MMアルゴリズムによる行列式点過程の学習

川島貴大（総合研究大学院大学 統計科学専攻 博士課程5年）

背景：行列式点過程モデリング

行列式点過程はある全体集合のもとその部分集合の生起確率を記述する確率モデルである．要素間の共起関係や“斥力”を柔軟に表現できるため，推薦システムや文書要約などに応用されている．

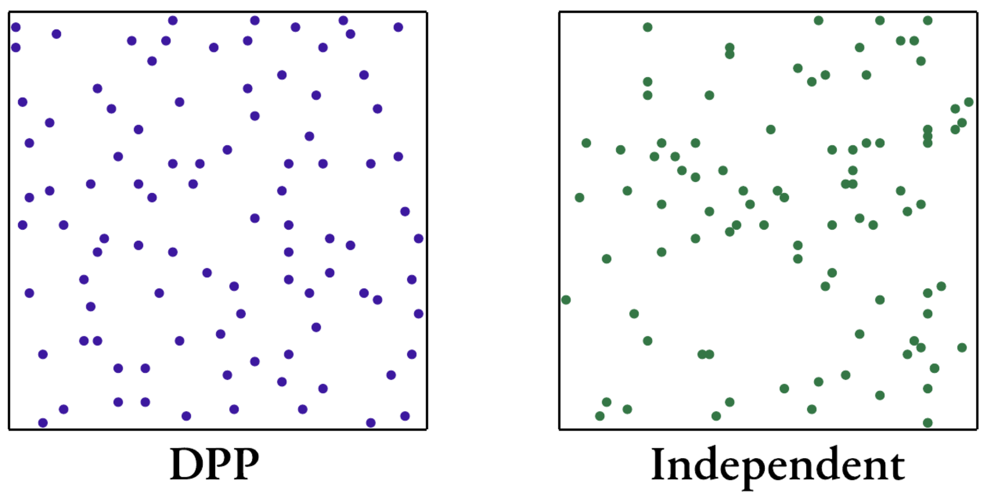


Figure 1. 行列式点過程からのサンプリングの様子（Kulesza+, 2012 より引用）

定義：行列式点過程の尤度関数

$$P(\mathcal{A}|\boldsymbol{L}) = \frac{\det([\boldsymbol{L}]_{\mathcal{A}})}{\det(\boldsymbol{L}+\boldsymbol{I})}$$

- $\mathcal{Y} = \{1, 2, \ldots, N\}$：全体集合
- $\mathcal{A} \subseteq \mathcal{Y}$：$\mathcal{Y}$のランダム部分集合
- $\boldsymbol{L} \in \mathbb{S}_{+}^{N}$：半正定値パラメータ行列
- $[\boldsymbol{L}]_{\mathcal{A}} = (L_{ij})_{i,j\in\mathcal{A}}$：パラメータ行列の主小行列 ($|\mathcal{A}| \times |\mathcal{A}|$)

提案法：MMアルゴリズムと代理関数

MM (Minorization-Maximization) アルゴリズムは目的関数$f(\theta)$の直接的な最大化が難しいとき，代理関数の逐次最適化によって局所解を目指すメタアルゴリズムのひとつである：

1. 実行可能領域内の任意の$\theta, \theta^{(t)}$に対して次を満たす代理関数$g(\cdot|\theta^{(t)})$を考える：
$$f(\theta) \geq g(\theta|\theta^{(t)}), \quad f(\theta^{(t)}) = g(\theta^{(t)}|\theta^{(t)}).$$
2. $\theta^{(t+1)} = \mathrm{argmax}_{\theta}\, g(\theta|\theta^{(t)})$を$t = 0, \ldots, T-1$について逐次的に解く．

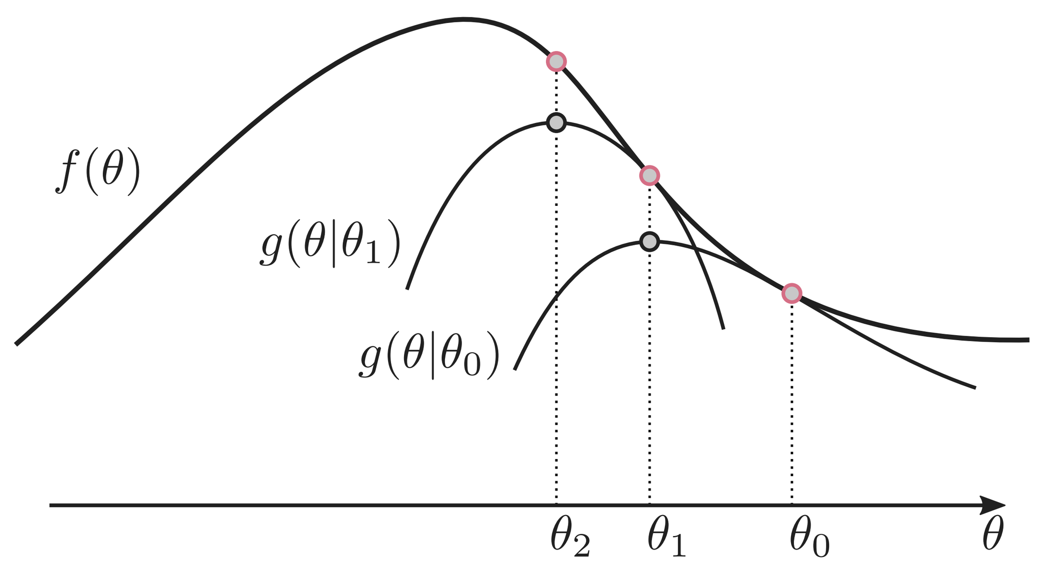


Figure 2. MMアルゴリズムの概念図．

本研究では，パラメータ$\boldsymbol{L}$が未知の行列式点過程から i.i.d. な観測$\mathcal{A}_1, \ldots, \mathcal{A}_M \subseteq \mathcal{Y}$が得られた状況を考える．

命題：MMアルゴリズムの代理関数

目的関数$f(\boldsymbol{L})$が行列式点過程の平均対数尤度であるとき，次の凹関数$g(\cdot|\boldsymbol{L}^{(t)})$は代理関数の条件を満たす：

$$g(\boldsymbol{L}|\boldsymbol{L}^{(t)}) = -\frac{1}{M}\sum_{m=1}^{M}\mathrm{tr}\{\boldsymbol{L}^{(t)}\boldsymbol{U}_{\mathcal{A}_m}^{\top}[\boldsymbol{L}^{(t)}]_{\mathcal{A}_m}^{-1}\boldsymbol{U}_{\mathcal{A}_m}\boldsymbol{L}^{(t)}\boldsymbol{L}^{-1}\} - \mathrm{tr}\{(\boldsymbol{L}^{(t)}+\boldsymbol{I})^{-1}\boldsymbol{L}\} + \mathrm{const.} \qquad (1)$$

提案法：代理関数の最適化と性質

MMアルゴリズムの代理関数は，その最適化問題が解きやすい形式であればベターである．提案する代理関数(1)の最大化問題について，次の命題が成り立つ：

命題：代理関数の最大化

代理関数(1)を大域的に最大化する$\boldsymbol{L}$は，次の方程式を満たす：

$$-\boldsymbol{L}(\boldsymbol{L}^{(t)}+\boldsymbol{I})^{-1}\boldsymbol{L} + \boldsymbol{Q}_M^{(t)} = \boldsymbol{O}, \qquad (2)$$

$$\boldsymbol{Q}_M^{(t)} = \boldsymbol{L}^{(t)}\left(\frac{1}{M}\sum_{m=1}^{M}\boldsymbol{U}_{\mathcal{A}_m}^{\top}[\boldsymbol{L}^{(t)}]_{\mathcal{A}_m}^{-1}\boldsymbol{U}_{\mathcal{A}_m}\right)\boldsymbol{L}^{(t)}.$$

式(2)はCARE (Continuous Algebraic Riccati Equation) とよばれる2次の行列方程式の一種であり，Schur method などで数値的に解ける．したがって提案法では，式(2)を満たす$\boldsymbol{L}$を逐次的に求めることで最尤推定を行う．

本研究と同様の問題設定のもと，Mariet+, 2015は不動点アルゴリズムとして$\boldsymbol{L}$の更新則を導いた．Mariet らの更新則は本研究とは異なる非凹な代理関数に基づくMMアルゴリズムとみなすことができ，提案法とは次のような関係がいえる：

命題：既存法との関連

式(1)で与えられる代理関数$g(\boldsymbol{L}|\boldsymbol{L}^{(t)})$は，$\boldsymbol{L}^{(t)}$の近傍に位置する$\boldsymbol{L}$について，Mariet らの代理関数よりもタイトな$f(\boldsymbol{L})$の下界を与える．

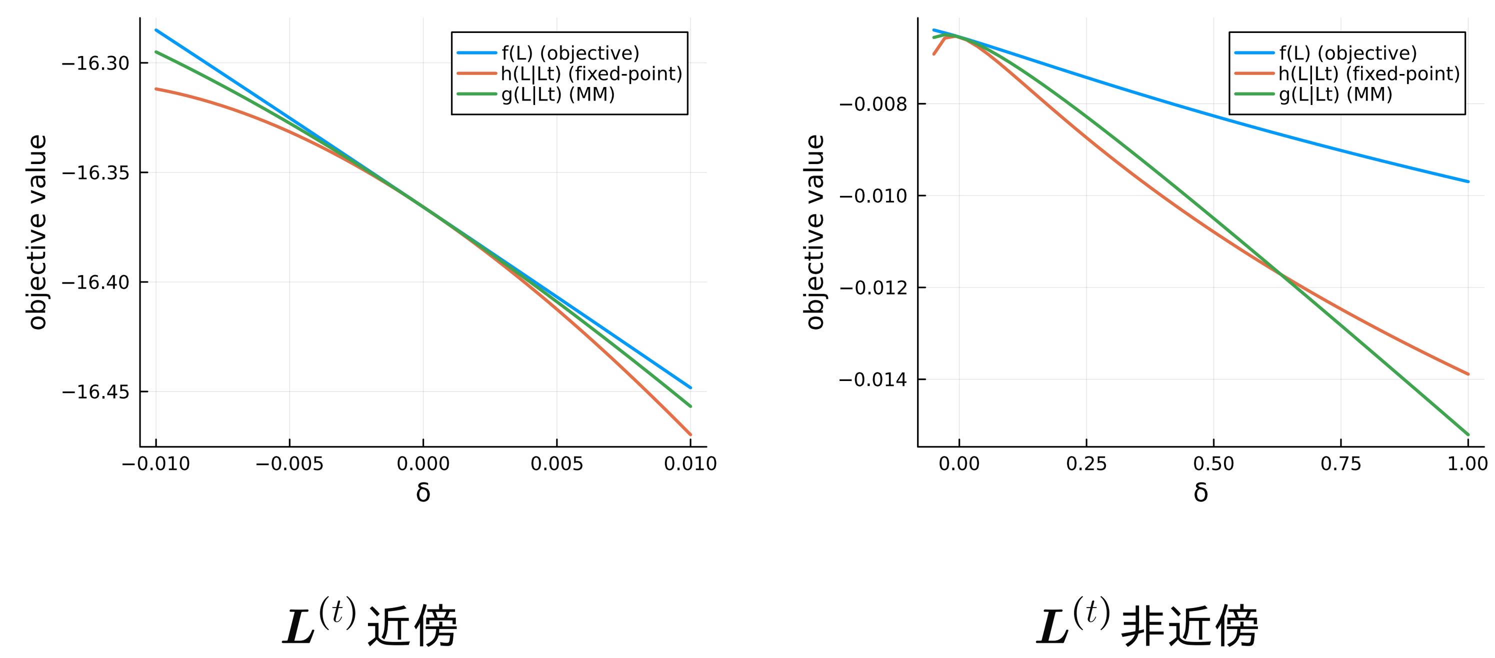
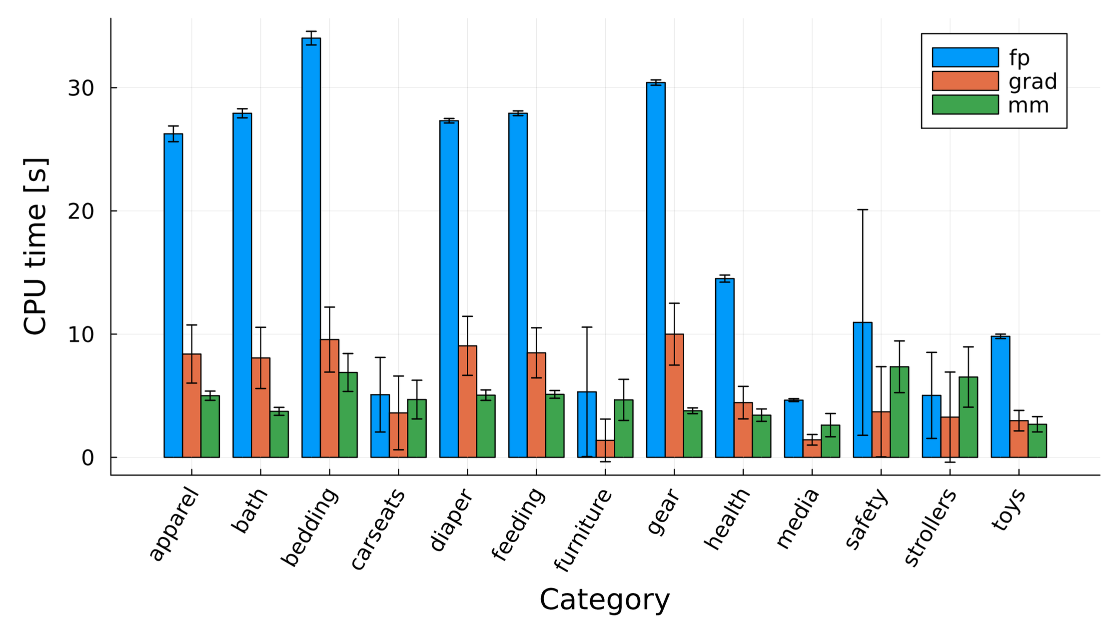


Figure 3. 代理関数の振る舞い．青：平均対数尤度（目的関数），橙：既存法の代理関数，緑：提案法の代理関数．

実験：Amazon Registry Dataset

Amazon Registry Dataset (Gillenwater+, 2014) に対して提案手法を適用し，提案手法が安定して高速であることを確認した．



Category	N	M
apparel	100	14,970
bath	100	14,542
bedding	100	16,370
carseats	34	7,566
diaper	100	16,759
feeding	100	19,001
furniture	32	7,093
gear	100	16,823
health	62	14,057
media	58	5,904
safety	36	8,892
strollers	40	7,393
toys	62	10,073

Figure 4. Amazon Baby Registry の各サブデータに対する実験結果とデータサイズ．異なる30の初期値から計算時間の平均と標準偏差を算出．fp：不動点アルゴリズム，grad：Adam，mm：提案法．

本研究は日野教授（統数研）との共同研究である．