* Estimação robusta

Situações que justificam estimadores robustos:

- Presença de medições incompatíveis (outliers)
- Dados que satisfazem mais de um modelo.

Os métodos que são capazes de lidar com estas situações são estudados em "estatística robusta". Alguns deles serão estudados aqui.

a) M-estimadores (M vem de máxima verossimilhança)

Função de custo:

$$V(z,\theta) = \sum_{k=1}^{N} \rho(r_k)$$

r_k: resíduos. ρ é uma função simétrica, positiva definida, cujo $\rho(0)=0$ é o único valor para o qual ρ é nula. r_k é o k-ésimo resíduo:

$$r_k = y_k - f(x_k, \theta).$$

Procura-se portanto

$$\hat{\theta} = \arg\min_{\theta} V(z,\theta).$$

Mas, $\hat{\theta}$ satisfaz a:

$$\left.\frac{\partial V(z,\theta)}{\partial \theta}\right|_{\theta=\hat{\theta}} = \sum_{k=1}^{N} \psi(r_k) \cdot \left.\frac{\partial r_k}{\partial \theta}\right|_{\theta=\hat{\theta}} = 0$$

com $\psi(r_k) = \dfrac{\partial \rho(r_k)}{\partial r_k}$ sendo a chamada função de influência. (na verdade, ψ é proporcional à função de influência: ... qualitativa)

A função ψ indica como as medidas influenciam no processo de estimação. Por exemplo, com $\rho(r_k)=\frac{r_k^2}{2}$, temos o critério soma dos quadrados. Sendo

$$\psi(r_k) = r_k \Rightarrow \text{Ponto de ruptura } \underline{0}.$$

que mostra que cada medição tem influência na estimativa. Já, se for usada a função de Huber:

$$\rho(r_k) = \begin{cases} r_k^2/2 & , \text{ se } |r_k| \leq c.s. \\ c.s.(|r_k| - c.s/2) & , \text{ caso contrário} \end{cases}$$

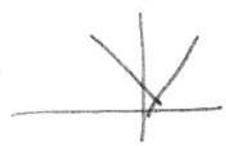
com $c = 1{,}345$ se o "bom erro" for Gaussiano, e s é uma estimativa de escala. Geralmente usa-se

$$s = MAD(r) = \underset{i}{med}\{|r_i - \underset{j}{med}(r_j)|\}$$

Mostrar como é robusto se comparado à média (mediana) e desvio padrão (MAD)

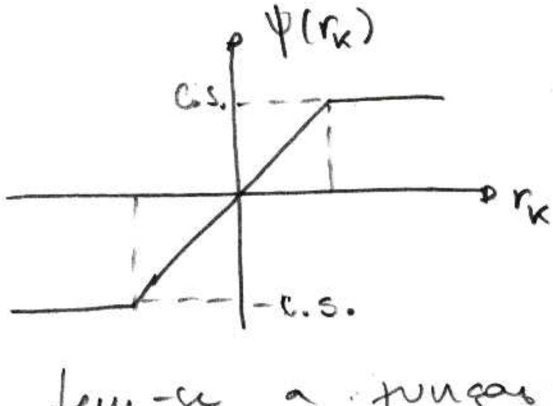
Tem-se então que

$$\psi(r_k) = \begin{cases} r_k & , \text{ se } |r_k| \leq c.s. \\ c.s.\, \underbrace{sign(r_k)}_{\frac{d|r_k|}{dr_k}} \end{cases}$$

$|r_k|$: 

$\frac{d|r_k|}{dr_k} =$ 

Portanto

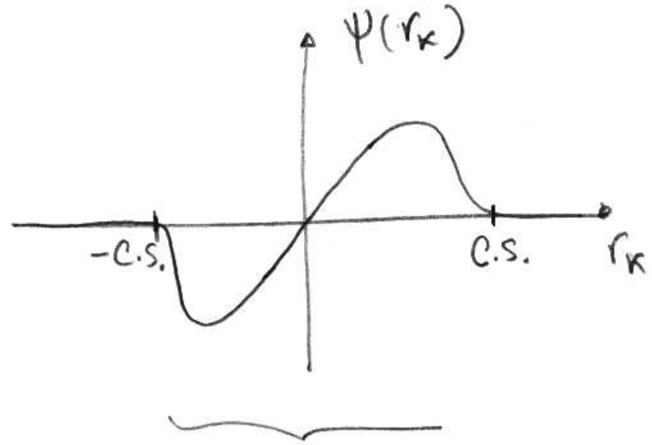


Ou ainda, tem-se a função Tukey biweight:

$$\rho(r_k) = \begin{cases} \frac{(c.s)^2}{6}\left(1 - [1 - (r_k/c.s.)^2]^3\right) & , |r_k| \leq c.s. \\ \frac{(c.s)^2}{6} & , \text{ caso contrário} \end{cases}$$

que leva a

$$\psi(r_k) = \begin{cases} r_k \cdot [1 - (r_k/_{c.s.})^2]^2 , & |r_k| \leq c.s. \\ 0 , & c.c. \end{cases}$$



o que mostra que $|r_k| \geq c.s$ não tem influência na estimação.

A resolução do problema pode ser através de um método iterativo pois estamos lidando com um problema não linear, ou por mínimos quadrados ponderados iterativos. O princípio é simples:

Seja

$$w(r_k) = \frac{\psi(r_k)}{r_k} \quad \text{ë peso associado a } r_k$$

então

$$\sum_{k=1}^{N} \psi(r_k) \cdot \frac{\partial r_k}{\partial \theta}\Big|_{\theta=\hat{\theta}} = \sum_{k=1}^{N} w(r_k) \cdot r_k \cdot \frac{\partial r_k}{\partial \theta}\Big|_{\theta=\hat{\theta}} = 0$$

que é o mesmo problema de

$$\hat{\theta} = \arg\min_{\theta} \sum_{k=1}^{N} w(r_k) \cdot \frac{r_k^2}{2}$$

que é o de mínimos quadrados ponderados.

Assim, o algoritmo se escreve como

(i) inicialize todos os pesos $w(r_k)$. Alternativos:

a) $w(r_k) = 1$ (sem info $\hat{\theta}_0$ a priori)

b) $w(r_k) = \dfrac{\psi(r_k)}{r_k}$ (com $\hat{\theta}_0$ a priori)

(ii) resolva o problema de mínimos quadrados ponderados:

$$\hat{\theta} = \arg\min_{\theta} \sum_{k=1}^{N} w(r_k) \cdot \frac{r_k^2}{2} \quad \Big\} \text{ solução fechada p/ } f(x_k, \theta) = \varphi_k^T \cdot \theta.$$

(iii) recalcule $w(r_k)$ considerando a nova estimativa

(iv) verifique se condição de parada é satisfeita. Caso não, retorne p/ o passo (ii).

Como exemplo, $w(r_k)$ pode ser:

- P/ $\rho(r_k) = \dfrac{r_k^2}{2}$ (mínimos quadrados),

$$w(r_k) = 1.$$

- P/ $\rho(r_k)$ sendo a função de Huber, então

$$w(r_k) = \begin{cases} 1, & |r_k| \leq c.s. \\ \dfrac{c.s.}{|r_k|}, & c.c. \end{cases}$$

$$= \min\left\{1, \frac{c.s.}{|r_k|}\right\}$$

- P/ $\rho(r_k)$ sendo Tukey's biweight

$$w(r_k) = \begin{cases} \left[1 - (r_k/c.s.)^2\right]^2, & |r_k| \leq c.s. \\ 0 \end{cases}$$

b) Mediana mínima dos quadrados (Least Median of squares, ou LMedS)*

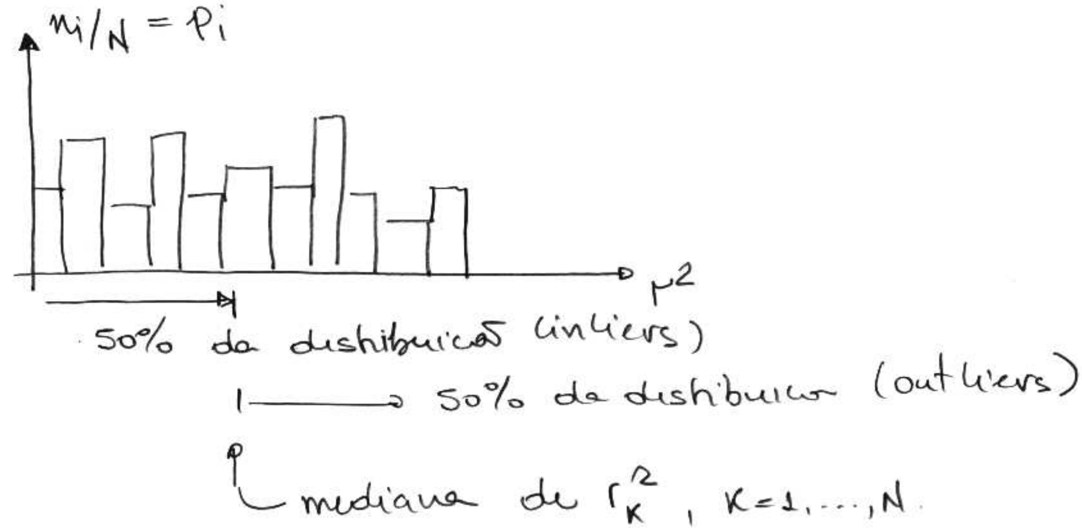
No LMedS, estamos interessados em resolver o seguinte problema

$$\hat{\theta} = \arg\min_{\theta} \operatorname{med}_{k} r_k^2 \quad , \quad k = 1, \ldots, N.$$

com $r_k = y_k - f(x_k, \theta)$. No caso, foi mostrado que o ponto de ruptura é dado por

$$\frac{\frac{N}{2} - d + 2}{N}$$

que, com $N \to \infty$ leva a 50% independente do número d de parâmetros. Mas, se considerarmos o problema of regressão linear p/ $d=2$ (reta), tem-se que o ponto de ruptura é 50%! Esse argumento é fácil de ser aceito, bastando considerar o histograma dos resíduos:



* Alguns autores denominam LMS, mas pode ser confundido com

O problema agora consiste em encontrar $\hat{\theta}$, ou seja, resolver o problema de otimização proposto pelo LMedS. Deve-se observar que o critério LMedS não é diferenciável. Portanto, não podemos usar os métodos Newtonianos. O seguinte algoritmo é então proposto:

Seja $Z = (x_1, y_1) \dots (x_N, y_N)$ N conjuntos de medições.

1. Para $s = 1$ até S, faça:
   - Amostre um subconjunto de d amostras do conjunto Z sem reposição Z_s
   - Resolva o problema para Z_s usando, por exemplo, mínimos quadrados:
     $$\hat{\theta}_s = \arg\min_{\theta} \sum_{(x_i, y_i) \in Z_s} (y_i - f(x_i, \theta))^2$$
   - Calcule $m_s = \operatorname{med} r_K^2$ com $r_K = y_K - f(x_K, \hat{\theta}_s)$, $K = 1, \dots, N$
2. Escolha como estimativa $\hat{\theta}_{LMedS}$ aquela associada à menor mediana dentre as S amostras anteriores.

O algoritmo acima faz uso de amostragem para determinar a estimativa $\hat{\theta}_{LMedS}$. Um ponto a ser resolvido consiste em determinar o número de amostras S, que deve ser grande o suficiente de forma que seja grande a chance de se amostrar pelo menos um subconjunto com "inliers". Sendo p a proporção de inliers no processo, a probabilidade de se ter pelo menos um inlier no conjunto de S amostras é

$$P_i = 1 - (1 - p^d)^S$$

do qual observa-se que

p^d : Probabilidade de todos os pontos de um subconjunto de d amostras serem "inliers" (Z_s)

$1-p^d$: Probabilidade de se ter gerado pelo menos um "outlier" em Z_s.

$(1-p^d)^s$: Probabilidade de todos os subconjuntos Z_s terem pelo menos um "outlier"

P_i : Probabilidade de pelo menos um subconjunto Z_s não possuir "outliers".

O número S pode, então ser obtido como

$$S = \frac{\ln(1-P_i)}{\ln(1-p^d)}$$

Para se ter uma idéia, seja $d=2$ (reta), $p=0,5$ (o mais robusto possível) e $P_i = 0,999$, teríamos $\underline{S \cong 25}$, que é um número razoável. Esse critério considera que a amostragem é realmente aleatória.

Por outro lado $\hat{\theta}_{LMeds}$ é ineficiente na presença de ruído Gaussiano (ver literatura), de forma que ele é mais utilizado para determinar os "inliers". Para tanto, sendo m_{LMeds} a mediana correspondente a $\hat{\theta}_{LMeds}$, usa-se a seguinte estimativa de desvio-padrão para os "inliers"

$$\hat{\sigma}_{LMeds} = 1,4826 \cdot \left\{1 + \frac{5}{N-d}\right\} \cdot \sqrt{m_{LMeds}}$$

e então, obtém-se como estimativa a solução

$$\hat{\theta} = \arg\min_{\theta} \sum_{k=1}^{N} w_k \cdot \left(y_k - f(x_k, \theta)\right)^2$$

com

$$w_k = \begin{cases} 1, & \text{se } |y_k - f(x_k, \hat{\theta}_{LMeds})| \leq 2{,}5 \cdot \hat{\sigma}_{LMeds} \\ 0, & \text{caso contrário} \end{cases}$$