Biblioteca Digital
F C E N - U B A

BIBLIOTECA CENTRAL LUIS F LELOIR
BIBLIOTECA CENTRAL LELOIR
FACULTAD DE CIENCIAS EXACTAS Y NATURALES UBA

Tesis de Maestría

Un S-estimador alternativo para el modelo de regresión lineal con datos censurados y distribución posiblemente asimétrica

Fragalá, Marina

2016-11-04

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

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

Cita tipo APA:
Fragalá, Marina. (2016-11-04). Un S-estimador alternativo para el modelo de regresión lineal con datos censurados y distribución posiblemente asimétrica. Facultad de Ciencias Exactas y Naturales. Universidad de Buenos Aires.

Cita tipo Chicago:
Fragalá, Marina. "Un S-estimador alternativo para el modelo de regresión lineal con datos censurados y distribución posiblemente asimétrica". Facultad de Ciencias Exactas y Naturales. Universidad de Buenos Aires. 2016-11-04.

EXACTAS UBA
Facultad de Ciencias Exactas y Naturales

UBA
Universidad de Buenos Aires

Dirección: Biblioteca Central Dr. Luis F. Leloir, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires. Intendente Güiraldes 2160 - C1428EGA - Tel. (++54 +11) 4789-9293

Contacto: digital@bl.fcen.uba.ar

UNIVERSIDAD DE BUENOS AIRES

Facultad de Ciencias Exactas y Naturales

Departamento de Matemática e Instituto de Cálculo

Un S-estimador alternativo para el modelo de regresión lineal con datos censurados y distribución posiblemente asimétrica

Lic. Marina Fragalá

Director: Dr. Víctor Yohai

Tesis de Maestría en Estadística Matemática

Buenos Aires, 4 de noviembre de 2016

Resumen

En esta tesis comenzamos haciendo una revisión detallada de resultados sobre los S estimadores para el modelo de regresión lineal con datos no censurados y errores que pueden tener distribución simétrica o asimétrica. Para estos estimadores aportamos una prueba diferente de la propiedad de consistencia de Fisher y estudiamos la consistencia fuerte a partir del concepto de funcionales de regresión.

Locatelli, Marazzi y Yohai (2010) desarrollaron S-estimadores para el modelo de regresión lineal con datos censurados y errores cuya distribución pertence a una familia paramétrica de posición y escala que también puede ser simétrica o asimétrica. En esta tesis proponemos y analizamos un nuevo procedimiento de estimación de tipo S para el mismo problema. Estos estimadores son altamente robustos pero ineficientes. Para mejorar la eficiencia, a partir de ellos definimos un estimador final utilizando un procedimiento de máxima verosimilitud truncada. Este estimador resulta altamente robusto y con una eficiencia asintótica alta si se compara con la del estimador de máxima verosimilitud para datos censurados. Las ventajas de este nuevo S-estimador para datos censurados sobre el propuesto en Locatelli et al. (2010) son las siguientes: (a) gracias a su estructura, podemos probar que el nuevo S-estimador tiene la propiedad de consistencia de Fisher y (b) el cálculo del nuevo S-estimador requiere un algoritmo de menor complejidad del que se usa para el S-estimador propuesto en Locatelli et al. (2010).

A mi hija Chiara,

por iluminar cada día de mi vida.

Quiero darle un especial agradecimiento a mi director Víctor Yohai, por sus enseñanzas a lo largo de estos años y por haberme guiado en la construcción de esta tesis.

Índice

1. Introducción **11**

Organización de la tesis 11

2. Datos censurados **12**

3. Modelo y notaciones **14**

Condición de identificabilidad 17

Esperanza condicional y distribución empírica para datos censurados 17

4. Estimadores de máxima verosimilitud para datos censurados **20**

5. S-estimadores de regresión lineal con datos no censurados **21**

Reparametrización del modelo 25

Definición clásica de S-estimador 26

Consistencia de Fisher 28

Consistencia fuerte del S-estimador cuando $\mu_0 = 0$ 29

Consistencia fuerte del S-estimador cuando $\mu_0 \neq 0$ 29

Formulación alternativa para los S-estimadores cuando $\mu_0 \neq 0$ 30

6. S-estimadores de regresión lineal para datos censurados propuestos por Locatelli, Marazzi y Yohai **31**

7. Nueva propuesta de S-estimadores de regresión lineal con datos censurados **34**

Definición de los nuevos estimadores 34

Consistencia de Fisher 38

8. Estimadores finales por máxima verosimilitud truncada **39**

9. Resultados numéricos **43**

Algoritmo para el cálculo de los S-estimadores alternativos 44

Resultados para el caso sin *outliers* 45

Comportamiento del S-estimador en presencia de outliers 46

10. Ejemplo **54**

11. Apéndice de la Sección 2 **66**

12. Apéndice de la Sección 3 **69**

13. Apéndice de la Sección 4 **71**

14. Apéndice de la Sección 5 **72**

S funcionales 82

Consistencia fuerte de estimadores definidos a partir de funcionales 84

Formulación alternativa para los S-estimadores de regresión 92

15. Apéndice de la Sección 6 **92**

16. Apéndice de la Sección 7 **94**

17. Apéndice de la Sección 9 **98**

1. Introducción

Un tópico importante en muchas áreas como por ejemplo la biomedicina, ingienería y ciencias sociales entre otras, es el análisis estadístico para datos que miden lo que se denomina "tiempo de vida útil", como en el caso de la durabilidad de artículos manufacturados. También "tiempo de vida" o "sobrevida" como por ejemplo, los períodos de tratamientos en enfermedades, etc.

Las situaciones a considerar son aquellas en las cuales el tiempo hasta la ocurrencia de algún evento para los individuos de una población, es el motivo de interés. Dichos eventos pueden ser muertes reales de individuos, reaparición de una enfermedad, curación, falla de un componente electrónico, etc. Por "tiempo de vida" entenderemos el período de tiempo hasta la ocurrencia del evento, que se denomina "falla" o "muerte", medido desde algún momento en particular.

El tiempo necesario para observar la "supervivencia" de todos los individuos en un estudio podría ser suficientemente largo como para que determinadas limitaciones prácticas impidiesen la observación completa, sólo pudiendo observarse parcialmente. Esto conduce a lo que se denomina "censura" en cuyo caso sólo se sabrá que el "tiempo de vida" del individio, excede un cierto valor (en este caso la censura se dice a derecha).

El modelo que se suele usar para este tipo de datos es el que se denomina *accelerated failure time regression* el cual asume una relación log-lineal entre la respuesta y las variables explicatorias cuando la distribución de los errores pertencece a una familia paramétrica dada (que puede ser simétrica o asimétrica).

Lawless (2003) obtuvo una expresión para las ecuaciones de máxima verosimilitud con datos censurados. Locatelli, Marazzi y Yohai (2010) dieron la versión paramétrica del S-estimador para datos censurados.

Los S-estimadores tienen un alto punto de ruptura, pero suelen ser bastante ineficientes comparados con el estimador de máxima verosimilitud. Por tal motivo, Locatelli, Marazzi y Yohai (2010) desarrollaron un estimador final para datos censurados conocido como estimador de máxima verosimilitud truncada, el cual mantiene el alto punto de ruptura del estimador inicial y la eficiencia asintótica del estimador de máxima verosimilitud.

Organización de la tesis

Esta tesis se divide en 10 secciones con 7 apéndices. En las dos primeras secciones introducimos el concepto de datos censurados y desarrollamos el modelo de regresión sobre el cual trabajaremos. Los errores serán variables aleatorias continuas con distribución simétrica o asimétrica pertenenciente a una familia paramétrica de posición y escala. Repasamos para ello, el concepto y propiedades de esperanza condicional al vector observado. Finalmente hacemos una revisión rápida del estimador clásico de máxima verosimilitud para datos censurados.

En la Sección 5 recordamos la definición de S-estimador para datos no censurados para luego probar su propiedad de consistencia fuerte utilizando el concepto de conti-

nuidad débil de un funcional. Marazzi, Villar y Yohai (2009) demostraron la propiedad de consistencia de Fisher en otro contexto teórico. En esta tesis damos una prueba diferente de dicha propiedad. Mostramos dos formas distintas de obtener S-estimadores consistentes cuando los errores son asimétricos.

En la Sección 6 estudiamos los S-estimadores de regresión lineal para datos censurados propuestos por Locatelli, Marazzi y Yohai (2010).

En la Sección 7, desarrollamos un procedimiento para obtener, a partir de las ecuaciones de los S-estimadores existentes para datos censurados, una definición alternativa de S-estimador. No hay prueba aún de que los S-estimadores para datos censurados sean consistentes ni que valga la consistencia de Fisher. En el presente trabajo probamos, gracias a la estructura del nuevo S-estimador, la consistencia de Fisher del mismo. En virtud de dicha estructura, el algoritmo para el cálculo del nuevo estimador tiene además, menor complejidad computacional con respecto al original.

Damos una revisión de los conocidos estimadores finales de máxima verosimilitud truncada para datos censurados y errores eventualmente asimétricos desarrollados por Locatelly, Marazzi y Yohai (2010), los cuales mantienen un alto punto de ruptura pero con la misma eficiencia asintótica que el de máxima verosimilitud. Los mismos son una extensión natural del estimador de máxima verosimilitud truncada para observaciones no censuradas propuesta por Marazzi y Yohai (2004).

La Sección 9 está destinada a analizar el comportamiento del nuevo S-estimador a través de un estudio de simulación de Monte Carlo. Será comparado con el S-estimador de Locatelli, Marazzi y Yohai (2010) y con el de máxima verosimilitud. También contrastaremos las versiones finales de ambos S-estimadores por máxima verosimilitud truncada. Todos estos estimadores serán comparados luego bajo contaminación.

Finalmente analizamos un ejemplo con datos reales. Estudiamos en un primer momento la distribución de los datos, utilizando para ello el estimador de Kaplan y Meier (1958). Luego aplicamos las herramientas vistas en esta tesis para obtener y comparar las estimaciones vistas.

En los respectivos apéndices se podrán encontrar las demostraciones de los resultados enunciados en las secciones previas, como también desarrollos teóricos necesarios para las mismas.

2. Datos censurados

El análisis de sobrevivencia consiste en un conjunto de técnicas para analizar el tiempo de seguimiento ("tiempo de vida") hasta la ocurrencia del evento de interés ("falla o muerte").

Al hablar de "supervivencia" o "tiempo de vida" no nos referimos necesariamente a la vida biológica, sino que en sentido amplio debe interpretarse como el tiempo necesario hasta que se produce la "falla". Estos términos son usados muchas veces en sentido figurado, pero no presentan dificultades para entenderlos en cada contexto.

La definición de "tiempo de vida" incluye una escala de tiempo, un tiempo inicial o de origen y la especificación del evento: muerte, falla, etc. En algunos situaciones es difícil determinar cuándo el evento ocurre, como en el ejemplo de la aparición de un tumor. Por otro lado, la escala de tiempo no siempre es cronológica. Por ejemplo, para los vehículos a motor la escala de tiempo debería ser los kilómetros recorridos y para una impresora o fotocopiadora podríamos considerar como escala de tiempo la cantidad de páginas impresas.

Existen varios mecanismos que no hacen posible la observación completa de los tiempos de seguimiento, como es el caso de la censura (que a su vez puede haber de varios tipos: derecha, izquierda, por intervalos).

La censura por la derecha se presenta cuando hasta el último control que se le hace al individuo, aún no ha ocurrido el evento que se desea observar. Existen varias razones para que se presente este tipo de censura:

- Que hasta el momento de la finalización del estudio no haya ocurrido el evento. Esto se da en los casos donde el período de seguimiento es finito.

- Que el individuo haya abandonado el estudio.

- Que haya ocurrido en el individuo otro evento que imposibilite la ocurrencia del evento que se desea observar.

La censura de los datos podría estar planificada de antemano cuando, por ejemplo, se toma la decisión de terminar el experimento en un determinado momento y esto podría ocurrir antes de que todos los artículos hayan fallado. O por el contrario, no estar planificada como en el caso de un paciente bajo tratamiento que no puede seguir siendo observado en virtud de un translado, mudanza, fuga o fallecimiento. El primer caso se trata de una censura fija y el segundo de una censura aleatoria. Por ejemplo, si el evento de interés que se quiere registrar es la muerte de un paciente, la censura podría estar dada por un translado, mudanza, fuga entre otros. Por el contrario, si el evento de interés es la reducción del tumor luego de un tratamiento específico, la censura podría ocurrir con la muerte o el translado del paciente.

En esta tesis sólo trabajaremos con datos censurados a derecha y con censura aleatoria. Sin embargo, también podría ocurrir la censura a izquierda que de hecho es menos común. Se presenta cuando el evento que se desea observar ha ocurrido antes de que el dispositivo o individuo se incluya en el estudio. Se puede encontrar un ejemplo en el Apéndice respectivo.

Para obtener un panorama general de los distintos tipos de censura se puede ver el libro de Andersen et al. (1993), el de Klein y Moeschberger (1997) o el de Lawless (2003) entre otros. Algunos ejemplos fueron resumidos en el Apéndice de esta sección.

Introducimos algunas notaciones para referirnos a los conjuntos de datos censurados. Tanto en el caso de variables aleatorias, como para sus valores observados (realizaciones de dichas variables una vez observada la muestra) usaremos letras minúsculas. Esto viola la convención, pero no debería generar confusión.

A continuación especificamos el modelo aleatorio de censura a la derecha. Supongamos que n individuos tienen "tiempos de vida" representados por variables aleatorias

$y_1, \ldots, y_n$. Cada individuo tiene una censura (a derecha) que notaremos $v_1, \ldots, v_n$. Si $y_i \leq v_i$ entonces y_i será observado, en caso contrario será observado v_i y sólo sabremos que $y_i > v_i$. Lo que siempre observaremos son las variables que llamaremos censuradas $y_1^*, \ldots, y_n^*$ dadas por

$$y_i^* = \text{mín}\{y_i, v_i\}.$$

Definimos también la variable indicadora $\delta_i = I(y_i \leq v_i)$, es decir $\delta_i = 1$ si no hay censura, en cuyo caso $y_i^* = y_i$ y $\delta_i = 0$ si hay censura es decir $y_i > y_i^*$.

3. Modelo y notaciones

Sea $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)$ una muestra aleatoria, donde $\mathbf{x}_i$ es un vector de p variables explicatorias e y_i es una variable respuesta real. Consideramos el modelo de regresión lineal

$$y_i = \mathbf{x}_i^t \boldsymbol{\beta}_0 + \sigma_0 u_i, \ i = 1, \ldots, n. \tag{1}$$

En caso de haber ordenada al origen entenderemos $(\mathbf{x}_i, y_i) = (1, \overline{\mathbf{x}}_i, y_i)$ donde $\overline{\mathbf{x}}_i$ es un vector de $p-1$ variables explicatorias, cuando $p \geq 2$. Luego, la muestra aleatoria será $(\overline{\mathbf{x}}_1, y_1), \ldots, (\overline{\mathbf{x}}_n, y_n)$. Si denotamos $\boldsymbol{\beta}_0 = (\alpha_0, \boldsymbol{\alpha}_1)$ con $\alpha_0 \in \mathbb{R}$ y $\boldsymbol{\alpha}_1 \in \mathbb{R}^{p-1}$ entonces

$$y_i = \alpha_0 + \overline{\mathbf{x}}_i^t \boldsymbol{\alpha}_1 + \sigma_0 u_i, \ i = 1, \ldots, n.$$

En el resto de la tesis evitaremos esta notación a menos que sea necesario. Aunque no debería generar confusión, de haber ordenada la origen, haremos abuso de notación y seguiremos escribiendo $(\mathbf{x}_i, y_i)$ en lugar de $(\overline{\mathbf{x}}_i, y_i)$. Es decir, según el contexto entenderemos que $(\mathbf{x}_i, y_i)$ se refiere a $(\overline{\mathbf{x}}_i, y_i)$.

Es habitual en situaciones como las planteadas en los ejemplos de las secciones anteriores y sus respectivos apéndices, que los "tiempos de vida" y_i, se representen en escala logarítmica.

La distribución de las covariables $\mathbf{x}_i$ podría ser desconocida, pero no depende de los parámetros del modelo. Los errores u_i son variables aleatorias continuas, independientes e idénticamente distribuídas (notaremos i.i.d.) con función de distribución acumulada F_0 e independientes de $\mathbf{x}_i$. Los parámetros $\boldsymbol{\beta}_0 \in \mathbb{R}^p$ y $\sigma_0 \in \mathbb{R}^+$ son desconocidos. El primero es un vector cuya primera coordenada podría corresponder a la ordenada al origen y el segundo es un parámetro de escala.

Consideraremos censura a la derecha donde $v_1, \ldots, v_n$ son variables aleatorias i.i.d. y cada u_i independiente del vector $(\mathbf{x_i}, v_i)$. Supondremos que $(\mathbf{x}_1, y_1, v_1), \ldots, (\mathbf{x}_n, y_n, v_n)$ es una muestra aleatorias. Los vectores observables no son $(\mathbf{x}_i, y_i)$ sino $(\mathbf{x}_i, y_i^*, \delta_i)$.

Notaremos $F_0(z) = F_{0,1}(z)$ donde $F_{0,1}$ es el miembro estándar de una familia paramétrica de posición y escala $F_{\mu,\sigma}$ (con distribución simétrica o asimétrica), es decir

$$F_{\mu,\sigma}(z) = F_{0,1}\left(\frac{z-\mu}{\sigma}\right).$$

Notamos con f_0 y $f_{\mu,\sigma}$ las correspondientes densidades, luego

$$f_{\mu,\sigma}(z) = \frac{1}{\sigma} f_0\left(\frac{z-\mu}{\sigma}\right).$$

Además notamos $H_{\boldsymbol{\beta},\sigma}$ a la función de distribución acumulada de $(\mathbf{x}_i, y_i)$ cuando $\boldsymbol{\beta}$ y σ son los parámetros verdaderos del modelo. Es decir, esta notación asume verdadera la relación $y = \mathbf{x}^t\boldsymbol{\beta} + \sigma u$. Bajo el modelo (1) corresponde usar $H_{\boldsymbol{\beta}_0,\sigma_0}$.

Algunos ejemplos de modelos de posición y escala para el error son los que se detallan a continuación:

i) el modelo gaussiano con función de densidad

$$f_{\mu,\sigma}(z) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(z-\mu)^2}{2\sigma^2}} \quad z \in \mathbb{R},$$

ii) el modelo log-Weibull (asimétrico) con densidad

$$f_{\mu,\sigma}(z) = \frac{1}{\sigma}\, e^{\frac{z-\mu}{\sigma}} - e^{\frac{z-\mu}{\sigma}} \quad z \in \mathbb{R},$$

iii) y el modelo logístico (también asimétrico) con densidad

$$f_{\mu,\sigma}(z) = \frac{e^{\frac{z-\mu}{\sigma}}}{\sigma\left(1 + e^{\frac{z-\mu}{\sigma}}\right)^2} \quad z \in \mathbb{R}.$$

Otros ejemplos son mencionados por Lawless (2003) en los Capítulos 1, 3 y 6.

En las figuras 1 y 2 se muestran los gráficos correspondientes a las densidades de las distribuciones log-Weibull y logística respectivamente, con $\mu = 0$ y diferentes valores de σ.

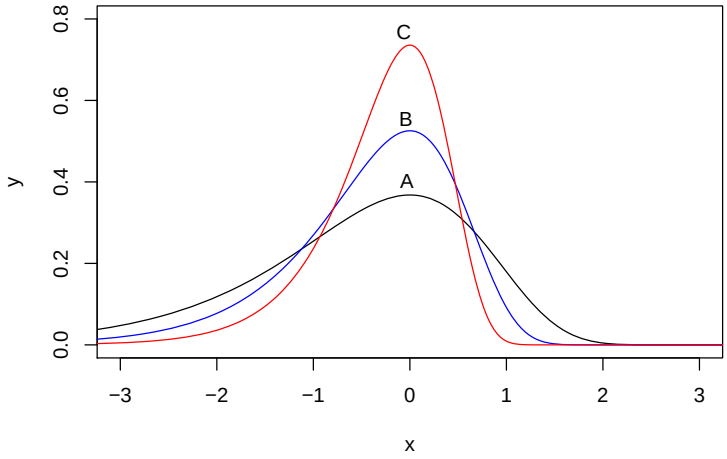

Figura 1: A, B y C corresponden a densidades de una distribución log-Weibull con $\sigma = 1$, $\sigma = 0{,}7$ y $\sigma = 0{,}5$ respectivamente y $\mu = 0$.

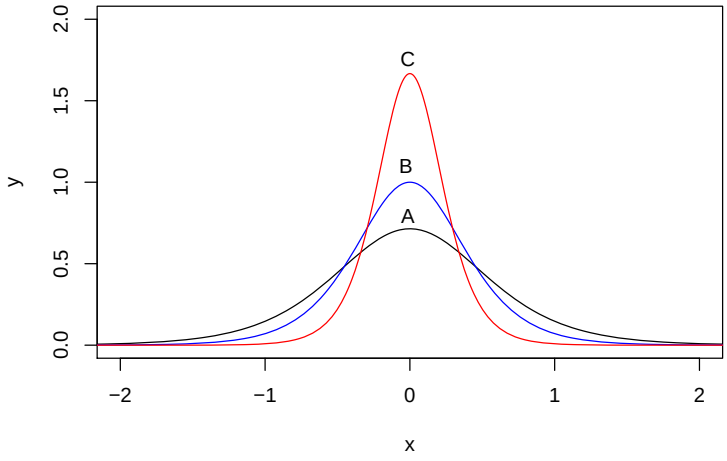

Figura 2: A, B y C corresponden a densidades de una distribución logística con $\sigma = 0{,}35$, $\sigma = 0{,}25$ y $\sigma = 0{,}15$ respectivamente y $\mu = 0$.

Condición de identificabilidad

La siguiente es una condición que se requiere para que los parámetros del modelo estén identificados.

CI *Para todo* $\mathbf{d} \neq \mathbf{0}$ *se tiene* que $P_{\boldsymbol{\beta}_0,\sigma_0}(\mathbf{x}^t\mathbf{d} = 0) < 1$.

La notación $P_{\boldsymbol{\beta}_0,\sigma_0}$ significa que estamos suponiendo verdadero el modelo (1). Esta condición dice que el vector $\mathbf{x}$ no está concentrado en ningún hiperplano y en particular se cumple, si la distribución de $\mathbf{x}$ tiene densidad. Si el modelo sólo tiene parámetros de posición y escala, la condición es superflua. En caso contrario y si la condición **CI** no se satisface, entonces existe un $\boldsymbol{\beta} \neq \boldsymbol{\beta}_0$ tal que $P_{\beta_0,\sigma_0}(\mathbf{x}^t\boldsymbol{\beta} = \mathbf{x}^t\boldsymbol{\beta}_0) = 1$. De esta forma, el parámetro $\boldsymbol{\beta}_0$ no estará identificado. Con la condición **CI**, $\boldsymbol{\beta}_0$ queda identificado y por lo tanto también los está σ_0.

Esperanza condicional y distribución empírica para datos censurados

Notaremos $E_0(g(u))$ a la esperanza de la función g basada en la distribución F_0. Análogamente, $E_{\boldsymbol{\beta},\sigma}(h(\mathbf{x}, y))$ será la esperanza de h basada en la distribución $H_{\boldsymbol{\beta},\sigma}$.

Notar que en general valen:

$$E(h(\mathbf{x}, y)|\mathbf{x}, y^*, \delta = 1) = h(\mathbf{x}, y^*) \tag{2}$$

$$E(h(\mathbf{x}, y)|\mathbf{x}, y^*, \delta = 0) = E(h(\mathbf{x}, y)|\mathbf{x}, y^*, y > y^*) \tag{3}$$

y por lo tanto, podemos escribir

$$E(h(\mathbf{x}, y)|\mathbf{x}, y^*, \delta) = \delta h(\mathbf{x}, y^*) + (1 - \delta)E(h(\mathbf{x}, y)|\mathbf{x}, y^*, y > y^*). \tag{4}$$

Usualmente trabajaremos con esta expresión para la esperanza condicional, aunque cuando $\delta = 1,\ h(\mathbf{x}, y^*) = h(\mathbf{x}, y)$ y

$$E(h(\mathbf{x}, y)|\mathbf{x}, y^*, \delta) = \delta h(\mathbf{x}, y) + (1 - \delta)E(h(\mathbf{x}, y)|\mathbf{x}, y^*, y > y^*).$$

A partir de (4) se puede obtener el siguiente lema, cuya prueba se encuentra en el Apéndice de esta sección.

Lema 3.1. *Sea* $h(\mathbf{x}, y)$ *una función medible. Bajo los supuestos del modelo (1) cuando* $\boldsymbol{\beta}$ *y* σ *son los parámetros verdaderos, se tiene*

$$E_{\boldsymbol{\beta},\sigma}(h(\mathbf{x}, y)|\mathbf{x}, y^*, y > y^*) = \frac{\displaystyle\int_{\frac{y^* - \mathbf{x}^t\boldsymbol{\beta}}{\sigma}}^{+\infty} h(\mathbf{x}, \mathbf{x}^t\boldsymbol{\beta} + \sigma u) f_0(u) du}{1 - F_0\left(\frac{y^* - \mathbf{x}^t\boldsymbol{\beta}}{\sigma}\right)}. \tag{5}$$

En la demostración del Lema 3.1 y haciendo uso de la independencia de u con el par $(\mathbf{x}, v)$ se dedujo que

$$E_{\boldsymbol{\beta},\sigma}(h(\mathbf{x}, y)|\mathbf{x}, y^*, y > y^*) = E\left(h(\mathbf{x}, \mathbf{x}^t\boldsymbol{\beta} + \sigma u)\,\middle|\, u > \frac{y^* - \mathbf{x}^t\boldsymbol{\beta}}{\sigma}\right). \tag{6}$$

Veamos cómo queda la esperanza condicional para un caso particular de h. Consideremos una función $\psi : \mathbb{R} \to \mathbb{R}$ y definamos para cada $\boldsymbol{\beta}$ y σ la función

$$h(\mathbf{x}, y) = \psi\left(\frac{y - \mathbf{x}^t\boldsymbol{\beta}}{\sigma}\right).$$

Luego de (6) y de la demostración de (5) se obtienen respectivamente (7) y (8)

$$E_{\boldsymbol{\beta},\sigma}\left(\psi\left(\frac{y - \mathbf{x}^t\boldsymbol{\beta}}{\sigma}\right)\middle| \mathbf{x}, y^*, y > y^*\right) = E\left(\psi(u)\middle| u > \frac{y^* - \mathbf{x}^t\boldsymbol{\beta}}{\sigma}\right) \tag{7}$$

$$= \frac{\displaystyle\int_{\frac{y^* - \mathbf{x}^t\boldsymbol{\beta}}{\sigma}}^{+\infty} \psi(u) f_0(u) du}{1 - F_0\left(\frac{y^* - \mathbf{x}^t\boldsymbol{\beta}}{\sigma}\right)}. \tag{8}$$

Observemos que la esperanza condicional para este caso particular, depende solamente del cociente $\frac{y^* - \mathbf{x}^t\boldsymbol{\beta}}{\sigma}$.

Sea $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)$ una muestra aleatoria verificando el modelo cuando $\boldsymbol{\beta}$ y σ son los parámetros verdaderos. Consideremos $(\mathbf{x}_1, y_1^*, \delta_1), \ldots, (\mathbf{x}_n, y_n^*, \delta_n)$ la correspondiente muestra aleatoria de datos censurados. Se define la **función de distribución empírica para datos censurados** como

$$H_{n,\boldsymbol{\beta},\sigma}(\mathbf{z}, z) =$$

$$= \frac{1}{n}\sum_{i=1}^{n} I(\mathbf{x}_i \leq \mathbf{z}) E_{\boldsymbol{\beta},\sigma}(I(y_i \leq z)|\mathbf{x}_i, y_i^*, \delta_i) \tag{9}$$

$$= \frac{1}{n}\sum_{i=1}^{n} (\delta_i I(\mathbf{x}_i \leq \mathbf{z}) I(y_i^* \leq z) + (1 - \delta_i) I(\mathbf{x}_i \leq \mathbf{z}) E_{\boldsymbol{\beta},\sigma}(I(y_i \leq z)|\mathbf{x}_i, y_i^*, y_i > y_i^*)). \tag{10}$$

La función de distribución empírica para datos censurados es efectivamente una distribución. En particular, a partir de (5), los términos correspondientes a las observaciones censuradas verifican

$$E_{\boldsymbol{\beta},\sigma}(I(y_i \leq z)|\mathbf{x}_i, y_i^*, y_i > y_i^*) = \frac{\displaystyle\int_{\frac{y_i^* - x_i^t\boldsymbol{\beta}}{\sigma}}^{+\infty} I(\mathbf{x}_i^t\boldsymbol{\beta} + \sigma u_i \leq z) f_0(u_i) du_i}{1 - F_0\left(\frac{y_i^* - \mathbf{x}_i^t\boldsymbol{\beta}}{\sigma}\right)}$$

$$= \frac{\displaystyle\int_{\frac{y_i^* - x_i^t\boldsymbol{\beta}}{\sigma}}^{\frac{z - \mathbf{x}_i^t\boldsymbol{\beta}}{\sigma}} f_0(u_i) du_i}{1 - F_0\left(\frac{y_i^* - \mathbf{x}_i^t\boldsymbol{\beta}}{\sigma}\right)}$$

$$= \frac{F_0\left(\frac{z - \mathbf{x}_i^t\boldsymbol{\beta}}{\sigma}\right) - F_0\left(\frac{y_i^* - \mathbf{x}_i^t\boldsymbol{\beta}}{\sigma}\right)}{1 - F_0\left(\frac{y_i^* - \mathbf{x}_i^t\boldsymbol{\beta}}{\sigma}\right)}.$$

Observación 3.1. Cuando no hay censura (es decir $\delta_i = 1$ o $y_i^* = y_i$ para todo $i = 1, \ldots, n$), $H_{n,\boldsymbol{\beta},\sigma}(\mathbf{z}, z)$ coincide con la función de distribución empírica usual

$$H_n(\mathbf{z}, z) = \frac{1}{n} \sum_{i=1}^{n} I(\mathbf{x}_i \leq \mathbf{z}) I(y_i \leq z).$$

Notar también que los términos de $H_{n,\boldsymbol{\beta},\sigma}(\mathbf{z}, z)$ son variables aleatorias cuya esperanza coincide con la función de distribución acumulada de $(\mathbf{x}_i, y_i)$, es decir

$$\begin{aligned} E_{\boldsymbol{\beta},\sigma}(I(\mathbf{x}_i \leq \mathbf{z}) E_{\boldsymbol{\beta},\sigma}(I(y_i \leq z)|\mathbf{x}_i, y_i^*, \delta_i)) &= E_{\boldsymbol{\beta},\sigma}(E_{\boldsymbol{\beta},\sigma}(I(\mathbf{x}_i \leq \mathbf{z}) I(y_i \leq z)|\mathbf{x}_i, y_i^*, \delta_i)) \\ &= E_{\boldsymbol{\beta},\sigma}(I(\mathbf{x}_i \leq \mathbf{z}) I(y_i \leq z)) \\ &= E_{\boldsymbol{\beta},\sigma}(I(\mathbf{x}_i \leq \mathbf{z}, y_i \leq z)) \\ &= P_{\boldsymbol{\beta},\sigma}(\mathbf{x}_i \leq \mathbf{z}, y_i \leq z) \\ &= H_{\boldsymbol{\beta},\sigma}(\mathbf{z}, z). \end{aligned}$$

Observación 3.2. Sea $(\mathbf{x}_1, y_1^*, \delta_1), \ldots, (\mathbf{x}_n, y_n^*, \delta_n)$ es una muestra aleatoria de datos censurados. Es fácil verificar a partir de la Ley de los Grandes Números que $H_{n,\boldsymbol{\beta},\sigma}(\mathbf{z}, z)$ resulta un estimador consistente para $H_{\boldsymbol{\beta},\sigma}(\mathbf{z}, z)$, es decir casi seguramente

$$H_{n,\boldsymbol{\beta},\sigma}(\mathbf{z}, z) \to H_{\boldsymbol{\beta},\sigma}(\mathbf{z}, z).$$

Notaremos $E_{n,\boldsymbol{\beta},\sigma}(h(\mathbf{x}, y))$ a la esperanza de la función medible h bajo $H_{n,\boldsymbol{\beta},\sigma}$ cuando $\boldsymbol{\beta}$ y σ son los parámetros verdaderos del modelo. Luego, a partir de la Observación 3.2, se obtiene

$$E_{n,\boldsymbol{\beta},\sigma}(h(\mathbf{x}, y)) \to E_{\boldsymbol{\beta},\sigma}(h(\mathbf{x}, y))$$

casi seguramente.

La demostración del siguiente lema se encuentra en el Apéndice de esta sección.

Lema 3.2. *Sea $h(\mathbf{x}, y)$ una función medible. Luego*

$$E_{n,\boldsymbol{\beta},\sigma}(h(\mathbf{x}, y)) = \frac{1}{n} \sum_{i=1}^{n} E_{\boldsymbol{\beta},\sigma}(h(\mathbf{x}_i, y_i)|\mathbf{x}_i, y_i^*, \delta_i). \tag{11}$$

La igualdad (11) se puede reescribir a partir de (4) y (5) de la siguiente manera:

$$E_{n,\boldsymbol{\beta},\sigma}(h(\mathbf{x}, y)) = \frac{1}{n} \sum_{i=1}^{n} \left(\delta_i h(\mathbf{x}_i, y_i^*) + (1 - \delta_i) \frac{\displaystyle\int_{\frac{y_i^* - x_i^t \boldsymbol{\beta}}{\sigma}}^{+\infty} h(\mathbf{x}_i, \mathbf{x}_i^t \boldsymbol{\beta} + \sigma u_i) f_0(u_i) du_i}{1 - F_0\left(\frac{y_i^* - \mathbf{x}_i^t \boldsymbol{\beta}}{\sigma}\right)} \right). \tag{12}$$

Si en particular

$$h(\mathbf{x}, y) = \psi\left(\frac{y - \mathbf{x}^t \boldsymbol{\beta}}{\sigma}\right),$$

entonces por (7), (8) y (12) se obtiene

$$E_{n,\boldsymbol{\beta},\sigma}\left(\psi\left(\frac{y-\mathbf{x}^t\boldsymbol{\beta}}{\sigma}\right)\right) =$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left(\delta_i\psi\left(\frac{y_i^*-\mathbf{x}_i^t\boldsymbol{\beta}}{\sigma}\right)+(1-\delta_i)E_{\beta,\sigma}\left(\psi(u_i)\,\middle|\,u_i>\frac{y_i^*-\mathbf{x}_i^t\boldsymbol{\beta}}{\sigma}\right)\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left(\delta_i\psi\left(\frac{y_i^*-\mathbf{x}_i^t\boldsymbol{\beta}}{\sigma}\right)+(1-\delta_i)\frac{\int_{\frac{y_i^*-\mathbf{x}_i^t\boldsymbol{\beta}}{\sigma}}^{+\infty}\psi(u_i)f_0(u_i)du_i}{1-F_0\left(\frac{y_i^*-\mathbf{x}_i^t\boldsymbol{\beta}}{\sigma}\right)}\right). \tag{13}$$

4. Estimadores de máxima verosimilitud para datos censurados

Consideremos una muestra aleatoria $(\mathbf{x}_1,y_1),\ldots,(\mathbf{x}_n,y_n)$ verificando el modelo (1) y sea $(\mathbf{x}_1,y_1^*,\delta_1),\ldots,(\mathbf{x}_n,y_n^*,\delta_n)$ la correspondiente muestra aleatoria observable de datos censurados. Para hallar los estimadores de máxima verosimilitud de $\boldsymbol{\beta}_0$ y σ_0 necesitamos una expresión para la función de verosimilitud. La distribución del vector $(\mathbf{x}_i,y_i^*,\delta_i)$ se puede descomponer en la distribución de dicho vector, condicional $\mathbf{x}_i$, por una función que sólo depende de la distribución de $\mathbf{x}_i$. Este último factor se puede despreciar porque si bien la distribución de $\mathbf{x}_i$ es desconocida, no depende de los parámetros $\boldsymbol{\beta}_0$ y σ_0.

Lawless (2003) deduce la función de verosimilitud $L(\boldsymbol{\beta}_0,\sigma_0)$ para la muestra aleatoria observada:

$$L(\boldsymbol{\beta}_0,\sigma_0)=\prod_{i=1}^{n}\left(\frac{1}{\sigma_0}f_0\left(\frac{y_i^*-\mathbf{x}_i^t\boldsymbol{\beta}_0}{\sigma_0}\right)\right)^{\delta_i}\left(1-F_0\left(\frac{y_i^*-\mathbf{x}_i^t\boldsymbol{\beta}_0}{\sigma_0}\right)\right)^{1-\delta_i}.$$

Si notamos

$$r=\sum_{i=1}^{n}\delta_i \tag{14}$$

$$l(\boldsymbol{\beta}_0,\sigma)=\ln L(\boldsymbol{\beta}_0,\sigma_0) \tag{15}$$

$$\rho_0=-\ln f_0, \tag{16}$$

obtenemos la siguiente expresión para el logaritmo de la verosimilitud:

$$l(\boldsymbol{\beta}_0,\sigma_0)=-r\ln(\sigma_0)-\sum_{i=1}^{n}\left(\delta_i\rho_0\left(\frac{y_i^*-\mathbf{x}_i^t\boldsymbol{\beta}_0}{\sigma_0}\right)+(1-\delta_i)\ln\left(1-F_0\left(\frac{y_i^*-\mathbf{x}_i^t\boldsymbol{\beta}_0}{\sigma_0}\right)\right)\right).$$

Lawless (2003) obtuvo una expresión para las ecuaciones de máxima verosimilitud, que en términos de las notaciones usadas por Locatelli, Marazzi, Yohai (2012), son las que se presentan en el siguiente lema. Para ello necesitamos definir las siguientes funciones:

$$\psi_0(t)=\rho_0'(t)=-\frac{f_0'(t)}{f_0(t)}$$

$$\psi_1(t)=t\psi_0(t).$$

La deducción de las ecuaciones de máxima verosimilitud se encuentra en el Apéndice de esta sección.

Lema 4.1. *Los estimadores de máxima verosimilitud para datos censurados verifican, en caso de existir, el siguiente sistema de ecuaciones:*

$$E_{n,\boldsymbol{\beta},\sigma}\left(\psi_0\left(\frac{y-\mathbf{x}^t\boldsymbol{\beta}}{\sigma}\right)\mathbf{x}\right)=0 \tag{17}$$

$$E_{n,\boldsymbol{\beta},\sigma}\left(\psi_1\left(\frac{y-\mathbf{x}^t\boldsymbol{\beta}}{\sigma}\right)\right)=1 \tag{18}$$

5. S-estimadores de regresión lineal con datos no censurados

El S-estimador es un tipo de estimador que puede ser calibrado para que tenga un alto punto de ruptura. Sin embargo, cuando esto ocurre suele ser bastante ineficiente. A menudo los S-estimadores se utilizan como estimadores iniciales en procedimientos computacionales para el cálculo de otro tipo de estimador más eficiente. Los S-estimadores de regresión para datos no censurados fueron introducidos por Rousseuw y Yohai (1984).

Sea $\rho : \mathbb{R} \to \mathbb{R}^+$ una función. Consideremos la siguiente lista de condiciones que podría satisfacer dicha función ρ.

C1 $\rho(0) = 0$,

C2 $\lim\limits_{z\to\pm\infty} \rho(z) = l$,

C3 ρ es par, es decir $\rho(z) = \rho(-z)$,

C4 ρ es continua,

C5 $0 \leq z_1 < z_2$ implica $\rho(z_1) \leq \rho(z_2)$,

C6 Existe $k \in \mathbb{R}$ tal que

- **C6-A** $\rho(z) = l$ si $|z| \geq k$,
- **C6-B** $\ln(l - \rho)$ es cóncava en $(-k, k)$,

C7 $0 \leq z_1 < z_2$ y $\rho(z_1) < l$ implica $\rho(z_1) < \rho(z_2)$.

Notar que las condiciones **C3** y **C5** implican que ρ es monótona no decreciente en módulo, es decir: $|z_1| < |z_2|$ entonces $\rho(z_1) \leq \rho(z_2)$. Análogamente las condiciones **C3** y **C7** implican que ρ es estrictamente creciente en módulo en un entorno del cero, es decir: $|z_1| < |z_2|$ y $\rho(z_1) < l$ entonces $\rho(z_1) < \rho(z_2)$.

Consideremos los valores de z para los cuáles ρ es derivable y llamamos $\psi(z) = \rho'(z)$. Si ρ es una función par (es decir vale **C3**) entonces ψ es impar. Si ρ verifica **C5** entonces $\psi(z) \geq 0$ para todo $z \geq 0$. Si ρ verifica **C2** entonces ψ es redescendiente, es decir $\lim_{z\to\pm\infty} \psi(z) = 0$.

Sin pérdida de generalidad supondremos $l = 1$.

Un ejemplo de función ρ satisfaciendo todas las condiciones anteriores, es un miembro de la familia bicuadrada de Tukey:

$$\rho_k(z) = \begin{cases} 3\left(\frac{z}{k}\right)^2 - 3\left(\frac{z}{k}\right)^4 + \left(\frac{z}{k}\right)^6 & \text{si} \quad |z| \leq k \\ 1 & \text{si} \quad |z| > k \end{cases}$$

donde k es un parámetro de ajuste. De esta forma

$$\psi_k(z) = \begin{cases} \frac{6}{k}\left(\frac{z}{k} - 2\left(\frac{z}{k}\right)^3 + \left(\frac{z}{k}\right)^5\right) & \text{si} \quad |z| \leq k \\ 0 & \text{si} \quad |z| > k \end{cases}.$$

Los siguientes gráficos corresponden a ρ_4 y a la correspondiente función ψ_4.

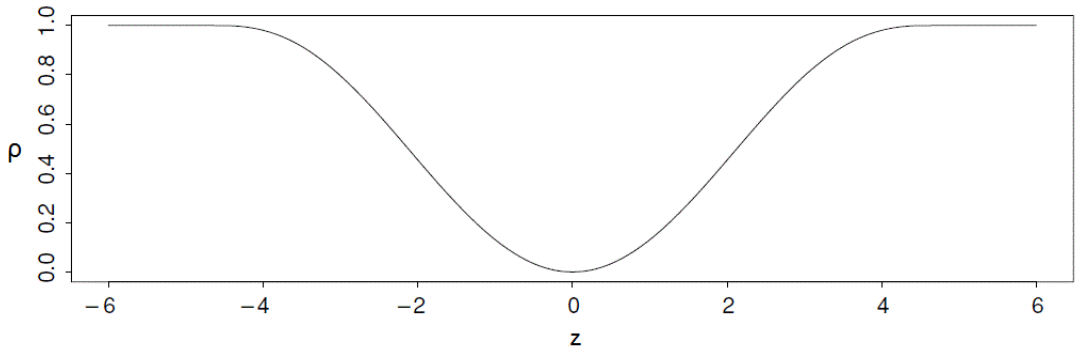

Figura 3: Gráfico de la función ρ_4

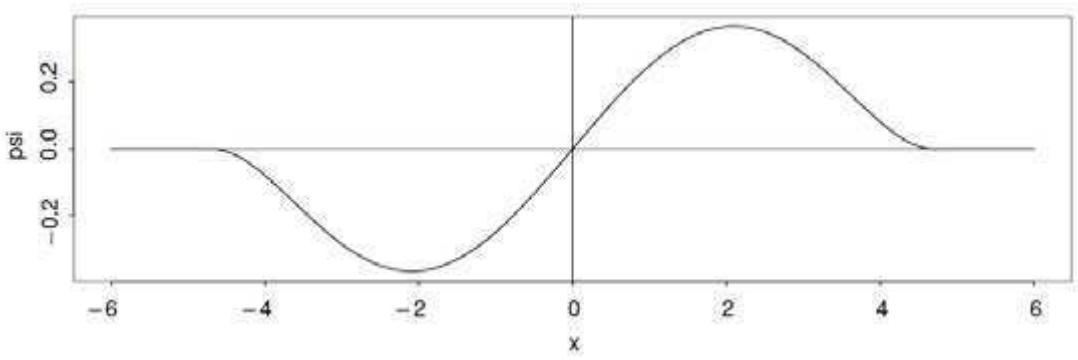

Figura 4: Gráfico de la función ψ_4

Huber (1981) llamó **M escala** a cada solución positiva s (en caso de existir) de la ecuación

$$E_0\left(\rho\left(\frac{u-\mu}{s}\right)\right) = b, \tag{19}$$

donde la esperanza está basada en F_0 y $b \in (0,1)$. Suele usarse $b = 0.5$ para que el estimador buscado alcance un punto de ruptura igual a 0.5.

En general, se puede definir la M escala de tal forma que siempre exista y sea única como

$$S(\mu) = \text{ínf}\left\{s \in \mathbb{R}^+ : E_0\left(\rho\left(\frac{u-\mu}{s}\right)\right) \leq b\right\}.$$

Bajo ciertas hipótesis (por ejemplo **C1** y **C2**) vale que

$$S(\mu) > 0 \text{ si y sólo si } P(u = \mu) < 1 - b.$$

La prueba de este resultado es muy similar a la que se hará en el siguiente Lema. Si ρ es continua (**C4**) y u tiene distribución continua (como estamos suponiendo) entonces podemos usar directamente la igualdad (19).

Veremos a continuación bajo qué condiciones existe y es única la M escala pensada como solución de la ecuación (19).

Lema 5.1. *Supongamos que ρ satisface las condiciones **C1**, **C2** y **C4**. Luego para cada $\mu \in \mathbb{R}$ y $b \in (0,1)$ fijos, la ecuación (19) tiene solución. Más aún, si la solución no es única y ρ verifica además **C3** y **C5**, el conjunto de soluciones de dicha ecuación es un intervalo cerrado.*

Para que no haya ambigüedades y dar una buena definición de M escala, se puede considerar para cada $\mu \in \mathbb{R}$, $S(\mu)$ como el mínimo valor que verifica la ecuación (19), que existe si suponemos las condiciones **C1**, **C2** y **C4**. Es decir

$$S(\mu) = \text{mín}\left\{s \in \mathbb{R}^+ : E_0\left(\rho\left(\frac{u-\mu}{s}\right)\right) = b\right\}.$$

Para cada $\mu \in \mathbb{R}$ tenemos $h_\mu : \mathbb{R}^+ \to \mathbb{R}$ definida como

$$h_\mu(s) = E_0\left(\rho\left(\frac{u-\mu}{s}\right)\right).$$

En la demostración del Lema 5.1 utilizamos que $h_\mu(s)$ es monótona no creciente, ya que $\rho(|z|)$ es monótona no decreciente, debido a las propiedades **C3** y **C5** de ρ. Si agregamos **C7**, $h_\mu(s)$ resulta estrictamente creciente que es lo que probamos en la demostración del próximo lema.

Lema 5.2. *Supongamos que ρ satisface las condiciones **C1** a **C5** y **C7**. Luego para cada $\mu \in \mathbb{R}$ y $b \in (0,1)$ fijos, la ecuación (19) tiene solución única $S(\mu)$.*

De hecho si ρ verifica las hipótesis del Lema 5.2, podemos considerar la función $S : \mathbb{R} \to \mathbb{R}$, definida en cada μ como la única solución de la ecuación (19).

Lema 5.3. *Supongamos que ρ verifica **C1** a **C5**. Si existen μ_0 y s_0 tales que*

$$\mu_0 = \arg\min_{\mu} S(\mu) \tag{20}$$

$$s_0 = \min_{\mu} S(\mu) \tag{21}$$

entonces

$$\mu_0 = \arg\min_{\mu} E_0\left(\rho\left(\frac{u-\mu}{s_0}\right)\right) \tag{22}$$

$$s_0 = S(\mu_0). \tag{23}$$

*La recíproca es cierta si ρ verifica además **C7**. En cualquier caso*

$$E_0\left(\rho\left(\frac{u-\mu_0}{s_0}\right)\right) = b. \tag{24}$$

Definición 5.1. Diremos que una función $f : \mathbb{R} \to \mathbb{R}$ es estrictamente unimodal si existe $a \in \mathbb{R}$ tal que f es monótona no decreciente para $t < a$, monótona no creciente para $t > a$ y f tiene un único máximo en $t = a$.

La primera parte del siguiente resultado se debe a Fasano, Maronna, Sued y Yohai (2012) y es una pequeña variante de la dada por Mizera (1993). La segunda parte fue probada por Yohai (1985).

Definimos para cada $s \in \mathbb{R}^+$, la función $q_s : \mathbb{R} \to \mathbb{R}$ como

$$q_s(\mu) = E_0\left(\rho\left(\frac{u-\mu}{s}\right)\right). \tag{25}$$

Lema 5.4. *Supongamos que la densidad f_0 de u es estrictamente unimodal. Si ρ verifica **C3** a **C6** entonces q_s alcanza un único mínimo absoluto. Si f_0 es simétrica alrededor de $t = a$, q_s alcanza su mínimo absoluto en $\mu = a$.*

Consideremos el conjunto no vacío y acotado inferiormente

$$\{S(\mu) : \mu \in \mathbb{R}\}.$$

Se llama **S escala** de F_0 al valor

$$s_0 = \inf_{\mu} S(\mu). \tag{26}$$

Lema 5.5. *Si ρ verifica **C1**, **C2** y **C4** entonces $s_0 > 0$. Si además ρ satisface **C3**, **C5** y **C7**, entonces la función $S(\mu)$ es continua y además la S escala de F_0 se realiza, es decir*

$$s_0 = \min_{\mu} S(\mu).$$

Lema 5.6. *Sea s_0 la S escala de F_0. Si ρ satisface **C1** a **C7** y f_0 es estrictamente unimodal entonces existe un único*

$$\mu_0 = \arg\min_{\mu} E_0\left(\rho\left(\frac{u-\mu}{s_0}\right)\right).$$

Además, dicho μ_0 es el único que verifica $S(\mu_0) = s_0$.

Bajo las hipótesis del Lema 5.6 y haciendo uso del Lema 5.3 se obtiene que existe un único μ_0 como en (20).

Tanto la existencia como la unicidad de μ_0 serán necesarios para obtener estimadores consistentes de los parámetros del modelo. Por comodidad y sin pérdida de generalidad asumiremos $s_0 = 1$. Esto es posible porque se puede considerar otra función que hereda las mismas propiedades que ρ, definida de la siguiente forma:

$$\rho_{s_0}(u) = \rho\left(\frac{u}{s_0}\right).$$

Si ρ verifica **C1** a **C7** y f_0 es estrictamente unimodal entonces por el Lema 5.6,

$$\mu_0 = 0 \Leftrightarrow E_0(\rho(u)) = b.$$

Así como se puede suponer $s_0 = 1$ cambiando la función ρ, se podrá considerar $\mu_0 = 0$ reparametrizando el modelo (1). Luego, se obtendrán S-estimadores consistentes de los parámetros de este nuevo modelo parametrizado. Estos mismos estimadores pueden ser combinados para obtener estimadores consistentes del modelo original. Veamos a continuación cómo reparametrizar el modelo.

Reparametrización del modelo

Sabemos que los errores u tienen función de distribución acumulada F_0 siendo F_0 un miembro estándar de la familia paramétrica de posición y escala $F_{\mu,\sigma}$, con densidades f_0 y $f_{\mu,\sigma}$ respectivamente. Luego, la variable aleatoria

$$z = \mu + \sigma u$$

tiene distribución $F_{\mu,\sigma}$ y densidad

$$f_{\mu,\sigma}(z) = \frac{1}{\sigma} f_0\left(\frac{z-\mu}{\sigma}\right).$$

Reescribimos el modelo como

$$z = \mu + \sigma\mu_0 + \sigma(u - \mu_0)$$

y notamos $\mu^* = \mu + \sigma\mu_0$ y $w = u - \mu_0$ de tal forma que

$$z = \mu^* + \sigma w.$$

Si llamamos G_0 y g_0 a las funciones de distribución acumulada y densidad respectivamente de w entonces

$$G_0(w) = F_0(w + \mu_0) \text{ y } g_0(w) = f_0(w + \mu_0).$$

Sea $G_{\mu^*,\sigma}$ la familia paramétrica de posición y escala para z con parámetros μ^* y σ y $f_{\mu^*,\sigma}$ la correspondiente función de densidad, entonces

$$g_{\mu^*,\sigma}(z) = \frac{1}{\sigma} g_0\left(\frac{z-\mu^*}{\sigma}\right) = \frac{1}{\sigma} f_0\left(\frac{z-\mu^*}{\sigma} + \mu_0\right) = \frac{1}{\sigma} f_0\left(\frac{z-\mu}{\sigma}\right) = f_{\mu,\sigma}(z),$$

como era de esperar. Además vale que

$$E_{G_0}(\rho(w)) = E_0(\rho(u - \mu_0)) = b.$$

Por otro lado, f_0 es estrictamente unimodal si y sólo si g_0 lo es.

Si el modelo (1) tiene ordenada al origen, vamos a considerar $\boldsymbol{\beta}_0 = (\alpha_0, \boldsymbol{\alpha}_1)$ con $\alpha_0 \in \mathbb{R}$ y $\boldsymbol{\alpha}_1 \in \mathbb{R}^{p-1}$ si $p \geq 2$. Luego

$$y_i = \alpha_0 + \mu_0\sigma_0 + \overline{\mathbf{x}}_i^t\boldsymbol{\alpha}_1 + \sigma_0 w_i,\ i = 1, \ldots, n$$

o bien

$$y_i = \alpha_0^* + \boldsymbol{\alpha}_1\overline{\mathbf{x}}_i^t + \sigma_0 w_i,\ i = 1, \ldots, n \tag{27}$$

donde $\alpha_0^* = \alpha_0 + \mu_0\sigma_0$. Si notamos $\boldsymbol{\beta}_0^* = (\alpha_0^*, \boldsymbol{\alpha}_1)$ entonces

$$y_i = \mathbf{x}_i^t\boldsymbol{\beta}_0^* + \sigma_0 w_i,\ i = 1, \ldots, n. \tag{28}$$

Si se obtienen estimadores consistentes de los parámetros del nuevo modelo parametrizado, es decir

$$\tilde{\beta}_{n,0} \to \alpha_0^*,\ \ \tilde{\boldsymbol{\beta}}_{n,1} \to \boldsymbol{\alpha}_1 \quad \text{y} \quad \tilde{\sigma}_n \to \sigma_0,$$

entonces tendremos estimadores consistentes de los parámetros del modelo (1) ya que

$$(\tilde{\beta}_{n,0} - \mu_0\tilde{\sigma}_n, \tilde{\boldsymbol{\beta}}_{n,1}) \to \beta_0.$$

Notar que para ambos modelos (original y parametrizado) los estimadores de la pendiente y la escala son los mismos.

En lo que sigue, recordaremos la definición de S-estimador para datos no censurados.

Definición clásica de S-estimador

Consideremos una muestra aleatoria $(\mathbf{x}_1.y_1), \ldots, (\mathbf{x}_n, y_n)$ y su función de distribución empírica H_n. Dado $b \in (0, 1)$, consideraremos en caso de existir, las soluciones positivas s de la ecuación

$$E_{H_n}\left(\rho\left(\frac{y - \mathbf{x}^t\boldsymbol{\beta}}{s}\right)\right) = b, \tag{29}$$

o de manera equivalente

$$\frac{1}{n}\sum_{i=1}^{n}\rho\left(\frac{y_i - \mathbf{x}_i^t\boldsymbol{\beta}}{s}\right) = b. \tag{30}$$

Para la consistencia fuerte del S-estimador será necesario considerar a b, de tal forma que

$$E_0\left(\rho\left(u\right)\right) = b.$$

Veamos condiciones necesarias y suficientes para que exista solución. Consideremos para cada $\boldsymbol{\beta}$ fijo el conjunto

$$A = \left\{i :\ y_i = \mathbf{x}_i^t\boldsymbol{\beta}\right\}. \tag{31}$$

Lema 5.7. *Supongamos que ρ satisface la condición **C1** y que está acotada superiormente por $l = 1$. Si $\#A^c < nb$ entonces la ecuación (30) no tiene solución.*

Supongamos que ρ verifica **C1** y $\#A^c = nb$. Si $\rho(z) < 1$ para todo z, entonces la ecuación (30) no tiene solución. Sin embargo, si ρ verifica **C6-A** entonces una solución posible es $s = \frac{1}{k} \text{mín} \left\{ |y_i - \mathbf{x}_i^t \boldsymbol{\beta}| : i \in A^c \right\}$.

A continuación veremos condiciones suficientes para que la solución de la ecuación (30) sea única.

Lema 5.8. *Supongamos que $\#A^c > nb$. Si ρ satisface las condiciones **C1**, **C2** y **C4** entonces para cada $\boldsymbol{\beta} \in \mathbb{R}^p$, la ecuación (30) tiene solución. Si además ρ satisface las condiciones **C3**, **C5** y **C7**, entonces la solución es única.*

Observación 5.1. En la demostración del Lema 5.8 se prueba que para cada $\boldsymbol{\beta}$, la función $h_{\boldsymbol{\beta}} : \mathbb{R}^+ \to \mathbb{R}$ definida como

$$h_{\boldsymbol{\beta}}(s) = \frac{1}{n} \sum_{i=1}^{n} \rho \left(\frac{y_i - \mathbf{x}_i^t \boldsymbol{\beta}}{s} \right)$$

es estrictamente decreciente.

La ecuación (30) podría tener múltiples soluciones, entonces podríamos considerar el ínfimo de todas ellas. Sin embargo, bajo ciertas hipótesis el ínfimo se realiza tal como se prueba en el siguiente lema.

Lema 5.9. *Supongamos que $\#A^c > nb$. Si ρ satisface las condiciones **C1**, **C2** y **C4** entonces para cada $\boldsymbol{\beta} \in \mathbb{R}^p$, existe*

$$s_n(\boldsymbol{\beta}) = \text{mín} \left\{ s \in \mathbb{R}^+ : \frac{1}{n} \sum_{i=1}^{n} \rho \left(\frac{y_i - \mathbf{x}_i^t \boldsymbol{\beta}}{s} \right) = b \right\}. \tag{32}$$

Se definen los S-estimadores (en caso de existir) como

$$\widetilde{\boldsymbol{\beta}}_n = \arg \min_{\boldsymbol{\beta}} s_n(\boldsymbol{\beta}) \tag{33}$$

$$\widetilde{\sigma}_n = \min_{\boldsymbol{\beta}} s_n(\boldsymbol{\beta}). \tag{34}$$

En particular $\widetilde{\sigma}_n = s_n(\widetilde{\boldsymbol{\beta}}_n)$ y

$$\frac{1}{n} \sum_{i=1}^{n} \rho \left(\frac{y_i - \mathbf{x}_i^t \widetilde{\boldsymbol{\beta}}_n}{\widetilde{\sigma}_n} \right) = b. \tag{35}$$

Lema 5.10. *Supongamos que ρ satisface **C1** a **C5** y que existen los S-estimadores $\widetilde{\boldsymbol{\beta}}_n$ y $\widetilde{\sigma}_n$ definidos por (33) y (34) (en particular $\widetilde{\sigma}_n = s_n(\widetilde{\boldsymbol{\beta}}_n)$). Luego $\widetilde{\boldsymbol{\beta}}_n$ es también un M estimador, es decir*

$$\widetilde{\boldsymbol{\beta}}_n = \arg \min_{\boldsymbol{\beta}} \frac{1}{n} \sum_{i=1}^{n} \rho \left(\frac{y_i - \mathbf{x}_i^t \boldsymbol{\beta}}{\widetilde{\sigma}_n} \right) \tag{36}$$

El siguiente resultado no es necesario para probar la consistencia fuerte de los S-estimadores, sin embargo refuerza la relación que hay entre S y M estimadores de regresión.

Lema 5.11. *Supongamos que ρ verifica **C1** a **C5** y **C7** y que existen $\widetilde{\boldsymbol{\beta}}_n$ definido como en (36) y $\widetilde{\sigma}_n = s_n(\widetilde{\boldsymbol{\beta}}_n)$, luego $\widetilde{\boldsymbol{\beta}}_n$ y $\widetilde{\sigma}_n$ resultan S-estimadores.*

Consistencia de Fisher

Fijados $\boldsymbol{\beta} \in \mathbb{R}^p$ y $b \in (0,1)$, se llama **escala residual** a cada solución positiva s (en caso de existir) de la siguiente ecuación

$$E_{\boldsymbol{\beta}_0,\sigma_0}\left(\rho\left(\frac{y-\mathbf{x}^t\boldsymbol{\beta}}{s}\right)\right) = b. \tag{37}$$

Veamos a continuación condiciones suficientes para que exista solución de (37).

Lema 5.12. *Supongamos que el vector $(\mathbf{x}, y)$ satisfce el modelo (1) y que ρ verifica **C1**, **C2** y **C4**. Luego, para cada $\boldsymbol{\beta} \in \mathbb{R}^p$ existe una escala residual.*

La ecuación (37) podría tener muchas soluciones. Podríamos considerar entonces, para cada $\boldsymbol{\beta}$, el ínfimo de todas ellas. Veamos que bajo ciertas condiciones, el ínfimo se realiza.

Lema 5.13. *Bajo las mismas hipótesis que en el Lema 5.12, existe*

$$s(\boldsymbol{\beta}) = \min\left\{s \in \mathbb{R}^+ : E_{\boldsymbol{\beta}_0,\sigma_0}\left(\rho\left(\frac{y-\mathbf{x}^t\boldsymbol{\beta}}{s}\right)\right) = b\right\}. \tag{38}$$

Lema 5.14. *Supongamos que ρ satisface **C1** a **C5** y **C7**, luego la ecuación (37) tiene solución única.*

Marazzi, Villar y Yohai (2009) obtuvieron el siguiente resultado en otro contexto teórico. Los supuestos son esencialmente los mismos. La prueba del siguiente Teorema se puede encontrar en el Apéndice de esta sección.

Teorema 5.1. *Supongamos que estamos bajo los supuestos del modelo (1) y la condición de identificabilidad **CI**. Sean la densidad f_0 de u estrictamente unimodal y ρ satisfaciendo **C1** a **C7**, tales que $E_0(\rho(u)) = b$. Luego*

$$\boldsymbol{\beta}_0 = \arg\min_{\beta} E_{\boldsymbol{\beta}_0,\sigma_0}\left(\rho\left(\frac{y-\mathbf{x}^t\boldsymbol{\beta}}{\sigma_0}\right)\right) \tag{39}$$

$$\sigma_0 = s(\boldsymbol{\beta}_0) \tag{40}$$

y además $\boldsymbol{\beta}_0$ es único.

Corolario 5.1. *Bajo los mismos supuestos del Teorema 5.1 se tiene que*

$$\boldsymbol{\beta}_0 = \arg\min_{\beta} s(\boldsymbol{\beta}) \tag{41}$$

$$\sigma_0 = \min_{\beta} s(\boldsymbol{\beta}) \tag{42}$$

y además $\boldsymbol{\beta}_0$ es único.

Consistencia fuerte del S-estimador cuando $\mu_0 = 0$

Una propiedad deseable para los S-estimadores $\widetilde{\boldsymbol{\beta}}_n$ y $\widetilde{\sigma}_n$ es que sean fuertemente consistentes de β_0 y σ_0 respectivamente. Mostraremos condiciones suficientes para que esto ocurra.

Notemos para cada distribución H

$$c(H) = \sup_{\mathbf{d} \in \mathbb{R}^p - \{\mathbf{0}\}} \{P_H(\mathbf{x}^t \mathbf{d} = 0)\}.$$

Teorema 5.2. *Sea $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)$ una muestra aleatoria satisfaciendo el modelo (1). Consideremos la densidad f_0 de u estrictamente unimodal y $b < 1 - c(H_{\boldsymbol{\beta}_0, \sigma_0})$ con $c(H_{\boldsymbol{\beta}_0, \sigma_0}) < 1$. Supongamos que ρ verifica **C1** a **C7** y $E_0(\rho(u)) = b$. Luego los S-estimadores resultan (en caso de existir) fuertemente consistentes de los parámetros del modelo $\boldsymbol{\beta}_0$ y σ_0 respectivamente, es decir $\widetilde{\boldsymbol{\beta}}_n \to \boldsymbol{\beta}_0$ y $\widetilde{\sigma}_n \to \sigma_0$ casi seguramente.*

Notar que la condición de identificabilidad **CI** es una condición necesaria, pero no suficiente de $c(H_{\boldsymbol{\beta}_0, \sigma_0}) < 1$.

Consistencia fuerte del S-estimador cuando $\mu_0 \neq 0$

Cuando $\mu_0 \neq 0$ el estimador de la ordenada al origen calculado a partir de (33) no resulta consistente. En lo que sigue veremos cómo obtener estimadores consistentes de todos los parámetros del modelo (1).

Si el modelo tiene ordenada al origen denotamos $\boldsymbol{\beta}_0 = (\alpha_0, \boldsymbol{\alpha}_1)$ con $\alpha_0 \in \mathbb{R}$ y $\boldsymbol{\alpha}_1 \in \mathbb{R}^{p-1}$. Luego

$$y_i = \alpha_0^* + \boldsymbol{\alpha}_1 \overline{\mathbf{x}}_i^t + \sigma_0 w_i, \ i = 1, \ldots, n$$

donde $\alpha_0^* = \alpha_0 + \mu_0 \sigma_0$ y $w_i = u_i - \mu_0$. Si consideramos $\beta_0^* = (\alpha_0^*, \boldsymbol{\alpha}_1)$ entonces

$$y_i = \mathbf{x}_i^t \boldsymbol{\beta}_0^* + \sigma_0 w_i, \ i = 1, \ldots, n.$$

Además $E_{G_0}(\rho(w)) = E_0(\rho(u - \mu_0)) = b$, siendo G_0 la función de distribución acumulada del nuevo error w.

Luego bajo las hipótesis del Teorema 5.2 se deduce inmediatamente que los S-estimadores $\tilde{\beta}_n$ y $\tilde{\sigma}_n$ convergen casi seguramente a los parámetros del modelo parametrizado, es decir:

$$\tilde{\boldsymbol{\beta}}_n \to \boldsymbol{\beta}_0^* \quad \text{y} \quad \tilde{\sigma}_n \to \sigma_0.$$

Si notamos $\tilde{\boldsymbol{\beta}}_n = (\tilde{\beta}_{n,0}, \tilde{\boldsymbol{\beta}}_{n,1})$ con $\tilde{\beta}_{n,0}$ el estimador de la ordenada al origen y $\tilde{\boldsymbol{\beta}}_{n,1}$ el estimador de la pendiente entonces

$$\tilde{\beta}_{n,0} \to \alpha_0 + \mu_0 \sigma_0 \quad \text{y} \quad \tilde{\boldsymbol{\beta}}_{n,1} \to \alpha_1.$$

Para obtener un estimador consistente de la ordenada al origen α_0 basta considerar $\tilde{\beta}_{n,0} - \mu_0 \tilde{\sigma}_n$.

Formulación alternativa para los S-estimadores cuando $\mu_0 \neq 0$

Sea $(\mathbf{x}_1.y_1), \ldots, (\mathbf{x}_n, y_n)$ una muestra aleatoria y H_n su función de distribución empírica. Fijado $b \in (0,1)$ consideramos, en caso de existir, las soluciones positivas s de la ecuación

$$E_{H_n}\left(\rho\left(\frac{y - \mathbf{x}^t\boldsymbol{\beta}}{s} - \mu_0\right)\right) = b, \tag{43}$$

o de manera equivalente

$$\frac{1}{n}\sum_{i=1}^{n}\rho\left(\frac{y_i - \mathbf{x}_i^t\boldsymbol{\beta}}{s} - \mu_0\right) = b. \tag{44}$$

Supongamos por simplicidad que $s_n(\boldsymbol{\beta})$ es única, luego se definen los S-estimadores como

$$\widetilde{\boldsymbol{\beta}}_n^* = \arg\min_{\boldsymbol{\beta}} s_n^*(\boldsymbol{\beta}) \tag{45}$$

$$\widetilde{\sigma}_n^* = \min_{\boldsymbol{\beta}} s_n^*(\boldsymbol{\beta}). \tag{46}$$

En particular $\widetilde{\sigma}_n^* = s_n^*(\widetilde{\boldsymbol{\beta}}_n^*)$ y además

$$\frac{1}{n}\sum_{i=1}^{n}\rho\left(\frac{y_i - \mathbf{x}_i^t\widetilde{\boldsymbol{\beta}}_n^*}{\widetilde{\sigma}_n^*} - \mu_0\right) = b,$$

siendo b de tal forma que

$$E_0\left(\rho\left(u - \mu_0\right)\right) = b.$$

Sin embargo, esta definición de S-estimadores genera una alta complejidad en los cálculos a la hora de probar la consistencia de los mismos, debido entre otras cosas a que la función

$$h_{\boldsymbol{\beta}}(s) = \rho\left(\frac{y - \mathbf{x}^t\boldsymbol{\beta}}{s} - \mu_0\right)$$

no es monótona si $\mu_0 \neq 0$.

En lo que sigue, buscaremos la relación entre los S-estimadores $\widetilde{\boldsymbol{\beta}}_n^*$ y $\widetilde{\sigma}_n^*$ definidos por (45) y (46) y aquellos definidos por (33) y (34), es decir $\widetilde{\boldsymbol{\beta}}_n$ y $\widetilde{\sigma}_n$. Consideremos por simplicidad un modelo que sólo tiene parámetros de posición y escala para la muestra aleatoria $y_1, \ldots, y_n$, es decir

$$y_i = \beta_0 + \sigma_0 u_i, \quad i = 1, \ldots, n. \tag{47}$$

Vamos a suponer unicidad de $s_n(\beta)$ para facilitar los cálculos. A partir de

$$\frac{1}{n}\sum_{i=1}^{n}\rho\left(\frac{y_i - \beta}{s_n^*(\beta)} - \mu_0\right) = \frac{1}{n}\sum_{i=1}^{n}\rho\left(\frac{y_i - \beta - \mu_0 s_n^*(\beta)}{s_n^*(\beta)}\right) = b,$$

se deduce que

$$s_n(\beta + \mu_0 s_n^*(\beta)) = s_n^*(\beta). \tag{48}$$

Esta relación motiva el siguiente resultado el cual muestra que $\widetilde{\beta}_n^* + \mu_0\widetilde{\sigma}_n^*$ y $\widetilde{\sigma}_n^*$ son los S-estimadores clásicos para los parámetros del modelo (47).

Lema 5.15. *Consideremos un modelo que sólo tiene parámetros de posición y escala. Supongamos que para cada $\boldsymbol{\beta}$, existe y es única $s_n^*(\beta)$ y sean $\widetilde{\beta}_n^*$ y $\widetilde{\sigma}_n^*$ los S-estimadores correspondientes definidos por (45) y (46). Del mismo modo, asumamos que para cada β, existe y es única $s_n(\boldsymbol{\beta})$ y sean $\widetilde{\beta}_n$ y $\widetilde{\sigma}_n$ los estimadores definidos por (33) y (34). Luego*

$$\begin{aligned}\widetilde{\beta}_n - \mu_0\widetilde{\sigma}_n &= \arg\min_{\beta} s_n^*(\beta)\\ \widetilde{\beta}_n^* + \mu_0\widetilde{\sigma}_n^* &= \arg\min_{\beta} s_n(\boldsymbol{\beta})\\ \widetilde{\sigma}_n &= \widetilde{\sigma}_n^*.\end{aligned}$$

Si $s_0 \neq 1$ se tiene $\widetilde{\sigma}_n^* = s_0\widetilde{\sigma}_n$.

En el caso general de regresión con ordenada al origen, notemos $\widetilde{\boldsymbol{\beta}}_n = (\widetilde{\beta}_{n,0}, \widetilde{\boldsymbol{\beta}}_{n,1})$ y $\widetilde{\boldsymbol{\beta}}_n^* = (\widetilde{\beta}_{n,0}^*, \widetilde{\boldsymbol{\beta}}_{n,1}^*)$ donde $\widetilde{\beta}_{n,0}$ y $\widetilde{\beta}_{n,0}^*$ corresponden a los estimadores de las ordenadas al origen y $\widetilde{\boldsymbol{\beta}}_{n,1}$ y $\widetilde{\boldsymbol{\beta}}_{n,1}^*$ los estimadores de las pendientes. Luego es fácil ver que

$$\begin{aligned}(\widetilde{\beta}_{n,0} - \mu_0\widetilde{\sigma}_n, \widetilde{\boldsymbol{\beta}}_{n,1}) &= \arg\min_{\boldsymbol{\beta}} s_n^*(\boldsymbol{\beta})\\ (\widetilde{\beta}_{n,0}^* + \mu_0\widetilde{\sigma}_n^*, \widetilde{\boldsymbol{\beta}}_{n,1}^*) &= \arg\min_{\boldsymbol{\beta}} s_n(\boldsymbol{\beta}).\end{aligned}$$

Si (45) tiene solución única entonces $\widetilde{\boldsymbol{\beta}}_n^* = (\widetilde{\beta}_{n,0} - \mu_0\widetilde{\sigma}_n, \widetilde{\boldsymbol{\beta}}_{n,1})$ y por lo tanto $\widetilde{\boldsymbol{\beta}}_n = (\widetilde{\beta}_{n,0}^* + \mu_0\widetilde{\sigma}_n^*, \widetilde{\boldsymbol{\beta}}_{n,1}^*)$. Ocurre lo mismo si (33) tiene solución única.

6. S-estimadores de regresión lineal para datos censurados propuestos por Locatelli, Marazzi y Yohai

Al considerar ahora el caso censurado ya no contamos con la función de distribución empírica convencional H_n, sino con la función de distribución empírica para datos censurados $(\mathbf{x}_1, y_1^*, \delta_1), \ldots, (\mathbf{x}_n, y_n^*, \delta_n)$. Recordemos su definición que fue dada en (9) y (10).

$$\begin{aligned}H_{n,\boldsymbol{\beta},\sigma}(\mathbf{z}, z) &= \frac{1}{n}\sum_{i=1}^{n} I(\mathbf{x}_i \leq \mathbf{z}) E_{\boldsymbol{\beta},\sigma}(I(y_i \leq z)|\mathbf{x}_i, y_i^*, \delta_i)\\ &= \frac{1}{n}\sum_{i=1}^{n} \left(\delta I(\mathbf{x}_i \leq \mathbf{z}) I(y_i^* \leq z) + (1-\delta) I(\mathbf{x}_i \leq \mathbf{z}) E_{\boldsymbol{\beta},\sigma}(I(y_i \leq z)|\mathbf{x}_i, y_i^*, y_i > y_i^*)\right).\end{aligned}$$

Hemos notado $E_{n,\boldsymbol{\beta},\sigma}(h(\mathbf{x}, y))$ a la esperanza de $h(\mathbf{x}, y)$ basada en $H_{n,\boldsymbol{\beta},\sigma}$. A partir del Lema 3.2, junto con (4) obtuvimos las siguientes formulaciones equivalentes:

$$\begin{aligned}E_{n,\boldsymbol{\beta},\sigma}(h(\mathbf{x}, y)) &= \frac{1}{n}\sum_{i=1}^{n} E_{\boldsymbol{\beta},\sigma}(h(\mathbf{x}_i, y_i)|\mathbf{x}_i, y_i^*, \delta_i)\\ &= \frac{1}{n}\sum_{i=1}^{n} \left(\delta_i h(\mathbf{x}_i, y_i^*) + (1-\delta_i) E_{\boldsymbol{\beta},\sigma}(h(\mathbf{x}_i, y_i)|\mathbf{x}_i, y_i^*, y_i > y_i^*)\right).\end{aligned}$$

Para el caso particular

$$h(\mathbf{x}, y) = \rho\left(\frac{y - \mathbf{x}^t\boldsymbol{\beta}}{\sigma}\right)$$

quedaría

$$E_{n,\boldsymbol{\beta},\sigma}\left(\rho\left(\frac{y - \mathbf{x}^t\boldsymbol{\beta}}{\sigma}\right)\right) =$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left(\delta_i\rho\left(\frac{y_i^* - \mathbf{x}_i^t\boldsymbol{\beta}}{\sigma}\right) + (1-\delta_i)E_{\boldsymbol{\beta},\sigma}\left(\rho\left(u_i\right)\middle| u_i > \frac{y_i^* - \mathbf{x}_i^t\boldsymbol{\beta}}{\sigma}\right)\right) \tag{49}$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left(\delta_i\rho\left(\frac{y_i^* - \mathbf{x}_i^t\boldsymbol{\beta}}{\sigma}\right) + (1-\delta_i)\frac{\int_{\frac{y_i^* - \mathbf{x}_i^t\boldsymbol{\beta}}{\sigma}}^{+\infty}\rho\left(u_i\right)f_0(u_i)du_i}{1 - F_0\left(\frac{y_i^* - \mathbf{x}_i^t\boldsymbol{\beta}}{\sigma}\right)}\right). \tag{50}$$

La versión no paramétrica del S-estimador para datos censurados fue estudiada por Salibian-Barrera y Yohai (2008). Locatelli, Marazzi y Yohai (2010) dieron la versión paramétrica del S-estimador para datos censurados que es la que desarrollaremos a continuación.

Consideremos el modelo (1) y una muestra aleatoria $(\mathbf{x}_1.y_1), \ldots, (\mathbf{x}_n, y_n)$ con censura a la derecha. Generalizando la definición de escala para el caso no censurado a partir de la ecuación (30), parecería natural definir para cada $\boldsymbol{\gamma} \in \mathbb{R}^p$, la escala $s_n(\boldsymbol{\gamma})$ como una solución (en caso de existir) de la ecuación

$$E_{n,\boldsymbol{\gamma},s_n(\boldsymbol{\gamma})}\left(\rho\left(\frac{y - \mathbf{x}^t\boldsymbol{\gamma}}{s_n(\boldsymbol{\gamma})}\right)\right) = b, \tag{51}$$

y por lo tanto definir

$$\overline{\boldsymbol{\beta}}_n = \arg\min_{\boldsymbol{\gamma}} s_n(\boldsymbol{\gamma}).$$

A partir de (50) y haciendo uso del Lema 3.1, la ecuación (51) resulta equivalente a

$$\frac{1}{n}\sum_{i=1}^{n}\left(\delta_i\rho\left(\frac{y_i^* - \mathbf{x}_i^t\boldsymbol{\gamma}}{s_n(\boldsymbol{\gamma})}\right) + (1-\delta_i)\frac{\int_{\frac{y_i^* - \mathbf{x}_i^t\boldsymbol{\gamma}}{s_n(\boldsymbol{\gamma})}}^{+\infty}\rho\left(u_i\right)f_0(u_i)du_i}{1 - F_0\left(\frac{y_i^* - \mathbf{x}_i^t\boldsymbol{\gamma}}{s_n(\boldsymbol{\gamma})}\right)}\right) = b.$$

Sin embargo esta definición de estimador presenta algunas dificultades: hay razones para creer que el estimador resultante no es consistente. Se conocen resultados empíricos que muestran que el estimador definido de esa forma no es consistente. Se debe a que $s_n(\boldsymbol{\gamma})$ no es la escala verdadera y la esperanza resulta distorcionada. Uno querría tomar la esperanza para la distribución verdadera y no para un valor cualquiera de $s_n(\boldsymbol{\gamma})$.

Locatelli, Marazzi y Yohai (2010) procedieron de otra forma para llegar a definir un estimador que resulte al menos "intuitivamente consistente". Computacionalmente pudieron observar lo que sería equivalente a la consistencia de Fisher.

Supongamos por un momento que conocemos $\boldsymbol{\beta}_0$. Luego se define la escala $s_n^*(\gamma)$ como una solución, en caso de existir, de la siguiente ecuación

$$E_{n,\boldsymbol{\beta}_0,s_n(\boldsymbol{\beta}_0)}\left(\rho\left(\frac{y-\mathbf{x}^t\boldsymbol{\gamma}}{s_n^*(\boldsymbol{\gamma})}\right)\right)=b, \tag{52}$$

donde $s_n(\boldsymbol{\beta})$ está definido a partir de (51) y el "pseudo" estimador como

$$\overline{\boldsymbol{\beta}}_n^* = \arg\min_{\boldsymbol{\gamma}} s_n^*(\boldsymbol{\gamma}). \tag{53}$$

En este caso (52) toma la forma:

$$\frac{1}{n}\sum_{i=1}^{n}\left(\delta_i\rho\left(\frac{y_i^*-\mathbf{x}_i^t\boldsymbol{\gamma}}{s_n^*(\boldsymbol{\gamma})}\right)+(1-\delta_i)\frac{\displaystyle\int_{\frac{y_i^*-\mathbf{x}_i^t\boldsymbol{\beta}_0}{s_n(\boldsymbol{\beta}_0)}}^{+\infty}\rho\left(\frac{\mathbf{x}_i^t(\boldsymbol{\beta}_0-\boldsymbol{\gamma})+s_n(\boldsymbol{\beta}_0)u_i}{s_n^*(\boldsymbol{\gamma})}\right)f_0(u_i)du_i}{1-F_0\left(\frac{y_i^*-\mathbf{x}_i^t\boldsymbol{\beta}_0}{s_n(\boldsymbol{\beta}_0)}\right)}\right)=b.$$

En el Apéndice correspondiente se encuentra una prueba heurística de la consistencia de este estimador. Desafortunadamente $\overline{\boldsymbol{\beta}}_n^*$ no sirve como estimador, pero sugiere un procedimiento para encontrar otro que podría ser consistente.

Para cada $\boldsymbol{\beta}\in\mathbb{R}^p$ y $\boldsymbol{\gamma}\in\mathbb{R}^p$ se define $S_n(\boldsymbol{\beta},\boldsymbol{\gamma})$ como una solución, en caso de existir, de la siguiente ecuación

$$E_{n,\boldsymbol{\beta},s_n(\boldsymbol{\beta})}\left(\rho\left(\frac{y-\mathbf{x}^t\boldsymbol{\gamma}}{S_n(\boldsymbol{\beta},\boldsymbol{\gamma})}\right)\right)=b, \tag{54}$$

donde $s_n(\boldsymbol{\beta})$ está definido como en (51). En particular $S_n(\boldsymbol{\gamma},\boldsymbol{\gamma})=s_n(\boldsymbol{\gamma})$.

Para cada $\boldsymbol{\beta}\in\mathbb{R}^p$ sea

$$\widetilde{\boldsymbol{\gamma}}_n(\boldsymbol{\beta}) = \arg\min_{\boldsymbol{\gamma}} S_n(\boldsymbol{\beta},\boldsymbol{\gamma}).$$

Notar que $S_n(\boldsymbol{\beta}_0,\boldsymbol{\gamma})=s_n^*(\boldsymbol{\gamma})$ y por lo tanto $\widetilde{\boldsymbol{\gamma}}_n(\boldsymbol{\beta}_0)=\overline{\boldsymbol{\beta}}_n^*$. Luego $\widetilde{\boldsymbol{\gamma}}_n(\boldsymbol{\beta}_0)\to\boldsymbol{\beta}_0$ y esto haría de $\boldsymbol{\beta}_0$ "casi un punto fijo"de $\widetilde{\boldsymbol{\gamma}}_n$. De esta forma resulta natural definir a $\widehat{\beta}_n$ como la ecuación de un punto fijo. Se definen los S-estimadores $\widehat{\beta}_n$ y $\widehat{\sigma}_n$ de $\boldsymbol{\beta}_0$ y σ_0 como

$$\widehat{\boldsymbol{\beta}}_n = \widetilde{\boldsymbol{\gamma}}_n(\widehat{\boldsymbol{\beta}}_n) \tag{55}$$

$$\widehat{\sigma}_n = s_n(\widehat{\boldsymbol{\beta}}_n). \tag{56}$$

Lema 6.1. *Supongamos que ρ es una función continuamente diferenciable y notemos por ψ a la derivada de ρ. Asumamos que existen los S-estimadores de $\boldsymbol{\beta}_0$ y σ_0 para datos censurados definidos por (55) y (56). Luego dichos estimadores verifican el siguiente sistema de ecuaciones:*

$$E_{n,\boldsymbol{\beta},\sigma}\left(\psi\left(\frac{y-\mathbf{x}^t\boldsymbol{\beta}}{\sigma}\right)\mathbf{x}\right)=0 \tag{57}$$

$$E_{n,\boldsymbol{\beta},\sigma}\left(\rho\left(\frac{y-\mathbf{x}^t\boldsymbol{\beta}}{\sigma}\right)\right)=b. \tag{58}$$

Este sistema de ecuaciones tiene como solución el S-estimador buscado, pero podría tener muchas otras soluciones como ocurre, por ejemplo, cuando ψ es redescendientes. Sin embargo, no contamos con un criterio para decidir cuál será el S-estimador que es solución de la ecuación (55) de punto fijo.

El algoritmo para calcular computacionalmente el S-estimador dado por (55) se describe en el trabajo de Locatelli, Marazzi y Yohai (2010). Ellos mostraron computacionalmente que es único. Sin embargo, este algoritmo es de alta complejidad computacional. En la siguiente sección propondremos otro estimador que también está dado por una ecuación de punto fijo, pero de menor complejidad computacional. Esta propuesta parte de considerar a (57) como la ecuación de un punto crítico. Luego bastaría integrarla para obtener la función a optimizar.

7. Nueva propuesta de S-estimadores de regresión lineal con datos censurados

En esta sección propondremos otro estimador que se define a partir de las ecuaciones de los S-estimadores analizados en la sección anterior. Lo desarrollaremos tanto para el caso de errores simétricos como asimétricos y siempre con datos censurados. Esta nueva propuesta permitirá estudiar una propiedad deseable que es la consistencia de Fisher, la que podría facilitar la prueba de la consistencia fuerte de un estimador.

Definición de los nuevos estimadores

Consideremos en primer lugar el modelo de posición y escala para la muestra aleatoria $y_1, \ldots, y_n$

$$y_i = \beta_0 + \sigma_0 u_i, \quad i = 1, \ldots, n,$$

Observemos que este es un caso especial del modelo de regresión correspondiente a una única variable explicativa constantemente igual a uno. Recordemos que y_i es una variable real y los errores u_i son variables aleatorias continuas, idénticamente distribuídas con función de distribución acumulada F_0, siendo F_0 el miembro estándar de una familia paramétrica de posición y escala (con distribución simétrica o asimétrica) y función de densidad f_0. Los parámetros $\boldsymbol{\beta}_0 \in \mathbb{R}$ y $\sigma_0 \in \mathbb{R}$ son desconocidos.

Todos los estimadores propuestos los calcularemos bajo el supuesto $\mu_0 = 0$. En caso contrario, habría que realizar las mismas correcciones propuestas en la Sección 5.

En este caso las ecuaciones (57) y (58) de los S-estimadores $\widehat{\beta}_n$ y $\widehat{\sigma}_n$ de β_0 y σ_0 respectivamente, habiendo observado datos censurados $(y_1^*, \delta_1), \ldots, (y_n^*, \delta_n)$ están dadas por

$$E_{n,\beta,\sigma}\left(\psi\left(\frac{y-\beta}{\sigma}\right)\right) = 0 \tag{59}$$

$$E_{n,\beta,\sigma}\left(\rho\left(\frac{y-\beta}{\sigma}\right)\right) = b. \tag{60}$$

La idea en lo que sigue es tratar a (59) como la ecuación de un punto crítico en β con σ un número real positivo fijo. Definamos para cada σ fijo y cada muestra aleatoria de datos censurados $y_1^*, \ldots, y_n^*$ una función de variable t

$$Q_\sigma(t) = E_{n,t,\sigma}\left(\psi\left(\frac{y-t}{\sigma}\right)\right) = \frac{1}{n}\sum_{i=1}^{n} E_{t,\sigma}\left(\psi\left(\frac{y_i-t}{\sigma}\right)\middle| y_i^*, \delta_i\right).$$

Observemos que (59) es equivalente a $Q_\sigma(\beta) = 0$ y por lo tanto $Q_{\widehat{\sigma}_n}(\widehat{\beta}_n) = 0$. El objetivo será integrar $Q_{\widehat{\sigma}_n}(t)$ con respecto a t y así encontrar una función para la cual $\widehat{\beta}_n$ es punto crítico. Es decir, nos interesa calcular

$$L_{\widehat{\sigma}_n}(\beta) = \int_0^\beta Q_{\widehat{\sigma}_n}(t)dt,$$

ya que

$$\frac{\partial L_{\widehat{\sigma}_n}(\beta)}{\partial\beta}|_{\beta=\widehat{\beta}_n} = E_{n,\widehat{\beta}_n,\widehat{\sigma}_n}\left(\psi\left(\frac{y-\widehat{\beta}_n}{\widehat{\sigma}_n}\right)\right) = Q_{\widehat{\sigma}_n}(\widehat{\beta}_n) = 0.$$

Luego $\widehat{\beta}_n$ y $\widehat{\sigma}_n$ verifican la ecuación (59). Para abreviar notación usaremos σ en lugar de $\widehat{\sigma}_n$, así

$$L_\sigma(\beta) = \frac{1}{n}\sum_{i=1}^{n}\left(\delta_i\int_0^\beta \psi\left(\frac{y_i^*-t}{\sigma}\right)dt + (1-\delta_i)\int_0^\beta E_{t,\sigma}\left(\psi\left(\frac{y_i-t}{\sigma}\right)\middle| y_i^*, y_i > y_i^*\right)dt\right).$$

Para los "términos no censurados"

$$\begin{aligned}\int_0^\beta \psi\left(\frac{y_i^*-t}{\sigma}\right)dt &= -\sigma\left(\rho\left(\frac{y_i^*-\beta}{\sigma}\right) - \rho\left(\frac{y_i^*}{\sigma}\right)\right)\\ &= -\sigma\rho\left(\frac{y_i^*-\beta}{\sigma}\right) + \kappa_1(\sigma, y_i^*).\end{aligned}$$

En virtud de la ecuación (5), para los "términos censurados" se tiene que

$$\begin{aligned}E_{t,\sigma}\left(\psi\left(\frac{y_i-t}{\sigma}\right)\middle| y_i > y_i^*\right) &= \frac{\int_{\frac{y_i^*-t}{\sigma}}^{+\infty}\psi(u_i)f_0(u_i)du_i}{1-F_0\left(\frac{y_i^*-t}{\sigma}\right)}\\ &= E_0\left(\psi(u_i)\middle| u_i > \frac{y_i^*-t}{\sigma}\right).\end{aligned}$$

Con lo cual

$$\int_0^\beta E_{t,\sigma}\left(\psi\left(\frac{y_i-t}{\sigma}\right)\middle| y_i > y_i^*\right)dt = \int_0^\beta \frac{\int_{\frac{y_i^*-t}{\sigma}}^{+\infty}\psi(u_i)f_0(u)du}{1-F_0\left(\frac{y_i^*-t}{\sigma}\right)}dt.$$

A partir de esta igualdad y haciendo la sustitución $s = \frac{y_i^* - t}{\sigma}$ se obtiene

$$\begin{aligned}\int_0^\beta E_{t,\sigma}\left(\psi\left(\frac{y_i - t}{\sigma}\right)\middle| y_i > y_i^*\right) dt &= -\sigma \int_{\frac{y_i^*}{\sigma}}^{\frac{y_i^* - \beta}{\sigma}} \frac{\int_s^{+\infty} \psi(u_i) f_0(u_i) du_i}{1 - F_0(s)} ds \\ &= -\sigma \int_{\frac{y_i^*}{\sigma}}^{\frac{y_i^* - \beta}{\sigma}} E_0(\psi(u_i)|u_i > s) ds \qquad (61) \\ &= -\sigma \int_0^{\frac{y_i^* - \beta}{\sigma}} E_0(\psi(u_i)|u_i > s) ds + \kappa_2(\sigma, y_i^*).\end{aligned}$$

Luego

$$\begin{aligned}L_\sigma(\beta) = -\frac{\sigma}{n} \sum_{i=1}^n \left(\delta_i \rho\left(\frac{y_i^* - \beta}{\sigma}\right) + (1 - \delta_i) \int_0^{\frac{y_i^* - \beta}{\sigma}} E_0\left(\psi(u_i)|u_i > s\right) ds\right) + \\ + \kappa(\sigma, y_1^*, \ldots, y_n^*).\end{aligned}$$

Consideremos la función

$$\begin{aligned}L_\sigma^*(\beta) &= -\frac{L_\sigma(\beta) - \kappa(\sigma, y_1^*, \ldots, y_n^*)}{\sigma} \\ &= \frac{1}{n} \sum_{i=1}^n \left(\delta_i \rho\left(\frac{y_i^* - \beta}{\sigma}\right) + (1 - \delta_i) \int_0^{\frac{y_i^* - \beta}{\sigma}} E_0\left(\psi(u_i)|u_i > s\right) ds\right).\end{aligned}$$

Estamos en condiciones de definir los S-estimadores alternativos como

$$\tilde{\beta}_n = \arg\min_{\beta \in \mathbb{R}} L_{\tilde{\sigma}_n}^*(\beta) \qquad (62)$$

$$\tilde{\sigma}_n = s_n(\tilde{\beta}_n) \qquad (63)$$

donde $s_n(\beta)$ está definida por la ecuación

$$E_{n,\beta,s_n(\beta)}\left(\rho\left(\frac{y - \beta}{s_n(\beta)}\right)\right) = b. \qquad (64)$$

Por lo tanto $\tilde{\beta}_n$ y $\tilde{\sigma}_n$ son soluciones de las ecuaciones (59) y (60).

Veamos que $L_{\tilde{\sigma}_n}^*(\beta)$ está acotada como función de β, lo cual garantizaría la existencia del ínfimo (pero no necesariamente del mínimo).

Lema 7.1. *Supongamos que ρ es una función continuamente diferenciable tal que satisface las condiciones **C1** a **C6-A**. Luego $L_\sigma^*(\beta)$ está acotada como función de β.*

Hasta acá hemos trabajado con el modelo de posición y escala. Sin embargo, podemos extender la definición del estimador al caso general de regresión lineal.

Consideremos cada muestra aleatoria de datos censurados $(\mathbf{x}_1, y_1^*, \delta_1), \ldots, (\mathbf{x}_n, y_n^*, \delta_n)$ y para cada σ fijo, la función

$$Q_\sigma(\boldsymbol{t}) = E_{n,\boldsymbol{t},\sigma}\left(\psi\left(\frac{y - \mathbf{x}^t \boldsymbol{t}}{\sigma}\right)\mathbf{x}\right).$$

Desarrollando esta expresión obtenemos

$$\begin{aligned} Q_\sigma(\boldsymbol{t}) &= \frac{1}{n}\sum_{i=1}^{n} E_{t,\sigma}\left(\psi\left(\frac{y_i-\mathbf{x}_i^t\boldsymbol{t}}{\sigma}\right)\mathbf{x}_i\middle|\,\mathbf{x}_i, y_i^*, \delta_i\right) \\ &\quad \frac{1}{n}\sum_{i=1}^{n}\left(\delta_i\psi\left(\frac{y_i^*-\mathbf{x}_i^t\boldsymbol{t}}{\sigma}\right)\mathbf{x}_i + (1-\delta_i)E_{\boldsymbol{t},\sigma}\left(\psi\left(\frac{y_i-\mathbf{x}_i^t\boldsymbol{t}}{\sigma}\right)\mathbf{x}_i\middle|\,\mathbf{x}_i, y_i^*, y_i > y_i^*\right)\right) \\ &= \frac{1}{n}\sum_{i=1}^{n}\left(\delta_i\psi\left(\frac{y_i^*-\mathbf{x}_i^t\boldsymbol{t}}{\sigma}\right)\mathbf{x}_i + (1-\delta_i)\mathbf{x}_i E_0\left(\psi(u_i)\middle|u_i > \frac{y_i^*-\mathbf{x}_i^t\boldsymbol{t}}{\sigma}\right)\right). \end{aligned}$$

Análogamente al caso de posición y escala, denotamos para $\sigma \in \mathbb{R}^+$ y $\beta \in \mathbb{R}^p$

$$L_\sigma^*(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n}\left(\delta_i\rho\left(\frac{y_i^*-\mathbf{x}_i^t\boldsymbol{\beta}}{\sigma}\right) + (1-\delta_i)\int_0^{\frac{y_i^*-\mathbf{x}_i^t\boldsymbol{\beta}}{\sigma}} E_0\left(\psi\left(u_i\right)|u_i > s\right)ds\right).$$

Luego podemos definir los S-estimadores alternativos para el caso general de regresión lineal y datos censurados, de la siguiente manera:

Consideremos

$$g(\sigma) = \arg\min_{\boldsymbol{\beta}} L_\sigma^*(\boldsymbol{\beta}). \tag{65}$$

Luego definimos el estimador de σ_0 por el valor $\tilde{\sigma}_n$ que satisface la siguiente ecuación de punto fijo

$$\tilde{\sigma}_n = s_n(g(\tilde{\sigma}_n)), \tag{66}$$

donde $s_n(\boldsymbol{\beta})$ está dado por la ecuación

$$E_{n,\boldsymbol{\beta},s_n(\boldsymbol{\beta})}\left(\rho\left(\frac{y-\mathbf{x}^t\boldsymbol{\beta}}{s_n(\boldsymbol{\beta})}\right)\right) = b. \tag{67}$$

El estimador $\widetilde{\boldsymbol{\beta}}_n$ de $\boldsymbol{\beta}_0$ se define por

$$\widetilde{\boldsymbol{\beta}}_n = g(\tilde{\sigma}_n),$$

o de manera equivalente:

$$\tilde{\boldsymbol{\beta}}_n = \arg\min_{\boldsymbol{\beta}\in\mathbb{R}} L_{\tilde{\sigma}_n}^*(\boldsymbol{\beta}) \tag{68}$$

$$\tilde{\sigma}_n = s_n(\widetilde{\boldsymbol{\beta}}_n). \tag{69}$$

Notar que $\tilde{\boldsymbol{\beta}}_n$ y $\tilde{\sigma}_n$ verifican (58). Como la ecuación (57) es equivalente a $Q_\sigma(\boldsymbol{\beta}) = 0$, los estimadores también verifican dicha ecuación, es decir $Q_{\tilde{\sigma}_n}(\tilde{\boldsymbol{\beta}}_n) = 0$. Esto se deduce de la definición de $\widetilde{\boldsymbol{\beta}}_n$ ya que

$$\frac{\partial L_{\tilde{\sigma}_n}^*(\boldsymbol{\beta})}{\partial\beta_j} = \frac{1}{n}\sum_{i=1}^{n}\left(\delta_i\psi\left(\frac{y_i^*-\mathbf{x}_i^t\boldsymbol{\beta}}{\tilde{\sigma}_n}\right)x_{ij} + (1-\delta_i)x_{ij}E_0\left(\psi\left(u_i\right)\middle|u_i > \frac{y_i^*-\mathbf{x}_i^t\boldsymbol{\beta}}{\tilde{\sigma}_n}\right)\right).$$

Luego

$$\left.\frac{\partial L^*_{\tilde{\sigma}_n}(\boldsymbol{\beta})}{\partial \beta_j}\right|_{\boldsymbol{\beta}=\widetilde{\boldsymbol{\beta}}_n} =$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left(\delta_i \psi\left(\frac{y_i^* - \mathbf{x}_i^t\widetilde{\boldsymbol{\beta}}_n}{\tilde{\sigma}_n}\right)x_{ij} + (1-\delta_i)E_0\left(\psi\left(u_i - u_0\right)x_{ij}\,\middle|\,\mathbf{x}_i, y_i^*, u_i > \frac{y_i^* - \mathbf{x}_i^t\widetilde{\boldsymbol{\beta}}_n}{\tilde{\sigma}_n}\right)\right)$$

$$= 0.$$

Como

$$E_0\left(\psi\left(u_i - u_0\right)x_{ij}\,\middle|\,\mathbf{x}_i, y_i^*, u_i > \tfrac{y_i^* - \mathbf{x}_i^t\widetilde{\boldsymbol{\beta}}_n}{\tilde{\sigma}_n}\right) = E_{\widetilde{\boldsymbol{\beta}}_n,\widehat{\sigma}_n}\left(\psi\left(\tfrac{y_i^* - \mathbf{x}_i^t\widetilde{\boldsymbol{\beta}}_n}{\tilde{\sigma}_n}\right)x_{ij}\,\middle|\,\mathbf{x}_i, y_i^*, y_i > y_i^*\right),$$

entonces

$$\begin{aligned}\left.\frac{\partial L^*_{\tilde{\sigma}_n}(\boldsymbol{\beta})}{\partial \beta_j}\right|_{\boldsymbol{\beta}=\widetilde{\boldsymbol{\beta}}_n} &= \frac{1}{n}\sum_{i=1}^{n} E_{\widetilde{\boldsymbol{\beta}}_n,\tilde{\sigma}_n}\left(\psi\left(\frac{y_i - \mathbf{x}_i^t\widetilde{\boldsymbol{\beta}}_n}{\tilde{}_n}\right)x_{ij}\,\middle|\,\mathbf{x}_i, y_i^*, \delta_i\right)\\ &= E_{n,\widetilde{\boldsymbol{\beta}}_n,\widehat{\sigma}_n}\left(\psi\left(\frac{y - \mathbf{x}^t\widetilde{\boldsymbol{\beta}}_n}{\tilde{\sigma}_n}\right)x_j\right)\\ &= \left(Q_{\tilde{\sigma}_n}(\widetilde{\boldsymbol{\beta}}_n)\right)_j\\ &= 0.\end{aligned}$$

Es decir, ambos estimadores, el propuesto en esta sección a partir de las ecuaciones (68) y (69) y el del "punto fijo"dado en (55) verifican el mismo sistema de ecuaciones (57) y (58). Si el sistema tuviera solución única, ambos estimadores coincidirían. Sin embargo, esto no se puede garantizar cuando ψ es redescendiente.

Consistencia de Fisher

Recordemos que para cada $\boldsymbol{\beta} \in \mathbb{R}^p$, la escala $s_n(\boldsymbol{\beta})$ fue definida en (51) como cada solución (en caso de existir) de la ecuación

$$E_{n,\boldsymbol{\beta},s_n(\boldsymbol{\beta})}\left(\rho\left(\frac{y - \mathbf{x}^t\boldsymbol{\beta}}{s_n(\boldsymbol{\beta})}\right)\right) = b, \tag{70}$$

o de manera equivalente

$$\frac{1}{n}\sum_{i=1}^{n}\left(\delta_i\rho\left(\frac{y_i^* - \mathbf{x}_i^t\boldsymbol{\beta}}{s_n(\boldsymbol{\beta})}\right) + (1-\delta_i)E_{\boldsymbol{\beta},s_n(\boldsymbol{\beta})}\left(\rho\left(u_i\right)\,\middle|\,u_i > \frac{y_i^* - \mathbf{x}_i^t\boldsymbol{\beta}}{s_n(\boldsymbol{\beta})}\right)\right) = b.$$

Luego resulta natural, fijados $\boldsymbol{\beta} \in \mathbb{R}^p$ y $b \in (0,1)$, llamar **escala residual** a cada solución positiva $s(\boldsymbol{\beta})$ (en caso de existir) de la siguiente ecuación

$$E\left(\delta\rho\left(\frac{y^* - \mathbf{x}^t\boldsymbol{\beta}}{s(\boldsymbol{\beta})}\right) + (1-\delta)E_{\boldsymbol{\beta},s(\boldsymbol{\beta})}\left(\rho\left(u\right)\,\middle|\,u > \frac{y^* - \mathbf{x}^t\boldsymbol{\beta}}{s(\boldsymbol{\beta})}\right)\right) = b. \tag{71}$$

A continuación veremos un importante resultado que sin el mismo sería poco probable que el estimador que acabamos de definir, diera como origen a un estimadoror consistente.

Lema 7.2. *Supongamos que estamos bajo el modelo (1) y que* $E_0(\rho(u)) = b$. *Luego* $s(\beta_0) = \sigma_0$

Teorema 7.1. *Supongamos que estamos bajo el modelo (1) para el caso de posición y escala, es decir* $y = \beta_0 + \sigma_0 u$ *y que la densidad de los errores* f_0 *es estrictamente unimodal. Asumamos que* ρ *es una función continuamente diferenciable la cual satisface las condiciones* ***C1*** *a* ***C7*** *y* $E_0(\rho(u)) = b$. *Notemos por* ψ *a la derivada de* ρ *y supongamos que para cada* $\boldsymbol{\beta} \in \mathbb{R}^p$ *existe la escala residual definida como en (71) en el caso particular de posición y escala. Luego*

$$\beta_0 = \arg\min_{\beta \in \mathbb{R}} E\left(\delta\rho\left(\frac{y^* - \beta}{\sigma_0}\right) + (1-\delta)\int_0^{\frac{y^*-\beta}{\sigma_0}} E_0\left(\psi\left(u\right)|u > s\right) ds\right)$$
$$\sigma_0 = s(\beta_0)$$

y además β_0 *es único.*

Del teorema anterior se deduce el caso general, el cual enunciamos a continuación.

Corolario 7.1. *Supongamos que estamos bajo el modelo (1) con la condición de identificabilidad* ***CI*** *y que la densidad de los errores* f_0 *es estrictamente unimodal. Asumamos que* ρ *es una función continuamente diferenciable la cual satisface las condiciones* ***C1*** *a* ***C7*** *y* $E_0(\rho(u)) = b$. *Notemos por* ψ *a la derivada de* ρ *y supongamos que para cada* $\boldsymbol{\beta} \in \mathbb{R}^p$ *existe la escala residual definida como en (71). Luego*

$$\boldsymbol{\beta}_0 = \arg\min_{\boldsymbol{\beta} \in \mathbb{R}} E\left(\delta\rho\left(\frac{y^* - \mathbf{x}^t\boldsymbol{\beta}}{\sigma_0}\right) + (1-\delta)\int_0^{\frac{y^*-\mathbf{x}^t\boldsymbol{\beta}}{\sigma_0}} E_0\left(\psi\left(u\right)|u > s\right) ds\right)$$
$$\sigma_0 = s(\boldsymbol{\beta}_0)$$

y además $\boldsymbol{\beta}_0$ *es único.*

8. Estimadores finales por máxima verosimilitud truncada

Los S-estimadores tienen un alto punto de ruptura pero como veremos en la Sección 9 son bastante ineficientes comparados con el estimador de máxima verosimilitud. Por tal motivo, Locatelli, Marazzi y Yohai (2010), consideraron y desarrollaron un estimador final conocido como "Truncated Maximum Likelihood", que notaremos y llamaremos MVT (máxima verosimilitud truncada) el cual es altamente robusto (si el estimador inicial lo es) y alcanza una eficiencia asintótica alta con respecto al estimador de máxima verosimilitud para datos censurados. Eso se logra rechazando observaciones que aparecen como sospechosas de ser *outliers* y luego aplicar un estimador de máxima verosimilitud "corregido" a través de una regla de corte adaptada al modelo y a los datos. Serán rechazadas aquellas observaciones que sean menores a un determinado valor de la verosimilitud para el modelo. Este estimador final que llamaremos estimador de máxima

verosimilitud truncada, fue estudiado inicialmente por Gervini y Yohai (2002) para el modelo de regresión y errores simétricos. Ellos probaron bajo hipótesis generales que estos estimadores, asintóticamente mantienen el punto de ruptura del estimador inicial y son tan eficientes como el de máxima verosimilitud bajo errores normales. Posteriorente fue extendido a errores asimétricos por Marazzi y Yohai (2004). En el 2010 Locatelli, Marazzi y Yohai lo generalizan a datos censurados. Dicho estimador es una extensión natural del estimador de máxima verosimilitud truncada para observaciones no censuradas propuesta por Marazzi y Yohai (2004).

Recordemos en una primera instancia, la versión original de Gervini y Yohai para comprender mejor el origen de las generalizaciones:

Consideremos inicialmente estimadores de regresión robustos $\tilde{\boldsymbol{\beta}}_n$ y $\tilde{\sigma}_n$ como por ejemplo los S-estimadores. Tomemos los residuos estandarizados definidos como

$$r_i = \frac{y_i - \mathbf{x}_i^t \tilde{\boldsymbol{\beta}}_n}{\tilde{\sigma}_n}.$$

Un valor grande de $|r_i|$ sugeriría que $(\mathbf{x}, y)$ es un *outlier*. Asumiendo un modelo de errores normales, parecería razonable considerar *outliers* aquellos datos con $|r_i| \geq 2{,}5$. Siguiendo esta idea, Rousseeuw y Leroy (1987) definieron

$$w_i = \begin{cases} 1 & \text{si } |r_i| < t_0 \\ 0 & \text{si } |r_i| \geq t_0 \end{cases}$$

con $t_0 = 2{,}5$ y calcularon el estimador de mínimos cuadrados pesados haciendo

$$\tilde{\tilde{\boldsymbol{\beta}}}_n = (X^t W X)^{-1} X^t W Y,$$

donde $W = \text{diag}(w_1, \ldots, w_n)$ e $Y = (y_1, \ldots, y_n)^t$. Se sabe que este procedimiento reduce la variabilidad, sin embargo el estimador resultante no logra ser asintóticamente eficientes bajo errores normales. Por tal motivo Gervini y Yohai propusieron un valor de corte "adaptado" a los residuos que son construidos de tal forma que la proporción de datos eliminados tiende a 0 cuando n tiende a infinito, bajo normalidad de los errores. El estimador de mínimos cuadrados con pesos w_i adaptados, resulta asintóticamente eficiente bajo el modelo de errores normales y robusto bajo desviaciones del modelo lineal. En particular mantiene el punto de ruptura del estimador inicial. El valor de corte se define como sigue:

Consideremos la función de distribución empírica de $|r_1|, \ldots, |r_n|$,

$$F_n^+(t) = \frac{1}{n} \sum_{i=1}^{n} I(|r_i| \leq t).$$

Para detectar *outliers*, se compara $F_n^+(t)$ con la función de distribución de los errores en valor absoluto (es decir $|u|$) bajo el modelo, que notaremos $F_0^+(t)$. Si $F_n^+(t) < F_0^+(t)$, o lo que es lo mismo $\#\{i/|r_i| > t\}/n > P(|u| > t)$, la proporción muestral de residuos en valor absoluto que supera a t es más grande que la proporción teórica. Si esto sucede

para un valor de t grande, significa que los *outliers* están presentes en la muestra. Como una medida de *outliers* en la muestra, los autores definieron

$$d_n = \sup_{t \geq \eta} \{F_0^+(t) - F_n^+(t)\}^+$$

donde $\{\ \}^+$ denota la parte positiva y η es algún percentil grande de F_0^+. Un valor de $\eta = 2{,}5$ parece razonable. Ellos proponen eliminar la cantidad de $[nd_n]$ observaciones que corresponden a los residuos en valor absoluto mayores, donde $[\ ]$ denota la parte entera. Finalmente tomar como valor de corte $t_n = |r|_{(i_n)}$ con $i_n = n - [nd_n]$ donde $|r|_{(i)}$ es el i-ésimo estadístico de orden de los errores estandarizados en valor absoluto. De esta manera se definen los pesos de la siguiente forma:

$$w_i = \begin{cases} 1 & \text{si } |r_i| < t_n \\ 0 & \text{si } |r_i| \geq t_n \end{cases}.$$

Luego, las observaciones con residuos en valor absoluto más grandes son eliminadas completamente. La función de pesos que fue utilizada arriba es $w(u) = I(u < 1)$ aplicada a $|r_i|/t_n$, pero en general sirve cualquiera que satisfaga las siguientes condiciones:

$w : [0, +\infty) \to [0, 1]$ no creciente, continua a derecha, continua en un entorno de 0, $w(0) = 1,\ w(u) > 0$ para $0 < u < 1$ y $w(u) = 0$ para $u \geq 1$.

Este método funciona bien cuando los errores son simétricos. Para errores asimétricos Marazzi y Yohai (2004) se basaron en esta idea y propusieron una regla de corte adaptado al grado de contaminación basándose en la log-verosimilitud negativa de los errores, que notaremos $\rho(u)$ como describiremos a continuación.

Consideremos los residuos $r_1, \ldots, r_n$ y calculemos la log-verosimilitud negativa empírica de los mismos que notaremos $\rho_i = \rho(r_i)$ para $i = 1, \ldots, n$. Un valor grande de ρ_i corresponde a una observación $(\mathbf{x}_i, y_i)$ con una pequeña verosimilitud bajo el modelo F_0 de los errores.

Sea F_0^+ la función de distribución de $\rho(u)$ bajo el modelo y sea η algún percentil grande de F_0^+. Por ejemplo, un valor posible podría ser $\eta = F_0^{-1}(0{,}99)$. Luego si consideramos $w_i = I(\rho_i < \eta)$ y rechazamos las observaciones $(\mathbf{x}_i, y_i)$ tales que $w_i = 0$, entonces los valores de corte superior e inferior fijos, estarían dados por las soluciones de la ecuación $\rho(z) = \eta$.

Sin embargo, si se quiere que los valores de corte estén adaptados al grado de contaminación habría que proceder de otra forma. Consideremos la función de distribución empírica F_n^+ de $\rho_1, \ldots, \rho_n$ la cual se va a comparar con F_0^+ de la siguiente manera:

Sea $\tilde{F}_{n,t}^+$ la distribución empírica truncada en t, es decir

$$\tilde{F}_{n,t}^+(z) = \begin{cases} F_n^+(z)/F_n^+(t) & \text{si } z \leq t \\ 1 & \text{si } z > t \end{cases}$$

y sean $\bar{t}_n$ y $\underline{t}_n$ tales que

$$\begin{aligned}\bar{t}_n &= \sup_t\{\tilde{F}^+_{n,t}(z) \geq F^+_0(z)\ \forall z \geq \eta\} \\ &= \sup_t\{F^+_n(z)/F^+_0(z) \geq F^+_n(t)\ \forall z \geq \eta\} \\ &= \sup_t\{\inf_{z\geq\eta}\{F^+_n(z)/F^+_0(z)\} \geq F^+_n(t)\},\end{aligned}$$

y análogamente

$$\underline{t}_n = \inf_t\{\inf_{z\geq\eta}\{F^+_n(z)/F^+_0(z)\} \leq F^+_n(t)\}.$$

Notar que $\bar{t}_n \geq \eta$. Llamamos $\alpha_n = \inf_{z\geq\eta}\{F^+_n(z)/F^+_0(z)\}$. Es fácil ver que si $F^+_n(t) \neq \alpha_n$ entonces $\bar{t}_n = \underline{t}_n = t_n$. Si existe un intervalo para el cual $F^+_n(t) = \alpha_n$ entonces $\bar{t}_n > \underline{t}_n$ y por lo tanto se utiliza alguna regla de interpolación adecuada para definir t_n. En cualquier caso se consideran los pesos $w_i = I(\rho_i < t_n)$ y se rechazan las observaciones $(\mathbf{x}_i, y_i)$ con $w_i = 0$.

Los valores de corte adaptados superior e inferior para los residuos se obtiene resolviendo la ecuación $\rho(z) = t_n$. Intuitivamente, la distribución empírica de los residuos restantes tienen colas que son comparables con la distribución F_0, de los errores bajo el modelo.

Para datos censurados y errores posiblemente asimétricos, Locatelli, Marazzi y Yohai lo generalizaron como indicaremos a continuación.

De la misma forma que para datos no censurados se quiere rechazar observaciones cuya verosimilitud bajo el modelo inicial sean más pequeñas que un cierto valor de corte.

En este caso un estimador consistente de F^+_0 está dado por

$$\begin{aligned}H^+_{n,\tilde{\boldsymbol{\beta}}_n,\tilde{\sigma}_n}(z) &= \frac{1}{n}\sum_{i=1}^n \Big(\delta_i I(\rho^*_i \leq z) + (1-\delta_i)E_{\tilde{\boldsymbol{\beta}}_n,\tilde{\sigma}_n}\left(I(\rho_i \leq z)|\mathbf{x}_i, y^*_i, y_i > y^*_i\right)\Big) \\ &= \frac{1}{n}\sum_{i=1}^n \Big(\delta_i I(\rho^*_i \leq z) + (1-\delta_i)P_{\tilde{\boldsymbol{\beta}}_n,\tilde{\sigma}_n}(\rho_i \leq z|\mathbf{x}_i, y^*_i, y_i > y^*_i)\Big)\end{aligned}$$

donde $\rho^*_i = \rho\left(\frac{y^*_i - \mathbf{x}^t_i\tilde{\boldsymbol{\beta}}_n}{\tilde{\sigma}_n}\right)$.

Luego se procede de manera similar al caso anterior: sea $\tilde{H}^+_{n,\tilde{\boldsymbol{\beta}}_n,\tilde{\sigma}_n}$ la distribución empírica truncada en t, es decir

$$\tilde{H}^+_{n,\tilde{\boldsymbol{\beta}}_n,\tilde{\sigma}_n,t}(z) = \begin{cases} H^+_{n,\tilde{\boldsymbol{\beta}}_n,\tilde{\sigma}_n}(z)/H^+_{n,\tilde{\boldsymbol{\beta}}_n,\tilde{\sigma}_n}(t) & \text{si } z \leq t \\ 1 & \text{si } z > t \end{cases}$$

y se toma $\bar{t}_n$ como antes

$$\bar{t}_n = \sup_t\{\tilde{H}^+_{n,\tilde{\boldsymbol{\beta}}_n,\tilde{\sigma}_n,t}(z) \geq F^+_0(z)\ \forall z \geq \eta\}.$$

Gervini y Yohai probaron que si la muestra no contiene *outliers*, $t_n \to +\infty$ casi seguramente. Este resultado también es válido en las generalizaciones siguientes para errores asimétricos y para datos censurados.

9. Resultados numéricos

Esta Sección contiene los resultados de un estudio de simulación por Monte Carlo diseñado con el objetivo de evaluar el procedimiento de estimación propuesto en la Sección 7. Analizaremos el comportamiento de dicho estimador y será comparado con otros, bajo el modelo, con y sin presencia de *outliers*.

Consideremos nuevamente el modelo de posición y escala para la muestra aleatoria $y_1, \ldots, y_n$

$$y_i = \beta_0 + \sigma_0 u_i, \quad i = 1, \ldots, n.$$

Para estudiar la eficiencia de los distintos estimadores consideramos los casos en que todas las observaciones son generadas por este modelo con F_0 igual a las distribuciones normal y log-Weibull estándar. En ambos casos tomaremos $\beta_0 = 1$ y $\sigma_0 = 1$. En el caso normal, los tiempos de censura son elegidos con distribución normal con parámetros $\mu = \mu_1$ y $\sigma = 1$. En el caso log-Weibull, la distribución de los tiempos de censura también será log-Weibull con $\mu = \mu_2$ y $\sigma = 1$. Los valores μ_1 y μ_2 son hallados de manera que el porcentaje de puntos censurados sea del 35 %. Simples cálculos (que se pueden encontrar en el anexo) muestran que $\mu_1 = 1{,}5445$ y $\mu_2 = 1{,}619$.

Para estudiar la robustez se utilizarán las mismas dos distribuciones que para el análisis de la eficiencia, pero reemplazando el 10 % de las observaciones por *outliers* iguales a un único valor y_0. El valor y_0 variará en una grilla en el intervalo $[-6, 7]$ con paso 0.5.

Los S-estimadores son calculados usando una función $\rho = \rho_k$ elegida dentro de la familia bicuadrada de Tukey, para ambos modelos, junto con μ_0 y s_0 verificando $s_0 = 1$, $E_0(\rho_k(u - \mu_0)) = 0{,}5$ y $\mu_0 = \arg\min_\mu E_0\left(\rho_k\left(u - \mu\right)\right)$. En el caso normal y por ser f_0 simétrica resulta $\mu_0 = 0$; luego $k = 1{,}548$. En el caso log-Weibull resultan $k = 1{,}718$ y $\mu_0 = -0{,}135$.

Cada simulación es realizada con $l = 1000$ repeticiones de n observaciones cada una. Para el caso sin *outliers* se consideran cuatro tamaños de muestra: $n = 100, 200, 500$ y 1000. En la simulaciones con *outliers* usamos solamente $n = 100$, dado que hay que repetir las simulaciones con muchas posiciones del *outlier*, para todos los estimadores.

La evaluación de los distintos estimadores se hace calculando su error cuadrático medio. Recordemos que dada una muestra aleatoria $y_1, \ldots, y_n$ y estimadores $\tilde{\beta}_n(y_1, \ldots, y_n)$ y $\tilde{\sigma}_n(y_1, \ldots, y_n)$ de β_0 y σ_0 respectivamente, basados en dicha muestra, se define el error cuadrático medio (ECM) de cada estimador como

$$\mathrm{ECM}(\tilde{\beta}_n) = E((\tilde{\beta}_n(y_1, \ldots, y_n) - \beta_0)^2),$$

$$\mathrm{ECM}(\tilde{\sigma}_n) = E((\tilde{\sigma}_n(y_1, \ldots, y_n) - \sigma_0)^2).$$

Dichos errores cuadráticos medios se estiman considerando l muestras independientes de n observaciones cada una $y_1^j, \ldots, y_n^j$ $(j = 1, \ldots, l)$ mediante

$$\widehat{\mathrm{ECM}}(\tilde{\beta}_n) = \frac{1}{l} \sum_{j=1}^{l} (\tilde{\beta}_n(y_1^j, \ldots, y_n^j) - \beta_0)^2,$$

$$\widehat{\mathrm{ECM}}(\tilde{\sigma}_n) = \frac{1}{l}\sum_{j=1}^{l}(\tilde{\sigma}_n(y_1^j,\ldots,y_n^j) - \sigma_0)^2.$$

Notaremos con S_1 el S-estimador propuesto por Locatelli, Marazzi y Yohai (2010) y por S_2 el S-estimator propuesto en la Sección 7. Los correspondientes estimadores de máxima verosimilitud truncada son denotados por MVT_1 y MVT_2 respectivamente. El estimador de máxima verosimilitud es llamado MV, el cual se calcula usando la función `survreg` del programa R dentro del paquete `survival`.

Los estimadores S_1 y MVT_1 se calculan usando el programa R con el comando `TML.censored` dentro de la librería `RobustAFT` que se puede descargar de CRAN.

En lo que sigue describiremos el algoritmo utilizado para hallar los estimadores del modelo propuesto.

Algoritmo para el cálculo de los S-estimadores alternativos

La ecuación (66) provee un mecanismo para construir el algoritmo de los S-estimadores en cuestión. Notar que $\tilde{\sigma}_n$ actúa como punto fijo de $s_n(g(\sigma))$. A continuación describimos los 4 pasos del algoritmo para el caso del modelo de posición y escala:

Paso 1: dados $y_1^*,\ldots,y_n^*$ calculamos para cada $\sigma > 0$

$$i_0(\sigma) = \arg\min_{1\le i\le n} L_\sigma^*(y_i^*)$$

y llamamos $\beta(\sigma) = y^*_{i_0(\sigma)}$.

Paso 2: Para cada $\sigma > 0$ fijo, calculamos

$$s_n(\beta(\sigma)) = \arg\min_s \left(E_{n,\beta(\sigma),s}\left(\rho\left(\frac{y-\beta(\sigma)}{s}\right)\right) - b\right)^2.$$

Paso 3: Calculamos el S-estimador para el parámetro de escala

$$\tilde{\sigma}_n = \arg\min_\sigma \left(\sigma - s_n(\beta(\sigma))\right)^2.$$

Paso 4: Obtenemos el S-estimador para el parámetro de posición

$$\tilde{\beta}_n = \arg\min_\beta L^*_{\tilde{\sigma}_n}(\beta).$$

En los Pasos 2, 3 y 4 optamos por usar el comando de R `optim` que requiere entre sus argumentos, un punto inicial para iniciar con el proceso iterativo. Por ejemplo, para el Paso 4, el punto inicial que tomamos es $\beta(\tilde{\sigma}_n)$, definido en el Paso 1. Para los puntos iniciales de los Pasos 2 y 3 utilizamos la mediana (normalizada) del valor absoluto de los desvíos con respecto a la mediana, conocido con las siglas MAD. En el anexo correspondiente especificamos mejor este punto.

A continuación daremos un posible algoritmo para el caso general de regresión. El mismo aún no fue programado.

Paso 1: Se toman N submuestras de tamaño p, y se encuentran para cada submuestra el valor de $\boldsymbol{\beta}$ que realiza un ajuste perfecto. Sean estos valorres $\boldsymbol{\beta}_1, ...\boldsymbol{\beta}_N$.

Paso 2: Para cada uno de estos valores se calcula un estimador de un paso de reweighted least squares obteniéndose $\boldsymbol{\beta}_1^*, ...\boldsymbol{\beta}_N^*$.

Paso 3: Se calculan las escalas de estos estimadores. Sea $\boldsymbol{\beta}^*$ el valor que tiene la mínima escala y σ^* la escala respectiva

Paso 4: Se calculan $\gamma_1 = g(\sigma^*)$ (por optimizacion partiendo de $\boldsymbol{\beta}^*$) y la correspondiente escala $s_n(\gamma_1)$. Supongamos que $s_n(g(\sigma^*)) < \sigma^*$. Llamamos $\sigma_1 = \sigma^*$.

Paso 5: Calculamos recursivamente σ_h y γ_h para $h > 1$ hasta que $s_n(\gamma_h) > \sigma_h$ de la siguiente manera: $\sigma_h = \sigma_{h-1} - \varepsilon$, $\gamma_h = g(\sigma_h)$, para algún $\epsilon > 0$ y pequeño. Para calcular γ_h se usa un optimizador comenzando en $\gamma_{h-1.}$

Paso 6: Supongamos que el proceso descripto en 5 se detiene en el paso h_0, . Luego se inicia un algoritmo de biseccción para resolver la ecuación $\sigma = s_n(g(\sigma))$ en el intervalo $[\sigma_{h_0-1}, \sigma_{h_0}]$.

Paso 7: Si en en el paso 4 resultase $s_n(\gamma(\sigma^*)) > \sigma^*$ se realiza un proceso similar pero en el paso 5 se hace $\sigma_h = \sigma_{h-1} + \varepsilon$

Paso 8: El valor de ε se podría tomar igual $\sigma_1/10$

Resultados para el caso sin *outliers*

En los Cuadros 1 y 2 se representan las eficiencias estimadas de los distintos estimadores con respecto al de MV. Dicha eficiencia se define como el cociente entre el ECM del estimador de MV y el ECM del estimador considerado.

Parámetro	*Estimador*	*Cantidad de observaciones*			
		100	*200*	*500*	*1000*
β_0	S_1	0.4471	0.4287	0.4262	0.4234
	S_2	0.4281	0.4273	0.4329	0.4233
	MVT_1	0.9846	0.9905	0.9881	0.9951
	MVT_2	0.9598	0.9852	0.9890	0.9956
σ_0	S_1	0.6374	0.6426	0.6354	0.5850
	S_2	0.6501	0.6350	0.6503	0.5865
	MVT_1	0.7958	0.8347	0.8850	0.8927
	MVT_2	0.7735	0.8234	0.8804	0.8895

Cuadro 1: Eficiencias simuladas para el modelo gaussiano basadas en 1000 muestras.

Parámetro	*Estimador*	*Cantidad de observaciones*			
		100	*200*	*500*	*1000*
β_0	S_1	0.4542	0.4798	0.4732	0.4823
	S_2	0.4524	0.4783	0.4733	0.4837
	MVT_1	0.9376	0.9886	0.9989	0.9986
	MVT_2	0.9534	0.9898	1.0000	0.9983
σ_0	S_1	0.5793	0.6622	0.5866	0.5905
	S_2	0.6138	0.6895	0.6068	0.6036
	MVT_1	0.7480	0.8472	0.8735	0.8851
	MVT_2	0.7427	0.8494	0.8689	0.8867

Cuadro 2: Eficiencias simuladas para el modelo log-Weibull basadas en 1000 muestras.

Como era de esperar para ambos modelos y parámetros, los estimadores de máxima verosimilitud MV son los más eficientes, seguidos por los estimadores de máxima verosimilitud truncada MVT_1 y MVT_2, con eficiencias casi idénticas entre ambos. Los estimadores S_1 y S_2, también tienen eficiencias muy parecidas entre ellos, aunque mucho menores a sus correspondientes versiones finales MVT_1 y MVT_2. Cuando el tamaño de la muestra crece los estimadores de posición MVT_1 y MVT_2 son tan eficientes como el de máxima verosimilitud aunque no ocurre lo mismo con el estimador de escala.

Para cuantificar estas diferencias y distinguiendo por estimador, podemos observar que en el caso del estimador de posición y en ambos modelos, los estimadores finales MVT_1 y MVT_2 son al menos el doble de eficientes que S_1 y S_2, aunque esta diferencia no es tan amplia en la escala.

Comportamiento del S-estimador en presencia de outliers

A continuación se muestran los resultados de las simulaciones correspondienres a muestras contaminadas de 100 observaciones cada una y se calcularon los errores cuadráticos medios empíricos de cada parámetro estimado para diferentes niveles de contaminación. En todos los casos se utilizaron 1000 muestras que fueron contaminadas con una fracción fija del 10 % de las observaciones elegidas al azar, reemplazadas por *outliers* todos iguales a un valor m. Se hicieron las simulaciones con m variando sobre una grilla uniforme de valores entre -6 y 7 con paso de 0.5.

Al final de cada tabla se resume la información considerando solamente el error cuadrático medio máximo de cada parámetro. Es decir se elige el valor de m que produce mayor error cuadrático medio simulado y se muestra sólo este error. Como siempre, se estudia el modelo gaussiano y el log-Weibull para los errores y se compararon los resultados obtenidos para los estimadores S_2 y MVT_2 versus MV, S_1 y MVT_1.

En la Cuadro 3 mostramos los errores cuadráticos medios simulados para los distintos valores de m de los estimadores de posición, con datos contaminados, bajo el modelo gaussiano.

Contaminación	- *Estimadores de* β_0 -				
m	S_1	S_2	MVT_1	MVT_2	MV
-6.0	0.0272	0.0277	0.0135	0.0135	0.0576
-5.5	0.0272	0.0277	0.0135	0.0135	0.0560
-5.0	0.0272	0.0277	0.0135	0.0135	0.0544
-4.5	0.0272	0.0277	0.0135	0.0135	0.0527
-4.0	0.0272	0.0277	0.0135	0.0135	0.0508
-3.5	0.0272	0.0277	0.0135	0.0135	0.0487
-3.0	0.0272	0.0277	0.0136	0.0137	0.0462
-2.5	0.0273	0.0277	0.0153	0.0150	0.0433
-2.0	0.0273	0.0277	0.0256	0.0243	0.0397
-1.5	0.0273	0.0277	0.0329	0.0325	0.0353
-1.0	0.0296	0.0297	0.0292	0.0292	0.0301
-0.5	0.0524	0.0519	0.0236	0.0236	0.0242
0.0	0.0732	0.0739	0.0188	0.0190	0.0180
0.5	0.0378	0.0381	0.0135	0.0136	0.0129
1.0	0.0174	0.0177	0.0118	0.0118	0.0114
1.5	0.0617	0.0645	0.0210	0.0214	0.0172
2.0	**0.0849**	**0.0888**	0.0386	0.0389	0.0345
2.5	0.0561	0.0574	0.0692	0.0691	0.0670
3.0	0.0308	0.0308	0.1141	0.0128	0.1177
3.5	0.0258	0.0260	**0.1439**	**0.1372**	0.1887
4.0	0.0255	0.0258	0.1115	0.1005	0.2816
4.5	0.0255	0.0258	0.0477	0.0410	0.3973
5.0	0.0255	0.0258	0.0217	0.0209	0.5366
5.5	0.0255	0.0258	0.0136	0.0136	0.7000
6.0	0.0255	0.0258	0.0135	0.0135	0.8879
6.5	0.0255	0.0258	0.0134	0.0134	1.1008
7.0	0.0255	0.0258	0.0134	0.0134	**1.3381**
ECM máx	0.0849	0.0888	0.1439	0.1372	1.3381

Cuadro 3: Error cuadrático medio simulado bajo el modelo gaussiano con 10 % de contaminación, para los diferentes estimadores del parámetro de posición. Simulaciones basadas en 1000 muestras de 100 observaciones cada una.

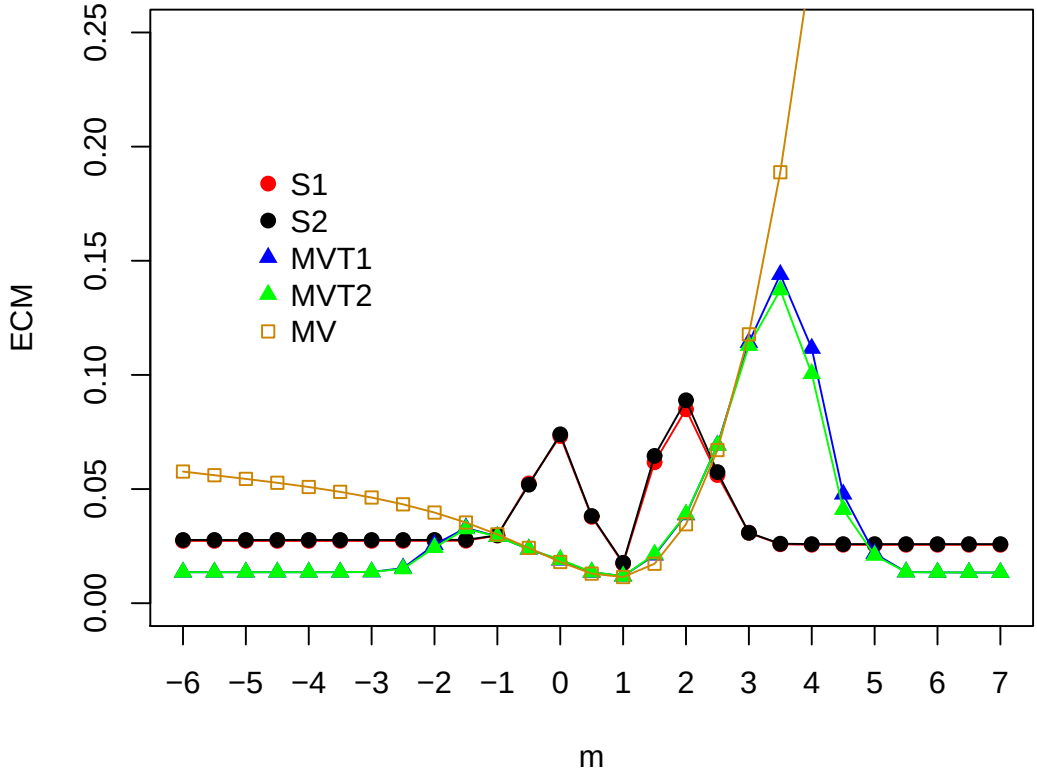

Figura 5: Gráfico correspondiente al Cuadro 3

Se observa inmediatamente comportamientos casi idénticos entre los estimadores S_1 y S_2, como también entre los estimadores finales MVT_1 y MVT_2. Estos últimos resultaron ser los más resistentes para contaminaciones extremas (que podrían ser muy grandes o muy pequeñas). Para valores de m entre -1 y 3, los estimadores finales y el estimador de máxima verosimilitud dan casi los mismos resultados. Esto último parecería razonable porque en definitiva, los estimadores finales son esencialmente estimadores de máxima verosimilitud aplicados a un conjunto de datos sin potenciales *outliers*.

Como era de esperar, todos los estimadores tienen prácticamente el mismo (y bajo) ECM para $m = 1$ en virtud de cómo se eligió el modelo, más específicamente el parámetro de posición $\beta_0 = 1$.

A continuación mostramos los errores cuadráticos medios simulados de los estimadores de escala, con contaminación, bajo el modelo gaussiano.

Contaminación	- *Estimadores de* σ_0 -				
m	S_1	S_2	MVT_1	MVT_2	MV
-6.0	0.0374	0.0313	0.0125	0.0126	1.8216
-5.5	0.0374	0.0313	0.0125	0.0126	1.4532
-5.0	0.0374	0.0313	0.0125	0.0126	1.1310
-4.5	0.0374	0.0313	0.0125	0.0127	0.8542
-4.0	0.0374	0.0313	0.0126	0.0127	0.6214
-3.5	0.0374	0.0313	0.0127	0.0129	0.4311
-3.0	0.0374	0.0313	0.0140	0.0135	0.2811
-2.5	0.0373	0.0312	0.0232	0.0209	0.1687
-2.0	0.0372	0.0311	0.0543	0.0496	0.0903
-1.5	0.0368	0.0318	0.0455	0.0444	0.0414
-1.0	0.0352	0.0295	0.0177	0.0174	0.0161
-0.5	0.0238	0.0201	0.0083	0.0083	0.0077
0.0	0.0111	0.0108	0.0101	0.0105	0.0085
0.5	0.0200	0.0227	0.0187	0.0200	0.0112
1.0	0.0227	0.0259	0.0208	0.0223	0.0109
1.5	0.0124	0.0123	0.0117	0.0124	0.0064
2.0	0.0312	0.0264	0.0110	0.0108	0.0107
2.5	0.0636	0.0549	0.0300	0.0297	0.0287
3.0	**0.0706**	**0.0603**	0.0726	0.0711	0.0735
3.5	0.0702	0.0598	**0.1225**	**0.1165**	0.1543
4.0	0.0702	0.0598	0.1187	0.1056	0.2784
4.5	0.0702	0.0598	0.0547	0.0476	0.4508
5.0	0.0702	0.0598	0.0245	0.0236	0.6752
5.5	0.0702	0.0598	0.0150	0.0151	0.9544
6.0	0.0702	0.0598	0.0147	0.0149	1.2901
6.5	0.0702	0.0598	0.0146	0.0149	1.6838
7.0	0.0702	0.0598	0.0145	0.0148	**2.1366**
ECM máx	0.0706	0.0603	0.1225	0.1165	2.1366

Cuadro 4: Error cuadrático medio simulado bajo el modelo gaussiano con 10 % de contaminación, para los diferentes estimadores del parámetro de escala. Simulaciones basadas en 1000 muestras de 100 observaciones cada una.

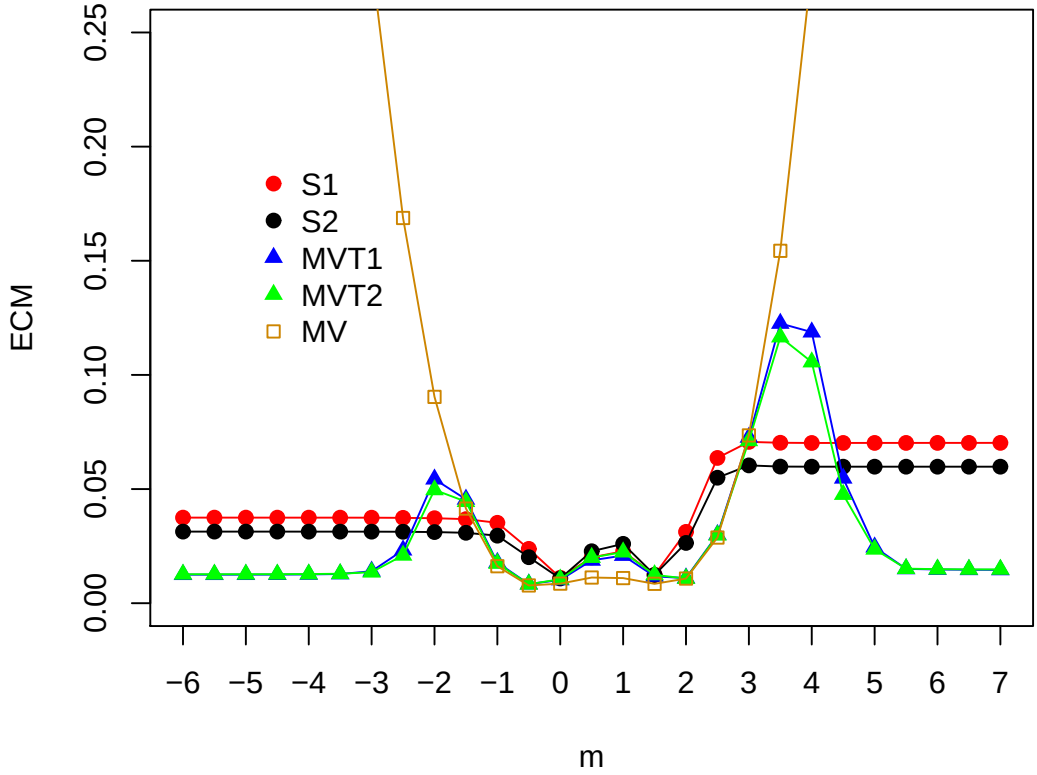

Figura 6: Gráfico correspondiente al Cuadro 4

Similar a lo que sucede en el caso de posición, los dos estimadores finales tienen casi el mismo comportamientos para todos los valores de m y son los más resistentes para *outliers* severos. Se observa también, para esta clase de *outliers* y para casi todos los valores de m, que el estimador S_2 tiene menor error cuadrático medio que S_1, aunque esta diferencia es relativamente menor.

Para los valores de $m = 0$ y $m = 2$ prácticamente todos los resultados coinciden. Esto podría deberse a cómo fueron elegidos los parámetros, $\beta_0 = 1$ y $\sigma_0 = 1$. Para esos valores de m la escala del error es razonable que esté cerca de 1.

A continuación mostramos los errores cuadráticos medios simulados de los estimadores de posición, con contaminación, bajo el modelo log-Weibull.

Contaminación	- *Estimadores de* β_0 -				
m	S_1	S_2	MVT_1	MVT_2	MV
-6.0	0.0347	0.0347	0.0164	0.0162	0.0344
-5.5	0.0347	0.0347	0.0173	0.0171	0.0342
-5.0	0.0347	0.0347	0.0199	0.0193	0.0340
-4.5	0.0347	0.0347	0.0248	0.0236	0.0033
-4.0	0.0347	0.0347	0.0293	0.0287	0.0332
-3.5	0.0347	0.0347	0.0322	0.0318	0.0326
-3.0	0.0347	0.0347	0.0316	0.0316	0.0317
-2.5	0.0347	0.0348	0.0304	0.0303	0.0305
-2.0	0.0348	0.0348	0.0289	0.0287	0.0290
-1.5	0.0365	0.0362	0.0268	0.0267	0.0270
-1.0	0.0563	0.0547	0.0255	0.0252	0.0245
-0.5	0.0962	0.0943	0.0262	0.0247	0.0212
0.0	0.0881	0.0862	0.0258	0.0235	0.0174
0.5	0.0303	0.0300	0.0157	0.0153	0.0138
1.0	0.0259	0.0264	0.0152	0.0152	0.0141
1.5	0.0826	0.0856	0.0335	0.0337	0.0296
2.0	**0.1078**	**0.1116**	0.0877	0.0876	0.0840
2.5	0.0777	0.0781	**0.1885**	0.1910	0.2041
3.0	0.0424	0.0416	0.1807	**0.2423**	0.4077
3.5	0.0359	0.0339	0.0488	0.1076	0.7043
4.0	0.0360	0.0336	0.0171	0.0268	1.0991
4.5	0.0361	0.0336	0.0161	0.0161	1.5953
5.0	0.0361	0.0336	0.0161	0.0154	2.1947
5.5	0.0361	0.0336	0.0161	0.0154	2.8988
6.0	0.0361	0.0336	0.0161	0.0154	3.7084
6.5	0.0361	0.0336	0.0161	0.0154	4.6244
7.0	0.0361	0.0336	0.0161	0.0154	**5.6471**
ECM máx	0.1078	0.1116	0.1885	0.2423	5.6471

Cuadro 5: Error cuadrático medio simulado bajo el modelo log-Weibull con 10 % de contaminación, para los diferentes estimadores del parámetro de posición. Simulaciones basadas en 1000 muestras de 100 observaciones cada una.

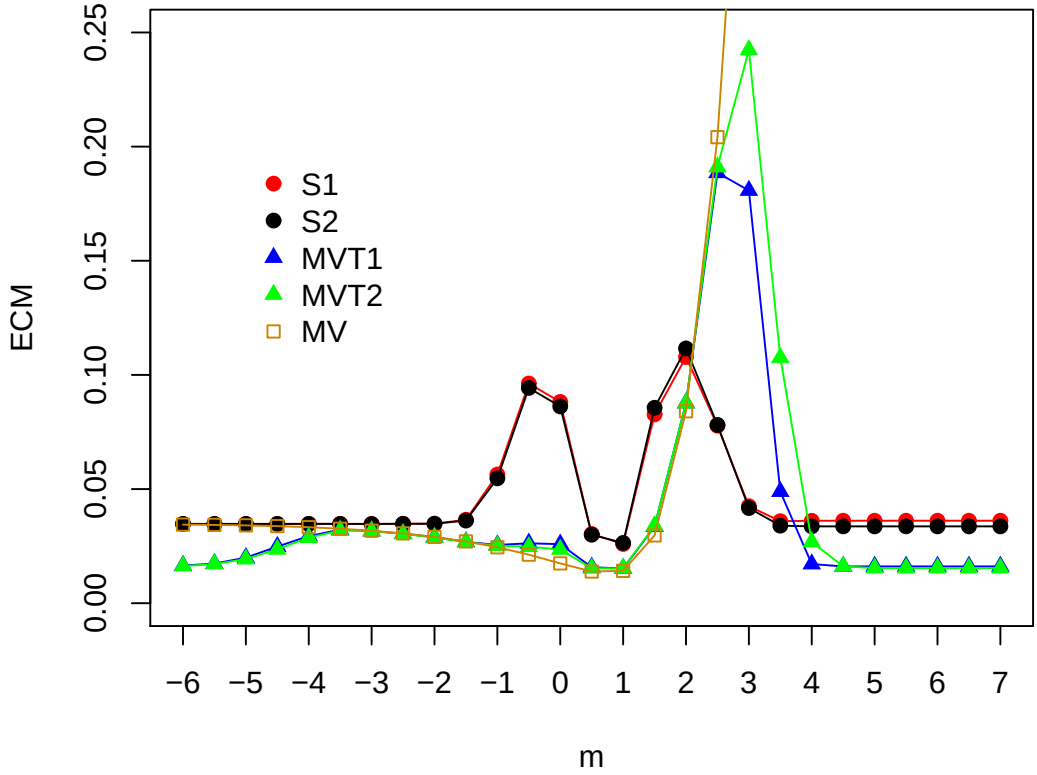

Figura 7: Gráfico correspondiente al Cuadro 5

Notar que S_1 y S_2 tienen resultados casi idénticos. Para valores de m entre 2 y 4, MVT_2 parecería ser menos resistentes que MVT_1. Notar que para los valores de m entre -3 y 2.5 los resultados para los estimadores de máxima verosimilitud y finales son prácticamente los mismos.

Luego en términos generales, las comparaciones qie se puedan hacer entre todos los estimadores son muy parecidas al caso normal.

A continuación mostramos los errores cuadráticos medios simulados de los estimadores de escala, con contaminación, bajo el modelo log-Weibull.

Contaminación	- *Estimadores de* σ_0 -				
m	S_1	S_2	MVT_1	MVT_2	MV
-6.0	0.0428	0.0360	0.0187	0.0173	0.1588
-5.5	0.0427	0.0360	0.0246	0.0233	0.1319
-5.0	0.0426	0.0359	0.0396	0.0345	0.1078
-4.5	0.0425	0.0358	0.0557	0.0505	0.0863
-4.0	0.0423	0.0356	0.0636	0.0610	0.0675
-3.5	0.0421	0.0354	0.0537	0.0532	0.0513
-3.0	0.0417	0.0351	0.0385	0.0384	0.0376
-2.5	0.0411	0.0346	0.0254	0.0260	0.0265
-2.0	0.0402	0.0339	0.0179	0.0177	0.0180
-1.5	0.0382	0.0324	0.0122	0.0121	0.0121
-1.0	0.0285	0.0244	0.0092	0.0092	0.0089
-0.5	0.0138	0.0118	0.0103	0.0098	0.0083
0.0	0.0202	0.0180	0.0211	0.0187	0.0100
0.5	0.0338	0.0341	0.0302	0.0292	0.0129
1.0	0.0258	0.0284	0.0260	0.0272	0.0135
1.5	0.0156	0.0147	0.0132	0.0138	0.0096
2.0	0.0421	0.0358	0.0137	0.0139	0.0136
2.5	**0.0769**	0.0661	0.0559	0.0532	0.0576
3.0	0.0632	**0.0686**	**0.0904**	**0.1091**	0.1694
3.5	0.0309	0.0667	0.0420	0.0715	0.3630
4.0	0.0250	0.0664	0.0233	0.0282	0.6442
4.5	0.0245	0.0664	0.0221	0.0199	1.0153
5.0	0.0245	0.0664	0.0221	0.0188	1.4775
5.5	0.0245	0.0664	0.0221	0.0188	2.0314
6.0	0.0245	0.0664	0.0221	0.0188	2.6774
6.5	0.0245	0.0664	0.0221	0.0188	3.4156
7.0	0.0245	0.0664	0.0221	0.0188	**4.2461**
ECM máx	0.0769	0.0686	0.0904	0.1091	4.2461

Cuadro 6: Error cuadrático medio simulado bajo el modelo log-Weibull con 10 % de contaminación, para los diferentes estimadores del parámetro de escala. Simulaciones basadas en 1000 muestras de 100 observaciones cada una.

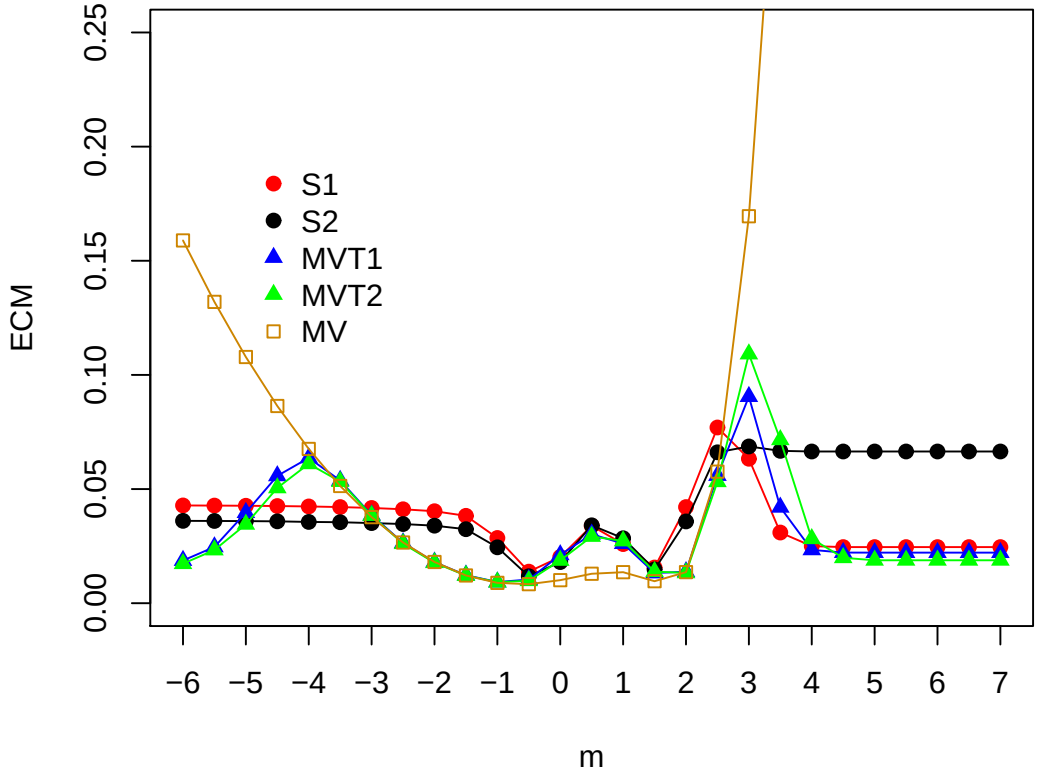

Figura 8: Gráfico correspondiente al Cuadro 6

En este caso sorprende la gran diferencia que hay entre los estimadores S_1 y S_2 para valores mayores a $m = 3$. Sin embargo, los estimadores finales dan prácticamente los mismos resultados en toda la grilla de valores de m.

10. Ejemplo

Esta sección está destinada a analizar un conjunto de datos reales con pacientes a los que se les mide el tiempo de supervivencia en días desde que se inician en un tratamiento. Supondremos un modelo con dos parámetros: ordenada al origen y escala. Analizaremos visualmente los datos para optar por una familia paramétrica adecuada de posición y escala entre dos posibles (gaussiana y log-Weibull), para los errores.

El nombre original del trabajo cuyos datos analizaremos es "*A clinical trial in the treatment of carcinoma of the oropharynx*", disponible en Kalbfleisch y Prentice (1997).

El archivo de datos correspondiente PHARYNX.DAT se puede encontrar en la base de datos de *software* estadístico de Análisis de Supervivencia de la Universidad de Massachusetts, `http://www.umass.edu/statdata/statdata/stat-survival.html`.

Aproximadamente el 27 % de los tiempos de supervivencia son censurados debido principalmente a los pacientes que sobrevivieron al tiempo de análisis. Algunos pacientes se perdieron en el seguimiento debido a que fueron transladados o transferidos a una institución que no participaba del estudio, aunque estos casos son relativamente raros. Desde un punto de vista estadístico, una característica importante de estos datos es la considerable falta de homogeneidad entre los individuos estudiados.

Este estudio incluyó mediciones de muchas covariables que cabría esperar se relacionen con la experiencia de supervivencia. Fueron consideradas variables como sexo, edad, el estado general del paciente, el sitio del tumor primario, las posibles diferencias entre las lesiones, el tipo de tratamiento (tratamiento combinado o terapia de radiación convencional), el tamaño del tumor, la institución, entre otras.

A continuación mostramos parte de la tabla original de 195 pacientes. Nosotros trabajaremos con la variable respuesta, es decir tiempo de supervivencia y su correspondiente información de censura.

N°	Inst.	Sexo	Trat.	Edad	Cond.	...	Censura	Tiempo
1	2	2	1	51	1	...	1	631
2	2	1	2	65	1	...	1	270
3	2	1	1	64	2	...	1	327
4	2	1	1	72	1	...	1	242
5	5	1	2	64	1	...	1	916
6	4	1	2	61	1	...	0	1823
7	4	1	1	65	1	...	1	637
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
193	4	1	2	59	2	...	1	173
194	5	1	2	47	1	...	0	413
195	3	1	2	57	2	...	1	274

Las instituciones participantes fueron 16, identificadas con la variable `Inst.` Para la variable `Sexo`, el valor 1 corresponde a "hombre" y el valor 2 corresponde a "mujer". A la variable `Trat.`(tratamiento) se le asignó valor 1 si el paciente fue sometido a una terapia de radiación convencional estándar y 2 si se le aplicó un tratamiento combinado. La variable `Cond.`(condición) está relacionada con condición general del paciente, a saber: 1: discapacidad, 2: trabajo restringido, 3: requiere asistencia para su cuidado y 4: postrado. Se asigna valor 0 al dato censurado y 1 al paciente que falleció mientras participaba del estudio. El `Tiempo` (de supervivencia desde el ingreso al estudio) fue medido en días. Hay también otras variables que tuvieron que ver con el tamaño y ubicación del tumor así como también la presencia y tipo de metástasis.

Veamos cómo se agrupan los datos totales en la Figura 9. A continuación en la Figura 10 los distinguimos entre "censurados" y "no censurados" a través de gráficos del tipo

boxplot. Las observaciones no censuradas suman un total de 142 quedando para las censuradas 53 (27 %).

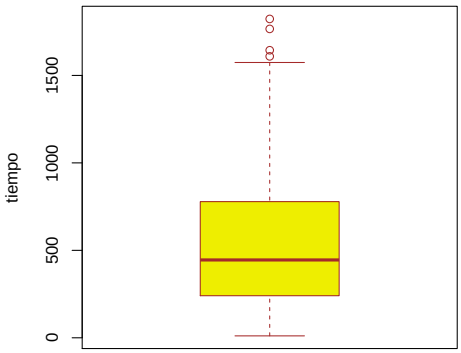

Figura 9: *Boxplot* con todos los datos.

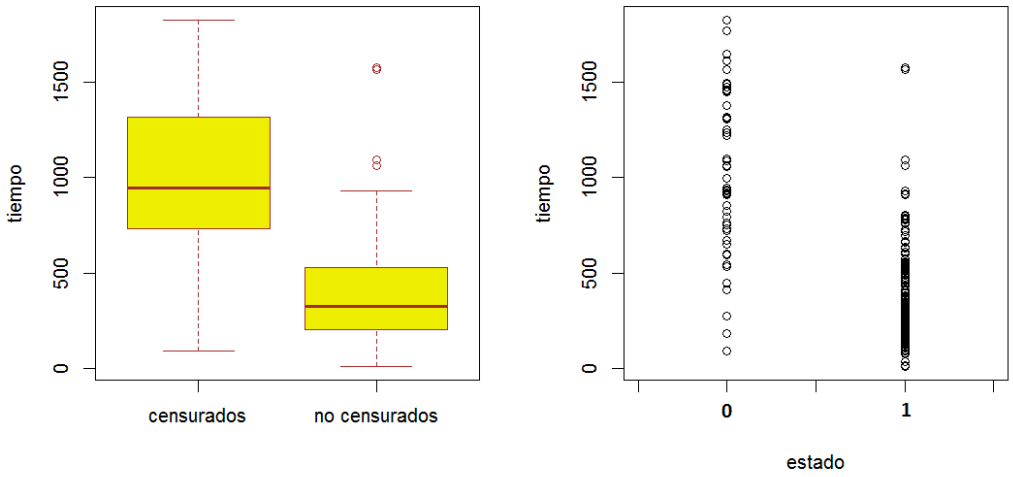

Figura 10: Datos censurados y no censurados por separado.

Se observa en la Figura 9 que la mitad de los datos totales se encuentran entre los 250 y 750 días aproximadamente de supervivencia. Luego, en los *boxplots* de la Figura 10, diferenciando por `Censura` se puede observar que la mitad de los datos censurados son

superiores a casi todos los datos no censurados. Esto podría deberse a que justamente los datos no censurados corresponden a personas que murieron durante el estudio, mientras que los censurados sobrevivieron más allá del tiempo que duró el estudio salvo aquellos que se perdieron, aunque estos casos fueron realmente raros.

Como es habitual el tiempo de supervivencia se suele trabajar en escala logarítmica. De esta forma, ajustamos un modelo de posición y escala para la muestra aleatoria $\log(y_1), \ldots, \log(y_{195})$:

$$\log(y_i) = \beta_0 + \sigma_0 u_i, \quad i = 1, \ldots, 195$$

siendo (y_i^*, δ_i) los valores observados. La densidad f_0 de los errores u_i pertenece a una familia de distribuciónes de posición y escala, como por ejemplo la distribución normal, log-Weibull o logística entre otras, presentadas en las primeras secciones de esta tesis. Notaremos $z_i = \log(y_i)$.

La identificación del modelo paramétrico más adecuado a través de comparaciones visuales son a menudo muy riesgosas y requieren apoyo de técnicas como la bondad de ajuste, entre otras, sobre todo cuando no se puede identificar una única distribución.

Trataremos de identificar un modelo entre dos posibles, uno simétrico como es el normal y otro asimétrico como es el caso de la distribución log-Weibull. Comenzaremos graficando histogramas (Figura 11) y *boxplots* (Figura 13) utilizando todos los datos y luego solamente los datos no censurados. Ninguno de ellos será representativo de la distribución original. Por tal motivo, apelaremos luego a un estimador de dicha distribución para construir los gráficos del tipo *qqplots*.

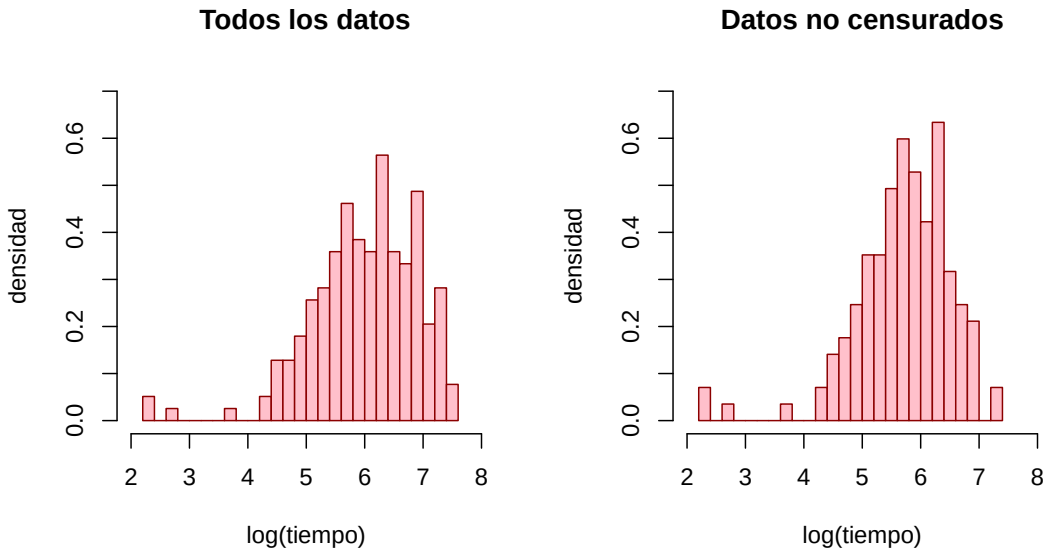

Figura 11: Histograma de área

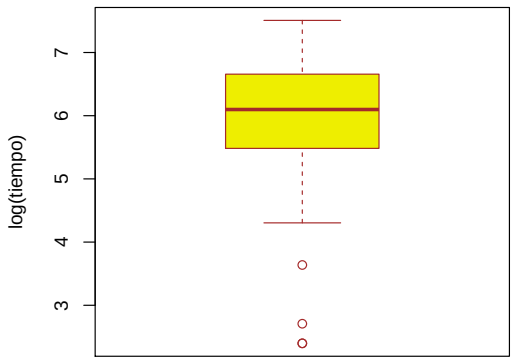

Figura 12: *Boxplot* con todos los datos para los logaritmos de los tiempos de supervivencia

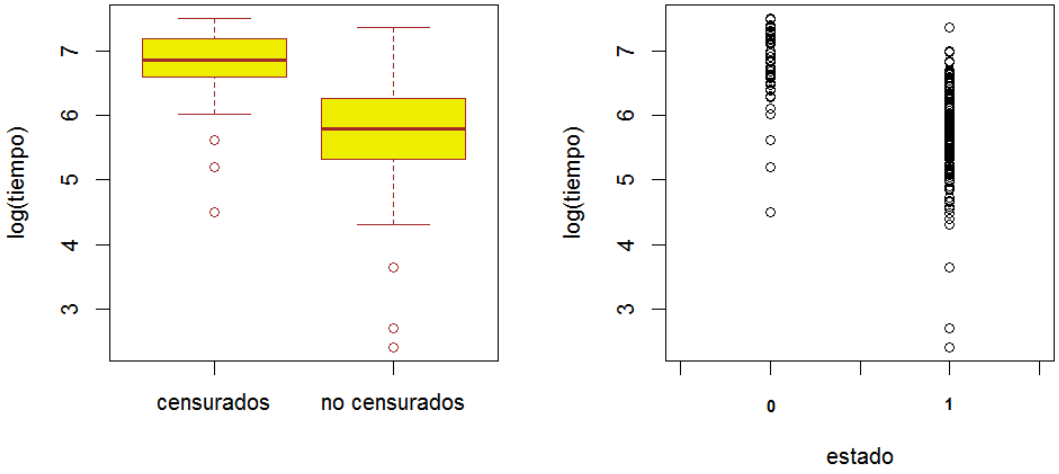

Figura 13: Datos censurados y no censurados por separado para los logaritmos de los tiempos de supervivencia

Observamos en los *boxplots* de la Figura 13 la presencia de cuatro datos atípicos pequeños (hay dos iguales) y por lo tanto distribuciones que parecerían asimétricas a izquierda. Descartando estos puntos, el resto de los datos parecerían tener una distri-

bución bastante simétrica. Entre los bigotes del *boxplot* y dentro de la caja se observa simetría.

Hay dos posibilidades: ajustar un modelo normal debido a la distribución simétrica de los datos, excluyendo los cuatro datos pequeños y considerarlos atípicos para dicha distribución o ajustar un modelo log-Weibull considerando esos cuatro datos pequeños, parte de una distribución asimétrica a derecha.

Reforcemos lo hecho hasta acá con el estudio de los *qqplots* de cada distribución. Buscaremos estimar la distribución verdadera y utilizarla para calcular los cuantiles teóricos correspondientes a los tiempos de supervivencia.

La presencia de datos censurados (es decir, al no poder observar todos los z_i) hace que la distribución empírica convencional de la muestra $z_1^*, \ldots, z_{195}^*$ no sea una distribución consistente de la distribución acumulada F_z de z_i. Esto nos conduce a tener que estimar la distribución verdadera de otra forma.

En realidad lo que se estima es la función de supervivencia S que se define en el tiempo t como la probabilidad de que una persona sobreviva (no le ocurra el evento de interés, que en este caso es el fallecimiento) al menos hasta el tiempo t. Más formalmente:

$$S(t) = P(z > t) = 1 - F_z(t).$$

Existen varias formas de estimar la función de sobrevivencia, pero quizás la más práctica sea el estimador de Kaplan y Meier (1958), porque los mismos tiempos de observación $z_i^* = (\log(y_i))^*$ van contribuyendo a la estimación de la función de supervivencia. Puede probarse que se trata del estimador no paramétrico de máxima verosimilitud de S.

Para el caso en que los datos presenten censura a la derecha, el estimador de Kaplan y Meier se define de la siguiente forma: sean $t_1 < \ldots < t_N$ los datos observados (tiempo de supervivencia) no repetidos y no censurados. Luego

$$\widehat{S}_{KM}(t) = \prod_{t_i \leq t} \frac{r(t_i) - d(t_i)}{r(t_i)},$$

donde $d(t_i)$ es el número de individuos muertos en el instante t_i mientras que $r(t_i)$ es el número de sujetos en riesto justo antes de t_i: sólo se observan los sujetos vivos que no se han "perdido" en el estudio. De esta forma cada dato t_i queda asociado a un valor p_i tal que

$$\widehat{S}_{KM}(t_i) = p_i \text{ con } p_i \geq p_{i+1}.$$

Considerando para cada t_i los pesos acumulados $q_i = 1 - p_i$ obtenemos una distribución empírica aproximada para la distribución acumulada F_z de z_i. Si no hay censura el estimador de Kaplan y Meier coincide con la distribución empírica usual. Por otro lado, notar que los datos censurados reciben el mismo peso que el no censurado anterior. Por lo tanto, para graficar la distribución empírica por Kaplan y Meier (Figura 14) utilizaremos solamente los datos no censurados, ya que el gráfico tiene la misma forma al que daría si consideráramos todos los datos.

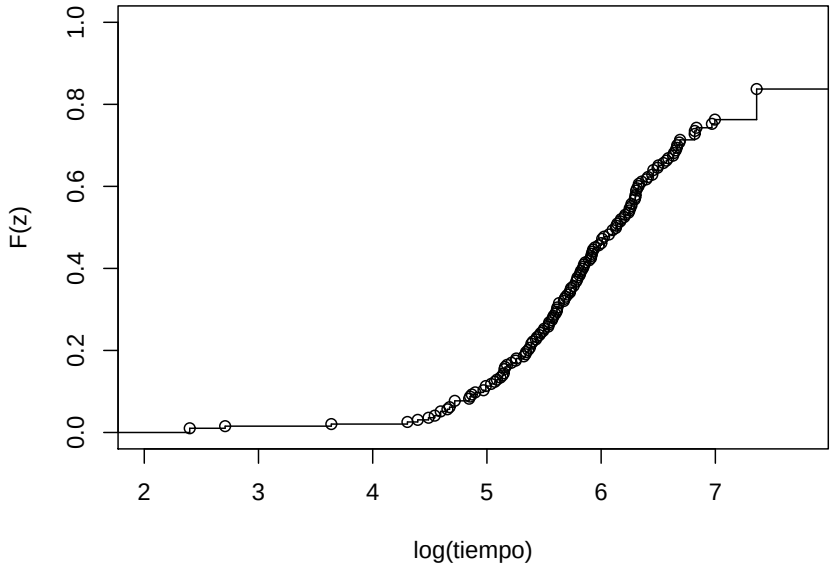

Figura 14: Estimación de la función de distribución acumulada por Kaplan y Meier

En lo que sigue trataremos de identificar el modelo paramétrico más adecuado para las variables z_i. Compararemos los cuantiles empíricos con sus respectivos cuantiles teóricos suponiendo una distribución específica F_z, es decir plotearemos:

$$t_i \text{ versus } F_z^{-1}(q_i),$$

siendo q_i los pesos acumulados estimados por Kaplan y Meier (1958). Por tratarse de familias de distribuciones de posición y escala, no perdemos generalidad si utilizamos a F_z en lugar del miembro estándar de la familia.

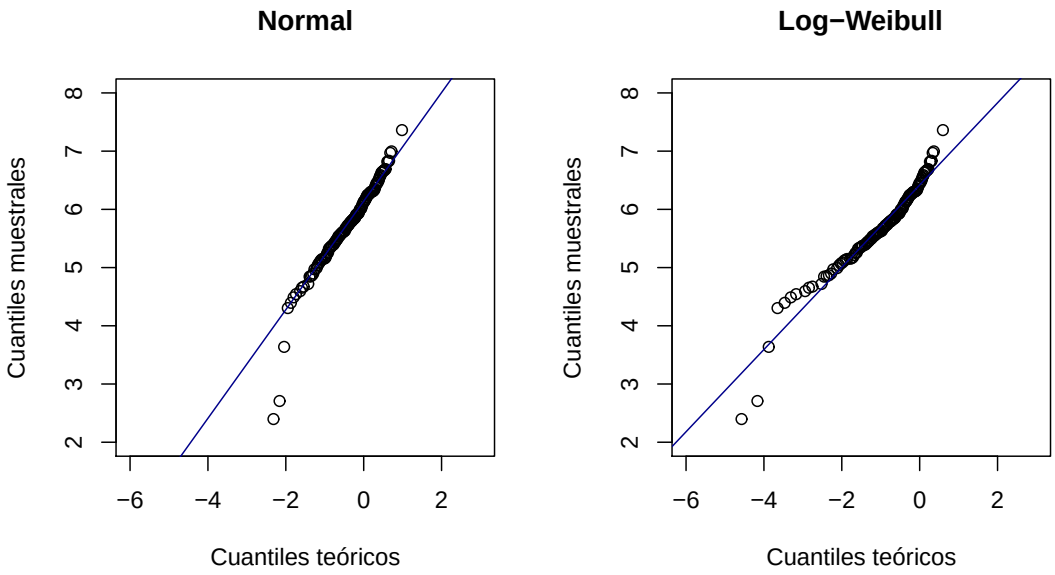

Figura 15: Qqplots construidos con los datos no censurados y cuantiles estimados por Kaplan y Meier

A partir de lo observado en la Figura 15 vamos a optar por la distribución normal y al final la contrastaremos con la log-Weibull. Supondremos entonces que los errores u_i y por lo tanto las variables z_i tienen distribución normal. En función de esto haremos las estimaciones correspondientes de los parámetros.

Parámetro	*Estimador*	
posición	S_1	5.966
	S_2	5.967
	MVT_1	6.231
	MVT_2	6.231
	MV	6.206
escala	S_1	1.015
	S_2	1.007
	MVT_1	0.989
	MVT_2	0.987
	MV	1.132

Cuadro 7: Estimaciones con todos los datos y ajuste normal

Estudiamos los residuos estandarizados en valor absoluto $|r_i|$ y utilizamos para ello el estimador MVT_2. Un valor grande de $|r_i|$ sugeriría que y_i es un *outlier*. Asumiendo un

modelo normal parecería razonable considerar como *outlier* aquellos datos con $|r_i| > 2{,}5$, aunque algunos autores sugieren usar 3. Se observa que los datos con números de orden 89, 99, 115 (estos dos últimos tienen el mismo valor) y 188 tienen residuos en valor absoluto mayores a 2.5. De hecho, corresponden en los cuatro casos, a observaciones no censuradas y coinciden con los valores que se encuentran alejados de la recta en el gráfico cuantil-cuantil (*qqplot*). Si los excluimos del análisis y calculamos nuevamente las estimaciones obtenemos:

Parámetro	*Estimador*	
posición	S_1	5.965
	S_2	5.967
	MVT_1	6.249
	MVT_2	6.249
	MV	6.248
escala	S_1	0.983
	S_2	0.976
	MVT_1	0.981
	MVT_2	0.981
	MV	0.977

Cuadro 8: Estimaciones sin *outliers* y ajuste normal

Si comparamos los Cuadros 7 y 8, el estimador que resultó más perturbado ante la presencia de *outliers* (como era de esperar aunque muy levemente) es el de máxima verosimilitud (para la escala), mientras que los otros estimadores robustos se mantuvieron aproximadamente en los mismos valores luego de eliminar *outliers*. Al excluir estos datos el estimador de máxima verosimilitud MV se acercó a las estimaciones finales de MVT_1 y MVT_2 tal cual esperábamos.

Los estimadores finales y el estimador de máxima verosimilitud dieron resultados muy parecidos tanto en posición como en escala. Los S-estimadores iniciales, que corresponden a estimadores robustos, dieron entre ellos casi lo mismos resultados con o sin presencia de *outliers*. Por otro lado, los S-estimadores de posición, quedaron relativamente lejos de las estimaciones finales, debido quizás a su baja eficiencia.

Nuevamente considerando todos los datos originales, observamos que los *outliers* no lograron perturbar demasiado al estimador de máxima verosimilitud MV. Por tal motivo y para seguir estudiando el comportamiento de los estimadores, sustituiremos la observación más grande (de valor 7.5 aproximadamente) por un *outliers* severo (de valor 20). Si bien se trata de una contaminación artificial, podremos comparar con más detalle todos los estimadores. Estimamamos con este conjunto de datos.

Parámetro	*Estimador*	
posición	S_1	5.966
	S_2	5.967
	MVT_1	6.196
	MVT_2	6.196
	MV	7.043
escala	S_1	1.014
	S_2	1.007
	MVT_1	0.952
	MVT_2	0.952
	MV	3.383

Cuadro 9: Estimaciones con *outlier* artificial y ajuste normal

Observamos en el Cuadro 9, que se hace evidente la influencia del *outlier*, en el estimador de máxima verosimilitud. Los S-estimadores iniciales y finales no se muestran perturbados con respecto a lo obtenido con los datos originales. La diferencia entre los iniciales y finales puede explicarse por la falta de eficiencia de los S-estimadores, como sucedía también con los datos originales.

Retomamos nuevamente el estudio con los datos originales y completos. El siguiente gráfico (Figura 16) muestra la distribución acumulada estimada por el método de Kaplan y Meier, al cual se le superpuso las funciones de distribución usando como parámetros, las estimaciones obtenidas con el estimador final que notaremos MVT (MVT_1 o MVT_2 porque cualquiera de ellos da los mismo), el S-estimador y el de máxima verosimilitud MV (todos con ajustes normales). Observar que en el primer caso, el ajuste es mejor que en los otros dos, como era de esperar.

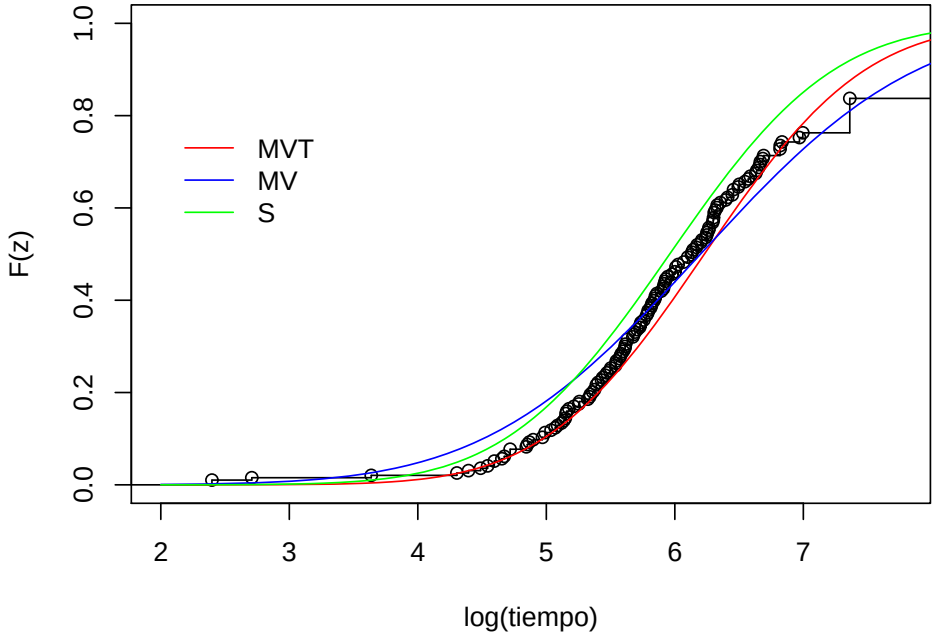

Figura 16: Kaplan y Meier y estimaciones de la distribución con todos los datos bajo modelo gaussiano

Veamos finalmente qué ocurre si suponemos un modelo log-Weibull, y comparémoslo con el normal.

Parámetro	*Estimador*	
posición	S_1	6.168
	S_2	6.169
	MVT_1	6.639
	MVT_2	6.639
	MV	6.640
escala	S_1	0.884
	S_2	0.877
	MVT_1	0.924
	MVT_2	0.923
	MV	0.920

Cuadro 10: Estimaciones con todos los datos y ajuste log-Weibull

Notemos que los estimadores S_1 y S_2 dieron estimaciones casi iguales entre ellos. Lo

mismo ocurre con los estimadores finales y el de máxima verosimilitud.

Notar que en el ajuste normal se observan mayores diferencias entre el estimador de máxima verosimilitud MV y los estimadores finales MVT de lo que ocurre con el ajuste log-Weibull, para el parámetro de posición. Esto se debe quizás a que los datos menores que son considerados *outliers* en el caso normal y son rechazados por el estimador MVT, no son tales en el caso log-Weibull debido a la forma asimétrica de la distribución.

Por otro lado, en el ajuste normal hay menos diferencias entre los S-estimadores y el estimador de máxima verosimilitud MV de lo que ocurre en el caso log-Weibull. Los S-estimadores usan ρ pares, es decir penalizan de la misma forma residuos grandes positivos o negativos con tal de que tengan el mismo valor absoluto. Quizás, para distribuciones asimétricas, esto puede ser un inconveniente porque se podrían perder datos que no son potenciales "outliers" para la distribución.

El siguiente gráfico muestra la distribución acumulada estimada por el método de Kaplan y Meier, a la cual se le superpuso las funciones de distribución usando como parámetros las estimaciones obtenidas con el estimador final (todas con ajuste log-Weibull) que notaremos MVT (MVT_1 o MVT_2), el de máxima verosimilitud MV y el S-estimador.

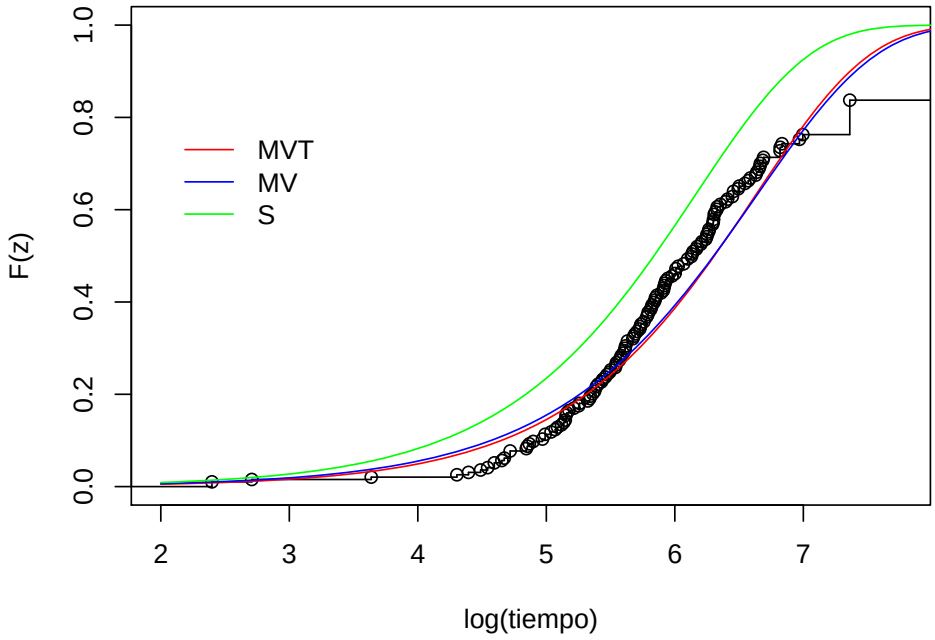

Figura 17: Kaplan y Meier y estimaciones de la distribución con todos los datos bajo el modelo log-Weibull

Observar que el S-estimador da un peor ajuste (debido quizás a su baja eficiencia) comparado con los otros dos, que son prácticamente iguales.

Graficamos a continuación la distribución acumulada estimada por el método de Kaplan Meier, a la cual se le superpuso las funciones de distribución usando como parámetros, las estimaciones obtenidas con el estimador final MVT bajo el modelo normal y bajo el modelo log-Weibull. El modelo normal parece ajustar mejor.

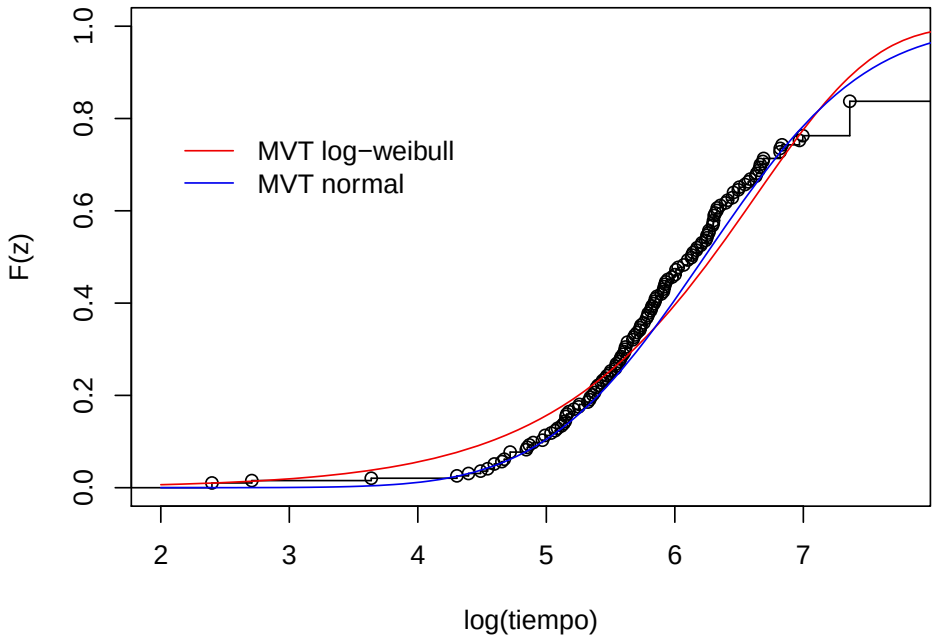

Figura 18: Kaplan y Meier y estimaciones con ambos ajustes: normal y log-Weibull

11. Apéndice de la Sección 2

A continuación damos una serie de ejemplos relativos a "tiempo de supervivencia" y datos censurados. Los mismos ilustran algunos de los conjuntos de datos que son de interés analizar. El siguiente es un ejemplo dado por Lawless (2003) para comprender mejor cómo actúa la censura a derecha.

Ejemplo 11.1. Nelson (1990) describió los resultados de un experimento en el cual muestras de un tipo de fluido de aislamiento eléctrico fueron sometidas a una tensión constante. Se observó el tiempo transcurrido desde que se lo sometió a dicha tensión hasta que el fluido falló o "se rompió". La tabla expuesta abajo (Cuadro 11) muestra los

resultados de siete grupos de muestras que se sometieron a distintos voltajes entre 26 y 38 kilovoltios.

voltage (kv)	n_i	*tiempo transcurrido hasta la falla*
26	3	5.79, 1579.52, 2323.7
28	5	68.85, 426.07, 110.29, 108.29, 1067.6
30	11	17.05, 22.66, 21.02, 175.88, 139.07, 144.12,20.46, 43.40, 194.90, 47.30, 7.74
32	15	0.40, 82.85, 9.88, 89.29, 215.10, 2.75, 0.79, 15.93, 3.91, 0.27, 0.69, 100.58, 27.80, 13.95, 53.24
34	19	0.96, 4,15, 0.19, 0,78, 8.01, 31.75, 7.35, 6.50, 8.27, 33.91, 32.52, 3.16, 4.85, 2.78, 4.67, 1.31, 12.06, 36.71, 72.89
36	15	1.97, 0.59, 2.58, 1.69, 2.71, 25.50, 0.35, 0.99, 3.99, 3.67, 2.07, 0.96, 5.35, 2,90, 13.77
38	8	0.47, 0.73, 1.40, 0.74, 0.39, 1.13, 0.09, 2.38

Cuadro 11: Tiempo de ruptura o falla en minutos para los siete niveles de voltage

Este experimento fue suficientemente largo como para poder observar la falla en todos los casos testeados. Sin embargo, no siempre se puede esperar tanto tiempo y es necesario terminar el estudio antes de que todos fallen dando origen a una observación censurada a derecha. Si, por ejemplo, sabemos de antemano que sólo se puede observar hasta los 180 minutos en los voltajes 26, 28, 30 y hasta los 200 minutos en el resto, tendríamos seis observaciones censuradas a derecha y sólo se sabría de ellas que las cinco primeras son mayores a 180 y la última mayor a 200. En este caso la finalización del estudio estaría planificada con anterioridad al inicio del mismo y la censura no sería aleatoria.

Ejemplo 11.2. Freireich y otros (1963) reportaron los resultados de una prueba clínica, en el cual la droga 6-mercaptopurine (6-MP) fue comparada con un placebo con respecto a la habilidad para mantener la remisión en niños con leucemia aguda. La prueba fue llevada a cabo en 11 hospitales norteamericanos, con 42 niños seleccionados de tal manera que al momento de iniciar la prueba tenían una remisión parcial o total de su leucemia, en virtud de haberse tratado con la droga Prednisone. La remisión parcial o total significa que la mayoría o todos los signos de la enfermedad habían desaparecido de la médula osea. La prueba fue diseñada de a pares de pacientes de un mismo hospital y con el mismo status de enfermedad (parcial o total). Dentro de cada par se eligió al azar quién iba a recibir placebo y quién 6-MP. Los pacientes fueron observados desde que se les administró la droga o el placebo, hasta que la enfermedad reapareció o hasta que por algún motivo inesperado se terminó el estudio. El tiempo se midió en meses. Los datos están resumidos en la siguiente tabla:

par	*tipo de remisión*	*pacientes con placebo*	*pacientes con 6-MP*
1	parcial	1	10
2	completa	22	7
3	completa	3	32*
4	completa	12	23
5	completa	8	22
6	parcial	17	6
7	completa	2	16
8	completa	11	34*
92	completa	8	32*
10	completa	12	25*
11	completa	2	11*
12	parcial	5	20*
13	completa	4	19*
14	completa	15	6
15	completa	8	17*
16	parcial	23	35*
17	parcial	5	6
18	completa	11	13
19	completa	4	9*
20	completa	1	6*
21	completa	8	10*

*Observaciones censuradas

Cuadro 12: Duración de la remisión con 6-MP versus pacebo en niños con leucemia aguda

Las observaciones con estrella son datos censurados pues estos pacientes aún seguían en estado de remisión cuando se los dejó de observar. La censura es común en pruebas clínicas debido a que éstas a menudo finalizan antes de que los individuos "hayan fallado".

Los Ejemplos 11.1 y 11.2 muestran dos tipos diferentes de censura. En el primero se censura con valores fijos especificados o planificados con anterioridad al experimento. Se podría dar cuando, por ejemplo, se conoce de antemano los costos del procedimiento del tal forma que imposibilite al investigador prolongar el experimento más allá de un determinado tiempo. En el Ejemplo 11.2 la censura es aleatoria a derecha y se produce cuando en el transcurso del estudio, algunos individuos experimentan otros sucesos independientes al de interés que provocan la salida del estudio, es decir el estudio es finalizado de una manera no planificada en general. Es el modelo más realista en este tipo de pruebas debido a que podría ocurrir que el paciente por algún motivo inesperado abandone el estudio, o se desenlace otro suceso que imposibilite la ocurrencia del evento que se desea observar.

En el Ejemplo 11.1, y_i representa el tiempo de duración hasta la falla, mientras que el tiempo de censura es $v_i = 180$ minutos ó 200 minutos según el caso. En el Ejemplo 11.2, y_i es el tiempo que dura la remisión del tumor desde el suministro de la droga o placebo según el caso. En todos los casos el dato censurado y_i^* es la duración real del

estudio que pudo finalizar por motivos inesperados. Lo que se observa, es aquello que ocurre primero.

Ejemplo 11.3. Klein y Moeschberger (1997) analizaron un estudio en adolescentes que empiezan a fumar marihuana, como ejemplo de censura aleatoria a la izquierda. Se seleccionó una muestra de adolescentes con 16 años que ya fumaban marihuana y se les preguntó a qué edad comenzaron. La variable de interés es la edad que tenían cuando fumaron por primera vez. Algunos lo recuerdan con exactitud (a los 14 años, a los 15 años, etc) por lo que proporcionan información completa sobre la variable en estudio. Sin embargo, hay otros jóvenes que no recuerdan en qué momento fumaron por primera vez, pero dado que ahora fuman y tienen 16 años, la información que aportan es un dato censurado a la izquierda, ya que se sabe que la variable de interés es menor o igual a 16.

12. Apéndice de la Sección 3

Modelo y notaciones

Demostración del Lema 3.1

Sean $g : \mathbb{R} \to \mathbb{R}$ una función medible y u_0 es un valor real fijo. Veamos primero que

$$E(g(u)|u > u_0) = \frac{\int_{u_0}^{+\infty} g(u) f_0(u) du}{1 - F_0(u_0)}. \tag{72}$$

Consideremos la variable aleatoria $z(u) = I(u > u_o)$ y calculemos la función de distribución acumulada de la variable aleatoria u condicionada al evento $z(u) = 1$,

$$F_{u|z(u)=1}(t) = \frac{P(u \leq t,\ z(u) = 1)}{P(z = 1)} = \frac{P(u_o < u \leq t)}{P(u > u_0)} = \frac{F_0(t) - F_0(u_0)}{1 - F_0(u_0)} I(t \geq u_0).$$

Luego derivando, obtenemos la función de densidad condicional

$$f_{u|z(u)=1}(t) = \frac{f_0(t)}{1 - F_0(u_0)} I(t \geq u_0).$$

Finalmente

$$E_0(g(u)|z(u) = 1) = \int_{-\infty}^{+\infty} g(u) f_{u|z(u)=1}(u) du = \frac{\int_{u_0}^{+\infty} g(u) f_0(u) du}{1 - F_0(u_0)}.$$

Recordemos que $y^* = \text{mín}\{y, v\}$, siendo v el valor de censura. En particular, $\delta = 0$ es equivalente a $y^* = v$ ó $y > v$. Para cada $\tilde{\mathbf{x}}$ y $\tilde{v}$ fijos de la misma dimensión que $\mathbf{x}$ y v respectivamente, calcularemos

$$E_{\boldsymbol{\beta},\sigma}(h(\mathbf{x}, y)|\mathbf{x} = \tilde{\mathbf{x}}, y^* = \tilde{v}, y > y^*) = E_{\boldsymbol{\beta},\sigma}(h(\mathbf{x}, y)|\mathbf{x} = \tilde{\mathbf{x}}, v = \tilde{v}, \delta = 0).$$

Observemos que, suponiendo verdadero $y = \mathbf{x}^t \boldsymbol{\beta} + \sigma u$ vale:

$$y > v \Leftrightarrow u > \frac{v - \mathbf{x}^t \boldsymbol{\beta}}{\sigma}.$$

Consideremos ahora la variable aleatoria

$$w(\mathbf{x}, v, u) = I\left(u > \frac{v - \mathbf{x}^t\boldsymbol{\beta}}{\sigma}\right).$$

Notar que $w(\mathbf{x}, v, u) = 1$ es equivalente a $\delta = 0$. Recordemos el siguiente resultado de esperanza condicional:

$$E(h(\mathbf{x}, u))|\mathbf{x} = \tilde{\mathbf{x}}, v = \tilde{v}, w(\mathbf{x}, v, u) = 1) = E(h(\tilde{\mathbf{x}}, u))|\mathbf{x} = \tilde{\mathbf{x}}, v = \tilde{v}, w(\tilde{\mathbf{x}}, \tilde{v}, u) = 1)$$

y como u es independiente del par $(\mathbf{x}, v)$ entonces

$$E(h(\mathbf{x}, u))|\mathbf{x} = \tilde{\mathbf{x}}, v = \tilde{v}, w(\mathbf{x}, v, u) = 1) = E(h(\tilde{\mathbf{x}}, u))|w(\tilde{\mathbf{x}}, \tilde{v}, u) = 1). \tag{73}$$

Si tomamos

$$u_0 = \frac{\tilde{v} - \tilde{\mathbf{x}}^t\boldsymbol{\beta}}{\sigma} \quad \text{y} \quad \tilde{g}(u) = h(\tilde{\mathbf{x}}, \tilde{\mathbf{x}}^t\boldsymbol{\beta} + \sigma u)$$

y haciendo uso de (72) y (73) obtenemos bajo el modelo

$$\begin{aligned}
E_{\boldsymbol{\beta},\sigma}(h(\mathbf{x}, y)|\mathbf{x} = \tilde{\mathbf{x}}, y^* = \tilde{v}, y > y^*) &= E(h(\mathbf{x}, \mathbf{x}^t\boldsymbol{\beta} + \sigma u)|\mathbf{x} = \tilde{\mathbf{x}}, v = \tilde{v}, w(\mathbf{x}, v, u) = 1) \\
&= E(\tilde{g}(u)|w(\tilde{\mathbf{x}}, \tilde{v}, u) = 1) \\
&= E(\tilde{g}(u)|u > u_0) \\
&= \frac{\displaystyle\int_{u_0}^{+\infty} \tilde{g}(u) f_0(u) du}{1 - F_0(u_0)} \\
&= \frac{\displaystyle\int_{\frac{\tilde{v} - \tilde{\mathbf{x}}^t\boldsymbol{\beta}}{\sigma}}^{+\infty} h(\tilde{\mathbf{x}}, \tilde{\mathbf{x}}^t\boldsymbol{\beta} + \sigma u) f_0(u) du}{1 - F_0\left(\frac{\tilde{v} - \tilde{\mathbf{x}}^t\boldsymbol{\beta}}{\sigma}\right)}.
\end{aligned}$$

□

Demostración del Lema 3.2

Por definición

$$E_{n,\boldsymbol{\beta},\sigma}(h(\mathbf{x}, y)) = \int\int h(\mathbf{x}, y) dH_{n,\boldsymbol{\beta},\sigma}(\mathbf{x}, y),$$

y en virtud de la ecuación (9) se obtiene $E_{n,\boldsymbol{\beta},\sigma}(h(\mathbf{x}, y)) =$

$$= \frac{1}{n}\sum_{i=1}^{n}\left(\delta_i h(\mathbf{x}_i, y_i^*) + (1 - \delta_i)\int h(\mathbf{x}, y) dE_{\boldsymbol{\beta},\sigma}(I(\mathbf{x}_i \leq \mathbf{x})I(y_i \leq y)|\mathbf{x}_i, y_i^*, y_i > y_i^*)\right).$$

Por otro lado

$$\begin{aligned}
E_{\boldsymbol{\beta},\sigma}(I(\mathbf{x}_i \leq \mathbf{x})I(y_i \leq y)|\mathbf{x}_i, y_i^*, y_i > y_i^*) &= E_{\boldsymbol{\beta},\sigma}(I(\mathbf{x}_i \leq \mathbf{x}, y_i \leq y)|\mathbf{x}_i, y_i > y_i^*) \\
&= P_{\boldsymbol{\beta},\sigma}(\mathbf{x}_i \leq \mathbf{x}, y_i \leq y|\mathbf{x}_i, y_i^*, y_i > y_i^*).
\end{aligned}$$

Luego $E_{n,\boldsymbol{\beta},\sigma}(h(\mathbf{x},y)) =$

$$
\begin{aligned}
&= \frac{1}{n}\sum_{i=1}^{n}\left(\delta_i h(\mathbf{x}_i,y_i^*) + (1-\delta_i)\int h(\mathbf{x},y)dP_{\boldsymbol{\beta},\sigma}(\mathbf{x}_i\le\mathbf{x}, y_i\le y|\mathbf{x}_i,y_i^*,y_i>y_i^*)\right)\\
&= \frac{1}{n}\sum_{i=1}^{n}\left(\delta_i h(\mathbf{x}_i,y_i^*) + (1-\delta_i)E_{\boldsymbol{\beta},\sigma}(h(\mathbf{x}_i,y_i)|\mathbf{x}_i,y_i^*,y_i>y_i^*)\right) \qquad (74)\\
&= \frac{1}{n}\sum_{i=1}^{n}E_{\boldsymbol{\beta},\sigma}(h(\mathbf{x}_i,y_i)|\mathbf{x}_i,y_i^*,\delta_i).
\end{aligned}
$$

□

13. Apéndice de la Sección 4

Estimadores de máxima verosimilitud para datos censurados

Demostración del Lema 4.1

Calculemos las derivadas parciales del logaritmo de la verosimilitud cuya expresión para cada $\boldsymbol{\beta}$ y σ está dada por

$$l(\boldsymbol{\beta},\sigma) = -r\ln(\sigma) - \sum_{i=1}^{n}\left(\delta_i\rho_0\left(\frac{y_i^*-\mathbf{x}_i^t\boldsymbol{\beta}}{\sigma}\right) + (1-\delta_i)\ln\left(1-F_0\left(\frac{y_i^*-\mathbf{x}_i^t\boldsymbol{\beta}}{\sigma}\right)\right)\right).$$

Comencemos con las derivadas parciales con respecto a $\boldsymbol{\beta}_j$, para $j:1,\ldots,p$.

$$\frac{\partial l(\boldsymbol{\beta},\sigma)}{\partial\beta_j} = \frac{1}{\sigma}\sum_{i=1}^{n}\left(\delta_i\psi_0\left(\frac{y_i^*-\mathbf{x}_i^t\boldsymbol{\beta}}{\sigma}\right)x_{ij} + (1-\delta_i)\frac{f_0\left(\frac{y_i^*-\mathbf{x}_i^t\boldsymbol{\beta}}{\sigma}\right)x_{ij}}{1-F_0\left(\frac{y_i^*-\mathbf{x}_i^t\boldsymbol{\beta}}{\sigma}\right)}\right).$$

Si desarrollamos solamente los términos censurados y usamos (7) y (8) obtenemos:

$$
\begin{aligned}
\frac{f_0\left(\frac{y_i^*-\mathbf{x}_i^t\boldsymbol{\beta}}{\sigma}\right)}{1-F_0\left(\frac{y_i^*-\mathbf{x}_i^t\boldsymbol{\beta}}{\sigma}\right)} &= \frac{\int_{\frac{y_i^*-\mathbf{x}_i^t\boldsymbol{\beta}}{\sigma}}^{+\infty} -f_0'(u_i)du_i}{1-F_0\left(\frac{y_i^*-\mathbf{x}_i^t\boldsymbol{\beta}}{\sigma}\right)}\\
&= \frac{\int_{\frac{y_i^*-\mathbf{x}_i^t\boldsymbol{\beta}}{\sigma}}^{+\infty} -\frac{f_0'(u_i)}{f_0(u_i)}f_0(u_i)du_i}{1-F_0\left(\frac{y_i^*-\mathbf{x}_i^t\boldsymbol{\beta}}{\sigma}\right)}\\
&= E_{\boldsymbol{\beta},\sigma}\left(\psi_0\left(\frac{y_i-\mathbf{x}_i^t\boldsymbol{\beta}}{\sigma}\right)\middle|\,\mathbf{x}_i, y_i>y_i^*\right).
\end{aligned}
$$

Luego

$$
\begin{aligned}
\frac{\partial l(\boldsymbol{\beta},\sigma)}{\partial\beta_j} &= \frac{1}{\sigma}\sum_{i=1}^{n}E_{\boldsymbol{\beta},\sigma}\left(\psi_0\left(\frac{y_i-\mathbf{x}_i^t\boldsymbol{\beta}}{\sigma}\right)x_{ij}\middle|\,\mathbf{x}_i,y_i^*,\delta_i\right)\\
&= \frac{n}{\sigma}E_{n,\boldsymbol{\beta},\sigma}\left(\psi_0\left(\frac{y-\mathbf{x}^t\boldsymbol{\beta}}{\sigma}\right)x_j\right).
\end{aligned}
$$

Para la derivada parcial con respecto a σ se tiene

$$\frac{\partial l(\boldsymbol{\beta},\sigma)}{\partial \sigma} =$$

$$\begin{aligned}
&= -\frac{r}{\sigma} + \sum_{i=1}^{n}\left(\delta_i \psi_0\left(\frac{y_i^* - \boldsymbol{x}_i^t\boldsymbol{\beta}}{\sigma}\right)\frac{y_i^* - \boldsymbol{x}_i^t\boldsymbol{\beta}}{\sigma^2} + (1-\delta_i)\frac{f_0\left(\frac{y_i^*-\boldsymbol{x}_i^t\boldsymbol{\beta}}{\sigma}\right)\frac{y_i^*-\boldsymbol{x}_i^t\boldsymbol{\beta}}{\sigma^2}}{1-F_0\left(\frac{y_i^*-\boldsymbol{x}_i^t\boldsymbol{\beta}}{\sigma}\right)}\right)\\
&= -\frac{r}{\sigma} + \sum_{i=1}^{n}\left(\frac{\delta_i}{\sigma}\psi_1\left(\frac{y_i^* - \boldsymbol{x}_i^t\boldsymbol{\beta}}{\sigma}\right) + \frac{(1-\delta_i)}{\sigma}\left(-\frac{\int_{\frac{y_i^*-\boldsymbol{x}_i^t\boldsymbol{\beta}}{\sigma}}^{+\infty} u_i f_0'(u_i)du_i}{1-F_0\left(\frac{y_i^*-\boldsymbol{x}_i^t\boldsymbol{\beta}}{\sigma}\right)} - 1\right)\right).
\end{aligned}$$

Por otro lado, si analizamos los términos censurados obtenemos:

$$\begin{aligned}
-\frac{\int_{\frac{y_i^*-\mathbf{x}_i^t\boldsymbol{\beta}}{\sigma}}^{+\infty} u_i f_0'(u_i)du_i}{1-F_0\left(\frac{y_i^*-\mathbf{x}_i^t\boldsymbol{\beta}}{\sigma}\right)} &= \frac{\int_{\frac{y_i^*-\mathbf{x}_i^t\boldsymbol{\beta}}{\sigma}}^{+\infty} -u_i\frac{f_0'(u_i)}{f_0(u_i)}f_0(u_i)du_i}{1-F_0\left(\frac{y_i^*-\mathbf{x}_i^t\boldsymbol{\beta}}{\sigma}\right)}\\
&= E_{\boldsymbol{\beta},\sigma}\left(\psi_1\left(\frac{y_i-\mathbf{x}_i^t\boldsymbol{\beta}}{\sigma}\right)\middle|\, \mathbf{x}_i, y_i > y_i^*\right).
\end{aligned}$$

Luego $\dfrac{\partial l(\boldsymbol{\beta},\sigma)}{\partial \sigma} =$

$$\begin{aligned}
&= -\frac{r}{\sigma} + \frac{1}{\sigma}\sum_{i=1}^{n}\left(\delta_i\psi_1\left(\frac{y_i^*-\mathbf{x}_i^t\boldsymbol{\beta}}{\sigma}\right) + (1-\delta_i)\left(E_{\beta,\sigma}\left(\psi_1\left(\frac{y_i-\mathbf{x}_i^t\boldsymbol{\beta}}{\sigma}\right)\middle|\, \mathbf{x}_i, y_i > y_i^*\right) - 1\right)\right)\\
&= \frac{n}{\sigma}\left(-1 + \frac{1}{n}\sum_{i=1}^{n}\left(\delta_i\psi_1\left(\frac{y_i^*-\mathbf{x}_i^t\boldsymbol{\beta}}{\sigma}\right) + (1-\delta_i)E_{\beta,\sigma}\left(\psi_1\left(\frac{y_i-\mathbf{x}_i^t\boldsymbol{\beta}}{\sigma}\right)\middle|\, \mathbf{x}_i, y_i > y_i^*\right)\right)\right)\\
&= \frac{n}{\sigma}\left(-1 + \frac{1}{n}\sum_{i=1}^{n}E_{\beta,\sigma}\left(\psi_1\left(\frac{y_i-\mathbf{x}_i^t\boldsymbol{\beta}}{\sigma}\right)\middle|\, \mathbf{x}_i, y_i^*, \delta_i\right)\right)\\
&= \frac{n}{\sigma}\left(-1 + E_{n,\boldsymbol{\beta},\sigma}\left(\psi_1\left(\frac{y-\mathbf{x}^t\boldsymbol{\beta}}{\sigma}\right)\right)\right).
\end{aligned}$$

De existir los estimadores de máxima verosimilitud deben anular las derivadas parciales y por lo tanto verifican las ecuaciones (17) y (18). □

14. Apéndice de la Sección 5

S estimadores de regresión con datos no censurados

Demostración del Lema 5.1

Fijemos $\mu \in \mathbb{R}$ y consideremos $h_\mu : \mathbb{R}^+ \to \mathbb{R}$ definida como

$$h_\mu(s) = E_0\left(\rho\left(\frac{u-\mu}{s}\right)\right) = E_0\left(\rho\left(\frac{u-\mu}{s}\right) I(|u-\mu|>0)\right).$$

Luego haciendo uso del Teorema de Convergencia Mayorada y suponiendo que ρ tiene las propiedades **C1**, **C2** y **C4** resulta

$$\lim_{s\to 0^+} h_\mu(s) = \lim_{s\to 0} \int_{-\infty}^{+\infty} \rho\left(\frac{u-\mu}{s}\right) I(|u-\mu|>0) f_0(u)du = P(|u-\mu|>0) = 1, \tag{75}$$

$$\lim_{s\to +\infty} h_\mu(s) = \lim_{s\to +\infty} \int_{-\infty}^{+\infty} \rho\left(\frac{u-\mu}{s}\right) f_0(u)du = 0. \tag{76}$$

Como $h_\mu(s)$ es continua y $b \in (0,1)$ se deduce que la ecuación $h_\mu(s) = b$ tiene solución como queríamos mostrar.

Supongamos que la ecuación (19) tiene más de una solución y notemos al conjunto no vacío de soluciones positivas $S(\mu)$ como

$$A_\mu^b = \{s \in \mathbb{R}^+ : h_\mu(s) = b\}.$$

Veamos que A_μ^b es un intervalo cerrado. Observar que $h_\mu(s)$ es monótona no creciente porque $\rho(|z|)$ es monótona no decreciente debido a las propiedades **C3** y **C5** de ρ. Luego si s_1 y s_2 son elementos de A_μ^b tales que $s_1 < s_2$ entonces cualquier valor $s \in (s_1, s_2)$ pertenecerá a A_μ^b ya que $b = h_\mu(s_2) \le h_\mu(s) \le h_\mu(s_1) = b$. Luego el conjunto A_μ^b es un intervalo. Falta mostrar que A_μ^b es cerrado.

A_μ^b está acotado superiormente porque en caso contrario, de la ecuación (76) resultaría $b = 0$. Como también A_μ^b está acotado inferiormente, entonces dicho conjunto tiene ínfimo y supremo. Llamemos

$$s^* = \inf_{s>0} A_\mu^b \quad \text{y} \quad s^{**} = \sup_{s>0} A_\mu^b.$$

Por definición de ínfimo existe una sucesión $(s_n)_{n\in\mathbb{N}}$ de elementos de A_μ^b que converge a s^*. Observar que s^* no puede ser 0 pues si no, por la ecuación (75) resultaría $b = 1$. Como $h_\mu(s)$ es continua entonces

$$b = \lim_{n\to\infty} h_\mu(s_n) = h_\mu(s^*).$$

En definitiva s^* es una solución de la ecuación (19) y por lo tanto

$$s^* = \min_{s>0} A_\mu^b.$$

Análogamente se procede con el supremo. Luego $A_\mu^b = [s^*, s^{**}]$. □

Demostración del Lema 5.2

Supongamos por el absurdo que existe $\mu \in \mathbb{R}$ tal que la ecuación (19) no tiene solución única. Luego existen $s_1 < s_2$ tales que

$$E_0\left(\rho\left(\frac{u-\mu}{s_1}\right)\right) = E_0\left(\rho\left(\frac{u-\mu}{s_2}\right)\right) = b.$$

Veamos a continuación que existe $\tilde{u}$ tal que

$$\rho\left(\frac{\tilde{u}-\mu}{s_2}\right) < 1 \text{ y } f_0(\tilde{u}) \neq 0.$$

Supongamos por el absurdo que esto no es cierto. Consideremos los conjuntos

$$A = \left\{u \in \mathbb{R} : \rho\left(\frac{u-\mu}{s_2}\right) < 1\right\} \text{ y } B = \{u \in \mathbb{R} : f_0(u) \neq 0\}.$$

Luego $A \cap B = \emptyset$. Si $u \in A$ entonces $f_0(u) = 0$. De esta forma se tiene

$$b = E_0\left(\rho\left(\frac{u-\mu}{s_2}\right)\right) = \int_{A^c} \rho\left(\frac{u-\mu}{s_2}\right) f_0(u)du.$$

Si $A^c = \emptyset$ entonces $b = 0$ y no es posible. Si $A^c \neq \emptyset$ entonces $b = 1$ y tampoco es posible.

Sea $\tilde{u} \in A \cap B$. Como

$$\left|\frac{\tilde{u}-\mu}{s_2}\right| < \left|\frac{\tilde{u}-\mu}{s_1}\right|, \quad \rho\left(\frac{\tilde{u}-\mu}{s_2}\right) < 1 \text{ y } f_0(\tilde{u}) \neq 0$$

entonces por la condición **C3** y **C7** se tiene que

$$\rho\left(\frac{\tilde{u}-\mu}{s_2}\right) f_0(\tilde{u}) < \rho\left(\frac{\tilde{u}-\mu}{s_1}\right) f_0(\tilde{u}),$$

y por lo tanto

$$b = E_0\left(\rho\left(\frac{u-\mu}{s_2}\right)\right) < E_0\left(\rho\left(\frac{u-\mu}{s_1}\right)\right) = b$$

llegando así a una contradicción. Luego la solución de la ecuación (19) es única □

Demostración del Lema 5.3

Supongamos que existe μ_1 tal que

$$E_0\left(\rho\left(\frac{u-\mu_1}{s_0}\right)\right) < E_0\left(\rho\left(\frac{u-\mu_0}{s_0}\right)\right) = b.$$

Por el Lema 5.1 existe $S(\mu_1)$ tal que

$$E_0\left(\rho\left(\frac{u-\mu_1}{s_0}\right)\right) < E_0\left(\rho\left(\frac{u-\mu_1}{S(\mu_1)}\right)\right) = b.$$

Consideremos la función definida por

$$h_{\mu_1}(s) = E_0\left(\rho\left(\frac{u-\mu_1}{s}\right)\right).$$

Como h_{μ_1} es monótona no creciente por las condiciones **C3** y **C5** resulta que $S(\mu_1) < s_0$, llegando así a una contradicción ya que $s_0 \leq S(\mu)$ para todo μ.

Recíprocamente, queremos ver que $S(\mu_0) = s_0 \leq S(\mu)$ para todo μ. Supongamos que existe $\tilde{\mu}_0$ tal que $S(\tilde{\mu}_0) < s_0$. Luego

$$b = E_0\left(\rho\left(\frac{u-\mu_0}{s_0}\right)\right) \leq E_0\left(\rho\left(\frac{u-\tilde{\mu}_0}{s_0}\right)\right) < E_0\left(\rho\left(\frac{u-\tilde{\mu}_0}{S(\tilde{\mu}_0)}\right)\right) = b,$$

lo cual es un absurdo. □

El siguiente resultado que se debe a Ibragimov (1956) es necesario para probar el Lema 5.4.

Lema 14.1. *Si f_0 es una densidad estrictamente unimodal y φ es una densidad tal que $\ln \varphi$ es cóncava en su soporte, entonces la convolución*

$$\lambda(t) = \int_{-\infty}^{\infty} \varphi(u-t) f_0(u) du = E_0(\varphi(u-t))$$

es estrictamente unimodal.

Demostración del Lema 5.4

Sea k el parámetro de ajuste de la función ρ. Denotemos

$$K = \int_{-k}^{k} \rho(z) dz \leq 2k$$

y consideremos

$$\varphi(u) = \frac{1-\rho(u)}{2k-K}.$$

Notar que φ es una densidad y que se anula si $|u| > k$ por la condición **C6-A** de ρ. Por otro lado

$$\ln \varphi = \ln\left(\frac{1-\rho}{2k-K}\right)$$

es cóncava en su soporte por la condición **C6-B** de ρ. Luego por el Lema 14.1

$$\lambda(\mu) = E_0(\varphi(u-\mu))$$

resulta estictamente unimodal. Recordemos la función $q : \mathbb{R} \to \mathbb{R}$ definida por

$$q(\mu) = E_0\left(\rho\left(u-\mu\right)\right).$$

Finalmente

$$q(\mu) = 1 - E_0(1-\rho(u-\mu)) = 1 - E_0((2k-K)\varphi(u-\mu)) = 1 - (2k-K)\lambda(\mu).$$

Por el Lema 14.1, $\lambda(\mu)$ es estrictamente unimodal. Luego $q(\mu)$ tiene un único mínimo abosluto.

La segunda parte del resultado se puede ver en Yohai (1985). Nosotros lo mostraremos en un caso particular, supondremos que f_0 es estrictamente decreciente en módulo y simétrica alrededor de $t = 0$. Esta prueba se debe a Maronna, Martin y Yohai (2007).

Como ρ es par, su derivada ψ es impar. Usando la hipótesis de simetría de f_0 se deduce una expresión alternativa para la derivada de $q(\mu)$:

$$\begin{aligned} q'(\mu) &= -\int_{-\infty}^{+\infty} \psi(u-\mu) f_0(u) du \\ &= -\int_{-\infty}^{0} \psi(u) f_0(u+\mu) du - \int_{0}^{+\infty} \psi(u) f_0(u+\mu) du \\ &= \int_{+\infty}^{0} \psi(-u) f_0(-u+\mu) du - \int_{0}^{+\infty} \psi(u) f_0(u+\mu) du \\ &= \int_{0}^{+\infty} \psi(u) f_0(u-\mu) du - \int_{0}^{+\infty} \psi(u) f_0(u+\mu) du \\ &= \int_{0}^{+\infty} \psi(u)(f_0(u-\mu) - f_0(u+\mu)) du. \end{aligned}$$

Con cálculos similares se obtiene que $q(\mu)$ es par. Por eso es suficiente probar que $q'(\mu) > 0$ si $\mu > 0$. Se sigue de las propiedades de la función ρ que $\psi(u) \geq 0$ si $u \geq 0$ y $\psi(u) > 0$ si $u \in (0, \epsilon)$ para algún $\epsilon > 0$. Como u y μ son positivos entonces $|u-\mu| < |u+\mu|$ y de allí $f_0(u-\mu) > f_0(u+\mu)$, lo cual implica que la última integral de arriba es positiva. □

Demostración del Lema 5.5

Veamos que $s_0 > 0$. Por definición de s_0 existe una sucesión $(\mu_n)_{n\in\mathbb{N}}$ tal que $\lim_{n\to+\infty} S(\mu_n) = s_0$. Además $(\mu_n)_{n\in\mathbb{N}}$ resulta una sucesión acotada pues en caso contrario, si existiera una subsucesión $(\mu_{n_j})_{j\in\mathbb{N}}$ divergente se tendría por el Teorema de Convergencia Mayorada

$$b = \lim_{j\to+\infty} E_0\left(\rho\left(\frac{u-\mu_{n_j}}{S(\mu_{n_j})}\right)\right) = 1$$

y no es posible.

Supongamos que $s_0 = 0$. Luego existe $(\mu_n)_{n\in\mathbb{N}}$ tal que $\lim_{n\to+\infty} S(\mu_n) = 0$. Como $(\mu_n)_{n\in\mathbb{N}}$ está acotada, tiene una subsucesión convergente $\lim_{j\to+\infty} \mu_{n_j} = \mu$. Luego por el Teorema de Convergencia Mayorada

$$b = \lim_{j\to+\infty} E_0\left(\rho\left(\frac{u-\mu_{n_j}}{S(\mu_{n_j})}\right)\right) = \lim_{j\to+\infty} \int_{-\infty}^{+\infty} \rho\left(\frac{u-\mu_{n_j}}{S(\mu_{n_j})}\right) f_0(u) du = 1,$$

llegando nuevamente a un absurdo.

En lo que sigue probaremos que $S(\mu)$ es continua. Sea $(\mu_n)_{n\in\mathbb{N}}$ tal que $\lim_{n\to+\infty} \mu_n = \mu$ para algún μ. Queremos ver que $\lim_{n\to+\infty} S(\mu_n) = S(\mu)$. Supongamos que $(S(\mu_n))_{n\in\mathbb{N}}$ no es convergente. Si $(S(\mu_n))_{n\in\mathbb{N}}$ tuviera una subsucesión divergente $(S(\mu_{n_j}))_{j\in\mathbb{N}}$ entonces

$$b = \lim_{j\to+\infty} E_0\left(\rho\left(\frac{u-\mu_{n_j}}{S(\mu_{n_j})}\right)\right) = 0,$$

llegando así a un absurdo. Luego $(S(\mu_n))_{n\in\mathbb{N}}$ está acotada. Como estamos suponiendo que $(S(\mu_n))_{n\in\mathbb{N}}$ no es convergente entonces existen dos subsucesiones convergentes $(S(\mu_{n_{j_1}}))_{j_1\in\mathbb{N}}$ y $(S(\mu_{n_{j_2}}))_{j_2\in\mathbb{N}}$ tales que $\lim_{j_1\to+\infty} S(\mu_{n_{j_1}}) = s_1$ y $\lim_{j_2\to+\infty} S(\mu_{n_{j_2}}) = s_2$ con $s_1 \neq s_2$. Luego

$$\begin{aligned} b = \lim_{j_1\to+\infty} E_0\left(\rho\left(\frac{u-\mu_{n_{j_1}}}{S(\mu_{n_{j_1}})}\right)\right) &= E_0\left(\rho\left(\frac{u-\mu}{s_1}\right)\right) \\ b = \lim_{j_1\to+\infty} E_0\left(\rho\left(\frac{u-\mu_{n_{j_2}}}{S(\mu_{n_{j_2}})}\right)\right) &= E_0\left(\rho\left(\frac{u-\mu}{s_2}\right)\right). \end{aligned}$$

En virtud de las hipótesis la solución es única, luego $s_1 = s_2 = S(\mu)$ lo cual contradice el supuesto. Finalmente $\lim_{n\to+\infty} S(\mu_n) = S(\mu)$.

Por último debemos ver que $s_0 = \min_{\mu} S(\mu)$. Por definición de s_0 existe una sucesión $(\mu_n)_{n\in\mathbb{N}}$ tal que $\lim_{n\to+\infty} S(\mu_n) = s_0$. Como ya vimos $(\mu_n)_{n\in\mathbb{N}}$ está acotada, luego tiene una subsucesión convergente, es decir $\lim_{j\to+\infty} \mu_{n_j} = \mu$ para algún μ. Como S es continua $\lim_{j\to+\infty} S(\mu_{n_j}) = S(\mu) = s_0$. □

Demostración del Lema 5.6

Dado s_0, la función q_{s_0} tiene un único mínimo absoluto μ_0 por el Lema 5.4. Sea u_1 tal que $S(\mu_1) = s_0$ y supongamos $\mu_1 \neq \mu_0$. Luego

$$E_0\left(\rho\left(\frac{u-\mu_0}{s_0}\right)\right) < E_0\left(\rho\left(\frac{u-\mu_1}{s_0}\right)\right) = b = E_0\left(\rho\left(\frac{u-\mu_0}{S(\mu_0)}\right)\right),$$

con lo cual $S(\mu_0) < s_0$ y esto contradice la definición de s_0. Luego por el Lema 5.5 $S(\mu_0) = s_0$ y además μ_0 resulta único en este sentido. □

Demostración del Lema 5.7

En virtud de (31) obtenemos

$$\frac{1}{n}\sum_{i=1}^{n} \rho\left(\frac{y_i - \mathbf{x}_i^t\boldsymbol{\beta}}{s}\right) = \frac{1}{n}\sum_{i\notin A} \rho\left(\frac{y_i - \mathbf{x}_i^t\boldsymbol{\beta}}{s}\right) \leq \frac{1}{n}\#A^c < b.$$

□

Demostración del Lema 5.8

Por continuidad de ρ y a partir de (31) obtenemos:

$$\lim_{s\to+\infty} \frac{1}{n}\sum_{i=1}^{n} \rho\left(\frac{y_i - \mathbf{x}_i^t\boldsymbol{\beta}}{s}\right) = 0$$

$$\lim_{s\to 0^+} \frac{1}{n}\sum_{i=1}^{n} \rho\left(\frac{y_i - \mathbf{x}_i^t\boldsymbol{\beta}}{s}\right) = \lim_{s\to 0^+} \frac{1}{n}\sum_{i\notin A} \rho\left(\frac{y_i - \mathbf{x}_i^t\boldsymbol{\beta}}{s}\right) = \frac{\#A^c}{n} > b.$$

Luego para cada $\boldsymbol{\beta} \in \mathbb{R}^p$ existe $s_n(\boldsymbol{\beta}) > 0$.

Llamemos $a_i = y_i - \mathbf{x}_i^t\boldsymbol{\beta}$ y supongamos que existen s_1, s_2 tales que $s_1 < s_2$ y

$$\frac{1}{n}\sum_{i\notin A}\rho\left(\frac{a_i}{s_1}\right) = b = \frac{1}{n}\sum_{i\notin A}\rho\left(\frac{a_i}{s_2}\right).$$

Luego existe i_0 tal que $\rho\left(a_{i_0}/s_2\right) < 1$ pues en caso contrario, si $\rho\left(a_i/s_2\right) = 1$ para todo i resultaría $b = \#A^c/n$ y esto no es posible.

Como $|a_{i_0}|/s_1 > |a_{i_0}|/s_2$ y $\rho(a_{i_0}/s_2) < 1$ entonces por **C3** y **C7** obtenemos $\rho(a_{i_0}/s_1) > \rho(a_{i_0}/s_2)$. A partir del supuesto **C5** se tiene que $\rho(a_i/s_1) \geq \rho(a_i/s_2)$ para todo i, luego

$$\frac{1}{n}\sum_{i\notin A}\rho\left(\frac{a_i}{s_1}\right) > \frac{1}{n}\sum_{i\notin A}\rho\left(\frac{a_i}{s_2}\right),$$

lo cual es un absurdo. □

Demostración del Lema 5.9

Por el Lema 5.8 el conjunto

$$\left\{s \in \mathbb{R}^+ : \frac{1}{n}\sum_{i=1}^{n}\rho\left(\frac{y_i - \mathbf{x}_i^t\boldsymbol{\beta}}{s}\right) = b\right\}$$

es no vacío y por estar acotado inferiormente, tiene ínfimo s^*.

Observar que s^* no puede ser 0 pues es caso contario resultaría

$$b = \lim_{s\to 0^+}\frac{1}{n}\sum_{i=1}^{n}\rho\left(\frac{y_i - \mathbf{x}_i^t\boldsymbol{\beta}}{s}\right) = \frac{\#A^c}{n}$$

y esto no es posible. Además

$$b = \lim_{s\to s^*}\frac{1}{n}\sum_{i=1}^{n}\rho\left(\frac{y_i - \mathbf{x}_i^t\boldsymbol{\beta}}{s}\right) = \frac{1}{n}\sum_{i=1}^{n}\rho\left(\frac{y_i - \mathbf{x}_i^t\boldsymbol{\beta}}{s^*}\right)$$

y por lo tanto s^* es solución de la ecuación (30). □

Demostración del Lema 5.10

Supongamos por el absurdo que $\widetilde{\boldsymbol{\beta}}_n$ no minimiza, luego existe $\widetilde{\boldsymbol{\beta}}_n^*$ tal que

$$\sum_{i=1}^{n}\rho\left(\frac{y_i - \mathbf{x}_i^t\widetilde{\boldsymbol{\beta}}_n^*}{\widetilde{\sigma}_n}\right) < \sum_{i=1}^{n}\rho\left(\frac{y_i - \mathbf{x}_i^t\widetilde{\boldsymbol{\beta}}_n}{\widetilde{\sigma}_n}\right) = nb,$$

pero entonces

$$\sum_{i=1}^{n}\rho\left(\frac{y_i - \mathbf{x}_i^t\widetilde{\boldsymbol{\beta}}_n^*}{\widetilde{\sigma}_n}\right) < \sum_{i=1}^{n}\rho\left(\frac{y_i - \mathbf{x}_i^t\widetilde{\boldsymbol{\beta}}_n^*}{s_n(\widetilde{\boldsymbol{\beta}}_n^*)}\right) = nb$$

de donde resulta $s_n(\widetilde{\boldsymbol{\beta}}_n^*) < \widetilde{\sigma}_n$ en virtud de que la función

$$h_{\widetilde{\boldsymbol{\beta}}_n^*}(s) = \sum_{i=1}^{n} \rho\left(\frac{y_i - \mathbf{x}_i^t \widetilde{\boldsymbol{\beta}}_n^*}{s}\right)$$

es monótona no creciente como función de $s > 0$ por las condiciones **C3** y **C5**. Esto contradice el hecho que $\widetilde{\sigma}_n$ sea la escala mínima. □

Demostración del Lema 5.11

Queremos ver que $\widetilde{\sigma}_n = \min_{\boldsymbol{\beta}} s_n(\boldsymbol{\beta})$. Supongamos que existe $\widetilde{\boldsymbol{\beta}}_n^*$ tal que $s_n(\widetilde{\boldsymbol{\beta}}_n^*) < \widetilde{\sigma}_n = s_n(\widetilde{\boldsymbol{\beta}}_n)$. Luego

$$\sum_{i=1}^{n} \rho\left(\frac{y_i - \mathbf{x}_i^t \widetilde{\boldsymbol{\beta}}_n^*}{\widetilde{\sigma}_n}\right) < \sum_{i=1}^{n} \rho\left(\frac{y_i - \mathbf{x}_i^t \widetilde{\boldsymbol{\beta}}_n^*}{s_n(\widetilde{\boldsymbol{\beta}}_n^*)}\right) = b$$

y por otro lado

$$\sum_{i=1}^{n} \rho\left(\frac{y_i - \mathbf{x}_i^t \widetilde{\boldsymbol{\beta}}_n^*}{\widetilde{\sigma}_n}\right) < \sum_{i=1}^{n} \rho\left(\frac{y_i - \mathbf{x}_i^t \widetilde{\boldsymbol{\beta}}_n}{\widetilde{\sigma}_n}\right) = b.$$

Esto contradice el supuesto de que $\widetilde{\boldsymbol{\beta}}_n$ minimiza (36). □

Demostración del Lema 5.12

Consideremos la función

$$h_{\boldsymbol{\beta}}(s) = E_{\boldsymbol{\beta}_0,\sigma_0}\left(\rho\left(\frac{y - \mathbf{x}^t\boldsymbol{\beta}}{s}\right)\right).$$

Por el Teorema de Convergencia Mayorada se tiene $\lim_{s\to+\infty} h_{\boldsymbol{\beta}}(s) = \rho(0) = 0$.

Por otro lado notar que

$$h_{\beta}(s) = E_{\boldsymbol{\beta}_0,\sigma_0}\left(\rho\left(\frac{y - \mathbf{x}^t\boldsymbol{\beta}}{s}\right) I(|y - \mathbf{x}^t\boldsymbol{\beta}| > 0)\right).$$

Luego

$\lim_{s\to 0^+} h_{\beta}(s) = E_{\boldsymbol{\beta}_0,\sigma_0}(I(|y - \mathbf{x}^t\boldsymbol{\beta}| > 0)) = 1 - P(y - \mathbf{x}^t\boldsymbol{\beta} = 0)$.

Suponiendo verdadero el modelo (1) tenemos

$\lim_{s\to 0^+} h_{\beta}(s) = 1 - P(\mathbf{x}^t\boldsymbol{\beta}_0 + \sigma_0 u - \mathbf{x}^t\boldsymbol{\beta} = 0) = 1 - P(\sigma_0 u = \mathbf{x}^t(\boldsymbol{\beta} - \boldsymbol{\beta}_0))$.

Si el modelo es el de posición y escala, la probabilidad anterior es 0. Si no

$P(\sigma_0 u = \mathbf{x}^t(\boldsymbol{\beta} - \boldsymbol{\beta}_0)) = E(I(\sigma_0 u = \mathbf{x}^t(\boldsymbol{\beta} - \boldsymbol{\beta}_0))) = E(E(I(\sigma_0 u = \mathbf{x}^t(\boldsymbol{\beta} - \boldsymbol{\beta}_0))|\mathbf{x}))$.

Sea $\tilde{\mathbf{x}}$ un vector fijo de la misma dimensión que $\mathbf{x}$. Luego usando el hecho que u y $\mathbf{x}$ son independientes resulta

$$\begin{aligned} E(I(\sigma_0 u = \mathbf{x}^t(\boldsymbol{\beta} - \boldsymbol{\beta}_0))|\mathbf{x} = \tilde{\mathbf{x}}) &= E(I(\sigma_0 u = \tilde{\mathbf{x}}^t(\boldsymbol{\beta} - \boldsymbol{\beta}_0))|\mathbf{x} = \tilde{\mathbf{x}}) \\ &= E(I(\sigma_0 u = \tilde{\mathbf{x}}^t(\boldsymbol{\beta} - \boldsymbol{\beta}_0))) \\ &= P(\sigma_0 u = \tilde{\mathbf{x}}^t(\boldsymbol{\beta} - \boldsymbol{\beta}_0)) = 0 \end{aligned}$$

pues u es una variable aleatoria continua. Luego $\lim_{s\to 0^+} h_\beta(s) = 1$.

Como h_β es continua entonces para todo $b \in (0,1)$ y $\boldsymbol{\beta} \in \mathbb{R}^p$ existe la escala residual $s(\beta)$ y además es positiva. □

Demostración del Lema 5.13

Por el Lema 5.12 el conjunto

$$\left\{ s \in \mathbb{R}^+ . : E_{\boldsymbol{\beta}_0,\sigma_0}\left(\rho\left(\frac{y - \mathbf{x}^t\boldsymbol{\beta}}{s}\right)\right)\right\}$$

es no vacío y por estar acotado inferiormente, tiene ínfimo s^*.

Observar que s^* no puede ser 0 pues si no, repitiendo los mismos procedimientos hechos en la demostración del Lema 5.12, resultaría

$$b = \lim_{s\to 0^+} h_\beta(s) = 1.$$

Además apelando nuevamente a la continuidad de $h_\beta(s)$, tenemos que

$$h_\beta(s^*) = \lim_{s\to s^*} h_\beta(s) = b$$

y por lo tanto s^* es una escala residual. □

Demostración del Lema 5.14

La demostración consiste en probar que la función

$$H_\beta(s) = E_{\boldsymbol{\beta}_0,\sigma_0}\left(\rho\left(\frac{y - \mathbf{x}^t\boldsymbol{\beta}}{s}\right)\right)$$

es estrictamente decreciente. Supongamos que $s_1 < s_2$ y consideremos

$$H_\beta(s) = E\left(E\left(\rho\left(\frac{y - \mathbf{x}^t\boldsymbol{\beta}}{s}\right)\middle|\mathbf{x}\right)\right).$$

Por otro lado, bajo el modelo (1) y sabiendo que $\mathbf{x}$ y u son independientes queda

$$E\left(\rho\left(\frac{y - \mathbf{x}^t\boldsymbol{\beta}}{s}\right)\middle|\mathbf{x} = \tilde{\mathbf{x}}\right) = E_0\left(\rho\left(\frac{\sigma_0 u - \tilde{\mathbf{x}}^t(\boldsymbol{\beta} - \boldsymbol{\beta}_0)}{s}\right)\right).$$

Bajo las hipótesis del Lema sabemos que la función

$$h_\mu(s) = E_0\left(\rho\left(\frac{u - \mu}{s}\right)\right)$$

es estrictamente decreciente. Luego

$$E\left(\rho\left(\frac{y - \mathbf{x}^t\boldsymbol{\beta}}{s_2}\right)\middle|\mathbf{x} = \tilde{\mathbf{x}}\right) < E\left(\rho\left(\frac{y - \mathbf{x}^t\boldsymbol{\beta}}{s_1}\right)\middle|\mathbf{x} = \tilde{\mathbf{x}}\right)$$

y por lo tanto

$$H_\beta(s_2) < H_\beta(s_1).$$

□

Demostración del Teorema 5.1

Sea $\boldsymbol{\beta} \neq \boldsymbol{\beta}_0$ y consideramos

$$R(\boldsymbol{\beta}) = E_{\boldsymbol{\beta}_0,\sigma_0}\left(\rho\left(\frac{y-\mathbf{x}^t\boldsymbol{\beta}}{\sigma_0}\right)\right) \quad \text{y} \quad H(\mathbf{x}) = E\left(\rho\left(\frac{y-\mathbf{x}^t\boldsymbol{\beta}}{\sigma_0}\right)\middle|\,\mathbf{x}\right).$$

Observar que $R(\boldsymbol{\beta}_0) = q(0)$ siendo $q(\mu) = E_0(\rho(u-\mu))$ y $R(\beta) = E(H(\mathbf{x}))$. Bastaría probar que $R(\beta) > q(0)$. Si el modelo es el de posición y escala, es decir si $y = \beta_0 + \sigma_0 u$ se tendrá

$$R(\boldsymbol{\beta}) = E_0\left(\rho\left(\frac{\beta_0-\beta}{\sigma_0}+u\right)\right)$$

y el resultado es inmediato debido al Lema 5.4. Bajo el modelo (1) tenemos

$$H(\mathbf{x}) = E\left(\rho\left(\frac{\mathbf{x}^t(\boldsymbol{\beta}_0-\boldsymbol{\beta})}{\sigma_0}+u\right)\middle|\,\mathbf{x}\right).$$

Notaremos por $\tilde{\mathbf{x}}$ una representación cualquiera del vector aleatorio $\mathbf{x}$. Como u y $\tilde{\mathbf{x}}$ son independientes entonces

$$H(\tilde{\mathbf{x}}) = E_0\left(\rho\left(\frac{\tilde{\mathbf{x}}^t(\boldsymbol{\beta}_0-\boldsymbol{\beta})}{\sigma_0}+u\right)\right).$$

Consideremos el conjunto $A = \{\tilde{\mathbf{x}}/\tilde{\mathbf{x}}^t(\boldsymbol{\beta}_0-\boldsymbol{\beta}) = 0\}$. Por la condición **CI** se tiene $P_{\beta_0,\sigma_0}(A) < 1$. Luego $R(\beta) = E(H(\mathbf{x})) = \int_A H(\tilde{\mathbf{x}})dF_{\mathbf{x}}(\tilde{\mathbf{x}}) + \int_{A^c} H(\tilde{\mathbf{x}})dF_{\mathbf{x}}(\tilde{\mathbf{x}})$, donde $F_{\mathbf{x}}$ es la función de distribución acumulada de $\mathbf{x}$. Si $\tilde{\mathbf{x}} \in A$ entonces $H(\tilde{\mathbf{x}}) = q(0)$ y si $\tilde{\mathbf{x}} \notin A$ resulta $H(\tilde{\mathbf{x}}) > q(0)$ debido al Lema 5.4. Sabiendo que $P_{\beta_0,\sigma_0}(A^c) > 0$ se deduce que $R(\boldsymbol{\beta}) > q(0)$.

Veamos a continuación que $s(\boldsymbol{\beta}_0) = \sigma_0$. Por definición de escala residual sabemos que

$$E_{\boldsymbol{\beta}_0,\sigma_0}\left(\rho\left(\frac{y-\mathbf{x}^t\boldsymbol{\beta}_0}{s(\boldsymbol{\beta}_0)}\right)\right) = b.$$

Bajo el modelo (1) se obtiene

$$E_0\left(\rho\left(\frac{\sigma_0 u}{s(\boldsymbol{\beta}_0)}\right)\right) = E_0\left(\rho\left(\frac{u}{c}\right)\right) = b,$$

siendo $c = s(\boldsymbol{\beta}_0)/\sigma_0$. Por el Lema 5.2 y usando el hecho que $\mu_0 = 0$ resulta $c = s_0 = 1$ como queríamos probar. □

Demostración del Corolario 5.1

Consideremos $\boldsymbol{\beta} \neq \boldsymbol{\beta}_0$, luego a partir del Teorema 5.1 obtenemos

$$E_{\boldsymbol{\beta}_0,\sigma_0}\left(\rho\left(\frac{y-\mathbf{x}^t\boldsymbol{\beta}}{s(\boldsymbol{\beta})}\right)\right) = b = E_{\boldsymbol{\beta}_0,\sigma_0}\left(\rho\left(\frac{y-\mathbf{x}^t\boldsymbol{\beta}_0}{\sigma_0}\right)\right) < E_{\boldsymbol{\beta}_0,\sigma_0}\left(\rho\left(\frac{y-\mathbf{x}^t\boldsymbol{\beta}}{\sigma_0}\right)\right).$$

Luego para todo $\boldsymbol{\beta} \neq \boldsymbol{\beta}_0$ se tiene que $s(\boldsymbol{\beta}_0) = \sigma_0 < s(\boldsymbol{\beta})$ debido a que la función

$$h_{\boldsymbol{\beta}}(s) = E_{\boldsymbol{\beta}_0,\sigma_0}\left(\rho\left(\frac{y-\mathbf{x}^t\boldsymbol{\beta}}{s}\right)\right)$$

es monótona no creciente como función de s.

□

En lo que sigue comenzaremos a desarrollar los conceptos teóricos necesarios para probar el Teorema 5.2.

S funcionales

La definición de escala residual $s(\boldsymbol{\beta})$ se puede generalizar de la siguiente forma:

Sea H la función de distribución del vector aleatorio $(\mathbf{x}, y)$. Fijados $\boldsymbol{\beta} \in \mathbb{R}^p$ y $b \in (0,1)$ se llama **funcional de escala residual** $S(H, \boldsymbol{\beta})$ a cada solución positiva (en caso de existir) de la ecuación

$$E_H\left(\rho\left(\frac{y-\mathbf{x}^t\boldsymbol{\beta}}{S(H,\boldsymbol{\beta})}\right)\right) = b. \tag{77}$$

Como la ecuación (77) podría tener más de una solución, supondremos que para cada $\boldsymbol{\beta} \in \mathbb{R}^p$,

$$S(H,\boldsymbol{\beta}) = \text{ínf}\left\{s \in \mathbb{R}^+ : E_H\left(\rho\left(\frac{y-\mathbf{x}^t\boldsymbol{\beta}}{s}\right)\right)\right\}.$$

Hay que proceder con cuidado porque este ínfimo podría ser cero. Sin embargo, bajo ciertas condiciones, el ínfimo se realiza y por lo tanto, es distinto de cero.

Observación 14.1. Si $H_{\boldsymbol{\beta}_0,\sigma_0}$ es la función de distribución acumulada del vector aleatorio $(\mathbf{x}, y)$ bajo el modelo (1) entonces $S(H_{\boldsymbol{\beta}_0,\sigma_0}, \boldsymbol{\beta}) = s(\boldsymbol{\beta})$.

El **M funcional de escala del error** se define como

$$S(H) = \inf_{\beta \in \mathbb{R}^p} S(H,\boldsymbol{\beta}), \tag{78}$$

y el **S funcional de regresión** T_S (en caso de existir) por

$$T_S(H) = \arg\min_{\beta \in \mathbb{R}^p} S(H,\boldsymbol{\beta}). \tag{79}$$

En este caso

$$S(H) = S(H, T_S(H)) = \min_{\beta \in \mathbb{R}^p} S(H,\boldsymbol{\beta}) \tag{80}$$

y

$$E_H\left(\rho\left(\frac{y-\mathbf{x}^t T_S(H)}{S(H)}\right)\right)=b.$$

Supondremos que el funcional T_S está bien definido por la elección de un único valor. Sin embargo, los siguientes resultados no dependerán de la elección de dicho valor.

En lo que sigue hasta terminar la sección asumiremos para ρ las condiciones **C1** a **C7** necesarias en cada caso. No indicaremos en las hipótesis de cada resultado cuáles son, en virtud de que estos resultados serán utilizados más tarde suponiendo todas las condiciones para ρ. De todas formas, son básicamente las necesarias para garantizar la monotonía de ρ.

Veamos a continuación que T_S es además, un **M funcional de regresión**, es decir

$$T_S(H)=\arg\min_{\beta\in\mathbb{R}^p} E_H\left(\rho\left(\frac{y-\mathbf{x}^t\boldsymbol{\beta}}{S(H)}\right)\right). \tag{81}$$

Supongamos por el absurdo que existe $\boldsymbol{\beta}_1$ tal que

$$E_H\left(\rho\left(\frac{y-\mathbf{x}^t\boldsymbol{\beta}_1}{S(H)}\right)\right)<E_H\left(\rho\left(\frac{y-\mathbf{x}^t T_S(H)}{S(H)}\right)\right)=b.$$

Por otro lado sabemos que

$$E_H\left(\rho\left(\frac{y-\mathbf{x}^t\boldsymbol{\beta}_1}{S(H,\boldsymbol{\beta}_1)}\right)\right)=b$$

de donde resulta $S(H,\boldsymbol{\beta}_1)<S(H)$ debido a la monotonía de la función

$$h_{\beta_1}(s)=E_H\left(\rho\left(\frac{y-\mathbf{x}^t\boldsymbol{\beta}_1}{s}\right)\right).$$

Esto contradice la definición de $S(H)$. Observar que no estamos probando que $T_S(H)$ sea único.

Recíprocamente si existe $S(H)$ y $T_S(H)$ como en (80) y (81) respectivamente entonces $T_S(H)$ resulta un S funcional de regresión, es decir verifica (79). Veamos este resultado y asumamos para ello que la función

$$h_{T_S(H)}(s)=E_H\left(\rho\left(\frac{y-\mathbf{x}^t T_S(H)}{s}\right)\right)$$

es estrictamente decreciente.

Queremos ver que $S(H)=S(H,T_S(H))$. Como $S(H)=\min_{\beta\in\mathbb{R}^p} S(H,\boldsymbol{\beta})$ entonces vamos a suponer $S(H)<S(H,T_S(H))$. Sea además $\boldsymbol{\beta}_1$ tal que $S(H)=S(H,\boldsymbol{\beta}_1)$. Luego

$$E_H\left(\rho\left(\frac{y-\mathbf{x}^t T_S(H)}{S(H,T_S(H))}\right)\right)<E_H\left(\rho\left(\frac{y-\mathbf{x}^t T_S(H)}{S(H)}\right)\right)\leq E_H\left(\rho\left(\frac{y-\mathbf{x}^t\boldsymbol{\beta}_1}{S(H)}\right)\right).$$

Como los extremos son iguales a b, llegamos de esta forma a una contradicción. A partir de ahora desarrollaremos la teoría necesaria para probar el Teorema 5.2

Definición 14.1. Sean $\mathcal{F} = \{F_{\boldsymbol{\theta}}\}$ con $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p$ una familia paramétrica de distribuciones. Sea $\mathbf{T}$ un funcional definido sobre un conjunto de distribuciones en $\mathbb{R}^p$. Diremos que $\mathbf{T}$ es Fisher consistente de $\boldsymbol{\theta}$ si $\mathbf{T}(F_\theta) = \boldsymbol{\theta}$

El siguiente resultado es una consecuencia inmediata del Teorema 5.1.

Lema 14.2. *Supongamos que estamos bajo los supuestos del modelo (1) y que se verifica la condición de identificabilidad **CI**. Supongamos que ρ satisface las condiciones **C1** a **C7**. Asumamos también que la densidad de los errores f_0, es estrictamente unimodal y que $E_0(\rho(u)) = b$. Luego (T_S, S) es Fisher consistente de $(\boldsymbol{\beta}_0, \sigma_0)$, considerando a T_S un M o S funcional de regresión.*

Demostración

Por el Corolario 5.1 y la observación 14.1 se tiene que

$$S(H_{\boldsymbol{\beta}_0,\sigma_0}) = \min_{\boldsymbol{\beta}\in\mathbb{R}^p} S(H_{\boldsymbol{\beta}_0,\sigma_0}, \boldsymbol{\beta}) = \min_{\boldsymbol{\beta}\in\mathbb{R}^p} s(\boldsymbol{\beta}) = s(\boldsymbol{\beta}_0) = \sigma_0$$

y

$$T_S(H_{\boldsymbol{\beta}_0,\sigma_0}) = \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^p} S(H_{\boldsymbol{\beta}_0,\sigma_0}, \boldsymbol{\beta}) = \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^p} s(\boldsymbol{\beta}) = \boldsymbol{\beta}_0$$

y además es único.

Por el Teorema 5.1

$$T_S(H_{\boldsymbol{\beta}_0,\sigma_0}) = \arg\min_{\beta} E_{\boldsymbol{\beta}_0,\sigma_0}\left(\rho\left(\frac{y - \mathbf{x}^t\boldsymbol{\beta}}{\sigma_0}\right)\right) = \boldsymbol{\beta}_0$$

y además es el único. □

Consistencia fuerte de estimadores definidos a partir de funcionales

En lo que sigue comenzaremos a estudiar resultados auxiliares que nos conducirán a probar la consistencia fuerte de los S-estimadores de regresión.

Para ellos recordaremos la definición de continuidad débil de un funcional. Veremos condiciones suficientes para que los funcionales definidos en la sección anterior sean débilmente continuos. La consistencia fuerte de los S estimadores será deducida a partir de la continuidad débil de los funcionales que dan origen a dichos estimadores.

Definición 14.2. Para cada $\alpha \in \Delta$ sea $\boldsymbol{z}_\alpha$ un vector aleatorio con función de distribución H_α. Se dice que $\{H_\alpha\}_{\alpha\in\Delta}$ es *tight* o rígido si para todo $\epsilon > 0$ existe $K > 0$ tal que $P(||\boldsymbol{z}_\alpha|| > K) < \epsilon$ para toda $\alpha \in \Delta$.

Lema 14.3. *Sean $\boldsymbol{z}_n$ con $n \in \mathbb{N}$ vectores aleatorias con función de distribución H_n y $\boldsymbol{z}$ un vector aleatorio con función de distribución H. Si $H_n \xrightarrow{D} H$ entonces $\{(H_n)_{n\in\mathbb{N}}\} \cup H$ es un conjunto tight o rígido.*

Demostración

Haremos la prueba para el caso de variables aleatorias reales y positivas en casi todo punto. Notaremos $H_0 = H$. Queremos ver que dado $\epsilon > 0$ existe K tal que $H_n(K) > 1-\epsilon$ para todo $n \in \mathbb{N}_0$. Como $\lim_{t\to\infty} H_0(t) = 1$ entonces dado $\epsilon > 0$ existe t_0 tal que para todo $t \geq t_0$, $H_0(t) \geq 1 - \frac{\epsilon}{2}$. Como los puntos de continuidad de H son densos en $\mathbb{R}$, podemos suponer que t_0 es un punto de continuidad de H_0. Como $H_n \xrightarrow{D} H$ entonces para cada t punto de continuidad de H se tiene $\lim_{n\to\infty} H_n(t) = H_0(t)$. Luego dado t_0 existe $n_0 \in \mathbb{N}$ tal que para todo $n \geq n_0$, $|H_n(t_0) - H_0(t_0)| \leq \frac{\epsilon}{2}$ de donde resulta que

$$\forall n \geq n_0 \;\; H_n(t_0) \geq H_0(t_0) - \frac{\epsilon}{2} \geq 1 - \epsilon.$$

En particular como H_n es una función no decreciente obtenemos

$$H_n(t) \geq 1 - \epsilon \;\; \forall t \geq t_0 \text{ y } \forall n \geq n_0.$$

Para cada $0 < i < n_0$ como el $\lim_{x\to\infty} H_i(t) = 1$ y H_i es no decreciente entonces dado $\epsilon > 0$ existe t_i tal que $\forall x \geq t_i,\; H_i(t) \geq 1 - \epsilon$. Finalmente la prueba se termina considerando $K = \text{máx}\{t_0, t_1, \ldots, t_{n_0-1}\}$. □

Las demostraciones desde aquí hasta el Teorema 14.1, se deben a Fasano, Maronna, Sued y Yohai (2010).

Consideremos H, $(H_n)_{n\in\mathbb{N}}$ distribuciones en $\mathbb{R}^{p+1}$ y $(\boldsymbol{\beta}_n)_{n\in\mathbb{N}}$, $(\sigma_n)_{n\in\mathbb{N}}$ sucesiones en $\mathbb{R}^p$, $\mathbb{R}^+$ respectivamente.

Lema 14.4. *Supongamos que $\boldsymbol{\beta}_n \to \boldsymbol{\beta}$, $\sigma_n \to \sigma > 0$ y $H_n \xrightarrow{D} H$, luego*

$$\lim_{n\to\infty} E_{H_n}\left(\rho\left(\frac{y - \mathbf{x}^t\boldsymbol{\beta}_n}{\sigma_n}\right)\right) = E_H\left(\rho\left(\frac{y - \mathbf{x}^t\boldsymbol{\beta}_n}{\sigma_n}\right)\right)$$

Demostración

Como $H_n \xrightarrow{D} H$ y ρ es continua y acotada entonces

$$\lim_{n\to\infty} E_{H_n}\left(\rho\left(\frac{y - \mathbf{x}^t\boldsymbol{\beta}}{\sigma}\right)\right) = E_H\left(\rho\left(\frac{y - \mathbf{x}^t\boldsymbol{\beta}}{\sigma}\right)\right).$$

Luego será suficiente probar

$$\lim_{n\to\infty}\left(E_{H_n}\left(\rho\left(\frac{y - \mathbf{x}^t\boldsymbol{\beta}_n}{\sigma_n}\right)\right) - E_{H_n}\left(\rho\left(\frac{y - \mathbf{x}^t\boldsymbol{\beta}}{\sigma}\right)\right)\right) = 0.$$

Como $\{(H_n)_{n\in\mathbb{N}}\} \cup H$ es un conjunto tight de distribuciones, es suficiente mostrar que si P es una conjunto tight de distribuciones de $(\mathbf{x}, y)$ entonces

$$\lim_{n\to\infty} \sup_{H\in P}\left|E_H\left(\rho\left(\frac{y - \mathbf{x}^t\boldsymbol{\beta}_n}{\sigma_n}\right)\right) - E_H\left(\rho\left(\frac{y - \mathbf{x}^t\boldsymbol{\beta}}{\sigma}\right)\right)\right| = 0.$$

Sea $M \in \mathbb{R}_{>0}$, luego

$$\left| E_H\left(\rho\left(\frac{y-\mathbf{x}^t\boldsymbol{\beta}_n}{\sigma_n}\right)\right) - E_H\left(\rho\left(\frac{y-\mathbf{x}^t\boldsymbol{\beta}}{\sigma}\right)\right)\right| \leq$$

$$\leq \left| \int_{||z||>M} \rho\left(\frac{y-\mathbf{x}^t\boldsymbol{\beta}_n}{\sigma_n}\right) dH - \int_{||z||>M} \rho\left(\frac{y-\mathbf{x}^t\boldsymbol{\beta}}{\sigma}\right) dH \right| +$$

$$+ \left| \int_{||z||\leq M} \rho\left(\frac{y-\mathbf{x}^t\boldsymbol{\beta}_n}{\sigma_n}\right) dH - \int_{||z||\leq M} \rho\left(\frac{y-\mathbf{x}^t\boldsymbol{\beta}}{\sigma}\right) dH \right| \leq$$

$$\leq 2E_F(I(||z|| > M)) + E_F\left(\left|\rho\left(\frac{y-\mathbf{x}^t\boldsymbol{\beta}_n}{\sigma_n}\right) - \rho\left(\frac{y-\mathbf{x}^t\boldsymbol{\beta}}{\sigma}\right)\right| I(||z|| \leq M)\right).$$

Por otro lado tenemos

$$\begin{aligned}
\left|\frac{y-\mathbf{x}^t\boldsymbol{\beta}_n}{\sigma_n} - \frac{y-\mathbf{x}^t\boldsymbol{\beta}}{\sigma}\right| &= \left|\frac{y\sigma - \mathbf{x}^t\boldsymbol{\beta}_n\sigma - y\sigma_n + \sigma_n\mathbf{x}^t\boldsymbol{\beta}_n}{\sigma\sigma_n}\right| \\
&= \left|\frac{y(\sigma-\sigma_n) + (\sigma_n-\sigma)\mathbf{x}^t\boldsymbol{\beta} + \sigma(\mathbf{x}^t\boldsymbol{\beta} - \mathbf{x}^t\boldsymbol{\beta}_n)}{\sigma\sigma_n}\right| \\
&\leq \frac{|y||\sigma-\sigma_n| + |\sigma_n-\sigma||\mathbf{x}^t\boldsymbol{\beta}| + |\sigma||\mathbf{x}^t\boldsymbol{\beta} - \mathbf{x}^t\boldsymbol{\beta}_n|}{\sigma\sigma_n} \\
&= A_n.
\end{aligned}$$

Notar que $\lim_{n\to\infty} A_n = 0$. Por ser P *tight*, dado $\epsilon > 0$ existe $K > 0$ tal que

$$\sup_{F\in P} P_F(||(\mathbf{x}, y)|| > K) \leq \frac{\epsilon}{4}.$$

Tomamos $M = K$. Fijando el ϵ anterior, por continuidad de ρ existe $\delta > 0$ tal que si $|u-v| < \delta$ entonces $|\rho(u)-\rho(v)| < \frac{\epsilon}{2}$. Fijado δ existe n_0 tal que para todo $n \geq n_0$, $A_n \leq \delta$. Luego

$$\left| E_F\left(\rho\left(\frac{y-\mathbf{x}^t\boldsymbol{\beta}_n}{\sigma_n}\right)\right) - E_F\left(\rho\left(\frac{y-\mathbf{x}^t\boldsymbol{\beta}}{\sigma}\right)\right)\right| \leq$$

$$\leq 2\frac{\epsilon}{4} + \frac{\epsilon}{2}P(||(\mathbf{x}, y)|| \leq K) \leq \epsilon.$$

□

Lema 14.5. *Sean H y $(H_n)_{n\in\mathbb{N}}$ tales que $H_n \xrightarrow{D} H$. Supongamos que $T_S(H_n)$ existe para cada $n \in \mathbb{N}$ y además $(T_S(H_n))_{n\in\mathbb{N}}$ está acotada. Si $T_S(H)$ existe y es único y $S(H_n) \longrightarrow S(H)$ entonces $T_S(H_n) \longrightarrow T_S(H)$.*

Demostración

Como la sucesión $(T_S(H_n))_{n\in\mathbb{N}}$ está acotada entonces tiene una subsucesión convergente. Luego bastaría ver que $T_S(H)$ es el único punto límite de dicho conjunto. Es decir

debemos ver que si tomamos cualquier subsucesión convergente $T_S(H_{n_k}) \to \beta^*$ entonces $\beta^* = T_S(H)$. Por definición de $T_S(H_{n_k})$ y $S(H_{n_k})$ tenemos que

$$E_{H_{n_k}}\left(\rho\left(\frac{y-\mathbf{x}^t T_S(H_{n_k})}{S(H_{n_k})}\right)\right) \leq E_{H_{n_k}}\left(\rho\left(\frac{y-\mathbf{x}^t T_S(H)}{S(H_{n_k})}\right)\right).$$

En virtud del Lema 14.4 estamos en condiciones de tomar límite a ambos miembros. De esta forma queda

$$E_H\left(\rho\left(\frac{y-\mathbf{x}^t \beta^*}{S(H)}\right)\right) \leq E_H\left(\rho\left(\frac{y-\mathbf{x}^t T_S(H)}{S(H)}\right)\right).$$

Como $T_S(H)$ es el único mínimo entonces $\beta^* = T_S(H)$. □

Lema 14.6. *Supongamos que* $H_n \xrightarrow{D} H$, $||\boldsymbol{\beta}_n|| \to \infty$ *y* $(\sigma_n)_{n\in\mathbb{R}}$ *acotada. Luego*

$$\liminf_{n\to\infty} E_{H_n}\left(\rho\left(\frac{y-\mathbf{x}^t\boldsymbol{\beta}_n}{\sigma_n}\right)\right) \geq 1 - c(H).$$

Demostración

El límite inferior de una sucesión acotada es el ínfimo de los puntos límites de la sucesión. Consideremos una subsucesión convergente

$$E_{H_{n_k}}\left(\rho\left(\frac{y-\mathbf{x}^t\boldsymbol{\beta}_{n_k}}{\sigma_{n_k}}\right)\right) \to L$$

y probemos que $L \geq 1 - c(H)$. Consideremos la sucesión acotada $\boldsymbol{\gamma}_{n_k} = \frac{\boldsymbol{\beta}_{n_k}}{||\boldsymbol{\beta}_{n_k}||}$, luego existe una subsucesión convergente $\boldsymbol{\gamma}_{n_{k_j}} = \frac{\boldsymbol{\beta}_{n_{k_j}}}{||\boldsymbol{\beta}_{n_{k_j}}||} \to \boldsymbol{\gamma}$. En particular

$$E_{H_{n_{k_j}}}\left(\rho\left(\frac{y-\mathbf{x}^t\boldsymbol{\beta}_{n_{k_j}}}{\sigma_{n_{k_j}}}\right)\right) \to L.$$

En lo que sigue haremos un abuso de notación y en vez de notar n_{k_j} denotamos simplemente j. Como $(\sigma_n)_{n\in\mathbb{N}}$ está acotada, existe un σ^* tal que $\sigma_n \leq \sigma^*$. Por otro lado para todo $\epsilon > 0$ existe d_ϵ tal que si $|u| \geq d_\epsilon$ entonces $\rho(u) > 1 - \epsilon$. Luego

$$\begin{aligned}
E_{H_j}\left(\rho\left(\frac{y-\mathbf{x}^t\boldsymbol{\beta}_j}{\sigma_j}\right)\right) &\geq E_{H_j}\left(\rho\left(\frac{y-\mathbf{x}^t\boldsymbol{\beta}_j}{\sigma^*}\right)\right) \\
&\geq (1-\epsilon)P_{H_j}\left(\left|\frac{y-\mathbf{x}^t\boldsymbol{\beta}_j}{\sigma^*}\right| \geq d_\epsilon\right) \\
&= (1-\epsilon)P_{H_j}\left(\left|\frac{y-\mathbf{x}^t\boldsymbol{\gamma}_j||\boldsymbol{\beta}_j||}{||\boldsymbol{\beta}_j||}\right| \geq \frac{d_\epsilon\sigma^*}{||\boldsymbol{\beta}_j||}\right).
\end{aligned}$$

Sean $(\mathbf{x}_0, y_0)$ con distribución H y $(\mathbf{x}_j, y_j)$ con distribución H_j. Sabemos que $(\mathbf{x}_j, y_j) \xrightarrow{D} (\mathbf{x}_0, y_0)$ y que $\frac{y_j}{||\beta_j||} \to 0$, luego componiendo con funciones continuas adecuadas se obtiene $\frac{y_j}{||\boldsymbol{\beta}_j||} - \mathbf{x}_j^t\boldsymbol{\gamma}_j \xrightarrow{D} -\mathbf{x}_0^t\boldsymbol{\gamma}$. Consideramos $\alpha > 0$. Podemos suponer que α es un punto de continuidad de H pues los puntos de continuidad de H son densos. Luego

$$\begin{aligned}
\lim_{j\to\infty} E_{H_j}\left(\rho\left(\frac{y-\mathbf{x}^t\boldsymbol{\beta}_j}{\sigma_j}\right)\right) &\geq \lim_{j\to\infty}(1-\epsilon)P_{H_j}\left(\left|\frac{y-\mathbf{x}^t\boldsymbol{\gamma}_j||\boldsymbol{\beta}_j||}{||\boldsymbol{\beta}_j||}\right| \geq \frac{d_\epsilon\sigma^*}{||\boldsymbol{\beta}_j||}\right) \\
&\geq (1-\epsilon)\lim_{j\to\infty} P_{H_j}\left(\left|\frac{y-\mathbf{x}^t\boldsymbol{\gamma}_j||\boldsymbol{\beta}_j||}{||\boldsymbol{\beta}_j||}\right| \geq \alpha\right) \\
&= (1-\epsilon)P_H(|\mathbf{x}_0^t\boldsymbol{\gamma}| > \alpha).
\end{aligned}$$

Esta última desigualdad vale para todo $\alpha > 0$ y $\epsilon > 0$. Entonces

$$\lim_{j\to\infty} E_{H_j}\left(\rho\left(\frac{y-\mathbf{x}^t\boldsymbol{\beta}_j}{\sigma_j}\right)\right) \geq P_H(|\mathbf{x}_0^t\boldsymbol{\gamma}| > 0) = 1 - P_H(\mathbf{x}_0^t\boldsymbol{\gamma} = 0) \geq 1 - c(H).$$

□

Lema 14.7. *Supongamos que $H_n \xrightarrow{D} H$, $S(H) > 0$ y $b < 1 - c(H)$ con $c(H) < 1$. Luego existe n_0 tal que $S(H_n) > 0$ para todo $n \geq n_0$.*

Demostración

Supongamos por el absurdo que existe una subsucesión $(H_{n_k})_{k\in\mathbb{N}}$ tal que $S(H_{n_k}) = 0$ para todo $k \in \mathbb{N}$. Veamos qué quiere decir $S(H^*) = 0$. Como $S(H^*) = \inf_{\boldsymbol{\beta}\in\mathbb{R}^p} S(H^*, \boldsymbol{\beta})$ entonces para todo $\epsilon > 0$ existe β_ϵ tal que $0 < S(H^*, \boldsymbol{\beta}_\epsilon) < \epsilon$. Sabemos que

$$E_{H^*}\left(\rho\left(\frac{y-\mathbf{x}^t\boldsymbol{\beta}_\epsilon}{S(H^*,\boldsymbol{\beta}_\epsilon)}\right)\right) = b$$

y por lo tanto

$$E_{H^*}\left(\rho\left(\frac{y-\mathbf{x}^t\boldsymbol{\beta}_\epsilon}{\epsilon}\right)\right) \leq b.$$

Luego para todo $k \in \mathbb{N}$ y $\epsilon > 0$, existe $\beta_{k,\epsilon}$ tal que $0 < S^*(H_{n_k}, \boldsymbol{\beta}_{k,\epsilon}) < \epsilon$ para el cual

$$E_{H_{n_k}}\left(\rho\left(\frac{y-\mathbf{x}^t\boldsymbol{\beta}_{k,\epsilon}}{\epsilon}\right)\right) \leq b.$$

Veamos que $(\boldsymbol{\beta}_{k,\epsilon})_{k\in\mathbb{N}}$ está acotada. Supongamos que no, es decir, existe una subsucesión $(\boldsymbol{\beta}_{k_j,\epsilon})_{j\in\mathbb{N}}$ tal que $||\boldsymbol{\beta}_{k_j,\epsilon}|| \to \infty$. Luego por el Lema 14.6

$$1 - c(H) \leq \liminf_{j\to\infty} E_{H_{n_{k_j}}}\left(\rho\left(\frac{y-\mathbf{x}^t\boldsymbol{\beta}_{k_j,\epsilon}}{\epsilon}\right)\right) \leq b$$

y esto es absurdo por hipótesis. Como $(\boldsymbol{\beta}_{k,\epsilon})_{k\in\mathbb{N}}$ está acotada posee una subsucesión convergente $\boldsymbol{\beta}_{k_j,\epsilon} \to \boldsymbol{\beta}^*_\epsilon$. Finalmente por el Lema 14.4 y pasando al límite la desigualdad

$$E_{H_{n_{k_j}}}\left(\rho\left(\frac{y-\mathbf{x}^t\boldsymbol{\beta}_{k_j,\epsilon}}{\epsilon}\right)\right) \leq b$$

se obtiene

$$E_H\left(\rho\left(\frac{y-\mathbf{x}^t\boldsymbol{\beta}^*_\epsilon}{\epsilon}\right)\right) \leq b = E_H\left(\rho\left(\frac{y-\mathbf{x}^t\boldsymbol{\beta}^*_\epsilon}{S(H,\boldsymbol{\beta}^*_\epsilon)}\right)\right).$$

Como estamos suponiendo que $S(H,\boldsymbol{\beta}^*_\epsilon)$ es la menor escala entre todas las que verifican

$$E_H\left(\rho\left(\frac{y-\mathbf{x}^t\boldsymbol{\beta}^*_\epsilon}{s}\right)\right) = b$$

se concluye que $S(H) \leq S(H,\boldsymbol{\beta}^*_\epsilon) \leq \epsilon$. Como esto vale para todo $\epsilon > 0$, se obtiene que $S(H) = 0$, arribando de este modo a una contradicción. □

Notar que el resultado anterior es independiente de la existencia de $T_S(H_n)$.

Definición 14.3. Decimos que un funcional T es débilmente continuo en H si cada vez que $H_n \xrightarrow{D} H$ entonces $T(H_n) \to T(H)$.

En esta definición estamos asumiendo implícitamente que existen $T(H)$ y $T(H_n)$ para cada $n \in \mathbb{N}$.

Lema 14.8. *Consideremos el funcional S débilmente continuo en H y supongamos que $T_S(H)$ existe y es único. Si $b < 1-c(H)$ con $c(H) < 1$ entonces T_S es débimente continuo en H.*

Demostración

Sea $H_n \xrightarrow{D} H$, queremos ver que $T_S(H_n) \longrightarrow T_S(H)$ (supondremos que $T_S(H_n)$ existe para cada $n \in \mathbb{N}$). Como S es débilmente continuo en H entonces $S(H_n) \longrightarrow S(H)$. En particular $(S(H_n))_{n\in\mathbb{N}}$ está acotada. Por el Lema 14.7 existe n_0 tal que $S(H_n) > 0$ para todo $n \geq n_0$.

Veamos primero que $(T_S(H_n))_{n\in\mathbb{N}}$ está acotada. Supongamos que $(T_S(H_n))_{n\in\mathbb{N}}$ no está acotada. Luego existe una subsucesión $(T_S(H_{n_k}))_{k\in\mathbb{N}}$ tal que $||T_S(H_{n_k})|| \to \infty$. Por el Lema 14.6 tenemos que

$$b = \liminf_{k\to\infty} E_{H_{n_k}}\left(\rho\left(\frac{y-\mathbf{x}^t T_S(H_{n_k})}{S(H_{n_k})}\right)\right) \geq 1 - c(H),$$

arribando de esta forma a una contradicción. Luego por el Lema 14.5 se obtiene el resultado. □

Teorema 14.1. *Supongamos que $T_S(H)$ existe y es único. Sea $b < 1-c(H)$ con $c(H) < 1$. Luego S y T_S son débilmente continuos en H.*

Demostración

Sea $H_n \xrightarrow{D} H$. Primero veamos que $(S(H_n))_{n\in\mathbb{N}}$ es una sucesión acotada. Sabemos que

$$E_H\left(\rho\left(\frac{y-\mathbf{x}^tT_S(H)}{S(H)}\right)\right)=b.$$

Como

$$\lim_{s\to+\infty} E_H\left(\rho\left(\frac{y-\mathbf{x}^tT_S(H)}{s}\right)\right)=0<b$$

entonces existe $\sigma^*>S(H)$ tal que

$$E_H\left(\rho\left(\frac{y-\mathbf{x}^tT_S(H)}{\sigma^*}\right)\right)<b.$$

Luego

$$\lim_{n\to\infty} E_{H_n}\left(\rho\left(\frac{y-\mathbf{x}^tT_S(H)}{\sigma^*}\right)\right)<b$$

y por lo tanto existe n_0 tal que si $n\geq n_0$

$$E_{H_n}\left(\rho\left(\frac{y-\mathbf{x}^tT_S(H)}{\sigma^*}\right)\right)<E_{H_n}\left(\rho\left(\frac{y-\mathbf{x}^tT_S(H)}{S(H_n,T_S(H))}\right)\right)=b,$$

de donde resulta $S(H_n,T_S(H))<\sigma^*$ para todo $n\geq n_0$, y por lo tanto $S(H_n)<\sigma^*$ para todo $n\geq n_0$. Observemos que $S(H)>0$ porque $T_S(H)$ existe y en particular por el Lema 14.7 existe n_1 tal que $S(H_n)>0$ para todo $n\geq n_1$. Probemos a continuación que $(T_S(H_n))_{n\in\mathbb{N}}$ está acotada. Supongamos que no, luego existe una subsucesión $(T_S(H_{n_k}))_{k\in\mathbb{N}}$ tal que $||T_S(H_{n_k})||\to\infty$. Por el Lema 14.6

$$b=\lim\inf_{k\to\infty} E_{H_{n_k}}\left(\rho\left(\frac{y-\mathbf{x}^tT_S(H_{n_k})}{S(H_{n_k})}\right)\right)\geq 1-c(H),$$

contradiciendo la hipótesis del teorema. Luego $(T_S(H_n))_{n\in\mathbb{N}}$ está acotada. Veamos que S es débilmente continua en H. Es decir probemos que $S(H_n)\to S(H)$. Si esto no ocurriese existiría una subsucesión $(H_{n_k})_{k\in\mathbb{N}}$ tal que $S(H_{n_k})\to\tilde{\sigma}\neq S(H)$. Como $(T_S(H_{n_k}))_{k\in\mathbb{N}}$ está acotada entonces sin pérdida de generalidad podremos suponer que $T_S(H_{n_k})\to\widetilde{\boldsymbol{\beta}}$. Primero veamos que $\tilde{\sigma}\neq 0$ y para ello supongamos que $\tilde{\sigma}=0$. Luego $\forall\ \epsilon>0$ tal que $\epsilon<S(H)$ $\exists\ n_{2,\epsilon}$ para el cual $S(H_{n_k})<\epsilon<S(H)$ $\forall n\geq n_{2,\epsilon}$ Sabemos que para cada $k\in\mathbb{N}$

$$E_{H_{n_k}}\left(\rho\left(\frac{y-\mathbf{x}^tT_S(H_{n_k})}{S(H_{n_k})}\right)\right)=b.$$

Luego $\forall n_k\geq n_{2,\epsilon}$

$$E_{H_{n_k}}\left(\rho\left(\frac{y-\mathbf{x}^tT_S(H_{n_k})}{\epsilon}\right)\right)\leq b$$

con lo cual tomando límite resulta

$$E_H\left(\rho\left(\frac{y-\mathbf{x}^t\widetilde{\boldsymbol{\beta}}}{\epsilon}\right)\right)\leq b=E_H\left(\rho\left(\frac{y-\mathbf{x}^t\widetilde{\boldsymbol{\beta}}}{S(H,\widetilde{\boldsymbol{\beta}})}\right)\right).$$

De allí $S(H) \leq S(H, \widetilde{\boldsymbol{\beta}}) \leq \epsilon$ (acá estamos usando que $S(H, \widetilde{\boldsymbol{\beta}})$ se define como la menor entre todas las soluciones de (77)) llegando así a una contradicción. Luego $\tilde{\sigma} \neq 0$. Por el Lema 14.4 tenemos

$$b = \lim_{k\to\infty} E_{H_{n_k}}\left(\rho\left(\frac{y - \mathbf{x}^t T_S(H_{n_k})}{S(H_{n_k})}\right)\right) = E_H\left(\rho\left(\frac{y - \mathbf{x}^t\widetilde{\boldsymbol{\beta}}}{\tilde{\sigma}}\right)\right)$$

de donde resulta $S(H, \tilde{\beta}) \leq \tilde{\sigma}$, con lo cual $S(H) \leq \tilde{\sigma}$. Por otro lado y por la definición de $T_S(H_{n_k})$ se tiene

$$b = E_{H_{n_k}}\left(\rho\left(\frac{y - \mathbf{x}^t T_S(H_{n_k})}{S(H_{n_k})}\right)\right) \leq E_{H_{n_k}}\left(\rho\left(\frac{y - \mathbf{x}^t T_S(H)}{S(H_{n_k})}\right)\right),$$

luego pasando al límite resulta

$$b = E_H\left(\rho\left(\frac{y - \mathbf{x}^t\widetilde{\boldsymbol{\beta}}}{\tilde{\sigma}}\right)\right) \leq E_H\left(\rho\left(\frac{y - \mathbf{x}^t T_S(H)}{\tilde{\sigma}}\right)\right).$$

Usando el hecho que

$$E_H\left(\rho\left(\frac{y - \mathbf{x}^t T_S(H)}{S(H)}\right)\right) = b$$

se obtiene $\tilde{\sigma} \leq S(H)$ y por lo tanto $\tilde{\sigma} = S(H)$ arribando así a una contradicción pues habíamos supuesto que eran distintos. Finalmente $\tilde{\sigma} = S(H)$. De esta forma hemos probado que S es débilmente continua. Como estamos bajo las hipótesis del Lema 14.8 se obtiene que T_S es débilmente continua. □

Corolario 14.1. *Sea $H_{\boldsymbol{\theta}}$ la función de distribución acumulada del vector aleatorio $(\mathbf{x}, y)$. Supongamos que (T_S, S) es Fisher consistente de $\boldsymbol{\theta}$ y que $b < 1 - c(H_{\boldsymbol{\theta}})$ con $c(H_{\boldsymbol{\theta}}) < 1$. Consideremos además H_n una sucesión de distribuciones aleatorias tal que $H_n \xrightarrow{D} H_{\boldsymbol{\theta}}$ casi seguramente. Luego la sucesión $(T_S(H_n), S(H_n))_{n\in\mathbb{N}}$ resulta fuertemente consistente de $\boldsymbol{\theta}$ siempre que $T_S(H_n)$ exista.*

Demostración

Como el par (T_S, S) es Fisher consistente de $\boldsymbol{\theta}$ se tiene que $T_S(H_{\boldsymbol{\theta}})$ es único y $(T_S(H_{\boldsymbol{\theta}}), S(H_{\boldsymbol{\theta}})) = \boldsymbol{\theta}$. Por el Teorema 14.1 T_S y S son funcionales débilmente continuos. Como $H_n \xrightarrow{D} H_{\boldsymbol{\theta}}$ casi seguramente entonces $T_S(H_n) \to T_S(H_{\boldsymbol{\theta}})$ y $S(H_n) \to S(H_{\boldsymbol{\theta}})$ casi seguramente. Finalmente $(T_S(H_n), S(H_n)) \to \boldsymbol{\theta}$ casi seguramente. □

Corolario 14.2. *Consideremos la muestra aleatoria $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)$ satisfaciendo el modelo (1) y sea H_n la distribución empírica de dicha muestra. Supongamos que la densidad de los errores f_0 es estrictamente unimodal y $b < 1 - c(H_{\boldsymbol{\beta}_0,\sigma_0})$ con $c(H_{\boldsymbol{\beta}_0,\sigma_0}) < 1$. Sea ρ tal que satisface **C1** a **C7** y $E_0(\rho(u)) = b$. Luego las sucesiones aleatorias $T_S(H_n)$ y $S(H_n)$ resultan fuertemente consistentes de $\boldsymbol{\beta}_0$ y σ_0 siempre que $T_S(H_n)$ exista.*

Demostración

Sabemos que $H_n \xrightarrow{D} H_{\boldsymbol{\beta}_0,\sigma_0}$ casi seguramente. La condición de identificabilidad **CI** está garantizada a partir de $c(H_{\boldsymbol{\beta}_0,\sigma_0}) < 1$. Por el Lema 5.12 existe para cada $\boldsymbol{\beta}$, el funcional de escala residual $S(H_{\boldsymbol{\beta}_0,\sigma_0}, \boldsymbol{\beta}) = s(\boldsymbol{\beta})$. Luego por el Teorema 14.2 se tiene que el par (T_S, S) es Fisher consistente de $(\boldsymbol{\beta}_0, \sigma_0)$. Finalmente, a partir del Corolario 14.1 obtenemos $T_S(H_n) \to \boldsymbol{\beta}_0$ y $S(H_n) \to \sigma_0$ casi seguramente. □

Demostración del Teorema 5.2

Es una consecuencia inmediata del resultado anterior ya que el S-estimador $\tilde{\boldsymbol{\beta}}_n$ es también un M estimador y además $T_S(H_n) = \tilde{\boldsymbol{\beta}}_n$ y $S(H_n) = \tilde{\sigma}_n$. □

Formulación alternativa para los S-estimadores de regresión

Demostración del Lema 5.15

Veamos que $\widetilde{\sigma}_n^* = \widetilde{\sigma}_n$ y para eso hagamos uso de la ecuación (48). Como $\widetilde{\beta}_n$ minimiza $s_n(\beta)$ obtenemos

$$\widetilde{\sigma}_n = s_n(\widetilde{\beta}_n) \leq s_n(\widetilde{\beta}_n^* + \mu_0\widetilde{\sigma}_n^*) = s_n(\widetilde{\beta}_n^* + \mu_0 s_n^*(\widetilde{\beta}_n)) = s_n^*(\widetilde{\beta}_n^*) = \widetilde{\sigma}_n^*.$$

Por otro lado $\widetilde{\beta}_n^*$ minimiza la escala $s_n^*(\beta)$ entonces

$$\widetilde{\sigma}_n^* = s_n^*(\widetilde{\beta}_n^*) \leq s_n^*(\widetilde{\beta}_n - \mu_0\widetilde{\sigma}_n) = s_n(\widetilde{\beta}_n - \mu_0\widetilde{\sigma}_n + \mu_0 s_n^*(\widetilde{\beta}_n - \mu_0\widetilde{\sigma}_n)).$$

Como

$$\frac{1}{n}\sum_{i=1}^{n} \rho\left(\frac{y_i - \widetilde{\beta}_n + \mu_0\widetilde{\sigma}_n}{\widetilde{\sigma}_n} - \mu_0\right) = \frac{1}{n}\sum_{i=1}^{n} \rho\left(\frac{y_i - \widetilde{\beta}_n}{\widetilde{\sigma}_n}\right) = b,$$

entonces por la unicidad de $s_n^*(\beta)$ resulta $s_n^*(\widetilde{\beta}_n - \mu_0\widetilde{\sigma}_n) = \widetilde{\sigma}_n$. Luego

$$\widetilde{\sigma}_n^* \leq s_n(\widetilde{\beta}_n - \mu_0\widetilde{\sigma}_n + \mu_0\widetilde{\sigma}_n) = s_n(\widetilde{\beta}_n) = \widetilde{\sigma}_n$$

con lo cual $\widetilde{\sigma}_n = \widetilde{\sigma}_n^*$.

Como $s_n(\widetilde{\beta}_n^* + \mu_0\widetilde{\sigma}_n^*) = \widetilde{\sigma}_n = \widetilde{\sigma}_n^* = s_n^*(\widetilde{\beta}_n - \mu_0\widetilde{\sigma}_n)$ entonces $\widetilde{\beta}_n^* + \mu_0\widetilde{\sigma}_n^*$ y $\widetilde{\beta}_n - \mu_0\widetilde{\sigma}_n$ resultan S-estimadores. □

15. Apéndice de la Sección 6

S estimadores de regresión para datos censurados propuestos por Locatelli, Marazzi y Yohai

Presentaremos un argumento heurístico para mostrar que $\overline{\boldsymbol{\beta}}_n^* \to \boldsymbol{\beta}_0$ adaptando el marco teórico de las secciones anteriores a las sucesiones

$$(T_S(H_{n,\boldsymbol{\beta}_0,s_n(\boldsymbol{\beta}_0)}))_{n\in\mathbb{N}} \text{ y } (S(H_{n,\boldsymbol{\beta}_0,s_n(\boldsymbol{\beta}_0)}))_{n\in\mathbb{N}}.$$

Consideremos a β_0 y σ_0 los parámetros verdaderos del modelo, luego

$$\lim_{n\to\infty} E_{n,\boldsymbol{\beta}_0,\sigma_0}\left(\rho\left(\frac{y-\mathbf{x}^t\boldsymbol{\beta}_0}{\sigma_0}\right)\right) = E_{\boldsymbol{\beta}_0,\sigma_0}\left(\rho\left(\frac{y-\mathbf{x}^t\boldsymbol{\beta}_0}{\sigma_0}\right)\right) = b.$$

Es decir para n suficientemente grande es de esperar que

$$E_{n,\boldsymbol{\beta}_0,\sigma_0}\left(\rho\left(\frac{y-\mathbf{x}^t\boldsymbol{\beta}_0}{\sigma_0}\right)\right) \cong b.$$

Por otro lado, por definición, la escala $s_n(\beta_0)$ es tal que

$$E_{n,\boldsymbol{\beta}_0,s_n(\boldsymbol{\beta}_0)}\left(\rho\left(\frac{y-\mathbf{x}^t\boldsymbol{\beta}_0}{s_n(\boldsymbol{\beta}_0)}\right)\right) = b$$

lo cual sugiere que $s_n(\boldsymbol{\beta}_0) \to s(\boldsymbol{\beta}_0) = \sigma_0$.

Luego estaríamos en condiciones de justificar heurísticamente que, casi seguramente

$$H_{n,\boldsymbol{\beta}_0,s_n(\boldsymbol{\beta}_0)} \to H_{\boldsymbol{\beta}_0,\sigma_0}.$$

$H_{n,\boldsymbol{\beta}_0,\sigma_0}$ es un estimador consistente de $H_{\boldsymbol{\beta}_0,\sigma_0}$, por la Observación 3.2. Si aceptamos que $s_n(\boldsymbol{\beta}_0) \to \sigma_0$, entonces $H_{n,\boldsymbol{\beta}_0,s_n(\boldsymbol{\beta}_0)} - H_{n,\boldsymbol{\beta}_0,\sigma_0} \to 0$ casi seguramente.

Consideremos

$$H_{n,\boldsymbol{\beta}_0,s_n(\boldsymbol{\beta}_0)} - H_{\boldsymbol{\beta}_0,\sigma_0} = (H_{n,\boldsymbol{\beta}_0,s_n(\boldsymbol{\beta}_0)} - H_{n,\boldsymbol{\beta}_0,\sigma_0}) + (H_{n,\boldsymbol{\beta}_0,\sigma_0} - H_{\boldsymbol{\beta}_0,\sigma_0}).$$

Observar que los dos paréntesis tienden a 0 en casi todo punto y por lo tanto bajo supuestos adecuados, se puede pensar que

$$T_S(H_{n,\boldsymbol{\beta}_0,s_n(\boldsymbol{\beta}_0)}) \to T_S(H_{\beta_0,\sigma_0}),$$

es decir $\overline{\boldsymbol{\beta}}_n^* \to \beta_0$.

Observación 15.1. Sea $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)$ una muestra aleatoria satisfaciendo el modelo (1) y $(\mathbf{x}_1, y_1^*, \delta_1), \ldots, (\mathbf{x}_n, y_n^*, \delta_n)$ la correpondiente muestra aleatoria de datos censurados. Supongamos que f_0 es estrictamente unimodal, ρ satisface las condiciones **C1** a **C7** y $b < 1 - c(H_{\boldsymbol{\beta}_0,\sigma_0})$ con $c(H_{\boldsymbol{\beta}_0,\sigma_0}) < 1$. Se observa a partir del Corolario 14.1, que los estimadores $\overline{\boldsymbol{\beta}}_n^* = T_S(H_{n,\boldsymbol{\beta}_0,s_n(\boldsymbol{\beta}_0)})$ y $s_n^*(\overline{\boldsymbol{\beta}}_n^*) = S(H_{n,\boldsymbol{\beta}_0,s_n(\boldsymbol{\beta}_0)})$ resultan fuertemente consistentes, es decir $\overline{\boldsymbol{\beta}}_n^* \to \boldsymbol{\beta}_0$ y $s_n^*(\overline{\boldsymbol{\beta}}_n^*) \to \sigma_0$ casi seguramente.

Demostración del Lema 6.1

Reemplazando la ecuación (51) por $\boldsymbol{\gamma} = \widehat{\boldsymbol{\beta}}_n$ y $s_n(\widehat{\boldsymbol{\beta}}_n) = \widehat{\sigma}_n$ se concluye que los estimadores verifican la ecuación (58).

Para verificar (57) debemos derivar la ecuación (54) con respecto a γ_i,

$$E_{n,\boldsymbol{\beta},s_n(\boldsymbol{\beta})}\left(\psi\left(\frac{y-\mathbf{x}^t\boldsymbol{\gamma}}{S_n(\boldsymbol{\beta},\boldsymbol{\gamma})}\right)\frac{-x_i S_n(\boldsymbol{\beta},\boldsymbol{\gamma}) - (y-\mathbf{x}^t\gamma)\frac{\partial}{\partial\gamma_i}S_n(\boldsymbol{\beta},\gamma)}{S_n^2(\boldsymbol{\beta},\boldsymbol{\gamma})}\right) = 0.$$

Luego evaluamos en $\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}}_n$ y $\boldsymbol{\gamma} = \widetilde{\boldsymbol{\gamma}}_n(\widehat{\boldsymbol{\beta}}_n) = \widehat{\boldsymbol{\beta}}_n$ pues de esta forma valen las siguientes igualdades $S_n(\widehat{\boldsymbol{\beta}}_n, \widehat{\boldsymbol{\beta}}_n) = s_n(\widehat{\boldsymbol{\beta}}_n) = \widehat{\sigma}_n$. Como $\widehat{\boldsymbol{\beta}}_n = \arg\min_{\boldsymbol{\gamma}} S_n(\widehat{\boldsymbol{\beta}}_n, \boldsymbol{\gamma})$ entonces $\frac{\partial}{\partial \gamma_i} S_n(\widehat{\boldsymbol{\beta}}_n, \gamma)\Big|_{\boldsymbol{\gamma}=\widehat{\boldsymbol{\beta}}_n} = 0$ y de esta forma se verifica (57). □

16. Apéndice de la Sección 7

Nueva propuesta de S-estimadores de regresión con datos censurados

Demostración del Lema 7.1

Para ello bastaría ver que cada término de la suma finita está acotado. Recordemos que

$$L^*_{\widehat{\sigma}_n}(\beta) = \frac{1}{n}\sum_{i=1}^{n}\left(\delta_i \rho\left(\frac{y_i^* - \beta}{\widehat{\sigma}_n}\right) + (1-\delta_i)\int_0^{\frac{y_i^*-\beta}{\widehat{\sigma}_n}} E_0\left(\psi(u_i)|u_i > s\right) ds\right).$$

Claramente los términos no censurados están acotados. Para los términos censurados debemos ver que, para cada i,

$$\int_0^{\frac{y_i^*-\beta}{\widehat{\sigma}_n}} E_0\left(\psi(u_i)|u_i > s\right) ds$$

está acotado como función de β. Consideremos la función

$$g(s) = E_0\left(\psi(u_i)|u_i > s\right) = \frac{\displaystyle\int_s^{+\infty} \psi(u_i) f_0(u_i) du_i}{1 - F_0(s)}.$$

Por definición de ψ resulta que $g(s) = 0\ \forall s \geq k$, donde k es el parámetro de ajuste de ψ. Por otro lado como estamos suponiendo $\mu_0 = 0$ (el lema sigue valiendo si $\mu_0 \neq 0$),

$$0 = \arg\min_{\mu} E_0\left(\rho\left(u - \mu\right)\right).$$

Llamemos $q(\mu) = E_0\left(\rho\left(u-\mu\right)\right)$. Luego

$$q'(0) = -\int_{-\infty}^{\infty} \psi(u) f_0(u) du = 0.$$

Como $\psi(u)$ se anula fuera del compacto $\{u/|u| \leq k\}$, entonces

$$\int_{-k}^{k} \psi(u) f_0(u) du = 0.$$

En consecuencia $g(s) = 0\ \ \forall s$ tal que $|s| \geq k$.

Como $\psi(u)f_0(u) \geq 0 \ \forall u \geq 0, \quad \psi(u)f_0(u) \leq 0 \ \forall u \leq 0$ y usando el hecho que $\int_{-\infty}^{\infty} \psi(u)f_0(u)du = 0$ se deduce que

$$\int_{s}^{\infty} \psi(u)f_0(u)du \geq 0.$$

De esta forma queda probado también que $g(s)$ es una función no negativa y que

$$\int_{0}^{\frac{y_i^* - \beta}{\tilde{\sigma}_n}} g(s)ds$$

es monótona no creciente y acotada como función de β. □

Demostración del Lema 7.2

Queremos ver que

$$E\left(\delta\rho\left(\frac{y^* - \mathbf{x}^t\boldsymbol{\beta}_0}{\sigma_0}\right) + (1-\delta)E_{\boldsymbol{\beta}_0,\sigma_0}\left(\rho\left(u\right)\middle| u > \frac{y^* - \mathbf{x}^t\boldsymbol{\beta}_0}{\sigma_0}\right)\right) = b.$$

Sea

$$G(\mathbf{x}, y^*, \delta) = \delta\rho\left(\frac{y^* - \mathbf{x}^t\boldsymbol{\beta}_0}{\sigma_0}\right) + (1-\delta)E_{\boldsymbol{\beta}_0,\sigma_0}\left(\rho\left(u\right)\middle| u > \frac{y^* - \mathbf{x}^t\boldsymbol{\beta}_0}{\sigma_0}\right).$$

Observemos que $E(G(\mathbf{x}, y^*, \delta)|\delta = 1) = E(\rho(u)|\delta = 1)$. Por otro lado, bajo el modelo (1) y haciendo uso de las siguientes equivalencias

$$u > \frac{y^* - \mathbf{x}^t\boldsymbol{\beta}_0}{\sigma_0} \Leftrightarrow y > y^* \Leftrightarrow \delta = 0$$

se tiene

$$\begin{aligned} E(G(\mathbf{x}, y^*, \delta)|\delta = 0) =& E\left(E_{\boldsymbol{\beta}_0,\sigma_0}\left(\rho\left(u\right)\middle| u > \frac{y^* - \mathbf{x}^t\boldsymbol{\beta}_0}{\sigma_0}\right)\middle|\, \delta = 0\right) \\ =& E\left(E\left(\rho\left(u\right)|\delta = 0\right)|\, \delta = 0\right) \\ =& E(\rho(u)|\delta = 0). \end{aligned}$$

Luego

$$E(G(\mathbf{x}, y^*, \delta)|\delta) = \delta E(\rho(u)|\delta = 1) + (1-\delta)E(\rho(u)|\delta = 0) = E(\rho(u)|\delta)$$

y finalmente

$$E(G(\mathbf{x}, y^*, \delta) = E(E(G(\mathbf{x}, y^*, \delta)|\delta)) = E(E(\rho(u)|\delta)) = E(\rho(u)) = b.$$

□

Demostración del Teorema 7.1

Notar que la primera ecuación es equivalente a

$$\beta_0 = \arg\min_{\beta\in\mathbb{R}} E\left(\delta\rho\left(\frac{y^*-\beta}{\sigma_0}\right) + (1-\delta)\int_{\frac{y^*}{\sigma_0}}^{\frac{y^*-\beta}{\sigma_0}} E_0\left(\psi(u)|u>s\right)ds\right)$$

porque las funciones a minimizar difieren en una constante que no depende de β. Consideremos para cada β fijo,

$$H_\beta(y^*,\delta) = \delta\rho\left(\frac{y^*-\beta}{\sigma_0}\right) + (1-\delta)\int_{\frac{y^*}{\sigma_0}}^{\frac{y^*-\beta}{\sigma_0}} E_0\left(\psi(u)|u>s\right)ds.$$

Sabemos que $E(E(H_\beta(y^*,\delta)|\delta)) = E(H_\beta(y^*,\delta))$, luego comenzaremos obteniendo una expresión para $E(H_\beta(y^*,\delta)|\delta)$. De (61) se obtiene

$$\begin{aligned}
E(H_\beta(y^*,\delta)|\delta=1) &= E\left(\rho\left(\frac{y^*-\beta}{\sigma_0}\right)\middle|\,\delta=1\right)\\
&= E\left(\rho\left(\frac{y-\beta}{\sigma_0}\right)\middle|\,\delta=1\right),\\
E(H_\beta(y^*,\delta)|\delta=0) &= E\left(\int_{\frac{y^*}{\sigma_0}}^{\frac{y^*-\beta}{\sigma_0}} E_0\left(\psi(u)|u>s\right)ds|\delta=0\right)\\
&= -\frac{1}{\sigma_0}E\left(\int_0^\beta E_{t,\sigma_0}\left(\psi\left(\frac{y-t}{\sigma_0}\right)\middle|\,y^*, y>y^*\right)dt\middle|\,\delta=0\right).\\
&= -\frac{1}{\sigma_0}E\left(\int_0^\beta E_{t,\sigma_0}\left(\psi\left(\frac{y-t}{\sigma_0}\right)\middle|\,y^*, \delta=0\right)dt\middle|\,\delta=0\right).
\end{aligned}$$

La función $g(s) = E_0\left(\psi(u)|u>s\right)$ es continua. Además por la demostración del Lema 7.1 $g(s)$ es no negativa y se anula fuera de un compacto. Luego

$$\int_{\frac{y^*}{\sigma_0}}^{\frac{y^*-\beta}{\sigma_0}} E_0\left(\psi(u)|u>s\right)ds$$

está acotada como función de y^* y por lo tanto, tiene la esperanza condicional a $\delta=0$, finita.

Notemos

$$h(t,y^*) = E_{t,\sigma_0}\left(\psi\left(\frac{y-t}{\sigma_0}\right)\middle|\,y^*,\delta=0\right),$$

que resulta continua y acotada debido a que

$$h(t,y^*) = g\left(\frac{y^*-t}{\sigma_0}\right).$$

$$\begin{aligned}
E(H_\beta(y^*,\delta)|\delta=0) &= -\frac{1}{\sigma_0}E\left(\int_0^\beta h(t,y^*)dt|\delta=0\right)\\
&= -\frac{1}{\sigma_0}\int_{-\infty}^{+\infty}\int_0^\beta h(t,y^*)\,dt\,dF_{y^*|\delta=0}(y^*).
\end{aligned}$$

Como

$$\int_{-\infty}^{+\infty} h(t, y^*)\, dF_{y^*|\delta=0}(y^*) = E(h(t, y^*)|\delta = 0)$$

existe para cada t fijo, luego aplicando el Teorema de Fubini para integrales de Riemann-Stieltjes y del hecho que dadas las variables aleatorias X, Y y Z vale $E\left(E(X|Y, Z)|Y\right) = E(X|Y)$, obtenemos

$$\begin{aligned}
E(H_\beta(y^*, \delta)|\delta = 0) &= -\frac{1}{\sigma_0}\int_0^\beta \int_{-\infty}^{+\infty} h(t, y^*)\, dF_{y^*|\delta=0}(y^*)\, dt \\
&= -\frac{1}{\sigma_0}\int_0^\beta E\left(E_{t,\sigma_0}\left(\psi\left(\frac{y-t}{\sigma_0}\right)\middle| y^*, \delta = 0\right)\middle| \delta = 0\right) dt \\
&= -\frac{1}{\sigma_0}\int_0^\beta E\left(\psi\left(\frac{y-t}{\sigma_0}\right)\middle| \delta = 0\right) dt \\
&= -\frac{1}{\sigma_0} E\left(\int_0^\beta \psi\left(\frac{y-t}{\sigma_0}\right) dt\middle| \delta = 0\right) \\
&= E\left(\rho\left(\frac{y-\beta}{\sigma_0}\right) - \rho\left(\frac{y}{\sigma_0}\right)\middle| \delta = 0\right).
\end{aligned}$$

Finalmente

$$\begin{aligned}
&E(H_\beta(y^*, \delta)|\delta) = \\
&= \delta E\left(\rho\left(\frac{y-\beta}{\sigma_0}\right)\middle| \delta = 1\right) + (1-\delta)E\left(\rho\left(\frac{y-\beta}{\sigma_0}\right) - \rho\left(\frac{y}{\sigma_0}\right)\middle| \delta = 0\right) \\
&= E\left(\delta\rho\left(\frac{y-\beta}{\sigma_0}\right) + (1-\delta)\left(\rho\left(\frac{y-\beta}{\sigma_0}\right) - \rho\left(\frac{y}{\sigma_0}\right)\right)\middle| \delta\right).
\end{aligned}$$

Tomando esperanza en la ecuación anterior resulta

$$E(H_\beta(y^*, \delta)) = E\left(\delta\rho\left(\frac{y-\beta}{\sigma_0}\right) + (1-\delta)\left(\rho\left(\frac{y-\beta}{\sigma_0}\right) - \rho\left(\frac{y}{\sigma_0}\right)\right)\right).$$

Como $E\left(\rho\left(\frac{y}{\sigma_0}\right)\right)$ no depende de β entonces la ecuación para β_0 es equivalente a

$$\begin{aligned}
\beta_0 &= \arg\min_{\beta\in\mathbb{R}} E\left(\delta\rho\left(\frac{y-\beta}{\sigma_0}\right) + (1-\delta)\rho\left(\frac{y-\beta}{\sigma_0}\right)\right) \\
&= \arg\min_{\beta\in\mathbb{R}} E\left(\rho\left(\frac{y-\beta}{\sigma_0}\right)\right).
\end{aligned}$$

Luego por el Teorema 5.1 sabemos que en β_0 se alcanza el mínimo y es único. □

Demostración del Corolario 7.1

Análogamente al caso de posición y escala definamos para cada β

$$H_{\boldsymbol{\beta}}(\mathbf{x}, y^*, \delta) = \left(\delta\rho\left(\frac{y^* - \mathbf{x}^t\boldsymbol{\beta}}{\sigma_0}\right) + (1-\delta)\int_0^{\frac{y^*-\mathbf{x}^t\boldsymbol{\beta}}{\sigma_0}} E_0\left(\psi\left(u\right)|u > s\right)\right).$$

Analicemos $E\left(H_{\boldsymbol{\beta}}(\mathbf{x}, y^*, \delta)|\mathbf{x}\right)$ para cada valor $\tilde{\mathbf{x}}$ del rango de $\mathbf{x}$ y notemos $\beta_{\tilde{\mathbf{x}}} = \tilde{\mathbf{x}}^t\boldsymbol{\beta}$, luego $E\left(H_{\boldsymbol{\beta}}(\mathbf{x}, y^*, \delta)|\mathbf{x} = \tilde{\mathbf{x}}\right) =$

$$= E\left(\delta\rho\left(\frac{y^* - \boldsymbol{\beta}_{\tilde{\mathbf{x}}}}{\sigma_0}\right) + (1-\delta)\int_0^{\frac{y^*-\boldsymbol{\beta}_{\tilde{\mathbf{x}}}}{\sigma_0}} E_0\left(\psi\left(u\right)|u > s\right)|\mathbf{x} = \tilde{\mathbf{x}}\right).$$

Por lo hecho en la demostración para el caso de posición y escala se tiene

$E\left(H_{\boldsymbol{\beta}}(\mathbf{x}, y^*, \delta)|\mathbf{x} = \tilde{\mathbf{x}}\right) =$

$$\begin{aligned}
&= E\left(\delta\rho\left(\frac{y - \boldsymbol{\beta}_{\tilde{\mathbf{x}}}}{\sigma_0}\right) + (1-\delta)\left(\rho\left(\frac{y - \boldsymbol{\beta}_{\tilde{\mathbf{x}}}}{\sigma_0}\right) - \rho\left(\frac{y}{\sigma_0}\right)\right)\middle|\, \mathbf{x} = \tilde{\mathbf{x}}\right)\\
&= E\left(\delta\rho\left(\frac{y - \boldsymbol{\beta}_{\tilde{\mathbf{x}}}}{\sigma_0}\right) + (1-\delta)\rho\left(\frac{y - \boldsymbol{\beta}_{\tilde{\mathbf{x}}}}{\sigma_0}\right) + K(\tilde{\mathbf{x}}, y^*, \delta)\middle|\, \mathbf{x} = \tilde{\mathbf{x}}\right)\\
&= E\left(\delta\rho\left(\frac{y - \boldsymbol{\beta}_{\tilde{\mathbf{x}}}}{\sigma_0}\right) + (1-\delta)\rho\left(\frac{y - \boldsymbol{\beta}_{\tilde{\mathbf{x}}}}{\sigma_0}\right)\middle|\, \mathbf{x} = \tilde{\mathbf{x}}\right) + C(x)\\
&= E\left(\rho\left(\frac{y - \boldsymbol{\beta}_{\tilde{\mathbf{x}}}}{\sigma_0}\right)\middle|\, \mathbf{x} = \tilde{\mathbf{x}}\right) + C(\tilde{\mathbf{x}}).
\end{aligned}$$

De esta forma se obtiene

$$E(H_{\boldsymbol{\beta}}(\mathbf{x}, y^*, \delta)|\mathbf{x})) = E\left(\rho\left(\frac{y - \mathbf{x}^t\boldsymbol{\beta}}{\sigma_0}\right)\middle|\, \mathbf{x}\right) + C(\mathbf{x}),$$

y tomando esperanza resulta

$$E\left(H_{\boldsymbol{\beta}}(\mathbf{x}, y^*, \delta)\right) = E\left(\rho\left(\frac{y - \mathbf{x}^t\boldsymbol{\beta}}{\sigma_0}\right)\right) + E(C(\mathbf{x})).$$

Como $E(C(\mathbf{x}))$ no depende de $\boldsymbol{\beta}$ entonces la ecuación para $\boldsymbol{\beta}_0$ es equivalente a

$$\boldsymbol{\beta}_0 = \arg\min_{\beta\in\mathbb{R}} E\left(\rho\left(\frac{y - \mathbf{x}^t\boldsymbol{\beta}}{\sigma_0}\right)\right).$$

Luego por el Teorema 5.1 sabemos que en $\boldsymbol{\beta}_0$ se alcanza el mínimo y es único. □

17. Apéndice de la Sección 9

Resultados numéricos

En lo que sigue desarrollaremos el entorno para el cual haremos las simulaciones.

Supongamos primero que el término del error u_i tiene densidad f_0 correspondiente a una distribución normal estándar. Luego y_i tiene distribución $N(1,1)$. Las variables que censuran v_i tendrán distribución $N(\mu_1, 1)$. La media μ_1 la elegimos de tal forma que la probabilidad de censura esté en torno al 35 % es decir $P(\delta_i = 0) = P(y_i > v_i) = 0{,}35$. De esta forma el valor de $\mu_1 = 1{,}5445$. Como f_0 es una densidad estrictamente decreciente

en $|u|$ y simétrica entonces $\mu_0 = 0$. La función $\rho = \rho_k$ fue elegida dentro de la familia bicuadrada de Tukey verificando $E_0(\rho(u)) = 0{,}5$. Luego $k = 1{,}548$ con $s_0 = 1$.

Cuando el término del error u_i tiene densidad f_0 correspondiente a una distribución log-Weibull estándar, y_i tiene distribución log-Weibull con parámetros $\mu = 1$ y $\sigma = 1$. Las variables que censuran v_i tendrán distribución log-Weibull con parámetros $\mu = \mu_2$ y $\sigma = 1$. Del mismo modo que para el caso gaussiano, el parámetro μ_2 lo elegimos de tal forma que la probabilidad de censura esté en torno al 35 % es decir $P(\delta_i = 0) = P(y_i > v_i) = 0{,}35$. Veamos cómo podemos calcular el valor del parámetro μ_2.

Recordar que la densidad de una log-Weibull está dada por

$$f_{\mu,\sigma}(z) = \frac{1}{\sigma}\, e^{\dfrac{z-\mu}{\sigma} \;-\; e^{\dfrac{z-\mu}{\sigma}}}, \quad z \in \mathbb{R}.$$

luego es fácil verificar que la la función de distribución acumulada es

$$F_{\mu,\sigma}(z) = 1 - e^{-e^{\dfrac{z-\mu}{\sigma}}}.$$

En virtud de la independencia de y_i con v_i se tiene

$$\begin{aligned} P(y > v) &= \int_{-\infty}^{+\infty} f_{\mu,1}(v)(1 - F_{1,1})(v)dv \\ &= \int_{-\infty}^{+\infty} \int_{v}^{+\infty} f_{\mu,1}(v) f_{1,1}(y) dy dv \\ &= \int_{-\infty}^{+\infty} e^{v-\mu} - e^{v-\mu} - e^{v-1} dv. \end{aligned}$$

Renombrando $a = e^{-\mu}$ y $b = e^{-1}$ obtenemos

$$P(y > v) = \int_{-\infty}^{+\infty} a e^{v-}\, e^{v}(a+b)dv = -\frac{a}{a+b} e^{-e^{v}(a+b)}\Big|_{-\infty}^{+\infty} = \frac{a}{a+b} = 0{,}35.$$

Luego despejando resulta $\mu_2 = 1{,}619$.

La función $\rho = \rho_k$ (de la familia bicuadrada de Tukey) y el valor μ_0 fueron elegidos como para que $\mu_0 = \arg\min_\mu E_0\left(\rho\left(u - \mu\right)\right)$, $E_0(\rho(u - \mu_0)) = 0{,}5$ con $s_0 = 1$. De esta forma obtuvimos $k = 1{,}718$ y $\mu_0 = -0{,}135$.

Algoritmo para el cálculo de los S-estimadores alternativos

A continuación especificaremos los puntos iniciales para los Pasos 2 y 3 del algoritmo de búscqueda de los S-estimadores alternativos $\tilde{\boldsymbol{\beta}}_n$ y $\tilde{\sigma}_n$. Recordemos que una alternativa robusta al desvío estándar es la mediana (normalizada) del valor absoluto de los desvíos respecto de la mediana (MAD) definida como

$$\mathrm{MAD}(y) = \frac{\mathrm{Med}|y - \mathrm{Med}(y)|}{a},$$

donde a es elegido según el modelo, de tal forma que si la variable aleatoria y tiene distribución N(μ,σ^2) o tiene distribución log-Weibull con parámetros μ y σ entonces MAD(y) = σ. En el primer caso para hallar el valor de a debemos resolver la siguiente ecuación

$$P\left(\frac{|y - \text{Med}(y)|}{a} \leq \sigma\right) = P\left(\frac{|y-\mu|}{\sigma} \leq a\right) = \frac{1}{2}.$$

Luego $a = 0{,}6748$.

Para la distribución log-Weibull procedemos de manera similar. A partir de la función de distribución acumulada despejamos el valor de la mediana Med(y)= $\mu + \sigma \ln(\ln 2)$ y hallamos a tal que

$$P\left(\frac{|y - \mu - \sigma \ln(\ln 2)|}{a} \leq \sigma\right) = P\left(-a + \ln(\ln 2) \leq u \leq a + \ln(\ln 2)\right) = \frac{1}{2},$$

siendo u una variable aleatoria con distribución log-Weibull de parámetros 0 y 1. De esta forma resulta $a = 0{,}7671$.

La versión empírica de la MAD(y) la notaremos MAD($y_1, \ldots, y_n$) y está dada por

$$\text{MAD}(y_1, \ldots, y_n) = \frac{\text{Med}\{|y_1 - \tilde{\mu}_n|, \ldots, |y_n - \tilde{\mu}_n|\}}{a}, \tag{82}$$

donde $\tilde{\mu}_n = \text{Med}\{y_1, \ldots, y_n\}$.

Como $\tilde{\mu}_n \to \text{Med}(y)$ casi seguramente entonces para las distribuciones en cuestión MAD($y_1, \ldots, y_n$) $\to \sigma$ casi seguramente.

Para el paso 3 tomamos como punto inicial MAD($y_1^*, \ldots, y_n^*$), es decir la MAD aplicado a la muestra de datos censurados.

En el Paso 2 procedemos de la siguiente manera: para cada $\sigma > 0$ fijo tenemos por el Paso 1, como modelo verdadero $y = \beta(\sigma) + \sigma u$, luego Med($y$) es igual a $\beta(\sigma)$ en el caso normal y $\beta(\sigma) + \sigma \ln(\ln 2)$ en el caso log-Weibull. Como conocemos el valor de la mediana (teórica), preferimos usarla en la ecuación (82), en lugar de $\tilde{\mu}_n$. Luego los puntos iniciales que tomamos en el modelo normal y en el modelo log-Weibull son respectivamente

$$\begin{aligned}
\text{MAD}_\sigma(y_1^*, \ldots, y_n^*) =& \frac{\text{Med}\{|y_1^* - \beta(\sigma)|, \ldots, |y_n - \beta(\sigma)|\}}{0{,}6748} \\
\text{MAD}_\sigma(y_1^*, \ldots, y_n^*) =& \frac{\text{Med}\{|y_1^* - \beta(\sigma) - \sigma \ln(\ln 2)|, \ldots, |y_n - \beta(\sigma) - \sigma \ln(\ln 2)|\}}{0{,}7671}.
\end{aligned}$$

Vale aclarar que computacionalmente no encontramos diferencias al hallar los S-estimadores alternativos cuando usamos como puntos iniciales los propuestos recién o simplemente MAD($y_1^*, \ldots, y_n^*$). Sin embargo el tiempo de cálculo es significativamente inferior usando este último punto inicial, lo cual parece razonable.

Referencias

[1] Fasano, M., Maronna, R., Sued, M., Yohai, V., (2012). "Continuity and differentiability of regression M functionals", *Bernoulli*, **18**, No. 4, 1284-1309. Disponible en `http://arxiv.org/pdf/1004.4314.pdf`.

[2] Freireich, E. J., Gehan, E., Frei E., Schroeder L.R., Wolman J.I., Lee S., Moon J.R., Gendel B.R., Spurr Ch.L., Storrs R., Haurani F., Hoogstraten H., Anbari R., Burgert O., Mills S.D., Pinkel D., Selawry O.S., (1963). "The Effect of 6-Mercaptopurine on the Duration of Steroid-induced Remissions in Acute Leukemia: A Model for Evaluation of Other Potentially Useful Therapy", *Proc. Am. 4. Assoc. Cancer Res*, **21**, No. 6, 699-716.

[3] Gervini D., Yohai V., (2002). "A class of robust and fully efficient regression estimators", *The Annals of Statistics*, **30**, No. 2, 583-616.

[4] Huber, P.J., (1981). "Robust Statistics", *Wiley*, New York.

[5] Ibragimov, I. A. (1956). "On the Composition of Unimodal Distributions", *Theory of Probability and its Applications*, **1**, 255-260.

[6] Kalbfleisch J.D., Prentice R.L., (1980). "The Statistical Analysis of Failure Time Data", *John Wiley & Sons*.

[7] Klein J. P., Moeschberger M.L., (1997). "Survival Analysis: Techniques for Censored and Truncated Data", segunda edición, *Springer*.

[8] Lawless, J.F., (2003). "Statisticals Models and Methods for Lifetime Data", tercera edidión, *Wiley*, New York.

[9] Locatelli I., Marazzi A., Yohai V., (2010). "Robust accelerated failure time regression", *Computational Statistics and Data Analysis* **55**, 874-887, journal homepage: `www.elsevier.com/locate/csda`.

[10] Marazzi, A., Villar, A.J., Yohai, V.J., (2009). "Supplemental material for Robust response transformations based on optimal prediction", *Jasa Site*, supplemental material.

[11] Marazzi A., Yohai V., (2004). "Adaptively truncated maximum likelihood regression with asymmetric errors", *Journal of Statistical Planning and inference*, **122**, 271-291.

[12] Maronna, R.A., Martin, R.D., Yohai, V.J., (2007). "Robust Statistics: Theory and Methods", *Wiley*, New York.

[13] Mizera, I., (1993). "On consistent M-estimators: tuning constants, unimodality and breakdown", *Kybernetika* **30**, 289-300.

[14] Mizera, I., (1993). "Weak continuity and identifiability of M-functionals", Thesis, Charles University, Prague.

[15] Nelson W.B., (1990). "Accelerated Testing, Statistical Models, Test Plans, and Data Analyses", *Wiley*, New York.

[16] Rousseeuw, P.J., Leroy, A., (1987). "Robust Regression and Outlier Detection", *Wiley*, New York.

[17] Rousseeuw, P.J., Yohai, V.J., (1987). "Robust regression by means of S-estimates", in: Robust and Nonlinear Time Series, Lecture Notes in Statistics, *Springer*, New York, **26**, 256-272.

[18] Salibian-Barrera, M., Yohai, V.J., (2008). "High breakdown point robust regression with censored data". *The Annals of Statistics* **36**, 118-146.

[19] Yohai, V.J. (1985). "High Breakdown Point and High Efficiency Robust Estimates for Regression". *Department of Statistics, University of Washington*, technical report, No. 66, disponible en
`http://www.stat.washington.edu/research/reports/1985/tr066.pdf`.