

UNIVERSIDAD CENTROCCIDENTAL
“LISANDRO ALVARADO”

Decanato de Ciencias y Tecnología
Licenciatura en Ciencias Matemáticas



“ESTIMACIÓN POR MÁXIMA VEROSIMILITUD DE LOS
PARÁMETROS DE COVARIANZA ESPACIAL”

TRABAJO ESPECIAL DE GRADO PRESENTADO POR

BR. BETTY USECHE

COMO REQUISITO FINAL
PARA OBTENER EL TÍTULO DE LICENCIADO
EN CIENCIAS MATEMÁTICAS
ÁREA DE CONOCIMIENTO: ESTADÍSTICA.
TUTOR: PROF. LUZ RODRÍGUEZ

Barquisimeto, Venezuela. Febrero de 2014



Universidad Centroccidental
 "Lisandro Alvarado"
 Decanato de Ciencias y Tecnología
 Licenciatura en Ciencias Matemáticas



ACTA
 TRABAJO ESPECIAL DE GRADO

Los suscritos miembros del Jurado designado por el Jefe del Departamento de Matemáticas del Decanato de Ciencias y Tecnología de la Universidad Centroccidental "Lisandro Alvarado", para examinar y dictar el veredicto sobre el Trabajo Especial de Grado titulado:

"ESTIMACIÓN POR MÁXIMA VEROSIMILITUD DE LOS PARÁMETROS DE COVARIANZA ESPACIAL"

presentado por el ciudadano BR. BETTY USECHE titular de la Cédula de Identidad No. 18998729, con el propósito de cumplir con el requisito académico final para el otorgamiento del título de Licenciado en Ciencias Matemáticas.

Luego de realizada la Defensa y en los términos que imponen los Lineamientos para el Trabajo Especial de Grado de la Licenciatura en Ciencias Matemáticas, se procedió a discutirlo con el interesado habiéndose emitido el veredicto que a continuación se expresa:

¹ _____

Con una calificación de _____ puntos.

En fe de lo expuesto firmamos la presente Acta en la Ciudad de Barquisimeto a los _____ días del mes de _____ de _____.

TUTOR

FIRMA

PRINCIPAL

FIRMA

PRINCIPAL

FIRMA

OBSERVACIONES:

¹ Aprobado ó Reprobado

Primeramente gracias Dios todopoderoso y a todos aquellos que hicieron posible el logro de mi trabajo, en especial a mi madre Belkis y a mi abuela Cecilia.

AGRADECIMIENTOS

Primeramente doy GRACIAS A DIOS por haberme acompañado y guiado a lo largo de mis estudios, por ser mi fortaleza en los momentos de debilidad, por guiarme y permitirme el haber llegado hasta este momento tan importante de mi formación.

En forma especial, agradezco la confianza, el apoyo y cuidados brindado por parte de mi madre Belkis Fernández y mi abuela Cecilia Ortiz, por siempre guiarme por el mejor camino, por los valores que me han inculcado y ser un gran modelo a seguir. Sin ustedes, mis logros no serían posibles...MIL GRACIAS, LAS QUIERO Y AMO.

A mi familia, mi tío Guzman Bladimir Fernández, que ha sido el mejor modelo de padre a seguir, a mi tía Dora Useche, por quererme y apreciarme tanto, a mi primo Bladimir y Zurima Fernández, por siempre brindarme su cariño, apoyo, y ser unos hermanos para mí, a mi niña hermosa, mi cami, mi pequeña prima que la quiero y amo como una sobrina, que con sus besos y abrazos me llenan de alegría.

A ti cielo, por ser una parte importante de mi vida, por apoyarme en todo momento, por tu paciencia y amor incondicional...TE AMO.

A mi tutora, prof Luz Rodruíguez, GRACIAS por su guía, sus consejos y haberme dado la oportunidad de desarrollar mi tesis a su lado. Mis respetos y admiración. Dios la bendiga siempre a usted y su familia.

A mi abuelita Lucia, abuelo Magdaleno y a mi amigo Elio que ya no se encuentra con nosotros físicamente, siempre estarán presente en mi corazón, por haberme apoyado, creído en mí siempre y estar cuidado desde el cielo.

A mis amigos, Lilibeth, Silmaris, Dionny, Ana y Yogeidys, por confiar, creer y brindarme su amistad incondicional en todos los momentos de la carrera. Gracias muchachos.

Son muchas las personas que han formado parte de mi a las que me encantaría agradecerles su amistad, consejos, apoyo, ánimo y compañía en los momentos más difíciles de mi vida. Algunas están aquí conmigo y otras en mis recuerdos y en mi corazón, sin importar en donde estén quiero darles las gracias por formar parte de mí, por todo lo que me han brindado y por todas sus bendiciones.

A TODOS GRACIAS...

RESUMEN

En este trabajo, el método de máxima verosimilitud para inferir sobre los parámetros de covarianzas espaciales son examinado. Las ventajas de la estimación por máxima verosimilitud son discutidas y se muestra que este método, se deriva de la suposición de una distribución gaussiana multivariada para los datos, dando un criterio robusto para el ajuste de los modelos de covarianza independientemente de la distribución multivariada de los datos. Sin embargo, esta distribución es imposible verificar en la práctica cuando sólo una realización de la función aleatoria está disponible. Entonces, el método de máxima entropía es el único criterio robusto de asignación de probabilidades en ausencia de información.

Debido a que la distribución gaussiana multivariada tiene la propiedad de máxima entropía para un vector de medias y matriz de covarianza fijos, la distribución multinormal es la opción más lógica, como una distribución predeterminada para los datos experimentales. Sin embargo, debería ser claro que la suposición de una distribución Gaussiana multivariada sólo se mantiene para la inferencia de los parámetros de covarianza espacial y no necesariamente para otras operaciones, tales como la interpolación espacial, simulación o estimación de distribuciones espaciales. Algunos resultados de simulación serán presentados para apoyar la afirmación de que el uso simultáneo del método de máxima verosimilitud y el método clásico no paramétrico de los momentos pueden mejorar considerablemente los resultados de la estimación de parámetros geoestadísticos.

ÍNDICE

Agradecimientos	i
Resumen	iii
1. Introducción.	1
1.1. Inferencia Basada en Máxima Verosimilitud.	2
1.2. Máxima Verosimilitud.	2
1.3. Naturaleza del Problema de Estimación.	3
2. Conceptos Fundamentales.	4
2.1. Conceptos Básicos.	4
2.2. Derivada de una Función Escalar de una Matriz.	5
2.2.1. Propiedades de la Derivada de una Función.	5
2.3. Variables Aleatorias.	6
2.4. Función de Distribución Acumulada (Fda).	6
2.5. Función de Densidad.	9
2.5.1. Propiedades de la Función de Densidad.	9
2.6. Esperanza.	10
2.6.1. Esperanza de una Variable Aleatoria Discreta.	10
2.6.2. Esperanza de una Variable Aleatoria Continua.	10
2.6.3. Propiedades de la Esperanza.	11
2.7. Varianza.	11
2.7.1. Propiedades de la Varianza.	12
2.8. Covarianza y Correlación.	12
2.8.1. Propiedades de la Covarianza.	12
2.8.2. Propiedades de la Correlación.	13
2.9. Muestra Aleatoria.	14

2.10. Estimadores y Estimaciones.	14
2.10.1. Estimación Puntual.	14
2.10.2. Propiedades de los Estimadores.	14
2.11. Distribuciones de Vectores Normales.	15
2.12. Momento de segundo orden.	17
2.13. La distribución Normal Mutivariada.	18
2.13.1. Densidad general.	19
2.13.2. Media y Covarianza muestral.	19
3. Campo Aleatorio.	20
3.1. Estimación De Variograma.	24
3.2. Máxima Verosimilitud.	25
3.3. Máxima Verosimilitud Restringida.	25
3.4. Mínimos Cuadrados.	26
4. El Metodo Inferencial de Máxima Verosimilitud	28
4.1. Estimador para β	30
4.2. Estimador para σ^2	31
4.3. Estimador para los Parámetros de Covarianza.	31
4.4. Caracterización del Modelo.	32
4.5. Ejemplo.	34
Conclusiones	41
Referencias Bibliográficas	42

CAPÍTULO 1

INTRODUCCIÓN.

El método de máxima verosimilitud no fué formulado o usado como una técnica general de estimación hasta que R.A. Fisher en 1912 introdujo una forma generalizada y rigurosa de éste. Fisher, más que introducir el concepto simple, publicó una serie de artículos en los cuales extendió estos conceptos a un comprensivo y unificado sistema de estadística matemática al igual que una filosofía de inferencia estadística la cual ha tenido un profundo y ancho desenvolvimiento.

Los problemas teóricos inherentes a los estimadores de máxima verosimilitud son principalmente aquellos concernientes a las propiedades de varianza de los estimadores, particularmente cuando el tamaño de la muestra es pequeña. Para tamaños de muestras grandes y cuando las observaciones son independientes, la teoría ha sido bien desarrollada, y existen una variedad de resultados para las propiedades asintóticas de los estimadores, en particular para la varianza de los mismos. A pesar de algunas desventajas, particularmente cuando los tamaños muestrales son pequeños, el método de máxima verosimilitud es atractivo, casi como una técnica práctica universal para formular ecuaciones de estimación.

La estimación por máxima verosimilitud (ML) es uno de los métodos más importantes para la estimación de parámetros en campos aleatorios Gaussianos. Ello es así por la versatilidad y buenas propiedades estadísticas que en general poseen los métodos inferenciales basados en la verosimilitud.

En la práctica, algunos investigadores prefieren hacer inferencia por ML (Mardia y Marshall, 1984; Vecchia, 1988), mientras que otros prefieren usar otros métodos (Cressie, 1993). Estos últimos argumentan que los estimadores ML de parámetros de covarianza

son sesgados, sobre todo cuando se tienen muestras pequeñas y/o cuando el modelo incluye varios parámetros de regresión. Sin embargo, el sesgo es sólo un aspecto de un estimador, y para establecer conclusiones más sólidas, la varianza del estimador debe ser considerada, o mejor aún, el error cuadrático medio del estimador.

1.1. Inferencia Basada en Máxima Verosimilitud.

Los métodos de inferencia basados de alguna forma en verosimilitud, entre estos el método de máxima verosimilitud, son tal vez los métodos más versátiles para ajustar datos de modelos estadísticos. En aplicaciones típicas, la meta es usar un modelo paramétrico para describir un conjunto de datos o un proceso que genere un conjunto de datos. Aparte de su fuerte motivación intuitiva, el mayor atractivo de los métodos estadísticos basados en alguna forma de verosimilitud consiste en que éstos pueden ser aplicados a una gran variedad de modelos y clases de datos (continuos, discretos, categóricos, censurados, truncados, etc.), donde otros métodos populares, tales como mínimos cuadrados, no proveen en general un método satisfactorio para hacer inferencia estadística.

Kitanidis (1983) fue (aparentemente) el primero en proponer el uso de métodos basados en la verosimilitud para estimación de parámetros en campos aleatorios Gaussianos. El uso del método de máxima verosimilitud para la estimación de parámetros en campos aleatorios Gaussianos fue desarrollado y estudiado por Kitanidis y Lane (1985), Mardia y Marshall (1984), Mardia y Watkins (1989) y Vecchia (1988), entre otros.

1.2. Máxima Verosimilitud.

Existen métodos de estimación que están basados en distribuciones iniciales, pero es útil poder aplicar un método relativamente sencillo para construir un estimador sin tener que especificar una distribución inicial. Un método con éstas características es el que se denomina Método de Máxima Verosimilitud que se puede aplicar a la mayoría de los problemas, tiene un fuerte atractivo intuitivo y usualmente proporciona una estimación razonable del parámetro de θ .

Además, si la muestra es grande, el método de máxima verosimilitud es quizás el método de estimación más ampliamente utilizado en estadística.

1.3. Naturaleza del Problema de Estimación.

Supongamos que se va a seleccionar una muestra aleatoria X_1, \dots, X_n de una distribución cuya función de densidad de probabilidad es $f(x|\theta)$ donde el valor del parámetro θ es desconocido.

Supóngase además que el valor de θ debe pertenecer a un intervalo concreto Ω sobre la recta real, y que el valor de θ se debe estimar a partir de los valores observados de la muestra.

Un estimador del parámetro θ , basado en variables aleatorias X_1, \dots, X_n es una función $\delta(X_1, \dots, X_n)$ que especifica el valor esperado de θ para cada conjunto de valores posibles de X_1, \dots, X_n . En otras palabras, si los valores observados de X_1, \dots, X_n son x_1, \dots, x_n entonces el valor del estimador de θ es $\delta(x_1, \dots, x_n)$ puesto que el valor de θ debe pertenecer al intervalo Ω .

Es conveniente distinguir los términos estimador y estimación.

Un estimador $\delta(X_1, \dots, X_n)$ es una función de los variables aleatorias X_1, \dots, X_n y el es una variable aleatoria cuya distribución de probabilidad se puede obtener a partir de la distribución conjunta de X_1, \dots, X_n .

Por otro lado, una estimación es un valor específico $\delta(x_1, \dots, x_n)$ del estimador que se determina utilizando valores observados específicos x_1, \dots, x_n de la muestra aleatoria (X_1, \dots, X_n) .

CAPÍTULO 2

CONCEPTOS FUNDAMENTALES.

Este capítulo tiene como objetivo dar una breve introducción a la teoría de probabilidad y estadística, haciendo énfasis en el método de máxima verosimilitud, estableciendo definiciones, propiedades y notaciones, que son necesarias para el desarrollo del trabajo.

2.1. Conceptos Básicos.

Definición 2.1. Si A es una matriz $m \times n$, entonces la transpuesta de A , denotada por A' o A^t , es una matriz $n \times m$ cuyo i -ésimo renglón es la i -ésima columna de A y cuya j -ésima columna es el j -ésimo renglón de A . En símbolos, si $A = (a_{ij})$, entonces $A' = (a_{ji})$.

Definición 2.2. Una matriz cuadrada $A : n \times n$ es simétrica, si $A = A'$.

Definición 2.3. Una matriz $A : n \times n$ es idempotente, si y sólo si $A^2 = A$.

Definición 2.4. Un conjunto $S = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k\}$ en \mathbb{R}^n es un conjunto ortogonal si $\mathbf{u}_i \cdot \mathbf{u}_j = 0$, siempre que $i \neq j$.

Definición 2.5. Un conjunto $S = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k\}$ de vectores en \mathbb{R}^n es ortonormal si, S es ortogonal y todo vector en S tiene longitud unitaria, es decir, $|\mathbf{u}_i| = 1$ para toda i .

Definición 2.6. Una matriz $A : n \times n$ es ortogonal, si sus columnas forman un conjunto ortonormal de vectores.

Definición 2.7. Sea $A = (a_{ij})_{n \times n}$. Diremos que la matriz A es definida positiva si

$$\sum_i \sum_j a_{ij} > 0.$$

Y es semidefinida positiva si

$$\sum_i \sum_j a_{ij} \geq 0.$$

2.2. Derivada de una Función Escalar de una Matriz.

Definición 2.8. La derivada de una función escalar f de una matriz $X = (x_{ij})$, con $i = 1, \dots, p$ y $j = 1, \dots, n$, esta definida como:

$$\frac{\partial f(X)}{\partial X} = \left(\frac{\partial f(X)}{\partial x_{ij}} \right), \quad \begin{array}{l} i = 1, \dots, p \\ j = 1, \dots, n \end{array}$$

2.2.1. Propiedades de la Derivada de una Función.

1. Sea $X : p \times p$, $|X| \neq 0$

Para $X \neq X'$,

$$\frac{\partial}{\partial X}|X| = |X|(X^{-1})', \text{ para } X : p \times p, |X| \neq 0.$$

Para $X = X'$,

$$\frac{\partial}{\partial X}|X| = 2|X|X^{-1} - \text{diag}(|X|X^{-1}).$$

2. Para $A' : p \times q$, $X : q \times p$

$$\frac{\partial}{\partial X} \text{tr}(A'X) = A \quad \text{y} \quad \frac{\partial}{\partial X} \text{tr}(X'A) = A.$$

3. Para $A : p \times p$, $X : p \times 1$

$$\frac{\partial}{\partial X}(X'AX) = 2AX.$$

4. Para $A : p \times q$, $X : q \times p$

$$\frac{\partial}{\partial X}(AX) = A'.$$

5. Si $X : p \times p$ y $X = X'$,

$$\frac{\partial}{\partial X} \text{tr} X^2 = 2X.$$

La prueba es inmediata de la definición.

2.3. Variables Aleatorias.

La relación entre los sucesos del espacio muestral y el valor numérico que se les asigna se establece a través de variables aleatorias.

Definición 2.9. Una variable aleatoria es una función que asigna un valor numérico a cada suceso elemental del espacio muestral. Es decir, una variable aleatoria es una variable cuyo valor numérico está determinado por el resultado del experimento aleatorio. La variable aleatoria la denotaremos con letras mayúsculas X, Y, \dots y con las letras minúsculas x, y, \dots sus valores.

La variable aleatoria puede tomar un número numerable o no numerable de valores, dando lugar a dos tipos de variables aleatorias, discretas y continuas.

Definición 2.10. Se dice que una variable aleatoria X es discreta si puede tomar un número finito o infinito, pero numerable, de posibles valores.

Definición 2.11. Se dice que una variable aleatoria X es continua si puede tomar un número infinito no numerable de valores, o bien, si puede tomar un número infinito de valores correspondientes a los puntos de uno más intervalos de la recta real.

2.4. Función de Distribución Acumulada (Fda).

Sean X, Y dos variables aleatorias definidas conjuntamente, es decir; X e Y tienen una distribución de probabilidad conjunta cuya Función de distribución acumulada (Fda) conjunta está dada por:

$$F(x, y) = P\{X \leq x, Y \leq y\}.$$

De manera general, cuando $X' = (X_1, \dots, X_p)$ es un vector de variable aleatorias que son distribuidas conjuntamente la Fda está dada por:

$$F(x) = F(x_1, \dots, x_p) = P\{X_1 \leq x_1, \dots, X_p \leq x_p\}.$$

Toda Fda multivariada F satisface las siguientes propiedades:

1. F es monótona no-decreciente en cada componente de X .

Basta probar que $F(E) \geq 0$ con $E \subset X \subset \mathbb{R}$.

Sabemos que $P\{a \leq X \leq b\} = F(b) - F(a) = F(E)$.

Definiendo $X = \Omega$ (el espacio muestral) tenemos que $E \subset X$ es un evento, luego por axioma de probabilidad

$$P\{E\} \geq 0 \Rightarrow F(E) \geq 0.$$

Así se cumple la propiedad.

2. $0 \leq F(x) \leq 1$.

Sea $S \subset X = \Omega$ (evento).

Por lo anterior $F(S) \geq 0$. Falta probar que $F(S) \leq 1$ ahora $\Omega = S \cup S^c$.

Entonces:

$$\begin{aligned} P\{\Omega\} &= P\{S \cup S^c\} = P\{S\} + P\{S^c\} = 1 \\ \implies P\{S\} &= 1 - P\{S^c\} \leq 1 \quad (\text{ya que } P\{S^c\} \geq 0) \end{aligned}$$

luego, $0 \leq P\{S\} \leq 1$, y como S es un evento arbitrario, se cumple la propiedad.

3. $F(-\infty, x_2, \dots, x_p) = F(x_1, -\infty, \dots, x_p) = \dots = F(x_1, x_2, \dots, -\infty)$.

Sabemos que

$F(x_1, x_2, \dots, x_p) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \dots \int_{-\infty}^{x_p} f(x_1, x_2, \dots, x_p) dx_p \dots dx_2 dx_1$ donde f es la función de densidad de $X = (X_1, X_2, \dots, X_p)$, como f es continua

$$\begin{aligned} F(-\infty, x_2, \dots, x_p) &= \int_{-\infty}^{-\infty} \int_{-\infty}^{x_2} \dots \int_{-\infty}^{x_p} f(x_1, x_2, \dots, x_p) dx_p \dots dx_2 dx_1 \\ &= \int_{-\infty}^{x_p} \int_{-\infty}^{x_2} \dots \int_{-\infty}^{-\infty} f(x_1, x_2, \dots, x_p) dx_1 \dots dx_2 dx_p \\ &= 0. \end{aligned}$$

Análogamente se prueba para $F(x_1, -\infty, \dots, x_p)$.

$$\therefore F(-\infty, x_2, \dots, x_p) = F(x_1, -\infty, \dots, x_p) = \dots = F(x_1, x_2, \dots, -\infty).$$

4. $F(\infty, \infty, \dots, \infty) = 1$.

$$\begin{aligned} F(\infty, \infty, \dots, \infty) &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(x_1, \dots, x_p) dx_p \dots dx_1 \\ &= 1, \end{aligned}$$

ya que f es una función de densidad de X .

5. La probabilidad de un rectángulo de dimensión p es no-negativo. Probaremos para $p = 2$.

$$\begin{aligned} P\{x_1 \leq X \leq x_2, y_1 \leq Y \leq y_2\} &= F(x_2, y_2) - F(x_2, y_1) - F(x_1, y_2) + F(x_1, y_1) \\ &\geq 0. \end{aligned}$$

Todas las propiedades son análogas al caso univariado excepto la última propiedad, existen funciones que cumplen las primeras 4 propiedades y la última no la cumple, así no son fdc.

Ejemplo 2.1. Supongamos que:

$$F(x_1, x_2) = \begin{cases} 0, & \text{si } x_1 \leq 0 \text{ o } x_2 \leq 0 \text{ o } x_1 + x_2 \leq 1; \\ 1, & \text{en otro caso.} \end{cases}$$

esta función satisface las primeras 4 propiedades, lo cual es suficiente para una fdc en el caso univariado, pero dado que

$$F(1, 1) - F(1, \frac{1}{2}) - F(\frac{1}{2}, 1) + F(\frac{1}{2}, \frac{1}{2}) = -1$$

no es un fdc ya que no cumple la última propiedad, así $F(x_1, x_2)$ no puede ser una fdc bivariada.

Asumiremos que todas las funciones $F(x)$ serán continuas, en consecuencia ésta será expresada como la integral de una función $f(x)$ llamada densidad, es decir:

$$F(X) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \dots \int_{-\infty}^{x_p} f(x) dx. \quad (2.1)$$

2.5. Función de Densidad.

Supongamos que $F(X)$ es continua; entonces de (2.1) la función de densidad conjunta (fdc) de X es:

$$f(x) = f(x_1, \dots, x_p) = \frac{\partial^p F(x)}{\partial x_1 \dots \partial x_p}. \quad (2.2)$$

Hay conjuntos donde los valores de x en (2.2) no existen. Análogo al caso univariado, esto es una relación para la probabilidad de un evento (o conjunto de valores en el espacio de dimensión p) en términos de la densidad conjunta para $X : p \times 1$

$$P\{X \subseteq R\} = \int \dots \int_R f(x) dx \quad (2.3)$$

para una región R .

2.5.1. Propiedades de la Función de Densidad.

Si X es una variable aleatoria de distribución continua entonces $P(X = x) = 0$ si x es un valor particular. Por esta razón, una función de densidad se puede redefinir en un número finito de puntos de un intervalo sin alterar el valor de la integral sobre dicho intervalo, es decir, sin modificar las probabilidades referidas a la variable X . En este sentido la función de densidad asociada a una variable X no es única; de aquí que, preferiblemente, se utilizan funciones de densidad continuas.

El conjunto

$$\{x \in \mathbb{R} : f(x) \neq 0\}$$

se llama soporte de la función de densidad. En general, algunas de las propiedades que debe satisfacer una función de densidad $f(x)$ son:

- No negativa, es decir, $f(x) \geq 0$.
- Consistente con la probabilidad de S , es decir,

$$\int_{-\infty}^{\infty} f(x)dx = 1.$$

2.6. Esperanza.

2.6.1. Esperanza de una Variable Aleatoria Discreta.

El concepto de esperanza es fácil de extender. Si X denota una variable aleatoria discreta que puede tomar los valores x_1, x_2, \dots, x_n con sus respectivas probabilidades p_1, p_2, \dots, p_n , donde $p_1 + p_2 + \dots + p_n = 1$, la *esperanza de X* denotada por $E(X)$, se define como:

$$E(X) = p_1x_1 + p_2x_2 + \dots + p_kx_k = \sum_{j=1}^k p_jx_j.$$

2.6.2. Esperanza de una Variable Aleatoria Continua.

Si una variable aleatoria X tiene una distribución continua cuya función de densidad es $f(x)$, entonces la *Esperanza de X* , denotada por $E(X)$ se define por:

$$E(x) = \int_{-\infty}^{\infty} xf(x)dx.$$

La condición matemática para la existencia de $E(X)$ es la convergencia absoluta de la integral que la define, es decir, $E(X)$ existe, si y sólo si, existe la integral

$$E(x) = \int_{-\infty}^{\infty} |x| f(x).$$

Aunque hay semejanza entre la esperanza de una variable aleatoria y el concepto físico de centro de gravedad, una función de densidad $f(x)$ no necesariamente tiene esperanza.

Si X es una variable aleatoria con función de densidad $f(x)$, entonces la esperanza (media, valor esperado) de una función $Y = h(X)$ se puede evaluar sin encontrar la función de densidad $g(y)$ de Y . El valor $E(h(X))$ se calcula por:

$$\begin{aligned} E(h(X)) &= \int_{-\infty}^{\infty} h(x)f(x)dx \\ &= \int_{-\infty}^{\infty} yg(y)dy, \end{aligned}$$

y la esperanza $E(h(X))$ existe, si y sólo si, cumple

$$\int_{-\infty}^{\infty} |h(x)| f(x)dx < \infty.$$

2.6.3. Propiedades de la Esperanza.

La esperanza es un operador lineal, ya que:

- $E(X + c) = E(X) + c$.
- $E(X + Y) = E(X) + E(Y)$.
- $E(aX) = aE(X)$.

Combinando estas propiedades, podemos ver que:

- $E(aX + b) = aE(X) + b$.
- $E(aX + bY) = aE(X) + bE(Y)$.

donde X e Y son variable aleatorias y a , b y c son tres constantes cualquiera.

2.7. Varianza.

Si X es una variable aleatoria con media $E(X) = \mu$, la varianza de una variable aleatoria X se define como el valor esperado de $(X - \mu)^2$. Esto es,

$$\begin{aligned} V(X) &= E[(X - \mu)^2] \\ &= E(X)^2 - \mu^2. \end{aligned}$$

La *desviación estándar* de X es la raíz cuadrada positiva de $V(X)$.

Como la varianza es la media de una variable aleatoria no negativa, entonces $V(X) \geq 0$.

2.7.1. Propiedades de la Varianza.

- $V(X)=0$ si, y sólo si, existe una constante c tal que $P(X=c) = 1$.
- Si a y b son constantes,

$$V(aX + b) = a^2V(X).$$

- $V(X) = E(X^2) - (E(X))^2$.
- $V(X_1 + X_2) = V(X_1) + V(X_2) + 2[E(X_1X_2) - E(X_1)E(X_2)]$.

2.8. Covarianza y Correlación.

Si X_1 y X_2 son variables aleatorias con medias μ_1 y μ_2 , respectivamente, la *covarianza* de X_1 y X_2 es

$$Cov(X_1, X_2) = E[(X_1 - \mu_1)(X_2 - \mu_2)].$$

Cuanto mayor sea el valor absoluto de la covarianza de X_1 y X_2 , mayor será la dependencia lineal entre X_1 y X_2 . Los valores positivos indican que X_1 aumenta cuando X_2 aumenta; los valores negativos indican que X_1 disminuye cuando X_2 aumenta. Un valor cero de la covarianza indica que las variables son *no correlacionadas* y que no hay dependencia lineal entre X_1 y X_2 .

2.8.1. Propiedades de la Covarianza.

Si X, X_1, X_2 y X_3 son variables aleatorias y a y b son constantes, se cumple:

- $Cov(a,b) = 0$ a, b constantes.
- $Cov(X,X)=V(X)$.

- $Cov(X_1, X_2) = Cov(X_2, X_1)$.
- $Cov(X, a) = 0$.
- $Cov(X_1 + a, X_2 + b) = Cov(X_1, X_2)$.
- $Cov(aX_1, bX_2) = ab Cov(X_1, X_2)$.
- $Cov(X_1 + X_2, X_3) = Cov(X_1, X_3) + Cov(X_2, X_3)$.

Nótese que si X_1 y X_2 son variables aleatorias con media μ_1 y μ_2 , respectivamente, entonces

$$\begin{aligned} Cov(X_1, X_2) &= E[(X_1 - \mu_1)(X_2 - \mu_2)] \\ &= E(X_1 X_2) - E(X_1)E(X_2). \end{aligned}$$

Desafortunadamente, es difícil utilizar la covarianza como medida absoluta de dependencia porque su valor depende de la escala de medición. En consecuencia, es difícil determinar a primera vista si una covarianza particular es grande o pequeña. Este problema se puede eliminar al estandarizar su valor y usar el *coeficiente de correlación*, ρ , una cantidad con la varianza y que se define como:

$$\rho = \frac{Cov(X_1 X_2)}{\sigma_1 \sigma_2}$$

donde σ_1 y σ_2 son desviaciones estándar de X_1 y X_2 , respectivamente.

2.8.2. Propiedades de la Correlación.

La correlación ρ no tiene dimensión y tiene las propiedades siguientes:

- $\rho(X, Y) = \rho(Y, X)$.
- $-1 \leq \rho \leq 1$.
- $\rho(X, X) = 1$, $\rho(X, -X) = -1$.
- $\rho(aX + b, cY + d) = \rho(X, Y)$ si $a, c \neq 0$.

2.9. Muestra Aleatoria.

Sea X una variable aleatoria con cierta distribución de probabilidad. Sean X_1, \dots, X_n las n variables aleatorias independientes que tiene cada una la misma distribución que X . Llamamos entonces a (X_1, \dots, X_n) una muestra aleatoria de la variable aleatoria X .

2.10. Estimadores y Estimaciones.

Los parámetros poblacionales o simplemente parámetros son características numéricas de la población.

La mayoría de las veces estos parámetros son desconocidos y es necesario estimarlos mediante una muestra de la población. Las variables aleatorias utilizadas para estimar los parámetros reciben el nombre de *estimadores*, y los valores específicos de estas variables se llaman *estimaciones* de los parámetros.

2.10.1. Estimación Puntual.

Las estimaciones que especifican un valor único para el parámetro se denominan *estimadores puntuales*.

Sea X_1, \dots, X_n una muestra aleatoria de una v.a X con función de densidad $f(x, \theta)$ con θ desconocido. Para dar una estimación del valor de θ procedemos así:

- Definimos $T = u(X_1, \dots, X_n)$ una v.a que depende de la muestra X_1, \dots, X_n . Así, el estadístico T es un estimador de θ .
- Una vez definido T , observamos los valores de las variables X_1, \dots, X_n , digamos $X_1 = x_1, \dots, X_n = x_n$ y consideraremos el valor $u(x_1, \dots, x_n)$ como una estimación de θ .

2.10.2. Propiedades de los Estimadores.

Las siguientes son propiedades y definiciones asociadas con los estimadores de los parámetros de cierta función de densidad

- El estimador $\hat{\theta}$ de θ es *insesgado* si $E(\hat{\theta}) = \theta$.

- El *sesgo* del estimador $\hat{\theta}$ es $B(\hat{\theta}) = E(\hat{\theta}) - \theta$.

- El *error cuadrático medio* (ECM) de $\hat{\theta}$ es,

$$\begin{aligned} ECM(\hat{\theta}) &= E[(\hat{\theta} - \theta)^2] \\ &= var(\hat{\theta}) + B(\hat{\theta})^2. \end{aligned}$$

- El *error de estimación* es $\epsilon = |\hat{\theta} - \theta|$.

- Sean $\hat{\theta}_1$ y $\hat{\theta}_2$ dos estimadores de θ ,

(a) Si $var(\hat{\theta}_1) < var(\hat{\theta}_2)$ entonces el estimador $\hat{\theta}_1$ es *relativamente más eficiente* que el estimador $\hat{\theta}_2$.

(b) La *eficiencia relativa* de $\hat{\theta}_2$ relativa a $\hat{\theta}_1$ es,

$$Eficiencia = \frac{\hat{\theta}_1}{\hat{\theta}_2}.$$

- $\hat{\theta}_1$ es un *estimador consistente* de θ si para cada $\epsilon > 0$,

$$\begin{aligned} \lim_{n \rightarrow \infty} P(|\hat{\theta} - \theta| \leq \epsilon) &= 1, \text{ o, en forma equivalente} \\ \lim_{n \rightarrow \infty} P(|\hat{\theta} - \theta| > \epsilon) &= 0. \end{aligned}$$

- $\hat{\theta}$ es un *estimador suficiente* para θ si para cada valor de $\hat{\theta}$, la distribución condicional de X_1, \dots, X_n dado que $\hat{\theta} = \theta_0$ es independiente de θ .

- Un estimador $\hat{\theta}$ de θ es un *estimador insesgado de mínima varianza* (EIMV) si es insesgado y posee la mínima varianza.

2.11. Distribuciones de Vectores Normales.

Sea X una matriz $n \times p$ que contiene n observaciones para cada una de las p variables aleatorias X_1, X_2, \dots, X_p , es decir,

$$X = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p)$$

donde

$$\mathbf{x}_j = \begin{pmatrix} x_{1j} \\ x_{2j} \\ \vdots \\ x_{nj} \end{pmatrix}, \quad j = 1, \dots, p.$$

Nótese que cada columna de X es un vector $n \times 1$ de observaciones para cada una de las p variables, y cada fila contiene observaciones para las p variables X_1, X_2, \dots, X_p correspondientes al i -ésimo individuo, $i = 1, \dots, n$.

Las p variables forman un vector de variables aleatorias o *variable aleatoria multivariada* \mathbf{X}

$$\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{pmatrix}^T$$

y la matriz de n observaciones de la variable aleatoria multivariada \mathbf{X} es:

$$X = \begin{pmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ \vdots \\ \mathbf{x}_n^T \end{pmatrix}.$$

Para una variable aleatoria n -dimensional $\mathbf{X} = (X_1, X_2, \dots, X_n)'$ la esperanza se define como:

$$E(\mathbf{X}) = \begin{pmatrix} E(X_1) \\ E(X_2) \\ \vdots \\ E(X_n) \end{pmatrix}$$

y la matriz de varianzas de \mathbf{X} como la matriz que contiene todas las varianzas y covarianzas de los elementos de \mathbf{X} , es decir,

$$Var(\mathbf{X}) = E\{\mathbf{X}\mathbf{X}' - \mathbf{X}[E(\mathbf{X})]' - \mathbf{X}'E(\mathbf{X}) + E(\mathbf{X})[E(\mathbf{X})]'\}.$$

La covarianza entre dos vectores aleatorios \mathbf{X} y \mathbf{Y} donde $\mathbf{X} = (X_1, X_2, \dots, X_n)'$ y $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)'$ se define como:

$$\begin{aligned} Cov(\mathbf{X}, \mathbf{Y}) &= E\{(\mathbf{X} - E(\mathbf{X}))(\mathbf{Y} - E(\mathbf{Y}))'\} \\ &= \begin{pmatrix} cov(X_1, Y_1) & cov(X_1, Y_2) & \cdots & cov(X_1, Y_n) \\ cov(X_2, Y_1) & cov(X_2, Y_2) & \cdots & cov(X_2, Y_n) \\ \vdots & \vdots & \cdots & \vdots \\ cov(X_n, Y_1) & cov(X_n, Y_2) & \cdots & cov(X_n, Y_n) \end{pmatrix}. \end{aligned}$$

Propiedades de la Distribuciones de Vectores Normales.

Sea $\mathbf{X} = (X_1, X_2, \dots, X_n)'$ y $A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$,

1. $E(A\mathbf{X}) = AE(\mathbf{X})$.
2. $Var(A\mathbf{X}) = AVar(\mathbf{X})A'$.

2.12. Momento de segundo orden.

La covarianza entre dos variables aleatorias Y y Z con momento de segundo orden finito, está definido como:

$$Cov(Y, Z) = E[(Y - E(Y))(Z - E(Z))],$$

ésto cuantitativamente puede ser positivo, negativo o cero, la covarianza matricial de un vector X está dada como sigue:

$$\Sigma = (\sigma_{ij}) = E[(X - E(X))(X - E(X))'],$$

para $i, j = 1, \dots, p$. Un elemento típico de Σ es $\sigma_{ij} = E[(X_i - E(X_i))(X_j - E(X_j))']$, $i, j = 1, \dots, p$ cuando $j = i$ los elementos están ubicados a lo largo de la diagonal de Σ y es llamada la varianza X .

Recordemos que:

$$Var(X_i) = E(X_i - E(X_i))^2,$$

cuando $i \neq j$, σ_{ij} es la covarianza de X_i y X_j . El coeficiente de correlación entre dos variables aleatorias escalares Y y Z con momento de segundo orden finito está definido como:

$$\rho = corr(Y, Z) = \frac{Cov(Y, Z)}{[(VarY)(VarZ)]^{\frac{1}{2}}}.$$

Esta es una medida de causa y efecto asociada con Y y Z . En general, $-1 \leq \rho \leq 1$, aunque en algunos casos, ρ es restringido a un intervalo más pequeño.

Una matriz de correlación $R = (\rho_{ij})$, $i, j = 1, \dots, p$; es útil para estudiar todas las asociaciones entre las componentes de un vector de variables simultáneamente. La matriz de correlación es calculada en muchos modelos usados en análisis de datos multivariados ya que la matriz R provee frecuentemente un rápido entendimiento dentro de muchas relaciones insospechadas.

Los elementos de la diagonal, ρ_{ij} , de la matriz de correlación deberían ser todos uno, y los elementos fuera de la diagonal dados por:

$$\rho_{ij} = corr(x_i, x_j) = \frac{Cov(x_i, x_j)}{[(Var(x_i))(Var(x_j))]^{\frac{1}{2}}}, \quad \text{con } i \neq j,$$

además, los ρ_{ij} deberían también satisfacer siempre la inecuación $-1 \leq \rho_{ij} \leq 1$ para $i, j = 1, \dots, p$.

2.13. La distribución Normal Mutivariada.

La distribución más importante y fundamental de análisis multivariado aplicado es la distribución normal multivariada. Su mayor papel se debe al hecho que estandariza sumas de vector de datos independientes que siguen distribuciones multivariadas arbitrarias, en muestras grandes, para seguir la distribución normal multivariada.

Este resultado es una generalización del teorema del límite central univariado a las dimensiones más altas. El papel central que juega la distribución también es atribuible, y no de manera pequeña, al hecho que pueden obtenerse a menudo los resultados para la distribución Normal y no para otras distribuciones que, por muchas razones, no son atractivas como la Normal.

2.13.1. Densidad general.

Sea $X : p \times 1$ un vector aleatorio con función de densidad $f(x)$. Se dirá que X sigue una distribución Normal (p-variante) multivariada no singular con vector de media $\theta : p \times 1$ y matriz de covarianza $\Sigma : p \times p$ si

$$f(x) = \frac{1}{(2\pi)^{\frac{p}{2}} |\Sigma|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} (x - \theta)' \Sigma^{-1} (x - \theta) \right\},$$

para $\Sigma > 0$. Si $|\Sigma| = 0$, la distribución de x es llamada singular o Normal degenerada y la densidad no existe.

2.13.2. Media y Covarianza muestral.

Sean x_1, \dots, x_N vectores observados $p \times 1$ independientes con una distribución $N(\theta, \Sigma)$.

Denotemos el vector de la *media muestral* por \bar{x} y definámoslo como:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$$

y denotemos la matriz de *covarianza muestral* como V , la cual estará determinada por:

$$V = \sum_{i=1}^N [(x_i - \bar{x})(x_i - \bar{x})'].$$

CAPÍTULO 3

CAMPO ALEATORIO.

Un *campo aleatorio* es una colección de variables aleatorias $\{Z(\mathbf{s}), \mathbf{s} \in D\}$ indexadas por los elementos de $D \subset \mathbb{R}^l, l \geq 1$, mientras que una *realización del campo aleatorio* es una función $Z(., \omega_0) : D \rightarrow \mathbb{R}$, donde ω_0 es un resultado particular de un experimento aleatorio; una realización se denota por $\{z(\mathbf{s}), \mathbf{s} \in D\}$. En general la teoría sobre campos aleatorios aparece en Christensen (1991), Cressie (1993) y Ripley (1981). En este trabajo se supondrá que los campos aleatorios satisfacen que $E\{Z^2(\mathbf{s})\} < \infty, \forall \mathbf{s} \in D$.

Los datos espaciales comúnmente disponibles en las ciencias de la tierra constan de una muestra de n observaciones $\mathbf{z} = (z(\mathbf{s}_1), \dots, z(\mathbf{s}_n))'$, provenientes de una realización del campo aleatorio de interés, donde $\mathbf{s}_1, \dots, \mathbf{s}_n$ son localizaciones muestrales conocidas en D .

Las funciones descriptivas básicas asociadas a un campo aleatorio son la función de media:

$$\mu(\mathbf{s}) = E\{Z(\mathbf{s})\}; \quad \mathbf{s} \in D$$

la cual puede ser una función arbitraria y la función de covarianza:

$$\begin{aligned} C(\mathbf{s}, \mathbf{u}) &= Cov\{Z(\mathbf{s}), Z(\mathbf{u})\} \\ &= E\{(Z(\mathbf{s}) - \mu(\mathbf{s}))(Z(\mathbf{u}) - \mu(\mathbf{u}))\}; \quad \mathbf{s}, \mathbf{u} \in D \end{aligned}$$

la cual debe ser positiva semidefinida. Esto significa que $\forall m \in \mathbb{N}, \mathbf{s}_1, \dots, \mathbf{s}_m \in D$, y $a_1, \dots, a_m \in \mathbb{R}$, se verifica

$$\sum_{i=1}^m \sum_{j=1}^m a_i a_j C(\mathbf{s}_i, \mathbf{s}_j) \geq 0.$$

La función $\mu(\mathbf{s})$ describe la tendencia global de las realizaciones del campo aleatorio, mientras que $C(\mathbf{s}, \mathbf{u})$ describe asociaciones (lineales) locales entre $Z(\mathbf{s})$ y $Z(\mathbf{u})$. Otra

medida de asociación local es la función de semivariograma:

$$\gamma(\mathbf{s}, \mathbf{u}) = \frac{1}{2} [Var\{Z(\mathbf{s}) - Z(\mathbf{u})\}],$$

la cual provee información similar a $C(\mathbf{s}, \mathbf{u})$, y está relacionada a ésta por

$$\gamma(\mathbf{s}, \mathbf{u}) = \frac{1}{2} [C(\mathbf{s}, \mathbf{s}) + C(\mathbf{u}, \mathbf{u}) - 2C(\mathbf{s}, \mathbf{u})]. \quad (3.1)$$

Los datos espaciales usualmente provienen de una única realización, de modo que no se disponen de réplicas. Por ello es necesario hacer algunas suposiciones de simetría o estacionaridad para poder realizar inferencia sobre el modelo. Las suposiciones más usuales sobre el modelo se dan a continuación:

- $Z(\cdot)$ se dice *estrictamente estacionario* si $\forall m \in \mathbb{N}$, $\mathbf{s}_1, \dots, \mathbf{s}_m \in D$, y $\mathbf{h} \in \mathbb{R}^l$ se tiene

$$(Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_m)) \stackrel{d}{=} (Z(\mathbf{s}_1 + \mathbf{h}), \dots, Z(\mathbf{s}_m + \mathbf{h})).$$

- $Z(\cdot)$ se dice *débilmente estacionario* o *estacionario* de segundo orden si

$$\mu(\mathbf{s}) = \mu \quad \text{y} \quad C(\mathbf{s}, \mathbf{u}) = \tilde{C}(\mathbf{s} - \mathbf{u}).$$

- $Z(\cdot)$ se dice *intrínsecamente estacionario* si

$$\mu(\mathbf{s}) = \mu \quad \text{y} \quad \gamma(\mathbf{s}, \mathbf{u}) = \tilde{\gamma}(\mathbf{s} - \mathbf{u}).$$

- Una función de covarianza se dice *isotrópica* si

$$C(\mathbf{s}, \mathbf{u}) = \tilde{C}(\|\mathbf{s} - \mathbf{u}\|),$$

donde $\|\mathbf{s} - \mathbf{u}\|$ denota la distancia Euclideana entre \mathbf{s} y \mathbf{u} .

En general, un proceso débilmente estacionario es también intrínsecamente estacionario, pero el recíproco no es cierto. En efecto:

Supongamos que el proceso es débilmente estacionario.

Tenemos que $\mu(\mathbf{s})$ es igual tanto para el proceso débilmente estacionario como para el proceso intrínsecamente estacionario.

Notemos que $\gamma(s, u) = \tilde{\gamma}(s, u)$:

$$\begin{aligned}\gamma(\mathbf{s}, \mathbf{u}) &= \frac{1}{2}[C(\mathbf{s}, \mathbf{s}) + C(\mathbf{u}, \mathbf{u}) - 2C(\mathbf{s}, \mathbf{u})] \\ &= \frac{1}{2}[\tilde{C}(\mathbf{s}-\mathbf{s}) + \tilde{C}(\mathbf{u}, \mathbf{u}) - 2\tilde{C}(\mathbf{s}-\mathbf{u})] \\ &= \tilde{\gamma}(\mathbf{s}, \mathbf{u}).\end{aligned}$$

Por otro lado, sea $\mathbf{h}=(\mathbf{s}, \mathbf{u})$, si suponemos que el proceso es débilmente estacionario, entonces usando (3.1) observamos que

$$\begin{aligned}\gamma(\mathbf{h}) &= \gamma(\mathbf{s}, \mathbf{u}) \\ &= \frac{1}{2}[C(\mathbf{s}, \mathbf{s}) + C(\mathbf{u}, \mathbf{u}) - 2C(\mathbf{s}, \mathbf{u})] \\ &= \frac{1}{2}[\tilde{C}(\mathbf{s}-\mathbf{s}) + \tilde{C}(\mathbf{u}-\mathbf{u}) - 2\tilde{C}(\mathbf{s}-\mathbf{u})] \\ &= \frac{1}{2}[\tilde{C}(\mathbf{0}) + \tilde{C}(\mathbf{0}) - 2\tilde{C}(\mathbf{s}-\mathbf{u})] \\ &= \frac{1}{2}[2\tilde{C}(\mathbf{0}) - 2\tilde{C}(\mathbf{s}-\mathbf{u})] \\ &= \tilde{C}(\mathbf{0}) - \tilde{C}(\mathbf{s}-\mathbf{u}) \\ &= C(\mathbf{0}, \mathbf{0}) - C(\mathbf{s}, \mathbf{u}) \\ &= C(\mathbf{0}) - C(\mathbf{h}).\end{aligned}$$

La descripción probabilística completa de un campo aleatorio requiere en general la especificación de la familia de distribuciones finito-dimensionales formada por todas las distribuciones conjuntas de la forma

$$F_{\mathbf{s}_1, \dots, \mathbf{s}_m}(x_1, \dots, x_m) = P\{Z(\mathbf{s}_1) \leq x_1, \dots, Z(\mathbf{s}_m) \leq x_m\},$$

$\forall m \in \mathbb{N}$ y $\mathbf{s}_1, \dots, \mathbf{s}_m \in D$.

Un campo aleatorio se dice *Gaussiano* si todos los miembros de la familia de distribuciones finito-dimensionales son normales multivariadas. En este caso, el campo aleatorio está descrito completamente por las funciones de media $\mu(\mathbf{s})$ y covarianza $C(\mathbf{s}, \mathbf{u})$. En este caso, las nociones de estacionaridad débil y estricta coinciden.

Algunos ejemplos (paramétricos) de función de media son:

$$\beta_1 \quad (\text{constante}),$$

$$\beta_1 + \beta_2 x + \beta_3 y,$$

$$\beta_1 + \beta_2 x + \beta_3 y + \beta_4 x^2 + \beta_5 xy + \beta_6 y^2,$$

$$\beta_1 + \beta_2 \|\mathbf{s} - \mathbf{s}^*\|; \quad \mathbf{s}^* \text{ localización fija,}$$

$$\beta_1 + \beta_2 X(\mathbf{s}); \quad X(\cdot) \text{ proceso auxiliar.}$$

Algunos ejemplos (paramétricos) de funciones de covarianza isotrópicas, con $d = \|\mathbf{s} - \mathbf{u}\|$ son :

Esférico

$$C^E(d) = \sigma^2 \begin{cases} 1 - \frac{3}{2} \left(\frac{d}{\theta_1}\right) + \frac{1}{2} \left(\frac{d}{\theta_1}\right)^3, & \text{si } 0 \leq d \leq \theta_1 \\ 0, & \text{si } d > \theta_1, \end{cases} \quad \sigma^2, \theta_1 > 0$$

Potencia Exponencial

$$C^{PE}(d) = \sigma^2 \exp\left(-\left(\frac{d}{\theta_1}\right)^{\theta_2}\right); \quad \sigma^2, \theta_1 > 0, \theta_2 \in (0, 2].$$

Cuadrático Racional

$$C^{RQ}(d) = \sigma^2 \left(1 + \left(\frac{d}{\theta_1}\right)^2\right)^{-\theta_2}; \quad \sigma^2, \theta_1, \theta_2 > 0.$$

Matérn

$$C^M(d) = \frac{1}{2^{\theta_2-1}\Gamma(\theta_2)} \left(\frac{d}{\theta_1}\right)^{\theta_2} k_{\theta_2}\left(\frac{d}{\theta_1}\right); \quad \theta_1, \theta_2 > 0,$$

donde $k_{\theta_2}(\cdot)$ es la función de Bessel de segundo tipo y orden θ_2 .

Sea $\{Z(\mathbf{s}) : \mathbf{s} \in D\}$, $D \subset \mathbb{R}^l$, y $l \geq 1$, un campo aleatorio Gaussiano, con funciones de media y covarianza dadas por

$$\mu(\mathbf{s}) = \sum_{j=1}^p \beta_j f_j(\mathbf{s}) = \boldsymbol{\beta}' \mathbf{f}(\mathbf{s}),$$

$$C(\mathbf{s}, \mathbf{u}) = \sigma^2 K_\vartheta(\mathbf{s}, \mathbf{u}); \quad \mathbf{s}, \mathbf{u} \in D,$$

donde $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)' \in \mathbb{R}^p$ son parámetros de regresión, $\mathbf{f}(\mathbf{s}) = (f_1(\mathbf{s}), \dots, f_p(\mathbf{s}))'$ son covariables conocidas dependientes de la localización \mathbf{s} , $\sigma > 0$ es un parámetro de escala, $K_\vartheta(\mathbf{s}, \mathbf{u})$ es una función positiva definida en $D \times D$ y $\boldsymbol{\vartheta} \in \Theta \subseteq \mathbb{R}^q$ son parámetros que controlan características geométricas del campo aleatorio. Frecuentemente se tendrá que $\sigma^2 = \text{var}\{Z(\mathbf{s})\}$ y $K_\vartheta(\mathbf{s}, \mathbf{u}) = \text{corr}\{Z(\mathbf{s}), Z(\mathbf{u})\}$, pero esto no tiene que ser necesariamente así. En la siguiente sección denotaremos por $\boldsymbol{\eta} = (\boldsymbol{\beta}, \boldsymbol{\gamma})'$ los parámetros del modelo, donde $\boldsymbol{\beta}$ son parámetros de esperanza y $\boldsymbol{\gamma} = (\sigma^2, \boldsymbol{\vartheta})$ son parámetros de covarianza.

Supongamos que los datos constan de una muestra (aleatoria) de n observaciones $\mathbf{Z} = (Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n))'$, provenientes de una realización del campo aleatorio de interés, donde $\mathbf{S}_n = \{\mathbf{s}_1, \dots, \mathbf{s}_n\}$ son localizaciones muestrales conocidas en D . Bajo este modelo se tiene entonces que $E\{\mathbf{Z}\} = X\boldsymbol{\beta}$, $\text{var}(\mathbf{Z}) = \boldsymbol{\Sigma}_\gamma = \sigma^2 \boldsymbol{\Sigma}_\vartheta$ y

$$\mathbf{Z} \sim N_n(X\boldsymbol{\beta}, \sigma^2 \boldsymbol{\Sigma}_\vartheta), \quad (3.2)$$

donde X es una matriz $n \times p$ conocida de rango completo ($p < n$), definida por $X_{ij} = f_j(\mathbf{s}_i)$, y $\boldsymbol{\Sigma}_\vartheta$ es una matriz $n \times n$ positiva definida, con $\Sigma_{\vartheta,ij} = K_\vartheta(\mathbf{s}_i, \mathbf{s}_j)$.

3.1. Estimación De Variograma.

Considerando los principales conceptos de covarianzas y variogramas espaciales, y el escenario general que consiste de un proceso $\{Z(\mathbf{s}), \mathbf{s} \in D\}$ observado en un número finito de puntos $\mathbf{s}_1, \dots, \mathbf{s}_N$.

El estimador más simple es: *El Estimador por el Método de los Momentos (MoM)*. En el caso en que los puntos muestrales $\mathbf{s}_1, \dots, \mathbf{s}_N$ están sobre una malla regular, éste se define como:

$$2\hat{\gamma}(h) = \frac{1}{|N(h)|} \sum_{(\mathbf{s}_i, \mathbf{s}_j) \in N(h)} \{Z(\mathbf{s}_i) - Z(\mathbf{s}_j)\}^2. \quad (3.3)$$

Donde $N(h)$ denota los pares $(\mathbf{s}_i, \mathbf{s}_j)$ para el cual $\mathbf{s}_i - \mathbf{s}_j = h$ y $|N(h)|$ denota el cardinal de $N(h)$.

Si los puntos no están sobre una malla regular, cambiamos la definición de $N(h)$ a:

$$N(h) = \{(\mathbf{s}_i, \mathbf{s}_j) : \mathbf{s}_i - \mathbf{s}_j \in T(h)\},$$

siendo $T(h)$ alguna vecindad pequeña o región tolerante alrededor de h . Una objeción del MoM es, como en muchos métodos basados en promedios muestrales, la no-robustez y las distribuciones muestrales sesgadas.

3.2. Máxima Verosimilitud.

La función de verosimilitud provee, en general, una poderosa herramienta para cuantificar la información que los datos observados poseen acerca de los parámetros desconocidos. En el marco del modelo y los datos descritos antes, la función de log-verosimilitud del vector de parámetros

$$\eta = (\beta, \sigma^2, \vartheta) \in \Omega = \mathbb{R}^p \times (0, \infty) \times \Theta,$$

basada en los datos observados $z = (z_1, \dots, z_n)'$, $z_i = z(s_i)$, viene dada, salvo por una constante aditiva, por

$$l(\eta; z) = -\frac{n}{2} \log(\sigma^2) - \frac{1}{2} \log(|\Sigma_\vartheta|) - \frac{1}{2\sigma^2} (\mathbf{z} - X\beta)' \Sigma_\vartheta^{-1} (\mathbf{z} - X\beta).$$

El estimador de máxima verosimilitud de η , suponiendo que éste existe, es el vector $\hat{\eta}^{ml} = (\hat{\beta}, \hat{\sigma}^2, \hat{\vartheta}) \in \Omega$ que maximiza $l(\eta; z)$ como función de η , para z fijo, es decir, $\hat{\eta}^{ml} = \arg \max_{\eta \in \Omega} l(\eta; z)$.

3.3. Máxima Verosimilitud Restringida.

El método de máxima verosimilitud restringida (REML) fue originalmente propuesto por Patterson y Thompson (1971) para la estimación de componentes de varianza en modelos lineales mixtos. Sin embargo, el método es igualmente aplicable para la estimación de parámetros en campos aleatorios Gaussianos. La idea del método REML consiste en aplicar el método de máxima verosimilitud usando una verosimilitud para los parámetros de covariancia γ , en nuestro caso $\gamma = (\sigma^2, \vartheta)$, basada en ciertas

combinaciones lineales de los datos llamadas contrastes, en vez de aquella basada en los datos mismos.

Una combinación lineal de los datos $\mathbf{a}'\mathbf{Z}$, con $\mathbf{a} \neq \mathbf{0}$, se denomina un *contraste* si

$$E_{\eta}\{\mathbf{a}'\mathbf{Z}\} = \mathbf{0}, \quad \forall \eta \in \Omega.$$

Sea $T_2 = (\mathbf{t}_1 \mid \dots \mid \mathbf{t}_{n-p})$ una matriz $n \times (n-p)$, cuyas columnas \mathbf{t}_i son vectores linealmente independientes en \mathbb{R}^n , cada uno de los cuales define un contraste para \mathbf{Z} , y por tanto $T_2'X = \mathbf{0}$. Ejemplos de tales matrices son las formadas por cualesquiera $n-p$ filas de la matriz de proyección $P = I_n - X(X'X)^{-1}X'$. En particular, $T_2 = (I_{n-p} \mid \mathbf{0})P$. Ahora, sea \mathbf{W} el vector aleatorio $(n-p) \times 1$ definido por $\mathbf{W} = T_2'\mathbf{Z}$. Entonces de (3.2) se tiene que

$$\mathbf{W} \sim N_{n-p}(\mathbf{0}, \sigma^2 T_2' \Sigma_{\theta} T_2), \quad (3.4)$$

de modo que la distribución de \mathbf{W} depende de γ pero no de β . De (3.4), la función de log-verosimilitud de γ basada en \mathbf{W} viene dada por

$$l(\gamma; \mathbf{w}) = -\frac{1}{2} \left[(n-p) \log(\sigma^2) + \log(|T_2' \Sigma_{\theta} T_2|) + \frac{1}{\sigma^2} \mathbf{w}' (T_2' \Sigma_{\theta} T_2)^{-1} \mathbf{w} \right]. \quad (3.5)$$

El estimador de máxima verosimilitud restringida (REML) de $\gamma = (\sigma^2, \boldsymbol{\theta})$, suponiendo que éste existe, se define como el vector $\tilde{\gamma}^{rl} = (\tilde{\sigma}^2, \tilde{\boldsymbol{\theta}}) \in \Gamma$ que maximiza $l(\gamma; \mathbf{w})$ como función de γ para \mathbf{w} fijo. El siguiente resultado muestra que $\tilde{\gamma}^{rl}$ no depende de la matriz de contrastes T_2 que se use para construir \mathbf{W} .

3.4. Mínimos Cuadrados.

Supongamos que tenemos la estimación del semivariograma $\hat{\gamma}(h)$ en un conjunto finito de valores de h y deseamos ajustar un modelo especificado por la función paramétrica $\gamma(h; \theta)$ en términos de un vector de parámetros finito θ .

Existen tres versiones bien usadas de estimadores de mínimos cuadrados:

- Mínimos Cuadrados Ordinario (*OLS*): Seleccionamos θ que minimice

$$\{\hat{\gamma} - \gamma(\theta)\}^T \{\hat{\gamma} - \gamma(\theta)\}.$$

- Mínimos Cuadrados Generalizados (*GLS*): Seleccionamos θ que minimice

$$\{\hat{\gamma} - \gamma(\theta)\}^T V(\theta)^{-1} \{\hat{\gamma} - \gamma(\theta)\}.$$

Aquí $V(\theta)$ denota la matriz de covarianza de $\hat{\gamma}$ la cual, debido a que el problema es no-lineal, depende del parámetro desconocido θ .

- Mínimos Cuadrados Pesados (*WLS*): Seleccionamos θ que minimice

$$\{\hat{\gamma} - \gamma(\theta)\}^T W(\theta)^{-1} \{\hat{\gamma} - \gamma(\theta)\}.$$

$W(\theta)$ es una matriz diagonal cuyas entradas diagonales son las mismas que las de $V(\theta)$. Así *WLS* toma en cuenta las varianzas de $\hat{\gamma}$ pero no las covarianzas, mientras *GLS* toma en cuenta ambas.

En general, se espera que los tres estimadores *OLS*, *WLS*, *GLS* sean en orden creciente de eficientes pero en orden decreciente de convenientes al usarlos. Nótese, en particular, que *OLS* es implementable de manera inmediata por un procedimiento de mínimos cuadrados no lineal, mientras que *WLS* y *GLS* requieren especificar la matrices $W(\theta)$ y $V(\theta)$.

Para un proceso Gaussiano, sin embargo, se tienen las expresiones

$$\text{var}[\{Z(s+h) - Z(s)\}^2] = 2\{2\gamma(h)\}, \quad (3.6)$$

$$\begin{aligned} & \text{corr}[\{Z(s_1+h_1) - Z(s_1)\}^2, \{Z(s_2+h_2) - Z(s_2)\}^2] \\ &= \frac{\{\gamma(s_1 - s_2 + h_1) + \gamma(s_1 - s_2 + h_2) - \gamma(s_1 - s_2 + h_1 - h_2) - \gamma(s_1 - s_2)\}^2}{2\gamma(h_1) \times 2\gamma(h_2)} \end{aligned} \quad (3.7)$$

las cuales pueden ser usadas para evaluar las matrices $W(\theta)$ y $V(\theta)$. Así *GLS* es posible de obtener en principio, pero complicado de implementar. Por ejemplo, no existe garantía que el resultado de la minimización del problema tenga una única solución.

CAPÍTULO 4

EL METODO INFERENCIAL DE MÁXIMA VEROSIMILITUD

Consideremos el modelo universal (modelo lineal en estadística):

$$\mathbf{Z} = \mu + \varepsilon, \quad (4.1)$$

donde:

- $\mu = X \beta$: ($n \times 1$) es llamado vector de medias o vector de tendencia,
- X : $n \times p$, es llamada matriz de funciones de tendencia básicas (monomios en la tendencia polinomial),
- β : $p \times 1$, es llamado vector de coeficientes de tendencia desconocidas,
- ε : $n \times 1$, es llamado vector de residuos, el cual posee media cero y varianza estacionaria de segundo orden V ($E\{\varepsilon\} = 0$ y $E\{\varepsilon \varepsilon'\} = V$), y la distribución Gaussiana multivariada (mGd) de Z esta dada por

$$p(Z; \theta) = (2\pi)^{-\frac{n}{2}} |V|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(Z - \mu)'V^{-1}(Z - \mu)\right\}$$

donde θ : $m \times 1$ es el vector de parámetros que definen el vector de media y V : $n \times n$ es la matriz de covarianza. Dada las observaciones de Z obtenemos la funcion de verosimilitud:

$$L(\theta; z) = (2\pi)^{-\frac{n}{2}} |V|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(z - \mu)'V^{-1}(z - \mu)\right\},$$

aplicando logaritmo obtenemos:

$$\ln[L(\theta; z)] = -\frac{n}{2} \ln(2\pi) - \frac{1}{2} \ln |V| - \frac{1}{2}(z - \mu)'V^{-1}(z - \mu).$$

Donde el negativo del logaritmo de la verosimilitud (NLLF) es:

$$-\ln[L(\theta; z)] = \frac{n}{2} \ln(2\pi) + \frac{1}{2} \ln |V| + \frac{1}{2} (z - \mu)' V^{-1} (z - \mu),$$

como se puede observar en Kitanidis(1983).

No obstante, es útil introducir la factorización (como en Mardia y Marshall, 1984):

$$V = \sigma^2 Q, \quad (4.2)$$

donde tenemos que σ^2 es la varianza y Q es la matriz de correlación.

Con esta factorización la varianza es estimada de forma independiente de los otros parámetros de covarianza(efecto nugget, sill, entre otros).

Con la factorización (4.2) y sabiendo que $\mu = X\beta$ el NLLF se convierte en:

$$\begin{aligned} L(\beta, \sigma^2, \theta; z) &= \frac{n}{2} \ln(2\pi) + \frac{1}{2} \ln |\sigma^2 Q| + \frac{1}{2} (z - X\beta)' (\sigma^2 Q)^{-1} (z - X\beta) \\ &= \frac{n}{2} \ln(2\pi) + \frac{1}{2} \ln [(\sigma^2)^n |Q|] + \frac{1}{2} (z - X\beta)' (\sigma^{-2} Q^{-1}) (z - X\beta) \\ &= \frac{n}{2} \ln(2\pi) + \frac{1}{2} \ln [(\sigma^2)^n + \ln |Q|] + \frac{1}{2} (z - X\beta)' (\sigma^{-2} Q^{-1}) (z - X\beta) \\ &= \frac{n}{2} \ln(2\pi) + \frac{1}{2} (2n \ln(\sigma) + \ln |Q|) + \frac{1}{2\sigma^2} (z - X\beta)' Q^{-1} (z - X\beta) \\ &= \frac{n}{2} \ln(2\pi) + n \ln(\sigma) + \frac{1}{2} \ln |Q| + \frac{1}{2\sigma^2} (z - X\beta)' Q^{-1} (z - X\beta). \end{aligned}$$

Así concluimos que:

$$L(\beta, \sigma^2, \theta; z) = \frac{n}{2} \ln(2\pi) + n \ln(\sigma) + \frac{1}{2} \ln |Q| + \frac{1}{2\sigma^2} (z - X\beta)' Q^{-1} (z - X\beta). \quad (4.3)$$

En la ecuación (4.3), θ representa los parámetros de covarianza, pero no incluye la varianza. Todos los parámetros serán estimados por ML, igualando a cero las derivadas de (4.3) con respecto a los diferentes parámetros.

4.1. Estimador para β .

A partir de la ecuación (4.3), nótese que

$$\begin{aligned}
L(\beta, \sigma^2, \theta; z) &= \frac{n}{2} \ln(2\pi) + n \ln(\sigma) + \frac{1}{2} \ln |Q| + \frac{1}{2\sigma^2} (z - X\beta)' Q^{-1} (z - X\beta) \\
&= \frac{n}{2} \ln(2\pi) + n \ln(\sigma) + \frac{1}{2} \ln |Q| + \frac{1}{2\sigma^2} (z' - \beta' X') (Q^{-1} z - Q^{-1} X\beta) \\
&= \frac{n}{2} \ln(2\pi) + n \ln(\sigma) + \frac{1}{2} \ln |Q| + \frac{1}{2\sigma^2} (z' Q^{-1} z - z' Q^{-1} X\beta - \beta' X' Q^{-1} z \\
&\quad + \beta' X' Q^{-1} X\beta) \\
&= \frac{n}{2} \ln(2\pi) + n \ln(\sigma) + \frac{1}{2} \ln |Q| + \frac{1}{2\sigma^2} (z' Q^{-1} z - z' Q^{-1} X\beta - z' Q^{-1} X\beta \\
&\quad + \beta' X' Q^{-1} X\beta) \\
&= \frac{n}{2} \ln(2\pi) + n \ln(\sigma) + \frac{1}{2} \ln |Q| + \frac{1}{2\sigma^2} (z' Q^{-1} z - 2z' Q^{-1} X\beta + \beta' X' Q^{-1} X\beta).
\end{aligned}$$

Así tenemos que:

$$L(\beta, \sigma^2, \theta; z) = \frac{n}{2} \ln(2\pi) + n \ln(\sigma) + \frac{1}{2} \ln |Q| + \frac{1}{2\sigma^2} (z' Q^{-1} z - 2z' Q^{-1} X\beta + \beta' X' Q^{-1} X\beta). \quad (4.4)$$

Derivando (4.4) respecto a β obtenemos:

$$\begin{aligned}
\frac{\partial [L(\beta, \sigma^2, \theta; z)]}{\partial \beta} &= \frac{\partial}{\partial \beta} \left\{ \frac{1}{2\sigma^2} [\beta' (X' Q^{-1} X) \beta - (2z' Q^{-1} X) \beta] \right\} \\
&= \frac{1}{2\sigma^2} [(2X' Q^{-1} X) \beta - 2(z' Q^{-1} X)'] \\
&= \frac{1}{\sigma^2} [(X' Q^{-1} X) \beta - X' Q^{-1} z].
\end{aligned}$$

Por lo tanto,

$$\frac{\partial [L(\beta, \sigma^2, \theta; z)]}{\partial \beta} = \frac{1}{\sigma^2} [X' Q^{-1} X \beta - X' Q^{-1} z]. \quad (4.5)$$

Igualando a cero la ecuación (4.5) tenemos:

$$\begin{aligned}
\frac{1}{\sigma^2} (X' Q^{-1} X \hat{\beta} - X' Q^{-1} z) = 0 &\Rightarrow X' Q^{-1} X \hat{\beta} - X' Q^{-1} z = 0 \\
&\Rightarrow X' Q^{-1} X \hat{\beta} = X' Q^{-1} z \\
&\Rightarrow \hat{\beta} = (X' Q^{-1} X)^{-1} X' Q^{-1} z.
\end{aligned}$$

Luego, el estimador de Máxima Verosimilitud para β es:

$$\hat{\beta} = (X'Q^{-1}X)^{-1}X'Q^{-1}Z. \quad (4.6)$$

4.2. Estimador para σ^2 .

Derivando la ecuación (4.3) respecto a σ^2 obtenemos:

$$\begin{aligned} \frac{\partial[L(\beta, \sigma^2, \theta; z)]}{\partial\sigma^2} &= \frac{\partial}{\partial\sigma^2} \left[\frac{n}{2} \ln(2\pi) + n \ln(\sigma) + \frac{1}{2} \ln |Q| + \frac{1}{2\sigma^2} (z - X\beta)'Q^{-1}(z - X\beta) \right] \\ &= \frac{\partial}{\partial\sigma^2} \left[\frac{n}{2} \ln(2\pi) + n \ln(\sqrt{\sigma^2}) + \frac{1}{2} \ln |Q| + \frac{1}{2\sigma^2} (z - X\beta)'Q^{-1}(z - X\beta) \right] \\ &= \frac{\partial}{\partial\sigma^2} \left[\frac{n}{2} \ln(2\pi) + \frac{n}{2} \ln(\sigma^2) + \frac{1}{2} \ln |Q| + \frac{1}{2\sigma^2} (z - X\beta)'Q^{-1}(z - X\beta) \right] \\ &= \frac{n}{2\sigma^2} - \frac{1}{2\sigma^4} (z - X\beta)'Q^{-1}(z - X\beta). \end{aligned}$$

Así,

$$\frac{\partial[L(\beta, \sigma^2, \theta; z)]}{\partial\sigma^2} = \frac{n}{2\sigma^2} - \frac{1}{2\sigma^4} (z - X\beta)'Q^{-1}(z - X\beta). \quad (4.7)$$

Igualando a cero la ecuación (4.7) tenemos:

$$\begin{aligned} \frac{n}{2\hat{\sigma}^2} - \frac{1}{2\hat{\sigma}^4} (z - X\hat{\beta})'Q^{-1}(z - X\hat{\beta}) = 0 &\Rightarrow \frac{n}{2\hat{\sigma}^2} = \frac{1}{2\hat{\sigma}^4} (z - X\hat{\beta})'Q^{-1}(z - X\hat{\beta}) \\ &\Rightarrow \frac{2\hat{\sigma}^4}{2\hat{\sigma}^2} = \frac{1}{n} (z - X\hat{\beta})'Q^{-1}(z - X\hat{\beta}) \\ &\Rightarrow \hat{\sigma}^2 = \frac{1}{n} (z - X\hat{\beta})'Q^{-1}(z - X\hat{\beta}). \end{aligned}$$

Por lo tanto, el estimador de máxima verosimilitud de σ^2 es:

$$\hat{\sigma}^2 = \frac{1}{n} (z - X\hat{\beta})'Q^{-1}(z - X\hat{\beta}). \quad (4.8)$$

4.3. Estimador para los Parámetros de Covarianza.

Sustituyendo (4.6) y (4.8) en (4.3), observamos que NLLF es una función solamente de los parámetros de covarianza y viene dada por:

$$\begin{aligned}
L'(\hat{\beta}, \hat{\sigma}^2, \theta; z) &= \frac{n}{2} \ln(2\pi) + n \ln \left[\sqrt{\frac{1}{n} (z - X\hat{\beta})' Q^{-1} (z - X\hat{\beta})} \right] + \frac{1}{2} \ln |Q| \\
&+ \frac{1}{2 \frac{1}{n} (z - X\hat{\beta})' Q^{-1} (z - X\hat{\beta})} (z - X\hat{\beta})' Q^{-1} (z - X\hat{\beta}) \\
&= \frac{n}{2} \ln(2\pi) + \frac{n}{2} \ln \left[\frac{1}{n} (z - X\hat{\beta})' Q^{-1} (z - X\hat{\beta}) \right] + \frac{1}{2} \ln |Q| + \frac{n}{2} \\
&= \frac{n}{2} \ln(2\pi) + \frac{n}{2} \ln \left(\frac{1}{n} \right) + \frac{n}{2} \ln [(z - X\hat{\beta})' Q^{-1} (z - X\hat{\beta})] + \frac{1}{2} \ln |Q| + \frac{n}{2} \\
&= \frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(n) + \frac{n}{2} \ln [(z - X\hat{\beta})' Q^{-1} (z - X\hat{\beta})] + \frac{1}{2} \ln |Q| + \frac{n}{2} \\
&= \frac{n}{2} [\ln(2\pi) + 1 - \ln(n)] + \frac{1}{2} \ln |Q| + \frac{n}{2} \ln [(z - X\hat{\beta})' Q^{-1} (z - X\hat{\beta})].
\end{aligned}$$

Por lo tanto, tenemos que:

$$L'(\hat{\beta}, \hat{\sigma}^2, \theta; Z) = \frac{n}{2} [\ln(2\pi) + 1 - \ln(n)] + \frac{1}{2} \ln |Q| + \frac{n}{2} \ln [(Z - X\hat{\beta})' Q^{-1} (Z - X\hat{\beta})]. \quad (4.9)$$

Los estimadores ML de los parámetros de covarianza θ se obtienen como los valores que minimizan la ecuación (4.9).

La minimización de (4.9) debe ser realizada numéricamente y pueden ser utilizados diferentes procedimientos. Por ejemplo, en el programa de dominio MLREML (Pardo-Iguzquiza, 1997) cinco diferentes procedimientos han sido implementados: búsqueda directa, métodos de puntuación (Kitanidis y Lane, 1985), búsqueda axial, métodos simplex (Nelder y Mead, 1965) y el recorrido simulado.

4.4. Caracterización del Modelo.

De (4.6), (4.8) y (4.9) es claro como los parámetros de interés son estimados directamente una vez que un modelo ha sido adoptado, por ejemplo, un modelo esférico con efecto nugget y rango isotrópico. Con el fin de evitar el procedimiento engorroso de comprobar cada posible modelo y la combinación de los parámetros, el variograma experimental obtenido por el método clásico de los momentos es utilizado con el

procedimiento ML. El modelo propuesto por el variograma experimental es el único usado por el procedimiento ML. Sin embargo, si el variograma experimental bruto es ruidoso varios modelos podrían encajar igual de bien, ML será usado para la selección del modelo como se explica a continuación.

ML puede ser usado en la selección del modelo en dos situaciones diferentes: Primero donde los modelos alternativos tienen el mismo número de parámetros y segundo donde ellos aplican un número diferente. Un ejemplo del primer caso es cuando el modelo podría ser esférico o exponencial; los parámetros de ambos modelos (en igual número) son estimados y el modelo con la menor NLLF es seleccionado. La segunda posibilidad un número diferente de parámetros entre los modelos, por ejemplo, diferentes ordenes para la tendencia o una covarianza isotrópica contra una anisotrópica. El modelo con la mayoría de parámetros es el más flexible, y una reducción en la NLLF es de esperarse. Este equilibrio entre la verosimilitud y parsimonia debería ser resuelta por un criterio tal como el criterio de información Aaike (Akaike, 1974):

$$AIC = 2 NLLF + 2 l, \quad (4.10)$$

donde l es el número de párametros en el modelo. El modelo que ofrece el mínimo AIC es el seleccionado.

Otra de la ventajas de ML ha sido la sencilla evaluación de la incertidumbre de las estimaciones. Como se ha señalado por Bard(1974) las buenas prácticas en la inferencia estadística implica no sólo la obtención de las estimaciones sino también una evaluación necesaria de la fiabilidad de las estimaciones. La incertidumbre de un estimador insesgado se caracteriza por la varianza de la distribución muestral. El método ML se adapta bien a la evaluación de la incertidumbre de los estimadores en al menos tres formas diferentes: la matriz de información de Fisher (Kitanidis, 1983), los intervalos de probabilidad y el estadístico de razón de verosimilitud (Kalbfleisch, 1979). Otra ventaja es que ML es preferible en términos del error cuadrático medio. Un estimador se caracteriza por su distribución muestral y el resumen estadístico de la distribución muestral significa:

$$\bar{\theta} = E\{\hat{\theta}\} \quad (4.11)$$

y la varianza:

$$Var(\hat{\theta}) = E\{(\hat{\theta} - \bar{\theta})(\hat{\theta} - \bar{\theta})'\}. \quad (4.12)$$

Buenas propiedades de un estimador son insesgamiento, es decir, el sesgo b es cero

$$b = \bar{\theta} - \theta, \quad (4.13)$$

donde θ es el vector de los parámetros reales, y varianza muestral pequeña (4.12). Cabe señalar que un estimador insesgado con una varianza muestral de gran tamaño puede ser menos deseable que el estimador sesgado con varianza muestral pequeña, esto depende de la magnitud del sesgo y la varianza. Esto se cuantifica mediante el error cuadrático medio (mse) el cual da un equilibrio entre el sesgo y la varianza (alternativamente, el mse se puede definir como la varianza de los dos estimadores alrededor de los verdaderos valores de los parámetros):

$$mse(\hat{\theta}) = Var(\hat{\theta}) + bb'. \quad (4.14)$$

4.5. Ejemplo.

En un experimento de simulación, el estimador ML es comparado con otros estimadores populares en geoestadística. La comparación de los estimadores se lleva a cabo utilizando sus distribuciones muestrales obtenidas a través de simulación Monte Carlo. Así, dada una realización de un campo aleatorio sobre una malla regular (50x50) con un variograma exponencial de rango 5 y varianza 10, 100 muestras de tamaño n se tomaron aleatoriamente, los diferentes estimadores son aplicados y los histogramas muestrales son comparados.

Uno de los estimadores utilizados es el método clásico de los momentos donde se usa un procedimiento de mínimos cuadrados para ajustar el modelo a la nube de puntos $\hat{\gamma}(h_{ij})$ (ver ecuación (3.3)) con dos variantes:

- En la primera LS-h, denota sólo los pares de puntos que satisfacen $h_{ij} \leq \frac{L}{2}$, donde L es la dimensión longitudinal de la malla.
- En la segunda variante, LS-t denota, todos los puntos usados en el ajuste.

Los estimadores restantes son máxima verosimilitud restringida (REML) y ML. REML, básicamente, es ML aplicada sobre incrementos generados de los datos (llamados contrastes de error en estadística) (Harville, 1977; Kitanidis, 1983; Dietrich y Osbome, 1991).

Estos métodos LS-h, LS-t y REML, se comparan con respecto a la estimación del intervalo en 100 muestras aleatorias de datos de tamaño 15 (muestra pequeña) y de tamaño 90 (muestra grande). Debería ser claro que una muestra aleatoria digamos 15 puntos implica la selección aleatoria de 15 localizaciones en la malla de 50x50 donde los valores en estas localizaciones son seleccionadas como una muestra aleatoria. Los histogramas (aproximación discreta a la distribución muestral) de las 100 estimaciones para los estimadores. En la figura 4.1 para muestra de tamaño 15 y en la figura 4.2 para muestra de tamaño 90.

Distribución Muestral con Tamaño de Muestra 15.

- En los histogramas todos los rangos estimados de más de 30 han sido asignados a la clase [25.5,30.5) centrada en 30.
- ML y REML nunca dan una estimación mayor de 30. LS-h da la mayor parte de estas estimaciones (13 %).
- Todas las estimaciones de ML están en el intervalo (0,14.5) y todas las estimaciones REML están en el intervalo (0,25.5).
- Los porcentajes de estimaciones fuera del rango (0,10.5) son LS-t (16 %), LS-h (23 %), ML (2 %) y REML(10 %).
- ML es el método que da más estimaciones con un rango cercano a cero (rango < 0.05). Los porcentajes son LS-t (8 %), LS-h (11 %), ML (21 %), REML (12 %).

- Los histogramas claramente son asimétricos y lejos de una forma gaussiana.
- La media y la varianza de las distribuciones de muestreo se ve afectada por los valores extremos de la cola que proporcionan resultados malos para LS-t, LS-h.
- Un procedimiento, aunque arbitrario, es considerar como estimaciones aceptables los valores que difieren en menos de 50 % del valor verdadero del parámetro, es decir las estimaciones en el intervalo $[2.5, 7.5]$. Los resultados son LS-t (42 %), LS-h (33 %).

Distribución Muestral con Tamaño de Muestra de 90.

- Los porcentajes de los valores en el intervalo $[2.5, 7.5]$ son LS-t (95 %), LS-h (97 %), ML (100 %), REML (98 %).
- Si se restringe el intervalo anterior a el intervalo que difiere en un 30 % de valor real, que es el intervalo $[3.5, 6.5]$, los porcentajes son: LS-t (47 %), LS-h (71 %), ML (82 %) y REML (83 %). ML y REML dan resultados comparables.
- Las estimaciones ML y REML se aproximan a una distribución gaussiana según lo predicho por la teoría de la muestra grande.

La siguiente tabla muestra el coeficiente de correlación entre las estimaciones de los diferentes estimadores para ambos tamaños de muestra (15 y 90).

	LSt	LSh	ML	REML
Muestra de 15				
LSt		0.37	0.33	0.47
LSh			0.37	0.37
ML				0.94
Muestra de 90				
LSt		0.50	0.39	0.37
LSh			0.39	0.37
ML				0.99

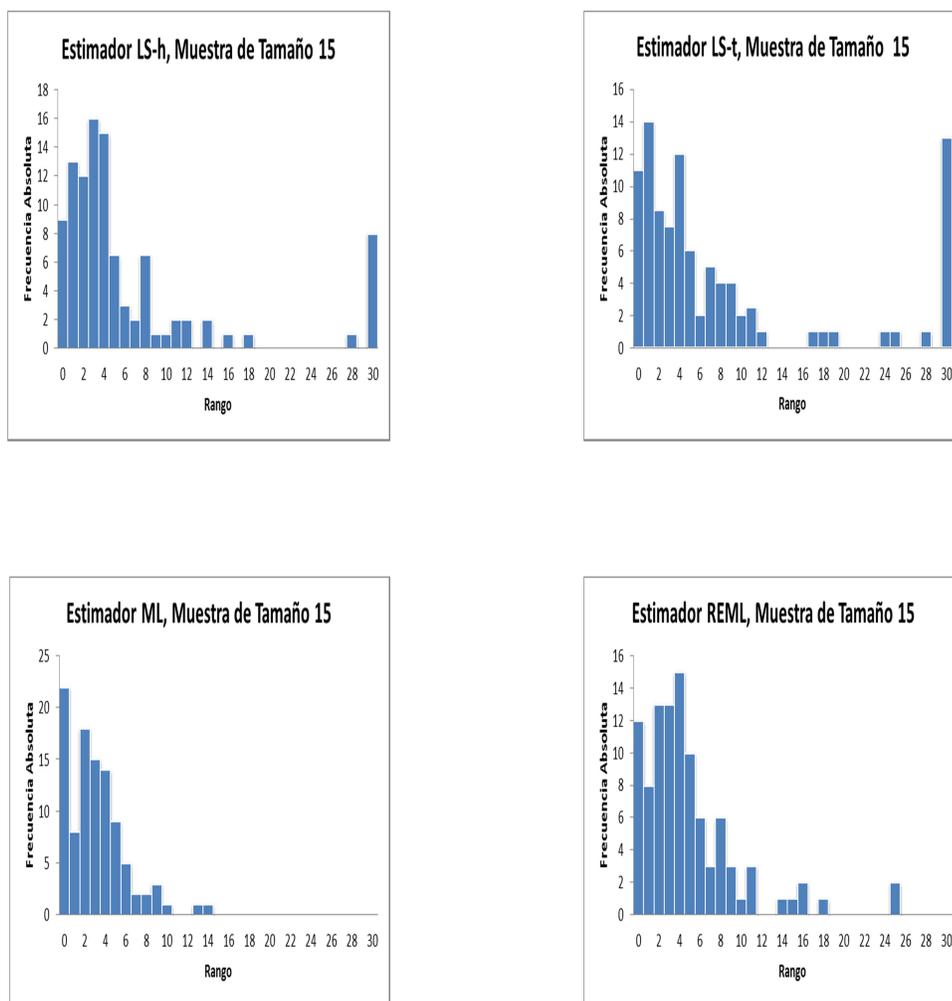


Figura 4.1: Histogramas de las distribuciones muestrales para $n = 15$.

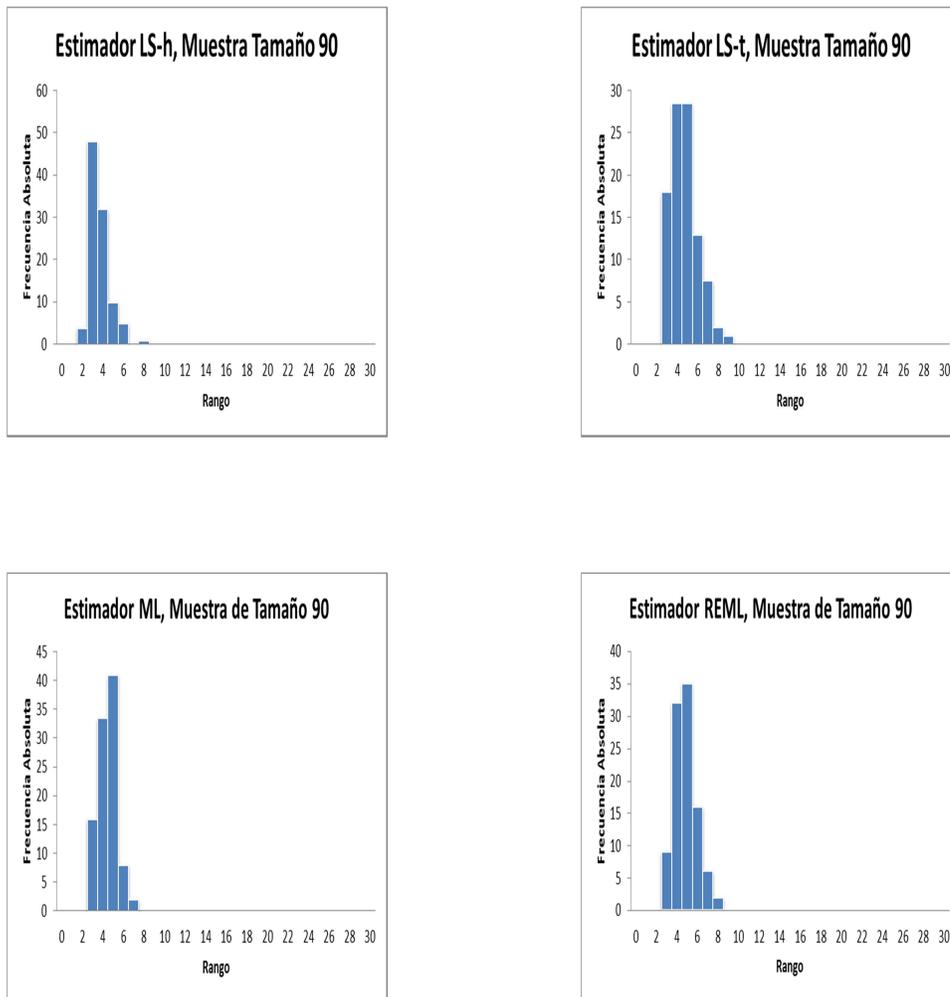


Figura 4.2: Histogramas de las distribuciones muestrales para $n = 90$.

- La correlación más alta es entre ML y REML, 0.94 para tamaño de muestra 15 y 0.90 para tamaño de muestra 90, respectivamente.
- La correlación de ML y REML con LS-t y LS-h siempre es menor que 0.5 y mayor que 0.3
- La correlación es baja en todas las demás situaciones, excepto 0.5 entre LS-t y LS-h con un tamaño de muestra de 90.

El problema del diseño de la red cuando se trata de inferencia paramétrica es evaluado fácilmente por ML de una manera similar al diseño experimental en estadísticas. La incertidumbre de los parámetros estimados es expresada por la varianza de la estimación (Kitanidis, 1983). La incertidumbre no depende del valor real de la variable en las localizaciones experimentales, pero si sobre el modelo de covarianza, el número de datos experimentales y la disposición espacial de estos datos. Esto es, la incertidumbre de las estimaciones puede ser evaluada sin tener que calcular las estimaciones por sí mismas. Este procedimiento puede ser utilizado fácilmente en la práctica de dos formas principales:

- Una comparación de diferentes planes de muestreo.
- Aumento de una red existente.

Finalmente, la justificación de utilizar ML independientemente de la distribución conjunta de los datos es que el NLLF en (4.9) puede ser simplificado después de descartar constantes irrelevantes (pero obsérvese que la verosimilitud completa es necesaria para el AIC) como:

$$L'(\hat{\beta}, \hat{\sigma}^2, \theta; z) \propto \ln |Q| + n \ln [(Z - X\hat{\beta})'Q^{-1}(Z - X\hat{\beta})]. \quad (4.15)$$

En (4.15), los dos términos tienen un significado físico. El primer término $\ln |Q|$ puede ser interpretado como una medida global de incertidumbre (Bard, 1974; Kitanidis, 1991). $\ln |Q|$ alcanza su máximo valor, $\ln |Q| = 0$, cuando $Q=I$ la matriz identidad (puro efecto nugett en la función de correlación, es decir incertidumbre máxima en la

función aleatoria), y $\ln |Q| < 0$ en el resto de las situaciones, esto es, donde existe una correlación mayor que cero.

El segundo término $(Z - X\hat{\beta})'Q^{-1}(Z - X\hat{\beta})$ es un ajuste de mínimos cuadrados pesados del modelo de covarianza que se podría utilizar sin hacer referencia a la mGd de los datos. [De hecho, el método de estimación de mínima varianza insesgada (MVUE) busca la matriz F en la forma cuadrática $(Z - \mu)'F(Z - \mu)$ la cual da la estimación de mínima varianza sujeto a las condiciones de insesgades].

Hay un intercambio entre los dos términos en la minimización de (4.15). El primer término tiende a dar grandes correlaciones y el segundo término se mueve en la dirección opuesta. El peso del logaritmo de cada término (1 y n respectivamente) es la selección de máxima entropía, que en promedio es la selección más conveniente. Así, incluso si el mGd es supuesta para producir el estimador ML el método todavía tiene una interpretación coherente.

Una desventaja de la utilización del método ML es su cálculo computacional cuando el número de datos es grande (en cada iteración de la minimización de (4.9) una matriz $n \times n$ debe ser invertida). Este problema ha sido resuelto de manera eficiente con el uso de la aproximación del método de máxima verosimilitud (Vecchina, 1988; Pardo-Igúzquiza y Dowd, 1997).

Conclusiones

El método de estimación ML proporciona ventajas adicionales al método clásico de los momentos para la estimación del variograma, siendo la más importante, la fácil evaluación de la incertidumbre de las estimaciones. Sin embargo, estos métodos no son contradictorios; ya que pueden utilizarse de forma complementaria. El modelo sugerido por el método de momentos puede ser utilizado en el procedimiento de ML y si hay un conflicto entre los diferentes modelos esto se puede resolver por el modelo de selección utilizando el procedimiento ML. El punto débil del ML frecuentemente es el considerar que éste es un derivado de la mGd, pero se ha demostrado que debido a que es imposible inferir la función de densidad de probabilidad (pdf) conjunta de los datos, hay buenas razones para seleccionar la mGd, como la distribución predeterminada. Además, la ecuación de verosimilitud de ML obtenida tiene una clara interpretación heurística y podría obtenerse de otras consideraciones independientes de la distribución multivariada de los datos. De acuerdo con los resultados de un experimento simulado, el método ML parece ser el más conveniente en términos de *mse*. Así los resultados empíricos parecen favorecer la consistencia, eficiencia y la normalidad asintótica de las estimaciones ML cuando los datos están correlacionados, es decir, el mismo que con variables aleatorias independientes e idénticamente distribuidos.

REFERENCIAS BIBLIOGRÁFICAS

- [1] Bard, Y. (1974), *Nonlinear parameter estimation* .Academic Press.New York,341.
- [2] Christensen, R. (1991), *Linear Models for Multivariate, Time Series and Spatial Data*. New York: Springer-Verlag.
- [3] Cressie, N. (1993). *Statistics for Spatial Data*. John Wiley, New York.
- [4] Cressie, N. y Lahiri, S.N. (1996), *Asymptotics for REML Estimation of Spatial Covariance Parameters*, *Journal of Statistical Planning and Inference*, **50**, 327-341.
- [5] Dietrich, C. y Osborne, M. (1991), *Estimation of covariance parameters in Kriging via restricted maximum likelihood*. *Math Geology*, **23**, 655-672.
- [6] Eulogio Pardo-Igúzquiza. (1998). *Maximun Likelihood Estimation of Spatial Covariance Parameters*. *Mathematical Geology*, Vol.30, No.1.
- [7] Harville, H. (1977), *Maximum likelihood approaches to variance component estimation and to related problems*. *Jour. Am. Stat. Assoc.*, **72**, 320-388.
- [8] Kalbfleisch (1979), *Probability and statistical inference II* . Spinger Verlag. New York, 316
- [9] Kitanidis, P.K. (1983), *Statistical Estimation of Polynomial Generalized Covariance Functions and Hydrologic Applications*, *Water Resources Research*, **19**, 909-921.
- [10] Kitanidis, P.K. y Lane, R.W. (1985), *Maximum Likelihood Parameter Estimation of Hydrologic Spatial Processes by the Gauss-Newton Method*. *Journal of Hydrology*, **79**, 53-71.

-
- [11] Mardia, K.V. y Marshall, R.J. (1984), *On multimodality of the likelihood in the spatial linear model*, **76**, 289-295.
- [12] Mardia, K.V. y Watkins. (1989), *Maximum Likelihood Estimation of Models for Residual Covariance in Spatial Regression*, *Biometrika*, **71**, 135-146.
- [13] Mardia, K.V. y Watkins. (1989), *On multimodality of the likelihood in the spatial linear model*. *Biometrika*, **76**, 389-295.
- [14] Nelder, J.A. y Mead, R. (1965), *A simplex method for function minimization*. *Computer Jour*, **7**, 308-313.
- [15] Pardo-Iguzquiza, E.,(1997), *MLREML: a computer program for the inference of spatial covariance parameters by maximum likelihood and restricted maximum likelihood*. *Computers & Geosciences*, **23**, 153-162.
- [16] Pardo-Iguzquiza, E., y Dowd, P. A.,(1997), *AMLE3D: a computer program of spatial covariance parameters by maximum likelihood and restricted maximum likelihood*. *Computers & Geosciences*, in press.
- [17] Ripley, B.D. (1981), *Spatial Statistics* . New York: John Wiley.
- [18] Rodríguez, L. (2007), *Métodos Geoestadísticos para Campos Aleatorios Gaussianos*. Trabajo de Ascenso, DCYT, UCLA, Venezuela.
- [19] Vechia, A. V., (1988), *Estimation and model identification for continuous spatial processes* : *Jour Roy . Stat.*, Ser. **50**, 297-312.