

Capítulo 2

Kriging Simple

2.1. Introducción

En este capítulo se asume que la estructura de covarianza ya ha sido determinada y que la media del proceso en cualquier punto del dominio es conocida. Sea $A = \{(z_1, x_1), \dots, (z_j, x_j), \dots, (z_n, x_n)\}$ una muestra del campo aleatorio (proceso estocástico) en el espacio R^p , es decir: $z_1, \dots, z_j, \dots, z_n$ las variables aleatorias definidas en los puntos $x_1, \dots, x_j, \dots, x_n$ del campo en consideración.

El propósito de Kriging Simple es predecir -a partir de la muestra escasa A - los valores de z en cualquier punto del dominio mediante una combinación lineal de los valores de la muestra. Puesto que el método se fundamenta en minimizar la varianza del error -diferencia entre los valores de z y su predicción - simultáneamente se obtiene una medida del error en todos los puntos de predicción.

2.1.1. Hipótesis

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

Por ser estacionario la estructura de covarianza entre las variables $z_1, \dots, z_j, \dots, z_n$ está dada por la posición relativa de las $x_1, \dots, x_j, \dots, x_n$, es decir fijadas éstas, la covarianza entre aquéllas es conocida. Más precisamente, existe una función $Cov(h)$ de $R^p \rightarrow R$ función del vector h :

$$cov(z_j, z_k) = Cov(h) \quad \text{donde } h = x_j - x_k$$

2.1.2. Propósito

Dada la muestra A y un nuevo punto arbitrario x_0 , donde se desconoce z_0 , construir un predictor **lineal insesgado** \hat{z}_0 de z_0 a partir de los valores de la muestra $z_1, \dots, z_j, \dots, z_n$ de modo de **minimizar la varianza del error** cometido, donde se entiende por $Error_0 = z_0 - \hat{z}_0$, la variable aleatoria diferencia entre z_0 y el predictor \hat{z}_0 .

Nótese que tanto \hat{z}_0 como $Error$ son variables aleatorias!! Sólo cuando se toma una muestra y $z_1, \dots, z_j, \dots, z_n$ son valores muestrales específicos de una realización, Kriging Simple genera un mapa de predicción de la realización y otro de la varianza del error.

Nota: los resultados que siguen corresponden a propiedades de las variables aleatorias, es decir antes del proceso de tomar valores específicos de $z_1, \dots, z_j, \dots, z_n$.

2.1.3. Por qué asumir media nula?

Como la media, $\mu(x)$, es conocida basta con considerar la variable $z(x) - \mu(x)$ cuyo valor esperado es 0. En lo que sigue se asume que el proceso estocástico tiene media nula, $\mu = 0$. En el caso genérico, basta con sumar la media $\mu(x)$ para restablecer la variable original.

Nótese, además, que Kriging Simple no presupone que la media $\mu(x)$ sea **constante!** sólo que es conocida.

2.2. Notación vectorial

Para simplificar el desarrollo y las demostraciones se usará la siguiente notación vectorial:

las variables aleatorias de la muestra A se incluyen en un vector aleatorio $Z = (z_1, \dots, z_j, \dots, z_n)^T$, entonces la matriz de covarianza $C = cov(Z)$ contiene la covarianza entre las variables aleatorias asociadas a la muestra (ver figura 2.1 a la izquierda)

$$C = Cov(Z) = \begin{bmatrix} cov(z_1, z_1) & \dots & cov(z_1, z_j) & \dots & cov(z_1, z_n) \\ \dots & \dots & \dots & \dots & \dots \\ cov(z_i, z_1) & \dots & cov(z_i, z_j) & \dots & cov(z_i, z_n) \\ \dots & \dots & \dots & \dots & \dots \\ cov(z_n, z_1) & \dots & cov(z_n, z_j) & \dots & cov(z_n, z_n) \end{bmatrix} \quad (2.1)$$

La diagonal principal de C está formada por el mismo valor, $V(z_j) = \sigma^2$ (conocida) y constante por la estacionaridad. La matriz de correlación entre las variables Z , está dada por: $R(Z) = \frac{C(Z)}{\sigma^2}$.

2.3. PREDICCIÓN DE Z_0 Y VARIANZA DEL ERROR

Las covarianzas entre las variables de la muestra y la variable z_0 en un punto a predecir, juega un rol esencial en las predicciones y se incluyen en el vector w , (ver figura 2.1 a la derecha).

El vector w contiene las covarianzas entre z_0 valor de la variable en el punto x_0 de predicción y las variables aleatorias de la muestra: $z_1, \dots, z_j, \dots, z_n$:

$$w = (\text{cov}(z_0, z_1), \dots, \text{cov}(z_0, z_j), \dots, \text{cov}(z_0, z_n))^T \quad (2.2)$$

Nótese que el vector de las correlaciones $\text{corr}(z_0, Z) = \frac{w}{\sigma^2}$.

Por último, se notará $L = (1, 1, 1, \dots, 1)^T$ al vector de unos de igual dimensión que Z . Con esta notación, por ejemplo, la media muestral de Z está dada por $\text{media} = \frac{L^T Z}{n}$.

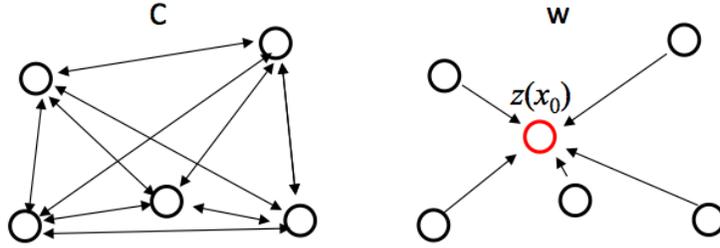


Figura 2.1: Covarianza entre puntos de la muestra y entre la muestra y el punto a predecir

2.3. Predicción de z_0 y varianza del error

Para predecir en un nuevo punto x_0 el valor z_0 del proceso, se asumirá que el predictor es una combinación lineal óptima (Best Linear Unbiased Predictor) de los valores de la muestra $z_1, \dots, z_j, \dots, z_n$, $\hat{z}_0 = \gamma^T Z = \sum_{h=1}^n \gamma_h z_h$; dónde óptimo hace referencia a que de todos los posibles predictores lineales insesgados se hallará aquel que haga mínima la *varianza del error*, $\text{Error}_0 = z_0 - \hat{z}_0$. Nótese que Error_0 es una variable aleatoria ya que para cada realización del proceso, si bien los puntos $x_1, \dots, x_j, \dots, x_n$ y x_0 son fijos los valores de la propiedad $z_1, \dots, z_j, \dots, z_n$ z_0 son aleatorios y en consecuencia el error error_0 es aleatorio y tiene una varianza asociada.

Puesto que $E(\gamma^T Z) = \gamma^T E(Z) = E(z_0) = 0$ toda combinación lineal es insesgada; luego, es la minimización de la varianza del error, $\text{Error}_0 = z_0 - \hat{z}_0$, la que conduce a la determinación de γ .

La siguiente proposición resume los resultados de Kriging Simple: primero, que la predicción óptima depende de la estructura de covarianza de la muestra, C , y de la covarianza entre los puntos de la muestra y el punto donde se desea predecir, w ; y segundo, que la varianza del error depende de estos últimos y de la varianza del proceso, σ^2 .

2.3.1. Proposición

El predictor óptimo \hat{z}_0 es:

$$\hat{z}_0 = \gamma^T Z = w^T \text{inv}(C) Z \quad (2.3)$$

con

$$\bar{\gamma} = \text{inv}(C)w \quad (2.4)$$

y la varianza del error:

$$\text{Var}(\text{Error}) = \text{Var}(z_0 - \gamma^T Z) = \sigma^2 - w^T \text{inv}(C) w \quad (2.5)$$

Demostración

▼ Se desea minimizar el mean square error del $\text{Error} = z_0 - \gamma^T Z$, que es igual a la varianza del Error ya que $E(\text{Error}) = 0$

$$\text{mín } H(\gamma) = \text{Var}(\text{Error}) = \text{Min } \text{Var}(z_0 - \gamma^T Z)$$

La función H a minimizar es función del vector γ , que contiene los parámetros a determinar de la combinación lineal.

Como

$$H(\lambda) = \text{Var}(z_0) + \gamma^T C \gamma - 2\gamma^T w \quad (2.6)$$

para obtener su mínimo se anulará el gradiente de H respecto de γ (ver Capítulo 5, 5.2.2.1):

$$\frac{\partial H}{\partial \lambda} = 2C\gamma - 2w = 0$$

Puesto que $\gamma^T C \gamma$ es una forma cuadrática definida positiva, por ser C una matriz de covarianza, existe la inversa de C y despejando en $C\gamma = w$ se obtienen los resultados ecu.2.4 y ecu. 2.3.

Sustituyendo el valor de ecu. 2.4 en ecu. 2.6 se obtiene ecu. 2.5▲

2.3.2. Corolarios

2.3.2.1. $\bar{\gamma}$ es fijo pero \hat{z}_0 sí es aleatorio

Nótese que $\bar{\gamma}$ es fijo y depende de la posición relativa de la muestra respecto del punto a predecir, pero no depende de los valores de z . En cambio el predictor \hat{z}_0 sí es aleatorio ya que depende de los valores de la propiedad en los puntos de la muestra.

2.3.2.2. La información contenida en la muestra permite reducir la incertidumbre del proceso

Como C es definida positiva también lo es $\text{inv}(C)$, entonces la varianza del error $0 \leq \text{Var}(\text{Error}) \leq \sigma^2$; la información contenida en la estructura de covarianza, (C, w) , permite reducir el error intrínseco del proceso.

2.3.2.3. La varianza del error es fija y depende de la posición de los puntos

La varianza del error no es un estimador, no depende de Z , se conoce a partir de la estructura de covarianza que depende exclusivamente de la posición relativa de los x y de los parámetros del modelo de estructura de covarianza. Sin embargo, en la realidad estos parámetros suelen ser estimados a partir de los valores de Z , lo que induce a un sesgo en la estimación de la varianza verdadera del error. En general, la varianza del error dada por la ecu. 2.5 subestima el verdadero error.

2.3.2.4. Los resultados anteriores son expresables en término de correlaciones:

$$\bar{\gamma} = \text{corr}(z_0, Z)^T R^{-1}$$

$$\text{Var}(\text{Error}) = \sigma^2 (1 - \text{corr}(z_0, Z)^T R^{-1} \text{corr}(z_0, Z))$$

2.3.2.5. Los pesos, $\bar{\gamma}$, no dependen de la varianza, σ^2 , del proceso

Del corolario anterior se deduce que los coeficientes de la combinación lineal, $\bar{\gamma}$, no dependen de la varianza σ^2 del proceso ya que son función de las correlaciones de los puntos de Z entre sí y de las correlaciones ente éstos y el nuevo punto z_0 . Es decir si se considera la variable $Y = kZ$, k constante los coeficientes en la estimación de Y serán los mismos de los de Z . En cambio la varianza del error tiene el factor σ^2 delante, es decir la varianza del error depende del ruido del proceso.

2.4. Propiedades relevantes

2.4.1. Kriging es interpolante

La predicción en un punto de la muestra coincide con el valor en la muestra.
Demostración:

▼ Si z_0 coincide con z_1 , entonces w es la primera columna de C luego, $\text{inv}(C)w = (1, 0, \dots, 0)^T$ ya que es la primera columna del producto $\text{inv}(C)*C = I$; resulta entonces que $\hat{z}_0 = \bar{\gamma}^T Z = w^T \text{inv}(C) Z = z_1$ ▲

2.4.2. En puntos de la muestra la varianza del error es cero

La varianza del error en un punto de la muestra es nula.

Demostración:

▼ puesto que si z_0 coincide con z_1 , $\text{inv}(C)w = (1, 0, \dots, 0)^T$ entonces $w^T \text{inv}(C)w = \sigma^2$

Sustituyendo en ecu. 2.5 queda $\text{Var}(\text{error}) = \sigma^2 - w^T \text{inv}(C)w = 0$ ▲

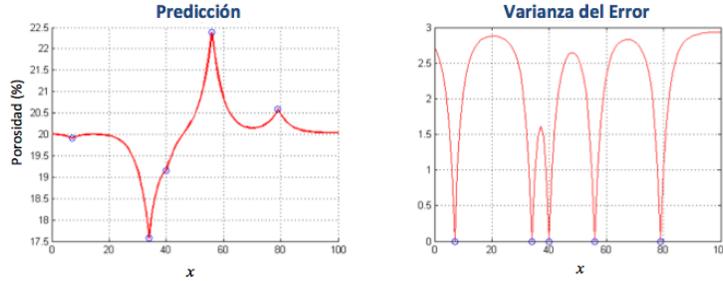


Figura 2.2: Kriging Simple es interpolante y la varianza del error es nula en la muestra

En la figura 2.2, a la izquierda se observa que la predicción pasa por los puntos de la muestra, y a la derecha que en dichos puntos la varianza del error es 0.

2.4.3. Independencia del Error y la muestra

Demostración



La covarianza entre la predicción $\hat{z}_0 = \mu + w^T C^{-1}(Z - \mu L)$ y Z es: $cov(\hat{z}_0, Z) = w^T C^{-1} cov(Z, Z) = w^T C^{-1} C = w = cov(z_0, Z)$

Luego, la diferencia entre el primero y último término da $cov(Error_0, Z) = 0$ ▲

2.4.4. Independencia de la predicción y el error en la predicción

Para cualquier punto arbitrario la predicción y el error cometido, ambos aleatorios!! son independientes ya que por el punto anterior el error es independiente de la muestra y la predicción es una combinación lineal de la muestra.

2.4.5. La suma de las varianzas del error y de la predicción es constante, σ^2

Como $z_0 = \hat{z}_0 + z_0 - \hat{z}_0 = \hat{z}_0 + Error$, entonces por la independencia recién probada y recordando que la varianza de la suma es igual a la suma de las varianzas en el caso de independencia, se cumple:

$$\sigma^2 = Var(z_0) = Var(\hat{z}_0) + Var(Error) \quad (2.7)$$

con ambos sumandos no negativos.

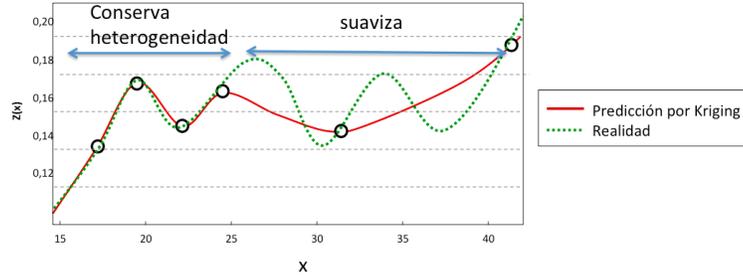


Figura 2.3: Kriging Simple suaviza pero no es un filtro pasa bajos

2.4.6. Suavidad de Kriging Simple

Por la ecu. 2.7 la predicción de Kriging Simple siempre tiene una varianza del error acotada superiormente por la del proceso, σ^2 , en consecuencia Kriging Simple reduce la incertidumbre. Además, en zonas más allá del rango la predicción es muy suave ya que es la media, constante. Sin embargo, Kriging no es un filtro pasa bajos ya que en zonas donde la muestra es abundante la heterogeneidad de la realización y la de Kriging Simple es la misma por ser interpolante.

En la figura 2.3 se observa que por la proximidad de los primeros cuatro puntos (a la izquierda en la figura) la verdadera función y la predicción son coincidentes y la heterogeneidad es la misma; en cambio, en la zona de la derecha donde hay pocos puntos la predicción por Kriging es suave.

2.5. Algunos casos especiales

2.5.1. Cuando los puntos de la muestra están muy alejados entre sí

Si los puntos de la muestra están suficientemente alejados como para que la correlación entre ellos sea nula (más allá del rango) la predicción está dada por:

$$\hat{z}_0 = \frac{1}{\sigma^2} w^T Z = \text{corr}(z_0, Z)^T Z$$

y la varianza del error:

$$\text{Var}(\text{error}) = \sigma^2 - \frac{1}{\sigma^2} w^T w$$

Demostración:

▼ Si $\text{corr}(z_j, z_k) = 0$ $j \neq k$, entonces $C = C = \sigma^2 I_n$ e $\text{inv}(C) = \frac{1}{\sigma^2} I_n$. Sustituyendo en ecu. 2.3 y ecu. 2.5 se obtiene el resultado. ▲

Obsérvese que los pesos están dados por la influencia de cada variable de la muestra $z_1, \dots, z_j, \dots, z_n$ en el nuevo punto (x_0, z_0) medida por la correlación existente entre cada z_j y z_0 , que a su vez esta correlación es función del vector $h_j = x_j - x_0$.

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

Si x_0 está suficientemente alejado de la muestra como para que las correlaciones entre z_0 y los valores muestrales $z_1, \dots, z_j, \dots, z_n$ sean nulas, Kriging Simple predice con la media del proceso y la varianza del error es máxima, σ^2 .

Demostración:

▼ Si $cov(z_0, z_j) = 0 \quad j = 1 \dots n$, entonces $w = 0$ y sustituyendo en ecu. 2.3 y ecu.2.5, $\hat{z}_0 = 0$, y $Var(error) = \sigma^2$.▲

2.5.3. Influencia de las observaciones más allá de su rango

Si el punto x_0 donde se desea predecir z está fuera del alcance de las observaciones excepto de la x_k , es decir, todas las covarianzas w_j son nulas, salvo la w_k podría pensarse que la predicción sólo dependerá del punto (z_k, x_k) , sin embargo, la predicción es $\hat{z}_0 = w_k inv(C)_{(k, \cdot)} Z$, donde $inv(C)_{(k, \cdot)}$ es la fila k de $inv(C)$. En consecuencia, todos los valores de la muestra influyen en la determinación de la predicción a través de la matriz inversa de la covarianza, y los valores de Z de la muestra, a pesar de que sólo z_k está correlacionado con z_0 . En la figura 2.4 se desea predecir en x_0 ; x_1 es el único punto que se encuentra a menos del rango de x_0 , sin embargo toda la muestra influye en la predicción.

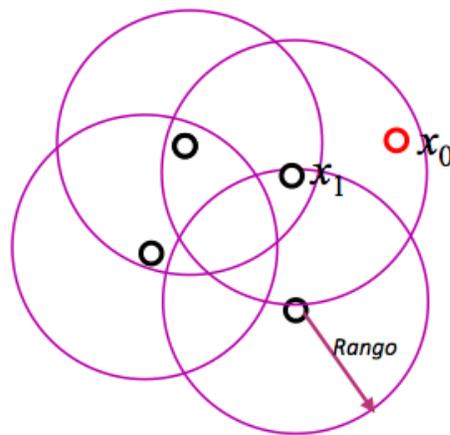


Figura 2.4: Influencia de las observaciones más allá del rango

2.6. Desagrupamiento (Declustering)

Cuando un subconjunto de la muestra está agrupado -formando un “cluster” - Kriging Simple les asigna menos peso del que tendrían por su posición relativa respecto del punto a estimar.

2.7. KRIGING SIMPLE CUANDO LA MEDIA $\mu(x)$ ES CONOCIDA Y NO NULA

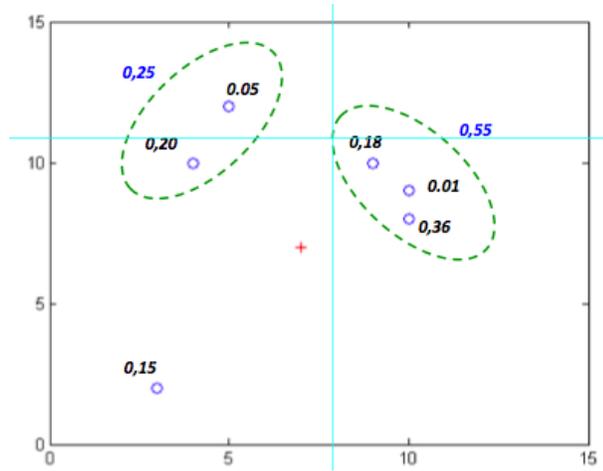


Figura 2.5: Kriging Simple reconoce agrupamientos y reduce sus pesos

En la figura 2.5 se muestran los pesos asociados a los puntos de la muestra. En el agrupamiento de la derecha se observa un punto con un peso de 0,01 a pesar de estar más cerca del punto a predecir (+) que el ubicado abajo a la izquierda cuyo peso es 0,15. Es la conformación de $inv(C)$ la responsable de reducir los pesos individuales y asignarle un peso a la clase.

2.7. Kriging Simple cuando la media $\mu(x)$ es conocida y no nula

Si consideramos el proceso de media conocida, $\mu(x) \neq 0$, la nueva variable $y = z - \mu(x)$ tiene media 0, y si se le aplica lo antes expuesto: $\hat{y}_0 = w^T inv(C) Y$, luego, se modifica el predictor de z_0 y la varianza del error es la misma.

$$\hat{z}_0 = w^T inv(C) (Z - \mu(x) L) + \mu(x) \quad (2.8)$$

Este resultado indica que la predicción es la media del proceso más una combinación de los desvíos de Z respecto de la media $\mu(x) \neq 0$, y que los coeficientes son los mismos $\bar{\gamma}$ del problema original de media nula.

Una expresión alternativa es la siguiente:

$$\hat{z}_0 = w^T inv(C) Z + coef \mu(x) \quad (2.9)$$

con

$$coef = 1 - w^T inv(C) L \quad (2.10)$$

la predicción es la misma expresión del caso de media nula más el producto de una constante, $coef$, por la media que permite que $E(\hat{z}_0) = \mu(x)$. Este coeficiente $coef$ reaparecerá en el capítulo sobre Kriging Ordinario.

2.8. Kriging Simple como predictor determinístico

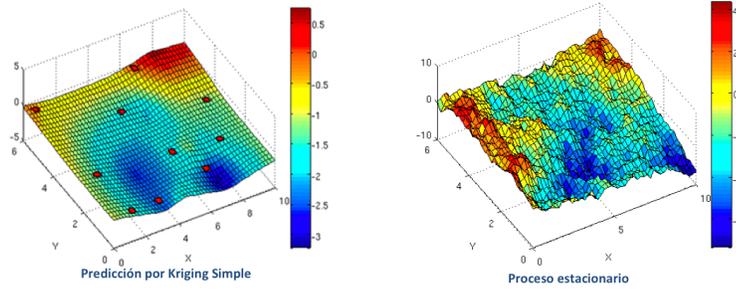


Figura 2.6: Kriging Simple como predictor determinístico

Cuando la muestra $A = \{(z_1, x_1), \dots, (z_j, x_j), \dots, (z_n, x_n)\}$ toma valores fijos de una realización, es decir los z_j son números dados y no variables aleatorias, la predicción $\hat{z}(x)$ en un punto arbitrario x del dominio, puede considerarse como un predictor determinístico:

$$\hat{z}(x) = w^T(x) \text{inv}(C)Z + \text{coef} \mu(x) = Z^T \text{inv}(C)w(x) + \text{coef} \mu(x)$$

Como $d = Z^T \text{inv}(C)$ es constante la predicción es una combinación lineal $d^T w(x)$ de las covarianzas entre los valores de z en el punto a predecir y los de la muestra dada, más la constante $\text{coef} \mu(x)$.

o su equivalente:

$$\hat{z}(x) = w^T(x) \text{inv}(C) (Z - \mu L) + \mu(x)$$

siendo $\text{inv}(C) (Z - \mu L)$ constante la predicción es la media del proceso más una combinación lineal de las covarianzas entre los valores de z en el punto a predecir y los de la muestra dada.

Nota: el predictor determinístico suave, e interpolante, no es lineal en x ! -inclusive si la media es constante- ya que $w(x)$ no es lineal en x . Por ejemplo, en la figura 2.5 se observa que la realización tiene las características de una spline. En cuanto a la suavidad, ésta depende del tamaño de la muestra y del modelo de covarianza elegido.

2.9. Hipótesis de Normalidad

Si se presupone que para cualquier conjunto $x_1, \dots, x_j, \dots, x_k$ 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.

2.9.1. Distribución de \hat{z}_0

\hat{z}_0 se distribuye $N(\hat{z}_0, Var(Error))$ y es posible obtener intervalos de confianza de z_0 , a partir de la función inversa de la distribución acumulada de la Normal.

Demostración

▼

El predictor de z_0 , \hat{z}_0 es insesgado y es una combinación lineal de Z

$$\hat{z}_0 = w^T inv(C) Z$$

donde la varianza del error, $\hat{z}_0 - z_0$, es constante:

$$Var(Error) = \sigma^2 - w^T inv(C) w$$

Luego, $\hat{z}_0 - z_0$ se distribuye Normal $N(0, Var(Error))$ ▲

2.9.2. distribución condicional multivariada

Por lo expuesto en el apéndice acerca de la distribución Normal multivariada la estimación de la media y la varianza del error coinciden con las de la distribución Normal condicionada de z_0 dado Z .

Demostración

▼Si

$$Zamp = \begin{bmatrix} z_0 \\ Z \end{bmatrix}$$

μ y Σ particionados:

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \text{ de tamaño } \begin{bmatrix} 1 \\ N \end{bmatrix}, \text{ respectivamente,}$$

$$\Sigma = \begin{vmatrix} \sigma^2 & w^T \\ w & C \end{vmatrix}$$

de tamaño

$$\Sigma = \begin{vmatrix} 1 \times 1 & 1 \times n \\ n \times 1 & n \times n \end{vmatrix}$$

entonces la distribución de z_0 condicional a Z (*conocido*) es multivariada normal $N(\bar{\mu}, \bar{\Sigma})$

donde

$\bar{\mu} = \mu_1 + w^T inv(C) (Z - \mu_2)$ y varianza $\bar{\Sigma} = \sigma^2 - w^T inv(C) w$ que para el caso de estacionaridad queda:

$$\bar{\mu} = \mu + w^T inv(C) (Z - \mu L)$$

Estos resultados coinciden con los de Kriging Simple sin la hipótesis de Normalidad. ▲

Resumiendo, cuando el proceso subyacente es gaussiano Kriging Simple es equivalente a una predicción normal multivariada condicionada a la muestra.

2.10. Kriging no es estacionario

Aunque el campo aleatorio original sea estacionario la predicción por Kriging Simple no lo es.

Si a y b son dos puntos del campo tal que $b - a = h$

$$\text{cov}(\hat{z}_a, \hat{z}_b) = w_a^T \text{inv}(C) \text{cov}(Z) \text{inv}(C) w_b = w_a^T \text{inv}(C) w_b$$

depende de la posición de los puntos; por ejemplo, se anula si cualquiera de los puntos está fuera del rango de la muestra. Tampoco es constante la varianza de la predicción ver ecu. 2.7 que oscila entre 0 y σ^2 .