DOI: 10.22199/S07160917.1984.0007.00003

REVISTA PROYECCIONES N° 7: 61-68
JUNIO 1984 - I.S.S.N. 0716-0917

GEOESTADÍSTICA Y REGRESIÓN LINEAL EN PROCESOS ESTOCÁSTICOS

PATRICIO ROJAS M.(*)

RESUMEN.

En este artículo se presentan algunos resultados obtenidos en la estimación y predicción de parámetros, usando la regresión lineal de procesos estocásticos. Cuando corresponde se establece una relación con la terminología usada en Geoestadística.

Los resultados presentados se obtuvieron usando un enfoque basado en espacios de Hilbert con núcleo reproductor, enfoque que permite ver la similitud entre las fórmulas obtenidas tanto para los casos finitos como infinitos, de las funciones de regresión con coeficientes no aleatorios y aleatorios, aunque no se muestran los resultados para el caso aleatorio.

(*) Departamento Matemáticas, Facultad de Ciencias, Universidad del Norte.

I. INTRODUCCION.

El problema fundamental de la Geoestadística, es el de estimar el comportamiento espacial de parámetros geológicos, tales como, Ley de Mineral, tonelaje de mineral, potencia, etc. en una cierta región. La estimación se basa en un modelo probabilístico, bajo la forma de un proceso estocástico multidimensional.

En general el procedimiento se realiza de la siguiente forma: se determina la región donde se realizarán las observaciones y el método con que se medirán los parámetros geológicos que interesan (etapa de muestreo). A partir de las observaciones se obtiene el variograma experimental, al cual se le ajusta un modelo probabilístico (etapa de ajuste del modelo). Una vez ajustado el modelo, se determinan los parámetros, con lo cual se obtiene el modelo probabilítico estimado (etapa de estimación de parámetros). Finalmente en base al modelo probabilístico estimado, se estiman los parámetros geológicos, considerando que el modelo probabilístico estimado, es el modelo real (etapa de aplicación) para estudiar el proceso estocástico subyacente que denotaremos por $[Z(x), x \in T]$ consideraremos sus características de segundo orden dada por dos funciones:

(a) La función media $m : T \to \mathbb{R}$, tal que

$$m(x) = E[Z(x)] \quad , \quad x \in T.$$

(b) La función covarianza $K : T \times T \to \mathbb{R}$, tal que

$$K(x,y) = \operatorname{cov}[Z(x), Z(y)] \quad x, y \in T$$

Para simplificar el problema supondremos K conocida y m como función lineal dada por

$$m(x) = \sum_{i=1}^{k} \beta_i f_i(x) \quad , \quad x \in T \qquad (1)$$

tal que las funciones $f_1, f_2, \ldots f_k$ son todas conocidas y los coeficientes $\beta_1, \beta_2, \ldots \beta_k$ son desconocidos, denotaremos por β al vector $(\beta_1, \beta_2, \ldots, \beta_k)$

Los problemas más interesantes en Geoestadística son la estimación del vector β, para determinar los valores medios y la estimación de una combinación lineal de la $Z(x)$ en los casos en que la función de regresión es conocida y también cuando los coeficientes de regresión son desconocidos. Todo esto se hace más útil para el caso en que el conjunto de observaciones es finito, lo cual hace técnicamente mas simple el problema matemático, el caso infinito, tiene un interés más bien académico, por la aplicación de los espacios de Hilbert, de núcleo reproductor.

II. CASO FINITO.

En este caso usamos la notación matricial de modo que (1) se puede anotar como

$$m = F\beta$$

con F una matriz de $n \times k$ con columnas $f_1, f_2, \dots f_k$

o sea $F_{mj} = f_j(x_m)$ $j = 1,2 \dots k;$ $m = 1, 2, \dots n.$

Denotaremos la varianza de $Z(x)$ por Σ es decir

$$\mathrm{var}\ [Z(x)] = \Sigma$$

y además supondremos Σ no singular.

La estimación de los coeficientes de regresión, se conoce en Geoestadística como deriva. Si M es un subconjunto finito de T, denotaremos por H_M al conjunto de todas las combinaciones lineales de las variables aleatorias $Z(t_i)$, $t_i \in M$. Se puede probar que H_M es un espacio vectorial y además que definiendo como producto interno la covarianza K, H_M es un espacio de Hilbert.

II.A. PROBLEMA DE ESTIMACION DE COEFICIENTES.

El problema de estimar los coeficientes de regresión se reducen a estimar $b'\beta$ para un vector fijo b, en estadística clásica, esta es una condición de insesgamiento, que en geoestadística Matheron la llama condición de universalidad y consiste en encontrar una combinación lineal de los $Z(X_i)$ de modo que:

$$\mathrm{var}\,(a'Z) = \min_{E}\ \mathrm{var}\,(b'\beta)$$

donde

$$E = \{w = a'Z \in H_s \ / \ E\,(a'Z) = b'\beta,\ \forall\ \beta\} \qquad (2)$$

Es bien conocida que esta condición se satisface para

$$\hat{\beta} = (F'\,\Sigma^{-1}\,F)^{-1}\,F'\,\Sigma^{-1}\,Z \qquad (3)$$

y tambien se puede deducir que

$$\mathrm{var}\,(\hat{\beta}) = (\,F'\,\Sigma^{-1}\,F)^{-1} \qquad (4)$$

y que

$$\mathrm{var}\,(b'\,\hat{\beta}) = b'\,(F'\,\Sigma^{-1}\,F)^{-1}\,b$$

Además si el proceso $[Z(x),\ x \in S]$ es Gaussiano se pueden calcular intervalos de confianza para $b'\beta$.

II.B. PROBLEMA DE PREDICCION CUANDO LA FUNCION DE REGRESION ES CONOCIDA.

El problema consiste en predecir $w^* \in H_T$ (T infinito) usando las combinaciones de E_D, donde E es el conjunto definido en (2) , puesto que los coeficientes β son conocidos, entonces se trata de encontrar el mejor predictor lineal de $w^* \in H_T$, este predictor resulta ser

$$w^* = b'\beta + d'\,\Sigma^{-1}\,(Z - F\beta) \qquad (5)$$

donde

$$d = \mathrm{cov}\,(a'Z,\ Z)$$

II.C. PROBLEMA DE PREDICCION CUANDO LOS COEFICIENTES DE REGRESION SON DESCONOCIDOS.

El problema es el mismo que en II.B., con la salvedad que ahora se trata de encontrar el mejor predictor de $w^* \in H_T$, como un estimador lineal insesgado de la esperanza de $w \in E$. Este estimador resulta ser:

$$\hat{w}^* = b' \hat{\beta} + d' \Sigma^{-1} (Z - F\hat{\beta}) \qquad (6)$$

donde

$\hat{\beta}$ está dado por (3)

Comparando (5) y (6) se ve que la única diferencia es el cambio de β por $\hat{\beta}$, como se esperaba, podemos notar además que como $a'Z - d' \Sigma^{-1} Z$ no está correlacionado con Z, tampoco está con $\hat{\beta}$, de modo que:

$$\text{cov } [a'Z - d' \Sigma^{-1} Z, \ (b' - d' \Sigma^{-1} F)\hat{\beta}] = 0$$

Por lo tanto

$$\text{var } (w^* - \hat{w}^*) = \text{var}(w^* - d' \Sigma^{-1} Z) + \text{var } [(b' - d' \Sigma^{-1} F)\hat{\beta}] \qquad (7)$$

A las ecuaciones (6) y (7) Matheron las llama, en Geoestadística, Teorema de Aditividad.

III. CASO INFINITO.

Si S es ahora un conjunto infinito, ya no podemos operar matricialmente, los elementos básicos ya no serán matrices, sino que funciones definidas sobre S.

A cada variable $w \in H_S$ asociemos la función g_w definida por

$$g_w(x) = \text{cov } [w, Z(x)] \quad , \quad x \in S$$

Sea $H_S(k) = \{g_w \ / \ w \in H_S\}$ $\qquad (8)$

Se puede demostrar que:

1.- $H_S(k)$ es un espacio vectorial y si definimos un producto interno sobre $H_S(k)$ se transforma en un espacio de Hilbert isométrico con H_S definido en II. El producto interno definido es de la forma

$$L : Z(y) \rightarrow K(°, y) \qquad y \in S$$

Tal que

$$\begin{aligned} \langle K(°, x), K(°, y)\rangle &= \mathrm{cov}[Z(x), Z(y)] \\ &= K(x, y) \quad x, y \in S \end{aligned} \qquad (9)$$

El elemento de H_S correspondiente a $g \in H_S(k)$ lo denotaremos por $\langle g, Z\rangle$.

2.- $H_S(k)$ es un espacio de Hilbert con núcleo reproductor, ya que

$$\langle , \rangle : H_S \rightarrow H_S(k)$$

Tal que

$$\langle w_1, w_2\rangle = \langle g_{w_1}, g_{w_2}\rangle$$

III.A. PROBLEMA DE ESTIMACION DE LOS COEFICIENTES.

Tal como en el caso finito se trata de estimar $b'\Theta$ para un vector fijo b, aquí se trata de encontrar una combinación lineal de los $Z(x_i)$ de modo que:

$$\mathrm{var}\ (a'Z) = \min_{H_S(k)} \mathrm{var}\ (b'\Theta)$$

Donde $H_S(k)$ es el conjunto definido en (8)

Usando multiplicadores de Lagrange se puede probar que la condición anterior se satisface para

$$\hat{\Theta} = \langle f,f\rangle^{-1} \langle f, Z\rangle \qquad (10)$$

donde $<f,f>$ es la matriz grammiana de $(f_1, f_2, \ldots, f_k)$

Se puede probar que

$$\text{var}\ (\hat{\Theta}) = <f,f>^{-1} \qquad (11)$$

También se puede probar que si S es finito entonces (10) se transforma en (3) y (11) en (4) ya que en este caso las funciones g_w se asimilan a vectores en $\mathbb{R}^n$ y se puede usar la notación matricial y la matriz grammiana $<f,f>$ se transforma en la matriz $F'\ \Sigma^{-1}\ F$.

II.B. PROBLEMA DE PREDICCION CUANDO LA FUNCION DE REGRESION ES CONOCIDA.

El problema consiste en predecir ahora $V^* \in H_T$ usando las combinaciones lineales del conjunto de funciones $H_S(k)$ definido en (8) , puesto que los coeficientes Θ son conocidos, se trata de encontrar el mejor predictor lineal de $V^* \in H_T$, este predictor resulta ser:

$$V^* = b'\Theta + <d,z> - <d,f>\ \Theta \qquad (12)$$

donde $<°,°>$ es el producto interno definido por (9) .

Si escribimos (5) como

$$w^* = b'\beta + d'\ \Sigma^{-1}\ Z - d'\ \Sigma^{-1}\ F \qquad (13)$$

Nótese la similitud de (12) y (13) .

Además en el caso finito

$$<d,Z> = d'\ \Sigma^{-1}\ Z$$

y

$$<d,f> = d'\ \Sigma^{-1}\ F$$

De modo que (12) se transforma en (13) para el caso finito.

III.C. PROBLEMA DE PREDICCION CUANDO LOS COEFICIENTES DE REGRESION SON DESCONOCIDOS.

El problema es el mismo que el citado en III.B. solo que ahora se trata de encontrar el mejor predictor lineal de $V^* \in H_T$, como un estimador lineal insesgado de la esperanza de $V \in H_S(k)$, este estimador resulta ser:

$$\hat{V}^* = b'\hat{\Theta} + \langle d,z\rangle - \langle d,f\rangle\hat{\Theta} \qquad (14)$$

Notemos que la única diferencia de (14) y (12) es el cambio de Θ por $\hat{\Theta}$ y nuevamente para el caso finito (14) se transforma en (6).

IV. FUNCION DE REGRESION CON COEFICIENTES ALEATORIOS.

Si ahora consideramos tanto para el caso infinito como para el caso finito, el problema de estimación y de predicción con funciones de regresión con coeficientes aleatorios, continua la similitud de fórmu - las y además a partir de estas fórmulas se obtienen como casos particulares las fórmulas que se detallan en este artículo; para el caso de funciones de regresión con coeficientes no aleatorios.

BIBLIOGRAFIA.

[1] David, Michel (1982): Geostatical are reserve estimation elsevier, New York.

[2] Del Pino, G.E. (1984): "Regresion en procesos estocásticos con aplicaciones a Geoestadística". XI Jornadas de Estadística. Valdivia.

[3] Matheron, G. (1965): "Les variables regionalisses et leur estimation", Paris, Masson.

[4] Journel, A.G. (1978): "Mining Geostatics", Academic Press, New York.