Contrastive Divergence 法の漸近理論と情報幾何

藤平 燎

2019 年 2 月

目次

第 1 章　序論　2

第 2 章　準備　6

2.1　Z 推定量の漸近理論 6

2.2　力学系の解の安定性 17

2.3　マルコフ連鎖モンテカルロ法 18

2.4　捩れをもつ統計多様体 20

第 3 章　主結果　26

3.1　推定力学系 26

3.2　Contrastive Divergence 法の漸近理論 27

3.3　Contrastive Divergence 法の情報幾何 31

第 4 章　結言　38

参考文献　40

第 1 章

序論

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 [2] 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 $(X_1,\ldots,X_N)\in\mathcal{X}^N$ 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 [10, pp.179-183]．本論文では，観測データ $(X_1,\ldots,X_N)$ 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 1.1. 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ばならない．

この問題を回避するため，Hinton [2] 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．

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

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

定理 1.1 と定理 1.2 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 1.3. $\hat{\theta}_N^{\mathrm{CD}}$ 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 [1]．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 [6, pp.333-339] 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．

本論文の構成

本論文の構成は以下の通りである．第 2 章では，準備として本論文の主題に関わる数学的概念について論じる．まず 2.1 節では，Z 推定量を定義し，その存在と漸近的性質について述べる．2.2 節では，$\dot{\theta} = f(\theta)$ の形で表せる力学系の平衡点が局所漸近安定となるための十分条件について述べる．2.3 節では，マルコフ連鎖モンテカルロ法の定義を述べ，適切な仮定のもとで，マルコフ連鎖が不変分布に収束することを証明する．2.4 節では，統計的モデル上に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が定まり，双対幾何構造が誘導されることについて述べる．

第 3 章が，本論文で新たに得られた結果である．まず 3.1 節で，推定力学系を定義し，その平衡点の局所漸近安定性について述べる．3.2 節では，Contrastive Divergence 法に対応する推定力学系を導出し，その平衡点が局所漸近安定である確率が，データ数 $N \to \infty$ のとき 1 に収束することを証明する．3.3 節では，Contrastive Divergence 法から 2.4 節で述べた方法で，捩れをもつ統計多様体構造が誘導されることを具体例を用いて説明する．

第 2 章

準備

2.1　Z 推定量の漸近理論

本論文を通して，状態空間 $\mathcal{X}$ は有限集合，パラメータ空間 Θ は $\mathbb{R}^d$ の開部分集合とし，$d < \#\mathcal{X}$ とする．また，$\mathcal{X}$ 上の確率分布族 $\{p(x;\theta)\}_{\theta\in\Theta}$ は以下を満たすと仮定し，これを統計的モデルと呼ぶ．

1. 任意の $x \in \mathcal{X}$，$\theta \in \Theta$ に対し，$p(x;\theta) > 0$.
2. 任意の $x \in \mathcal{X}$ に対し，$\theta \mapsto p(x;\theta)$ は C^∞ 級.
3. 任意の $x \in \mathcal{X}$ に対して $p(x;\theta_1) = p(x;\theta_2)$ ならば，$\theta_1 = \theta_2$.
4. 任意の $\theta \in \Theta$ に対し，
$$1, \frac{\partial \log p}{\partial \theta^1}(x;\theta), \ldots, \frac{\partial \log p}{\partial \theta^d}(x;\theta)$$
が $\mathcal{X}$ 上の関数として 1 次独立.

真のパラメータ値を $\theta_* \in \Theta$ とするとき，$p(x;\theta_*)$ に従う独立な確率変数列 $(X_1, \ldots, X_N) \in \mathcal{X}^N$ をサイズ N の観測データと呼ぶ．統計的モデル $\{p(x;\theta)\}_{\theta\in\Theta}$ に対し，観測データ $(X_1, \ldots, X_N)$ をもとに，真のパラメータ値 θ_* 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．

また，実行列 $A \in \mathbb{R}^{p\times q}$ に対して，A^{T} で A の転置行列を表し，ノルムを以下で定義する．

$$\|A\| := \sqrt{\mathrm{tr} A^{\mathrm{T}} A} = \sqrt{\sum_{i=1}^{p} \sum_{j=1}^{q} A_{ij}^2}$$

補題 2.1. 任意の行列 $A \in \mathbb{R}^{p\times q}$，$B \in \mathbb{R}^{q\times r}$ に対して，

$$\|AB\| \leq \|A\|\|B\|$$

が成り立つ．

証明． Cauchy-Schwarz の不等式より，

$$
\begin{aligned}
\|AB\|^2 &= \sum_{i=1}^{p}\sum_{j=1}^{r}\left(\sum_{k=1}^{q} A_{ik}B_{kj}\right)^2 \\
&\leq \sum_{i=1}^{p}\sum_{j=1}^{r}\left(\sum_{k=1}^{q} A_{ik}^2\right)\left(\sum_{\ell=1}^{q} B_{\ell j}^2\right) \\
&= \left(\sum_{i=1}^{p}\sum_{k=1}^{q} A_{ik}^2\right)\left(\sum_{\ell=1}^{q}\sum_{j=1}^{r} B_{\ell j}^2\right) \\
&= \|A\|^2\|B\|^2
\end{aligned}
$$

□

定義 2.2. 以下の 3 条件を満たす $u : \mathcal{X} \times \Theta \to \mathbb{R}^d$ を推定関数という．

1. 任意の $x \in \mathcal{X}$ に対して，$\theta \mapsto u(x;\theta)$ は C^∞ 級．
2. 任意の $\theta \in \Theta$ に対して，
$$\sum_{x\in\mathcal{X}} p(x;\theta)u(x;\theta) = 0$$
3. 任意の $\theta \in \Theta$ に対して，
$$A(\theta) := -\sum_{x\in\mathcal{X}} p(x;\theta)\frac{\partial u}{\partial\theta}(x;\theta)$$
は正則．

ここで，

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

とおくと，推定関数の定義より $\psi_*(\theta_*) = 0$ であり，また $\dfrac{\partial\psi_*}{\partial\theta}(\theta_*) = -A(\theta_*)$ であるから，$\dfrac{\partial\psi_*}{\partial\theta}(\theta_*)$ は正則である．このことから，θ_* は局所的に ψ_* の唯一の零点であることが示せる．証明には次の補題を用いる．

補題 2.3 (縮小写像の不動点定理 [7, p220])**.** D を $\mathbb{R}^d$ のコンパクト部分集合とする．$\varphi : D \to D$ が縮小写像であるとき，φ は D に唯一の不動点をもつ．すなわち，ある $\gamma \in (0,1)$ が存在して，任意の $\theta_1, \theta_2 \in D$ に対して

$$\|\varphi(\theta_1) - \varphi(\theta_2)\| \leq \gamma\,\|\theta_1 - \theta_2\|$$

が成り立つとき，ある $\theta_0 \in D$ が一意的に存在して，$\varphi(\theta_0) = \theta_0$ が成り立つ．

命題 2.4. ある $\delta > 0$ が存在して，θ_* は $\overline{B_\delta(\theta_*)} := \{\theta \in \Theta \mid \|\theta - \theta_*\| \leq \delta\}$ 上で ψ_* の唯一の零点である．

証明． $\dfrac{\partial\psi_*}{\partial\theta}$ の連続性より，ある $\delta > 0$ が存在して，任意の $\theta \in \overline{B_\delta(\theta_*)}$ に対して，

$$\left\| \frac{\partial\psi_*}{\partial\theta}(\theta) - \frac{\partial\psi_*}{\partial\theta}(\theta_*) \right\| \leq \frac{1}{4} \left\| \left(\frac{\partial\psi_*}{\partial\theta}(\theta_*) \right)^{-1} \right\|^{-1}$$

が成り立つ．

$$\varphi_*(\theta) := \theta - \left(\frac{\partial\psi_*}{\partial\theta}(\theta_*) \right)^{-1} \psi_*(\theta)$$

とおくと，各 $\theta \in \Theta$ に対し，$\psi_*(\theta) = 0$ と $\varphi_*(\theta) = \theta$ は同値である．さらに，任意の $\theta \in \overline{B_\delta(\theta_*)}$ に対して，

$$\begin{aligned}
\left\| \frac{\partial\varphi_*}{\partial\theta}(\theta) \right\| &= \left\| I - \left(\frac{\partial\psi_*}{\partial\theta}(\theta_*) \right)^{-1} \frac{\partial\psi_*}{\partial\theta}(\theta) \right\| \\
&= \left\| \left(\frac{\partial\psi_*}{\partial\theta}(\theta_*) \right)^{-1} \left(\frac{\partial\psi_*}{\partial\theta}(\theta_*) - \frac{\partial\psi_*}{\partial\theta}(\theta) \right) \right\| \\
&\leq \left\| \left(\frac{\partial\psi_*}{\partial\theta}(\theta_*) \right)^{-1} \right\| \left\| \frac{\partial\psi_*}{\partial\theta}(\theta_*) - \frac{\partial\psi_*}{\partial\theta}(\theta) \right\| \\
&\leq \left\| \left(\frac{\partial\psi_*}{\partial\theta}(\theta_*) \right)^{-1} \right\| \cdot \frac{1}{4} \left\| \left(\frac{\partial\psi_*}{\partial\theta}(\theta_*) \right)^{-1} \right\|^{-1} \\
&= \frac{1}{4}
\end{aligned}$$

が成り立つ．

次に，任意の $\theta_1, \theta_2 \in \overline{B_\delta(\theta_*)}$ に対して，

$$\begin{aligned}
\varphi_*(\theta_1) - \varphi_*(\theta_2) &= \int_0^1 \frac{\mathrm{d}}{\mathrm{d}t} \Big(\varphi_*(t\theta_1 + (1-t)\theta_2) \Big) \mathrm{d}t \\
&= \left(\int_0^1 \frac{\partial\varphi_*}{\partial\theta} \Big(t\theta_1 + (1-t)\theta_2 \Big) \mathrm{d}t \right) (\theta_1 - \theta_2)
\end{aligned}$$

であり，任意の $t \in [0,1]$ に対して $t\theta_1 + (1-t)\theta_2 \in \overline{B_\delta(\theta_*)}$ であることに注意して，両辺のノルムをとると，

$$\begin{aligned}
\|\varphi_*(\theta_1) - \varphi_*(\theta_2)\| &= \left\| \left(\int_0^1 \frac{\partial\varphi_*}{\partial\theta} \Big(t\theta_1 + (1-t)\theta_2 \Big) \mathrm{d}t \right) (\theta_1 - \theta_2) \right\| \\
&\leq \left\| \int_0^1 \frac{\partial\varphi_*}{\partial\theta} \Big(t\theta_1 + (1-t)\theta_2 \Big) \mathrm{d}t \right\| \|\theta_1 - \theta_2\| \\
&\leq \int_0^1 \left\| \frac{\partial\varphi_*}{\partial\theta} \Big(t\theta_1 + (1-t)\theta_2 \Big) \right\| \mathrm{d}t \cdot \|\theta_1 - \theta_2\| \\
&\leq \frac{1}{4} \|\theta_1 - \theta_2\|
\end{aligned}$$

を得る．

さらに，$\varphi_*(\theta_*) = \theta_*$ を用いると，任意の $\theta \in \overline{B_\delta(\theta_*)}$ に対して，

$$\|\varphi_*(\theta) - \theta_*\| = \|\varphi_*(\theta) - \varphi_*(\theta_*)\| \leq \frac{1}{4}\|\theta - \theta_*\| < \delta$$

となることより，$\varphi_*(\theta) \in \overline{B_\delta(\theta_*)}$ であるので，$\varphi_*\left(\overline{B_\delta(\theta_*)}\right) \subset \overline{B_\delta(\theta_*)}$ が成り立つ．

以上により，φ_* は $\overline{B_\delta(\theta_*)}$ 上の縮小写像であることが分かったので，縮小写像の不動点定理 (補題 2.3) より，θ_* は $\overline{B_\delta(\theta_*)}$ 上で φ_* の唯一の不動点となる．いいかえれば，θ_* は $\overline{B_\delta(\theta_*)}$ 上で ψ_* の唯一の零点である． □

定義 2.5 (Z 推定量)**.**

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

とおく．ψ_N の零点として得られる推定量を $\hat{\theta}_N$ で表し，Z 推定量という．

Z 推定量の存在に関する命題を述べるため，次の補題を準備する．

補題 2.6 (一様大数の強法則)**.** $f : \mathcal{X} \times \Theta \to \mathbb{R}^m$ が各 $x \in \mathcal{X}$ で $\theta \in \Theta$ に関して C^∞ 級で，$X_1, \ldots, X_N$ が独立に確率分布 p に従うとすると，Θ の任意のコンパクト凸部分集合 D に対して，

$$\lim_{N\to\infty} \sup_{\theta \in D} \left\| \frac{1}{N}\sum_{n=1}^{N} f(X_n; \theta) - \sum_{x \in \mathcal{X}} p(x) f(x; \theta) \right\| = 0 \quad \text{a.s.}$$

が成り立つ．

証明. $\mathcal{X}$ が有限集合，D がコンパクトであることから，ある $\eta > 0$ が存在し，任意の $x \in \mathcal{X}$, $\theta \in D$ に対して，

$$\left\| \frac{\partial f}{\partial \theta}(x; \theta) \right\| < \eta$$

が成り立つ．

$\varepsilon > 0$ を任意にとる．D がコンパクトであることを用いて，

$$D \subset \bigcup_{k=1}^{K} \left(D \cap \overline{B_{\varepsilon/\eta}(\theta_k)} \right)$$

となるように $K \in \mathbb{N}$ と $\theta_1, \ldots, \theta_K \in D$ をとる．

各 $k = 1, \ldots, K$ について，任意の $x \in \mathcal{X}$, $\theta \in D \cap \overline{B_{\varepsilon/\eta}(\theta_k)}$ に対して，

$$\begin{aligned} f(x; \theta) - f(x; \theta_k) &= \int_0^1 \frac{\mathrm{d}}{\mathrm{d}t}\Big(f(x; t\theta + (1-t)\theta_k) \Big) \mathrm{d}t \\ &= \left(\int_0^1 \frac{\partial f}{\partial \theta}(x; t\theta + (1-t)\theta_k) \mathrm{d}t \right)(\theta - \theta_k) \end{aligned}$$

であり，D が凸であるから，任意の $t \in [0,1]$ に対して $t\theta + (1-t)\theta_k \in D$ が成り立つことに注意して，両辺のノルムをとると，

$$\begin{aligned}
\|f(x;\theta) - f(x;\theta_k)\| &= \left\| \left(\int_0^1 \frac{\partial f}{\partial \theta}(x; t\theta + (1-t)\theta_k) \mathrm{d}t \right) (\theta - \theta_k) \right\| \\
&\leq \left\| \int_0^1 \frac{\partial f}{\partial \theta}(x; t\theta + (1-t)\theta_k) \mathrm{d}t \right\| \|\theta - \theta_k\| \\
&\leq \int_0^1 \left\| \frac{\partial f}{\partial \theta}(x; t\theta + (1-t)\theta_k) \right\| \mathrm{d}t \cdot \|\theta - \theta_k\| \\
&< \eta \cdot \frac{\varepsilon}{\eta} \\
&= \varepsilon
\end{aligned}$$

また，各 $k = 1, \ldots, K$ について，$f(X_1;\theta_k), \ldots, f(X_N;\theta_k)$ は有限集合 $\{f(x;\theta_k) \mid x \in \mathcal{X}\}$ に値をとる独立な確率変数列であるから，大数の強法則より，

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} f(X_n;\theta_k) = \sum_{x \in \mathcal{X}} p(x) f(x;\theta_k) \quad \text{a.s.}$$

が成り立つ．

各 $k = 1, \ldots, K$ について，任意の $\theta \in D \cap \overline{B_{\varepsilon/\eta}(\theta_k)}$ に対して，

$$\begin{aligned}
&\left\| \frac{1}{N} \sum_{n=1}^{N} f(X_n;\theta) - \sum_{x \in \mathcal{X}} p(x) f(x;\theta) \right\| \\
\leq &\left\| \frac{1}{N} \sum_{n=1}^{N} f(X_n;\theta) - \frac{1}{N} \sum_{n=1}^{N} f(X_n;\theta_k) \right\| \\
&\qquad + \left\| \frac{1}{N} \sum_{n=1}^{N} f(X_n;\theta_k) - \sum_{x \in \mathcal{X}} p(x) f(x;\theta_k) \right\| \\
&\qquad\qquad + \left\| \sum_{x \in \mathcal{X}} p(x) f(x;\theta_k) - \sum_{x \in \mathcal{X}} p(x) f(x;\theta) \right\| \\
\leq &\frac{1}{N} \sum_{n=1}^{N} \|f(X_n;\theta) - f(X_n;\theta_k)\| \\
&\qquad + \left\| \frac{1}{N} \sum_{n=1}^{N} f(X_n;\theta_k) - \sum_{x \in \mathcal{X}} p(x) f(x;\theta_k) \right\| \\
&\qquad\qquad + \sum_{x \in \mathcal{X}} p(x) \|f(x;\theta_k) - f(x;\theta)\| \\
< &\, 2\varepsilon + \left\| \frac{1}{N} \sum_{n=1}^{N} f(X_n;\theta_k) - \sum_{x \in \mathcal{X}} p(x) f(x;\theta_k) \right\|
\end{aligned}$$

であることを用いると，

$$
\begin{aligned}
&\limsup_{N\to\infty}\sup_{\theta\in D}\left\|\frac{1}{N}\sum_{n=1}^{N}f(X_n;\theta)-\sum_{x\in\mathcal{X}}p(x)f(x;\theta)\right\|\\
&\le 2\varepsilon+\limsup_{N\to\infty}\max_{1\le k\le K}\left\|\frac{1}{N}\sum_{n=1}^{N}f(X_n;\theta_k)-\sum_{x\in\mathcal{X}}p(x)f(x;\theta_k)\right\|\\
&\le 2\varepsilon+\sum_{k=1}^{K}\limsup_{N\to\infty}\left\|\frac{1}{N}\sum_{n=1}^{N}f(X_n;\theta_k)-\sum_{x\in\mathcal{X}}p(x)f(x;\theta_k)\right\|\\
&= 2\varepsilon \quad \text{a.s.}
\end{aligned}
$$

となるから，任意の $\varepsilon>0$ に対して，

$$
P\left(\limsup_{N\to\infty}\sup_{\theta\in D}\left\|\frac{1}{N}\sum_{n=1}^{N}f(X_n;\theta)-\sum_{x\in\mathcal{X}}p(x)f(x;\theta)\right\|\le 2\varepsilon\right)=1
$$

が成り立つ．

各 $m\in\mathbb{N}$ に対して，

$$
\mathcal{A}_m=\left\{\limsup_{N\to\infty}\sup_{\theta\in D}\left\|\frac{1}{N}\sum_{n=1}^{N}f(X_n;\theta)-\sum_{x\in\mathcal{X}}p(x)f(x;\theta)\right\|\le\frac{1}{m}\right\}
$$

とおくと，上の事実より $P(\mathcal{A}_m^c)=0$ が成り立ち，

$$
\bigcap_{m=1}^{\infty}\mathcal{A}_m=\left\{\lim_{N\to\infty}\sup_{\theta\in D}\left\|\frac{1}{N}\sum_{n=1}^{N}f(X_n;\theta)-\sum_{x\in\mathcal{X}}p(x)f(x;\theta)\right\|=0\right\}
$$

である．$\mathcal{B}_1=\mathcal{A}_1^c$，$\mathcal{B}_m=\mathcal{A}_m^c\setminus\mathcal{A}_{m-1}^c$ $(m\ge 2)$ とおくと，

$$
P\left(\bigcup_{m=1}^{\infty}\mathcal{A}_m^c\right)=P\left(\bigcup_{m=1}^{\infty}\mathcal{B}_m\right)=\sum_{m=1}^{\infty}P\left(\mathcal{B}_m\right)\le\sum_{m=1}^{\infty}P\left(\mathcal{A}_m^c\right)=0
$$

が成り立つ．したがって，

$$
\begin{aligned}
&P\left(\lim_{N\to\infty}\sup_{\theta\in D}\left\|\frac{1}{N}\sum_{n=1}^{N}f(X_n;\theta)-\sum_{x\in\mathcal{X}}p(x)f(x;\theta)\right\|=0\right)\\
=&P\left(\bigcap_{m=1}^{\infty}\mathcal{A}_m\right)\\
&=1
\end{aligned}
$$

が成り立つ． □

命題 2.7. 命題 2.4 でとった $\delta>0$ に対して，ψ_N が $\overline{B_\delta(\theta_*)}$ に唯一の零点をもつ確率は，$N\to\infty$ のとき 1 に収束する．

証明． まず，

$$\varphi_N(\theta) := \theta - \left(\frac{\partial \psi_*}{\partial \theta}(\theta_*)\right)^{-1} \psi_N(\theta)$$

とおくと，各 $\theta \in \Theta$ に対し，$\psi_N(\theta) = 0$ と $\varphi_N(\theta) = \theta$ は同値である．そして，$\delta > 0$ は，任意の $\theta \in \overline{B_\delta(\theta_*)}$ に対して，

$$\left\| \frac{\partial \psi_*}{\partial \theta}(\theta) - \frac{\partial \psi_*}{\partial \theta}(\theta_*) \right\| \leq \frac{1}{4} \left\| \left(\frac{\partial \psi_*}{\partial \theta}(\theta_*)\right)^{-1} \right\|^{-1}$$

が成り立つように選んだことに注意すると，

$$\begin{aligned}
\left\| \frac{\partial \varphi_N}{\partial \theta}(\theta) \right\| &= \left\| I - \left(\frac{\partial \psi_*}{\partial \theta}(\theta_*)\right)^{-1} \frac{\partial \psi_N}{\partial \theta}(\theta) \right\| \\
&= \left\| \left(\frac{\partial \psi_*}{\partial \theta}(\theta_*)\right)^{-1} \left(\frac{\partial \psi_*}{\partial \theta}(\theta_*) - \frac{\partial \psi_N}{\partial \theta}(\theta)\right) \right\| \\
&\leq \left\| \left(\frac{\partial \psi_*}{\partial \theta}(\theta_*)\right)^{-1} \right\| \left\| \frac{\partial \psi_*}{\partial \theta}(\theta_*) - \frac{\partial \psi_N}{\partial \theta}(\theta) \right\| \\
&\leq \left\| \left(\frac{\partial \psi_*}{\partial \theta}(\theta_*)\right)^{-1} \right\| \left(\left\| \frac{\partial \psi_*}{\partial \theta}(\theta_*) - \frac{\partial \psi_*}{\partial \theta}(\theta) \right\| + \left\| \frac{\partial \psi_*}{\partial \theta}(\theta) - \frac{\partial \psi_N}{\partial \theta}(\theta) \right\| \right) \\
&\leq \frac{1}{4} + \left\| \left(\frac{\partial \psi_*}{\partial \theta}(\theta_*)\right)^{-1} \right\| \left\| \frac{\partial \psi_*}{\partial \theta}(\theta) - \frac{\partial \psi_N}{\partial \theta}(\theta) \right\| \qquad (2.1)
\end{aligned}$$

一方，任意の $\theta \in \overline{B_\delta(\theta_*)}$ に対して，

$$\|\varphi_*(\theta) - \theta_*\| = \|\varphi_*(\theta) - \varphi_*(\theta_*)\| \leq \frac{1}{4} \|\theta - \theta_*\| \leq \frac{\delta}{4}$$

であったことを思い出すと，

$$\begin{aligned}
\|\varphi_N(\theta) - \theta_*\| &\leq \|\varphi_N(\theta) - \varphi_*(\theta)\| + \|\varphi_*(\theta) - \varphi_*(\theta_*)\| \\
&\leq \left\| \left(\frac{\partial \psi_*}{\partial \theta}\right)^{-1} \right\| \|\psi_N(\theta) - \psi_*(\theta)\| + \frac{\delta}{4} \qquad (2.2)
\end{aligned}$$

が成り立つ．

さらに，一様大数の強法則 (補題 2.6) より，

$$\begin{aligned}
&\lim_{N\to\infty} \sup_{\theta \in \overline{B_\delta(\theta_*)}} \left\| \frac{\partial \psi_N}{\partial \theta}(\theta) - \frac{\partial \psi_*}{\partial \theta}(\theta) \right\| \\
&= \lim_{N\to\infty} \sup_{\theta \in \overline{B_\delta(\theta_*)}} \left\| \frac{1}{N} \sum_{n=1}^{N} \frac{\partial u}{\partial \theta}(X_n; \theta) - \sum_{x \in \mathcal{X}} p(x; \theta_*) \frac{\partial u}{\partial \theta}(x; \theta) \right\| \\
&= 0 \quad \text{a.s.}
\end{aligned}$$

および

$$\begin{aligned}
&\lim_{N\to\infty} \sup_{\theta\in\overline{B_\delta(\theta_*)}} \|\psi_N(\theta)-\psi_*(\theta)\| \\
&= \lim_{N\to\infty} \sup_{\theta\in\overline{B_\delta(\theta_*)}} \left\| \frac{1}{N}\sum_{n=1}^{N} u(X_n;\theta) - \sum_{x\in\mathcal{X}} p(x;\theta_*)u(x;\theta) \right\| \\
&= 0 \quad \text{a.s.}
\end{aligned}$$

が成り立つので，特に，任意の $\varepsilon>0$ に対して

$$\lim_{N\to\infty} P\left(\sup_{\theta\in\overline{B_\delta(\theta_*)}} \left\| \frac{\partial\psi_N}{\partial\theta}(\theta) - \frac{\partial\psi_*}{\partial\theta}(\theta) \right\| < \varepsilon \right) = 1 \tag{2.3}$$

および

$$\lim_{N\to\infty} P\left(\sup_{\theta\in\overline{B_\delta(\theta_*)}} \|\psi_N(\theta)-\psi_*(\theta)\| < \varepsilon \right) = 1 \tag{2.4}$$

を得る．

ここで，

$$\begin{aligned}
&P\left(\psi_N \text{が} \overline{B_\delta(\theta_*)} \text{に唯一の零点をもつ}\right) \\
&= P\left(\varphi_N \text{が} \overline{B_\delta(\theta_*)} \text{に唯一の不動点をもつ}\right) \\
&\geq P\left(\varphi_N \text{が} \overline{B_\delta(\theta_*)} \text{上の縮小写像} \wedge \varphi_N\left(\overline{B_\delta(\theta_*)}\right) \subset \overline{B_\delta(\theta_*)}\right) \\
&\geq P\left(\sup_{\theta\in\overline{B_\delta(\theta_*)}} \left\| \frac{\partial\varphi_N}{\partial\theta}(\theta) \right\| \leq \frac{1}{2} \wedge \sup_{\theta\in\overline{B_\delta(\theta_*)}} \|\varphi_N(\theta)-\theta_*\| \leq \delta \right) \\
&\geq P\left(\mathcal{A}_N \cap \mathcal{B}_N\right)
\end{aligned}$$

が成り立つことに注意する．ここに，(2.1) および (2.2) を念頭に

$$\begin{aligned}
\mathcal{A}_N &:= \left\{ \sup_{\theta\in\overline{B_\delta(\theta_*)}} \left\| \frac{\partial\psi_N}{\partial\theta}(\theta) - \frac{\partial\psi_*}{\partial\theta}(\theta) \right\| < \frac{1}{4} \left\| \left(\frac{\partial\psi_*}{\partial\theta}(\theta_*) \right)^{-1} \right\|^{-1} \right\} \\
\mathcal{B}_N &:= \left\{ \sup_{\theta\in\overline{B_\delta(\theta_*)}} \|\psi_N(\theta)-\psi_*(\theta)\| < \frac{\delta}{4} \left\| \left(\frac{\partial\psi_*}{\partial\theta}(\theta_*) \right)^{-1} \right\|^{-1} \right\}
\end{aligned}$$

とおいた．すると，(2.3) (2.4) より

$$\begin{aligned}
1 - P(\mathcal{A}_N \cap \mathcal{B}_N) &= P(\mathcal{A}_N^{\mathrm{c}} \cup \mathcal{B}_N^{\mathrm{c}}) \\
&\leq P(\mathcal{A}_N^{\mathrm{c}}) + P(\mathcal{B}_N^{\mathrm{c}}) \\
&\to 0 \qquad (N\to\infty)
\end{aligned}$$

が成り立つから，

$$\lim_{N\to\infty} P\left(\psi_N \text{が} \overline{B_\delta(\theta_*)} \text{に唯一の零点をもつ}\right) = 1$$

を得る． □

以下，命題 2.7 の $\delta > 0$ を固定する．そして，$\hat{\theta}_N, \theta_*$ が $\overline{B_\delta(\theta_*)}$ 上でそれぞれ ψ_N, ψ_* の唯一の零点であると仮定し，Z 推定量の強一致性と漸近正規性を証明する．

命題 2.8 (Z 推定量の強一致性)**.** $\hat{\theta}_N$ は，$N \to \infty$ のとき θ_* に概収束する．

証明. $\varepsilon \in (0, \delta)$ を任意にとり，

$$\gamma := \min \left\{ \|\psi_*(\theta)\| \,\middle|\, \theta \in \overline{B_\delta(\theta_*)} \wedge \theta \notin B_\varepsilon(\theta_*) \right\}$$

とおくと，θ_* が $\overline{B_\delta(\theta_*)}$ 上で ψ_* の唯一の零点であることから，$\gamma > 0$ であり，任意の $\theta \in \overline{B_\delta(\theta_*)}$ に対して，$\|\theta - \theta_*\| \geq \varepsilon$ ならば $\|\psi_*(\theta)\| \geq \gamma$ となる．したがって，任意の $\theta \in \overline{B_\delta(\theta_*)}$ に対して，$\|\psi_*(\theta)\| < \gamma$ ならば $\|\theta - \theta_*\| < \varepsilon$ が成り立つことに注意すれば，

$$\begin{aligned}
&P\left(\limsup_{N\to\infty} \|\hat{\theta}_N - \theta_*\| < \varepsilon\right) \\
&= P\left(\bigcup_{N_0 \in \mathbb{N}} \bigcap_{N \geq N_0} \left\{\|\hat{\theta}_N - \theta_*\| < \varepsilon\right\}\right) \\
&\geq P\left(\bigcup_{N_0 \in \mathbb{N}} \bigcap_{N \geq N_0} \left\{\|\psi_*(\hat{\theta}_N)\| < \gamma\right\}\right) \\
&= P\left(\bigcup_{N_0 \in \mathbb{N}} \bigcap_{N \geq N_0} \left\{\|\psi_*(\hat{\theta}_N) - \psi_N(\hat{\theta}_N)\| < \gamma\right\}\right) \\
&\geq P\left(\bigcup_{N_0 \in \mathbb{N}} \bigcap_{N \geq N_0} \left\{\sup_{\theta \in \overline{B_\delta(\theta_*)}} \|\psi_*(\theta) - \psi_N(\theta)\| < \gamma\right\}\right) \\
&= P\left(\limsup_{N\to\infty} \sup_{\theta \in \overline{B_\delta(\theta_*)}} \|\psi_*(\theta) - \psi_N(\theta)\| < \gamma\right) \\
&= P\left(\limsup_{N\to\infty} \sup_{\theta \in \overline{B_\delta(\theta_*)}} \left\| \frac{1}{N}\sum_{n=1}^{N} u(X_n;\theta) - \sum_{x\in\mathcal{X}} p(x;\theta_*)u(x;\theta) \right\| < \gamma\right) \\
&= 1
\end{aligned}$$

を得る．ここで，最後の等号では一様大数の強法則 (補題 2.6) を用いた．したがって，

$$P\left(\lim_{N\to\infty} \|\hat{\theta}_N - \theta_*\| = 0\right) = 1$$

が成り立つ． □

漸近正規性の証明のために，4 つの補題を用意する．

補題 2.9 ([8, pp.6-7])**.** E をユークリッド空間の開部分集合，α_N, α を E に値を取る確率変数とする．α_N が α に分布収束することと，E の任意の閉集合 F に対して，

$$\limsup_{N\to\infty} P(\alpha_N \in F) \leq P(\alpha \in F)$$

が成り立つことは同値である．

以下では，確率変数 α_N が α に概収束，確率収束，分布収束することをそれぞれ，

$$\alpha_N \to \alpha \quad \text{a.s.} \qquad \alpha_N \xrightarrow{p} \alpha \qquad \alpha_N \xrightarrow{d} \alpha$$

で表す．

補題 2.10 (連続写像定理 [8, pp.7-8])**.** E をユークリッド空間の開部分集合，$f : E \to \mathbb{R}^d$ を連続写像とする．α_N, α を E に値を取る確率変数とすると，$N \to \infty$ のとき以下が成り立つ．

1. $\alpha_N \to \alpha \quad \text{a.s.}$ ならば，$f(\alpha_N) \to f(\alpha) \quad \text{a.s.}$
2. $\alpha_N \xrightarrow{p} \alpha$ ならば，$f(\alpha_N) \xrightarrow{p} f(\alpha)$
3. $\alpha_N \xrightarrow{d} \alpha$ ならば，$f(\alpha_N) \xrightarrow{d} f(\alpha)$

補題 2.11 ([10, p.58 定理 1.57])**.** $\alpha_N, \alpha \in \mathbb{R}^d$, $\Gamma_N \in \mathbb{R}^{d\times d}$ を確率変数，$\Gamma \in \mathbb{R}^{d\times d}$ を定数行列とする．$N \to \infty$ のとき，$\alpha_N \xrightarrow{d} \alpha$ かつ $\Gamma_N \xrightarrow{p} \Gamma$ ならば，$(\alpha_N, \Gamma_N) \xrightarrow{d} (\alpha, \Gamma)$ が成り立つ．

補題 2.12. 補題 2.11 の仮定のもとで，Γ が正則行列であり，確率変数 $\beta_N \in \mathbb{R}^d$ が任意の $N \in \mathbb{N}$ に対して，$\Gamma_N \beta_N = \alpha_N$ を満たすならば，$N \to \infty$ のとき $\beta_N \xrightarrow{d} \Gamma^{-1}\alpha$ が成り立つ．

証明. 正則行列全体の集合は，$\mathbb{R}^{d\times d}$ において，ノルム $\|A\| := \sqrt{\mathrm{tr} A^{\mathrm{T}} A}$ に関して開集合であることから，ある $\gamma > 0$ が存在して，$\|A - \Gamma\| < \gamma$ ならば A が正則となる，

$\Gamma_N \xrightarrow{p} \Gamma$ より，

$$\begin{aligned}
&\limsup_{N\to\infty} P(\Gamma_N \text{が正則でない}) \\
\leq &\limsup_{N\to\infty} P(\|\Gamma_N - \Gamma\| > \gamma) \\
= &0
\end{aligned}$$

が成り立つ．

$$\widetilde{\Gamma}_N := \begin{cases} \Gamma_N & (\Gamma_N \text{が正則のとき}) \\ I & (\Gamma_N \text{が正則でないとき}) \end{cases}$$

とおくと，$\widetilde{\Gamma}_N$ は正則である．

任意の $\varepsilon > 0$ に対して，

$$\begin{aligned}
&\limsup_{N\to\infty} P(\|\widetilde{\Gamma}_N - \Gamma\| > \varepsilon) \\
\leq &\limsup_{N\to\infty} P(\|\Gamma_N - \Gamma\| > \varepsilon \ \wedge \ \Gamma_N \text{が正則}) + \limsup_{N\to\infty} P(\Gamma_N \text{が正則でない}) \\
\leq &\limsup_{N\to\infty} P(\|\Gamma_N - \Gamma\| > \varepsilon) + \limsup_{N\to\infty} P(\Gamma_N \text{が正則でない}) \\
= &\ 0
\end{aligned}$$

より，$\widetilde{\Gamma}_N \xrightarrow{p} \Gamma$ であるから，$\alpha_N \xrightarrow{d} \alpha$ と補題 2.11 より，$(\widetilde{\Gamma}_N, \alpha_N) \xrightarrow{d} (\Gamma, \alpha)$ が成り立つ．さらに，E を正則行列全体として，逆行列をかける写像 $E \times \mathbb{R}^d \to \mathbb{R}^d$ は連続であるから，連続写像定理 (補題 2.10) より，$\widetilde{\Gamma}_N^{-1}\alpha_N \xrightarrow{d} \Gamma^{-1}\alpha$ が成り立つ．

したがって，任意の閉集合 $F \subset \mathbb{R}^d$ に対して，補題 2.9 より，

$$\begin{aligned}
&\limsup_{N\to\infty} P(\beta_N \in F) \\
\le&\limsup_{N\to\infty} P(\beta_N \in F \ \wedge\ \Gamma_N\text{が正則}) + \limsup_{N\to\infty} P(\Gamma_N\text{が正則でない}) \\
=&\limsup_{N\to\infty} P(\widetilde{\Gamma}_N^{-1}\alpha_N \in F) + \limsup_{N\to\infty} P(\Gamma_N\text{が正則でない}) \\
\le&P(\Gamma^{-1}\alpha \in F)
\end{aligned}$$

が成り立つから，$N \to \infty$ のとき $\beta_N \xrightarrow{d} \Gamma^{-1}\alpha$ が成り立つ．

命題 2.13 (Z 推定量の漸近正規性)**.** $\sqrt{N}(\hat{\theta}_N - \theta_*)$ は d 次元正規分布 $\mathcal{N}(0, A(\theta_*)^{-1}V(\theta_*)A(\theta_*)^{-\mathrm{T}})$ に分布収束する．ただし，

$$\begin{aligned}
A(\theta_*) &= -\sum_{x\in\mathcal{X}} p(x;\theta_*)\frac{\partial u}{\partial\theta}(x;\theta_*) \\
V(\theta_*) &= \sum_{x\in\mathcal{X}} p(x;\theta_*)u(x;\theta_*)u(x;\theta_*)^{\mathrm{T}}
\end{aligned}$$

である．

証明. 任意の $N \in \mathbb{N}$ に対して，

$$\begin{aligned}
-\psi_N(\theta_*) &= \psi_N(\hat{\theta}_N) - \psi_N(\theta_*) \\
&= \int_0^1 \frac{\mathrm{d}}{\mathrm{d}t}\psi_N(t\hat{\theta}_N + (1-t)\theta_*)\mathrm{d}t \\
&= \left(\int_0^1 \frac{\partial\psi_N}{\partial\theta}(t\hat{\theta}_N + (1-t)\theta_*)\mathrm{d}t\right)\left(\hat{\theta}_N - \theta_*\right)
\end{aligned}$$

両辺に $\sqrt{N}$ をかけて，

$$\left(\int_0^1 \frac{\partial\psi_N}{\partial\theta}(t\hat{\theta}_N + (1-t)\theta_*)\mathrm{d}t\right)\left(\sqrt{N}(\hat{\theta}_N - \theta_*)\right) = -\sqrt{N}\psi_N(\theta_*) \tag{2.5}$$

を得る．

$$\begin{aligned}
&\left\|\int_0^1 \frac{\partial\psi_N}{\partial\theta}(t\hat{\theta}_N + (1-t)\theta_*)\mathrm{d}t - \frac{\partial\psi_*}{\partial\theta}(\theta_*)\right\| \\
\le&\int_0^1 \left\|\frac{\partial\psi_N}{\partial\theta}(t\hat{\theta}_N + (1-t)\theta_*) - \frac{\partial\psi_*}{\partial\theta}(\theta_*)\right\|\mathrm{d}t \\
\le&\int_0^1 \left\|\frac{\partial\psi_N}{\partial\theta}(t\hat{\theta}_N + (1-t)\theta_*) - \frac{\partial\psi_*}{\partial\theta}(t\hat{\theta}_N + (1-t)\theta_*)\right\|\mathrm{d}t \\
&\qquad + \int_0^1 \left\|\frac{\partial\psi_*}{\partial\theta}(t\hat{\theta}_N + (1-t)\theta_*) - \frac{\partial\psi_*}{\partial\theta}(\theta_*)\right\|\mathrm{d}t
\end{aligned}$$

$$\le \sup_{\theta\in\overline{B_\delta(\theta_*)}} \left\| \frac{\partial\psi_N}{\partial\theta}(\theta) - \frac{\partial\psi_*}{\partial\theta}(\theta) \right\| + \int_0^1 \left\| \frac{\partial\psi_*}{\partial\theta}(t\hat{\theta}_N + (1-t)\theta_*) - \frac{\partial\psi_*}{\partial\theta}(\theta_*) \right\| \mathrm{d}t$$

一様大数の強法則 (補題 2.6) より，

$$\lim_{N\to\infty} \sup_{\theta\in\overline{B_\delta(\theta_*)}} \left\| \frac{\partial\psi_N}{\partial\theta}(\theta) - \frac{\partial\psi_*}{\partial\theta}(\theta) \right\| = 0 \quad \text{a.s.}$$

また，

$$\theta \mapsto \int_0^1 \left\| \frac{\partial\psi_*}{\partial\theta}(t\theta + (1-t)\theta_*) - \frac{\partial\psi_*}{\partial\theta}(\theta_*) \right\| \mathrm{d}t$$

は連続であり，Z 推定量の強一致性 (命題 2.8) より，

$$\lim_{N\to\infty} \hat{\theta}_N = \theta_* \quad \text{a.s.}$$

が成り立つから，連続写像定理 (補題 2.10) より，

$$\lim_{N\to\infty} \int_0^1 \left\| \frac{\partial\psi_*}{\partial\theta}(t\hat{\theta}_N + (1-t)\theta_*) - \frac{\partial\psi_*}{\partial\theta}(\theta_*) \right\| \mathrm{d}t = 0 \quad \text{a.s.}$$

よって，

$$\lim_{N\to\infty} \int_0^1 \frac{\partial\psi_N}{\partial\theta}(t\hat{\theta}_N + (1-t)\theta_*)\mathrm{d}t = \frac{\partial\psi_*}{\partial\theta}(\theta_*) \qquad \text{a.s.}$$

特に，

$$\int_0^1 \frac{\partial\psi_N}{\partial\theta}(t\hat{\theta}_N + (1-t)\theta_*)\mathrm{d}t \xrightarrow{p} \frac{\partial\psi_*}{\partial\theta}(\theta_*) \quad (N\to\infty)$$

を用いる．

一方，中心極限定理より，$\mathcal{N}(0, V(\theta_*))$ に従うある確率変数 Z に対して，

$$-\sqrt{N}\psi_N(\theta_*) = -\frac{1}{\sqrt{N}} \sum_{n=1}^N u(X_n;\theta_*) \xrightarrow{d} Z$$

が成り立つ．

式 (2.5) と補題 2.12 より，

$$\sqrt{N}(\hat{\theta}_N - \theta_*) \xrightarrow{d} \left(\frac{\partial\psi_*}{\partial\theta}(\theta_*) \right)^{-1} Z = -A(\theta_*)^{-1}Z$$

ここで，$-A(\theta_*)^{-1}Z$ は $\mathcal{N}(0, A(\theta_*)^{-1}V(\theta_*)A(\theta_*)^{-T})$ に従うので，$\sqrt{N}(\hat{\theta}_N - \theta_*)$ は $\mathcal{N}(0, A(\theta_*)^{-1}V(\theta_*)A(\theta_*)^{-\mathrm{T}})$ に分布収束する． □

2.2 力学系の解の安定性

$f : \mathbb{R}^d \to \mathbb{R}^d$ を C^1 級写像とし，$\dot{\theta} = f(\theta)$ の形で表せる力学系について考える．ここに，$\dot{\theta}$ は時刻 t に関する θ の微分である．$f(\theta_*) = 0$ を満たす θ_* を力学系 $\dot{\theta} = f(\theta)$ の平衡点という．$\theta(t) \equiv \theta_*$（定値関数）は $\dot{\theta} = f(\theta)$ の解の 1 つである．

定義 2.14. 平衡点 θ_* が局所安定であるとは，任意の $\varepsilon > 0$ に対してある $\delta > 0$ が存在し，任意の解 $\theta(t)$ に対して，$\|\theta(0) - \theta_*\| < \delta$ ならば，任意の $t \geq 0$ に対して $\|\theta(t) - \theta_*\| < \varepsilon$ が成り立つことをいう．

定義 2.15. 平衡点 θ_* が局所漸近安定であるとは，それが局所安定であり，かつ，ある $\delta > 0$ が存在して，任意の解 $\theta(t)$ に対して，$\|\theta(0) - \theta_*\| < \delta$ ならば，$\lim_{t\to\infty} \theta(t) = \theta_*$ が成り立つことをいう．

定義 2.16. 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う．

次の命題は，与えられた平衡点 θ_* が局所漸近安定となるための十分条件を与えるものである．

命題 2.17. [9, p.11] θ_* を力学系 $\dot{\theta} = f(\theta)$ の平衡点とする．$\dfrac{\partial f}{\partial \theta}(\theta_*)$ が安定行列ならば，θ_* は局所漸近安定である．

2.3 マルコフ連鎖モンテカルロ法

$p : \mathcal{X} \to \mathbb{R}$ が $\mathcal{X}$ 上の確率分布であるとは，任意の $x \in \mathcal{X}$ に対して $p(x) > 0$ かつ，$\sum_{x\in\mathcal{X}} p(x) = 1$ が成り立つことをいうものとする．

$\mathcal{X}$ 上の確率分布 p と $f : \mathcal{X} \to \mathbb{R}$ に対して，f の p に関する期待値を計算するとき，p に従う独立な確率変数列 $X_1, \ldots, X_N$ を用いて

$$\frac{1}{N}\sum_{n=1}^{N} f(X_n) \approx \sum_{x\in\mathcal{X}} p(x)f(x)$$

と近似的に計算する方法をモンテカルロ法といい，N が十分大きいときには大数の法則によって近似の精度が保証される．p に従う確率変数を直接生成することが困難な場合に，p に分布収束するマルコフ連鎖を構成し，ランダムに選んだ初期状態を十分な回数遷移させた状態を p に従う確率変数とみなしてモンテカルロ法を用いる方法を，マルコフ連鎖モンテカルロ法という．

各 $y \in \mathcal{X}$ に対して，$T(\,\cdot\,|y)$ が $\mathcal{X}$ 上の確率分布であるとき，T を遷移核といい，p に対し，

$$p(x)T(y|x) = p(y)T(x|y)$$

が任意の $x, y \in \mathcal{X}$ で成り立つとき，T は p に対して詳細つりあい条件を満たすという．

確率分布 q に対し，確率分布 $\sum_{y\in\mathcal{X}} T(\,\cdot\,|y)q(y)$ を対応させる作用素を T によって生成されるマルコフ作用素といい，Φ_T で表す．すなわち

$$(\Phi_T\, q)(x) := \sum_{y\in\mathcal{X}} T(x|y)q(y)$$

確率変数列 $Y_0, Y_1, \ldots$ について，各 k で Y_{k+1} が $T(\,\cdot\,|Y_k)$ に従うとき，$Y_0, Y_1, \ldots$ を T によって生成されるマルコフ連鎖という．

詳細つりあい条件の式

$$p(x)T(y|x) = p(y)T(x|y)$$

の両辺で y について和をとると，

$$p(x) = \sum_{y\in\mathcal{X}} T(x|y)p(y)$$

すなわち，$p = \Phi_T\, p$ を得る．このとき，p は Φ_T に関する不変分布であるという．次の定理はマルコフ連鎖が不変分布に収束することを示している．

命題 2.18. Φ_T が T によって生成されるマルコフ作用素で，p がその不変分布であるならば，任意の確率分布 q と $x \in \mathcal{X}$ に対して，

$$\lim_{k\to\infty} (\Phi_T^k\, q)(x) = p(x)$$

が成り立つ．

証明. $q_k := \Phi_T^k\, q$ とおく．また，定数 $c \in (0,1)$ を，任意の $x, y \in \mathcal{X}$ に対して，$(1-c)p(x) < T(x|y)$ を満たすようにとる．

$$q_1(x) = \sum_{y\in\mathcal{X}} T(x|y)q(y) > (1-c)p(x)$$

より，

$$r_1(x) := \frac{1}{c}\Big(q_1(x) - (1-c)p(x)\Big)$$

とおくと，r_1 は確率分布で，

$$q_1(x) = (1-c)p(x) + cr_1(x) \tag{2.6}$$

と表せる．同様に，ある確率分布 r_2 が存在して，

$$(\Phi_T\, r_1)(x) = (1-c)p(x) + cr_2(x) \tag{2.7}$$

と表せる．(2.6) の両辺に Φ_T を作用させて，

$$q_2(x) = (1-c)p(x) + c(\Phi_T\, r_1)(x)$$

(2.7) を代入して，

$$q_2(x) = (1-c)(1+c)p(x) + c^2 r_2(x)$$

以上の操作を繰り返して，

$$\begin{aligned} q_k(x) &= (1-c)(1+c+\cdots+c^{k-1})p(x) + c^k r_k(x) \\ &= (1-c^k)p(x) + c^k r_k(x) \\ &\to p(x) \qquad (k\to\infty) \end{aligned}$$

□

与えられた確率分布 p から，詳細つりあい条件を満たす遷移核 T を具体的に構成する方法の一例を挙げる．

例 2.19 (ギブス・サンプリング)**.** $\mathcal{X} = \{0,1\}^2$ とする．

$$
\begin{aligned}
T(x|y) :=& \frac{1}{2}\Big(\frac{p(x_1,x_2)}{p(0,x_2)+p(1,x_2)} \cdot \frac{p(y_1,x_2)}{p(y_1,0)+p(y_1,1)} \\
&\qquad + \frac{p(x_1,x_2)}{p(x_1,0)+p(x_1,1)} \cdot \frac{p(x_1,y_2)}{p(0,y_2)+p(1,y_2)}\Big)
\end{aligned}
$$

とおくと，

$$
\begin{aligned}
&T(x|y)p(y) \\
=&\frac{1}{2}\Big(\frac{p(x_1,x_2)}{p(0,x_2)+p(1,x_2)} \cdot \frac{p(y_1,x_2)}{p(y_1,0)+p(y_2,1)} \cdot p(y_1,y_2) \\
&\qquad + \frac{p(x_1,x_2)}{p(x_1,0)+p(x_1,1)} \cdot \frac{p(x_1,y_2)}{p(0,y_2)+p(1,y_2)} \cdot p(y_1,y_2)\Big) \\
=&\frac{1}{2}\Big(p(x_1,x_2) \cdot \frac{p(y_1,x_2)}{p(0,x_2)+p(1,x_2)} \cdot \frac{p(y_1,y_2)}{p(y_1,0)+p(y_2,1)} \\
&\qquad + p(x_1,x_2) \cdot \frac{p(x_1,y_2)}{p(x_1,0)+p(x_1,1)} \cdot \frac{p(y_1,y_2)}{p(0,y_2)+p(1,y_2)}\Big) \\
=&p(x)T(y|x)
\end{aligned}
$$

を得る．

2.4 捩れをもつ統計多様体

本節は，参考文献 [1] のレビューである．多様体とその上の各量はすべて C^∞ 級を仮定する．(M,g) を Riemann 多様体，∇ を M 上のアファイン接続，$\mathfrak{X}(M)$ を M 上のベクトル場全体とする．任意の $X,Y,Z \in \mathfrak{X}(M)$ に対して

$$
Xg(Y,Z) = g(\nabla_X Y, Z) + g(Y, \nabla^*_X Z)
$$

を満たすアファイン接続 ∇^* が一意的に定まり，これを g に関する ∇ の双対接続という．また，4 つ組 (M,g,∇,∇^*) を統計多様体という．∇ の曲率 R と捩率 T をそれぞれ

$$
\begin{aligned}
R(X,Y,Z) &:= \nabla_X\nabla_Y Z - \nabla_Y\nabla_X Z - \nabla_{[X,Y]}Z \\
T(X,Y) &:= \nabla_X Y - \nabla_Y X - [X,Y]
\end{aligned}
$$

で定義し，恒等的に $T=0$ のとき，∇ には捩れがないという．∇^* の曲率 R^* と捩率 T^* も同様に定義される．$R=0 \Leftrightarrow R^*=0$ は成り立つが，$T=0$ でも $T^*=0$ とは限らない．R,R^*,T,T^* がすべて 0 となる場合，(M,g,∇,∇^*) を双対平坦多様体という．$p \in M$ のまわりで座標系 $(\theta^1,\ldots,\theta^d)$ をとり，∇,∇^* に対する接続係数をそれぞれ ${\Gamma_{ij}}^k, {\Gamma^*_{ij}}^k$ とし，

$$
\begin{aligned}
\Gamma_{ij,k} &:= {\Gamma_{ij}}^{\ell} g_{\ell k} \\
\Gamma^*_{ij,k} &:= {\Gamma^*_{ij}}^{\ell} g_{\ell k}
\end{aligned}
$$

とする．

$$
\begin{aligned}
{R_{ijk}}^{\ell}\partial_\ell &:= R(\partial_i, \partial_j, \partial_k) \\
{T_{ij}}^{k}\partial_k &:= T(\partial_i, \partial_j)
\end{aligned}
$$

で ${R_{ijk}}^{\ell}, {T_{ij}}^{k}$ を定義すると，

$$
\begin{aligned}
{R_{ijk}}^{\ell} &= \partial_i {\Gamma_{jk}}^{\ell} - \partial_j {\Gamma_{ik}}^{\ell} + {\Gamma_{jk}}^{m}{\Gamma_{im}}^{\ell} - {\Gamma_{ik}}^{m}{\Gamma_{jm}}^{\ell} \\
{T_{ij}}^{k} &= {\Gamma_{ij}}^{k} - {\Gamma_{ji}}^{k}
\end{aligned}
$$

と表されるので，∇ に捩率がないことは ${\Gamma_{ij}}^{k} = {\Gamma_{ji}}^{k}$ と同値である．このことは座標系のとり方によらない．

定義 2.20 (コントラスト関数)**.** $\phi \in C^\infty(M \times M)$ と $X_1, \ldots, X_m, Y_1, \ldots, Y_n \in \mathfrak{X}(M)$ に対して，$\phi(X_1 \cdots X_m | Y_1 \cdots Y_n) \in C^\infty(M)$ を

$$
\phi(X_1 \cdots X_m | Y_1 \cdots Y_n)(r) := (X_1)_p \cdots (X_m)_p (Y_1)_q \cdots (Y_n)_q \phi(p, q)|_{p=r, q=r}
$$

で定義する．$\phi \in C^\infty(M \times M)$ がコントラスト関数であるとは，

1. 任意の $p, q \in M$ に対して，$\phi(p, q) \geq 0$ かつ，$\phi(p, q) = 0$ ならば $p = q$
2. 任意の $X \in \mathfrak{X}(M)$ に対して，$\phi(X|X) = 0$ ならば $X = 0$

を満たすことをいう．

命題 2.21. 多様体 M 上にコントラスト関数 ϕ が与えられたとき，

$$
\begin{aligned}
g(Y, Z) &= -\phi(Y|Z) \\
g(\nabla_X Y, Z) &= -\phi(XY|Z) \\
g(Y, \nabla^*_X Z) &= -\phi(Y|XZ)
\end{aligned}
$$

によって g, ∇, ∇^* が定まり，(M, g, ∇, ∇^*) は捩れをもたない統計多様体になる．

証明. $X, Y, Z \in \mathfrak{X}(M)$ を任意にとる．任意の $p \in M$ に対して，関数 $\phi(p, \,\cdot\,) \in C^\infty(M)$ が p で最小値をとることから，

$$
\phi(\,\cdot\,|Z) = 0 \qquad \phi(\,\cdot\,|ZZ) \leq 0
$$

が成り立つ．$\phi(\,\cdot\,|Z) = 0$ の両辺に Y を作用させて，

$$
\phi(Y|Z) + \phi(\,\cdot\,|YZ) = 0
$$

同様に，

$$\phi(Z|Y) + \phi(\,\cdot\,|ZY) = 0$$

$[Y,Z] := YZ - ZY \in \mathfrak{X}(M)$ に対しても，$\phi(\,\cdot\,|[Y,Z]) = 0$ が成り立つから，$\phi(\,\cdot\,|YZ) = \phi(\,\cdot\,|ZY)$ より，

$$g(Y,Z) = -\phi(Y|Z) = -\phi(Z|Y) = g(Z,Y)$$

よって，g は対称である．また，

$$g(Z,Z) = -\phi(Z|Z) = -\phi(\,\cdot\,|ZZ) \geq 0$$

で，等号は $Z = 0$ のときのみ成り立つので，g は正定値である．

任意の $f \in C^\infty(M)$ に対して，

$$\begin{aligned} g(\nabla_X(fY), Z) &= -\phi(X(fY)|Z) \\ &= -\phi((Xf)Y + fXY|Z) \\ &= -\phi((Xf)Y|Z) - \phi(fXY|Z) \\ &= g((Xf)Y, Z) + g(f\nabla_X Y, Z) \\ &= g((Xf)Y + f\nabla_X Y, Z) \end{aligned}$$

が成り立つことなどから，∇, ∇^* はアファイン接続である．

$$\begin{aligned} g(T(X,Y), Z) &= g(\nabla_X Y - \nabla_Y X - [X,Y], Z) \\ &= g(\nabla_X Y, Z) - g(\nabla_Y X, Z) - g([X,Y], Z) \\ &= -\phi(XY|Z) - \phi(YX|Z) - \phi([X,Y]|Z) \\ &= 0 \end{aligned}$$

より，$T = 0$ である．同様に，$T^* = 0$ が成り立つ． □

定義 2.22 (プレコントラスト関数)**.** TM を M の接束とする．
$\phi \in C^\infty(M \times TM)$ と $X_1, \ldots, X_m, Y_1, \ldots, Y_n, Z \in \mathfrak{X}(M)$ に対して，
$\rho(X_1 \cdots X_m | Y_1 \cdots Y_n Z) \in C^\infty(M)$ を

$$\rho(X_1 \cdots X_m | Y_1 \cdots Y_n Z)(r) = (X_1)_p \cdots (X_m)_p (Y_1)_q \cdots (Y_n)_q \rho(p, Z_q)|_{p=r, q=r}$$

で定義する．$\rho \in C^\infty(M \times TM)$ がプレコントラスト関数であるとは，

1. 任意の $f_1, f_2 \in C^\infty(M)$, $X_1, X_2 \in \mathfrak{X}(M)$ に対して，

$$\rho(\,\cdot\,|f_1X_1 + f_2X_2) = f_1\rho(\,\cdot\,|X_1) + f_2\rho(\,\cdot\,|X_2)$$

2. 任意の $X \in \mathfrak{X}(M)$ に対して，$\rho(\,\cdot\,|X) = 0$
3. 任意の $X \in \mathfrak{X}(M)$ に対して，$\rho(X|X) \geq 0$ かつ，$\rho(X|X) = 0$ ならば $X = 0$

を満たすことをいう．

コントラスト関数が与えられると，自明なプレコントラスト関数が定まる．

命題 2.23. コントラスト関数 $\phi \in C^\infty(M \times M)$ に対して，

$$\rho(p, X_q) := X_q \phi(p, \cdot)$$

によって定義される $\rho \in C^\infty(M \times TM)$ はプレコントラスト関数である．

証明. 任意の $f_1, f_2 \in C^\infty(M)$, $X_1, X_2 \in \mathfrak{X}(M)$ に対して，

$$\begin{aligned}\rho(\,\cdot\,|f_1X_1 + f_2X_2) &= \phi(\,\cdot\,|f_1X_1 + f_2X_2) \\ &= f_1\phi(\,\cdot\,|X_1) + f_2\phi(\,\cdot\,|X_2) \\ &= f_1\rho(\,\cdot\,|X_1) + f_2\rho(\,\cdot\,|X_2)\end{aligned}$$

任意の $X \in \mathfrak{X}(M)$ に対して，

$$\rho(\,\cdot\,|X) = \phi(\,\cdot\,|X) = 0$$

また，

$$\rho(X|X) = \phi(X|X) \geq 0$$

で，等号は $X = 0$ のときのみ成り立つ． □

命題 2.24. 多様体 M 上にプレコントラスト関数 ρ が与えられたとき，

$$\begin{aligned}g(Y, Z) &= -\rho(Y|Z) \\ g(\nabla_X Y, Z) &= -\rho(XY|Z) \\ g(Y, \nabla^*_X Z) &= -\rho(Y|XZ)\end{aligned}$$

によって g, ∇, ∇^* が定まり，(M, g, ∇, ∇^*) は一般に捩れを許す統計多様体になる．

証明. g が Riemann 計量，∇, ∇^* がアファイン接続であることはコントラスト関数の場合と全く同様に示せる．

また，

$$\begin{aligned}g(T(X, Y), Z) &= g(\nabla_X Y - \nabla_Y X - [X, Y], Z) \\ &= g(\nabla_X Y, Z) - g(\nabla_Y X, Z) - g([X, Y], Z) \\ &= -\rho(XY|Z) - \rho(YX|Z) - \rho([X, Y]|Z) \\ &= 0\end{aligned}$$

より，$T = 0$ である．一方，一般に

$$\rho(Y|XZ) - \rho(Y|ZX) = \rho(Y|XZ - ZX)$$

とはならないので，T^* は 0 とは限らない． □

定義 2.25. 統計的モデル $\{p(x;\theta)\}_{\theta\in\Theta}$ と推定関数 $u(x;\theta)$ が与えられたとき．

$$A(\theta) = -\sum_{x\in\mathcal{X}} p(x;\theta)\frac{\partial u}{\partial\theta}(x;\theta)$$
$$V(\theta) = \sum_{x\in\mathcal{X}} p(x;\theta)u(x;\theta)u(x;\theta)^{\mathrm{T}}$$

とおく．ただし，ここでは任意の $\theta\in\Theta$ に対して $V(\theta)$ が正則であることを仮定する．

$$\widetilde{u}(x;\theta) := A(\theta)^{\mathrm{T}}V(\theta)^{-1}u(x;\theta)$$

で定義される $\widetilde{u}(x;\theta)$ を標準化された推定関数という．

$M=\{p(x;\theta)\}_{\theta\in\Theta}$ とすると，M は θ を座標系とする多様体とみなせる．p_1,p_2 の座標をそれぞれ θ_1,θ_2 とすると，推定関数 $u(x;\theta)$ が与えられたとき，

$$\rho\left(p_1,\left(\frac{\partial}{\partial\theta^j}\right)_{p_2}\right) := \sum_{x\in\mathcal{X}} p(x;\theta_1)\widetilde{u}_j(x;\theta_2)$$

によって M 上のプレコントラスト関数が定まる．対応する計量と接続は

$$\begin{aligned}
g_{ij}(\theta) &= \rho(\partial_i|\partial_j)(\theta)\\
&= \sum_{x\in\mathcal{X}} \frac{\partial p}{\partial\theta^i}(x;\theta)\widetilde{u}_j(x;\theta)\\
&= \sum_{x\in\mathcal{X}} p(x;\theta)\widetilde{u}_i(x;\theta)\widetilde{u}_j(x;\theta)\\
\Gamma_{ij,k}(\theta) &= \rho(\partial_i\partial_j|\partial_k)(\theta)\\
&= \sum_{x\in\mathcal{X}} \frac{\partial^2 p}{\partial\theta^i\partial\theta^j}(x;\theta)\widetilde{u}_k(x;\theta)\\
\Gamma^*_{ij,k}(\theta) &= \rho(\partial_k|\partial_i\partial_j)(\theta)\\
&= \sum_{x\in\mathcal{X}} \frac{\partial p}{\partial\theta^k}(x;\theta)\frac{\partial\widetilde{u}_j}{\partial\theta^i}(x;\theta)
\end{aligned}$$

となり，(M,g,∇,∇^*) は一般には捩れを許す統計多様体となる．ただし，$\widetilde{u}(x;\theta)$ が各 $x\in\mathcal{X}$ について θ に関して可積分ならば，すなわち

$$\frac{\partial\widetilde{u}_j}{\partial\theta^i}(x,\theta) = \frac{\partial\widetilde{u}_i}{\partial\theta^j}(x,\theta) \qquad (x\in\mathcal{X},\ i,j=1,\dots,d)$$

が成り立つならば，$\Gamma^*_{ij,k}(\theta)=\Gamma^*_{ji,k}(\theta)$ となるため，∇^* にも捩れは生じない．

なお，$g_{ij}(\theta)$ を ij 成分にもつ行列を $G(\theta)$ とすると，

$$
\begin{aligned}
G(\theta) &= \sum_{x\in\mathcal{X}} p(x;\theta)\widetilde{u}(x;\theta)\widetilde{u}(x;\theta)^{\mathrm{T}} \\
&= \sum_{x\in\mathcal{X}} p(x;\theta)(A(\theta)^{\mathrm{T}}V(\theta)^{-1}u(x;\theta))(A(\theta)^{\mathrm{T}}V(\theta)^{-1}u(x;\theta))^{\mathrm{T}} \\
&= A(\theta)^{\mathrm{T}}V(\theta)^{-1}\left(\sum_{x\in\mathcal{X}} p(x;\theta)u(x;\theta)u(x;\theta)^{\mathrm{T}}\right)V(\theta)^{-1}A(\theta) \\
&= A(\theta)^{\mathrm{T}}V(\theta)^{-1}A(\theta)
\end{aligned}
$$

となる．これは定理 2.13 において，Z 推定量が収束する正規分布の分散の逆行列に一致し，特に推定関数として $s(x;\theta) = \dfrac{\partial \log p}{\partial \theta}(x;\theta)$ を用いた場合は Fisher 情報行列に一致する．$G(\theta)$ は Fisher 情報行列の一般化であり，Godambe 情報行列と呼ばれている．

第 3 章

主結果

3.1 推定力学系

本節では推定力学系を導入し，その解の局所漸近安定性について述べる．なめらかにパラメトライズされた統計的モデル $\{p(x;\theta)\}_{\theta\in\Theta}$，推定関数 $u(x;\theta)$，および未知の分布 $p(x;\theta_*)$ に従う i.i.d. 観測データ $(X_1,\ldots,X_N)$ が与えられたとして，

$$\psi_N(\theta) := \frac{1}{N}\sum_{n=1}^{N} u(X_n;\theta)$$
$$\psi_*(\theta) := \sum_{x\in\mathcal{X}} p(x;\theta_*)u(x;\theta)$$

とおく．推定関数の定義より，$\psi_*(\theta_*) = 0$ であるから，θ_* は力学系 $\dot{\theta} = \psi_*(\theta)$ の平衡点である．また，命題 2.17 より，$\dfrac{\partial\psi_*}{\partial\theta}(\theta_*)$ が安定行列ならば，θ_* は $\dot{\theta} = \psi_*(\theta)$ の局所漸近安定な平衡点となる．

一方，力学系 $\dot{\theta} = \psi_N(\theta)$ を推定力学系と呼ぶことにする．また，$\delta > 0$ を固定し，$\hat{\theta}_N, \theta_*$ が $\overline{B_\delta(\theta_*)}$ 上でそれぞれ ψ_N, ψ_* の唯一の零点であることを仮定する．

推定力学系について以下の定理が成り立つ．

定理 3.1. $\dfrac{\partial\psi_*}{\partial\theta}(\theta_*)$ 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．

証明． $\dfrac{\partial\psi_*}{\partial\theta}(\theta_*)$ が安定行列であることと行列から固有値への対応が連続であることから，ある $\gamma > 0$ が存在して，$\left\|A - \dfrac{\partial\psi_*}{\partial\theta}(\theta_*)\right\| < \gamma$ ならば A は安定行列となる．

補題 2.6 より，

$$\lim_{N\to\infty}\sup_{\theta\in\overline{B_\delta(\theta_*)}}\left\|\frac{\partial\psi_N}{\partial\theta}(\theta) - \frac{\partial\psi_*}{\partial\theta}(\theta)\right\| = 0 \quad \text{a.s.}$$

が成り立つ．

また，Z 推定量の強一致性 (命題 2.8) より $\hat{\theta}_N \to \theta_*$ a.s. であるので，$\dfrac{\partial \psi_*}{\partial \theta}$ の連続性と連続写像定理 (補題 2.10) より，

$$\lim_{N\to\infty} \left\| \frac{\partial \psi_*}{\partial \theta}(\hat{\theta}_N) - \frac{\partial \psi_*}{\partial \theta}(\theta_*) \right\| = 0 \quad \text{a.s.}$$

が成り立つ．

以上より，

$$\begin{aligned}
&\left\| \frac{\partial \psi_N}{\partial \theta}(\hat{\theta}_N) - \frac{\partial \psi_*}{\partial \theta}(\theta_*) \right\| \\
&\le \left\| \frac{\partial \psi_N}{\partial \theta}(\hat{\theta}_N) - \frac{\partial \psi_*}{\partial \theta}(\hat{\theta}_N) \right\| + \left\| \frac{\partial \psi_*}{\partial \theta}(\hat{\theta}_N) - \frac{\partial \psi_*}{\partial \theta}(\theta_*) \right\| \\
&\le \sup_{\theta \in \overline{B_\delta(\theta_*)}} \left\| \frac{\partial \psi_N}{\partial \theta}(\theta) - \frac{\partial \psi_*}{\partial \theta}(\theta) \right\| + \left\| \frac{\partial \psi_*}{\partial \theta}(\hat{\theta}_N) - \frac{\partial \psi_*}{\partial \theta}(\theta_*) \right\| \\
&\to 0 \qquad \text{a.s.}
\end{aligned}$$

が成り立つので，特に，

$$\lim_{N\to\infty} P\left(\left\| \frac{\partial \psi_N}{\partial \theta}(\hat{\theta}_N) - \frac{\partial \psi_*}{\partial \theta}(\theta_*) \right\| < \gamma \right) = 1$$

したがって，

$$\begin{aligned}
&P\left(\frac{\partial \psi_N}{\partial \theta}(\hat{\theta}_N) \text{ が安定行列} \right) \\
&\ge P\left(\left\| \frac{\partial \psi_N}{\partial \theta}(\hat{\theta}_N) - \frac{\partial \psi_*}{\partial \theta}(\theta_*) \right\| < \gamma \right) \\
&\to 1 \quad (N \to \infty)
\end{aligned}$$

が成り立つ．この事実と命題 2.17 より，$\hat{\theta}_N$ 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． □

3.2 Contrastive Divergence 法の漸近理論

$p(x;\theta_*)$ に従う i.i.d. 観測データ $(X_1, \dots, X_N)$ が与えられたとき，θ の関数

$$\prod_{n=1}^{N} p(X_n;\theta)$$

を尤度といい，尤度を最大にするパラメータとして定義される推定量を最尤推定量という．最尤推定量は θ の方程式

$$\frac{1}{N}\sum_{n=1}^{N} \frac{\partial \log p}{\partial \theta}(X_N;\theta) = 0$$

の解としても得られるので，推定関数として

$$s(x;\theta)=\frac{\partial \log p}{\partial \theta}(x;\theta)$$

をとったときの Z 推定量に一致する．このとき，

$$\begin{aligned}A(\theta)&=-\sum_{x\in\mathcal{X}}p(x;\theta)\frac{\partial s}{\partial \theta}(x;\theta)\\ V(\theta)&=\sum_{x\in\mathcal{X}}p(x;\theta)s(x;\theta)s(x;\theta)^{\mathrm{T}}\end{aligned}$$

はともに Fisher 情報行列

$$\sum_{x\in\mathcal{X}}p(x;\theta)\left(\frac{\partial \log p}{\partial \theta}(x;\theta)\right)\left(\frac{\partial \log p}{\partial \theta}(x;\theta)\right)^{\mathrm{T}}$$

に一致する．

以下，統計的モデルが

$$p(x;\theta)=\frac{\exp(-H(x;\theta))}{Z(\theta)}$$

の形で表されている場合を考える．このような統計的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，

$$Z(\theta)=\sum_{x\in\mathcal{X}}\exp(-H(x;\theta))$$

を正規化定数または分配関数という．例えば，$\mathcal{X}=\{0,1\}^L$ のとき，$Z(\theta)$ の計算量は $O(2^L)$ となることからもわかるように，$Z(\theta)$ の評価には，計算量的な困難がしばしば伴う．エネルギーベースモデルにおける最尤推定について考えよう．

$$\begin{aligned}s(x;\theta)&=\frac{\partial \log p}{\partial \theta}(x;\theta)\\ &=\frac{\partial}{\partial \theta}\Big(-H(x;\theta)-\log Z(\theta)\Big)\\ &=-\frac{\partial H}{\partial \theta}(x;\theta)-\frac{1}{Z(\theta)}\sum_{y\in\mathcal{X}}\frac{\partial}{\partial \theta}\exp(-H(y;\theta))\\ &=\sum_{y\in\mathcal{X}}p(y;\theta)\frac{\partial H}{\partial \theta}(y;\theta)-\frac{\partial H}{\partial \theta}(x;\theta)\end{aligned}$$

より，推定力学系は

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

と表せる．ここで，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，$Z(\theta)$ の評価が必要となる．ここでは $Z(\theta)$ の直接計算を回避し，近似的に期待値を計算するために，2.3 節で導入したマルコフ連鎖モンテカルロ法を用いる．任意の $x, y \in \mathcal{X}$, $\theta \in \Theta$ に対して，詳細つりあい条件

$$p(x;\theta)T(y|x;\theta) = p(y;\theta)T(x|y;\theta)$$

を満たす θ に関してなめらかな遷移核の族 $\{T(x|y;\theta)\}_{\theta\in\Theta}$ が与えられたとし，$T(x|y;\theta)$ から生成されるマルコフ作用素を Φ_θ とする．命題 2.18 より，任意の確率分布 q と $x \in \mathcal{X}$ に対して，

$$\lim_{k\to\infty} (\Phi_\theta^k q)(x) = p(x;\theta)$$

が成り立つので，θ を更新するごとに十分大きい k をとり，ランダムに選んだ初期状態 Y_0 を k 回遷移させて得られる $Y_k(\theta)$ を $p(y;\theta)$ に従う確率変数とみなせばよい．

Hinton [2] はマルコフ連鎖の初期状態をランダムに選んだ初期状態ではなく，観測データ $(X_n)_{n=1}^N$ とし，遷移回数 k を 1 とする近似計算手法を提案した．この手法を Contrastive Divergence (CD) 法という．X_n を 1 回遷移させて得られる確率変数は $T(y|X_n)$ に従うので，CD 法におけるパラメータ更新則は

$$\dot{\theta} = \frac{1}{N}\sum_{n=1}^{N}\sum_{y\in\mathcal{X}} T(y|X_n;\theta)\frac{\partial H}{\partial\theta}(y;\theta) - \frac{1}{N}\sum_{n=1}^{N}\frac{\partial H}{\partial\theta}(X_n;\theta) \tag{3.1}$$

となる．ここで，

$$u^{\mathrm{CD}}(x;\theta) := s(x;\theta) - \sum_{y\in\mathcal{X}} T(y|x;\theta)s(y;\theta)$$

とおくと，$u^{\mathrm{CD}}(x;\theta)$ は後の定理 3.3 に示すように推定関数となる．これに，

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

を代入すると，

$$\begin{aligned} u^{\mathrm{CD}}(x;\theta) &= \sum_{y\in\mathcal{X}} p(y;\theta)\frac{\partial H}{\partial\theta}(y;\theta) - \frac{\partial H}{\partial\theta}(x;\theta) - \sum_{y\in\mathcal{X}} p(y;\theta)\frac{\partial H}{\partial\theta}(y;\theta) + \sum_{y\in\mathcal{X}} T(y|x;\theta)\frac{\partial H}{\partial\theta}(y;\theta) \\ &= \sum_{y\in\mathcal{X}} T(y|x;\theta)\frac{\partial H}{\partial\theta}(y;\theta) - \frac{\partial H}{\partial\theta}(x;\theta) \end{aligned}$$

を得る．推定関数 $u^{\mathrm{CD}}(x;\theta)$ を用いて得られる Z 推定量 $\hat{\theta}_N^{\mathrm{CD}}$ を CD 推定量と呼ぶことにする．なお，

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

とおくと，CD 法のパラメータ更新則 (3.1) は推定力学系 $\dot{\theta} = \psi_N^{\mathrm{CD}}(\theta)$ に一致する．

補題 3.2. 任意の $\theta \in \Theta$ に対して，

$$A^{\mathrm{CD}}(\theta) := -\sum_{x \in \mathcal{X}} p(x;\theta) \frac{\partial}{\partial \theta} u^{\mathrm{CD}}(x;\theta)$$

は正定値行列である．

証明.

$$p(y;\theta) = \sum_{x \in \mathcal{X}} T(y|x;\theta) p(x;\theta)$$

の両辺を微分して

$$\frac{\partial p}{\partial \theta}(y;\theta) = \sum_{x \in \mathcal{X}} T(y|x;\theta) \frac{\partial p}{\partial \theta}(x;\theta) + \sum_{x \in \mathcal{X}} p(x;\theta) \frac{\partial T}{\partial \theta}(y|x;\theta)$$

を得る．詳細つりあい条件 $p(x;\theta)T(y|x;\theta) = p(y;\theta)T(x|y;\theta)$ に注意して，

$$\begin{aligned}
A^{\mathrm{CD}}(\theta) &= -\sum_{x \in \mathcal{X}} p(x;\theta) \frac{\partial}{\partial \theta} \Big(s(x;\theta) - \sum_{y \in \mathcal{X}} T(y|x;\theta) s(y;\theta) \Big) \\
&= \sum_{y \in \mathcal{X}} s(y;\theta) \Big(\sum_{x \in \mathcal{X}} p(x;\theta) \frac{\partial T}{\partial \theta}(y|x;\theta) \Big)^{\mathrm{T}} \\
&= \sum_{y \in \mathcal{X}} s(y;\theta) \Big(\frac{\partial p}{\partial \theta}(y;\theta) - \sum_{x \in \mathcal{X}} T(y|x;\theta) \frac{\partial p}{\partial \theta}(x;\theta) \Big) \\
&= \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{X}} p(y;\theta) T(x|y;\theta) s(y;\theta) \Big(s(y;\theta) - s(x;\theta) \Big)^{\mathrm{T}} \\
&= -\frac{1}{2} \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{X}} p(x;\theta) T(y|x;\theta) s(y;\theta) \Big(s(x;\theta) - s(y;\theta) \Big)^{\mathrm{T}} \\
&\qquad + \frac{1}{2} \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{X}} p(x;\theta) T(y|x;\theta) s(x;\theta) \Big(s(x;\theta) - s(y;\theta) \Big)^{\mathrm{T}} \\
&= \frac{1}{2} \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{X}} p(x;\theta) T(y|x;\theta) \Big(s(x;\theta) - s(y;\theta) \Big) \Big(s(x;\theta) - s(y;\theta) \Big)^{\mathrm{T}}
\end{aligned}$$

より，任意の $v \in \mathbb{R}^d$ に対して

$$v^{\mathrm{T}} A^{\mathrm{CD}}(\theta) v = \frac{1}{2} \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{X}} p(x;\theta) T(y|x;\theta) \Big(v^{\mathrm{T}} (s(x;\theta) - s(y;\theta)) \Big)^2 \geq 0$$

が成り立つ．また，等号は任意の $x, y \in \mathcal{X}$ に対して

$$v^{\mathrm{T}} (s(x;\theta) - s(y;\theta)) = 0$$

のとき，すなわち

$$v^1 s_1(x;\theta) + \cdots + v^d s_d(x;\theta) = v^1 s_1(y;\theta) + \cdots + v^d s_d(y;\theta)$$

のときのみ成り立つ．$1, s_1(x;\theta), \ldots, s_d(x;\theta)$ は x の関数として 1 次独立なので，このとき $v = 0$ となる．よって，$A^{\mathrm{CD}}(\theta)$ は正定値行列である． □

定理 3.3. $u^{\mathrm{CD}}(x;\theta)$ は推定関数である．

証明. 定義 2.2 の 3 条件が成り立つことを確認すればよい．まず，仮定より $\theta \mapsto u^{\mathrm{CD}}(x;\theta)$ は C^∞ 級．次に，詳細つりあい条件より

$$\begin{aligned}\sum_{x\in\mathcal{X}} p(x;\theta)u^{\mathrm{CD}}(x;\theta) &= \sum_{x\in\mathcal{X}} p(x;\theta)\Big(s(x;\theta) - \sum_{y\in\mathcal{X}} T(y|x;\theta)s(y;\theta)\Big)\\ &= \sum_{x\in\mathcal{X}} p(x;\theta)s(x;\theta) - \sum_{y\in\mathcal{X}} p(y;\theta)s(y;\theta)\\ &= 0\end{aligned}$$

さらに，補題 3.2 より，$A^{\mathrm{CD}}(\theta)$ は正則行列である． □

系 3.4. CD 推定量 $\hat{\theta}_N^{\mathrm{CD}}$ が推定力学系 $\dot{\theta} = \psi_N^{\mathrm{CD}}(\theta)$ の局所漸近安定な平衡点である確率は，$N \to \infty$ のとき 1 に収束する．

証明. 関数 ψ_*^{CD} を

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

で定義すると，補題 3.2 より，$A^{\mathrm{CD}}(\theta_*)$ が正定値なので，$\dfrac{\partial \psi_*^{\mathrm{CD}}}{\partial \theta}(\theta_*) = -A^{\mathrm{CD}}(\theta_*)$ は安定行列である．よって，定理 3.1 より，結論が従う． □

3.3 Contrastive Divergence 法の情報幾何

2.4 節で，標準化された推定関数が非可積分な場合には，誘導されるアファイン接続の一方に捩れが生じ得ることについて述べた．本節では，CD 法で用いる標準化された推定関数が一般には非可積分となり，誘導されるアファイン接続の一方に捩れが生じることを例示する．

例 3.5. $\mathcal{X} = \{0,1\}^2$, $\Theta = \mathbb{R}^2$ とし，エネルギー関数

$$H(x;\theta) = -ax_1 - wx_1x_2 \qquad (x = (x_1, x_2),\ \theta = (a, w))$$

をもつエネルギーベースモデル

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

を考える．このモデルは片方のユニットにのみバイアスのあるユニット数 2 のボルツマンマシンであり，a がバイアス，w がユニット間の重みパラメータを表している．マルコフ連鎖の遷移核とし

て，例 2.19 で導入した

$$T(x|y;\theta) := \frac{1}{2}\Big(\frac{p(x_1,x_2;\theta)}{p(0,x_2;\theta)+p(1,x_2;\theta)}\cdot\frac{p(y_1,x_2;\theta)}{p(y_1,0;\theta)+p(y_1,1;\theta)} + \frac{p(x_1,x_2;\theta)}{p(x_1,0;\theta)+p(x_1,1;\theta)}\cdot\frac{p(x_1,y_2;\theta)}{p(0,y_2;\theta)+p(1,y_2;\theta)}\Big)$$

を用いると，CD 法に対応する推定関数は

$$u(x;\theta) = \sum_{y\in\mathcal{X}} T(y|x;\theta)\frac{\partial H}{\partial\theta}(y;\theta) - \frac{\partial H}{\partial\theta}(x;\theta)$$

と表され，$A(\theta), V(\theta)$ は以下で定義されるのであった．

$$A(\theta) := -\sum_{x\in\mathcal{X}} p(x;\theta)\frac{\partial u}{\partial\theta}(x;\theta)$$
$$V(\theta) := \sum_{x\in\mathcal{X}} p(x;\theta)u(x;\theta)u(x;\theta)^{\mathrm{T}}$$

まず，この設定において，$V(\theta)$ が正則であることを示そう．Mathematica を用いた計算により，

$$\det V(\theta) = \frac{E(\theta)}{64(e^a+1)^4(e^w+1)^4(e^{a+w}+1)^4(e^{a+w}+e^a+2)^2}$$

ここに，

$$\begin{aligned}E(\theta) = e^{2a+w}(&1064e^{a+w} + 1536e^{2a+w} + 922e^{3a+w} + 193e^{4a+w} + 2352e^{a+2w} + 5232e^{2a+2w} \\ &+ 5250e^{3a+2w} + 2370e^{4a+2w} + 378e^{5a+2w} + 2352e^{a+3w} + 7680e^{2a+3w} + 11708e^{3a+3w} \\ &+ 8751e^{4a+3w} + 2994e^{5a+3w} + 352e^{6a+3w} + 1064e^{a+4w} + 5232e^{2a+4w} + 11708e^{3a+4w} \\ &+ 13212e^{4a+4w} + 7380e^{5a+4w} + 1792e^{6a+4w} + 128e^{7a+4w} + 168e^{a+5w} + 1536e^{2a+5w} \\ &+ 5250e^{3a+5w} + 8751e^{4a+5w} + 7380e^{5a+5w} + 2880e^{6a+5w} + 384e^{7a+5w} + 144e^{2a+6w} \\ &+ 922e^{3a+6w} + 2370e^{4a+6w} + 2994e^{5a+6w} + 1792e^{6a+6w} + 384e^{7a+6w} + 40e^{3a+7w} \\ &+ 193e^{4a+7w} + 378e^{5a+7w} + 352e^{6a+7w} + 128e^{7a+7w} + 168e^a + 144e^{2a} + 40e^{3a} \\ &+ 256e^w + 384e^{2w} + 256e^{3w} + 64e^{4w} + 64)\end{aligned}$$

これより，行列式は常に正，よって $V(\theta)$ は正則である．

次に，$A(\theta), V(\theta)$ によって標準化された推定関数

$$\widetilde{u}(x;\theta) = A(\theta)^{\mathrm{T}}V(\theta)^{-1}u(x;\theta)$$

が $\theta = (\theta^1,\theta^2) = (a,w)$ に関して非可積分であることを示そう．そのためには，

$$\frac{\partial\widetilde{u}_1}{\partial\theta^2}(x;\theta) \neq \frac{\partial\widetilde{u}_2}{\partial\theta^1}(x;\theta)$$

を示せばよい．特に $x = (0,0)$ の場合を計算してみると，

$$\frac{\partial\widetilde{u}_1}{\partial\theta^2}((0,0);\theta) = \frac{\widetilde{E}_{12}(\theta)}{\widetilde{D}(\theta)}, \qquad \frac{\partial\widetilde{u}_2}{\partial\theta^1}((0,0);\theta) = \frac{\widetilde{E}_{21}(\theta)}{\widetilde{D}(\theta)}$$

となる．ここに，

$$\widetilde{D}(\theta) = (64 + 168e^{a} + 144e^{2a} + 40e^{3a} + 256e^{w} + 384e^{2w} + 256e^{3w} + 64e^{4w} + 1064e^{a+w}$$
$$+ 1536e^{2a+w} + 922e^{3a+w} + 193e^{4a+w} + 2352e^{a+2w} + 5232e^{2a+2w} + 5250e^{3a+2w}$$
$$+ 2370e^{4a+2w} + 378e^{5a+2w} + 2352e^{a+3w} + 7680e^{2a+3w} + 11708e^{3a+3w} + 8751e^{4a+3w}$$
$$+ 2994e^{5a+3w} + 352e^{6a+3w} + 1064e^{a+4w} + 5232e^{2a+4w} + 11708e^{3a+4w} + 13212e^{4a+4w}$$
$$+ 7380e^{5a+4w} + 1792e^{6a+4w} + 128e^{7a+4w} + 168e^{a+5w} + 1536e^{2a+5w} + 5250e^{3a+5w}$$
$$+ 8751e^{4a+5w} + 7380e^{5a+5w} + 2880e^{6a+5w} + 384e^{7a+5w} + 144e^{2a+6w} + 922e^{3a+6w}$$
$$+ 2370e^{4a+6w} + 2994e^{5a+6w} + 1792e^{6a+6w} + 384e^{7a+6w} + 40e^{3a+7w} + 193e^{4a+7w}$$
$$+ 378e^{5a+7w} + 352e^{6a+7w} + 128e^{7a+7w})^2$$

$$\widetilde{E}_{12}(\theta) = -2e^{a+w}(512 + 1152e^{a} - 256e^{2a} - 2464e^{3a} - 2004e^{4a} - 328e^{5a} + 108e^{6a}$$
$$+ 4096e^{w} + 14336e^{2w} + 28672e^{3w} + 35840e^{4w} + 28672e^{5w} + 14336e^{6w} + 4096e^{7w}$$
$$+ 512e^{8w} + 17920e^{a+w} + 20736e^{2a+w} - 12352e^{3a+w} - 44208e^{4a+w} - 33744e^{5a+w}$$
$$- 9104e^{6a+w} - 304e^{7a+w} + 93696e^{a+2w} + 214528e^{2a+2w} + 162016e^{3a+2w}$$
$$- 119940e^{4a+2w} - 296408e^{5a+2w} - 199765e^{6a+2w} - 54204e^{7a+2w} - 4058e^{8a+2w}$$
$$+ 252416e^{a+3w} + 857600e^{2a+3w} + 1364736e^{3a+3w} + 818936e^{4a+3w} - 447536e^{5a+3w}$$
$$- 990906e^{6a+3w} - 590100e^{7a+3w} - 146156e^{8a+3w} - 11360e^{9a+3w} + 405248e^{a+4w}$$
$$+ 1847296e^{2a+4w} + 4339904e^{3a+4w} + 5437400e^{4a+4w} + 2981000e^{5a+4w}$$
$$- 646637e^{6a+4w} - 1836520e^{7a+4w} - 981542e^{8a+4w} - 211688e^{9a+4w} - 14368e^{10a+4w}$$
$$+ 410112e^{a+5w} + 2400256e^{2a+5w} + 7517312e^{3a+5w} + 13669760e^{4a+5w}$$
$$+ 14355200e^{5a+5w} + 7532072e^{6a+5w} + 216592e^{7a+5w} - 1921200e^{8a+5w}$$
$$- 951200e^{9a+5w} - 170368e^{10a+5w} - 8704e^{11a+5w} + 263680e^{a+6w} + 1957376e^{2a+6w}$$
$$+ 7853248e^{3a+6w} + 18870232e^{4a+6w} + 28067632e^{5a+6w} + 25392294e^{6a+6w}$$
$$+ 12602520e^{7a+6w} + 1962748e^{8a+6w} - 1042912e^{9a+6w} - 522240e^{10a+6w}$$
$$- 71680e^{11a+6w} - 2048e^{12a+6w} + 103936e^{a+7w} + 997888e^{2a+7w} + 5089024e^{3a+7w}$$
$$+ 15627408e^{4a+7w} + 30436224e^{5a+7w} + 38055332e^{6a+7w} + 29847400e^{7a+7w}$$
$$+ 13554840e^{8a+7w} + 2717792e^{9a+7w} - 178048e^{10a+7w} - 144896e^{11a+7w}$$
$$- 12288e^{12a+7w} + 22656e^{a+8w} + 303360e^{2a+8w} + 2013408e^{3a+8w} + 7877628e^{4a+8w}$$
$$+ 19574696e^{5a+8w} + 31805306e^{6a+8w} + 33785664e^{7a+8w} + 22790580e^{8a+8w}$$
$$+ 9070544e^{9a+8w} + 1772736e^{10a+8w} + 68608e^{11a+8w} - 14336e^{12a+8w} + 2048e^{a+9w}$$
$$+ 48896e^{2a+9w} + 457152e^{3a+9w} + 2341360e^{4a+9w} + 7433168e^{5a+9w}$$
$$+ 15402232e^{6a+9w} + 21154368e^{7a+9w} + 19051408e^{8a+9w} + 10814368e^{9a+9w}$$
$$+ 3554176e^{10a+9w} + 561664e^{11a+9w} + 24576e^{12a+9w} + 3072e^{2a+10w} + 51808e^{3a+10w}$$
$$+ 376492e^{4a+10w} + 1572328e^{5a+10w} + 4175855e^{6a+10w} + 7322884e^{7a+10w}$$

$$+ 8510846e^{8a+10w} + 6398368e^{9a+10w} + 2928384e^{10a+10w} + 716800e^{11a+10w} + 67584e^{12a+10w} + 2048e^{3a+11w} + 27064e^{4a+11w} + 161200e^{5a+11w} + 565638e^{6a+11w} + 1276556e^{7a+11w} + 1900916e^{8a+11w} + 1848320e^{9a+11w} + 1119616e^{10a+11w} + 378368e^{11a+11w} + 53248e^{12a+11w} + 512e^{4a+12w} + 5592e^{5a+12w} + 28103e^{6a+12w} + 84168e^{7a+12w} + 162258e^{8a+12w} + 203608e^{9a+12w} + 160864e^{10a+12w} + 72704e^{11a+12w} + 14336e^{12a+12w})$$

$$\widetilde{E}_{21}(\theta) = e^{a+w}(3072 + 11264e^{a} + 15168e^{2a} + 8416e^{3a} + 944e^{4a} - 576e^{5a} - 80e^{6a} + 22528e^{w} + 71680e^{2w} + 129024e^{3w} + 143360e^{4w} + 100352e^{5w} + 43008e^{6w} + 10240e^{7w} + 1024e^{8w} + 118784e^{a+w} + 242240e^{2a+w} + 237440e^{3a+w} + 107000e^{4a+w} + 11152e^{5a+w} - 5106e^{6a+w} - 720e^{7a+w} + 507904e^{a+2w} + 1430784e^{2a+2w} + 2056992e^{3a+2w} + 1583072e^{4a+2w} + 595776e^{5a+2w} + 60144e^{6a+2w} - 16892e^{7a+2w} - 1941e^{8a+2w} + 1187840e^{a+3w} + 4405504e^{2a+3w} + 8591872e^{3a+3w} + 9535448e^{4a+3w} + 5989200e^{5a+3w} + 1911242e^{6a+3w} + 186864e^{7a+3w} - 27031e^{8a+3w} - 1680e^{9a+3w} + 1693696e^{a+4w} + 8074624e^{2a+4w} + 20521664e^{3a+4w} + 30589344e^{4a+4w} + 27406624e^{5a+4w} + 14321864e^{6a+4w} + 3884540e^{7a+4w} + 354845e^{8a+4w} - 21600e^{9a+4w} + 912e^{10a+4w} + 1531904e^{a+5w} + 9328512e^{2a+5w} + 30311680e^{3a+5w} + 58546928e^{4a+5w} + 70031136e^{5a+5w} + 51817556e^{6a+5w} + 22664752e^{7a+5w} + 5169759e^{8a+5w} + 413808e^{9a+5w} - 8048e^{10a+5w} + 2304e^{11a+5w} + 876544e^{a+6w} + 6897920e^{2a+6w} + 28580928e^{3a+6w} + 70398720e^{4a+6w} + 108782848e^{5a+6w} + 107080032e^{6a+6w} + 66001896e^{7a+6w} + 24025422e^{8a+6w} + 4491136e^{9a+6w} + 286336e^{10a+6w} - 2048e^{11a+6w} + 1024e^{12a+6w} + 303104e^{a+7w} + 3202304e^{2a+7w} + 17185280e^{3a+7w} + 53996656e^{4a+7w} + 106330400e^{5a+7w} + 135036148e^{6a+7w} + 110536544e^{7a+7w} + 56569146e^{8a+7w} + 16805600e^{9a+7w} + 2453760e^{10a+7w} + 107520e^{11a+7w} - 1024e^{12a+7w} + 56320e^{a+8w} + 876864e^{2a+8w} + 6368864e^{3a+8w} + 26027824e^{4a+8w} + 65510784e^{5a+8w} + 106098048e^{6a+8w} + 112050488e^{7a+8w} + 76131258e^{8a+8w} + 31792576e^{9a+8w} + 7395104e^{10a+8w} + 768000e^{11a+8w} + 17408e^{12a+8w} + 4096e^{a+9w} + 122944e^{2a+9w} + 1341312e^{3a+9w} + 7499032e^{4a+9w} + 24699216e^{5a+9w} + 51298294e^{6a+9w} + 69210000e^{7a+9w} + 60726526e^{8a+9w} + 33689696e^{9a+9w} + 11026848e^{10a+9w} + 1834496e^{11a+9w} + 105472e^{12a+9w} + 6144e^{2a+10w} + 134816e^{3a+10w} + 1160352e^{4a+10w} + 5302976e^{5a+10w} + 14529520e^{6a+10w} + 25250260e^{7a+10w} + 28369943e^{8a+10w} + 20352896e^{9a+10w} + 8857600e^{10a+10w} + 2074624e^{11a+10w} + 191488e^{12a+10w} + 4096e^{3a+11w} + 76216e^{4a+11w} + 556944e^{5a+11w} + 2161570e^{6a+11w} + 5010224e^{7a+11w}$$

$$
\begin{aligned}
&+ 7289037e^{8a+11w} + 6709680e^{9a+11w} + 3778432e^{10a+11w} + 1176576e^{11a+11w} \\
&+ 152576e^{12a+11w} + 1024e^{4a+12w} + 19616e^{5a+12w} + 131528e^{6a+12w} \\
&+ 447756e^{7a+12w} + 882777e^{8a+12w} + 1056672e^{9a+12w} + 761552e^{10a+12w} \\
&+ 305152e^{11a+12w} + 52224e^{12a+12w} + 1352e^{6a+13w} + 10384e^{7a+13w} + 32819e^{8a+13w} \\
&+ 54576e^{9a+13w} + 50512e^{10a+13w} + 24832e^{11a+13w} + 5120e^{12a+13w})
\end{aligned}
$$

例えば，$\theta = (a, w) = (0, \log 2)$ のとき，

$$
\begin{aligned}
\frac{\partial \widetilde{u}_1}{\partial \theta^2}((0,0);(0,\log 2)) &= -\frac{150438473044}{1597501349929} \\
\frac{\partial \widetilde{u}_2}{\partial \theta^1}((0,0);(0,\log 2)) &= -\frac{263090566216}{1597501349929}
\end{aligned}
$$

であるので，一般に

$$
\frac{\partial \widetilde{u}_1}{\partial \theta^2}((0,0);\theta) \neq \frac{\partial \widetilde{u}_2}{\partial \theta^1}((0,0);\theta)
$$

となることがわかる．

このことから，誘導されるアファイン接続 ∇^* の捩率は 0 とはならない．実際，直接計算により

$$
\Gamma^*_{12,1}(\theta) - \Gamma^*_{21,1}(\theta) = \frac{4(e^a+1)(e^w-1)^3(e^w+1)^3 e^{4a+w}(e^{a+w}+1)}{D^*(\theta)}
$$

を得る．ここに，

$$
\begin{aligned}
D^*(\theta) = (e^{a+w} + e^a + 2)^2 (&1064e^{a+w} + 1536e^{2a+w} + 922e^{3a+w} + 193e^{4a+w} + 2352e^{a+2w} \\
&+ 5232e^{2a+2w} + 5250e^{3a+2w} + 2370e^{4a+2w} + 378e^{5a+2w} + 2352e^{a+3w} + 7680e^{2a+3w} \\
&+ 11708e^{3a+3w} + 8751e^{4a+3w} + 2994e^{5a+3w} + 352e^{6a+3w} + 1064e^{a+4w} + 5232e^{2a+4w} \\
&+ 11708e^{3a+4w} + 13212e^{4a+4w} + 7380e^{5a+4w} + 1792e^{6a+4w} + 128e^{7a+4w} + 168e^{a+5w} \\
&+ 1536e^{2a+5w} + 5250e^{3a+5w} + 8751e^{4a+5w} + 7380e^{5a+5w} + 2880e^{6a+5w} + 384e^{7a+5w} \\
&+ 144e^{2a+6w} + 922e^{3a+6w} + 2370e^{4a+6w} + 2994e^{5a+6w} + 1792e^{6a+6w} + 384e^{7a+6w} \\
&+ 40e^{3a+7w} + 193e^{4a+7w} + 378e^{5a+7w} + 352e^{6a+7w} + 128e^{7a+7w} + 168e^{a} + 144e^{2a} \\
&+ 40e^{3a} + 256e^{w} + 384e^{2w} + 256e^{3w} + 64e^{4w})
\end{aligned}
$$

例えば，$\theta = (a, w) = (0, \log 2)$ のとき，

$$
\Gamma^*_{12,1}(0,\log 2) - \Gamma^*_{21,1}(0,\log 2) = \frac{648}{31598075}
$$

である．

例 3.6. $\mathcal{X} = \{0,1\}^2$，$\check{\Theta} = \mathbb{R}^3$ とし，エネルギー関数

$$
H(x;\check{\theta}) = -ax_1 - wx_1x_2 - bx_2 \qquad \big(x = (x_1, x_2),\ \check{\theta} = (a, w, b)\big)
$$

をもつエネルギーベースモデルを考える．このモデルを $\check{\theta} = (a, w, b)$ を座標系とする多様体とみなすと，例 3.5 は $b = 0$ で表される部分多様体になっている．先程と同様に，$V(\check{\theta})$ の行列式を計算すると，

$$\det V(\check{\theta}) = \frac{E(\check{\theta})}{16(e^a+1)^4(e^b+1)^4(e^{a+w}+1)^4(e^{b+w}+1)^4(e^{a+b+w}+e^a+e^b+1)^2}$$

ここに，

$$\begin{aligned} E(\check{\theta}) = e^{2a+2b+w}&(e^{a+b+w}+e^{a+w}+e^{b+w}+1)^2(10e^{a+b+w}+4e^{2a+b+w}+4e^{a+2b+w}+3e^{a+b+2w} \\ &+4e^{2a+b+2w}+4e^{a+2b+2w}+4e^{2a+2b+2w}+3e^{a+b}+4e^{a+w}+4e^{2a+w}+4e^a+4e^{b+w} \\ &+4e^{2b+w}+4e^b+4)^2 \end{aligned}$$

これより，行列式は常に正，よって $V(\check{\theta})$ は正則である．次に，標準化された推定関数

$$\widetilde{u}(x;\check{\theta}) = A(\check{\theta})^{\mathrm{T}} V(\check{\theta})^{-1} u(x;\check{\theta})$$

を計算すると，

$$\begin{aligned}
\widetilde{u}_1((0,0);\check{\theta}) &= -\frac{e^a(e^{b+w}+1)}{e^{a+b+w}+e^a+e^b+1} \\
\widetilde{u}_2((0,0);\check{\theta}) &= -\frac{e^{a+b+w}}{e^{a+b+w}+e^a+e^b+1} \\
\widetilde{u}_3((0,0);\check{\theta}) &= -\frac{e^b(e^{a+w}+1)}{e^{a+b+w}+e^a+e^b+1} \\
\widetilde{u}_1((0,1);\check{\theta}) &= -\frac{e^a(e^{b+w}+1)}{e^{a+b+w}+e^a+e^b+1} \\
\widetilde{u}_2((0,1);\check{\theta}) &= -\frac{e^{a+b+w}}{e^{a+b+w}+e^a+e^b+1} \\
\widetilde{u}_3((0,1);\check{\theta}) &= \frac{e^a+1}{e^{a+b+w}+e^a+e^b+1} \\
\widetilde{u}_1((1,0);\check{\theta}) &= \frac{e^b+1}{e^{a+b+w}+e^a+e^b+1} \\
\widetilde{u}_2((1,0);\check{\theta}) &= -\frac{e^{a+b+w}}{e^{a+b+w}+e^a+e^b+1} \\
\widetilde{u}_3((1,0);\check{\theta}) &= -\frac{e^b(e^{a+w}+1)}{e^{a+b+w}+e^a+e^b+1} \\
\widetilde{u}_1((1,1);\check{\theta}) &= \frac{e^b+1}{e^{a+b+w}+e^a+e^b+1} \\
\widetilde{u}_2((1,1);\check{\theta}) &= \frac{e^a+e^b+1}{e^{a+b+w}+e^a+e^b+1} \\
\widetilde{u}_3((1,1);\check{\theta}) &= \frac{e^a+1}{e^{a+b+w}+e^a+e^b+1}
\end{aligned}$$

これより，すべての $x \in \{0,1\}^2$ に対して，

$$\widetilde{u}(x;\check{\theta}) = \frac{\partial \log p}{\partial \check{\theta}}(x;\check{\theta})$$

となることが確認できる．すなわち，このモデルにおいて，CD 法に対応する標準化された推定関数は，最尤推定に対応する推定関数に一致する．よって，誘導される双対幾何構造は最尤推定から誘導されるものと一致し，モデルは双対平坦多様体となる．

補足

例 3.5，例 3.6 のモデルが定める統計多様体をそれぞれ $(M, g, \nabla, \nabla^*), (N, h, \check{\nabla}, \check{\nabla}^*)$ としよう．M は N の部分多様体であるから，N の双対構造 $(h, \check{\nabla}, \check{\nabla}^*)$ から M に双対構造を誘導することができる．では，この誘導された双対構造は (g, ∇, ∇^*) に一致するだろうか．答えは否である．実際，h を M に制限して得られる計量は Fisher 計量であり，g とは異なる．

第 4 章

結言

結論

本論文では，CD 法におけるパラメータ更新則を表す力学系と，そこから誘導される双対幾何構造について研究し，以下の結論を得た．

1. CD 法におけるパラメータ更新則は，ある推定関数に対応する推定力学系に一致し，平衡点が局所漸近安定である確率は，データ数 $N \to \infty$ のとき 1 に収束する．なお，その力学系の平衡点として得られる推定量は Z 推定量であるので，一致性と漸近正規性を有する．この結果は，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明らかにするものである．
2. ある非対称性をもつボルツマンマシンにおいて，ギブスサンプリングを用いた CD 法に対応する標準化された推定関数は非可積分であり，プレコントラスト関数を通して誘導されるアファイン接続の一方には，捩れが生じる．

今後の課題

以下に，今後の課題を挙げる．

1. 本論文では，統計的モデルに識別可能性条件を仮定した．一方で，CD 法は隠れ変数をもつボルツマンマシン等の，識別可能性条件を満たさない特異モデルにも応用されている．このようなモデルに対して，本論文で述べた結果はどの程度成立するかを研究することは，重要な未解決問題である．
2. 統計的モデルに推定関数から誘導される双対幾何構造について，捩れが生じることにはどのような統計的意味があるのかを明らかにしたい．

謝辞

まず，本論文執筆にあたる全ての面において，終始懇切丁寧なご指導，ご鞭撻を頂いた藤原彰夫教授に，心より感謝申し上げます．

藤原教授には，学部の 1 年間と修士の 2 年間にわたり，日々のセミナーやディスカッションにおいて，多くの厳しくも的確なアドバイスを頂きました．太田慎一教授をはじめとする，理化学研究所革新知能統合研究センター数理解析チームの皆様には，セミナーでの発表の機会を頂き，多くの貴重なアドバイスを頂きました．研究室の先輩である筒井大二氏には，研究の進め方等について，親身になって相談に乗って頂きました．数学専攻の同期の皆様には，多くの刺激を頂きました．研究生活に関わるすべての方々に，深くお礼申し上げます．

最後に，大学院への進学を許可して頂き，経済面および精神面において支えて頂いた家族に，深く感謝の意を表します．

参考文献

[1] 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).

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

[3] 伊庭幸人, “マルコフ連鎖モンテカルロ法の基礎,” 計算統計 II, pp.3-106, 岩波書店 (2005).

[4] 笠原晧司, 微分方程式の基礎, 朝倉書店 (1982).

[5] L. H. Loomis and S. Sternberg, *Advanced Calculus*, Jones and Bartlett Publishers (1990).

[6] P. McCullagh, *Generalized Linear Models (2nd Edition)*, Chapman and Hall (1989).

[7] W. Rudin, *Principles of Mathematical Analysis (3rd Edition)*, McGraw-Hill (1976).

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

[9] S. Wiggins, *Introduction to Applied Nonlinear Dynamical Systems and Chaos (2nd Edition)*, Springer (2003).

[10] 吉田朋広, 数理統計学, 朝倉書店 (2006).

[11] K. H. Yuan and R. I. Jennrich, “Asymptotics of Estimating Equations under Natural Conditions,” *Journal of Multivariate Analysis*, **65**, pp.245-260 (1998)