12.04.2016.

EM - algorithm

If complete, we could find the MLE using

$$\sum_{ij} \left(y_{ij} - \underbrace{(\mu + \alpha_i + \beta_j)}_{\gamma_{ij}} \right)^2$$

$$\Rightarrow \quad (X^T X)^{-1} X^T y$$

The design matrix would be

$$X = \begin{array}{c} \begin{array}{cccc} \mu & \alpha_1 & \beta_1 & \beta_2 \end{array} \\ \begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & -1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & -1 & 0 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & -1 \end{pmatrix} \end{array} \begin{array}{c} \\ y_{11} \\ y_{21} \\ y_{12} \\ y_{22} \\ y_{13} \\ y_{23} \end{array}$$

$$\begin{pmatrix} y_{11} \\ y_{21} \\ y_{12} \\ y_{22} \\ y_{13} \\ y_{23} \end{pmatrix} = \begin{pmatrix} \quad & \quad & \quad & \quad \end{pmatrix} \begin{pmatrix} \mu \\ \alpha_1 \\ \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{21} \\ \varepsilon_{12} \\ \varepsilon_{22} \\ \varepsilon_{13} \\ \varepsilon_{22} \end{pmatrix}$$

$$\underbrace{\begin{pmatrix} \frac{1}{6} & 0 & 0 & 0 \\ 0 & \frac{1}{6} & 0 & 0 \\ 0 & 0 & \frac{1}{3} & -\frac{1}{6} \\ 0 & 0 & -\frac{1}{6} & \frac{1}{3} \end{pmatrix}}_{(X^T X)^{-1}} \underbrace{\begin{pmatrix} \sum_{ij} y_{ij} \\ \sum_j y_{1j} - \sum_j y_{2j} \\ y_{11} + y_{21} - y_{13} - y_{23} \\ y_{12} + y_{22} - y_{13} - y_{23} \end{pmatrix}}_{X^T y} = \begin{pmatrix} \hat{\mu} \\ \hat{\alpha}_1 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix}$$

Our example:

$$y = (y_{11}, y_{12}, y_{13}, y_{21}, y_{22})$$

$$x = (y_{11}, y_{12}, y_{13}, y_{21}, y_{22}, y_{23})$$

$$\Theta = (\mu, \alpha_1, \alpha_2, \beta_1, \beta_2, \beta_3)$$

$$\gamma_{ij} = \mu + \alpha_i + \beta_j$$

log likelihood

$$\Rightarrow \quad l(\Theta; x) = \text{const} - \frac{1}{2\sigma^2} \sum_{i,j} (y_{ij} - \gamma_{ij})^2$$

E-step

$$Q(\Theta) = E\left(l(\Theta, x) \mid \vec{y}, \Theta^{(i)}\right)$$

$$= \text{const} - E\left[\frac{1}{2\sigma^2} \sum_{i,j} (y_{ij} - \gamma_{ij})^2 \mid \vec{y}, \Theta^{(i)}\right]$$

$$= \text{const} - \frac{1}{2\sigma^2}\left[E\left((y_{23} - \gamma_{23})^2 \mid \vec{y}, \Theta^{(i)}\right) + \sum_{(i,j) \in (2,3)} (y_{ij} - \gamma_{ij})^2\right]$$

$$E[(X-c)^2] = \underbrace{E[(X - EX)^2] + (EX - c)^2}$$

$$E(X^2 - 2X E(X) + (EX)^2) + (EX)^2 - 2c EX + c^2$$

$$= E(X^2) - 2\cancel{(EX)^2} + \cancel{(EX)^2} + \cancel{(EX)^2} - 2cEX + c^2$$

$$= E(X^2) - 2cEX + c^2 = E[(X-c)^2]$$

$$\Rightarrow \quad E(y_{23} \mid y, \Theta^{(i)}) = \gamma_{23}^{(i)} = \hat{Y}_{23}$$

$$\Rightarrow \quad E[(y_{23} - \gamma_{23})^2 \mid \vec{y}, \Theta^{(i)}] = \underbrace{E[(y_{23} - \hat{y}_{23})^2]}_{\text{indep. of } \gamma_{23}} + (\hat{y}_{23} - \gamma_{23})^2$$

$\Rightarrow$ indep of Θ $\Rightarrow$ not used in optimisation

$\Rightarrow$ $Q(\theta)$ is the usual log-likelihood with imputated data for y_{23} $\Rightarrow$ we can maximise it (M-step)

If required, the MLE for σ^2 is

$$\hat{\sigma}^2 = \frac{1}{5} \sum_{i,j} (y_{ij} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j)^2$$

↑ goes over all available data.

Rao example

Complete data $x = (x_1, x_2, x_3, x_4, x_5)$

$$y = h(x) = (x_1 + x_2, x_3, x_4, x_5)$$
$$= (y_1, y_2, y_3, y_4)$$

log-likelihood

$$l(\theta, x) = x_1 \log\left(\frac{1}{2}\right) + x_2 \log\left(\frac{\theta}{4}\right) + x_3 \log\left(\frac{1-\theta}{4}\right) +$$
$$x_4 \log\left(\frac{1-\theta}{4}\right) + x_5 \log\left(\frac{\theta}{4}\right) + \text{const.}$$

$$= (x_2 + x_5)\log(\theta) + (x_3 + x_4)\log(1-\theta) + \text{const.}$$

$$\Rightarrow l'(\theta, x) = \frac{x_2 + x_5}{\theta} + \frac{x_3 + x_4}{1-\theta} \cdot (-1) \overset{!}{=} 0$$

$$(x_2 + x_5)(1-\theta) = (x_3 + x_4) \cdot \theta$$

$$x_2 + x_5 = \theta(x_3 + x_4 + x_2 + x_5)$$

$$\Rightarrow \hat{\theta} = \frac{x_2 + x_5}{x_2 + x_3 + x_4 + x_5}$$

E-step

$$Q(\theta) = E[l(\theta; x) \mid \vec{y}, \theta^{(i)}]$$

$$= (\underbrace{E(x_2 \mid \vec{y}, \theta^{(i)})}_{\text{only thing we need to compute}} + x_5)\log\theta + (x_3 + x_4)\log(1-\theta)$$

$$y_1 = x_1 + x_2 \Rightarrow x_2 \mid y_1, \theta^{(i)} \sim \text{Bin}\left(y_1, \underbrace{\frac{\frac{\theta^{(i)}}{4}}{\frac{1}{2} + \frac{\theta^{(i)}}{4}}}_{\frac{\theta^{(i)}}{2+\theta^{(i)}}}\right)$$

$$= \Big(\underbrace{y_1 \cdot \frac{\theta^{(i)}}{2+\theta^{(i)}}}_{\hat{x}_2} + x_5 \Big) \log(\theta) + (x_3 + x_4) \log(1-\theta)$$

M-step: Maximise $Q(\theta)$ for θ

$$\Rightarrow \quad \hat{\theta} = \frac{\hat{x}_2 + x_5}{\hat{x}_2 + x_3 + x_4 + x_5}$$

iterate between those steps until convergence

Illustration of a hidden Markov model (HMM)