



UNIVERSIDAD DE BUENOS AIRES  
Facultad de Ciencias Exactas y Naturales  
Departamento de Matemática

ESTIMADORES DE PROYECCIÓN PARA MODELOS DE REGRESIÓN  
BINOMIAL

Tesis presentada para optar al título de Doctor de la Universidad de  
Buenos Aires en el área Ciencias Matemáticas

**Andrea Claudia Bergesio**

Director de tesis: Víctor J. Yohai

Lugar de trabajo: Instituto de Cálculo - FCEN - UBA

Buenos Aires, 2006

# Estimadores de Proyección para Modelos de Regresión Binomial

## Resumen

Presentamos una clase de estimadores para modelos de regresión binomial que es una extensión de los estimadores de proyección definidos por Maronna y Yohai (1993) para el modelo lineal. Consideramos una variable dependiente  $y$ , que está relacionada con un vector  $\mathbf{x}$  de dimensión  $p$ . Suponemos que la distribución de  $y$  dado  $\mathbf{x}$  es binomial  $\text{Bi}(1, \theta)$  con probabilidad  $\theta$  dependiendo de  $\mathbf{x}$  a través de una función de enlace  $\pi$

$$\theta = \pi(\boldsymbol{\beta}'_0 \mathbf{x}) \quad (1)$$

para algún  $\boldsymbol{\beta}_0 \in \mathbb{R}^p$ .

Consideramos el siguiente modelo auxiliar con un solo parámetro  $\gamma_0$ . Se observa  $(y, z, w)$ , donde la distribución de  $y$  condicional a las variables  $z$  y  $w$  es  $\text{Bi}(1, \theta)$  y  $\theta$  es de la forma

$$\theta = \pi(\gamma_0 w + z). \quad (2)$$

Supongamos que dada una muestra aleatoria de tamaño  $n$ ,  $(y_1, z_1, w_1), \dots, (y_n, z_n, w_n)$  para el modelo (2) se tiene un estimador altamente robusto  $\hat{\gamma}_n((y_1, z_1, w_1), \dots, (y_n, z_n, w_n))$  de  $\gamma_0$ . Este estimador  $\hat{\gamma}_n$  va a ser utilizado para definir estimadores para  $\boldsymbol{\beta}_0$  en el modelo (1) de la siguiente manera. Consideramos una muestra aleatoria  $(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)$  del modelo (1), luego el estimador de proyección  $\hat{\boldsymbol{\beta}}_n$  se define como el argumento que minimiza

$$\sup_{\boldsymbol{\lambda}: \|\boldsymbol{\lambda}\|=1} |\hat{\gamma}_n((y_1, z_1(\boldsymbol{\lambda}), w_1(\boldsymbol{\beta})), \dots, (y_n, z_n(\boldsymbol{\lambda}), w_n(\boldsymbol{\beta})))|_s(z_1(\boldsymbol{\lambda}), \dots, z_n(\boldsymbol{\lambda}))$$

donde  $w_i(\boldsymbol{\beta}) = \boldsymbol{\beta}' \mathbf{x}_i$ ,  $z_i(\boldsymbol{\lambda}) = \boldsymbol{\lambda}' \mathbf{x}_i$ ,  $1 \leq i \leq n$ , y  $s$  es un estimador de escala robusto.

Los principales resultados de la tesis son:

- Se encontró el estimador  $\hat{\gamma}_n$  para el modelo (2) con menor sensibilidad a errores groseros en la clase de M-estimadores que son condicionalmente consistentes en el sentido de Fisher. Este estimador fue usado como estimador inicial para definir los estimadores de proyección.
- Se obtuvieron cotas superiores para el máximo sesgo de los estimadores de proyección en función del máximo sesgo del estimador inicial  $\hat{\gamma}_n$ .

- Se demostró que los estimadores de proyección tienen orden de convergencia  $n^{-1/2}$ .
- Se desarrolló un algoritmo que permite calcular los estimadores de proyección propuestos.
- Se realizó un estudio de Monte Carlo, donde se comparan los estimadores de proyección con otros estimadores robustos. Este estudio muestra la superioridad en robustez de los estimadores de proyección.

**Palabras claves:** regresión logística, estimadores robustos, estimadores de proyección, máximo sesgo.

# Projection Estimates for Binomial Regression Models

## Abstract

We present a class of estimates for binomial regression models which are an extension of the projection estimates introduced by Maronna and Yohai (1983) for the linear model. Suppose that we observe  $(y, \mathbf{x})$ , where  $\mathbf{x}$  is a random vector of dimension  $p$  and the distribution of  $y$  given  $\mathbf{x}$  is binomial  $\text{Bi}(1, \theta)$  with probability  $\theta$  and where

$$\theta = \pi(\boldsymbol{\beta}'_0 \mathbf{x}) \quad (1)$$

for some  $\boldsymbol{\beta}_0 \in \mathbb{R}^p$ .

Consider now the following binomial regression model with only one parameter  $\gamma_0$ . We observe  $(y, z, w)$ , where the distribution of  $y$  conditional  $(z, w)$  is  $\text{Bi}(1, \theta)$  and  $\theta$  is of the form

$$\theta = \pi(\gamma_0 w + z). \quad (2)$$

Suppose that given a random sample of size of  $n$  of model (2)  $(y_1, z_1, w_1), \dots, (y_n, z_n, w_n)$ , we have a highly robust estimate  $\hat{\gamma}_n((y_1, z_1, w_1), \dots, (y_n, z_n, w_n))$  of  $\gamma_0$ . We will use  $\hat{\gamma}_n$  to define an estimate of  $\boldsymbol{\beta}_0$  in model (1) as follows. Consider a random sample  $(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)$  of model (1), then the projection estimate  $\hat{\boldsymbol{\beta}}_n$  is defined by the argument that minimizes

$$\sup_{\boldsymbol{\lambda}: \|\boldsymbol{\lambda}\|=1} |\hat{\gamma}_n((y_1, z_1(\boldsymbol{\lambda}), w_1(\boldsymbol{\beta})), \dots, (y_n, z_n(\boldsymbol{\lambda}), w_n(\boldsymbol{\beta})))| s(z_1(\boldsymbol{\lambda}), \dots, z_n(\boldsymbol{\lambda}))$$

where  $w_i(\boldsymbol{\beta}) = \boldsymbol{\beta}' \mathbf{x}_i$ ,  $z_i(\boldsymbol{\lambda}) = \boldsymbol{\lambda}' \mathbf{x}_i$ ,  $1 \leq i \leq n$ , and where  $s$  is a robust scale estimate.

The main achievements in this thesis are

- The estimate  $\hat{\gamma}_n$  for model (2) with smallest gross error sensitivity in the class of M-estimates which are conditionally Fisher consistent. This estimate was used as initial estimate to define the projection estimates.
- Upper bounds for the maximum bias of the projection estimates as a function of the maximum bias of the initial estimate  $\hat{\gamma}_n$ .
- The order of consistency of the projection estimates, which is  $n^{-1/2}$ .
- An algorithm to compute the projection estimates.

- A Monte Carlo study where several robust estimates for the logistic model are compared. This study shows that projection estimates compare favorably with other robust estimates.

**Key words:** logistic regression, robust estimates, projection estimates, maximum bias.

# Índice

<b>1. Introducción</b>	<b>2</b>
<b>2. Algunas definiciones previas</b>	<b>6</b>
<b>3. Estimadores de regresión basados en proyecciones</b>	<b>7</b>
<b>4. Elección del estimador univariado</b>	<b>10</b>
<b>5. Cota para el sesgo</b>	<b>14</b>
<b>6. Consistencia</b>	<b>16</b>
6.1. Consistencia uniforme del estimador univariado . . . . .	16
6.2. Tasa de consistencia . . . . .	20
<b>7. Algoritmo</b>	<b>26</b>
<b>8. Simulaciones</b>	<b>29</b>
<b>9. Ejemplos</b>	<b>36</b>
9.1. Datos de Cook y Weisberg . . . . .	36
9.2. Datos de Finney . . . . .	42

# 1. Introducción

Consideramos el modelo de regresión binomial. Suponemos que observamos una variable dependiente Bernoulli  $y$ , que toma valores 0 ó 1, que está relacionada con  $p$  variables  $x_1, \dots, x_p$  por medio de una función de enlace  $\pi$ , de manera tal que si denotamos a  $(x_1, \dots, x_p)' = \mathbf{x}$ , tenemos que

$$P(y = 1 | \mathbf{x}) = \pi(\boldsymbol{\beta}'_0 \mathbf{x}) \quad (1)$$

para algún  $\boldsymbol{\beta}_0 \in \mathbb{R}^p$ , que es el vector de parámetros que nos proponemos estimar. Una de las funciones  $\pi$  más utilizadas es la función logística

$$\pi(u) = \frac{\exp(u)}{1 + \exp(u)}.$$

El modelo correspondiente se llama modelo de regresión logística.

El procedimiento clásico para estimar  $\boldsymbol{\beta}_0$  es el método de máxima verosimilitud. Denotemos  $V_n = \{v_1, \dots, v_n\}$  una muestra de tamaño  $n$  donde  $v_i = (\mathbf{x}_i, y_i) \in \mathbb{R}^p \times \{0, 1\}$ ,  $1 \leq i \leq n$ , satisfacen (1). El estimador de máxima verosimilitud  $\widehat{\boldsymbol{\beta}}_0$  se define como

$$\widehat{\boldsymbol{\beta}}_0 = \arg \max_{\boldsymbol{\beta}} L(\boldsymbol{\beta}, V_n) = \arg \min_{\boldsymbol{\beta}} \sum_{i=1}^n d(\boldsymbol{\beta}, v_i)$$

donde  $L(\boldsymbol{\beta}, V_n)$  es el logaritmo de la función de verosimilitud

$$L(\boldsymbol{\beta}, V_n) = \sum_{i=1}^n [y_i \ln \pi(\mathbf{x}'_i \boldsymbol{\beta}) + (1 - y_i) \ln (1 - \pi(\mathbf{x}'_i \boldsymbol{\beta}))]$$

y  $d(\boldsymbol{\beta}, v_i) = -y_i \ln \pi(\mathbf{x}'_i \boldsymbol{\beta}) - (1 - y_i) \ln (1 - \pi(\mathbf{x}'_i \boldsymbol{\beta}))$  es la deviance de la  $i$ -ésima observación. El estimador  $\widehat{\boldsymbol{\beta}}_0$  así definido es asintóticamente eficiente para estimar  $\boldsymbol{\beta}_0$ , pero es muy sensible a la presencia de respuestas atípicas. Pregibon (1981) y Cook y Weisberg (1982) muestran esta sensibilidad a partir de un conjunto de datos reales que analizaremos en la sección 9. En particular, Pregibon (1981) estudia los métodos basados en diagnósticos para medir el efecto de observaciones atípicas sobre el estimador de máxima verosimilitud y Cook y Weisberg (1982) definen una medida de influencia como una "distancia entre verosimilitudes"

$$LD_i = 2 \left[ L(\widehat{\boldsymbol{\beta}}, V_n) - L(\widehat{\boldsymbol{\beta}}_{(i)}, V_n) \right]$$

donde  $\widehat{\boldsymbol{\beta}}_{(i)}$  es el estimador de máxima verosimilitud que se obtiene cuando se excluye de la muestra la  $i$ -ésima observación, para analizar cuánto influye la observación  $i$  en la estimación de  $\boldsymbol{\beta}_0$ . Más recientemente, Croux et al. (2002) muestran que si la muestra  $V_n$  contiene puntos en el "overlap", es

decir cuando no es posible hallar un hiperplano que separe perfectamente las observaciones de  $y$ , el estimador de máxima verosimilitud está uniformemente acotado cuando se contamina la muestra agregando observaciones arbitrarias. Además, estos autores muestran que agregando  $2(p-1)$  outliers convenientemente elegidos el estimador de máxima verosimilitud puede estar tan cerca de 0 como se quiera. Por lo tanto se puede decir que el punto de ruptura a 0 es  $2(p-1)/(n+2(p-1))$ .

Pregibon (1982) propuso modificar los estimadores de máxima verosimilitud para aumentar su robustez. Su propuesta fue definir estimadores que minimizan

$$\sum_{i=1}^n \rho(d(\boldsymbol{\beta}, v_i)) \quad (2)$$

donde  $\rho$  es una función de pérdida que crece más suavemente que la identidad. Él sugirió usar como  $\rho$  a una función en la familia

$$\rho_c(t) = \begin{cases} t & \text{si } t \leq c \\ 2(tc)^{1/2} - c & \text{si } t > c \end{cases} \quad (3)$$

Sin embargo, los estimadores obtenidos de la minimización de (2) no resultan consistentes. Copas (1988) halló una expresión aproximada para el sesgo de estos estimadores y mostró que puede ser no despreciable. Más aún, si la función  $\rho$  es de la familia (3) el estimador resultante no es robusto frente a datos atípicos muy influyentes.

Bianco y Yohai (1996) proponen un M-estimador que puede ser pensado como una versión consistente en el sentido de Fisher de (2). Estos estimadores se definen a partir de la minimización de

$$\sum_{i=1}^n [\rho(d(\boldsymbol{\beta}, v_i)) + G(\pi(\mathbf{x}'_i \boldsymbol{\beta})) + G((1 - \pi(\mathbf{x}'_i \boldsymbol{\beta})))] \quad (4)$$

donde  $\rho$  es una función acotada, derivable y no decreciente y

$$G(t) = \int_0^t \rho'(-\ln u) du. \quad (5)$$

Bianco y Yohai (1996) sugieren utilizar la siguiente función

$$\rho(t) = \begin{cases} t - \frac{t^2}{2c} & \text{si } t \leq c \\ c/2 & \text{si } t > c \end{cases} \quad (6)$$

En ese trabajo se muestra que estos estimadores son consistentes y asintóticamente normales. También se muestra en algunos ejemplos que se pueden combinar buenas propiedades de robustez (medida por ejemplo por el máximo sesgo) y eficiencia (medida por la varianza asintótica). Croux y Haesbroeck (2003) confirman estas propiedades numéricamente, pero encuentran



que cuando se trabaja con la función de pérdida (6), frecuentemente el estimador no existe. En este trabajo se derivan algunos criterios para la existencia de estos estimadores para muestras finitas y se propone una función  $\rho$  alternativa que asegura la existencia del estimador cuando hay “overlap” en las observaciones.

Künsch et al. (1989) estudian estimadores óptimos para modelos lineales generalizados dentro de la clase de M estimadores. De dicha propuesta nos restringimos al caso  $(\mathbf{x}, y)$  satisfaciendo el modelo (1). Los estimadores que se proponen en el citado trabajo son de la forma

$$\sum_{i=1}^n \Psi(y_i, \mathbf{x}_i, \hat{\boldsymbol{\beta}}_n) = 0 \quad (7)$$

donde  $\Psi$  es una función acotada y que cumple con la restricción de ser *condicionalmente insesgado dado  $\mathbf{x}$* , es decir que si  $(\mathbf{x}, y)$  satisface el modelo (1) entonces

$$E_{\boldsymbol{\beta}}(\Psi(y, \mathbf{x}, \boldsymbol{\beta})|\mathbf{x}) = 0. \quad (8)$$

La función de influencia del M-estimador definido por (7) está dada por

$$IC_{\Psi}(y, \mathbf{x}, \boldsymbol{\beta}) = D(\Psi, \boldsymbol{\beta})^{-1}\Psi(y, \mathbf{x}, \boldsymbol{\beta})$$

con

$$D(\Psi, \boldsymbol{\beta}) = -\frac{\partial}{\partial \boldsymbol{\theta}} \int \int \Psi(y, \mathbf{x}, \boldsymbol{\theta}) dP_{\boldsymbol{\beta}}(y|\mathbf{x}) dF(\mathbf{x}) \Big|_{\boldsymbol{\theta}=\boldsymbol{\beta}}.$$

Bajo condiciones de regularidad, la matriz de covarianza asintótica del estimador definido por (7) está dada por

$$\begin{aligned} V(\Psi) &= E_{\boldsymbol{\beta}} [IC_{\Psi}(y, \mathbf{x}, \boldsymbol{\beta}) IC_{\Psi}(y, \mathbf{x}, \boldsymbol{\beta})'] \\ &= D(\Psi, \boldsymbol{\beta})^{-1} W(\Psi, \boldsymbol{\beta}) D(\Psi, \boldsymbol{\beta})'^{-1} \end{aligned}$$

donde

$$W(\Psi, \boldsymbol{\beta}) = E_{\boldsymbol{\beta}} [\Psi(y, \mathbf{x}, \boldsymbol{\beta}) \Psi(y, \mathbf{x}, \boldsymbol{\beta})'].$$

Por otro lado, la sensibilidad autoestandarizada del M-estimador se define por  $s(\Psi)$  donde

$$\begin{aligned} s^2(\Psi) &= \sup_{y, \mathbf{x}} IC_{\Psi}(y, \mathbf{x}, \boldsymbol{\beta})' V^{-1}(\Psi) IC_{\Psi}(y, \mathbf{x}, \boldsymbol{\beta}) \\ &= \sup_{y, \mathbf{x}} \Psi(y, \mathbf{x}, \boldsymbol{\beta})' W(\Psi, \boldsymbol{\beta})^{-1} \Psi(y, \mathbf{x}, \boldsymbol{\beta}). \end{aligned}$$

Luego Künsch et al. (1989) resuelven el problema de encontrar la función  $\Psi$  satisfaciendo (8) que minimice  $\text{traza}(V(\Psi))$  sujeto a  $s(\Psi) \leq b$ . La solución a este problema es de la siguiente forma

$$\Psi^*(y, \mathbf{x}, \boldsymbol{\beta}, B) = d(y, \mathbf{x}, \boldsymbol{\beta}, B) w_b(|d(y, \mathbf{x}, \boldsymbol{\beta}, B)| (\mathbf{x}' B^{-1} \mathbf{x})^{1/2}) \mathbf{x}$$

donde  $d(y, \mathbf{x}, \boldsymbol{\beta}, B) = y - \pi(\boldsymbol{\beta}'\mathbf{x}) - c(\pi(\boldsymbol{\beta}'\mathbf{x}), b/(\mathbf{x}'B^{-1}\mathbf{x})^{1/2})$  y  $w_b(a) = H_b(a)/a$  con  $H_b$  la función de Huber dada por  $H_b(a) = \max(-b, \min(a, b))$ . La función  $c(\boldsymbol{\beta}, a)$  se elige de modo que se cumpla (8). La función  $c$  resulta

$$c(\pi, a) = \begin{cases} a(\pi/(1-\pi)) - \pi & \text{si } \pi < 0.5, a < 1 - \pi \\ 1 - \pi - a((1-\pi)/\pi) & \text{si } \pi > 0.5, a < \pi \\ 0 & \text{en otro caso} \end{cases}$$

y la matriz  $B$  satisface

$$E_{\theta} [\Psi^*(\mathbf{x}, y, \boldsymbol{\beta}, B)\Psi^*(\mathbf{x}, y, \boldsymbol{\beta}, B)'] = B.$$

En esta tesis se proponen métodos robustos alternativos a los descriptos anteriormente para la estimación de parámetros en modelos de regresión logística. Estos métodos se basan en proyecciones y pueden ser considerados extensiones de los estimadores de proyección propuestos por Maronna y Yohai (1993).

La definición general de estimadores basados en proyecciones para el modelo (1) se introduce en la sección 3 y en la sección 4 presentamos algunos estimadores de esta familia especialmente interesantes. En la sección 5 se obtiene una cota para el sesgo. En la sección 6 se muestra que la tasa de consistencia para los estimadores propuestos es  $n^{-1/2}$ . Un algoritmo para la obtención de este estimador es descripto en la sección 7. En la sección 8 se presentan los resultados de un estudio de Monte Carlo donde se compara el estimador propuesto con otros estimadores robustos. En la sección 9 se analizan algunos conjuntos de datos reales.

## 2. Algunas definiciones previas

Consideremos un modelo paramétrico  $\{F_\theta : \theta \in \Theta\}$ , donde  $F_\theta$  es una distribución sobre  $\mathbb{R}^k$  y  $\Theta \subset \mathbb{R}^p$ , y un funcional  $\mathbf{T}$  definido en  $\mathbb{R}^p$ . Dado un funcional  $\mathbf{T} : \mathcal{F}_k \rightarrow \mathbb{R}^p$ , donde  $\mathcal{F}_k$  es un conjunto de distribuciones en  $\mathbb{R}^k$  que contiene las distribuciones empíricas y  $F_\theta$ ,  $\theta \in \Theta$ , podemos definir un estimador de  $\theta$  de la siguiente manera. Dada una muestra aleatoria de tamaño  $n$  de  $F_\theta$ ,  $x_1, \dots, x_n$ , el estimador correspondiente se define por  $\mathbf{T}(x_1, \dots, x_n) = \mathbf{T}(F_n)$ , donde  $F_n$  es la distribución empírica de las observaciones.

Un funcional  $\mathbf{T}$  es *consistente en el sentido de Fisher* si

$$\mathbf{T}(F_\theta) = \theta \quad \forall \theta \in \Theta.$$

Cuando ésto ocurre y el funcional es continuo en la métrica de la convergencia en distribución, el estimador  $\mathbf{T}_n$  convergerá al valor verdadero  $\theta$  en probabilidad.

Un entorno de contaminación de una distribución  $F_\theta$  definida sobre  $\mathbb{R}^k$  de tamaño  $\varepsilon$  se define como

$$\nu_{F_\theta, \varepsilon} = \{F : F = (1 - \varepsilon)F_\theta + \varepsilon F^*\},$$

donde  $F^*$  es una distribución arbitraria sobre  $\mathbb{R}^k$ .

Sea  $\mathbf{T}$  un estimador consistente en el sentido de Fisher para  $\theta$ . Si  $F \in \nu_{F_\theta, \varepsilon}$  definimos el *sesgo* de  $\mathbf{T}$  en  $F$  como

$$b(\mathbf{T}, \theta, F) = \|\mathbf{T}(F) - \mathbf{T}(F_\theta)\|_V = \|\mathbf{T}(F) - \theta\|_V$$

donde  $V$  es una matriz definida positiva de  $p \times p$  y si  $\mathbf{x} \in \mathbb{R}^p$ , entonces  $\|\mathbf{x}\|_V^2 = \mathbf{x}'V\mathbf{x}$ .

Una forma de medir la robustez de un estimador  $\mathbf{T}$  es a través del máximo sesgo en  $F \in \nu_{F_\theta, \varepsilon}$ , que denotamos

$$\delta(\mathbf{T}, F_\theta, \varepsilon) = \sup \{b(\mathbf{T}, \theta, F) : F \in \nu_{F_\theta, \varepsilon}\}.$$

### 3. Estimadores de regresión basados en proyecciones

Maronna y Yohai (1993) introducen una nueva clase de estimadores para el modelo de regresión lineal basado en proyecciones que también llamaremos estimadores de tipo P. Consideremos el modelo de regresión múltiple

$$y = \beta'_0 \mathbf{x} + u, \quad (9)$$

donde  $\mathbf{x} = (x_1, \dots, x_p)'$  es un vector aleatorio de  $\mathbb{R}^p$ ,  $\beta_0 = (\beta_{10}, \dots, \beta_{p0})'$  es el vector de los verdaderos parámetros de la regresión y el error  $u$  es una variable aleatoria independiente de  $\mathbf{x}$ . También supondremos que la distribución de  $u$  es simétrica.

Consideremos el modelo de regresión univariado que pasa por el origen

$$y = \beta_0 x + u, \quad (10)$$

donde  $x$  y  $u$  son variables independientes y la distribución de  $u$  es simétrica.

Para definir estimadores de tipo P se necesitan un funcional de estimación  $T(H)$  consistente en el sentido de Fisher para el modelo (10) donde  $H$  es la distribución bivariada de  $(x, y)$  y un funcional  $s(F)$ , donde  $F$  es una distribución sobre  $\mathbb{R}$ , asociado a un estimador de escala univariado.

El hecho de que el estimador  $T$  sea consistente en el sentido de Fisher significa que si  $(x, y)$  satisface el modelo (10) y  $H$  es su distribución se tendrá que

$$T(H) = \beta_0.$$

Dado el vector aleatorio  $(\mathbf{x}, y)$  de dimensión  $p + 1$ , y una función  $g$  de  $\mathbb{R}^{p+1} \rightarrow \mathbb{R}^q$  con  $q$  entero positivo, denotamos por  $\mathcal{L}(g(\mathbf{x}, y), H)$  la distribución de  $g(\mathbf{x}, y)$  cuando  $(\mathbf{x}, y)$  tiene distribución  $H$ .

Para  $\beta$  y  $\lambda \in \mathbb{R}^p$  se define

$$A(\beta, \lambda, H) = |T(\mathcal{L}(y - \beta' \mathbf{x}, \lambda' \mathbf{x}, H))| s(\mathcal{L}(\lambda' \mathbf{x}, H))$$

y

$$C(\beta, H) = \sup_{\|\lambda\|=1} A(\beta, \lambda, H),$$

donde  $\|\cdot\|$  denota la norma euclídea. El estimador de proyección  $\mathbf{T}$  cuando  $(\mathbf{x}, y)$  tiene distribución  $H$  se define por

$$\mathbf{T}(H) = \arg \min_{\beta \in \mathbb{R}^p} C(\beta, H).$$

La motivación heurística de este estimador es simple. Si  $(\mathbf{x}, y)$  satisface el modelo (9), entonces  $y - \beta'_0 \mathbf{x} = u$  es independiente de la proyección  $\lambda' \mathbf{x}$  para todo  $\lambda$ . Como  $T$  es consistente en el sentido de Fisher como un estimador de regresión univariado, si  $H$  es la distribución de  $(\mathbf{x}, y)$  satisfaciendo el modelo

(9) entonces  $T(\mathcal{L}(y - \beta'_0 \mathbf{x}, \lambda' \mathbf{x}, H)) = 0$  para todo  $\lambda$  y luego  $C(\beta_0, H) = 0$ . Cuando  $H$  no corresponde exactamente a la distribución de  $(\mathbf{x}, y)$  la ecuación  $C(\beta, H) = 0$  no tiene en general solución, entonces el estimador se define minimizando  $C$ . En el trabajo citado se proponen además algunas variantes en la definición de este tipo de estimadores.

Inspirados por estas ideas, introducimos estimadores basados en proyecciones para el modelo logístico aunque en este caso carecemos de errores independientes de  $\mathbf{x}$ .

Sean  $(\mathbf{x}_i, y_i)$ ,  $1 \leq i \leq n$  observaciones independientes donde  $y_i$  son variables Bernoulli cuyas probabilidades de éxito dependen de las variables  $\mathbf{x}_i$  de  $\mathbb{R}^p$  a través de la función

$$P(y_i = 1 \mid \mathbf{x}_i) = \pi(\beta'_0 \mathbf{x}_i). \quad (11)$$

Muy frecuentemente el modelo tendrá intercept y en este caso  $\mathbf{x}_i = (1, x_{i,1}, \dots, x_{i,p-1})$ . La función  $\pi$  es denominada función de enlace y generalmente es una función de distribución simétrica, es decir se cumple  $\pi(-t) = 1 - \pi(t)$ . Una de las funciones de enlace más populares es la logística dada por

$$\pi(t) = \frac{e^t}{1 + e^t}.$$

Un funcional  $\mathbf{T}$  para estimar  $\beta_0$  será afin equivariante si dada una matriz  $A$  de  $p \times p$  no singular, se tiene

$$\mathbf{T}(\mathcal{L}(y, A\mathbf{x}, H)) = A^{-1}\mathbf{T}(H).$$

Consideremos ahora el siguiente modelo de regresión binomial con solo un parámetro desconocido. Se observan  $(y, z, w)$ , donde  $y$  es Bernoulli y se tiene

$$P(y = 1 \mid z, w) = \pi(\gamma_0 z + w) \quad (12)$$

donde la función de enlace  $\pi$  es la de antes,  $w$  es una variable observada y  $\gamma$  es el parámetro desconocido. Consideramos ahora un funcional de estimación para  $\gamma_0$  consistente en el sentido de Fisher y que denominaremos  $T_0$ . La condición de consistencia de Fisher significa que si  $F_{\gamma_0}$  es la distribución de  $(y, z, w)$  satisfaciendo (12), entonces se tendrá que  $T_0(F_{\gamma_0}) = \gamma_0$ . Será natural pedir que  $T_0$  cumpla las siguientes propiedades de equivarianza.

**Afin equivarianza.** Sea  $F$  la distribución de  $(y, z, w)$ . Para todo  $a \neq 0$ ,

$$T_0(\mathcal{L}(y, az, w, F)) = \frac{1}{a}T_0(\mathcal{L}(y, z, w, F)). \quad (13)$$

**Regresión equivarianza.** Para todo  $\gamma$

$$T_0(\mathcal{L}(y, z, w + \gamma z, F)) = T_0(\mathcal{L}(y, z, w, F)) - \gamma. \quad (14)$$

En las dos definiciones anteriores el significado de  $\mathcal{L}(y, g(z, w), h(z, w), F)$  es el de distribución de  $(y, g(z, w), h(z, w))$  cuando  $(y, z, w)$  tiene distribución  $F$ . Consideremos ahora  $(\mathbf{x}, y)$  satisfaciendo (1) y llamemos a su distribución conjunta  $H_{\beta_0}$ . Tomemos las proyecciones  $z = \boldsymbol{\lambda}'\mathbf{x}$  y  $w = \boldsymbol{\beta}'_0\mathbf{x}$ , luego  $(y, z, w)$  satisface el modelo (12) con  $\gamma_0 = 0$  y por lo tanto si  $T_0$  es consistente en el sentido de Fisher se tendrá

$$T_0(\mathcal{L}(y, \boldsymbol{\lambda}'\mathbf{x}, \boldsymbol{\beta}'_0\mathbf{x}, H_{\beta_0})) = 0, \quad \forall \boldsymbol{\lambda} \in \mathbb{R}^p.$$

Luego, definimos de la misma manera que para el caso de regresión lineal

$$C(\boldsymbol{\beta}, H) = \sup_{\boldsymbol{\lambda}: s[\mathcal{L}(\boldsymbol{\lambda}'\mathbf{x}, H)] = 1} |T_0(\mathcal{L}(y, \boldsymbol{\lambda}'\mathbf{x}, \boldsymbol{\beta}'\mathbf{x}, H))|$$

donde  $s$  es un funcional de escala satisfaciendo

$$s[\mathcal{L}(az, F)] = |a| s[\mathcal{L}(z, F)]. \quad (15)$$

Finalmente definimos como estimador de proyección al funcional  $\mathbf{T}$  a través de

$$C(\mathbf{T}(H), H) = \min_{\boldsymbol{\beta}} C(\boldsymbol{\beta}, H)$$

donde  $H$  es la distribución de  $(\mathbf{x}, y)$ .

Es inmediato mostrar que el estimador  $\mathbf{T}$  así definido es afín equivariante.

**Observación 3.1** Una definición equivalente de  $\mathbf{T}$  se obtiene poniendo

$$C(\boldsymbol{\beta}, H) = \sup_{\|\boldsymbol{\lambda}\|=1} |T_0(\mathcal{L}(y, \boldsymbol{\lambda}'\mathbf{x}, \boldsymbol{\beta}'\mathbf{x}, H))| s(\mathcal{L}(\boldsymbol{\lambda}'\mathbf{x}, H)).$$

En efecto, tomemos  $\boldsymbol{\lambda}$  tal que  $s(\mathcal{L}(\boldsymbol{\lambda}'\mathbf{x}, H)) = 1$  por (13) y (15) se tiene

$$s(\mathcal{L}(\boldsymbol{\lambda}'\mathbf{x}, H)) = \|\boldsymbol{\lambda}\| s\left(\mathcal{L}\left(\frac{\boldsymbol{\lambda}'\mathbf{x}}{\|\boldsymbol{\lambda}\|}, H\right)\right) = 1$$

y por la afín equivarianza de  $T_0$

$$|T_0(\mathcal{L}(y, \boldsymbol{\lambda}'\mathbf{x}, \boldsymbol{\beta}'\mathbf{x}, H))| = \frac{1}{\|\boldsymbol{\lambda}\|} \left| T_0\left(\mathcal{L}\left(y, \frac{\boldsymbol{\lambda}'\mathbf{x}}{\|\boldsymbol{\lambda}\|}, \boldsymbol{\beta}'\mathbf{x}, H\right)\right) \right|.$$

Luego

$$|T_0(\mathcal{L}(y, \boldsymbol{\lambda}'\mathbf{x}, \boldsymbol{\beta}'\mathbf{x}, H))| = \left| T_0\left(\mathcal{L}\left(y, \frac{\boldsymbol{\lambda}'\mathbf{x}}{\|\boldsymbol{\lambda}\|}, \boldsymbol{\beta}'\mathbf{x}, H\right)\right) \right| s\left(\mathcal{L}\left(\frac{\boldsymbol{\lambda}'\mathbf{x}}{\|\boldsymbol{\lambda}\|}, H\right)\right).$$

## 4. Elección del estimador univariado

Para elegir el estimador inicial  $T_0$  para el modelo (12) que utilizaremos en el estimador de proyección resolveremos un problema de Hampel en la clase de M-estimadores condicionalmente consistentes.

Un funcional asociado a un M-estimador se define resolviendo

$$E_F(\psi(y, z, w, T(F))) = 0, \quad (16)$$

donde  $\psi : \mathbb{R}^4 \rightarrow \mathbb{R}$  es una función de scores. El M-estimador es condicionalmente consistente en el sentido de Fisher si cuando  $F_{\gamma_0}$  es la distribución de  $(y, z, w)$  satisfaciendo (12), se tiene

$$E_{F_{\gamma_0}}(\psi(y, z, w, \gamma_0) | z, w) = 0 \text{ para todo } z \text{ y } w. \quad (17)$$

Claramente un M-estimador condicionalmente consistente en el sentido de Fisher es también consistente en el sentido de Fisher cualquiera sea la distribución de  $(z, w)$ .

Planteamos el siguiente problema de Hampel: Encontrar el M-estimador condicionalmente consistente en el sentido de Fisher con mínima varianza asintótica sujeto a la restricción que

$$\text{SEG}(T_0, \gamma_0) \leq G(\gamma_0)$$

donde SEG indica la sensibilidad a errores groseros dada por

$$\text{SEG}(T, \gamma) = \frac{\max_{y,z,w} |\psi(y, z, w, \gamma)|}{B(\gamma, \psi)}$$

con

$$B(\gamma, \psi) = E \left( \frac{\partial}{\partial \gamma} \psi(y, z, w, \gamma) \right)$$

y  $G(\gamma)$  es una cota dada según el grado de robustez deseado. Sea  $\psi_k^H$  la función de Huber definida por

$$\psi_k^H(u) = \text{signo}(u) \min(|u|, k)$$

con  $k > 0$  y sea  $\psi_{\text{ML}}$  la función de score de máxima verosimilitud para el modelo (12) dada por

$$\psi_{\text{ML}}(y, z, w, \gamma) = (y - p(\gamma z + w))z.$$

Definamos  $r(z, w, \gamma, k)$  a través de

$$E_{F_{\gamma_0}}(\psi_k^H(\psi_{\text{ML}}(y, z, w, \gamma_0) - r(z, w, \gamma_0, k))) = 0,$$

y llamemos

$$\psi_k^*(y, z, w, \gamma) = \psi_k^H(\psi_{\text{ML}}(y, z, w, \gamma) - r(z, w, \gamma, k)),$$

y

$$B(\gamma, \psi^*) = E \left( \frac{\partial}{\partial \gamma} \psi^*(y, z, w, \gamma) \right).$$

Es posible mostrar usando una demostración similar al Teorema 3.2 en Maronna et al. (2006) que  $\psi_k^*(y, z, w, \gamma)$  es el M-estimador condicionalmente consistente óptimo en el sentido de Hampel cuando

$$G(\gamma_0) = \frac{k}{B(\gamma_0, \psi^*)}.$$

Vamos a encontrar ahora la función de score cuando  $k \rightarrow 0$ . Más precisamente, como multiplicando la función de score por una constante no cambia el M-estimador asociado, vamos a calcular

$$\psi_0(y, z, w, \gamma) = \lim_{k \rightarrow 0} \frac{1}{k} \psi_k^*(y, z, w, \gamma)$$

y mostraremos que

$$\psi_0(y, z, w, \gamma) = \eta_0(y, \pi(\gamma z + w)) \kappa_0(z) \quad (18)$$

donde

$$\eta_0(y, \pi) = \frac{(y - \pi)}{\max\{\pi, 1 - \pi\}} \quad (19)$$

y

$$\kappa_0(z) = \text{signo}(z). \quad (20)$$

Supongamos en primer lugar que  $z > 0$ . En este caso se tiene que

$$z(-\pi(\gamma z + w)) - r(z, w, \gamma, k) < z(1 - \pi(\gamma z + w)) - r(z, w, \gamma, k). \quad (21)$$

Para que se verifique (17) es necesario que el lado izquierdo de (21) tenga distinto signo que el derecho. Luego se debe cumplir

$$z(-\pi(\gamma z + w)) - r(z, w, \gamma, k) < 0 < z(1 - \pi(\gamma z + w)) - r(z, w, \gamma, k).$$

Observemos en primer lugar que si

$$|z(y - \pi(\gamma z + w)) - r(z, w, \gamma, k)| < k, \quad y = 0, 1 \quad (22)$$

para que la condición (17) se satisfaga se necesita que  $r(z, w, \gamma, k) = 0$ . Pero entonces si  $k$  es suficientemente pequeño (22) no se puede cumplir.

Supongamos ahora

$$|z(y - \pi(\gamma z + w)) - r(z, w, \gamma, k)| > k, \quad y = 0, 1. \quad (23)$$

Entonces la condición (17) no valdrá, excepto que  $\pi(\gamma z + w) = 1/2$ . En ese caso

$$\psi_k^*(y, z, w, \gamma) = \begin{cases} k & \text{si } y = 1 \\ -k & \text{si } y = 0 \end{cases}$$

y luego se cumple (18).

En consecuencia, si  $\pi(\gamma z + w) \neq 1/2$  y  $k$  es suficientemente pequeño se podrán dar solamente los siguientes dos casos.



(a)

$$\pi(\gamma z + w) > 1/2,$$

$$z(-\pi(\gamma z + w)) - r(z, w, \gamma, k) \leq -k$$

y

$$0 < z(1 - \pi(\gamma z + w)) - r(z, w, \gamma, k) \leq k.$$

En este caso para que se cumpla (17) debe ocurrir que

$$-(1 - \pi(\gamma z + w))k + \pi(\gamma z + w)(z(1 - \pi(\gamma z + w)) - r(z, w, \gamma, k)) = 0$$

es decir,

$$z(1 - \pi(\gamma z + w)) - r(z, w, \gamma, k) = \frac{(1 - \pi(\gamma z + w))k}{\pi(\gamma z + w)}.$$

Entonces

$$\psi_k^*(y, z, w, \gamma) = \begin{cases} \frac{(1 - \pi(\gamma z + w))k}{\pi(\gamma z + w)} & \text{si } y = 1 \\ -k & \text{si } y = 0 \end{cases}.$$

Luego se cumple (18).

(b)

$$\pi(\gamma z + w) < 1/2,$$

$$-k \leq z(-\pi(\gamma z + w)) - r(z, w, \gamma, k) < 0$$

y

$$z(1 - \pi(\gamma z + w)) - r(z, w, \gamma, k) \geq k.$$

En este caso para que se cumpla (17) debe ocurrir que

$$(1 - \pi(\gamma z + w))z(-\pi(\gamma z + w)) - r(z, w, \gamma, k) + \pi(\gamma z + w)k = 0$$

es decir,

$$z(-\pi(\gamma z + w)) - r(z, w, \gamma, k) = -\frac{(\pi(\gamma z + w))k}{1 - \pi(\gamma z + w)}.$$

Entonces

$$\psi_k^*(y, z, w, \gamma) = \begin{cases} k & \text{si } y = 1 \\ -\frac{\pi(\gamma z + w)}{1 - \pi(\gamma z + w)}k & \text{si } y = 0 \end{cases}.$$

Luego se cumple (18).

En forma similar se demuestra que (18) vale para  $z < 0$  y es evidente para  $z = 0$ .

**Observación 4.1** Es fácil demostrar que el  $M$ -estimador de  $\gamma_0$  correspondiente a  $\psi_0(y, z, w, \gamma)$  definida por (18) es afín y regresión equivariante.

**Lema 4.1** La función  $\eta_0(y, \pi)$  dada por (19) satisface:

(a)  $\eta_0(y, \pi)$  es no creciente respecto de  $\pi$ .

(b) Si  $y$  es  $Bi(1, \pi)$ , entonces la derivada de  $\eta_0(y, \pi)$  respecto de  $\pi$  tiene esperanza negativa y es acotada.

**Demostración.** Tenemos que,

$$\eta_0(y, \pi) = \frac{y - \pi}{\max\{\pi, 1 - \pi\}} = \begin{cases} \frac{y}{\pi} - 1 & \text{si } \pi > 1/2 \\ \frac{y - \pi}{1 - \pi} & \text{si } \pi < 1/2 \end{cases} \quad (24)$$

es decreciente como función de  $\pi$ . Además

$$\frac{\partial \eta_0(y, \pi)}{\partial \pi} = \begin{cases} -\frac{y}{\pi^2} & \text{si } \pi > 1/2 \\ \frac{y - 1}{(1 - \pi)^2} & \text{si } \pi < 1/2 \end{cases}$$

y por lo tanto tenemos que

$$\frac{\partial \eta_0(y, \pi)}{\partial \pi} \leq 0.$$

Por otro lado resulta

$$E\left(\frac{\partial \eta_0(y, \pi)}{\partial \pi}\right) = \begin{cases} -\frac{1}{\pi} & \text{si } \pi > 1/2 \\ -\frac{1}{(1 - \pi)} & \text{si } \pi < 1/2 \end{cases}$$

y (b) se cumple.  $\square$

## 5. Cota para el sesgo

Sea ahora  $(\mathbf{x}, y)$  satisfaciendo (1). Llamemos  $G_0$  a la distribución de  $\mathbf{x}$  y  $H_{\beta_0}$  a la distribución de  $(\mathbf{x}, y)$ . Definimos el entorno de contaminación de  $H_{\beta_0}$  de tamaño  $\varepsilon$  por

$$\mathcal{V}_{H_{\beta_0}, \varepsilon} = \{H : H = (1 - \varepsilon)H_{\beta_0} + \varepsilon H^*\},$$

donde  $H^*$  es cualquier distribución sobre  $\mathbb{R}^p \times \{0, 1\}$ .

Sea  $S(G)$  un funcional para estimar la covarianza definido sobre las distribuciones en  $\mathbb{R}^p$  con la propiedad de ser afín equivariante. Luego si  $\mathbf{z} = \mathbf{A}\mathbf{x}$ , donde  $A$  es una matriz  $p \times p$ , se debe cumplir  $S(\mathcal{L}(\mathbf{z})) = AS(\mathcal{L}(\mathbf{x}))A'$ .

Sea  $\mathbf{T}$  un estimador para  $\beta_0$  en el modelo (1) afín equivariante y consistente en el sentido de Fisher. Luego  $\mathbf{T}(H_{\beta_0}) = \beta_0$ . En cambio si  $H \in \mathcal{V}_{H_{\beta_0}, \varepsilon}$ , en general podrá ocurrir que  $\mathbf{T}(H) \neq \beta_0$ . Entonces definimos el sesgo de  $\mathbf{T}$  en  $H$  por

$$\begin{aligned} b(\mathbf{T}, \beta_0, H) &= \|\mathbf{T}(H) - \beta_0\|_{S^{-1}(G_0)} \\ &= ((\mathbf{T}(H) - \beta_0)' S^{-1}(G_0) (\mathbf{T}(H) - \beta_0))^{1/2} \end{aligned}$$

donde  $S$  es un funcional de covarianza.

La robustez del estimador se puede medir a través del máximo sesgo de  $\mathbf{T}$  en  $H \in \mathcal{V}_{H_{\beta_0}, \varepsilon}$  definido por

$$\delta(\mathbf{T}, \beta_0, G_0, \varepsilon) = \sup \left\{ b(\mathbf{T}, \beta_0, H) : H \in \mathcal{V}_{H_{\beta_0}, \varepsilon} \right\}.$$

Se puede mostrar que si  $\mathbf{T}$  es afín equivariante, entonces  $\delta$  no cambia si en vez de observar  $\mathbf{x}$  se observa  $\mathbf{x}^* = A\mathbf{x}$ , donde  $A$  es una matriz  $p \times p$  no singular.

Sea  $(y, z, w)$  satisfaciendo el modelo (12) y  $F_{\gamma_0}$  a la distribución de  $(y, z, w)$ . Sea  $T_0$  un funcional consistente en el sentido de Fisher para estimar el parámetro  $\gamma_0$ . El entorno de contaminación de  $F_{\gamma_0}$  de tamaño  $\varepsilon$  se llamará  $\mathcal{V}_{F_{\gamma_0}, \varepsilon}$  y dado  $F \in \mathcal{V}_{F_{\gamma_0}, \varepsilon}$ , el sesgo de  $T_0$  en  $F$  está dado por

$$b(T_0, \gamma_0, F) = |T_0(F) - \gamma_0|$$

y además

$$\delta(T, \gamma_0, F_{\gamma_0}, \varepsilon) = \sup \left\{ b(T, \gamma_0, F) : F \in \mathcal{V}_{F_{\gamma_0}, \varepsilon} \right\}.$$

En lo que sigue derivaremos una cota para el máximo sesgo de los estimadores de proyección.

Sea  $G_0$  la distribución de  $\mathbf{x}$  bajo el modelo central. Definamos

$$d^+(G_0, \varepsilon) = \sup_{\|\lambda\|=1} \sup_{G \in \mathcal{V}_{G_0, \varepsilon}} s(\mathcal{L}(\lambda' \mathbf{x}, G)), \quad (25)$$

y

$$d^-(G_0, \varepsilon) = \inf_{\|\lambda\|=1} \inf_{G \in \mathcal{V}_{G_0, \varepsilon}} s(\mathcal{L}(\lambda' \mathbf{x}, G)). \quad (26)$$

También definimos

$$\delta^*(T_0, \beta_0, G_0, \varepsilon) = \sup_{\|\lambda\|=1} \delta(T_0, 0, M_{\lambda, \beta_0}, \varepsilon),$$

donde  $M_{\lambda, \beta_0} = \mathcal{L}(\lambda' \mathbf{x}, \beta_0' \mathbf{x}, G_0)$

Luego tenemos el siguiente Teorema.

**Teorema 5.1** *Supongamos que  $(\mathbf{x}, y)$  satisfacen el modelo (1) donde  $\mathbf{x} \in \mathbb{R}^p$ . Consideremos un estimador de proyección  $\mathbf{T}$  basado en un estimador inicial  $T_0$  para el modelo (12) y una escala  $s$ . Supongamos además que  $T_0$  es consistente en el sentido de Fisher y que es afín y regresión equivariante. Entonces*

$$\delta(\mathbf{T}, \beta_0, G_0, \varepsilon) \leq \delta^*(T_0, \beta_0, G_0, \varepsilon) \left[ 1 + \frac{d^+(G_0, \varepsilon)}{d^-(G_0, \varepsilon)} \right]. \quad (27)$$

**Demostración.** Sea  $H \in \mathcal{V}_{H_{\beta_0, \varepsilon}}$ ,  $\beta = \mathbf{T}(H)$  y  $\lambda_\beta = (\beta - \beta_0) / \|\beta - \beta_0\|$ . Como  $T_0$  es regresión equivariante se tiene que

$$|T_0(\mathcal{L}(y, \lambda_\beta' \mathbf{x}, \beta_0' \mathbf{x}, H)) - \|\beta - \beta_0\| | = |T_0(\mathcal{L}(y, \lambda_\beta' \mathbf{x}, \beta' \mathbf{x}, H))|. \quad (28)$$

Como bajo  $H_\beta$  se tiene que  $(y, \lambda_\beta' \mathbf{x}, \beta_0' \mathbf{x})$  satisface el modelo (12) con  $\gamma = 0$ , resulta

$$|T_0(\mathcal{L}(y, \lambda_\beta' \mathbf{x}, \beta_0' \mathbf{x}, H))| \leq \delta^*(T_0, \beta_0, G_0, \varepsilon),$$

entonces de (28) resulta

$$\|\beta - \beta_0\| \leq |T_0(\mathcal{L}(y, \lambda_\beta' \mathbf{x}, \beta' \mathbf{x}, H))| + \delta^*(T, \beta_0, G_0, \varepsilon). \quad (29)$$

Además como  $\mathbf{T}(H) = \beta$ , de la definición del estimador de proyección se tiene

$$s(\mathcal{L}(\lambda_\beta' \mathbf{x}, G_0)) |T_0(\mathcal{L}(y, \lambda_\beta' \mathbf{x}, \beta' \mathbf{x}, H))| \quad (30)$$

$$\leq \sup_{\|\lambda\|=1} s(\mathcal{L}(\lambda' \mathbf{x}, G_0)) \sup_{\|\lambda\|=1} |T_0(\mathcal{L}(y, \lambda' \mathbf{x}, \beta_0' \mathbf{x}, H))|, \quad (31)$$

y como bajo  $H_{\beta_0}$  satisfacen el modelo (12) con  $\gamma = 0$ , resulta

$$\begin{aligned} \sup_{\|\lambda\|=1} |T_0(\mathcal{L}(y, \lambda' \mathbf{x}, \beta_0' \mathbf{x}, H))| &\leq \sup_{\|\lambda\|=1} \delta(T_0, 0, M_{\lambda, \beta_0}, \varepsilon). \\ &= \delta^*(T_0, \beta_0, G_0, \varepsilon). \end{aligned} \quad (32)$$

De (29), (30) y (32) se obtiene (27).  $\square$

## 6. Consistencia

### 6.1. Consistencia uniforme del estimador univariado

Supongamos que  $(\mathbf{x}, y)$  satisface el modelo (1). Si realizamos las proyecciones  $z = \boldsymbol{\lambda}'\mathbf{x}$  y  $w = \boldsymbol{\beta}'_0\mathbf{x}$ , donde  $\boldsymbol{\beta}_0$  es el verdadero valor del parámetro, es fácil ver que  $(y, w, z)$  satisface el modelo (12) con  $\gamma_0 = 0$ .

Vamos a considerar ahora la clase de estimadores para el modelo (12) donde  $\gamma_0$  se estima por la solución de

$$\sum_{i=0}^n \eta(y_i, \pi(\gamma z_i + w_i)) \kappa(z_i) = 0. \quad (33)$$

La versión funcional de los estimadores dados por (33) es  $T_0(F)$  y viene dada por la ecuación

$$E_F \eta(y, \pi(T_0(F)z + w)) \kappa(z) = 0.$$

Es inmediato verificar que estos estimadores son afín y regresión equivariantes.

Las hipótesis que vamos a hacer sobre  $\eta$ ,  $\kappa$ ,  $p$  y  $G_0$  son las siguientes.

**P1.**  $\eta$  es continuamente derivable, no creciente respecto de  $\pi$  y acotada. Además si  $y$  es  $\text{Bi}(1, \pi)$  se tiene

$$E(\eta(y, \pi)) = 0 \quad (34)$$

y

$$E(\eta_\pi(y, \pi)) < 0,$$

donde  $\eta_\pi$  es la derivada parcial de  $\eta(y, \pi)$  respecto de  $\pi$ .

**P2.**  $\kappa$  es Lipschitz, impar y acotada. Además si  $u > 0$ ,  $\kappa(u) \geq 0$  y existe  $c > 0$  tal que si  $0 < u < c$ , entonces  $\kappa(u)$  es estrictamente creciente y  $P(0 < \boldsymbol{\lambda}'\mathbf{x} < c) > 0$ .

**P3.**  $\pi(t)$  es una función de distribución estrictamente creciente con dos derivadas continuas acotadas. Además  $\pi(-t) = 1 - \pi(t)$  y  $\pi'(t) > 0$ .

**P4.**  $E(\|\mathbf{x}\|^2) < \infty$

Es fácil demostrar que **P1-P3** implican que  $\psi(y, \mathbf{x}, \boldsymbol{\lambda}, \gamma) = \eta(y, \pi(\gamma \boldsymbol{\lambda}'\mathbf{x} + \boldsymbol{\beta}'_0\mathbf{x})) \kappa(\boldsymbol{\lambda}'\mathbf{x})$  es no creciente respecto de  $\gamma \in \mathbb{R}$ .

Obsérvese que el estimador obtenido en la sección 4 es de la forma (33) donde  $\eta$  y  $\kappa$  satisfacen las condiciones **P1** y **P2** excepto que  $\eta$  sea derivable respecto de  $\pi$ . Sin embargo esta función puede ser aproximada tanto como se quiera por funciones derivables.

**Lema 6.1** Sea  $g(\boldsymbol{\lambda}, \mathbf{u})$  una función a valores reales continua en  $\boldsymbol{\lambda}$  definida sobre  $\mathbb{R}^p \times \mathbb{R}^q$  y sea  $\{\mathbf{v}_i, i \geq 1\}$  una sucesión de vectores aleatorios i.i.d. Sea

$K$  un conjunto compacto incluído en  $\mathbb{R}^p$  y supongamos además que para todo  $\lambda_0 \in K$  existe  $\varepsilon(\lambda_0)$  tal que

$$E \left( \sup_{B(\lambda_0, \varepsilon(\lambda_0))} |g(\lambda, \mathbf{v})| \right) < \infty$$

donde  $B(\lambda, \varepsilon)$  es la bola de centro  $\lambda$  y radio  $\varepsilon$ . Luego

(a) Si  $E(g(\lambda, \mathbf{v})) > 0$  para todo  $\lambda \in K$  se tiene

$$\underline{\lim}_{n \rightarrow \infty} \inf_{\lambda \in K} \frac{1}{n} \sum_{i=1}^n g(\lambda, \mathbf{v}_i) > 0 \text{ c.s.} \quad (35)$$

(b) Si  $E(g(\lambda, \mathbf{v})) < 0$  para todo  $\lambda \in K$  se tiene

$$\overline{\lim}_{n \rightarrow \infty} \sup_{\lambda \in K} \frac{1}{n} \sum_{i=1}^n g(\lambda, \mathbf{v}_i) < 0 \text{ c.s.}$$

(c)

$$\underline{\lim}_{n \rightarrow \infty} \sup_{\lambda \in K} \frac{1}{n} \sum_{i=1}^n (g(\lambda, \mathbf{v}_i) - E(g(\lambda, \mathbf{v}_i))) = 0 \text{ c.s.}$$

**Demostración.** Probemos primero (a). Sean  $\lambda_0 \in K$  y

$$u_n = \inf_{\lambda \in B(\lambda_0, 1/n)} g(\lambda, \mathbf{v}).$$

Luego  $u_n \rightarrow g(\lambda_0, \mathbf{v})$ , y por el Teorema de Convergencia Dominada

$$E(u_n) \rightarrow E(g(\lambda_0, \mathbf{v})) > 0$$

por hipótesis. Entonces existe un  $n_0 = n_0(\lambda_0)$  tal que para todo  $n \geq n_0$ ,  $E(u_n) > 0$ .

Como  $B(\lambda, 1/n(\lambda))$  forman un cubrimiento de  $K$  existe un subcubrimiento finito de  $K$   $B(\lambda_1, \delta_1), \dots, B(\lambda_J, \delta_J)$  donde  $\delta_j = 1/n_0(\lambda_j)$ . Luego

$$\begin{aligned} \inf_{\lambda \in K} \frac{1}{n} \sum_{i=1}^n g(\lambda, \mathbf{v}_i) &\geq \inf_{1 \leq j \leq J} \inf_{\lambda \in B(\lambda_j, \delta_j)} \frac{1}{n} \sum_{i=1}^n g(\lambda, \mathbf{v}_i) \\ &\geq \inf_{1 \leq j \leq J} \frac{1}{n} \sum_{i=1}^n \inf_{\lambda \in B(\lambda_j, \delta_j)} g(\lambda, \mathbf{v}_i). \end{aligned} \quad (36)$$

Sean

$$\eta_j = E \left( \inf_{\lambda \in B(\lambda_j, \delta_j)} g(\lambda, \mathbf{v}) \right) > 0$$

y  $\eta_0 = \inf_{1 \leq j \leq J} \eta_j$ . Entonces por la Ley de los Grandes Números para  $1 \leq j \leq J$  con probabilidad 1 existe  $n_j$  tal que para todo  $n \geq n_j$

$$\frac{1}{n} \sum_{i=1}^n \inf_{\boldsymbol{\lambda} \in B_0(\boldsymbol{\lambda}_j, \delta_j)} g(\boldsymbol{\lambda}, \mathbf{v}_i) > \frac{\eta_0}{2}.$$

Sea  $n_0 = \max_{1 \leq j \leq J} n_j$ . Luego por (36), para todo  $n \geq n_0$  se tiene

$$\inf_{\boldsymbol{\lambda} \in K} \frac{1}{n} \sum_{i=1}^n g(\boldsymbol{\lambda}, \mathbf{v}_i) > \frac{\eta_0}{2}.$$

Esto prueba (a). Análogamente se prueba (b).

Probaremos (c). Sea  $\varepsilon > 0$ , como  $E(g(\boldsymbol{\lambda}, \mathbf{v})) - E(g(\boldsymbol{\lambda}, \mathbf{v})) - \varepsilon) < 0$  por la parte (b) con probabilidad 1 existe  $n_1$  tal que para todo  $n \geq n_1$

$$\sup_{\boldsymbol{\lambda} \in K} \frac{1}{n} \sum_{i=1}^n (g(\boldsymbol{\lambda}, \mathbf{v}_i) - E(g(\boldsymbol{\lambda}, \mathbf{v}_i)) - \varepsilon) < 0.$$

Similarmente por parte (a), con probabilidad 1, existe  $n_2$  tal que para todo  $n \geq n_2$

$$\inf_{\boldsymbol{\lambda} \in K} \frac{1}{n} \sum_{i=1}^n (g(\boldsymbol{\lambda}, \mathbf{v}_i) - E(g(\boldsymbol{\lambda}, \mathbf{v}_i)) + \varepsilon) > 0.$$

Si definimos  $n_3 = \max(n_1, n_2)$ , luego para todo  $n \geq n_3$  y para todo  $\boldsymbol{\lambda} \in K$  se tiene

$$-\varepsilon \leq \frac{1}{n} \sum_{i=1}^n g(\boldsymbol{\lambda}, \mathbf{v}_i) - E(g(\boldsymbol{\lambda}, \mathbf{v}_i)) \leq \varepsilon$$

o equivalentemente para todo  $n \geq n_3$  con probabilidad 1 se cumple que

$$\sup_{\boldsymbol{\lambda} \in K} \left| \frac{1}{n} \sum_{i=1}^n g(\boldsymbol{\lambda}, \mathbf{v}_i) - E(g(\boldsymbol{\lambda}, \mathbf{v}_i)) \right| \leq \varepsilon$$

y esto prueba (c).  $\square$

Llamemos  $\widehat{\gamma}_n(\boldsymbol{\lambda})$  el valor obtenido que resuelve (33) cuando  $z_i = \boldsymbol{\lambda}' \mathbf{x}_i$  y  $w_i = \boldsymbol{\beta}'_0 \mathbf{x}_i$ ,  $1 \leq i \leq n$ . Mostraremos que

$$\sup_{\|\boldsymbol{\lambda}\|=1} |\widehat{\gamma}_n(\boldsymbol{\lambda})| \xrightarrow{c.s.} 0.$$

En lo que sigue denotamos

$$K_p = \{\boldsymbol{\lambda} \in \mathbb{R}^p : \|\boldsymbol{\lambda}\| = 1\}. \quad (37)$$

**Teorema 6.1** *Supongamos que se cumplen P1-P4 y que  $(\mathbf{x}, y)$  satisface el modelo (1). Entonces*

$$\sup_{\boldsymbol{\lambda} \in K_p} |\widehat{\gamma}_n(\boldsymbol{\lambda})| \longrightarrow_{c.s.} 0.$$

**Demostración.** Llamemos

$$\psi(y, \mathbf{x}, \boldsymbol{\lambda}, \gamma) = \eta(y, \pi(\gamma \boldsymbol{\lambda}' \mathbf{x} + \boldsymbol{\beta}'_0 \mathbf{x})) \kappa(\boldsymbol{\lambda}' \mathbf{x}).$$

Definamos

$$r_n(\gamma, \boldsymbol{\lambda}) = \frac{1}{n} \sum_{i=1}^n \psi(y_i, \mathbf{x}_i, \boldsymbol{\lambda}, \gamma).$$

Luego por P1-P4 y el Lema 6.1 (c)

$$r_n(\gamma, 0) = \frac{1}{n} \sum_{i=1}^n \psi(y_i, \mathbf{x}_i, \boldsymbol{\lambda}, 0) \longrightarrow_{c.s.} E(\psi(y, \mathbf{x}, \boldsymbol{\lambda}, 0)) = 0$$

uniformemente para todo  $\boldsymbol{\lambda} \in K_p$ .

Llamemos

$$\eta_\pi(y, \pi) = \frac{\partial \eta(y, \pi)}{\partial \pi}, \quad \psi_\gamma(y, \mathbf{x}, \boldsymbol{\lambda}, \gamma) = \frac{\partial \psi(y, \mathbf{x}, \boldsymbol{\lambda}, \gamma)}{\partial \gamma}.$$

Usando P1-P4 se obtiene que

$$\eta_\pi(y, \pi(\boldsymbol{\beta}'_0 \mathbf{x})) \kappa(\boldsymbol{\lambda}' \mathbf{x}) \pi'(\boldsymbol{\beta}'_0 \mathbf{x})(\boldsymbol{\lambda}' \mathbf{x}) \leq 0$$

y

$$P(\eta_\pi(y, \pi(\boldsymbol{\beta}'_0 \mathbf{x})) \kappa(\boldsymbol{\lambda}' \mathbf{x}) \pi'(\boldsymbol{\beta}'_0 \mathbf{x})(\boldsymbol{\lambda}' \mathbf{x}) < 0) > 0$$

y por lo tanto para todo  $\boldsymbol{\lambda} \in K_p$ ,

$$E(\psi_\gamma(y, \mathbf{x}, \boldsymbol{\lambda}, 0)) = E(\eta_\pi(y, \pi(\boldsymbol{\beta}'_0 \mathbf{x})) \kappa(\boldsymbol{\lambda}' \mathbf{x}) \pi'(\boldsymbol{\beta}'_0 \mathbf{x})(\boldsymbol{\lambda}' \mathbf{x})) < 0. \quad (38)$$

Luego, dado  $\varepsilon > 0$  tenemos

$$E(\psi(y, \mathbf{x}, \boldsymbol{\lambda}, \varepsilon)) < 0, \quad E(\psi(y, \mathbf{x}, \boldsymbol{\lambda}, -\varepsilon)) > 0.$$

Luego por Lema 6.1 existe  $n_0$  tal que para todo  $n \geq n_0$

$$\sup_{\boldsymbol{\lambda} \in K_p} r_n(\varepsilon, \boldsymbol{\lambda}) < 0 \quad \text{y} \quad \inf_{\boldsymbol{\lambda} \in K_p} r_n(-\varepsilon, \boldsymbol{\lambda}) > 0.$$

Como  $r_n(\gamma, \boldsymbol{\lambda})$  es no creciente como función de  $\gamma$  obtenemos

$$\sup_{\boldsymbol{\lambda} \in K_p} \sup_{\gamma \geq \varepsilon} r_n(\gamma, \boldsymbol{\lambda}) < 0, \quad \inf_{\boldsymbol{\lambda} \in K_p} \inf_{\gamma \leq -\varepsilon} r_n(\gamma, \boldsymbol{\lambda}) > 0$$

y luego como  $\widehat{\gamma}_n(\boldsymbol{\lambda})$  está definido por  $r_n(\widehat{\gamma}_n(\boldsymbol{\lambda}), \boldsymbol{\lambda}) = 0$ , el Teorema queda demostrado.  $\square$



## 6.2. Tasa de consistencia

Dado un espacio métrico compacto  $S$  con métrica  $d$ , llamaremos  $C(S)$  al espacio de funciones continuas a valores reales definidas en  $S$  y  $C(S)^*$ , al espacio de los funcionales lineales continuos sobre  $C(S)$ .

Dada una métrica  $\rho$  sobre  $S$  se denota por  $N_\rho(S, u)$  al menor número de esferas de radio menor o igual que  $u$  que cubren a  $S$ .

Sea  $X$  una variable definida sobre un espacio de probabilidad  $(\Omega, \mathcal{F}, P)$  con valores en  $C(S)$ , el espacio de funciones continuas a valores reales sobre  $S$ . Se dirá que la distribución de  $X$  satisface el Teorema Central del Límite si dada una sucesión de variables aleatorias independientes  $\{X_i, i \geq 2\}$  con la misma distribución que  $X$  entonces se tiene que

$$\frac{1}{n^{1/2}} \sum_{i=1}^n X_i$$

converge en distribución a un proceso gaussiano.

Necesitaremos el siguiente Teorema de Jain y Marcus (1975).

**Teorema 6.2** *Sea  $S$  un espacio métrico compacto con métrica  $d$ . Sea  $X$  una variable aleatoria definida sobre un espacio de probabilidad  $(\Omega, \mathcal{F}, P)$  con valores en  $C(S)$  que satisface*

$$E(f(X)) = 0 \text{ para todo } f \in C(S)^* \quad (39)$$

y

$$\sup_{\lambda \in S} E[X^2(\lambda)] < \infty. \quad (40)$$

Supongamos que existe una variable aleatoria no negativa  $M, E(M^2) < \infty$ , y una métrica  $\rho$  en  $S$ , que es continua con respecto a  $d$ , tal que dados  $s, t \in S, \omega \in \Omega$ ,

$$|X(s, \omega) - X(t, \omega)| \leq M(\omega)\rho(s, t). \quad (41)$$

Si además

$$\int \sqrt{\ln N_\rho(S, u)} du < \infty, \quad (42)$$

entonces la distribución de  $X$  satisface el Teorema Central del Límite.

**Observación 6.1** *La condición (42) es equivalente a*

$$\sum_{n=1}^{\infty} 2^{-n} \sqrt{\ln N_\rho(S, 2^{-n})} < \infty.$$

**Lema 6.2** Sea  $(\mathbf{x}, y)$  un vector aleatorio en  $\mathbb{R}^{p+1}$  satisfaciendo el modelo (1). Supongamos que se cumplen **P1-P4**. Dada una muestra aleatoria de tamaño  $n$ ,  $\{(\mathbf{x}_i, y_i), 1 \leq i \leq n\}$  con la misma distribución que  $(\mathbf{x}, y)$ , se verifica que

$$\sqrt{n} \sup_{\|\boldsymbol{\lambda}\|=1} \left| \frac{1}{n} \sum_{i=1}^n \eta(y_i, \pi(\boldsymbol{\beta}'_0 \mathbf{x}_i)) \kappa(\boldsymbol{\lambda}' \mathbf{x}_i) \right| \quad (43)$$

está acotada en probabilidad.

**Demostración.** La prueba de este lema surge de la aplicación del Teorema 6.2. Consideramos el compacto  $K_p$  definido en (37) y sea  $C(K_p)$  el espacio de las funciones continuas a valores reales definidas sobre  $K_p$ .

La sucesión  $g_i(\boldsymbol{\lambda}) = \eta(y_i, \pi(\boldsymbol{\beta}'_0 \mathbf{x}_i)) \kappa(\boldsymbol{\lambda}' \mathbf{x}_i)$  es un conjunto de variables i.i.d. con valores en  $C(K_p)$  donde  $K_p$  está provisto la métrica Euclídea. Sea  $g = \eta(y, \pi(\boldsymbol{\beta}'_0 \mathbf{x})) \kappa(\boldsymbol{\lambda}' \mathbf{x})$ , demostraremos que  $g$  cumple las condiciones del Teorema 6.2.

Por el Teorema de Riesz (Kolmogórov y Fomín, 1975, pág. 413) dado  $f$  un funcional lineal continuo existe una medida  $\nu$  sobre  $K_p$  tal que

$$f(g) = \int g(\boldsymbol{\lambda}) d\nu(\boldsymbol{\lambda}).$$

Luego

$$\begin{aligned} Ef(g) &= E \left( \int \eta(y, \pi(\boldsymbol{\beta}'_0 \mathbf{x})) \kappa(\boldsymbol{\lambda}' \mathbf{x}) d\nu(\boldsymbol{\lambda}) \right) \\ &= E \left( \eta(y, \pi(\boldsymbol{\beta}'_0 \mathbf{x})) \int \kappa(\boldsymbol{\lambda}' \mathbf{x}) d\nu(\boldsymbol{\lambda}) \right) \\ &= E \left( E \left( \eta(y, \pi(\boldsymbol{\beta}'_0 \mathbf{x})) \left[ \int \kappa(\boldsymbol{\lambda}' \mathbf{x}) d\nu(\boldsymbol{\lambda}) \right] \middle| \mathbf{x} \right) \right) \\ &= E \left( \left[ \int \kappa(\boldsymbol{\lambda}' \mathbf{x}) d\nu(\boldsymbol{\lambda}) \right] E(\eta(y, \pi(\boldsymbol{\beta}'_0 \mathbf{x})) | \mathbf{x}) \right). \end{aligned}$$

Como por **P1**,  $E(\eta(y, \pi(\boldsymbol{\beta}'_0 \mathbf{x})) | \mathbf{x}) = 0$  resulta  $Ef(g) = 0$ . Luego se cumple la condición (39).

Como  $\eta$  y  $\kappa$  son acotadas también resulta  $\sup_{\boldsymbol{\lambda} \in K_p} E(g^2(\boldsymbol{\lambda})) < \infty$  y luego se verifica (40).

Por otro lado por la condición **P2** existe  $m_0$  tal que  $|\kappa(t_2) - \kappa(t_1)| \leq m_0 |t_2 - t_1|$  entonces

$$\begin{aligned} |g(\boldsymbol{\lambda}_2) - g(\boldsymbol{\lambda}_1)| &\leq |\eta(y, \pi(\boldsymbol{\beta}'_0 \mathbf{x}))| |\kappa(\boldsymbol{\lambda}'_2 \mathbf{x}) - \kappa(\boldsymbol{\lambda}'_1 \mathbf{x})| \\ &\leq m_0 \|\eta\|_\infty \|\mathbf{x}\| \|\boldsymbol{\lambda}_2 - \boldsymbol{\lambda}_1\|. \end{aligned}$$

Luego por **P4** la condición (41) del Teorema 6.2 también se cumple.

Además si  $\rho$  es la distancia Euclídea se tiene que  $N(K_p, 2^{-n}) \leq C_p 2^{pn}$  donde  $C_p$  es una constante. Luego

$$\sum_{i=1}^n 2^{-n} \sqrt{\ln N_\rho(K_p, 2^{-n})} < \infty$$

y por la Observación 6.1 se cumple la condición (42).

Entonces por el Teorema 6.2

$$A_n = \frac{1}{n^{1/2}} \sum_{i=1}^n g_i(\boldsymbol{\lambda})$$

converge en distribución a un proceso  $D$  Gaussiano en el espacio de funciones continuas sobre  $K_p$ . Obsérvese que dado  $h$  en  $C(K_p)$ , la función  $b(h) = \sup_{\boldsymbol{\lambda} \in K_p} h(\boldsymbol{\lambda})$  es continua. Luego

$$\sup_{\boldsymbol{\lambda} \in K_p} A_n(\boldsymbol{\lambda}) \rightarrow_D \sup_{\boldsymbol{\lambda} \in K_p} D(\boldsymbol{\lambda})$$

y por lo tanto  $\sup_{\boldsymbol{\lambda} \in K_p} A_n(\boldsymbol{\lambda})$  está acotado en probabilidad. Esto demuestra el Lema 6.2.  $\square$

**Teorema 6.3** *Sea  $(\mathbf{x}, y)$  un vector aleatorio en  $\mathbb{R}^{p+1}$  satisfaciendo el modelo (1). Supongamos que se cumplen **P1-P4**. Dada una muestra aleatoria de tamaño  $n$ ,  $\{(\mathbf{x}_i, y_i), 1 \leq i \leq n\}$  para  $(\mathbf{x}, y)$  definamos  $\hat{\gamma}_n(\boldsymbol{\lambda})$  como el valor de  $\gamma$  que satisface*

$$\frac{1}{n} \sum_{i=1}^n \eta(y_i, \pi(\gamma \boldsymbol{\lambda}' \mathbf{x}_i + \boldsymbol{\beta}'_0 \mathbf{x}_i)) \kappa(\boldsymbol{\lambda}' \mathbf{x}_i) = 0. \quad (44)$$

Luego se verifica que

$$\sqrt{n} \sup_{\boldsymbol{\lambda} \in K_p} |\hat{\gamma}_n(\boldsymbol{\lambda})| = O_p(1).$$

**Demostración.** Desarrollando por Taylor (44) alrededor de  $\gamma = 0$ , se obtiene

$$A_n(\boldsymbol{\lambda}) + B_n(\boldsymbol{\lambda}) \hat{\gamma}_n(\boldsymbol{\lambda}) + \frac{1}{2} C_n(\boldsymbol{\lambda}) \hat{\gamma}_n(\boldsymbol{\lambda})^2 = 0,$$

donde

$$A_n(\boldsymbol{\lambda}) = \frac{1}{n} \sum_{i=1}^n \eta(y_i, \pi(\boldsymbol{\beta}'_0 \mathbf{x}_i)) \kappa(\boldsymbol{\lambda}' \mathbf{x}_i),$$

$$B_n(\boldsymbol{\lambda}) = \frac{1}{n} \sum_{i=1}^n \eta_\pi(y_i, \pi(\boldsymbol{\beta}'_0 \mathbf{x}_i)) \kappa(\boldsymbol{\lambda}' \mathbf{x}_i) \pi'(\boldsymbol{\beta}'_0 \mathbf{x}_i) (\boldsymbol{\lambda}' \mathbf{x}_i)$$

y

$$\begin{aligned}
C_n(\boldsymbol{\lambda}) &= \frac{1}{n} \sum_{i=1}^n \left[ \eta_{\pi\pi}(y_i, \pi(\gamma^*(\gamma, \boldsymbol{\lambda})\boldsymbol{\lambda}'\mathbf{x}_i + \boldsymbol{\beta}'_0\mathbf{x}_i))\kappa(\boldsymbol{\lambda}'\mathbf{x}_i) \cdot \right. \\
&\quad \left. \pi'^2(\gamma^*(\gamma, \boldsymbol{\lambda})\boldsymbol{\lambda}'\mathbf{x}_i + \boldsymbol{\beta}'_0\mathbf{x}_i)(\boldsymbol{\lambda}'\mathbf{x}_i)^2 \right] \\
&+ \frac{1}{n} \sum_{i=1}^n \left[ \eta_{\pi}(y_i, \pi(\gamma^*(\gamma, \boldsymbol{\lambda})\boldsymbol{\lambda}'\mathbf{x}_i + \boldsymbol{\beta}'_0\mathbf{x}_i))\kappa(\boldsymbol{\lambda}'\mathbf{x}_i) \cdot \right. \\
&\quad \left. \pi''(\gamma^*(\gamma, \boldsymbol{\lambda})\boldsymbol{\lambda}'\mathbf{x}_i + \boldsymbol{\beta}'_0\mathbf{x}_i)(\boldsymbol{\lambda}'\mathbf{x}_i)^2 \right]
\end{aligned}$$

con  $\gamma^*(\gamma, \boldsymbol{\lambda})$  entre  $\widehat{\gamma}_n(\boldsymbol{\lambda})$  y 0 y donde  $\eta_{\pi}$  y  $\eta_{\pi\pi}$  denotan la derivada primera y segunda de  $\eta$  respecto de  $\pi$  respectivamente.

A partir de este desarrollo se tiene que

$$\widehat{\gamma}_n(\boldsymbol{\lambda}) = -\frac{A_n(\boldsymbol{\lambda})}{B_n(\boldsymbol{\lambda}) + \frac{1}{2}C_n(\boldsymbol{\lambda})\widehat{\gamma}_n(\boldsymbol{\lambda})},$$

luego

$$\begin{aligned}
\sqrt{n} \sup_{\boldsymbol{\lambda} \in K_p} |\widehat{\gamma}_n(\boldsymbol{\lambda})| &= \sqrt{n} \sup_{\boldsymbol{\lambda} \in K_p} \left| \frac{A_n(\boldsymbol{\lambda})}{-B_n(\boldsymbol{\lambda}) - \frac{1}{2}C_n(\boldsymbol{\lambda})\widehat{\gamma}_n(\boldsymbol{\lambda})} \right| \\
&\leq \frac{\sqrt{n} \sup_{\boldsymbol{\lambda} \in K_p} |A_n(\boldsymbol{\lambda})|}{\inf_{\boldsymbol{\lambda} \in K_p} \left| -B_n(\boldsymbol{\lambda}) - \frac{1}{2}C_n(\boldsymbol{\lambda})\widehat{\gamma}_n(\boldsymbol{\lambda}) \right|}.
\end{aligned}$$

Por Lema 6.2 se tiene que

$$\sqrt{n} \sup_{\boldsymbol{\lambda} \in K_p} |A_n(\boldsymbol{\lambda})| = O_p(1).$$

Luego, en virtud del Teorema 6.1 bastará mostrar que

$$\lim_{n \rightarrow \infty} \inf_{\boldsymbol{\lambda} \in K_p} (-B_n(\boldsymbol{\lambda})) > 0 \text{ c.s.} \quad (45)$$

y

$$\sup_{\|\boldsymbol{\lambda}\|=1} |C_n(\boldsymbol{\lambda})| = O_p(1). \quad (46)$$

Por (38)

$$E(\eta_{\pi}(y, \pi(\boldsymbol{\beta}'_0\mathbf{x}))\kappa(\boldsymbol{\lambda}'\mathbf{x})\pi'(\boldsymbol{\beta}'_0\mathbf{x})(\boldsymbol{\lambda}'\mathbf{x})) < 0.$$

(45) resulta de (38) y Lema 6.1.

Por otra parte existe  $k_0$  tal que

$$\begin{aligned}
|C_n(\boldsymbol{\lambda})| &\leq \frac{k_0}{n} \sum_{i=1}^n (\boldsymbol{\lambda}'\mathbf{x}_i)^2 \\
&\leq \frac{k_0}{n} \sum_{i=1}^n \|\mathbf{x}_i\|^2.
\end{aligned}$$

Como

$$\lim_{n \rightarrow \infty} \sum_{i=1}^n \|\mathbf{x}_i\|^2 = E(\|\mathbf{x}\|^2) < \infty$$

obtenemos (46). Esto prueba el Teorema.  $\square$

Necesitamos la siguiente hipótesis sobre la escala  $s$  usada en la definición del estimador de proyección.

**P5.** Existen  $a > 0$  y  $b > 0$  tal que  $a \leq s(\mathcal{L}(\boldsymbol{\lambda}'\mathbf{x}, G_0)) \leq b$  para todo  $\boldsymbol{\lambda} \in K_p$ .

Por ejemplo se puede demostrar que **P5** se cumple si  $s$  es una M-escala con punto de ruptura 0.5 y  $P_{G_0}(\boldsymbol{\lambda}'\mathbf{x} = 0) < 0.5$  para todo  $\boldsymbol{\lambda} \in K_p$ .

**Teorema 6.4** *Sea  $(\mathbf{x}, y)$  un vector aleatorio en  $\mathbb{R}^{p+1}$  satisfaciendo el modelo (1). Supongamos que se cumplen **P1-P4**. Consideremos el estimador  $\mathbf{T}$  de proyección basado en el funcional inicial  $T_0(F)$  para el modelo (12) definido por el valor  $\gamma$  tal que*

$$E_F \eta(y, \pi(\gamma z + w)) \kappa(z) = 0$$

*y por un funcional de escala  $s$  satisfaciendo **P5**. Sea  $F_n$  la distribución empírica correspondiente a una muestra aleatoria de tamaño  $n$ ,  $\{(\mathbf{x}_i, y_i), 1 \leq i \leq n\}$  y  $\hat{\boldsymbol{\beta}}_n = \mathbf{T}(F_n)$ . Luego  $\zeta_n = n^{1/2}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)$  está acotado en probabilidad.*

**Demostración.** Como  $T_0$  es regresión equivariante, dados  $\boldsymbol{\lambda}, \boldsymbol{\beta} \in \mathbb{R}^p, \gamma \in \mathbb{R}$ ,

$$T_0 [\mathcal{L}(y, \boldsymbol{\lambda}'\mathbf{x}, \boldsymbol{\beta}'\mathbf{x}, H_n)] - \gamma = T_0 [\mathcal{L}(y, \boldsymbol{\lambda}'\mathbf{x}, (\boldsymbol{\beta} + \gamma\boldsymbol{\lambda})'\mathbf{x}, H_n)]. \quad (47)$$

Podemos escribir  $\hat{\boldsymbol{\beta}}_n = \boldsymbol{\beta}_0 + \frac{k_n}{\sqrt{n}} \boldsymbol{\mu}_n$ , con  $\|\boldsymbol{\mu}_n\| = 1$  y  $k_n = \|\zeta_n\|$ . Tenemos que mostrar que  $k_n$  está acotado en probabilidad. Sea  $H_n$  la distribución empírica de la muestra. Como de acuerdo al Teorema 6.3 se tiene que

$$n^{1/2} \sup_{\boldsymbol{\lambda} \in K_p} |T_0 (\mathcal{L}(y, \boldsymbol{\lambda}'\mathbf{x}, \boldsymbol{\beta}'_0\mathbf{x}, H_n))|$$

está acotada en probabilidad, por la definición del estimador de proyección y P5 se tiene que

$$U_n = n^{1/2} T_0 \left( \mathcal{L} \left( y, \boldsymbol{\mu}'_n \mathbf{x}, \left( \boldsymbol{\beta}_0 + \frac{k_n}{n^{1/2}} \boldsymbol{\mu}_n \right)' \mathbf{x}, H_n \right) \right)$$

está acotado en probabilidad. Usando la regresión equivarianza de  $T_0$  tenemos que

$$U_n = n^{1/2} T_0 (\mathcal{L}(y, \boldsymbol{\mu}'_n \mathbf{x}, \boldsymbol{\beta}'_0 \mathbf{x}, H_n)) - k_n$$

y

$$k_n = n^{1/2}T_0(\mathcal{L}(y, \boldsymbol{\mu}'_n \mathbf{x}, \boldsymbol{\beta}'_0 \mathbf{x}, H_n)) - U_n.$$

Como por el Teorema 6.3

$$n^{1/2}T_0(\mathcal{L}(y, -\boldsymbol{\mu}'_n \mathbf{x}, \boldsymbol{\beta}'_0 \mathbf{x}, H_n))$$

está acotado en probabilidad, resulta  $k_n$  acotado en probabilidad.  $\square$

## 7. Algoritmo

Para obtener el estimador de proyección, utilizamos un algoritmo basado en submuestras que nos permitirá calcular un valor aproximado. No poseemos algoritmos para el cálculo del valor exacto de estos estimadores. El esquema del algoritmo utilizado es similar al usado por Maronna y Yohai (1993) para calcular los P estimadores de un modelo lineal. Vamos a suponer que el estimador inicial es de la forma (33) y que se satisfacen las propiedades **P1-P4**.

Sea  $\{(\mathbf{x}_i, y_i), 1 \leq i \leq n\}$  una muestra de tamaño  $n$ . Una forma de encontrar un valor aproximado al estimador de proyección será determinando un conjunto finito de posibles candidatos para  $\boldsymbol{\beta}$  que llamaremos  $D_n \subseteq \mathbb{R}^p$  y para cada  $\boldsymbol{\beta} \in D_n$ , un conjunto finito de direcciones de norma 1 en  $\mathbb{R}^p$  que llamaremos  $D_n^*(\boldsymbol{\beta})$ . Tanto  $D_n$  como  $D_n^*(\boldsymbol{\beta})$  se obtendrán utilizando la muestra.

Supongamos que el estimador inicial resuelve una ecuación como la dada en (33). Luego para cada  $\boldsymbol{\beta} \in \mathbb{R}^p$  y cada  $\boldsymbol{\lambda} \in D_n^*(\boldsymbol{\beta})$ , aplicamos el estimador inicial a la muestra  $\{(y_i, \boldsymbol{\lambda}'\mathbf{x}_i, \boldsymbol{\beta}'\mathbf{x}_i), 1 \leq i \leq n\}$ , y el valor obtenido lo notaremos por  $\gamma_n(\boldsymbol{\beta}, \boldsymbol{\lambda})$ . Este estimador resulta como solución de

$$\sum_{i=1}^n \eta(y_i, \pi(\gamma \boldsymbol{\lambda}'\mathbf{x}_i + \boldsymbol{\beta}'\mathbf{x}_i)) \kappa(\boldsymbol{\lambda}'\mathbf{x}_i) = 0.$$

Además calculamos una escala  $s$  de  $\boldsymbol{\lambda}'\mathbf{x}_i$ ,  $1 \leq i \leq n$  que llamaremos  $s_n(\boldsymbol{\lambda})$ .

Sean

$$A_n(\boldsymbol{\beta}, \boldsymbol{\lambda}) = |\gamma_n(\boldsymbol{\beta}, \boldsymbol{\lambda})| s_n(\boldsymbol{\lambda})$$

y

$$C_n(\boldsymbol{\beta}) = \max_{\boldsymbol{\lambda} \in D_n^*(\boldsymbol{\beta})} A_n(\boldsymbol{\beta}, \boldsymbol{\lambda}).$$

Luego el estimador aproximado de proyección se obtiene como

$$\hat{\boldsymbol{\beta}}_n = \arg \min_{\boldsymbol{\beta} \in D_n} C_n(\boldsymbol{\beta}). \quad (48)$$

Es claro que si  $D_n = \mathbb{R}^p$  y  $D_n^*(\boldsymbol{\beta}) = \{\boldsymbol{\lambda} \in \mathbb{R}^p : \|\boldsymbol{\lambda}\| = 1\}$  para todo  $\boldsymbol{\beta}$ , el estimador definido en (48) es el estimador de proyección exacto.

Para cada  $\boldsymbol{\beta} \in D_n$  consideremos

$$D_n^*(\boldsymbol{\beta}) = \{\boldsymbol{\lambda} = (\boldsymbol{\beta}^* - \boldsymbol{\beta}) / \|\boldsymbol{\beta}^* - \boldsymbol{\beta}\|, \boldsymbol{\beta}^* \in D_n, \boldsymbol{\beta}^* \neq \boldsymbol{\beta}\}. \quad (49)$$

Cada elemento de  $D_n$  se obtiene de la siguiente manera:

1. Tomamos al azar una submuestra de tamaño  $n^*$ ,  $n > n^* > p$  del conjunto de datos originales, y calculamos el estimador de máxima verosimilitud con esta submuestra que llamaremos  $\boldsymbol{\beta}^{(0)}$ .

2. Corregimos el estimador obtenido en el paso 1 multiplicándolo por un valor de  $\gamma$  que resulta de la solución de

$$\sum_{i=0}^n \eta(y_i, \pi(\gamma \boldsymbol{\beta}^{(0)' \mathbf{x}_i})) \kappa(\boldsymbol{\beta}^{(0)' \mathbf{x}_i}) = 0$$

Luego  $\boldsymbol{\beta}^{(1)} = \gamma \boldsymbol{\beta}^{(0)}$ .

3. Calculamos las “deviances”

$$D_i = -y_i \ln(\pi(\boldsymbol{\beta}^{(1)' \mathbf{x}_i}) - (1 - y_i) \ln(1 - \pi(\boldsymbol{\beta}^{(1)' \mathbf{x}_i}))$$

y las ordenamos obteniendo  $D_{i_1} \leq \dots \leq D_{i_n}$ . Sea  $m = \lfloor n/2 \rfloor$ . Luego obtenemos el estimador de máxima verosimilitud correspondiente a la submuestra  $(\mathbf{x}_{i_1}, y_{i_1}), \dots, (\mathbf{x}_{i_m}, y_{i_m})$ , que llamamos  $\boldsymbol{\beta}^{(2)}$ . Este es el candidato obtenido.

Repetiendo los pasos 1-3  $N$  veces obtenemos  $N$  candidatos que forman el conjunto  $D_n$ . El valor de  $N$  se calcula de manera que la probabilidad de que alguna de las submuestras obtenidas en el paso 1 no contenga un outlier sea por lo menos  $1 - \alpha$  donde  $\alpha$  es un número pequeño, por ejemplo  $\alpha=0.01$ . En Maronna et al. (2006, cap. 5, pág. 137) se muestra que este número depende de  $p$  y de la proporción  $\varepsilon$  de outliers que supongamos puede haber en la muestra y que  $N$  debe elegirse mayor que  $|\log \alpha| / (1 - \varepsilon)^p$ . Llamemos  $\boldsymbol{\beta}_i$  al  $i$ -ésimo candidato hallado y

$$\lambda_{ij} = \frac{\boldsymbol{\beta}_j - \boldsymbol{\beta}_i}{\|\boldsymbol{\beta}_j - \boldsymbol{\beta}_i\|}.$$

Luego  $D_n = \{\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_N\}$  y  $D_n^*(\boldsymbol{\beta}_i) = \{\boldsymbol{\lambda}_{ij}, : 1 \leq j \leq N, i \neq j\}$ .

En realidad para calcular  $\hat{\boldsymbol{\beta}}_n$  no es necesario calcular los  $N(N - 1)$  valores de  $A_n(\boldsymbol{\beta}_i, \boldsymbol{\lambda}_{ij})$ ,  $1 \leq i, j \leq N$ ,  $i \neq j$ . El procedimiento inductivo es el siguiente: se comienza con  $\boldsymbol{\beta}_1$  y obtenemos  $C_n(\boldsymbol{\beta}_1)$ . Luego se calculan los valores  $A_n(\boldsymbol{\beta}_2, \boldsymbol{\lambda}_j)$  hasta que para algún  $j$  se tiene

$$A_n(\boldsymbol{\beta}_2, \boldsymbol{\lambda}_j) \geq C_n(\boldsymbol{\beta}_1). \quad (50)$$

Cuando esto ocurre, sin necesidad de calcular  $C_n(\boldsymbol{\beta}_2)$ , sabemos que  $C_n(\boldsymbol{\beta}_2) \geq C_n(\boldsymbol{\beta}_1)$  y que  $r_2 = \min_{1 \leq i \leq 2} C_n(\boldsymbol{\beta}_i) = C_n(\boldsymbol{\beta}_1)$ . Si (50) nunca ocurre se tendrá

$$r_2 = \min_{1 \leq j \leq N, j \neq 2} A_n(\boldsymbol{\beta}_2, \boldsymbol{\lambda}_j).$$

El valor  $i_2$  se define por

$$i_2 = \arg \min_{1 \leq i \leq 2} C_n(\boldsymbol{\beta}_i)$$



o equivalentemente por

$$C_n(\beta_{i_2}) = \min_{1 \leq i \leq 2} C_n(\beta_i).$$

Supongamos ahora que para  $h < N$  tenemos calculados

$$r_h = \min_{1 \leq i \leq h} C_n(\beta_i)$$

y

$$i_h = \arg \min_{1 \leq i \leq h} C_n(\beta_i).$$

Luego se calculan  $A_n(\beta_{h+1}, \lambda_j)$  hasta que para algún  $j$  se tiene

$$A_n(\beta_{h+1}, \lambda_j) \geq r_h. \quad (51)$$

Cuando esto ocurre, sin necesidad de calcular  $C_n(\beta_{h+1})$  sabemos que  $r_{h+1} = r_h$  y  $i_{h+1} = i_h$ . Si (51) nunca ocurre se tendrá

$$r_{h+1} = \max_{1 \leq j \leq N, j \neq h+1} A_n(\beta_{h+1}, \lambda_j)$$

y  $i_{h+1} = h + 1$ . Este procedimiento inductivo lo continuamos hasta obtener  $r_N$  y  $i_N$ . El estimador de proyección computado con este algoritmo será  $\beta_{i_N}$ , es decir el valor que minimiza  $C_n(\beta_i)$  para  $1 \leq i \leq N$ .

## 8. Simulaciones

Con la finalidad de estudiar la eficiencia y robustez de los estimadores de proyección en el modelo logístico y compararlo con otras propuestas se realizó un estudio de Monte Carlo.

Se generaron muestras del modelo (1) con

$$\pi(\mathbf{x}'\boldsymbol{\beta}_0) = \frac{e^{\mathbf{x}'\boldsymbol{\beta}_0}}{1 + e^{\mathbf{x}'\boldsymbol{\beta}_0}}$$

usando como valor del parámetro  $\boldsymbol{\beta}_0 = (0, 2, 2)'$ . Se tomó  $\mathbf{x} = (1, x_1, x_2)'$ , donde  $(x_1, x_2)$  tiene una distribución  $N_2(0, I_2)$ . (Se denota por  $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  la distribución normal multivariada de dimensión  $p$  con media  $\boldsymbol{\mu}$  y matriz de covarianza  $\boldsymbol{\Sigma}$ , y por  $I_p$  la matriz identidad  $p \times p$ ). La  $i$ -ésima observación de la muestra será  $(\mathbf{x}_i, y_i)$  con  $\mathbf{x}_i = (1, x_{i2}, x_{i3})$ . Llamaremos  $\mathbf{v}_i^* = (x_{i2}, x_{i3})'$ .

Se analizan los siguientes estimadores:

- (i) Máxima verosimilitud (**ML**), es decir, la solución de las siguientes ecuaciones

$$\sum_{i=1}^n (y_i - \pi(\mathbf{x}_i'\boldsymbol{\beta}))x_{ij} = 0, j = 1, 2, 3.$$

- (ii) Máxima verosimilitud pesado (**WML**), es decir, la solución de las siguientes ecuaciones

$$\sum_{i=1}^n w(\mathbf{x}_i)(y_i - \pi(\mathbf{x}_i'\boldsymbol{\beta}))x_{ij} = 0, j = 1, 2, 3$$

con  $w(\mathbf{x}) = W(RD_i)$  y  $W(t) = I(t^2 \leq \chi_{2, 0.975}^2)$  donde  $\chi_{p, \alpha}^2$  es el cuantil  $\alpha$  de una distribución  $\chi^2$  con  $p$  grados de libertad, y  $RD_i$  es la distancia de Mahalanobis dada por

$$RD_i^2 = (\mathbf{v}_i^* - \hat{\boldsymbol{\mu}})' \hat{\boldsymbol{\Sigma}}^{-1} (\mathbf{v}_i^* - \hat{\boldsymbol{\mu}})$$

donde  $\hat{\boldsymbol{\mu}}$  y  $\hat{\boldsymbol{\Sigma}}$  son los estimadores de posición y covarianza obtenidos aplicando el estimador de *mínimo determinante de la covarianza* (MCD) a la muestra  $\mathbf{v}_1^*, \dots, \mathbf{v}_n^*$  (Rousseeuw y Leroy, 1987, pág. 262) usando el 75 % de las observaciones.

- (iii) M-estimador propuesto por Bianco-Yohai (**BY**). Estos estimadores se definen por (4) y (5). Se utilizó la siguiente función  $\rho$  propuesta por Croux y Haesbroeck (2003)

$$\rho(t) = \begin{cases} te^{-\sqrt{d}} & \text{si } t \leq d \\ -2e^{-\sqrt{t}}(1 + \sqrt{t}) + e^{-\sqrt{d}}(2(1 + \sqrt{d}) + d) & \text{si } t > d \end{cases}$$

con  $d = 0.5$ . Para el cálculo de este estimador se utilizó el programa BYlogreg propuesto por Croux y Haesbroeck (<http://www.econ.kuleuven.be/public/NDBAE06/programs/roblog/BYlogreg.txt>).

(iv) M-estimador pesado (**WBY**). Se define por

$$\sum_{i=1}^n w(\mathbf{x}_i) [\rho(d(\boldsymbol{\beta}, y_i, \mathbf{x}_i)) + G(p(\mathbf{x}'_i \boldsymbol{\beta})) + G((1 - p(\mathbf{x}'_i \boldsymbol{\beta})))] \quad (52)$$

utilizando la misma función  $\rho$  que en BY y los mismos pesos descriptos para WLM. Se usaron dos estimadores del centro y la matriz de covarianza: un MCD con 75 % de las observaciones (**WBY**<sub>0.75</sub>) y un MCD con 50 % de las observaciones (**WBY**<sub>0.5</sub>). Para el cálculo de este estimador se utilizó el programa WBYlogreg propuesto por Croux y Haesbroeck (<http://www.econ.kuleuven.be/public/NDBAE06/programs/roblog/WBYlogreg.txt>).

(v) Estimadores de proyección donde el estimador inicial  $T_0$  es de la forma (17), con función  $\psi(y, z, w, \gamma)$  dada por (18), con  $\eta(y, \pi)$  de la forma (19) y usando para  $\kappa$  la función de Huber

$$\kappa(z) = \min(|z|, k) \text{signo}(z). \quad (53)$$

La escala utilizada fue

$$s(z_1, \dots, z_n) = \frac{\text{median}(z_1, \dots, z_n)}{0.6745}.$$

Se usaron dos valores de  $k$  en (53),  $k = 1$  (**PR**<sub>1</sub><sup>HU</sup>) y  $k = 0.8$  (**PR**<sub>0.8</sub><sup>HU</sup>).

(vi) Estimadores de proyección iguales que en (v) pero usando para  $\kappa$  una función redescendiente en la familia propuesta por Hampel

$$k(z) = \begin{cases} z & 0 \leq |z| \leq c_1 \\ c_1 \text{sign}(z) & c_1 < |z| \leq c_2 \\ \frac{c_1}{c_3 - c_2} (c_3 - |z|) \text{sign}(z) & c_2 < |z| \leq c_3 \\ 0 & |z| > c_3 \end{cases}.$$

Para este estimador (**PR**<sup>HA</sup>) se tomó  $c_1 = 1.1$ ,  $c_2 = 1.5$  y  $c_3 = 2.3$ .

El número de replicaciones en la simulación fue  $B = 500$ . Para poder comparar los comportamientos de los estimadores descriptos antes se usaron los siguientes criterios:

1. **Error cuadrático medio** (ECM). Dado un estimador  $\widehat{\boldsymbol{\beta}}$  del parámetro  $\boldsymbol{\beta}_0$  el ECM se define por

$$\text{ECM} = E(\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|^2)$$

y se estima por

$$\widehat{\text{ECM}} = \frac{1}{B} \sum_{i=1}^B \|\widehat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_0\|^2$$

donde  $\widehat{\boldsymbol{\beta}}_i$  es el estimador obtenido a partir de la  $i$ -ésima muestra generada.

2. **Error de clasificación** (EC). El error de clasificación de un estimador  $\widehat{\beta}$  se define por

$$EC = P\left(y = 1, \widehat{\beta}' \mathbf{x} < 0\right) + P\left(y = 0, \widehat{\beta}' \mathbf{x} > 0\right)$$

donde  $(\mathbf{x}, y)$  es una observación independiente de la muestra utilizada para calcular  $\widehat{\beta}$  y generada sin outliers. Para estimarlo, además de las  $B$  muestras de tamaño 100 utilizadas para calcular los valores  $\widehat{\beta}_i, 1 \leq i \leq B$ , generamos una muestra de tamaño  $M = 1000$  sin outliers que llamaremos  $(\mathbf{x}_j^*, y_j^*), 1 \leq j \leq M$ . Luego, EC es estimado por

$$\widehat{EC} = \frac{1}{BM} \sum_{i=1}^B \sum_{j=1}^M \left[ I\left(y_j^* = 1, \widehat{\beta}_i' \mathbf{x}_j^* < 0\right) + I\left(y_j^* = 0, \widehat{\beta}_i' \mathbf{x}_j^* > 0\right) \right].$$

Se utilizaron muestras sin contaminar de tamaño 100 y se le agregaron un número de outliers de modo que el porcentaje de contaminación se aproximase al 5% y al 10%.

Para obtener muestras de aproximadamente un 5% de contaminación se le agregaron a la muestra original 6 observaciones iguales de la forma  $(\mathbf{x}_i, 0)$  donde  $\mathbf{x}_i' = (1, l, l), 101 \leq i \leq 106$ , con  $l \in \{0.5 * j, 1 \leq j \leq 11\}$ . Los valores de  $\widehat{ECM}$  se muestran en la Tabla 1 y los de  $\widehat{EC}$  en la Tabla 2. La columna SC muestran los valores para una muestra de tamaño 100 sin outliers. Los valores en negrita corresponden a los máximos para distintos tamaños del outliers. A los efectos comparativos, el valor el porcentaje de errores de clasificación si se usara para clasificar el valor verdadero  $\beta_0$  es 17.70.

**Tabla 1. Errores cuadráticos medios con 5% de contaminación**

Estimador	Tamaño de la contaminación ( $l$ )					
	SC	0.5	1	1.5	2	2.5
ML	0.770	0.789	1.853	2.969	3.966	4.833
WML	0.815	0.843	1.944	<b>3.077</b>	2.345	0.912
BY	1.081	0.898	1.489	1.742	1.777	1.730
WBY <sub>0.75</sub>	1.133	0.947	1.565	<b>1.822</b>	1.296	1.136
WBY <sub>0.5</sub>	1.164	<b>1.695</b>	1.648	1.230	1.193	1.116
PR <sub>1</sub> <sup>HU</sup>	0.820	0.938	1.444	1.442	1.440	1.445
PR <sub>0.8</sub> <sup>HU</sup>	0.838	1.009	1.250	1.236	1.237	1.239
PR <sup>HA</sup>	0.853	1.014	<b>1.545</b>	0.822	0.831	0.844

Estimador	Tamaño de la contaminación ( $l$ )				
	3	3.5	4	4.5	5
ML	5.569	6.173	6.651	7.023	<b>7.307</b>
WML	0.804	0.804	0.804	0.804	0.804
BY	1.682	1.621	1.557	1.503	<b>1.858</b>
WBY <sub>0.75</sub>	1.118	1.118	1.118	1.118	1.118
WBY <sub>0.5</sub>	1.108	1.108	1.108	1.108	1.108
PR <sub>1</sub> <sup>HU</sup>	<b>1.449</b>	1.442	1.441	1.442	1.436
PR <sub>0.8</sub> <sup>HU</sup>	1.239	1.244	1.245	<b>1.250</b>	1.247
PR <sup>HA</sup>	0.851	0.836	0.832	0.840	0.844

**Tabla 2. Errores de clasificación con 5 % de contaminación**

Estimador	Tamaño de la contaminación ( $l$ )					
	SC	0.5	1	1.5	2	2.5
ML	18.17	18.82	19.48	20.31	21.49	23.20
WML	18.21	18.94	19.66	<b>20.59</b>	19.929	18.33
BY	18.20	18.78	18.81	18.77	18.72	18.66
WBY <sub>0.75</sub>	18.24	18.91	<b>18.95</b>	18.91	18.51	18.24
WBY <sub>0.5</sub>	18.26	19.00	<b>19.15</b>	18.86	18.46	18.28
PR <sub>1</sub> <sup>HU</sup>	18.23	<b>18.95</b>	18.66	18.59	18.59	18.61
PR <sub>0.8</sub> <sup>HU</sup>	18.25	<b>19.01</b>	18.61	18.56	18.56	18.56
PR <sup>HA</sup>	18.30	<b>19.19</b>	18.78	18.52	18.57	18.48

Estimador	Tamaño de la contaminación ( $l$ )				
	3	3.5	4	4.5	5
ML	25.58	28.64	32.12	35.65	<b>38.87</b>
WML	18.20	18.20	18.20	18.20	18.20
BY	18.66	18.64	18.83	19.05	<b>20.70</b>
WBY <sub>0.75</sub>	18.23	18.23	18.23	18.23	18.23
WBY <sub>0.5</sub>	18.24	18.23	18.23	18.23	18.23
PR <sub>1</sub> <sup>HU</sup>	18.63	18.60	18.61	18.60	18.58
PR <sub>0.8</sub> <sup>HU</sup>	18.58	18.57	18.58	18.59	18.60
PR <sup>HA</sup>	18.47	18.41	18.39	18.38	18.37

Para obtener muestras con aproximadamente un 10 % de contaminación se le agregaron a la muestra original 12 observaciones iguales de la forma  $(\mathbf{x}_i, 0)$  donde  $\mathbf{x}'_i = (1, l, l)$ ,  $101 \leq i \leq 112$ , con  $l \in \{0.5 * j, 1 \leq j \leq 11\}$ . Los valores de  $\widehat{\text{ECM}}$  y  $\widehat{\text{EC}}$  se muestran en Tablas 3 y 4 respectivamente

**Tabla 3. Errores cuadráticos medios con 10 % de contaminación**

Estimador	Tamaño de la contaminación ( $l$ )					
	SC	0.5	1	1.5	2	2.5
ML	0.849	1.436	3.284	4.816	5.956	6.768
WML	0.873	1.545	3.468	<b>5.009</b>	4.254	1.116
BY	1.077	1.700	3.225	4.427	5.555	6.523
WBY <sub>0.75</sub>	1.104	1.807	3.417	<b>4.679</b>	4.007	1.289
WBY <sub>0.5</sub>	1.183	1.956	3.739	<b>4.772</b>	3.481	1.960
PR <sup>HU</sup> ( $k = 1$ )	0.903	1.907	3.062	3.252	3.329	3.364
PR <sup>HU</sup> ( $k = .8$ )	0.931	2.100	2.799	2.861	2.915	2.953
PR <sup>HA</sup>	0.932	2.146	<b>3.265</b>	2.024	1.180	1.267

Estimador	Tamaño de la contaminación ( $l$ )				
	3	3.5	4	4.5	5
ML	7.329	7.712	7.975	8.158	<b>8.288</b>
WML	0.871	0.858	0.858	0.858	0.858
BY	7.242	7.718	7.997	8.170	<b>8.283</b>
WBY <sub>0.75</sub>	1.100	1.088	1.088	1.088	1.088
WBY <sub>0.5</sub>	1.274	1.109	1.112	1.112	1.111
PR <sup>HU</sup> ( $k = 1$ )	3.383	3.393	3.397	<b>3.418</b>	3.373
PR <sup>HU</sup> ( $k = .8$ )	2.963	2.958	2.961	<b>2.990</b>	2.957
PR <sup>HA</sup>	1.212	1.224	1.065	1.084	1.042

**Tabla 4. Errores de clasificación con 10 % de contaminación**

Estimador	Tamaño de la contaminación ( $l$ )					
	SC	0.5	1	1.5	2	2.5
ML	17.90	19.95	22.22	25.81	30.90	36.84
WML	17.94	20.19	22.76	<b>26.81</b>	26.13	18.59
BY	17.93	19.98	20.98	22.84	27.18	33.10
WBY <sub>0.75</sub>	17.97	20.25	21.55	<b>23.99</b>	23.27	18.36
WBY <sub>0.5</sub>	17.99	20.67	22.79	<b>25.14</b>	23.06	20.28
PR <sub>1</sub> <sup>HU</sup>	17.95	<b>20.65</b>	19.52	19.56	19.84	19.98
PR <sub>0.8</sub> <sup>HU</sup>	17.97	<b>21.00</b>	19.30	19.38	19.54	19.70
PR <sup>HA</sup>	18.02	<b>21.50</b>	20.12	18.99	19.29	19.75

Estimador	Tamaño de la contaminación ( $l$ )				
	3	3.5	4	4.5	5
ML	42.35	46.45	49.37	51.47	<b>53.20</b>
WML	17.96	17.91	17.91	17.91	17.91
BY	40.08	46.03	49.41	51.48	<b>52.94</b>
WBY <sub>0.75</sub>	17.98	17.94	17.94	17.94	17.94
WBY <sub>0.5</sub>	18.76	18.07	17.95	17.95	17.95
PR <sub>1</sub> <sup>HU</sup>	20.04	20.18	20.23	20.30	20.16
PR <sub>0.8</sub> <sup>HU</sup>	19.74	19.76	19.79	19.93	19.76
PR <sup>HA</sup>	19.43	19.26	18.83	18.69	18.61

Se observa que para los dos porcentajes de contaminación presentados los estimadores que tienen menor máximo ECM son los estimadores de proyección.

Para el 5 % de contaminación los estimadores de proyección tienen el mínimo máximo error de clasificación, pero muy parecido al del WBY<sub>0.75</sub>. Para el 10 % de contaminación los estimadores de proyección tienen máximo error de clasificación claramente menor que el resto de los estimadores. La ventaja del PR<sup>HA</sup> con respecto al PR<sub>1</sub><sup>HU</sup> y al PR<sub>0.8</sub><sup>HU</sup> es que debido a su carácter redescendiente tanto el ECM como el EC disminuyen cuando el tamaño de los outliers ( $l$ ) crece.



## 9. Ejemplos

Con el objetivo de estudiar el comportamiento de los estimadores de proyección en datos reales, presentamos dos ejemplos con conjuntos de datos bien conocidos. En cada caso se indicarán los estimadores que fueron considerados así como también se presenta una medida de bondad de ajuste y un qq-plot para las deviances.

La medida de bondad de ajuste surgida a partir de una aproximación de Kordzakhia et al. (2001), utilizada por Croux y Haesbroeck (2003), de estadísticos chi-cuadrado basados en la transformación arcoseno  $\chi_{arc}^2$  se define como

$$\chi_{arc}^2 = \sum_{i=1}^n 4 [\arcsin \sqrt{y_i} - \arcsin \sqrt{\pi_i}]^2$$

donde  $\pi_i = G(\mathbf{x}_i' \hat{\boldsymbol{\beta}})$  son las probabilidades ajustadas para  $i = 1, \dots, n$ . Un valor más pequeño de  $\chi_{arc}^2$  indica mejor bondad de ajuste.

García Ben y Yohai (2004) introducen un estimador de la distribución de la deviance. Ellos proponen calcular la distribución teórica de las deviances suponiendo que el verdadero valor del parámetro es  $\hat{\boldsymbol{\beta}}$  y así definen un nuevo QQ-plot donde se grafican los cuantiles de la distribución empírica de las deviances versus los cuantiles de la distribución teórica correspondiente al valor de  $\hat{\boldsymbol{\beta}}$  estimado.

### 9.1. Datos de Cook y Weisberg

Estos datos han sido considerados por Cook y Weisberg (1982, cap. 5, pág. 193) y Johnson (1985) para mostrar la identificación de datos influyentes. Los datos (Tabla 5) corresponden a 33 pacientes con leucemia. La variable de respuesta es 1 cuando el paciente sobrevive más de 52 semanas y 0 en caso contrario. Se consideran dos covariables:

WBC: cantidad de glóbulos blancos

AG: Presencia o ausencia de cierta característica morfológica de los glóbulos blancos.

El modelo también incluye intercept.

Cook y Weisberg (1982) notaron que la observación 15 es una observación muy influyente cuando se estiman los parámetros del modelo usando máxima verosimilitud y que, eliminando esta observación, el ajuste que se obtiene es mucho mejor. En la Tabla 6 se presentan los estimadores obtenidos usando máxima verosimilitud (ML), máxima verosimilitud sin considerar la observación número 15 (ML<sub>-15</sub>), el M-estimador Bianco-Yohai (BY) con  $d = 0.5$  y los estimadores de proyección  $PR_1^{HU}$  y  $PR_{0.7}^{HU}$ . En la Tabla 7 se muestran las probabilidades ajustadas correspondientes a los distintas observaciones.

La Tabla 7 muestra que el estimador  $PR_1^{HU}$  clasifica las observaciones de la misma manera que lo hace el M-estimador usado (BY), respecto del

estimador de máxima verosimilitud (MV), clasifica correctamente la observación 8. Los valores de  $\chi_{\text{arc}}^2$  para los estimadores de proyección son apenas superiores al alcanzado por el estimador BY, que de acuerdo a esta medida resulta mejor.

En las Figuras 1, 2 y 3 se grafican los QQ-plots de las deviances para los estimadores ML, BY y  $\text{PR}_1^{\text{HU}}$  respectivamente. Se observa que tanto el BY como el  $\text{PR}_1^{\text{HU}}$  detectan la observación 15 como outlier.

**Tabla 5. Datos de Cook y Weisberg**

Obs.	Y	WBC	AG
1	1	2300	1
2	1	750	1
3	1	4300	1
4	1	2600	1
5	0	6000	1
6	1	10500	1
7	1	10000	1
8	0	17000	1
9	0	5400	1
10	1	7000	1
11	1	9400	1
12	0	32000	1
13	0	35000	1
14	0	52000	1
15	1	100000	1
16	0	100000	1
17	0	100000	1
18	1	4400	0
19	1	3000	0
20	0	4000	0
21	0	1500	0
22	0	9000	0
23	0	5300	0
24	0	10000	0
25	0	19000	0
26	0	27000	0
27	0	28000	0
28	0	31000	0
29	0	26000	0
30	0	21000	0
31	0	79000	0
32	0	100000	0
33	0	100000	0

**Tabla 6. Estimadores para los datos de Cook y Weisberg**

	ML	ML <sub>-15</sub>	BY	PR <sub>1</sub> <sup>HU</sup>	PR <sub>0.7</sub> <sup>HU</sup>
constante	-1.3073	0.2119	0.1595	-0.4396	-0.3733
WBC	-0.0000	-0.0002	-0.0002	-0.0001	-0.0001
AC	2.2611	2.5581	1.9274	1.9830	2.1069
$\chi_{arc}^2$	52.1576	42.3890	44.0546	46.0527	45.4068

**Tabla 7. Probabilidades ajustadas para los datos de Cook y Weisberg**

Obs.	$y$	ML	ML <sub>-15</sub>	BY	PR <sub>1</sub> <sup>HU</sup>	PR <sub>0.7</sub> <sup>HU</sup>
1	1	0.7070	0.9028	0.8428	0.7841	0.8119
2	1	0.7171	0.9304	0.8759	0.8117	0.8382
3	1	0.6936	0.8529	0.7899	0.7445	0.7732
4	1	0.7050	0.8964	0.8356	0.7785	0.8064
5	0	0.6820	0.7953	0.7355	0.7073	0.7361
6	1	0.6502	0.5739	0.5559	0.5954	0.6213
7	1	0.6538	0.6024	0.5777	0.6086	0.6351
8	0	0.6020	0.2257	0.2833	0.4183	0.4326
9	0	0.6861	0.8174	0.7557	0.7208	0.7496
10	1	0.6751	0.7543	0.6996	0.6840	0.7126
11	1	0.6582	0.6357	0.6034	0.6243	0.6513
12	0	0.4843	0.0085	0.0269	0.1210	0.1150
13	0	0.4605	0.0042	0.0160	0.0900	0.0836
14	0	0.3322	0.0001	0.0008	0.0150	0.0121
15	1	0.0977	0.0000	0.0000	0.0001	0.0000
16	0	0.0977	0.0000	0.0000	0.0001	0.0000
17	0	0.0977	0.0000	0.0000	0.0001	0.0000
18	1	0.1904	0.3049	0.3496	0.2841	0.2906
19	1	0.1974	0.3789	0.4079	0.3164	0.3258
20	0	0.1924	0.3252	0.3659	0.2931	0.3005
21	0	0.2051	0.4647	0.4734	0.3532	0.3658
22	0	0.1689	0.1293	0.1920	0.1929	0.1924
23	0	0.1861	0.2619	0.3142	0.2643	0.2693
24	0	0.1645	0.1050	0.1660	0.1763	0.1747
25	0	0.1289	0.0139	0.0388	0.0736	0.0682
26	0	0.1029	0.0021	0.0097	0.0318	0.0277
27	0	0.1000	0.0017	0.0081	0.0286	0.0247
28	0	0.0918	0.0008	0.0048	0.0207	0.0175
29	0	0.1059	0.0027	0.0115	0.0354	0.0311
30	0	0.1219	0.0087	0.0275	0.0599	0.0547
31	0	0.0215	0.0000	0.0000	0.0001	0.0001
32	0	0.0112	0.0000	0.0000	0.0000	0.0000
33	0	0.0112	0.0000	0.0000	0.0000	0.0000

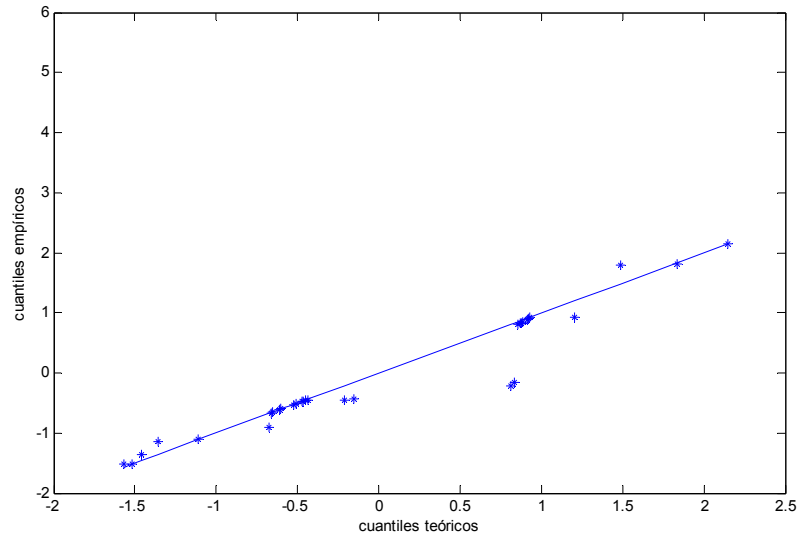


Figura 1. QQ-plot de las deviancias del estimador ML para los datos de Cook-Weisberg

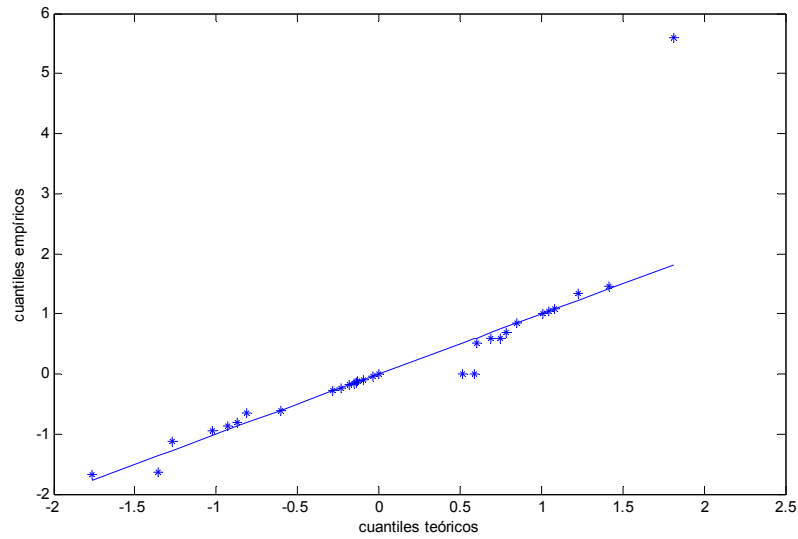


Figura 2. QQ-plot de las deviancias del estimador BY para los datos de Cook-Weisberg

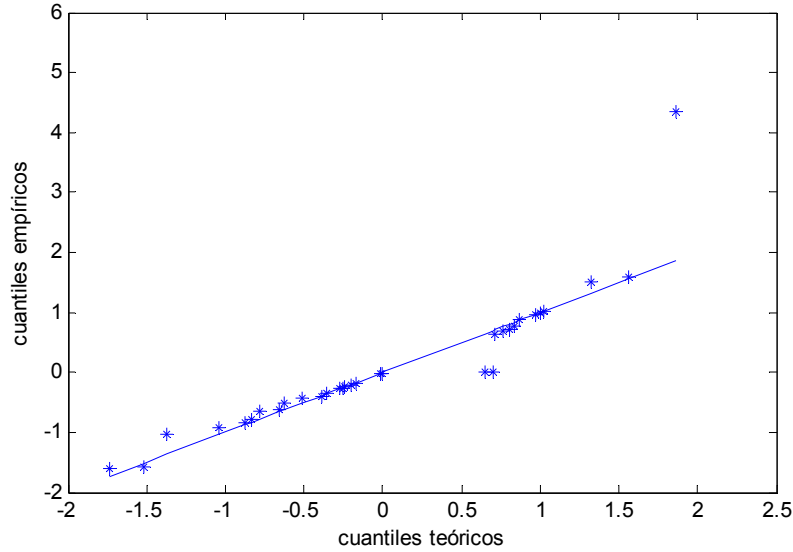


Figura 3. QQ-plot de las deviancias del estimador  $PR_1^{HU}$  para los datos de Cook-Weisberg

## 9.2. Datos de Finney

Consideramos un conjunto de datos introducido por Finney (1947), estudiado por Pergibon (1981) para mostrar el impacto de observaciones potencialmente influenciadas en la regresión logística, y por Künsch et al. (1989). En este ejemplo se estudia el efecto de la velocidad y el volumen del aire inspirado (ambos en logaritmos) sobre la respuesta binaria  $Y$  que representa la presencia ( $Y = 1$ ) o ausencia ( $Y = 0$ ) de vasoconstricción en la piel de los dedos. Las observaciones se presentan en la Tabla 8. El modelo que se considera es el siguiente

$$\ln \frac{\pi}{1 - \pi} = \beta_0 + \beta_1 \ln(\text{volumen}) + \beta_2 \ln(\text{velocidad}).$$

Este conjunto de datos presenta dos observaciones centrales (4 y 18) que han sido detectadas como influyentes para los estimadores clásicos. Si estas dos observaciones son extraídas y se utiliza máxima verosimilitud para la estimación, sólo queda un punto en el conjunto de overlap. Los procedimientos robustos aplicados (BY y  $PR^{HU}$ ) arrojan idéntica clasificación, excepto la observación 24 que es clasificada correctamente sólo con  $PR^{HU}(k = 0.7)$  (ver Tabla 10).

**Tabla 8**

Obs.	Y	Volumen	Velocidad
1	1	3.7	0.82
2	1	3.5	1.09
3	1	1.2	2.5
4	1	0.75	1.5
5	1	0.8	3.2
6	1	0.7	3.5
7	0	0.6	0.75
8	0	1.1	1.7
9	0	0.9	0.75
10	0	0.9	0.45
11	0	0.8	0.57
12	0	0.55	2.75
13	0	0.6	3
14	1	1.4	2.33
15	1	0.75	3.75
16	1	2.3	1.64
17	1	3.2	1.6
18	1	0.85	1.42
19	0	1.7	1.06
20	1	1.8	1.8
21	0	0.4	2.0
22	0	0.95	1.36
23	0	1.35	1.35
24	0	1.5	1.36
25	1	1.6	1.78
26	0	0.6	1.5
27	1	1.8	1.5
28	0	0.95	1.9
29	1	1.9	0.95
30	0	1.6	0.4
31	1	2.7	0.75
32	0	2.35	0.03
33	0	1.1	1.83
34	1	1.1	2.2
35	1	1.2	2
36	1	0.8	3.33
37	0	0.95	1.9
38	0	0.75	1.9
39	1	1.3	1.63



**Tabla 9**

	Estimador				
	ML	ML <sub>{-4,-18}</sub>	BY	PR <sup>HU</sup> ( $k = 1$ )	PR <sup>HU</sup> ( $k = .7$ )
constante	-2.8871	-4.7793	-6.8935	-4.1397	-5.1982
Volumen	5.1942	8.3599	10.7936	6.2240	7.4913
Velocidad	4.5777	6.9749	9.4142	5.7339	6.8168
$\chi_{arc}^2$	48.3175	43.9778	40.7735	46.4903	45.1932

Tabla 10

Y obs.	Estimador				
	ML Prob.Est.	ML <sub>{-4,-18}</sub> Prob.Est.	BY Prob.Est.	PR <sup>HU</sup> (k = 1) Prob.Est.	PR <sup>HU</sup> (k = .7) Prob.Est.
1	0.9526	0.9916	0.9953	0.9461	0.9627
1	0.9823	0.9982	0.9994	0.9845	0.9916
1	0.9218	0.9700	0.9843	0.9244	0.9382
1	0.0741	0.0127	0.0021	0.0265	0.0101
1	0.7822	0.8128	0.8386	0.7578	0.7425
1	0.7301	0.7265	0.7409	0.6950	0.6614
0	0.0011	0.0000	0.0000	0.0001	0.0000
0	0.5093	0.4301	0.2954	0.3766	0.2959
0	0.0086	0.0005	0.0000	0.0016	0.0004
0	0.0008	0.0000	0.0000	0.0001	0.0000
0	0.0013	0.0000	0.0000	0.0002	0.0000
0	0.2040	0.0617	0.0214	0.1130	0.0584
0	0.3749	0.1999	0.1126	0.2651	0.1771
1	0.9389	0.9808	0.9910	0.9429	0.9564
1	0.8415	0.8844	0.9202	0.8387	0.8398
1	0.9760	0.9964	0.9988	0.9798	0.9880
1	0.9951	0.9997	1.0000	0.9970	0.9988
1	0.1066	0.0243	0.0047	0.0415	0.0175
0	0.5339	0.5158	0.3504	0.3768	0.3045
1	0.9457	0.9857	0.9932	0.9473	0.9613
0	0.0113	0.0005	0.0000	0.0028	0.0007
0	0.1486	0.0446	0.0104	0.0632	0.0297
0	0.5114	0.4558	0.3039	0.3656	0.2882
0	0.6517	0.6803	0.5933	0.5367	0.4838
1	0.8997	0.9598	0.9736	0.8901	0.9049
0	0.0245	0.0020	0.0002	0.0067	0.0019
1	0.8831	0.9509	0.9633	0.8634	0.8775
0	0.4464	0.3249	0.1971	0.3146	0.2302
1	0.5528	0.5570	0.3897	0.3920	0.3231
0	0.0096	0.0007	0.0000	0.0015	0.0004
1	0.7221	0.8202	0.7538	0.5969	0.5699
0	0.0000	0.0000	0.0000	0.0000	0.0000
0	0.5925	0.5579	0.4562	0.4796	0.4098
1	0.7716	0.8201	0.8261	0.7260	0.7090
1	0.7743	0.8291	0.8320	0.7250	0.7094
1	0.8116	0.8514	0.8832	0.7972	0.7909
0	0.4464	0.3249	0.1971	0.3146	0.2302
0	0.1911	0.0625	0.0188	0.0954	0.0484
1	0.6709	0.6946	0.6313	0.5731	0.5244

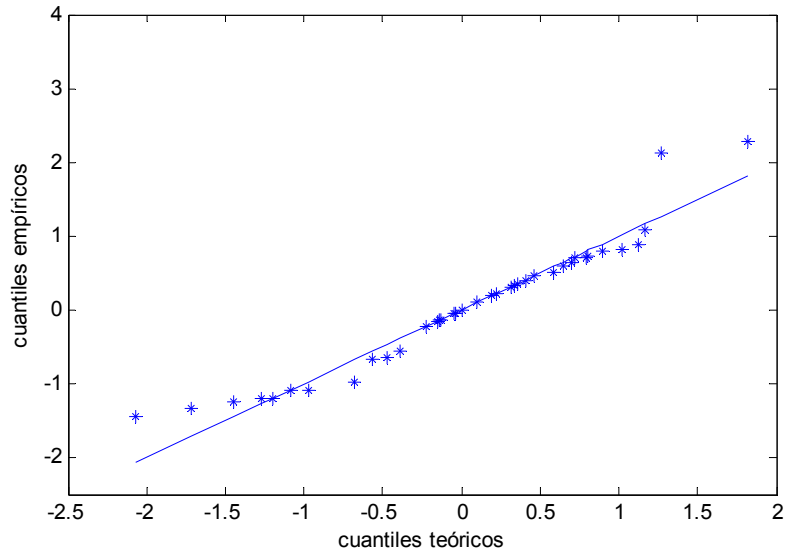


Figura 4. QQ-plot de las deviancias del estimador ML para los datos de Finney

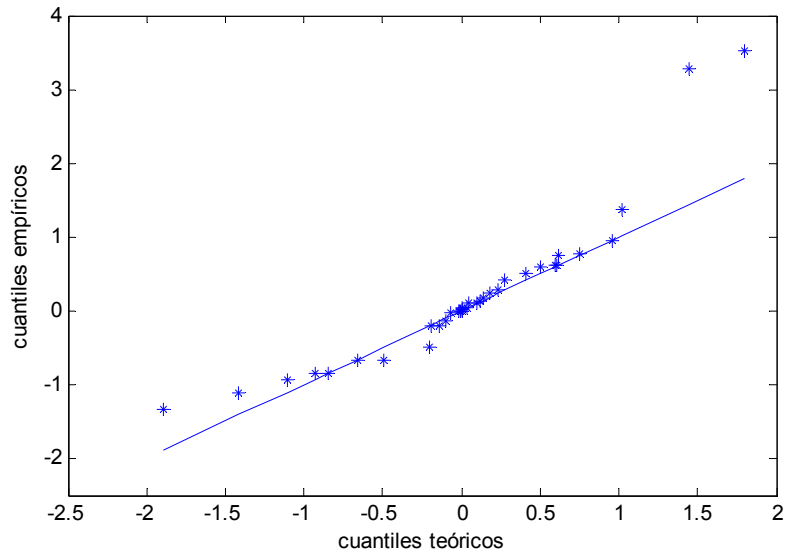


Figura 5. QQ-plot de las deviancias del estimador BY para los datos de Finney

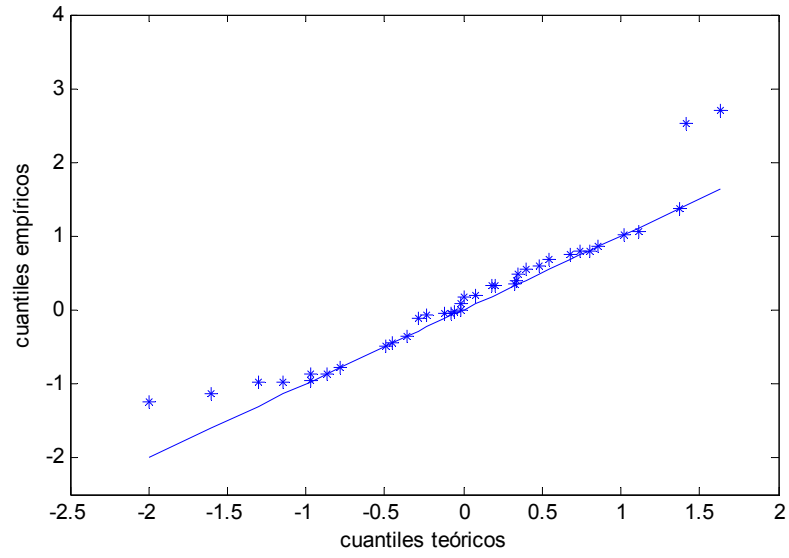


Figura 6. QQ-plot de las deviancias del estimador  $PR_1^{HU}$  para los datos de Finney

## Referencias

- Bianco, A. y Yohai, V. (1996). Robust Estimation in the Logistic Regression Model. En *Robust Statistics, Data Analysis and Computer Intensive Methods*, Proceedings of the workshop in honor of Peter Huber, editor H. Rieder, Lecture notes in Statistics, **109**, 17-34, Springer-Verlag, New York.
- Cook, R. y Weisberg, S. (1982). *Residuals and Influence in Regression*. Chapman and Hall, London.
- Croux, C. y Haesbroeck, G. (2003). Implementing the Bianco and Yohai Estimator for Logistic Regression. *Computational Statistics & Data Analysis*, **44**, no 1-2, 273-295.
- García Ben, M. y Yohai, V. (2004). Quantile-Quantile Plot for Deviance Residuals in Generalized Linear Model. *Journal of Computational & Graphical Statistics*, **13**, no 1, 36-47.
- Hampel, F., Ronchetti, E., Rousseeuw, E., Stahel, W. (1986). *Robust Statistics The Approach Based on Influence Functions*. John Wiley & Sons.
- Jain, N. y Marcus, M. (1975). Central Limit Theorems for C(S)-valued Random Variables. *Journal of Functional Analysis*, **19**, 216-231.
- Johnson, W. (1985). Influence Measures for Logistic Regression: Another Point of View. *Biometrika*, **72**, 1, 59-65.
- Kolmogorov, A.N y Fomín (1975). *Elementos de la teoría de funciones y del análisis funcional*. Editorial Mir-Moscú.
- Kordzakhia, H., Mishsa, G. y Reiersolmoen, L. (2001). Robust Estimation in Logistic Regression Models. *Journal of Statistical Planning and Inference*, **98**, 211-223.
- Krasker, W., y Welsch, R. (1982). Efficient Bounded-Influence Regression Estimation. *Journal of the American Statistical Association*, **77**, 595-604.
- Künsch, H., Stefanski, L. y Carrol, R. (1989). Conditionally Unbiased Bounded-Influence Estimation in General Regression Models, with Applications to Generalized Linear Models. *Journal of the American Statistical Association*, **84**, 460-466.
- Maronna, R. y Yohai, V. (1993). Bias-Robust Estimates of Regression Based on Projections. *Annals of Statistics*, **21**, 965-990.
- Maronna, R., Martin, D. y Yohai, V. (2006). *Robust Statistics. Theory and Methods*. John Wiley & Sons, Ltd.
- Pregibon, D. (1981). Logistic Regression Diagnostics. *Annals of Statistics*, **9**, 705-724.
- Rousseeuw, P y Leroy, A. (1987). *Robust Regression and Outlier Detection*. John Wiley & Sons.