情報とエントロピー入門

21/05/29 マルレク基礎

はじめに

小論は、技術者に「エントロピー」とは何かについて、基本的な知識を提供することを目的にしています。

熱力学の中で登場した「エントロピー」という概念は、熱力学から統計力学、統計力学から量子力学という、19世紀から20世紀にかけての科学の枠組みの歴史的変化の中で中心的役割を果たします。

重要なことは、「エントロピー」概念は、もっぱら抽象的な「科学の枠組み」にかかわる概念ではなかったということです。それは、それぞれの時代の具体的で実践的な「技術」と密接なつながりを持っていました。

はじめに

蒸気機関が先端技術であった時代にも（熱効率の最大化）、電信・電話・ラジオ・テレビが先端技術であった時代にも（通信効率の最適化）、エントロピーを最小にしようという試みが、技術的チャレンジの中心課題でした。今日の最先端技術である、いわゆる「AI」技術の中核も、クロス・エントロピーの最小化と考えることができます。

科学者だけでなく、技術者も、実践的な概念としてエントロピーを理解することが重要だと、僕は考えています。

はじめに

小論では、ボルツマンとシャノンのエントロピーを紹介します。

ボルツマンが明らかにしたことは、可能なミクロな状態を全て数え上げた時、その「数」がエントロピーを決め、それが「温度」や「圧力」といったマクロな状態を規定するということです。ただ、ミクロの視点からマクロの視点に移る時、ミクロの状態が持っていた多くの情報は失われます。

シャノンが対象とした情報の世界には、もちろん、「温度」や「圧力」があるわけではありません。ただ、そこでもエントロピーは定義できます。確率分布が与えられるものには、その分布に対応したエントロピーが存在することをシャノンの仕事は示しています。

はじめに

エントロピーについては、興味深い話題がたくさんあります。
エントロピーの増大による宇宙の熱的死、エントロピーの増大に抗うものとしての生命、時間の流れとエントロピー ...

21世紀の科学は、エネルギーとエントロピー、すなわち、物質と情報の科学として発展すると僕は考えています。

残念ながら、今回は、そうした話題に触れることができませんでした。こうしたトピックについては、別の機会に紹介したいと思います。

Agenda
情報とエントロピー入門

Part I
エントロピーの理論を振り返る

Part II
統計力学とボルツマン・エントロピー

Part III
情報理論とシャノン・エントロピー

Part IV
シャノン・エントロピーの通信技術への応用

Part I
エントロピーの理論を振り返る

- □ 熱力学的エントロピー
- □ 統計力学的エントロピー
- □ 量子論的エントロピー
- □ 情報理論的エントロピー
- □ エントロピー概念の拡大

Part II
統計力学とボルツマン・エントロピー

- ボルツマンのエントロピーの定義
- ある系の持つ可能な状態の数を考える
- ボルツマンの S = k log W を計算する
- ボルツマンの悲劇

Part III
情報理論とシャノン・エントロピー

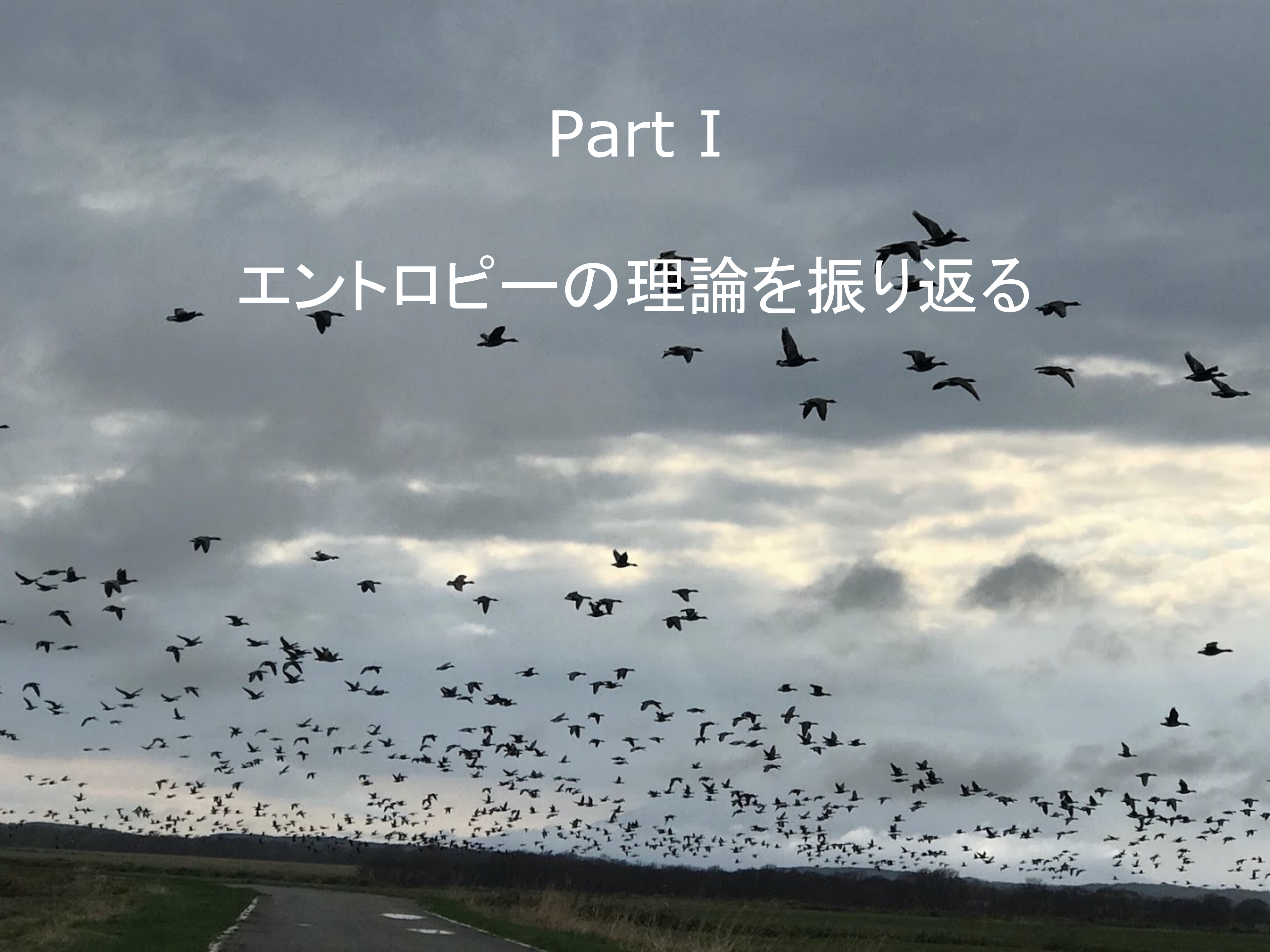
- □ シャノンのエントロピー
- □ シャノンが考えたこと
- □ シャノンのエントロピーを計算する
- □ フォン・ノイマンがシャノンに囁いたこと
- □ シャノン・エントロピーの性質について

Part IV
シャノン・エントロピーの通信技術への応用

- □ 通信の世界でメッセージのエントロピーを考える
- □ ボルツマンとシャノンのエントロピーの比較
- □ エントロピーをグラフィカルに表現する
- □ クロスエントロピーをグラフィカルに表現する
- □ 相対エントロピー

Part I

エントロピーの理論を振り返る

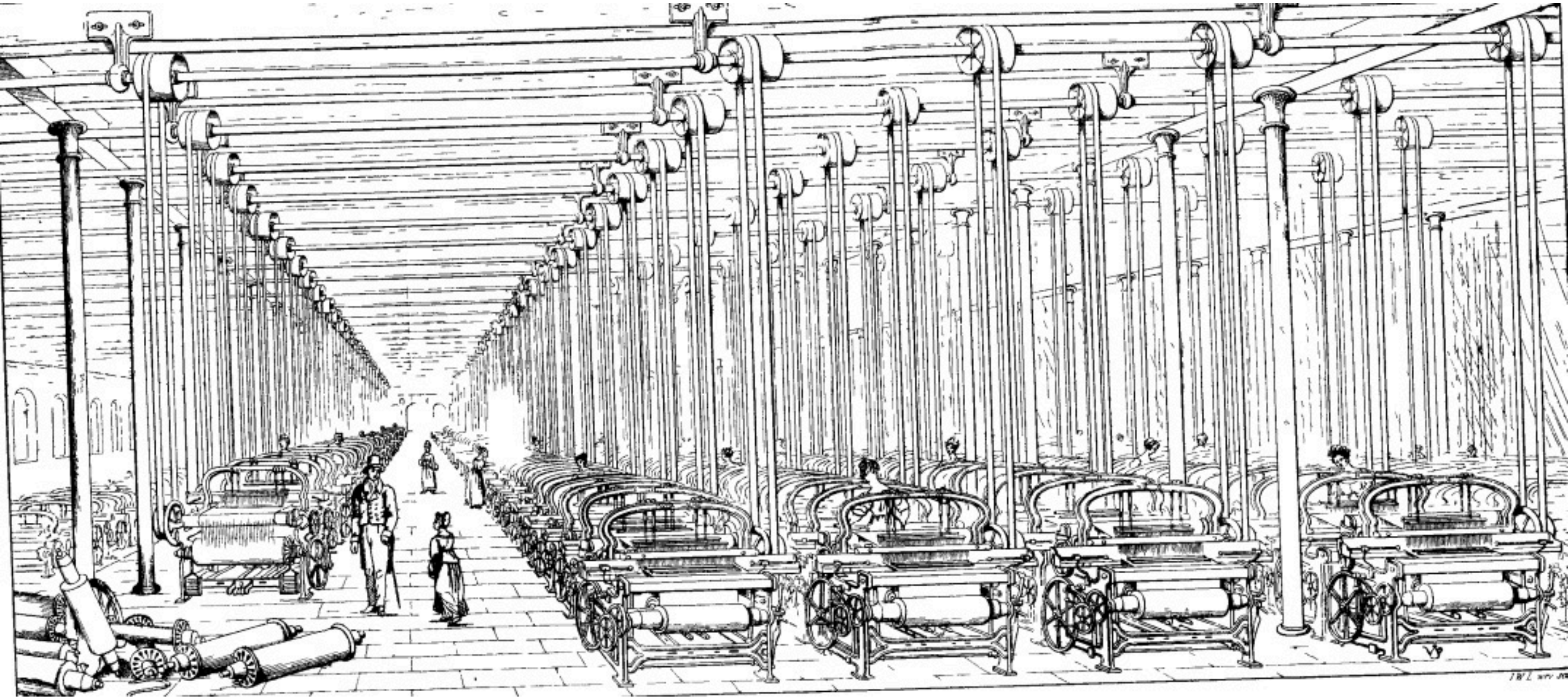


エントロピーの理論を振り返る

- 熱力学的エントロピー
  - カルノー
  - クラシウス
- 統計力学的エントロピー
  - ボルツマン
  - ギブス
- 量子論的エントロピー
  - フォン・ノイマン
- 情報理論的エントロピー
  - シャノン
- エントロピー概念の拡大

熱力学的エントロピー

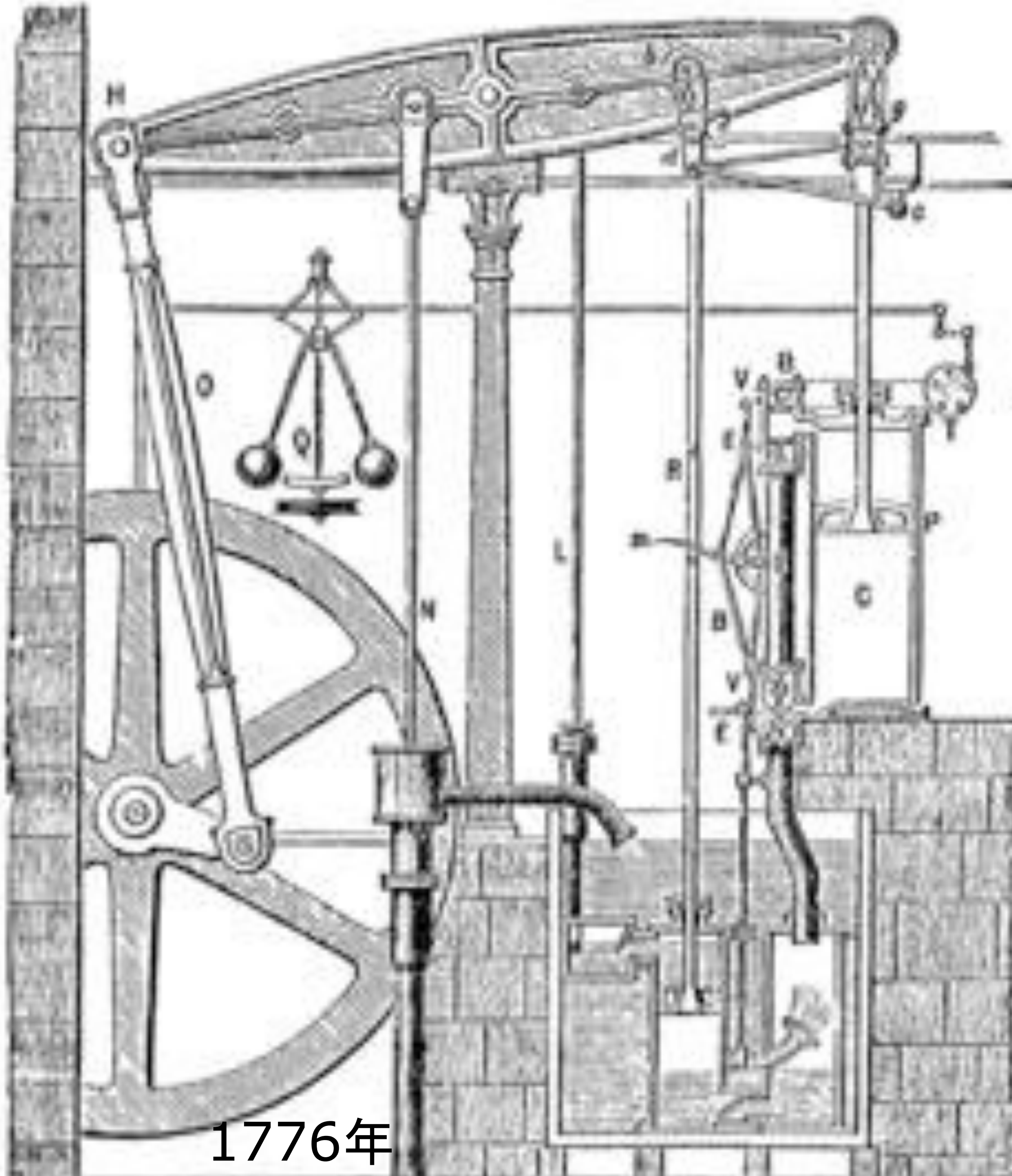
蒸気機関の登場



http://socserv.mcmaster.ca/econ/ugcm/3ll3/ure/PhilosophyManufactures.pdf

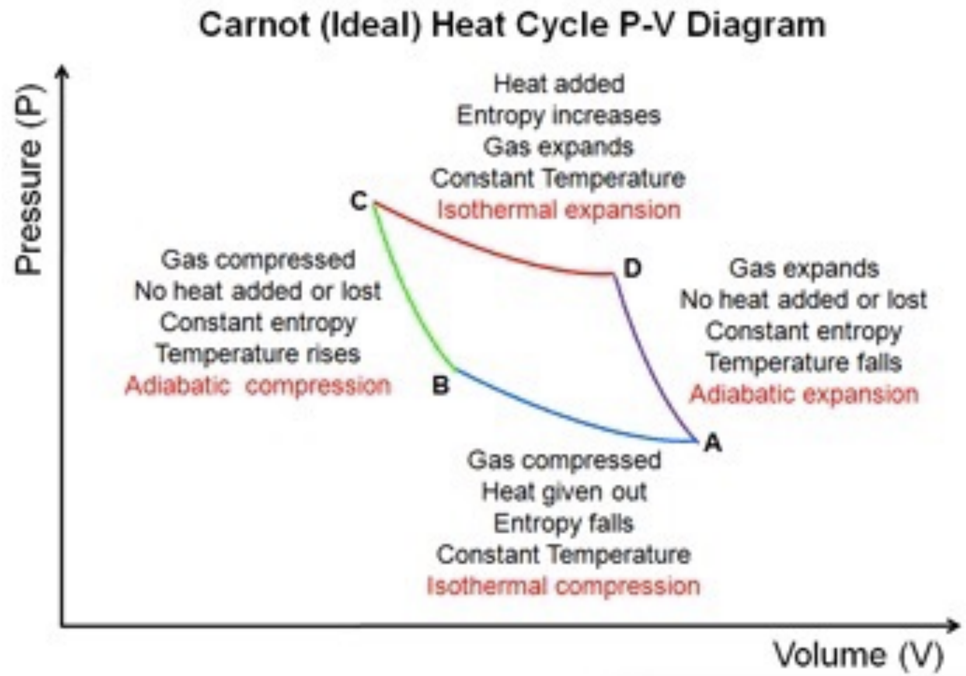
エントロピーとは、何か？

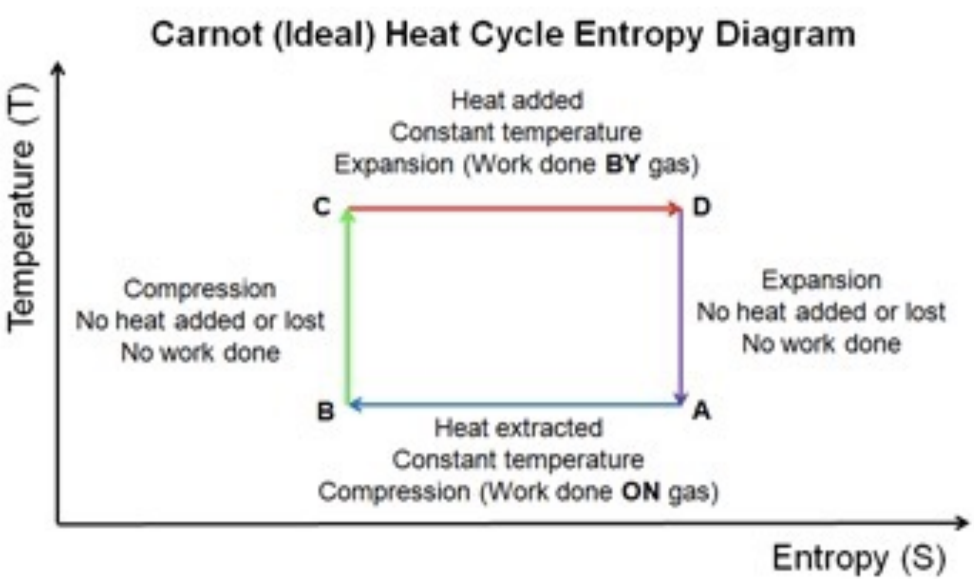
- Entropy という言葉は、19世紀半ば(1856年?)に、Clausiusによって名付けられた。それは、「熱」や「温度」「仕事」「エネルギー」といった量に関連づけられ、「熱力学」の中で生まれた概念であった。
- Wattの最初の蒸気機関が完成したのは、1776年。熱力学の基礎を作ったCarnotの仕事は、1828年である。彼らの実践的関心は、産業革命を可能にした当時の最先端技術である「蒸気機関」の効率化に向けられていた。
- 彼らは、蒸気機関に熱として与えられる蒸気機関内部のエネルギーの全てが、仕事として外部に取り出せるわけではないことを、経験的に発見する。



1736 - 1819

Carnot Cycle





1796-1832

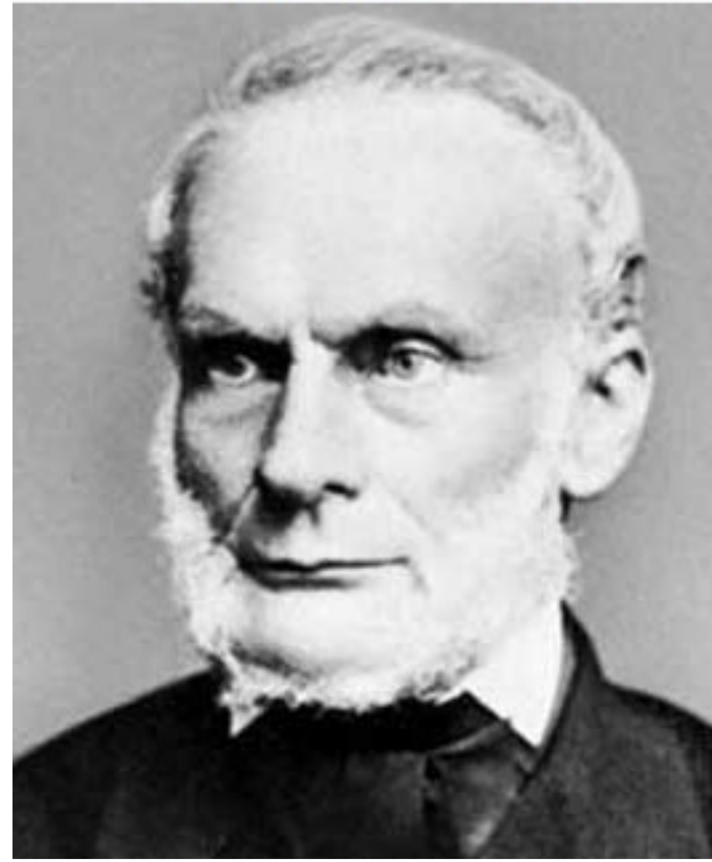
エントロピー $S = \dfrac{Q}{T}$ 1865年

I propose to name the quantity *S* the entropy of the system, after the Greek word [τροπη *trope*], the transformation.
I have deliberately chosen the word entropy to be as similar as possible to the word energy:

$$\sum_{i=1}^{n} \frac{Q_i}{T_i} \leq 0$$

$$TdS = dU + PdV$$

Rudolf Clausius



1822 - 1888

熱力学的エントロピーの単位

- 19世紀末に熱機関の熱力学を定式化した Helmholtzや Gibbsは、仕事として外部に取出せるエネルギーを「自由エネルギー」と呼ぶのだが、エントロピーは、仕事として取り出すことのできないエネルギーに関係付けられている。
- 内部エネルギーを U、温度を T、エントロピー S とすると、ヘルムホルツの自由エネルギーFは、F=U-TSで定義される。
- この式から分かるように、TS すなわち、温度とエントロピーを掛けたものは、エネルギーである。逆に言えば、エントロピーとは、単位的には、エネルギーを温度で割ったものである。

エネルギーを温度で割ったものが、なぜ、情報の理論と関係がつくのか？　こうした関係は、古典的な熱力学の議論からは、まったく出てこないのである。　ボルツマンが鍵を握っている。

統計力学的エントロピー


S = k. log W
LVDWIG
BOLTZMANN
1844 – 1906
DR PHIL. PAULA
BOLTZMANN

ボルツマンのエントロピー　S = k log W
1877年

このボルツマンのエントロピーの式

$$S = k \log W$$

は、エントロピーが、ミクロな状態の数の対数 log W に「比例」するという彼の偉大な発見を表しているのだが、情報の理論と結びつくのは、この log W の部分なのである。この部分は、単なる整数の対数なので、次元（単位のこと）は無い。

先の関係式で、「エネルギー 割る 温度」というエントロピーの単位（次元のこと）は、すべて、k というボルツマンの定数の中に閉じ込められている。

Boltzmann`s *H Theorem*
1872年

In classical statistical mechanics, the H-theorem, introduced by Ludwig Boltzmann in 1872, describes the tendency to decrease in the quantity H (defined below) in a nearly-ideal gas of molecules.

$$H(t) = \int_0^\infty f(E,t)\left(\ln\frac{f(E,t)}{\sqrt{E}} - 1\right) dE.$$

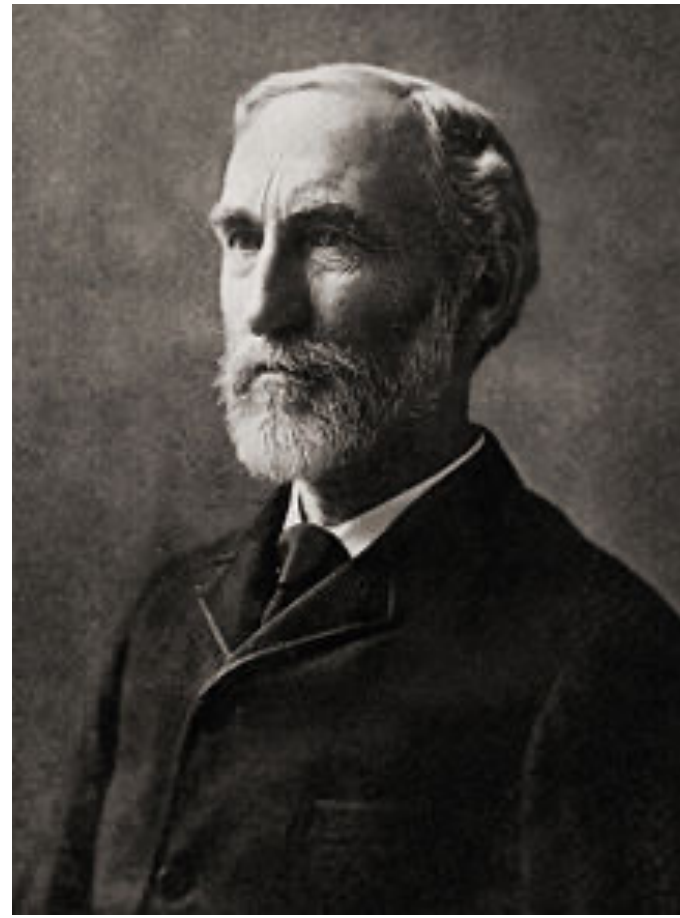
https://en.wikipedia.org/wiki/H-theorem

ギブスのエントロピー
1902年

Gibbs generalized Boltzmann's statistical interpretation of entropy S by defining the entropy of an arbitrary ensemble as

$$S = -k_{\mathrm{B}} \sum_i p_i \ln p_i$$

$$S_{\mathrm{G}}(\rho) = -k \int_{\mathscr{X}} dx\, \rho(x)\, \log \rho(x)\,,$$



量子論的エントロピー

Mathematical Foundations of Quantum Mechanics:
von Neumann
1932年

フォン･ノイマンによるエントロピーの定式化

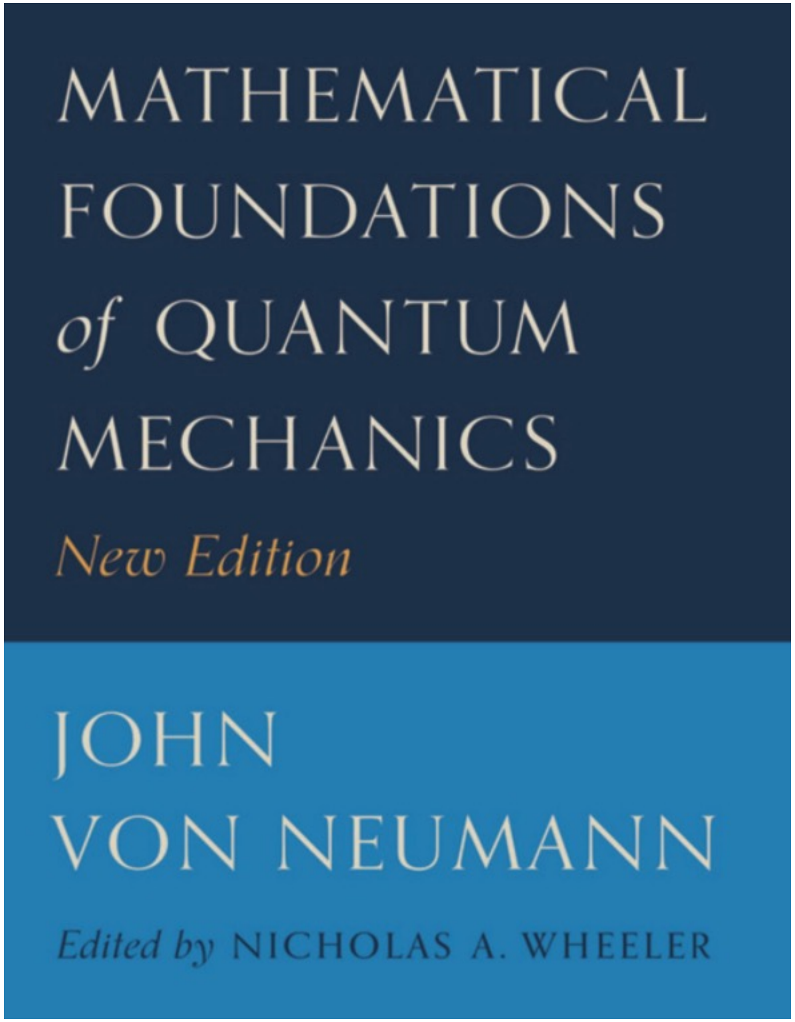
エントロピーの量子論的定式化を行ったのは、フォン･ノイマンである。もちろん、シャノンの登場のはるか以前である。驚くほどの慧眼である。

すでに、1932年の「量子力学の数学的基礎」の中で、次のようなエントロピーの定式化を与えている。

$$S = -\mathrm{Tr}(\rho \log \rho)$$

ここで、ρ は密度行列、Trは行列のTrace（対角要素の和をとったもの）である。この式の説明は、別の機会にゆずるが、ギブスの与えた密度関数 ρ を用いたエントロピーの式と、基本的には同じ形をしている。（ΣもTrも、和をとる演算である）

$$S = -\Sigma(\rho \log \rho)$$


MATHEMATICAL
FOUNDATIONS
of QUANTUM
MECHANICS
New Edition
JOHN
VON NEUMANN
Edited by NICHOLAS A. WHEELER

CONTENTS

CHAPTER IV

Deductive Development of the Theory

CHAPTER V

General Considerations

CHAPTER VI

The Measuring Process

CHAPTER IV

DEDUCTIVE DEVELOPMENT OF THE THEORY

CHAPTER V

GENERAL CONSIDERATIONS

We have at this point completed the desired reversible process. The entropy has increased by $N\kappa \sum_n w_n \ln w_n$ and since it must be zero in the final state it was

$$\boxed{-N\kappa \sum_{n=1}^{\infty} w_n \ln w_n}$$

in the initial state.

Since U has the eigenfunctions $\phi_1, \phi_2, \ldots$ with eigenvalues $w_1, w_2, \ldots$ the operator $\mathsf{U} \ln \mathsf{U}$ has the same eigenfunctions but eigenvalues $w_1 \ln w_1, w_2 \ln w_2, \ldots$ Consequently

$$\boxed{\mathrm{Tr}(\mathsf{U} \ln \mathsf{U}) = \sum_{n=1}^{\infty} w_n \ln w_n}$$

We observe that from $0 \leqslant w_n \leqslant 1$ it follows that $w_n \ln w_n \leqslant 0$, with equality only for $w_n = 0$ or 1; note also that for $w_n = 0$ we take $w_n \ln w_n$ to be zero, as follows from the fact that in the preceding discussion the vanishing w_n were not considered at all. One is led to those same conclusions by continuity considerations.

So we have determined that the entropy of a U-ensemble, consisting of N individual systems, is $-N\kappa \mathrm{Tr}(\mathsf{U} \ln \mathsf{U})$. The preceding remarks concerning $w_n \ln w_n$ establish that this is always $\geqslant 0$, and $= 0$ only if all w_n are either 0 or 1. From $\mathrm{Tr}\mathsf{U} = \sum_n w_n = 1$ it follows that in such cases only one $w_n = 1$ and all the others are 0, therefore $\mathsf{U} = \mathsf{P}_{[\phi]}$. That is, the states have entropy $= 0$, while mixtures have entropies > 0.

3. REVERSIBILITY AND EQUILIBRIUM PROBLEMS

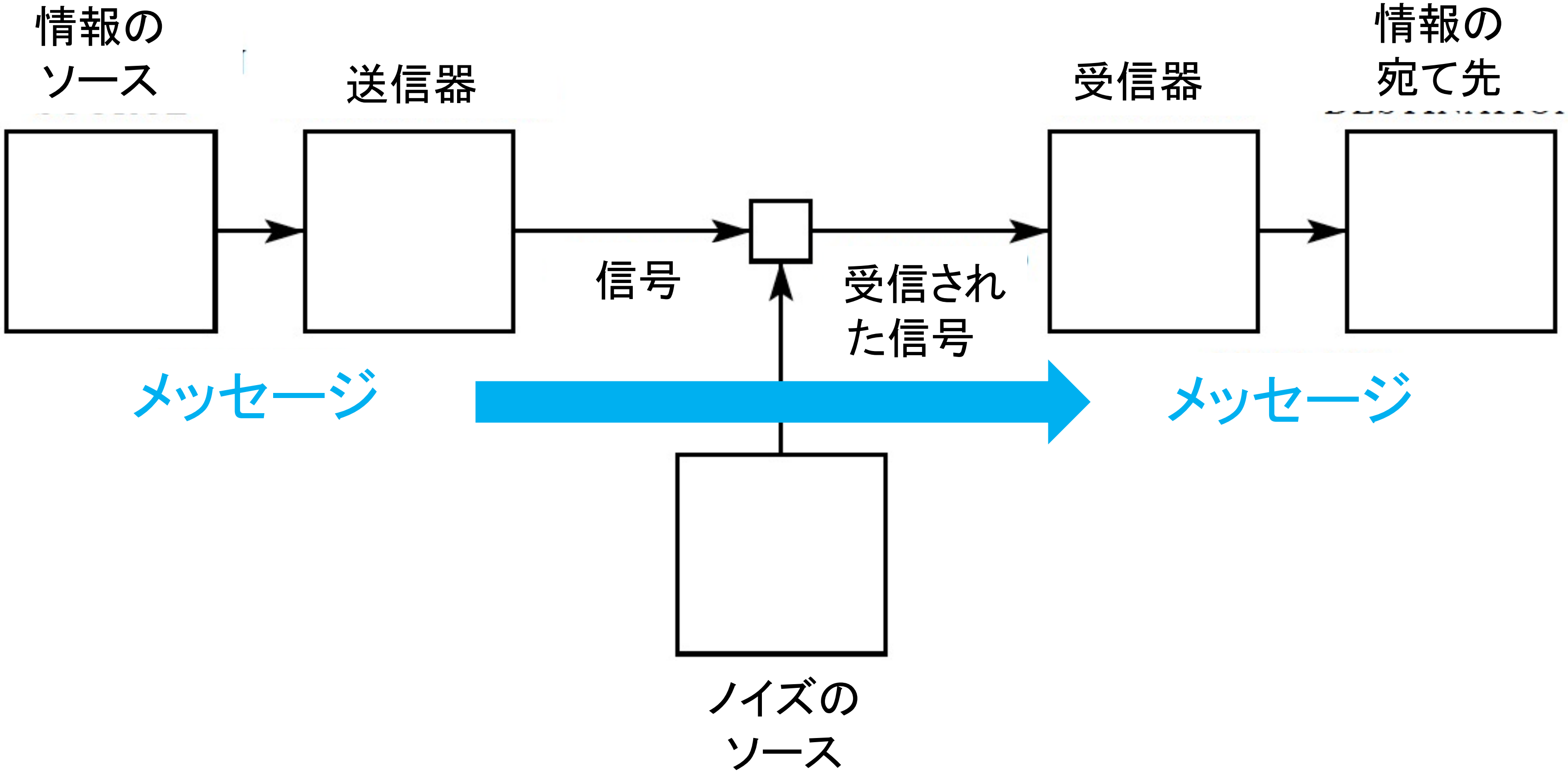
We can now prove the irreversibility of the measurement process, as asserted in **V.1**. For example, if U is a state then $\mathsf{U} = \mathsf{P}_{[\phi]}$ and in the measurement of a

Mathematical Foundations of Quantum Mechanics
By John von Neumann

情報理論的エントロピー

A Mathematical Theory of Communication
C. E. SHANNON
http://www.qsl.net/n9zia/pdf/shannon1948.pdf
1948年

一般的な通信システムの図式と通信の基本問題



通信の基本的な問題は、情報のソースで選択されたメッセージを、正確であれ近似的であれ、情報の宛て先で再生産することである。

3. If a choice be broken down into two successive choices, the original H should be the weighted sum of the individual values of H. The meaning of this is illustrated in Fig. 6. At the left we have three

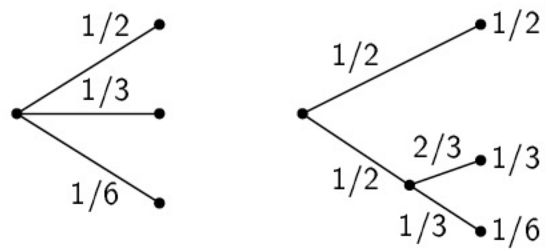


Fig. 6—Decomposition of a choice from three possibilities.

possibilities $p_1 = \frac{1}{2}, p_2 = \frac{1}{3}, p_3 = \frac{1}{6}$. On the right we first choose between two possibilities each with probability $\frac{1}{2}$, and if the second occurs make another choice with probabilities $\frac{2}{3}, \frac{1}{3}$. The final results have the same probabilities as before. We require, in this special case, that

$$H(\tfrac{1}{2}, \tfrac{1}{3}, \tfrac{1}{6}) = H(\tfrac{1}{2}, \tfrac{1}{2}) + \tfrac{1}{2}H(\tfrac{2}{3}, \tfrac{1}{3}).$$

The coefficient $\frac{1}{2}$ is the weighting factor introduced because this second choice only occurs half the time.

In Appendix 2, the following result is established:

Theorem 2: The only H satisfying the three above assumptions is of the form:

$$\boxed{H = -K \sum_{i=1}^{n} p_i \log p_i}$$

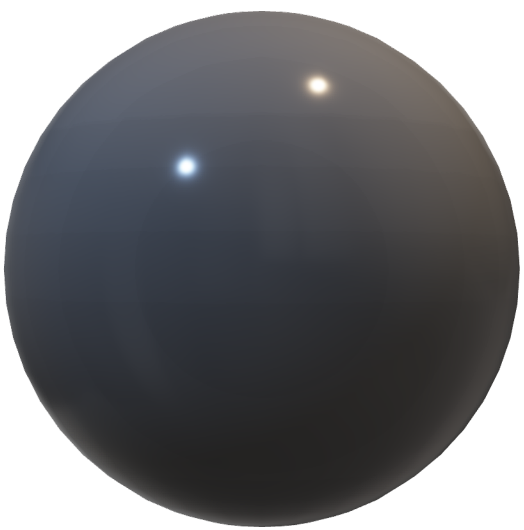
where K is a positive constant.

A Mathematical Theory of Communication
By C. E. SHANNON

エントロピー概念の拡大

1973年
ベッケンシュタイン　ブラックホールのエントロピー

- 観測可能な物理量としては、質量・電荷・角運動量の三つしか持たないと信じられていた。1973年、ベッケンシュタインは、これを覆す発見をする。彼は、ブラックホールが、先の三つの物理量の他に、エントロピー を持つこと、しかも、そのエントロピー S_{BH}が、ブラックホールの「地平」の表面積Aに比例することを見出す。この比例定数は、ホーキングが 1/4 であることを発見する。



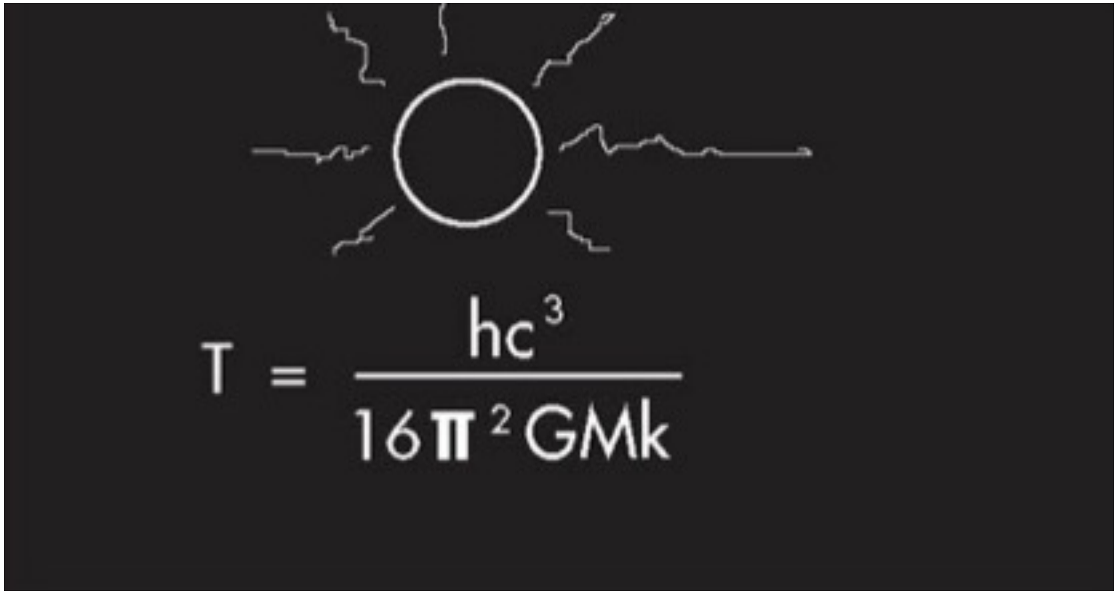
ブラックホールのエントロピー

$$S_{BH} = \frac{A}{4}$$

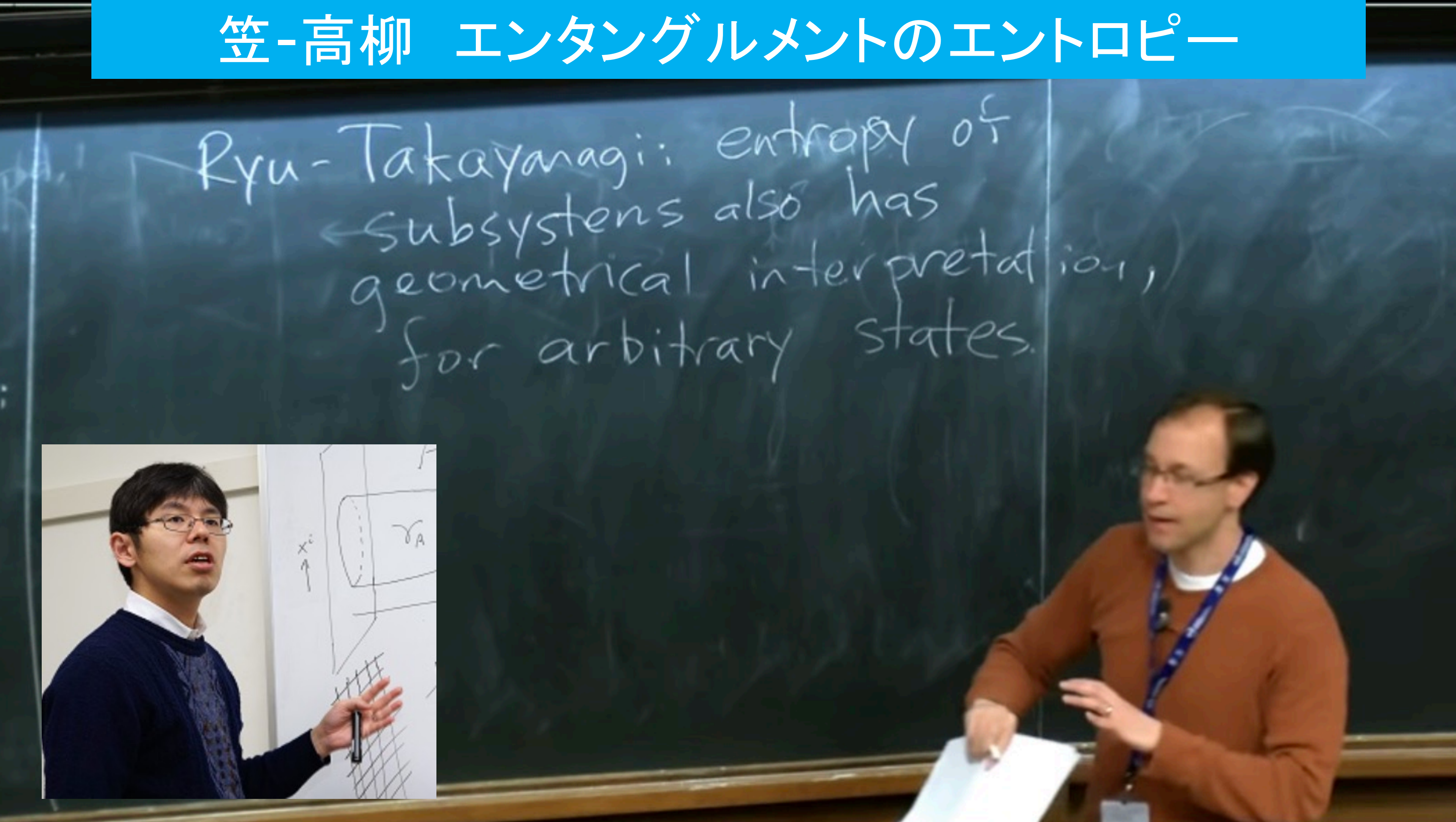
球の表面積 A=4πr^2

ブラックホールは、「温度」も持つ

- この輻射により、ブラックホールは、温度も持つことになる。ホーキングは、その温度も導出した。ブラックホールは質量 *M* が小さければ小さいほど高温であるといえる。とはいえその温度は、例えば太陽の数倍の質量を持つブラックホールの場合、100万分の1 K 程度となり、通常の恒星質量クラスのブラックホールでは宇宙背景放射の温度（3 K）よりもずっと低い。



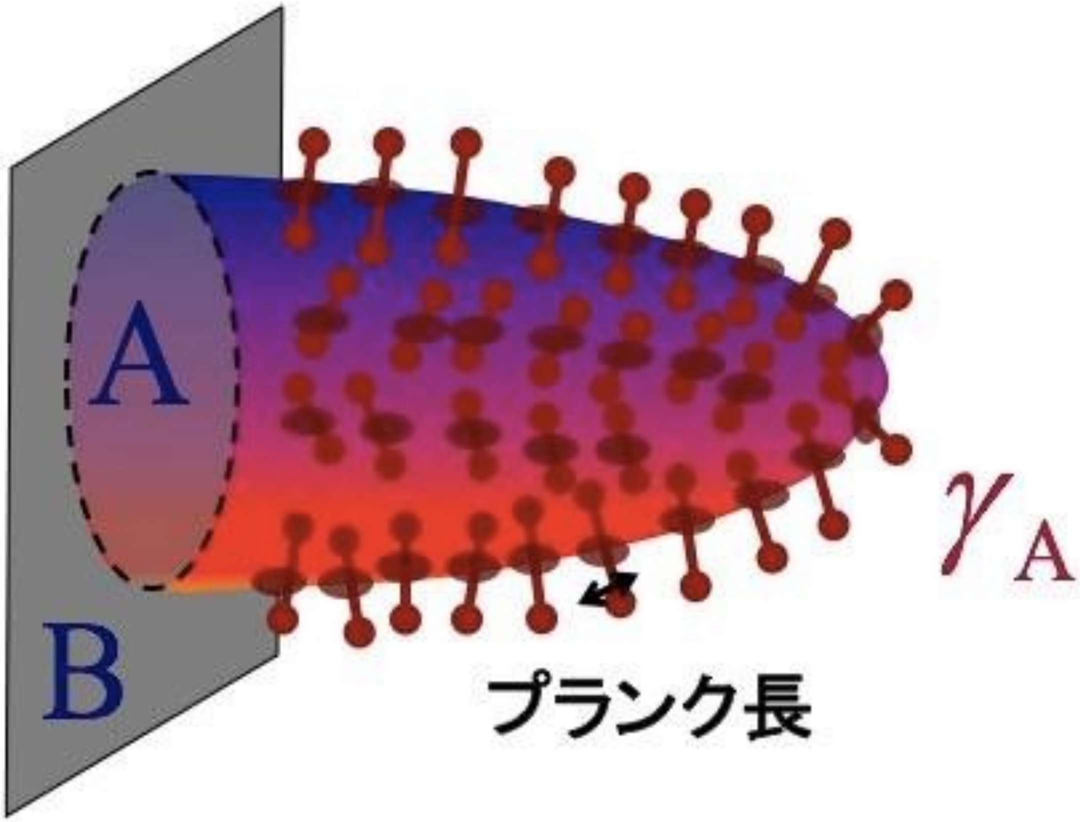
2010年
笠-高柳　エンタングルメントのエントロピー



Tadashi Takayanagi

Van Raamsdonk

γ_A AdS時空での面積を最小にする曲面（極小曲面）

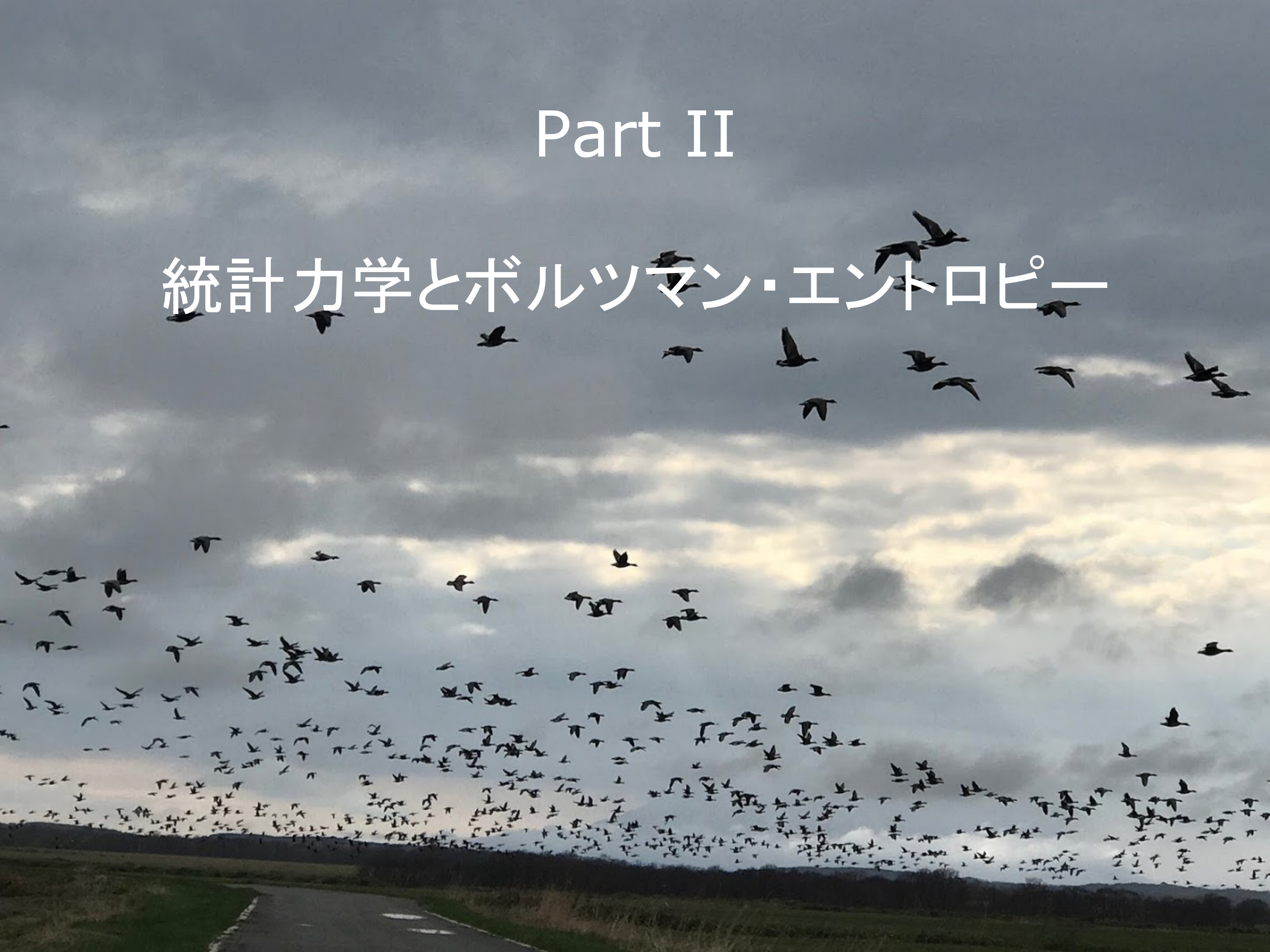


$$S_A = \frac{\mathrm{Area}(\gamma_A)}{4G_N} \approx \frac{\mathrm{Area}(\gamma_A)}{l_{pl}^2}.$$

ブラックホールの表面積のかわりに、Aを境界とする
極小曲面の表面積を考えれば、それがAdS時空での
エントロピーを与える

Part II

統計力学とボルツマン・エントロピー



Part II
統計力学とボルツマン・エントロピー

ボルツマンのエントロピーの定義

ある系の持つ可能な状態の数を考える

ボルツマンの $S = k \log W$ を計算する

ボルツマンの悲劇

ボルツマンのエントロピーの定義


S = k. log W
LVDWIG
BOLTZMANN
1844 – 1906
DR PHIL. PAULA
BOLTZMANN
GEB. CHIARI
1891 – 1977

ボルツマンのエントロピー　S = k log W
1877年

ボルツマンの墓には、彼が発見した、次のエントロピーの式が刻まれている。

$$S = k \log W$$

ここで、Sはエントロピーで、kはボルツマンの定数、Wはミクロな状態の数である。

熱力学的には エネルギーを温度で割った単位を持つ量であるエントロピーが、情報の理論と結びつくの、この式が鍵である。

ボルツマンのエントロピー　S = k log W
1877年

このボルツマンの式は、エントロピーが、ミクロな状態の数の対数 log W に「比例」するという彼の偉大な発見を表しているのだが、情報の理論と結びつくのは、この log W の部分なのである。この部分は、単なる整数の対数なので、次元（単位のこと）は無い。

先の関係式で、「エネルギー 割る 温度」というエントロピーの単位（次元のこと）は、すべて、k というボルツマンの定数の中に閉じ込められている。

ボルツマンのエントロピー　S = k log W
1877年

もしも、 k = 1 なら（実際、そういう表記をすることもある）

$$S = \log W$$

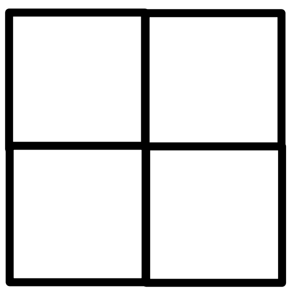
となって、エントロピーと E/T(Joul/Kelvin)という熱力学的単位との関係は、ほとんど見えなくなる。

その意味では、ボルツマンの定数k は、熱力学的エントロピーと情報論的エントロピーを結びつける、役割を持つのである。
（少なくとも、「単位・次元」を揃えるという意味では）

ある系の持つ可能な状態の数を考える

簡単な例

- 4つの箱からできているシステムを考える。
  箱は、「赤」と「青」のいずれかの状態を持つとしよう。

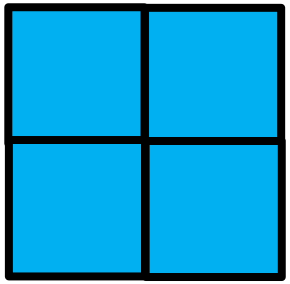


- このシステムで赤の状態の箱の数を $n_{\text{赤}}$ とし、
  このシステムで青の状態の箱の数を $n_{\text{青}}$ としよう。
  $n_{\text{赤}} + n_{\text{青}} = 4$ である。

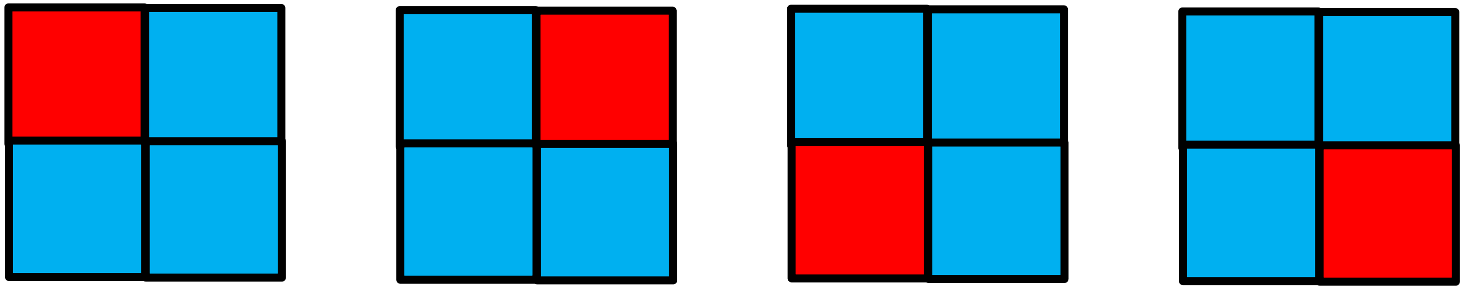
- このシステムが、どのような状態を取りうるかは、次のように、
  との場合わけをしてみればわかる。

可能な状態の数を
場合わけで数え上げる

1. $n_{赤}=0,\ n_{青}=4$ の場合　　1通り　　$\dfrac{4!}{0!\,4!}=1$

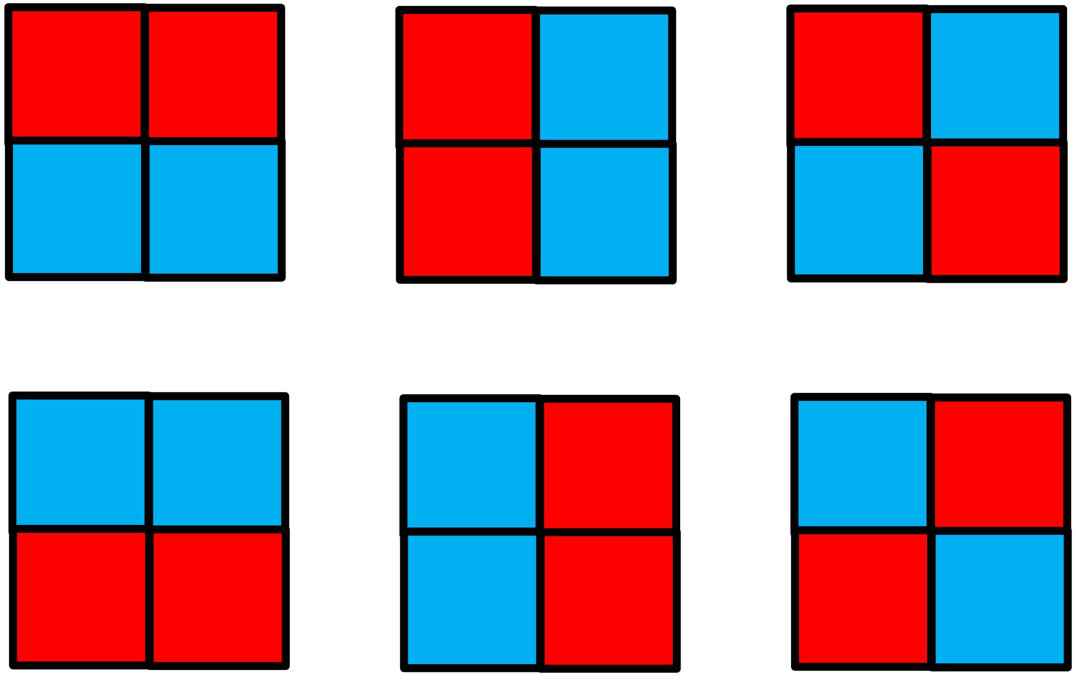


2. $n_{赤}=1,\ n_{青}=3$ の場合　　4通り　　$\dfrac{4!}{1!\,3!}=4$



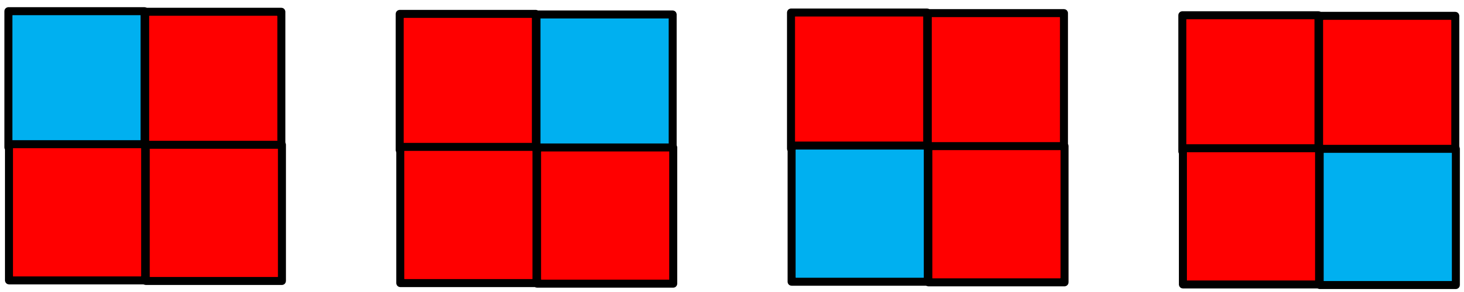
可能な状態の数を
場合わけで数え上げる

3. $n_{赤}$＝2, $n_{青}$＝2 の場合　　6通り　　$\frac{4!}{2!\,2!} = 6$

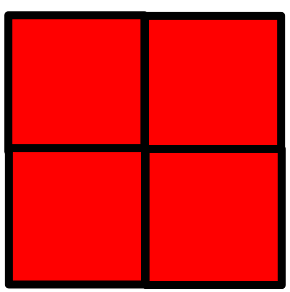


可能な状態の数を
場合わけで数え上げる

4. $n_{赤}=3$, $n_{青}=1$ の場合　　4通り　　$\dfrac{4!}{3!\,1!}=4$

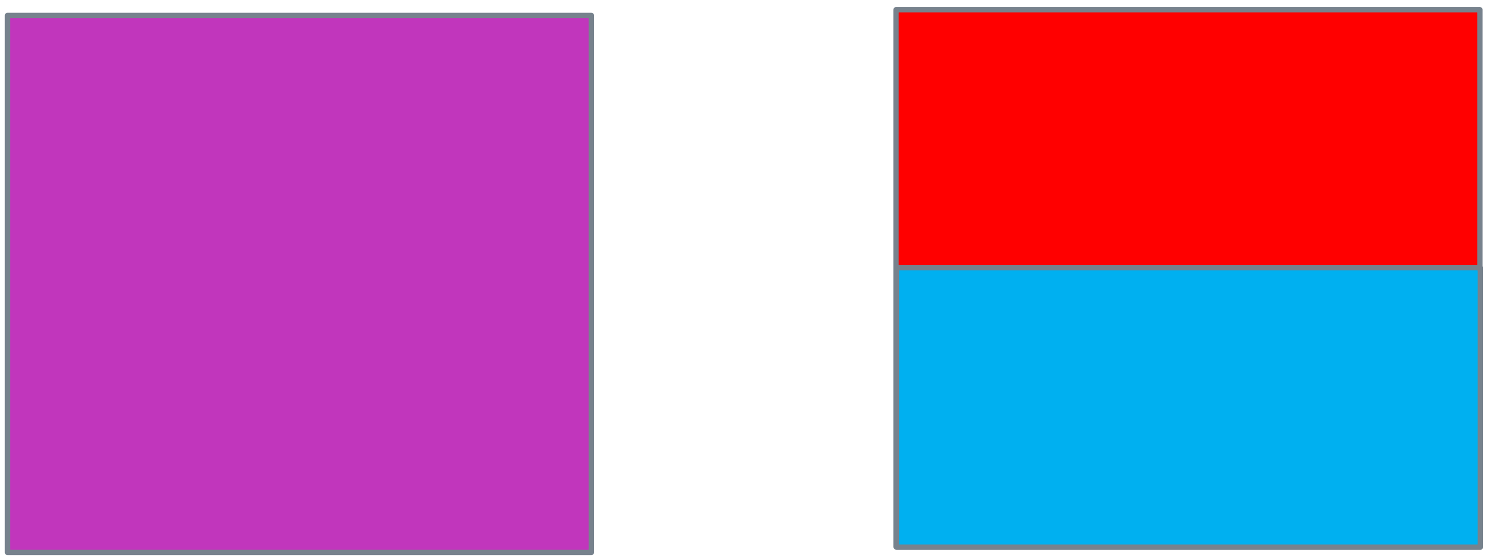


5. $n_{赤}=4$, $n_{青}=0$ の場合　　1通り　　$\dfrac{4!}{4!\,0!}=1$



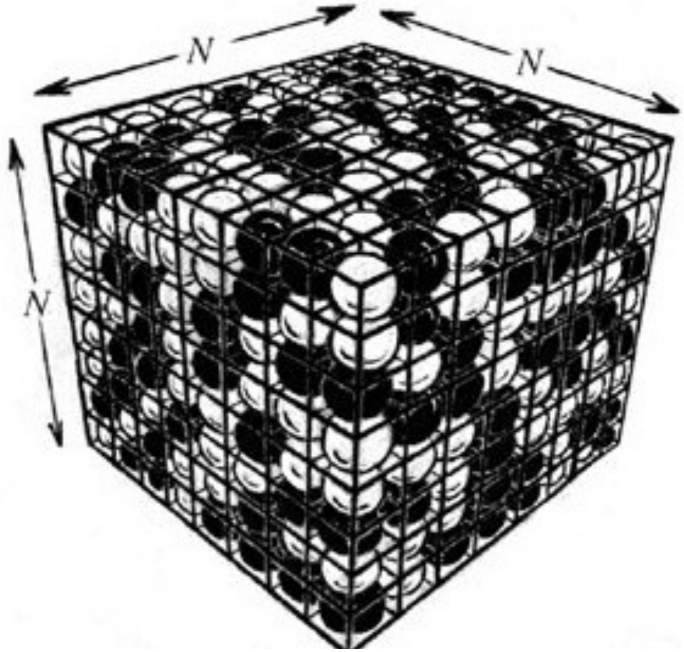
二色のペンキを混ぜる

- 同じ密度の、赤と青のペンキを、「よくかき混ぜて」みよう。そうすれば、紫の色になる。なぜ、右のようにならないのだろう？ 「よくかき混ぜた」から？



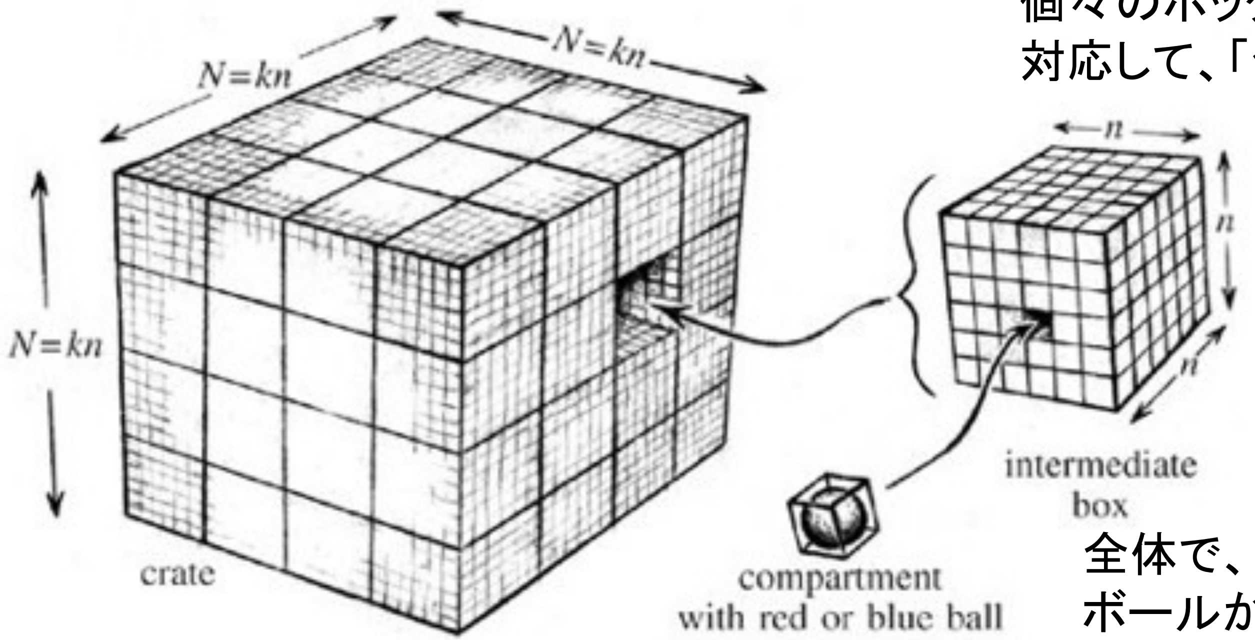
- ちなみに、青のペンキの比重が、赤のペンキより重ければ、時間がたてば、分離して、右のようになることはありうる。

赤のボールrと
青のボールbが
NxNxNのボックス
に入っているとする
ただし、rとbは、同じ
数だとしよう。



Nが小さいと、
赤のボールと
青のボールは、
それぞれ、はっきり
見える。

$1 << n << N$
と、大きさを増やしていくと、
個々のボックスも、r/bに
対応して、「色」を持つ



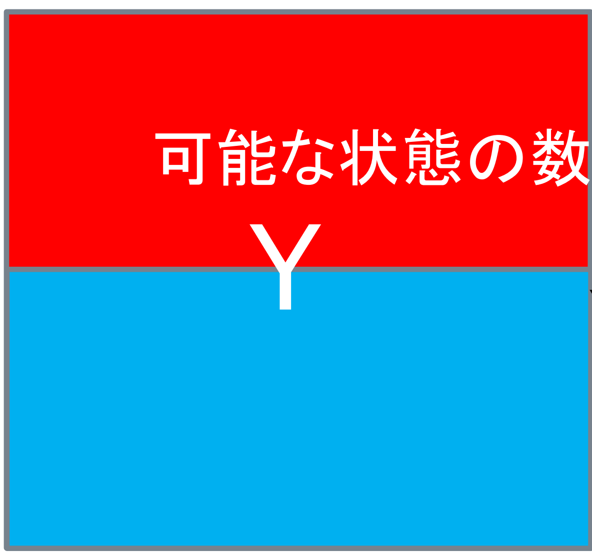
アボガドロ数
6×10^{23}

全体で、10^{24}ほどの
ボールがあるとしよう
$N = 10^8$, $k = 10^3$, $n = 10^5$

$X=10^{23\,570000\,000000\,000000\,000000}$

$Y=10^{46\,500000\,000000}$

$N=10^{24}$として、赤のボールと青のボールの可能な配置の数を計算して見る



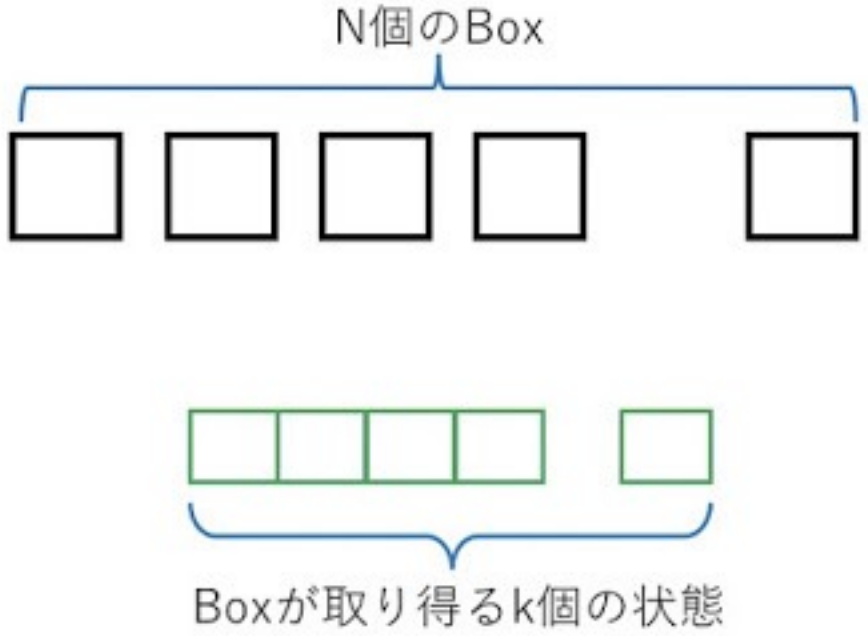
$Y=10^{46\,500000\,000000}$

$Y/X=10^{-23\,570000\,000000\,000000\,000000}$

Xである確率が、圧倒的に高い。

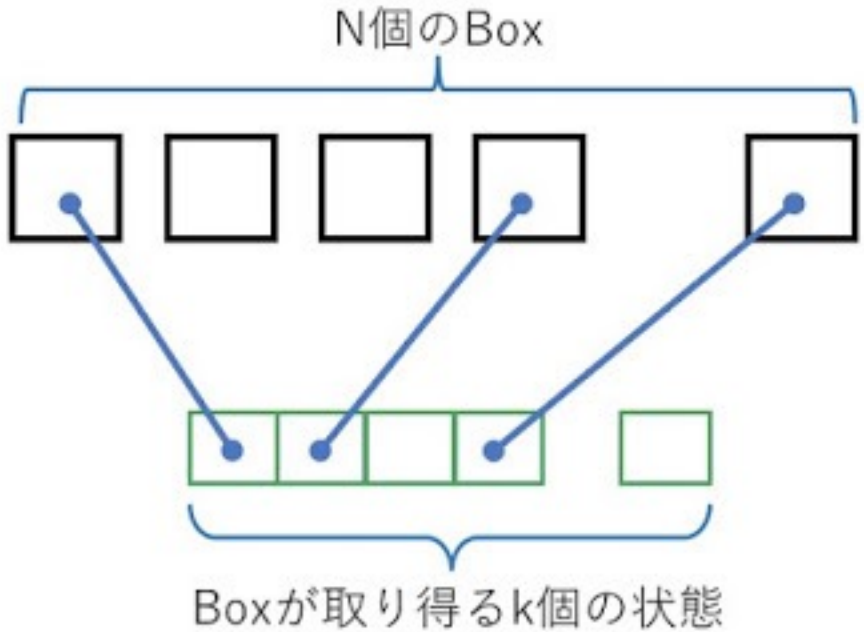
より一般に、
ある系の持つ可能な状態の数を考える

同じN個のBoxからなる系を考えよう。それぞれのBoxはk個の状態を持つとしよう。

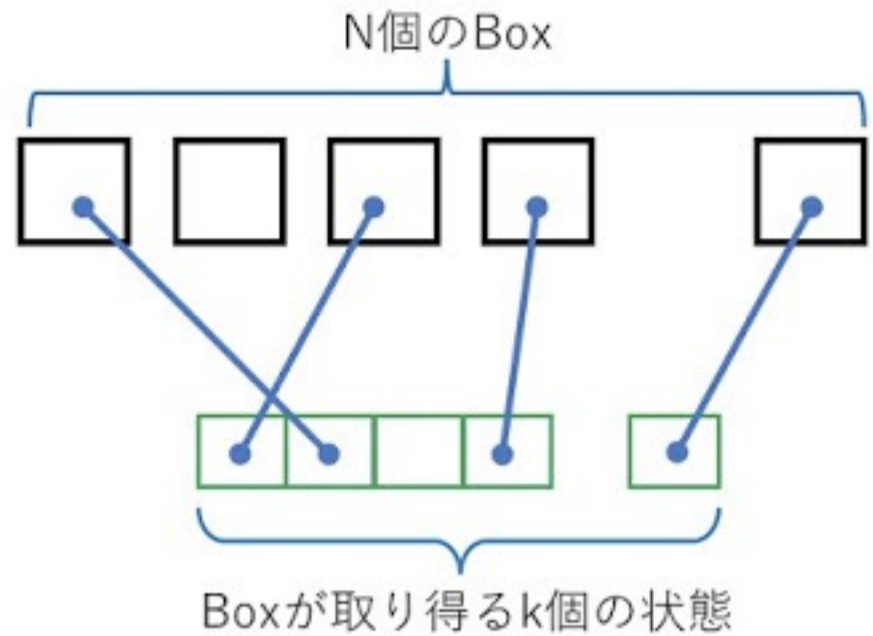


あるBoxがある状態iを取るとき、次のように線で結ぶ。

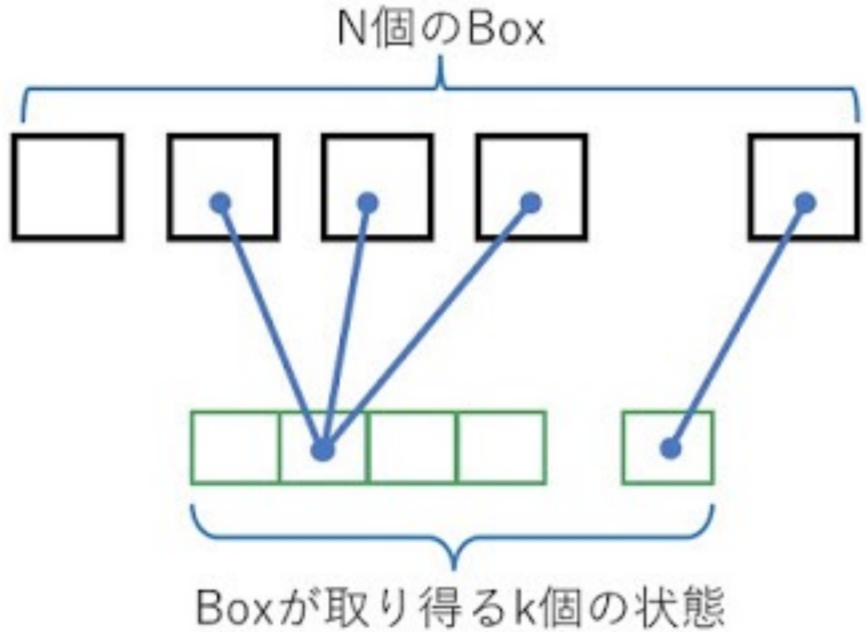
この例は、Box1が状態1を取り、Box4が状態2を取り、Box Nが状態4を取ることを表している。Box2, Box3は、この図では線で状態と結ばれていないが、それは、表記上省略しているだけである。



次の図では、Box1が状態2を取り、Box3が状態1を取り、Box4が状態4を取り、Box Nが状態kを取ることを表している。Box2の状態は、この図では省略されている。



全てのBoxは、必ず、何らかの状態を持つのだが（例では省略されていても）、全ての状態が、あるBoxの状態になるとは限らない。
次の例では、Box2, Box3, Box4 は、同じ状態2を持っている。



状態iが、何個のBoxで共有されているかを n_iで表すことにしよう。
先の例では、$n_2=3$ということになる。
$n_1+n_2+n_3+\cdots+n_k=N$ である

この時、N個のBox、k個の状態からなるシステムが取り得る可能な組み合わせの総数Wは、次のようにして計算できる。

$$\boxed{W=\frac{N!}{n_1!n_2!n_{!}\cdots n_k!}}$$

ボルツマンの

$S = k \log W$ を計算してみよう！

log Wを計算してみる

階乗は、少し扱いにくいので、次のスターリングの近似式を用いる。

$$\boxed{\log n! \simeq n \log n - n}$$

log Wを計算してみる

階乗は、少し扱いにくいので、次のスターリングの近似式を用いる。

$$\log n! \simeq n \log n - n$$

$$\boxed{\begin{aligned}\log W &= \log\left(\frac{N!}{n_1!n_2!n_! \cdots n_k!}\right) = \log N! - \sum \log n_i! = \\ &= (N \log N - N) - \sum (n_i \log n_i - n_i)\end{aligned}}$$

log Wを計算してみる

- 階乗は、少し扱いにくいので、次のスターリングの近似式を用いる。

$$\log n! \simeq n \log n - n$$

$$\log W = \log\left(\frac{N!}{n_1!n_2!n_l \cdots n_k!}\right) = \log N! - \sum \log n_i! =$$

$$= (N \log N - N) - \sum (n_i \log n_i - n_i)$$

先に見たように、$\sum n_i$ = Nなので、左の -Nと右の$\sum n_i$は、打ち消しあって消える。

$$\log W = N \log N - \sum n_i \log n_i$$

$$\log W = N \log N - \sum n_i \log n_i$$

ここで、$p_i = \frac{n_i}{N}$ とおく。 $n_i = p_i N$ である。
また、$\sum n_i = N$ なので、$\sum p_i = 1$ である。

$$\log W = N \log N - \sum n_i \log n_i$$

ここで、$p_i = \frac{n_i}{N}$ とおく。 n_i=$p_i N$ である。
また、$\sum n_i = N$ なので、$\sum p_i = 1$ である。

$$\boxed{\begin{aligned} \frac{1}{N} \log W &= \log N - \frac{1}{N}\left(\sum n_i \log n_i\right) \\ &= \log N - \sum p_i \log (p_i N) \\ &= \log N - \left(\sum p_i \log p_i + \sum p_i \log N\right) \end{aligned}}$$

$$\log W = N \log N - \sum n_i \log n_i$$

ここで、$p_i = \frac{n_i}{N}$ とおく。 $n_i = p_i N$ である。
また、$\sum n_i = N$ なので、$\sum p_i = 1$ である。

$$\frac{1}{N}\log W = \log N - \frac{1}{N}\left(\sum n_i \log n_i\right)$$
$$= \log N - \sum p_i \log(p_i N)$$
$$= \log N - \left(\sum p_i \log p_i + \sum p_i \log N\right)$$

$\sum p_i = 1$なので、$\sum p_i \log N$は、$\log N$ となって打ち消しあう。

$$\boxed{\frac{1}{N}\log W = -\sum p_i \log p_i}$$

ボルツマンのエントロピーと
シャノンのエントロピー

こうして、ボルツマンのエントロピーは、次の形に表すことができることがわかった。

$$S_B = k \log W$$
$$= -Nk \sum_i p_i \log p_i$$

これは、シャノンのエントロピー

$$\mathrm{H(X)} = -K \sum_i p_i \log p_i$$

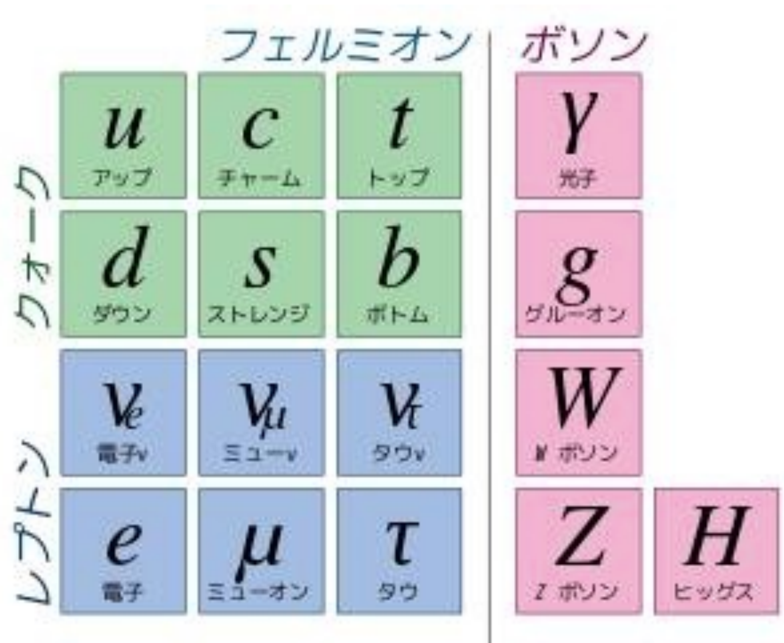
と極めて近い形をしていることがわかる。

ボルツマンの悲劇


S = k. log W
LVDWIG
BOLTZMANN
1844 – 1906
DR PHIL. PAULA
BOLTZMANN

現代物理学と原子論

- 物理現象を、より基本的な要素の実在とその運動で説明しようという「原子論」は、現代物理学の基本的な態度である。
- 20世紀の物理学の大きな成果である「素粒子の標準モデル」は、現代の原子論に他ならない。

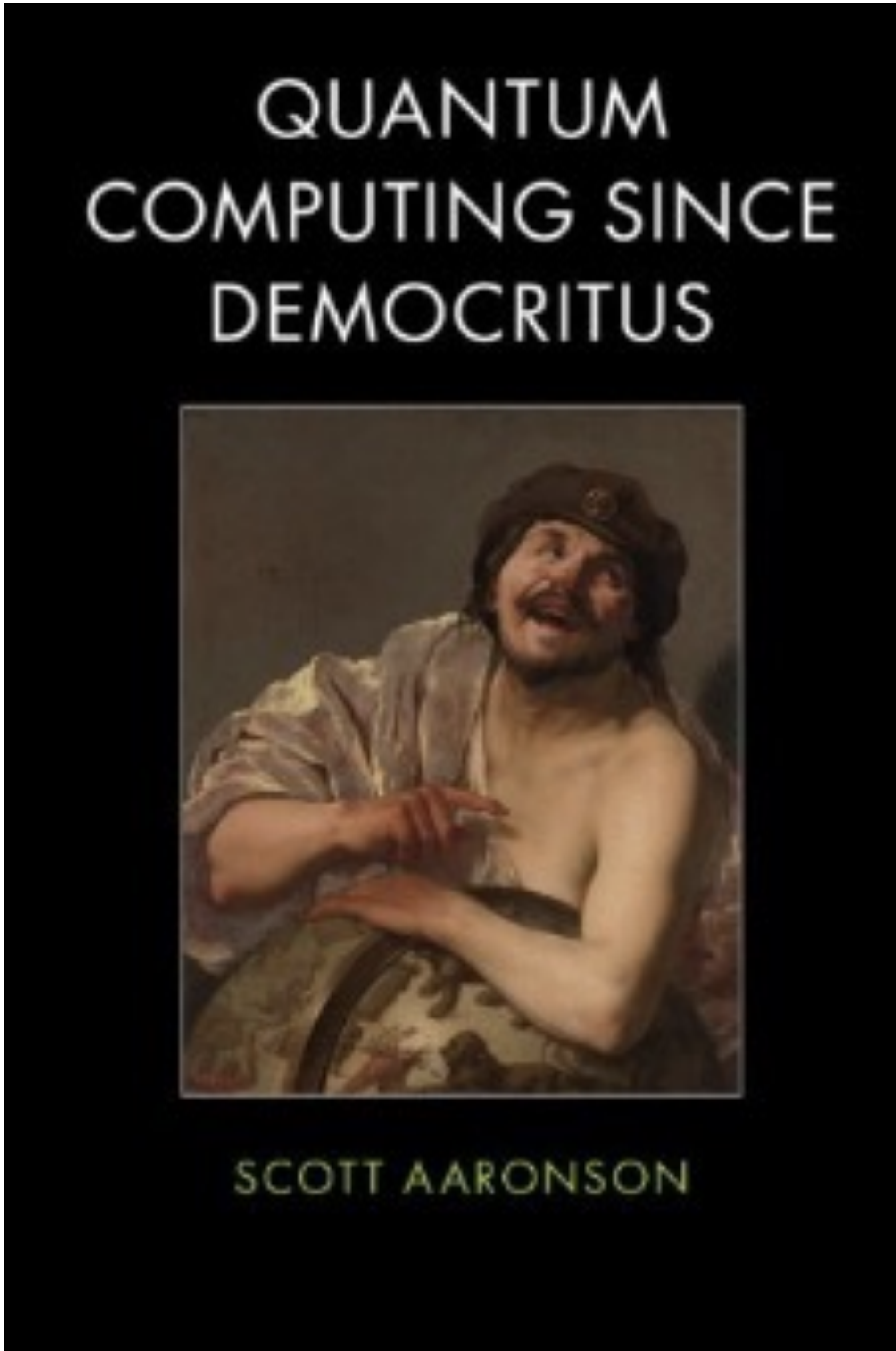


現代物理学の「標準モデル」での素粒子のリスト

現代の計算科学者で量子物理学者のScott Aaronsonは、自書

“Quantum Computing Since Democritus”
「デモクリトス以来の量子コンピューティング」

の表紙を古代ギリシャで原子論を提唱したデモクリトスで飾った。



異端の学説　原子論

- しかし、デモクリトスの著作をプラトンは、全部、焼かせたという。

異端の学説　原子論

- しかし、デモクリトスの著作をプラトンは、全部、焼かせたという。
- 次は、3世紀のキケロの原子論批判である。

異端の学説　原子論

- しかし、デモクリトスの著作をプラトンは、全部、焼かせたという。
- 次は、3世紀のキケロの原子論批判である。
- 「しかし、そのこと〔粒子の偶然の衝突で宇宙が誕生したということ〕が起こりえたと考える者は、どうして次のことに考えが至らないのか、わたしは理解に苦しむ。すなわち、黄金製であれ何であれ、二十一種類のアルファベットの文字を数えきれないほど集めて何かある容器の中に投げ込み、それらを攪拌して地面に投げ出すと、たとえばエンニウスの『年代記』のように、読者にとってちゃんと読める形になって並ぶとはどうして考えないのか。もっとも、わたしは、幸運の助けを借りたとしても、たった一行の詩句さえまともにできるかどうか、いぶかしく思っているのだが。」

https://ja.wikipedia.org/wiki/無限の猿定理

19世紀でも、原子の存在は、疑われていた

- 19世紀の物理学でも、「原子論」は「異端」の学説だったらしい。（化学の世界では、原子論は、早くから受け入れられていたようなのだが）

19世紀でも、原子の存在は、疑われていた

- 19世紀の物理学でも、「原子論」は「異端」の学説だったらしい。（化学の世界では、原子論は、早くから受け入れられていたようなのだが）
- ボルツマンは、日本でいえば、江戸時代末期に生まれ、明治維新の頃に論文を書いていると思っていい。彼は、原子論者だった。

19世紀でも、原子の存在は、疑われていた

- 19世紀の物理学でも、「原子論」は「異端」の学説だったらしい。（化学の世界では、原子論は、早くから受け入れられていたようなのだが）
- ボルツマンは、日本でいえば、江戸時代末期に生まれ、明治維新の頃に論文を書いていると思っていい。彼は、原子論者だった。
- 原子レベルのミクロな現象を日常的なマクロな現象を結びつける有力な手段は、ボルツマンが大きな貢献をした統計力学だったのだが、彼の業績は、当時、正当に評価されなかった。

ボルツマンの悲劇

- 19世紀の物理学者の双璧は、ボルツマンとマックスウェルだと思うが、二人のアカデミーでの人生は、はっきりと明暗を分けている。

ボルツマンの悲劇

- 19世紀の物理学者の双璧は、ボルツマンとマックスウェルだと思うが、二人のアカデミーでの人生は、はっきりと明暗を分けている。
- 当時の物理学会で大きな発言力を持っていたマッハとその取り巻きは、執拗にボルツマンの「原子論」を攻撃した。ボルツマンは、追い詰められ、次第に精神を病んでゆく。

ボルツマンの悲劇

- 19世紀の物理学者の双璧は、ボルツマンとマックスウェルだと思うが、二人のアカデミーでの人生は、はっきりと明暗を分けている。
- 当時の物理学会で大きな発言力を持っていたマッハとその取り巻きは、執拗にボルツマンの「原子論」を攻撃した。ボルツマンは、追い詰められ、次第に精神を病んでゆく。
- アインシュタインの画期的な論文「熱の分子論から要求される静止液体中の懸濁粒子の運動について」が出たのは、1905年5月である。ボルツマンも、この20代の若者の論文を読んだと思うのだが。

ボルツマンの悲劇

- 19世紀の物理学者の双璧は、ボルツマンとマックスウェルだと思うが、二人のアカデミーでの人生は、はっきりと明暗を分けている。
- 当時の物理学会で大きな発言力を持っていたマッハとその取り巻きは、執拗にボルツマンの「原子論」を攻撃した。ボルツマンは、追い詰められ、次第に精神を病んでゆく。
- アインシュタインの画期的な論文「熱の分子論から要求される静止液体中の懸濁粒子の運動について」が出たのは、1905年5月である。ボルツマンも、この20代の若者の論文を読んだと思うのだが。
- 翌1906年、彼は、自ら命を絶つ。

Part III

情報理論とシャノン・エントロピー

Part III
情報理論とシャノン・エントロピー

シャノンのエントロピー

シャノンが考えたこと

シャノンのエントロピーを計算する

フォン・ノイマンがシャノンに囁いたこと

シャノン・エントロピーの性質について

シャノンのエントロピー

ボルツマンのエントロピー

ボルツマンの墓には、彼が発見した、次のエントロピーの式が刻まれている。

$$S_B = k \log W$$

ここで、S_Bはボルツマンのエントロピーで、kはボルツマンの定数、Wはミクロな状態の数である。

フォン・ノイマンのエントロピー

フォン・ノイマンは、エントロピー S_Nを次の式で定義した。

$$\mathrm{S_N} = -Tr\,(\rho \log \rho)$$

ただし、Trは行列のTrace(対角成分の和)で、ρは、密度行列である。

シャノンのエントロピー

シャノンは、シャノンのエントロピーを次の式で定義した。

$$\mathrm{H(X)} = -K \sum_{i} p_i \log p_i$$

ここでKは正の定数で、p_iは、確率分布Xが与えられた時の、事象i の確率である。

シャノンが考えたこと

先に、ボルツマンのエントロピーからシャノンのエントロピーが導かれることを見てきたが、シャノンはこのようにして情報量としてのエントロピーを導いたわけではない。

シャノンは、通信システムの分析を通じて、情報量＝エントロピーの概念に到達した。ここでは、彼が、どういうことを考えていたかを振り返ってみよう。

一般的な通信システムの図式

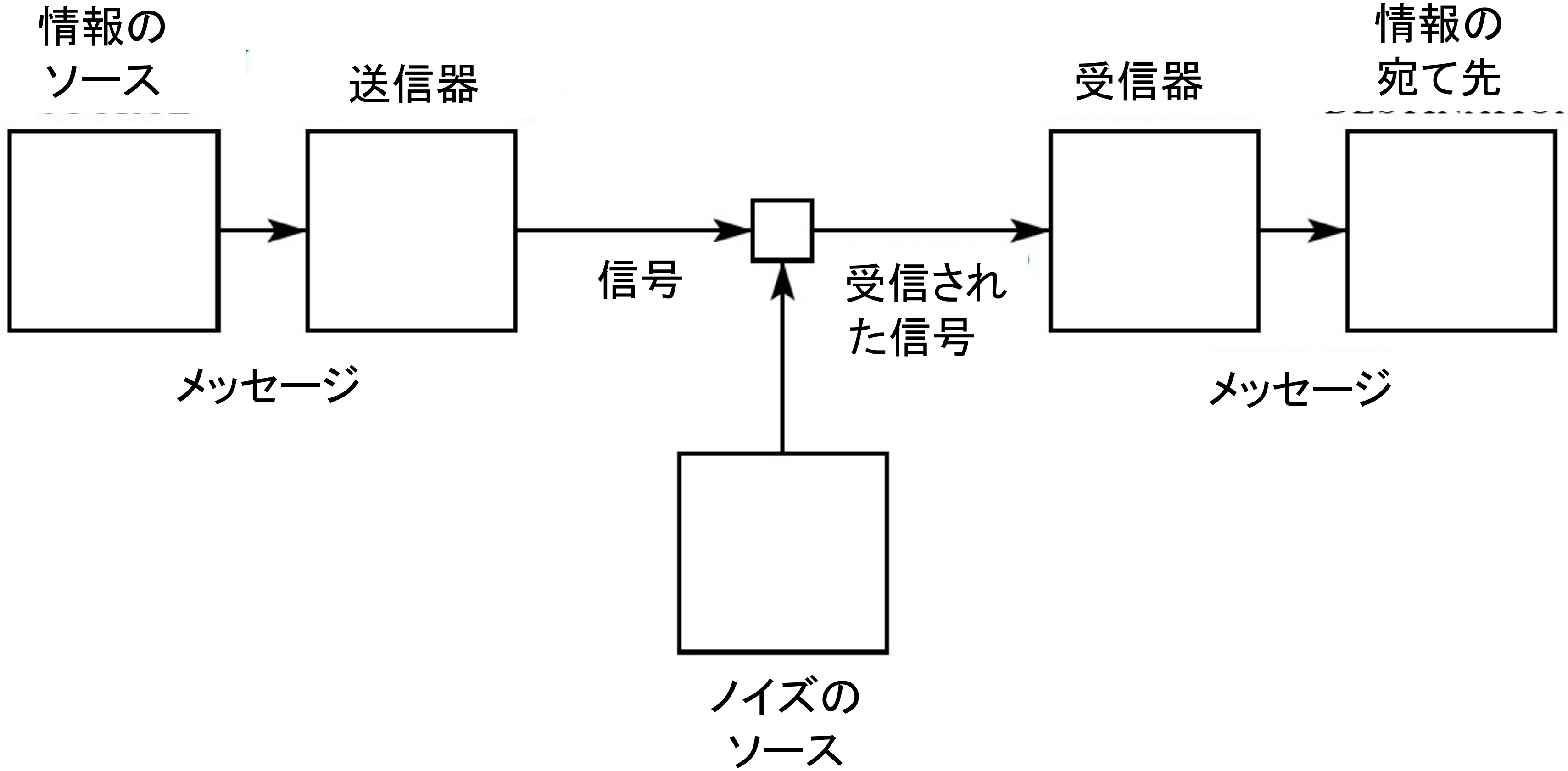
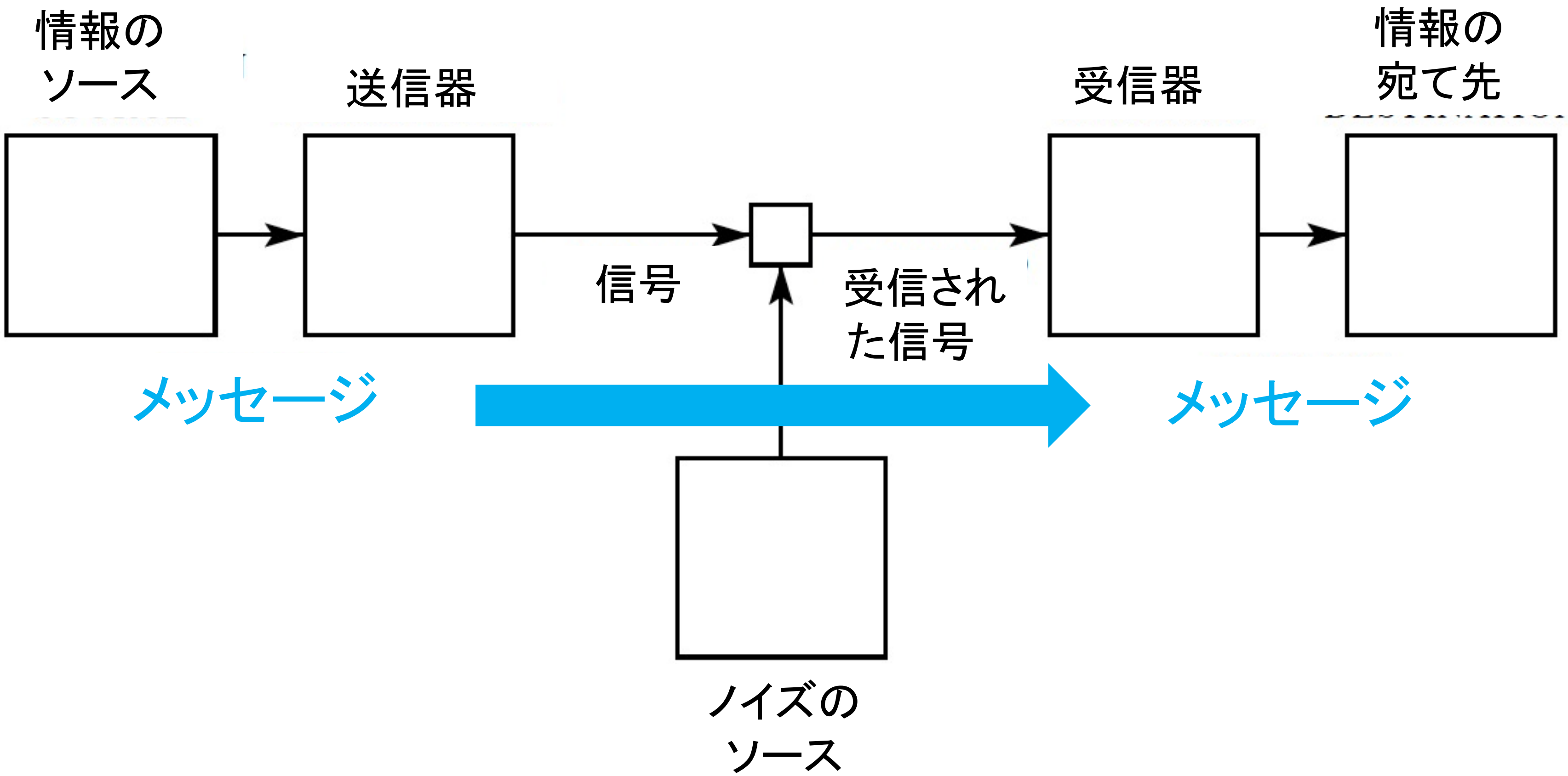


Fig. 1—Schematic diagram of a general communication system.

一般的な通信システムの図式と通信の基本問題



通信の基本的な問題は、情報のソースで選択されたメッセージを、正確であれ近似的であれ、情報の宛て先で再生産することである。

メッセージの「意味」は、重要ではない

多くの場合、そのメッセージは意味を持っている。すなわち、そのメッセージは、あるシステムに従って確かな物理的実体あるいは概念的実体を、参照するかそれに関連している。ただ、こうした通信の意味論的な諸側面は、工学的問題とは無関係である。

メッセージの「意味」は、重要ではない

多くの場合、そのメッセージは意味を持っている。すなわち、そのメッセージは、あるシステムに従って確かな物理的実体あるいは概念的実体を、参照するかそれに関連している。ただ、こうした通信の意味論的な諸側面は、工学的問題とは無関係である。

重要な側面は、実際のメッセージは可能なメッセージの集合の中から選択された一つのメッセージだと言うことである。

メッセージの「意味」は、重要ではない

多くの場合、そのメッセージは意味を持っている。すなわち、そのメッセージは、あるシステムに従って確かな物理的実体あるいは概念的実体を、参照するかそれに関連している。ただ、こうした通信の意味論的な諸側面は、工学的問題とは無関係である。

重要な側面は、実際のメッセージは可能なメッセージの集合の中から選択された一つのメッセージだと言うことである。

通信システムは、実際に選択されたメッセージだけに対してではなく、全ての可能な選択に対して機能するように設計されていなければならない。なぜなら、設計の時点では、どのメッセージが選ばれるかは、わからないからである。

情報の尺度として「対数」を利用する

もしも、選択されるメッセージの集合の要素の数が有限であるなら、この集合から一つのメッセージが選ばれた時、その選択はすべて同じようにありうるので、この数あるいはこの数の単調な関数はすべて、生成されたこの情報の尺度と見なすことができる。

情報の尺度として「対数」を利用する

もしも、選択されるメッセージの集合の要素の数が有限であるなら、この集合から一つのメッセージが選ばれた時、その選択はすべて同じようにありうるので、この数あるいはこの数の単調な関数はすべて、生成されたこの情報の尺度と見なすことができる。

Hartlyが指摘したように、この関数として最も自然なのは、対数を選ぶことである。

この定義は、メッセージの統計的性質の影響や、メッセージの値が連続的なものである場合など、よく考えた上で一般化されねばならないのだが、我々は全ての場合で、情報の尺度として、本質的には対数的な尺度を利用するだろう。

対数的尺度が便利であるいくつかの理由

工学的に重要な、時間。帯域、リレーの数といったパラメーターは、可能な数の対数について、線形に変化する傾向がある。

例えば、リレーを一個追加すると、可能な状態の数は2倍になる。元の状態を数をNとすれば、リレー一個を追加した時可能な状態の数は2Nとなる。状態の数 N→2Nの変化は、2を底とする対数で考えれば、$\log N \rightarrow \log 2N = \log 2 + \log N = 1 + \log N$ である。

対数的尺度が便利であるいくつかの理由

工学的に重要な、時間。帯域、リレーの数といったパラメーターは、可能な数の対数について、線形に変化する傾向がある。

例えば、リレーを一個追加すると、可能な状態の数は2倍になる。元の状態を数をNとすれば、リレー一個を追加した時可能な状態の数は2Nとなる。状態の数 N→2Nの変化は、2を底とする対数で考えれば、log N → log 2N = log 2 + log N = 1 + log N である。

時間が2倍になれば、可能なメッセージの数は、元の可能なメッセージの数の2乗になるが、対数で考えれば、2倍になる。

対数的尺度が便利であるいくつかの理由

我々は、直感的には、ある実体の尺度を、共通の基準と線形で比較する。

二枚のパンチカードは一枚のカードより、2倍の情報容量を持つと感じる。

二つの同じ通信チャンネルがあれば、一つの通信チャンネルより、2倍の情報容量を持つと感じる。

情報の尺度としての bit

対数の底の選択は、情報の尺度の単位の選択に対応している。
対数の底に 2を選んだ時、その単位は bit と呼ばれる。

二つの状態を取るデバイスは、log 2 = 1 bitの情報を蓄える。
こうしたデバイスががN個あった時、N個のデバイスが取りうる状態の総数は、2^Nである。$\log 2^N = N$ だから、これらのデバイスは N bitの情報を蓄えることになる。

選択と不確かさとエントロピー

我々は、離散的な情報源をマルコフ過程として表現してきた。

我々は、ある意味で、どれほどの情報がこうしたプロセスによって生成されたかを、さらには、どのような割合で情報が生成されたかを測定する量を定義できるだろうか？

選択と不確かさとエントロピー

我々は、離散的な情報源をマルコフ過程として表現してきた。

我々は、ある意味で、どれほどの情報がこうしたプロセスによって生成されたかを、さらには、どのような割合で情報が生成されたかを測定する量を定義できるだろうか？

それが生起する確率が、p1, p2, … , pnの可能なイベントの集合が与えられたとしよう。これらの確率は知られているが、我々が知っていることのすべては、どれかのイベントが起きるだろうと言うことのみである。

選択と不確かさとエントロピー

我々は、離散的な情報源をマルコフ過程として表現してきた。

我々は、ある意味で、どれほどの情報がこうしたプロセスによって生成されたかを、さらには、どのような割合で情報が生成されたかを測定する量を定義できるだろうか？

それが生起する確率が、p1, p2, ... , pnの可能なイベントの集合が与えられたとしよう。これらの確率は知られているが、我々が知っていることのすべては、どれかのイベントが起きるだろうと言うことのみである。

そのイベントを選び出すのに、どれほどの「選択」が含まれているのか、あるいは、その出力について、我々はいかに不確かなのかを測定する尺度を、我々は見つけ出すことができるだろうか？

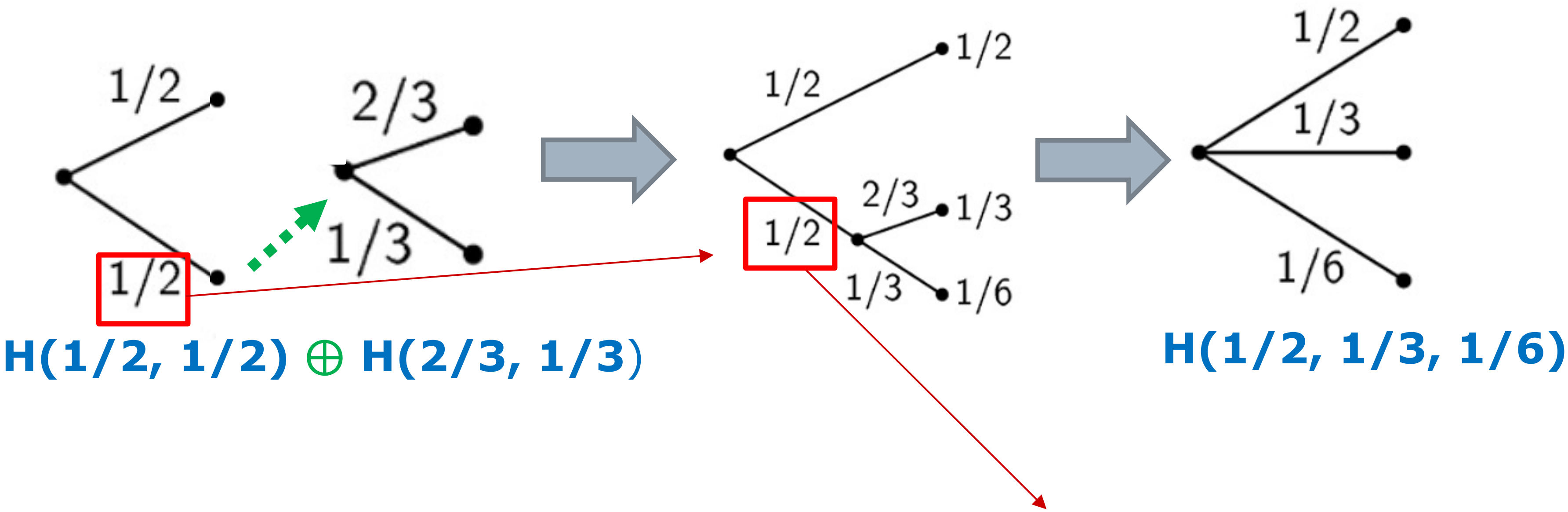
Hが満たすべき条件

もし、そうした尺度があるとして、それを H(p1, p2, … ,pn) と表そう。この時、この関数Hが、次のような性質を満たすことを要求するのは合理的である。

1. H は、pi について連続的である。
2. もし全てのpiが等しく、pi ＝ 1/n であるとすれば、Hは、nについて単調増加の関数である。
   （同じように見えるイベントが与えられ、より多くの可能なイベントが存在する時、より多くの選択、より多くの不確かさが存在することになる。）
3. もし選択が、連続的に行われる二つの選択に分解されるのなら、求めるHは、個々の選択のHの値の、重みづけられた和になるべきである。

先の条件 3. の説明

H(1/2, 1/2)で表される選択と、H(2/3, 1/3)で表される選択を、次のように連続して行なったとする。



H(1/2, 1/3, 1/6) = H(1/2, 1/2) + 1/2 H(2/3, 1/3)

シャノンが証明したこと

先の 1,2,3の条件を満たす関数 Hは、ただ一つで、次の形をしている。(ただし、Kは正の定数)

$$\mathrm{H}(p_1, p_2, \ldots, p_n) = -K \sum_i p_i \log p_i$$

フォン・ノイマンがシャノンに囁いたこと

フォン・ノイマンのもう一つの「貢献」

エントロピーの歴史の中で、フォン・ノイマンは、エントロピーの量子化だけでなく、もう一つの重要な「貢献」をしている。

1940年、それまでの熱力学的エントロピー概念の系譜とは全く独立にシャノンが発見した「不確実さの関数」＝「情報量」の概念に、「エントロピー」という名前をつけることを勧めたのは、フォン・ノイマンだと言われている。
http://www.eoht.info/page/Neumann-Shannon%20anecdote

Wikipedia の "Talk:History of entropy" にも、詳しい記事がある。
https://en.wikipedia.org/wiki/Talk%3AHistory_of_entropy

フォン・ノイマンのもう一つの「貢献」

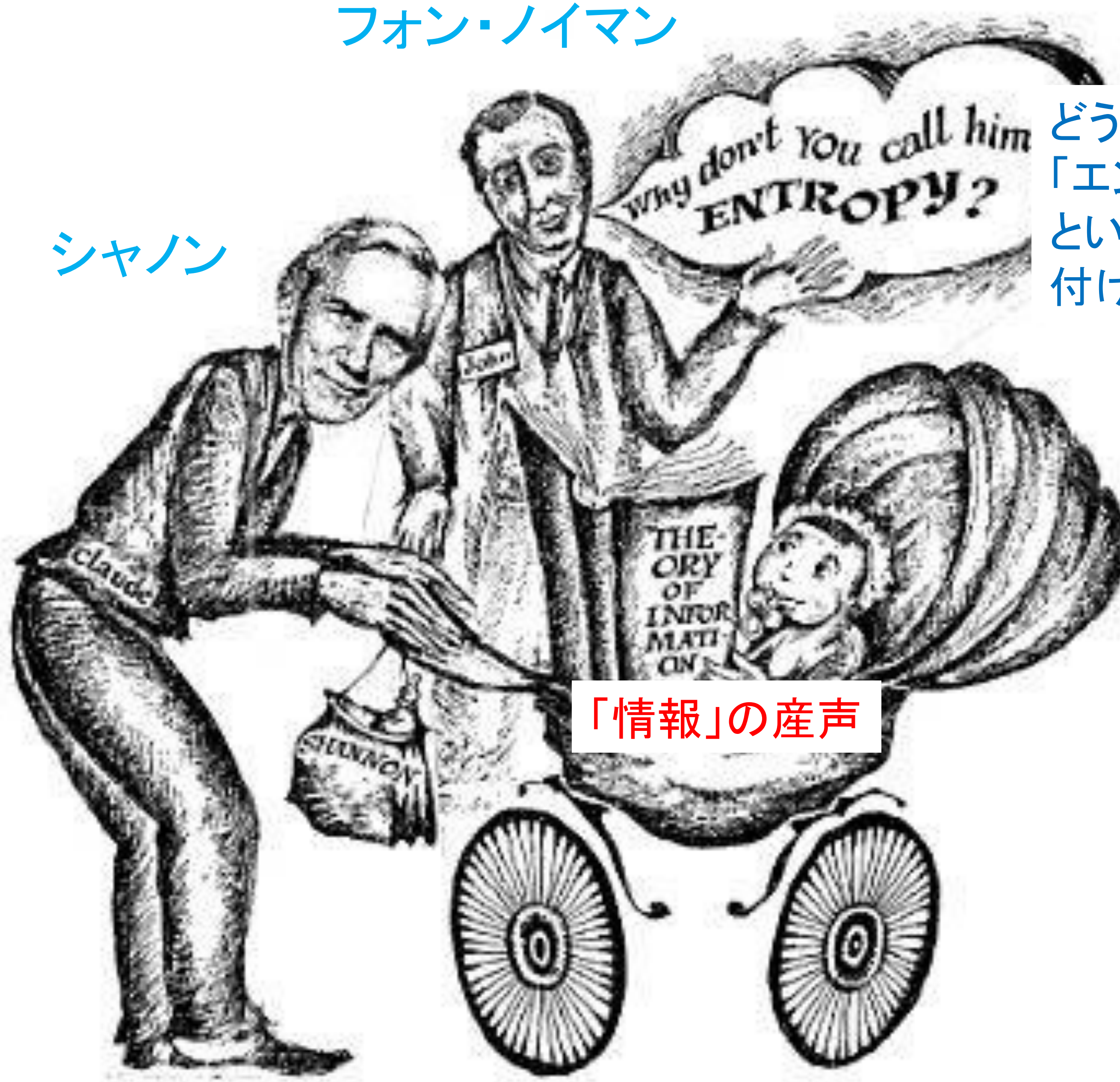
“You should call it ‘entropy’ for two reasons:
First, the function is already used
in thermodynamics under that name;

second, and more importantly, most people don’t
know what entropy really is, and if you use the word
‘entropy’ in an argument you will win ever time.”

“nobody really knows what entropy is anyway.”
「どうせ誰もエントロピーがどんなものかわかっていないんだから」

フォン・ノイマン

シャノン



どうして彼に
「エントロピー」
という名前を
付けないんだ？

「情報」の産声


フォン・ノイマン
シャノン
どうせ誰も「エントロピー」がどんなものかわかっていないんだから
THE-
ORY
OF
INFOR
MATI-
ON
Claude
John
SHANNON


フォン・ノイマン
「エントロピー」という
言葉を使えば、いつだって
議論に勝てるだろう。
シャノン
John
Claude
THEORY OF INFORMATION
SHANNON

シャノン・エントロピーの性質について

シャノンのエントロピー H(X)の定義

シャノンは、シャノンのエントロピーを次の式で定義した。

$$\mathrm{H(X)} = -\sum_{i} p_i \log_2 p_i$$

p_iは、確率分布Xが与えられた時の、事象i の確率である。

$\mathrm{H(X)} \geq 0$

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

と変形できる。（ $-\log a = \log a^{-1} = \log \frac{1}{a}$ ）

$$\mathrm{H(X)} \geq 0$$

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

と変形できる。（ $-\log a = \log a^{-1} = \log \frac{1}{a}$ ）

p_iは確率だから、0 $\leq p_i \leq 1$ で $\frac{1}{p_i} \geq 1$ である。

$$\mathrm{H(X)} \geq 0$$

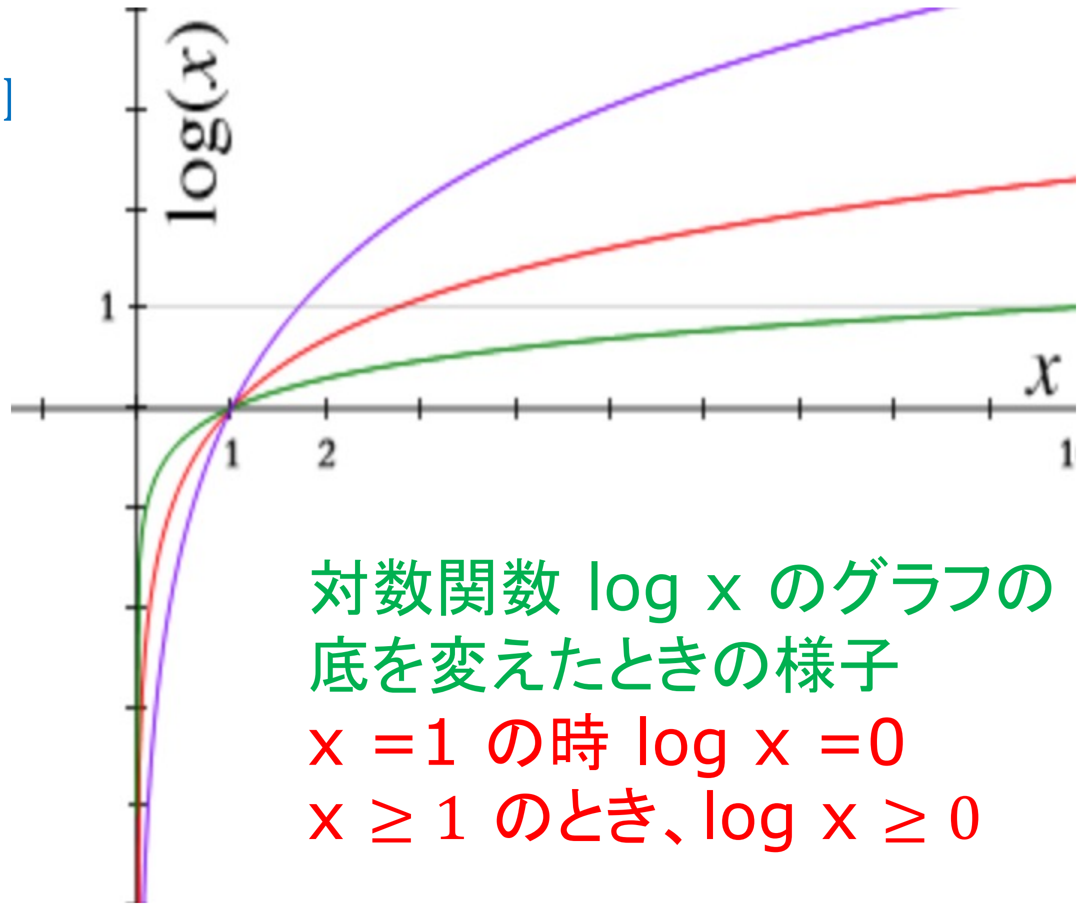
$$\mathrm{H(X)} = -\sum_i p_i \log_2 p_i = \sum_i p_i \log_2 \frac{1}{p_i}$$

と変形できる。（ $-\log a = \log a^{-1} = \log \frac{1}{a}$ ）

p_iは確率だから、$0 \leq p_i \leq 1$ で $\frac{1}{p_i} \geq 1$ である。

この時、対数の性質から、$\log_2 \frac{1}{p_i} \geq 0$

対数の性質



対数関数 log x のグラフの
底を変えたときの様子

x =1 の時 log x =0

x ≥ 1 のとき、log x ≥ 0

$\mathrm{H(X)} \geq 0$

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

と変形できる。（ $-\log\ a = \log a^{-1} = \log\ \frac{1}{a}$ ）

p_iは確率だから、$0 \leq p_i \leq 1$ で $\frac{1}{p_i} \geq 1$ である。

この時、対数の性質から、$\log_2 \frac{1}{p_i} \geq 0$

よって、p_i も $\log_2 \frac{1}{p_i}$ も ≥ 0 であり $p_i \log_2 \frac{1}{p_i} \geq 0$

これから $\mathrm{H(X)} \geq 0$がわかる。

$H(X) = 0$ となる場合

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

が 0となるのは、$p_i = 0$ あるいは $\log_2 \frac{1}{p_i} = 0$ すなわち、$p_i = 1$ の場合である。ただし、

$$\sum_i p_i = 1$$

である。

H(X) = 0 となる場合

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

が 0となるのは、$p_i = 0$ あるいは $\log_2 \frac{1}{p_i}$ = 0 すなわち、$p_i = 1$ の場合である。ただし、

$$\sum_i p_i = 1$$

であるので、

H(X) = 0 となるのは、ある一つの i についてのみ、$p_i = 1$ となり、それ以外の全ての jについて、$p_j = 0$ になる場合 である。

$\mathrm{H(X)} = 0$ となる場合

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

が 0となるのは、$p_i = 0$ あるいは $\log_2 \frac{1}{p_i} = 0$ すなわち、$p_i = 1$ の場合である。ただし、

$$\sum_i p_i = 1$$

であるので、

H(X) = 0 となるのは、ある一つの i についてのみ、$p_i = 1$ となり、それ以外の全ての jについて、$p_j = 0$ になる場合 である。

ある事象が確実に起こるとき、そのエントロピーはゼロになる。

H(X)が最大の値をとる場合

逆に、ある事象が起きることが不確実なものになるにつれて、エントロピーは増大する。

n個の事象 X=$\{x_1, x_2,\ x_3, \ldots,\ \ x_n\}$ に対してX上の
確率分布　P=$\{p_1, p_2,\ p_3, \ldots,\ \ p_n\}$ が与えられているとしよう。
この時、H(X) ＝ H($p_1, p_2,\ p_3, \ldots,\ \ p_n$)と表すことにする。

H(X)が最大の値をとる場合

逆に、ある事象が起きることが不確実なものになるにつれて、エントロピーは増大する。

n個の事象 X=$\{x_1, x_2,\ x_3, \ldots,\ \ x_n\}$ に対してX上の
確率分布　P=$\{p_1, p_2,\ p_3, \ldots,\ \ p_n\}$ が与えられているとしよう。
この時、H(X) = H($p_1, p_2,\ p_3, \ldots,\ \ p_n$)と表すことにする。

H(X)が最大になるのは、どのx_i が選択されるか確実には分からない場合、すなわち、$p_1 = p_2 = p_3 = \cdots = p_n$の場合である。

H(X)が最大の値をとる場合

逆に、ある事象が起きることが不確実なものになるにつれて、エントロピーは増大する。

n個の事象 X={$x_1, x_2, x_3, \dots, x_n$} に対してX上の
確率分布　P={$p_1, p_2, p_3, \dots, p_n$} が与えられているとしよう。
この時、H(X) = H($p_1, p_2, p_3, \dots, p_n$)と表すことにする。

H(X)が最大になるのは、どのx_i が選択されるか確実には分からない場合、すなわち、$p_1 = p_2 = p_3 = \cdots = p_n$の場合である。

これは、$p_1 + p_2 + p_3 + \cdots + p_n = 1$ だから、全てのiについて
p_i=1/n の場合である。

H(X)が最大の値をとる場合

逆に、ある事象が起きることが不確実なものになるにつれて、エントロピーは増大する。

n個の事象 X={$x_1, x_2, x_3, \ldots, x_n$} に対してX上の
確率分布　P={$p_1, p_2, p_3, \ldots, p_n$} が与えられているとしよう。
この時、H(X) = H($p_1, p_2, p_3, \ldots, p_n$)と表すことにする。

H(X)が最大になるのは、どのx_i が選択されるか確実には分からない場合、すなわち、$p_1 = p_2 = p_3 = \cdots = p_n$の場合である。
これは、$p_1 + p_2 + p_3 + \cdots + p_n = 1$ だから、全てのiについてp_i=1/n の場合である。この時、エントロピーの最大値は、

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

シャノンエントロピーの単位としての bit

X={0,1} の時、H(X)の最大エントロピーは、

$H(1/2,\ 1/2) = \log_2 2 = 1$

これをエントロピーの単位として bit と呼ぶ。

シャノンエントロピーの単位としての bit

X={0,1} の時、H(X)の最大エントロピーは、
H(1/2, 1/2) = $\log_2 2 = 1$
これをエントロピーの単位として bit と呼ぶ。
bit は、二つの状態を持つものの最大エントロピーである。

シャノンエントロピーの単位としての bit

X={0,1} の時、H(X)の最大エントロピーは、

$$H(1/2,\ 1/2) = \frac{1}{2}\log_2 2 + \frac{1}{2}\log_2 2 = \log_2 2 = 1$$

これをエントロピーの単位として bit と呼ぶ。

bit は、二つの状態を持つものの最大エントロピーである。

X={0,1,2,3}の時、H(X)の最大エントロピーは

$$H(1/4,\ 1/4,\ 1/4,\ 1/4) = \log_2 4 = 2\log_2 2 = 2\ bit$$

四つの状態を持つものの最大エントロピーは 2bit である。

シャノンエントロピーの単位としての bit

X={0,1} の時、H(X)の最大エントロピーは、

H(1/2, 1/2) $= \log_2 2 = 1$

これをエントロピーの単位として bit と呼ぶ。

bit は、二つの状態を持つものの最大エントロピーである。

X={0,1,2,3}の時、H(X)の最大エントロピーは

H(1/4, 1/4, 1,4, 1/4) $= \log_2 4 = 2\log_2 2 = 2\,bit$

四つの状態を持つものの最大エントロピーは 2bit である。

一般に 2^nの状態を持つものの最大エントロピーは

$$\log_2 2^n = n\log_2 2 = n\,bit$$

シャノンエントロピーの単位としての bit

X={0,1} の時、H(X)の最大エントロピーは、
H(1/2, 1/2) = $\log_2 2 = 1$
これをエントロピーの単位として bit と呼ぶ。
bit は、二つの状態を持つものの最大エントロピーである。

X={0,1,2,3}の時、H(X)の最大エントロピーは
H(1/4, 1/4, 1,4, 1/4) = $\log_2 4 = 2\log_2 2 = 2\,bit$
四つの状態を持つものの最大エントロピーは 2bit である。

一般に 2^nの状態を持つものの最大エントロピーは

$$\log_2 2^n = n\log_2 2 = n\,bit$$

これは、我々のbit についての直観と一致している。


Part IV
シャノン・エントロピーの
通信技術への応用

Part IV
シャノン･エントロピーの通信技術への応用

通信の世界でメッセージのエントロピーを考える

ボルツマンとシャノンのエントロピーの比較

エントロピーをグラフィカルに表現する

クロスエントロピーをグラフィカルに表現する

相対エントロピー

通信の世界でメッセージの
エントロピーを考える

この節の目標

シャノンエントロピーの基本的な性格を見てきたが、まだ抽象的で、現実の情報や通信の世界とはかけ離れていると思われたかもしれない。

この節の目標は、「通信の世界」をモデルに、シャノンエントロピーを、改めて導出することである。

具体的には、ある言語(例えば「英語」)で書かれた N文字のメッセージ(例えば、"I love you.")のエントロピーを考えてみることにする。

基本的な想定

通信のモデルを考える上で、次のような想定をしている。

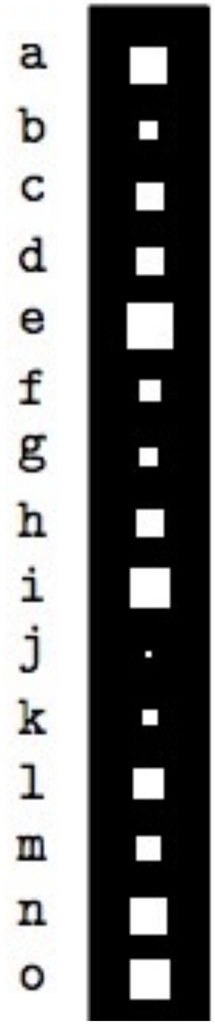
1. ある言語 L には、「アルファベット」 a_iが l 個あるとする。
2. ある言語 L の「メッセージ」は、L の「アルファベット」 a_iの並びからなる文字列である。
3. ある言語 L の文字列の中で、「アルファベット」の一文字 a_iが使われる出現確率をp_i とする。

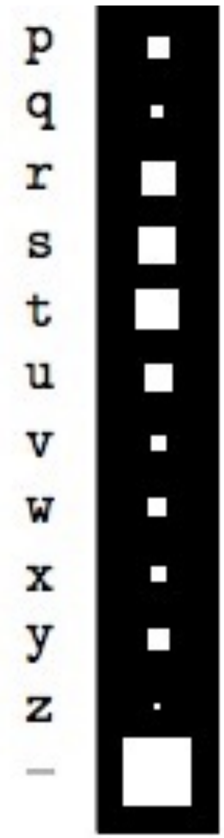
次に、英語のテキストの場合の「アルファベット」 a_iとその出現確率p_iを示す。(ここでのアルファベットの個数は l = 27 である。)

英文テキストの場合でのアルファベット27文字の出現確率

（大文字・小文字の区別なし。句読点も数字も無視している）

i	a_i	p_i
1	a	0.0575
2	b	0.0128
3	c	0.0263
4	d	0.0285
5	e	0.0913
6	f	0.0173
7	g	0.0133
8	h	0.0313
9	i	0.0599
10	j	0.0006
11	k	0.0084
12	l	0.0335
13	m	0.0235
14	n	0.0596
15	o	0.0689
16	p	0.0192
17	q	0.0008
18	r	0.0508
19	s	0.0567
20	t	0.0706
21	u	0.0334
22	v	0.0069
23	w	0.0119
24	x	0.0073
25	y	0.0164
26	z	0.0007
27	–	0.1928





出現確率が一番高いのは、スペース

e, t, o,.. がそれに続く。

シャノンのエントロピーを計算する

N個の長さの文字列のメッセージのエントロピーを考える

先の想定に従って、言語 L には、l 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。

ここでは、まず、Lの長さNのメッセージ（文字列）の持つエントロピー Mを次のように定義して、それを計算してみよう。

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. $M = \log_2 W$ で定義する。

この M の定義は、ボルツマンのエントロピーの定義によく似ている。両者の関係については、後で述べる。

「典型的なメッセージ」？

言語 Lのアルファベットa_iから構成される長さNのメッセージには、例えば、英語の例で言えば、'a' がN個連なったものや、 'b' がN個連なったものが含まれている。これらは、メッセージ中での'a'や'b'の出現が突出して多いので、「典型的なメッセージ」とは言わない。

「典型的なメッセージ」とは、メッセージ中のアルファベットa_iの出現頻度も、出現確率 p_iに従うようなメッセージである。

N文字からなる「典型的なメッセージ」では、

アルファベットa_1はp_1N 個出現し、アルファベットa_2はp_2N 個出現し、… 、アルファベットa_lはp_lN 個出現する。

M = $\log_2$ W を計算する

この時、

$$W = \left(\text{典型的な、異なるメッセージの数} \right) = \boxed{\frac{N!}{(p_1N)!(p_2N)!\cdots(p_\ell N)!}}$$

だから

$$M = \log_2 \left(\text{典型的な、異なるメッセージの数} \right) = \log_2 \left(\frac{N!}{(p_1N)!(p_2N)!\cdots(p_\ell N)!} \right)$$

M = $\log_2$ W を計算する

Mに、スターリングの公式: log(N!) ≈ N log N – N を使う

$$\log_2\left(\frac{N!}{(p_1N)!(p_2N)!\cdots(p_\ell N)!}\right)=\log_2(N!)-\{\log_2((p_1N)!)+\ldots+\log_2((p_\ell N)!)\}$$

$$\approx(N\log_2 N-N)-\{(p_1N\log_2 p_1N-p_1N)+\ldots+(p_\ell N\log_2 p_\ell N-p_\ell N)\}$$

M ＝ $\log_2$ W を計算する

Nを括り出す

$$\log_2\left(\frac{N!}{(p_1N)!(p_2N)!\cdots(p_\ell N)!}\right)=\log_2(N!)-\{\log_2((p_1N)!)+\ldots+\log_2((p_\ell N)!)\}$$

$$\approx(N\log_2 N-N)-\{(p_1N\log_2 p_1N-p_1N)+\ldots+(p_\ell N\log_2 p_\ell N-p_\ell N)\}$$

$$=N\{\log_2 N-1-p_1\log_2 p_1-p_1\log_2 N+\underline{p_1}-\ldots-p_\ell\log_2 p_\ell-p_\ell\log_2 \underline{N}+p_\ell\}$$

M = $\log_2$ W を計算する

青い線の部分は打ち消しあって、消える

$$\log_2\left(\frac{N!}{(p_1N)!(p_2N)!\cdots(p_\ell N)!}\right) = \log_2(N!) - \{\log_2((p_1N)!) + \ldots + \log_2((p_\ell N)!)\}$$

$$\approx (N\log_2 N - N) - \{(p_1N\log_2 p_1N - p_1N) + \ldots + (p_\ell N\log_2 p_\ell N - p_\ell N)\}$$

$$= N\{\underline{\log_2 N} - 1 - p_1\log_2 p_1 - \underline{p_1\log_2 N} + p_1 - \ldots - p_\ell\log_2 p_\ell - \underline{p_\ell\log_2 N} + p_\ell\}$$

$\log N - (p_1\log N + p_2\log N + \cdots + p_i\log N)$
$= \log N - (p_1 + p_2 + \cdots + p_i)\log N = 0$
よって、この項は消える。

M = $\log_2$ W を計算する

青い線の部分は打ち消しあって、消える

$$\log_2\left(\frac{N!}{(p_1N)!(p_2N)!\cdots(p_\ell N)!}\right) = \log_2(N!) - \{\log_2((p_1N)!) + \ldots + \log_2((p_\ell N)!)\}$$

$$\approx (N\log_2 N - N) - \{(p_1N\log_2 p_1N - p_1N) + \ldots + (p_\ell N\log_2 p_\ell N - p_\ell N)\}$$

$$= N\{\cancel{\log_2 N} - 1 - p_1\log_2 p_1 - \cancel{p_1\log_2 N} + p_1 - \ldots - p_\ell\log_2 p_\ell - \cancel{p_\ell\log_2 N} + p_\ell\}$$

logN - ($p_1$logN + $p_2$logN + ･･･ +$p_i$logN)
= logN - (p_1 + p_2 + ･･･ + p_i) log N = 0
よって、この項は消える。

M ＝ $\log_2$ W を計算する

赤い線の部分は打ち消しあって、消える

$$\log_2\left(\frac{N!}{(p_1N)!(p_2N)!\cdots(p_\ell N)!}\right)=\log_2(N!)-\{\log_2((p_1N)!)+\ldots+\log_2((p_\ell N)!)\}$$

$$\approx(N\log_2 N-N)-\{(p_1N\log_2 p_1N-p_1N)+\ldots+(p_\ell N\log_2 p_\ell N-p_\ell N)\}$$

$$=N\{\cancel{\log_2 N}\underline{-1}-p_1\log_2 p_1-\cancel{p_1\log_2 N}+\underline{p_1}-\ldots-p_\ell\log_2 p_\ell-\cancel{p_\ell\log_2 N}+\underline{p_\ell}\}$$

-1+(p_1 ＋ p_2 ＋ ･･･ ＋ p_i)= 0
この項も消える

$M = \log_2 W$ を計算する

赤い線の部分は打ち消しあって、消える

$$\log_2\left(\frac{N!}{(p_1N)!(p_2N)!\cdots(p_\ell N)!}\right) = \log_2(N!) - \{\log_2((p_1N)!) + \ldots + \log_2((p_\ell N)!)\}$$

$$\approx (N\log_2 N - N) - \{(p_1 N \log_2 p_1 N - p_1 N) + \ldots + (p_\ell N \log_2 p_\ell N - p_\ell N)\}$$

$$= N\{\cancel{\log_2 N} - \cancel{1} - p_1 \log_2 p_1 - p_1 \cancel{\log_2 N} + \cancel{p_1} - \ldots - p_\ell \log_2 p_\ell - p_\ell \cancel{\log_2 N} + \cancel{p_\ell}\}$$

$-1+(p_1 + p_2 + \cdots + p_i) = 0$
この項も消える

M = log₂ W を計算する

緑の線の項が残る

$$\log_2\left(\frac{N!}{(p_1N)!(p_2N)!\cdots(p_\ell N)!}\right)=\log_2(N!)-\{\log_2((p_1N)!)+\ldots+\log_2((p_\ell N)!)\}$$

$$\approx(N\log_2 N-N)-\{(p_1N\log_2 p_1N-p_1N)+\ldots+(p_\ell N\log_2 p_\ell N-p_\ell N)\}$$

$$=N\{\cancel{\log_2 N}-\cancel{1}-\underline{p_1\log_2 p_1}-\cancel{p_1\log_2 N}+\cancel{p_1}-\ldots-\underline{p_\ell\log_2 p_\ell}-\cancel{p_\ell\log_2 N}+\cancel{p_\ell}\}$$

残るのは、この項だけである

M ＝ $\log_2$ W を計算する

緑の線の項が残る

$$\log_2\left(\frac{N!}{(p_1N)!(p_2N)!\cdots(p_\ell N)!}\right)=\log_2(N!)-\{\log_2((p_1N)!)+\ldots+\log_2((p_\ell N)!)\}$$

$$\approx(N\log_2 N-N)-\{(p_1N\log_2 p_1N-p_1N)+\ldots+(p_\ell N\log_2 p_\ell N-p_\ell N)\}$$

$$=N\{\cancel{\log_2 N}-\cancel{1}-\underline{p_1\log_2 p_1}-\cancel{p_1\log_2 N}+\cancel{p_1}-\ldots-\underline{p_\ell\log_2 p_\ell}-\cancel{p_\ell\log_2 N}+\cancel{p_\ell}\}$$

$$=-N\sum_{j=1}^{\ell}p_j\log_2 p_j$$

M ＝ $\log_2$ W を計算する

$-\sum_i p_i \log_2 p_i$ は、シャノンエントロピー H(X)に他ならない。

$$\log_2\left(\frac{N!}{(p_1N)!(p_2N)!\cdots(p_\ell N)!}\right) = \log_2(N!) - \{\log_2((p_1N)!) + \ldots + \log_2((p_\ell N)!)\}$$

$$\approx (N\log_2 N - N) - \{(p_1N\log_2 p_1N - p_1N) + \ldots + (p_\ell N\log_2 p_\ell N - p_\ell N)\}$$

$$= N\{\cancel{\log_2 N} - \cancel{1} - \underline{p_1\log_2 p_1} - \cancel{p_1\log_2 N} + \cancel{p_1} - \ldots - \underline{p_\ell\log_2 p_\ell} - \cancel{p_\ell\log_2 N} + \cancel{p_\ell}\}$$

$$= -N\sum_{j=1}^{\ell} p_j\log_2 p_j$$

M = log$_2$ W を計算する

$-\sum_{i} p_i \log_2 p_i$ は、シャノンエントロピー H(X)に他ならない。

$$\log_2\left(\frac{N!}{(p_1N)!(p_2N)!\cdots(p_\ell N)!}\right)=\log_2(N!)-\{\log_2((p_1N)!)+\ldots+\log_2((p_\ell N)!)\}$$

$$\approx (N\log_2 N-N)-\{(p_1N\log_2 p_1N-p_1N)+\ldots+(p_\ell N\log_2 p_\ell N-p_\ell N)\}$$

$$=N\{\cancel{\log_2 N}-\cancel{1}-\underline{p_1\log_2 p_1}-\cancel{p_1\log_2 N}+\cancel{p_1}-\ldots-\underline{p_\ell\log_2 p_\ell}-\cancel{p_\ell\log_2 N}+\cancel{p_\ell}\}$$

$$=-N\sum_{j=1}^{\ell}p_j\log_2 p_j$$

$$=NH(X)$$

N個の長さの文字列のメッセージのエントロピーを考える

言語 L には、l 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。

この時、Lの長さNの異なる典型的なメッセージのエントロピーは、確率分布 p_i のシャノンエントロピー がH(X)の時、NH(X)で与えられる。

$$\log_2 \left(\text{典型的な、異なるメッセージの数} \right) = NH(X)$$

だから

$$\left(\text{典型的な、異なるメッセージの数} \right) = 2^{NH(X)}$$

だから、N文字のメッセージは、$2^{NH(X)}$ 個しか存在しない。このことは、メッセージのレベルでは、$2^{NH(X)}$ bit を使ってエンコードするしかないということを意味する。

2 possible messages require 1 bit: **0, 1**.
4 possible messages require 2 bits: **00, 01, 10, 11**.

それでは、文字レベルでは、n文字のアルファベットから、N文字のメッセージをエンコードするのに、何 bit 必要になるか？

ボルツマンとシャノンの
エントロピーの比較

ボルツマンとシャノンのエントロピーの比較

$$\frac{N!}{(p_1N)!(p_2N)!\cdots(p_\ell N)!}$$

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
- $\{n_1,\cdots,n_l\}$:l-cellへの分布
- $\{p_1,\cdots,p_l\}$: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_iN: cell内のミクロ状態の数

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
- $\{a_1,\cdots,a_l\}$:l文字のアルファベット
- $\{p_1,\cdots,p_l\}$: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_iN: メッセージ内のa_iの数

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

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

シャノンの情報量/エントロピーと他のエントロピー概念との関係

Thermodynamic entropy: $$\Delta S = S_f - S_i = {}_R\int_i^f \frac{\delta Q_R}{T}$$

Boltzmann entropy: $$S_B(\Gamma_{D_i}) = -Nk\sum_{j=1}^{\ell} p_j \ln p_j + const.$$

典型的なN文字のメッセージを
エンコードするのに、何bit が必要か？

Gibbs entropy: $$S_G(\rho) = -k\int_\Omega \rho(x,t)\ln(\rho(x,t))\,dx$$

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

エントロピーを
グラフィカルに表現する

Christopher Olah
“Visual Information Theory”
http://colah.github.io/posts/2015-09-Visual-Information/

シャノンのエントロピー　H(X)

$$\mathrm{H(X)} = H(x_1, x_2, \ldots, x_n) = H(p_1, p_2, \ldots, p_n)$$

$$= -\sum_i p_i \log_2 p_i = \sum_i p_i \log_2 \frac{1}{p_i}$$

シャノンのエントロピー H(X)

$$\mathrm{H(X)} = H(x_1, x_2, \dots, x_n) = H(p_1, p_2, \dots, p_n)$$

$$= -\sum_i p_i \log_2 p_i = \sum_i p_i \log_2 \frac{1}{p_i}$$

$p(x_i) = p_i,\ L(x_i) = \log_2 \frac{1}{p_i}$ と置くと、

$H(X) = \sum_i p(x_i) L(x_i)$ と書ける

シャノンのエントロピー　H(X)

$$\mathrm{H(X)} = H(x_1, x_2, \ldots, x_n) = H(p_1, p_2, \ldots, p_n)$$

$$= -\sum_i p_i \log_2 p_i = \sum_i \boxed{p_i} \boxed{\log_2 \frac{1}{p_i}}$$

$p(x_i) = p_i,\ L(x_i) = \log_2 \frac{1}{p_i}$ と置くと、

$$H(X) = \sum_i \boxed{p(x_i)} \boxed{L(x_i)}$$ と書ける

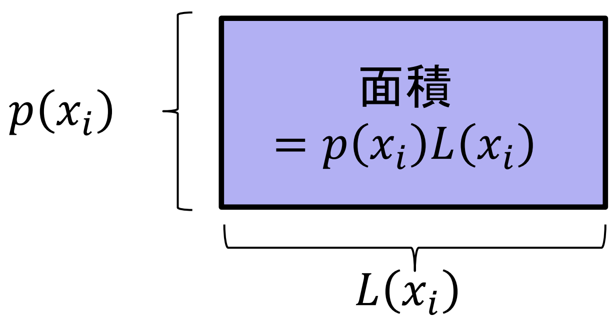
シャノンのエントロピー　H(X)
を長方形の面積の和として考える

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

$p(x_i) = p_i,\ L(x_i) = \log_2 \frac{1}{p_i}$ と置くと、

$H(X) = \sum_i p(x_i) L(x_i)$ と書ける

ここで、高さ $\mathrm{p}(x_i)$ 、幅 $L(x_i)$ の長方形の面積の和としてエントロピーを考える。

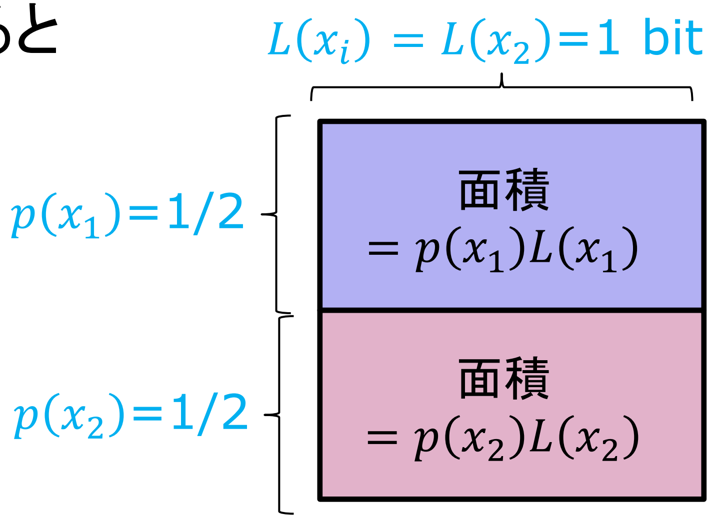


シャノンのエントロピー H(X) を長方形の面積の和として考える

$$\mathrm{H(X)} = H(x_1, x_2) = H\left(\frac{1}{2}, \frac{1}{2}\right)$$ としよう。

$$p(x_1) = p(x_2) = \frac{1}{2}, \qquad L(x_1) = L(x_2) = \log_2 2 = 1$$

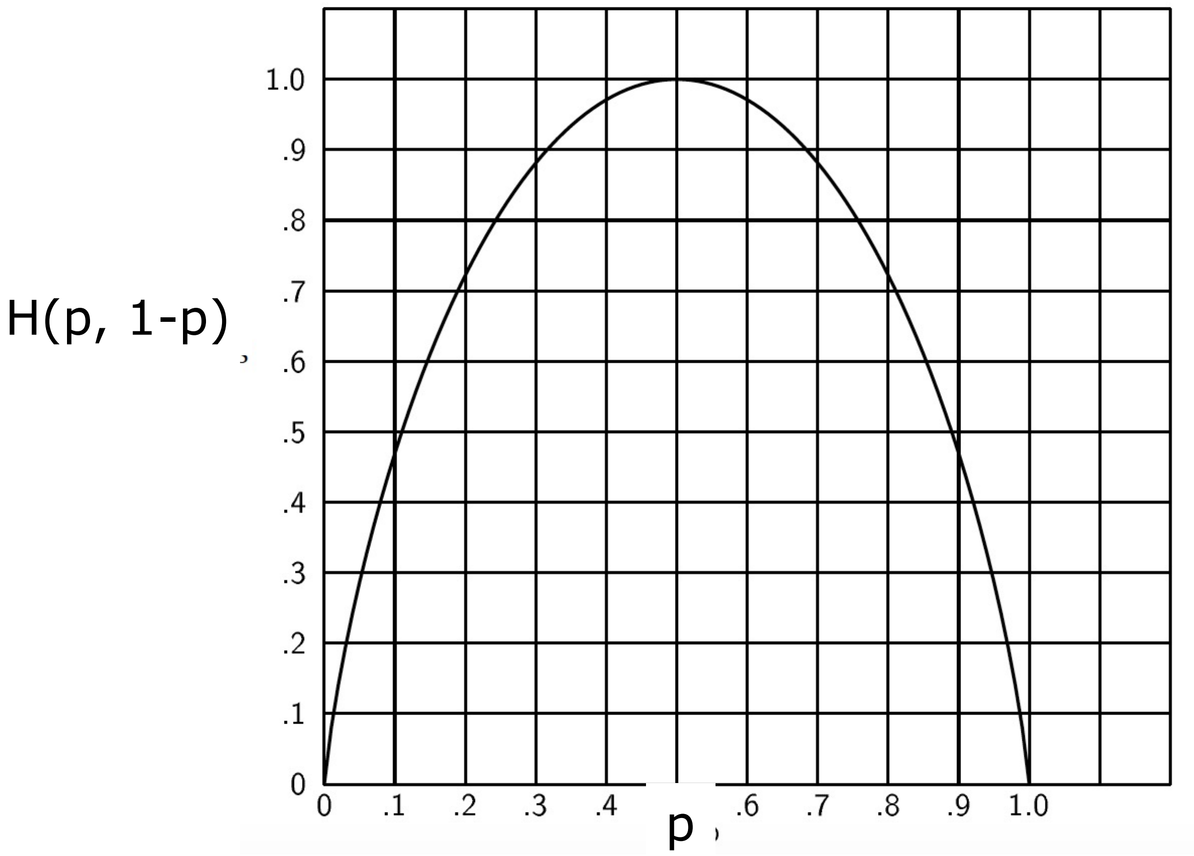
ここで、高さ $p(x_i)$ 、幅 $L(x_i)$ の長方形の面積の和としてエントロピーを考えると



これは、エントロピー $H\left(\frac{1}{2}, \frac{1}{2}\right)$ の値が1であることを表している。

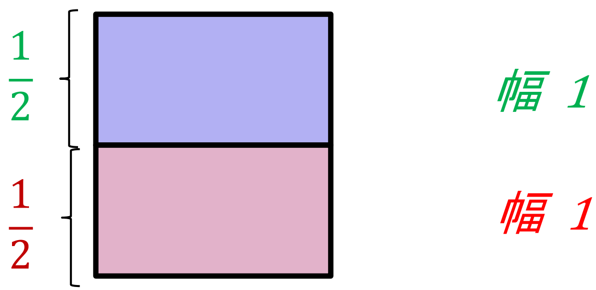
$H(p.q)=H(p.1\text{-}p)$ のグラフ

$$H(X) = H(x_1, x_2) = H(p, q)$$
$$= H(p, 1 - p)$$

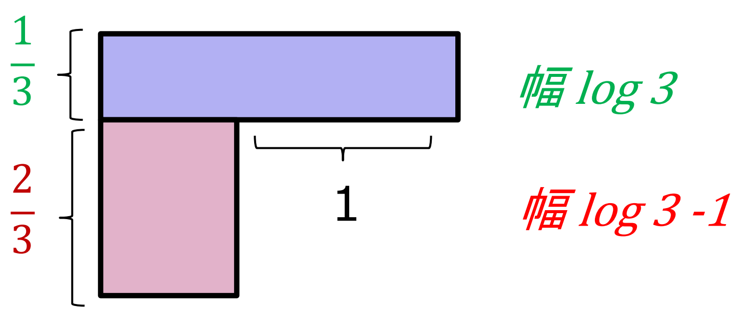


$H(p.q)=H(p.1\text{-}p)$の図形

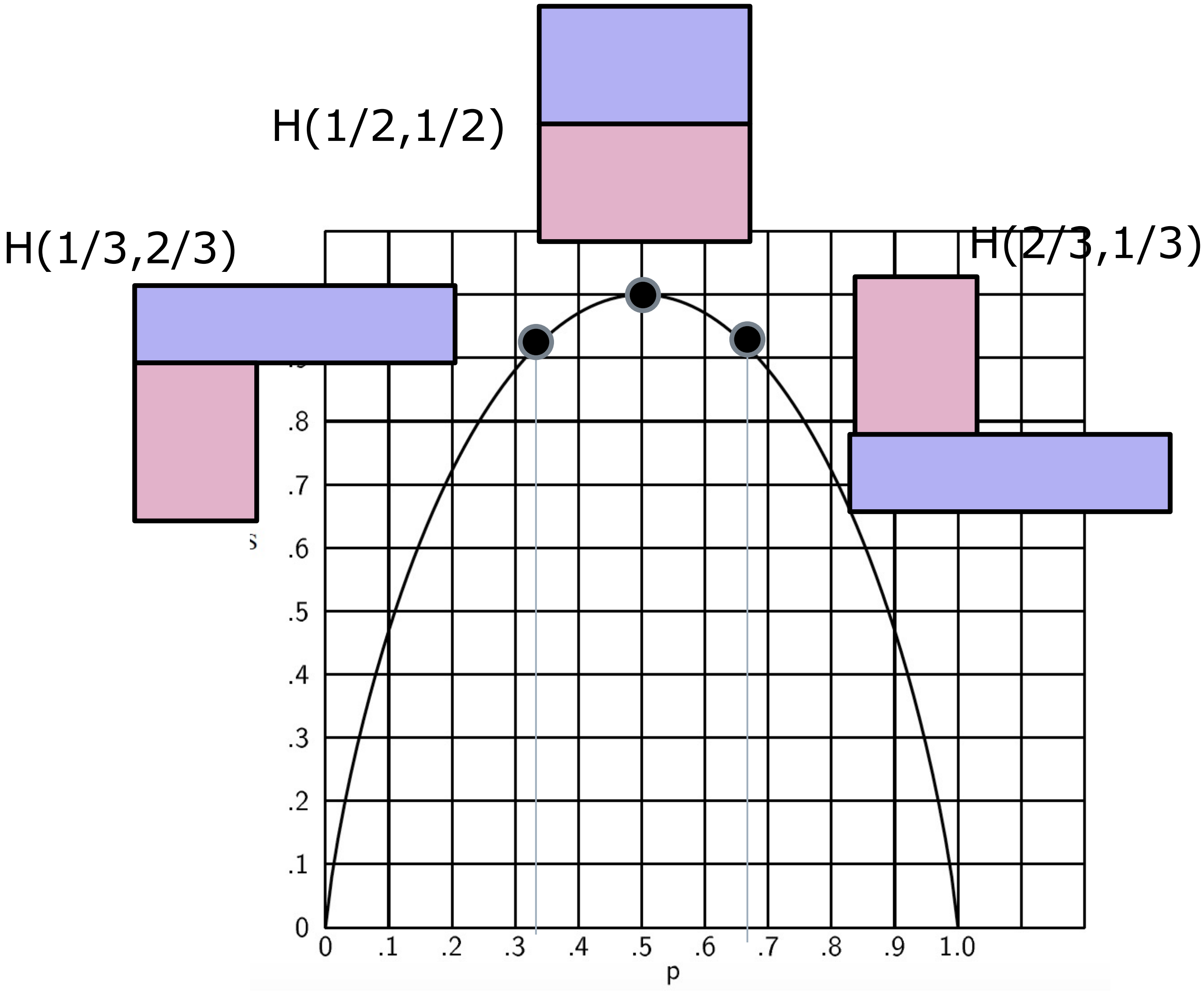
$$H(X)=H(x_1,x_2)=H\left(\frac{1}{2},\frac{1}{2}\right)=\frac{1}{2}\log_2 2+\frac{1}{2}\log_2 2=\log_2 2=1\ \ bit$$



$$H(X)=H(x_1,x_2)=H\left(\frac{1}{3},\frac{2}{3}\right)=\frac{1}{3}\log 3+\frac{2}{3}\log\frac{3}{2}$$



H(1/2. 1/2), H(1/3, 2/3), H(2/3, 1/3) の値



n文字のアルファベットをエンコードするのに何bit 必要か？

n = アルファベットの数

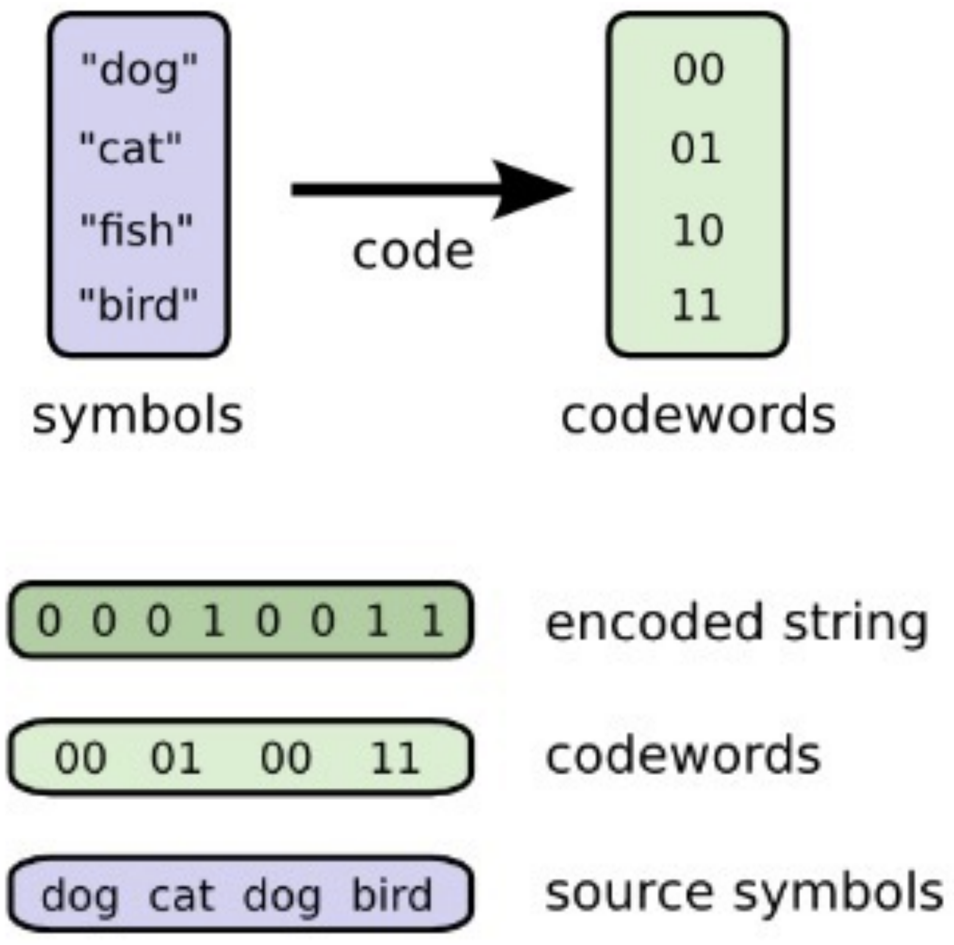
x = アルファベット一文字のbit数

2 文字 1 bit: 0, 1
4 文字 2 bits: 00, 01, 10, 11
8 文字 3 bits: 000, 001, 010, 011, 100, 101, 110, 111

だから: $n = 2^x$, $x = \log_2 n$.

1文字 $\log_2 n$ bits ということは、N文字長のメッセージについては、N $\log_2 n$ bits が必要ということ。。

固定長のコード　4 words　2bit



N文字のメッセージをエンコードするのに、何bit が必要か？

$X = \{A, B, C, D\}$ とする ($n = 4$)
この時、1文字につき $\log_2 4 = 2$ bits が必要である
例えば、: A = 00, B = 01, C = 10, D = 11.

N文字のメッセージをエンコードするのに、2N bits が必要

N文字のメッセージをエンコードするのに、何bit が必要か？

X = {A, B, C, D} とする (n = 4)
この時、1文字につき $\log_2 4 = 2$ bits が必要である
例えば、: A = 00, B = 01, C = 10, D = 11.

N文字のメッセージをエンコードするのに、2N bits が必要

ただ、それは各文字が同じ確率で出現する時である。
全ての文字が等しい確率で出現するとすると

$$NH(X) = -N\left(\tfrac{1}{4}\log_2\tfrac{1}{4} + \tfrac{1}{4}\log_2\tfrac{1}{4} + \tfrac{1}{4}\log_2\tfrac{1}{4} + \tfrac{1}{4}\log_2\tfrac{1}{4}\right) = 2N.$$

典型的なN文字のメッセージを エンコードするのに、何bit が必要か？

それぞれの文字が、典型的なN文字のメッセージに出現する確率が、次のようだとしよう。

$p_A = 1/2,\ p_B = 1/4,\ p_C = p_D = 1/8$

この時、こうした典型的なN文字のメッセージの数は、

$$NH(X) = -N\left(\frac{1}{2}\log_2\frac{1}{2} + \frac{1}{4}\log_2\frac{1}{4} + \frac{1}{8}\log_2\frac{1}{8} + \frac{1}{8}\log_2\frac{1}{8}\right) = 1.75N$$

典型的なN文字のメッセージを
エンコードするのに、何bit が必要か？

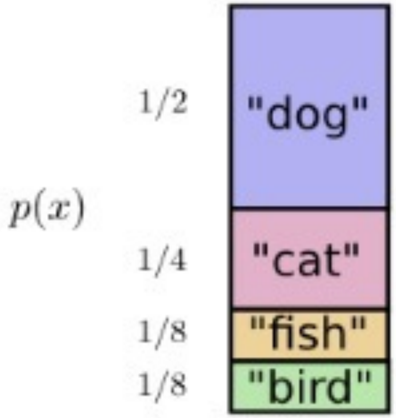
それぞれの文字が、典型的なN文字のメッセージに出現する確率が、次のようだとしよう。
$p_A = 1/2$, $p_B = 1/4$, $p_C = p_D = 1/8$

この時、こうした典型的なN文字のメッセージの数は、

$$NH(X) = -N\left(\frac{1}{2}\log_2\frac{1}{2} + \frac{1}{4}\log_2\frac{1}{4} + \frac{1}{8}\log_2\frac{1}{8} + \frac{1}{8}\log_2\frac{1}{8}\right) = 1.75N$$

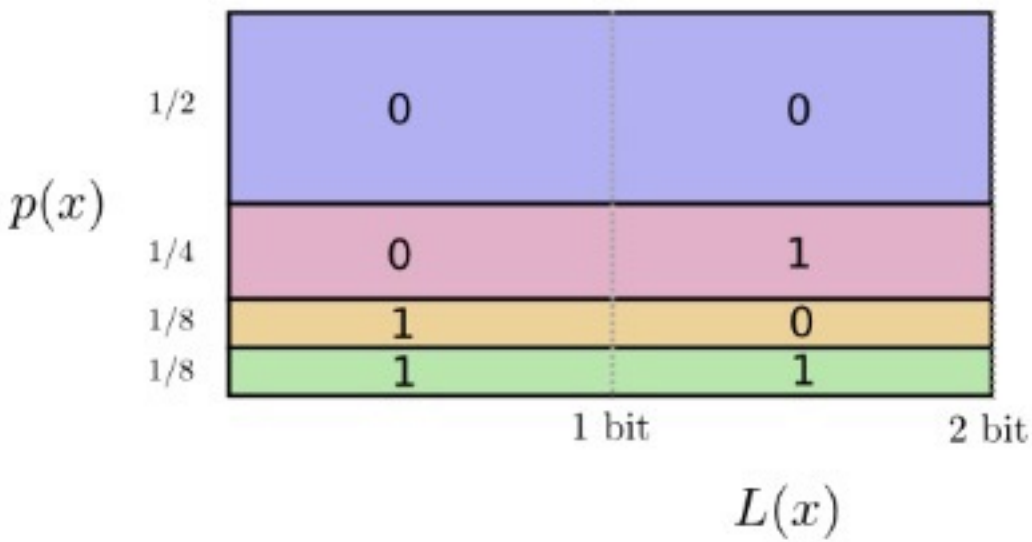
この時、N文字のメッセージをエンコードするのに、2N bits ではなく、1.75N bitsでいいことになる。

語の出現頻度に違いがある場合



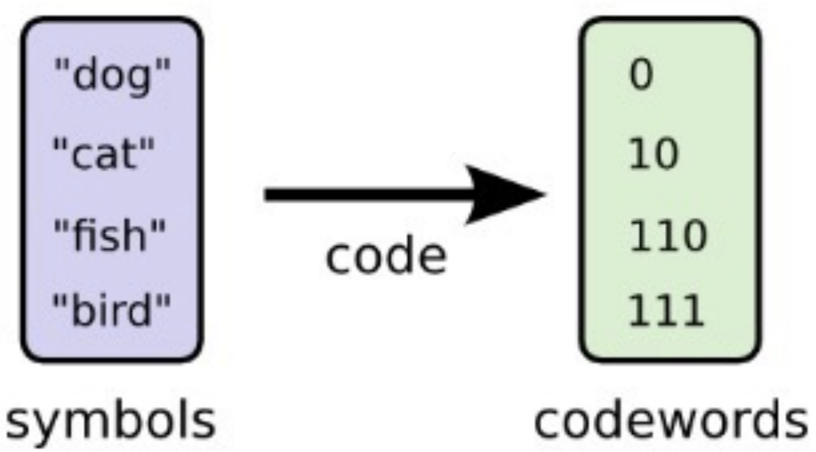
犬好きな人の
語の出現頻度

固定長コード　2bit

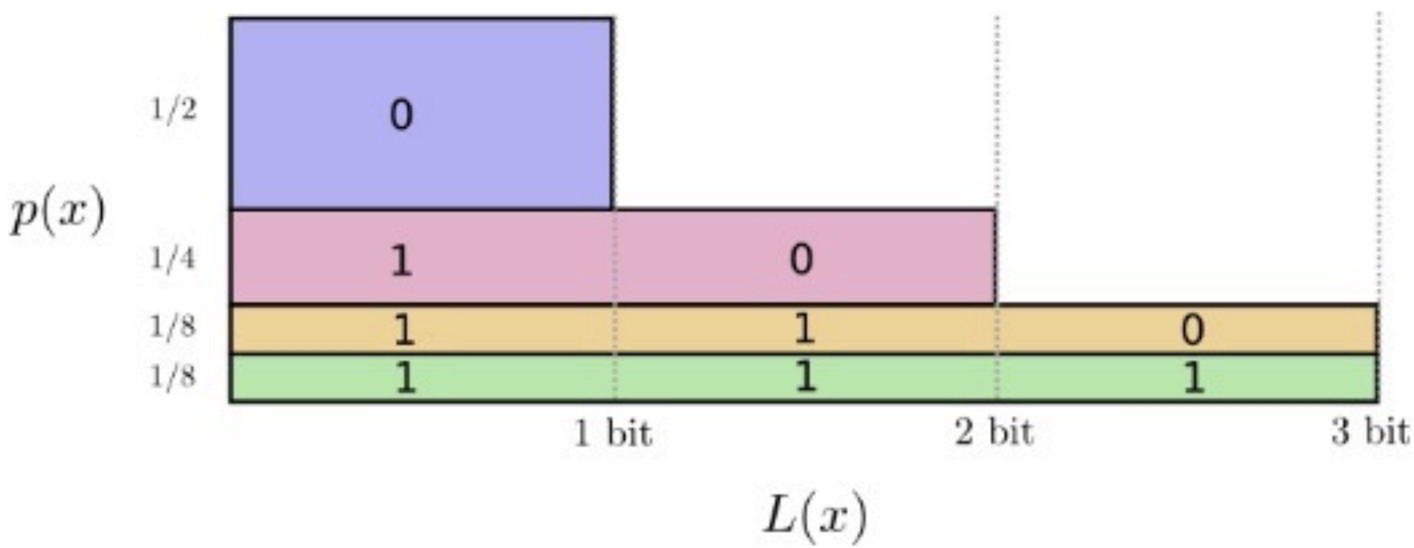


ハフマン・コード

語の出現頻度に違いがある場合

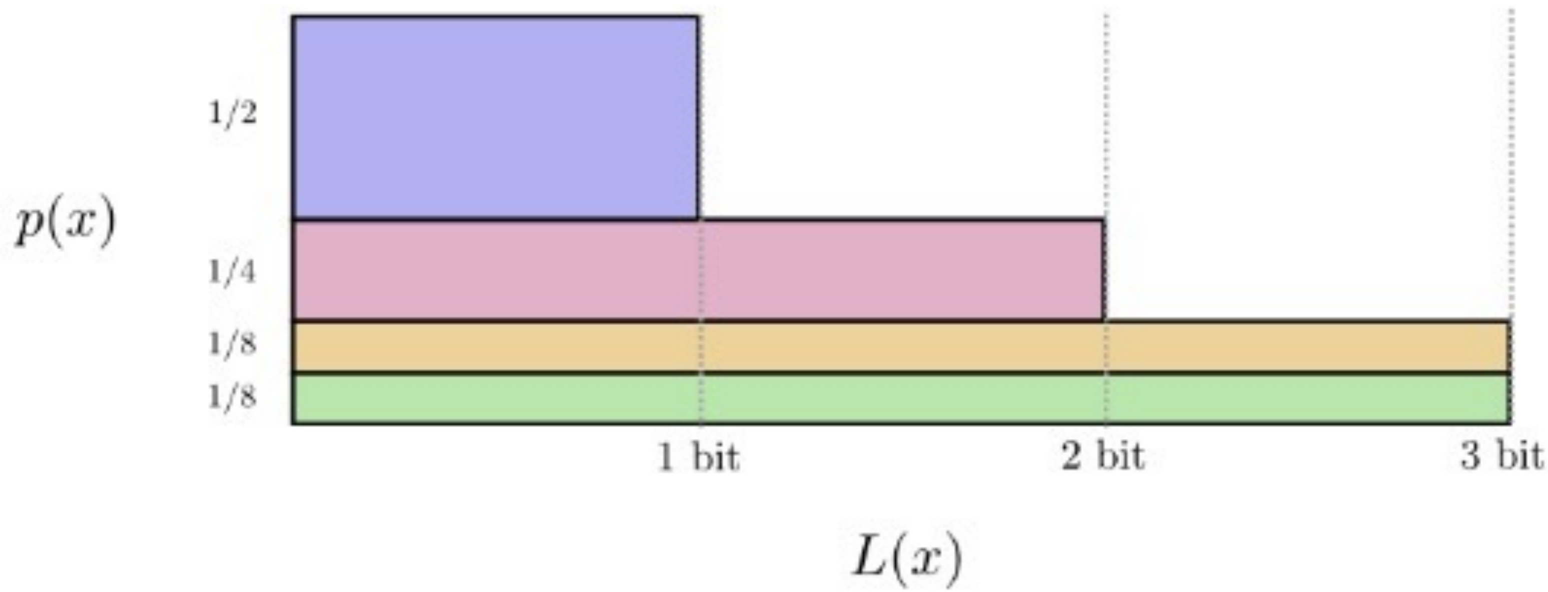


可変長コード　1.75 bit



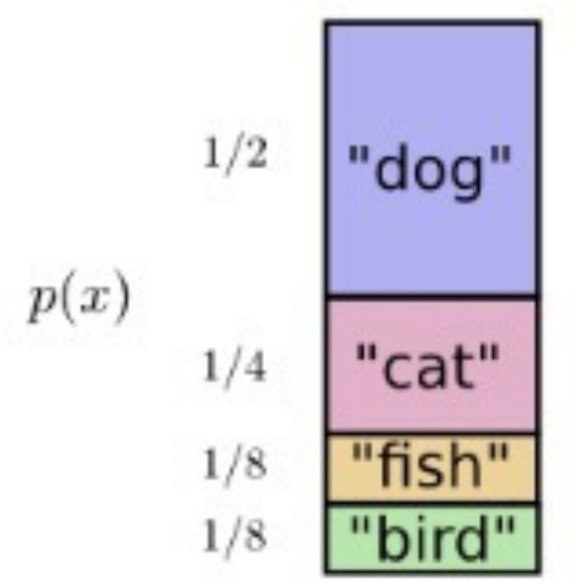
エントロピー

H(X) ＝ 最適なコードの平均長
＝ 下の図の面積

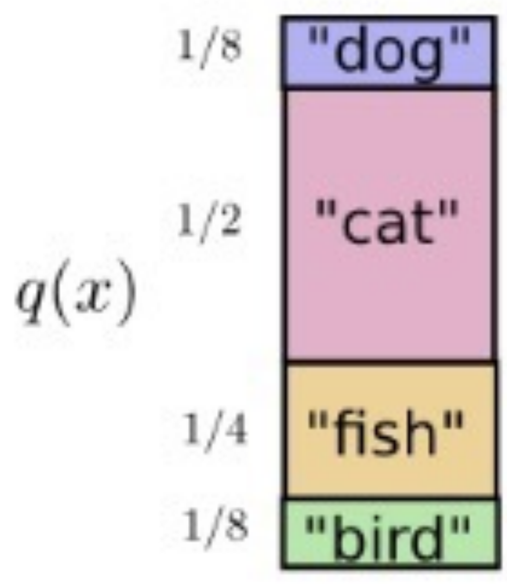


クロスエントロピーを
グラフィカルに表現する

語の出現頻度の異なる二つのグループ

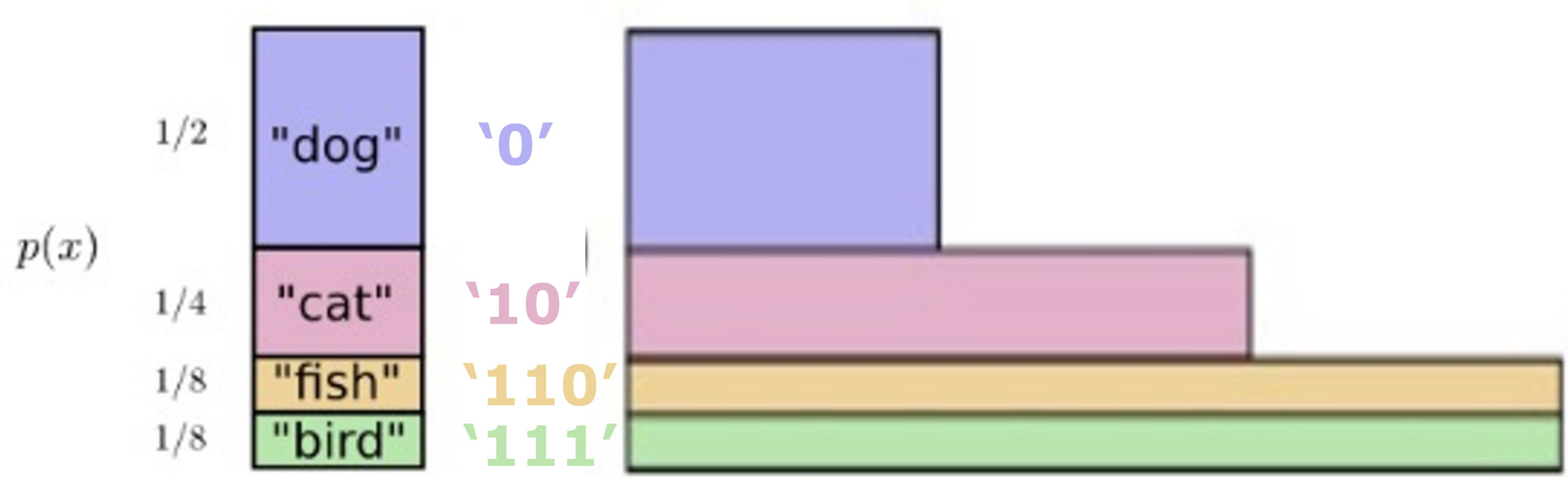


犬好きの人の
語の出現頻度
Bob



猫好きの人の
語の出現頻度
Alice

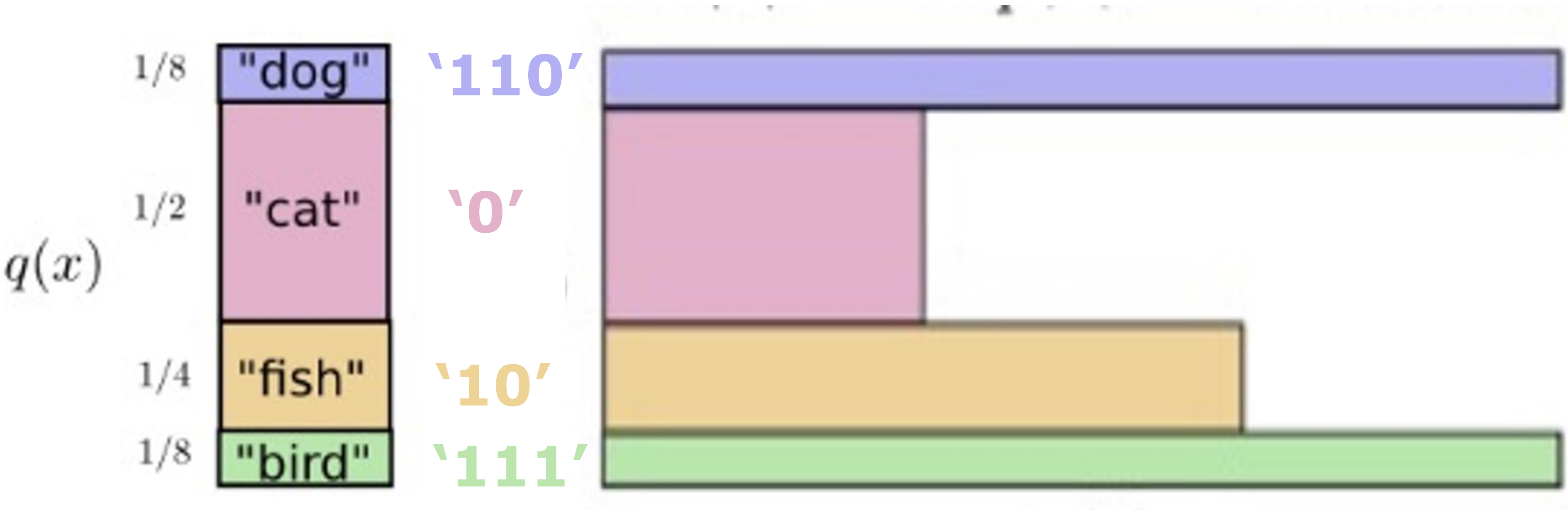
犬好きのBobのグループ内の
コミュニケーションに最適のコード



犬好きの人の
語の出現頻度
Bob

$$H(p) = 1/2 \times 1 + 1/4 \times 2 + 1/8 \times 3 + 1/8 \times 3$$
$$= 1 + 3/4 = 1.75$$

猫好きのAliceのグループ内のコミュニケーションに最適のコード

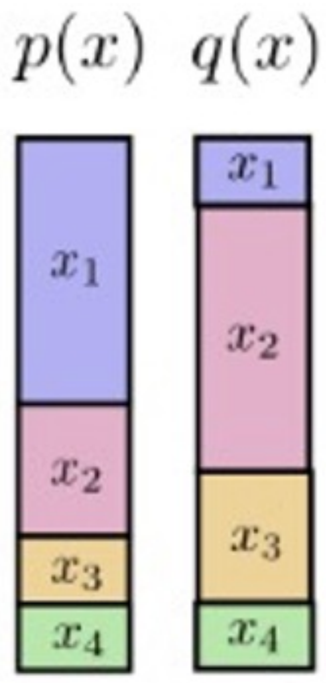


猫好きの人の語の出現頻度
Alice

H(q) = 1/8x3 + 1/2x1 + 1/4x2 + 1/8x3
= 1 + 3/4 = 1.75

Cross-Entropy $H_p(q)$

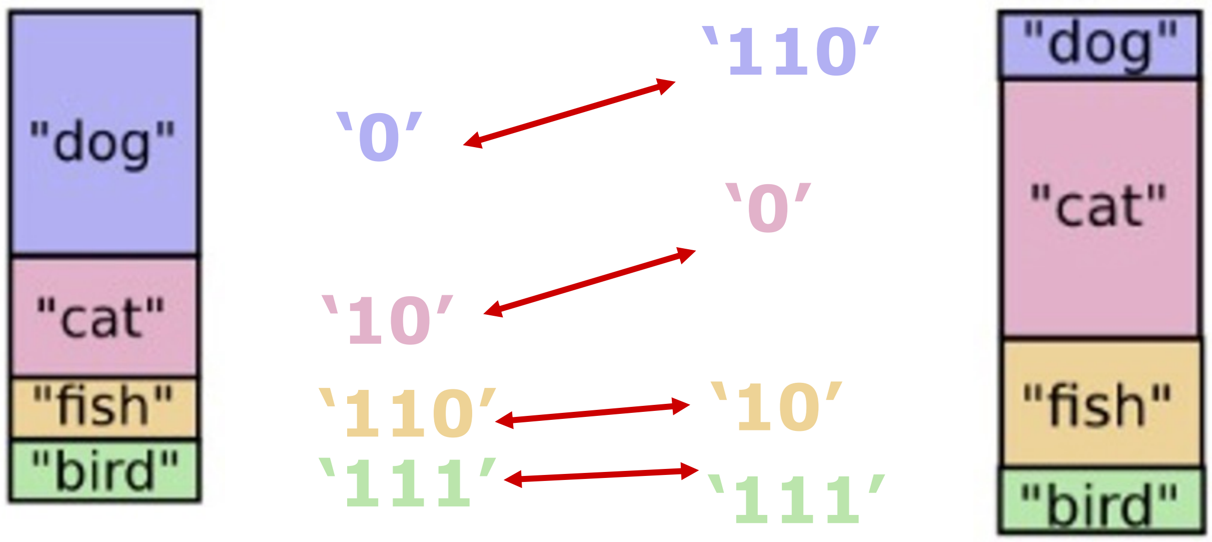
$$H_p(q) = \sum_x q(x) \log_2\left(\frac{1}{p(x)}\right)$$



Cross-Entropy: $H_p(q)$

p(x)のコードを使って
q(x)からメッセージを送る時の
メッセージの長さの平均

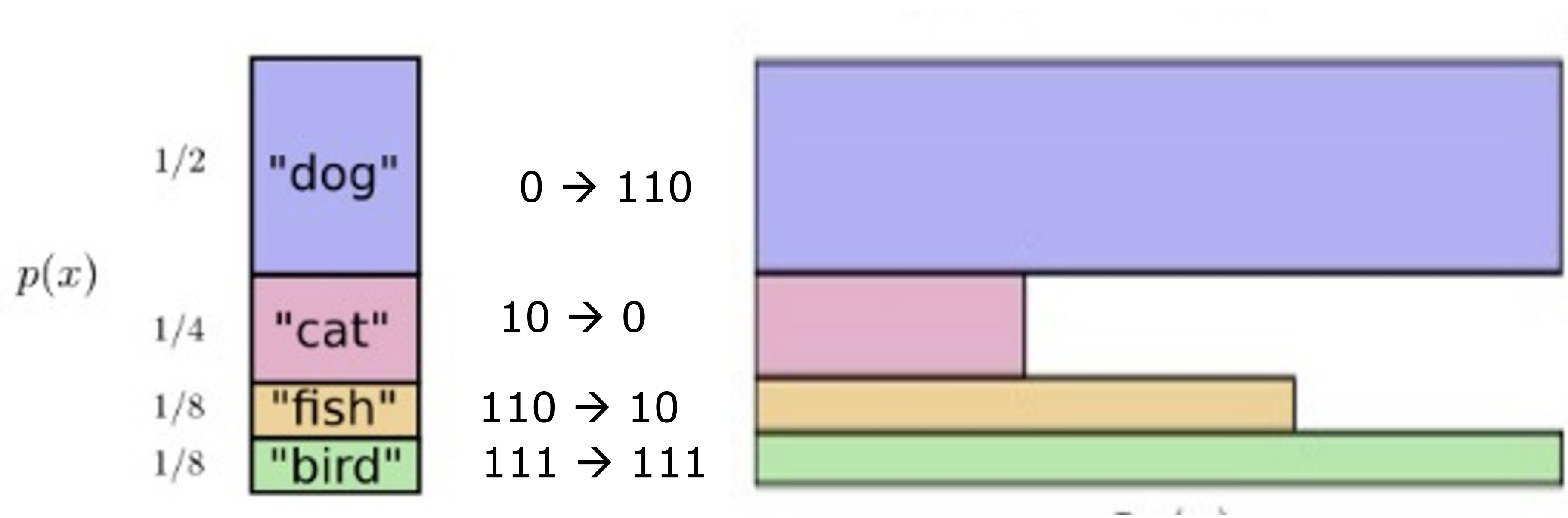
異なる二つのグループのコミュニケーション
犬語と猫語



犬語 p

猫語 q

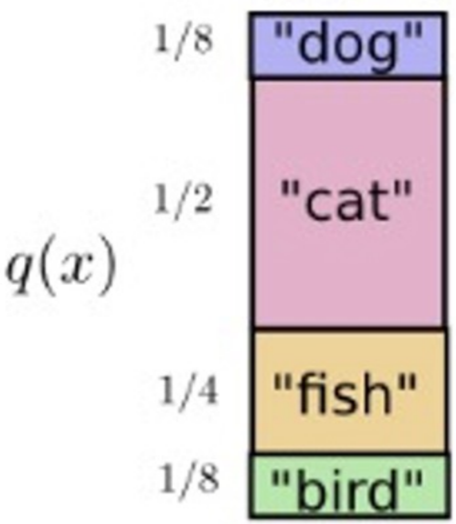
Bob(犬)のグループpからAlice（猫）のグループqへ
猫語qを使ってメッセージを送る　$H_q(p)$



犬好きの人の
語の出現頻度
Bob

$H_q(p) = 1/2 \times 3 + 1/4 \times 1 + 1/8 \times 2 + 1/8 \times 3$
$= 2.375$

Alice(猫)のグループqからBob（犬）のグループpへ
犬語pを使ってメッセージを送る　$H_p(q)$

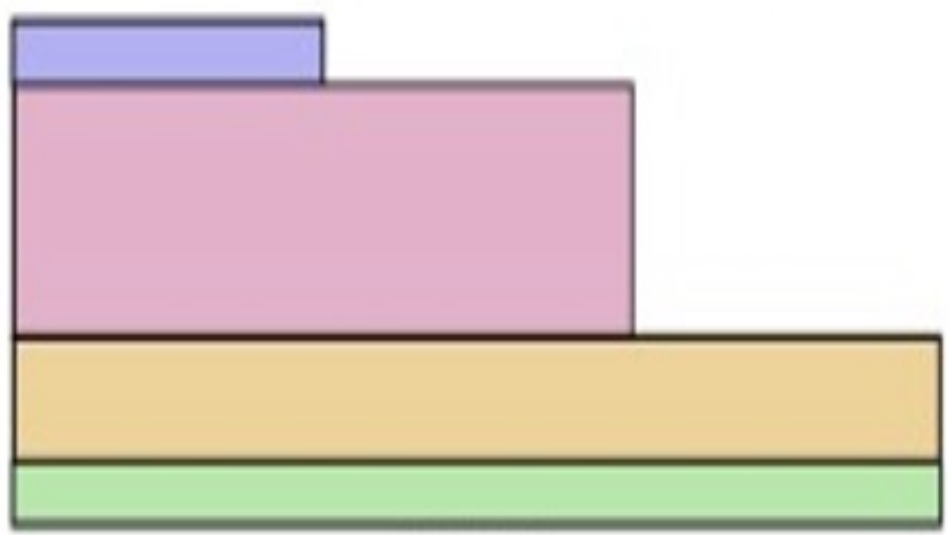


110 → 0

0 → 10

10 → 110

111 → 111



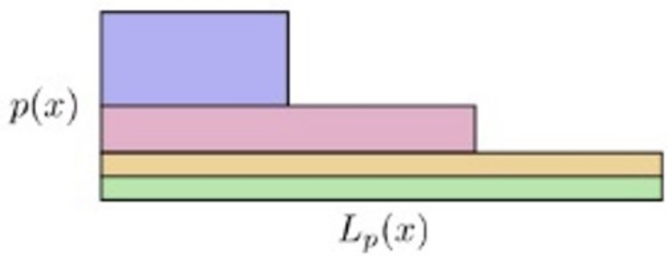
猫好きの人の
語の出現頻度
Alice

$H_q(p)$ = 1/8x1 + 1/2x2 + 1/8x3 + 1/8x3
= 2.25

Cross Entropyのグラフィカルな表現の例

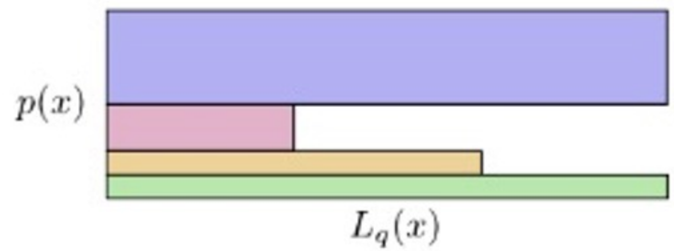
Bob pが自分のコードpを使う

$$H(p) = H_p(p) = 1.75 \text{ bits}$$



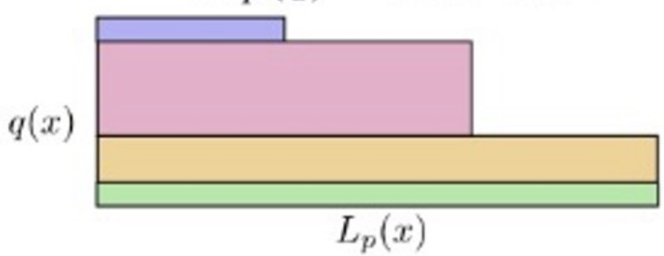
Bob pがAliceのコードqを使う

$$H_q(p) = 2.375 \text{ bits}$$



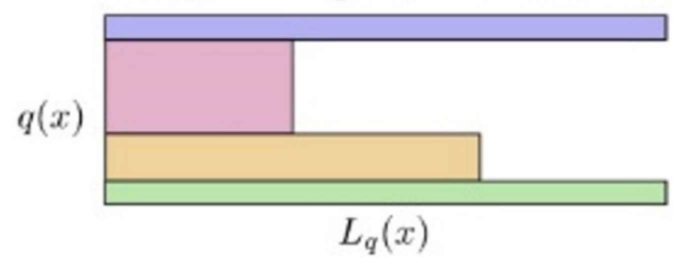
Alice qがBobのコードpを使う

$$H_p(q) = 2.25 \text{ bits}$$

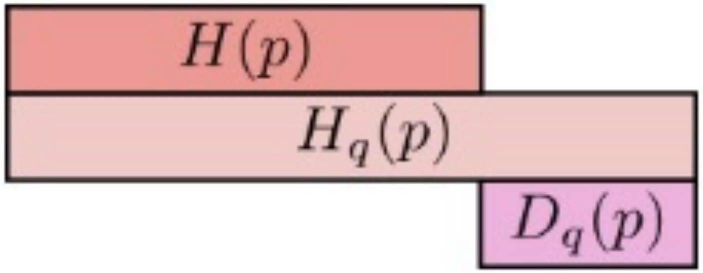


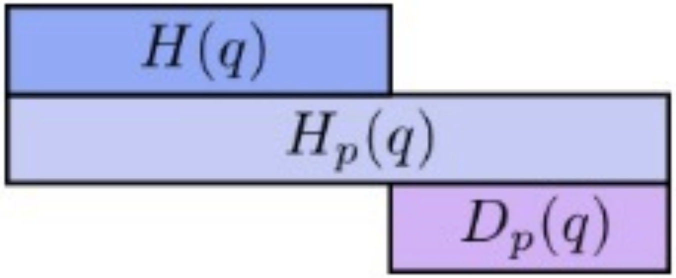
Alice qが自分のコードqを使う

$$H(q) = H_q(q) = 1.75 \text{ bits}$$



相対エントロピー $D_q(p)$ (Kullback–Leibler divergence)





$$D_q(p) = H_q(p) - H(p)$$

$$D_p(q) = H_p(q) - H(q)$$