



# COLEGIO DE POSTGRADUADOS

---

---

INSTITUCIÓN DE ENSEÑANZA E INVESTIGACIÓN  
EN CIENCIAS AGRÍCOLAS

## CAMPUS MONTECILLO

POSTGRADO EN SOCIOECONOMÍA, ESTADÍSTICA  
E INFORMÁTICA-ESTADÍSTICA

**Pruebas estadísticas para la modelación del  
semivariograma en campos espaciales Gaussianos  
estacionarios**

María Guzmán Martínez

TESIS

PRESENTADA COMO REQUISITO PARCIAL PARA  
OBTENER EL GRADO DE

**DOCTORA EN CIENCIAS**

MONTECILLO, TEXCOCO, EDO. DE MÉXICO  
2015

---

---

La presente tesis titulada: **Pruebas estadísticas para la modelación del semivariograma en campos espaciales Gaussianos estacionarios**, realizada por la alumna: **María Guzmán Martínez**, bajo la dirección del Consejo Particular indicado, ha sido aprobada por el mismo y aceptada como requisito parcial para obtener el grado de:

DOCTORA EN CIENCIAS  
SOCIOECONOMÍA ESTADÍSTICA E INFORMÁTICA  
ESTADÍSTICA

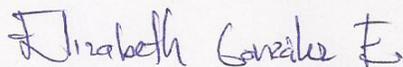
CONSEJO PARTICULAR

CONSEJERO



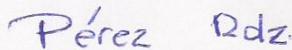
Dr. José Aurelio Villaseñor Alva

ASESOR



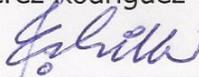
Dra. Elizabeth González Estrada

ASESOR



Dr. Paulino Pérez Rodríguez

ASESOR



Dr. Enrique Villa Diharce

ASESOR



Dr. Eduardo Castaño Tostado

## DEDICATORIA

A mi familia, a mis amigos, a la Dra. Elizabeth González Estrada y al Dr. José A. Villaseñor Alva.

## AGRADECIMIENTOS

Al Consejo Nacional de Ciencia y Tecnología, por el apoyo económico brindado para los estudios de mi Doctorado en Estadística en el Colegio de Postgraduados, Campus Montecillo.

A los integrantes de mi Consejo Particular:

Dr. José A. Villaseñor Alva, por su tiempo, paciencia, y conocimiento compartido, gracias Doctor.

Dra. Elizabeth González Estrada por su apoyo incondicional, gracias Doctora.

Dr. Paulino Pérez Rodríguez, por su apoyo, gracias Doctor.

Dr. Enrique Raúl Villa Diharce: por su apoyo y por tomarse el tiempo para formar parte de mi consejo, gracias Doctor.

Dr. Eduardo Castaño Tostado, por sus palabras de aliento y por su apoyo, gracias Doctor.

A la Sociedad Matemática Mexicana y a la Fundación Kovalévskaja por el apoyo financiero complementario para el desarrollo de la investigación matemática en México que se me ha otorgado.

# Pruebas estadísticas para la modelación del semivariograma en campos espaciales Gaussianos estacionarios

María Guzmán Martínez, DC

Colegio de Postgraduados, 2015

La Geoestadística es considerada como una rama de la estadística aplicada. Sus principales objetivos son la estimación, predicción y simulación de fenómenos en el espacio. La Geoestadística se da a la tarea de analizar y modelar la variabilidad de fenómenos espaciales utilizando la función del semivariograma. Para esto se requieren tres cosas: *i*) elegir un modelo para el semivariograma dentro de una familia de modelos que existen en la literatura para el semivariograma; *ii*) estimar los parámetros del modelo y *iii*) validar el ajuste del modelo propuesto. En la literatura existen varias propuestas con respecto a cada uno de estos tres puntos.

El objetivo de esta investigación fue determinar la estructura de dependencia de un proceso espacial Gaussiano estacionario. Para ello se proponen cuatro pruebas estadísticas. Estas pruebas ayudan en la toma de decisión al momento de proponer un modelo paramétrico para el semivariograma que modele la variabilidad del proceso espacial Gaussiano. Con la finalidad de estudiar el comportamiento de las pruebas estadísticas, se estiman los parámetros del semivariograma utilizando tres procedimientos de estimación: *mínimos cuadrados ponderados*, *máxima verosimilitud* y *máxima verosimilitud restringida*.

La primera prueba construye una envoltura gráfica para el semivariograma empírico de la muestra, por medio de remuestreo paramétrico. La segunda prueba también genera una envoltura gráfica para el semivariograma empírico de la muestra. La diferencia entre esta prueba estadística y la anterior, es el número de muestras que requiere el algoritmo para realizar el remuestreo paramétrico. La tercera prueba estadística utiliza como estadístico de prueba el *promedio del cuadrado del error normalizado y validación cruzada*. La cuarta prueba estadística utiliza como estadístico de prueba el coeficiente de correlación muestral. Esta prueba es para probar exclusivamente dependencia espacial exponencial.

El tamaño y la potencia de las pruebas estadísticas se estudiaron por medio de simulación Monte Carlo. Se realizó una aplicación con datos de lluvia acumulada de Suiza.

**Palabras clave:** Geoestadística, Semivariograma, Proceso Gaussiano, Pruebas estadísticas.

# Statistical tests for modeling the semivariogram in stationary Gaussian spatial fields

María Guzmán Martínez, DC

Colegio de Postgraduados, 2015

Geostatistics is considered a branch of Applied Statistics. Its main objectives are the estimation, prediction and simulation of phenomena in the space. The spatial variability of these phenomena is modeled through the semivariogram function.

In this manuscript, four statistical tests are proposed in order to determine the dependence structure (semivariogram function) of a stationary spatial Gaussian process. The parameters of the semivariogram function were estimated using three different procedures: weighted least squared, maximum likelihood and restricted maximum likelihood.

The first test builds a graphical envelope for empirical semivariogram by means of parametric bootstrap. The second test also generates a graphical envelope for the empirical semivariogram of the sample. The difference between this statistical test and the previous one is the number of samples required by the algorithm for performing the parametric resampling. The third statistical test uses as test statistic the mean square normalized error criteria and cross validation. The fourth statistical test uses as test statistic the sample correlation coefficient, and is special for testing exponential spatial dependence exclusively.

The size and power of the statistical tests were studied by means of Monte Carlo simulation. An application to cumulative rainfall data of Switzerland was performed.

**Key words:** Geostatistics, Semivariogram, Gaussian process, Statistical tests.

# Contenido

<b>1. Introducción</b>	<b>1</b>
<b>2. Marco Teórico</b>	<b>4</b>
2.1. Procesos estocásticos . . . . .	4
2.2. Proceso Gaussiano espacial estacionario . . . . .	6
2.3. Semivariogramas . . . . .	8
2.3.1. Modelos para el semivariograma . . . . .	9
2.3.2. Semivariogramas empíricos . . . . .	10
2.4. Estimación del semivariograma paramétrico . . . . .	12
<b>3. Pruebas empíricas</b>	<b>15</b>
3.1. Primera prueba empírica . . . . .	15
3.1.1. Nivel de la prueba empírica . . . . .	17
3.1.2. Estudio de simulación . . . . .	18
3.1.3. Resultados . . . . .	20
3.2. Segunda prueba estadística . . . . .	24
3.2.1. Estudio de simulación . . . . .	25
3.2.2. Resultados . . . . .	27

## Contenido

---

3.3. Aplicación . . . . .	29
<b>4. Prueba basada en el criterio del <i>MSNE</i></b>	<b>35</b>
4.1. Prueba estadística modificada . . . . .	36
4.2. Estudios de simulación . . . . .	37
4.3. Resultados . . . . .	38
4.4. Aplicación . . . . .	40
<b>5. Prueba de correlación muestral</b>	<b>41</b>
5.1. Estudio de simulación . . . . .	42
5.2. Resultados . . . . .	43
<b>6. Conclusiones</b>	<b>45</b>
<b>Bibliografía</b>	<b>47</b>
<b>A. Conceptos teóricos</b>	<b>50</b>
A.1. Probabilidad de variables aleatorias . . . . .	50
A.2. Función teórica del semivariograma . . . . .	50
A.3. Kriging ordinario . . . . .	52
A.4. Método de bootstrap paramétrico . . . . .	55
<b>B. Programas en R</b>	<b>56</b>
B.1. Primera prueba empírica . . . . .	56
B.2. Segunda prueba empírica . . . . .	58
B.3. Prueba con el criterio <i>MSNE</i> . . . . .	61

## Contenido

---

B.4. Prueba bootstrap basada en el criterio <i>MSNE</i> . . . . .	63
B.5. Prueba de correlación muestral . . . . .	66

# Lista de Figuras

2.1. Función de semivariograma. . . . .	10
3.1. Mapa de Suiza con 467 estaciones. . . . .	30
3.2. Mapa de Suiza con 100 estaciones. . . . .	31
3.3. Semivariogramas empíricos. . . . .	31
3.4. Estimación del semivariograma Matérn . . . . .	33

# Lista de Cuadros

2.1. Algunas funciones de correlación. . . . .	8
3.1. Tamaño y potencia estimados de la primera prueba en una malla regular	20
3.2. Tamaño y potencia estimados de la primera prueba para una malla irregular	21
3.3. Tamaño y potencia estimados de la primera prueba para una malla regular	22
3.4. Tamaño y potencia estimados de la primera prueba en una malla irregular	23
3.5. Tamaño y potencia estimados de la primera prueba en una malla irregular	24
3.6. Tamaño y potencia estimados de la segunda prueba en una malla regular	27
3.7. Tamaño y potencia estimados de la segunda prueba en una malla irregular	28
3.8. Tamaño y potencia estimados de la segunda prueba empírica en una malla irregular . . . . .	29
3.9. Estimación de $\theta$ y su error estándar bajo $H_0$ . . . . .	32
3.10. Aplicación de la prueba empírica para probar $H_0$ en (3.3) con $\alpha = 0.05$ . .	33
4.1. Tamaño y potencia estimados de la prueba basada en el criterio $MSNE$ .	39
4.2. Tamaño y potencia estimados de la prueba bootstrap basada en el criterio $MSNE$ . . . . .	40
5.1. Tamaño y potencia estimados de la prueba de correlación en una malla regular . . . . .	44

5.2. Tamaño y potencia estimados de la prueba de correlación en una malla irregular . . . . .	44
---	----

# Capítulo 1

## Introducción

La Geoestadística es considerada como una rama de la estadística aplicada que estudia fenómenos espaciales (Journel y Huijbregts, 1978, Matheron, 1962). La Geoestadística emergió a principios de 1980 como una mezcla híbrida de cuatro disciplinas: ingeniería minera, geología, matemática y estadística (Cressie, 1993). Dentro de los objetivos de la Geoestadística están la estimación, predicción y simulación de fenómenos en el espacio (Cressie y Wikle, 2011). La Geoestadística también se ocupa del análisis y modelación de la variabilidad espacial.

La Geoestadística es ampliamente utilizada en áreas como: Biología, Geografía, Economía, Procesamiento de imágenes, Ciencias de la Tierra, Ciencias del Medio Ambiente, Ecología, Geografía, Epidemiología, Agronomía, Recursos Forestales, Prospección de Minerales (Bustos y Guerrero (2011), < [http : //www.inpe.br/](http://www.inpe.br/) > [27/09/2015], < [http : //www.conae.gov.ar/index.php/espanol/](http://www.conae.gov.ar/index.php/espanol/) > [27/09/2015]).

Los fenómenos espaciales,  $Y(s)$ , que estudia la Geoestadística son de variación continua en un subconjunto fijo,  $D \subset \mathbb{R}^d$  (espacio euclidiano  $d$ -dimensional), es decir  $s$  varía continuamente en  $D$ . A estos fenómenos se les suele clasificar como fenómenos espaciales *isotrópicos* si su variabilidad espacial solo depende de la distancia, y *anisotrópicos*, cuando su variabilidad espacial depende de la distancia y de la dirección.

Un fenómeno espacial  $Y(s)$  puede ser considerado como una realización de un proceso estocástico (campo aleatorio)  $\{S(s) : s \in D \subset \mathbb{R}^d\}$ .

En Geoestadística los procesos espaciales Gaussianos son de gran interés por dos razones. La primera es que la predicción, estimación y la distribución teórica son relativamente más fáciles de manejar. La segunda es que se puede hacer uso de las consideraciones asintóticas referentes a la distribución Gaussiana que existen en la literatura (Cressie, 1993). Un ejemplo de trabajos realizados para campos aleatorios Gaussianos son los de Kent y Mohammadzadeh (1999), Zimmerman y Cressie (1992), Gregori y Mateu (2014)

## 1. Introducción

---

y [Menezes y Febrero-Bande \(2005\)](#).

Modelar la variabilidad presente en los fenómenos espaciales es de gran interés. Si estos tienen estacionaridad intrínseca, la función del semivariograma dada por

$$\gamma(h) = \frac{1}{2} \text{Var}(Y(s_i) - Y(s_j)), s_i, s_j \in D, \quad (1.1)$$

es utilizada para modelar su variabilidad espacial ([Müller y Zimmerman, 1997](#)). En la literatura existe una gran cantidad de funciones (modelos) para la expresión (1.1) ([Banerjee y Gelfand, 2014](#)). En general, los modelos para el semivariograma involucran tres parámetros: la pepita,  $\tau^2$ ; la meseta,  $\sigma^2$ ; y el rango,  $\phi$ ; el cual es el límite de la correlación espacial. La siguiente ecuación

$$\gamma(h) = \tau^2 + \sigma^2(1 - \rho(h, \phi)), \quad (1.2)$$

expresa de forma general el modelo para la expresión (1.1).

Determinar o *validar* un modelo para el semivariograma, con la finalidad de modelar la variabilidad presente en un proceso espacial Gaussiano estacionario, es una tarea de la cual varios investigadores se han ocupado ([Glatzer y Müller, 2004](#), [Zhang et al., 2005](#)). Entre ellos están [Clark y Allingham \(2011\)](#) que modelan la variabilidad presente en los campos aleatorios Gaussianos, con la construcción de intervalos de confianza para el semivariograma empírico de la muestra. [Barry \(1996\)](#) propuso un diagnóstico gráfico para evaluar el ajuste del modelo para el semivariograma dado un conjunto de datos. [Oliver y Webster \(2014\)](#) enumeran varias cuestiones que se deben tener en cuenta cuando se trata de validar algún modelo para el semivariograma. Entre los criterios de validación que recomiendan [Oliver y Webster \(2014\)](#) incluyen el *el criterio de información de Akaike* (AIC) ([Akaike, 1998](#)) y el *promedio del cuadrado del error normalizado*.

Entre los métodos que existen en la literatura para estimar los parámetros del modelo para el semivariograma están: *mínimos cuadrados ponderados (wls)* ([Cressie, 1985](#)), *máxima verosimilitud (ml)* y *máxima verosimilitud restringida (reml)* ([Cressie, 2015](#)).

En este contexto, el objetivo de esta investigación es proponer una prueba estadística que ayude a determinar un modelo para el semivariograma con la finalidad de modelar la variabilidad presente en los campos espaciales Gaussianos estacionarios.

Bajo este objetivo se proponen cuatro pruebas estadísticas:

La primera prueba estadística se propone utilizando bootstrap paramétrico ([Gaetan y Bleakley, 2010](#)).

La segunda prueba estadística es una adaptación al caso espacial de la metodología que genera envolturas gráficas ([Davison y Hinkley, 1997](#)).

## 1. Introducción

---

La tercera prueba estadística está planteada en términos de dos criterios ampliamente utilizados en la Geoestadística, la *validación cruzada* y el *promedio del cuadrado del error normalizado* (Gaetan y Bleakley, 2010).

La cuarta prueba estadística se planteó para probar exclusivamente dependencia exponencial en un campo espacial Gaussiano estacionario.

Dentro de los objetivos específicos de esta investigación están: estudiar el tamaño y la potencia de las cuatro pruebas estadísticas, así como conocer el efecto de los tres métodos de estimación (*wls*, *ml* y *reml*) en las propiedades de las pruebas estadísticas que se proponen. Estos estudios se realizaron por medio de simulación.

La estructura del trabajo es como sigue. En el Capítulo 2, se describe el marco teórico de la Geoestadística. En el Capítulo 3, se describen las primeras dos pruebas estadísticas, que denotamos con el nombre de *pruebas empíricas*. En el Capítulo 4, se plantea la prueba estadística que utiliza el criterio *MSNE* como estadístico de prueba. Después en el Capítulo 5, se presenta la prueba de correlación muestral para probar dependencia exponencial. Por último en el Capítulo 6, se dan las conclusiones.

# Capítulo 2

## Marco Teórico

### 2.1. Procesos estocásticos

**Definición 1 (Proceso estocástico)** *Un proceso estocástico  $\{S(t), t \in T\}$  es una colección de variables aleatorias. Es decir, para cada  $t \in T$ ,  $S(t)$  es una variable aleatoria (Ross, 1996).*

Si el proceso estocástico  $S$ , está indizado por un espacio de coordenadas  $s \in D \subset \mathbb{R}^d$ , es decir  $\{S(s), s \in D\}$ , se le llama *campo aleatorio* (Grigoriu, 2013).

Una realización del campo aleatorio  $\{S(s), s \in D\}$  con  $n$  observaciones, está dada por la secuencia  $\{S(s_1) = y(s_1), \dots, S(s_n) = y(s_n), s_i \in D\}$ . Generalmente, a partir de la muestra  $y(s_1), \dots, y(s_n)$  se intenta conocer las características del proceso  $S$  en  $s_i, i = 1, \dots, n$ , y con esta información hacer inferencia del proceso  $S(s)$  en todo  $D$ .

**Definición 2** *Un proceso estocástico  $\{S(s), s \in D\}$  es estrictamente estacionario si dado  $n \geq 1$ ,  $\{s_1, \dots, s_n\}$ , cualquier conjunto de  $n$  sitios en  $D$ , y para cualquier  $h \in \mathbb{R}^2$ , la distribución conjunta de  $\{S(s_1), \dots, S(s_n)\}$  es la misma que  $\{S(s_1 + h), \dots, S(s_n + h)\}$  (Banerjee y Gelfand, 2014).*

Es decir, si un proceso estocástico es estrictamente estacionario, entonces la distribución conjunta de las variables aleatorias es invariante en  $D$ .

Nótese que la estacionaridad estricta en un proceso estocástico no dice nada sobre la estructura de correlación que existe entre las variables  $\{S(s_1), \dots, S(s_n)\}$ ; lo que sí se sabe

## 2.1. Procesos estocásticos

---

es que la correlación entre  $S(s_i)$  y  $S(s_j)$  sólo depende de la diferencias de las coordenadas,  $\|s_i - s_j\|$ .

**Definición 3** *Un proceso estocástico  $\{S(s), s \in D\}$  es estacionario de segundo orden (estacionaridad débil) si*

$$\begin{aligned} E(S(s)) &= \mu, \\ Cov(S(s), S(s+h)) &= C(h), \quad s, s+h \in D, \end{aligned}$$

donde  $C(\cdot)$ , es la covarianza de  $S(s)$  y  $S(s+h)$ . En Geoestadística a  $C(\cdot)$  se le conoce como *covariograma*. A un proceso estocástico estacionario de segundo orden también se le llama proceso estocástico de covarianza estacionaria.

**Definición 4** *Un proceso estocástico  $\{S(s), s \in D\}$  tiene estacionaridad intrínseca si*

$$E(S(s_i) - S(s_j)) = 0, \quad s_i, s_j \in D$$

y

$$Var(S(s_i) - S(s_j)) = E[S(s_i) - S(s_j)]^2 = 2\gamma(h),$$

donde  $h = \|s_i - s_j\|$ ,  $s_i, s_j \in D$ .

A  $2\gamma(h)$  se le conoce como *variograma* y a  $\gamma(h)$  como *semivariograma*. Generalmente, se emplea el semivariograma para modelar la variabilidad espacial del campo  $S(s)$ .

Tener estacionaridad de segundo orden en un proceso estocástico implica tener también estacionaridad intrínseca. Es decir, la estacionaridad de segundo orden es más fuerte que la estacionaridad intrínseca.

Si  $\{S(s), s \in D\}$  es un campo aleatorio con estacionaridad intrínseca, entonces

$$E(S(s_i) - S(s_j)) = 0, \quad s_i, s_j \in D,$$

y

$$2\gamma(s_i, s_j) = Var(S(s_i) - S(s_j)) \quad s_i, s_j \in D.$$

Desarrollando para  $2\gamma(s_i, s_j)$ , se tiene

$$\begin{aligned} Var(S(s_i) - S(s_j)) &= E[(S(s_i) - S(s_j)) - E(S(s_i) - S(s_j))]^2 \\ &= E[S(s_i) - S(s_j)]^2 \\ &= E[S^2(s_i) + S^2(s_j) - 2S(s_i)S(s_j)] \\ &= Var(S(s_i)) + Var(S(s_j)) - 2Cov(S(s_i), S(s_j)), \quad \forall s_i, s_j \in D. \end{aligned}$$

## 2.2. Proceso Gaussiano espacial estacionario

---

Es decir,

$$2\gamma(h) = \text{Var}(S(s_i)) + \text{Var}(S(s_j)) - 2\text{Cov}(S(s_i), S(s_j)), \quad \forall s_i, s_j \in D.$$

El semivariograma está dado por

$$\gamma(h) = \frac{1}{2} \{ \text{Var}(S(s_i)) + \text{Var}(S(s_j)) - 2\text{Cov}(S(s_i), S(s_j)) \}, \quad \forall s_i, s_j \in D.$$

La relación entre variograma y covariograma está dada por

$$\begin{aligned} 2\gamma(h) &= \text{Var}(S(s_i)) + \text{Var}(S(s_j)) - 2\text{Cov}(S(s_i), S(s_j)) \\ &= C(0) + C(0) - 2C(h), \end{aligned}$$

es decir,

$$\gamma(h) = C(0) - C(h). \quad (2.1)$$

Nótese que dado  $C$  se puede recuperar  $\gamma$ . En cambio, si se tiene  $\gamma$  y se quiere recuperar  $C$ , se requiere que el campo espacial sea ergódico, es decir, que  $C(h) \rightarrow 0$  cuando  $h \rightarrow \infty$ . Es decir, que la covarianza entre los valores de dos puntos tiende a cero cuando los puntos tienden a separarse en el espacio ([Banerjee y Gelfand, 2014](#)).

## 2.2. Proceso Gaussiano espacial estacionario

**Definición 5 (Campo espacial Gaussiano)** *Un campo estocástico  $\{S(s) : s \in D \subset \mathbb{R}^2\}$  donde  $s$  varía continuamente en  $D$ , un subconjunto fijo de  $\mathbb{R}^2$ , es un campo espacial Gaussiano si para cualquier colección de localizaciones  $s_1, \dots, s_n$  con  $s_i \in D$ , la distribución conjunta de  $\{S(s_1), \dots, S(s_n)\}$  es Gaussiana Multivariada ([Diggle y Ribeiro, 2007](#)).*

**Definición 6 (Proceso Espacial Gaussiano Estacionario)** *El campo espacial Gaussiano  $\{S(s) : s \in D \subset \mathbb{R}^2\}$ , es estacionario si*

$$E(S(s)) \text{ es constante para toda } s \in D,$$

$$\text{Var}(S(s)) \text{ es constante para toda } s \in D,$$

y la función de correlación

$$\rho(h) = \text{Corr}(S(s_i), S(s_j)),$$

sólo depende de  $h = \|s_i - s_j\|$ ,  $s_i, s_j \in D$ .

Generalmente  $\{S(s) : s \in D \subset \mathbb{R}^2\}$  no es directamente observable, en su lugar se tiene un conjunto disponible de observaciones  $y(s_1), \dots, y(s_n)$ , que provienen de las variables

## 2.2. Proceso Gaussiano espacial estacionario

---

aleatorias  $Y(s_1), \dots, Y(s_n)$ ,  $s_i \in D$ . Es decir,  $Y(s_i)$  es la versión ruidosa de  $S(s_i)$  donde  $s_i \in D$  ([Diggle y Ribeiro, 2007](#), [Diggle y Christensen, 2003](#)).

Sea

$$\{S(s) : s \in D \subset \mathbb{R}^2\}, \quad (2.2)$$

un campo espacial Gaussiano estacionario (*CEGE*). Un modelo para (2.2) puede escribirse como la suma de dos variables aleatorias,

$$Y(s) = S(s) + Z(s), \quad s \in D, \quad (2.3)$$

donde  $S(s)$  es el campo espacial Gaussiano estacionario con distribución

$$S(s) \sim \mathcal{N}(\mu, \sigma^2),$$

y  $Z(s)$  es el término de error con distribución

$$Z(s) \sim \mathcal{N}(0, \tau^2).$$

Sea  $\mathbf{Y}(s) = \{Y(s_1), \dots, Y(s_n)\}$  una realización del campo (2.2) y

$$Y(s_i) = S(s_i) + Z(s_i); \quad i = 1, \dots, n, \quad s_i \in D,$$

donde  $Z(s_1), \dots, Z(s_n)$ , son independientes e idénticamente distribuidos ([Diggle y Christensen, 2003](#)). La distribución conjunta de  $\mathbf{Y}(s)$ , que es Normal Multivariada de dimensión  $n$ , está dada por

$$\mathbf{Y}(s) \sim \mathcal{NM}(\mu \mathbf{1}, \sigma^2 \mathbf{R}(\phi) + \tau^2 \mathbf{I}), \quad (2.4)$$

donde  $\mu$  es el promedio del proceso,  $\mathbf{1}$  es un vector de dimensión  $n \times 1$ ,  $\tau^2 + \sigma^2$  es la varianza del campo,  $\phi$  es un parámetro de escala,  $\mathbf{I}$  es la matriz identidad de dimensión  $n \times n$  y  $\mathbf{R}(\phi)$  es la matriz de correlaciones de dimensión  $n \times n$  cuyos elementos están dados por

$$(\mathbf{R}(\phi))_{ij} = \rho_Y(h_{ij}, \phi),$$

es decir,  $(\mathbf{R}(\phi))_{ij}$  es la correlación que existe entre  $Y(s_i)$  y  $Y(s_j)$ ,  $s_i, s_j \in D$ ,  $h_{ij} = \|s_i - s_j\|$ .

En la literatura existen varias expresiones paramétricas para la función  $\rho_Y(h, \phi)$  ([Gaetan y Bleakley, 2010](#)). En el Cuadro 2.1 se muestran algunas de ellas.

## 2.3. Semivariogramas

**Cuadro 2.1:** Algunas funciones de correlación.

Correlación	Expresión	Rango
Exponencial	$\rho_Y(h, \phi) = \begin{cases} \exp\left(-\frac{h}{\phi}\right) \\ 0 \end{cases}$	$\begin{cases} h \neq 0, \phi > 0 \\ h = 0, \phi > 0 \end{cases}$
Esférica	$\rho_Y(h, \phi) = \begin{cases} 1 - 1.5\left(\frac{h}{\phi}\right) + 0.5\left(\frac{h}{\phi}\right)^3 \\ 0 \end{cases}$	$\begin{cases} 0 < h \leq \phi \\ h > \phi \end{cases}$
Circular	$\rho_Y(h, \phi) = \frac{2}{\pi} \left\{ \cos^{-1}\left(\frac{h}{\phi}\right) - \frac{h}{\phi} \sqrt{1 - \left(\frac{h}{\phi}\right)^2} \right\}$	$\{ 0 < h \leq \phi \}$
Cúbica	$\rho_Y(h, \phi) = \begin{cases} 1 - \left[ 7\left(\frac{h}{\phi}\right)^2 - \frac{35}{4}\left(\frac{h}{\phi}\right)^3 + \frac{7}{2}\left(\frac{h}{\phi}\right)^5 - \frac{3}{4}\left(\frac{h}{\phi}\right)^7 \right] \\ 0 \end{cases}$	$\begin{cases} 0 \leq h \leq \phi \\ h > \phi \end{cases}$

La función de semivariograma que corresponde al proceso (2.4), está dada por

$$\gamma(h) = \tau^2 + \sigma^2 \left( 1 - \frac{\rho_S(h)}{\sigma^2 + \tau^2} \right). \quad (2.5)$$

Si  $\tau^2 = 0$  en el modelo (2.4), entonces

$$\gamma(h) = \sigma^2(1 - \rho_S(h)),$$

es la función de semivariograma que corresponde para el modelo (2.4).

La deducción de la expresión (2.5) se encuentra en el Apéndice (A).

## 2.3. Semivariogramas

Webster y Oliver (2007) y Oliver y Webster (2014) describen dos tipos de semivariogramas: el *semivariograma regional* y el *semivariograma experimental*.

El *semivariograma regional* se puede calcular si se cuenta con la información completa de la región bajo estudio. Pero dependiendo del tipo de estructura de correlación que tenga el campo espacial, el *semivariograma regional* podría no explicar totalmente la información del campo espacial, aún cuando se cuente con la información completa de la región. Un ejemplo de ello es un campo estacionario de segundo orden no limitado en una región pequeña. También podría pasar que la región de correlación del semivariograma sea más grande que la distancia máxima de la región. Brus y De Gruijter (1994), llaman a este tipo de semivariogramas no-ergódicos.

## 2.3. Semivariogramas

---

El *semivariograma experimental* o también conocido como *semivariograma muestral*, es el semivariograma que se calcula a partir de una realización,  $\{y(s_1), \dots, y(s_n)\}$ ,  $s_i \in D$ , del campo espacial  $\{S(s), s \in D\}$ . El semivariograma muestral estima al semivariograma regional.

### 2.3.1. Modelos para el semivariograma

Existen en la literatura varios modelos (o semivariogramas paramétricos) para el semivariograma,  $\gamma(h)$ . Algunos modelos que suelen utilizarse en la práctica son:

Semivariograma exponencial:

$$\gamma_{exp}(h, \boldsymbol{\theta}) = \tau^2 + \sigma^2 \left(1 - \exp\left(-\frac{h}{\phi}\right)\right), \quad h > 0, \quad \phi > 0.$$

Semivariograma esférico:

$$\gamma_{esf}(h, \boldsymbol{\theta}) = \begin{cases} \tau^2 + \sigma^2 \left(1.5 \left(\frac{h}{\phi}\right) - 0.5 \left(\frac{h}{\phi}\right)^3\right), & \text{si } 0 \leq h \leq \phi, \\ \tau^2 + \sigma^2, & \text{si } h > \phi. \end{cases}$$

Semivariograma circular:

$$\gamma_{cir}(h, \boldsymbol{\theta}) = \begin{cases} \tau^2 + \sigma^2 \left\{1 - \frac{2}{\pi} \cos^{-1}\left(\frac{h}{\phi}\right) + \frac{2h}{\pi\phi} \sqrt{1 - \left(\frac{h}{\phi}\right)^2}\right\}, & \text{si } 0 \leq h \leq \phi, \\ \tau^2 + \sigma^2, & \text{si } h > \phi. \end{cases}$$

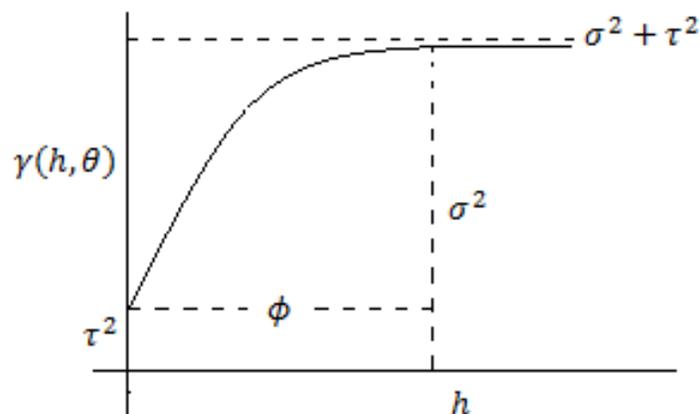
Semivariograma cúbico:

$$\gamma_{cub}(h, \boldsymbol{\theta}) = \begin{cases} \tau^2 + \sigma^2 \left(7 \left(\frac{h}{\phi}\right)^2 - 8.75 \left(\frac{h}{\phi}\right)^3 + 3.5 \left(\frac{h}{\phi}\right)^5 - 0.75 \left(\frac{h}{\phi}\right)^7\right), & \text{si } h \leq \phi, \\ \tau^2 + \sigma^2, & \text{si } h > \phi. \end{cases}$$

Aquí  $\boldsymbol{\theta} = (\tau^2, \sigma^2, \phi)^t$ . El parámetro  $\tau^2$  es la pepita, y representa una discontinuidad puntual en el origen. Ésta puede ser debida a errores de medición o escala del proceso espacial. La suma  $\tau^2 + \sigma^2$  es la meseta (la varianza del campo) y representa la cota superior del semivariograma. El parámetro  $\phi$  es el rango del semivariograma paramétrico que representa la distancia a partir de la cual dos observaciones son independientes. En la Figura (2.1) se muestra la gráfica típica de la función de semivariograma.

## 2.3. Semivariogramas

**Figura 2.1:** Función de semivariograma.



El *rango efectivo* (o *rango práctico*) del semivariograma paramétrico está definido como la distancia en la cual el semivariograma paramétrico explica el 95 % de la variabilidad total del campo espacial Gaussiano estacionario (Webster y Oliver, 2007).

En el modelo

$$Y(s_i) = S(s_i) + Z(s_i), \quad i = 1, \dots, n; \quad s_i \in D,$$

$S(s_i)$  introduce  $\sigma^2$  (una parte de la *meseta*) y  $\phi$ , mientras que  $Z(s_i)$  introduce  $\tau^2$ , (la otra parte de la *meseta*) (Banerjee y Gelfand, 2014).

### 2.3.2. Semivariogramas empíricos

Existen en Geoestadística varios métodos para estimar empíricamente a  $\gamma(h)$ .

Un estimador empírico de  $\gamma(h)$  que hasta cierto punto es intuitivo, se da a continuación.

A partir de

$$\gamma(h) = \frac{1}{2} \text{Var}(S(s_i) - S(s_j)),$$

donde  $h := \|s_i - s_j\|$ , que es equivalente a

$$\gamma(h) = \frac{1}{2} E(S(s_i) - S(s_j))^2, \quad (2.6)$$

se construye un estimador empírico para  $\gamma(h)$ , el cual denotaremos como  $\gamma_n(h)$ .

Como

$$Y(s_i) = S(s_i) + Z(s_i), \quad i = 1, \dots, n; \quad s_i \in D, \quad (2.7)$$

### 2.3. Semivariogramas

se sustituyen (2.7) en (2.6), entonces

$$\gamma(h) = \frac{1}{2} E (Y(s_i) - Y(s_j))^2, \quad (2.8)$$

A partir de  $(Y(s_1), \dots, Y(s_n))$ , una realización del campo espacial (2.2), se puede obtener una estimación para (2.8) con las  $Y(s_i)$  que se encuentren a una distancia  $h$ . Es decir,

$$\gamma_n(h) = \frac{1}{2N_h} \sum_{(s_i, s_j) \in N(h)} (Y(s_i) - Y(s_j))^2, \quad (2.9)$$

donde  $N(h)$  contiene todos los pares de puntos  $(s_i, s_j)$  a una distancia  $h$ , con una cierta tolerancia  $\Delta$ , es decir,

$$N(h) = \{(s_i, s_j) : h - \Delta \leq \|s_i - s_j\| \leq h + \Delta; i, j = 1, \dots, n\},$$

$N_h$  denota la cardinalidad de  $N(h)$  y  $h \in \mathcal{H} = \{h_1, \dots, h_r\}$ .

El semivariograma empírico (2.9) fue propuesto por Matheron (1962). Aunque este estimador es insesgado, su comportamiento es pobre cuando existen valores atípicos en la muestra debido al término al cuadrado que aparece en la suma (Menezes y Febrero-Bande, 2005). Por otra parte, Webster y Oliver (1992), mostraron que el valor del término  $N_h$  debe de ser mayor o igual a 100.

Cressie y Hawkins (1980) proponen un estimador empírico robusto ante valores atípicos, el cual está dado por

$$\gamma_n(h) = \frac{\left(\frac{1}{N_h} \sum_{(s_i, s_j) \in N(h)} |Y(s_i) - Y(s_j)|^{1/2}\right)^4}{0.914 + \frac{0.988}{N_h} + \frac{0.09}{N_h^2}}. \quad (2.10)$$

El denominador de la expresión (2.10) es una corrección al asumir que  $Y(s_i) - Y(s_j)$  tiene distribución Normal (Cressie, 1993).

Otro estimador empírico para  $\gamma(h)$  que proponen Cressie y Hawkins (1980) es

$$\gamma_n(h) = \frac{\left[ \text{mediana} \left\{ |Y(s_i) - Y(s_j)|^{1/2} \right\} \right]^4}{0.914}. \quad (2.11)$$

Dowl (1984) propone como estimador empírico para  $\gamma(h)$  a

$$\gamma_n(h) = 2.198 \{ \text{mediana}(|V_i(h)|) \}^2, \quad (2.12)$$

donde  $V_i(h) = Y(s_i) - Y(s_i + h)$  para  $i = 1, 2, \dots, N(h)$ . Este estimador supone que las diferencias están simétricamente distribuidas, tal que la esperanza de la mediana de las diferencias es cero.

## 2.4. Estimación del semivariograma paramétrico

---

Genton (1998) propone como estimador empírico para  $\gamma(h)$  a

$$\gamma_n(h) = \left( 2.219 \{ |V_i(h) - V_j(h)| ; i < j \}_{(k)} \right)^2, \quad (2.13)$$

donde

$$k = \binom{[N_h/2] + 1}{2},$$

y  $[N_h/2]$  denota la parte entera de  $N_h/2$ . En la expresión (2.13), primero se ordenan el conjunto de las diferencias absolutas  $|V_i(h) - V_j(h)|$  para  $i < j$ , después se calcula el  $k$ -ésimo cuantil de este conjunto. La constante 2.219 se incluye con la finalidad de tener consistencia con la distribución Gaussiana. Lo que permite este estimador es calcular la  $k$ -ésima estadística de orden de las combinaciones  $\binom{[N_h/2]+1}{2}$  de las distancias.

El estimador de Genton (1998), está basado en el estimador de escala,  $Q_N$ , de Rousseeuw y Croux (1992).

Journal y Huijbregts (1978) hacen varias sugerencias para calcular un estimador empírico para  $\gamma(h)$ , entre ellas:

1.  $\gamma_n(h)$  se estima con valores de  $|h|$  menores que la mitad de la distancia máxima en  $D$ , es decir,

$$|h| < \frac{d_{\text{máx}}}{2}.$$

Si  $D$  es un rectángulo de lados  $a$  y  $b$ , entonces  $d_{\text{máx}} = \frac{\sqrt{a^2+b^2}}{2}$ .

2. La constante  $N_h$  debe de estar entre

$$10 \leq N_h \leq 25.$$

Lark (2000) realizó un estudio en el que compara los semivariogramas empíricos de Matheron (1962), Cressie y Hawkins (1980), Dowd (1984) y Genton (1998). Utilizando la mediana del error cuadrado normalizado, Lark (2000) concluyó que el mejor estimador empírico para  $\gamma(h)$  cuando se tiene muestras con valores atípicos es el que propone Dowd (1984). Pero cuando la muestra no incluye valores atípicos, el estimador de Dowd (1984) es el menos eficiente.

En los estudios de simulación que se realizarán, se trabaja con datos sin observaciones atípicas.

## 2.4. Estimación del semivariograma paramétrico

Para modelar la dependencia espacial en (2.4), es necesario ajustar un modelo paramétrico a  $\gamma(h)$ . Métodos como el de mínimos cuadrados ordinarios (*ols*), mínimos cuadrados

## 2.4. Estimación del semivariograma paramétrico

ponderados (*wls*) o mínimos cuadrados generalizados (*gls*) ayudan a estimar el vector de parámetros  $\boldsymbol{\theta}$  de la función  $\gamma(h, \boldsymbol{\theta})$  (Gaetan y Bleakley, 2010, Menezes y Febrero-Bande, 2005). Estos tres métodos siguen el siguiente esquema:

1. Se calcula el semivariograma empírico,  $\hat{\gamma}_n(h)$ ,  $h \in \mathcal{H}$ .
2. Con la ayuda de  $\hat{\gamma}_n(h)$  el investigador elige, de entre una familia de modelos paramétricos, el modelo  $\gamma(h, \boldsymbol{\theta})$  para  $\gamma(h)$  (Gaetan y Bleakley, 2010).
3. A partir de  $\hat{\gamma}_n(h)$  y  $\gamma(h, \boldsymbol{\theta})$ , el método de *wls* (*ols* o *gls*) estima a  $\boldsymbol{\theta}$ .

Aquí la estimación de  $\boldsymbol{\theta}$  se ve afectada por el estimador empírico que se elija para  $\gamma(h)$ .

A continuación se muestran los métodos *ols*, *wls* y *gls*.

Mínimos cuadrados ordinarios:

$$\boldsymbol{\theta}_{ols} = \min_{\boldsymbol{\theta} \in \Theta} \sum_{k=1}^r (\hat{\gamma}_n(h_k) - \gamma(h_k, \boldsymbol{\theta}))^2,$$

este criterio necesita que los  $\hat{\gamma}_n(h_k)$ ,  $h_k \in \mathcal{H} = \{h_1, \dots, h_r\}$  sean independientes y que tengan la misma varianza. Dado que difícilmente se cumplen estos supuestos se recurre a la técnica de mínimos cuadrados generalizados:

$$\boldsymbol{\theta}_{gls} = \min_{\boldsymbol{\theta} \in \Theta} (\hat{\gamma}_n - \gamma(\boldsymbol{\theta})) \{Cov_{\boldsymbol{\theta}}(\hat{\gamma}_n)\}^{-1} (\hat{\gamma}_n - \gamma(\boldsymbol{\theta})),$$

donde  $\hat{\gamma}_n = \{\hat{\gamma}_n(h_1), \dots, \hat{\gamma}_n(h_r)\}$  y  $\gamma(\boldsymbol{\theta}) = \{\gamma(h_1, \boldsymbol{\theta}), \dots, \gamma(h_r, \boldsymbol{\theta})\}$ . Aquí lo complicado es calcular  $Cov_{\boldsymbol{\theta}}(\hat{\gamma}_n)$ . Bajo el supuesto de que  $Y(s)$  es un *CEGE*, entonces  $Var_{\boldsymbol{\theta}}(\hat{\gamma}_n(h_k)) \approx 2\gamma^2(h_k, \boldsymbol{\theta})/N_{h_k}$ , utilizando esta aproximación se tiene el método de mínimos cuadrados ponderados:

$$\boldsymbol{\theta}_{wls} = \min_{\boldsymbol{\theta} \in \Theta} \sum_{k=1}^r \frac{N_{h_k}}{\hat{\gamma}_n^2(h_k, \boldsymbol{\theta})} (\hat{\gamma}_n(h_k) - \gamma(h_k, \boldsymbol{\theta}))^2.$$

Otras formas de estimar  $\boldsymbol{\theta}$ , sin utilizar  $\hat{\gamma}_n(h)$ , es con máxima verosimilitud (*ml*) o máxima verosimilitud restringida (*reml*) (Cressie, 2015).

Dado que la distribución de  $\mathbf{Y}(s)$ , está dada por

$$\mathbf{Y}(s) \sim \mathcal{NM}(\mu \mathbf{1}, \sigma^2 \mathbf{R}(\phi) + \tau^2 \mathbf{I}), \quad (2.14)$$

la verosimilitud de (2.14) está dada por

$$l_{ml}(\mu, \boldsymbol{\theta}) \propto -\frac{1}{2} \left\{ \log |\sigma^2 \mathbf{R}(\phi) + \tau^2 \mathbf{I}| + (\mathbf{y} - \mu \mathbf{1})^t (\sigma^2 \mathbf{R}(\phi) + \tau^2 \mathbf{I})^{-1} (\mathbf{y} - \mu \mathbf{1}) \right\}. \quad (2.15)$$

## 2.4. Estimación del semivariograma paramétrico

---

Maximizando numéricamente a (2.15) (Mardia y Marshall, 1984), se obtiene  $\hat{\boldsymbol{\theta}}$  un estimador de  $\boldsymbol{\theta}$ . Luego el modelo para el semivariograma estaría dado por  $\gamma(h, \hat{\boldsymbol{\theta}})$ .

Nótese que cuando se estima  $\boldsymbol{\theta}$  con el método de *ml*, se está estimando al mismo tiempo  $\mu$ , esto trae como consecuencia una estimación sesgada de  $\boldsymbol{\theta}$  (Cressie, 1993). Para solucionarlo se utiliza el método de máxima verosimilitud restringida (*reml*) (Patterson y Thompson, 1971).

En este caso, para estimar  $\boldsymbol{\theta}$  se plantea una transformación lineal de los datos con  $\mathbf{Y}^* = \mathbf{A}\mathbf{Y}$ , de tal forma que la distribución de  $\mathbf{Y}^*$  no dependa de  $\mu$ . Aunque se desconozca  $\mu$  y  $\boldsymbol{\theta}$  se puede encontrar una representación para  $\mathbf{A}$ , por ejemplo una proyección de mínimos cuadrados ordinarios de los residuales,

$$\mathbf{A} = \mathbf{I} - \mathbf{1}(\mathbf{1}^t\mathbf{1})^{-1}\mathbf{1}^t = \mathbf{I} - \mathbf{1}_{n \times n}$$

El principio de *reml* consiste en estimar  $\boldsymbol{\theta}$  aplicando *ml* a  $\mathbf{Y}^*$ .

Se define  $\mathbf{V} := \text{Var}(\mathbf{Y}) = \mathbf{R}(\boldsymbol{\phi}) + \nu^2\mathbf{I}$ , donde  $\nu^2 = \tau^2/\sigma^2$ . La función de verosimilitud a maximizar está dada por

$$l_{reml}(\nu^2, \boldsymbol{\phi}) \propto -\frac{1}{2} \left\{ \log |\sigma^2(\mathbf{V})| + \log \left| \mathbf{1}^t (\sigma^2(\mathbf{V}))^{-1} \mathbf{1} \right| \right\} - \frac{1}{2} \left\{ (\mathbf{y} - \mathbf{1}\hat{\boldsymbol{\mu}}(\mathbf{V}))^t (\sigma^2(\mathbf{V}))^{-1} (\mathbf{y} - \mathbf{1}\hat{\boldsymbol{\mu}}(\mathbf{V})) \right\}, \quad (2.16)$$

donde

$$\hat{\sigma}^2(\mathbf{V}) = n^{-1}(\mathbf{y} - \mathbf{1}\hat{\boldsymbol{\mu}}(\mathbf{V}))^t \mathbf{V}^{-1} (\mathbf{y} - \mathbf{1}\hat{\boldsymbol{\mu}}(\mathbf{V})),$$

y

$$\hat{\boldsymbol{\mu}}(\mathbf{V}) = (\mathbf{1}^t \mathbf{V}^{-1} \mathbf{1})^{-1} \mathbf{1}^t \mathbf{V}^{-1} \mathbf{y}.$$

Maximizando (2.16), se obtiene  $\hat{\boldsymbol{\theta}}$ .

Kerry y Oliver (2007) mostraron que el método de *reml* da mejores estimaciones para  $\boldsymbol{\theta}$  que el método de *wls*, en particular para tamaños de muestras pequeñas ( $n = 50$ ).

# Capítulo 3

## Pruebas empíricas

En algunas ocasiones, para validar el ajuste de un modelo teórico a un conjunto de datos, se recurre a una prueba gráfica. Existen varios tipos de pruebas gráficas, una de ellas consiste en graficar los cuantiles muestrales versus los cuantiles del modelo teórico, con la finalidad de observar un comportamiento lineal. En otros casos la prueba gráfica consiste en generar una *envoltura gráfica* para algún gráfico empírico generado a partir de la muestra.

Las dos pruebas estadísticas que se proponen a continuación ayudan a probar si un modelo dado para el semivariograma, modela la dependencia espacial de un *CEGE*.

La primera prueba estadística que se propone se basa en el criterio de validación (*Método de Monte Carlo: remuestreo paramétrico*) para el modelo del semivariograma (Gaetan y Bleakley, 2010). La finalidad de replantear este criterio de validación en términos de una prueba de hipótesis formal es el poder calcular la probabilidad de cometer el error tipo I. La segunda prueba estadística es una adaptación al caso espacial de la teoría de pruebas gráficas que generan envolturas gráficas (Davison y Hinkley, 1997).

### 3.1. Primera prueba empírica

El *Método de Monte Carlo: remuestreo paramétrico* (Gaetan y Bleakley, 2010) es tomado en cuenta para plantear una prueba estadística que ayude a determinar el modelo apropiado para modelar la variabilidad de un proceso espacial Gaussiano estacionario.

Sea  $\mathbf{Y} = \{Y(s_1), \dots, Y(s_n)\}$  una realización del campo (2.2) con estructura de dependencia espacial dada por  $\gamma(h, \boldsymbol{\theta})$ .

### 3.1. Primera prueba empírica

---

Nótese que al ser  $\mathbf{Y}$  un conjunto de variables aleatorias entonces su dependencia espacial puede ser modelada con la función del semivariograma. Recordemos también que la distribución conjunta de  $\mathbf{Y}$  está dada por (2.4).

Ahora bien, a partir de  $\mathbf{Y}$  se quiere probar

$$H_0 : \gamma(h, \boldsymbol{\theta}) = \gamma_0(h, \boldsymbol{\theta}) \text{ vs } H_1 : \gamma(h, \boldsymbol{\theta}) \neq \gamma_0(h, \boldsymbol{\theta}),$$

donde  $\gamma_0(h, \boldsymbol{\theta})$  es una estructura de dependencia espacial conocida, la cual depende de un vector de parámetros desconocidos  $\boldsymbol{\theta}$ .

Bajo  $H_0$  y a partir de  $\mathbf{y} = \{y(s_1), \dots, y(s_n)\}$ , se calcula:  $\hat{\boldsymbol{\theta}}_0$  un estimador de  $\boldsymbol{\theta}$  y el semivariograma empírico,  $\hat{\gamma}_n(h)$ ,  $h \in \mathcal{H} = \{h_1, \dots, h_r\}$ .

En una prueba gráfica la idea es comparar  $\hat{\gamma}_n(h)$  con el probable comportamiento de  $\gamma_n(h)$  para toda  $h \in \mathcal{H}$  cuando  $H_0$  es cierta. Recordemos que  $\gamma_n(h)$  puede ser expresada por cualquiera de expresiones que se vieron en la Subsección (2.3.2).

Por bootstrap paramétrico (Davison y Hinkley, 1997) se puede obtener una envoltura gráfica para  $\hat{\gamma}_n(h)$ . A continuación se muestra el procedimiento.

A partir del modelo (2.4) con estructura de dependencia  $\gamma_0(h, \hat{\boldsymbol{\theta}}_0)$  se generan  $m$  muestras aleatorias en las localizaciones  $\{s_1, \dots, s_n\}$ . Sean  $\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(m)}$ , las  $m$  muestras aleatorias generadas, donde  $\mathbf{y}^{(j)} = \{y^{(j)}(s_1), \dots, y^{(j)}(s_n)\}$ .

Para cada muestra  $\mathbf{y}^{(j)}$  se calcula  $\hat{\gamma}_n^{(j)}(h)$ ,  $h \in \mathcal{H}$ ,  $j = 1, \dots, m$ .

Bajo  $H_0$ ,  $\gamma_n(h), \gamma_n^{(1)}(h), \gamma_n^{(2)}(h), \dots, \gamma_n^{(m)}(h)$ , son independientes e idénticamente distribuidos para algún  $h$  fijo en  $\mathcal{H}$ . Entonces

$$P \left( \gamma_n(h) < \min_{j=1, \dots, m} \gamma_n^{(j)}(h) \mid H_0 \right) \leq \frac{1}{m+1},$$

y

$$P \left( \gamma_n(h) > \max_{j=1, \dots, m} \gamma_n^{(j)}(h) \mid H_0 \right) \leq \frac{1}{m+1}.$$

Definiendo  $\hat{\gamma}_{\text{inf}}^*(h) = \min_{j=1, \dots, m} \hat{\gamma}_n^{(j)}(h)$  y  $\hat{\gamma}_{\text{sup}}^*(h) = \max_{j=1, \dots, m} \hat{\gamma}_n^{(j)}(h)$ ,  $h \in \mathcal{H}$ , los valores críticos  $\hat{\gamma}_{\text{inf}}^*(h)$  y  $\hat{\gamma}_{\text{sup}}^*(h)$  definen aproximadamente un intervalo de confianza empírico de nivel  $1 - 2/(m+1)$  para  $\hat{\gamma}_n(h)$  (Apéndice A).

Se rechaza  $H_0$  con un nivel de significancia  $2/(m+1)$  si para algún  $h \in \mathcal{H}$ ,  $\hat{\gamma}_n(h) \notin [\hat{\gamma}_{\text{inf}}^*(h), \hat{\gamma}_{\text{sup}}^*(h)]$ . No se rechaza  $H_0$  si para todo  $h \in \mathcal{H}$ ,  $\hat{\gamma}_n(h) \in [\hat{\gamma}_{\text{inf}}^*(h), \hat{\gamma}_{\text{sup}}^*(h)]$ .

Fijando a  $m = 39$  se espera tener una prueba de nivel 0.05.

### 3.1. Primera prueba empírica

---

#### 3.1.1. Nivel de la prueba empírica

Si existe al menos un  $h \in \mathcal{H} = \{h_1, \dots, h_r\}$ , tal que

$$\{\hat{\gamma}_n(h) < \hat{\gamma}_{\text{inf}}^*(h) \text{ ó } \hat{\gamma}_n(h) > \hat{\gamma}_{\text{sup}}^*(h)\},$$

entonces el semivariograma  $\gamma_0(h, \hat{\theta}_0)$  no explica la dependencia espacial en  $\mathbf{Y}$ , y por lo tanto no modela la dependencia espacial del campo (2.2).

Nótese que esta prueba de hipótesis es del tipo *intersección-uniión*, es decir,

$$H_0 : \bigcap_{k=1}^r H_{0_k} \text{ vs } H_1 : \bigcup_{k=1}^r H_{1_k},$$

en donde

$$H_{0_k} : \gamma(h_k, \boldsymbol{\theta}) = \gamma_0(h_k, \boldsymbol{\theta}) \text{ vs } H_{1_k} : \gamma(h_k, \boldsymbol{\theta}) \neq \gamma_0(h_k, \boldsymbol{\theta}).$$

Se rechaza  $H_0$  si alguna de las  $H_{0_k}$  es rechazada.

Sea

$$A_k := \{\gamma_{\text{inf}}^*(h_k) \leq \gamma_n(h_k) \leq \gamma_{\text{sup}}^*(h_k), h_k \in \mathcal{H}\},$$

entonces la región de rechazo para  $H_{0_k}$  es de la forma

$$A_k^c := \{\gamma_n(h_k) < \gamma_{\text{inf}}^*(h_k) \text{ ó } \gamma_n(h_k) > \gamma_{\text{sup}}^*(h_k), h_k \in \mathcal{H}\},$$

luego la región total de rechazo para  $H_0$  está dada por

$$A^c := \bigcup_{k=1}^r A_k^c.$$

Por otra parte, como para cada  $h_k \in \mathcal{H}$ ,

$$P(\gamma_n(h_k) < \gamma_{\text{inf}}^*(h_k)) = P(\gamma_n(h_k) > \gamma_{\text{sup}}^*(h_k)) \leq \frac{1}{m+1},$$

entonces la región de rechazo para  $H_{0_k}$  tiene una probabilidad de a lo más  $2/(m+1)$ , ya que

$$\begin{aligned} P(A_k^c) &= P(\gamma_n(h_k) < \gamma_{\text{inf}}^*(h_k) \text{ ó } \gamma_n(h_k) > \gamma_{\text{sup}}^*(h_k)) \\ &\leq P(\gamma_n(h_k) < \gamma_{\text{inf}}^*(h_k)) + P(\gamma_n(h_k) > \gamma_{\text{sup}}^*(h_k)) \\ &\leq \frac{2}{m+1}. \end{aligned}$$

### 3.1. Primera prueba empírica

---

Por otra parte, nótese que

$$P(A^c) = P\left(\bigcup_{k=1}^r A_k^c\right) = P\left(\left(\bigcap_{k=1}^r A_k\right)^c\right),$$

Caso 1: si existe una  $A_j$  tal que  $A_j \subset \bigcap_{k=1}^r A_k$ , entonces

$$P(A^c) = P(A_j^c) \leq \frac{2}{m+1},$$

es decir, se rechaza  $H_0$  con un nivel de  $2/(m+1)$ .

Caso 2: si  $\{A_k^c, k = 1, \dots, r\}$  son eventos no necesariamente disjuntos, es decir,  $A_i^c \cap A_j^c \neq \emptyset$ , entonces

$$\begin{aligned} P(A^c) &= P\left(\bigcup_{k=1}^r A_k^c\right) \\ &= \sum_{k=1}^r P(A_k^c) - \sum_{i < j=2}^r P(A_i^c \cap A_j^c) + \dots + (-1)^{r+1} P(A_1 \cap \dots \cap A_r) \leq \frac{2r}{m+1}. \end{aligned}$$

Aquí se rechaza  $H_0$  con nivel del  $2r/(m+1)$ . En este caso si  $r = 10$  y  $m = 39$ , entonces se rechaza  $H_0$  con nivel del 0.5. Para este caso la prueba empírica, no conservaría un nivel de  $\alpha = 0.05$ .

#### 3.1.2. Estudio de simulación

Dado  $\mathbf{Y}(s) = \{Y(s_1), \dots, Y(s_n)\}$ , una realización del campo (2.2), se quiere probar que éste tiene estructura de dependencia exponencial. Para ello nos apoyamos en la distribución conjunta de  $\mathbf{Y}(s)$ , que como ya sabemos, está dada por

$$\mathbf{Y}(s) \sim \mathcal{NM}(\mu \mathbf{1}, \sigma^2 \mathbf{R}(\phi) + \tau^2 \mathbf{I}). \quad (3.1)$$

Entonces la hipótesis a probar está dada por:

$$H_0 : \gamma(h, \boldsymbol{\theta}) = \gamma_{exp}(h, \boldsymbol{\theta}) \text{ vs } H_1 : \gamma(h, \boldsymbol{\theta}) \neq \gamma_{exp}(h, \boldsymbol{\theta}).$$

Algoritmo para probar  $H_0$ :

1. Se calcula  $\hat{\gamma}_n(h)$ ,  $h \in \mathcal{H} = \{h_1, \dots, h_r\}$ , a partir de la muestra disponible  $\mathbf{y} = \{y(s_1), \dots, y(s_n)\}$ .
2. Bajo  $H_0$  y a partir de  $\mathbf{y}$ , se calcula  $\hat{\boldsymbol{\theta}}_0$  un estimador de  $\boldsymbol{\theta}$ .

### 3.1. Primera prueba empírica

---

3. A partir del modelo (3.1) con  $\hat{\boldsymbol{\theta}}_0$ , se generan  $m$  muestras aleatorias,  $\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(m)}$ , donde  $\mathbf{y}^{(j)} = \{y^{(j)}(s_1), \dots, y^{(j)}(s_n)\}$ . Para cada muestra generada  $j = 1, \dots, m$ , se calcula  $\hat{\gamma}_n^{(j)}(h)$ ,  $h \in \mathcal{H}$ .
4. Se calcula  $\hat{\gamma}_{\text{inf}}^*(h) = \min_{j=1, \dots, m} \hat{\gamma}_n^{(j)}(h)$ ,  $h \in \mathcal{H}$ .
5. Se calcula  $\hat{\gamma}_{\text{sup}}^*(h) = \max_{j=1, \dots, m} \hat{\gamma}_n^{(j)}(h)$ ,  $h \in \mathcal{H}$ .
6. Se rechaza  $H_0$  con un nivel de prueba  $2/(m+1)$  si para algún  $h \in \mathcal{H}$ ,  $\hat{\gamma}_n(h) \notin [\hat{\gamma}_{\text{inf}}^*(h), \hat{\gamma}_{\text{sup}}^*(h)]$ .

Condiciones generales para la simulación:

- La simulación se realiza en mallas irregulares y regulares con diferentes tamaños de muestra,  $n$ ,  $\{50, 100, 200, 300, 500\}$ .
- Las muestras son simuladas bajo cuatro estructuras de dependencia espacial: exponencial, esférico, circular y cúbico.
- Los parámetros del modelo (3.1) se fijan en  $\sigma^2 = 4$ ,  $\tau^2 = 0$ ,  $\phi = 2$  y  $\mu = 0.5$ .
- El cálculo del semivariograma empírico se realiza con los métodos de [Cressie y Hawkins \(1980\)](#) y el de [Genton \(1998\)](#). La cardinalidad de  $\mathcal{H}$  se fija en  $r = 10$  ([Journel y Huijbregts \(1978\)](#)).
- La estimación de  $\boldsymbol{\theta}$  se realiza con *wls*, *ml* y *reml*, con la finalidad de conocer su efecto en la prueba estadística.
- El nivel de prueba se fija en  $\alpha = 0.05$ .
- Se generaron  $m = 39$  muestras bootstrap.
- Para la simulación Monte Carlo se fijó a  $N = 1000$ .

Los resultados del estudio de simulación para esta y las siguientes pruebas estadísticas se muestran bajo la siguiente notación: En  $H_0 : \gamma(h, \boldsymbol{\theta}) = \gamma_{\text{exp}}(h, \boldsymbol{\theta})$  se tiene el tamaño estimado de la prueba. Cuando  $H_1 : \gamma(h, \boldsymbol{\theta}) = \gamma_{\text{esf}}(h, \boldsymbol{\theta})$ ,  $H_1 : \gamma(h, \boldsymbol{\theta}) = \gamma_{\text{cir}}(h, \boldsymbol{\theta})$  o  $H_1 : \gamma(h, \boldsymbol{\theta}) = \gamma_{\text{cub}}(h, \boldsymbol{\theta})$  se tiene la potencia estimada de la prueba.

En este trabajo de investigación todos los estudios de simulación son realizados en R ([Team, 2015](#)). Para generar procesos espaciales Gaussianos, estimar  $\boldsymbol{\theta}$  y realizar el cálculo del semivariograma empírico se utiliza el Paquete *geoR* ([Ribeiro Jr et al., 2015](#)).

### 3.1. Primera prueba empírica

#### 3.1.3. Resultados

El Cuadro 3.1 contiene los resultados para el nivel y la potencia para la prueba empírica en una malla regular. De los tres métodos que se utilizan para estimar  $\theta$ , se observa que el método *ml* genera resultados favorables para la prueba. Ya que para  $n = 200$  la potencia empieza a crecer, además de conservar el nivel.

El hecho de que la prueba no conserve el nivel con el método de *wls*, puede ser consecuencia de lo que comentan, Kerry y Oliver (2007), que este no es bueno para estimar  $\theta$ .

**Cuadro 3.1:** Tamaño y potencia estimados de la prueba. Estudio de simulación realizado con el modelo (3.1) en una malla regular de  $(0, 10) \times (0, 10)$ , con  $\sigma^2 = 4$ ,  $\phi = 2$ ,  $\tau^2 = 0$ ,  $\mu = 0.5$  y semivariograma empírico (2.10).

Hipótesis	Tamaño de la muestra				
	$n = 50$	$n = 100$	$n = 200$	$n = 300$	$n = 500$
$H_0 : \gamma(h, \theta) = \gamma_{exp}(h, \theta)$					
wls	0.007	0.005	0.118	0.095	0.181
ml	0.048	0.036	0.031	0.023	0.033
reml	0.025	0.013	0.018	0.011	0.016
$H_1 : \gamma(h, \theta) = \gamma_{esf}(h, \theta)$					
wls	0.034	0.026	0.806	0.685	0.699
ml	0.054	0.025	0.074	0.181	0.284
reml	0.067	0.013	0.033	0.112	0.172
$H_1 : \gamma(h, \theta) = \gamma_{cir}(h, \theta)$					
wls	0.027	0.048	0.699	0.526	0.686
ml	0.057	0.042	0.130	0.256	0.356
reml	0.048	0.021	0.078	0.159	0.227
$H_1 : \gamma(h, \theta) = \gamma_{cub}(h, \theta)$					
wls	0.031	0.009	0.905	0.808	0.879
ml	0.059	0.020	0.113	0.650	0.989
reml	0.045	0.017	0.055	0.532	0.980

El Cuadro 3.2 contiene los resultados para el nivel y la potencia de la prueba empírica bajo una malla irregular. Aquí los mejores resultados para la prueba se obtienen cuando se estima  $\theta$  con *reml*. A partir de  $n = 100$  la prueba tiene nivel y potencia.

El hecho de que la prueba no tenga nivel para  $n = 50$ , se puede deber a que el tamaño de la muestra es pequeño.

Por otra parte, también se observa que la prueba no conserva el nivel cuando se estima  $\theta$  con *wls*, esto puede deberse a que este criterio toma en cuenta la estimación empírica del semivariograma.

### 3.1. Primera prueba empírica

Los resultados de los estudios de simulación para la prueba empírica en una malla irregular son importantes porque en la realidad generalmente los fenómenos espaciales son observados en mallas irregulares.

**Cuadro 3.2:** Tamaño y potencia estimados de la prueba. Estudio de simulación con el modelo (3.1) en una malla irregular de  $(0, 10) \times (0, 10)$ , con  $\sigma^2 = 4$ ,  $\phi = 2$ ,  $\tau^2 = 0$ ,  $\mu = 0.5$  y semivariograma empírico (2.10).

Hipótesis	Tamaño de la muestra				
	$n = 50$	$n = 100$	$n = 200$	$n = 300$	$n = 500$
$H_0 : \gamma(h, \boldsymbol{\theta}) = \gamma_{exp}(h, \boldsymbol{\theta})$					
wls	0.095	0.088	0.925	0.979	1.0
ml	0.210	0.114	0.099	0.083	0.074
reml	0.147	0.061	0.064	0.061	0.052
$H_1 : \gamma(h, \boldsymbol{\theta}) = \gamma_{esf}(h, \boldsymbol{\theta})$					
wls	0.179	0.456	1.0	1.0	1.0
ml	0.237	0.156	0.230	0.259	0.339
reml	0.203	0.117	0.181	0.190	0.235
$H_1 : \gamma(h, \boldsymbol{\theta}) = \gamma_{cir}(h, \boldsymbol{\theta})$					
wls	0.188	0.427	1.0	1.0	1.0
ml	0.259	0.191	0.275	0.298	0.371
reml	0.216	0.137	0.195	0.205	0.263
$H_1 : \gamma(h, \boldsymbol{\theta}) = \gamma_{cub}(h, \boldsymbol{\theta})$					
wls	0.199	0.587	1.0	1.0	1.0
ml	0.290	0.354	0.913	0.989	1.0
reml	0.234	0.268	0.893	0.970	1.0

Para identificar si los resultados anteriores dependen del método que se ha utilizado para calcular  $\gamma_n(h)$ , se realizará nuevamente el estudio de simulación, pero ahora utilizando el estimador que propone [Genton \(1998\)](#).

En el Cuadro 3.3 se tienen los resultados del nivel y la potencia de la prueba empírica bajo una malla regular. Los resultados de este estudio muestran que la prueba tiene potencia a partir de  $n = 200$  además de conservar el nivel, cuando se estima  $\boldsymbol{\theta}$  con *ml*.

Nótese que estos resultados son similares a los obtenidos del Cuadro (3.1). Esto puede indicar que el comportamiento de la prueba es independiente del criterio que se elija para estimar  $\gamma_n(h)$ , en este caso el método de [Genton \(1998\)](#) y el de [Cressie y Hawkins \(1980\)](#).

### 3.1. Primera prueba empírica

**Cuadro 3.3:** Tamaño y potencia estimados de la prueba. Estudio de simulación con el modelo (3.1) en una malla regular de  $(0, 10) \times (0, 10)$ , con  $\sigma^2 = 4$ ,  $\phi = 2$ ,  $\tau^2 = 0$  y  $\mu = 0.5$  y semivariograma empírico (2.13).

Hipótesis	Tamaño de la muestra				
	$n = 50$	$n = 100$	$n = 200$	$n = 300$	$n = 500$
$H_0 : \gamma(h, \boldsymbol{\theta}) = \gamma_{exp}(h, \boldsymbol{\theta})$					
wls	0.009	0.013	0.019	0.136	0.204
ml	0.052	0.036	0.027	0.032	0.039
reml	0.043	0.021	0.016	0.022	0.019
$H_1 : \gamma(h, \boldsymbol{\theta}) = \gamma_{esf}(h, \boldsymbol{\theta})$					
wls	0.030	0.104	0.690	0.763	0.798
ml	0.057	0.026	0.074	0.149	0.300
reml	0.048	0.014	0.040	0.097	0.203
$H_1 : \gamma(h, \boldsymbol{\theta}) = \gamma_{cir}(h, \boldsymbol{\theta})$					
wls	0.019	0.182	0.495	0.763	0.835
ml	0.052	0.036	0.162	0.149	0.367
reml	0.045	0.020	0.098	0.097	0.245
$H_1 : \gamma(h, \boldsymbol{\theta}) = \gamma_{cub}(h, \boldsymbol{\theta})$					
wls	0.034	0.040	0.753	0.944	0.995
ml	0.061	0.018	0.095	0.651	0.995
reml	0.046	0.011	0.050	0.538	0.991

El Cuadro 3.4 contiene los resultados del nivel y la potencia estimados para la prueba empírica pero ahora bajo una malla irregular. La estimación de  $\gamma_n(h)$  se realizó también con el método de [Genton \(1998\)](#). Aquí la prueba tiene nivel a partir de  $n = 100$ , ya sea que se estime  $\boldsymbol{\theta}$  con *ml* o *reml*. Nótese también que la prueba tiene potencia.

### 3.1. Primera prueba empírica

**Cuadro 3.4:** Tamaño y potencia estimados para la prueba. Estudio de simulación con el modelo (3.1) en una malla irregular de  $(0, 10) \times (0, 10)$ , con  $\sigma^2 = 4$ ,  $\phi = 2$ ,  $\tau^2 = 0$ ,  $\mu = 0.5$  y semivariograma empírico (2.13).

Hipótesis	Tamaño de la muestra				
	$n = 50$	$n = 100$	$n = 200$	$n = 300$	$n = 500$
$H_0 : \gamma(h, \boldsymbol{\theta}) = \gamma_{exp}(h, \boldsymbol{\theta})$					
wls	0.220	0.064	0.124	0.150	0.223
ml	0.324	0.084	0.064	0.045	0.056
reml	0.218	0.064	0.036	0.029	0.030
$H_1 : \gamma(h, \boldsymbol{\theta}) = \gamma_{esf}(h, \boldsymbol{\theta})$					
wls	0.404	0.502	0.562	0.583	0.695
ml	0.358	0.176	0.208	0.206	0.287
reml	0.286	0.126	0.156	0.136	0.211
$H_1 : \gamma(h, \boldsymbol{\theta}) = \gamma_{cir}(h, \boldsymbol{\theta})$					
wls	0.390	0.416	0.482	0.592	0.728
ml	0.368	0.204	0.256	0.248	0.341
reml	0.310	0.150	0.186	0.169	0.241
$H_1 : \gamma(h, \boldsymbol{\theta}) = \gamma_{cub}(h, \boldsymbol{\theta})$					
wls	0.488	0.740	0.918	0.950	0.996
ml	0.428	0.434	0.758	0.933	1.0
reml	0.366	0.392	0.704	0.898	1.0

De los resultados anteriores se puede concluir lo siguiente. El nivel y la potencia estimados de la prueba se ven afectados en general por dos cuestiones: *i*) el tipo de malla que se utiliza para el estudio de simulación, en este caso malla regular y malla irregular, *ii*) por el método que se utiliza para la estimación de  $\boldsymbol{\theta}$ .

Los resultados para la prueba en malla irregular son de interés, ya que es en la malla irregular cuando la prueba tiene nivel, y generalmente en Geoestadística los fenómenos espaciales se presentan en mallas irregulares. En cuanto a los tres métodos utilizados para estimar  $\boldsymbol{\theta}$ , la prueba tiene nivel cuando se emplea *reml* o *ml*, criterios que reporta la literatura como los más indicados para estimar  $\boldsymbol{\theta}$  no así el método de *wls*.

Por otra parte, las conclusiones en cuanto al nivel y la potencia estimados de la prueba no se ven afectadas cuando se cambia del método de Cressie y Hawkins (1980) al método de Genton (1998) para estimar  $\gamma_n(h)$ .

De los resultados anteriores, también se puede observar que la potencia de la prueba empírica es mayor cuando en la hipótesis alternativa se tiene un modelo cúbico que cuando se tiene el modelo esférico o circular.

Con la finalidad de seguir estudiando la prueba empírica, se realiza otro estudio de simulación fijando a  $\sigma^2 = 4$ ,  $\phi = 2$ ,  $\tau^2 = 0.5$  y  $\mu = 0.5$ . En el Cuadro 3.5 se muestran los

## 3.2. Segunda prueba estadística

resultados de la simulación.

**Cuadro 3.5:** Tamaño y potencia estimados de la prueba. Estudio de simulación con el modelo (3.1) en una malla irregular de  $(0, 10) \times (0, 10)$ , con  $\sigma^2 = 4$ ,  $\phi = 2$ ,  $\tau^2 = 0.5$ ,  $\mu = 0.5$  y semivariograma empírico (2.10).

Hipótesis	Tamaño de la muestra				
	$n = 50$	$n = 100$	$n = 200$	$n = 300$	$n = 500$
$H_0 := \gamma(h, \boldsymbol{\theta}) = \gamma_{exp}(h, \boldsymbol{\theta})$					
wls	0.103	0.045	0.257	0.479	0.831
ml	0.191	0.079	0.064	0.046	0.043
reml	0.130	0.055	0.040	0.036	0.028
$H_1 := \gamma(h, \boldsymbol{\theta}) = \gamma_{esf}(h, \boldsymbol{\theta})$					
wls	0.186	0.217	0.602	0.877	0.997
ml	0.246	0.097	0.093	0.104	0.112
reml	0.185	0.074	0.063	0.077	0.087
$H_1 := \gamma(h, \boldsymbol{\theta}) = \gamma_{cir}(h, \boldsymbol{\theta})$					
wls	0.170	0.202	0.635	0.890	0.994
ml	0.243	0.131	0.114	0.120	0.146
reml	0.175	0.087	0.083	0.076	0.101
$H_1 := \gamma(h, \boldsymbol{\theta}) = \gamma_{cub}(h, \boldsymbol{\theta})$					
wls	0.170	0.298	0.910	0.995	1.0
ml	0.266	0.156	0.263	0.336	0.492
reml	0.233	0.132	0.207	0.289	0.400

De acuerdo a este estudio de simulación, se observa (Cuadro 3.5) que a partir de  $n = 100$  la prueba tiene nivel cuando se estima  $\boldsymbol{\theta}$  con *reml* o *ml*. Bajo estas condiciones de simulación, se observa pérdida de potencia en la prueba.

## 3.2. Segunda prueba estadística

Para que la primera prueba estadística fuera de nivel  $\alpha = 0.05$  fue necesario fijar el número de muestras bootstrap,  $m$ , en 39. Como Solberg (2004) sugiere que el número mínimo de muestras para realizar remuestreo es de 500, entonces se implementa otra metodología para generar una prueba gráfica (Davison y Hinkley, 1997).

Sea  $\mathbf{Y} = \{Y(s_1), \dots, Y(s_n)\}$  una realización del campo (2.2) con estructura de dependencia  $\gamma(h, \boldsymbol{\theta})$ . A partir de ésta realización se busca determinar la forma paramétrica de  $\gamma(h, \boldsymbol{\theta})$ . Para ello se plantea el siguiente juego de hipótesis:

$$H_0 : \gamma(h, \boldsymbol{\theta}) = \gamma_0(h, \boldsymbol{\theta}) \text{ vs } H_1 : \gamma(h, \boldsymbol{\theta}) \neq \gamma_0(h, \boldsymbol{\theta}).$$

## 3.2. Segunda prueba estadística

---

Aquí  $\gamma_0(h, \boldsymbol{\theta})$  es la estructura de dependencia espacial que se quiere probar para el campo (2.2).

A continuación, por bootstrap paramétrico se obtendrá un intervalo de confianza para  $\gamma_n(h)$ .

Primero, bajo  $H_0$  y a partir  $\{y(s_1), \dots, y(s_n)\}$  se calcula  $\hat{\boldsymbol{\theta}}_0$ , un estimador de  $\boldsymbol{\theta}$  y  $\hat{\gamma}_n(h)$ ,  $h \in \mathcal{H}$ .

Tomando en cuenta la estructura de dependencia espacial  $\gamma_0(h, \hat{\boldsymbol{\theta}}_0)$  para el modelo (3.1), se generan  $m$  muestras iid en las localizaciones  $\{s_1, \dots, s_n\}$ . Sean  $\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(m)}$  las  $m$  muestras generadas, donde  $\mathbf{y}^{(j)} = \{y^{(j)}(s_1), \dots, y^{(j)}(s_n)\}$ .

Para cada muestra  $\mathbf{y}^{(j)}$  se calcula el semivariograma empírico  $\hat{\gamma}_n^{(j)}(h)$ ,  $h \in \mathcal{H}$ .

Nótese que bajo  $H_0$ ,  $\hat{\gamma}_n(h), \hat{\gamma}_n^{(1)}(h), \dots, \hat{\gamma}_n^{(m)}(h)$  son valores igualmente probables de  $\gamma_n(h)$  para  $h$  fijo. Asumiendo que  $\gamma_n(h)$  es continua, entonces

$$P\left(\gamma_n(h) < \gamma_{n(q)}^{(*)}(h) \mid H_0\right) = \frac{q}{m+1},$$

donde  $\gamma_{n(q)}^{(*)}(h)$  denota el  $q$ -ésimo valor ordenado, y

$$P\left(\gamma_n(h) > \gamma_{n(m+1-q)}^{(*)}(h) \mid H_0\right) = \frac{q}{m+1}.$$

donde  $\gamma_{n(m+1-q)}^{(*)}(h)$  denota el  $(m+1-q)$ -ésimo valor ordenado. Los valores de  $m$  y  $q$  se escogen de tal manera que,

$$\alpha = \frac{q}{m+1}.$$

Los valores  $\hat{\gamma}_{n(q)}^{(*)}(h), \hat{\gamma}_{n(m+1-q)}^{(*)}(h)$  definen un intervalo de confianza de nivel  $1 - 2\alpha$  para  $\hat{\gamma}_n(h)$ .

Como  $h \in \mathcal{H} = \{h_1, \dots, h_r\}$ , entonces se tendrá un total  $r$  parejas de valores críticos. Este conjunto de parejas definen una envoltura del nivel  $1 - 2\alpha$  para  $\hat{\gamma}_n(h)$ . Luego se rechaza  $H_0$  con un nivel de  $2\alpha$  ( $\alpha \in (0, 1)$ ) si para alguna  $h \in \mathcal{H}$ ,  $\hat{\gamma}_n(h) \notin \left[\hat{\gamma}_{n(q)}^{(*)}(h), \hat{\gamma}_{n(m+1-q)}^{(*)}(h)\right]$ . No se rechaza  $H_0$  si para todo  $h \in \mathcal{H}$ ,  $\hat{\gamma}_n(h) \in \left[\hat{\gamma}_{n(q)}^{(*)}(h), \hat{\gamma}_{n(m+1-q)}^{(*)}(h)\right]$ .

### 3.2.1. Estudio de simulación

A partir de  $\mathbf{Y}(s) = \{Y(s_1), \dots, Y(s_n)\}$ , una realización del proceso espacial (2.2) con estructura de dependencia espacial  $\gamma(h, \boldsymbol{\theta})$ , se desea que  $\gamma(h, \boldsymbol{\theta})$  tiene una estructura

## 3.2. Segunda prueba estadística

---

específica, dada por  $\gamma_0(h, \boldsymbol{\theta})$ . El juego de hipótesis está dado por:

$$H_0 : \gamma(h, \boldsymbol{\theta}) = \gamma_0(h, \boldsymbol{\theta}) \text{ vs } H_1 : \gamma(h, \boldsymbol{\theta}) \neq \gamma_0(h, \boldsymbol{\theta}).$$

Algoritmo para probar  $H_0$ :

1. Se calcula  $\hat{\gamma}_n(h)$ ,  $h \in \mathcal{H} = \{h_1, \dots, h_r\}$  a partir de la muestra disponible  $\mathbf{y} = \{y(s_1), \dots, y(s_n)\}$ .
2. Bajo  $H_0$  y a partir de la muestra  $\mathbf{y}$  se calcula  $\hat{\boldsymbol{\theta}}_0$  un estimador de  $\boldsymbol{\theta}$ .
3. Bajo  $H_0$  y con  $\hat{\boldsymbol{\theta}}_0$  se generan  $m$  muestras iid del modelo (3.1),  $\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(m)}$ , donde  $\mathbf{y}^{(j)} = \{y^{(j)}(s_1), \dots, y^{(j)}(s_n)\}$ .  
Para cada muestra,  $j = 1, \dots, m$ , se calcula  $\hat{\gamma}_n^{(j)}(h)$ ,  $h \in \mathcal{H}$ .
4. Para cada  $h \in \mathcal{H}$ , se ordenan los  $\hat{\gamma}_n^{(j)}(h)$ ,  $j = 1, \dots, m$ , es decir,

$$\hat{\gamma}_n^{(*)}(h) \leq \hat{\gamma}_n^{(*)}(h) \leq \dots \leq \hat{\gamma}_n^{(*)}(h),$$

y se seleccionan los valores críticos  $\hat{\gamma}_n^{(*)}(h)$  y  $\hat{\gamma}_n^{(*)}(h)$ , los cuales definen un intervalo de confianza del nivel  $1 - 2\alpha$  para  $\hat{\gamma}_n(h)$ .

5. Se rechaza  $H_0$  con un tamaño de prueba  $2\alpha$  ( $\alpha \in (0, 1)$ ) si para alguna  $h \in \mathcal{H}$ ,  $\hat{\gamma}_n(h) \notin [\hat{\gamma}_n^{(*)}(h), \hat{\gamma}_n^{(*)}(h)]$ .

Condiciones generales para el estudio de simulación:

- Las simulaciones se realizaron en mallas irregulares y regulares de  $(0, 10) \times (0, 10)$ , para diferentes tamaños de muestra,  $n$ ,  $\{50, 100, 200, 300, 500\}$ .
- El modelo (3.1) fue simulado bajo cuatro estructuras de dependencia: exponencial, esférico, circular y cúbico.
- Los parámetros del modelo (3.1), se fijan en  $\sigma^2 = 4$ ,  $\tau^2 = 0$ ,  $\phi = 2$  y  $\mu = 0.5$ .
- El semivariograma empírico se calcula con el método de [Cressie y Hawkins \(1980\)](#). El valor de  $r$  se fija en 10 ([Journel y Huijbregts \(1978\)](#)).
- Se estima  $\boldsymbol{\theta}$  con tres métodos, *wls*, *ml* y *reml*.
- Para que la prueba tenga un nivel de  $\alpha = 0.05$ , se fija el número de muestras bootstrap en  $m = 399$ . En cuanto al cuantil se considera a  $q = 10$ .
- Para la simulación Monte Carlo, se realizan  $N = 1000$  replicaciones.

## 3.2. Segunda prueba estadística

### 3.2.2. Resultados

El Cuadro 3.6 contiene los resultados del tamaño y la potencia estimados de la prueba empírica para una malla regular. Este estudio de simulación muestra que la prueba tiene nivel cuando se estima  $\theta$  con *reml* o *ml*. Pero a partir de  $n = 300$  es cuando la prueba tiene potencia.

**Cuadro 3.6:** Tamaño y potencia estimados para la prueba. Estudio de simulación con el modelo (3.1) en una malla regular de  $(0, 10) \times (0, 10)$ , con  $\sigma^2 = 4$ ,  $\phi = 2$ ,  $\tau^2 = 0$  y  $\mu = 0.5$ .

Hipótesis	Tamaño de la muestra				
	$n = 50$	$n = 100$	$n = 200$	$n = 300$	$n = 500$
$H_0 : \gamma(h, \theta) = \gamma_{exp}(h, \theta)$					
wls	0.004	0.002	0.100	0.101	0.192
ml	0.034	0.010	0.005	0.011	0.016
reml	0.013	0.002	0.001	0.004	0.005
$H_1 : \gamma(h, \theta) = \gamma_{esf}(h, \theta)$					
wls	0.027	0.014	0.825	0.673	0.691
ml	0.035	0.013	0.045	0.158	0.261
reml	0.029	0.007	0.031	0.108	0.154
$H_1 : \gamma(h, \theta) = \gamma_{cir}(h, \theta)$					
wls	0.023	0.037	0.723	0.562	0.706
ml	0.041	0.028	0.140	0.252	0.368
reml	0.037	0.011	0.063	0.158	0.238
$H_1 : \gamma(h, \theta) = \gamma_{cub}(h, \theta)$					
wls	0.025	0.003	0.917	0.822	0.888
ml	0.037	0.003	0.090	0.761	1.0
reml	0.039	0.003	0.053	0.602	1.0

El Cuadro 3.7 contiene los resultados del tamaño y la potencia estimados de la prueba empírica bajo una malla irregular. Aquí los mejores resultados para la prueba se tienen cuando se estima  $\theta$  con *ml*. Ya que a partir de  $n = 100$  la prueba tiene nivel y potencia. Al igual que en la primera prueba empírica, esta segunda prueba no tiene nivel cuando se estima  $\theta$  con *wls*.

### 3.2. Segunda prueba estadística

**Cuadro 3.7:** Tamaño y potencia estimados de la prueba. Estudio de simulación con el modelo (3.1) en una malla irregular de  $(0, 10) \times (0, 10)$ , con  $\sigma^2 = 4$ ,  $\phi = 2$ ,  $\tau^2 = 0$  y  $\mu = 0.5$ .

Hipótesis	Tamaño de la muestra				
	$n = 50$	$n = 100$	$n = 200$	$n = 300$	$n = 500$
$H_0 : \gamma(h, \boldsymbol{\theta}) = \gamma_{exp}(h, \boldsymbol{\theta})$					
wls	0.081	0.071	0.961	0.986	0.999
ml	0.164	0.080	0.056	0.050	0.051
reml	0.118	0.032	0.052	0.039	0.031
$H_1 : \gamma(h, \boldsymbol{\theta}) = \gamma_{esf}(h, \boldsymbol{\theta})$					
wls	0.165	0.533	1.0	1.0	1.0
ml	0.207	0.126	0.200	0.239	0.338
reml	0.165	0.097	0.151	0.180	0.245
$H_1 : \gamma(h, \boldsymbol{\theta}) = \gamma_{cir}(h, \boldsymbol{\theta})$					
wls	0.169	0.465	1.0	1.0	1.0
ml	0.225	0.169	0.244	0.302	0.383
reml	0.161	0.133	0.183	0.221	0.277
$H_1 : \gamma(h, \boldsymbol{\theta}) = \gamma_{cub}(h, \boldsymbol{\theta})$					
wls	0.186	0.630	1.0	1.0	1.0
ml	0.230	0.360	0.955	0.996	1.0
reml	0.191	0.286	0.929	0.997	1.0

El Cuadro 3.8 contiene los resultados del tamaño y la potencia estimados de la prueba empírica bajo una malla irregular con  $\sigma^2 = 4$ ,  $\phi = 2$ ,  $\tau^2 = 0.5$  y  $\mu = 0.5$ .

La prueba tiene nivel y potencia a partir de  $n = 100$ , cuando se estima  $\boldsymbol{\theta}$  con *ml*. En este caso al estimar  $\tau^2$ , la potencia estimada de la prueba disminuye en comparación con los resultados anteriores.

### 3.3. Aplicación

**Cuadro 3.8:** Tamaño y potencia estimados de la prueba empírica. Estudio de simulación con el modelo (3.1) en una malla irregular de  $(0, 10) \times (0, 10)$ , con  $\sigma^2 = 4$ ,  $\phi = 2$ ,  $\tau^2 = 0.5$  y  $\mu = 0.5$ .

Hipótesis	Tamaño de la muestra				
	$n = 50$	$n = 100$	$n = 200$	$n = 300$	$n = 500$
$H_0 := \gamma(h, \boldsymbol{\theta}) = \gamma_{exp}(h, \boldsymbol{\theta})$					
wls	0.097	0.023	0.290	0.512	0.888
ml	0.154	0.048	0.029	0.030	0.025
reml	0.109	0.021	0.015	0.022	0.019
$H_1 := \gamma(h, \boldsymbol{\theta}) = \gamma_{esf}(h, \boldsymbol{\theta})$					
wls	0.145	0.210	0.655	0.928	0.999
ml	0.182	0.085	0.087	0.082	0.110
reml	0.141	0.052	0.067	0.054	0.073
$H_1 := \gamma(h, \boldsymbol{\theta}) = \gamma_{cir}(h, \boldsymbol{\theta})$					
wls	0.148	0.201	0.661	0.926	0.999
ml	0.222	0.083	0.096	0.122	0.156
reml	0.173	0.066	0.069	0.082	0.100
$H_1 := \gamma(h, \boldsymbol{\theta}) = \gamma_{cub}(h, \boldsymbol{\theta})$					
wls	0.162	0.288	0.955	0.999	1.0
ml	0.216	0.134	0.253	0.376	0.594
reml	0.178	0.106	0.209	0.294	0.494

Para esta segunda prueba empírica, se observa el mismo comportamiento que se tuvo con la primera prueba, en cuanto nivel y potencia estimada, para las dos mallas consideradas, regular e irregular.

Con el método *ml* la prueba tiene nivel. El aumento en el número de muestras bootstrap ayudo a que se obtuvieran mejores resultados para la prueba cuando el estudio de simulación se realiza en una malla irregular.

### 3.3. Aplicación

El 26 de Abril de 1986 hubo un accidente en la plata nuclear Vladímir Ilich Lenin (a 18 km de la ciudad de Chernóbil, actual Ucrania). Con el objetivo de evaluar el riesgo de contaminación por lluvia radiactiva en Suiza, varios investigadores (Dubois y De Cort, 2003) analizaron los registros de lluvia acumulada el 8 de mayo de 1986, reportado por el servicio meteorológico de Suiza. 100 de las 467 observaciones disponibles fueron analizadas por los investigadores, con el objetivo de proponer un modelo que estimará la variabilidad espacial de los datos, y así poder predecir las 367 observaciones restantes (la base de datos se puede encontrar en *geoR* (Ribeiro Jr et al., 2015)).

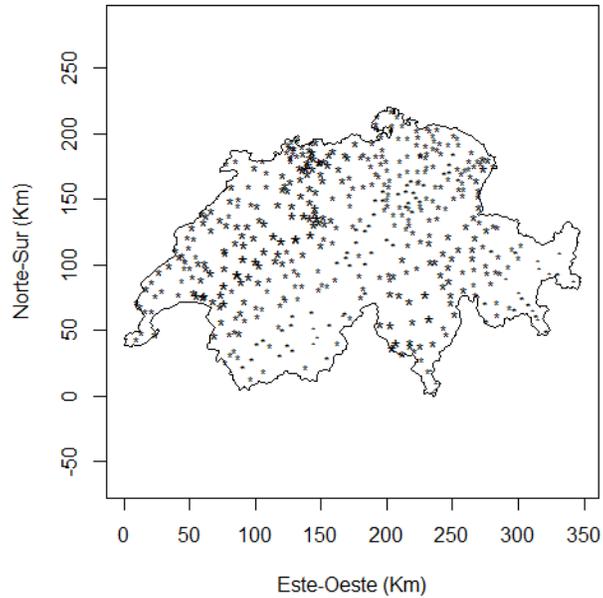
### 3.3. Aplicación

---

Nuestro objetivo en esta aplicación es analizar la variabilidad espacial de las observaciones que corresponden a lluvia radiactiva en Suiza, a partir de las mismas 100 observaciones que se dispusieron para los investigadores en ese momento.

La Figura 3.1 muestra la ubicación de las 467 estaciones meteorológicas en Suiza.

**Figura 3.1:** Mapa de Suiza con 467 estaciones.

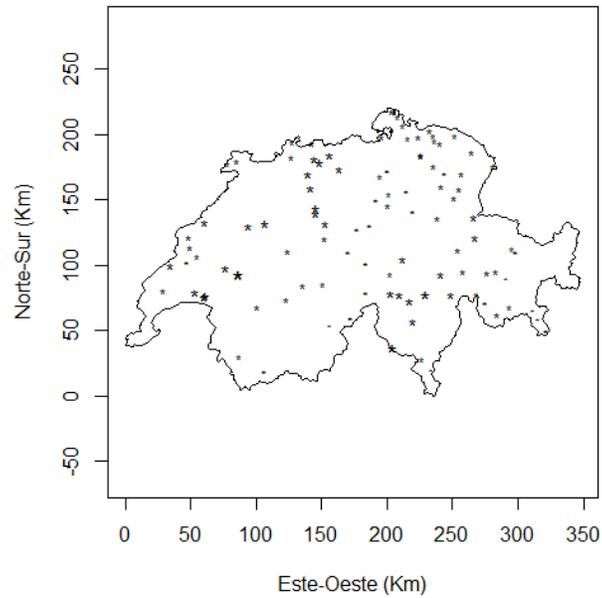


La Figura 3.2 muestra la ubicación de las 100 estaciones meteorológicas escogidas para el estudio.

### 3.3. Aplicación

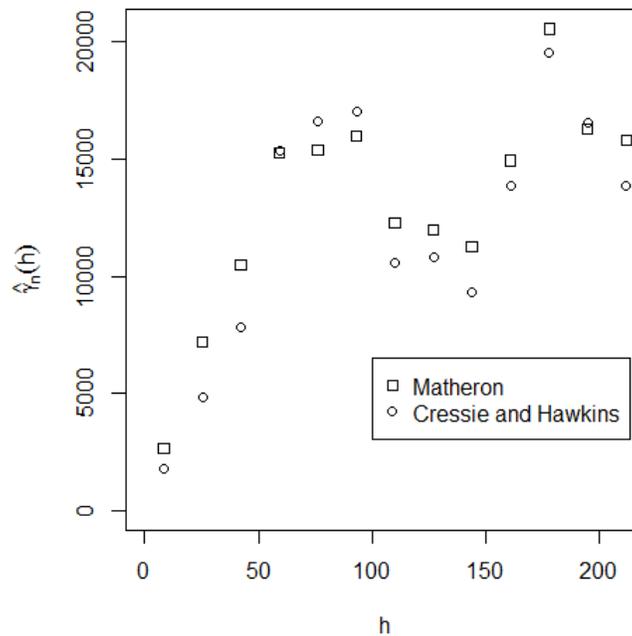
---

**Figura 3.2:** Mapa de Suiza con 100 estaciones.



La Figura 3.3 muestra los semivariogramas empíricos calculados con el método de [Matheron \(1962\)](#) y el de [Cressie y Hawkins \(1980\)](#), a partir de las 100 observaciones disponibles.

**Figura 3.3:** Semivariogramas empíricos.



### 3.3. Aplicación

Gaetan y Bleakley (2010) analizaron este conjunto de datos. Para explicar la variabilidad espacial de las observaciones, calcularon el semivariograma empírico con el método de Matheron (1962) y empleando *wls*, ajustaron el semivariograma de Matérn, definido como

$$\gamma_{mat}(h, \boldsymbol{\theta}) = \tau^2 + \sigma^2 \left[ 1 - \frac{(h/\phi)^\nu}{2^{\nu-1}\Gamma(\nu)} K_\nu(h/\phi) \right], \quad (3.2)$$

donde  $\nu > 0$ ,  $\phi > 0$  y  $K_\nu(\cdot)$  denota la función de Bessel de orden  $\nu$ . Uno de los criterios que utilizaron para validar el semivariograma paramétrico fue el *Método Monte Carlo: remuestreo paramétrico* (Gaetan y Bleakley, 2010).

Fijando a  $\tau^2$  en cero, Gaetan y Bleakley (2010) concluyen que el semivariograma Matérn con parámetros  $\hat{\phi} = 18.19$ ,  $\hat{\sigma}^2 = 15000.57$  y  $\hat{\nu} = 1$ , explica la variabilidad espacial de la muestra.

Empleando la prueba empírica propuesta en la Sección 3.1, se va analizar la variabilidad espacial de las 100 observaciones. Nótese que las 100 observaciones se encuentran dispersos en una malla irregular de  $(0, 250) \times (0, 350)$  kilómetros cuadrados. Es decir, se tienen pocos datos para un área tan grande.

El juego de hipótesis a ser probado es:

$$H_0 : \gamma(h, \boldsymbol{\theta}) = \gamma_{mat}(h, \boldsymbol{\theta}) \text{ vs } H_1 : \gamma(h, \boldsymbol{\theta}) \neq \gamma_{mat}(h, \boldsymbol{\theta}). \quad (3.3)$$

Para estimar los parámetros del semivariograma Matérn, se usan los métodos de *wls*, *ml* y *reml*. Al igual que Gaetan y Bleakley (2010), aquí se fijó a  $\tau^2$  en cero. El Cuadro 3.9 muestra las estimaciones de los parámetros y sus errores estándar.

**Cuadro 3.9:** Estimación de  $\boldsymbol{\theta}$  y su error estándar bajo  $H_0$ .

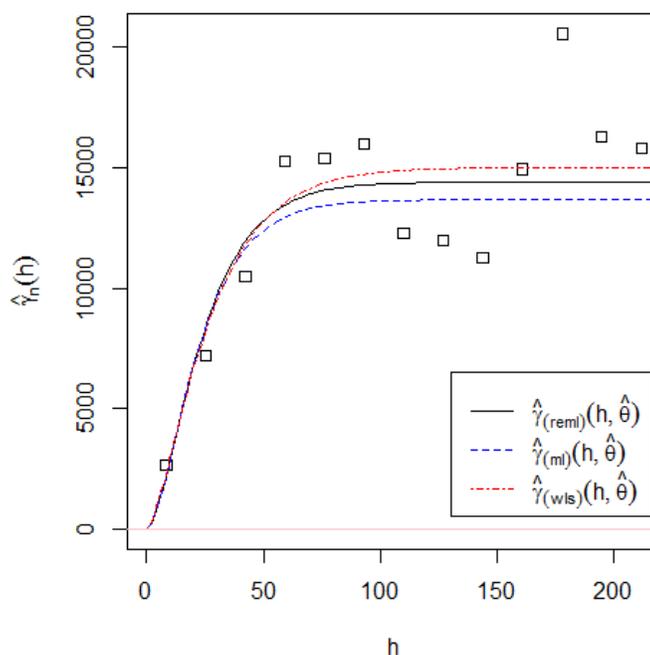
Método	$\hat{\sigma}^2$	$\hat{\phi}$	$\hat{\mu}$	$\hat{\nu}$
wls	15000.570 (3121.53)	18.1939 (6.663)	--	1 (0)
ml	13664.772 (3157.563)	13.404 (4.781)	173.025 (0)	1.306 (1.029)
reml	14396.173 (3490.667)	14.556 (6.004)	171.788 (28.002)	1.242 (0.625)

El método de *wls* no estima a  $\mu$ .

La Figura 3.4 muestra la estimación del semivariograma Matérn con el método de *wls*, *reml* y *ml*.

### 3.3. Aplicación

Figura 3.4: Estimación del semivariograma Matérn



Aunque la afirmación hecha por Gaetan y Bleakley (2010) de que el semivariograma Matérn con parámetros  $\hat{\phi} = 18.19$ ,  $\hat{\sigma}^2 = 15000.57$  y  $\hat{\nu} = 1$  explica la variabilidad espacial de las 100 observaciones parece ser correcta debemos mencionar que utilizar el método de *wls* para estimar los parámetros del semivariograma Matérn no es la mejor opción (Menezes y Febrero-Bande, 2005). Por otra parte, utilizar el método de Matheron (1962) para calcular el semivariograma empírico tampoco es una buena elección (Kerry y Oliver, 2007, Marchant y Lark, 2007, Roustant y Helbert, 2007).

La Tabla 3.10 presenta los resultados obtenidos al aplicar la prueba empírica con los tres métodos diferentes de estimación de  $\theta$  para probar  $H_0$  en (3.3) con un tamaño de prueba  $\alpha = 0.05$ . De los resultados se concluye que el semivariograma Matérn no modela la dependencia espacial de las observaciones.

Cuadro 3.10: Aplicación de la prueba empírica para probar  $H_0$  en (3.3) con  $\alpha = 0.05$ .

Método para estimar $\theta$	Decisión
wls	Rechaza $H_0$ .
ml	Rechaza $H_0$ .
reml	Rechaza $H_0$ .

Finalmente podemos concluir que la prueba empírica produce resultados adecuados cuando es usada como herramienta para escoger el modelo para el semivariograma que explique

### 3.3. Aplicación

---

correctamente la variabilidad espacial de un conjunto de datos dado.

# Capítulo 4

## Prueba basada en el criterio del *MSNE*

A partir de una realización  $\mathbf{Y}(s) = \{Y(s_1), \dots, Y(s_n)\}$  del campo (2.2), se desea modelar la estructura de dependencia espacial de este. Para ello hacemos uso del semivariograma y de la distribución conjunta de  $\mathbf{Y}(s)$ , la cual está dada por,

$$\mathbf{Y}(s) \sim \mathcal{NM}(\mu \mathbf{1}, \sigma^2 \mathbf{R}(\phi) + \tau^2 \mathbf{I}), \quad (4.1)$$

donde  $\mu$  es una constante desconocida.

Probar que el campo (2.2) tiene estructura de dependencia  $\gamma_0(h, \boldsymbol{\theta})$  es equivalente a probar que la estructura de dependencia en  $\mathbf{Y}(s)$  es  $\gamma_0(h, \boldsymbol{\theta})$ . Entonces el conjunto de hipótesis a probar está dado por

$$H_0 : \gamma(h, \boldsymbol{\theta}) = \gamma_0(h, \boldsymbol{\theta}) \text{ vs } H_1 : \gamma(h, \boldsymbol{\theta}) \neq \gamma_0(h, \boldsymbol{\theta}).$$

Para probar estas hipótesis se puede utilizar como estadístico de prueba el promedio del cuadrado del error normalizado (*MSNE*) (Gaetan y Bleakley, 2010), el cual se define como:

$$n = \sum_{i=1}^n \frac{\left(Y(s_i) - \hat{Y}(s_i)\right)^2}{\hat{\sigma}_{ok}^2(s_i)},$$

donde  $\hat{\sigma}_{ok}^2(s_i)$  es la varianza del kriging ordinario (Apéndice A) y  $\hat{Y}(s_i)$  se calcula utilizando validación cruzada y kriging ordinario. Es decir, a partir de  $Y(s_1), \dots, Y(s_n)$  primero se omite el valor  $Y(s_i)$  de la muestra y luego se calcula  $\hat{Y}(s_i)$  con las  $n - 1$  observaciones restantes utilizando kriging ordinario, esto se realiza para  $i = 1, \dots, n$ .

## 4.1. Prueba estadística modificada

---

Si se asume que los residuales normalizados

$$\frac{\left(Y(s_1) - \hat{Y}(s_1)\right)}{\hat{\sigma}_{ok}(s_1)}, \dots, \frac{\left(Y(s_n) - \hat{Y}(s_n)\right)}{\hat{\sigma}_{ok}(s_n)},$$

son variables normales independientes con varianza uno (Gaetan y Bleakley, 2010) entonces:

$$\frac{\left(Y(s_1) - \hat{Y}(s_1)\right)^2}{\hat{\sigma}_{ok}^2(s_1)}, \dots, \frac{\left(Y(s_n) - \hat{Y}(s_n)\right)^2}{\hat{\sigma}_{ok}^2(s_n)},$$

son variables independientes, cada una con distribución  $\chi_1^2$ . Luego

$$n \times MSNE \sim \chi_n^2.$$

Se rechaza  $H_0$  con un tamaño de prueba  $\alpha \in (0, 1)$  si

$$n \times MSNE \leq q(n, \alpha/2) \text{ ó } n \times MSNE \geq q(n, 1 - \alpha/2),$$

aquí  $q(n, \alpha/2)$  y  $q(n, 1 - \alpha/2)$  son los cuantiles de una distribución  $\chi_n^2$  (Bustos y Guerrero, 2011).

## 4.1. Prueba estadística modificada

Para considerar al  $MSNE$  como un estadístico de prueba es necesario asumir que las variables

$$\frac{\left(Y(s_1) - \hat{Y}(s_1)\right)}{\hat{\sigma}_{ok}(s_1)}, \dots, \frac{\left(Y(s_n) - \hat{Y}(s_n)\right)}{\hat{\sigma}_{ok}(s_n)}.$$

son independientes. Esto implica que las variables

$$\frac{\left(Y(s_1) - \hat{Y}(s_1)\right)^2}{\hat{\sigma}_{ok}^2(s_1)}, \dots, \frac{\left(Y(s_n) - \hat{Y}(s_n)\right)^2}{\hat{\sigma}_{ok}^2(s_n)},$$

sean independientes con distribución  $\chi_1^2$ .

En este trabajo de investigación se sugiere calcular por remuestreo paramétrico la distribución del estadístico de prueba,  $n \times MSNE$ . Así bajo  $H_0$  y con  $\hat{\theta}_0$  se generan  $m$  muestras del modelo (3.1). A partir de cada muestra se calcula  $n \times MSNE^{(1)}, \dots, n \times MSNE^{(m)}$ . Estos valores proporcionan una aproximación de la distribución nula de  $n \times MSNE$ .

Para definir un intervalo de confianza para  $n \times MSNE$  con coeficiente  $(1 - \alpha)100\%$ , se determinan los valores críticos  $q(\alpha/2)$  y  $q(1 - \alpha/2)$ , a partir de la distribución aproximada de

## 4.2. Estudios de simulación

---

$n \times MSNE$ , tal que

$$q_{(\alpha/2)} \leq n \times MSNE \leq q_{(1-\alpha/2)}. \quad (4.2)$$

## 4.2. Estudios de simulación

Algoritmo para probar  $H_0$ :

1. Bajo  $H_0$  y a partir  $\mathbf{y} = \{y(s_1), \dots, y(s_n)\}$ , la muestra disponible, se calcula  $\hat{\boldsymbol{\theta}}_0$  un estimador de  $\boldsymbol{\theta}$ .
2. Bajo la estructura de dependencia  $\gamma_0(h, \hat{\boldsymbol{\theta}}_0)$  se calcula  $\hat{y}(s_i)$ ,  $i = 1, \dots, n$ , utilizando kriging ordinario y validación cruzada.
3. Se calcula

$$n \times MSNE = \sum_{i=1}^n \frac{(y(s_i) - \hat{y}(s_i))^2}{\hat{\sigma}_{ok}^2(s_i)},$$

4. Se rechaza  $H_0$  con un tamaño de prueba  $\alpha \in (0, 1)$  si

$$n \times MSNE < q(n, \alpha/2) \text{ ó } n \times MSNE > q(n, 1 - \alpha/2).$$

Algoritmo para probar  $H_0$  cuando la distribución de  $n \times MSNE$  es estimada por remuestreo paramétrico:

1. Bajo  $H_0$  y a partir de la muestra disponible  $\mathbf{y} = \{y(s_1), \dots, y(s_n)\}$ , se calcula  $\hat{\boldsymbol{\theta}}_0$  un estimador de  $\boldsymbol{\theta}$ .
2. Bajo la estructura de dependencia  $\gamma_0(h, \hat{\boldsymbol{\theta}}_0)$  se calcula  $\hat{y}(s_i)$ ,  $i = 1, \dots, n$ , utilizando kriging ordinario y validación cruzada, se calcula

$$n \times MSNE = \sum_{i=1}^n \frac{(y(s_i) - \hat{y}(s_i))^2}{\hat{\sigma}_{ok}^2(s_i)},$$

3. Bajo  $H_0$ , y usando  $\hat{\boldsymbol{\theta}}_0$ ; se generan  $m$  muestras aleatorias iid  $\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(m)}$  del proceso (3.1), donde  $\mathbf{y}^{(j)} = \{y^{(*)}(s_1), \dots, y^{(*)}(s_n)\}$ . Para cada muestra generada  $j = 1, \dots, m$ , se calcula

$$n \times MSNE^{(j)} = \sum_{i=1}^n \frac{(y^{(*)}(s_i) - \hat{y}^{(*)}(s_i))^2}{\hat{\sigma}_{ok}^{(*)2}(s_i)},$$

4. Se ordenan los  $n \times MSNE^{(j)}$ ,  $j = 1, \dots, m$  y se obtienen los valores críticos  $q_{(\alpha/2)}$  y  $q_{(1-\alpha/2)}$ .

### 4.3. Resultados

---

5. Se rechaza  $H_0$  con un tamaño de prueba  $\alpha \in (0, 1)$  si

$$q_{(\alpha/2)} < n \times MSNE \text{ ó } n \times MSNE > q_{(1-\alpha/2)}.$$

Condiciones generales para el estudio de simulación:

- La simulación de las muestras del modelo (2.4) se realiza en mallas irregulares de  $(0, 10) \times (0, 10)$  con diferentes tamaños de muestra,  $n$ ,  $\{50, 100, 200, 300\}$ .
- Las muestras son simuladas bajo cuatro estructuras de dependencia: exponencial, esférica, circular y cúbica.
- Los parámetros del modelo (2.4) se fijan en  $\sigma^2 = 4$ ,  $\tau^2 = 0$ ,  $\phi = 2$  y  $\mu = 0.5$ .
- La estimación de  $\theta$  se realiza con *ml* y *reml*.
- El tamaño de prueba se fija en  $\alpha = 0.05$ .
- Para calcular la distribución de  $n \times MSNE$  se generan 120 muestras.
- Para la simulación Monte Carlo se realizaron  $N = 200$  replicaciones.
- El semivariograma empírico se calcula con el método de [Cressie y Hawkins \(1980\)](#).

### 4.3. Resultados

El Cuadro 4.1 muestra los resultados del estudio de simulación para probar  $H_0$ . Independientemente del criterio que se elija para estimar  $\theta$ , la prueba no tiene nivel, para los tamaños de muestras que se están considerando. Esta prueba tiene potencia solo cuando en la hipótesis alternativa se considera el semivariograma cúbico y para  $n$  igual a 200 y 300. Lo anterior puede ser el resultado de suponer que  $n \times MSNE^{(j)}$  tiene distribución  $\chi_n^2$ .

### 4.3. Resultados

**Cuadro 4.1:** Tamaño y potencia estimados para la prueba basada en el criterio  $MSNE$ . Estudio de simulación con el modelo (2.4) en una malla irregular de  $(0, 10) \times (0, 10)$ , con  $\sigma^2 = 4$ ,  $\phi = 2$ ,  $\tau^2 = 0$  y  $\mu = 0.5$ .

Hipótesis	Tamaño de la muestra			
	$n = 50$	$n = 100$	$n = 200$	$n = 300$
$H_0 : \gamma(h, \boldsymbol{\theta}) = \gamma_{exp}(h, \boldsymbol{\theta})$				
ml	0.0	0.0	0.0	0.0
reml	0.0	0.0	0.0	0.0
$H_1 : \gamma(h, \boldsymbol{\theta}) = \gamma_{esf}(h, \boldsymbol{\theta})$				
ml	0.0	0.0	0.015	0.035
reml	0.0	0.0	0.010	0.03
$H_1 : \gamma(h, \boldsymbol{\theta}) = \gamma_{cir}(h, \boldsymbol{\theta})$				
ml	0.0	0.005	0.040	0.035
reml	0.0	0.005	0.035	0.025
$H_1 : \gamma(h, \boldsymbol{\theta}) = \gamma_{cub}(h, \boldsymbol{\theta})$				
ml	0.0	0.06	0.990	1.0
reml	0.0	0.06	0.985	1.0

El Cuadro 4.2 muestra los resultados del estudio de simulación para la prueba estadística, cuando la distribución de  $n \times MSNE$  se estima por remuestreo paramétrico.

Conclusiones de los resultados del Cuadro 4.2:

- Independientemente del método que se elija para la estimación de  $\boldsymbol{\theta}$ , el tamaño de la prueba está cercano al valor nominal de  $\alpha = 0.05$ .
- La prueba empieza a tener potencia a partir de  $n = 100$ . Cuando en  $H_1$  se tiene el semivariograma cúbico, la potencia es mucho mayor.

## 4.4. Aplicación

**Cuadro 4.2:** Tamaño y potencia estimados de la prueba bootstrap basada en el criterio *MSNE*. Estudio de simulación con el modelo (2.4) en una malla irregular de  $(0, 10) \times (0, 10)$ , con  $\sigma^2 = 4$ ,  $\phi = 2$ ,  $\tau^2 = 0$  y  $\mu = 0.5$ .

Hipótesis	Tamaño de la muestra			
	$n = 50$	$n = 100$	$n = 200$	$n = 300$
$H_0 : \gamma(h, \boldsymbol{\theta}) = \gamma_{exp}(h, \boldsymbol{\theta})$				
ml	0.08	0.065	0.07	0.06
reml	0.045	0.055	0.04	0.06
$H_1 : \gamma(h, \boldsymbol{\theta}) = \gamma_{esf}(h, \boldsymbol{\theta})$				
ml	0.075	0.175	0.295	0.360
reml	0.07	0.150	0.290	0.365
$H_1 : \gamma(h, \boldsymbol{\theta}) = \gamma_{cir}(h, \boldsymbol{\theta})$				
ml	0.07	0.22	0.335	0.310
reml	0.04	0.22	0.425	0.310
$H_1 : \gamma(h, \boldsymbol{\theta}) = \gamma_{cub}(h, \boldsymbol{\theta})$				
ml	0.125	0.705	1.0	1.0
reml	0.155	0.66	1.0	1.0

Como se puede observar, la prueba bootstrap basada en el criterio *MSNE* tiene nivel y potencia a partir de  $n = 100$ , hecho que no sucede con la prueba basada en el criterio *MSNE*.

## 4.4. Aplicación

Para la aplicación de esta prueba se considera el ejemplo de la Sección 3.3, recordemos que el modelo que se propone para explicar la variabilidad de las 100 observaciones está dado por:

$$\gamma(h, \boldsymbol{\theta})_{mat} = \tau^2 + \sigma^2 \left[ 1 - \frac{(h/\phi)^\nu}{2^{\nu-1}\Gamma(\nu)} K_\nu(h/\phi) \right]. \quad (4.3)$$

Se fijará nuevamente el valor de  $\tau^2$  en cero y se estiman los parámetros del Modelo Matérn con *ml*. Obteniendo como resultado  $\hat{\phi} = 13.404$ ,  $\hat{\sigma}^2 = 13664.772$ ,  $\mu = 173.025$  y  $\hat{\nu} = 1.306$ .

Lo que corresponde ahora es probar el juego de hipótesis dado por

$$H_0 : \gamma(h, \boldsymbol{\theta}) = \gamma_{mat}(h, \boldsymbol{\theta}) \text{ vs } H_1 : \gamma(h, \boldsymbol{\theta}) \neq \gamma_{mat}(h, \boldsymbol{\theta}). \quad (4.4)$$

Este juego de hipótesis es probado con la prueba bootstrap basada en el criterio *MSNE*. Con un  $\alpha = 0.05$  se rechaza la hipótesis nula dada en (4.4).

Este resultado es consistente con el obtenido en la Sección 3.3.

# Capítulo 5

## Prueba de correlación muestral

La prueba de correlación muestral está dirigida a probar dependencia exponencial en el campo (2.2). La construcción de esta prueba depende de la forma paramétrica que tiene la función del semivariograma exponencial. Se le llama prueba de correlación muestral porque el estadístico de prueba es el coeficiente de correlación muestral. Cabe mencionar que la distribución del estadístico de prueba se estima por remuestreo paramétrico.

A continuación se presenta la argumentación del planteamiento de la prueba de correlación muestral.

Sea  $\mathbf{Y}(s) = \{Y(s_1), \dots, Y(s_n)\}$  una realización del campo (2.2). A partir de  $\mathbf{Y}(s)$  se desea probar el siguiente juego de hipótesis

$$H_0 : \gamma(h, \boldsymbol{\theta}) = \gamma_{exp}(h, \boldsymbol{\theta}) \text{ vs } H_1 : \gamma(h, \boldsymbol{\theta}) \neq \gamma_{exp}(h, \boldsymbol{\theta}). \quad (5.1)$$

Bajo  $H_0$  se calcula  $\hat{\boldsymbol{\theta}}_0$ , un estimador de  $\boldsymbol{\theta}$ , entonces

$$\gamma_{exp}(h_k, \hat{\boldsymbol{\theta}}_0) = \hat{\tau}^2 + \hat{\sigma}^2 \left( 1 - \exp \left( -\frac{h_k}{\hat{\phi}} \right) \right), \quad h_k \in \mathcal{H}. \quad (5.2)$$

Aproximando (5.2) con su estimador empírico, se tiene

$$\hat{\gamma}_n(h_k) \approx \gamma_{exp}(h_k, \hat{\boldsymbol{\theta}}_0) = \hat{\tau}^2 + \hat{\sigma}^2 \left( 1 - \exp \left( -\frac{h_k}{\hat{\phi}} \right) \right), \quad h_k \in \mathcal{H}. \quad (5.3)$$

Despejando  $h_k/\hat{\phi}$  en la expresión (5.3), se define

$$\hat{g}_n(h_k) := -\log \left( 1 - \frac{\hat{\gamma}_n(h_k) - \hat{\tau}^2}{\hat{\sigma}^2} \right) = \frac{h_k}{\hat{\phi}}, \quad h_k \in \mathcal{H}.$$

Es decir,  $\hat{g}_n(h_k)$  es una función lineal de  $h_k$ . Entonces para probar el conjunto de hipótesis

## 5.1. Estudio de simulación

---

en (5.1), se propone usar el coeficiente de correlación muestral de  $\hat{g}_n(h_k)$  y  $h_k$ , es decir,

$$R_{\mathbf{h},g_n(\mathbf{h})} := \widehat{Corr}(\mathbf{h}, \hat{g}_n(\mathbf{h})),$$

como una estadística de prueba.

Bajo  $H_0$ , se espera que  $R_{\mathbf{h},g_n(\mathbf{h})} \doteq 1$ .

La distribución nula de  $R_{\mathbf{h},g_n(\mathbf{h})}$  se estima por bootstrap paramétrico. Bajo  $H_0$  y dado  $\hat{\boldsymbol{\theta}}_0$  se generan  $m$  muestras iid del proceso (2.4). Para la  $j$ -ésima muestra,  $j = 1, \dots, m$  se calcula  $\hat{\boldsymbol{\theta}}_0^{(j)}$ , y  $\hat{\gamma}_n^{(j)}(h_k)$ ,  $h_k \in \mathcal{H}$  para estimar  $R_{\mathbf{h}}^{(j)}$ ,  $j = 1, \dots, m$ . Los valores  $R_{\mathbf{h},g_n(\mathbf{h})}^{(1)}, \dots, R_{\mathbf{h},g_n(\mathbf{h})}^{(m)}$  proporcionan una aproximación de la distribución nula de  $R_{\mathbf{h},g_n(\mathbf{h})}$  de la cual se obtiene el valor crítico  $\hat{c}_\alpha$ .

La prueba rechaza  $H_0$  con un tamaño de prueba  $\alpha \in (0, 1)$  si  $R_{\mathbf{h},g_n(\mathbf{h})} < \hat{c}_\alpha$ .

## 5.1. Estudio de simulación

Algoritmo para probar

$$H_0 : \gamma(h, \boldsymbol{\theta}) = \gamma_{exp}(h, \boldsymbol{\theta}) \text{ vs } H_1 : \gamma(h, \boldsymbol{\theta}) \neq \gamma_{exp}(h, \boldsymbol{\theta}).$$

1. Dado  $\mathbf{y} = \{y(s_1), \dots, y(s_n)\}$  se calcula  $\gamma_n(h_k)$ ,  $h_k \in \mathcal{H}$ .
2. Bajo  $H_0$  se calcula  $\hat{\boldsymbol{\theta}}_0$ , un estimador de  $\boldsymbol{\theta}$ .
3. Se calcula

$$R = \widehat{Corr}(h_k, \hat{g}_k(h_k)), \quad h_k \in \mathcal{H}.$$

4. Bajo  $H_0$  y considerando  $\hat{\boldsymbol{\theta}}_0$  se generan  $m$  muestras iid  $\mathbf{y}^{(1)}, \mathbf{y}^{(2)}, \dots, \mathbf{y}^{(m)}$  usando el modelo (2.4), donde  $\mathbf{y}^{(j)} = \{y^{(j)}(s_1), \dots, y^{(j)}(s_n)\}$ . Para cada muestra se calcula
  - a)  $\hat{\gamma}_n^{(j)}(h_k)$ ,  $h_k \in \mathcal{H}$ ,  $j = 1, \dots, m$ .
  - b)  $R^{(j)} = Corr(h_k, g_k(h_k))$ ,  $h_k \in \mathcal{H}$ .
5. Se ordenan los  $m$  valores  $R^{(j)}$  en forma no decreciente.
6. El valor de la constante crítica  $c(\alpha)$  es el que corresponde a la  $R^{[(\alpha m)]}$ , es decir,  $\hat{c}(\alpha) := R^{[(\alpha m)]}$ .
7. Se rechaza  $H_0$  con un tamaño de prueba  $\alpha \in (0, 1)$  si  $R < \hat{c}(\alpha)$ .

Consideraciones para la simulación:

## 5.2. Resultados

---

- La simulación de las muestras del modelo (2.4) se realiza en mallas regulares e irregulares con diferentes tamaños de muestras,  $n$ ,  $\{50, 100, 200, 300\}$ .
- Para el cálculo de  $\gamma_n(h_k)$  se utiliza en método de Cressie y Hawkins (1980),  $h_k \in \mathcal{H} = \{h_1, \dots, h_r\}$ , se fija  $r = 10$ .
- Se estima  $\theta$  con *ml* y *reml*.
- Para obtener una aproximación de la distribución nula del estadístico de prueba, se simularon  $m = 200$  muestras bootstrap.
- Para la simulación Monte Carlo se realizaron  $N = 500$  replicaciones.

## 5.2. Resultados

El Cuadro 5.1, muestra los resultados del estudio de simulación, en el caso de tener una malla regular para el campo (2.2). Como se puede observar, el tamaño estimado de la prueba converge al valor fijado  $\alpha = 0.05$  cuando el tamaño de la muestra aumenta. Este comportamiento se observa independientemente de si se estima  $\theta$  con *reml* o *ml*.

Los resultados muestran que cuando  $n = 200$  y  $n = 300$ , la potencia de la prueba es significativamente buena para las tres estructuras de dependencia que se están considerando en la hipótesis alternativa. Nótese que esta situación no ocurría con las tres pruebas estadísticas antes planteadas.

Otro punto importante es que aun cuando este estudio de simulación sea para una malla regular, el tamaño y la potencia estimados de la prueba estadística no se ven afectados como ocurría con las tres primeras pruebas estadísticas.

## 5.2. Resultados

**Cuadro 5.1:** Tamaño y potencia estimados de la prueba de correlación. Estudios de simulación con el modelo (2.4) en una malla regular de  $(0, 10) \times (0, 10)$ , con  $\sigma^2 = 4$ ,  $\phi = 2$ ,  $\tau^2 = 0$  y  $\mu = 0.5$ .

Hipótesis	Tamaño de la muestra			
	$n = 50$	$n = 100$	$n = 200$	$n = 300$
$H_0 : \gamma(h, \boldsymbol{\theta}) = \gamma_{exp}(h, \boldsymbol{\theta})$				
ml	0.116	0.116	0.066	0.088
reml	0.118	0.108	0.09	0.064
$H_1 : \gamma(h, \boldsymbol{\theta}) = \gamma_{esf}(h, \boldsymbol{\theta})$				
ml	0.116	0.740	0.678	0.704
reml	0.696	0.714	0.656	0.736
$H_1 : \gamma(h, \boldsymbol{\theta}) = \gamma_{cir}(h, \boldsymbol{\theta})$				
ml	0.612	0.674	0.600	0.656
reml	0.692	0.698	0.622	0.664
$H_1 : \gamma(h, \boldsymbol{\theta}) = \gamma_{cub}(h, \boldsymbol{\theta})$				
ml	0.720	0.804	0.734	0.828
reml	0.734	0.802	0.772	0.822

Los resultados de la Cuadro 5.2 corresponden a mallas irregulares. Se observa que el tamaño estimado de la prueba estadística oscila alrededor de 0.05 cuando  $n > 100$ . La prueba tiene potencia para los tres semivariogramas que se están considerando en la hipótesis alternativa. La potencia estimada crece conforme el tamaño de la muestra aumenta.

**Cuadro 5.2:** Tamaño y potencia estimados de la prueba de correlación. Estudio de simulación con el modelo (2.4) en una malla irregular de  $(0, 10) \times (0, 10)$ , con  $\sigma^2 = 4$ ,  $\phi = 2$ ,  $\tau^2 = 0$  y  $\mu = 0.5$ .

Hipótesis	Tamaño de la muestra			
	$n = 50$	$n = 100$	$n = 200$	$n = 300$
$H_0 : \gamma(h, \boldsymbol{\theta}) = \gamma_{exp}(h, \boldsymbol{\theta})$				
ml	0.090	0.054	0.052	0.072
reml	0.096	0.106	0.070	0.048
$H_1 : \gamma(h, \boldsymbol{\theta}) = \gamma_{esf}(h, \boldsymbol{\theta})$				
ml	0.426	0.488	0.548	0.594
reml	0.482	0.540	0.496	0.590
$H_1 : \gamma(h, \boldsymbol{\theta}) = \gamma_{cir}(h, \boldsymbol{\theta})$				
ml	0.394	0.422	0.458	0.468
reml	0.414	0.482	0.456	0.494
$H_1 : \gamma(h, \boldsymbol{\theta}) = \gamma_{cub}(h, \boldsymbol{\theta})$				
ml	0.464	0.566	0.592	0.620
reml	0.470	0.618	0.566	0.622

# Capítulo 6

## Conclusiones

En este trabajo de investigación se realizaron estudios de simulación para el estudio de cuatro pruebas estadísticas que se proponen con el objetivo de ayudar a modelar la dependencia espacial en campos espaciales Gaussianos estacionarios.

Los estudios de simulación para la prueba empírica proporcionan evidencia suficiente para concluir que el criterio de *wls* no es un buen método para estimar  $\theta$ . De estos estudios de simulación se observó también que el tamaño y la potencia de la prueba empírica, se ven directamente afectados por el tipo de malla que se esté considerando para la simulación del campo espacial Gaussiano estacionario. Por otra parte, el desempeño de la prueba empírica no se ve afectado por el método que se elija para la estimación el semivariograma empírico, en este caso el método de [Cressie y Hawkins \(1980\)](#) y el método de [Genton \(1998\)](#).

Los estudios de simulación para la prueba estadística basada en el *MSNE* como estadístico de prueba, muestran que sí es conveniente aproximar la distribución de éste por remuestreo paramétrico, ya que de esta manera la prueba tiene un tamaño estimado cercano al tamaño nominal.

Las dos pruebas empíricas y la prueba estadística que utiliza al *MSNE* como estadístico de prueba permiten probar cualquier estructura de dependencia espacial en los campos espaciales Gaussianos estacionarios. De los estudios de simulación en mallas irregulares se observó que la prueba estadística basada en el *MSNE* es más potente que las pruebas empíricas. Sin embargo, el tiempo computacional requerido es mayor para esta prueba.

En cuanto a la prueba de correlación muestral, se observó que tiene potencia contra los tres semivariogramas que se están considerando en la hipótesis alternativa. Hecho que no ocurre con las otras pruebas estadísticas que se han propuesto en este trabajo.

Esta prueba puede ser de ayuda para el investigador ya que el semivariograma exponencial es uno de los semivariogramas paramétricos que más se emplea en las aplicaciones ([Oliver](#)

## 6. Conclusiones

---

y Webster, 2014).

# Bibliografía

- Akaike, H. 1998. Information theory and an extension of the maximum likelihood principle. *Pages 199–213 of: Selected Papers of Hirotugu Akaike*. Springer.
- Banerjee, S., Carlin B.P., y Gelfand, A. E. 2014. *Hierarchical modeling and analysis for spatial data*. CRC Press.
- Barry, R.P. 1996. A diagnostic to assess the fit of a variogram model to spatial data. *Journal of Statistical Software*, **1**(1), 1–11.
- Brus, D.J., y De Gruijter, J.J. 1994. Estimation of non-ergodic variograms and their sampling variance by design-based sampling strategies. *Mathematical geology*, **26**(4), 437–454.
- Bustos, O., y Guerrero, A. 2011. Breve introducción a la Matemática de la Estadística Espacial. *Ensaïos Matemáticos*, **20**, 1–115.
- Clark, R.G., y Allingham, S. 2011. Robust resampling confidence intervals for empirical variograms. *Mathematical Geosciences*, **43**(2), 243–259.
- Cressie, N. 1985. Fitting variogram models by weighted least squares. *Journal of the International Association for Mathematical Geology*, **17**(5), 563–586.
- Cressie, N. 1993. *Statistics for spatial data*. John Wiley & Sons.
- Cressie, N. 2015. *Statistics for spatial data*. John Wiley & Sons.
- Cressie, N., y Hawkins, D.M. 1980. Robust estimation of the variogram: I. *Journal of the International Association for Mathematical Geology*, **12**(2), 115–125.
- Cressie, N., y Wikle, C.K. 2011. *Statistics for spatio-temporal data*. John Wiley & Sons.
- Davison, A.C., y Hinkley, D.V. 1997. *Bootstrap methods and their application*. Vol. 1. Cambridge University Press.
- Davison, A.C., y Kuonen, D. 2002. An Introduction to the Bootstrap with Applications in R. *Statistical Computing and Statistical Graphics Newsletter*, **13**(1), 6–11.
- Diggle, P., y Ribeiro, P.J. 2007. *Model-based geostatistics*. Springer Science & Business Media.

## Bibliografía

---

- Diggle, P.J., Ribeiro Jr P.J., y Christensen, O.F. 2003. An introduction to model-based geostatistics. *Pages 43–86 of: Spatial statistics and computational methods*. Springer New York.
- Dowd, P.A. 1984. The variogram and kriging: robust and resistant estimators. *Pages 91–106 of: Geostatistics for natural resources characterization*. Springer Netherlands.
- Dubois, G., Malczewski J., y De Cort, M. 2003. *Mapping radioactivity in the environment: spatial interpolation comparison 97*. Office for Official Publications of the European Communities.
- Efron, B. 1979. Bootstrap methods: another look at the Jackknife. *The annals of Statistics*, 1–26.
- Gaetan, C., Guyon X., y Bleakley, K. 2010. *Spatial statistics and modeling*. Vol. 271. New York: Springer.
- Genton, M.G. 1998. Highly robust variogram estimation. *Mathematical Geology*, **30**(2), 213–221.
- Glatzer, E., y Müller, W.G. 2004. Residual diagnostics for variogram fitting. *Computers & Geosciences*, **30**(8), 859–866.
- Gregori, P., Porcu E., y Mateu, J. 2014. Models of covariance functions of Gaussian random fields escaping from isotropy, stationarity and non negativity. *Image Analysis & Stereology*, **33**(1), 75–81.
- Grigoriu, M. 2013. *Stochastic calculus: applications in science and engineering*. Springer Science & Business Media.
- Journel, A.G., y Huijbregts, C.J. 1978. *Mining geostatistics*. Academic press.
- Kent, J.T., y Mohammadzadeh, M. 1999. Spectral approximation to the likelihood for an intrinsic Gaussian random field. *Journal of multivariate analysis*, **70**(1), 136–155.
- Kerry, R., y Oliver, M.A. 2007. Comparing sampling needs for variograms of soil properties computed by the method of moments and residual maximum likelihood. *Geoderma*, **140**(4), 383–396.
- Lark, R.M. 2000. A comparison of some robust estimators of the variogram for use in soil survey. *European Journal of Soil Science*, **51**(1), 137–157.
- Marchant, B.P., y Lark, R.M. 2007. Robust estimation of the variogram by residual maximum likelihood. *Geoderma*, **140**(1), 62–72.
- Mardia, K.V., y Marshall, R.J. 1984. Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika*, **71**(1), 135–146.
- Matheron, G. 1962. *Traité de géostatistique appliquée, tome i: Mémoires du bureau de recherches géologiques et minières*. *Pairs: Editions Technip*, **14**(6).

## Bibliografía

---

- Menezes, R., Garcia-Soidán P., y Febrero-Bande, M. 2005. A comparison of approaches for valid variogram achievement. *Computational statistics*, **20**(4), 623–642.
- Müller, W., y Zimmerman, D.L. 1997. Optimal design for variogram estimation.
- Oliver, M.A., y Webster, R. 2014. A tutorial guide to geostatistics: Computing and modelling variograms and kriging. *Catena*, **113**, 56–69.
- Patterson, H.D., y Thompson, R. 1971. Recovery of inter-block information when block sizes are unequal. *Biometrika*, **58**(3), 545–554.
- Ribeiro Jr, Paulo J, Diggle, Peter J, y Ribeiro Jr, Maintainer Paulo J. 2015. Package geoR.
- Ross, S.M. 1996. *Stochastic processes*. Vol. 2. New York: John Wiley & Sons.
- Rousseeuw, P.J., y Croux, C. 1992. Explicit scale estimators with high breakdown point. *L1-Statistical analysis and related methods*, **1**, 77–92.
- Roustant, O., Dupuy D., y Helbert, C. 2007. Robust estimation of the variogram in computer experiments. *Rapport interne, Département 3MI, Ecole des Mines de Saint-Etienne*.
- Solanas, Antoni, y Olivera, Vicenta. 1992. Bootstrap: fundamentos e introducción a sus aplicaciones. *Anuario de psicología The UB Journal of psychology*, 143–154.
- Solberg, Helge Erik. 2004. The IFCC recommendation on estimation of reference intervals. The RefVal program. *Clinical Chemistry and Laboratory Medicine*, **42**(7), 710–714.
- Team, R Core. 2015. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Webster, Richard, y Oliver, Margaret A. 1992. Sample adequately to estimate variograms of soil properties. *Journal of soil science*, **43**(1), 177–192.
- Webster, Richard, y Oliver, Margaret A. 2007. *Geostatistics for environmental scientists*. John Wiley & Sons.
- Zhang, Lianjun, Gove, Jeffrey H, y Heath, Linda S. 2005. Spatial residual analysis of six modeling techniques. *Ecological Modelling*, **186**(2), 154–177.
- Zimmerman, Dale L, y Cressie, Noel. 1992. Mean squared prediction error in the spatial linear model with estimated covariance parameters. *Annals of the institute of statistical mathematics*, **44**(1), 27–43.

# Apéndice A

## Conceptos teóricos

### A.1. Probabilidad de variables aleatorias

Dadas  $Z_1, \dots, Z_{m+1}$  variables aleatorias independientes e idénticamente distribuidas, entonces

$$P(Z_1 = \min(Z_1, \dots, Z_{m+1})) \leq \frac{1}{m+1}.$$

Esto es equivalente a dadas  $Z, Z_1, \dots, Z_m$  variables aleatorias independientes e idénticamente distribuidas, entonces

$$P(Z < \min(Z_1, \dots, Z_m)) \leq \frac{1}{m+1}$$

y

$$P(Z > \max(Z_1, \dots, Z_m)) \leq \frac{1}{m+1}.$$

### A.2. Función teórica del semivariograma

Deducción de la función de semivariograma para el proceso (2.4).

Como

$$Y(s_i) = S(s_i) + Z(s_i), \quad i = 1, \dots, n, \quad s_i \in D,$$

entonces

$$\begin{aligned} E(Y(s_i)) &= \mu, \\ \text{Var}(Y(s_i)) &= \sigma^2 + \tau^2. \end{aligned}$$

## A.2. Función teórica del semivariograma

---

De donde

$$E(Y(s_i) - Y(s_j)) = 0.$$

Como todos los campos Gaussianos estacionarios (estrictamente o débilmente) son intrínsecamente estacionarios, entonces se puede trabajar con la varianza de las diferencias de las variables aleatorias, es decir,

$$\begin{aligned} \text{Var}(Y(s_i) - Y(s_j)) &= E \{ \{ [Y(s_i) - Y(s_j)] - E[Y(s_i) - Y(s_j)] \}^2 \} \\ &= E \{ \{ Y(s_i) - Y(s_j) \}^2 \} \\ &= \text{Var}(Y(s_i)) + \text{Var}(Y(s_j)) - 2\text{Cov}(Y(s_i), Y(s_j)), \quad s_i, s_j \in D, \end{aligned}$$

nótese que

$$\begin{aligned} \text{Cov}(Y(s_i), Y(s_j)) &= E \{ [Y(s_i) - E(Y(s_i))] [Y(s_j) - E(Y(s_j))] \} \\ &= E \{ [S(s_i) + Z(s_i) - E(Y(s_i))] [S(s_j) + Z(s_j) - E(Y(s_j))] \} \\ &= E \{ [S(s_i) + Z(s_i) - \mu] [S(s_j) + Z(s_j) - \mu] \} \\ &= E \{ [(S(s_i) - \mu) + Z(s_i)] [(S(s_j) - \mu) + Z(s_j)] \}, \end{aligned}$$

entonces

$$\begin{aligned} \text{Cov}(Y(s_i), Y(s_j)) &= E \{ (S(s_i) - \mu) (S(s_j) - \mu) \} \\ &\quad + E \{ Z(s_i) (S(s_j) - \mu) + Z(s_j) (S(s_i) - \mu) + Z(s_i) Z(s_j) \}, \end{aligned}$$

como los  $Z(s_i)$  son independientes e idénticamente distribuidos, entonces

$$\text{Cov}(Y(s_i), Y(s_j)) = \text{Cov}(S(s_i), S(s_j)).$$

Entonces

$$\begin{aligned} \rho_Y(h_{ij}) &= \rho(\|s_i - s_j\|) \\ &= \text{Corr}(Y(s_i), Y(s_j)) \\ &= \frac{\text{Cov}(Y(s_i), Y(s_j))}{\sqrt{\text{Var}(Y(s_i))\text{Var}(Y(s_j))}} \\ &= \frac{\text{Cov}(S(s_i), S(s_j))}{\sqrt{\text{Var}(Y(s_i))\text{Var}(Y(s_j))}} \\ &= \frac{\text{Cov}(S(s_i), S(s_j))}{\sigma^2 + \tau^2}. \end{aligned}$$

Por lo tanto, la varianza de la diferencia de dos variables está dada por

$$\text{Var}(Y(s_i) - Y(s_j)) = 2(\tau^2 + \sigma^2) - 2 \left( \frac{\text{Cov}(S(s_i), S(s_j))}{\sigma^2 + \tau^2} \right),$$

### A.3. Kriging ordinario

---

luego

$$2\gamma(h_{ij}) = 2(\tau^2 + \sigma^2) - 2 \left( \frac{Cov(S(s_i), S(s_j))}{\sigma^2 + \tau^2} \right),$$

simplificando

$$\gamma(h_{ij}) = \tau^2 + \sigma^2 - \left( \frac{Cov(S(s_i), S(s_j))}{\sigma^2 + \tau^2} \right).$$

Como

$$\begin{aligned} Corr(S(s_i) - S(s_j)) &= \frac{Cov(S(s_i) - S(s_j))}{\sqrt{Var(S(s_i))Var(S(s_j))}} \\ &= \frac{Cov(S(s_i) - S(s_j))}{\sigma^2}, \end{aligned}$$

entonces

$$\gamma(h_{ij}) = \tau^2 + \sigma^2 - \left( \frac{\sigma^2 Corr(S(s_i), S(s_j))}{\sigma^2 + \tau^2} \right),$$

simplificando

$$\gamma(h_{ij}) = \tau^2 + \sigma^2 - \frac{\sigma^2 \rho_S(h_{ij})}{\sigma^2 + \tau^2},$$

donde  $\rho_S(h_{ij})$  es la correlación que existe entre  $S(s_i)$  y  $S(s_j)$ .

La función de semivariograma que corresponde al proceso (2.4), está dado por

$$\gamma(h) = \tau^2 + \sigma^2 \left( 1 - \frac{\rho_S(h)}{\sigma^2 + \tau^2} \right).$$

### A.3. Kriging ordinario

Sea  $Y(s)$  la versión ruidosa del campo  $\{S(s) : s \in D \subset \mathbb{R}^2\}$ . Supongamos que  $Y(s)$  puede ser modelado como la suma de un componente sistemático y un componente de residuo espacial, es decir,

$$Y(s) = \mu(s) + Z(s); \quad s \in D. \quad (\text{A.1})$$

Dependiendo de las características de  $\mu(s)$ , se tienen variantes del kriging, por ejemplo:

- Kriging simple si:  $\mu(s) = \mu, \forall s \in D \subset \mathbb{R}^2$ , aquí  $\mu$  es un valor conocido.
- Kriging ordinario si:  $\mu(s) = \mu$  es constante en una vecindad  $B$  de  $s$ , donde  $B \subset D$ , aquí  $\mu$  es un valor constante pero desconocido.

### A.3. Kriging ordinario

---

El kriging ordinario tiene como objetivo estimar  $Y$  en  $s_0$  usando los valores de la muestra  $Y(s_i)$ ,  $i = 1, \dots, n$ , que se encuentren cerca de la ubicación  $s_0$ . Esta estimación se calcula con una combinación lineal de los valores de la muestra, es decir,

$$Y^*(s_0) = \sum_{i=1}^n \lambda_i Y(s_i),$$

donde los pesos  $\lambda$  son elegidos de tal manera que el estimador  $Y^*(s_0)$  sea

1. Insesgado, es decir,  $E(Y^*(s_0) - Y(s_0)) = 0$ .
2. Minimice el *error de predicción cuadrado medio*:  $E(Y^*(s_0) - Y(s_0))^2$ .

El error de predicción cuadrado medio, también se le conoce como *varianza de la estimación* y se denota como:

$$\sigma_E^2 = \text{Var}(Y^*(s_0) - Y(s_0)).$$

La propiedad de insesgadez es garantizada por el hecho de que  $\sum_{i=1}^n \lambda_i = 1$ , entonces

$$\begin{aligned} E(Y^*(s_0) - Y(s_0)) &= E\left(\sum_{i=1}^n \lambda_i Y(s_i) - Y(s_0) \sum_{i=1}^n \lambda_i\right) \\ &= \sum_{i=1}^n \lambda_i E(Y^*(s_0) - Y(s_0)) = 0, \end{aligned}$$

luego la esperanza de los incrementos es cero, ya que se está asumiendo que se tiene estacionaridad intrínseca.

Por otra parte,

$$\begin{aligned} \sigma_E^2 &= \text{Var}(Y^*(s_0) - Y(s_0)) \\ &= E((Y^*(s_0) - Y(s_0))^2) \\ &= \text{Var}(Y^*(s_0)) + \text{Var}(Y(s_0)) - 2\text{Cov}(Y^*(s_0), Y(s_0)). \end{aligned} \quad (\text{A.2})$$

Recordando que  $\gamma(h) = \sigma^2 - \text{Cov}(h)$ , se desarrolla cada término de la expresión anterior,

$$\begin{aligned} \text{Var}(Y^*(s_0)) &= \sigma^2 - \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \text{Cov}(Y(s_i), Y(s_j)) \\ &= \sigma^2 - \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \gamma(Y(s_i), Y(s_j)). \end{aligned}$$

Por otra parte

$$\text{Var}(Y(s_0)) = \sigma^2.$$

### A.3. Kriging ordinario

---

Para el último término se tiene

$$\begin{aligned}
 2Cov(Y^*(s_0), Y(s_0)) &= 2Cov\left(\sum_{i=1}^n \lambda_i Y(s_i), Y(s_0)\right) \\
 &= 2\sum_{i=1}^n \lambda_i Cov(Y(s_i), Y(s_0)) \\
 &= 2\sum_{i=1}^n \lambda_i Cov(Y(s_i), Y(s_0)) \\
 &= 2\sum_{i=1}^n \lambda_i (\sigma^2 - \gamma(Y(s_i), Y(s_0))) \\
 &= 2\sigma^2 - 2\sum_{i=1}^n \lambda_i \gamma(Y(s_i), Y(s_0)).
 \end{aligned}$$

Sustituyendo en (A.2) se tiene

$$\begin{aligned}
 \sigma_E^2 &= Var(Y^*(s_0)) + Var(Y(s_0)) - 2Cov(Y^*(s_0), Y(s_0)) \\
 &= \sigma^2 - \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \gamma(Y(s_i), Y(s_j)) + \sigma^2 - 2\sigma^2 + 2\sum_{i=1}^n \lambda_i \gamma(Y(s_i), Y(s_0)) \\
 &= 2\sum_{i=1}^n \lambda_i \gamma(Y(s_i), Y(s_0)) - \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \gamma(Y(s_i), Y(s_j)).
 \end{aligned}$$

Lo que corresponde ahora es minimizar  $\sigma_E^2$  sujeto a la restricción  $\sum_{i=1}^n \lambda_i = 1$ . El problema de minimización con restricciones se resuelve mediante el método de multiplicadores de la Lagrange, entonces

$$\sigma_E^2 = 2\sum_{i=1}^n \lambda_i \gamma(Y(s_i), Y(s_0)) - \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \gamma(Y(s_i), Y(s_j)) - 2m_{ok} \left( \sum_{i=1}^n \lambda_i - 1 \right).$$

Siguiendo el procedimiento acostumbrado para obtener valores extremos de una función, se deriva  $\sigma_E^2$  con respecto a  $\lambda_i$ ,  $i = 1, \dots, n$  y se iguala a cero. Después se deriva  $\sigma_E^2$  con respecto a  $m_{ok}$  y se iguala a cero. Del procedimiento anterior resulta un sistema de  $n + 1$  ecuaciones con  $n + 1$  incógnitas,

$$\begin{cases} \sum_{j=1}^n \lambda_j \gamma(Y(s_i), Y(s_j)) + m_{ok} = \gamma(Y(s_i), Y(s_0)), i = 1, \dots, n \\ \sum_{i=1}^n \lambda_i = 1 \end{cases}$$

## A.4. Método de bootstrap paramétrico

---

que matricialmente se representa como

$$\begin{pmatrix} \gamma(Y(s_1), Y(s_1)) & \cdots & \gamma(Y(s_1), Y(s_n)) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \gamma(Y(s_n), Y(s_1)) & \cdots & \gamma(Y(s_n), Y(s_n)) & 1 \\ 1 & \cdots & 1 & 0 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \\ m_{ok} \end{pmatrix} = \begin{pmatrix} \gamma(Y(s_1), Y(s_0)) \\ \vdots \\ \gamma(Y(s_n), Y(s_0)) \\ 1 \end{pmatrix}$$

lo cual es equivalente a

$$\mathbf{\Gamma}\boldsymbol{\lambda} = \boldsymbol{\gamma}$$

donde  $\mathbf{\Gamma}$  es una matriz simétrica de  $(n + 1) \times (n + 1)$ .

Entonces los pesos  $\lambda_1, \dots, \lambda_n$  que minimizan el error de predicción se obtienen por medio del sistema

$$\boldsymbol{\lambda} = \mathbf{\Gamma}^{-1}\boldsymbol{\gamma}.$$

el cual involucra el semivariograma.

La minimización de la *varianza del estimador*,  $\sigma_E^2$ , es llamada *varianza del kriging ordinario* y está dada por

$$\sigma_{ok}^2(s_0) = \boldsymbol{\lambda}'\boldsymbol{\gamma} = \sum_{i=1}^n \lambda_i \gamma(Y(s_i), Y(s_0)) + m_{ok} \quad (\text{A.3})$$

## A.4. Método de bootstrap paramétrico

[Efron \(1979\)](#) introdujo el método de bootstrap. Los métodos bootstrap son técnicas de remuestreo para evaluar la incertidumbre ([Davison y Kuonen, 2002](#)).

Existen varios tipos de bootstrap. Entre ellos está el método de bootstrap paramétrico.

Cuando el investigador conoce la función de distribución de las variables aleatorias bajo estudio, aun cuando desconozca los parámetros de la función, se tiene el método conocido como bootstrap paramétrico ([Solanas y Olivera, 1992](#)).

Cuando se conoce la función de distribución  $F$ , y la familia paramétrica a la cual pertenece, sólo se estima  $\theta$  el vector de parámetros de  $F$ , a partir de la muestra dada,  $x_1, \dots, x_n$ , teniendo de esta forma  $F_{\hat{\theta}}$ . Ahora a partir de  $F_{\hat{\theta}}$  se pueden generar muestras aleatorias independientes.

# Apéndice B

## Programas en R

### B.1. Primera prueba empírica

```
rm(list=ls())
library(geoR) # Este Paquete permite utilizar las funciones
              # \emph{grf}, \emph{variog} y \emph{variofit}.

prueba<-function(n, m, model_H1, entre)
{
  ###--- Simulación del campo Gaussiano ---###

  cag <- grf(n, nsim=1, grid = "irreg",xlims=c(0,10),ylims=c(0,10),
            cov.model=model_H1, method="eigen",cov.pars = c(4, 2),
nugget = 0.5, mean=0.5)

  ###--- Estimación del semivariograma empírico ---###

  dis<-sqrt(2*10*10)/2

  uvec<-seq(0,dis,l=entre)

  v_bin<-variog(cag, nugget.tolerance=0, trend="cte", option = "bin",
                estimator.type = "modulus", uvec=uvec, pairs.min=30)

  gamma_h_k<-v_bin$v
  h_k<-v_bin$u

  ###--- Estimación de theta ---###
```

## B.1. Primera prueba empírica

---

```
sill<-var(cag$data)
phi<-mean(v_bin$v)
nugget<-0.1

model_H0<-"exp"

wls<-variofit(v_bin, ini.cov.pars=c(sill, phi), cov.model=model_H0,
             fix.nugget = FALSE, nugget=nugget, max.dist=dis,
weights="cressie")

ml<-likfit(cag, cov.model=model_H0, ini.cov.pars=c(sill, phi),
          fix.nugget =FALSE, nugget=nugget, lik.method="ML")

reml<-likfit(cag, cov.model=model_H0, ini.cov.pars=c(sill, phi),
            fix.nugget = FALSE, nugget=nugget, lik.method="REML")

wmr<-matrix(c(wls$cov.pars[1], wls$cov.pars[2],0,wls$nugget,
             ml$sigmaq, ml$phi,ml$beta,ml$tausq,
             reml$sigmaq, reml$phi,reml$beta,reml$tausq),4,3)

###--- Intervalo de confianza con supremo e ínfimo ---###

m_intervalo<- function(still_e,phi_e,beta_e,tau_e,m)
{
  rbin<-matrix(0,length(v_bin$v),m)

  m_cag<-grf(n,nsim=m,grid =cag$coords,xlims=c(0,10),
            ylims=c(0,10),cov.model=model_H0,
method="eigen",cov.pars = c(still_e,phi_e),
nugget =tau_e, mean=beta_e)

  mv_bin<- variog(m_cag, nugget.tolerance=0, trend="cte",
option = "bin",estimator.type="modulus",
uvec=uvec, pairs.min=30)

  rbin<- mv_bin$v

  return(rbin)
}

###--- Aceptación vs Rechazo ---###

L<-length(h_k)
total<-matrix(NA,1,3)
```

## B.2. Segunda prueba empírica

---

```
      for(i in 1:3)
      {
        intervalo<-m_intervalo(still_e=wmr[1,i], phi_e=wmr[2,i],
beta_e=wmr[3,i], tau_e=wmr[4,i],m)

        gmin<-apply(intervalo, 1, min)
        gmax<-apply(intervalo, 1, max)

        res<-ifelse(gmin<=gamma_h_k & gamma_h_k<=gmax,1,0)
        total[1,i]<-ifelse(sum(res)==L,0,1)
      }

    return(total)
  }

###--- Simulación Monte Carlo ---###

N<-1000
m<-39
n<-500
entre<-11
model_H1<-"exp"

r_total<-matrix(0,N,3)

for(j in 1:N)
{
  r_total[j,]<-prueba(n=n, m=m, model_H1, entre=entre)

  cat(j,"\n")
}

apply(r_total,2,sum)/N
```

## B.2. Segunda prueba empírica

```
rm(list=ls())
library(geoR) # Este Paquete permite utilizar las funciones
              # \emph{grf}, \emph{variog} y \emph{variofit}.
```

## B.2. Segunda prueba empírica

---

```
prueba<-function(n, m, model_H1, entre)
{
  ###--- Simulación del campo Gaussiano ---###

  cag <- grf(n,nsim=1, grid = "irreg",xlims=c(0,10),ylims=c(0,10),
            cov.model=model_H1, method="eigen",
            cov.pars = c(4, 2), nugget = 0.5, mean=0.5)

  ###--- Estimación del semivariograma empírico ---###

  dis<-sqrt(2*10*10)/2
  uvec<-seq(0,dis,l=entre)

  v_bin<-variog(cag, nugget.tolerance=0, trend="cte", option = "bin",
               estimator.type = "modulus", uvec=uvec, pairs.min=30)

  gamma_h_k<-v_bin$v
  h_k<-v_bin$u

  ###--- Estimación de theta ---###

  sill<-var(cag$data)
  phi<-mean(v_bin$v)
  nugget<-0.1

  model_H0<-"exp"

  wls<-variofit(v_bin, ini.cov.pars=c(sill, phi), cov.model=model_H0,
               fix.nugget = FALSE, nugget=nugget, max.dist=dis,
               weights="cressie")

  ml<-likfit(cag, cov.model=model_H0, ini.cov.pars=c(sill, phi),
            fix.nugget =FALSE, nugget=nugget, lik.method="ML")

  reml<-likfit(cag, cov.model=model_H0, ini.cov.pars=c(sill, phi),
              fix.nugget = FALSE, nugget=nugget, lik.method="REML")

  wmr<-matrix(c(wls$cov.pars[1], wls$cov.pars[2],0,wls$nugget,
               ml$sigma_sq, ml$phi, ml$beta,ml$tausq,
               reml$sigma_sq, reml$phi,reml$beta,reml$tausq),4,3)

  ###--- Intervalo de confianza con supremo e ínfimo ---###

  m_intervalo<- function(still_e, phi_e, beta_e,tau_e, m)
```

## B.2. Segunda prueba empírica

---

```
{
  cag_m<-grf(n, nsim=m, grid =cag$coords, xlims=c(0,10),
  ylims=c(0,10),cov.model=model_H0, method="eigen",
  cov.pars = c(still_e, phi_e), nugget =tau_e,
mean=beta_e)

  v_bin_m<- variog(cag_m, nugget.tolerance=0, trend="cte",
  option = "bin",estimator.type="modulus",
  uvec=uvec, pairs.min=30)

  return(v_bin_m$v)
}

###--- Aceptación vs Rechazo ---###

L<-length(h_k)

total<-matrix(NA,1,3)

for(i in 1:3)
{
  intervalo<-m_intervalo(still_e=wmr[1,i], phi_e=wmr[2,i],
  beta_e=wmr[3,i],tau_e=wmr[4,i], m)

  tt<-t(apply(intervalo,1,sort))

  gmin<-tt[,10]
  gmax<-tt[,390]

  res<-ifelse(gmin<=gamma_h_k & gamma_h_k<=gmax,1,0)

  total[1,i]<-ifelse(sum(res)==L,0,1)
}

return(total)
}

###--- Simulación Monte Carlo ---###

N<-1000
m <-399

model_H1<-"exp"

n<-500
```

### B.3. Prueba con el criterio *MSNE*

---

```
entre<-11

r_total<-matrix(0,N,3)

for(j in 1:N)
{
  r_total[j,]<-prueba(n=n, m=m, model_H1,entre=entre)
  cat(j,"\n")
}

apply(r_total,2,sum)/N
```

### B.3. Prueba con el criterio *MSNE*

```
rm(list=ls())
library(geoR) # Este Paquete permite utilizar las funciones \emph{grf},
              # \emph{variog}, \emph{variofit} y \emph{xvalid}.

library(stats) # Paquete para el cálculo de los cuantiles de la distribución
               #  $\chi^2$ .

prueba<-function(n, model_H1 , entre)
{
  ###--- Simulación del campo Gaussiano ---###

  estimator.type<-"modulus"

  cag <- grf(n, nsim=1, grid = "irreg",xlims=c(0,10),ylims=c(0,10),
            cov.model=model_H1 , method="eigen",
            cov.pars = c(4, 2), nugget = 0, mean=0.5)

  ###--- Estimación del semivariograma empírico ---###

  dis<-sqrt(2*10*10)/2
  uvec<-seq(0,dis,l=entre)

  v_bin_c<-variog(cag, nugget.tolerance=0, trend="cte",
                 option = "bin", estimator.type = estimator.type,
  uvec=uvec, pairs.min=30)

  gamma_h_k<-v_bin_c$v
```

### B.3. Prueba con el criterio *MSNE*

---

```
h_k<-v_bin_c$u

###--- Estimación de theta ---###

sill<-var(cag$data)
phi<-mean(v_bin_c$v)
nugget<-0

model_H0 <-"exp"

ml<-likfit(cag, cov.model=model_H0, ini.cov.pars=c(sill, phi),
           fix.nugget =TRUE,  lik.method="ML")

reml<-likfit(cag, cov.model=model_H0, ini.cov.pars=c(sill, phi),
             fix.nugget = TRUE,  lik.method="REML")

###--- Validación cruzada ---###

m_ml<-  xvalid(cag, model = ml)
m_reml<- xvalid(cag, model = reml)

msne.ml<-mean(m_ml$std.error^2)
msne.reml<-mean(m_reml$std.error^2)

n_ml<-n*msne.ml
n_reml<-n*msne.reml

datos<-matrix(0,1,2)
alfa<-0.05

###--- Rechaza H0: 1, No rechaza H0: 0 ---###

datos[1,1]<-ifelse(qchisq(alfa/2,n)<=n_ml & n_ml <= qchisq(1-alfa/2,n)
,0,1)
datos[1,2]<-ifelse(qchisq(alfa/2,n)<=n_reml & n_reml <= qchisq(1-alfa/2,n)
,0,1)

return(datos)
}

###--- Simulación Monte Carlo ---###

N<-200
n<-50
```

#### B.4. Prueba bootstrap basada en el criterio *MSNE*

---

```
entre<-18

model_H1<-"exp"

msne_ml<-as.numeric()
msne_reml<-as.numeric()

for(j in 1:N)
  {
    datos<-prueba(n, model_H1 ,entre)

    msne_ml[j]<-datos[1,1]
    msne_reml[j]<-datos[1,1]
    cat(j,"\n")
  }

sum(msne_ml)/N
sum(msne_reml)/N
```

#### B.4. Prueba bootstrap basada en el criterio *MSNE*

```
rm(list=ls())

library(geoR) # Este Paquete permite utilizar las funciones \emph{grf},
             # \emph{variog}, \emph{variofit} y \emph{xvalid}.

library(stats) # Paquete para el cálculo de los cuantiles de la distribución
              #  $\chi^2$ .

library(snowfall) # Paquete para correr en paralelo.

N<-200
m<-120
n<-300

entre<-11

model_H1<-"exp"

datos<-as.numeric()

for (i in 1:N)
```

## B.4. Prueba bootstrap basada en el criterio *MSNE*

---

```
{

###--- Simulación del campo Gaussiano ---###

estimator.type<-"modulus"

cag <- grf(n,nsim=1, grid = "irreg",xlims=c(0,10),ylims=c(0,10),
          cov.model=model_H1 , method="eigen",
          cov.pars = c(4, 2), nugget = 0, mean=0.5)

###--- Estimación del semivariograma empírico ---###

dis<-sqrt(2*10*10)/2
uvec<-seq(0,dis,l=entre)

v_bin_c<-variog(cag, nugget.tolerance=0, trend="cte", option = "bin",
               estimator.type = estimator.type, uvec=uvec, pairs.min=30)

###--- Estimación de theta ---###

sill<-var(cag$data)
phi<-mean(v_bin_c$v)
nugget<-0

model_H0 <-"exp"

ml<-likfit(cag, cov.model=model_H0, ini.cov.pars=c(sill, phi),
          fix.nugget =TRUE, nugget=nugget, lik.method="ML")

###--- Validación cruzada ---###

xva<- xvalid(cag, model = ml)

msne <- n*mean(xva$std.error^2)

###---Simulación en Paralelo ---###
#####

sfInit(parallel=TRUE, cpus=4)
sfLibrary(snowfall)
sfLibrary(geoR)
sfExportAll( )

#####
```

## B.4. Prueba bootstrap basada en el criterio *MSNE*

---

```
cuanti<-function(a)
{
  cag_m<- grf(n, nsim=1, grid =cag$coords,xlims=c(0,10),
  ylims=c(0,10),cov.model=model_H0 , method="eigen",
  cov.pars = c(ml$sigmasq, ml$phi), nugget = 0,
  mean=ml$beta)

  v_bin_m<-variog(cag_m, nugget.tolerance=0, trend="cte",
  option = "bin",estimator.type = estimator.type,
  uvec=uvec, pairs.min=30)

  sill_m<-var(cag_m$data)
  phi_m<-mean(v_bin_m$v)

  ml_m<-likfit(cag_m, cov.model=model_H0, ini.cov.pars=
  c(sill_m, phi_m),fix.nugget =TRUE, nugget=nugget,
  lik.method="ML")

  xva_m<- xvalid(cag_m, model = ml_m)

  msne_m <-n*mean(xva_m$std.error^2)

  return(msne_m)
}

resultado<-sfLapply(1:m,cuanti)
sfStop()
#####

cuantiles<- unlist(rbind(resultado))

a<-sort(cuantiles)[round(0.025*m)]
b<-sort(cuantiles)[round(0.975*m)]

datos[i]<-ifelse( msne<a | msne > b ,1,0)

cat(i,"\n")
}

sum(datos)/N
```

### B.5. Prueba de correlación muestral

```
rm(list=ls())
library(geoR) # Este Paquete permite utilizar las funciones \emph{grf},
              # \emph{variog}, \emph{variofit} y \emph{xvalid}.

library(snowfall) # Paquete para correr en paralelo.

fun_sigma<-function(cag, v_bin)
  {
    fun_ml<-function(cag, v_bin, c)
      {
        ml<-likfit(cag, cov.model="exp",
ini.cov.pars=c(var(cag$data), mean(v_bin$v)),
              fix.nugget =TRUE, nugget=0, lik.method="ml",
              limits=pars.limits(phi = c(mean(v_bin$v)+c,
100),sigmasq = c(var(cag$data), 100)))

        return(ml)
      }
    c<-0
    re_f<-fun_ml(cag, v_bin, c)

    while(re_f$sigmasq<=max(v_bin$v))
      {
        c<-c+1
        re_f<-fun_ml(cag, v_bin, c)

        re_f$sigmasq<-re_f$sigmasq
      }
    return(re_f)
  }

###--- Bajo H_0 ---###

N<-500
m<-200
n<-300
entre<-11

model_H1<-"exp"

uvec<-seq(0,sqrt(2*10*10)/2,l=entre)
```

## B.5. Prueba de correlación muestral

---

```
tipo<-"modulus"

recha<-as.numeric(0)

for(j in 1:N)
  {
    ###--- Simulación del campo Gaussiano ---###

    cag_s<-grf(n, nsim=1, grid="reg", xlims=c(0,10),
              ylims=c(0,10),cov.model=model_H1, method="eigen",
              cov.pars = c(4, 2), nugget =0, mean=0.5)

    ###--- Calculo del semivariograma empírico ---###

    v_bin_s<-variog(cag_s, option="bin", estimator.type=tipo,
                   uvec=uvec, pairs.min=30,nugget.tolerance=0)

    sig_s<-fun_sigma(cag=cag_s, v_bin=v_bin_s)

    g_h=-1*log(1-v_bin_s$v/sig_s$sigmasq)

    r<-cor(g_h, v_bin_s$u)

    ###---Simulaciones en Paralelo ---###
    #####
    sfInit(parallel=TRUE, cpus=4)
    sfLibrary(snowfall)
    sfLibrary(geoR)
    sfExportAll( )
    #####

    fun_dirst_R<-function(a)
      {
        cag_m<-grf(n, nsim=1, grid =cag_s$coords, xlims=c(0,10),
                  ylims=c(0,10),cov.model="exp", method="eigen",
                  cov.pars = c(sig_s$sigmasq, sig_s$phi),
nugget =0, mean=sig_s$beta)

        v_bin_m<-variog(cag_m, option="bin", estimator.type=tipo,
                       uvec=uvec, pairs.min=30,nugget.tolerance=0)

        sig_m<-fun_sigma(cag=cag_m, v_bin=v_bin_m)

        g_h=-1*log(1-v_bin_m$v/sig_m$sigmasq)
```

## B.5. Prueba de correlación muestral

---

```
        r<-cor(g_h, v_bin_m$u)

        return(r)
    }

    resultado<-sfLapply(1:m, fun_dirst_R)
    sfStop()

R<-matrix(unlist(rbind(resultado)),1,m)
ca<-sort(R)[m*(0.05)]

recha[j]<-ifelse(r< ca, 1, 0)
cat(j,"\n")

}
sum(recha)/N
```