

パラメータ分割による不完全データフィッシャースコアリング

関西大学 高井 啓二

人間に関わる研究を行なう社会科学や医学では、欠測データを始めとした不完全データを避けることができない。不完全データがあると、当初計画したデータサイズを確保できないだけでなく、想定しているパラメータの推定にも困難が生じてしまう。それは、各ケースに対するパラメータのサイズが欠測データパターンに応じて変化し、データを構成するケース全体が共通のパラメータを持たないからである。このような状況でパラメータ (特に最尤推定値;MLE と略) を推定するために、Dempster, Laird and Rubin(1977) は EM アルゴリズム (以降 EM と略) を提唱した。その提唱以来、EM は不完全データから MLE を推定するための標準的手法となった。

とは言え、EM にはいくつかの欠点があることも指摘されてきた (McLachlan and Krishnan, 2008)。その指摘のうち代表的なものを三つ挙げると、(i) 収束が遅いこと、(ii) MLE の標準誤差を推定できないこと、(iii) EM による更新式が明示的に得られず結局は非線形方程式を解かざるを得ないこと、である。これらの欠点を克服する方法として、発表者は昨年の本学会においてフィッシャースコアリングにもとづくパラメータの推定法 (Incomplete-data Fisher scoring; IFS) を提唱した (高井, 2017)。その方法の更新式は、

$$\boldsymbol{\theta}_{\text{new}} = \boldsymbol{\theta}_{\text{old}} + s \frac{1}{n} J_{\text{com}}(\boldsymbol{\theta}_{\text{old}})^{-1} \nabla \ell_{\text{obs}}(\boldsymbol{\theta}_{\text{old}})$$

である。ただし、 n はサンプルサイズ、 $\boldsymbol{\theta}$ はパラメータであり、 $J_{\text{com}}(\boldsymbol{\theta})$ は完全データの情報行列であり、 $\ell_{\text{obs}}(\boldsymbol{\theta})$ は完全データの尤度関数、 s はステップ幅である。例えば、正規分布の場合には、パラメータを平均 $\boldsymbol{\mu}$ と分散 Σ とすると、

$$\boldsymbol{\mu}_{\text{new}} = \boldsymbol{\mu}_{\text{old}} + s \frac{1}{n} \Sigma_{\text{old}}^{-1} \frac{\partial \ell_{\text{obs}}(\boldsymbol{\theta}_{\text{old}})}{\partial \boldsymbol{\mu}}, \quad \Sigma_{\text{new}} = \Sigma_{\text{old}} + s \frac{2}{n} (\Sigma_{\text{old}} \otimes \Sigma_{\text{old}})^{-1} \frac{\partial \ell_{\text{obs}}(\boldsymbol{\theta}_{\text{old}})}{\partial \text{vec}(\Sigma)},$$

となる。ステップ幅 $s = 1$ のとき、この更新式は EM の更新式と正確に一致する。従って、ステップ幅 s を適切に決めることにより、EM よりも早く収束させることができる。ステップ幅の計算に工夫をすることにより MLE の分散を得ることができる。

従来の IFS では、ステップ幅は全パラメータ共通として計算していた。しかし、それでは因子分析モデルなどのパラメータが多い場合には、ステップ幅の更新量が少なくなってしまうことがある。上の正規分布の例であるように、しばしば完全データの情報行列はブロック対角となる。そこで、本発表ではブロック対角の性質を用いた場合の IFS を導入し、その性質を調べる。

参考文献 Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *JRSS. Series B (Methodological)* **39**, 1–38. / McLachlan, G. & Krishnan, T. (2008). *The EM Algorithm and Extensions*. Wiley-Interscience, 2nd ed.