Contrastive Divergence 法の漸近理論と情報幾何

藤平 燎

Contrastive Divergence (CD) 法とは，計算量的に困難な正規化定数の評価が必要となる統計的モデルのパラメータ推定問題において，近似的に最尤推定を行うために Hinton [1] によって提案された手法である．CD 法は，ある更新則に従ってパラメータを逐次更新する手法であり，最尤推定量に近い推定量が得られることが経験的に知られている．しかしながら，パラメータの更新則を表す力学系の収束性や，その収束先として得られる推定量の一致性の証明は，これまで与えられていなかった．本研究の目的は，CD 法におけるパラメータ更新則を表す力学系の研究を通じ，CD 法によって得られる推定量の漸近的性質を明らかにすることにある．

状態空間 $\mathcal{X}$ を有限集合，パラメータ空間 Θ を $\mathbb{R}^d$ の開部分集合とする．$\mathcal{X}$ 上の統計的モデル $\{p(x;\theta)\}_{\theta\in\Theta}$ と未知の確率分布 $p(x;\theta_*)$ に従う i.i.d. 観測データ $(X_1,\dots,X_N)\in\mathcal{X}^N$ が与えられたとき，真のパラメータ値 $\theta_*\in\Theta$ を推定する問題をパラメータ推定問題といい，観測データをもとに推定値を与える関数 $\hat{\theta}_N:\mathcal{X}^N\to\Theta$ をパラメータ θ の推定量という．統計的モデル $\{p(x;\theta)\}_{\theta\in\Theta}$ に対して，$\displaystyle\sum_{x\in\mathcal{X}}p(x;\theta)u(x;\theta)=0$ を満たし，かつ $\displaystyle\sum_{x\in\mathcal{X}}p(x;\theta)\frac{\partial u}{\partial\theta}(x;\theta)$ が正則行列となるような関数 $u:\mathcal{X}\times\Theta\to\mathbb{R}^d$ を推定関数といい，

$$\psi_N(\theta):=\frac{1}{N}\sum_{n=1}^{N}u(X_n;\theta)$$

の零点として定義される推定量 $\hat{\theta}_N$ を Z 推定量という．上記設定のもとで，Z 推定量は $N\to\infty$ のとき強一致性と漸近正規性をもつことが知られている．本論文では，観測データ $(X_1,\dots,X_N)$ が与えられたとき，θ を時刻 t の関数 $\theta(t)$ と見て，$\hat{\theta}_N$ を平衡点とする力学系

$$\dot{\theta}=\psi_N(\theta)$$

を考え，これを推定力学系と呼ぶことにする．ここに，$\dot{\theta}$ は t に関する θ の微分である．固有値の実部がすべて負であるような正方行列を安定行列といい，$\dfrac{\partial\psi_N}{\partial\theta}(\hat{\theta}_N)$ が安定行列ならば，$\hat{\theta}_N$ は局所漸近安定な平衡点となる．このとき，$\hat{\theta}_N$ に十分近い初期点をとって推定力学系に従ってパラメータを更新することで，$t\to\infty$ のとき，パラメータは $\hat{\theta}_N$ に収束する．

本研究の第 1 の主結果は以下の定理である．

定理 1. 関数 $\psi_*:\Theta\to\mathbb{R}^d$ を

$$\psi_*(\theta):=\sum_{x\in\mathcal{X}}p(x;\theta_*)u(x;\theta)$$

で定義する．$\dfrac{\partial\psi_*}{\partial\theta}(\theta_*)$ が安定行列ならば，$\dfrac{\partial\psi_N}{\partial\theta}(\hat{\theta}_N)$ が安定行列となる確率は，$N\to\infty$ のとき 1 に収束する．特に，$\hat{\theta}_N$ が推定力学系 $\dot{\theta}=\psi_N(\theta)$ の局所漸近安定な平衡点となる確率は，$N\to\infty$ のとき 1 に収束する．

統計的モデルが

$$p(x;\theta)=\frac{\exp(-H(x;\theta))}{Z(\theta)},\qquad\left(Z(\theta):=\sum_{x\in\mathcal{X}}\exp(-H(x;\theta))\right)$$

と表される場合を考えよう．このようなモデルをエネルギーベースモデルといい，$H(x;\theta)$ をエネルギー関数，$Z(\theta)$ を正規化定数または分配関数という．エネルギーベースモデルは，様々な分野，例えば画像処理などで用いられるマルコフ確率場を含む重要な統計的モデルである．しかしながら，例えば $\mathcal{X}=\{0,1\}^L$ のとき，$Z(\theta)$ の計算量は $O(2^L)$ となることからも分かるように，$Z(\theta)$ の評価には，しばしば計算量的な困難が伴う．

エネルギーベースモデルを用いた場合の最尤推定における推定関数は

$$s(x;\theta):=\frac{\partial\log p}{\partial\theta}(x;\theta)=\sum_{y\in\mathcal{X}}p(y;\theta)\frac{\partial H}{\partial\theta}(y;\theta)-\frac{\partial H}{\partial\theta}(x;\theta),$$

であり，対応する推定力学系は

$$\dot{\theta} = \sum_{y \in \mathcal{X}} p(y;\theta) \frac{\partial H}{\partial \theta}(y;\theta) - \frac{1}{N}\sum_{n=1}^{N} \frac{\partial H}{\partial \theta}(X_n;\theta)$$

となる．ここで，右辺第 1 項は $p(y;\theta)$ に関する期待値であり，それを厳密に評価するためには計算量的に困難な $Z(\theta)$ の評価が必要となる．$Z(\theta)$ の直接計算を回避し，近似的に期待値を評価する方法の一つにマルコフ連鎖モンテカルロ法がある．しかしながら，マルコフ連鎖モンテカルロ法では，θ の値を更新するごとに，$p(y;\theta)$ に収束するマルコフ連鎖を新たに構成し，十分な回数の遷移を行わなければならない．

この問題を回避するため，Hinton [1] は，マルコフ連鎖の初期状態を観測データ $(X_1, \ldots, X_n)$ とし，遷移回数を 1 回だけとするパラメータ更新則：

$$\dot{\theta} = \frac{1}{N}\sum_{n=1}^{N} \left(\sum_{y \in \mathcal{X}} T(y|X_n;\theta) \frac{\partial H}{\partial \theta}(y;\theta) - \frac{\partial H}{\partial \theta}(X_n;\theta) \right)$$

を提案し，Contrastive Divergence (CD) 法と名付けた．ここに，$T(y|x;\theta)$ は，マルコフ連鎖の遷移核である．このパラメータ更新則に従って推定を行うと，最尤推定値に近い値が得られることが経験的には知られていたが，それを数学的に正当化する研究はこれまでなかった．

本研究では，まず，以下の量を導入する．

$$\begin{aligned} u^{\mathrm{CD}}(x;\theta) &:= s(x;\theta) - \sum_{y \in \mathcal{X}} T(y|x;\theta) s(y;\theta) \\ \psi_N^{\mathrm{CD}}(\theta) &:= \frac{1}{N}\sum_{n=1}^{N} u^{\mathrm{CD}}(X_N;\theta) \\ \psi_*^{\mathrm{CD}}(\theta) &:= \sum_{x \in \mathcal{X}} p(x;\theta_*) u^{\mathrm{CD}}(x;\theta) \end{aligned}$$

すると，$u^{\mathrm{CD}}(x;\theta)$ は推定関数となり，対応する推定力学系

$$\dot{\theta} = \psi_N^{\mathrm{CD}}(\theta)$$

は，CD 法のパラメータ更新則に一致することが証明できる．従って，ψ_N^{CD} の零点を $\hat{\theta}_N^{\mathrm{CD}}$ とおくと，$\hat{\theta}_N^{\mathrm{CD}}$ は Z 推定量であり，$N \to \infty$ のとき強一致性と漸近正規性をもつ．さらに以下の事実が証明できる．これが本研究の第 2 の主結果である．

定理 2. $\dfrac{\partial \psi_*^{\mathrm{CD}}}{\partial \theta}(\theta_*)$ は安定行列である．

定理 1 と定理 2 を組み合わせることで，以下の系を得る．これは，CD 法が経験的に成功をおさめてきたことに対する一つの数学的根拠を与えるものである．

系 3. $\hat{\theta}_N^{\mathrm{CD}}$ が推定力学系 $\dot{\theta} = \psi_N^{\mathrm{CD}}(\theta)$ の局所漸近安定な平衡点となる確率は，$N \to \infty$ のとき 1 に収束する．

本研究ではさらに，推定関数 $u^{\mathrm{CD}}(x;\theta)$ から誘導される双対幾何構造についても考察を行った．推定関数が与えられたとき，自然なプレコントラスト関数を通して，統計的モデルに誘導される統計多様体構造には，捩率が生じ得ることが知られている [3]．ただし，そのような統計的モデルと推定関数の具体例は，ややアドホックなものしか知られていなかった．本論文では，より自然に捩率が登場する例として，ある非対称な構造をもつボルツマンマシンに CD 法を適用したとき，対応する推定関数が非可積分になり，誘導される双対アファイン接続の一方に捩率が生じることを見出した．

参考文献

[1] G. E. Hinton, “Training Products of Experts by Minimizing Contrastive Divergence,” *Neural Computation*, **14**, pp.1771-1800 (2002).

[2] A. W. van der Vaart, *Asymptotic Statistics*, Cambridge University Press (1998).

[3] 逸見昌之, “推定関数と捩れを許す統計多様体,” 数理解析研究所講究録 **1916**, pp.18-36 (2014).