $\widehat{\mu} = \hat{\beta}^T Z$ se obtiene ecu. 3.5.

En cuanto a la varianza del error mínima, se deduce de las ecuaciones 3.7, 3.8 y 3.9:

$$Var(Error) = Var(z_0) - \alpha^T w - \lambda$$

Sustituyendo α ecu. 3.12 y λ ecu.3.11 se tiene

$$Var(Error) = Var(z_0) - (\bar{\gamma} + h\, coef\, inv(C)\, L)^T w + h\, coef$$

$$Var(Error) = \sigma^2 - w^T inv(C) w - h\, coef\, L^T inv(C) w + h\, coef$$

$$Var(Error) = \sigma^2 - w^T inv(C) w + hcoef^2 \tag{3.14}$$

y sustituyendo h y $coef$ por su valor se obtiene ecu. 3.6▲

corolarios

3.3.1. El predictor es similar al de Kriging Simple

Nótese que la predicción ecu. 3.5 es la media del proceso más un promedio ponderado de los desvíos de los valores de Z respecto de la media $\widehat{\mu}$. Estos pesos son los mismos obtenidos en Kriging Simple ecu. 2.8.

3.3.2. La varianza del error tiene un término adicional a la expresión de Kriging Simple

Los dos primeros términos de la ecu. 3.6 corresponden a la fórmula de Kriging Simple a la que se le agrega el término positivo: $Var(\hat{\mu})(1 - w^T inv(C) L)^2$, es decir se tiene un incremento en la varianza del error que se deriva de la incertidumbre en el predictor de la media $\hat{\mu}$.

3.3.3. El predictor depende de la estructura de correlación

$\widehat{z_0} = \widehat{\alpha}^T \mathrm{Z} = corr(z_0, Z)^T R(Z)^{-1} (Z - L\widehat{\mu}) + \widehat{\mu}$ es decir, los coeficientes de predicción dependen de la estructura de correlación de los puntos de la muestra Z entre sí, y de la correlación entre la nueva variable z_0 y los puntos de la muestra Z.

3.3.4. Caso x_0 alejado de la muestra (ausencia de información)

Si x_0 está suficientemente alejado como para que $w = 0$ entonces el modelo predice con la media $\widehat{\mu}$. En cuanto a la varianza del error: $Var(Error) = \sigma^2 + Var(\hat{\mu})$. A diferencia de Kriging Simple la varianza del error puede superar la varianza del proceso!

3.4. Solución numérica eficiente

Los resultados establecidos en las proposiciones fundamentales de las secciones 3,2 y 3,3 , se obtienen eficientemente usando las ventajas de la solución de sistemas lineales cuando la matriz es definida positiva, como lo son las matrices de covarianza.

Entrada:
La muestra $\{(z_1, x_1), ...(z_j, x_j), ...(z_n, x_n)\}$ y el punto x_o donde predecir z_o
La matriz de covarianza C entre las variables aleatorias de la muestra
El vector de covarianza w entre las variables de la muestra y z_o

El **algoritmo** es el siguiente:

1. Resolver el sistema $C\,aux = L$ (es decir hallar $aux = inv(C)\,L$)
2. Resolver el sistema $C\,\bar{\gamma} = w$ (es decir hallar $\bar{\gamma} = inv(C)\,w$)
3. Calcular $coef = 1 - L^T\bar{\gamma}$
4. Calcular $h = (L^T\,aux)^{-1}$ (Varianza de $\hat{\mu}$)
5. Calcular $\hat{\beta} = h\,aux$
6. Calcular $\hat{\alpha} = \bar{\gamma} + coef\,\hat{\beta}$
7. Calcular $Var(error) = \sigma^2 - w^T\bar{\gamma} + hcoef^2$
8. predecir $\hat{\mu} = \hat{\beta}^T Z$ (media muestral)
9. predecir $\hat{z_o} = \hat{\alpha}^T Z$ (predicción del proceso en el punto x_o)

Nótese

- que los sistemas $1\,y\,2$ se resuelven simultaneamente, y como C es definida positiva la solución de dichos sistemas es eficiente (Cholesky, por ejemplo)
- que los primeros 7 pasos se realizan previos al muestreo donde se determinan la varianza de $\hat{\mu}$ y la varianza del error!
- Que los valores de z obtenidos en el muestreo son sólo necesarios en los pasos $8\,y\,9$ para predecir la media muestral $\hat{\mu}$ y la predicción $\hat{z_o}$.

3.5. Kriging es interpolante y la varianza del error es nula en la muestra

Si z_0 coincide con z_1, entonces w es la primera columna de C luego, $inv(C)\,w = (1, 0, ...,0)^T$, $L^T inv(C)\,w = 1$ y $w^T inv(C)\,w = \sigma^2$ entonces

$$\widehat{\alpha} = (1, 0, ...,0)^T$$

lo que implica que el predictor $\widehat{\alpha}^T \mathrm{Z} = z_1$, es decir es interpolante ya que el modelo de pronóstico asigna el valor muestral.

En cuanto a la varianza del error de la ec. 3.14:

$$Var(Error) = \sigma^2 - \sigma^2 = 0$$

es decir el modelo predice con exactitud en los puntos de la muestra.

3.6. Hipótesis de Normalidad

Si se presupone que para cualquier conjunto $x_1, ...x_j, ...x_k$el vector Z asociado es normal multivariado es posible expresar las distribuciones de los predictores hallados en este capítulo así como obtener intervalos de confianza, test de hipótesis, etc. Sin embargo, debido a la estimación de μ, los predictores de Kriging Ordinario no coinciden con los resultados de la distribución normal multivariada válidos para Kriging Simple.

3.6.1. Distribución de $\widehat{\mu}$

Por lo expresado en la estimación de μ, Ecu. 3.1, $\widehat{\mu}$ es una combinación lineal de Z y es insesgado

$$\widehat{\mu} = \beta^T Z = \frac{L^T C Z}{L^T C L}$$

de varianza conocida, Ecu. 3.2

$$Var(\widehat{\mu}) = Var(\beta^T Z) = \frac{1}{L^T C L}$$

Luego, $\widehat{\mu}$ se distribuye normal $N\left(\mu, Var(\mu)\right)$ y es posible obtener intervalos de confianza de la media, μ, del proceso a partir de la función inversa de la distribución acumulada de la Normal.

3.6.2. Distribución de z_0

En cuanto al predictor de z_0 por la ecu. 3.5 $\widehat{z_0}$ es insesgado y es una C. Lineal de Z

$$\widehat{z_0} = \widehat{\alpha}^T \mathrm{Z} = w^T inv(C)\ (Z - L\widehat{\mu}) + \widehat{\mu}$$

donde la varianza del error, $\widehat{z_0} - z_0$, es constante:

Luego, $z_0 - \widehat{z_0}$ se distribuye normal $N\left(0, Var(Error)\right)$ y es posible obtener intervalos de confianza de z_0, a partir de la función inversa de la distribución acumulada de la Normal.