

多変量歪正規分布における 単純なアルゴリズムによるパラメータ推定

南山大学 阿部 俊弘
統計数理研究所 藤澤 洋徳

Azzalini & Capitanio (1999) では様々な手法で多変量歪正規分布を導いている．まずは次の確率変数を用意する：

$$\begin{pmatrix} X_0 \\ \mathbf{X} \end{pmatrix} \sim N_{p+1}(\mathbf{0}, \Omega^*), \quad \Omega^* = \begin{pmatrix} 1 & \delta^T \Omega^{1/2} \\ \Omega^{1/2} \delta & \Omega \end{pmatrix}$$

とする．ここで、 $X_0 \in \mathbb{R}$, $\mathbf{X} \in \mathbb{R}^p$, $\delta^T \delta < 1$, $\Omega \in \mathbb{R}^p \times \mathbb{R}^p$ は共分散行列である．このとき、

$$\mathbf{Z} = \begin{cases} \mathbf{X} & X_0 > 0, \\ -\mathbf{X} & \text{otherwise} \end{cases}$$

とすることにより、密度関数

$$f(\mathbf{z}) = 2\Phi(\boldsymbol{\alpha}^T \mathbf{z})\phi(\mathbf{z}; \mathbf{0}, \Omega), \quad \boldsymbol{\alpha} = \frac{\delta}{(1 - \delta^T \delta)^{1/2}},$$

が得られる．これは、 p 次元歪正規分布と呼ばれ、 $SN_p(\Omega, \boldsymbol{\alpha})$ と表すことにする．

ここで最尤推定を考えよう．確率変数 $\mathbf{Z}$ は、確率変数 X_0 を通した潜在変数によって表現されている．そのため、それを利用して、EM アルゴリズムを作ることが考えられる．ところが、その表現を利用しても、陽な表現は得られない．たとえば、歪正規分布の最尤推定を行う R パッケージ `sn` においては、オプションは一般の最適化関数 `optim` とほぼ同じで、EM アルゴリズムは組み込まれていない．

本研究では、冗長なパラメータ表現を利用して、陽な形で得られる EM アルゴリズムを導出する．上記の表現を参考にして、 $\tau > 0$ を導入し、

$$\Sigma = \begin{pmatrix} \tau^2 & \tau \delta^T \Omega^{1/2} \\ \tau \Omega^{1/2} \delta & \Omega \end{pmatrix}$$

とし、先ほどと同様に、 $\mathbf{Z}$ を構成すると、同じ p 次元歪正規分布 $SN_p(\Omega, \boldsymbol{\alpha})$ が得られる．結果としての歪正規分布に τ は現れていないため、冗長なパラメータ τ をあえて導入してることになっている．パラメータ τ の導入によって、分散共分散行列 Σ が制約のない正定値行列になるため、後で EM アルゴリズムを構成するのが楽になり、陽な表現が得られる．詳細は当日に報告する．

参考文献

- [1] Azzalini, A. & Capitanio, A. (1999). Statistical applications of the multivariate skew normal distribution. *J. Roy. Statist. Soc. Ser. B*, **61**, 579–602.