CÁLCULO EFICIENTE DEL ESTIMADOR *JACKKNIFE* PARA MÍNIMOS CUADRADOS LINEALES DE RANGO COMPLETO

por

Héctor Jairo Martínez R.[1] **& Ana María Sanabria R.**[2]

Resumen

Martínez R., Héctor Jairo & Ana María Sanabria R.: Cálculo eficiente del estimador *jackknife* para mínimos lineales de rango completo. Rev. Acad. Colomb. Cienc. **30** (116): 361–365, 2006. ISSN 0370-3908.

En este artículo extendemos, al problema de mínimos cuadrados lineales de rango completo, el resultado presentado por **Martínez & Sanabria** [3], en el cual se reduce el costo del algoritmo estándar para calcular el estimador jackniffe para mínimos cuadrados lineales cuando, además del problema prinicipal, todos los subproblemas involucrados son de rango completo.

Palabras clave: Estimador *jackknife*, mínimos cuadrados lineales, rango completo.

Abstract

In this article, we extend to the full rank linear least square problem a **Martínez & Sanabria** result [3], which reduces the cost of the standard algorithm to compute the jackknife estimator for the linear least square problem when, beside the initial problem, all the involved subproblems are full rank.

Key words: Jackknife estimator, Linear least square problem, Full rank.

[1]Profesor Titular, Depto. de Matemáticas, Universidad del Valle, A.A. 25360. Email: hector@univalle.edu.co
[2]Profesor Asociado, Depto. de Matemáticas, Universidad del Valle, A.A. 25360. Email: anamasan@univalle.edu.co
AMS Classification 2000: Primary: . Secondary: .

1. Introducción.

En las dos últimas décadas, los estimadores basados en técnicas de re-muestreo han ganado importancia debido a las facilidades computacionales de los nuevos tiempos. Pero, como paralelamente ha aumentado la dimensión de los problemas a resolver, las facilidades computacionales no eximen de la necesidad de buscar algoritmos más eficientes para los cálculos. En [3], usando convenientemente propiedades básicas del álgebra lineal, se propone un algoritmo mucho más eficiente que el algoritmo estándar para el cálculo del *estimador jackknife para mínimos cuadrados lineales* (EJMCL), el cual funciona cuando el problema de estimación inicial y los subproblemas involucrados son de rango completo.

En este artículo, proponemos una modificación al algoritmo anterior, de tal manera que conserve la eficiencia sin requerir condición alguna sobre los subproblemas involucrados en la estimación *jackknife*. Para lograr este propósito, inicialmente, recordamos la definición del estimador *jackknife* para el problema de mínimos cuadrados lineales de rango completo, luego presentamos el algoritmo estándar para el cálculo del EJMC y los aportes, para mejorar dicho algoritmo, hechos por **Martínez** y **Sanabria**. Posteriormente, presentamos una nueva caracterización de la o las soluciones de los subproblemas de mínimos cuadrados (no necesariamente de rango completo), requeridas para el cálculo del EJMCL. Finalmente, con base en el resultado anterior, proponemos la modificación al algoritmo planteada anteriormente, la cual es el objetivo central de este artículo.

2. Estimador *jackknife* para mínimos cuadrados lineales (EJMCL).

Como se planteó en [3], dado un parámetro θ y un estimador $T = t_m(Y_1, \ldots, Y_m)$ de este parámetro, se puede construir otro estimador, utilizando la técnica de jackniffe, la cual consiste en corregir el estimador inicial con base en el promedio de los m estimadores que se obtienen al aplicar el procedimiento inicial de estimación a cada una de las submuestras que resultan al eliminar una observación de la muestra inicial.

Formalmente, dada $Y_1, Y_2, \ldots, Y_m$, una muestra aleatoria de una población caracterizada por un parámetro θ y $T = t_m(Y_1, \ldots, Y_m)$ un estimador de θ, se calculan los estimadores

$$T_i = t_{m-1}(Y_1, \ldots, Y_{i-1}, Y_{i+1}, \ldots, Y_m)\,,$$

para $i = 1, \ldots, m$, y luego se calcula el estimador

$$\begin{aligned} T_J &= \frac{1}{m}\sum_{i=1}^{m}(mT - (m-1)T_i) \\ &= mT - (m-1)\sum_{i=1}^{m}\frac{T_i}{m} \\ &= T + (m-1)\left(T - \sum\frac{T_i}{m}\right), \end{aligned} \tag{1}$$

llamado estimador jackniffe [1].

En particular, dado el conjunto de observaciones (a_i^T, α_i), donde $a_i \in \mathbb{R}^n$, $m \geq n$ y $\alpha_i \in \mathbb{R}$, el problema de estimar x tal que $\alpha_i = a_i^T x$, para $i = 1, \cdots, m$, por el método de los mínimos cuadrados, se reduce a encontrar $\widehat{x}$ tal que

$$||A\widehat{x} - y||_2 = \min_{x\in\mathbb{R}^n} ||Ax - y||_2\ ,$$

donde $A = [a_1, \cdots, a_m]^T$ y $y = (\alpha_1, \cdots, \alpha_m)^T$, lo cual se conoce como el *problema de mínimos cuadrados lineales* (PMCL). Un PCML es de *rango completo* si las columnas de A forman un conjunto de vectores linealmente independientes; en caso contrario, se dirá de *rango deficiente*.

Así, el estimador jackniffe para mínimos cuadrados lineales es

$$x_J = m\widehat{x} - (m-1)\sum_{i=1}^{m}\frac{\widehat{x}_i}{m}\ ,$$

donde $\widehat{x}_i$ es tal que

$$||A_i\widehat{x}_i - y_i||_2 = \min_{x\in\mathbb{R}^n} ||A_i x - y_i||_2\ ,$$

con

$$A_i = [a_1, \cdots, a_{i-1}, a_{i+1}, \cdots, a_m]^T$$

y

$$y_i = (\alpha_1, \cdots, \alpha_{i-1}, \alpha_{i+1}, \cdots, \alpha_m)^T\,.$$

3. Algoritmos para calcular el EJMCL.

Por lo descrito en la sección anterior, el algoritmo para calcular el EJMCL de un PMCL de rango completo se divide en tres pasos:

Algoritmo Estándar.

0. Dados $A \in \mathbb{R}^{m\times n}$ de rango completo , $y \in \mathbb{R}^m$,
1. Resolver $\min_{x\in\mathbb{R}^n} ||Ax - y||_2$,
   Salida: $\widehat{x}$.

2. Para $i = 1, \cdots, m$
   Resolver $\min_{x \in \mathbb{R}^n} ||A_i x - y_i||_2^2$,
   SALIDA: $\widehat{x}_i$.
3. Calcular $x_J = m\widehat{x} - (m-1)\sum_{i=1}^{m} \frac{\widehat{x}_i}{m}$.
   SALIDA: x_J.

Si las matrices A y A_i, para $i = 1, \cdots, m$ son todas de rango completo, los problemas de los pasos 1 y 2 se reducen a encontrar las soluciones únicas de las ecuaciones

$$A^T A x = A^T y \quad \text{y} \quad A_i^T A_i x = A_i^T y_i \quad \text{para } i = 1, \cdots, m.$$

En otras palabras, a calcular

$$\widehat{x} = (A^T A)^{-1} A^T y \quad \text{y} \quad \widehat{x}_i = (A_i^T A_i)^{-1} A_i^T y_i$$

para $i = 1, \cdots, m$, respectivamente [2].

En [3], bajo estos supuestos, los autores proponen un algoritmo que reduce el costo de los cálculos de $\widehat{x}_i$, una vez $\widehat{x}$ está calculado.

ALGORITMO MODIFICADO 1.

Para $i = 1, \cdots, m$.
{*Resolver* $A_i^T A_i x = A_i^T y_i$.}
- Resolver $S z_i = a_i$.
- Calcular $\delta_i = a_i^T z_i$.
- Calcular $\sigma_i = 1 - \delta_i$.
- Calcular $\beta_i = z_i^T d - \alpha_i \delta_i$.
- Calcular $\widehat{x}_i = \widehat{x} + \left(\frac{\beta_i}{\sigma_i} - \alpha_i\right) z_i$.

SALIDA: $\widehat{x}_i$.

Para este efecto, con base en ciertas relaciones entre las matrices y los vectores involucrados y la propiedad de matrices inversas dada por **Shermann, Morrison & Woodbury** [2], **Martínez & Sanabria** obtuvieron el siguiente resultado [3].

Teorema 1. [3] *Dadas las matrices*

$$A = [a_1, \cdots, a_m]^T \in \mathbb{R}^{m \times n},$$

$$A_i = [a_1, \cdots, a_{i-1}, a_{i+1}, \cdots, a_m]^T \in \mathbb{R}^{(m-1) \times n},$$

de rango completo y los vectores $y = (\alpha_1, \cdots, \alpha_m)^T \in \mathbb{R}^m$, $y_i = (\alpha_1, \cdots, \alpha_{i-1}, \alpha_{i+1}, \cdots, \alpha_m)^T \in \mathbb{R}^{(m-1)}$, *la solución de* $A_i^T A_i x = A_i^T y_i$ *está dada por*

$$\widehat{x}_i = \widehat{x} + \left(\frac{z_i^T d_i}{\sigma_i} - \alpha_i\right) z_i \ ,$$

donde $\widehat{x}$ *es la solución de* $A^T A x = A^T y$, z_i *es la solución de* $A^T A z = a_i$, $\sigma_i = 1 - a_i^T z_i$ *y* $d_i = A_i^T y_i$.

Desafortunadamente, el algoritmo mencionado sólo se puede implementar si tanto la matriz A, como las matrices A_i son de rango completo ($\sigma_i \neq 0$). No basta con que A sea de rango completo, puesto que esto no implica que las matrices A_i lo sean, como se puede ver en el siguiente ejemplo.

Si

$$A = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ -1 & -1 & -1 & -2 \\ 3 & 3 & 3 & 1 \end{pmatrix},$$

es fácil ver que las matrices A, A_1, A_2, A_5 y A_6 son de rango completo, pero A_3 y A_4 no lo son.

Nos propusimos entonces, encontrar una caracterización de la o las soluciones de $A_i^T A_i x = A_i^T y_i$, basada en la solución de $A^T A x = A^T y$, independientemente de si A_i es o no de rango completo, para modificar el algoritmo anterior manteniendo su eficiencia en el cálculo del estimador jackniffe.

4. Caracterización de la o las soluciones de $\mathbf{A_i^T A_i x = A_i^T y_i}$.

Lema. *Dadas la matriz* $A = [a_1, \cdots, a_m]^T \in \mathbb{R}^{m \times n}$ *de rango completo y la solución* $\widehat{x}$ *de* $A^T A x = A^T y$, *donde* $y = (\alpha_1, \cdots, \alpha_m)^T \in \mathbb{R}^m$, *si* $\sigma_i = 1 - a_i^T z_i = 0$, *entonces* $a_i^T \widehat{x} - \alpha_i = 0$, *donde* z_i *es la solución de* $A^T A z = a_i$.

Demostración. Si

$$A_i = [a_1, \cdots, a_{i-1}, a_{i+1}, \cdots, a_m]^T \in \mathbb{R}^{(m-1) \times n},$$
$$y_i = (\alpha_1, \cdots, \alpha_{i-1}, \alpha_{i+1}, \cdots, \alpha_m)^T \in \mathbb{R}^{(m-1)},$$
$$d_i = A_i^T y_i \ , \ S = A^T A, S_i = A_i^T A_i \ \text{ y } \ S_i \widehat{x}_i = d_i \ ,$$

tenemos que, como $d = A^T y = d_i + \alpha_i a_i$ y $S = S_i + a_i a_i^T$ (véanse los lemas 1 y 2 en [3]),

$$\begin{aligned}
a_i^T \widehat{x} - \alpha_i &= a_i^T S^{-1}(d_i + \alpha_i a_i) - \alpha_i \\
&= a_i^T S^{-1} d_i + \alpha_i a_i^T S^{-1} a_i - \alpha_i \\
&= z_i^T d_i + \alpha_i (z_i^T a_i - 1) \\
&= z_i^T S_i \widehat{x}_i + \alpha_i (z_i^T a_i - 1) \\
&= z_i^T (S - a_i a_i^T) \widehat{x}_i + \alpha_i (z_i^T a_i - 1) \\
&= z_i^T S \widehat{x}_i - z_i^T a_i a_i^T \widehat{x}_i + \alpha_i (z_i^T a_i - 1) \\
&= a_i^T \widehat{x}_i - z_i^T a_i a_i^T \widehat{x}_i + \alpha_i (z_i^T a_i - 1) \\
&= (1 - z_i^T a_i) a_i^T \widehat{x}_i + \alpha_i (z_i^T a_i - 1) \\
&= (1 - z_i^T a_i)(a_i^T \widehat{x}_i - \alpha_i) \ ,
\end{aligned}$$

de donde resulta obvio que si $\sigma_i = 1 - z_i^T a_i = 0$, entonces $a_i^T \widehat{x} - \alpha_i = 0$. □

En [3] se demuestra que si $\sigma_i = 1 - a_i^T z_i \neq 0$, entonces A_i es de rango completo y, por el teorema 1,

$$\widehat{x}_i = \widehat{x} + \left(\frac{z_i^T d_i}{\sigma_i} - \alpha_i \right) z_i \ .$$

Veamos que en caso de que $\sigma_i = 0$, utilizando el lema anterior, podemos determinar algunas soluciones de $A_i^T A_i x = A_i^T y_i$.

Teorema 2. *Bajo las mismas condiciones del lema anterior, si $\sigma_i = 1 - a_i^T z_i = 0$, se tiene que $\widehat{x}_i = \widehat{x} + \gamma z_i$ es solución de $A_i^T A_i x = A_i^T y_i$, para todo $\gamma \in \mathbb{R}$.*

Demostración. Sea $\gamma \in \mathbb{R}$,

$$\begin{aligned} S_i(\widehat{x} + \gamma z_i) &= (S - a_i a_i^T)(\widehat{x} + \gamma z_i) \\ &= S\widehat{x} + \gamma S z_i - a_i a_i^T \widehat{x} - \gamma a_i a_i^T z_i \\ &= d + \gamma a_i - a_i a_i^T \widehat{x} - \gamma a_i a_i^T z_i \\ &= d + \gamma a_i (1 - a_i^T z_i) - a_i a_i^T \widehat{x} \ . \end{aligned}$$

Usando la hipótesis $\sigma_i = 0$, por el lema anterior, $a_i^T \widehat{x} = \alpha_i$ y por lo tanto $S_i(\widehat{x} + \gamma z_i) = d - a_i \alpha_i = d_i$. □

Para completar la caracterización de las soluciones de $S_i x = d_i$, obtuvimos el siguiente resultado.

Teorema 3. *Bajo los supuestos del teorema 2, si $\widehat{x}_i$ es una solución de $S_i x = d_i$, entonces $\widehat{x}_i = \widehat{x} + \gamma_i z_i$ para algún $\gamma_i \in \mathbb{R}$. Además, si $\sigma_i \neq 0$, $\gamma_i = \frac{z_i^T d_i}{\sigma_i} - \alpha_i$ y si $\sigma_i = 0$, γ_i es cualquier número real.*

Demostración.

$$\begin{aligned} S_i \widehat{x}_i &= d_i \\ (S - a_i a_i^T)\widehat{x}_i &= d_i \\ S\widehat{x}_i - a_i a_i^T \widehat{x}_i &= (d - \alpha_i a_i) \\ S\widehat{x}_i &= d - \alpha_i a_i + a_i a_i^T \widehat{x}_i \\ &= d - \alpha_i a_i + \beta_i a_i \\ &= d + (\beta_i - \alpha_i) a_i \ . \end{aligned}$$

Así que

$$\begin{aligned} \widehat{x}_i &= S^{-1}(d + \gamma_i a_i) \\ &= S^{-1} d + \gamma_i S^{-1} a_i \\ &= \widehat{x} + \gamma_i z_i \ . \end{aligned}$$

Si $\sigma_i \neq 0$, S_i es invertible, por lo tanto $\widehat{x}_i$ es único y por el teorema 1, $\gamma_i = \frac{z_i^T d_i}{\sigma_i} - \alpha_i$. Si $\sigma_i = 0$, por el teorema 2, γ_i es cualquier número real. □

5. Algoritmo modificado y mejorado para calcular el EJMCL.

El anterior resultado, nos permite implementar una ligera modificación del algoritmo propuesto en [3] para ser eficientes en el cálculo del estimador jackniffe para mínimos cuadrados lineales con la condición única de que A sea de rango completo. El algoritmo modificado y mejorado para calcular $\widehat{x}_i$ aparece a continuación.

ALGORITMO MODIFICADO 2.
Para $i = 1, \cdots, m$.
{*Resolver* $A_i^T A_i x = A_i^T y_i$.}
- Resolver $S z_i = a_i$.
- Calcular $\delta_i = a_i^T z_i$.
- Calcular $\sigma_i = 1 - \delta_i$.

Si $\sigma_i \neq 0$,
{ Solución única}
$\beta_i = z_i^T d - \alpha_i \delta_i$.
$\gamma_i = \dfrac{\beta_i}{\sigma_i} - \alpha_i$.
Si no,
{Escoja una de las infinitas soluciones}
$\gamma_i = 0$.
end si
$\widehat{x}_i = \widehat{x} + \gamma_i z_i$.
end para
SALIDA: $\widehat{x}_i$.

Al comparar este algoritmo con el propuesto en la sección 7 de [3] (algoritmo modificado 1), podemos concluir que la diferencia aparece por la posibilidad abierta que existe de que $\sigma_i = 0$ (subproblema de rango deficiente), en cuyo caso el subproblema i tiene infinitas soluciones. En este caso, sugerimos tomar como solución la misma del problema inicial ($\gamma_i = 0 \Rightarrow \widehat{x}_i = \widehat{x}$).

En consecuencia, como se demostró en [3], este algoritmo reduce el costo de solución de los subproblemas aunque no sean de rango completo; más precisamente, reduce de

$$m \left[\frac{n^2(m-1)}{2} + n(m-1) + \frac{n^3}{6} \right]$$

a $m[3n + n^2]$ el número de *flops* requeridos para el cálculo del estimador *jackknife* de mínimos cuadrados lineales (EJMCL) para un modelo lineal de rango completo, siendo n el número de parámetros a estimar y m el tamaño de la muestra.

Es más, como se mostró en las conclusiones de [3], si no se necesitan los EMCL de cada una de las submuestras, se puede obtener el EJMCL sin calcular los EMCL,

mediante la expresión (véase la sección 9 de [3])

$$x_J = \widehat{x} + \frac{1}{m}\sum_{i=1}^{m} \gamma_i z_i \ ,$$

con la ventaja adicional de saber que algunos γ_i son cero, lo cual simplifica la sumatoria de la expresión anterior.

6. Conclusiones

En este artículo se ha caracterizado completamente el conjunto solución de los subproblemas de mínimos cuadrados lineales que resultan en el cálculo del estimador jackniffe de mínimos cuadrados lineales bajo el supuesto de que el problema inicial es de rango completo.

Este resultado permite mejorar el algoritmo propuesto en [3] para calcular el mencionado estimador, sin requerir que los problemas involucrados sean de rango completo, manteniendo la misma eficiencia de cómputo.

Al igual que en [3], este resultado permite hacer cálculos más eficientes siempre que el algoritmo que se utilice para resolver los diferentes subproblemas de mínimos cuadrados lineales que el estimador jackniffe requiere resolver sea el mismo que se utilice para resolver el problema de mínimos cuadrados lineales inicial.

Queda como un reto, para nosotros y nuestros lectores, el estudio de las soluciones de los subproblemas de mínimos cuadrados lineales para el caso de rango inicial deficiente

Referencias

[1] **R. Behar & M. Yepes.** *Sobre algunas técnicas de remuestreo: El método de jackknife.* Heurística **5** (6) (1991), 49-58.

[2] **J. E. Dennis & R. B. Schnabel.** *Numerical Methods for Unconstrained Optimization and Nonlinear Equations.* Prentice Hall, Engelwood Cliffs NJ, 1983.

[3] **H. J. Martínez & A. M. Sanabria.** *Cálculo eficiente del estimador jackknife para mínimos cuadrados lineales bajo condiciones de unicidad.* Matemáticas: Enseñanza Universitaria, **8** (1-2) (2000), 29–43.

Recibido el 10 de noviembre de 2005

Aceptado para su publicación en mayo de 2006