Estadística Espacial

Kenneth Roy Cabrera Torres

17 de marzo de 2015

Tabla de contenido

Introducción a la Estadística Espacial

Definición

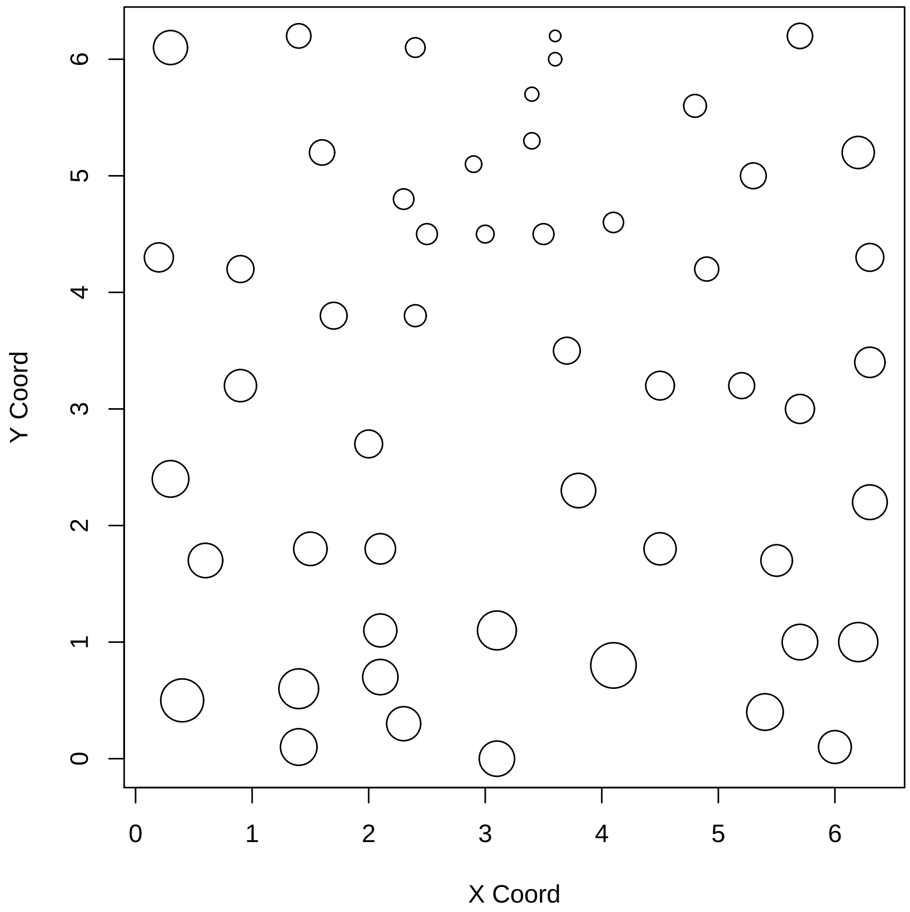
El término *estadística espacial* se usa para describir una amplia gama de modelos estadísticos que procuran analizar datos espacialmente referenciados o georreferenciados.

La *geostadística* se refiere a modelos y métodos de datos que siguen las siguientes características:

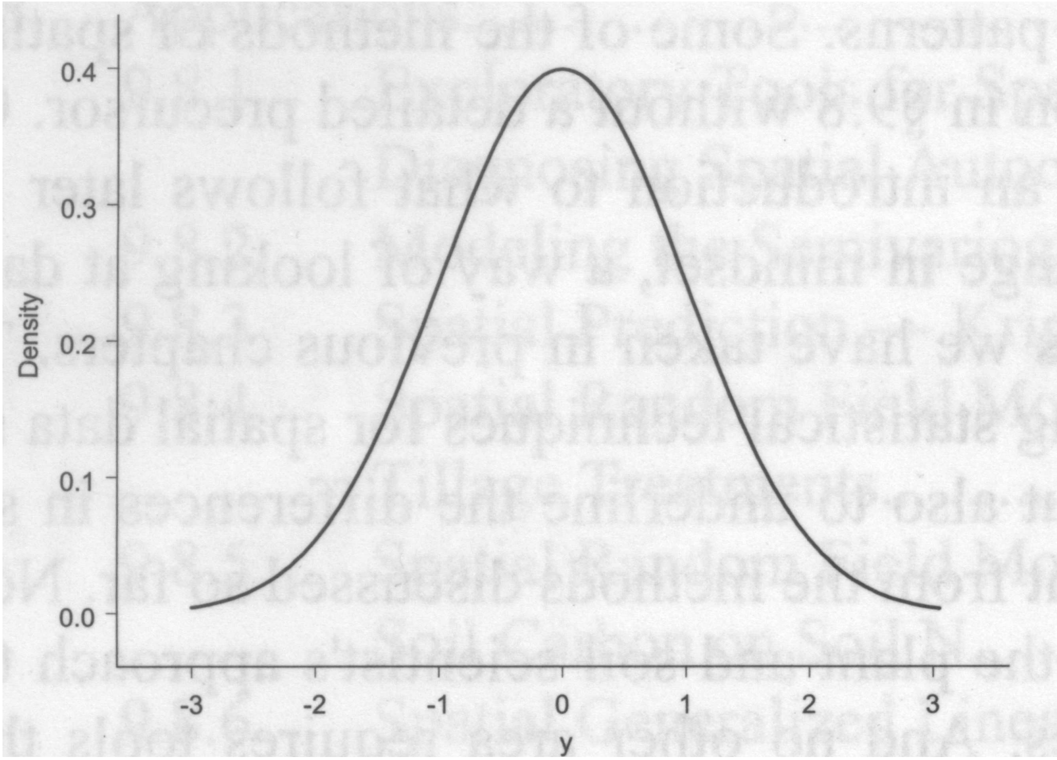
Primero, los valores de Y_i: $i = 1, \ldots, n$ son observados en una conjunto discreto de lugares x_i al interior de una región espacial A.

Segundo, cada valor observado de Y_i es ya sea una medida directa, o estadística relacionada con, el valor de un fenómeno espacial subyacente, $S(x)$ en los correspondientes sitios de muestreo x_i.

Datos de elevación

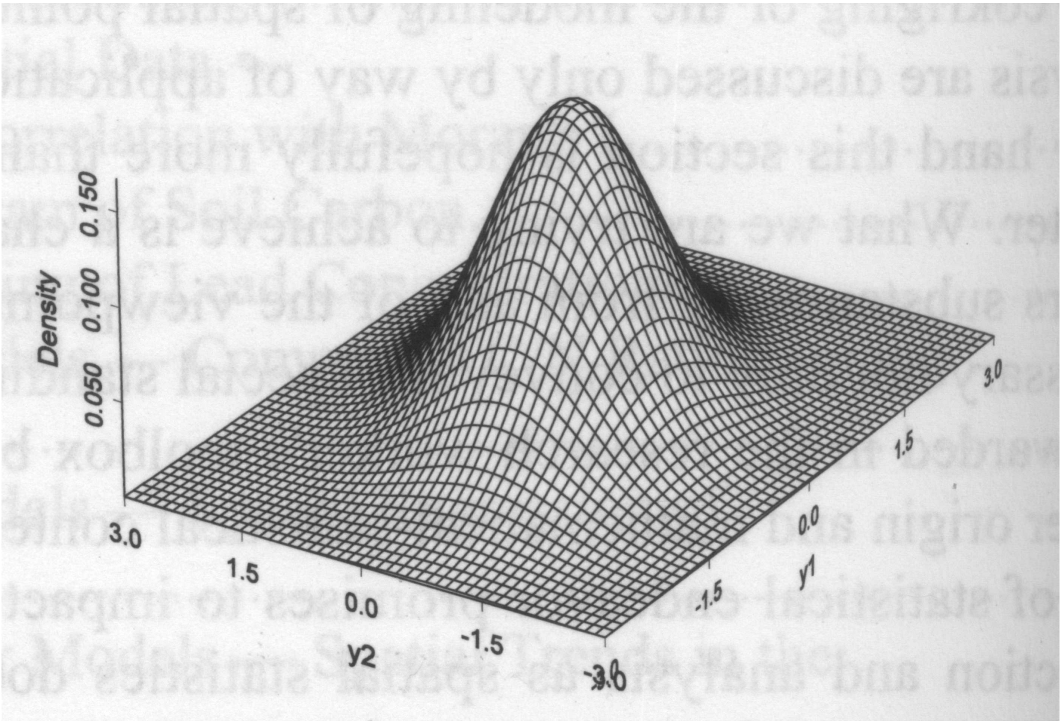


Distribución univariada



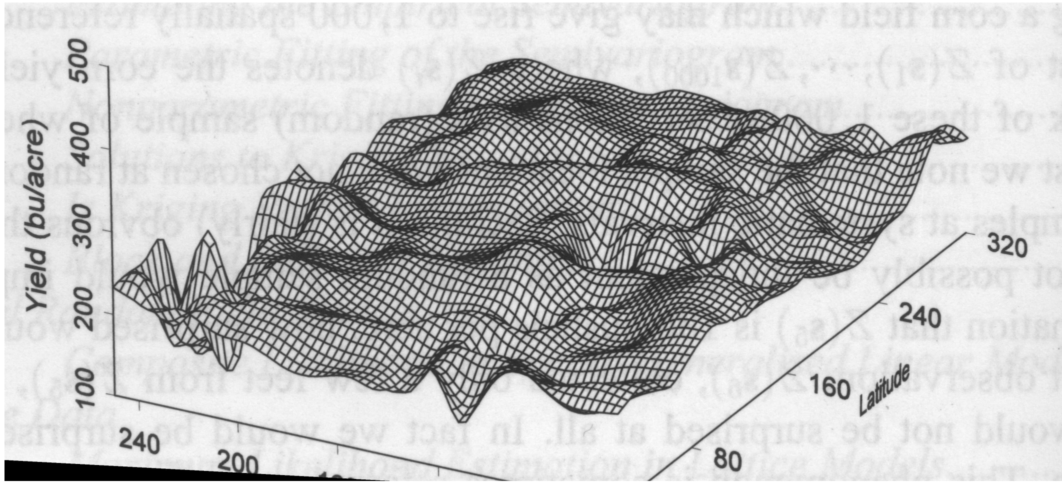
$$X \sim \mathcal{N}(0,1)$$

Distribución bivariada



$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0{,}3 \\ 0{,}3 & 1 \end{bmatrix}\right)$$

Un campo aleatorio



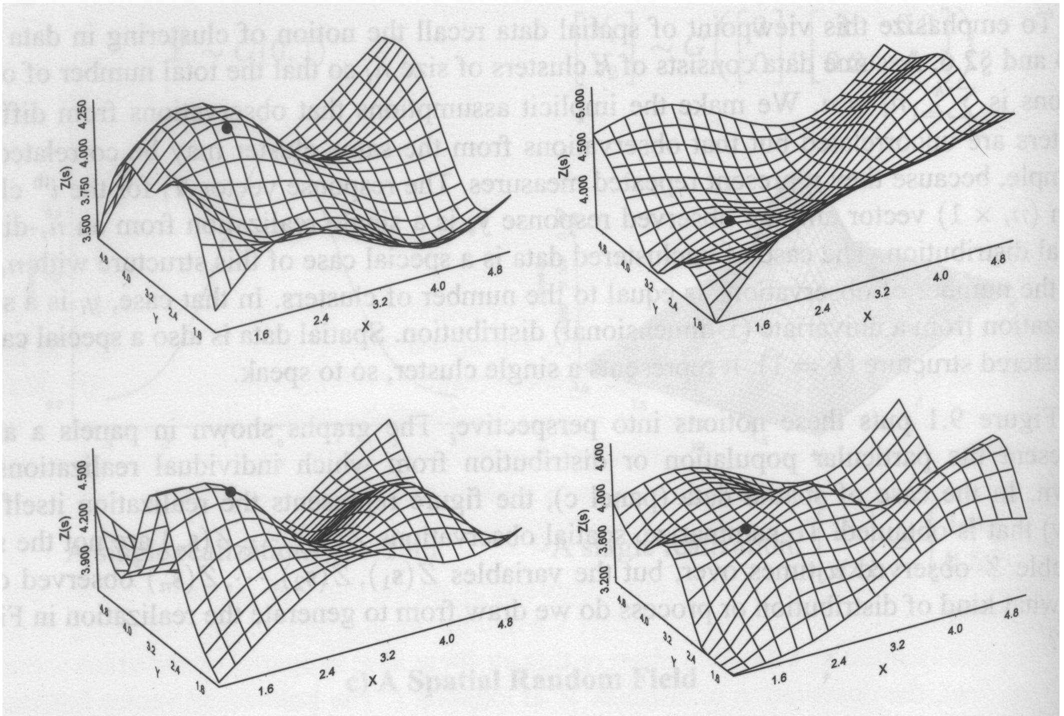
$$\{S(x) : x \in A \subset \mathcal{R}^2\}$$

Una sola realización

Campos aleatorios

- Un conjunto de datos espaciales se consideran una realización de un experimento aleatorio. Sólo se obtiene una realización de $S(x)$ en el punto x. Esta es una realización de un campo aleatorio, es decir un proceso estocástico.
- $S(x_0)$ es una variable aleatoria si se considera la distribución de todas las realizaciones posibles en el punto x_0.
- Cuando se muestrea un campo aleatorio, estas muestras se toman de una realización particular de un experimento aleatorio.

Realizaciones de un campo aleatorio



Cuatro realizaciones de un campo aleatorio.

Terminología y notación

Para efectos prácticos los datos geoestadísticos univariados se puede tomar las siguiente notación:

$$(x_i, y_i) : i = 1, \ldots, n.$$

Donde x_i denota la localización espacial (generalmente un espacio bidimensional) y y_i es un valor escalar asociado a la localización en x_i. Generalmente a y se le conoce como una variable de respuesta o variable medida.
Cada y_i es una realización de la variable aleatorio Y_i cuya distribución depende de la localización de x_i de un proceso estocástico continuo espacial subyacente $S(x)$ que no es directamente observable.
En muchos casos se puede asumir que $Y_i = S(x_i)$, pero en general no son iguales.

Conceptos y supuestos

Modelo Geoestadístico

El *Modelo Geostadístico* incorpora al menos dos elementos:

1. Un proceso estocástico real $\{S(x) : x \in A\}$ que se considera una realización parcial del proceso estocástico $\{S(x) : x \in \mathcal{R}^2\}$ en todo el plano.
2. Una distribución multivariada de la variable aleatoria $Y = (Y_1, \ldots, Y_n)$ condicionado en $S(\cdot)$.

Se suele denominar a $S(x)$ la *señal* y a Y_i la *respuesta*. A menudo Y_i se puede pensar como una versión ruidosa de $S(x_i)$ y la Y_i se puede asumir condicionalmente independiente dado $S(\cdot)$.

Supuestos Modelo Geoestadístico

Los supuestos básico del modelo clásico son:

1. $\{S(x) : x \in \mathcal{R}^2\}$ es un proceso gaussiano con media μ, y varianza $\sigma^2 = Var\{S(x)\}$ y función de correlación $\rho(u) = Corr\{S(x), S(x')\}$, donde $u = ||x - x'||$ y $||\cdot||$ denota la distancia;
2. Condicionado en $\{S(x) : x \in \mathcal{R}^2\}$, las y_i son realizaciones de variables mutuamente independientes Y_i, distribuidas normalmente con media $E[Y_i|S(\cdot)] = S(x_i)$ y varianza condicional τ^2.

Una forma equivalente es:

$$Y_i = S(x_i) + \varepsilon_i : i = 1, \ldots, n$$

donde $\{S(x) : x \in \mathcal{R}^2\}$ se define como el supuesto 1 y ε_i son variables aleatorias mutuamente independientes $\mathcal{N}(0, \tau^2)$.

Función de correlación Matérn

Una función muy común de correlacion definida en geoestadística es la función Matérn definida como:

$$\rho(u;\phi,\kappa) = \{2^{\kappa-1}\Gamma(k)\}^{-1}(u/\phi)^{\kappa}K_{\kappa}(u/\phi)$$

Donde $K_{\kappa}(\cdot)$ es una función de Bessel modificada de segunda clase, de orden κ. El parámetro $\phi > 0$ determina la tasa a la cual decae a cero la correlación en la medida que decrece u. El parámetro $\kappa > 0$ es el orden del modelo Matérn y está asociado a la forma del descenso de la función.

Funciones asociada a la función Matérn

Siguiendo la función Matérn se tienen dos situaciones muy comunes de funciones de correlación

- Función de correlación exponencial:
  Si $\kappa = \frac{1}{2}$ entonces se tiene que la función queda reducida a:

$$\rho(u;\phi) = e^{-\frac{u}{\phi}}$$

- Función de correlación gausiana:
  Si $\kappa \to \infty$ entonces la función Matérn tiene como límite:

$$\rho(u;\phi) = e^{-\left(\frac{u}{\phi}\right)^2}$$

Rangos prácticos

El rango práctico se define como el valor de u en donde la función $\rho(u)$ tiene un valor de 0.05. O en otras palabras es el valor de u a partir del cual se puede considerar que $\rho(u) \approx 0$.

- Para la función exponencial se tiene el rango práctico es: $\approx 3\phi$.
- Para la función gausiana se tiene que el rango práctico es: $\approx \sqrt{3}\phi$.

Rango práctico para función Matérn

Para la función Matérn a un rango práctico de $\rho_0 = 0{,}05$, se debe hallar la raiz de la siguiente ecuación:

$$\rho(u; \phi, \kappa) - \rho_0 = 0$$
$$\rho(u; \phi, \kappa) - 0{,}05 = 0$$

Ya que ϕ es un parámetro de escala, la solución será de la forma $r\phi$ donde r es el factor multiplicador que se haya de la solución de la ecuación:

$$\rho(u; \phi, 1) - 0{,}05 = 0$$

Funciones exponencial y gausiana

Otra manera de expresar las funciones de correlación exponencial y gausiana es:

- Exponencial:

$$\rho(u;\alpha) = e^{-3\frac{u}{\alpha}}$$

- Gausiana:

$$\rho(u;\alpha) = e^{-3\left(\frac{u}{\alpha}\right)^2}$$

En donde α es el rango práctico.

Función potencia

Se define como:

$$\rho(u; \phi, \kappa) = e^{-\left(\frac{u}{\phi}\right)^{\kappa}}$$

Donde el parámetro de escala $\phi > 0$ y el de forma κ está limitado por $0 < \kappa \leq 2$. Note que si $\kappa = 2$ es la misma gausiana.

Función esférica

Esta función se define como:

$$\rho(u;\phi) = \begin{cases} 1 - \frac{3}{2}(\frac{u}{\phi}) + \frac{1}{2}\left(\frac{u}{\phi}\right)^3 & : \quad 0 \leq u \leq \phi \\ 0 & : \quad u \geq 0 \end{cases}$$

En este caso el rango absoluto es ϕ dado que si $u \geq \phi$ se tiene que $\rho(u) = 0$, por eso no tendría rango práctico.

Semivariograma empírico

Para un conjunto de datos $(x_i, y_i) : i = 1, \ldots, n$, el *las ordenadas del variograma empírico* es el valor $v_{ij} = \frac{1}{2}(y_i - y_j)^2$. Algunos autores lo prefieren denominar *ordenadas del semivariograma*.

Si y_i es estacionario en media y varianza. El valor esperado de v_{ij} es $\sigma^2\{1 - \rho(x_i, x_j)\}$ donde σ^2 es la varianza y $\rho(x_i, x_j)$ es la correlación entre y_i y y_j.

Si y_i es un proceso estacionario, entonces $\rho(\cdot)$ depende sólo de la distancia entre x_i y x_j, y además tiende a cero a grandes distancias, entonces v_{ij} tiende a σ^2 cuando la distancia entre $u_{ij} = ||x_i - x_j||$ tiende a infinito.

Semivariograma teórico

El semivariograma de un espacio estocástico $S(x)$ es la función:

$$V(x,x') = \frac{1}{2}\text{Var}\{S(x) - S(x')\}$$

Note que $V(x,x') = \frac{1}{2}\left[\text{Var}\{S(x)\} + \text{Var}\{S(x')\} - 2\text{Cov}\{S(x), S(x')\}\right]$. En el caso estacionario, simplifica a $V(u) = \sigma^2\{1-\rho(u)\}$.

Recordemos que:

$$\rho(x,x') = \frac{\text{Cov}(S(x)S(x'))}{\sqrt{\text{Var}\{S(x)\}}\sqrt{\text{Var}\{S(x')\}}}$$

Gráfica de semivariograma exponencial