

Predicción Geoestadística: Kriging Simple, Ordinario y con Tendencia

Salvador Pintos ¹

noviembre 2012

¹Applied Computing Institute, University of Zulia

Resumen

Este reporte contienen una formulación alternativa basada en vectores aleatorios de los resultados fundamentales de predicción geoestadística. Para evitar los complejos sistemas lineales y el exceso de sumatorias que les está asociado -tal como aparecen en los textos habituales- se recurre al poder de síntesis del álgebra lineal. Si bien su aplicación directa es en ciencias de la tierra, los resultados de Kriging se presentan para cualquier dimensión finita, permitiendo su uso en problemas de aproximación y optimización en (R^p) . El capítulo **1** es una introducción a procesos estocásticos y a los conceptos de continuidad espacial y estacionaridad. El capítulo **2** contiene los resultados fundamentales de Kriging Simple. El **3**, Kriging Ordinario; el **4**, Kriging con Tendencia. En el apéndice se incluye un resumen de álgebra lineal y vectores aleatorios con los instrumentos necesarios para derivar los resultados fundamentales de Kriging. Pensando en aquel lector a quien las demostraciones le provocan pánico, éstas aparecen deliberadamente entre los símbolos \blacktriangledown y \blacktriangle para que, en una primera lectura, pueda conocer los resultados y sus consecuencias sin necesidad de atascarse en el cómo.

Capítulo 1

Procesos estocásticos en ciencias de la tierra

Una de las características observables de las variables que se consideran en ciencias de la tierra, como la porosidad o la permeabilidad en la caracterización de yacimientos petrolíferos, es la de presentar patrones de continuidad espacial debido al proceso crono-depositacional. Por otra parte sólo se conocen escasas mediciones -directas o indirectas- de estas propiedades, por ejemplo, mediciones en los pozos.

Es así que la predicción de estas variables en puntos arbitrarios del campo en estudio deberá tener en cuenta la incertidumbre intrínseca de esas predicciones y establecer una medida de la misma.

Los métodos de predicción determinísticos como interpolación inversa, el k-vecino próximo o splines hacen énfasis en la predicción; en cambio la propuesta geoestadística, Kriging -que se fundamenta en los patrones de continuidad- privilegia el control de la incertidumbre.

Antes de explicar los fundamentos de Kriging se presentarán algunos conceptos básicos de procesos estocásticos.



Figura 1.1: Imagen geológica

1.1. El paradigma estocástico

Sea D un subconjunto del plano, R^2 o del espacio R^3 y $z(x)$ una función (propiedad petrofísica, por ejemplo) objeto de estudio definida en D . El *paradigma estocástico* consiste en considerar a $z(x)$ como una realización de un proceso estocástico $Z(x, w)$. Se entiende que es un proceso estocástico si para cada $x_o \in D$, fijo, $Z(x_o, w)$ es una variable aleatoria, y para cada w_o fijo $Z(x, w_o)$ -realización- es una función de x . Más precisamente, existe w_k tal que la función de interés, $z(x)$, es la realización $z(x) = Z(x, w_k)$. En lo que sigue se notará z_k a la variable aleatoria $Z(x_k, w)$ salvo que sea necesario enfatizar la posición x_k y en cuyo caso se indicará el $z(x)$.

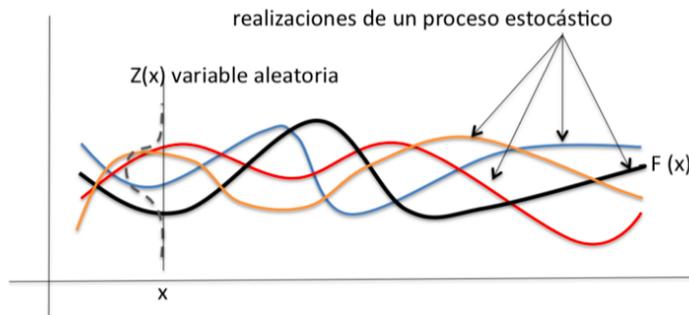


Figura 1.2: Realización de un proceso estocástico

En la Figura 1.2 una variable $F(x)$, por ejemplo saturación de agua, se la identifica con una realización de un proceso estocástico; para un x fijo la incertidumbre acerca del verdadero valor de la saturación queda contemplada en la variable aleatoria $Z(x)$.

1.1.1. La continuidad espacial

Lejos de ser un ruido blanco, estos procesos asociados a variables geofísicas presentan una alta correlación en puntos, x_1, x_2 , cercanos del dominio, y la covarianza entre las variables aleatorias $Z(x_1, w)$ y $Z(x_2, w)$ asociadas,

$$\text{Cov}(Z(x_1, w), Z(x_2, w)), \quad x_1, x_2 \text{ arbitrarios} \quad (1.1)$$

caracteriza la continuidad espacial. La posibilidad de predecir en un punto arbitrario a partir de una muestra escasa dada -tal como es común en ciencias de la tierra- y reducir la incertidumbre inherente al proceso estocástico depende del alcance de la covarianza.

1.1.2. Geoestadística

Si el dominio D es una región compacta y conexa del plano o el espacio, ($E = R^2$, o R^3), como es el caso de una región de la tierra, entonces los proce-

Los procesos estocásticos fundamentales se les define como Campos Aleatorios (Random Fields) y son objeto de estudio de la Geoestadística. Otro ejemplo de procesos estocásticos son las Series de Tiempo donde el conjunto D es un conjunto de tiempos igualmente espaciados $D = \{t : t = n \times t_o\}$.

1.2. Procesos estacionarios

Si la función de covarianza, ecu. 1.1, es conocida para cualquier par de puntos del dominio es posible realizar predicciones lineales óptimas tal como se presentan en los capítulos 2 y 3. Sin embargo, puesto que en general sólo se dispone de una muestra escasa de n pares $A = \{(z_k, x_k), 1 \leq k \leq n\}$ de una realización $z(x)$, la inferencia de la estructura de covarianza ecu. 1.1, se fundamenta en una hipótesis adicional, *estacionaridad*.

Se define que el proceso $Z(x, w)$ es *estacionario* si para cualquier conjunto finito $\{x_1, \dots, x_j, \dots, x_m\}$ arbitrario la distribución conjunta de las variables aleatorias asociadas $\{z_1, \dots, z_j, \dots, z_m\}$ es invariante a traslaciones del conjunto $\{x_1, \dots, x_j, \dots, x_m\}$.

Más precisamente para cualquier vector h , los conjuntos $\{x_1, \dots, x_j, \dots, x_n\}$ y $\{x_1+h, \dots, x_j+h, \dots, x_n+h\}$ tienen sus variables aleatorias asociadas $\{z_1, \dots, z_j, \dots, z_n\}$ con idéntica distribución conjunta de probabilidad.

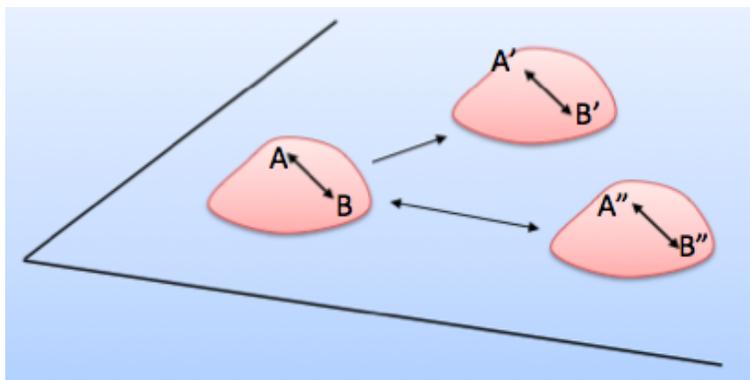


Figura 1.3: Proceso estacionario

La figura 1.3 presenta tres regiones que se corresponden en una traslación. Si el proceso es estacionario se cumple: que las medias y varianzas en puntos homólogos son iguales $E(Z(A)) = E(Z(A')) = E(Z(A''))$; $Var(Z(A)) = Var(Z(A')) = Var(Z(A''))$; que las covarianzas entre pares homólogos son iguales: $Cov(Z(A), Z(B)) = Cov(Z(A'), Z(B')) = Cov(Z(A''), Z(B''))$. Estos tres resultados dan origen a una condición más débil que la estacionaridad estricta que se presenta a continuación.

1.2.1. Procesos estacionarios de segundo orden

Si el conjunto finito se reduce a un punto, $m = 1$ todas las variables z_j están igualmente distribuidas entonces:

- $E(z_j) = \mu$ constante $\forall z_j$
- $Var(z_j)$ constante $\forall z_j$

Si el conjunto finito se reduce a dos puntos $\{x_1, x_2\}$, $m = 2$, las covarianzas son iguales ante traslaciones:

$$cov(z(x_1), z(x_2)) = cov(z(x_1 + h), z(x_2 + h)) \quad \forall h \quad (1.2)$$

El proceso es *estacionario de segundo orden* si se cumplen las 3 propiedades arriba indicadas.

Si bien es una hipótesis más débil que la estacionaridad es suficiente para múltiples propósitos en geoestadística, por ejemplo predicción por Kriging. Además, si el proceso es gaussiano (las distribuciones conjuntas Normalmente distribuidas), estacionario de segundo orden y estacionario son equivalentes ya que la distribución Normal Multivariada queda caracterizada por los momentos de primer y segundo orden.

1.2.2. Existencia de una función de covarianza $Cov(h)$

La estacionaridad de segundo orden implica la existencia de una función de covarianza $Cov(h)$ tal que:

$$cov(z(x_1), z(x_2)) = Cov(h) \quad h = x_2 - x_1 \quad (1.3)$$

siendo h el vector del espacio diferencia entre los puntos dados.

Demostración

▼ Fijado un punto arbitrario del dominio, x_{fijo} , entonces, para $k = x_{fijo} - x_1$ sustituyendo en ecu. 1.2 $cov(z(x_1), z(x_2)) = cov(z(x_1 + k), z(x_2 + k))$

luego, $cov(z(x_{fijo}), z(x_2 - x_1 + x_{fijo})) = cov(z(x_{fijo}), z(x_{fijo} + h))$

Esta expresión sólo depende de h ; luego, la covarianza entre dos variables $z(x_1), z(x_2)$ depende del vector $h = x_2 - x_1$ y no de las posiciones en sí de x_1, x_2

▲

1.2.3. Positividad

Si el espacio es el habitual R^3 , entonces la función Cov es de $R^3 \Rightarrow R$. $Cov(h)$ no es cualquier función ya que debe tener la característica de generar matrices definidas positivas. Cuando Cov es de $R \Rightarrow R$ (caso isotrópico, ver punto siguiente) son modelos válidos: exponencial, esférico, gaussiano, Matern, etc.

1.2.4. Isotropía

El proceso es *isotrópico* si la dependencia es más simple aún y sólo depende de la distancia $\|h\|$ y no de su dirección. Es decir $cov(z(x_1), z(x_2)) = Cov(\|h\|)$ y es una gran simplificación ya que la función Cov en lugar de ser de $R^3 \Rightarrow R$, es de $R \Rightarrow R$, y es una función más simple y por ende más fácil de modelar. La isotropía implica que las iso-superficies son esferas en R^3 o circunferencias en R^2 . Si Cov es dependiente de la dirección del vector además de su módulo se afirma que el campo aleatorio es *anisotrópico* y existen distintos tipos de anisotropía.

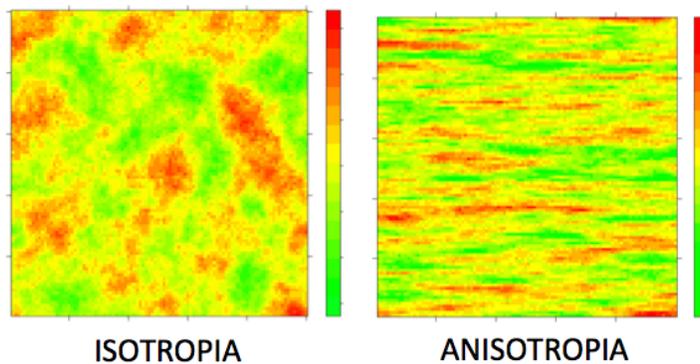


Figura 1.4: Realizaciones de un proceso isotrópico (izq) y anisotrópico (der)

La figura 1.4 muestra a la izquierda una realización de un proceso isotrópico, y a la derecha un anisotrópico donde se observan extensas zonas horizontales casi constantes en contraste con los rápidos cambios en vertical.

1.3. Análisis y modelado de la continuidad espacial

Este tema, fundamental en geostatística, que permite inferir un modelo de la estructura de covarianza a partir de una muestra, no se incluye en estas notas orientadas a predicción por Kriging.

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 $\text{media} = \frac{L^T Z}{n}$.

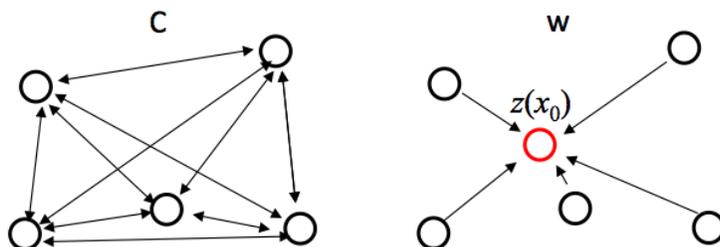


Figura 2.1: Covarianza entre puntos de la muestra y entre la muestra y el punto a predecir

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 $\text{Var}(\text{error}) = \sigma^2 - w^T \text{inv}(C)w = 0$ ▲

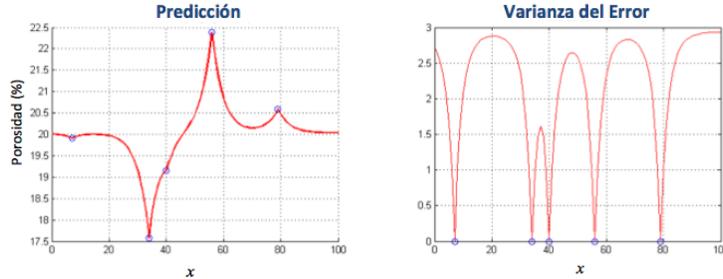


Figura 2.2: Kriging Simple es interpolante y la varianza del error es nula en la muestra

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.

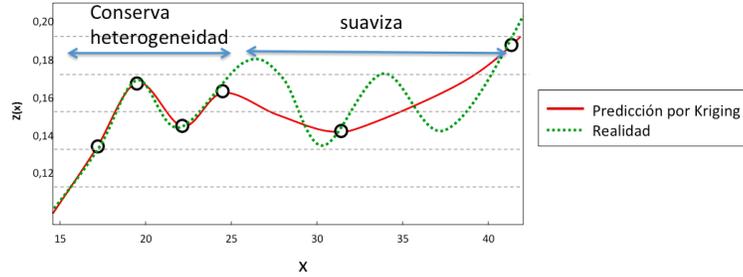


Figura 2.3: Kriging Simple suaviza pero no es un filtro pasa bajos

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 $C = \sigma^2 I_n$ 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.

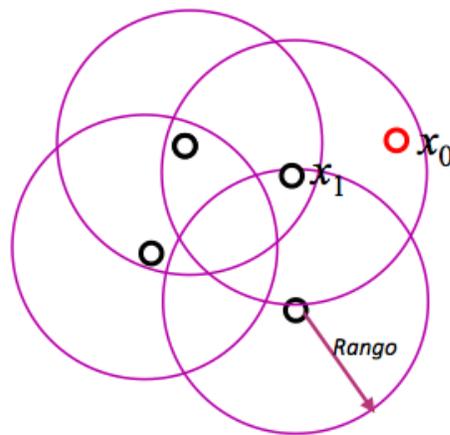


Figura 2.4: Influencia de las observaciones más allá del rango

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

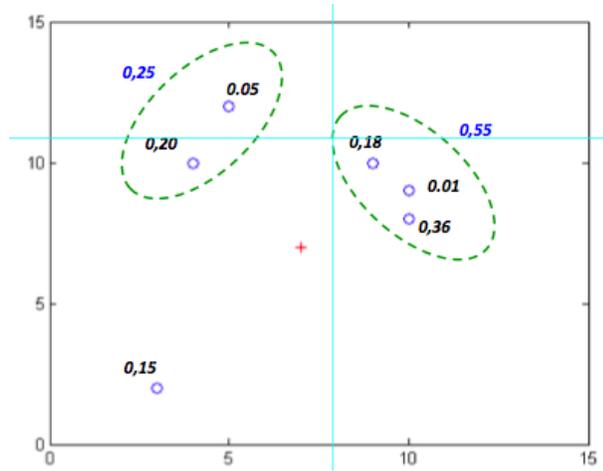


Figura 2.5: Kriging Simple reconoce agrupamientos y reduce sus pesos

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

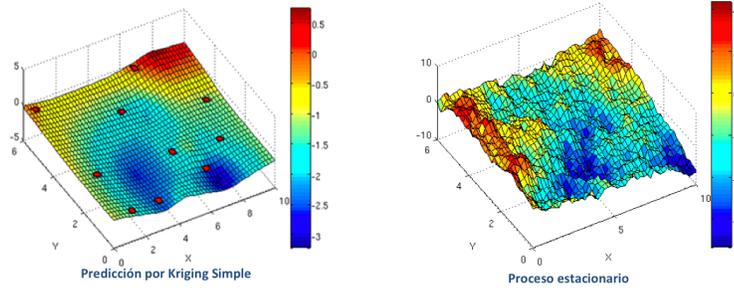


Figura 2.6: Kriging Simple como predictor determinístico

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

$$Z_{amp} = \begin{bmatrix} z_0 \\ Z \end{bmatrix}$$

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 .

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 $A = \{(z_1, x_1), \dots, (z_j, x_j), \dots, (z_n, x_n)\}$ del campo aleatorio 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, 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 $z_1, \dots, z_j, \dots, z_n$ son valores muestrales específicos de una realización, Kriging Simple genera un mapa de predicción de la realización y otro de la varianza del error. Los resultados que siguen corresponden a propiedades de las variables aleatorias, es decir antes del proceso de tomar valores específicos de $z_1, \dots, z_j, \dots, z_n$.

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:

$$\hat{\mu} = \beta^T Z \quad \text{con} \quad \bar{\beta} = \frac{\text{inv}(C) L}{L^T \text{inv}(C) L} \quad (3.1)$$

Siendo la varianza mínima del error

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

Demostración:

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

$$E(\beta^T Z) = \beta^T E(Z) = \beta^T \mu L = \mu, \text{ luego}$$

$$\beta^T L = 1 \quad (3.3)$$

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

$$\text{La varianza del estimador es: } \text{Var}(\beta^T Z) = \beta^T C \beta$$

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

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

3.2. ESTIMACIÓN DE LA MEDIA μ DEL PROCESO Y SU VARIANZA

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 :

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

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

de la primera:

$$\beta = -\delta \text{inv}(C)L$$

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

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

entonces:

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

En cuanto a su varianza

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

3.2.2. Corolarios

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

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

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

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

- 3.2.2.3.** Los pesos de $\bar{\beta}$ dependen de la posición relativa entre los puntos x y de los parámetros del modelo de covarianza
- 3.2.2.4.** $Var(\hat{\mu})$ 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 $Var(\hat{\mu}) = \frac{\sigma^2}{n}$ y $\hat{\mu} = m = \frac{L^T Z}{n}$ 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 $Var(\hat{\mu})$ 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:

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

o su equivalente sólo en función de C y w :

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

el predictor por:

$$\hat{z}_0 = \hat{\alpha}^T Z = \hat{\mu} + w^T \text{inv}(C) (Z - L\hat{\mu}) \quad (3.5)$$

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

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

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:

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

3.3. PREDICCIÓN DE Z_0 Y VARIANZA DEL ERROR

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

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

El primer término es:

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

Luego se minimiza:

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

Si anulamos los gradientes de la función $H(\alpha, \lambda)$ respecto de α y λ se tiene :

$$\nabla_{\alpha} = 2C\alpha - 2w + 2\lambda L = 0$$

$$\nabla_{\lambda} = 2(\alpha^T L - 1) = 0$$

quedando el sistema con dos ecuaciones:

$$C\alpha + \lambda L = w \quad (3.8)$$

$$L^T \alpha = 1 \quad (3.9)$$

Este sistema se puede expresar matricialmente:

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

Para simplificar los cálculos que siguen se notará $h = (L^T \text{inv}(C) L)^{-1} = \text{Var}(\hat{\mu})$.

El sistema 3.10 es:

$$\begin{cases} C\alpha + \lambda L = w \\ L^T \alpha = 1 \end{cases}$$

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

$$1 + \lambda L^T \text{inv}(C) L = L^T \text{inv}(C) w$$

usando 2.4 y despejando:

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

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

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

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

Luego, el predictor óptimo de z_0 es:

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

y sustituyendo la media del proceso $\hat{\mu} = \hat{\beta}^T Z$ 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:

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

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

$$\begin{aligned} \text{Var}(\text{Error}) &= \text{Var}(z_0) - (\bar{\gamma} + h \text{coef inv}(C) L)^T w + h \text{coef} \\ \text{Var}(\text{Error}) &= \sigma^2 - w^T \text{inv}(C) w - h \text{coef} L^T \text{inv}(C) w + h \text{coef} \end{aligned}$$

$$\text{Var}(\text{Error}) = \sigma^2 - w^T \text{inv}(C) w + h \text{coef}^2 \quad (3.14)$$

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 $\hat{\mu}$. 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: $\text{Var}(\hat{\mu})(1 - w^T \text{inv}(C) L)^2$, 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

$\hat{z}_0 = \hat{\alpha}^T Z = \text{corr}(z_0, Z)^T R(Z)^{-1} (Z - L\hat{\mu}) + \hat{\mu}$ 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 $\hat{\mu}$. En cuanto a la varianza del error: $\text{Var}(\text{Error}) = \sigma^2 + \text{Var}(\hat{\mu})$. 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 $\{(z_1, x_1), \dots, (z_j, x_j), \dots, (z_n, x_n)\}$ 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 simultáneamente, 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_o coincide con z_1 , entonces w es la primera columna de C luego, $inv(C) w = (1, 0, \dots, 0)^T$, $L^T inv(C) w = 1$ y $w^T inv(C) w = \sigma^2$ entonces

$$\hat{\alpha} = (1, 0, \dots, 0)^T$$

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

En cuanto a la varianza del error de la ecu. 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 $x_1, \dots, x_j, \dots, x_k$ 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. 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 $\hat{\mu}$

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

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

de varianza conocida, ecu. 3.2

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

Luego, $\hat{\mu}$ se distribuye normal $N(\mu, Var(\mu))$ 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 \hat{z}_0 es insesgado y es una C. Lineal de Z

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

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

Luego, $z_0 - \hat{z}_0$ se distribuye normal $N(0, Var(Error))$ y es posible obtener intervalos de confianza de z_0 , a partir de la función inversa de la distribución acumulada de la Normal.

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 $A = \{(z_1, x_1), \dots, (z_j, x_j), \dots, (z_n, x_n)\}$, 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:

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

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 $E = (e_1, \dots, e_j, \dots, e_n)^T$. 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 $F : n \times p$:

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

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:

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

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

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

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

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

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 : $AZ = AF\beta + AE \quad con \quad Var(AE) = I$. Se

tiene así un modelo lineal clásico $Y = X\beta + \epsilon$ con $AZ = Y$ $AF = X$ $AE = \epsilon$ cuya solución clásica es:

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

o su equivalente:

$$\hat{\beta} = (F^T A^T A F)^{-1} (F^T A^T) A Z = (F^T \text{inv}(C) F)^{-1} F^T \text{inv}(C) Z$$

▲

4.4. Predicción de z_0 y varianza del error

La proposición que sigue muestra que el predictor óptimo tiene una estructura similar al de Kriging Ordinario; en la ecu. 4.6 el primer término tiene los mismos pesos $w^T \text{inv}(C)$ 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:

$$\hat{\alpha} = \text{inv}(C) (I - FhF^T \text{inv}(C)) w + \text{inv}(C) FhF_0 \quad (4.4)$$

con

$$h = (F^T \text{inv}(C) F)^{-1} \quad (4.5)$$

El predictor óptimo:

$$\hat{\alpha}^T Z = w^T \text{inv}(C) (Z - F\hat{\beta}) + F_0^T \hat{\beta} \quad (4.6)$$

donde $\hat{\beta}$ es la expresión ecu. 4.3 que representa la Tendencia en x_0 .

La varianza del error asociada:

$$\text{Var}(\text{Error}) = \sigma_0^2 - w^T \text{inv}(C) w + h \|F_0 - F^T \text{inv}(C) w\|^2 \quad (4.7)$$

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:

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

$$\hat{\alpha} = inv(C) (I - FhF^T inv(C)) w + inv(C) FhF_0$$

$$\hat{\lambda} = h (F^T inv(C) w - F_0) \quad (4.9)$$

De donde el predictor óptimo de z_0 es:

$$\hat{\alpha}^T Z = w^T inv(C) (I - FhF^T inv(C)) Z + F_0^T hF^T inv(C) Z$$

4.4. PREDICCIÓN DE Z_0 Y VARIANZA DEL ERROR

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

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

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

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

queda expresada como:

$$\text{Var}(\text{Error}) = \sigma_0^2 - w^T \text{inv}(C) w + (F_0^T - w^T \text{inv}(C) F) h (F_0 - F^T \text{inv}(C) w)$$

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 $\text{Var}(\text{Error}) \geq \sigma_0^2$.

Apéndice A

Álgebra lineal y vectores aleatorios

A.1. Vectores aleatorios

A.1.1. Definiciones

Sean $X \in R^p$ $Y \in R^q$ vectores aleatorios, es decir X un vector columna cuyos componentes son variables aleatorias, se define *valor esperado del vector* X al vector columna: $E(X) = [E(x_1), \dots, E(x_p)]^T$ y matriz de covarianza entre X y Y :

$$Cov(X, Y) = \begin{bmatrix} cov(x_1, y_1) & \dots & cov(x_1, y_q) \\ \dots & \dots & \dots \\ cov(x_p, y_1) & \dots & cov(x_p, y_q) \end{bmatrix} = E[(X - \mu_x)(Y - \mu_y)^T]$$

o su equivalente: $Cov(X, Y) = E[XY^T] - \mu_x \mu_y$

Nótese que $Cov(X, Y) = Cov(Y, X)^T$

Si $Y = X$ entonces se notará $Cov(X) = Cov(X, X)$.

A.1.2. Estimación

Si se dispone de una muestra de tamaño n de los vectores aleatorios X e Y donde la muestra de X está dada por la matriz $n \times p$, $Xobs$, y la muestra de Y dada por la matriz $n \times q$, $Yobs$, las medias muestrales m_x y m_y se obtienen promediando las columnas de las matrices:

$$m_x = \frac{1}{n} L_n^T Xobs \quad m_y = \frac{1}{n} L_n^T Yobs$$

donde se nota al vector de n unos así: $L_n^T = [1 \ 1 \ \dots \ 1]$.

En cuanto a la covarianza muestral está dada por:

$$Cov(X, Y) = \frac{1}{n} X_{obs}^T Y_{obs} - m_x m_y^T$$

A.1.3. Propiedades

Si A y B son matrices constantes: $Cov(AX, BY) = ACov(X, Y)B^T$
 si además $X = Y$

- $Cov(AX, BX) = ACov(X)B^T$
- $Cov(AX) = ACov(X)A^T$

Si $a \in R^p$ y $b \in R^q$ son vectores columnas constantes entonces:

- $V(a^T X) = a^T Cov(X)a \geq 0$
- $cov(a^T X, b^T Y) = a^T Cov(X, Y)b$
- Si $X = Y$ entonces $cov(a^T X, b^T X) = a^T Cov(X)b$

Si A, B son matrices $m \times p$ y $m \times q$ constantes entonces:

$$Cov(AX+BY) = ACov(X)A^T + BCov(Y)B^T + ACov(X, Y)B^T + BCov(Y, X)A^T$$

Un caso particular es el siguiente: si $a \in R^p$ y $b \in R^q$ son vectores columnas constantes entonces:

$$Cov(a^T X + b^T Y) = a^T Cov(X)a + b^T Cov(Y)b + 2a^T Cov(X, Y)b$$

Si u es una variable aleatoria y $c \in R^n$, entonces:

$$Cov(u c) = Var(u)c c^T$$

Esta expresión se aplica a las proyecciones de un vector aleatorio $X \in R^p$ sobre la dirección de un vector c , fijo. Si se supone c unitario $\|c\| = 1$ la proyección está dada por $Px = (c^T X)c$ entonces:

$$Cov(Px) = c c^T Var(c^T X) = c c^T (c^T Cov(X)c)$$

nótese que el último paréntesis es un número real.

Si c es vector propio de $Cov(X)$ de valor propio λ , entonces: $Cov(Px) = c c^T \lambda c^T c$.

Y como $\|c\| = 1$ entonces $Cov(Px) = c c^T \lambda$.

A.2. Algunas fórmulas de derivación

A.2.1. Definición de gradiente

Si $x \in R^n$ $z \in R^q$ se define como gradiente del vector $z(x)$, ∇z la matriz $n \times q$ donde la columna k está formada por el vector gradiente $\nabla z_k(x)$ del k componente de z .

Regla de la cadena Si $y \in R^p$ es función de z , es decir $y(z)$ entonces el gradiente de la función compuesta $y(z(x))$, respecto de x , está dado por:

$$\nabla_x y = \nabla_x z \times \nabla_z y$$

nótese que el producto es conforme: $n \times p = (n \times q)(q \times p)$.

Ejemplo: Si la matriz A ($m \times q$) es constante entonces $\nabla Ax = A^T$; luego, si $y(x) \in R^q$ es función de x , por la regla de la cadena:

$$\nabla Ay(x) = \nabla_x y(x) A^T.$$

A.2.2. propiedades:

Si $y(x) \in R^q$ entonces $\nabla(z^T y) = \nabla z^T y + \nabla y^T z$

A.2.2.1. Corolarios

- si $z = y$ entonces $\nabla(\|y\|^2) = 2\nabla y^T y$
- $\nabla(\|x\|^2) = 2x$
- si $z = \text{constante}$ entonces $\nabla(z^T y) = \nabla y^T z$
- $\nabla(z^T x) = z$
- $\nabla \|Ay\|^2 = 2\nabla y A^T Ay$
- $\nabla \|Ax\|^2 = 2A^T Ax$
- $\text{Hessiano}(\|Ax\|^2) = 2A^T A$
- $\nabla(x^T Ax) = (A^T + A)x$, si A es simétrica: $\nabla(x^T Ax) = 2Ax$
- $\text{Hessiano}(x^T Ax) = A^T + A$, si A es simétrica: $\text{Hessiano}(x^T Ax) = 2A$

A.3. Solución de un sistema frecuente en Kriging

Sea el sistema cuadrado ya particionado de $n + m$ ecuaciones, $n > m$, de solución $\begin{pmatrix} \alpha \\ \lambda \end{pmatrix}$:

$$\begin{array}{l} C \quad F \quad = \quad e \\ F^T \quad 0 \quad = \quad g \end{array}$$

A.3. SOLUCIÓN DE UN SISTEMA FRECUENTE EN KRIGING

con C ($n * n$) definida positiva, F $n * m$ de rango m .

Este sistema tiene solución única, la matriz es invertible; si no lo fuera existiría una solución del sistema homogéneo:

$$\begin{aligned} C\alpha + F\lambda &= 0 \\ F^T\alpha &= 0 \end{aligned}$$

con al menos uno de los dos, $\alpha \neq 0$ o $\lambda \neq 0$. Entonces por la primera ecuación si $\alpha = 0 \Rightarrow F\lambda = 0$ pero como F es de rango m , $\lambda = 0$.

Por otra parte si $\alpha \neq 0$, multiplicando la primera por α^T queda $\alpha^T C\alpha + \alpha^T F\lambda = 0$, pero por la segunda $\alpha^T F\lambda = 0$; luego, $\alpha^T C\alpha = 0$ que no puede ser ya que C es definida positiva.

Habiéndose probado la existencia, resolvamos el sistema. Multiplicando la primera por C^{-1} queda:

$$\begin{aligned} I \quad C^{-1}F &= C^{-1}e \\ F^T \quad 0 &= g \end{aligned}$$

Multiplicando la primera por $-F^T$ y sumando a la segunda queda:

$$\begin{aligned} I \quad C^{-1}F &= C^{-1}e \\ 0 \quad -F^T C^{-1}F &= g - F^T C^{-1}e \end{aligned}$$

Luego, la solución del sistema está dada por

$$\lambda = \text{inv}(F^T C^{-1}F) (F^T C^{-1}e - g)$$

y

$$\alpha = C^{-1}e - C^{-1}F \lambda$$

$$\left\{ \begin{array}{l} \alpha = q \\ q = q \\ q = q \end{array} \right\}$$

Si $F = L$ entonces $\lambda = \frac{L^T C^{-1}e - g}{L^T C^{-1}L}$ es un escalar. En cuanto a $\alpha = C^{-1}e - C^{-1}\lambda L$.