

1 MPF と CD1 の対応関係

1.1 MPF と CD1 の対応関係

この章では MPF と CD1 の対応関係について記述します。ただし仮定するモデルは、 θ をパラメータ、 $g(x)$ を十分統計量とす以下のようなギブス分布で与えられるとします。

前の章までの議論で CD の評価関数は以下のように表現でき

$$\frac{\partial}{\partial \theta_{ij}} CD(\theta) = \sum_{x \in \mathcal{D}} \sum_{x' \in \partial x} (g_{ij}(x) - g_{ij}(x')) [e^{\theta(g(x) - g(x'))} + 1]^{-1} d^{-1} |\mathcal{D}|^{-1} p(x') \quad (1)$$

$$= \sum_{x \in \mathcal{D}} \sum_{x' \in \partial x} (g_{ij}(x) - g_{ij}(x')) T_{x'x}^{CD1}(\theta) p(x') \quad (2)$$

MPF の場合は以下のように書くことができます。

$$\frac{\partial}{\partial \theta_{ij}} MPF(\theta) = \sum_{x \in \mathcal{D}} \sum_{x' \in \partial x} (g_{ij}(x) - g_{ij}(x')) T_{x'x}^{MPF}(\theta) p(x') \quad (3)$$

従って、CD1, MPF の評価関数の勾配は遷移確率に対応する項が異なるだけでほとんど同じ形になります。以下の説明では CD1 と MPF の評価関数の勾配は全く同一の量になることがわかります。上の説明とコンシステントではないことは後ほど、調整します。ただし、遷移確率は状態空間全域で定義されており、状態空間を χ とし、 $x, x' \in \chi$ とするとき

$$p(x'|x; \theta) = \begin{cases} > 0 & x \in \mathcal{D} \text{ and } x' \in \partial x \\ 0 & (\text{otherwise}) \end{cases}$$

$$MPF(\theta) = \sum_{x \in \mathcal{D}} \sum_{x' \notin \mathcal{D}} p(x'|x; \theta) p(x) \quad (4)$$

$$= \sum_{x \in \mathcal{D}} \sum_{x' \in \partial x} p(x'|x; \theta) p(x) \quad (5)$$

2 行目に移るところでは遷移先が humming 距離 1 に制限してあることを利用しました。評価関数をパラメータについて微分すると

$$\frac{\partial}{\partial \theta} MPF(\theta) = \text{const} \sum_{x \in \mathcal{D}} \sum_{x' \in \partial x} (g(x') - g(x)) \sum_{x' \in \partial x} p(x'|x; \theta) p(x) \quad (6)$$

$$= \text{const} \sum_{x \in \mathcal{D}} \sum_{x' \in \partial x + \{x\}} (g(x') - g(x)) p(x'|x; \theta) p(x) \quad (7)$$

$$= \text{const} \sum_{x \in \chi} \sum_{x' \in \chi} (g(x') - g(x)) p(x'|x; \theta) p(x) \quad (8)$$

$$\propto \sum_{x \in \chi} \sum_{x' \in \chi} g(x') p(x'|x; \theta) p(x) - \sum_{x \in \chi} \sum_{x' \in \chi} g(x) p(x'|x; \theta) p(x) \quad (9)$$

$$= \sum_{x' \in \chi} g(x') \sum_{x \in \chi} p(x', x; \theta) - \sum_{x \in \chi} g(x) p(x) \sum_{x' \in \chi} p(x'|x; \theta) \quad (10)$$

$$= \sum_{x' \in \chi} g(x') p^{(1)}(x'; \theta) - \sum_{x \in \chi} g(x) p^{(0)}(x) \quad (11)$$

第一行目の式は、前章までの結果を書きました。第二行目の式は x から $x' = x$ への遷移の場合は $g(x) - g(x') = 0$ よりキャンセルするため、同じ遷移先も含めて和をとっています。第3行目は確率分布が状態空間全体で定義されていることを利用しました。最後の式の変形は、確率分布が状態空間全域で定義されていることを利用して、同時確率の周辺化をそれぞれの確率変数 x, x' に対して行いました。

1.2 得られた評価関数の勾配の考察

前章までに得られた結果を考察します。Shor らの提案した MPF[?] ではデータの状態点から、データ以外の状態点へ遷移する確率を最小にすることを評価関数に取り入れていましたが、評価関数のパラメータ微分を、十分統計量の差を始状態と次の状態の同時確率で平均化した量として定義することで、自然な形で、MFP の評価関数と対応が付き、CD 1 と一致します (すると MPF のパラメータ微分は CD1 のそれと同一のものになっていなければならないのではないか?)。

方法では MPF の評価関数は遷移確率の選び方によりますが、遷移行列に対する対角成分については寄与しないことが分かります。

2 CD1 の一致性の証明

最尤推定は一致性 (標本数が無限の極限で真のパラメータに任意の精度で近くなる性質) をもっており、また最尤推定解は不偏推定量 (標本数 N を固定した下で、得られる標本を真の標本で平均したときに推定解が真のパラメータに) に成っていることが保証されています。この章では、CD1 と MPF による推定が一致性を持つことの証明を示します (MPF については CD とほとんど同じであるため省略します)。不偏推定量になっているか

2.1 CD1 の一致性証明

$$\frac{\partial}{\partial \theta} CD(\theta_0) = -\frac{\partial CD(\hat{\theta})}{\partial \theta} + \frac{\partial CD(\theta_0)}{\partial \theta} \quad (12)$$

$$= -\frac{\partial^2 CD(\theta_0)}{\partial \theta^2} (\hat{\theta} - \theta_0) + O(\|\hat{\theta} - \theta_0\|^2) \quad (13)$$

$$(14)$$

ここで、 $\hat{\theta}$ は CD1 を用いて推定解であるとし、 θ_0 は真値であるとし、CD 法を用いた解が θ_0 に対して近いと仮定して、一致性を確かめる議論では二次以上の項は無視できるとします。まず右辺に対して CD1 の表現を代入します。

$$\frac{\partial^2 CD(\theta_0)}{\partial \theta^2} = \lim_{\theta \rightarrow \theta_0} \frac{\partial}{\partial \theta} |\mathcal{D}|^{-1} \sum_{x \in \mathcal{D}} \sum_{x' \in \partial x} (g_{ij}(x) - g_{ij}(x')) T_{x'x}^{CD1}(\theta) p(x') \quad (15)$$

$$\rightarrow \lim_{\theta \rightarrow \theta_0} \frac{\partial}{\partial \theta} \sum_{x \in \mathcal{X}} \sum_{x' \in \partial x} (g_{ij}(x) - g_{ij}(x')) T_{x'x}^{CD1}(\theta) p(x') p^{(\infty)}(x|\theta_0) \quad (16)$$

$$= \lim_{\theta \rightarrow \theta_0} \frac{\partial}{\partial \theta} \mathbb{E}_{x|\theta_0} \left[\sum_{x' \in \partial x} (g_{ij}(x) - g_{ij}(x')) T_{x'x}^{CD1}(\theta) p(x') \right] \quad (17)$$

$$(18)$$

1 行目から 2 行目へ移るところでは大数の法則を利用しました ($|\mathcal{D}|$ の極限)。