

Majorization

大上由人

2024 年 5 月 13 日

目次

1	リソース理論と確率熱力学の位置づけ	2
1.1	確率熱力学的アプローチ	2
1.2	リソース理論的アプローチ	2
1.3	確率熱力学とリソース理論の関係	3
2	古典的エントロピー及びダイバージェンス	4
2.1	古典的状态及び系	4
2.2	シャノンエントロピー及び KL ダイバージェンス	6
2.3	Rényi エントロピー及びダイバージェンス	8
2.4	フィッシャー情報量	15
3	Classical Majorization	19
3.1	Majorization	19
3.2	d-Majorization と Thermo-Majorization	21
3.3	連続変数の場合	24
4	古典熱力学への適応	26
4.1	第二法則と KL ダイバージェンス	26
4.2	非平衡状態への遷移の要する仕事	29
4.3	熱平衡状態への緩和で取り出せる仕事	31
4.4	平衡状態間の遷移における仕事	31

1 リソース理論と確率熱力学の位置づけ

非平衡状態の熱力学第二法則を考えるとときに、これまで二つのアプローチがとられてきた。

1.1 確率熱力学のアプローチ

確率熱力学においては、仕事の揺らぎを許した系を考え、ゆらぎの定理の系として、エントロピー生成の熱力学第二法則が導かれる。

1.2 リソース理論的アプローチ

リソース理論においては、外部からの仕事は完全に制御されたものとし、系の状態を変化させるために必要なリソースを考える。

Def: free state/free operation

何のコストもなしに得られる状態を free state、何のコストもなしに行える操作を free operation という。

熱力学においては、仕事や、非平衡状態が resource である。というのも、仕事によって非平衡状態を作り出すことができ、また、非平衡状態から仕事を取り出すことができるからである。

また、Gibbs 分布 (カノニカル分布) は、free state であり、緩和過程 (GPM) は free operation である。

さらに、monotone と呼ばれる概念も重要である。

Def: monotone

free operation に対して、単調に変化する量を monotone という。とくに、ある操作が可能かどうかの必要十分条件を与える monotone を complete monotone という。

ex: 従来の熱力学

従来の熱力学における resource、free state、free operation、(complete) monotone は以下のように対応する。

- resource: 仕事

- free state: 熱平衡状態
- free operation: 断熱操作
- (complete) monotone: エントロピー

1.3 確率熱力学とリソース理論の関係

両者は、出所や set up が異なるが、マクロ極限をとると、同じ第二法則が導かれる。

2 古典的エントロピー及びダイバージェンス

2.1 古典的状态及び系

必要な量を定義する。

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

$$p = (p_1, p_2, \dots, p_d)^\top \quad (2.1)$$

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、

$$u = \left(\frac{1}{d}, \frac{1}{d}, \dots, \frac{1}{d} \right)^\top \quad (2.2)$$

と表す。

また、異なる確率分布の積を、

$$p \otimes q \in \mathcal{P}_{dd'} \quad p \in \mathcal{P}_d, q \in \mathcal{P}_{d'} \quad (2.3)$$

と表し、とくに、同じ確率分布の累乗を、

$$p^{\otimes n} \in \mathcal{P}_{d^n} \quad p \in \mathcal{P}_d \quad (2.4)$$

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、

$$\text{supp}(p) = \{i \in [d] | p_i > 0\} \subset \{1, 2, \dots, d\} \quad (2.5)$$

と表す。また、

$$\text{rank}(p) = |\text{supp}(p)| \quad (2.6)$$

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。

$$p'_i = \sum_{j=1} T_{ij} p_j \quad (2.7)$$

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。

$$\sum_{i=1}^d T_{ij} = 1 \quad (2.8)$$

Prf

略 (確率の規格化を利用する。)

□

Def: 二重確率遷移行列

確率遷移行列 T が、

$$\sum_{j=1} T_{ij} = 1 \quad (2.9)$$

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

$$p'_i = \sum_{j=1} T_{ij} u_j \quad (2.10)$$

$$= \frac{1}{d} \sum_{j=1}^d T_{ij} \quad (2.11)$$

$$= \frac{1}{d} \cdot d \quad (2.12)$$

$$= 1 \quad (2.13)$$

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、

$$D(p, q) = \frac{1}{2} \sum_{i=1}^d |p_i - q_i| \quad (2.14)$$

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、

$$D(p, q) \geq D(Tp, Tq) \quad (2.15)$$

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。

2.2 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、

$$S_1(p) = - \sum_{i=1}^d p_i \log p_i \quad (2.16)$$

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、

$$S_1(p||q) = \sum_{i=1}^d p_i \log \frac{p_i}{q_i} \quad (2.17)$$

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、

$$S_1(p) = \log(d) - S_1(p||q) \quad (2.18)$$

Prf

$$S_1(p||u) = \sum_{i=1}^d p_i \log \frac{p_i}{\frac{1}{d}} \quad (2.19)$$

$$= \sum_{i=1}^d p_i \log dp_i \quad (2.20)$$

$$= \sum_{i=1}^d p_i \log d + \sum_{i=1}^d p_i \log p_i \quad (2.21)$$

$$= \log d - S_1(p) \quad (2.22)$$

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、 $\log \frac{1}{1-x}$ のテイラー展開を用いて*¹、以下のようになる。

$$S_1(p||p - \Delta p) = \frac{1}{2} \sum_i \frac{(\Delta p_i)^2}{p_i} + O(\Delta p^3) \quad (2.23)$$

$$(2.24)$$

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、

$$S_1(p||q) \geq S_1(p'||q') \quad (2.25)$$

が成り立つ。

Prf

後に一般に示す。

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、

$$S_1(p) \leq S_1(Tp) \quad (2.26)$$

*¹ $x + \frac{x^2}{2} + \dots$

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、

$$I_1(p_{AB})_{A:B} = S_1(p_A) + S_1(p_B) - S_1(p_{AB}) = S_1(p_{AB} || p_A \otimes p_B) \geq 0 \quad (2.27)$$

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、

$$I_1(p_{AB})_{A:B} = 0 \Leftrightarrow p_{AB} = p_A \otimes p_B \quad (2.28)$$

が成り立つ。また、KL ダイバージェンスの単調性から、

$$I_1(p_{AB})_{A:B} \geq I_1(T_A \otimes T_B p_{AB})_{A:B} \quad (2.29)$$

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

略 (過去のゼミ資料を参考せよ)

□

2.3 Rényi エントロピー及びダイバージェンス

シャノンエントロピーを包含する概念として、Rényi エントロピーがある。

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、

$$S_\alpha(p) = \frac{1}{1-\alpha} \log \left(\sum_{i=1}^d p_i^\alpha \right) \quad (2.30)$$

で定義される。

また、ダイバージェンスについても、Rényi ダイバージェンスがある。

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、

$$S_\alpha(p||q) = \frac{1}{\alpha - 1} \log \left(\sum_{i=1}^d p_i^\alpha q_i^{1-\alpha} \right) \quad (2.31)$$

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。

これらの量が、たしかにシャノンエントロピーと KL ダイバージェンスを包含していることを示す。

Prop: Rényi エントロピーとシャノンエントロピーの関係

Rényi-1 エントロピーは、シャノンエントロピーに一致する。すなわち、

$$S_1(p) = S_\alpha(p)|_{\alpha=1} \quad (2.32)$$

が成り立つ。

Prf

$$S_\alpha(p)|_{\alpha=1} = -\lim_{\alpha \rightarrow 1} \frac{1}{1 - \alpha} \log \left(\sum_{i=1}^d p_i^\alpha \right) \quad (2.33)$$

$$= -\frac{d}{d\alpha} \log \left(\sum_{i=1}^d p_i^\alpha \right) |_{\alpha=1} \quad (2.34)$$

$$= -\frac{\sum_{i=1}^d p_i \log p_i}{\sum_{i=1}^d p_i} \quad (2.35)$$

$$= -\sum_{i=1}^d p_i \log p_i \quad (2.36)$$

$$= S_1(p) \quad (2.37)$$

であることからわかる。 □

Prop: Rényi ダイバージェンスとダイバージェンスの関係

Rényi-1 ダイバージェンスは、KL ダイバージェンスに一致する。すなわち、

$$S_\alpha(p||q)|_{\alpha=1} = S_1(p||q) \quad (2.38)$$

が成り立つ。

Prf

$$S_\alpha(p||q)|_{\alpha=1} = \lim_{\alpha \rightarrow 1} \frac{1}{\alpha - 1} \log \left(\sum_{i=1}^d p_i^\alpha q_i^{1-\alpha} \right) \quad (2.39)$$

$$= \frac{d}{d\alpha} \log \left(\sum_{i=1}^d p_i^\alpha q_i^{1-\alpha} \right) \Big|_{\alpha=1} \quad (2.40)$$

$$= \frac{\sum_{i=1}^d q_i (p_i/q_i)^\alpha \log(p_i/q_i)}{\sum_{i=1}^d p_i^\alpha q_i^{1-\alpha}} \Big|_{\alpha=1} \quad (2.41)$$

$$= \sum_{i=1}^d p_i \log \frac{p_i}{q_i} \quad (2.42)$$

$$= S_1(p||q) \quad (2.43)$$

であることからわかる。 □

また、 $\alpha = 0, \infty$ の場合は重要らしいので、以下で確認する。

Prop: Rényi エントロピー/ダイバージェンスの極限

Rényi-0 エントロピーは、

$$S_0(p) = \log(\text{rank}(p)) \quad (2.44)$$

で定義される。また、Rényi- ∞ エントロピーは、

$$S_\infty(p) = -(\log \max_i p_i) \quad (2.45)$$

で定義される。また、Rényi-0 ダイバージェンスは、

$$S_0(p||q) = -\log \left(\sum_{i:p_i>0} q_i \right) \quad (2.46)$$

で定義される。また、Rényi- ∞ ダイバージェンスは、

$$S_\infty(p||q) = \log \left(\max_i \frac{p_i}{q_i} \right) \quad (2.47)$$

で定義される。

Prf

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、

$$S_\alpha(p) = \frac{1}{1-\alpha} \log d - S_\alpha(p||u) \quad (2.48)$$

が成り立つ。

以下、Rényi ダイバージェンスについての性質を示す。

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、

$$S_\alpha(p||q) \geq 0 \quad (2.49)$$

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

$$S_\alpha(p||q) = 0 \Leftrightarrow p = q \quad (2.50)$$

であり、また、 $\alpha = 0$ のとき、

$$S_0(p||q) = 0 \Leftrightarrow \text{supp}(p) \subset \text{supp}(q) \quad (2.51)$$

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、

$$S_\alpha(p||q) \geq S_\alpha(p'||q') \quad (2.52)$$

が成り立つ。また、二重確率遷移行列に対して、

$$S_\alpha(p) \leq S_\alpha(Tp) \quad (2.53)$$

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、

$$S_\alpha(p||q) \leq S_{\alpha'}(p||q) \quad (2.54)$$

が成り立ち、また、

$$S_\alpha(p) \geq S_{\alpha'}(p) \quad (2.55)$$

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、

$$\sum_{i=1}^d q'_i f\left(\frac{p'_i}{q'_i}\right) \leq \sum_{i=1}^d q_i f\left(\frac{p_i}{q_i}\right) \quad (2.56)$$

が成り立つ。

Prf

Jensen の不等式より、

$$\sum_{j=1}^d q'_j f\left(\frac{p'_j}{q'_j}\right) \leq \sum_{j=1}^d \sum_{i=1}^d \frac{T_{ji} q_i}{q'_j} f\left(\frac{p_i}{q_i}\right) = \sum_{i=1}^d q_i f\left(\frac{p_i}{q_i}\right) \quad (2.57)$$

である。 □

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

$$S_1(p||q) \geq S_1(p'||q') \quad (2.58)$$

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

$$\sum_{i=1}^d q_i f\left(\frac{p_i}{q_i}\right) \geq f\left(\sum_{i=1}^d q_i \frac{p_i}{q_i}\right) = f(1) = 0 \quad (2.59)$$

である。両辺対数をとることにより、

$$\log\left(\sum_{i=1}^d q_i \frac{p_i}{q_i}\right) \geq 0 \quad (2.60)$$

である。これを両辺 $\frac{1}{\alpha-1}$ かけることにより、

$$\frac{1}{\alpha-1} \log\left(\sum_{i=1}^d q_i \frac{p_i}{q_i}\right) \geq 0 \quad (2.61)$$

である。したがって、

$$S_\alpha(p||q) \geq 0 \quad (2.62)$$

である。また、 $0 < \alpha < 1$ のときは、上の Jensen の不等式で不等号が逆になり、 $\frac{1}{\alpha-1}$ をかけるときにもう一度符号が逆になることに注意して、同様に示すことができる。

$\alpha = 0, 1, \infty$ の場合は、それぞれの定義から自明である。

□

Prf:(Rényi ダイバージェンスの単調性)

$1 < \alpha < \infty$ のとき、Lem で、 $f(x) = x^\alpha$ として示した不等式を用いると、

$$\sum_{i=1}^d q_i' \frac{p_i'^\alpha}{q_i'^\alpha} \leq \sum_{i=1}^d q_i \frac{p_i^\alpha}{q_i^\alpha} \quad (2.63)$$

である。この両辺対数をとることにより、

$$\log \left(\sum_{i=1}^d p_i'^\alpha q_i^{1-\alpha} \right) \leq \log \left(\sum_{i=1}^d p_i^\alpha q_i^{1-\alpha} \right) \quad (2.64)$$

である。したがって、

$$\frac{1}{\alpha-1} \log \left(\sum_{i=1}^d p_i'^\alpha q_i^{1-\alpha} \right) \leq \frac{1}{\alpha-1} \log \left(\sum_{i=1}^d p_i^\alpha q_i^{1-\alpha} \right) \quad (2.65)$$

である。したがって、

$$S_\alpha(p||q) \geq S_\alpha(p'||q') \quad (2.66)$$

である。

$0 < \alpha < 1$ のときは、同様に示すことができる。

$\alpha = 0, 1, \infty$ の場合は、それぞれの定義から自明である。

□

Prf:(Rényi ダイバージェンスの単調性 (2))

$\alpha \leq \alpha'$ に対して、 $f(x) = x^{\frac{\alpha-1}{\alpha'-1}}$ とすると、この関数は、 $1 < \alpha < \alpha' < \infty$ に対して下に凸な関数であり、 $0 < \alpha < \alpha' < 1$ および $0 < \alpha < 1 < \alpha'$ に対して上に凸な関数である。

したがって、Jensen の不等式より、

$$S_\alpha(p||q) = \frac{1}{\alpha - 1} \log \left(\sum_{i=1} p_i \left(\frac{p_i}{q_i} \right)^{\alpha-1} \right) \quad (2.67)$$

$$= \frac{1}{\alpha - 1} \log \left(\sum_{i=1} p_i \left(\frac{p_i}{q_i} \right)^{(\alpha'-1)(\frac{\alpha-1}{\alpha'-1})} \right) \quad (2.68)$$

$$\leq \frac{1}{\alpha' - 1} \log \left(\sum_{i=1} p_i \left(\frac{p_i}{q_i} \right)^{\alpha'-1} \right) \quad (2.69)$$

$$= S_{\alpha'}(p||q) \quad (2.70)$$

である。

また、 $\alpha = 0, 1, \infty$ の場合は、それぞれの定義から自明である。

□

さらに、f-ダイバージェンスという概念がある。

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、

$$D_f(p||q) = \sum_{i=1}^d q_i f \left(\frac{p_i}{q_i} \right) \quad (2.71)$$

で定義される。

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。また、Rényi ダイバージェンスは、 $f(x) = x^\alpha$ として \log をとって $\frac{1}{\alpha-1}$ をかけたものである。

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、

$$D_f(p||q) \geq 0 \quad (2.72)$$

が成り立つ。また、

$$D_f(p||q) = 0 \Leftrightarrow p = q \quad (2.73)$$

が成り立つ。

Prf

Jensen の不等式より、

$$\sum_{i=1}^d q_i f \left(\frac{p_i}{q_i} \right) \geq f \left(\sum_{i=1}^d q_i \frac{p_i}{q_i} \right) = f(1) = 0 \quad (2.74)$$

である。ことからわかる。 □

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、

$$D_f(p||q) \geq D_f(p'||q') \quad (2.75)$$

が成り立つ。

Prf

よくよく見ると、これは Lem で示した不等式と同じである。 □

2.4 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、

$$J_{p(\theta),kl} = \sum_{i=1}^d p_i(\theta) \partial_k [\log p_i(\theta)] \partial_l [\log p_i(\theta)] = \sum_{i=1}^d \frac{\partial_k p_i(\theta) \partial_l p_i(\theta)}{p_i(\theta)} \quad (2.76)$$

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。

Prop: フィッシャー情報量の単調性

任意の確率遷移行列 T に対して、

$$J_{p(\theta)} \geq J_{Tp(\theta)} \quad (2.77)$$

が成り立つ。

Prf

$p' = Tp$ とし、 $c = (c^1, \dots, c^m) \in \mathbb{R}^m$ として、 $\partial = \sum_k c^k \partial_k$ とする。このとき、

$$c^\top J_{p(\theta)} c = \sum_i p_i \left(\frac{\partial p_i}{p_i} \right)^2 \quad (2.78)$$

$$c^\top J_{p'(\theta)} c = \sum_i p'_i \left(\frac{\partial p'_i}{p'_i} \right)^2 \quad (2.79)$$

$$(2.80)$$

である。したがって、Lemma で $f = x^2$ として示した不等式より、

$$c^\top J_{p(\theta)} c \geq c^\top J_{p'(\theta)} c \quad (2.81)$$

である。したがって、

$$J_{p(\theta)} \geq J_{Tp(\theta)} \quad (2.82)$$

である。

□

フィッシャー情報量の操作的な意味付けとして、Cramér-Rao の不等式がある。

Thm:Cramér-Rao の不等式

あるパラメータ θ に対する不偏推定量 θ_{est} に対して、不偏条件 $\sum_{i=1}^d p_i(\theta) \theta_{est}(i) = \theta$ が成り立つとする。このとき、正確さは共分散行列

$$Cov_{\theta}^{kl}(\theta_{est}) = \sum_{i=1}^d p_i(\theta) (\theta_{est}^k(i) - \theta^k) (\theta_{est}^l(i) - \theta^l) \quad (2.83)$$

により表現される。このとき、

$$Cov_{\theta}^{kl}(\theta_{est}) \geq (J_{p(\theta)})_{kl}^{-1} \quad (2.84)$$

が成り立つ。

すなわち、フィッシャー情報量は、 θ の不偏推定量が、フィッシャー情報量によって制限されることを示している。

例として、指数型分布族とよばれる確率分布の集合族を考える。簡単のために、パラメータを θ のみとし、

$$p_i(\theta) = h_i \exp(\theta T_i - A(\theta)) \quad (2.85)$$

であるとする。ただし、 $A(\theta)$ は、 θ のなめらかな関数である。このとき、簡単な計算により、

$$\sum_{i=1} T_i p_i(\theta) = A'(\theta) \quad \sum_{i=1} T_i^2 p_i(\theta) = A''(\theta) + A'(\theta)^2 \quad (2.86)$$

である。したがって

$$J_{p(\theta)} = A''(\theta) \quad (2.87)$$

である。

熱力学の文脈では、 $p_i(\theta)$ をギブス分布、 T_i をエネルギーとして、 $-\theta$ を逆温度として解釈することができる。そして、 $\theta^{-1}A(\theta)$ は、自由エネルギーである。すなわち、

$$p_i(\theta) = \frac{1}{Z(\theta)} \exp(-\theta E_i) \quad (2.88)$$

である。

また、(2.89) は、 T_i が、 $A'(\theta)$ の不偏推定量であることを示している。それに対応するフィッシャー情報量は、 $J_{p(\theta)} = A''(\theta)^{-1}$ である。というのも、 $\frac{d}{d\theta} = (A''(\theta))^{-1} \frac{d}{d\theta}$ であるからである。

また、情報幾何の文脈では、フィッシャー情報量は、確率分布空間の計量として考えられる。以下では、monotone 計量という概念を導入する。

Def: Monotone 計量

$G_p : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ が、フルランクの $p \in \mathcal{P}_d$ に対して、以下を満たすとき、 G_p は Monotone 計量であるという。

- G_p は双線形である。
- $G_p(a, a) \geq 0$ であり、 $G_p(a, a) = 0 \Leftrightarrow a = 0$ である。
- $p \mapsto G_p(a, a)$ は任意の a に対して連続である。
- 任意の T, a, p に対して、 $G_p(a, a) \geq G_{Tp}(Ta, Ta)$ が成り立つ。

このとき、特にフィッシャー情報計量は、

$$G_p^F(a, b) = \sum_{i=1}^d \frac{a_i b_i}{p_i} \quad (2.89)$$

で定義される。ただし、 $a = (a_1, \dots, a_d), b = (b_1, \dots, b_d)$ である。

この計量は、フィッシャー情報量行列と、

$$J_{p(\theta), kl} = G_{p(\theta)}^F(\partial_k p(\theta), \partial_l p(\theta)) \quad (2.90)$$

との関係がある。

Prop: フィッシャー情報量計量の単調性

任意の確率遷移行列 T に対して、

$$G_p^F(a, a) \geq G_{Tp}^F(Ta, Ta) \quad (2.91)$$

が成り立つ。

Prf

逆に、任意の monotone 計量は、フィッシャー情報量計量を用いて、以下のように表現できる。

Thm:Chentsov の定理

任意の monotone 計量 G_p は、フィッシャー情報量計量を用いて、

$$G_p(a, b) = kG_p^F(a, b) + k' \left(\sum_{i=1}^d a_i \right) \left(\sum_{i=1}^d b_i \right) \quad (2.92)$$

と表現できる。ただし、 $k, k' \geq 0$ である。

Prf

略

3 Classical Majorization

状態の遷移可能性は、majorization によって、必要十分に特徴づけられる。

3.1 Majorization

Def:Majorization

$p, p' \in \mathcal{P}_d$ に対して、 p が p' を majorize することを、 $p' \prec p$ と書き、

$$\forall k \in \{1, \dots, d\} \quad \sum_{i=1}^k p_i'^{\downarrow} \leq \sum_{i=1}^k p_i^{\downarrow} \quad (3.1)$$

が成り立つとき、 p が p' を majorize するという。ただし、 p^{\downarrow} は、 p の要素を降順に並べたものである。

Majorization は、 p と p' の確率分布を比較したときに、ばらつきの大きさを比較する量である。

Majorization を図に表すと、以下のローレンツ曲線で表現される。

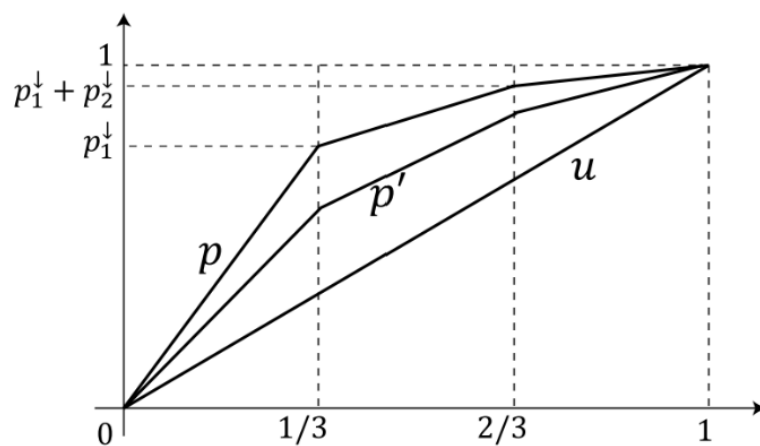


図 1 Majorization の図示

図でいうところの、下に来ている曲線が、よりばらつきが小さい確率分布であり、とくに一様分布のときは直線になることがわかる。すなわち、

$$u \prec p \quad \forall p \in \mathcal{P}_d \quad (3.2)$$

である。

注意すべきこととして、この \prec は全順序ではない。すなわち、ある確率分布の組 p, p' に対して、 $p \prec p'$ か $p' \prec p$ のどちらも成り立たない場合がある。このとき、ローレンツ曲線は交わる。

Thm:Majorization の特徴づけ

$p, p' \in \mathcal{P}_d$ に対して、以下は同値である。

1. $p' \prec p$
2. $\forall t \in \mathbb{R} \quad \sum_{i=1}^d |p'_i - t| \leq \sum_{i=1}^d |p_i - t|$
3. 任意の下に凸な関数 f に対して、

$$\sum_{i=1}^d f(p'_i) \leq \sum_{i=1}^d f(p_i) \quad (3.3)$$

4. $p' = Tp$ となる二重確率遷移行列 T が存在する。

Prf

この証明は後で行う。

3 についての補足

また、一様分布は、二重確率遷移行列のもとで不変である。熱力学の文脈では、 u は高温極限をとった時の Gibbs 分布に対応している。このとき、二重確率遷移行列は、そのような Gibbs 分布を保つものである。

Thm:Birkhoff の定理

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. T は、置換行列の凸結合で表現できる。すなわち、

$$T = \sum_k r_k P_k \quad (3.4)$$

である。ただし、 $r_k \geq 0$ であり、 $\sum_k r_k = 1$ である。また、 P_k は、置換行列である。

Prf

教科書で証明されていないので一旦飛ばす。

Prop: 二重確率遷移行列の性質

$$\text{「} T \text{ が二重確率遷移行列である」} \Leftrightarrow Tp \prec p \quad \forall p \in \mathbb{R}^d \quad (3.5)$$

Prf

Def:Schur 凸性

$F : \mathbb{R}^d \rightarrow \mathbb{R}$ が Schur 凸であるとは、

$$\forall p, p' \in \mathcal{P}_d \quad p' \prec p \Rightarrow F(p') \leq F(p) \quad (3.6)$$

が成り立つことである。言い換えると、Schur 凸な関数は、majorization に関して単調である。

Prop:

$F : \mathbb{R}^d \rightarrow \mathbb{R}$ が Schur 凸かつ微分可能なことと、 F が転置に対して不変なこと、すなわち、

$$F(p) = F(Pp) \quad \forall p, P \quad (3.7)$$

が成り立つことは同値であり、また、

$$\forall p \in \mathcal{P}_d \quad (p_i p_j) \left(\frac{\partial F}{\partial p_i} - \frac{\partial F}{\partial p_j} \right) \geq 0 \quad (3.8)$$

も同値である。

Prf

3.2 d-Majorization と Thermo-Majorization

先ほどまでは、一つの分布についての遷移について考えていたが、次は、二つの分布の組の遷移を考える。

このとき、 $p^* = (p_1^*, \dots, p_d^*)^\top, q^* = (q_1^*, \dots, q_d^*)^\top$ を、 $p = (p_1, \dots, p_d)^\top, q = (q_1, \dots, q_d)^\top$ を $\frac{p_1^*}{q_1^*} \geq \frac{p_2^*}{q_2^*} \geq \dots \geq \frac{p_d^*}{q_d^*}$ が成り立つように並べ替えたものとして定義する。この新しく定義した p^*, q^* に対して横軸を $\sum_{i=1}^k q_i^*$ 、縦軸を $\sum_{i=1}^k p_i^*$ としたローレンツ曲線を考える。図は以下のようになる。

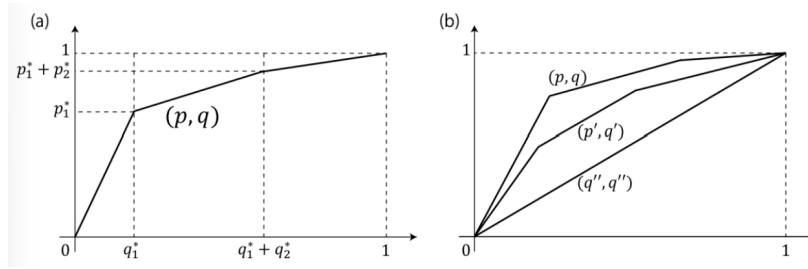


図 2 d-Majorization の図示

Def:d-Majorization

$p, q, p', q' \in \mathcal{P}_d$ に対して、

$$(p', q') \prec (p, q) \Leftrightarrow \text{「}(p, q) \text{ のローレンツ曲線が } (p', q') \text{ のローレンツ曲線の上に来る」} \quad (3.9)$$

が成り立つとき、 (p, q) は (p', q') を d-majorize するという。

d-majorization は、以下のような言い換えが可能である。

Thm:Blackwell の定理

$p, q, p', q' \in \mathcal{P}_d$ かつ q, q' がフルランクであるとき、以下は同値である。

1. $(p', q') \prec (p, q)$
- 2.

$$\forall t \in \mathbb{R} \quad \sum_{i=1}^d |p'_i - tq'_i| \leq \sum_{i=1}^d |p_i - tq_i| \quad (3.10)$$

3. 任意の下に凸な関数 f に対して、

$$\sum_{i=1}^d q'_i f\left(\frac{p'_i}{q'_i}\right) \leq \sum_{i=1}^d q_i f\left(\frac{p_i}{q_i}\right) \quad (3.11)$$

4. ある二重確率遷移行列 T が存在して、

$$p' = Tp \quad q' = Tq \quad (3.12)$$

が成り立つ。

Prf

後で証明を行う。

Def: Thermo-Majorization

$p, q, p' \in \mathcal{P}_d$ に対して、 p が q について p' を thermo-majorize するとは、

$$(p', q) \prec (p, q) \quad (3.13)$$

が成り立つことである。

(q, q) のローレンツ曲線は、直線であるため、 q は任意の p について thermo-majorize される。すなわち、

$$\forall p \quad (q, q) \prec (p, q) \quad (3.14)$$

である。したがって、thermo-majorization は、確率分布 p が q にどれだけ近いかを表す。

とくに、熱力学の文脈では q はあるハミルトニアン H の Gibbs 分布であり、 $q = p^G$ とかける。^{*2} このとき、 $q = Tq$ であるとは、 T は、Gibbs 分布を変えない遷移行列であるということである。このような遷移行列を、Gibbs-preserving map という。

リソース理論の枠組みでは、GPM は、free operation として扱われ、Gibbs 分布は、free state として扱われる。

Thm: 遷移可能条件

$p, q, p', q' \in \mathcal{P}_d$ とし、 q, q' がフルランクであるとする。このとき、以下の二つが成り立つ。

$$1. \quad (p', q') \prec (p, q) \Leftrightarrow S_0(p||q) \geq S_0(p'||q') \quad S_\infty(p||q) \leq S_\infty(p'||q') \quad (3.15)$$

$$2. \quad S_\infty(p'||q') \leq S_0(p||q) \Leftrightarrow (p', q') \prec (p, q) \quad (3.16)$$

Prf

以下の図より明らかである。

^{*2} 任意のフルランクな q はあるハミルトニアンの Gibbs 分布であるらしい。

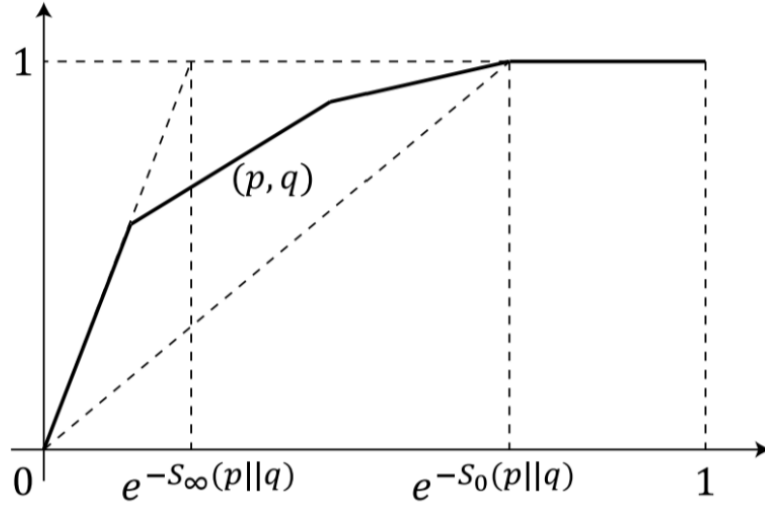


図3 遷移可能条件の図示

上の定理から明らかなように、Rényi divergence の観点からは、状態遷移と d-majorization に関する必要条件と十分条件は一致しない。

3.3 連続変数の場合

Majorization および d-majorization の連続変数の場合を考える。
簡単のため、 $x \in [0, 1]$ とする。

Def: 連続変数における Majorization

$p, p' \in L^1([0, 1])$ に対して、 p が p' を majorize することを、 $p' \prec p$ と書き、

$$\forall y \in [0, 1] \quad \int_0^y p'(x)^{\downarrow} dx \leq \int_0^y p(x)^{\downarrow} dx \quad (3.17)$$

が成り立つとき、 p が p' を majorize するという。ただし、

$$p^{\downarrow}(x) = \sup\{y : m_p(y) > x\} \quad (3.18)$$

$$m_p(y) = \mu[x : p(x) > y] \quad (3.19)$$

$$\mu : \text{Lebesgue measure} \quad (3.20)$$

である。^a

^a ルベグ積分が分からないせいで全然納得できない。

Def: 二重確率遷移写像

$T : L^1([0, 1]) \rightarrow L^1([0, 1])$ が二重確率遷移写像であるとは、

$$\forall p \in L^1([0, 1]) \quad Tp \prec p \tag{3.21}$$

が成り立つことである。

4 古典熱力学への適応

古典熱力学との対応を見ていく。

4.1 第二法則と KL ダイバージェンス

熱力学第二法則が、KL ダイバージェンスの単調性に由来することを確認する、 E_i を、系の i 番目の状態のエネルギーとし、Gibbs 状態 $p^G \in \mathcal{P}_d$ を、 $p_i^G = \frac{e^{-\beta E_i}}{Z}$ で定義する。ただし、 Z は正規化定数である。また、平衡状態での自由エネルギーは、 $F = -\beta^{-1} \log Z$ である。また、GPM は、

$$p^G = T p^G \quad (4.1)$$

であった。

Prop: 詳細つり合い条件と GPM

系が詳細つり合い条件

$$T_{ji} e^{-\beta E_i} = T_{ij} e^{-\beta E_j} \quad (4.2)$$

を満たすならば、 T は GPM である。

Prf

$$\sum_j T_{ij} p_j = \sum_j T_{ji} p_i \quad (4.3)$$

$$= p_i \quad (\because \sum_j T_{ji} = 1) \quad (4.4)$$

からわかる。 □

以下、 $\beta \geq 0$ を逆温度とする。KL ダイバージェンスの単調性を用いると、

$$S_1(p \| p^G) \geq S_1(Tp \| p^G) \quad (4.5)$$

$$(4.6)$$

となることがわかる。KL ダイバージェンスの定義より、

$$S_1(p||p^G) = \sum_{i=1}^d p_i \log \left(\frac{p_i}{p_i^G} \right) \quad (4.7)$$

$$= \sum_{i=1}^d p_i \log(p_i) + \beta \sum_{i=1}^d p_i E_i \quad (4.8)$$

$$= -S(p) + \beta \sum_{i=1}^d p_i E_i \quad (4.9)$$

同様にして、

$$S_1(Tp||p^G) = -S(Tp) + \beta \sum_{i=1}^d (Tp)_i E_i \quad (4.10)$$

$$(4.11)$$

である。したがって、

$$S(Tp) - S(p) \geq \beta \left(\sum_{i=1}^d E_i (Tp)_i - \sum_{i=1}^d E_i p_i \right) \quad (4.12)$$

である。右辺を観察してみると、遷移前後のエネルギーの期待値の差が現れていることがわかる。ここで、

$$Q = \sum_{i=1}^d E_i (Tp_i - \sum_{i=1}^d E_i p_i) \quad (4.13)$$

と定義すると、

$$\Delta S = S(Tp) - S(p) \geq \beta Q \quad (4.14)$$

と書くことができる。これは、クラウジウス不等式と呼ばれ、熱力学第二法則を表す。ただし、注意したいこととしては、左辺はシャノンエントロピーという、情報理論の文脈でのエントロピーである。

以上の議論を踏まえると、 T という、Gibbs 分布を保つ遷移がおこるとき、分布を保つために熱浴との相互作用が起こり、エントロピーが増大することがわかる。

Rem

上の熱について、 $Q_{ji} = E_j - E_i$ を、状態 j から i への遷移の際の熱の変化量とすると、

$$Q = \sum_{i,j} T_{ji} p_i Q_{ji} \quad (4.15)$$

と書くことができる。

確率熱力学の文脈では、

$$\Sigma = S_1(p||p^G) - S_1(Tp||p^G) = \Delta S - \beta Q \geq 0 \quad (4.16)$$

をエントロピー生成と呼ぶ。とくに、 $\Delta S - \beta Q$ は、熱浴のエントロピー増加ととらえられる。

つぎに、仕事と自由エネルギーについての第二法則を見ていく。いま、系のハミルトニアンは時間依存するとし、系への仕事はハミルトニアンの時間変化により実現されるとする。ここで、ハミルトニアンは、仕事浴めによって変化するのではなく、外部操作によって駆動すると考える。

まず、クエンチによって駆動される系を考える。いま、 $E = \sum_i E_i p_i$, $E' = \sum_i E'_i p_i$ がクエンチ前後での系のエネルギーの期待値あるとする。また、クエンチによってなされる仕事の平均は、

$$W = E' - E \quad (4.17)$$

と書かれる。

Rem

このセットアップでは、仕事は揺らぐ量であることに注意したい。すなわち、確率的な仕事を $w_i = E'_i - E_i$ として、

$$W = \sum_i w_i p_i \quad (4.18)$$

である。

次に、より一般的な熱力学過程を考える。全体の過程が、複数のクエンチおよび緩和過程によって構成されたと考える。クエンチとクエンチの間では、ハミルトニアンは一定であるとする。また、各緩和過程は確率的に独立であるとし、そのときのハミルトニアンに対応する GPM によって記述されると仮定する。以上の過程を図示すると以下ようになる。

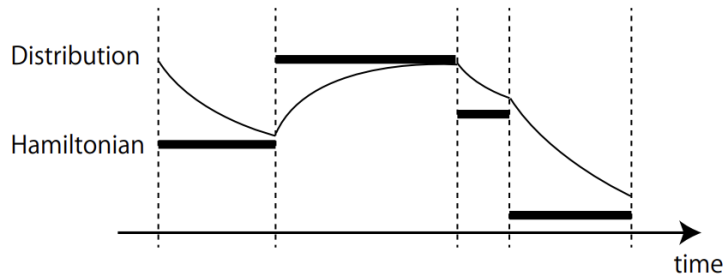


図4 熱力学過程の図示

全体の過程についての始状態と終状態を p, p' とし、ハミルトニアンを H, H' とする。また、Gibbs 状態を p^G, p'^G として、ヘルムホルツ自由エネルギーを F, F' とする。^{*3}

上の過程における、平均の仕事と熱について考える。 $\Delta E = E' - E$, $E = \sum_i E_i p_i$, $E' = \sum_i E'_i p_i$ とすると熱力学第一法則が成り立つ。すなわち、

$$\Delta E = W + Q \quad (4.19)$$

^{*3} この自由エネルギーは、熱平衡状態にある時の自由エネルギーであることに注意されたい。

が成立する。このとき、クエンチは、一瞬で行われるので、その間に確率分布は変化せず、また、熱の交換も起こらないとみなせる。したがって、この過程において、熱力学第一法則に影響はない。この第一法則を用いると、第二法則は、

$$W \geq \Delta E - \beta^{-1} \Delta S_1 \quad (4.20)$$

と書くことができる。ここで、非平衡状態における自由エネルギーを以下のように定義する。

$$F_1(p; H) = \Delta E_i - \beta^{-1} \Delta S_1(p) = \beta^{-1} S_1(p || p^G) + F \quad (4.21)$$

このとき、

$$F_1(p; H) = F \Leftrightarrow p = p^G \quad (4.22)$$

である。この自由エネルギーの表式を用いることにより、

$$W \geq F_1(p'; H') - F_1(p; H) \quad (4.23)$$

と書くことができる。この不等式は、仕事の揺らぎを許すときの仕事の下限を与える。

4.2 非平衡状態への遷移の要する仕事

非平衡状態へ遷移する際に必要な仕事を、Thermo-Majorization を用いて考える。

p^G を、系の Gibbs 分布とし、 $p_i^G = \frac{e^{-\beta E_i}}{Z}$ で定義する。また、 p を、系の任意の状態とする。また、今、仕事浴が、二つのエネルギー準位 $0, w$ を持つとする。仕事浴の Gibbs 分布を r^G とし、 $r^G = (\frac{1}{1+e^{-\beta w}}, \frac{e^{-\beta w}}{1+e^{-\beta w}})^T$ とし、また、仕事浴の任意の状態を $r = (r_0, r_w)^T$ $r_0 + r_w = 1$ とする。また、はじめと最後の仕事浴のエネルギーが 0 か w で与えられているとして、エネルギーの変化はすべて w か $-w$ であるとする。

以下、非平衡状態を作るための最小仕事を求める。ここで、系のハミルトニアンが初めと終わりで等しいことを仮定する。

始状態の系の確率分布は p^G であり、仕事浴の確率分布は、 $r^{up} = (0, 1)$ であるとする。このとき、SW 系 (着目系と仕事浴の合成系) の初期状態は、 $p^G \otimes r^{up}$ である。

また、終状態について、終状態の系の状態を p 、仕事浴の状態を $r^{down} = (1, 0)$ とする。このとき、SW 系の終状態は、 $p \otimes r^{down}$ である。

このとき、それぞれのテンソル積を書き下すと以下ようになる。

$$p^G \otimes r^{up} = (0, p_1^G, 0, p_2^G, \dots, 0, p_n^G) \quad (4.24)$$

$$p \otimes r^{down} = (p_1, 0, p_2, 0, \dots, p_n, 0) \quad (4.25)$$

また、ローレンツ曲線を書くために、 $p^G \otimes r^G$ を書き下すと、

$$p^G \otimes r^G = (p_1^G r^{G,down}, p_1^G r^{G,up}, p_2^G r^{G,down}, p_2^G r^{G,up}, \dots, p_n^G r^{G,down}, p_n^G r^{G,up}) \quad (4.26)$$

である。ここで、 $r^{G,up} = \frac{1}{1+e^{-\beta w}}$ 、 $r^{G,down} = \frac{e^{-\beta w}}{1+e^{-\beta w}}$ である。

以上の準備の下、ローレンツ曲線を描くことを考える。まず、初期状態についてのローレンツ曲線は、要素が0でないところから大きい順に足していけばよいことがわかる。また、終状態については、折れ線となる。これらを図示すると以下ようになる。

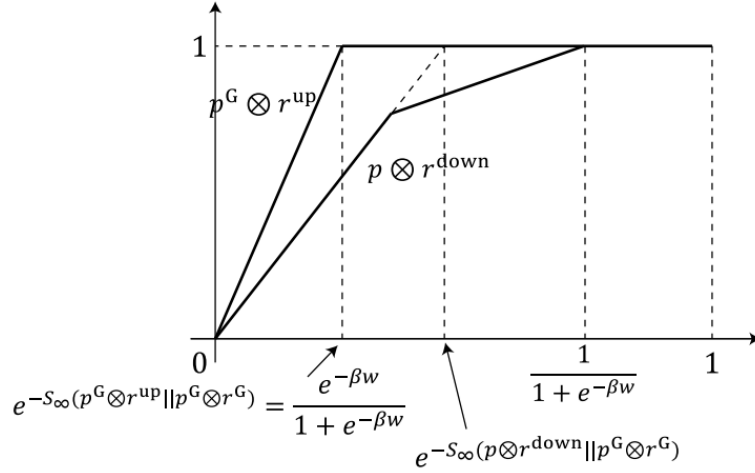


図5 非平衡状態への遷移の図示

したがって、非平衡状態への遷移が可能な必要十分条件は、

$$\frac{e^{-\beta w}}{1+e^{-\beta w}} \leq e^{-S_{\infty}(p \otimes r^{down} || p^G \otimes r^G)} \quad (4.27)$$

である。ここで、

$$S_{\infty}(p \otimes r^{down} || p^G \otimes r^G) = \ln \left(\max_i \frac{p_i \otimes r^{down}}{p_i^G \otimes r^{G,down}} \right) \quad (4.28)$$

$$= \ln \left(\max_i \frac{p_i}{p_i^G} \right) - \ln(r^{G,down}) \quad (4.29)$$

$$= S_{\infty}(p || p^G) - \ln(r^{G,down}) \quad (4.30)$$

であることを用いると、

$$\frac{e^{-\beta w}}{1+e^{-\beta w}} \leq \frac{1}{1+e^{-\beta w}} e^{-S_{\infty}(p || p^G)} \quad (4.31)$$

となる。整理して、

$$w \geq \beta^{-1} S_{\infty}(p || p^G) \quad (4.32)$$

となる。これが、平衡系から非平衡系への遷移に必要な最小仕事である。

4.3 熱平衡状態への緩和で取り出せる仕事

上と同様のことを考える。疲れたので図と結果だけを書く。

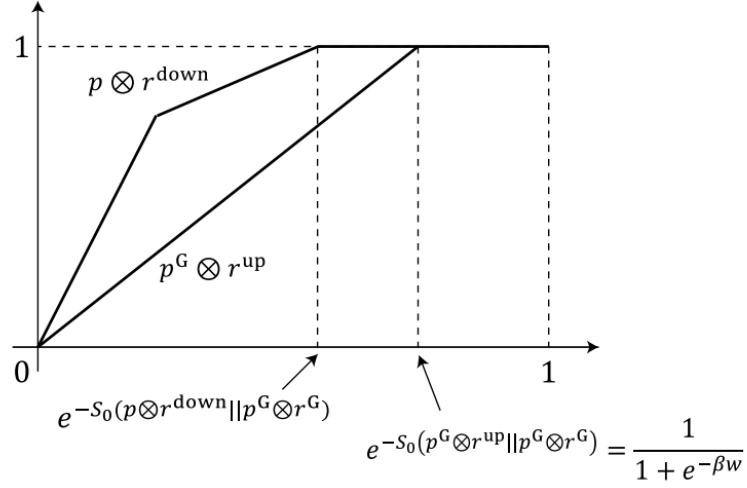


図 6 熱平衡状態への緩和の図示

$$-w \leq \beta^{-1} S_0(p || p^G) \quad (4.33)$$

である。これが、非平衡系から平衡系への遷移において取り出せる最大仕事である。

この等号が成り立つようなプロトコルが本文にあるが、クエンチしている時点でセットアップがよくわからん。外部系から操作しているように見えてしまう。

ただまあ、仕事の揺らぎがある系 (i.e. 確率熱力学に従う系) は、その揺らぎを許さない形よりも効率が悪い気がする (カルノーサイクルの隔離る熱力学での話を思い出す) ので、原理的限界としてとりあえずリソース理論を用いていて、たまたまそれを達成する操作を確率熱力学の文脈からも出せるという話である気もしてくる。

4.4 平衡状態間の遷移における仕事

最後に、Gibbs 分布間の状態の遷移を考える。このとき、clock と呼ばれる機構により、ハミルトニアンを時間変化させる。

系のハミルトニアンについて、clock が 0 のときは、 E_i 、clock が 1 のときは、 E'_i であるとする。

また、はじめと最後の確率分布を p, p' とする。ただし、

$$p_i = \frac{e^{-\beta E_i}}{Z} \quad (4.34)$$

$$p'_i = \frac{e^{-\beta E'_i}}{Z'} \quad (4.35)$$

である。

また、それに対応する自由エネルギーは、

$$F = -\beta^{-1} \log Z \quad (4.36)$$

$$F' = -\beta^{-1} \log Z' \quad (4.37)$$

である。

以下、簡単のために、clock は 0 か 1 の値のみを取るとする。初め、 $C(clock) = (1, 0)$ であり、終わりには $C(clock) = (0, 1)$ であるとする。

このとき、保存する Gibbs 分布は、

$$p_{SC}^G = \frac{Z}{Z + Z'} p \otimes c + \frac{Z'}{Z + Z'} p' \otimes c' \quad (4.38)$$

である。

∴

以上の準備の下、Majorization を用いる。 $p_{SC}^G \otimes r^{up}$ と $p_{SC}^G \otimes r^{down}$ について、Majorization を考える。このとき、どちらも Gibbs 分布であるからどちらも直線となる：

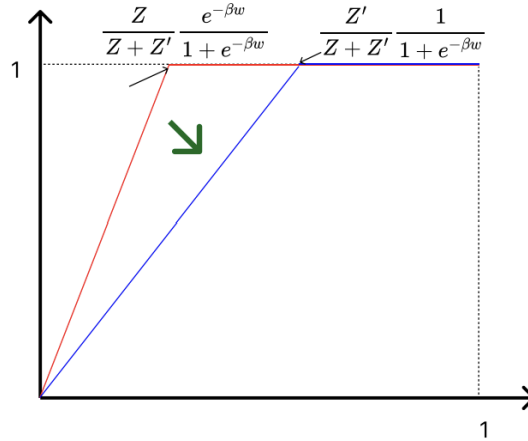


図 7 平衡状態間の遷移の図示

このとき p が p' を thermo-majorize する条件は、

$$\frac{Z}{Z + Z'} \frac{e^{-\beta w}}{1 + e^{-\beta w}} \leq \frac{Z'}{Z + Z'} \frac{1}{1 + e^{-\beta w}} \quad (4.39)$$

である。これを整理することで、

$$w \geq \Delta F \quad (4.40)$$

となる。これが、平衡状態間の遷移における最小仕事であり、たしかに熱力学第二法則が再現されることがわかる。