Technical Appendix: `MultiScale` EM Algorithm

February 21, 2018

The Parametric Model

N voters and J binary questions to vote on. The vote matrix is $Y \in \{\pm 1\}^{N \times J}$. Note the use of -1 instead of 0 for nea, which is non-standard. For some $d \in \mathbb{N}$, Let $\alpha_j \in \mathbb{R}$, $\beta_j \in \mathbb{R}^d$ and $\gamma_i \in \mathbb{R}^d$ for each $j = 1, \ldots, J$ and $i = 1, \ldots, N$. We assume the following latent variable model generates the sign matrix Y.

$$
\begin{aligned}
y_{ij} &= \text{sign}(z_{ij}) \\
z_{ij} &= \alpha_j + \beta_j^\top \gamma_i + \epsilon_{ij}, \quad \epsilon_{ij} \overset{\text{ind.}}{\sim} \mathcal{N}(0, 1)
\end{aligned}
$$

Note that for $\theta = (\{\alpha_j\}_{j=1}^J, \{\beta_j\}_{j=1}^J, \{\gamma_i\}_{i=1}^N)$ this implies the reduced form likelihood

$$
p_\theta(Y) = \prod_{i=1}^N \prod_{j=1}^J \Phi\left(y_{ij}(\alpha_j + \beta_j^\top \gamma_i)\right).
$$

Let $O \in \{0, 1\}^{N \times J}$ denote the matrix of observation statuses. That is $o_{ij} = 1$ if the (i, j)th cell of Y is observed and $o_{ij} = 0$ if it is missing. Then we make the standard assumption that missingness is at random such that

$$
p_\theta(Y_{\text{obs}} \mid O) = \prod_{i=1}^N \prod_{j=1}^J \left[\Phi\left(y_{ij}(\alpha_j + \beta_j^\top \gamma_i)\right)\right]^{o_{ij}}.
$$

We assume standard priors on θ; specifically,

$$
\xi(\theta) = \prod_{j=1}^J \mathcal{N}\left(\begin{bmatrix}\alpha_j \\ \beta_j\end{bmatrix}; \mu_{ab}, \Sigma_{ab}\right) \prod_{i=1}^N \mathcal{N}\left(\gamma_i; \mu_\gamma, \Sigma_\gamma\right),
$$

where $\mu_{ab} \in \mathbb{R}^{d+1}$, $\mu_\gamma \in \mathbb{R}^d$ and $\Sigma_{ab} \in \mathbb{R}^{(d+1)\times(d+1)}$, $\Sigma_\gamma \in \mathbb{R}^{d \times d}$ are positive definite matrices.

EM Algorithm

We consider just the log likelihood to illustrate how we extend the EM algorithm of Imai, Lo and Olmsted (2016). Let $m_{ij} = \alpha_j + \beta_j^\top \gamma_i$. The complete data (except for the values for which $o_{ij} = 0$) log likelihood is given by

$$
\begin{aligned}
&\log p(Y_{\text{obs}}, Z_{\text{obs}} \mid O) \\
&= \prod_{i=1}^{N} \prod_{j=1}^{J} \Big[\mathcal{N}(z_{ij} \mid m_{ij}, 1)\Big]^{\mathbb{I}[y_{ij}=\text{sign}(z_{ij})]\, o_{ij}} \\
&= \sum_{i=1}^{N} \sum_{j=1}^{J} o_{ij} \mathbb{I}(y_{ij} = \text{sign}(z_{ij})] \log \mathcal{N}(z_{ij}; m_{ij}, 1) \\
&= \sum_{i=1}^{N} \sum_{j=1}^{J} o_{ij} \Big[\mathbb{I}(y_{ij} = +1)\mathbb{I}(z_{ij} \geq 0) + \mathbb{I}(y_{ij} = -1)\mathbb{I}(z_{ij} < 0)\Big] \log \mathcal{N}(z_{ij}; m_{ij}, 1).
\end{aligned}
$$

But this is the same log likelihood found in Imai, Lo and Olmsted (2016, Appendix A), except for the insistence on only using the observed data as observations. Therefore we can take their update equations and restrict ourselves to only using observed data. Specifically,

$$
z_{ij} \leftarrow m_{ij} + y_{ij} \frac{\phi(y_{ij})}{\Phi\left(y_{ij} m_{ij}\right)}
$$

$$
\gamma_i \leftarrow \left(\Sigma_\gamma^{-1} + \sum_{j=1}^{J} o_{ij} \beta_j \beta_j^\top\right)^{-1} \left(\Sigma_\gamma^{-1} \mu_\gamma + \sum_{j=1}^{J} o_{ij} \beta_j (z_{ij} - \alpha_j)\right)
$$

$$
\begin{bmatrix} \alpha_j \\ \beta_j \end{bmatrix} \leftarrow \left(\Sigma_{ab}^{-1} + \sum_{i=1}^{N} o_{ij} \begin{bmatrix} 1 \\ \gamma_i \end{bmatrix} \begin{bmatrix} 1 \\ \gamma_i \end{bmatrix}^\top\right)^{-1} \left(\Sigma_{ab}^{-1} \mu_{ab} + \sum_{i=1}^{N} o_{ij} z_{ij} \begin{bmatrix} 1 \\ \gamma_i \end{bmatrix}\right).
$$

References

Imai, Kosuke, James Lo and Jonathan Olmsted. 2016. "Fast Estimation of Ideal Points with Massive Data." *American Political Science Review* 110(4):1–20.