

Tesis de Posgrado

Estimación robusta de locación y dispersión multivariadas

Maronna, Ricardo Antonio

1974

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

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

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

Cita tipo APA:

Maronna, Ricardo Antonio. (1974). Estimación robusta de locación y dispersión multivariadas. Facultad de Ciencias Exactas y Naturales. Universidad de Buenos Aires.
http://digital.bl.fcen.uba.ar/Download/Tesis/Tesis_1457_Maronna.pdf

Cita tipo Chicago:

Maronna, Ricardo Antonio. "Estimación robusta de locación y dispersión multivariadas". Tesis de Doctor. Facultad de Ciencias Exactas y Naturales. Universidad de Buenos Aires. 1974.
http://digital.bl.fcen.uba.ar/Download/Tesis/Tesis_1457_Maronna.pdf

UNIVERSIDAD DE BUENOS AIRES
FACULTAD DE CIENCIAS EXACTAS Y NATURALES

ESTIMACION ROBUSTA DE LOCALIZACION
Y DISPERSION MULTIVARIADAS

Tesis presentada por
RICARDO ANTONIO MARONNA
para optar al grado de Doctor en Matemática

1977

1977

agradezco la ayuda brindada por el Dr. Víctor YCHAI, director de la Tesis y amigo muy querido, por el Dr. Luis Antonio SANTALO, consejero de estudios, y por el Dr. Rafael PANZCHE, a quien sin haber participado en este trabajo, le debo más de un escalearazo oportuno en momentos de flacurza.

43

INDICE

0.- NOTACION	1
1.- INTRODUCCION	
1.1.- El problema de la robustez	2
1.2.- El caso multivariado	6
1.3.- Plan de la obra	10
2.- EXISTENCIA Y UNICIDAD	
2.1.- Condiciones generales	12
2.2.- Existencia de las soluciones	16
2.3.- Unicidad	25
3.- PROPIEDADES ASINTOTICAS	
3.1.- Consistencia	30
3.2.- Normalidad asintótica	34
3.3.- Robustez	42
3.4.- Eficiencia asintótica	47
4.- MUESTRAS FINITAS	
4.1.- Cálculo numérico de los estimadores	55
4.2.- Método de Monte Carlo para variables "Normal/Independiente"	57
4.3.- Resultados de la simulación	63
4.4.- Cuestiones pendientes	67
REFERENCIAS	70

0 - NOTACION

En este trabajo, los vectores de R^m se indican con minúsculas subrayadas: \underline{x} , \underline{y} , ... y las matrices de $m \times m$ con mayúsculas subrayadas: \underline{U} , \underline{V} , ... \underline{I} es la matriz identidad, y \underline{U}' la traspuesta \underline{U} . Las normas de vectores, y las normas de matrices (como operadores) se escriben con $|\cdot|$. Se escribe $\underline{A} \geq \underline{B}$ si $\underline{A} - \underline{B}$ es semidefinida positiva.

En general P indica una medida de probabilidad en R^m y E_P el correspondiente operador de expectación, que se escribirá E cuando no haya lugar a dudas. \underline{x} es siempre la "variable muda" de la expectación, es decir, si h es una función de R^m en R , se escribe $E_P h(\underline{x}) = \int h(\underline{x}) P(d\underline{x})$. $N(\underline{a}, \underline{V})$ es la distribución normal m -variada con media \underline{a} y matriz de covarianzas \underline{V} .

B_r es la bola cerrada de centro \underline{a} y radio r , y S_r su superficie. La función indicatriz del conjunto A se escribe $I(\underline{x} \in A)$ ó $I(A)$. A' es el complemento de A .

Se abreviará: "cqs" y "t.c." en vez de "cualquiera sea" y "tal que", \lim_n en vez de $\lim_{n \rightarrow \infty}$ y $d^2(\underline{x}, \underline{y}, \underline{z})$ en vez de $(\underline{x}-\underline{y})' \underline{z} (\underline{x}-\underline{y})$

1 INTRODUCCION

1.1.- El problema de la robustez

Un hecho que inevitablemente debe ser tenido en cuenta en el trabajo estadístico, es que nunca se conoce exactamente el tipo de distribución de los datos. En 1821, Gauss observó que la suposición de normalidad traía como consecuencia lógica el uso de la media -en el problema de locación- y en general, del método de cuadrados mínimos en el de regresión (ver Huber, 1972), es decir procedimientos lineales en los datos, que por lo tanto permitían una gran simplificación de la teoría y de la práctica. El uso prolongado de estos procedimientos ha terminado por hacer que la normalidad sea frecuentemente considerada, no como una suposición cómoda, sino como una propiedad que efectivamente poseen los datos estadísticos. Sin embargo, el análisis de grandes muestras de datos (Romanowski y Green 1965) muestra que, aún en aquellos casos donde la parte central de la distribución muestra la clásica forma de campana, las "colas" suelen ser mucho más pesadas que lo que correspondería a una normal. Por otra parte, los procedimientos adecuados para la distribución normal, suelen ser pésimos aún para distribuciones aproximadamente normales con colas más pesadas.

Esta realidad ha sido asumida claramente en lo referente a tests, con el desarrollo, desde fines de la década del 30, de tests "no paramétricos" que tienen un error de tipo I constante, y una potencia no muy inferior a la óptima para una amplia gama

de distribuciones. En el problema de estimación, fué J. W. Tukey quien, desde mediados de la década del 40 (Tukey, 1946) comenzó a predicar el uso de estimadores "robustos", es decir, que tengan una razonable eficiencia para la distribución normal y para distribuciones aproximadamente normales-contrariamente a la media y la varianza, que siendo estimadores óptimos de locación y escala para la normal, pueden ser muy deficientes para distribuciones próximas a ésta (una definición rigurosa de robustez puede verse en (Hampel, 1971)). Tukey propuso el uso de medias y varianzas "podadas" y "winsorizadas" (ver (1.1.1) y (1.1.2)) y estudió sistemáticamente su comportamiento para distribuciones de tipo "normal contaminada".

La aceptación de estas ideas en el trabajo estadístico fue mucho más lenta que la de los tests no paramétricos, recién en (Tukey 1960) se publica una exposición más accesible y difundida.

El desarrollo del tema en la década del 60 ha sido mucho más rápido. Para el problema de locación y/o escala, a fines de la década se distinguen tres tipos fundamentales de estimadores robustos (para referencias generales, ver (Huber 1972) y (Andrews y otros, 1972)):

a) Combinaciones lineales de estadísticos de orden, o "L-estimadores". Sean $x_{(1)}, \dots, x_{(n)}$ los valores ordenados de la muestra. Un L-estimador t es de la forma:

$$t = \sum_{i=1}^n a_i x_{(i)}$$

con $\sum_i a_i = 1$.

Ejemplos son las "medias α -podadas" ($0 < \alpha < 1$), donde

$$a_i = \begin{cases} 1/(n-2[n\alpha]) & \text{si } [n\alpha] < i < n-[n\alpha] \\ 0 & \text{si no} \end{cases} \quad (1.1.1)$$

y las "medidas α -winsorizadas" donde

$$a_i = \begin{cases} 1/n & \text{si } [n\alpha] < i < n-[n\alpha] \\ [n\alpha]/n & \text{si } i = [n\alpha] \text{ ó } i = n-[n\alpha] \\ 0 & \text{los demás } i. \end{cases} \quad (1.1.2)$$

Análogamente se pueden definir varianzas "podadas" o "winsorizadas". Las propiedades asintóticas de este tipo de estimadores han sido estudiadas, entre otros, por Chernoff, Gastwirth y Johns (1967).

b) Estimadores basados en tests de rangos, o "R-estimadores". Hodges y Lehmann (1963) probaron que de cada test de rangos para locación se puede obtener un estimador de locación que tiene la misma eficiencia asintótica. El uso efectivo de estos estimadores parece en general muy limitado por dificultades de cálculo.

c) Estimadores de tipo de máxima verosimilitud, o "ii-estimadores". Si se tiene una muestra x_1, \dots, x_n de una distribución con densidad $g(x) = f((x-t)/s)$, los estimadores de máxima verosimilitud de los parámetros de locación y escala t y s son soluciones de las ecuaciones:

$$(1/n) \sum_{i=1}^n \psi_1((x_i-t)/s) = 0 \quad (1.1.3)$$

$$(1/n) \sum_{i=1}^n \psi_2((x_i - t)/s) = 1 \quad (1.1.4)$$

donde ψ_1 y ψ_2 se definen como

$$\psi_1(z) = -d \log f(z)/dz \quad \text{y} \quad \psi_2(z) = z \psi_1(z). \quad (1.1.5)$$

Huber (1964) estudió estimadores que se definen como soluciones de la forma (1.1.3)-(1.1.4), pero donde las ψ_i pueden ser cualesquiera. Para el caso en que se supone s conocido, Huber probó que para la familia de "normales ϵ -contaminadas" -es decir, distribuciones de la forma

$$F = (1-\epsilon)\phi + \epsilon H \quad (1.1.6)$$

donde ϕ es la distribución normal standard, y H puede ser cualquier distribución simétrica- el estimador de locación asintóticamente minimax es solución de (1.1.3) con $\psi_1(z) = \psi(z, k)$, donde

$$\psi(z, k) = \max(-k, \min(z, k)) \quad (1.1.7)$$

siendo k una constante positiva que depende de ϵ . Para el caso de s desconocido, Huber propone tres métodos, de los cuales el más practicable parece ser su "Propuesta 2", que consiste en resolver las ecuaciones (1.1.3)-(1.1.4) con $\psi_1(z) = \psi(z, k)$ y $\psi_2(z) = \psi_1^2(z)/\beta$ donde $\beta = E_{\phi} \psi_1^2(x)$ es una constante elegida para que en caso normal s^2 sea un estimador consistente de la varianza.

1.2.- El caso multivariado

Sea f una densidad radial en R^m (es decir, $f(\underline{x})$ depende sólo de $|\underline{x}|$). Se tiene una muestra $\underline{x}_1, \dots, \underline{x}_n$, donde para cierta matriz no singular \underline{S} y cierto vector \underline{t} , las variables $\underline{S}^{-1}(\underline{x}_i - \underline{t})$ ($i = 1, \dots, n$) tienen densidad f : e interesa estimar los parámetros de locación \underline{t} y de dispersión $\underline{V} = \underline{SS}'$ (nótese que si la variable \underline{y} tiene distribución radial, y si $\underline{SS}' = \underline{TT}'$, entonces la distribución de \underline{Sy} es igual a la de \underline{Ty} , de modo que en general \underline{S} no es identificable).

El limitarse a muestras de distribuciones radiales, aunque algo simplista, tiene algunas ventajas, en particular respecto a "qué es lo que se está estimando". Si g es una densidad obtenida de una densidad radial por una transformación afín, el vector de locación que se quiere estimar es el centro de simetría de g , o sea, el \underline{t} tal que $g(\underline{x}-\underline{t})$ es una función simétrica de \underline{x} , y la matriz de dispersión es una \underline{V} tal que, si $\underline{V} = \underline{SS}'$, sea $g(\underline{S}^{-1}(\underline{x}-\underline{t}))$ una función radial de \underline{x} .

Esta \underline{V} está, definida a menos de un múltiplo escalar. Nótese que si la g tiene momentos de 2° orden, cualquier \underline{V} de esta clase es un múltiplo escalar de la matriz de covarianzas. Hay en Análisis Multivariado varias situaciones en las que interesa la estimación "robusta" del vector de locación y de la matriz de dispersión a menos de un múltiplo escalar (o algún múltiplo escalar de la matriz de covarianzas cuando ésta existe). La más obvia es cuando simplemente interesa estimar locación, y la estimación de la matriz de escala aparece sólo como un paso auxiliar

para obtener un estimador invariante (como en la "Propuesta 2" de Huber). Otros casos son (ver Anderson (1953)):

a) Discriminación: Si se tienen n distribuciones con densidades $f_i(\underline{x}) = f(\underline{S}^{-1}(\underline{x} - \underline{t}_i))$ ($i = 1, \dots, n$) -- es decir, las f_i difieren sólo en locación- y si $f(\underline{x})$ es una función decreciente de $|\underline{x}|$, la regla de discriminación óptima en el sentido de Bayes -cuando se conocen \underline{S} y los \underline{t}_i - consiste en asignar el punto \underline{x} a la población i_0 tal que minimice

$$c_i(\underline{x} - \underline{t}_i)' \underline{V}^{-1}(\underline{x} - \underline{t}_i) \tag{1.2.1}$$

donde $\underline{V} = \underline{S}\underline{S}'$, y las constantes c_i dependen de las probabilidades a priori y de los costos de clasificación errónea. Si f es normal (o en general si tiene momentos de 2º orden) el resultado es la conocida regla en que los \underline{t}_i son las medias y la \underline{V} (cualquier múltiplo escalar de) la matriz de covarianzas común de las f_i . En el caso más realista en que los parámetros no se conocen sino que deben ser estimados a partir de muestras, se obtienen una regla asintóticamente óptima reemplazando en (1.2.1) los parámetros desconocidos por sus estimadores de máxima verosimilitud (Hoel y Peterson, 1949). De modo, que en el caso normal es conveniente estimar los \underline{t}_i y la \underline{V} respectivamente con las medias y la matriz de covarianzas muestrales. Pero este procedimiento puede ser pésimo si la f no es exactamente, sino sólo aproximadamente normal, y por lo tanto para ser más realista conviene utilizar estimadores robustos de \underline{V} y de los \underline{t}_i .

b) Componentes principales: Sea \underline{x} un vector aleatorio con media \underline{t} y matriz de covarianzas \underline{V} . Para cada variedad lineal H , sea \underline{T}_H el operador de proyección ortogonal sobre H . Entonces la variedad lineal de dimensión $p \leq m$ que minimiza $E |\underline{x} - \underline{T}_H \underline{x}|^2$ es la que pasa por \underline{t} y tiene la dirección de los autovectores correspondientes a los p mayores autovalores de \underline{V} (de modo que aquí también basta conocer cualquier múltiplo escalar de \underline{V}).

El procedimiento tradicional para estimar H consiste en estimar \underline{t} y \underline{V} con las medias y covarianzas muestrales, pero por los mismos motivos que para discriminación puede ser más prudente utilizar estimadores robustos de los parámetros.

c) Detección de observaciones anómalas ("outliers"): Si la muestra m -variada $\underline{x}_1, \dots, \underline{x}_n$ proviene de una distribución $N(\underline{t}, \underline{V})$, las variables $v_i = (\underline{x}_i - \underline{t})' \underline{V}^{-1} (\underline{x}_i - \underline{t})$ ($i = 1, \dots, n$) son una muestra de una distribución χ^2 con m grados de libertad. Si \underline{t} y \underline{V} se estiman de la muestra -y ésta no es muy pequeña- es de esperar que las v_i ordenadas se comporten aproximadamente como estadísticos de orden de χ^2 . Si se supone que la muestra original es una mezcla de una mayoría de observaciones "buenas" distribuidas normalmente, con algunas observaciones "anómalas" con una distribución más dispersa, aquel hecho puede utilizarse para detectar estas últimas (ver referencias en (Gnanadesikan y Kattenring, 1972)). Dado que se supone explícitamente que la distribución es sólo aproximadamente normal, es naturalmente deseable utilizar estimadores de \underline{t} y \underline{V} que sean menos sensibles a las observaciones "anómalas" que las media y las covarianzas. Aquí también basta con un múltiplo de \underline{V} .

Los métodos robustos multivariados propuestos hasta ahora en la literatura corresponden principalmente al caso en que se desea estimar solamente la locación. Ellos son:

a) Aplicación de un estimador robusto de locación univariado, coordinada a coordinada (Bickel, 1964). En particular son de este tipo los R -estimadores correspondientes a test no paramétricos de locación multivariada (Sen y Puri, 1971). Estos estimadores son invariantes por traslación, pero no por transformaciones afines cualesquiera.

a) Un análogo multivariado de los L -estimadores está dado por una propuesta de J. Tukey mencionada en (Huber, 1972). El procedimiento -que Tukey llama "polar" la muestra, y que en el caso univariado se reduce a lo que el mismo Tukey llamó "podar" la muestra- consiste en eliminar los puntos extremos de ésta, y repetir la operación un número fijo de veces, o hasta eliminar una cantidad prefijada de puntos. luego se toma la media y/o la matriz de covarianzas de los puntos que quedan. Este método es invariante por transformaciones afines pero computacionalmente parece muy complicado para su aplicación práctica, y casi nada se sabe sobre sus propiedades teóricas.

c) Gentleman (1965) propuso un L -estimador definido como el \underline{t} que minimiza $\sum_{i=1}^n |\underline{x}_i - \underline{t}|^p$ (el caso $p = 2$ corresponde a la media muestral).

d) Gnanadesikan y Kettenring (1972) proponen algunos métodos heurísticos, basados en la eliminación o modificación de "observaciones extremas", pero no hacen un estudio sistemático de sus propiedades.

1.3.- Plan de la obra

El objetivo de esta tesis es estudiar la estimación simultánea de localización y dispersión por medio de ψ -estimadores que se definen a continuación. En la situación descrita al comienzo de la sección anterior, las ecuaciones para los estimadores de máxima verosimilitud de \underline{t} y \underline{S} son

$$(1/n) \sum_i \psi_1 (\underline{S}^{-1}(\underline{x}_i - \underline{t})) = \underline{0} \quad (1.3.1)$$

$$(1/n) \sum_i \psi_2 (\underline{S}^{-1}(\underline{x}_i - \underline{t})) = \underline{I} \quad (1.3.2)$$

donde las funciones $\psi_1: \mathbb{R}^m \rightarrow \mathbb{R}^m$ $\psi_2: \mathbb{R}^{m \times m} \rightarrow \mathbb{R}^{m \times m}$ se definen como

$$\psi_1(\underline{x}) = -\dot{L}(\log f(\underline{x})) \quad (1.3.3)$$

$$\psi_2(\underline{x}) = \psi_1(\underline{x}) \underline{x}' \quad (1.3.4)$$

Para f radial pongamos $f(\underline{x}) = h(|\underline{x}|)$, $\psi_1(r) = -h'(r)/h(r)$ ($r \geq 0$). $\underline{v} = \underline{S}\underline{S}'$, $d_i^2 = d^2(\underline{x}_i, \underline{t}, \underline{v}^{-1}) = |\underline{S}^{-1}(\underline{x}_i - \underline{t})|^2$. Entonces $\psi_1(\underline{x}) = \psi_1(|\underline{x}|) \underline{x}/|\underline{x}|$ y (1.3.1)-(1.3.2) se transforman respectivamente en

$$(1/n) \sum_i u_1(d_i) (\underline{x}_i - \underline{t}) = \underline{0} \quad (1.3.5)$$

$$(1/n) \sum_i u_2(d_i^2) (\underline{x}_i - \underline{t})(\underline{x}_i - \underline{t})' = \underline{v} \quad (1.3.6)$$

donde para $r > 0$

$$u_1(r) = \psi_1(r)/r \quad (1.3.7)$$

$$u_2(r^2) = u_1(r) , \quad (1.3.8)$$

Notar que si f es normal, u_1 y u_2 son constantes, y por lo tanto \underline{t} y \underline{V} son respectivamente la media y la matriz de covarianzas. Las funciones u_1 y u_2 podrían considerarse como medidas del peso de las observaciones en las ecuaciones respectivas.

Los M -estimadores que se estudiarán serán soluciones de ecuaciones de la forma (1.3.5)-(1.3.6) donde las u_j pueden ser funciones cualesquiera. En el capítulo siguiente se prueba la existencia y unicidad de las soluciones de (1.3.5)-(1.3.6) para una clase bastante general de funciones u_1 y u_2 . En el capítulo 3 se prueba la consistencia y la normalidad asintótica de los estimadores, se examinan otras cualidades de robustez y se compara el comportamiento asintótico de algunos estimadores particulares para varias distribuciones; y por último en el capítulo 4 se muestra un procedimiento para la resolución numérica de las ecuaciones, y se estudia mediante simulación el comportamiento de los estimadores para muestras finitas.

Deberé dejar para más adelante el estudio de la aplicación de estos métodos para Discriminación y Componentes Principales, que constituían mi interés inicial. La teoría asintótica resulta difícil, y la simulación para muestras pequeñas demandará experimentos cuidadosamente planificados y bastante tiempo.

2 - EXISTENCIA Y UNICIDAD

2.1.- Condiciones generales

En este capítulo se considerará la existencia y unicidad de soluciones de ecuaciones de la forma

$$E_P u_1(d(\underline{x}, \underline{t}, \underline{V})^{-1}) (\underline{x} - \underline{t}) = 0 \quad (2.1.1)$$

$$E_P u_2(d^2(\underline{x}, \underline{t}, \underline{V})^{-1}) (\underline{x} - \underline{t})(\underline{x} - \underline{t})' = \underline{y} \quad (2.1.2)$$

donde P es una medida de probabilidad cualquiera en R^m , y las u_i son funciones de R_+ en R_+ . Si $\{\underline{x}_1, \dots, \underline{x}_n\}$ es una muestra, y P es la correspondiente distribución empírica -es decir, la medida atómica tal que $P(\{\underline{x}_i\}) = 1/n$ ($i = 1, \dots, n$)- entonces (2.1.1)-(2.1.2) se transforman en (1.3.5)-(1.3.6), o sea en las ecuaciones que definen los estimadores. Si P es la distribución subyacente de los datos, entonces (2.1.1)-(2.1.2) definen los parámetros a estimar.

Aquí conviene remarcar que el espacio de los parámetros es

$\theta = \{(\underline{t}, \underline{V})\} = \theta_1 \times \theta_2$, con $\theta_1 = R^m$ y $\theta_2 = \{\text{conjunto de las matrices definidas positivas simétricas de } n \times n\}$

Las soluciones de las ecuaciones son obviamente invariantes por transformaciones afines: si se define la medida P° como $P^\circ(A) = P\{\underline{x} | \underline{TX} + \underline{a} \in A\}$, y $(\underline{t}, \underline{V})$ es solución de (2.1.1)-(2.1.2),

entonces $\underline{t}^\circ = \underline{Tt} + \underline{a}$ y $\underline{V}^\circ = \underline{TVT}'$ son soluciones del mismo sistema cuando P es reemplazada por P° .

Si \underline{S} es cualquier matriz tal que

$$\underline{V} = \underline{SS}' \quad (2.1.3)$$

entonces (2.1.1)-(2.1.2) se pueden escribir

$$E u_1 (|\underline{S}^{-1}(\underline{x}-\underline{t})|) \underline{S}^{-1}(\underline{x}-\underline{t}) = 0 \quad (2.1.4)$$

$$E u_2 (|\underline{S}^{-1}(\underline{x}-\underline{t})|^2)(\underline{S}^{-1}(\underline{x}-\underline{t}))(\underline{S}^{-1}(\underline{x}-\underline{t}))' = \underline{1}. \quad (2.1.5)$$

ecuaciones que determinan \underline{t} y \underline{V} a través de (2.1.3). Si se quiere definir unívocamente \underline{S} , se puede agregar a (2.1.4)-(2.1.5) la condición de que \underline{S} sea simétrica definida positiva.

Para facilitar la notación y la comparación con el caso univariado, se define $\psi_i(s) = s u_i(s)$ ($i = 1, 2$). Se supondrán las siguientes condiciones sobre las u_i :

- (A) Las u_i son no crecientes, y $u_i(0) < \infty$ ($i = 1, 2$).
- (B) Las ψ_i son continuas y acotadas. Sean $K_i = \sup \{\psi_i(s) | s > 0\}$
- (C) ψ_2 es no decreciente, y es estrictamente creciente en el intervalo donde $\psi_2 < K_2$ (o sea, $\psi_2(a) = \psi_2(b) \rightarrow \psi_2(a) = K_2$ si $a \neq b$)
- (D) Existe un s_0 tal que $\psi_2(s_0^2) > m$, y que $\psi_1(s) > 0$ para $0 \leq s \leq s_0$ (y por lo tanto $K_2 > m$).
- (E) Existe un $\epsilon > 0$ tal que para todo hiperplano H es
 $\Gamma(H) \leq 1 - m/K_2 - \epsilon$.

Ejemplos: Es fácil verificar (A)-(D) para dos estimadores que se utilizarán más adelante:

i) Estimador de tipo "Huber". Sean k_1 y k_2 constantes positivas, $\psi_1(s) = \psi(s, k_1)$ (ver (1.1.7)) de modo que $K_1 = k_1$. Sean $\beta = E_{\phi}(\psi^2 |x|^2, k_2^2)/m$, donde ϕ es la distribución m -variada $N(\underline{0}, \underline{I})$, y $\psi_2(s) = \psi(s, k_2^2)/\beta$.

Esta elección, igual que la hecha por Huber (1964) tiene por objeto "calibrar" la matriz de escala para que coincida exactamente con la matriz de covarianza en el caso normal.

Aquí $K_2 = k_2^2/\beta$ y es fácil verificar que $K_2 > m$, de modo que (D) se cumple evidentemente. En la "Propuesta 2" de Huber (1964) para $m = 1$, es $k_1 = k_2$. En cuanto a (A), (B) y (C), son obvias.

Nótese que el caso $k_1 = k_2 = \infty$ corresponde a la media y matriz de covarianzas.

ii) Estimador de Máxima Verosimilitud para la distribución de Student.

Sea \underline{x} un vector aleatorio con distribución $N(\underline{0}, \underline{I})$ en R^m , y z una variable real independiente de \underline{x} , con distribución χ^2 con n grados de libertad. Se llama distribución de Student m -variada radial con n grados de libertad a la distribución del vector $\underline{y} = \underline{x} / (z/n)^{1/2}$ que tiene densidad

$$f_{m,n}(\underline{y}) = C_{m,n} / (n + |\underline{y}|^2)^{(m+n)/2}$$

donde $C_{m,n}$ es una constante. Obviamente las marginales de $f_{m,n}$ también son Student. El caso $n = 1$ es la distribución de Cauchy.

Para el correspondiente estimador de máxima verosimilitud, es

$$\psi(\underline{x}) = (m+n) \underline{x} / (n + |\underline{x}|^2)$$

y por lo tanto

$$\psi_1(s) = (m+n) s / (n + s^2)$$

$$\psi_2(s) = (m+n) s / (n + s)$$

Las condiciones (A) - (D) son fáciles de verificar. Nótese que cuando $n \rightarrow \infty$, $f_{m,n} \rightarrow N(\underline{0}, \underline{I})$, y correspondientemente u_1 y u_2 tienden a sendas constantes, como era de esperar. La ecuación para escala podría modificarse recalibrando ψ_2 -multiplicándola por una constante adecuada- para que \underline{V} coincida asintóticamente con la matriz de covarianza en el caso normal, pero esto no es imprescindible.

Observación: La condición (E) establece una relación entre las ψ_i y P . Como se verá en continuación, esta condición resulta esencial para la existencia y la consistencia de los estimadores. Generalmente se supone que los datos provienen de una distribución absolutamente continua, de modo que (E) se

verificará siempre para la distribución subyacente. Puede en cambio haber inconvenientes para muestras de tamaño n finito como en R^m cualquier conjunto de n puntos forma un hiperplano H , con la distribución empírica P será $P(H) = m/n$ y por lo tanto para que se pueda cumplir (E) deberá ser $m/n \ll 1 - m/K_2$. Para muestras no muy grandes de dimensión elevada, esto impone una restricción sobre el estimador, pues K_2 deberá ser suficientemente grande, pero como se verá en el la sección 3.3 esto implicará una pérdida de robustez.

NOTA BENE: Desde ahora se supondrá siempre que se cumplen las condiciones (A) (B)-(C)-(D). Además, como la ecuación (2.1.1) no cambia si se multiplica u_1 por una constante, se supondrá que $u_1(0) = 1$.

En el resto del Capítulo 2 se supondrá que F cumple (E).

2.2.- Existencia de las soluciones

Para simplificar se escribirá en lo sucesivo

$$F(\underline{t}, \underline{V}) = E u_2(d^2(\underline{x}, \underline{t} \underline{V}^{-1})) (\underline{x} - \underline{t})(\underline{x} - \underline{t})' \quad (2.2.1)$$

de modo que (2.1.2) se escribe

$$F(\underline{t} \underline{V}) = \underline{V} \quad (2.2.2)$$

Para cada \underline{t} , F es una función de θ_2 en θ_2 .

Se tratará primero la ecuación (2.2.2) separadamente.
 Conviene separar un resultado para uso posterior.

Lema 1: Para cada $A > 0$, existe un $r_0 > 0$ tal que,
cgs $\underline{t} \in \mathbb{B}_A$ y cgs $r \geq r_0$, es $F(\underline{t}, rI) \leq rI$.

Demostración: Por ser ψ_2 acotada, resulta

$$\lim_{n \rightarrow \infty} \sup_{|\underline{z}|=1} \sup_{|\underline{t}| \leq A} \underline{z}' F(\underline{t}, rI) \underline{z} / r = 0$$

de modo que hay un r_0 tal que, cgs $\underline{t} \in \mathbb{B}_A$ y cgs $r \geq r_0$ es

$$\underline{z}' F(\underline{t}, rI) \underline{z} \leq r \underline{z}' I \underline{z} \quad \text{cgs } \underline{z}$$

o sea, $F(\underline{t}, rI) \leq rI$.

QED

Teorema 1: Para cada \underline{t} existe una única solución $\underline{V} = \underline{V}(\underline{t})$
de la ecuación $F(\underline{t}, \underline{V}) = \underline{V}$.

Demostración: No hay pérdida de generalidad en suponer
 $\underline{t} = \underline{0}$. Para simplificar se escribirá en la demostración $F(\underline{V})$
 en lugar de $F(\underline{0}, \underline{V})$.

Es fácil ver que si \underline{U} es definida positiva, $F(\underline{U})$ también
 lo es, pues si no, habría un $\underline{z} \neq \underline{0}$ tal que

$$0 = \underline{z}' F(\underline{U}) \underline{z} = E u_2 (\underline{z}' \underline{U}^{-1} \underline{z}) (\underline{z}' \underline{x})^2$$

y como u_2 es positiva por (C) debería ser $P(\underline{z}'\underline{x} = 0) = 1$, lo que contradice (E). Usando ahora el Lema 1 con $\underline{t} = \underline{0}$, sea $\underline{V}_0 = r_0 \underline{I}$, de modo que $F(\underline{V}_0) \leq \underline{V}_0$ y definamos recursivamente $\underline{V}_{n+1} = F(\underline{V}_n)$, cosa que es posible pues por el razonamiento anterior cada \underline{V}_n es no singular.

Si $\underline{U}, \underline{U}' \in \Theta_2$ y $\underline{U} \leq \underline{U}'$, es $\underline{U}^{-1} \leq \underline{U}'^{-1}$, y por ser u_2 no creciente resulta $F(\underline{U}) \leq F(\underline{U}')$ de modo que $\underline{V}_{n+1} \leq \underline{V}_n$. Por lo tanto existe el $\lim_{n \rightarrow \infty} \underline{V}_n$. Este límite, que llamaremos \underline{V} , será la solución buscada si se prueba que no es singular. Para ello supongamos que existe \underline{z} tal que $\underline{z}'\underline{V}\underline{z} = 0$. Sea $H = \{\underline{x} \mid \underline{z}'\underline{x} = 0\}$. Sea $b > 0$ cualquiera, por (D) existe s_b tal que $s \geq s_b$ implica $\psi_2(s) \geq K_2 - b$. Sean los conjuntos $C_{bn} = \{\underline{x} \mid \underline{x}'\underline{V}_n^{-1}\underline{x} \leq s_b\}$ y $C_b = \bigcap_n C_{bn}$. Obviamente $C_{b,n+1} \subseteq C_{b,n}$. Por otra parte, $C_b \subseteq H$. En efecto sea $\underline{x} \in C_b$, poniendo $\underline{y}_n = \underline{S}_n \underline{S}_n^{-1} \underline{x} = \underline{S}_n^{-1} \underline{x}$, resulta $|\underline{y}_n|^2 = \underline{x}'\underline{V}_n^{-1}\underline{x} \leq s_b$, por lo tanto $|\underline{z}'\underline{x}| = |(\underline{S}_n \underline{z})'\underline{y}_n| \leq s_b |\underline{z}'\underline{V}_n \underline{z}|^{1/2} \rightarrow 0$ cuando $n \rightarrow \infty$ de modo que $\underline{x} \in H$.

En consecuencia, dado $c > 0$ existe n_{bc} tal que $n \geq n_{bc}$ implica $P(C_{bn}^c) \geq 1 - P(H) - c$. Además

$$\begin{aligned} \underline{V}_n &\geq F(\underline{V}_n) \geq E I(\underline{x} \in C_{bn}^c) u_2(\underline{x}'\underline{V}_n^{-1}\underline{x}) \underline{xx}' \\ &\geq E I(\underline{x} \in C_{bn}^c) [(K_2 - b) / (\underline{x}'\underline{V}_n^{-1}\underline{x})] \underline{xx}'. \end{aligned}$$

multiplicando la desigualdad anterior a derecha e izquierda por $(\underline{S}_n^{-1})'$ y por \underline{S}_n^{-1} respectivamente, resulta

$$\underline{I} \geq (K_2 - b) E I(\underline{x} \in C'_{bn})(\underline{S}_n^{-1} \underline{x} / |\underline{S}_n^{-1} \underline{x}|)(\underline{S}_n^{-1} \underline{x} / |\underline{S}_n^{-1} \underline{x}|)',$$

y tomando traza resulta que cuando $n \geq n_{bc}$

$$m \geq (K_2 - b)P(C'_{bn}) \geq (K_2 - b)(1 - P(H)) \quad c)$$

y como esto vale cts $b, c > 0$, se obtiene $m/K_2 \geq 1 - P(H)$, lo que contradice (E), demostrándose así la existencia de la solución.

No hay inconveniente en suponer -mediante un cambio de coordenadas- que \underline{I} es solución, o sea que $F(\underline{I}) = \underline{I}$. Para demostrar la unicidad se probará que en estas condiciones $\underline{U} \neq \underline{I}$ implica $|F(\underline{U}) - \underline{I}| < |\underline{U} - \underline{I}|$. Para ello sean respectivamente $a_1 \leq \dots \leq a_m$ y $b_1 \leq \dots \leq b_m$ los autovalores de \underline{U} y de $F(\underline{U})$. Bastará probar que $a_1 < 1$ implica $b_1 > a_1$ y que $a_m > 1$ implica $b_m < a_m$. Se probará sólo la primera implicación, pues la demostración de la segunda es exactamente igual.

Sea $\underline{R} = E \psi_2(\underline{x}'\underline{x} / a_1) \underline{xx}' / (\underline{x}'\underline{x})$. Como $\underline{x}'\underline{U}^{-1} \underline{x} \leq \underline{x}'\underline{x}/a_1$ es $F(\underline{U}) \geq a_1 \underline{R}$. De modo que basta probar que el menor autovalor de \underline{R} es mayor que 1, o equivalentemente, que $\underline{z}' \underline{R} \underline{z} > \underline{z}' \underline{I} \underline{z}$ cts $\underline{z} \neq \underline{0}$. Como $F(\underline{I}) = \underline{I}$, ello equivale a probar que

$$0 < \underline{z}'(\underline{R}-\underline{I})\underline{z} = E (\psi_2(\underline{x}'\underline{x} / a_1) - \psi_2(\underline{x}'\underline{x})) (\underline{z}'\underline{x})^2 / (\underline{x}'\underline{x}).$$

Por ser $a_1 < 1$ y ψ_2 no decreciente el integrando es no negativo. Supongamos que para algún $\underline{z} \neq \underline{0}$ fuera el último miembro nulo. Por (C) $\psi_2(\underline{x}'\underline{x} / a_1) = \psi_2(\underline{x}'\underline{x}) = K_2$ implica $\psi_2(\underline{x}'\underline{x}) = K_2$.

Por lo tanto $F(\psi_2(\underline{x}'\underline{x}) = K_2) + P(\underline{z}'\underline{x} = 0) = 1$. El segundo sumando es menor que $1 - m/K_2$ por (E), tomando traza en la ecuación $F(\underline{I}) = \underline{I}$ se deduce que el primer sumando es $\leq m/K_2$. Esta contradicción completa el teorema.

Lema 2: La función $\underline{V}(\underline{t})$ tiene las siguientes propiedades:

- i) Existe A_1 tal que para todo \underline{t} : $|\underline{V}(\underline{t})^{-1}| \leq A_1^2$
- ii) Existe A_2 tal que para todo \underline{t} $\underline{t}' \underline{V}(\underline{t})^{-1} \underline{t} \leq A_2^2$
- iii) $\lim_{\underline{t} \rightarrow \infty} |\underline{V}(\underline{t})| / |\underline{t}|^2 = A_3 < \infty$
- iv) $\underline{V}(\underline{t})$ es una función continua de \mathbb{R}^m en Θ_2 . (tomando en Θ_2 la topología inducida por $\mathbb{R}^{m \times m}$).

Demostración:

- i) De (E) se deduce fácilmente (por absurdo) la existencia de $c > 0$ tal que para todo $\underline{z} \in S_1$ y todo \underline{t}

$$P\{x \mid |\underline{z}'(\underline{x}-\underline{t})| \leq c\} \leq 1 - m/K_2 - a/2 \quad (2.2.3)$$

donde a es la constante que aparece en (E). Sea $b > 0$ tq

$$(K_2 - b) (m/K_2 + a/2) > m, \text{ y } A_1 > 0 \text{ tq } \psi_2[(cA_1)^2] > K_2 - b$$

Dado ahora cualquier \underline{t} sea $e = e(\underline{t})$ el menor autovalor de $\underline{V} = \underline{V}(\underline{t})$, y $\underline{z} = \underline{z}(\underline{t})$ el correspondiente autovector normalizado. Se probará que $1/e \leq A_1^2$, y esto probará (i) pues $1/e = |\underline{V}^{-1}|$.

Supongamos entonces que $1/e > A_1^2$, tomando traza en (2.1.5) se obtiene

$$m = E \psi_2(d^2(\underline{x}, \underline{t}, \underline{V}^{-1})) \quad (2.2.4)$$

y como $d^2(\underline{x}, \underline{t}, \underline{V}^{-1}) \geq |\underline{z}'(\underline{x}-\underline{t})|^2 / e$, resulta

$$\begin{aligned} m &\geq E I(|\underline{z}'(\underline{x}-\underline{t})| > c) \psi_2(|\underline{z}'(\underline{x}-\underline{t})|^2 / e) \geq \\ &\geq P\{\underline{x} \mid |\underline{z}'(\underline{x}-\underline{t})| > c\} \psi_2[(cA_1)^2] > (m/K_2 + a/2)(K_2 - b) > m. \end{aligned}$$

ii) Sea $\underline{s} = \underline{s}(\underline{t})$ cualquier raíz de $\underline{V} = \underline{V}(\underline{t})$. Entonces (i) equivale a $|\underline{s}(\underline{t})^{-1}| \leq A_1$ y (2.2.4) se puede escribir

$$m = E \psi_2(|\underline{s}^{-1}\underline{x} - \underline{s}^{-1}\underline{t}|^2) \quad (2.2.4')$$

Dado que para cada \underline{x} se cumple

$$\lim_{|\underline{s}| \rightarrow \infty} \inf_{|\underline{z}| \leq A_1} \psi_2(|\underline{z}'\underline{x} - \underline{s}'|) = K_2$$

resulta que por el Teorema de Convergencia de Lebesgue y por ser $K_2 > m$ existen b y A_2 positivos tal que $|\underline{s}| > A_2$ implica

$$E \inf_{|\underline{z}| \leq A_1} \psi_2(|\underline{z}'\underline{x} - \underline{s}'|) \geq m + 2b \quad (2.2.5)$$

y por lo tanto, si para algún \underline{t} fuera $|\underline{s}(\underline{t})^{-1}\underline{t}| > A_2$ poniendo en (2.2.5) $\underline{z} = \underline{s}(\underline{t})^{-1}$ y $\underline{s} = \underline{s}(\underline{t})^{-1}\underline{t}$, se obtendría una contradicción con (2.2.4').

iii) Sean ahora $c(\underline{t})$ el mayor autovalor de $\underline{V}(\underline{t})$ ($= |\underline{V}(\underline{t})|$) y $\underline{z}(\underline{t})$ el correspondiente autovector normalizado, de modo que con \underline{v} se cumple $\underline{v}' \underline{V}(\underline{t})^{-1} \underline{v} \geq |\underline{v}|^2 / c(\underline{t})$. Aplicando esto en (2.1.2) queda

$$\begin{aligned} \underline{v} &= E u_2(d^2(\underline{x}, \underline{t}, \underline{V}(\underline{t})^{-1})) (\underline{x} - \underline{t})(\underline{x} - \underline{t})' \leq \\ &\leq E u_2(|\underline{x} - \underline{t}|^2 / c(\underline{t})) (\underline{x} - \underline{t})(\underline{x} - \underline{t})' \end{aligned}$$

multiplcando a izquierda y derecha por \underline{z}' y por \underline{z} respectivamente, y aplicando la desigualdad de Cauchy-Schwarz.

$$1 = \underline{z}(\underline{t})' \underline{V}(\underline{t}) \underline{z}(\underline{t}) / c(\underline{t}) \leq E \psi_2[|\underline{x} - \underline{t}|^2 / c(\underline{t})] \quad (2.2.6)$$

De (ii) se deduce que $A_2^2 \geq |\underline{t}|^2 / c(\underline{t})$, de modo que $\liminf_{\underline{t} \rightarrow \infty} c(\underline{t}) = \infty$, por lo tanto para que (2.2.6) valga debe ser $\liminf_{\underline{t} \rightarrow \infty} |\underline{t}|^2 / c(\underline{t}) > 0$, lo que equivale a la tesis.

iv) Es fácil ver que cuando \underline{t} recorre un compacto de \mathbb{R}^m , $\underline{V}(\underline{t})$ se mantiene en un compacto de θ_2 , o sea, $|\underline{V}(\underline{t})|$ y $|\underline{V}(\underline{t})^{-1}|$ permanecen acotadas. Lo segundo es consecuencia de (i). Para probar lo primero, dado $r > 0$, existe por el Lema 1 un r tal que $F(\underline{t}, r\underline{I}) \leq r\underline{I}$ con $\underline{t} \in \mathcal{B}_r$. Procediendo recursivamente como al principio del Teorema 1, resulta que para todo $\underline{t} \in \mathcal{B}_r$ es $\underline{V}(\underline{t}) \leq r\underline{I}$, y a fortiori $|\underline{V}(\underline{t})| \leq r$.

Sea ahora una sucesión $\underline{t}_n \rightarrow \underline{t}_0$, hay que probar que $\underline{V}(\underline{t}_n) \rightarrow \underline{V}(\underline{t}_0)$. Sea $\{n'\}$ cualquier subsucesión, como los $\{\underline{t}_{n'}\}$ están acotados, las $\underline{V}(\underline{t}_{n'})$ están en un compacto, de modo que hay una subsucesión $\{n''\} \subseteq \{n'\}$ y una $\underline{V}_0 \in \theta_2$ tales que $\underline{V}(\underline{t}_{n''}) \rightarrow \underline{V}_0$. Como la función $F(\underline{t}, \underline{V})$ de θ en θ_2 es continua, y $F(\underline{t}_{n''}, \underline{V}(\underline{t}_{n''})) = \underline{I}$ para todo n'' , resulta tomando límite que $F(\underline{t}_0, \underline{V}_0) = \underline{I}$, o sea, que $\underline{V}_0 = \underline{V}(\underline{t}_0)$. QED

Teorema 2: Existe una solución $(\underline{t}_0, \underline{V}_0)$ del sistema (2.1.1)-(2.1.2). Además, \underline{t}_0 pertenece a la cápsula convexa del soporte de P.

Demostración. Hay que probar la existencia de un \underline{t}_0 tal que

$$E \ u_1(d(\underline{x}, \underline{t}_0, \underline{V}(\underline{t}_0)^{-1})) (\underline{x} - \underline{t}_0) = \underline{0}. \quad (2.2.7)$$

Sea entonces $\underline{S}(\underline{t})$ la raíz cuadrada simétrica definida positiva de $\underline{V}(\underline{t})$, por el Lema 2 (iv) es $\underline{S}(\underline{t})$ una función continua de \underline{t} . Pongamos para simplificar $\psi(\underline{x}) = :u_1(|\underline{x}|) \underline{x}$ y definamos la función $G: \mathbb{R}^n \rightarrow \mathbb{R}^n$:

$$G(\underline{t}) = \underline{t} + c (|\underline{t}| + 1) E \ \psi(\underline{S}(\underline{t})^{-1} (\underline{x} - \underline{t}))$$

donde c es una constante real a determinar luego convenientemente. La función G es continua, y ahora se probará que se puede hallar c, \underline{t}_0 para alguna bola cerrada B sea $G(B) \subseteq B$.

Por el teorema de punto fijo de Brower (Dunford-Schwartz, 1957) existirá entonces un $\underline{t}_0 \in B$ tal que $G(\underline{t}_0) = \underline{t}_0$, y que por lo tanto será solución de (2.2.7).

Notemos en primer lugar que $\lim_{|\underline{t}| \rightarrow \infty} \underline{t}' \underline{\Psi}(\underline{t})^{-1} \underline{t} / |\underline{t}| = 0$ por Lema 2(ii); y que por Lema 2(iii) es

$$\liminf_{|\underline{t}| \rightarrow \infty} \underline{t}' \underline{S}(\underline{t})^{-1} \underline{t} / |\underline{t}| \geq \liminf_{|\underline{t}| \rightarrow \infty} |\underline{t}| / |\underline{S}(\underline{t})| = b > 0$$

y por lo tanto

$$\begin{aligned} \limsup_{|\underline{t}| \rightarrow \infty} \underline{t}' E \psi(\underline{S}(\underline{t})^{-1} (\underline{x} - \underline{t})) / |\underline{t}| < \\ < -b \limsup_{|\underline{t}| \rightarrow \infty} E u_1(|\underline{S}(\underline{t})^{-1} (\underline{x} - \underline{t})|). \end{aligned} \quad (2.2.8)$$

Por (2.2.4') hay un $\eta > 0$ tal que para cada \underline{t} es $P(|\underline{S}(\underline{t})^{-1} (\underline{x} - \underline{t})| \leq s_0) \geq \eta$, donde s_0 es la constante que figura en (D) (basta tomar cualquier $\eta \leq 1 - m/\psi_2(s_0^2)$). Como $\psi_1(s_0) \rightarrow 0$, se deduce de (2.2.8) que

$$\limsup_{|\underline{t}| \rightarrow \infty} \underline{t}' E \psi(\underline{S}(\underline{t})^{-1} (\underline{x} - \underline{t})) / |\underline{t}| \leq -b \eta u_1(s_0) < 0. \quad (2.2.9)$$

Al primer miembro lo llamaremos $-c$ (con $c > 0$).

Por lo tanto, como $|\psi(\underline{x})| \leq K_1$ para cada \underline{x} , para cada c es

$$\begin{aligned} \limsup_{|\underline{t}| \rightarrow \infty} |\underline{t}' / |\underline{t}| + c E \psi(\underline{S}(\underline{t})^{-1} (\underline{x} - \underline{t}))|^2 &\leq \\ &\leq 1 + c^2 K_1^2 - 2c(b \eta u_1(s_0)). \end{aligned} \quad (2.2.10)$$

Tomando c bastante chico, el segundo miembro de (2.2.10) se hace < 1 , de modo que existen $d > 0$ y c tales que

$$\limsup_{\underline{t} \rightarrow \infty} |\underline{t} / |\underline{t}| + c E \psi(S(\underline{t})^{-1} (\underline{x} - \underline{t}))| \leq 1 - d \quad (2.2.11)$$

En consecuencia existe A tal que $|\underline{t}| \leq A$ implica $|G(\underline{t})| \leq A$, pues si no, existiría una sucesión $\{\underline{t}_n\}$ con $|\underline{t}_n| \leq n$ tal que $|G(\underline{t}_n)| > n$, como G es continua, debe ser $\limsup_n |\underline{t}_n| = \infty$, y además $|G(\underline{t}_n)| / |\underline{t}_n| > 1$, lo cual contradice (2.2.11). Por lo tanto, para algún A es $G(B_A) \subseteq B_A$, lo cual termina la demostración de la existencia.

Para probar la segunda afirmación del Teorema, sean C la cápsula convexa del soporte de P , $(\underline{t}_0, \underline{V}_0)$ una solución del sistema, y $b = \inf_{\underline{x} \in C} d^2(\underline{x}, \underline{t}_0; \underline{V}_0^{-1})$ de modo que por (2.2.4) es $m > \psi_2(b)$.

Si $\underline{t}_0 \notin C$, existen un hiperplano H que separa \underline{t}_0 y C , un vector \underline{z} ortogonal a H , y un $c > 0$ tales que $\underline{z}'(\underline{x} - \underline{t}) \geq c$ cqs $\underline{x} \in C$. Multiplicando (2.1.1) por \underline{z}' resulta $0 \geq c E u_1(d(\underline{x}, \underline{t}_0; \underline{V}_0^{-1}))$, y por (D) es $P(d(\underline{x}, \underline{t}_0; \underline{V}_0^{-1}) > s_0) = 1$, y esto implica que $\psi_2(b) > m$. QED

2.3.- Unicidad

La unicidad de las soluciones es fácil de probar para las ecuaciones que definen los parámetros del modelo, o sea, para el caso en que P es la distribución subyacente. Se puede obtener mayor generalidad en la distribución con condiciones

más restrictivas sobre ψ_1 .

Teorema. Para que la solución del sistema (2.2.1)-(2.1.2) sea única es suficiente cualquiera de las dos condiciones siguientes.

- i) La distribución P tiene una densidad f tal que $f(\underline{x})$ es una función decreciente de $|\underline{x}|$.
- ii) La distribución P es simétrica (o sea, $P(\underline{x} \in A) = P(-\underline{x} \in A)$ para todo borliano A), ψ_1 es no decreciente, y $m > 1$.

Demostración. Para (i) basta probar que para todo $\underline{t} \neq \underline{u}$ y toda \underline{S} no singular

$$\underline{t}' \int_{\mathbb{R}^m} u_1(|\underline{S}^{-1}(\underline{x}-\underline{t})|) (\underline{x}-\underline{t}) f(\underline{x}) d\underline{x} < 0. \quad (2.3.1)$$

Sea $A = \{\underline{x} \mid \underline{t}'(\underline{x}-\underline{t}) > 0\}$. Haciendo la transformación $\underline{y} = \underline{T}\underline{x} = \underline{t}'(\underline{x}-\underline{t}) + \underline{t}$ y llamando $\underline{z}(\underline{x})$ al integrando de (2.3.1) resulta

$$\begin{aligned} \int_{A'} \underline{t}' \underline{z}(\underline{x}) d\underline{x} &= \int_{A'} f(\underline{x}) \underline{t}'(\underline{x}-\underline{t}) u_1(|\underline{S}^{-1}(\underline{x}-\underline{t})|) d\underline{x} = \\ &= - \int_A f(2\underline{t}-\underline{y}) \underline{t}'(\underline{y}-\underline{t}) u_1(|\underline{S}^{-1}(\underline{y}-\underline{t})|) d\underline{y} \end{aligned}$$

y por lo tanto el primer miembro de (2.3.1) es

$$\int_A \underline{t}' \underline{z}(\underline{x}) \, d\underline{x} + \int_{A'} \underline{t}' \underline{z}(\underline{x}) \, d\underline{x} =$$

$$= \int_A \underline{t}'(\underline{x}-\underline{t}) u_1(|S^{-1}(\underline{x}-\underline{t})|) (f(\underline{x}) - f(2\underline{t}-\underline{x})) \, d\underline{x}$$

y esto es negativo, pues $\underline{x} \in A$ implica $|\underline{x}| > |2\underline{t}-\underline{x}|$.

Como la simetría de una distribución se conserva por transformaciones lineales, para (ii) basta probar que cqs $\underline{t} \neq \underline{u}$

$$\underline{t}' E_p \Psi(\underline{x}-\underline{t}) < 0 \tag{2.3.2}$$

donde $\Psi(\underline{x}) = u_1(|\underline{x}|) \underline{x}$. Aplicando Cauchy-Schwarz se verifica que cos \underline{a} y \underline{b}

$$(\underline{a}-\underline{b})' (\Psi(\underline{a}) - \Psi(\underline{b})) =$$

$$= \psi_1(|\underline{a}|) |\underline{a}| + \psi_1(|\underline{b}|) |\underline{b}| - \underline{a}'\underline{b} (\psi_1(|\underline{a}|)/|\underline{a}| + \psi_1(|\underline{b}|)/|\underline{b}|)$$

$$\geq (\psi_1(|\underline{a}|) - \psi_1(|\underline{b}|))(|\underline{a}| - |\underline{b}|) \geq 0 \tag{2.3.3}$$

de modo que el primer miembro de (2.3.3) es nulo sólo si $\underline{a}'\underline{b} = |\underline{a}| |\underline{b}|$, o sea, si \underline{a} y \underline{b} son linealmente dependientes. Por lo tanto, tomando $\underline{a} = \underline{x}-\underline{t}$, $\underline{b} = \underline{x}$, el integrando de (2.3.2) es ≤ 0 , y si el primer miembro es nulo, debe ser \underline{x} linealmente dependiente con \underline{t} con probabilidad 1, lo que contradice (E) si $m > 1$.

Observación: Para el caso simétrico univariado, es fácil verificar que una condición suficiente para la unicidad es que ψ_1 sea no decreciente, y que exista s_1 tq

$$\psi_2(s_1^2) > m, \text{ y } \psi_1' > 0, \text{ en } [0, s_1]$$

Para probar la unicidad de la solución para las ecuaciones que definen los estimadores -o sea, cuando P es la distribución empírica- no parece aprovechable el que P sea atómica, y por lo tanto es necesario probarlo para P cualquiera. Naturalmente esto requiere condiciones más fuertes sobre las ψ_1 . Una pauta sobre el tipo de condiciones requeridas se puede obtener del caso univariado.

Lema: Sea $\chi(s) = d \psi_2(s^2)/ds$. Las siguientes condiciones (F) son suficientes para la unicidad de las soluciones, si P cumple (E), en el caso $m = 1$.

(F1) $\psi_1'(s) \geq 0$ para todo $s \geq 0$.

(F2) Sea $s_1 = \sup \{s | \psi_1'(s) > 0\}$. Entonces $\chi(s) = 0$ si $s \geq s_1$.

(F3) Hay un $s_2 < s_1$ tal que $\psi_2(s_2^2) > m$ y tal que $\chi(s) / \psi_1'(s)$ es estrictamente creciente en $[0, s_2]$, y no decreciente en $[0, s_1]$.

Demostración: En esta demostración se considerará ψ_1

extendida a toda \mathbb{R} como función impar. Supongamos que haya dos soluciones distintas (t_0, S_0) y (t_1, S_1) . No hay inconveniente en suponer que $t_0 = 0$ y $S_0 = 1$. Sea $g(t) = E \psi_1(x-t)$. Entonces $g(0) = 0$, $g'(t) \leq 0$ para todo t , y como por (2.2.5) es $P(|x-t| \leq s_1) > 0$, resulta $g'(0) < 0$. Por lo tanto no puede ser $S_1 = 1$ y $t_1 \neq 0$, y en consecuencia es $t_1 \neq 0$ y $S_1 \neq 1$.

Para $r \in [0, 1]$, definimos $t(r) = t_1 r$, $S(r) = (1-r) + r S_1$, $z = z(r) = (x-t(r))/S(r)$. Sea $c = -t_1 / (S_1 - 1) = -t'(r)/S'(r)$. Sea $h(s)$ una función impar no decreciente igual a $\chi(s) / \psi_1'(s)$ para $s \in [0, s_1]$ y definamos

$$f(r) = E\{-h(c) \psi_1(z(r)) + [\psi_2(z^2(r)) - 1]\}$$

Dado que $f(0) = f(1) = 0$, se obtendrá un absurdo si se prueba que f monótona no constante. Derivando se obtiene:

$$\begin{aligned} f'(r) &= E\{z'(r) [(\chi(z) - h(c) \psi_1'(z))]\} = \\ &= -((S_1 - 1)/S(r)) E\{\psi_1'(s) (h(z) - h(c)) (z-c)\}. \end{aligned}$$

Para todo r , el integrando es no negativo, y en $r = 0$ la expectación es > 0 , pues $P(|z(0)| \leq s_2) > 0$. QED

Las condiciones (F) se cumplen obviamente si $\psi_2(s^2) = C[\psi_1(s)]^2$ -donde C es cualquier constante- que es el caso de la "Propuesta 2" de Huber. La demostración dada aquí es esencialmente un calco de la de Huber (1964), pág. 97. La misma idea parece ser válida para el caso multivariado general, pero algunos detalles de la demostración resultan difíciles.

3 - PROPIEDADES ASINTOTICAS

3.1.- Consistencia

Sea P una distribución en R^m que cumple (E) y para la cual hay solución única $(\underline{t}, \underline{v})$ de (2.1.1)-(2.1.2), sean $\underline{x}_1, \dots, \underline{x}_n, \dots$ variables independientes, todas con distribución P . Para cada n sean P_n la distribución empírica correspondiente a la muestra $\underline{x}_1, \dots, \underline{x}_n$ -es decir, la medida atómica tal que $P_n(\{\underline{x}_i\}) = 1/n$ - y $(\underline{t}_n, \underline{v}_n)$ una solución de (2.1.1)-(2.1.2) cuando la medida es P_n .

Teorema: $\lim_{n \rightarrow \infty} (\underline{t}_n, \underline{v}_n) = (\underline{t}, \underline{v})$ con probabilidad 1.

Demostración: De acuerdo con las condiciones del "caso B" de (Huber, 1967) basta probar que existe un compacto $K \subseteq \theta$ tal que con probabilidad 1 la sucesión de estimadores $(\underline{t}_n, \underline{v}_n)$ estén en K desde un cierto n en adelante, o sea

$$P \{ \liminf_n \{ (\underline{t}_n, \underline{v}_n) \in K \} \} = 1$$

Esto equivale a la existencia de constantes C_1, C_2, C_3 tales que con probabilidad 1

$$\limsup_n |\underline{v}_n^{-1}| \leq C_1, \quad \limsup_n |\underline{v}_n| \leq C_2, \quad \limsup_n |\underline{t}_n| \leq C_3 \quad (3.1.1)$$

Para cada n sean S_n la raíz simétrica definida positiva de V_n , $e_{1n} \leq \dots \leq e_{mn}$ los autovalores de V_n , z_{1n}, \dots, z_{mn} los correspondientes autovectores normalizados y $E_n = E_{P_n}$. En adelante se omitirá el subíndice n en todas partes, excepto E_n y P_n . Recordemos que $|V| = e_m$ y que $|V^{-1}| = 1/e_1$.

La generalización dada por Holfowitz (1954) del Teorema de Glivenko-Cantelli establece que, para los semiespacios $S \subseteq R^m$ se cumple con probabilidad 1 que $\lim_n P_n(S) = P(S)$ uniformemente en S . De aquí, repitiendo la demostración del Lema 2 de la sección 2.2 se prueba fácilmente la existencia de constantes A_1^1 y A_2^1 tales que con probabilidad 1 (pues $V_n = V(t_n)$)

$$\limsup_n |\underline{V}(t)^{-1}| \leq A_1^1{}^2, \quad (2.1.2)$$

$$\limsup_n \underline{t}' \underline{V}(t)^{-1} \underline{t} \leq A_2^1{}^2. \quad (3.1.3)$$

Para ello basta tener en cuenta que para cada $\underline{x}, \underline{t}$ y $c > 0$, los conjuntos de la forma $\{\underline{x} \mid |\underline{z}'(\underline{x}-\underline{t})| \leq c\}$ son intersecciones de dos semiespacios, por lo tanto existe c' tal que con probabilidad 1

$$\begin{aligned} & \limsup_n \sup_{\underline{t}} \sup_{|\underline{z}|=1} P_n\{\underline{x} \mid |\underline{z}'(\underline{x}-\underline{t})| \leq c'\} \\ & \leq 1 - m/K_2 \quad c/2. \end{aligned}$$

y de aquí se deduce (3.1.2) como en (i) del Lema 2 reemplazando

c por c'. Análogamente, se deduce del Teorema de Holfowitz que para cada ϵ existe un cubo $Q \subseteq \mathbb{R}^m$ tal que probabilidad 1

$$\liminf_{n \rightarrow \infty} P_n(Q) \geq 1 - \epsilon \quad (3.1.4)$$

y por lo tanto (2.2.5) vale uniformemente; o sea, existe A_2' tal que $|s| > A_2'$ implica que con probabilidad 1

$$\liminf_{n \rightarrow \infty} E_n \inf_{|u| \leq A_2'} \psi_2(|\underline{u}x - s|) \geq m + 2b$$

y (3.1.3) se demuestra entonces igual que (ii) del Lema 2.

Con (3.1.2) tenemos ya probada la primera desigualdad de (3.1.1), pues $\underline{v} = \underline{v}(\underline{t})$. Ahora basta probar la segunda, pues de ésta y de (3.1.3) sale la última.

Sean $\underline{b} = |\underline{v}|^{-1/2}$ y $\underline{z} = \underline{z}_m$, de modo que $\underline{S}\underline{z} = \underline{z}/\underline{b}$. Se probará que hay una constante $c > 0$ tal que con probabilidad 1, es $\liminf_n t \geq c$. Sean $\underline{w} = \underline{S}^{-1} \underline{t}$, ya se vió en (3.1.3) que $\limsup_n |\underline{w}| \leq A_2'$. Multiplicando (2.1.5) por \underline{z}' y por \underline{z} resulta

$$1 = E_n u_2(|\underline{S}^{-1}\underline{x} - \underline{w}|^2) (L \underline{z}'\underline{x} - \underline{z}'\underline{w})^2 \quad (3.1.4)$$

Sea $\delta \in (0,1)$ que luego se fijará convenientemente: sea $\epsilon < \delta/(K_2 + A_2' + K_1)$. Del teorema de Holfowitz se deduce la existencia de un $r > 0$ tal que con probabilidad 1 es $\limsup_n P_n(B_r^c) < \epsilon$. Poniendo para abreviar $q = |\underline{S}^{-1}\underline{x} - \underline{w}|$ resulta que para n grande es, con probabilidad 1

$$|1 - E_n I(B_r) u_2(q^2) (b \underline{z}'\underline{x} - \underline{z}'\underline{w})^2| \leq K_2 \epsilon < \delta \quad (3.1.5)$$

$$\begin{aligned} & |E_n I(B_r) u_2(q^2) (b \underline{z}'\underline{x} - \underline{z}'\underline{w})^2 - E_n I(B_r) u_2(q^2) (\underline{z}'\underline{w})^2| \leq \\ & \leq E_n I(B_r) u_2(q^2) (b^2 r^2 + 2 b r \lambda_2') \end{aligned} \quad (3.1.6)$$

$$E_n I(B_r) u_2(q^2) (\underline{z}'\underline{w})^2 \leq \epsilon \lambda_2' < \delta. \quad (3.1.7)$$

Sea $c_0 > 0$ tal que para $b < c_0$ el segundo miembro de (3.1.6) sea menor que δ . Sea $c = \min(c_0, \delta/r)$, y supongamos que $\liminf_n b < c$. Pasando a una subsucesión $\{n'\}$ tq $b_{n'} < c$, resulta de (3.1.5)-(3.1.6)-(3.1.7)

$$|1 - (\underline{z}'\underline{w})^2 E_{n'} u_2(q^2)| < 3 \delta \quad (3.1.8)$$

y tomando $\delta \leq 1/6$ se obtiene de (3.1.8) y de (A)

$$1/2 < (\underline{z}'\underline{w})^2 E_{n'} u_2(q^2) \leq (\underline{z}'\underline{w})^2 u_2(0). \quad (3.1.9)$$

Sea ahora s_0 la constante de (D) y $p = 1 - m / \psi_2(s_0^2)$, que por (2.2.4') cumple $P_n(q \leq s_0) \leq p$ para todo n , con probabilidad 1. De (3.1.9) se deduce

$$\begin{aligned} & |E_{n'} u_1(q) (\underline{z}'\underline{w})| \geq P_n(q \leq s_0) u_1(s_0) / (2 u_2(0))^{1/2} \geq \\ & \geq p u_1(s_0) / (2 u_2(0))^{1/2} \end{aligned} \quad (3.1.10)$$

Por otra parte, se definen las tres variables

$$a_1 = E_n I(B_r) u_1(c) (\delta \underline{z}'\underline{x} - \underline{z}'\underline{w})$$

$$a_2 = E_n I(D_c) u_1(c) \delta \underline{z}'\underline{x}$$

$$a_3 = E_n I(B_r) u_1(c) \underline{z}'\underline{w}.$$

De sus respectivas definiciones se obtiene que

$|a_1| \leq \epsilon K_1 < \delta$ y que $|a_2| \leq \delta r < \delta$. Multiplicando (2.1.4) por \underline{z}' resulta

$$\delta = a_1 + a_2 - a_3, \text{ y por lo tanto } |a_3| \leq 2 \delta$$

Ahora restando y sumando a_3 dentro del módulo, y recordando (3.1.3):

$$\limsup_n |E_n u_1(c) \underline{z}'\underline{w}| \leq \epsilon |\underline{z}'\underline{w}| + |a_3| \leq 3 \delta \quad (3.1.11)$$

De modo que tomando, por ejemplo, δ igual al último miembro de (3.1.10) dividido por ϵ , se obtiene una contradicción entre (3.1.11) y (3.1.10). QED

3.2.- Normalización sintáctica

Para regularizar la notación con la de (Huber, 1967) se escribirá genéricamente $\theta = (\underline{t}, \underline{v})$. Sean θ_3 el conjunto de los

matrices simétricas de $m \times m$, y $\theta^0 = \theta_1 \times \theta_3 \supseteq \theta$. Se utilizará en θ^0 (y por lo tanto en θ) la norma $|\theta| = \max(|\underline{t}|, |\underline{V}|)$.

Sea Ψ la función de $\mathbb{R}^m \times \theta$ en θ . $\Psi(\underline{x}, \theta) = (\Psi_1(\underline{x}, \theta), \Psi_2(\underline{x}, \theta))$ definida por

$$\Psi_1(\underline{x}, \theta) = u_1(d(\underline{x}, \underline{t}, \underline{V}^{-1})) (\underline{x} - \underline{t})$$

$$\Psi_2(\underline{x}, \theta) = u_2(d^2(\underline{x}, \underline{t}, \underline{V}^{-1})) (\underline{x} - \underline{t})(\underline{x} - \underline{t})' - \underline{V}.$$

Sean P la distribución subyacente y $\lambda(\theta) = (\lambda_1(\theta), \lambda_2(\theta)) = E_P \Psi(\underline{x}, \theta)$, de modo que el parámetro θ a estimar está definido por $\lambda(\theta) = \lambda$. Si F_n es la distribución empírica correspondiente a una muestra de tamaño n , el estimador $\theta_n^* = (\underline{t}_n, \underline{V}_n)$ queda definido por $E_{F_n} \Psi(\underline{x}, \theta_n^*) =$

Para $j = 1, 2$ sean

$$U_j(\underline{x}, \theta, \delta) = \sup\{|\Psi_j(\underline{x}, \theta_1) - \Psi_j(\underline{x}, \theta)| \mid |\theta_1 - \theta| < \delta\}.$$

Para probar la normalidad asintótica, hay que verificar las condiciones (H-3) de (Huber, 1967) pág. 227, o saber, que existen a, b, c, δ_0 positivos tales que para $j = 1, 2$

- i) $|\lambda(\theta)| \geq |\theta - \theta^0|$ para $|\theta - \theta^0| \leq \delta$
- ii) $E U_j(\underline{x}, \theta, \delta) \leq b \delta$ para $|\theta - \theta^0| + \delta \leq \delta_0$ ($\delta \geq b$)
- iii) $E U_j^2(\underline{x}, \theta, \delta) \leq c \delta$ para $|\theta - \theta^0| + \delta \leq \delta_0$ ($\delta \geq c$).

Sea $(D\lambda)_\theta$ la derivada de λ en θ , o sea el operador lineal de θ° en θ° tal que para $(\underline{h}, \underline{t}) \in \theta^\circ$,

$$\lambda[\theta + (\underline{h}, \underline{t})] - \lambda(\theta) = (D\lambda)_\theta (\underline{h}, \underline{t}) + o(|(\underline{h}, \underline{t})|).$$

El Corolario del Lema 3 de Huber establece que si se cumplen (i), (ii) y (iii), y si $(D\lambda)_\theta$ es no singular, la distribución límite de $n^{1/2}(\theta_n^* - \theta)$ es normal con media λ y matriz de covarianza $(D\lambda)_\theta^{-1} C (\lambda)_\theta^{-1}$, donde C es la matriz de covarianza de la variable $\Psi(\underline{x}, \theta)$.

Para simplificar la escritura, se considerará sólo el caso de P radial.

Teorema: Si las funciones $s\psi_j^*(s)$ ($j = 1, 2$) son acotadas, y P tiene densidad radial f tal que $E_P \psi_1^*(|\underline{x}|) / s > \delta$ para todo $s > \delta$, entonces $n^{1/2}(\theta_n^* - \theta)$ tiene distribución límite normal y \underline{t}_n y \underline{v}_n son asintóticamente independientes.

Demostración: Para obtener (ii) y (iii) se calcularán las derivadas de ψ_1 y ψ_2 . Se escribirá d en vez de $d(\underline{x}, \underline{t}, \underline{v}^{-1})$

$$\begin{aligned} (D\psi_1)_\theta (\underline{h}, \underline{t}) &= \\ &= -[2 \underline{h}' \underline{v}^{-1} (\underline{x} - \underline{t}) + (\underline{x} - \underline{t})' \underline{v}^{-1} \underline{v}^{-1} (\underline{x} - \underline{t})] (\underline{x} - \underline{t})' u_1'(d) / (2d) - \\ &= u_1(d) \underline{h} \end{aligned} \tag{3.2.1}$$

$$\begin{aligned}
 (D \psi_2)_\theta (h, \underline{t}) &= \\
 &= - u_2'(d^2) [2 \underline{h}' \underline{V}^{-1}(\underline{x}-\underline{t}) + (\underline{x}-\underline{t})' \underline{V}^{-1} \underline{M} \underline{V}^{-1}(\underline{x}-\underline{t}) (\underline{x}-\underline{t})'] - \\
 &- u_2(d^2) [\underline{h}(\underline{x}-\underline{t})' + (\underline{x}-\underline{t}) \underline{h}'] - \underline{M} \quad . \quad (3.2.2)
 \end{aligned}$$

Como las funciones $s\psi_j'(s)$ y $\psi_j(s)/s$ son acotadas, también las ψ_j' lo son ($j = 1, 2$). Teniendo en cuenta que $|\underline{x}-\underline{t}|/d \leq |\underline{V}|^{1/2}$, es fácil ver que hay constantes A y B que dependen sólo de $|\underline{V}|$ y de $|\underline{V}^{-1}|$ tales que

$$|(D \psi_j)_\theta (h, \underline{t})| \leq A |\underline{h}| + B |\underline{t}| \quad (j = 1, 2)$$

y del Teorema del Valor Medio (Cartan, 1972) resulta que si θ permanece en un compacto, existe $b > 0$ tal que, para todo \underline{x} , $U_j(\underline{x}, \theta, \delta) \leq b \delta$, de donde resultan obviamente (ii) y (iii).

Ahora, si P es radial, es $\theta_j = (\underline{u}, s^2 \underline{I})$ donde s es alguna constante. De (3.2.1) y (3.2.2) se obtiene, usando la simetría de P y el hecho de que $u_j'(r) = (x_j'(r) - u_j(r))/r$ para todo $r > 0$

$$\begin{aligned}
 E (D \psi_1)_\theta (h, \underline{x}) &= \{E u_1(|\underline{x}|/s) [\underline{h} - (\underline{h}'\underline{x}) \underline{x} / |\underline{x}|^2] + \\
 &+ E \psi_j'(|\underline{x}|/s) (\underline{u}'\underline{x}) \underline{x} / |\underline{x}|^2\} \quad (3.2.3)
 \end{aligned}$$

$$\begin{aligned}
 E (D \psi_2)_{\theta_0}(\underline{h}, \underline{h}) &= \\
 &= - \{ E u_2'(|\underline{x}|^2/s^2) \underline{x}' \underline{ix} \underline{xx}' / s^4 + \text{II} \} \quad (3.2.4)
 \end{aligned}$$

Para probar (i) basta demostrar que $(D\Psi)_{\theta_0}$ es no singular. De (3.2.3) y (3.2.4) se deduce que las "derivadas parciales" $\partial\lambda_1/\partial\underline{v}$ y $\partial\lambda_2/\partial\underline{t}$ son nulas en θ_0 , de modo que es suficiente probar que los operadores $\underline{D}_{1\underline{t}} = (\partial\lambda_1/\partial\underline{t})_{\theta_0}$ y $\underline{D}_{2\underline{v}} = (\partial\lambda_2/\partial\underline{v})_{\theta_0}$ son no singulares.

Como f es radial, $|\underline{x}|$ es independiente de $\underline{z} = \underline{x} / |\underline{x}|$, y este último vector tiene distribución uniforme sobre la superficie esférica $\{|\underline{z}| = 1\}$. Multiplicando (3.2.3) por \underline{h}' , y teniendo en cuenta que $E(\underline{h}'\underline{z})^2 = |\underline{h}|^2/m$ para todo \underline{h} , resulta

$$\begin{aligned}
 \underline{h}' \underline{D}_{1\underline{t}} \underline{h} &= -[(1 - 1/m) E u_1'(|\underline{x}|/s) + \\
 &+ (1/m) E \psi_1'(|\underline{x}|/s)] |\underline{h}|^2 \quad (3.2.5)
 \end{aligned}$$

y esto es negativo para todo $\underline{h} \neq 0$, lo que prueba que $\underline{D}_{1\underline{t}}$ es no singular.

Ahora llamaremos $-\underline{M}$ a la expresión en (3.2.4), es decir

$$\underline{M} = -\underline{D}_{2\underline{v}} \underline{v} = \underline{U} + \underline{H}$$

donde

$$\underline{U} = E u_2'(|\underline{x}|^2/s^2) |\underline{x}|^4/s$$

$$\underline{H} = E \underline{z}' \underline{iz} \underline{zz}'$$

Debemos probar que $\underline{u} = \underline{0}$ sólo si $\underline{u} = \underline{0}$. Usando (2.2.4) resulta

$$\Lambda = E \psi_2'(|\underline{x}|^2/s^2) |\underline{x}|^2/s^2 - E \psi_2(|\underline{x}|^2/s^2) = c - m \quad (3.2.6)$$

donde c es algún número positivo, por (2.2.4) y (D).

Para calcular \underline{u} , probaremos que sus autovectores coinciden con los de \underline{A} , y calcularemos los correspondientes autovalores. Conviene previamente disponer de un resultado. Si \underline{x} tiene distribución normal radial, y \underline{e}_1 es cualquier vector de longitud 1, la variable $y = (\underline{x}'\underline{e}_1)^2/|\underline{x}|^2$ tiene distribución Beta con parámetros $(m-1)/2$ y $1/2$, y este resultado sigue siendo válido cualquiera sea la distribución (radial) de \underline{x} , pues $y = (\underline{z}'\underline{e}_1)^2$, y la distribución de \underline{z} es siempre la uniforme. De aquí se obtiene entonces que

$$E (\underline{z}'\underline{e}_1)^4 = 3 / (m(m+2)). \quad (3.2.7)$$

De este último se deduce todavía que si \underline{e}_2 es ortogonal a \underline{e}_1 y tiene longitud 1 es

$$E (\underline{z}'\underline{e}_1)^2 (\underline{z}'\underline{e}_2)^2 = 1 / (n(n+2)). \quad (3.2.8)$$

Sean entonces \underline{e}_i y λ_i ($i = 1, \dots, m$) los autovectores de \underline{A} y los correspondientes autovalores, de modo que poniendo $z_i = \underline{z}'\underline{e}_i$ ($i = 1, \dots, m$), la coordenada k -ésima del vector \underline{u} \underline{e}_j en la base $\{\underline{e}_1, \dots, \underline{e}_m\}$ resulta

$$\underline{u}_k^T (\underline{U} \underline{u}_j) = \sum_{i=1}^m a_i E z_i^2 z_j z_k .$$

Como la distribución de \underline{z} es radial, la distribución conjunta de $(\pm z_1, \pm z_2, \dots, \pm z_m)$ es igual a la de \underline{z} , para cualquier combinación de los signos \pm en consecuencia, cuando $j \neq k$ se anula la expectativa dentro de la sumatoria, y por ende el primer miembro.

Esto prueba que cada \underline{u}_j es autovector de \underline{U} , con autovalor

$$u_j = \sum_{i=1}^m a_i E z_i^2 z_j^2 = \left(\sum_{i=1}^m a_i + 2 a_j \right) / (m(m+2)) .$$

Uniendo este último con (3.2.0) se obtiene que los \underline{u}_j son también los autovectores de \underline{U} , con autovalores

$$v_j = [(c-m) \sum_{i=1}^m a_i + (m^2+2c) a_j] / [m(m+2)] . \quad (3.2.9)$$

Hay que probar que si los v_j son todos nulos, los a_j también lo son. Si $c = m$, este hecho es obvio por (3.2.9).

Si $c \neq m$, el sistema de ecuaciones (en las a_j) dado por

$$v_j = 0 \quad (j = 1, \dots, m) \text{ es equivalente a}$$

$$\sum_{i=1}^m a_i + C a_j = 0 \quad (j = 1, \dots, m)$$

con $C = (m^2 + 2c) / (c-m)$. La matriz \underline{a} de este sistema tiene

coeficientes $b_{ij} = 1 + C\delta_{ij}$. Por consideraciones de simetría se deduce que los autovalores de \underline{b} son $C+m$ y C (con multiplicidades 1 y $m-1$, respectivamente). Como obviamente ambos son distintos de cero, el sistema tiene solución única y esto termina con (i).

Para completar el teorema: como la distribución de \underline{z} es igual a la de $-\underline{z}$ cualesquiera sean $i, j, k \in \{1, \dots, m\}$ es $E z_i z_j z_k = 0$. Por lo tanto la covarianza entre cualquier componente de $\underline{\Psi}_1(\underline{x}, \theta_1)$ y cualquier elemento de $\underline{\Psi}_2(\underline{x}, \theta_2)$ es nula, y en consecuencia la matriz \underline{C} de covarianzas de $\underline{\Psi}$ y la $(\partial \lambda)_{\theta}$ respectivamente, se pueden representar en bloques como

$$\begin{bmatrix} \underline{A} & 0 \\ \underline{B} & \underline{B} \end{bmatrix} \quad \begin{bmatrix} \underline{E} \underline{1t} & 0 \\ 0 & \underline{L} \underline{2V} \end{bmatrix}$$

donde \underline{A} y \underline{E} son las matrices de covarianzas de $\underline{\Psi}_1$ y $\underline{\Psi}_2$ respectivamente. Por lo tanto la matriz de covarianza asintótica de $(\underline{t}_n, \underline{V}_n)$ se puede escribir en bloque como

$$\begin{bmatrix} \underline{D} \underline{1t}^{-1} \underline{A} \underline{B} \underline{1t}^{-1'} & 0 \\ 0 & \underline{L} \underline{2V}^{-1} \underline{B} \underline{L} \underline{2V}^{-1'} \end{bmatrix}$$

y esto prueba que \underline{t}_n y \underline{V}_n son asintóticamente independientes. QED

Para calcular en el próximo capítulo la distribución asintótica de \underline{t}_n , notemos que el segundo miembro de (3.2.5) depende sólo de $|\underline{h}|$, y por lo tanto $\underline{D}_{1\underline{t}}$ es un múltiplo escalar de \underline{I} , y lo mismo sucede con $E \underline{\psi}_1 \underline{\psi}_1'$ por simetría. O sea

$$\underline{D}_{1\underline{t}} = -b \underline{I} \quad \text{y} \quad E \underline{\psi}_1 \underline{\psi}_1' = a \underline{I} \quad (3.2.10)$$

con

$$a = s^2 E \psi_1^2 (|\underline{x}|^2/s^2) / m \quad (3.2.11)$$

$$b = E [u_1(|\underline{x}|/s) (1 - 1/m) + \psi_1' (|\underline{x}|/s) / m] \quad (3.2.12)$$

y la matriz de covarianzas asintótica de \underline{t}_n es $c \underline{I}$, con $c = a/b^2$.

3.3.- Robustez

Para comparar cuantitativamente el comportamiento de distintos estimadores para nuestras grandes, se utilizarán, además de las varianzas asintóticas calculadas para varias distribuciones, dos índices de robustez introducidos por Hampel (1968): la función de influencia ("influence curve") y la cota de derrumbe ("breakdown bound").

Considerando un estimador T cualquiera como un funcional $T(P)$ sobre el espacio de las medidas de probabilidad P en

$\Omega = \mathbb{R}^m$, con valores en un espacio de parámetros θ (que coincide con \mathbb{R}^m cuando T es un estimador de localización) se define la función de influencia de T , para la medida P , calculada en el punto $\underline{x} \in \Omega$:

$$FI(\underline{x}) = FI(\underline{x}, P, T) = \lim_{\epsilon \rightarrow 0^+} \{T[(1-\epsilon)P + \epsilon\delta_{\underline{x}}] - T(P)\} / \epsilon$$

donde $\delta_{\underline{x}}$ es la probabilidad concentrada en \underline{x} . Es decir que $FI(\underline{x})$ da una idea de la influencia que puede tener sobre T una pequeña desviación del "modelo" P , en la forma de una pequeña contaminación en \underline{x} .

Para calcular las funciones de influencia de nuestros θ -estimadores se escribirá, para una P fija

$$Q = Q(\underline{x}, \epsilon) = (1-\epsilon) P + \epsilon \delta_{\underline{x}} \tag{3.3.1}$$

Desde ahora supondremos que se cumplen todas las hipótesis del Teorema de la sección anterior. Expresando (2.1.1) en la notación de dicha sección, y teniendo en cuenta que $E_P \underline{\psi}_1(\underline{x}, \theta^*(P)) = \underline{0}$, y que $\theta^*(P) = \theta_0 = (\underline{0}, s^2 \underline{I})$, resulta para cada \underline{x}_0 (poniendo para abreviar Q en vez de $Q(\underline{x}_0, \epsilon)$)

$$\begin{aligned} \underline{0} &= E_Q \underline{\psi}_1(\underline{x}_0, \theta^*(Q)) = (1-\epsilon) E_P \underline{\psi}_1(\underline{x}_0, \theta^*(P)) + \epsilon \underline{\psi}_1(\underline{x}_0, \theta^*(Q)) \\ &= (1-\epsilon) \underline{0}_{1t} \underline{t}(\theta_0) + \theta(\epsilon) + \epsilon \underline{\psi}_1(\underline{x}_0, \theta^*(Q)) \end{aligned}$$

y teniendo en cuenta (3.2.10) se obtiene, haciendo $\epsilon \rightarrow 0+$

$$FI(\underline{x}_0, \underline{t}, P) = - \frac{D_{\underline{t}}^{-1}}{1} \underline{\psi}_1(\underline{x}, \theta_0) = \psi_1(|\underline{\psi}_0|/s)(s/b) \underline{x}_0/|\underline{x}_0|. \quad (3.2.2)$$

La función de influencia de \underline{v} se obtiene de igual manera, pero la expresión no resulta muy manejable.

Una medida numérica de robustez que se obtiene de la FI es la llamada por Hampel "sensitividad para errores groseros" ("cross error sensitivity") definida como

$$SEG = SEG(T, P) = \sup_{\underline{x}} |FI(\underline{x}, T, P)|$$

que expresa la máxima influencia que puede tener una pequeña desviación del modelo P. En nuestro caso resulta de (3.3.2) que $SEG = K_1 s/b$.

La cota de derrumbe de T en P expresa cuánto se puede uno apartar del modelo P sin que el estimador se desquicie, es decir, deje de dar información sobre el parámetro (como caso típico, para estimadores de locación éste equivale a la proporción de masa que se puede colocar en ∞ sin que T también se vaya a ∞). En nuestro caso, considerando el estimador como un funcional sobre las probabilidades, la definición de Hampel (1971) se puede expresar como sigue. Sea $\#$ la distancia de Prohorov sobre las probabilidades en $\Omega = \mathcal{R}^b$ entonces la cota de derrumbe $\delta^* = \delta^*(T, P)$ es el supremo de los $\epsilon \in [0, 1]$ tales que existe un compacto K_ϵ contenido estrictamente en θ tal que

$\Pi(P,Q) < \epsilon$ implica $T(\theta) \in K_\epsilon$. Se buscará una cota para δ^* .

Sea $\epsilon^* = \epsilon^*(T,P)$ el supremo de los $\epsilon \in]0,1[$ tales que existe un compacto K_ϵ contenido estrictamente en θ , tal que $T(\theta(\underline{x}_j, \epsilon)) \in K_\epsilon$ para todo \underline{x}_j . Es fácil probar que $\Pi[P,Q(\underline{x}_j, \epsilon)] \leq \epsilon$, y por lo tanto $\delta^* \leq \epsilon^*$.

Para acotar el ϵ^* de $\theta^* = (\underline{t}, \underline{V})$ se usará la ecuación (2.1.5). Poniendo $\underline{t} = \underline{t}(\theta(\underline{x}_j, \epsilon))$, $\underline{S} = \underline{S}(\theta(\underline{x}_j, \epsilon))$, $\underline{z} = \underline{S}^{-1}(\underline{x}_j - \underline{t})$ y recordando que la matriz del integrando es semidefinida positiva:

$$\begin{aligned} \underline{I} &= E_{\theta(\underline{x}_j, \epsilon)} u_2(|\underline{S}^{-1}(\underline{x}_j - \underline{t})|^2) (\underline{S}^{-1}(\underline{x}_j - \underline{t})) (\underline{S}^{-1}(\underline{x}_j - \underline{t}))' \\ &\geq \epsilon u_2(|\underline{z}|^2) \underline{z}\underline{z}'. \end{aligned}$$

Multiplicando por \underline{z}' y por \underline{z} , se obtiene que

$$1 \geq \epsilon \psi_2(|\underline{z}|^2).$$

Ahora, si $\epsilon < \epsilon^*$, cuando $\underline{x}_j \rightarrow \infty$, $(\underline{t}, \underline{S})$ permanece en un compacto, y por lo tanto $\underline{z} \rightarrow \infty$, de modo que de la última fórmula resulta que ϵ debe ser menor que $1/K_2$, y por lo tanto

$$\epsilon^* \leq 1 / K_2. \quad (3.3.3)$$

Por otra parte, tomando $\underline{x}_j = \underline{t}$ y notando que en este caso es $\underline{t}(\theta(\underline{x}_j, \epsilon)) = \underline{t}$ de (2.2.4) se obtiene

$$\begin{aligned}
 m &= E_{O(\underline{x}_T, \epsilon)} \psi_2[d^2(\underline{x}, \underline{t}; \underline{V}^{-1})] = \\
 &= (1-\epsilon) E_P \psi_2[d^2(\underline{x}, \underline{t}; \underline{V}^{-1})] \leq (1-\epsilon) K_2
 \end{aligned}$$

y por lo tanto- debe ser

$$\epsilon^* \leq 1 - m/K_2 \quad (3.3.4)$$

De (3.3.3)-(3.3.4) resulta

$$\delta^* \leq \epsilon^* \leq \min\{1/K_2, 1-m/K_2\} \quad (3.3.5)$$

y como $K_2 \geq m$, resulta en definitiva

$$\delta^* \leq 1/(m+1) \quad (3.3.6)$$

De modo que hay una cota para δ^* que no depende de las funciones ψ_j elegidas. Si bien la cota de derrumbe expresa el comportamiento del estimador en condiciones muy extremas, y por lo tanto su valor exacto no tiene por qué ser tomado al pie de la letra, el hecho de que para m grande δ^* sea inevitablemente pequeño señala una dificultad cuyo significado preciso habrá que investigar más detalladamente.

Observación: De los resultados expuestos por Hampel (1968 y 1971) se desprende que al menos para $m = 1$

$$\delta^* = \epsilon^* = \min \{1/K_2, 1-n/K_2, 1/2\} \quad (3.3.6)$$

Infelizmente Hampel no explica los detalles de la demostración.

3.4.- Eficiencia asintótica

En esta sección se comparan los comportamientos asintóticos de distintos estimadores de localización para varias distribuciones radiales. Se presentarán dos familias de estimadores:

- i) La "Propuesta 2" de Huber, descrita en 2.1. Para no aumentar excesivamente el número de estimadores a considerar, se tomó $k_2 = k_1$. Para permitir alguna comparación entre distintas dimensiones, se consideró mejor parametrizar la familia con $p = P(|x| \geq k_1)$ (donde P es la normal radial standard) o sea, la "proporción de winsorización". El estimador se indica con $H(p, m) = H(p)$. Se consideraron los valores de p : 0.5, 0.3 y 0.2 (que cuando $m = 1$ corresponden aproximadamente a los valores de k_1 : 1.7, 1.6 y 1.2, tres de los valores examinados en (Andrews y otros, 1972)). También se incluyó el caso límite $p = 1$ (o sea $k_1 = \infty$), es decir, la "mediana", con $k_2 \neq \infty$ cualquiera.

Como esta familia está "calibrada" para la normal, para esta distribución será siempre $s^2 = 1$.

- ii) Estimador de máxima verosimilitud para la distribución de Student radial con g grados de libertad, definido en 2.1.

se lo indica con $MVST(g,m)$ o $MVST(g)$, para los valores de g : 1, 2, 3 y 5.

Por comodidad se ha preferido no calibrarlo para la normal.

El comportamiento de estas dos familias se describe en la Tabla 1.

Para completar el panorama, se incluirán en la comparación:

- iii) Estimadores de localización basados en la aplicación coordenada a coordenada de un estimador univariante. Se utilizarán de esta forma $MVST(5,1)$ y $H(3,3,1)$ (elegidos por tener comportamiento semejante para la normal). Los resultados se exhiben en Tabla 2.

Las distribuciones elegidas son la normal, indicada con $N(m)$ y la familia de Student radial con g grados de libertad, indicada con $ST(g)$ ó $ST(g,m)$. Esta familia se eligió por varios motivos:

- a) tiene una gama completa de "pesadez de colas", con la normal en un extremo ($g = \infty$) y Cauchy en el otro ($g = 1$).
- b) por ser de la forma "Normal/Independiente", las marginales son de la misma familia, lo que permite algunas comparaciones entre distintos valores de m .
- c) la forma analítica sencilla de las densidades hace los cálculos numéricos bastante fáciles, y permite calcular la cota de Rao-Cramér.

Tales los estimadores en consideración tienen distribución asintótica normal con media μ y matriz de covarianzas

$\underline{I} = c\underline{I}$, de modo que este número $c = c(\underline{t})$ -que coincide con la varianza asintótica de cada coordenada de \underline{t} , y con $\text{tr } \underline{I}/m$ y con $(\det \underline{I})^{1/m}$ - contiene toda la información necesaria para comparar los estimadores, será llamada "varianza" de \underline{t} . En las tablas que siguen se exhiben las "deficiencias" de los estimadores, obtenidas dividiendo la varianza de \underline{t} por la del estimador de máxima verosimilitud de la distribución correspondiente.

Este procedimiento es perfectamente natural para comparar los \underline{I} -estimadores entre sí, pues por ser éstos equivariantes por transformaciones afines en general, la eficiencia relativa de dos de ellos para una distribución radial, depende sólo del tipo de ésta, y de lo mismo usando como criterio $\text{tr } \underline{I}$ o $\det \underline{I}$. En cambio, la eficiencia de los estimadores (iii) dependerá en general del criterio que se use y de la relación entre las variables. según propuso Dickel (1964) para la mediana y el R-estimador de Hodges-Lehmann aplicada a coordenada a coordenada en dos dimensiones, si se utiliza el criterio del determinante, la eficiencia de estos estimadores respecto a la media muestral en el caso normal tiende a cero cuando la correlación tiende a uno. Por lo tanto, al limitarnos al caso radial nos colocamos en la situación más favorable para estos estimadores, si se usa el criterio del determinante.

En la Tabla 1, junto con las deficiencias se da para δ^* el valor (3.3.5) (o el (3.3.6) para $m=1$), y una medida de sensibilidad equivalente a $SE\theta$, dada por $SEIS = SE\theta^2/m$.

Esta transformación puede facilitar algún grado de comparación entre distintos m . Desconozco la manera de calcular la función de influencia -y por lo tanto la SENS- para estimadores de tipo (iii).

Las varianzas se pueden obtener de las tablas multiplificando las eficiencias por la varianza de $WVST(g,m)$ para $ST(g,m)$, que es $1 + 2/(m+g)$. Para comparar con la media: la varianza de ésta para $ST(g,m)$ es $g/(g-2)$.

Observación 1: Calculando (3.2.11)-(3.2.12) para $H(k,m)$ en el caso normal -recomendándose $s = 1-$ y haciendo luego $k \rightarrow 0$ se obtiene que la varianza de $H(1,0)$ ($m \geq 2$) para la normal es

$$\text{Var} = 2 [\Gamma(m/2) / \Gamma((m-1)/2)]^2 / [m(1 - 1/m)^2]$$

que tiende a 1 con $m \rightarrow \infty$, lo cual explica la alta eficiencia de $H(1,0,m)$ -y a fortiori la de los demás $H(k,m)$ - con m grande. Además, no es difícil verificar que $\text{VAR} = \text{SENS}$.

No se consigna el valor de $\delta^*(H(1,0))$ por depender éste del estimador de escala usado.

Observación 2: Véase que para cada m , al variar g , los valores de SENS y de la varianza asintótica de $WVST(g,m)$ en la normal, disminuyen simultáneamente, hasta alcanzar un punto en que se comienza a esperar SENS vuelva a aumentar (ese punto está en algunos casos bastante más allá de $g = 5$ y por eso no aparece aquí en todos los valores de m). Descartando la posibilidad de un error de cálculo -esto hecho resulta bastante singular-

TABLE 1

Deficiencia y robustez asintóticas de MVST(g,m) y H(p,m)

	ST(1)	ST(2)	ST(3)	ST(5)	NORMAL	SEHS	δ^*
<u>m = 1</u>							
MVST(1)	1.000	1.033	1.173	1.296	1.820	2.862	0.500
MVST(2)	1.125	1.000	1.019	1.073	1.314	2.119	0.333
MVST(3)	1.333	1.023	1.000	1.020	1.184	2.128	0.250
MVST(5)	1.800	1.112	1.024	1.000	1.092	2.459	0.167
H(1.0)	1.234	1.200	1.234	1.301	1.571	1.562	0.500
H(0.5)	1.247	1.035	1.017	1.037	1.195	1.820	0.500
H(0.3)	1.547	1.095	1.024	1.005	1.099	2.192	0.500
H(0.2)	1.860	1.172	1.056	1.009	1.060	2.566	0.412
<u>m = 2</u>							
MVST(1)	1.000	1.054	1.116	1.206	1.537	2.278	0.333
MVST(2)	1.033	1.000	1.015	1.069	1.279	1.639	0.250
MVST(3)	1.228	1.018	1.000	1.017	1.179	1.533	0.200
MVST(5)	1.563	1.034	1.022	1.000	1.097	1.609	0.143
H(1.0)	1.200	1.081	1.071	1.094	1.273	1.273	---
H(0.5)	1.460	1.102	1.031	1.005	1.093	1.515	0.361
H(0.3)	1.826	1.208	1.081	1.018	1.048	1.802	0.291
H(0.2)	2.161	1.300	1.129	1.037	1.029	2.070	0.248

	ST(1)	ST(2)	ST(3)	ST(5)	NORMAL	SEMS	δ^*
<u>m = 3</u>							
RVST(1)	1.899	1.839	1.886	1.158	1.440	2.292	0.250
RVST(2)	1.861	1.800	1.812	1.059	1.258	1.566	0.200
RVST(3)	1.172	1.015	1.000	1.015	1.176	1.392	0.166
RVST(5)	1.433	1.380	1.319	1.000	1.101	1.360	0.125
H(1.0)	1.234	1.071	1.041	1.041	1.178	1.178	
H(0.5)	1.578	1.158	1.062	1.012	1.001	1.399	0.253
H(0.3)	1.952	1.279	1.125	1.038	1.031	1.640	0.209
H(0.2)	2.280	1.372	1.177	1.052	1.019	1.953	0.183

<u>m = 4</u>							
RVST(1)	1.909	1.839	1.866	1.122	1.377	2.423	0.200
RVST(2)	1.847	1.800	1.810	1.042	1.239	1.589	0.167
RVST(3)	1.136	1.013	1.000	1.013	1.170	1.362	0.143
RVST(5)	1.359	1.369	1.117	1.000	1.103	1.261	0.111
H(1.0)	1.270	1.081	1.037	1.023	1.132	1.132	---
H(0.5)	1.650	1.190	1.088	1.023	1.045	1.336	0.105
H(0.3)	2.016	1.321	1.150	1.055	1.023	1.547	0.165
H(0.2)	2.333	1.417	1.210	1.001	1.014	1.739	0.146

	ST(1)	ST(2)	ST(3)	ST(5)	NORMAL	SENS	δ^*
<u>m = 6</u>							
RVST(1)	1.000	1.019	1.043	1.081	1.287	2.792	0.143
RVST(2)	1.132	1.000	1.027	1.031	1.203	1.734	0.125
RVST(3)	1.194	1.009	1.000	1.010	1.155	1.413	0.111
RVST(5)	1.249	1.053	1.013	1.000	1.101	1.210	0.090
H(1.0)	1.327	1.107	1.047	1.015	1.080	1.080	---
H(0.5)	1.728	1.252	1.126	1.044	1.030	1.264	0.136
H(0.3)	2.073	1.375	1.198	1.081	1.015	1.440	0.117
H(0.2)	2.369	1.469	1.252	1.109	1.009	1.591	0.106

<u>m = 10</u>							
RVST(1)	1.000	1.010	1.022	1.043	1.189	3.679	0.090
RVST(2)	1.018	1.000	1.004	1.018	1.151	2.145	0.083
RVST(3)	1.055	1.006	1.000	1.006	1.125	1.652	0.077
RVST(5)	1.152	1.034	1.009	1.000	1.091	1.289	0.066
H(1.0)	1.397	1.107	1.073	1.024	1.051	1.051	---
H(0.5)	1.787	1.305	1.168	1.073	1.018	1.198	0.084
H(0.3)	2.099	1.422	1.240	1.112	1.009	1.335	0.075
H(0.2)	2.352	1.500	1.212	1.100	1.005	1.451	0.069

Tabla 2Deficiencia asintótica de $WVST(5, 1)$ y $H(0, 3, 1)$

	ST(1)	ST(2)	ST(3)	ST(5)
<u>$m = 2$</u>				
WVST	2.100	1.230	1.157	1.137
H	1.856	1.217	1.157	1.142
<u>$m = 3$</u>				
WVST	2.410	1.324	1.152	1.167
H	2.063	1.304	1.152	1.172
<u>$m = 4$</u>				
WVST	2.571	1.390	1.195	1.151
H	2.210	1.361	1.195	1.196
<u>$m = 6$</u>				
WVST	2.850	1.483	1.257	1.128
H	2.400	1.460	1.257	1.134
<u>$m = 10$</u>				
WVST	3.040	1.589	1.331	1.170
H	2.610	1.564	1.331	1.152

Nota: De la Tabla 1 se obtiene que las deficiencias de $WVST(5, 1)$ y de $H(0, 3, 1)$ para la normal m -variada son respectivamente, 1.152 y 1.150 para todo m , y los respectivos δ^* son 1.157 y 1.150, también para todo m .

4 - MUESTRAS FINITAS

4.1.- Cálculo numérico de los estimadores

Definamos $F(\underline{t}, \underline{v})$ como en (2.2.1), y además las funciones F° y G de θ en θ_2 y de θ en θ_1 respectivamente:

$$F^\circ(\underline{t}, \underline{v}) = F(\underline{t}, \underline{v}) \underline{v}^{-1} F(\underline{t}, \underline{v}) \quad (4.1.1)$$

$$G(\underline{t}, \underline{v}) = E u_1(d(\underline{x}, \underline{t}, \underline{v}^{-1})) \underline{x} / E u_1(d(\underline{x}, \underline{t}, \underline{v}^{-1})) \quad (4.1.2)$$

de modo que $(\underline{t}, \underline{v})$ es solución de (2.1.1)-(2.1.2) si y sólo si cumple

$$G(\underline{t}, \underline{v}) = \underline{t} \quad \text{y} \quad F(\underline{t}, \underline{v}) = \underline{v} \quad (4.1.3)$$

o bien

$$G(\underline{t}, \underline{v}) = \underline{t} \quad \text{y} \quad F^\circ(\underline{t}, \underline{v}) = \underline{v} . \quad (4.1.3')$$

Para calcular $(\underline{t}, \underline{v})$ cuando P es una distribución empírica se eligió el siguiente método. Se parte de una aproximación inicial $(\underline{t}_0, \underline{v}_0)$ y en forma recurrente se calculan $(\underline{t}_n, \underline{v}_n)$ mediante las fórmulas

$$\underline{t}_{n+1} = G(\underline{t}_n, \underline{v}_n) \quad (4.1.4)$$

$$\underline{v}_{n+1} = F(\underline{t}_n, \underline{v}_n) \quad (4.1.5)$$

o bien, en vez de (4.1.5)

$$\underline{v}_{n+1} = F^\circ(\underline{t}_n, \underline{v}_n) . \quad (4.1.5')$$

Obviamente, si $(\underline{t}_n, \underline{v}_n)$ tiene límite, esto será necesariamente una solución. No he podido probar esta convergencia, pero la he verificado empíricamente para ambos procedimientos en todos los experimentos numéricos realizados. El método basado en (4.1.5') mostró ser bastante más rápido que el que utiliza (4.1.5), especialmente en muestras de distribuciones de colas largas, y por este motivo fue adoptado para la simulación descrita en la sección siguiente.

Esta forma de resolver la ecuación está inspirada en el algoritmo dado en página 17 de (Andrews y otros, 1972) para MVST(1,1). Probablemente este método es un tanto "ingenuo" y puede ser mejorado, pero no he tenido tiempo de hallar un algoritmo más sofisticado.

Como todas las matrices que aparecen son simétricas y definitivas positivas, el cálculo de las inversas y de las distancias $d^2(\underline{x}, \underline{t}, \underline{V}^{-1})$ se hace mediante la factorización de Choleski.

Se tomarán como $(\underline{t}_0, \underline{v}_0)$ la media y la matriz de covarianzas muestrales. El procedimiento iterativo continúa hasta que los valores absolutos de todos los elementos de $\underline{t}_{n+1} - \underline{t}_n$ y de $\underline{v}_{n+1} - \underline{v}_n$ se hacen menores que una cota prefijada.

4.2.- Método de Monte Carlo para variables Normal/Independiente

Para comparar los comportamientos de varios $\hat{\mu}$ -estimadores para muestras de tamaño n de distintas distribuciones radiales, hay que tener en cuenta que, por simetría, la distribución de $\underline{t} = \underline{t}_n$ es radial, y por lo tanto basta estudiar la distribución de $|t|$ además (si n no es muy chico y la distribución de la muestra no tiene colas demasiado pesadas) la distribución de $n^{1/2}\underline{t}$ debiera ser aproximadamente normal, de modo que $n|t|^2$ será aproximadamente un múltiplo de una χ^2 con n grados de libertad. Para comparar los estimadores bastará en consecuencia calcular algún parámetro de escala de la distribución de $n|t|^2$.

Los valores que se decidió estimar son.

i) La varianza: $v = E n|t|^2 / m$.

ii) Percentiles extremos de la distribución convenientemente normalizados. Sean ξ_p el percentil $1-p$ de $n|t|^2$, y χ_p^2 el correspondiente percentil de la distribución χ^2 con n grados de libertad. Se define la p -pseudo varianza como $PV_p = \xi_p / \chi_p^2$.

Obviamente, para estimar ξ_p es necesario estimar primero de alguna forma la función de distribución G de $n|t|^2$ y luego interpolar.

iii) Para tener una idea de la normalidad de las colas del estimador, se define el índice de no normalidad

$III_{pc} = PV_p / PV_q$, donde $p < q$. Este índice vale 1 si \underline{t} es normal, y será mayor que 1 si su distribución tiene

colas pesadas. Su importancia radica en que la comparación a través de parámetros de escala tiene un sentido claro sólo si las distribuciones de los estimadores son aproximadamente del mismo tipo, y por eso importa verificar esta suposición.

iv) Para la matriz de escala, interesa ver la aproximación de la "forma" de \underline{V} a la "correcta", que en este caso es radial. Se buscó para medir la deformación la distribución de una función de los autovalores de \underline{V} , $DEF(\underline{V})$, que alcanzara sus extremos cuando todos los autovalores son iguales y cuando alguno es nulo, y tal que $DEF(c\underline{V}) = DEF(\underline{V})$ para toda $c > 0$. Además conviene que DEF se pueda calcular sin tener que calcular explícitamente los autovalores de \underline{V} , lo que consumiría demasiado tiempo de máquina. La única función que se pudo hallar con esas características fue

$$DEF(\underline{V}) = (\det \underline{V})^{1/m} / (\text{tr } \underline{V}/m)$$

que está entre 0 y 1, y alcanza estos extremos cuando \underline{V} es singular, y cuando es un múltiplo de \underline{I} , respectivamente.

Las distribuciones usadas corresponden a variables m-dimensionales de la forma $\underline{x} = \underline{z} / y$, donde \underline{z} e y son independientes, \underline{z} es $N(\underline{0}, \underline{I})$, e y es positiva. Estas variables son fáciles de generar, comprenden una amplia gama de distribuciones (Normal, Student, Normal con distintos tipos de contaminación) y

-lo que es muy importante- las particulares características de la normal permiten mejorar la precisión de la simulación, según las ideas expuestas en (Podges, 1967) y (Andrews y otros, 1972), es decir, mejorando el muestreo mediante una adecuada estratificación, de la forma siguiente.

Sean: z_1, \dots, z_n variables independientes en R^m , todas $N(\underline{0}, I)$, y_1, \dots, y_n variables independientes, equidistribuidas, positivas, independientes de las z_i , $x_i = z_i / y_i$ ($i=1, \dots, n$), e \underline{y} el vector de R^n con coordenadas (y_1, \dots, y_n) . La distribución de las x_i condicional en \underline{y} (o sea, con las y_i fijas) es normal, con densidad

$$f_{\underline{y}}(x_1, \dots, x_n) = K \exp \left\{ (-1/2) \sum_{i=1}^n y_i^2 |x_i|^2 \right\}$$

Considerando la familia de distribuciones de las variables $\alpha x_i + \underline{\mu}$ (donde los parámetros α y $\underline{\mu}$ recorren R y R^m respectivamente) los estadísticos suficientes de la distribución condicional en \underline{y} son

$$\underline{x}_{\cdot} = \sum_{i=1}^n y_i^2 x_i / \sum_{i=1}^n y_i^2 \quad (4.2.1)$$

$$s_{\cdot}^2 = \sum_{i=1}^n y_i^2 |x_i - \underline{x}_{\cdot}|^2 / (n-1) \quad (4.2.2)$$

Condicionales en \underline{y} la distribución marginal de \underline{x}_{\cdot} es $N(\underline{0}, a_{\underline{y}} I)$, con $a_{\underline{y}} = 1 / \sum_{i=1}^n y_i^2$ la de $(n-1) s_{\cdot}^2$ es χ^2 con $(n-1)m$ grados de libertad y \underline{x}_{\cdot} y s_{\cdot}^2 son independientes.

Sean $\underline{c}_i = (\underline{x}_i - \underline{x}_0)/s_0$ ($i = 1, \dots, n$) y \underline{c} el vector de $(\mathbb{R}^m)^n$ cuyas coordenadas son $(\underline{c}_1, \dots, \underline{c}_n)$ (lo que en (Andrews y otros, 1972) es llamado "configuración invariante"). Como la distribución condicional de \underline{c} no depende del parámetro $(\alpha, \underline{\mu})$, \underline{c} es condicionalmente independiente del estadístico suficiente (\underline{x}_0, s_0^2) por el Lema de Basu (Lehmann 1959)

Consideremos ahora a \underline{t} aplicado a la muestra $\underline{x}_1, \dots, \underline{x}_n$ como una función de n variables vectoriales. Por la invariancia de \underline{t} :

$$\underline{t}(\underline{x}_1, \dots, \underline{x}_n) = s_0 \underline{t}(\underline{c}_1, \dots, \underline{c}_n) + \underline{x}_0 \quad (4.2.3)$$

Para estimar v se usa

$$\begin{aligned} E\{|\underline{t}(\underline{x}_1, \dots, \underline{x}_n)|^2 \mid \underline{c}, \underline{y}\} &= E\{s_0^2 |\underline{t}(\underline{c}_1, \dots, \underline{c}_n)|^2 + \\ &+ |\underline{x}_0|^2 + 2 s_0 \underline{x}_0' \underline{t}(\underline{c}_1, \dots, \underline{c}_n) \mid \underline{c}, \underline{y}\} = \\ &= m |\underline{t}(\underline{c}_1, \dots, \underline{c}_n)|^2 + m a_{\underline{y}} + 0 \end{aligned}$$

Y tomando expectación en la igualdad anterior

$$E |\underline{t}(\underline{x}_1, \dots, \underline{x}_n)|^2 = m E |\underline{t}(\underline{c}_1, \dots, \underline{c}_n)|^2 + m E a_{\underline{y}}$$

El cálculo de $E a_{\underline{y}}$ se hace analíticamente, conociendo la distribución de las \underline{y}_i . Para estimar $E |\underline{t}(\underline{c}_1, \dots, \underline{c}_n)|$ se

generan N muestras, y de cada una se obtiene un valor de $\underline{t}(\underline{c}_1, \dots, \underline{c}_n)$, que llamaremos \underline{t}_j ($j = 1, \dots, N$).

De esta forma el estimador de v , v^* , es

$$v^* = n \left\{ (1/N) \sum_{j=1}^N |\underline{t}_j|^2 + E \varepsilon_{\underline{v}} \right\}$$

Para estimar la precisión de v^* se requiere

$$\text{Var}(v^*) = n^2 \text{Var}(|\underline{t}(\underline{c}_1, \dots, \underline{c}_n)|^2) / N$$

lo cual se estima usando

$$\text{Var}(|\underline{t}(\underline{c}_1, \dots, \underline{c}_n)|^2) \approx \sum_{j=1}^N |\underline{t}_j|^4 / N - \left(\sum_{j=1}^N |\underline{t}_j|^2 / N \right)^2 \quad (4.2.4)$$

En cuanto a la función de distribución, mucho mejor que la estimación "ingenua" de G mediante la función de distribución empírica -especialmente imprecisa en las colas, o sea en los percentiles que interesan- es usar métodos del tipo descrito anteriormente. Como $G(r) = E I(n|\underline{t}|^2 < r)$, la idea es descomponer esta media tomando primero medias condicionales.

Sean esta vez $\underline{b}_i = \underline{x}_i - \underline{x}_0$ ($i = 1, \dots, n$) y \underline{b} el vector de $(\mathbb{R}^m)^n$ con coordenadas $(\underline{b}_1, \dots, \underline{b}_n)$ de modo que

$$\underline{t}(\underline{x}_1, \dots, \underline{x}_n) = \underline{t}(\underline{b}_1, \dots, \underline{b}_n) + \underline{x}_0$$

Condicionamente en $(\underline{b}, \underline{y})$ la distribución del segundo miembro es $N(\underline{t}(\underline{b}_1, \dots, \underline{b}_n) / a_{\underline{y}} \mid I)$ y por lo tanto la de $|\underline{t}(\underline{x}_1, \dots, \underline{x}_n)|^2 / a_{\underline{y}}$ es χ^2 no central, con m grados de libertad y parámetro de no-centralidad $\lambda = \lambda(\underline{b}, \underline{y}) = |\underline{t}(\underline{b}_1, \dots, \underline{b}_n)|^2 / a_{\underline{y}}$. De modo que si $F_{m, \lambda}$ es la función de distribución de la χ^2 no central con m grados de libertad y parámetro de no-centralidad λ

$$P \{n|\underline{t}(\underline{x}_1, \dots, \underline{x}_n)|^2 < r \mid \underline{b}, \underline{y}\} = F_{m, \lambda}(r/na_{\underline{y}})$$

Por lo tanto, para cada r

$$G(r) = E F_{m, \lambda}(\underline{b}, \underline{y}) (r/na_{\underline{y}})$$

que se puede estimar muestreando sobre los \underline{b}_i y los \underline{y}_i . En el experimento se fijaron unos pocos valores de r , determinados de antemano para que estuvieran cerca de los percentiles deseados, los que se obtuvieron luego por interpolación. La $F_{m, \lambda}$ se calculó mediante la aproximación de Pearson (1959). La varianza del estimador $G^*(r)$ se estimó usando la misma idea que en (4.2.4).

Naturalmente, mejor hubiera sido aprovechar la invariancia y trabajar con la descomposición (4.2.3), pero la distribución condicional del cuadrado de la norma del segundo miembro dados \underline{c} , \underline{y} resulta muchísimo más complicada (Guenther y Terragno, 1964) y el aumento del tiempo de computación hubiera compensado la menor cantidad de muestras requerida.

Para estimar la distribución de DEF no parece posible aprovechar la estructura de estas distribuciones, de modo que se utilizó el método "ingenuo", calculando directamente la función de distribución empírica.

4.3.- Resultados de la simulación

Para que el tiempo de máquina utilizado para el experimento no fuera excesivo, hubo que limitarse al caso bidimensional. Para poder realizar futuras simulaciones en dimensión mayor, será imprescindible entre otras cosas mejorar notablemente el algoritmo para el cálculo del estimador.

Se utilizaron los estimadores de Huber $H(p)$ con los tres valores de $p = 0$ (media y matriz de covarianzas muestrales), 0.30 y 0.50 las distribuciones Normal y ST(3), y los tamaños de muestra $n = 10$ y $n = 20$. El número N de replicaciones fue 400 para la Normal, 1500 para ST(3) con $n = 20$, y 2000 para ST(3) con $n = 10$. El tiempo total requerido fue de 3 horas de máquina (IBM 360/50).

Se estimaron los percentiles ξ_p para los cuatro valores de p : 0.25, 0.10, 0.50 y 0.025. Para ello se estimó primero la función de distribución $G(r)$ en los cuatro puntos $r = v_\infty \chi_p^2$ (donde v_∞ es la varianza asintótica de \underline{t} , ya calculada en la sección 3.4) dado que debiera ser $v_\infty \chi_p^2 \approx \xi_p$. Para los valores obtenidos, se observó que $n^{-1} \log(1 - G^*(r))$ era prácticamente constante, cosa razonable si se considera que G debiera ser aproximadamente una χ^2 con $m = 2$ grados de libertad, o sea una distribución exponencial. Por lo tanto se consideró que se obtendría un estimador aproximadamente insesgado ξ_p^* interpolando

Finalmente los valores de $\log(1 - G^*(r))$. Dado que ξ_p^* es una función sencilla de las variables $G^*(r)$ -cuya distribución conjunta es aproximadamente normal- se estimó la varianza de ξ_p^* mediante una conocida fórmula asintótica (Rao (1965), pág. 321).

En la Tabla 3 se exhiben los valores de v_p y los de PV_p y $INN_{p\alpha}$ para $p = 0.05$ y $q = 0.25$. En el caso de la media para $ST(3)$ debió calcularse $\xi_{0.25}$ por extrapolación. Luego de cada valor se da entre paréntesis su desviación standard en unidades de la última cifra decimal de modo que por ejemplo, 1.609(14) significa un valor observado de 1.609 con una $\sigma = 0.014$.

Para DEF se calculó la función de distribución empírica $H(r)$ con $r \in (0,1)$ a intervalos de 0.02. En la Tabla 3 se exhiben los percentiles 0.25 y 0.50 de H , DEF_{25} y DEF_{50} , que fueron calculados directamente por interpolación lineal. La desviación standard de éstos, estimada con el mismo método que para los ξ_p , es en todos los casos menor que 0.01 pero desconoce la magnitud del sesgo introducido por la interpolación lineal.

TABLA 3

Comparación de estimadores para muestras finitas

	<u>v</u>	<u>PV</u>	<u>IEH</u>	<u>DEF₅₀</u>	<u>DEF₂₅</u>
<u>Normal, n = 10</u>					
H(0.0)	1.0 (0)	1.0 (0)	1.0	0.83	0.68
H(0.3)	1.051(3)	1.050(3)	1.0	0.80	0.64
H(0.5)	1.103(6)	1.100(5)	1.0	0.77	0.59
<u>ST(3), n = 10</u>					
H(0.0)	2.78(11)	2.90(7)	1.27	0.74	0.56
H(0.3)	1.663(14)	1.753(21)	1.09	0.78	0.62
H(0.5)	1.572(12)	1.651(18)	1.08	0.76	0.57
<u>Normal, n = 20</u>					
H(0.0)	1.0(0)	1.0(0)	1.0	0.93	0.87
H(0.3)	1.048(3)	1.045(3)	1.0	0.91	0.85
H(0.5)	1.094(5)	1.093(4)	1.0	0.90	0.82
<u>ST(3), n = 20</u>					
H(0.0)	2.92(9)	3.06(6)	1.18	0.80	0.71
H(0.3)	1.609(14)	1.647(17)	1.04	0.89	0.80
H(0.5)	1.528(11)	1.564(15)	1.0	0.87	0.79

De los resultados se puede observar que $H(0.3)$ y $H(0.5)$, sin perder mucha eficiencia para la normal, dan mucho mejores resultados que la media para $ST(3)$. Aquí $H(0.5)$ no resulta notablemente mejor que $H(0.3)$ para esta última distribución, pero haría falta incluir en el estudio otras distribuciones con colas más pesadas para ver mejor la diferencia entre ambos estimadores de localización. Nótese también que el INH es bastante bajo, y que v y PV son muy semejantes, aún para $n = 10$, lo que indica que la teoría asintótica da aproximaciones bastante buenas para muestras pequeñas -al menos para la Normal y $ST(3)$.

En cuanto a los percentiles de DEF, no se observan diferencias notables entre los estimadores de dispersión. Parece inevitable concluir que, o bien los comportamientos de éstos son efectivamente parecidos, o bien que hubiera sido necesario tomar percentiles más extremos -y por lo tanto más imprecisos si se usa la función de distribución empírica- o bien que DEF no es un estadístico adecuado para hacer la comparación.

Una idea precisa de la aproximación de los resultados obtenidos la da el hecho de que para la media aplicada a $ST(3)$, el valor exacto de v es 3, para todo n .

Los resultados también ponen de manifiesto las ventajas del procedimiento de estimación expuesto en la sección anterior. Sea T_{ij} el estimador "ingenuo" de v , o sea, el promedio de los m valores observados de $n |t|^{2j} / m$. Suponiendo t aproximadamente normal, debe ser $\text{Var}(|t|^{2j}) \approx (2/m) (E |t|^{2j})^2$, por una conocida propiedad de la χ^2 . Por lo tanto, para $m = 2$ es

$\sigma = (\text{Var } T_H)^{1/2} \approx N^{-1/2}$ v. Por ejemplo si se desea una pre cisión comparable a la obtenida para $H(0.3)$ aplicado a la Hor mal con $n = 1-$, debería ser $\sigma \approx 1.05 N^{-1/2} \approx 0.003$, lo que implica $N \approx 100000$. Para la media muestral aplicada a $ST(3)$ es todavía $\text{Var } T_H = \infty$. Cifras análogas valen para la preci sión de los percentiles.

4.4.- Cuestiones pendientes

Hay varias preguntas que han quedado sin respuesta, y que requerirán trabajo futuro, a saber:

1) Es fundamental hallar un algoritmo realmente eficiente para el cálculo numérico de los estimadores, tanto para facili tar y hacer atractivo el uso de éstos en aplicaciones, como para posibilitar la prosecución del estudio de sus propiedades, parti cularmente para dimensión elevada y muestras no muy grandes, don de la teoría asintótica podría ser poco útil. Además haría falta examinar los problemas numéricos que puedan aparecer cuando la matriz V es "casi singular".

2) Como ya se mencionara al principio del trabajo, uno de mis objetivos era estudiar la aplicación de estos estimadores a Discriminación Lineal y Componentes Principales. Muy probablemente sea el de Monte Carlo el único método posible, ya que la teoría asintótica resulta ya bastante difícil para el caso más simple de media y matriz de covarianzas, con distribuciones normales.

3) Para la realización de tests aproximados de locación interesa definir un estimador adecuado, \underline{t} , de la matriz de covarianzas asintótica de \underline{t} , y estudiar la distribución del estadístico "Studentizado" (o más bien "Hotellingzador") $T^2 = \underline{t}'\underline{V}^{-1}\underline{t}$ para muestras pequeñas.

4) Como se observó en la sección 3.3, los estimadores estudiados tienen una cota de derrumbe muy pequeña para dimensión elevada. Para entender el significado práctico de este hecho será necesario experimentar una versión finita de δ^* , contaminando muestras finitas con observaciones ubicadas a distancia grande (finita). Queda además el problema teórico de ver si alguna modificación de los estimadores permite suisanar radicalmente el inconveniente.

5) Al final de la sección 2.1 se observó que para m grande, la matriz \underline{V} podría ser singular si el tamaño de muestra n no era bastante mayor que m . Hace falta analizar cuándo se presenta realmente esta dificultad.

6) Hay que recordar que todo lo realizado se refiere al caso más simple: el radial. Dado que los datos multivariados suelen tener estructuras más complicadas, sería útil ver lo que sucede al apartarse de este caso, y comprender claramente qué es lo que se quería estimar entonces. El problema no resulta muy claro aún para el caso univariado.

7) No se han considerado aquí las cuestiones de optimalidad, que no parecen simples. Huber (1964) halló los estimadores

asintóticamente minimax de localización con escala fija, y de escala con localización fija, para distribuciones normales contaminadas simétricamente, en el caso univariado. En un manuscrito inédito, el mismo autor exhibe un procedimiento heurístico para hallar el estimador minimax de localización en el caso multivariado radial. La solución involucra funciones de Bessel y de Heumann (según la paridad de n) y por lo tanto no resulta muy manejable salvo para $m = 3$.

El problema de la estimación óptima de la matriz de dispersión es interesante, pero primero habría que definir adecuadamente cuál es lo que se quiere optimizar.

REFERENCIAS

Nota: se usará la abreviatura A.M.S. para "Annals of Mathematical Statistics".

Anderson, T.W. (1958). An Introduction to Multivariate Statistical Analysis. Wiley, New York.

Andrews, D.F., Bickel, P.J., Hannaford, F.R., Huffer, P.J., Rogers, M.H., y Tukey, J.W. (1972). Robust Estimates of Location: Survey and Advances. Princeton University Press.

Bickel, P.J. (1964). On some alternative estimates for shift in the p -variate one sample problem. A.M.S. 35, 1075-90.

Cartan, H. (1972). Cálculo Diferencial. Ed. Omega, Barcelona.

Chernoff, H., Gastwirth, J.L., y Johns, M.V. (1967). Asymptotic distribution of linear combinations of functions of order statistics with applications to estimation. A.M.S. 38, 52-72

Dunford, N. y Schwartz, J. (1958). Linear Operators. Interscience, New York.

Gentleman, W.H. (1965). Robust estimation of multivariate location by minimizing p -th power deviations. Tesis de doctorado. Princeton University.

Gnanadesikan, R. y Kettenring, J.R. (1972). Robust estimates, residuals, and outlier rejection with multivariate data. Biometrics 28, 81-124.

-)
Guentner, H.C., y Terragno, P.J. (1964). A review of the literature on a class of coverage problems. A.M.S. 35, 232-260.
- Hampel, F.R. (1968). Contributions to the theory of robust estimation. Tesis Doctoral. University of California, Berkeley.
- (1971). A general qualitative definition of robustness. A.M.S. 42, 1887-96.
- Hodges, J.L. (1967). Efficiency in normal samples and tolerance of extreme values for some estimates of location. Proc. Fifth Berkeley Symp. Math. Stat. and Prob. 1, 163-186.
- y Lehmann, E.L. (1963). Estimates of location based on rank tests. A.M.S. 34, 598-611
- Hoel, P.G. y Peterson, R.P. (1949). A solution to the problem of optimum allocation. A.M.S. 20, 433-438.
- Muler, P.J. (1964). Robust estimation of a location parameter. A.M.S. 35, 73-101.
- (1967). The behavior of maximum likelihood estimates under nonstandard conditions. Proc. Fifth Berkeley Symp. Math. Stat. Prob. 1, 221-233.
- (1972). Robust statistics: A review. A.M.S. 43, 1041-1067.
- Lehmann, E.L. (1959). Testing Statistical Hypotheses. Wiley. New York.
- Pearson, E.S. (1959). Note on an approximation to the distribution of non-central χ^2 . Biometrika 46, 364.

- Rao, C.R. (1965). Linear Statistical Inference and Its Applications. Wiley, New York.
- Relles, D.A. (1968). Robust regression by modified least squares. Tesis Doctoral. Yale University.
- Romanowski, H. y Green, E. (1965). Practical applications of the modified normal distribution. Bull. Géodésique 76, 1-20.
- Sen, P.K. y Puri, M.L. (1971). Nonparametric Methods in Multivariate Analysis. Wiley. New York.
- Tukey, J.W. (1940). Sampling from contaminated distributions, preliminary report (abstract). A.M.S. 17, 501.
- (1960). A survey of sampling from contaminated distributions. Contributions to Probability and Statistics ed. I. Olkin. Stanford University Press.
- Helfowitz, J. (1954). Generalization of the theorem of Glivenko-Cantelli. A.M.S. 25, 131-138.