



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

Métodos robustos basados en transformaciones con aplicaciones al modelo lineal generalizado

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

Marina Silvia Valdora

Director de tesis: Dr. Víctor Yohai
Consejero de estudios: Dr. Víctor Yohai

Lugar de trabajo: Instituto de cálculo, FCEN-UBA

Buenos Aires, mayo 2014

Fecha de defensa: 2 de septiembre de 2014

Métodos robustos basados en transformaciones con aplicaciones al modelo lineal generalizado

Resumen

Esta tesis consta de 2 partes que corresponden a los Capítulos 1 y 2. En el Capítulo 1 proponemos una familia de estimadores robustos para modelos lineales generalizados. En el Capítulo 2 se introduce una familia de estimadores robustos para distribuciones dependientes de un parámetro.

Los estimadores definidos en el Capítulo 1 son M-estimadores redescendentes basados en transformaciones (MT-estimadores) y, más generalmente, M-estimadores pesados basados en transformaciones (WMT-estimadores). La idea principal es usar un M-estimador después de aplicar una función estabilizadora de la varianza a las respuestas. Mostramos la consistencia y la normalidad asintótica de estos estimadores. También calculamos una cota inferior para su punto de ruptura asintótico. Un estudio de Monte Carlo muestra que estos estimadores se comparan favorablemente con otros estimadores robustos para modelos lineales generalizados con respuesta Poisson y log link. Por último, consideramos un ejemplo de datos reales y comparamos el ajuste dado por el MT-estimador con los ajustes correspondientes a otros estimadores existentes.

Los estimadores definidos en el Capítulo 2 son estimadores para datos univariados basados en la transformación integral de probabilidad (MI-estimadores). Estos estimadores tienen una definición simple y son muy simples de calcular. Mostramos la consistencia y normalidad asintótica de los mismos y hallamos cotas para su punto de ruptura asintótico. Estudiamos en especial el caso de la distribución de Poisson, en el que probamos que el punto de ruptura es óptimo. Para el caso de la distribución de Poisson realizamos un estudio de Monte Carlo para comparar el desempeño de estos estimadores con el de otros estimadores robustos para datos univariados.

Robust Methods Based on Transformations with Applications to General Linear Models

Abstract

This thesis consists of two parts corresponding to Chapters 1 and 2. In Chapter 1 we propose a family of robust estimators for generalized linear models. In Chapter 2 we introduce a family of robust estimators for distributions that depend on a single parameter.

The estimators defined in Chapter 1 are redescending M-estimators based on transformations (MT-estimators) and, more generally, weighted M-estimators based on transformations (WMT-estimators). The main idea is to use an M-estimator after applying a variance stabilizing function to the responses. We show the consistency and asymptotic normality of these estimators. We also compute a lower bound for their asymptotic breakdown point. A Monte Carlo study shows that these estimators compare favourably to other robust estimators for generalized linear models with Poisson response and log link. Finally, we consider an example of real data and compare the fit obtained using the MT-estimator to the fits corresponding to other existing estimators.

The estimators defined in Chapter 2 are estimators for univariate data based on the probability integral transformation (MI-estimators). These estimators have a simple definition and are very easy to compute. We show their consistency and asymptotic normality and find bounds for their asymptotic breakdown point. We study in particular the case of the Poisson distribution, in which we prove that the breakdown point is optimum. For the case of the Poisson distribution we perform a Monte Carlo study to compare the performance of MI-estimators to that of other robust estimators for univariate data.

Key words: M-estimators, transformations, breakdown point, generalized linear models.

Agradecimientos

Quiero agradecerles a todos los que me ayudaron en estos años y que hicieron posible que escribiera esta tesis.

A Javier por ser mi compañero de vida, por la contención, la confianza y el apoyo incansable sin los cuales nada de esto hubiese sido posible.

A mis hijos Fede y Agus por sus risas y su amor incondicional y por recordarme cada día las cosas verdaderamente importantes de la vida.

A mis hermanos Gaby y Germán por haber sido y seguir siendo tan buenos compañeros.

A mis padres, por haberme enseñado el placer por el estudio y por haberme dado tanto en tan poco tiempo.

A Norma y Choqui por cuidar tantas horas a mis hijos y sobre todo por hacerlos tan felices.

Al pueblo argentino por mantener con su esfuerzo la universidad pública que hizo posibles mis estudios.

A mis compañeros de oficina Julieta, Paula, Agustín, Andrés y Lucas por las charlas, los consejos y por su amistad.

A toda la gente del Instituto de Cálculo por el compañerismo y la buena onda que hacen que sea un placer trabajar allí.

A Mariela, Dani, Maru, Ana, Marta y Graciela por todo lo que me permitieron aprender de ellas.

A Graciela por su valiosísimo apoyo y porque sus comentarios y correcciones mejoraron notablemente esta tesis.

Especialmente a Víctor, por haberme dado la oportunidad aprender de él, por su confianza, su paciencia y su generosidad.

Índice general

1. M-estimadores basados en transformaciones para modelos lineales generalizados	9
1.1. Introducción	9
1.1.1. Modelos lineales generalizados	9
1.1.2. Algunas propuestas para la estimación robusta de MLG . . .	10
1.1.3. Organización del capítulo	13
1.2. Definición de M-estimadores basados en transformaciones	14
1.3. Algunos ejemplos de transformaciones para estabilizar la varianza .	16
1.3.1. Regresión de Poisson.	17
1.3.2. Regresión exponencial.	17
1.3.3. Regresión binomial.	17
1.4. Consistencia	18
1.5. Normalidad Asintótica	20
1.6. Punto de ruptura asintótico	23
1.6.1. Regresión de Poisson	24
1.6.2. Regresión exponencial	24
1.6.3. Ejemplo	25
1.7. Método computacional	25
1.7.1. Cálculo de $m(\lambda)$	25
1.7.2. Cálculo de los WMT-estimadores	26
1.7.3. Tiempos de cómputo	27
1.8. Estudio de Monte Carlo	28
1.9. Una aplicación a un conjunto de datos reales: datos de epilepsia . .	36
2. Estimadores robustos para un parámetro unidimensional	38
2.1. Introducción	38
2.2. Definición de M-estimadores basados en la transformación integral de probabilidad	38
2.3. Consistencia	40
2.4. Normalidad asintótica	41
2.5. Medidas de robustez	42

2.5.1.	Función de influencia	42
2.5.2.	Punto de ruptura asintótico.	44
2.6.	Aplicación a la distribución de Poisson	45
2.6.1.	Varianza asintótica	45
2.6.2.	Función de influencia	48
2.6.3.	Punto de ruptura	49
2.6.4.	Estudio de Monte Carlo	52
A.	Demostraciones de los Teoremas del capítulo 1	61
A.1.	Consistencia	61
A.1.1.	Un resultado general de consistencia	62
A.1.2.	Consistencia de los WMT-estimadores	67
A.2.	Distribución Asintótica	71
A.2.1.	Normalidad asintótica de los WMT-estimadores	74
A.3.	Punto de ruptura asintótico	77
A.4.	Método computacional	81
B.	Demostraciones de los Teoremas del capítulo 2	83
B.1.	Consistencia	83
B.2.	Punto de ruptura asintótico	84

Capítulo 1

M-estimadores basados en transformaciones para modelos lineales generalizados

1.1. Introducción

1.1.1. Modelos lineales generalizados

Los modelos lineales generalizados (MLG) son ampliamente utilizados en el análisis de datos. Estos modelos suponen que la variable respuesta y sigue una distribución F_λ donde λ se relaciona con las variables explicativas x_1, \dots, x_p a través de una función g llamada función de enlace de manera que

$$g(\lambda) = \mathbf{x}^T \boldsymbol{\beta}_0,$$

donde $\mathbf{x} = (x_1, \dots, x_p)$ es el vector de covariables o variables explicativas y $\boldsymbol{\beta}_0 \in \mathbb{R}^p$ es un parámetro desconocido. Asumimos que F_λ pertenece a una familia exponencial de manera que tiene una densidad de la forma

$$f(y, \lambda) = \exp(l(\lambda)y - H(l(\lambda)) - C(y)), \quad (1.1)$$

donde l , H y C son funciones dadas. Si $l = g$ la función de enlace se llama canónica.

Sea $\hat{\lambda}(y)$ el estimador de máxima verosimilitud de λ en un modelo con una sola observación y . La función deviance se define como

$$d(y, \lambda) = 2 \left(\log f(y, \hat{\lambda}(y)) - \log f(y, \lambda) \right)$$

Dada una muestra aleatoria $(y_1, \mathbf{x}_1) \dots (y_n, \mathbf{x}_n)$ de un MLG con función de

enlace canónica, la log-verosimilitud tiene la forma

$$L(\boldsymbol{\beta}) = \sum_{i=1}^n \log f(y_i, g^{-1}(\mathbf{x}_i^T \boldsymbol{\beta})) = \sum_{i=1}^n (\mathbf{x}_i^T \boldsymbol{\beta} y_i - H(\mathbf{x}_i^T \boldsymbol{\beta}) - C(y_i)).$$

El estimador de máxima verosimilitud maximiza la log-verosimilitud o, equivalentemente, minimiza

$$\sum_{i=1}^n d(y_i, g^{-1}(\mathbf{x}_i^T \boldsymbol{\beta})).$$

Derivando, se obtienen las ecuaciones para el EMV:

$$\sum_{i=1}^n (y_i - h(\mathbf{x}_i^T \boldsymbol{\beta})) \mathbf{x}_i = 0, \quad (1.2)$$

donde h es la derivada de H .

En el caso de regresión de Poisson se tiene que $\hat{\lambda}(y) = y$ y entonces

$$d(y, \lambda) = 2 \left(y \log \left(\frac{y}{\lambda} \right) - (y - \lambda) \right). \quad (1.3)$$

Las ecuaciones de estimación para el EMV son en este caso

$$\sum_{i=1}^n (y_i - e^{\mathbf{x}_i^T \boldsymbol{\beta}}) \mathbf{x}_i = 0.$$

Es conocido que el estimador de máxima verosimilitud es el más eficiente cuando todas las observaciones siguen el modelo pero es muy sensible a la presencia de datos atípicos o outliers. Para solucionar este problema se han propuesto varios estimadores robustos.

1.1.2. Algunas propuestas para la estimación robusta de MLG

Estimadores condicionalmente insesgados de influencia acotada

Estos estimadores fueron propuestos por Künsch, Stefanski y Carroll (1989). Ellos consideraron M-estimadores de la forma

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

donde $\Psi : \mathbb{R}^{2p+1} \longrightarrow \mathbb{R}^p$ es tal que

$$E(\Psi(y_i, \mathbf{x}_i, \boldsymbol{\beta}) | \mathbf{x}_i) = 0.$$

Los estimadores que satisfacen esta última ecuación se denominan condicionalmente Fisher-consistentes.

Ellos hallaron el estimador de esta clase que es solución de un problema de optimización similar al considerado por Hampel (1974). Este estimador maximiza una medida de la eficiencia sujeto a una cota en una medida de la sensibilidad infinitesimal similar a la GES (sensibilidad a errores groseros).

La función de score Ψ óptima tiene la forma

$$\Psi_{cond}(y, \mathbf{x}, \boldsymbol{\beta}, b, B) = D(y, \mathbf{x}, \boldsymbol{\beta}, b, B) W_b \left(|D(y, \mathbf{x}, \boldsymbol{\beta}, b, B)| (\mathbf{x}^T B^{-1} \mathbf{x})^{1/2} \right) \mathbf{x}, \quad (1.4)$$

donde b es la cota en la medida de sensibilidad infinitesimal, la función D viene dada por

$$D(y, \mathbf{x}, \boldsymbol{\beta}, b, B) = y - h(\mathbf{x}^T \boldsymbol{\beta}) - c \left(\mathbf{x}^T \boldsymbol{\beta}, \frac{b}{(\mathbf{x}^T B^{-1} \mathbf{x})^{1/2}} \right)$$

y $W_b(u) = \psi_b^H(u)/u$ donde ψ_b^H es la función de Huber, definida por

$$\psi_b^H(u) = \text{sign}(u) \min(|u|, b), \quad (1.5)$$

Por otro lado, B es una matriz de dispersión que satisface

$$E(\Psi_{cond}(y, \mathbf{x}, \boldsymbol{\beta}, b, B) \Psi_{cond}(y, \mathbf{x}, \boldsymbol{\beta}, b, B)^T) = B.$$

Comparando (1.4) con (1.2) se observa que ambas funciones de score son muy parecidas. La función W_b penaliza las observaciones atípicas y hace que Ψ_{cond} sea acotada y el término $c(\mathbf{x}^T \boldsymbol{\beta}, b/(\mathbf{x}^T B^{-1} \mathbf{x})^{1/2})$ se agrega para asegurar la condición de Fisher consistencia condicional. Estos estimadores son altamente robustos para pequeños porcentajes de contaminación. Sin embargo, Maronna, Bustos y Yohai (1979) probaron que, en el caso del modelo lineal, el punto de ruptura de estos estimadores tiende a cero cuando p tiende a infinito.

Estimadores de quasi-verosimilitud robustos

Supongamos ahora que la función de distribución de y , F_λ , es desconocida pero que se conocen funciones m y V tales que $\text{var}_\lambda(y) = V(\lambda)$ y $E_\lambda(y) = m(\lambda)$.

La función de quasi-verosimilitud, introducida por Wedderburn (1974), para una única observación con distribución F_λ se define como

$$Q(\lambda, y) = \int_0^\lambda \frac{y - m(t)}{V(t)} m'(t) dt.$$

La ventaja de la función de quasi-verosimilitud es que no requiere conocer la familia de distribuciones F_λ , sino solamente la media y la varianza. Los estimadores de quasi-verosimilitud se definen de manera similar a los estimadores de máxima verosimilitud, reemplazando la función de verosimilitud por la de quasi-verosimilitud. Es decir, se definen por

$$\arg \text{máx} \sum_{i=1}^n Q(g(\mathbf{x}_i^T \boldsymbol{\beta}), y_i).$$

Derivando con respecto a $\boldsymbol{\beta}$ se obtiene

$$\sum_{i=1}^n \frac{y_i - m(g(\mathbf{x}_i^T \boldsymbol{\beta}))}{V(g(\mathbf{x}_i^T \boldsymbol{\beta}))} m'(g(\mathbf{x}_i^T \boldsymbol{\beta})) g'(\mathbf{x}_i^T \boldsymbol{\beta}) \mathbf{x}_i = 0. \quad (1.6)$$

Podemos observar que estos estimadores son M-estimadores con función de score

$$\frac{y_i - m(g(\mathbf{x}_i^T \boldsymbol{\beta}))}{V(\mathbf{x}_i^T \boldsymbol{\beta})} m'(g(\mathbf{x}_i^T \boldsymbol{\beta})) g'(\mathbf{x}_i^T \boldsymbol{\beta}) \mathbf{x}_i.$$

Estos estimadores no son robustos porque Q no es acotada. Para solucionar esta falta de robustez Cantoni y Ronchetti (2001) propusieron una robustificación de estos estimadores definiendo los estimadores de quasi-verosimilitud robustos mediante la ecuación.

$$\sum_{i=1}^n w(\mathbf{x}_i, \hat{\boldsymbol{\mu}}_n, \hat{\boldsymbol{\Sigma}}_n) (v(y_i, m(g(\mathbf{x}_i^T \boldsymbol{\beta}))) - a(g(\mathbf{x}_i^T \boldsymbol{\beta}))) m'(g(\mathbf{x}_i^T \boldsymbol{\beta})) g'(\mathbf{x}_i^T \boldsymbol{\beta}) \mathbf{x}_i = 0, \quad (1.7)$$

donde

$$v(y, \lambda) = \psi \left(\frac{y - m(\lambda)}{V^{1/2}(\lambda)} \right) \frac{1}{V^{1/2}(\lambda)},$$

$$a(\lambda) = E_\lambda (v(y, \lambda)),$$

ψ es una función monótona y acotada, por ejemplo en la familia de Huber dada por (1.5), $\hat{\boldsymbol{\mu}}_n$ y $\hat{\boldsymbol{\Sigma}}_n$ son estimadores robustos de posición y matriz de covarianza de \mathbf{x} respectivamente basados en $\mathbf{x}_1, \dots, \mathbf{x}_n$ y w es una función de la distancia de Mahalanobis de la forma

$$w(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \omega \left(\sqrt{(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})} \right). \quad (1.8)$$

El propósito de la función ψ es acotar la influencia de las observaciones con residuos grandes mientras que el de la función w es acotar la influencia de las observaciones de alta palanca (leverage). El propósito del término $a(\mathbf{x}_i^T \boldsymbol{\beta})$ es asegurar la condición (1.6) de consistencia de Fisher.

Como veremos en el estudio de Monte Carlo en la Sección 1.8 el uso de una función de score monótona hace que la robustez de estos estimadores sea muy limitada.

Otras propuestas robustas

Bergesio y Yohai (2011) introdujeron estimadores de proyección para modelos lineales generalizados que son una generalización de los estimadores de proyección para modelos lineales introducidos por Maronna y Yohai (1993). Estos estimadores son altamente robustos pero no son asintóticamente normales. Por lo tanto, proponen usar M-estimadores a un paso usando como estimadores iniciales los P-estimadores. Los estimadores a un paso son asintóticamente normales y conservan las propiedades de robustez de los P-estimadores. Una desventaja de estos estimadores es que su cálculo requiere algoritmos de alta complejidad.

Otra clase de estimadores robustos es la de los M-estimadores propuestos por Bianco y Yohai (1996) y luego estudiados por Croux y Haesbroeck (2003). Los estimadores propuestos por Bianco y Yohai (1996) fueron generalizados por Bianco, Boente y Rodrigues (2012) y aplicados al caso de regresión de Poisson. Estos estimadores se definen como

$$\widehat{\beta}_n = \arg \min_{\beta} \sum_{i=1}^n \rho(y_i, \mathbf{x}_i^T \beta, t) \omega_1(\mathbf{x}_i),$$

donde

$$\rho(y, u) = \phi(d(y, g^{-1}(u))) + G(g^{-1}(u)),$$

donde d es la función deviance definida en (1.3) y g es la función de enlace. Típicamente ϕ es una función que se asemeja a la identidad en un entorno de cero. El término $G(g^{-1}(u))$ es necesario para asegurar la Fisher consistencia y verifica

$$G'(t) = E \left(\phi'(-\log(y, t) + D(y)) \frac{\frac{\partial}{\partial t} f(y, t)}{f(y, t)} \right).$$

Morgenthaler (1992) propuso una robustificación de los estimadores de quasi verosimilitud usando una medida de bondad de ajuste basada en la norma L^1 . Por lo tanto, la función de score correspondiente, la función signo, es monótona como en el caso de los estimadores descritos en la Sección 1.1.2 con la consiguiente falta de robustez. Para MLG en los que F_λ pertenece a la familia de distribuciones de Bernoulli podemos citar también los trabajos de Christmann (1994) y de Rousseeuw y Christmann (2003).

1.1.3. Organización del capítulo

En la Sección 1.2, definimos los M-estimadores basados en transformaciones (MT) y los M-estimadores pesados basados en transformaciones (WMT). En la Sección 1.3, damos ejemplos de transformaciones estabilizadoras de la varianza para

algunas familias de distribuciones. En las secciones 1.4 y 1.5, estudiamos la consistencia y la distribución asintótica, respectivamente, de los WMT-estimadores. Por otra parte, en la Sección 1.6 obtenemos una cota inferior para el punto de ruptura de los MT-estimadores y la calculamos para los casos particulares de regresión de Poisson y regresión exponencial. En la Sección 1.7, describimos el método computacional que utilizamos para calcular los estimadores propuestos mientras que en la Sección 1.8, mostramos los resultados de un estudio de Monte Carlo en los que se observa que los estimadores propuestos se comparan favorablemente con otros estimadores existentes para el modelo de regresión de Poisson. Por último, en la Sección 1.9, mostramos una aplicación a un conjunto de datos reales. Los resultados enunciados en este capítulo se demuestran en el Apéndice A, donde damos teoremas generales de consistencia, normalidad asintótica y punto de ruptura para una clase de M-estimadores muy general que incluye a los WMT-estimadores. Luego, mostramos cómo estos resultados se aplican a los WMT-estimadores. Todos los lemas y proposiciones del Apéndice A aparecen numerados como A.1.1, A.2.1, etc para facilitar su cita.

1.2. Definición de M-estimadores basados en transformaciones

En esta sección, introducimos una familia de estimadores para el MLG. Estos estimadores tienen la ventaja de tener una definición simple y al mismo tiempo ser altamente robustos. Los estimadores que proponemos son M-estimadores redescendentes aplicados a las respuestas transformadas (MT-estimadores). El propósito de transformar las respuestas es estabilizar sus varianzas a un valor aproximadamente constante y así lograr un correcto escalamiento de la función de pérdida usada para definir el M-estimador.

Un ejemplo importante de MLG es el caso en que y sólo toma los valores 0 y 1, es decir, cuando F_λ es la familia de distribuciones de Bernoulli. Los estimadores propuestos no son aplicables en este caso. Es fácil ver que, si y_1, \dots, y_n siguen la distribución de Bernoulli entonces el estimador propuesto coincide con un M-estimador ordinario con las variables no transformadas.

Una dificultad que se encuentra al definir los M-estimadores es que, en general, la dispersión de F_λ depende de λ , y, por lo tanto, las variables de respuesta correspondientes a diferentes observaciones de la muestra pueden tener dispersiones diferentes. Por lo tanto, el escalamiento de la función ρ puede ser diferente para cada observación, pero este escalamiento depende de β , que es desconocido. Proponemos solucionar este problema transformando las respuestas. Para ello utilizamos una función $t(y)$ tal que la varianza de $t(y)$ no dependa o dependa

muy poco de λ . En la Sección 1.3 daremos algunos ejemplos de cómo obtener una función t con esta propiedad.

Consideremos un modelo lineal generalizado (GLM) donde $y \in \mathbb{R}$ es la respuesta y $\mathbf{x} = (x_1, \dots, x_p)^T$ es un vector de variables explicativas. Asumimos que

$$y|\mathbf{x} \sim F_\lambda, \quad (1.9)$$

donde F_λ es una familia exponencial de distribuciones discreta o continua en \mathbb{R} con el mismo soporte D , $\lambda \in \Lambda \subset \mathbb{R}$ y

$$g(\lambda) = \boldsymbol{\beta}_0^T \mathbf{x}, \quad (1.10)$$

donde $\boldsymbol{\beta}_0 \in \mathbb{R}^p$ es desconocido y $g : \mathbb{R} \rightarrow \mathbb{R}$ es una función de enlace conocida. Supondremos que λ toma valores en un intervalo $(\lambda^{(1)}, \lambda^{(2)})$ donde $\lambda^{(1)}$ puede ser $-\infty$ y $\lambda^{(2)}$ puede ser $+\infty$. Asumiremos también que $g^{-1} : \mathbb{R} \rightarrow (\lambda^{(1)}, \lambda^{(2)})$ es continua y estrictamente creciente y que

$$\lim_{u \rightarrow -\infty} g^{-1}(u) = \lambda^{(1)}, \quad \lim_{u \rightarrow \infty} g^{-1}(u) = \lambda^{(2)}. \quad (1.11)$$

Supongamos que $t : \mathbb{R} \rightarrow \mathbb{R}$ es tal que la varianza $t(y)$ es aproximadamente constante cuando y tiene distribución F_λ . Sea $\rho : \mathbb{R} \rightarrow \mathbb{R}$ una función continua y acotada con un único mínimo en 0 y definamos $m(\lambda)$ por

$$m(\lambda) = \arg \min_u E_\lambda(\rho(t(y) - u)). \quad (1.12)$$

Vamos a suponer que $m(\lambda)$ es continua y esta unívocamente definida para todo λ . Dada una muestra aleatoria $(y_1, \mathbf{x}_1), \dots, (y_n, \mathbf{x}_n)$ definimos el M-estimador pesado basado en transformaciones (WMT-estimador) de $\boldsymbol{\beta}_0$ por

$$\widehat{\boldsymbol{\beta}}_n = \arg \min_{\boldsymbol{\beta}} L_n(\boldsymbol{\beta}), \quad (1.13)$$

donde

$$L_n(\boldsymbol{\beta}) = \frac{1}{n} \sum_{i=1}^n \rho(t(y_i) - m(g^{-1}(\boldsymbol{\beta}^T \mathbf{x}_i))) w(\mathbf{x}_i, \widehat{\boldsymbol{\mu}}_n, \widehat{\boldsymbol{\Sigma}}_n), \quad (1.14)$$

$\widehat{\boldsymbol{\mu}}_n$ y $\widehat{\boldsymbol{\Sigma}}_n$ son estimadores robustos de posición y matriz de covarianza de \mathbf{x} respectivamente basados en $\mathbf{x}_1, \dots, \mathbf{x}_n$ y w es una función de la distancia de Mahalanobis de la forma

$$w(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \omega(\sqrt{(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})}), \quad (1.15)$$

con ω una función no negativa y no creciente. Usamos estimadores consistentes $\widehat{\boldsymbol{\mu}}_n$ y $\widehat{\boldsymbol{\Sigma}}_n$ de manera que $\widehat{\boldsymbol{\mu}}_n \rightarrow \boldsymbol{\mu}_0$ c.s. y $\widehat{\boldsymbol{\Sigma}}_n \rightarrow \boldsymbol{\Sigma}_0$ c.s., donde $\boldsymbol{\Sigma}_0$ es definida positiva.

El propósito de la función de peso $w(\mathbf{x}, \widehat{\boldsymbol{\mu}}_n, \widehat{\boldsymbol{\Sigma}}_n)$ es penalizar las observaciones de alto leverage.

Tenemos que

$$\begin{aligned} & E_{\beta_0} (\rho(t(y_i) - m(g^{-1}(\boldsymbol{\beta}^T \mathbf{x}_i))) w(\mathbf{x}_i, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)) \\ &= E [E_{\beta_0} \{ \rho(t(y_i) - m(g^{-1}(\boldsymbol{\beta}^T \mathbf{x}_i))) | \mathbf{x}_i \} w(\mathbf{x}_i, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)] \end{aligned}$$

y $E_{\beta_0} (\rho(t(y_i) - m(g^{-1}(\boldsymbol{\beta}^T \mathbf{x}_i))) | \mathbf{x}_i)$ se minimiza cuando $\boldsymbol{\beta} = \boldsymbol{\beta}_0$ para todo i . Entonces $E_{\beta_0} (\rho(t(y_i) - m(g^{-1}(\boldsymbol{\beta}^T \mathbf{x}_i))) w(\mathbf{x}_i, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0))$ también se minimiza cuando $\boldsymbol{\beta} = \boldsymbol{\beta}_0$ y, por lo tanto, como hemos asumido que la función m está unívocamente definida, los WMT-estimadores son Fisher consistentes. Esta, y otras condiciones necesarias para la consistencia, se listan en la Sección 1.4.

Como la varianza de $t(y)$ dado \mathbf{x} es aproximadamente constante, no es necesario usar una escala en la definición del M-estimador. Como ρ es acotada, el estimador definido por (1.13) es robusto aún cuando $\omega \equiv 1$. Sin embargo, en los casos en que se esperan outliers de alto leverage la robustez del estimador puede aumentarse usando una función de peso. En algunos casos el uso de pesos puede disminuir, no sólo la eficiencia, sino también la robustez del estimador. Esto ocurre cuando hay observaciones buenas de alto leverage, es decir cuando hay observaciones (\mathbf{x}, y) tales que $(\mathbf{x} - \widehat{\boldsymbol{\mu}}_n) \widehat{\boldsymbol{\Sigma}}_n^{-1} (\mathbf{x} - \widehat{\boldsymbol{\mu}}_n)$ es grande pero la respuesta y es generada de acuerdo al GLM. En estos casos la función de peso penalizará observaciones buenas y, por lo tanto, puede aumentar la influencia de outliers con bajo leverage que tienen pesos más grandes. Por esta razón, no debe descartarse la posibilidad de usar una función de peso constante, $w(u) = 1$. A los WMT-estimadores con función de peso constante los llamaremos, para abreviar, MT-estimadores. Claramente todos los resultados que probaremos para los WMT-estimadores se aplican también a los MT-estimadores.

1.3. Algunos ejemplos de transformaciones para estabilizar la varianza

Si notamos $\mu(\lambda)$ y $\nu(\lambda)$ a la media y la varianza de F_λ respectivamente, una expansión de Taylor de orden uno muestra que, tomando

$$t(y) = \int_0^y \frac{1}{[\nu(\mu^{-1}(u))]^{1/2}} du, \quad (1.16)$$

se obtiene que $\text{var}(t(y))$ es aproximadamente constante. Para ver esto, sea $y \sim F_{\lambda_0}$ y sea $\mu_0 = \mu(\lambda_0)$. Entonces bajo condiciones de regularidad apropiadas se tiene

$$t(y) \approx t(\mu_0) + \frac{1}{[\nu(\mu^{-1}(\mu_0))]^{1/2}} (y - \mu_0).$$

Por lo tanto,

$$V(t(y)) \approx \frac{1}{\nu(\mu^{-1}(\mu_0))} V(y) = 1,$$

como queríamos demostrar.

En muchos casos, la relación entre la esperanza y la varianza viene dada por

$$\nu(\lambda) = \mu(\lambda)^q. \quad (1.17)$$

En estos casos, (1.16) da como resultado

$$t(y) = \begin{cases} y^{-(q/2)+1} & \text{if } q \neq 2 \\ \log(y) & \text{if } q = 2 \end{cases}. \quad (1.18)$$

1.3.1. Regresión de Poisson.

En este caso, p_λ tiene como soporte al conjunto \mathbb{N}_0 de los enteros no negativos y la función de probabilidad puntual es

$$p(y, \lambda) = \exp(-\lambda)\lambda^y/y!.$$

Se tiene $\mu(\lambda) = \lambda$ y $\nu(\lambda) = \lambda$ y, por lo tanto, $t(y) = y^{1/2}$. La Figura 1.1, que muestra un gráfico de $\text{var}(y^{1/2})$ como función de λ , confirma que esta función es casi constante. De ahora en más, para el modelo de regresión de Poisson, la transformación que usaremos es $t(y) = \sqrt{y}$.

1.3.2. Regresión exponencial.

Consideremos ahora el caso en que F_λ tiene soporte en el conjunto \mathbb{R}^+ de los números reales positivos con densidad

$$p(y, \lambda) = \frac{1}{\lambda} \exp\left(-\frac{y}{\lambda}\right) I(y > 0).$$

En este caso, $\mu(\lambda) = \lambda$ y $\nu(\lambda) = \lambda^2$, y entonces, de acuerdo a (1.17) y (1.18), $q = 2$ y $t(y) = \log(y)$. Luego $\log(y) = w + \log(\lambda)$ donde $\exp(w)$ es una variable aleatoria con distribución F_1 . Por lo tanto, $\text{var}_\lambda(\log(y))$ es constante.

1.3.3. Regresión binomial.

Supongamos ahora que F_λ es una distribución $\text{Bi}(\lambda, n)$, es decir, su función de probabilidad puntual es

$$p(y, \lambda) = \binom{n}{k} \lambda^y (1 - \lambda)^{n-y}, \quad y = 0, 1, \dots, n, \quad 0 \leq \lambda \leq 1.$$

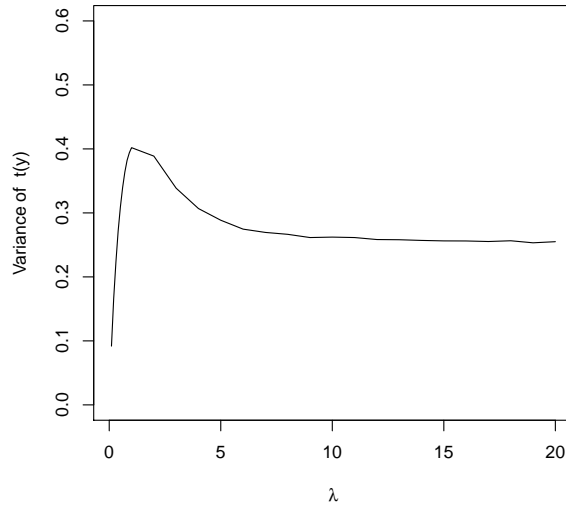


Figura 1.1: Varianza de $t(y) = \sqrt{y}$ para $y \sim P(\lambda)$

Entonces $\mu(\lambda) = n\lambda$, $\nu(\lambda) = n\lambda(1 - \lambda)$, y (1.16) da como resultado $t(y) = \arcsin\left(\sqrt{y/n}\right)$. La Figura 1.2 muestra un gráfico de $\text{var}_\lambda\left(\arcsin\sqrt{y/n}\right)$ para $n = 5$, que es casi constante.

1.4. Consistencia

En esta sección, estudiamos la consistencia de los WMT-estimadores. Necesitamos suponer que se cumplen las siguientes condiciones.

- A1** $\sup_\lambda \text{var}_\lambda(t(y)) = A < \infty$.
- A2** $m(\lambda)$ está unívocamente definida para todo λ y $\lambda_1 < \lambda_2$ implica $m(\lambda_1) < m(\lambda_2)$.
- A3** F_λ es continua en λ .
- A4** Si $\lambda_1 > \lambda_2$, $z_1 \sim F_{\lambda_1}$ y $z_2 \sim F_{\lambda_2}$ entonces z_1 es estocásticamente mayor que z_2 .
- A5** La función t es creciente y continua.
- B1** $\rho(u) \geq 0$, $\rho(0) = 0$ y $\rho(u) = \rho(-u)$.

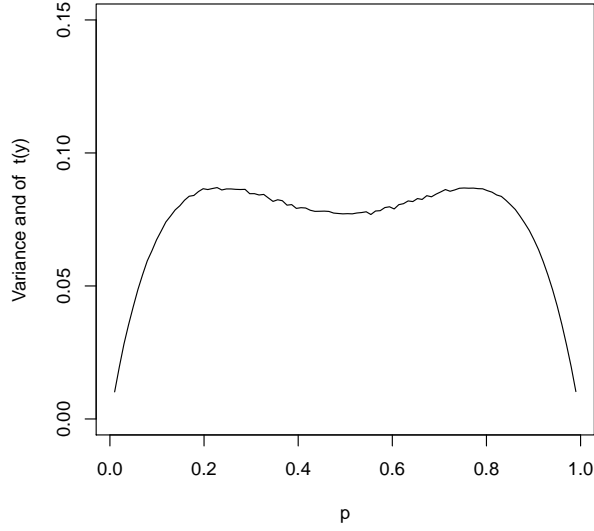


Figura 1.2: Varianza de $t(y) = \arcsin\left(\sqrt{y/5}\right)$ para $y \sim Bi(5, \lambda)$

B2 $\lim_{u \rightarrow \infty} \rho(u) = a < \infty$. Sin pérdida de generalidad asumiremos que $a = 1$.

B3 $0 \leq u < v$ implica $\rho(u) \leq \rho(v)$.

B4 $0 \leq u < v$ y $\rho(u) < 1$ implica $\rho(u) < \rho(v)$.

B5 ρ es continua.

B6 Existe η tal que $\rho(A^{1/2} + \eta) < 1$.

B7 Existe $\boldsymbol{\mu}_0 \in \mathbb{R}^p$ y una matriz definida positiva $\boldsymbol{\Sigma}_0$ tal que $\hat{\boldsymbol{\mu}}_n \rightarrow \boldsymbol{\mu}_0$ c.s. y $\hat{\boldsymbol{\Sigma}}_n \rightarrow \boldsymbol{\Sigma}_0$ c.s con $P(w(\mathbf{x}, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) > 0) > 0$.

B8 La función de peso $\omega : \mathbb{R} \rightarrow \mathbb{R}$ es continua, acotada, no creciente y no negativa y $\sup_{v \in \mathbb{R}} \omega(v) = 1$.

Sea $S = \{\mathbf{t} \in \mathbb{R}^p : \|\mathbf{t}\| = \mathbf{1}\}$. Entonces

$$\inf_{\mathbf{t} \in S} P(\{\mathbf{t}^T \mathbf{x} \neq 0\} \cap \{w(\mathbf{x}, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) > 0\}) > 0.$$

Notar que las condiciones **B7-B9** no son necesarias en el caso de los MT-estimadores.

Sean

$$m_1 = \inf_{\lambda} m(\lambda) = \lim_{\lambda \rightarrow \lambda^{(1)}} m(\lambda), \quad (1.19)$$

$$m_2 = \sup_{\lambda} m(\lambda) = \lim_{\lambda \rightarrow \lambda^{(2)}} m(\lambda), \quad (1.20)$$

con $\lambda^{(1)}$ y $\lambda^{(2)}$ definidas por (1.11) y $m_3 = m(g^{-1}(0))$.

Sea

$$\Phi_0(y, \mathbf{x}, \boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = w(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) \rho(t(y) - m(g^{-1}(\boldsymbol{\beta}^T \mathbf{x}))), \quad (1.21)$$

entonces, para todo $\mathbf{t} \in \mathbb{R}^p$,

$$\lim_{\gamma \rightarrow +\infty} \Phi_0(y, \mathbf{x}, \gamma \mathbf{t}, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) = \Phi_0^*(y, \mathbf{x}, \text{sign}(\mathbf{t}^T \mathbf{x})),$$

donde

$$\Phi_0^*(y, \mathbf{x}, j) = \begin{cases} w(\mathbf{x}, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) \rho(t(y) - m_1) & \text{if } j = -1 \\ w(\mathbf{x}, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) \rho(t(y) - m_3) & \text{if } j = 0 \\ w(\mathbf{x}, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) \rho(t(y) - m_2) & \text{if } j = 1 \end{cases}. \quad (1.22)$$

Teorema 1.4.1 *Sea (y_i, \mathbf{x}_i) , $i \in \mathbb{N}$ una sucesión de variables aleatorias i.i.d. que satisfacen (1.9) y (1.10). Supongamos que se cumplen **A1-A5** y **B1-B9** y sea $\widehat{\boldsymbol{\beta}}_n$ el estimador definido por (1.13). Sea*

$$\tau = \inf_{\mathbf{t} \in S} E_{\boldsymbol{\beta}_0}(\Phi_0^*(y, \mathbf{x}, \text{sign}(\mathbf{t}^T \mathbf{x}))) - E_{\boldsymbol{\beta}_0}(\Phi_0(y, \mathbf{x}, \boldsymbol{\beta}_0, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)) \quad (1.23)$$

con Φ_0 y Φ_0^* definidas por (1.21) y (1.22) respectivamente. Entonces (i) $\tau > 0$ y (ii) si $P(\mathbf{t}^T \mathbf{x} = 0) < \tau$ para todo $\mathbf{t} \in \mathbb{R}^p$ entonces $\widehat{\boldsymbol{\beta}}_n \rightarrow \boldsymbol{\beta}_0$ c.s..

Observemos que, por **B2** y **B8**, $\tau < 1$.

La demostración de este teorema se obtendrá como consecuencia de un resultado más general y se da en el Apéndice A.

1.5. Normalidad Asintótica

En esta sección, estudiamos la normalidad asintótica de los WMT-estimadores. Necesitamos asumir que se cumplen las siguientes condiciones, además de las condiciones **A1-A5** y **B1-B9**, necesarias para la consistencia.

C1 F_λ tiene tres derivadas continuas y acotadas como función de λ y la función de enlace g es dos veces continuamente diferenciable.

C2 ρ tiene tres derivadas continuas y acotadas. Notamos $\psi = \rho'$.

C3 Se cumple que $\sqrt{n}(\widehat{\boldsymbol{\mu}}_n - \boldsymbol{\mu}_0) = O_p(1)$ y $\sqrt{n}(\widehat{\boldsymbol{\Sigma}}_n - \boldsymbol{\Sigma}_0) = O_p(1)$, donde $\boldsymbol{\mu}_0$ y $\boldsymbol{\Sigma}_0$ están dados por **B7**.

C4 Existe $\varepsilon > 0$ tal que $|E_\lambda(\psi'(t(y) - m(\lambda)))| > \varepsilon$ para todo λ .

Sea $\boldsymbol{\Psi} = (\Psi_1, \dots, \Psi_p) : \mathbb{R} \times \mathbb{R}^p \times \mathbb{R}^p \times \mathbb{R}^p \times \mathbb{R}^{p \times p} \rightarrow \mathbb{R}^p$ definida por

$$\begin{aligned} \Psi_j(y, \mathbf{x}, \boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) &= w(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) \frac{\partial}{\partial \beta_j} \rho(t(y) - m(g^{-1}(\boldsymbol{\beta}^T \mathbf{x}))) \\ &= w(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) \psi\left(t(y) - m(g^{-1}(\boldsymbol{\beta}^T \mathbf{x}))\right) m'(g^{-1}(\boldsymbol{\beta}^T \mathbf{x})) g^{-1'}(\boldsymbol{\beta}^T \mathbf{x}) x_j. \end{aligned}$$

Notamos $\mathbf{J}_\Psi(y, \mathbf{x}, \boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = (J_\Psi^{j,k}(y, \mathbf{x}, \boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\Sigma}))_{1 \leq j, k \leq p}$ a la matriz Jacobiana de $\boldsymbol{\Psi}$ respecto de $\boldsymbol{\beta}$ es decir,

$$J_\Psi^{j,k}(y, \mathbf{x}, \boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{\partial}{\partial \beta_k} \Psi_j(y, \mathbf{x}, \boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\Sigma}), \quad 1 \leq j, k \leq p.$$

Observemos que las condiciones **C1**, **C2** y el Lema A.2.3 implican que $\boldsymbol{\Psi}$ y \mathbf{J}_Ψ están bien definidas.

C5 (i) Existe $\varepsilon > 0$ tal que

$$E \left(\sup_{\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| \leq \varepsilon} \left| J_\Psi^{j,k}(y, \mathbf{x}, \boldsymbol{\beta}, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) \right| \right) < \infty,$$

para todo $1 \leq j, k \leq p$, donde $\|\cdot\|$ denota la norma l_2 .

(ii) $E(\mathbf{J}_\Psi(y, \mathbf{x}, \boldsymbol{\beta}_0, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0))$ es no singular.

Una familia de funciones que verifica las condiciones **B1-B6** y **C2** es

$$\rho_k(u) = \begin{cases} 1 - (1 - (\frac{u}{k})^2)^4 & \text{if } |u| \leq k \\ 1 & \text{if } |u| \geq k \end{cases} \quad (1.24)$$

con $k > A^{1/2}$. Debe notarse que las funciones en la familia bicuadrada tienen una expresión similar. La única diferencia es que el exponente 3 se reemplaza aquí por un 4. La ventaja de la familia definida por (1.24) es que, mientras que las funciones pertenecientes a la familia bicuadrada tienen sólo dos derivadas en $u = \pm k$, éstas tienen tres, como pide la condición **C2**. La Figura 1.3 muestra un gráfico de ρ_k con $k = 2.9$.

Observemos que **C5** se satisface claramente cuando la función ω tiene soporte compacto o cuando \mathbf{x} toma valores en un conjunto compacto.

Consideremos el caso de regresión de Poisson con $\omega = 1$ y ψ con soporte compacto. Probaremos en la Proposición A.2.1, que se encuentra en el Apéndice

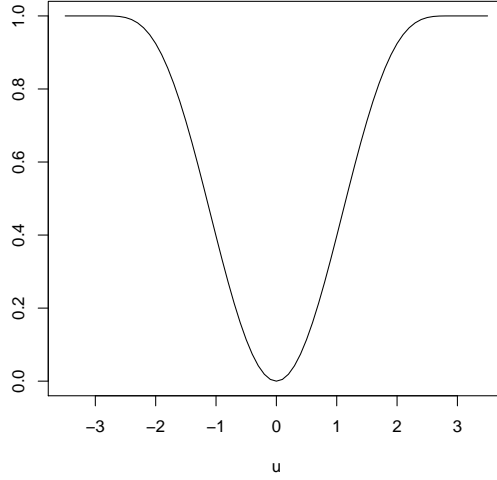


Figura 1.3: Gráfico de ρ_k con $k = 2,9$

A, que una condición suficiente para que se cumpla **C5** (i) es que $E(y|x_k x_j) < \infty$ para $1 \leq k, j \leq p$. En la misma proposición, se muestra que, para todo $\mathbf{a} \neq 0$,

$$\mathbf{a}^T E(\mathbf{J}_\Psi(y, \mathbf{x}, \beta_0, \mu_0, \Sigma_0)) \mathbf{a} > 0,$$

lo que implica que se cumple **C5** (ii).

El siguiente teorema da la distribución asintótica de los WMT-estimadores.

Teorema 1.5.1 *Supongamos que se cumplen **A1-A5**, **B1-B9**, **C1-C5** y que $P(\mathbf{t}^T \mathbf{x} = 0) < \tau$ para todo $\mathbf{t} \in S$, donde τ está dado por (1.23). Sea $(y_i, \mathbf{x}_i), i \in \mathbb{N}$, una sucesión de vectores aleatorios i.i.d. que satisfacen (1.9) y (1.10) y sea $\hat{\beta}_n$ definido por (1.13). Entonces,*

$$\sqrt{n}(\hat{\beta}_n - \beta_0) \xrightarrow{D} \mathcal{N}_p(0, \mathbf{B}^{-1} \mathbf{A} \mathbf{B}^{T-1}),$$

donde $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denota la distribución normal multivariada p -dimensional con media $\boldsymbol{\mu}$ y matriz de covarianza $\boldsymbol{\Sigma}$ y

$$\mathbf{A} = E_{\beta_0}(\Psi(y, \mathbf{x}, \beta_0, \mu_0, \Sigma_0) \Psi(y, \mathbf{x}, \beta_0, \mu_0, \Sigma_0)^T), \quad \mathbf{B} = E_{\beta_0}(\mathbf{J}_\Psi(y, \mathbf{x}, \beta_0, \mu_0, \Sigma_0)).$$

Para usar este resultado para hacer inferencia asintótica necesitamos estimar las matrices \mathbf{A} y \mathbf{B} . Para ello, sea

$$\hat{\mathbf{A}}_n = E_{H_n}(\Psi(y, \mathbf{x}, \hat{\beta}_n, \hat{\mu}_n, \hat{\Sigma}_n) \Psi(y, \mathbf{x}, \hat{\beta}_n, \hat{\mu}_n, \hat{\Sigma}_n)^T)$$

y

$$\widehat{\mathbf{B}}_n = E_{H_n}(\mathbf{J}_\psi(y, \mathbf{x}, \widehat{\boldsymbol{\beta}}_n, \widehat{\boldsymbol{\mu}}_n, \widehat{\boldsymbol{\Sigma}}_n)),$$

donde H_n es la distribución empírica de (y, \mathbf{x}) . Bajo las hipótesis del Teorema A.2.1, por el Lema A.2.1, $\widehat{\mathbf{A}}_n \rightarrow \mathbf{A}$ c.s. y $\widehat{\mathbf{B}}_n \rightarrow \mathbf{B}$ c.s.. Entonces, la matriz de covarianza asintótica $\mathbf{B}^{-1}\mathbf{A}\mathbf{B}^{T-1}$ puede ser estimada en forma consistente por $\widehat{\mathbf{B}}_n^{-1}\widehat{\mathbf{A}}_n\widehat{\mathbf{B}}_n^{T-1}$.

1.6. Punto de ruptura asintótico

En esta sección, estudiamos el punto de ruptura asintótico de los MT-estimadores, es decir de los estimadores definidos por (1.13) con $\omega = 1$. Dada una muestra $(y_1, \mathbf{x}_1), \dots, (y_n, \mathbf{x}_n)$, el MT- estimador $\widehat{\boldsymbol{\beta}}_n$ también puede escribirse como $\mathbf{T}(H_n)$ donde

$$\mathbf{T}(H) = \arg \min_{\boldsymbol{\beta} \in \mathbb{R}^p} E_H (\rho(t(y) - m(g^{-1}(\boldsymbol{\beta}^T \mathbf{x})))) \quad (1.25)$$

y H_n es la distribución empírica de la muestra.

Hablando informalmente, el punto de ruptura es la máxima proporción de datos atípicos que puede contener la muestra de manera tal que el estimador siga dando alguna información sobre $\boldsymbol{\beta}_0$. Para ello el porcentaje de outliers no debe ser capaz de hacer que el estimador tienda a infinito.

Definition 1.6.1 *Sea \mathcal{D} el conjunto de todas las distribuciones en $\mathbb{R} \times \mathbb{R}^p$. El punto de ruptura asintótico del funcional \mathbf{T} se define como*

$$\varepsilon^*(\mathbf{T}, H_0) = \sup_{\varepsilon} \{\varepsilon \in (0, 1) : \sup_{H^* \in \mathcal{D}} \{\|\mathbf{T}((1 - \varepsilon)H_0 + \varepsilon H^*)\| < \infty\}\}.$$

El siguiente teorema da una cota inferior para el punto de ruptura los MT-estimadores. Su demostración se da en la Sección A.3.

Teorema 1.6.1 *Sea (y, \mathbf{x}) un vector aleatorio con distribución H_0 tal que $P_{H_0}(\alpha^T \mathbf{x} = 0) = 0$ para todo $\alpha \in \mathbb{R}^p$. Supongamos $y|\mathbf{x} \sim F_{g^{-1}(\boldsymbol{\beta}_0^T \mathbf{x})}$. Sea*

$$\varepsilon_0 = \frac{E_{H_0}(\min(\rho(t(y) - m_1), \rho(t(y) - m_2))) - E_{H_0}(\rho(t(y) - m(g^{-1}(\boldsymbol{\beta}_0^T \mathbf{x}))))}{1 + E_{H_0}(\min(\rho(t(y) - m_1), \rho(t(y) - m_2))) - E_{H_0}(\rho(t(y) - m(g^{-1}(\boldsymbol{\beta}_0^T \mathbf{x}))))},$$

con m_1 y m_2 dadas por (1.19) y (1.20) respectivamente. Entonces el punto de ruptura asintótico del funcional \mathbf{T} definido por (1.25) en H_0 satisface

$$\varepsilon^*(\mathbf{T}, H_0) \geq \varepsilon_0.$$

A continuación calculamos ε_0 en algunos casos particulares.

1.6.1. Regresión de Poisson

En esta sección, calculamos ε_0 , la cota inferior del punto de ruptura asintótico de los MT-estimadores dada por el Teorema 1.6.1, para el caso de regresión de Poisson. Como esta cota depende de m_1 y m_2 usaremos el Lema A.3.2 que se encuentra en el Apéndice A. Este lema establece que, en el caso de regresión de Poisson, $m_1 = 0$ y $m_2 = +\infty$.

Se tiene entonces que

$$\min(\rho(t(y) - m_1), \rho(t(y) - m_2)) = \rho(t(y))$$

y luego

$$\varepsilon_0 = \frac{E_{H_0}(\rho(t(y)) - \rho(t(y) - m(g^{-1}(\beta_0^T \mathbf{x}))))}{1 + E_{H_0}(\rho(t(y)) - \rho(t(y) - m(g^{-1}(\beta_0^T \mathbf{x}))))}.$$

En este caso, si $m(g^{-1}(\beta_0^T \mathbf{x}))$ está cerca de cero también ε_0 estará cerca de cero. Para el modelo Poisson, esto ocurre si $\beta_0^T \mathbf{x}$ es negativo y tiene valor absoluto grande, es decir cuando $P(y = 0)$ es grande. Observemos que, en estos casos, una pequeña proporción de inliers iguales a cero puede hacer que la proporción de ceros sea igual a 0.5 y las observaciones buenas no nulas pueden ser erróneamente tomadas por outliers.

1.6.2. Regresión exponencial

Para hallar ε_0 en el caso de regresión exponencial usaremos el Lema A.3.3. Este lema se encuentra en el Apéndice A y dice que, en el caso de regresión exponencial $m_1 = -\infty$ y $m_2 = +\infty$.

La cota inferior para el punto de ruptura resulta ser entonces

$$\varepsilon_0 = \frac{E_{H_0}(1 - \rho(t(y) - m(g^{-1}(\beta_0^T \mathbf{x}))))}{1 + E_{H_0}(1 - \rho(t(y) - m(g^{-1}(\beta_0^T \mathbf{x}))))}.$$

En este caso, si tomamos como función de enlace a $g(u) = \log(u)$, se tiene que $m(\lambda) - \log(\lambda) = m(1)$. Para ver esto escribimos

$$\begin{aligned} E_\lambda(\rho(\log(y) - m(\lambda))) &= E_\lambda(\rho(\log(y) - \log(\lambda) - (m(\lambda) - \log(\lambda)))) \\ &= E_1(\rho(\log(z) - (m(\lambda) - \log(\lambda)))) . \end{aligned}$$

Por lo tanto, para todo x

$$E_1(\rho(\log(z) - (m(\lambda) - \log(\lambda)))) \leq E_1(\rho(\log(z) - (x - \log(\lambda)))) ,$$

y entonces $m(\lambda) - \log(\lambda) = m(1)$. Por lo tanto,

$$\varepsilon_0 = \frac{E(1 - \rho(t(y) - m(1)))}{1 + E(1 - \rho(t(y) - m(1)))},$$

donde y es una variable aleatoria con distribución $\mathcal{E}(1)$, y luego es independiente de β_0 . Por ejemplo si usamos un MT-estimador con ρ en la familia dada en (1.24) con $k = 6$, valor con el que se obtiene una eficiencia de 0.85, se tiene $\varepsilon_0 = 0.463$.

1.6.3. Ejemplo

Para ejemplificar, estudiaremos ahora el punto de ruptura en un modelo de regresión de Poisson donde $\mathbf{x} = (1, \mathbf{x}^*)$ y \mathbf{x}^* tiene distribución $N_{p-1}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Si escribimos $\beta_0 = (\beta_0, \beta_0^*)$ entonces la distribución de $\beta_0^T \mathbf{x}$ es $N(\mu, \sigma^2)$ donde $\mu = \beta_0^{*T} \boldsymbol{\mu} + \beta_0$ y $\sigma^2 = \beta_0^{*T} \boldsymbol{\Sigma} \beta_0^*$ y, por lo tanto, ε_0 depende sólo de μ y σ^2 . La cota inferior ε_0 depende de μ y σ^2 y en gran parte de $P(y = 0)$. Sean $\varepsilon_{\min}(p)$ y $\varepsilon_{\max}(p)$ definidos por

$$\varepsilon_{\min}(p) = \inf_{\mu, \sigma^2} \{ \varepsilon_0(\mu, \sigma^2) : P_{\mu, \sigma^2}(y = 0) = p \}$$

y

$$\varepsilon_{\max}(p) = \sup_{\mu, \sigma^2} \{ \varepsilon_0(\mu, \sigma^2) : P_{\mu, \sigma^2}(y = 0) = p \}.$$

La Figura 1.4 muestra las curvas $\varepsilon_{\min}(p)$ y $\varepsilon_{\max}(p)$. Estas curvas son cotas superiores e inferiores para ε_0 como función de $P(y = 0)$ cuando \mathbf{x}^* tiene una distribución normal multivariada.

1.7. Método computacional

1.7.1. Cálculo de $m(\lambda)$

Para obtener la función $m(\lambda)$ que interviene en la definición de los MT-estimadores observemos que si y tiene distribución $P(\lambda)$ entonces

$$y^{1/2} - \lambda^{1/2} \rightarrow_d N(0, 1/4)$$

cuando $\lambda \rightarrow \infty$. Esto se demuestra en el Apéndice A, en la proposición A.4.1. Luego, para valores grandes de λ , $E(\sqrt{y})$ está cerca de $\sqrt{\lambda}$ y podremos aproximar $m(\lambda)$ por $\sqrt{\lambda}$. Entonces procedemos como sigue: para $0 < \lambda \leq 3$ ajustamos un spline cúbico usando una grilla con paso 0.1. El valor de $m(\lambda)$ en cada punto de la grilla se calcula usando la función `optimize` en R. Para $\lambda > 3$, comprobamos

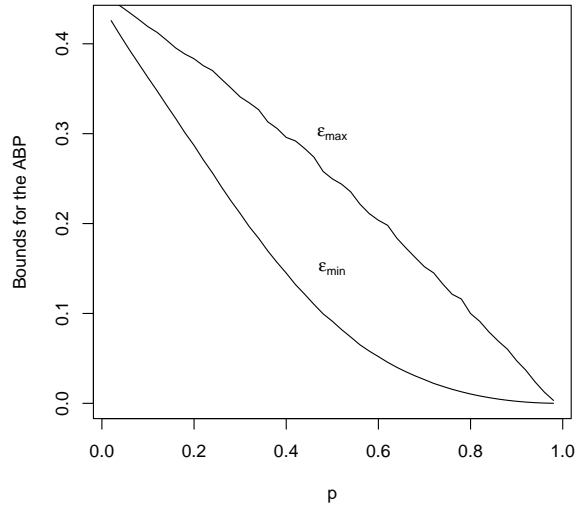


Figura 1.4: ε_{min} y ε_{max} como funciones de $p = P(y = 0)$

numéricamente que $m(\lambda)$ está muy cerca de $\sqrt{\lambda} y$, por lo tanto, una buena aproximación para $m(\lambda)$ se obtiene a usando un método de Newton-Raphson a un paso empezando en $\lambda^{1/2}$. Es decir se aproxima $m(\lambda)$ por

$$\sqrt{\lambda} + \frac{E_{\lambda}(\psi(\sqrt{y} - \sqrt{\lambda}))}{E_{\lambda}(\psi'(\sqrt{y} - \sqrt{\lambda}))}.$$

1.7.2. Cálculo de los WMT-estimadores

Para el cómputo de los WMT-estimadores, es necesario optimizar una función que puede tener múltiples puntos críticos. Esto presenta la dificultad de que, dependiendo del punto desde donde se comienza la iteración, se puede converger a diferentes mínimos locales. Es por eso que la elección de los valores iniciales de los parámetros es una parte crucial del cálculo del estimador. Si estos valores iniciales se encuentran muy lejos del mínimo absoluto, el método iterativo puede converger a un punto crítico distinto del mínimo absoluto que buscamos. Para evitar esto nos aseguramos de elegir como valor inicial de β un valor que esté lo más cerca posible del valor verdadero β_0 .

El valor inicial fue calculado por el método de submuestreo seguido de un paso de concentración similar al propuesto por Rousseeuw y Van Zomeren (1990). Este método consiste en calcular un conjunto finito de candidatos a mínimo de la función

objetivo y luego reemplazar la minimización sobre todo \mathbb{R}^p por una minimización sobre este conjunto. Más precisamente, dada una muestra $(\mathbf{x}_1, y_1) \dots (\mathbf{x}_n, y_n)$, con $y_i \in \mathbb{R}$ y $\mathbf{x}_i \in \mathbb{R}^p$, se quiere minimizar en $\boldsymbol{\beta}$ la función $L_n(\boldsymbol{\beta})$ definida en (1.14). Para ello se eligen al azar N submuestras de tamaño p . Para cada submuestra se calcula un primer candidato a mínimo absoluto calculando el estimador de máxima verosimilitud basado en esa submuestra. Si N es grande, es de esperar que haya varias submuestras que no contengan outliers y entonces este candidato a mínimo no se verá afectado por la contaminación. Para mejorar este candidato realizamos un paso de concentración. Para ello calculamos las n deviances correspondientes al estimador de máxima verosimilitud calculado en el paso anterior y luego calculamos el estimador de máxima verosimilitud basado en las $[n/2]$ observaciones con menor deviance. Finalmente evaluamos la función L_n en los N candidatos mejorados y elegimos como estimador inicial aquel en el que L_n alcanza el valor mínimo.

El número N de submuestras puede determinarse como en el caso de regresión lineal; ver, por ejemplo, el Capítulo 5 de Maronna, Martin y Yohai (2006). Este valor depende del número de covariables p , de la fracción esperada de outliers ε y de la probabilidad deseada de obtener al menos una submuestra libre de outliers. Supongamos que la muestra contiene una proporción ε de outliers. Entonces la probabilidad de que una submuestra no contenga outliers es $(1 - \varepsilon)^p$ y la probabilidad de que al menos una de las N submuestras no contenga outliers es $1 - (1 - (1 - \varepsilon)^p)^N$. Si queremos que esta probabilidad sea mayor a $1 - \alpha$ tenemos que elegir

$$N > \frac{\log(\alpha)}{\log(1 - (1 - \varepsilon)^p)}.$$

Por ejemplo para $\alpha = 0,01$, si $p = 6$ y queremos protegernos de hasta un 50 % de outliers habría que tomar al menos 293 submuestras. Este número asciende a 4714 si $p = 10$.

Para el cálculo del mínimo definido en (1.13) usamos el método de optimización L-BFGS-B que se encuentra en R dentro de las opciones de la función `optim`. El valor inicial es el estimador basado en submuestras descrito en esta sección.

1.7.3. Tiempos de cómputo

La Tabla 1.1 da los tiempos de cómputo expresados en minutos y segundos para MT-estimadores en una PC con un procesador AMD Athlon II X3 450 con una velocidad de 3.20 GHz y 8 GB de memoria RAM para N igual a 500, 1000 y 1500 y diferentes valores de p . En esa misma Tabla mostramos los correspondientes valores de ε cuando $1 - \alpha = 0.99$. Usamos un programa escrito completamente en R. Se observa que para $p = 50$, si la fracción esperada de outliers es menor que 0.1, el estimador puede computarse en un tiempo razonable, aún utilizando una

p	$N=500$		$N=1000$		$N=1500$	
	t	ϵ	t	ϵ	t	ϵ
10	0:51	0.37	1:29	0.42	2:06	0.44
20	1:21	0.21	2:02	0.24	2:35	0.25
30	1:42	0.15	2:29	0.16	3:10	0.17
40	2:14	0.11	2:59	0.13	3:55	0.14
50	2:42	0.09	3:46	0.10	4:32	0.11

Tabla 1.1: Tiempos de cómputo (t) y máximo porcentaje de outliers tolerado (ϵ) para diferentes valores de N y p

computadora no muy rápida. Estos tiempos pueden mejorarse si se usa una PC más poderosa y un código escrito parcialmente en C. Sin embargo, el valor de N aumenta exponencialmente con p y ϵ y, por lo tanto, este algoritmo puede volverse inaplicable cuando estos valores son grandes.

1.8. Estudio de Monte Carlo

Se realizó un estudio de Monte Carlo para comparar el comportamiento de los estimadores propuestos con el de los otros estimadores robustos existentes para regresión de Poisson y función de enlace $g(\lambda) = \log(\lambda)$.

Dado un estimador $\widehat{\beta}$, notamos ECM al error cuadrático medio definido por $E(\|\widehat{\beta} - \beta_0\|^2)$, donde $\|\cdot\|$ denota la norma l_2 . Estimamos el ECM por

$$\widehat{\text{ECM}} = \frac{1}{N} \sum_{i=1}^N \|\widehat{\beta}_i - \beta_0\|^2,$$

donde $\widehat{\beta}_i$ es el valor del estimador en la replicación i -ésima y N es el número de repeticiones que se tomó igual a 1000.

El vector de covariables es $\mathbf{x} = (1, \mathbf{x}^*)$ con $\mathbf{x}^* \sim N_5(0, \mathbf{I})$ e $y_i | \mathbf{x}_i \sim \mathcal{P}(e^{\beta_0^T \mathbf{x}_i})$. Consideramos tres modelos diferentes, con los siguientes valores de β_0 : $\beta_{0,1} = (0, 1, 0, 0, 0, 0)$, $\beta_{0,2} = (2, 1, 0, 0, 0, 0)$ y $\beta_{0,3} = (2, 1.5, 0, 0, 0, 0)$. Para cada uno de estos modelos simulamos el caso en que las muestras no contienen outliers y los casos en que las muestras contienen un 10 % de outliers idénticos de la forma (\mathbf{x}_0, y_0) . Consideramos el caso de outliers de bajo leverage con $\mathbf{x}_{0,1} = (1, 3, 0, 0, 0, 0)$ y de alto leverage con $\mathbf{x}_{0,2} = (1, 3, 4, 0, 0, 0)$. Los valores de y_0 se toman en una grilla de números enteros entre 0 y $\mu_0 + K$ donde $\mu_0 = e^{\beta_0^T \mathbf{x}_0} = E(y | \mathbf{x} = \mathbf{x}_0)$. El valor de K es distinto para los distintos modelos considerados y se elige de manera tal que llegue a observarse que el comportamiento de los estimadores se estabiliza. Cuando la contaminación se ubica en valores de y_0 pequeños, se observa que el

comportamiento de los estimadores es más variable; es por eso que la grilla es más fina cerca de cero y más gruesa lejos de cero. Para los valores de y_0 cercanos a $\boldsymbol{\mu}_0$ el ECM es pequeño para todos los estimadores considerados. Por razones de espacio estos valores no fueron incluidos en las tablas. En la Tabla 1.2, correspondiente al caso en que $\boldsymbol{\beta}_0 = \boldsymbol{\beta}_{0,1}$ y $\mathbf{x}_0 = \mathbf{x}_{0,1}$ y en el que $\boldsymbol{\mu}_0 = 20.06$, omitimos los valores de y_0 tales que $15 \leq y_0 \leq 26$. Para estos valores, el ECM resulta menor o igual a 0.09 para los siete estimadores considerados. En la Tabla 1.3, correspondiente al caso en que $\boldsymbol{\beta}_0 = \boldsymbol{\beta}_{0,2}$, $\mathbf{x}_0 = \mathbf{x}_{0,1}$ y $\boldsymbol{\mu}_0 = 148.41$, los valores de y_0 omitidos son $80 \leq y_0 \leq 240$. El ECM también resulta menor o igual a 0.09 para todos los estimadores considerados y estos valores de y_0 . En la Tabla 1.4, correspondiente al caso en que $\boldsymbol{\beta}_0 = \boldsymbol{\beta}_{0,3}$, $\mathbf{x}_0 = \mathbf{x}_{0,1}$ y $\boldsymbol{\mu}_0 = 665.13$, omitimos los valores de y_0 con $400 \leq y_0 \leq 1000$. El ECM resulta, en estos casos, menor o igual a 0.13 para los siete estimadores calculados. Por último, en la Tabla 1.5, correspondiente al caso en que $\boldsymbol{\beta}_0 = \boldsymbol{\beta}_{0,1}$, $\mathbf{x}_0 = \mathbf{x}_{0,2}$ y $\boldsymbol{\mu}_0 = 20.06$, los valores de y_0 omitidos son $10 \leq y_0 \leq 40$, para los cuales los ECM de los siete estimadores resultan menores o iguales a 0.11. Simulamos los siguientes estimadores: el estimador de máxima verosimilitud (ML), los estimadores de quasi verosimilitud robustos propuestos por Cantoni y Ronchetti (2001) sin pesos (QL) y con pesos (WQL), los estimadores condicionalmente insesgados de influencia acotada (CUBIF) propuestos por Künsch, Stefanski y Carroll (1989), los M-estimadores a un paso empezando en el estimador de proyección (PM) propuestos por Bergesio y Yohai (2011) y los MT- y WMT-estimadores. Los MT- y WMT-estimadores fueron calculados usando la ρ -función dada en (1.24). En el caso de los WMT-estimadores la función de peso usada es

$$\omega(t) = \begin{cases} 1 & \text{si } t^2 \leq \chi_{0.965,5}^2 \\ \frac{\chi_{0.975,5}^2 - t^2}{\chi_{0.975,5}^2 - \chi_{0.965,5}^2} & \text{si } \chi_{0.965,5}^2 < t^2 \leq \chi_{0.975,5}^2 \\ 0 & \text{si } t^2 > \chi_{0.975,5}^2 \end{cases}, \quad (1.26)$$

donde $\chi_{\alpha,p}^2$ es tal que $P(x \leq \chi_{\alpha,p}^2) = \alpha$ si x tiene una distribución chi cuadrado con p grados de libertad.

Es conocido que si $\mathbf{x} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ entonces $(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \sim \chi_p^2$. Por lo tanto, como $\mathbf{x}_i \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ para todo i y $\hat{\boldsymbol{\mu}}_n$ y $\hat{\boldsymbol{\Sigma}}_n$ son estimadores de $\boldsymbol{\mu}$ y $\boldsymbol{\Sigma}$ respectivamente entonces el efecto de esta función de peso es eliminar aproximadamente un 2,5% de las observaciones. Si hay outliers en el conjunto de las variables predictoras, es decir, puntos de alto leverage, y los estimadores $\hat{\boldsymbol{\mu}}_n$ y $\hat{\boldsymbol{\Sigma}}_n$ son robustos, esta función de peso les dará peso cero a estos outliers ya que la distancia de Mahalanobis $(\mathbf{x}_i - \hat{\boldsymbol{\mu}}_n)^T \hat{\boldsymbol{\Sigma}}_n^{-1} (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_n)$ será muy grande para esos valores de i . Los estimadores de posición $\hat{\boldsymbol{\mu}}_n$ y de dispersión $\hat{\boldsymbol{\Sigma}}_n$ usados para calcular los pesos son S-estimadores con punto de ruptura asintótico igual a 0.5 basados en la ρ -función bicuadrada. Estos S-estimadores fueron calculados usando la función `SestCov`, `method="bisquare"` que se encuentra en el paquete `rrcov` de R.

Las constantes de calibración de los estimadores fueron elegidas para obtener una eficiencia entre 0.8 y 0.9. Esto no siempre fue posible para los PM-estimadores ya que para algunos modelos la eficiencia permanecía baja aún para valores altos de la constante de calibración. Los valores de las constantes de calibración que usamos son 0.9 para los QL-estimadores, 2.8 para los WQL-estimadores, 2 para los PM-estimadores, 1.6 para los estimadores CUBIF, 2.3 para los MT-estimadores y 2.9 para los WMT-estimadores.

La Tabla 1.2 muestra las eficiencias con respecto al estimador ML cuando no hay outliers para los tres modelos. Las Tablas 1.3 a 1.6 y las Figuras 1.5 a 1.8 dan los valores de \widehat{ECM} en las muestras contaminadas para los tres modelos. En las Figuras 1.5 a 1.7 no mostramos el \widehat{ECM} de WQL ni de WTM ya que tienen comportamientos similares o peores a los de QL y MT respectivamente. En la Figura 1.8, en cambio, mostramos el \widehat{ECM} de WQL ni de WTM y no de sus análogos sin pesos ya que, en este caso, son las alternativas pesadas las que se comportan mejor.

Para resumir los resultados de este estudio podemos decir que todos los estimadores considerados excepto el PM tienen una eficiencia similar para los tres modelos considerados. El PM estimador tiene una eficiencia más baja, especialmente para $\beta_0 = \beta_{0,3}$. La Tabla 1.3 y la Figura 1.5 muestran que, cuando $\beta_0 = \beta_{0,1}$ y la contaminación se ubica en $\mathbf{x}_{0,1}$, el más robusto de los estimadores es el PM seguido del MT. Sin embargo, cuando $\beta_0 = \beta_{0,2}$ y $\mathbf{x}_0 = \mathbf{x}_{0,1}$, de acuerdo a la Tabla 1.4 y a la Figura 1.6 los estimadores más robustos son el PM y el MT, que tienen un comportamiento similar. En la Tabla 1.5 y la Figura 1.7 vemos que el PM, el MT y el CUBIF son los estimadores más robustos cuando $\beta_0 = \beta_{0,3}$ y $\mathbf{x}_0 = \mathbf{x}_{0,1}$. Por último, en la Tabla 1.6 y la Figura 1.8 vemos que, para $\beta_0 = \beta_{0,1}$ y $\mathbf{x}_0 = \mathbf{x}_{0,2}$ los estimadores más robustos son el WMT y el PM. Para los tres valores de β_0 considerados el MT-estimador tiene un buen comportamiento con y sin outliers, especialmente cuando los outliers son de bajo leverage. En el caso de outliers de alto leverage, el WMT es más robusto que el MT-estimador. El PM-estimador para $\beta_0 = \beta_{0,1}$ se comporta un poco mejor bajo contaminación con outliers pero su eficiencia para conjuntos de datos no contaminados puede ser mucho menor. Cuando los outliers se ubican $\mathbf{x}_0 = \mathbf{x}_{0,1}$, el QL y el MT-estimador se comportan mejor que el WQL y el WMT-estimador respectivamente, es decir que el agregado de los pesos empeora el comportamiento de los estimadores. Cuando la contaminación se ubica en $\mathbf{x}_0 = \mathbf{x}_{0,2}$ y $\beta_0 = \beta_{0,1}$, en cambio, el WQL y el WMT-estimador se comportan mejor que sus respectivas versiones sin pesos debido a que los outliers son de alto leverage.

	QL	WQL	CUB	PM	MT	WMT
$\beta_{0,1}$	0.88	0.77	0.86	0.75	0.74	0.74
$\beta_{0,2}$	0.88	0.78	0.88	0.77	0.86	0.80
$\beta_{0,3}$	0.89	0.78	0.82	0.47	0.89	0.82

Tabla 1.2: Eficiencias sin outliers con respecto al estimador ML

y_0	0	1	2	3	4	5	10	30	35	40	45
ML	0.58	0.47	0.39	0.32	0.27	0.23	0.10	0.08	0.12	0.17	0.22
QL	0.71	0.63	0.50	0.40	0.33	0.27	0.12	0.10	0.14	0.20	0.25
WQL	0.93	0.69	0.54	0.44	0.36	0.30	0.14	0.10	0.15	0.20	0.25
CUBIF	0.73	0.54	0.42	0.34	0.28	0.23	0.10	0.08	0.13	0.15	0.19
PM	0.31	0.31	0.28	0.27	0.23	0.20	0.11	0.09	0.12	0.15	0.17
MT	0.11	0.51	0.41	0.33	0.27	0.23	0.11	0.10	0.14	0.19	0.25
WMT	0.13	0.57	0.44	0.35	0.29	0.24	0.12	0.11	0.15	0.20	0.25

Tabla 1.3: ECM bajo contaminación para $\beta_0 = \beta_{0,1}$ y $\mathbf{x}_0 = (1, 3, 0, 0, 0, 0)$. Para $15 \leq y_0 \leq 26$, $ECM \leq 0.09$, $\mu_0 = e^{\beta_0^T \mathbf{x}_0} = E(y|\mathbf{x} = \mathbf{x}_0) = 20.06$

y_0	0	10	20	30	40	50	60	70	280	320	360	400
ML	0.54	0.41	0.31	0.24	0.19	0.14	0.11	0.08	0.10	0.15	0.20	0.26
QL	0.36	0.36	0.33	0.26	0.20	0.15	0.11	0.08	0.10	0.14	0.18	0.23
WQL	1.01	0.60	0.41	0.30	0.22	0.17	0.13	0.09	0.11	0.15	0.20	0.26
CUBIF	0.24	0.37	0.34	0.25	0.18	0.14	0.10	0.08	0.06	0.07	0.07	0.07
PM	0.04	0.04	0.05	0.05	0.05	0.04	0.04	0.04	0.02	0.02	0.02	0.02
MT	0.01	0.01	0.01	0.02	0.05	0.08	0.10	0.08	0.09	0.12	0.14	0.13
WMT	0.01	0.01	0.01	0.03	0.06	0.10	0.10	0.09	0.09	0.13	0.15	0.14

Tabla 1.4: ECM bajo contaminación para $\beta_0 = \beta_{0,2}$ y $\mathbf{x}_0 = (1, 3, 0, 0, 0, 0)$. Para $80 \leq y_0 \leq 240$, $ECM \leq 0.09$, $\mu_0 = e^{\beta_0^T \mathbf{x}_0} = E(y|\mathbf{x} = \mathbf{x}_0) = 148.41$

y_0	0	50	100	150	200	1200	1400	1600	1800
ML	1.35	0.90	0.64	0.47	0.34	0.17	0.28	0.41	0.55
QL	1.01	0.80	0.51	0.36	0.26	0.13	0.21	0.28	0.36
WQL	2.13	0.91	0.59	0.42	0.31	0.15	0.23	0.32	0.41
CUBIF	0.07	0.07	0.07	0.07	0.07	0.06	0.07	0.07	0.07
PM	0.11	0.11	0.11	0.11	0.11	0.08	0.08	0.08	0.08
MT	0.01	0.01	0.01	0.01	0.04	0.11	0.15	0.16	0.14
WMT	0.01	0.01	0.01	0.02	0.06	0.11	0.15	0.17	0.17

Tabla 1.5: ECM bajo contaminación para $\beta_0 = \beta_{0,3}$ y $\mathbf{x}_0 = (1, 3, 0, 0, 0, 0)$. Para $400 \leq y_0 \leq 1000$, $ECM \leq 0.13$, $\mu_0 = e^{\beta_0^T \mathbf{x}_0} = E(y|\mathbf{x} = \mathbf{x}_0) = 665.13$

y_0	0	1	2	3	4	5	60	80	100
ML	0.39	0.29	0.22	0.19	0.15	0.13	0.12	0.17	0.22
QL	0.54	0.34	0.25	0.20	0.16	0.14	0.14	0.20	0.25
WQL	0.39	0.29	0.23	0.18	0.16	0.14	0.15	0.20	0.25
CUBIF	0.50	0.32	0.24	0.19	0.16	0.13	0.13	0.17	0.21
PM	0.28	0.21	0.17	0.14	0.12	0.11	0.11	0.11	0.11
MT	0.62	0.39	0.27	0.21	0.18	0.15	0.16	0.22	0.27
WMT	0.27	0.17	0.14	0.12	0.11	0.10	0.10	0.12	0.15

Tabla 1.6: ECM bajo contaminación para $\beta_0 = \beta_{0,1}$ y $\mathbf{x}_0 = (1, 3, 4, 0, 0, 0)$. Para $10 \leq y_0 \leq 40$, $MSE \leq 0.11$, $\mu_0 = e^{\beta_0^T \mathbf{x}_0} = E(y|\mathbf{x} = \mathbf{x}_0) = 20.06$

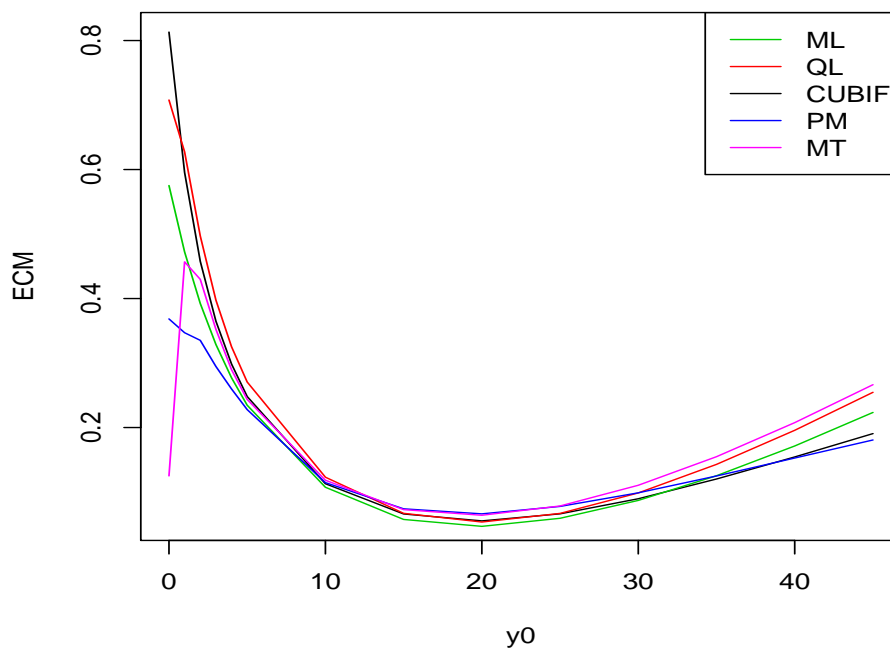


Figura 1.5: ECM bajo contaminación para $\beta_0 = \beta_{0,1}$, $\mathbf{x}_0 = \mathbf{x}_{0,1}$, $\mu_0 = 20.06$

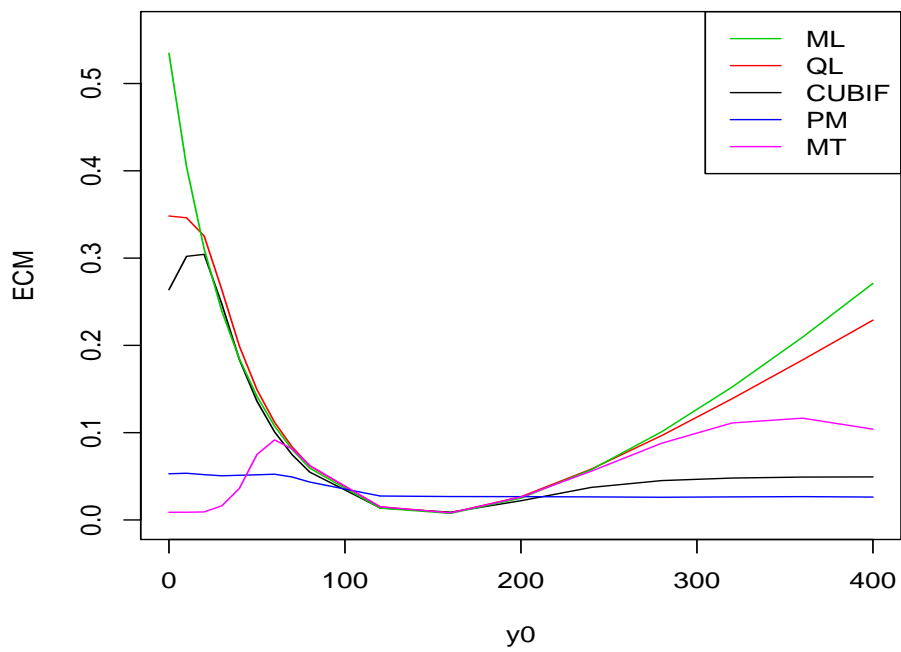


Figura 1.6: ECM bajo contaminación para $\beta_0 = \beta_{0,2}$, $\mathbf{x}_0 = \mathbf{x}_{0,1}$, $\mu_0 = 148.41$

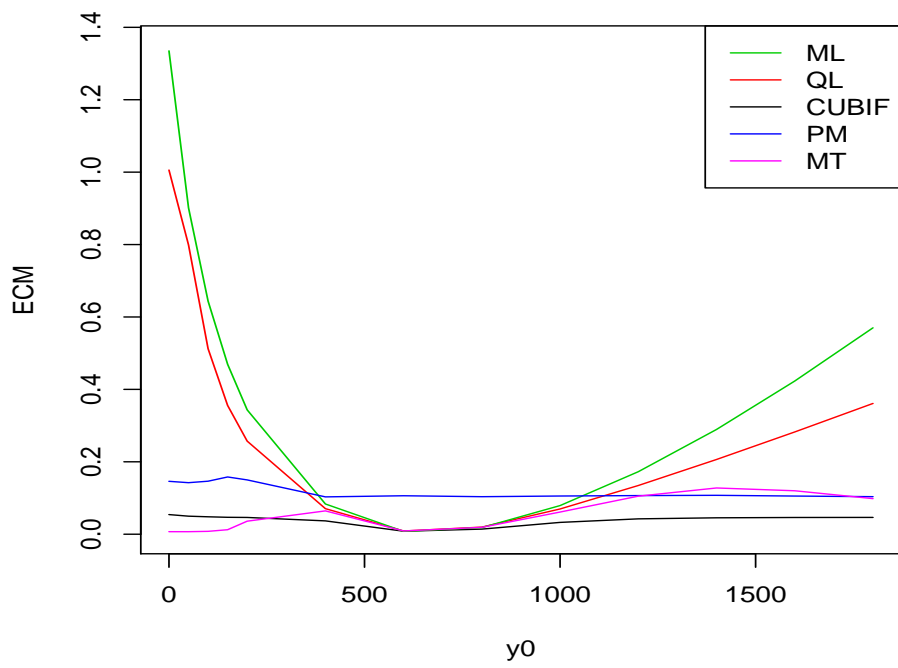


Figura 1.7: ECM bajo contaminación para $\beta_0 = \beta_{0,3}$, $\mathbf{x}_0 = \mathbf{x}_{0,1}$, $\mu_0 = 665.134$.

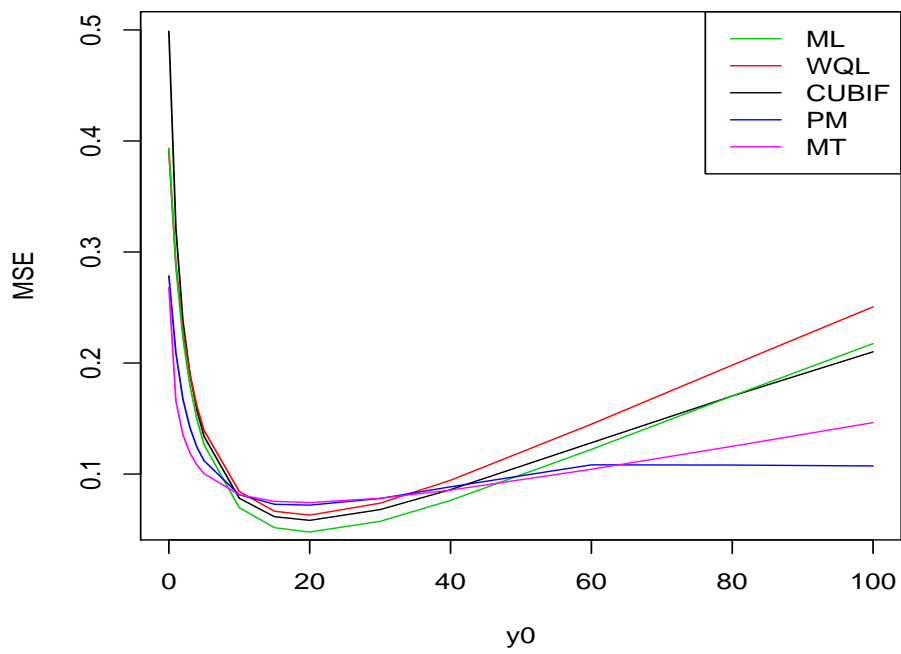


Figura 1.8: ECM bajo contaminación para $\beta_0 = \beta_{0,1}$, $\mathbf{x}_0 = \mathbf{x}_{0,2}, \mu_0 = 20.06$

1.9. Una aplicación a un conjunto de datos reales: datos de epilepsia

Breslow (1996) usó un modelo lineal generalizado con respuesta Poisson y log link para estudiar el efecto de drogas en pacientes epilépticos. Consideró datos de un ensayo clínico de 59 pacientes con epilepsia, 31 de los cuales fueron elegidos al azar para recibir la droga antiepiléptica Prograbide. Los restantes 28 recibieron un placebo. La variable de respuesta es SumY: el número de ataques durante cuatro semanas en un intervalo de tiempo dado. Las variables explicativas son Age 10: la edad del paciente dividida por diez, Base4: el número de ataques en las cuatro semanas previas al estudio, Trt: una variable dummy que toma los valores 1 o 0 si el paciente recibió la droga o un placebo respectivamente y Base4*Trt: para tomar en cuenta la interacción entre ambas variables.

Ajustamos un modelo lineal generalizado con respuesta Poisson y log link usando los mismos estimadores que se consideraron en el estudio de simulación. Para cada ajuste calculamos los residuos deviance, es decir

$$r_i = \left(2y_i \log \left(\frac{y_i}{e^{\hat{\beta}^T \mathbf{x}_i}} \right) - y_i + e^{\hat{\beta}^T \mathbf{x}_i} \right)^{1/2} \text{signo} \left(y_i - e^{\hat{\beta}^T \mathbf{x}_i} \right).$$

La Figura 1.9 muestra boxplots de los valores absolutos de los residuos deviance para los estimadores considerados. Los correspondientes a los estimadores WMT y WQL no fueron incluidos, ya que dan un ajuste un poco peor que sus análogos sin pesos. En el gráfico de la izquierda consideramos todos los residuos mientras que, para que las cajas y los bigotes se vean más claramente, en el gráfico de la derecha, eliminamos los outliers detectados por los boxplots en cada ajuste. Es claro que el MT-estimador da el mejor ajuste para la mayoría de los datos.

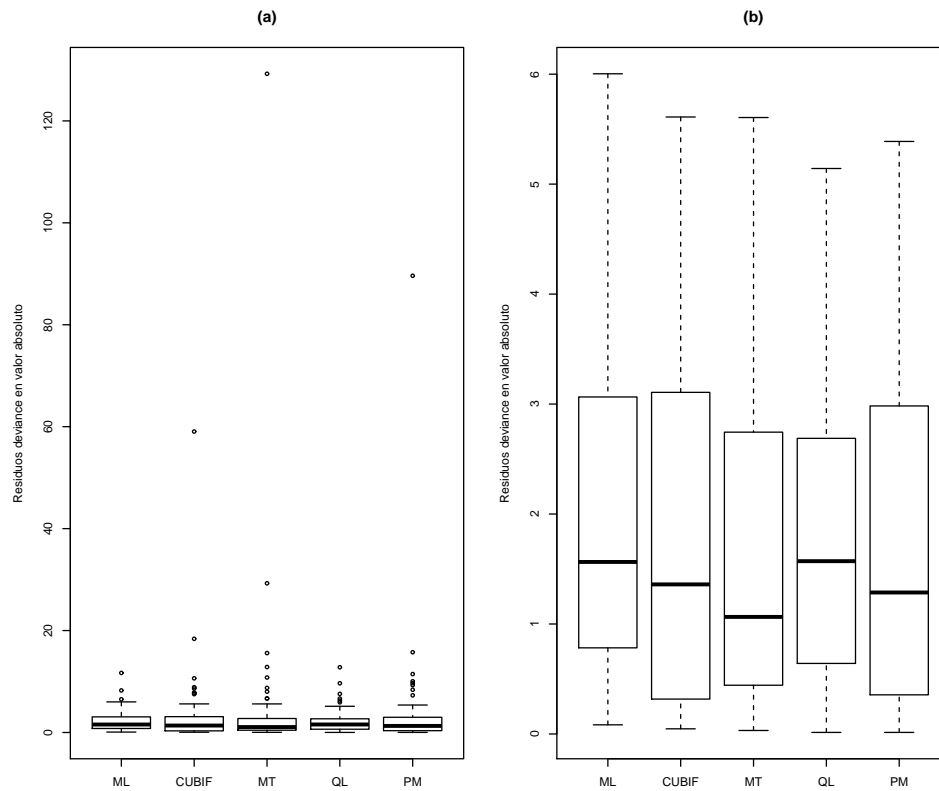


Figura 1.9: Boxplots de los valores absolutos de los residuos deviance para (a) todas las observaciones y (b) las observaciones sin los outliers.

Capítulo 2

Estimadores robustos para un parámetro unidimensional

2.1. Introducción

En este capítulo, introducimos una familia de estimadores robustos para familias de distribuciones de una variable aleatoria unidimensional dependientes de un solo parámetro basados en la transformación integral de probabilidad (MI-estimadores). En la Sección 2.2, damos la definición de los MI-estimadores. En la Sección 2.3, estudiamos su consistencia y en la Sección 2.4, su distribución asintótica. En la Sección 2.5, estudiamos dos medidas de robustez para los MI-estimadores, la función de influencia y el punto de ruptura asintótico. En la Sección 2.6, estudiamos el caso particular de la distribución de Poisson; calculamos la varianza asintótica, la función de influencia y el punto de ruptura asintótico de los MI-estimadores para este caso particular y comparamos la función de influencia de los MI-estimadores con la de los estimadores óptimos de Hampel. Por último, realizamos un estudio de Monte Carlo que compara el desempeño de estos estimadores con el de los estimadores óptimos de Hampel (1974) y el de los estimadores de cuasi-verosimilitud robustos propuestos por Cantoni y Ronchetti (2001). Las demostraciones de los resultados enunciados en este capítulo se encuentran en el Apéndice B y se citarán como B.1.1, B.1.2, etc.

2.2. Definición de M-estimadores basados en la transformación integral de probabilidad

En esta sección, definimos una familia de M-estimadores para un parámetro unidimensional basados en una transformación de los datos.

Consideramos una familia de distribuciones $F_\theta(x)$, $\theta \in \Theta$, sobre \mathbb{R} discreta o continua con probabilidad puntual o densidad $p_\theta(x)$. Los estimadores que se estudiarán son M-estimadores de la forma siguiente. Dada una muestra x_1, \dots, x_n , un M-estimador $\hat{\theta}_n$ está dado por la ecuación

$$\sum_{i=1}^n \psi(x_i, \hat{\theta}_n) = 0, \quad (2.1)$$

donde $\psi : \mathbb{R}^2 \rightarrow \mathbb{R}$. Sea \mathcal{F} un conjunto de funciones de distribución sobre \mathbb{R} y sea F_n la función de distribución empírica de la muestra. Luego también se tiene que

$$\hat{\theta}_n = T_\psi(F_n),$$

donde $T_\psi : \mathcal{F} \rightarrow \mathbb{R}$ es el funcional definido por

$$E_F(\psi, T_\psi(F)) = 0.$$

Asumimos que el funcional T_ψ está unívocamente definido, es decir, que existe una única raíz de la función $\theta \mapsto E_F(\psi(x, \theta))$. El funcional T_ψ se dice que tiene la propiedad de consistencia de Fisher si satisface

$$T_\psi(F_\theta) = \theta,$$

para todo $\theta \in \Theta$ o, lo que es equivalente en el caso de M-funcionales,

$$E_\theta(\psi(x, \theta)) = 0 \text{ y } E_\theta(\psi(x, \theta_1)) \neq 0 \text{ para todo } \theta_1 \neq \theta. \quad (2.2)$$

donde, por $E_\theta(\psi(x, \theta))$, denotamos la esperanza de $\psi(x, \theta)$ cuando x tiene distribución F_θ .

Los estimadores que se propondrán en este capítulo son M-estimadores basados en la transformación integral de probabilidad. Si x es una variable aleatoria continua con distribución F entonces $u = F(x)$ es una variable aleatoria con distribución uniforme en el intervalo $[0, 1]$. Esta transformación se conoce como transformación integral de probabilidad. Si, en cambio, x es una variable aleatoria discreta que toma valores en los enteros no negativos, la aplicación que la transforma en una variable aleatoria uniforme en $[0, 1]$ se obtiene por $u = F(x) - vp(x)$ donde v es una variable uniforme en $[0, 1]$, independiente de x , y p es la función de probabilidad puntual de x . La demostración de este hecho se puede encontrar, por ejemplo, en el Lema 1 de la Sección 5.3 de Ferguson (1967). De aquí resulta que, si x es una variable aleatoria continua, entonces

$$E_\theta \left(F_\theta(x) - \frac{1}{2} \right) = 0, \quad (2.3)$$

y, si x es una variable aleatoria discreta,

$$E_{\theta} \left(F_{\theta}(x) - \frac{1}{2}p_{\theta}(x) - \frac{1}{2} \right) = 0. \quad (2.4)$$

Estas ecuaciones motivan la idea de estudiar los siguientes M-estimadores. Dada x_1, \dots, x_n una muestra aleatoria definimos el M-estimador basado en la transformación integral de probabilidad (MI-estimador) para distribuciones discretas como el valor $\hat{\theta}_n$, solución de la ecuación de estimación

$$\sum_{i=1}^n F_{\hat{\theta}_n}(x_i) - \frac{1}{2}p_{\hat{\theta}_n}(x_i) - \frac{1}{2} = 0. \quad (2.5)$$

Si, en cambio, x_1, \dots, x_n es una muestra aleatoria proveniente de una distribución continua F_{θ} , $\theta \in \Theta$, se define el MI-estimador para distribuciones continuas como la solución, $\hat{\theta}_n$, de

$$\sum_{i=1}^n F_{\hat{\theta}_n}(x_i) - \frac{1}{2} = 0. \quad (2.6)$$

Por (2.4) y (2.3) estos estimadores resulta que tienen la propiedad de consistencia de Fisher. Además, como son M-estimadores con función de score acotada, resultarán robustos.

En lo que sigue notaremos $\psi_I(x, \theta)$ a la función de score de los MI-estimadores. Es decir que

$$\psi_I(x, \theta) = F_{\theta}(x) - \frac{1}{2}p_{\theta}(x) - \frac{1}{2}$$

si F_{θ} es una distribución discreta y

$$\psi_I(x, \theta) = F_{\theta}(x) - \frac{1}{2}$$

si F_{θ} es continua.

2.3. Consistencia

Sea x_1, \dots, x_n una sucesión de variables aleatorias con función de distribución F_{θ_0} y sea un M-estimador dado por (2.1). Supongamos que se verifican las siguientes condiciones.

A1 $E_{\theta}(\psi(x, \theta)) = 0$ para todo $\theta \in \Theta$.

A2 $\psi(x, \theta)$ es continua y monótona en θ . Además existe un intervalo $[a, b]$ tal que $P_{\theta}[a, b] > 0$ y tal que para, $x \in [a, b]$, la función $\psi(x, \theta)$ es estrictamente monótona en θ .

La condición **A2** asegura que la solución de la ecuación (2.1) es única y, por lo tanto, que el funcional T_ψ está unívocamente definido.

El siguiente teorema muestra que los M-estimadores que cumplen **A1** y **A2** son fuertemente consistentes.

Teorema 2.3.1 *Sea x_1, \dots, x_n una sucesión de variables aleatorias i.i.d. con función de distribución $F_{\theta_0}(x)$ y sea $\hat{\theta}_n$ definido por*

$$\sum_{i=1}^n \psi(y_i, \hat{\theta}_n) = 0. \quad (2.7)$$

*Entonces, si se cumplen **A1** y **A2**, $\hat{\theta}_n$ es fuertemente consistente.*

Para demostrar este teorema, es necesario observar que, por **A2**, la función $\theta \mapsto E_{\theta_0}(\psi(x, \theta))$ es estrictamente creciente y, por lo tanto, existe un único $\theta \in \Theta$ que es solución de la ecuación $E_{\theta_0}(\psi(x, \theta)) = 0$. El teorema se deduce entonces de los resultados en la Sección 3.2 de Huber (1981).

El siguiente teorema muestra que, bajo condiciones muy generales, ψ_I cumple las condiciones **A1** y **A2** y, por lo tanto, los MI-estimadores son fuertemente consistentes. Su demostración se encuentra en el apéndice B.

Teorema 2.3.2 *(a) ψ_I satisface siempre la condición **A1** (b) Una condición suficiente para que se satisfaga **A2** es que, si $\theta_1 < \theta_2$, entonces $F_{\theta_1}(x) \geq F_{\theta_2}(x)$ para todo x y, para algún x , se cumple que $F_{\theta_1}(x) > F_{\theta_2}(x)$.*

Nota. La familias exponenciales a un parámetro dadas en (1.1) con $l(\lambda)$ estrictamente creciente cumplen la condición (b). La demostración es la misma que la del Lema 3 de Bianco, García Ben y Yohai (2005) cambiando el parámetro λ por $\alpha = -l(\lambda)$.

2.4. Normalidad asintótica

Sea x_1, \dots, x_n una sucesión de variables aleatorias i.i.d. con función de distribución $F_{\theta_0}(x)$ y función de densidad o función de probabilidad puntual $p_{\theta_0}(x)$. Consideremos primero la normalidad asintótica de M-estimadores generales. Necesitamos asumir que, además de **A1** y **A2**, se cumplen las siguientes condiciones.

A3 $E_{\theta_0}((\psi(x, \theta))^2) < \infty$

A4 $\psi(x, \theta)$ es dos veces derivable respecto de θ y se cumple que

(a) Existe una función $K(x)$ con $E_{\theta_0}(K(x)) < \infty$ tal que, para todo $x \in \mathbb{R}$,

$$\left| \frac{\partial}{\partial \theta} \psi(x, \theta) \right| \leq K(x).$$

(b) Existe $\delta > 0$ tal que

$$E_{\theta_0} \left(\sup_{|\theta - \theta_0| < \delta} \left| \frac{\partial^2}{\partial \theta^2} \psi(x, \theta) \right| \right) < \infty.$$

El siguiente teorema da la distribución asintótica de los M-estimadores. Su demostración se encuentra, por ejemplo, en la Sección 10.3 de Maronna, Martín y Yohai (2006).

Teorema 2.4.1 *Sea $x_i, i \in \mathbb{N}$ una sucesión de variables aleatorias i.i.d. Sea $\hat{\theta}_n$ definido por (2.7) con ψ tal que se cumplen las condiciones **A1-A4**. Entonces*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{D} \mathcal{N}(0, v(\theta_0)),$$

donde

$$v(\theta) = \frac{E_{\theta}((\psi(x, \theta))^2)}{(E_{\theta}(\frac{\partial}{\partial \theta} \psi(x, \theta)))^2}.$$

Observemos que la expresión que define a $v(\theta)$ tiene sentido ya que, por **A2** se tiene

$$E_{\theta} \left(\frac{\partial}{\partial \theta} \psi(x, \theta) \right) \neq 0.$$

2.5. Medidas de robustez

2.5.1. Función de influencia

Una de las medidas de robustez más usadas es la curva de influencia. Propuesta por Hampel (1974), da una medida de cuánto varía un estimador al reemplazar una pequeña proporción ε de elementos de la muestra por outliers idénticos. Se define de la siguiente manera. Sea (x_1, \dots, x_n) una muestra aleatoria de una distribución F_{θ} , sea F_n la distribución empírica de la muestra y consideremos un estimador $\hat{\theta}_n$ de θ . Supongamos que el estimador puede representarse como un funcional de la

distribución empírica, es decir que $\widehat{\theta}_n = T(F_n)$. Entonces la función de influencia del funcional T , o del estimador $\widehat{\theta}_n$, es

$$IF(x_0, \theta, T) = \lim_{\varepsilon \rightarrow 0} \frac{T((1 - \varepsilon)F_\theta + \varepsilon\delta_{x_0}) - T(F_\theta)}{\varepsilon}.$$

Hampel también definió la sensibilidad a errores groseros (*GES*) como el supremo en x_0 de la función de influencia, es decir, viene dada por

$$\gamma^*(T, \theta) = \sup_x |IC(x, \theta, T)|.$$

En el caso de M-estimadores dados por (2.1) con ψ derivable, bajo condiciones generales, se tiene que

$$IF(x_0, \theta, T_\psi) = \frac{\psi(x_0, \theta)}{E_\theta(\frac{\partial}{\partial \theta} \psi(x, \theta))} \quad (2.8)$$

(ver Maronna, Martin y Yohai, 2006, Sección 3.1). Además, si ψ es acotada se tiene

$$\gamma^*(T_\psi, \theta) = \frac{\max_{x_0} |\psi(x_0, \theta)|}{E_\theta(\frac{\partial}{\partial \theta} \psi(x, \theta))}.$$

Estimadores óptimos de Hampel

Hampel definió M-estimadores óptimos de manera que existiese un balance entre robustez y eficiencia. Esto se obtiene buscando la función ψ tal que el funcional T_ψ tenga la propiedad de consistencia de Fisher, es decir,

$$E_\theta(\psi(x, \theta)) = 0 \quad (2.9)$$

y minimice la varianza asintótica $v_\psi(\theta)$ sujeto a que

$$\gamma^*(T_\psi, \theta) \leq G(\theta), \quad (2.10)$$

donde $G(\theta)$ es una función determinada.

Hampel (1974) demuestra que la función ψ que es solución de este problema viene dada por

$$\psi^*(x, \theta) = \psi_{k(\theta)}^H(\psi_0(y, \theta) - r(\theta)),$$

donde ψ_k^H es la función de score de Huber dada por (1.5), ψ_0 es la derivada de la log verosimilitud, es decir,

$$\psi_0(x, \theta) = \frac{\partial}{\partial \theta} \log(p_\theta(x))$$

ya k y r son funciones que resultan de las condiciones (2.9) y (2.10).

2.5.2. Punto de ruptura asintótico.

En el caso en que el espacio de parámetros es un intervalo de la forma (θ_1, θ_2) , donde θ_1 puede ser $-\infty$ y θ_2 puede ser $+\infty$, el punto de ruptura se define de manera similar a la Definición 1.6.1. La diferencia es que ahora, para que el estimador de alguna información sobre el parámetro θ , la proporción de outliers no debe ser capaz de hacer que el estimador tienda a ninguno de los bordes del intervalo.

Definition 2.5.1 *Sea \mathcal{D} el conjunto de todas las distribuciones en \mathbb{R} que tienen el mismo rango que F_θ y sea T un funcional asociado a un estimador de un parámetro unidimensional $\theta \in (\theta_1, \theta_2)$. El punto de ruptura asintótico del funcional T se define como*

$$\varepsilon^*(T, \theta) = \sup_{\varepsilon} \left\{ \varepsilon \in (0, 1) : \theta_1 < \inf_{H \in \mathcal{D}} \{T((1 - \varepsilon)F_\theta + \varepsilon H)\} \leq \sup_{H \in \mathcal{D}} \{T((1 - \varepsilon)F_\theta + \varepsilon H)\} < \theta_2 \right\}.$$

Llamamos punto de ruptura asintótico a θ_1 de T a

$$\sup_{\varepsilon} \left\{ \varepsilon \in (0, 1) : \inf_{H \in \mathcal{D}} \{T((1 - \varepsilon)F_\theta + \varepsilon H)\} > \theta_1 \right\}$$

y punto de ruptura asintótico a θ_2 de T a

$$\sup_{\varepsilon} \left\{ \varepsilon \in (0, 1) : \sup_{H \in \mathcal{D}} \{T((1 - \varepsilon)F_\theta + \varepsilon H)\} < \theta_2 \right\}.$$

El siguiente teorema da el punto de ruptura asintótico de un M-estimador definido por medio de una ψ continua, estrictamente monótona y acotada. Su demostración se encuentra en el Apéndice B.

Teorema 2.5.1 *Consideremos una familia de distribuciones sobre \mathbb{R} y T el funcional definido por*

$$E_\theta(\psi(x, T(F_\theta))) = 0. \quad (2.11)$$

*Supongamos que se cumplen **A1** y **A2**. Supongamos que $k_1 = \sup_{x, \theta} \psi(x, \theta)$ y $-k_2 = \inf_{x, \theta} \psi(x, \theta)$ son finitos. Sean*

$$\lim_{\theta \rightarrow \theta_1} \psi(x, \theta) = K_1(x), \quad - \lim_{\theta \rightarrow \theta_2} \psi(x, \theta) = K_2(x).$$

Entonces el punto de ruptura asintótico del estimador definido por el funcional T es mayor o igual a

$$\min \left\{ \frac{E_\theta K_1(x)}{E_\theta K_1(x) + k_2}, \frac{E_\theta K_2(x)}{E_\theta K_2(x) + k_1} \right\}.$$

En el caso de los MI-estimadores se tiene que

$$\frac{1}{2}p_{\theta}(x) - \frac{1}{2} \leq \psi_I(x, \theta) \leq F_{\theta}(x) - \frac{1}{2},$$

y, por lo tanto,

$$-\frac{1}{2} \leq \psi_I(x, \theta) \leq \frac{1}{2}.$$

Se tiene entonces que $K_1(x)$, $K_2(x)$, k_1 y k_2 son menores o iguales que $1/2$. También podemos asegurar que son mayores que cero ya que si no $\psi(x, \theta)$ no cambiaría de signo y no podría cumplirse la propiedad de Fisher-consistencia. Es fácil ver que en el caso de la distribución exponencial y en el caso de la distribución normal con varianza conocida se cumple que

$$\lim_{\theta \rightarrow \infty} F_{\theta}(x) = 1 \text{ y } \lim_{\theta \rightarrow 0} F_{\theta}(x) = 0,$$

y, por lo tanto,

$$K_1(x) = K_2(x) = k_1 = k_2 = \frac{1}{2}.$$

Esto implica que para estas distribuciones se alcanza el máximo punto de ruptura de $\varepsilon = 1/2$. En el caso de la distribución de Poisson, en cambio, $E_{\theta}(K_1(x))$ depende de θ y, por lo tanto, la cota dada por el Teorema 2.5.1 depende de θ . El caso particular de esta distribución será estudiado en detalle en la próxima sección.

2.6. Aplicación a la distribución de Poisson

2.6.1. Varianza asintótica

En el caso de la distribución de Poisson, por la ecuación (A.40) del Apéndice A, se tiene que

$$\frac{\partial}{\partial \theta} F_{\theta}(x) = \begin{cases} F_{\theta}(x-1) - F_{\theta}(x) & \text{si } x > 0 \\ -F_{\theta}(0) & \text{si } x = 0 \end{cases}.$$

De esto se deduce que

$$\frac{\partial}{\partial \theta} \psi_I(x, \theta) = \begin{cases} \psi_I(\theta, x-1) - \psi_I(x, \theta) & \text{si } x > 0 \\ -\psi_I(\theta, 0) - \frac{1}{2} & \text{si } x = 0 \end{cases}.$$

Observemos que esta función es acotada. Un calculo directo muestra que la derivada segunda de $\psi_I(x, \theta)$ respecto de θ también es acotada. Por lo tanto se cumple

la condición **A3**.

Se tiene

$$\begin{aligned}
E_{\theta} \left(\frac{\partial}{\partial \theta} \psi_I(x, \theta) \right) &= E_{\theta} \left((\psi_I(x-1, \theta) - \psi_I(x, \theta)) I(x > 0) + \left(-\psi_I(\theta, 0) - \frac{1}{2} \right) I(x = 0) \right) \\
&= E_{\theta} (\psi_I(x-1, \theta) I(x > 0)) - E_{\theta} (\psi_I(x, \theta)) - \frac{1}{2} p_{\theta}(0) \\
&= E_{\theta} (\psi_I(x-1, \theta) I(x > 0)) - \frac{1}{2} p_{\theta}(0).
\end{aligned}$$

Veamos que esta esperanza también puede escribirse de la forma

$$E_{\theta} \left(\frac{\partial}{\partial \theta} \psi_I(x, \theta) \right) = -\frac{1}{2} E_{\theta} (p_{\theta}(x)) - \frac{1}{2} E_{\theta} (p_{\theta}(x-1) I(x > 0)). \quad (2.12)$$

Se tiene

$$\begin{aligned}
E_{\theta} \left(\frac{\partial}{\partial \theta} \psi_I(x, \theta) \right) &= E_{\theta} (\psi_I(x-1, \theta) I(x > 0)) - \frac{1}{2} p_{\theta}(0) \\
&= E_{\theta} (F_{\theta}(x-1) I(x > 0)) - \frac{1}{2} E_{\theta} (p_{\theta}(x-1) I(x > 0)) \\
&\quad - \frac{1}{2} E_{\theta} (I(x > 0)) - \frac{1}{2} p_{\theta}(0) \\
&= E_{\theta} (F_{\theta}(x) - p_{\theta}(x)) - \frac{1}{2} E_{\theta} (p_{\theta}(x-1) I(x > 0)) - \frac{1}{2} \\
&= E_{\theta} \left(\frac{1}{2} p_{\theta}(x) + \frac{1}{2} - p_{\theta}(x) \right) - \frac{1}{2} E_{\theta} (p_{\theta}(x-1) I(x > 0)) - \frac{1}{2} \\
&= -\frac{1}{2} E_{\theta} (p_{\theta}(x)) - \frac{1}{2} E_{\theta} (p_{\theta}(x-1) I(x > 0)),
\end{aligned}$$

como queríamos demostrar.

Esto, junto con el Teorema 2.4.1, implica que la varianza asintótica de los MI-estimadores para el caso de la distribución de Poisson es

$$v(\theta) = \frac{4E \left(\left(F_{\theta}(x) - \frac{1}{2} p_{\theta}(x) - \frac{1}{2} \right)^2 \right)}{\left(E_{\theta} (p_{\theta}(x) + p_{\theta}(x-1) I(x > 0)) \right)^2}.$$

La eficiencia asintótica del MI-estimador es, por definición

$$eff_{MI}(\theta) = \frac{v_0(\theta)}{v(\theta)} = \frac{\theta \left(E_{\theta} (p_{\theta}(x) + p_{\theta}(x-1) I(x > 0)) \right)^2}{4E_{\theta} \left(\left(F_{\theta}(x) - \frac{1}{2} p_{\theta}(x) - \frac{1}{2} \right)^2 \right)},$$

donde $v_0(\theta)$ es la varianza asintótica del estimador de máxima verosimilitud, que en este caso es igual a θ . Calculamos $eff_{MI}(\theta)$ para varios valores de θ . Los resultados se muestran en el la figura 2.1.

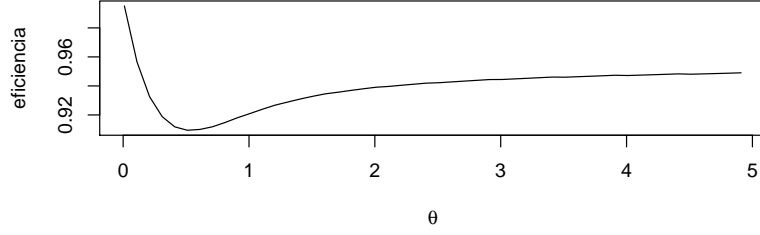


Figura 2.1: Eficiencia asintótica del MI-estimador para muestras provenientes de una distribución $\mathcal{P}(\theta)$ como función de θ

Se puede observar que la eficiencia asintótica del MI-estimador en el caso Poisson es muy alta para todos los valores de θ . Calculando $v(\theta)$ para una grilla de valores de θ pudimos comprobar que esta eficiencia es mayor que 0,90 para todo θ . Veamos además que tiende a 1 cuando $\theta \rightarrow 0$, como puede sospecharse a partir de la figura.

Por un lado, se tiene que

$$\begin{aligned} \lim_{\theta \rightarrow 0} (E_{\theta} (p_{\theta}(x) + p_{\theta}(x-1)I(x > 0)))^2 &= \lim_{\theta \rightarrow 0} p_{\theta}(0)^2 + E_{\theta} ((p_{\theta}(x) + p_{\theta}(x-1)) I(x > 0))^2 \\ &= 1. \end{aligned} \quad (2.13)$$

Además,

$$\begin{aligned} \frac{E_{\theta} \left(\left(F_{\theta}(x) - \frac{1}{2}p_{\theta}(x) - \frac{1}{2} \right)^2 \right)}{\theta} &= \frac{\left(\frac{1}{2}p_{\theta}(0) - \frac{1}{2} \right)^2 e^{-\theta}}{\theta} \\ &+ \left(p_{\theta}(0) + \frac{1}{2}p_{\theta}(1) - \frac{1}{2} \right)^2 e^{-\theta} \\ &+ \sum_{k=2}^{\infty} \left(F_{\theta}(k) - \frac{1}{2}p_{\theta}(k) - \frac{1}{2} \right)^2 \frac{e^{-\theta}\theta^{k-1}}{k!}. \end{aligned} \quad (2.14)$$

El tercer término del segundo miembro tiende a cero cuando θ tiende a cero mien-

tras que

$$\begin{aligned} \lim_{\theta \rightarrow 0} \left(\frac{1}{2} p_{\theta}(0) - \frac{1}{2} \right)^2 \frac{e^{-\theta}}{\theta} + \left(p_{\theta}(0) + \frac{1}{2} p_{\theta}(1) - \frac{1}{2} \right)^2 e^{-\theta} &= \lim_{\theta \rightarrow 0} \frac{(e^{-\theta} - 1)^2 e^{-\theta}}{4\theta} \\ &+ \frac{(2e^{-\theta} + e^{-\theta}\theta - 1)^2}{4} \\ &= \frac{1}{4}. \end{aligned} \quad (2.15)$$

De (2.13), (2.14) y (2.15) se obtiene

$$\lim_{\theta \rightarrow 0} \text{eff}_{MI}(\theta) = 1.$$

2.6.2. Función de influencia

Usando (2.8) y (2.12), se obtiene que la función de influencia tiene la forma

$$IF_{\hat{\theta}_H}(x_0, \theta_0) = \frac{\psi_{k(\theta_0)} \left(\frac{x_0}{\theta_0} - 1 - r(\theta_0) \right)}{E_{\theta_0} \left(\psi_{k(\theta_0)} \left(\frac{x+1}{\theta_0} - 1 - r(\theta_0) \right) \right)}$$

para los estimadores de Hampel y

$$IF_{\hat{\theta}_{MI}}(x_0, \theta_0) = -\frac{F_{\theta_0}(x_0) - \frac{1}{2} p_{\theta_0}(x_0) - \frac{1}{2}}{\frac{1}{2} E_{\theta_0} (p_{\theta_0}(x) + p_{\theta_0}(x-1) I(x > 0))}$$

para los MI-estimadores.

Ambas funciones de influencia son acotadas y se tiene

$$\gamma^*(\hat{\theta}_H, \theta_0) = \frac{k(\theta_0)}{E_{\theta_0} \left(\psi_{k(\theta_0)} \left(\frac{x+1}{\theta_0} - 1 - r(\theta_0) \right) \right)},$$

mientras que

$$\gamma^*(\hat{\theta}_{MI}, \theta_0) = \frac{1}{E_{\theta_0} (p_{\theta_0}(x) + p_{\theta_0}(x-1) I(x > 0))}.$$

Para comparar el MI-estimador con el estimador óptimo de Hampel calculamos, para cada θ , la varianza asintótica del MI-estimador, luego elegimos $k(\theta)$ y $r(\theta)$ para que el estimador de Hampel tenga esa misma varianza asintótica y se cumpla la condición de Fisher-consistencia. Luego comparamos curvas de influencia de

θ	0.1	0.5	1	2	3	4	5
MI	1.100	1.484	1.909	2.592	3.138	3.602	4.014
H	1.095	1.378	1.758	2.375	2.920	3.345	3.750

Tabla 2.1: Sensibilidad a errores groseros del MI-estimador y el estimador de Hampel para diferentes valores de θ

ambos estimadores para distintos valores de θ (ver Figura 2.2). En realidad la función de influencia puede en este caso considerarse con dominio en los números enteros no negativos. Si un conjunto de datos se supone que proviene de una distribución de Poisson, cualquier valor no entero o negativo sería inmediatamente detectado como erróneo y eliminado o corregido. Sin embargo, graficamos ambas funciones como curvas continuas para poder visualizarlas mejor. Por la definición del estimador de Hampel se obtendrá que el máximo de la función de influencia es menor para dicho estimador que para el MI-estimador. Sin embargo, observamos que ambas curvas de influencia son muy similares. A continuación calculamos la GES para diferentes valores de θ . Los resultados se muestran en la Tabla 2.1 y la Figura 2.3.

2.6.3. Punto de ruptura

En el caso de la distribución de Poisson, se tiene que $\theta_1 = 0$ y $\theta_2 = +\infty$. Para todo $x \in \mathbb{N}_0$, se tiene que $\lim_{\theta \rightarrow +\infty} p_\theta(x) = 0$ y, por lo tanto,

$$K_2(x) = \lim_{\theta \rightarrow +\infty} F_\theta(x) - \frac{1}{2}p_\theta(x) - \frac{1}{2} = -\frac{1}{2}.$$

Por otro lado,

$$\lim_{\theta \rightarrow 0} p_\theta(x) = \begin{cases} 0 & \text{si } x > 0 \\ 1 & \text{si } x = 0 \end{cases},$$

de donde

$$K_1(x) = -\lim_{\theta \rightarrow 0} F_\theta(x) - \frac{1}{2}p_\theta(x) - \frac{1}{2} = \begin{cases} \frac{1}{2} & \text{si } x > 0 \\ 0 & \text{si } x = 0 \end{cases}.$$

Se tiene entonces que $E_\theta(K_2(x)) = 1/2$ y $E_\theta(K_1(x)) = 1/2(1 - P_\theta(x = 0)) = 1/2(1 - e^{-\theta})$. Por lo tanto, se deduce del Teorema 2.5.1 que el punto de ruptura es mayor o igual que

$$\varepsilon_\theta = \frac{1 - e^{-\theta}}{2 - e^{-\theta}}.$$

En realidad lo que se ve es que el punto de ruptura a ∞ es $1/2$ mientras que el punto de ruptura a 0 es mayor o igual a ε_θ . Observemos que $\varepsilon_\theta \rightarrow \frac{1}{2}$ cuando

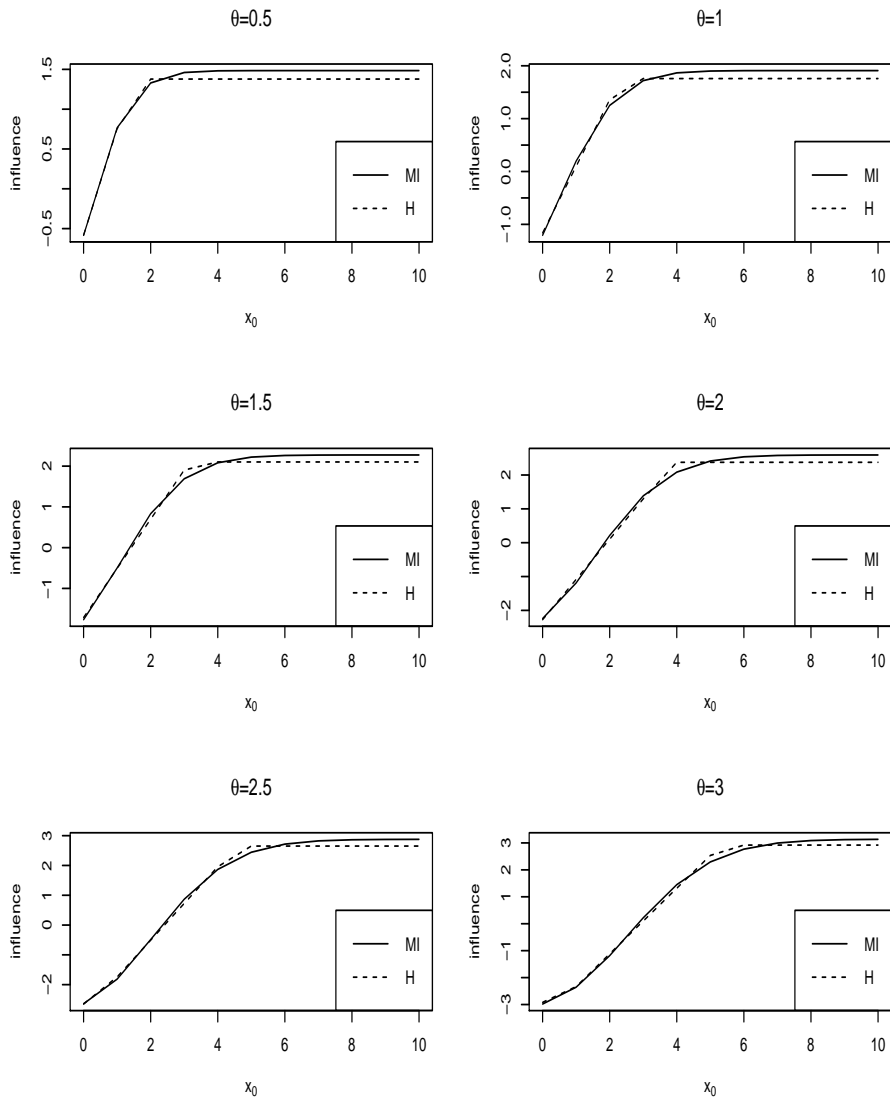


Figura 2.2: Funciones de influencia del estimador de Hampel y del MI-estimador para diferentes valores de θ

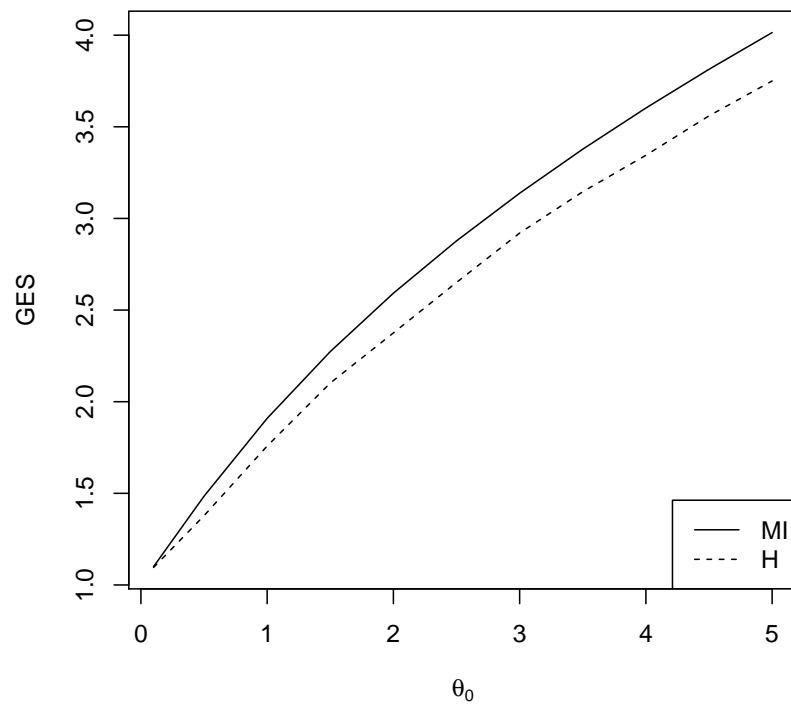


Figura 2.3: Sensibilidad a errores groseros para diferentes valores de θ .

$\theta \rightarrow \infty$ pero, lamentablemente, $\varepsilon_0 \rightarrow 0$ cuando $\theta \rightarrow 0$. Esto significa que, para valores pequeños de θ , esta cota no es informativa. Sin embargo, veremos que, en el caso de la distribución de Poisson esta cota inferior está lejos de ser óptima, ya que el verdadero punto de ruptura a cero de los MI-estimadores es 1. Esto también sucede con el estimador de máxima verosimilitud, aunque este último tiene punto de ruptura igual a cero en $+\infty$. En particular, se tiene que los MI-estimadores sólo pueden dar cero si todas las observaciones son iguales a cero. El siguiente teorema afirma que, en el caso de la distribución de Poisson, el punto de ruptura a cero del MI estimador es uno. Su demostración se encuentra en el apéndice B.

Teorema 2.6.1 *Sea T el funcional correspondiente al MI-estimador y $F_n = (1 - \varepsilon)F_\theta + \varepsilon H_n$ donde H_n es tal que $P_{H_n}(\mathbb{N}_0) = 1$. Si $T(F_n) \rightarrow 0$ entonces $\varepsilon = 1$.*

2.6.4. Estudio de Monte Carlo

Realizamos un estudio de Monte Carlo para comparar los estimadores propuestos con otros estimadores existentes. Consideramos muestras provenientes de la distribución de Poisson con media θ para cinco valores diferentes de θ : 0.1, 0.5, 1, 2, 3. Para cada valor de θ considerado, calculamos la varianza asintótica del MI-estimador. Luego calculamos $\hat{\theta}_0$, el MT-estimador introducido en el Capítulo 1. El valor de este estimador se toma como estimador inicial para calcular el k y r , de manera que se eligen k y r tales que el estimador de Hampel tenga la misma varianza asintótica que el MI-estimador cuando el verdadero valor del parámetro es $\hat{\theta}_0$ y tales que que el estimador de Hampel sea Fisher consistente. Consideramos muestras de tamaño 100 sin contaminar y muestras contaminadas con un 5%, un 10% y un 20% de outliers idénticos. Los valores considerados para estos outliers son números enteros x_0 entre 0 y 10. Calculamos los siguientes estimadores: el estimador de máxima verosimilitud (MV), el QL-estimador propuesto en Cantoni y Ronchetti (2001), el MT-estimador (MT), el estimador óptimo de Hampel (H) y el MI-estimador aquí propuesto (MI). Dado un estimador $\hat{\theta}_n$ de θ notamos ECM al error cuadrático medio definido por $E\left((\hat{\theta}_n - \theta)^2\right)$. Lo estimamos como

$$\widehat{\text{ECM}} = \frac{1}{n} \sum_{i=1}^n (\hat{\theta}_n - \theta)^2.$$

En la Tabla 2.2 se muestran las eficiencias de los estimadores considerados con respecto al estimador de máxima verosimilitud. Como era de esperar por la forma en que elegimos $k(\theta)$, estas eficiencias son casi idénticas. Las Tablas 2.3 a 2.7 dan el $\widehat{\text{ECM}}$ en muestras contaminadas con un 5% de outliers, las Tablas 2.8 a 2.12

	0.1	0.5	1	2	3
MI	0.96	0.90	0.93	0.94	0.95
MT	0.90	0.73	0.81	0.86	0.83
H	0.96	0.91	0.93	0.94	0.95
QL	0.94	0.91	0.93	0.94	0.95

Tabla 2.2: Eficiencias con respecto al estimador de máxima verosimilitud

dan el $\widehat{\text{ECM}}$ en muestras contaminadas con un 10 % de outliers y las Tablas 2.13 a 2.17 dan el $\widehat{\text{ECM}}$ en muestras contaminadas con un 20 % de outliers. Las Figuras 2.4, 2.5 y 2.6 muestran el máximo error cuadrático medio como función de θ para los estimadores MI, QL y H en muestras contaminadas con un 5 %, un 10 % y un 20 % de outliers respectivamente. El ECM del estimador de máxima verosimilitud no fue incluido en estos gráficos por razones de escala.

De los resultados de este estudio podemos concluir que, para los modelos considerados, los MI-estimadores tienen una eficiencia muy alta cuando todas las observaciones siguen el modelo y son también muy robustos, resultando poco influenciados por un porcentaje de observaciones atípicas de hasta 20 %. En las tablas se puede observar que, para los estimadores MI, H y QL, definidos a partir de una función de score monótona, el ECM aumenta al aumentar el valor de x_0 convergiendo, cuando $x_0 \rightarrow +\infty$, a un valor finito. En cambio, para el MT-estimador, definido a partir de una función de score redescendiente, el ECM disminuye a partir de un cierto valor de x_0 y toma valores muy pequeños para valores grandes de x_0 . Por lo tanto, si bien el máximo valor del ECM correspondiente al MT-estimador, no es muy diferente del del resto de los estimadores robustos, tiene la ventaja de que éste máximo ocurre en unos pocos valores de x_0 . De esta manera, resulta que los outliers grandes tienen una influencia mucho menor en el MT-estimador que en el resto de los estimadores.

Podemos decir que el QL-estimador es el que obtiene un mejor balance entre robustez y eficiencia en los modelos considerados. Sin embargo, la ventaja del MI-estimador es que no requiere calcular ninguna corrección para que sea Fisher consistente y, por lo tanto, es más simple de computar.

x_0	0	1	2	3	4	5	6	7	8	9	10
MV	0.10	0.29	0.98	2.22	3.87	6.06	8.75	11.9	15.6	19.8	24.5
MI	0.10	0.36	0.46	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47
MT	0.11	0.40	0.27	0.16	0.13	0.12	0.12	0.12	0.12	0.12	0.12
H	0.10	0.38	0.38	0.38	0.38	0.38	0.38	0.38	0.38	0.38	0.38
QL	0.10	0.37	0.38	0.38	0.38	0.38	0.38	0.38	0.38	0.38	0.38

Tabla 2.3: Error cuadrático medio multiplicado por 10^2 bajo un 5% de contaminación para $\theta = 0.1$.

x_0	0	1	2	3	4	5	6	7	8	9	10
MV	0.058	0.057	0.10	0.20	0.35	0.55	0.80	1.09	1.44	1.84	2.29
MI	0.064	0.072	0.11	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13
MT	0.077	0.099	0.11	0.11	0.093	0.085	0.080	0.080	0.078	0.078	0.078
H	0.063	0.071	0.12	0.12	0.12	0.12	0.13	0.13	0.13	0.13	0.13
QL	0.065	0.067	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12

Tabla 2.4: Error cuadrático medio multiplicado por 10 bajo un 5% de contaminación para $\theta = 0.5$

x_0	0	1	2	3	4	5	6	7	8	9	10
MV	0.12	0.098	0.12	0.20	0.32	0.50	0.72	0.99	1.32	1.69	2.12
MI	0.14	0.10	0.15	0.20	0.22	0.23	0.23	0.23	0.23	0.23	0.23
MT	0.17	0.12	0.17	0.21	0.21	0.19	0.17	0.15	0.14	0.14	0.13
H	0.13	0.11	0.15	0.21	0.21	0.22	0.22	0.22	0.22	0.22	0.22
QL	0.14	0.10	0.15	0.22	0.22	0.22	0.22	0.22	0.22	0.22	0.22

Tabla 2.5: Error cuadrático medio multiplicado por 10 bajo un 5% de contaminación para $\theta = 1$

x_0	0	1	2	3	4	5	6	7	8	9	10
MV	0.31	0.23	0.20	0.22	0.29	0.41	0.58	0.80	1.07	1.40	1.76
MI	0.35	0.24	0.21	0.26	0.33	0.40	0.42	0.43	0.43	0.43	0.43
MT	0.35	0.27	0.22	0.29	0.39	0.46	0.46	0.43	0.37	0.33	0.30
H	0.35	0.24	0.21	0.25	0.36	0.40	0.40	0.40	0.40	0.40	0.40
QL	0.35	0.24	0.21	0.24	0.36	0.41	0.41	0.41	0.41	0.41	0.41

Tabla 2.6: Error cuadrático medio multiplicado por 10 bajo un 5% de contaminación para $\theta = 2$

x_0	0	1	2	3	4	5	6	7	8	9	10
MV	0.53	0.41	0.33	0.30	0.32	0.39	0.51	0.69	0.91	1.18	1.50
MI	0.57	0.46	0.34	0.31	0.36	0.46	0.54	0.60	0.62	0.64	0.64
MT	0.46	0.58	0.40	0.35	0.42	0.54	0.65	0.69	0.67	0.62	0.55
H	0.56	0.45	0.34	0.31	0.35	0.46	0.59	0.60	0.60	0.60	0.60
QL	0.55	0.45	0.34	0.31	0.35	0.46	0.60	0.61	0.61	0.61	0.61

Tabla 2.7: Error cuadrático medio multiplicado por 10 bajo un 5% de contaminación para $\theta = 3$

x_0	0	1	2	3	4	5	6	7	8	9	10
MV	0.010	0.089	0.37	0.85	1.52	2.40	3.48	4.76	6.24	7.91	9.79
MI	0.011	0.11	0.17	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18
MT	0.011	0.13	0.082	0.030	0.015	0.013	0.013	0.013	0.013	0.013	0.013
H	0.011	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12
QL	0.011	0.12	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13

Tabla 2.8: Error cuadrático medio multiplicado por 10 bajo un 10% de contaminación para $\theta = 0.1$

x_0	0	1	2	3	4	5	6	7	8	9	10
MV	0.075	0.071	0.27	0.66	1.26	2.06	3.05	4.25	5.64	7.24	9.04
MI	0.084	0.11	0.27	0.35	0.37	0.38	0.38	0.38	0.38	0.38	0.38
MT	0.10	0.17	0.23	0.19	0.13	0.10	0.087	0.082	0.081	0.081	0.081
H	0.083	0.11	0.32	0.34	0.35	0.37	0.38	0.38	0.38	0.38	0.38
QL	0.087	0.098	0.32	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33

Tabla 2.9: Error cuadrático medio multiplicado por 10 bajo un 10% de contaminación para $\theta = 0.5$

x_0	0	1	2	3	4	5	6	7	8	9	10
MV	0.20	0.093	0.19	0.49	0.99	1.68	2.58	3.68	4.98	6.48	8.17
MI	0.24	0.098	0.27	0.48	0.59	0.63	0.64	0.64	0.64	0.64	0.64
MT	0.31	0.12	0.30	0.45	0.46	0.37	0.26	0.20	0.16	0.15	0.14
H	0.22	0.10	0.28	0.54	0.55	0.56	0.58	0.59	0.60	0.61	0.61
QL	0.23	0.098	0.26	0.57	0.57	0.57	0.57	0.57	0.57	0.57	0.57

Tabla 2.10: Error cuadrático medio multiplicado por 10 bajo un 10% de contaminación para $\theta = 1$

x_0	0	1	2	3	4	5	6	7	8	9	10
MV	0.61	0.30	0.19	0.28	0.57	1.06	1.75	2.64	3.73	5.02	6.51
MI	0.75	0.33	0.19	0.38	0.69	0.93	1.06	1.12	1.14	1.15	1.15
MT	0.69	0.39	0.19	0.44	0.84	1.14	1.19	1.02	0.76	0.53	0.39
H	0.73	0.31	0.19	0.34	0.76	1.03	1.03	1.04	1.05	1.07	1.09
QL	0.74	0.31	0.19	0.34	0.77	1.04	1.04	1.04	1.04	1.04	1.04

Tabla 2.11: Error cuadrático medio multiplicado por 10 bajo un 10% de contaminación para $\theta = 2$

x_0	0	1	2	3	4	5	6	7	8	9	10
MV	0.12	0.069	0.039	0.028	0.038	0.067	0.12	0.19	0.27	0.38	0.51
MI	0.13	0.087	0.041	0.028	0.049	0.087	0.12	0.15	0.16	0.16	0.17
MT	0.064	0.12	0.051	0.031	0.058	0.10	0.15	0.17	0.16	0.14	0.10
H	0.13	0.082	0.041	0.029	0.045	0.088	0.14	0.15	0.15	0.15	0.15
QL	0.12	0.084	0.041	0.029	0.044	0.087	0.15	0.15	0.15	0.15	0.15

Tabla 2.12: Error cuadrático medio bajo un 10% de contaminación para $\theta = 3$

x_0	0	1	2	3	4	5	6	7	8	9	10
MV	0.012	0.33	1.45	3.37	6.08	9.602	13.9	19.03	25.0	31.7	39.2
MI	0.012	0.45	0.87	0.10	1.02	1.02	1.02	1.02	1.02	1.02	1.02
MT	0.013	0.55	0.40	0.112	0.023	0.014	0.014	0.014	0.014	0.014	0.014
H	0.013	0.51	0.51	0.51	0.51	0.51	0.51	0.51	0.51	0.51	0.51
QL	0.013	0.47	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98

Tabla 2.13: Error cuadrático medio multiplicado por 10 bajo un 20% de contaminación para $\theta = 0.1$

x_0	0	1	2	3	4	5	6	7	8	9	10
MV	0.014	0.014	0.093	0.25	0.49	0.81	1.21	1.69	2.25	2.89	3.61
MI	0.017	0.027	0.10	0.16	0.18	0.19	0.19	0.19	0.19	0.19	0.19
MT	0.020	0.046	0.083	0.073	0.038	0.017	0.011	0.009	0.009	0.009	0.009
H	0.017	0.025	0.13	0.18	0.23	0.32	0.39	0.40	0.40	0.40	0.40
QL	0.018	0.022	0.13	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16

Tabla 2.14: Error cuadrático medio bajo un 20% de contaminación para $\theta = 0.5$

x_0	0	1	2	3	4	5	6	7	8	9	10
MV	0.048	0.008	0.048	0.17	0.37	0.65	1.01	1.45	1.97	2.57	3.25
MI	0.065	0.009	0.076	0.18	0.25	0.29	0.30	0.30	0.30	0.31	0.31
MT	0.089	0.014	0.084	0.18	0.23	0.19	0.100	0.043	0.023	0.017	0.016
H	0.060	0.009	0.075	0.22	0.23	0.24	0.28	0.33	0.36	0.37	0.37
QL	0.061	0.009	0.069	0.22	0.26	0.26	0.26	0.26	0.26	0.26	0.26

Tabla 2.15: Error cuadrático medio bajo un 20% de contaminación para $\theta = 1$

x_0	0	1	2	3	4	5	6	7	8	9	10
MV	0.18	0.057	0.016	0.055	0.17	0.37	0.65	1.01	1.45	1.97	2.57
MI	0.25	0.067	0.017	0.090	0.22	0.36	0.45	0.50	0.52	0.53	0.53
MT	0.27	0.086	0.016	0.11	0.28	0.46	0.57	0.53	0.38	0.20	0.092
H	0.23	0.063	0.016	0.079	0.25	0.43	0.44	0.44	0.45	0.48	0.52
QL	0.24	0.063	0.016	0.076	0.24	0.46	0.48	0.48	0.48	0.48	0.48

Tabla 2.16: Error cuadrático medio bajo un 20% de contaminación para $\theta = 2$

x_0	0	1	2	3	4	5	6	7	8	9	10
MV	0.39	0.19	0.065	0.024	0.064	0.18	0.38	0.66	1.02	1.46	1.98
MI	0.48	0.25	0.072	0.0242	0.10	0.26	0.43	0.57	0.67	0.73	0.75
MT	0.16	0.37	0.097	0.026	0.12	0.31	0.52	0.68	0.72	0.63	0.45
H	0.48	0.23	0.071	0.025	0.087	0.25	0.51	0.65	0.65	0.66	0.68
QL	0.46	0.24	0.073	0.024	0.085	0.25	0.51	0.68	0.69	0.69	0.69

Tabla 2.17: Error cuadrático medio bajo un 20% de contaminación para $\theta = 3$

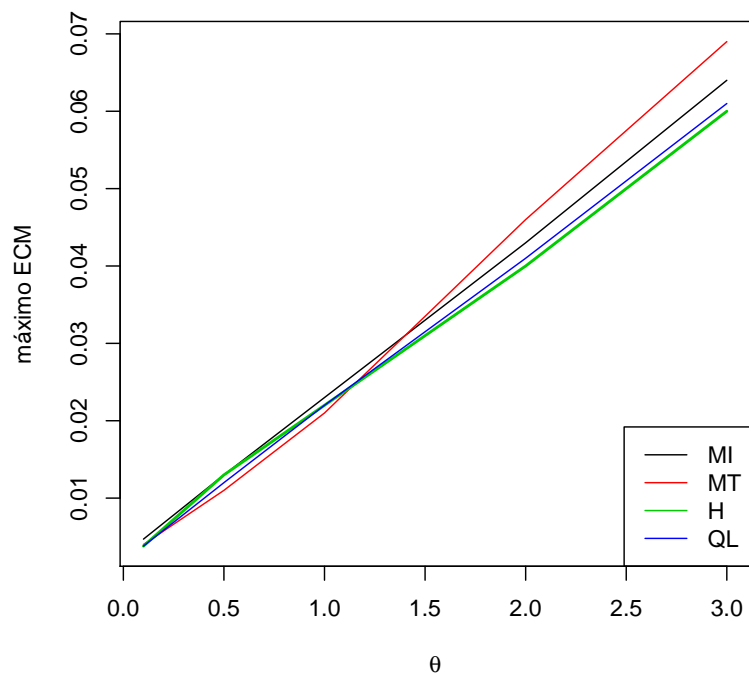


Figura 2.4: Maximo ECM en muestras con un 5% de contaminaci3n.

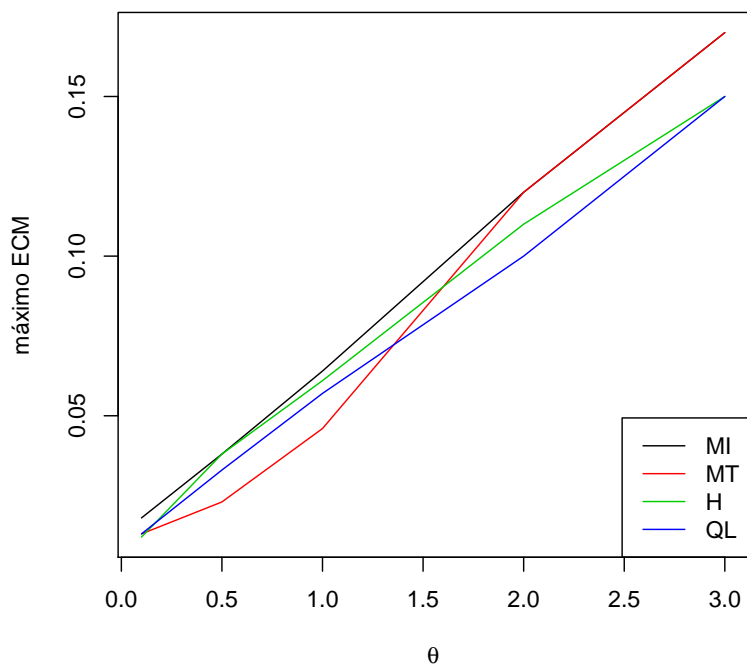


Figura 2.5: Mximo ECM en muestras con un 10 % de contaminacin.

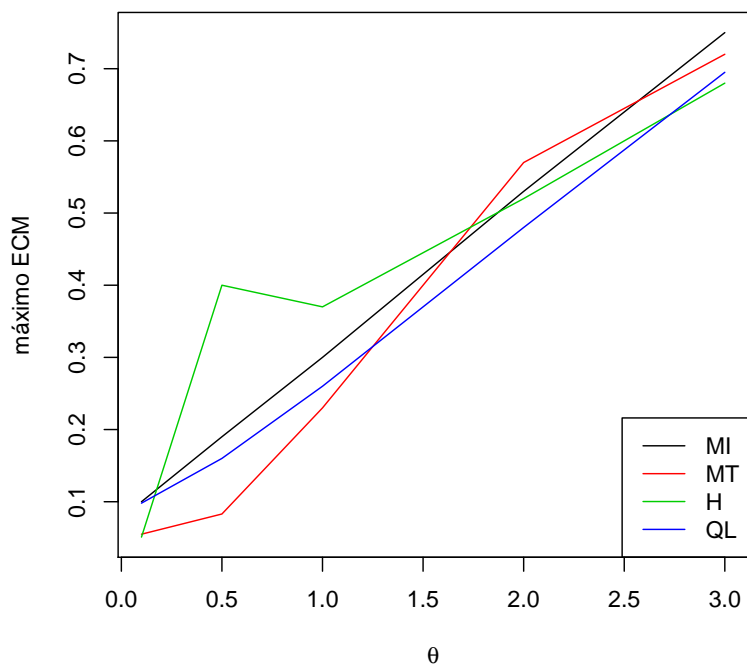


Figura 2.6: Máximo ECM en muestras con un 20 % de contaminación.

Apéndice A

Demostraciones de los Teoremas del capítulo 1

En este apéndice consideramos una clase de M-estimadores que incluye como caso particular a los WMT-estimadores. Sean $(y_1, \mathbf{x}_1), \dots, (y_n, \mathbf{x}_n)$ vectores aleatorios i.i.d con $y_i \in \mathbb{R}, \mathbf{x}_i \in \mathbb{R}^p$ y sea $\Phi : \mathbb{R} \times \mathbb{R}^p \times \mathbb{R}^p \times \mathbb{R}^q \rightarrow \mathbb{R}$. Consideremos el estimador $\hat{\boldsymbol{\beta}}_n$ definido por

$$\hat{\boldsymbol{\beta}}_n = \arg \min_{\boldsymbol{\beta} \in \mathbb{R}^p} \sum_{i=1}^n \Phi(y_i, \mathbf{x}_i, \boldsymbol{\beta}, \hat{\theta}_n), \quad (\text{A.1})$$

con $\hat{\theta}_n$ una sucesión de estimadores de θ_0 donde θ_0 es un parámetro de ruido con valores en \mathbb{R}^q .

A.1. Consistencia

En esta sección estudiamos la consistencia de los estimadores definidos por (A.1) que nos permitirá deducir la consistencia de los WMT-estimadores.

Supondremos que se cumplen las siguientes condiciones

P0 Existe θ_0 tal que la sucesión de estimadores $\hat{\theta}_n$ converge casi seguramente a θ_0 .

P1 La función Φ es continua y acotada y existe una función $\vartheta(\mathbf{x}, \theta) : \mathbb{R}^p \times \mathbb{R}^q \rightarrow \mathbb{R}$ y una constante C tal que

$$|\Phi(y_i, \mathbf{x}_i, \boldsymbol{\beta}, \theta_2) - \Phi(y_i, \mathbf{x}_i, \boldsymbol{\beta}, \theta_1)| \leq C |\vartheta(\mathbf{x}, \theta_2) - \vartheta(\mathbf{x}, \theta_1)|$$

para todo $y, \mathbf{x}, \boldsymbol{\beta}, \theta_1$ y θ_2 . Más aún, si θ_0 es como en **P0**, $\theta_n \rightarrow \theta_0$ implica $\sup_{\mathbf{x}} |\vartheta(\mathbf{x}, \theta_n) - \vartheta(\mathbf{x}, \theta_0)| \rightarrow 0$, es decir, $\vartheta(\mathbf{x}, \theta)$ es equicontinua en θ_0 .

P2 Sea θ_0 como en **P0**. Existe β_0 tal que

$$E_{\beta_0}(\Phi(y, \mathbf{x}, \beta_0, \theta_0)) < E_{\beta_0}(\Phi(y, \mathbf{x}, \beta, \theta_0)) \quad (\text{A.2})$$

para todo $\beta \neq \beta_0$.

P3 Sea $S = \{\mathbf{t} \in \mathbb{R}^p, \|\mathbf{t}\| = 1\}$, θ_0 como en **P0** y $\widehat{\beta}_0$ como en **P2**. Entonces existe una función $\Phi^*(y, \mathbf{x}, j)$, $j = -1, 0, 1$ tal que, para todo $\mathbf{t} \in S$,

$$\lim_{\gamma \rightarrow +\infty} \Phi(y, \mathbf{x}, \gamma \mathbf{t}, \theta_0) = \Phi^*(y, \mathbf{x}, \text{sign}(\mathbf{t}^T \mathbf{x})),$$

y, si $\mathbf{t}^T \mathbf{x} \neq 0$, entonces existe un entorno de \mathbf{t} donde esta convergencia es uniforme en \mathbf{t} , y y \mathbf{x} . Más aún,

$$\tau = \inf_{\|\mathbf{t}\|=1} [E_{\beta_0}(\Phi^*(y, \mathbf{x}, \text{sign}(\mathbf{t}^T \mathbf{x})) - E_{\beta_0}(\Phi(y, \mathbf{x}, \beta_0, \theta_0))] > 0. \quad (\text{A.3})$$

A.1.1. Un resultado general de consistencia

Dado $\mathbf{t} \in \mathbb{R}^p$ y $\varepsilon > 0$, notaremos $B(\mathbf{t}, \varepsilon)$ a la bola de centro \mathbf{t} y radio ε , es decir $B(\mathbf{t}, \varepsilon) = \{\mathbf{s} \in \mathbb{R}^p : \|\mathbf{s} - \mathbf{t}\| \leq \varepsilon\}$.

Para demostrar la consistencia de los estimadores definidos en A.1 necesitamos el siguiente lema.

Lema A.1.1 *Supongamos que \mathbf{t} y δ son tales que $P(\mathbf{x}^T \mathbf{t} = 0) < \delta$. Entonces existen $\varepsilon > 0$ y un conjunto compacto $C_{\mathbf{t}} \subset \mathbb{R}^p$ tales que (i) $P(\mathbf{x} \in C_{\mathbf{t}}) > 1 - \delta$ y (ii) para todo $\mathbf{s} \in B(\mathbf{t}, \varepsilon)$ y $\mathbf{x} \in C_{\mathbf{t}}$ se tiene que $\text{sign}(\mathbf{x}^T \mathbf{t}) = \text{sign}(\mathbf{x}^T \mathbf{s})$ y $|\mathbf{x}^T \mathbf{s}| > \varepsilon$.*

Demostración

Como $\lim_{n \rightarrow \infty} P(|\mathbf{x}^T \mathbf{t}| \geq 1/n) = P(|\mathbf{x}^T \mathbf{t}| > 0) > 1 - \delta$ entonces existen $\xi > 0$ y $n_0 \in \mathbb{N}$ tal que $P(|\mathbf{x}^T \mathbf{t}| \geq 1/n) > 1 - \delta + \xi$ para todo $n \geq n_0$. Sea $\varsigma = 1/n_0$ y $K \geq 1$ tal que $P(\|\mathbf{x}\| \leq K) > 1 - \xi$. Como

$$P(|\mathbf{x}^T \mathbf{t}| \geq \varsigma \text{ y } \|\mathbf{x}\| \leq K) \geq P(|\mathbf{x}^T \mathbf{t}| \geq \varsigma) - P(\|\mathbf{x}\| > K) > 1 - \delta + \xi - \xi = 1 - \delta,$$

entonces tomando $C_{\mathbf{t}} = \{\mathbf{x} \in \mathbb{R}^p : |\mathbf{x}^T \mathbf{t}| \geq \varsigma, \|\mathbf{x}\| \leq K\}$ se obtiene (i). Para demostrar (ii) tomemos $\varepsilon = \varsigma/(2K)$, $\mathbf{x} \in C_{\mathbf{t}}$ y $\mathbf{s} \in B(\mathbf{t}, \varepsilon)$. Luego,

$$|\mathbf{x}^T \mathbf{t} - \mathbf{x}^T \mathbf{s}| \leq \|\mathbf{x}\| \|\mathbf{t} - \mathbf{s}\| < K \frac{\varsigma}{2K} = \frac{\varsigma}{2}.$$

Si $\mathbf{x}^T \mathbf{t} > 0$, como $\mathbf{x}^T \mathbf{t} \geq \varsigma$ resulta

$$\mathbf{x}^T \mathbf{s} > \mathbf{x}^T \mathbf{t} - \frac{\varsigma}{2} > 0. \quad (\text{A.4})$$

Similarmente si $\mathbf{x}^T \mathbf{t} < 0$, como $\mathbf{x}^T \mathbf{t} \leq -\varsigma$ resulta

$$\mathbf{x}^T \mathbf{s} < \mathbf{x}^T \mathbf{t} + \frac{\varsigma}{2} < 0. \quad (\text{A.5})$$

Es decir, $\text{sign}(\mathbf{x}^T \mathbf{t}) = \text{sign}(\mathbf{x}^T \mathbf{s})$. Por otra parte,

$$|\mathbf{x}^T \mathbf{s}| \geq |\mathbf{x}^T \mathbf{t}| - |\mathbf{x}^T \mathbf{s} - \mathbf{x}^T \mathbf{t}| > \varsigma - \frac{\varsigma}{2} = \frac{\varsigma}{2} > \varepsilon. \quad (\text{A.6})$$

Luego, de (A.4), (A.5) y (A.6) se obtiene (ii) ■

Lema A.1.2 *Sea (y_i, \mathbf{x}_i) , $i \in \mathbb{N}$ una sucesión de vectores aleatorios i.i.d y sea $\widehat{\boldsymbol{\beta}}_n$ el estimador definido por (A.1). Supongamos que se cumplen las condiciones **P0**, **P1** y **P3** y que $P(\mathbf{t}^T \mathbf{x} = 0) < \tau/M$ para todo $\mathbf{t} \in S$, donde τ está definido en (A.3) y $M = \sup_{y, \mathbf{x}, \beta, \theta} \Phi(y, \mathbf{x}, \beta, \theta)$. Entonces, existe un conjunto compacto $C \subset \mathbb{R}^p$ y $n_0 \in \mathbb{N}$ tales que, si $n \geq n_0$, $\widehat{\boldsymbol{\beta}}_n \in C$ casi seguramente.*

Demostración

Alcanza con probar que existe $K_0 \in \mathbb{N}$ tal que con probabilidad uno, existe $n_0 \in \mathbb{N}$ que satisface

$$\sum_{i=1}^n \Phi_n(y_i, \mathbf{x}_i, \boldsymbol{\beta}_0, \widehat{\theta}_n) < \inf_{\|\boldsymbol{\beta}\| > K_0} \sum_{i=1}^n \Phi_n(y_i, \mathbf{x}_i, \boldsymbol{\beta}, \widehat{\theta}_n) \quad \forall n \geq n_0.$$

Esto es equivalente a

$$\overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \Phi_n(y_i, \mathbf{x}_i, \boldsymbol{\beta}_0, \widehat{\theta}_n) < \underline{\lim}_{n \rightarrow \infty} \frac{1}{n} \inf_{\|\boldsymbol{\beta}\| > K_0} \sum_{i=1}^n \Phi_n(y_i, \mathbf{x}_i, \boldsymbol{\beta}, \widehat{\theta}_n) \quad \text{c.s.} \quad (\text{A.7})$$

Por **P0** y **P1**

$$\left| \frac{1}{n} \sum_{i=1}^n \Phi(y_i, \mathbf{x}_i, \boldsymbol{\beta}, \widehat{\theta}_n) - \frac{1}{n} \sum_{i=1}^n \Phi(y_i, \mathbf{x}_i, \boldsymbol{\beta}, \theta_0) \right| \leq C \sup_{\mathbf{x}} \left| \vartheta(\mathbf{x}, \widehat{\theta}_n) - \vartheta(\mathbf{x}, \theta_0) \right| \rightarrow 0 \quad \text{c.s.} \quad (\text{A.8})$$

para todo $\boldsymbol{\beta} \in \mathbb{R}^p$. Entonces basta probar que existe K_0 tal que

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \Phi(y_i, \mathbf{x}_i, \boldsymbol{\beta}_0, \theta_0) < \underline{\lim}_{n \rightarrow \infty} \frac{1}{n} \inf_{\|\boldsymbol{\beta}\| > K_0} \sum_{i=1}^n \Phi(y_i, \mathbf{x}_i, \boldsymbol{\beta}, \theta_0) \quad \text{c.s.}, \quad (\text{A.9})$$

y esto es equivalente a

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \Phi(y_i, \mathbf{x}_i, \boldsymbol{\beta}_0, \theta_0) < \underline{\lim}_{n \rightarrow \infty} \inf_{\gamma > K_0} \inf_{\mathbf{s} \in S} \frac{1}{n} \sum_{i=1}^n \Phi(y_i, \mathbf{x}_i, \gamma \mathbf{s}, \theta_0) \quad \text{c.s.} \quad (\text{A.10})$$

Empezamos probando que, para todo $\mathbf{t} \in S$, existe ε tal que

$$E_{\beta_0}(\Phi(y, \mathbf{x}, \beta_0, \theta_0)) < E_{\beta_0} \left(\lim_{\gamma \rightarrow \infty} \inf_{\mathbf{s} \in B(\mathbf{t}, \varepsilon)} \Phi(y, \mathbf{x}, \gamma \mathbf{s}, \theta_0) \right). \quad (\text{A.11})$$

Por el Lema A.1.1, dado $\mathbf{t} \in S$ existe un conjunto compacto $C_{\mathbf{t}} \subset \mathbb{R}^p$ y $\varepsilon > 0$ tales que $P(\mathbf{x} \in C_{\mathbf{t}}) > 1 - \tau/M$ y, para todo $\mathbf{s} \in B(\mathbf{t}, \varepsilon)$ y $\mathbf{x} \in C_{\mathbf{t}}$, $|\mathbf{s}^T \mathbf{x}| > \varepsilon$ y $\text{sign}(\mathbf{s}^T \mathbf{x}) = \text{sign}(\mathbf{t}^T \mathbf{x})$. Dados $\mathbf{x} \in \mathbb{R}^p$, $y \in \mathbb{R}$ se tiene

$$\lim_{\gamma \rightarrow \infty} \inf_{\mathbf{s} \in B(\mathbf{t}, \varepsilon)} \Phi(y, \mathbf{x}, \gamma \mathbf{s}, \theta_0) \leq \lim_{\gamma \rightarrow \infty} \Phi(y, \mathbf{x}, \gamma \mathbf{t}, \theta_0) = \Phi^*(y, \mathbf{x}, \text{sign}(\mathbf{t}^T \mathbf{x})).$$

Supongamos que se cumple la desigualdad estricta en algún $y \in \mathbb{R}$, $\mathbf{x} \in C_{\mathbf{t}}$, es decir

$$\lim_{\gamma \rightarrow \infty} \inf_{\mathbf{s} \in B(\mathbf{t}, \varepsilon)} \Phi(y, \mathbf{x}, \gamma \mathbf{s}, \theta_0) < \Phi^*(y, \mathbf{x}, \text{sign}(\mathbf{t}^T \mathbf{x})).$$

Entonces existen $\zeta > 0$ y sucesiones de números positivos $\gamma_n \rightarrow \infty$ y $\mathbf{s}_n \in B(\mathbf{t}, \varepsilon)$ tales que

$$\Phi(y, \mathbf{x}, \gamma_n \mathbf{s}_n, \theta_0) < \Phi^*(y, \mathbf{x}, \text{sign}(\mathbf{t}^T \mathbf{x})) - \zeta. \quad (\text{A.12})$$

Podemos asumir que $\mathbf{s}_n \rightarrow \mathbf{s}_0$, donde $\mathbf{s}_0 \in B(\mathbf{t}, \varepsilon)$ y $\mathbf{s}_0^T \mathbf{x} \neq 0$. Más aún, como para n suficientemente grande, el signo de $\mathbf{s}_n^T \mathbf{x}$ es igual al signo de $\mathbf{s}_0^T \mathbf{x}$ y de $\mathbf{t}^T \mathbf{x}$, por P3 se tiene

$$\lim_{n \rightarrow \infty} \Phi(y, \mathbf{x}, \gamma_n \mathbf{s}_n, \theta_0) = \Phi^*(y, \mathbf{x}, \text{sign}(\mathbf{s}_0^T \mathbf{x})) = \Phi^*(y, \mathbf{x}, \text{sign}(\mathbf{t}^T \mathbf{x})),$$

lo que contradice (A.12). Entonces

$$\lim_{\gamma \rightarrow \infty} \inf_{\mathbf{s} \in B(\mathbf{t}, \varepsilon)} \Phi(y, \mathbf{x}, \gamma \mathbf{s}, \theta_0) = \Phi^*(y, \mathbf{x}, \text{sign}(\mathbf{t}^T \mathbf{x}))$$

para todo $y \in \mathbb{R}_{\geq 0}$, $\mathbf{x} \in C_{\mathbf{t}}$.

Dado un conjunto A , notamos A^c a su complemento. Entonces, usando que $P(\mathbf{x} \in C_{\mathbf{t}}^c) < \tau/M$, $\sup_{y, \mathbf{x}, j} \Phi^*(y, \mathbf{x}, j) \leq M$ y (A.3) se obtiene

$$\begin{aligned} E_{\beta_0} \left(\lim_{\gamma \rightarrow \infty} \inf_{\mathbf{s} \in B(\mathbf{t}, \varepsilon)} \Phi(y, \mathbf{x}, \gamma \mathbf{s}, \theta_0) \right) &\geq E_{\beta_0} \left(\lim_{\gamma \rightarrow \infty} \inf_{\mathbf{s} \in B(\mathbf{t}, \varepsilon)} \Phi(y, \mathbf{x}, \gamma \mathbf{s}, \theta_0) I_{C_{\mathbf{t}}}(\mathbf{x}) \right) \\ &= E_{\beta_0} \left(\Phi^*(y, \mathbf{x}, \text{sign}(\mathbf{t}^T \mathbf{x})) I_{C_{\mathbf{t}}}(\mathbf{x}) \right) \\ &\geq E_{\beta_0} \left(\Phi^*(y, \mathbf{x}, \text{sign}(\mathbf{t}^T \mathbf{x})) \right) \\ &= E_{\beta_0} \left(\Phi^*(y, \mathbf{x}, \text{sign}(\mathbf{t}^T \mathbf{x})) I_{C_{\mathbf{t}}^c}(\mathbf{x}) \right) \\ &> E_{\beta_0} \left(\Phi(y, \mathbf{x}, \beta_0, \theta_0) \right), \end{aligned}$$

lo que prueba (A.11).

Finalmente mostraremos como obtener (A.10) de (A.11) usando la ley de los grandes números y la compacidad de S . Por (A.11) y el Teorema de convergencia mayorada, para todo $\mathbf{t} \in S$ existen $\zeta_{\mathbf{t}}$, $\varepsilon_{\mathbf{t}}$ y $K_{\mathbf{t}}$ tales que

$$E_{\beta_0} \left(\inf_{\gamma > K_{\mathbf{t}}} \inf_{\mathbf{s} \in B(\mathbf{t}, \varepsilon_{\mathbf{t}})} \Phi(y, \mathbf{x}, \gamma \mathbf{s}, \theta_0) \right) > E_{\beta_0} \Phi(y, \mathbf{x}, \beta_0, \theta_0) + \zeta_{\mathbf{t}}. \quad (\text{A.13})$$

Como S es compacto, existe un conjunto finito $\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_j$ de elementos de S tal que $S \subset \bigcup_{h=1}^j B(\mathbf{t}_h, \varepsilon_{\mathbf{t}_h})$. Sea $K_0 = \max\{K_{\mathbf{t}_1}, \dots, K_{\mathbf{t}_j}\}$, y $\zeta_0 = \min\{\zeta_{\mathbf{t}_1}, \dots, \zeta_{\mathbf{t}_j}\}$. Entonces

$$\inf_{\gamma > K_0} \inf_{\mathbf{s} \in S} \frac{1}{n} \sum_{i=1}^n \Phi(y_i, \mathbf{x}_i, \gamma \mathbf{s}, \theta_0) \geq \inf_{1 \leq h \leq j} \inf_{\gamma > K_h} \inf_{\mathbf{s} \in B(\mathbf{t}_h, \varepsilon_{\mathbf{t}_h})} \frac{1}{n} \sum_{i=1}^n \Phi(y_i, \mathbf{x}_i, \gamma \mathbf{s}, \theta_0),$$

y, por la ley de los grandes números, y (A.13) obtenemos

$$\begin{aligned} \lim_{n \rightarrow \infty} \inf_{\gamma > K_0} \inf_{\mathbf{s} \in S} \frac{1}{n} \sum_{i=1}^n \Phi(y_i, \mathbf{x}_i, \gamma \mathbf{s}, \theta_0) &\geq \inf_{1 \leq h \leq j} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \inf_{\gamma > K_h} \inf_{\mathbf{s} \in B(\mathbf{t}_h, \varepsilon_{\mathbf{t}_h})} \Phi(y_i, \mathbf{x}_i, \gamma \mathbf{s}, \theta_0) \\ &= \inf_{1 \leq h \leq j} E \left(\inf_{\gamma > K_h} \inf_{\mathbf{s} \in B(\mathbf{t}_h, \varepsilon_{\mathbf{t}_h})} \Phi(y_i, \mathbf{x}_i, \gamma \mathbf{s}, \theta_0) \right) \\ &\geq E_{\beta_0} (\Phi(y, \mathbf{x}, \beta_0, \theta_0)) + \zeta_0 \text{ c.s..} \end{aligned}$$

Como

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \Phi(y_i, \mathbf{x}_i, \beta_0, \theta_0) = E_{\beta_0} (\Phi(y, \mathbf{x}, \beta_0, \theta_0)),$$

(A.10) se deduce de (A.11). ■

Teorema A.1.1 *Sea (y_i, \mathbf{x}_i) , $i \in \mathbb{N}$ una sucesión de vectores aleatorios i.i.d con $y_i \in \mathbb{R}$, $\mathbf{x}_i \in \mathbb{R}^p$ y sea $\hat{\beta}_n$ el estimador definido por (A.1). Supongamos que se cumplen las propiedades **P0-P3** y que $P(\mathbf{x}^T \mathbf{t} = 0) < \tau/M$ para todo $\mathbf{t} \in S$, con τ definido por (A.3) y $M = \sup_{y, \mathbf{x}, \beta} \Phi(y, \mathbf{x}, \beta, \theta_0)$. Entonces, la sucesión $\hat{\beta}_n$ converge a β_0 casi seguramente.*

Demostración

Vamos a probar que, para todo U , entorno abierto de β_0 , existe $n_0 \in \mathbb{N}$ tal que, si $n \geq n_0$ entonces $\hat{\beta}_n \in U$. Sea C el compacto dado por el Lema A.1.2. Entonces, de acuerdo a este lema, es suficiente probar que para todo U , entorno abierto de β_0 existe $n_0 \in \mathbb{N}$ tal que si $n \geq n_0$ se tiene

$$\inf_{\beta \in C-U} \frac{1}{n} \sum_{i=1}^n \Phi_n(y_i, \mathbf{x}_i, \beta, \hat{\theta}_n) > \frac{1}{n} \sum_{i=1}^n \Phi_n(y_i, \mathbf{x}_i, \beta_0, \hat{\theta}_n) \text{ c.s.,}$$

o equivalentemente

$$\underline{\lim}_{n \rightarrow \infty} \inf_{\beta \in C-U} \frac{1}{n} \sum_{i=1}^n \Phi(y_i, \mathbf{x}_i, \beta, \hat{\theta}_n) > \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \Phi(y_i, \mathbf{x}_i, \beta_0, \hat{\theta}_n) \text{ c.s..} \quad (\text{A.14})$$

Por (A.8), para probar esto basta mostrar que

$$\underline{\lim}_{n \rightarrow \infty} \inf_{\beta \in C-U} \frac{1}{n} \sum_{i=1}^n \Phi(y_i, \mathbf{x}_i, \beta, \theta_0) > \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \Phi(y_i, \mathbf{x}_i, \beta_0, \theta_0) \text{ c.s..} \quad (\text{A.15})$$

Usando el Teorema de convergencia mayorada y la compacidad de $C - U$ se puede encontrar $\mathbf{t}_j \in \mathbb{R}^p$, $\varepsilon_j > 0$, $\eta_j > 0$, $1 \leq j \leq h$ tales que

$$\cup_{j=1}^h B(\mathbf{t}_j, \varepsilon_j) \supset C - U$$

y

$$E(\Phi(y, \mathbf{x}, \beta_0, \theta_0)) \leq E\left(\inf_{\beta \in B(\mathbf{t}_j, \varepsilon_j)} \Phi(y, \mathbf{x}, \beta, \theta_0)\right) - \eta_j, 1 \leq j \leq h. \quad (\text{A.16})$$

También se tiene

$$\inf_{\beta \in C-U} \frac{1}{n} \sum_{i=1}^n \Phi(y_i, \mathbf{x}_i, \beta, \theta_0) \geq \inf_{1 \leq j \leq h} \inf_{\beta \in S(\mathbf{t}_j, \varepsilon_j)} \frac{1}{n} \sum_{i=1}^n \Phi(y_i, \mathbf{x}_i, \beta, \theta_0).$$

Entonces, si $\eta_0 = \min_{1 \leq j \leq h} \eta_j$, por la ley de los grandes números y (A.16) podemos escribir

$$\begin{aligned} \underline{\lim}_{n \rightarrow \infty} \inf_{\beta \in C-U} \frac{1}{n} \sum_{i=1}^n \Phi(y_i, \mathbf{x}_i, \beta, \theta_0) &\geq \inf_{1 \leq j \leq h} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \inf_{\beta \in S(\mathbf{t}_j, \varepsilon_j)} \Phi(y_i, \mathbf{x}_i, \beta, \theta_0) \\ &\geq \inf_{1 \leq j \leq h} E\left(\inf_{\beta \in S(\mathbf{t}_j, \varepsilon_j)} \Phi(y_i, \mathbf{x}_i, \beta, \theta_0)\right) \\ &\geq E(\Phi(y, \mathbf{x}, \beta_0, \theta_0)) + \eta_0, \end{aligned}$$

y, como se tiene

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \Phi(y_i, \mathbf{x}_i, \beta_0, \theta_0) = E(\Phi(y_i, \mathbf{x}_i, \beta_0, \theta_0)),$$

se obtiene(A.15).■

A.1.2. Consistencia de los WMT-estimadores

En esta sección se demuestra la consistencia de los WMT-estimadores. La demostración consiste en verificar que se cumplen las hipótesis del Teorema A.1.1. Necesitamos algunos lemas previos.

Lema A.1.3 *Supongamos que se cumplen **A1-A2** y **B1-B6**. Entonces existe $\kappa_0 > 0$ tal que*

$$E_\lambda(\rho(t(y)) - m(\lambda)) < 1 - \kappa_0 \quad (\text{A.17})$$

para todo λ .

Demostración

Por la desigualdad de Chebyshev se tiene que, para todo $\delta > 0$

$$\begin{aligned} E_\lambda(\rho(t(y)) - m(\lambda)) &\leq E_\lambda(\rho(t(y)) - E_\lambda(t(y))) \\ &\leq \rho(\delta)P(|t(y) - E_\lambda(t(y))| < \delta) + P(|t(y) - E_\lambda(t(y))| \geq \delta) \\ &= 1 - (1 - \rho(\delta))P(|t(y) - E_\lambda(t(y))| < \delta) \\ &\leq 1 - (1 - \rho(\delta))(1 - \frac{A}{\delta^2}). \end{aligned}$$

Entonces, tomando $\delta = A^{1/2} + \eta$ y usando por **B6** se obtiene

$$E_\lambda(\rho(t(y)) - m(\lambda)) \leq 1 - \kappa_0$$

con

$$\kappa_0 = (1 - \rho(A^{1/2} + \eta)) \left(1 - \frac{A}{(A^{1/2} + \eta)^2} \right) > 0,$$

con lo que el lema queda demostrado. ■

Lema A.1.4 *Supongamos que se cumplen **B1- B6**. Entonces **A1-A5** implican que $m(\lambda)$ es continua.*

Demostración

Sea $\lambda^{(0)} \in \mathbb{R}_{>0}$ y λ_i una sucesión que converge a $\lambda^{(0)}$. Queremos probar que $m(\lambda_i) \rightarrow m(\lambda^{(0)})$. Supongamos que no es así. Entonces, pasando a una subsucesión si es necesario, podemos suponer que $m(\lambda_i)$ converge a un valor m_0 , posiblemente igual a $\mp\infty$. Supongamos que $m_0 = \mp\infty$ y sea $\lambda^{(1)} < \lambda^{(0)}$. Entonces, por el Teorema de convergencia mayorada, se tiene $E_{\lambda^{(1)}}(\rho(t(y)) - m(\lambda_i)) \rightarrow 1$. Luego, usando **A5** y **A4** se obtiene que $E_{\lambda_i}(\rho(t(y)) - m(\lambda_i)) \rightarrow 1$ contradiciendo el hecho de que, por el Lema A.1.3, $E_\lambda(\rho(t(y)) - m(\lambda)) < 1 - \kappa_0$ para todo λ . Esto prueba que m_0 no puede ser $+\infty$ ni $-\infty$. Por la definición de $m(\lambda)$ se tiene

$$E_{\lambda_n}(\rho(t(y)) - m(\lambda_n)) \leq E_{\lambda_n}(\rho(t(y)) - m(\lambda^{(0)}))$$

para todo $n \in \mathbb{N}$. Tomando límites resulta

$$E_{\lambda^{(0)}}(\rho(t(y) - m_0)) \leq E_{\lambda^{(0)}}(\rho(t(y) - m(\lambda^{(0)}))).$$

Entonces, por la unicidad de $m(\lambda^{(0)})$, se obtiene que $m(\lambda^{(0)}) = m_0$. Esto prueba la continuidad $m(\lambda)$. ■

Lema A.1.5 *Sea $\boldsymbol{\mu}_0 \in \mathbb{R}^p$ y $\boldsymbol{\Sigma}_0 \in \mathbb{R}^{p \times p}$ una matriz definida positiva. Supongamos que se cumple **B8**. Entonces $\boldsymbol{\mu}_n \rightarrow \boldsymbol{\mu}_0$ y $\boldsymbol{\Sigma}_n \rightarrow \boldsymbol{\Sigma}_0$ implican*

$$\sup_{\mathbf{x}} |w(\mathbf{x}, \boldsymbol{\mu}_n, \boldsymbol{\Sigma}_n) - w(\mathbf{x}, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)| \rightarrow 0. \quad (\text{A.18})$$

Demostración

Supongamos que $\sup_{\mathbf{x}} |w(\mathbf{x}, \boldsymbol{\mu}_n, \boldsymbol{\Sigma}_n) - w(\mathbf{x}, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)|$ no tiende a cero. Entonces existe una sucesión \mathbf{x}_n y $\varepsilon > 0$ tal que para todo n

$$|w(\mathbf{x}_n, \boldsymbol{\mu}_n, \boldsymbol{\Sigma}_n) - w(\mathbf{x}_n, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)| > \varepsilon. \quad (\text{A.19})$$

Podemos asumir sin pérdida de generalidad que $\mathbf{x}_n \rightarrow \mathbf{x}_0$ o $\|\mathbf{x}_n\| \rightarrow \infty$. En el primer caso

$$\lim_{n \rightarrow \infty} (\mathbf{x}_n - \boldsymbol{\mu}_n)^T \boldsymbol{\Sigma}_n^{-1} (\mathbf{x}_n - \boldsymbol{\mu}_n) = (\mathbf{x}_0 - \boldsymbol{\mu}_0)^T \boldsymbol{\Sigma}_0^{-1} (\mathbf{x}_0 - \boldsymbol{\mu}_0) = \lim_{n \rightarrow \infty} (\mathbf{x}_n - \boldsymbol{\mu}_0)^T \boldsymbol{\Sigma}_0^{-1} (\mathbf{x}_n - \boldsymbol{\mu}_0).$$

Por lo tanto, como por **B8**, ω es continua se obtiene

$$\lim_{n \rightarrow \infty} w(\mathbf{x}_n, \boldsymbol{\mu}_n, \boldsymbol{\Sigma}_n) = \lim_{n \rightarrow \infty} w(\mathbf{x}_n, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) = w(\mathbf{x}_0, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0),$$

lo que contradice (A.19). En el segundo caso se tiene que

$$\lim_{n \rightarrow \infty} (\mathbf{x}_n - \boldsymbol{\mu}_n)^T \boldsymbol{\Sigma}_n^{-1} (\mathbf{x}_n - \boldsymbol{\mu}_n) = \lim_{n \rightarrow \infty} (\mathbf{x}_n - \boldsymbol{\mu}_0)^T \boldsymbol{\Sigma}_0^{-1} (\mathbf{x}_n - \boldsymbol{\mu}_0) = \infty,$$

y **B8** implica que existe a tal que $\lim_{d \rightarrow \infty} \omega(d) = a$. Por lo tanto, resulta

$$\lim_{n \rightarrow \infty} w(\mathbf{x}_n, \boldsymbol{\mu}_n, \boldsymbol{\Sigma}_n) = \lim_{n \rightarrow \infty} w(\mathbf{x}_n, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) = a$$

contradiendo (A.19) también en este caso. ■

Demostración del Teorema 1.4.1

Sea

$$\Phi_0(y, \mathbf{x}, \boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = w(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) \rho(t(y) - m(g^{-1}(\boldsymbol{\beta}^T \mathbf{x}))). \quad (\text{A.20})$$

Por el Teorema A.1.1, alcanza con probar que se cumplen las condiciones **P0-P3** con $\theta = (\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $\hat{\theta}_n = (\hat{\boldsymbol{\mu}}_n, \hat{\boldsymbol{\Sigma}}_n)$, $\Phi = \Phi_0$ y $\vartheta(\mathbf{x}, \theta) = w(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$.

P1 se deduce del Lema A.1.5. Sea

$$D(\lambda, u) = E_\lambda(\rho(t(y) - u)),$$

entonces, por **A2**, se tiene que para cada $\beta \neq \beta_0$ resulta

$$D(g^{-1}(\beta_0^T \mathbf{x}), m(g^{-1}(\beta^T \mathbf{x}))) - D(g^{-1}(\beta_0^T \mathbf{x}), m(g^{-1}(\beta_0^T \mathbf{x}))) \geq 0 \quad (\text{A.21})$$

y

$$D(g^{-1}(\beta_0^T \mathbf{x}), m(g^{-1}(\beta^T \mathbf{x}))) - D(g^{-1}(\beta_0^T \mathbf{x}), m(g^{-1}(\beta_0^T \mathbf{x}))) > 0 \text{ if } \beta_0^T \mathbf{x} \neq \beta^T \mathbf{x}. \quad (\text{A.22})$$

Además se tiene que

$$E_{\beta_0}(\Phi_0(y, \mathbf{x}, \beta, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) | \mathbf{x}) = D(g^{-1}(\beta_0^T \mathbf{x}), m(g^{-1}(\beta^T \mathbf{x}))) w(\mathbf{x}, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0). \quad (\text{A.23})$$

Sea $V = \{\mathbf{x} : (\beta - \beta_0)^T \mathbf{x} \neq 0\} \cap \{\mathbf{x} : w(\mathbf{x}, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) > 0\}$, entonces, por **B9** resulta $P(V) > 0$. De (A.21) y (A.23) se deduce que

$$\begin{aligned} & E_{\beta_0}(\Phi_0(y, \mathbf{x}, \beta, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)) - E_{\beta_0}(\Phi_0(y, \mathbf{x}, \beta_0, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)) \\ & \geq E\{[D(g^{-1}(\beta_0^T \mathbf{x}), m(g^{-1}(\beta^T \mathbf{x}))) - D(g^{-1}(\beta_0^T \mathbf{x}), m(g^{-1}(\beta_0^T \mathbf{x})))] w(\mathbf{x}, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) I(V)\}. \end{aligned}$$

Como, para $\mathbf{x} \in V$, se tiene

$$[D(g^{-1}(\beta_0^T \mathbf{x}), m(g^{-1}(\beta^T \mathbf{x}))) - D(g^{-1}(\beta_0^T \mathbf{x}), m(g^{-1}(\beta_0^T \mathbf{x})))] w(\mathbf{x}, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) > 0,$$

resulta que $E_{\beta_0}(\Phi_0(y, \mathbf{x}, \beta, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)) - E_{\beta_0}(\Phi_0(y, \mathbf{x}, \beta_0, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)) > 0$ y, por lo tanto, se verifica **P2**.

Para probar **P3** basta mostrar que $\tau > 0$, donde τ viene dada por (1.23). Por **B9**

$$\delta = \inf_{\mathbf{t} \in S} P(\{\mathbf{t}^T \mathbf{x} \neq 0\} \cap \{w(\mathbf{x}, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) > 0\}) > 0. \quad (\text{A.24})$$

Vamos a mostrar que existe $\zeta > 0$ tal que

$$\inf_{\mathbf{t} \in S} P(\{\mathbf{t}^T \mathbf{x} \neq 0\} \cap \{w(\mathbf{x}, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) > \zeta\}) \geq \delta/2. \quad (\text{A.25})$$

Supongamos que no es así. Entonces existen sucesiones $\mathbf{t}_n \in S$ y $\zeta_n \rightarrow 0$ tales que

$$P(\{\mathbf{t}_n^T \mathbf{x} \neq 0\} \cap \{w(\mathbf{x}, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) > \zeta_n\}) < \delta/2.$$

Sin pérdida de generalidad podemos asumir que $\mathbf{t}_n \rightarrow \mathbf{t}_0 \in S$. Entonces se tiene que

$$P(\{\mathbf{t}_0^T \mathbf{x} \neq 0\} \cap \{w(\mathbf{x}, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) > 0\}) \leq \delta/2,$$

lo que contradice (A.24). También podemos encontrar K_1 y K_2 tales que

$$P(\{\boldsymbol{\beta}_0^T \mathbf{x} \in [K_1, K_2]\}) > 1 - \delta/4. \quad (\text{A.26})$$

Por lo tanto, si escribimos

$$V_{\mathbf{t}} = \{\mathbf{t}^T \mathbf{x} \neq 0\} \cap \{w(\mathbf{x}, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) > \zeta\} \cap \{\boldsymbol{\beta}_0^T \mathbf{x} \in [K_1, K_2]\}, \quad (\text{A.27})$$

por (A.25) y (A.26) se obtiene

$$P(V_{\mathbf{t}}) > \delta/4. \quad (\text{A.28})$$

para todo $\mathbf{t} \in S$. Por **A2**

$$D(g^{-1}(\boldsymbol{\beta}_0^T \mathbf{x}), m(g^{-1}(\boldsymbol{\beta}_0^T \mathbf{x}))) < D(g^{-1}(\boldsymbol{\beta}_0^T \mathbf{x}), m_i), \quad i = -1, 1 \quad (\text{A.29})$$

y

$$D(g^{-1}(\boldsymbol{\beta}_0^T \mathbf{x}), m(g^{-1}(\boldsymbol{\beta}_0^T \mathbf{x}))) \leq D(g^{-1}(\boldsymbol{\beta}_0^T \mathbf{x}), m_3). \quad (\text{A.30})$$

Las ecuaciones (A.29) y (A.30) implican que, para todo $\mathbf{x} \in \mathbb{R}^p$ y $\mathbf{t} \in S$

$$E_{\beta_0}((\Phi_0^*(y, \mathbf{x}, \text{sign}(\mathbf{t}^T \mathbf{x})) - \Phi_0(y, \mathbf{x}, \boldsymbol{\beta}_0^T \mathbf{x}, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)) | x) \geq 0. \quad (\text{A.31})$$

Definamos $C_i(\lambda) = D(\lambda, m_i) - D(\lambda, m(\lambda))$, $i = -1, 1$. Entonces para todo $\lambda \in [g^{-1}(K_1), g^{-1}(K_2)]$ e $i = -1, 1$, tenemos que $C_i(\lambda) > 0$ y C_i es continua. Luego resulta

$$c_i = \min_{\lambda \in [g^{-1}(K_1), g^{-1}(K_2)]} C_i(\lambda) > 0, \quad i = -1, 1.$$

Sea $c_0 = \min(c_{-1}, c_1)$; entonces, por (A.31), (A.27) y (A.28) se tiene que

$$\begin{aligned} & E_{\beta_0}(\Phi_0^*(y, \mathbf{x}, \text{sign}(\mathbf{t}^T \mathbf{x})) - \Phi_0(y, \mathbf{x}, \boldsymbol{\beta}_0^T \mathbf{x}, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)) \\ & \geq E(E_{\beta_0}((\Phi_0^*(y, \mathbf{x}, \text{sign}(\mathbf{t}^T \mathbf{x})) - \Phi_0(y, \mathbf{x}, \boldsymbol{\beta}_0^T \mathbf{x}, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)) | \mathbf{x}) I(\mathbf{x} \in V_{\mathbf{t}})) \\ & \geq E\left(\min_{\lambda \in [g^{-1}(K_1), g^{-1}(K_2)]} \min_{i \in \{-1, 1\}} C_i(\lambda) w(\mathbf{x}, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) I(\mathbf{x} \in V_{\mathbf{t}})\right) \\ & \geq c_0 E(w(\mathbf{x}, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) I(\mathbf{x} \in V_{\mathbf{t}})) \\ & \geq c_0 \zeta \frac{\delta}{4}. \end{aligned}$$

Esto implica que $\tau \geq c_0 \zeta \delta/4$ y, por lo tanto, se cumple **P3**.

Finalmente **P0** se deduce de **B7**. ■

A.2. Distribución Asintótica

En esta sección estudiamos la distribución asintótica de los estimadores definidos en (A.1). Necesitamos asumir que se cumplen las siguientes condiciones.

F1 Existen funciones $S : \mathbb{R} \times \mathbb{R}^p \times \mathbb{R}^p \rightarrow \mathbb{R}$ y $w : \mathbb{R}^p \times \mathbb{R}^q$ tales que

$$\Phi(y, \mathbf{x}, \boldsymbol{\beta}, \theta) = S(y, \mathbf{x}, \boldsymbol{\beta})w(\mathbf{x}, \theta).$$

y

$$E_{\beta_0} \left(\frac{\partial}{\partial \boldsymbol{\beta}} \Big|_{\beta_0} S(y, \mathbf{x}, \boldsymbol{\beta}) \Big| \mathbf{x} \right) = 0$$

F2 Para cada $y \in \mathbb{R}$, $\mathbf{x} \in \mathbb{R}^p$ fijos, $S(y, \mathbf{x}, \cdot) \in C^2(U)$, con U un entorno de $\boldsymbol{\beta}_0$.

F3 $\sqrt{n} (\hat{\theta}_n - \theta_0) = O_p(1)$, con θ_0 como en **P1**

F4 Definamos

$$\Psi_j(y, \mathbf{x}, \boldsymbol{\beta}, \theta) = \frac{\partial}{\partial \beta_j} S(y, \mathbf{x}, \boldsymbol{\beta})w(\mathbf{x}, \theta).$$

Sea $\mathbf{J}_{\Psi}(y, \mathbf{x}, \boldsymbol{\beta}, \theta)$ la matriz Jacobiana de Ψ respecto de $\boldsymbol{\beta}$. Luego, existe $\varepsilon > 0$ tal que

$$E_{\beta_0} \left(\sup_{\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| \leq \varepsilon} \left| \mathbf{J}_{\Psi}^{j,k}(y, \mathbf{x}, \boldsymbol{\beta}, \theta_0) \right| \right) < \infty$$

para todo $1 \leq j, k \leq q$ y $E_{\beta_0} (J_{\Psi}(y, \mathbf{x}, \boldsymbol{\beta}_0, \theta_0))$ es no singular.

F5 La función w es continua, acotada, no creciente y no negativa.

Necesitaremos los dos lemas siguientes cuyas demostraciones pueden encontrarse en Yohai (1985) y en el suplemento de Bergesio y Yohai (2011), disponible online, respectivamente.

Lema A.2.1 Sean $\mathbf{z}_1, \dots, \mathbf{z}_n$ vectores aleatorios i.i.d. con distribución F y sea $\phi : \mathbb{R}^p \times \mathbb{R}^h \rightarrow \mathbb{R}$ una función continua que satisface que, para algún $\varepsilon > 0$,

$$\sup_{\|\theta - \theta_0\| \leq \varepsilon} |\phi(\mathbf{z}, \theta)|$$

bajo F tiene esperanza finita. Supongamos además que $\hat{\xi}_n \rightarrow \theta_0$ c.s.. Entonces

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \phi(\mathbf{z}_i, \hat{\xi}_n) = E(\phi(\mathbf{z}, \theta_0)) \text{ c.s..}$$

Lema A.2.2 Sea $c(\mathbf{z}, \boldsymbol{\beta}) : \mathbb{R}^p \times \mathbb{R}^k \rightarrow \mathbb{R}^k$ una función continuamente diferenciable en $\boldsymbol{\beta}$ y sean $\mathbf{z}_1, \dots, \mathbf{z}_n$ vectores aleatorios i.i.d. de dimensión p . Consideremos una sucesión de estimadores $\widehat{\boldsymbol{\beta}}_n$ tal que $n^{1/2}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) = O_p(1)$. Supongamos que existe $\zeta > 0$ tal que

$$\sup_{\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| \leq \zeta} |c(\mathbf{z}, \boldsymbol{\beta})|$$

tiene esperanza finita y que

$$E \left[\left. \frac{\partial c(\mathbf{z}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \right|_{\boldsymbol{\beta} = \boldsymbol{\beta}_0} \right] = \mathbf{0}.$$

Entonces

$$\frac{1}{n^{1/2}} \sum_{i=1}^n c(\mathbf{z}_i, \widehat{\boldsymbol{\beta}}_n) - \frac{1}{n^{1/2}} \sum_{i=1}^n c(\mathbf{z}_i, \boldsymbol{\beta}_0) = o_p(1).$$

Teorema A.2.1 Sea (y_i, \mathbf{x}_i) , $i \in \mathbb{N}$ una sucesión de vectores aleatorios i.i.d con $y_i \in \mathbb{R}$, $\mathbf{x}_i \in \mathbb{R}^p$ y sea $\widehat{\boldsymbol{\beta}}_n$ el estimador definido por (A.1). Supongamos que se cumplen las propiedades **P0-P3** y **F1-F5** y que $P(\mathbf{t}^T \mathbf{x} = 0) < \tau/M$ para todo $\mathbf{t} \in S$, con τ definido por (A.3) y $M = \sup_{y, \mathbf{x}, \boldsymbol{\beta}, \theta_0} \Phi(y, \mathbf{x}, \boldsymbol{\beta}, \theta_0)$. Entonces

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) \xrightarrow{D} \mathcal{N}(0, B^{-1}AB^{T-1}),$$

donde

$$A = E_{\boldsymbol{\beta}_0} (\Psi(y, \mathbf{x}, \boldsymbol{\beta}_0, \theta_0) \Psi(y, \mathbf{x}, \boldsymbol{\beta}_0, \theta_0)^T) \quad \text{y} \quad B = E_{\boldsymbol{\beta}_0} (\mathbf{J}_{\Psi}(y, \mathbf{x}, \boldsymbol{\beta}_0, \theta_0)).$$

Demostración

El estimador $\widehat{\boldsymbol{\beta}}_n$ satisface

$$\sum_{i=1}^n \Psi(\mathbf{x}_i, y_i, \widehat{\boldsymbol{\beta}}_n, \widehat{\theta}_n) = 0.$$

Mediante una expansión de Taylor obtenemos

$$\frac{1}{n} \sum_{i=1}^n \mathbf{J}_{\Psi}(\mathbf{x}_i, y_i, \xi_n, \widehat{\theta}_n) (\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) = -\frac{1}{n} \sum_{i=1}^n \Psi(\mathbf{x}_i, y_i, \boldsymbol{\beta}_0, \widehat{\theta}_n), \quad (\text{A.32})$$

donde ξ_n es un punto intermedio entre $\widehat{\boldsymbol{\beta}}_n$ y $\boldsymbol{\beta}_0$. Como, por el Teorema 1.4.1, $\widehat{\boldsymbol{\beta}}_n$ es fuertemente consistente a $\boldsymbol{\beta}_0$, entonces ξ_n también es fuertemente consistente a $\boldsymbol{\beta}_0$. Entonces, por **F4** y el Lema A.2.1, obtenemos

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \mathbf{J}_{\Psi}(\mathbf{x}_i, y_i, \xi_n, \widehat{\theta}_n) = B \text{ c.s.} \quad (\text{A.33})$$

Como, por **F4**, la matriz B es no singular, para n suficientemente grande,

$$\frac{1}{n} \sum_{i=1}^n J_{\Psi}(\mathbf{x}_i, y_i, \xi_n, \hat{\theta}_n)$$

es también no singular. Luego, de (A.32), obtenemos

$$\begin{aligned} \sqrt{n}(\hat{\beta}_n - \beta_0) &= - \left(\frac{1}{n} \sum_{i=1}^n J_{\Psi}(\mathbf{x}_i, y_i, \xi_n, \hat{\theta}_n) \right)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n \Psi(\mathbf{x}_i, y_i, \beta_0, \hat{\theta}_n) \\ &= - \left(\frac{1}{n} \sum_{i=1}^n J_{\Psi}(\mathbf{x}_i, y_i, \xi_n, \hat{\theta}_n) \right)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n \left[\Psi(\mathbf{x}_i, y_i, \beta_0, \hat{\theta}_n) - \Psi(\mathbf{x}_i, y_i, \beta_0, \theta_0) + \Psi(\mathbf{x}_i, y_i, \beta_0, \theta_0) \right]. \end{aligned} \quad (\text{A.34})$$

Por el Teorema central del límite se tiene

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \Psi(\mathbf{x}_i, y_i, \beta_0, \theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, A). \quad (\text{A.35})$$

Sea $c_i(y, x, \theta) = w(x, \theta) \frac{\partial S}{\partial \beta_i}(y, x, \beta_0)$. Por **F1**, se tiene

$$E \left(\frac{\partial c_i}{\partial \theta}(y, \mathbf{x}, \theta) \right) = E \left(\frac{\partial w}{\partial \theta}(\mathbf{x}, \theta) E \left(\frac{\partial S}{\partial \beta_i}(y, \mathbf{x}, \beta_0) | \mathbf{x} \right) \right) = 0.$$

Entonces, por el Lema A.2.2, obtenemos

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \left(w(\mathbf{x}_i, \theta_0) - w(\mathbf{x}_i, \hat{\theta}_n) \right) S(y_i, \mathbf{x}_i, \beta_0) = o_p(1)$$

y, por lo tanto,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \Psi(y_i, \mathbf{x}_i, \beta_0, \hat{\theta}_n) - \Psi(y_i, \mathbf{x}_i, \beta_0, \theta_0) = o_p(1). \quad (\text{A.36})$$

En consecuencia (A.34) es igual a

$$- \left(\frac{1}{n} \sum_{i=1}^n J_{\Psi}(\mathbf{x}_i, y_i, \xi_n, \hat{\theta}_n) \right)^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \Psi(\mathbf{x}_i, y_i, \beta_0, \theta_0) + o_p(1) \right). \quad (\text{A.37})$$

Luego, el teorema se deduce de (A.33), (A.35), (A.37) y el Lema de Slutsky. ■

A.2.1. Normalidad asintótica de los WMT-estimadores

Para obtener la normalidad asintótica de los WMT estimadores necesitamos verificar que se cumplen las condiciones **F1** a **F5**. **F1** es inmediata. **F2** se deduce del siguiente lema.

Lema A.2.3 *Si **A2**, **C1**, **C2** y **C4** se cumplen entonces m es dos veces diferenciable.*

Demostración

Para cada $\lambda \in (\lambda^{(1)}, \lambda^{(2)})$ $m(\lambda)$ se define como el mínimo en m de $E_\lambda(\rho(t(y) - m))$. Entonces, para todo λ , se tiene

$$E_\lambda(\psi(t(y) - m(\lambda))) = 0. , \quad (\text{A.38})$$

donde $\psi = \rho'$.

Sea $\psi_0(y, \lambda) = \frac{\partial}{\partial \lambda} f_\lambda(y)$. Por **C4** y el Teorema de la función implícita m es diferenciable y

$$m'(\lambda) = \frac{\int \psi(t(y) - m(\lambda)) \psi_0(y, \lambda) dy}{E_\lambda[\psi'(t(y) - m(\lambda))]}$$

en el caso continuo y

$$m'(\lambda) = \frac{\sum \psi(t(y) - m(\lambda)) \psi_0(y, \lambda)}{E_\lambda[\psi'(t(y) - m(\lambda))]} \quad (\text{A.39})$$

en el caso discreto. luego usando **C1**, **C2** y **C4** resulta que m' también es continuamente diferenciable. ■

La normalidad asintótica de los WMT-estimadores enunciada en el Teorema 1.5.1 es ahora consecuencia directa del Teorema A.2.1.

A continuación se establecen condiciones suficientes para que se cumplan **F4** y **F5** en el caso de regresión de Poisson. Necesitamos el siguiente lema.

Lema A.2.4 *Supongamos que se cumplen **A1-A2** y **B1-B6**. Sea m definida por (1.12), sea κ_0 como en el Lema A.1.3 y k la constante de calibración de la función ρ utilizada en la definición de m . Entonces*

$$\sqrt{\lambda} - k - \frac{1}{\sqrt{\kappa_0}} < m(\lambda) < \sqrt{\lambda} + k + \frac{1}{\sqrt{\kappa_0}}.$$

Demostración

Observemos primero que, por la desigualdad de Chebyshev, se tiene

$$\begin{aligned}
P\left(\left|\sqrt{y} - \sqrt{\lambda}\right| < \delta\right) &= P\left(\frac{|y - \lambda|}{\sqrt{y} + \sqrt{\lambda}} < \delta\right) \\
&\geq P\left(\frac{|y - \lambda|}{\sqrt{\lambda}} < \delta\right) \\
&\geq 1 - \frac{\text{Var}(y - \lambda)}{\delta^2 \lambda} \\
&= 1 - \frac{1}{\delta^2}.
\end{aligned}$$

Supongamos que $m(\lambda) \leq \sqrt{\lambda} - k - 1/\sqrt{\kappa_0}$. Entonces $|\sqrt{y} - m(\lambda)| < k$ implica

$$m(\lambda) - \sqrt{\lambda} - k < \sqrt{y} - \sqrt{\lambda} < m(\lambda) - \sqrt{\lambda} + k < 0,$$

y, por lo tanto, que $|\sqrt{y} - \sqrt{\lambda}| > -m(\lambda) + \sqrt{\lambda} - k$. Entonces,

$$\begin{aligned}
E_\lambda(\rho(\sqrt{y} - m(\lambda))) &\geq \sum_{\{y/|\sqrt{y}-m(\lambda)| \geq k\}} \rho(\sqrt{y} - m(\lambda))p_\lambda(y) \\
&= 1 - P(|\sqrt{y} - m(\lambda)| < k) \\
&\geq 1 - P\left(|\sqrt{y} - \sqrt{\lambda}| \geq -m(\lambda) + \sqrt{\lambda} - k\right) \\
&> 1 - \frac{1}{\left(-m(\lambda) + \sqrt{\lambda} - k\right)^2} \\
&\geq 1 - \kappa_0.
\end{aligned}$$

Esto es un absurdo por el Lema A.1.3. Para probar la otra desigualdad hacemos algo similar. Supongamos que $m(\lambda) \geq \sqrt{\lambda} + k + 1/\sqrt{\kappa_0}$. Entonces $|\sqrt{y} - m(\lambda)| < k$ implica que

$$0 < m(\lambda) - \sqrt{\lambda} - k < \sqrt{y} - \sqrt{\lambda} < m(\lambda) - \sqrt{\lambda} + k,$$

y, por lo tanto, que $|\sqrt{y} - \sqrt{\lambda}| > m(\lambda) - \sqrt{\lambda} + k$ contradiciendo nuevamente el Lema A.1.3, lo que concluye la demostración del lema. ■

Proposición A.2.1 *Supongamos que se cumple el modelo dado por (1.9) y (1.10) con F_λ la distribución de Poisson. Supongamos que $\omega = 1$ y que la función $\psi = \rho'$ tiene soporte compacto en $[-k, k]$. Entonces (a) una condición suficiente para que se cumpla **C5** (i) es que $E(y|x_k x_j) < \infty$, $1 \leq j, k \leq p$, (b) una condición suficiente para que se cumpla **C5** (ii) es que la matriz $E_{\beta_0}(\mathbf{J}_\Psi(y, \mathbf{x}, \boldsymbol{\beta}_0, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0))$ sea definida positiva.*

Demostración

Observemos primero que, si $y \sim \mathcal{P}(\lambda)$ y llamamos p_λ a la función de probabilidad puntual de y , entonces

$$\frac{d}{d\lambda}p_\lambda(y) = \begin{cases} p_\lambda(y-1) - p_\lambda(y) & \text{si } y \neq 0 \\ -p_\lambda(0) & \text{si } y = 0 \end{cases}. \quad (\text{A.40})$$

De (A.38), (A.40) y (A.39) se deduce que si $y \sim \mathcal{P}(\lambda)$ entonces

$$m'(\lambda) = \frac{E_\lambda(\psi(t(y+1)) - m(\lambda))}{E_\lambda(\psi(t(y)) - m(\lambda))}.$$

Notemos, para abreviar $E_k^i = E_\lambda(\psi^{(i)}(t(y+k) - m(\lambda)))$ donde el supraíndice (i) denota la derivada i -ésima. Entonces mediante un cálculo directo se obtiene

$$m''(\lambda) = \frac{-E_1' E_1 E_0' + E_2 E_0'^2 + E_1 E_0'^2 + E_1^2 E_0'' - E_1 E_1' E_0' + E_1 E_0'^2}{E_0'^3},$$

y con esta notación resulta

$$m'(\lambda) = \frac{E_1}{E_0'}.$$

Por **C2** y **C4** m y m' son acotadas como funciones de λ . Entonces se tiene que

$$\begin{aligned} J_{\Psi}^{j,k}(y, \mathbf{x}, \boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) &= w(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) \left(-\psi' \left(t(y) - m \left(e^{\boldsymbol{\beta}^T \mathbf{x}} \right) \right) \right) \left(m' \left(e^{\boldsymbol{\beta}^T \mathbf{x}} \right) \right)^2 \left(e^{\boldsymbol{\beta}^T \mathbf{x}} \right)^2 x_k x_j \\ &\quad + \psi \left(t(y) - m \left(e^{\boldsymbol{\beta}^T \mathbf{x}} \right) \right) m'' \left(e^{\boldsymbol{\beta}^T \mathbf{x}} \right) \left(e^{\boldsymbol{\beta}^T \mathbf{x}} \right)^2 x_k x_j \\ &\quad + \psi \left(t(y) - m \left(e^{\boldsymbol{\beta}^T \mathbf{x}} \right) \right) m' \left(e^{\boldsymbol{\beta}^T \mathbf{x}} \right) e^{\boldsymbol{\beta}^T \mathbf{x}} x_k x_j. \end{aligned}$$

Por lo tanto, una condición suficiente para que se cumpla **C5** (i) es que

$$E \left(\sup_{\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| \leq \varepsilon} \left| \left(e^{\boldsymbol{\beta}^T \mathbf{x}} \right)^2 x_k x_j \right| + \left| e^{\boldsymbol{\beta}^T \mathbf{x}} x_k x_j \right| \right) < \infty$$

para todo $1 \leq j, k \leq p$. Del Lema A.2.4 se deduce que

$$\lambda^2 < \left(m(\lambda) + k + \frac{1}{\sqrt{\kappa_0}} \right)^4$$

y, por lo tanto,

$$e^{2\boldsymbol{\beta}^T \mathbf{x}} < \left(m(e^{\boldsymbol{\beta}^T \mathbf{x}}) + k + \frac{1}{\sqrt{\kappa_0}} \right)^4$$

Ahora $J_{\Psi}^{j,k}(y, \mathbf{x}, \boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = 0$ si $|t(y) - m(e^{\boldsymbol{\beta}^T \mathbf{x}})| > k$. Por lo tanto, al acotar $|J_{\Psi}^{j,k}(y, \mathbf{x}, \boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\Sigma})|$ podemos asumir que $|t(y) - m(e^{\boldsymbol{\beta}^T \mathbf{x}})| \leq k$, es decir, que $\sqrt{y} - k \leq m(e^{\boldsymbol{\beta}^T \mathbf{x}}) \leq \sqrt{y} + k$. Entonces,

$$\begin{aligned} & \left| \left(e^{\boldsymbol{\beta}^T \mathbf{x}} \right)^2 x_k x_j \right| + \left| e^{\boldsymbol{\beta}^T \mathbf{x}} x_k x_j \right| \\ & \leq \left(\left(m(e^{\boldsymbol{\beta}^T \mathbf{x}}) + k + \frac{1}{\sqrt{\kappa_0}} \right)^4 + \left(m(e^{\boldsymbol{\beta}^T \mathbf{x}}) + k + \frac{1}{\sqrt{\kappa_0}} \right)^2 \right) |x_k x_j| \\ & \leq \left(\left(\sqrt{y} + 2k + \frac{1}{\sqrt{\kappa_0}} \right)^4 + \left(\sqrt{y} + 2k + \frac{1}{\sqrt{\kappa_0}} \right)^2 \right) |x_k x_j|. \end{aligned}$$

Luego es fácil ver que existen constantes M_1 y M_2 tales que

$$\left| \left(e^{\boldsymbol{\beta}^T \mathbf{x}} \right)^2 x_k x_j \right| + \left| e^{\boldsymbol{\beta}^T \mathbf{x}} x_k x_j \right| \leq M_1 y^2 |x_j x_k| + M_2.$$

Para que se cumpla **C5** (i) es suficiente entonces que

$$E_{\beta_0}(y^2 x_k x_j) < \infty \text{ para } 1 \leq j, k \leq p.$$

Veamos ahora que la matriz Jacobiana de Ψ respecto de $\boldsymbol{\beta}$ es definida positiva. Claramente, esto implica que es no singular. Observemos que, por A.38,

$$\begin{aligned} & E_{\beta_0}(\mathbf{J}_{\Psi}(y, \mathbf{x}, \boldsymbol{\beta}_0, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)) \\ & = E_{\beta_0} \left(w(\mathbf{x}, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) \left(-\psi' \left(t(y) - m \left(e^{\boldsymbol{\beta}_0^T \mathbf{x}} \right) \right) \right) \left(m' \left(e^{\boldsymbol{\beta}_0^T \mathbf{x}} \right) \right)^2 \left(e^{\boldsymbol{\beta}_0^T \mathbf{x}} \right)^2 \mathbf{xx}^T \right) \\ & = E \left(E_{\beta_0} \left(-\psi' \left(t(y) - m \left(e^{\boldsymbol{\beta}_0^T \mathbf{x}} \right) \right) \right) \middle| \mathbf{x} \right) w(\mathbf{x}, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) \left(m' \left(e^{\boldsymbol{\beta}_0^T \mathbf{x}} \right) \right)^2 \left(e^{\boldsymbol{\beta}_0^T \mathbf{x}} \right)^2 \mathbf{xx}^T. \end{aligned}$$

Por lo tanto, usando **A2**, **B7** y **C4** se tiene que

$$\begin{aligned} & \mathbf{a}^T E(\mathbf{J}_{\Psi}(y, \mathbf{x}, \boldsymbol{\beta}_0, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)) \mathbf{a} \\ & = E \left(E_{\beta_0} \left(\left(-\psi' \left(t(y) - m \left(e^{\boldsymbol{\beta}_0^T \mathbf{x}} \right) \right) \right) \middle| \mathbf{x} \right) w(\mathbf{x}, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) \left(m' \left(e^{\boldsymbol{\beta}_0^T \mathbf{x}} \right) \right)^2 e^{2\boldsymbol{\beta}_0^T \mathbf{x}} \mathbf{a}^T \mathbf{xx}^T \mathbf{a} \right) \\ & > 0. \end{aligned}$$

y, por lo tanto, **C5** se cumple. ■

A.3. Punto de ruptura asintótico

En esta sección estudiamos el punto de ruptura asintótico de los estimadores propuestos en (A.1). Consideraremos solamente el caso en que Φ sólo depende de

y, \mathbf{x} y $\boldsymbol{\beta}$. Es decir

$$\widehat{\boldsymbol{\beta}}_n = \arg \min_{\boldsymbol{\beta} \in \mathbb{R}^p} \sum_{i=1}^n \Phi(y_i, \mathbf{x}_i, \boldsymbol{\beta}). \quad (\text{A.41})$$

Como la mayoría de los estimadores conocidos, estos estimadores pueden verse como funcionales. Sea (y, \mathbf{x}) un vector aleatorio en $\mathbb{R} \times \mathbb{R}^p$, \mathcal{D} el conjunto de todas las distribuciones en $\mathbb{R} \times \mathbb{R}^p$ y $\mathbf{T} : \mathcal{D} \rightarrow \mathbb{R}^p$ el funcional definido por

$$\mathbf{T}(H) = \arg \min_{\boldsymbol{\beta} \in \mathbb{R}^p} E_H(\Phi(y, \mathbf{x}, \boldsymbol{\beta})). \quad (\text{A.42})$$

Dada una muestra $(y_1, \mathbf{x}_1), \dots, (y_n, \mathbf{x}_n)$, correspondiente al modelo (1.9)-(1.10), sea H_n su distribución empírica. Luego $\widehat{\boldsymbol{\beta}}_n = \mathbf{T}(H_n)$. Recordemos que el punto de ruptura asintótico del funcional \mathbf{T} en $H_0 \in \mathcal{D}$ se define como

$$\varepsilon^*(\mathbf{T}, H_0) = \sup\{\varepsilon \in (0, 1) : \sup_{H^* \in \mathcal{D}} \{\mathbf{T}(1 - \varepsilon)H_0 + \varepsilon H^*\} < \infty\}. \quad (\text{A.43})$$

El siguiente teorema da una cota inferior para el punto de ruptura asintótico de este funcional.

Teorema A.3.1 *Sea (\mathbf{x}, y) un vector aleatorio con distribución H_0 tal que $P_{H_0}(\boldsymbol{\alpha}^T \mathbf{x} = 0) = 0$ para todo $\boldsymbol{\alpha} \in \mathbb{R}^p$. Supongamos $y|\mathbf{x} \sim F_{g^{-1}(\boldsymbol{\beta}_0^T \mathbf{x})}$. Sea \mathbf{T} el funcional definido por (1.25) donde Φ verifica las condiciones **P0-P3**. Sea*

$$\varepsilon_0 = \frac{E_{H_0}(\min(\Phi^*(y, \mathbf{x}, 1), \Phi^*(y, \mathbf{x}, -1))) - E_{H_0}(\Phi(y, \mathbf{x}, \boldsymbol{\beta}_0))}{E_{H_0}(\min(\Phi^*(y, \mathbf{x}, 1), \Phi^*(y, \mathbf{x}, -1))) - E_{H_0}(\Phi(y, \mathbf{x}, \boldsymbol{\beta}_0))} + M$$

con $M = \sup_{y, \mathbf{x}, \boldsymbol{\beta}} \Phi(y, \mathbf{x}, \boldsymbol{\beta})$. Entonces

$$\varepsilon^*(\mathbf{T}, H_0) \geq \varepsilon_0.$$

Demostración

Supongamos que ε es un número real positivo para el cual existe una sucesión de funciones de distribución H_k tales que $\|\mathbf{T}((1 - \varepsilon)H_0 + \varepsilon H_k)\| \rightarrow \infty$ cuando $k \rightarrow \infty$. Mostraremos que $\varepsilon \geq \varepsilon_0$. Escribimos $\boldsymbol{\beta}_k = \mathbf{T}((1 - \varepsilon)H_0 + \varepsilon H_k)$ para abreviar. Entonces se tiene

$$\begin{aligned} (1 - \varepsilon)E_{H_0}(\Phi(y, \mathbf{x}, \boldsymbol{\beta}_k)) &\leq (1 - \varepsilon)E_{H_0}(\Phi(y, \mathbf{x}, \boldsymbol{\beta}_k) + \varepsilon E_{H_k}(\Phi(y, \mathbf{x}, \boldsymbol{\beta}_k))) \\ &\leq (1 - \varepsilon)E_{H_0}(\Phi(y, \mathbf{x}, \boldsymbol{\beta}_0) + \varepsilon E_{H_k}(\Phi(y, \mathbf{x}, \boldsymbol{\beta}_0))) \\ &\leq (1 - \varepsilon)E_{H_0}(\Phi(y, \mathbf{x}, \boldsymbol{\beta}_0) + \varepsilon M). \end{aligned} \quad (\text{A.44})$$

Sea $\alpha_k = \beta_k / \|\beta_k\|$, entonces podemos asumir sin pérdida de generalidad que $\alpha_k \rightarrow \alpha$. Luego tenemos

$$\begin{aligned} & \lim_{k \rightarrow \infty} (1 - \varepsilon) E_{H_0}(\Phi(y, \mathbf{x}, \beta_k)) \\ &= \lim_{k \rightarrow \infty} (1 - \varepsilon) E_{H_0} [\Phi(y, \mathbf{x}, \beta_k) I(\alpha^T \mathbf{x} < 0) + \Phi(y, \mathbf{x}, \beta_k) I(\alpha^T \mathbf{x} > 0)] \\ &\geq (1 - \varepsilon) E_{H_0}(\min(\Phi^*(y, \mathbf{x}, 1), \Phi^*(y, \mathbf{x}, -1))). \end{aligned} \tag{A.45}$$

Combinando las desigualdades (A.44) y (A.45) se obtiene

$$(1 - \varepsilon) E_{H_0}(\min(\Phi^*(y, \mathbf{x}, 1), \Phi^*(y, \mathbf{x}, -1))) \leq (1 - \varepsilon) E_{H_0}(\Phi(y, \mathbf{x}, \beta_0)) + \varepsilon M,$$

y entonces

$$\varepsilon \geq \frac{E_{H_0}(\min(\Phi^*(y, \mathbf{x}, 1), \Phi^*(y, \mathbf{x}, -1))) - E_{H_0}(\Phi(y, \mathbf{x}, \beta_0))}{E_{H_0}(\min(\Phi^*(y, \mathbf{x}, 1), \Phi^*(y, \mathbf{x}, -1))) - E_{H_0}(\Phi(y, \mathbf{x}, \beta_0)) + M}.$$

Esto prueba el teorema. ■

El Teorema 1.6.1 es un caso particular de este último teorema y su demostración es inmediata.

A continuación demostramos los resultados necesarios para calcular ε_0 en los casos de regresión exponencial y de Poisson.

Lema A.3.1 *Si $y \sim P(\lambda)$ y $t(\lambda) = \sqrt{\lambda}$ entonces $\sup_{\lambda} V(t(y)) \leq 1$*

Demostración

Tenemos que

$$\begin{aligned} V(t(y)) &= E_{\lambda} \left((\sqrt{y} - E_{\lambda}(\sqrt{y}))^2 \right) \\ &\leq E \left((\sqrt{y} - \sqrt{\lambda})^2 \right) = E \left(\left(\frac{y - \lambda}{\sqrt{y} + \sqrt{\lambda}} \right)^2 \right) \\ &\leq E \left(\left(\frac{y - \lambda}{\sqrt{\lambda}} \right)^2 \right) = 1. \end{aligned}$$

■

Lema A.3.2 *Si $y \sim P(\lambda)$ entonces $m_1 = 0$ y $m_2 = +\infty$.*

Demostración

El Lema A.2.4 implica inmediatamente que $m_2 = +\infty$. Veamos ahora que $m_1 = 0$. Sea λ_n una sucesión de números reales positivos tales que $\lim_{n \rightarrow \infty} \lambda_n = 0$.

Por el Lema A.2.4 podemos asumir que $m(\lambda_n)$ converge a un punto $m_1 \in \mathbb{R}_{\geq 0}$. Por definición de m tenemos

$$E_{\lambda_n}(\rho(\sqrt{y}) - m(\lambda_n)) \leq E_{\lambda_n}(\rho(\sqrt{y})). \quad (\text{A.46})$$

Usando que $\rho \leq 1$ y la función de probabilidad puntual de las variables Poisson obtenemos

$$e^{-\lambda} \rho(-m(\lambda)) \leq E_{\lambda}(\rho(\sqrt{y} - m(\lambda))) \leq e^{-\lambda} \rho(-m(\lambda)) + 1 - e^{-\lambda}, \quad (\text{A.47})$$

y

$$E_{\lambda}(\rho(\sqrt{y})) \leq 1 - e^{-\lambda}. \quad (\text{A.48})$$

Las ecuaciones (A.47) y (A.48) implican que

$$\lim_{\lambda \rightarrow 0} E_{\lambda}(\rho(\sqrt{y} - m(\lambda))) = \rho(-m_1) \quad \text{y} \quad \lim_{\lambda \rightarrow 0} E_{\lambda}(\rho(\sqrt{y})) = 0.$$

Tomando límites en (A.46) y usando que ρ es par y no negativa se obtiene $\rho(m_1) = 0$ y, por lo tanto, $m_1 = 0$. ■

Lema A.3.3 *Si $y \sim \mathcal{E}(\lambda)$ entonces $m_1 = -\infty$ y $m_2 = +\infty$.*

Demostración

Como $y/\lambda \sim \mathcal{E}(1)$ entonces

$$P(|\log y - \log \lambda| < \delta) = c_{\delta}, \quad (\text{A.49})$$

donde $c_{\delta} = e^{-e^{-\delta}} - e^{-e^{\delta}}$ es una constante positiva independiente de λ que tiende a uno cuando δ tiende a infinito. Sea δ_0 tal que $c_{\delta} > 1 - \kappa_0$ para todo $\delta \geq \delta_0$. Veremos que

$$\log \lambda - k - \delta_0 < m(\lambda) < \log \lambda + k + \delta_0. \quad (\text{A.50})$$

Supongamos que $m(\lambda) \leq \log \lambda - k - \delta_0$. Entonces $|\log y - m(\lambda)| < k$ implica

$$m(\lambda) - \log \lambda - k < \log y - \log \lambda < m(\lambda) - \log \lambda + k < 0.$$

Entonces

$$\begin{aligned} E_{\lambda}(\rho(\log y - m(\lambda))) &\geq 1 - P(|\log y - m(\lambda)| < k) \\ &\geq 1 - P(|\log y - \log \lambda| \geq \log \lambda - m(\lambda) - k) \\ &= c_{\log \lambda - m(\lambda) - k} \\ &\geq 1 - \kappa_0. \end{aligned}$$

Esto es un absurdo por el Lema A.1.3. Para mostrar la otra desigualdad supongamos que $m(\lambda) \geq \log \lambda + k + \delta$. Entonces $|\log y - m(\lambda)| < k$ implica

$$0 < m(\lambda) - \log \lambda - k < \log y - \log \lambda < m(\lambda) - \log \lambda + k.$$

Luego

$$\begin{aligned} E_\lambda(\rho(\log y - m(\lambda))) &\geq 1 - P(|\log y - m(\lambda)| < k) \\ &\geq 1 - P(|\log y - \log \lambda| \geq m(\lambda) - \log \lambda - k) \\ &= c_{\log \lambda - m(\lambda) - k} \\ &\geq 1 - \kappa_0 \end{aligned}$$

contradiciendo nuevamente el Lema A.1.3. Esto prueba (A.50). Como δ_0 no depende de λ (A.50) implica que $m_1 = -\infty$ y $m_2 = +\infty$. ■

A.4. Método computacional

Proposición A.4.1 *Si y tiene distribución $\mathcal{P}(\lambda)$ entonces $y^{1/2} - \lambda^{1/2} \xrightarrow{d} N(0, 1/4)$ cuando $\lambda \rightarrow \infty$.*

Demostración

Mediante una expansión de Taylor se obtiene

$$y^{1/2} - \lambda^{1/2} = \frac{y - \lambda}{2\sqrt{\lambda}} - \frac{(y - \lambda)^2}{4\xi^{3/2}}.$$

donde ξ es un punto intermedio entre y y λ . El primer término tiende en distribución a una variable aleatoria $N(0, 1/4)$. Para ver que el segundo término tiende a cero en probabilidad cuando λ tiende a ∞ basta ver que tanto $(y - \lambda)^2/4\lambda^{3/2}$ como $(y - \lambda)^2/4y^{3/2}$ tienden cero en probabilidad cuando λ tiende a ∞ . Por la desigualdad de Chebyshev

$$\begin{aligned} P\left(\frac{(y - \lambda)^2}{4\lambda^{3/2}} > \varepsilon\right) &= P(|y - \lambda| > 2\lambda^{3/4}\varepsilon^{1/2}) \\ &\leq \frac{1}{4\lambda^{3/2}\varepsilon} \rightarrow 0 \text{ cuando } \lambda \rightarrow \infty, \end{aligned}$$

y luego $(y - \lambda)^2/4\lambda^{3/2} \rightarrow 0$ cuando $\lambda \rightarrow \infty$. Sea $k \in (0, 1)$, entonces

$$\begin{aligned}
 P\left(\frac{(y - \lambda)^2}{4y^{3/2}} > \varepsilon\right) &< P\left(\frac{(y - \lambda)^2}{4y^{3/2}} > \varepsilon \cap \left|\frac{y - \lambda}{\lambda}\right| < k\right) + P\left(\left|\frac{y - \lambda}{\lambda}\right| \geq k\right) \\
 &\leq P\left(\frac{(y - \lambda)^2}{4\lambda^{3/2}(1 - k)^{3/2}} > \varepsilon\right) + \frac{1}{\lambda k^2} \\
 &= P(|y - \lambda| > 2\lambda^{3/4}\varepsilon^{1/2}(1 - k)^{3/4}) + \frac{1}{\lambda k^2} \\
 &\leq \frac{1}{4\lambda^{1/2}\varepsilon(1 - k)^{3/2}} + \frac{1}{\lambda k^2},
 \end{aligned}$$

y, por lo tanto, $(y - \lambda)^2/4y^{3/2} \rightarrow 0$ cuando $\lambda \rightarrow \infty$.

Usando el Lemma de Slutsky, se llega a que $y^{1/2} - \lambda^{1/2} \xrightarrow{d} N(0, 1/4)$ cuando $\lambda \rightarrow \infty$, como queríamos demostrar. ■

Apéndice B

Demostraciones de los Teoremas del capítulo 2

B.1. Consistencia

Demostración del Teorema 2.3.2

La afirmación (a) se deduce inmediatamente de (2.5) y (2.6). Para probar (b) veamos primero que, para todo k del rango de x y $\theta_1 < \theta_2$, se tiene

$$\psi_I(k, \theta_1) \geq \psi_I(k, \theta_2). \quad (\text{B.1})$$

Esto es inmediato para variables aleatorias continuas. Si x es discreta, tenemos

$$\psi_I(k, \theta) = P_\theta(x < k) + \frac{1}{2}p_\theta(k) - \frac{1}{2}.$$

Por hipótesis se tiene que

$$P_{\theta_1}(x < k) + p_{\theta_1}(k) = F_{\theta_1}(k) \geq F_{\theta_2}(k) = P_{\theta_2}(x < k) + p_{\theta_2}(k). \quad (\text{B.2})$$

Además, como en general $P_\theta(x < k) = \lim_{t \rightarrow k^-} F_\theta(t)$, se tiene que

$$P_{\theta_1}(x < k) \geq P_{\theta_2}(x < k). \quad (\text{B.3})$$

Observemos que (B.1) es equivalente a

$$P_{\theta_1}(x < k) + \frac{1}{2}p_{\theta_1}(k) \geq P_{\theta_2}(x < k) + \frac{1}{2}p_{\theta_2}(k). \quad (\text{B.4})$$

Luego, la ecuación (B.4) es inmediata de (B.3) si $p_{\theta_1}(k) \geq p_{\theta_2}(k)$. En caso contrario, por (B.2) se tiene que

$$P_{\theta_1}(x < k) - P_{\theta_2}(x < k) \geq p_{\theta_2}(k) - p_{\theta_1}(k) > \frac{1}{2}(p_{\theta_2}(k) - p_{\theta_1}(k)), \quad (\text{B.5})$$

y también (B.4) se cumple. Supongamos ahora que existe k_0 en el rango de x tal que $F_{\theta_1}(k_0) > F_{\theta_2}(k_0)$. Entonces, como F_θ es continua a derecha, existe $\varepsilon > 0$ tal que $F_{\theta_1}(k) > F_{\theta_2}(k)$ para todo $k \in [k_0, k_0 + \varepsilon]$. Por lo tanto, para esos valores de k , la desigualdad (B.2) es estricta y, por lo tanto, la desigualdad (B.5) también lo es. Se llega así a que $\psi(k, \theta_1) > \psi(k, \theta_2)$ para todo $k \in [k_0, k_0 + \varepsilon]$. Esto prueba (b). ■

B.2. Punto de ruptura asintótico

Demostración del Teorema 2.5.1

Consideremos una distribución F de la forma $F = (1 - \varepsilon) F_\theta + \varepsilon H$. Se tiene

$$\begin{aligned} 0 = E_F(\psi(x, T(F))) &= (1 - \varepsilon) E_\theta(\psi(x, T(F))) + \varepsilon E_H(\psi(x, T(F))) \\ &\leq (1 - \varepsilon) E_\theta(\psi(x, T(F))) + \varepsilon k_1. \end{aligned} \quad (\text{B.6})$$

Supongamos que existe una sucesión de funciones de distribución F_n tal que $T(F_n) \rightarrow \theta_2$. Entonces $E_\theta(\psi(x, T(F_n))) \rightarrow -E_\theta(K_2(x))$. Tomando límite en (B.6) se llega a que

$$-(1 - \varepsilon) E_\theta(K_2(x)) + \varepsilon k_1 \geq 0,$$

y esto implica que

$$\varepsilon \geq \frac{E_\theta K_2(x)}{E_\theta K_2(x) + k_1}.$$

Un razonamiento análogo permite calcular una cota inferior para el punto de ruptura a θ_1 . Se tiene que

$$\begin{aligned} 0 = E_F(\psi_I(x, T(F))) &= (1 - \varepsilon) E_\theta(\psi_I(x, T(F))) + \varepsilon E_H(\psi_I(x, T(F))) \\ &\geq (1 - \varepsilon) E_\theta(\psi_I(x, T(F))) - \varepsilon k_2. \end{aligned} \quad (\text{B.7})$$

Supongamos que existe una sucesión de funciones de distribución F_n tal que $T(F_n) \rightarrow \theta_1$. Entonces $E_\theta(\psi(x, T(F_n))) \rightarrow E_\theta(K_1(x))$ y tomando límite en (B.7) se llega a que

$$(1 - \varepsilon) E_\theta(K_1(x)) - \varepsilon k_2 \leq 0,$$

y, por lo tanto,

$$\varepsilon \geq \frac{E_\theta K_1(x)}{E_\theta K_1(x) + k_2}.$$

■

Demostración del Teorema 2.6.1

Se tiene

$$\begin{aligned} 0 &= E_{F_n} \psi_I(x, T(F_n)) = E_\theta \psi_I(x, T(F_n)) (1 - \varepsilon) + E_{H_n}(\psi_I(x, T(F_n))) \varepsilon \\ &= (1 - \varepsilon) E_\theta \psi_I(x, T(F_n)) + \varepsilon \sum_{x=1}^{\infty} (F_{T(F_n)}(x) - \frac{1}{2} p_{T(F_n)}(x) - \frac{1}{2}) p_{H_n}(x) \\ &\quad + \varepsilon \left(\frac{1}{2} p_{T(F_n)}(0) - \frac{1}{2} \right) p_{H_n}(0). \end{aligned}$$

Veamos primero que $\sum_{x=1}^{\infty} p_{H_n}(x) \rightarrow 0$. Supongamos que esto no es cierto, entonces existe $\delta > 0$ tal que $\sum_{x=1}^{\infty} p_{H_n}(x) > \delta$. Para $x > 0$, como $T(F_n) \rightarrow 0$, tenemos

$$F_{T(F_n)}(x) - \frac{1}{2}p_{T(F_n)}(x) - \frac{1}{2} \rightarrow \frac{1}{2},$$

y luego existe n_0 tal que para, $n \geq n_0$,

$$F_{T(F_n)}(x) - \frac{1}{2}p_{T(F_n)}(x) - \frac{1}{2} > \frac{1}{4}.$$

Observemos que

$$\lim_{\theta \rightarrow 0} \psi_I(x, \theta) = \begin{cases} \frac{1}{2} & \text{si } x > 0 \\ 0 & \text{si } x = 0 \end{cases}$$

Luego, obtenemos que

$$\begin{aligned} 0 &> (1 - \varepsilon) E_{\theta} \psi_I(x, T(F_n)) + \varepsilon \frac{1}{4} \delta + \varepsilon \left(\frac{1}{2} p_{T(F_n)}(0) - \frac{1}{2} \right) p_{H_n}(0) \\ &\rightarrow (1 - \varepsilon) \frac{1}{2} (1 - e^{-\theta}) + \frac{\varepsilon}{4} \delta > 0 \end{aligned}$$

pues $T(F_n) \rightarrow 0$. Esto es un absurdo. Por lo tanto, $P_{H_n}(x \geq 1) = \sum_{x=1}^{\infty} p_{H_n}(x) \rightarrow 0$.

Como

$$-\frac{1}{2} \sum_{x=1}^{\infty} p_{H_n}(x) \leq \sum_{x=1}^{\infty} \left(F_{T(F_n)}(x) - \frac{1}{2} p_{T(F_n)}(x) - \frac{1}{2} \right) p_{H_n}(x) \leq \frac{1}{2} \sum_{x=1}^{\infty} p_{H_n}(x),$$

usando de nuevo que $T(F_n) \rightarrow 0$, se tiene que

$$\begin{aligned} 0 &= \lim_{n \rightarrow \infty} \left[(1 - \varepsilon) E_{\theta} \psi_I(x, T(F_n)) + \varepsilon \sum_{x=1}^{\infty} \left(F_{T(F_n)}(x) - \frac{1}{2} p_{T(F_n)}(x) - \frac{1}{2} \right) p_{H_n}(x) \right. \\ &\quad \left. + \varepsilon \left(\frac{1}{2} p_{T(F_n)}(0) - \frac{1}{2} \right) p_{H_n}(0) \right] \\ &= (1 - \varepsilon) \frac{1}{2} (1 - e^{-\theta}), \end{aligned}$$

y, por lo tanto, $\varepsilon = 1$. ■

Bibliografía

- [1] Bergesio, A. and Yohai, V. J. (2011). “Projection estimators for generalized linear models”. *Journal of the American Statistical Association*, 106, 661-671.
- [2] Bianco, A., García Ben M. y Yohai V.J.(2005). “Robust estimation for regression with asymmetric errors”, *Canadian Journal of Statistics*, 33, 511-528
- [3] Bianco, A. y Yohai, V. J. (1996), “Robust Estimation in the Logistic Regression Model”, in *Robust Statistics, Data Analysis and Computer Intensive Methods, Proceedings of the workshop in honor of Peter Huber*, ed. H. Rieder, *Lecture Notes in Statistics*, 109, 7–34, New York, Springer-Verlag.
- [4] Bianco, A. M., Boente, G. y Rodrigues, I. M. (2013). “Resistant Estimators in Poisson and Gamma models with missing responses and an application to outlier detection” *Journal of Multivariate Analysis*, 114, 209-226,
- [5] Bondell, H. D. (2005), “Minimum Distance Estimation for the Logistic Regression Model,” *Biometrika*, 92, 724–731.
- [6] Breslow N. E. (1996) “Generalized linear models: Checking assumptions and strengthening conclusions”. *Statistica Applicata* 8, 23–41.
- [7] Cantoni, E. y Ronchetti, E. (2001), “Robust Inference for Generalized Linear Models,” *Journal of the American Statistical Association*, 96, 1022–1030
- [8] Carroll, R. J., y Pederson, S. (1993), “On Robustness in the Logistic Regression Model,” *Journal of the Royal Statistical Society, Ser. B*, 55, 693–706.
- [9] Christmann, A. (1994). “Least median of weighted squares in logistic regression with large strata” *Biometrika*, 81, 413-417.

- [10] Čížek, P. (2008), “Robust and Efficient Adaptive Estimation of Binary-Choice Regression Models”, *Journal of the American Statistical Association*, 103, 687–696.
- [11] Croux, C. y Haesbroeck, G. (2003), “Implementing the Bianco and Yohai Estimator for Logistic Regression,” *Computational Statistics and Data Analysis*, 44, 273–295.
- [12] Ferguson, T.S. (1967), *Mathematical Statistics. A decision theoretic approach*, Academic Press.
- [13] Hampel, F.R. (1974), ”The influence curve and its role in robust estimation ”, The influence curve and its role in robust estimation ”. *Journal of the American Statistical Association*, 69, 383-93.
- [14] Huber P.J. (1981), *Robust Statistics*, Wiley.
- [15] Künsch, H., Stefanski, L. y Carroll, R. (1989), “Conditionally Unbiased Bounded-Influence Estimation in General Regression Models, with Applications to Generalized Linear Models,” *Journal of the American Statistical Association*, 84, 460–466.
- [16] Maronna, R. A., Bustos, O. y Yohai, V. J. (1979), “Bias and Efficiency Robustness of General M-estimators for Regression with Random Carriers,” in *Smoothing techniques for curve estimation*, eds. T. Gasser and M. Rosenblatt, Lectures Notes in Mathematics No. 757, 91–111, New York, Springer Verlag.
- [17] Maronna, R. A., Martin, R. D. y Yohai, V. J. (2006), *Robust Statistics: Theory and Methods*, Chichister, Wiley.
- [18] Morgenthaler, S. (1992). “Least-absolute-deviations for generalized linear models”. *Biometrika*, 79, 747-754.
- [19] Rousseeuw, P. J. y Van Zomeren, B. C. (1990). “Unmasking multivariate outliers and leverage points”. *Journal of the American Statistical Association*, 85, 633–651
- [20] Rousseeuw, P.J. y Christmann, A. (2003). “Robustness against separation and outliers in logistic regression”. *Computational Statistics and Data Analysis*, 43, 315-332.

- [21] Wedderburn, R. W. M. (1974), “Quasi-likelihood Functions, Generalized Linear Models, and the Gauss–Newton Method,” *Biometrika*, 61, 439–447.
- [22] Yohai, V.J.(1985) “High Breakdown Point and High Efficiency Robust Estimates for Regression,” Technical Report No. 66. Department of Statistics, University of Washington. Available at <http://www.stat.washington.edu/research/reports/1985/tr066.pdf>.