

# MAT-206: Algoritmo EM

**Felipe Osorio**

`fosorios.mat.utfsm.cl`

Departamento de Matemática, UTFSM



## Observaciones:

- ▶ El problema de **datos perdidos** (ausentes) surge en diversos contextos en estadística, por nombrar algunos:
  - ▶ Análisis de sobrevivencia y confiabilidad.
  - ▶ Análisis de series de tiempo.
  - ▶ Modelos con efectos mixtos.
  - ▶ Modelos de análisis factorial / modelos de ecuaciones estructurales (SEM).
  - ▶ Modelos de variable latente (IRT, modelos con errores en las variables).
- ▶ Existe varios enfoques para analizar modelos con datos perdidos (ver Little y Rubin, 1987)<sup>1</sup>

---

<sup>1</sup> *Statistical Analysis with Missing Data*. Wiley, New York.



# 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”**.



## Algoritmo EM (Dempster, Laird y Rubin, 1977)<sup>2</sup>

Sea  $\mathbf{Y}_{\text{com}} = (\mathbf{Y}_{\text{obs}}, \mathbf{Y}_{\text{mis}})$  vector de datos completos con función de densidad

$$p(\mathbf{y}_{\text{com}}; \boldsymbol{\theta}),$$

donde  $\mathbf{Y}_{\text{obs}}$  y  $\mathbf{Y}_{\text{mis}}$  denotan los datos observados y perdidos, respectivamente.

En modelos con datos incompletos, la log-verosimilitud de datos observados

$$\ell_o(\boldsymbol{\theta}; \mathbf{Y}_{\text{obs}}) = \log p(\mathbf{y}_{\text{obs}}; \boldsymbol{\theta}) = \log \int p(\mathbf{y}_{\text{com}}; \boldsymbol{\theta}) d\mathbf{y}_{\text{mis}},$$

puede ser difícil de maximizar directamente.

El algoritmo EM aumenta los datos  $\mathbf{Y}_{\text{obs}}$  con variables latentes permitiendo que la log-verosimilitud de datos completos

$$\ell_c(\boldsymbol{\theta}; \mathbf{Y}_{\text{com}}) = \log p(\mathbf{y}_{\text{com}}; \boldsymbol{\theta}),$$

sea bastante simple para muchas aplicaciones en Estadística.

---

<sup>2</sup>Journal of the Royal Statistical Society, Series B **39**, 1-38.



---

**Algoritmo 1:** Algoritmo EM (Esperanza-Maximización).

---

**Entrada:** Conjunto de datos observados  $\mathbf{Y}_{\text{obs}}$  y estimación inicial  $\boldsymbol{\theta}^{(0)}$ .

**Salida :** Estimación ML de  $\boldsymbol{\theta}$ .

1 **begin**

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

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

3   **Paso-M:** actualizar  $\boldsymbol{\theta}^{(k+1)}$ , como:

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

4   Iterar entre pasos-E y M hasta alcanzar convergencia.

5 **end**

---

# Algoritmo EM generalizado

---

## Algoritmo 2: Algoritmo GEM.

---

**Entrada:** Conjunto de datos observados  $\mathbf{Y}_{\text{obs}}$  y estimación inicial  $\boldsymbol{\theta}^{(0)}$ .

**Salida :** Estimación ML de  $\boldsymbol{\theta}$ .

1 **begin**

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

$$\begin{aligned} Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(k)}) &= \mathbb{E}\{\ell_c(\boldsymbol{\theta}; \mathbf{Y}_{\text{com}}) | \mathbf{Y}_{\text{obs}}, \boldsymbol{\theta}^{(k)}\} \\ &= \int \ell_c(\boldsymbol{\theta}; \mathbf{Y}_{\text{com}}) f(\mathbf{y}_{\text{mis}} | \mathbf{y}_{\text{obs}}; \boldsymbol{\theta}^{(k)}) d\mathbf{y}_{\text{mis}}. \end{aligned}$$

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

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

4 Iterar entre pasos-E y M hasta alcanzar convergencia.

5 **end**

---

### Sugerencia:

Considerar una **única** iteración Newton en la optimización de  $Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(k)})$  requerida en el **Paso-M\***.



# Propiedades del algoritmo EM

La clave del algoritmo EM está basada en el hecho que mediante maximizar  $Q(\theta; \theta^{(k)})$  permite incrementar la verosimilitud en cada etapa.

La identidad básica del algoritmo EM es

$$p(\mathbf{y}_{\text{mis}}|\mathbf{y}_{\text{obs}}; \theta) = \frac{p(\mathbf{y}_{\text{com}}; \theta)}{p(\mathbf{y}_{\text{obs}}; \theta)} \left( \frac{p(\mathbf{y}_{\text{obs}}, \mathbf{y}_{\text{mis}}; \theta)}{p(\mathbf{y}_{\text{obs}}; \theta)} \right),$$

de ahí que

$$\log p(\mathbf{y}_{\text{obs}}; \theta) = \log p(\mathbf{y}_{\text{com}}; \theta) - \log p(\mathbf{y}_{\text{mis}}|\mathbf{y}_{\text{obs}}; \theta).$$

De este modo, para cualquier  $\theta_0$

$$\begin{aligned} \log p(\mathbf{y}_{\text{obs}}; \theta) &= E\{\log p(\mathbf{y}_{\text{com}}; \theta) | \mathbf{y}_{\text{obs}}, \theta_0\} - E\{\log p(\mathbf{y}_{\text{mis}}|\mathbf{y}_{\text{obs}}; \theta) | \mathbf{y}_{\text{obs}}, \theta_0\} \\ &= Q(\theta; \theta_0) - H(\theta; \theta_0). \end{aligned}$$

(estas esperanzas son tomadas con relación a la distribución condicional  $p(\mathbf{y}_{\text{mis}}|\mathbf{y}_{\text{obs}}, \theta_0)$ ). De este modo, se tiene que

$$\begin{aligned} \ell_o(\theta^{(k+1)}) - \ell_o(\theta^{(k)}) &= Q(\theta^{(k+1)}; \theta^{(k)}) - Q(\theta^{(k)}; \theta^{(k)}) \\ &\quad - \{H(\theta^{(k+1)}; \theta^{(k)}) - H(\theta^{(k)}; \theta^{(k)})\}. \end{aligned}$$



## Propiedades del algoritmo EM

La primera diferencia es no negativa pues

$$Q(\boldsymbol{\theta}^{(k+1)}; \boldsymbol{\theta}^{(k)}) \geq Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(k)}), \quad \forall \boldsymbol{\theta},$$

para cualquier  $\boldsymbol{\theta}$

$$\begin{aligned} H(\boldsymbol{\theta}^{(k+1)}; \boldsymbol{\theta}^{(k)}) - H(\boldsymbol{\theta}^{(k)}; \boldsymbol{\theta}^{(k)}) &= \mathbb{E} \left\{ \log \frac{p(\mathbf{y}_{\text{mis}} | \mathbf{y}_{\text{obs}}; \boldsymbol{\theta})}{p(\mathbf{y}_{\text{mis}} | \mathbf{y}_{\text{obs}}; \boldsymbol{\theta}^{(k)})} \middle| \mathbf{y}_{\text{obs}}, \boldsymbol{\theta}^{(k)} \right\} \\ &\leq \log \mathbb{E} \left\{ \frac{p(\mathbf{y}_{\text{mis}} | \mathbf{y}_{\text{obs}}; \boldsymbol{\theta})}{p(\mathbf{y}_{\text{mis}} | \mathbf{y}_{\text{obs}}; \boldsymbol{\theta}^{(k)})} \middle| \mathbf{y}_{\text{obs}}, \boldsymbol{\theta}^{(k)} \right\} \\ &= \log \int p(\mathbf{y}_{\text{mis}} | \mathbf{y}_{\text{obs}}; \boldsymbol{\theta}) d\mathbf{y}_{\text{mis}} = 0. \end{aligned}$$

De este modo la secuencia  $\{\boldsymbol{\theta}^{(k)}\}$  satisface

$$\ell_o(\boldsymbol{\theta}^{(k+1)}; \mathbf{y}_{\text{obs}}) \geq \ell_o(\boldsymbol{\theta}^{(k)}; \mathbf{y}_{\text{obs}}).$$

*Observación:*

Ésta propiedad aún es válida si se selecciona un  $\boldsymbol{\theta}^{(k+1)}$  tal que:

$$Q(\boldsymbol{\theta}^{(k+1)}; \boldsymbol{\theta}^{(k)}) \geq Q(\boldsymbol{\theta}^{(k)}; \boldsymbol{\theta}^{(k)})$$





## Teorema (Dempster, Laird y Rubin, 1977)

Todo algoritmo EM o GEM **incrementa la log-verosimilitud de datos observados**  $\ell_o(\boldsymbol{\theta}; \mathbf{y}_{\text{obs}})$  en cada iteración, esto es,

$$\ell_o(\boldsymbol{\theta}^{(k+1)}; \mathbf{y}_{\text{obs}}) \geq \ell_o(\boldsymbol{\theta}^{(k)}; \mathbf{y}_{\text{obs}}).$$

## 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  $\ell_o(\boldsymbol{\theta}; \mathbf{Y}_{\text{obs}})$ .



## *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.<sup>3</sup>
- ▶ 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).

---

<sup>3</sup> puede ser **extremamente** lento.



## Estimación ML en la distribución t multivariada

Un vector aleatorio  $\mathbf{y}$  tiene una distribución  $t$  multivariada con posición  $\boldsymbol{\mu}$  y escala  $\boldsymbol{\Sigma}$ , esto es  $\mathbf{y} \sim t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ , si su función de densidad asume la forma

$$f(\mathbf{y}) = \frac{\Gamma(\frac{\nu+p}{2})}{\Gamma(\frac{\nu}{2})\pi^{p/2}} \nu^{-p/2} |\boldsymbol{\Sigma}|^{-1/2} \left(1 + \frac{u}{\nu}\right)^{-\frac{1}{2}(\nu+p)},$$

con  $u = (\mathbf{y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\mu})$ . Esta distribución puede ser escrito como

$$\mathbf{Y}|W \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}/\omega), \quad W \sim \text{Gamma}\left(\frac{\nu}{2}, \frac{\nu}{2}\right).$$

### Observación:

Recuerde que  $\text{Gamma}(a, b)$  tiene densidad

$$f(\omega) = \frac{b^a}{\Gamma(a)} \omega^{a-1} \exp(-b\omega), \quad \omega \geq 0, a, b > 0.$$



## Estimación ML en la distribución $t$ multivariada

Note que, la función de log-verosimilitud de datos observados para una muestra aleatoria de tamaño  $n$ , desde  $t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$  adopta la forma:

$$\begin{aligned}\ell_o(\boldsymbol{\theta}; \mathbf{y}_{\text{obs}}) &= \sum_{i=1}^n \log f(\mathbf{y}_i; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) \\ &= n \log \Gamma\left(\frac{\nu + p}{2}\right) - n \log \Gamma\left(\frac{\nu}{2}\right) - \frac{np}{2} \log \nu + \frac{np}{2} \log \pi \\ &\quad - \frac{n}{2} \log |\boldsymbol{\Sigma}| - \frac{\nu + p}{2} \sum_{i=1}^n \log \left(1 + \frac{u_i}{\nu}\right).\end{aligned}$$

A continuación revisaremos la estimación ML en la distribución  $t$  usando el algoritmo EM.



## Estimación ML en la distribución $t$ multivariada

Es conveniente revisar la estimación ML para la distribución  $t$  como un problema de datos incompletos. Sea,

$$\mathbf{y}_{\text{com}} = (\mathbf{y}^\top, \boldsymbol{\omega}^\top)^\top, \quad \mathbf{y} = (\mathbf{y}_1^\top, \dots, \mathbf{y}_n^\top)^\top,$$

con  $\boldsymbol{\omega} = (\omega_1, \dots, \omega_n)^\top$  representa los datos perdidos. De este modo,

$$\ell_c(\boldsymbol{\theta}; \mathbf{y}_{\text{com}}) = \log p(\mathbf{y}|\boldsymbol{\omega}; \boldsymbol{\theta}) + \log p(\boldsymbol{\omega}; \boldsymbol{\theta}),$$

donde

$$\log p(\mathbf{y}|\boldsymbol{\omega}; \boldsymbol{\theta}) = -\frac{np}{2} \log 2\pi - \frac{n}{2} \log |\boldsymbol{\Sigma}| - \frac{1}{2} \sum_{i=1}^n \omega_i (\mathbf{y}_i - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{y}_i - \boldsymbol{\mu})$$

$$\log p(\boldsymbol{\omega}; \boldsymbol{\theta}) = -n \log \Gamma\left(\frac{\nu}{2}\right) + \frac{n\nu}{2} \log\left(\frac{\nu}{2}\right) + \frac{\nu}{2} \sum_{i=1}^n (\log \omega_i - \omega_i) - \sum_{i=1}^n \log \omega_i$$



Por otro lado, se puede mostrar de forma sencilla que la distribución condicional de  $\omega_i | \mathbf{y}_i$  es

$$\omega_i | \mathbf{y}_i \sim \text{Gamma}\left(\frac{\nu + p}{2}, \frac{\nu + u_i}{2}\right), \quad i = 1, \dots, n,$$

y sigue que

$$E(\omega_i | \mathbf{y}_i) = \frac{\nu + p}{\nu + u_i}.$$

De este modo,

$$\omega_i^{(k)} = E(\omega_i | \mathbf{y}_i, \boldsymbol{\theta}^{(k)}) = \frac{\nu + p}{\nu + u_i^{(k)}}, \quad i = 1, \dots, n,$$

donde  $u_i^{(k)} = (\mathbf{y}_i - \boldsymbol{\mu}^{(k)})^\top \{\boldsymbol{\Sigma}^{(k)}\}^{-1} (\mathbf{y}_i - \boldsymbol{\mu}^{(k)})$ .



## Estimación ML en la distribución t multivariada

Por tanto, la parte relevante de la función  $Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(k)})$  está dada por (estamos asumiendo  $\nu$  conocido):

$$Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(k)}) = -\frac{n}{2} \log |\boldsymbol{\Sigma}| - \frac{1}{2} \sum_{i=1}^n \omega_i^{(k)} (\mathbf{y}_i - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_i - \boldsymbol{\mu}).$$

Diferenciando con relación a  $\boldsymbol{\mu}$  y  $\boldsymbol{\Sigma}$  y resolviendo la condición de primer orden, tenemos

$$\begin{aligned}\boldsymbol{\mu}^{(k+1)} &= \frac{1}{\sum_{i=1}^n \omega_i^{(k)}} \sum_{i=1}^n \omega_i^{(k)} \mathbf{y}_i, \\ \boldsymbol{\Sigma}^{(k+1)} &= \frac{1}{n} \sum_{i=1}^n \omega_i^{(k)} (\mathbf{y}_i - \boldsymbol{\mu}^{(k+1)}) (\mathbf{y}_i - \boldsymbol{\mu}^{(k+1)})^\top.\end{aligned}$$



---

**Algoritmo 3:** Algoritmo EM en la  $t$  multivariada.

---

**Entrada:** Conjunto de datos observados  $\mathbf{y}_1, \dots, \mathbf{y}_n$  y estimación inicial  $\boldsymbol{\theta}^{(0)}$ .

**Salida :** Estimación ML de  $\boldsymbol{\mu}$  y  $\boldsymbol{\Sigma}$ .

1 **begin**

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

$$\omega_i^{(k)} = \frac{\nu + p}{\nu + u_i^{(k)}}, \quad i = 1, \dots, n.$$

3   **Paso-M:** actualizar  $\boldsymbol{\mu}^{(k+1)}$  y  $\boldsymbol{\Sigma}^{(k+1)}$  como:

$$\boldsymbol{\mu}^{(k+1)} = \frac{1}{\sum_{i=1}^n \omega_i(\boldsymbol{\theta}^{(k)})} \sum_{i=1}^n \omega_i(\boldsymbol{\theta}^{(k)}) \mathbf{x}_i, \quad (3)$$

$$\boldsymbol{\Sigma}^{(k+1)} = \frac{1}{n} \sum_{i=1}^n \omega_i(\boldsymbol{\theta}^{(k)}) (\mathbf{x}_i - \boldsymbol{\mu}^{(k+1)})(\mathbf{x}_i - \boldsymbol{\mu}^{(k+1)})^\top. \quad (4)$$

4   Iterar entre pasos-E y M hasta alcanzar convergencia.

5 **end**

---

A la convergencia del algoritmo hacemos  $\boldsymbol{\mu} = \hat{\boldsymbol{\mu}}$  y  $\boldsymbol{\Sigma} = \hat{\boldsymbol{\Sigma}}$ .





## Una curiosa propiedad de la distribución $t$ multivariada

Desde (4), debemos tener que a la convergencia del algoritmo:

$$\widehat{\Sigma} = \frac{1}{n} \sum_{i=1}^n \widehat{\omega}_i (\mathbf{x}_i - \widehat{\mu})(\mathbf{x}_i - \widehat{\mu})^\top,$$

así premultiplicando por  $\widehat{\Sigma}^{-1}$  y aplicando traza, tenemos:

$$\begin{aligned} \text{tr } \mathbf{I}_p &= \frac{1}{n} \sum_{i=1}^n \widehat{\omega}_i \text{tr } \Sigma^{-1} (\mathbf{x}_i - \widehat{\mu})(\mathbf{x}_i - \widehat{\mu})^\top \\ p &= \frac{1}{n} \sum_{i=1}^n \widehat{\omega}_i \widehat{u}_i, \end{aligned} \tag{5}$$

usando la función de pesos asociada a la distribución  $t$ , tenemos que:

$$\nu + p = \widehat{\omega}_i (\nu + \widehat{u}_i) = \widehat{\omega}_i \nu + \widehat{\omega}_i \widehat{u}_i,$$

promediando y usando (5), lleva a:

$$\nu + p = \nu \frac{1}{n} \sum_{i=1}^n \widehat{\omega}_i + p, \quad \Rightarrow \quad \frac{1}{n} \sum_{i=1}^n \widehat{\omega}_i = 1.$$



# Estimación ML en la distribución t multivariada

La consideración anterior llevó a Kent, Tyler y Vardi (1994)<sup>4</sup> a proponer la siguiente variante del Algoritmo 3:

---

**Algoritmo 4:** Algoritmo EM en la  $t$  multivariada.

---

**Entrada:** Conjunto de datos observados  $\mathbf{y}_1, \dots, \mathbf{y}_n$  y estimación inicial  $\boldsymbol{\theta}^{(0)}$ .

**Salida :** Estimación ML de  $\boldsymbol{\mu}$  y  $\boldsymbol{\Sigma}$ .

1 **begin**

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

$$\omega_i^{(k)} = \frac{\nu + p}{\nu + u_i^{(k)}}, \quad i = 1, \dots, n.$$

3   **Paso-M:** actualizar  $\boldsymbol{\mu}^{(k+1)}$  y  $\boldsymbol{\Sigma}^{(k+1)}$  como:

$$\boldsymbol{\mu}^{(k+1)} = \frac{1}{\sum_{i=1}^n \omega_i(\boldsymbol{\theta}^{(k)})} \sum_{i=1}^n \omega_i(\boldsymbol{\theta}^{(k)}) \mathbf{x}_i,$$

$$\boldsymbol{\Sigma}^{(k+1)} = \frac{1}{\sum_{i=1}^n \omega_i(\boldsymbol{\theta}^{(k)})} \sum_{i=1}^n \omega_i(\boldsymbol{\theta}^{(k)}) (\mathbf{x}_i - \boldsymbol{\mu}^{(k+1)})(\mathbf{x}_i - \boldsymbol{\mu}^{(k+1)})^\top.$$

4   Iterar entre pasos-E y M hasta alcanzar convergencia.

5 **end**

---

Posteriormente, Liu, Rubin y Wu (1998)<sup>5</sup> identificaron esta variante en la clase de algoritmos EM (PX-EM) de parámetros-expandidos.

---

<sup>4</sup>Communications in Statistics - Simulation and Computation **23**, 441-453.

<sup>5</sup>Biometrika **85**, 755-770.



## Estimación ML en la distribución t multivariada

Cuando los grados de libertad  $\nu$  son desconocidos, debemos añadir la siguiente sub-etapa al Paso-M del Algoritmo 3 ó 4.

$$\nu^{(k+1)} = \arg \max_{\nu} Q_2(\nu; \boldsymbol{\theta}^{(k)}),$$

donde

$$Q_2(\nu; \boldsymbol{\theta}^{(k)}) = \frac{n\nu}{2} \log \left( \frac{\nu}{2} \right) - n \log \Gamma \left( \frac{\nu}{2} \right) + \frac{n\nu}{2} \left\{ \frac{1}{n} \sum_{i=1}^n (\log \omega_i^{(k)} - \omega_i^{(k)}) \right. \\ \left. + \psi \left( \frac{\nu^{(k)} + p}{2} \right) - \log \left( \frac{\nu^{(k)} + p}{2} \right) \right\},$$

con  $\psi(z) = d \log \Gamma(z) / dz$  la función digama.

### *Observación:*

Podemos actualizar  $\nu^{(k+1)}$  usando un método de optimización uni-dimensional.



# Referencias bibliográficas



Dempster, A.P., Laird, N.M., Rubin, D.B. (1977).

Maximum likelihood from incomplete data via the EM algorithm (with discussion)  
*Journal of the Royal Statistical Society, Series B* **39**, 1-38.



Kent, J.T., Tyler, D.E., Vardi, Y. (1994).

A curious likelihood identity for the multivariate  $t$ -distribution.  
*Communication in Statistics: Simulation and Computation* **23**, 441-453.



Lange, K., Sinsheimer, J.S. (1993).

Normal/independent distributions and their applications in robust regression.  
*Journal of Computational and Graphical Statistics* **2**, 175-198.



Little, R.J.A. (1988).

Robust estimation of the mean and covariance matrix from data with missing values.  
*Applied Statistics* **37**, 23-38.



Little, R.J.A., Rubin, D.B. (1987).  
*Statistics Analysis with Missing Data*.  
Wiley, New York.



Liu, C., Rubin, D.B., Wu, Y.N. (1998).

Parameter expansion to accelerate EM: The PX-EM algorithm.  
*Biometrika* **85**, 775-770.

