EM Algorithm Details

Xiang Liu
Educational Testing Service

Derivative identities

We first establish some identities for derivatives that will be used repeatedly. The probability of observing a tool use is given by the following response function,

$$P(W_{ij} = 1|\theta_{i2}) = \frac{1}{1 + \exp(-\alpha_j(\theta_{i2} - \beta_j))} = P_{ij}. \tag{1}$$

The probability of not observing a tool use is then

$$P(W_{ij} = 0|\theta_{i2}) = 1 - P_{ij} = Q_{ij}. \tag{2}$$

The first-order partial derivatives are

$$\frac{\partial P_{ij}}{\partial \alpha_j} = \frac{\theta_{i2} \exp(-\alpha_j(\theta_{i2} - \beta_j))}{(1 + \exp(-\alpha_j(\theta_{i2} - \beta_j)))^2} = (\theta_{i2} - \beta_j)P_{ij}Q_{ij}, \tag{3}$$

and

$$\frac{\partial P_{ij}}{\partial \beta_j} = -\frac{\alpha_j \exp(-\alpha_j(\theta_{i2} - \beta_j))}{(1 + \exp(-\alpha_j(\theta_{i2} - \beta_j)))^2} = -\alpha_j P_{ij}Q_{ij}. \tag{4}$$

The probability of getting a correct response conditional on $W_{ij} = w_{ij}$ is

$$P(X_{ij} = 1|\theta_{i1}, w_{ij}) = \frac{1}{1 + \exp(-a_j(\theta_{i1} - b_j + w_{ij}\gamma_j))} = \dot{P}_{ij}. \tag{5}$$

Then the probability of getting an incorrect response is

$$P(X_{ij} = 0|\theta_{i1}, w_{ij}) = 1 - \dot{P}_{ij} = \dot{Q}_{ij}. \tag{6}$$

Taking first-order partial derivatives, it leads to

$$\frac{\partial \dot{P}_{ij}}{\partial a_j} = (\theta_{i1} - b_j + w_{ij}\gamma_j)\dot{P}_{ij}\dot{Q}_{ij}, \tag{7}$$

$$\frac{\partial \dot{P}_{ij}}{\partial b_j} = -a_j\dot{P}_{ij}\dot{Q}_{ij}, \tag{8}$$

and

$$\frac{\partial \dot{P}_{ij}}{\partial \gamma_j} = a_j w_{ij}\dot{P}_{ij}\dot{Q}_{ij}. \tag{9}$$

Likelihood Equations

Given $\boldsymbol{X} = \boldsymbol{x}$, $\boldsymbol{W} = \boldsymbol{w}$, and $\boldsymbol{\Theta} = \boldsymbol{\theta}$ the complete data likelihood of the model parameters are

$$L = \prod_{i=1}^{N} \phi(\boldsymbol{\theta}_i; \rho) \prod_{j=1}^{J} P_{ij}^{w_{ij}} Q_{ij}^{1-w_{ij}} \dot{P_{ij}}^{x_{ij}} \dot{Q_{ij}}^{1-x_{ij}}. \tag{10}$$

Equivalently, the compete data log-likelihood is

$$\log L = \sum_{j=1}^{J} l_j(\alpha_j, \beta_j) + \sum_{j=1}^{J} l_j(a_j, b_j, \gamma_j) + \sum_{i=1}^{N} \log \phi(\boldsymbol{\theta}_i; \rho), \tag{11}$$

where

$$l_j(\alpha_j, \beta_j) = \sum_{i=1}^{N} w_{ij} \log P_{ij} + (1 - w_{ij}) \log Q_{ij}, \tag{12}$$

and

$$l_j(a_j, b_j, \gamma_j) = \sum_{i=1}^{N} x_{ij} \log \dot{P_{ij}} + (1 - x_{ij}) \log \dot{Q_{ij}}. \tag{13}$$

Since the three sets of item parameters, $\{\alpha_j, \beta_j\}$, $\{a_j, b_j, \gamma_j\}$, and $\{\rho\}$, are separated in different sums in the complete data log-likelihood, we take gradient separately, i.e.

$$\nabla l_j(\alpha_j, \beta_j) = \begin{pmatrix} \frac{\partial l_j}{\partial \alpha_j} \\ \\ \frac{\partial l_j}{\partial \beta_j} \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{N} (w_{ij} - P_{ij})(\theta_{i2} - \beta_j) \\ \\ -\sum_{i=1}^{N} (w_{ij} - P_{ij}) \alpha_j \end{pmatrix}, \tag{14}$$

and

$$\nabla l_j(a_j, b_j, \gamma_j) = \begin{pmatrix} \frac{\partial l_j}{\partial a_j} \\ \\ \frac{\partial l_j}{\partial b_j} \\ \\ \frac{\partial l_j}{\partial \gamma_j} \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{N} (x_{ij} - P_{ij})(\theta_{i1} - b_j + w_{ij}\gamma_j) \\ \\ -\sum_{i=1}^{N} (x_{ij} - P_{ij}) a_j \\ \\ \sum_{i=1}^{N} (x_{ij} - P_{ij}) a_j w_{ij} \end{pmatrix}. \tag{15}$$

Taking the first order partial derivative w.r.t ρ leads to a third order polynomial in ρ,

$$\frac{\partial l(\rho)}{\partial \rho} = N\rho^3 - \left(\sum_{i=1}^{N} \theta_{i1}\theta_{i2} \right) \rho^2 - \left(N - \sum_{i=1}^{N} (\theta_{i1}^2 + \theta_{i2}^2) \right) \rho - \sum_{i=1}^{N} \theta_{i1}\theta_{i2} \tag{16}$$

Jacobian Matrix

In finding the maximum likelihood estimates, we need to solve the likelihood equations,

$$E_{\boldsymbol{\theta}|\ldots} \left[\nabla l_j(\alpha_j, \beta_j) \right] = \mathbf{0}, \tag{17}$$

$$E_{\boldsymbol{\theta}|\ldots} \left[\nabla l_j(a_j, b_j, \gamma_j) \right] = \mathbf{0}, \tag{18}$$

, and

$$E_{\boldsymbol{\theta}|\ldots}\left[\frac{\partial l(\rho)}{\partial \rho}\right] = 0. \tag{19}$$

These nonlinear systems of equations can be solved numerically by the derivative based Newton-Raphson method. The Jacobian matrix of the gradients are

$$\boldsymbol{J}_{\nabla l_j(\alpha_j,\beta_j)} = \begin{pmatrix} \frac{\partial^2 l_j}{\partial \alpha_j^2} & \frac{\partial^2 l_j}{\partial \alpha_j \partial \beta_j} \\ \frac{\partial^2 l_j}{\partial \beta_j \partial \alpha_j} & \frac{\partial^2 l_j}{\partial \beta_j^2} \end{pmatrix} = \begin{pmatrix} -\sum_i (\theta_{i2} - \beta_j)^2 P_{ij} Q_{ij} & -\sum_i (w_{ij} - P_{ij}) - (\theta_{i2} - \beta_j)\alpha_j P_{ij} Q_{ij} \\ -\sum_i (w_{ij} - P_{ij}) - (\theta_{i2} - \beta_j)\alpha_j P_{ij} Q_{ij} & -\sum_i \alpha_j^2 P_{ij} Q_{ij} \end{pmatrix} \tag{20}$$

and

$$\boldsymbol{J}_{\nabla l_j(a_j,b_j,\gamma_j)} = \begin{pmatrix} \frac{\partial^2 l_j}{\partial a_j^2} & \frac{\partial^2 l_j}{\partial a_j \partial b_j} & \frac{\partial^2 l_j}{\partial a_j \partial \gamma_j} \\ \frac{\partial^2 l_j}{\partial b_j \partial a_j} & \frac{\partial^2 l_j}{\partial b_j^2} & \frac{\partial^2 l_j}{\partial b_j \partial \gamma_j} \\ \frac{\partial^2 l_j}{\partial \gamma_j \partial a_j} & \frac{\partial^2 l_j}{\partial \gamma_j \partial b_j} & \frac{\partial^2 l_j}{\partial \gamma_j^2} \end{pmatrix} = \begin{pmatrix} -\sum_i (\theta_{i1} - b_j + w_{ij}\gamma_j)^2 \dot{P}_{ij}\dot{Q}_{ij} & \sum_i (\theta_{i1} - b_j + w_{ij}\gamma_j) a_j \dot{P}_{ij}\dot{Q}_{ij} - (x_{ij} - \dot{P}_{ij}) & -\sum_i (\theta_{i1} - b_j + w_{ij}\gamma_j) a_j w_{ij} \dot{P}_{ij}\dot{Q}_{ij} - w_{ij}(x_{ij} - \dot{P}_{ij}) \\ \sum_i (\theta_{i1} - b_j + w_{ij}\gamma_j) a_j \dot{P}_{ij}\dot{Q}_{ij} - (x_{ij} - \dot{P}_{ij}) & -\sum_i a_j^2 \dot{P}_{ij}\dot{Q}_{ij} & \sum_i a_j^2 w_{ij} \dot{P}_{ij}\dot{Q}_{ij} \\ -\sum_i (\theta_{i1} - b_j + w_{ij}\gamma_j) a_j w_{ij} \dot{P}_{ij}\dot{Q}_{ij} - w_{ij}(x_{ij} - \dot{P}_{ij}) & \sum_i a_j^2 w_{ij} \dot{P}_{ij}\dot{Q}_{ij} & -\sum_i a_j^2 w_{ij}^2 \dot{P}_{ij}\dot{Q}_{ij} \end{pmatrix} \tag{21}$$