TÉCNICA DE JACKKNIFE Y ESTIMADORES EN UN MODELO LINEAL

Jackknife Technique and Linear Model Estimators

RESUMEN

En este artículo estudiamos los efectos que tiene el utilizar esta técnica en estimadores eficientes, pero sesgados de un vector de parámetros en modelos lineales, buscando reducir el sesgo sin perder la eficiencia de estos estimadores.

PALABRAS CLAVES: Técnica de Jackknife, sesgo, Modelos lineales.

ABSTRACT

In this article study this technique effects over efficient estimators but in skew vectors in linear model parameters so reducing the skew without losing the estimators efficiency.

KEYWORDS: Technique of jackknife, slant.

EDGAR ALIRIO VALENCIA ANGULO
Profesor Auxiliar, Magíster en Ciencias Matemáticas
Departamento de Matemáticas
Facultad de Ciencias Básicas
Universidad Tecnológica de Pereira
evalencia@utp.edu.co

FERNANDO MESA
Profesor Titular, Magister en Instrumentación Física
Departamento de Matemáticas
Facultad de Ciencias Básicas
Universidad Tecnológica de Pereira
femesa@utp.edu.co

1. INTRODUCCIÓN

El modelo lineal poblacional es de los modelos más utilizados cuando se desea establecer relaciones entre variables aleatorias asociadas con el comportamiento de un proceso o fenómeno. Una representación de este modelo es el siguiente**:**

$$Y = A\beta + E, \qquad (1)$$

Donde Y es un vector observable de dimensión n, X es una matriz de dimensión $n \times (p+1)$ cuyos elementos son valores de variables no aleatorias, β vector de parámetros desconocidos y estimables y E es el vector de error no observable. La estimación de β se puede hacer utilizando varios métodos, el más conocidos es el de mínimos cuadrados y se puede probar que este estimador es insesgado lineal y de mínima varianza.

Se ha observado que cuando el modelo tiene problema de multicolinealidad, el estimador $\hat{\beta} = \left(A^t A\right)^{-1} A^t Y$ de mínimos cuadrados de β, tiene error cuadrático medio demasiado grande. Por esta razón, se han considerado estimadores que reduzcan el error cuadrático medio, mejorando la eficiencia mediante la introducción de un pequeño sesgo en sus expresiones. Ejemplos de lo anterior son los estimadores $\hat{\beta}^*(k) = \left(A^t A + kI\right)^{-1} A^t Y$ $\hat{\beta}^*(c) = c\hat{\beta}$ donde $k \geq 0$ y $0 < c \leq 1$ el estimador de ridge y estimador encogido respectivamente.

En este artículo se demuestra que no es aconsejable usar la técnica de jackknife en estos dos estimadores.

2. MÉTODO DE JACKKNIFE

Los métodos de remuestreo son técnicas basadas en tomar diferentes muestras de una muestra dada y hacer una estimación de un parámetro en cada una de ellas, luego relacionando todas las estimaciones se obtiene un nuevo estimador, que con frecuencia tiene mejores propiedades que el estimador inicial.

2.1 DESCRIPCIÓN DEL MÉTODO

El método de jackknife fue presentado por Quenouille en 1949. Este método consiste en lo siguiente: Supongamos que tenemos una muestra aleatoria $Y_1, Y_2, ..., Y_n$ y un estimador t del parámetro θ basado en la muestra de tamaño n. Sea t_i al estimador t evaluado en los $n-1$ elementos que quedan después de separar el k-ésimo elemento de la muestra $t_k = t(Y_1, Y_2, ..., Y_{k-1}, Y_{k+1}, ..., Y_n)$. Ahora se construye la expresión $s_k = nt - (n-1)t_k$; donde $k = 1, 2, \ldots, n$ la cual recibe el nombre de pseudovalor. Quenouille define el estimador de jackknife t_j simple de θ asociado al estimador de t y a la muestra $Y_1, Y_2, ..., Y_n$ como el promedio de los pseudovalores

$$t_j = \frac{1}{n}\sum_{k=1}^{n} s_k = nt - \frac{n-1}{n}\sum_{k=1}^{n} t_k. \qquad (2)$$

2.2 REDUCCIÓN DEL SESGO

Una de las aplicaciones más importantes de la técnica de jackknife es la reducción del sesgo.

Definición 1 Sean $Y_1, Y_2, ..., Y_n$ una muestra aleatoria $t = t(Y_1, Y_2, ..., Y_n)$ un estimador del parámetro θ basado en la muestra de tamaño n. El estimador t se

Fecha de Recepción: 26 de enero de 2009
Fecha de Aceptación: 14 de mayo de 2009

llama estimador insesgado de θ si $E(t)=\theta$ para todo θ. De lo contrario expresión $\theta - E(t) = sesgo\ (t)$ se llama sesgo de t.
A continuación se presentan algunos estimadores de un modelo lineal y se calcula el estimador de jackknife y el sesgo en cada estimador.

3. MODELOS LINEALES

El propósito de esta sección es estimar los parámetros σ^2 y β del modelo lineal $Y = A\beta + E$ donde

$$Y=\begin{pmatrix} Y_1\\ Y_2\\ \vdots\\ Y_n\end{pmatrix}, \qquad A=\begin{pmatrix} 1 & a_{11} & \cdots & a_{1p}\\ 1 & a_{12} & \cdots & a_{2p}\\ \cdots & \cdots & \cdots & \cdots\\ 1 & a_{n1} & \cdots & a_{np}\end{pmatrix},$$

$$\beta=\begin{pmatrix} \beta_1\\ \beta_2\\ \vdots\\ \beta_p\end{pmatrix} \quad \text{y} \quad E=\begin{pmatrix} E_1\\ E_2\\ \vdots\\ E_n\end{pmatrix}.$$

Para el análisis de este modelo debemos tener encuenta los siguientes aspectos:

1. La matriz A de dimensión $n\times(p+1)$ es de rango $p+1$ (modelo de rango completo).
2. $Es(E)=\mathrm{O}$ (la esperanza o valor esperado del vector E es el vector cero).
3. $Cov(E)=Es(E^tE)=\sigma^2 I$ donde (I es la matriz idéntica).
4. $E \sim N(\mathrm{O},\sigma^2 I)$, es decir, E tiene distribución normal multivariada con $Es(E)=\mathrm{O}$ y $Cov(E)=Es(E^tE)=\sigma^2 I$.

3.1 ESTIMACIÓN DE PARÁMETROS

El problema en el modelo lineal, es estimar los parámetros que intervienen. En esta sección desarrollaremos el método de mínimos cuadrados.
Este método consiste en encontrar un valor β de tal forma que la distancia entre el valor estimado βA y el valor observado Y sea mínima; es decir, dada una matriz de datos de dimensión $n\times(p+1)$ y un vector de observación aleatorio $Y \in \Re^n$, $n \geq p+1$ encontrar un vector β tal que

$$\min_{\beta\in\Re^{p+1}} \|A\beta - Y\|^2 = \min_{\beta\in\Re^{p+1}} E^tE. \qquad (3)$$

El valor de β que minimiza E^tE puede ser obtenido derivando E^tE con respecto a β e igualando a cero

$$\frac{\partial(E^tE)}{\partial\beta} = \frac{\partial(Y^tY - 2\beta^tA^tY + \beta^tA^tA\beta)}{\partial\beta}$$
$$= -2A^tY + 2A^tA\beta = \mathrm{O}.$$

El método de mínimos cuadrados nos proporciona un estimador de σ^2 el cual se denota como $\hat{\sigma}^2$ y viene dado por

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(Y_i = \bar{Y})^2}{n-p-1} \qquad (4).$$

El siguiente teorema muestra las bondades del estimador de mínimos cuadrados $\hat{\beta}$ en el modelo lineal.

Teorema 2 Si se tiene el modelo lineal $Y = A\beta + E$, donde $Es(E)=\mathrm{O}$, $Cov(E)=\sigma^2 I$ y A es una matriz de rango completo. Entonces el estimador $\hat{\beta}$ de β es el estimador lineal de mínima varianza dentro de la clase de estimadores lineales insesgados.
La demostración se puede ver en **[3]**.

3.2 MULTICOLINEALIDAD

En algunas situaciones las columnas de A están casi linealmente relacionadas, y en tales casos las inferencias basadas sobre los modelos de regresión pueden ser engañosas ó erróneas. Cuando existe una casi dependencia lineal entre estas columnas, se dice que existe un problema de multicolinealidad o que la matriz del modelo lineal es inestable. La multicolinealidad se refiere a una situación en que, debido a una fuerte interrelación entre las variables independientes; se hace difícil desligar sus efectos individuales sobre la variable dependiente; la cuestión reside en el grado que esas interrelaciones se presentan para constituir un problema.
En esta sección estudiaremos algunos estimadores sesgados de β que buscan reducir los efectos de inestabilidad en el modelo lineal. Muchos de estos estimadores son de la forma $\hat{\beta}^* = (A^tA + B)^{-1}A^tY$ donde B es una matriz definida positiva, A^tA y B conmutan. Estudiaremos dos clases de estimadores los cuales dependen de la forma de la matriz B. Un primer caso es cuando $B = kI$ donde $k \geq 0$, el estimador que se obtiene es $\hat{\beta}^*(k) = (A^tA + kI)^{-1}A^tY$, el cual se conoce con el nombre de estimador de ridge. El segundo caso se tiene cuando $B = cI - A^tA$, donde $0 \leq c \leq 1$, el estimador que resulta se conoce con el nombre de estimador encogido y su expresión matemática es $\hat{\beta}^*(c) = c\hat{\beta}$ donde es el estimador de mínimos

cuadrados de β. La importancia de estudiar estos estimadores, se debe a que a pesar de ser estimadores sesgados, tienen error cuadrático medio menor que el error cuadrático medio de $\hat{\beta}$, es decir, estos estimadores reducen los efectos de la multicolinealidad.

4. EL EFECTO DE LA APLICACIÓN DE LA TÉCNICA DE JACKKNIFE EN LOS ESTIMADORES DE RIDGE Y ENCOGIDO

En esta sección nos proponemos estudiar el efecto de la aplicación de la técnica de jackknife en el sesgo del estimador de ridge y el estimador encogido.

4.1 EL EFECTO DE LA DE LA TÉCNICA DE JACKKNIFE EN EL ESTIMADOR DE RIDGE

Se mostró anteriormente que para un modelo lineal, el estimador de mínimos cuadrados $\hat{\beta}$ es el estimador lineal de insesgado de mínima varianza dentro de la clase de estimadores insesgados de β. Una medida adicional de la bondad de un estimador es la de error cuadrático medio, el cual mide la dispersión del estimador $\hat{\beta}$ con respecto al parámetro β. El error cuadrático medio se define como

$$ECM\left(\hat{\beta}\right) = Es\left[\left(\hat{\beta}-\beta\right)^t\left(\hat{\beta}-\beta\right)\right]$$

$$= \left(Es\left(\hat{\beta}\right)-\beta\right)^t\left(Es\left(\hat{\beta}-\beta\right)\right) + V\left(\hat{\beta}\right) \quad (5)$$

donde $V\left(\hat{\beta}\right)$ es la varianza del estimador $\hat{\beta}$ y se define como $V\left(\hat{\beta}\right) = Cov\left(\hat{\beta}\right) = \sigma^2\left(A^tA\right)^{-1}$. Como $\hat{\beta}$ es el estimador lineal insesgado de minima varianza, entonces

$$ECM\left(\hat{\beta}\right) = \sigma^2\left(A^tA\right)^{-1} \quad (6).$$

Esto no significa que $ECM\left(\hat{\beta}\right)$ sea pequeño. Esto lo podemos ver cuando el modelo tiene problemas de multicolinealidad en un modelo lineal. En este caso los elementos de la diagonal de la matriz $\left(A^tA\right)^{-1}$ son usados para detectar multicolinealidad en el modelo lineal y vienen dados

$$c_{jj} = \frac{1}{\left(1-R_j^2\right)\left(\sum_{j=0}^{p}\left(a_{jj}-\overline{a}_j\right)^2\right)} \quad (7)$$

donde R_j^2 es el coeficiente de determinación múltiple de la columna a_j sobre las p columnas restantes, c_{jj} se le conoce como factor de inflación de la varianza ver **[2]**.

Si a_j es casi linealmente dependiente sobre algún subconjunto de las p columnas restantes, entonces R_j^2 es casi uno y c_{jj} puede ser muy grande, por lo tanto $ECM\left(\hat{\beta}\right)$ puede ser demasiado grande. Para controlar esto, Hoerl y Kennard introducen el estimador de ridge

$$\hat{\beta}^*(k) = \left(A^tA + kI\right)^{-1}A^tY$$

$$= \left[I + k\left(A^tA\right)^{-1}\right]^{-1}\hat{\beta} = X\hat{\beta} \quad (8)$$

Observemos que $\hat{\beta}$ es un estimador sesgado de β cuando $k \geq 0$ esto es, $Es\left(\hat{\beta}^*(k)\right) = X\beta.$

La transformación del estimador de ridge estudiada por Hoerl y Kennard reduce A^tA a una matriz diagonal, por aplicación de una matriz ortogonal P es decir, $\Lambda = P^tA^tAP$, donde Λ es una matriz diagonal de valores propios λ_i, $i = 1,2,...,p$ y las columnas de P son los correspondientes vectores propios.

Si escribimos $A^* = AP^t$ y $\alpha = P\beta$ donde $PP^t = P^tP = I$ el modelo lineal $Y = A\beta + E$ se puede escribir $Y = A^*\alpha + E$ donde $A^*A^{*t} = \Lambda$, modelo lineal canónico. El estimador de mínimos cuadrados de α es $\hat{\alpha}^* = \left(A^{*t}A\right)^{-1}A^{*t}Y = \Lambda^{-1}A^{*t}Y$ y el estimador de ridge $\hat{\alpha}^*(k) == \left(\Lambda + kI\right)^{-1}A^{*t}Y.$

Uno de los conceptos más importantes para decidir que tan bueno es un estimador es el error cuadrático medio. $ECM\left(\hat{\beta}^*(k)\right)$ con $ECM\left(\hat{\alpha}^*(k)\right)$.

$$ECM\left(\hat{\beta}^*(k)\right) = Es\left[\left(\hat{\beta}-\beta\right)^tX^tX\left(\hat{\beta}-\beta\right)\right] \quad (9)$$

$$+ \left(X\beta-\beta\right)^t\left(X\beta-\beta\right).$$

Teniendo encuenta que $X = I - k\left(A^tA - kI\right)^{-1}$,

$$ECM\left(\hat{\beta}^*(k)\right) = \sigma^2\left[tr\left(A^tA - kI\right)^{-1} + ktr\left(A^tA\right)^{-2}\right] \quad (10)$$

$$+ k^2\beta^t\left(A^tA - kI\right)^{-2}\beta = \gamma_1(k) + \gamma_2(k).$$

Así que puede ser considerado como el sesgo al cuadrado del estimador $\gamma_2(k)$ y se define como la traza de la matriz de covarianzas del estimador $\hat{\beta}^*(k)$ y $\gamma_1(k)$ se define como la matriz de covarianzas del estimador . Esta matriz se calcula de la siguiente manera:

$$Cov\left(\hat{\beta}^*(k)\right) = \sigma^2 X\left(A^t A - kI\right)^{-1} X^t . \qquad (11)$$

A continuación presentamos algunos resultados que permiten comparar los estimadores $\hat{\beta}^*(k)$ y $\hat{\beta}$, sus demostraciones se pueden ver en **[4].**

Proposición 3 Si los $\lambda_i, i = 1,2,...,p$ son los valores propios de $A^t A$, entonces $\mu_i(W) = \frac{1}{\lambda_i + k}$ y $\mu_i(X) = \frac{\lambda_i}{\lambda_i + k}$ son los valores propios de $W = \left(A^t A + kI\right)^{-1}$ y $X = \left(I + k\left(A^t A\right)^{-1}\right)^{-1}$.

Proposición 4 El sego al cuadrado de $\gamma_2(k)$ y la varianza total de $\gamma_1(k)$ son funciones continuas monótonamente crecientes y monótonamente decreciente respectivamente.

Teorema 5 Existe siempre un $k > 0$ tal que

$$ECM\left(\hat{\beta}^*(k)\right) < ECM\left(\hat{\beta}\right). \qquad (12)$$

Presentamos algunas observaciones sobre el estimador de ridge en el modelo lineal y en el modelo canónico.

Observación 6 La relación del estimador de ridge en el modelo lineal general y en el modelo canónico esta dada por $\hat{\alpha}^*(k) = P\hat{\beta}^*(k)$. Además

$$ECM\left(\hat{\beta}^*(k)\right) < ECM\left(\hat{\alpha}^*(k)\right). \qquad (13)$$

Una ventaja de trabajar en el modelo canónico es que podemos hallar en forma explicita el valor del sesgo al cuadrado del estimador de ridge

$$E\left(\hat{\alpha}^*(k)\right) = \left(k\Lambda^{-1} + I\right)^{-1} Es\left(\hat{\alpha}\right) = \left(k\Lambda^{-1} + I\right)^{-1}\alpha \quad (14)$$

$$\begin{aligned} sesgo\left(\hat{\alpha}^*(k)\right)^2 &= \left(Es\left(\hat{\alpha}^*(k)\right) - \alpha\right)^t \left(Es\left(\hat{\alpha}^*(k)\right) - \alpha\right) \\ &= \alpha^t\left(k\Lambda^{-1} + I\right)^t\left(k\Lambda^{-1} + I\right)\alpha \qquad (15) \\ &= \alpha^t\left(k\Lambda^{-1} + I\right)^2\alpha, \end{aligned}$$

como $\left(k\Lambda^{-1} + I\right)^2$ es una matriz diagonal, por lo tanto

$$\begin{aligned} sesgo\left(\hat{\alpha}^*(k)\right)^2 &= \alpha^t diag\left(\frac{\lambda_0}{\lambda_0 + k} - 1, \cdots, \frac{\lambda_p}{\lambda_p + k} - 1\right)\alpha \\ &= k^2 \sum_{i=0}^{p} \frac{\alpha_i^2}{\left(\lambda_i + k\right)^2}. \qquad (16) \end{aligned}$$

Definición 7 Sea $\hat{\alpha}$ el estimador de mínimos cuadrados en el modelo canónico. El estimador de jackknife de mínimos cuadrados de $\hat{\alpha}$ viene dado por

$$\hat{\alpha}_J = n\hat{\alpha} - \frac{n-1}{n}\sum_{j=1}^{n}\hat{\alpha}_j \qquad (17).$$

Una forma de hallar estos estimadores $\hat{\alpha}_j$ es suponer que la matriz A^* del modelo $Y = A^*\alpha + E$ y las matrices A_j^* sean de rango completo (A_j^* la matriz obtenida al eliminar la fila j de la matriz A^*), y resolver las ecuaciones normales $A^{*t}A^*\alpha = A^{*t}Y$ y $A_j^{*t}A_j^*\alpha_j = A_j^{*t}Y_j$ para obtener los estimadores $\hat{\alpha}_j$ y $\hat{\alpha}$ los cuales determinan por completo al estimador de jackknife de mínimos cuadrados.

Los siguientes resultados establecen la manera de calcular el valor de los $\hat{\alpha}_j$.

Lema 8 Sea $A^* = \left[a_1, a_2, \cdots, a_n\right] \in \Re^{n\times(p+1)}$. Si la matriz que resulta de eliminar la fila j en la matriz A^* es $A_j^* = \left[a_1, a_2, \cdots a_{j-1}, a_{j+1}, \cdots, a_n\right]^t \in \Re^{(n-1)\times(p+1)}$, entonces $A_j^{*t}A_j^*\alpha_j = A^{*t}A^* - a_j a_j^t$.

Un segundo resultado, nos permite simplificar $A_j^{*t}Y_j$.

Lema 9 Si además de los supuestos del lema anterior, tenemos $Y = \left(y_1, y_2, \cdots, y_n\right)^t \in \Re^n$, $y_j = a_j^t\hat{\alpha}$ y $Y_j = \left(y_1, y_2, \cdots y_{j-1}, y_{j+1}, \cdots, y_n\right)^t \in \Re^{n-1}$ es el vector que resulta de eliminar la componente j en el vector Y , entonces $A_j^{*t}Y_j = A^{*t}Y - y_j a_j^t$.

Lema 10 Supongamos que la matriz A^* es de rango completo y que $\hat{\alpha}$ es la solución de $A^{*t}A^*\alpha = A^{*t}Y$. Si se tiene que, $1 - a_j^t z_j = 1 - z_j^t a_j$, entonces $a_j^t\hat{\alpha} - y_j = 0$ donde z_j es solución de $A^{*t}A^*z = a_j$.

El resultado contenido en **[5]**, dice: si $\sigma_j = 1 - z_j^t a_j$ es diferente de cero si y sólo si A_j^* es de rango completo y

$$\hat{\alpha}_j = \hat{\alpha} + \left(\frac{z_j^t S\hat{\alpha} - y_j\delta_j}{\sigma_j} - y_j\right)z_j \qquad (18)$$

es solución de $A_j^{*t} A_j^* \alpha_j = A_j^{*t} Y_j$ donde $\delta_j = 1 - \sigma_j$ y $S = A^{*t} A^*$. Veamos que cuando $\delta_j = 0$ y usando el Lema 8, encontramos soluciones de $A_j^{*t} A_j^* \alpha_j = A_j^{*t} Y_j$.

Teorema 11 Bajo los supuestos del Lema 9. Si $\hat{\alpha}$ es solución de $A_j^{*t} A_j^* \alpha_j = A_j^{*t} Y_j$, entonces $\hat{\alpha}_j = \hat{\alpha} + \psi z_j$, para alguna ψ. Para calcular el estimador de jackknife de mínimos cuadrados utilizamos

$$\psi = \left(\frac{z_j^t S \hat{\alpha} - y_j \delta_j}{\sigma_j} - y_j \right) z_j. \quad (19)$$

La demostración se puede ver en **[5]**.
Utilizando el Teorema 11, definimos el estimador

$$\hat{\alpha}_j^*(k) = \hat{\alpha}^*(k) + \left(\frac{z_j^t(k) S(k) \hat{\alpha} - y_j \delta_j(k)}{\sigma_j} - y_j \right) z_j(k). \quad (20)$$

donde $S(k) = A^{*t} A^* + kI$, $z_j = S^{-1}(k) a_j$ y $\delta_j(k) = z_j^t(k) a_j$ y demostramos los siguientes resultados:

Proposición 12 $E\left(\hat{\alpha}_J^*(k)\right) = E\left(\hat{\alpha}(k)\right)$ para todo α.

Demostración. Hallemos $E\left(\hat{\alpha}_j^*(k)\right)$

$$\begin{aligned} E\left(\hat{\alpha}_j^*(k)\right) &= E\left(\hat{\alpha}^*(k)\right) + \left(\frac{z_j^t(k) S(k) E\left(\hat{\alpha}\right) - y_j \delta_j(k)}{\sigma_j} - y_j \right) z_j(k) \quad (21) \\ &= E\left(\hat{\alpha}^*(k)\right) + \left(\frac{z_j^t(k) S(k) \alpha - y_j \delta_j(k)}{\sigma_j} - y_j \right) z_j(k) \\ &= E\left(\hat{\alpha}^*(k)\right) + \left(\frac{a_j^t \alpha - y_j \delta_j(k)}{\sigma_j} - y_j \right) z_j(k) \\ &= E\left(\hat{\alpha}^*(k)\right). \end{aligned}$$

Finalmente obtenemos

$$\begin{aligned} E\left(\hat{\alpha}_J^*\right) &= nE\left(\hat{\alpha}^*(k)\right) - \frac{n-1}{n} \sum_{j=1}^{n} E\left(\hat{\alpha}_j^*(k)\right) \quad (22) \\ &= nE\left(\hat{\alpha}^*(k)\right) - (n-1)E\left(\hat{\alpha}^*(k)\right) = E\left(\hat{\alpha}^*(k)\right). \end{aligned}$$

Proposición 13 $V\left(\hat{\alpha}_J^*(k)\right) = V\left(\hat{\alpha}^*(k)\right)$ para todo $n > 1$.

Demostración. Calculemos la varianza de $\hat{\alpha}_J^*(k)$.

$$V\left(\hat{\alpha}_J^*(k)\right) = n^2 V\left(\hat{\alpha}^*(k)\right) - \frac{(n-1)^2}{n^2} \sum_{j=1}^{n} V\left(\hat{\alpha}_j^*(k)\right).$$

Como $\left(\frac{(n-1)^2}{n^2}\right) > 0$ y $\sum_{j=1}^{n} V\left(\hat{\alpha}_j^*(k)\right) > 0$, pora todo $n > 1$, consiguiente $V\left(\hat{\alpha}_J^*(k)\right) \geq V\left(\hat{\alpha}^*(k)\right)$.

Proposición 14 $ECM\left(\hat{\alpha}_J^*(k)\right) \geq ECM\left(\hat{\alpha}^*(k)\right)$

La demostración se sigue de las Proposiciones 12 y 13.
La importancia de la Proposición 14 es que nos permite ver que no siempre la técnica de jackknife logra reducir el sesgo cuadrático y que en ocasiones puede hasta desmejorar otras bondades importantes de un estimador, como aumentar su error cuadrático medio.

4.1 EL EFECTO DE LA DE LA TÉCNICA DE JACKKNIFE EN EL ESTIMADOR ENCOGIDO

Supongamos que estamos trabajando en el modelo lineal $Y = A\beta + E$, con las suposiciones dadas anteriormente. Definamos el siguiente estimador:

Definición 15 Sea c una constante $0 < c \leq 1$ y $\hat{\beta}$ el estimador de mínimos cuadrado de β, $\hat{\beta}^*(c) = c\hat{\beta}$ es llamado estimador encogido del parámetro β.
Antes de estudiar con detalle este estimador, hallemos el error cuadrático medio de $\hat{\beta}^*(c) = c\hat{\beta}$. Por definición

$$ECM\left(\hat{\beta}^*(c)\right) = \left(sesgo\left(\hat{\beta}^*(c)\right)\right)^2 + V\left(\hat{\beta}^*(c)\right). (23)$$

Haciendo un cálculo sencillo llegamos a que

$$\begin{aligned} \left(sesgo\left(\hat{\beta}^*(c)\right)\right)^2 &= \left(E\left(\hat{\beta}^*(c)\right) - \beta\right)^t \left(E\left(\hat{\beta}^*(c)\right) - \beta\right) \\ &= (c-1)^2 \beta^t \beta \quad (24) \end{aligned}$$

y $V\left(\hat{\beta}^*(c)\right) = c^2 \sigma^2 \sum_{i=0}^{p} \frac{1}{\lambda_i}$ donde λ_i son los valores propios de $A^t A$. Por lo tanto tenemos que

$$ECM\left(\hat{\beta}^*(c)\right) = (c-1)^2 \beta^t \beta + c^2 \sigma^2 \sum_{i=0}^{p} \frac{1}{\lambda_i}. \quad (25)$$

Proposición 16 $ECM\left(\hat{\beta}^*(c)\right) < ECM\left(\hat{\beta}\right)$ si y sólo si $\dfrac{\beta^t\beta - \sigma^2 \sum_{i=0}^{p} \frac{1}{\lambda_i}}{\beta^t\beta + \sigma^2 \sum_{i=0}^{p} \frac{1}{\lambda_i}} < c < 1$.

La demostración se puede ver en **[4]**.

Presentamos el estimador jackknife encogido $\hat{\beta}_J^*(c)$ y algunos resultados sobre su error cuadrático medio.

Definición 17 Definimos los estimadores $\hat{\beta}_j^*(c) = c\,\hat{\beta}_j$ y $\hat{\beta}_J^*(c) = n\,\hat{\beta}^*(c) - \dfrac{n-1}{n}\sum_{j=1}^{n} \hat{\beta}_j^*(c)$.

Proposición 18 $Es\left(\hat{\beta}_J^*(c)\right) = Es\left(\hat{\beta}(c)\right)$ para todo n

Demostración

$$\begin{aligned} Es\left(\hat{\beta}_J^*(c)\right) &= nEs\left(\hat{\beta}^*(c)\right) - \frac{n-1}{n}\sum_{j=1}^{n} Es\left(\hat{\beta}_j^*(c)\right) \\ &= nEs\left(\hat{\beta}^*(c)\right) - (n-1)Es\left(\hat{\beta}^*(c)\right) \\ &= Es\left(\hat{\alpha}^*(c)\right) = c\beta. \qquad (26) \end{aligned}$$

Una consecuencia inmediata de este resultado es

$$\left(sesgo\left(\hat{\beta}_J^*(c)\right)\right)^2 = \left(sesgo\left(\hat{\beta}^*(c)\right)\right)^2. \quad (27)$$

Proposición 19 $V\left(\hat{\beta}_J^*(c)\right) = V\left(\hat{\beta}^*(c)\right)$ para todo valor de σ^2, $n > 1$ y $0 < c \leq 1$.

Demostración. Hallemos la varianza de $\hat{\beta}_J^*(c)$.

$$V\left(\hat{\beta}_J^*(c)\right) = n^2 V\left(\hat{\beta}^*(c)\right) - \frac{(n-1)^2}{n^2}\sum_{j=1}^{n} V\left(\hat{\beta}_j^*(k)\right). \quad (28)$$

Como $\left(\dfrac{(n-1)^2}{n^2}\right)$ para todo $n > 1$ y $\sum_{j=1}^{n} V\left(\hat{\beta}_j^*(c)\right) > 0$, entonces $V\left(\hat{\beta}_J^*(k)\right) \geq V\left(\hat{\beta}^*(k)\right)$.

Proposición 20 $ECM\left(\hat{\alpha}_J^*(k)\right) \geq ECM\left(\hat{\alpha}^*(k)\right)$

La demostración se sigue de las Proposiciones 18 y 19.

5. CONCLUSIONES

El estimador de mínimos cuadrados $\hat{\beta}$ en el modelo lineal estudiado, suele tener un error cuadrático medio muy grande cuando el modelo tiene problemas de multicolinealidad. Para estos casos se aconseja utilizar los estimadores sesgados $\hat{\beta}^*(c)$ y $\hat{\beta}^*(k)$ de β para combatirla. La razón es que estos estimadores tienen error cuadrático medio menor que el de $\hat{\beta}$ para algunos valores de k y c. Se probo que la técnica de jackknife no reduce el sesgo cuadrático y aumenta la varianza en los estimadores $\hat{\beta}^*(c)$ y $\hat{\beta}^*(k)$, y en consecuencia aumenta el error cuadrático medio, por esta razón no es aconsejable utilizar esta técnica en estos estimadores.

6. BIBLIOGRAFÍA

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

[2] M. F. Egerton and P. J. Loycoc. Same Criticisms of Scochastics, Shrinkage and Ridge regression, with counterexamples. Technometrics, 23 (1981), 155-158.

[3] F. A. Graybill. An introduction to linear statistical models. MacGraw-Hill. E.U.A, (1961).

[4] A. Hoerl y R. Kennard. Ridge Regression. Biased estimation for non ortogonal problems. Technometrics, 12 (1970), 55-67.

[5] H. J. Martínez y A. M Sanabria. Calculo eficiente del estimador de jackknife para mínimos cuadrados lineales bajo condiciones de unicidad. Matemáticas: Enseñanza universitaria, 8 (1-2) (2000), 29-43.