Parameter estimation with the incomplete-data Fisher scoring method

Keiji Takai
Faculty of Business and Commerce, Kansai University

The EM algorithm (EM for short) has been the standard method for parameter estimation of a model with incomplete data. The major problems of EM, however, which are often pointed out, are (1) its slow convergence, and (2) unavailability of the standard error. Another problem, which is often not pointed out but important, is (3) that EM has to resort to numerical methods such as the Newton-Raphson method and its variants when the Q-function of EM cannot be explicitly solved. For these three problems, various methods have been proposed (e.g., McLachlan and Krishnan, 2008).

In this talk, I propose another iterative method to overcome these problems, which I call the "Incomplete-data Fisher scoring" method (IFS for short). The IFS method has properties such that (1) it converges faster than EM, (2) it obtains the standard error only using the program developed for the faster convergence, and (3) it takes the form of the Newton-Raphson type methods (thus can make a smooth transition to them when necessary). The iterative equation is written as

$$\boldsymbol{\alpha}^{(t+1)} = \boldsymbol{\alpha}^{(t)} + q\frac{1}{n} J_{\mathrm{com}}(\boldsymbol{\alpha}^{(t)})^{-1} \nabla \ell_{\mathrm{obs}}(\boldsymbol{\alpha}^{(t)}),$$

where $\boldsymbol{\alpha}^{(t)}$ is the tth estimate of parameter $\boldsymbol{\alpha}$, q is the steplength, n is the sample size $J_{\mathrm{com}}(\boldsymbol{\alpha}^{(t)})$ is the complete-data information matrix, $\nabla \ell_{\mathrm{obs}}(\boldsymbol{\alpha}^{(t)})$ is the first derivative of the observed log-likelihood function $\ell_{\mathrm{obs}}(\boldsymbol{\alpha})$ w.r.t. $\boldsymbol{\alpha}$, which is evaluated at $\boldsymbol{\alpha} = \boldsymbol{\alpha}^{(t)}$.

This IFS is closely related to the EM. The EM gradient algorithm (Lange, 1995) takes the same form as IFS when the distribution is an exponential family with the natural parameter. However, IFS is not limited to a exponential family distribution. An approximation to EM also takes the same form as IFS (Meilijson, 1989). However, IFS with $q = 1$ and EM take exactly the same form of the iterative equation for a normal distribution. In the mixture of two Poisson distributions, the iterative equation of EM becomes close to that of IFS. It follows from these examples that EM and IFS behave similarly.

Despite IFS's close resemblance with EM, it can be derived without the idea of EM. Application of the Fisher scoring method to $\ell_{\mathrm{obs}}(\boldsymbol{\alpha})$ results in $E[\nabla^2 \ell_{\mathrm{obs}}(\boldsymbol{\alpha}^{(t)})]$ which is actually difficult to compute. Instead, with the Lower-bound algorithm (Böhning and Lindsay, 1988), the expectation can be replaced with $-n^{-1} J_{\mathrm{com}}(\boldsymbol{\alpha}^{(t)})$ which is a lot easier to compute. Therefore, we have the IFS iterative equation.

In my talk, I am going to describe additional issues: an acceleration method by steplength adjustment, a convergence theorem, a convergence rate, and three comparisons with (PX-)EM (Liu, Rubin and Wu, 1998).