ALGORITMOS PARA DADOS AUMENTADOS

1. INTRODUÇÃO

Dois algoritmos baseados na consideração de dados latentes. Temos os dados efetivamente observados, Y, e de uma maneira conveniente aumentamos esses dados, introduzindo os dados, Z, chamamos latentes ou não observados, de modo a facilitar o procedimento de cálculo da verossimilhança ou da densidade a posteriori.

O principio de "aumento dos dados" pode ser enunciado da seguinte forma: "aumentar" ao dados observados Y, com dados latentes Z de tal forma que P(θ/y,z) ou L(θ/y,z) sejam simples.

ALGORITMO EM: maximizar a verossimilhança dos dados observados utilizando-se os dados completos X={Y,Z} de maneira conveniente.

ALGORITMO DE DADOS AUMENTAODS: maximizar P(θ/y) utilizando P(θ/y,z).

EXEMPLO 1: suponha que se observe um processo periodicamente(por ex., peixes presos em armadilhas) que se acumule em intervalos regulares:

$x_1, x_2, \ldots, x_7$: quantidade nos dias 1, 2, ..., 7.

$y_1 = x_1 + x_2$

$y_2 = x_3$ $\rightarrow$ $X = (x_1, x_2, \ldots, x_7)$: dados completos (o que deveria ter sido observado)

$y_3 = x_4 + x_5$ $Y = (y_1, y_2, \ldots, y_4)$: dados incompletos (o que foi observado)

$y_4 = x_6 + x_7$

2. ALGORITMO EM

$X = (x_1, x_2, ..., x_n)'$ do modelo $f(x/\theta) \rightarrow L(\theta/x)$.

Suponha que não observamos X completamente, mas alguma função de X, digamos Y=h(X). Dizemos que

X: dados completos

Y=h(X): dados incompletos

A verossimilhança dos dados observados(incompletos) Y é dada por

$$L(\theta / Y) = \int_{\psi(y)} L(\theta / X) dX \qquad (1)$$

Onde $\psi(y)$: parte do espaço amostral ψ de X determinada pela equação y=h(x).

O algoritmo EM (E: esperança, M: maximização) é um procedimento iterativo segundo o qual encontramos o valor de θ que maximiza a verossimilhança dos dados observados, L(θ/Y), usando L(θ/X) de maneira conveniente. "conveniente" significa escolher L(θ/X) que fornece L(θ/Y), utilizando (1), de modo a tornar o problema fácil.

EXEMPLO 2: problema genético estudado.

197 animais distribuídos em 4 classes:
(y1, y2, y3, y4)´ segundo as prob. $\frac{\theta+2}{4}$, $\frac{1-\theta}{4}$, $\frac{1-\theta}{4}$, $\frac{\theta}{4}$,
$0 < \theta < 1$.

Y = (125, 18, 20, 34)´ = (y1, y2, y3, y4)´

A verossimilhança dos dados incompletos, é dada por

$$L(\theta / Y) = \frac{(y_1 + y_2 + y_3 + y_4)}{y_1! y_2! y_3! y_4!} \left(\frac{2+\theta}{4}\right)^{y_1} \left(\frac{1-\theta}{4}\right)^{y_2+y_3} \left(\frac{\theta}{4}\right)^{y_4}$$

Ou

$$L(\theta / Y) \propto \left(\frac{2+\theta}{4}\right)^{y_1} \left(\frac{1-\theta}{4}\right)^{y_2+y_3} \left(\frac{\theta}{4}\right)^{y_4}$$ → posteriori observada

Suponha que consideremos X=(x1, ..., x5) como sendo os dados aumentados (completos) onde y1 = x1+x2, y2=x3, y3=x4 e y4=x5, com probabilidades ($\frac{1}{2}, \frac{\theta}{4}, \frac{1-\theta}{4}, \frac{1-\theta}{4}, \frac{\theta}{4}$), de modo que

$$L(\theta / X) \propto \left(\frac{1-\theta}{4}\right)^{x_3+x_4} \left(\frac{\theta}{4}\right)^{x_2+x_5}$$ → posteriori aumentado

Podemos simplificar e utilizar

$$L(\theta / Y) \propto (2+\theta)^{y_1} (1-\theta)^{y_2+y_3} \theta^{y_4}$$

E

$$L(\theta / X) \propto \theta^{x_2+x_5} (1-\theta)^{x_3+x_4}$$ (mais simples)

A formula (1) fica

$$L(\theta / Y) = \sum_{x_1, x_2} L(\theta / x_1, x_2, 18.20.34)$$

Onde a soma é estendida a todos os pares (x1, x2) tais que x1+x2=125, xi≥0, i=1,2.

ALGORITMO EM:

1) tomemos um valor inicial, $\theta^{(0)}$, como estimador de θ.

2) Encontremos a esperança condicional de X, dado Y, (Y=(125, 18, 20, 34), y1=x1+x2), ou seja, estimamos os dados por suas esperanças condicionais, dado Y e $\theta^{(0)}$.

Passo E:

E[x_3/Y, $\theta^{(0)}$]=18

E[x_4/Y, $\theta^{(0)}$] = 20

E[x_5/Y, $\theta^{(0)}$] = 34

E[x_1/ Y, $\theta^{(0)}$] = E[x_1/x_1+x_2=125, $\theta^{(0)}$] = $x_1^{(0)}$

E[x_2/ Y, $\theta^{(0)}$] = E[x_2/x_1+x_2=125, $\theta^{(0)}$] = $x_2^{(0)}$

mas, x_1/x_1+x_2=125 ~ binomial, n=125,
$P_1 = \frac{1/2}{(2+\theta^{(0)})/4} = \frac{2}{2+\theta^{(0)}}$

x_2/x_1+x_2 = 125 ~ binomial, n=125, $P_2 = \frac{\theta^{(0)}/4}{(2+\theta^{(0)})/4} = \frac{\theta^{(0)}}{2+\theta^{(0)}}$

→ $x_1^{(0)}$ = np_1 = $\frac{250}{2+\theta^{(0)}}$ e $x_2^{(0)}$ = np_2 = $\frac{125\theta^{(0)}}{2+\theta^{(0)}}$

DADOS LATENTES → $X^{(0)} = (x_1^{(0)}, x_2^{(0)}, 18, 20, 34)'$: dados completos estimados

3. **O passo M**: consiste em maximizar a verossimilhança dos dados completos L(θ/$X^{(0)}$).

Temos

$$l(\theta / X^{(0)}) = x_1^{(0)} \log(1/2) + x_2^{(0)} \log(\theta/4) + x_3 \log(\frac{1-\theta}{4}) + x_4 \log(\frac{1-\theta}{4}) + x_5 \log(\theta/4)$$

$$l(\theta / X^{(0)}) \propto (x_2^{(0)} + x_5)\log(\theta) + (x_3 + x_4)\log(1-\theta)$$

Derivando em relação a θ, obtemos

$$\frac{dl}{d\theta} = \frac{x_2^{(0)} + x_5}{\theta} - \frac{x_3 + x_4}{1-\theta}$$

$$\Rightarrow \quad \hat{\theta}^{(1)} = \frac{x_2^{(0)} + x_5}{x_2^{(0)} + x_3 + x_4 + x_5} = \frac{x_2^{(0)} + 34}{x_2^{(0)} + 72}$$

De um modo geral, dada a estimativa na iteração (i), $\theta^{(i)}$, estimados os dados latentes por

$$x_1^{(i)} = \frac{250}{2+\theta^{(i)}}, \quad x_2^{(i)} = \frac{125\theta^{(i)}}{2+\theta^{(i)}}$$

E atualizamos o estimador de θ por

$$\theta^{(i+1)} = \frac{x_2^{(i)} + 34}{x_2^{(i)} + 72}$$

Se $\theta^{(0)} = 0{,}5$

Iteração(i)	$\theta^{(i)}$
0	0,5
1	0,60800
2	0,62400
3	0,62648
4	0,62677
5	0,62681
6	0,62682
7	0,62682

$$\Rightarrow \quad \hat{\theta} = 0{,}62682, \quad \hat{\pi}_1 = \frac{\hat{\theta}+2}{4} = 0{,}6567, \quad \hat{\pi}_3 = \hat{\pi}_4 = \frac{1-\hat{\theta}}{4} = 0{,}0933, \quad \hat{\pi}_4 = \frac{\hat{\theta}}{4} = 0{,}1567$$

3. ALGORITMO EM GERAL

Podemos utilizar o algoritmo EM para maximizar L(θ/Y) ou então a posteriori P(θ/Y). neste caso, temos a posteriori aumentada P(θ/Y,Z)= P(θ/X). neste contexto bayesiano, temos que considerar a densidade

$P(Z/Y, \theta^{(i)})$: distribuição condicional preditora dos dados latentes Z, condicional ao valor atual da moda ou aos dados observados.

No exemplo 2, esta distribuição é a binomial com parâmetros n = 125 e $p = \theta^{(i)}/(2+\theta^{(i)})$.

O algoritmo descrito está em termos da verossimilhança. Para o caso da densidade a posteriori, faz-se as modificações necessárias.

ALGORITMO EM:

1. **passo E**: calculamos

$$Q[\theta, \theta^{(i)}] = E[l(\theta/X)/Y, \theta^{(i)}]$$

Ou seja, a esperança condicional da log-verossimilhança aumentada, supondo dados Y e o valor atual $\theta^{(i)}$.

2. **Passo M**: escolhemos o valor $\theta^{(i+1)}$ no espaço paramétrico que maximiza $Q[\theta, \theta^{(i)}]$.

3. itere até convergência, ou seja, até que

$$|| \theta^{(i+1)} - \theta^{(i)} || \text{ ou } || Q[\theta, \theta^{(i_1)}] - Q[\theta, \theta^{(i)}] ||$$

sejam suficientemente pequenos.

No caso da posteriori consideramos

$$Q(\theta, \theta^{(i)}) = \int_Z \log p(\theta / Y, Z)\, p(Z / Y, \theta^{(i)}) dz$$

$\Updownarrow$

dist. Cond. Preditor dos dados latentes

EXEMOLO 3: VOLTAMOS AO EXEMPLO DO MODELO GENETICO

$$Q(\theta,\theta^{(i)}) = E[(x_2 + x_5)\log(\theta) + (x_3 + x_4)\log(1-\theta)/Y,\theta^{(i)}] =$$
$$= [E(x_2 / Y,\theta^{(i)}) + x_5]\log(\theta) + (x_3 + x_4)\log(1-\theta)$$

Sabemos que

$$E[x_2 / Y,\theta^{(i)}] = \frac{125\theta^{(i)}}{2+\theta^{(i)}}$$

Maximizando Q[θ, $\theta^{(i)}$], obtemos

$$\theta^{(i+1)} = \frac{E(x_2 / Y,\theta^{(i)}) + x_5}{E(x_2 / Y,\theta^{(i)}) + x_3 + x_4 + x_5}$$

PROPOSIÇÃO 1: seja $l(\theta)$ a verossimilhança dos dados observados. Então

$$l(\theta^{(i+1)} / Y) \geq l(\theta^{(i)} / Y)$$

Toda iteração do algoritmo EM aumenta a log-verossimilhança. O mesmo vale para a densidade a posteriori.

PROPOSIÇÃO 2: supunha que uma sequencia de iterações do algoritmo EM satisfaça

i) $\frac{\partial Q(\theta,\theta^{(i)})}{\partial\theta}|_{\theta=Q^{(i+1)}} = 0$

ii) $\theta^{(i)} \to \theta^{*}$

então, as iterações convergem para um ponto estacionário de L(θ/Y).

4. MONTE CARLO EM

No passo E do algoritmo EM, temos que calcular

$$Q(\theta,\theta^{(i)}) = \int_Z \log p(\theta / Y, Z)\, p(Z / Y, \theta^{(i)})\, dz$$

que pode ser complicado. Podemos utilizar MMC para facilitar esse passo. Wei and tanner (1990)

Algoritomo MCEM para o passo E:

1) simule uma amostra iid $z_1, z_2, \ldots, z_m$ de $P(Z/Y,\theta^{(i)})$
2) calcule $\hat{Q}_{i+1}(\theta) = \frac{1}{m}\sum_{j=1}^{m} \log P(\theta / z_j, Y)$
3) no passo M, $\hat{Q}$ é maximizado para obter $\theta^{(i+1)}$.

Um problema é especificar o valor de m. Uma solução é aumentar o valor de m à medida que o numero de iterações cresce e monitorar a convergência, por meio de um gráfico de $\theta^{(i)}$x i.

EXEMPLO 3: no problema genético, $(x_2/\theta^{(i)},Y)$~binomial com parâmetros n=125 e $p = \frac{\theta^{(i)}}{2+\theta^{(i)}}$. Geramos $z_1, z_2, \ldots, z_m$ dessa distribuição e fazemos

$$\hat{E}(x_2 / \theta^{(i)}, Y) = \frac{1}{m}\sum_{i=1}^{m} z_i = \bar{z}$$

O passo M continua como antes, obtendo-se

$$\theta^{(i+1)} = (\bar{z} + x_5)/((\bar{z} + x_3 + x_4 + x_5)$$

ALGORITMOS PARA DADOS AUMENTADOS

1. INTRODUÇÃO

Dois algoritmos baseados na consideração de dados latentes. Temos os dados efetivamente observados, Y, e de uma maneira conveniente aumentamos esses dados, introduzindo os dados, Z, chamamos latentes ou não observados, de modo a facilitar o procedimento de cálculo da verossimilhança ou da densidade a posteriori.

O principio de "aumento dos dados" pode ser enunciado da seguinte forma: "aumentar" ao dados observados Y, com dados latentes Z de tal forma que $P(\theta/y,z)$ ou $L(\theta/y,z)$ sejam simples.

ALGORITMO EM: maximizar a verossimilhança dos dados observados utilizando-se os dados completos X={Y,Z} de maneira conveniente.

ALGORITMO DE DADOS AUMENTAODS: maximizar $P(\theta/y)$ utilizando $P(\theta/y,z)$.

EXEMPLO 1: suponha que se observe um processo periodicamente(por ex., peixes presos em armadilhas) que se acumule em intervalos regulares:

$x_1, x_2, \ldots, x_7$: quantidade nos dias 1, 2, ..., 7.

$y_1 = x_1 + x_2$

$y_2 = x_3$ $\rightarrow$ $X = (x_1, x_2, \ldots, x_7)$: dados completos (o que deveria ter sido observado)

$y_3 = x_4 + x_5$ $Y = (y_1, y_2, \ldots, y_4)$: dados incompletos (o que foi observado)

$y_4 = x_6 + x_7$

2. ALGORITMO EM

X = (x_1, x_2, ..., x_n)´ do modelo f(x/θ) → L(θ/x).

Suponha que não observamos X completamente, mas alguma função de X, digamos Y=h(X). Dizemos que

X: dados completos

Y=h(X): dados incompletos

A verossimilhança dos dados observados(incompletos) Y é dada por

$$L(\theta / Y) = \int_{\psi(y)} L(\theta / X) dX \qquad (1)$$

Onde $\psi(y)$: parte do espaço amostral ψ de X determinada pela equação y=h(x).

O algoritmo EM (E: esperança, M: maximização) é um procedimento iterativo segundo o qual encontramos o valor de θ que maximiza a verossimilhança dos dados observados, L(θ/Y), usando L(θ/X) de maneira conveniente. "conveniente" significa escolher L(θ/X) que fornece L(θ/Y), utilizando (1), de modo a tornar o problema fácil.

EXEMPLO 2: problema genético estudado.

197 animais distribuídos em 4 classes:
(y1, y2, y3, y4)´ segundo as prob. $\frac{\theta+2}{4}$, $\frac{1-\theta}{4}$, $\frac{1-\theta}{4}$, $\frac{\theta}{4}$,
0 < θ < 1.

Y = (125, 18, 20, 34)´ = (y1, y2, y3, y4)´

A verossimilhança dos dados incompletos, é dada por

$$L(\theta / Y) = \frac{(y_1 + y_2 + y_3 + y_4)}{y_1! y_2! y_3! y_4!} \left(\frac{2+\theta}{4}\right)^{y_1} \left(\frac{1-\theta}{4}\right)^{y_2+y_3} \left(\frac{\theta}{4}\right)^{y_4}$$

Ou

$$L(\theta/Y) \propto \left(\frac{2+\theta}{4}\right)^{y_1} \left(\frac{1-\theta}{4}\right)^{y_2+y_3} \left(\frac{\theta}{4}\right)^{y_4}$$ → posteriori observada

Suponha que consideremos X=(x1, ..., x5) como sendo os dados aumentados (completos) onde y1 = x1+x2, y2=x3, y3=x4 e y4=x5, com probabilidades ($\frac{1}{2}, \frac{\theta}{4}, \frac{1-\theta}{4}, \frac{1-\theta}{4}, \frac{\theta}{4}$), de modo que

$$L(\theta/X) \propto \left(\frac{1-\theta}{4}\right)^{x_3+x_4} \left(\frac{\theta}{4}\right)^{x_2+x_5}$$ → posteriori aumentado

Podemos simplificar e utilizar

$$L(\theta/Y) \propto (2+\theta)^{y_1} (1-\theta)^{y_2+y_3} \theta^{y_4}$$

E

$$L(\theta/X) \propto \theta^{x_2+x_5} (1-\theta)^{x_3+x_4}$$ (mais simples)

A formula (1) fica

$$L(\theta/Y) = \sum_{x_1, x_2} L(\theta/x_1, x_2, 18.20.34)$$

Onde a soma é estendida a todos os pares (x1, x2) tais que x1+x2=125, xi≥0, i=1,2.

ALGORITMO EM:

1) tomemos um valor inicial, $\theta^{(0)}$, como estimador de θ.

2) Encontremos a esperança condicional de X, dado Y, (Y=(125, 18, 20, 34), y1=x1+x2), ou seja, estimamos os dados por suas esperanças condicionais, dado Y e $\theta^{(0)}$.

Passo E:

$E[x_3/Y, \theta^{(0)}]=18$

$E[x_4/Y, \theta^{(0)}] = 20$

$E[x_5/Y, \theta^{(0)}] = 34$

$E[x_1/ Y, \theta^{(0)}] = E[x_1/x_1+x_2=125, \theta^{(0)}] = x_1^{(0)}$

$E[x_2/ Y, \theta^{(0)}] = E[x_2/x_1+x_2=125, \theta^{(0)}] = x_2^{(0)}$

mas, $x_1/x_1+x_2=125$ ~ binomial, n=125,

$$P_1 = \frac{1/2}{(2+\theta^{(0)})/4} = \frac{2}{2+\theta^{(0)}}$$

$x_2/x_1+x_2 = 125$ ~ binomial, n=125, $P_2 = \frac{\theta^{(0)}/4}{(2+\theta^{(0)})/4} = \frac{\theta^{(0)}}{2+\theta^{(0)}}$

→ $x_1^{(0)} = np_1 = \frac{250}{2+\theta^{(0)}}$ e $x_2^{(0)} = np_2 = \frac{125\theta^{(0)}}{2+\theta^{(0)}}$

DADOS LATENTES → $X^{(0)} = (x_1^{(0)}, x_2^{(0)}, 18, 20, 34)'$: dados completos estimados

3. **O passo M**: consiste em maximizar a verossimilhança dos dados completos $L(\theta/X^{(0)})$.

Temos

$$l(\theta/X^{(0)}) = x_1^{(0)}\log(1/2) + x_2^{(0)}\log(\theta/4) + x_3\log(\frac{1-\theta}{4}) + x_4\log(\frac{1-\theta}{4}) + x_5\log(\theta/4)$$

$$l(\theta/X^{(0)}) \propto (x_2^{(0)} + x_5)\log(\theta) + (x_3 + x_4)\log(1-\theta)$$

Derivando em relação a θ, obtemos

$$\frac{dl}{d\theta} = \frac{x_2^{(0)} + x_5}{\theta} - \frac{x_3 + x_4}{1-\theta}$$

$$\Rightarrow \quad \hat{\theta}^{(1)} = \frac{x_2^{(0)} + x_5}{x_2^{(0)} + x_3 + x_4 + x_5} = \frac{x_2^{(0)} + 34}{x_2^{(0)} + 72}$$

De um modo geral, dada a estimativa na iteração (i), $\theta^{(i)}$, estimados os dados latentes por

$$x_1^{(i)} = \frac{250}{2+\theta^{(i)}}, \quad x_2^{(i)} = \frac{125\theta^{(i)}}{2+\theta^{(i)}}$$

E atualizamos o estimador de θ por

$$\theta^{(i+1)} = \frac{x_2^{(i)} + 34}{x_2^{(i)} + 72}$$

Se $\theta^{(0)} = 0,5$

Iteração(i)	$\theta^{(i)}$
0	0,5
1	0,60800
2	0,62400
3	0,62648
4	0,62677
5	0,62681
6	0,62682
7	0,62682

$$\Rightarrow \quad \hat{\theta} = 0{,}62682, \quad \hat{\pi}_1 = \frac{\hat{\theta}+2}{4} = 0{,}6567, \quad \hat{\pi}_3 = \hat{\pi}_4 = \frac{1-\hat{\theta}}{4} = 0{,}0933, \quad \hat{\pi}_4 = \frac{\hat{\theta}}{4} = 0{,}1567$$

3. ALGORITMO EM GERAL

Podemos utilizar o algoritmo EM para maximizar L(θ/Y) ou então a posteriori P(θ/Y). neste caso, temos a posteriori aumentada P(θ/Y,Z)= P(θ/X). neste contexto bayesiano, temos que considerar a densidade

$P(Z/Y, \theta^{(i)})$: distribuição condicional preditora dos dados latentes Z, condicional ao valor atual da moda ou aos dados observados.

No exemplo 2, esta distribuição é a binomial com parâmetros n = 125 e $p= \theta^{(i)}/(2+ \theta^{(i)})$.

O algoritmo descrito está em termos da verossimilhança. Para o caso da densidade a posteriori, faz-se as modificações necessárias.

ALGORITMO EM:

1. **passo E**: calculamos

$$Q[\theta, \theta^{(i)}] = E[l(\theta/X)/Y, \theta^{(i)}]$$

Ou seja, a esperança condicional da log-verossimilhança aumentada, supondo dados Y e o valor atual $\theta^{(i)}$.

2. **Passo M**: escolhemos o valor $\theta^{(i+1)}$no espaço paramétrico que maximiza $Q[\theta, \theta^{(i)}]$.

3. itere até convergência, ou seja, até que

$$|| \theta^{(i+1)} - \theta^{(i)}|| \text{ ou } || Q[\theta, \theta^{(i_1)}]- Q[\theta, \theta^{(i)}]||$$

sejam suficientemente pequenos.

No caso da posteriori consideramos

$$Q(\theta,\theta^{(i)}) = \int_Z \log p(\theta/Y,Z)\, p(Z/Y,\theta^{(i)})dz$$

$\Updownarrow$

dist. Cond. Preditor dos dados latentes

EXEMOLO 3: VOLTAMOS AO EXEMPLO DO MODELO GENETICO

$$Q(\theta,\theta^{(i)}) = E[(x_2 + x_5)\log(\theta) + (x_3 + x_4)\log(1-\theta)/Y,\theta^{(i)}] =$$
$$= [E(x_2/Y,\theta^{(i)}) + x_5]\log(\theta) + (x_3 + x_4)\log(1-\theta)$$

Sabemos que

$$E[x_2/Y,\theta^{(i)}] = \frac{125\theta^{(i)}}{2+\theta^{(i)}}$$

Maximizando Q[θ, $\theta^{(i)}$], obtemos

$$\theta^{(i+1)} = \frac{E(x_2/Y,\theta^{(i)}) + x_5}{E(x_2/Y,\theta^{(i)}) + x_3 + x_4 + x_5}$$

PROPOSIÇÃO 1: seja $l(\theta)$ a verossimilhança dos dados observados. Então

$$l(\theta^{(i+1)}/Y) \geq l(\theta^{(i)}/Y)$$

Toda iteração do algoritmo EM aumenta a log-verossimilhança. O mesmo vale para a densidade a posteriori.

PROPOSIÇÃO 2: supunha que uma sequencia de iterações do algoritmo EM satisfaça

i) $\frac{\partial Q(\theta,\theta^{(i)})}{\partial\theta}\Big|_{\theta=Q^{(i+1)}} = 0$

ii) $\theta^{(i)} \to \theta^{*}$

então, as iterações convergem para um ponto estacionário de L(θ/Y).

4. MONTE CARLO EM

No passo E do algoritmo EM, temos que calcular

$$Q(\theta,\theta^{(i)}) = \int_Z \log p(\theta / Y, Z) p(Z / Y, \theta^{(i)}) dz$$

que pode ser complicado. Podemos utilizar MMC para facilitar esse passo. Wei and tanner (1990)

Algoritomo MCEM para o passo E:

1) simule uma amostra iid z_1, z_2, ..., z_m de $P(Z/Y,\theta^{(i)})$
2) calcule $\hat{Q}_{i+1}(\theta) = \frac{1}{m}\sum_{j=1}^{m} \log P(\theta / z_j, Y)$
3) no passo M, $\hat{Q}$ é maximizado para obter $\theta^{(i+1)}$.

Um problema é especificar o valor de m. Uma solução é aumentar o valor de m à medida que o numero de iterações cresce e monitorar a convergência, por meio de um gráfico de $\theta^{(i)}$x i.

EXEMPLO 3: no problema genético, $(x_2/\theta^{(i)},Y)$~binomial com parâmetros n=125 e $p = \frac{\theta^{(i)}}{2+\theta^{(i)}}$. Geramos z_1, z_2, ..., z_m dessa distribuição e fazemos

$$\hat{E}(x_2 / \theta^{(i)}, Y) = \frac{1}{m}\sum_{i=1}^{m} z_i = \bar{z}$$

O passo M continua como antes, obtendo-se

$$\theta^{(i+1)} = (\bar{z} + x_5)/((\bar{z} + x_3 + x_4 + x_5)$$

5. CÁLCULO DOS ERROS PADRÕES

ALGORITMO EM → Calcula a moda da distribuição à posteriori: P(θ/Y) ou da fv L((θ/Y)

Para obter erros padrões devemos calcular a matriz Lessiana. Há 3 formas possíveis:

a) cálculo direto ou numérico

obtido o EMV $\hat{\theta}$, podemos calcular as derivadas segundas de log L((θ/Y) ou de log P((θ/Y), no ponto $\hat{\theta}$.

$$Var_{\theta_0}(\hat{\theta}) \quad \to \quad I(\theta_0)^{-1}$$

$I(\theta) = -E_\theta[\ddot{l}(\theta / Y)]$: informação de Fisher

Na prática, isto pode ser difícil e um enfoque alternativo é calcular as derivadas numericamente.

b) método de Louis

sabemos que

$$\log[P(\theta / Y)] = \log[p(\theta / Y, Z)] - \log[P(Z / \theta, Y)] + \log P(Z / Y)$$

$$\frac{-\partial^2 \log[p(\theta / Y)]}{\partial\theta_i \partial\theta_j} = \frac{-\partial^2 \log[P(\theta / Y, Z)]}{\partial\theta_i \theta_j} + \frac{\partial^2 \log[P(Z / Y, \theta)}{\partial\theta_i \partial\theta_j}$$

Integrando ambos os lados com respeito a P(Z/Y,θ), obtemos

$$\frac{-\partial^2 \log[p(\theta / Y)]}{\partial\theta_i \partial\theta_j} = \frac{-\partial^2}{\partial\theta_i \theta_j} Q(\theta, \theta^{(i)}) |_{\theta^{(i)}=\theta} + \frac{\partial^2}{\partial\theta_i \partial\theta_j} H(\theta, \theta^{(i)}) |_{\theta^{(i)}=\theta}$$

Esta identidade constitui o principio da informação faltante:

Informação observada =
= informação completa – informação faltante

RESULTADO: $\frac{\partial^2}{\partial\theta^2}H = \text{var}\left[\frac{\partial \log P(\theta/Y,Z)}{\partial\theta}\right]$: informação faltante

Utilizando esse resultado e a expressão acima, podemos obter o erro padrão de $\hat{\theta}$.

EXEMPLO 5: voltamos ao problema genético

$$P(\theta/Y,Z) \propto \theta^{x_2+x_5}(1-\theta)^{x_3+x_4}$$

$$\frac{\partial \log P(\theta/Y,Z)}{\partial\theta} = \frac{x_2+x_5}{\theta} - \frac{x_3+x_4}{1-\theta}$$

inf. completa

$$\Rightarrow \quad \frac{-\partial^2 Q}{\partial\theta^2}|_{\hat{\theta}} = \frac{E(x_2/Y,\hat{\theta})+x_5}{\hat{\theta}^2} + \frac{x_3+x_4}{1-\theta} = \frac{29{,}83+34}{(0{,}6268)^2} + \frac{18+20}{1-(0{,}6268)^2} = 435{,}3$$

informação faltante:

$\rightarrow$

$$\frac{\partial^2}{\partial\theta^2}H = \text{var}\left[\frac{\partial \log P(\theta/Y,Z)}{\partial\theta}\right] = \frac{\text{var}(X_2/\hat{\theta})}{\hat{\theta}^2} = \frac{np(1-p)}{\hat{\theta}^2} = \frac{125x\frac{\hat{\theta}}{2+\hat{\theta}}x\frac{2}{2+\hat{\theta}}}{\hat{\vartheta}^2} = 57{,}8$$

Informação observada:

$$\frac{-\partial^2 \log P(\theta/Y)}{\partial\theta^2} = 435{,}3 - 57{,}8 = 377{,}5$$

$$\Rightarrow \quad ep(\hat{\theta}) = \sqrt{\frac{1}{377{,}5}} = 0{,}05$$

c) simulação

o cálculo da informação completa pode ser complicado

$$-\int_z \frac{\partial^2 \log P(\theta/Y,Z)}{\partial\theta^2} P(Z/\hat{\theta}/Y)dZ$$: informação completa

Se pudemos amostrar da densidade P(Z/θ,Y), a integral pode ser aproximada por

$$\frac{1}{m}\sum_{j=1}^{m}\frac{\partial^2 \log P(\theta/Y,Z_j)}{\partial\theta^2}$$: informação completa simulada,

Z: valor simulado, z_1, z_2, ..., z_m ~iid $P(Z/\hat{\theta},Y)$

De modo análogo, podemos aproximar a informação faltante:

$$\frac{\partial^2}{\partial\theta^2}H = \mathrm{var}\left[\frac{\partial \log P(\theta/Y,Z)}{\partial\theta}\right]$$ por

$$\frac{1}{m}\sum_{j=1}^{m}\left(\frac{\partial \log P(\theta/Y,z_j)}{\partial\theta}\right)^2 - \left[\frac{1}{m}\sum_{i=1}^{m}\frac{\partial \log P(\theta/Y,z_j)}{\partial\theta}\right]^2$$: informação faltante simulada

No exemplo genético,

$$(Z/\hat{\theta},Y) \sim \quad binomial(125,\frac{\hat{\theta}}{2+\hat{\theta}}) \quad com\,\hat{\theta} \to 0{,}6268$$

6. ALGORITMO DE DADOS AUMENTADOS

O algoritmo EM utiliza a simplicidade da verossimilhança (ou densidade a posteriori) dos dados aumentados. No algoritmo de dados aumentados(ADA), em vez de obter o maxímo dessas funções, o objetivo é obter a verossimilhança ou distribuição a posteriori, a fim de obter outras informações como intervalos de confiança (ou credibilidade). Trabalharemos, aqui, com a distribuição a posteriori. A idéia é obter uma estimativa de P(θ/Y), baseada na distribuição aumentada P((θ/Y,Z).

O ADA é motivado por duas identidades:

$P(\theta / Y) = \int_z P(\theta / Y, Z) P(Z / Y) dz$: identidade da posteriori

↓

P(θ/Y): densidade da posteriori de θ

P(θ/Y,Z): posteriori aumentada

P(Z/Y): densidade preditiva dos dados latentes dado Y

$P(Z / Y) = \int_\theta P(z / \phi, Y) P(\phi / Y) d\phi$: identidade da preditora

P(Z/Φ,Y): preditora dos dados latentes

O ADA é um algoritmo iterativo entre essas duas identidades, aproximando sucessivamente a densidade a posteriori.

ALGORITMO ADA:

[1] simule z_1, z_2, ..., z_m da estimativa atual P(Z/Y)

[2] atualize a estimativa P(θ/Y) por meio de

$$\frac{1}{m}\sum_{j=1}^{m} P(\theta / Y, Z_j)$$

[3] itera.

No passo precisamos simular da preditora P(Z/Y). Pela identidade da preditora: $P(Z / Y) = \int_\theta P(z / \phi, Y) P(\phi / Y) d\phi$, ela é uma mistura de preditoras aumentadas em relação à posteriori observada, esse passo pode ser implementado por meio da iteração:

[1´] simule θ^* da estimativa atual P(θ/Y)

[2´] amostre z de $P(z/ \theta^*, Y)$

Por sua vez, o passo [1´], simulação da posteriori observada atual, é realizado selecionando-se j aleatoriamente dos inteiros 1, 2, ..., m e então simulando de $P(\theta/z_j, Y)$, dada a forma discreta da estimativa de $P(\theta/Y)$ em [2] (no passo [2] do ADA.

Comparando esse algoritmo com o EM, temos:

Substituímos os passos E e M pelos passos que chamamos S e I, onde,

S: estamos simulando dos dados latentes da estimativa atual da preditora.

I: depois integramos para obter a posteriori.

Podemos então chamar o ADA de algoritmo SI (S de simulação e I de integração).

EXEMPLO 6: continuando o exemplo do problema genético

Posteriori aumentada

$$P(\theta / X) \propto \theta^{x_2+x_5}(1-\theta)^{x_3+x_4} \quad \sim \quad Beta(x_2+x_5+1, x_3+x_4+1)$$

$(X_2/Y,\theta) \sim$ Binomial(125, $\theta/(\theta+2)$), onde $Z= X_2$

↓

preditora aumentada

ALGORITMO SI:

[1] simule $x_2^1, \cdots, x_2^m$ da estimativa atual de $P(X_2/Y)$

[2] atualize a estimativa de $P(\theta/Y)$ por meio de

$$\frac{1}{m}\sum_{j=1}^{m} Beta(\alpha_j, \beta_j)(\theta)$$

Com $\alpha_j = x_2^j + x_5 + 1 \quad e \quad \beta_j = x_3 + x_4 + 1$

O passo [1] do algoritmo consiste em repetir m vezes:

[1´] gere j uniformemente de 1, 2, ..., m

[2´] simule θ^* da Beta(α_j, β_j)

[3´] simule x da binomial(125, $\theta^*/(2+ \theta^*)$)