EM algorithm for Composite Likelihood

with application to two-way data array

Huy Dang

November 27, 2018

Motivation and Definition

Motivation: High dimensional response variables make likelihood inferences difficult by rendering the computation of likelihoods infeasible.
Thus, a class of likelihoods, called *Composite likelihoods /Pseudo-likelihoods* is often used in place of the full likelihood.
Definition: (Varin et. al., 2011) Consider a vector of random variable Y from the density $f(y;\theta)$ for some unknown p-dim parameter $\theta \in \Theta$. Let $(\mathcal{A}_1, \cdots, \mathcal{A}_K)$ be a set of marginal or conditional events with associated likelihoods $\mathcal{L}_k(\theta; y) \propto f(y \in \mathcal{A}_k; \theta)$.
Composite likelihood is defined as the weighted product:

$$\mathcal{L}_C(\theta; y) = \prod_{k=1}^{K} \mathcal{L}_k(\theta; y)^{w_k}$$

Examples

Examples:

- Composite conditional likelihood: pairwise cond. densities

$$\mathcal{L}_C(\theta; y) = \prod_{r=1}^{m} \prod_{s=1}^{m} f(y_r | y_s; \theta)$$

- Composite marginal likelihood:

$$\mathcal{L}_C(\theta; y) = \prod_{r=1}^{m} f(y_r | \theta)$$

Properties: There are results on asymptotic properties, efficiency, robustness of composite likelihood based estimators. But they vary case by case, and are somewhat limited.

Problem Statement

My simplified version: 2-way data array. U and V are i.i.d row and column discrete latent variables.

	V_1	V_2	$\cdots$	V_s
U_1	Y_{11}	Y_{12}	$\cdots$	Y_{1s}
U_2	Y_{21}	Y_{22}	$\cdots$	Y_{2s}
$\vdots$	$\vdots$		$\vdots$	
U_r	Y_{r1}	Y_{r2}	$\cdots$	Y_{rs}

$$\lambda_u = P(U_i = u), u = 1, \cdots, k_1$$
$$\rho_v = P(V_j = v), v = 1. \cdots, k_2$$
$$Y_{ij} | U_i = u, V_j = v \sim N(\psi_{uv}, \sigma^2)$$

	1	2	$\cdots$	k_2
1	ψ_{11}	ψ_{12}	$\cdots$	ψ_{1k_2}
2	ψ_{21}	ψ_{22}	$\cdots$	ψ_{2k_2}
$\vdots$	$\vdots$		$\vdots$	
k_1	$\psi_{k_1 1}$	$\psi_{k_1 2}$	$\cdots$	$\psi_{k_1 k_2}$

In reality: Problems can be made more complicated by allowing V to be generated from a Markov chain with k_2 states. It accommodates certain types of data: genomics, economics, etc.

Full and Composite Likelihood

Let $\boldsymbol{y}_i^{(r)}$ be the ith row of data, and $\boldsymbol{y}_j^{(c)}$ be the jth column.
The full likelihood is:

$$L(\theta; \boldsymbol{Y}) = p(\boldsymbol{Y}) = \sum_{\boldsymbol{u}} p(\boldsymbol{Y}|\boldsymbol{u})p(\boldsymbol{u})$$

where $p(\boldsymbol{Y}|\boldsymbol{u})$ is computed using a well-known recursion in HM literature (Baum et. al. 1970, Welch 2003).
Row Composite Likelihood: assuming that the rows are independent.

$$L_C(\theta; \boldsymbol{Y}) = \prod_i (\boldsymbol{y}_i^{(r)}) = \prod_i \sum_u \lambda_u p(\boldsymbol{y}_i^{(r)}|U_i = u)$$

where $p(\boldsymbol{y}_i^{(r)}|U_i = u)$ is computed using a well-known recursion in HM literature (Baum et. al. 1970, Welch 2003).
Flops(full) $= O(k_1^r k_2 s)$, Flops(Composite) $= \mathrm{O}(\mathrm{k}_1 r k_2 s)$

EM Algorithm for Full Likelihood

$$L^*(\theta; \boldsymbol{Y}, \boldsymbol{U}, \boldsymbol{V}) = P(\boldsymbol{U} = \boldsymbol{u}) \cdot P(\boldsymbol{V} = \boldsymbol{v}) \cdot \prod_{i=1}^{r} \prod_{j=1}^{s} N(y_{ij}; \psi_{u_i v_j}, \sigma^2)$$

$$= \left(\prod_{i=1}^{r} \prod_{u=1}^{k_1} \lambda_u^{w_{iu}} \right) \left(\prod_{j=1}^{s} \prod_{v=1}^{k_2} \rho_v^{z_{jv}} \right) \left(\prod_{i=1}^{r} \prod_{j=1}^{s} \prod_{u=1}^{k_1} \prod_{v=1}^{k_2} N(y_{ij}; \psi_{uv}, \sigma^2) \right)^{w_{iu} z_{jv}}$$

where $w_{iu} = I(U_i = u)$; $z_{jv} = I(V_i = v)$

$$l^*(\theta; \boldsymbol{Y}, \boldsymbol{U}, \boldsymbol{V}) = \sum_{i=1}^{r} \sum_{u=1}^{k_1} w_{iu} log(\lambda_u) + \sum_{j=1}^{s} \sum_{v=1}^{k_2} z_{jv} log(\rho_v)$$

$$+ \sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{u=1}^{k_1} \sum_{v=1}^{k_2} w_{iu} z_{jv} log(N(y_{ij}; \psi_{uv}, \sigma^2))$$

The conditional expectation involves terms such as:
$E_{\theta^{(n-1)}}(w_{iu} | \boldsymbol{Y}) = P(U_i = u | \boldsymbol{Y}; \theta^{(n-1)}) = \frac{1}{p(\boldsymbol{Y})} \sum_{\boldsymbol{u}: u_i = u} p(\boldsymbol{Y} | \boldsymbol{u}) p(\boldsymbol{u})$

EM for Full (and Composite) Likelihood

For Composite Likelihood: Z_{ijv} in place of z_{jv}.

$$l_C^*(\theta; \boldsymbol{Y}_i^{(r)}, \boldsymbol{U}, \boldsymbol{V}) = \sum_{i=1}^{r}\sum_{u=1}^{k_1} w_{iu} log(\lambda_u) + \sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{v=1}^{k_2} z_{ijv} log(\rho_v)$$

$$+ \sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{u=1}^{k_1}\sum_{v=1}^{k_2} w_{iu} z_{jv} log(N(y_{ij}; \psi_{uv}, \sigma^2))$$

$$E_{\theta^{(n-1)}}(w_{iu}|\boldsymbol{Y}) = P(U_i = u|\boldsymbol{Y}_i^{(r)}) = \frac{1}{p(\boldsymbol{Y}_i^{(r)})} p(\boldsymbol{Y}_i^{(r)}|U_i = u)p(u)$$

Updates:

$$\lambda_u = \frac{1}{r}\sum_i \hat{w_{iu}}; \quad \rho_v = \frac{1}{s}\sum_j \hat{z_{jv}};$$

$$\mu_{uv} = \frac{(\widehat{w_{iu} z_{jv}}) y_{ij}}{\sum_i \sum_j \widehat{w_{iu} z_{jv}}}; \quad \sigma^2 = \frac{1}{rs}\sum_i\sum_j\sum_u\sum_v (\widehat{w_{iu} z_{jv}})(y_{ij} - \mu_{uv})^2$$

Simulation

$$k_1 = 2, k_2 = 2, \rho = (0.39, 0.61), \lambda = (0.4, 0.6), \Psi = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \sigma^2 = 0.5$$

r = 10, s = 15

EM w/ Full Likelihood:

$\hat{\rho} = (0.47, 0.53), \hat{\lambda} = (0.5, 0.5)$

$\hat{\Psi} = \begin{bmatrix} 0.9845 & 1.9034 \\ 3.0959 & 3.9929 \end{bmatrix}, \hat{\sigma}^2 = 0.2039$

computation time:$4596.17s(76mins)$

EM w/ Composite Likelihood:

$\hat{\rho} = (0.47, 0.53), \hat{\lambda} = (0.5, 0.5)$

$\hat{\Psi} = \begin{bmatrix} 0.9845 & 1.8702 \\ 3.0959 & 3.97 \end{bmatrix}, \hat{\sigma}^2 = 0.1970$

computation time:$1.27s$

r = 50, s = 100

EM w/ Full Likelihood: doesn't run

EM w/ Composite Likelihood:

$\hat{\rho} = (0.52, 0.48), \hat{\lambda} = (0.48, 0.52)$

$\hat{\Psi} = \begin{bmatrix} 0.9807 & 1.9852 \\ 3.0093 & 4.0182 \end{bmatrix}, \hat{\sigma}^2 = 0.25$

computation time:$551.58s$

Some comments

- When does it work? When does it misbehave?
- Starting value
- Possible next steps