

Majorization

大上由人

2024 年 5 月 9 日

1 古典的エントロピー及びダイバージェンス

1.1 古典的状态及び系

Def: 状態を表す確率分布

古典的系における状態は確率分布

$$p = (p_1, p_2, \dots, p_d)^\top \quad (1.1)$$

で表される。ここで、 $p_i \geq 0$ かつ $\sum_{i=1}^d p_i = 1$ である。また、 d 次元の確率分布全体の集合を、 \mathcal{P}_d と表す。

また、その集合に属する一様分布を、

$$u = \left(\frac{1}{d}, \frac{1}{d}, \dots, \frac{1}{d} \right)^\top \quad (1.2)$$

と表す。

また、異なる確率分布の積を、

$$p \otimes q \in \mathcal{P}_{dd'} \quad p \in \mathcal{P}_d, q \in \mathcal{P}_{d'} \quad (1.3)$$

と表し、とくに、同じ確率分布の累乗を、

$$p^{\otimes n} \in \mathcal{P}_{d^n} \quad p \in \mathcal{P}_d \quad (1.4)$$

と表す。

Def:Supp

確率分布 $p = p_{i_i} \in \mathcal{P}_d$ に対して、 p の台を、

$$\text{spp}(p) = \{i \in [d] | p_i > 0\} \subset \{1, 2, \dots, d\} \quad (1.5)$$

と表す。また、

$$\text{rank}(p) = |\text{spp}(p)| \quad (1.6)$$

を、 p のランクという。とくに、 $\text{rank}(p) = d$ のとき、 p はフルランクであるという。

要するに、確率が 0 でないようなインデックスの集合を台と呼び、その要素数をランクと呼ぶ。

Def: 確率遷移行列

古典的な確率分布の時間発展は、確率遷移行列 T を用いて以下のように表される。

$$p'_i = \sum_{j=1} T_{ij} p_j \quad (1.7)$$

Prop: 確率遷移行列の性質

確率遷移行列 T は以下の性質を持つ。

$$\sum_{i=1}^d T_{ij} = 1 \quad (1.8)$$

Prf

Def: 二重確率遷移行列

確率遷移行列 T が、

$$\sum_{j=1} T_{ij} = 1 \quad (1.9)$$

をみたすとき、二重確率遷移行列という。

Prop: 二重確率遷移行列の特徴づけ

以下の二つは同値である。

1. T は二重確率遷移行列である。
2. 一様分布 u は T に対して不変である。すなわち、 $u = Tu$ である。

Prf

$$p'_i = \sum_{j=1} T_{ij} u_j \quad (1.10)$$

$$= \frac{1}{d} \sum_{j=1}^d T_{ij} \quad (1.11)$$

$$= \frac{1}{d} \cdot d \quad (1.12)$$

$$= 1 \quad (1.13)$$

であることからわかる。

Def: トレース距離

二つの確率分布 p, q のトレース距離は、

$$D(p, q) = \frac{1}{2} \sum_{i=1}^d |p_i - q_i| \quad (1.14)$$

で定義される。

Prop: トレース距離の性質

トレース距離は、 T に対して非増加である。すなわち、

$$D(p, q) \geq D(Tp, Tq) \quad (1.15)$$

が成り立つ。

Prf

後により一般の証明をするため、ここでは省略する。

1.2 シャノンエントロピー及び KL ダイバージェンス

Def: シャノンエントロピー

確率分布 $p \in \mathcal{P}_d$ のシャノンエントロピーは、

$$S_1(p) = - \sum_{i=1}^d p_i \log p_i \quad (1.16)$$

で定義される。

Def:KL ダイバージェンス

二つの確率分布 $p, q \in \mathcal{P}_d$ の KL ダイバージェンスは、

$$S_1(p||q) = \sum_{i=1}^d p_i \log \frac{p_i}{q_i} \quad (1.17)$$

で定義される。ただし、 $\text{supp}(p) \subset \text{supp}(q)$ でないときは、 $S_1(p||q) = \infty$ とする。

このとき、エントロピーと KL ダイバージェンスの関係がわかる。

Prop: エントロピーと KL ダイバージェンスの関係

任意の $p, q \in \mathcal{P}_d$ に対して、

$$S_1(p) = \ln(d) - S_1(p||u) \quad (1.18)$$

Prf

$$S_1(p||u) = \sum_{i=1}^d p_i \log \frac{p_i}{\frac{1}{d}} \quad (1.19)$$

$$= \sum_{i=1}^d p_i \log dp_i \quad (1.20)$$

$$= \sum_{i=1}^d p_i \log d + \sum_{i=1}^d p_i \log p_i \quad (1.21)$$

$$= \log d - S_1(p) \quad (1.22)$$

であることからわかる。 □

これより、 $S_1(p) \leq \log d$ であることがわかる。

このとき、KL ダイバージェンスのテイラー展開は以下ようになる。

$$S_1(p||p - \Delta p) = \frac{1}{2} \sum_i \frac{(\Delta p_i)^2}{p_i} + O(\Delta p^3) \quad (1.23)$$

$$(1.24)$$

ここで、 $\sum_i \Delta p_i = 0$ を用いている。

Prop:KL ダイバージェンスの単調性

KL ダイバージェンスは、 $p' = Tp$ および $q' = Tq$ に対して、

$$S_1(p||q) \geq S_1(p'||q') \quad (1.25)$$

が成り立つ。

Prf

注意されたいこととして、KL ダイバージェンスの単調性の逆はいえない。すなわち、単調性を満たすが、 $p' = Tp$ および $q' = Tq$ を満たすような T が存在しない場合がある。

次に、二重確率遷移行列について考える。このとき、KL ダイバージェンスの単調性と、

$$S_1(p) \leq S_1(Tp) \quad (1.26)$$

が成り立つ。

すなわち、二重確率遷移行列による時間発展は、エントロピーを増加させる。

Def: 相互情報量

二つの確率分布 $p, q \in \mathcal{P}_d$ の相互情報量は、

$$I_1(p_{AB})_{A:B} = S_1(p_A) + S_1(p_B) - S_1(p_{AB}) = S_1(p_{AB} || p_A \otimes p_B) \geq 0 \quad (1.27)$$

で定義される。

この量は、A と B の相関を表す量である。

Prop: 相互情報量の性質

任意の $p, q \in \mathcal{P}_d$ に対して、

$$I_1(p_{AB})_{A:B} = 0 \Leftrightarrow p_{AB} = p_A \otimes p_B \quad (1.28)$$

が成り立つ。また、KL ダイバージェンスの単調性から、

$$I_1(p_{AB})_{A:B} \geq I_1(T_A \otimes T_B p_{AB})_{A:B} \quad (1.29)$$

が成り立つ。ただし、 $T_A \otimes T_B$ は、各 A,B に独立に作用する確率遷移行列である。

Prf

1.3 Rényi エントロピー及びダイバージェンス

Def: Rényi エントロピー

確率分布 $p \in \mathcal{P}_d$ の Rényi エントロピーは、 $0 \leq \alpha \leq \infty$ 、 $p \in \mathcal{P}_d$ に対して、

$$S_\alpha(p) = \frac{1}{1-\alpha} \log \left(\sum_{i=1}^d p_i^\alpha \right) \quad (1.30)$$

で定義される。

Def: Rényi ダイバージェンス

二つの確率分布 $p, q \in \mathcal{P}_d$ の Rényi ダイバージェンスは、 $0 \leq \alpha \leq \infty$ 、 $p \in \mathcal{P}_d$ に対して、

$$S_\alpha(p||q) = \frac{1}{\alpha-1} \log \left(\sum_{i=1}^d p_i^\alpha q_i^{1-\alpha} \right) \quad (1.31)$$

で定義される。ただし、 $\text{supp}(p) \subset \text{supp}(q)$ でないときは、 $S_\alpha(p||q) = \infty$ とする。

Prop: Rényi エントロピーと KL ダイバージェンスの関係

任意の $p, q \in \mathcal{P}_d$ に対して、

$$S_\alpha(p) = \frac{1}{1-\alpha} \log d - S_\alpha(p||u) \quad (1.32)$$

が成り立つ。

Prop: Rényi ダイバージェンスの非負性

任意の $p, q \in \mathcal{P}_d$ に対して、

$$S_\alpha(p||q) \geq 0 \quad (1.33)$$

が成り立つ。また、 $0 < \alpha \leq \infty$ に対して、

$$S_\alpha(p||q) = 0 \Leftrightarrow p = q \quad (1.34)$$

であり、また、 $\alpha = 0$ のとき、

$$S_0(p||q) = 0 \Leftrightarrow \text{supp}(p) \subset \text{supp}(q) \quad (1.35)$$

が成り立つ。

Prop: Rényi ダイバージェンスの単調性

任意の $p' = Tp, q' = Tq \in \mathcal{P}_d$ に対して、

$$S_\alpha(p||q) \geq S_\alpha(p'||q') \quad (1.36)$$

が成り立つ。また、二重確率遷移行列に対して、

$$S_\alpha(p) \leq S_\alpha(Tp) \quad (1.37)$$

が成り立つ。

Prop: Rényi ダイバージェンスの単調性 (2)

$\alpha \leq \alpha'$ に対して、

$$S_\alpha(p||q) \leq S_{\alpha'}(p||q) \quad (1.38)$$

が成り立ち、また、

$$S_\alpha(p) \geq S_{\alpha'}(p) \quad (1.39)$$

が成り立つ。

Lem:

f を下に凸な関数であるとし、 $p, q, p', q' \in \mathbb{R}^d$ がすべて正であるとする。もし、 $p' = Tp, q' = Tq$ であるとき、

$$\sum_{i=1}^d q'_i f\left(\frac{p'_i}{q'_i}\right) \leq \sum_{i=1}^d q_i f\left(\frac{p_i}{q_i}\right) \quad (1.40)$$

が成り立つ。

Prf

Jensen の不等式より、

$$\sum_{j=1}^d q'_j f\left(\frac{p'_j}{q'_j}\right) \leq \sum_{j=1}^d \sum_{i=1}^d \frac{T_{ji} q_i}{q'_j} f\left(\frac{p_i}{q_i}\right) = \sum_{i=1}^d q_i f\left(\frac{p_i}{q_i}\right) \quad (1.41)$$

である。 □

ここで、 $p, q \in \mathcal{P}_d$ として、 $f(x) = x \log x$ とすると、

$$S_1(p||q) \geq S_1(p'||q') \quad (1.42)$$

が成り立つ。これは、KL ダイバージェンスの単調性を示している。

Prf(Rényi ダイバージェンスの非負性)

$f_\alpha(x) = x^\alpha$ とすると、 $1 < \alpha \leq \infty$ に対して、 $f_\alpha(x)$ は下に凸な関数である。

Jensen の不等式より

$$\sum_{i=1}^d q_i f\left(\frac{p_i}{q_i}\right) \geq f\left(\sum_{i=1}^d q_i \frac{p_i}{q_i}\right) = f(1) = 0 \quad (1.43)$$

である。

Def:f-ダイバージェンス

$f(0, \infty) \rightarrow \mathbb{R}$ を下に凸な関数とし、 $x = 1$ で $f(x)$ が狭義凸かつ $f(1) = 0$ であるとする。このとき、 $p, q \in \mathcal{P}_d$ に対して、

$$D_f(p||q) = \sum_{i=1}^d q_i f\left(\frac{p_i}{q_i}\right) \quad (1.44)$$

で定義される。

Prop:f-ダイバージェンスの非負性

任意の $p, q \in \mathcal{P}_d$ に対して、

$$D_f(p||q) \geq 0 \quad (1.45)$$

が成り立つ。また、

$$D_f(p||q) = 0 \Leftrightarrow p = q \quad (1.46)$$

が成り立つ。

Prf

Prop:f-ダイバージェンスの単調性

任意の $p' = Tp, q' = Tq \in \mathcal{P}_d$ に対して、

$$D_f(p||q) \geq D_f(p'||q') \quad (1.47)$$

が成り立つ。

Prf

KL ダイバージェンスは、 $f(x) = x \log x$ のときの f-ダイバージェンスである。しかし、Rényi ダイバージェンスは、f-ダイバージェンスの形では表されない。

1.4 フィッシャー情報量

以下、我々は、なめらかにパラメータ化された確率分布 $p(\theta)$ について考える。ただし、 θ の取りうる領域は、 \mathbb{R}^m の開部分集合である。

Def: フィッシャー情報量

$p(\theta) \in \mathcal{P}_d$ がフルランクであるとし、 $\theta \in \mathbb{R}^m$ をパラメータとする。このとき、フィッシャー情報量は $m \times m$ 行列で、

$$J_{p(\theta),kl} = \sum_{i=1}^d p_i(\theta) \partial_k [\log p_i(\theta)] \partial_l [\log p_i(\theta)] = \sum_{i=1}^d \frac{\partial_k p_i(\theta) \partial_l p_i(\theta)}{p_i(\theta)} \quad (1.48)$$

で定義される。

フィッシャー情報量は、f-ダイバージェンスの極限として得られる。