Algoritmo EM: Parte 1

Prof. Caio Azevedo

Exemplo de Rao (tetranomial)

- Considere o exemplo descrito em Rao (1965, pp. 368-369), em que 197 animais são distribuídos multinomialmente em 4 categorias. Os dados observados consistem em

  $y = (y_1, y_2, y_3, y_4) = (125, 18, 20, 34)$.
- Se $Y \sim \text{Multonomial}_{p-1}(m, \theta_1, \theta_2, ..., \theta_{p-1})$, então

$$f(y;\boldsymbol{\theta}) = \frac{m!}{\prod_{i=1}^{p-1} y_i!(m - \sum_{i=1}^{p-1} y_i)!} \left(\prod_{i=1}^{p-1} \theta_i^{y_i}\right) \left(1 - \sum_{i=1}^{p-1} \theta_i\right)^{m - \sum_{i=1}^{m-1}} \prod_{i=1}^{m-1} \mathbb{1}_{\{0,1,\ldots,m\}}(y_i) \mathbb{1}_{\{0,1,\ldots,m\}}(\sum_{i=1}^{m-1} y_i)$$

Cont.

- Neste caso, tem-se o interesse em se ajuster o seguinte modelo genético:

$$\left(\frac{1}{2}+\frac{1}{4}\theta, \frac{1}{4}(1-\theta), \frac{1}{4}(1-\theta), \frac{\theta}{4}\right)$$

- Assim, tem-se que $(y_4 = 197 - y_1 - y_2 - y_3)$:

$$\begin{aligned} f(\mathbf{y};\theta) &= \frac{(y_1+y_2+y_3+y_4)!}{y_1!y_3!y_3!y_4!}\left(\frac{1}{2}+\frac{1}{4}\theta\right)^{y_1}\left(\frac{1}{4}(1-\theta)\right)^{y_2} \\ &\quad \left(\frac{1}{4}(1-\theta)\right)^{y_3}\left(\frac{\theta}{4}\right)^{y_4} \end{aligned}$$

- **Exercício: desenvolver e implementar o algoritmo Escore de Fisher para estimar θ, bem como implementar o TRV para testar a valdade do modelo proposto.**

Cont.

- Para maximizar $l(\theta) = \ln L(\theta)$, vamos "aumentar" a dimensão do modelo multinomial, considerando

  $y_1 = x_1 + x_2, y_3 = x_2, y_4 = x_3, y_5 = x_4$, assim

$$
\begin{aligned}
f(\mathbf{x};\theta) &= \frac{(x_1+x_2+x_3+x_4+x_5)!}{x_1!x_3!x_3!x_4!x_5!}\left(\frac{1}{2}\right)^{x_1}\left(\frac{1}{4}\theta\right)^{x_2}\left(\frac{1}{4}(1-\theta)\right)^{x_3} \\
&\quad \left(\frac{1}{4}(1-\theta)\right)^{x_4}\left(\frac{\theta}{4}\right)^{x_5}
\end{aligned}
$$

- Resultado importante: pode-se provar que $f_{(y;\theta)} = \sum_{x\in Y(\Omega)} f(x;\theta)$. Neste caso, equivale a somar $f_{(\mathbf{x};\theta)}$ nos pares ordenados (x_1, x_2), com a restrição de que $x_1 + x_2 = y$.

Cont.

- Temos que:

$$
\begin{aligned}
l(\theta) &= const. + x_1 \ln(1/2) + x_2 \ln\left(\frac{\theta}{4}\right) + x_3 \ln\left(\frac{1}{4} - \frac{\theta}{4}\right) \\
&+ x_4 \ln\left(\frac{1}{4} - \frac{\theta}{4}\right) + x_5 \ln\left(\frac{\theta}{4}\right)
\end{aligned}
$$

- Note que x_1, x_2 não são observáveis. Tomaremos a $\mathcal{E}_{(X_1,X_2|(y_1,y_2,y_3,y_4,\theta^{(t)}))}(l(\theta)|y_1,y_2,y_3,y_4,\theta^{(t)})$, em que $\theta^{(t)}$ é uma estimativa provisória de θ.

Cont.

- Note que $X_1|y_1,y_2,y_3,y_4,\theta^{(t)} \sim \text{binomial}\left(y_1 = 125, \frac{1/2}{1/2+\theta^{(t)}/4}\right)$ e $X_2|y_1,y_2,y_3,y_4,\theta^{(t)} \sim \text{binomial}\left(y_1 = 125, \frac{\theta^{(t)}/4}{1/2+\theta^{(t)}/4}\right)$.
- Seja $x_i^{(t)} = \mathcal{E}_{(X_1,X_2|(y_1,y_2,y_3,y_4,\theta^{(t)}))}(X_i|y_1,y_2,y_3,y_4,\theta^{(t)})$. Assim, temos que

$$x_1^{(t)} = 125\frac{1/2}{1/2+\theta^{(t)}/4}, x_2^{(t)} = 125\frac{\theta^{(t)}/4}{1/2+\theta^{(t)}/4}$$

- Dado $x_i^{(t)}$, a maximização com respeito a θ de $l(\theta)$, nos leva à:

$$\theta^{(t+1)} = \frac{x_2^{(t)} + 34}{x_2^{(t)} + 34 + 18 + 20}$$

Cont.

- Algoritmo EM. Dado um valor inicial para θ, digamos $\theta^{(0)}$, repita

Passo E: Calcule:

$$x_1^{(t)} = 125\frac{1/2}{1/2+\theta^{(t)}/4}, x_2^{(t)} = 125\frac{\theta^{(t)}/4}{1/2+\theta^{(t)}/4}$$

Passo M: Calcule:

$$\theta^{(t+1)} = \frac{x_2^{(t)}+34}{x_2^{(t)}+34+18+20}$$

até que algum critério de convergência seja obtido.

Cont.

- Histórico de iterações

t	$\theta^{(t)}$	$\theta^{(t)} - \theta^{(*)}$	$(\theta^{(t+1)} - \theta^{(*)})/(\theta^{(t+1)} - \theta^{(*)})$
0	0.500000000	0.126821498	0.1465
1	0.608247423	0.018574075	0.1346
2	0.624321051	0.002500447	0.1330
3	0.626488879	0.000332619	0.1328
4	0.626777323	0.000044176	0.1328
5	0.626815632	0.000005866	0.1328
6	0.626820719	0.000000779	-
7	0.626821395	0.000000104	-
8	0.626821484	0.000000014	-

- Seja $\mathbf{Y}_1, ..., \mathbf{Y}_n$ uma a.a. de $\mathbf{Y} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Desejamos estimar $\boldsymbol{\mu}$ e $\boldsymbol{\Sigma}$ por máxima verossimilhança.

$$\mathbf{Y} = \begin{bmatrix} Y_{11} & Y_{21} & \ldots & Y_{p1} \\ Y_{21} & Y_{22} & \ldots & Y_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ Y_{1n} & Y_{2n} & \ldots & Y_{pn} \end{bmatrix}$$

- Se tivermos todas as observações, então $\widehat{\boldsymbol{\mu}} = \frac{1}{n}\sum_{j=1}^{n} \mathbf{Y}_j$ e $\boldsymbol{\Sigma} = \frac{1}{n}\sum_{j=1}^{n} \left(\mathbf{Y}_j - \overline{\mathbf{Y}}\right)' \left(\mathbf{Y}_j - \overline{\mathbf{Y}}\right)$.
- Suponha que $p = 2$ e que, para $m <<< n$ indivíduos, exatamente uma das observações esteja faltando (Y_{1j} ou Y_{j2}). Por simplicidade, suponha que $\boldsymbol{\Sigma}$ é conhecida.

- Matriz de dados.

$$\mathbf{Y} = \begin{bmatrix} Y_{11} & - \\ Y_{21} & Y_{22} \\ \vdots & \vdots \\ - & Y_{pn} \end{bmatrix}$$

- Imputar as observações perdidas usando o fato de que

  $Y_{j1}|y_{j2}, \boldsymbol{\mu}, \boldsymbol{\Sigma} \sim N_1(\overline{\mu}_1, \overline{\sigma}_1^2)$, em que $\overline{\mu}_1 = \mu_1 + \sigma_{12}\left(\sigma_2^2\right)^{-1}(y_{j2} - \mu_2)$ e $\overline{\sigma}_1^2 = \sigma_1^2 - (\sigma_{12})^2\left(\sigma_2^2\right)^{-1}$.

  $Y_{j2}|y_{j1}, \boldsymbol{\mu}, \boldsymbol{\Sigma} \sim N_1(\overline{\mu}_2, \overline{\sigma}_2^2)$, em que $\overline{\mu}_2 = \mu_2 + \sigma_{12}\left(\sigma_1^2\right)^{-1}(y_{j1} - \mu_1)$ e $\overline{\sigma}_2^2 = \sigma_2^2 - (\sigma_{12})^2\left(\sigma_2^1\right)^{-1}$.

- Log-verossimilhança

$$l(\boldsymbol{\mu}) = const. - 0.5\sum_{j=1}^{n}\left(\mathbf{y}'\boldsymbol{\Sigma}^{-1}\mathbf{y} + \boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - 2\mathbf{y}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\right)$$

$$\frac{\partial l(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}} = -0.5\sum_{j=1}^{n}\left(2\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - 2\boldsymbol{\Sigma}^{-1}\mathbf{y}\right)$$

- "Máxima verossimilhança com imputação". Inicie o processo com valores iniciais $\boldsymbol{\mu}^{(t)}$.
- Cálculo da esperança (Passo E): Calcular, conforme a observação faltante para cada indivíduo (caso necessário),
$y_{j1}^{*} = \mathcal{E}(Y_{j1}|y_{j2}, \boldsymbol{\mu}^{(t)}, \boldsymbol{\Sigma}^{(t)}) = \mu_1^{(t)} + \sigma_{12}^{(t)}\left(\sigma_2^{2(t)}\right)^{-1}\left(y_{j2} - \mu_2^{(t)}\right)$
$y_{j2}^{*} = \mathcal{E}(Y_{j2}|y_{j1}, \boldsymbol{\mu}^{(t)}, \boldsymbol{\Sigma}^{(t)}) = \mu_2^{(t)} + \sigma_{12}^{(t)}\left(\sigma_1^{2(t)}\right)^{-1}\left(y_{j1} - \mu_1^{(t)}\right)$
- Maximização da log-verossimilhança (Passo M): Com a matriz de dados completa, calcular
$\widehat{\boldsymbol{\mu}}^{(t+1)} = \frac{1}{n}\sum_{j=1}^{n}\mathbf{Y}_j^{(*)} = \overline{\mathbf{Y}}^{(t)}$ e

Algorito EM

- Seja $\mathbf{Y}$ o conjunto de variáveis observadas e $\mathbf{Y}^*$ o conjunto de variáveis não observadas (não observáveis, dados faltantes).
- Em geral, $\mathbf{Y}$ é chamado de dados incompletos e $(\mathbf{Y}, \mathbf{Y}^*)$ são os dados completos.
- Seja $l(\boldsymbol{\theta})$ a logveerossimilhança em que $\boldsymbol{\theta} \in \Theta \subset \mathcal{R}^p$ é o conjunto de parâmetros a ser estimado.

Cont.

- Seja $\boldsymbol{\theta}^{(0)}$ valores iniciais apropriados como estimativas de $\boldsymbol{\theta}$. Repita o processo a seguir:

Passo E: Calcule a esperança condicional (na log-verossimilhança) dos dados faltantes condicionado as varláveis observadas e à estimativas provisórias de $\boldsymbol{\theta}^{(t)}$, ou seja $\mathcal{E}_{\mathbf{Y}^*|Yb,\boldsymbol{\theta}^{(t)}}(l(\boldsymbol{\theta})|\mathbf{y},\boldsymbol{\theta}^{(t)})$.

Passo M: Maximixar a esperança acima em relação à $\boldsymbol{\theta}$.

até que que algum critério de convergência seja alcançado.

Estrutura do algoritmo EM

- Seja $l(\boldsymbol{\theta}, \mathbf{y}, \mathbf{y}^*)$ a log-verossimilhança aumentada e $\boldsymbol{\theta}^{(t)}$ estimativas provisórias para $\boldsymbol{\theta}$. O algoritmo EM pode ser resumido nos seguintes passos

Passo E: Calcule a esperança condicional (na log-verossimilhança) dos dados faltantes condicionado as varláveis observadas e à estimativas provisórias de $\boldsymbol{\theta}^{(t)}$, ou seja

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) = \mathcal{E}\left[l(\boldsymbol{\theta}, \mathbf{y}, \mathbf{y}^*)|\mathbf{y}, \boldsymbol{\theta}\right]$$

Passo M: Maximixar a esperança acima em relação à $\boldsymbol{\theta}$ ou seja, obter

$$\boldsymbol{\theta}^{(t+1)} = \text{argmax}_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$$

até que que algum critério de convergência seja alcançado.

Observações

- O cálculo das esperanças necessárias podem ser complicadas, necessitando do emprego de aproximações analítivas ou numéricas.
- Para a maximização pode ser necessário o emprego de métodos numéricos.
- Na família exponencial as contas ficam mais simples.
- A utilização de valores iniciais apropriados para $\boldsymbol{\theta}$ auxilia na convergência do algoritmo EM.
- A introdução de variáveis "aumentadas" apropriadas, é outro fator de importância na convergência e obtenção de estimativas acuradas.

Regressão Probito

- Seja $Y_1, ..., Y_n$ uma a.a. de $Y \sim \text{Bernoulli}(p_i), p_i = \Phi\left(\beta_0 + X_i\beta_1\right)$. Desejamos estimar (β_0, β_1).
- Verossimilhança:

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} p_i^{y_i}(1-p_i)^{1-y_i}$$

- Defina:

$$\begin{aligned} Y_i &= I_{(Z_i>0)} \\ Z_i &\sim N(\beta_0 + X_i\beta_1, 1) \end{aligned}$$

Regressão Probito

- Assim, pode-se provar que

$$Z_i|y_i, x_i, \boldsymbol{\beta} = \begin{cases} N_{(0,\infty)}\left(\beta_0 + X_i\beta_1, 1\right) & \text{se } y_i = 1 \\ N_{(-\infty,0)}\left(\beta_0 + X_i\beta_1, 1\right) & \text{se } y_i = 0 \end{cases} \tag{1}$$

- Assim, pode-se trabalhar com a seguinte verossimilhança

$$L(\boldsymbol{\beta}, \mathbf{z}, \mathbf{y}) \propto \prod_{i=1}^{n} exp\left[-0,5\left(z_i - \beta_0 - X_i\beta_1\right)^2\right]$$

Cont.

- Resumindo:

$$\begin{aligned} Y_i &\sim \text{Bernoulli}(p_i) \\ Y_i|z_i &= 1, \text{se } z_i \geq 0, 0 \text{se } z_i < 0 \\ Z_i|y_i &\sim N_{(\mathbb{1}_{(z_i \geq 0, y_i = 1, z_i < 0, y_i = 0)})}(\beta_0 + X_I \beta_1, 1) \end{aligned}$$

- Logo, a verossimilhança completa (dados aumentados) é dada por

$$f(\mathbf{z}, \mathbf{y}, \boldsymbol{\beta}) = L(\boldsymbol{\beta}, \mathbf{z}, \mathbf{y}) \propto \prod_{i=1}^{n} exp\left[-0,5\left(z_i - \beta_0 - X_i \beta_1\right)^2\right]$$

Cont.

- Log-verossimilhança:

$$L(\beta) = const. - \exp\left[-0.5\left(\mathbf{z} - \mathbf{X}\beta\right)'\left(\mathbf{z} - \mathbf{X}\beta\right)\right]$$

- É necessário calcular apenas $\mathcal{E}(Z_i|y_i, x_i, \beta^{(t)})$

Cont.

- Algoritmo EM:

Passom E: Calcular

$$\mathcal{E}(Z_i|y_i, x_i, \boldsymbol{\beta}^{(t)}) = \begin{cases} \mu_i(t) + \frac{\phi(-\mu_i(t))}{1-\Phi(-\mu_i(t))}, & \text{se } y_i = 1 \\ \mu_i(t) - \frac{\phi(-\mu_i(t))}{\Phi(-\mu_i(t))}, & \text{se } y_i = 0 \end{cases}$$

em que $\mu_i^{(t)} = \beta_0^{(t)} + X_i\beta_1^{(t)}$, ϕ e Φ são a densidade e a fda da distribuição $N(0,1)$.

Passom E: Calcular

$$\boldsymbol{\beta}^{(t+1)} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Z}$$

Mistura finita de normais

- Amostra aleatória de $Y \sim MFN_2\left(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, p\right)$, ou seja

$$f(y;\boldsymbol{\theta}) = pf_{Y_1}(y;\mu_1,\sigma_1^2) + (1-p)f_{Y_2}(y;\mu_2,\sigma_2^2)$$

$$\boldsymbol{\theta} = (p,\mu_1,\mu_2,\sigma_1^2,\sigma_2^2), p \in (0,1), \mu_i \in \mathcal{R}, \sigma_i^2 \in \mathcal{R}^2, i = 1,2$$

em que $f_{Y_i}(.;.)$ é a fdp de uma distribuição $N(\mu_i,\sigma_i^2), i = 1,2$.

Aplicação do algoritmo EM

- Defina uma variável latente Z, tal que $Z \sim \text{Bernoulli}(p)$ e:

$$Y = \begin{cases} Y_1, & \text{se } Z = 1 \\ Y_2 & \text{se } Z = 0 \end{cases}$$

- Assim, $Y|z = 1 \sim N(\mu_1, \sigma_1^2)$ e $Y|z = 0 \sim N(\mu_2, \sigma_2^2)$.
- Logo, $Z|y \sim \text{Bernoulli}(p^*)$, em que $p^* = \frac{pf_{Y_1}(y;\mu_1,\sigma_1^2)}{pf_{Y_1}(y;\mu_1,\sigma_1^2)+(1-p)f_{Y_2}(y;\mu_2,\sigma_2^2)}$
- Considerando, agora uma amostra aleatória de Y, $Y_1, Y_2, ..., Y_n$ e definindo Z_i para cada Y_i.

Cont.

- Logo:

  $f(\mathbf{y},\mathbf{z}|\boldsymbol{\theta}) = \prod_{i=1}^{n}\left\{\left(pf_{Y_1}(y;\mu_1,\sigma_1^2)\right)^{z_i}\left((1-p)f_{Y_2}(y;\mu_2,\sigma_2^2)\right)^{1-z_i}\right\}$

- A log-verossimilhança é dada por:

$$
\begin{aligned}
l(\boldsymbol{\theta},\mathbf{y},\mathbf{z}) &= \sum_{i=1}^{n}\{z_i\left(\ln p + \ln f_{Y_1}(y_i;\mu_1,\sigma_1^2)\right) \\
&+ (1-z_i)\left(\ln(1-p) + \ln f_{Y_2}(y_i;\mu_2,\sigma_2^2)\right)\}
\end{aligned}
$$

Cont.

- Portanto

$$
\begin{aligned}
Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) &= \sum_{i=1}^{n}\{z_i^{(t)}\left(\ln p + \ln f_{Y_1}(y_i;\mu_1,\sigma_1^2)\right) \\
&+ (1-z_i^*)\left(\ln(1-p) + \ln f_{Y_2}(y_i;\mu_2,\sigma_2^2)\right)\}
\end{aligned}
$$

em que

$$
z_i^{(t)} = \mathcal{E}(Z_i|y_i,\boldsymbol{\theta}^{(t)}) = p_i^{(t)} = \frac{p^{(t)}f_{Y_1}(y;\mu_1^{(t)}),\sigma_1^{2(t)}}{p^{(t)}f_{Y_1}(y_i;\mu_1^{(t)},\sigma_1^{2(t)}) + (1-p^{(t)})f_{Y_2}(y_i;\mu_2^{(t)},\sigma_2^{2(t)})}
$$

Cont.

- Por outro lado, maximizando-se $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$, tem-se que

$$
\begin{aligned}
\mu_j^{(t+1)} &= \frac{\sum_{i=1}^n z_i^{(t)} y_i}{\sum_{i=1}^n z_i^{(t)}} \\
\sigma_j^{2(t+1)} &= \frac{\sum_{i=1}^n z_i^{(t)} \left(y_i - \boldsymbol{\mu}_j^{(t+1)}\right)^2}{\sum_{i=1}^n z_i^{(t)}} \\
p^{(t+1)} &= \frac{1}{n}\sum_{i=1}^n z_i^{(t)}
\end{aligned}
$$

Modelo linear misto

- Suponha que $i = 1, .., n$ indivíduos sejam estudados durante $j = 1, ..., t$ instantes de avaliação, em relação à uma característica Y_{ij}.
- Seja

$$
\begin{aligned}
Y_{ij} &= \mu_j + b_i + \xi_{ij} \\
b_i &\overset{iid}{\sim} N(0, \psi) \\
\xi_{ij} &\overset{iid}{\sim} N(0, \sigma^2)
\end{aligned}
$$

$\xi_{ij} \perp b_i, \forall i, j$

Cont.

- Desejamos estimar $\boldsymbol{\theta} = (\mu_j, \psi, \sigma^2)$ e predizer b_i (variáveis latentes).
- Uma possibilidade: verossimilhança marginal. Necessário obter a distribuição marginal de $\mathbf{Y}$ em relação à $\mathbf{b}$.
- Outra possibilidade: considerar $\mathbf{b}$ como os dados faltantes.

Cont.

- Verossimilhança aumentada:

$$
\begin{aligned}
L(\boldsymbol{\theta}; \mathbf{y}, \mathbf{b}) &\propto \left\{ \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i,j} (y_{ij} - \mu_j - b_i)^2 \right\} \right\} \\
&\times \left\{ \exp\left\{ -\frac{1}{2\psi} \sum_{i} (b_i)^2 \right\} \right\} (\sigma^2)^{-nt/2} \psi^{-n/2}
\end{aligned}
$$

Cont.

- Log-verossimilhança:

$$
\begin{aligned}
l(\boldsymbol{\theta}; \mathbf{y}, \mathbf{b}) &= -\frac{1}{2\sigma^2}\sum_{i,j}\left((y_{ij}-\mu_j)^2 - 2(y_{ij}-\mu_j)b_i + b_i^2\right) \\
&- \frac{1}{2\psi}\sum_i (b_i)^2 - \frac{nt}{2}\ln\sigma^2 - \frac{n}{2}\ln\psi + const
\end{aligned}
$$

- Portanto:

$$
\begin{aligned}
Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) &= -\frac{1}{2\sigma^2}\sum_{i,j}\left((y_{ij}-\mu_j)^2 - 2(y_{ij}-\mu_j)b_i^{(t)} + b_i^{2(t)}\right) \\
&- \frac{1}{2\psi}\sum_i (b_i)^{2(t)} - nt\ln\sigma^2 - t\ln\psi + const
\end{aligned}
$$

Cont.

- Em que

$$b_i^{(t)} = \mathcal{E}(b_i|\mathbf{y}_{i.}, \boldsymbol{\mu}^{(t)}, \sigma^{2(t)}, \psi^{(t)})$$

$$b_i^{2(t)} = \mathcal{E}(b_i^{(2)}|\mathbf{y}_{i.}, \boldsymbol{\mu}^{(t)}, \sigma^{2(t)}, \psi^{(t)})$$

Cont.

- Além disso

$$b_i|\mathbf{y}_{i.}, \boldsymbol{\mu}, \sigma^2, \psi \overset{ind.}{\sim} N(\mu^*_{b_i}, \psi^*),$$

em que

$$\psi^* = \left(\frac{1}{\psi} + \frac{t}{\sigma^2}\right)^{-1}$$
$$\mu^*_{b_i} = (\psi^*)^{-1}\left(\frac{t}{\sigma^2}(\overline{y}_{i.} - \overline{\mu})\right)$$

$$\overline{y}_{i.} = \frac{1}{t}\sum_{j=1}^{t} y_{ij}; \overline{\mu} = \frac{1}{t}\sum_{j=1}^{t} \mu_j$$

Cont.

- Por fim, maximizando-se a log-verossimilhança em relação à $\boldsymbol{\theta}$, temos

$$
\begin{aligned}
\mu_j^{(t+1)} &= \frac{1}{n}\sum_i \left(y_{ij} - b_i^{(t)}\right) \\
\sigma^{2(t+1)} &= \frac{1}{nt}\left((y_{ij} - \mu_j^{(t+1)})^2 - 2(y_{ij} - \mu_j^{(t+1)})b_i^{(t)} + b_i^{2(t)}\right) \\
\psi^{(t+1)} &= \frac{1}{n}\sum_j \left(b_i^{2(t)}\right)
\end{aligned}
$$

Cálculo do Erro-padrão

- Sabemos que, sob certas condições de regularidade o emv de $\boldsymbol{\theta}$, $\widehat{\boldsymbol{\theta}}$ é tal que, para n suficientemente grande $\widehat{\boldsymbol{\theta}} \approx N_p\left(\boldsymbol{\theta}, \mathbf{I}\left(\boldsymbol{\theta}\right)^{-1}\right)$.
- O estimador obtido pelo EM, é um estimador de máxima verossimilhança. Portanto, sob as condições mencionadas, apresenta a propriedade acima.
- Como obter a Informação de Fisher (observada) no contexto do algoritmo EM.

Cont.

- Sejam $l(\boldsymbol{\theta}, \mathbf{y}) = l$ e $l(\boldsymbol{\theta}, \mathbf{y}, \mathbf{y}^*) = l^*$, respectivamente, a log-versossimilhança associada aos dados incompletos e aos dados completos.
- Sejam ainda

$$S(\boldsymbol{\theta}) = \frac{\partial^p}{\partial \boldsymbol{\theta}^p} l; S(\boldsymbol{\theta})^* = \frac{\partial^p}{\partial \boldsymbol{\theta}^p} l^*$$

$$H(\boldsymbol{\theta}) = \frac{\partial^{2p}}{\partial \boldsymbol{\theta}^p \partial \boldsymbol{\theta}^{p'}} l; H(\boldsymbol{\theta})^* = \frac{\partial^{2p}}{\partial \boldsymbol{\theta}^p \partial \boldsymbol{\theta}^{p'}} l^*$$

Cont.

- Pode-se provar que, a Informação de Fisher observada $I(\boldsymbol{\theta})^o$ é pode ser aproximada por (idendidade de Louis)

$$
\begin{aligned}
I(\boldsymbol{\theta})^o &= \mathcal{E}\left(-H(\boldsymbol{\theta})^*|\mathbf{y},\boldsymbol{\theta}\right) - \mathcal{E}\left(S(\boldsymbol{\theta})^* S(\boldsymbol{\theta})^{*'}|\mathbf{y},\boldsymbol{\theta}\right) \\
&+ S(\boldsymbol{\theta})S(\boldsymbol{\theta})'
\end{aligned}
$$

- Seja $\widetilde{\boldsymbol{\theta}}$ a estimativa obtida via algoritmo EM, então:

$$
I(\widetilde{\boldsymbol{\theta}})^o = \mathcal{E}\left(-H(\widetilde{\boldsymbol{\theta}})^*|\mathbf{y},\boldsymbol{\theta}\right) - \mathcal{E}\left(S(\widetilde{\boldsymbol{\theta}})^* S(\widetilde{\boldsymbol{\theta}})^{*'}|\mathbf{y},\boldsymbol{\theta}\right)
$$

Cont.

- Pois $S(\widetilde{\boldsymbol{\theta}}) = \mathbf{0}$.
- Como, também $S(\widetilde{\boldsymbol{\theta}})^* = \mathbf{0}$, temos que

$$I(\widetilde{\boldsymbol{\theta}})^o = \mathcal{E}\left(-H(\widetilde{\boldsymbol{\theta}})^*|\mathbf{y},\boldsymbol{\theta}\right) - Cov\left(S(\widetilde{\boldsymbol{\theta}})^*|\mathbf{y},\boldsymbol{\theta}\right)$$