QUINTA
JORNADA DE
MATEMATICA
SEPTIEMBRE
TEMUCO

METODOS ROBUSTOS

(CURSO B2)

Dr. Héctor Allende
(U.T.F.S.M.)

PROCEDIMIENTOS ESTADISTICOS ROBUSTOS

1.- INTRODUCCION

El concepto de robustez fue introducido por primera vez por G. Box (1953), en un trabajo sobre estimación de la varianza para el caso no gaussiano. El se ocupa de la insensibilidad de los procedimientos estadísticos con respecto a pequeñas desviaciones del modelo estadístico supuesto y muy pocas veces contrastado. A partir de este trabajo se han dedicado, en casi todas las áreas de la estadística, grandes esfuerzos para precisar este nuevo concepto y para desarrollar procedimientos estadísticos robustos.

La estadística paramétrica, asume que sus datos provienen de una distribución conocida desarrollando procedimientos óptimos para modelos paramétricos exactos, y además supone que el comportamiento sigue siendo adecuado cuando el modelo es válido sólo aproximadamente. Este supuesto de continuidad no se cumple en muchos procedimientos clásicos, por lo que es necesario desarrollar métodos estadísticos robustos.

En el problema de estimación paramétrica puntual se asumen 3 supuestos (el supuesto de independencia, el supuesto de identidad de la distribución y el supuesto de la distribución). En la práctica cada uno de estos supuestos puede ser válido sólo aproximadamente. Por lo tanto es necesario obtener métodos robustos frente a cada uno de estos supuestos, es decir, robustez con respecto al supuesto de independencia, robustez con respecto al supuesto de identidad de la distribución y robustez con respecto al supuesto de una distribución (robustez distribucional).

En estas notas el concepto de robustez tratado será robustez distribucional, es decir robustez en el sentido de Huber (1981), este es el caso más importante y mejor desarrollado a la fecha. Con respecto a la robustez de los supuestos de independencia, distribución idéntica y aleatoriedad se conocen muy pocos trabajos, entre otros motivos es debido a que la falta de robustez en tal sentido es menos grave que en el caso de robustez distribucional (ver Portnoy (1977)).

1.1.- ALGUNOS ESTIMADORES DEL PARAMETRO DE LOCALIZACION

a) $\bar{X}_n = (1/n)\cdot \sum_{i=1}^{n} X_i$ (media aritmética)

b) $\mathrm{Med} = \begin{cases} X(\tfrac{1}{2}(n+1)) , & \text{si } n \text{ es impar} \\ \tfrac{1}{2}[X(\tfrac{1}{2}n) + X(\tfrac{1}{2}(n+2))] , & \text{si } n \text{ es par} \end{cases}$ (mediana)

c) $M_k = (1/(n-2k))\cdot \sum_{i=k+1}^{n-k} X_{(i)}$ (media truncada de orden k)

d) $W_k = (1/n)\{k\cdot X_{(k+1)} + k\cdot X_{(n-k)} + \sum_{i=k+1}^{n-k} X_{(i)}\}$ (media winzorizada)

En un sentido amplio podemos decir que $\bar{X}_n$ no es robusta, debida a su alta sensibilidad a los valores extremos o atípicos, en cambio la mediana, la media truncada de orden k y la media winzorizada de orden k son robustas ya que no son afectadas por unos pocos valores extremos de los datos.

1.2.- ALGUNOS ESTIMADORES DEL PARAMETRO DE ESCALA

Es conocida disputa entre Eddington (1914) y Fisher (1920) sobre los méritos de dos medidas de dispersión con respecto al promedio, la comparación de estimadores del parámetro de escala puede ser estudiada basada en algunas ideas de robustez.

Sea $X_1, X_2, \cdots, X_n$ una muestra aleatoria proveniente de una distribución

$$F(x) = (1-\epsilon)\cdot\Phi((x-\mu)/\sigma) + \epsilon\cdot\Phi((x-\mu)/3\sigma),$$

donde Φ, es la función de distribución de la normal estándar.

La medida de dispersión propuesta por Eddington fue la desviación media

$$d_n = (1/n)\cdot \sum_{i=1}^{n} |X_i - \bar{X}|,$$

mientras que la propuesta por Fisher fue la desviación estándar,

$$S_n = \left[(1/n)\cdot \sum_{i=1}^{n} (X_i - \bar{X})^2\right]^{1/2}$$

Fisher concluyó con la disputa demostrando que bajo el supuesto de normalidad S_n es un 12% más eficiente que d_n. Puesto que si $\epsilon=0$, tenemos

$$S_n \xrightarrow{n \to \infty} \sigma \quad , \text{ y}$$

$$d_n \xrightarrow{n \to \infty} \sqrt{(2/\pi)}\cdot\sigma$$

Comparando estos estimadores sobre la base de la eficiencia relativa asintótica (ERA) obtenemos:

$$ERA(\epsilon) = \lim_{n \to \infty} \frac{VAR[S_n]/E^2[S_n]}{VAR[d_n]/E^2[d_n]} = \frac{\frac{1}{4}(3(1+80\epsilon)/(1+8\epsilon)^2 - 1)}{\frac{1}{2}\pi(1+8\epsilon)/(1+2\epsilon)^2 - 1}$$

ϵ	0	0.002	0.005	0.01	0.02	0.05	0.1	0.15	0.25	0.5	1.0
ERA	0.876	1.016	1.198	1.439	1.752	2.035	1.903	1.689	1.371	1.017	0.876

(Ejemplo tomado de Huber (1981))

La tabla muestra que bastará $\epsilon=0.002$ (2 observaciones en 1000) para que la ventaja de S_n respecto de d_n desaparezca. Por otro lado para $\epsilon=0.05$, la ventaja de d_n con respecto a S_n es evidente.

Muchos datos en las Ciencias Físicas son modelados adecuadamente por una distribución de la forma

$$F(x) = (1-\epsilon)\cdot\Phi(x) + \epsilon\cdot\Phi(x/a), \qquad a >> 1$$

donde Φ es la distribución acumulada de la $N(0,1)$.

El ejemplo anterior muestras que alargando las colas de la distribución, S_n se ve muy afectado pero no así d_n. Es decir, d_n es más resistente a la presencia de datos atipicos (outliers). Para muchos efectos prácticos es posible intercambiar "robustez distribucional" y resistencia.

Finalmente, otra alternativa para estimar robustamente el parámetro de escala es la Mediana de las desviaciones absoluta MAD (median absolute desviation) propuesta por Andrews (1972)

$$MAD_n = (1/0.6745)(MAD(X_i)) = (1/0.6745)(Med_i(|X_i - Med_j(X_j)|)),$$

donde $0.6745 = \Phi^{-1}(3/4)$ es un factor de normalización para lograr consistencia del estimador bajo distribución normal.

2.- CONTINUIDAD Y ROBUSTEZ CUALITATIVA

Hampel (1971) introdujo el concepto de robustez cualitativa asociado al problema de estimación puntual: la idea básica es que un pequeño cambio en la distribución del modelo F asumido, debería causar sólo pequeños cambios en la distribución de la función de estimacion $L_F(T_n)$.

DEF 2.1 : Sean $X_1, X_2, \cdots, X_n$ sucesión de variables aleatorias independientes e idénticamente distribuidas con distribución común F_0, $T_n = T_n(X_1, X_2, \cdots, X_n)$ sucesión de estimadores o test. Se dice que (T_n) es cualitativamente robusta en F_0 si para cada $\epsilon>0$, $\exists\, \delta>0$, $\forall\, n_0>0$: $\forall\, F$, $\forall\, n>n_0$ $d_L(F_0,F)<\delta \implies d_L(L_{F_0}(T_n), L_F(T_n))<\epsilon$, donde d_L es una métrica adecuada.

OBS 1 : Si (T_n) se deriva de un funcional $T_n = T(F_n)$, entonces esta definición es equivalente a la continuidad débil de T.

OBS 2 : La métrica d_L es una métrica adecuada en el espacio de todas las medidas de Probabilidad (Distribuciones).

3.- LA NOTACION FUNCIONAL DE ESTIMADORES

En la mayoría de los casos comunes los Test estadísticos y estimadores dependen de una muestra $(X_1, X_2, \cdots, X_n)$ sólamente a través de la distribución empírica F_n.

DEF 3.1 : Sean $x_1, x_2, \cdots, x_n \in \mathbb{R}$, valores de una muestra aleatoria $X_1, X_2, \cdots, X_n$ sobre $\mathbb{R}$. Se llama Distribución empírica a

$$F_n(x) = (1/n)\cdot \sum_{i=1}^{n} I_{(x_i<x)}$$

donde I es la función indicatriz.

OBS 1 : Para espacios muestrales más generales F_n puede ser definida a través de la medida empírica

$$F_n = (1/n)\cdot \sum_{i=1}^{n} \delta_{x_i}$$

donde δ_{x_i} es una distribución que asigna probabilidad 1 al punto x_i (Medida de Dirac).

OBS 2 : En caso que $T_n = T_n(X_1, X_2, \cdots, X_n)$ dependa de la muestra sólo a través de F_n, podemos escribir

$$T_n(X_1, X_2, \cdots, X_n) = T(F_n)$$

Ejemplos :

$$\text{i) } T(F) = \int X\cdot dF = E[X]$$

$$T(F_n) = \int X\cdot dF_n = \overline{X}_n$$

ii) $T(F) = \int (X^2 - E[X])^2 \cdot dF = V[X]$

$T(F_n) = \int (X - E[X])^2 \cdot dF_n = (^1/_n) \cdot \sum_{i=1}^{n} (X_i)^2 - (\bar{X}_n)^2$

iii) $T(F) = F^{-1}(½)$

$T(F_n) = F_n^{-1}(½) = \text{Mediana } \{x_1, \cdots, x_n\}$

iv) $T(F) = (^1/_{(1-2\alpha)}) \cdot \int_{\alpha}^{1-\alpha} F^{-1}(t) \cdot dt$

$T(F_n) = (^1/_{(1-2\alpha)}) \cdot \int_{\alpha}^{1-\alpha} F_n^{-1}(t) \cdot dt = \bar{X}_\alpha$

v) El test estadístico de el Lema de Neyman-Pearson.
El test de potencia máxima entre dos densidades f_0 y f_1 está basado en una estadística de la forma

$$T(F_n) = \int \Psi(X) \cdot dF_n = (^1/_n) \cdot \Sigma\, \Psi(x_i)$$

$$\text{donde } \Psi(X) = \log\left(\frac{f_1(x)}{f_0(x)}\right)$$

iv) El estimador máximo verosímil de θ de una familia de densidades $f(x,\theta)$ es la solución de

$$T(F_n) = \int \Psi(X) \cdot dF_n = 0$$

$$\text{donde } \Psi(X) = \frac{d}{d\theta} \log f(x,\theta)$$

4.- ROBUSTEZ CUANTITATIVA

4.1.- FUNCION DE INFLUENCIA

Hampel (1968) introdujo una medida de robustez local basada en la derivada de Gâteaux de Funcionales llamada función de influencia.

DEF 4.1 : Si $T_n = T_n(x_1, \cdots, x_n)$ admite una representación $T_n = T(F_n)$, diremos que T_n está definido por un funcional T definido sobre un subconjunto del espacio de todas las probabilidades ($Z(\mathbb{R})$).

DEF 1.2 : Sea $T(F_n)$ un estimador o test, la función de influencia de T en F está dada por

$$IC(X,F,T) = \frac{d}{dt} T(F_t)\Big|_{t=0}$$

donde $F_t = (1-t)\cdot F + t\cdot\delta_x$, con δ_x distribución que asigna probabilidad 1 al origen.

OBS 1 : La función de influencia es la derivada de Gâteaux de T en F en la dirección de δ_x.

Ejemplos :

i) $T(F) = \int X\cdot dF$

$$IC(X,F,T) = \frac{d}{dt} T(F_t)\Big|_{t=0} = \frac{d}{dt}\left((1-t)\cdot T(F) + t\cdot T(\delta_x)\right)\Big|_{t=0}$$

$$= -T(F) + X = X - \int X\cdot dF = X - E_F[X]$$

ii) $T(F) = F^{-1}(\frac{1}{2})$

$$IC(Y,F,T) = \frac{d}{dt} T(F_t)\Big|_{t=0} = \frac{d}{dt} F_t^{-1}(\tfrac{1}{2})\Big|_{t=0}$$

$$\tfrac{1}{2} = F_t(F_t^{-1}(\tfrac{1}{2}))$$

$$0 = \frac{d}{dt}\left(F_t(F_t^{-1}(\tfrac{1}{2}))\right)\Big|_{t=0}$$

$$0 = \frac{d}{dt}\left\{(1-t)\cdot F(F_t^{-1}(\tfrac{1}{2})) + t\cdot\delta_y(F_t^{-1}(\tfrac{1}{2}))\right\}$$

$$0 = -F(F^{-1}(\tfrac{1}{2})) + f(F^{-1}(\tfrac{1}{2}))\cdot\frac{d}{dt} F_t^{-1}(\tfrac{1}{2})\Big|_{t=0} + \delta_y(F^{-1}(\tfrac{1}{2}))$$

$$IC(Y,F,T) = \frac{\frac{1}{2} - \delta_y(F^{-1}(\frac{1}{2}))}{f(F^{-1}(\frac{1}{2}))}$$

$$= \begin{cases} -\dfrac{1}{2f(F^{-1}(\frac{1}{2}))} & y \leq F^{-1}(\frac{1}{2}) \\ \dfrac{1}{2f(F^{-1}(\frac{1}{2}))} & y > F^{-1}(\frac{1}{2}) \end{cases}$$

1.2.- APLICACION DE LA FUNCION DE INFLUENCIA

a) Bajo condiciones de regularidad es posible linealizar T en las cercanías de F en términos de IC(X,F,T)

$$T(F_n) - T(F) = \int IC(X,F,T)\cdot(dF_n - dF) + R_n$$

como $\int IC(X,F,T)\cdot dF = E_F(IC) = 0$, podemos escribir

$$\sqrt{n}\cdot(T(F_n) - T(F)) \approx (1/\sqrt{n})\cdot \sum_{i=1}^{n} IC(X_i,F,T)$$

Luego usando el teorema central del limite

$$\sqrt{n}\cdot(T(F_n) - T(F)) \xrightarrow[D]{n \to \infty} N(0,V(F,T))$$

donde $V(F,T) = \int IC^2(X,F,T)\cdot dF$

Ejemplo :

$$T(F) = V(X) = E[X^2] - E[X]$$

$$T(F_n) = (1/n)\cdot \sum_{i=1}^{n} X_i^2 - \bar{X}_n^2 = S_n^2$$

$$IC(X,F,T) = (X - E_F(X))^2 - V_F(X)$$

$$V(F,T) = \int [(X - E_F(X))^2 - V_F(X)]^2\cdot dF = \mu_4 - \mu_2^2$$

a) Función de influencia caso finito. $T_n = T(F_n) = T_n(X_1,\cdots,X_n)$. El impacto de una observación suplementaria adicional de X en T está dado por

$$I(T_n,X) = T_{n+1}(X_1,\cdots,X_n,X) - T_n(X_1,\cdots,X_n)$$

$$= T[(1 - (1/n+1))\cdot F_n + (1/n+1)\cdot \delta_x] - T(F_n)$$

DEF 4.3 : La sensibilidad de T_n para una muestra $(X_1,\cdots,X_n)$ dada está definida por

$$SC(T_n,X) = \sup_{X} |I(T_n,X)| \qquad \text{(Tukey 1970)}$$

Ejemplos :

i) $T(F) = E[X]$

$T_n = \bar{X}_n$

$T_{n+1} = (1/n+1)\cdot(X + \sum_{i=1}^{n} X_i)$

$I(T_n,X) = (1/n+1)\cdot(X - \bar{X}_n)$

$SC(T_n,X) = +\infty$

ii) $T(F) = F^{-1}(\frac{1}{2})$, $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ donde $n = 2m+1$

$\text{Med}\,(X_i) = X_{(m+1)}$

$$T_{n+1}(X_1,\cdots,X_n,X) = \begin{cases} \frac{1}{2}\cdot(X_{(m)} + X_{(m+1)}) & \text{si } X < X_{(m)} \\ \frac{1}{2}\cdot(X_{(m+1)} + X_{(m+2)}) & \text{si } X \geq X_{(m+2)} \\ \frac{1}{2}\cdot(X + X_{(m+1)}) & \text{si } X_{(m)} \leq X \leq X_{(m+1)} \end{cases}$$

$$I(T_n,X) = \begin{cases} \frac{1}{2}\cdot(X_{(m)} - X_{(m+1)}) & \text{si } X < X_{(m)} \\ \frac{1}{2}\cdot(X_{(m+2)} - X_{(m+1)}) & \text{si } X \geq X_{(m+2)} \\ \frac{1}{2}\cdot(X - X_{(m+1)}) & \text{si } X_{(m)} \leq X \leq X_{(m+1)} \end{cases}$$

$$SC(T_n,X) = \sup_x |I(T_n,X)| = \frac{1}{2}\cdot(X_{(m+2)} - X_{(m+1)})$$

4.3.- OTRAS MEDIDAS DE ROBUSTEZ DERIVADAS DE LA CURVA DE INFLUENCIA

DEF 4.4 : Sea T un estimador o test en F, se llama valor extremo de sensibilidad a

$$GES(T,F) = \sup_x |IC(X,F,T)|$$

OBS 1 : Si $GES(T,F) < \infty$, entonces decimos que T es B-robusto. (Rouseeuw 1981).

OBS 2 : Un estimador T se dice Fisher consistente ssi $T(F_\theta)=\theta$, $\forall\ \theta \in \Theta$

OBS 3 : Un estimador Fisher consistente T con $GES(T,F)$ mínimo se llama B-robusto maximal. (Hampel 1986).

DEF 4.5 : Sea T un estimador o test con función de influencia $IC(X,F,T)$, entonces se llama sensibilidad local a

$$\lambda^*(T,F) = \sup_{\substack{x,y \\ x \neq y}} \frac{|IC(X,F,T) - IC(y,F,T)|}{x-y}$$

Ejemplos :

1) $T = T(F) = \int X\cdot dF$

$$GES(T,F) = \sup_x |IC(X,F,T)| = \infty$$

$\lambda^*(T,F) = 1$

2) $T = T(F) = F^{-1}(½)$

$GES(T,F) = 1/\{2\cdot f(F-1(½))\}$

$\lambda^*(T,F) = 0$

4.4.- CONFIABILIDAD GLOBAL O PUNTO DE RUPTURA

Hemos visto el concepto de función de influencia y su utilidad para formalizar varias ideas en la teoría de la robustez. Pero una limitación de esta medida es que es un concepto local. Para completar esta medida necesitamos un concepto o medida de confiabilidad global del estimador o Test. Dos medidas globales de robustez son: punto de ruptura y robustez cualitativa.

DEF 4.5 : El punto de ruptura de una sucesión de estimadores $(T_n : n \geq 1)$ en F está dada por

$$\epsilon^* := \text{Sup} \{ \epsilon \leq 1 \ ; \ \exists \text{ un conjunto compacto } K_\epsilon \subset \Theta : \pi_d(F,G) < \epsilon \Longrightarrow G(T_n \in K_\epsilon) \xrightarrow{n \to \infty} 1 \}$$

donde $\pi_d(F,G) := \text{Inf} \{ \epsilon \ ; \ F(A) \leq G(A^\epsilon) + \epsilon, \ \forall A \in \mathcal{O}\!\iota \}$

con $G(A^\epsilon) := \{ X \in \Omega : \inf_{y \in A} d(x,y) < \epsilon \}$

OBS 1 : $\pi_d(F,G)$ es la distancia de Prokorov entre dos Distribuciones de Probabilidad en $Z(\mathbb{R})$.

OBS 2 : $\Theta = R$, entonces $\epsilon^* = \text{Sup} \{ \epsilon \leq 1 \ ; \ \exists r_\epsilon : \pi_d(F,G) < \epsilon \Longrightarrow G(\{|T_n| \leq r_\epsilon\}) \xrightarrow{n \to \infty} 1 \}$

OBS 3 : El punto de ruptura debería ser formalmente denotado por $\epsilon^*((T_n \ ; \ n\geq 1),F)$ pero usualmente no depende de F.

OBS 4 : Existen algunas variantes de esta definición que reemplazan la distancia de Prokorov por la distancia de Levy o la distancia de Lipschitz.

4.5.- VERSION FINITA DEL PUNTO DE RUPTURA (Donoho-Huber (1983))

Sea $X_n := (X_1, X_2, \cdots, X_n)$ una muestra aleatoria de tamaño n y sea $T_n(X_n)$ un estimador basado en la muestra de X_n.

DEF 4.6 : El punto de ruptura de la muestra finita ϵ_n^* del estimador T_n de la muestra X_n está dado por

$$\epsilon_n^*(T_n,X_n) := (1/n)\cdot \mathrm{Max}\{ m \,; \max_{i_1,\cdots,i_m} \cdot \mathrm{Sup}_{y_1,\cdots,y_m} \cdot |T_n(z_1,\cdots z_n)| < \infty\}$$

donde la muestra $(z1,\cdots,zn)$ es obtenida reemplazando los m puntos $X_{i1},\cdots,X_{im}$ por valores arbitrarios $y_1,\cdots,y_m$.

OBS 1 : Sea $X_n^{(m)}$ la muestra contaminada obtenida a partir de X_n, donde se cambian m datos en la muestra original por $y_1,\cdots,y_m$.

Si $\beta(m,T_n,X_n) = \sup\limits_{X_n^{(m)}} \|T(X_n^{(m)}) - T(X_n)\|$

siendo $\|T(X_n^{(m)}) - T(X_n)\|$ la norma sobre $\mathbb{R}^k$.

Entonces $\epsilon_n^*(T_n,X_n) = \mathrm{Min}\{ (m/n) : \beta(m,T_n,X_n) = \infty \}$

Ejemplos :

1) Sea $T(X_n) = \sum_{i=1}^{n} (X_i/n)$

$\beta(1,T_n,X_n) = (1/n)\cdot(y_1 - X_{i1})$, y_1 arbitrario.

$\epsilon_n^*(T_n,X_n) = (1/n)$

2) Sea $T(X_n) = F_n^{-1}(½)$

$\epsilon_n^*(T_n,X_n) = (½)$

3) Sea $T(X_n) = \overline{X}_\alpha$

$\epsilon_n^*(T_n,X_n) = \alpha$

4.7.- CONTINUIDAD Y ROBUSTEZ CUALITATIVA

DEF 4.7 : Se dice que una sucesión de estimadores $\{ T_n \,;\, n\geq 1 \}$ es cualitativamente robusto en F si $\forall\ \epsilon>0$, $\exists\ \delta>0$, $\forall\ G\in Z(\mathbb{R})$, $\forall\ n \in N$: $\pi_d(F,G)<\delta \Longrightarrow \pi_d(L_F(T_n),L_G(T_n))<\epsilon$, donde π_d es la distancia de Prokorov entre dos distribuciones en $Z(\mathbb{R})$.

DEF 4.8 : Una sucesión de estimadores $\{ T_n \,;\, n\geq 1 \}$ se dice continua en F, si $\forall\ \epsilon>0$, $\exists\ \delta>0, \exists\ n_0\in N$, $\forall\ n,m\geq n_0$ $\forall\ F_n,F_m\in Z_n(\mathbb{R})$:

$$\left.\begin{array}{l} \pi_d(F,F_n)<\delta \\ \pi_d(F,F_m)<\delta \end{array}\right\} \Longrightarrow |T_n(F_n) - T_m(F_m)| < \epsilon$$

Estas dos definiciones son claramente no equivalentes, pero pueden ser relacionadas por los siguientes resultados (Hampel (1986)).

TEO 4.1 : Una sucesión de estimadores $\{T_n; n\geq 1\}$, la cual es continua en F y todos los T_n son funciones continuas de las observaciones es cualitativamente robusto (Hampel (1986)).

TEO 4.2 : Sea T_n una sucesión de estimadores generados por el funcional T ($T_n(F_n):=T(F_n)$), entonces T es continuo con respecto a la distancia de Prokorov $\forall F\in Z(\mathbb{R})$ ssi $\{T_n; n\geq 1\}$ es cualitativamente robusto $\forall F\in Z(\mathbb{R})$.

5.- ESTIMADORES ROBUSTOS

Sean $X_1, X_2, \cdots, X_n$ variables aleatorias independientes con distribución común F_θ, $\theta\in\Theta\subseteq\mathbb{R}$ y T_n un estimador que puede ser obtenido de un funcional $\theta=T(F)$ definido sobre la familia de distribuciones por evaluación de T en la distribución empírica F_n: $T_n=T(F_n)$. A continuación se presentan 3 clases de estimadores robustos T_n de θ: estimadores tipo M, estimadores tipo L y estimadores tipo R.

5.1.- ESTIMADORES TIPO M

La clase de estimadores tipo M fue introducida por Huber (1964) y sus principales propiedades son expuestas por Huber (1981).

DEF 5.1 : Un M-estimador T_n es la solución del problema de minimización

$$\sum_{i=1}^{n} \rho(x_i,\theta) = \min_{\theta} ! \qquad (5.1)$$

OBS 1 : Si ρ es una función derivable en θ, T_n es la solución de la ecuación lineal de primer orden:

$$\sum_{i=1}^{n} \psi(x_i,\theta) = 0 \qquad (5.2)$$

donde $\psi(x,\theta) = \frac{\partial}{\partial\theta}\rho(x,\theta)$.

Asi T_n verifica $(1/n)\cdot\sum_{i=1}^{n}\psi(x_i,T_n) = \int (x,T_n)\cdot dF_n = 0$

O bien $\int \psi(x,T(F))\cdot dF = 0$ (forma funcional).

OBS 2 : Es bien conocido que el estimador máximo verosimil es definido como el valor $T_n = T_n(x_1,\cdots,x_n)$ que maximiza

$$\prod_{i=1}^{n} f(x_i,T_n)$$

o equivalentemente

$$\sum_{i=1}^{n} -\ln f(x_i,T_n) = \min_{T_n} !$$

Huber propuso una generalización a:

$$\sum_{i=1}^{n} \rho(x_i, T_n) = \min_{T_n} !$$

donde $\rho : \mathcal{X} \times \Theta \dashrightarrow \mathbb{R}$ diferenciable. Por tal motivo se llama a estos estimadores tipo máxima verosimilitud.

CALCULO DE LA FUNCION DE INFLUENCIA

Sea $F_t = (1-t)\cdot F + t\cdot \delta_x$

$$T(F_t) = \int \Psi(x, T(F_t))\cdot dF_t =$$

$$= (1-t)\cdot \int \Psi(x, T(F_t))\cdot dF + t\cdot \Psi(x, T(F_t)) = 0$$

derivando con respecto a t

$$-\int \Psi(x, T(F_t))\cdot dF + (1-t)\cdot \left[\int \frac{\partial}{\partial \theta} \Psi(x, \theta)\Big|_{\theta = T(F_t)} \cdot dF \right] \cdot \frac{dT(F_t)}{dt} +$$

$$+ \Psi(x, T(F_t)) + t\cdot \frac{\partial}{\partial \theta} \Psi(x, \theta)\Big|_{\theta = T(F_t)} \cdot \frac{dT(F_t)}{dt} = 0$$

evaluando en t=0 tenemos

$$\Psi(x, T(F)) - \int \Psi(x, T(F))\cdot dF + IC(X,F,T)\cdot \int \frac{\partial}{\partial \theta} \Psi(x, \theta)\Big|_{\theta = T(F)} \cdot dF = 0$$

$$IC(X,F,T) = \frac{\Psi(x, T(F))}{-\int \frac{\partial}{\partial \theta} \Psi(x, \theta)\Big|_{\theta = T(F)} \cdot dF} \qquad (5.3)$$

OBS 1 : En caso de estimar el parámetro de localización, tenemos $\mathcal{X} = \mathbb{R}$, $\Theta = \mathbb{R}$ y $F(x, \theta) = F(x - \theta)$. Luego, es natural considerar el uso de funciones Ψ del tipo:

$$\Psi(x, \theta) = \Psi(x - \theta)$$

donde

$$IC(X,F,T) = \frac{\Psi(x - \theta)}{\int \Psi'(x - \theta)\cdot dF} = \frac{\Psi}{\int \Psi' \cdot dF}$$

Ejemplos :

1.- Esperanza

$$\theta = E[X] \quad ; \quad \rho(t) = t^2 \quad ; \quad T_n = \bar{X}_n$$

$$IC(X,F,T) = X - E[X]$$

2.- Mediana

$$\theta = F^{-1}(1/2) \quad ; \quad \rho(t) = |t| \quad ; \quad T_n = \min_{\theta} \sum_{i=1}^{n} |x_i - \theta|$$

$$IC(X,F,T) = \frac{\text{sign}[x - F^{-1}(1/2)]}{2 \cdot f(F^{-1}(1/2))}$$

ELECCION DE Ψ.

Si θ es un parámetro de localización y la distribución F supuesta es simétrica, la función debe ser impar y no decreciente.

Huber (1964) sugiere la función siguiente

$$\Psi_C(x) = \text{sign}[x] \cdot \min(|x| , C)$$

donde C es una constante de ajuste.

Por otro lado Tukey (1972) sugiere la función

$$\Psi_B(x) = x \cdot [1 - (x/b)^2]^2 \cdot I_{[-b,b]}(x)$$

donde $I_{[-b,b]}(x)$ es la función caracteristica sobre el conjunto $[-b,b]$ y b es una constante de ajuste.

PROPIEDADES ASINTOTICAS (Serfling (1980))

Si $E_F[\Psi(x,\theta)] = 0$ admite una solución única $\theta = T(F)$ y $\lim_{n\to\infty} T(F_n) = T(F)$. Entonces se puede demostrar que

$$\sqrt{n} \cdot (T(F_n) - T(F)) \text{ -----> } N(0, V(T,F))$$

$$\text{donde } V(T,F) = \int IC^2(X,F,T) \cdot dF = \frac{E_F[\Psi^2(x,\theta)]}{\{\int \frac{\partial}{\partial\theta}\Psi(x,\theta) \cdot dF\}^2}$$

para el caso de parámetro de localización, podemos escribir

$$V(T,F) = \frac{\int \Psi^2 \cdot dF}{\{\int \Psi' \cdot dF\}^2}$$

OBS 1 : Si es una función acotada, entonces se puede demostrar que los M estimadores son B-robustos y cualitativamente robustos y que su punto de ruptura es $\epsilon^* = ½$. Por otro lado si Ψ es una función no acotada se puede demostrar que los estimadores tipo M no son cualitativamente robustos, además tienen un punto de ruptura $\epsilon^* = 0$.

OBS 2 : Desafortunadamente los estimadores tipo M no son invariantes ante transformaciones de escala. Este problema puede ser resuelto definiendo T_n por

$$\sum_{i=1}^{n} \Psi\left(\frac{x_i - T_n}{S_n}\right) = 0$$

donde S_n es un estimador robusto del parámetro de escala, tal como

$$S_n = 1.483 \cdot MAD(x_i) = 1.483 \cdot med_i\{|x_i \cdot med_j(x_j)|\}$$

Bickel (1975) propuso un tipo de estimador M llamado estimador M de un paso (m-estimador) definido por

$$T^{(0)} = med\{x_i\}$$

$$S(0) = med\{|x_i - T^{(0)}|\}$$

$$T^{(m+1)} = T^{(m)} + \frac{\sum_{i=1}^{n} \Psi\left(\frac{x_i - T^{(m)}}{S^{(0)}}\right) \cdot S^{(0)}}{\sum_{i=1}^{n} \Psi'\left(\frac{x_i - T^{(m)}}{S^{(0)}}\right)}$$

Una segunda alternativa conocida como la proposición 2 de Huber estima conjuntamente el parámetro de localización y escala mediante la solución del sistema.

$$\sum_{i=1}^{n} \Psi\left(\frac{x_i - T}{S}\right) = 0$$

$$\sum_{i=1}^{n} \Psi^2\left(\frac{x_i - T}{S}\right) = (n-1) \cdot \beta$$

donde $\beta = E_\Phi[\Psi^2]$.

Un algoritmo para determinar conjuntamente T y S aparece en Huber (1981) y está dado por

$$[S^{(m+1)}]^2 = \frac{1}{(n-1) \cdot \beta} \cdot \sum_{i=1}^{n} \Psi^2\left(\frac{x_i - T^{(m)}}{S^{(m)}}\right) \cdot [S^{(m)}]^2$$

$$T^{(m+1)} = T^{(m)} + \frac{\sum_{i=1}^{n} \Psi\left(\frac{x_i - T^{(m)}}{S^{(m)}}\right) \cdot S^{(m)}}{\sum_{i=1}^{n} \Psi'\left(\frac{x_i - T^{(m)}}{S^{(m)}}\right)}$$

5.2.- ESTIMADORES TIPO L

Recordemos la definición de estadística de posición.
Sea $Ro^n = \{ (z_1,\cdots,z_n) \in \mathbb{R}^n : z_1 \le \cdots \le z_n \}$, entonces la función

$$O_n : \mathbb{R}^n \longrightarrow \mathbb{R}o^n$$

definida por $O_n((y_1,\cdots,y_n)') = (y_{(1)},\cdots,y_{(n)})'$ siendo $(y_{(1)},\cdots,y_{(n)})'$ el vector $(y_1,\cdots,y_n)'$ ordenado de manera que $y_{(1)} \le \cdots \le y_{(n)}$, se llama vector aleatorio de estadísticos de orden.

DEF 5.2 : Sean $a_1,\cdots,a_n \in \mathbb{R}$ $\sum_{i=1}^{n} a_i = 1$. Entonces se llama estimador tipo L inducido por $a_1,\cdots,a_n$ basado en $X_1,\cdots,X_n$ a

$$L_n = L_n(X_1,\cdots,X_n) = \sum_{i=1}^{n} a_i \cdot h(X_{(i)}) \qquad (5.4)$$

OBS 1 : En general los coeficientes a_i dependen del tamaño de la muestra y pueden ser engendrados a través de una medida M en $[0,1]$, es decir,

$$a_{ni} = \int_{\frac{(i-1)}{n}}^{i/n} dM(s)$$

En tal caso $T_n = T(F_n) = \int h[F_n^{-1}(s)]\cdot dM_s$,

donde $F_n = (1/n)\cdot \sum_{i=1}^{n} \delta x_i$,

y por lo tanto su función inversa F está definida por

$$F_n^{-1}(s) = \begin{cases} x(i) & \text{si } \frac{(i-1)}{n} < s \le \frac{i}{n}, \quad i=1,2,\cdots,n-1 \\ x(n) & \text{si } 1-\frac{1}{n} < s < 1 \end{cases}$$

Luego la forma funcional del estimador es

$$T(F) = \int h[F^{-1}(s)]\cdot dM_s \qquad (5.5)$$

CALCULO DE LA FUNCION DE INFLUENCIA

Sea $\cdot\ F_t = (1-t)\,F + t\delta_x$;

$$\frac{dT(F_t)}{dt} = \int_0^1 \frac{dF_t^{-1}(s)}{dt}\, h'[F_t^{-1}(s)]\, dM(s)\,.$$

Si $u < x$: $s = F_t(u) = (1-t)F(u)$ donc $u = F_t^{-1}(s) = F^{-1}(\frac{s}{1-t})$

$$\Longrightarrow \qquad \frac{dF_t^{-1}(s)}{dt} = \frac{s}{(1-t)^2} \cdot \frac{1}{f(F^{-1}(\frac{s}{1-t}))} \quad ;$$

si $u \geq x$: $s = F_t(u) = (1-t)F(u) + t$ donc $u = F_t^{-1}(s) = F^{-1}(\frac{s-t}{1-t})$

$$\Longrightarrow \qquad \frac{dF_t^{-1}(s)}{dt} = \frac{s-1}{(1-t)^2} \cdot \frac{1}{f\left[F^{-1}(\frac{s-t}{1-t})\right]} \ ,$$

Luego

$$\frac{dT(F_t)}{dt} = \int_0^{F_t(x)} \frac{s}{(1-t)^2} \cdot \frac{h'\left[F^{-1}(\frac{s}{1-t})\right]}{f\left[F^{-1}(\frac{s}{1-t})\right]} \, dM(s) + \int_{F_t(x)}^1 \frac{s-1}{(1-t)^2} \cdot \frac{h'\left[F^{-1}(\frac{s-t}{1-t})\right]}{f\left[F^{-1}(\frac{s-t}{1-t})\right]} \, dM(s)$$

y

$$IC(x \ ; F,T) = \int_0^{F(x)} s \, \frac{h'\left[F^{-1}(s)\right]}{f\left[F^{-1}(s)\right]} \, dM(s) + \int_{F(x)}^1 (s-1) \, \frac{h'\left[F^{-1}(s)\right]}{f\left[F^{-1}(s)\right]} \, dM(s) \ ;$$

ó bien

$$IC(x \ ; F,T) = \int_0^1 \frac{s h'\left[F^{-1}(s)\right]}{f\left[F^{-1}(s)\right]} \, dM(s) - \int_{F(x)}^1 \frac{h'\left[F^{-1}(s)\right]}{f\left[F^{-1}(s)\right]} \, dM(s)$$

Si M admite una densidad m

$$IC(x \ ; F,T) = \int_{-\infty}^{+\infty} F(y) h'(y) m\left[F(y)\right] dy - \int_x^{+\infty} h'(y) \, m\left[F(y)\right] dy$$

$$IC(x \ ; F,T) = \int_{-\infty}^{+\infty} \left[F(y) - 1_{[x,+\infty[}(y)\right] h'(y) \, m\left[F(y)\right] dy \ .$$

$$\Longrightarrow \qquad \frac{d}{dx} IC(x \ ; F,T) = h'(x) . m\left[F(x)\right] .$$

En particular si $h(x) = x$ $\wedge$ F es simetrica.

$$IC(x \ ; F,T) = \int_0^x m\left[F(y)\right] dy .$$

Ejemplos :

a) Media Truncada :

Sea $\alpha \in\ ]0,\frac{1}{2}]$, entonces la media α-truncada se obtiene suprimiendo las $[n\alpha]$ primeras y últimas observaciones de la muestra ordenada

$$\bar{X}_{n,\alpha} = \frac{1}{n-2[n\alpha]} \cdot \sum_{i=[n\alpha]+1}^{n-[n\alpha]} X(i)$$

S.p.d.g. supongamos para simplificar que $\alpha\cdot n \in N$

$$T_n = \bar{X}_{n,\alpha} = \sum_{i=1}^{n} a_{ni}\cdot h(x_{(i)}) \ , \quad \text{donde } h(x) = x \ , \quad y$$

$$a_{ni} = \begin{cases} \dfrac{1}{n(1-2\alpha)} & \text{si } 1+\alpha\cdot n \le i \le n\cdot(1-\alpha) \\ 0 & \text{e.t.o.c.} \end{cases}$$

$$\text{asi} \quad a_{ni} = \frac{1}{1-2\alpha} \cdot \int_{\frac{i-1}{n}}^{i/n} I_{[\alpha,1-\alpha]}(t) \cdot dt = \int_{\frac{i-1}{n}}^{i/n} dM(s)$$

Luego M admite densidad $m(t) = \dfrac{1}{1-2\alpha} \cdot I_{[\alpha,1-\alpha]}(t)$

$$T_n = T(F_n) = \frac{1}{1-2\alpha} \cdot \int_{\alpha}^{1-\alpha} F_n^{-1}(t)\cdot dt$$

siendo su forma funcional

$$T = T(F) = \frac{1}{1-2\alpha} \cdot \int_{\alpha}^{1-\alpha} F^{-1}(s)\cdot ds$$

FUNCION DE INFLUENCIA

$$IC(x\ ;\ F,T) = \int_{-\infty}^{+\infty} F(y)\ m[F(y)]\ dy - \int_{x}^{+\infty} m[F(y)]\ dy = \frac{1}{1-2\alpha}\left(\int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} F(y)dy - \int_{x}^{+\infty} 1_{[\alpha,1-\alpha]}(F(y))dy\right)$$

$$(1-2\alpha)\ IC = (1-\alpha)\ F^{-1}(1-\alpha) - \alpha\ F^{-1}(\alpha) - \int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} y dF(y) - \int_{x}^{+\infty} 1_{[\alpha,1-\alpha]}(F(y))dy$$

$$= (1-\alpha)\ F^{-1}(1-\alpha) - \alpha\ F^{-1}(\alpha) - \int_{\alpha}^{1-\alpha} F^{-1}(s)\ ds - \int_{x}^{+\infty} 1_{[\alpha,1-\alpha]}(F(y))dy\ .$$

de donde

si $x < F^{-1}(\alpha)$: $(1-2\alpha)\, IC = (1-\alpha)\, F^{-1}(1-\alpha) - \alpha\, F^{-1}(\alpha) - \int_\alpha^{1-\alpha} F^{-1}(s)ds - \left[F^{-1}(1-\alpha) - F^{-1}(\alpha)\right]$

si $F^{-1}(\alpha) \leq x \leq F^{-1}(1-\alpha)$: $(1-2\alpha) IC = (1-\alpha)F^{-1}(1-\alpha) - \alpha\, F^{-1}(\alpha) - \int_\alpha^{1-\alpha} F^{-1}(s)ds - \left[F^{-1}(1-\alpha) - x\right]$

si $F^{-1}(1-\alpha) < x$: $(1-2\alpha)\, IC = (1-\alpha)\, F^{-1}(1-\alpha) - \alpha\, F^{-1}(\alpha) - \int_\alpha^{1-\alpha} F^{-1}(s)\, ds.$

$$IC(x\,;\,F,\,T) = \begin{cases} -\dfrac{1}{1-2\alpha}\left[\alpha\, F^{-1}(1-\alpha) - (1-\alpha)\, F^{-1}(\alpha) + \displaystyle\int_\alpha^{1-\alpha} F^{-1}(s)\, ds\right] & \text{si } x < F^{-1}(\alpha) \\ \dfrac{1}{1-2\alpha}\left[x - \alpha\, F^{-1}(\alpha) - \alpha\, F^{-1}(1-\alpha) - \displaystyle\int_\alpha^{1-\alpha} F^{-1}(s)ds\right] & F^{-1}(\alpha) \leq x \leq F^{-1}(1-\alpha) \\ -\dfrac{1}{1-2\alpha}\left[\alpha\, F^{-1}(\alpha) - (1-\alpha)\, F^{-1}(1-\alpha) + \displaystyle\int_\alpha^{1-\alpha} F^{-1}(s)ds\right] & \text{si } x > F^{-1}(1-\alpha)\,. \end{cases}$$

Pero IC(X,F,T) es una función en X monótona, luego si F es simétrica T(F) = 0, se verifica que

$$(1-2\alpha)\cdot IC(X,F,T) = \begin{cases} F^{-1}(\alpha) & \text{si } x < F^{-1}(\alpha) \\ X & \text{si } F^{-1}(\alpha) \leq x \leq F^{-1}(1-\alpha) \\ F^{-1}(1-\alpha) & \text{si } x > F^{-1}(1-\alpha) \end{cases}$$

<u>OBS 1 :</u> $\alpha \to ½$ se tiene $T(F) = F^{-1}(½)$

$$IC(X,F,T) = \frac{\text{sign}[x - F^{-1}(½)]}{2\cdot f[F^{-1}(½)]}$$

<u>OBS 2 :</u> $\alpha \to 0$ se tiene $T(F) = E[X]$

$$IC(X,F,T) = X - E[X]$$

<u>b) Media Winsorizada</u>

Sea $\alpha \in\,]0,½]$, la media winsorizada de orden α es la media aritmética simple obtenida reemplazando las primeras $[\alpha n]$ observaciones por $X_{([\alpha n]+1)}$ y las últimas $[\alpha n]$ observaciones por $X_{(n-[n\alpha])}$:

$$\overline{W}_{n,\alpha} = \frac{1}{n}\cdot\left[\sum_{i=[n\alpha]+1}^{n-[n\alpha]} X_{(i)} + [\alpha n]\cdot\left(X_{([\alpha n]+1)} + X_{(n-[\alpha n])}\right)\right]$$

S.p.d.g. supongamos $\alpha n \in \mathbb{N}$:

$$T_n = \bar{W}_{n,\alpha} = \sum_{i=1}^{n} a_{ni} \cdot h(X_{(i)}) \quad , \quad \text{donde}$$

$$a_{ni} = \begin{cases} \frac{1}{n} & \text{si} \quad 1+\alpha \cdot n \leq i \leq n \cdot (1-\alpha) \\ 0 & \text{e.t.o.c.} \end{cases} \quad , \quad y$$

$$h(X_{(i)}) = \begin{cases} X_{(i)} & \text{si} \quad 1+\alpha n \leq i \leq n(1-\alpha) \\ X_{(1+\alpha n)} & \text{si} \quad i \leq n\alpha \\ X_{(n-\alpha n)} & \text{si} \quad i \geq (n-\alpha n) \end{cases}$$

es decir

$$T_n = T(F_n) = \alpha \cdot F^{-1}(\alpha) + \int_{\alpha}^{1-\alpha} F_n^{-1}(t) \cdot dt + \alpha \cdot F_n^{-1}(1-\alpha)$$

$$T(F) = \int_{\alpha}^{1-\alpha} F^{-1}(s) \cdot ds + \alpha \cdot F^{-1}(\alpha) + \alpha \cdot F^{-1}(1-\alpha).$$

CALCULO DE LA FUNCION DE INFLUENCIA

$$m(t) = 1_{[\alpha, 1-\alpha]}(t) + \alpha\, \delta_{\alpha} + \alpha\, \delta_{1-\alpha} \qquad h'(x) = 1_{[F^{-1}(\alpha),\, F^{-1}(1-\alpha)]}(x)$$

$$\int_0^1 \frac{s h'[F^{-1}(s)]}{f[F^{-1}(s)]}\, dM(s) = \alpha \frac{\alpha}{f[F^{-1}(\alpha)]} + \int_{\alpha}^{1-\alpha} \frac{s m(s)}{f[F^{-1}(s)]}\, ds + \alpha \frac{1-\alpha}{f[F^{-1}(1-\alpha)]}$$

$$= \frac{\alpha^2}{f[F^{-1}(\alpha)]} + (1-\alpha)\, F^{-1}(1-\alpha) - \alpha F^{-1}(\alpha) - \int_{\alpha}^{1-\alpha} F^{-1}(s)\, ds + \frac{\alpha-\alpha^2}{f[F^{-1}(1-\alpha)]}$$

$$\text{Si } x<F^{-1}(\alpha) : \int_{F(x)}^{1} \frac{h'[F^{-1}(s)]}{f[F^{-1}(s)]}\, dM(s) = \frac{\alpha}{f[F^{-1}(\alpha)]} + \int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} ds + \frac{\alpha}{f[F^{-1}(1-\alpha)]}$$

$$x \; ; F,T) = F^{-1}(\alpha) - \alpha F^{-1}(\alpha) - \bar{\alpha}F^{-1}(1-\alpha) - \int_{\alpha}^{1-\alpha} F^{-1}(s)\,ds - \frac{\alpha}{f[F^{-1}(\alpha)]} + \frac{\alpha^2}{f[F^{-1}(\alpha)]} - \frac{\alpha^2}{f[F^{-1}(1-\alpha)]}$$

$$\text{Si } F^{-1}(\alpha) \leq x \leq F^{-1}(1-\alpha) : \int_{F(x)}^{1} \frac{h'[F^{-1}(s)]}{f[F^{-1}(s)]}\,dH(s) = \int_{x}^{F^{-1}(1-\alpha)} ds + \frac{\alpha}{f[F^{-1}(1-\alpha)]}$$

$$IC(x \; ; F,T) = x - \alpha F^{-1}(\alpha) - \alpha F^{-1}(1-\alpha) - \int_{\alpha}^{1-\alpha} F^{-1}(s)\,ds + \frac{\alpha^2}{f[F^{-1}(\alpha)]} - \frac{\alpha^2}{f[F^{-1}(1-\alpha)]}$$

$$\text{Si } \quad F^{-1}(1-\alpha) < x : \int_{F(x)}^{1} \frac{h'[F^{-1}(s)]}{f[F^{-1}(s)]}\,dH(s) = 0 \,,$$

$$IC(x \; ; F,T) = (1-\alpha) F^{-1}(1-\alpha) - \alpha F^{-1}(\alpha) - \int_{\alpha}^{1-\alpha} F^{-1}(s)\,ds + \frac{\alpha^2}{f[F^{-1}(\alpha)]} + \frac{\alpha^2}{f[F^{-1}(1-\alpha)]}$$

Si F es simétrica, entonces $T(F) = 0$, luego se verifica

$$IC(X,F,T) = \begin{cases} F^{-1}(\alpha) - \dfrac{\alpha}{f[F^{-1}(\alpha)]} & \text{si } x < F^{-1}(\alpha) \\ X & \text{si } F^{-1}(\alpha) \leq x \leq F^{-1}(1-\alpha) \\ F^{-1}(1-\alpha) + \dfrac{\alpha}{f[F^{-1}(\alpha)]} & \text{si } x > F^{-1}(1-\alpha) \end{cases}$$

5.3.- ESTIMADORES TIPO R

Los estimadores tienen su origen en el problema de la comparación de 2 muestras en estadística no paramétrica es decir se construyen a partir de una combinación lineal de las estadísticas de rango

$$S_N = \sum_{i=1}^{N} c_i \cdot a_N(R_i)$$

donde N es el número total de observaciones $(z_1, \cdots, z_N)$, R_i es el rango de z_i en la muestra ordenada

$$R_i = \sum_{j=1}^{N} I_{\mathbb{R}^+}(z_i - z_j)$$

Los $a_N(i)$, $1 \leq i \leq N$ son los "scores" construidos en general a partir de una función $J:(0,1) \dashrightarrow \mathbb{R}$ no decreciente: $J(1-t) = -J(t)$.

Las elecciones usuales de "scores" $a_N(i)$ son:

$$J((i-½)/N) \;;\quad N\cdot\int_{(i-1)/n}^{i/n} J(t)\cdot dt \;;\quad J(i/(N+1)) \;;\quad (N+1)\cdot\int_{i/(N+1)}^{(i+1)/(N+1)} J(t)\cdot dt \;;\quad \text{etc.}$$

TEST DE RANGOS

Sean $x_1,\cdots,x_m$ e $y_1,\cdots,y_n$ dos muestras de variables aleatorias independientes provenientes de las distribuciones F y G respectivamente. Supongamos que deseamos contrastar la hipótesis nula H_0: $F=G$ v/s la alternativa H_1: $\forall x\in R,\ G(x) = F(x - \theta),\ \theta>0$. Para esta prueba en métodos no paramétricos se usa una estadística lineal de rangos del tipo:

$$S_{m,n} = (1/m)\sum_{i=1}^{m} a_N(R_i) \quad \text{es decir} \quad c_i = \begin{cases} 1/m & 1 \le i \le m \\ 0 & m+1 \le i \le N \end{cases}$$

donde $N=n+m$, y R_i es el rango de la muestra global $(x_1,\cdots,x_m,y_1,\cdots,y_n)$ de tamaño N.

Para simplificar la estadística supongamos $m=n$. Entonces:

$$S_n = (1/n)\cdot\sum_{i=1}^{n} a_{2n}(R_i) \;,\ \text{donde } a_{2n} = J((i-½)/2n),\ \text{luego}$$

$$R_i = \sum_{j=1}^{n} I_{]-\infty,x_i]}(x_j) + \sum_{j=1}^{n} I_{]-\infty,x_i]}(y_j) = n\cdot F_n(x_i) + G_n(x_i)$$

Sea:

$$H = ½\cdot(F+G) \text{ y en los puntos de discontinuidad sea}$$

$$H(x) = ½\cdot(H(x-0) + H(x+0)),\ x\in R$$

$$n\cdot[F_n(x_i-0) + F_n(x_i+0)] = 2n\cdot F_n(x_i) - 1 \;,$$

$$n\cdot[G_n(x_i-0) + G_n(x_i+0)] = 2n\cdot G_n(x_i)$$

Obteniendo:

$$R_i = n\cdot H_n(x_i) + ½ \;;\quad ½\cdot H_n(x_i) = (R_i-½)/2n$$

$$S_n = \int J[½\cdot F_n(t) + ½\cdot G_n(t)]\cdot dF_n(t) = S(F_n,G_n)$$

luego, la forma funcional es:

$$S(F,G) = \int J[½\cdot F(t) + ½\cdot G(t)]\cdot dF(t)$$

Ahora veamos en el contexto de la estimación de parámetros de localización esta estadística lineal de rangos.

DEF 5.3 : Un estimador-R de localización puede ser definido como $T_n = T_n(x_1,\cdots,x_n)$, donde T_n es elegido de modo que,

$S_n = (1/N)\cdot \sum_{i=1}^{n} a_N(R_i)$, sea lo más próxima a cero posible

cuando calculamos la estadística de rangos de la muestra $x_1,\cdots,x_n$ $2T_n-x_1,\cdots,2T_n-x_n$.

OBS 1 : La idea de esta definición es la siguiente: a partir de la muestra original x1,,xn podemos construir una muestra imagen reemplazando cad xi por Tn-(xi-Tn). Tn se elige de manera Sn no detecte cambio del parámetro de localización, es decir Sn debe ser lo más cercano a cero posible. La forma de Tn es

$$\int J[\tfrac{1}{2}\cdot F_n(t) + \tfrac{1}{2}\cdot(1 - F_n(2T_n-t))]\cdot dF_n = 0$$

o bien

$$\int J[\tfrac{1}{2}\cdot F(t) + \tfrac{1}{2}\cdot(1 - F(2T-t))]\cdot dF = 0$$

La función de influencia para T, cuando F es simétrico, está dado por

$$IC(X,F,T) = \frac{J(F(x))}{\int [J(F(t))]'\cdot dF}$$

(Ver Hampel (1986)).

BIBLIOGRAFIA

- Allende, H. (1988)
  "Robuste Schätzung von ARMA-Prozessen"
  Ph.D. Thesis
  Depto of Statistics, University of Dortmund, West Germany

- Allende, H.; Heiler, S. (1987)
  "A Robust Approach to Parameter Estimation in Time Series Models"
  H.Dinges and S.Kounias Ed.Proc.
  17th European Meeting of Statisticians
  Vol. 1, pp 45-49

- Box, G.E.P. (1953)
  "Non-Normality and Test on Variances"
  Biometrika 40, S 318-353

- Bustos, H.O.; Maronna, R.A.; Yohai, V.J. (1979)
  "Bias and Efficiency-robustness of general m-estimators for regression with random carries"
  Smoothing Techniques for Curves Estimation.
  T.H. Gasser, M. Rossenblatt,
  Eds. Proc. Heidelberg-New York: Springer.

- Clark, D.I.; Osborne, M.R. (1986)
  "Finite Algorithms for Huber's m-estimator"
  SIAM J. SCI. Statis. Comput. Vol. 7, No. 1, 72-85

- Hampel, F.R. (1971)
  "A general qualitative definition of robustness"
  Ann. Math. Statist. 42, 1887-1896

- Hampel, F.R. (1974)
  "A general qualitative definition of robustness"
  Ann. Math. Statist. 42, 1887-1896

- Hampel, F.R. (1971)
  "The influence curve and its role in robust estimation"
  JASA, Vol. 69, 383-393.

- Hampel, F.R.; Ronchetti, E.; Rousseeuw, P.J.; Stahel, W. (1986)
  "Robust Statistics - The approach based on influence curve"
  New York, John Wiley.

- Huber, P.J. (1964)
  "Robust Estimation of a location parameter"
  Ann. Math. Statist. 35, 73-101

- Huber, P.J. (1973)
  "Robust regression: Asymptotic conjetures and Monte Carlo"
  Ann. Math. Statist. 1, 799-821

- Huber, P.J. (1981)
  "Robust Statistics"
  New York, John Wiley.

- Huber, P.J. (1983)
  "Maximax aspects of bounded-influence regression"
  JASA Vol. 77, No. 379, 595-604

- Maronna, R.A.; Yohai, V.J. (1981)
  "Asymptotic behavior of general M-estimates for regression and scale with ranfom carriers"
  Z. Wahrscheinlichkeitstheory und verw. Gebiete, 58, 7-20

- Portnoy, S.L. (1977)
  "Robust estimation in dependent situations"
  Ann. Statist. 5, 22-43

- Serfling, R.J. (1980)
  "Approximation theorems of Mathematical Statistics"
  New York, John Wiley

- Tukey, J.W. (1970)
  "Explorative Data Analysis"
  Addison Wesley

- Wolke, R.(1985)
  "Iterative Verfahren zur numerische Berechnung von M-Schätzungen"
  Dissertation, Universität Halle.