Revista Colombiana de Estadística
Nº 11 - 1985

SOBRE LA OPTIMIZACION DE ESTIMADORES DE MINIMO CUADRADO

Ramón Antonio Matos Moreno

U.Lomonósov. Moscú. 1985

Resumen. Se demuestra una variante del teorema Gauss-Markov, la cual permite estimar asintóticamente las fronteras inferiores de la función de riesgo en el caso de regresión lineal.

Abstract. A modification of the Gauss-Markov theorem, which allows the asymptotic estimation of lower bounds for the risk function in the case of linear regression is found.

1. Introducción. Uno de los resultados clásicos de la estadística matemática lo constituye el llamado teorema Gauss-Markov. Este teorema estipula que en la clase de estimadores lineales insesgados de las formas lineales que permiten estimación, el estimador de mínimo cuadrado tiene varianza mínima, (Scheffe (1958)).

Son conocidas algunas modificaciones de este resultado (J.Dr. Hodges, Lehman E.L. (1950), Rao, C.R. (1973), p.319).

El objetivo del presente trabajo es demostrar una variante de este teorema, la cual permite con el mínimo de alteraciones obtener el resultado correspondiente para las generalizaciones de los modelos de regresión no lineal y los casos de diseño secuencial de experimentos. Este último caso no lo trataremos en esta publicación. Anotamos sin embargo que en él, la propiedad de optimización debe ser asintótica, local y extenderse a una clase más general de estimadores que los insesgados, cuya existencia en estas condiciones es dudosa.

2. Afirmación I.

Una variante sencilla de frontera inferior no asintótica para los estimadores lineales del modelo lineal es la siguiente:

Sean y_i variables aleatorias (v.a.) tales que $y_i = f_i^T\theta + e_i$, $i = 1,\ldots,N$, donde $\theta \in \mathbb{R}^p$ es un parámetro desconocido, f_i son vectores dados pertenecientes a $\mathbb{R}$ tales que $M = \sum_{i=1}^{N} f_i f_i^T$ es una matriz no degenerada, $E\,e_i = 0$ y el vector Eee^T, donde $e = (e_1,\ldots,e_N)^T$, tiene la matriz unitaria como matriz de covarianza, T denota el signo de transposición, E - la esperanza, D - la varianza. Entonces para todo $a \in \mathbb{R}^p$ y $L \in \mathbb{R}^p$

$$\sup_{\theta\in\mathbb{R}^p} E(a^T y - L^T\theta)^2 > L^T M^{-1} L \qquad (1)$$

Además la igualdad sólo se alcanza para el estimador de mínimo cuadrado

$$L^T\hat{\theta} = L^T M^{-1} F^T y, \text{ donde } \quad F = (f_1,\ldots,f_N)^T,$$

$$y = (y_1,\ldots,y_N)^T.$$

Demostración. Utilizaremos la conocida desigualdad:

$$E\xi^2 = (E\xi)^2 + D\xi, \text{ en donde } \xi\text{-}V.a., \ E\xi < \infty. \qquad (2)$$

Bajo las condiciones citadas

$$E(a^T y - L^T\theta) = (Fa-L)^T\theta.$$

De donde, sino se verifica

$$Fa-L = 0, \qquad (3)$$

se cumple

$$\sup_{\theta\in\mathbb{R}^p} |E(a^T y - L^T\theta)|^2 = \infty$$

Por tanto de acuerdo con (2)

$$\sup_{\theta\in\mathbb{R}^p} E(a^T y - L^T\theta)^2 = \infty$$

Queda por examinar el caso en el cual (3) es verdadero, o sea cuando se cumple la condición:

$$Ea^T y = L^T\theta$$

en cuyo caso la demostración se realiza idénticamente a la demostración clásica (Scheffe (1958)).

3. Afirmación II.

El riesgo que aparece en la parte izquierda de (1) no es conveniente para la generalización de los modelos no lineales. En este caso es más racional examinar el supremo del riesgo elevado al cuadrado del estimador en un dominio no muy grande.

El significado de este tratamiento es que se tiene información del experimento anterior que permite hacer más pequeño el dominio donde se busca el parámetro desconocido θ. Sin embargo procediendo de esta manera sólo se pueden establecer asintóticamente las fronteras del riesgo para un volumen grande de muestras.

Supongamos que para un entero N, las mediciones de las funciones F_N en el punto $x \in X$ dan los valores $y_N = F_N^T\theta + e_N$, $y_N(y_1,\ldots,y_N)^T$, $F_N = (f(x_1),\ldots,f(x_N))^T$, $f(x) = (f_1(x),\ldots,f_p(x))^T$, $e_N = (e_1,\ldots,e_N)$, $Ee_i = 0$, $Ee_ie_j = (i,j)$. $E_N(x)$ es la medida probabilística correspondiente a la N-ésima serie de mediciones, $E_N(x) = \frac{1}{N}\sum_{i=1}^{N} \mathbb{I}_x(x_i)$, $\mathbb{I}_x(x_i) = \delta(x,x_i)$, δ-*símbolo de Kroneker*. El parámetro desconocido θ pertenece al conjunto abierto Θ_N.

Sea $\tau_N = \frac{1}{N}\sum_{i=1}^{N} a_N(x_i)y_i$ el estimador escalar para $L^T\theta$ (donde $L \in \mathbb{R}^p$ y a_N son funciones definidas en $\mathbb{X}$).

Las funciones definidas deben cumplir las condiciones siguientes:

I $f(x)$ es una función continua y acotada del espacio topológico $\mathbb{X}$ en $\mathbb{R}^p$.

II Θ_N contiene una esfera $B_{(r_N)}$ de radio r_N, $\lim_{N\to\infty} r_N\sqrt{N} = \infty$.

III Las funciones a_N son continuas, acotadas uniformemente y convergen uniformemente a $a(x)$ en $\mathbb{X}$ si N tiende a ∞.

IV La familia de medidas probabilísticas $E_N(x)$ converge débilmente a la medida probabilística $E(x)$.

V La matriz de información de la medida probabilística límite es positiva, o sea

$$m = \int_X f(x) f^T(x)\, dE(x) > 0.$$

Definimos el riesgo asintótico de la siguiente manera:

$$R_L = \lim_{k\to\infty} \overline{\lim_{N\to\infty}} \sup_{B(k/\sqrt{N})} NE(\tau_N - L^T\theta)^2.$$

Teorema. *Para todo* $L \in \mathbb{R}^p$ *y para toda sucesión* τ_N *que cumpla las condiciones* I-V

$$R_L \geqslant L^T m^{-1} L.$$

Demostración. Damos por conocida la igualdad

$$RL = \lim_{k\to\infty} \overline{\lim_{N\to\infty}} \sup_{B(k/\sqrt{N})} \left[\|\sqrt{N}(E\tau_N - L^T\theta)\|^2 + ND\tau_N\right] \qquad (4)$$

Si para alguna subsucesión N_i suponemos que

$$\lim_{k\to\infty} \overline{\lim_{N_i\to\infty}} \sup_{B(k/\sqrt{N_i})} \|\sqrt{N_i}(E\tau_{N_i} - L^T\theta)\| = \infty,$$

entonces $R_L = \infty$ y con ello la afirmación del teorema se verifica.

Por definición $E\tau_N = \frac{1}{N} a_N^T F_N\theta$, entonces

$$\|\sqrt{N}(E\tau_N - L^T\theta)\| = \|\sqrt{N}\left(\frac{a_N^T F_N}{N} - L\right)\theta\| = \gamma_\theta.$$

Por definición de convergencia débil y la condición IV obtenemos que

$$\lim_{N\to\infty} \frac{1}{N} a_N^T F_N = \int_X a(x) f(x) dE(x).$$

De las dos conclusiones anteriores y la condición III se deduce que es posible encontrar una esfera $B_{(k/\sqrt{N})}$ de radio $k/\sqrt{N}$ tal que

$$\lim_{N\to\infty} \sup_{B(k/\sqrt{N})} \gamma_\theta > k\left|\int_X (a(x) f(x) - L) dE(x)\right|$$

De esta igualdad es claro que si para todo N, $\lim_{N\to\infty} \sup \gamma_\theta < \infty$, entonces

$$\left|\int_X (a(x) f(x) - L) dE(x)\right| = 0 \tag{5}$$

Examinando el segundo sumando de R_L en (4) obtenemos que

$$\lim_{k\to\infty} \underline{\lim}_{N\to\infty} {}_{B(k/\sqrt{N})} ND\tau_N = \int_X a^2(x) dE(x)$$

Por otro lado no es difícil comprobar que

$$\lim_{N\to\infty} ND(L^T\hat{\theta}) = L^T m^{-1} L$$

donde $\hat{\theta}$ es el estimador de mínimo cuadrado.

Utilizando la igualdad (5) concluimos que el vector $a(x) - Lf^{-1}(x) \in L_2(\mathbb{X}, E)$, ($L_2(\mathbb{X}, E)$ es el espacio de funciones integrables al cuadro en $\mathbb{X}$

con medida E) es perpendicular al vector $b^T f(x) \in L_2$ para todo $b \in \mathbb{R}^p$. Es efecto

$$\left|\int_X (a(x)-Lf^{-1}(x))b^T f(x)dE(x)\right| = \left|\int_X (a(x)b^T f(x)-Lb^T)dE(x)\right|$$
$$= 0$$

por lo tanto

$$\int_X a^2(x)dE(x) = L^T m^{-1}L+\int_X (a(x)-Lf^{-1}(x))^2 dE(x) \geq L^T m^{-1}L.$$

lo que demuestra que

$$R_L \geq L^T m^{-1}L.$$

* *

BIBLIOGRAFIA

Hodges, J. Jr., Lehamn, E.L., "Some problems in minimax point estimation", Ann. Math. Statist. Vol. 21, Nº 2, (1950), p.182-197.

Rao, C.R., Linear statistical inference and its applications, John Wiley e Hijos, New York (1973).

Scheffe, H., The analysis of variance; John Wiley e Hijos, Ney York,)1958).

**