MAT-266: Métodos resistentes a outliers en regresión, distribuciones con colas pesadas

Felipe Osorio

fosorios.mat.utfsm.cl

Departamento de Matemática, UTFSM

Observación:

Debemos resaltar que diferentemente al caso de la distribución normal, en el caso general de la familia elíptica, podemos tener los siguientes enfoques:

(a) Modelo dependiente: Supondremos $Y_1, \dots, Y_n$ tal que su densidad conjunta $\boldsymbol{Y} = (Y_1, \dots, Y_n)^\top \sim \mathsf{EC}_n(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2 \boldsymbol{I}_n; g)$, sigue una distribución de contornos elipticos.[1]

(b) Modelo independiente: Considere $Y_1, \dots, Y_n$ variables aleatorias independientes cada una con distribución $\mathsf{EC}_1(\boldsymbol{x}_i^\top \boldsymbol{\beta}, \sigma^2; g)$.

[1] El estimador de $\boldsymbol{\beta}$ en el modelo dependiente es equivalente al LSE y por tanto NO es robusto.

En la siguientes diapositivas asumiremos el modelo independiente,

$$Y_i \overset{\text{ind}}{\sim} \mathsf{EC}_1(\boldsymbol{x}_i^\top \boldsymbol{\beta}, \sigma^2; g), \qquad i = 1, \dots, n,$$

donde

$$f(y_i; \boldsymbol{\theta}) = \frac{1}{\sigma} g\big((Y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta})^2 / \sigma^2\big),$$

con $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2)^\top$, y la función generadora de densidades g debe satisfacer:

$$\int_0^\infty u^{-1/2} g(u) \, \mathsf{d}\, u < +\infty.$$

Estimación ML bajo la familia de contornos elípticos

Para obtener los estimadores máximo verosímiles de $\boldsymbol{\beta}$ y σ^2 asumiremos que $g(\cdot)$ es conocido. La función de log-verosimilitud adopta la forma:

$$\ell(\boldsymbol{\theta}) = -\frac{n}{2}\log\sigma^2 + \sum_{i=1}^{n}\log g(u_i),$$

donde $u_i = (Y_i - \boldsymbol{x}_i^\top\boldsymbol{\beta})^2/\sigma^2$, para $i = 1,\dots,n$. Diferenciando $\ell(\boldsymbol{\theta})$ con relación a $\boldsymbol{\beta}$, obtenemos

$$\begin{aligned}
\mathsf{d}_\beta\,\ell(\boldsymbol{\theta}) &= \sum_{i=1}^{n}\mathsf{d}_\beta\log g(u_i) = \sum_{i=1}^{n}\frac{g'(u_i)}{g(u_i)}\,\mathsf{d}_\beta u_i \\
&= -\frac{2}{\sigma^2}\sum_{i=1}^{n}\frac{g'(u_i)}{g(u_i)}(Y_i - \boldsymbol{x}_i^\top\boldsymbol{\beta})\boldsymbol{x}_i^\top\,\mathsf{d}\,\boldsymbol{\beta} \\
&= \frac{1}{\sigma^2}\sum_{i=1}^{n}W_i(\boldsymbol{\theta})(Y_i - \boldsymbol{x}_i^\top\boldsymbol{\beta})\boldsymbol{x}_i^\top\,\mathsf{d}\,\boldsymbol{\beta}
\end{aligned}$$

donde $W_i(\boldsymbol{\theta}) = -2g'(u_i)/g(u_i)$, para $i = 1,\dots,n$.

Análogamente, diferenciando $\ell(\boldsymbol{\theta})$ con relación a σ^2, tenemos

$$\begin{aligned}
\mathsf{d}_{\sigma^2}\,\ell(\boldsymbol{\theta}) &= -\frac{n}{2\sigma^2}\,\mathsf{d}\,\sigma^2 + \sum_{i=1}^n \frac{g'(u_i)}{g(u_i)}\,\mathsf{d}_{\sigma^2}\,u_i \\
&= -\frac{n}{2\sigma^2}\,\mathsf{d}\,\sigma^2 - \frac{1}{\sigma^4}\sum_{i=1}^n \frac{g'(u_i)}{g(u_i)}(Y_i - \boldsymbol{x}_i^\top\boldsymbol{\beta})\,\mathsf{d}\,\sigma^2 \\
&= -\frac{n}{2\sigma^2}\,\mathsf{d}\,\sigma^2 + \frac{1}{2\sigma^4}\sum_{i=1}^n W_i(\boldsymbol{\theta})(Y_i - \boldsymbol{x}_i^\top\boldsymbol{\beta})\,\mathsf{d}\,\sigma^2 \\
&= -\frac{n}{2\sigma^2}\,\mathsf{d}\,\sigma^2 + \frac{1}{2\sigma^4}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^\top\boldsymbol{W}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})\,\mathsf{d}\,\sigma^2
\end{aligned}$$

donde

$$\boldsymbol{W} = \operatorname{diag}\big(W_1(\boldsymbol{\theta}, \dots, W_n(\boldsymbol{\theta}))\big).$$

La condición de primer orden, lleva al siguiente sistema de ecuaciones:

$$\boldsymbol{X}^\top \widehat{\boldsymbol{W}}(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}) = \boldsymbol{0},$$
$$\widehat{\sigma}^2 = \frac{1}{n}(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})\widehat{\boldsymbol{W}}(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}),$$

que no tiene solución en forma explícita y por tanto métodos iterativos son requeridos.

Por ejemplo, usando una estimación inicial $\boldsymbol{\theta} = \boldsymbol{\theta}^{(k)}$, actualizamos las estimaciones para $\boldsymbol{\beta}$ y σ^2, como:

$$\boldsymbol{\beta}^{(k+1)} = (\boldsymbol{X}^\top \boldsymbol{W}^{(k)} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{W}^{(k)} \boldsymbol{Y},$$
$$\sigma^{2(k+1)} = \frac{1}{n}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}^{(k)})\boldsymbol{W}^{(k)}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}^{(k)}),$$

a la convergencia del algoritmo[2], hacemos $(\widehat{\boldsymbol{\beta}}, \widehat{\sigma}^2)$.

[2] Esto es, cuando la secuencia $\{\boldsymbol{\beta}^{(k)}, \sigma^{2(k)}\}$ se 'estabiliza'.

EX UMBRA IN SOLEM

Funciones de pesos $W(\theta)$ para algunas distribuciones elípticas

- Normal: $Y_1, \dots, Y_n \overset{\text{ind}}{\sim} \mathsf{N}(\boldsymbol{x}_i^\top \boldsymbol{\beta}, \sigma^2)$, tenemos:

$$W_i(\boldsymbol{\theta}) = 1, \qquad i = 1, \dots, n.$$

- t-Student: $Y_1, \dots, Y_n \overset{\text{ind}}{\sim} t(\boldsymbol{x}_i^\top \boldsymbol{\beta}, \sigma^2, \nu)$, $\nu > 0$,

$$W_i(\boldsymbol{\theta}) = \frac{\nu + 1}{\nu + u_i}, \qquad i = 1, \dots, n,$$

la distribución Cauchy$(\boldsymbol{x}_i^\top \boldsymbol{\beta}, \sigma^2)$, es obtenida para $\nu = 1$.

- Normal contaminada: $Y_1, \dots, Y_n \overset{\text{ind}}{\sim} \mathsf{CN}(\boldsymbol{x}_i^\top \boldsymbol{\beta}, \sigma^2, \epsilon, \gamma)$, con $\epsilon \in [0, 1)$ y $\gamma > 0$,

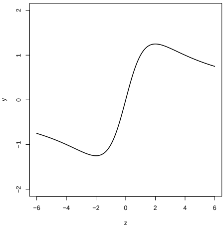
$$W_i(\boldsymbol{\theta}) = \frac{(1 - \epsilon)\exp(-u/2) + \epsilon\gamma^{-3/2}\exp(-u/(2\gamma))}{(1 - \epsilon)\exp(-u/2) + \epsilon\gamma^{-1/2}\exp(-u/(2\gamma))}, \qquad i = 1, \dots, n.$$

- Exponencial Potencia: $Y_1, \dots, Y_n \overset{\text{ind}}{\sim} \mathsf{PE}(\boldsymbol{x}_i^\top \boldsymbol{\beta}, \sigma^2, \lambda)$, $\lambda > 0$,

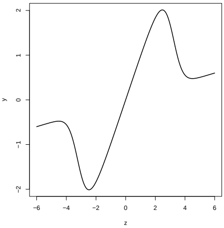
$$W_i(\boldsymbol{\theta}) = \lambda u_i^{\lambda - 1}, \qquad i = 1, \dots, n.$$

EX UMBRA IN SOLEM

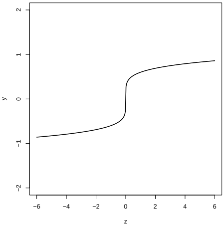
Funciones de influencia para algunas distribuciones elípticas



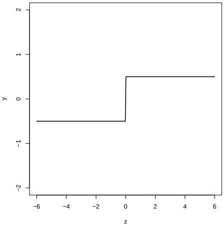
(a) t de Student, $\nu = 4$



(b) CN, $\epsilon = 0.05, \gamma = 10$



(c) PE, $\lambda = 0.6$



(d) PE, $\lambda = 0.5$

Estimación ML bajo mezclas de escala normal

Sea $Y_1, \dots, Y_n$ variable aleatorias independientes con distribución $\mathsf{SMN}(\boldsymbol{x}_i^\top \boldsymbol{\beta}, \sigma^2; \mathsf{H})$ cada una con densidad

$$f(y_i; \boldsymbol{\theta}) = (2\pi\sigma)^{-1/2} \int_0^\infty \omega^{1/2} \exp\big(-\omega u_i/2\big) \, \mathrm{d}\,\mathsf{H}(\omega),$$

donde $u_i = (Y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta})^2/\sigma^2$, $i = 1, \dots, n$.

Observación:

Una variable aleatoria $Y_i \sim \mathsf{SMN}(\boldsymbol{x}_i^\top \boldsymbol{\beta}, \sigma^2; \mathsf{H})$, admite tu representación:

$$Y_i | W = \omega \sim \mathsf{N}(\boldsymbol{x}_i^\top \boldsymbol{\beta}, \sigma^2/\omega), \qquad W \sim \mathsf{H}(\boldsymbol{\delta}),$$

lo que permite abordar la estimación ML usando el algoritmo EM.

Ejemplo: Distribución Slash

Una variable aleatoria Y_i tiene distribución Slash si su función de densidad es de la forma:

$$f(y_i) = \nu(2\pi\sigma^1)^{-1/2} \int_0^1 \omega^{\nu-1/2} \exp(-\omega u/2)\,\mathsf{d}\omega.$$

Tenemos que $h(\omega) = \nu\omega^{\nu-1}$, para $\omega \in (0,1)$ y $\nu > 0$. Es decir, $W \sim \mathsf{Beta}(\nu, 1)$.

Ejemplo: Distribución t de Student

Para $Y_i \sim t(\boldsymbol{x}_i^\top \boldsymbol{\beta}, \sigma^2, \nu)$, con $\nu > 0$, podemos escribir

$$Y_i|W \sim \mathsf{N}(\boldsymbol{x}_i^\top \boldsymbol{\beta}, \sigma^2/\omega), \qquad W \sim \mathsf{Gamma}(\nu/2, \nu/2),$$

es decir,

$$h(\omega;\nu) = \frac{(\nu/2)^{\nu/2}\omega^{\nu/2-1}}{\Gamma(\nu/2)} \exp(-\nu\omega/2).$$

Algoritmo EM (Esperanza-Maximización)

Consideraciones:

- Algoritmo para el cálculo iterativo de estimadores ML en modelos con datos incompletos.
- Requiere de una formulación de datos aumentados.
- Reemplaza una optimización "compleja" (estimación ML) por una serie de maximizaciones "simples".

EX UMBRA IN SOLEM

Formulación de datos aumentados:

Sea $\boldsymbol{Y}_{\text{obs}}$ vector de datos observados con función de densidad $f(\boldsymbol{y}_{\text{obs}}; \boldsymbol{\theta})$.

El objetivo es aumentar los datos observados $\boldsymbol{Y}_{\text{obs}}$ con variables latentes $\boldsymbol{Y}_{\text{mis}}$ (datos perdidos). Esto es, se considera el vector de datos completos

$$\boldsymbol{Y}_{\text{com}} = (\boldsymbol{Y}_{\text{obs}}^{\top}, \boldsymbol{Y}_{\text{mis}}^{\top})^{\top},$$

tal que la densidad $f(\boldsymbol{y}_{\text{com}}; \boldsymbol{\theta})$ sea simple.

Algoritmo EM (Dempster, Laird y Rubin, 1977)[3]

El algoritmo EM es útil cuando la función de log-verosimilitud

$$\begin{aligned}\ell_{\mathsf{o}}(\boldsymbol{\theta}; \boldsymbol{Y}_{\mathsf{obs}}) &= \log f(\boldsymbol{y}_{\mathsf{obs}}; \boldsymbol{\theta}) \\ &= \log \int f(\boldsymbol{y}_{\mathsf{com}}; \boldsymbol{\theta})\, \mathsf{d}\boldsymbol{y}_{\mathsf{mis}},\end{aligned}$$

es difícil de maximizar directamente.

El algoritmo EM es un procedimiento iterativo que permite realizar la estimación ML basandose en la log-verosimilitud de datos completos:

$$\ell_{\mathsf{c}}(\boldsymbol{\theta}; \boldsymbol{Y}_{\mathsf{com}}) = \log f(\boldsymbol{y}_{\mathsf{com}}; \boldsymbol{\theta}).$$

[3] Journal of the Royal Statistical Society, Series B **39**, 1-38.

Algoritmo EM (Esperanza-Maximización)

El algoritmo EM permite obtener los MLE en problemas con datos incompletos por medio de las etapas:

Paso E: para $\boldsymbol{\theta}^{(k)}$ estimación de $\boldsymbol{\theta}$ en la k-ésima iteración, calcular la Q-función,

$$\begin{aligned} Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)}) &= \mathsf{E}\{\ell_{\mathsf{c}}(\boldsymbol{\theta}; \boldsymbol{Y}_{\mathsf{com}})|\boldsymbol{Y}_{\mathsf{obs}}, \boldsymbol{\theta}^{(k)}\} \\ &= \int \ell_c(\boldsymbol{\theta}; \boldsymbol{Y}_{\mathsf{com}}) f(\boldsymbol{y}_{\mathsf{mis}}|\boldsymbol{y}_{\mathsf{obs}}; \boldsymbol{\theta}^{(k)})\, \mathsf{d}\boldsymbol{y}_{\mathsf{mis}}. \qquad (1) \end{aligned}$$

Paso M: determinar $\boldsymbol{\theta}^{(k+1)}$ como

$$\boldsymbol{\theta}^{(k+1)} = \arg\max_{\theta} Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)}). \qquad (2)$$

Una variante del Algoritmo EM

Dempster, Laird y Rubin (1977) definieron el Algoritmo EM generalizado (GEM), mediante la siguiente modificación del paso M:

Paso M*: seleccionar $\boldsymbol{\theta}^{(k+1)}$ satisfaciendo,

$$Q(\boldsymbol{\theta}^{(k+1)}|\widehat{\boldsymbol{\theta}}^{(k)}) > Q(\boldsymbol{\theta}^{(k)}|\boldsymbol{\theta}^{(k)}).$$

Sugerencia: considerar sólo un paso Newton en la optimización de $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)})$.

Teorema (Dempster, Laird y Rubin, 1977)

Todo algoritmo EM o GEM incrementa la log-verosimilitud de datos observados $\ell_{\mathsf{o}}(\boldsymbol{\theta};\boldsymbol{Y}_{\mathsf{obs}})$ en cada iteración, esto es,

$$\ell_{\mathsf{o}}(\boldsymbol{\theta}^{(k+1)};\boldsymbol{Y}_{\mathsf{obs}}) \geq \ell_{\mathsf{o}}(\boldsymbol{\theta}^{(k)};\boldsymbol{Y}_{\mathsf{obs}}).$$

Convergencia (Wu, 1983; Little y Rubin, 1987)

Bajo condiciones suaves, la secuencia $\{\boldsymbol{\theta}^{(k)}\}_{k\geq 0}$ generada por el algoritmo EM (GEM). Converge a un punto estacionario de $\ell_{\mathsf{o}}(\boldsymbol{\theta};\boldsymbol{Y}_{\mathsf{obs}})$.

EX UMBRA IN SOLEM

Propiedades del algoritmo EM:

- Frecuentemente el algoritmo EM es simple, de bajo costo computacional y numéricamente estable.
- Dempster, Laird y Rubin (1977) mostraron que el algoritmo EM converge con velocidad lineal, que depende de la proporción de información perdida.[4]
- Para modelos con datos aumentados con densidad en la familia exponencial, el algoritmo EM reduce a actualizar las estadísticas suficientes.
- Errores estándar pueden ser obtenidos por cálculo directo, diferenciación numérica o usando el Principio de Información Perdida (Louis, 1982).

[4] puede ser **extremadamente** lento.

Estimación ML usando mezclas de escala normal

Sea $Y_1, \dots, Y_n$ variables aleatorias independientes $\mathsf{SMN}(\boldsymbol{x}_i\boldsymbol{\beta}, \sigma^2; \mathsf{H})$. Se llevará a cabo la estimación ML usando el algoritmo EM.

De este modo, tenemos el siguiente modelo jerárquico:

$$Y_i | W_i \sim \mathsf{N}(\boldsymbol{x}_i^\top\boldsymbol{\beta}, \sigma^2/\omega_i), \qquad W_i \sim \mathsf{H}(\boldsymbol{\delta}), \qquad i = 1, \dots, n.$$

En este caso el vector de datos completos es $\boldsymbol{Y}_{\mathsf{com}} = (\boldsymbol{Y}^\top, \boldsymbol{\omega}^\top)^\top$, donde

$$\boldsymbol{Y} = (Y_1, \dots, Y_n)^\top, \qquad \boldsymbol{\omega} = (\omega_1, \dots, \omega_n)^\top.$$

En este contexto, $\boldsymbol{Y}$ corresponde a los datos observados, mientras que $\boldsymbol{\omega}$ serán asumidos como datos perdidos.

Asuma $\boldsymbol{\delta}$ conocido, la función de log-verosimilitud de datos completos adopta la forma:

$$
\begin{aligned}
\ell_{\mathsf{c}}(\boldsymbol{\theta}; \boldsymbol{y}_{\mathsf{com}}) &= \log f(\boldsymbol{y}_{\mathsf{com}}; \boldsymbol{\theta}) = \sum_{i=1}^{n} \log f(y_i, \omega_i; \boldsymbol{\theta}) \\
&= \sum_{i=1}^{n} \log f(y_i|\omega_i; \boldsymbol{\theta}) + \sum_{i=1}^{n} \log h(\omega_i; \boldsymbol{\delta}) \\
&= -\frac{n}{2} \log 2\pi\sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^{n} \omega_i (Y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta})^2 - \frac{1}{2} \sum_{i=1}^{n} \log \omega_i + \log h^{(n)}(\boldsymbol{\omega}; \boldsymbol{\delta}),
\end{aligned}
$$

donde $h^{(n)}(\boldsymbol{\omega}; \boldsymbol{\delta})$ denota la densidad conjunta para las variables de mezcla $\boldsymbol{\omega} = (\omega_1, \dots, \omega_n)^\top$.

Estimación ML usando mezclas de escala normal

Considere una estimación para $\boldsymbol{\theta} = \boldsymbol{\theta}^{(k)}$, entonces

$$Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(k)}) = \mathsf{E}\{\ell_{\mathsf{c}}(\boldsymbol{\theta}; \boldsymbol{y}_{\mathsf{com}})|\boldsymbol{y}; \boldsymbol{\theta}^{(k)}\} = Q_1(\boldsymbol{\beta}, \sigma^2; \boldsymbol{\theta}^{(k)}) + Q_2(\boldsymbol{\delta}; \boldsymbol{\theta}^{(k)}),$$

donde

$$Q_1(\boldsymbol{\beta}, \sigma^2; \boldsymbol{\theta}^{(k)}) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\omega_i^{(k)}(Y_i - \boldsymbol{x}_i^\top\boldsymbol{\beta})^2,$$

$$Q_2(\boldsymbol{\delta}; \boldsymbol{\theta}^{(k)}) = \mathsf{E}\{\log h^{(n)}(\boldsymbol{\omega}; \boldsymbol{\delta})|\boldsymbol{y}; \boldsymbol{\theta}^{(k)}\},$$

con $\omega_i^{(k)} = \mathsf{E}(\omega_i|\boldsymbol{x}_i; \boldsymbol{\theta}^{(k)})$ para $i = 1, \dots, n$. En general, la forma para la esperanza condicional requerida en el paso-E del algoritmo EM es dada por:

$$\mathsf{E}(\omega_i|\boldsymbol{x}_i; \boldsymbol{\theta}) = \frac{\int_0^\infty \omega_i^{3/2}\exp(-\omega_i u_i/2)\,\mathsf{dH}(\boldsymbol{\delta})}{\int_0^\infty \omega_i^{1/2}\exp(-\omega_i u_i/2)\,\mathsf{dH}(\boldsymbol{\delta})},$$

con $u_i = (Y_i - \boldsymbol{x}_i^\top\boldsymbol{\beta})^2/\sigma^2$, $i = 1, \dots, n$.

EX UMBRA IN SOLEM

Estimación ML usando mezclas de escala normal

- *t*-Student: $Y_i \overset{\text{ind}}{\sim} t(\boldsymbol{x}_i^\top \boldsymbol{\beta}, \sigma^2, \nu)$, $i = 1, \dots, n$, en cuyo caso

$$\mathsf{E}(\omega_i | Y_i; \boldsymbol{\theta}) = \frac{\nu + 1}{\nu + u_i}.$$

- Slash: $Y_i \overset{\text{ind}}{\sim} \mathsf{Slash}(\boldsymbol{x}_i^\top \boldsymbol{\beta}, \sigma^2, \nu)$, para $i = 1, \dots, n$. De este modo,

$$\mathsf{E}(\omega_i | Y_i; \boldsymbol{\theta}) = \Big(\frac{2\nu + 1}{u_i}\Big) \frac{P_1(\nu + 3/2, u_i/2)}{P_1(\nu + 1/2, u_i/2)},$$

  donde

$$P_z(a, b) = \frac{b^a}{\Gamma(a)} \int_0^z t^{a-1} e^{-bt} \, \mathsf{d}t,$$

  es la función gama incompleta (regularizada).

- Exponencial Potencia: $Y_i \overset{\text{ind}}{\sim} \mathsf{PE}(\boldsymbol{x}_i^\top \boldsymbol{\beta}, \sigma^2, \lambda)$, $i = 1, \dots, n$, donde

$$\mathsf{E}(\omega_i | Y_i; \boldsymbol{\theta}) = \lambda u_i^{\lambda - 1}, \qquad u_i \neq 0, \lambda \in (0, 1].$$

Finalmente, el algoritmo EM obtener los MLE en el modelo

$$Y_i \overset{\text{ind}}{\sim} \mathsf{SMN}(\boldsymbol{x}_i^\top \boldsymbol{\beta}, \sigma^2; \mathsf{H}), \qquad i = 1, \dots, n,$$

adopta la forma:

Paso E: para $\boldsymbol{\theta}^{(k)}$, calcular:

$$\omega_i^{(k)} = \mathsf{E}(\omega_i | Y_i; \boldsymbol{\theta}^{(k)}), \qquad i = 1, \dots, n.$$

Paso M: actualizar $\boldsymbol{\beta}^{(k+1)}$ y $\sigma^{2(k+1)}$ como:

$$\boldsymbol{\beta}^{(k+1)} = (\boldsymbol{X}^\top \boldsymbol{W}^{(k)} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{W}^{(k)} \boldsymbol{Y}$$

$$\sigma^{2(k+1)} = \frac{1}{n} (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}^{(k)})^\top \boldsymbol{W}^{(k)} (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}^{(k)}),$$

donde $\boldsymbol{W}^{(k)} = \operatorname{diag}(\omega_1(\boldsymbol{\theta}^{(k)}), \dots, \omega_n(\boldsymbol{\theta}^{(k)}))$.

A la convergencia del algoritmo hacemos $\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}}$ y $\sigma^2 = \widehat{\sigma}^2$.

Llamadas telefónicas en Bélgica 1950-1973

```
# carga datos de llamadas
> library(MASS)
> data(phones)

# carga biblioteca 'heavy'
> library(heavy)

# Ajuste usando distribución Cauchy
> fm <- heavyLm(calls ~ year, data = phones, family = Cauchy())

# Salida
> fm
Call:
heavyLm(formula = calls ~ year, data = phones, family = Cauchy())
Converged in 85 iterations

Coefficients:
 (Intercept)        year
   -53.8761      1.1229

Degrees of freedom: 24 total; 22 residual
Scale estimate: 1.394029
```

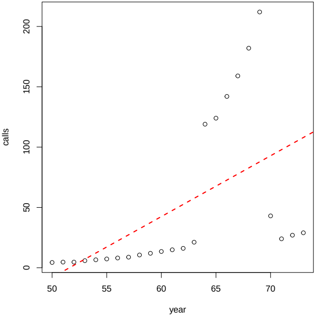
Llamadas telefónicas en Bélgica 1950-1973

```
# Salida
> summary(fm)
Linear model under heavy-tailed distributions
 Data: phones; Family: Cauchy()

Residuals:
    Min       1Q   Median       3Q      Max
 -1.848   -0.213    0.319   38.959  188.397

Coefficients:
             Estimate Std.Error Z value  p-value
(Intercept) -53.8761    3.0473  -17.6801   0.0000
year          1.1229    0.0492   22.8051   0.0000

Degrees of freedom: 24 total; 22 residual
Scale estimate: 1.394029
Log-likelihood: -102.28 on 3 degrees of freedom

# 'pesos' estimados
> fm$weights
 [1] 0.4694 0.8970 1.9514 1.9051 1.9642 1.6087 1.2597 0.8829 1.5338
[10] 1.8178 2.0000 1.8932 1.8320 0.1381 0.0003 0.0003 0.0002 0.0001
[19] 0.0001 0.0001 0.0083 0.5796 1.9988 1.2589
```

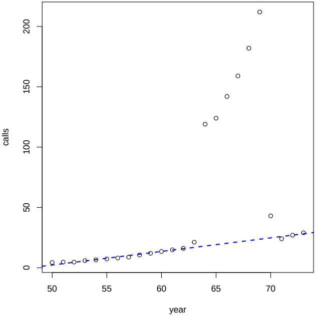
Llamadas telefónicas en Bélgica 1950-1973

```
# Estimando los 'grados de libertad'
> fm <- heavyLm(calls ~ year, data = phones, family = Student(df = 4))
> fm
Call:
heavyLm(formula = calls ~ year, data = phones,
        family = Student(df = 0.35881))
Converged in 170 iterations

Coefficients:
 (Intercept)        year
   -54.0142      1.1260

Degrees of freedom: 24 total; 22 residual
Scale estimate: 0.2661921

# 'pesos' estimados
> fm$weights
 [1] 0.0791 0.2057 2.9578 2.3797 2.7616 0.7649 0.3692 0.1843 0.6294
[10] 1.3377 3.7112 2.4381 1.9253 0.0197 0.0000 0.0000 0.0000 0.0000
[19] 0.0000 0.0000 0.0011 0.0947 3.6673 0.4728
```

Llamadas telefónicas en Bélgica 1950-1973

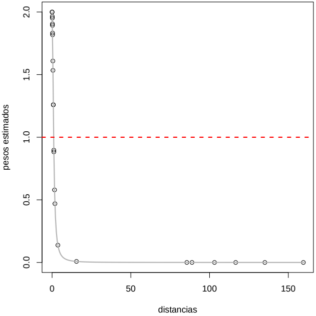


(a) normal

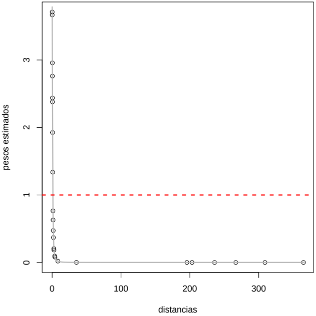


(b) Student, $\widehat{\nu} = 0.3588$

Llamadas telefónicas en Bélgica 1950-1973



(a) Cauchy



(b) Student, $\widehat{\nu} = 0.3588$