

情報とエントロピー

21/06/26 マルゼミ

Statistical mechanics is not really modern physics. It is pre-modern physics; it is modern physics; and the reader can be assured that it will be post-modern physics. The second law of thermodynamics will probably outlast anything that comes up.

-- Susskind

Agenda

情報とエントロピー

第一部 確率分布とエントロピー

第二部 エネルギーとエントロピー

第三部 ディープ・ラーニングとエントロピー

第四部 量子情報とエントロピー

Agenda

情報とエントロピー

第一部 確率分布とエントロピー

- シャノンのエントロピー
- ボルツマンのエントロピー

第二部 エネルギーとエントロピー

- 分配関数 Z
- 分配関数の利用

第三部 ディープ・ラーニングとエントロピー

- 相対エントロピーとその解釈
- ディープ・ラーニングと分配関数 Z の利用

第四部 量子情報とエントロピー

- ノイマン・エントロピー
- エネルギーと量子エントロピー

第一部

確率分布とエントロピー

1. シヤノン・エントロピー

2. ボルツマン・エントロピー

Agenda

情報とエントロピー

第一部 確率分布とエントロピー

- シャノンのエントロピー
- ボルツマンのエントロピー

第二部 エネルギーとエントロピー

- 分配関数 Z
- 分配関数の利用

第三部 ディープ・ラーニングとエントロピー

- 相対エントロピーとその解釈
- ディープ・ラーニングと分配関数 Z の利用

第四部 量子情報とエントロピー

- ノイマン・エントロピー
- エネルギーと量子エントロピー

第一部

確率分布とエントロピー 1

シャノン・エントロピー



シャノン・エントロピー

シャノン・エントロピー $H(X)$

$$\begin{aligned} H(X) &= H(x_1, x_2, \dots, x_n) = H(p_1, p_2, \dots, p_n) \\ &= - \sum_i p_i \log_2 p_i = \sum_i p_i \log_2 \frac{1}{p_i} \end{aligned}$$

シャノン・エントロピー $H(X)$

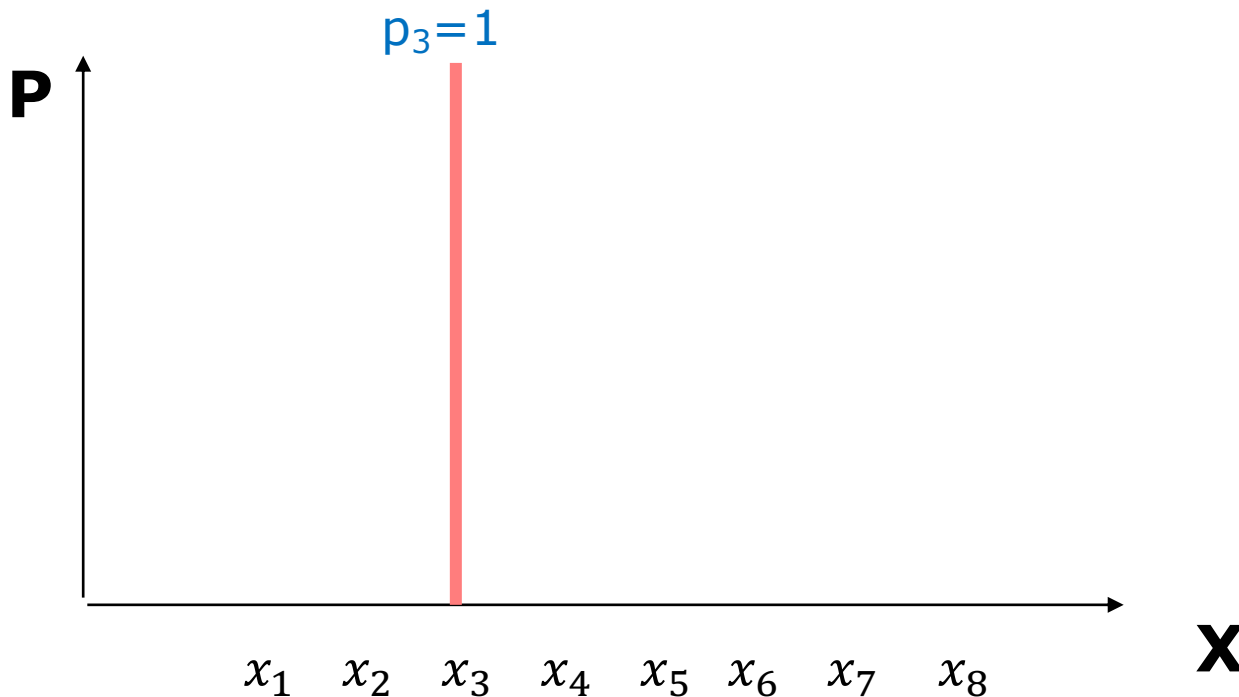
$$\begin{aligned} H(X) &= H(x_1, x_2, \dots, x_n) = H(p_1, p_2, \dots, p_n) \\ &= - \sum_i p_i \log_2 p_i = \sum_i p_i \log_2 \frac{1}{p_i} \end{aligned}$$

n 個の確率変数を取る値を $X = \{x_1, x_2, \dots, x_n\}$ 、その確率分布を $P = (p_1, p_2, \dots, p_n)$ とする時、上の $H(X)$ はこの確率分布のエントロピーを与える。

p_i は確率なので、 $\sum p_i = 1$ であることに注意しよう。

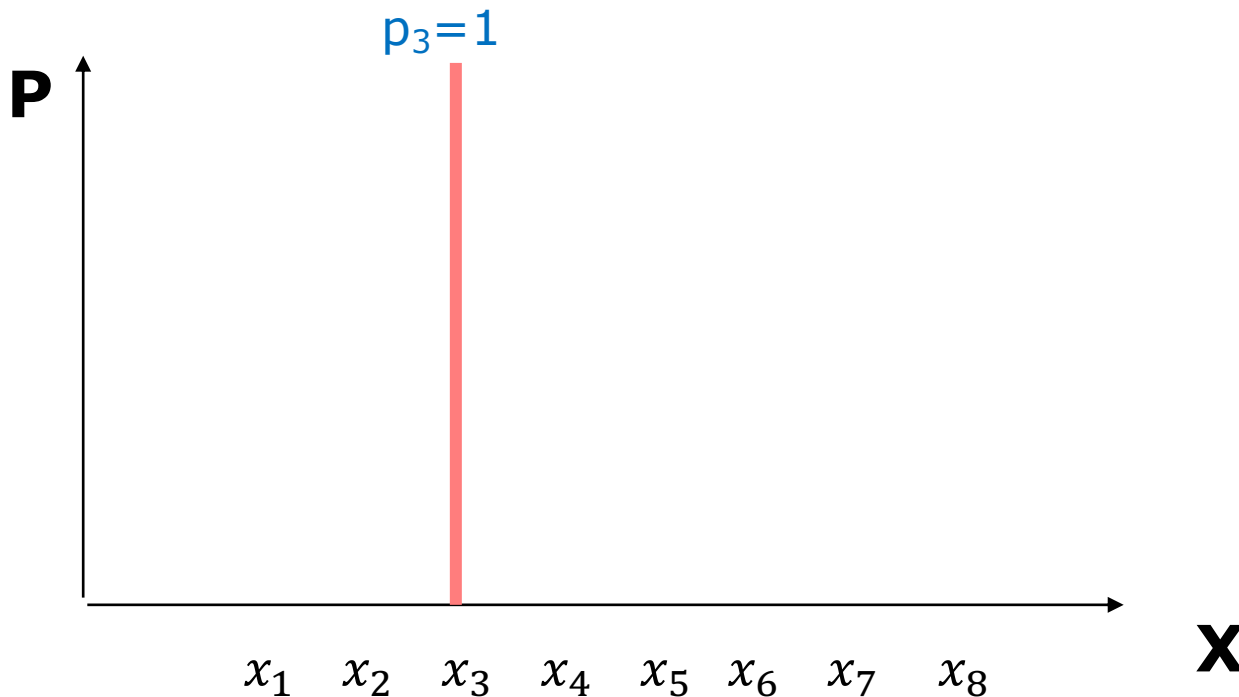
確率分布が与えるシャノン・エントロピーの例

一つの p_i のみが1の場合



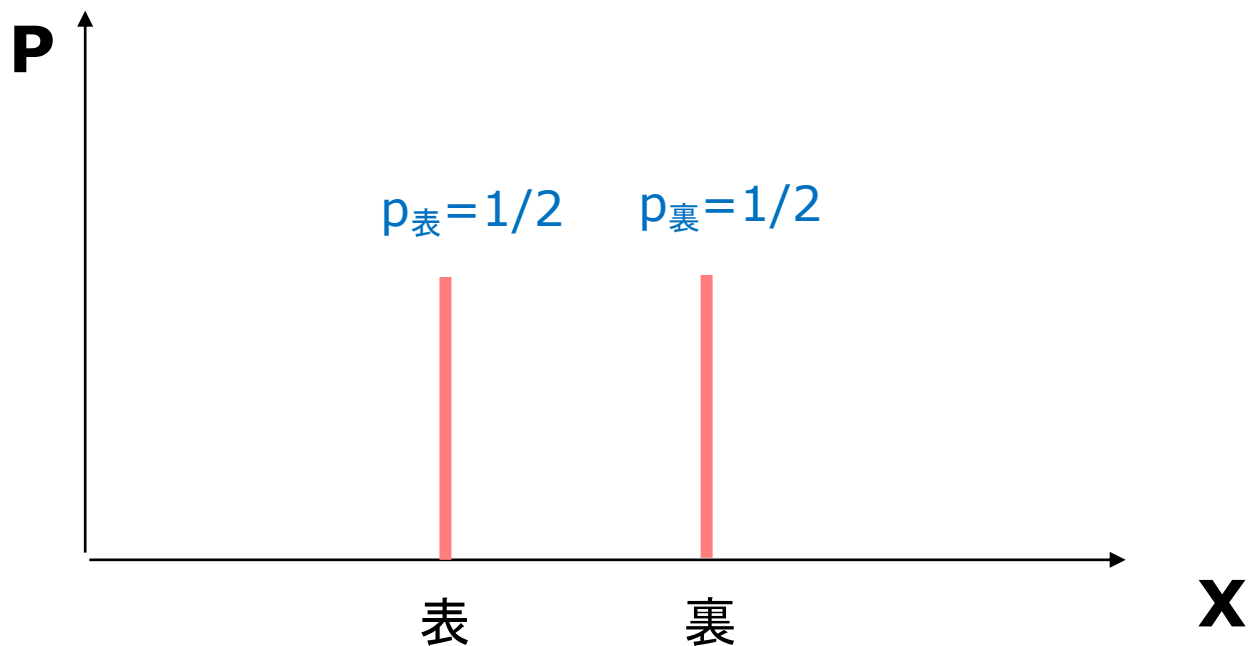
この例では、 X が x_3 の値を取る場合のみ、 $p_3=1$ である。
この時、 $H(X)=1 \cdot \log(1/1)=1 \cdot \log 1=1 \cdot 0=0$
(ただし、 $0 \cdot \log(1/0)=0$ とした)

一つの p_i のみが1の場合
それは、必ず x_i が起きることを意味する



ある事象 x_i が起きることの確率が1なら、それは必ず x_i が起きることを意味する。その時、エントロピーの値はゼロになる。
確実な情報のエントロピーは、ゼロである。

コイン・トスの確率分布

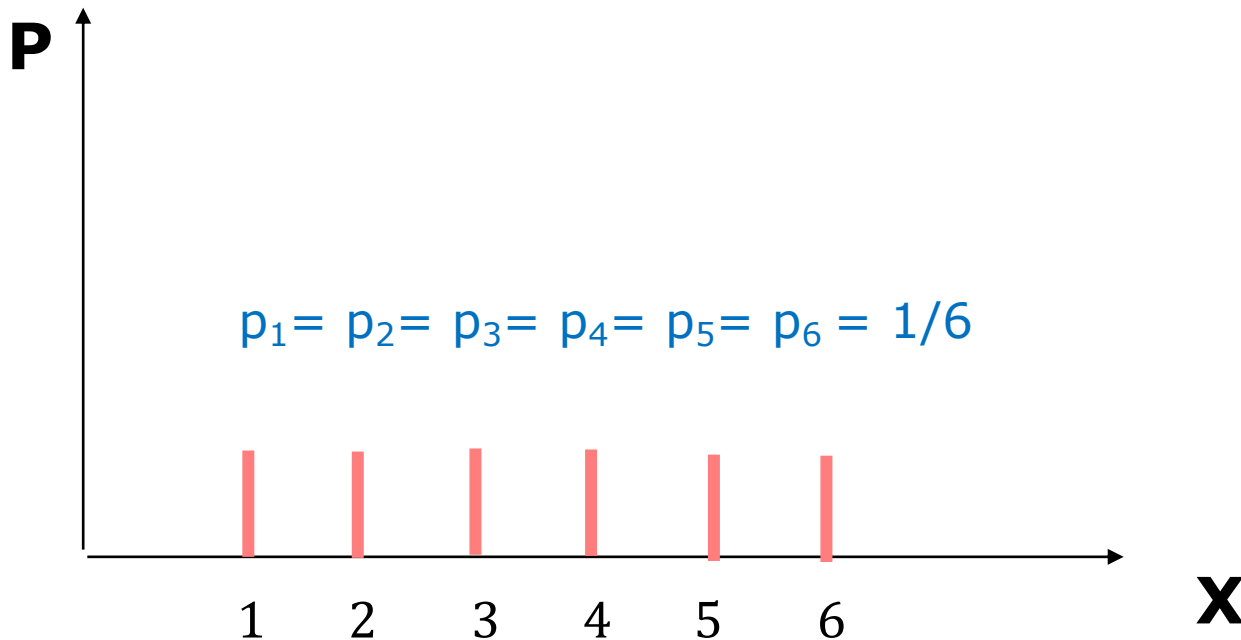


コインを投げた時、表が出る確率は $1/2$ 、裏が出る確率は $1/2$ 。

$$H(X) = 1/2 \cdot \log 2 + 1/2 \cdot \log 2 = \log 2 = 1$$

コイン・トスの確率分布のエントロピーは、1 bitである。

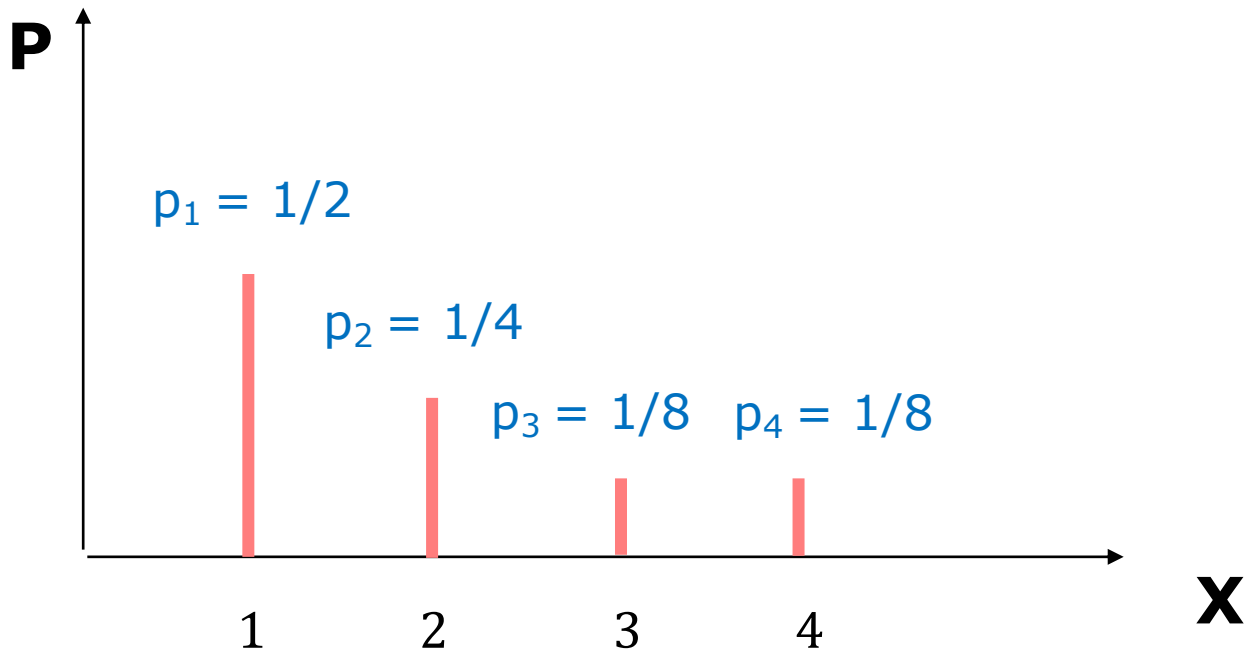
サイコロの出る目の確率分布



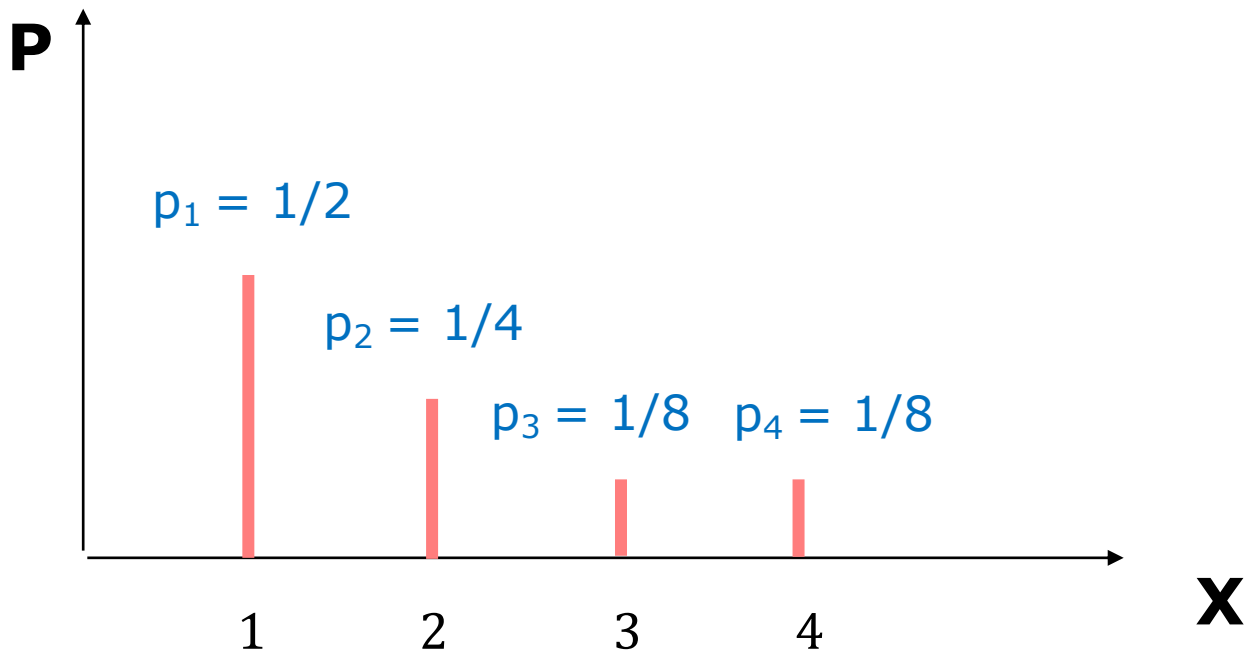
$p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = 1/6$ だから

$$H(X) = 1/6 \cdot (\log 6 + \log 6 + \log 6 + \log 6 + \log 6 + \log 6) = \log 6 \\ = 1 + \log 3 \text{ bit}$$

次のような確率分布が与えられた時の
エントロピーは？

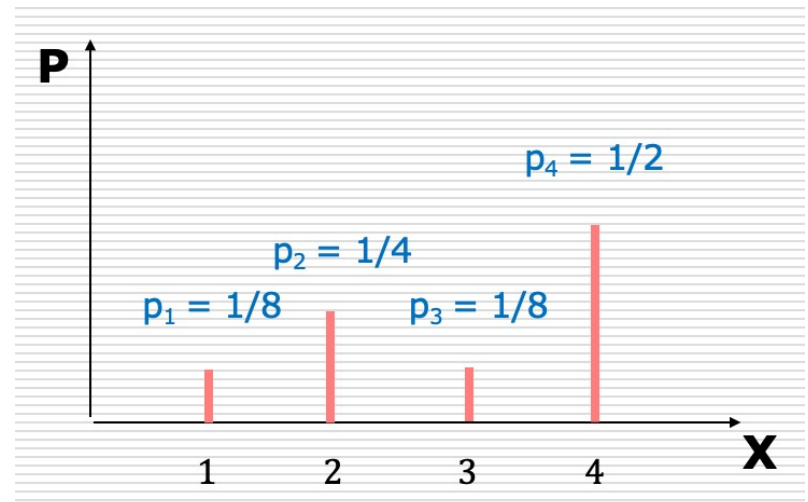
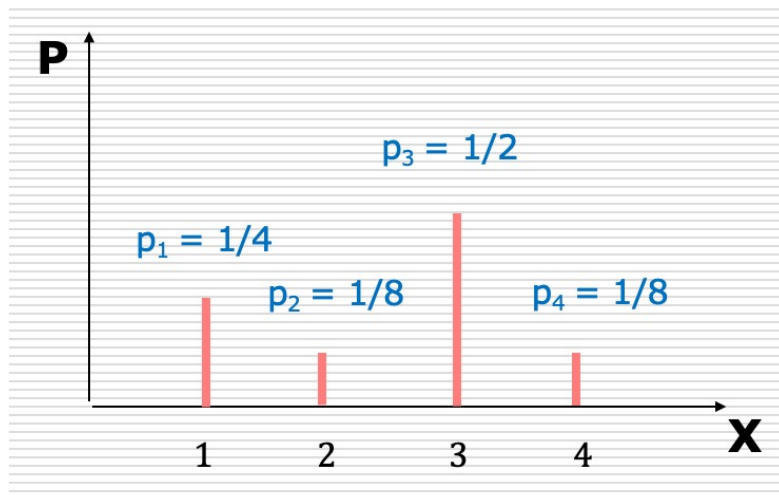
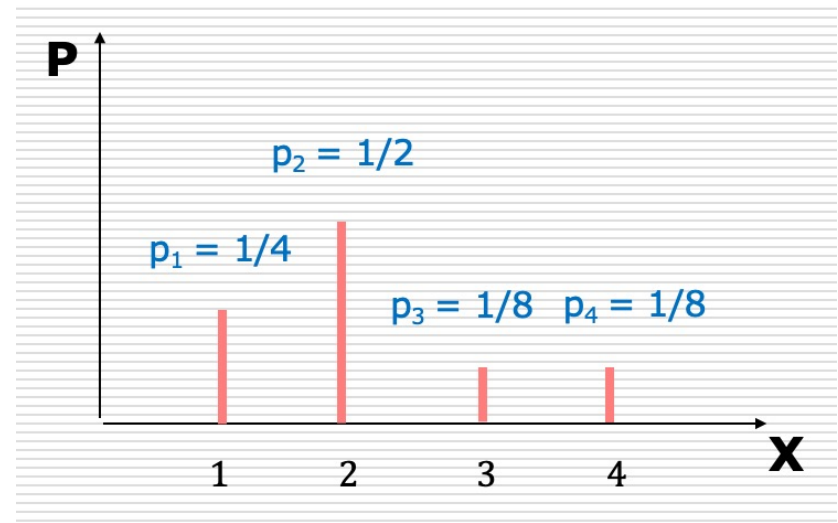
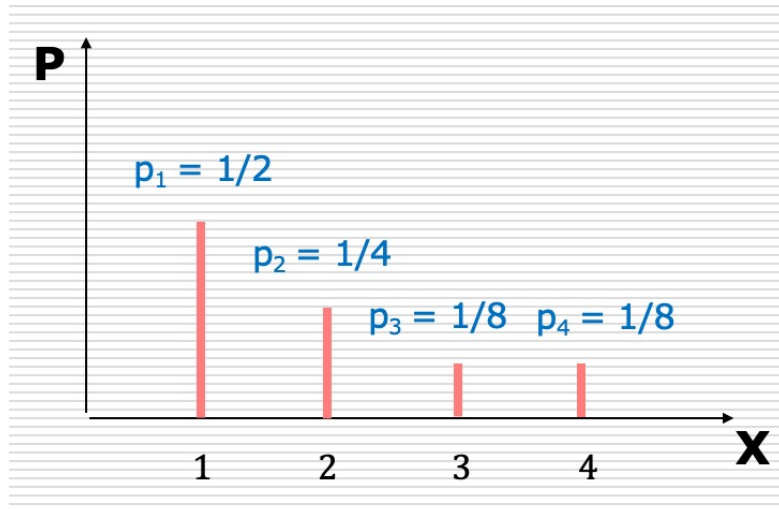


次のような確率分布が与えられた時の
エントロピーは 1.75 bit



$$\begin{aligned} H(X) &= 1/2 \cdot \log 2 + 1/4 \cdot \log 4 + 1/8 \cdot \log 8 + 1/8 \cdot \log 8 \\ &= 1/2 \cdot \log 2 + 1/4 \cdot \log 2^2 + 1/8 \cdot \log 2^3 + 1/8 \cdot \log 2^3 \\ &= 1/2 \cdot 1 + 1/4 \cdot 2 + 1/8 \cdot 3 + 1/8 \cdot 3 = 1/2 + 1/2 + 2 \cdot 3/8 \\ &= 1.75 \text{ bit} \end{aligned}$$

次の確率分布は、分布の形は違っていても
同じエントロピーを持つ



シャノン・エントロピーの導出

シャノン・エントロピー $H(X)$

$$H(X) = - \sum_i p_i \log_2 p_i = \sum_i p_i \log_2 \frac{1}{p_i}$$

シャノン・エントロピー $H(X)$ の定義

$$H(X) = - \sum_i p_i \log_2 p_i = \sum_i p_i \log_2 \frac{1}{p_i}$$

言語 L には、 I 個のアルファベット a_i が存在し、 L のテキスト内で、アルファベットの一文字 a_i が使われる出現確率を p_i が与えられているとしよう。

シャノン・エントロピー $H(X)$ の定義

$$H(X) = - \sum_i p_i \log_2 p_i = \sum_i p_i \log_2 \frac{1}{p_i}$$

言語 L には、 I 個のアルファベット a_i が存在し、 L のテキスト内で、アルファベットの一文字 a_i が使われる出現確率を p_i が与えられているとしよう。

L の長さ N のメッセージ(文字列)の持つエントロピー H を次のように定義する。

1. L の異なる典型的なメッセージを数え上げ、その数を W とする。
2. $H = \log_2 W$ で定義する。

異なる典型的なメッセージの数 W

N をメッセージの文字数とする。

メッセージを構成するのは、 l 個の文字からなるアルファベット $\{a_1, \dots, a_l\}$ である。

それぞれの文字の確率分布は、 $\{p_1, \dots, p_l\}$ で与えられている。

ある文字 a_i がメッセージに現れる個数を n_i とする。

ある文字 a_i がメッセージに現れる確率は $p_i = n_i/N$ なので、

メッセージ内の a_i の数 n_i は、 $n_i = p_i N$ であたえられる。

異なる典型的なメッセージの数 W

N をメッセージの文字数とする。

メッセージを構成するのは、 l 個の文字からなるアルファベット $\{a_1, \dots, a_l\}$ である。

それぞれの文字の確率分布は、 $\{p_1, \dots, p_l\}$ で与えられている。

ある文字 a_i がメッセージに現れる個数を n_i とする。

ある文字 a_i がメッセージに現れる確率は $p_i = n_i/N$ なので、

メッセージ内の a_i の数 n_i は、 $n_i = p_i N$ であたえられる。

この時、異なる典型的なメッセージの数 W は、次の式で与えられる。

$$\frac{N!}{(p_1 N)! (p_2 N)! \cdots (p_\ell N)!}$$

シャノンのエントロピー $H = \log_2 W$

シャノンのエントロピー $H = \log_2 W$ は、次の式で与えられる。

$$H = \log_2 W = \log_2 \left(\frac{N!}{(p_1 N)! (p_2 N)! \cdots (p_\ell N)!} \right)$$

シャノンのエントロピー $H = -\sum_i p_i \log_2 p_i$

実は、この式から、

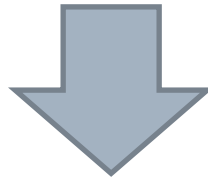
$$H = \log_2 W = \log_2 \left(\frac{N!}{(p_1 N)! (p_2 N)! \cdots (p_\ell N)!} \right)$$

次の式が導かれる。

$$H(X) = -\sum_i p_i \log_2 p_i = \sum_i p_i \log_2 \frac{1}{p_i}$$

シャノンのエントロピー Hの導出

$$H = \log_2 W = \log_2 \left(\frac{N!}{(p_1 N)! (p_2 N)! \cdots (p_\ell N)!} \right)$$



$$H(X) = - \sum_i p_i \log_2 p_i = \sum_i p_i \log_2 \frac{1}{p_i}$$

次の資料を参考にされたい。「情報とエントロピー入門」
<https://www.marulabo.net/docs/info-entropy/>

第一部

確率分布とエントロピー 2

ボルツマン・エントロピー



ミクロな状態の数と ボルツマン・エントロピー

ボルツマン・エントロピー S_B

$$S_B = k \log W$$

k は、ボルツマン定数

$$k = 1.38 \times 10^{-23}$$

W は、ミクロな状態の数

ボルツマン・エントロピー S_B

$$S_B = k \log W$$

k は、ボルツマン定数

$$k = 1.38 \times 10^{-23}$$

W は、ミクロな状態の数

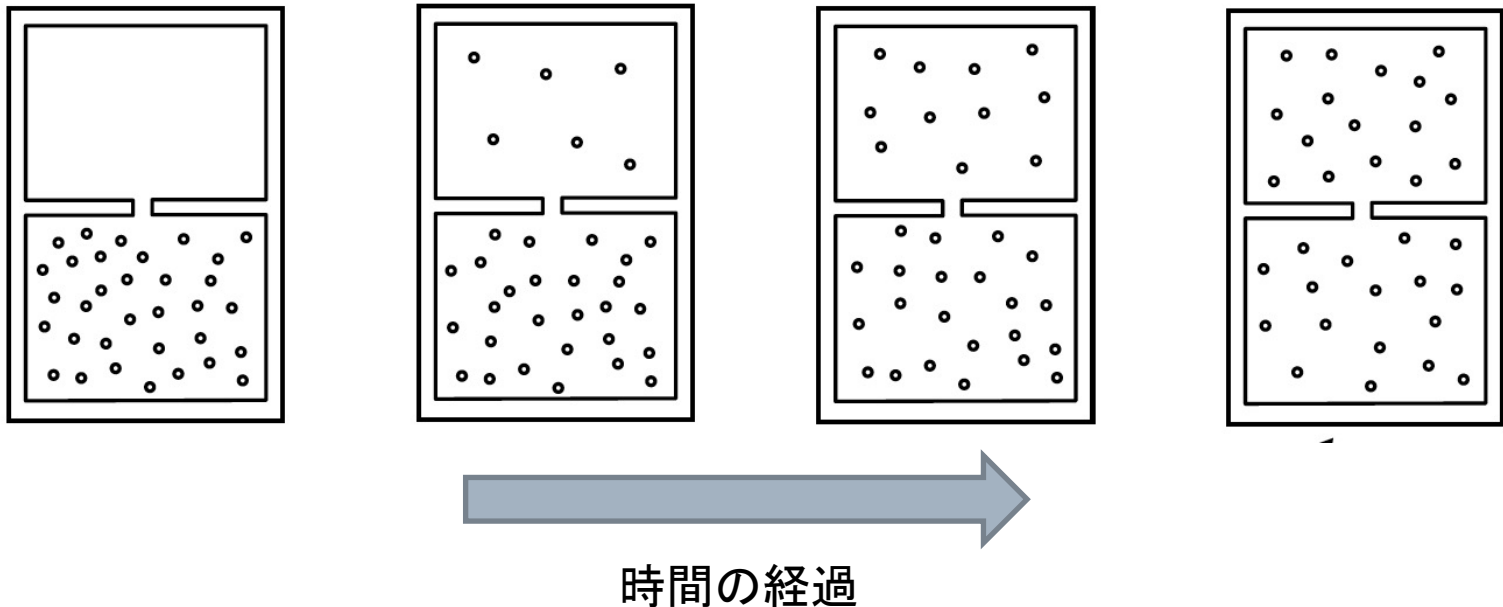
ミクロな状態を表す量と
マクロな状態を表す量

熱平衡状態

最初、二つの部分に分けられた一つの容器の一方にガスが溜まっていたとする。

時間とともに、ガスは容器の全体に広がっていく。

最終的には、ガスは容器全体に一様に広がり、平衡状態に達する。



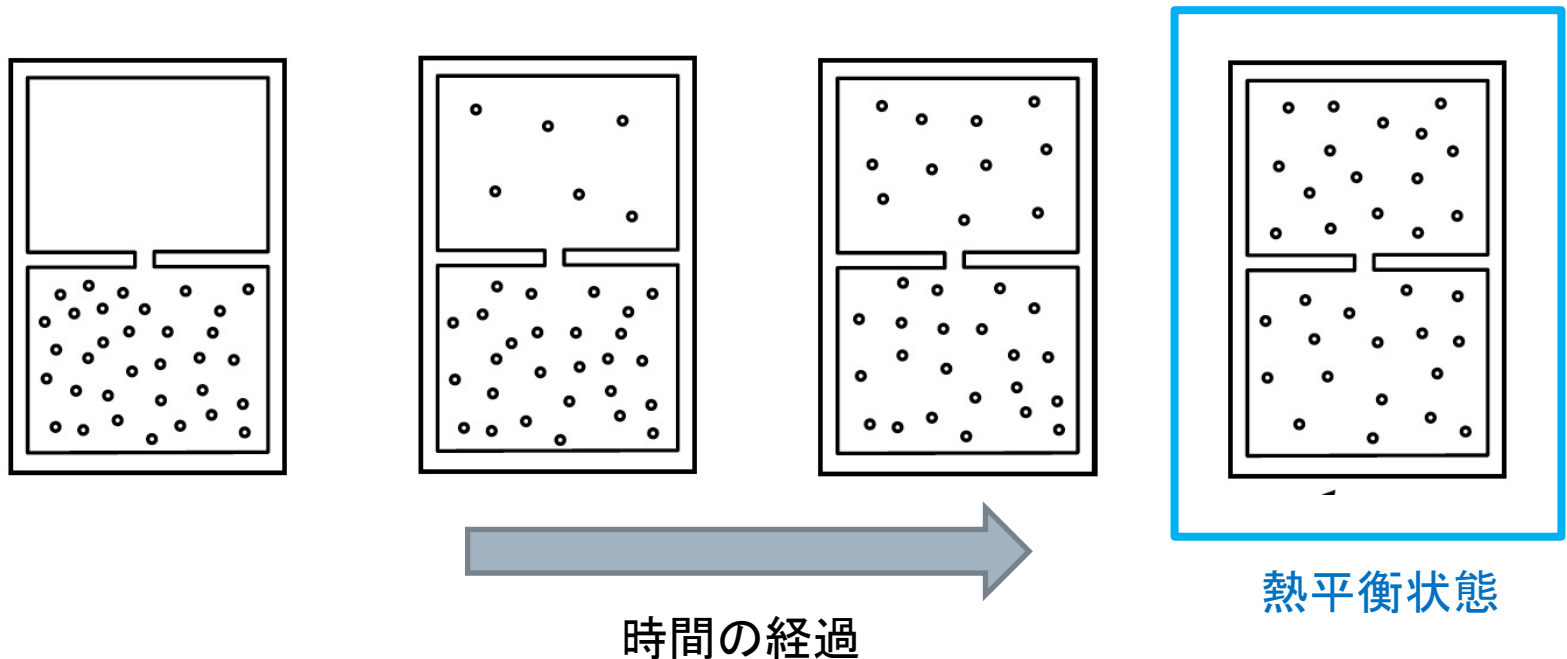
熱平衡状態

最初、二つの部分に分けられた一つの容器の一方にガスが溜まっていたとする。

時間とともに、ガスは容器の全体に広がっていく。

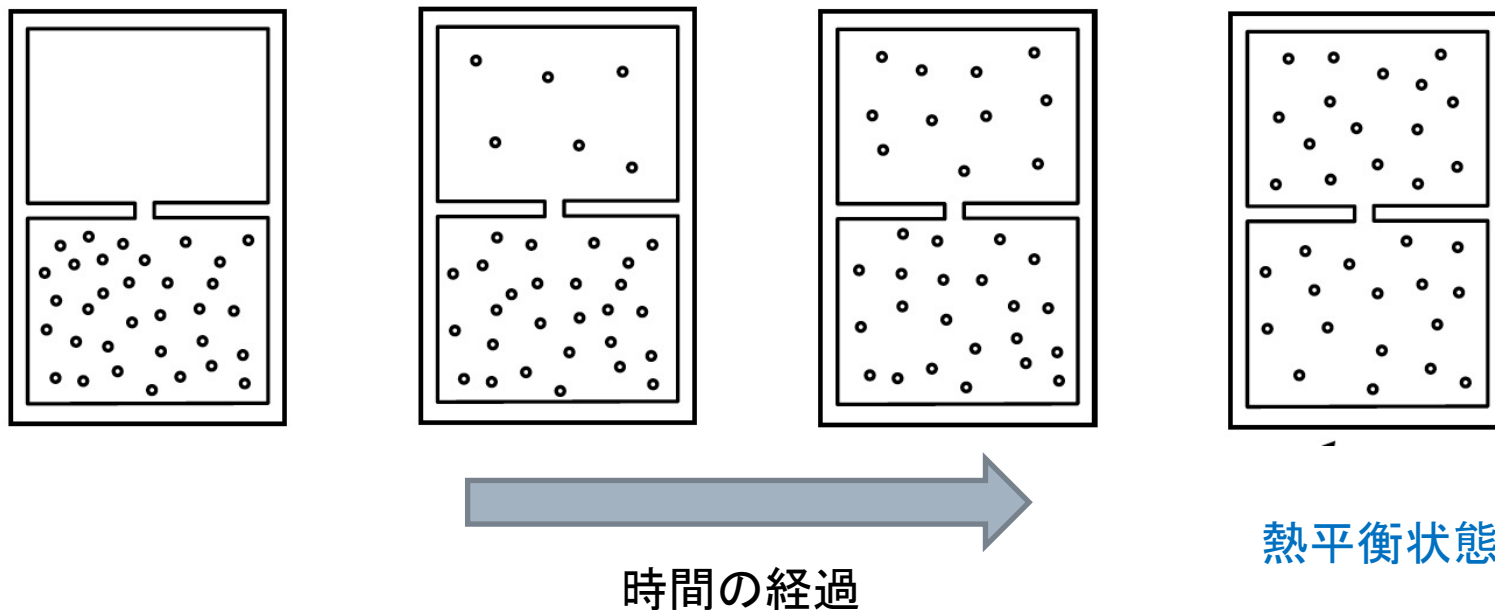
最終的には、ガスは容器全体に一様に広がり、平衡状態に達する。

この状態を「熱平衡状態」という。



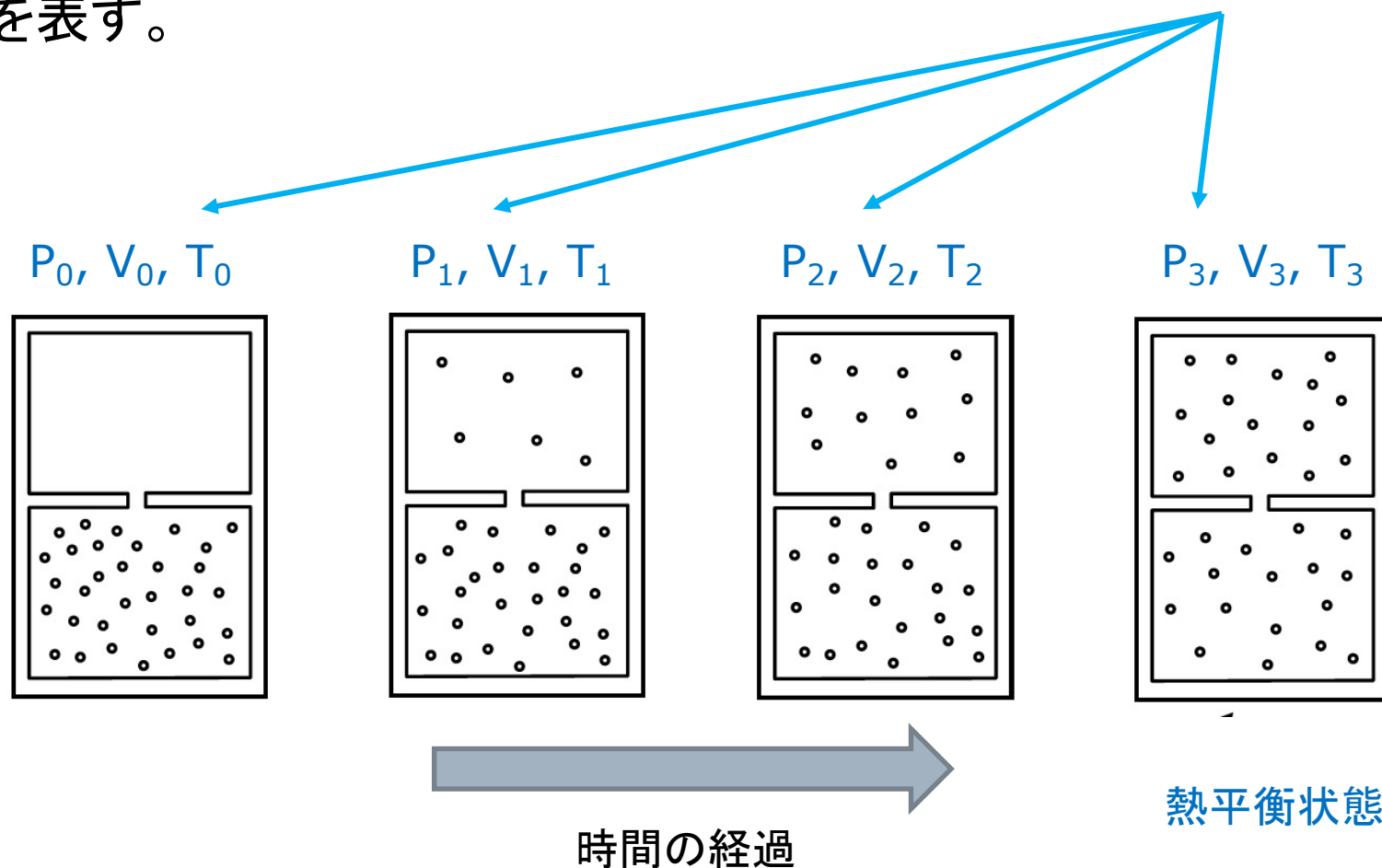
マクロな状態を表す量

容器の体積 V , ガスの圧力 P , ガスの温度 T 等が、マクロな状態を表す。



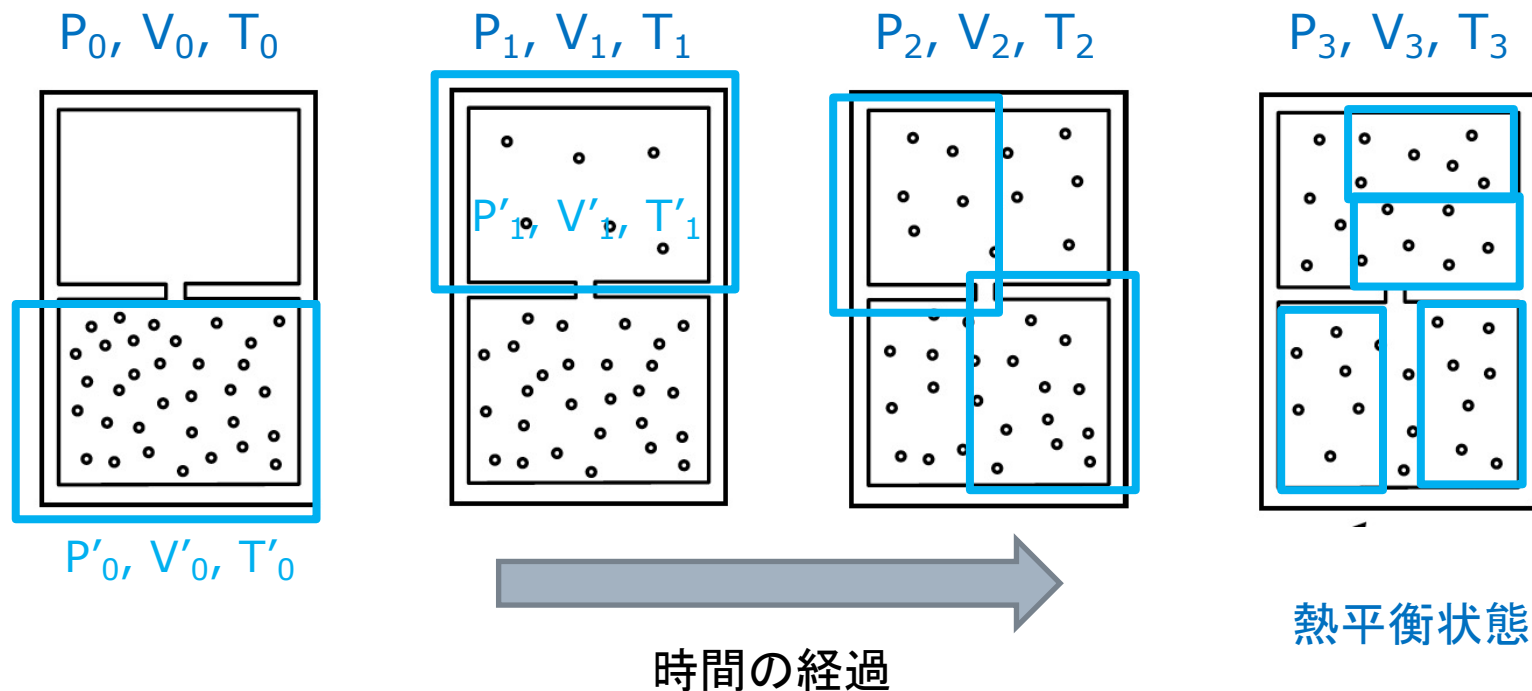
マクロな状態を表す量

容器の体積 V , ガスの圧力 P , ガスの温度 T 等が、マクロな状態を表す。



マクロな状態を表す量

容器の体積 V , ガスの圧力 P , ガスの温度 T 等が、マクロな状態を表す。マクロな量は、容器の部分でも観測できる。



ミクロな状態を表す量

直接には観測できないが、ガスを構成している一つ一つの分子の持つ状態を表す量をミクロな量という。

ミクロな状態を表す量

直接には観測できないが、ガスを構成している一つ一つの分子の持つ状態を表す量をミクロな量という。

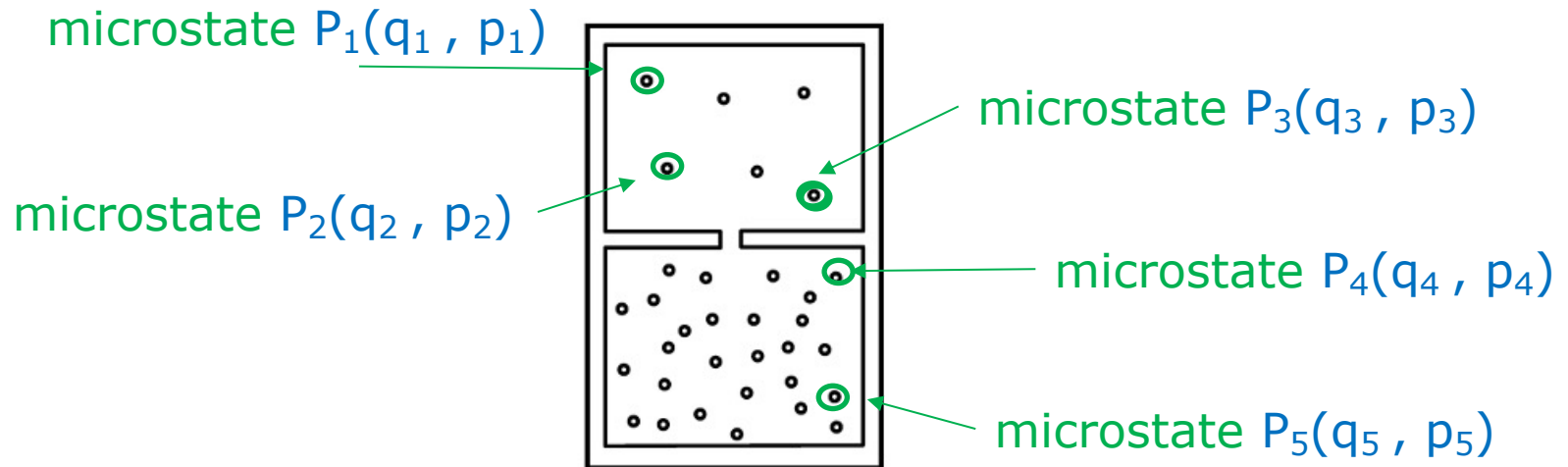
例えば、一つ一つの分子の「位置」 q 、一つ一つの分子の持つ「運動量」 p はミクロな状態を表す量である。

ミクロな状態を表す量

直接には観測できないが、ガスを構成している一つ一つの分子の持つ状態を表す量をミクロな量という。

例えば、一つ一つの分子の「位置」 q 、一つ一つの分子の持つ「運動量」 p はミクロな状態を表す量である。

一つ一つの分子のミクロな状態 P を (q, p) の組みで表すことができる。

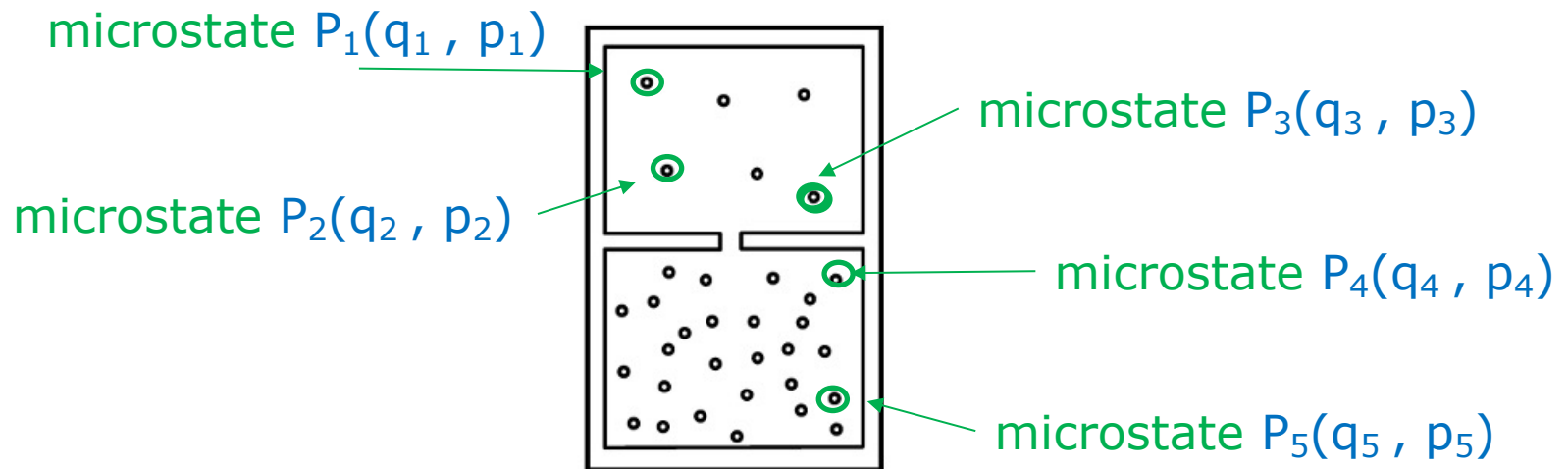


ミクロな状態を表す量

直接には観測できないが、ガスを構成している一つ一つの分子の持つ状態を表す量をミクロな量という。

例えば、一つ一つの分子の「位置」 q 、一つ一つの分子の持つ「運動量」 p はミクロな状態を表す量である。

一つ一つの分子のミクロな状態 P を (q, p) の組みで表すことができる。



ここでは5つの分子のマイクロ状態を図示したが、
全部の分子について、マイクロ状態を考える

ボルツマンが考えたこと

「ミクロな状態の上にマクロな状態が生まれる」

ボルツマンが考えたこと

原子論者ボルツマンが考えたことは、マクロな状態を扱う熱力学をミクロな状態の理論から導こうということだった。

ボルツマンが考えたこと

原子論者ボルツマンが考えたことは、マクロな状態を扱う熱力学をミクロな状態の理論から導こうということだった。

かけ離れているように見えるミクロとマクロの世界を、「ミクロな状態の上にマクロな状態が生まれる」という観点で結びつけようということだ。

ボルツマンが考えたこと

原子論者ボルツマンが考えたことは、マクロな状態を扱う熱力学をミクロな状態の理論から導こうということだった。

かけ離れているように見えるミクロとマクロの世界を、「ミクロな状態の上にマクロな状態が生まれる」という観点で結びつけようということだ。

次のように考える。

ボルツマンが考えたこと

原子論者ボルツマンが考えたことは、マクロな状態を扱う熱力学をミクロな状態の理論から導こうということだった。

かけ離れているように見えるミクロとマクロの世界を、「ミクロな状態の上にマクロな状態が生まれる」という観点で結びつけようということだ。

次のように考える。

- 一つ一つのミクロな状態は、正確に、一つのマクロの状態に対応している。

ボルツマンが考えたこと

原子論者ボルツマンが考えたことは、マクロな状態を扱う熱力学をミクロな状態の理論から導こうということだった。

かけ離れているように見えるミクロとマクロの世界を、「ミクロな状態の上にマクロな状態が生まれる」という観点で結びつけようということだ。

次のように考える。

- 一つ一つのミクロな状態は、正確に、一つのマクロの状態に対応している。
- 多くの異なるミクロな状態が、一つのマクロな状態に対応づけられる。

ミクロな状態とマクロな状態の関係

- 一つ一つのミクロな状態は、正確に、一つのマクロの状態に対応している。
- 多くの異なるミクロな状態が、一つのマクロな状態に対応づけられる。

ミクロな状態とマクロな状態の関係

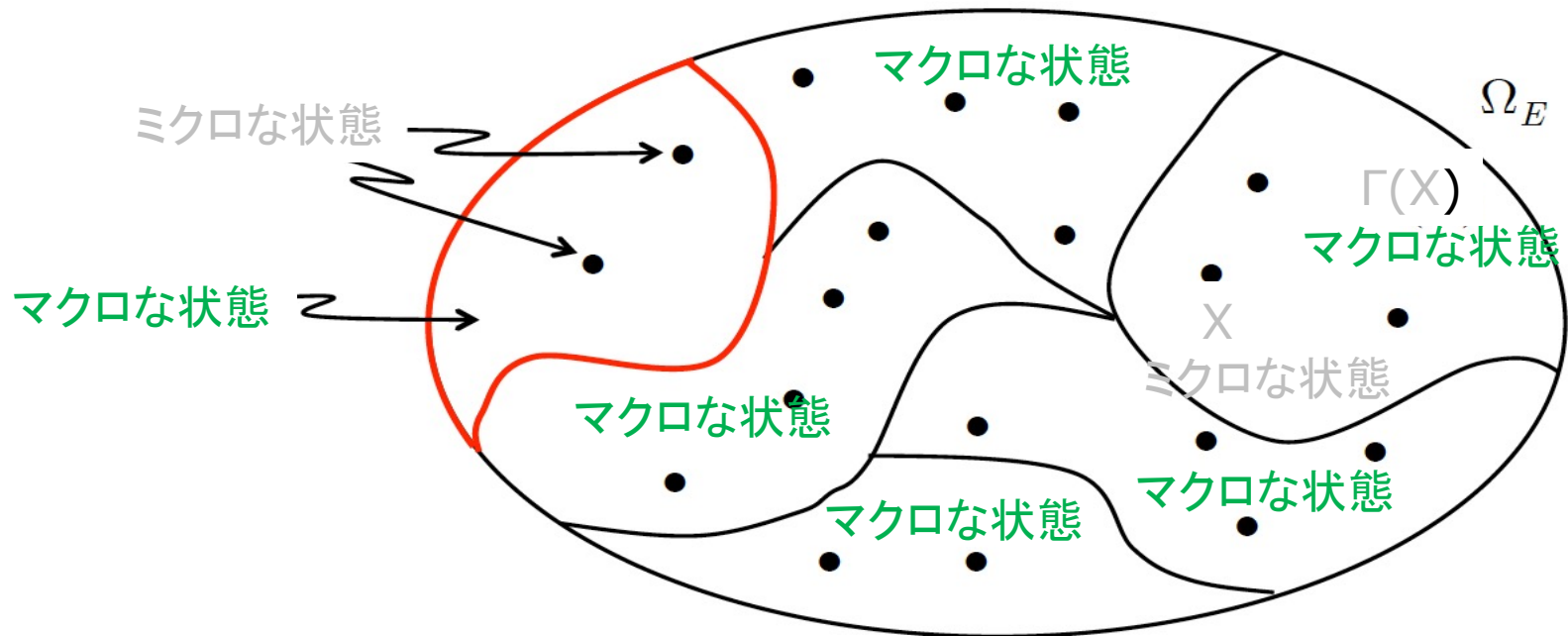
- 一つ一つのミクロな状態は、正確に、一つのマクロの状態に対応している。
- 多くの異なるミクロな状態が、一つのマクロな状態に対応づけられる。

領域を異なるマクロ状態に対応する領域に分解する。

ミクロな状態とマクロな状態の関係

- 一つ一つのミクロな状態 X は、正確に、一つのマクロの状態 $\Gamma(X)$ に対応している。
- 多くの異なるミクロな状態が、一つのマクロな状態に対応づけられる。

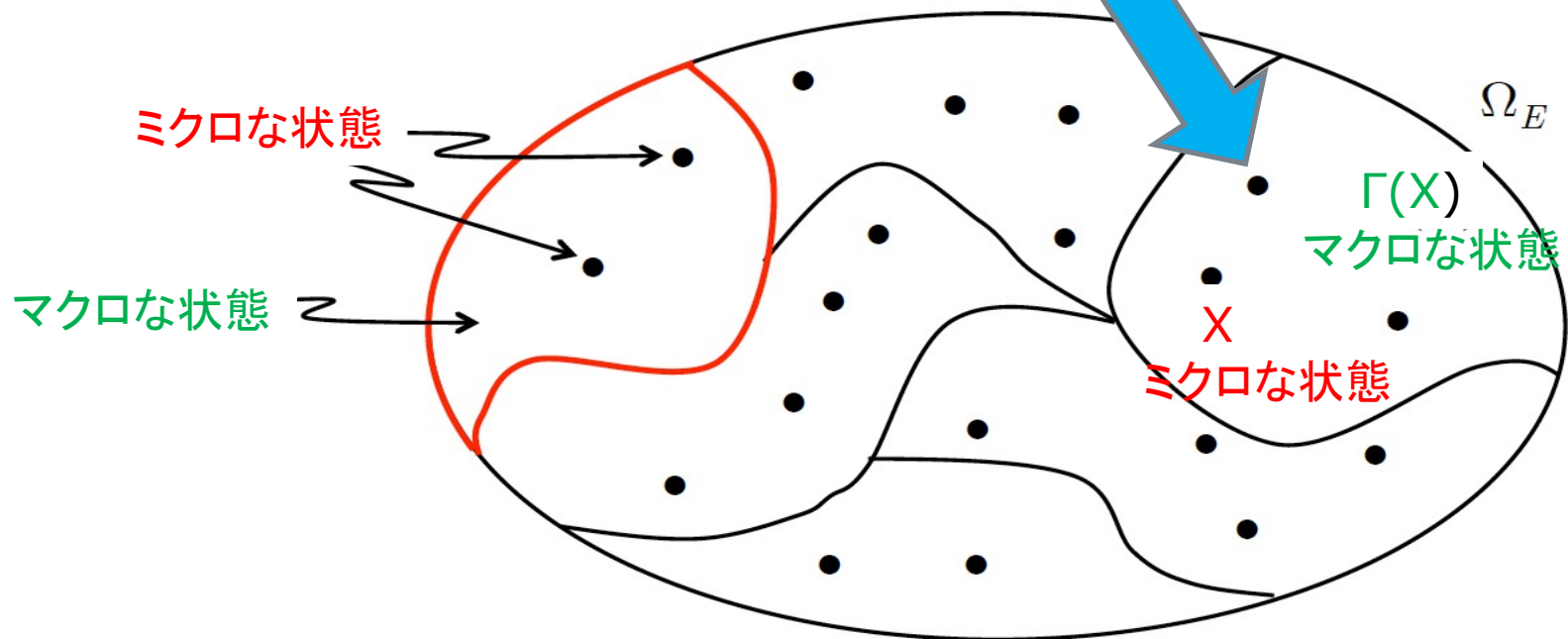
領域を異なるマクロ状態に対応する領域に分解する。



ミクロな状態とマクロな状態の関係

- 一つ一つのミクロな状態 X は、正確に、一つのマクロの状態 $\Gamma(X)$ に対応している。
- 多くの異なるミクロな状態が、一つのマクロな状態に対応づけられる。

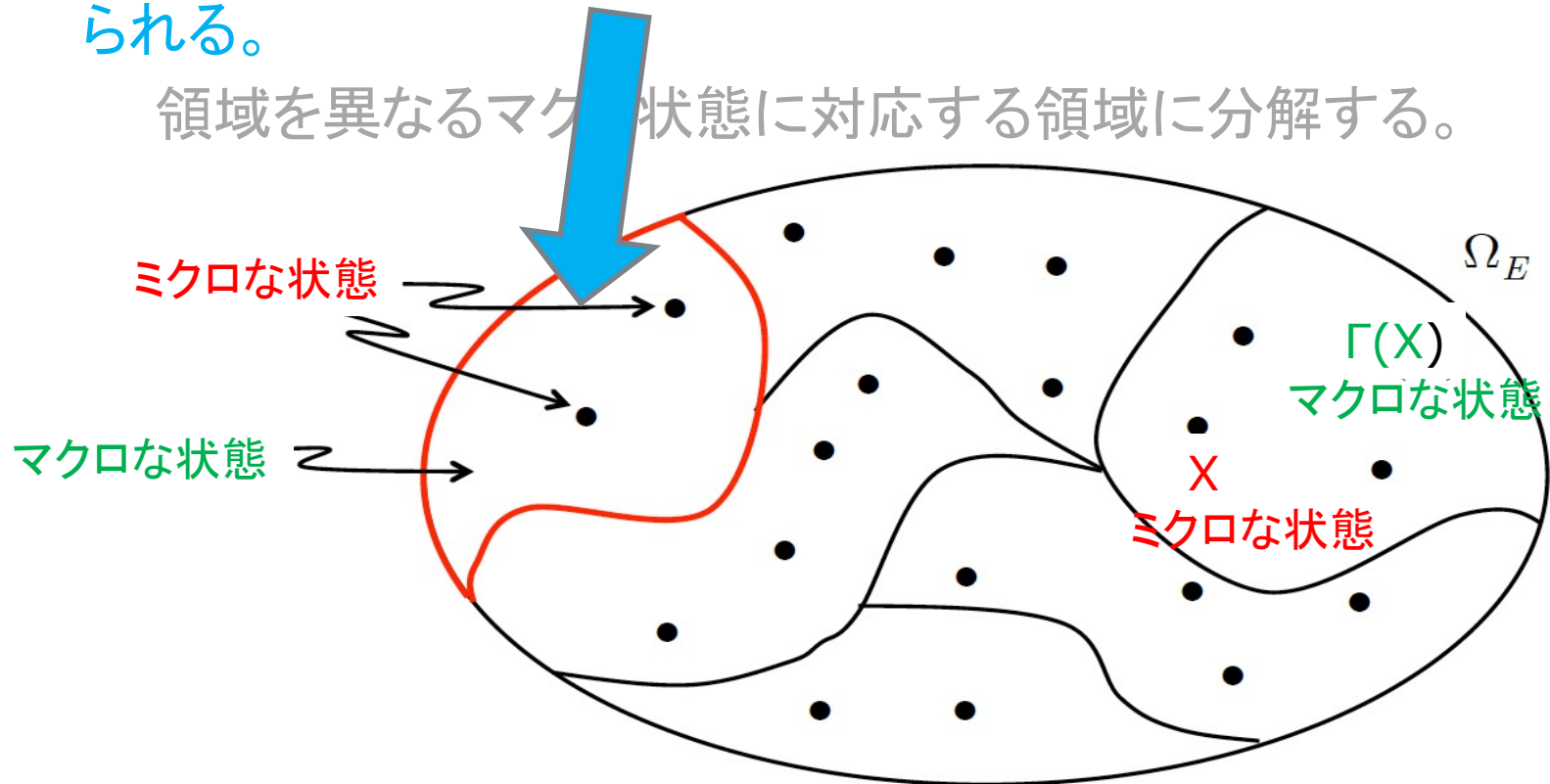
領域を異なるマクロ状態に対応する領域に分解する。



ミクロな状態とマクロな状態の関係

- 一つ一つのミクロな状態 X は、正確に、一つのマクロの状態 $\Gamma(X)$ に対応している。
- 多くの異なるミクロな状態が、一つのマクロな状態に対応づけられる。

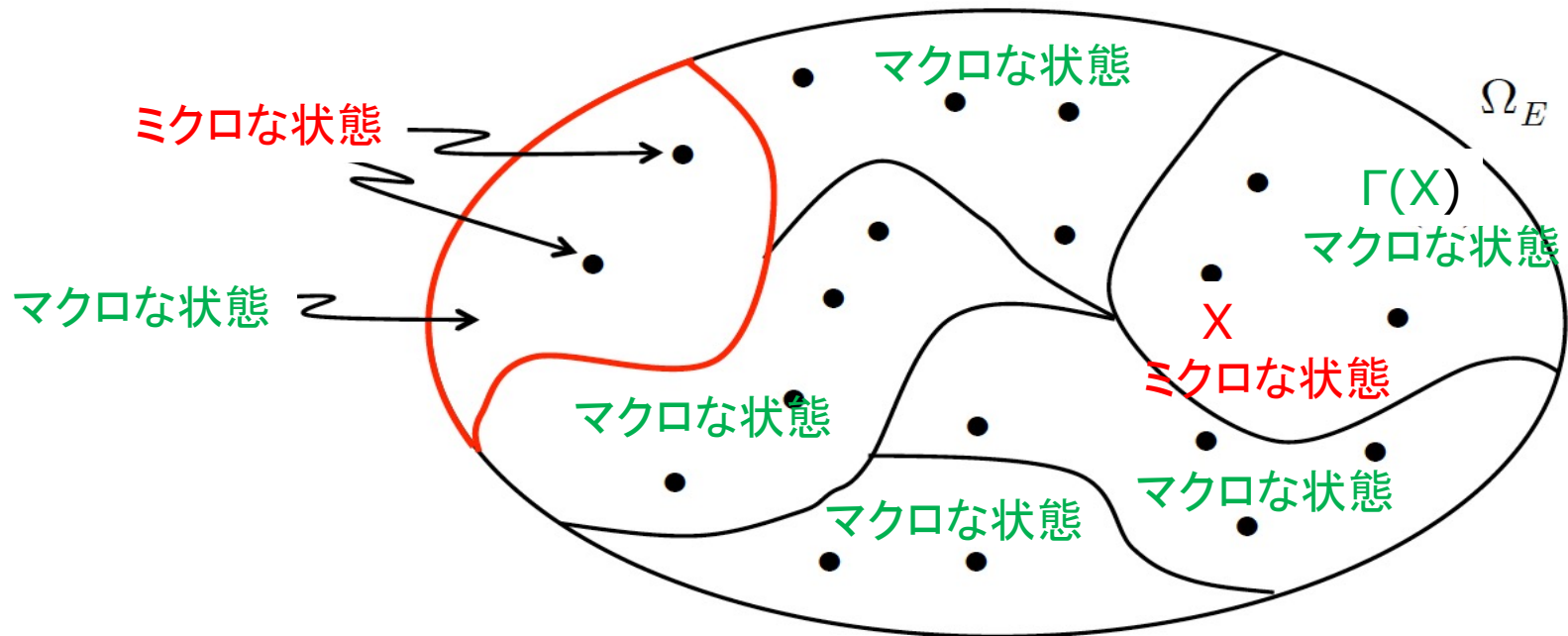
領域を異なるマクロ状態に対応する領域に分解する。



ミクロな状態とマクロな状態の関係

- 一つ一つのミクロな状態 X は、正確に、一つのマクロの状態 $\Gamma(X)$ に対応している。
- 多くの異なるミクロな状態が、一つのマクロな状態に対応づけられる。

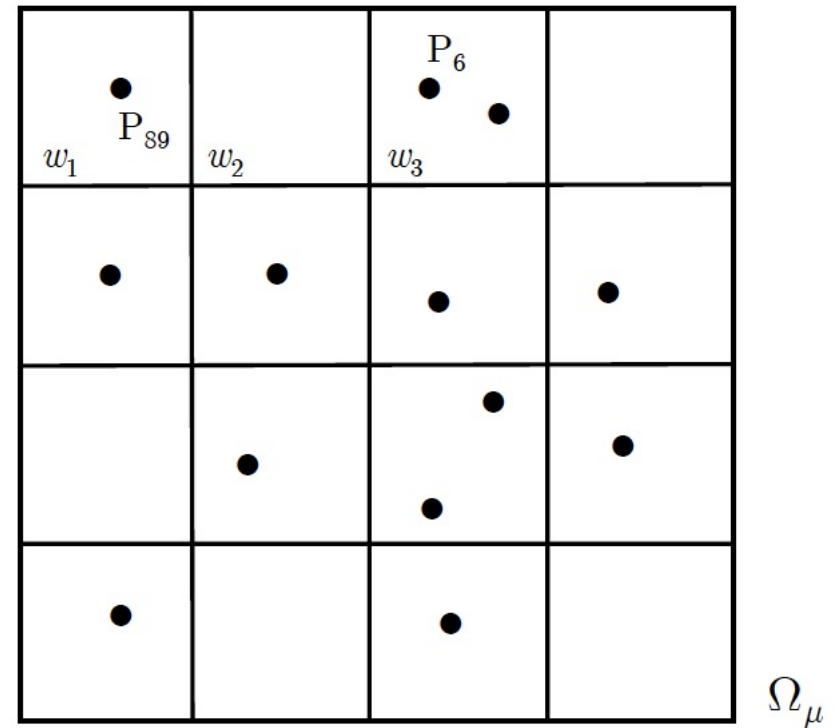
領域を異なるマクロ状態に対応する領域に分解する。



系がとりうる異なる
「ミクロの状態」の数 W を求める

N個の分子を、 ℓ 個のマクロ状態に配分する

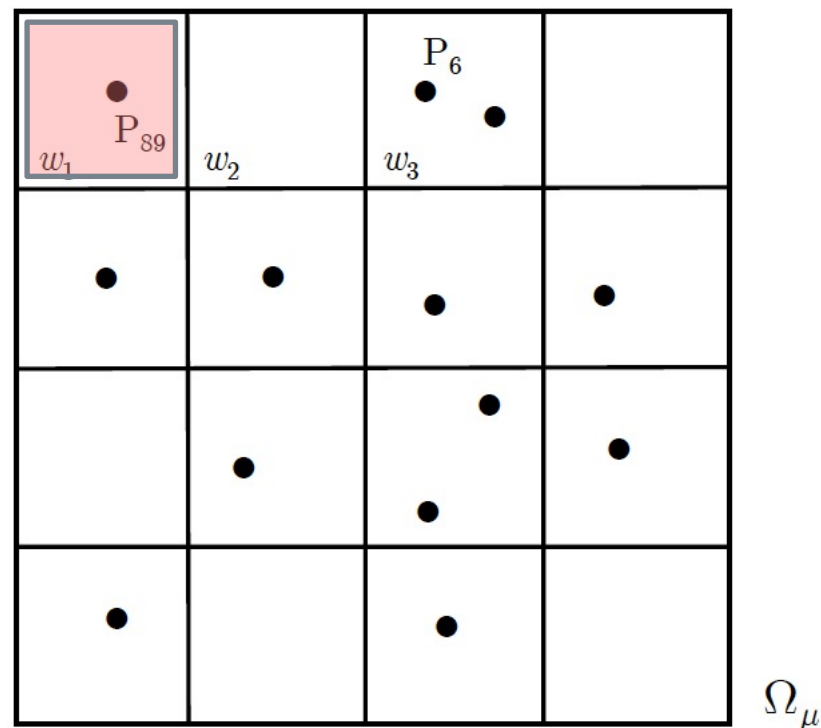
ℓ 個のマクロ状態を、 ℓ 個の小箱
 ℓ -cell で表そう。



N個の分子を、 ℓ 個のマクロ状態に配分する

ℓ 個のマクロ状態を、 ℓ 個の小箱
 ℓ -cell で表そう。

この図では、マクロ状態 w_1 に、
89番目の分子が、ミクロ状態 P_{89}
で含まれている。

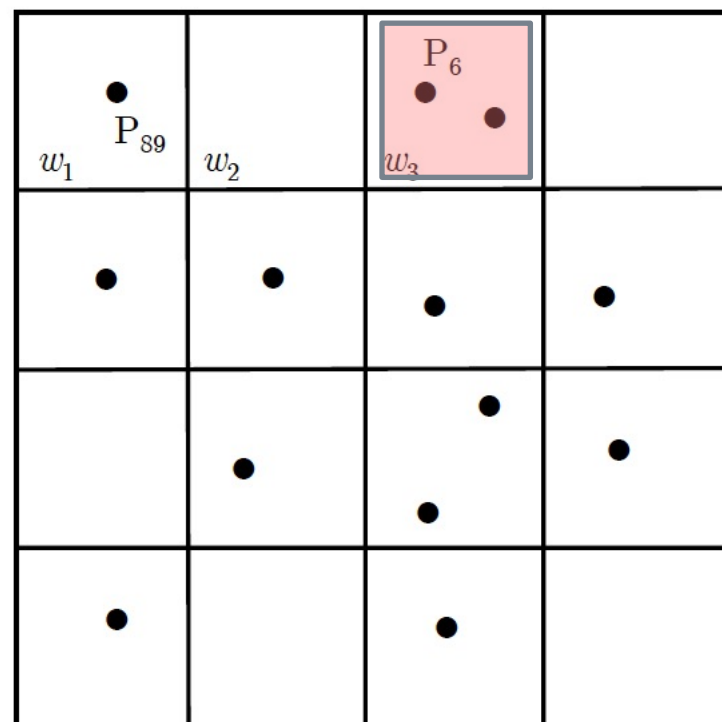


N個の分子を、 ℓ 個のマクロ状態に配分する

ℓ 個のマクロ状態を、 ℓ 個の小箱
 ℓ -cell で表そう。

この図では、マクロ状態 w_1 に、
89番目の分子が、ミクロ状態 P_{89}
で含まれている。

マクロ状態 w_3 には、
6番目の分子が、ミクロ状態 P_6
で含まれている。マクロ状態 w_3
には、もう一つの分子も含まれている。



Ω_μ

系がとりうる異なる 「ミクロの状態」の数 W を求める

N を、ミクロな分子の数とする。

マクロには、系を構成するのは、 ℓ 個のマクロ状態である。

$\{n_1, \dots, n_\ell\}$ は、それぞれのcellに含まれる分子の数である。

$p_i = n_i/N$ は、分子が i 番目のcellにある確率。

$\{p_1, \dots, p_\ell\}$ は、 ℓ -cell上の確率分布を与える。

i 番目のcell に含まれる分子の数は、 $n_i = p_i N$ で与えられる。

系がとりうる異なる 「ミクロの状態」の数 W を求める

N を、ミクロな分子の数とする。

マクロには、系を構成するのは、 ℓ 個のマクロ状態である。

$\{n_1, \dots, n_\ell\}$ は、それぞれのcellに含まれるミクロ状態の数である。

$p_i = n_i/N$ は、分子が i 番目のcellにある確率。

$\{p_1, \dots, p_\ell\}$ は、 ℓ -cell上の確率分布を与える。

i 番目のcell に含まれるミクロ状態の数は、 $n_i = p_i N$ で与えられる。

この時、系がとりうる異なる「ミクロの状態」の数 W は、次の式で与えられる。

$$\frac{N!}{(p_1 N)! (p_2 N)! \cdots (p_\ell N)!}$$

ボルツマン・エントロピー

ボルツマン・エントロピー S_B

$$S_B = k \log W$$

k は、ボルツマン定数

$$k = 1.38 \times 10^{-23}$$

W は、ミクロな状態の数

ボルツマン・エントロピー S_B

$$S_B = k \log W$$

k は、ボルツマン定数

$$k = 1.38 \times 10^{-23}$$

W は、ミクロな状態の数

$$W = \frac{N!}{(p_1 N)!(p_2 N)! \dots (p_l N)!}$$

ボルツマン・エントロピー S_B

$$S_B = k \log W = k \log \left\{ \frac{N!}{(p_1 N)! (p_2 N)! \dots (p_l N)!} \right\}$$

ボルツマン・エントロピー S_B

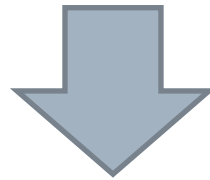
$$S_B = k \log W = k \log \left\{ \frac{N!}{(p_1 N)! (p_2 N)! \dots (p_l N)!} \right\}$$

この式から、
次の式が導かれる

$$S_B = -Nk \sum p_i \log p_i$$

ボルツマン・エントロピー S_B

$$S_B = k \log W = k \log \left\{ \frac{N!}{(p_1 N)! (p_2 N)! \dots (p_l N)!} \right\}$$



$$S_B = -Nk \sum p_i \log p_i$$

この導出については次の資料を参考にされたい。「情報とエントロピー入門」
<https://www.marulabo.net/docs/info-entropy/>

ボルツマン・エントロピーと シャノン・エントロピーの比較

ボルツマンとシャノンの エントロピーの比較

ボルツマン

- ガス
- 異なるミクロな状態を数え上げる
- N : ミクロな粒子の数
- $\{n_1, \dots, n_l\}$: l -cellへの分布
- $\{p_1, \dots, p_l\}$: cell上の確率分布
- $p_i = n_i/N$: ミクロな状態がcellにある確率
- $p_i N$: cell内のミクロ状態の数

$$S_B(\Gamma_{D_i}) = -Nk \sum_{j=1}^{\ell} p_j \ln p_j$$

$$\frac{N!}{(p_1 N)! (p_2 N)! \cdots (p_{\ell} N)!}$$

シャノン

- メッセージ
- 異なるメッセージを数え上げる
- N : メッセージの文字数
- $\{a_1, \dots, a_l\}$: l 文字のアルファベット
- $\{p_1, \dots, p_l\}$: 文字の確率分布
- $p_i = n_i/N$: a_i がメッセージにある確率
- $p_i N$: メッセージ内の a_i の数

$$H(X) = -\sum_{j=1}^{\ell} p_j \log_2 p_j$$





第二部

エネルギーとエントロピー

1. 分配関数 Z

2. 分配関数の利用

Agenda

情報とエントロピー

第一部 確率分布とエントロピー

- シャノンのエントロピー
- ボルツマンのエントロピー

第二部 エネルギーとエントロピー

- 分配関数 Z
- 分配関数の利用

第三部 ディープ・ラーニングとエントロピー

- 相対エントロピーとその解釈
- ディープ・ラーニングと分配関数 Z の利用

第四部 量子情報とエントロピー

- ノイマン・エントロピー
- エネルギーと量子エントロピー

第二部

エネルギーとエントロピー 1

分配関数 Z



マクロな状態にミクロな分子が含まれる確率

マクロな状態にミクロな分子が含まれる確率

N個の分子が、 ℓ 個のマクロな状態 $\{ w_1, w_2, \dots, w_\ell \}$ に分割されているとしよう。

w_i に含まれる分子の数を n_i とすると、

$$\sum_i n_i = N$$

w_i に分子が含まれる確率 p_i は、次のように与えられる。

$$p_i = \frac{n_i}{N}$$

また、

$$\sum_i p_i = \sum_i \frac{n_i}{N} = \frac{1}{N} \sum_i n_i = 1$$

である。

マクロな状態にミクロな分子が含まれる確率 と、そのエントロピー

よって、

$$\sum_i p_i = 1$$

この確率分布 $\{p_1, \dots, p_\ell\}$ が与えられた時、この系のエントロピー S は、次のようになる。

$$S = - \sum_i p_i \log p_i$$

系のエネルギーを考える

系のエネルギーを考える

今、 ℓ 個のマクロな状態 $\{ w_1, w_2, \dots, w_\ell \}$ の中で、一つの分子の持つ平均エネルギーを $\{ E_1, E_2, \dots, E_\ell \}$ とし、系全体の中で一つの分子の持つ平均エネルギーを E とする。

系のエネルギーを考える

今、 ℓ 個のマクロな状態 $\{ w_1, w_2, \dots, w_\ell \}$ の中で、一つの分子の持つ平均エネルギーを $\{ E_1, E_2, \dots, E_\ell \}$ とし、系全体の中で一つの分子の持つ平均エネルギーを E とする。

それぞれのマクロな状態に含まれる分子の数は n_i だから、それぞれのエネルギーは $\{ n_1 E_1, n_2 E_2, \dots, n_\ell E_\ell \}$ で、また、全体のエネルギーは、分子の総数は N だから NE になる。

系のエネルギーを考える

今、 ℓ 個のマクロな状態 $\{ w_1, w_2, \dots, w_\ell \}$ の中で、一つの分子の持つ平均エネルギーを $\{ E_1, E_2, \dots, E_\ell \}$ とし、系全体の中で一つの分子の持つ平均エネルギーを E とする。

それぞれのマクロな状態に含まれる分子の数は n_i だから、それぞれのエネルギーは $\{ n_1 E_1, n_2 E_2, \dots, n_\ell E_\ell \}$ で、また、全体のエネルギーは、分子の総数は N だから NE になる。

前者の和は、後者に等しいので

$$\sum_i n_i E_i = NE$$

系のエネルギーを考える

今、 ℓ 個のマクロな状態 $\{ w_1, w_2, \dots, w_\ell \}$ の中で、一つの分子の持つ平均エネルギーを $\{ E_1, E_2, \dots, E_\ell \}$ とし、系全体の中で一つの分子の持つ平均エネルギーを E とする。

それぞれのマクロな状態に含まれる分子の数は n_i だから、それぞれのエネルギーは $\{ n_1 E_1, n_2 E_2, \dots, n_\ell E_\ell \}$ で、また、全体のエネルギーは、分子の総数は N だから NE になる。

前者の和は、後者に等しいので

$$\sum_i n_i E_i = NE$$

両辺を N で割って

$$\sum_i \frac{n_i}{N} E_i = \sum_i p_i E_i = E$$

系のエネルギーを考える

今、 ℓ 個のマクロな状態 $\{ w_1, w_2, \dots, w_\ell \}$ の中で、一つの分子の持つ平均エネルギーを $\{ E_1, E_2, \dots, E_\ell \}$ とし、系全体の中で一つの分子の持つ平均エネルギーを E とする。

それぞれのマクロな状態に含まれる分子の数は n_i だから、それぞれのエネルギーは $\{ n_1 E_1, n_2 E_2, \dots, n_\ell E_\ell \}$ で、また、全体のエネルギーは、分子の総数は N だから NE になる。

前者の和は、後者に等しいので

$$\sum_i n_i E_i = NE$$

両辺を N で割って

$$\sum_i \frac{n_i}{N} E_i = \sum_i p_i E_i = E$$

エントロピーが最大になる条件を求める

エントロピーが最大になる条件と 拘束条件

エントロピー S が最大になる条件を考える。

$$S = - \sum_i p_i \log p_i$$

計算の都合上、次の $-S$ が最小になる条件を考える。

$$-S = \sum_i p_i \log p_i$$

ただし、次のような二つの拘束条件がある。

$$\sum_i p_i = 1$$
$$\sum_i p_i E_i = E$$

エントロピーが最大になる条件と 拘束条件

二つの拘束条件は、次の形に書ける。

$$\sum_i p_i - 1 = 0$$

$$\sum_i p_i E_i - E = 0$$

エントロピーが最大になる条件と 拘束条件

二つの拘束条件は、次の形に書ける。

$$\sum_i p_i - 1 = 0$$

$$\sum_i p_i E_i - E = 0$$

この時、任意の α, β について、 $[\]$ の中はゼロだから、次の式が成り立つ。

$$-S = \sum_i p_i \log p_i$$

$$-S = \sum_i p_i \log p_i + \alpha \left[\sum_i p_i - 1 \right] + \beta \left[\sum_i p_i E_i - E \right]$$

ラグランジェの未定乗数法

ラグランジェの未定乗数法

$-S$ は、 p_i, α, β についての関数である。 $-S$ が p_i について最小値をとるのは、次の条件を満たす時である。

$$\frac{\partial(-S)}{\partial p_i} = 0$$

ラグランジェの未定乗数法

$-S$ は、 p_i, α, β についての関数である。 $-S$ が p_i について最小値をとるのは、次の条件を満たす時である。

$$\frac{\partial(-S)}{\partial p_i} = 0$$

$$\frac{\partial(-S)}{\partial p_i} = \frac{\partial}{\partial p_i} \left\langle \sum_i p_i \log p_i + \alpha \left[\sum_i p_i - 1 \right] + \beta \left[\sum_i p_i E_i - E \right] \right\rangle$$

ラグランジェの未定乗数法

$-S$ は、 p_i, α, β についての関数である。 $-S$ が p_i について最小値をとるのは、次の条件を満たす時である。

$$\frac{\partial(-S)}{\partial p_i} = 0$$

$$\frac{\partial(-S)}{\partial p_i} = \frac{\partial}{\partial p_i} \left\langle \sum_i p_i \log p_i + \alpha \left[\sum_i p_i - 1 \right] + \beta \left[\sum_i p_i E_i - E \right] \right\rangle$$

最初の項は、 $\frac{\partial x \log x}{\partial x} = \log x + 1$ を使うと $\log p_i + 1$ に、

二番目の項は α に、三番目の項は βE_i になる。

$$\frac{\partial(-S)}{\partial p_i} = \log p_i + 1 + \alpha + \beta E_i = 0$$

これから、 $\log p_i + 1 + \alpha + \beta E_i = 0$ がわかる。

分配関数 Z

分配関数 Z

$$\log p_i + 1 + \alpha + \beta E_i = 0$$

$$\log p_i = -(1 + \alpha + \beta E_i)$$

だから、

$$p_i = e^{-(1+\alpha)} e^{-\beta E_i}$$

$Z = e^{1+\alpha}$ とすると、

$$p_i = \frac{1}{Z} e^{-\beta E_i}$$

分配関数 Z

これを

$$\sum_i p_i = 1$$

に代入すると、

$$\sum_i \frac{1}{Z} e^{-\beta E_i} = 1$$

よって、

$$\frac{1}{Z} \sum_i e^{-\beta E_i} = 1$$

これから、

$$Z = \sum_i e^{-\beta E_i}$$

分配関数 Z

$$Z(\beta) = \sum_i e^{-\beta E_i}$$

を、分配関数(partition function)と呼ぶ。

第二部

エネルギーとエントロピー 2

分配関数 Z の利用



ふりかえり：分配関数Zの導出

ラグランジェの未定乗数法を使って

$$\frac{\partial(-S)}{\partial p_i} = \frac{\partial}{\partial p_i} \left\langle \sum_i p_i \log p_i + \alpha \left[\sum_i p_i - 1 \right] + \beta \left[\sum_i p_i E_i - E \right] \right\rangle = 0$$

これから

$$\begin{aligned} \log p_i + 1 + \alpha + \beta E_i &= 0 \\ p_i &= e^{-(1+\alpha)} e^{-\beta E_i} \end{aligned}$$

$Z = e^{1+\alpha}$ と置いて

$$p_i = \frac{1}{Z} e^{-\beta E_i}$$

$\sum_i p_i = 1$ を使うと

$$Z = \sum_i e^{-\beta E_i}$$

エネルギー E を分配関数 Z で表す

エネルギーEを分配関数Zで表す

出発点

$$\sum_i p_i E_i = E$$

$$p_i = \frac{1}{Z} e^{-\beta E_i}$$

$$\sum_i \frac{1}{Z} e^{-\beta E_i} E_i = E$$

エネルギーEを分配関数Zで表す

$$Z = \sum_i e^{-\beta E_i}$$

$$\frac{\partial Z}{\partial \beta} = \sum_i -E_i e^{-\beta E_i}$$

$$\sum_i \frac{1}{Z} e^{-\beta E_i} E_i = E$$

$$\frac{1}{Z} \frac{\partial Z}{\partial \beta} = - \sum_i \frac{1}{Z} E_i e^{-\beta E_i} = -E$$

$$E = - \frac{1}{Z} \frac{\partial Z}{\partial \beta}$$

エネルギーEを分配関数Zで表す

$$E = -\frac{1}{Z} \frac{\partial Z}{\partial \beta}$$

$$E = -\frac{\partial \log Z}{\partial \beta}$$

エントロピー S を分配関数 Z で表す

エントロピーSを分配関数Zで表す

出発点

$$S = - \sum_i p_i \log p_i$$

$$p_i = \frac{1}{Z} e^{-\beta E_i}$$

$$\log p_i = -(1 + \alpha + \beta E_i)$$

$$\begin{aligned} &= - \sum_i \frac{1}{Z} e^{-\beta E_i} [-(1 + \alpha + \beta E_i)] \\ &= \sum_i \frac{1}{Z} e^{-\beta E_i} [1 + \alpha + \beta E_i] \end{aligned}$$

エントロピー S を分配関数 Z で表す

$$S = \sum_i \frac{1}{Z} e^{-\beta E_i} [1 + \alpha + \beta E_i]$$

$Z = e^{1+\alpha}$ と置いたから、 $\log Z = 1 + \alpha$

$$S = \sum_i \frac{1}{Z} e^{-\beta E_i} [\log Z + \beta E_i]$$

$$\sum_i \frac{1}{Z} e^{-\beta E_i} E_i = E$$

$$= \sum_i \frac{1}{Z} e^{-\beta E_i} \log Z + \beta \sum_i \frac{1}{Z} e^{-\beta E_i} E_i$$

$$\frac{1}{Z} \sum_i e^{-\beta E_i} = 1$$

$$= \frac{1}{Z} \log Z \sum_i e^{-\beta E_i} + \beta E$$

$$= \log Z + \beta E$$

$$S = \beta E + \log Z$$

温度 T と β の関係

古典的な熱力学での エントロピー S の表現

古典的な熱力学では、エントロピーは、次のように表現されていた。

$$dE = TdS$$

$$\frac{dE}{dS} = T$$

$$dS = \frac{dE}{T}$$

エントロピー S の分配関数 Z を使った表現

出発点

$$S = \beta E + \log Z$$

$$\frac{dS}{dE} = \beta + E \frac{d\beta}{dE} + \frac{d \log Z}{dE}$$

$$\frac{d \log Z}{dE} = \frac{\partial \log Z}{\partial \beta} \frac{d\beta}{dE}$$

$$= \beta + \frac{d\beta}{dE} \left[E + \frac{\partial \log Z}{\partial \beta} \right]$$

$$E = - \frac{\partial \log Z}{\partial \beta}$$

$$\frac{dS}{dE} = \beta$$

温度Tと β の関係

$$\frac{dE}{dS} = T$$

$$\frac{dS}{dE} = \beta$$

$$T = \frac{1}{\beta}$$

β は、温度Tの逆数である

分配関数 Z から導かれる式たち

$$Z = \sum_i e^{-\beta E_i}$$

$$p_i = \frac{1}{Z} e^{-\beta E_i}$$

$$E = - \frac{\partial \log Z}{\partial \beta}$$

$$S = \beta E + \log Z$$

$$T = \frac{1}{\beta}$$





第三部

ディープ・ラーニングとエントロピー

1. 相対エントロピーとその解釈

2. ディープ・ラーニングと分配関数 Z の利用

Agenda

情報とエントロピー

第一部 確率分布とエントロピー

- シャノンのエントロピー
- ボルツマンのエントロピー

第二部 エネルギーとエントロピー

- 分配関数 Z
- 分配関数の利用

第三部 ディープ・ラーニングとエントロピー

- 相対エントロピーとその解釈
- ディープ・ラーニングと分配関数 Z の利用

第四部 量子情報とエントロピー

- ノイマン・エントロピー
- エネルギーと量子エントロピー

第三部

ディープ・ラーニングとエントロピー 1

相対エントロピーとその解釈



確率分布とシャノン・エントロピー

確率分布とエントロピー

ある確率分布 p_i が与えられた時、そのエントロピー S は次の式で与えられる。

$$S = - \sum_i p_i \log p_i$$

確率分布とエントロピー

ただ、どんな確率分布についても、アприオリに一つのエントロピーが先の公式で天下りの的に定まるということに、すこし違和感を持つ人がいるかもしれない。そもそも、確率分布がアприオリに与えられるものは、自明ではない。

確率分布とエントロピー

ただ、どんな確率分布についても、アприオリに一つのエントロピーが先の公式で天下りの的に定まるということに、すこし違和感を持つ人がいるかもしれない。そもそも、確率分布がアприオリに与えられるものは、自明ではない。

そういう人には、次の「相対的なエントロピー」という考え方の方が、納得が行きやすいと思う。得られるエントロピーは、絶対的な確定したものではなく、事前に知っていたこととの関係で決まる、相対的なものだと考えるのだ。

相対エントロピー そのBayesian的解釈

相対エントロピー

pに対するqの「相対エントロピー」 $H_{rel}(q, p)$ を、次の式で定義する。

$$H_{rel}(q, p) = \sum_x q(x) \log \frac{q(x)}{p(x)}$$

$H_{rel}(q, p) = 0$ となるのは、 $q = p$ の場合だけであることはすぐわかる。

$H_{rel}(q, p)$ を $H(q(x) || p(x))$ と表すこともある。

相対エントロピーの Bayesian的解釈

「相対エントロピー」 $H_{\text{rel}}(q, p)$ に次のような解釈を与えよう。

事前に知っていた(多分、それは正確な知識ではないかもしれないので「仮説」といってもいい)確率分布を p_i としよう。実際に、観測して得られた新しい確率分布 q_i とする。

こうしたアプローチは、Bayesianのものである。「相対エントロピー」というのは、アприオリな「シャノンのエントロピー」を、Bayesianの考え方で、相対化したエントロピーと考えることができる。

認識と学習のBayesian的解釈

エントロピー＝情報量のこの相対的な解釈は、人間の認識で得られる情報量の解釈には、とても向いている。認識や学習のモデルを、この情報量を使って解釈できる。

例えば、先の $H_{rel}(q, p) = 0$ の場合の解釈では、仮説 p と実験結果 q が一致した場合には、実験で得られた情報量は0 だと考えればいい。

認識と学習のBayesian的解釈

認識の順番を t で表してみよう。

先の例では、事前の仮説 $p(t - 1)$ が、実験によって、事後に $q(t)$ に置き換わるのだが、この実験で得られた情報は、 $H_{rel}(q(t), p(t - 1))$ で表される。

ここで、 $q(t)$ を新たな $p(t)$ として、 $H_{rel}(q(t + 1), p(t))$ を考える操作を繰り返すことができる。

これは、「認識の発展」のモデルと考えることができる。この「認識の発展」は、 $H_{rel}(q, p) = 0$ になるときに終わる。

認識と学習のBayesian的解釈

逆に、もしも、最初から正しい分布 q を、何らかの方法で我々が知っていて、実験 $p(t)$ を繰り返すのなら、 $H_{rel}(q, p(t))$ は、実測値 $p(t)$ から、正しい答え q に至るために「学習しなければいけない情報量」を表すことになる。 $(q$ は、「常に正しい」と仮定しているので、それは時間には依存しない。 q は、 t を含まないことに注意。)

ここでは、 $H_{rel}(q, p(t)) = 0$ は、「もはや、学習すべき情報が残されていない」ことを意味して、その状態で、学習は終わる。

クロス・エントロピー

クロス・エントロピー

ディープラーニングでコスト関数として利用される「クロス・エントロピー」は、こうした「相対エントロピー」の一種である。

「正しい」分布を q 、実測値を p としたとき、クロス・エントロピー $H_{cross}(q, p)$ は、次の式で定義される。

$$H_{cross}(q, p) = \sum q_i \log p_i$$

クロス・エントロピーと相対エントロピー

先の $H_{rel}(q, p)$ の定義から、次のことがわかる。

$$\begin{aligned} H_{rel}(q, p) &= \sum q_i \log \frac{q_i}{p_i} = \sum q_i \log q_i - \sum q_i \log p_i \\ &= -H(q) - H_{cross}(q, p) \end{aligned}$$

クロス・エントロピー

個人的には、シャノンのエントロピーの定義も、ぶっくらぼうだと思うのだが、クロス・エントロピーの定義の意味は、式の形からはわかりにくいように感じている。

ディープラーニングでの「学習」の直接の目的は、クロス・エントロピーを最小にすることなのだが、相対エントロピーの言葉で言えば、それは、「正しい」認識に至るために、「残された学習すべき情報を最小のものにすること」となる。

こちらの方がわかりやすいように思う。

第三部

ディープ・ラーニングとエントロピー 2

ディープ・ラーニングと分配関数 Z の利用



ディープ・ラーニングの メタ・パラメーターとしての「温度」

ディープ・ラーニングの メタ・パラメーターとしての「温度」

ディープ・ラーニングの「学習」では、様々な「メタ・パラメーター」が登場する。このパラメーターの設定で、学習の精度やスピードが、大きく変わる。

こうしたメタ・パラメーターの一つに「温度」が登場することがある。大抵は、Softmaxに確率分布を放り込む時に、一律に一定の定数をかける処理をすることがあるのだが、その定数を「温度」ということがあるのだ。

最初は、この言葉づかいが、奇妙に思えたのだが、Softmax自体が、エントロピーの理論を借用しているのだと思うと、こうした言葉遣いには、実は、何の不思議もない。

ディープ・ラーニングの メタ・パラメーターとしての「温度」の例

だいぶ前に見かけた例では、GoogleのDeep MindチームのDNC (Differentiable Neural Computer)論文に、次のような関数 $C(M, k, \beta)[i]$ が定義されていた。

$$C(M, k, \beta)[i] = \frac{\exp\{\mathcal{D}(k, M[i, \cdot])\beta\}}{\sum_j \exp\{\mathcal{D}(k, M[i, \cdot])\beta\}}$$

この関数 $C(M, k, \beta)[i]$ は、DNCで、もっとも基本的な関数の一つである。ご覧のように、分子でも分母でも、冪乗部分に一律に定数 β がかかっている。Nature誌での解説では、この β は、「キーの強さ」とされている。

where $\underline{k} \in \mathbb{R}^W$ is a lookup key, $\beta \in [1, \infty)$ is a scalar representing **key strength** and D is the cosine similarity:

ディープ・ラーニングの メタ・パラメーターとしての「温度」の例

でも、この式は、Softmaxの原型である、次の式、すなわち、ある分配関数 $Z = \sum_i e^{-\beta E_i}$ が与えられた時、系がある状態*i* を取る確率 $p_i = \frac{1}{Z} e^{-\beta E_i}$ を表す式の変化系である。

$$p_i = \frac{1}{Z} e^{-\beta E_i} = \frac{e^{-\beta E_i}}{\sum_i e^{-\beta E_i}}$$

ディープ・ラーニングと確率分布

ディープ・ラーニングのBayesian的解釈

もしも、我々が、最初から正しい確率分布 q を、何らかの方法で知っていて、確率分布 $p(t)$ の観測を繰り返すのなら、相対エントロピー $H_{rel}(q, p(t))$ は、観測された確率分布 $p(t)$ から、正しい確率分布 q に至るために「学習しなければいけない情報量」を表すことになる。

ディープ・ラーニングが繰り返し行っている計算は、バック・プロパゲーションで確率分布 $p(t)$ を少しずつ変化させて、相対エントロピー $H_{rel}(q, p(t))$ を、ゼロに近づけようとしていることに他ならない。

確率分布は、分配関数で定義される

これまで、エネルギーとエントロピーの関係から、分配関数を導出したのだが、重要なことは、どんな確率分布も、分配関数で定義できるということだ。

確率分布が与えられれば、それが何の確率分布だろうと、そこにエントロピーを見出すことができるように、我々は、与えられた確率分布を、それがどんなものであろうと、分配関数で再定義できる。

ただし、この再定義の際、元々の分配関数の定義で用いられていた「エネルギー」や「逆温度 β 」には、別の解釈が与えられることになる。

確率分布を、「温度」で表す

確率分布を、分配関数を使って「温度」で表す

$\beta = 1/k_B T$ と、明示的に温度 T が見える形にした時、 T の値をあげたり下げたりした時、確率分布 p_i がどう変化するかを見てみよう。

$\beta = 1/k_B T$ とおくと、分配関数 $Z = \sum_i e^{-\beta E_i} = \sum_i \exp\left(\frac{E_i}{k_B T}\right)$

E_i の分布は変わらないとして

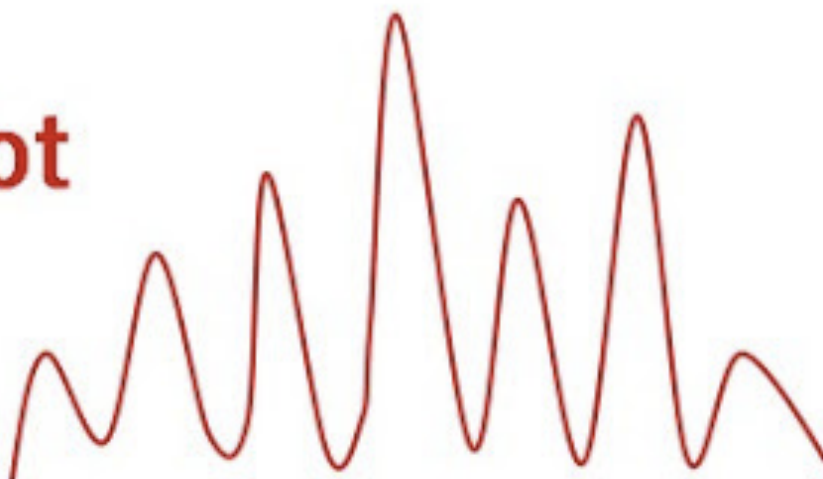
$$p_i = \frac{\exp\left(\frac{E_i}{k_B T}\right)}{\sum_i \exp\left(\frac{E_i}{k_B T}\right)}$$

である。

$$p_i = \frac{\exp\left(\frac{E_i}{k_B T}\right)}{\sum_i \exp\left(\frac{E_i}{k_B T}\right)}$$

の変化

Hot

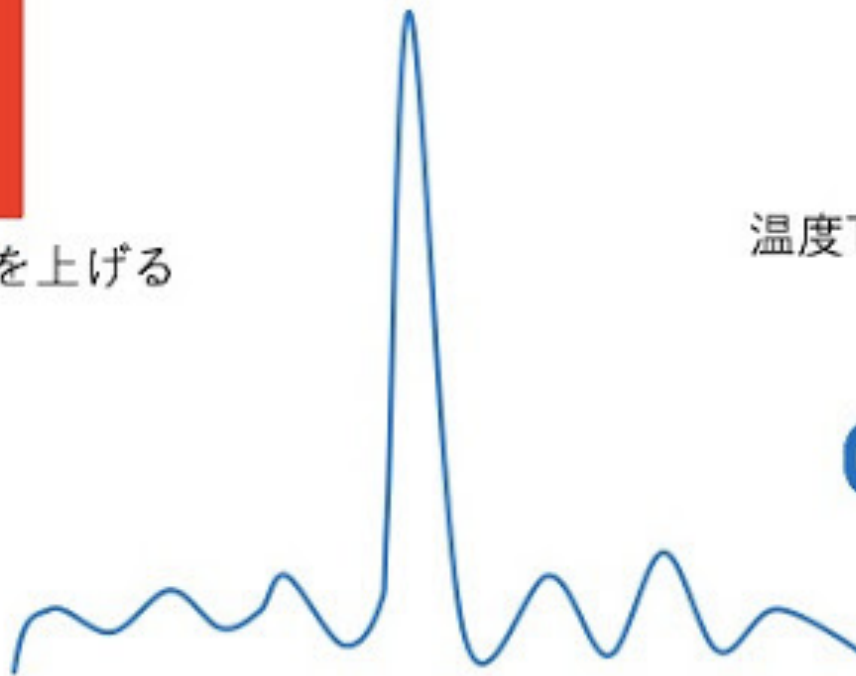


温度Tを上げる



温度Tを下げる

Cool



「温度」の解釈を考える

ディープ・ラーニングでの メタ・パラメーター「温度」の役割

ディープ・ラーニングの学習でメタ・パラメーター「温度」の役割を考えてみる。先の図で見たように、温度が変わると、確率分布が変わる。「温度」は、先の β の逆数だ。

要するに、「温度」が高いと、いろんなところに複数の比較的高い確率が見られるようになり、「温度」が低いと、一つのピークがはっきりと現れる。(山の凸凹のパターンは、基本的に同じなのだが、コントラストが強調される)

特徴抽出には、後者の方が良さそうに見えるかもしれないが、意外と単純ではない。前者のほうが、収束には時間がかかりそうだが、思いもかけない特徴を発見する可能性があるかもしれないし、後者の方は、実は、つまらない特徴しか見つけられないということになるのかもしれない。

「ホットな議論」と「クールな議論」

面白いのは、我々の日常の議論にも、こうした類型的には二つのパターンがあるのではないかということである。

「ホットな議論」は、いろんなトピックに話が跳んで、なかなか議論がまとまらない。ただ、議論百出は、議論に参加する人の視点を拡大してくれるかもしれない。

「クールな議論」は、小学生の学級委員長の議論のまとめみたいなもので、間違っていないかもしれないが、面白みに欠ける。飛躍がなく、新しい発見に至らない。

おそらく、人間の思考にも、「温度」に相当するメタ・パラメーターがあって、我々は、それをうまくコントロールしているのに違いない。
(と、テキトーなことを言ってみる。)





第四部

量子情報とエントロピー

1. ノイマン・エントロピー

2. エネルギーと量子エントロピー

Agenda

情報とエントロピー

第一部 確率分布とエントロピー

- シャノンのエントロピー
- ボルツマンのエントロピー

第二部 エネルギーとエントロピー

- 分配関数 Z
- 分配関数の利用

第三部 ディープ・ラーニングとエントロピー

- 相対エントロピーとその解釈
- ディープ・ラーニングと分配関数 Z の利用

第四部 量子情報とエントロピー

- ノイマン・エントロピー
- エネルギーと量子エントロピー

第四部

量子情報とエントロピー 1

ノイマン・エントロピー



密度行列

密度行列(密度演算子)

状態 $|\varphi_i\rangle$ が確率 p_i で起きる時、次の式で定義される行列を、密度行列という。

$$\rho = \sum_i p_i |\varphi_i\rangle \langle \varphi_i|$$

これは、状態 $|\varphi_i\rangle$ の「確率分布」を与えるものと考えてよい。

Trace

- traceは、行列の対角要素の和である。
行列Aのtraceを $\text{tr}(A)$ と表す。
- 密度行列 ρ のtraceは、次のように定義される。

$$\text{tr}(\rho) = \sum_k \langle k | \rho | k \rangle$$

密度行列(密度演算子)の例 1

$|\varphi\rangle = \alpha|0\rangle + \beta|1\rangle$ としよう。

$$\begin{aligned} |\varphi\rangle\langle\varphi| &= (\alpha|0\rangle + \beta|1\rangle)(\alpha^*\langle 0| + \beta^*\langle 1|) \\ &= \alpha\alpha^*|0\rangle\langle 0| + \alpha\beta^*|0\rangle\langle 1| + \beta\alpha^*|1\rangle\langle 0| + \beta\beta^*|1\rangle\langle 1| \\ &= |\alpha|^2|0\rangle\langle 0| + \alpha\beta^*|0\rangle\langle 1| + \beta\alpha^*|1\rangle\langle 0| + |\beta|^2|1\rangle\langle 1| \\ &= \begin{bmatrix} |\alpha|^2 & \alpha\beta^* \\ \beta\alpha^* & |\beta|^2 \end{bmatrix} \end{aligned}$$

$$\text{tr}(|\varphi\rangle\langle\varphi|) = |\alpha|^2 + |\beta|^2 = 1$$

外積 $|i\rangle\langle j|$ と行列の i 行・ j 列の要素の対応

$$\begin{pmatrix} |0\rangle\langle 0| \text{の係数} & |0\rangle\langle 1| \text{の係数} \\ |1\rangle\langle 0| \text{の係数} & |1\rangle\langle 1| \text{の係数} \end{pmatrix}$$

pureな密度行列のTrace

pureな密度行列は、 $\rho = |\psi\rangle\langle\psi|$ である。

$|\psi\rangle = \sum_i \alpha_i |i\rangle$ とすると、pureな密度行列のTrace は、

$$\begin{aligned}\text{tr}(|\psi\rangle\langle\psi|) &= \sum_k \langle k| \sum_{i,j} \alpha_i |i\rangle \alpha_j^* \langle j| |k\rangle \\ &= \sum_{i,j,k} \alpha_i \alpha_j^* \boxed{\langle k|i\rangle \langle j|k\rangle} \\ &= \sum_k |\alpha_k|^2 \\ &= 1.\end{aligned}$$

pureな密度行列のTrace は、常に 1 である。

mixed な密度行列のTrace

mixedな密度行列 ρ は、次のように表される。

$$\rho = \sum_i p_i |\psi_i\rangle \langle \psi_i|$$

$$\text{tr}(\rho) = \sum_k \langle k | \sum_i p_i |\psi_i\rangle \langle \psi_i | | k \rangle$$

$$= \sum_i p_i \text{tr}(|\psi_i\rangle \langle \psi_i|)$$

$$= \sum_i p_i$$

$$= 1.$$

mixedな密度行列のTrace も 1である。

traceと内積

pureな密度行列、 $\rho = |\psi\rangle\langle\varphi|$ を考える

$$|\psi\rangle = \sum_i \alpha_i |i\rangle, |\varphi\rangle = \sum_j \beta_j |j\rangle \text{ とすると、}$$

$$\text{tr}(|\psi\rangle\langle\varphi|) = \sum_k \langle k| \sum_{i,j} \alpha_i |i\rangle \beta_j^* \langle j| |k\rangle$$

$$= \sum_{i,j,k} \alpha_i \beta_j^* \boxed{\langle k|i\rangle \langle j|k\rangle}$$

$$= \sum_i \alpha_i \beta_i^* = \langle\varphi|\psi\rangle$$

$$\text{tr}(|\psi\rangle\langle\varphi|) = \langle\varphi|\psi\rangle$$

$$\text{tr}(A|\psi\rangle\langle\varphi|) = \langle\varphi|A|\psi\rangle$$

行列の対数

Aが行列の時、 e^A の定義

$$\exp(x) \equiv e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \dots$$

と同様に、Aが行列の時、 e^A を次のように定義する。

$$e^A \equiv \sum_{n=0}^{\infty} \frac{A^n}{n!}$$

Aが行列の時、 $\log A$ の定義

$$\exp(x) \equiv e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \dots$$

と同様に、 A が行列の時、 e^A を次のように定義する。

$$e^A \equiv \sum_{n=0}^{\infty} \frac{A^n}{n!}$$

$e^A = B$ の時、 $\log B = A$ とする。

$\log(A)$ の計算

$$\log(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \dots$$

を使う。

$\log(A) \equiv \log(\lambda(I+K)) = \log(\lambda I) + \log(I+K)$
とすると、

$$\log(A) = (\log \lambda)I + K - \frac{K^2}{2} + \frac{K^3}{3} - \frac{K^4}{4} + \dots$$

Kの次元をnとすると、 $K^n=0$ となるので、この級数は有限の項しか持たない。

log Aの基本的性質

- $AB=BA$ なら

$$\log(AB) = \log(A) + \log(B)$$

- $\log(A^{-1}) = -\log(A)$

- A が正定値なら

$$\text{tr} \log(AB) = \text{tr} \log(A) + \text{tr} \log(B)$$

ノイマン・エントロピー

ノイマン・エントロピー

密度行列で表された量子状態 ρ のノイマン・エントロピー $S(\rho)$ は、次の式で与えられる。

$$S(\rho) = -\text{tr}(\rho \log \rho)$$

ノイマン・エントロピー

密度行列 ρ の固有値を λ_i とする。この時、 ρ は、対角形式で次のように表せる。

$$\rho = \sum_i \lambda_i |i\rangle\langle i|$$

この時、ノイマン・エントロピー $S(\rho)$ は、次の式で与えられる。

$$S(\rho) = - \sum_i \lambda_i \log \lambda_i$$

ノイマン・エントロピー

密度行列 ρ の固有値を λ_i とする。この時、 ρ は、対角形式で次のように表せる。

$$\rho = \sum_i \lambda_i |i\rangle\langle i|$$

この時、ノイマン・エントロピー $S(\rho)$ は、次の式で与えられる。

$$S(\rho) = - \sum_i \lambda_i \log \lambda_i$$

この定義は、シャノン・エントロピーと一致する。

ノイマン・エントロピー

密度行列 ρ の固有値を λ_i とすれば、ノイマン・エントロピー $S(\rho)$ は、次の式で与えられる。

$$S(\rho) = - \sum_i \lambda_i \log \lambda_i$$

この定義は、シャノン・エントロピーと一致する。

ノイマン・エントロピーの例

d次元の単位行列をIとする時、 I/d のノイマン・エントロピーを計算してみよう。

行列 I/d の固有値は、 $1/d$ なので、

$$S(I/d) = - \sum_i \frac{1}{d} \log \frac{1}{d} = \sum_i \frac{1}{d} \log d = \log d$$

これは、次元dのノイマンエントロピーの最大値を与える。

第四部

量子情報とエントロピー 2

エネルギーと量子エントロピー

21/06/26 マルゼミ ショートムービー #08

ノイマン・エントロピーの導出

密度行列と分配関数

系のエネルギーがハミルトニアン H で与えられる系の密度行列 ρ は次の式で与えられる。

$$\rho = \frac{e^{-\beta H}}{Z}$$

分配関数 Z は、次の式で与えられる。

$$Z = \text{tr}(e^{-\beta H})$$



$$Z = \sum_i e^{-\beta E_i}$$

密度行列の対数をとる

密度行列 ρ は次の式で与えられる。

$$\rho = \frac{e^{-\beta H}}{Z}$$

$$-\log \rho = \beta H + \log Z$$

両辺に ρ をかける。

$$-\rho \log \rho = \left(\frac{e^{-\beta H}}{Z} \right) (\beta H + \log Z)$$

両辺のTraceをとる。

$$-\rho \log \rho = \left(\frac{e^{-\beta H}}{Z} \right) (\beta H + \log Z)$$

$$-tr(\rho \log \rho) = tr \left[\left(\frac{e^{-\beta H}}{Z} \right) (\beta H + \log Z) \right]$$

$$= tr \left[\left(\frac{e^{-\beta H}}{Z} \right) (\beta H + \log Z) \right]$$

$$= \beta tr \left[\left(\frac{e^{-\beta H}}{Z} \right) H \right] + \log Z tr \left[\frac{e^{-\beta H}}{Z} \right]$$



$$\sum_i \frac{1}{Z} e^{-\beta E_i} E_i = E$$



$$\frac{1}{Z} \sum_i e^{-\beta E_i} = 1$$

$$S = \beta E + \log Z$$

$$\begin{aligned} -\text{tr}(\rho \log \rho) &= \beta \text{tr} \left[\left(\frac{e^{-\beta H} H}{Z} \right) \right] + \log Z \text{tr} \left[\frac{e^{-\beta H}}{Z} \right] \\ &= \beta H + \log Z = S \end{aligned}$$

よって、 $-\text{tr}(\rho \log \rho)$ はエントロピー S を表す。

ハミルトニアン 時間発展の演算子

時間発展の演算子

時刻 t での量子の状態を、 $|\Psi(t)\rangle$ で表してみよう。

$t=0$ の時の状態を $|\Psi(0)\rangle$ とすると、ある時間発展を表現する演算子 $U(t)$ を使って、次の式が成り立つとする。

$$|\Psi(t)\rangle = U(t)|\Psi(0)\rangle$$

二つの状態が「区別」できることは、それが「直交」していることに等しい

$t=0$ の時に、 $|\Psi(0)\rangle$ と $|\Phi(0)\rangle$ は、直交していたとする。

すなわち、 $\langle\Psi(0)|\Phi(0)\rangle = 0$ かつ $\langle\Psi(t)|\Phi(t)\rangle = 0$

$t=0$ での直交関係は、時間がたっても、直交したままで、変わらない。

この時、 $\langle\Psi(t)| = \langle\Psi(0)|U^\dagger(t)$ だから、

$$\langle\Psi(t)|\Phi(t)\rangle = \langle\Psi(0)|U^\dagger(t)U(t)|\Phi(0)\rangle = 0.$$

時間発展の演算子

Ψ 、 Φ の異なる直交基底 i, j をとると、先の式から、

$$\langle i | \mathbf{U}^\dagger(t) \mathbf{U}(t) | j \rangle = 0 \quad (i \neq j)$$

もし、 $i=j$ なら、 $\mathbf{U}(t) | i \rangle$ と $\mathbf{U}(t) | j \rangle$ の内積は、1である。

$$\langle i | \mathbf{U}^\dagger(t) \mathbf{U}(t) | j \rangle = \delta_{ij}$$

$$\mathbf{U}^\dagger \mathbf{U} = I$$

時間発展の演算子は、ユニタリである。

ハミルトニアン

微小な時間変化 ϵ を考える。

時間発展のユニタリ性から、 $U^\dagger(\epsilon)U(\epsilon) = I$

ϵ が、極めて小さいとすれば、この演算子の作用は、恒等作用 I とごく近いと考えて、 $U(\epsilon) = I - i\epsilon H$ とおく。

この時、 $U^\dagger(\epsilon) = I + i\epsilon H^\dagger$ である。

これを、先のユニタリ性の式に代入すると、

$$(I + i\epsilon H^\dagger)(I - i\epsilon H) = I$$

ϵ の二次の項を無視すると、 $H^\dagger - H = 0$

よって、この H は、エルミートである。

この H をハミルトニアンと呼ぶ。

ここでの $-i\epsilon H$
という設定は、
「天下り」のもの
である。

シュレジンガー方程式

$$U(\epsilon) = I - i\epsilon H$$

$|\Psi(\epsilon)\rangle = |\Psi(0)\rangle - i\epsilon H|\Psi(0)\rangle$ が成り立つので、

$$\frac{|\Psi(\epsilon)\rangle - |\Psi(0)\rangle}{\epsilon} = -iH|\Psi(0)\rangle$$

これから、

$$\frac{\partial |\Psi\rangle}{\partial t} = -iH|\Psi\rangle$$

両辺が、エネルギー(例えば、
ジュール: $\text{kg}\cdot\text{m}^2/\text{s}^2$)を表す
ように、次元を揃える

これを、時間発展のシュレジンガー方程式と呼ぶ。

$$\hbar \frac{\partial |\Psi\rangle}{\partial t} = -iH|\Psi\rangle.$$

$$\hbar = \frac{h}{2\pi} = 1.054571726 \cdots \times 10^{-34} \text{ kg m}^2/\text{s}$$

ここでは、学ばないこと 量子力学は、ここから始まるのだが

- ハミルトニアン H は、古典論でのエネルギーに相当する Observable である。
- 観測されるエネルギーの値は、 H の固有値 E_j である。

$$H|E_j\rangle = E_j|E_j\rangle$$

この式を、**時間に依存しないシュレジンガー方程式**と呼ぶ。

- 先の $\frac{\partial|\Psi\rangle}{\partial t} = -iH|\Psi\rangle$ は、

次の、よく知られたシュレジンガー方程式と同値である。

$$i\hbar\frac{\partial\Psi(x)}{\partial t} = -\frac{\hbar^2}{2m}\frac{\partial^2\Psi(x)}{\partial x^2} + U(x)\Psi(x)$$

ここでは、学ばないこと 量子力学は、ここから始まるのだが

- ある系 $\Psi(t)$ のオブザーバブル L の観測値の期待値 $\langle L \rangle$ は、 $\langle L \rangle = \langle \Psi(t) | L | \Psi(t) \rangle$ で与えられる。 t で微分すると、次の式が得られる。ドットは、時間微分を表す。

$$\frac{d}{dt} \langle \Psi(t) | \mathbf{L} | \Psi(t) \rangle = \langle \dot{\Psi}(t) | \mathbf{L} | \Psi(t) \rangle + \langle \Psi(t) | \mathbf{L} | \dot{\Psi}(t) \rangle$$

- この時、

$$\frac{d}{dt} \langle \Psi(t) | \mathbf{L} | \Psi(t) \rangle = \frac{i}{\hbar} \langle \Psi(t) | \mathbf{H} \mathbf{L} | \Psi(t) \rangle - \frac{i}{\hbar} \langle \Psi(t) | \mathbf{L} \mathbf{H} | \Psi(t) \rangle$$

ここでは、 $\hbar \frac{\partial |\Psi\rangle}{\partial t} = -i \mathbf{H} |\Psi\rangle$ と、その複素共役を使った。

ここでは、学ばないこと 量子力学は、ここから始まるのだが

□ 先の式を整理すると、

$$\frac{d}{dt} \langle \Psi(t) | \mathbf{L} | \Psi(t) \rangle = \frac{i}{\hbar} \langle \Psi(t) | [\mathbf{H}\mathbf{L} - \mathbf{L}\mathbf{H}] | \Psi(t) \rangle$$

□ $\mathbf{H}\mathbf{L} - \mathbf{L}\mathbf{H} = [\mathbf{H}, \mathbf{L}]$ で表す (これを、 $\mathbf{H}$ と $\mathbf{L}$ の交換関係という) と、次の式が得られる。

$$\frac{d}{dt} \langle \mathbf{L} \rangle = \frac{i}{\hbar} \langle [\mathbf{H}, \mathbf{L}] \rangle$$

□ $\mathbf{L} = \mathbf{H}$ とすると、 $[\mathbf{H}, \mathbf{H}] = 0$ だから、 $\mathbf{H}$ の期待値の時間微分は0であることがわかる。**これは、エネルギーの保存則**を表している。