Curso intensivo de Krigeado

- Es un método de Interpolación
- Lo hemos citado y lo citaremos en:
  - Imputación de ausencias (obvio…)
  - Detección de errores
  - Estimación de sensibilidad de modelos
- Base estadística
- Incorporado en algunos GIS (¿malamente? ¿parcialmente?...)

¿Qué es la Geoestadística?

- Def.: *Aplicación de la teoría de las variables regionalizadas a la estimación de procesos en el espacio*
- Si z(x) es el valor de z en el punto x, z(x) es una variable regionalizada
  - Concepto no probabilístico
  - Quizá función continua
- Usualmente z(x) está compuesta de
  - Componentes aleatorios y
  - Componentes estructurados
  - No luce "suave"

➔Conviene considerar a z(x) como una función aleatoria

Algunas consecuencias…

- La *realidad* es simplemente una realización o instancia de un experimento aleatorio
- Sólo tenemos <u>una</u> realidad; hay que hacer inferencia estadística sólo con ello
  - En general no sería posible
  - Requerirá hipótesis adicionales
  - Ej.: homogeneidad espacial
- Las funciones aleatorias son sólo un modelo posible de la realidad

Definiciones…

Momentos de la distribución

- 1er. orden: Esperanza E(Z(x))=m(x)
  - m(x) es llamada "deriva" o "tendencia"
- 2do. orden:
  - Varianza Var(Z(x))=E([Z(x)-m(x)]^2)
  - Covarianza C(x_i,x_j)

    C(x_i,x_j)= E([Z(x_i)-m(x_i)].[Z(x_j)-m(x_j)])
  - Semivariograma γ(x_i,x_j)

    γ(x_i,x_j)= 0.5*E([Z(x_i)-Z(x_j)]^2)
- Var(Z(x))≥0; γ(x_i,x_j)≥0 pero C(x_i,x_j) no se sabe

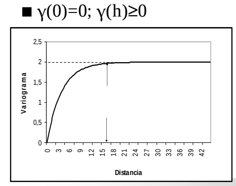
Más definiciones

- Def.: Z(x) estacionaria de segundo orden si
  - E(Z(x)) existe y no depende de x
  - C(x+h,x)=C(h) (sólo depende de la separación)
- Implica
  - Var(Z(x))=C(0)
  - γ(x+h,x)= γ(h)=0.5*E([Z(x+h)-Z(x)]^2)
- h es en general un vector; suele asumirse isotropía, por lo que γ(h)= γ(| h |)
- γ(h)=var(Z)-C(h) sólo si la media es estrictamente constante; en otro caso, usar γ(h) es más conveniente que usar C(h)

Más sobre variogramas…

Def.: γ(h)=0.5*E([Z(x+h)-Z(x)]^2)

- γ(0)=0; γ(h)≥0



Rango ***(Range):***

Distancia a la cual el variograma se estabiliza

Meseta ***(Sill)*** **:**

Valor constante que toma el variograma en distancias mayores al rango

Fórmula del Variograma

- El variograma debe cumplir algunas condiciones matemáticas restrictivas
- Salen de imponer que Var(Y)≥0, siendo

  Y=Σ λ_iZ(x_i), λ_i y x_i conjunto arbitrario
- Hay algunos *modelos* de variogramas que se ajustan a los datos
  - Esférico, Exponencial, Gaussiano, Pepita, etc.
  - Hay otros menos populares
- Todos dependen de la meseta **S** y el rango **a**, excepto el denominado Pepita *(nugget)*

Estimación del Variograma

- Un tópico en sí mismo
- "*Left to the user…*"
- Métodos:
  - A sentimiento (!)
  - Mínimos cuadrados
  - Jacknife
  - Máxima Verosimilitud
  - Validación Cruzada
  - Validación Cruzada de Máxima Verosimilitud
  - …
- Sin variograma…

Krigeado

- Del geólogo sudafricano D. G. Krige
- Hay muchas variantes y casos particulares
- Caso Puntual: se modela el estimador con

$$Z^*(x)=\sum_{i=1}^{N}\lambda_i(x)Z(x_i)$$

eligiéndose los pesos λ_i(x) para que sea insesgado

$$E\left(Z^*(x)\right)=m=E\left(Z(x)\right)$$

y de varianza mínima

$$\operatorname{var}\left[Z(x)-Z^*(x)\right]$$

Algunos detalles

- Se asume m constante; hay variantes para otro caso
- Los pesos son función del punto
- Salen del sistema:

$$\begin{bmatrix} 0 & \gamma_{12} & \gamma_{13} & \cdots & \gamma_{1n} & 1 \\ \gamma_{21} & 0 & \gamma_{23} & \cdots & \gamma_{2n} & 1 \\ \gamma_{31} & \gamma_{32} & 0 & & \vdots & \vdots \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \gamma_{n1} & \gamma_{n2} & \gamma_{n3} & \cdots & 0 & 1 \\ 1 & 1 & 1 & \cdots & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \vdots \\ \lambda_n \\ \mu \end{bmatrix} = \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \vdots \\ \gamma_n \\ 1 \end{bmatrix}$$

Algunos detalles(2)

- Donde $\gamma_{ij} = \gamma(x_i - x_j); \gamma_i = \gamma(x_i - x)$
- Nótese que:
  - ➤ El variograma depende de los datos
  - ➤ Los coeficientes λ_i dependen del variograma, pero no de los datos mismos
  - ➤ Idem con la varianza, mediante la expresión

$$Var(Z^* - Z) = \sum_i \lambda_i \gamma(x_i - x) + \mu$$

  - ➤ La matriz del sistema es constante; puede usarse LU
- El resultado es perfectamente determinista; lo *estocástico* reside en los datos mismos

Algunos detalles(3)

- El de Krigeado es un estimador BLUE
  - ➤ Sólo si el variograma es "exacto"
  - ➤ Sólo si la función aleatoria es normal
    - – En ese caso, es el *Best* incluso comparando con los no lineales
    - – Difícil de verificar la normalidad en la práctica (por lo multivariado…)
- El Krigeado es interpolante
  - ➤ Sólo si se asumen datos sin error
- Bajo ciertas hipótesis
  - ➤ error $\sim N^{(-2/d)}$; N nro. de puntos y d dimensión del espacio (típicamente 2)
  - ➤ ¡Incluso con el variograma erróneo!
    - – Pero en este caso la varianza no es consistente

Algunos detalles(3)

- Si los datos *tienen un error* cuya varianza es ε^2 el sistema cambia levemente

$$\begin{bmatrix} \varepsilon^2 & \gamma_{12} & \gamma_{13} & \cdots & \gamma_{1n} & 1 \\ \gamma_{21} & \varepsilon^2 & \gamma_{23} & \cdots & \gamma_{2n} & 1 \\ \gamma_{31} & \gamma_{32} & \varepsilon^2 & & \vdots & \vdots \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \gamma_{n1} & \gamma_{n2} & \gamma_{n3} & \cdots & \varepsilon^2 & 1 \\ 1 & 1 & 1 & \cdots & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \vdots \\ \lambda_n \\ \mu \end{bmatrix} = \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \vdots \\ \gamma_n \\ 1 \end{bmatrix}$$

Simulación

- Def.:
  - ➤ No condicionada: Consiste en generar realizaciones con igual media y varianza que la disponible
  - ➤ Condicionada: Idem, pero obligando a que además adopte valores específicos en ciertos puntos
- Tres tipos de métodos
  - ➤ Espectrales, Bandas Rotantes y Matricial
  - ➤ Sólo presentaremos el Matricial

Método Matricial de Simulación

- No es el más eficaz si se necesitan muchos puntos
  - ➤ Matricial $O(n^3)$
  - ➤ Bandas rotantes $O(n^{1/2})$
- Implícitamente se asume normalidad
- La fórmula para Z_S es:

$$Z_S = Z^* + \mathbf{M.u}; \mathbf{M.M}^T = \mathbf{C}; \mathbf{u_i} \sim N(0,1)$$

- La simulación se logra generando diversos **u**
  - ➤ Problema estándar
  - ➤ Muchas librerías disponibles

¿Para qué se usa la Simulación?

- Generar realizaciones
  - ➤ Compatibles con las medidas disponibles
  - ➤ Compatibles con el variograma asumido
- Ejemplo: un MDE
  - ➤ Generar N realizaciones del raster buscado
  - ➤ Delinear zona de visibilidad a un mástil
  - ➤ Calcular área A_i de esa zona
  - ➤ Calcular valor esperado, promedio, máximos, etc. del conjunto A_i y sus niveles de confianza
- Se comentarán más casos luego

Literatura & Software

- Digital:
  - ➤ Rudolf Dutter, Vienna Inst. of Technology
    CD del curso; http://www.statistik.tuwien.ac.at/public/dutt/vorles/geost_03/geo.html
  - ➤ Denis Marcotte, École Polytechnique de Montréal
    CD del curso; http://geo.polymtl.ca/~marcotte/glq3401geo.html
  - ➤ Oscar Rondón, Venezuela
    CD del curso
- Papel:
  - ➤ Samper, F.J. y Carrera, J. 1990. Geoestadística: Aplicaciones a la hidrología subterránea, CIMNE, ISBN 84-404-6045-7
- Biblioteca GSLIB
- Matlab+EasyKrig
  ftp://globec.whoi.edu/pub/software/kriging/V2.1/easy_krig2.1