情報幾何学的な関数族における パラメータ推定特性

A Study of Parameter Estimation Property for the Information Geometric Families of Functions

北陸先端科学技術大学院大学 ○ 林 幸雄

Yukio Hayashi

Japan Advanced Institute of Science and Technology, Hokuriku

Abstract We study an estimation property in the families of parametric functions called α-affine manifolds. It has been shown that the estimation results are continuously changed according to the values of $\alpha \in R$; however the case of $\alpha \approx 1$ is numerical instable with very slow convergence, and discontinuous for the case of $\alpha = 1$ as well-known exponential family.

1 はじめに

情報幾何学[1]は, パラメトリックなモデルやシステムの族に共通な性質を解明するための方法論で, 統計学をはじめ, 情報理論, 数理物理学, システム理論, 数理計画など, さまざまな分野に広がって進展している. 例えば, ニューラルネットや独立成分分析における自然勾配法[2]は, (オンラインで)効率的な学習が行なえる点で注目を集めている. この手法では, 原理的には勾配系の反復計算毎に計量の逆行列を求める必要があるが, 解くべき問題の構造をうまく活かした解法[2] や近似法[9] などが提案されている. 一方, 同様な勾配系で表される半正定値問題に対する内点法では, その許容領域が自己平行の場合, 反復なしで陽に解が求められる[8].

我々は, α-分布族と呼ばれる確率分布族 S_α (及びその正規化条件を外した関数族 : α-アファイン多様体$\tilde{S}_\alpha$) においても, 同様な自己平行性が成り立つ場合は, 反復計算なしで測地線 (勾配系の解曲線) が求められることを示している[6] [7]. このような測地線は, ある観測データと多様体上の関数値との距離に相当する評価基準 (ダイバージェンス) を最小化する自然勾配系に従うことから, その収束値がパラメータ推定に利用できる. 特に, $\alpha = 1$ の指数型分布族では, これは良く知られた最ゆう推定となる[3].

一方, 一般のα-分布族やα-アファイン多様体に関する推定問題を扱った研究は少ないが, この関数族はファジィ平均演算子に相当する[5] 点で興味深い. 本論文では, 上記のパラメータ推定問題において, 種々の$\alpha \in R$の値に対する推定特性を実験的に比較検討する.

2 ダイバージェンス最小化に基づく推定法

まず, 有限離散変数 $x \in \mathcal{X} = \{1, 2, \ldots, m\}$ 上の関数族$\tilde{M} = \{\tilde{q}(x)\}$ の中で, パラメータ $\tilde{\theta} = (\tilde{\theta}_1, \tilde{\theta}_2, \ldots, \tilde{\theta}_n)$ によって規定される n 次元部分多様体$\tilde{S}_\alpha = \{m_\alpha(x;\ \tilde{\theta})\}$ を考える[1].

$$\frac{2}{1-\alpha}\left[m_\alpha(x;\ \tilde{\theta})\right]^{\frac{1-\alpha}{2}} \stackrel{\text{def}}{=} C(x) + \sum_{i=1}^{n} \tilde{\theta}^i F_i(x),$$

但し, $n < m$, $\alpha \neq \pm 1$, $m_\alpha(x;\ \tilde{\theta}) > 0$, $\sum_{x\in\mathcal{X}} m_\alpha(x;\ \tilde{\theta}) < \infty$ とする. また, $C(x) = 0$ で, $\mathcal{X}$上の任意の基底関数 $F_i(x)$ は互いに一次独立とする[1].

さて, ある観測データ$\tilde{q}(x)$ を最も良く推定する$\tilde{S}_\alpha$上の関数 m_αと, そのパラメータ値$\tilde{\theta}$を求める問題を考えよう. 時系列なデータに対しては, ある期間に蓄えたデータに対して, そのサンプリング時刻毎に同様な推定を行なうものとする.

評価基準としてα-ダイバージェンス D_αの最小化を考えると, 以下の(自然)勾配系の解が推定値を与える. 又, $D_\alpha \geq 0$ の凸性から, 任意の初期値から唯一の解が得られる.

$$\frac{d\tilde{\theta}^i}{dt} = -\sum_{j=1}^{n} g^{ij}(\tilde{\theta})\frac{\partial D_\alpha(m_\alpha||\tilde{q})}{\partial\tilde{\theta}^j}, \qquad (1)$$

ここで, $[g^{ij}]$ は Fisher 計量 $[g_{ij}]$ の逆行列であるが, Hessian $g_{ij} = \frac{\partial D_\alpha}{\partial\tilde{\theta}^i\partial\tilde{\theta}^j}$でもある. また,

$$D_\alpha(m_\alpha||\tilde{q}) \stackrel{\text{def}}{=} E_{m_\alpha}\left[f_\alpha\left(\frac{\tilde{q}(x)}{m_\alpha(x;\ \tilde{\theta})}\right)\right], \qquad (2)$$

$$f_\alpha(u) \stackrel{\text{def}}{=} \frac{4}{1-\alpha^2}\left[\frac{1-\alpha}{2} + \frac{1+\alpha}{2}u - u^{\frac{1+\alpha}{2}}\right], \qquad (3)$$

$$\frac{\partial D_\alpha}{\partial\tilde{\theta}^j} = \frac{2}{1+\alpha}\sum_x\left\{F_j\left(m_\alpha^{\frac{1+\alpha}{2}} - \tilde{q}^{\frac{1+\alpha}{2}}\right)\right\},$$

である(E_{m_α}は m_αによる期待値).

図 1の点線は, 勾配系 (1) を正規化したα-分布族の測地線を摸式的に表す[2].

3 実験結果

$n = 9$, $m = 125$ の例で勾配系 (1) を 4 次の Runge-Kutta 法で解き, αの値に従った推定結果の変化を調べた($\alpha = \pm 1$ の場合は, 指数型及び混合型分布族に対する推定[6]を行なった). ここで, 式 (2)(3) による D_αは異なるαの値に対してそれぞれ別々の評価基準を与えるので, p_αと q との平均 2 乗誤差 MSE で推定結果の比較を行なった.

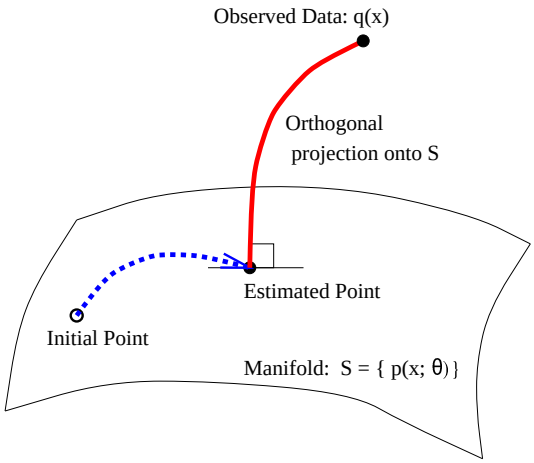

図 1: 観測データ $q(x)$ に対するα-分布族 $S_\alpha = \{p_\alpha(x;\ \theta)\}$ 上のパラメータ推定

実験結果[3]を以下にまとめる.

推定誤差 **MSE:** $0 \leq \alpha \leq 1$ では, αの値が大きくなる程, 収束特性グラフが上にシフトし, $\alpha = 1 - \varepsilon$付近から急激な数値的発散を起こす(図 2, $\varepsilon \approx 0.2$). 逆に, $\alpha > 1$ では, $\alpha = 1 + \varepsilon$付近の発散状況から, グラフが下にシフトする(図 3, 4). $\alpha < -1$ では, ほとんど変化はない[4].

パラメータ値 $\tilde{\theta}^i$: $\alpha = 1 \pm \varepsilon$付近で値が負正に増大(図 5, 6). 但し, $\alpha = 1$ の場合を除いて, グラフは連続的に変化. 又, $\alpha > 0$ の値が大きくなるほど, $\tilde{\theta}^i \to 0_-$ (図 5).

関数値 p_α: 同様に, $\alpha = 1 - \varepsilon$付近で振幅が急増するが(図 7), $\alpha = 1$ の場合を除いてグラフは連続的に変化(図 7, 8). $\alpha > 1$,

[1] $\mathcal{X}$の有限離散性には, 無限次元の議論が避けられる, 変数 x による積分が和になりその逆演算が行列計算に帰着する, 計算機システム上の事象への応用に適当[5]などのメリットがある.

[2] 正規化: $p_\alpha(x;\ \theta) = m_\alpha(x;\ \tilde{\theta})/\sum_x m_\alpha(x;\ \tilde{\theta})$, $q(x) = \tilde{q}(x)/\sum_x \tilde{q}(x)$.

[3] 観測データや基底関数(スケーリングの違いも含む)を換えた場合にも, 全く同様な結果.

[4] D_αのグラフもこれらと同様な収束特性.

$\alpha < -1$ の場合もそれぞれ同様.

又, 観測データ q (所々に平坦部分を持つグラフ) に対して, $\alpha = \pm 1$ の場合の推定値は比較的良くマッチして妥当 (図 9).

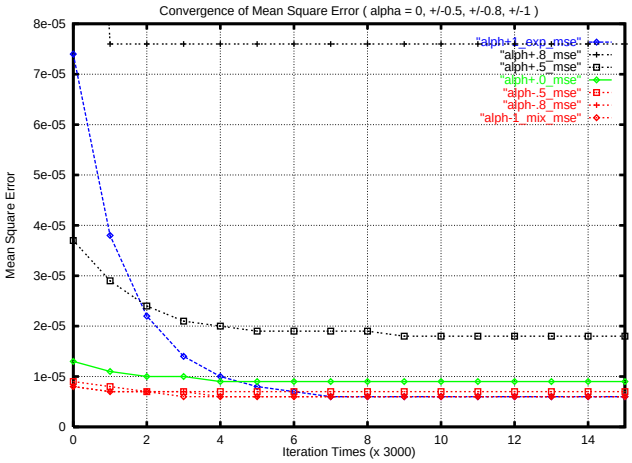

図 2: 推定誤差の収束性 $(0 \leq \alpha \leq 1)$

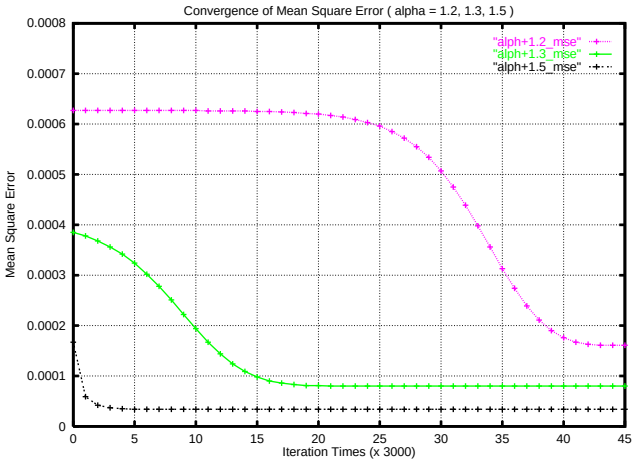

図 3: 推定誤差の収束性 $(\alpha \approx 1 + \varepsilon)$

4 おわりに

本論文では, α-分布族の測地線による基本的なパラメータ推定特性を実験的に比較検討した. 一方, 勾配系を介さないより効率的な計算方法に関しては, (自己平行でない場合を含めて) 別の機会で報告したい. 我々は, 情報幾何学とアルゴリズム論の密接な関係に着目しているが, 可積分系からのアプローチ[4]を

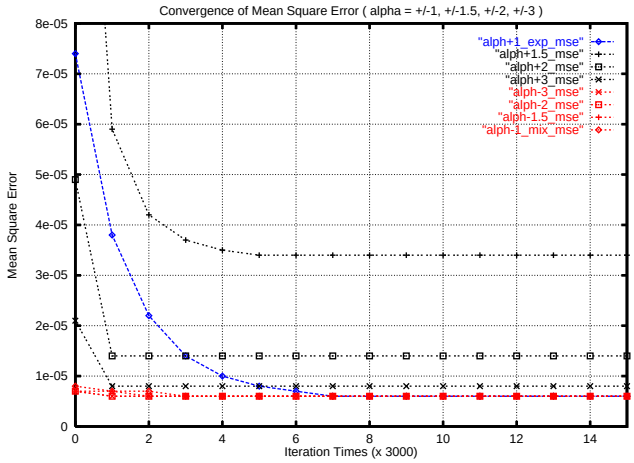

図 4: 推定誤差の収束性 $(\alpha \geq 1, \alpha \leq -1)$

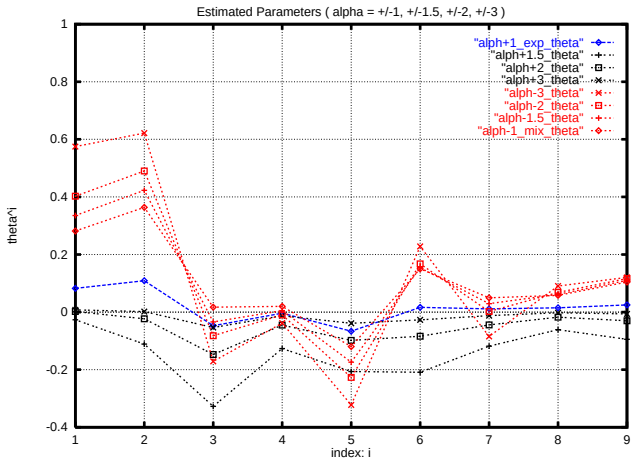

図 5: 推定された $\tilde{\theta}^i$ $(\alpha \geq 1, \alpha \leq -1)$

含めて, 効率的なアルゴリズムの背後に潜む共通の数理構造が徐々に明らかにされつつあり, これらの今後の進展に大いに期待したい.

謝辞: 本研究の一部は, (財) 人工知能研究振興財団, 及び, 文部省科学研究費 10878045 の援助を受けている.

参考文献

1) 甘利 俊一, 長岡 浩司. "情報幾何の方法," 岩波講座 応用数学 [対象 12], 岩波書店, 1994.

2) S. Amari, "Natural Gradient Works Efficiently in Learning," *Neural Computation*, Vol. 10, pp. 251-276, 1998.

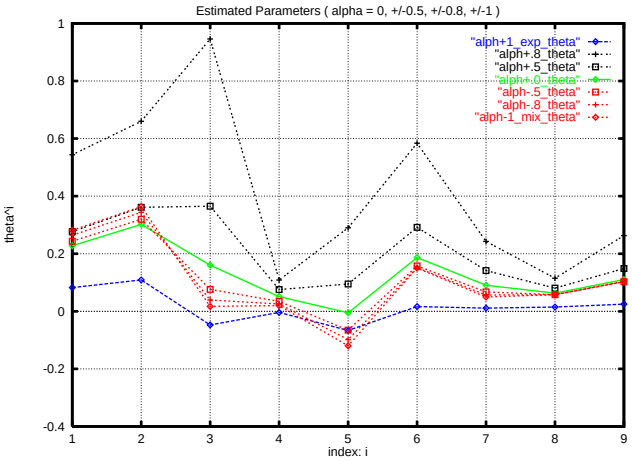

図 6: 推定された $\tilde{\theta}^i$ $(0 \leq \alpha \leq 1)$

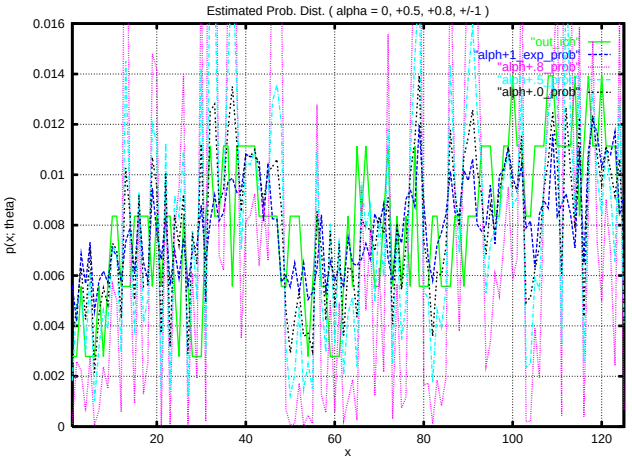

図 7: 推定された関数 p_α $(0 \leq \alpha \leq 1)$

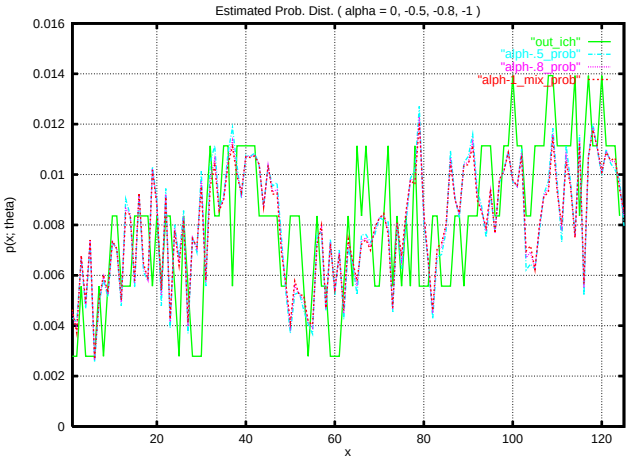

図 8: 推定された関数 p_α $(-1 \leq \alpha \leq 0)$

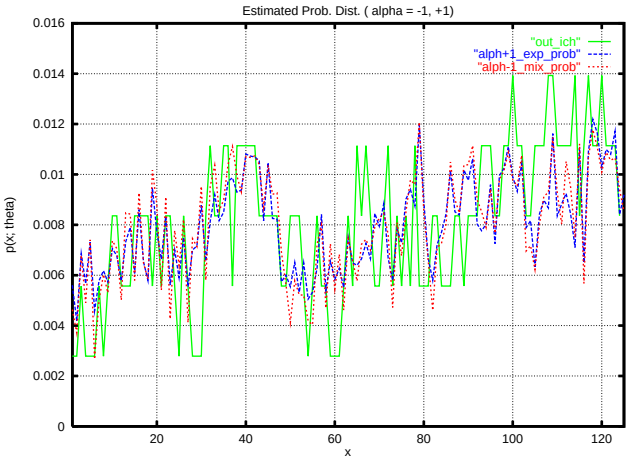

図 9: 推定された関数 p_α $(\alpha = \pm 1)$

3) A. Fujiwara and S. Amari, "Gradient systems in view of information geometry," *Physica D*, Vol. 80, pp. 317-327, 1995.

4) 中村 佳正, "アルゴリズムと可積分系," システム/情報/数理, Vol. 43, No. 11, pp. 584-591, 1999.

5) 林 幸雄, "個人選好による情報アクセスに適したデータモデルについて," 情処研報, 98-DBS-116, pp. 381-388, 1998.

6) Y. Hayashi, "Direct Calculation Methods for Parameter Estimation in Statistical Manifolds of Finite Discrete Distributions," *IECE Trans. on Fundamentals*, Vol. E81-A, No. 7, pp. 1486-1492, 1998.

7) Y. Hayashi, "Enhancement of a Gradient Descent or Ascent with a Function Weight for the Minimum α-Divergence," *Proc. of IEEE SMC'99*, Vol. V, pp. 98-103, 1999.

8) A. Ohara, "Information Geometric Analysis of a Interior-Point Method for Semidefinite Programming," 数理解析研講究録, 1004, pp. 71-89, 1997.

9) H. Park, S. Amari, and K. Fukumizu, "Natural Gradient Learning Algorithms for Various Stochastic Models," Submitted