R Computations with T. Louis EM-Information Formula In a Contingency-Table Example from Ex. I in HW2

Eric Slud, STAT 818M 2/24/2019

We consider here the implementation of formula (2.2) in Theorem 2.7 for the whole-sample information matrix computed as part of the EM algorithm, applied to the second of two 'Observed-data' formulations (the one that was not part of the assigned problem) in the handout for HW2 Problem I.

Recall that in this example, the underlying complete data were $\mathbf{X}_{com} = \{X_{ij}, \;\; i = 1, 2, \; j = 1, \ldots, 3\}$, with $N = \sum_{i=1}^{2} \sum_{j=1}^{3} X_{ij}$ fixed and known, and unknown parameters

$$\pi_{11} \;=\; \alpha \pi_1, \;\; \pi_{ij} \;=\; \gamma \, p_j \qquad \text{for} \quad (i,j) \neq (1,1)$$

subject to the two constraints

$$p_1 + p_2 + p_3 = 1, \qquad \alpha p_1 + \gamma(p_1 + 2p_2 + 2p_3) = 2\gamma + (\alpha - \gamma) p_1 = 1 \tag{1}$$

leaving the 3-dimensional unknown parameter to be estimated, as $\theta = (\alpha, p_1, p_2)$. Then $p_3 = 1 - p_1 - p_2$ and $\gamma = (1 - p_1 \alpha)/(2 - p_1)$, and the associated log-likelihood is

$$\log L_{com}(\theta) \;=\; X_{11} \log(\alpha) + \sum_{j=1}^{2} X_{+j} \log p_j \,+\, X_{+3} \, \log(1 - p_1 - p_2) \,+\, (N - X_{11}) \log(\frac{1 - \alpha p_1}{2 - p_1})$$

Maximizing this exprssion (previously done more conveniently using the constrained parameters $(\alpha, \gamma, p_1, p_2, p_3)$ via Lagrange multipliers) yields the complete-data MLEs $\hat{\theta}$ as

$$\hat{\alpha}\hat{p}_1 \;=\; \frac{X_{11}}{N} \quad , \quad \hat{p}_1 \;=\; \frac{2\, X_{21}}{N - X_{11} + X_{21}} \quad , \quad \hat{p}_2 \;=\; \frac{X_{+2}}{N - X_{11} + X_{21}} \tag{2}$$

The observed-data setting considered here (not the one assigned in HW2 Problem I), was

X_{11}	X_{12}	X_{13}
$X_{21} + X_{22}$		X_{23}

For these data, the observed -data log-likelihood is

$$\log L_{obs,2}(\theta) \;=\; X_{11} \log(\alpha p_1) \;+\; (X_{21} + X_{22}) \log(p_1 + p_2) \;+\; X_{12} \log(p_2)$$
$$+ \, X_{13} \log(p_3) \;+\; X_{23} \log(p_3) \;+\; (N - X_{11}) \log \gamma$$

Now according to formula (2.2) in Kim and Shao, there are three matrices to calculate toward the (whole-sample version rather than the per-observation version) observed-data information matrix $I_{obs,2}(\hat{\theta}^{EM})$, where $\hat{\theta}^{EM}$ is the same as the observed-data MLE, calculated using EM-algorithm iterative steps. The three matrices are

$$E_{\theta_0}(\, I_{com}(\theta)\,|\,\mathbf{Y}_{obs})\Big|_{\theta=\theta_0=\hat{\theta}^{EM}} \;,\quad S_{obs}(\theta)^{\otimes 2}\Big|_{\theta=\theta_0=\hat{\theta}^{EM}} \;,\quad E_{\theta_0}(\, S_{com}(\theta)^{\otimes 2}\,|\,\mathbf{Y}_{obs})\Big|_{\theta=\theta_0=\hat{\theta}^{EM}} \tag{3}$$

where

$$I_{com}(\theta) \;=\; \nabla_\theta^{\otimes 2}\log L_{com}(\theta) \;\;,\quad S_{com}(\theta) \;=\; \nabla_\theta \log L_{com}(\theta) \;\;,\quad S_{obs}(\theta) \;=\; E_\theta(S_{com}(\theta)\,|\,\mathbf{Y}_{obs})$$

In comparing these expressions recall that general formulas (the *Bartlett identity*) show that the expected values of the first and third matrices of (3) at θ with respect to the model P_θ must be equal, but that does not hold true for the conditional expected matrices in (3).

We provide concrete formulas for all of these matrices in our example. Begin by writing the score

$$S_{com}(\theta) \;=\; \begin{pmatrix} X_{11}/\alpha - (N-X_{11})p_1/(1-\alpha p_1) \\ X_{+1}/p_1 - X_{+3}/p_3 + (N-X_{11})(1-2\alpha)/((2-p_1)(1-\alpha p_1)) \\ X_{+2}/p_2 \;-\; X_{+3}/p_3 \end{pmatrix}$$

Recall that conditionally given the observed data $\mathbf{Y}_{obs}$,

$$(X_{21}+X_{22}) - X_{22} \;=\; X_{21} \;\sim\; \text{Binom}(X_{21}+X_{22},\, \frac{p_1}{p_1+p_2})$$

Therefore

$$S_{obs}(\theta) \;=\; \begin{pmatrix} X_{11}/\alpha \;-\; (N-X_{11})p_1/(1-\alpha p_1) \\ \frac{X_{11}}{p_1} \;+\; \frac{X_{21}+X_{22}}{p_1+p_2} \;-\; \frac{X_{+3}}{p_3} \;+\; \frac{(N-X_{11})(1-2\alpha)}{(2-p_1)(1-\alpha p_1)} \\ X_{12}/p_2 \;+\; (X_{21} \;+\; X_{22})/(p_1+p_2) \;-\; X_{+3}/p_3 \end{pmatrix} \tag{4}$$

and similarly

$$E_\theta(\, I_{com}(\theta)\,|\,\mathbf{Y}_{obs}) \;=\; \begin{pmatrix} A_{11} & A_{12} & 0 \\ A_{12} & A_{22} & A_{23} \\ 0 & A_{23} & A_{33} \end{pmatrix} \tag{5}$$

where

$$A_{11} \;=\; \frac{X_{11}}{\alpha^2} \;+\; \frac{(N-X_{11})\,p_1^2}{(1-\alpha p_1)^2} \;\;,\quad A_{12} \;=\; \frac{N-X_{11}}{(1-\alpha p_1)^2} \;\;,\quad A_{33} \;=\; \frac{X_{12}}{p_2^2} \;+\; \frac{X_{21}+X_{22}}{p_2\,(p_1+p_2)} \;+\; \frac{X_{+3}}{p_3^2}$$

$$A_{22} \;=\; \frac{X_{11}}{p_1^2} \;+\; \frac{X_{21} \;+\; X_{22}}{p_1\,(p_1+p_2)} \;+\; \frac{X_{+3}}{p_3^2} \;+\; (N - X_{11})\Big(\frac{\alpha^2}{(1-\alpha p_1)^2} \;-\; \frac{1}{(2-p_1)^2}\Big), \;\; A_{23} \;=\; \frac{X_{+3}}{p_3^2}$$

Up to this point, formulas (5) and (4) respectively provide us with the first and second matrices on the right-hand side of equation (3), while the third is obtained from the conditional expectation of $S_{com}(\theta)^{\otimes 2}$ as

$$E_\theta(\, S_{com}(\theta)^{\otimes 2} \,|\, \mathbf{Y}_{obs}) \;=\; S_{obs}(\theta)^{\otimes 2} \;+\; \begin{pmatrix} 0 & 0 & 0 \\ 0 & B_{22} & B_{23} \\ 0 & B_{23} & B_{33} \end{pmatrix} \tag{6}$$

where

$$B_{22} \;=\; \mathrm{Var}_\theta\Big(\frac{X_{21}}{p_1}\,\Big|\,\mathbf{Y}_{obs}\Big) \;=\; (X_{21}+X_{22})\,\frac{p_2}{p_1\,(p_1+p_2)^2}$$

and similarly

$$B_{33} \;=\; (X_{21}+X_{22})\,\frac{p_1}{p_2\,(p_1+p_2)^2} \quad , \quad B_{23} \;=\; -\,\frac{X_{21}+X_{22}}{(p_1+p_2)^2}$$

So from (3) or from formula (2.23) of Kim and Shao, the final formula for the observed information at θ in this example is

$$I_{obs}(\theta) \;=\; \begin{pmatrix} A_{11} & A_{12} & 0 \\ A_{12} & A_{22}-B_{22} & A_{23}-B_{23} \\ 0 & A_{23}-B_{23} & A_{33}-B_{33} \end{pmatrix} \tag{7}$$

This formula, with $\hat\theta \;=\; \hat\theta^{EM}$ substituted for θ, is the one implemented in the R script `RscriptsEM1.RLog`.