Tarea 2 - MAT467

ESTUDIANTE: CLEMENTE FERRER

1) Demuestre que si $C_j(\mathbf{h})$ son funciones de covarianza, j = 1, ..., k, entonces $\prod_{j=1}^k C_j(\mathbf{h})$ es una función de covarianza válida.

Solución: Definamos $C(h) = \prod_{j=1}^k C_j(h)$ y mostremos que es una función de covarianza válida. Es decir, debemos verificar la paridad y la condición de ser semidefinida positiva.

Procederemos por inducción en j. El caso base debemos verificarlo para j = 2, es decir, C(h) = C₁(h)C₂(h). Para ello, consideremos a Σ la matriz cuyas entradas son C(s_i - s_j), i, j = 1,...,n. Como C es un producto, la matriz de covarianza Σ puede ser escrita como un producto de Kronecker Σ = C₁⊗C₂, donde C₁, C₂ ∈ M_{n×n}(ℝ). En particular, los elementos de C₁ y C₂ son de la forma C_{1,2}(s_i - s_j), i, j = 1,...,n. Como C₁ y C₂ son funciones de covarianza válidas, sus respectivas matrices de covarianza poseen valores propios λ₁,...,λ_n y μ₁,...,μ_n respectivamente, los cuales son todos positivos por ser semidefinidas positivas. Por lo anterior el producto entre ambas es semidefinido positivo, es decir, Σ lo es y por ende C(h) es semidefinida positiva como se quería probar.

Asumamos cierto ahora para k y probemos para k+1, esto es

$$C(oldsymbol{h}) = \prod_{j=1}^{k+1} C_j(oldsymbol{h}) = \prod_{j=1}^k C_j(oldsymbol{h}) \cdot C_k(oldsymbol{h}).$$

Por hipótesis inductiva $\prod_{j=1}^k C_j(\boldsymbol{h})$ es semidefinida positiva. Luego, renombrando $C_1(\boldsymbol{h}) := \prod_{j=1}^k C_j(\boldsymbol{h})$ y $C_2(\boldsymbol{h}) := C_k(\boldsymbol{h})$ notamos que todo se resume a probar que el producto de dos funciones semidefinidas positivas es semidefinida positiva, pero esto fue probado en el caso base de la inducción. Por lo tanto concluimos que

$$C(\boldsymbol{h}) = \prod_{j=1}^k C_j(\boldsymbol{h}) \cdot C_k(\boldsymbol{h})$$

es semidefinida positiva.

• Por otro lado, cada C_j es una función de covarianza y por ende $C_j(\mathbf{h}) = C_j(-\mathbf{h})$, para todo $j = 1, \ldots, k$. Luego

$$C(\mathbf{h}) = \prod_{j=1}^{k} C_j(\mathbf{h}) = \prod_{j=1}^{k} C_j(-\mathbf{h}) = C(-\mathbf{h}),$$

lo que comprueba la paridad de C(h).

Por lo anterior, concluimos que C(h) es una función de covarianza válida.

- 2) Para los variogramas Powered Exponential y Wave,
 - a) Identifique el efecto nugget, el sill y el rango.

Solución:

• Powered Exponential: El variograma para este modelo está dado por

$$\gamma(t) = \begin{cases} \tau^2 + \sigma^2 \left(1 - \exp\left(-|\phi t|^p\right) \right) & \text{si } t > 0, \\ 0 & \text{en otro caso} \end{cases}$$

Luego, los efecto nugget, sill y rango son respectivamente:

Nugget:
$$\lim_{t \to 0^+} \gamma(t) = \lim_{t \to 0^+} \tau^2 + \sigma^2 (1 - \exp(-|\phi t|^p))$$

= $\tau^2 + \sigma^2 \lim_{t \to 0^+} (1 - \exp(-|\phi t|^p))$
= τ^2 .

Sill:
$$\lim_{t \to \infty} \gamma(t) = \lim_{t \to \infty} \tau^2 + \sigma^2 \left(1 - \exp\left(-|\phi t|^p\right) \right)$$
$$= \tau^2 + \sigma^2 \lim_{t \to \infty} \left(1 - \exp\left(-|\phi t|^p\right) \right)$$
$$= \tau^2 + \sigma^2$$

Range:
$$\min\{t: \gamma(t) = \text{Sill}\} = \infty$$

• Wave: El variograma para este modelo está dado por

$$\gamma(t) = \begin{cases} \tau^2 + \sigma^2 \left(1 - \frac{\sin(\phi t)}{\phi t} \right) & \text{if } t > 0 \\ 0 & \text{otherwise} \end{cases}$$

Luego, los efecto nugget, sill y rango son respectivamente:

Nugget:
$$\lim_{t \to 0^+} \gamma(t) = \lim_{t \to 0^+} \tau^2 + \sigma^2 \left(1 - \frac{\sin(\phi t)}{\phi t} \right)$$
$$= \tau^2 + \sigma^2 \lim_{t \to 0^+} \left(1 - \frac{\sin(\phi t)}{\phi t} \right)$$
$$= \tau^2$$

Sill:
$$\lim_{t \to \infty} \gamma(t) = \lim_{t \to \infty} \tau^2 + \sigma^2 \left(1 - \frac{\sin(\phi t)}{\phi t} \right)$$
$$= \tau^2 + \sigma^2 \lim_{t \to \infty} \left(1 - \frac{\sin(\phi t)}{\phi t} \right)$$
$$= \tau^2 + \sigma^2$$

Range:
$$\min\{t: \gamma(t) = \text{Sill}\} = \frac{\pi}{\phi}$$

b) Encuentre la función de covarianza C(t) correspondiente a los variogramas usados en la parte a.

Solución: Por proposición vista en clases, la función de covarianza puede ser recuperada a partir del variograma mediante la relación

$$C(\boldsymbol{h}) = \lim_{\|\boldsymbol{u}\| \to \infty} [\gamma(\boldsymbol{u}) - \gamma(\boldsymbol{h})]$$

Además, $C(0) = C(\mathbf{h}) + \gamma(\mathbf{h})$. Ahora, procedamos a obtener las funciones de covarianzas pedidas.

• Powered Exponential: Sea t > 0 con t = ||h||

$$\begin{split} C(t) &= \lim_{\|\boldsymbol{u}\| \to \infty} [\gamma(\boldsymbol{u}) - \gamma(t)] \\ &= \lim_{\|\boldsymbol{u}\| \to \infty} \gamma(\boldsymbol{u}) - [\tau^2 + \sigma^2 \left(1 - \exp\left(-|\phi t|^p\right)\right)] \\ &= \tau^2 + \sigma^2 - [\tau^2 + \sigma^2 \left(1 - \exp\left(-|\phi t|^p\right)\right)] \\ &= \sigma^2 \exp\left(-|\phi t|^p\right) \end{split}$$

Luego,

$$C(0) = C(t) + \gamma(t) = \sigma^2 \exp(-|\phi t|^p) + \tau^2 + \sigma^2 (1 - \exp(-|\phi t|^p)) = \tau^2 + \sigma^2$$

Así, concluimos que

$$C(t) = \begin{cases} \sigma^2 \exp\left(-|\phi t|^p\right) & \text{si } t > 0, \\ \tau^2 + \sigma^2 & \text{en otro caso.} \end{cases}$$

• Wave: Sea t > 0 con $t = ||\boldsymbol{h}||$

$$\begin{split} C(t) &= \lim_{\|\boldsymbol{u}\| \to \infty} [\gamma(\boldsymbol{u}) - \gamma(t)] \\ &= \lim_{\|\boldsymbol{u}\| \to \infty} \gamma(\boldsymbol{u}) - \tau^2 - \sigma^2 \left(1 - \frac{\sin(\phi t)}{\phi t} \right) \\ &= \tau^2 + \sigma^2 - \tau^2 - \sigma^2 \left(1 - \frac{\sin(\phi t)}{\phi t} \right) \\ &= \sigma^2 \frac{\sin(\phi t)}{\phi t} \end{split}$$

Luego,

$$C(0) = C(t) + \gamma(t) = \sigma^2 \frac{\sin(\phi t)}{\phi t} + \tau^2 + \sigma^2 \left(1 - \frac{\sin(\phi t)}{\phi t}\right) = \tau^2 + \sigma^2$$

Así, concluimos que

$$C(t) = \begin{cases} \sigma^2 \frac{\sin(\phi t)}{\phi t} & \text{si } t > 0, \\ \tau^2 + \sigma^2 & \text{en otro caso.} \end{cases}$$

3) El estimador de momentos de Matheron del semivariograma puede ser escrito como una forma cuadrática en términos de los datos observados.

$$\widehat{\gamma}(\boldsymbol{h}) = \frac{1}{2} X(\boldsymbol{s})^T \boldsymbol{A}(\boldsymbol{h}) X(\boldsymbol{s}).$$

La matriz A es llamada matriz de diseño espacial.

a) Imagine n puntos regularmente espaciados en \mathbb{R} . Describa la forma de la matriz A(h) en este caso.

Solución: Consideraremos t = ||h|| y d = 1 para considerar n puntos regularmente espaciados en \mathbb{R} . Ahora, a partir de la estimación del semivariograma basado en los momentos

$$\widehat{\gamma}(t) = \frac{1}{2|N(t)|} \sum_{N(t)} \left(X\left(s_i\right) - X\left(s_j\right) \right)^2,$$

donde $N(t) = \{(s_i, s_j) : ||s_i - s_j|| = t, 1 \le i, j \le n\}$. Definiendo

$$X(\mathbf{s}) = (Y(\mathbf{s}_1), \dots, Y(\mathbf{s}_n))^T$$

tenemos que

$$A(\boldsymbol{h}) = \frac{1}{n - \|\boldsymbol{h}\|} \begin{pmatrix} \boldsymbol{I}_{n-\|\boldsymbol{h}\|} & -\boldsymbol{I}_{n-\|\boldsymbol{h}\|} \\ -\boldsymbol{I}_{n-\|\boldsymbol{h}\|} & \boldsymbol{I}_{n-\|\boldsymbol{h}\|} \end{pmatrix}$$

b) De la forma explícita de la matriz A(h) cuando los datos están dispuestos en una grilla regular de tamaño 3×3 en \mathbb{Z}^2 .

Solución: En este caso n=9 y ||h|| puede tomar los valores 1, 2, $\sqrt{2}$, $2\sqrt{2}$ y $\sqrt{5}$. Luego, la forma explícita de $A(\mathbf{h})$ es

$$A(1) = \frac{1}{12} \begin{pmatrix} 2 & -2 & 0 & -2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & -2 & 0 & -2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & -2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & -2 & 0 & -2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 4 & -2 & 0 & -2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 3 & 0 & 0 & -2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2 & -2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3 & -2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 \end{pmatrix}$$

4) Un campo aleatorio Gamma puede ser construido como sigue. Sean X_i , $i = 1, \dots, n$ variables aleatorias independientes Gamma (α_i, β) tal que $\mathbb{E}[X_i] = \alpha_i \beta$. Considere el proceso $Z(s_i) = X_0 + X_i$, $i = 1, \dots, n$. Encuentre la función de covarianza $C(i, j) = \mathbb{E}[Z(s_i), Z(s_j)]$ y la función de correlación. Es $Z(s_i)$ un proceso estacionario?, Cuál es la distribución de $Z(s_i)$?

Solución: Considerando la independencia de X_i se tiene que para $i \neq j$

$$C(i,j) = Cov(Z(s_i), Z(s_j))$$

$$= Cov(X_0 + X_i, X_0 + X_j)$$

$$= Cov(X_0, X_0) + Cov(X_0, X_j) + Cov(X_i, X_j) + Cov(X_i, X_0)$$

$$= Var(X_0) + E[X_0X_j] - E(X_0)E(X_j) + E[X_iX_j] - E(X_i)E(X_j) + E[X_iX_0] - E[X_0]E[X_i]$$

$$= Var(X_0) + E[X_0X_i] + E[X_0X_j] - E[X_0]\alpha_j\beta - E[X_0]\alpha_i\beta + E[X_iX_j] - \alpha_i\alpha_j\beta^2$$

$$= Var[X_0] + E[X_0X_i] + E[X_0X_j] - E[X_0]\beta(\alpha_j + \alpha_i) + E[X_iX_j] - \alpha_i\alpha_j\beta^2$$

$$= \alpha_0\beta^2 + \alpha_0\alpha_i\beta^2 + \alpha_0\alpha_j\beta^2 - \alpha_0(\alpha_j + \alpha_i)\beta^2 + E[X_iX_j] - \alpha_i\alpha_j\beta^2$$

$$= \begin{cases} \alpha_0\beta^2, & i \neq j \\ \alpha_0\beta^2 + \alpha_i\beta^2, & i = j \end{cases}$$

Por lo anterior, concluimos que

$$Var[Z(s_i)] = C(i,i) = \alpha_0 \beta^2 + \alpha_i \beta^2$$

y por ende

$$\rho(Z(s_i), Z(s_j)) = \frac{Cov(Z(s_i), Z(s_j))}{\sqrt{Var[Z(s_i)]Var[Z(s_j)]}}$$
$$= \frac{\alpha_0 \beta^2}{\sqrt{(\alpha_0 \beta^2 + \alpha_i \beta^2)(\alpha_0 \beta^2 + \alpha_j \beta^2)}}$$

Finalmente, como la media no es constante para i = j, deducimos que $Z(s_i)$ no es un proceso estacionario. Más aún, dada la independencia, $Z(s_i) \sim \text{Gamma}(a_0 + \alpha_i, \beta)$.

5) Sean $T_{ij} = Z(\mathbf{s}_i) - Z(\mathbf{s}_j)$ y $\mathbf{h}_{ij} = \mathbf{s}_i - \mathbf{s}_j$. Asuma que $T_{ij} \sim \mathcal{N}(0, 2\gamma(\mathbf{h}_{ij}, \boldsymbol{\theta}))$. Demuestre que

$$\rho\left(T_{ij}^{2}, T_{kl}^{2}\right) = \frac{\left\{\gamma\left(\boldsymbol{h}_{il}, \boldsymbol{\theta}\right) + \gamma\left(\boldsymbol{h}_{jk}, \boldsymbol{\theta}\right) - \gamma\left(\boldsymbol{h}_{jl}, \boldsymbol{\theta}\right) - \gamma\left(\boldsymbol{h}_{ik}, \boldsymbol{\theta}\right)\right\}^{2}}{4\gamma\left(\boldsymbol{h}_{ij}, \boldsymbol{\theta}\right)\gamma\left(\boldsymbol{h}_{kl}, \boldsymbol{\theta}\right)}.$$

Para solucionar el problema, consideraremos los siguientes lemas:

Lema. Sea $X \sim \mathcal{N}(0, \sigma^2)$, entonces $E[X^4] = 3\sigma^4$.

Lema. Sean $X, Y \sim \mathcal{N}(0, \sigma^2)$ tales que $\rho(X, Y) = \rho$, entonces $\rho(X^2, Y^2) = \rho^2$.

Solución: A partir del enunciado, notamos que

$$E[T_{ij}^2] = E[T_{ij}]^2 + Var[T_{ij}] = 2\gamma(\boldsymbol{h}_{ij}, \boldsymbol{\theta})$$

y
$$Var\left[T_{ij}^{2}\right] = E[T_{ij}^{4}] - E[T_{ij}^{2}]^{2} = 3Var[T_{ij}]^{2} - (2\gamma(\boldsymbol{h}_{ij}, \boldsymbol{\theta}))^{2} = 2(2\gamma(\boldsymbol{h}_{ij}, \boldsymbol{\theta}))^{2}.$$

Luego

$$\begin{split} \rho(T_{ij}^2, T_{kl}^2) &= \rho(T_{ij}, T_{kl})^2 \\ &= \frac{\operatorname{Cov}(T_{ij}, T_{kl})^2}{\operatorname{Var}(T_{ij}) \operatorname{Var}(T_{kl})} \\ &= \frac{\operatorname{Cov}(Z(s_j + \mathbf{h}_{ij}) - Z(s_j), Z(s_l + \mathbf{h}_{kl}) - Z(s_l))}{2\gamma(\mathbf{h}_{ij}, \boldsymbol{\theta}) 2\gamma(\mathbf{h}_{kl}\boldsymbol{\theta})} \\ &= \frac{\left\{ \gamma((s_j - s_l + \mathbf{h}_{ij}), \boldsymbol{\theta}) + \gamma((s_j - s_l - \mathbf{h}_{kl}), \boldsymbol{\theta}) - \gamma((s_j - s_l), \boldsymbol{\theta}) - \gamma((s_j - s_l + \mathbf{h}_{ij} - \mathbf{h}_{kl}), \boldsymbol{\theta}) \right\}^2}{2\gamma(\mathbf{h}_{ij}, \boldsymbol{\theta}) 2\gamma(\mathbf{h}_{kl}\boldsymbol{\theta})} \\ &= \frac{\left\{ \gamma(\mathbf{h}_{il}, \boldsymbol{\theta}) + \gamma(\mathbf{h}_{jk}, \boldsymbol{\theta}) - \gamma(\mathbf{h}_{jl}, \boldsymbol{\theta}) - \gamma(\mathbf{h}_{ik}, \boldsymbol{\theta}) \right\}^2}{4\gamma(\mathbf{h}_{ij}, \boldsymbol{\theta}) \gamma(\mathbf{h}_{kl}, \boldsymbol{\theta})}. \end{split}$$

6) Considere el conjunto de datos Santiago . csv el cual contiene 4 variables y 38 casos: Las variables son latitud, Longitud, KPI y Potencia. Cada caso corresponde a una planta de paneles solares instalada. Las localizaciones son las coordenadas de la planta, el KPI es un indicador del costo que tiene la construcción de cada planta, a menor KPI menor costo de construcción para la empresa. La variable potencia corresponde a la potencia nominal de la planta en kW.

Realice un análisis descriptivo espacial de la variable KPI de acuerdo a su georeferenciación. Determine si existe correlación espacial. En caso de existir correlación espacial verifique esto usando el variograma. Comente sobre anisotropía, ajuste del variograma usando algún modelo isotrópico paramétrico, rango de dependencia, efecto nugget y calidad del ajuste.

Solución:

El conjunto de datos Santiago contiene el número ostiones capturados en ciertas localizaciones. Las variables de interes son

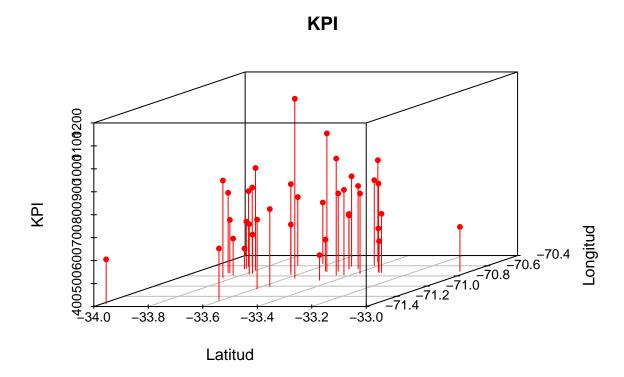
• Longitud: Longitud

• Latitud: Latitud.

• KPI: Indicador del costo asociado a la construcción de cada planta.

Nos interesa saber si existe autocorrelación espacial en este conjunto de datos y de haber, verificarlo con el variograma.

Haciendo uso de la libreria scatterplot3d podemos graficar los datos como sigue



Mediante la libreria ape calculamos el índice de Moran y su esperanza:

observed: -0.02346849

expected: -0.02702703

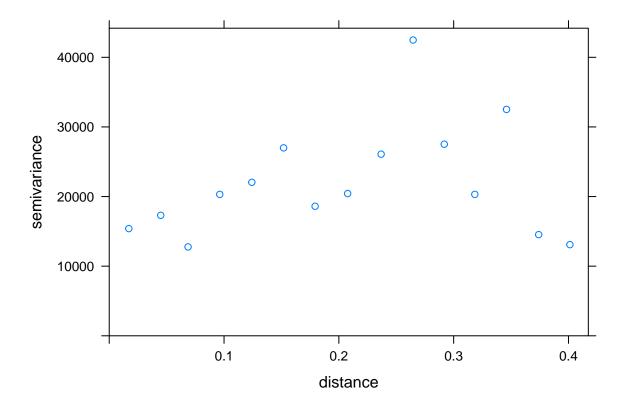
sd: 0.05029767

p-value: 0.9435972

Así, concluimos que

$$I = -0.02346849$$
 y $\mathbb{E}[I] = -0.02702703$.

Considerando $\alpha = 0.05$, podemos NO rechazar la Hipótesis $H_0 : \text{cor}[Y(s_i), Y(s_j)] = 0$, para todo i, j, pues p-value $> \alpha$. Es decir, NO existe autocorrelación espacial en la variable KPI.



De la misma manera, los **Cloud Variogram** y **Cressie-Hawkings Variogram** son respectivamente

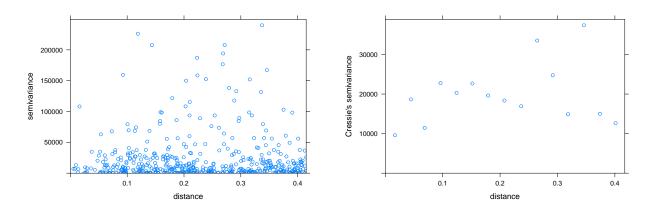
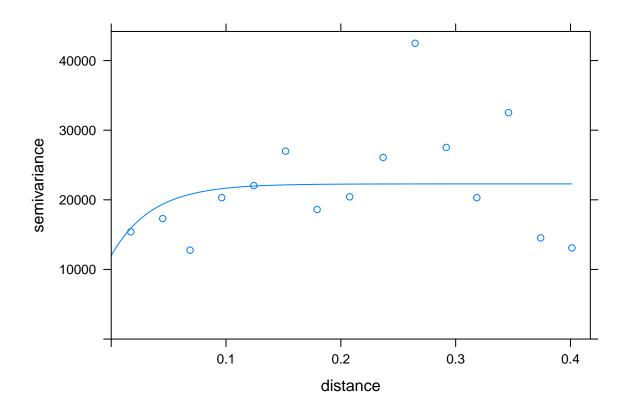


Figura 1: Cloud Variogram

Figura 2: Cressie Hawkings Variogram

Finalmente, ajustamos modelos al variograma empírico usando el comando fit.variogram y obtenemos

model psill range kappa 1 Nug 15042.83 0.00000000 0.0 2 Mat 10287.33 0.03560912 4.7 Superponiendo un modelo de Matern en el variograma empírico se vislumbra la siguiente gráfica



Anexo: Códigos

Los códigos usados pueden ser hallados en el siguiente GitHub