

CONTRIBUCIONES A LA TEORÍA DE ROBUSTEZ  
RESPECTO AL SESGO

TESIS DOCTORAL

Autor: José Ramón Berrendero Díaz

Directores: Profs. Juan J. Romo y Rubén H. Zamar

UNIVERSIDAD CARLOS III DE MADRID  
Departamento de Estadística y Econometría

Getafe, junio de 1996

DONACIÓN

A la memoria de mi madre y a mi padre

# Índice

<b>Resumen</b>	<b>5</b>
<b>1 Introducción a la teoría de robustez respecto al sesgo</b>	<b>10</b>
1.1 Conceptos generales . . . . .	10
1.2 Robustez respecto al sesgo para estimadores de escala . . . . .	15
1.2.1 M-estimadores de escala . . . . .	15
1.2.2 Curva de sesgo de M-estimadores de escala . . . . .	17
1.2.3 Teoría minimax para M-estimadores de escala . . . . .	19
1.3 Robustez respecto al sesgo para estimadores de dispersión y localización	20
1.3.1 Estimadores de dispersión . . . . .	21
1.3.2 Estimadores de localización . . . . .	25
1.4 Robustez respecto al sesgo en el modelo de regresión lineal . . . . .	27
1.4.1 Curva de sesgo en el modelo de regresión lineal . . . . .	27
1.4.2 Estimadores robustos de regresión con residuos admisibles . . . .	28
1.4.3 Ejemplos de estimadores con residuos admisibles . . . . .	30

---

1.4.4	Otros estimadores robustos de regresión . . . . .	35
<b>2</b>	<b>La tasa de ruptura de la curva de sesgo asintótico máximo</b>	<b>39</b>
2.1	Introducción . . . . .	39
2.2	Tasa de ruptura de M-estimadores de escala . . . . .	44
2.2.1	Tasa de ruptura respecto a outliers . . . . .	44
2.2.2	Tasa de ruptura respecto a inliers . . . . .	55
2.3	Tasa de ruptura de $\tau$ -estimadores de escala . . . . .	57
2.4	Tasa de ruptura de estimadores de dispersión . . . . .	59
2.4.1	Tasa de ruptura de la MEDA . . . . .	59
2.4.2	Tasa de ruptura de algunas alternativas a la MEDA . . . . .	63
2.5	Tasa de ruptura de M-estimadores de localización . . . . .	67
2.6	Estimación simultánea de localización y dispersión . . . . .	75
<b>3</b>	<b>Sesgo de estimadores de regresión con residuos admisibles</b>	<b>79</b>
3.1	Introducción e hipótesis básicas . . . . .	79
3.2	Cálculo del sesgo de estimadores de residuos admisibles . . . . .	82
3.3	Ejemplos de aplicación . . . . .	88
3.3.1	S-estimadores . . . . .	88
3.3.2	$\tau$ -estimadores . . . . .	91
3.3.3	Sesgo del LMS-estimador con regresores no elípticos . . . . .	95
3.3.4	Estimadores basados en rangos absolutos . . . . .	95

---

3.4	Curva de sesgo para M-estimadores con escala general . . . . .	101
3.5	Una aplicación al análisis robusto de componentes principales . . . . .	108
3.5.1	Notación y resultados básicos . . . . .	110
3.5.2	Cálculo de la curva de sesgo asintótico máximo . . . . .	114
3.5.3	Ejemplo: M-estimadores de escala . . . . .	117
<b>4</b>	<b>Sesgo de estimadores de regresión con residuos admisibles generalizados</b>	<b>121</b>
4.1	Introducción e hipótesis básicas . . . . .	121
4.2	Curva de sesgo de estimadores con residuos admisibles generalizados . .	124
4.2.1	Lemas previos . . . . .	124
4.2.2	Curva de sesgo asintótico máximo . . . . .	132
4.3	Ejemplos de aplicación . . . . .	139
4.3.1	GS-estimadores . . . . .	139
4.3.2	$G\tau$ -estimadores . . . . .	143
4.3.3	GR-estimadores . . . . .	146
4.4	Teoría minimax en la clase de estimadores con residuos admisibles generalizados . . . . .	150
4.4.1	Lemas previos . . . . .	150
4.4.2	Estimadores de sesgo minimax . . . . .	155
	<b>Referencias</b>	<b>160</b>

## Resumen

La curva de sesgo asintótico máximo mide la mayor discrepancia posible entre el valor al que converge un estimador y el valor del parámetro que se desea estimar cuando hay datos atípicos en la muestra. El objetivo central del presente trabajo es el de investigar nuevos aspectos y problemas de la teoría de robustez, cuando la estabilidad de los estimadores se evalúa mediante esta curva, así como unificar y extender algunos de los resultados que ya existían previamente.

En primer lugar, se propone una nueva medida de robustez, la tasa de ruptura, que mide la estabilidad de los estimadores cuando la proporción de datos atípicos en la muestra es alta. Esta medida matiza la información proporcionada por medidas de robustez ya existentes, como el punto de ruptura. También se obtiene un método general para calcular la curva de sesgo asintótico máximo, válido para estimadores de regresión que se construyen buscando aquel vector de parámetros que minimiza una medida robusta de la escala de los residuos. Finalmente, aunque se prueba que no es posible aplicar el método anterior a estimadores de regresión basados en minimizar la escala de las diferencias de los residuos, también se proporciona una teoría adecuada a este caso.

El capítulo uno contiene una introducción a la teoría de robustez respecto al sesgo. Las ideas, conceptos y resultados que se utilizan en los capítulos posteriores se discuten

más detalladamente. Los contextos en los que se centra la atención son el modelo de localización y dispersión, y el modelo de regresión lineal. Las aportaciones originales de la tesis se concentran en los capítulos dos, tres y cuatro.

En el capítulo dos se discute la conveniencia de disponer de medidas sencillas que resuman la información que la curva de sesgo asintótico máximo contiene sobre la estabilidad de los estimadores. Se propone una nueva medida de este tipo, la tasa de ruptura, y se proporcionan los resultados que permiten calcularla fácilmente a partir de la definición de los estimadores. Diversos ejemplos numéricos en los modelos de escala, localización y dispersión muestran la utilidad de esta nueva medida. También se ponen de manifiesto en este capítulo algunas relaciones entre la manera de definir los estimadores y las propiedades de robustez de los mismos.

El capítulo tres está dedicado esencialmente a la obtención de un método para calcular la curva de sesgo asintótico máximo de estimadores de regresión con residuos admisibles. Estos estimadores son tales que la distribución de los valores absolutos de los residuos que generan no puede ser mejorada uniformemente por la distribución de los residuos proporcionados por otro vector de parámetros. Los resultados pueden ser aplicados a diversas clases de estimadores que cumplen esta propiedad. En concreto, se darán fórmulas para S-,  $\tau$ - y R-estimadores de regresión. El método también es válido si la distribución de los regresores no es elíptica.

Cuando en lugar de residuos, se consideran las proyecciones de los datos según una determinada dirección, y se busca la dirección que hace menor una medida robusta de la escala de estas proyecciones, obtenemos un método robusto de hacer análisis de componentes principales. El método general mencionado en el párrafo anterior se aplica también en este contexto para evaluar la robustez de las componentes principales robustas desde el punto de vista de la curva de sesgo asintótico máximo.

Finalmente, en el capítulo cuatro se introduce una nueva clase general de estimadores robustos de regresión, la de estimadores con residuos admisibles generalizados. La idea consiste en considerar cualquier estimador con residuos admisibles pero midiendo la escala de las diferencias entre pares de residuos en lugar de medir la escala de los residuos directamente. Aunque la teoría del capítulo tres no puede aplicarse en este caso, se proporcionan también fórmulas para las curvas de sesgo asintótico máximo de estos estimadores. También en este capítulo se obtiene el estimador que, dentro de esta clase, tiene menor curva de sesgo asintótico máximo (estimador de sesgo minimax).

Algunas conclusiones del trabajo descrito previamente son:

- Las propiedades de robustez ante una proporción elevada de *outliers* de los M-estimadores de escala, localización y dispersión se pueden deducir fácilmente a partir de las funciones utilizadas para definir los estimadores. Por ejemplo, para un M-estimador de escala, cuanto más “plana” sea la función que lo define en un entorno de cero, más robusto es el estimador. Propiedades análogas a la anterior pueden también probarse para estimadores de localización y regresión.
- La pérdida de robustez de un M-estimador de localización en relación a la mediana cuando la proporción de *outliers* es alta, se debe en gran parte a que es necesario estimar el parámetro de dispersión desconocido.
- Muchos estimadores robustos de regresión están basados en elegir el vector de parámetros que minimiza una medida robusta de la escala de los residuos. Un procedimiento general para obtener su curva de sesgo asintótico máximo es el siguiente:
  - (a) Situar una proporción dada de contaminaciones puntuales en una dirección adecuada, determinada por un vector de parámetros, y llevar los datos contaminados suficientemente lejos de los datos no contaminados.

- (b) Seleccionar la ecuación de regresión que ajusta exactamente estas contaminaciones y calcular la escala de los residuos resultantes.
  - (c) Seleccionar la ecuación de regresión correspondiente a los datos no contaminados, es decir, ignorar por completo los *outliers*, y calcular la escala de los residuos resultantes.
  - (d) La norma del vector de parámetros del modelo para el que se da la igualdad entre las escalas obtenidas en los apartados (b) y (c) es el valor del sesgo asintótico máximo para la proporción de contaminación dada.
- Cuando los estimadores de regresión se basan en encontrar el vector de parámetros que minimiza una medida robusta de la escala de las diferencias de pares de residuos, el procedimiento para calcular la curva de sesgo asintótico máximo es similar al anterior. La diferencia estriba en el tipo de contaminaciones. En esta nueva situación, las contaminaciones son tales que, a medida que se alejan de los datos no contaminados, la dispersión entre ellas aumenta sin límite.
  - El estimador que minimiza un cuantil de las diferencias en valor absoluto de pares de residuos, tiene propiedades de optimalidad respecto al sesgo en una clase amplia de estimadores robustos de regresión. El cuantil óptimo a minimizar depende de la proporción de datos contaminados en la muestra.

## AGRADECIMIENTOS

Quiero agradecer en primer lugar a los profesores Juan Romo y Rubén Zamar su imprescindible ayuda y sus ánimos constantes durante la realización de este trabajo. También quiero mostrar mi agradecimiento a los miembros del Departamento de Estadística y Econometría de la Universidad Carlos III de Madrid que colaboraron para crear un ambiente realmente propicio. Me gustaría mencionar especialmente a Santiago Mira, Marta Moreno y Manuel Moreno, con los que compartí conversaciones y horas de trabajo en el mismo despacho. En último lugar, aunque no en importancia, agradezco a Susana su presencia.

Durante la realización de este trabajo, el autor se ha beneficiado de la financiación parcial del proyecto de investigación DGICYT PB93-0232.

Getafe, junio de 1996.

# Capítulo 1

## Introducción a la teoría de robustez respecto al sesgo

### 1.1 Conceptos generales

Las inferencias estadísticas son el resultado de suponer un modelo sobre los datos disponibles y aplicar al conjunto formado por los datos y el modelo algún tipo de procedimiento. Usualmente, los datos sólo se ajustan de forma aproximada al modelo por lo que implícitamente se asume que el procedimiento estadístico utilizado goza de una propiedad de continuidad que podría enunciarse de la siguiente manera: si el modelo supuesto se aproxima al modelo que realmente genera los datos, los resultados bajo el modelo supuesto también se aproximarán a los resultados del procedimiento bajo el modelo real. Sin embargo, algunos de los métodos estadísticos más clásicos, tales como la estimación de máxima verosimilitud bajo normalidad o la estimación de mínimos cuadrados en el modelo de regresión lineal, no verifican esta propiedad pues se muestran inestables ante pequeñas desviaciones del modelo supuesto o nominal. La teoría de robustez en Estadística se ocupa de diseñar procedimientos estables o *robustos* y de evaluar la estabilidad o *robustez* de dichos procedimientos.

Para poder evaluar los métodos estadísticos desde este punto de vista necesitamos, en primer lugar, formalizar matemáticamente la idea de proximidad entre el modelo real y el modelo nominal. En segundo lugar, precisamos criterios que nos permitan cuantificar la estabilidad de los distintos métodos.

Con respecto al primer punto, en el presente trabajo utilizaremos los llamados *entornos de contaminación*. Supondremos que tenemos una muestra de datos  $y_1, \dots, y_n$  procedentes de una distribución  $F$  perteneciente al conjunto

$$V_\epsilon = \{F : F = (1 - \epsilon)F^\theta + \epsilon H, H \text{ distribución arbitraria}\}, \quad 0 \leq \epsilon \leq 1,$$

es decir, una proporción aproximada  $1 - \epsilon$  de los datos de la muestra procede de la distribución nominal  $F^\theta$ , que es un modelo conocido salvo el valor del parámetro  $\theta$  que se desea estimar, mientras que la proporción  $\epsilon$  restante procede de una distribución arbitraria  $H$ . Estas últimas observaciones se pueden interpretar como datos atípicos o *outliers*. La utilización de estos entornos para modelizar la incertidumbre sobre el modelo se remonta a Tukey (1958, 1960) y a Huber (1964).

Se han propuesto muchos criterios para medir la estabilidad de un procedimiento estadístico. Algunos de ellos son apropiados para evaluar la robustez cuando la proporción  $\epsilon$  de datos contaminados en la muestra es pequeña mientras que otros indican la proporción máxima de datos contaminados que el método puede soportar. Nosotros estudiaremos aquí un criterio de robustez global, en el sentido de que proporciona información sobre la robustez en ambas situaciones: *la curva de sesgo asintótico máximo*.

Pensaremos en los distintos métodos estadísticos como en funcionales,  $T$ , definidos sobre un conjunto de distribuciones  $\mathcal{M}$  que incluye a  $V_\epsilon$ , y con valores en  $\mathbb{R}^p$ , donde  $p$  es la dimensión del parámetro a estimar. Cuando se dispone de una muestra concreta, la estimación del parámetro se obtiene aplicando  $T$  a la función de distribución empírica de los datos,  $F_n$ , por lo que  $\mathcal{M}$  también debe contener al conjunto de todas las posibles

distribuciones empíricas. Si  $F \in V_\epsilon$  es la distribución de los datos, es posible probar en muchas ocasiones que

$$\hat{\theta}_n = T(F_n) \rightarrow T(F) \quad \text{c.s.}$$

Como en general  $F \neq F^\theta$ , si definimos una medida de discrepancia adecuada,  $d$ , en  $\mathbb{R}^p$ , la estabilidad asintótica de  $T$  en  $F$  con respecto a  $d$  viene dada por

$$b(T, F) = d[T(F), T(F^\theta)],$$

es decir, la discrepancia existente entre el resultado asintótico obtenido bajo el modelo real y bajo el nominal. La curva de sesgo asintótico máximo es la máxima discrepancia sobre el conjunto de todas las distribuciones admitidas como posibles, es decir, sobre el entorno de contaminación:

$$B_T(\epsilon) = \sup_{F \in V_\epsilon} d[T(F), T(F^\theta)].$$

Dado que al aumentar  $\epsilon$ , el entorno  $V_\epsilon$  sobre el que tomamos el supremo es mayor, la curva  $B_T(\epsilon)$  es creciente como función de la proporción de datos contaminados en la muestra. Además, nos da información sobre la estabilidad de  $T$  para cualquier valor de  $\epsilon$ . Por sencillez y como no hay posibilidad de confusión, llamaremos curva de sesgo, o simplemente sesgo, a la curva de sesgo asintótico máximo.

Esta curva fue definida inicialmente por Huber (1964) en el contexto de estimación de un parámetro de localización. Posteriormente, Martin y Zamar (1989) la introdujeron y calcularon en el caso de estimación de un parámetro de escala y Martin, Yohai y Zamar (1989) la estudiaron en el modelo de regresión lineal. Algunas de sus propiedades en el método de estimación de mínima distancia han sido investigadas por Donoho y Liu (1988). He y Simpson (1993) han probado algunas propiedades de la curva de sesgo válidas para un grupo bastante general de modelos.

A veces es conveniente disponer de medidas sencillas que resuman el comportamiento de  $B_T(\epsilon)$ . Daremos algún ejemplo de esta afirmación en la sección 2.1. Dos medidas ampliamente utilizadas que cumplen este propósito son:

- (a) La *sensitividad*,  $\gamma_T^*$ , que es la derivada de la curva de sesgo evaluada en  $\epsilon = 0$ . Un desarrollo de Taylor de primer orden implica

$$B_T(\epsilon) = \gamma_T^* \epsilon + o(\epsilon), \quad \text{si } \epsilon \rightarrow 0.$$

Bajo condiciones de regularidad, la sensitividad coincide con el supremo de la función de influencia, o sensitividad a errores grandes, definida por Hampel (1974). Si  $\gamma_T^* < \infty$ , se dice que  $T$  tiene *influencia acotada*.

- (b) El *punto de ruptura*,  $\epsilon_T^*$ , introducido también por Hampel (1974) y ampliamente discutido en el trabajo de Donoho y Huber (1983), se define como

$$\epsilon_T^* = \sup\{\epsilon : B_T(\epsilon) < \infty\}.$$

El punto de ruptura representa la máxima proporción de datos contaminados que un funcional  $T$  puede resistir ya que si  $\epsilon > \epsilon^*$ , entonces  $B_T(\epsilon) = \infty$  y es posible obtener un valor asintótico del estimador que no guarda ninguna relación con el valor del parámetro que se desea estimar. En relación con  $\epsilon_T^*$  se propondrá y estudiará una medida complementaria de robustez en el capítulo dos.

Una analogía física que ayuda a comprender mejor los conceptos de sensitividad y punto de ruptura puede encontrarse en Rousseeuw y Leroy (1987), página 187. Estos autores comparan la sensitividad con la elasticidad de una barra cuando se aplica sobre ella una fuerza pequeña, y el punto de ruptura con la fuerza necesaria para poder romper la barra. Si identificamos la barra con un estimador y la fuerza con los datos atípicos, obtenemos una descripción bastante gráfica de los dos conceptos anteriores.

El cálculo efectivo de la curva de sesgo es a veces difícil. Contribuciones a la resolución de este problema para clases amplias de funcionales pueden encontrarse en

los capítulos tres y cuatro. Otros problemas de interés en relación con la curva de sesgo consisten en seleccionar una clase de estimadores  $\mathcal{T}$  y encontrar, para cada  $\epsilon$  fijo, el estimador  $T^* \in \mathcal{T}$  que verifica

$$B_{T^*}(\epsilon) \leq B_T(\epsilon), \quad \text{para todo } T \in \mathcal{T}.$$

Se dice entonces que  $T^*$  es el *estimador de sesgo minimax* en  $\epsilon$  para la clase  $\mathcal{T}$ . Un ejemplo de resolución de este tipo de problemas se encuentra en el capítulo cuatro.

En las páginas siguientes se pretende un objetivo doble: en primer lugar, precisar los conceptos introducidos en esta sección para los modelos que van a ser estudiados en los capítulos posteriores y, en segundo lugar, revisar los resultados más importantes que existen en la literatura, algunos de los cuales serán utilizados más adelante en este trabajo. En la sección 1.2 se formalizan las ideas de robustez respecto al sesgo en el modelo de estimación de un parámetro de escala, que es un modelo sencillo pero con interesantes implicaciones en modelos más complicados. Fundamentalmente seguiremos aquí el trabajo de Martin y Zamar (1989). Algunas complicaciones aparecen cuando se ha de estimar tanto el parámetro de dispersión como el de localización. La sección 1.3 trata este problema siguiendo a Martin y Zamar (1993a, 1993b). Por último, en la sección 1.4 se discute la teoría de robustez respecto al sesgo para el modelo de regresión lineal. Aunque la literatura aquí es muy extensa, los artículos en los que se apoya nuestro trabajo de forma esencial son los de Martin, *et al.* (1989), Yohai y Zamar (1993) y Croux, Rousseeuw y Hössjer (1994).

## 1.2 Robustez respecto al sesgo para estimadores de escala

### 1.2.1 M-estimadores de escala

Supongamos que se dispone de una muestra  $y_1, \dots, y_n$  de observaciones independientes e idénticamente distribuidas según  $F^\sigma(y) = F_0(y/\sigma)$ . A lo largo de esta sección y de la siguiente supondremos que  $F_0$  es estrictamente creciente, con densidad  $f_0$  simétrica y estrictamente unimodal (los resultados de esta sección son también válidos si  $F_0$  es una distribución estrictamente creciente en  $(0, \infty)$  de una variable aleatoria no negativa). Para fijar ideas, podemos suponer que  $F_0(y) = \Phi(y)$  es la función de distribución de la normal estándar. Si calculamos en este caso el estimador de máxima verosimilitud de  $\sigma$  obtenemos  $S_n = [\sum_{i=1}^n y_i^2/n]^{1/2}$ . Es inmediato comprobar que  $S_n$  satisface la ecuación

$$\frac{1}{n} \sum_{i=1}^n \chi(y_i/S_n) = b, \quad (1.1)$$

donde  $\chi(y) = y^2$  y  $b = 1$ . El estimador de máxima verosimilitud  $S_n$  no es robusto ya que un único dato atípico puede hacerle tomar valores aberrantes. En su trabajo fundamental, Huber (1964) propone transformar convenientemente la función  $\chi$  de forma que se obtengan estimadores más robustos. Esta modificación puede consistir, por ejemplo, en considerar en la ecuación (1.1) una parábola truncada en lugar de  $\chi(y) = y^2$ . Es decir, tomar  $\chi(y) = y^2 I\{|y| \leq c\} + c^2 I\{|y| > c\}$ , con el fin de que los datos atípicos influyan menos en el resultado final. También es necesario elegir convenientemente la constante  $b$ .

Supongamos que  $\chi$  verifica la siguiente hipótesis:

**Hipótesis 1**  $\chi(y)$  es par, no decreciente para valores positivos de  $y$ , acotada con  $\chi(\infty) = 1$ ,  $\chi(0) = 0$ , continua en 0 y con a lo más un número finito de discontinuidades.

Las consideraciones anteriores motivan la siguiente definición: se dice que  $S_n$  es un *M-estimador de escala* si

$$S_n = \inf\left\{s : \frac{1}{n} \sum_{i=1}^n \chi\left(\frac{y_i}{s}\right) < b\right\}, \quad (1.2)$$

donde  $\chi$  verifica la hipótesis 1 y  $0 < b < 1$ .

El ínfimo de la ecuación (1.2) es necesario para asegurar la existencia del estimador en algunos ejemplos importantes en los que  $\chi$  no es continua. Si  $\chi$  es continua se puede sustituir (1.2) por la expresión más sencilla (1.1).

Asociado a  $S_n$  podemos definir el funcional  $S(F)$  como

$$S(F) = \inf\left\{s : E_F \chi\left(\frac{y}{s}\right) < b\right\}. \quad (1.3)$$

Sobre las relaciones entre  $S(F)$  y  $S_n$  se pueden hacer las siguientes precisiones:

- (a) Como se comprueba de manera inmediata,  $S_n$  se obtiene aplicando el funcional  $S(F)$  a la función de distribución empírica de los datos,  $F_n$ . Es decir,  $S_n = S(F_n)$ .
- (b) Bajo la hipótesis 1,  $S_n$  converge al valor del funcional aplicado a la verdadera distribución de los datos, es decir,

$$S_n \rightarrow S(F^\sigma) \quad \text{c.s.} \quad (1.4)$$

Una comprobación de (1.4) se puede encontrar en la Nota A.1 de Martin y Zamar (1989).

- (c) Es deseable la convergencia hacia el valor del parámetro que se desea estimar,  $\sigma$ . En virtud de (1.4), es necesario seleccionar  $\chi$  y  $b$  de forma que  $S(F^\sigma) = \sigma$ . Cuando esto sucede decimos que el funcional  $S$  es *consistente en el sentido de Fisher* o Fisher-consistente. Para que se cumpla esta condición es suficiente tomar  $\chi$  y  $b$  tales que

$$E_{F^\sigma} \chi\left(\frac{y}{\sigma}\right) = b. \quad (1.5)$$

En este caso, cuando no hay contaminaciones en la muestra el estimador  $S_n$  converge al valor del parámetro que se desea estimar, es decir, es consistente. Si queremos que se cumpla esta importante propiedad, no es posible elegir  $\chi$  y  $b$  independientemente en la definición de un M-estimador de escala, sino que ambas deben estar ligadas por la condición (1.5).

### 1.2.2 Curva de sesgo de M-estimadores de escala

Supongamos ahora que la muestra está contaminada, es decir, la distribución de la que proceden los datos es  $F$ , donde

$$F \in V_\epsilon = \{F : F = (1 - \epsilon)F^\sigma + \epsilon H, H \text{ distribución arbitraria}\}.$$

Dado que

$$S_n \rightarrow S(F) \quad \text{c.s.,}$$

y como  $F \neq F^\sigma$ , se tiene en general que  $S(F) \neq F(F^\sigma) = \sigma$ . Debido a la presencia de datos contaminados en la muestra se produce una diferencia entre el valor asintótico del estimador y el valor del parámetro que queremos estimar.

Como medida de discrepancia entre  $S(F)$  y  $\sigma$  elegimos el cociente  $S(F)/\sigma$ . Como los M-estimadores de escala son equivariantes por cambios de escala, podemos suponer sin pérdida de generalidad para estudiar la curva de sesgo que  $\sigma = 1$ .

Debemos considerar dos tipos de sesgo: el sesgo respecto a *outliers*,  $B_\chi^+(\epsilon)$ , causado por observaciones grandes que llevan a sobreestimar el parámetro, y el sesgo respecto a *inliers*,  $B_\chi^-(\epsilon)$ , producido por observaciones pequeñas que llevan a infraestimar el parámetro. Las respectivas definiciones son:

$$B_\chi^+(\epsilon) = \sup_{F \in V_\epsilon} S(F),$$

y

$$B_x^-(\epsilon) = \inf_{F \in V_\epsilon} S(F).$$

Definimos la clase de funciones  $C_b$  como la formada por aquellas funciones  $\chi$  tales que verifican la hipótesis 1 y además cumplen la condición de consistencia de Fisher  $E_{F_0}\chi(y) = b$ . Puede probarse (es obvio cuando  $\chi$  es continua) que  $B_x^+(\epsilon)$  se obtiene tomando la distribución contaminante  $H = \delta_x$ , donde  $\delta_x$  asigna probabilidad uno al punto  $x$ , y haciendo  $x \rightarrow \infty$ . En este caso, debe cumplirse, para  $\chi \in C_b$ ,

$$(1 - \epsilon)E_{F_0}\chi\left(\frac{y}{B_x^+(\epsilon)}\right) + \epsilon = b.$$

Por otra parte,  $B_x^-(\epsilon)$  se obtiene contaminando en  $H = \delta_x$  y haciendo  $x \rightarrow 0$ , con lo que para  $\chi \in C_b$ ,

$$(1 - \epsilon)E_{F_0}\chi\left(\frac{y}{B_x^-(\epsilon)}\right) = b.$$

Resolviendo las dos ecuaciones no lineales anteriores para cada  $\epsilon$  fijo, se obtienen los valores de  $B_x^+(\epsilon)$  y  $B_x^-(\epsilon)$ . Podemos resumir estos resultados en la siguiente proposición:

**Proposición 1 (Martin y Zamar (1989), teorema 1)** *Sea  $\chi \in C_b$  y sea  $h_x(s) = E_{F_0}\chi(y/s)$ . Entonces,*

$$B_x^+(\epsilon) = h_x^{-1}\left(\frac{b - \epsilon}{1 - \epsilon}\right), \quad (1.6)$$

y

$$B_x^-(\epsilon) = h_x^{-1}\left(\frac{b}{1 - \epsilon}\right), \quad (1.7)$$

En relación con los sesgos  $B_x^+(\epsilon)$  y  $B_x^-(\epsilon)$ , podemos definir los correspondientes puntos de ruptura respecto a *outliers* e *inliers* de la siguiente manera:

$$\epsilon^+ = \sup\{\epsilon : B_x^+(\epsilon) < \infty\},$$

y

$$\epsilon^- = \sup\{\epsilon : B_{\chi}^-(\epsilon) > 0\}.$$

Dado que  $\lim_{s \rightarrow \infty} h_{\chi}(s) = 0$  y  $\lim_{s \rightarrow 0} h_{\chi}(s) = 1$ , se sigue que  $\epsilon^+ = b$  y  $\epsilon^- = 1 - b$ . Por tanto, el punto de ruptura de un M-estimador de escala basado en  $\chi \in C_b$  verifica

$$\epsilon^* = \min\{\epsilon^+, \epsilon^-\} = \min\{b, 1 - b\}.$$

Esta fórmula coincide con la obtenida por Huber (1981), página 110. Cuando se selecciona una función  $\chi$  se determina, a través de la condición (1.5), el valor de  $b$  y es este valor el que caracteriza el punto de ruptura del M-estimador de escala.

### 1.2.3 Teoría minimax para M-estimadores de escala

Es natural preguntarse por las funciones  $\chi$  que dan lugar a estimadores con el menor sesgo asintótico máximo. Para todo  $0 < b < 1$ , existe una única constante  $a > 0$  tal que la función  $\chi$  de salto

$$\chi_a(y) = \begin{cases} 0, & \text{si } |y| < a \\ 1, & \text{si } |y| \geq a \end{cases}$$

pertenece a  $C_b$ . Resulta que esta función es la más robusta respecto a  $B_{\chi}^+(\epsilon)$  y a  $B_{\chi}^-(\epsilon)$ , para todo  $\epsilon$ , en la clase  $C_b$ , tal y como se enuncia a continuación.

**Proposición 2 ( Martin y Zamar (1989), lema A.3 )** *Sea  $F_0$  una distribución con densidad  $f_0$  con la propiedad de que  $f_0(sy)/f_0(y)$  es decreciente en  $y$  para todo  $s > 1$ , entonces para toda función  $\chi \in C_b$ ,*

$$B_{\chi}^+(\epsilon) \geq B_{\chi_a}^+(\epsilon), \quad \text{para todo } 0 < \epsilon < b,$$

y

$$B_{\chi}^-(\epsilon) \leq B_{\chi_a}^-(\epsilon), \quad \text{para todo } 0 < \epsilon < 1 - b,$$

donde  $\chi_a$  es la única función de salto que pertenece a  $C_b$ .

El M-estimador de escala con sesgo minimax basado en la función de salto  $\chi_a$  es un estadístico de orden reescalado, es decir, si  $S_n$  es el M-estimador basado en  $\chi_a$ , entonces

$$S_n = \frac{1}{a} y_{(n-[nb])}, \quad (1.8)$$

donde  $[nb]$  denota la parte entera de  $nb$ . Por ejemplo, para  $b = 1/2$ , el estimador de sesgo minimax no es más que la mediana convenientemente reescalada para obtener consistencia bajo el modelo  $F_0$ .

### 1.3 Robustez respecto al sesgo para estimadores de dispersión y localización

Supongamos ahora que la distribución central del entorno es el modelo clásico de localización y dispersión,

$$F^{\mu,\sigma}(y) = F_0\left(\frac{y - \mu}{\sigma}\right),$$

y que estamos interesados en estimar el parámetro de dispersión,  $\sigma$ , y el de localización,  $\mu$ . Para el parámetro de dispersión, la definición del sesgo es exactamente la misma que la que hemos visto en el modelo de escala de la sección anterior. Sea ahora  $T$  un funcional que define un estimador para el parámetro de localización. La curva de sesgo de  $T$  se define como

$$B_T(\epsilon) = \sup_{F \in \mathcal{V}_\epsilon} \frac{|T(F) - \mu|}{\sigma}.$$

En este caso no es necesario distinguir entre *outliers* e *inliers* ya que el supremo y el ínfimo sobre el entorno tienen el mismo valor absoluto.

Dado que sólo consideraremos estimadores equivariantes por traslaciones y cambios de escala, para estudiar las curvas de sesgo podemos suponer sin pérdida de generalidad que  $\mu = 0$  y  $\sigma = 1$

### 1.3.1 Estimadores de dispersión

Veamos algunos resultados sobre robustez respecto al sesgo de estimadores de dispersión. Los estimadores de dispersión se pueden contemplar como estimadores de escala de los datos centrados  $y_1 - T_n, \dots, y_n - T_n$ , donde el estimador de localización utilizado para centrar los datos se puede seleccionar siguiendo distintos procedimientos.

Por ejemplo,  $T_n$  puede elegirse previamente *ad-hoc* y aplicar posteriormente un M-estimador de escala a los datos centrados. Tal vez el estimador de dispersión más conocido en esta clase sea la MEDA (*mediana de las desviaciones absolutas a la mediana*):

$$M_n = c \operatorname{med}_i\{|y_i - \operatorname{med}_j y_j|\}, \quad c = 1/F_0^{-1}(0.75). \quad (1.9)$$

La MEDA corresponde a seleccionar previamente  $T_n = \operatorname{med}_i\{y_i\}$  y aplicar a los datos centrados mediante  $T_n$  un M-estimador de escala basado en la función  $\chi$  de salto que cumple  $b = 1/2$ .

Otro posible método consiste en definir  $T_n$  como el valor de  $t$  que minimiza un M-estimador de escala,  $D_n(t)$ , de los valores  $y_1 - t, \dots, y_n - t$ . Es decir:

$$T_n = \arg \min_{t \in \mathbb{R}} D_n(t).$$

Este método corresponde a aplicar al modelo de localización y dispersión el concepto de S-estimador de regresión introducido por Rousseeuw y Yohai (1984). El *S-estimador de dispersión* correspondiente es

$$D_n = D_n(T_n).$$

Si  $D_n(t)$  está basado en la misma función  $\chi$  de salto utilizada para definir la MEDA, puede probarse (véase Rousseeuw y Leroy, 1987) que  $T_n$  es el punto medio del intervalo

más pequeño que contiene al menos el 50% de los datos mientras que

$$D_n = c \min_{t \in \mathbb{R}} \text{med}_i \{|y_i - t|\}, \quad c = 1/F_0^{-1}(0.75), \quad (1.10)$$

es la longitud de dicho intervalo convenientemente reescalada para obtener consistencia bajo el modelo  $F_0$ . De ahí que a este estimador de dispersión se le denomine SHORTH (de la expresión en inglés *shortest half*).

Comparemos las expresiones (1.9) y (1.10). En (1.9) se fija la localización y se mide la dispersión de los datos en torno a esta localización. En (1.10), localización y dispersión se determinan simultáneamente de forma que se minimice la dispersión, medida a través de un M-estimador robusto de la escala de los “residuos”. En cuanto al sesgo de S-estimadores de dispersión se tiene el siguiente resultado:

**Proposición 3 (Martin y Zamar (1993b), teorema 4)** *Sea  $\chi \in C_b$ . Las curvas de sesgo respecto a outliers e inliers del S-estimador de dispersión basado en  $\chi$  son las mismas que las del M-estimador de escala basado en la misma función  $\chi$ . Además, la curva de sesgo respecto a inliers del S-estimador de dispersión es la misma que la del M-estimador de dispersión basado en la misma función  $\chi$  y calculado sobre los datos centrados previamente por un estimador de localización.*

Otra manera de estimar la dispersión y la localización simultáneamente es análoga a la obtención de los estimadores de máxima verosimilitud. Se resuelve el siguiente sistema de dos ecuaciones con dos incógnitas:

$$(1/n) \sum_{i=1}^n \psi((y_i - T_n)/D_n) = 0 \quad (1.11)$$

$$(1/n) \sum_{i=1}^n \chi((y_i - T_n)/D_n) = b, \quad (1.12)$$

donde  $\chi \in C_b$  y la función  $\psi$  verifica la hipótesis:

**Hipótesis 2**  $\psi$  es impar, no decreciente, acotada con  $\psi(\infty) = 1$ , y continua salvo quizá en un número finito de puntos.

Este método fue propuesto ya por Huber (1964) y es conocido con el nombre de *propuesta 2 de Huber*. La MEDA es un caso particular de resolución del sistema formado por (1.11) y (1.12) cuando  $\psi(y) = \text{sgn}(y)$ , la función que toma el valor 1 si  $y \geq 0$  y toma el valor -1 si  $y < 0$ , y  $\chi$  es una función de salto.

Sobre las curvas de sesgo de estimadores simultáneos de localización y dispersión se tiene el siguiente resultado:

**Proposición 4 (Martin y Zamar (1993b), lema 2)** *Supongamos que se verifican las mismas hipótesis que en la proposición 3 y que la función  $\psi$  satisface la hipótesis 2. Sean  $t^*$  y  $s^*$  los sesgos de los estimadores simultáneos de localización y dispersión obtenidos al resolver el sistema formado por las ecuaciones (1.11) y (1.12). Entonces  $t^*$  y  $s^*$  verifican*

$$(1 - \epsilon)E_{F_0}\chi\left(\frac{y - t^*}{s^*}\right) + \epsilon = b,$$

y

$$(1 - \epsilon)E_{F_0}\psi\left(\frac{y - t^*}{s^*}\right) + \epsilon = 0.$$

Finalmente, existen algunos estimadores de dispersión que no requieren el uso de un estimador de localización previo o simultáneo. Estos estimadores no miden la dispersión en torno a un centro sino que recogen más bien la magnitud de las diferencias entre los distintos pares de datos  $y_i - y_j$ . En concreto, Rousseeuw y Croux (1993) proponen las siguientes alternativas:

$$S_n = k \text{ med}_i\{\text{med}_j|y_i - y_j|\}, \quad (1.13)$$

y

$$Q_n = d \{|y_i - y_j| : i < j\}_{(p)}, \quad (1.14)$$

donde  $p = \binom{h}{2}$  con  $h = \lfloor \frac{n}{2} \rfloor + 1$ , y las constantes  $k$  y  $d$  son las adecuadas para que los estimadores sean Fisher-consistentes.

En la ecuación (1.13) se define un estimador que se calcula de la siguiente forma: para cada  $i$  fijo, calcular la mediana de  $\{|y_i - y_j| : j = 1, \dots, n\}$ . Esta operación da lugar a  $n$  valores, la mediana de los cuales, multiplicada por la constante  $k$ , es  $S_n$ . La idea de aplicar medianas repetidamente fue introducida por Tukey (1977) y utilizada por Siegel (1982) para definir estimadores robustos de regresión. Rousseeuw y Basset (1990) proponen la utilización de medianas repetidas para estimar la localización en grandes conjuntos de datos.

Por otra parte, la idea que lleva a definir  $Q_n$  es similar a la que conduce a la definición del estimador de localización de Hodges-Lehmann (1963),

$$\text{med}\left\{\frac{y_i + y_j}{2} : i < j\right\}.$$

Este estimador puede considerarse una versión “suavizada” de la mediana pues consiste en la mediana de promedios de pares de observaciones. Un estimador de dispersión “suave” se obtiene sustituyendo en el estimador de Hodges-Lehmann los promedios por las diferencias en valor absoluto entre pares de observaciones; este es precisamente el estimador  $Q_n$ .

Sean  $B_S^+(\epsilon)$  y  $B_S^-(\epsilon)$  las curvas de sesgo respecto a *outliers* e *inliers* del estimador  $S_n$ . El siguiente resultado nos permite calcularlas:

**Proposición 5 (Rousseeuw y Croux (1993), teorema 4)** *Se verifica que*

$$B_S^+(\epsilon) = kg^+ \left[ F_0^{-1} \left( \frac{3 - 2\epsilon}{4(1 - \epsilon)} \right) \right],$$

donde la función  $g^+$  se define implícitamente por la relación

$$F_0(x + g^+(x)) - F_0(x - g^+(x)) = \frac{1}{2(1 - \epsilon)}.$$

La curva de sesgo respecto a *inliers* viene dada por

$$B_S^-(\epsilon) = kg^- \left[ F_0^{-1} \left( \frac{3 - 4\epsilon}{4(1 - \epsilon)} \right) \right],$$

donde la función  $g^-$  se define implícitamente por la relación

$$F_0(x + g^-(x)) - F_0(x - g^-(x)) = \frac{1 - 2\epsilon}{2(1 - \epsilon)}.$$

Sobre los sesgos respecto a *outliers* e *inliers* de  $Q_n$ , que denotaremos  $B_Q^+(\epsilon)$  y  $B_Q^-(\epsilon)$  respectivamente, Rousseeuw y Croux (1993) dan el siguiente resultado:

**Proposición 6 (Rousseeuw y Croux (1993), teorema 7)** Sea  $F_0^*$  la convolución  $F_0 * F_0$ . Entonces el sesgo respecto a *outliers* verifica

$$B_Q^+(\epsilon) = d(F_0^*)^{-1} \left( \frac{5 - 8\epsilon + 4\epsilon^2}{8(1 - \epsilon)^2} \right),$$

y el sesgo respecto a *inliers* viene dado por la solución de la ecuación

$$(1 - \epsilon)^2 F_0^*[d^{-1}B_Q^-(\epsilon)] + 2\epsilon(1 - \epsilon)F_0[B_Q^-(\epsilon)] + \epsilon^2 = 5/8.$$

### 1.3.2 Estimadores de localización

Nuestro interés se centra ahora en el estudio del sesgo de *M-estimadores de localización*. Esta clase de estimadores se define (véase Huber, 1964) como

$$T_n = \inf \left\{ t : \sum_{i=1}^n \psi \left( \frac{y_i - t}{D_n} \right) < 0 \right\},$$

donde  $D_n$  es un estimador robusto del parámetro de dispersión,  $\sigma$ , equivariante por cambios de escala, y la función  $\psi$  satisface la hipótesis 2. La inclusión del estimador de dispersión  $D_n$  es necesaria para obtener estimadores de localización que sean equi-variantes por traslaciones y por cambios de escala.

La versión funcional correspondiente a un M-estimador de localización es

$$T_\psi(F) = \inf\{t : E_F \psi\left(\frac{y-t}{D(F)}\right) < 0\},$$

donde  $D(F)$  es el funcional de dispersión a partir del cual se define  $D_n$ .

Martin y Zamar (1993a) prueban que el sesgo máximo de  $T$ ,  $B_\psi(\epsilon)$ , se obtiene contaminando en  $H = \delta_x$  y haciendo  $x \rightarrow \infty$ . Por lo que se debe verificar

$$(1 - \epsilon)E_{F_0} \psi\left(\frac{y - B_\psi(\epsilon)}{D^+(\epsilon)}\right) + \epsilon = 0,$$

donde  $D^+(\epsilon)$  es el sesgo respecto a *outliers* del estimador de dispersión utilizado.

Se puede enunciar el siguiente resultado:

**Proposición 7 (Martin y Zamar (1993a), lema 1)** *Definamos la siguiente función:*

$$g_\psi(t, s) = -E_{F_0} \psi\left(\frac{y-t}{s}\right) = s \int_0^\infty \psi(y)[f_0(sy-t) - f_0(sy+t)]dy.$$

Si  $\psi$  verifica la hipótesis 2 y  $0 < D^+(\epsilon) < \infty$ , entonces

$$B_\psi(\epsilon) = g_\psi^{-1}\left(\frac{\epsilon}{1-\epsilon}, D^+(\epsilon)\right),$$

donde  $g_\psi^{-1}(x, s)$  es la inversa de la función  $g_\psi(\cdot, s)$ .

De la proposición 7 se deduce que el punto de ruptura de  $T$  es  $\epsilon^* = 1/2$ . También de esta proposición se obtiene que el sesgo de la mediana es

$$B_0(\epsilon) = F_0^{-1}\left(\frac{1}{2(1-\epsilon)}\right), \quad (1.15)$$

que es la expresión encontrada utilizando otros procedimientos por Huber (1981). Recordemos que Huber (1964) también probó que para todo estimador de localización,  $T$ , equivariante por traslaciones y cambios de escala, se cumple

$$B_0(\epsilon) \leq B_T(\epsilon), \quad \text{para todo } 0 < \epsilon < 1/2.$$

Esto significa que para cualquier proporción de contaminación, la mediana tiene sesgo minimax entre todos los estimadores de localización equivariantes.

## 1.4 Robustez respecto al sesgo en el modelo de regresión lineal

De aquí en adelante escribimos en **negrita** aquellas variables que toman valores vectoriales. También se escriben en **negrita** los parámetros vectoriales y sus estimadores, así como los funcionales que los definen. Las letras mayúsculas en cursiva se reservan para matrices y para funciones de distribución.

### 1.4.1 Curva de sesgo en el modelo de regresión lineal

Sean  $(y_1, \mathbf{x}_1), \dots, (y_n, \mathbf{x}_n)$ ,  $\mathbf{x}_i \in \mathbb{R}^p$ ,  $y_i \in \mathbb{R}$ , observaciones independientes que satisfacen el siguiente modelo lineal:

$$y_i = \boldsymbol{\theta}'_0 \mathbf{x}_i + u_i, \quad 1 \leq i \leq n, \quad (1.16)$$

donde los errores  $u_i$  se distribuyen según  $F_0$  y son independientes de las variables regresoras  $\mathbf{x}_i$ . Supongamos que las variables  $\mathbf{x}_i$ , para  $i = 1, \dots, n$ , son vectores aleatorios independientes con distribución  $G_0$  tal que existe  $E_{G_0} \mathbf{x} \mathbf{x}' = A$ . A partir de  $F_0$ ,  $G_0$ , y la independencia entre ambas se puede obtener la distribución de  $(y_i, \mathbf{x}_i)$  bajo el modelo (1.16), que denotamos por  $H_0$ . Esta distribución es

$$H_0(y, \mathbf{x}) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_p} F_0(y - \boldsymbol{\theta}'_0 \mathbf{s}) dG_0(\mathbf{s}).$$

Dado un vector de parámetros  $\boldsymbol{\theta}$ , los residuos correspondientes se denotarán por  $r_i(\boldsymbol{\theta}) = y_i - \boldsymbol{\theta}' \mathbf{x}_i$ , para  $i = 1, \dots, n$ . Cuando quede claro por el contexto cuál es el vector de parámetros, los residuos se denotarán por  $r_i$ . La distribución de  $|y_i - \boldsymbol{\theta}' \mathbf{x}_i|$  cuando  $(y_i, \mathbf{x}_i)$  se distribuye según  $H$ , se denotará por  $F_{H, \boldsymbol{\theta}}$ .

Con el fin de modelizar la aparición de datos atípicos bajo el modelo  $H_0$ , supondremos que la verdadera distribución de los datos pertenece al entorno de contami-

nación

$$V_\epsilon = \{H : H = (1 - \epsilon)H_0 + \epsilon\tilde{H}, \tilde{H} \text{ distribución arbitraria}\}.$$

La mayoría de los estimadores de regresión robustos se pueden obtener a partir de un funcional  $\mathbf{T}$  definido sobre un conjunto amplio,  $\mathcal{H}$ , de distribuciones sobre  $\mathbb{R}^{p+1}$  que incluye el entorno  $V_\epsilon$  y la clase de todas las funciones de distribución empíricas. Los estimadores del vector de parámetros  $\theta_0$  se obtienen aplicando el funcional  $\mathbf{T}$  a la distribución empírica correspondiente a la muestra de que se disponga.

Todos los funcionales que vamos a manejar se comportan de manera conveniente cuando hacemos una transformación afín de los regresores, es decir, son *equivariantes por transformaciones afines*. Más formalmente, verifican la siguiente propiedad: sea  $y^* = y + \mathbf{x}'\mathbf{b}$ , donde  $\mathbf{b} \in \mathbb{R}^{p+1}$ , y sea  $\mathbf{x}^* = C'\mathbf{x}$ , donde  $C$  es una matriz  $p \times p$  no singular. Si  $H^*$  es la distribución de  $(y^*, \mathbf{x}^*)$ , entonces  $\mathbf{T}(H^*) = C^{-1}[\mathbf{T}(H) + \mathbf{b}]$ .

El sesgo asintótico de  $\mathbf{T}$  en  $H$ ,  $b_A(\mathbf{T}, H)$ , se define de manera que sea invariante ante transformaciones afines,

$$b_A(\mathbf{T}, H) = \{[\mathbf{T}(H) - \theta_0]'A[\mathbf{T}(H) - \theta_0]\}^{1/2}.$$

Como sólo vamos a considerar estimadores equivariantes por transformaciones afines, para estudiar la curva de sesgo podemos suponer sin pérdida de generalidad que  $A$  es la matriz identidad  $I$  y  $\theta_0 = \mathbf{0}$ . Por lo tanto,  $b_A(\mathbf{T}, H) = b(\mathbf{T}, H) = \|\mathbf{T}(H)\|$ . La curva de sesgo de  $\mathbf{T}$  se define entonces como

$$B_{\mathbf{T}}(\epsilon) = \sup_{H \in V_\epsilon} b(\mathbf{T}, H) = \sup_{H \in V_\epsilon} \|\mathbf{T}(H)\|. \quad (1.17)$$

#### 1.4.2 Estimadores robustos de regresión con residuos admisibles

Yohai y Zamar (1993) introdujeron la clase de estimadores de regresión con residuos admisibles. Informalmente hablando, esta clase está formada por estimadores que

proporcionan unos residuos cuyos valores absolutos tienen una distribución empírica que no es mejorada por la correspondiente a los residuos de ningún otro vector de parámetros. La definición formal es la siguiente:

**Definición 1** *Se dice que un funcional  $\mathbf{T}$  que define un estimador de regresión, tiene residuos admisibles sobre  $\mathcal{H}$  si se verifica la siguiente condición: sea  $H_n \in \mathcal{H}$  y  $\theta^* \in \mathbb{R}^p$  tales que  $F_{H_n, \mathbf{T}(H_n)}$  y  $F_{H_n, \theta^*}$  son continuas sobre  $(0, \infty)$  y*

$$\lim_{n \rightarrow \infty} F_{H_n, \mathbf{T}(H_n)}(u) = F_1(u),$$

y

$$\lim_{n \rightarrow \infty} F_{H_n, \theta^*}(u) = F_2(u),$$

para todo  $u > 0$ , donde  $F_1$  y  $F_2$  son dos funciones de distribución posiblemente subestocásticas y continuas sobre  $(0, \infty)$ . Entonces  $F_1(u) = F_2(u)$  para todo  $u$  o existe  $u > 0$  tal que  $F_1(u) > F_2(u)$ .

Para entender mejor esta definición, consideremos el caso  $H_n = H$  para todo  $n$ . Si fuese  $F_1(u) = F_{H, \mathbf{T}(H)}(u) \leq F_{H, \theta^*}(u) = F_2(u)$ , para todo  $u > 0$ , con desigualdad estricta en algún valor de  $u$ , los residuos en valor absoluto correspondientes al vector de parámetros  $\theta^*$  serían estocásticamente menores que los correspondientes a  $\mathbf{T}(H)$ , con lo que  $\theta^*$  daría lugar a un ajuste “mejor” que  $\mathbf{T}(H)$ . La definición 1 excluye esta posibilidad.

Muchos estimadores robustos de regresión se basan en minimizar un funcional  $J(F_{H, \theta})$  que mide la escala de los residuos en valor absoluto  $|r_i(\theta)|$ ,  $i = 1, \dots, n$ , es decir,

$$\mathbf{T}(H) = \arg \min_{\theta} J(F_{H, \theta}). \quad (1.18)$$

Por ejemplo, si  $H_n$  es la función de distribución empírica de la muestra y

$$J(F_{H_n}, \boldsymbol{\theta}) = \sum_{i=1}^n (y_i - \boldsymbol{\theta}' \mathbf{x}_i)^2,$$

entonces  $\mathbf{T}(H_n)$  es el estimador de mínimos cuadrados ordinario. Con el fin de conseguir mejores propiedades de robustez, podemos elegir  $J(F_H, \boldsymbol{\theta}) = S(F_H, \boldsymbol{\theta})$ , un M-estimador de escala como los que hemos estudiado en la sección 1.2. Más adelante, veremos otros ejemplos de estimadores definidos a través de la expresión (1.18).

Yohai y Zamar (1993), teorema 5.1, probaron que si el funcional  $J$  que mide la escala de los residuos verifica ciertas condiciones que se comentarán ampliamente en el capítulo tres, entonces el funcional  $\mathbf{T}$  definido a través de la expresión (1.18) da lugar a estimadores de regresión con residuos admisibles. Esencialmente, lo que se requiere es que el funcional  $J$  sea monótono frente a la relación de dominancia estocástica.

Otra propiedad interesante de los estimadores con residuos admisibles (véase Yohai y Zamar (1993), lema 5.3) es que no tienen influencia acotada. De hecho, se verifica

$$B_{\mathbf{T}}(\epsilon) = c\sqrt{\epsilon} + o(\sqrt{\epsilon})$$

tal y como demuestran Yohai y Zamar (1992), teorema 2.2. El comportamiento de la curva de sesgo para proporciones de contaminación pequeñas no es aproximadamente lineal en este caso, como ocurre cuando la sensibilidad  $\gamma^*$  es finita, sino que se asemeja al comportamiento de  $\sqrt{\epsilon}$ . De este hecho también se desprende que cualquier estimador de regresión robusto con influencia acotada no puede tener residuos admisibles.

### 1.4.3 Ejemplos de estimadores con residuos admisibles

En este apartado vamos a definir y estudiar algunas propiedades de clases de estimadores robustos de regresión con residuos admisibles.

### S-estimadores

Como se adelantó en la sección 1.3, Rousseeuw y Yohai (1984) definieron un *S-estimador de regresión* como el vector de parámetros que minimiza un M-estimador de la escala de los residuos, es decir,

$$\mathbf{T}(H) = \arg \min_{\boldsymbol{\theta}} S(F_{H,\boldsymbol{\theta}}), \quad (1.19)$$

donde el funcional  $S(F)$  se define como en la ecuación (1.3).

Martin *et al.* (1989) obtuvieron la curva de sesgo de estos estimadores. En el caso de regresores gaussianos, su teorema 3.1 proporciona la expresión

$$B_S^2(\epsilon) = \left[ \frac{B_\chi^+(\epsilon)}{B_\chi^-(\epsilon)} \right]^2 - 1, \quad (1.20)$$

donde  $B_\chi^+(\epsilon)$  y  $B_\chi^-(\epsilon)$  son las curvas de sesgo respecto a *outliers* e *inliers* del M-estimador de escala basado en  $\chi$ . La fórmula (1.20) será obtenida como un caso particular de los métodos obtenidos en el capítulo tres y es importante porque al relacionar las curvas de sesgo correspondientes al M-estimador de escala con el sesgo del S-estimador de regresión, permite extender al contexto de regresión conclusiones obtenidas previamente en el modelo de escala, que es más sencillo. Por ejemplo, de (1.20) se deduce que el punto de ruptura de un S-estimador de regresión es el mismo que el del M-estimador de escala basado en la misma función  $\chi$ , es decir,  $\epsilon^* = \min\{b, 1-b\}$ .

Un caso particular importante de S-estimador de regresión es el LMS-estimador (iniciales procedentes de la expresión en inglés *least median of squares*) definido por Rousseeuw (1984) como

$$\hat{\boldsymbol{\theta}}_n = \arg \min_{\boldsymbol{\theta}} \text{med}_i (y_i - \boldsymbol{\theta}'\mathbf{x}_i)^2. \quad (1.21)$$

Como se vio en la expresión (1.8), las funciones  $\chi$  de salto definen estimadores que son cuantiles reescalados de la muestra. Por lo tanto, el S-estimador correspondiente

a  $b = 1/2$  y a una función  $\chi$  de salto, coincide con el LMS-estimador de Rousseeuw. Combinando la expresión (1.20) con la proposición 2, se tiene que el LMS-estimador de regresión tiene curva de sesgo minimax entre todos los S-estimadores con punto de ruptura  $\epsilon^* = 1/2$ , para cualquier valor de  $\epsilon < 1/2$ .

Dentro de la clase de S-estimadores de regresión no es posible obtener simultáneamente eficiencia y alto punto de ruptura ya que, si su punto de ruptura es  $\epsilon^* = 1/2$ , la mayor eficiencia que puede tener un S-estimador de regresión es del 32.9% (véase Hössjer (1992), tabla 1).

### $\tau$ -estimadores

Sea  $S_n$  un M-estimador de escala de los valores absolutos de los residuos basado en una función  $\chi_1 \in C_{b_1}$  y sea  $\chi_2 \in C_{b_2}$ . Se dice que  $\hat{\theta}_n$  es un  $\tau$ -estimador de regresión si

$$\hat{\theta}_n = \arg \min_{\theta} S_n^2 \sum_{i=1}^n \chi_2 \left( \frac{r_i(\theta)}{S_n} \right).$$

Esta clase de estimadores fue introducida por Yohai y Zamar (1988) con el fin de obtener buenas propiedades de eficiencia y robustez simultáneamente. La idea intuitiva en la que se basan los  $\tau$ -estimadores es la siguiente: se trata de tomar una función  $\chi_2$  que sea aproximadamente cuadrática cerca de cero. En tal caso, si son pequeños los residuos normalizados por el M-estimador de escala,  $r_i(\theta)/S_n$ , se verifica

$$S_n^2 \sum_{i=1}^n \chi_2 \left( \frac{r_i(\theta)}{S_n} \right) \approx S_n^2 \sum_{i=1}^n \frac{r_i(\theta)^2}{S_n^2} = \sum_{i=1}^n r_i(\theta)^2,$$

con lo que  $\hat{\theta}_n$  es semejante al estimador de mínimos cuadrados ordinarios, por lo que será eficiente cuando no hay datos atípicos. Por el contrario, si  $r_i(\theta)/S_n$  es grande para algún  $i$ , la influencia en la estimación de este residuo no es decisiva ya que, en tal caso,

$$\chi_2 \left( \frac{r_i(\theta)}{S_n} \right) \approx 1.$$

Este carácter adaptativo es el que confiere a los  $\tau$ -estimadores sus buenas propiedades de robustez y de eficiencia bajo el modelo. En particular, Yohai y Zamar (1988) probaron que los  $\tau$ -estimadores heredan el punto de ruptura del S-estimador basado en  $\chi_1$ , es decir,  $\epsilon^* = \min\{b_1, 1 - b_1\}$ . Sin embargo, su eficiencia bajo el modelo puede ser tan alta como se desee eligiendo convenientemente la función  $\chi_2$ .

La versión funcional de los  $\tau$ -estimadores es

$$\mathbf{T}(H) = \arg \min_{\boldsymbol{\theta}} \tau(F_{H,\boldsymbol{\theta}}), \quad (1.22)$$

donde

$$\tau(F)^2 = S(F)^2 E_F \chi_2 \left( \frac{y}{S(F)} \right). \quad (1.23)$$

Obtendremos expresiones para la curva de sesgo de los  $\tau$ -estimadores en el capítulo tres.

### Estimadores basados en rangos absolutos o R-estimadores.

Sea  $R_i^+(\boldsymbol{\theta})$  el rango absoluto o número de orden que ocupa el residuo  $|r_i(\boldsymbol{\theta})|$  entre el conjunto de residuos  $|r_1(\boldsymbol{\theta})|, \dots, |r_n(\boldsymbol{\theta})|$ . Consideremos también una función  $a(u)$  definida en el intervalo  $[0, 1]$  y con valores en  $\mathbb{R}$ . Los *estimadores de regresión basados en rangos absolutos* admiten la siguiente formulación general:

$$\hat{\boldsymbol{\theta}}_n = \arg \min_{\boldsymbol{\theta}} \sum_{i=1}^n a \left( \frac{R_i^+(\boldsymbol{\theta})}{n} \right) |y_i - \boldsymbol{\theta}' \mathbf{x}_i|^k. \quad (1.24)$$

Por lo tanto, los R-estimadores son promedios de potencias de los residuos ponderadas por una función que depende del valor de los rangos absolutos. Hössjer (1994) ha estudiado algunas de las propiedades de robustez así como la distribución asintótica de estos estimadores.

Las particularizaciones más interesantes de la fórmula general (1.24) se obtienen tomando  $k = 1$  ó  $k = 2$  y la función

$$a(u) = \begin{cases} 1, & |u| \leq 1 - \alpha \\ 0, & |u| > 1 - \alpha \end{cases}, \quad \alpha \geq 1/2. \quad (1.25)$$

En lo que sigue nos restringimos por sencillez a considerar el caso de esta función. Para  $k = 1$ , el estimador resultante es:

$$\hat{\theta}_n = \arg \min_{\theta} \sum_{l=1}^{[n(1-\alpha)]} |r(\theta)_{(l)}|,$$

es decir, el vector de parámetros que minimiza la media recortada de los valores absolutos de los residuos una vez que han sido desechados los  $[n\alpha]$  mayores. Este estimador se denomina  $\alpha$ -LTAV (nombre procedente de la expresión en inglés *least trimmed absolute values*).

Análogamente, el estimador resultante para  $k = 2$  es

$$\hat{\theta}_n = \arg \min_{\theta} \sum_{l=1}^{[n(1-\alpha)]} r^2(\theta)_{(l)},$$

que llamaremos  $\alpha$ -LTS (nombre procedente de la expresión en inglés *least trimmed squares*). Este estimador fue propuesto por Rousseeuw (1985).

Se puede ver fácilmente de manera intuitiva que cualquier estimador definido como en la ecuación (1.24), donde  $a(u)$  venga dada por (1.25), tiene como punto de ruptura  $\epsilon^* = 1 - \alpha$ .

La versión funcional de los R-estimadores es

$$\mathbf{T}(H) = \arg \min R(F_{H,\theta}), \quad (1.26)$$

donde

$$R(F) = \int_0^\infty a[F(u)]u^k dF(u). \quad (1.27)$$

Las curvas de sesgo de los R-estimadores también se obtienen en el capítulo tres.

#### 1.4.4 Otros estimadores robustos de regresión

Todos los estimadores definidos en el apartado anterior tienen residuos admisibles, tal y como prueban Yohai y Zamar (1993). Hay otras clases de estimadores robustos que no los tienen. En este apartado veremos primero los S-estimadores generalizados. Esta clase será ampliada en el capítulo cuatro, en el que también se obtiene un método para calcular la curva de sesgo de cualquier estimador perteneciente a la clase generalizada. Más adelante en este apartado se introducen y comentan algunas propiedades de los M- y GM-estimadores de regresión. En general, estos estimadores tampoco tienen residuos admisibles. Para terminar, se reseñarán muy brevemente otras estrategias para estimar de manera robusta los parámetros del modelo de regresión lineal.

##### S-estimadores generalizados

Recogiendo una idea ya presente en Serfling (1984), Croux *et al.* (1994) definen la clase de S-estimadores generalizados o GS-estimadores. Esencialmente, un GS-estimador es un S-estimador calculado a partir de las diferencias de los residuos  $\{|r_i(\boldsymbol{\theta}) - r_j(\boldsymbol{\theta})| : i < j\}$  en lugar de hacerlo directamente sobre los residuos. Más formalmente, puede darse la siguiente definición: sea  $\chi \in C_b$ , se dice que  $\hat{\boldsymbol{\theta}}$  es un *GS-estimador de regresión* si

$$\hat{\boldsymbol{\theta}}_n = \arg \min_{\boldsymbol{\theta}} S_n(\boldsymbol{\theta}), \quad (1.28)$$

donde  $S_n(\boldsymbol{\theta})$  resuelve la ecuación

$$\binom{n}{2}^{-1} \sum_{i < j} \chi \left( \frac{|r_i(\boldsymbol{\theta}) - r_j(\boldsymbol{\theta})|}{S_n(\boldsymbol{\theta})} \right) = b. \quad (1.29)$$

Para evitar problemas con funciones  $\chi$  discontinuas, puede sustituirse la ecuación (1.29) por la definición

$$S_n(\boldsymbol{\theta}) = \inf\left\{s : \binom{n}{2}^{-1} \sum_{i < j} \chi\left(\frac{|r_i(\boldsymbol{\theta}) - r_j(\boldsymbol{\theta})|}{S_n(\boldsymbol{\theta})}\right) < b\right\}.$$

Con el fin de que el correspondiente estimador de escala sea consistente, debe elegirse  $b = E_{H_0} \chi(y_i - y_j)$ . Por otra parte, el valor de  $b$  determina, al igual que ocurría en el caso de S-estimadores de regresión, el punto de ruptura del estimador. De hecho, se prueba que

$$\epsilon^* = \min\{(1 - b)^{1/2}, 1 - (1 - b)^{1/2}\}.$$

Por lo tanto, hay que escoger  $b = 3/4$  para obtener  $\epsilon^* = 1/2$ .

El estimador de dispersión  $Q_n$  definido en (1.14) es un caso de GS-estimador de dispersión basado en una función  $\chi$  de salto. Al definir  $Q_n$ , se eligió el cuantil  $p$  de una forma aparentemente arbitraria. Sin embargo, aquella era la elección apropiada para obtener  $\epsilon^* = 1/2$ .

Como hemos visto, los S-estimadores no pueden ser muy eficientes cuando el punto de ruptura es  $\epsilon^* = 1/2$ . Sin embargo, pueden obtenerse GS-estimadores con punto de ruptura igual a  $1/2$  y sin embargo bastante eficientes. Además, los GS-estimadores son asintóticamente normales incluso cuando los S-estimadores definidos por la misma función  $\chi$  no lo son. Otra ventaja añadida de esta modificación consiste en que la función objetivo a minimizar no depende de la inclusión de un término independiente en el modelo (1.16) por lo que si consideramos el modelo lineal con término independiente

$$y_i = \alpha_0 + \boldsymbol{\theta}'_0 \mathbf{x}_i + u_i, \quad 1 \leq i \leq n,$$

el parámetro  $\alpha_0$  puede ser estimado fácilmente en una segunda etapa tras haber obtenido una estimación de  $\boldsymbol{\theta}_0$ .

En el capítulo cuatro se probará que los GS-estimadores no tienen residuos admisibles. Además, se estudiarán las curvas de sesgo cuando calculamos S-,  $\tau$ - y R-estimadores sobre las diferencias de los residuos en lugar de sobre los residuos directamente.

### M- y GM-estimadores

Por extensión directa de los M-estimadores de localización se obtienen los M-estimadores de regresión. Diremos que  $\hat{\theta}_n$  es un *M-estimador de regresión* si

$$\hat{\theta}_n = \arg \min_{\theta} \sum_{i=1}^n \rho \left( \frac{r_i(\theta)}{S_n} \right), \quad (1.30)$$

donde  $S_n$  es un estimador de escala de los residuos. Si la función  $\rho$  es diferenciable, con derivada  $\rho' = \psi$ , entonces la condición de primer orden que debe verificar  $\hat{\theta}_n$  se puede escribir

$$\sum_{i=1}^n \psi \left( \frac{r_i(\hat{\theta}_n)}{S_n} \right) \mathbf{x}_i = 0. \quad (1.31)$$

De esta ecuación se pueden derivar las condiciones que debe cumplir  $\rho$  para que un M-estimador sea robusto. Si, como es el caso habitual en localización,  $\rho$  es una función convexa no acotada (por ejemplo, cuadrática para argumentos pequeños y lineal para argumentos grandes) entonces, la presencia del factor  $\mathbf{x}_i$  en la ecuación (1.31) implica que el punto de ruptura es  $\epsilon^* = 0$ , debido a la influencia de los puntos con alto potencial (*high leverage points*). Para asegurar que  $\epsilon^* > 0$ , la función  $\rho$  debe ser acotada y, como consecuencia, la derivada  $\psi$  ha de ser redescendiente de manera que para alguna constante  $c > 0$ ,  $\psi(u) = 0$  si  $|u| > c$ . En este caso, el cálculo de  $\hat{\theta}_n$  es más difícil porque se debe minimizar una función que no es convexa.

Puede probarse que si  $\rho$  cumple la hipótesis 1, entonces  $\hat{\theta}_n$ , tal y como ha sido definido en (1.30), tiene residuos admisibles. Los MM-estimadores definidos por Yohai

(1987) se encuentran en esta situación. Algunas propiedades de su curva de sesgo se demuestran en la sección 3.4 del capítulo tres. Si la función  $\rho$  no es acotada (por ejemplo, para el caso de mínimos cuadrados ordinarios) entonces  $\hat{\theta}$  no tiene residuos admisibles (véase Yohai y Zamar, 1993).

Otra posibilidad para robustificar la condición (1.31) consiste en utilizar los M-estimadores generalizados o GM-estimadores. Se dice que  $\hat{\theta}$  es un *GM-estimador de regresión* si verifica la condición

$$\sum_{i=1}^n w(\|\mathbf{x}_i\|) \psi \left( \frac{r_i(\hat{\theta})}{S_n v(\|\mathbf{x}_i\|)} \right) \mathbf{x}_i = 0,$$

donde  $v$  y  $w$  son funciones definidas en  $\mathbb{R}^+$  con valores en  $\mathbb{R}^+$  tales que  $w(u) \rightarrow 0$  si  $u \rightarrow \infty$ . Se trata de amortiguar la influencia de observaciones con alto potencial mediante las funciones de ponderación  $v$  y  $w$ . Las posibilidades  $v(u) = 1$ ,  $v(u) = w(u)$  y  $v(u) = 1/w(u)$  han sido sugeridas respectivamente por Mallows (1975), Schweppe, y Hill y Ryan (Hill, 1977).

Los GM-estimadores tienen influencia acotada y por lo tanto no tienen residuos admisibles. Por otra parte, Maronna, Bustos y Yohai (1979) probaron que el punto de ruptura de los GM-estimadores converge a cero rápidamente cuando la dimensión de los regresores aumenta.

Una forma de resolver este problema consiste en efectuar un sólo paso del algoritmo Newton-Raphson utilizado para calcular el GM-estimador  $\hat{\theta}_n$ . Las propiedades de este procedimiento han sido investigadas por Simpson, Ruppert y Carroll (1992) y Coakley y Hettmansperger (1993). Simpson y Yohai (1993) obtienen algunas propiedades de la curva de sesgo de estos estimadores.

Finalmente, otros estimadores de regresión que no tienen residuos admisibles se basan en proyecciones y han sido propuestos por Maronna y Yohai (1993).

## Capítulo 2

# La tasa de ruptura de la curva de sesgo asintótico máximo

### 2.1 Introducción

En la literatura sobre robustez se han considerado problemas diversos en los que el objetivo es encontrar el estimador más eficiente, bajo el modelo central de un entorno de contaminación, dentro de una clase de estimadores dada en la que se verifican ciertas restricciones sobre las propiedades de robustez. Por ejemplo, Hampel (1968) propuso y resolvió el siguiente problema:

$$\min_{T \in \mathcal{T}} V(T, F_0) \quad \text{s.a.} \quad \gamma_T^* \leq \gamma_0, \quad (2.1)$$

donde  $\mathcal{T}$  es la clase de M-estimadores de escala o, alternativamente, la de M-estimadores de localización, y  $V(T, F_0)$  es la varianza asintótica de  $T$  bajo la distribución nominal  $F_0$ . En este caso, la restricción  $\gamma_T^* \leq \gamma_0$  implica que  $T$  debe tener influencia acotada y además su sensibilidad debe estar por debajo del umbral  $\gamma_0$ . Problemas análogos al de Hampel en el contexto de regresión han sido estudiados por Krasker (1980) y Krasker y Welsch (1982).

Otro ejemplo de este tipo de problemas se debe a Hössjer (1992) y ya fue mencionado en la introducción. Se trata de resolver

$$\min_{T \in \mathcal{T}} V(T, F_0) \quad \text{s.a.} \quad \epsilon_T^* = 1/2, \quad (2.2)$$

donde ahora  $\mathcal{T}$  es la clase de S-estimadores de regresión. En este caso, la restricción sobre las propiedades de robustez de los estimadores involucra al punto de ruptura y, por lo tanto, se refiere al comportamiento de  $T$  para proporciones de contaminación altas.

Para plantear estos problemas ha sido necesario disponer de medidas, tales como la sensibilidad o el punto de ruptura, que resuman el comportamiento de la curva de sesgo. Hemos necesitado concentrar la información sobre las propiedades de robustez del estimador en cantidades sencillas que se obtuvieran fácilmente a partir del propio estimador.

En este capítulo estamos interesados en el problema de resumir el comportamiento de la curva de sesgo. Cabría esperar que estimadores con los mismos valores de  $\gamma^*$  y  $\epsilon^*$  tuviesen curvas de sesgo similares puesto que  $\gamma^*$  se relaciona con  $B(\epsilon)$  para  $\epsilon \approx 0$  y  $\epsilon^*$  se relaciona con  $B(\epsilon)$  para  $\epsilon$  grande. Sin embargo, esto no es necesariamente así, tal y como se ilustra en el siguiente ejemplo en el contexto de estimación de un parámetro de escala.

Consideremos que la distribución central del entorno es  $F^\sigma(y) = \Phi(y/\sigma)$ , donde  $\Phi$  denota la función de distribución correspondiente a la distribución normal estándar. Hemos considerado tres M-estimadores de escala  $\hat{\sigma}_H$ ,  $\hat{\sigma}_L$ , y  $\hat{\sigma}_T$ . Los tres tienen punto de ruptura  $\epsilon^* = 1/2$  y sensibilidades muy similares. Las correspondientes funciones  $\chi$  se recogen en la tabla 2.1. La función  $\chi_H$  recibe el nombre de función de Huber y no es más que una parábola truncada, es decir, corresponde a la función “score” utilizada para obtener el estimador de máxima verosimilitud bajo normalidad pero modificada para

obtener mejores propiedades de robustez. La función  $\chi_T$  se denomina en la literatura función bicuadrada de Tukey (véase Beaton y Tukey, 1974) y proporciona al estimador buenas propiedades de robustez y eficiencia. Por último,  $\chi_L$  es la función valor absoluto truncada y corresponde a la función “score” del estimador de máxima verosimilitud de la distribución doble exponencial (o de Laplace) modificada convenientemente para obtener robustez.

Las tres funciones dependen de una constante que ha sido ajustada de forma que  $\chi_H, \chi_T, \chi_L \in C_{1/2}$ . Estas constantes, así como las sensibilidades de los tres estimadores, se recogen también en la tabla 2.1.

Dado que los tres estimadores tienen sensibilidades parecidas, sus curvas de sesgo respecto a *outliers* son similares para pequeños valores de  $\epsilon$  (véase la figura 2.1). Por otra parte, las curvas de sesgo de  $\hat{\sigma}_H$  y  $\hat{\sigma}_T$  son bastante parecidas para todo  $\epsilon$ , mientras que la de  $\hat{\sigma}_L$  es bastante diferente para valores de  $\epsilon$  mayores que, digamos, 0.3.

Lo que ocurre es que el punto de ruptura sólo informa sobre la localización de la asíntota vertical de  $B_T(\epsilon)$  pero no da información sobre el comportamiento de  $B_T(\epsilon)$  cuando  $\epsilon \approx \epsilon^*$ .

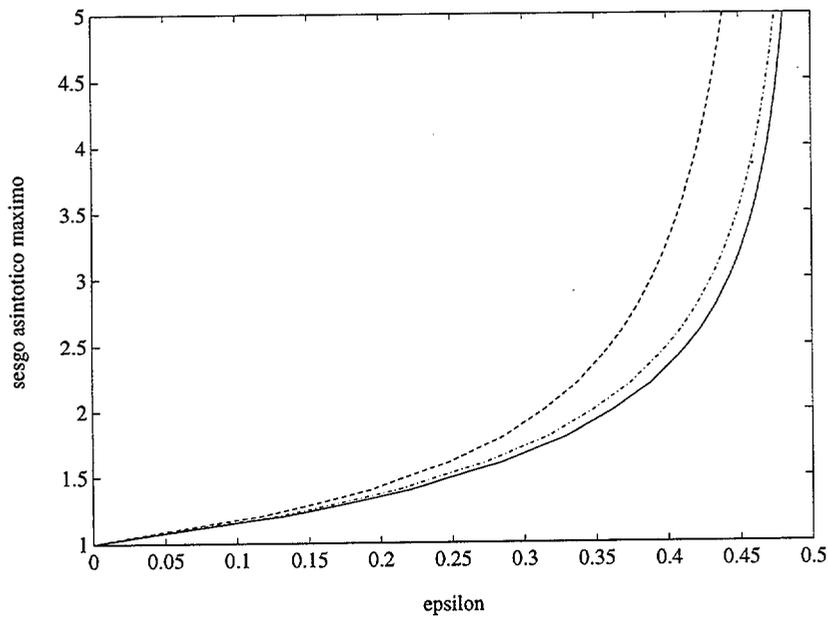
Dados dos estimadores,  $T_1$  y  $T_2$ , con el mismo punto de ruptura  $\epsilon^*$ , resulta interesante comparar las velocidades con las que sus respectivas curvas de sesgo  $B_1(\epsilon)$  y  $B_2(\epsilon)$ , convergen a infinito cuando  $\epsilon \rightarrow \epsilon^*$ . Esta comparación matizará la información proporcionada por  $\epsilon^*$  respecto al comportamiento de  $T_1$  y  $T_2$  cuando hay muchos datos contaminados en la muestra.

Las consideraciones anteriores motivan la siguiente definición:

**Definición 2** *Dados dos estimadores,  $T_1$  y  $T_2$ , con el mismo punto de ruptura,  $\epsilon^*$ , y con curvas de sesgo  $B_{T_1}(\epsilon)$  y  $B_{T_2}(\epsilon)$ , la tasa de ruptura relativa entre  $T_1$  y  $T_2$  se define*

Función	Constante	$\gamma^*$
$\chi_H(y) = \min(y^2/c^2, 1)$	1.041	1.23
$\chi_T(y) = \min(3y^2/c^2 - 3y^4/c^4 + y^6/c^6, 1)$	1.547	1.28
$\chi_L(y) = \min( y/c , 1)$	1.470	1.39

**Tabla 2.1:** Funciones que definen tres conocidos M-estimadores de escala junto con las constantes adecuadas para obtener punto de ruptura 1/2 y las correspondientes sensibilidades.



**Figura 2.1:** Curvas de sesgo asimótico máximo respecto a *outliers* para  $\hat{\sigma}_L$  (línea discontinua),  $\hat{\sigma}_T$  (línea de puntos y rayas), and  $\hat{\sigma}_H$  (línea continua).

como

$$TRR(T_1, T_2) = \lim_{\epsilon \rightarrow \epsilon^*} \frac{B_{T_1}(\epsilon)}{B_{T_2}(\epsilon)}.$$

A veces resulta útil considerar un estimador base  $T_0$  y definir la tasa de ruptura de  $T$  como

$$TR(T) \equiv TRR(T, T_0)$$

Una posible elección de  $T_0$  es considerar el estimador de sesgo minimax con punto de ruptura  $\epsilon^*$ . Si  $TR(T_1)$  ó  $TR(T_2)$  es finita, entonces

$$TRR(T_1, T_2) = \frac{TR(T_1)}{TR(T_2)}$$

y la tasa de ruptura relativa no contiene ninguna información adicional. Sin embargo, hay situaciones en las que tanto  $TR(T_1)$  como  $TR(T_2)$  son ambas iguales a infinito, en cuyo caso  $TRR(T_1, T_2)$  permite una comparación directa entre  $T_1$  y  $T_2$ .

Volviendo al ejemplo anterior, puede probarse fácilmente utilizando los resultados de la sección 2.2, que  $TRR(\chi_T, \chi_H) = 1.16$  y  $TRR(\chi_L, \chi_H) = \infty$  de manera que la tasa de ruptura recoge las analogías y diferencias observadas en la figura 2.1. Además, veremos que estos resultados están determinados por el comportamiento cerca de cero de las funciones  $\chi$  correspondientes. En otras palabras, podemos conocer el comportamiento del sesgo de un M-estimador de escala cerca del punto de ruptura observando la forma de la función que lo define cerca de cero.

El presente capítulo está estructurado de la siguiente forma: en la sección 2.2 se obtienen las tasas de ruptura correspondientes a M-estimadores de escala. En la sección 2.3 se hacen algunas observaciones sobre la tasa de ruptura de  $\tau$ -estimadores de escala. En las secciones 2.4 y 2.5 se estudian las tasas de ruptura entre algunos de los estimadores de dispersión y localización definidos en el capítulo uno. Finalmente, la sección 2.6 trata la cuestión de cuál es el punto de ruptura cuando la localización y la dispersión se estiman simultáneamente.



## 2.2 Tasa de ruptura de M-estimadores de escala

Consideremos la clase de M-estimadores de escala definida en el capítulo uno y supongamos que los datos proceden de la distribución  $F \in V_\epsilon$ , donde  $F^\sigma(y) = F_0(y/\sigma)$  y  $F_0$  satisface la siguiente hipótesis.

**Hipótesis 3**  $F_0$  corresponde a una variable aleatoria positiva, es estrictamente creciente sobre  $(0, \infty)$  y derivable con función de densidad decreciente  $f_0$ .

Consideremos la clase  $C_b$  de funciones  $\chi$  que verifican la hipótesis 1 y la condición de consistencia de Fisher,  $E_{F_0}\chi(y) = b$ . Todas las funciones de  $C_b$  definen M-estimadores de escala con los mismos puntos de ruptura respecto a *outliers* e *inliers*.

**Nota:** Todos los resultados de esta sección son válidos si reemplazamos la hipótesis 3 por la siguiente:  $F_0$  es estrictamente creciente y derivable con función de densidad  $f_0$  simétrica y estrictamente unimodal. De esta forma, podemos aplicar los resultados a muchas variables aleatorias que también toman valores negativos como la normal o la doble exponencial. Sólo son necesarios algunos mínimos ajustes: la función de salto tomada como base debe saltar en los puntos  $\pm F_0^{-1}(1 - b/2)$  y la constante  $c_k$  que aparece en la proposición 8 debe ser multiplicada por dos.

### 2.2.1 Tasa de ruptura respecto a outliers

La mayoría de las funciones  $\chi$  propuestas en la literatura pueden escribirse de la siguiente forma

$$\chi(y) = ry^k + o(y^k), \quad \text{cuando } y \rightarrow 0^+ \quad (2.3)$$

para algún  $k > 0$  y algún  $r > 0$ . Este es el tipo de comportamiento que se intenta recoger en la siguiente definición:

**Definición 3** Se dice que  $k$  es el orden local de  $\chi$  en cero si para algún  $\delta > 0$ ,  $\chi$  es derivable en  $[0, \delta)$  al menos  $k + 1$  veces (en el punto 0, basta con que las derivadas existan por la derecha), y además  $\chi^{(j)}(0^+) = 0$  para todo  $j < k$  y  $\chi^{(k)}(0^+) > 0$ .

Como es obvio, la magnitud del orden local refleja el comportamiento de  $\chi$  en un entorno de cero. Si  $k$  es el orden local de  $\chi$  en cero, podemos escribir  $\chi$  como en (2.3) con  $r = \chi^{(k)}(0^+)/k! > 0$ .

Puede verificarse fácilmente que para las funciones del ejemplo de la introducción de este capítulo,  $\chi_L$  tiene orden local  $k = 1$  mientras que  $\chi_H$  y  $\chi_T$  tienen orden local  $k = 2$ . Consideremos ahora dos ejemplos más tomados de Croux (1994),  $\chi_C(y) = y^2/(y^2 + c^2)$  y  $\chi_W(y) = 1 - \exp(-y^2/c)$ , que, al contrario que  $\chi_L$ ,  $\chi_H$  y  $\chi_T$  no son polinomios cerca de cero. Estas funciones definen estimadores que serán utilizados como ejemplos más adelante. Dado que  $\chi'_C(0^+) = \chi'_W(0^+) = 0$ , y  $\chi_C^{(2)}(0^+) = 2/c^2$ ,  $\chi_W^{(2)}(0^+) = 2/c$ , el orden local de  $\chi_C$  y  $\chi_W$  es  $k = 2$ . Los correspondientes estimadores de escala,  $\hat{\sigma}_C$  y  $\hat{\sigma}_W$  tienen punto de ruptura  $1/2$  si  $c = 0.61$  y  $c = 0.66$ , respectivamente.

En la siguiente proposición se da una aproximación, válida para proporciones de contaminación  $\epsilon$  cercanas al punto de ruptura, de la curva de sesgo respecto a *outliers* de M-estimadores de escala.

**Proposición 8** Sea  $\chi \in C_b$  tal que su orden local en cero es  $k < \infty$  y supongamos que  $F_0$  satisface la hipótesis 3 y que  $\lim_{y \rightarrow \infty} y^{k+1} f_0(y) = 0$ . Entonces

$$\lim_{\epsilon \rightarrow b} \left( \frac{1 - \epsilon}{b - \epsilon} \right)^{-1/k} B_{\chi}^+(\epsilon) = c_k(\chi), \quad (2.4)$$

donde  $c_k$  es una constante, independiente de  $\epsilon$ , dada por

$$c_k = c_k(\chi) = \left[ \frac{\chi^{(k)}(0^+)}{k!} \int_0^{\infty} y^k f_0(y) dy \right]^{1/k}. \quad (2.5)$$

**Demostración:**

Dado que  $k$  es el orden local de  $\chi$  en cero, existe  $\delta > 0$  tal que el siguiente desarrollo de Taylor es válido si  $y < \delta s$ :

$$\chi\left(\frac{y}{s}\right) = \chi^{(k)}(0^+) \frac{y^k}{k!s^k} + o\left(\frac{y^k}{s^k}\right), \quad \text{si } y/s \rightarrow 0.$$

Por tanto, la función  $h_\chi$ , a partir de la cual se obtiene el sesgo, puede escribirse como

$$\begin{aligned} h_\chi(s) &= \int_0^{\delta s} \chi\left(\frac{y}{s}\right) f_0(y) dy + \int_{\delta s}^\infty \chi\left(\frac{y}{s}\right) f_0(y) dy \\ &= \frac{\chi^{(k)}(0^+)}{k!s^k} \int_0^{\delta s} y^k f_0(y) dy + \int_0^{\delta s} o\left(\frac{y^k}{s^k}\right) f_0(y) dy \\ &\quad + \int_{\delta s}^\infty \chi\left(\frac{y}{s}\right) f_0(y) dy = I_1(s) + I_2(s) + I_3(s). \end{aligned}$$

Si multiplicamos ambos miembros de esta ecuación por  $s^k$  y tomamos límites cuando  $s \rightarrow \infty$ , obtenemos

$$\lim_{s \rightarrow \infty} s^k I_1(s) = \frac{\chi^{(k)}(0^+)}{k!} \int_0^\infty y^k f_0(y) dy < \infty, \quad (2.6)$$

ya que  $y^{k+1} f_0(y) \rightarrow 0$  si  $y \rightarrow \infty$ . Además

$$\lim_{s \rightarrow \infty} s^k I_2(s) = \lim_{s \rightarrow \infty} \int_0^{\delta s} \frac{s^k}{y^k} o\left(\frac{y^k}{s^k}\right) y^k f_0(y) dy = 0, \quad (2.7)$$

por el teorema de la convergencia dominada, y

$$\lim_{s \rightarrow \infty} s^k I_3(s) = \lim_{s \rightarrow \infty} \int_{\delta s}^\infty s^k \chi\left(\frac{y}{s}\right) f_0(y) dy \leq \lim_{s \rightarrow \infty} s^k [1 - F_0(\delta s)] = 0, \quad (2.8)$$

aplicando la regla de L'Hôpital y dado que  $y^{k+1} f_0(y) \rightarrow 0$  si  $y \rightarrow \infty$ .

Por (2.6), (2.7) y (2.8)

$$\lim_{s \rightarrow \infty} s^k h_\chi(s) = \frac{\chi^{(k)}(0^+)}{k!} \int_0^\infty y^k f_0(y) dy \quad (2.9)$$

Reemplazamos ahora  $s$  por  $B_x^+(\epsilon)$  en la ecuación (2.9). Por la proposición 1

$$\lim_{\epsilon \rightarrow b} [B_x^+(\epsilon)]^k \frac{b - \epsilon}{1 - \epsilon} = \frac{\chi^{(k)}(0^+)}{k!} \int_0^\infty y^k f_0(y) dy = [c_k(\chi)]^k.$$

El resultado se obtiene de forma inmediata a partir de esta ecuación.

□

Los casos  $k = 1$  y  $k = 2$  son particularmente importantes porque la mayoría de las funciones  $\chi$  que aparecen en la literatura tienen estos órdenes. Si el orden local de  $\chi$  es  $k = 1$ , entonces, salvo constantes,

$$B_x^+(\epsilon) \approx \frac{1}{b - \epsilon},$$

para  $\epsilon$  cercano al punto de ruptura. Si  $k = 2$ , entonces, salvo constantes,

$$B_x^+(\epsilon) \approx \left( \frac{1}{b - \epsilon} \right)^{1/2},$$

para  $\epsilon$  cercano al punto de ruptura. En general, funciones  $\chi$  con órdenes locales grandes producen estimadores con mejor comportamiento de la curva de sesgo cerca del punto de ruptura. Como las funciones  $\chi$  con órdenes locales grandes son más “planas” en un entorno de cero, se deduce que el comportamiento de la curva de sesgo respecto a *outliers* será mejor cuanto más “plana” sea la función  $\chi$  que define el estimador en un entorno de cero.

Una vez fijado el orden local en cero de  $\chi$ , las hipótesis que debe cumplir  $F_0$  no son restrictivas en cuanto al comportamiento de las colas se refiere. Por ejemplo, para el importante caso en que  $k = 2$ ,  $F_0$  podría ser una distribución  $t$  de Student con tres grados de libertad.

Una cuestión de interés teórico es lo que sucede cuando el orden local de  $\chi$  no es un entero. Por ejemplo, podríamos considerar una función cuyo comportamiento

local cerca de cero sea aproximadamente  $y^{1/2}$ . De hecho, cuando  $\chi(y) = ry^\alpha + o(y^\alpha)$  en un entorno de cero, para algún  $\alpha > 0$ , se puede probar siguiendo las líneas de la demostración de la proposición 8, que

$$\lim_{\epsilon \rightarrow b} \left( \frac{1-\epsilon}{b-\epsilon} \right)^{-1/\alpha} B_\chi^+(\epsilon) = c_\alpha(\chi),$$

donde  $c_\alpha = (r \int_0^\infty y^\alpha f_0(y) dy)^{1/\alpha}$ .

De la proposición 8 se deducen los siguientes corolarios relativos a las tasas de ruptura respecto a *outliers* de los M-estimadores de escala.

**Corolario 1** *Supongamos que el orden local de  $\chi_i$  es  $k_i < \infty$  para  $i = 1, 2$  y que se verifican las hipótesis de la proposición 8. Entonces:*

- (a) *Si  $k_1 = k_2 = k$ ,  $TRR^+(\chi_1, \chi_2) = [\chi_1^{(k)}(0^+)/\chi_2^{(k)}(0^+)]^{1/k}$ .*
- (b) *Si  $k_1 < k_2$ ,  $TRR^+(\chi_1, \chi_2) = \infty$ .*

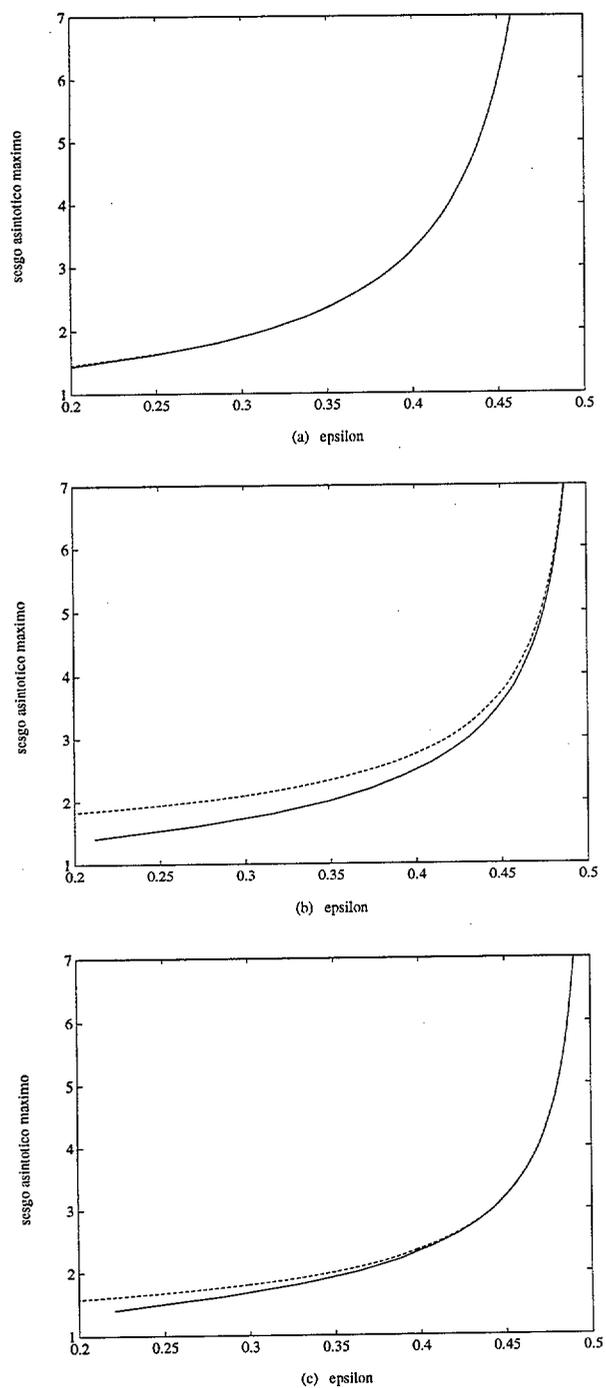
Para todo  $0 < b < 1$ , la función de salto

$$\chi_a(y) = \begin{cases} 0, & \text{si } |y| < a \\ 1, & \text{si } |y| \geq a \end{cases},$$

con  $a = F_0^{-1}(1-b)$ , pertenece a  $C_b$ . Teniendo en cuenta la proposición 2, parece natural elegir  $\chi_a$  para definir el estimador base en la definición de tasa de ruptura. Respecto a esta tasa de ruptura, el siguiente corolario también es inmediato:

**Corolario 2** *Supongamos que el orden local de  $\chi$  es  $k < \infty$  y que estamos bajo las hipótesis de la proposición 8. Entonces,  $TR^+(\chi) = \infty$ .*

Las aproximaciones dadas en la proposición 8 para los estimadores  $\hat{\sigma}_L$ ,  $\hat{\sigma}_T$ , y  $\hat{\sigma}_H$  se han representado en la figura 2.2



**Figura 2.2:** Aproximaciones (línea discontinua) de las curvas de sesgo (línea continua) para  $\hat{\sigma}_L$  (a),  $\hat{\sigma}_T$  (b), y  $\hat{\sigma}_H$  (c), obtenidas a partir de la proposición 8.

Por la proposición 8, la curva de sesgo respecto a *outliers* de  $\hat{\sigma}_L$  se comporta (salvo constantes y para  $\epsilon$  grande) como el cuadrado de la curva de sesgo respecto a *outliers* de  $\hat{\sigma}_T$  y  $\hat{\sigma}_H$ . Más precisamente, si  $F_0$  es la distribución normal estándar, se puede deducir de la proposición 8 que

$$\lim_{\epsilon \rightarrow 1/2} \frac{[B_{\chi_H}^+(\epsilon)]^2}{B_{\chi_L}^+(\epsilon)} = 1.35,$$

y

$$\lim_{\epsilon \rightarrow 1/2} \frac{[B_{\chi_T}^+(\epsilon)]^2}{B_{\chi_L}^+(\epsilon)} = 1.83.$$

En la tabla 2.2 comparamos siete M-estimadores de escala con punto de ruptura  $1/2$ . Las funciones  $\chi_L$ ,  $\chi_T$ ,  $\chi_H$ ,  $\chi_W$  y  $\chi_C$  fueron definidas anteriormente. Las nuevas funciones  $\chi_4(y) = \min(y^4/c^4, 1)$ ,  $c = 0.85$ , y

$$\chi_{CR}(y) = \begin{cases} 1, & \text{si } |y| \geq k \\ by^2 + c, & \text{si } \delta \leq |y| < k \\ a|y|, & \text{si } 0 \leq |y| < \delta \end{cases},$$

se incluyen para ilustrar la relación entre eficiencia y tasa de ruptura. La función  $\chi_{CR}$  fue inicialmente introducida por Croux (1994) para probar que no existe una relación de compromiso entre alto punto de ruptura y eficiencia en la clase de M-estimadores de escala. Eligiendo las constantes  $a$ ,  $b$ ,  $c$ ,  $k$  y  $\delta$  apropiadamente (véanse los detalles en Croux, 1994) este estimador alcanza un punto de ruptura  $1/2$  y una eficiencia arbitrariamente alta. Croux también señala que el compromiso que realmente existe es entre la curva de sesgo máximo y la eficiencia. Dado que la función de Croux tiene orden local  $k = 1$  y  $\chi_H$  (la función tomada como base en la tabla 2.2) tiene orden local  $k = 2$ , aplicamos el corolario 1(b) para concluir que  $TRR^+(\chi_{CR}, \chi_H) = \infty$ . Por lo tanto, el compromiso existente entre la eficiencia y la curva de sesgo es reflejado por el valor de la tasa de ruptura. En términos de tasa de ruptura, la función  $\chi_{CR}$  es sólo

Función	$k$	$TRR^+(\chi, \chi_H)$	$TRR^-(\chi, \chi_H)$	$\gamma^*$	$EF$
$\chi_{CR}$	1	$\infty$	0.73	2.69	76.7
$\chi_L$	1	$\infty$	0.94	1.39	61.5
$\chi_C$	2	1.70	0.72	1.59	52.2
$\chi_W$	2	1.27	0.96	1.33	54.9
$\chi_T$	2	1.16	0.99	1.28	53.9
$\chi_H$	2	1	1	1.23	50.6
$\chi_4$	4	0	1.03	1.19	43.9

**Tabla 2.2:** Tasas de ruptura respecto a *outliers* e *inliers* de varios M-estimadores de escala tomando el M-estimador de Huber como base. Se muestran también las sensibilidades y las eficiencias asintóticas.

comparable a la función  $\chi_L$  entre todas las funciones consideradas en la tabla 2.2 (ambas tienen orden local  $k = 1$ ). Utilizando el corolario 1 se prueba que  $TRR^+(\chi_{CR}, \chi_L) = 6.18$ . Para las funciones de Croux correspondientes a las eficiencias 92% y 94%, las tasas de ruptura relativas son  $TRR^+(\chi_{CR}, \chi_L) = 60.2$  y  $TRR^+(\chi_{CR}, \chi_L) = 602.7$ , respectivamente. Para completar la presente discusión sobre las relaciones entre tasa de ruptura y eficiencia obsérvese que, como  $\chi_4$  tiene orden local  $k = 4$ , su tasa de ruptura respecto a *outliers* refleja un comportamiento del sesgo muy bueno pero, por otra parte, la eficiencia de este estimador es muy baja. Puede conjeturarse que una tasa de ruptura pequeña respecto a *outliers* y una alta eficiencia no pueden alcanzarse simultáneamente en la clase de M-estimadores de escala.

Hasta este punto hemos considerado el caso en que el orden local de  $\chi$  sea finito. Puede darse el caso de que  $k = \infty$ , es decir, que  $\chi^{(j)}(0) = 0$  para todo  $j$ . Esto ocurre

para funciones que son extremadamente “planas” o para funciones que son constantes e iguales a cero en un entorno de cero. Un ejemplo del primer tipo viene dado por la función

$$\chi(y) = \begin{cases} \exp(-y^{-2}), & \text{si } y \neq 0 \\ 0, & \text{si } y = 0 \end{cases},$$

mientras que un ejemplo de la segunda situación viene dado por las funciones de salto,  $\chi_a$ , que definen los estimadores de sesgo minimax.

La siguiente proposición caracteriza la clase de funciones  $\chi$  para las que la tasa de ruptura (respecto a las funciones de salto que definen los estimadores con sesgo minimax) es finita. Para que esto ocurra es necesario y suficiente que la función sea constante e igual a cero en un entorno de cero.

**Proposición 9** *Supongamos que  $F_0$  verifica la hipótesis 3 y que además*

$$\lim_{y \rightarrow \infty} \frac{f_0(y)}{1 - F_0(y)} > 0. \quad (2.10)$$

*Supongamos que  $\chi \in C_b$ . Entonces,  $TR(\chi) < \infty$  si y sólo si*

$$m = \sup\{y : \chi(y) = 0\} > 0.$$

**Demostración:**

En primer lugar, vamos a suponer  $m = \sup\{y : \chi(y) = 0\} > 0$  y vamos a probar que la tasa de ruptura es finita. En este caso, nótese que

$$\begin{aligned} h_\chi(s) &= \int_0^\infty \chi(y/s) dF_0(y) = \int_{ms}^\infty \chi(y/s) f_0(y) dy \\ &\leq \int_{ms}^\infty f_0(y) dy = 1 - F_0(ms). \end{aligned}$$

Sustituyendo  $s$  por  $B_x^+(\epsilon)$  y operando, tenemos que

$$B_x^+(\epsilon) \leq \frac{1}{m} F_0^{-1} \left( 1 - \frac{b - \epsilon}{1 - \epsilon} \right) = \frac{1}{m} F_0^{-1} \left( \frac{1 - b}{1 - \epsilon} \right).$$

Por otra parte,

$$B_{x_a}^+(\epsilon) = \frac{1}{a} F_0^{-1} \left( 1 - \frac{b - \epsilon}{1 - \epsilon} \right) = \frac{1}{a} F_0^{-1} \left( \frac{1 - b}{1 - \epsilon} \right), \quad (2.11)$$

donde  $a = F_0^{-1}(1 - b)$ . Como consecuencia

$$TR^+(\chi) = \lim_{\epsilon \rightarrow b} \frac{B_x^+(\epsilon)}{B_{x_a}^+(\epsilon)} \leq \frac{a}{m} < \infty, \quad (2.12)$$

dado que  $m > 0$ .

Para probar el recíproco, supongamos por reducción al absurdo que  $\chi(y) > 0$  para todo  $y > 0$ . Veamos que en esta situación debe cumplirse  $TR^+(\chi) = \infty$ .

Sea  $d > 0$ , definamos

$$\chi_d(y) = \begin{cases} 0, & \text{si } |y| < d \\ 1, & \text{si } |y| \geq d \end{cases}$$

Nótese que

$$\begin{aligned} h_\chi(s) - h_{\chi_d}(s) &= s \int_0^\infty \chi(y) f_0(ys) dy - s \int_d^\infty f_0(ys) dy \\ &= s \int_0^d \chi(y) f_0(ys) dy - s \int_d^\infty [1 - \chi(y)] f_0(ys) dy \\ &\geq s f_0(ds) \int_0^d \chi(y) dy - [1 - \chi(d)] [1 - F_0(ds)] \\ &= [1 - F_0(ds)] \left[ \frac{s f_0(ds)}{1 - F_0(ds)} \int_0^d \chi(y) dy - [1 - \chi(d)] \right], \end{aligned}$$

donde hemos utilizado que  $f_0$  es decreciente y la monotonía de la función  $\chi$ . Dado que se cumple (2.10),

$$\lim_{s \rightarrow \infty} \frac{s f_0(sd)}{1 - F_0(sd)} = \infty,$$

por lo que se verifica

$$h_{\chi}(s) - h_{\chi_d}(s) > 0,$$

para  $s$  suficientemente grande.

Tomando ahora  $s = B_{\chi}^+(\epsilon)$ , aplicando la proposición 1 y teniendo en cuenta que  $h_{\chi_d}(s) = 1 - F_0(ds)$ . Para todo  $d > 0$ , si  $\epsilon$  es suficientemente próximo a  $b$  se verifica

$$\frac{b - \epsilon}{1 - \epsilon} - 1 + F_0[dB_{\chi}^+(\epsilon)] > 0.$$

Operando se tiene que

$$B_{\chi}^+(\epsilon) > \frac{1}{d} F_0^{-1} \left( \frac{1 - b}{1 - \epsilon} \right),$$

si  $\epsilon$  es suficientemente próximo a  $b$ . Por (2.11) y esta última desigualdad,  $TR^+(\chi) \geq a/d$ . Pero esto es válido para  $d > 0$  arbitrario. Haciendo  $d \rightarrow 0$ , se obtiene que  $TR^+(\chi) = \infty$ .

□

La hipótesis 2.10 en el enunciado de la proposición anterior es equivalente a pedir que la tasa de fallo correspondiente a  $F_0$  no se anule cuando  $y \rightarrow \infty$ , por lo tanto, no es una condición restrictiva. Por ejemplo, tanto la distribución normal como la exponencial la verifican.

La demostración de la proposición 9 permite además dar una fórmula para el valor de la tasa de ruptura en aquellos casos en los que es finita.

**Corolario 3** *Bajo las hipótesis de la proposición 9 y si  $m = \sup\{y : \chi(y) = 0\} > 0$ , se tiene que*

$$TR^+(\chi) = \frac{a}{m}.$$

**Demostración:**

El resultado se obtiene si se tiene en cuenta (2.12) y, para la desigualdad recíproca, se reproduce la segunda parte de la demostración de la proposición 9, lo que se puede hacer para cualquier  $d > m$ . Se deduce  $TR^+(\chi) \geq a/d$  para todo  $d > m$ . Haciendo  $d \rightarrow m$ , se concluye que  $TR^+(\chi) \geq a/m$ .

□

**2.2.2 Tasa de ruptura respecto a inliers**

La siguiente proposición proporciona una fórmula para calcular la tasa de ruptura respecto a *inliers* de los M-estimadores de escala.

**Proposición 10** *Supongamos que  $F_0$  cumple la hipótesis 3 y que la función  $\chi \in C_b$  es tal que  $\int_0^\infty [1 - \chi(y)]dy < \infty$ . Entonces*

$$TR^-(\chi) = a \left[ \int_0^\infty [1 - \chi(y)]dy \right]^{-1}, \quad \text{donde } a = F_0^{-1}(1 - b).$$

**Demostración:**

Sea  $D = \int_0^\infty [1 - \chi(y)]dy < \infty$ . Obsérvese que

$$\begin{aligned} \int_0^D \chi(y)dy - \int_D^\infty [1 - \chi(y)]dy &= \int_0^D \chi(y)dy - \left[ \int_0^\infty [1 - \chi(y)]dy - \int_0^D [1 - \chi(y)]dy \right] \\ &= \int_0^D \chi(y)dy - D + D - \int_0^D \chi(y)dy = 0. \end{aligned}$$

Es decir,

$$\int_0^D \chi(y)dy - \int_D^\infty [1 - \chi(y)]dy = 0. \quad (2.13)$$

Definamos la función de salto

$$\chi_D(y) = \begin{cases} 0, & \text{si } |y| < D \\ 1, & \text{si } |y| \geq D \end{cases}.$$

Nótese que

$$\begin{aligned} h_\chi(s) - h_{\chi_D}(s) &= s \left[ \int_0^D \chi(y) f_0(sy) dy - \int_D^\infty [1 - \chi(y)] f_0(sy) dy \right] \\ &\geq s f_0(sD) \left[ \int_0^D \chi(y) dy - \int_D^\infty [1 - \chi(y)] dy \right] = 0, \end{aligned}$$

dado que  $f_0$  es decreciente y por la igualdad (2.13). Tomando  $s = B_\chi^-(\epsilon)$ , se sigue que

$$B_\chi^-(\epsilon) > \frac{1}{D} F_0^{-1} \left[ 1 - \frac{b}{1 - \epsilon} \right]$$

y, como consecuencia,  $TR^-(\chi) \geq a/D$ .

Para probar la desigualdad recíproca, obsérvese que, debido a (2.13), para todo  $d < D$

$$\int_0^d \chi(y) dy < \int_0^D \chi(y) dy = \int_D^\infty [1 - \chi(y)] dy < \int_d^\infty [1 - \chi(y)] dy. \quad (2.14)$$

Definamos, para  $d < D$ , la función de salto

$$\chi_d(y) = \begin{cases} 0, & \text{si } |y| < d \\ 1, & \text{si } |y| \geq d \end{cases}.$$

Se tiene que

$$h_\chi(s) - h_{\chi_d}(s) = s \left\{ \int_0^d \chi(y) f_0(sy) dy - \int_d^\infty [1 - \chi(y)] f_0(sy) dy \right\}.$$

Por el teorema de la convergencia dominada, el término entre llaves converge, cuando  $s \rightarrow 0$ , a

$$f_0(0) \left\{ \int_0^d \chi(y) dy - \int_d^\infty [1 - \chi(y)] dy \right\}$$

que es estrictamente negativo por (2.14). Se sigue que, dado  $d < D$ , para  $s$  suficientemente próximo a cero  $h_\chi(s) - h_{\chi_a}(s) < 0$ . Por lo tanto, tomando  $s = B_\chi^-(\epsilon)$ , y dado  $d < D$ , para  $\epsilon$  suficientemente próximo a  $1 - b$ ,

$$B_\chi^-(\epsilon) \leq \frac{1}{d} F_0^{-1} \left[ 1 - \frac{b}{1 - \epsilon} \right].$$

Se tiene que  $TR^-(\chi) \leq a/d$  para todo  $d < D$ . Haciendo  $d \rightarrow D$ , se deduce  $TR^-(\chi) \leq a/D$ .

□

La condición  $\int_0^\infty [1 - \chi(y)] dy < \infty$  se verifica para todos los ejemplos que hemos estudiado en el apartado anterior. De hecho si existe  $A$  tal que  $\chi(y) = 1$  para todo  $y \geq A$ , la condición se satisface automáticamente. Aplicando la proposición 10 se obtienen los valores de las tasas de ruptura respecto a *inliers* de todos los estimadores que hemos considerado hasta ahora. Los resultados se presentan en la tabla 2.2 de la página 51.

### 2.3 Tasa de ruptura de $\tau$ -estimadores de escala

La clase de  $\tau$ -estimadores de escala combina buenas propiedades de eficiencia y punto de ruptura. Como hemos visto en el capítulo uno, estos estimadores fueron definidos en el contexto de regresión robusta. Sean las funciones  $\chi_i \in C_{b_i}$ ,  $i = 1, 2$ . La  $\tau$ -escala de  $F$  se define como

$$\tau(F) = S(F) \left[ \frac{1}{b_2} E_F \chi_2 \left( \frac{y}{S(F)} \right) \right]^{1/2},$$

donde  $S(F)$  es el M-estimador de escala basado en  $\chi_1$ . Para calcular el  $\tau$ -estimador de escala, basta reemplazar  $F$  por la función de distribución empírica  $F_n$  en la definición anterior.

El punto de ruptura de un  $\tau$ -estimador es  $\epsilon^* = \{b_1, 1 - b_1\}$ . Esto es, un  $\tau$ -estimador hereda el punto de ruptura del M-estimador de escala basado en  $\chi_1$ . Rousseeuw y Croux (1994) encontraron las curvas de sesgo máximo de los  $\tau$ -estimadores con respecto a *outliers* e *inliers* cuando la función  $s \rightarrow s^2 E_{F_0} \chi_2(y/s)$  es creciente para  $s \geq 0$ . En este caso, una aplicación directa de las fórmulas (3.9) y (3.10) del citado trabajo permite deducir

$$TRR^+(\tau, S) = TRR^-(\tau, S) = (b_1/b_2)^{1/2}.$$

Por lo tanto, los  $\tau$ -estimadores no heredan la tasa de ruptura de los M-estimadores iniciales de escala en que se basan. De hecho, dado que  $b_2$  se selecciona para obtener eficiencia mientras que  $b_1$  se selecciona para obtener alto punto de ruptura, en general se tiene que  $b_1/b_2 > 1$ . Nótese que en la misma medida que empeora la tasa de ruptura respecto a *outliers*, mejora la tasa de ruptura respecto a *inliers*.

Por ejemplo, si  $\chi_1$  y  $\chi_2$  son dadas por  $\chi_T$  (véase tabla 2.1), donde las constantes correspondientes se ajustan para obtener  $b_1 = 1/2$  (alto punto de ruptura) y  $b_2 = 0.16$  (para que la eficiencia del  $\tau$ -estimador sea del 95% bajo el modelo normal), entonces  $TRR^+(\tau, S) = TRR^-(\tau, S) = 1.76$ . Al pasar del M-estimador de escala basado en  $\chi_1$  con el 53.9% de eficiencia (véase tabla 2.2) al  $\tau$ -estimador con el 95%, se paga un precio en robustez que no recoge el punto de ruptura pero sí la tasa de ruptura respecto a *outliers*.

En el capítulo tres estudiaremos aspectos relacionados con la curva de sesgo asintótico máximo de  $\tau$ -estimadores de regresión.

## 2.4 Tasa de ruptura de estimadores de dispersión

Supongamos ahora que la distribución central del entorno es el modelo clásico de localización y dispersión:

$$F^{\mu,\sigma}(y) = F_0\left(\frac{y - \mu}{\sigma}\right),$$

donde  $F_0$  es una función de distribución que verifica la siguiente hipótesis:

**Hipótesis 4** (a)  $F_0$  es estrictamente creciente y derivable con función de densidad  $f_0$  simétrica y estrictamente unimodal.

(b)  $\lim_{y \rightarrow \infty} f_0(y)/[1 - F_0(y)] > 0$ .

En esta sección se obtienen las tasas de ruptura respecto a *outliers* e *inliers* de los estimadores de dispersión que fueron introducidos en la sección 1.3.

En el apartado 2.4.1 se estudian las tasas de ruptura relativas entre la MEDA y el SHORTH, definidos respectivamente en las ecuaciones (1.9) y (1.10). En el apartado 2.4.2 se calculan las tasas de ruptura de las alternativas a la MEDA que no requieren un estimador de localización previo.

### 2.4.1 Tasa de ruptura de la MEDA

En esta sección el objetivo es cuantificar el aumento de sesgo que se produce cuando elegimos previamente el estimador de localización y aplicamos un M-estimador a los datos centrados, frente al caso de los S-estimadores, en que se elige la localización que minimiza la M-escala de los datos centrados. Por la primera parte de la proposición 3, este aumento de sesgo es debido a la necesidad de aplicar un estimador de localización previo, antes de aplicar el M-estimador de escala. Lo que haremos será elegir un

estimador importante de cada tipo, la MEDA y el SHORTH, y calcular sus tasas de ruptura relativas.

Por la segunda parte de la proposición 3 la curva de sesgo de la MEDA respecto a *inliers* es la misma que la del SHORTH. Como consecuencia,

$$TRR^-(\text{MEDA}, \text{SHORTH}) = 1.$$

La diferencia aparece en el comportamiento de la curva de sesgo respecto a *outliers*. La tasa de ruptura correspondiente se da en la proposición 11. Previamente, se dan dos lemas que comparan los sesgos respecto a *outliers* de la MEDA,  $B_M^+(\epsilon)$ , y del SHORTH,  $B_{SH}^+(\epsilon)$ , con la curva de sesgo asintótico máximo de la mediana,  $B_0(\epsilon)$ .

**Lema 1** Sea  $c = 1/F_0^{-1}(0.75)$ . Supongamos que  $F_0$  satisface la hipótesis 4. Se verifica

$$\lim_{\epsilon \rightarrow 1/2} \frac{B_{SH}^+(\epsilon)}{B_0(\epsilon)} = c.$$

**Demostración:**

Sea  $\chi_a$  la función de salto utilizada para definir el SHORTH, es decir,

$$\chi_a(y) = \begin{cases} 0, & \text{si } |y| < a \\ 1, & \text{si } |y| \geq a \end{cases}, \quad a = F_0^{-1}(0.75). \quad (2.15)$$

Sea  $g_a(y) = E_{F_0} \chi_a(y/s) = 2[1 - F_0(as)]$ . Por las proposiciones 1 y 3, se tiene que

$$B_{SH}^+(\epsilon) = g_a^{-1} \left( \frac{b - \epsilon}{1 - \epsilon} \right) = g_a^{-1} \left( \frac{1 - 2\epsilon}{2(1 - \epsilon)} \right),$$

puesto que, en este caso,  $b = 1/2$ . Por lo tanto,

$$B_{SH}^+(\epsilon) = cF_0^{-1} \left( \frac{3 - 2\epsilon}{4(1 - \epsilon)} \right).$$

El límite que queremos calcular es

$$\lim_{\epsilon \rightarrow 1/2} \frac{B_{SH}^+(\epsilon)}{B_0(\epsilon)} = c \lim_{\epsilon \rightarrow 1/2} \frac{F_0^{-1}[(3-2\epsilon)/(4(1-\epsilon))]}{F_0^{-1}[1/(2(1-\epsilon))]},$$

ya que el sesgo de la mediana,  $B_0(\epsilon)$ , es  $F_0^{-1}[1/(2(1-\epsilon))]$  por la fórmula (1.15).

Es suficiente probar que el límite de la anterior expresión es igual a uno. Para ello definamos  $G_1(\epsilon) = F_0^{-1}[(3-2\epsilon)/(4(1-\epsilon))]$  y  $G_2(\epsilon) = F_0^{-1}[1/(2(1-\epsilon))]$ . Por el teorema del valor medio, existe  $\xi > 0$  con

$$\frac{1}{2(1-\epsilon)} < \xi < \frac{3-2\epsilon}{4(1-\epsilon)},$$

tal que

$$G_1(\epsilon) - G_2(\epsilon) = \frac{1}{f_0[F_0^{-1}(\xi)]} \left( \frac{3-2\epsilon}{4(1-\epsilon)} - \frac{1}{2(1-\epsilon)} \right) = \frac{1}{f_0[F_0^{-1}(\xi)]} \frac{1-2\epsilon}{4(1-\epsilon)}.$$

Dado que  $f_0$  es unimodal y simétrica, sustituyendo  $\xi$  por su cota superior,

$$0 < \frac{G_1(\epsilon) - G_2(\epsilon)}{G_2(\epsilon)} < \frac{1-2\epsilon}{4(1-\epsilon)G_2(\epsilon)f_0[G_1(\epsilon)]} = \frac{1-F_0[G_1(\epsilon)]}{2G_2(\epsilon)f_0[G_1(\epsilon)]}, \quad (2.16)$$

ya que

$$1 - F_0[G_1(\epsilon)] = 1 - \frac{3-2\epsilon}{4(1-\epsilon)} = \frac{1-2\epsilon}{4(1-\epsilon)}.$$

El último cociente de la expresión (2.16) converge a 0 cuando  $\epsilon \rightarrow 1/2$  por la hipótesis 4(b). Por lo tanto,

$$\lim_{\epsilon \rightarrow 1/2} \frac{G_1(\epsilon) - G_2(\epsilon)}{G_2(\epsilon)} = 0,$$

de donde se deduce que

$$\lim_{\epsilon \rightarrow 1/2} \frac{G_1(\epsilon)}{G_2(\epsilon)} = 1,$$

como queríamos probar. □

**Lema 2** Sea  $c = 1/F_0^{-1}(0.75)$ . Supongamos que  $F_0$  satisface la hipótesis 4. Se verifica

$$\lim_{\epsilon \rightarrow 1/2} \frac{B_M^+(\epsilon)}{B_0(\epsilon)} = 2c.$$

**Demostración:**

Por la proposición 4, el sesgo de la MEDA respecto a *outliers* verifica

$$E_{F_0} \chi_a \left( \frac{y - B_0(\epsilon)}{B_M^+(\epsilon)} \right) = \frac{1 - 2\epsilon}{2(1 - \epsilon)}, \quad (2.17)$$

donde  $\chi_a$  es la función de salto definida por la ecuación (2.15) con  $a = 1/c$ . Es inmediato comprobar que

$$E_{F_0} \chi_a \left( \frac{y - B_0(\epsilon)}{B_M^+(\epsilon)} \right) = F_0[B_0(\epsilon) - aB_M^+(\epsilon)] - F_0[B_0(\epsilon) + aB_M^+(\epsilon)] + 1.$$

Reordenando términos, la condición (2.17) es equivalente a

$$F_0[B_0(\epsilon) + aB_M^+(\epsilon)] - F_0[B_0(\epsilon) - aB_M^+(\epsilon)] = \frac{1}{2(1 - \epsilon)}. \quad (2.18)$$

Definamos ahora  $H_1(\epsilon) = 2cB_0(\epsilon)$  y  $H_2(\epsilon) = cB_0(\epsilon) + B_{SH}^+(\epsilon)$ . Vamos a probar

$$H_1(\epsilon) < B_M^+(\epsilon) < H_2(\epsilon). \quad (2.19)$$

Para demostrar la primera desigualdad sustituimos  $H_1(\epsilon)$  por  $B_M^+(\epsilon)$  en (2.18) y tenemos en cuenta que la distribución  $F_0$  es simétrica. Se sigue que

$$\begin{aligned} F_0[3B_0(\epsilon)] - F_0[-B_0(\epsilon)] &= F_0[3B_0(\epsilon)] - 1 + F_0[B_0(\epsilon)] \\ &= F_0[3B_0(\epsilon)] - 1 + \frac{1}{2(1 - \epsilon)} < \frac{1}{2(1 - \epsilon)}. \end{aligned}$$

Dado que la función  $l(x) = F_0[B_0(\epsilon) + ax] - F_0[B_0(\epsilon) - ax]$  es estrictamente creciente (ya que  $F_0$  lo es por hipótesis), comparando la desigualdad anterior con la igualdad (2.18),

$$H_1(\epsilon) < B_M^+(\epsilon).$$

Para probar la desigualdad restante, sustituimos  $H_2(\epsilon)$  por  $B_M^+(\epsilon)$  en la ecuación (2.18). Se sigue que

$$F_0 \left[ 2B_0(\epsilon) + F_0^{-1} \left( \frac{3-2\epsilon}{4(1-\epsilon)} \right) \right] - \left[ 1 - \frac{3-2\epsilon}{4(1-\epsilon)} \right] > 2 \frac{3-2\epsilon}{4(1-\epsilon)} - 1 = \frac{1}{2(1-\epsilon)}.$$

Por lo tanto,  $B_M^+(\epsilon) < H_2(\epsilon)$ .

A partir de (2.19) y aplicando el lema 1 se obtiene

$$2c = \lim_{\epsilon \rightarrow 1/2} \frac{H_1(\epsilon)}{B_0(\epsilon)} \leq \lim_{\epsilon \rightarrow 1/2} \frac{B_M^+(\epsilon)}{B_0(\epsilon)} \leq \lim_{\epsilon \rightarrow 1/2} \frac{H_2(\epsilon)}{B_0(\epsilon)} = 2c,$$

por lo que

$$\lim_{\epsilon \rightarrow 1/2} \frac{B_M^+(\epsilon)}{B_0(\epsilon)} = 2c$$

como queríamos probar. □

A partir de los dos lemas anteriores, la siguiente proposición es inmediata:

**Proposición 11** *Bajo las hipótesis de los lemas 1 y 2 se verifica:*

$$TRR^+(\text{MEDA}, \text{SHORTH}) \equiv TR^+(\text{MEDA}) = 2.$$

Por lo tanto, de forma aproximada y para  $\epsilon \approx 1/2$ , la curva de sesgo de la MEDA respecto a *outliers* es el doble que la del SHORTH.

#### 2.4.2 Tasa de ruptura de algunas alternativas a la MEDA

En este apartado derivamos las tasas de ruptura de los estimadores  $S_n$  y  $Q_n$  ya definidos en las ecuaciones (1.13) y (1.14) como alternativas a la MEDA y al SHORTH. Los resultados se reúnen en la siguiente proposición.

**Proposición 12** Sea  $c = 1/F_0^{-1}(0.75)$ ,

- (a) si  $F_0$  satisface la hipótesis 4, entonces  $TR^+(S) = 2k/c$  y  $TR^-(S) = k/c$ .  
 (b) Si  $F_0$  es la función de distribución correspondiente a la distribución normal estándar,  $TR^+(Q) = 2^{1/2}d/c$  y  $TR^-(Q) = [c(1 + 2^{-3/2}d^{-1})]^{-1}$ .

**Demostración:**

- (a) Por las proposiciones 1 y 3,

$$B_{SH}^+(\epsilon) = cF_0^{-1}\left(\frac{3-2\epsilon}{4(1-\epsilon)}\right).$$

En la proposición 5 se muestra que el sesgo respecto a *outliers* de  $S_n$ , que denotaremos por  $B_S^+(\epsilon)$ , verifica

$$F_0\left[\frac{B_{SH}^+(\epsilon)}{c} + \frac{B_S^+(\epsilon)}{k}\right] - F_0\left[\frac{B_{SH}^+(\epsilon)}{c} - \frac{B_S^+(\epsilon)}{k}\right] = \frac{1}{2(1-\epsilon)}. \quad (2.20)$$

Basta sustituir  $x$  por  $B_S^+(\epsilon)/c$  en la definición implícita de la función  $g^+$ .

Sea  $B_0(\epsilon)$  la curva de sesgo de la mediana. Definamos  $H_1(\epsilon) = 2kB_0(\epsilon)$  y  $H_2(\epsilon) = (2k/c)B_{SH}^+(\epsilon)$ . Si sustituimos  $B_S^+(\epsilon)$  por  $H_1(\epsilon)$  en la ecuación (2.20) se obtiene

$$\begin{aligned} & F_0[B_S^+(\epsilon)/c + 2B_0(\epsilon)] - F_0[B_S^+(\epsilon)/c - 2B_0(\epsilon)] \\ &= F_0[B_S^+(\epsilon)/c + 2B_0(\epsilon)] - 1 + F_0[2B_0(\epsilon) - B_S^+(\epsilon)/c] \\ &< F_0[2B_0(\epsilon) - B_S^+(\epsilon)/c] < F_0[B_0(\epsilon)] = \frac{1}{2(1-\epsilon)}. \end{aligned}$$

Dado que la función  $l(x) = F_0[B_0(\epsilon) + ax] - F_0[B_0(\epsilon) - ax]$  es estrictamente creciente (ya que  $F_0$  lo es por hipótesis) se deduce

$$H_1(\epsilon) < B_S^+(\epsilon).$$

Si sustituimos ahora  $B_S^+(\epsilon)$  por  $H_2(\epsilon)$  en la ecuación (2.20) se obtiene

$$F_0[3B_{SH}^+(\epsilon)/c] - F_0[-B_{SH}^+(\epsilon)/c] > F_0[B_{SH}^+(\epsilon)/c] - \left(1 - \frac{3-2\epsilon}{4(1-\epsilon)}\right) = \frac{1}{2(1-\epsilon)}.$$

Por lo tanto,

$$B_S^+(\epsilon) < H_2(\epsilon).$$

Hemos probado  $H_1(\epsilon) < B_S^+(\epsilon) < H_2(\epsilon)$ .

Dividiendo en estas desigualdades por  $B_{SH}^+(\epsilon)$ , tomando límites cuando  $\epsilon \rightarrow 1/2$  y aplicando el lema 1 se deduce que  $TR^+(S) = 2k/c$ .

Vamos a obtener a continuación  $TR^-(S)$ . Sabemos, por las proposiciones 1 y 3, que

$$B_{SH}^-(\epsilon) = cF_0^{-1}\left(\frac{3-4\epsilon}{4(1-\epsilon)}\right).$$

Por la proposición 5, el sesgo respecto a *inliers* de  $S_n$ , que denotamos por  $B_S^-(\epsilon)$ , verifica

$$F_0\left[\frac{B_{SH}^-(\epsilon)}{c} + \frac{B_S^-(\epsilon)}{k}\right] - F_0\left[\frac{B_{SH}^-(\epsilon)}{c} - \frac{B_S^-(\epsilon)}{k}\right] = \frac{1-2\epsilon}{2(1-\epsilon)}.$$

Aplicamos el teorema del valor medio al miembro de la izquierda. Se sigue que para todo  $\epsilon \in (0, 1/2)$  existe

$$\alpha = \alpha(\epsilon) \in \left(\frac{B_{SH}^-(\epsilon)}{c} - \frac{B_S^-(\epsilon)}{k}, \frac{B_{SH}^-(\epsilon)}{c} + \frac{B_S^-(\epsilon)}{k}\right),$$

tal que

$$2f_0[\alpha(\epsilon)]\frac{B_S^-(\epsilon)}{k} = \frac{1-2\epsilon}{2(1-\epsilon)},$$

donde  $f_0$  es la función de densidad correspondiente a  $F_0$ . Despejando  $B_S^-(\epsilon)$ ,

$$B_S^-(\epsilon) = \frac{k(1-2\epsilon)}{4(1-\epsilon)f_0[\alpha(\epsilon)]}.$$

Nótese que  $\lim_{\epsilon \rightarrow 1/2} \alpha(\epsilon) = 0$ . Aplicando la regla de L'Hôpital,

$$\lim_{\epsilon \rightarrow 1/2} \frac{B_S^-(\epsilon)}{B_{SH}^-(\epsilon)} = \frac{k}{2cf_0(c)} \lim_{\epsilon \rightarrow 1/2} \frac{1-2\epsilon}{F_0^{-1}\left(\frac{3-4\epsilon}{4(1-\epsilon)}\right)} = \frac{k}{c}.$$

(b) Para calcular la tasa de ruptura respecto a *outliers* del estimador  $Q_n$  aplicamos la proposición 6. El sesgo respecto a *outliers* de  $Q_n$ , que denotamos por  $B_Q^+(\epsilon)$ , verifica

$$B_Q^+(\epsilon) = d(F_0^*)^{-1} \left[ \frac{5 - 8\epsilon + 4\epsilon^2}{8(1 - \epsilon)^2} \right],$$

donde  $F_0^*$  denota la convolución  $F_0 * F_0$ . Si  $F_0$  es la función de distribución de la normal estándar,  $\Phi$ , entonces  $\Phi^*(y) = \Phi(2^{-1/2}y)$  y  $(\Phi^*)^{-1}(t) = 2^{1/2}\Phi^{-1}(t)$ . Por lo tanto, en este caso,

$$B_Q^+(\epsilon) = d2^{1/2}\Phi^{-1} \left( \frac{5 - 8\epsilon + 4\epsilon^2}{8(1 - \epsilon)^2} \right).$$

Como consecuencia,

$$TR^+(Q) = \lim_{\epsilon \rightarrow 1/2} \frac{B_Q^+(\epsilon)}{B_{SH}^+(\epsilon)} = \frac{2^{1/2}}{c} \lim_{\epsilon \rightarrow 1/2} \frac{G_1(\epsilon)}{G_2(\epsilon)},$$

donde

$$G_1(\epsilon) = \Phi^{-1} \left( \frac{5 - 8\epsilon + 4\epsilon^2}{8(1 - \epsilon)^2} \right), \quad \text{y} \quad G_2(\epsilon) = \Phi^{-1} \left( \frac{3 - 2\epsilon}{4(1 - \epsilon)} \right).$$

Análogamente a como se razonó en la demostración del lema 2, se prueba que el límite vale uno, de donde se obtiene el resultado.

Para completar la demostración, sólo queda calcular  $TR^-(Q)$  en el caso  $F_0 \equiv \Phi$ . Para ello aplicamos de nuevo la proposición 6 según la cual el sesgo respecto a *inliers* de  $Q_n$ , que denotamos por  $B_Q^-(\epsilon)$ , verifica

$$(1 - \epsilon)^2 \Phi[B_Q^-(\epsilon)/(2^{1/2}d)] + 2\epsilon(1 - \epsilon)\Phi[B_Q^-(\epsilon)] + \epsilon^2 = 5/8. \quad (2.21)$$

Utilizando el teorema del valor medio, se tiene que para todo  $\epsilon \in (0, 1/2)$  existen  $\alpha = \alpha(\epsilon) \in (0, B_Q^-(\epsilon)/(2^{1/2}d))$  y  $\beta = \beta(\epsilon) \in (0, B_Q^-(\epsilon))$  tales que

$$\Phi[B_Q^-(\epsilon)/(2^{1/2}d)] - \Phi(0) = \varphi(\alpha)B_Q^-(\epsilon)/(2^{1/2}d),$$

y

$$\Phi[B_Q^-(\epsilon)] - \Phi(0) = \varphi(\beta)B_Q^-(\epsilon),$$

donde  $\varphi$  es la función de densidad correspondiente a  $\Phi$ . Sustituyendo estas expresiones en la ecuación (2.21) y dado que  $\Phi(0) = 1/2$  obtenemos

$$(1 - \epsilon)^2 \left[ 1/2 + \varphi(\alpha) \frac{B_Q^-(\epsilon)}{2^{1/2}d} \right] + 2\epsilon(1 - \epsilon)[1/2 + \varphi(\beta)B_Q^-(\epsilon)] + \epsilon^2 = 5/8.$$

Despejando  $B_Q^-(\epsilon)$  llegamos a la expresión

$$B_Q^-(\epsilon) = \frac{2^{1/2}d(1 - 4\epsilon^2)}{8(1 - \epsilon)^2\varphi(\alpha) + 2^{1/2}16d\epsilon(1 - \epsilon)\varphi(\beta)}.$$

Nótese que  $\lim_{\epsilon \rightarrow 1/2} \alpha(\epsilon) = \lim_{\epsilon \rightarrow 0.5} \beta(\epsilon) = 0$ . Una aplicación de la regla de L'Hôpital implica

$$\lim_{\epsilon \rightarrow 1/2} \frac{B_Q^-(\epsilon)}{B_{SH}^-(\epsilon)} = \frac{2^{1/2}d}{2c\varphi(0)[1 + 2^{3/2}d]} \lim_{\epsilon \rightarrow 1/2} \frac{1 - 4\epsilon^2}{F_0^{-1}\left(\frac{3-4\epsilon}{4(1-\epsilon)}\right)} = [c(1 + 2^{-3/2}d^{-1})]^{-1}.$$

□

La tabla 2.3 muestra las tasas de ruptura resultantes cuando las constantes se ajustan de forma que los estimadores sean consistentes bajo el modelo normal. Aunque el estimador  $Q_n$  sea el más eficiente, el comportamiento de su curva de sesgo es el peor de todos tanto para *outliers* como para *inliers* y tanto para proporciones de contaminación pequeñas como para proporciones de contaminación grandes. Por otra parte,  $S_n$  parece un compromiso razonable entre eficiencia y buen comportamiento del sesgo. Nótese que los estimadores MEDA y SHORTH sólo difieren en su tasa de ruptura respecto a *outliers*.

## 2.5 Tasa de ruptura de M-estimadores de localización

Supongamos que el modelo central del entorno es el mismo que el de la sección anterior, es decir, un modelo paramétrico de localización y dispersión. Nuestro interés se centra



Estimador	$TR^+$	$TR^-$	$\gamma^*$	$EF$
SHORTH	1	1	1.16	36.74
MEDA	2	1	1.16	36.74
$S$	1.60	0.80	1.62	58.23
$Q$	2.12	0.58	2.06	82.27

**Tabla 2.3:** Tasas de ruptura respecto a *outliers* e *inliers*, sensitividades y eficiencias de cuatro estimadores de dispersión robustos.

ahora en el estudio de tasas de ruptura para M-estimadores de localización.

En vista de que la mediana tiene sesgo minimax en la clase de M-estimadores de localización, resulta natural tomar el sesgo de la mediana como base y definir la tasa de ruptura de cualquier M-estimador de localización respecto a la mediana.

En el siguiente lema se obtiene la tasa de ruptura de una clase especial de M-estimadores de localización. El lema se utilizará en la demostración del resultado más importante de esta sección.

**Lema 3** Sea  $c = 1/F_0^{-1}(0.75)$  y sea  $TR^+(D)$  la tasa de ruptura respecto a *outliers* del estimador de dispersión  $D_n$ . Supongamos que se verifica la hipótesis 4 y definamos

$$\psi_A(y) = \begin{cases} 0, & \text{si } |y| < A \\ \text{sgn}(y), & \text{si } |y| \geq A \end{cases} \quad (2.22)$$

Entonces, la tasa de ruptura del M-estimador de localización dado por  $\psi_A$ , cuando la dispersión se estima previamente por  $D_n$ , es

$$TR(\psi_A) = 1 + cTR^+(D)A.$$

**Demostración:**

Sea  $B_A(\epsilon)$  la curva de sesgo máximo del M-estimador dado por  $\psi_A$ . Por la proposición 7,  $B_A(\epsilon) = g_A^{-1}[\epsilon/(1 - \epsilon), D^+(\epsilon)]$ , donde

$$\begin{aligned} g_A(t, s) &= - \int_{-\infty}^{\infty} \psi_A\left(\frac{y-t}{s}\right) dF_0(y) = s \int_0^{\infty} \psi_A(y)[f_0(sy-t) - f_0(sy+t)] dy \\ &= F_0(As+t) - F_0(As-t). \end{aligned}$$

Tomando en esta ecuación  $t = B_A(\epsilon)$  y  $s = D^+(\epsilon)$  se sigue que

$$\frac{\epsilon}{1 - \epsilon} = F_0[AD^+(\epsilon) + B_A(\epsilon)] - F_0[AD^+(\epsilon) - B_A(\epsilon)].$$

Definamos  $H_1(\epsilon) = AD^+(\epsilon) + F_0^{-1}[\epsilon/(1 - \epsilon)]$  y  $H_2(\epsilon) = AD^+(\epsilon) + B_0(\epsilon)$ . Los mismos argumentos que en la demostración del lema 2 prueban que  $H_1(\epsilon) < B_A(\epsilon) < H_2(\epsilon)$ . Si dividimos ahora por  $B_0(\epsilon)$  en estas desigualdades, tomamos límites cuando  $\epsilon \rightarrow 1/2$  y aplicamos el lema 1, obtenemos

$$cTR^+(D)A + \lim_{\epsilon \rightarrow 1/2} \frac{F_0^{-1}[\epsilon/(1 - \epsilon)]}{B_0(\epsilon)} \leq TR(\psi_A) \leq cTR^+(D) + 1.$$

Por tanto, basta probar

$$\lim_{\epsilon \rightarrow 1/2} \frac{F_0^{-1}[\epsilon/(1 - \epsilon)]}{B_0(\epsilon)} = 1.$$

Para ello, definimos  $G_1(\epsilon) = F_0^{-1}[1/(2(1 - \epsilon))]$  y  $G_2(\epsilon) = F_0^{-1}[\epsilon/(1 - \epsilon)]$  y repetimos el mismo argumento que en la demostración del lema 1.

□

La siguiente proposición es el resultado más importante de esta sección. En ella se dan cotas para la tasa de ruptura de M-estimadores de localización cuando la dispersión se estima previamente mediante un estimador  $D_n$ .

**Proposición 13** Sea  $c = 1/F_0^{-1}(0.75)$  y sea  $TR^+(D)$ , la tasa de ruptura respecto a outliers del estimador de dispersión. Supongamos que se verifican las hipótesis 2 y 4.

Entonces:

(a) Si existe  $A > 0$  tal que  $\psi(y) = 1$  para todo  $y \geq A$ , entonces

$$\max\{1, cTR^+(D)A\} \leq TR(\psi) \leq 1 + cTR^+(D)A.$$

(b) Si  $\psi(y) < 1$  para todo  $y \in \mathbb{R}$ , entonces

$$TR(\psi) = \infty.$$

**Demostración:**

(a) En primer lugar probaremos que  $1 + cTR^+(D)A$  es una cota superior para  $TR(\psi)$ . Sea  $\psi_A$  la función definida en la ecuación (2.22). Es inmediato que  $\psi_A(y) \leq \psi(y)$  para todo  $y \geq 0$ . Esto implica que  $g_A(t, s) \leq g_\psi(t, s)$  para  $s > 0$  y  $t > 0$  y por lo tanto,  $g_\psi^{-1}(x, s) \geq g_A^{-1}(x, s)$  para todo  $s > 0$  y  $0 < x < 1$ . Por la proposición 7,  $B_\psi(\epsilon) \leq B_A(\epsilon)$ . Si aplicamos el lema 3,

$$TR(\psi) \leq TR(\psi_A) = 1 + cTR^+(D)A.$$

Ahora veamos que  $cTR^+(D)A$  es una cota inferior para  $TR(\psi)$ . Definamos para  $0 < r < A$  la siguiente función auxiliar

$$\psi_r(y) = \begin{cases} \text{sgn}(y), & \text{si } |y| > r \\ \psi(r)\text{sgn}(y), & \text{si } |y| \leq r \end{cases}.$$

Operando, es fácil comprobar que, para  $s > 0$  y  $t > 0$ ,

$$\begin{aligned} g_r(t, s) &= s \int_0^\infty \psi_r(y)(f_0(sy - t) - f_0(sy + t))dy \\ &= [1 - \psi(r)][F_0(rs + t) - F_0(rs - t)] + \psi(r)[2F_0(t) - 1]. \end{aligned}$$

Sustituyendo  $t$  por  $B_r(\epsilon)$  y  $s$  por  $D^+(\epsilon)$  en esta ecuación y aplicando la proposición 7,

$$\begin{aligned} \frac{\epsilon}{1-\epsilon} &= [1 - \psi(r)][F_0(rD^+(\epsilon) + B_r(\epsilon)) - F_0(rD^+(\epsilon) - B_r(\epsilon))] \\ &+ \psi(r)[2F_0(B_r(\epsilon)) - 1]. \end{aligned}$$

Ahora tomamos límites cuando  $\epsilon \rightarrow 1/2$  en ambos lados de la ecuación. El miembro de la izquierda converge a uno y lo mismo debe suceder con el segundo miembro. Esto implica en particular que

$$\lim_{\epsilon \rightarrow 1/2} [rD^+(\epsilon) - B_r(\epsilon)] = -\infty.$$

Por consiguiente, existe  $\epsilon_0$  tal que si  $\epsilon_0 < \epsilon < 1/2$ , entonces  $B_r(\epsilon) > rD^+(\epsilon)$ . Se sigue, aplicando el lema 1,

$$TR(\psi_r) \geq cTR^+(D)r. \quad (2.23)$$

Si  $0 < r < A$ , por la definición de  $\psi_r$  se tiene que  $\psi(y) \leq \psi_r(y)$ , para todo  $y \geq 0$ . Por lo tanto, por la proposición 7,  $B_\psi(\epsilon) \geq B_r(\epsilon)$ . Luego, si  $0 < r < A$ ,

$$TR(\psi) \geq TR(\psi_r) \geq cTR^+(D)r.$$

Si hacemos  $r \rightarrow A$ , obtenemos el resultado.

(b) Dado que  $\psi(y) < 1$  para todo  $y$ , se verifica que  $\psi(y) \leq \psi_r(y)$  para todo  $y \geq 0$  y para todo  $r \geq 0$ . Por lo tanto,  $B_\psi(\epsilon) \geq B_r(\epsilon)$  para todo  $r \geq 0$ . Aplicando la desigualdad (2.23) se obtiene, para todo  $r \geq 0$ ,

$$TR(\psi) \geq TR(\psi_r) \geq cTR^+(D)r.$$

Si hacemos  $r \rightarrow \infty$  se obtiene el resultado.

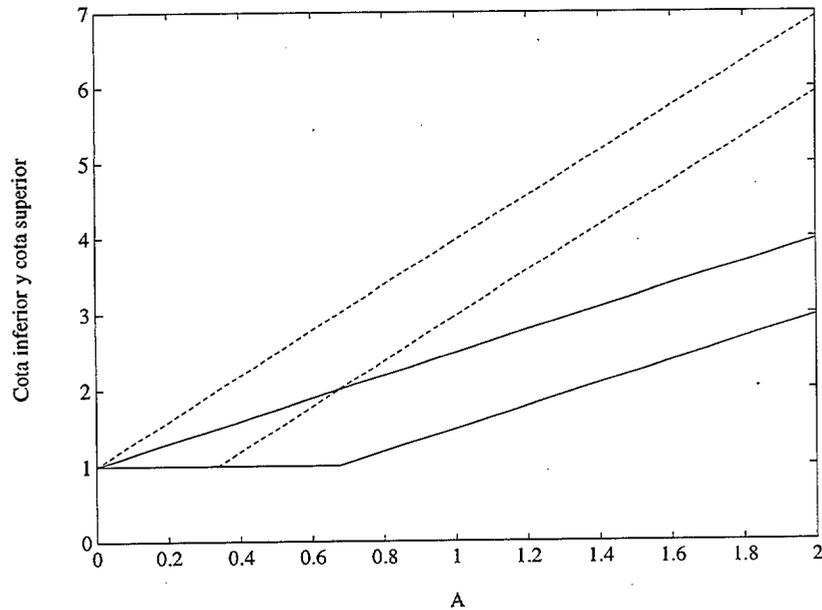
□

Nótese que en el lema 3 hemos dado un ejemplo de M-estimador de localización cuya tasa de ruptura es exactamente igual a  $1 + cTR^+(D)A$ . Por lo tanto, la cota superior obtenida en la proposición 13(a) no puede ser mejorada.

Los M-estimadores de localización basados en funciones  $\psi$  monótonas se pueden contemplar como soluciones al problema de minimizar una función de pérdida convexa. Una función  $\psi$  que alcanza su supremo, como las incluidas en el apartado (a) de la proposición anterior, corresponde a una función de pérdida que es lineal para valores grandes de su argumento. La proposición 13(a) indica que la tasa de ruptura correspondiente a estos estimadores crece de manera aproximadamente lineal con el valor de  $A$  a partir del cual la función de pérdida es lineal. La pendiente de este crecimiento lineal es determinada por la tasa de ruptura respecto a *outliers* del estimador de dispersión elegido previamente. En la figura 2.3 se representan las cotas inferiores y superiores como funciones de  $A$  en el caso gaussiano, cuando tomamos la MEDA y el SHORTH como estimadores previos de dispersión.

Por otra parte, una función  $\psi$  que siempre es estrictamente creciente está ligada a una función de pérdida convexa que siempre crece más lentamente que una función lineal. En este caso, la tasa de ruptura es infinito como se prueba en la parte (b) de la proposición.

Hemos establecido dos clases de funciones de pérdida convexas con un comportamiento bien diferente respecto a la tasa de ruptura. Como hemos visto, un buen comportamiento de la curva de sesgo asintótico para proporciones de contaminación grandes requiere una función de pérdida finalmente lineal. Un ejemplo importante de M-estimadores en esta clase viene dado por los estimadores de Huber de varianza minimax. De hecho, todos los M-estimadores de localización con alguna propiedad de optimalidad relacionada con la robustez tienen esta propiedad, por lo que la proposición



**Figura 2.3:** Cotas inferiores y superiores de  $TR(\psi)$  cuando el estimador de dispersión es la MEDA (línea discontinua) o el SHORT (línea continua).

13(a) confirma las buenas propiedades de robustez de estos estimadores.

En la tabla 2.4 se presenta el punto medio entre la cota inferior y la superior para diversos valores frecuentes de  $A$  y para la MEDA y el SHORT como estimadores de dispersión.

En el caso poco realista en que el parámetro de dispersión es conocido es posible obtener el siguiente resultado.

**Proposición 14** *Supongamos que  $F_0$  verifica la hipótesis 4(a) y que la función  $\psi$  cumple las hipótesis del apartado (a) de la proposición 13. Supongamos que el pa-*

A	SHORTH	MEDA
0.5	1.37	1.99
1	1.99	3.48
1.25	2.36	4.22
1.5	2.73	4.97
1.75	3.10	5.71

**Tabla 2.4:** Punto medio entre las cotas inferiores y superiores de la tasa de ruptura para varios M-estimadores de localización

*rámetro de dispersión es conocido. Entonces:*

$$TR(\psi) = 1.$$

**Demostración:**

Definimos la función auxiliar  $\psi_A$  como en la ecuación (2.22). Dado que  $\psi_A(y) \leq \psi(y)$  para todo  $y \geq 0$ , se verifica  $g_A(t) \leq g_\psi(t)$  para todo  $t \geq 0$  y, por lo tanto,  $B_A(\epsilon) \geq B_\psi(\epsilon)$ .

Operando, es fácil comprobar que

$$\begin{aligned} g_A(t) &= \int_0^\infty \psi_A(y)[f_0(y-t) - f_0(y+t)]dy \\ &= F_0(t-A) + F_0(t+A) - 1 > 2F_0(t-A) - 1. \end{aligned}$$

Tomando en esta desigualdad  $t = g_A^{-1}[\epsilon/(1-\epsilon)] \equiv B_A(\epsilon)$  y operando, se obtiene

$$\frac{1}{2(1-\epsilon)} > F_0[B_A(\epsilon) - A].$$

Si aplicamos  $F_0^{-1}$  a los dos miembros de esta desigualdad se deduce  $B_0(\epsilon) > B_A(\epsilon) - A$ , y por consiguiente,

$$B_A(\epsilon) < B_0(\epsilon) + A.$$

Dado que, como hemos visto,  $B_A(\epsilon) \geq B_\psi(\epsilon)$ , aplicando la propiedad minimax del sesgo de la mediana se tiene que  $B_\psi(\epsilon) \geq B_0(\epsilon)$ . Uniendo todas las desigualdades,

$$B_0(\epsilon) \leq B_\psi(\epsilon) < B_0(\epsilon) + A \quad (2.24)$$

de donde se sigue que  $TR(\psi) = 1$ .

□

La desigualdad (2.24) nos da información adicional sobre el comportamiento del sesgo de  $T_\psi$ . Resulta que  $B_\psi(\epsilon)$  difiere del sesgo de la mediana a lo más en la constante  $A$ . Esta proposición, junto con el resto de resultados de esta sección, permite concluir que la pérdida de robustez respecto a la tasa de ruptura que se produce al estimar la localización se debe principalmente a la estimación del parámetro de dispersión desconocido.

## 2.6 Estimación simultánea de localización y dispersión

En este apartado consideramos de nuevo la función de salto (2.15), utilizada para definir tanto la MEDA como el SHORTH. La estrategia de estimación que consideramos ahora es la de estimar conjuntamente la dispersión y la localización mediante la resolución conjunta del sistema formado por las ecuaciones (1.11) y (1.12).

Huber (1981), páginas 141–143, estudia el punto de ruptura de los estimadores obtenidos por este método. Sin embargo, sus resultados no se aplican al importante

caso de la función  $\chi_a$  puesto que sus supuestos exigen que las funciones  $\psi$  y  $\chi$  sean estrictamente crecientes para argumentos positivos. La siguiente proposición complementa el resultado de Huber y muestra cómo esta estrategia de estimación simultánea lleva en los casos más interesantes a puntos de ruptura drásticamente más bajos que los de los métodos considerados en las secciones 2.4 y 2.5. Lo que significa esta proposición es que si modificamos ligeramente la función correspondiente a la localización,  $\psi(y) = \text{sgn}(y)$ , que define la mediana, el punto de ruptura se reduce a  $\epsilon^* = 1/4$  como máximo, incluso tomando para la dispersión una función tan robusta como  $\chi_a$ .

**Proposición 15** *Consideremos las funciones  $\chi_a$ , definida en (2.15) y  $\psi$  tal que verifica la hipótesis 2. Supongamos que  $F_0$  verifica la hipótesis 4.*

(a) *Si  $\psi(a) < 1$ , entonces  $\epsilon^* \leq 1/4$ .*

(b) *Si  $\psi(k) = 1$  para algún  $k < a$ , entonces  $\epsilon^* = 1/2$ ,*

*donde  $\epsilon^*$  es el punto de ruptura común de los estimadores simultáneos de localización y dispersión basados en  $\chi$  y  $\psi$ .*

### Demostración:

Sea  $B(\epsilon)$  el sesgo del estimador de localización y sea  $D^+(\epsilon)$  el sesgo respecto a *outliers* del estimador de dispersión. Por la proposición 4, estos sesgos satisfacen las ecuaciones

$$(1 - \epsilon)E_{F_0}\chi_a\left(\frac{y - B(\epsilon)}{D^+(\epsilon)}\right) + \epsilon = \frac{1}{2}, \quad (2.25)$$

$$(1 - \epsilon)E_{F_0}\psi\left(\frac{y - B(\epsilon)}{D^+(\epsilon)}\right) + \epsilon = 0. \quad (2.26)$$

Operando, la ecuación (2.25) se puede escribir como

$$F_0[aD^+(\epsilon) + B(\epsilon)] + F_0[aD^+(\epsilon) - B(\epsilon)] = \frac{3 - 2\epsilon}{2(1 - \epsilon)} \quad (2.27)$$

En primer lugar, veamos que los puntos de ruptura correspondientes a la dispersión y a la localización, respectivamente  $\epsilon_D^*$  y  $\epsilon_T^*$ , son iguales. En efecto, por reducción al absurdo supongamos que  $\epsilon_D^* < \epsilon_T^*$ . Existe entonces  $\epsilon_0 \in (\epsilon_D^*, \epsilon_T^*)$ ,  $\epsilon_0 \neq 1/2$ . Tomamos límites en (2.27) cuando  $\epsilon \rightarrow \epsilon_0$ . Dado que  $\lim_{\epsilon \rightarrow \epsilon_0} D^+(\epsilon) = \infty$  y  $\lim_{\epsilon \rightarrow \epsilon_0} B(\epsilon) < \infty$ , obtenemos

$$2 = \frac{3 - 2\epsilon_0}{2(1 - \epsilon_0)},$$

de donde  $\epsilon_0 = 1/2$ , lo que constituye una contradicción. Análogamente, se llega a una contradicción si se parte de suponer  $\epsilon_T^* < \epsilon_D^*$ .

(a) Probamos ahora el primer apartado de la proposición. Operando, la condición (2.26) puede escribirse como

$$D^+(\epsilon) \int_0^\infty \psi(y)[f_0(D^+(\epsilon)y - B(\epsilon)) - f_0(D^+(\epsilon)y + B(\epsilon))]dy = \frac{\epsilon}{1 - \epsilon}.$$

Definamos  $\psi_a$  de la siguiente forma

$$\psi_a(y) = \begin{cases} \operatorname{sgn}(y), & \text{si } |y| > a \\ \psi(a)\operatorname{sgn}(y), & \text{si } |y| \leq a \end{cases}$$

Como  $\psi(a) < 1$ , se tiene que  $\psi(y) \leq \psi_a(y)$  para  $y > 0$ . Por lo tanto

$$\begin{aligned} \frac{\epsilon}{1 - \epsilon} &\leq s \int_0^\infty \psi_a(y)[f_0(D^+(\epsilon)y - B(\epsilon)) - f_0(D^+(\epsilon)y + B(\epsilon))]dy \\ &= [1 - \psi(a)][F_0(aD^+(\epsilon) + B(\epsilon)) - F_0(aD^+(\epsilon) - B(\epsilon))] \\ &\quad + \psi(a)[2F_0(B(\epsilon)) - 1]. \end{aligned} \quad (2.28)$$

Multiplicando la ecuación (2.27) por  $1 - \psi(a)$ , sumando al resultado la ecuación (2.28) y reordenando términos se obtiene

$$[1 - \psi(a)]F_0[aD^+(\epsilon) + B(\epsilon)] + \psi(a)[F_0(B(\epsilon)) - 1/2] \geq \frac{3}{4(1 - \epsilon)} - \psi(a)\frac{3 - 2\epsilon}{4(1 - \epsilon)} \quad (2.29)$$

Supongamos que  $\epsilon^* > 1/4$ . Si tomamos límites en (2.27) cuando  $\epsilon \rightarrow 1/4$ , el término de la derecha converge a  $1 - \psi(a)/2$  mientras que el límite del miembro de la izquierda,  $L$ , verifica

$$L < 1 - \psi(a) + \psi(a)/2 = 1 - \psi(a)/2,$$

ya que  $F_0$  es estrictamente creciente. Pero esto es imposible por la desigualdad (2.29). Por consiguiente,  $\epsilon^* \leq 1/4$ .

(b) Bajo las hipótesis de este apartado se cumple la condición (2.27), que no depende de la elección de la función  $\psi$ . Además, si definimos  $\psi_k$  como en la ecuación (2.22), dado que  $\psi_k(y) \leq \psi(y)$  para  $y > 0$ , la ecuación (2.26) implica

$$F_0[kD^+(\epsilon) + B(\epsilon)] + F_0[kD^+(\epsilon) - B(\epsilon)] \leq \frac{\epsilon}{1 - \epsilon}. \quad (2.30)$$

Sea  $\epsilon^*$  el punto de ruptura y denotemos por  $L$  a  $\lim_{\epsilon \rightarrow \epsilon^*} [aD^+(\epsilon) - B(\epsilon)]$ . De la ecuación (2.27) se sigue que

$$F_0(L) = \frac{1}{2(1 - \epsilon^*)}.$$

Si  $L = \infty$ , entonces  $\epsilon^* = 1/2$  como queríamos probar. Si  $L < \infty$ , entonces debe cumplirse  $\lim_{\epsilon \rightarrow \epsilon^*} [kD^+(\epsilon) - B(\epsilon)] = -\infty$ , ya que  $k < a$ . Pero entonces de la ecuación (2.30) se deduce

$$\frac{\epsilon^*}{1 - \epsilon^*} \geq 1,$$

de donde  $\epsilon^* \geq 1/2$ , es decir  $\epsilon^* = 1/2$ .

□

## Capítulo 3

### Sesgo de estimadores de regresión con residuos admisibles

#### 3.1 Introducción e hipótesis básicas

Consideremos el modelo de regresión lineal descrito en la sección 1.4. En este capítulo se obtiene un método general que permite calcular curvas de sesgo asintótico máximo para una clase amplia de estimadores de regresión robustos; en concreto, consideraremos la clase de estimadores de regresión con residuos admisibles que ya fue introducida en la definición 1.

Los resultados obtenidos amplían la literatura existente sobre el tema en dos direcciones: en primer lugar, sólo en el caso de S-estimadores de regresión se disponía de una fórmula para la curva de sesgo. Nosotros obtendremos las curvas de sesgo de otras clases de estimadores con residuos admisibles. Además, el método que permite estos cálculos se basa en un principio común, válido para todas estas clases de estimadores. En segundo lugar, los resultados obtenidos hasta ahora requerían que los regresores tuvieran una distribución elíptica, lo que es una hipótesis ciertamente

restrictiva. Nuestros resultados pueden aplicarse a regresores no elípticos, a cambio de tener que resolver numéricamente un problema de optimización que puede resultar complicado.

Al final del capítulo veremos dos modificaciones de los resultados que permiten su aplicación a M-estimadores de regresión con escala general, sección 3.4, y al cálculo de sesgo en dirección de algunos métodos de análisis robusto de componentes principales, sección 3.5.

Los estimadores de regresión que vamos a considerar vienen definidos por la expresión (1.18), donde el funcional  $J$ , que mide la escala de los residuos, verifica la siguiente hipótesis:

**Hipótesis 5** (a) Si  $F$  y  $G$  son dos funciones de distribución sobre  $[0, \infty)$  tales que

$$F(u) \leq G(u) \text{ para todo } u \in \mathbb{R}, \text{ entonces } J(F) \geq J(G).$$

(b) ( $\eta$ -monotonía). Dadas dos sucesiones de funciones de distribución,  $F_n$  y  $G_n$ , sobre  $[0, \infty)$ , continuas en  $(0, \infty)$  y tales que  $F_n(u) \rightarrow F(u)$  y  $G_n(u) \rightarrow G(u)$ , donde  $F$  y  $G$  son posiblemente subestocásticas y continuas en  $(0, \infty)$ , con  $G(\infty) \geq 1 - \eta$  y

$$G(u) \geq F(u), \text{ para todo } u > 0, \quad (3.1)$$

entonces

$$\lim_{n \rightarrow \infty} J(F_n) \geq \lim_{n \rightarrow \infty} J(G_n). \quad (3.2)$$

Además, si (3.1) se cumple estrictamente, entonces (3.2) también se cumple estrictamente.

La hipótesis 5(a) requiere que  $J$  sea monótono respecto a la relación de dominancia estocástica para distribuciones de variables aleatorias no negativas.

La hipótesis 5(b) de  $\eta$ -monotonía fue introducida por Yohai y Zamar (1993) y es suficiente que se verifique para algún  $\eta > 0$  para que el estimador  $\mathbf{T}$  tenga residuos admisibles. Todos los ejemplos interesantes de estimadores con residuos admisibles son también  $\eta$ -monótonos y por lo tanto, en lo que se refiere a las aplicaciones de los resultados, no es importante la diferencia entre asumir admisibilidad de los residuos o  $\eta$ -monotonía.

Si consideramos  $F_n = F$  y  $G_n = G$  para todo  $n$  en la hipótesis 5(b) y si  $F$  y  $G$  son funciones de distribución sobre  $[0, \infty)$ , continuas sobre  $(0, \infty)$  y tales que  $G(u) > F(u)$  para  $u > 0$ , entonces  $J(F) > J(G)$ . Por tanto, también podemos contemplar la  $\eta$ -monotonía como una condición de monotonía estricta para distribuciones que reúnan ciertas condiciones.

Para el cálculo de la curva de sesgo de los estimadores de este capítulo en el punto  $\epsilon$ ,  $B_{\mathbf{T}}(\epsilon)$ , es suficiente que se verifique la  $\eta$ -monotonía para  $\eta = \epsilon$ . La situación es ligeramente distinta en el capítulo cuatro.

Será necesario suponer también la siguiente condición sobre el funcional  $J$ :

**Hipótesis 6** *Si  $F$  y  $G$  son dos funciones de distribución sobre  $[0, \infty)$ , con  $F$  continua, entonces*

$$\lim_{n \rightarrow \infty} J[(1 - \epsilon)F + \epsilon U_n] \geq J[(1 - \epsilon)F + \epsilon G],$$

donde  $U_n$  denota la función de distribución uniforme sobre  $[n - (1/n), n + (1/n)]$ .

Esta hipótesis es de carácter técnico y se puede comprobar fácilmente para todos los ejemplos importantes. Lo que se exige esencialmente es que si contaminamos en un punto y llevamos el punto tan lejos como queramos, se alcance una escala mayor que la obtenida contaminando de cualquier otra manera. La utilización de distribuciones

uniformes en lugar de distribuciones degeneradas se debe a que en algunas situaciones es necesario que las distribuciones sean continuas.

Es inmediato demostrar que si la hipótesis 5(a) se verifica para distribuciones posiblemente subestocásticas y se cumple la  $\eta$ -monotonía para  $\eta = \epsilon$ , entonces también se verifica la hipótesis 6. Veamos la prueba: las distribuciones de la sucesión  $\{(1 - \epsilon)F + \epsilon U_n\}$  son continuas. Además

$$\lim_{n \rightarrow \infty} (1 - \epsilon)F(u) + \epsilon U_n(u) = (1 - \epsilon)F(u), \text{ para todo } u > 0.$$

Por la  $\epsilon$ -monotonía,

$$\lim_{n \rightarrow \infty} J[(1 - \epsilon)F + \epsilon U_n] = J[(1 - \epsilon)F].$$

Pero por otra parte,

$$J[(1 - \epsilon)F] \geq J[(1 - \epsilon)F + \epsilon G],$$

ya que estamos suponiendo que la hipótesis 5(a) también se verifica para distribuciones subestocásticas.

Finalmente, hemos de imponer la siguiente condición sobre las distribuciones nominales de los regresores,  $G_0$ , y de los errores,  $F_0$ :

**Hipótesis 7** *La distribución de los errores  $F_0$  tiene una densidad  $f_0$  que es simétrica y estrictamente unimodal. La distribución de los regresores  $G_0$  tiene densidad y  $P_{G_0}\{\theta'x = 0\} = 0$ , para todo  $\theta \neq 0$ .*

Nótese que no es necesario suponer que  $G_0$  sea una distribución elíptica.

### 3.2 Cálculo del sesgo de estimadores de residuos admisibles

El siguiente lema es necesario para probar el resultado más importante de este capítulo:

**Lema 4** Definamos, para todo  $t \geq 0$ , la función

$$m(t) = \inf_{\|\boldsymbol{\theta}\|=t} J[(1 - \epsilon)F_{H_0, \boldsymbol{\theta}} + \epsilon\delta_0].$$

Entonces, bajo la hipótesis 5(b) para algún  $\eta > 0$  y la hipótesis 7 se tiene que

(a) Existe  $\boldsymbol{\theta}_t \in \mathbb{R}^p$  tal que  $\|\boldsymbol{\theta}_t\| = t$ , y  $m(t) = J[(1 - \epsilon)F_{H_0, \boldsymbol{\theta}_t} + \epsilon\delta_0]$ .

(b)  $m(t)$  es estrictamente creciente.

**Demostración:**

De la hipótesis 5(b) se deduce que si  $F_n$  y  $F$  son funciones de distribución sobre  $[0, \infty)$ , continuas sobre  $(0, \infty)$  y tales que  $F_n(u) \rightarrow F(u)$  para todo  $u > 0$ , entonces

$$\lim_{n \rightarrow \infty} J(F_n) = J(F).$$

Supongamos que  $\{\boldsymbol{\theta}_n\}$  es una sucesión tal que  $\boldsymbol{\theta}_n \rightarrow \boldsymbol{\theta}$ . Definamos

$$F_n = (1 - \epsilon)F_{H_0, \boldsymbol{\theta}_n} + \epsilon\delta_0.$$

Por la observación anterior,  $\lim_{n \rightarrow \infty} J(F_n) = J[(1 - \epsilon)F_{H_0, \boldsymbol{\theta}} + \epsilon\delta_0]$ , es decir, la función  $f(\boldsymbol{\theta}) = J[(1 - \epsilon)F_{H_0, \boldsymbol{\theta}} + \epsilon\delta_0]$  es continua. Como el conjunto  $\{\boldsymbol{\theta} : \|\boldsymbol{\theta}\| = t\}$  es compacto para todo  $t$ , el ínfimo en la definición de  $m(t)$  es de hecho un mínimo. Esto prueba la parte (a).

En el lemma A.1 de Yohai y Zamar (1993) se demuestra que, bajo la hipótesis 7,  $F_{H_0, \lambda \boldsymbol{\theta}}(u)$  es estrictamente decreciente como función de  $\lambda$ . Sean  $t_1$  y  $t_2$  tales que  $t_1 > t_2 > 0$ . Por la parte (a), existe  $\boldsymbol{\theta}_1$  tal que  $m(t_1) = J[(1 - \epsilon)F_{H_0, \boldsymbol{\theta}_1} + \epsilon\delta_0]$ . Como  $F_{H_0, \boldsymbol{\theta}_1}(u) < F_{H_0, (t_2/t_1)\boldsymbol{\theta}_1}(u)$ , se sigue de la hipótesis 5(b) que

$$m(t_1) = J[(1 - \epsilon)F_{H_0, \boldsymbol{\theta}_1} + \epsilon\delta_0] > J[(1 - \epsilon)F_{H_0, (t_2/t_1)\boldsymbol{\theta}_1} + \epsilon\delta_0].$$

Pero, por definición de  $m(t)$ ,

$$J[(1 - \epsilon)F_{H_0, (t_2/t_1)\theta_1} + \epsilon\delta_0] \geq m(t_2).$$

Al unir las dos últimas desigualdades, se obtiene la parte (b). □

Estamos ya en disposición de enunciar y demostrar el resultado más importante de este capítulo.

**Proposición 16** *Sea  $\mathbf{T}$  un estimador definido según la ecuación (1.18) a partir de un funcional  $J$  que verifica la hipótesis 5 para  $\eta = \epsilon$  y la hipótesis 6. Sea  $t^* \in \mathbb{R}$  tal que*

$$m(t^*) \equiv \inf_{\|\theta\|=t^*} J[(1 - \epsilon)F_{H_0, \theta} + \epsilon\delta_0] = \lim_{n \rightarrow \infty} J[(1 - \epsilon)F_{H_0, 0} + \epsilon U_n], \quad (3.3)$$

donde  $U_n$  denota la distribución uniforme sobre el intervalo  $(n - 1/n, n + 1/n)$ . Entonces, bajo la hipótesis 7, se tiene que

$$B_{\mathbf{T}}(\epsilon) = t^*.$$

### Demostración:

Supongamos que  $\tilde{\theta} \in \mathbb{R}^p$  es tal que  $t = \|\tilde{\theta}\| > t^*$ . Se probará en primer lugar que  $J(F_{H, \tilde{\theta}}) > J(F_{H, 0})$  para cualquier  $H \in V_\epsilon$ . Este hecho implica que  $\tilde{\theta}$  no puede ser solución del problema de minimización (1.18) para ninguna distribución del entorno de contaminación y, como consecuencia,  $B_{\mathbf{T}}(\epsilon) \leq t^*$ .

Sea  $H \in V_\epsilon$ . Para todo  $u > 0$  se verifica

$$F_{H, \tilde{\theta}}(u) \leq (1 - \epsilon)F_{H_0, \tilde{\theta}}(u) + \epsilon\delta_0(u).$$

Aplicando la hipótesis 5(a), la definición de la función  $m(t)$  y el lema 4 junto con el hecho de que  $t > t^*$ , se tiene para todo  $H \in V_\epsilon$ ,

$$J(F_{H,\tilde{\theta}}) \geq J[(1 - \epsilon)F_{H_0,\tilde{\theta}} + \epsilon\delta_0] \geq m(t) > m(t^*). \quad (3.4)$$

Por otra parte, por la ecuación (3.3),

$$m(t^*) = \lim_{n \rightarrow \infty} J[(1 - \epsilon)F_{H_0,0} + \epsilon U_n] \quad (3.5)$$

y, por la hipótesis 6,

$$\lim_{n \rightarrow \infty} J[(1 - \epsilon)F_{H_0,0} + \epsilon U_n] \geq J(F_{H,0}). \quad (3.6)$$

Teniendo en cuenta (3.4), (3.5) y (3.6), se obtiene

$$J(F_{H,\tilde{\theta}}) > J(F_{H,0}), \quad \text{para todo } H \in V_\epsilon,$$

que es lo que queríamos probar.

Estableceremos a continuación la desigualdad recíproca,  $B_{\mathbf{T}}(\epsilon) \geq t^*$ . Para ello, sea  $t < t^*$  y  $\theta_t \in \mathbb{R}^p$  tal que  $\|\theta_t\| = t$  y

$$m(t) = J[(1 - \epsilon)F_{H_0,\theta_t} + \epsilon\delta_0].$$

Tal vector  $\theta_t$  existe en aplicación del lema 4(a). Definamos la siguiente sucesión de contaminaciones:  $\tilde{H}_n$  es la distribución de  $(y_n, \mathbf{x}_n)$ , donde  $\mathbf{x}_n = n\theta_t$  e  $y_n$  se distribuye uniformemente sobre el intervalo  $[nt^2 - (1/n), nt^2 + (1/n)]$ . Si  $F_n$  es la función de distribución uniforme sobre el intervalo  $[-1/n, 1/n]$ , entonces para todo  $\beta \in \mathbb{R}^p$  y  $u > 0$ , se cumple

$$F_{\tilde{H}_n,\beta}(u) = F_n[u - n(t^2 - \beta'\theta_t)] - F_n[-u - n(t^2 - \beta'\theta_t)]. \quad (3.7)$$

Sea ahora  $H_n = (1 - \epsilon)H_0 + \epsilon\tilde{H}_n$  y denotemos  $\mathbf{T}_n = \mathbf{T}(H_n)$ . Supongamos, por reducción al absurdo, que  $\sup_n \|\mathbf{T}_n\| < t$ . En este caso, existirá una subsucesión de  $\{\mathbf{T}_n\}$ , que denotamos también por  $\{\mathbf{T}_n\}$ , tal que  $\lim_{n \rightarrow \infty} \mathbf{T}_n = \tilde{\boldsymbol{\theta}}$ , donde  $\tilde{t} = \|\tilde{\boldsymbol{\theta}}\| < t$ .

Nótese que  $0 \leq |\boldsymbol{\theta}'_t \tilde{\boldsymbol{\theta}}| \leq \|\boldsymbol{\theta}_t\| \|\tilde{\boldsymbol{\theta}}\| = t\tilde{t} < t^2$ . Por lo tanto,  $|t^2 - |\boldsymbol{\theta}'_t \tilde{\boldsymbol{\theta}}|| > 0$ . Se sigue de (3.7) que

$$\lim_{n \rightarrow \infty} F_{\tilde{H}_n, \mathbf{T}_n}(u) = 0, \quad \text{para todo } u > 0. \quad (3.8)$$

Además,  $t^2 - \boldsymbol{\theta}'_t \boldsymbol{\theta}_t = 0$ . Se sigue de (3.7) que

$$\lim_{n \rightarrow \infty} F_{\tilde{H}_n, \boldsymbol{\theta}_t}(u) = 1, \quad \text{para todo } u > 0. \quad (3.9)$$

Teniendo en cuenta (3.8) y el lema A.1 de Yohai y Zamar (1993), tenemos que para todo  $u > 0$ ,

$$\begin{aligned} \lim_{n \rightarrow \infty} F_{H_n, \mathbf{T}_n}(u) &= (1 - \epsilon)F_{H_0, \tilde{\boldsymbol{\theta}}}(u) \\ &\leq (1 - \epsilon)F_{H_0, 0}(u) \\ &= \lim_{n \rightarrow \infty} [(1 - \epsilon)F_{H_0, 0}(u) + \epsilon U_n(u)]. \end{aligned} \quad (3.10)$$

Por la hipótesis 5(b), la desigualdad (3.10) y la propiedad (3.3) que verifica  $t^*$ ,

$$\lim_{n \rightarrow \infty} J(F_{H_n, \mathbf{T}_n}) \geq \lim_{n \rightarrow \infty} J[(1 - \epsilon)F_{H_0, 0} + \epsilon U_n] = m(t^*). \quad (3.11)$$

A partir de (3.9), para todo  $u > 0$  se verifica

$$\lim_{n \rightarrow \infty} F_{H_n, \boldsymbol{\theta}_t}(u) = (1 - \epsilon)F_{H_0, \boldsymbol{\theta}_t}(u) + \epsilon \delta_0(u). \quad (3.12)$$

Por la hipótesis 5(b) y la ecuación (3.12), junto con el lema 4(b) y el hecho de que  $t < t^*$ ,

$$\lim_{n \rightarrow \infty} J[F_{H_n, \boldsymbol{\theta}_t}] = J[(1 - \epsilon)F_{H_0, \boldsymbol{\theta}_t} + \epsilon \delta_0] = m(t) < m(t^*). \quad (3.13)$$

Al unir las desigualdades (3.11) y (3.13), se deduce que

$$\lim_{n \rightarrow \infty} J(F_{H_n}, \mathbf{T}_n) > \lim_{n \rightarrow \infty} J(F_{H_n}, \boldsymbol{\theta}_t),$$

y, por lo tanto, existe  $n$  suficientemente grande para el cual

$$J(F_{H_n}, \mathbf{T}_n) > J(F_{H_n}, \boldsymbol{\theta}_t),$$

en contradicción con el hecho de que

$$\mathbf{T}_n = \arg \min_{\boldsymbol{\theta}} J(F_{H_n}, \boldsymbol{\theta}).$$

La contradicción procede de suponer  $\sup_n \|\mathbf{T}(H_n)\| < t$  y por lo tanto se verifica que  $\sup_n \|\mathbf{T}(H_n)\| \geq t$ .

Para cualquier  $t < t^*$ , existe una distribución  $H_n \in V_\epsilon$  tal que  $\|\mathbf{T}(H_n)\| \geq t$ . Se sigue que  $B_{\mathbf{T}}(\epsilon) \geq t^*$ . □

Cuando la distribución de los regresores,  $G_0$ , es esférica es fácil comprobar que  $F_{H, \boldsymbol{\theta}}$  depende de  $\boldsymbol{\theta}$  únicamente a través del valor de  $\|\boldsymbol{\theta}\|$ . Por lo tanto, podemos eliminar el ínfimo en la ecuación (3.3) y obtener el siguiente corolario:

**Corolario 4** *Bajo las hipótesis y la notación de la proposición 16 y suponiendo además que  $G_0$  es esférica se verifica*

$$B_{\mathbf{T}}(\epsilon) = \|\boldsymbol{\theta}^*\|,$$

donde  $\boldsymbol{\theta}^* \in \mathbb{R}^p$  es tal que

$$J[(1 - \epsilon)F_{H_0, \boldsymbol{\theta}^*} + \epsilon\delta_0] = \lim_{n \rightarrow \infty} J[(1 - \epsilon)F_{H_0, 0} + \epsilon U_n]. \quad (3.14)$$

Sea  $\delta_n$  la distribución que asigna probabilidad uno al punto  $n$ . En la mayoría de los ejemplos, se cumple que

$$\lim_{n \rightarrow \infty} J[(1 - \epsilon)F_{H_0,0} + \epsilon U_n] = \lim_{n \rightarrow \infty} J[(1 - \epsilon)F_{H_0,0} + \epsilon \delta_n] \equiv J[(1 - \epsilon)F_{H_0,0} + \epsilon \delta_\infty].$$

Por lo tanto, la igualdad (3.14) puede escribirse

$$J[(1 - \epsilon)F_{H_0, \theta^*} + \epsilon \delta_0] = J[(1 - \epsilon)F_{H_0,0} + \epsilon \delta_\infty].$$

Podemos utilizar esta última expresión para dar una interpretación intuitiva del corolario 4. Supongamos que hay una proporción  $\epsilon$  de *outliers* en el punto  $(\theta^{*\prime} \mathbf{x}, \mathbf{x})$ . Entonces,  $(1 - \epsilon)F_{H_0, \theta^*} + \epsilon \delta_0$  es la distribución de los residuos cuando esta proporción de *outliers* está perfectamente ajustada y  $J[(1 - \epsilon)F_{H_0, \theta^*} + \epsilon \delta_0]$  es la escala de estos residuos. Por otra parte,  $J[(1 - \epsilon)F_{H_0,0} + \epsilon \delta_\infty]$  representa la escala de los residuos cuando los *outliers* son completamente ignorados y  $\|\theta^* \mathbf{x}\| \rightarrow \infty$ . En el corolario 4 se ha probado que el sesgo máximo es el valor de  $\|\theta^*\|$  tal que ambas escalas coinciden.

### 3.3 Ejemplos de aplicación

En esta sección se muestran distintas aplicaciones de los resultados de la sección 3.2 para obtener las curvas de sesgo de distintos estimadores robustos de regresión que aparecen con frecuencia en la literatura. Para simplificar los cálculos numéricos suponemos, salvo que se indique lo contrario, que  $G_0$  es una distribución normal  $p$ -variante y que  $F_0$  es una distribución normal estándar.

#### 3.3.1 S-estimadores

En primer lugar, consideremos un S-estimador de regresión,  $\mathbf{T}$ , que minimiza un M-estimador de escala basado en una función  $\chi$ . Estos estimadores fueron definidos en la



ecuación (1.19).

Los  $S$ -estimadores cumplen las hipótesis que permiten aplicar la proposición 16. Ello es consecuencia del lema siguiente:

**Lema 5** *Si  $\chi \in C_b$ , entonces el funcional  $S(F)$ , definido mediante (1.3), verifica las hipótesis 5 y 6.*

**Demostración:**

Yohai y Zamar (1993), lema 5.1., prueban que  $S(F)$  cumplen la hipótesis de  $\eta$ -monotonía, para todo  $\eta > 0$ .

Veamos que  $S(F)$  cumple la hipótesis 5(a). Para ello consideremos dos funciones de distribución,  $F$  y  $G$ , sobre  $[0, \infty)$  tales que  $F(u) \leq G(u)$  para todo  $u \geq 0$ . Es posible construir dos variables aleatorias  $X$  e  $Y$  distribuidas como  $F$  y  $G$  respectivamente y tales que  $X \geq Y$  (c.s.). Dado que  $\chi$  es no decreciente en  $(0, \infty)$ , para todo  $s > 0$  se tiene que  $\chi(X/s) \geq \chi(Y/s)$  (c.s.). Por lo tanto, si  $s > 0$  es tal que  $E\chi(X/s) \leq b$ , entonces también  $E\chi(Y/s) \leq b$ . Como consecuencia,

$$S(F) = \inf\{s > 0 : E\chi(X/s) \leq b\} \geq \inf\{s > 0 : E\chi(Y/s) \leq b\} = S(G).$$

Veamos ahora que

$$s_\infty \equiv \lim_{n \rightarrow \infty} s_n = \lim_{n \rightarrow \infty} S[(1 - \epsilon)F + \epsilon U_n] \geq S[(1 - \epsilon)F + \epsilon G] \equiv s_0.$$

Por definición de  $M$ -estimador de escala,

$$(1 - \epsilon) \int_{-\infty}^{\infty} \chi\left(\frac{y}{s_n}\right) dF(y) + \epsilon \frac{n}{2} \int_{n^{-1/n}}^{n^{+1/n}} \chi\left(\frac{y}{s_n}\right) dy \leq b,$$

y haciendo  $n \rightarrow \infty$ ,

$$(1 - \epsilon) \int_{-\infty}^{\infty} \chi\left(\frac{y}{s_{\infty}}\right) dF(y) + \epsilon \leq b. \quad (3.15)$$

Supongamos ahora que

$$(1 - \epsilon) \int_{-\infty}^{\infty} \chi\left(\frac{y}{s_0}\right) dF(y) + \epsilon < b.$$

Entonces, existe  $s < s_0$  tal que

$$(1 - \epsilon) \int_{-\infty}^{\infty} \chi\left(\frac{y}{s}\right) dF(y) + \epsilon < b,$$

lo que implica que

$$(1 - \epsilon) \int_{-\infty}^{\infty} \chi\left(\frac{y}{s}\right) dF(y) + \epsilon \int_{-\infty}^{\infty} \chi\left(\frac{y}{s}\right) dG(y) \leq (1 - \epsilon) \int_{-\infty}^{\infty} \chi\left(\frac{y}{s}\right) dF(y) + \epsilon < b,$$

en contradicción con la definición de  $s_0$ . Por lo tanto

$$(1 - \epsilon) \int_{-\infty}^{\infty} \chi\left(\frac{y}{s_0}\right) dF(y) + \epsilon \geq b.$$

Teniendo en cuenta (3.15), se deduce que  $s_{\infty} \geq s_0$ .

□

Veamos cómo utilizar el corolario 4 para obtener la curva de sesgo de los S-estimadores de regresión. Definamos  $g(s) = E_{\Phi} \chi(y/s)$ , donde  $\Phi$  denota la función de distribución normal estándar. Sea  $S_1 = S[(1 - \epsilon)F_{H_0, \theta^*} + \epsilon \delta_0]$ . Como la distribución de  $y - \theta^{*'} \mathbf{x}$  es normal con media 0 y varianza  $1 + \|\theta^*\|^2$ ,  $S_1$  verifica

$$(1 - \epsilon) E_{H_0} \chi\left(\frac{y - \theta^{*'} \mathbf{x}}{S_1}\right) = (1 - \epsilon) g\left(\frac{S_1}{(1 + \|\theta^*\|^2)^{1/2}}\right) = b.$$

Por lo tanto

$$S_1 = (1 + \|\theta^*\|^2)^{1/2} g^{-1}\left(\frac{b}{1 - \epsilon}\right). \quad (3.16)$$

Sea ahora  $S_2 = \lim_{n \rightarrow \infty} S[(1 - \epsilon)F_{H_0,0} + \epsilon\delta_n]$ . Se tiene que  $S_2$  cumple

$$(1 - \epsilon)E_{H_0}\chi(Y/S_2) + \epsilon = (1 - \epsilon)g(S_2) + \epsilon = b,$$

de donde resulta

$$S_2 = g^{-1}\left(\frac{b - \epsilon}{1 - \epsilon}\right). \quad (3.17)$$

Para obtener el valor del sesgo aplicando el corolario 4, calculamos el valor de  $\|\theta^*\|$  que resulta al despejar en la ecuación  $S_1 = S_2$ . Usando las expresiones (3.16) y (3.17) resulta

$$B_S^2(\epsilon) = \|\theta^*\|^2 = \left[ \frac{g^{-1}\left(\frac{b - \epsilon}{1 - \epsilon}\right)}{g^{-1}\left(\frac{b}{1 - \epsilon}\right)} \right]^2 - 1. \quad (3.18)$$

Nótese que esta expresión coincide con la fórmula (3.18) de Martin *et al.* (1989), que ya se mencionó en el capítulo uno, expresión (1.20).

### 3.3.2 $\tau$ -estimadores

Calculamos en este apartado la curva de sesgo de un  $\tau$ -estimador de regresión. Para la definición de estos estimadores véanse las expresiones (1.22) y (1.23).

En el siguiente lema se comprueba que la proposición 16 se puede aplicar para calcular la curva de sesgo de los  $\tau$ -estimadores de regresión.

**Lema 6** Sean  $\chi_1 \in C_b$  y  $\chi_2 \in C_{b_2}$ . Supongamos que  $\chi_2$  es tal que la función  $f(s) = s^2 E_F \chi_2(y/s)$  es creciente para cualquier  $F$ . Entonces, el funcional  $\tau(F)$  verifica las hipótesis 5 y 6.

**Demostración:**

Para comprobar la hipótesis de  $\eta$ -monotonía, para todo  $\eta > 0$ , véase el trabajo de Yohai y Zamar (1993), lema 5.1.

Probemos que  $\tau(F)$  cumple la hipótesis 5(a). Para ello consideremos dos funciones de distribución,  $F$  y  $G$ , sobre  $[0, \infty)$  tales que  $F(u) \leq G(u)$  para todo  $u \geq 0$ . Por el lema 5 tenemos que  $S^2(F) \geq S^2(G)$ . Por lo tanto,

$$\begin{aligned} \tau(F) &= S^2(F)E_{F\chi_2}[y/S(F)] \geq S^2(G)E_{F\chi_2}[y/S(G)] \\ &\geq S^2(G)E_{G\chi_2}[y/S(G)] = \tau^2(G), \end{aligned}$$

donde la última desigualdad ha sido obtenida aplicando los mismos razonamientos que en la demostración del lema 5.

Para comprobar la hipótesis 6, denotemos  $s_\infty = \lim_{n \rightarrow \infty} S[(1 - \epsilon)F + \epsilon U_n]$  y  $s_0 = S[(1 - \epsilon)F + \epsilon G]$ . Por el lema 5, se tiene que

$$\begin{aligned} \tau_\infty^2 &= \lim_{n \rightarrow \infty} \tau^2[(1 - \epsilon)F + \epsilon U_n] \geq s_0^2 [(1 - \epsilon)E_{F\chi_2}(y/s_0) + \epsilon] \\ &\geq s_0^2 [(1 - \epsilon)E_{F\chi_2}(y/s_0) + \epsilon E_{G\chi_2}(y/s_0)] = \tau^2[(1 - \epsilon)F + \epsilon G]. \end{aligned}$$

□

De manera análoga a como se hizo en el apartado anterior definamos

$$\begin{aligned} \tau_1 &= \tau^2[(1 - \epsilon)F_{H_0, \theta^*} + \epsilon \delta_0], \\ \tau_2 &= \lim_{n \rightarrow \infty} \tau^2[(1 - \epsilon)F_{H_0, 0} + \epsilon \delta_n], \end{aligned}$$

y  $g_i(s) = E_{\Phi\chi_i}(Y/s)$ , para  $i = 1, 2$ . Tras algunos cálculos similares a los que ya hemos hecho en el caso de S-estimadores, obtenemos

$$\tau_1 = (1 + \|\theta^*\|^2) \left[ g_1^{-1} \left( \frac{b}{1 - \epsilon} \right) \right]^2 (1 - \epsilon) g_2 \left[ g_1^{-1} \left( \frac{b}{1 - \epsilon} \right) \right],$$

y

$$\tau_2 = \left[ g_1^{-1} \left( \frac{b-\epsilon}{1-\epsilon} \right) \right]^2 \left[ (1-\epsilon)g_2 \left[ g_1^{-1} \left( \frac{b}{1-\epsilon} \right) \right] + \epsilon \right].$$

Imponemos ahora la condición  $\tau_1 = \tau_2$  para obtener

$$B_\tau^2(\epsilon) = \left[ \frac{g_1^{-1} \left( \frac{b-\epsilon}{1-\epsilon} \right)}{g_1^{-1} \left( \frac{b}{1-\epsilon} \right)} \right]^2 \left[ \frac{g_2 \left[ g_1^{-1} \left( \frac{b-\epsilon}{1-\epsilon} \right) \right]}{g_2 \left[ g_1^{-1} \left( \frac{b}{1-\epsilon} \right) \right]} + \frac{\epsilon}{1-\epsilon} \frac{1}{g_2 \left[ g_1^{-1} \left( \frac{b}{1-\epsilon} \right) \right]} \right] - 1. \quad (3.19)$$

Si denotamos

$$H(\epsilon) = \left[ \frac{g_2 \left[ g_1^{-1} \left( \frac{b-\epsilon}{1-\epsilon} \right) \right]}{g_2 \left[ g_1^{-1} \left( \frac{b}{1-\epsilon} \right) \right]} + \frac{\epsilon}{1-\epsilon} \frac{1}{g_2 \left[ g_1^{-1} \left( \frac{b}{1-\epsilon} \right) \right]} \right],$$

de (3.18) y (3.19) se deduce

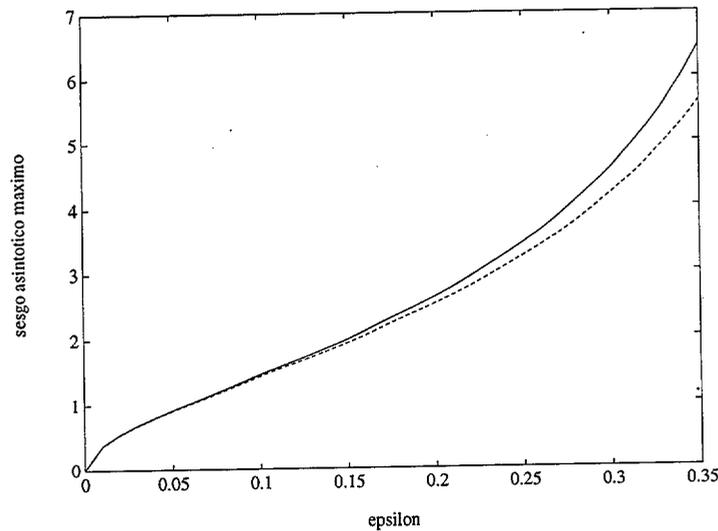
$$1 + B_\tau^2(\epsilon) = [1 + B_S^2(\epsilon)]H(\epsilon). \quad (3.20)$$

Esta es la relación entre el sesgo de un S-estimador basado en la función  $\chi_1$  y el sesgo del  $\tau$ -estimador basado en  $\chi_1$  y  $\chi_2$ . Dado que  $0 < H(\epsilon) < \infty$  para  $\epsilon < \min\{b, 1-b\}$ , el  $\tau$ -estimador hereda el punto de ruptura del S-estimador basado en  $\chi_1$  independientemente de la elección de  $\chi_2$ . Esta función se puede seleccionar con el fin de alcanzar una eficiencia tan alta como se desee. (Véanse más detalles sobre este aspecto en Yohai y Zamar, 1988).

En la tabla 3.1 de la página 96 se presentan los valores numéricos de la curva de sesgo para distintas proporciones de contaminación de algunos S- y  $\tau$ -estimadores, todos ellos con punto de ruptura  $\epsilon^* = 0.5$ . Hemos considerado el S-estimador basado en una función bicuadrada de Tukey (estimador STUKEY) que ya fue definida en la tabla 2.1. Posteriormente, hemos combinado este estimador con otra función bicuadrada de Tukey con el fin de obtener un  $\tau$  estimador con punto de ruptura  $\epsilon^* = 0.5$  y eficiencia del 95% (estimador TAUTUKEY). El sesgo máximo de STUKEY es considerablemente menor que el de TAUTUKEY si bien su eficiencia es mucho más baja. Basta recordar que

Hössjer (1992) probó que la eficiencia de un S-estimador con punto de ruptura  $\epsilon^* = 0.5$  es a lo sumo del 35%. Otro S-estimador importante es el LMS-estimador, que ya se introdujo en el capítulo uno. Su sesgo para distintas proporciones de contaminación también se puede encontrar en la tabla 3.1. Combinando este estimador con la función bicuadrada de Tukey, obtenemos otro  $\tau$ -estimador (TAU-LMS-TUKEY) cuyo sesgo es menor que el de TAU-TUKEY.

Los sesgos de los estimadores TAU-TUKEY y TAU-LMS-TUKEY se han representado gráficamente en la figura 3.1



**Figura 3.1:** Curvas de sesgo de los estimadores TAU-TUKEY (línea continua) y TAU-LMS-TUKEY (línea discontinua)

### 3.3.3 Sesgo del LMS-estimador con regresores no elípticos

Supongamos que la distribución de los errores,  $F_0$ , es normal estándar. Consideremos un modelo de regresión lineal en el que hay dos regresores,  $X_1$  y  $X_2$ , y tomemos el LMS-estimador. En este apartado se obtienen algunos resultados en este caso cuando los regresores no son esféricos. Para ello es preciso resolver numéricamente la ecuación (3.3).

En la tabla 3.2 pueden encontrarse algunos valores para las curvas de sesgo de este estimador en los siguientes casos:

- (a) El vector  $(X_1, X_2)$  se distribuye como una normal bivalente con vector de medias  $\mathbf{0}$  y matriz de varianzas y covarianzas  $I$ . Con respecto a esta situación se pueden comparar el resto de casos considerados.
- (b) Para  $i = 1, 2$ ,  $X_i$  sigue una distribución  $t$  de Student estandarizada con tres grados de libertad y ambos regresores son independientes. Con este caso se pretende estudiar el efecto de colas pesadas en la distribución de los regresores.
- (c) Para  $i = 1, 2$ ,  $X_i$  se distribuye como una distribución  $\chi^2$  estandarizada con cuatro grados de libertad y ambos regresores son independientes. Este ejemplo permite estudiar el efecto que puede tener en el sesgo la presencia de regresores asimétricos.

En la tabla 3.2 se observa que en los casos (b) y (c) el sesgo es mayor que en el caso elíptico (a). Además los regresores asimétricos del caso (c) proporcionan unos sesgos menores que los regresores con colas pesadas del caso (b).

### 3.3.4 Estimadores basados en rangos absolutos

Nos ocupamos en este apartado de los estimadores de regresión basados en rangos absolutos. Estos estimadores se definieron en las expresiones (1.26) y (1.27).

$\epsilon$	0.05	0.1	0.15	0.2	0.25	0.3
LMS	0.52	0.82	1.13	1.51	2.01	2.73
STUKEY	0.55	0.87	1.23	1.65	2.17	3.02
0.5-LTS	0.63	1.02	1.45	2.02	2.85	4.19
0.5-LTAV	0.74	1.34	1.84	2.61	3.98	6.62
TAU-LMS-TUKEY	0.90	1.41	1.92	2.51	3.23	4.18
TAUTUKEY	0.92	1.45	2.00	2.65	3.39	4.49

**Tabla 3.1:** Sesgo máximo para algunos estimadores robustos de regresión

$\epsilon$	0.05	0.1	0.15	0.2	0.25	0.3
Caso (a)	0.52	0.82	1.13	1.51	2.01	2.73
Caso (b)	0.62	1.12	1.63	2.21	3.00	4.33
Caso (c)	0.52	0.92	1.23	1.71	2.31	3.23

**Tabla 3.2:** Curva de sesgo del LMS-estimador para distribuciones no elípticas de los regresores.

Supondremos que  $a(u)$  es una función definida en  $[0, 1]$  con valores reales sobre la que hacemos la siguiente hipótesis:

**Hipótesis 8** (a)  $a(u)$  es continua sobre  $[0, 1 - \alpha]$  con  $\alpha \geq 1/2$ .

(b)  $a(u) = 0$  si  $1 - \alpha < u \leq 1$ .

(c)  $a(u) > 0$  si  $0 < u < 1 - \alpha$ .

El siguiente lema garantiza que la curva de sesgo de R-estimadores de regresión se puede obtener aplicando la proposición 16.

**Lema 7** Si la función  $a(u)$  cumple la hipótesis 8, entonces el funcional  $R(F)$ , definido en la ecuación (1.27), cumple la hipótesis 5 para  $\eta < \alpha$  y la hipótesis 6 para  $\epsilon < \alpha$ .

**Demostración:**

Para la hipótesis 5, véase el lema A.4 de Yohai y Zamar (1993) para la parte (a) y el teorema 5.2 del mismo trabajo para la parte (b).

También el citado lema A.4(b) implica que si  $\epsilon < \alpha$ ,

$$\lim_{n \rightarrow \infty} R[(1 - \epsilon)F + \epsilon U_n] = R[(1 - \epsilon)F],$$

mientras que, por el lema A.4.(a),

$$R[(1 - \epsilon)F] \geq R[(1 - \epsilon)F + \epsilon G],$$

de donde se concluye que  $R(F)$  satisface la hipótesis 6.

□

En lo que sigue nos restringimos por sencillez a considerar el caso de la función:

$$a(u) = \begin{cases} 1, & |u| \leq 1 - \alpha \\ 0, & |u| > 1 - \alpha \end{cases}$$

Para calcular  $B_R(\epsilon)$  denotemos

$$F_{\theta^*} = (1 - \epsilon)F_{H_0, \theta^*} + \epsilon\delta_0,$$

y

$$F_n = (1 - \epsilon)F_{H_0, 0} + \epsilon U_n.$$

Sea  $\gamma = (1 + \|\theta^*\|^2)^{1/2}$  y sean  $\Phi$  y  $\varphi$  las funciones de distribución y de densidad de una normal estándar. Supongamos que  $\epsilon < \alpha$ . Sean

$$c_1 = \Phi^{-1} \left[ 1 - \frac{\alpha}{2(1 - \epsilon)} \right], \quad y \quad c_2 = \Phi^{-1} \left[ 1 - \frac{\alpha - \epsilon}{2(1 - \epsilon)} \right],$$

de manera que  $F_{\theta^*}(\gamma c_1) = 1 - \alpha$  y  $F_{H_0, 0}(c_2) = 1 - \alpha$ . Introduzcamos también la notación

$$I_k(c) = \int_0^c u^k \varphi(u) du.$$

Con estas definiciones y aplicando el corolario 4 se trata de resolver en  $\theta^*$  la ecuación

$$R(F_{\theta^*}) = \lim_{n \rightarrow \infty} R(F_n).$$

Calculemos primero  $R(F_{\theta^*})$ . Obsérvese que

$$F_{\theta^*}(u) = (1 - \epsilon)[2\Phi(u/\gamma) - 1] + \epsilon,$$

de donde se sigue que

$$R(F_{\theta^*}) = \int_0^{\gamma c_1} u^k dF_{\theta^*}(u) = 2\gamma^k(1 - \epsilon) \int_0^{c_1} u^k \varphi(u) du = 2\gamma^k(1 - \epsilon)I_k(c_1). \quad (3.21)$$

Por otra parte, el lema A.4(b) de Yohai y Zamar (1993) permite mostrar que

$$\lim_{n \rightarrow \infty} R(F_n) = (1 - \epsilon) \int_0^\infty a[(1 - \epsilon)F_{H_0,0}(u)]u^k dF_{H_0,0}(u).$$

Dado que estamos suponiendo normalidad,  $F_{H_0,0}(u) = 2\Phi(u) - 1$ , lo que lleva a

$$\lim_{n \rightarrow \infty} R(F_n) = 2(1 - \epsilon) \int_0^{c_2} u^k \varphi(u) du = 2(1 - \epsilon) I_k(c_2). \quad (3.22)$$

Igualando las expresiones (3.21) y (3.22), se tiene

$$[1 + B_R^2(\epsilon)]^{k/2} = \frac{I_k(c_2)}{I_k(c_1)} \quad (3.23)$$

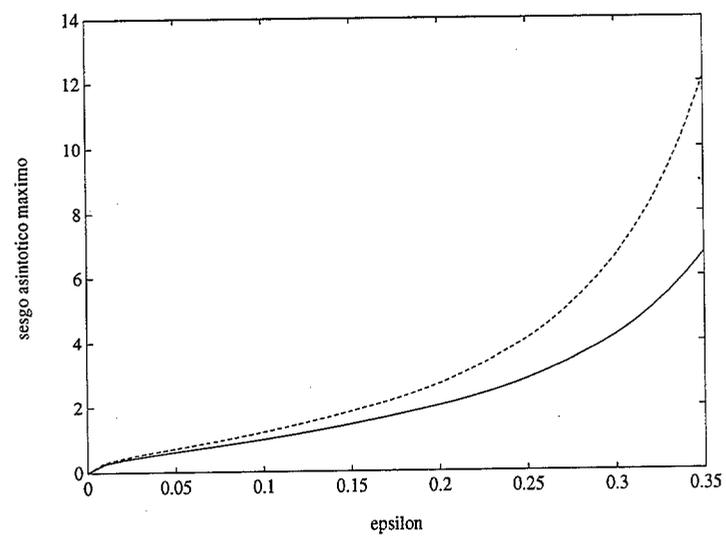
Particularizando al caso  $k = 1$ , tenemos que la curva de sesgo máximo del estimador  $\alpha$ -LTAV viene dada por la relación

$$[1 + B_R^2(\epsilon)]^{1/2} = \frac{I_1(c_2)}{I_1(c_1)} = \frac{\varphi(0) - \varphi(c_2)}{\varphi(0) - \varphi(c_1)}.$$

Para el caso  $k = 2$ , encontramos la siguiente expresión que debe cumplir el sesgo del estimador  $\alpha$ -LTS:

$$[1 + B_R^2(\epsilon)]^{1/2} = \frac{I_2(c_2)}{I_2(c_1)} = \frac{\Phi(c_2) - 0.5 - c_2\varphi(c_2)}{\Phi(c_1) - 0.5 - c_1\varphi(c_1)}.$$

En la tabla 3.1 se encuentran los valores del sesgo para estos dos casos particulares cuando  $\alpha = 0.5$ . Se observa que el estimador 0.5-LTS es más robusto que el 0.5-LTAV, especialmente para proporciones de contaminación grandes. En general, si consideramos la sucesión de estimadores generada cuando  $k$  recorre los números naturales, se comprueba numéricamente que, a medida que  $k$  crece, la curva de sesgo es menor, uniformemente en  $\epsilon$ . Este hecho no resulta sorprendente si tenemos en cuenta que cuando  $k \rightarrow \infty$ , estos estimadores se aproximan al LMS-estimador. Nótese también que para valores pequeños de  $\epsilon$ , las curvas de sesgo de los estimadores LTS y LTAV son mejores que las correspondientes a los  $\tau$ -estimadores. Sin embargo, se debe recordar



**Figura 3.2:** Curvas de sesgo de los estimadores LTS (línea continua) y LTAV (línea discontinua) para  $\alpha = 0.5$

que en general los  $R$ -estimadores con alto punto de ruptura no son eficientes (véase Hössjer, 1994).

La representación gráfica de las curvas de sesgo de los estimadores LMS y LTAV cuando  $\alpha = 0.5$  se encuentra en la figura 3.2.

### 3.4 Curva de sesgo para M-estimadores con escala general

Los estimadores de regresión con escala general son otro ejemplo de estimadores de residuos admisibles. Martin *et al.* (1989) los definieron como

$$\mathbf{T}(H) = \arg \min_{\boldsymbol{\theta}} E_H \rho \left( \frac{y - \boldsymbol{\theta}' \mathbf{x}}{S(H)} \right),$$

donde  $S(H)$  es un estimador robusto de la escala de los residuos y la función  $\rho$  verifica la hipótesis 1. Los MM-estimadores definidos por Yohai (1987) son un ejemplo de estimadores incluidos en esta clase que no pertenecen a ninguna de las clases de estimadores consideradas previamente.

Los M-estimadores con escala general no sólo dependen de la distribución de los residuos  $|r_i(\boldsymbol{\theta})|$  sino de la distribución de los residuos reescalados por el funcional  $S(H)$ . Por ello, no es posible aplicar los resultados de la sección 3.2 para obtener su curva de sesgo. En general, esta curva dependerá de las propiedades de robustez respecto al sesgo de  $S(H)$ .

En esta sección obtenemos una cota superior y otra inferior para la curva de sesgo de M-estimadores de regresión con escala general. Veremos que en algunos casos importantes y para valores pequeños de  $\epsilon$ , ambas cotas coinciden, por lo que es posible dar el valor exacto del sesgo en estos casos particulares.

Utilizaremos la siguiente notación: sea  $s_1 = \inf_{H \in \mathcal{V}_\epsilon} S(H)$ ,  $s_2 = \sup_{H \in \mathcal{V}_\epsilon} S(H)$ , y

$$h(t, s) = \inf_{\|\boldsymbol{\theta}\|=t} E_{H_0} \rho \left( \frac{y - \boldsymbol{\theta}' \mathbf{x}}{s} \right) - E_{H_0} \rho \left( \frac{y}{s} \right).$$

Las dos funciones siguientes van a jugar un papel importante en las cotas de la proposición 17.

$$h_1(t) = h(t, s_2), \text{ y } h_2(t) = \inf_{s_1 \leq s \leq s_2} h(t, s).$$

Como estimador de la escala, consideremos la escala estimada a partir de los residuos correspondientes a S-estimadores de regresión, es decir, supondremos la siguiente hipótesis:

**Hipótesis 9** Sea  $\chi$  una función que cumple la hipótesis 1 y  $S(H) = \min_{\boldsymbol{\theta}} S(F_{H, \boldsymbol{\theta}})$ , donde  $S(F_{H, \boldsymbol{\theta}})$  se define como en la ecuación (1.2) con  $b = b_\chi$ . Se verifica que la curva del correspondiente S-estimador de regresión,  $B_\chi(\epsilon)$ , es tal que  $B_\chi(\epsilon) < h_1^{-1}[\epsilon/(1 - \epsilon)]$ .

La hipótesis anterior se cumple siempre que se escoja una función  $\chi$  que proporcione estimadores robustos respecto al sesgo. Por ejemplo, el uso de funciones  $\chi$  de salto garantizará que la hipótesis se verifica.

**Proposición 17** Sea  $\mathbf{T}$  un M-estimador de regresión con escala general, donde  $\rho$  satisface la hipótesis 1 y  $S(H)$  cumple la hipótesis 9. Si además, las distribuciones  $F_0$  y  $G_0$  verifican la hipótesis 7, entonces

$$h_1^{-1} \left( \frac{\epsilon}{1 - \epsilon} \right) \leq B_{\mathbf{T}}(\epsilon) \leq h_2^{-1} \left( \frac{\epsilon}{1 - \epsilon} \right). \quad (3.24)$$

**Demostración:**

Por Yohai y Zamar (1993), lema 5.1, para todo  $s > 0$ , el funcional  $J(F) = E_F \rho(y/s)$  cumple la hipótesis 5(b). Por el lema 4, para todo  $s > 0$ ,  $h(t, s)$  es una función

estrictamente creciente en  $t$ . Además, también por el lema 4, para todo  $t > 0$ , existe  $\theta_t \in \mathbb{R}^p$  tal que

$$h_1(t) = E_{H_0} \rho \left( \frac{y - \theta_t' \mathbf{x}}{s_2} \right) - E_{H_0} \rho \left( \frac{y}{s_2} \right)$$

y  $h_1(t)$  es también estrictamente creciente.

Veamos ahora que  $B_{\mathbf{T}}(\epsilon) \leq t_2$ , donde  $t_2$  es tal que  $h_2(t_2) = \epsilon/(1 - \epsilon)$ . Sea  $\tilde{\theta} \in \mathbb{R}^p$  tal que  $\tilde{t} = \|\tilde{\theta}\| > t_2$ . Probaremos que

$$E_H \rho \left( \frac{y - \tilde{\theta}' \mathbf{x}}{S(H)} \right) > E_H \rho \left( \frac{y}{S(H)} \right), \text{ para todo } H \in V_\epsilon. \quad (3.25)$$

Sea  $H = (1 - \epsilon)H_0 + \epsilon\tilde{H}$ . Se verifican las siguientes desigualdades:

$$h[\tilde{t}, S(H)] > h[t_2, S(H)] \geq \inf_{s_1 \leq s \leq s_2} h(t_2, s) = h_2(t_2) = \frac{\epsilon}{1 - \epsilon}.$$

Por lo tanto, para todo  $H \in V_\epsilon$ ,

$$E_{H_0} \rho \left( \frac{y - \tilde{\theta}' \mathbf{x}}{S(H)} \right) - E_{H_0} \rho \left( \frac{y}{S(H)} \right) > \frac{\epsilon}{1 - \epsilon},$$

es decir,

$$(1 - \epsilon)E_{H_0} \rho \left( \frac{y - \tilde{\theta}' \mathbf{x}}{S(H)} \right) > (1 - \epsilon)E_{H_0} \rho \left( \frac{y}{S(H)} \right) + \epsilon.$$

Se sigue que, para todo  $H \in V_\epsilon$ ,

$$\begin{aligned} E_H \rho \left( \frac{y - \tilde{\theta}' \mathbf{x}}{S(H)} \right) &\geq (1 - \epsilon)E_{H_0} \rho \left( \frac{y - \tilde{\theta}' \mathbf{x}}{S(H)} \right) \\ &> (1 - \epsilon)E_{H_0} \rho \left( \frac{y}{S(H)} \right) + \epsilon \geq E_H \rho \left( \frac{y}{S(H)} \right), \end{aligned}$$

luego se verifica la desigualdad (3.25).

Ahora veamos que  $B_{\mathbf{T}}(\epsilon) \geq t_1$ , donde  $t_1$  es tal que  $h_1(t_1) = \epsilon/(1 - \epsilon)$ . Sea  $t > 0$  tal que  $B_X(\epsilon) < t < t_1$ , donde  $B_X(\epsilon) < t_1$  aplicando la hipótesis 9. Basta probar que  $B_{\mathbf{T}}(\epsilon) \geq t$ .

Existe  $\theta_t \in \mathbb{R}^p$  tal que

$$h_1(t) = h(t, s_2) = E_{H_0} \rho \left( \frac{y - \theta_t' \mathbf{x}}{s_2} \right) - E_{H_0} \rho \left( \frac{y}{s} \right).$$

Nótese que, como  $h_1$  es estrictamente creciente,  $h_1(t) < h_1(t_1)$ . Se deduce que

$$(1 - \epsilon) E_{H_0} \rho \left( \frac{y - \theta_t' \mathbf{x}}{s_2} \right) < (1 - \epsilon) E_{H_0} \rho \left( \frac{y}{s_2} \right) + \epsilon. \quad (3.26)$$

Se define ahora la siguiente sucesión de distribuciones contaminantes  $\tilde{H}_n = \delta_{(y_n, \mathbf{x}_n)}$ , donde  $\mathbf{x}_n = n\theta_t$  e  $y_n = \mathbf{x}_n' \theta_t = nt^2$ . Sea  $H_n = (1 - \epsilon)H_0 + \epsilon\tilde{H}_n$ . Supongamos que  $\sup_n \|\mathbf{T}(H_n)\| < t$  con el fin de llegar a una contradicción. Bajo esta suposición, existe una subsucesión convergente, denotada por  $\{\mathbf{T}(H_n)\}$ , tal que

$$\lim_{n \rightarrow \infty} \mathbf{T}(H_n) = \lim_{n \rightarrow \infty} \mathbf{T}_n = \tilde{\theta}, \text{ donde } \|\tilde{\theta}\| = \tilde{t} < t.$$

Tenemos que

$$\lim_{n \rightarrow \infty} \left| \frac{y_n - \mathbf{T}_n' \mathbf{x}_n}{S(H_n)} \right| = \infty, \text{ y } \left| \frac{y_n - \theta_t' \mathbf{x}_n}{S(H_n)} \right| = 0, \text{ para todo } n. \quad (3.27)$$

Probamos ahora que  $\lim_{n \rightarrow \infty} S(H_n) = s_2$ . Sabemos que

$$(1 - \epsilon) E_{H_0} \chi(y/s_2) + \epsilon = b_\chi.$$

Por lo tanto, si el S-estimador de regresión,  $T_\chi(H)$ , satisface  $\beta = \lim_{n \rightarrow \infty} T_\chi(H_n) = 0$ , entonces  $\lim_{n \rightarrow \infty} S(H_n) = s_2$ . Como  $\|\beta\| \leq B_\chi(\epsilon) < b$ , entonces  $\lim_{n \rightarrow \infty} |y_n - \beta' \mathbf{x}_n| = \infty$ . Si  $l = \lim_{n \rightarrow \infty} S(H_n)$ , obtenemos

$$(1 - \epsilon) E_{H_0} \chi \left( \frac{y - \beta' \mathbf{x}}{l} \right) + \epsilon = b_\chi. \quad (3.28)$$

Supongamos que  $\|\beta\| > 0$ . Entonces,

$$(1 - \epsilon) E_{H_0} \chi \left( \frac{y - \beta' \mathbf{x}}{s_2} \right) + \epsilon > (1 - \epsilon) E_{H_0} \chi(Y/s_2) + \epsilon = b_\chi. \quad (3.29)$$

Por (3.28) y (3.29),  $l > s_2$ . Esto contradice el hecho de que  $s_2 = \sup_{H \in V_\epsilon} S(H)$ . Por lo tanto  $\|\beta\| = 0$  y  $\lim_{n \rightarrow \infty} S(H_n) = s_2$ . Usaremos este hecho para obtener las ecuaciones (3.30) y (3.31).

Las ecuaciones (3.26) y (3.27) implican

$$\begin{aligned} \lim_{n \rightarrow \infty} E_{H_n} \rho \left( \frac{y - \mathbf{T}'_n \mathbf{x}}{S(H_n)} \right) &= (1 - \epsilon) E_{H_0} \rho \left( \frac{y - \tilde{\theta}' \mathbf{x}}{s_2} \right) + \epsilon \\ &\geq (1 - \epsilon) E_{H_0} \rho \left( \frac{y}{s_2} \right) + \epsilon > (1 - \epsilon) E_{H_0} \rho \left( \frac{y - \theta'_t \mathbf{x}}{s_2} \right) \end{aligned} \quad (3.30)$$

Por otra parte, aplicando (3.27),

$$\lim_{n \rightarrow \infty} E_{H_n} \rho \left( \frac{y - \theta'_t \mathbf{x}}{S(H_n)} \right) = (1 - \epsilon) E_{H_0} \rho \left( \frac{y - \theta'_t \mathbf{x}}{s_2} \right). \quad (3.31)$$

Por lo tanto, para  $n$  suficientemente grande,

$$E_{H_n} \rho \left( \frac{y - \mathbf{T}'_n \mathbf{x}}{S(H_n)} \right) > E_{H_n} \rho \left( \frac{y - \theta'_t \mathbf{x}}{S(H_n)} \right).$$

Esta última desigualdad es una contradicción ya que

$$\mathbf{T}_n = \arg \min_{\theta} E_{H_n} \rho \left( \frac{y - \theta' \mathbf{x}}{S(H_n)} \right).$$

Para todo  $t > 0$  tal que  $B_X(\epsilon) < t < t_1$ , hemos encontrado una sucesión de distribuciones  $\{H_n\}$  en el entorno de contaminación  $V_\epsilon$  tal que  $\sup_n \|\mathbf{T}(H_n)\| \geq t$ . Se sigue que  $B_{\mathbf{T}}(\epsilon) \geq t_1$ .

□

El valor exacto de la curva de sesgo puede hallarse si conseguimos probar que, para algunos valores de  $\epsilon$ , las cotas obtenidas en la proposición 17 coinciden. Vamos a dar a continuación una condición suficiente para que esto ocurra cuando la distribución de

los regresores es normal. Como veremos, se puede comprobar numéricamente que esta condición se cumple en algunos casos importantes. Este hecho probará que las cotas encontradas en la proposición 17 no se pueden mejorar.

Nótese también que no es difícil obtener numéricamente el valor de  $s_2$  y lo mismo sucede con la función  $h_1$ . Sin embargo, el cálculo de la función  $h_2$  puede involucrar un problema de optimización difícil de resolver en general. Si logramos probar que en algunos casos las dos cotas coinciden, sólo necesitaremos calcular la función  $h_1$ .

Cuando los regresores son esféricos podemos eliminar el ínfimo en la definición de la función  $h(t, s)$  y, por lo tanto, tiene sentido escribir  $h(\theta, s)$ ,  $h_1(\theta)$  y  $h_2(\theta)$ . Sea  $\theta^* = h_1^{-1}[\epsilon/(1 - \epsilon)]$ . Si la función  $h(\theta^*, s)$  es decreciente para todo  $s_1 < s < s_2$ , entonces  $h_1(\theta^*) = h_2(\theta^*)$ . En este caso, la cota inferior y la superior tomarían el mismo valor. Por lo tanto, una condición suficiente para que esto ocurra es

$$\frac{\partial h(\theta^*, s)}{\partial s} < 0, \quad \text{para todo } s \in (s_1, s_2). \quad (3.32)$$

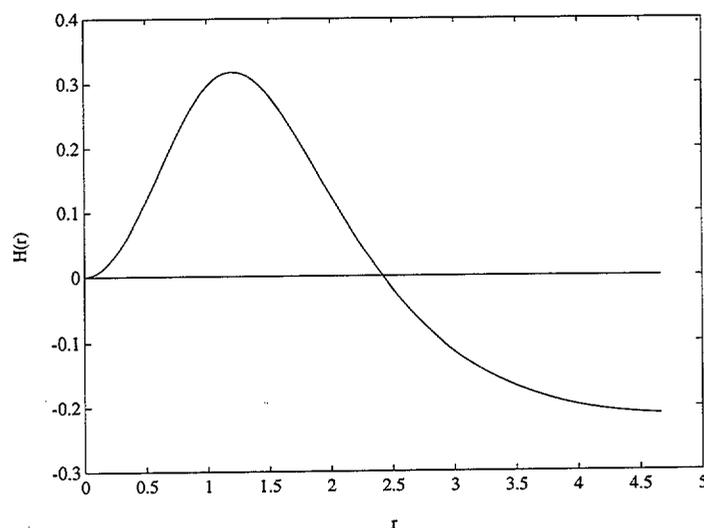
Si suponemos que los regresores son normales, denotamos  $\psi = \rho'$  y  $\gamma = (1 + \|\theta^*\|^2)^{1/2}$ , y además suponemos que es posible derivar bajo el signo integral, la condición (3.32) es equivalente a

$$G(\gamma, s) = E_{\Phi}[(\gamma Z/s)\psi(\gamma Z/s)] - E_{\Phi}[(Z/s)\psi(Z/s)] > 0, \quad \text{para todo } s \in (s_1, s_2) \quad (3.33)$$

Como  $\gamma \geq 1$  y  $G(1, s) = 0$ , una condición suficiente para (3.33) es que  $G(r, s)$  sea creciente en  $r$  para  $s \in (s_1, s_2)$  y  $1 \leq r \leq \gamma$ . Por lo tanto, una nueva condición suficiente es

$$\begin{aligned} H(r/s) &= r \frac{\partial G(r, s)}{\partial r} \\ &= E_{\Phi}[(r/s)Z\psi(rZ/s)] + E_{\Phi}[(r^2 Z^2/s^2)\psi'(rZ/s)] \\ &> 0, \quad \text{para todo } s \in (s_1, s_2), \quad 1 \leq r \leq \gamma. \end{aligned} \quad (3.34)$$

En la figura 3.3, se ha representado la función  $H(r/s)$  correspondiente al caso en que  $\rho$  es la función bicuadrada de Tukey. Podemos comprobar que hay un rango de valores de  $r/s$  para el que la función es positiva. Para el valor  $r/s = 2.45$ , se tiene que  $H(r/s) = 0$  y, por lo tanto,  $H(r/s) > 0$  siempre que  $r/s < 2.45$ .



**Figura 3.3:** Gráfico de la función  $H(r)$  cuando  $\rho$  es la función bicuadrada de Tukey.

Para verificar la condición (3.34), se calculan  $\theta^*$  y  $s_1$  para una proporción de contaminación fija  $\epsilon$ . Como para todo  $s \in (s_1, s_2)$  se cumple  $\gamma/s_1 > r/s$ , entonces  $H(\gamma/s_1) > 0$  implica  $H(r/s) > 0$  para todo  $s \in (s_1, s_2)$  y  $1 \leq r \leq \gamma$ . Por lo tanto, la condición suficiente (3.34) se satisface.

En el ejemplo de la función bicuadrada de Tukey (donde la constante  $c$  se ajusta para obtener una eficiencia del 95%), los resultados obtenidos se presentan en la tabla 3.3. Hemos considerado dos funcionales de escala diferentes: el S-estimador basado en una función de salto y el basado en una función bicuadrada de Tukey. Ambos han sido ajus-

tados para obtener un punto de ruptura del 50%. Llamemos  $\mathbf{T}_1$  y  $\mathbf{T}_2$  respectivamente a los MM-estimadores resultantes. Tal y como se ha descrito, para  $\epsilon < .15$ , se puede comprobar numéricamente que la condición (3.34) se cumple para ambos estimadores. Las dos cotas de la proposición 17 coinciden en los dos casos.

$\epsilon$	.01	.03	.05	.07	.09	0.11	0.13	0.15
$B_{\mathbf{T}_1}(\epsilon)$	0.31	0.57	0.77	0.96	1.15	1.34	1.55	1.77
$B_{\mathbf{T}_2}(\epsilon)$	0.31	0.57	0.77	0.95	1.14	1.32	1.52	1.73

Tabla 3.3: Valores que toma la curva de sesgo para los MM-estimadores  $\mathbf{T}_1$  y  $\mathbf{T}_2$ .

### 3.5 Una aplicación al análisis robusto de componentes principales

El análisis de componentes principales (ACP) tiene como objetivo buscar combinaciones lineales de las variables observadas que sean útiles para resumir los datos, reduciendo su dimensión, y perdiendo la menor cantidad de información posible en el proceso. El ACP clásico alcanza este objetivo encontrando las combinaciones lineales de las variables originales que tienen máxima varianza. Se toma así la varianza como criterio para medir la información contenida en los datos.

En esta sección se estudia una adaptación de los resultados obtenidos anteriormente en este capítulo que permite estudiar el sesgo en dirección de algunos métodos de ACP robustos. Comenzamos con una revisión breve de las distintas formas de realizar ACP robusto.

Si  $\mathbf{x} \in \mathbb{R}^p$  es un vector aleatorio, el ACP clásico se basa en resolver la siguiente sucesión de problemas:

$$\begin{aligned}
 \mathbf{v}_1 &= \arg_{\mathbf{a}} \min \text{Var}(\mathbf{a}'\mathbf{x}) \quad \text{s.a.} \quad \|\mathbf{a}\| = 1, \\
 \mathbf{v}_2 &= \arg_{\mathbf{a}} \min \text{Var}(\mathbf{a}'\mathbf{x}) \quad \text{s.a.} \quad \|\mathbf{a}\| = 1, \mathbf{a} \perp \mathbf{v}_1, \\
 &\dots\dots\dots \\
 \mathbf{v}_p &= \arg_{\mathbf{a}} \min \text{Var}(\mathbf{a}'\mathbf{x}) \quad \text{s.a.} \quad \|\mathbf{a}\| = 1, \mathbf{a} \perp \mathbf{v}_1, \dots, \mathbf{v}_{p-1}.
 \end{aligned} \tag{3.35}$$

Si  $\mathbf{x}$  se distribuye como una normal  $p$ -dimensional con media  $\boldsymbol{\mu}$  y matriz de varianzas covarianzas  $V$ , los vectores  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p$ , solución de (3.35), son los autovectores con norma igual a uno de  $V$  correspondientes a los autovalores  $\lambda_1 < \lambda_2 < \dots < \lambda_p$ . Las componentes principales clásicas se obtienen a partir de los datos sustituyendo la distribución poblacional por la empírica correspondiente a la muestra, lo que lleva a calcular los autovalores y autovectores de la matriz de varianzas y covarianzas muestrales.

Ante la presencia de datos atípicos o *outliers*, la inestabilidad o falta de robustez de la varianza se traduce en inestabilidad y falta de robustez del ACP clásico. Se han propuesto diferentes alternativas para obtener componentes principales más robustas. Un primer grupo de estrategias se basan en estimar la matriz de varianzas y covarianzas  $V$  de forma robusta en lugar de hacerlo mediante la matriz de varianzas y covarianzas muestral. Es decir, se plantea el mismo problema clásico (3.35) pero se estima la solución del problema de forma robusta. Algunos resultados incluidos en este enfoque pueden encontrarse en Maronna (1976), Devlin *et al.* (1981), Huber (1981), Rousseeuw (1985) y Maronna *et al.* (1992).

El segundo grupo de alternativas para obtener componentes principales robustas se basa en transformar el problema (3.35). En lugar de resolver (3.35) se sustituye la varianza por un estimador de dispersión más robusto con el fin de que la estabilidad de

la estimación de la dispersión se traslade también a la estimación de las componentes principales. Este método fue propuesto por Li y Chen (1985) quienes se basaron en ideas procedentes de las técnicas *projection pursuit*.

Li y Chen (1985) estudian algunas de las propiedades básicas del método, así como su robustez con respecto al tamaño de las componentes. En esta sección, retomamos el estudio de estas propiedades teóricas, analizando la robustez del método respecto a la dirección de las componentes. Con el fin de simplificar, supondremos que la dimensión del problema es  $p = 2$ .

### 3.5.1 Notación y resultados básicos

Supongamos que se observa una muestra  $\mathbf{x}_1, \dots, \mathbf{x}_n$  procedente de una distribución bidimensional  $F$  contenida en el entorno de contaminación

$$V_\epsilon = \{F : F = (1 - \epsilon)F_0 + \epsilon H, H \text{ distribución arbitraria}\},$$

donde  $F_0$  es la distribución normal bivalente con vector de medias  $\mathbf{0}$  y matriz de varianzas y covarianzas  $V$ . Supongamos también que  $V$  tiene 2 autovalores distintos  $\lambda_1 < \lambda_2$  y denotemos por  $\mathbf{v}_1$  y  $\mathbf{v}_2$  los autovectores de norma igual a uno correspondientes a estos autovalores.

Sea  $\mathbf{a} \in \mathbb{R}^2$  con  $\|\mathbf{a}\| = 1$ . La distribución de  $|\mathbf{a}'\mathbf{x}|$  cuando  $\mathbf{x}$  se distribuye según  $F$  se denotará mediante  $F^{\mathbf{a}}$ .

Sea  $J(F)$  un funcional que define un estimador de escala robusto. El método de Li y Chen (1985) para obtener ACP robusto se basa en los siguientes funcionales, que generalizan el problema planteado en (3.35).

$$A_1(F) = \arg_{\mathbf{a}} \min J(F^{\mathbf{a}}) \quad \text{s.a.} \quad \|\mathbf{a}\| = 1, \quad (3.36)$$

$$A_2(F) \perp A_1(F), \quad \|A_2(F)\| = 1. \quad (3.37)$$

Un ejemplo especialmente importante resulta cuando  $J(F) = S(F)$  es un M-estimador de escala.

Con el fin de estudiar la robustez en dirección de los funcionales  $A_1(F)$  y  $A_2(F)$  definimos la curva de sesgo asintótico máximo siguiendo a Zamar (1989, 1992). Fijada una distribución  $F \in V_\epsilon$ , el sesgo asintótico en dirección en  $F$  viene determinado por el coseno del ángulo que forma el valor del funcional en  $F$  con la componente principal poblacional que queremos estimar. Es decir,

$$b_i(\epsilon, F) = 1 - |A'_i(F)\mathbf{v}_i|, \quad i = 1, 2.$$

La curva de sesgo asintótico máximo se define ahora como el mayor sesgo posible dentro del entorno de contaminación  $V_\epsilon$

$$B_i(\epsilon) = \sup_{F \in V_\epsilon} b_i(\epsilon, F) = 1 - \inf_{F \in V_\epsilon} |A'_i(F)\mathbf{v}_i|, \quad i = 1, 2. \quad (3.38)$$

En relación con el concepto de curva de sesgo asintótico máximo puede también definirse el punto de ruptura

$$\epsilon_i^* = \sup\{\epsilon : B_i(\epsilon) < 1\}, \quad i = 1, 2. \quad (3.39)$$

Una proporción de datos atípicos mayor que  $\epsilon_i^*$  nos puede conducir a una dirección perpendicular a la que queremos estimar con lo que el estimador no proporciona información útil.

Nótese que podemos prescindir del valor absoluto que aparece en la definición de la curva de sesgo ya que si  $\mathbf{b} = -\mathbf{a}$ , entonces  $F^{\mathbf{a}} = F^{\mathbf{b}}$ . También obsérvese que, como  $A_1(F) \perp A_2(F)$  y  $\mathbf{v}_1 \perp \mathbf{v}_2$ , se tiene que  $B_1(\epsilon) = B_2(\epsilon)$ , por lo que será suficiente estudiar la curva de sesgo correspondiente al funcional  $A_1(F)$ , que denotaremos  $B(\epsilon)$ , prescindiendo del subíndice.

Los dos resultados siguientes se refieren a sendas propiedades básicas del metodo que estamos analizando. El primero pone de manifiesto la falta de robustez del ACP clásico en relación a los conceptos que acabamos de definir.

**Lema 8** *Consideremos el funcional  $A_1$  definido en (3.36), donde  $J(F^{\mathbf{a}}) = \text{Var}_F(\mathbf{a}'\mathbf{x})$  (ACP clásico). Se verifica que  $B(\epsilon) = 1$  para todo  $\epsilon > 0$ . Por tanto,  $\epsilon^* = 0$ .*

**Demostración:**

Sea  $\epsilon > 0$  y  $\mathbf{x}_n = \alpha_n \mathbf{v}_1$ , donde  $\{\alpha_n\}$  es una sucesión tal que  $\alpha_n \rightarrow \infty$ . Sea  $F_n = (1 - \epsilon)F_0 + \epsilon\delta_{\mathbf{x}_n}$ , donde  $\delta_{\mathbf{x}_n}$  es la distribución que asigna probabilidad uno al punto  $\mathbf{x}_n$ . Es fácil comprobar que la matriz de varianzas y covarianzas correspondiente a la distribución  $F_n$  es

$$\text{Var}_{F_n}(\mathbf{x}) = E_{F_n}[\mathbf{x}\mathbf{x}'] - \epsilon^2 \mathbf{x}_n \mathbf{x}_n' = (1 - \epsilon)[V + \epsilon \mathbf{x}_n \mathbf{x}_n'], \quad (3.40)$$

donde  $V$  es la matriz de varianzas y covarianzas de  $F_0$ .

Se verifica, aplicando (3.40), que

$$\text{Var}_{F_n}(\mathbf{v}_2'\mathbf{x}) = (1 - \epsilon)(\mathbf{v}_2'V\mathbf{v}_2).$$

Por otra parte, sea  $\mathbf{b} \in \mathbb{R}^p$  tal que  $\mathbf{b}'\mathbf{v}_1 \neq 0$ , entonces

$$\text{Var}_{F_n}(\mathbf{b}'\mathbf{x}) = (1 - \epsilon)[\mathbf{b}'V\mathbf{b} + \epsilon(\alpha_n \mathbf{b}'\mathbf{v}_1)^2].$$

Se sigue que

$$\begin{aligned} \text{Var}_{F_n}(\mathbf{b}'\mathbf{x}) - \text{Var}_{F_n}(\mathbf{v}_2'\mathbf{x}) &= (1 - \epsilon)(\mathbf{b}'V\mathbf{b} - \mathbf{v}_2'V\mathbf{v}_2) + \epsilon(1 - \epsilon)(\alpha_n \mathbf{b}'\mathbf{v}_1)^2 \\ &> (1 - \epsilon)(\lambda_1 - \lambda_2) + \epsilon(1 - \epsilon)(\alpha_n \mathbf{b}'\mathbf{v}_1)^2 > 0, \end{aligned}$$

para  $n$  suficientemente grande. Por lo tanto, en este caso,  $A_1(F_n) = \mathbf{v}_2$ . Para  $n$  suficientemente grande,

$$\inf_{F \in V_\epsilon} |A_1(F)' \mathbf{v}_1| = 0, \text{ para todo } \epsilon > 0,$$

y por lo tanto  $B(\epsilon) = 1$ , para todo  $\epsilon > 0$ .

□

El siguiente resultado básico se refiere a la consistencia de Fisher. Lo que se prueba en el lema 9 es que cuando los funcionales se aplican a la distribución libre de contaminaciones, obtenemos las componentes principales nominales. Dicho de otra forma,  $B(0) = 0$ .

**Lema 9** *Supongamos que el funcional  $J(F^{\mathbf{a}})$  es equivariante ante cambios de escala. Bajo esta condición, se verifica*

$$A_i(F_0) = \mathbf{v}_i, \quad i = 1, 2.$$

**Demostración:**

Si  $\mathbf{x}$  se distribuye según  $F_0$ , entonces  $\mathbf{a}'\mathbf{x}$  se distribuye según una normal univariante de media 0 y desviación típica  $(\mathbf{a}'V\mathbf{a})^{1/2}$ . Dado que  $J$  es equivariante ante cambios de escala,  $J(F_0^{\mathbf{a}}) = (\mathbf{a}'V\mathbf{a})^{1/2}J(\Phi)$ , donde  $\Phi$  denota la función de distribución de la normal estándar. Como consecuencia, el problema de la ecuación (3.36) que define el funcional  $A_1(F)$  se transforma, cuando  $F = F_0$ , en el problema que define al ACP clásico, cuya solución es  $\mathbf{v}_1$ .

□

### 3.5.2 Cálculo de la curva de sesgo asintótico máximo

En este apartado obtenemos un método general para el cálculo de  $B(\epsilon)$  análogo al que se dio para el caso de estimadores de regresión con residuos admisibles. Para ello, suponemos que el funcional  $J$  verifica las hipótesis 5 y 6 que aseguran la admisibilidad de los residuos.

La siguiente notación será útil

$$\begin{aligned} \mathbf{a}(\alpha) &\equiv (1 - \alpha)\mathbf{v}_1 + [1 - (1 - \alpha)^2]^{1/2}\mathbf{v}_2, \quad \alpha \in (0, 1), \\ h(\alpha) &\equiv (1 - \alpha)^2\lambda_1 + [1 - (1 - \alpha)^2]\lambda_2, \quad \alpha \in (0, 1). \end{aligned}$$

**Proposición 18** *Supongamos que  $J$  verifica la hipótesis 5, con  $\eta = \epsilon$ , y la hipótesis 6. Sea  $\alpha_0 \in (0, 1)$  tal que*

$$J[(1 - \epsilon)F_0^{\mathbf{a}(\alpha_0)} + \epsilon\delta_0] = \lim_{n \rightarrow \infty} J[(1 - \epsilon)F_0^{\mathbf{V}_1} + \epsilon U_n], \quad (3.41)$$

donde  $U_n$  denota la distribución uniforme sobre el intervalo  $[n, 2n]$ . Entonces,  $B(\epsilon) = \alpha_0$ .

Si para todo  $\alpha_0 \in (0, 1)$ ,

$$J[(1 - \epsilon)F_0^{\mathbf{a}(\alpha_0)} + \epsilon\delta_0] < \lim_{n \rightarrow \infty} J[(1 - \epsilon)F_0^{\mathbf{V}_1} + \epsilon U_n], \quad (3.42)$$

entonces  $B(\epsilon) = 1$ .

#### Demostración:

Sea  $\alpha_0$  tal que verifica (3.41). Veamos primero que  $B_1(\epsilon) \leq \alpha_0$ . Basta establecer que para todo  $F \in V_\epsilon$ ,  $A'_1(F)\mathbf{v}_1 \geq 1 - \alpha_0$  ó, equivalentemente, que  $A'_1(F)\mathbf{v}_1 \geq 1 - \alpha$ , para todo  $\alpha > \alpha_0$  y para cualquier  $F \in V_\epsilon$ .

Sea  $\alpha > \alpha_0$  y  $\mathbf{a} = \mathbf{a}(\alpha) \in \mathbb{R}^2$ . Es inmediato que, para todo  $u \geq 0$  y  $F \in V_\epsilon$ ,

$$F^{\mathbf{a}}(u) = (1 - \epsilon)F_0^{\mathbf{a}}(u) + \epsilon H^{\mathbf{a}}(u) \leq (1 - \epsilon)F_0^{\mathbf{a}}(u) + \epsilon \delta_0(u).$$

Ahora, la hipótesis 5 implica que, para cualquier  $F \in V_\epsilon$ ,

$$J(F^{\mathbf{a}}) \geq J[(1 - \epsilon)F_0^{\mathbf{a}} + \epsilon \delta_0]. \quad (3.43)$$

Además, para todo  $u \geq 0$ ,

$$\begin{aligned} F_0^{\mathbf{a}}(u) &= 2\Phi\left(\frac{u}{h(\alpha)}\right) - 1 \\ &< 2\Phi\left(\frac{u}{h(\alpha_0)}\right) - 1 = F_0^{\mathbf{a}(\alpha_0)}, \end{aligned}$$

puesto que la función  $h(\alpha)$  es estrictamente creciente. La condición de  $\epsilon$ -monotonía estricta implica ahora que

$$J[(1 - \epsilon)F_0^{\mathbf{a}(\alpha)} + \epsilon \delta_0] > J[(1 - \epsilon)F_0^{\mathbf{a}(\alpha_0)} + \epsilon \delta_0]. \quad (3.44)$$

Para todo  $F \in V_\epsilon$ , por (3.41), (3.43), (3.44) y por la hipótesis 6,

$$\begin{aligned} J(F^{\mathbf{a}}) &> J[(1 - \epsilon)F_0^{\mathbf{a}(\alpha_0)} + \epsilon \delta_0] \\ &= \lim_{n \rightarrow \infty} J[(1 - \epsilon)F_0^{\mathbf{v}_1} + \epsilon U_n] \geq J(F^{\mathbf{v}_1}). \end{aligned}$$

Por lo tanto,  $\mathbf{a} \neq A_1(F)$  para todo  $F \in V_\epsilon$  lo que implica que  $A_1'(F)\mathbf{v}_1 > 1 - \alpha$  para todo  $F \in V_\epsilon$ .

A continuación probaremos que  $B(\epsilon) \geq \alpha_0$ . Basta establecer que, para todo  $\alpha < \alpha_0$ , existe  $F \in V_\epsilon$  tal que  $A_1'(F)\mathbf{v}_1 \leq 1 - \alpha$ .

Sea  $\mathbf{a}^\perp(\alpha)$  ortogonal a  $\mathbf{a}(\alpha)$ , es decir

$$\mathbf{a}^\perp(\alpha) = [1 - (1 - \alpha)^2]^{1/2}\mathbf{v}_1 - (1 - \alpha)^2\mathbf{v}_2.$$

Definamos la siguiente sucesión de contaminaciones

$$H_n \equiv U \left[ \left( n - \frac{1}{n} \right) \mathbf{a}^\perp(\alpha), \left( n + \frac{1}{n} \right) \mathbf{a}^\perp(\alpha) \right], \quad (3.45)$$

donde  $U[a, b]$  denota la distribución uniforme en el intervalo  $(a, b)$ .

Sea  $F_n = (1 - \epsilon)F_0 + \epsilon H_n$ . Denotemos  $\mathbf{a}_n = A_1(F_n)$ . Supongamos, por reducción al absurdo, que  $\inf_n \mathbf{a}'_n \mathbf{v}_1 > 1 - \alpha$ . Debe existir una subsucesión  $\{\mathbf{a}_n\}$  tal que  $\lim_{n \rightarrow \infty} \mathbf{a}_n = \tilde{\mathbf{a}} = \mathbf{a}(\tilde{\alpha})$  con  $\tilde{\mathbf{a}}' \mathbf{v}_1 = 1 - \tilde{\alpha} > 1 - \alpha$ .

Nótese que

$$\lim_{n \rightarrow \infty} \mathbf{a}^\perp(\alpha)' \mathbf{a} = [1 - (1 - \alpha)^2]^{1/2} (1 - \tilde{\alpha})^2 - (1 - \alpha)^2 [1 - (1 - \tilde{\alpha})^2]^{1/2} > 0.$$

Como consecuencia, a partir de la ecuación (3.45) en que se definen las contaminaciones se deduce que, para todo  $u > 0$ ,

$$\lim_{n \rightarrow \infty} H_n^{\tilde{\mathbf{a}}_n}(u) = 0.$$

Por lo tanto,

$$\begin{aligned} \lim_{n \rightarrow \infty} F_n^{\mathbf{a}_n}(u) &= (1 - \epsilon) F_0^{\mathbf{a}(\tilde{\alpha})}(u) \\ &\leq (1 - \epsilon) F_0^{\mathbf{v}_1}(u) = \lim_{n \rightarrow \infty} [(1 - \epsilon) F_0^{\mathbf{v}_1}(u) + \epsilon U_n(u)]. \end{aligned}$$

Aplicando ahora la hipótesis de  $\epsilon$ -monotonía y la condición (3.41) del enunciado,

$$\lim_{n \rightarrow \infty} J(F_n^{\mathbf{a}_n}) \geq J[(1 - \epsilon) F_0^{\mathbf{a}(\alpha_0)} + \epsilon \delta_0]. \quad (3.46)$$

Dado que  $\mathbf{a}(\alpha)' \mathbf{a}^\perp(\alpha) = 0$ , se deduce que  $\lim_{n \rightarrow \infty} H_n^{\mathbf{a}(\alpha)}(u) = 1$  para todo  $u \geq 0$ .

Como consecuencia de este hecho y dado que  $\alpha < \alpha_0$ ,

$$\lim_{n \rightarrow \infty} F_n^{\mathbf{a}(\alpha)}(u) = (1 - \epsilon) F_0^{\mathbf{a}(\alpha)}(u) + \epsilon \delta_0(u) > (1 - \epsilon) F_0^{\mathbf{a}(\alpha_0)}(u) + \epsilon \delta_0(u).$$

La condición de  $\epsilon$ -monotonía estricta implica ahora

$$\lim_{n \rightarrow \infty} J(F_n^{\mathbf{a}(\alpha)}) < J[(1 - \epsilon)F_0^{\mathbf{a}(\alpha_0)} + \epsilon\delta_0]. \quad (3.47)$$

Aplicando (3.46) y (3.47),

$$\lim_{n \rightarrow \infty} J(F_n^{\mathbf{a}_n}) > \lim_{n \rightarrow \infty} J(F_n^{\mathbf{a}(\alpha)}),$$

lo que es una contradicción ya que  $\mathbf{a}_n = \arg \min J(F_n^{\mathbf{a}})$ .

Hemos probado que  $\inf_n A'_1(F_n)\mathbf{v}_1 \leq 1 - \alpha$ , para cualquier  $\alpha < \alpha_0$ , es decir,

$$\inf_n A'_1(F_n)\mathbf{v}_1 \leq 1 - \alpha_0.$$

Se sigue que  $B(\epsilon) \geq \alpha_0$ .

Finalmente, supongamos que se verifica (3.42) y sea  $\alpha \in (0, 1)$ . Definimos la misma sucesión de contaminaciones  $H_n$ , tal y como se hizo en la ecuación (3.45). Una prueba totalmente análoga a la que acabamos de hacer da lugar a la conclusión  $B(\epsilon) = 1$  en este caso. □

### 3.5.3 Ejemplo: M-estimadores de escala

Para ilustrar cómo obtener sesgos en dirección a partir de la proposición 18 consideramos el caso en que  $J(F) = S(F)$  es un M-estimador de escala.

Introducimos la función  $g(s) = E_{\Phi}\chi(Y/s)$ . Sea  $S_1 = S[(1 - \epsilon)F_0^{\mathbf{a}(\alpha_0)} + \epsilon\delta_0]$ . El valor  $S_1$  debe verificar que,

$$(1 - \epsilon)g[S_1/h(\alpha_0)] + \epsilon = b.$$

Por tanto,

$$S_1 = h(\alpha_0)g^{-1} \left( \frac{b}{1-\epsilon} \right). \quad (3.48)$$

Por otra parte,  $S_2 = \lim_{n \rightarrow \infty} S[(1-\epsilon)F_0^{V_1} + \epsilon U_n]$  verifica

$$(1-\epsilon)g[S_2/\lambda_1^{1/2}] + \epsilon = b,$$

de donde se deduce que

$$S_2 = \lambda_1^{1/2} g^{-1} \left( \frac{b-\epsilon}{1-\epsilon} \right). \quad (3.49)$$

En aplicación de la fórmula (3.41), imponemos  $S_1 = S_2$ . De esta igualdad y teniendo en cuenta las expresiones (3.48) y (3.49), despejamos el valor de  $\alpha_0$ . Tras algunos cálculos obtenemos:

$$B(\epsilon) = 1 - \left( \frac{r - Q(\epsilon)}{r - 1} \right), \text{ para } r > Q(\epsilon), \quad (3.50)$$

$$B(\epsilon) = 1, \text{ para } r \leq Q(\epsilon), \quad (3.51)$$

donde  $r = \lambda_2/\lambda_1$  y

$$Q(\epsilon) = \left[ \frac{g^{-1} \left( \frac{b-\epsilon}{1-\epsilon} \right)}{g^{-1} \left( \frac{b}{1-\epsilon} \right)} \right]^2$$

es el cociente al cuadrado entre los sesgos respecto a *outliers* e *inliers* del M-estimador de escala basado en  $\chi$ .

Las expresiones (3.50) y (3.51) permiten deducir que el punto de ruptura es:

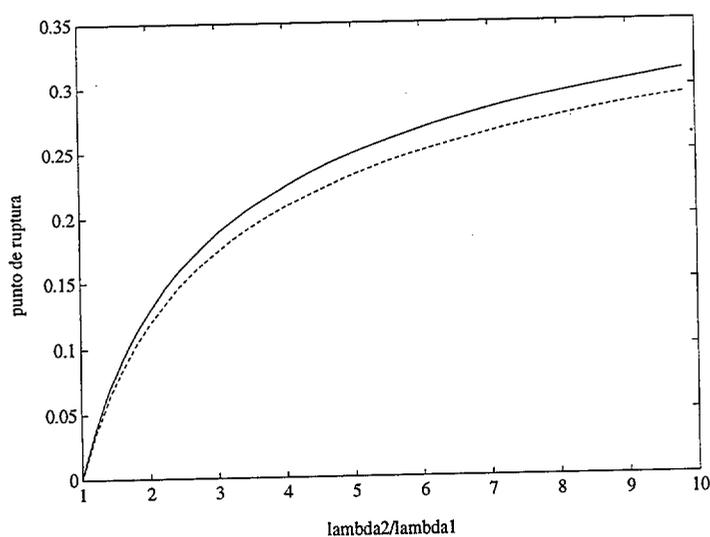
$$\epsilon^* = Q^{-1}(r). \quad (3.52)$$

Como se observa, el punto de ruptura es el resultado de dos factores:

- (a) El cociente  $r$  entre el segundo autovalor y el primero. Cuanto mayor es la diferencia en magnitud entre las componentes principales  $v_1$  y  $v_2$ , más fácil resulta distinguir entre ellas por lo que el estimador resulta más robusto independientemente de la elección de  $\chi$ .

- (b) El cociente entre los sesgos respecto a *outliers* e *inliers* del estimador de escala,  $Q(\epsilon)$ . Cuanto más robusto sea el M-estimador de escala, más robusto es el estimador de componentes principales basado en él.

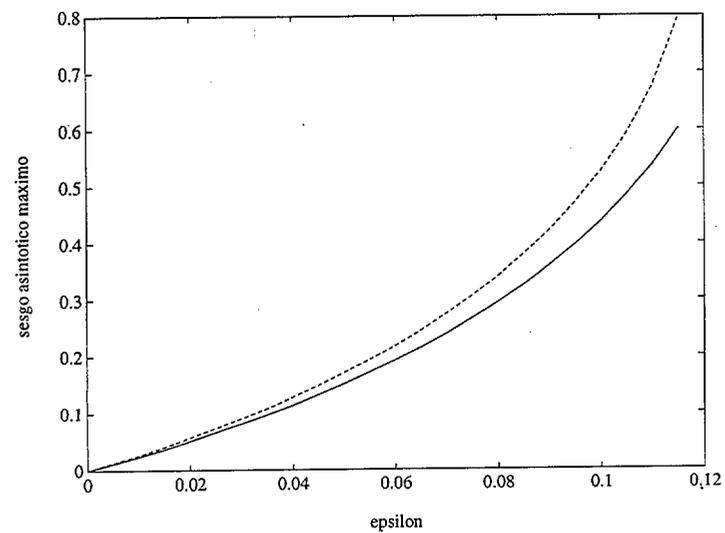
Como hemos visto, un M-estimador de escala importante desde el punto de vista teórico viene dado por la elección  $b = 1/2$  y  $\chi(y)$  una función de salto. Otra función  $\chi$  que se ha utilizado en ejemplos anteriores es la función bicuadrada de Tukey. Los valores de los puntos de ruptura de estos funcionales como función de  $r$  se hallan representados en la figura 3.4. Se observa que a pesar de haber tomado M-estimadores de escala muy robustos, el punto de ruptura no es muy alto para valores moderados de  $r$ .



**Figura 3.4:** Punto de ruptura de  $A_1(F)$  como función de  $r$  cuando  $\chi$  es una función de salto (línea continua) o cuando es la función bicuadrada de Tukey (línea discontinua).

Las respectivas curvas de sesgo asintótico máximo cuando  $r = 2$  se han representado

en la figura 3.5.



**Figura 3.5:** Curvas de sesgo asimótico máximo de  $A_1(F)$  cuando  $\chi$  es una función de salto (línea continua) o cuando es la función bicuadrada de Tukey (línea discontinua).

## Capítulo 4

### Sesgo de estimadores de regresión con residuos admisibles generalizados

#### 4.1 Introducción e hipótesis básicas

En el capítulo tres se estudió la curva de sesgo asintótico máximo de una clase amplia de estimadores de regresión robustos que incluía, entre otros, a S-,  $\tau$ - y R-estimadores. Todos estos estimadores son asintóticamente normales bajo condiciones de regularidad. Estas condiciones son ligeramente restrictivas en el sentido que se explica a continuación. Tomemos el ejemplo de la clase de S-estimadores. Por la proposición 2 y por la fórmula (1.20), los S-estimadores más robustos respecto al sesgo son aquellos basados en funciones  $\chi$  de salto. Sin embargo, la distribución asintótica de este tipo de S-estimadores no es normal (véase Rousseeuw y Yohai, 1984; y Davies, 1990). Este es el caso, por ejemplo, del LMS-estimador de Rousseeuw (1984), que converge a una tasa  $n^{-1/3}$  hacia una distribución no normal. Parecidos razonamientos se pueden hacer en la clase de los  $\tau$ -estimadores.

Como vimos en el capítulo uno, una manera de resolver este problema consiste en

aplicar la definición de S-estimador a las diferencias de los residuos en vez de hacerlo a los residuos (GS-estimadores). Pero la idea que conduce a definir los GS-estimadores puede ser aplicada a otros grupos de estimadores de regresión robustos. Así, en la sección 4.3 se definirán los  $G\tau$ - y los GR-estimadores. En general, dado un estimador de regresión robusto con residuos admisibles como los considerados en el capítulo anterior, se puede considerar su versión generalizada. En este capítulo se estudian las propiedades de robustez asintótica de la clase de los estimadores de regresión generalizados construidos de esta forma. En concreto, tras probar que no es posible aplicar la teoría desarrollada en el capítulo anterior, se proporciona un método general para calcular su sesgo asintótico máximo. Además, se encuentra el estimador de sesgo minimax en esta clase y se comparan sus propiedades con el estimador de sesgo minimax en la clase de estimadores con residuos admisibles.

Sean  $X$  e  $Y$  dos variables aleatorias independientes e idénticamente distribuidas según  $F$ . Denotaremos en este capítulo por  $F^*$  la función de distribución de la variable aleatoria  $|X - Y|$ . Así, si en este capítulo  $F_{H,\theta}$  denota la distribución de los residuos  $r_i(\theta)$  cuando  $(y_i, x_i)$  se distribuye según  $H$ , entonces  $F_{H,\theta}^*$  es la distribución de las diferencias de residuos  $|r_i(\theta) - r_j(\theta)|$  cuando  $i \neq j$ .

En el capítulo anterior vimos que un estimador robusto de regresión con residuos admisibles puede definirse mediante (1.18) a partir de un funcional  $J$  que verifica la hipótesis 5(b).

En la siguiente definición, vamos a precisar la clase de estimadores robustos de regresión que estudiaremos en este capítulo.

**Definición 4** *Se dice que el funcional  $\mathbf{T}$  define un estimador con residuos admisibles*

generalizado del vector de parámetros  $\theta$  si

$$\mathbf{T}(H) = \arg \min_{\theta} J(F_{H,\theta}^*), \quad (4.1)$$

donde  $J$  es un funcional que verifica la hipótesis 5, para algún  $\eta > 0$ .

A partir de una muestra concreta, los estimadores se calculan aplicando  $\mathbf{T}$  a la función de distribución empírica de los datos. Si, dado un estimador con residuos admisibles, lo calculamos sobre los valores absolutos de las diferencias de los residuos en lugar de sobre los residuos, obtenemos un estimador con residuos admisibles generalizado. Los GS-estimadores de regresión, definidos en (1.28) y (1.29), constituyen un primer ejemplo de estimador con residuos admisibles generalizado. El resto de ejemplos a los que podremos aplicar nuestros resultados se encuentran en la sección 4.3.

El funcional  $J^*$  asociado a  $J$  se define de la siguiente forma: dada una función de distribución  $F$ ,

$$J^*(F) \equiv J(F^*).$$

Según convenga, se escribirá  $J(F_{H,\theta}^*)$ , como en la ecuación (4.1), ó  $J^*(F_{H,\theta})$ .

Cuando sea necesario, también se utilizará la notación

$$F_{H_1 \times H_2, \theta}(u) \equiv P_{H_1 \times H_2} \{|r_1(\theta) - r_2(\theta)| \leq u\},$$

por lo que  $F_{H,\theta}^*(u) = F_{H \times H, \theta}(u)$ .

Finalmente, utilizaremos  $\delta_0$  para denotar la distribución unidimensional degenerada en 0 y  $\Delta_0$  para denotar la distribución  $(p+1)$ -dimensional degenerada en  $\mathbf{0}$ .

Además de la hipótesis 5 serán necesarios los siguientes supuestos para probar los resultados más importantes de este capítulo:

**Hipótesis 10** Sea  $V_n$  la función de distribución uniforme en el intervalo  $(n, 2n)$ . Entonces

$$\lim_{n \rightarrow \infty} J^*[(1 - \epsilon)F_{H_0,0} + \epsilon V_n] \geq J^*[F_{H,0}],$$

para toda distribución  $H \in V_\epsilon$ .

La hipótesis 10 es una condición técnica que cumple una función análoga a la que cumpliría la hipótesis 6 en el capítulo anterior. Si el funcional  $J$  verifica la hipótesis 5(a) para distribuciones posiblemente subestocásticas y es  $\eta$ -monótono para  $\eta = \epsilon(2 - \epsilon)$ , entonces  $J^*$  cumple la hipótesis 10.

La siguiente hipótesis que vamos a introducir impone condiciones sobre las distribuciones de los errores y de las variables regresoras. Nótese que no es necesario que la distribución  $G_0$  sea elíptica.

**Hipótesis 11** La distribución de los errores  $F_0$  tiene una densidad  $f_0$  que es simétrica y estrictamente unimodal. La distribución de los regresores,  $G_0$ , es tal que  $\theta'x$  tiene una densidad que es simétrica y estrictamente unimodal para todo  $\theta \neq 0$ . En particular,  $P_{G_0}\{\theta'x = 0\} = 0$ , para todo  $\theta \neq 0$ .

## 4.2 Curva de sesgo de estimadores con residuos admisibles generalizados

### 4.2.1 Lemas previos

En este apartado se incluyen algunos resultados auxiliares que serán importantes en las demostraciones de los resultados de este capítulo.

En el primer lema se obtiene la distribución de la diferencia de dos distribuciones uniformes independientes. Su demostración se reduce a un simple cambio de variable por lo que será omitida.

**Lema 10** Sean  $X$  e  $Y$  dos variables aleatorias independientes distribuidas uniformemente sobre el intervalo  $(a, b)$ . La variable aleatoria  $X - Y$  tiene función de densidad

$$f(x) = \begin{cases} (b - a + x)/(b - a)^2, & \text{si } a - b < x < 0 \\ (b - a - x)/(b - a)^2, & \text{si } 0 < x < b - a \end{cases}$$

El siguiente lema es básico para poner de manifiesto las discrepancias respecto a la curva de sesgo asintótico máximo entre los estimadores con residuos admisibles y los estimadores con residuos admisibles generalizados. En él se analiza la diferencia de comportamiento en el límite de las diferencias de residuos en valor absoluto si: (1) los residuos son cada vez mayores en magnitud pero cada vez más concentrados, apartado (a) del lema, o (2) los residuos son cada vez mayores en magnitud y además su dispersión aumenta de manera indefinida, apartado (c) del lema.

**Lema 11** Sea  $\{c_n\}$  una sucesión de números reales tal que  $\lim_{n \rightarrow \infty} c_n = \infty$ . Sea  $\{\theta_n\}$  una sucesión convergente en  $\mathbb{R}^p$ . Sea  $U_n$  una distribución de  $(y_n, \mathbf{x}_n)$  tal que los residuos  $r(\theta_n)$ , bajo  $U_n$ , siguen una distribución uniforme sobre el intervalo  $(c_n - 1/c_n, c_n + 1/c_n)$ . Finalmente, sea  $V_n$  una distribución de  $(y_n, \mathbf{x}_n)$  tal que los residuos  $r(\theta_n)$ , bajo  $V_n$ , siguen una distribución uniforme sobre el intervalo  $(c_n, 2c_n)$ . Bajo la hipótesis 11, se verifica

$$(a) \lim_{n \rightarrow \infty} F_{U_n \times U_n, \theta_n}(u) = 1, \quad \text{para todo } u > 0.$$

$$(b) \lim_{n \rightarrow \infty} F_{H_0 \times U_n, \theta_n}(u) = 0, \quad \text{para todo } u > 0.$$

$$(c) \lim_{n \rightarrow \infty} F_{V_n \times V_n, \theta_n}(u) = 0, \quad \text{para todo } u > 0.$$

$$(d) \lim_{n \rightarrow \infty} F_{H_0 \times V_n, \theta_n}(u) = 0, \text{ para todo } u > 0.$$

**Demostración:**

Para probar (a), nótese que

$$F_{U_n \times U_n, \theta_n}(u) = P_{U_n \times U_n} \{|r_1(\theta_n) - r_2(\theta_n)| \leq u\} = 2 \int_0^u \frac{c_n^2}{4} \left[ \frac{2}{c_n} - x \right] I\{0 < x < \frac{2}{c_n}\} dx,$$

por la simetría de la distribución de  $r_1 - r_2$  y por el lema 10. Como consecuencia, para  $n$  suficientemente grande y para todo  $u > 0$ ,

$$F_{U_n \times U_n, \theta_n}(u) = 2 \int_0^{2/c_n} \frac{c_n^2}{4} \left[ \frac{2}{c_n} - x \right] dx = \frac{c_n^2}{2} \left[ \frac{4}{c_n^2} - \frac{4}{2c_n^2} \right] = 1.$$

Para probar (b) hay que tener en cuenta que

$$\begin{aligned} F_{H_0 \times U_n, \theta_n}(u) &= P_{H_0 \times U_n} \{|r_1(\theta_n) - r_2(\theta_n)| \leq u\} = P_{H_0 \times U_n} \{r_2 - u \leq r_1 \leq r_2 + u\} \\ &= \frac{c_n}{2} \int_{c_n - 1/c_n}^{c_n + 1/c_n} P_{H_0} \{r_2 - u \leq r_1 \leq r_2 + u\} dr_2 \\ &= F_{H_0, \theta_n}(\xi_n + u) - F_{H_0, \theta_n}(\xi_n - u), \end{aligned}$$

donde  $\xi_n \in (c_n - 1/c_n, c_n + 1/c_n)$ , por el teorema del valor medio para integrales de funciones continuas y dado que, por la hipótesis 11, la función de distribución  $F_{H_0, \theta_n}$  es continua. Ahora bien, puesto que  $\{\theta_n\}$  es convergente, es claro que

$$\lim_{n \rightarrow \infty} F_{H_0, \theta_n}(\xi_n + u) - F_{H_0, \theta_n}(\xi_n - u) = 0,$$

para todo  $u > 0$ .

Para probar (c) utilizamos de nuevo el lema 10 y hacemos un cálculo análogo al del apartado (a) para obtener

$$F_{V_n \times V_n, \theta_n}(u) = P_{V_n \times V_n} \{|r_1(\theta_n) - r_2(\theta_n)| \leq u\} = 2 \int_0^u \frac{c_n - x}{c_n^2} I\{0 < x < c_n\} dx.$$

Para  $n$  suficientemente grande,

$$F_{V_n \times V_n, \theta_n}(u) = 2 \int_0^u \frac{c_n - x}{c_n^2} dx = \frac{2}{c_n^2} \left[ c_n u - \frac{u^2}{2} \right],$$

expresión que converge a 0 cuando  $n \rightarrow \infty$  para todo  $u > 0$ .

La demostración del apartado (d) es totalmente análoga a la de (b) por lo que se omite.

□

Siempre podemos encontrar sucesiones  $\{U_n\}$  y  $\{V_n\}$  que cumplen las condiciones del enunciado. Por ejemplo, para obtener una sucesión  $\{U_n\}$  basta tomar variables  $y_n$  distribuidas uniformemente en el intervalo  $(c_n - 1/c_n, c_n + 1/c_n)$  y variables  $\mathbf{x}_n$  degeneradas en  $\mathbf{0}$ . Análogamente para obtener  $\{V_n\}$ .

En el siguiente lema se estudia la forma de  $F_{H, \theta}^*$  cuando la distribución de las observaciones  $H$  pertenece a un entorno de contaminación.

**Lema 12** *Sea el entorno de contaminación*

$$V_\epsilon = \{H : H = (1 - \epsilon)H_0 + \epsilon\tilde{H}, \tilde{H} \text{ distribución arbitraria}\}.$$

*Para todo  $u \geq 0$ , para todo  $\theta \in \mathbb{R}^p$  y para toda distribución  $H \in V_\epsilon$ , se verifica*

$$F_{H, \theta}^*(u) = (1 - \epsilon)^2 F_{H_0, \theta}^*(u) + 2\epsilon(1 - \epsilon) F_{H_0 \times \tilde{H}, \theta}(u) + \epsilon^2 F_{\tilde{H}, \theta}^*(u).$$

**Demostración:**

Como consecuencia de la independencia entre los residuos  $r_1$  y  $r_2$  y aplicando el teorema de Fubini, se tiene que

$$\begin{aligned}
F_{\tilde{H},\boldsymbol{\theta}}^*(u) &= P_{H \times H, \boldsymbol{\theta}}\{|r_1 - r_2| \leq u\} \\
&= \int \int I\{|r_1 - r_2| \leq u\} dF_{H,\boldsymbol{\theta}}(r_1) dF_{H,\boldsymbol{\theta}}(r_2) \\
&= \int \left[ (1 - \epsilon) \int I\{|r_1 - r_2| \leq u\} dF_{H_0,\boldsymbol{\theta}}(r_1) \right. \\
&\quad \left. + \epsilon \int I\{|r_1 - r_2| \leq u\} dF_{\tilde{H},\boldsymbol{\theta}}(r_1) \right] dF_{H,\boldsymbol{\theta}}(r_2) \\
&= (1 - \epsilon)^2 \int \int I\{|r_1 - r_2| \leq u\} dF_{H_0,\boldsymbol{\theta}}(r_1) dF_{H_0,\boldsymbol{\theta}}(r_2) \\
&\quad + 2\epsilon(1 - \epsilon) \int \int I\{|r_1 - r_2| \leq u\} dF_{H_0,\boldsymbol{\theta}}(r_1) dF_{\tilde{H},\boldsymbol{\theta}}(r_2) \\
&\quad + \epsilon^2 \int \int I\{|r_1 - r_2| \leq u\} dF_{\tilde{H},\boldsymbol{\theta}}(r_1) dF_{\tilde{H},\boldsymbol{\theta}}(r_2) \\
&= (1 - \epsilon)^2 F_{H_0,\boldsymbol{\theta}}^*(u) + 2\epsilon(1 - \epsilon) F_{H_0 \times \tilde{H},\boldsymbol{\theta}}(u) + \epsilon^2 F_{\tilde{H},\boldsymbol{\theta}}^*(u)
\end{aligned}$$

que es la expresión dada en el enunciado. □

Los lemas siguientes son resultados esencialmente técnicos pero serán importantes en las demostraciones de las proposiciones 19 y 20 de la sección siguiente.

**Lema 13** *Si se verifica la hipótesis 11,*

$$F_{H_0 \times \tilde{H},\boldsymbol{\theta}}(u) \leq F_{H_0 \times \Delta_0,\boldsymbol{\theta}}(u)$$

para todo  $u \geq 0$ , para todo  $\boldsymbol{\theta} \in \mathbb{R}^p$  y para toda distribución  $\tilde{H}$ .

**Demostración:**

Por la hipótesis 11, la distribución de los residuos  $r_i(\boldsymbol{\theta})$  es simétrica y estrictamente unimodal para cualquier  $\boldsymbol{\theta} \in \mathbb{R}^p$ . Teniendo en cuenta este hecho y la independencia

entre  $r_1$  y  $r_2$ , se tiene que

$$\begin{aligned}
 F_{H_0 \times \tilde{H}, \boldsymbol{\theta}}(u) &= P_{H_0 \times \tilde{H}}\{|r_1(\boldsymbol{\theta}) - r_2(\boldsymbol{\theta})| \leq u\} \\
 &= \int P_{H_0}\{r_2 - u \leq r_1 \leq r_2 + u\} dF_{\tilde{H}, \boldsymbol{\theta}}(r_2) \\
 &\leq \int P_{H_0}\{-u \leq r_1 \leq u\} dF_{\tilde{H}, \boldsymbol{\theta}}(r_2) \\
 &= P_{H_0}\{|r_1(\boldsymbol{\theta})| \leq u\} = F_{H_0 \times \Delta_0, \boldsymbol{\theta}}(u).
 \end{aligned}$$

□

Nótese que para que se cumpla el lema anterior es suficiente que se verifique la desigualdad

$$P_{H_0}\{r_2 - u \leq r_1(\boldsymbol{\theta}) \leq r_2 + u\} \leq P_{H_0}\{-u \leq r_1(\boldsymbol{\theta}) \leq u\},$$

para todo  $r_2 > 0$ . Puede relajarse la hipótesis 11 y pedir que  $G_0$  sea tal que se verifique la desigualdad anterior en lugar de simetría y unimodalidad estricta de la distribución de  $\boldsymbol{\theta}'\mathbf{x}$ . Por ejemplo, supongamos que  $r_1(\boldsymbol{\theta})$  se distribuye uniformemente en el intervalo  $(-K_{\boldsymbol{\theta}}, K_{\boldsymbol{\theta}})$  para algún  $K_{\boldsymbol{\theta}} > 0$ . En este caso se verifican los resultados de este capítulo aunque no sea cierto que la distribución de  $\boldsymbol{\theta}'\mathbf{x}$  sea simétrica y estrictamente unimodal para todo  $\boldsymbol{\theta} \neq 0$ .

**Lema 14** *Sea  $J$  un funcional que verifica la hipótesis 5(a) y supongamos que se cumple la hipótesis 11. Entonces:*

$$J^*[F_{H, \boldsymbol{\theta}}] \geq J^*[(1 - \epsilon)F_{H_0, \boldsymbol{\theta}} + \epsilon\delta_0],$$

para todo  $\boldsymbol{\theta} \in \mathbb{R}^p$  y todo  $H \in V_\epsilon$ .

**Demostración:**

Aplicando el lema 12 y la definición del funcional  $J^*$ ,

$$J^*(F_{H,\theta}) = J[(1-\epsilon)^2 F_{H_0,\theta}^* + 2\epsilon(1-\epsilon)F_{H_0 \times \tilde{H},\theta} + \epsilon^2 F_{\tilde{H},\theta}^*],$$

para todo  $\theta \in \mathbb{R}^p$  y todo  $H \in V_\epsilon$ .

Es claro que para todo  $\theta \in \mathbb{R}^p$  y todo  $u \geq 0$ ,

$$F_{\tilde{H},\theta}^*(u) = F_{\tilde{H} \times \tilde{H},\theta}(u) \leq 1 = F_{\Delta_0 \times \Delta_0,\theta}(u) = \delta_0(u). \quad (4.2)$$

Por otra parte, aplicando el lema 13 se deduce que, para todo  $\theta \in \mathbb{R}^p$  y para todo  $u \geq 0$ ,

$$F_{H_0 \times \tilde{H},\theta}(u) \leq F_{H_0 \times \Delta_0,\theta}(u). \quad (4.3)$$

Aplicando la hipótesis 5(a) y teniendo en cuenta (4.2) y (4.3), se obtiene, para todo  $\theta \in \mathbb{R}^p$ ,

$$\begin{aligned} J^*(F_{H,\theta}) &= J[(1-\epsilon)^2 F_{H_0,\theta}^* + 2\epsilon(1-\epsilon)F_{H_0 \times \tilde{H},\theta} + \epsilon^2 F_{\tilde{H},\theta}^*] \\ &\geq J[(1-\epsilon)^2 F_{H_0,\theta}^* + 2\epsilon(1-\epsilon)F_{H_0 \times \Delta_0,\theta} + \epsilon^2 F_{\Delta_0,\theta}^*] \\ &= J^*[(1-\epsilon)F_{H_0,\theta} + \epsilon\delta_0]. \end{aligned}$$

□

**Lema 15** *Sea  $J$  un funcional que verifica la hipótesis 5(b) para algún  $\eta > 0$ . Definamos, para todo  $t > 0$ ,*

$$m^*(t) = \inf_{\|\theta\|=t} J^*[(1-\epsilon)F_{H_0,\theta} + \epsilon\delta_0].$$

*Bajo la hipótesis 11 se verifica*

- (a) Existe  $\theta_t \in \mathbb{R}^p$  tal que  $\|\theta_t\| = t$ , y  $m^*(t) = J^*[(1 - \epsilon)F_{H_0, \theta_t} + \epsilon\delta_0]$ .
- (b)  $m^*(t)$  es estrictamente creciente.

### Demostración:

Esta demostración, con algunas complicaciones técnicas añadidas, sigue el mismo esquema que la demostración del lema 4.

Para todo  $\theta \in \mathbb{R}^p$  y para todo  $u \geq 0$ ,  $F_{\Delta_0, \theta}(u) = \delta_0(u)$ . Sea  $\{\theta_n\}$  una sucesión tal que  $\lim_{n \rightarrow \infty} \theta_n = \theta$ . Por la definición de  $J^*$  y por el lema 12,

$$\begin{aligned} J^*[(1 - \epsilon)F_{H_0, \theta_n} + \epsilon\delta_0] &= J^*[(1 - \epsilon)F_{H_0, \theta_n} + \epsilon F_{\Delta_0, \theta}] \\ &= J[(1 - \epsilon)^2 F_{H_0, \theta_n}^* + 2\epsilon(1 - \epsilon)F_{H_0 \times \Delta_0, \theta_n} + \epsilon^2 F_{\Delta_0, \theta_n}^*]. \end{aligned}$$

Dado que  $J$  satisface la hipótesis 5(b),

$$\begin{aligned} \lim_{n \rightarrow \infty} J^*[(1 - \epsilon)F_{H_0, \theta_n} + \epsilon\delta_0] &= J[(1 - \epsilon)^2 F_{H_0, \theta}^* + 2\epsilon(1 - \epsilon)F_{H_0 \times \Delta_0, \theta} + \epsilon^2 F_{\Delta_0, \theta}^*] \\ &= J^*[(1 - \epsilon)F_{H_0, \theta} + \epsilon\delta_0]. \end{aligned}$$

Por lo tanto, la función  $f^*(\theta) = J^*[(1 - \epsilon)F_{H_0, \theta} + \epsilon\delta_0]$  es continua. Dado que el conjunto  $\{\theta : \|\theta\| = t\}$  es compacto, se alcanza el ínfimo en la definición de la función  $m^*(t)$ , con lo que queda probado el apartado (a).

Para probar (b), consideramos  $t_1 > t_2 > 0$ . Como consecuencia del apartado (a), existe  $\theta_1 \in \mathbb{R}^p$  tal que

$$m^*(t_1) = J^*[(1 - \epsilon)F_{H_0, \theta_1} + \epsilon\delta_0].$$

Bajo la hipótesis 11, podemos aplicar el lema A.3 de Yohai y Zamar (1993). Por lo tanto, para todo  $\theta \in \mathbb{R}^p$ ,  $\lambda > 0$  y  $u > 0$  se verifica que  $F_{H_0, \lambda \theta}(u)$  es estrictamente

decreciente en  $\lambda$ . Si sustituimos la distribución de los residuos por la de la diferencia de residuos

$$r_1(\boldsymbol{\theta}) - r_2(\boldsymbol{\theta}) = y_1 - y_2 - \boldsymbol{\theta}'(\mathbf{x}_1 - \mathbf{x}_2),$$

también se verifica que  $F_{H_0, \lambda}^*(u)$  es estrictamente decreciente en  $\lambda$ . La demostración es análoga a la del lema citado ya que si  $F_0$  y  $G_0$  cumplen el supuesto 11, también lo cumplen las distribuciones de  $y_1 - y_2$  y de  $\mathbf{x}_1 - \mathbf{x}_2$  en la expresión anterior.

Como consecuencia de esta propiedad se verifica que, para todo  $u > 0$ ,

$$\begin{aligned} & (1 - \epsilon)^2 F_{H_0, \boldsymbol{\theta}_1}^*(u) + 2\epsilon(1 - \epsilon) F_{H_0 \times \Delta_0, \boldsymbol{\theta}_1}(u) + \epsilon^2 F_{\Delta_0, \boldsymbol{\theta}_1}^*(u) \\ < & (1 - \epsilon)^2 F_{H_0, (t_2/t_1) \boldsymbol{\theta}_1}^*(u) + 2\epsilon(1 - \epsilon) F_{H_0 \times \Delta_0, (t_2/t_1) \boldsymbol{\theta}_1}(u) + \epsilon^2 F_{\Delta_0, (t_2/t_1) \boldsymbol{\theta}_1}^*(u). \end{aligned}$$

Aplicando la desigualdad anterior y teniendo en cuenta la hipótesis 5(b) y la definición de  $m^*(t)$ ,

$$m^*(t_1) > J^*[(1 - \epsilon) F_{H_0, (t_2/t_1) \boldsymbol{\theta}_1} + \epsilon \delta_0] \geq m^*(t_2),$$

por lo que  $m^*(t)$  es estrictamente creciente.

□

#### 4.2.2 Curva de sesgo asintótico máximo

En este apartado probaremos en primer lugar que no es posible aplicar la teoría del capítulo anterior para calcular la curva de sesgo asintótico máximo de los estimadores con residuos admisibles generalizados. Posteriormente, se dará un resultado que permite obtener dicha curva. El punto esencial que distingue la teoría de este capítulo de la teoría del capítulo anterior estriba en la diferencia de comportamiento entre ambas

clases de estimadores cuando no sólo la localización de las contaminaciones es más lejana sino que también aumenta su dispersión. Esta diferencia de comportamiento se refleja en los apartados (a) y (c) del lema 11 y tiene como consecuencia la siguiente proposición en la que se utiliza la hipótesis de  $\epsilon$ -monotonía definida en el enunciado de la hipótesis 5(b) de la página 81.

**Proposición 19** *Sea  $J$  un funcional  $\epsilon$ -monótono. Entonces, el funcional  $J^*$ , definido por  $J^*(F) \equiv J(F^*)$ , no verifica dicha propiedad por lo que el estimador con residuos admisibles generalizado asociado a  $J$  no es  $\epsilon$ -monótono.*

**Demostración:**

Sea  $U_n$  la distribución de  $(y_n, \mathbf{x}_n)$ , donde  $y_n$  se distribuye uniformemente en el intervalo  $(n - 1/n, n + 1/n)$  y  $\mathbf{x}_n$  es degenerado en el vector  $\mathbf{0}$ . Sea  $V_n$  la distribución de  $(y_n, \mathbf{x}_n)$ , donde  $y_n$  se distribuye uniformemente en el intervalo  $(n, 2n)$  y  $\mathbf{x}_n$  es degenerado en el vector  $\mathbf{0}$ . Por reducción al absurdo, supongamos que  $J^*$  es  $\epsilon$ -monótono. Consideremos las dos sucesiones siguientes de funciones de distribución:

$$F_n = (1 - \epsilon)F_{H_0,0} + \epsilon F_{U_n,0},$$

y

$$G_n = (1 - \epsilon)F_{H_0,0} + \epsilon F_{V_n,0}.$$

Las contaminaciones de la mixtura  $F_n$  consisten básicamente en un punto cuya localización es cada vez más lejana respecto al grupo de datos no contaminados. Por otra parte, en las contaminaciones correspondientes a  $G_n$ , no sólo la localización es más lejana sino que la dispersión también aumenta.

Dado que para todo  $u \geq 0$ , se verifica

$$\lim_{n \rightarrow \infty} F_n(u) = (1 - \epsilon)F_{H_0,0},$$

y

$$\lim_{n \rightarrow \infty} G_n(u) = (1 - \epsilon)F_{H_0,0},$$

la  $\epsilon$ -monotonía de  $J^*$  implicaría

$$\lim_{n \rightarrow \infty} J^*(F_n) = \lim_{n \rightarrow \infty} J^*(G_n). \quad (4.4)$$

Sin embargo, por el lema 12,

$$\begin{aligned} F_n^*(u) &= (1 - \epsilon)^2 F_{H_0,0}^*(u) + 2\epsilon(1 - \epsilon)F_{H_0 \times U_n,0}(u) + \epsilon^2 F_{U_n,0}^*(u), \\ G_n^*(u) &= (1 - \epsilon)^2 F_{H_0,0}^*(u) + 2\epsilon(1 - \epsilon)F_{H_0 \times V_n,0}(u) + \epsilon^2 F_{V_n,0}^*(u), \end{aligned}$$

y aplicando el lema 11 (con  $\theta_n = \mathbf{0}$  para todo  $n$ ), se deduce

$$\lim_{n \rightarrow \infty} F_n^*(u) = (1 - \epsilon)^2 F_{H_0 \times H_0,0}(u) + \epsilon^2 > (1 - \epsilon)^2 F_{H_0 \times H_0,0}(u) = \lim_{n \rightarrow \infty} G_n^*(u).$$

De esta última desigualdad se obtiene, ya que  $J$  es  $\epsilon$ -monótona,

$$\lim_{n \rightarrow \infty} J(F_n^*) < \lim_{n \rightarrow \infty} J(G_n^*),$$

o equivalentemente,

$$\lim_{n \rightarrow \infty} J^*(F_n) < \lim_{n \rightarrow \infty} J^*(G_n),$$

en contradicción con la igualdad (4.4).

□

El siguiente resultado es el más importante de este capítulo. En él se proporciona el método para calcular la curva de sesgo de cualquier estimador con residuos admisibles generalizado. Por razones técnicas, para calcular  $B_{\mathbf{T}}(\epsilon)$  se debe pedir  $\eta$ -monotonía para  $\eta = \epsilon(2 - \epsilon)$ , si bien esto no supondrá una restricción importante.

**Proposición 20** *Sea  $\mathbf{T}$  un estimador con residuos admisibles generalizado tal que el funcional asociado  $J$  verifica la hipótesis 5 con  $\eta = \epsilon(2 - \epsilon)$ . Sea  $t^* \in \mathbb{R}$  tal que*

$$m^*(t^*) \equiv \inf_{\|\tilde{\theta}\|=t^*} J^*[(1 - \epsilon)F_{H_0, \tilde{\theta}} + \epsilon\delta_0] = \lim_{n \rightarrow \infty} J^*[(1 - \epsilon)F_{H_0, 0} + \epsilon V_n], \quad (4.5)$$

donde  $V_n$  denota la distribución uniforme sobre el intervalo  $(n, 2n)$ . Entonces, bajo las hipótesis 10 y 11, se tiene que

$$B_{\mathbf{T}}(\epsilon) = t^*.$$

**Demostración:**

Supongamos que  $\tilde{\theta} \in \mathbb{R}^p$  es tal que  $t = \|\tilde{\theta}\| > t^*$ . Se probará en primer lugar que  $J^*(F_{H, \tilde{\theta}}) > J^*(F_{H, 0})$  para cualquier  $H \in V_\epsilon$ . Este hecho implica que  $\tilde{\theta}$  no puede ser solución del problema de minimización (4.1) para ninguna distribución del entorno de contaminación y, como consecuencia,  $B_{\mathbf{T}}(\epsilon) \leq t^*$ .

Sea  $H \in V_\epsilon$ . Por el lema 14,

$$J^*(F_{H, \tilde{\theta}}) \geq J^*[(1 - \epsilon)F_{H_0, \tilde{\theta}} + \epsilon\delta_0]. \quad (4.6)$$

Por definición de la función  $m^*(t)$  y, dado que  $t > t^*$ , aplicando el lema 15,

$$J^*[(1 - \epsilon)F_{H_0, \tilde{\theta}} + \epsilon\delta_0] \geq m^*(t) > m^*(t^*). \quad (4.7)$$

Pero, por (4.5),  $t^*$  verifica

$$m^*(t^*) = \lim_{n \rightarrow \infty} J^*[(1 - \epsilon)F_{H_0, 0} + \epsilon V_n], \quad (4.8)$$

y por la hipótesis 10,

$$\lim_{n \rightarrow \infty} J^*[(1 - \epsilon)F_{H_0, 0} + \epsilon V_n] \geq J^*[F_{H, 0}]. \quad (4.9)$$

Teniendo en cuenta (4.6), (4.7), (4.8) y (4.9), se obtiene

$$J^*(F_{H,\hat{\theta}}) > J^*(F_{H,0}), \text{ para todo } H \in V_\epsilon,$$

que es lo que queríamos probar.

Estableceremos a continuación la desigualdad contraria,  $B_{\mathbf{T}}(\epsilon) \geq t^*$ . Para ello, sea  $t < t^*$  y  $\theta_t \in \mathbb{R}^p$  tal que  $\|\theta_t\| = t$  y

$$m^*(t) = J^*[(1 - \epsilon)F_{H_0, \theta_t} + \epsilon\delta_0].$$

Tal vector  $\theta_t$  existe en aplicación del lema 15(a). Definamos la siguiente sucesión de contaminaciones:  $\tilde{H}_n$  es la distribución de  $(y_n, \mathbf{x}_n)$ , donde  $y_n = \theta_t' \mathbf{x}_n$  y  $\mathbf{x}_n$  se distribuye uniformemente en el intervalo  $(n\theta_t, 2n\theta_t)$ . Dado un vector de parámetros  $\beta \in \mathbb{R}^p$ , bajo la distribución  $\tilde{H}_n$  que acabamos de definir se verifica

$$(\theta_t - \beta)' \mathbf{x}_n \equiv U[n(t^2 - \beta' \theta_t), 2n(t^2 - \beta' \theta_t)]. \quad (4.10)$$

Sea ahora  $H_n = (1 - \epsilon)H_0 + \epsilon\tilde{H}_n$  y denotemos  $\mathbf{T}_n = \mathbf{T}(H_n)$ . Supongamos, por reducción al absurdo, que  $\sup_n \|\mathbf{T}_n\| < t$ . En este caso, existirá una subsucesión de  $\{\mathbf{T}_n\}$ , que denotamos también por  $\{\mathbf{T}_n\}$ , tal que  $\lim_{n \rightarrow \infty} \mathbf{T}_n = \tilde{\theta}$ , donde  $\tilde{t} = \|\tilde{\theta}\| < t$ .

A continuación probaremos los dos hechos siguientes:

$$\lim_{n \rightarrow \infty} F_{H_0 \times \tilde{H}, \mathbf{T}_n}(u) = 0, \quad (4.11)$$

y

$$\lim_{n \rightarrow \infty} F_{\tilde{H} \times \tilde{H}, \mathbf{T}_n}(u) = 0, \quad (4.12)$$

para todo  $u \geq 0$ .

Para demostrar (4.11), se debe tener en cuenta que

$$F_{H_0 \times \tilde{H}, \mathbf{T}_n}(u) = P_{H_0 \times \tilde{H}}\{|r_1(\mathbf{T}_n) - r_2(\mathbf{T}_n)| \leq u\},$$

y por la ecuación (4.10),  $r_2(\mathbf{T}_n) = (\boldsymbol{\theta}_t - \mathbf{T}_n)' \mathbf{x}_n$  se distribuye uniformemente en el intervalo  $(c_n, 2c_n)$  donde  $c_n = n(t^2 - \mathbf{T}_n' \boldsymbol{\theta}_t)$  converge a infinito ya que  $\lim_{n \rightarrow \infty} \mathbf{T}_n = \tilde{\boldsymbol{\theta}}$  y  $\tilde{\boldsymbol{\theta}}' \boldsymbol{\theta}_t \leq \tilde{t}t < t^2$ . Podemos aplicar el lema 11(d) con  $V_n = \tilde{H}_n$  y  $\boldsymbol{\theta}_n = \mathbf{T}_n$  para deducir (4.11).

Para probar (4.12), obsérvese que, por la ecuación (4.10), los residuos  $r_1(\mathbf{T}_n)$  y  $r_2(\mathbf{T}_n)$  se distribuyen ahora uniformemente sobre el intervalo  $(c_n, 2c_n)$ . De nuevo, aplicando el lema 11(c), con  $V_n = \tilde{H}_n$  y  $\boldsymbol{\theta}_n = \mathbf{T}_n$ , se obtiene (4.12).

Por (4.11), (4.12) y el lema 12,

$$\lim_{n \rightarrow \infty} F_{H_n \times H_n, \mathbf{T}_n}(u) = (1 - \epsilon)^2 F_{H_0 \times H_0, \tilde{\boldsymbol{\theta}}}(u), \quad \text{para todo } u \geq 0. \quad (4.13)$$

Por otra parte, aplicando de nuevo los lemas 11 y 12,

$$\lim_{n \rightarrow \infty} F_{H_n \times H_n, 0}(u) = (1 - \epsilon)^2 F_{H_0 \times H_0, 0}(u), \quad \text{para todo } u \geq 0. \quad (4.14)$$

Nótese que  $F_{H_0 \times H_0, \tilde{\boldsymbol{\theta}}}(u) \leq F_{H_0 \times H_0, 0}(u)$  para todo  $u \geq 0$ . Además, por (4.13), (4.14) y dado que  $J$  es  $\eta$ -monótono para  $\eta = \epsilon(2 - \epsilon)$ ,

$$\lim_{n \rightarrow \infty} J^*(F_{H_n, \mathbf{T}_n}) \geq \lim_{n \rightarrow \infty} J^*(F_{H_n, 0}). \quad (4.15)$$

Por (4.5) y por el lema 15, dado que  $t < t^*$ ,

$$\begin{aligned} \lim_{n \rightarrow \infty} J^*(F_{H_n, 0}) &= \lim_{n \rightarrow \infty} J^*[(1 - \epsilon)F_{H_0, 0} + \epsilon V_n] = m^*(t^*) \\ &> m^*(t) = J^*[(1 - \epsilon)F_{H_0, \boldsymbol{\theta}_t} + \epsilon \delta_0], \end{aligned} \quad (4.16)$$

ya que  $\boldsymbol{\theta}_t$  se eligió desde un principio en la dirección adecuada para verificar la última igualdad.

Nótese que, bajo  $\tilde{H}_n$ , se tiene que  $r(\boldsymbol{\theta}_t) = \boldsymbol{\theta}_t' \mathbf{x}_n - \boldsymbol{\theta}_t' \mathbf{x}_n = 0$  y, por lo tanto,

$$F_{\tilde{H}_n \times \tilde{H}_n, \boldsymbol{\theta}_t}(u) = F_{\Delta_0 \times \Delta_0, \boldsymbol{\theta}_t}(u) = 1,$$

y

$$F_{H_0 \times \tilde{H}_n, \theta_t}(u) = F_{H_0 \times \Delta_0, \theta_t}(u),$$

para todo  $u \geq 0$  y para todo  $n$ . Como consecuencia,

$$\lim_{n \rightarrow \infty} F_{H_n \times H_n, \theta_t} = (1 - \epsilon)^2 F_{H_0, \theta_t}^* + 2\epsilon(1 - \epsilon) F_{H_0 \times \Delta_0, \theta_t} + \epsilon^2 F_{\Delta_0, \theta_t}^*$$

y, por la  $\eta$ -monotonía del funcional  $J$ , con  $\eta = \epsilon(2 - \epsilon)$ ,

$$\lim_{n \rightarrow \infty} J^*(F_{H_n, \theta_t}) = J^*[(1 - \epsilon)F_{H_0, \theta_t} + \epsilon F_{\Delta_0, \theta_t}] = J^*[(1 - \epsilon)F_{H_0, \theta_t} + \epsilon \delta_0]. \quad (4.17)$$

Uniendo las cadenas de igualdades y desigualdades (4.15), (4.16) y (4.17) deducimos que

$$\lim_{n \rightarrow \infty} J^*(F_{H_n, \mathbf{T}_n}) > \lim_{n \rightarrow \infty} J^*(F_{H_n, \theta_t}),$$

y, por lo tanto, existe  $n$  suficientemente grande para el cual

$$J^*(F_{H_n, \mathbf{T}_n}) > J^*(F_{H_n, \theta_t})$$

en contradicción con el hecho de que

$$\mathbf{T}_n = \arg \min_{\theta} J^*(F_{H_n, \theta}).$$

La contradicción procede de suponer  $\sup_n \|\mathbf{T}(H_n)\| < t$  y, por lo tanto, se verifica que  $\sup_n \|\mathbf{T}(H_n)\| \geq t$ .

Para cualquier  $t < t^*$ , existe una distribución  $H_n \in V_\epsilon$  tal que  $\|\mathbf{T}(H_n)\| \geq t$ . Se sigue que  $B_{\mathbf{T}}(\epsilon) \geq t^*$ .

□

Cuando la distribución de los regresores,  $G_0$ , es esférica, es fácil comprobar que  $F_{H, \theta}^*$  depende de  $\theta$  únicamente a través del valor de  $\|\theta\|$ . Por lo tanto, podemos eliminar el ínfimo en la ecuación (4.5) y obtener el siguiente corolario:

**Corolario 5** *Bajo las hipótesis y la notación de la proposición 20 y suponiendo además que  $G_0$  es esférica, se verifica*

$$B_{\mathbf{T}}(\epsilon) = \|\theta^*\|,$$

donde  $\theta^* \in \mathbb{R}^p$  es tal que

$$J^*[(1 - \epsilon)F_{H_0, \theta^*} + \epsilon\delta_0] = \lim_{n \rightarrow \infty} J^*[(1 - \epsilon)F_{H_0, 0} + \epsilon V_n]. \quad (4.18)$$

### 4.3 Ejemplos de aplicación

En esta sección mostramos cómo utilizar el corolario 5 para obtener las curvas de sesgo de distintas clases de estimadores con residuos admisibles generalizados. Para simplificar los cálculos numéricos suponemos a lo largo de toda la sección que  $G_0$  es una distribución normal  $p$ -variante y que  $F_0$  es una normal estándar.

#### 4.3.1 GS-estimadores

Esta clase de estimadores ya fue definida en el capítulo uno. En este caso,  $J^*(F) = S^*(F) = S(F^*)$ , donde  $S$  es el funcional definido en (1.3). En el siguiente lema se comprueba que se verifican las condiciones para poder aplicar la proposición 20. En el lema 5 ya se probó que  $S$  satisface la hipótesis 5. Basta comprobar la hipótesis 10.

**Lema 16** *Si  $\chi \in C_b$ , entonces el funcional  $S^*$  cumple la hipótesis 10.*

**Demostración:**

Se trata de probar

$$s_\infty = \lim_{n \rightarrow \infty} s_n = \lim_{n \rightarrow \infty} S^*[(1 - \epsilon)F_{H_0, 0} + \epsilon V_n] \geq S^*[(1 - \epsilon)F_{H_0, 0} + \epsilon G] = s_0.$$

Por definición del funcional  $S^*$ ,

$$(1 - \epsilon)^2 \int_{-\infty}^{\infty} \chi\left(\frac{y}{s_n}\right) dF(y) + 2\epsilon(1 - \epsilon) \int_{-\infty}^{\infty} \chi\left(\frac{u}{s_n}\right) dF_{H_0 \times V_n, 0}(u) + \epsilon^2 \int_{-\infty}^{\infty} \chi\left(\frac{u}{s_n}\right) dF_{V_n \times V_n, 0}(u) \leq b. \quad (4.19)$$

Ahora, para todo  $k > 0$ ,

$$\begin{aligned} \int_{-\infty}^{\infty} \chi(u) dF_{V_n \times V_n, 0}(u) &= 2 \int_0^{\infty} \chi(u) dF_{V_n \times V_n, 0}(u) \\ &\geq 2\chi(k) \int_k^{\infty} dF_{V_n \times V_n, 0}(u) = 2\chi(k)[1 - F_{V_n \times V_n, 0}(k)] \end{aligned}$$

y, aplicando el lema 11,

$$\lim_{n \rightarrow \infty} \int_{-\infty}^{\infty} \chi(u) dF_{V_n \times V_n, 0}(u) \geq \chi(k),$$

para todo  $k > 0$ . Haciendo  $k \rightarrow \infty$ , se obtiene

$$\lim_{n \rightarrow \infty} \int_{\infty}^{\infty} \chi(u) dF_{V_n \times V_n, 0}(u) = 1. \quad (4.20)$$

Análogamente se puede comprobar que

$$\lim_{n \rightarrow \infty} \int_{\infty}^{\infty} \chi(u) dF_{H_0 \times V_n, 0}(u) = 1. \quad (4.21)$$

Por lo tanto, por (4.20) y (4.21), si tomamos límites en (4.19) cuando  $n \rightarrow \infty$ ,

$$(1 - \epsilon)^2 \int_{-\infty}^{\infty} \chi\left(\frac{y}{s_{\infty}}\right) dF(y) + 2\epsilon(1 - \epsilon) + \epsilon^2 \leq b. \quad (4.22)$$

Supongamos ahora que

$$(1 - \epsilon)^2 \int_{-\infty}^{\infty} \chi\left(\frac{y}{s_0}\right) dF(y) + 2\epsilon(1 - \epsilon) + \epsilon^2 < b.$$

Entonces, existe  $s < s_0$  tal que

$$(1 - \epsilon)^2 \int_{-\infty}^{\infty} \chi\left(\frac{y}{s}\right) dF(y) + 2\epsilon(1 - \epsilon) + \epsilon^2 < b.$$

lo que implica que

$$(1 - \epsilon)^2 \int_{-\infty}^{\infty} \chi\left(\frac{y}{s}\right) dF(y) + 2\epsilon(1 - \epsilon) \int_{-\infty}^{\infty} \chi\left(\frac{u}{s}\right) dF_{H_0 \times G, 0}(u) + \epsilon^2 \int_{-\infty}^{\infty} \chi\left(\frac{u}{s}\right) dF_{G \times G, 0}(u) < b,$$

en contradicción con la definición de  $s_0$ . Por lo tanto

$$(1 - \epsilon)^2 \int_{-\infty}^{\infty} \chi\left(\frac{y}{s_0}\right) dF(y) + 2\epsilon(1 - \epsilon) + \epsilon^2 \geq b.$$

Teniendo en cuenta (4.22) se deduce que  $s_{\infty} \geq s_0$ .

□

Veamos cómo aplicar el corolario 5. Definamos la siguientes funciones:

$$g(s) = E_{\Phi} \chi(y/s),$$

y

$$h(s, \epsilon) = (1 - \epsilon)^2 g(2^{-1/2}s) + 2\epsilon(1 - \epsilon)g(s).$$

Sea  $S_1 = S^*[(1 - \epsilon)F_{H_0, \theta^*} + \epsilon\delta_0]$ . Entonces,  $S_1$  debe cumplir

$$(1 - \epsilon)^2 E_{H_0 \times H_0} \chi\left(\frac{r_1(\theta^*) - r_2(\theta^*)}{S_1}\right) + 2\epsilon(1 - \epsilon) E_{H_0} \chi\left(\frac{r(\theta^*)}{S_1}\right) = b$$

o, equivalentemente,

$$h\left(\frac{S_1}{(1 + \|\theta^*\|^2)^{1/2}}, \epsilon\right) = b.$$

Despejando  $S_1$  se obtiene:

$$S_1 = (1 + \|\theta^*\|^2)^{1/2} h^{-1}(b, \epsilon). \quad (4.23)$$

Por otra parte, sea  $S_2 = \lim_{n \rightarrow \infty} S^*[(1 - \epsilon)F_{H_0, 0} + \epsilon V_n]$ . Entonces  $S_2$  verifica, por las ecuaciones (4.20) y (4.21),

$$(1 - \epsilon)^2 E_{H_0 \times H_0} \chi\left(\frac{y_1 - y_2}{S_2}\right) + 2\epsilon - \epsilon^2 = b$$

o, equivalentemente,

$$(1 - \epsilon)^2 g(2^{-1/2} S_2) = b - 2\epsilon + \epsilon^2.$$

Despejando  $S_2$  se obtiene

$$S_2 = 2^{1/2} g^{-1} \left( \frac{b - 2\epsilon + \epsilon^2}{(1 - \epsilon)^2} \right). \quad (4.24)$$

De la ecuación  $S_1 = S_2$ , teniendo en cuenta (4.23) y (4.24) y despejando  $\|\theta^*\| = B_S(\epsilon)$ ,

$$B_S(\epsilon) = \left[ 2 \left[ \frac{g^{-1} \left( \frac{b - 2\epsilon + \epsilon^2}{(1 - \epsilon)^2} \right)}{h^{-1}(b, \epsilon)} \right]^2 - 1 \right]^{1/2}. \quad (4.25)$$

De esta expresión se deduce que  $B_S(\epsilon)$  tiende a infinito cuando el numerador de la fracción

$$\frac{g^{-1} \left( \frac{b - 2\epsilon + \epsilon^2}{(1 - \epsilon)^2} \right)}{h^{-1}(b, \epsilon)}$$

converge a infinito o cuando el denominador converge a cero. Dado que lo primero ocurre cuando  $\epsilon \rightarrow 1 - (1 - b)^{1/2}$  y lo segundo ocurre cuando  $\epsilon \rightarrow (1 - b)^{1/2}$ , se tiene que el punto de ruptura de los GS-estimadores es

$$\epsilon^* = \min\{1 - (1 - b)^{1/2}, (1 - b)^{1/2}\}.$$

El mayor valor que puede alcanzar el punto de ruptura es  $\epsilon^* = 0.5$ , para  $b = 0.75$ .

Cuando se considera una función de salto y se toma  $b = 0.75$ , el GS-estimador resultante se reduce a calcular el vector de parámetros que minimiza el primer cuartil de las diferencias de los residuos en valor absoluto. Llamaremos a este vector de parámetros el estimador LQD (de las iniciales de la expresión en inglés *least quartile of differences*). Otra posible elección  $\chi$  es la función bicuadrada de Tukey que, para el valor  $c = 0.9958$  de la constante de ajuste, proporciona un estimador cuyo punto de ruptura es  $\epsilon^* = 0.5$ . A este estimador lo llamaremos GSTUKEY.

En la tabla 4.1, se muestran los resultados obtenidos al aplicar la expresión (4.25) a los estimadores LQD y GSTUKEY. Se observa que el sesgo del LQD es ligeramente menor que el del GSTUKEY. Por otra parte, según muestran Hössjer, Croux y Rousseeuw (1994), la eficiencia del LQD es también ligeramente menor que la del GSTUKEY. En la figura 4.1 se representan las curvas de sesgo asintótico máximo correspondientes a ambos estimadores.

### 4.3.2 $G\tau$ -estimadores

Sea  $S^*(F)$  un funcional basado en una función  $\chi_1$  tal que define un GS-estimador (véase el apartado anterior). Sea  $\chi_2$  otra función que, al igual que  $\chi_1$ , verifica la hipótesis 1. Consideremos en este apartado el funcional  $\tau^*$  definido por

$$\tau^*(F) = S^*(F) \left[ E_{F \times F} \chi_2 \left( \frac{y_1 - y_2}{S^*(F)} \right) \right]^{1/2}.$$

Podemos dar ahora la siguiente definición:

**Definición 5** *Se dice que el funcional  $T(H)$  define un  $G\tau$ -estimador ( $\tau$ -estimador generalizado) si*

$$T(H) = \arg \min_{\theta} \tau^*(F_H, \theta).$$

Para una muestra concreta, aplicamos el funcional a la función de distribución empírica correspondiente, con lo que resulta el estimador

$$\hat{\theta}_n = \arg \min_{\theta} S_n(\theta) \left[ \sum_{i < j} \chi_2 \left( \frac{|r_i(\theta) - r_j(\theta)|}{S_n(\theta)} \right) \right]^{1/2},$$

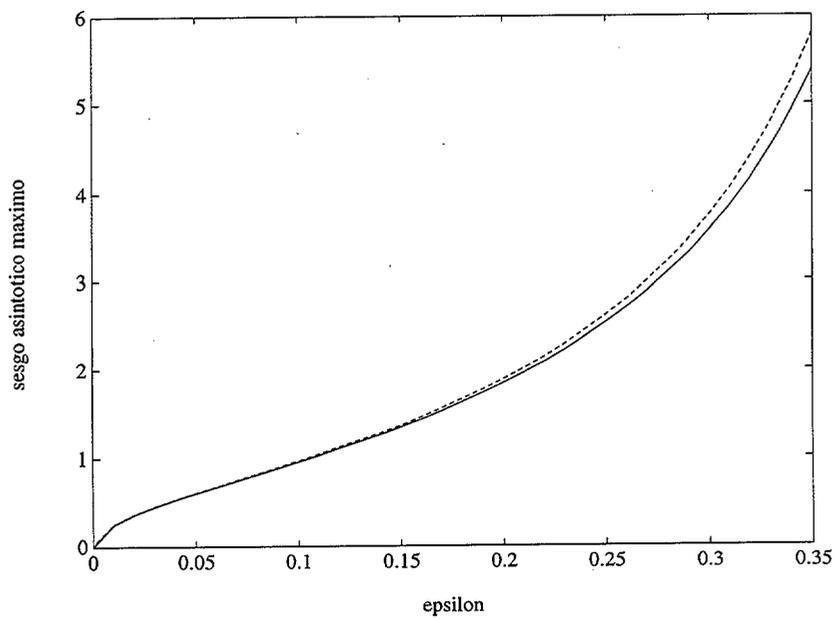
donde  $S_n(\theta)$  verifica la ecuación (1.29) con  $\chi = \chi_1$ .

Para estar en las hipótesis del corolario 5, la función  $\chi_2$  debe ser tal que

$$f(s) = s^2 E_{F \times F} \chi_2 \left( \frac{y_1 - y_2}{s} \right)$$

epsilon	0.05	0.1	0.15	0.2	0.25	0.3
LQD	0.59	0.95	1.34	1.83	2.51	3.56
GSTUKEY	0.60	0.96	1.36	1.88	2.60	3.73

**Tabla 4.1:** Sesgos asintóticos máximos de los estimadores LQD y GSTUKEY para distintos valores de la proporción de contaminación.



**Figura 4.1:** Curvas de sesgo asintótico máximo de los estimadores LQD (línea continua) y GSTUKEY (línea discontinua).

sea no decreciente para todo  $F$ . En este caso, si  $S^*$  cumple la hipótesis 10 (lo que se ha probado en el lema 16) entonces  $\tau^*$  también la cumple. Como vimos en el capítulo anterior, el funcional  $\tau$  verifica la hipótesis 5 para todo  $\eta > 0$ , por lo que podemos aplicar el corolario 5.

Definamos

$$\tau_1 = \left( \tau^*[(1 - \epsilon)F_{H_0, \theta^*} + \epsilon\delta_0] \right)^2,$$

y

$$\tau_2 = \left( \lim_{n \rightarrow \infty} \tau^*[(1 - \epsilon)F_{H_0, 0} + \epsilon V_n] \right)^2.$$

Para  $i = 1, 2$ , sea  $g_i(s) = E_{\Phi} \chi_i(y/s)$ , y sea

$$h_i(s, \epsilon) = (1 - \epsilon)^2 g_i(2^{-1/2}s) + 2\epsilon(1 - \epsilon)g_i(s).$$

Algunos cálculos semejantes a los ya realizados para GS-estimadores, dan

$$\tau_1 = (1 + \|\theta^*\|^2)^{1/2} \left[ h_1^{-1}(b, \epsilon) \right]^2 h_2 \left[ h_1^{-1}(b, \epsilon), \epsilon \right],$$

y

$$\tau_2 = S_2^2 \left[ (1 - \epsilon)^2 g_2(2^{-1/2}S_2) + 2\epsilon - \epsilon^2 \right],$$

donde

$$S_2 = 2^{1/2} g_1^{-1} \left( \frac{b - 2\epsilon + \epsilon^2}{(1 - \epsilon)^2} \right).$$

Dado que  $B_{\tau}(\epsilon) = \|\theta^*\|$ , si se impone la igualdad  $\tau_1 = \tau_2$  se tiene

$$1 + B_{\tau}^2(\epsilon) = 2 \left[ \frac{g_1^{-1} \left( \frac{b - 2\epsilon + \epsilon^2}{(1 - \epsilon)^2} \right)}{h_1^{-1}(b, \epsilon)} \right]^2 H(\epsilon),$$

donde

$$H(\epsilon) = \frac{(1 - \epsilon)^2 g_2 \left[ g_1^{-1} \left( \frac{b - 2\epsilon + \epsilon^2}{(1 - \epsilon)^2} \right) \right] + 2\epsilon - \epsilon^2}{h_2 \left[ h_1^{-1}(b, \epsilon), \epsilon \right]}.$$

Expresado de forma equivalente,

$$1 + B_{\tau}^2(\epsilon) = [1 + B_S^2(\epsilon)]H(\epsilon)$$

es la ecuación que relaciona el sesgo de un GS-estimador basado en  $\chi_1$  con el sesgo de un  $G\tau$ -estimador basado en  $\chi_1$  y  $\chi_2$ . Dado que la función  $H(\epsilon)$  es acotada para todo  $\epsilon < \min\{(1-b)^{1/2}, 1 - (1-b)^{1/2}\}$ , el punto de ruptura de un  $G\tau$ -estimador basado en dos funciones  $\chi_1$  y  $\chi_2$  es el mismo que el del GS-estimador correspondiente a  $\chi_1$ . De esta manera es factible elegir una función  $\chi_2$  apropiada para que el estimador sea eficiente sin que se vea afectado el punto de ruptura.

### 4.3.3 GR-estimadores

Estos estimadores resultan de considerar

$$J^*(F) = R^*(F) = \int a[F^*(u)]u^k dF^*(u)$$

en la definición 4. Más formalmente, se puede dar la siguiente definición:

**Definición 6** *Se dice que el funcional  $T(H)$  define un GR-estimador (R-estimador generalizado) si*

$$T(H) = \arg \min_{\theta} R^*(F_{H,\theta}).$$

De nuevo los casos más interesantes se obtienen tomando  $k = 1$  ó  $k = 2$  y definiendo la función

$$a(u) = \begin{cases} 1, & |u| \leq 1 - \alpha \\ 0, & |u| > 1 - \alpha \end{cases}$$

En lo que sigue nos restringimos a considerar el caso de esta función. Sea  $h_n = n(n-1)/2$ . Para  $k = 1$  el estimador resultante es

$$\hat{\theta}_n = \arg \min_{\theta} \sum_{l=1}^{[h_n(1-\alpha)]} \{|(r_i(\theta) - r_j(\theta))| : i < j\}_{(l)},$$

es decir, el vector de parámetros que minimiza la media recortada de las diferencias  $\{|(r_i(\boldsymbol{\theta}) - r_j(\boldsymbol{\theta}))| : i < j\}$  una vez que han sido desechadas las  $[h_n\alpha]$  mayores.

Análogamente, para  $k = 2$  el estimador resultante es

$$\hat{\boldsymbol{\theta}}_n = \arg \min_{\boldsymbol{\theta}} \sum_{l=1}^{[h_n(1-\alpha)]} \{|(r_i(\boldsymbol{\theta}) - r_j(\boldsymbol{\theta}))|^2 : i < j\}_{(l)}.$$

Estos dos estimadores son las respectivas generalizaciones de los estimadores  $\alpha$ -LTAV y  $\alpha$ -LTS considerados en el capítulo anterior, y los llamaremos  $\alpha$ -GLTAV y  $\alpha$ -GLTS respectivamente.

Vimos en la sección 3.3 que los funcionales  $R$  satisfacen la hipótesis 5 para todo  $\eta < \alpha$ , siempre que la función  $a(u)$  verifique la siguiente hipótesis

**Hipótesis 12** (a)  $a(u)$  es continua en el intervalo  $[0, 1 - \alpha]$ , con  $\alpha \geq 0.5$ .

(b)  $a(u) = 0$ , si  $1 - \alpha < u \leq 1$ .

(c)  $a(u) > 0$ , si  $0 < u < 1 - \alpha$ .

Dado que la proposición 20 exige que  $\eta = \epsilon(2 - \epsilon)$  y que el lema A4(a) de Yohai y Zamar (1993) prueba que se verifica la hipótesis 11 siempre que  $\epsilon(2 - \epsilon) < \alpha$ , es posible calcular  $B_R(\epsilon)$  para aquellos valores de  $\epsilon$  tales que  $\epsilon(2 - \epsilon) < \alpha$  o, dicho de otra forma, para aquellos valores de  $\epsilon$  tales que  $\epsilon < 1 - (1 - \alpha)^{1/2}$ . Veremos más adelante que esta restricción no limita el interés del resultado.

Para calcular  $B_R(\epsilon)$  denotemos

$$F_{\boldsymbol{\theta}^*} = (1 - \epsilon)F_{H_0, \boldsymbol{\theta}^*} + \epsilon\delta_0,$$

y

$$F_n = (1 - \epsilon)F_{H_0, 0} + \epsilon V_n.$$

Sea  $\gamma = (1 + \|\theta^*\|^2)^{1/2}$  y sean  $\Phi$  y  $\varphi$  las funciones de distribución y de densidad de una normal estándar. Sean  $0 < c_1 < \infty$  y  $0 < c_2 < \infty$  dos números reales tales que  $F_{\theta^*}^*(\gamma c_1) = 1 - \alpha$  y  $(1 - \epsilon)^2 F_{H_0,0}^*(2^{1/2} c_2) = 1 - \alpha$ , (es fácil comprobar que para que existan hemos de suponer  $\epsilon < \min\{(1 - \alpha)^{1/2}, 1 - (1 - \alpha)^{1/2}\}$ ). Introduzcamos también la notación

$$I_k(c) = \int_0^c u^k \varphi(c) du. \quad (4.26)$$

Con estas definiciones y aplicando el corolario 5 se trata de resolver en  $\theta^*$  la ecuación

$$R^*(F_{\theta^*}) = \lim_{n \rightarrow \infty} R^*(F_n). \quad (4.27)$$

Calculemos primero  $R^*(F_{\theta^*})$ . Operando se obtiene

$$\begin{aligned} R^*(F_{\theta^*}) &= \int_0^\infty a[F_{\theta^*}^*(u)] u^k dF_{\theta^*}^*(u) = \int_0^{\gamma c_1} u^k dF_{\theta^*}^*(u) \\ &= (1 - \epsilon)^2 \int_0^{\gamma c_1} u^k dF_{H_0, \theta^*}^*(u) + 2\epsilon(1 - \epsilon) \int_0^{\gamma c_1} u^k dF_{H_0, \theta^*}^*(u) \\ &= \gamma^k (1 - \epsilon) [(1 - \epsilon) 2^{\frac{2+k}{2}} I_k(2^{-1/2} c_1) + 4\epsilon I_k(c_1)]. \end{aligned} \quad (4.28)$$

Para calcular  $\lim_{n \rightarrow \infty} R^*(F_n)$ , sea  $c_n$  tal que  $F_n^*(c_n) = 1 - \alpha$ . Obsérvese que  $\lim_{n \rightarrow \infty} c_n = 2^{1/2} c_2$  ya que, aplicando el lema 11, para todo  $u > 0$ ,

$$\lim_{n \rightarrow \infty} F_n^*(u) = (1 - \epsilon)^2 F_{H_0,0}^*(u).$$

Se tiene que

$$R^*(F_n) = (1 - \epsilon)^2 \int_0^{c_n} u^k dF_{H_0,0}^*(u) + \epsilon(1 - \epsilon) \int_0^{c_n} u^k dF_{H_0 \times V_n,0}(u) + \epsilon^2 \int_0^{c_n} u^k dF_{V_n \times V_n,0}^*(u).$$

Ahora,

$$\int_0^{c_n} u^k dF_{H_0 \times V_n,0}(u) \leq c_n^k F_{H_0 \times V_n,0}(c_n) \leq C^k F_{H_0 \times V_n,0}(C),$$

puesto que la sucesión  $\{c_n\}$  es convergente. Por el lema 11(d),

$$\lim_{n \rightarrow \infty} C^k F_{H_0 \times V_n,0}(C) = 0.$$

Análogamente,

$$\lim_{n \rightarrow \infty} \int_0^{c_n} u^k dF_{V_n \times V_n, 0}^*(u) = 0.$$

Como consecuencia,

$$\lim_{n \rightarrow \infty} R^*(F_n) = (1 - \epsilon)^2 \int_0^{2^{1/2} c_2} u^k dF_{H_0, 0}^*(u) = (1 - \epsilon)^2 2^{\frac{2+k}{2}} I_k(c_2). \quad (4.29)$$

De la ecuación (4.27), teniendo en cuenta (4.28) y (4.29), se deduce que

$$[1 + B_R^2(\epsilon)]^{k/2} = \frac{(1 - \epsilon) I_k(c_2)}{(1 - \epsilon) I_k(2^{-1/2} c_1) + 2^{\frac{2-k}{2}} \epsilon I_k(c_1)}. \quad (4.30)$$

Esta fórmula es válida para  $\epsilon < \min\{(1 - \alpha)^{1/2}, 1 - (1 - \alpha)^{1/2}\}$ .

Los sesgos asintóticos máximos para los casos  $k = 1$  y  $k = 2$  se obtienen aplicando la fórmula (4.30), teniendo en cuenta que por (4.26),

$$I_1(c) = \varphi(0) - \varphi(c),$$

y

$$I_2(c) = \Phi(c) - \frac{1}{2} - c\varphi(c).$$

A partir del sesgo asintótico máximo obtenido, es posible obtener el punto de ruptura de los estimadores de esta clase. Es claro que  $B_R(\epsilon)$  tenderá a infinito si el denominador de la expresión (4.30) tiende a cero ó si el numerador tiende a infinito. Lo primero ocurre cuando  $c_1 \rightarrow 0$  y es fácil comprobar que esto ocurre si y sólo si  $\epsilon \rightarrow (1 - \alpha)^{1/2}$ . Lo segundo ocurre cuando  $c_2 \rightarrow \infty$  ó, equivalentemente, cuando  $\epsilon \rightarrow 1 - (1 - \alpha)^{1/2}$ . Como consecuencia, el punto de ruptura es

$$\epsilon^* = \min\{(1 - \alpha)^{1/2}, 1 - (1 - \alpha)^{1/2}\}.$$

Obsérvese que la fórmula (4.30) es válida para todo  $\epsilon < \epsilon^*$ , por lo que las restricciones introducidas sobre  $\epsilon$  son irrelevantes.

En la tabla 4.2 se muestran los resultados numéricos obtenidos para los estimadores  $\alpha$ -GLTAV y  $\alpha$ -GLTS con  $\alpha = 0.75$  para que el punto de ruptura sea  $\epsilon^* = 0.5$ . Al igual que sucede con las versiones no generalizadas, se observa que el estimador  $\alpha$ -GLTS es más robusto que el  $\alpha$ -GLTAV. Las correspondientes curvas han sido representadas en la figura 4.2.

#### 4.4 Teoría minimax en la clase de estimadores con residuos admisibles generalizados

En esta sección calculamos el estimador cuya función de sesgo máximo es mínima (sesgo minimax) dentro de la clase de estimadores con residuos admisibles generalizados, cuando la distribución de los regresores,  $G_0$ , es elíptica. Lo que probaremos es que el estimador minimax  $\mathbf{T}^*$  viene dado por:

$$\mathbf{T}^*(H) = \arg \min_{\boldsymbol{\theta}} F_{H \times H, \boldsymbol{\theta}}^{-1}(\alpha^*),$$

es decir, es el vector de parámetros que minimiza un cuantil  $\alpha^*$  de la distribución de la diferencia de los residuos en valor absoluto,  $F_{H, \boldsymbol{\theta}}^*$ . El valor de  $\alpha^*$  depende de la proporción de datos atípicos  $\epsilon$  y se dará explícitamente a lo largo de las demostraciones.

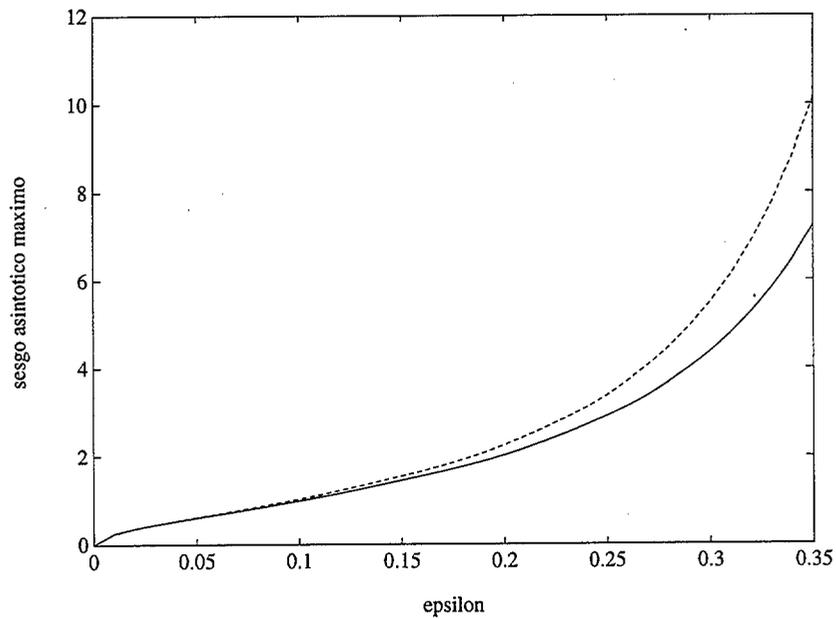
##### 4.4.1 Lemmas previos

En este apartado se prueban distintos lemas necesarios para las demostraciones de esta sección.

**Lema 17** *Sea  $J$  un funcional que verifica la hipótesis 5(b). Sean  $G_n$ , y  $F$  funciones de distribución sobre el intervalo  $[0, \infty)$ , continuas sobre  $(0, \infty)$ , tales que  $G_n(u) \rightarrow G(u)$ ,*

$\epsilon$	0.05	0.1	0.15	0.2	0.25	0.3
$\alpha$ -GLTAV	0.61	1.02	1.52	2.24	3.38	5.49
$\alpha$ -GLTS	0.60	0.98	1.42	2.02	2.90	4.37

**Tabla 4.2:** Sesgos asintóticos máximos de los estimadores  $\alpha$ -GLTAV y  $\alpha$ -GLTS donde  $\alpha = 0.75$ , para distintos valores de la proporción de contaminación.



**Figura 4.2:** Curvas de sesgo asintótico máximo de los estimadores  $\alpha$ -GLTAV (línea discontinua) y  $\alpha$ -GLTS (línea continua), donde  $\alpha = 0.75$ .

donde  $G$  es posiblemente subestocástica y es continua en  $(0, \infty)$ , con  $G(\infty) > 1 - \epsilon$ .

Entonces, si

$$\lim_{n \rightarrow \infty} J(G_n) = J(F),$$

existe  $\tilde{u} > 0$  tal que  $F(\tilde{u}) = G(\tilde{u})$ .

### Demostración:

Supongamos, por reducción al absurdo, que para todo  $u > 0$ ,  $F(u) > G(u)$ . Dado que  $J$  es  $\epsilon$ -monótono, se tiene que

$$J(F) < \lim_{n \rightarrow \infty} J(G_n),$$

que es una contradicción ya que habíamos supuesto  $\lim_{n \rightarrow \infty} J(G_n) = J(F)$ .

Análogamente llegamos a una contradicción si suponemos que para todo  $u > 0$ ,  $F(u) < G(u)$ .

□

**Lema 18** *Supongamos que se verifica la hipótesis 11 y que  $G_0$  es esférica. Entonces, para todo  $u > 0$ ,*

$$(a) \lim_{\|\theta\| \rightarrow \infty} F_{H_0 \times H_0, \theta}(u) = 0.$$

(b)  $F_{H_0 \times H_0, \theta}(u)$  es continua en  $\theta$ , uniformemente en  $u$ .

### Demostración:

La demostración consiste en seguir la prueba del lema A.2 de Yohai y Zamar (1993). Para ello, hay que tener en cuenta que

$$F_{H_0 \times H_0, \theta}(u) = P_{H_0 \times H_0} \{ |(y_1 - y_2) - \theta'(x_1 - x_2)| \leq u \},$$

y que las distribuciones de  $y_1 - y_2$  y de  $\mathbf{x}_1 - \mathbf{x}_2$  verifican las propiedades de  $F_0$  y  $G_0$  exigidas en la hipótesis 11, es decir, la distribución de  $y_1 - y_2$  es simétrica, estrictamente unimodal, la distribución de  $\mathbf{x}_1 - \mathbf{x}_2$  es esférica y la distribución de  $\boldsymbol{\theta}'(\mathbf{x}_1 - \mathbf{x}_2)$  es continua para todo  $\boldsymbol{\theta} \neq \mathbf{0}$  por lo que se cumple  $P\{\boldsymbol{\theta}'(\mathbf{x}_1 - \mathbf{x}_2) = 0\} = 0$ , para todo  $\boldsymbol{\theta} \neq \mathbf{0}$ .

□

**Lema 19** Sea  $\tilde{H} = (1 - \epsilon)H_0 + \epsilon\Delta_0$ . Definamos

$$\|\boldsymbol{\theta}^*\| = \sup\{\|\boldsymbol{\theta}\| : F_{\tilde{H},\boldsymbol{\theta}}^*(u) \geq (1 - \epsilon)^2 F_{H_0,0}^*(u), \text{ para todo } u \geq 0\}. \quad (4.31)$$

Bajo la hipótesis 11 se verifica:

(a)  $0 < \|\boldsymbol{\theta}^*\| < \infty$ , para todo  $\epsilon \in (0, 1/2)$ .

(b) Existe  $u^* > 0$  tal que,

$$F_{\tilde{H},\boldsymbol{\theta}^*}^*(u^*) = (1 - \epsilon)^2 F_{H_0,0}^*(u^*). \quad (4.32)$$

**Demostración:**

Dado que para todo  $\epsilon \in (0, 1/2)$ , el vector  $\mathbf{0}$  pertenece al conjunto de la ecuación (4.31), se sigue que  $\|\boldsymbol{\theta}^*\|$  está bien definido. Además, el supremo se alcanza ya que, por el lema 18(b), las funciones  $F_{H_0 \times H_0, \boldsymbol{\theta}}(u)$  y  $F_{H_0, \boldsymbol{\theta}}(u)$  son continuas en  $\boldsymbol{\theta}$  y, por lo tanto, también lo es la función  $F_{\tilde{H}, \boldsymbol{\theta}}^*(u)$ .

Aplicando ahora el lema 18(b), se deduce que existe  $\delta > 0$  tal que si  $\|\boldsymbol{\theta}\| < \delta$ ,

$$F_{H_0 \times H_0, \boldsymbol{\theta}}(u) \geq F_{H_0 \times H_0, 0}(u) - \frac{\epsilon^2}{2}, \text{ para todo } u \geq 0.$$

Entonces

$$\begin{aligned} F_{\tilde{H}, \theta}^*(u) &= (1 - \epsilon)^2 F_{H_0 \times H_0, \theta}(u) + 2\epsilon(1 - \epsilon) F_{H_0 \times \Delta_0, \theta}(u) + \epsilon^2 \\ &\geq (1 - \epsilon)^2 F_{H_0 \times H_0, 0}(u) - (1 - \epsilon)^2 \frac{\epsilon^2}{2} + \epsilon^2 \\ &\geq (1 - \epsilon)^2 F_{H_0 \times H_0, 0}(u), \text{ para todo } u \geq 0. \end{aligned}$$

Como consecuencia,  $\|\theta^*\| \geq \delta > 0$ .

Probaremos a continuación que  $\|\theta^*\| < \infty$ . Para ello, nótese que, por el lema 18(a), para todo  $\epsilon < 1/2$  existe  $\delta > 0$  tal que, para todo  $u \geq 0$ ,

$$\lim_{\|\theta\| \rightarrow \infty} [(1 - \epsilon)^2 F_{H_0 \times H_0, \theta}(u) + 2\epsilon(1 - \epsilon) F_{H_0 \times \Delta_0, \theta}(u) + \epsilon^2] = \epsilon^2 < (1 - \epsilon)^2 - \delta. \quad (4.33)$$

Además, para este  $\delta > 0$ , existe  $u_0$  tal que

$$(1 - \epsilon)^2 F_{H_0 \times H_0, 0}(u_0) > (1 - \epsilon)^2 - \frac{\delta}{2}. \quad (4.34)$$

Como consecuencia de (4.33) y (4.34),

$$\begin{aligned} \lim_{\|\theta\| \rightarrow \infty} F_{\tilde{H}, \theta}^*(u_0) &= \lim_{\|\theta\| \rightarrow \infty} [(1 - \epsilon)^2 F_{H_0 \times H_0, \theta}(u_0) + 2\epsilon(1 - \epsilon) F_{H_0 \times \Delta_0, \theta}(u_0) + \epsilon^2] \\ &< (1 - \epsilon)^2 - \delta/2 < (1 - \epsilon)^2 F_{H_0 \times H_0, 0}(u_0). \end{aligned}$$

Por lo tanto,  $\|\theta^*\| \leq c$ , para algún  $c \in \mathbb{R}$ .

Para probar la parte (b) del lema supongamos, por reducción al absurdo, que

$$F_{\tilde{H}, \theta^*}^*(u) > (1 - \epsilon)^2 F_{H_0, 0}^*(u), \text{ para todo } u > 0. \quad (4.35)$$

Existe  $u_0$  tal que

$$F_{\tilde{H}, \theta^*}^*(u) > 1 - \left(\epsilon + \frac{\epsilon^2}{2}\right) > (1 - \epsilon)^2 \geq (1 - \epsilon)^2 F_{H_0, 0}^*(u), \text{ para todo } u \geq u_0.$$

Por lo tanto, existe  $\delta_1 > 0$  tal que

$$F_{\tilde{H}, \theta^*}^*(u) > (1 - \epsilon)^2 F_{H_0, 0}^*(u) + \delta_1, \text{ para todo } u \geq u_0. \quad (4.36)$$

Ahora, por (4.35) y dado que las funciones  $F_{H_0 \times H_0, \theta^*}(u)$ ,  $F_{H_0, \theta^*}(u)$  y  $F_{H_0 \times H_0, 0}(u)$  son uniformemente continuas en el intervalo  $[0, u_0]$ , existe  $\delta_2 > 0$  tal que

$$F_{\tilde{H}, \theta^*}^*(u) > (1 - \epsilon)^2 F_{H_0, 0}^*(u) + \delta_2, \quad \text{para todo } u \in [0, u_0]. \quad (4.37)$$

Sea  $\delta = \min\{\delta_1, \delta_2\}$ . Entonces, por las desigualdades (4.36) y (4.37),

$$F_{\tilde{H}, \theta^*}^*(u) > (1 - \epsilon)^2 F_{H_0, 0}^*(u) + \delta, \quad \text{para todo } u \geq 0.$$

Como por el lema 18(b) las funciones  $F_{H_0, \theta}^*(u)$  y  $F_{H_0, \theta}(u)$  son continuas en  $\theta$  uniformemente en  $u$ , la desigualdad anterior implica que debe existir  $\tilde{\theta} \in \mathbb{R}^p$ , con  $\|\tilde{\theta}\| \geq \|\theta^*\|$ , tal que

$$F_{\tilde{H}, \tilde{\theta}}^*(u) \geq (1 - \epsilon)^2 F_{H_0, 0}^*(u), \quad \text{para todo } u \geq 0,$$

en contradicción con la definición de  $\|\theta^*\|$ .

□

#### 4.4.2 Estimadores de sesgo minimax

En este apartado se dan dos resultados. En la proposición 21 se establece una cota inferior para el sesgo de cualquier estimador con residuos admisibles generalizado. Dicha cota se alcanza, como se prueba en la proposición 22, para ciertos estimadores que tendrán por tanto el mínimo sesgo posible dentro de esta clase.

**Proposición 21** *Sea  $\|\theta^*\|$  tal y como se define en el enunciado del lema 19. Sea  $\mathbf{T}$  un estimador con residuos admisibles generalizado. Supongamos que se verifica la hipótesis 5 con  $\eta = \epsilon(2 - \epsilon)$ , y las hipótesis 10 y 11. Supongamos también que  $G_0$  es esférica. Se tiene que*

$$B_{\mathbf{T}}(\epsilon) \geq \|\theta^*\|.$$

**Demostración:**

Por el corolario 5,  $B_{\mathbf{T}}(\epsilon) = \|\boldsymbol{\theta}\|$ , donde  $\boldsymbol{\theta}$  verifica la ecuación (4.18). Por reducción al absurdo, supongamos que  $\|\boldsymbol{\theta}\| < \|\boldsymbol{\theta}^*\|$ .

Aplicando el lema 17 y teniendo en cuenta (4.18) se deduce que existe  $\tilde{u} > 0$  tal que

$$F_{\tilde{H},\boldsymbol{\theta}}^*(\tilde{u}) = (1 - \epsilon)^2 F_{H_0,0}^*(\tilde{u}), \quad (4.38)$$

donde  $\tilde{H} = (1 - \epsilon)H_0 + \epsilon\Delta_0$ .

Pero como estamos suponiendo que  $\|\boldsymbol{\theta}\| < \|\boldsymbol{\theta}^*\|$ , por definición de  $\|\boldsymbol{\theta}^*\|$ , se verifica

$$F_{\tilde{H},\boldsymbol{\theta}}^*(\tilde{u}) > F_{\tilde{H},\boldsymbol{\theta}^*}^*(\tilde{u}) \geq (1 - \epsilon)^2 F_{H_0,0}^*(\tilde{u}),$$

en contradicción con la ecuación (4.38).

□

**Proposición 22** *Supongamos que un funcional  $J$  verifica la hipótesis 5 con  $\eta = \epsilon(2 - \epsilon)$ , y la hipótesis 10. Supongamos también que  $G_0$  es esférica y que se verifica la hipótesis 11. Sea  $\alpha^* = (1 - \epsilon)^2 F_{H_0 \times H_0,0}(u^*)$ , donde  $u^*$  fue definido en el enunciado del lema 19. El funcional*

$$\mathbf{T}_{\alpha^*}(H) = \arg \min_{\boldsymbol{\theta}} F_{H \times H,\boldsymbol{\theta}}^{-1}(\alpha^*)$$

*verifica que  $B_{\mathbf{T}_{\alpha^*}}(\epsilon) \leq B_{\mathbf{T}}(\epsilon)$  para cualquier estimador  $\mathbf{T}$  con residuos admisibles generalizado basado en  $J$ .*

**Demostración:**

Es suficiente probar que  $\|\mathbf{T}_{\alpha^*}(H)\| \leq \|\theta^*\|$ , para cualquier distribución  $H$  del entorno de contaminación  $V_\epsilon$ . Por reducción al absurdo, supongamos que existe  $H = (1 - \epsilon)H_0 + \epsilon H^*$  tal que  $\theta = \mathbf{T}_{\alpha^*}(H)$  y  $\|\theta\| > \|\theta^*\|$ .

Sea  $\lambda = \|\theta\|/\|\theta^*\| > 1$  y definamos  $\tilde{\theta} = \lambda\theta^*$ . Obsérvese que  $\|\tilde{\theta}\| = \|\theta\|$ . Por los lemas 13 y 19(b), y por la definición de  $\alpha^*$ ,

$$\begin{aligned} F_{H \times H, \theta}(u^*) &= (1 - \epsilon)^2 F_{H_0, \theta}^*(u^*) + 2\epsilon(1 - \epsilon) F_{H_0 \times H^*, \theta}(u^*) + \epsilon^2 F_{H^* \times H^*, \theta}(u^*) \\ &\leq (1 - \epsilon)^2 F_{H_0, \theta}^*(u^*) + 2\epsilon(1 - \epsilon) F_{H_0 \times \Delta_0, \theta}(u^*) + \epsilon^2 \\ &= (1 - \epsilon)^2 F_{H_0, \tilde{\theta}}^*(u^*) + 2\epsilon(1 - \epsilon) F_{H_0 \times \Delta_0, \tilde{\theta}}(u^*) + \epsilon^2 \\ &< (1 - \epsilon)^2 F_{H_0, \theta^*}^*(u^*) + 2\epsilon(1 - \epsilon) F_{H_0 \times \Delta_0, \theta^*}(u^*) + \epsilon^2 \\ &= (1 - \epsilon)^2 F_{H_0 \times H_0, 0}(u^*) = \alpha^*. \end{aligned}$$

Como consecuencia,

$$F_{H \times H, \theta}^{-1}(\alpha^*) > u^*. \tag{4.39}$$

Por otra parte,

$$F_{H \times H, 0}(u^*) \geq (1 - \epsilon)^2 F_{H_0 \times H_0, 0}(u^*) = \alpha^*,$$

con lo que

$$F_{H \times H, 0}^{-1}(\alpha^*) \leq u^*. \tag{4.40}$$

De las desigualdades (4.39) y (4.40) se deduce que

$$F_{H \times H, \theta}^{-1}(\alpha^*) > F_{H \times H, 0}^{-1}(\alpha^*),$$

en contradicción con el hecho de que  $\theta = \mathbf{T}_{\alpha^*}(H)$ . Por lo tanto  $\|\mathbf{T}_{\alpha^*}(H)\| \leq \|\theta^*\|$ , como queríamos probar.

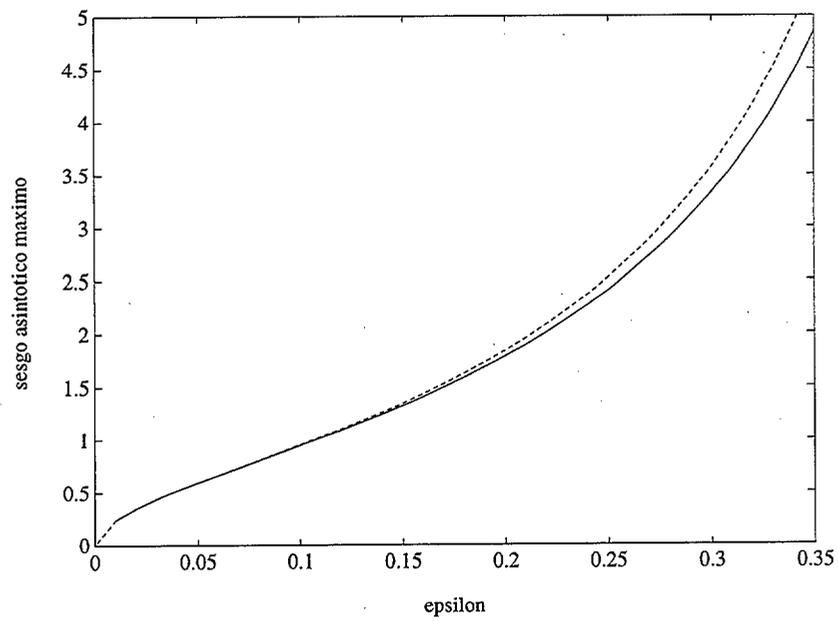
□

Para el caso en que  $G_0$  sea una distribución normal  $p$ -variante esférica, la tabla 4.3 recoge los resultados obtenidos al aplicar la proposición anterior. En ella se da el estimador de sesgo minimax en términos del cuantil a minimizar,  $\alpha^*$ , para cada proporción de contaminación, así como su curva de sesgo,  $B^*(\epsilon)$ , que es el menor sesgo posible en esta clase.

epsilon	0.05	0.1	0.15	0.2	0.25	0.3
$\alpha^*$	0.27	0.32	0.33	0.34	0.34	0.33
$B^*(\epsilon)$	0.59	0.94	1.32	1.78	2.40	3.31

**Tabla 4.3:** Sesgos y estimadores minimax en la clase de estimadores con residuos admisibles generalizados.

Compárense los sesgos óptimos de la tabla 4.3 con los sesgos del estimador LQD de la tabla 4.1. Se observa que la diferencia no es muy grande. En la clase de estimadores con residuos admisibles generalizados, el estimador LQD juega un papel semejante al estimador LMS en la clase de estimadores con residuos admisibles o a la mediana en la clase de estimadores de localización equivariantes ante traslaciones. En el gráfico 4.3 se representa el sesgo del estimador LQD y una curva formada por los sesgos minimax correspondientes a cada proporción de contaminación. Este gráfico permite apreciar la cercanía al óptimo del estimador LQD para todas las proporciones de contaminación.



**Figura 4.3:** Curva de sesgos óptimos (línea discontinua) y curva de sesgo del estimador LQD (línea continua).

## Referencias

- Beaton, A.E. y Tukey, J.W. (1974). The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data, *Technometrics*, **24**, 147–185.
- Coakley, C.W. y Hettmansperger, T.P. (1993). A bounded influence, high breakdown, efficient regression estimator. *J. Amer. Statist. Assoc.* **88**, 872–880.
- Croux, C. (1994). Efficient high-breakdown M-estimators of scale. *Statist. Probab. Lett.* **19**, 371–379.
- Croux, C., Rousseeuw, P.J. y Hössjer, O. (1994). Generalized S-Estimators. *J. Amer. Statist. Assoc.* **89**, 1271–1281.
- Davies, L. (1990). The asymptotics of S-estimators in the linear regression model. *Ann. Statist.*, **18**, 1651–1675.
- Devlin, S.J., Gnanadesikan, R., y Kettenring, J.R. (1981). Robust estimation of dispersion matrices and principal components. *J. Amer. Statist. Assoc.*, **76**, 354–362.
- Donoho, D.L. y Huber, P.J. (1983). The notion of breakdown point, en *A Festschrift for Erich A. Lehmann*, eds. P. Bickel, K. Doksum, J.L. Hodges Jr.. Wadsworth, Belmont, Calif., 157–184.
- Donoho, D.L. y Liu, R.C. (1988). The automatic robustness of minimum distance functionals. *Ann. Statist.* **16**, 552–586.

- Hampel, F.R. (1968). Contributions to the theory of robust estimation. Ph.D. thesis. University of California, Berkeley.
- Hampel, F.R. (1974). The influence curve and its role in robust estimation. *J. Amer. Statist. Assoc.* **69**, 383-393.
- He, X. y Simpson, D.G. (1993). Lower bounds for contamination bias: globally minimax versus locally linear estimation. *Ann. Statist.* **21**, 314-337.
- Hill, R.W. (1977). Robust regression when there are outliers in the carriers. Ph.D. thesis. Harvard University, Cambridge, Mass.
- Hodges, J.L., Jr. y Lehmann, E.L. (1963). Estimates of location based on rank tests. *Ann. Math. Statist.* **34**, 598-611.
- Hössjer, O. (1992). On the optimality of S-estimators. *Statist. Probab. Lett.* **14**, 413-419.
- Hössjer, O. (1994). Rank-based estimates in the linear model with high breakdown point. *J. Amer. Statist. Assoc.* **89**, 149-158.
- Hössjer, O., Croux, C. y Rousseeuw, P. (1994). Asymptotics of generalized S-estimators. *J. Multivariate Analysis.* **51**, 148-177.
- Huber, P.J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.* **35**, 73-101.
- Huber, P.J. (1981). *Robust Statistics*. Wiley, New York.
- Krasker, W.S. (1980). Estimation in linear regression models with disparate data points. *Econometrica*, **48**, 1333-1346.
- Krasker, W.S. y Welsch, R.E. (1982). Efficient bounded-influence regression estimation. *J. Amer. Statist. Assoc.*, **77**, 595-604.
- Li, G. y Chen, Z. (1985). Projection-pursuit approach to robust dispersion matrices and principal components: primary theory and Monte Carlo. *J. Amer. Statist. Assoc.*, **80**, 759-766.

- Mallows, C.L. (1975). On some topics in robustness. Technical Memorandum. Bell Telephone Laboratories, Murray Hill, N.J.
- Maronna, R.A. (1976). Robust M-estimators of multivariate location and scatter, *Ann. Statist.*, **4**, 51-67.
- Maronna, R.A., Bustos, O. y Yohai, V.J. (1979). Bias and efficiency robustness of general M-estimators for regression with random carriers, en *Smoothing techniques for curve estimation*, eds. T. Gasser y M. Rosenblatt, Springer-Verlag, New-York, 91-116.
- Maronna, R.A., Stahel, W.A. y Yohai, V.J. (1992) Bias-robust estimators of multivariate scatter based on projections. *J. Multivariate Analysis*, **42**, 141-161.
- Maronna, R.A. y Yohai, V.J. (1993). Bias-robust estimates of regression based on projections. *Ann. Statist.* **21**, 965-990.
- Martin, R.D., Yohai, V.J. y Zamar, R.H. (1989). Min-max bias robust regression. *Ann. Statist.* **17**, 1608-1630.
- Martin, R.D. y Zamar, R.H. (1989). Asymptotically Min-max bias robust M-estimates of scale for positive random variables. *J. Amer. Statist. Assoc.* **17**, 494-501.
- Martin, R.D. y Zamar, R.H. (1993a). Efficiency-constrained bias-robust estimation of location. *Ann. Statist.* **21**, 338-354.
- Martin, R.D. y Zamar, R.H. (1993b). Bias robust estimation of scale. *Ann. Statist.* **21**, 991-1017.
- Rousseeuw, P.J. (1984). Least median of squares regression. *J. Amer. Statist. Assoc.* **79**, 871-880.
- Rousseeuw, P.J. (1985). Multivariate estimation with high breakdown point, en: *Mathematical Statistics and Applications*. Eds. W. Grossmann, G. Pflug, I. Vincze, y W. Wertz, Reidel, Dordrecht, The Netherlands, 283-297.

- Rousseeuw, P.J. y Basset, G.W., Jr. (1990). The remedian: a robust averaging method for large data sets. *J. Amer. Statist. Assoc.* **85**, 97–104.
- Rousseeuw, P.J. y Croux, C. (1993). Alternatives to the median absolute deviation. *J. Amer. Statist. Assoc.* **88**, 1273–1283.
- Rousseeuw, P.J. y Croux, C. (1994). The bias of k-step M-estimators. *Statist. Probab. Lett.* **20**, 411–420.
- Rousseeuw, P.J. y Leroy, A.M. (1987). *Robust regression and outlier detection*. Wiley, New York.
- Rousseeuw, P.J. y Yohai, V.J. (1984). Robust regression by means of S-estimators. En *Robust and Nonlinear Time Series Analysis*, eds. J. Franke, W. Härdle y R.D. Martin. Lecture Notes in Statistics 26, Springer, New-York, 256–272.
- Serfling, R.J. (1984). Generalized L-, M- and R-statistics. *Ann. Statist.* **12**, 76–86.
- Siegel, A.F. (1982). Robust regression using repeated medians. *Biometrika*, **69**, 242–244.
- Simpson, D.G., Ruppert, D. y Carroll, R.J. (1992). On one-step GM-estimates and stability of inferences in linear regression. *J. Amer. Statist. Assoc.* **87**, 439–450.
- Simpson, D.G. y Yohai, V. (1993). Functional stability of one-step GM-estimators in linear regression. Technical Report 71, Department of Statistics, University of Illinois, Urbana-Champaign.
- Tukey, J.W. (1958). Bias and confidence in not-quite large samples (Abstract), *Ann. Math. Statist.*, **29**, p. 614.
- Tukey, J.W. (1960). A survey of sampling from contaminated distributions, en: *Contributions to Probability and Statistics*, I. Olkin, Ed., Stanford University Press, Stanford, California.
- Tukey, J.W. (1977). *Exploratory Data Analysis*, Addison-Wesley, Reading, MA.

- Yohai, V.J. (1987). High breakdown point and high efficiency robust estimates for regression. *Ann. Statist.* **15**, 642–656.
- Yohai, V.J. y Zamar, R.H. (1988). High breakdown point estimates of regression by means of the minimization of an efficient scale. *J. Amer. Statist. Assoc.* **83**, 406–413.
- Yohai, V.J. y Zamar, R.H. (1992). Optimally bounding a generalized gross error sensitivity of unbounded influence M-estimates of regression. Working paper 92-44, Departamento de Estadística y Econometría. Universidad Carlos III, Madrid.
- Yohai, V.J. y Zamar, R.H. (1993). A minimax-bias property of the least  $\alpha$ -quantile estimates. *Ann. Statist.* **21**, 1824–1842.
- Zamar, R.H. (1989). Robust estimation in the errors-in-variables model, *Biometrika*, **76**, 149–160.
- Zamar, R.H. (1992). Bias robust estimation in orthogonal regression. *Ann. Statist.* **20**, 1875–1888.