

## Tesis de Posgrado

# M- estimadores robustos para regresión binominal y multinomial

Bianco, Ana María

1990

Tesis presentada para obtener el grado de Doctor en Ciencias  
Matemáticas de la Universidad de Buenos Aires

Este documento forma parte de la colección de tesis doctorales y de maestría de la Biblioteca Central Dr. Luis Federico Leloir, disponible en [digital.bl.fcen.uba.ar](http://digital.bl.fcen.uba.ar). Su utilización debe ser acompañada por la cita bibliográfica con reconocimiento de la fuente.

This document is part of the doctoral theses collection of the Central Library Dr. Luis Federico Leloir, available in [digital.bl.fcen.uba.ar](http://digital.bl.fcen.uba.ar). It should be used accompanied by the corresponding citation acknowledging the source.

**Cita tipo APA:**

Bianco, Ana María. (1990). M- estimadores robustos para regresión binominal y multinomial. Facultad de Ciencias Exactas y Naturales. Universidad de Buenos Aires.  
[http://digital.bl.fcen.uba.ar/Download/Tesis/Tesis\\_2316\\_Bianco.pdf](http://digital.bl.fcen.uba.ar/Download/Tesis/Tesis_2316_Bianco.pdf)

**Cita tipo Chicago:**

Bianco, Ana María. "M- estimadores robustos para regresión binominal y multinomial". Tesis de Doctor. Facultad de Ciencias Exactas y Naturales. Universidad de Buenos Aires. 1990.  
[http://digital.bl.fcen.uba.ar/Download/Tesis/Tesis\\_2316\\_Bianco.pdf](http://digital.bl.fcen.uba.ar/Download/Tesis/Tesis_2316_Bianco.pdf)

**EXACTAS** UBA

Facultad de Ciencias Exactas y Naturales



**UBA**

Universidad de Buenos Aires

**Universidad de Buenos Aires**  
**Facultad de Ciencias Exactas y Naturales**

**TEMA DE TESIS**  
**M-ESTIMADORES ROBUSTOS PARA REGRESION BINOMIAL Y MULTINOMIAL**

**AUTORA:**  
**Ana Maria Bianco**


**DIRECTOR:**  
**Dr. Victor J. Yohai**

2316  
ej 2

**LUGAR DE TRABAJO**  
**Facultad de Ciencias Exactas y Naturales**

**Tesis presentada para optar al título de Doctor en Ciencias Matemáticas**

**1990**



1  
1  
1  
1  
1  
1  
1  
1  
1  
1  
1

**Agradezco al Dr. Victor J. Yohai por -  
haberme guiado con tanta dedicación  
en la realización de este trabajo.**

1  
1  
1  
1  
1  
1  
1  
1  
1  
1  
1

## 1. INTRODUCCIÓN

En el siguiente trabajo se proponen métodos robustos para estimar el parámetro de regresión en los modelos de regresión logística binomial y multinomial. La motivación para este tipo de estimadores surge del hecho de que el método de máxima verosimilitud es muy sensible a la presencia de datos anómalos en la muestra.

Es bien conocido que en el modelo lineal el método de máxima verosimilitud puede ser muy ineficiente cuando la hipótesis de normalidad se cumple sólo aproximadamente. Esto condujo a la búsqueda de otros procedimientos, llamados robustos, que fueran estables aún cuando la suposición de normalidad se verificase sólo aproximadamente.

Consideremos el modelo de regresión binomial. Supongamos que observamos una variable dependiente Bernoulli  $Y$ , que sólo toma valores 0 ó 1, que está relacionada con  $q$  variables  $X_1, \dots, X_q$  por medio de una función de enlace  $H$ . De manera tal que, si con  $\mathbf{X}$  notamos a  $(X_1, \dots, X_q)'$ , dado  $\mathbf{x} \in \mathfrak{R}^q$  tenemos que

$$P(Y = 1 | \mathbf{X} = \mathbf{x}) = H(\mathbf{x}' \underline{\beta}_0), \quad (1.1)$$

para algún  $\underline{\beta}_0 \in \mathfrak{R}^q$ , que es el vector de parámetros desconocidos que nos proponemos estimar.

Supondremos que  $H$  es una función de distribución continua y estrictamente creciente. Por ejemplo,  $H$  puede ser la función logística

$$H(t) = \frac{\exp(t)}{1 + \exp(t)}. \quad (1.2)$$

o la función probit  $\Phi(t)$  que corresponde a la función de distribución normal standard.

Bajo el modelo dado por (1.1), la función de frecuencia de  $Y$  dado  $\mathbf{X} = \mathbf{x}$  es

$$f(y | \mathbf{x}' \underline{\beta}_0) = \exp(y \ln(H(\mathbf{x}' \underline{\beta}_0)) + (1 - y) \ln(1 - H(\mathbf{x}' \underline{\beta}_0))) \quad (1.3)$$

El procedimiento clásico para estimar a  $\underline{\beta}_0$  es el método de máxima verosimilitud. Si  $\mathbf{Z}_1 = (\mathbf{X}'_1, Y_1)', \dots, \mathbf{Z}_n = (\mathbf{X}'_n, Y_n)'$  es una muestra aleatoria que satisface (1.1), donde

$\mathbf{X}_i = (X_{i1}, \dots, X_{iq})'$ ,  $i = 1, \dots, n$ , el estimador de máxima verosimilitud (E.M.V.) de  $\underline{\beta}_0$  es el valor  $\underline{\beta}$  que maximiza

$$L(\mathbf{Z}_1, \dots, \mathbf{Z}_n, \underline{\beta}) = L(\bar{\mathbf{Z}}, \underline{\beta}) = \frac{1}{n} \sum_{i=1}^n l(Y_i, \mathbf{X}_i' \underline{\beta}), \quad (1.4)$$

donde  $l(y, \theta) = \ln f(y, \theta)$ . Este método provee un estimador asintóticamente eficiente para estimar  $\underline{\beta}_0$ , pero como ya dijimos es extremadamente sensible a la presencia de respuestas atípicas o puntos de diseño extremos que llamaremos de alto "leverage".

Pregibon (1981) y Cook y Weisberg (1982) ejemplificaron esta sensibilidad mediante el análisis de datos reales. Estos autores estudiaron conjuntos de datos reales (que analizamos en la Sección 6) con el fin de mostrar que la estimación del parámetro por máxima verosimilitud puede estar altamente influenciada por una única observación. Esto revela la falta de robustez del método y por lo tanto la necesidad de encontrar técnicas que nos protejan de las observaciones atípicas que pueden ser altamente influyentes.

Se puede hallar en la literatura distintos métodos propuestos para la identificación de aquellas observaciones que son influyentes y también diversos estimadores robustos del parámetro.

Pregibon (1981) mostró los beneficios de los métodos de diagnóstico para medir el efecto de las observaciones anómalas sobre el estimador de máxima verosimilitud. Algunos de estos métodos son generalizaciones de las técnicas de diagnóstico empleadas para mínimos cuadrados en el modelo lineal, mientras que otras están basadas en la medida de discrepancia introducida por Cook (1977) y en la "deviance". Esta última fue definida por Nelder y Wedderburn (1972) como

$$D = \sum_{i=1}^n d(Y_i, \mathbf{X}_i' \underline{\beta}) = \sum_{i=1}^n d_i \quad (1.5)$$

donde las componentes  $d_i$  de la "deviance" a su vez están definidas por

$$d_i = -2\{l(y_i, \mathbf{x}_i' \underline{\beta}) - l(y_i, \hat{\theta}_i)\}$$

donde  $\hat{\theta}_i$  es el estimador de  $\theta = \mathbf{x}_i' \underline{\beta}$  basado en la observación  $y_i$  solamente.

Pregibon (1982) propuso un estimador alternativo al de máxima verosimilitud basado en la deviance. Como el E.M.V. equivale al estimador que se obtiene de minimizar  $D$ , Pregibon sugiere minimizar una función "deviance" modificada, menos sensible que la original  $D$ . Considera una función robusta de las componentes  $d(\theta, y)$ , por ejemplo cualquier función de pérdida robusta utilizada en regresión lineal. La desventaja de estos estimadores es que en general no son consistentes.

Con el fin de realizar diagnóstico, Cook y Weisberg definieron una medida de influencia, que llamaron "distancia entre las verosimilitudes", como

$$LD_i = 2[L(\bar{Z}, \hat{\underline{\beta}}) - L(\bar{Z}, \hat{\underline{\beta}}_{(i)})]$$

donde  $\hat{\underline{\beta}}_{(i)}$  es el E.M.V. que se obtiene cuando se excluye de la muestra la observación  $i$ . Esta medida da una idea de cuanto se deteriora la función de verosimilitud cuando no se considera la observación  $i$  y por lo tanto de cuanto influye esta observación en la estimación de  $\underline{\beta}_0$ .

Johnson (1985) propuso identificar las observaciones influyentes midiendo el efecto de una observación sobre la estimación de las probabilidades y la clasificación de un individuo en una de las dos clases.

Por otra parte, Stefanski, Carroll y Ruppert (1986) presentaron una clase de estimadores robustos para el modelo lineal generalizado, extendiendo los resultados obtenidos por Krasker y Welsh (1982) para el modelo lineal. Ellos aplican estos estimadores al caso particular de regresión logística. Obtienen estimadores óptimos de influencia acotada del tipo de los de Hampel, minimizando cierto funcional de la matriz de covarianza asintótica. Con la misma idea de hallar estimadores que minimizan un funcional de la matriz de covarianza asintótica sujetos a una cota en la medida de sensibilidad, Künsch, Stefanski y Carroll (1987) hallan estimadores de influencia acotada, que a diferencia de los anteriores



son consistentes en el sentido de Fisher cuando se los condiciona a  $\mathbf{X}$ . Estas dos clases de estimadores son computacionalmente complejos de calcular y además requieren el cómputo de la matriz de covarianza de las  $\mathbf{X}$  para su cálculo. aunque los propuestos por Künsch et al. son algo más sencillos y están menos afectados por la estimación de la matriz de covarianza.

Sin embargo, es sabido que en el caso del modelo lineal los estimadores óptimos de influencia acotada no son muy robustos cuando la dimensión  $q$  es grande. Por ejemplo, su punto de ruptura tiende a 0 cuando  $q$  es grande, como lo probaron Maronna, Bustos y Yohai (1979). Varios estimadores de alto punto de ruptura han sido propuestos para el modelo lineal, como por ejemplo Siegel (1982), Rousseeuw (1982), Rousseeuw y Yohai (1984), Yohai (1987) y Yohai y Zamar (1988). Excepto el primero, todos los estimadores no tienen función de influencia acotada y a pesar de esto, cuando  $q$  es grande son más robustos que los estimadores de influencia acotada. Esto muestra que la curva de influencia, que mide la sensibilidad del estimador bajo contaminaciones infinitesimales puede no ser adecuada para medir los efectos de una contaminación positiva, aunque pequeña. Los estimadores propuestos por Yohai (1987) son M-estimadores con función de pérdida acotada y escala conveniente. Esto nos sugiere estudiar M-estimadores con función de pérdida acotada para estimar  $\underline{\beta}_0$  bajo el modelo (1.1).

Veamos como podemos definir los estimadores que proponemos a partir de los máxima verosimilitud. Si llamamos

$$Y_{i1} = Y_i, \quad Y_{i2} = 1 - Y_i, \quad H_1 = H, \quad H_2 = 1 - H$$

y cambiamos de signo en (1.4), el E.M.V. puede definirse como aquel  $\underline{\beta}$  que minimiza

$$\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^2 Y_{ij} (-\ln(H_j(\mathbf{X}_i' \underline{\beta}))). \quad (1.6)$$

Se propone la siguiente clase de M-estimadores como una robustificación natural del

E.M.V.: se define  $\hat{\underline{\beta}}$  como aquel valor de  $\underline{\beta}$  que minimiza

$$\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^2 [Y_{ij} \rho(-\ln(H_j(\mathbf{X}_i' \underline{\beta}))) + G(H_j(\mathbf{X}_i' \underline{\beta}))]. \quad (1.7)$$

donde  $\rho$  es una función acotada, no decreciente con derivada  $\psi$  y

$$G(t) = \int_0^t \psi(-\ln u) du. \quad (1.8)$$

Como veremos en la Sección 2 la elección de  $G$  en (1.8) hace que el estimador resulte consistente en el sentido de Fisher.

En la Sección 2 estudiamos la consistencia del M-estimador que minimiza (1.7). La Proposición 2.1 prueba la consistencia en el sentido de Fisher del estimador y a partir de ella se deduce la consistencia fuerte en el Teorema 2.1.

En la Sección 3 analizamos la distribución asintótica del método propuesto. En el Teorema 3.1 probamos que el estimador converge débilmente a una distribución normal en el caso en que la función de enlace  $H$  es de la forma (1.2) y en el Teorema 3.2 probamos un resultado análogo para funciones de enlace  $H$  generales.

En la Sección 4 presentamos dos familias de funciones  $\rho$  y calculamos numéricamente la eficiencia asintótica del M-estimador para estas dos familias y para diversos valores del parámetro.

En la primera parte de la Sección 5 calculamos numéricamente el sesgo asintótico del procedimiento que proponemos y en la segunda parte comparamos los valores asintóticos obtenidos con los correspondientes al estimador propuesto por Künsch, Stefanski y Carroll (1987).

En la Sección 6 se presenta un estimador inicial para el cómputo del M-estimador que minimiza (1.7) y se ilustra el comportamiento del estimador con dos ejemplos de datos reales.

En la Sección 7 se extienden los M-estimadores al modelo de regresión multinomial. Se prueba la consistencia en el sentido de Fisher del estimador, se calcula su matriz de covarianza asintótica y se ejemplifica el comportamiento del estimador extendido por medio de ejemplos simulados y reales.

## 2. CONSISTENCIA

En la presente sección demostraremos la consistencia en el sentido de Fisher de los M-estimadores definidos por (1.7) y la consistencia fuerte de los mismos.

Se harán las siguientes hipótesis sobre  $\rho$

A1.  $\rho(t)$  es acotada, con derivada acotada  $\psi$  en  $[0, \infty)$ . Notaremos  $K = \sup_t \rho(t)$  y  $L = \sup_t \psi(t)$

A2.  $\psi(t) \geq 0$  y existe  $c > -\ln \frac{1}{2}$  tal que  $\psi(t) > 0$  para todo  $0 < t < c$

A3.  $\rho(0) = 0$

A1 y A2 aseguran que las funciones  $\psi$  utilizadas son redescendientes y que si se anulan a partir de algún punto, éste debe ser lo suficientemente grande como para que al acotar  $(-\ln p)$  mediante  $\rho$  en (1.7), el M-estimador sea consistente en el sentido de Fisher (ver Lema 2.1).

### Ejemplos.

En las secciones siguientes se considerarán dos familias de funciones  $\rho$  que satisfacen A1 a A3. Estas son

$$\rho_c(t) = \begin{cases} t - (t^2/2c) & \text{si } 0 \leq t < c \\ c/2 & \text{si } t \geq c \end{cases}$$
$$\rho_\lambda^* = \frac{1 - e^{-\lambda t}}{\lambda} \quad \text{si } t \geq 0$$

donde  $c$  y  $\lambda$  son constantes positivas. Es claro que cuando  $c$  tiende a  $\infty$  la función  $\rho_c$  converge a la identidad y que lo mismo ocurre con  $\rho_\lambda^*$  cuando  $\lambda$  tiende a 0. Por lo tanto, bajo estas condiciones los M-estimadores correspondientes convergen al estimador de máxima verosimilitud.

Serán necesarias las siguientes hipótesis sobre  $H$ :

B1.  $H(t)$  es una función estrictamente creciente.

B2.  $\lim_{t \rightarrow \infty} H(t) = 1$ .

B3.  $\lim_{t \rightarrow -\infty} H(t) = 0$ .

B4.  $H(t)$  es continua.

B5.  $H(t)$  es dos veces diferenciable.

En esta sección se supondrá que  $H$  satisface las hipótesis B1 a B4.

### Ejemplos.

Las funciones logística y probit de la sección anterior son ejemplos de funciones de enlace que satisfacen el conjunto B de hipótesis.

Para simplificar la notación, se definirá para  $y \in [0, 1]$ ,  $\mathbf{x} \in \mathbb{R}^q$ ,  $\underline{\beta} \in \mathbb{R}^q$

$$\phi(y, \mathbf{x}, \underline{\beta}) = \sum_{j=1}^2 y_j \rho(-\ln(H_j(\mathbf{x}'\underline{\beta}))) + G(H_j(\mathbf{x}'\underline{\beta})), \quad (2.1)$$

donde  $Y_j$ ,  $H_j$  y  $G$  fueron definidas en la Sección 1.

La proposición 2.1 probará la consistencia en sentido de Fisher del estimador definido por (1.7). Se probará primero el siguiente lema.

LEMA 2.1. Sea  $Y$  una variable Bernoulli tal que  $P(Y = 1) = p_0$ . luego

$$L(p_0, p) = E_{p_0}(Y \rho(-\ln p) + G(p) + (1 - Y) \rho(-\ln(1 - p)) + G(1 - p)) \quad (2.2)$$

tiene un único mínimo en  $p = p_0$ .

DEMOSTRACIÓN:

Es suficiente mostrar que

$$\frac{\partial L(p_0, p)}{\partial p} < 0 \quad \text{para todo } p < p_0 \quad (2.3)$$

y

$$\frac{\partial L(p_0, p)}{\partial p} > 0 \quad \text{para todo } p > p_0. \quad (2.4)$$

Dado que

$$L(p_0, p) = p_0 \rho(-\ln p) + G(p) + (1 - p_0) \rho(-\ln(1 - p)) + G(1 - p),$$

se tiene que

$$\begin{aligned}\frac{\partial L(p_0, p)}{\partial p} &= -\frac{p_0}{p} \psi(-\ln p) + \psi(-\ln p) + \frac{1-p_0}{1-p} \psi(-\ln(1-p)) - \psi(-\ln(1-p)) \\ &= (p-p_0) \left( \frac{\psi(-\ln p)}{p} + \frac{\psi(-\ln(1-p))}{1-p} \right).\end{aligned}\quad (2.5)$$

Como por A2 existe  $c > -\ln \frac{1}{2}$  tal que  $\psi$  es estrictamente positiva si  $0 < t < c$ . entonces  $\psi(-\ln p)$  es positivo o  $\psi(-\ln(1-p))$  lo es. En consecuencia

$$\frac{\psi(-\ln p)}{p} + \frac{\psi(-\ln(1-p))}{1-p} > 0$$

y de (2.5) se obtiene que el signo de  $\partial L(p_0, p)/\partial p$  depende solamente de la diferencia  $(p-p_0)$ , verificándose (2.3) y (2.4).

**PROPOSICIÓN 2.1.** *Sea  $\rho$  una función satisfaciendo A1 a A3. Si para todo  $\underline{\alpha} \in \mathbb{R}^q, \underline{\alpha} \neq 0$ , se cumple que  $P(\mathbf{X}'\underline{\alpha} = 0) < 1$ . entonces para todo  $\underline{\beta} \in \mathbb{R}^q, \underline{\beta} \neq \underline{\beta}_0$*

$$E_{\underline{\beta}_0}(\phi(Y, \mathbf{X}, \underline{\beta}_0)) < E_{\underline{\beta}_0}(\phi(Y, \mathbf{X}, \underline{\beta})).$$

**DEMOSTRACIÓN:**

Tomando esperanza condicional con respecto a  $\mathbf{X}$ , es inmediato que

$$\begin{aligned}E_{\underline{\beta}_0}(\phi(Y, \mathbf{X}, \underline{\beta})) &= E(H(\mathbf{X}'\underline{\beta}_0)\rho(-\ln(H(\mathbf{X}'\underline{\beta}))) + (1-H(\mathbf{X}'\underline{\beta}_0))\rho(-\ln(1-H(\mathbf{X}'\underline{\beta}))) \\ &\quad + G(H(\mathbf{X}'\underline{\beta})) + G(1-H(\mathbf{X}'\underline{\beta}))) \\ &= E(L(H(\mathbf{X}'\underline{\beta}_0), H(\mathbf{X}'\underline{\beta}))).\end{aligned}$$

donde  $L$  fue definida en (2.2). Por el Lema 2.1, se tiene que  $L(H(\mathbf{X}'\underline{\beta}_0), H(\mathbf{X}'\underline{\beta}))$  alcanza su mínimo en  $H(\mathbf{X}'\underline{\beta}) = H(\mathbf{X}'\underline{\beta}_0)$  y dado que  $H$  es estrictamente creciente, esto es equivalente a afirmar que  $L$  alcanza el mínimo para  $\underline{\beta}$  tal que  $\mathbf{X}'\underline{\beta} = \mathbf{X}'\underline{\beta}_0$ .

Dado  $\underline{\beta} \in \mathfrak{R}^q$  se define el conjunto  $A_{\underline{\beta}} = \{\mathbf{x} \in \mathfrak{R}^q : \mathbf{x}'\underline{\beta} = \underline{\mathbf{x}}'\underline{\beta}_0\}$ . Por hipótesis  $P(A_{\underline{\beta}}) < 1$ . Sea  $P_X$  es la función de distribución de  $\mathbf{X}$ , entonces por el Lema 2.1 se tiene que

$$\begin{aligned} E_{\underline{\beta}_0}(\phi(Y, \mathbf{X}, \underline{\beta}_0)) &= \int_{A_{\underline{\beta}}} L(H(\mathbf{X}'\underline{\beta}_0), H(\mathbf{X}'\underline{\beta}_0)) dP_X + \int_{A_{\underline{\beta}}^c} L(H(\mathbf{X}'\underline{\beta}_0), H(\mathbf{X}'\underline{\beta}_0)) dP_X \\ &< \int_{A_{\underline{\beta}}} L(H(\mathbf{X}'\underline{\beta}_0), H(\mathbf{X}'\underline{\beta}_0)) dP_X + \int_{A_{\underline{\beta}}^c} L(H(\mathbf{X}'\underline{\beta}_0), H(\mathbf{X}'\underline{\beta})) dP_X \\ &= E_{\underline{\beta}_0}(\phi(Y, \mathbf{X}, \underline{\beta})), \end{aligned}$$

donde  $A_{\underline{\beta}}^c$  es el complemento de  $A_{\underline{\beta}}$ , de donde se obtiene el resultado de la Proposición.

A continuación se estudiará la consistencia de los M-estimadores propuestos.

Sea  $\mathbf{Z}_1 = (\mathbf{X}'_1, Y_1)', \dots, \mathbf{Z}_n = (\mathbf{X}'_n, Y_n)'$  una muestra de vectores aleatorios en  $\mathfrak{R}^{q+1}$  independientes e idénticamente distribuidos tales que  $\mathbf{Z}_i$  tiene distribución  $P_Z = P_{XY}$  que satisface

$$P(Y = 1 | \mathbf{X} = \mathbf{x}) = H(\mathbf{x}'\underline{\beta}_0).$$

Se supondrá además

C1.  $P(\mathbf{X}'\underline{\alpha} = 0) = 0$  para todo  $\underline{\alpha}$ .

Se probará que la sucesión de M-estimadores  $\hat{\underline{\beta}}_n$  basados en  $\mathbf{Z}_1, \dots, \mathbf{Z}_n$  que se obtiene de minimizar (1.7) converge a  $\underline{\beta}_0$  en casi todo punto.

La demostración de la consistencia del M-estimador sigue en líneas generales el Lema 1 y el Teorema 1 de Huber (1967). La diferencia esencial radica en que la hipótesis A5 del Lema mencionado no se cumple en el presente caso. Por medio de esta hipótesis Huber prueba que la sucesión de estimadores yace ultimamente en un compacto en casi todo punto. Se probará esta propiedad en forma independiente en la Proposición 2.2 y a partir de ella se podrá deducir la consistencia en el Teorema 2.1.

Los siguientes lemas prueban algunos resultados previos que serán necesarios.

LEMA 2.2. Sea  $\rho$  una función que satisface las hipótesis A1 a A3. entonces  $\phi(y, \mathbf{x}, \underline{\beta})$  definida en (2.1) cumple

- i)  $\phi(y, \mathbf{x}, \underline{\beta}) > 0$ , para todo  $y \in [0, 1]$ ,  $\mathbf{x} \in \mathfrak{R}^q$  y  $\underline{\beta} \in \mathfrak{R}^q$
- ii) Existe  $M > 0$  tal que  $\phi(y, \mathbf{x}, \underline{\beta}) < M$
- iii)  $\phi(y, \mathbf{x}, \underline{\beta})$  es continua en cada uno de sus argumentos.

DEMOSTRACIÓN:

- i) Como por A2  $\psi$  es una función no negativa, se tiene que si  $0 < t < 1$

$$G(t) = \int_0^1 \psi(-\ln s) ds$$

también lo es.

Además, como por A1  $\rho$  es positiva, por lo menos alguno de los términos que involucra a  $\rho$  también lo es, con lo que queda demostrado i).

- ii) Por A1 se verifica que  $G(1) = \int_0^1 \psi(-\ln s) ds \leq L$ , luego

$$\phi(y, \mathbf{x}, \underline{\beta}) \leq yK + (1 - y)K + 2G(1) \leq K + 2L$$

de manera que basta tomar  $M = K + 2L$  para que se cumpla ii).

- iii)  $\phi(y, \mathbf{x}, \underline{\beta})$  es continua por ser composición de funciones continuas.

LEMA 2.3. Si  $\rho$  satisface A1 a A3, entonces dado  $\mathbf{t} \in \mathfrak{R}^q$  tal que  $\|\mathbf{t}\| = 1$  y  $\lambda \in \mathfrak{R}$

$$E_{\underline{\beta}_0} \left( \lim_{\lambda \rightarrow \infty} \phi(Y, \mathbf{X}, \lambda \mathbf{t}) \right) = \lim_{\lambda \rightarrow \infty} E_{\underline{\beta}_0} \left( \phi(Y, \mathbf{X}, \lambda \mathbf{t}) \right) = E_{\underline{\beta}_0} \left( \lim_{\lambda \rightarrow \infty} E(\phi(Y, \mathbf{X}, \lambda \mathbf{t}) | \mathbf{X}) \right).$$

DEMOSTRACIÓN:

Como por el Lema 2.2  $\phi(Y, \mathbf{X}, \underline{\beta})$  es una función acotada y continua, el Teorema de Convergencia Mayorada implica la tesis.

LEMA 2.4. Sea  $\mathbf{X}$  un vector aleatorio en  $\mathfrak{R}^q$  satisfaciendo la condición C1. Luego, dados  $\mathbf{t} \in \mathfrak{R}^q$  tal que  $\|\mathbf{t}\| = 1$  y  $\delta > 0$ , existen números positivos  $\epsilon$  y  $A$  tales que

$$P(|\mathbf{X}'\mathbf{t}| > \epsilon \quad \text{y} \quad \|\mathbf{X}\| < A) > 1 - \delta.$$



DEMOSTRACIÓN: Dado  $\mathbf{t} \in \mathfrak{R}^q$ , se define  $B_n = \{\mathbf{x} : |\mathbf{x}'\mathbf{t}| > \frac{1}{n}\}$  para  $n \in \mathcal{N}$  y  $B = \{\mathbf{x} : |\mathbf{x}'\mathbf{t}| > 0\}$ . Como los conjuntos  $B_n$  forman una sucesión creciente de conjuntos en  $\mathfrak{R}^q$  tal que  $B_n \uparrow B$  y  $\mathbf{X}$  satisface la condición C1, se tiene que

$$\lim_{n \rightarrow \infty} P(\mathbf{X} \in B_n) = P(\mathbf{X} \in B) = 1.$$

Por lo tanto, dado  $\delta > 0$  existe  $\epsilon > 0$  tal que

$$P(|\mathbf{X}'\mathbf{t}| > \epsilon) > 1 - \frac{\delta}{2} \quad (2.6)$$

Por otra parte, dado  $\delta > 0$  existe  $A > 0$  tal que

$$P(\|\mathbf{X}\| < A) > 1 - \frac{\delta}{2} \quad (2.7)$$

De (2.6) y (2.7) resulta la tesis del Lema, pues

$$\begin{aligned} P(\|\mathbf{X}'\mathbf{t}\| > \epsilon \text{ y } \|\mathbf{X}\| < A) &= 1 - P(\|\mathbf{X}'\mathbf{t}\| \leq \epsilon \text{ ó } \|\mathbf{X}\| \geq A) \\ &\geq 1 - P(\|\mathbf{X}'\mathbf{t}\| \leq \epsilon) - P(\|\mathbf{X}\| \geq A) > 1 - \frac{\delta}{2} - \frac{\delta}{2}. \end{aligned}$$

LEMA 2.5. Sean  $\mathbf{t} \in \mathfrak{R}^q$  tal que  $\|\mathbf{t}\| = 1$  y  $\lambda \in \mathfrak{R}$ . Si  $\rho$  cumple las hipótesis A1 a A3 y  $\mathbf{X}$  es un vector aleatorio que cumple C1, entonces

$$E_{\underline{\beta}_0}(\lim_{\lambda \rightarrow \infty} \phi(Y, \mathbf{X}, \lambda \mathbf{t})) - E_{\underline{\beta}_0}(\phi(Y, \mathbf{X}, \underline{\beta}_0)) > 0.$$

DEMOSTRACIÓN:

Por el Lema 2.3 se tiene que

$$E_{\underline{\beta}_0}(\lim_{\lambda \rightarrow \infty} \phi(Y, \mathbf{X}, \lambda \mathbf{t})) = E(\lim_{\lambda \rightarrow \infty} E(\phi(Y, \mathbf{X}, \lambda \mathbf{t})|\mathbf{X})) \quad (2.8)$$

Por otro lado, se obtiene que

$$\begin{aligned} E(\phi(Y, \mathbf{X}, \lambda \mathbf{t})|\mathbf{X}) &= \\ &= H(\mathbf{X}'\underline{\beta}_0)\rho(-\ln(H(\mathbf{X}'\lambda \mathbf{t}))) + (1 - H(\mathbf{X}'\underline{\beta}_0))\rho(-\ln(1 - H(\mathbf{X}'\lambda \mathbf{t}))) \\ &\quad + G(H(\mathbf{X}'\lambda \mathbf{t})) + G(1 - H(\mathbf{X}'\lambda \mathbf{t})) \\ &= L(H(\mathbf{X}'\underline{\beta}_0), H(\lambda \mathbf{X}'\mathbf{t})). \end{aligned}$$

siendo  $L$  la función definida en (2.2). El Lema 2.1 y la hipótesis B1 implican que  $L$  alcanza su único mínimo en  $\mathbf{X}'\underline{\beta}_0$ .

En virtud de B2, B3 y C1 se tiene que

$$\lim_{\lambda \rightarrow \infty} H(\lambda \mathbf{x}'\mathbf{t}) = \begin{cases} 1 & \text{si } \mathbf{x}'\mathbf{t} > 0 \\ 0 & \text{si } \mathbf{x}'\mathbf{t} < 0 \end{cases}$$

con probabilidad uno. En consecuencia, como  $L$  es una función continua, por (2.3) y (2.4) resulta que

$$\begin{aligned} \lim_{\lambda \rightarrow \infty} E(\phi(Y, \mathbf{X}, \lambda \mathbf{t}) | \mathbf{X}) &= L(H(\mathbf{X}'\underline{\beta}_0), \lim_{\lambda \rightarrow \infty} H(\lambda \mathbf{X}'\mathbf{t})) \\ &> L(H(\mathbf{X}'\underline{\beta}_0), H(\mathbf{X}'\underline{\beta}_0)) = E(\phi(Y, \mathbf{X}, \underline{\beta}_0) | \mathbf{X}) \end{aligned} \quad (2.9)$$

con probabilidad uno. Finalmente, dado que  $\mathbf{X}$  cumple C1, de (2.8) y (2.9) se tiene que

$$E_{\underline{\beta}_0}(\lim_{\lambda \rightarrow \infty} \phi(Y, \mathbf{X}, \lambda \mathbf{t})) > E(E(\phi(Y, \mathbf{X}, \underline{\beta}_0) | \mathbf{X})) = E_{\underline{\beta}_0}(\phi(Y, \mathbf{X}, \underline{\beta}_0))$$

como se quería probar.

Dados  $\underline{\alpha} \in \mathbb{R}^q$  y  $\epsilon > 0$ , definimos  $B(\underline{\alpha}, \epsilon) = \{\underline{\beta} \in \mathbb{R}^q : \|\underline{\beta} - \underline{\alpha}\| < \epsilon\}$ .

**LEMA 2.6.** *Sea  $\mathbf{X} \in \mathbb{R}^q$  es un vector aleatorio que cumple la condición C1. Si  $\phi$  es una función que satisface las hipótesis A1 a A3, entonces*

$$\lim_{\epsilon \rightarrow 0} E(\inf_{\underline{\beta} \in B(\underline{\alpha}, \epsilon)} \phi(Y, \mathbf{X}, \underline{\beta}) - \phi(Y, \mathbf{X}, \underline{\alpha})) = 0.$$

**DEMOSTRACIÓN:**

Por ser  $\phi$  una función positiva cumple que

$$0 \leq \phi(Y, \mathbf{X}, \underline{\alpha}) - \inf_{\underline{\beta} \in B(\underline{\alpha}, \epsilon)} \phi(Y, \mathbf{X}, \underline{\beta}) \leq \phi(Y, \mathbf{X}, \underline{\alpha}) < M$$

siendo  $M$  la constante del Lema 2.2 parte ii). Dado que  $\phi$  es una función continua de sus argumentos, el Lema se obtiene aplicando el Teorema de Convergencia Mayorada.

**LEMA 2.7.** *Sea  $\phi$  una función que cumple las condiciones A1 a A3 y  $\mathbf{X} \in \mathbb{R}^q$  un vector aleatorio que satisface la hipótesis C1. Dado  $\mathbf{t} \in \mathbb{R}^q$  tal que  $\|\mathbf{t}\| = 1$ , existe  $\epsilon > 0$  tal que :*

$$E_{\underline{\beta}_0}(\liminf_{\lambda \rightarrow \infty} \inf_{\mathbf{s} \in B(\mathbf{t}, \epsilon)} \phi(Y, \mathbf{X}, \lambda \mathbf{s})) > E_{\underline{\beta}_0}(\phi(Y, \mathbf{X}, \underline{\beta}_0)).$$

DEMOSTRACIÓN:

Sea  $\mathbf{t} \in \mathbb{R}^q$  tal que  $\|\mathbf{t}\| = 1$ . Dado  $\delta > 0$ , por el Lema 2.4, existen  $\epsilon > 0$  y  $A > 0$  tales que

$$P(|\mathbf{X}'\mathbf{t}| > \epsilon \text{ y } \|\mathbf{X}\| < A) > 1 - \delta,$$

y sea

$$C_\epsilon = \{\mathbf{x} : |\mathbf{x}'\mathbf{t}| > \epsilon \text{ y } \|\mathbf{x}\| < A\}.$$

Tomando  $\eta = \epsilon/2A$ , se verá que si  $\mathbf{x} \in C_\epsilon$  y  $\mathbf{s} \in B(\mathbf{t}, \eta)$  se cumple que

$$\text{sg}(\mathbf{x}'\mathbf{t}) = \text{sg}(\mathbf{x}'\mathbf{s}) \text{ y } |\mathbf{x}'\mathbf{s}| > \epsilon/2 \quad (2.10)$$

donde con  $\text{sg}(r)$  notamos el signo de  $r$ .

Si  $\mathbf{x} \in C_\epsilon$  y  $\mathbf{s} \in B(\mathbf{t}, \eta)$  entonces

$$|\mathbf{x}'\mathbf{t} - \mathbf{x}'\mathbf{s}| \leq \|\mathbf{x}\| \|\mathbf{t} - \mathbf{s}\| < A\eta < A \frac{\epsilon}{2A} = \epsilon/2.$$

Por lo tanto

$$\mathbf{x}'\mathbf{t} - \epsilon/2 < \mathbf{x}'\mathbf{s} < \mathbf{x}'\mathbf{t} + \epsilon/2$$

y queda demostrado (2.10).

En virtud de que  $\lim_{\lambda \rightarrow \infty} \phi(y, \mathbf{x}, \lambda \mathbf{t})$  sólo depende del signo de  $\mathbf{x}'\mathbf{t}$ , dado  $\mathbf{x} \in C_\epsilon$  por (2.10) resulta que

$$\liminf_{\lambda \rightarrow \infty} \inf_{\mathbf{s} \in B(\mathbf{t}, \eta)} \phi(Y, \mathbf{X}, \lambda \mathbf{s}) = \lim_{\lambda \rightarrow \infty} \phi(Y, \mathbf{X}, \lambda \mathbf{t}). \quad (2.11)$$

Como  $\phi$  es una función positiva, utilizando la igualdad (2.11) se deduce que

$$\begin{aligned} E_{\underline{P}_0}(\liminf_{\lambda \rightarrow \infty} \inf_{\mathbf{s} \in B(\mathbf{t}, \eta)} \phi(Y, \mathbf{X}, \lambda \mathbf{s})) &= \\ &= \int_{C_\epsilon} \liminf_{\lambda \rightarrow \infty} \inf_{\mathbf{s} \in B(\mathbf{t}, \eta)} \phi(Y, \mathbf{X}, \lambda \mathbf{s}) \, dP_{XY} + \int_{C_\epsilon^c} \liminf_{\lambda \rightarrow \infty} \inf_{\mathbf{s} \in B(\mathbf{t}, \eta)} \phi(Y, \mathbf{X}, \lambda \mathbf{s}) \, dP_{XY} \\ &\geq \int_{C_\epsilon} \liminf_{\lambda \rightarrow \infty} \inf_{\mathbf{s} \in B(\mathbf{t}, \eta)} \phi(Y, \mathbf{X}, \lambda \mathbf{s}) \, dP_{XY} = \int_{C_\epsilon} \lim_{\lambda \rightarrow \infty} \phi(Y, \mathbf{X}, \lambda \mathbf{t}) \, dP_{XY} \end{aligned} \quad (2.12)$$

Tomando  $\delta < D/M$ , donde  $M$  es la constante del Lema 2.2 y

$D = E_{\underline{\beta}_0}(\lim_{\lambda \rightarrow \infty} \phi(Y, \mathbf{X}, \lambda t)) - E_{\underline{\beta}_0}(\phi(Y, \mathbf{X}, \beta_0)) > 0$  por el Lema 2.5, usando que  $\phi$  es positiva se obtiene que

$$\begin{aligned}
 E_{\underline{\beta}_0}(\lim_{\lambda \rightarrow \infty} \phi(Y, \mathbf{X}, \lambda t)) &= \int_{C_\delta} \lim_{\lambda \rightarrow \infty} \phi(Y, \mathbf{X}, \lambda t) \, dP_{XY} + \int_{C_\delta^c} \lim_{\lambda \rightarrow \infty} \phi(Y, \mathbf{X}, \lambda t) \, dP_{XY} \\
 &\leq \int_{C_\delta} \lim_{\lambda \rightarrow \infty} \phi(Y, \mathbf{X}, \lambda t) \, dP_{XY} + M \int_{C_\delta^c} dP_{XY} \\
 &\leq \int_{C_\delta} \lim_{\lambda \rightarrow \infty} \phi(Y, \mathbf{X}, \lambda t) \, dP_{XY} + M\delta \\
 &< \int_{C_\delta} \lim_{\lambda \rightarrow \infty} \phi(Y, \mathbf{X}, \lambda t) \, dP_{XY} + D
 \end{aligned} \tag{2.13}$$

De (2.13) resulta que

$$E_{\underline{\beta}_0}(\phi(Y, \mathbf{X}, \beta_0)) < \int_{C_\delta} \lim_{\lambda \rightarrow \infty} \phi(Y, \mathbf{X}, \lambda t) \, dP_{XY} \tag{2.14}$$

Finalmente, de (2.12) y (2.14) se tiene que

$$E_{\underline{\beta}_0}(\liminf_{\lambda \rightarrow \infty} \inf_{\mathbf{s} \in B(\mathbf{t}, \eta)} \phi(Y, \mathbf{X}, \lambda \mathbf{s})) > E_{\underline{\beta}_0}(\phi(Y, \mathbf{X}, \beta_0))$$

como se quería probar.

**PROPOSICIÓN 2.2.** *Si se cumplen las hipótesis A1 a A3 y C1, la sucesión de estimadores  $\widehat{\underline{\beta}}_n$  que se obtiene de minimizar (1.7) yace ultimamente en un compacto en casi todo punto.*

**DEMOSTRACIÓN:**

Sea  $S = \{\mathbf{t} \in \mathbb{R}^q : \|\mathbf{t}\| = 1\}$ . Dado  $\mathbf{t} \in S$ , por el Lema 2.7 y el Teorema de Convergencia Mayorada existen números positivos  $\varepsilon_t$ ,  $\eta_t$  y  $M_t$  tales que

$$E_{\underline{\beta}_0}(\inf_{\lambda > M_t} \inf_{\mathbf{s} \in B(\mathbf{t}, \varepsilon_t)} \phi(Y, \mathbf{X}, \lambda \mathbf{s})) > E_{\underline{\beta}_0}(\phi(Y, \mathbf{X}, \underline{\beta}_0)) + 3\eta_t.$$

La Ley de los Grandes Números implica que existe  $n_t$  tal que si  $n \geq n_t$ , entonces

$$\begin{aligned}
 \inf_{\lambda > M_t} \inf_{\mathbf{s} \in B(\mathbf{t}, \varepsilon_t)} \frac{1}{n} \sum_{i=1}^n \phi(Y_i, \mathbf{X}_i, \lambda \mathbf{s}) &\geq \frac{1}{n} \sum_{i=1}^n \inf_{\lambda > M_t} \inf_{\mathbf{s} \in B(\mathbf{t}, \varepsilon_t)} \phi(Y_i, \mathbf{X}_i, \lambda \mathbf{s}) \\
 &> E_{\underline{\beta}_0}(\phi(Y, \mathbf{X}, \underline{\beta}_0)) + 2\eta_t.
 \end{aligned}$$

en casi todo punto.

Sean  $B(t_i, \varepsilon_{t_i})$   $1 \leq i \leq N$  un subcubrimiento finito del compacto  $S$ . Sean  $M = \max(M_1, \dots, M_N)$ ,  $n_1^* = \max(n_1, \dots, n_N)$ ,  $\varepsilon = \min(\varepsilon_1, \dots, \varepsilon_N)$  y  $\eta = \min(\eta_1, \dots, \eta_N)$ . Luego, si  $n \geq n_1^*$

$$\inf_{\lambda > M} \inf_{\mathbf{s} \in S} \frac{1}{n} \sum_{i=1}^n \phi(Y_i, \mathbf{X}_i, \lambda \mathbf{s}) > E_{\underline{\beta}_0}(\phi(Y, \mathbf{X}, \underline{\beta}_0)) + 2\eta \quad (2.15)$$

en casi todo punto.

Por otro lado, por la Ley de los Grandes Números existe  $n_2^*$  tal que si  $n \geq n_2^*$ , entonces

$$\frac{1}{n} \sum_{i=1}^n \phi(Y_i, \mathbf{X}_i, \underline{\beta}_0) < E_{\underline{\beta}_0}(\phi(Y, \mathbf{X}, \underline{\beta}_0)) + \eta \quad (2.16)$$

en casi todo punto. En consecuencia tomando  $n^* = \max(n_1^*, n_2^*)$  y

$C = \{\underline{\beta} \in \mathbb{R}^q : \|\underline{\beta}\| \leq M\}$ , por (2.15) y (2.16) se obtiene

$$\frac{1}{n} \sum_{i=1}^n \phi(Y_i, \mathbf{X}_i, \underline{\beta}_0) < \inf_{\underline{\beta} \in C} \frac{1}{n} \sum_{i=1}^n \phi(Y_i, \mathbf{X}_i, \underline{\beta})$$

en casi todo punto. Por lo tanto

$$P(\exists n^* : n \geq n^* \quad \widehat{\underline{\beta}}_n \in C) = 1$$

y la proposición queda demostrada.

**TEOREMA 2.1.** Si se satisfacen A1 a A3 y C1, entonces la sucesión de estimadores  $\widehat{\underline{\beta}}_n$  que se obtiene de minimizar (1.7) converge a  $\underline{\beta}_0$  en casi todo punto.

**DEMOSTRACIÓN:**

En virtud de la Proposición 2.2, existe un compacto  $C$  tal que

$$P(\exists n^* : n \geq n^* \quad \widehat{\underline{\beta}}_n \in C) = 1.$$

Sea  $U$  un entorno abierto de  $\underline{\beta}_0$ . Dado que  $E_{\underline{\beta}}(\phi(Y, \mathbf{X}, \underline{\beta}))$  es una función continua en  $\underline{\beta}$ , ésta alcanza su ínfimo en el compacto  $C - U$ . Como por la Proposición 2.1, dado  $\underline{\beta} \neq \underline{\beta}_0$ ,  $E_{\underline{\beta}_0}(\phi(Y, \mathbf{X}, \underline{\beta})) < E_{\underline{\beta}}(\phi(Y, \mathbf{X}, \underline{\beta}))$ , entonces existe  $\eta > 0$  tal que

$$\inf_{\underline{\beta} \in C-U} E_{\underline{\beta}_0}(\phi(Y, \mathbf{X}, \underline{\beta})) = \min_{\underline{\beta} \in C-U} E_{\underline{\beta}}(\phi(Y, \mathbf{X}, \underline{\beta})) > E_{\underline{\beta}_0}(\phi(Y, \mathbf{X}, \underline{\beta}_0)) + 4\eta. \quad (2.17)$$

El Lema 2.6 implica que dado  $\underline{\alpha} \in C - U$  existe  $\varepsilon_\alpha$  tal que

$$E_{\underline{\beta}_0} \left( \inf_{\underline{\beta} \in B(\underline{\alpha}, \varepsilon_\alpha)} \phi(Y, \mathbf{X}, \underline{\beta}) \right) > E_{\underline{\beta}_0} (\phi(Y, \mathbf{X}, \underline{\alpha})) - \eta. \quad (2.18)$$

(2.17) y (2.18) implican que si  $\underline{\alpha} \in C - U$ , entonces existe  $\varepsilon_\alpha$  tal que

$$E_{\underline{\beta}_0} \left( \inf_{\underline{\beta} \in B(\underline{\alpha}, \varepsilon_\alpha)} \phi(Y, \mathbf{X}, \underline{\beta}) \right) > E_{\underline{\beta}_0} (\phi(Y, \mathbf{X}, \underline{\beta}_0)) + 3\eta. \quad (2.19)$$

Dada la compacidad de  $C - U$ , existen  $B(\underline{\alpha}_1, \varepsilon_{\alpha_1}), \dots, B(\underline{\alpha}_N, \varepsilon_{\alpha_N})$  que cubren a  $C - U$ .

Por la Ley de los Grandes Números para casi todo punto se puede encontrar  $n_j, j = 1, \dots, N$  tal que si  $n > n_j$

$$\frac{1}{n} \sum_{i=1}^n \inf_{\underline{\beta} \in B(\underline{\alpha}_j, \varepsilon_{\alpha_j})} \phi(Y_i, \mathbf{X}_i, \underline{\beta}) > E_{\underline{\beta}_0} \left( \inf_{\underline{\beta} \in B(\underline{\alpha}_j, \varepsilon_{\alpha_j})} \phi(Y, \mathbf{X}, \underline{\beta}) \right) - \eta. \quad (2.20)$$

De (2.19) y (2.20) se deduce que si  $n > n_j, j = 1, \dots, N$ , entonces

$$\begin{aligned} \inf_{\underline{\beta} \in B(\underline{\alpha}_j, \varepsilon_{\alpha_j})} \frac{1}{n} \sum_{i=1}^n \phi(Y_i, \mathbf{X}_i, \underline{\beta}) &\geq \frac{1}{n} \sum_{i=1}^n \inf_{\underline{\beta} \in B(\underline{\alpha}_j, \varepsilon_{\alpha_j})} \phi(Y_i, \mathbf{X}_i, \underline{\beta}) \\ &> E_{\underline{\beta}_0} \left( \inf_{\underline{\beta} \in B(\underline{\alpha}_j, \varepsilon_{\alpha_j})} \phi(Y, \mathbf{X}, \underline{\beta}) \right) - \eta > E_{\underline{\beta}_0} (\phi(Y, \mathbf{X}, \underline{\beta}_0)) + 2\eta \end{aligned} \quad (2.21)$$

en casi todo punto.

Por otro lado, usando la Ley de los Grandes Números con probabilidad 1 existe  $n_0$  tal que  $n \geq n_0$  implica

$$\inf_{\underline{\beta}} \frac{1}{n} \sum_{i=1}^n \phi(Y_i, \mathbf{X}_i, \underline{\beta}) \leq \frac{1}{n} \sum_{i=1}^n \phi(Y_i, \mathbf{X}_i, \underline{\beta}_0) < E_{\underline{\beta}_0} (\phi(Y, \mathbf{X}, \underline{\beta}_0)) + \eta, \quad (2.22)$$

Si  $n_1^* = \max(n_0, n_1, \dots, n_N)$ , de (2.21) y (2.22) se tiene

$$\inf_{\underline{\beta} \in C-U} \frac{1}{n} \sum_{i=1}^n \phi(Y_i, \mathbf{X}_i, \underline{\beta}) > E_{\underline{\beta}_0} (\phi(Y_i, \mathbf{X}_i, \underline{\beta}_0)) + 2\eta > \inf_{\underline{\beta}} \frac{1}{n} \sum_{i=1}^n \phi(Y_i, \mathbf{X}_i, \underline{\beta}) + \eta \quad (2.23)$$

en casi todo punto. para todo  $n \geq n_1^*$ . Luego,  $\widehat{\underline{\beta}}_n \notin U$  para  $n \geq n_1^*$  implicaría que en casi todo punto  $\widehat{\underline{\beta}}_n \in C - U$ , es decir

$$\frac{1}{n} \sum_{i=1}^n \phi(Y_i, \mathbf{X}_i, \widehat{\underline{\beta}}_n) > \inf_{\underline{\beta}} \frac{1}{n} \sum_{i=1}^n \phi(Y_i, \mathbf{X}_i, \underline{\beta}) + \eta.$$

absurdo que provino de suponer que  $\widehat{\underline{\beta}}_n \notin U$ .

### 3. NORMALIDAD ASINTÓTICA

En esta sección se estudiará la distribución asintótica del M-estimador propuesto. Se analizarán dos casos, en el primero supondremos que  $H$  es la función logística, es decir

$$H(t) = \frac{\exp(t)}{1 + \exp(t)}$$

y en el segundo se considerarán funciones de enlace  $H$  más generales.

En esta sección se supondrá

$$C2. P(\mathbf{X}'\underline{\beta} = d) = 0, d \in \mathfrak{R}, \underline{\beta} \in \mathfrak{R}^q.$$

Se define el vector de dimensión  $q$  de derivadas parciales de  $\phi$  con respecto a  $\underline{\beta}$

$$\Psi(y, \mathbf{x}, \underline{\beta}) = \left( \frac{\partial \phi(y, \mathbf{x}, \underline{\beta})}{\partial \underline{\beta}} \right)'$$

y la matriz de  $q \times q$  de derivadas parciales segundas

$$J(y, \mathbf{x}, \underline{\beta}) = \frac{\partial^2 \phi(y, \mathbf{x}, \underline{\beta})}{\partial \underline{\beta}^2} = \frac{\partial \Psi(y, \mathbf{x}, \underline{\beta})}{\partial \underline{\beta}}.$$

Por A5  $\psi$  será derivable salvo a lo sumo en un número finito de puntos. Podemos definir  $\psi'$  arbitrariamente en estos puntos para que esté definida por lo menos formalmente en todo punto (aunque no sea la derivada de los mismos) y así también  $J(y, \mathbf{x}, \underline{\beta})$  estará definida formalmente en todo punto.

En esta sección se supondrá que la función de enlace  $H$  satisface las hipótesis B1 a B5.

### 3.1 Caso logístico.

Se harán las siguientes hipótesis sobre  $\psi$ :

A4.  $\psi(-\ln p)$  es una función lipschitziana de  $p \in (0, 1)$  de constante  $K^*$ .

A5.  $\psi$  tiene derivada  $\psi'$ , salvo a lo sumo en un conjunto finito de puntos  $c_i$ ,  $1 \leq i \leq k$  y  $\psi'$  es acotada y continua en  $[0, c_1)$ ,  $(c_i, c_{i+1})$ ,  $1 \leq i < k$  y  $(c_k, \infty)$ .

#### Ejemplos.

Las familias  $\rho_c$  y  $\rho_\lambda^*$ ,  $\lambda \geq 1$  definidas en la sección anterior satisfacen A4 y A5. En el caso de  $\rho_c$ , se tiene que  $k = 1$  pues  $c$  es el único punto donde  $\rho$  no es derivable, mientras que la derivada de  $\rho_\lambda^*$  existe en todo punto.

Los resultados obtenidos sobre la distribución asintótica se hallan en el Teorema 3.1.

Será necesario probar previamente los siguientes resultados.

LEMA 3.1. Si  $\psi$  satisface A4, entonces

$$q(y, p) = (p - y)[\psi(-\ln p)(1 - p) + \psi(-\ln(1 - p))p] \quad (3.1)$$

definida para  $y, p \in (0, 1)$  es lipschitziana de  $p$  independiente de  $y$ .

DEMOSTRACIÓN:

Se quiere mostrar que existe una constante  $M_q$  tal que

$$|q(y, p) - q(y, p^*)| < M_q |p - p^*| \quad \text{si } p, p^* \in (0, 1). \quad (3.2)$$

Se tiene que

$$\begin{aligned} |q(y, p) - q(y, p^*)| &\leq \\ &\leq |y(1 - p)\psi(-\ln p) - y(1 - p^*)\psi(-\ln p^*)| \\ &\quad + |p(1 - p)\psi(-\ln p) - p^*(1 - p^*)\psi(-\ln p^*)| \\ &\quad + |yp\psi(-\ln(1 - p)) - yp^*\psi(-\ln(1 - p^*))| \\ &\quad + |p^2\psi(-\ln(1 - p)) - (p^*)^2\psi(-\ln(1 - p^*))| \end{aligned} \quad (3.3)$$



En lo que sigue se acotan cada uno de los términos que figura en el miembro derecho de (3.3).

Como  $\psi(-\ln p)$  es lipschitziana de constante  $K^*$ , entonces

$$|\psi(-\ln p) - \psi(-\ln p^*)| \leq K^* |p - p^*|. \quad (3.4)$$

i) Sumando y restando  $y(1 - p^*)\psi(-\ln p)$  resulta

$$\begin{aligned} & |y(1 - p)\psi(-\ln p) - y(1 - p^*)\psi(-\ln p^*)| \leq \\ & \leq |y||1 - p^*||\psi(-\ln p) - \psi(-\ln p^*)| + |\psi(-\ln p)||y|(1 - p) - (1 - p^*)| \\ & \leq K^* |p - p^*| + L|p - p^*| = (K^* + L)|p - p^*| \end{aligned} \quad (3.5)$$

donde  $L$  es la constante de A1.

ii) Sumando y restando  $p^*(1 - p^*)\psi(-\ln p)$  se obtiene que

$$\begin{aligned} & |p(1 - p)\psi(-\ln p) - p^*(1 - p^*)\psi(-\ln p^*)| \leq \\ & \leq |p(1 - p)\psi(-\ln p) - p^*(1 - p^*)\psi(-\ln p)| \\ & \quad + |p^*(1 - p^*)\psi(-\ln p) - p^*(1 - p^*)\psi(-\ln p^*)| \\ & \leq |\psi(-\ln p)||p - p^2 - p^* + (p^*)^2| + |p^*(1 - p^*)||\psi(-\ln p) - \psi(-\ln p^*)| \\ & \leq L(|p^* - p| + |(p^*)^2 - p^2|) + 1/4K^*|p - p^*| \\ & = L(|p - p^*| + |p - p^*||p + p^*|) + 1/4K^*|p - p^*| \\ & \leq (3L + K^*/4)|p - p^*| \end{aligned} \quad (3.6)$$

donde  $L$  es la constante de A1.

iii) Dado que  $|y| \leq 1$ , sumando y restando  $p^* \psi(-\ln(1-p))$  se obtiene que

$$\begin{aligned}
& |yp\psi(-\ln(1-p)) - yp^*\psi(-\ln(1-p^*))| \leq \\
& \leq |p\psi(-\ln(1-p)) - p^*\psi(-\ln(1-p))| \\
& + |p^*\psi(-\ln(1-p)) - p^*\psi(-\ln(1-p^*))| \\
& \leq |\psi(-\ln(1-p))||p - p^*| + |p^*||\psi(-\ln(1-p)) - \psi(-\ln(1-p^*))| \\
& \leq L|p - p^*| + K^*|p - p^*| = (L + K^*)|p - p^*|
\end{aligned} \tag{3.7}$$

iv) Nuevamente sumando y restando convenientemente se tiene que

$$\begin{aligned}
& |p^2\psi(-\ln(1-p)) - (p^*)^2\psi(-\ln(1-p^*))| \leq \\
& \leq |p^2\psi(-\ln(1-p)) - (p^*)^2\psi(-\ln(1-p))| + |(p^*)^2\psi(-\ln(1-p)) - (p^*)^2\psi(-\ln(1-p^*))| \\
& = |\psi(-\ln(1-p))||p^2 - (p^*)^2| + |p^*|^2|\psi(-\ln(1-p)) - \psi(-\ln(1-p^*))| \\
& \leq L|p - p^*||p + p^*| + K^*|p - p^*| \leq (2L + K^*)|p - p^*|
\end{aligned} \tag{3.8}$$

De (3.5), (3.6), (3.7) y (3.8) se obtiene que

$$|q(y, p) - q(y, p^*)| \leq (2(L + K^*) + (3L + K^*/4) + (2L + K^*))|p - p^*|,$$

de manera que basta tomar  $M_q = 2(L + K^*) + (3L + K^*/4) + (2L + K^*)$  para que se cumpla (3.2).

**PROPOSICIÓN 3.1.** Si  $\mathbf{X} \in \mathbb{R}^q$  es un vector aleatorio que cumple la condición C2, entonces  $E_{\underline{\beta}_0}(J(Y, \mathbf{X}, \underline{\beta}_0))$  es una matriz invertible.

**DEMOSTRACIÓN:**

Sea

$$\Sigma = E_{\underline{\beta}_0}(J(Y, \mathbf{X}, \underline{\beta}_0)). \tag{3.9}$$

Tomando esperanza condicional se tiene que

$$\Sigma_{ij} = E(X_i X_j H(\mathbf{X}' \underline{\beta}_0)(1 - H(\mathbf{X}' \underline{\beta}_0))W'(H(\mathbf{X}' \underline{\beta}_0))),$$

donde

$$W(p) = \psi(-\ln p)(1-p) + \psi(-\ln(1-p))p > 0, \quad p \in (0, 1).$$

Se define  $\mathbf{T} = (T_1, \dots, T_q)$  por

$$T_i = X_i [H(\mathbf{X}'\underline{\beta}_0)(1 - H(\mathbf{X}'\underline{\beta}_0))W(H(\mathbf{X}'\underline{\beta}_0))]^{1/2}.$$

Luego, se tiene que

$$\Sigma_{ij} = E(T_i T_j).$$

Se probará que  $\Sigma$  es no singular, con lo que quedará demostrada la tesis.

Dado  $\mathbf{l} \in \mathbb{R}^q, \mathbf{l} \neq 0$ , se tiene

$$E((\mathbf{l}'\mathbf{T})^2) = \sum_{i=1}^q \sum_{j=1}^q l_i l_j E(T_i T_j) = \mathbf{l}'\Sigma\mathbf{l} > 0$$

a menos que  $P(\mathbf{l}'\mathbf{T} = 0) = 1$ .

Dado que  $p(1-p)W(p) > 0$  si  $0 < p < 1$  y  $\mathbf{X}$  cumple la hipótesis C2, se verifica que

$$P(\mathbf{l}'\mathbf{T} = 0) = P(\mathbf{l}'\mathbf{X}[H(\mathbf{X}'\underline{\beta}_0)(1 - H(\mathbf{X}'\underline{\beta}_0))W(H(\mathbf{X}'\underline{\beta}_0))]^{1/2} = 0) = P(\mathbf{l}'\mathbf{X} = 0) = 0$$

En consecuencia,  $\Sigma$  es definida positiva, con lo que la proposición queda demostrada.

**TEOREMA 3.1.** *Supóngase que se verifican las condiciones A1 a A5 y sea  $\mathbf{X}$  un vector aleatorio en  $\mathbb{R}^q$  tal que  $E(\|\mathbf{X}\|^3) < \infty$  y satisface C2. Si  $\widehat{\underline{\beta}}_n$  es una sucesión de estimadores definidos por la minimización de (1.7), entonces  $\sqrt{n}(\widehat{\underline{\beta}}_n - \underline{\beta}_0)$  converge en distribución a un vector normal multidimensional con media 0 y matriz de covarianza  $C(\underline{\beta}_0)$  dada por*

$$C(\underline{\beta}_0) = E(J(Y, \mathbf{X}, \underline{\beta}_0))^{-1} E(\Psi(Y, \mathbf{X}, \underline{\beta}_0)\Psi(Y, \mathbf{X}, \underline{\beta}_0)') (E(J(Y, \mathbf{X}, \underline{\beta}_0))^{-1})'.$$

**DEMOSTRACIÓN:**

Sean  $\lambda(\underline{\beta}) = E(\Psi(Y, \mathbf{X}, \underline{\beta})) = E(\mathbf{X}q(Y, H(\mathbf{X}'\underline{\beta})))$ , siendo  $q(y, p)$  la función definida en (3.1) y

$$U(y, \mathbf{x}, \underline{\beta}, d) = \sup_{\|\underline{\tau} - \underline{\beta}\| \leq d} \|\Psi(y, \mathbf{x}, \underline{\beta}) - \Psi(y, \mathbf{x}, \underline{\tau})\|.$$

De acuerdo con el Teorema 3 de Huber (1967) y su Corolario, es suficiente probar que existen números positivos  $a, b, c, d_0$  tales que

$$(i) \|\lambda(\underline{\beta})\| \geq a\|\underline{\beta} - \underline{\beta}_0\| \text{ si } \|\underline{\beta} - \underline{\beta}_0\| \leq d_0$$

$$(ii) E(U(Y, \mathbf{X}, \underline{\beta}, d)) \leq b.d$$

$$(iii) E(U^2(Y, \mathbf{X}, \underline{\beta}, d)) \leq c.d$$

y que  $E(J(Y, \mathbf{X}, \underline{\beta}_0))$  es no singular.

Se comenzará probando que  $\lambda(\underline{\beta})$  es diferenciable en  $\underline{\beta} = \underline{\beta}_0$  y que

$$\left[ \frac{\partial \lambda(\underline{\beta})}{\partial \underline{\beta}} \right]_{\underline{\beta}=\underline{\beta}_0} = E_{\underline{\beta}_0}(J(Y, \mathbf{X}, \underline{\beta}_0)) = \Sigma. \quad (3.10)$$

Por el Lema 3.1 existe  $M_q$  tal que para todo  $y$

$$|q(y, p) - q(y, p^*)| < M_q|p - p^*|. \quad (3.11)$$

Si  $P_{\mathbf{X}Y}$  es la distribución conjunta del vector aleatorio  $(X, Y)$ , se tiene que

$$\begin{aligned} \frac{\partial \lambda_i(\underline{\beta})}{\partial \beta_j} &= \lim_{h \rightarrow 0} \frac{\lambda_i(\underline{\beta} + \mathbf{e}_j h) - \lambda_i(\underline{\beta})}{h} \\ &= \lim_{h \rightarrow 0} \int \mathbf{x}_i \frac{[q(y, H(\mathbf{x}'(\underline{\beta} + \mathbf{e}_j h))) - q(y, H(\mathbf{x}'\underline{\beta}))]}{h} dP_{\mathbf{X}Y}. \end{aligned} \quad (3.12)$$

donde con  $\mathbf{e}_1, \dots, \mathbf{e}_q$  notamos la base canónica de  $\mathbb{R}^q$ . Dado que con probabilidad 1 el integrando de (3.12) converge a  $\partial \Psi_i(y, \mathbf{x}, \underline{\beta})/\partial \beta_j$ , y  $q(y, p)$  es lipschitziana en  $p$  y  $H$  es lipschitziana de constante 1 y  $E(\|\mathbf{X}\|^2) < \infty$ , el Teorema de Convergencia Dominada implica que

$$\frac{\partial \lambda_i(\underline{\beta})}{\partial \beta_j} = \int \lim_{h \rightarrow 0} \mathbf{x}_i \frac{[q(y, H(\mathbf{x}'(\underline{\beta} + \mathbf{e}_j h))) - q(y, H(\mathbf{x}'\underline{\beta}))]}{h} dP_{\mathbf{X}Y}.$$

Por lo tanto existen las derivadas parciales en un entorno de  $\underline{\beta}_0$ .

Para demostrar la diferenciable de  $\lambda(\underline{\beta})$  en  $\underline{\beta} = \underline{\beta}_0$ , es suficiente ver que  $\partial \lambda_i(\underline{\beta})/\partial \beta_j$

es continua en  $\underline{\beta} = \underline{\beta}_0$ . Se tiene que

$$\begin{aligned}
& \frac{\partial \lambda_i(\underline{\beta})}{\partial \beta_j} \\
&= \int \mathbf{x}_i \mathbf{x}_j \{ H(\mathbf{x}' \underline{\beta})(1 - H(\mathbf{x}' \underline{\beta})) [\psi(-\ln(H(\mathbf{x}' \underline{\beta}))(1 - H(\mathbf{x}' \underline{\beta}))) \\
&\quad + \psi(-\ln(1 - H(\mathbf{x}' \underline{\beta}))) H(\mathbf{x}' \underline{\beta})] + (H(\mathbf{x}' \underline{\beta}) - Y) [-\psi'(-\ln(H(\mathbf{x}' \underline{\beta}))(1 - H(\mathbf{x}' \underline{\beta})))^2 \\
&\quad - \psi'(-\ln(H(\mathbf{x}' \underline{\beta}))) H(\mathbf{x}' \underline{\beta})(1 - H(\mathbf{x}' \underline{\beta})) + \psi'(-\ln(1 - H(\mathbf{x}' \underline{\beta}))) (H(\mathbf{x}' \underline{\beta}))^2 \\
&\quad + \psi'(-\ln(1 - H(\mathbf{x}' \underline{\beta}))) H(\mathbf{x}' \underline{\beta})(1 - H(\mathbf{x}' \underline{\beta}))] \} dP_{\mathbf{X}Y} \\
&= \int \frac{\partial \Psi_i(\mathbf{y}, \mathbf{x}, \underline{\beta})}{\partial \beta_j} dP_{\mathbf{X}Y}. \tag{3.13}
\end{aligned}$$

En virtud de las hipótesis A1, A5, B's y C2, el factor entre llaves de (3.13) es acotado y continuo en casi todo punto. Aplicando el Teorema de Convergencia Dominada se deduce que  $\partial \lambda_i(\underline{\beta})/\partial \beta_j$  es continua en  $\underline{\beta}_0$ . Por lo tanto se puede intercambiar la diferenciación con la integración, con lo que queda demostrada la igualdad (3.10).

La diferenciabilidad de  $\lambda$  implica que  $\|\lambda(\underline{\beta}) - \Sigma(\underline{\beta} - \underline{\beta}_0)\| = o(\|\underline{\beta} - \underline{\beta}_0\|)$ . Como por la Proposición 3.1  $\Sigma$  es invertible, se obtiene que

$$\|\lambda(\underline{\beta})\| \geq \|\Sigma(\underline{\beta} - \underline{\beta}_0)\| - o(\|\underline{\beta} - \underline{\beta}_0\|) \geq \|\Sigma^{-1}\|^{-1} \|\underline{\beta} - \underline{\beta}_0\| - o(\|\underline{\beta} - \underline{\beta}_0\|). \tag{3.14}$$

donde con  $\|A\|$  se nota el módulo del máximo autovalor de  $A \in \mathbb{R}^{q \times q}$

Por otra parte, existe  $d_0 > 0$  tal que si  $\|\underline{\beta} - \underline{\beta}_0\| < d_0$  entonces

$$o(\|\underline{\beta} - \underline{\beta}_0\|) < \frac{1}{2} \|\Sigma^{-1}\|^{-1} \|\underline{\beta} - \underline{\beta}_0\|. \tag{3.15}$$

Tomando  $a = \frac{1}{2} \|\Sigma^{-1}\|^{-1}$ , de (3.14) y (3.15) se obtiene i), ya que

$$\|\lambda(\underline{\beta})\| \geq 1/2 \|\Sigma^{-1}\|^{-1} \|\underline{\beta} - \underline{\beta}_0\| = a \|\underline{\beta} - \underline{\beta}_0\|.$$

Para demostrar ii) y iii) se acotará  $U(Y, \mathbf{X}, \underline{\beta}, d)$ .

De acuerdo con (3.11) se tiene que

$$\begin{aligned} U(y, \mathbf{x}, \underline{\beta}, d) &= \sup_{\|\underline{\tau} - \underline{\beta}\| \leq d} \|\mathbf{x}q(y, H(\mathbf{x}'\underline{\beta})) - \mathbf{x}q(y, H(\mathbf{x}'\underline{\tau}))\| \\ &\leq \sup_{\|\underline{\tau} - \underline{\beta}\| \leq d} M_q \|\mathbf{x}\| \|H(\mathbf{x}'\underline{\beta}) - H(\mathbf{x}'\underline{\tau})\| \end{aligned}$$

Dado que  $H(t)$  también es lipschitziana en  $t$  de constante 1 se verifica

$$U(y, \mathbf{x}, \underline{\beta}, d) \leq M_q \|\mathbf{x}\|^2 d. \quad (3.16)$$

Eligiendo  $b = E(M_q \|\mathbf{X}\|^2)$ , se obtiene ii).

Finalmente, como  $|H| < 1$

$$U(y, \mathbf{x}, \underline{\beta}, d) \leq \sup_{\|\underline{\tau} - \underline{\beta}\| \leq d} \|\mathbf{x}\| M_q |H(\mathbf{x}'\underline{\beta}) - H(\mathbf{x}'\underline{\tau})| \leq 2M_q \|\mathbf{x}\| \quad (3.17)$$

Tomando  $c = 2(M_q)^2 E(\|\mathbf{X}\|^3)$  de (3.16) y (3.17) resulta iii), ya que

$$E(U^2(Y, \mathbf{X}, \underline{\beta}, d)) \leq 2M_q E(\|\mathbf{X}\|^3) d$$

**Observación:**

Si  $\psi$  es una función derivable en todo punto, entonces se puede relajar la hipótesis C2, ya que en este caso el integrando de (3.12) converge en todo punto.

### 3.2 Caso H general.

En esta sección se demostrará la normalidad asintótica del M-estimador propuesto para el caso en que H es una función de enlace general. El resultado, cuya demostración es análoga a la del caso I, se encuentra en el Teorema 3.2.

Se necesitarán las siguientes hipótesis sobre H y  $\psi$

D1.  $H'(t)$  está acotada por una constante  $C_{H'}$  y es una función lipschitziana de constante  $M_{H'}$ .

D2.  $\psi$  es una función derivable salvo en un número finito de puntos  $c_i, 1 \leq i \leq k$ , con derivada  $\psi'$  continua en  $[0, c_1), (c_i, c_{i+1}), i \leq i < k$  y  $(c_k, \infty)$ .

D3.  $\psi(-\ln p)/p$  es una función acotada y lipschitziana. Se llamará  $C_\psi$  a la constante que acota a  $\psi(-\ln p)/p$  y  $M_\psi$  a la constante de Lipschitz.

D4. Sea

$$r(y, z) = H'(z)(H(z) - y) \left[ \frac{\psi'(-\ln H(z))}{H(z)} + \frac{\psi(-\ln(1 - H(z)))}{1 - H(z)} \right]$$

definida para  $y \in (0, 1)$ ,  $z \in \mathfrak{R}$ , entonces  $E(\|\mathbf{X}\|^2 | \partial r(y, z) / \partial z |_{z=\mathbf{x}'_0}) < \infty$ .

#### Ejemplos.

Si se toma  $H(t) = \Phi(t)$ , donde con  $\Phi$  se nota la función de distribución normal standard y  $\rho$  perteneciente a la familia  $\rho_c$  de la sección anterior, se satisfacen las condiciones D1, D2, D3 y D4.

Serán necesarios los siguientes resultados previos.

**LEMA 3.2.** *Bajo la hipótesis D1,  $H'(z)H(z)$  es una función lipschitziana.*

DEMOSTRACIÓN:

Se tiene que

$$\begin{aligned} |H'(z)H(z) - H'(z^*)H(z^*)| &\leq \\ &\leq |H'(z)H(z) - H'(z^*)H(z)| + |H'(z^*)H(z) - H'(z^*)H(z^*)| \\ &\leq |H(z)||H'(z) - H'(z^*)| + |H'(z^*)||H(z) - H(z^*)| \end{aligned}$$

Como  $H'$  está acotada,  $H$  es lipschitziana con constante que se denominará  $M_H$ . Luego, dado que  $H$  está acotada por 1,  $H'$  por  $C_{H'}$  y ambas son lipschitzianas se obtiene que

$$|H'(z)H(z) - H'(z^*)H(z^*)| \leq (M_{H'} + C_{H'}M_H)|z - z^*|,$$

de manera que basta tomar constante de Lipschitz  $M_{H'H} = M_{H'} + C_{H'}M_H$  para probar el Lema.

LEMA 3.3. Si se satisfacen las hipótesis D1 a D3, entonces  $r(y, z)$  es una función lipschitziana en  $z$  independiente de  $y$ .

DEMOSTRACIÓN:

Se quiere probar que existe  $M_r$  tal que

$$|r(y, z) - r(y, z^*)| < M_r |z - z^*| \quad z, z^* \in \mathfrak{K} \quad (3.18)$$



Como

$$\begin{aligned}
|r(y, z) - r(y, z^*)| &= \tag{3.19} \\
&= |H'(z)(H(z) - y) \left[ \frac{\psi(-\ln H(z))}{H(z)} + \frac{\psi(-\ln(1 - H(z)))}{1 - H(z)} \right] \\
&\quad - H'(z^*)(H(z^*) - y) \left[ \frac{\psi(-\ln H(z^*))}{H(z^*)} + \frac{\psi(-\ln(1 - H(z^*)))}{1 - H(z^*)} \right]| \\
&\leq |H'(z)(H(z) - y)| \left| \frac{\psi(-\ln H(z))}{H(z)} - \frac{\psi(-\ln H(z^*))}{H(z^*)} \right| \\
&\quad + |H'(z)(H(z) - y)| \left| \frac{\psi(-\ln(1 - H(z)))}{1 - H(z)} - \frac{\psi(-\ln(1 - H(z^*)))}{1 - H(z^*)} \right| \\
&\quad + \left| \frac{\psi(-\ln H(z^*))}{H(z^*)} \right| |H'(z)(H(z) - y) - H'(z^*)(H(z^*) - y)| \\
&\quad + \left| \frac{\psi(-\ln(1 - H(z^*)))}{1 - H(z^*)} \right| |H'(z)(H(z) - y) - H'(z^*)(H(z^*) - y)|
\end{aligned}$$

Dado que  $|H'(z)(H(z) - y)| \leq 2|H'(z)| \leq 2C_{H'}$ , llamando  $C_\psi$  a la cota de  $\psi(-\ln p)/p$ , (3.19) se acota por

$$4C_{H'}M_\psi|H(z) - H(z^*)| + 2C_\psi|H'(z)H(z) - H'(z^*)H(z^*)| + 2C_\psi|H'(z) - H'(z^*)| \tag{3.20}$$

Por D1 H tiene derivada acotada, luego es lipschitziana con constante que se denominará  $M_H$  y en consecuencia se puede acotar (3.20) por

$$[4C_{H'}M_HM_\psi + 2C_\psi M_{H'H} + 2C_\psi M_{H'}]|z - z^*|.$$

Finalmente, tomando  $M_r = 4C_{H'}M_HM_\psi + 2C_\psi M_{H'H} + 2C_\psi M_{H'}$  se cumple (3.18).

**TEOREMA 3.2.** *Supóngase que se cumplen las hipótesis A1 a A4, C2, D1 a D4. Sea  $\mathbf{X}$  un vector aleatorio en  $\mathbb{R}^q$  tal que  $E(\|\mathbf{X}\|^3) < \infty$ . Supóngase además que*

i)  $\partial r(y, z)/\partial z$  es continua salvo en aquellos puntos donde  $\psi'$  no existe.

ii)  $E_{\underline{\beta}_0}(J(Y, \mathbf{X}, \underline{\beta}_0))$  es no singular.

Luego, si  $\widehat{\underline{\beta}}_n$  es una sucesión de estimadores definidos como aquel valor de  $\underline{\beta}$  que minimiza (1.7), entonces  $\sqrt{n}(\widehat{\underline{\beta}}_n - \underline{\beta}_0)$  converge en distribución a un vector normal multi-dimensional con media 0 y matriz de covarianza  $C(\underline{\beta}_0)$  dada por

$$C(\underline{\beta}_0) = E(J(Y, \mathbf{X}, \underline{\beta}_0))^{-1} E(\Psi(Y, \mathbf{X}, \underline{\beta}_0)\Psi(Y, \mathbf{X}, \underline{\beta}_0)') (E(J(Y, \mathbf{X}, \underline{\beta}_0))^{-1})'.$$

DEMOSTRACIÓN:

Sean  $\lambda(\underline{\beta})$  y  $U(y, \mathbf{x}, \underline{\beta}, d)$  definidas como en el Teorema 3.1. En este caso se verifica que  $\lambda(\underline{\beta}) = E(\Psi(Y, \mathbf{X}, \underline{\beta})) = E(\mathbf{X}r(Y, \mathbf{X}'\underline{\beta}))$ , siendo  $r(y, z)$  la función definida en D4. Nuevamente, de acuerdo con el Teorema 3 de Huber (1967) y su Corolario, basta probar que existen números positivos  $a, b, c, d_0$  tales que

$$(i) \|\lambda(\underline{\beta})\| \geq a\|\underline{\beta} - \underline{\beta}_0\| \text{ si } \|\underline{\beta} - \underline{\beta}_0\| \leq d_0$$

$$(ii) E(U(Y, \mathbf{X}, \underline{\beta}, d)) \leq b.d$$

$$(iii) E(U^2(Y, \mathbf{X}, \underline{\beta}, d)) \leq c.d$$

y que  $E(J(Y, \mathbf{X}, \underline{\beta}_0))$  es no singular.

Se verá que  $\lambda(\underline{\beta})$  es diferenciable en  $\underline{\beta} = \underline{\beta}_0$  y que

$$\left[ \frac{\partial \lambda(\underline{\beta})}{\partial \underline{\beta}} \right]_{\underline{\beta}=\underline{\beta}_0} = E_{\underline{\beta}_0}(J(Y, \mathbf{X}, \underline{\beta}_0)) = \Sigma \quad (3.21)$$

Por el Lema 3.2 existe  $M_r$  tal que

$$|r(y, z) - r(y, z^*)| < M_r|z - z^*| \quad (3.22)$$

Si  $P_{XY}$  es la distribución conjunta del vector aleatorio  $(X, Y)$ , se tiene que

$$\begin{aligned} \frac{\partial \lambda_i(\underline{\beta})}{\partial \beta_j} &= \lim_{h \rightarrow 0} \frac{\lambda_i(\underline{\beta} + \mathbf{e}_j h) - \lambda_i(\underline{\beta})}{h} \\ &= \lim_{h \rightarrow 0} \int \mathbf{x}_i \frac{[r(y, \mathbf{x}'(\underline{\beta} + \mathbf{e}_j h)) - r(y, \mathbf{x}'\underline{\beta})]}{h} dP_{XY} \end{aligned} \quad (3.23)$$

donde con  $\mathbf{e}_1, \dots, \mathbf{e}_q$  se nota la base canónica de  $\mathfrak{R}^q$ . Dado que  $r(y, z)$  es lipschitziana de  $z$ , se tiene que

$$|\mathbf{x}_i \frac{[r(y, \mathbf{x}'(\underline{\beta} + \mathbf{e}_j h)) - r(y, \mathbf{x}'\underline{\beta})]}{h}| \leq |\mathbf{x}_i| M_r |\mathbf{x}'(\underline{\beta} + \mathbf{e}_j h) - \mathbf{x}'\underline{\beta}| \leq M_r |\mathbf{x}_i| |\mathbf{x}_j|.$$

Dado que con probabilidad 1 el integrando de (3.23) converge a  $\partial \Psi_i(y, \mathbf{x}, \underline{\beta}) / \partial \beta_j$ , el Teorema de Convergencia Dominada implica que

$$\frac{\partial \lambda_i(\underline{\beta})}{\partial \beta_j} = \int \lim_{h \rightarrow 0} \mathbf{x}_i \frac{[q(y, H(\mathbf{x}'(\underline{\beta} + \mathbf{e}_j h))) - q(y, H(\mathbf{x}'\underline{\beta}))]}{h} dP_{XY}.$$

Por lo tanto, se puede intercambiar la diferenciación con la integración y de esta manera queda demostrada la igualdad (3.21).

Para demostrar la diferenciabilidad de  $\lambda(\underline{\beta})$  en  $\underline{\beta} = \underline{\beta}_0$ , es suficiente ver que  $\partial\lambda_i(\underline{\beta})/\partial\beta_j$  es continua en  $\underline{\beta} = \underline{\beta}_0$ . Se tiene que

$$\frac{\partial\lambda_i(\underline{\beta})}{\partial\beta_j} = \int \mathbf{x}_i \mathbf{x}_j \frac{\partial r(\mathbf{y}, \mathbf{x}'\underline{\beta})}{\partial z} dP_{XY}$$

Como  $\partial r(\mathbf{y}, z)/\partial z|_{z=\mathbf{x}'\underline{\beta}}$  es continua en casi todo punto por C2, D2 y B1 a B5 y además por D4  $E(\|\mathbf{X}\|^2 |\partial r(Y, \mathbf{X}'\underline{\beta})/\partial z|) < \infty$ , aplicando el Teorema de Convergencia Dominada se deduce que  $\partial\lambda_i(\underline{\beta})/\partial\beta_j$  es continua en  $\underline{\beta}_0$ .

Como la diferenciabilidad de  $\lambda$  implica que  $\|\lambda(\underline{\beta}) - \Sigma(\underline{\beta} - \underline{\beta}_0)\| = o(\|\underline{\beta} - \underline{\beta}_0\|)$  y por la hipótesis ii)  $\Sigma$  es invertible, resulta que

$$\|\lambda(\underline{\beta})\| \geq \|\Sigma(\underline{\beta} - \underline{\beta}_0)\| - o(\|\underline{\beta} - \underline{\beta}_0\|) \geq \|\Sigma^{-1}\|^{-1} \|\underline{\beta} - \underline{\beta}_0\| - o(\|\underline{\beta} - \underline{\beta}_0\|), \quad (3.24)$$

donde como en el Teorema 3.1 con  $\|A\|$  se nota la norma del máximo autovalor de  $A \in \mathfrak{R}^{q \times q}$ .

Por otra parte, existe  $d_0 > 0$  tal que si  $\|\underline{\beta} - \underline{\beta}_0\| < d_0$ , entonces

$$o(\|\underline{\beta} - \underline{\beta}_0\|) < \frac{1}{2} \|\Sigma^{-1}\|^{-1} \|\underline{\beta} - \underline{\beta}_0\|. \quad (3.25)$$

Si se elige  $a = \frac{1}{2} \|\Sigma^{-1}\|^{-1}$  de (3.24) y (3.25) usando que

$$\|\lambda(\underline{\beta})\| \geq 1/2 \|\Sigma^{-1}\|^{-1} \|\underline{\beta} - \underline{\beta}_0\| = a \|\underline{\beta} - \underline{\beta}_0\|$$

resulta i).

Se acotarán  $U(Y, \mathbf{X}, \underline{\beta}, d)$  para probar ii) y iii).

Dado que  $r(\mathbf{y}, z)$  es lipschitziana en  $z$  de constante  $M_r$ , se tiene que

$$\begin{aligned} U(\mathbf{y}, \mathbf{x}, \underline{\beta}, d) &= \sup_{\|\underline{\tau} - \underline{\beta}\| \leq d} \|\mathbf{x}r(\mathbf{y}, \mathbf{x}'\underline{\beta}) - \mathbf{x}r(\mathbf{y}, \mathbf{x}'\underline{\tau})\| \\ &\leq \sup_{\|\underline{\tau} - \underline{\beta}\| \leq d} M_r \|\mathbf{x}\| \|\mathbf{x}'\underline{\beta} - \mathbf{x}'\underline{\tau}\| \leq M_r d \|\mathbf{x}\|^2 \end{aligned} \quad (3.26)$$

Eligiendo  $b = E(M_r, \|X\|^2)$ , se obtiene ii).

Por otra parte, por las hipótesis B's, D1 y D3  $|r(y, z)|$  está acotada y a al correspondiente cota se la llamará  $C_r$ . Luego, se tiene que

$$U(y, \underline{x}, \underline{\beta}, d) = \sup_{\|\underline{\tau} - \underline{\beta}\| \leq d} \|\underline{x}r(y, \underline{x}'\underline{\beta}) - \underline{x}r(y, \underline{x}'\underline{\tau})\| \leq 2\|\underline{x}\| \sup_z |r(y, z)| \leq 2\|\underline{x}\|C_r \quad (3.27)$$

De (3.26) y (3.27) resulta que

$$(U(y, \underline{x}, \underline{\beta}, d))^2 \leq 2M_r C_r d \|\underline{x}\|^3.$$

de manera que basta tomar  $c = 2M_r C_r E(\|X\|^3)$  para obtener iii)

**Observación:** Como en el Teorema 3.1 cabe observar que si  $\psi$  es continua, se puede relajar hipótesis C2, ya que en este caso el integrando de(3.23) converge en todo punto.

#### 4. EFICIENCIA: RESULTADOS NUMÉRICOS

En esta sección se calcula numericamente la eficiencia asintótica del M-estimador cuando  $\rho$  pertenece a las familias

$$\rho_c(t) = \begin{cases} t - (t^2/2c) & \text{si } 0 \leq t < c \\ c/2 & \text{si } t \geq c \end{cases} \quad (4.1)$$

$$\rho_\lambda^*(t) = \frac{1 - e^{-\lambda t}}{\lambda} \quad \text{si } t \geq 0 \quad (4.2)$$

#### Observación.

Cuando  $\lambda = 1$  ó  $2$   $\rho_\lambda^*$  da origen a un estimador equivalente al de cuadrados mínimos, es decir al estimador que se obtiene de minimizar

$$\frac{1}{n} \sum_{i=1}^n (Y_i - H(\mathbf{X}'_i \underline{\beta}))^2$$

#### DEMOSTRACIÓN:

Si  $\lambda = 1$ , se tiene que

$$\rho_1^*(t) = 1 - \exp(-t), \quad \psi_1^*(t) = \exp(-t),$$

de manera que

$$\begin{aligned} & \min_{\underline{\beta}} \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^2 [Y_{ij} \rho_1^*(-\ln H_j(\mathbf{X}'_i \underline{\beta})) + G(H_j(\mathbf{X}'_i \underline{\beta}))] \\ & \min_{\underline{\beta}} \frac{1}{n} \sum_{i=1}^n [Y_i \rho_1^*(-\ln H(\mathbf{X}'_i \underline{\beta})) + G(H(\mathbf{X}'_i \underline{\beta})) \\ & \quad + (1 - Y_i) \rho_1^*(-\ln(1 - H(\mathbf{X}'_i \underline{\beta}))) + G(1 - H(\mathbf{X}'_i \underline{\beta}))] \\ & = \min_{\underline{\beta}} \frac{1}{n} \sum_{i=1}^n [Y_i^2 - 2Y_i H(\mathbf{X}'_i \underline{\beta}) + (H(\mathbf{X}'_i \underline{\beta}))^2 + \frac{1}{2}] \end{aligned} \quad (4.3)$$

Como  $y_i = 1$  ó  $0$ , (4.3) equivale a

$$\min_{\underline{\beta}} \frac{1}{n} \sum_{i=1}^n [Y_i^2 - 2Y_i H(\mathbf{X}'_i \underline{\beta}) + (H(\mathbf{X}'_i \underline{\beta}))^2 + \frac{1}{2}].$$

que es equivalente a

$$\min_{\underline{\beta}} \frac{1}{n} \sum_{i=1}^n (Y_i - H(\mathbf{X}'_i \underline{\beta}))^2,$$

como se quería probar.

Cuando  $\lambda = 2$  la demostración es análoga.

Es interesante observar que en el modelo de regresión binomial el estimador de mínimos cuadrados da origen a un estimador robusto, a diferencia de lo que ocurre en el modelo de regresión normal. Esto se debe a que bajo el modelo que estudiamos los residuos  $(y - H(\mathbf{x}'\underline{\beta}))$  permanecen acotados, mientras que en el modelo de regresión normal esto no sucede.

El cómputo de la eficiencia asintótica se realizó considerando que  $H$  es la función logística y que  $\mathbf{X}$  tiene distribución normal multivariada de media 0. Dado que el M-estimador es afín invariante, se puede suponer que  $\mathbf{X}$  es  $N_q(0, I)$ . Como en este caso la distribución de  $\mathbf{X}'\underline{\beta}_0$  depende solamente de  $\|\underline{\beta}_0\|$ , se puede suponer que  $\beta_{0i} = 0, i \geq 2$ .

Los valores de  $\beta_{01}$  fueron elegidos de manera tal que

$$P(H(\mathbf{X}'\underline{\beta}) > 0.95 \text{ ó } H(\mathbf{X}'\underline{\beta}) < 0.05) = \alpha$$

para distintos valores de  $\alpha$ .

La Tabla 1 muestra los valores de  $\beta_{01}$  obtenidos para los distintos valores de  $\alpha$  fijados.

Tabla 1

$\alpha$	$\beta_{01}$
0.025	1.314
0.05	1.502
0.10	1.795
0.20	2.300
0.50	4.394

La constante  $c$  de la familia  $\rho_c$  fue elegida como  $c = -\ln p^*$  para  $p^* = 0.025, 0.05, 0.10, 0.15$  y en el caso de la familia  $\rho_\lambda^*$  se tomó  $\lambda = 1$  y  $1.5$ . Si bien otros valores de  $\lambda$  fueron considerados sólo se presentan los resultados que corresponden a  $\lambda = 1$  y  $\lambda = 1.5$ , ya que otros valores dan resultados similares o poco interesantes desde el punto de vista de la robustez.

El estimador de máxima verosimilitud de  $\underline{\beta}_0$ , en el caso en que  $H$  es la función logística, es asintóticamente normal con matriz de covarianza dada por

$$V(\underline{\beta}_0) = E_{\underline{\beta}_0} (H(\mathbf{X}'\underline{\beta}_0)(1 - H(\mathbf{X}'\underline{\beta}_0))\mathbf{X}\mathbf{X}')^{-1}$$

La eficiencia se definió componente a componente como el cociente entre las varianzas asintóticas del estimador de máxima verosimilitud y la del M-estimador, es decir

$$e_{ij} = \frac{V(\underline{\beta}_0)_{ij}}{C(\underline{\beta}_0)_{ij}}$$

donde  $C(\underline{\beta}_0)$  es la matriz dada por el Teorema 3.1.

Bajo las hipótesis consideradas todas las componentes diferentes de la primera tienen la misma eficiencia.

La Tabla 2 muestra las eficiencias de la primera y segunda componentes de los M-estimadores para la familia  $\rho_c$  y la Tabla 3 muestra las correspondientes a la familia  $\rho_\lambda^*$ . Se observa que la eficiencia asintótica del M-estimador disminuye cuando  $\alpha$  crece. Esto es esperable ya que cuando  $\alpha$  crece la probabilidad de que  $H(\mathbf{X}'\underline{\beta}_0)$  ó  $1 - H(\mathbf{X}'\underline{\beta}_0)$  tomen valores cercanos a 1 ó 0 crece y a la vez, el M-estimador da menos peso que el estimador de máxima verosimilitud a las observaciones con probabilidad de éxito pequeña. Esta pérdida de eficiencia disminuye cuando  $p^*$  y  $\lambda$  decrecen dado que aumenta el peso de las observaciones con probabilidad de éxito pequeña.



Tabla 2. Eficiencias Asintóticas del M-estimador con  $\rho_{-\ln p^*}$

		$p^*$							
		0.025		0.05		0.10		0.15	
$\beta_{01}$		ef. $\hat{\beta}_1$	ef. $\hat{\beta}_2$	ef. $\hat{\beta}_1$	ef. $\hat{\beta}_2$	ef. $\hat{\beta}_1$	ef. $\hat{\beta}_2$	ef. $\hat{\beta}_1$	ef. $\hat{\beta}_2$
1.314		0.918	0.977	0.853	0.961	0.745	0.925	0.688	0.892
1.502		0.888	0.970	0.808	0.949	0.693	0.907	0.643	0.71
1.795		0.841	0.958	0.746	0.931	0.628	0.882	0.587	0.844
2.300		0.770	0.941	0.661	0.906	0.550	0.850	0.522	0.810
4.394		0.622	0.901	0.513	0.856	0.423	0.792	0.414	0.753

Tabla 3 Eficiencias asintóticas del M-estimador con  $\rho_\lambda^*$

$\lambda$				
1			1.5	
$\beta_1$	ef. $\hat{\beta}_1$	ef. $\hat{\beta}_2$	ef. $\hat{\beta}_1$	ef. $\hat{\beta}_2$
1.314	0.866	0.951	0.846	0.943
1.502	0.836	0.940	0.813	0.930
1.795	0.793	0.924	0.767	0.913
2.300	0.733	0.902	0.704	0.889
4.394	0.613	0.860	0.581	0.844

## 5. CALCULO DE LOS VALORES ASINTÓTICOS BAJO CONTAMINACIONES

### 5.1. Sesgos asintóticos.

En esta sección se estudia la sensibilidad de los M-estimadores propuestos cuando la distribución de las observaciones presenta contaminación .

Sea  $P_0 = P_{g_0}$  la distribución verdadera de  $(X, Y)$  y sea  $P_\epsilon$  la distribución

$$P_\epsilon = (1 - \epsilon)P_0 + \epsilon\delta_{xy} \quad (5.1)$$

donde con  $\delta_{xy}$  se nota la distribución de la masa puntual en  $(x, y)$ .

Se estudiará la sensibilidad del M-estimador analizando cómo varían los valores asintóticos del estimador suponiendo que la distribución subyacente es  $P_\epsilon$ . El caso límite, cuando  $\epsilon$  tiende a 0, corresponde al análisis basado en la curva de influencia introducida por Hampel. Sin embargo, analizando el caso en que la contaminación de tamaño  $\epsilon$  es positiva da una evaluación más realista de la robustez del estimador.

Se considerará el caso en que  $q = 1$ ,  $X$  tiene distribución  $N(0, 1)$ ,  $H$  es la función logística,  $\beta_0 = 2.300$  y  $\beta_0 = 4.394$  (que corresponden a  $\alpha = 0.20$  y  $\alpha = 0.50$  respectivamente de la Tabla 1 de la sección anterior). Se consideraron los valores  $\beta_0$  usados en la Tabla 1, pero sólo presentamos estos 2 casos, ya que representan el comportamiento del estimador para los otros valores del parámetro verdadero.

El siguiente lema será útil para limitar el tipo de contaminaciones que se estudiarán.

**LEMA 5.1.** Si  $P_\epsilon^* = (1 - \epsilon)P_0 + \epsilon\delta_{x0}$  para  $x > 0$  y  $P_\epsilon' = (1 - \epsilon)P_0 + \epsilon\delta_{-x1}$ , entonces si  $H$  es la función logística

$$E_{P_\epsilon^*}(\phi(Y, \mathbf{X}, \beta)) = E_{P_\epsilon'}(\phi(Y, \mathbf{X}, \beta)).$$

**DEMOSTRACIÓN:**

Se tiene que

$$E_{P_\epsilon^*}(\phi(Y, \mathbf{X}, \beta)) = (1 - \epsilon)E_{P_0}(\phi(Y, \mathbf{X}, \beta)) + \epsilon[\rho(-\ln(1 - H(x\beta))) + G(H(x\beta)) + G(1 - H(x\beta))] \quad (5.2)$$

Dado que  $H$  es la función logística  $H(x\beta) = 1 - H((-x)\beta)$ , reemplazando en (5.2) se obtiene que

$$\begin{aligned} E_{P_\epsilon}(\phi(Y, \mathbf{X}, \beta)) &= \\ &= (1 - \epsilon)E_{P_0}(\phi(Y, \mathbf{X}, \beta)) + \epsilon[\rho(-\ln H((-x)\beta)) + G(H((-x)\beta)) + G(1 - H((-x)\beta))] \\ &= E_{P_\epsilon}(\phi(Y, \mathbf{X}, \beta)) \end{aligned}$$

como se quería demostrar.

Este lema permite restringir el estudio a contaminaciones puntuales con  $y = 0, x > 0$ .

Las tablas que siguen muestran el valor estimado del parámetro  $\beta_0$  obtenido al minimizar

$$E_{P_\epsilon}(\phi(Y, \mathbf{X}, \beta))$$

para distintos valores de  $\epsilon$  utilizando las funciones  $\rho$  que pertenecen a las familias (4.1) y (4.2) de la sección anterior.

Las Tablas 4, 5, 6 y 7 muestran los valores estimados obtenidos para  $\beta = 2.300$ .  $\epsilon = 0.10$  para el M-estimador propuesto cuando se emplean las familias de funciones  $\rho$  dadas por (4.1) y (4.2) con  $c = -\ln p^*$ ,  $p^* = 0.10, 0.05$  y  $\lambda = 1, 1.5$  respectivamente. Las tablas muestran que el estimador se comporta en forma similar en los cuatro casos considerados. El valor estimado desciende en general hasta el 50 por ciento del valor del parámetro  $\beta_0$  para contaminaciones cercanas al punto medio del intervalo  $[0,3]$  y luego se recupera progresivamente hasta que el valor asintótico coincide con  $\beta_0$  para contaminaciones de masa puntual uno en  $x > 3, y = 0$ . En la Tabla 4 se observa que el máximo sesgo es mayor que en los demás casos. De la lectura de las Tablas 4 y 6 puede comprobarse que el M-estimador basado en  $\rho_1^*$  y el basado en  $\rho_{-\ln 0.05}$  tienen un comportamiento similar. La ventaja del primero es que su máximo sesgo es menor que el del M-estimador con  $\rho_{-\ln 0.05}$  mientras que la ventaja de éste es que el sesgo tiende más rápidamente a cero a medida que  $x$  crece.

La misma observación se puede hacer sobre los estimadores basados en  $\rho_{- \ln 0.10}$  y  $\rho_{1.5}^*$ .

De los cuatro estimadores estudiados, el M-estimador con  $\rho_{- \ln 0.10}$  es el de menor máximo sesgo.

Las Tablas 8 y 9 dan el valor asintótico para  $\beta_0 = 4.394$  y  $\epsilon = 0.10$  y  $0.15$  para el M-estimador con  $\rho_{- \ln 0.10}$ . Para  $\beta_0 = 4.394$  el comportamiento del M-estimador conserva las características observadas en el caso en que  $\beta_0 = 2.300$ .

El comportamiento del M-estimador basado en  $\rho_{- \ln 0.05}$ ,  $\rho_1^*$  y  $\rho_{1.5}^*$  es semejante al que revelan las Tablas 8 y 9.

También se analizaron el resto de los valores del parámetro dados en la Tabla 1, obteniéndose resultados similares a los aquí presentados. -

Bajo las suposiciones realizadas, puede notarse en los casos estudiados, que cuanto mayor es el valor absoluto del parámetro, mayor será el porcentaje de contaminación que resista el M-estimador.

Tabla 4 Valores Asintóticos cuando  $\epsilon = 0.10$ ,  $\beta = 2.300$ ,  $\rho = \ln 0.05$

$x$ contaminante										
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Estimador	2.054	1.863	1.710	1.585	1.481	1.392	1.315	1.248	1.189	1.136

$x$ contaminante										
	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0
Estimador	1.089	1.046	1.007	0.971	0.938	0.908	0.880	0.855	2.094	2.130

$x$ contaminante										
	2.1	2.2	2.3	2.4	2.5	2.6	2.7	2.8	2.9	3.0
Estimador	2.160	2.185	2.206	2.223	2.237	2.249	2.258	2.266	2.272	2.278

Tabla 5 Valores Asintóticos cuando  $\epsilon = 0.10$ ,  $\beta = 2.300$ ,  $\rho = \ln 0.10$

	$x$ contaminante									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Estimador	1.997	1.784	1.626	1.503	1.404	1.322	1.255	1.197	1.148	1.108

	$x$ contaminante									
	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0
Estimador	1.073	1.046	1.027	1.020	1.726	1.780	1.839	1.903	1.969	2.028

	$x$ contaminante									
	2.1	2.2	2.3	2.4	2.5	2.6	2.7	2.8	2.9	3.0
Estimador	2.077	2.119	2.152	2.180	2.203	2.221	2.236	2.248	2.258	2.265

Tabla 6 Valores Asintóticos cuando  $\epsilon = 0.10$ ,  $\beta = 2.300$ ,  $\rho_1^*$

$x$ contaminante										
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Estimador	2.047	1.850	1.694	1.569	1.467	1.381	1.309	1.247	1.193	1.146

$x$ contaminante										
	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0
Estimador	1.104	1.068	1.037	1.009	0.986	0.967	0.952	0.942	0.938	0.946

$x$ contaminante										
	2.1	2.2	2.3	2.4	2.5	2.6	2.7	2.8	2.9	3.0
Estimador	0.973	1.073	1.690	1.899	2.009	2.081	2.131	2.169	2.197	2.218



Tabla 7 Valores Asintóticos cuando  $\epsilon = 0.10$ ,  $\beta = 2.300$ ,  $\rho_{1.5}^*$

		$x$ contaminante									
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Estimador		2.033	1.830	1.674	1.550	1.450	1.367	1.297	1.238	1.187	1.143

		$x$ contaminante									
		1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0
Estimador		1.106	1.074	1.048	1.026	1.010	1.000	1.000	1.007	1.035	1.113

		$x$ contaminante									
		2.1	2.2	2.3	2.4	2.5	2.6	2.7	2.8	2.9	3.0
Estimador		1.427	1.775	1.934	2.030	2.095	2.142	2.177	2.203	2.222	2.239

Tabla 8 Valores Asintóticos cuando  $\epsilon = 0.10$ ,  $\beta = 4.394$ ,  $\rho^*_{-\ln 0.010}$

$x$ contaminante										
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Estimador	3.192	2.663	2.345	2.132	1.964	1.867	1.795	1.761	3.021	3.297

$x$ contaminante										
	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0
Estimador	3.611	3.876	4.060	4.176	4.251	4.299	4.331	4.352	4.366	4.375

$x$ contaminante										
	2.1	2.2	2.3	2.4	2.5	2.6	2.7	2.8	2.9	3.0
Estimador	4.382	4.386	4.389	4.391	4.393	4.394	4.394	4.394	4.394	4.394

Tabla 9 Valores Asintóticos cuando  $\epsilon = 0.15$ ,  $\beta = 4.394$ .  $\rho_{-\ln 0.10}^*$

$x$ contaminante										
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Estimador	2.852	2.316	1.991	1.766	1.600	1.471	1.366	1.280	1.207	1.145

$x$ contaminante										
	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0
Estimador	2.667	3.217	3.704	3.987	4.141	4.232	4.325	4.348	4.364	4.374

$x$ contaminante										
	2.1	2.2	2.3	2.4	2.5	2.6	2.7	2.8	2.9	3.0
Estimador	4.374	4.381	4.385	4.388	4.390	4.392	4.393	4.393	4.394	4.394

## 5.2 Comparación de sesgos asintóticos.

En esta sección se calculan los valores asintóticos para los M-estimadores propuestos y se comparan con los estimadores condicionalmente insesgados de influencia acotada (CIIA-estimador) obtenidos por Künsch, Stefanski y Carroll (1987) cuando la distribución subyacente es  $P_c$  dada por (5.1).

Bajo el modelo (1.1) Künsch et al. (1987) propusieron estimar  $\beta_0$  mediante la resolución del siguiente sistema

$$E(\psi_{cond}(Y, \mathbf{X}, \underline{\beta})|X) = 0$$

$$E(\psi_{cond}(Y, \mathbf{X}, \underline{\beta})\psi_{cond}(Y, \mathbf{X}, \underline{\beta})') = B(\underline{\beta})$$

donde  $\psi_{cond}(y, \mathbf{x}, \underline{\beta}) = (l - c\mathbf{x})\min^{1/2}(1, b^2(l - c)'B^{-1}(l - c))$ ,  $b$  es una constante a elegir convenientemente,  $l = l(y, \mathbf{x}, \underline{\beta}) = (y - H(\mathbf{x}'\underline{\beta}))\mathbf{x}$ ,  $H$  es la función logística y

$$c = c(\theta, a) = \begin{cases} aH(\theta)/(1 - H(\theta)) - H(\theta) & \text{si } \theta < 0 \quad a < 1 - H(\theta) \\ 1 - H(\theta) - a(1 - H(\theta))/H(\theta) & \text{si } \theta > 0 \quad a < H(\theta) \\ 0 & \text{caso contrario} \end{cases}$$

Este estimador tiene la propiedad de minimizar la  $tr(V(\psi)V^{-1}(\psi_{cond}))$  entre todos los estimadores definidos por

$$E(\psi(Y, \mathbf{X}, \underline{\beta})) = 0$$

con  $\psi$  que satisfacen

$$E(\psi(Y, \mathbf{X}, \underline{\beta}_0)|\mathbf{X}) = 0$$

$$\sup_{(y, \mathbf{x})} IC'_{\psi} V^{-1}(\psi_{cond}) IC'_{\psi} \leq b^2$$

donde con  $V(\psi)$  y  $IC'_{\psi}$  se nota la matriz de covarianza asintótica y la curva de influencia respectivamente del estimador.

El cálculo de los valores asintóticos fue realizado suponiendo las mismas hipótesis que en la parte I.

En la Tabla 10, se dan los valores asintóticos para distintos estimadores suponiendo que la distribución subyacente es  $P_\epsilon$  para  $\epsilon = 0.10$  y  $\beta_0 = 2.300$ . Los estimadores considerados son el M-estimador con  $\rho$  en la familia  $\rho_c$ ,  $c = -\ln 0.10$  y en la familia  $\rho_\lambda^*$  con  $\lambda = 1$  y  $1.5$  y el CIAA-estimador con  $b = 0.075$ . Este valor de  $b$  da la misma varianza asintótica que el M-estimador con  $\rho_{-\ln 0.10}$ .

En la Tabla 11 se dan los mismos resultados para  $\beta_0 = 4.394$  reemplazando  $\rho_{-\ln 0.10}$  por  $\rho_{-\ln 0.05}$  en el M-estimador y  $b = 0.037$  para el CIAA-estimador, ya que para estos valores las eficiencias asintóticas se asemejan.

Las Tablas 10 y 11 muestran que el máximo sesgo del CIAA-estimador es más pequeño que el correspondiente al M-estimador. Sin embargo, una ventaja del M-estimador es que cuando el valor contaminante  $x$  es grande su sesgo es muy pequeño o nulo mientras que el CIAA-estimador se estabiliza en su mínimo valor (que corresponde al máximo sesgo) para  $x$  grande.

Tabla 10. Valores asintóticos cuando  $\epsilon = 0.10$  y  $\beta_{01} = 2.30$

Estimador	$x$ contaminante								Var. Asin.
	0.1	0.5	1	1.5	2	2.5	3	5	
$\rho_{-\ln 0.10}$	1.99	1.40	1.11	1.73	2.03	2.20	2.27	2.30	38.4
$\rho_1^*$	2.05	1.47	1.15	0.99	0.95	2.01	2.22	2.30	28.8
$\rho_{1.5}^*$	2.03	1.45	1.14	0.81	0.82	2.10	2.24	2.30	30.0
$CH_{10,075}$	1.54	1.34	1.34	1.34	1.34	1.34	1.34	1.34	38.4

Tabla 11. Valores asintóticas cuando  $\epsilon = 0.10$  y  $\beta_{01} = 4.39$

Estimador	$x$ contaminante							Var. Asin.
	0.5	1	1.5	2	2.5	3	5	
$\rho_{-\ln 0.05}$	2.10	1.67	4.30	4.38	4.39	4.39	4.39	169.9
$\rho_1^*$	2.12	1.78	4.03	4.35	4.39	4.39	4.39	142.2
$\rho_{1.5}^*$	2.09	1.84	4.13	4.36	4.39	4.39	4.39	150.0
$CH_{10,037}$	2.27	2.27	2.27	2.27	2.27	2.27	2.27	169.9

## 6. EJEMPLOS

En esta sección ilustramos el comportamiento del M-estimador propuesto usando el modelo logístico para dos conjuntos de datos reales.

En ambos ejemplos se computó un estimador inicial por un procedimiento "stepwise" (paso a paso), que se describe en la parte I y el estimador final se obtuvo utilizando el método Downhill Simplex propuesto por Nelder y Mead (1965).

### I. Estimador Inicial.

El estimador inicial se computa por un procedimiento paso a paso que ajusta un modelo de una covariable cada vez. La minimización se lleva a cabo utilizando una grilla inicial y el método Golden Section Search. Se selecciona el modelo que minimiza (1.7), luego se fija el coeficiente correspondiente a la primera covariable seleccionada y se ajusta un nuevo modelo de dos variables como en el caso de un sola variable. Se continua de esta manera hasta ingresar todas las covariables.

El procedimiento realiza los siguientes pasos.

Supongamos que  $X$ , la matriz de diseño, es tal que  $X = (\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_{q-1}) \in \mathfrak{R}^{n \times q}$ , donde  $\mathbf{x}_0$  es un vector columna de 1's.

Paso 1.

Las columnas  $\mathbf{x}_i$ ,  $1 \leq q - 1$  son estandarizadas de manera robusta (por ejemplo, tomando la mediana y la desviación mediana absoluta). Para simplificar la notación llamaremos  $\mathbf{x}_i$  a las columnas transformadas.

Paso 2.

Se consideran los  $q$  modelos

$$P(Y = 1 | X_i = \mathbf{x}_i) = H(\mathbf{x}_i \beta_i), \quad 0 \leq i \leq q - 1,$$

se estima  $\beta_i$  por el M-estimador y se selecciona la variable que minimiza (1.7). Sin pérdida de generalidad podemos suponer que 1 es el índice de la covariable elegida.

### Paso 3

Antes de entrar una nueva variable se ortogonalizan robustamente las  $(q-1)$  columnas restantes  $\mathbf{x}_i$  con respecto a la columna  $\mathbf{x}_1$ , seleccionada en el paso anterior. Nuevamente, por razones de simplicidad, llamamos  $\mathbf{x}_i$  a las columnas transformadas.

### Paso 4

Se considera los siguientes  $q-1$  modelos

$$P(Y = 1 | (X_1 = x_1, X_i = x_i)) = H(x_1 \hat{\beta}_1 + x_i \beta_i)$$

donde con  $\hat{\beta}_1$  notamos el estimador de la componente  $\beta_1$  estimada en el Paso 2. Se estima  $\beta_i$  por el M-estimador propuesto y se selecciona la variable  $i$  que minimiza (1.7) para este modelo.

De aquí en más, las variables restantes entran una a una, una vez que han sido ortogonalizadas robustamente con respecto a la última variable que ingresó al modelo y se elige la nueva variable entrante fijando los coeficientes de las covariables ya seleccionadas y minimizando (1.7) para el modelo correspondiente como en el Paso 4.

En los ejemplos que siguen este procedimiento fue aplicado iterativamente y se comprobó que este estimador inicial converge al estimador final.



## II. Ejemplo: datos de Cook y Weisberg.

Estos datos fueron considerados por Cook y Weisberg (1982, Cap 5, pag.193) y Johnson (1985) con el fin de ejemplificar la identificación de observaciones influyentes.

Los datos, presentados en la Tabla 12, corresponden a 33 pacientes que padecen de leucemia. La variable de respuesta vale 1 cuando el paciente sobrevive más de 52 semanas y las dos covariables observadas son

WBC = número de glóbulos blancos

AG = presencia o ausencia de cierta característica morfológica en los glóbulos

Cook y Weisberg detectaron que la observación correspondiente a un paciente con  $WBC=100.000$  que sobrevive más de un año, es muy influyente cuando se estima el parámetro por el método de máxima verosimilitud. Ellos notaron que al excluir esta observación del conjunto de datos, el estimador de máxima verosimilitud basado en el resto de la muestra permite un mejoramiento del ajuste y que las probabilidades de sobrevivida estimadas de aquellos pacientes con valores de WBC pequeños aumentan significativamente.

En la Tabla 13 se dan los coeficientes estimados correspondientes a

- i) Estimador de máxima verosimilitud basado en la muestra completa ( $EMV_{33}$ ).
- ii) Estimador de máxima verosimilitud basado en la muestra sin la observación influyente ( $EMV_{32}$ ).
- iii) M-estimador con  $\rho = \ln 0.10$  usando la muestra completa y las probabilidades ajustadas correspondientes.

La Tabla 13 muestra que los coeficientes obtenidos aplicando el M-estimador a la muestra completa son muy parecidos a los que se obtienen por máxima verosimilitud después de excluir la observación influyente. Las probabilidades estimadas a partir de estos dos estimadores son muy parecidas y mejoran notablemente el ajuste basado en  $EMV_{33}$ .

Tabla 12. Datos de Cook y Weisberg

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

Tabla 13. Datos de Cook y Weisberg: coeficientes y probabilidades estimados.

	Estimador		
	$EMV_{33}$	$EMV_{32}$	$M_{0.10}$
constante	-1.30	0.21	0.10
WBC	$-3.18 \times 10^{-5}$	$-2.35 \times 10^{-4}$	$-1.83 \times 10^{-4}$
AG	2.26	2.56	2.36
obs.y	Prob.Est.	.Prob.Est.	Prob.Est.
1	0.707	0.903	0.885
1	0.717	0.930	0.911
1	0.694	0.853	0.841
1	0.705	0.897	0.879
0	0.682	0.796	0.795
1	0.650	0.575	0.630
1	0.654	0.603	0.651
0	0.602	0.227	0.341
0	0.686	0.818	0.813
1	0.675	0.755	0.764
1	0.658	0.637	0.676
0	0.484	0.009	0.032
0	0.460	0.004	0.019
0	0.332	0.000	0.001
1	0.097	0.000	0.000
0	0.097	0.000	0.000
0	0.097	0.000	0.000
1	0.190	0.305	0.331
1	0.197	0.379	0.390
0	0.192	0.326	0.347
0	0.205	0.465	0.457
0	0.169	0.130	0.175
0	0.186	0.262	0.295
0	0.165	0.105	0.150
0	0.129	0.014	0.033
0	0.103	0.002	0.008
0	0.100	0.002	0.006
0	0.092	0.001	0.004
0	0.106	0.003	0.009
0	0.122	0.009	0.023
0	0.021	0.000	0.000
0	0.011	0.000	0.000
0	0.011	0.000	0.000

### 6.3 Ejemplo: Datos de Finney.

Estos datos fueron originalmente presentados por Finney y analizados por Pregibon (1981) y Künsch, Stefanski y Carroll (1987) en el contexto de regresión logística binomial.

En este ejemplo se estudia el efecto de la velocidad y volumen del aire inspirado sobre una vasoconstricción pasajera en la piel de los dedos. La variable de respuesta vale 1 cuando se produce una vasoconstricción. Las observaciones se hallan en la Tabla 14.

El modelo considerado es

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

Pregibon observó que si se estima el parámetro por el método de máxima verosimilitud las observaciones 4 y 18 no son bien ajustadas y diagnosticó que ambas son altamente influyentes en la determinación de los coeficientes. También notó que si estas observaciones se extraen de la muestra es posible una perfecta discriminación de los datos restantes.

En la Tabla 15 se dan los coeficientes y probabilidades estimadas por el estimador de máxima verosimilitud (EMV), el M-estimador con  $\rho_{-\ln 0.10}(M_{0.10})$  y el CHA-estimador obtenido por Künsch et al. (1987) con  $b = 5.5$ .

La Tabla 15 muestra que M-estimador da estimaciones de los coeficientes mucho mayores que los provistos por los otros métodos, lo que es consecuencia de la perfecta discriminación que es posible realizar si las observaciones 4 y 18 son eliminadas de la muestra. Esto muestra que el M-estimador no es influenciado por la presencia de estos datos. Por otra parte, también se observa en la Tabla que el ajuste provisto por el M-estimador que proponemos es perfecto salvo en las dos observaciones influyentes y en la 24, que no es bien ajustada por ninguno de los tres procedimientos.

Tabla 14. Datos de Finney

obs.	Volumen	Velocidad	Respuesta
1	3.7	0.825	1
2	3.5	1.09	1
3	1.25	2.5	1
4	0.75	1.5	1
5	0.8	3.2	1
6	0.7	3.5	1
7	0.6	0.75	0
8	1.1	1.7	0
9	0.9	0.75	0
10	0.9	0.45	0
11	0.8	0.57	0
12	0.55	2.75	0
13	0.6	3.0	0
14	1.4	2.33	1
15	0.75	3.75	1
16	2.3	1.64	1
17	3.2	1.6	1
18	0.85	1.415	1
19	1.7	1.06	0
20	1.8	1.8	1
21	0.4	2.0	0
22	0.95	1.36	0
23	1.35	1.35	0
24	1.5	1.36	0
25	1.6	1.78	1
26	0.6	1.5	0
27	1.8	1.5	1
28	0.95	1.9	0
29	1.9	0.95	1
30	1.6	0.4	0
31	2.7	0.75	1
32	2.35	0.03	0
33	1.1	1.83	0
34	1.1	2.2	1
35	1.2	2.0	1
36	0.8	3.33	1
37	0.95	1.9	0
38	0.75	1.9	0
39	11.3	1.625	1

Tabla 15. Datos de Finney: coeficientes y probabilidades estimados.

	Estimador		
	<i>EMV</i>	<i>M</i> <sub>0.10</sub>	<i>CIA</i>
$\beta_0$	-2.92	-655	-6.41
$\beta_1$	5.22	1113	9.98
$\beta_2$	4.63	859	8.85
obs.y	Prob.Est.	Prob.Est.	Prob.Est.
1	0.953	1	0.993
1	0.982	1	0.999
1	0.923	1	0.981
4   1	0.073	0	0.003
1	0.786	1	0.840
1	0.735	1	0.753
0	0.001	0	0
8   0	0.509	0	0.318
0	0.008	0	0
0	0.001	0	0
0	0.001	0	0
0	0.205	0	0.032
0	0.378	0	0.114
1	0.940	1	0.988
1	0.845	1	0.918
1	0.976	1	0.998
1	0.995	1	1
1   1	0.103	0	0.007
0	0.530	0	0.355
1	0.946	1	0.991
0	0.011	0	0
0	0.146	0	0.015
0	0.509	0	0.319
24 → 0	0.650	1	0.589
1	0.901	1	0.967
0	0.024	0	0
1	0.883	1	0.955
0	0.446	0	0.224
1	0.584	1	0.387
0	0.009	0	0
1	0.718	1	0.722
0	0.017	0	0
0	0.593	0	0.472
1	0.773	1	0.820
1	0.776	1	0.824
1	0.815	1	0.882
0	0.446	0	0.224
0	0.190	0	0.027
1	0.668	1	0.624

## 7. EXTENSIÓN AL MODELO DE REGRESIÓN MULTINOMIAL

El M-estimador propuesto para el modelo de regresión binomial puede extenderse en forma natural al modelo de regresión multinomial.

Supóngase que se observa un vector de respuestas  $\mathbf{Y} = (Y_1, \dots, Y_k)'$  con distribución multinomial, donde  $Y_i = 0$  ó  $1$  y,  $\sum_{i=1}^k Y_i = 1$ , que está relacionado con el vector de covariables  $\mathbf{X} = (X_1, \dots, X_q)'$ . De manera tal que dado  $\mathbf{x} \in \mathbb{R}^q$

$$\begin{aligned} P(Y_j = 1 | \mathbf{X} = \mathbf{x}) &= \frac{H(\mathbf{x}' \underline{\beta}_0^j)}{1 + \sum_{i=1}^{k-1} H(\mathbf{x}' \underline{\beta}_0^i)} = H_j^*(\mathbf{x}, \underline{\mathbf{B}}_0) \text{ si } j = 1, \dots, k-1 \\ P(Y_k = 1 | \mathbf{X} = \mathbf{x}) &= \frac{1}{1 + \sum_{i=1}^{k-1} H(\mathbf{x}' \underline{\beta}_0^i)} = H_k^*(\mathbf{x}, \underline{\mathbf{B}}_0) \text{ si } j = k. \end{aligned} \quad (7.1)$$

donde  $\underline{\beta}_0^j \in \mathbb{R}^q, j = 1, \dots, k-1$ , son los  $(k-1)$  parámetros a estimar y donde  $\underline{\mathbf{B}}_0 = (\underline{\beta}_0^1, \dots, \underline{\beta}_0^{k-1})' \in \mathbb{R}^{q(k-1)}$ .

De aquí en más se supondrá que  $H$  verifica las siguientes hipótesis

D1.  $H(t)$  es una función continua

D2.  $H(t)$  es estrictamente creciente

D3.  $\lim_{t \rightarrow \infty} H(t) = \infty$

D4.  $\lim_{t \rightarrow -\infty} H(t) = 0$

En particular, cuando  $H(t) = \exp(t)$  resulta el modelo logístico multinomial y se tiene que

$$\begin{aligned} P(Y_j = 1 | \mathbf{X} = \mathbf{x}) &= \frac{\exp(\mathbf{x}' \underline{\beta}_0^j)}{1 + \sum_{i=1}^{k-1} \exp(\mathbf{x}' \underline{\beta}_0^i)} \text{ si } j = 1, \dots, k-1 \\ P(Y_k = 1 | \mathbf{X} = \mathbf{x}) &= \frac{1}{1 + \sum_{i=1}^{k-1} \exp(\mathbf{x}' \underline{\beta}_0^i)} \text{ si } j = k. \end{aligned} \quad (7.1^*)$$

Sea  $\mathbf{Z}_1 = (\mathbf{Y}'_1, \mathbf{X}'_1)', \dots, \mathbf{Z}_n = (\mathbf{Y}'_n, \mathbf{X}'_n)'$  una muestra de  $n$  observaciones independientes, donde  $\mathbf{X}_i = (X_{i1}, \dots, X_{iq})'$  y  $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{ik})'$ ,  $i = 1, \dots, n$ , satisfacen el modelo (7.1).

Los estimadores de máxima verosimilitud de  $\underline{\beta}_0^1, \dots, \underline{\beta}_0^{k-1}$  se obtienen de minimizar

$$\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^k Y_{ij} (-\ln(H_j^*(\mathbf{X}_i, \underline{\mathbf{B}}))).$$

Los M-estimadores de  $\underline{\beta}_0^1, \dots, \underline{\beta}_0^{k-1}$  pueden definirse análogamente al caso binomial minimizando

$$\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^k Y_{ij} \rho_j(-\ln(H_j^*(\mathbf{X}, \underline{\mathbf{B}}))) + G_j(H_j^*(\mathbf{X}, \underline{\mathbf{B}})), \quad (7.2)$$

donde las funciones  $G_j$  están definidas en (1.8).

En (7.2) se ha considerado una función  $\rho_j$  para cada clase  $j$ . En particular, se podrían tomar todas las funciones  $\rho_j$  iguales entre sí, pero si hubiera una clase mucho más pequeña que los demás o alguna evidencia de que en una clase hay algún dato atípico podría ser conveniente considerar distintas funciones  $\rho$ .

En las secciones que siguen se probará la consistencia, se calculará la matriz de covarianza asintótica de los M-estimadores definidos por la minimización de (7.2) y se estudiará el comportamiento del M-estimador a través de ejemplos reales y simulados.



## 7.1 Consistencia en el sentido de Fisher.

En esta sección se probará la consistencia en el sentido de Fisher del M-estimador definido por la minimización de (7.2).

Sea  $\rho$  una función que satisface las siguientes condiciones

D1.  $\rho$  es una función acotada con derivada acotada y continua  $\psi$

D2.  $\rho(0) = 0$

D3.  $\rho$  es una función creciente y  $\psi$  satisface alguna de las siguientes condiciones

D3.1  $\psi(t) > 0$  para todo  $t \geq 0$

D3.2  $\psi(t) > 0$  si  $0 \leq t < c$  y  $\psi(t) = 0$  si  $t \geq c$ , para  $c = -\ln p^*$

Se necesitará el siguiente lema para probar el resultado.

**LEMA 7.1.** *Sea  $\rho$  una función que satisface las hipótesis D1 a D3 y sea  $Y$  una variable aleatoria de Bernoulli de parámetro  $p_0$ . Supóngase que  $\psi = \rho'$  cumple D3.1 ó D3.2 con  $p^* < p_0$ . Luego, si se define*

$$M_\rho(p, p_0) = \rho(-\ln p)E_{p_0}(Y) + G(p) \quad p \in (0, 1), \quad (7.3)$$

entonces  $M_\rho(p, p_0)$  alcanza su único mínimo en  $p = p_0$ . Si en cambio  $\psi$  cumple D3.2 con  $p^* \geq p_0$ ,  $M_\rho(p, p_0)$  alcanza su mínimo en todo el intervalo  $(0, p^*]$ .

**DEMOSTRACIÓN:**

Dado que

$$M_\rho(p, p_0) = p_0\rho(-\ln p) + G(p) \quad (7.4)$$

se tiene que

$$\frac{\partial M_\rho(p, p_0)}{\partial p} = \psi(-\ln p)\left(1 - \frac{p_0}{p}\right) \quad (7.5)$$

De (7.5) resulta que  $\partial M_\rho(p, p_0)/\partial p$  se anula si y sólo si  $\psi(-\ln p) = 0$  ó  $p = p_0$ .

Si  $\psi$  cumple D3.1,  $\partial M_\rho(p, p_0)/\partial p$  se anula sólo si  $p = p_0$  y como

$$\frac{\partial M_\rho(p, p_0)}{\partial p} < 0 \text{ si } p < p_0 \quad (7.6)$$

$$\frac{\partial M_\rho(p, p_0)}{\partial p} > 0 \text{ si } p > p_0 \quad (7.7)$$

resulta que  $M_\rho(p, p_0)$  alcanza su mínimo en  $p = p_0$ .

Si  $\psi$  cumple D3.2 con  $c = -\ln p^* > -\ln p_0$  se obtiene que

$$\frac{\partial M_\rho(p, p_0)}{\partial p} < 0 \text{ si } p^* < p < p_0$$

$$\frac{\partial M_\rho(p, p_0)}{\partial p} > 0 \text{ si } p_0 < p < 1,$$

luego por la continuidad de  $M_\rho$  se cumple que  $M_\rho(p_0, p_0) < M_\rho(p, p_0)$  si  $p \in [p^*, 1)$ . Como  $M_\rho(p, p_0)$  es constante en  $(0, p^*]$ , resulta que la función tiene un único mínimo en  $p = p_0$ .

Finalmente, si  $\psi$  satisface D3.2 con  $c = -\ln p^* \leq -\ln p_0$  se tiene que

$$\frac{\partial M_\rho(p, p_0)}{\partial p} > 0 \text{ si } p^* < p < 1 \quad (7.8)$$

$$M_\rho(p, p_0) = M_\rho(p^*, p_0) \text{ si } 0 < p \leq p^* \quad (7.9)$$

De (7.8) y (7.9) se obtiene que  $M_\rho(p, p_0)$  alcanza su mínimo en todo el intervalo  $(0, p^*]$ .

Con el fin de simplificar la notación, se notará

$$\phi(\mathbf{Y}, \mathbf{X}, \underline{\mathbf{B}}) = \sum_{j=1}^k Y_j \rho_j(-\ln(H_j^*(\mathbf{X}, \underline{\mathbf{B}}))) + G_j(H_j^*(\mathbf{X}, \underline{\mathbf{B}})).$$

**PROPOSICIÓN 7.1.** Sean  $\rho_j, j = 1, \dots, k$  funciones satisfaciendo las condiciones D1, D2, D3.1 y supóngase que

$$P(\mathbf{X}'\beta = 0) < 1 \quad \beta \in \mathfrak{R}^q.$$

Luego, para todo  $\underline{\mathbf{B}} \neq \underline{\mathbf{B}}_0$ , se tiene que

$$E_{\underline{\mathbf{B}}_0}(\phi(\mathbf{Y}, \mathbf{X}, \underline{\mathbf{B}}_0)) < E_{\underline{\mathbf{B}}_0}(\phi(\mathbf{Y}, \mathbf{X}, \underline{\mathbf{B}})).$$

DEMOSTRACIÓN:

Dado  $\underline{\mathbf{B}} \in \mathfrak{X}^{q(k-1)}$  se define el conjunto

$$A(\underline{\mathbf{B}}) = \cap_{i=1}^k \{\mathbf{x} : H_i^*(\mathbf{x}, \underline{\mathbf{B}}) = H_i^*(\mathbf{x}, \underline{\mathbf{B}}_0)\}.$$

Dado que

$$H_i^*(\mathbf{x}, \underline{\mathbf{B}}) = H(\mathbf{x}'\beta^i)H_k^*(\mathbf{x}, \underline{\mathbf{B}}), \quad \text{si } i \neq k$$

$$H_k^*(\mathbf{x}, \underline{\mathbf{B}}) = 1/(1 + \sum_{j=1}^{k-1} H(\mathbf{x}'\beta^j)),$$

$\mathbf{x} \in A(\underline{\mathbf{B}})$  si y sólo si

$$H(\mathbf{x}'\beta^i) = H(\mathbf{x}'\beta_0^i), \quad i = 1, \dots, k-1,$$

que como  $H$  es estrictamente creciente es equivalente a

$$\mathbf{x}'(\beta^i - \beta_0^i) = 0, \quad i = 1, \dots, k-1.$$

Luego, como  $A(\underline{\mathbf{B}}) = \cap_{i=1}^{k-1} \{\mathbf{x} : \mathbf{x}'(\beta^i - \beta_0^i) = 0\}$ , se tiene que  $P(A(\underline{\mathbf{B}})) < 1$ .

Por otro lado, tomando esperanza condicional con respecto a  $\mathbf{X}$ , se tiene que

$$E_{\underline{\mathbf{B}}_0}(\phi(\mathbf{Y}, \mathbf{X}, \underline{\mathbf{B}}) | \mathbf{X} = \mathbf{x}) = E\left(\sum_{j=1}^k Y_j \rho_j(-\ln(H_j^*(\mathbf{x}, \underline{\mathbf{B}}))) + G_j(H_j^*(\mathbf{x}, \underline{\mathbf{B}}))\right)$$

$$= \sum_{j=1}^k M_{\rho_j}(H_j^*(\mathbf{x}, \underline{\mathbf{B}}), H_j^*(\mathbf{x}, \underline{\mathbf{B}}_0))$$

y en consecuencia

$$E_{\underline{\mathbf{B}}_0}(\phi(\mathbf{Y}, \mathbf{X}, \underline{\mathbf{B}})) = E\left(\sum_{j=1}^k M_{\rho_j}(H_j^*(\mathbf{X}, \underline{\mathbf{B}}), H_j^*(\mathbf{X}, \underline{\mathbf{B}}_0))\right) \quad (7.10)$$

donde las funciones  $M_{\rho_j}$  fueron definidas en (7.4). De (7.10) se tiene que

$$E_{\underline{\mathbf{B}}_0}(\phi(\mathbf{Y}, \mathbf{X}, \underline{\mathbf{B}}_0)) = \quad (7.11)$$

$$= \int_{A(\underline{\mathbf{B}})} \sum_{j=1}^k M_{\rho_j}(H_j^*(\mathbf{X}, \underline{\mathbf{B}}_0), H_j^*(\mathbf{X}, \underline{\mathbf{B}}_0)) dP_{\mathbf{X}} + \int_{A^c(\underline{\mathbf{B}})} \sum_{j=1}^k M_{\rho_j}(H_j^*(\mathbf{X}, \underline{\mathbf{B}}), H_j^*(\mathbf{X}, \underline{\mathbf{B}}_0)) dP_{\mathbf{X}}$$

$$= \int_{A(\underline{\mathbf{B}})} \sum_{j=1}^k M_{\rho_j}(H_j^*(\mathbf{X}, \underline{\mathbf{B}}), H_j^*(\mathbf{X}, \underline{\mathbf{B}}_0)) dP_{\mathbf{X}} + \int_{A^c(\underline{\mathbf{B}})} \sum_{j=1}^k M_{\rho_j}(H_j^*(\mathbf{X}, \underline{\mathbf{B}}_0), H_j^*(\mathbf{X}, \underline{\mathbf{B}}_0)) dP_{\mathbf{X}}$$

Como  $P(A^c(\underline{\mathbf{B}})) > 0$ , por el Lema 7.1 y por (7.11) se tiene que

$$\begin{aligned} E_{\underline{\mathbf{B}}_0}(\phi(\mathbf{Y}, \mathbf{X}, \underline{\mathbf{B}}_0)) &= \\ &< \int_{A(\underline{\mathbf{B}})} \sum_{j=1}^k M_{\rho_j}(H_j^*(\mathbf{X}, \underline{\mathbf{B}}), H_j^*(\mathbf{X}, \underline{\mathbf{B}}_0)) dP_{\mathbf{X}} + \int_{A^c(\underline{\mathbf{B}})} \sum_{j=1}^k M_{\rho_j}(H_j^*(\mathbf{X}, \underline{\mathbf{B}}), H_j^*(\mathbf{X}, \underline{\mathbf{B}}_0)) dP_{\mathbf{X}} \\ &= E_{\underline{\mathbf{B}}_0}(\phi(\mathbf{Y}, \mathbf{X}, \underline{\mathbf{B}})) \end{aligned}$$

como se quería probar.

Dado  $\underline{\mathbf{B}} \in \mathfrak{R}^{q(k-1)}$  se definen los conjuntos

$$C_i(\underline{\mathbf{B}}) = \{\mathbf{x} : H_i^*(\mathbf{x}, \underline{\mathbf{B}}) \neq H_i^*(\mathbf{x}, \underline{\mathbf{B}}_0)\} \cap \{[\mathbf{x} : H_i^*(\mathbf{x}, \underline{\mathbf{B}}) > p_i^*] \cup [\mathbf{x} : H_i^*(\mathbf{x}, \underline{\mathbf{B}}_0) > p_i^*]\}$$

donde  $p_i^*$  está dado por D3.2 para cada  $\rho_i$ .

**PROPOSICIÓN 7.2.** Sean que  $\rho_j, j = 1, \dots, k$  funciones que cumplen las condiciones D1, D2, D3.2 y supóngase que dado  $\underline{\mathbf{B}} \in \mathfrak{R}^{q(k-1)}$  existe  $i$  tal que

$$P(C_i(\underline{\mathbf{B}})) > 0.$$

Entonces, para todo  $\underline{\mathbf{B}} \neq \underline{\mathbf{B}}_0$ , se tiene que

$$E_{\underline{\mathbf{B}}_0}(\phi(\mathbf{Y}, \mathbf{X}, \underline{\mathbf{B}}_0)) < E_{\underline{\mathbf{B}}_0}(\phi(\mathbf{Y}, \mathbf{X}, \underline{\mathbf{B}})).$$

**DEMOSTRACIÓN:**

Dado  $\underline{\mathbf{B}} \in \mathfrak{R}^{q(k-1)}$ , considérese el conjunto  $C_i(\underline{\mathbf{B}})$  que por hipótesis tiene probabilidad positiva. Luego, por el Lema 7.1 se tiene que

$$\begin{aligned} E_{\underline{\mathbf{B}}_0}(M_{\rho_i}(H_i^*(\mathbf{X}, \underline{\mathbf{B}}_0), H_i^*(\mathbf{X}, \underline{\mathbf{B}}_0))) &= \tag{7.12} \\ &= \int_{C_i(\underline{\mathbf{B}})} M_{\rho_i}(H_i^*(\mathbf{X}, \underline{\mathbf{B}}_0), H_i^*(\mathbf{X}, \underline{\mathbf{B}}_0)) dP_{\mathbf{X}} + \int_{C_i^c(\underline{\mathbf{B}})} M_{\rho_i}(H_i^*(\mathbf{X}, \underline{\mathbf{B}}_0), H_i^*(\mathbf{X}, \underline{\mathbf{B}}_0)) dP_{\mathbf{X}} \\ &= \int_{C_i(\underline{\mathbf{B}})} M_{\rho_i}(H_i^*(\mathbf{X}, \underline{\mathbf{B}}_0), H_i^*(\mathbf{X}, \underline{\mathbf{B}}_0)) dP_{\mathbf{X}} + \int_{C_i^c(\underline{\mathbf{B}})} M_{\rho_i}(H_i^*(\mathbf{X}, \underline{\mathbf{B}}), H_i^*(\mathbf{X}, \underline{\mathbf{B}}_0)) dP_{\mathbf{X}}. \end{aligned}$$

Dado que  $P(C_i^c(\underline{\mathbf{B}})) > 0$ , por el Lema 7.1 y por (7.12) se tiene que

$$\begin{aligned}
 & E(M_{\rho_i}(H_i^*(\mathbf{X}, \underline{\mathbf{B}}_0), H_i^*(\mathbf{X}, \underline{\mathbf{B}}_0))) \\
 & < \int_{C_i(\underline{\mathbf{B}})} M_{\rho_i}(H_i^*(\mathbf{X}, \underline{\mathbf{B}}), H_i^*(\mathbf{X}, \underline{\mathbf{B}}_0)) dP_{\mathbf{X}} + \int_{C_i^c(\underline{\mathbf{B}})} M_{\rho_i}(H_i^*(\mathbf{X}, \underline{\mathbf{B}}), H_i^*(\mathbf{X}, \underline{\mathbf{B}}_0)) dP_{\mathbf{X}} \\
 & = E_{\underline{\mathbf{B}}_0}(M_{\rho_i}(H_i^*(\mathbf{X}, \underline{\mathbf{B}}), H_i^*(\mathbf{X}, \underline{\mathbf{B}}_0))). \tag{7.13}
 \end{aligned}$$

Finalmente, como por el Lema 7.1

$$E(M_{\rho_j}(H_j^*(\mathbf{X}, \underline{\mathbf{B}}_0), H_j^*(\mathbf{X}, \underline{\mathbf{B}}_0))) \leq E(M_{\rho_j}(H_j^*(\mathbf{X}, \underline{\mathbf{B}}), H_j^*(\mathbf{X}, \underline{\mathbf{B}}_0))) \quad j \neq i,$$

de (7.10) y (7.13) se obtiene que

$$\begin{aligned}
 E_{\underline{\mathbf{B}}_0}(\phi(\mathbf{Y}, \mathbf{X}, \underline{\mathbf{B}}_0)) &= \sum_{j=1}^k E(M_{\rho_j}(H_j^*(\mathbf{X}, \underline{\mathbf{B}}_0), H_j^*(\mathbf{X}, \underline{\mathbf{B}}_0))) \\
 &< \sum_{j=1}^k E(M_{\rho_j}(H_j^*(\mathbf{X}, \underline{\mathbf{B}}), H_j^*(\mathbf{X}, \underline{\mathbf{B}}_0))) = E_{\underline{\mathbf{B}}_0}(\phi(\mathbf{Y}, \mathbf{X}, \underline{\mathbf{B}}))
 \end{aligned}$$

como se quería demostrar.

A partir de la consistencia en el sentido de Fisher del estimador, se podría demostrar la consistencia del M-estimador siguiendo las técnicas empleadas en la Sección 3.

## 7.2 Matriz de covarianza asintótica.

En esta sección se calcula la matriz de covarianza asintótica de los M-estimadores definidos a partir de la minimización de (7.2).

Sea  $\mathbf{Z}_1 = (\mathbf{Y}'_1, \mathbf{X}'_1)', \dots, \mathbf{Z}_n = (\mathbf{Y}'_n, \mathbf{X}'_n)'$  una muestra de vectores en  $\mathbb{R}^{q+k}$  independientes, donde  $\mathbf{X}_i = (X_{i1}, \dots, X_{iq})'$  y  $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{ik})'$ ,  $i = 1, \dots, n$ , satisfacen el modelo (7.1). Sean  $\hat{\mathbf{B}}_n$  la sucesión de estimadores basados en  $\mathbf{Z}_1, \dots, \mathbf{Z}_n$  que se obtienen de minimizar (7.2). Con las técnicas usadas en la Sección 3 se podría demostrar que  $\sqrt{n}(\hat{\mathbf{B}}_n - \hat{\mathbf{B}}_0)$  tiene distribución asintótica normal multidimensional con matriz de covarianza asintótica  $C(\hat{\mathbf{B}}_0)$  dada por

$$C(\hat{\mathbf{B}}_0) = E(J(\mathbf{Y}, \mathbf{X}, \hat{\mathbf{B}}_0))^{-1} E\left(\frac{\partial \phi(\mathbf{Y}, \mathbf{X}, \hat{\mathbf{B}}_0)}{\partial \mathbf{B}} \frac{\partial \phi(\mathbf{Y}, \mathbf{X}, \hat{\mathbf{B}}_0)}{\partial \mathbf{B}}'\right) (E(J(\mathbf{Y}, \mathbf{X}, \hat{\mathbf{B}}_0))^{-1})'$$

donde

$$J(\mathbf{Y}, \mathbf{X}, \mathbf{B})_{ij} = \frac{\partial^2 \phi(\mathbf{Y}, \mathbf{X}, \mathbf{B})}{\partial B_i \partial B_j}$$

En el caso de particular de regresión logística multinomial, es decir cuando  $H(t) = \exp(t)$ ,  $C(\hat{\mathbf{B}}_0)$  tiene una expresión relativamente sencilla.

Por razones de simplicidad, se notará

$$\begin{aligned} \Pi_j &= \frac{\exp(\mathbf{x}' \underline{\beta}^j)}{1 + \sum_{i=1}^{k-1} \exp(\mathbf{x}' \underline{\beta}^i)} \quad \text{si } j = 1, \dots, k-1 \\ \Pi_k &= \frac{1}{1 + \sum_{i=1}^{k-1} \exp(\mathbf{x}' \underline{\beta}^i)} \quad \text{si } j = k. \end{aligned} \quad (7.14)$$

y  $\Pi_j^0$  cuando las  $\Pi_j$  cumplen (7.14) con  $\underline{\beta}^i = \underline{\beta}_0^i$ ,  $i = 1, \dots, k-1$ , vectores de parámetros verdaderos.

Dada  $B_l$ , componente  $l$ -ésima de  $\mathbf{B}$ , si  $l = m + q(h-1)$ ,  $m \in \{1, 2, \dots, q\}$ ,  $h \in \{1, \dots, k-1\}$ , se tiene que  $B_l$  corresponde a la  $m$ -ésima componente del vector  $\underline{\beta}^h$ .

En estas condiciones, si  $l = m + q(h-1)$  se tiene que

$$\frac{\partial \Pi_j}{\partial B_l} = \begin{cases} -X_m \Pi_j \Pi_h & \text{si } h \neq j \\ X_m \Pi_j (1 - \Pi_j) & \text{si } h = j \end{cases}$$

Luego, derivando con respecto a  $B_l$ , se obtiene que

$$\frac{\partial \phi(\mathbf{Y}, \mathbf{X}, \mathbf{B})}{\partial B_l} = X_m \left\{ \sum_{i \neq h} \Pi_h \psi_j(-\ln \Pi_j)(Y_j - \Pi_j) + (1 - \Pi_h) \psi_h(-\ln \Pi_h)(\Pi_h - Y_h) \right\}$$

Por lo tanto, si  $l = m + q(h - 1)$  y  $r = n + q(v - 1)$  se tiene que

$$E_{\mathbf{B}_0} \left( \frac{\partial \phi(\mathbf{Y}, \mathbf{X}, \mathbf{B}_0)}{\partial \mathbf{B}} \frac{\partial \phi(\mathbf{Y}, \mathbf{X}, \mathbf{B}_0)}{\partial \mathbf{B}} \right)'_{lr} = E_{\mathbf{B}_0} (X_m X_n L_{hv}),$$

donde

$$\begin{aligned} L_{hv} &= \sum_{i \neq h} \sum_{j \neq v} \Pi_h^0 \Pi_v^0 \psi_i(-\ln \Pi_i^0) \psi_j(-\ln \Pi_j^0) \text{Cov}(Y_i, Y_j) \\ &\quad - \Pi_h^0 (1 - \Pi_v^0) \sum_{i \neq h} \psi_v(-\ln \Pi_v^0) \psi_i(-\ln \Pi_i^0) \text{Cov}(Y_i, Y_v) \\ &\quad - \Pi_v^0 (1 - \Pi_h^0) \sum_{i \neq v} \psi_h(-\ln \Pi_h^0) \psi_i(-\ln \Pi_i^0) \text{Cov}(Y_i, Y_h) \\ &\quad + \psi_h(-\ln \Pi_h^0) \psi_v(-\ln \Pi_v^0) \text{Cov}(Y_h, Y_v) (1 - \Pi_h^0)(1 - \Pi_v^0). \end{aligned} \quad (7.16)$$

De manera que

$$E \left( \frac{\partial \phi(\mathbf{Y}, \mathbf{X}, \mathbf{B}_0)}{\partial \mathbf{B}} \frac{\partial \phi(\mathbf{Y}, \mathbf{X}, \mathbf{B}_0)}{\partial \mathbf{B}} \right)' = E(L \otimes \mathbf{X}\mathbf{X}'),$$

siendo  $L$  la matriz de componentes  $L_{hv}$  dadas por (7.16) y  $\otimes$  es el producto de Kronecker.

Por otra parte, si  $l = m + q(h - 1)$ ,  $r = n + q(v - 1)$  se tiene que

1)  $h \neq v$

$$\begin{aligned} \frac{\partial^2 \phi(\mathbf{Y}, \mathbf{X}, \mathbf{B})}{\partial B_l \partial B_r} &= \quad (7.17) \\ &= X_m X_n \left\{ \sum_{j \neq h, v} \Pi_h^0 \Pi_v^0 \{ (Y_j - \Pi_j^0) [\psi_j'(-\ln \Pi_j^0) - \psi_j(-\ln \Pi_j^0)] + \psi_j(-\ln \Pi_j^0) \Pi_j^0 \} \right. \\ &\quad - \Pi_h^0 \{ (Y_v - \Pi_v^0) [\psi_v'(-\ln \Pi_v^0)(1 - \Pi_v^0) + \psi_v(-\ln \Pi_v^0) \Pi_v^0] + \psi_v(-\ln \Pi_v^0) \Pi_v^0 (1 - \Pi_v^0) \} \\ &\quad \left. - \Pi_v^0 \{ (Y_h - \Pi_h^0) [\psi_h'(-\ln \Pi_h^0)(1 - \Pi_h^0) + \psi_h(-\ln \Pi_h^0) \Pi_h^0] + \psi_h(-\ln \Pi_h^0) \Pi_h^0 (1 - \Pi_h^0) \} \right\} \end{aligned}$$

Si se define para  $h \neq v$

$$T_{hv} = \Pi_h^0 \Pi_v^0 \left\{ \sum_{j \neq h, v} \Pi_j^0 \psi_j(-\ln \Pi_j^0) - (1 - \Pi_v^0) \psi_v(-\ln \Pi_v^0) - (1 - \Pi_h^0) \psi_h(-\ln \Pi_h^0) \right\}, \quad (7.18)$$

tomando esperanza condicional en (7.17) resulta que

$$E(J(\mathbf{Y}, \mathbf{X}, \underline{\mathbf{B}}_0)_{lr}) = E(X_m X_n T_{hv}). \quad (7.19)$$

2)  $h = v$

$$\begin{aligned} \frac{\partial^2 \phi(\mathbf{Y}, \mathbf{X}, \underline{\mathbf{B}})}{\partial B_l \partial B_l} &= \quad (7.20) \\ &= X_m X_n \left\{ \sum_{j \neq h} \Pi_h^0 \{ (Y_j - \Pi_j^0) [\psi_j'(-\ln \Pi_j^0) \Pi_h^0 + \psi_j(-\ln \Pi_j^0)(1 - \Pi_h^0)] + \psi_j(-\ln \Pi_j^0) \Pi_j^0 \Pi_h^0 \} \right. \\ &\quad \left. + (1 - \Pi_h^0) \{ (Y_h - \Pi_h^0) [\psi_h'(-\ln \Pi_h^0)(1 - \Pi_h^0) + \psi_h(-\ln \Pi_h^0) \Pi_h^0] + \psi_h(-\ln \Pi_h^0) \Pi_h^0 (1 - \Pi_h^0) \} \right\}. \end{aligned}$$

Luego, si se define

$$T_{hh} = (\Pi_h^0)^2 \sum_{j=1}^k \Pi_j^0 \psi_j(-\ln \Pi_j^0) + \Pi_h^0 (1 - 2\Pi_h^0) \psi_h(-\ln \Pi_h^0) \quad (7.21)$$

tomando esperanza condicional en (7.19) se tiene que si  $h = v$

$$E(J(\mathbf{Y}, \mathbf{X}, \underline{\mathbf{B}}_0)_{lr}) = E(X_m X_n T_{hh}). \quad (7.22)$$

Finalmente, de (7.19) y (7.22) se obtiene que

$$E(J(\mathbf{Y}, \mathbf{X}, \underline{\mathbf{B}}_0)) = E(T \otimes \mathbf{X}\mathbf{X}').$$

siendo  $T$  la matriz de componentes  $T_{hv}$  dadas por (7.18) y (7.21).



### 7.3 Ejemplos.

En esta sección se ilustra el comportamiento del M-estimador propuesto para el modelo de regresión logística multinomial a través del análisis de datos reales y simulados.

#### 7.3.1 Datos Simulados.

Con el fin de analizar el comportamiento del M-estimador se generaron muestras de vectores  $(\mathbf{Y}, \mathbf{X})$  que siguieran el modelo (7.1) y se las contaminó con un porcentaje de observaciones anómalas. En general, se observó que el M-estimador resiste la presencia de los datos atípicos, mientras que los estimadores obtenidos a partir del método de máxima verosimilitud pueden ser seriamente afectados.

El siguiente es uno de los ejemplos analizados.

En este ejemplo se generó una muestra de tamaño 150 de vectores  $(\mathbf{Y}, \mathbf{X})$  donde  $\mathbf{Y} = (Y_1, Y_2, Y_3)$ ,  $\mathbf{X} = (X_1, X_2, X_3)$  cumplen

$$\mathbf{Y}|\mathbf{X} = \mathbf{x} \sim M(1, H_1^*(\mathbf{x}, \underline{\mathbf{B}}^0), H_2^*(\mathbf{x}, \underline{\mathbf{B}}^0), H_3^*(\mathbf{x}, \underline{\mathbf{B}}^0))$$

donde  $H_i^*$  están dadas por (7.1\*) y  $\underline{\mathbf{B}}^0 = (\underline{\beta}_1^0, \underline{\beta}_2^0)$ , con  $\underline{\beta}_1^0 = (1, -2, 5)$  y  $\underline{\beta}_2^0 = (-1, -1, -3)$ . El vector  $\mathbf{X}$  se generó de manera tal que  $X_1$  fuera 1,  $X_2$  siguiera una distribución exponencial de parámetro  $\lambda = 1$  y  $X_3$  tuviera distribución normal standard independientes.

La muestra original fue contaminada con un 5 por ciento de datos atípicos. Estas contaminaciones se hicieron según cuatro criterios distintos. En la contaminación 1 la componente  $x_3$  fue elegida de manera tal que  $H_1^*(\mathbf{x}, \underline{\mathbf{B}}^0)$  fuera por lo menos 0.90 y a pesar de esto la observación correspondiente fue clasificada en la clase 3, es decir se tomó  $y_3 = 1$ . En contaminaciones 2, 3 y 4 se tomó un 5 por ciento de observaciones con  $x_3 = 3, 5$  y 10 respectivamente y aún cuando para todas estas observaciones  $H_1^*(\mathbf{x}, \underline{\mathbf{B}}^0) \cong 1$  se las clasificó en la clase 3 y no en la 1.

Las tablas que siguen dan los coeficientes estimados por el método de máxima verosimilitud (EMV) y por el M-estimador usando  $\rho_j = \rho_c$  con  $c = -\ln 0.05$  ( $M_{-\ln 0.05}$ ) para los

datos originales y los contaminados. Las tablas muestran que el estimador de máxima verosimilitud es sensiblemente afectado por la presencia de los datos atípicos, mientras que el M-estimador resiste la presencia de los mismos. En este sentido, el estimador de máxima verosimilitud cambia notoriamente a medida que la componente  $x_3$  de los datos contaminados aumenta. Por ejemplo, la tercera componente de vector  $\hat{\mathbf{B}}_{1,1}^0$ , es decir  $(\hat{\beta}_{13}^0)$ , cuando es estimado por máxima verosimilitud, disminuye progresivamente hasta ser negativa. Sin embargo, en las tres últimas contaminaciones los valores estimados por el M-estimador coinciden y son próximos a los parámetros verdaderos.

Tabla 16 Coeficientes estimados para los datos originales

Parámetros	EMV	$M_{-\ln 0.05}$
(1, -2, 5)	(0.22, -1.38, 4.69)	(0.53, -2.13, 5.91)
(-1, -1, -3)	(-1.18, -1.00, -3.13)	(-1.19, -1.23, -3.25)

Tabla 17 Coeficientes estimados para la contaminación 1

Parámetros	$EMV$	$M_{-\ln 0.05}$
(1, -2, 5)	(-0.33, -0.47, 2.26)	(0.29, -1.89, 5.67)
(-1, -1, -3)	(-1.34, -0.99, -3.38)	(-1.12, -1.22, -3.10)

Tabla 18 Coeficientes estimados para la contaminación 2

Parámetros	$EMV$	$M_{-\ln 0.05}$
(1, -2, 5)	(0.05, -0.30, 0.73)	(0.39, -1.99, 5.80)
(-1, -1, -3)	(-1.11, -1.02, -3.52)	(-1.24, -1.17, -3.17)

**Tabla 19 Coeficientes estimados para la contaminación 3**

<b>Parámetros</b>	<b><i>EMV</i></b>	<b><math>M_{-\ln 0.05}</math></b>
(1, -2, 5)	(0.33, -0.32, 0.18)	(0.39, -1.99, 5.80)
(-1, -1, -3)	(-0.94, -1.07, -3.71)	(-1.24, -1.17, -3.17)

**Tabla 20 Coeficientes estimados para la contaminación 4**

<b>Parámetros</b>	<b><i>EMV</i></b>	<b><math>M_{-\ln 0.05}</math></b>
(1, -2, 5)	(0.60, -0.39, -0.07)	(0.39, -1.99, 5.80)
(-1, -1, -3)	(-0.77, -1.15, -3.84)	(-1.24, -1.17, -3.17)

### 7.3.2 Datos de Lubischew.

En este ejemplo se muestra una aplicación del modelo de regresión logística multinomial al problema de discriminación, en el que se desea clasificar a un individuo en una de  $k$  clases distintas en función de  $q$  mediciones  $x_1, \dots, x_q$ .

En el presente caso se desea clasificar 74 escarabajos machos de tres especies distintas en sus correspondientes especies según dos variables:

$X_1$  : máximo ancho del aedeagus en la parte delantera

$X_2$  : ángulo frontal del aedagus

Se considerará que el vector  $\mathbf{Y} = (Y_1, Y_2, Y_3)$  es tal que  $Y_i = 1$  si la observación pertenece a la clase  $i$  y que

$$\mathbf{Y}|\mathbf{X} = \mathbf{x} \sim M(1, H_1^*(\mathbf{x}, \underline{\mathbf{B}}^0), H_2^*(\mathbf{x}, \underline{\mathbf{B}}^0), H_3^*(\mathbf{x}, \underline{\mathbf{B}}^0))$$

donde  $H_j^*$  están dadas por (7.1\*),  $\mathbf{X} = (1, X_1, X_2)$  y  $\underline{\mathbf{B}}^0 = (\underline{\beta}_1^0, \underline{\beta}_2^0)$  es el vector de parámetros a estimar.

La regla de asignación es la siguiente: un nuevo individuo será clasificado en la clase  $i$  si  $H_i^*(\mathbf{x}, \underline{\mathbf{B}}) \geq H_j^*(\mathbf{x}, \underline{\mathbf{B}})$  para todo  $j$ .

Con el fin de estudiar la resistencia del M-estimador propuesto, se contaminó la muestra de manera tal que 6 escarabajos de la especie 3 fueron clasificados como pertenecientes a la clase 1.

Los parámetros fueron estimados por el método de máxima verosimilitud (EMV) y por el M-estimador propuesto con  $\rho_j = \rho_c$  con  $c = -\ln 0.05$  ( $M_{-\ln 0.05}$ ). En la Tabla 21 se dan los valores estimados por ambos métodos.

Se puede observar que el estimador de máxima verosimilitud es muy sensible a la presencia de los datos alterados, ya que los valores estimados por este método cambian

notoriamente cuando se contamina la muestra. Por otro lado, el M-estimador no cambia ante al contaminar la muestra. Un comentario similar se puede hacer sobre las probabilidades estimadas, ya que en el caso de las estimadas por máxima verosimilitud conducirían en muchas de las observaciones a una clasificación errónea de las mismas, mientras que las estimadas por el M-estimador tienen un ajuste prácticamente perfecto. De hecho, cuando se clasifica a partir de las probabilidades estimadas por máxima verosimilitud, un 10 por ciento de los individuos (que no coinciden con los que fueron mal asignados inicialmente) son mal clasificados. También se observa que las probabilidades estimadas de la clase 3 en los individuos pertenecientes a la misma (que son aproximadamente un tercio de la muestra) descienden a valores próximos a  $1/2$ , mientras que cuando no hay contaminación estas probabilidades valen prácticamente uno.

Tabla 21 Coeficientes estimados

Datos	$EMV^*$	$M_{-\ln 0.05}$
Originales	(-209.73, 0.90, 6.89)	(-209.83, 0.92, 6.80)
	(59.35, -0.99, 5.82)	(58.54, -0.98, 5.83)
Contaminados	(-26.21, 0.15, 0.45)	(-209.83, 0.92, 6.80)
	(99.73, -0.91, 1.58)	(58.54, -0.98, 5.83)

## REFERENCIAS

- Cook, R.D. (1977), "Detection of influential observations in linear regression," *Technometrics* 19, 8-15.
- Cook, R.D., and Weisberg, S. (1982), *Residuals and Influence in regression*, London: Chapman and Hall.
- Huber, P.J. (1967), "The Behavior of Maximum Likelihood Estimates under Non-standard Conditions," *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1. Ed. L.M. LeCam and J. Neyman, 221-233, University of California Press.
- Johnson, W. (1985), "Influence Measures for Logistic Regression: Another Point of View," *Biometrika*, 72, 59-66.
- Krasker, W.S. and Welsch, R.E. (1982), "Efficient bounded influence regression estimation using alternative definitions of sensitivity," *Journal of the American Statistical Association* 77, 595-605.
- Künsch, H.R., Stefansky, L.A. , and Carroll, R.J. (1987), "Conditionally Unbiased Bounded Influence Estimation in General Regression Models, with Applications to Generalized Linear Models," Research report No. 54, Seminar für Statistik, Eidgenössische Technische Hochschule, Zurich.
- Maronna, R., Bustos, O., and Yohai, V.J. (1979), "Bias- and Efficiency-Robustness of General M-estimates for Regression with Random Carriers," in *Smoothing Techniques for Curve Estimation* (Lecture Note in Mathematics, 757), eds. T.Gasser and M.Rosemblatt, Berlin: Springer Verlag, pp. 91-116.
- Nelder, J.A., and Mead, R.(1965), *Computer Journal*, 7, 308.
- Pregibon, D.(1981), "Logistic Regression Diagnostics," *Annals of Statistics*, 9, 705-724.
- Pregibon, D. (1982), "Resistent Fits for some Commonly Used Logistic Models with Medical Applications," *Biometrics*, 38, 485-498.
- Rousseeuw, P.J. (1983), "Least Median of Squares Regression," *Journal of the American*

*Statistical Association*, 79, 871-880.

Rousseeuw, P.J. and Yohai, V.J. (1984), "Robust Regression by Means of S-estimates," in *Robust and Nonlinear Time Series* (Lecture Notes in Statistics, 26), eds. J.Franke, W.Hardle and R.D.Martin, New York: Springer Verlag, pp. 256-272

Siegel, A.F. (1982), "Robust Regression Using Repeated Medians," *Biometrika*, 69, 242-244.

Stefansky, L.A., Carroll, R.J. , and Ruppert D. (1986), "Optimally Bounded Score Functions for Generalized Linear Models with Applications to Logistic Regression," *Biomctrika*. 73, 413-425.

Yohai, V.J. (1987),"High Breakdown Point and High Efficiency Robust Estimates for Regression," *Annals of Statistics*, 15, 692-656.

Yohai, V.J., and Zamar, R.H. (1988), "High Breakdown-Point Estimates of Regression by Means of the Minimization of an Efficient Scale," *Journal of the American Statistical Association*, 83, 406-413.

∴  
∴  
∴