2022 年 6 月 17 日 統計数理研究所 オープンハウス

歪正規分布における加速EMアルゴリズム

藤澤 洋徳　数理・推論研究系 教授

はじめに

歪正規分布において最尤推定を得るためのEMアルゴリズムについて考える．通常の潜在変数表現をそのまま利用すると，EMアルゴリズムのMステップにおいて，適当な方程式の解を得る必要があり効率的ではない．Chen+ (2014) は上手いアイデアを利用して陽な形で得られるEMアルゴリズムを導出した．本研究では，通常の潜在変数表現に対して，あえて過剰パラメータを与えることで，陽な形で得られるEMアルゴリズムを導出した．このEMアルゴリズムは，Chen+ (2014) のEMアルゴリズムを加速するような形式になっており，数値例でもその加速を確認できた．本研究は，阿部（法政大）・川島（東工大）・Ley (Ghent U.) との共同研究である．

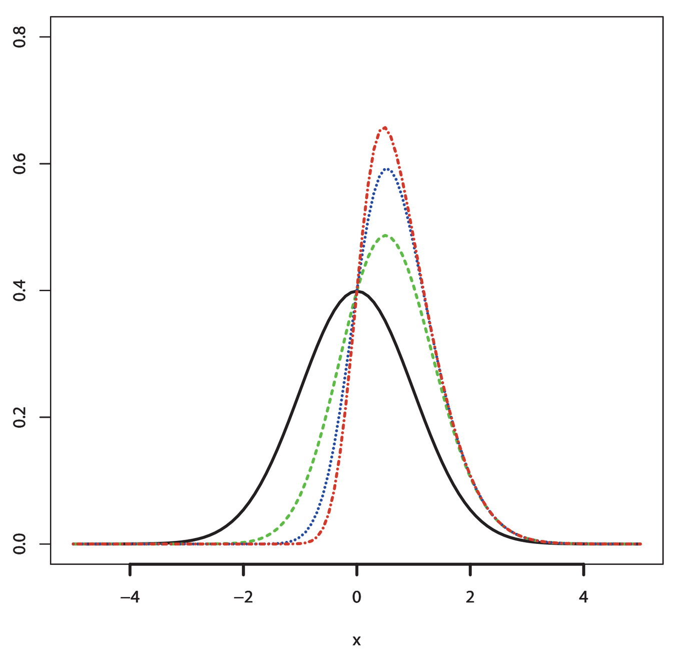
歪正規分布

$\boldsymbol{X} \sim \mathrm{SN}(\boldsymbol{\mu}, \Omega, \boldsymbol{\delta})$
(Skew Normal Distribution)

$$f(x) = 2\Phi(\boldsymbol{\alpha}^\top(x-\boldsymbol{\mu}))\phi(x; \boldsymbol{\mu}, \Omega),$$

$$\boldsymbol{\alpha} = \Omega^{-1}\boldsymbol{\delta}/(1-\boldsymbol{\delta}^\top\Omega^{-1}\boldsymbol{\delta})^{1/2}$$

$\Phi(z)$: cdf of $N(0,1)$
$\phi(z; \boldsymbol{\mu}, \Omega)$: pdf of $N(\boldsymbol{\mu}, \Omega)$



潜在変数表現

Azzalini & Capitanio (1999)

$$\begin{pmatrix} \boldsymbol{Y} \\ Y_0 \end{pmatrix} \sim N_{p+1}(\boldsymbol{0}, \Omega^*), \quad \Omega^* = \begin{pmatrix} \Omega & \boldsymbol{\delta} \\ \boldsymbol{\delta}^\top & 1 \end{pmatrix}.$$

そのとき$\boldsymbol{U} = \mathrm{sgn}(Y_0)\boldsymbol{Y} \sim \mathrm{SN}(\boldsymbol{0}, \Omega, \boldsymbol{\delta})$.

過剰パラメータ表現

$$\begin{pmatrix} \boldsymbol{Y} \\ Y_0 \end{pmatrix} \sim N_{p+1}(\boldsymbol{0}, \Sigma), \quad \Sigma = \begin{pmatrix} \Omega & \tau\Omega^{1/2}\boldsymbol{\delta} \\ \tau\boldsymbol{\delta}^\top\Omega^{1/2} & \tau^2 \end{pmatrix},$$

そのとき$\boldsymbol{X} = \mathrm{sgn}(Y_0)\boldsymbol{Y} \sim \mathrm{SN}(\boldsymbol{0}, \Omega, \Omega^{1/2}\boldsymbol{\delta}) = \mathrm{SN}^*(\boldsymbol{0}, \Omega, \boldsymbol{\delta})$.

Remark: Σがフリーパラメータ．
$\boldsymbol{X}$の分布はτに依存していない．

EM algorithm

データ：$x_1, \ldots, x_n \sim \mathrm{SN}^*(\boldsymbol{\mu}, \Omega, \boldsymbol{\delta})$

記号　$\boldsymbol{\xi} = (\boldsymbol{\mu}, \Sigma)$

$$S(x_i; \boldsymbol{\mu}, \boldsymbol{\xi}') = \mathrm{E}\left[\begin{pmatrix} x_i - \boldsymbol{\mu} \\ |y_{0i}| \end{pmatrix}\begin{pmatrix} x_i - \boldsymbol{\mu} \\ |y_{0i}| \end{pmatrix}^\top \middle| x_i; \boldsymbol{\xi}'\right]$$

条件付期待値は陽に表現可能．

EM algorithm　（τが現れない形で表現可能）

$$\hat{\boldsymbol{\mu}}^{(k+1)} = \bar{x} - \hat{\Sigma}_{12}^{(k)}\left\{\hat{\Sigma}_{22}^{(k)}\right\}^{-1}\frac{1}{n}\sum_{i=1}^n \mathrm{E}\left[|y_{0i}| \middle| x_i; \hat{\boldsymbol{\xi}}^{(k)}\right]$$

$$\hat{\Sigma}^{(k+1)} = \frac{1}{n}\sum_{i=1}^n S(x_i; \hat{\boldsymbol{\mu}}^{(k+1)}, \hat{\boldsymbol{\xi}}^{(k)})$$

過去のアルゴリズムとの違い

記号

$$\boldsymbol{\lambda} = \boldsymbol{\delta}/(1-\|\boldsymbol{\delta}\|^2) \qquad \hat{\boldsymbol{\psi}}_n^{(k)}(\boldsymbol{\mu}): \text{ omitted}$$

Chen+
(The update algorithm except for $\boldsymbol{\lambda}$ is very similar to our EM algorithm.)

$$\hat{\boldsymbol{\lambda}}^{(k+1)} = \{\hat{\Omega}^{(k)}\}^{-1}\hat{\boldsymbol{\psi}}_n^{(k)}(\hat{\boldsymbol{\mu}}^{(k)})$$

提案法

$$\hat{\boldsymbol{\lambda}}^{(k+1)} = a_n{}^{(k)}\{\hat{\Omega}^{(k+1)}\}^{-1}\hat{\boldsymbol{\psi}}_n^{(k)}(\hat{\boldsymbol{\mu}}^{(k+1)})$$

$$a_n{}^{(k)} = 1\Big/\sqrt{1 + \{\hat{\boldsymbol{\lambda}}^{(k)}\}^\top \boldsymbol{b}_1^{(k)} - \{\boldsymbol{b}_2^{(k+1)}\}^\top\{\Omega^{(k+1)}\}^{-1}\boldsymbol{b}_2^{(k+1)}}$$

$$\boldsymbol{b}_1^{(k)} = \hat{\boldsymbol{\psi}}_n^{(k)}(\hat{\boldsymbol{\mu}}^{(k)}), \quad \boldsymbol{b}_2^{(k+1)} = \hat{\boldsymbol{\psi}}_n^{(k)}(\hat{\boldsymbol{\mu}}^{(k+1)})$$

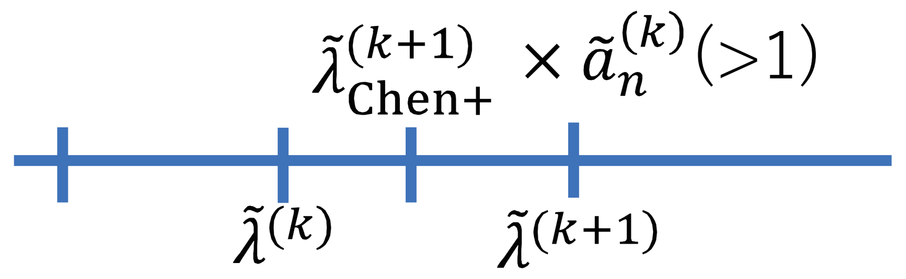
加速性：　a_nの効果

仮定：$p = 1$. $\boldsymbol{\xi}^{(k+1)} \approx \boldsymbol{\xi}^{(k)}$.　記号：$\tilde{o} = \hat{\Omega}^{1/2}\hat{o}$

$$\tilde{\lambda}_{Chen+}^{(k+1)} = \{\hat{\Omega}^{(k)}\}^{-1/2}\hat{\boldsymbol{\psi}}_n^{(k)}(\hat{\boldsymbol{\mu}}^{(k)}) \qquad \tilde{\lambda}^{(k+1)} \approx \tilde{a}_n^{(k)}\tilde{\lambda}_{Chen+}^{(k+1)}$$

$$\tilde{a}_n^{(k)} = 1\Big/\sqrt{1 - \tilde{\lambda}_{Chen+}^{(k+1)}\left(\tilde{\lambda}_{Chen+}^{(k+1)} - \tilde{\lambda}^{(k)}\right)}$$

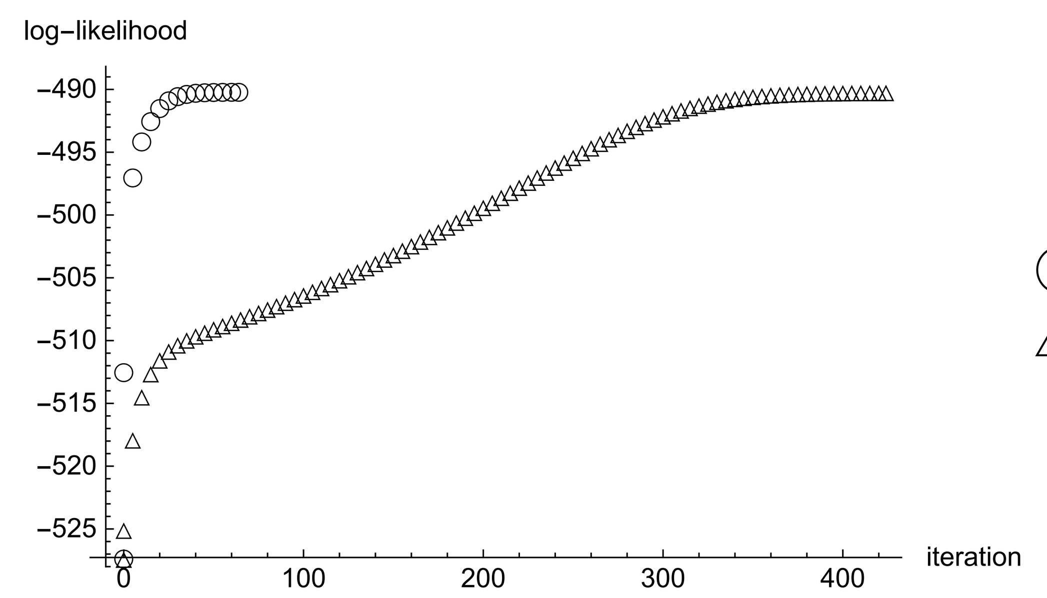
加速性の説明　（縮退の場合もある）



提案手法$\tilde{\lambda}^{(k+1)}$は$\tilde{\lambda}_{Chen+}^{(k+1)}$を$\tilde{a}_n^{(k)}$倍して（加速して）いる．

数値実験

Australian Institute of Sport (AIS) data with $n = 202$



○：提案手法
△：Chen+

最近の大学院生の研究テーマ

現学生
社会人A：状態空間モデル＋様々なタイプのデータ処理．
社会人B：HSIC Lasso．
社会人C：ロバスト推定．非漸近理論．収束レート．
修了生
フルタイムa：因果探索．因果推論．ロバスト推定．
社会人a：スパースモデリング（高相関・高欠測・転移学習に関連して）
フルタイムb：ロバスト性とスパース性を併せもつ回帰モデリング
フルタイムc：多重代入法

大学共同利用機関法人 情報・システム研究機構
統計数理研究所
The Institute of Statistical Mathematics