Abstract

9/12(Mon) のディスカッション後に 1MC 遷移で得られる Hamming Distance=1 の確率分布 $p^{(1)}(x)$ 及び、MPF,CD の評価関数 $MPF(\theta)$, $CD(p^{(0)}\|p^{(1)})$ の形式をまとめておきます。また、標本数 M 無限大の極限で $MPF(\theta)$ が 0 に漸近する条件を書きました (証明自体は、藤堂先生が示している通りですが、一部必要となる条件があるかと思います)。

1 Summary of $p^{(1)}$, $MPF(\theta)$ and $CD(p^{(0)}\|p^{(1)})$

経験分布は以下のように、データ数 M が大きくなると、同じ状態が複数現れるため、データ内で重複するものを同一視して、数密度で表現します。

$$\frac{q(x)}{M} := \frac{1}{M}\sum_{m=1}^{M}\delta(x, x^{(m)}) = p^{(0)}(x),\ M := |\mathcal{D}| \tag{1}$$

以降はこの経験分布の表現を用いて $p^{(1)}$ などを表現していきます。
$x \in \mathcal{D}$ の場合は

$$p^{(1)}(x) = \frac{q(x)}{M}(1 - \sum x' \in \partial x p(x'|x)) + \sum x' \in \partial x, x' \in \mathcal{D}\frac{q(x')}{M}p(x|x') \tag{2}$$

同様に、$x \notin \mathcal{D}$ の場合は $x \in \mathcal{D}$ の表現を使用します。

$$p^{(1)}(x) = \sum_{x' \in \partial x, x' \in \mathcal{D}} \frac{q(x')}{M}p(x'|x) \tag{3}$$

次に、$MPF(\theta)$ の表現です。基本的には上の議論で得られた $p^{(1)}$ を $MPF(\theta)$ に使用するだけです。ただし、$q(x)$ はデータに含まれている状態でのみ有限の値を持ち得るため、$p^{(1)}$ は

$$\begin{aligned}
MPF(\theta) &:= KL(p^{(0)}\|p^{(1)}|\theta) = \sum_{x:all} p^{(0)}(x)\log(\frac{p^{(0)}(x)}{p^{(1)}(x)}) && (4)\\
&= \sum_{x\in\mathcal{D}}\frac{q(x)}{M}\log\frac{q(x)}{M} - \sum_{x\in\mathcal{D}}\frac{q(x)}{M}\log\{p^{(1)}(x)\} && (5)\\
&= \sum_{x\in\mathcal{D}}\frac{q(x)}{M}\log\frac{q(x)}{M} - \sum_{x\in\mathcal{D}}\frac{q(x)}{M}\log\{\frac{q(x)}{M}(1-\sum_{x'\in\partial x}p(x'|x)) + \sum_{x'\in\partial x, x'\in\mathcal{D}}\frac{q(x')}{M}p(x|x')\} && (6)\\
& && (7)
\end{aligned}$$

CD の評価関数を書き下す前に、MPF の標本数が大きくなる極限の振る舞いについて以下では議論します。

2 MPF の標本数無限の性質

以下では MPF の評価関数に対して、標本数 M が漸近的に無限大になる場合に推定精度について考察します。
標本数無限の極限で真のパラメータ θ_0 に確率 1 で収束する、最尤推定法の場合と MPF を比較する事で、違いを理解したいと考えました。
まず、最尤推定ですが、本的には KL 距離を最小化することを考えます。

$$KL(p^{(0)}\|p^{(\infty)}(\theta) = \sum_x p^{(0)}(x)\frac{p^{(0)}(x)}{p^{(\infty)}(x)} \tag{8}$$

ですが、標本数が有限の場合、真の確率分布と $p^{(0)}$ 自身の "距離 "も有限の大きさで離れているため、上式の KL 距離を最小化しても、真のパラメータ θ_0 を得られたことにはなり得ません。しかしながら、標本数 M が無限大になることで

$$p^{(0)}(x) \to p^{(\infty)}(x|\theta_0) \tag{9}$$

となり、

$$KL(p^{(0)}\|p^{(\infty)}(\theta)) \to KL(p^{(\infty)}(\theta_0)\|p^{(\infty)}(\theta)) = \sum_x p^{(\infty)}(x|\theta_0)\frac{p^{(\infty)}(x|\theta_0)}{p^{(\infty)}(x|\theta)} \tag{10}$$

後は上式を、適当な最適化手法で、KL 距離の最小化が出来れば、確率 1 で真のパラメータを推定できます。
一方で MPF はというと、標本数 M の無限大で経験分布 $p^{(0)}(x)$ は $p^{(\infty)}(x|\theta_0)$ に漸近収束します。$p^{(1)}(x)$ の場合も同様に、$p^{(\infty)}(x|\theta_0)$ に漸近収束する事を示せばいいかと思います。
藤堂先生は、詳細釣り合いを満たす場合とそうではない場合について、$p^0(x) \to p^{(\infty)}(x|\theta*)$ の条件の下で、

$$\begin{aligned} p^{(1)}(x) &= p^{(0)}(x) - p^{(0)}(x)\sum_{x'\in\partial x} p(x'|x) + \sum_{x'\in\partial x} p^{(0)}(x')p(x|x') && (11)\\ &= p^{(0)}(x) - p^{(0)}(x)\sum_{x'\in\partial x} \frac{p^{(\infty)}(x')}{p^{(\infty)}(x)}p(x|x') + \sum_{x'\in\partial x} p^{(0)}(x')p(x|x') && (12)\end{aligned}$$

となり、$M \to \infty$ で $p^{(0)}(x) \to p^{(\infty)}(x|\theta_0)$ であるため、右辺の第 2 項目以降は消えて、第 1 項目のみ残り $p^{(0)}(x)$、$p^{(1)}(x) = p^{(0)}(x) = p^{(\infty)}(x)$ を示していたかと思います。
詳細釣り合いを満たさない場合についても同様に考えます。

$$\begin{aligned} p^{(1)}(x) &= p^{(0)}(x)\{1 - \sum_{x'\in\partial x} p(x'|x)\} + \sum_{x'\in\partial x, x'\in\mathcal{D}} p^{(0)}(x')p(x|x') && (13)\\ &= p^{(0)}(x) - \sum_{x'\in\partial x}\{p^{(0)}(x)p(x'|x) - p^{(0)}(x')p(x|x')\} && (14)\end{aligned}$$

ただし、$M \to \infty$ では、どの状態も必ずデータの中に含まれているとして和の条件を外しました。従って、上式 2 行目の第 2 項目を balance condition(任意の状態点に注目した時、流入流出の流量が釣り合っている条件) を満たすように遷移確率を作ればいいのだと思います。

ここまで考えてみて、よく分からなくなった事があります。
MPF において、標本数無限では、評価関数 ($p^{(0)} = p^{(\infty)}(x|\theta_0)$ と $p^{(1)}(\theta)$ の KL 距離) が 0 になる事が分かったのですが。

$$p^{(0)}(x) \to p^{(\infty)}(x|\theta_0),\ (\text{example. } p^{(\infty)}(x|\theta_0) = \frac{e^{-E(x|\theta_0)}}{Z(\theta_0)}) \tag{15}$$

最尤推定の場合は、関数形が同じだったので KL を最小化して得られる θ を、parameterize された確率分布に適応すれば、推定できたことになるのですが、$p^{(1)}(x|theta)$ の θ における関数形は $p^{(\infty)}(x|\theta)$ とは異なります。
$KL(p^{(0)\|p^{(1)}(\theta)})$ の最小化で得られたパラメータ θ^* を $p^{(\infty)}(\theta = \theta^*)$ とすると何が得られるのでしょうか。