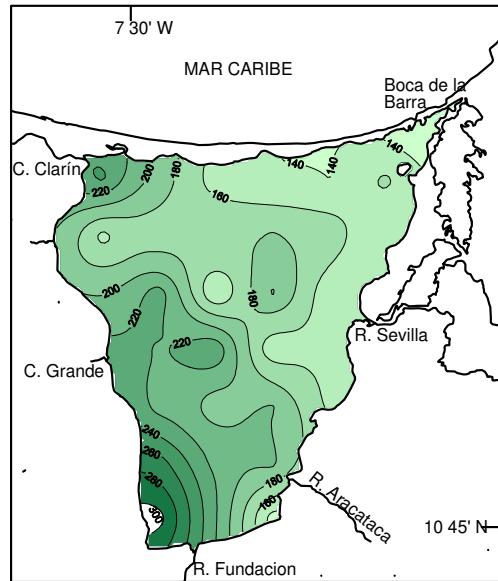


# Estadística Espacial



## NOTAS DE CLASE

Ramón Giraldo Henao

Profesor Asociado

Departamento de Estadística

Universidad Nacional de Colombia

Bogotá, Diciembre de 2011



# Prefacio

La necesidad de acudir a herramientas estadísticas para el análisis de datos en todas las áreas del conocimiento, ha hecho que aparezcan con el correr de los años nuevas metodologías que, no obstante se centran en fundamentos probabilísticos comunes, son específicas para cada una de las diversas disciplinas del saber. Algunos ejemplos son econometría, psicometría o bioestadística. Desde hace unas dos décadas se ha dedicado mucho interés a la modelación de datos espaciales, entendidos como aquellos en los que además de los valores de las variables de interés se registran también lo sitios donde estos ocurren. El análisis de datos espaciales se ha hecho frecuente en muchas áreas del conocimiento, incluyendo entre otras geografía, geología, ciencias ambientales, economía epidemiología o medicina. Las técnicas estadísticas clásicas suponen que al estudiar un fenómeno se realizan observaciones bajo circunstancias idénticas y además que las observaciones son independientes y por ello no son convenientes para analizar fenómenos que varían en tiempo o espacio. Cuando se tienen datos espaciales, intuitivamente se tiene la noción de que observaciones cercanas están correlacionadas y por ello es necesario acudir a herramientas de análisis que contemplen dicha estructura.

El presente documento ha sido preparado como guía de clase para estudiantes del curso de estadística espacial del pregrado en estadística de la Universidad Nacional de Colombia. Este cubre los elementos esenciales de cada una de las técnicas de análisis de datos espaciales, es decir, geoestadística, datos de áreas (regionales) y patrones puntuales. Estas notas también tienen el propósito servir de consulta a geólogos, biólogos, agrónomos, ingenieros, meteorólogos, ecólogos, epidemiólogos y todos aquellos profesionales que se encargan del estudio de información georreferenciada. El documento tiene un enfoque teórico-práctico.

Para el seguimiento completo de la teoría descrita se requiere tener conocimientos básicos de álgebra lineal, estadística matemática (univariable y multivariable) y de modelos lineales. Un resumen no exhaustivo de estos temas es hecho al final en el apéndice. Aquellos lectores que estén poco familiarizados con los métodos estadísticos, podrán obviar las secciones en las que se hacen desarrollos teóricos y centrar su atención en la filosofía de los métodos presentados y en las aplicaciones mostradas en cada uno de los capítulos del documento. No obstante en el escrito se cubren diversos temas de la estadística espacial y se hacen aplicaciones de métodos recientes, es necesario acudir a la lectura de artículos científicos y textos avanzados para lograr un buen dominio de estas metodologías.

En el texto no se hace una redacción en castellano riguroso, en el sentido de que en muchas ocasiones se prefiere evitar la traducción de palabras técnicas. Por ello en muchas partes del documento se encuentran expresiones como *kriging*, *test*, *nugget* o *kernel*, en vez de traducciones como *krigeage*, *prueba*, *pepita* o *núcleo*, respectivamente..

# Índice general

<b>1. Datos Espaciales y Análisis Exploratorio</b>	<b>1</b>
1.1. Conceptos básicos de probabilidad y procesos estocásticos . . . . .	1
1.2. Datos espaciales y áreas de la estadística espacial . . . . .	3
1.3. Medidas de dependencia espacial . . . . .	8
1.3.1. Test de Mantel . . . . .	9
1.3.2. Test de Moran . . . . .	10
1.3.3. Variograma . . . . .	12
1.4. Efectos de la correlación en inferencia estadística . . . . .	14
1.4.1. Efecto en la estimación . . . . .	14
1.4.2. Efecto en la predicción . . . . .	15
1.5. Implementación en R . . . . .	18
<b>2. Patrones Puntuales</b>	<b>22</b>
2.1. Tipos de Patrones . . . . .	23
2.2. Estimación de la intensidad . . . . .	23
2.2.1. Definición de Intensidad . . . . .	25
2.2.2. Estimación Kernel . . . . .	27
2.2.3. Estimación Kernel de la intensidad $\lambda(s)$ . . . . .	29
2.3. Métodos basados en cuadrantes . . . . .	30
2.3.1. Test de bondad de ajuste $\chi^2$ . . . . .	31
2.4. Métodos basados en distancias . . . . .	32
2.4.1. Función G(h) . . . . .	32

2.4.2. Función F(h) . . . . .	34
2.4.3. Función K de Ripley . . . . .	36
2.4.4. Función L . . . . .	37
2.5. Modelos . . . . .	37
2.5.1. Proceso Poisson no Homogéneo . . . . .	38
2.5.2. Proceso Poisson Homogéneo . . . . .	41
2.5.3. Proceso de Cox . . . . .	42
2.5.4. Proceso de Poisson Cluster . . . . .	43
2.5.5. Implementación en R . . . . .	44
<b>3. Datos de Áreas</b> . . . . .	<b>58</b>
3.1. Visualización de datos de áreas . . . . .	58
3.1.1. Mapas de datos originales y tasas . . . . .	58
3.1.2. Mapas de riesgos relativos . . . . .	59
3.1.3. Mapas Probabilísticos y Suavizados con Estimación Bayesiana . . . . .	62
3.2. Exploración de datos de áreas . . . . .	64
3.2.1. Medidas de proximidad espacial . . . . .	64
3.2.2. Medias móviles . . . . .	64
3.2.3. Estimación kernel . . . . .	64
3.3. Correlación espacial . . . . .	64
3.3.1. I de Moran y C de Geary . . . . .	67
3.3.2. Indice Bayesiano Empírico (IBE) . . . . .	68
3.3.3. Función de semivariograma . . . . .	69
3.4. Modelos de regresión espacial . . . . .	69
3.4.1. Modelo de regresión polinomial local y ponderado geográficamente .	69
3.4.2. Modelos lineales con errores correlacionados . . . . .	72
3.4.3. Modelos autorregresivos espaciales . . . . .	74
3.4.4. Modelo autorregresivo simultáneo (SAR) . . . . .	74
3.4.5. Modelo autorregresivo condicional (CAR) . . . . .	77

<b>4. Geoestadística</b>	<b>1</b>
4.1. Definición de Geoestadística . . . . .	1
4.2. Variable Regionalizada . . . . .	2
4.2.1. Momentos de una Variable Regionalizada . . . . .	2
4.2.2. Estacionariedad . . . . .	3
4.3. Correlación espacial muestral y ajuste de modelos . . . . .	6
4.3.1. Funciones de correlación espacial . . . . .	6
4.3.2. Modelos teóricos de semivarianza . . . . .	8
4.4. Estimación por mínimos cuadrados . . . . .	14
4.4.1. Estimación por Mínimos cuadrados . . . . .	14
4.4.2. Mínimos cuadrados ponderados . . . . .	15
4.4.3. Cuasi-verosimilitud . . . . .	16
4.5. Estimación asumiendo normalidad . . . . .	18
4.5.1. Estimación por Máxima Verosimilitud . . . . .	18
4.6. Predicción espacial . . . . .	21
4.6.1. Predicción espacial optima . . . . .	21
4.6.2. Definición de Kriging . . . . .	22
4.6.3. Kriging Ordinario . . . . .	22
4.6.4. Otros Métodos Kriging . . . . .	30
4.6.5. Kriging en Bloques . . . . .	33
4.6.6. Kriging Universal . . . . .	34
4.6.7. Kriging Residual . . . . .	37
4.6.8. Kriging Indicador . . . . .	37
4.6.9. Kriging Log-Normal y Multi-Gaussiano . . . . .	39
4.6.10. Cokriging Ordinario . . . . .	40

# Índice de figuras

1.1.	Distribución espacial de clorofila en la Ciénaga Grande de Santa Marta (costa norte de Colombia). Datos medidos en un jornada de muestreo realizada en marzo de 1997. . . . .	6
1.2.	Mapa cloroplético de la tasa de delitos en Colombia en el año 2003. . . . .	7
1.3.	Ubicación de deslizamientos en el corredor Caño Limón-Coveñas en 2008 ( <i>panel izquierdo</i> ) y ubicación de sismos de baja magnitud en Colombia en el periodo Julio a Diciembre de 2008 ( <i>panel derecho</i> ). . . . .	9
1.4.	Comportamiento típico de un semivariograma acotado con una representación de los parámetros básicos. SEMEXP corresponde al semivariograma experimental y MODELO al ajuste de un modelo teórico. . . . .	13
2.1.	Ejemplos simulados de patrones puntuales con distribución aleatoria ( <i>izquierda</i> ), regular ( <i>derecha</i> ) y agregada ( <i>centro</i> ). . . . .	24
2.2.	a.) Ejemplos de patrones que incluyen covariables y b.) efecto de borde ( <i>derecha</i> ). En la figura de la izquierda la escala de valores corresponde a la elevación en metros . . . . .	25
2.3.	Marcas asociadas a los datos de sismos en Colombia . . . . .	26
2.4.	Estimación de la intensidad de sismos en Colombia (Julio-Diciembre 2008). . . . .	30
2.5.	Gráficas del la función G para cada tipo de proceso . . . . .	34
2.6.	Gráficas del la función F para cada tipo de proceso . . . . .	35
2.7.	Gráficas del la función K para cada tipo de proceso . . . . .	38
2.8.	Gráficas del la función L para cada tipo de proceso . . . . .	39

3.1.	Mapas temáticos de mortalidad infantil en Colombia en 2003. . . . .	59
3.2.	Mapas temáticos de tasas de mortalidad infantil en Colombia en 2003. . .	60
3.3.	Distribución de las tasa de violencia familiar (arriba) y de hurto calificado (abajo) según unidades seccionales de la Fiscalía General de la Nación, en el año 2003. . . . .	61
3.4.	Dispersograma entre tasas de violencia familiar de 2003 y tasas suavizadas por el método bayesiano empírico. ). . . . .	63
3.5.	Las áreas sombreadas corresponden a los departamentos de la costa Caribe de Colombia (sin incluir San Andrés y Providencia). . . . .	65
4.1.	Representación de una superficie interpolada para una variable regionalizada estacionaria. . . . .	4
4.2.	Representación de una superficie interpolada para una variable regionalizada no estacionaria. . . . .	6
4.3.	Comportamiento típico de un semivariograma acotado con una representación de los parámetros básicos. SEMEXP corresponde al semivariograma experimental y MODELO al ajuste de un modelo teórico. . . . .	10
4.4.	Comparación de los modelos exponencial, esférico y Gaussiano. La línea punteada vertical representa el rango en el caso del modelo esférico y el rango efectivo en el de los modelos exponencial y gaussiano. Este tiene un valor de 210, respecto a una escala simulada entre 0 y 300. El valor de la meseta es 30 y el de la pepita 0. El 95 % de la meseta es igual a 28.5. . . . .	11
4.5.	Comportamiento típico de los modelos de semivarianza monómicos. . . . .	13
4.6.	Modelo de semivarianza teórico para variables sin correlación espacial. . . . .	13
4.7.	Representación del procedimiento kriging residual. La superficie interpolada (arriba) es igual a la suma de la tendencia lineal (centro) más la predicción de los errores (abajo). . . . .	38
4.8.	Representación de la transformación a scores normales. . . . .	40

# Capítulo 1

## Datos Espaciales y Análisis Exploratorio

En este capítulo se presentan conceptos básicos de probabilidad que permiten posteriormente enmarcar las áreas de la estadística espacial dentro del contexto de los procesos estocásticos. Se definen algunas medidas de autocorrelación espacial y se dan dos ejemplos de como la dependencia espacial afecta la inferencia estadística.

### 1.1. Conceptos básicos de probabilidad y procesos estocásticos

**Definition 1.1.** *Sea  $\Omega \neq \emptyset$ . Un sistema  $\mathcal{F}$  de subconjuntos de  $\Omega$  se llama  $\sigma$ -álgebra si satisface las siguientes condiciones*

1.  $\Omega \in \mathcal{F}$
2. Si  $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$
3. Si  $A_1, \dots, A_n \in \mathcal{F} \Rightarrow \bigcup_{i=1}^n A_i \in \mathcal{F}$ .

**Definition 1.2.** *Sea  $\Omega \neq \emptyset$ ,  $\mathcal{F}$  una  $\sigma$ -álgebra de subconjuntos de  $\Omega$ . La pareja  $(\Omega, \mathcal{F})$  se llama espacio medible.*

## 21.1. CONCEPTOS BÁSICOS DE PROBABILIDAD Y PROCESOS ESTOCÁSTICOS

**Definition 1.3.** Sea  $(\Omega, \mathcal{F})$  un espacio medible.  $P : \mathcal{F} \rightarrow [0, 1]$  se llama medida de probabilidad si satisface

1.  $P(A) \geq 0, \forall A \in \mathcal{F}$
2.  $P(\Omega) = 1$
3. Si  $A_1, \dots, A_n \in \mathcal{F}$  con  $A_i \cap A_j = \emptyset, \forall i \neq j \Rightarrow P(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i)$ .

**Propiedades de  $P$**

1.  $P(\emptyset) = 0$
2.  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
3. Si  $A \subseteq B \Rightarrow P(A) \leq P(B)$ .
4.  $P(A) = 1 - P(A^c)$ .

**Definition 1.4.** La tripla  $(\Omega, \mathcal{F}, P)$ , donde  $\Omega \neq \emptyset$ ,  $\mathcal{F}$   $\sigma$ -álgebra sobre  $\Omega$  y  $P$  es una medida de probabilidad sobre  $(\Omega, \mathcal{F})$ , se denomina espacio de probabilidad.

**Definition 1.5.** Sea  $(\Omega, \mathcal{F}, P)$ , un espacio de probabilidad.  $X : \Omega \rightarrow \mathbb{R}$  se llama variable aleatoria.

**Definition 1.6 (Proceso Estocástico).** Es una familia de variables aleatorias  $\{Z(s) : s \in D \subset \mathbb{R}^P\}$  definida sobre un espacio de probabilidad  $(\Omega, \mathcal{F}, P)$ . El conjunto  $D$  de índices del procesos se denomina espacio de parámetros. Los valores que toma  $Z(s)$  se llaman estados y el conjunto de todos los posibles valores de  $Z(s)$  se llama espacio de estados.

Los procesos estocásticos son clasificados de acuerdo con el espacio de parámetros (discreto y continuo) y el espacio de estados (discreto y continuo). Algunos ejemplos de procesos estocásticos no espaciales son los siguientes

1. Espacio de parámetros discreto y espacio de estados discreto
  - $Z(n)$ : Preferencia del consumidor en el  $n$ -ésimo mes, con  $n \in \mathbb{N}$ .

- $Z(n)$ : Número de individuos de la  $n$ -ésima generación de una población, con  $n \in \mathbb{N}$ .
2. Espacio de parámetros continuo y espacio de estados discreto
- $Z(t)$ : Número de partículas de una sustancia acuosa de volumen  $t$ , con  $t \in T \subset \mathbb{R}$ .
  - $Z(t)$ : Número de individuos que esperan el bus por periodo de tiempo  $t$ , con  $t \in T \subset \mathbb{R}$ .
3. Espacio de parámetros discreto y espacio de estados continuo
- $Z(n)$ : Tiempo de espera hasta que el  $n$ -ésimo estudiante arribe a la parada de bus, con  $n \in \mathbb{N}$ .
  - $Z(n)$ : Utilidad en pesos de un jugador después del  $n$ -ésimo lanzamiento de una moneda, con  $n \in \mathbb{N}$ .
4. Espacio de parámetros continuo y espacio de estados continuo
- $Z(t)$ : Contenido de un embalse sobre un periodo de tiempo  $t$ , con  $t \in T \subset \mathbb{R}$ .
  - $Z(t)$ : Temperatura en el instante  $t$ , con  $t \in T \subset \mathbb{R}$ .

## 1.2. Datos espaciales y áreas de la estadística espacial

Estadística espacial es la reunión de un conjunto de metodologías apropiadas para el análisis de datos que corresponden a la medición de variables aleatorias en diversos sitios (puntos del espacio o agregaciones espaciales) de una región. De manera más formal se puede decir que la estadística espacial trata con el análisis de realizaciones de un proceso estocástico  $\{Z(s) : s \in D \subset \mathbb{R}^P\}$ , en el que  $s$  es la ubicación en el espacio Euclíadiano  $P$ -dimensional y  $Z(s)$  es una variable aleatoria en la ubicación  $s$ .

### Observaciones

- El proceso estocástico  $\{Z(s) : s \in D \subset \mathbb{R}^P\}$ , en el que  $s$  es sitio del espacio, también se denomina *proceso aleatorio* o *campo aleatorio*.
- El proceso estocástico  $\{\mathbf{Z}(s) : s \in D \subset \mathbb{R}^P\}$ , en el que  $\mathbf{Z}(s)$  es un vector aleatorio se denomina en el contexto espacial *proceso aleatorio multivariable* o *campo aleatorio multivariable*
- El proceso estocástico  $\{Z(s) : s \in D \subset \mathbb{R}^P\}$  en el que tanto el espacio de estados como el espacio de parámetros es continuo (es decir que las variables aleatorias  $Z(s)$  son continuas y  $D \subset \mathbb{R}^P$  es un conjunto continuo) se denomina *variable regionalizada*. Este término es particularmente usado en aplicaciones de la estadística espacial en ingeniería y geología.
- Cuando se tiene una observación del proceso estocástico  $\{Z(s) : s \in D \subset \mathbb{R}^P\}$  se dispone de una muestra de tamaño  $\{\mathbf{Z}(s) = (Z(s_1), Z(s_2), \dots, Z(s_n))\}$  (con  $n$  el número de sitios donde se hace la medición de la variable aleatoria  $Z(s)$ ) y no de una muestra de tamaño  $n$  de una variable aleatoria. Por ello puede ser carente de sentido práctico hacer inferencia estadística clásica (intervalos de confianza, pruebas de normalidad, etc) con los datos obtenidos. Desconocer esto hace que se cometan errores intentando validar los supuestos necesarios para la aplicación de métodos estadísticos espaciales. En general en estadística espacial, como en el caso clásico, es deseable tener normalidad para hacer inferencia. Sin embargo lo que se asume en este contexto es que la muestra corresponde a la observación de vector aleatorio con distribución normal multivaluada y no que se tiene una muestra  $n$ -variada de una variable aleatoria con distribución normal. Usar una prueba de normalidad univariada (por ejemplo la de Shapiro-Wilk) para comprobar si los datos siguen una distribución normal es ciertamente equivocado en el contexto espacial, puesto que además de desconocer que no se tiene una muestra *iid* (puesto que hay dependencia espacial), lo que en realidad habría que probar es normalidad multivariada.

La estadística espacial se subdivide en tres grandes áreas. La pertinencia de cada una de ellas está asociada a las características del conjunto  $D$  de índices del proceso estocástico

de interés. A continuación se mencionan dichas áreas y se describen las propiedades de  $D$  en cada una de éstas.

**Geoestadística:** Estudia datos de procesos estocásticos en los que el espacio de parámetros  $D \subset \mathbb{R}^P$  es continuo. Algunos ejemplos de datos espaciales que son tratados con métodos geoestadísticos son

- $\{Z(s) : s \in D \subset \mathbb{R}^P\}$ , donde  $Z(s)$  mide el contenido de nitrógeno en sitios de un parcela experimental. En este caso los sitios pertenecen a  $D \subset \mathbb{R}^2$ .
- $\{Z(s) : s \in D \subset \mathbb{R}^P\}$ , donde  $Z(s)$  corresponde a la precipitación en sitios de Colombia.

En los dos ejemplos anteriores hay infinitos sitios donde medir y por ello el conjunto de parámetros es continuo. Sin embargo en la práctica es potestad del investigador seleccionar en qué sitios de la región de interés hace la medición de las variables, es decir, el investigador puede hacer selección de puntos del espacio a conveniencia o puede seleccionar los sitios bajo algún esquema de muestreo probabilístico. En este sentido se dice que el conjunto  $D \subset \mathbb{R}^P$  es *fijo*. Un ejemplo de un conjunto de datos analizado con métodos geoestadísticos es presentado en la Figura 1.1. Es importante resaltar que en geoestadística el propósito esencial es la interpolación y si no hay continuidad espacial pueden hacerse predicciones carentes de sentido.

**Datos de áreas o regionales:** En este caso el proceso estocástico tiene espacio de parámetros  $D \subset \mathbb{R}^P$  discreto y la selección de los sitios de medición depende del investigador ( $D$  fijo). Las ubicaciones de muestreo pueden estar regular o irregularmente espaciadas. Dos ejemplos de datos regionales son

- $\{Z(s) : s \in D \subset \mathbb{R}^P\}$ , donde  $Z(s)$  es la variable aleatoria correspondiente a la tasa de mortalidad y los sitios son departamentos de Colombia, es decir  $D$  es el conjunto discreto formado por los departamentos del país.

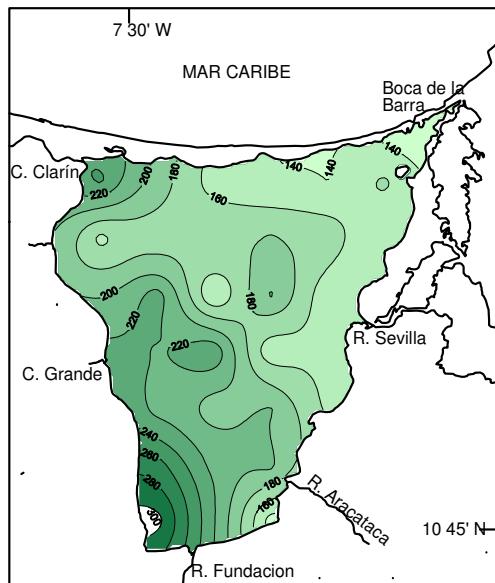


Figura 1.1: Distribución espacial de clorofila en la Ciénaga Grande de Santa Marta (costa norte de Colombia). Datos medidos en un jornada de muestreo realizada en marzo de 1997.

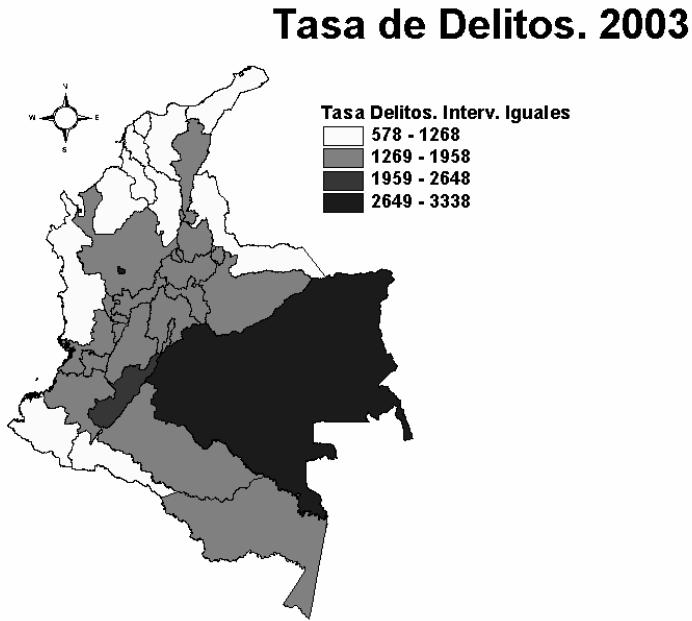


Figura 1.2: Mapa cloroplético de la tasa de delitos en Colombia en el año 2003.

- $\{Z(s) : s \in D \subset \mathbb{R}^P\}$ , donde  $Z(s)$  corresponde a la producción cafetera (en kilogramos) y  $D$  es el conjunto de todas las fincas productoras de café del país.

Nuevamente el investigador puede decidir donde (en que departamentos o en que fincas en los ejemplos) hace la medición de las variables de interés, es decir en datos de áreas también el conjunto  $D \subset \mathbb{R}^P$  es *fijo*. En la Figura 1.2 se presenta un ejemplo de un conjunto de datos que corresponde a la observación de un proceso aleatorio de datos regionales. Un ejemplo de datos de área con sitios regularmente espaciados es el de colores de pixeles en interpretación de imágenes de satélite. En ese caso el conjunto de ubicaciones de interés es discreto y estas corresponden a agregaciones espaciales más que a un conjunto de puntos del espacio. Es obvio que la interpolación espacial puede ser carente de sentido con este tipo de datos. Sus principales aplicaciones se encuentran en el campo epidemiológico.

**Patrones Puntuales:** La diferencia central del análisis de patrones puntuales con las

técnicas geoestadísticas y el análisis de datos de áreas radica en el hecho de que el conjunto  $D \subset \mathbb{R}^P$  es aleatorio, es decir que la decisión al respecto de donde se hace la medición no depende del investigador. Dicho conjunto puede ser discreto o continuo, pero la ubicación de los sitios donde ocurre el fenómeno a estudiar es dada por la naturaleza. En general el propósito de análisis en estos casos es el de determinar si la distribución de los individuos dentro de la región es aleatoria, agregada o uniforme. Algunos ejemplos de datos correspondientes a patrones puntuales son dados a continuación

- Ubicación de nidos de pájaros en una región dada.
- Localización de imperfectos en una placa metálica
- Sitios de terremotos en Colombia
- Municipios de Colombia con mayorías negras

En los tres primeros ejemplos  $D \subset \mathbb{R}^P$  es continuo y en el último es discreto. Cuando en cada sitio se hace medición de alguna variable (por ejemplo del número de huevos en los nidos, de la forma del imperfecto en la placa o de la tasa de analfabetismo de los municipios de mayorías negras) se dice que ese tiene un *patrón espacial marcado*. Dos ejemplos de datos correspondientes a patrones espaciales son dados en la Figura 1.3.

### 1.3. Medidas de dependencia espacial

La dependencia espacial hace referencia a la estructura de correlación de las variables aleatorias del proceso  $\{Z(s) : s \in D \subset \mathbb{R}^P\}$ . Cuando hay dependencia espacial los sitios cercanos tienen valores más similares que los distantes. Por el contrario la ausencia de correlación espacial se refleja en el hecho de que la distancia entre los sitios no tiene influencia en la relación de sus valores. A continuación se presentan algunos test y funciones que permiten establecer estadísticamente o de manera empírica si existe dependencia (correlación) espacial.

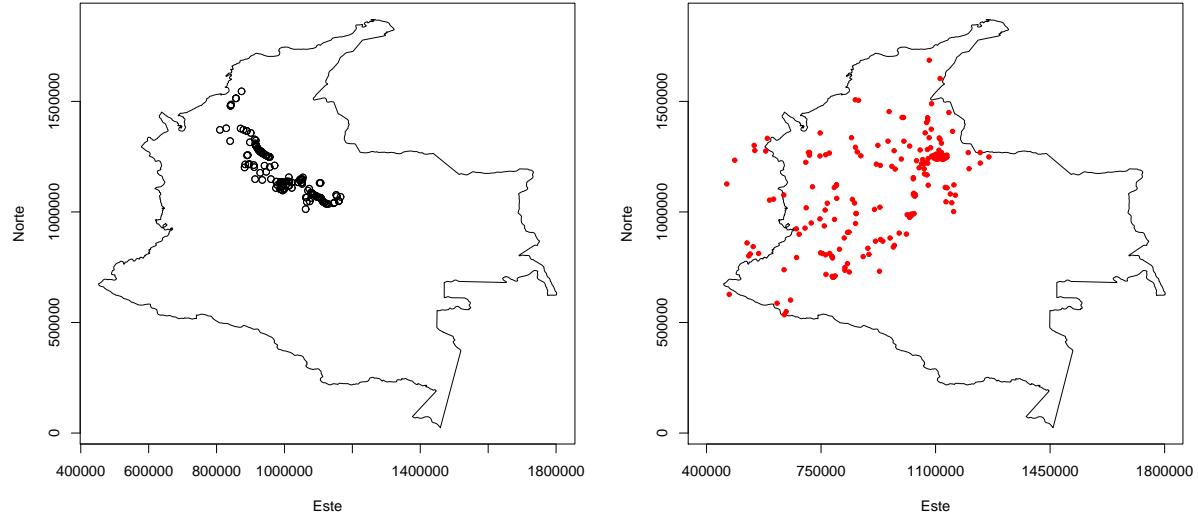


Figura 1.3: Ubicación de deslizamientos en el corredor Caño Limón-Coveñas en 2008 (*panel izquierdo*) y ubicación de sismos de baja magnitud en Colombia en el periodo Julio a Diciembre de 2008 (*panel derecho*).

### 1.3.1. Test de Mantel

Permite comprobar estadísticamente si las observaciones provienen de un proceso estocástico en el que las variables son correlacionadas espacialmente.

*Hipótesis*

$$H_0 : \text{Hay aleatoriedad espacial}$$

$$H_a : \text{Hay correlación espacial}$$

*Estadística de prueba*

$$M = \sum_{i=1}^n \sum_{j=1}^n W_{ij} U_{ij},$$

donde  $W_{ij} = \|s_i - s_j\|$  y  $U_{ij} = (Z(s_i) - Z(s_j))^2$ . La estadística de mantel está relacionada con la pendiente del modelo de regresión simple  $U_{ij} = \beta W_{ij} + e_{ij}$  a través de  $\beta = M / \sum_{i=1}^n \sum_{j=1}^n W_{ij}^2$ , es decir que intuitivamente se tiene que a mayor  $M$ , mayor dependencia espacial positiva. La significancia de la prueba puede establecerse por varios caminos. Puede emplearse un *test de permutaciones* en el que asumiendo aleatoriedad

se encuentran las  $n!$  posibles asignaciones de sitios a valores y con cada una de ellas se calcula  $M$ , obteniéndose por consiguiente su distribución bajo  $H_0$ . También en el caso de  $n$  grande puede usarse un *test de Monte Carlo* en el que solo se toman  $k$  de las asignaciones aleatorias de sitios a valores de la variable. En ambos casos (permutaciones, Monte Carlo) podría usarse una aproximación a la normal estimando  $E(M)$  y  $V(M)$  a través de  $\bar{M} = 1/n \sum_{i=1}^n M_i$  y  $s_M^2 = 1/(n-1) \sum_{i=1}^n (M_i - \bar{M})^2$ . En el caso de asumir normalidad y aleatoriedad, es decir si  $Z(s_1), \dots, Z(s_n)$  son *iid*, con  $Z(s_i) \sim N(\mu, \sigma^2)$ , pueden obtenerse expresiones para  $E(M)$  y  $V(M)$  y establecer el nivel de significancia basándose en un *test normal*.

### 1.3.2. Test de Moran

Este test es especialmente usado en datos de áreas. Sean  $Z(s_1), \dots, Z(s_n)$ , las variables medidas en las  $n$  áreas. La noción de autocorrelación espacial de estas variables está asociada con la idea de que valores observados en áreas geográficas adyacentes serán más similares que los esperados bajo el supuesto de independencia espacial. El índice de autocorrelación de Moran considerando la información de los vecinos más cercanos es definida como

$$I = \frac{\frac{n}{\sum_{i=1}^n \sum_{j=1}^n W_{ij}} \sum_{i=1}^n \sum_{j=1}^n W_{ij} (Z(s_i) - \bar{Z})(Z(s_j) - \bar{Z})}{\sum_{i=1}^n (Z(s_i) - \bar{Z})^2} \quad (1.1)$$

Valores positivos (entre 0 y 1) indican autocorrelación directa (similitud entre valores cercanos) y valores negativos (entre -1 y 0) indican autocorrelación inversa (disimilitud entre las áreas cercanas). Valores del coeficiente cercanos a cero apoyan la hipótesis de aleatoriedad espacial.

Para el cálculo del índice de Moran es necesario definir la proximidad entre las áreas. Lo anterior se lleva a cabo por medio del cálculo de una matriz de proximidad espacial. Dado un conjunto de  $n$  áreas ( $A_1, \dots, A_n$ ) se construye una matriz  $W^{(1)}$  de orden  $(n \times n)$  donde cada uno de los elementos  $W_{ij}$  representa una medida de proximidad entre  $A_i$  y  $A_j$ . Dicha medida puede ser calculada con alguno de los siguientes criterios:

- $W_{ij} = 1$  si el centro de  $A_i$  se encuentra a una distancia determinada de  $A_j$  o  $W_{ij} = 0$  en caso contrario.
- $W_{ij} = 1$  si  $A_i$  comparte frontera con  $A_j$  y en caso contrario  $W_{ij} = 0$ .
- $W_{ij} = I_{ij}/I_i$ , donde  $I_{ij}$  es la longitud de la frontera entre  $A_i$  y  $A_j$  y  $I_i$  es el perímetro de  $A_i$ .
- $W_{ij} = d_{ij}$ , con  $d_{ij}$  la distancia entre los centros de las dos áreas.

En todos los casos anteriores  $W_{ii} = 0$ . La idea de la matriz de proximidad espacial puede ser generalizada a vecinos de mayor orden (vecinos de vecinos) construyéndose así las matrices  $W^{(2)}, \dots, W^{(n)}$ . Se acostumbra a normalizar las filas de la matriz, es decir que la suma por fila de los  $W_{ij}$  sea igual a uno.

Una vez obtenido el valor del coeficiente es necesario evaluar su significancia estadística. En otras palabras se requiere probar la hipótesis de aleatoriedad espacial con base en el valor observado. Para llevar a cabo esto es necesario establecer la correspondiente distribución de probabilidad de la estadística de prueba  $I$ . Bajo normalidad, es decir asumiendo que  $Z(s_1), \dots, Z(s_n)$  son *iid* con  $Z(s_i) \sim N(\mu, \sigma^2)$ , la estadística

$$Z = \frac{I - \mathbb{E}(I)}{\sqrt{\mathbb{V}(I)}}$$

sigue una distribución normal estándar, en la que el valor esperado y la varianza están dados por

$$\mathbb{E}(I) = -\frac{1}{(n+1)}, \quad \mathbb{V}(I) = \frac{n^2 S_1 - n^2 S_2 + 3S_0^2}{(n^2 - 1)S_0^2} - \frac{1}{(n-1)^2},$$

donde

$$\begin{aligned} S_0 &= \sum_{i \neq j}^n W_{ij}, \quad S_1 = \sum_{i \neq j}^n (W_{ij} + W_{ji})^2, \quad S_2 = \sum_{i=1}^n (W_{i0} + W_{0i})^2, \\ W_{i0} &= \sum_{j=1}^n W_{ij}, \quad W_{0i} = \sum_{j=1}^n W_{ji}. \end{aligned}$$

Otra posibilidad para establecer la significancia estadística, con menos supuestos, es llevando a cabo un test de permutación o de Monte Carlo como los descritos para la estadística de Mantel.

### 1.3.3. Variograma

El variograma, denotado por  $2\gamma(h)$ , se define como la varianza de la diferencia entre variables separadas por una distancia  $h = \|s_i - s_j\|$ . Asumiendo que  $\mathbb{E}(Z(s)) = \mu$  se tiene

$$\begin{aligned} 2\gamma(h) &= \mathbb{V}(Z(s + h) - Z(s)) \\ &= \mathbb{E}(Z(s + h) - Z(s))^2. \end{aligned} \quad (1.2)$$

La mitad del variograma se llama semivariograma y caracteriza las propiedades de dependencia espacial de un fenómeno espacial. Esta función es usualmente empleada para tratar datos de un fenómeno con continuidad espacial (datos geoestadísticos). Usando el método de momentos se tiene que un estimador del semivariograma es

$$\bar{\gamma}(h) = \frac{1}{n(h)} \sum_{i=1}^{n(h)} (Z(s_i + h) - Z(s_i))^2, \quad (1.3)$$

donde  $n(h)$  representa el número de parejas de sitios  $(s_i, s_j)$  que se encuentran separados por una distancia  $h$ . En la práctica, debido a irregularidad en el muestreo y porque en las distancias entre los sitios, se toman intervalos de distancia  $\{[0, h], (h, 2h], (2h, 3h], \dots\}$  y el semivariograma experimental corresponde a una distancia promedio entre parejas de sitios dentro de cada intervalo y no a una distancia  $h$  específica. Obviamente el número de parejas de puntos  $n$  dentro de los intervalos no es constante. Para interpretar el semivariograma experimental se parte del criterio de que a menor distancia entre los sitios mayor similitud o correlación espacial entre las observaciones. Por ello en presencia de autocorrelación se espera que para valores de  $h$  pequeños el semivariograma experimental tenga magnitudes menores a las que este toma cuando las distancias se incrementan.

Como se verá en el capítulo 4 la solución del problema de predicción espacial requiere del conocimiento de la estructura de autocorrelación para cualquier posible distancia entre sitios dentro del área de estudio. De la ecuación (1.3) es claro que el semivariograma muestral es calculado sólo para algunas distancias promedios particulares. Por ello se hace necesario el ajuste de modelos que generalicen la dependencia espacial para cualquier distancia (Figura 1.3). Existen diversos modelos teóricos de semivarianza que pueden ajustarse al semivariograma muestral. En Cressie (1993) se presenta una discusión respecto

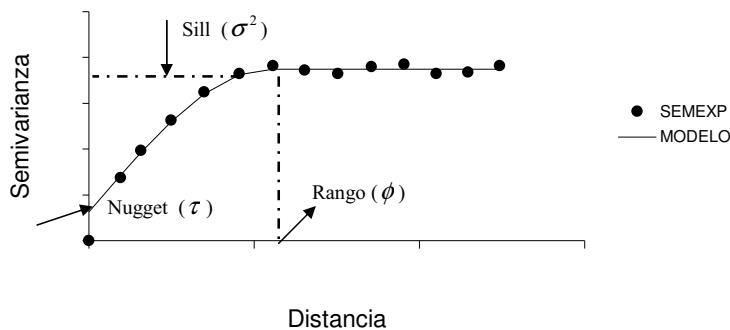


Figura 1.4: Comportamiento típico de un semivariograma acotado con una representación de los parámetros básicos. SEMEXP corresponde al semivariograma experimental y MODELO al ajuste de un modelo teórico.

a las características y condiciones que éstos deben cumplir. En general dichos modelos pueden dividirse en no acotados (lineal, logarítmico, potencial) y acotados (esférico, exponencial, Gaussiano) (?). Los del segundo grupo garantizan que la covarianza de los incrementos es finita, por lo cual son ampliamente usados cuando hay evidencia de que presentan buen ajuste. La mayoría de modelos empleados para ajustar el semivariograma muestral, tienen tres parámetros en común (Figura 1.4) que son descritos a continuación:

- *Nugget ( $\tau$ )*: Representa una discontinuidad puntual del semivariograma en el origen (Figura 1.3). Puede ser debido a errores de medición en la variable o a la escala de la misma. En algunas ocasiones puede ser indicativo de que parte de la estructura espacial se concentra a distancias inferiores a las observadas.
- *Sill ( $\sigma^2$ )*: Es un estimador de la varianza de las variables del proceso. También puede definirse como el límite del semivariograma cuando la distancia  $h$  tiende a infinito.
- *Rango( $\phi$ )*. En términos prácticos corresponde a la distancia a partir de la cual dos observaciones son independientes. El rango se interpreta como la zona de influencia. Existen algunos modelos de semivariograma en los que no existe una distancia finita para la cual dos observaciones sean independientes; por ello se llama rango efectivo a la distancia para la cual el semivariograma alcanza el 95 % de la meseta (*sill*).

## 1.4. Efectos de la correlación en inferencia estadística

Muchas métodos estadísticos están basados en el supuesto de que las variables aleatorias involucradas en la muestra son independientes. La violación de dicho supuesto tiene consecuencias en todos los procesos inferenciales. En esta sección se ilustra como la correlación entre las variables (por consiguiente la no independencia entre las mismas) afecta la estimación y la predicción en el modelo de regresión simple (sin covariables).

### 1.4.1. Efecto en la estimación

Sea  $Y_1, \dots, Y_n$  una muestra aleatoria de  $Y \sim N(\mu, \sigma^2)$ . El estimador de  $\mu$  es  $\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$ . El valor esperado y la varianza de este estimador son  $\mu$  y  $\sigma^2/n$ , respectivamente. Ahora suponga que las variables  $Y_1, \dots, Y_n$  son correlacionadas y que  $Cov(Y_i, Y_j) = \sigma^2 \rho$ . En este caso nuevamente el estimador de  $\mu$  es  $\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$  y su valor esperado es  $\mu$ , sin embargo la correlación aumenta (en este caso) la varianza del estimador. Veamos

$$\begin{aligned}
V(\bar{Y}) &= V\left(\frac{1}{n} \sum_{i=1}^n Y_i\right) \\
&= \frac{1}{n^2} \left( \sum_{i=1}^n \sum_{j=1}^n Cov(Y_i, Y_j) \right) \\
&= \frac{1}{n^2} [(\sigma^2 + \sigma^2 \rho, \dots, +\sigma^2 \rho), \dots, (\sigma^2 + \sigma^2 \rho, \dots, +\sigma^2 \rho)] \\
&= \frac{1}{n^2} (n\sigma^2 + (n-1)\sigma^2 \rho, \dots, (n-1)\sigma^2 \rho) \\
&= \frac{1}{n^2} (n\sigma^2 + n(n-1)\sigma^2 \rho) \\
&= \frac{\sigma^2}{n} (1 + (n-1)\rho). \tag{1.4}
\end{aligned}$$

Si  $\rho > 0$  en (1.4),  $V(\bar{Y}) > \sigma^2/n$ , es decir la varianza del estimador de  $\mu$  cuando hay correlación es mayor que la de este mismo cuando las variables son independientes.

### 1.4.2. Efecto en la predicción

Sean  $Y_1, \dots, Y_n$  variables aleatorias tales que  $Y_i \sim N(\mu, \sigma^2)$  y  $Cov(Y_i, Y_j) = \sigma^2\rho$ . Un modelo lineal para representar este escenario es

$$\begin{aligned} \mathbf{Y} &= \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} \mu \\ \vdots \\ \mu \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix} \\ &= \mu\mathbf{1} + \boldsymbol{\epsilon}, \end{aligned} \quad (1.5)$$

donde

$$V(\boldsymbol{\epsilon}) = \Sigma = \sigma^2 \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{pmatrix}.$$

Suponga que se quiere predecir una nueva observación  $Y_0$ . Definiendo el predictor por

$$Y_0^* = \sum_{i=1}^n \lambda_i Y_i, \quad (1.6)$$

los pesos  $\lambda_i$  se obtienen de tal forma que se minimice la esperanza de una función de pérdida. Bajo pérdida cuadrática, el mejor predictor (lineal en este caso), será el que minimiza la función

$$\min_{\lambda_1, \dots, \lambda_n} \mathbb{E}(Y_0^* - Y_0)^2, \text{ sujeto a } \mathbb{E}(Y_0^*) = \mathbb{E}(Y_0).$$

De acuerdo con lo anterior, la función objetivo es

$$\min_{\boldsymbol{\lambda}, m} \mathbb{V}(Y_0^* - Y_0), \text{ sujeto a } \sum_{i=1}^n \lambda_i = 1.$$

Se tiene que  $Cov(\mathbf{Y}, Y_0) = \sigma^2 \rho \mathbf{1}$ . Desarrollando la varianza e incluyendo un multiplicador de Lagrange para la condición de insesgadez la función a optimizar es

$$\begin{aligned} & \min_{\lambda, m} \mathbb{V}(Y_0^*) + \mathbb{V}(Y_0) - 2Cov(Y_0^*, Y_0) - 2m\left(\sum_{i=1}^n \lambda_i - 1\right) \\ & \min_{\lambda, m} \mathbb{V}(\boldsymbol{\lambda}^T \mathbf{Y}) + \sigma^2 - 2Cov(\boldsymbol{\lambda}^T \mathbf{Y}, Y_0) - 2m(\boldsymbol{\lambda}^T \mathbf{1} - 1) \\ & \min_{\lambda, m} \boldsymbol{\lambda}^T \Sigma \boldsymbol{\lambda} + \sigma^2 - 2\boldsymbol{\lambda}^T \mathbf{c} - 2m(\boldsymbol{\lambda}^T \mathbf{1} - 1), \quad \mathbf{c} = \begin{pmatrix} \sigma^2 \rho \\ \vdots \\ \sigma^2 \rho \end{pmatrix} = \sigma^2 \rho \mathbf{1}. \end{aligned}$$

Tomando derivadas respecto a  $\lambda$  y  $m$  se obtiene el siguiente sistema

$$\begin{aligned} \Sigma \boldsymbol{\lambda} - \mathbf{c} - m\mathbf{1} &= 0 \\ \boldsymbol{\lambda}^T \mathbf{1} - 1 &= 0. \end{aligned} \tag{1.7}$$

Despejando  $\boldsymbol{\lambda}$  en la primera ecuación del sistema (1.7), se obtiene

$$\boldsymbol{\lambda} = \Sigma^{-1}(\mathbf{c} + m\mathbf{1}). \tag{1.8}$$

Reemplazando esta expresión en la segunda ecuación del sistema (1.7) se encuentra

$$\begin{aligned} & (\Sigma^{-1}(\mathbf{c} + m\mathbf{1}))^T \mathbf{1} = 1 \\ & (\Sigma^{-1}\mathbf{c} + \Sigma^{-1}m\mathbf{1})^T \mathbf{1} = 1 \\ & \mathbf{1}^T (\Sigma^{-1}\mathbf{c}) + \mathbf{1}^T (\Sigma^{-1}m\mathbf{1}) = 1 \\ & \mathbf{1}^T (\Sigma^{-1}m\mathbf{1}) = 1 - \mathbf{1}^T (\Sigma^{-1}\mathbf{c}) \\ & m = (1 - \mathbf{1}^T (\Sigma^{-1}\mathbf{c})) (\mathbf{1}^T \Sigma^{-1} \mathbf{1})^{-1} \\ & m = \frac{1 - \mathbf{1}^T (\Sigma^{-1}\mathbf{c})}{\mathbf{1}^T \Sigma^{-1} \mathbf{1}} \end{aligned} \tag{1.9}$$

Sustituyendo (1.9) en la ecuación (1.8) se obtiene

$$\begin{aligned}
\boldsymbol{\lambda} &= \Sigma^{-1} \left( \mathbf{c} + \frac{1 - \mathbf{1}^T(\Sigma^{-1}\mathbf{c})}{\mathbf{1}^T\Sigma^{-1}\mathbf{1}} \mathbf{1} \right) \\
\boldsymbol{\lambda}^T &= \left( \mathbf{c} + \mathbf{1} \frac{1 - \mathbf{1}^T(\Sigma^{-1}\mathbf{c})}{\mathbf{1}^T\Sigma^{-1}\mathbf{1}} \right)^T (\Sigma^{-1})^T \\
\boldsymbol{\lambda}^T &= \left( \mathbf{c} + \mathbf{1} \frac{1 - \mathbf{1}^T(\Sigma^{-1}\mathbf{c})}{\mathbf{1}^T\Sigma^{-1}\mathbf{1}} \right)^T \Sigma^{-1}.
\end{aligned} \tag{1.10}$$

De acuerdo con la solución obtenida en (1.10), el predictor en (1.6) es definido por

$$\begin{aligned}
Y_0^* &= \sum_{i=1}^n \lambda_i Y_i \\
&= \boldsymbol{\lambda}^T \mathbf{Y} \\
&= \left[ \left( \mathbf{c} + \mathbf{1} \frac{1 - \mathbf{1}^T(\Sigma^{-1}\mathbf{c})}{\mathbf{1}^T\Sigma^{-1}\mathbf{1}} \right)^T \Sigma^{-1} \right] \mathbf{Y}
\end{aligned}$$

Haciendo algunas manipulaiones de álgebra se obtiene que

$$Y_0^* = \hat{\mu} + \mathbf{c}^T \Sigma^{-1} (\mathbf{Y} - \mathbf{1} \hat{\mu}), \tag{1.11}$$

donde  $\hat{\mu}$  es el estimador de mínimos cuadrados generalizados de  $\mu$  en la ecuación (1.5). La varianza del predictor en (1.11) está dada por

$$\sigma_p^2 = \sigma^2 - \mathbf{c}^T \Sigma^{-1} \mathbf{c} + \frac{(1 - \mathbf{1}^T \Sigma^{-1} \mathbf{c})^2}{(\mathbf{1}^T \Sigma^{-1} \mathbf{1})}. \tag{1.12}$$

### Observación

- Del modelo lineal general  $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$  se tiene que el estimador de mínimos cuadrados generalizados del vector de parámetros es  $\boldsymbol{\beta} = (\mathbf{X}^T \Sigma^{-1} \mathbf{X})^{-1} (\mathbf{X}^T \Sigma^{-1} \mathbf{Y})$ . Definiendo  $\mathbf{X} = \mathbf{1}$  y  $\boldsymbol{\beta} = \mu$ , se obtiene que  $\hat{\mu} = (\mathbf{1}^T \Sigma^{-1} \mathbf{1})^{-1} (\mathbf{1}^T \Sigma^{-1} \mathbf{Y})$ .

Ahora considérese el caso de predicción teniendo una muestra aleatoria. Sean  $Y_1, \dots, Y_n$  variables aleatorias independientes e idénticamente distribuidas, con  $Y_i \sim N(\mu, \sigma^2)$ . Planteando el mismo predictor dado en (1.6) y reemplazando  $\Sigma = \sigma^2 \mathbf{I}$  y  $\mathbf{c} = \mathbf{0}$  en las ecuaciones

(1.8) y (1.9), se encuentra que  $m = (\mathbf{1}^T(\sigma^2\mathbf{I})^{-1}\mathbf{1})^{-1}$  y que

$$\begin{aligned}\boldsymbol{\lambda} &= (\sigma^2\mathbf{I})^{-1}(\mathbf{1}^T(\sigma^2\mathbf{I})^{-1}\mathbf{1})^{-1}\mathbf{1} \\ &= \begin{pmatrix} 1/\sigma^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1/\sigma^2 \end{pmatrix} \begin{pmatrix} \sigma^2/n \\ \vdots \\ \sigma^2/n \end{pmatrix} = \begin{pmatrix} 1/n \\ \vdots \\ 1/n \end{pmatrix}. \end{aligned} \quad (1.13)$$

Al sustituir (1.13) en (1.6) se obtiene

$$\begin{aligned}Y_0^* &= \sum_{i=1}^n \lambda_i Y_i \\ &= \frac{1}{n} \sum_{i=1}^n Y_i = \bar{Y}. \end{aligned} \quad (1.14)$$

Tomando  $\Sigma = \sigma^2\mathbf{I}$  y  $\mathbf{c} = \mathbf{0}$  en (1.12) se obtiene que  $\sigma_p^2 = \sigma^2(1 + 1/n)$ , es decir la varianza de predicción del modelo bajo independencia.

## 1.5. Implementación en R

```
#####
# Ejemplos del test de Mantel para aleatoriedad versus
# asociación espacial
#####

library(ape) # Para el test de Mantel
library(spdep)
library(geoR)
library(gstat)
library(MASS) # Para simulación de normales

# Dos ejemplos bajo aleatoriedad

q1 <- matrix(runif(36), nrow = 6)
q2 <- matrix(runif(36), nrow = 6)
mantel.test(q1, q2, graph = TRUE,
            main = "Mantel test: a random example with 6 X 6 matrices",
            xlab = "Mantel-statistic", ylab = "Density",
            sub = "The vertical line shows the observed M-statistic")
```

```
q1 <- matrix(rnorm(36, 10, 1), nrow = 6)
q2 <- matrix(rnorm(36, 30, 3), nrow = 6)
mantel.test(q1, q2, graph = TRUE,
            main = "Mantel test: a random example with 6 X 6 matrices",
            xlab = "Mantel-statistic", ylab = "Density",
            sub = "The vertical line shows the observed M-statistic")

# Un ejemplo donde no hay aleatoriedad espacial
# Se simulan datos de una normal multivariada
# con una estructura de covarianza espacial

grid<-expand.grid(seq(0,9),seq(0,9))    # sitios
plot(grid)
distancia<-dist(grid,diag=T,upper=T)    # distancia entre los sitios ( $W_{ij}$ )
distancia<-as.matrix(distancia)
covarianza<-4*exp(-3*distancia/6)      # Modelo de covarianza espacial
covarianza<-as.matrix(covarianza)

medias.cte<-rep(20,100)                  # Vector de medias
medias.ncte<-2*grid[,1]+3*grid[,2]      # Vector de medias no constante
normal.cte<-mvrnorm(1,medias.cte,covarianza)
# Simulación de 100 valores de una normal multivariada
dist_cte=as.matrix(dist(normal.cte,diag=T,upper=T))
# Matriz de distancia entre los valores simulados

normal.ncte<-mvrnorm(1,medias.ncte,covarianza)
dist_ncte=as.matrix(dist(normal.ncte,diag=T,upper=T))
# Matriz de distancia entre los valores simulados

# Test de Mantel bajo correlación (se debe rechazar  $H_0$ )
# con datos estacionarios

mantel.test(distancia, dist_cte, graph = TRUE,
            main = "Mantel test: a simulated data set with spatial correlation",
            xlab = "M-statistic", ylab = "Density",
            sub = "The vertical line shows the observed M-statistic")

# Test de Mantel bajo correlación (se debe rechazar  $H_0$ )
# con datos no estacionarios (vector de medias no constante)

mantel.test(distancia, dist_ncte, graph = TRUE,
```

```

main = "Mantel test: a simulated data set with spatial correlation",
xlab = "M-statistic", ylab = "Density",
sub = "The vertical line shows the observed M-statistic")

#####
# Ejemplos Variogramas
#####
grid<-expand.grid(seq(0,9),seq(0,9))    # sitios
plot(grid)
distancia<-dist(grid,diag=T,upper=T)      # distancia entre los sitios (Wij)
distancia<-as.matrix(distancia)
covarianza<-4*exp(-3*distancia/6)        # Modelo de covarianza espacial
covarianza<-as.matrix(covarianza)

medias.cte<-rep(20,100)                  # Vector de medias
medias.ncte<-2*grid[,1]+3*grid[,2]       # Vector de medias no constante
normal.cte<-mvrnorm(1,medias.cte,covarianza)
# Simulación de 100 valores de una normal multivariada
dist_cte<-as.matrix(dist(normal.cte,diag=T,upper=T))
# Matriz de distancia entre los valores simulados

normal.ncte<-mvrnorm(1,medias.ncte,covarianza)
dist_ncte<-as.matrix(dist(normal.ncte,diag=T,upper=T))
# Matriz de distancia entre los valores simulados

datoscte<-cbind(grid[,1],grid[,2],normal.cte)
datosncte<-cbind(grid[,1],grid[,2],normal.ncte)

datoscte<-as.geodata(datoscte,coords=1:2,var=3)
datosncte<-as.geodata(datosncte,coords=1:2,var=3)

plot(datoscte)
plot(datosncte)

#Construcción del variograma cloud

max(dist_cte)

cloud1<-variog(datoscte,option="cloud",max.dist=12)
cloud2<-variog(datosncte,option="cloud",max.dist=12)
par(mfrow=c(1,2))
plot(cloud1,main="Datos con media constante")
plot(cloud2,main="Datos con media no constante")

```

```
#Construcción del variograma Bin

bin1<-variog(datoscte,option="bin",max.dist=9)
plot(bin1) #modelo exponencial c(h)=sigma^2*(-3h/phi)

cloud21<-variog(datosnkte,option="cloud",max.dist=12)
cloud22<-variog(datosnkte,option="cloud",max.dist=12,trend="1st")
par(mfrow=c(1,2))
plot(cloud21,main="Datos con media no constante con tendencia")
plot(cloud22,main="Datos con media no constante sin tendencia")

bin21<-variog(datosnkte,option="bin",max.dist=9)
bin22<-variog(datosnkte,option="bin",max.dist=9,trend="1st")
par(mfrow=c(1,2))
plot(bin21,main="Datos con media no constante con tendencia")
plot(bin22,main="Datos con media no constante sin tendencia")
```

# Capítulo 2

## Patrones Puntuales

La realización de un proceso puntual consiste en un arreglo (patrón) de puntos en un conjunto aleatorio  $D$ . Estos puntos se denominan los eventos del proceso. Si los eventos son observados sólo parcialmente, el patrón se llama un patrón muestreado (Sampled point mapped). Cuando todos los eventos de la realización se registran, se dice que este es puntual.

Sea  $D$  un conjunto aleatorio, el experimento que genera una realización particular puede ser visto como un sorteo de lugares en  $D$  de los eventos que son observados. Los patrones puntuales son realizaciones de experimentos aleatorios y se distingue entre patrones (completamente) aleatorios, agrupados (espacialmente agregados), y los normales, lo cual no debe conducir a la falsa impresión de que los dos últimos tipos de patrones no contienen ningún tipo de aleatoriedad.

Un patrón puntual se llama completamente aleatorio si se cumplen los siguientes requisitos. El promedio de eventos por unidad de área, la intensidad  $\lambda(s)$ , es homogénea a lo largo  $D$ , el número de eventos en dos subregiones que no se solapan,  $A_1$  y  $A_2$  son independientes, y el número de eventos en cualquier subregión sigue una distribución de Poisson. Por lo tanto, los eventos se distribuyen uniforme e independiente a lo largo del dominio, proceso que se reconoce matemáticamente como un proceso Poisson homogéneo, y que sirve como hipótesis nula para muchas investigaciones estadísticas en los patrones puntuales.

## 2.1. Tipos de Patrones

Muchos eventos pueden ser independientes en subregiones no traslapadas, pero la intensidad  $\lambda(s)$  no es homogénea a lo largo de  $D$ . Otros eventos pueden estar localizados en regiones donde la intensidad es grande, y unos pocos estarán localizados en regiones donde  $\lambda(s)$  es pequeña. Los eventos pueden ocurrir con una intensidad constante (promedio)  $\lambda(s) \equiv \lambda$  pero mostrar algún tipo de interacción. La presencia de un evento puede atraer o alejar otros eventos cercanos. Por lo tanto, derivaciones de un patron espacial completamente aleatorio son diferenciados en agregados (agrupados) o regulares (Figura 2.1). La razón de esta diferenciación puede ser la variación deterministica espacial de la función de intensidad  $\lambda(s)$  o los resultados de un elemento estocástico.

Diferentes modelos de patrones puntuales logran muchas derivaciones de los patrones espaciales completamente aleatorios en diferentes maneras. En ese punto se pueden distinguir patrones aleatorios, regulares, o agregados. En un patron agregado, la distancia media entre los eventos  $s_i$  y su vecino más cercano es más pequeña que la misma distancia media en un patron espacial completamente aleatorio. De manera similar, en un patron regular la distancia media entre un evento y su vecino más cercano es más grande que la esperada bajo aleatoriedad. Muchos de los patrones observados son realizaciones instantáneas de un proceso que envuelve el tiempo. La evolución temporal afecta el grado en el cual los eventos son agregados o agrupados. Por otro lado muchos de los patrones observados incluyen observaciones de covariables (Figura 2.2 a.)), patrones llamados *Patron puntual marcado*, por ejemplo magnitud y profundidad asociadas a los datos de sismos en Colombia (Figura 2.3), donde el radio del circulo es proporcional a la marca. Adicionalmente en muchas mediciones se pueden presentar el efecto borde (Figura 2.2 b.)), problema que lleva a sobreestimar o subestimar, debido a la inclusión o no de marcas en las estimaciones.

## 2.2. Estimación de la intensidad

Incluso para procesos homogéneos es útil estimar la intensidad de los eventos localizados. En la practica, las variables espaciales estiman  $\hat{\lambda}(s)$  como la intensidad en una

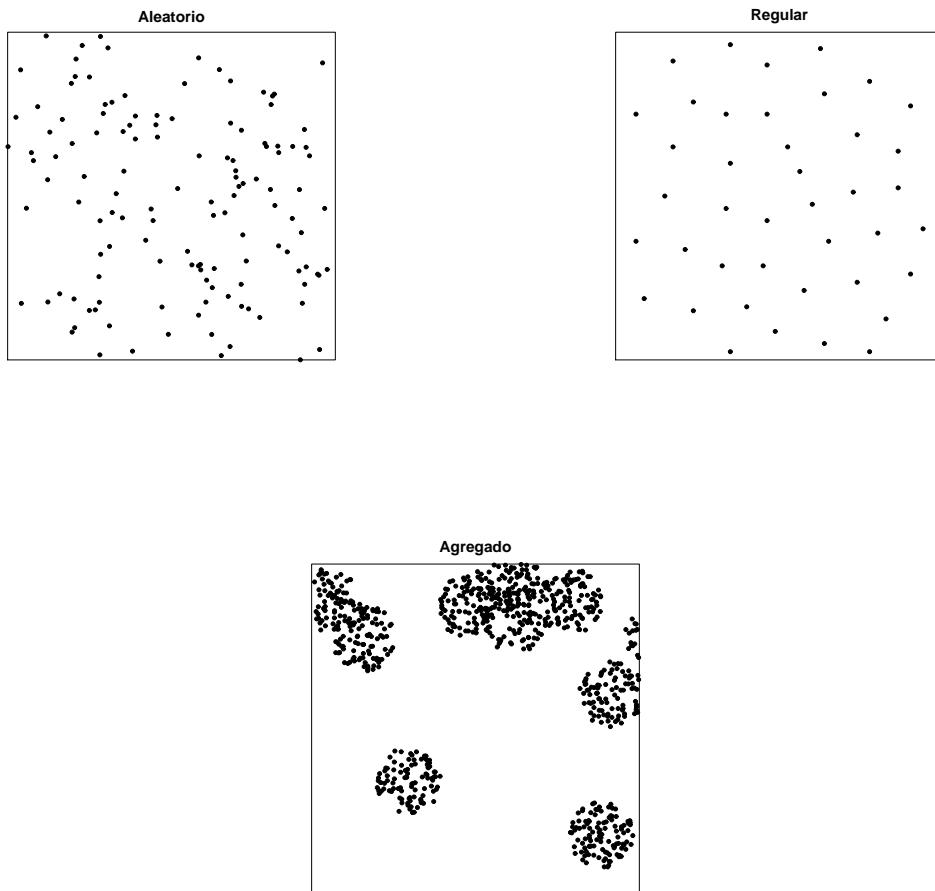


Figura 2.1: Ejemplos simulados de patrones puntuales con distribución aleatoria (*izquierda*), regular (*derecha*) y agregada (*centro*).

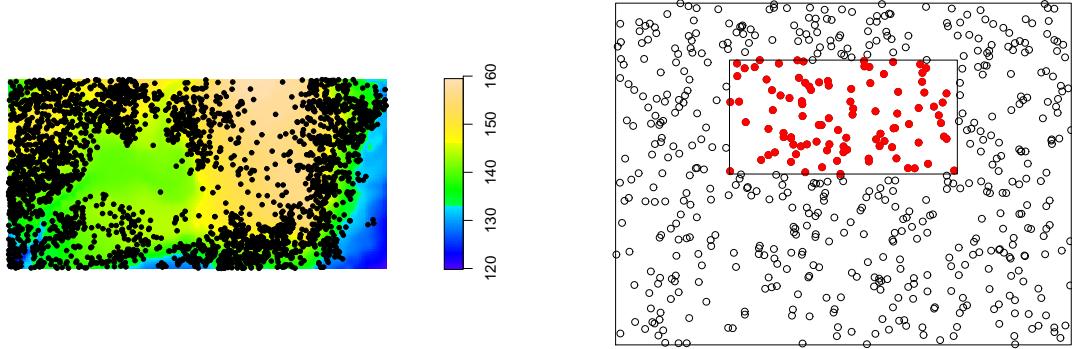


Figura 2.2: a.) Ejemplos de patrones que incluyen covariables y b.) efecto de borde (*derecha*). En la figura de la izquierda la escala de valores corresponde a la elevación en metros

localización ( $s$ ), la cual es obtenida por métodos no paramétricos, de suavizamiento de cuadrantes, o por métodos de estimación de la intensidad.

### 2.2.1. Definición de Intensidad

Como se mencionó anteriormente la intensidad es un factor importante en el desarrollo del análisis espacial de un área determinada, ya que es de mucha importancia conocer el número promedio de eventos ocurridos por unidad de área, por lo cual se define:

$$\lambda = \text{Número de eventos que ocurren por unidad de área.}$$

En general, si las áreas son muy pequeñas, se llega a la definición de intensidad

$$\lambda(s) = \lim_{|s| \rightarrow 0} \frac{E(N(|s|))}{|s|}$$

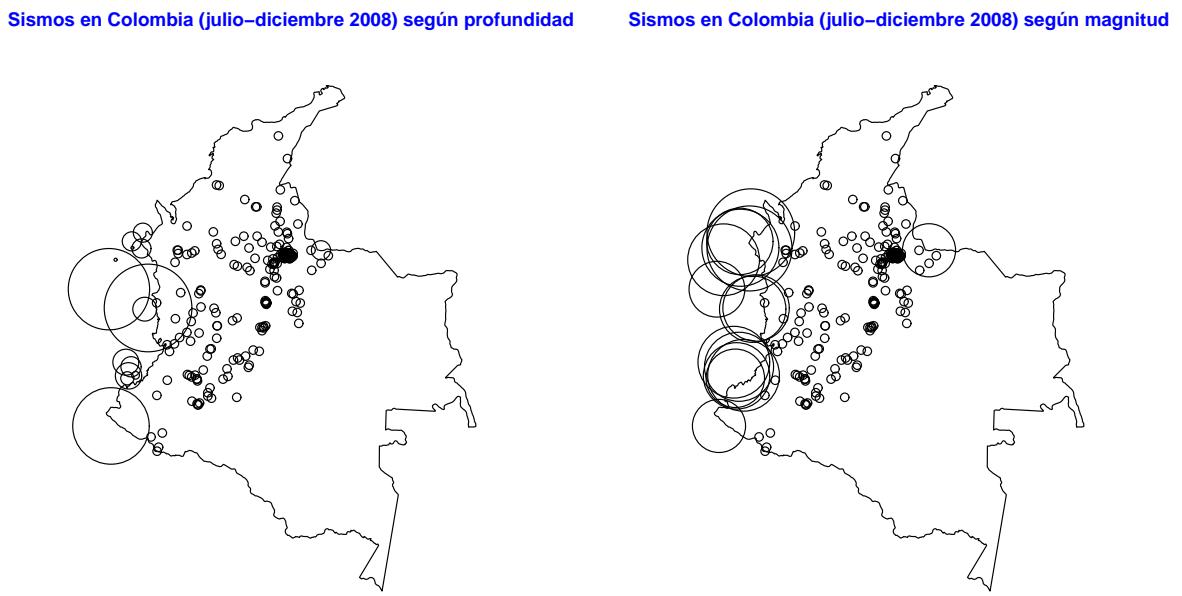


Figura 2.3: Marcas asociadas a los datos de sismos en Colombia

Si el patrón es completamente aleatorio:

$$\lambda(s) = \lambda = \frac{n}{|A|}$$

### 2.2.2. Estimación Kernel

La estimación Kernel es un método no paramétrico de estimación, en el cual, en la formación de función de densidad no se imponen asunciones tan rígidas a los datos como sucede en el planteamiento paramétrico.

Se usa la estimación tipo Kernel para:

- Densidad  $\hat{f}(x)$
- Regresión  $\hat{Y} = \hat{m}(x) + e$
- Intensidad  $\hat{\lambda}_s$

#### Función Kernel

$K(u)$ , función *Kernel* o *Núcleo*, es una función que satisface ciertas condiciones de regularidad, generalmente es un función de densidad simétrica alrededor de cero, que debe cumplir:

$$1. \int k(u) = 1$$

$$2. \int uk(u) = 0$$

Algunas funciones Kernel son:

- Gausiana

$$k(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

- Cuadrática

$$K(x) = \begin{cases} 0,75(1 - x^2) & |x| \leq 1 \\ 0 & \text{en otro caso} \end{cases}$$

- Mínima varianza

$$K(x) = \begin{cases} 3/8(3 - 5x^2) & |x| \leq 1 \\ 0 & \text{en otro caso} \end{cases}$$

### Kernel Bivariado

Kernel Producto

$$k(x_1, x_2) = k(x_1)k(x_2)$$

### Kernel en estimación de la densidad

$$f(x) = F'(x) = \lim_{h \rightarrow 0} \frac{F(x+h) - F(x)}{h}$$

Se divide  $\mathbb{R}$  en intervalos de igual amplitud  $b$ , y se llama a los extremos  $b_j$ , entonces cada  $x$  pertenece a uno de los intervalos  $(b_j, b_{j+1})$ .

$$\begin{aligned} \hat{f}(x) &= \frac{F(b_{j+1}) - F(b_j)}{b} \\ &= \frac{\frac{N^o\{x_i \leq b_{j+1}\}}{n} - \frac{N^o\{x_i \leq b_j\}}{n}}{b} \\ &= \frac{N^o\{b_j \leq x_i \leq b_{j+1}\}}{nb} \\ &= \frac{1}{nb} \sum_{i=1}^n I_{[x-b/2, x+b/2]}(x_i) \\ &= \frac{1}{nb} \sum_{i=1}^n I_{[-1/2, 1/2]}(\frac{x_i - x}{b}) \end{aligned}$$

$$I_{[x-b/2, x+b/2]} = \begin{cases} 1 & x - \frac{b}{2} < x_i < x + \frac{b}{2} \\ 0 & \text{en otro caso} \end{cases}$$

$$I_{[-1/2, 1/2]}(\frac{x_i - x}{b}) = \begin{cases} 1 & -\frac{1}{2} < \frac{x_i - x}{b} < \frac{1}{2} \\ 0 & \text{en otro caso} \end{cases}$$

$$\hat{f}(x) = \frac{1}{nb} \sum_{i=1}^n f(u)$$

$f(u)$  es una densidad uniforme  $U = \frac{x_i - x}{b} \sim U[-1/2, 1/2]$

Si  $h = b$  se tiene  $\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n k(\frac{x_i - x}{h})$

### 2.2.3. Estimación Kernel de la intensidad $\lambda(s)$

En el caso bi-dimensional, la función kernel univariada necesita ser reemplazada por una función que se acomode a las coordenadas. Un producto de funciones kernel puede ser obtenido multiplicando dos funciones kernel univariadas. La elección de estas puede hacerse a conveniencia, lo que implica la ausencia de interacción entre coordenadas. Los anchos de banda pueden ser elegidos diferente de las dos dimensiones, pero los más comunes son los kernels esféricos.

En un patrón completamente aleatorio,

$$\hat{\lambda} = \frac{n}{|A|} = \frac{\text{Nº eventos}}{\text{Área de A}}$$

si  $x_i$  y  $y_i$  son las coordenadas de localización del sitio  $s_i$ , entonces el producto kernel conduce a una estimador de la intensidad dado por:

$$\begin{aligned}\hat{\lambda}(s) &= \frac{1}{|A|h_x h_y} \sum_{i=1}^n K\left(\frac{x - x_i}{h_x}\right) K\left(\frac{y - y_i}{h_y}\right) \\ &= \frac{1}{|A|} \sum_{i=1}^n \frac{1}{h} K\left(\frac{s - s_i}{h}\right)\end{aligned}$$

donde,  $h_x$  y  $h_y$  son los anchos de banda en las direcciones respectivas del sistema de coordenadas.

Por otra parte si se presentan problemas con el efecto de borde  $\delta(S)$ , se obtiene un estimador que lo incluya:

$$\hat{\lambda}(s) = \frac{1}{\delta(S)} \sum_{i=1}^n \frac{1}{h^2} K\left(\frac{s - s_i}{h}\right)$$

$$\text{Con } \delta(S) = \int_A \frac{1}{h^2} K\left(\frac{s-u}{h}\right) du$$

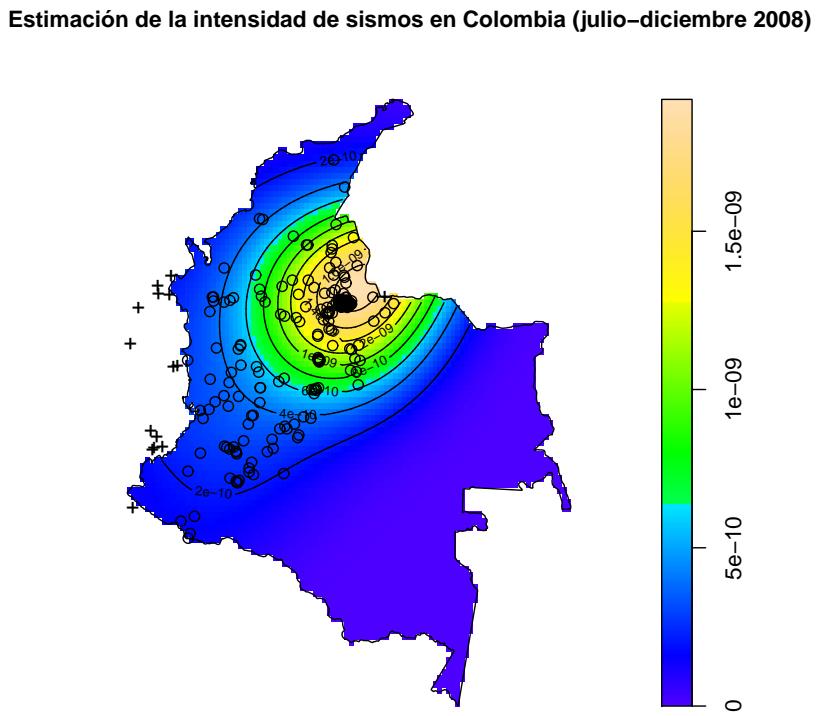
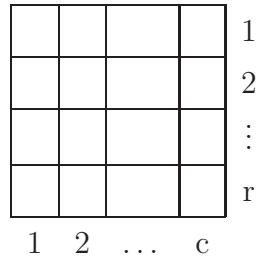


Figura 2.4: Estimación de la intensidad de sismos en Colombia (Julio-Diciembre 2008).

## 2.3. Métodos basados en cuadrantes

Otro tipo de métodos para probar la aleatoriedad espacial de un patrón son basados en la división del dominio  $D$  en subregiones no traslapadas (cuadrantes)  $A_1, \dots, A_k$  de igual tamaño.

Sean  $A_1, A_2, \dots, A_K$  tal que  $\bigcup_{i=1}^k A_i = D$  donde  $A_i$  y  $A_j$  no se traslanan.



$n$ : Número total de eventos

$n_{ij}$ : número total de eventos en la cuadrícula  $ij$ .

### 2.3.1. Test de bondad de ajuste $\chi^2$

Esta prueba de ajuste estadístico, prueba la hipótesis nula, que los  $n$  puntos son distribuidos uniformemente independientes a lo largo de  $D$ , o en otras palabras, los conteos de sitios por cuadrante son variables Poisson independientes con media común.

El estadístico de prueba esta dado por:

$$X^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \sim \chi^2_{(rc-1)} \quad (2.1)$$

Para el caso considerado

$$O_{ij} = n_{ij},$$

$$E_{ij} = \bar{n} = \frac{n}{rc},$$

$$\hat{\lambda} = \frac{n}{|A|},$$

$$N(A) \sim Poisson(\lambda A)$$

$$\Rightarrow X^2 = \sum_{i=1}^n \sum_{j=1}^n \frac{(n_{ij} - \bar{n})^2}{\bar{n}} \sim \chi^2_{(rc-1)}$$

### Índice de dispersión

Si  $X \sim Poisson(\lambda)$ ,  $E(X) = V(X) = \lambda$

$$I = \frac{V(X)}{E(X)} \approx 1$$

En un patrón completamente aleatorio:

$$N(A) \sim Poisson(\lambda A)$$

$$I = \frac{V(N(A))}{E(N(A))} = \frac{S^2}{\bar{n}} = \frac{\sum_{rc=1}^{rc-1} (n_{ij} - \bar{n})^2}{\bar{n}}$$

$(I - 1)$  = Índice de tamaño del cluster

$(I - 1) \approx 0$  el patrón es completamente aleatorio

$> 0$  el patrón es agregado

$< 0$  el patrón es regular

## 2.4. Métodos basados en distancias

La elección de la forma y el número de cuadrantes para probar la hipótesis de aleatoriedad basados en conteos por áreas es un elemento que puede influenciar los resultados. Las pruebas estadísticas basadas en distancias entre eventos o entre puntos muestrados y eventos elimina estas características. En esta sección se nombrar pruebas que tienen en cuenta tales distancias entre eventos.

### 2.4.1. Función $G(h)$

Esta función tiene en cuenta la mínima distancia entre eventos (distancia al vecino mas cercano). Se define:

$d_i$ =Mínima distancia entre un evento y sus vecinos

$D$  = Variable aleatoria distancia al sitio del evento más cercano

$G(D)$  = Función de distribución de D.

### Estimación de $G(D)$

Función de distribución empírica:

$$\hat{G}(D) = \frac{N^{\circ}\{d_i \leq d\}}{n}$$

### Distribución teórica de $G(D)$

$$G(d) = P(D \leq d) = 1 - P(D > d) = 1 - P(\text{No hay eventos en } A = \pi d^2)$$

Si  $N(A) \sim Poisson(\lambda \pi d^2)$  entonces  $P(N(A) = 0) = e^{-\lambda \pi d^2}$

por lo tanto

$$G(d) = 1 - e^{-\lambda \pi d^2}$$

de donde:

- Si  $\hat{G}(d)$  crece rápidamente en distancias cortas, los eventos son agregados.
- Si los eventos son regularmente espaciados,  $\hat{G}(d)$  crece lentamente hasta cierta distancia (espacio eventos) y después crece rápidamente.

### Bandas de confianza

Adicionalmente se hace un cálculo de las bandas de confianza para la función en donde se calcula:

1.  $\hat{G}_0(h)$  = Función  $G(h)$  calculada con los datos observados
2.  $\hat{G}_1(h), \hat{G}_2(h), \dots, \hat{G}_g(h)$  = Función  $G(h)$  calculada con  $g$  realizaciones de un proceso completamente aleatorio PCA.
3.  $G_L(h) = \min_{i=1,\dots,g} \hat{G}_i(h)$  y  $G_U(h) = \max_{i=1,\dots,g} \hat{G}_i(h)$

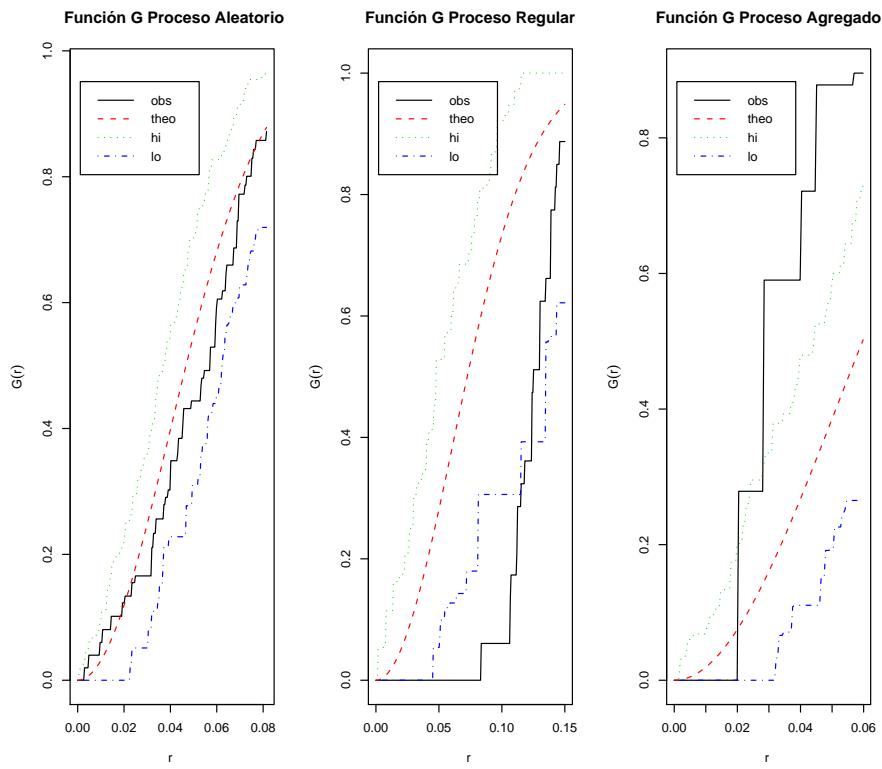


Figura 2.5: Gráficas del la función G para cada tipo de proceso

### 2.4.2. Función F(h)

Tiene en cuenta la mínima distancia punto-evento, donde un punto es un evento escogido aleatoriamente dentro de la región de estudio. Para su construcción se siguen los siguientes pasos:

1. Seleccione aleatoriamente m puntos  $\{p_1, p_2, \dots, p_m\}$ .
2. Calcule  $d_i = d(p_i, s_i)$ , la distancia de cada punto escogido al sitio del evento más cercano.

Sea  $D$  la variable aleatoria mínima distancia de un punto a un evento, entonces:

$$F(d) = P(D \leq d) \text{ y } \hat{F}(d) = \frac{N^{\circ}\{d_i \leq d\}}{m}$$

Luego la función teórica será

$$F(d) = 1 - e^{-\lambda\pi d^2}$$

de donde:

- Si  $\hat{F}(d)$  crece lentamente al comienzo y rápidamente para distancia largas, el patrón es agregado.
- Si  $\hat{F}(d)$  crece rápido al comienzo, el patrón es regular.

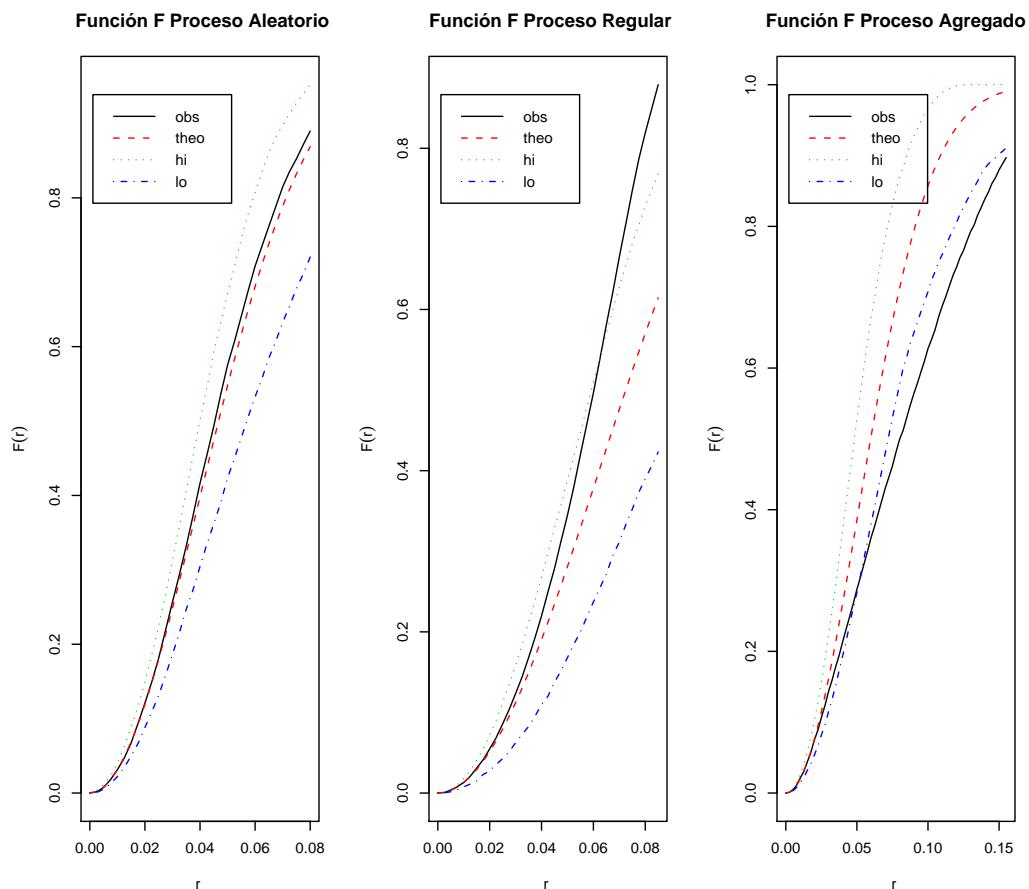


Figura 2.6: Gráficas de la función  $F$  para cada tipo de proceso

En un patrón completamente aleatorio

$$F(x) = 1 - e^{-\lambda\pi x^2}$$

donde,  $x$  :Distancia mínima de un punto a un evento.

$$E(x) = \frac{1}{2\sqrt{\lambda}} y V(x) = \frac{4 - \pi}{4\pi\lambda}$$

Si se tiene  $x_1, \dots, x_m$

$$E(\bar{x}) = \frac{1}{2\sqrt{\lambda}} y V(\bar{x}) = \frac{4 - \pi}{4\pi\lambda m}$$

Ahora si se define  $G(Y) = 1 - e^{-\lambda\pi d^2}$

Con  $Y$  =Miníma distancia al evento más cercano.

$Y_1, \dots, Y_m$

$$E(Y) = \frac{1}{2\sqrt{\lambda}} y V(Y) = \frac{4 - \pi}{4\pi\lambda}$$

La función de distribución de G(Y) y F(Y) son iguales bajo aleatoriedad.

### 2.4.3. Función K de Ripley

La función  $K$  de Ripley (1976) es una función de  $\lambda_2$  para procesos estacionarios e isotropicos. También es conocida como la medida reducida del segundo momento (Cressie (1993)), y es la función reducida del segundo momento. Está dada por:

$$k(d) = \frac{E(\text{Número de eventos adicionales a una distancia } d \text{ de un evento elegido aleatoriamente})}{\lambda}$$

$\lambda$  =Intensidad

$$\hat{K}(d) = \frac{\sum \sum_{i \neq j} I_d(d_{ij})}{\hat{\lambda}_n}$$

con

- $\hat{\lambda} = n/A$

- A=Area de la región R
- $d_{ij}$  =Distancia entre el i-ésimo y el j-ésimo evento
- $$\begin{cases} I_d(d_{ij}) = 1 & \text{si } d_{ij} < 0 \\ 0 & \text{en otro caso} \end{cases}$$

Bajo aleatoriedad

$$K(d) = \frac{\lambda\pi d^2}{\lambda} = \pi d^2$$

De donde

- Si  $\hat{K}(d) < \pi d^2$  hay regularidad
- Si  $\hat{K}(d) > \pi d^2$  hay agregación

#### 2.4.4. Función L

Otra función de  $\lambda_2$  usda en el estudio de dependencia en patrones puntuales es fácil de calcular basada en la función  $K(h)$ , y es llamada la función  $L$ , dad por:

$$L(d) = \sqrt{\frac{K(d)}{\pi}}$$

bajo aleatoriedad

$$L(d) = \sqrt{\frac{\pi d^2}{\pi}} = d$$

Si el patrón es aleatorio no debe haber desviación de la recta.

### 2.5. Modelos

El proceso Poisson homogéneo da un punto de inicio natural para investigación estadística de un patron puntual observado. El rechazo de la hipótesis de aleatoriedad no es

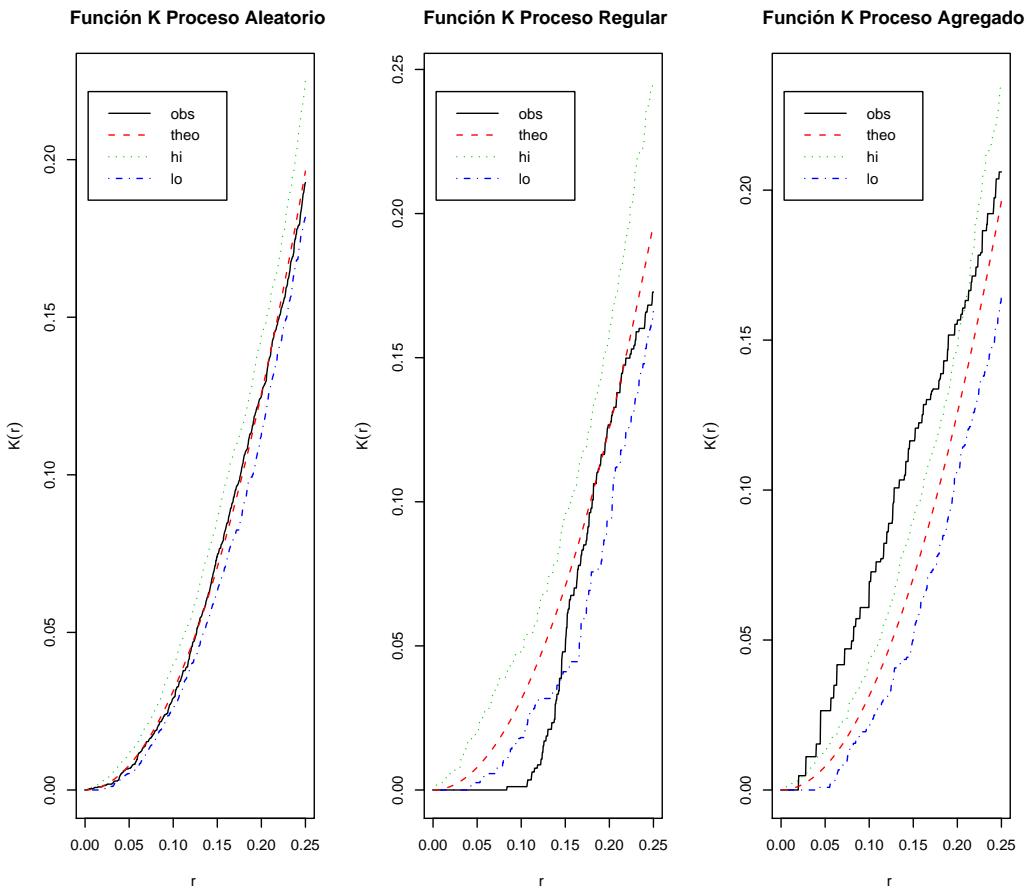


Figura 2.7: Gráficas del la función  $K$  para cada tipo de proceso

sorprendente en muchas aplicaciones y es natural preguntarse, Qué tipo de patrón es?, si las pruebas sugieren un patrón agregado. En estos casos se hace necesaria la búsqueda de un modelo apropiado para cada situación. A continuación se presentan un amplio número de modelos que son alternativas para procesos regulares y agregados.

### 2.5.1. Proceso Poisson no Homogéneo

A menudo son más realistas los modelos basados en procesos de Poisson no homogéneos. El modelo más simple cuando el proceso no es estacionario es el proceso de Poisson nohomogéneo, que se obtiene sustituyendo la intensidad constante  $\lambda$ , de un proceso de Poisson

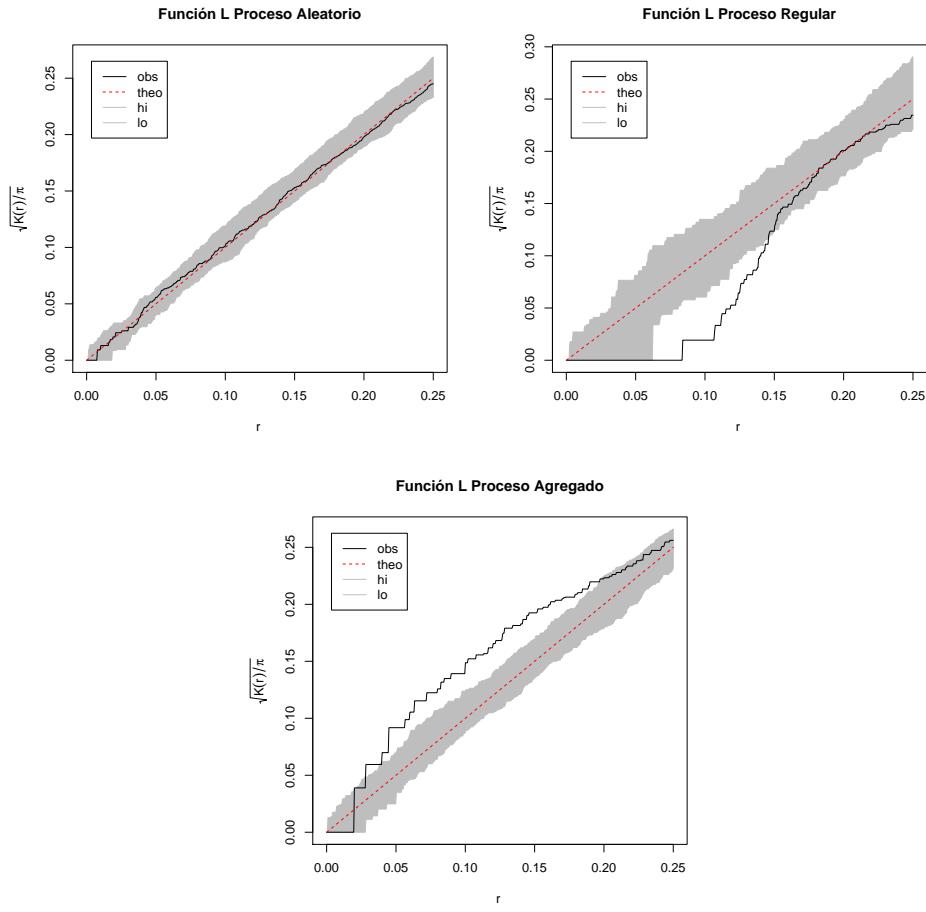


Figura 2.8: Gráficas del la función L para cada tipo de proceso

por una función de intensidad variable  $\lambda(x)$ . Así un proceso de Poisson no homogéneo está definido como sigue.

Sean:

- $N(A)$ : Número de eventos en la región A
- $|A|$ : Área de la región A
- $E(N(A)) = \lambda(A) = \int_A \lambda(s)ds$

- Intensidad:  $\lambda(s) = \lim_{|s| \rightarrow 0} \frac{E(N(s))}{|s|}$

Un proceso de Poisson es no homogéneo si:

1.  $N(A) \sim Poisson(\lambda(A))$ .
2. Si  $A_1$  y  $A_2$  son dos regiones no traslapadas,  $N(A_1)$  y  $N(A_2)$  son independientes.
3. Dado  $N(A) = n$ , las localizaciones  $\{S_1, S_2, \dots, S_n\}$  son realizaciones independientes distribuidas uniformemente.

$$f_A(s|n) = \frac{\lambda(s)}{\int_A \lambda(s) ds}$$

### Verosimilitud

$$\begin{aligned} P(N(A) = n) &= \frac{e^{-\lambda(A)} (\lambda(A))^n}{n!} \\ f_A(s_i|n) &= \frac{\lambda(s_i)}{\int_A \lambda(s) ds} \\ \Rightarrow f_A(s_1, s_2, \dots, s_n|n) &= \prod_{i=1}^n f_A(s_i|n) \\ &= \frac{\prod_{i=1}^n \lambda(s_i)}{(\int_A \lambda(s) ds)^n} = \frac{\prod_{i=1}^n \lambda(s_i)}{(\lambda(A))^n} \end{aligned}$$

Así la función de verosimilitud está dada por:

$$\begin{aligned} f_A(s_1, s_2, \dots, s_n, n) &= f_A(s_1, s_2, \dots, s_n|n)P(N(A) = n) \\ &= \frac{e^{-\lambda(A)} (\lambda(A))^n}{n!} \frac{\prod_{i=1}^n \lambda(s_i)}{(\lambda(A))^n} \\ &= \frac{e^{-\lambda(A)} \prod_{i=1}^n \lambda(s_i)}{n!} \\ \ln(f_A(s_1, s_2, \dots, s_n, n)) &= -\lambda(A) + \sum_{i=1}^n \ln(\lambda(s_i)) - \ln n! \\ &= \sum_{i=1}^n \ln(\lambda(s_i)) - \int_A \lambda(s) ds - \ln n! \end{aligned}$$

Encontrar el máximo de esta función es muy complicado, por lo que se encuentran estimaciones de los parámetros a través de un modelo lineal generalizado de tal manera que:

$$g(\lambda(s_i)) = X\beta \text{ con } g = \ln(\cdot)$$

Para estimar  $\beta$  se usa el método de Newton-Raphson. El modelo lineal generalizado más simple es:

$$\ln(\lambda(s_i)) = \beta_0 + \beta_1 x_i + \beta_2 y_i$$

donde  $(x_i, y_i)$  son coordenadas espaciales.

### 2.5.2. Proceso Poisson Homogéneo

Los procesos Poisson homogéneos, son la base a partir de la que se construye la teoría de los procesos puntuales espaciales. Representan el mecanismo estocástico más sencillo que puede generar un diseño puntual espacial y se aplican como modelo ideal de los procesos completamente aleatorios que, a pesar de no acercarse a la práctica, resulta muy útil como descripción aproximada de algunos patrones observados. Los procesos de Poisson están definidos como sigue.

$$\begin{aligned}
\lambda(s) &= \lambda \\
\lambda(A) &= \int_A \lambda(s)ds = \int_A \lambda ds = \lambda|A| \\
P(N(A) = n) &= \frac{e^{-\lambda|A|}(\lambda|A|)^n}{n!} \\
f_A(s|n) &= \frac{\lambda(s)}{\int_A \lambda(s)ds} = \frac{\lambda}{\lambda|A|} = \frac{1}{|A|} \\
f_A(s_1, s_2, \dots, s_n | n) &= \prod_{i=1}^n 1/|A| = 1/|A|^n \\
f_A(s_1, s_2, \dots, s_n, n) &= \frac{1}{|A|^n} \frac{e^{-\lambda|A|}(\lambda|A|)^n}{n!} = \frac{e^{-\lambda|A|}\lambda^n}{n!} \\
\ln(f_A(s_1, s_2, \dots, s_n, n)) &= -\lambda(|A|) + n \ln(\lambda) - \ln(n!)
\end{aligned}$$

Así un estimador máximo verosímil para la intensidad esta dado por:

$$\begin{aligned}
\frac{d \ln f_A}{d \lambda} &= -|A| + \frac{n}{\lambda} = 0 \\
|A| &= \frac{n}{\lambda} \\
\hat{\lambda} &= \frac{n}{|A|}
\end{aligned}$$

### 2.5.3. Proceso de Cox

Si se quieren modelar fenómenos donde la distribución espacial de los eventos puede ser el resultado de variaciones estocásticas, es razonable pensar en la función de intensidad de un proceso de Poisson,  $\lambda(x)$ , como una realización de un proceso estocástico.

Un proceso de Cox es un proceso de Poisson en donde la intensidad es aleatoria.

$$N(A) = \int_A \Lambda(s)ds$$

Donde  $\Lambda(s)$  es una variable aleatoria generalmente distribución Gamma.

### Proceso Poisson/Gamma

$$\lambda(s) \sim \text{Gamma}$$

$$f(\Lambda(s)) = \frac{1}{\beta^\alpha \Gamma(\alpha)} \lambda(s)^{\alpha-1} e^{(-\lambda(s)/\beta)}$$

$$E(\Lambda(s)) = \alpha\beta \quad \text{var}(\Lambda(s)) = \alpha\beta^2$$

Si  $\lambda$  no depende de  $s$ , entonces el proceso de Cox es homogéneo. Si  $\lambda$  depende de  $s$ , entonces el proceso de Cox es no homogéneo.

### Simulación de un proceso de Cox

- 1. Se obtiene una realización  $\Lambda(s)$ , es decir, se obtiene  $\lambda(s)$
- 2. Dado que  $\Lambda(s) = \lambda(s)$ , se simula un proceso Poisson de media  $\lambda(s)$
- Se usa el método de descomposición de Cholesky

$$P((A) = n) = E_\Lambda \left( \frac{e^{-\Lambda(A)} (\Lambda(A))^n}{n!} \right)$$

$$\Rightarrow \lambda^*(s) = E(\Lambda(s)) \text{ donde } \lambda^*(s) \text{ es la media de las realizaciones de } \Lambda(s)$$

### Proceso de Cox Log-Gaussiano

$$\Lambda(s) = e^{z(s)}; \ln(z(n)) \sim N(\mu, \sigma^2)$$

Nota: Este proceso tiene una función k-Ripley conocida.

## Verosimilitud

$$\begin{aligned}
 P(N(A)) &= \frac{e^{-\Lambda(A)}(\Lambda(A))^n}{n!} \\
 f(s_1, s_2, \dots, s_n | n) &= \text{frac} \prod_{i=1}^n \Lambda(s_i)(\Lambda(A))^n \\
 f(s_1, s_2, \dots, s_n, n) &= \frac{e^{-\Lambda(A)} \prod_{i=1}^n \Lambda(s_i)}{n!} \\
 \ln(f(s_1, s_2, \dots, s_n, n)) &= \sum_{i=1}^n \ln(\Lambda(s_i)) - \int_a \Lambda(s) ds
 \end{aligned}$$

El modelo lineal generalizado está dado por:

$$\ln(\Lambda(s)) = X\beta$$

y por métodos bayesianos (cadenas de Markov de Montecarlo) se obtiene la estimación de  $\beta$

### 2.5.4. Proceso de Poisson Cluster

Los procesos de Poisson con clústers son modelos que incorporan explícitamente la formación de grupos y, por tanto, proporcionan la base adecuada para modelar patrones puntuales con eventos agrupados.

Son procesos que cumplen las siguientes condiciones:

1. Hay un proceso con "padres" que siguen un proceso Poisson con intensidad  $\lambda$ .
2. Cada parente produce un número aleatorio de "descendientes"  $M$  de acuerdo a una distribución.
3. Las posiciones de los descendientes respecto a los padres es aleatoria y sigue una distribución bivariada.
4. El proceso Poisson-Cluster corresponde a la posición de los descendientes.

Nota: Si  $f(s)$  es una normal bivariada, el proceso se llama "Thomasz tiene función K conocida.

$$f(x, y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{x^2 + y^2}{2\sigma^2}\right\}; \quad k(h) = \pi h^2 + \left(\frac{1 + \exp(-h^2/4\sigma^2)}{\lambda}\right)$$

La estimación de los parámetros ( $\lambda$  y  $\sigma^2$ ) se hace por la minimización de la función de contrastes.

### 2.5.5. Proceso de Inhibición Simple (MATHERN I)

Los modelos alternativos al de Poisson vistos hasta este momento tienden a producir diseños agrupados. Para describir un patrón regular el camino más natural consiste en imponer una distancia mínima entre dos eventos. Estos modelos se denominan procesos de inhibición simple y hay distintos modos de formularlos. Para obtener un proceso de este tipo se sigue lo siguiente.

1. Se obtiene un proceso de Poisson con intensidad constante
2. Se quitan todas las parejas separadas por una distancia menor que  $d$

#### Propiedades:

- $\lambda^* = \lambda e^{-\lambda\pi d^2}$
- $k(h) = \frac{2\pi}{\lambda} \int_0^h u \lambda_2(u) du; \quad h \geq 0$

$$\lambda_2(h) = \lambda_2(s_i, s_j) = \lim_{|s_i| \rightarrow 0, |s_j| \rightarrow 0} \frac{E(N(S_i)N(s_j))}{|s_i||s_j|}$$

$\lambda_2(h)$  función de intensidad de segundo orden

$$C(h) = C(s_i s_j) = \lambda_2(s_i s_j) - \lambda(s_i)\lambda(s_j)$$

$\hat{k}(h)$ : Estimación de K-Ripley  $k(h)$ : Función teórica bajo un modelo

Los parámetros se estiman usando el método de la "función de contraste" dado por:

$$\int_0^{b \approx 0,25} (\hat{k}(h) - k(h))^2 dh$$

### 2.5.6. Implementación en R

```

rm(list=ls())
library(spatstat) #para trabajar patrones puntuales

#####
# Tipos de patrones (aleatorio, uniforme, agregado), Patrones marcados,
# Covariables.
#####

## Aleatorio
aleatorio<-rpoispp(100)
plot(aleatorio, main="Aleatorio", pch=20)
summary(aleatorio)
plot(density(aleatorio), main="Estimación kernel de la intensidad")
contour(density(aleatorio), add=TRUE, main="Estimación kernel de la intensidad")
points(aleatorio, pch=20)

cx<-runif(600, 0,4)#coordenada x
cy<-runif(600,0,3) #coordenada y
#se generan 600 sitios aleatorios
patron<-ppp(cx,cy,c(0,4),c(0,3))
plot(patron, main="Aleatorio")
summary(patron) #lamdna=50, porque el area=12 y 600/12=50
plot(density(patron),main="Intensidad Constante = 50")
contour(density(patron),add=TRUE)
plot(patron, add=TRUE, col=2)
plot(density(patron, bw=.25), main="Intensidad Constante = 50")
contour(density(patron, bw=.25),add=TRUE)

### No Homogéneo

lambda<-function(x,y)
{
  100*(x+y)
}
nohomo<-rpoispp(lambda)
plot(nohomo)

### Proceso Matern Cluster
Mat<-rMatClust(kappa=10, r=0.1, mu=100)

```

```
Mat
plot(Mat, main="Agregado", pch=20)
summary(Mat)
plot(density(Mat), main="Agregado")
contour(density(Mat), add=TRUE, main="Estimación kernel de la intensidad")
points(Mat, pch=20)

### Patrón Regular

data(cells)
plot(cells, main="Regular", pch=20)

#####
## Simulando un patrón marcado (aleatorio y con tendencia)
#####

x<-runif(0,1,n=100)
y<-runif(0,1,n=100)
m1<-rnorm(100,10,2)
P<-ppp(x, y, c(-0.1,1.1),c(-.1,1.1), marks =m1)
plot(P, main="Patrón Marcado")

m2<-x-y+rnorm(1,0,1)
P2<-ppp(x, y, c(-0.1,1.1),c(-.1,1.1), marks =m2)
plot(P2, main="")

#####
# Estimación kernel de la densidad. Se usa un kernel Gaussiano
# K-Ripley
#####

#####
### Datos de pinos en Suecia
#####

demo(data)
data(swedishpines)
x<-(swedishpines)
plot(density(x,10), main="Estimación kernel de la intensidad")
contour(density(x,10), add=TRUE, main="Estimación kernel de la intensidad")
points(x, pch=20)
summary(x)
```

```
u<-Kest(x)
plot(u,, main="K-Ripley para los datos Swedishpines")
u

#####
### Datos de arboles en Nueva Zelanda
#####

data(nztrees)
y<-(nztrees)
plot(density(y,10), main="Estimación kernel de la intensidad")
contour(density(y,10), add=TRUE)
points(y, pch=20)
summary(y)

#####
### Datos de arboles en una selva tropical
#####

data(bei)
z<-density(bei,sigma=70)
plot(z, main="Datos de árboles en una selva tropical")

plot(density(bei), main="Estimación de la intensidad")
#points(bei, pch=20)
contour(density(bei), add=TRUE)
contour(density(bei))

# Ejemplo de covariables

# Nota bei.extra$elev es formato im(pixel image)

elev<-bei.extra$elev
plot(elev, main="")
plot(bei, pch=20, main="Elevación como covariable", add=TRUE)
contour(density(bei), add=TRUE)

# Ejemplo de efecto de borde

plot(bei, add=TRUE)
```

```
#####
# Métodos basados en cuadrantes
#####

aleatorio<-rpoispp(100)
plot(aleatorio, main="Aleatorio", pch=20)
summary(aleatorio)
plot(density(aleatorio), main="Estimación kernel de la intensidad")
contour(density(aleatorio), add=TRUE, main="Estimación kernel de la intensidad")
points(aleatorio, pch=20)

Q<-quadratcount(aleatorio, nx=6,ny=3)
plot(aleatorio)
plot(Q, add=TRUE)

M<-quadrat.test(aleatorio, nx=3, ny=3)
plot(aleatorio)
plot(M, add=TRUE)
M

#####
# Simulación de la estadística X-squared para la prueba de aleatoriedad
# basado en el conteo de cuadículas.
#####

aleatorio<-rpoispp(100)
Mobs<-quadrat.test(aleatorio, nx=3, ny=3)
plot(aleatorio)
plot(Mobs, add=TRUE)
Mobs

s<- 1000 # Número de simulaciones del proceso aleatorio

Xsq<-matrix(0, nrow=s, ncol=1)
for (i in 1:s)
{
  aleatorio<-rpoispp(100)
  M<-quadrat.test(aleatorio, nx=3, ny=3)
  Xsq[i,]<-M$statistic
}

hist(Xsq, freq=FALSE) # Histograma de los datos simulados
```

```

x<-rchisq(10000, df=8)
curve(dchisq(x, df = 8), col = 2, lty = 2, lwd = 2, add = TRUE) # Curva teórica de la chi-cuadrado con 8 gl
lines(density(Xsq), col="blue") # Estimación kernel de la densidad
abline(v=Mobs$statistic, col=2, lty=2) # Valor de la estadística observada

# Lambda no homogéneo
lambda<-function(x,y)
{
  100*(x+y)
}

nohomogeneo<-rpoispp(lambda)
nohomogeneo

MobsNohomoge<-quadrat.test(nohomogeneo, nx=3, ny=3)
abline(v=MobsNohomoge$statistic, col=3, lty=2)

# Nota: los grados de libertad de la chi cuadrado son [(3x3)-1]
# y entonces el percentil 95 (valor crítico de la prueba es

critico<-qchisq(0.95,8)
critico

# Estimación del alpha (Pr rechazar H0 siendo cierto) a partir de las simulaciones.
# Es decir en cuantas ocasiones de las 1000 la estadística observada es mayor que el crítico

suma<-0
for (i in 1:s)
{
  if (Xsq[i,> critico) suma<-suma+1 else suma<-suma
}
suma
alpha.estimado<-suma/s
alpha.estimado

# Probabilidad de rechazo con los datos simulados del proceso no homogéneo de media lambda
# Es decir en cuantas ocasiones de las 100 rechazo cuando el proceso es no homogéneo

sumanh<-0
XNoH<- matrix(0,nrow=s, ncol=1)
for (i in 1:s)
{
  nohomogeneo<-rpoispp(lambda)
}

```

```
M2<-quadrat.test(nohomogeneo, nx=3, ny=3)
XNoH[i,]<-M2$statistic
if (XNoH[i,]> critico) sumanh<-sumanh+1 else sumanh<-sumanh
}

sumanh
probabilidad_rechazo<-sumanh/s
probabilidad_rechazo

#####
# Funciones G, F, K, L
#####

# Proceso completamente aleatorio

aleatorio<-rpoispp(100)
G1<-Gest(aleatorio)
plot(G1)
plot(envelope(aleatorio, Gest), main=" Función G con bandas de confianza")

F1<-Fest(aleatorio)
plot(F1)
plot(envelope(aleatorio, Fest), main=" Función F con bandas de confianza")

K1<-Kest(aleatorio)
plot(K1)
plot(envelope(aleatorio, Kest), main=" Función K con bandas de confianza")

E <- envelope(aleatorio, Kest)
plot(E, sqrt(. / pi) ~ r, main="Función L")



# Proceso de inhibición (regular)

data(cells)
plot(cells)

G2<-Gest(cells)
plot(G2)
plot(envelope(cells, Gest), main= "Función G con bandas de confianza")

F2<-Fest(cells)
plot(F2)
```

```
plot(envelope(cells, Fest), main=" Función F con bandas de confianza")

K2<-Kest(cells)
plot(K2)
plot(envelope(cells, Kest), main=" Función K con bandas de confianza")

E <- envelope(cells, Kest)
plot(E, sqrt(. / pi) ~ r, main="Función L")

# Proceso de agregado

data(redwood)
plot(redwood)
G3<-Gest(redwood)
plot(G3)
plot(envelope(redwood, Gest), main= "Función G con bandas de confianza")

F3<-Fest(redwood)
plot(F3)
plot(envelope(redwood, Fest), main= "Función F con bandas de confianza")

K3<-Kest(redwood)
plot(K3)
plot(envelope(redwood, Kest), main=" Función K con bandas de confianza")

E <- envelope(redwood, Kest)
plot(E, sqrt(. / pi) ~ r, main="Función L")

#####
# Simulacion de procesos
# Homogeneo, No homogeneo, Cluster, Cox, Matern
#####

#####
## Poisson Homogéneo
#####

aleatorio<-rpoispp(100)
plot(aleatorio, main="Aleatorio", pch=20)
```

```
summary(aleatorio)
plot(density(aleatorio), main="Estimación kernel de la intensidad")
contour(density(aleatorio), add=TRUE, main="Estimación kernel de la intensidad")
points(aleatorio, pch=20)
points(0.5, 0.8, col=2, pch=19)

fit_homo<-ppm(aleatorio, ~1)
names(fit_homo)
fit_homo$coef[[1]]
exp(fit_homo$coef[[1]]) # Estimación de la intensidad basada en el beta estimado.

#####
# Proceso Poisson No Homogéneo
#####

lambda<-function(x,y)
{
  100*(x+y)
} #en la medida que xy y aumenten la
#funcion de intensidad tamb aumenta
nohomo<-rpoispp(lambda) #va ir cambiando dependiendo de las coordenadas x y
summary(nohomo)
windows()
plot(nohomo)
plot(density(nohomo), main ="Patrón No Homogéneo")
contour(density(nohomo), add=TRUE)

# Ajuste de un modelo lineal a la intensidad

fit_nohomo<-ppm(nohomo, ~x+y) #lambda(x,y)= exp(beta_0+beta1*x+beta2*y)
names(fit_nohomo)
fit_nohomo$coef

# Función para calcular la intensidad en un punto (x,y)

lambda_lineal<-function(x,y,fit)
{
  lambda<-exp(fit[[1]]+fit[[2]]*x+fit[[3]]*y)
  return(lambda)
}
```

```

x<-0.7
y<-0.8
lambda_lineal(x,y,fit_no homo$coef) # Estimación de la función de intensidad en x=0.5 y y=0.5
# Comparar contra el gráfico

par(mfrow=c(1,2))
plot(density(no homo), main ="Patrón No Homogéneo. Ajuste No Paramétrico")
contour(density(no homo), add=TRUE)
points(x,y, col=2, pch=19)
plot(fit_no homo, se=FALSE)
points(x,y, col=2, pch=19)

# Ajuste de un modelo polinomial

fit_no homo2<-ppm(no homo, ~polynom(x,2)+y)
fit_no homo2
plot(fit_no homo2, se=FALSE)

fit_no homo3<-ppm(no homo, ~polynom(x,y,2))
fit_no homo3
plot(fit_no homo3, se=FALSE)

# Comparación de modelos
#(usando la tabla de Desviación basada en la razón de verosimilitud)

fit_no homo<-ppm(no homo, ~x+y)
fitnull<-update(fit_no homo,~1)
anova(fitnull, fit_no homo, test="Chi")

fit_no homo2<-ppm(no homo, ~polynom(x,2)+y)
fitnull2<-update(fit_no homo2, ~x+y)
anova(fitnull2, fit_no homo2, test="Chi")

# Interpretación: Si el p-valor es pequeño (p.ej menor del 5%)
# se rechaza la hipótesis nula (es decir que el modelo de
# menos parámetros es mejor)

lambda<-function(x,y)
{
  200*(x+y+x^2+y^2+x*y)
}
no homo<-rpoispp(lambda)
summary(no homo)
plot(no homo)

```

```
plot(density(nohomo), main ="Patrón No Homogéneo")
contour(density(nohomo), add=TRUE)

fit_nohomo3<-ppm(nohomo, ~polynom(x,y,2))
plot(fit_nohomo3, se=FALSE)
fitnull3<-update(fit_nohomo3, ~x+y)
anova(fitnull3, fit_nohomo3, test="Chi")

#####
# Realizaciones de un proceso de Cox con intensidad constante
# lambda (lambda cambia en cada iteración)
#####

par(mfrow=c(2,2))
vectorl<-c(rep(0,10000))
vectorn<-c(rep(0,10000))

for (i in 1:10000)
{
  alpha<-10
  beta<-5
  lambda<-rgamma(1, shape=alpha, scale=beta)
  vectorl[i]<-lambda
  Cox<-rpoispp(lambda)
  #plot(Cox)
  #print(Cox)
  vectorn[i]<-Cox$n
}
vectorl
vectorn
mean(vectorn)
mean(vectorl)

# Nota: Cuando se simula un proceso Poisson homogeneo,
# en cada iteración el lambda es el mismo,
# en el de Cox el lambda cambia!!!

#####
#Poisson no homogeneo aleatorio
#####

par(mfrow=c(2,2))
vectorl<-c(rep(0,100))
vectorn<-c(rep(0,100))
```

```
for (i in 1:2)
{
alpha<-10
beta<-5
lambda1<-rgamma(1, shape=alpha, scale=beta)
lambda<-function(x,y)
{
lambda1*(x+y)
}
Cox<-rpoispp(lambda)
plot(Cox)
plot(density(Cox), main ="Patrón No Homogéneo")
contour(density(Cox), add=TRUE)

print(Cox)
vectorn[i]<-Cox$n
}

mean(vectorn)
mean(vectorl)

# Cox Log_normal

par(mfrow=c(3,2))
for (i in 1:3)
{
lambda<-function(x,y)
{
exp(rnorm(1,5,1)*(x+y))
}
Cox_log<-rpoispp(lambda)
plot(Cox_log)
plot(density(Cox_log), main ="Patrón No Homogéneo")
contour(density(Cox_log), add=TRUE)
#print(Cox_log)

}

hist(e)

#####
# Realizaciones de un proceso de Cox con intensidad NO constante
# lambda (lambda cambia en cada iteración y depende de x,y)
#####
```

```
# Poisson/Gamma

par(mfrow=c(1,3))
for (i in 1:3)
{
  alpha<-10
  beta<-5
  lambda<-rgamma(1, shape=alpha, scale=beta)
  lambda1<-function(x,y)
  {
    lambda*(x^2+y^2)
  }
  Cox_no homo<-rpoispp(lambda1)
  plot(Cox_no homo)
}

# Cox Log_normal

par(mfrow=c(1,3))
for (i in 1:3)
{
  lambda<-function(x,y)
  {
    exp(rnorm(1,5,1)*(x+y))
  }
  Cox_log<-rpoispp(lambda)
  plot(Cox_log)
}

#####
# Proceso Poisson Agregado (cluster)
#####

# Proceso de Thomas

rThomas

pp_Thomas<-rThomas(10, 2, 10)
plot(pp_Thomas)
fit<-kppm(pp_Thomas, ~x+y, "Thomas")
```

```
fit  # Estima los parámetros de la función K(h) usando el método del mínimo contraste

#ensayo para ver los clusters
pp_Thomas<-rThomas(2, 0.02, 100)

plot(pp_Thomas)
fit<-kppm(pp_Thomas, ~1, "Thomas")
fit

#funcion k de ripley para proceso cluster

K1<-Kest(pp_Thomas, correction="none")
plot(K1)
pi
#teorica
clase<-function(h)
{
  k<-(pi*h^2)+(1-exp(-(h^2))/4)*(0.01^2))/2
  return(k)
}
clase(0.05)
dist=seq(0,12,by=0.01)
k_ripley<-clase(dist)
plot(dist,k_ripley, type="l")
```

# **Capítulo 3**

## **Datos de Areas**

### **3.1. Visualización de datos de áreas**

La forma más usual de visualización de datos de agregaciones espaciales es a través de mapas cloropléticos o temáticos donde las áreas son sombreadas (coloreadas) de acuerdo con sus valores en la variable de interés. El número de clases y las correspondientes clases pueden ser basadas en diferentes criterios.

#### **3.1.1. Mapas de datos originales y tasas**

Los sistemas de información geográfica usualmente disponen de tres métodos de corte de los valores de la variable: Intervalos iguales, cuantiles (cuartiles, deciles, percentiles) y desviaciones estándar. En el caso de intervalos iguales se toma la diferencia entre el mínimo y el máximo de la variable (rango) y se divide por el número de clases o intervalos deseados, obteniéndose así la amplitud de los mismos. Este procedimiento es similar a como se hacía hace unos años, cuando no había mucho acceso a computadores, para obtener los intervalos en la construcción de los histogramas de frecuencia. Si la distribución de frecuencias es asimétrica, esta división deja una gran parte del mapa sombreado con el mismo color y por consiguiente impide una buena descripción del fenómeno de estudio. De otro lado pueden emplearse las cuantiles. Tradicionalmente se usan la mediana y los cuartiles. Solo

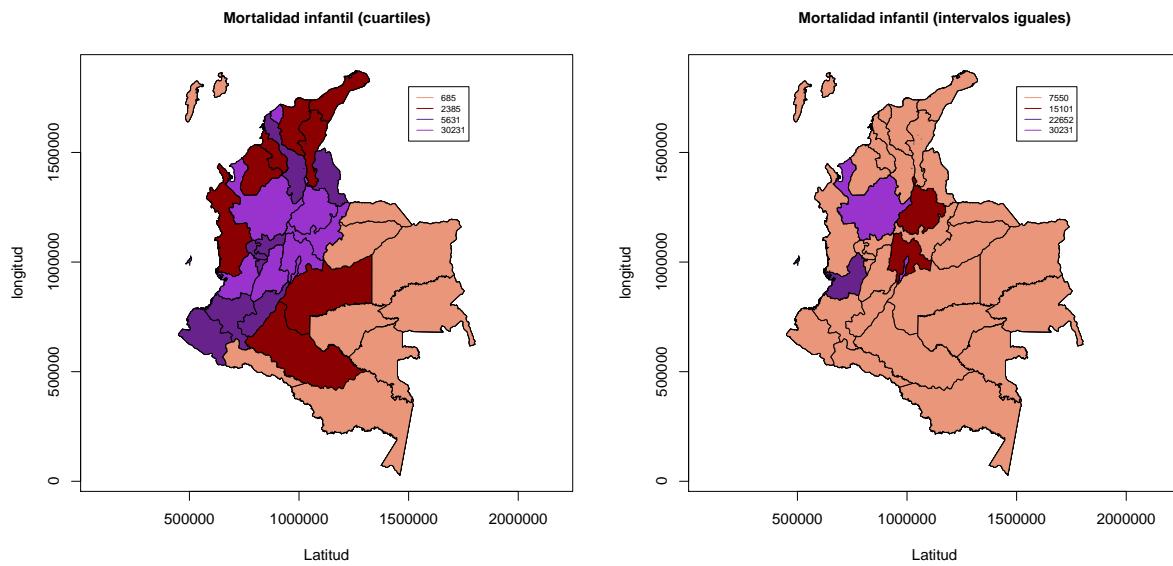


Figura 3.1: Mapas temáticos de mortalidad infantil en Colombia en 2003.

en aquellos casos en que haya un gran número de áreas en el mapa podría pensarse en el uso de deciles o percentiles. En estos casos hay un número igual de áreas sombreadas con cada color. La desventaja de este procedimiento es que los valores extremos son enmascarados, puesto que caen dentro de alguna categoría grande, lo cual en algunas situaciones prácticas podría impedir la detección de zonas críticas, por ejemplo en el caso de que las variables sean tasas. Finalmente en el caso de desviaciones estándar los límites de los intervalos son construidos con base en desviaciones estándar respecto a la media. Este procedimiento es válido sólo en aquellas situaciones en los que el supuesto de simetría pueda ser razonable. Las diferencias entre los distintos modos de construcción de mapas son ilustradas con base en información mortalidad infantil (número total y tasa) por departamentos de Colombia en el año 2003.

### 3.1.2. Mapas de riesgos relativos

Suponga que está interesado en hacer comparaciones de dos mapas de la misma región sobre la misma variable en períodos de tiempo distintos o en el mismo período pero con

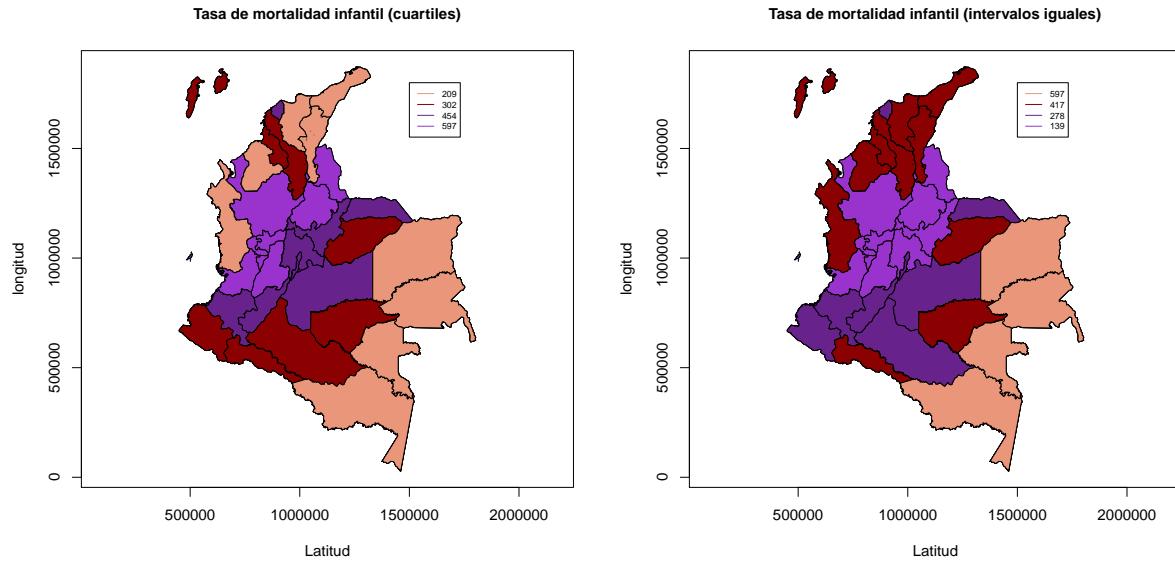


Figura 3.2: Mapas temáticos de tasas de mortalidad infantil en Colombia en 2003.

variables distintas. En esta situaciones como estas es común trabajar con medidas de riesgo relativo al nivel medio de la región. Asuma que la tasa es calculada como se describió en la sección anterior. En cada área  $i$  se calcula el número esperado de eventos asumiendo una tasa igual a la de toda la región de estudio. La tasa global es calculada como  $r = \left( \frac{\sum_{i=1}^n z_i}{\sum_{i=1}^n n_i} \right)$  y el número esperado de eventos como  $E_i = n_i r$ . El riesgo relativo ( $RR$ ) en el área  $i$ , es calculado como el cociente entre el número observado y el número esperado de eventos, es decir  $RR_i = \frac{z_i}{E_i}$ . De este modo un  $RR_i = 1$  indicaría que el área  $i$  tiene tantos casos como los esperados si su tasa fuera igual a la de toda la región. Es común multiplicar  $RR_i$  por 100, siendo entonces este valor el punto de referencia global. Un valor de 200 en el riesgo relativo del área  $i$  indicaría que esta tiene el doble de casos a los esperados si su riego fuera igual al de toda la región. Los riegos relativos también pueden calcularse teniendo en cuenta diferentes grupos etéreos o género. Sean  $z_{ij}$  el número de eventos que ocurren entre personas de la clase  $j$  en el área  $i$  y  $n_{ij}$  el número de personas de dicha clase en la misma área, entonces la tasa global para la categoría  $j$  es  $r_j = \frac{\sum_{i=1}^n z_{ij}}{\sum_{i=1}^n n_{ij}}$ , con  $n$  igual al número total de áreas. Entonces  $E_{ij} = r_j n_{ij}$  es el número esperado de eventos de la clase  $j$  en el área  $i$  si el riego de la clase  $j$  fuese constante en el espacio. El número esperado

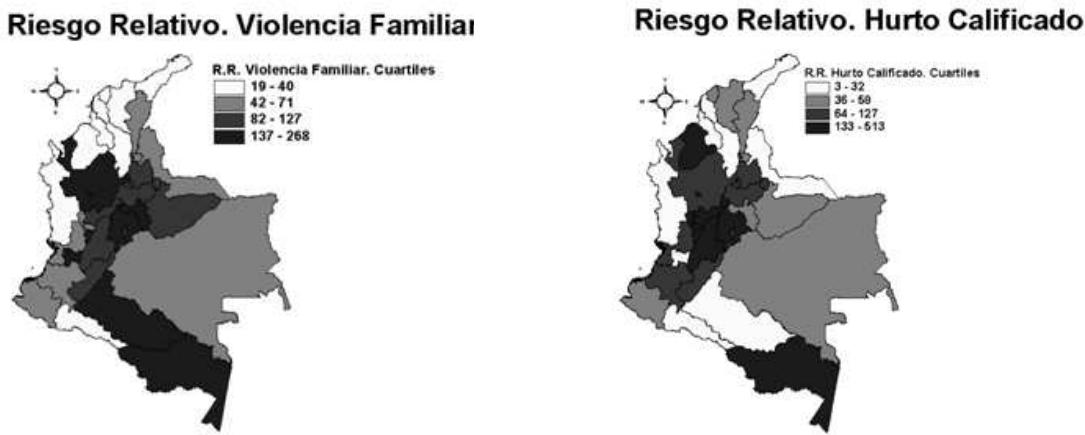


Figura 3.3: Distribución de las tasa de violencia familiar (arriba) y de hurto calificado (abajo) según unidades seccionales de la Fiscalía General de la Nación, en el año 2003.

de eventos en el área  $i$  es  $E_i = \sum_{j=1}^k E_{ij}$ , con  $k = 2$  si se tienen en cuenta el género o  $k$  igual al número de grupos de edad. El riesgo relativo del área  $i$  se calcula nuevamente como el cociente entre casos observados y esperados. En esta situación se asume que es una variable aleatoria Poisson con valor esperado  $E_i$ . En general cuando se construye un mapa se asume que las tasas no son iguales en toda la región. Así que un modelo que puede ser un poco más realista es aquel en que los riesgos varían en el espacio, es decir  $Z_i$  es una variable aleatoria Poisson con valor esperado  $\mu_i = \lambda_i E_i$ , donde el parámetro  $\lambda_i$  mide la diferencia del riesgo del área  $i$  relativa a la media de la región total. En la expresión anterior el riesgo relativo es  $\lambda_i$  y su estimador más verosímil es  $\frac{z_i}{E_i}$ .

Para ilustrar el uso de los mapas de riesgos relativos, se considera información de las tasas de violencia familiar y de hurto calificado por unidades seccionales de la Fiscalía en el año 2003. Los mapas son presentados en la figura 3.3.

La observación de los dos mapas de la figura 3.3 permite concluir que las tasas de violencia familiar y de hurto calificado presentan un patrón de distribución espacial similar, con tasas bajas, respecto a la tasa global del país, en las unidades seccionales de la costa

(Caribe y Pacífica), Villavicencio, Mocoa, Yopal y Cúcuta. También permiten inferir que son más marcadas las diferencias en las tasa de hurto calificado que en las de violencia familiar, puesto que los riegos relativos en el primer caso tiene valores de cuartil uno y mediana más bajos. Algunos aspectos puntuales de interés se reflejan en las seccionales de Cali y Florencia. Las tasas de violencia familiar resultan altas en dichas seccionales, respecto a la tasa global del país, mientras que en el caso de hurto calificado sucede lo contrario.

Se debe resaltar que una mejor comparación podría hacerse, teniendo en cuenta lo mencionado en último párrafo de la sección 3.1.1, si los límites de los dos mapas fueran los mismos. A pesar de que ambos están en la misma escala (riegos relativos) y de que ambos fueron construidos con base en cuartiles, la comparación debe ser cuidadosa puesto que el rango de variación en la escala de riesgos relativos es distinta.

### **3.1.3. Mapas Probabilísticos y Suavizados con Estimación Bayesiana**

De la sección 3.1.1, quedó claro que cuando la variable estudiada en cada área de la región de interés es el número de eventos (nacimientos, muertes, enfermos, etc.) se deben calcular tasas con el propósito de hacer comparables estas mismas. No obstante en algunos casos debe ser cuidadosa la interpretación, sobre todo en aquellos en que las poblaciones  $n_i$  de algunas de las áreas son muy pequeñas respecto a las restantes. Imagine que se estudia la tasa de mortalidad infantil por barrios de Bogotá. Una muerte en un barrio de la localidad de Kennedy con alto número de nacimientos no afecta considerablemente dicha tasa, como si ocurriría en otro como Quinta Paredes o Villa del Prado con un número considerablemente menor de nacimientos. Suponga que en Quinta Paredes hubo en un periodo de tiempo 200 nacimientos y una muerte esto indicaría que la tasa es de 500 por cada 100000 nacimientos. Un evento adicional duplicaría la tasa a 1000 por cada 100000, mientras que si en Kennedy se presentan 4000 nacimientos, la misma situación haría cambiar la tasa de mortalidad de 25 a 50 por cada 100000 nacimientos. Estos problemas son típicos cuando se estudian áreas con poblaciones en riesgo muy distintas. Algunas de las alternativas citadas en la literatura para tratar el problema de alta variabilidad en las

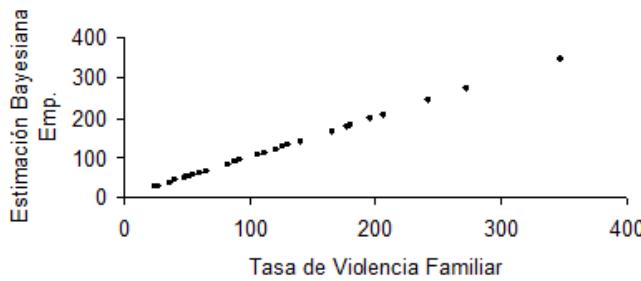


Figura 3.4: Dispersograma entre tasas de violencia familiar de 2003 y tasas suavizadas por el método bayesiano empírico. ).

tasas son: La construcción de mapas de probabilidades y el suavizamiento de las tasas a través del método bayesiano empírico. En el primer caso se asume que  $Z_i$  sigue un modelo de probabilidad de Poisson con valor esperado  $E_i$ . Si  $X$  es una variable aleatoria con el mismo modelo de  $Z_i$  se define:

$$P_i = \begin{cases} P(X \geq z_i) & \text{Si } z_i \geq E_i \\ P(X \leq z_i) & \text{Si } z_i < E_i \end{cases}$$

Un valor de  $P_i$  cercano a cero indica que el valor de la tasa en el área  $i$  es muy bajo o muy alto respecto a su correspondiente valor esperado. Este procedimiento tiene la desventaja, cuando la población  $n_i$  en cada área es muy alta, de arrojar en la mayoría de las áreas valores de probabilidad bajos. En el caso de las tasa de violencia familiar reportados por la seccionales de la Fiscalía para el año 2003, dichas tasas son considerablemente bajas en la mayoría de estas, porque la población cubierta es grande en cada una de ellas. En este caso (Fig. 3.4, abajo) este procedimiento no aporta mucha información. Una situación ideal, bajo el supuesto de que las tasas son estacionarias, ocurriría cuando las probabilidades estén cerca de 0.5.

De otro lado el suavizamiento bayesiano empírico asume que la tasa en el área  $i(R_i)$  es una variable aleatoria que tiene media  $\mu_i$  y varianza  $\sigma_i^2$ . El caso más simple ocurre cuando se asume estacionariedad (igual media). El procedimiento consiste en hacer una re-estimación ( $\hat{\theta}_i$ ) de la tasa en el área  $i$  usando la siguiente expresión:

$\hat{\theta}_i = w_i r_i + (1 - w_i) r$ , con  $r_i$  y  $r$  definidos como en la sección 3.1.2 y  $w_i = \frac{\hat{\sigma}^2}{\hat{\sigma}^2 + \frac{r}{n_i}}$ , donde  $\hat{\sigma}^2 = \frac{\sum_{i=1}^n n_i(r_i - r)^2}{\sum_{i=1}^n n_i} - \frac{r}{\bar{n}}$ .

Al igual que el procedimiento basado en el mapa de probabilidades, cuando la población cubierta en cada área, con base en la que se calculan las tasas, es muy alta, el efecto de suavizamiento es imperceptible. En estos casos los ponderadores  $w_i$  serán cercanos a uno y por consiguiente  $\hat{\theta}_i = r_i$ . Por ello el mapa de suavizamiento empírico de Bayes en el caso de las tasa de violencia familiar no muestran un efecto de suavizamiento importante respectos a las tasas originales. Solo en algunos casos puntuales, que pueden establecerse sólo al hacer las comparaciones numéricas en la base de datos, puede identificarse el efecto de la estimación. En la figura 3.4 se observa una correlación casi perfecta entre las tasas observadas y las estimadas, lo cual indica que el efecto de suavizamiento en este caso fue muy pequeño.

Una situación distinta a la de violencia familiar ocurre al hacer el suavizamiento de las tasas de mortalidad en Colombia (tomando como áreas los departamentos). Se observa al comparar las tasas observadas con las estimadas que las más bajas (menores de 100) son aumentadas en la estimación a más de 100, acercándolas así al promedio observado en Colombia.

## 3.2. Exploración de datos de áreas

### 3.2.1. Medidas de proximidad espacial

### 3.2.2. Medias móviles

### 3.2.3. Estimación kernel

## 3.3. Correlación espacial

Una etapa muy importante del análisis estadístico del análisis de datos de área es la determinación de la estructura de autocorrelación espacial. Análogo a como ocurre

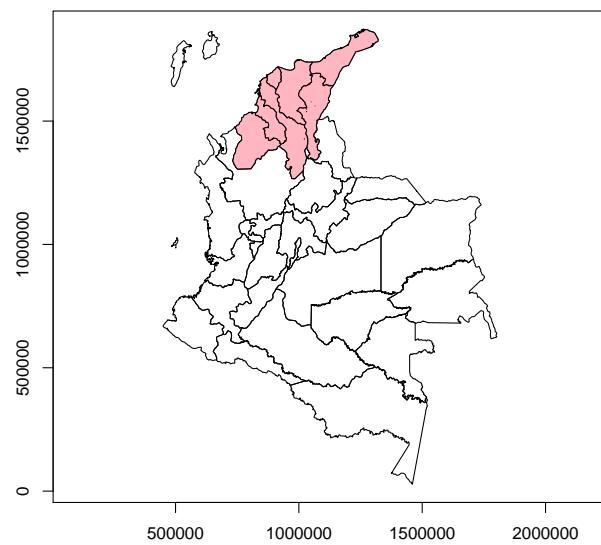


Figura 3.5: Las áreas sombreadas corresponden a los departamentos de la costa Caribe de Colombia (sin incluir San Andrés y Providencia).

en el análisis de series de tiempo con las funciones de autocorrelación simple y parcial, el establecimiento de la asociación o similitud de los valores en función de la distancia entre ellos es por si mismo un resultado que permite caracterizar la población de estudio y además es fundamental en el desarrollo de modelos de predicción. En este contexto se busca establecer la magnitud de la asociación entre las áreas. Las herramientas más utilizadas para este propósito son los índices de autocorrelación de Moran (I de Moran), de Geary (C de Geary), la función de semivariograma (que aunque propia de análisis de superficies aleatorias tiene también utilidad en el análisis de áreas) y el índice bayesiano empírico. Para el cálculo de estos índices es necesario definir la proximidad entre las áreas. Lo anterior se lleva a cabo por medio del cálculo de una matriz de proximidad espacial. Dado un conjunto de  $n$  áreas ( $A_1, \dots, A_n$ ) se construye una matriz  $W^{(1)}$  de orden ( $n \times n$ ) donde cada uno de los elementos  $w_{ij}$  representa una medida de proximidad entre  $A_i$  y  $A_j$ . Dicha medida puede ser calculada con alguno de los siguientes criterios:

- $w_{ij} = 1$  si el centro de  $A_i$  se encuentra a una distancia determinada de  $A_j$  o  $w_{ij} = 0$  en caso contrario.
- $w_{ij} = 1$  si  $A_i$  comparte frontera con  $A_j$ , en caso contrario  $w_{ij} = 0$ .
- $w_{ij} = I_{ij}/I_i$ , donde  $I_{ij}$  es la frontera entre  $A_i$  y  $A_j$  y  $I_i$  es el perímetro de  $A_i$ .
- $w_{ij} = d_{ij}$ , la distancia entre los centros de las dos áreas.

En todos los casos anteriores  $w_{ii} = 0$ . La idea de la matriz de proximidad espacial puede ser generalizada a vecinos de mayor orden (vecinos de vecinos) construyéndose así las matrices  $W^{(2)}, \dots, W^{(n)}$ . Se acostumbra a normalizar las filas de la matriz (la suma de los  $w_{ij}$  por fila iguales a uno).

Una forma simple de explorar la variación de la tendencia espacial de los datos es a través del cálculo de las medias de los valores en las áreas usando la información de los vecinos por medio de la matriz de proximidad espacial:

$$\bar{x}_i = \frac{1}{n} \sum_{j=1}^n w_{ij} z_j, \text{ con } w_{ij} = 1 \text{ ó } 0$$

Un mapa de los  $\bar{x}_i$  permitirá tener superficies más suavizadas, con menores fluctuaciones entre los valores.

### 3.3.1. I de Moran y C de Geary

Sean  $(Z_1, \dots, Z_n)$  variables aleatorias medidas en las  $n$  áreas. La noción de autocorrelación espacial de estas variables está asociada con la idea de que valores observados en áreas geográficas adyacentes serán más similares (o disímiles) que los esperados bajo el supuesto de independencia espacial. La autocorrelación espacial ocurre cuando áreas cercanas tienden El índice de autocorrelación de Moran considerando la información de los vecinos más cercanos es definida como:

$$I = \frac{n \sum_{i=1}^n \sum_{j=1}^n w_{ij}(z_i - \bar{z})(z_j - \bar{z})}{S_0 \sum_{i=1}^n (z_i - \bar{z})^2}, \text{ con } S_0 = \sum_{i=1}^n w_{ii}$$

Valores positivos (entre 0 y 1) indican autocorrelación directa (similitud entre valores cercanos) y valores negativos (entre -1 y 0) indican autocorrelación inversa (disimilitud entre las áreas cercanas). Valores del coeficiente cercanos a cero apoyan la hipótesis de no correlación espacial. El mismo cálculo realizado con matrices de proximidad de menor orden puede hacerse empleando las matrices  $W^{(2)}, \dots, W^{(n)}$  estimándose así la función autocorrelación. Sea  $k = 2, \dots, n$ , entonces en cada valor de  $k$  (cada rezago), se obtiene:

$$I^k = \frac{n \sum_{i=1}^n \sum_{j=1}^n w_{ij}^k(z_i - \bar{z})(z_j - \bar{z})}{S_0 \sum_{i=1}^n (z_i - \bar{z})^2}$$

Una vez obtenido el valor del coeficiente es necesario evaluar su validez estadística. En otras palabras se requiere probar la hipótesis de independencia espacial con base en el valor observado. Para llevar a cabo esto es necesario establecer la correspondiente estadística de prueba y su distribución de probabilidad. Aunque con algunas restricciones debido a supuestos fuertes que son asumidos, se emplea generalmente la estadística:

$$Z = \frac{I - E(I)}{\sqrt{V(I)}}, \text{ con } E(I) = \frac{-1}{n-1} \text{ y } V(I) = \frac{n^2 S_1 - n^2 S_2 + 3S_0^2}{(n^2 - 1)S_0^2} - \left(\frac{-1}{n-1}\right)^2$$

donde

$$S_1 = \sum_{i \neq j} (w_{ij} + w_{ji})^2, S_2 = \sum_i (w_{i\cdot} + w_{\cdot i})^2 \text{ con } w_{i\cdot} = \sum_j W_{ij} \text{ y } w_{\cdot i} = \sum_j w_{ji}$$

La estadística mencionada sigue asintóticamente (cuando es grande el número de áreas) distribución normal estándar y por consiguiente la decisión se toma usando valores extremos de dicha distribución. Otra posibilidad para establecer la significancia estadística, con menos supuestos, es a través de una prueba de pseudo significancia (Assuncao (2001)).

Un supuesto implícito en el cálculo del coeficiente de Moran es el de estacionariedad de primer y segundo orden (media y varianza constante). Cuando este supuesto es dudoso puede emplearse el coeficiente de autocorrelación de Geary, el cuál es similar a la función  $s$  de semivarianza (sección 3.3.3) usada en geoestadística. El coeficiente Geary es sin embargo menos utilizado que el de Moran. Su expresión es:

$$C = \frac{(n - 1)}{2S_0} \left( \frac{n \sum_{i \neq j} w_{ij}(z_i - z_j)^2}{\sum_{i=1}^n (z_i - \bar{z})^2} \right)$$

Los coeficientes de autocorrelación de Moran y de Geary pueden calcularse usando el software Crimestat (Levine & Associates, 2002). La función de autocorrelación de Moran puede calcularse con GS+(Gamma Design Software, 1999).

### 3.3.2. Índice Bayesiano Empírico (IBE)

El índice Bayesiano empírico es empleado para detectar autocorrelación espacial con tasas. Sean  $(\theta_1, \dots, \theta_n)$  las tasas de las áreas 1 a  $n$ . Suponga que el número de casos  $Z_i$  es una variable aleatoria con distribución de Poisson con esperanza condicional  $E(Z_i|\theta_i) = n_i\theta_i$ . La tasa estimada  $r_i = \frac{z_i}{n_i}$  tiene media condicional  $E(r_i|\theta_i) = \theta_i$  y varianza  $E(r_i|\theta_i) = \theta_i/n_i$ . Por lo tanto las tasas estimadas tienen diferentes medias y esperanzas condicionales. Suponga que las tasas  $\theta_1, \theta_2, \dots, \theta_n$  tienen distribución a-priori independiente con media y varianza  $\alpha$  y  $\beta$ , respectivamente. Así la esperanza marginal de  $r_i$  es  $\beta$  y la varianza marginal es  $\alpha + (\beta/n_i)$ . Ahora las medias son iguales y las varianzas difieren y la varianza aumenta en la medida que la población decrece. Usando estimadores de momentos propuestos por Marshall (1991) se obtiene:

$$\hat{\alpha} = s^2 - (\hat{\beta}/m), \hat{\beta} = \frac{\sum_{i=1}^n z_i}{\sum_{i=1}^n n_i}, \text{ con } m = \frac{\sum_{i=1}^n n_i}{n} \text{ y } s^2 = \frac{\sum_{i=1}^n n_i(r_i - \hat{\beta})^2}{\sum_{i=1}^n n_i}$$

Por lo tanto la esperanza y la varianza marginal de los  $r_i$  son estimadas por  $\hat{\beta}$  y  $v_i = \hat{\alpha} + \hat{\beta}/n_i$ .

Estandarizado se define  $y_i = \frac{r_i - \hat{\beta}}{\sqrt{v_i}}$  y el índice de Bayesiano empírico como:

$$IBE = \frac{n \sum_{i \neq j} w_{ij} y_i y_j}{\sum_{i \neq j} w_{ij} \sum (y_i - \bar{y})^2}$$

### 3.3.3. Función de semivariograma

La función de semivarianza, muy importante dentro de la teoría geoestadística, puede ser utilizada en el análisis de datos de áreas. Se requiere conocer para el cálculo de dicha función, además de los valores de la variable de interés su coordenada geográfica (la de su centroide). La función se define como:

$$\gamma(h) = \frac{\sum (Z(x+h) - Z(x))^2}{2n(h)}$$

donde  $h$  representa la distancia entre las áreas y  $n(h)$  es el número de áreas que se encuentran a una distancia  $h$ . Esta es una función monótona creciente. Obviamente cuando  $h = 0$  la función vale cero. Una consulta detallada al respecto de esta función y de cómo modelarla se encuentra en Giraldo (2002).

## 3.4. Modelos de regresión espacial

### 3.4.1. Modelo de regresión polinomial local y ponderado geométricamente

Una de las maneras para hacer uso de la información espacial y sus relaciones es haciendo uso de las covariables que son funciones polinomiales de las coordenadas espaciales

$s_i = [x_i, y_i]'$ . Por ejemplo, una tendencia lineal de superficie usa un polinomio de primer grado en  $[x, y]$  para describir la variación espacial en la respuesta, es decir

$$Z(s_i) = \beta_0 + \beta_1 x_i + \beta_2 y_i + \epsilon_i \text{ con } \epsilon_i \sim iid(0, \sigma^2)$$

La inferencia por mínimos cuadrados ordinarios puede ser usada para estimar los parámetros  $\beta$ . Sin embargo, cuando el análisis no es espacial:  $\mathbf{X}$  es clara y completamente ligado a la ubicación espacial. Adicionalmente, los parámetros estimados eran simplemente un medio para obtener una superficie de respuesta, los coeficientes  $\beta$  en sí mismos tienen una interpretación espacial, la medición de la fuerza de la gran tendencia espacial en  $Z(s_i)$

Los parámetros estimados de un análisis de una superficie de respuesta con tendencia proporciona una amplia escala de interpretación de la variación espacial en  $Z(s_i)$ . Sin embargo, esto es esencialmente espacial, desde que el modelo tenga un conjunto de parámetros que se aplique en todo lugar, sin tener en cuenta la localización espacial. Como se discutió anteriormente, se puede adaptar una regresión local polinomial al caso espacial ajustando un modelo polinomial a cualquier locación espacial especificada  $s_0$ . Este modelo es esencialmente una versión espacial de los procedimientos de estimación local comúnmente conocidos como LOESS o regresión no paramétrica, donde las covariables son funciones polinomiales de las coordenadas espaciales. En aplicaciones tradicionales del LOESS y los métodos no paramétricos de regresión donde las covariables forman  $\mathbf{X}$ , el término *local* se refiere al atributo o  $\mathbf{X}$ -espacio y no a la localización espacial. Los pesos son funciones de  $x_i - x_0$  y el análisis es espacial. Qué necesita realmente tener un modelo que es ajustado localmente en el mismo sentido, las covariables generales permiten que no se hagan necesarias funciones polinomiales para las coordenadas espaciales. El mismo modelo general descrito anteriormente puede ser usado, pero con covariables

$$\mathbf{Z}(s) = \mathbf{X}(s)\beta(s_0) + e_0 \text{ con } e_0 \sim (\mathbf{0}, \sigma^2 \mathbf{W}(s_0)^{-1})$$

donde las ponderaciones forman los elementos de  $\mathbf{W}(s_0)$ ,  $W(s_i, s_0)$ , determina el grado en el que  $Z(s_i)$  afecta la estimación de la relación entre  $\mathbf{X}(s)$  y  $\mathbf{Z}(s)$  hasta el punto de interés  $s_0$ . Las covariables en  $\mathbf{X}(s)$  no tienen funciones polinomiales para las coordenadas

espaciales. Estas pueden ser, precios, ingresos, elevación, raza, etc. Sin embargo, en contraste con los aportes del kernel tradicional, el aspecto local del modelo es basado en las localizaciones espaciales, no en los valores de las covariables. En contraste con los modelos para la estimación local, el interés no recae en el ajuste de una superficie de respuesta, pero en el ajuste de los mapas de los  $\beta_j$  que describen la variación espacial en la relación entre  $x_j(s)$  y  $Z(s)$ . En geografía este modelo es conocido como **Regresión ponderada geográficamente** (Fotheringham et al. (2002)).

Para evitar las discontinuidades no naturales en la estimación de los parámetros, la función espacial de los pesos tiene que ser escogida como una forma en la que la influencia de una observación decrece con el incremento de la distancia  $s_0$ . Las funciones de ponderación kernel son algunos ejemplos en las que se satisface esta condición. Cressie (1998) y Fotheringham et al. (2002) presentan una discusión más general sobre la escogencia de la función espacial y otras alternativas adicionales.

Una vez que la función espacial sea elegida,  $\beta(s_0)$  puede ser estimado por mínimos cuadrados ponderados

$$\hat{\beta}(s_0) = (\mathbf{X}(s)' \mathbf{W}(s_0) \mathbf{X}(s))^{-1} \mathbf{X}(s)' \mathbf{W}(s_0) \mathbf{Z}(s)$$

y la varianza del parámetro estimado es

$$Var[\hat{\beta}(s_0)] = \sigma^2 \mathbf{C} \mathbf{C}', \quad \mathbf{C} = (\mathbf{X}(s)' \mathbf{W}(s_0) \mathbf{X}(s))^{-1} \mathbf{X}(s)' \mathbf{W}(s_0)$$

si  $\beta$  es estimado como una muestra de localizaciones,  $s_i$ , entonces los valores ajustados serán dados por  $\hat{\mathbf{Z}}(s) = \mathbf{L} \mathbf{Z}(s)$ , donde la  $i$ -ésima columna de  $\mathbf{L}$  es  $x(s_i)(\mathbf{X}(s)' \mathbf{W}(s_i) \mathbf{X}(s))^{-1} \mathbf{X}(s)' \mathbf{W}(s_i)$ . El componente de varianza,  $\sigma^2$ , puede ser estimado de los residuales del ajuste usando Cressie (1998)

$$\hat{\sigma}^2 = \frac{(\mathbf{Z}(s) - \hat{\mathbf{Z}}(s))' (\mathbf{Z}(s) - \hat{\mathbf{Z}}(s))}{tr(\mathbf{I} - \mathbf{L})(\mathbf{I} - \mathbf{L})'}$$

Cressie (1998) da más detalles sobre como este modelo local puede ser usado en el análisis geoestadístico y Fotheringham et al. (2002) da algunos ejercicios prácticos de este modelo, y sobre como varias extensiones sobre el ajuste, pueden ser usadas en el análisis geográfico.

### 3.4.2. Modelos lineales con errores correlacionados

En esta sección, se continua asumiendo una relación lineal entre la variable de interés y las covariables, pero adicionalmente la variación espacial de las covariables, los componentes del término de error pueden no ser correlacionados.

Adicionalmente, debido a la existencia de numerosas técnicas de estimación para los parámetros de covarianza  $\theta$ , hay diferentes aportes que pueden ser utilizados para describir la dependencia de  $\Sigma(\theta)$  en  $\theta$ . Aquí se discuten dos aportes. El primero basado en los modelos para la función de covarianza y aplicados a datos espaciales continuos (geoestadística). El segundo es basado en las medidas de proximidad espacial y aplicados a datos regionales.

Primero se enfoca en modelos geoestadísticos para  $\Sigma(\theta)$  y en ellos se usan modelos de regresión lineal con errores linealmente correlacionados. Lo que concierne ahora es la estimación de  $\theta$  para hacer inferencia sobre  $\beta$  y en la predicción del proceso  $Z(s)$ .

Si  $\theta$  es conocida, el estimador mínimo generalizado

$$\hat{\beta}_{gls} = (X(s)'\Sigma(\theta)^{-1}X(s))^{-1}X(s)'\Sigma(\theta)^{-1}Z(s)$$

puede ser usado para estimar  $\beta$ . Con el fin de utilizar el estimador para inferencia estadística, se necesita estar seguro que es un estimador consistente de  $\beta$  y luego determinar sus propiedades distribucionales. Si se asume que los errores,  $e(s)$ , tienen una distribución normal, entonces el estimador máximo verosímil de  $\beta$  es equivalente al estimador de mínimos cuadrados generalizados. Entonces, el estimador de mínimos cuadrados cuadrados generalizados es consistente para  $\beta$  y  $\hat{\beta}_{gls} \sim G(\beta, (X(s)\Sigma(\theta)^{-1}X(s))^{-1})$ . Sin embargo, si los errores no son normales, estas propiedades no son garantizadas. La consistencia de  $\hat{\beta}_{gls}$  depende de la matriz  $X(s)$  y la matriz de covarianza  $\Sigma(\theta)$ . Una condición que asegurará la consistencia de  $\hat{\beta}_{gls}$  para  $\beta$  es

$$\lim_{n \rightarrow \infty} \frac{(X(s)'\Sigma(\theta)^{-1}X(s))}{n} = Q$$

donde  $Q$  es una matriz finita y no singular. Las propiedades asintóticas de  $\hat{\beta}_{gls}$  son

derivadas de las propiedades asintóticas de

$$\sqrt{n}(\hat{\beta}_{gls} - \beta) = \left( \frac{X(s)' \Sigma(\theta)^{-1} X(s)}{n} \right)^{-1} \frac{X(s)' \Sigma(\theta)^{-1} e(s)}{\sqrt{n}}$$

El teorema del límite central dado en casi todos los libros de estadística básica no es relevante directamente en este problema desde  $X(s)' \Sigma(\theta)^{-1} e(s)$  no es una suma de variables aleatorias independiente e idénticamente distribuidas. Una condición que puede ser aplicada en este caso es llamada el teorema del límite central de Lindberg-Feller, la cual puede ser usada para mostrar que (Schimidt (1976))

$$\hat{\beta}_{gls} \sim G(\beta, (X(s)' \Sigma(\theta)^{-1} X(s))^{-1}).$$

Si  $\theta$  es desconocido, mínimos cuadrados generalizados ponderados iterativamente pueden ser usados para estimar  $\beta$  y  $\theta$  en el modelo general. Este da un estimador de mínimos cuadrados generalizados para efectos fijos  $\beta$ ,

$$\hat{\beta}_{egls} = (X(s)' \Sigma(\tilde{\theta})^{-1} X(s))^{-1} X(s)' \Sigma(\tilde{\theta})^{-1} Z(s)$$

donde  $\tilde{\theta}$  es el estimador de los parámetros de covarianza, obtenido por métodos de estimación en el caso de diferentes medias espaciales. Se está interesado con los efectos del uso de los parámetros estimados de covarianza en las conexiones de algunas expresiones.

La derivación de las propiedades distribucionales de  $\hat{\beta}_{egls}$  se dificulta por que  $\Sigma(\tilde{\theta})$  y  $e(s)$  pueden estar correlacionadas. Entonces, se retoman los resultados asintóticos adicionales

$$\lim_{n \rightarrow \infty} n^{-1} X(s)' [\Sigma(\tilde{\theta}) - \Sigma(\theta)] X(s) \xrightarrow{p} 0$$

y

$$\lim_{n \rightarrow \infty} n^{-1/2} X(s)' [\Sigma(\tilde{\theta}) - \Sigma(\theta)] X(s) \xrightarrow{p} 0$$

entonces  $\hat{\beta}_{egls}$  tienen la misma distribución en el límite de  $\hat{\beta}_{gls}$  y entonces

$$\sqrt{n}(\hat{\beta}_{egls} - \beta) \xrightarrow{d} G(0, Q^{-1})$$

(Theil (1971), Schimidt (1976)). Entonces, si estas condiciones se mantienen

$$\hat{\beta}_{egls} \sim G(\beta, (X(s)'\Sigma(\theta)^{-1}X(s))^{-1})$$

En muchos casos, estas condiciones se reúnen si  $\tilde{\theta}$  es un estimador consistente de  $\theta$ . Sin embargo, la consistencia de  $\tilde{\theta}$  no es suficiente: condiciones suficientes son dadas en Fuller and Battese (1973), quienes muestran que estas condiciones son mantenidas para algunos modelos estructurales de error anidados y Theil (1971) da condiciones para un proceso de series temporales AR(1).

Se puede derivar el chequeo de estas condiciones usando métodos de máxima verosimilitud para estimar  $\beta$  y  $\theta$  simultáneamente.

### 3.4.3. Modelos autorregresivos espaciales

En el contexto de series temporales los modelos autorregresivos (AR) indican que existe una relación lineal entre lo que ocurre con la variable en un periodo de tiempo  $t$  y lo que ocurrió con ella en periodos anteriores, es decir si  $\{Z(t), t \in \mathbb{R}\}$  representa una serie de tiempo (estacionaria, con media cero), un modelo AR( $p$ ), indica que se tiene el modelo  $Z(t) = \rho_1 Z(t-1) + \dots + \rho_p Z(t-p) + \epsilon(t)$ , con  $\epsilon(t)$  un ruido blanco. El análogo a dicho modelo en el contexto espacial se presenta cuando para el proceso espacial  $\{Z(s), s \in D \subset \mathbb{R}^2\}$  se plantea el modelo lineal que relaciona  $Z(s_i)$  con las variables en los sitios vecinos ( $\mathbf{Z}(s)_{-i}$ ), donde  $\mathbf{Z}(s)_{-i}$  representa el vector de todas las variables excepto  $Z(s_i)$ . Existen varios tipos de modelos autorregresivos espaciales. Los dos más usados en la práctica son

- Modelo autorregresivo simultáneo
- Modelo autorregresivo condicional

A continuación se describen dichos modelos y se muestra como hacer estimación y pruebas de hipótesis en ambos casos.

### 3.4.4. Modelo autorregresivo simultáneo (SAR)

En este modelo se asigna una estructura de autocorrelación espacial a los residuales de un modelo de regresión lineal, usando para ello un matriz de proximidad espacial entre las

áreas. Este también es conocido como *modelo de errores correlacionados* o de *variables espacialmente rezagadas*. El modelo SAR es planteado como

$$\begin{aligned}\mathbf{Z}(s) &= \mathbf{X}(s)\boldsymbol{\beta} + \mathbf{U}(s), \\ \mathbf{U}(s) &= \rho\mathbf{W}\mathbf{U}(s) + \boldsymbol{\epsilon}(s),\end{aligned}\tag{3.1}$$

donde  $\mathbb{E}(\boldsymbol{\epsilon}(s)) = \mathbf{0}$  y  $\mathbb{V}(\boldsymbol{\epsilon}(s)) = \sigma^2\mathbf{I}$ . En la ecuación (3.1)  $\mathbf{W}$  es una matriz de proximidad (ver sección 3.2.1). Es claro que si  $\rho = 0$  se tiene un modelo de regresión tradicional. El modelo también puede ser expresado como

$$\begin{aligned}\mathbf{Z}(s) &= \mathbf{X}(s)\boldsymbol{\beta} + \rho\mathbf{W}\mathbf{U}(s) + \boldsymbol{\epsilon}(s) \\ &= \mathbf{X}(s)\boldsymbol{\beta} + \rho\mathbf{W}(\mathbf{Z}(s) - \mathbf{X}(s)\boldsymbol{\beta}) + \boldsymbol{\epsilon}(s) \\ &= \mathbf{X}(s)\boldsymbol{\beta} + \rho\mathbf{W}\mathbf{Z}(s) + \rho\mathbf{W}\mathbf{X}(s)\boldsymbol{\beta} + \boldsymbol{\epsilon}(s)\end{aligned}\tag{3.2}$$

$$\begin{aligned}\mathbf{Z}(s) - \rho\mathbf{W}\mathbf{Z}(s) &= \mathbf{X}(s)\boldsymbol{\beta} + \rho\mathbf{W}\mathbf{X}(s)\boldsymbol{\beta} + \boldsymbol{\epsilon}(s) \\ \boldsymbol{\epsilon}(s) &= (\mathbf{I} - \rho\mathbf{W})(\mathbf{Z}(s) - \mathbf{X}(s)\boldsymbol{\beta}).\end{aligned}\tag{3.3}$$

De (3.1), se tiene que  $\mathbb{E}(\mathbf{U}(s)) = (\mathbf{I} - \rho\mathbf{W})^{-1}\mathbb{E}(\boldsymbol{\epsilon}(s)) = \mathbf{0}$  y por consiguiente  $\mathbb{E}(\mathbf{Z}(s)) = \mathbf{X}(s)\boldsymbol{\beta}$ . Además la varianza de  $\mathbf{Z}(s)$  está dada por

$$\begin{aligned}\Sigma_{SAR} &= \mathbb{V}(\mathbf{Z}(s)) \\ &= \mathbb{V}(\mathbf{U}(s)) \\ &= \mathbb{V}[(\mathbf{I} - \rho\mathbf{W})^{-1}\boldsymbol{\epsilon}(s)] \\ &= (\mathbf{I} - \rho\mathbf{W})^{-1}\sigma^2\mathbf{I}((\mathbf{I} - \rho\mathbf{W})^{-1})^T \\ &= (\mathbf{I} - \rho\mathbf{W})^{-1}\sigma^2\mathbf{I}(\mathbf{I} - \rho\mathbf{W}^T)^{-1} \\ &= \sigma^2((\mathbf{I} - \rho\mathbf{W})^{-1}(\mathbf{I} - \rho\mathbf{W}^T)^{-1})\end{aligned}\tag{3.4}$$

En la ecuación (3.2) los términos  $\rho\mathbf{W}\mathbf{Z}(s)$  y  $\rho\mathbf{W}\mathbf{X}(s)\boldsymbol{\beta}$  son llamados variables espacialmente rezagadas y de ahí que el modelo SAR sea conocido con este nombre. Dos simplificaciones del modelo (3.2) son  $\mathbf{Z}(s) = \mathbf{X}(s)\boldsymbol{\beta} + \rho\mathbf{W}\mathbf{Z}(s) + \boldsymbol{\epsilon}(s)$  y  $\mathbf{Z}(s) = \rho\mathbf{W}\mathbf{Z}(s) + \boldsymbol{\epsilon}(s)$ . El segundo de ellos es conocido como *modelo autorregresivo puro*. El modelo (3.1) está bien

especificado si la matriz  $(\mathbf{I} - \rho \mathbf{W})$  es invertible. Esto implica que se pongan algunas condiciones sobre  $\mathbf{W}$  y sobre  $\rho$ . Si la matriz  $\mathbf{W}$  es definida de tal forma que las filas sumen uno (dividiendo cada  $w_{ij}$  por la suma  $\sum_j w_{ij}$ ), se garantiza que  $\rho < 1$ .

### Estimación e inferencia en el modelo SAR

En el caso de que se haya definido una matriz de proximidad entre las áreas ( $\mathbf{W}$ ) y asumiendo conocido  $\rho$  puede usarse el método de mínimos cuadrados generalizados para hacer la estimación del vector  $\beta$  y del parámetro  $\sigma^2$  en el modelo (3.1). Los correspondientes estimadores están dados por

$$\hat{\beta} = (\mathbf{X}(s)^T \Sigma_{SAR}^{-1} \mathbf{X}(s))^{-1} \mathbf{X}(s)^T \Sigma_{SAR}^{-1} \mathbf{Z}(s),$$

$$\hat{\sigma}^2 = \frac{(\mathbf{Z}(s) - \mathbf{X}(s)\hat{\beta})^T \Sigma_{SAR}^{-1} (\mathbf{Z}(s) - \mathbf{X}(s)\hat{\beta})}{n - k},$$

donde  $\Sigma_{SAR}^{-1}$  está definido como en (3.4).

### Estimación en el modelo SAR usando máxima verosimilitud

$$Z(s) \sim N(X(s)\beta, \sigma^2(I - \rho W)^{-1}((I - \rho W)^{-1})^t)$$

$$y \sim N(\mu, \sigma^2), f(y) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(y - \mu)^T \Sigma^{-1} (y - \mu)\right\}$$

$$\begin{aligned} f(z(s)) &= \frac{1}{(2\pi)^{n/2} |\Sigma_{SAR}|^{1/2}} \exp\left\{-\frac{1}{2}(z(s) - X(s)\beta)^T \Sigma_{SAR}^{-1} (z(s) - X(s)\beta)\right\} \\ \ln(f(z(s))) &= -\frac{n}{2} \ln(2\pi) + \frac{1}{2} \ln|\Sigma_{SAR}| - \frac{1}{2}(z(s) - X(s)\beta)^T \Sigma_{SAR}^{-1} (z(s) - X(s)\beta) \\ -2\ln(f(z(s))) &= n \ln(2\pi) - \ln|\Sigma_{SAR}| + (z(s) - X(s)\beta)^T \Sigma_{SAR}^{-1} (z(s) - X(s)\beta) \end{aligned}$$

Para encontrar los estimadores, se debe maximizar  $-2\ln(f(z(s)))$  respecto a  $\rho, \sigma^2$  y  $\beta$ , con  $\Sigma_{SAR} = \sigma^2(I - \rho W)^{-1}((I - \rho W)^{-1})^t$

### Procedimiento iterativo

1. Estime  $\beta$  usando mínimos cuadrados ordinarios con el modelo  $Z(s) = X(s)\beta + \epsilon(s)$

2. Maximize  $-2\log(f(z(s)))$  usando  $\hat{\beta}$  y obtenga  $\hat{\sigma}^2$  y  $\hat{\rho}$

3. Reestime  $\beta$  usando mínimos cuadrados generalizados

$$\hat{\beta} = (X^t(s)\Sigma_{SAR}^{-1}X(s))^t(X^t(s)\Sigma_{SAR}^{-1}Z(s))^t$$

4. Repita 2) y 3) hasta que haya convergencia en las estimaciones de  $\hat{\beta}, \hat{\sigma}^2$  y  $\hat{\rho}$ .

### 3.4.5. Modelo autorregresivo condicional (CAR)

$$E(Z(s_i)/Z(s)_{-i}) = X^t(s_i)\beta + \sum_{j=1}^n C_{ij}(Z(s_j) - X^t(s_i)\beta)$$

$$V(Z(s_i)/Z(s)_{-i}) = \sigma_i^2$$

Si  $Z(s_i)/Z(s)_{-i} \sim \text{Normal}$ , se puede demostrar que :

$$\begin{bmatrix} Z(1) \\ Z(2) \\ \vdots \\ Z(n) \end{bmatrix} \sim N_n(X(s)\beta, (I - C)^{-1}\Sigma_C), \text{ con } \Sigma_c = \text{Diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2)$$

Para asegurar que  $\Sigma_{CAR} = (I - C)^{-1}\Sigma_c$  sea simétrica, se requiere que  $\sigma^2 c_{ij} = \sigma_i^2 c_{ji}$ .

Nota

Si  $C = \rho W$ , entonces  $Z(s) = X(s)\beta + \rho W(Z(s) - X(s)\beta) + \epsilon(s)$

#### Estimación en el modelo CAR usando máxima verosimilitud

$$Z(s) \sim N(X(s)\beta, (I - C)^{-1}\Sigma_c)$$

$$y \sim N(\mu, \sigma^2), f(y) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(y - \mu)^t \Sigma^{-1} (y - \mu)\right\}$$

$$\begin{aligned}
 f(z(s)) &= \frac{1}{(2\pi)^{n/2}|\Sigma_{CAR}|^{1/2}} \exp\left\{-\frac{1}{2}(z(s) - X(s)\beta)^t \Sigma_{CAR}^{-1} (z(s) - X(s)\beta)\right\} \\
 \ln(f(z(s))) &= -\frac{n}{2}\ln(2\pi) + \frac{1}{2}\ln|\Sigma_{CAR}| - \frac{1}{2}(z(s) - X(s)\beta)^t \Sigma_{CAR}^{-1} (z(s) - X(s)\beta) \\
 -2\ln(f(z(s))) &= n\ln(2\pi) - \ln|\Sigma_{CAR}| + (z(s) - X(s)\beta)^t \Sigma_{CAR}^{-1} (z(s) - X(s)\beta)
 \end{aligned}$$

Para encontrar los estimadores, se debe maximizar  $-2\ln(f(z(s)))$  respecto a  $\rho, \sigma^2$  y  $\beta$ , con  $\Sigma_{CAR} = (I - C)^{-1}\Sigma_c$

# Capítulo 4

## Geoestadística

### 4.1. Definición de Geoestadística

Geoestadística es una rama de la estadística que trata fenómenos espaciales Journel and Huijbregts (1978a). Su interés primordial es la estimación, predicción y simulación de dichos fenómenos Myers (1987). Esta herramienta ofrece una manera de describir la continuidad espacial, que es un rasgo distintivo esencial de muchos fenómenos naturales, y proporciona adaptaciones de las técnicas clásicas de regresión para tomar ventajas de esta continuidad Isaaks and Srivastava (1989). Petitgas (1996), la define como una aplicación de la teoría de probabilidades a la estimación estadística de variables espaciales. La modelación espacial es la adición más reciente a la literatura estadística. Geología, ciencias del suelo, agronomía, ingeniería forestal, astronomía, o cualquier disciplina que trabaja con datos colectados en diferentes locaciones espaciales necesita desarrollar modelos que indiquen cuando hay dependencia entre las medidas de los diferentes sitios. Usualmente dicha modelación concierne con la predicción espacial, pero hay otras áreas importantes como la simulación y el diseño muestral Cressie (1989). Cuando el objetivo es hacer predicción, la geoestadística opera básicamente en dos etapas. La primera es el análisis estructural, en la cual se describe la correlación entre puntos en el espacio. En la segunda fase se hace predicción en sitios de la región no muestreados por medio de la técnica kriging (capítulo 4). Este es un proceso que calcula un promedio ponderado de las observaciones mues-

trales. Los pesos asignados a los valores muestrales son apropiadamente determinados por la estructura espacial de correlación establecida en la primera etapa y por la configuración de muestreo Petitgas (1996). Los fundamentos básicos de estas etapas son presentados a continuación.

## 4.2. Variable Regionalizada

Una variable medida en el espacio de forma que presente una estructura de correlación, se dice que es una variable regionalizada. De manera más formal se puede definir como un proceso estocástico con dominio contenido en un espacio euclíadiano  $d$ -dimensional  $\mathbb{R}^d, Z(x) : x \in D \subset \mathbb{R}^d$ . Si  $d = 2$ ,  $Z(x)$  puede asociarse a una variable medida en un punto  $x$  del plano Díaz-Francés (1993). En términos prácticos  $Z(x)$  puede verse como una medición de una variable aleatoria (p.ej. concentración de un contaminante) en un punto  $x$  de una región de estudio. Recuérdese que un proceso estocástico es una colección de variables aleatorias indexadas; esto es, para cada  $x$  en el conjunto de índices  $D$ ,  $Z(x)$  es una variable aleatoria. En el caso de que las mediciones sean hechas en una superficie, entonces  $Z(x)$  puede interpretarse como la variable aleatoria asociada a ese punto del plano ( $x$  representa las coordenadas, planas o geográficas, y  $Z$  la variable en cada una de ellas). Estas variables aleatorias pueden representar la magnitud de una variable ambiental medida en un conjunto de coordenadas de la región de estudio.

### 4.2.1. Momentos de una Variable Regionalizada

Sea  $Z(x) : x \in D \subset \mathbb{R}^d$  el proceso estocástico que define la variable regionalizada. Para cualquier  $n$  puntos  $x_1, x_2, \dots, x_n$ , el vector aleatorio  $\vec{Z}(x + h) = [Z(x_1), Z(x_2), \dots, Z(x_n)]^T$  está definido por su función de distribución conjunta  $F[z_1, z_2, \dots, z_n] = P[Z(x_1) \leq z_1, Z(x_2) \leq z_2, \dots, Z(x_n) \leq z_n]$  Conocidas las densidades marginales univariadas y bivariadas se pueden establecer los siguientes valores esperados (momentos univariados y bivariados):

- $E(Z(x_i)) = m(x_i)$
- $V(Z(x_i)) = E[Z(x_i) - m(x_i)]^2 = \sigma_i^2$

- $C(Z(x_i), Z(x_j)) = E[Z(x_i) - m(x_i)][Z(x_j) - m(x_j)]$ : Función de autocovarianza.
- $\gamma(Z(x_i), Z(x_j)) = \frac{1}{2}E[Z(x_i) - Z(x_j)]^2$ : Función de semivarianza.

### 4.2.2. Estacionariedad

La variable regionalizada es estacionaria si su función de distribución conjunta es invariante respecto a cualquier translación del vector  $h$ , o lo que es lo mismo, la función de distribución del vector aleatorio  $\vec{Z}(x) = [Z(x_1), Z(x_2), \dots, Z(x_n)]^T$  es idéntica a la del vector  $\vec{Z}(x + h) = [Z(x_1 + h), Z(x_2 + h), \dots, Z(x_n + h)]^T$  para cualquier  $h$ . La teoría geoestadística se basa en los momentos arriba descritos y la hipótesis de estacionariedad puede definirse en términos de estos:

#### Estacionariedad de Segundo Orden

Sea  $Z(x) : x \in D \subset \mathbb{R}^d$  una variable regionalizada definida en un dominio  $D$  contenido en  $\mathbb{R}^d$  (generalmente una variable medida en la superficie de una región) se dice que  $Z(x)$  es estacionario de segundo orden si cumple:

- a.  $E[Z(x)] = m, k \in \mathbb{R}, \forall x \in D \subset \mathbb{R}^d$ . El valor esperado de la variable aleatoria es finito y constante para todo punto en el dominio.
- b.  $COV[Z(x), Z(x + h)] = C(h) < \infty$  Para toda pareja  $Z(x), Z(x + h)$  la covarianza existe y es función única del vector de separación  $h$ .

En la figura 4.1 se muestra el gráfico de una variable regionalizada estacionaria. Exceptuando fluctuaciones aleatorias, el valor promedio de la variable no muestra una tendencia definida en alguna dirección. La existencia de la covarianza implica que la varianza existe, es finita y no depende de  $h$ , es decir  $V(Z(x_i)) = C(0) = \sigma^2$ . Así mismo la estacionariedad de segundo orden implica la siguiente relación entre la función de semivarianza y la de autocovarianza:

$$\begin{aligned}
 \gamma(Z(x+h), Z(x)) &= \gamma(h) = \frac{1}{2}E[Z(x+h) - m - Z(x) + m]^2 \\
 &= \frac{1}{2}E(Z(x+h) - m)^2 + E(Z(x) - m)^2 - 2E(Z(x+h) - m)(Z(x) - m) \\
 &= \frac{1}{2}\sigma^2 + \frac{1}{2}\sigma^2 - E(Z(x+h) - m)(Z(x) - m) \\
 &= \sigma^2 - C(h).
 \end{aligned}$$

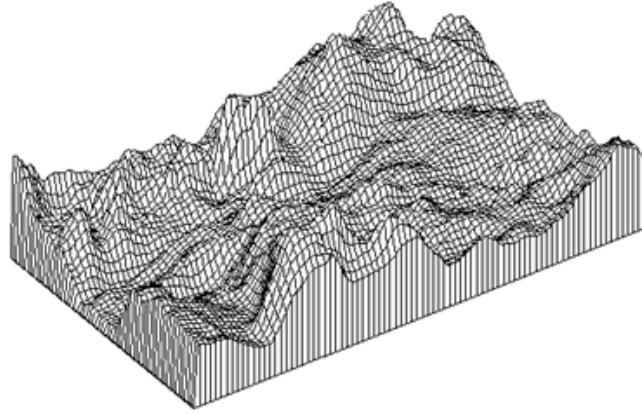


Figura 4.1: Representación de una superficie interpolada para una variable regionalizada estacionaria.

### Estacionariedad Débil o Intrínseca

Existen algunos fenómenos físicos reales en los que la varianza no es finita. En estos casos se trabaja sólo con la hipótesis que pide que los incrementos  $[Z(x+h) - Z(x)]$  sean estacionarios, esto es Clark (1979):

- a.  $Z(x)$  tiene esperanza finita y constante para todo punto en el dominio. Lo que implica que la esperanza de los incrementos es cero.

$$E[Z(x+h) - Z(x)] = 0$$

- b. Para cualquier vector  $h$ , la varianza del incremento está definida y es una función única de la distancia.

$$V[Z(x + h) - Z(x)] = E[Z(x + h) - Z(x)]^2 = 2\gamma(h)$$

Es claro que si una variable regionalizada es estacionaria fuerte entonces también será estacionaria débil. El concepto de estacionariedad es muy útil en la modelación de series temporales Box and Jenkins (1976). En este contexto es fácil la identificación, puesto que sólo hay una dirección de variación (el tiempo). En el campo espacial existen múltiples direcciones y por lo tanto se debe asumir que en todas el fenómeno es estacionario. Cuando la esperanza de la variable no es la misma en todas las direcciones o cuando la covarianza o correlación dependan del sentido en que se determinan, no habrá estacionariedad. Si la correlación entre los datos no depende de la dirección en la que esta se calcule se dice que el fenómeno es isotrópico, en caso contrario se hablará de anisotropía. En Isaaks and Srivastava (1989) se definen los posibles tipos de anisotropía y se proponen algunas soluciones. Cressie (1993) discute cual debe ser el tratamiento en caso de que la media no sea constante. En casos prácticos resulta compleja la identificación de la estacionariedad. Suelen emplearse gráficos de dispersión de la variable respecto a las coordenadas, de medias móviles y de valores clasificados según puntos de referencia, con el propósito de identificar posibles tendencias de la variable en la región de estudio. La isotropía es estudiada a través del cálculo de funciones de autocovarianza o de semivarianza muestrales (capítulo3) en varias direcciones. Si estas tienen formas considerablemente distintas puede no ser válido el supuesto de isotropía. Finalmente una variable regionalizada será no estacionaria si su esperanza matemática no es constante, esto es si . En la figura ?? se representa una variable regionalizada en la que existe tendencia en el valor promedio de la variable, lo cual es claro indicador de no estacionariedad.

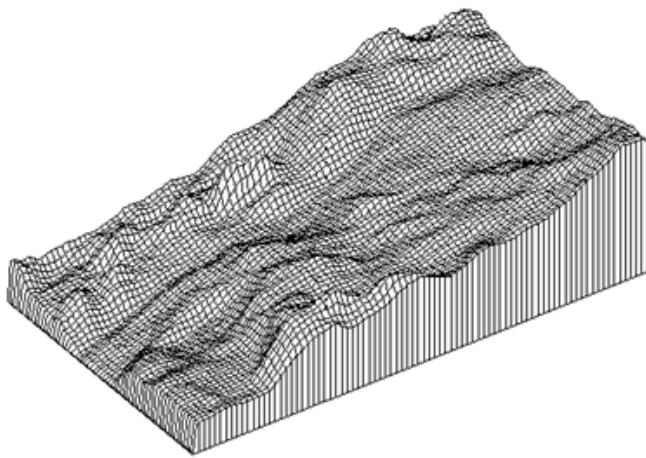


Figura 4.2: Representación de una superficie interpolada para una variable regionalizada no estacionaria.

## 4.3. Correlación espacial muestral y ajuste de modelos

### 4.3.1. Funciones de correlación espacial

La primera etapa en el desarrollo de un análisis geoestadístico es la determinación de la dependencia espacial entre los datos medidos de una variable. Esta fase es también conocida como análisis estructural. Para llevarla a cabo, con base en la información muestral, se usan tres funciones: El semivariograma, el covariograma y el correlograma. A continuación se hace una revisión de los conceptos asociados a cada una de ellas y se describen sus bondades y limitaciones.

#### Variograma y semivariograma

Cuando se definió la estacionariedad débil en el capítulo anterior se mencionó que se asumía que la varianza de los incrementos de la variable regionalizada era finita. A esta función denotada por  $2\gamma(h)$  se le denomina variograma. Utilizando la definición teórica de

la varianza en términos del valor esperado de una variable aleatoria, tenemos:

$$\begin{aligned} 2\gamma(h) &= V(Z(x+h) - Z(x)) \\ &= E((Z(x+h) - Z(x))^2) - (E(Z(x+h) - Z(x)))^2 \\ &= E((Z(x+h) - Z(x))^2) \end{aligned}$$

La mitad del variograma  $\gamma(h)$ , se conoce como la función de semivarianza y caracteriza las propiedades de dependencia espacial del proceso. Dada una realización del fenómeno, la función de semivarianza es estimada, por el método de momentos, a través del semivariograma experimental, que se calcula mediante Wackernagel (1995):

$$\bar{\gamma}(h) = \frac{\sum(Z(x+h) - Z(x))^2}{2n}$$

donde  $Z(x)$  es el valor de la variable en un sitio  $x$ ,  $Z(x+h)$  es otro valor muestral separado del anterior por una distancia  $h$  y  $n$  es el número de parejas que se encuentran separadas por dicha distancia. La función de semivarianza se calcula para varias distancia  $h$ . En la práctica, debido a irregularidad en el muestreo y por ende en las distancias entre los sitios, se toman intervalos de distancia  $[0, h], (h, 2h], (2h, 3h], \dots$  y el semivariograma experimental corresponde a una distancia promedio entre parejas de sitios dentro de cada intervalo y no a una distancia  $h$  específica. Obviamente el número de parejas de puntos  $n$  dentro de los intervalos no es constante.

Para interpretar el semivariograma experimental se parte del criterio de que a menor distancia entre los sitios mayor similitud o correlación espacial entre las observaciones. Por ello en presencia de autocorrelación se espera que para valores de  $h$  pequeños el semivariograma experimental tenga magnitudes menores a las que este toma cuando las distancias  $h$  se incrementan.

### Covariograma y correlograma

La función de covarianza muestral entre parejas de observaciones que se encuentran a una distancia  $h$  se calcula, empleando la formula clásica de la covarianza muestral, por:

$$\begin{aligned} C(h) = COV(Z(x+h), Z(x)) &= \frac{\sum_{i=1}^n (Z(x+h) - m)(Z(x) - m)}{n} \\ &= \frac{\sum_{i=1}^n (Z(x+h)Z(x))}{n} - m^2 \end{aligned}$$

donde  $m$  representa el valor promedio en todo punto de la región de estudio y  $n$  es el número de parejas de puntos que se encuentran a una distancia  $h$ . En este caso es también válida la aclaración respecto a las distancias dadas en el último párrafo de la página anterior. Asumiendo que el fenómeno es estacionario y estimando la varianza de la variable regionalizada a través de la varianza muestral, se tiene que el correlograma muestral está dado por:

$$r(h) = \frac{COV(Z(x+h), Z(x))}{S_{h+h}S_x} = \frac{C(h)}{S_x^2} = \frac{C(h)}{C(0)}$$

Bajo el supuesto de estacionariedad cualquiera de las tres funciones de dependencia espacial mencionadas, es decir semivariograma, covariograma o correlograma, puede ser usada en la determinación de la relación espacial entre los datos. Sin embargo como se puede observar en las fórmulas, la única que no requiere hacer estimación de parámetros es la función de semivarianza. Por esta razón, fundamentalmente, en la práctica se emplea el semivariograma y no las otras dos funciones.

#### 4.3.2. Modelos teóricos de semivarianza

Como se verá a partir del capítulo cuatro la solución del problema de predicción espacial kriging requiere del conocimiento de la estructura de autocorrelación para cualquier posible distancia entre sitios dentro del área de estudio. En la presentación del semivariograma experimental dada anteriormente se indicó que este es calculado sólo para algunas distancias promedios particulares. Por ello se hace necesario el ajuste de modelos que generalicen lo observado en el semivariograma experimental a cualquier distancia. Existen diversos modelos teóricos de semivarianza que pueden ajustarse al semivariograma experimental. En Samper and Carrera (1990) se presenta una discusión respecto a las características y condiciones que éstos deben cumplir. En general dichos modelos pueden

dividirse en no acotados (lineal, logarítmico, potencial) y acotados (esférico, exponencial, gaussiano) (Warrick (1986)). Los del segundo grupo garantizan que la covarianza de los incrementos es finita, por lo cual son ampliamente usados cuando hay evidencia de que presentan buen ajuste. Todos estos modelos tienen tres parámetros comunes (Fig. 4.3) que son descritos a continuación:

- **Efecto Pepita**

Se denota por  $C_0$  y representa una discontinuidad puntual del semivariograma en el origen (figura 4.3). Puede ser debido a errores de medición en la variable o a la escala de la misma. En algunas ocasiones puede ser indicativo de que parte de la estructura espacial se concentra a distancias inferiores a las observadas.

- **Meseta**

Es la cota superior del semivariograma. También puede definirse como el límite del semivariograma cuando la distancia  $h$  tiende a infinito. La meseta puede ser o no finita. Los semivariogramas que tienen meseta finita cumplen con la hipótesis de estacionariedad fuerte; mientras que cuando ocurre lo contrario, el semivariograma define un fenómeno natural que cumple sólo con la hipótesis intrínseca. La meseta se denota por  $C_1$  o por  $(C_0 + C_1)$  cuando la pepita es diferente de cero. Si se interpreta la pepita como un error en las mediciones, esto explica porque se sugiere que en un modelo que explique bien la realidad, la pepita no debe representar mas del 50 % de la meseta. Si el ruido espacial en las mediciones explica en mayor proporción la variabilidad que la correlación del fenómeno, las predicciones que se obtengan pueden ser muy imprecisas. En la figura 4.3 se representa este parámetro para el caso de uno de los modelos acotados.

- **Rango**

En términos prácticos corresponde a la distancia a partir de la cual dos observaciones son independientes. El rango se interpreta como la zona de influencia. Existen algunos modelos de semivariograma en los que no existe una distancia finita para la cual dos observaciones sean independientes; por ello se llama rango efectivo a la distancia

para la cual el semivariograma alcanza el 95 % de la meseta. Entre más pequeño sea el rango, más cerca se está del modelo de independencia espacial. El rango no siempre aparece de manera explícita en la fórmula del semivariograma. En el caso del modelo esférico, el rango coincide con el parámetro  $a$ , que se utilizará en las ecuaciones más adelante. Sin embargo, en el modelo exponencial, el rango efectivo es  $\frac{a}{3}$  y en el modelo gaussiano es  $\frac{a}{\sqrt{3}}$ .

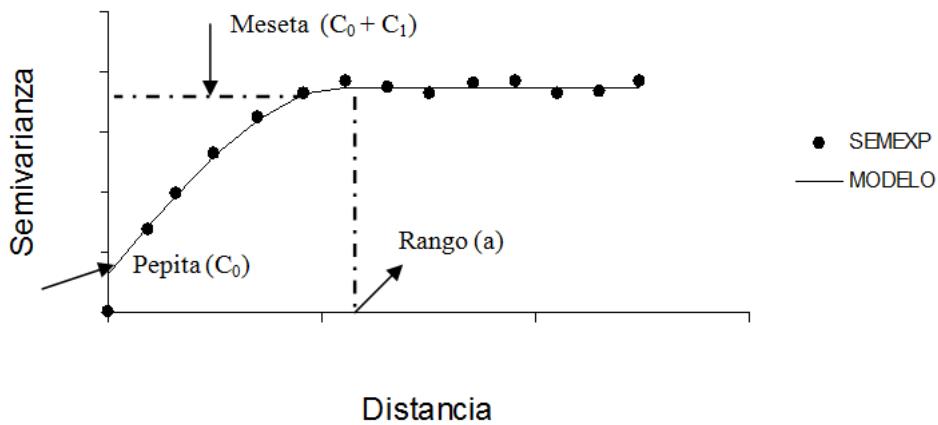


Figura 4.3: Comportamiento típico de un semivariograma acotado con una representación de los parámetros básicos. SEMEXP corresponde al semivariograma experimental y MODELO al ajuste de un modelo teórico.

### Modelo Esférico

Tiene un crecimiento rápido cerca al origen (Fig. 4.4), pero los incrementos marginales van decreciendo para distancias grandes, hasta que para distancias superiores al rango los incrementos son nulos. Su expresión matemática es la siguiente:

$$\gamma(h) = \begin{cases} C_0 + C_1 \left( \frac{3}{2} \left( \frac{h}{a} \right) - \frac{1}{2} \left( \frac{h}{a} \right)^3 \right) & h \leq a \\ C_0 + C_1 & h > a \end{cases}$$

En donde  $C_1$  representa la meseta,  $a$  el rango y  $h$  la distancia.

## Modelo exponencial

Este modelo se aplica cuando la dependencia espacial tiene un crecimiento exponencial respecto a la distancia entre las observaciones. El valor del rango es igual a la distancia para la cual el semivariograma toma un valor igual al 95 % de la meseta (Fig. 4.4). Este modelo es ampliamente usado. Su expresión matemática es la siguiente:

$$\gamma(h) = C - 0 + C_1 \left( 1 - \exp \left( \frac{-3h}{a} \right) \right)$$

## Modelo gaussiano

Al igual que en el modelo exponencial, la dependencia espacial se desvanece solo en una distancia que tiende a infinito. El principal distintivo de este modelo es su forma parabólica cerca al origen (Fig. 4.4). Su expresión matemática es:

$$\gamma(h) = C - 0 + C_1 \left( 1 - \exp \left( \frac{-h}{a^2} \right) \right)$$

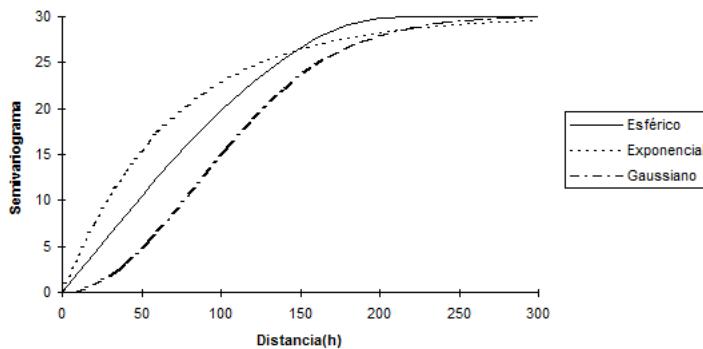


Figura 4.4: Comparación de los modelos exponencial, esférico y Gaussiano. La línea punteada vertical representa el rango en el caso del modelo esférico y el rango efectivo en el de los modelos exponencial y gaussiano. Este tiene un valor de 210, respecto a una escala simulada entre 0 y 300. El valor de la meseta es 30 y el de la pepita 0. El 95 % de la meseta es igual a 28.5.

### Modelo Matérn

Esta familia lleva el nombre de Matern Bertil, quien lo introdujo en su tesis doctoral sueca en 1960, posteriormente renombrado como Matern (1986). Se define por

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

donde  $(\phi, \kappa)$  son parámetros y  $K_\kappa(\cdot)$  denota la función modificada de Bessel de tercera clase de orden  $\kappa$ .

Esta familia es valida para  $\phi > 0$  y  $\kappa > 0$ . El caso  $\kappa = 0,5$  es el mismo de la función de correlación exponencial,  $\rho(u) = \exp(-u/\phi)$ . La función de correlación Gaussiana, es el caso donde  $\kappa \rightarrow \infty$ .

Un característica atractiva de este modelo es que el parámetro  $\kappa$  controla la diferenciabilidad del proceso subyacente  $S(x)$  de forma directa; la parte entera de  $\kappa$  da el número de veces que  $S(x)$  es diferenciable en media cuadrática. Por lo tanto, como ya hemos visto,  $\kappa = 0,05$  corresponde a un proceso continuo, pero no diferenciable,  $\kappa = 1$  es diferenciable, y mayores valores de  $\kappa$  se pueden elegir para dar derivados adicionales, lo cual es deseable.

### Modelo monómicos

Corresponden a los modelos que no alcanzan la meseta (Fig. 4.5). Su uso puede ser delicado debido a que en algunos casos indican la presencia de no estacionariedad en alguna dirección.

Su fórmula matemática es la siguiente:

$$\gamma(h) = kh^\theta \quad 0 < \theta < 2$$

Obviamente cuando el parámetro  $\theta$  es igual a uno, el modelo es lineal y  $k$  representa la pendiente de la ecuación de regresión con intercepto cero. Gráficamente se pueden representar así:

### Modelo de independencia (Pepita Puro).

Es indicativo de carencia de correlación espacial entre las observaciones de una variable (Fig. 4.6). Es común sumar este modelo a otro modelo teórico de semivarianza, para

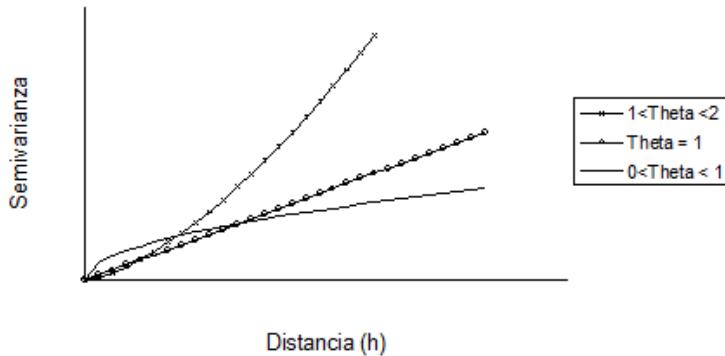


Figura 4.5: Comportamiento típico de los modelos de semivarianza monómicos.

obtener lo que se conoce como semivariograma anidado. Lo anterior se sustenta en una propiedad de los semivariogramas que dice que cualquier combinación lineal de semivariogramas con coeficientes positivos es un semivariograma. Su expresión matemática es:

$$\gamma(h) = \begin{cases} 0 & h = 0 \\ C_0 & h > 0 \quad \text{donde} \quad C_0 > 0 \end{cases}$$

Su representación gráfica es la siguiente:



Figura 4.6: Modelo de semivarianza teórico para variables sin correlación espacial.

La estimación de los parámetros de los modelos teóricos descritos puede ser llevada a cabo, entre otros métodos, por máxima verosimilitud (Cressie (1993)) o regresión no lineal (Gotway, 1991). Algunos paquetes de computo geoestadísticos como GS+ (Gamma Design

Software, 1999) traen incorporados procedimientos iterativos como el de Gauss-Newton para llevar a cabo la estimación. Otros como GeoEAS (Englund y Sparks, 1988) sólo permiten el ajuste a sentimiento por el método de ensayo y error. Como se mencionó anteriormente, cuando la autocorrelación no es igual en todas las direcciones entonces se dice que hay anisotropía. Esa puede ser geométrica o zonal. La primera se presenta cuando los semivariogramas calculados en varias direcciones tienen igual meseta pero varían en el rango. En el segundo caso todos los semivariogramas direccionales tiene igual rango pero diferente meseta. Algunas transformaciones apropiadas para solucionar la anistropía y hacer válida la construcción de un semivariograma omnidireccional se pueden encontrar en ?, Samper and Carrera (1990) y Cressie (1993).

## 4.4. Estimación por mínimos cuadrados

Tradicionalmente, los geoestadísticos se han acercado a la estimación paramétrica de la estructura de covarianza espacial como un ejercicio de ajuste de curva. La razón de esto es que el variograma teórico define una familia paramétrica de funciones, por ejemplo  $V(u, \theta)$ , mientras que el variograma empírico define una relación empírica entre distancias correspondientes entre lugar  $u_{ij}$  y las coordenadas del variograma  $v_{ij}$ . Desde  $E[v_{ij}] = V(u_{ij}, \theta)$ , el problema de la fijación de un modelo de variograma se puede ver como, en esencia, un problema de regresión no lineal, con  $u_{ij}$  la variable explicativa y  $v_{ij}$  la respuesta. La debilidad en esta analogía es que  $n$  datos originales de los puntos  $(y_i, x_i)$  producen  $\frac{1}{2}n(n - 1)$  coordenadas del variograma empírico  $v_{ij}$ , creando fuerte dependencia estadística inherente entre los  $v_{ij}$ , lo cual afecta negativamente el desempeño de conocer las técnicas de modelos de regresión.

### 4.4.1. Estimación por Mínimos cuadrados

Tal vez la más simple, y sin duda la más conocida, es la técnica estadística de ajuste de curva normal de estimación de mínimos cuadrados. Esta técnica se puede aplicar a la estimación del variograma de la siguiente manera. Las coordenadas del variograma empíri-

co  $u_{ij}$  son un estimador insesgado de  $V(u_{ij})$ , para cualquiera que sean las propiedades de distribución del proceso  $Y(\cdot)$ . Por lo tanto, un estimador consistente de  $\theta$  en un modelo paramétrico,  $V(u) = V(u, \theta)$ , se puede obtener al minimizar la cantidad

$$MCO_1(\theta) = \sum_{j>i} v_{ij} - V(u_{ij}; \theta)^2 \quad (4.1)$$

o, es decir, la contraparte muestral del variograma

$$MCO_2(\theta) = \sum_k \bar{v}_{ij} - V(u_{ij}; \theta)^2 \quad (4.2)$$

El variograma muestral (4.2) es más comúnmente utilizado en la práctica, pero aparte de la comodidad computacional, los autores no pueden pensar en ninguna buena razón por las cuales este debe ser así, y desde un punto de vista estadístico no es una desventaja obvia: para una distribución irregular de los lugares de muestreo, el promedio de las ordenadas del variograma empírico dentro de los rangos discretos significa que  $v_k$  sólo es aproximadamente insesgado para  $V(u_k; \theta)$ .

#### 4.4.2. Mínimos cuadrados ponderados

Una objeción posible para (4.1) o (4.2) como criterios para reducir al mínimo es que no tienen en cuenta la varianza no constante de  $v_{ij}$ . Recordemos que, para un proceso gaussiano  $Y(x)$ , la distribución muestral de cada variograma empírico  $v_{ij}$  está dada por

$$v_{ij} \sim V(u_{ij}; \theta) \chi_1^2 \quad (4.3)$$

de ello se deduce que

$$E[v_{ij}] = V(u_{ij}; \theta)$$

y

$$Var(v_{ij}) = 2V(u_{ij}; \theta)^2$$

Tenga en cuenta también que en el variograma muestral, los  $\bar{v}_\kappa$  son los promedios de número desigual,  $n_\kappa$ , es decir, de las coordenadas del variograma empírico. Esta observación llevó a Cressie (1985) para proponer un criterio de mínimos cuadrados ponderados,

$$MCP_2(\theta) = \sum_n n_\kappa [\bar{v}_\kappa - V(u_\kappa; \theta)/V(u_\kappa; \theta)]^2 \quad (4.4)$$

La recomendación práctica habitual con (4.4) es incluir en el resumen sólo los  $\bar{v}_\kappa$  para los cuales  $n_\kappa > 30$  Journel and Huijbregts (1978b). Sin embargo, desde una perspectiva estadística clásica, la estimación ponderada de mínimos cuadrados como se define por (4.4) no es en absoluto una cosa natural a hacer, fundamentalmente porque los pesos son definidos en términos de los parámetros desconocidos. Por la misma razón, si se define un criterio análogo a (4.4), pero basado en el variograma empírico, es decir,

$$MCP_1(\theta) = \sum_{j>i} [v_{ij} - V(u_{ij}; \theta)/V(u_{ij}; \theta)]^2 \quad (4.5)$$

la estimación de  $\theta$  reduciendo al mínimo (4.5) resulta ser una muy mala idea. Barry and Diggle (1997) muestran que la reducción mínima de (4.4) o (4.5) corresponde a la solución de una ecuación en polarización negativa para  $\theta$ , y que el sesgo en la estimación de la ecuación correspondiente a (4.4) es del orden de  $n_\kappa^{-1}$ . Mínimos cuadrados ordinarios aplicados a las coordenadas del variograma empírico no sufren de esta dificultad teórica. Proporciona una forma sencilla, pero ineficaz estadísticamente, de estimar los parámetros del variograma. Una modificación de (4.4) sería el peso sólo en función del número de  $n_\kappa$ , dando un criterio

$$MCP_3(\theta) = \sum_{\kappa} n_\kappa \bar{v}_\kappa - V(u_\kappa; \theta)^2 \quad (4.6)$$

Esto es aproximadamente equivalente a los mínimos cuadrados ordinarios aplicados a las coordenadas del variograma empírico, la diferencia entre los dos procedimientos que son únicamente consecuencia del sesgo en el cálculo del suavizador  $\bar{v}_\kappa$ . Llamamos a este último método  $n$ -mínimos cuadrados ponderados, para distinguirlo de los mínimos cuadrados ponderados en el sentido de Cressie (1985).

#### 4.4.3. Cuasi-verosimilitud

Cuasi-verosimilitud proporciona una aproximación a la estimación paramétrica de los parámetros del variograma que está inspirada en el método de mínimos cuadrados ponderados, pero en base a una ecuación de estimación insesgada. El método fue desarrollado

originalmente en el contexto de modelos lineales generalizados, para explotar una vinculación funcional entre la media y la varianza de la variable respuesta, confiando únicamente en los supuestos del primero y segundo momento para la validez de las inferencias resultantes. Muy en general, supongamos que un vector aleatorio  $Y$  tiene un vector de medias  $\mu(\theta)$  y la matriz de varianza  $\sum(\theta)$ . El estimador de cuasi-verosimilitud de  $\theta$  es la solución a la ecuación de estimación de  $Q(\theta) = 0$ , donde

$$Q(\theta) = \mu'(\theta)^T [\sum(\theta)]^{-1} y - \mu(\theta) \quad (4.7)$$

Hay que tener en cuenta que en el verdadero valor de  $\theta$ ,  $E[Q(\theta)] = 0$ . Esto muestra que (4.7) es una ecuación de estimación objetiva, y la teoría General de ecuaciones de estimación se pueden aplicar para mostrar que la solución de (4.7) da un estimador consistente de  $\theta$ . Si la consistencia tiene o no  $\sum(\theta)$  se especifica correctamente. Las soluciones de (4.7) por mínimos cuadrados ponderados iterativo es el método estándar de la estimación de  $\hat{\theta}$  en un clásico modelo lineal generalizado McCullagh and Nelder (1989).

Debido a la consistencia del estimador de cuasi-verosímil y que no requiere una especificación correcta de  $\sum(\theta)$ , también podríamos considerar una estimación de la ecuación más simple,

$$Q(\theta) = \sum_{i=1}^n \mu'_i(\theta) W_i(\theta) y_i - \mu_i(\theta) \quad (4.8)$$

en el que  $W(\theta) = diag W_i(\theta)$  es una matriz de ponderación, también llamado a veces el inverso de la varianza de la matriz de trabajo, ya que los estimadores resultantes son eficientes si el peso de la matriz se elige de manera que  $W(\theta)^{-1} = Var(y)$ .

Para aplicar estas ideas a la estimación del variograma, se utilizan las coordenadas del variograma  $v_{ij}$  como las respuestas y  $W_{ij}(\theta) = v(u_{ij}, \theta)^{-2}$  como los pesos. Los estimadores resultantes son consistentes con la única condición de que la familia paramétrica  $V(u, \theta)$  se ha especificado correctamente, y son potencialmente más eficientes que los estimadores de mínimos cuadrados ordinarios, en la medida en que reconocen la varianza no constante de la  $v_{ij}$ . Sin embargo, el resultado empírico en Barry and Diggle (1997) sugieren que en la práctica el aumento de la eficacia a menudo puede ser leve.

## 4.5. Estimación asumiendo normalidad

### 4.5.1. Estimación por Máxima Verosimilitud

Desde una perspectiva estadística clásica, la función de verosimilitud es la base sobre la que los métodos de inferencia se construyen. La estimación de los parámetros del modelo mediante la maximización de la función de verosimilitud bajo un modelo asumido ofrece, en condiciones muy generales, que los estimadores son insesgados y eficientes cuando se aplican a muestras grandes. Por lo tanto, si estamos dispuestos a utilizar a los modelos paramétricos, es natural utilizar la función de verosimilitud como base para la estimación de parámetros.

#### Máxima verosimilitud bajo los supuestos Gaussianos

Se asume que los datos  $Z = (Z_1, Z_2, \dots, Z_n)$  son generados de un modelo Gaussiano lineal,

$$Y_i = \mu(x_i) + S(x_i) + Z_i : i = 1, \dots, n \quad (4.9)$$

En (4.9)  $\mu(x)$  es determinado por un modelo de regresión lineal,

$$\mu(x_i) = \sum_{k=1}^p f_k(x_i)\beta_k \quad (4.10)$$

donde los  $f_k(\cdot)$  se observan, espacialmente referencian variables explicativas. También  $S(\cdot)$  es un proceso estacionario Gaussiano con varianza  $\sigma^2$  y función de correlación  $\rho(u, \phi)$ . Por último, los  $Z_i$  son independientes entre sí,  $Z_i \sim N(0, \gamma^2)$ . Para derivar la función de verosimilitud asociada, es útil escribir el modelo en forma de matriz. En primer lugar,  $E(Y) = F\beta$ , donde  $F$  es una matriz  $n$  por  $p$  cuya  $k$ -ésima columna se compone de los valores  $f_k(x_1), \dots, f_k(x_n)$  y  $\beta = (\beta_1, \dots, \beta_p)$  es un vector de  $p$  elementos de los parámetros de regresión. Ahora,  $\theta = (\tau^2, \sigma^2, \phi)$  denota el conjunto de parámetros que definen la matriz de covarianza de los datos observados  $Y$ , y  $Var(Y) = G(\theta)$ . Teniendo en cuenta que  $G(\theta) = \tau^2 I + \sigma^2 R(\phi)$ , donde  $I$  denota la matriz identidad  $n$  por  $n$  y  $R(\phi)$  la matriz  $n$  por  $n$  con el elemento ij-ésimo  $r_{ij} = \rho(\|x_i - x_j\|, \phi)$ .

Siguiendo esta notación, el modelo definido por (5.9) y (5.10) implica que  $Y \sim MVN\{F\beta, G(\theta)\}$ . De ello se deduce que el logaritmo de la verosimilitud de  $(\beta, \theta)$ , hasta una constante aditiva, es

$$l(\beta, \theta) = -\frac{1}{2}\{\log |G(\theta)| + (y - F\beta)'G(\theta)^{-1}(y - F\beta)\} \quad (4.11)$$

para maximizar el logaritmo de la verosimilitud (4.11), primero se elimina  $\beta$  de la maximización numérica de  $l(\cdot)$ , señalando que, para un  $\theta$  dado, el estimador de máxima verosimilitud de  $\beta$  es

$$\hat{\beta}(\theta) = [F'G(\theta)^{-1}F]^{-1}F'G(\theta)^{-1}y$$

Sustituyendo a  $\hat{\beta}(\theta)$  en (5.11), obtenemos el logaritmo de la verosimilitud reducida para  $\theta$ ,

$$L(\theta) = l(\hat{\beta}(\theta), \theta)$$

En casos especiales, la estructura algebraica de  $G(\theta)$  y  $F$  puede permitir una mayor simplificación. En particular, siempre se puede extraer un factor escalar desde  $G(\theta)$ , por ejemplo, escribiendo

$$G(\theta) = \sigma^2\{\nu^2 + R(\phi)\}$$

donde  $\nu^2 = \tau^2/\phi$  es una especie de relación de ruido entre la señal. Esto nos permite eliminar  $\sigma^2$  algebraicamente la base del criterio de maximización numérica, que es una simplificación útil porque la fiabilidad de la maximización de la función tienden a ser más alta cuando la dimensionalidad de la función a maximizar es más bajo.

### Máxima verosimilitud restringida

Una variante de los métodos de máxima verosimilitud, llamada máxima verosimilitud restringida (REML) es ampliamente considerada como una mejora en la máxima verosimilitud, ya que conduce a una menor estimador sesgado en muestras pequeñas, aunque su justificación original fue en un contexto muy diferente, el de la estimación de componentes de varianza en el diseño experimento Patterson and Thompson (1971). El principio del REML es el siguiente:

- Bajo el modelo asumido  $E[Y] = F\beta$ , transformar los datos de forma lineal a  $Y^* = AY$  de tan manera que la distribución de  $Y^*$  no dependa de  $\beta$ .
- Estimar  $\theta$  por máxima verosimilitud aplicada a los datos transformados  $Y^*$ .

Note que siempre podemos encontrar una matriz  $A$  adecuada sin los verdaderos valores de  $\beta$  o  $\theta$ , por ejemplo

$$A = I - F(F'F)^{-1}F'$$

El estimador REML de  $\theta$  puede ser calculando maximizando

$$L^*(\theta) = -\frac{1}{2}\{\log |G(\theta)| + \log |F'\{G(\theta)\}^{-1}F| + (y - F\tilde{\beta})'\{G(\theta)\}^{-1}(y - F\tilde{\beta})\}$$

donde  $\tilde{\beta} = \hat{\beta}(\theta)$ . Tenga en cuenta, en particular, que  $L^*(\theta)$  depende de  $F$ , y por lo tanto de una especificación correcta del modelo de  $\mu(x)$ , pero no de la elección de  $A$ . Por lo tanto, dado el modelo de  $\mu(\cdot)$ , el métodos goza de la misma objetividad que la estimación máxima verosímil.

REML es recomendado ampliamente como un método de estimación de parámetros para el modelo geoestadístico. Sin embargo, por experiencia se ha visto que puede ser más sensible que la estimación de máxima probabilidad de los errores de especificación para  $\mu(x)$ . En su contexto original de diseño de experimentos, una especificación incorrecta del modelo no es a menudo un problema, y la ventaja específica de REML de máxima verosimilitud es su capacidad para ajustar automáticamente el efecto de la estimación de parámetros en el modelo antes de estimar los parámetros de varianza. Por ejemplo, el estimador insesgado de la varianza del error en un modelo de regresión clásica es el estimador REML. Sin embargo, en el ámbito espacial, donde no hay distinción tajante entre  $\mu(x)$  y  $S(x)$ , pueden estar dispuestos a adoptar un enfoque más pragmático, que describe la variabilidad en los datos de  $Y$  a través de los efectos combinados de los supuestos acerca de  $\mu(x)$  y de  $S(x)$  sin la confianza de que sea parte de que el modelo asumido sea correcto. Cuando el modelo de tal que  $\mu(x)$  es constante, la estimación REML tiene una relación estrecha con el método de predicción espacial conocida como kriging ordinario, que a su vez, tiene una interpretación bayesiana. Cuando los modelos no-constantes para  $\mu(x)$  son usados, los méritos relativos de REML y la estimación máxima verosímil son menos claras.

## 4.6. Predicción espacial

### 4.6.1. Predicción espacial optima

De la teoría de la decisión se conoce que si  $Z_0$  es una cantidad aleatoria y  $Z_0^*$  es su predictor, entonces  $L(Z_0 : Z_0^*)$  representa la pérdida en que se predice  $Z_0$  con  $Z_0^*$  y el mejor predictor será el que minimice  $E(L(Z_0 : Z_0^*)/Z)$  con  $Z = Z_1, Z_2, \dots, Z_n'$  es decir el predictor óptimo es el que minimice la *esperanza condicional* de la función de pérdida. Si  $L(Z_0 : Z_0^*) = [Z_0 - Z_0^*]^2 \Rightarrow Z_0^* = E(Z_0/Z)$ . La expresión anterior indica que para encontrar el predictor óptimo se requiere conocer la distribución conjunta de las  $n + 1$  variables aleatorias.

**Teorema:** Sea  $Z = (Z_1, \dots, Z_n)$  cualquier conjunto de  $n$  variables aleatorias cuyos valores  $y = (y_1, \dots, y_n)$  son observados. Sea  $T$  cualquier otra variable aleatoria y  $\hat{T} = t(Y)$  por cualquier función de  $Y$ . Entonces,  $E[(T - \hat{T})^2]$  toma el mínimo valor cuando  $\hat{T} = E(T|Y)$ .

**Prueba:** Escribiendo

$$E[(T - \hat{T})^2] = E_Y[E_T[(T - \hat{T})^2|Y]] \quad (4.12)$$

donde los subíndices de los dos operadores esperanzas indican que las esperanzas se pueden expresar respecto a  $Y$  y  $T$ . Escribir la esperanza interna en (4.9) como

$$E_T[(T - \hat{T})^2|Y] = Var_T\{(T - \hat{T})|Y\} + \{E_T[(T - \hat{T})|Y]\}^2$$

condicionada a  $Y$ , cualquier función de  $Y$  es una constante, por lo que  $Var_T\{(T - \hat{T})|Y\} = Var_T(T|Y)$  y

$$E_T[(T - \hat{T})^2|Y] = Var_T(T|Y) + \{E_T(T|Y) - \hat{T}\}^2 \quad (4.13)$$

Ahora, tomando las esperanzas de ambos lados de la expresión (4.10) con respecto a los  $Y$ . Esto da

$$E[(T - \hat{T})^2] = E_Y[Var_T(T|Y)] + E_Y[\{E_T(T|Y) - \hat{T}\}^2] \quad (4.14)$$

El primer término del lado derecho de (4.11) no depende de la elección de los  $\hat{T}$ , mientras que el segundo es no negativo, e igual a cero si y sólo si  $\hat{T} = E(T|Y)$ .

### 4.6.2. Definición de Kriging

La palabra kriging (expresión anglosajona) procede del nombre del geólogo sudafricano D. G. Krige, cuyos trabajos en la predicción de reservas de oro, realizados en la década del cincuenta, suelen considerarse como pioneros en los métodos de interpolación espacial. Kriging encierra un conjunto de métodos de predicción espacial que se fundamentan en la minimización del error cuadrático medio de predicción. En el cuadro 4.1 se mencionan los tipos de kriging y algunas de sus propiedades. En las secciones 4.4.3 y 4.4.4, se hace una presentación detallada de ellos.

Cuadro 4.1: Tipos de predictores kriging y sus propiedades

TIPO DE PREDICTOR	NOMBRE	PROPIEDADES
LINEAL	Simple	Son óptimos si hay normalidad
	Ordinario	multivariada.
	Universal	Independiente de la distribución son los mejores predictores linealmente insesgados.
NO LINEAL	Indicador	Son predictores óptimos.
	Probabilístico	
	Log Normal	
	Trans-Gaussiano	
	Disyuntivo	

Los métodos kriging se aplican con frecuencia con el propósito de predicción, sin embargo estas metodologías tienen diversas aplicaciones, dentro de las cuales se destacan la simulación y el diseño de redes óptimas de muestreo.

### 4.6.3. Kriging Ordinario

Suponga que se hacen mediciones de la variable de interés  $Z$  en los puntos  $x_i, i = 1, 2, \dots, n$ , de la región de estudio, es decir se tienen realizaciones de las variables  $Z(x_1), \dots, Z(x_n)$ , y se desea predecir  $Z(x_o)$ , en el punto  $x_o$  donde no hubo medición. En esta circunstancia, el método kriging ordinario propone que el valor de la variable puede predecirse como una combinación lineal de

las  $n$  variables aleatorias así:

$$\begin{aligned} Z^*(x_0) &= \lambda_1 Z(x_1) + \lambda_2 Z(x_2) + \lambda_3 Z(x_3) + \dots + \lambda_n Z(x_n) \\ &= \sum_{i=1}^n \lambda_i Z(x_i) \end{aligned}$$

en donde los  $\lambda_i$  representan los pesos o ponderaciones de los valores originales. Dichos pesos se calculan en función de la distancia entre los puntos muestreados y el punto donde se va a hacer la correspondiente predicción. La suma de los pesos debe ser igual a uno para que la esperanza del predictor sea igual a la esperanza de la variable. Esto último se conoce como el requisito de insesgamiento. Estadísticamente la propiedad de insesgamiento se expresa a través de:

$$E(Z^*(x_0)) = E(Z(x_0))$$

Asumiendo que el proceso es estacionario de media  $m$  (desconocida) y utilizando las propiedades del valor esperado, se demuestra que la suma de las ponderaciones debe ser igual a uno:

$$\begin{aligned} E\left(\sum_{i=1}^n \lambda_i Z(x_i)\right) &= m \\ \sum_{i=1}^n \lambda_i E(Z(x_i)) &= m \\ \sum_{i=1}^n \lambda_i m &= m \\ m \sum_{i=1}^n \lambda_i &= m \Rightarrow \sum_{i=1}^n \lambda_i = 1 \end{aligned}$$

Se dice que  $Z^*(x_0)$  es el mejor predictor, lineal en este caso, porque los pesos se obtienen de tal manera que minimicen la varianza del error de predicción, es decir que minimicen la expresión:

$$V(Z^*(x_0) - Z(x_0))$$

Esta última es la característica distintiva de los métodos kriging, ya que existen otros métodos de interpolación como el de distancias inversas o el poligonal, que no garantizan varianza mínima de predicción (Samper and Carrera (1990)). La estimación de los pesos se obtiene minimizando  $V[Z^*(x_0) - Z(x_0)]$  sujeto a  $\sum_{i=1}^n \lambda_i = 1$ .

Se tiene que  $V[Z^*(x_0) - Z(x_0)] = V[Z^*(x_0)] - 2\text{cov}[Z^*(x_0), Z(x_0)] + V[Z(x_0)]$

Desagregando las componentes de la ecuación anterior se obtiene los siguiente:

$$\begin{aligned} V[Z^*(x_0)] &= V\left[\sum_{i=1}^n \lambda_i Z(x_i)\right] \\ &= \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \text{COV}[Z(x_i), Z(x_j)] \end{aligned}$$

En adelante se usará la siguiente notación:  $\text{COV}[Z(x_i), Z(x_j)] = C_{ij}$  y  $V[Z(x_0)] = \sigma^2$ .

De lo anterior

$$\begin{aligned} \text{COV}[Z^*(x_0), Z(x_0)] &= \text{COV}\left[\sum_{i=1}^n \lambda_i Z(x_i), Z(x_0)\right] \\ &= \sum_{i=1}^n \text{COV}[Z(x_i), Z(x_0)] \\ &= \sum_{i=1}^n \lambda_i C_{i0} \end{aligned}$$

Entonces reemplazando, se tiene que:

$$V[Z^*(x_0) - Z(x_0)] = \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j C_{ij} - 2 \sum_{i=1}^n \lambda_i C_{i0} + \sigma^2 \quad (4.15)$$

Luego se debe minimizar la función anterior sujeta a la restricción  $\sum_{i=1}^n \lambda_i = 1$ . Este problema de minimización con restricciones se resuelve mediante el método de multiplicadores de Lagrange.

$$\sigma_k^2 = \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j C_{ij} - 2 \sum_{i=1}^n \lambda_i C_{i0} + \sigma^2 + \underbrace{2\mu}_{\text{Multiplicador de Lagrange}} \underbrace{\left( \sum_{i=1}^n \lambda_i - 1 \right)}_0$$

Siguiendo el procedimiento acostumbrado para obtener valores extremos de una función, se deriva

e iguala a cero, en este caso con respecto a  $\lambda_i$  y  $\mu$ .

$$\begin{aligned}
 \frac{\partial(\sigma_k^2)}{\partial\lambda_1} &= \frac{\partial \left[ (\lambda_1^2 C_{11} + 2\lambda_1 \sum_{j=2}^n \lambda_j C_{1j} + \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j C_{ij}) \right]}{\partial\lambda_1} \\
 &- \frac{2 \sum_{i=1}^n \lambda_i C_{i0} + \sigma^2 + 2\mu (\sum_{i=1}^n \lambda_i - 1)}{\partial\lambda_1} \\
 &= \left( 2\lambda_1 C_{11} + 2 \sum_{j=1}^n \lambda_j C_{1j} \right) - 2C_{10} + 2\mu \\
 &= 2 \sum_{j=1}^n \lambda_j C_{1j} - 2C_{10} + 2\mu = 0 \\
 \Rightarrow \quad &\sum_{j=1}^n \lambda_j C_{1j} + \mu = C_{10}
 \end{aligned} \tag{4.16}$$

De manera análoga se determinan las derivadas con respecto a  $\lambda_2, \dots, \lambda_n$ :

$$\begin{aligned}
 \frac{\partial(\sigma_k^2)}{\partial\lambda_2} &= 2 \sum_{j=1}^n \lambda_j C_{2j} - 2C_{20} + 2\mu = 0 \\
 \Rightarrow \quad &\sum_{j=1}^n \lambda_j C_{2j} + \mu = C_{20}
 \end{aligned} \tag{4.17}$$

$$\begin{aligned}
 &\vdots \\
 \frac{\partial(\sigma_k^2)}{\partial\lambda_n} &= 2 \sum_{j=1}^n \lambda_j C_{nj} - 2C_{n0} + 2\mu = 0 \\
 \Rightarrow \quad &\sum_{j=1}^n \lambda_j C_{nj} + \mu = C_{n0}
 \end{aligned} \tag{4.18}$$

por último derivamos con respecto a  $\mu$ :

$$\begin{aligned}
 \frac{\partial(\sigma_k^2)}{\partial\mu} &= 2 \sum_{i=1}^n \lambda_i - 2 = 0 \\
 \Rightarrow \quad &\sum_{i=1}^n \lambda_i = 1
 \end{aligned} \tag{4.19}$$

De (4.2), (4.3), (4.4), (4.5) resulta un sistema de  $(n + 1)$  ecuaciones con  $(n + 1)$  incógnitas,

que matricialmente puede ser escrito como:

$$\begin{pmatrix} C_{11} & \cdots & C_{1n} & 1 \\ \vdots & \ddots & \vdots & \vdots \\ C_{n1} & \cdots & C_{nn} & 1 \\ 1 & \cdots & 1 & 0 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \\ \mu \end{pmatrix} = \begin{pmatrix} C_{10} \\ \vdots \\ C_{n0} \\ 1 \end{pmatrix}$$

$$C_{ij} \cdot \lambda = C_{i0}$$

por lo cual los pesos que minimizan el error de predicción se determinan mediante la función de covariograma a través de:

$$\lambda = C_{ij}^{-1} \cdot C_{i0}$$

Encontrando los pesos se calcula la predicción en el punto  $x_o$ . De forma análoga se procede para cada punto donde se quiera hacer predicción.

### Varianza de Predicción del Kriging Ordinario

Multiplicando (4.2), (4.3) y (4.4) por  $\lambda_i$  se obtiene:

$$\lambda_i \left( \sum_{j=1}^n \lambda_j C_{ij} + \mu \right) = \lambda_i C_{i0} \quad \forall i, i = 1, 2, \dots, n$$

Sumando las  $n$  ecuaciones

$$\sum_{i=1}^n \lambda_i \sum_{j=1}^n \lambda_j C_{ij} + \sum_{i=1}^n \lambda_i \mu = \sum_{i=1}^n \lambda_i C_{i0} \sum_{i=1}^n \sum_{j=1}^n \lambda_j C_{ij} = \sum_{i=1}^n \lambda_i C_{i0} - \sum_{i=1}^n \lambda_i \mu$$

Sustituyendo la expresión anterior en (4.1)

$$\begin{aligned} \sigma_k^2 &= \sigma^2 + \sum_{i=1}^n \lambda_i C_{i0} - \sum_{i=1}^n \lambda_i \mu - 2 \sum_{i=1}^n \lambda_i C_{i0} \\ \sigma_k^2 &= \sigma^2 + \sum_{i=1}^n \lambda_i C_{i0} - \mu \end{aligned} \tag{4.20}$$

### Estimación de Ponderaciones por medio de la Función de Semivarianza

Los pesos  $\lambda$  pueden ser estimados a través de la función de semivarianza, para lo cual se requiere conocer la relación entre las funciones de covariograma y de semivarianza. Antes de esto

conveniente tener en cuenta la siguiente notación:  $\sigma^2 = V(Z(x))$ ,  $\gamma(h)$ , donde  $h$  es la distancia entre los puntos  $i$  y  $j$  análogamente  $C_{ij} = C(h)$ . La relación entre las dos funciones en cuestión es la siguiente:

$$\begin{aligned}
 \gamma_{ij} &= \frac{1}{2}E[(Z(x_j) - Z(x_i))^2] \\
 &= \frac{1}{2}E[(Z(x_j))^2 - 2(Z(x_j)Z(x_i)) + (Z(x_i))^2] \\
 &= \frac{1}{2}E[(Z(x_j))^2] - E[Z(x_i)Z(x_i)] + \frac{1}{2}E[(Z(x_i))^2] \\
 &= \frac{1}{2}[E(Z(x_j))^2 - k^2] + \frac{1}{2}[E(Z(x_j))^2 - k^2] - [E(Z(x_i)Z(x_i)) - k^2] \\
 &= \frac{1}{2}[V(Z(x))] + \frac{1}{2}[V(Z(x))] - COV[Z(x_i)Z(x_i)] \\
 &= V[Z(x)] - COV[Z(x_j)Z(x_i)] \\
 &= \sigma^2 - C_{ij} \Rightarrow C_{ij} = \sigma^2 - \gamma_{ij}
 \end{aligned} \tag{4.21}$$

Reemplazando (4.7) en (4.2), (4.3) y (4.4) se determinan los pesos óptimos en términos de la función de semivarianza:

$$\begin{aligned}
 \frac{\partial \sigma_k^2}{\partial \lambda_1} &= \sum_{i=1}^n \lambda_j C_{1j} + \mu - C_{10} \\
 &= \sum_{i=1}^n \lambda_j (\sigma^2 - \gamma_{1j}) + \mu - (\sigma^2 - \gamma_{10}) \\
 &= \sigma^2 \sum_{i=1}^n \lambda_j - \sum_{j=1}^n \lambda_j \gamma_{1j} + \mu - \sigma^2 + \gamma_{10} \\
 &= \sigma^2 - \sum_{i=1}^n \lambda_j \gamma_{1j} + \mu - \sigma^2 + \gamma_{10} \\
 \Rightarrow \quad &\sum_{i=1}^n \lambda_j \gamma_{1j} - \mu = \gamma_{10}
 \end{aligned}$$

Similarmente,

$$\begin{aligned}\frac{\partial \sigma_k^2}{\partial \lambda_2} &= \sum_{i=1}^n \lambda_i \gamma_{2j} - \mu = \gamma_{20} \\ &\vdots \\ \frac{\partial \sigma_k^2}{\partial \lambda_2} &= \sum_{i=1}^n \lambda_i \gamma_{nj} - \mu = \gamma_{n0}\end{aligned}$$

El sistema de ecuaciones se completa con (4.5). De acuerdo con lo anterior los pesos se obtienen en términos del semivariograma a través del sistema de ecuaciones:

$$\begin{pmatrix} \gamma_{11} & \cdots & \gamma_{1n} & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \gamma_{n1} & \cdots & \gamma_{nn} & 1 \\ 1 & \cdots & 1 & 0 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \\ -\mu \end{pmatrix} = \begin{pmatrix} \gamma_{10} \\ \vdots \\ \gamma_{n0} \\ 1 \end{pmatrix}$$

Para establecer la expresión de la correspondiente varianza del error de predicción en términos de la función de semivarianza se reemplaza (4.7) en (4.6), de donde:

$$\begin{aligned}\sigma_k^2 &= \sigma^2 - \left[ \sum_{i=1}^n \lambda_i (\sigma^2 - \gamma_{ij}) \right] + \mu \\ \sigma_k^2 &= \sigma^2 - \sigma^2 \sum_{i=1}^n \lambda_i + \sum_{i=1}^n \lambda_i \gamma_{ij} + \mu \\ \sigma_k^2 &= \sum_{i=1}^n \lambda_i \gamma_{i0} + \mu\end{aligned}$$

Los pesos de kriging ordinario también pueden ser estimados mediante el uso del correlograma aplicando la siguiente relación:  $\rho_{ij} = \frac{C_{ij}}{\sigma^2}$ , caso en el que la correspondiente varianza de predicción estaría dada por Isaaks and Srivastava (1989):

$$\sigma_k^2 = \sigma^2 \left( 1 - \sum_{i=1}^n \lambda_i \gamma_{i0} + \mu \right)$$

### Validación del kriging

Existen diferentes métodos para evaluar la bondad de ajuste del modelo de semivariograma elegido con respecto a los datos muestrales y por ende de las predicciones hechas con kriging.

El más empleado es el de validación cruzada, que consiste en excluir la observación de uno de los  $n$  puntos muestrales y con los  $n-1$  valores restantes y el modelo de semivariograma escogido, predecir vía kriging el valor de la variable en estudio en la ubicación del punto que se excluyó. Se piensa que si el modelo de semivarianza elegido describe bien la estructura de autocorrelación espacial, entonces la diferencia entre el valor observado y el valor predicho debe ser pequeña. Este procedimiento se realiza en forma secuencial con cada uno de los puntos muestrales y así se obtiene un conjunto de  $n$  “errores de predicción”. Lo usual es calcular medidas que involucren a estos errores de predicción para diferentes modelos de semivarianza y seleccionar aquel que optimice algún criterio como por ejemplo el del mínimo error cuadrático medio (MECM). Este procedimiento es similar a la conocida técnica de remuestreo Jackknife ([Efron, 1982](#)) empleada en diversos contextos estadísticos para calcular varianzas de estimación, entre otros aspectos. Una forma descriptiva de hacer la validación cruzada es mediante un gráfico de dispersión de los valores observados contra los valores predichos. En la medida en que la nube de puntos se ajuste más a una línea recta que pase por el origen, mejor será el modelo de semivariograma utilizado para realizar el kriging.

## Representación de las predicciones

Una vez se ha hecho la predicción en un conjunto de puntos diferentes de los muestrales vía kriging, se debe elaborar un mapa que dé una representación global del comportamiento de la variable de interés en la zona estudiada. Los más empleados son los mapas de contornos, los mapas de residuos y los gráficos tridimensionales. En el caso de los mapas de contornos, en primer lugar se divide el área de estudio en un enmallado y se hace la predicción en cada uno de los nodos de éste mismo. Posteriormente se unen los valores predichos con igual valor, generando así las líneas de contorno (isolíneas de distribución). Este gráfico permite identificar la magnitud de la variable en toda el área de estudio. Es conveniente acompañar el mapa de interpolaciones de la variable con los correspondientes mapas de isolíneas de los errores y de las varianzas de predicción (posiblemente estimados a través de métodos matemáticos), con el propósito de identificar zonas de mayor incertidumbre respecto a las predicciones.

## Intervalos de Confianza

Asumiendo que los errores de predicción siguen una distribución normal estándar y que son independientes, un intervalo de confianza del  $100(1 - \alpha)\%$ ,  $0 < \alpha < 1$ , para  $Z(x)$  es:  $[Z^*(x) - Z_{1-\alpha/2}\sigma_k, Z^*(x) + Z_{1-\alpha/2}\sigma_k]$  con  $Z^*(x)$  el valor calculado de la predicción y  $Z_{1-\alpha/2}$  el percentil de una normal estándar.

### 4.6.4. Otros Métodos Kriging

A continuación se mencionan algunos aspectos generales de otros métodos de predicción espacial. Un estudio riguroso de ellos puede hacerse en Cressie (1993), Deutsch (1992) y Samper and Carrera (1990).

#### Kriging Simple

Suponga que hay una variable regionalizada estacionaria con media ( $m$ ) y covarianza conocidas. De manera análoga a como se define en modelos lineales (por ejemplo en diseño de experimentos) el modelo establecido en este caso es igual a la media más un error aleatorio con media cero. La diferencia es que en este caso los errores no son independientes. Sea  $Z(x)$  la variable de interés medida en el sitio  $x$ .

$$\begin{aligned} E[Z(x)] &= m \\ Z(x) &= m + \varepsilon(x), \quad \text{con} \quad E[\varepsilon(x)] = 0 \end{aligned}$$

El predictor de la variable de interés en un sitio  $x_0$  donde no se tiene información se define como:

$$Z^*(x_0) = m = \varepsilon^*(x_0)$$

con  $\varepsilon^*(x_0)$  que corresponde a la predicción del error aleatorio en el sitio  $x_0$ . Despejando de la ecuación anterior

$$\varepsilon^*(x_0) = Z^*(x_0) - m$$

El predictor del error aleatorio se define por:

$$\varepsilon^*(x_0) = \sum_{i=1}^n \lambda_i \varepsilon(x_i) = \sum_{i=1}^n \lambda_i (Z(x_i) - m)$$

de donde el predictor de la variable de estudio es:

$$Z^*(x_0) = m + \left[ \sum_{i=1}^n \lambda_i (Z(x_i) - m) \right] = m + \sum_{i=1}^n \lambda_i \varepsilon(x_i)$$

El predictor es insesgado si:

$E(Z^*(x_0)) = E(Z(x_0)) = m$ . Luego el predictor será insesgado cuando  $E(\varepsilon^*(x_0)) = 0$ .  $E(\varepsilon^*(x_0)) = \sum_{\lambda_i} \varepsilon(x_i) = \sum_{i=1}^n \lambda_i \varepsilon(x_i) = 0$ . Por consiguiente, a diferencia del kriging ordinario, en este caso no existen restricciones para las ponderaciones tendientes al cumplimiento de la condición de insesgamiento. La estimación de los pesos del método kriging ordinario se obtiene de tal forma que se minimice  $V(\varepsilon^*(x_0) - \varepsilon(x_0))$ .

$$\begin{aligned} V(\varepsilon^*(x_0) - \varepsilon(x_0)) &= E(\varepsilon^*(x_0) - \varepsilon(x_0))^2 \\ &= E \left( \left( \sum_{i=1}^n \lambda_i \varepsilon(x_i) \right) - \varepsilon(x_0) \right)^2 \\ &= E \left[ \left( \sum_{i=1}^n \lambda_i \varepsilon(x_i) \right) \left( \sum_{j=1}^n \lambda_j \varepsilon(x_j) \right) - 2E \left( \left( \sum_{i=1}^n \lambda_i \varepsilon(x_i) \right) (\varepsilon(x_0)) \right) + E(\varepsilon(x_0))^2 \right] \\ &= \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j E(\varepsilon(x_i) \varepsilon(x_j)) - 2 \sum_{i=1}^n \lambda_i E(\varepsilon(x_i) \varepsilon(x_0)) + E(\varepsilon(x_0))^2 \end{aligned}$$

Usando:

I.  $E[\varepsilon(x_0)] = 0$

II.  $E(\varepsilon(x_i) \varepsilon(x_j)) = COV(\varepsilon(x_i) \varepsilon(x_j)) = C_{ij}$

III.  $E[\varepsilon(x_0)]^2 = \sigma^2$

$$V(\varepsilon^*(x_0) - \varepsilon(x_0)) = \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j C_{ij} - 2 \sum_{i=1}^n \lambda_i C_{i0} + \sigma^2 \quad (4.22)$$

derivando respecto a  $\lambda_1$  se tiene:

$$\begin{aligned}
\frac{\partial V(\varepsilon^*(x_0) - \varepsilon(x_0))}{\partial \lambda_1} &= \frac{\partial}{\partial \lambda_1} \left( \lambda_1^2 C_{11} + 2\lambda_1 \sum_{j=2}^n \lambda_j C_{1j} + \sum_{i=2}^n \sum_{j=2}^n \lambda_i \lambda_j C_{ij} \right. \\
&\quad \left. - 2\lambda_1 C_{1j} - 2\lambda_1 C_{10} - 2 \sum_{i=2}^n \lambda_i C_{i0} + \sigma^2 \right) \\
&= 2\lambda_1 C_{11} + \sum_{i=2}^n \lambda_i C_{1j} - 2C_{10} \\
&= 2 \sum_{i=1}^n \lambda_i C_{1j} - 2C_{10}
\end{aligned}$$

igualando a cero

$$\sum_{i=1}^n \lambda_i C_{1j} = C_{10}$$

. En general para cualquier  $i, i = 1, 2, \dots, n$ , se obtiene:

$$\frac{\partial}{\partial \lambda_1} = \sum_{i=1}^n \lambda_i C_{ij} - 2C_{i0} \quad (4.23)$$

Con las  $n$  ecuaciones resultantes se construye el siguiente sistema de ecuaciones:

$$\begin{pmatrix} C_{11} & C_{12} & \cdots & C_{1n} \\ C_{21} & C_{22} & \cdots & C_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ C_{n1} & C_{n2} & \cdots & C_{nn} \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix} = \begin{pmatrix} C_{10} \\ C_{20} \\ \vdots \\ C_{n0} \end{pmatrix}$$

### Varianza de Predicción Kriging Simple

Se tiene de (4.8) que:

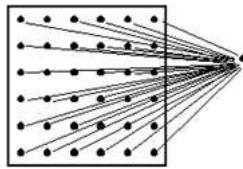
$$\begin{aligned}
V(\varepsilon^*(x_0) - \varepsilon(x_0)) &= \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j C_{ij} - 2 \sum_{i=1}^n \lambda_i C_{i0} + \sigma^2 \\
\sigma_k^2 &= \sum_{i=1}^n \lambda_i \sum_{j=1}^n C_{ij} - 2 \sum_{i=1}^n \lambda_i C_{i0} + \sigma^2 \\
&= 2 \sum_{i=1}^n \lambda_i C_{1j} - 2C_{10}
\end{aligned}$$

reemplazando (4.9) en (4.8)

$$\begin{aligned}\sigma_k^2 &= \sum_{i=1}^n \lambda_i C_{i0} - 2 \sum_{i=1}^n \lambda_i C_{i0} + \sigma^2 \\ \sigma_k^2 &= \sigma^2 - \sum_{i=1}^n \lambda_i C_{i0}\end{aligned}$$

#### 4.6.5. Kriging en Bloques.

En los dos métodos kriging hasta ahora descritos el objetivo ha estado centrado en la predicción puntual. A menudo, sin embargo, se requiere estimar un bloque, o más precisamente, estimar el valor promedio de la variable dentro de un área local. El valor promedio dentro del bloque es



estimado por:

$$\bar{Z}(A) = \sum_{i=1}^n \lambda_i Z(x_i)$$

Del sistema de ecuaciones para el kriging ordinario se tiene:

$$\begin{pmatrix} C_{11} & \cdots & C_{1n} & 1 \\ \vdots & \ddots & \vdots & \vdots \\ C_{n1} & \cdots & C_{nn} & 1 \\ 1 & \cdots & 1 & 0 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \\ \mu \end{pmatrix} = \begin{pmatrix} C_{10} \\ \vdots \\ C_{n0} \\ 1 \end{pmatrix}$$

Consecuentemente el vector del lado derecho de la igualdad en el sistema de arriba debe modificarse para incluir las covarianzas respecto al bloque. La covarianza de un punto al bloque corresponde a la covarianza promedio entre el punto muestreado  $i$  y todos los puntos dentro del bloque (en la práctica un enmallado regular de puntos dentro del bloque es usado como se muestra en la figura de la página anterior). El sistema de ecuaciones del kriging en bloques está dado por:

$$\begin{pmatrix} C_{11} & \cdots & C_{1n} & 1 \\ \vdots & \ddots & \vdots & \vdots \\ C_{n1} & \cdots & C_{nn} & 1 \\ 1 & \cdots & 1 & 0 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \\ \mu \end{pmatrix} = \begin{pmatrix} \bar{C}_{1A} \\ \vdots \\ \bar{C}_{nA} \\ 1 \end{pmatrix}$$

donde el vector de covarianzas al lado derecho de la igualdad en el sistema anterior es contiene las covarianzas entre las variables  $Z(x_1), Z(x_2), \dots, Z(x_n)$  y el bloque A donde se quiere hacer la estimación.

$$\bar{C}_{iA} = \frac{1}{|A|} \sum j / j \in AC_{iA}$$

La varianza del error de predicción del kriging en bloques está dada por:

$$\sigma_{kB}^2 = \bar{C}_{AA} - \left( \sum_{i=1}^n \lambda_i \bar{C}_{iA} + \mu \right) \quad \text{con} \quad \bar{C}_{AA} = \frac{1}{|A|} \sum j / j \in AC_{iA}$$

, igual a la covarianza entre pares de puntos dentro del bloque. Isaaks and Srivastava (1989) muestran a través de ejemplos que el kriging en bloques coincide con el promedio de predicciones hechas por kriging ordinario sobre cada uno de los puntos del enmallado dentro del bloque. Así mismo indican que en la práctica es suficiente con un enmallado cuadrado (6x6) para obtener estimaciones estables en los bloques.

#### 4.6.6. Kriging Universal

En los supuestos hechos hasta ahora respecto a los métodos kriging se ha asumido que la variable regionalizada es estacionaria (al menos se cumple con la hipótesis intrínseca). En muchos casos, la variable no satisface estas condiciones y se caracteriza por exhibir una tendencia. Por ejemplo en hidrología los niveles piezométricos de una acuífero pueden mostrar una pendiente global en la dirección del flujo Samper and Carrera (1990). Para tratar este tipo de variables es frecuente descomponer la variable  $Z(x)$  como la suma de la tendencia, tratada como una función determinística, más una componente estocástica estacionaria de media cero. Asuma que:

$$Z(x) = m(x) + \varepsilon(x)$$

con  $E(\varepsilon(x)) = 0, V(\varepsilon(x)) = \sigma^2$  y por consiguiente  $E(Z(x)) = m(x)$  La tendencia puede expresarse mediante:

$$m(x) = \sum_{l=1}^P a_l f_l(x)$$

donde las funciones  $f_l(x)$  son conocidas y  $P$  es el número de términos empleados para ajustar  $m(x)$ .

El predictor kriging universal se define como:

$$Z^*(x_0) = \sum_{i=1}^n \lambda_i Z(x_i)$$

este será insesgado si:

$$\begin{aligned} E(Z^*(x_0)) &= m(x_0) \\ E\left(\sum_{i=1}^n \lambda_i Z(x_i)\right) &= m(x_0) \\ \left(\sum_{i=1}^n \lambda_i m(x_i)\right) &= m(x_0) \\ \sum_{i=1}^n \lambda_i \left(\sum_{l=P}^n a_l f_l(x_i)\right) &= \sum_{l=1}^P a_l f_l(x_0) \\ \sum_{l=1}^P a_l \left(\sum_{i=1}^n \lambda_i f_l(x_i)\right) &= \sum_{l=1}^n a_l f_l(x_0) \\ \Rightarrow \sum_{i=1}^n \lambda_i f_l(x_0) &= \sum_{l=1}^P f_l(x_0) \end{aligned}$$

La obtención de los pesos en el kriging universal, análogo a los otros métodos kriging, se hace de tal forma que la varianza del error de predicción sea mínima.

$$\begin{aligned} V(Z^*(x_0) - Z(x_0)) &= E(Z^*(x_0) - Z(x_0))^2 \\ &= E\left(\left(\sum_{i=1}^n \lambda_i(m(x_i) - \varepsilon(x_i))\right) - (m(x_0) - \varepsilon(x_0))\right)^2 \\ &= E\left[\left(\sum_{i=1}^n \lambda_i(m(x_i) - m(x_i))\right) + \left(\sum_{i=1}^n \lambda_i(\varepsilon(x_i) - \varepsilon(x_0))\right)\right]^2 \\ &= E\left[\left(\sum_{i=1}^n \lambda_i(\varepsilon(x_i) - \varepsilon(x_0))\right)^2\right] \\ &= \sum_{i=1}^n \sum_{j=1}^n E(\varepsilon(x_i)\varepsilon(x_j)) - 2 \sum_{i=1}^n E(\varepsilon(x_i)\varepsilon(x_0)) + E(\varepsilon(x_0))^2 \end{aligned}$$

Usando

$$C_{ij} = COV(\varepsilon(x_i), \varepsilon(x_j))\sigma^2 = E(\varepsilon(x_0))^2$$

se tiene

$$V(Z^*(x_0) - Z(x_0)) = \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j C_{ij} - 2 \sum_{i=1}^n \lambda_i C_{i0} + \sigma^2$$

Luego incluyendo la restricción dada por la condición de insesgamiento, se debe minimizar:

$$\sigma_{ku}^2 = \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j C_{ij} - 2 \sum_{i=1}^n \lambda_i C_{i0} + \sigma^+ \sum_{l=1}^P \mu_l \left[ \sum_{i=1}^n \lambda_i f_l(x_i) - f_l(x_0) \right]$$

o en términos de la función de semivarianza

$$\sigma_{ku}^2 = \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \gamma_{ij} - 2 \sum_{i=1}^n \lambda_i \gamma_{i0} + \sigma^+ \sum_{l=1}^P \mu_l \left[ \sum_{i=1}^n \lambda_i f_l(x_i) - f_l(x_0) \right]$$

derivando la expresión anterior respecto a  $\lambda_1, \lambda_2, \dots, \lambda_n, \mu_1, \mu_2, \dots, \mu_P$ , e igualando a cero las correspondientes derivadas se obtienen las siguientes ecuaciones:

$$\sum_{j=i}^n \lambda_j \gamma_{ij} + \sum_{l=1}^P \mu_l f_l(x_i) = \gamma_{i0}, \quad i = 1, 2, \dots, n \quad \sum_{j=i}^n \lambda_j f_l(x_i) = f_l(x_0), \quad j = 1, 2, \dots, P$$

en términos matriciales

$$\begin{pmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1n} & f_{11} & \cdots & f_{p1} \\ \gamma_{21} & \gamma_{22} & \cdots & \gamma_{2n} & f_{12} & \cdots & f_{p2} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \gamma_{n1} & \gamma_{n2} & \cdots & \gamma_{nn} & f_{1n} & \cdots & f_{pn} \\ f_{11} & f_{12} & \cdots & f_{1n} & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ f_{p1} & f_{p2} & \cdots & f_{pn} & 0 & \cdots & 0 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \\ \mu_1 \\ \vdots \\ \mu_n \end{pmatrix} = \begin{pmatrix} \gamma_{10} \\ \gamma_{20} \\ \vdots \\ \gamma_{n0} \\ f_{10} \\ \vdots \\ f_{p0} \end{pmatrix}$$

donde  $f_{ij} = f_l(x_j)$  es la  $l$ -ésima función en el punto  $j$ -ésimo. La varianza de predicción del kriging universal está dada por Samper and Carrera (1990):

$$\sigma_{ku}^2 = \sum_{i=1}^n \lambda_i \gamma_{i0} + \sum_{l=1}^P \mu_l f_l(x_0)$$

Nótese que si  $p = l$  y  $f_l(x) = 1$ , el sistema de ecuaciones del kriging universal y la varianza de predicción coinciden con las del kriging ordinario. En este orden de ideas puede decirse que el kriging ordinario es un caso particular del kriging universal.

#### 4.6.7. Kriging Residual

La técnica kriging residuales empleada bajo las mismas circunstancias del kriging universal, es decir en aquellos casos en que la variable regionalizada no es estacionaria debido a la presencia de tendencia espacial en el valor promedio de la variable. La hipótesis central del kriging residual consiste en suponer conocida la tendencia  $m(x)$ . A partir de ella se calculan los residuos con base en los cuales se aplica kriging ordinario. La estimación de la tendencia es generalmente llevada a cabo por medio de mínimos cuadrados. La predicción en un sitio no muestreado es igual a la tendencia estimada más la predicción del error, es decir:

$$Z^*(x_0) = \hat{m}(x_0) = e^*(x_0)$$

$$e^*(x_0) = \sum_{i=1}^n \lambda_i e(x_i)$$

los pesos o ponderaciones son estimados por kriging ordinario como se muestra en la sección anterior. La varianza de predicción de la variable de interés coincide con la varianza de predicción de los errores. En la figura 4.7 se muestra un esquema con el procedimiento kriging residual en el caso de una tendencia lineal.

#### 4.6.8. Kriging Indicador

Suponga que se tiene una variable regionalizada  $\{Z(x) : x \in D \subset R^d\}$  estacionaria. Se define la siguiente transformación:

$$I(x_i, z_l) = \begin{cases} 1 & \text{Si } Z(x_i) \leq z_l \\ 0 & \text{Otro caso} \end{cases}$$

Algunas propiedades son las siguientes:

I.)  $Pr(I(x, z) = 1) = Pr(Z(x) \leq z_l) = F(z_l)$

II.)

$$\begin{aligned} E(I(x, z_l)) &= 1Pr(I(x, z_l) = 1) + 0Pr(I(x, z_l) = 0) \\ &= 1PR(I(x, z_l) = 1) = F(z_l) \end{aligned}$$

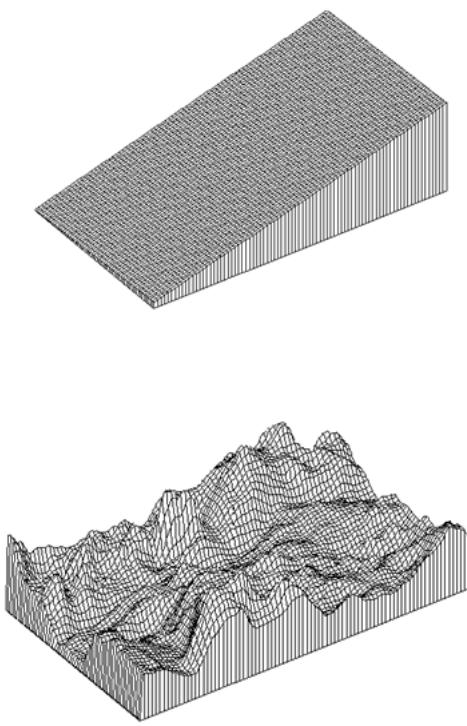


Figura 4.7: Representación del procedimiento kriging residual. La superficie interpolada (arriba) es igual a la suma de la tendencia lineal (centro) más la predicción de los errores (abajo).

El predictor kriging indicador es igual a:

$$I^*(x_0, z_l) = \sum_{i=1}^n \lambda_i(z_l) I(x_i, z_l)$$

es decir que la predicción de la función indicadora en el sitio  $x_0$  es igual a una combinación lineal de las  $n$  funciones indicadoras evaluadas en los sitios de medición. Samper and Carrera (1990) muestran que el kriging indicador es un estimador de la probabilidad acumulada hasta el límite  $z$  definido en la función indicadora. El predictor kriging indicador (dado que predice probabilidades acumuladas) tiene las siguientes restricciones:

- I.  $0 \leq I^*(x, z_l) \leq 1$
- II.  $I^*(x, z_l) \leq I^*(x, z'_l)$  si  $z_l \leq z'_l$

Una condición suficiente para que estas restricciones se cumplan es que  $\lambda_i(z_l) = \lambda_i$  con  $0 \leq \lambda_i \leq 1$ ,  $\forall i, \forall z_l$ . Sin embargo en la práctica las ponderaciones se estiman de tal forma que el predictor

sea insesgado de varianza mínima. Para la condición de insesgamiento:

$$E(I^*(x_0, z_l)) = E(I(x_0, z_l)) = F(z_l)$$

$$\sum_{i=1}^n \lambda_i(z_l) E(I(x_i, z_L)) = F(z_l)$$

$$\sum_{i=1}^n \lambda_i F(z_l) = F(z_l) \Rightarrow \sum_{i=1}^n \lambda_i(z_l) = 1$$

Después de llevar a cabo el proceso de derivación sobre la expresión de la varianza del error de predicción (obtenida de forma análoga a como se hizo en kriging ordinario), se obtiene el siguiente sistema de ecuaciones:

$$\sum_{i=1}^n \lambda_i(z_l) \gamma_{ij} + \mu = \gamma i 0, i = 1, 2, \dots, n$$

$$\sum_{i=1}^n \lambda_i = 1$$

donde  $\gamma_{ij} = \gamma(h)$ , la función de semivarianza evaluada para la distancia entre los sitios  $i, j$ . La varianza del error de predicción se encuentra de la misma forma a como se mencionó en la sección 4.2.

#### 4.6.9. Kriging Log-Normal y Multi-Gaussiano

En estos dos procedimientos se hacen transformaciones de la variable regionalizada con el propósito de normalizar en cada sitio de la región de estudio. El primero de estos consiste en aplicar kriging ordinario a la transformación logarítmica de los datos. Veamos: Sea  $\{Z(x) : x \in D\}$  una variable regionalizada log-normal. Es decir que  $Y(x) + \log(Z(x))$  tiene distribución normal. Algunas veces se requiere adicionar una constante positiva de tal forma que  $Y(x)$  esté definida. El predictor kriging log-normal es:

$$Y^* = \sum_{i=1}^n \lambda_i Y(x_i)$$

Los pesos se obtienen de manera análoga al kriging ordinario. El semivariograma usado es el de los valores transformados. La complicación práctica puede darse al hacer la retransformación a la escala original, puesto que  $Z^*(x_0) = \exp(Y^*(x_0))$  es un predictor sesgado. Se puede demostrar

que un predictor insesgado es (Cressie (1993)):  $Z^*(x_0) = \exp(Y^*(x_0)) + \frac{\sigma_{k0}^2}{2} - \mu$ , donde  $\sigma_{k0}^2$  es la varianza de predicción obtenida en el sitio  $x_0$  por medio de kriging ordinario sobre los valores transformados y  $\mu$  es el multiplicador de Lagrange empleado para la condición de insesgamiento sobre la escala de valores transformados. Respecto al kriging multi-gaussiano, suponga que se tiene una variable regionalizada  $\{Z(x) : x \in D\}$  estacionaria. Este procedimiento consiste en hacer una transformación de  $Z(x)$  tal que los valores transformados sigan una distribución normal estándar. En ese sentido es una generalización del kriging log-normal. Los pasos del método kriging multi-gaussiano son los siguientes:

- I. Se encuentra la función de probabilidad acumulada empírica  $F_n(Z(x))$ .
- II. Se calculan con base en  $F_n(Z(x))$  los “scores” normales estándar (Fig. 4.8), es decir los valores de una distribución de probabilidad normal estándar para los cuales la probabilidad acumulada corresponde a  $F_n(Z(x))$ . En otras palabras se encuentra  $U(x) = \Phi^{-1}(F_n(Z(x)))$ .
- III. Se realiza kriging simple sobre los valores transformados.

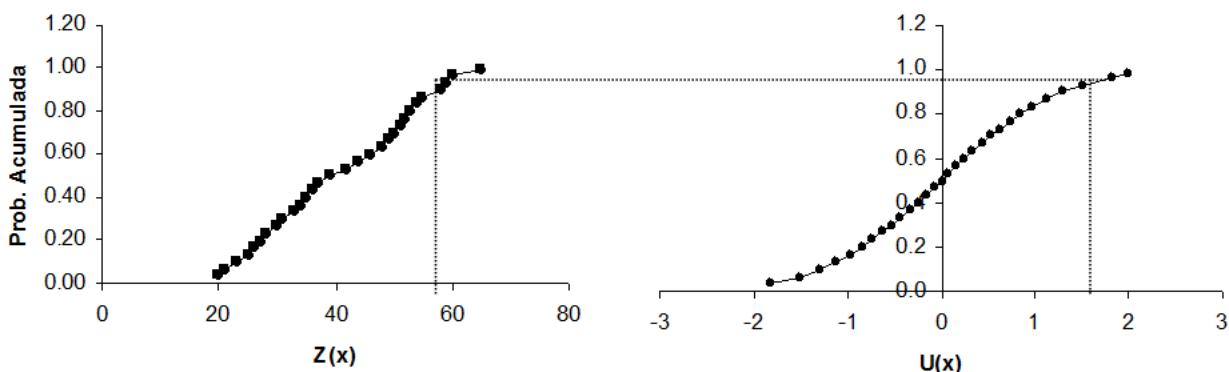


Figura 4.8: Representación de la transformación a scores normales.

#### 4.6.10. Cokriging Ordinario

Si se tienen dos variables regionalizadas  $Zv_1(x)$  y  $Zv_2(x)$  tomadas en sitios de la región de estudio, no necesariamente iguales, entonces el semivariograma cruzado entre ellas, se estima

por:

$$\gamma_{v_1 v_2}(h) = \frac{1}{2n_h} \sum_{k=1}^{n_h} \{Z_{v_1}(x+k) - Z_{v_1}(x)\} \{Z_{v_2}(x+k) - Z_{v_2}(x)\} \quad (4.24)$$

Donde  $n_h$  es el número de parejas de datos que se encuentran a una distancia  $h$  (Bogaert (1995)).

### Modelo Lineal de Corregionalización (MLC)

El MLC asume que todos los semivariogramas simples y cruzados pueden expresarse como una suma de modelos básicos (exponencial, esférico, gaussiano, etc.) idénticos. Para el caso de dos variables:

$$\begin{aligned} \gamma_{v_1}(h) &= \alpha_0 \gamma_0(h) + \dots + \alpha_m \gamma_m(h) \\ \gamma_{v_2}(h) &= \beta_0 \gamma_0(h) + \dots + \beta_m \gamma_m(h) \\ \gamma_{v_1 v_2}(h) &= \delta_0 \gamma_0(h) + \dots + \delta_m \gamma_m(h) \end{aligned} \quad (4.25)$$

donde  $\gamma_{v_1}(h)$  y  $\gamma_{v_2}(h)$  son los semivariogramas simples,  $\gamma_{v_1 v_2}(h)$  es el semivariograma cruzado.  $\gamma_0(h), \gamma_1(h), \dots, \gamma_m(h)$  son los modelos básicos de semivariograma y  $\alpha, \beta$  y  $\delta$  son constantes.

Matricialmente:

$$\Gamma(h) = \begin{pmatrix} \gamma_{v_1}(h) & \gamma_{v_1 v_2}(h) \\ \gamma_{v_1 v_2}(h) & \gamma_{v_2}(h) \end{pmatrix} = \sum_{s=0}^m B_s \lambda_s(h)$$

donde

$$B_s = \begin{pmatrix} \alpha_s & \delta_s \\ \delta_s & \beta_s \end{pmatrix} \quad \lambda_s(h) = \begin{pmatrix} \gamma_s(h) & 0 \\ 0 & \gamma_s(h) \end{pmatrix} \quad (4.26)$$

$\Gamma(h)$  se le conoce como matriz de corregionalización.

### Predictor Cokriging

El método de predicción espacial *cokriging* consiste en hacer predicción espacial de una variable con base en su información y en la de algunas variables auxiliares que estén correlacionadas espacialmente con ella. El predictor cokriging tiene la siguiente expresión en el caso en el que se considera una sola variable auxiliar:

$$\hat{Z}_{v_1}^*(x_0) = \sum_{i=1}^{n_1} a_i Z_{v_1}(x_i) + \sum_{j=1}^{n_2} b_j Z_{v_2}(x_j) \quad (4.27)$$

el lado izquierdo de la igualdad en la ecuación anterior representa la predicción de la variable de interés en el punto  $x_0$  no muestreado. Con  $i = 1, 2, \dots, n_1$ , representa la variable primaria. Así mismo, con  $j = 1, 2, \dots, n_2$ , representa la variable auxiliar.  $a_i$  y  $b_j$ , con  $i = 1, 2, \dots, n_1$  y  $j = 1, 2, \dots, n_2$  respectivamente, representan los pesos o ponderaciones de las observaciones de las variables primaria y auxiliar y se estiman con base en el MLC ajustado a los semivariogramas simples y cruzados. Los pesos  $a_i$  y  $b_j$  se estiman de manera análoga al proceso descrito para el método kriging ordinario, es decir estos serán los que minimizan la varianza del error de predicción sujeta a la restricción de que el predictor sea insesgado. La estimación de los parámetros se obtiene resolviendo el siguiente sistema de ecuaciones (Isaaks and Srivastava (1989)):

$$\left( \begin{array}{ccccccccc} \gamma_{v1}(1, 1) & \dots & \gamma_{v1}(n, 1) & \gamma_{v1 v2}(1, 1) & \dots & \gamma_{v1 v2}(m, 1) & 1 & 0 \\ \vdots & & \vdots & \vdots & & \vdots & \vdots & \vdots \\ \gamma_{v1}(1, n) & \dots & \gamma_{v1}(n, n) & \gamma_{v1 v2}(1, n) & \dots & \gamma_{v1 v2}(m, n) & 1 & 0 \\ \gamma_{v1 v2}(1, 1) & \dots & \gamma_{v1 v2}(n, 1) & \gamma_{v2}(1, 1) & \dots & \gamma_{v2}(m, 1) & 0 & 1 \\ \vdots & & \vdots & \vdots & & \vdots & \vdots & \vdots \\ \gamma_{v1 v2}(1, m) & \dots & \gamma_{v1 v2}(n, m) & \gamma_{v2}(1, m) & \dots & \gamma_{v2}(m, m) & 0 & 1 \\ 1 & \dots & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & \dots & 0 & 1 & \dots & 1 & 0 & 0 \end{array} \right) \begin{pmatrix} a_1 \\ \vdots \\ a_n \\ b_1 \\ \vdots \\ b_m \\ \mu_1 \\ \mu_2 \end{pmatrix} = \begin{pmatrix} \gamma_{v1}(0, 1) \\ \vdots \\ \gamma_{v1}(0, n) \\ \gamma_{v1 v2}(0, 1) \\ \vdots \\ \gamma_{v1 v2}(0, m) \\ 1 \\ 0 \end{pmatrix} \quad (4.28)$$

La matriz del lado izquierdo contiene los valores de las funciones de semivarianza y de semivarianza cruzada calculadas para todas las distancias entre las parejas de puntos consideradas. Las dos ultimas filas de dicha matriz son las correspondientes a la restricción de insesgamiento del predictor.  $a_i$  y  $b_j$  con  $i = 1, 2, \dots, n$  y  $j = 1, 2, \dots, m$ , son los parámetros a estimar,  $\mu_1$  y  $\mu_2$  son los multiplicadores de Lagrange empleados para la restricción de insesgamiento y el vector del lado derecho contiene los valores de la funciones de semivarianza y semivarianza cruzada evaluados para las distancia entre los sitios de muestreo (de ambas variables) y el sitio donde se desea hacer la predicción. Las dos últimas filas del vector están asociadas a la condición de insesgamiento. La correspondiente varianza de predicción del método cokriging se calcula como (Bogaert (1995)):

$$\begin{aligned} \sigma_k^2 &= Cov(Z_{v1}(x_0), Z_{v1}(x_0)) + \mu_1 + \sum_{i=1}^n a_i Cov(Z_{v1}(x_i), Z_{v1}(x_0)) \\ &\quad + \sum_{j=1}^m b_j Cov(Z_{v2}(x_j), Z_{v2}(x_0)) \end{aligned} \quad (4.29)$$

donde  $\mu_1$  es el multiplicador de Lagrange empleado para la restricción dado por la condición de insesgamiento ( $\sum_{i=1}^n a_i = 1$ ).  $Cov(Z_{vi}(x_k), Z_{vi}(x_l)) = \sigma_{vi}^2 - \gamma_{vivi}(k, l)$  la función de covarianza

espacial de la variable  $i, i = 1, 2$ , evaluada para la distancia entre los sitios de muestreo  $k, l$ . La ventaja del método cokriging frente al kriging radica en el hecho de que cuando la variable auxiliar está ampliamente correlacionada con la variable de interés se puede obtener un disminución en la varianza de predicción, no obstante dicha variable tenga menor densidad de muestreo. En situaciones en las que la variable objetivo tiene costos altos de muestreo se recomienda la aplicación de esta metodología (Bogaert (1995)).

### Kriging Probabilístico

Es un predictor basado en cokriging que utiliza como variables predictoras una variable indicadora y una variable generada a través de la *transformación uniforme*. Sea  $Z(x_i)$  la variable observada,  $i = 1, 2, \dots, n$ , entonces se definen las siguientes transformaciones:

- $I(x_i, z) = \begin{cases} 1 & \text{si } Z(x_i) \leq z \\ 0 & \text{otro caso} \end{cases}$
- $U(x_i) = \frac{R(Z(x_i))}{n}$  para todo  $i, i = 1, 2, \dots, n$ .

con  $R(Z(x_i))$  igual al rango (posición que ocupa dentro de los datos ordenados de menor a mayor) la  $i$ -ésima observación muestral. La predicción de probabilidad de éxito en el sitios de interés está dada por:

$$I^*(x_0) = \sum_{i=1}^n \lambda_i I(x_i, z) + \sum_{i=1}^n v_i U(x_i)$$

Los pesos  $\lambda_i$  y  $v_i$  se estiman mediante el sistema de ecuaciones del método cokriging.

### Predicción bajo Normalidad

$$\underline{S}(x) = \begin{pmatrix} S(x_1) \\ \vdots \\ S(x_n) \end{pmatrix} \sim N_n(\overrightarrow{\mu}, \sigma^2 \rho)$$

$$\rho_{(nxn)} \begin{pmatrix} 1 & \rho_{12} & \cdots & \rho_{1n} \\ & 1 & \cdots & \rho_{2n} \\ & & \ddots & \vdots \\ & & & 1 \end{pmatrix}$$

$$\rho_{ij} = \rho(||x_i - x_j||)$$

$$\underline{Z}(x) = \begin{pmatrix} Z(x_1) \\ \vdots \\ Z(x_n) \end{pmatrix} \sim N_n(\vec{0}, \Gamma^2 I)$$

$\underline{Z}(x)$  independiente de  $\underline{S}(x)$ .

$$\underline{Y}(x) = \begin{pmatrix} Y(x_1) \\ \vdots \\ Y(x_n) \end{pmatrix} \sim N_n(\vec{\mu}, V = [\sigma^2 \rho + \Gamma^2 I])$$

$$\begin{aligned} E(\underline{Y}(x)) &= E(\underline{S}(x) + \underline{Z}(x)) = \vec{\mu} \\ Cov(\underline{Y}(x)) &= Cov(\underline{S}(x) + \underline{Z}(x)) \\ &= \sigma^2 \rho + \Gamma^2 I \end{aligned}$$

$$\underline{Y}(x) | \underline{S}(x) \sim N_n(\underline{S}(x), \Gamma^2 I)$$

Porque

$$\begin{aligned} E(\underline{Y}(x) | \underline{S}(x)) &= \underline{S}(x) + E(\underline{Z}(x) | \underline{S}(x)) \\ &= \underline{S}(x) \\ Cov(\underline{Y}(x) | \underline{S}(x)) &= Cov[(\underline{S}(x) + \underline{Z}(x)) | \underline{S}(x)] \\ &= Cov(\underline{Z}(x) | \underline{S}(x)) \\ &= Cov(\underline{Z}(x)) \\ &= \Gamma^2 I \end{aligned}$$

Sea  $T = S(x_0)$ , entonces

$$\underline{S}(x) = \begin{pmatrix} S(x_0) \\ S(x_1) \\ \vdots \\ S(x_n) \end{pmatrix}_{(n+1) \times 1} \sim N_n(\vec{\mu}, \sigma^2 \rho)$$

$$\sigma^2 \rho = \begin{pmatrix} \sigma^2 & \sigma^2 \rho_{01} & \sigma^2 \rho_{02} & \dots & \sigma^2 \rho_{0n} \\ \sigma^2 \rho_{10} & \sigma^2 & \sigma^2 \rho_{12} & \dots & \sigma^2 \rho_{1n} \\ \sigma^2 \rho_{20} & \sigma^2 \rho_{21} & \sigma^2 & \dots & \sigma^2 \rho_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma^2 \rho_{n0} & \sigma^2 \rho_{n1} & \sigma^2 \rho_{n2} & \dots & \sigma^2 \end{pmatrix} = \begin{pmatrix} \sigma^2 & | & \sigma^2 r' \\ - & - & - \\ \sigma^2 r & | & \sigma^2 \rho \end{pmatrix} \quad \text{con} \quad r = \begin{pmatrix} \rho_{10} \\ \vdots \\ \rho_{n0} \end{pmatrix}$$

$$\underline{Z}(x) = \begin{pmatrix} 0 \\ Z(x_1) \\ \vdots \\ Z(x_n) \end{pmatrix} \sim N_{n+1}(\overrightarrow{0}, V), V = \begin{pmatrix} 0 & | & 0 & 0 & \dots & 0 \\ - & | & - & - & - & - \\ 0 & | & \Gamma^2 & 0 & \dots & 0 \\ 0 & | & 0 & \Gamma^2 & \dots & 0 \\ \vdots & | & \vdots & \vdots & \ddots & \vdots \\ 0 & | & 0 & 0 & \dots & \Gamma^2 \end{pmatrix}$$

$$V = \begin{pmatrix} 0 & | & \mathbf{0}' \\ - & - & \cdots \cdots \\ \mathbf{0} & | & (\Gamma^2 I)_{(n \times n)} \end{pmatrix} \quad \mathbf{0}' = (0, \dots, 0)_{(1 \times n)}$$

$$\underline{Y}(x)_{(n+1) \times 1} = \begin{pmatrix} Y(x_0) = S(x_0) \\ Y(x_1) \\ \vdots \\ Y(x_n) \end{pmatrix} = \begin{pmatrix} S(x_0) \\ S(x_1) + Z(x_1) \\ \vdots \\ S(x_n) + Z(X_n) \end{pmatrix} = \begin{pmatrix} \Gamma \\ Y \end{pmatrix}$$

$$\underline{Y}(x)_{(n+1) \times 1} \sim N_{(n+1)}(\overrightarrow{\mu}, \sum)$$

$$\begin{aligned}
\sigma^2 \rho &= \begin{pmatrix} \sigma^2 & \sigma^2 \rho_{01} & \sigma^2 \rho_{02} & \dots & \sigma^2 \rho_{0n} \\ \sigma^2 \rho_{10} & \sigma^2 & \sigma^2 \rho_{12} & \dots & \sigma^2 \rho_{1n} \\ \sigma^2 \rho_{20} & \sigma^2 \rho_{21} & \sigma^2 & \dots & \sigma^2 \rho_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma^2 \rho_{n0} & \sigma^2 \rho_{n1} & \sigma^2 \rho_{n2} & \dots & \sigma^2 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 & \dots & 0 \\ 0 & \Gamma^2 & 0 & \dots & 0 \\ 0 & 0 & \Gamma^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \Gamma^2 \end{pmatrix} \\
&= \begin{pmatrix} \sigma_{(1 \times 1)}^2 & | & (\sigma^2 r')_{(1 \times n)} \\ \hline \cdots & | & \cdots \\ (\sigma^2 r)_{(n \times 1)} & | & \sigma^2 \rho \end{pmatrix} + \begin{pmatrix} 0 & | & \mathbf{0}' \\ \hline - & | & \cdots \\ \mathbf{0} & | & (\Gamma^2 I)_{(n \times n)} \end{pmatrix} \\
&= \begin{pmatrix} \sigma^2 & | & (\sigma^2 r') \\ \hline \cdots & | & \cdots \\ \sigma^2 r & | & (\sigma^2 \rho + \Gamma^2 I)_{n \times n} \end{pmatrix}
\end{aligned}$$

De la normal multivariada Si  $\vec{X} = \begin{pmatrix} X^{(1)} \\ X^{(2)} \end{pmatrix} \sim N \left( \begin{pmatrix} \mu^{(1)} \\ \mu^{(2)} \end{pmatrix}, \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right)$

$X^{(1)}|X^{(2)} \sim (\mu^{(1)} + \sum_{12} \sum_{22}^{-1} (X^{(2)} - \mu^{(2)}), \sum_{11} - \sum_{12} \sum_{22}^{-1} \sum_{21})$  Usando el resultado anterior

$$T|Y_{(n \times 1)} \sim N(\mu(x_0) + \sigma^2 r'(\Gamma^2 I + \sigma^2 \rho)^{-1}(\underline{Y}(x)_{(n \times 1)} - \underline{\mu}^{(2)}), \sigma^2 - \sigma^2 r'(\Gamma^2 I + \sigma^2 \rho)^{-1} \sigma^2 r)$$

por consiguiente el predictor óptimo bajo pérdida cuadrática en el modelo Gaussiano estacionario es

$$\hat{T} = E(T|\underline{Y}(x)) = \mu(x_0) + \sigma^2 r'(\Gamma^2 I + \sigma^2 \rho)^{-1}(\underline{Y}(x) - \underline{\mu}^{(2)})$$

con varianza

$$V(T|\underline{Y}(x)) = \sigma^2 - \sigma^2 r'(\Gamma^2 I + \sigma^2 \rho)^{-1} \sigma^2 r$$

$\hat{T}$  es denominado en la literatura geoestadística Kriging simple.

# Bibliografía

- Assuncao, R. (2001). *Estatística Espacial con Aplicações em Epidemiologia, Economia e Sociología.* , UFScar. www.est.ufmg.br.
- Barry, J.T, C. M. and P. Diggle (1997). *Parametric estimation of the variogram.* Lancaster University Technical Report.
- Bogaert, P., P. M. . F. B. (1995). *The Spatial Interpolation of Agroclimatic Data.* Cokriging Software and Source Code.
- Box, G. and G. Jenkins (1976). *Time Series Analysis.* New York: Holden Day.
- Clark, I. (1979). *Practical Geostatistics.* Elsevier Publishing.
- Cressie, N. (1985). Fitting variogram models of weighted least squares. *Journal of the international Association of Mathematical Geology* 17, 563–86.
- Cressie, N. (1989). *Geostatistics.* The American Statistician 43, 197–202.
- Cressie, N. (1993). *Statistic for spatial data.* John Wiley Sons.
- Cressie, N. (1998). Fundamentals of spatial statistics. *Collecting Spatial Data: Optimum Design of Experiments for Random Field* W.G Muller, 9–33.
- Díaz-Francés, E. (1993). *Introducción a Conceptos Básicos de Geoestadística.* Guanajuato, México: Memorias Seminario Estadística y Medio Ambiente. Centro de Investigación en Matemáticas, CIMAT.
- Deutsch, C. V. . A. G. J. (1992). *GSLIB: Geostatistical Software Library and User's Guide.* New York: Oxford University Press.

- Fotheringham, A., C. Brunsdon, and M. Charlton (2002). *Geographically Weighted Regression.* John Wiley and Sons.
- Fuller, W. and G. Battece (1973). Transformations for estimations of linear models with nester error structure. *Journal of American Statistical Association* 68, 626–632.
- Isaaks, E. and M. Srivastava (1989). Applied geostatistics. *Oxford University Press.*
- Journel, A. and C. Huijbregts (1978a). Mining geostatistics. *Academic Press.*
- Journel, A. and C. Huijbregts (1978b). Mining Geostatistics. *Academic Press.*
- McCullagh, P. and J. Nelder (1989). Generlized linear Models. *Chapman and Hall.*
- Myers, D. (1987). Moptimization of sampling locations for variogram calculations. *Water Resources Research* 23, 283–293.
- Patterson, H. and R. Thompson (1971). recovery of inter-block information when block sizes are unequal. *Biometrika* 58, 545–554.
- Petitgas, P. (1996). *Geostatistics and their applications to fisheries survey data.* Computers and Fisheries Research 5, 114–142.
- Samper, F. and J. Carrera (1990). *Geoestadística. Aplicaciones a la hidrogeología subterránea.* UPC Barcelona: Centro Internacional de Métodos Numéricos en Ingeniería.
- Schimidt, P. (1976). *Econometrics.* New York: Marcel Dekker.
- Theil, H. (1971). *Principles of Econometrics.* Jhon Wiley and sons.
- Wackernagel, H. (1995). *Multivariable geostatistics:An introduction with applications.* Springer Verlag.
- Warrick, A. W., D. E. M. . D. R. N. (1986). Geostatistical methods applied to soil science. methods of soil analysis. part 1. *Physical and Mineralogical Methods- Agronomy Monograph* 9, 52–81.