

情報理論04

津山工業高等専門学校情報工学科 講師
電気通信大学先進理工学科 協力研究員
藤田一寿

情報量とエントロピー

情報の定量化

- ・ 良い情報とはなにか
- ・ 珍しい情報が良い情報でしょう
- ・ 珍しいとは確率が低いということに相当する
- ・ 確率を使って情報を定量化してみよう

確率を使った情報の量

確率が低いほうが情報が多い



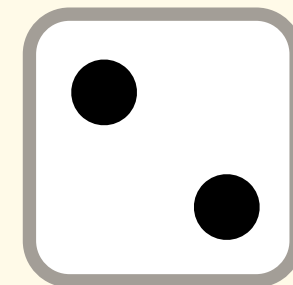
確率の逆数が情報の量になりそうだ

確率の逆数を情報量として使ってみる

サイコロの目が2であった時の情報量

$$P(X = 2) = \frac{1}{6}$$

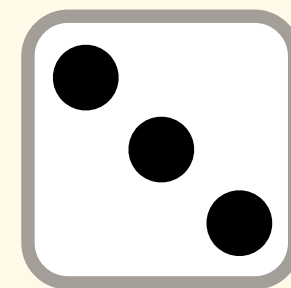
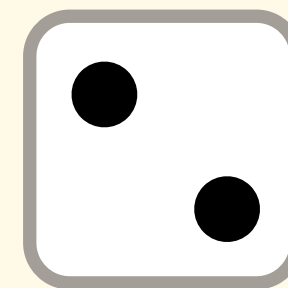
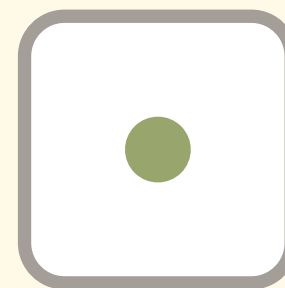
$$I(P(X = 2)) = 6$$



サイコロの目が3以下であった時の情報量

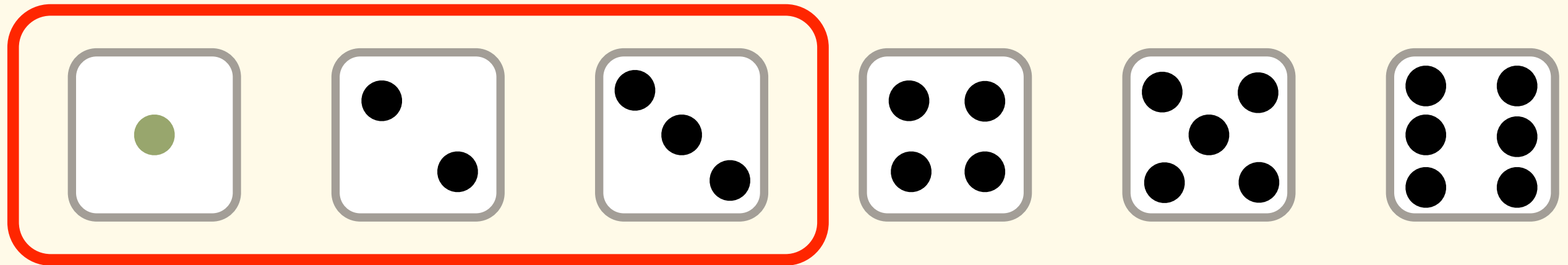
$$P(X \leq 3) = \frac{3}{6} = \frac{1}{2}$$

$$I(P(X \leq 3)) = 2$$

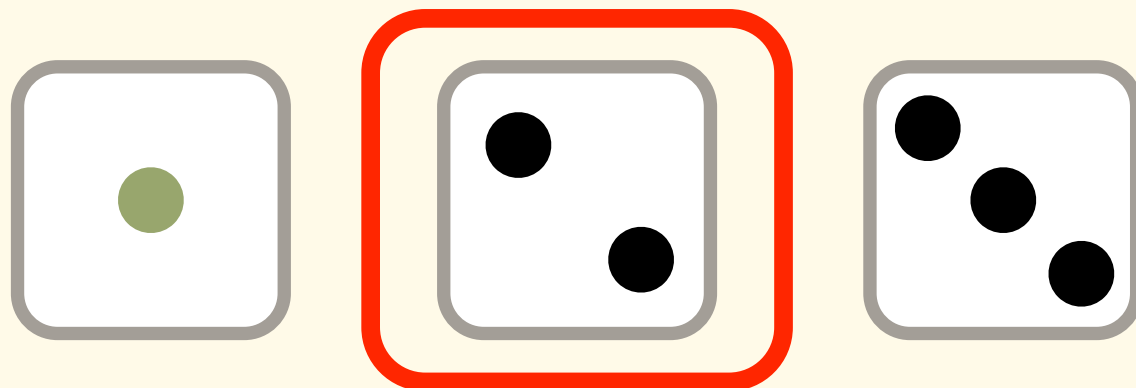


よさそうな気がする

目の選び方を変える

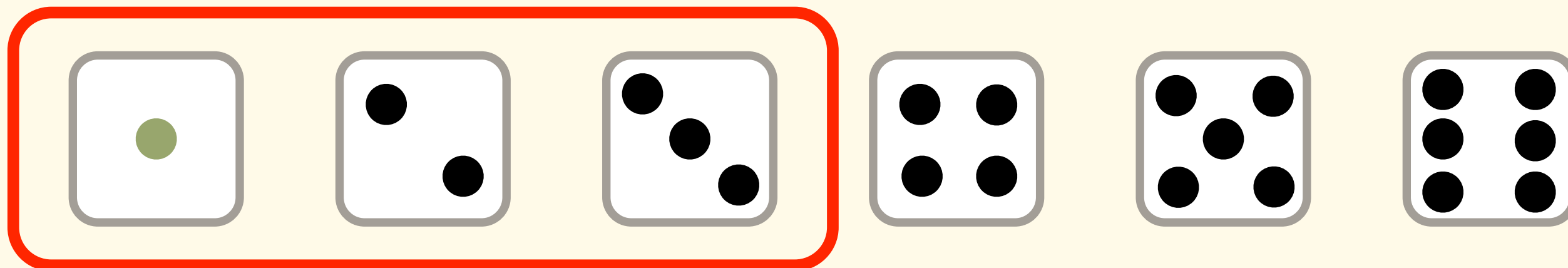


まずサイコロの目が3以下だった場合を考え、



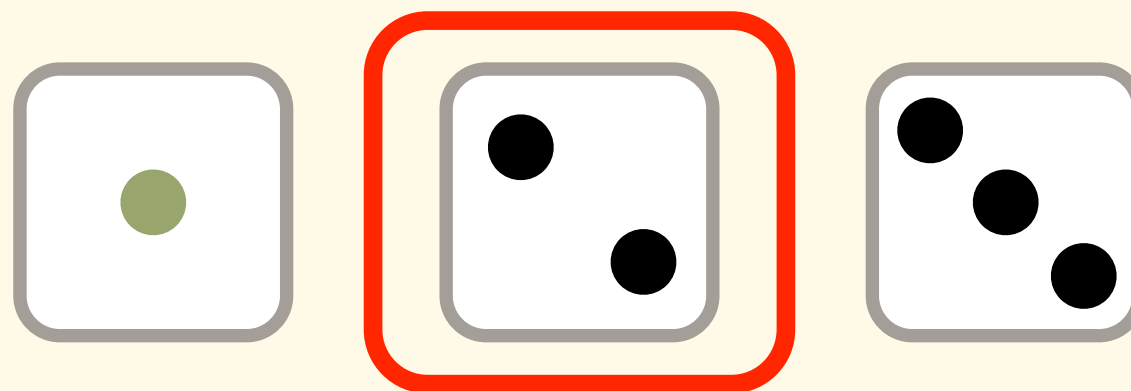
その後2をであった場合を考える

この場合も2を選んだことには変わらない。すなわち、最終的に得られる情報の量は同じでなければならな



サイコロの目が3以下だった時に得られるの情報量

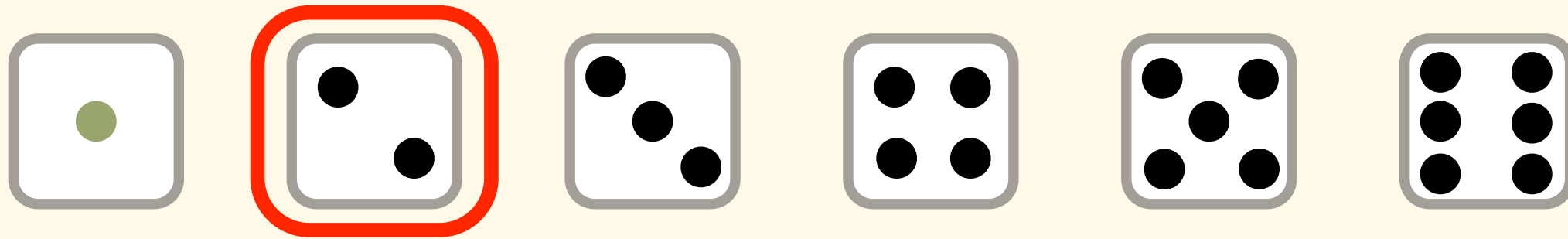
$$I\left(\frac{1}{2}\right) = 2$$



さらに3以下の目の中で2の目であった時に得られる情報量

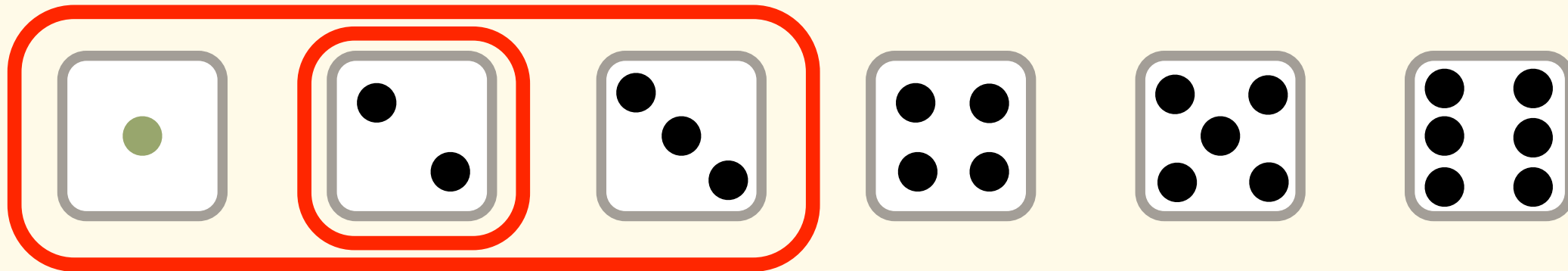
$$I\left(\frac{1}{3}\right) = 3$$

サイコロの目が2だった時の情報量



$$I\left(\frac{1}{6}\right) = 6$$

サイコロの目が3以下の目だったとき、更にその目が2であった時の情報量



$$I\left(\frac{1}{2}\right) + I\left(\frac{1}{3}\right) = 2 + 3 = 5$$

情報量が異なっている!!

最終的に2の目が出たという情報は同じであるにもかかわらず、情報量が異なっている。確率の逆数は情報量として使えない。

$$I\left(\frac{1}{6}\right) = I\left(\frac{1}{2} \times \frac{1}{3}\right) = I\left(\frac{1}{2}\right) + I\left(\frac{1}{3}\right) \quad \text{情報の加法性}$$

が成り立たなければならない。

加法性が成り立つとすると

$$f(xy) = f(x) + f(y)$$

が成り立つ。上記の式が成り立つとき

$$\begin{aligned}\frac{f(x + \varepsilon x) - f(x)}{\varepsilon x} &= \frac{f(1 + \varepsilon)}{\varepsilon x} \\ &= \frac{1}{x} \frac{f(1 + \varepsilon)}{\varepsilon}\end{aligned}$$

$\varepsilon \rightarrow 0$ の極限を取ると

$$\lim_{\varepsilon \rightarrow 0} \frac{f(x + \varepsilon x) - f(x)}{\varepsilon x} = \frac{c}{x}$$

ただし c は次の式で表す。

$$c = \lim_{\varepsilon \rightarrow 0} \frac{f(1 + \varepsilon)}{\varepsilon}$$

先ほどの $\varepsilon \rightarrow 0$ の極限は、導関数そのものなので

$$f'(x) = \frac{c}{x}$$

上記の微分方程式をとくと

$$f(x) = c \log x + d$$

確率が1の時、得られる情報は0とすると

$$f(1) = d = 0$$

$$f(x) = c \log x$$

x が確率の場合、 p は1以下で、情報量は p が減少すればするほど増えなければならないので、 c は負となる。

ビット、ニット、ディット

$$c = -1 / \log_2 e$$

logの底が2

ビット

$$c = -1$$

logの底がe

ニット、ナット

$$c = -1 / \log_{10} e$$

logの底が10

ディット

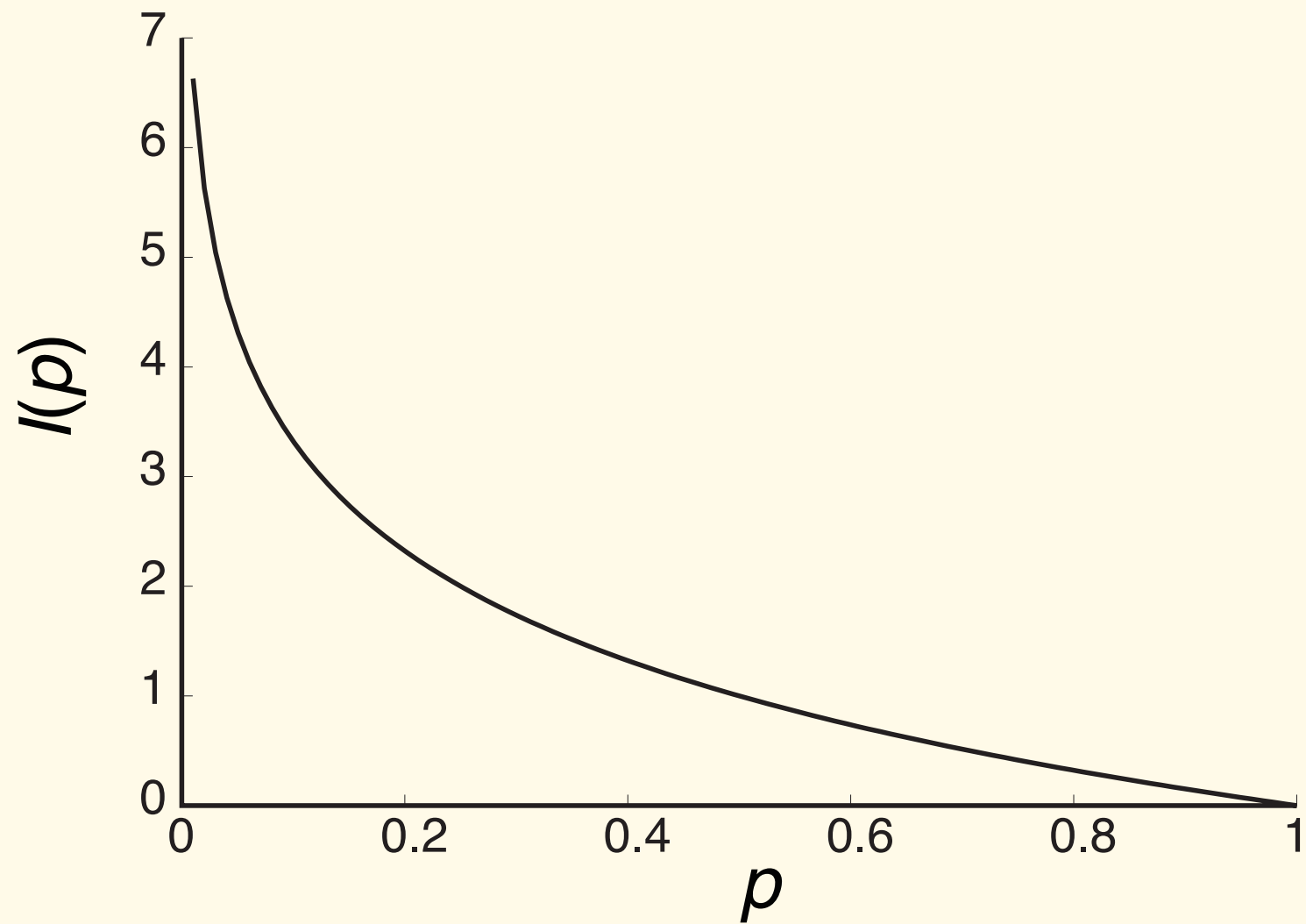
$$1\text{bit} = \log 2\text{nit} = \log_{10} 2\text{dit}$$

情報量

確率 p の事象が実際に起こったことを知らせる情報に含まれる情報量を

$$I(p) = -\log_2 p \text{ ビット}$$

と定義する。



演習

- ・ ジョーカーを除くトランプ52枚カードを引くとき
 - ・ スペードのAであった時に得られる情報量を求めよ。
 - ・ スペードであることのみ知った時に得られる情報量を求めよ。
 - ・ Aであることのみ知った時に得られる情報量を求めよ。
- ・ 台風が来るのは1年に15日、満潮は1日に2時間とする。このとき、台風が来てかつ満潮という危険度の情報量は何ビットか。
(それぞれの事象の確率分布は一様分布しているものとする。)
- ・ 2つのサイコロを降った時、その目の和が7であったとする。しかし、後日その時のサイコロの目がいくつであったか忘れてしまった。この場合、失われた情報量は何ビットか。

情報量の期待値

$$I(p) = -\log_2 p$$

- ・ 確率 p の事象が**起こった**時に得られる情報量
- ・ 将来得られる情報量ではない
- ・ 期待値を使えば良いのではないか

- ・ A_1, A_2, \dots, A_n の n 個の事象がある。
- ・ それぞれの事象が起こる確率は p_1, p_2, \dots, p_n である。
- ・ 得られる情報量の期待値は

$$\begin{aligned} H &= \sum_{i=1}^n (-\log_2 p_i) p_i \\ &= - \sum_{i=1}^n p_i \log_2 p_i \end{aligned}$$

情報量の期待値が高いということは、どの事象が起こるか予想がつかないので、将来得られる情報量は多いということ。言い換えれば不確実度が高い。

情報量の期待値が低いということは、どの事象が起こるかわかりきっているので、将来得られる情報量は少ないということ。言い換えれば、不確実度は低い。

エントロピー

n個の事象がそれぞれ確率 p_1, p_2, \dots, p_n で生じるとき、どの事象がどれだけ発生したかの不確実度を

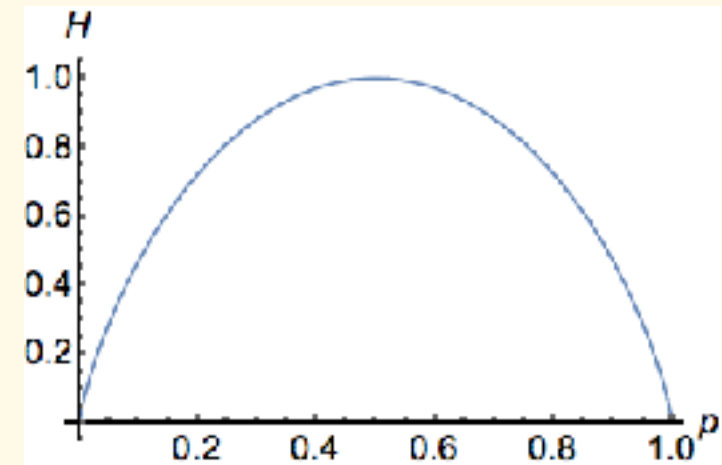
$$H(p_1, p_2, \dots, p_n) = - \sum_{i=1}^n p_i \log_2 p_i$$

とし、エントロピーと呼ぶ。

例 コイントス

- ・ 事象が2つの場合それぞれの事象が起きる確率は、 p と $q = (1 - p)$ である。
- ・ コイントスの場合、表が出る確率を p , 裏が出る確率を q と考えられる。
- ・ コイントスのエントロピーは

$$\begin{aligned} H &= -p \log_2 p - (1 - p) \log_2 (1 - p) \\ &= -p \log_2 (p / (1 - p)) - \log_2 (1 - p) \end{aligned}$$



表もしくは裏が出やすいコインはエントロピーが低いと言える。

問題

先ほどのコインスのエントロピーが最大値とする p の値を求めよ

エントロピーは非負

エントロピー H は

$$H \geq 0$$

であり、 $H=0$ が成り立つときは、 p_i のうち一つが1残りは0の場合である。

ただし、 $0 \log_2 0 = 0$ とする。

エントロピーの最大値

すべての事象が等確率に起こるとき、エントロピーは最大値となる。

$$p_i = \frac{1}{n}$$

$$\begin{aligned} H_{\max} &= - \sum_{i=1}^n \frac{1}{n} (-\log_2 n) \\ &= n \times \frac{1}{n} \log_2 n \\ &= \log_2 n \end{aligned}$$

最大値をどう求めるのか

$\sum_i p_i = 1$ の条件のもと、Hの最大化を行う。

ラグランジュの未定乗数法を使う。未定乗数 λ を導入すると

$$\frac{\partial}{\partial p_i} \left(- \sum_i p_i \log p_i - \lambda \sum_i p_i \right) = 0$$

$$-p_i \times \frac{1}{p_i} - \log p_i - \lambda = 0$$

$$\log p_i = -1 - \lambda$$

$$p_i = \exp(-1 - \lambda) \quad p_i \text{は定数と分かる}$$

ここでは、情報量にnitを用いている。

ここで $p_i = c$ とおくと(先ほどの式から定数だと分かっている)、

$$\sum_i c = 1$$

$$nc = 1$$

$$c = 1/n$$

よって、 H が最大値となる時の p_i は

$$p_i = \frac{1}{n}$$

演習

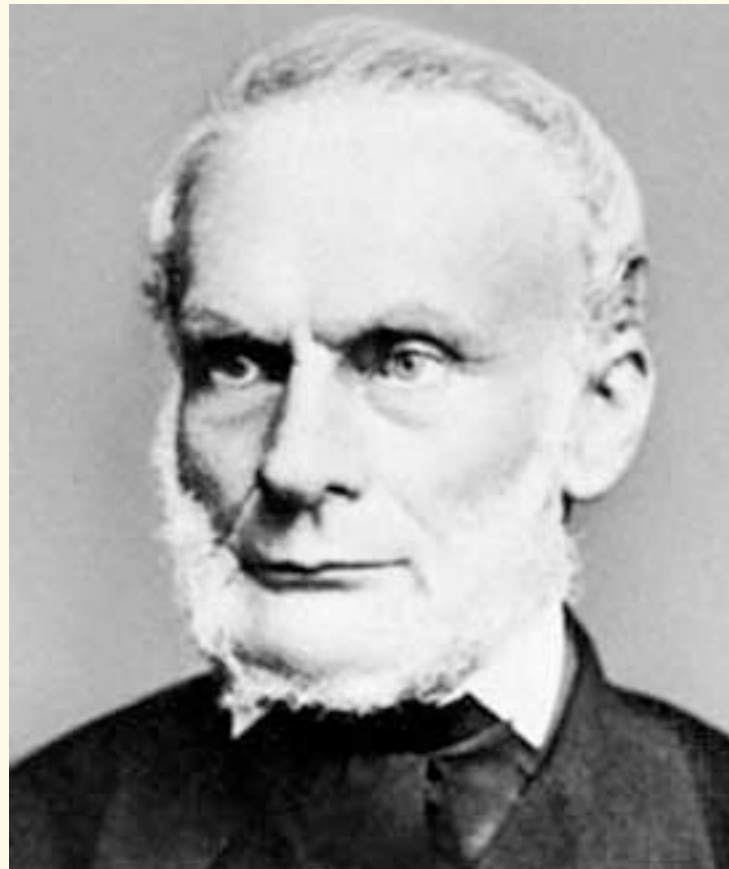
- ・ ある都市のある日の天気予報が、晴れ45%、曇35%、雨12%、雪8%のとき、エントロピー H を求めよ。
- ・ 「いろは」48文字の生起確率が全て等しいときと仮定した時のエントロピー H を求めよ。

エントロピーの歴史

- ・ 1850年クラウジウスがエントロピーという言葉を作る。熱力学で用いる物理量。
- ・ 1877年ボルツマンが状態の数をエントロピーと結びつけた。
- ・ 1948年シャノンが情報理論におけるエントロピーを導入する。
 - ・ 情報理論が確立

クラウジウス

- ・ 1822年生まれ
- ・ 1854年熱力学第二法則を確立
- ・ 1865年エントロピーを初めて使う



熱力学

- ・ 現象論的な学問
- ・ 要素を考えない
 - ・ 熱のやりとりのみに注目
 - ・ 原子が存在していようがまいが関係ない
- ・ 原子を前提としていないので、乱雑さとは無縁

$$\Delta Q = T \Delta S$$

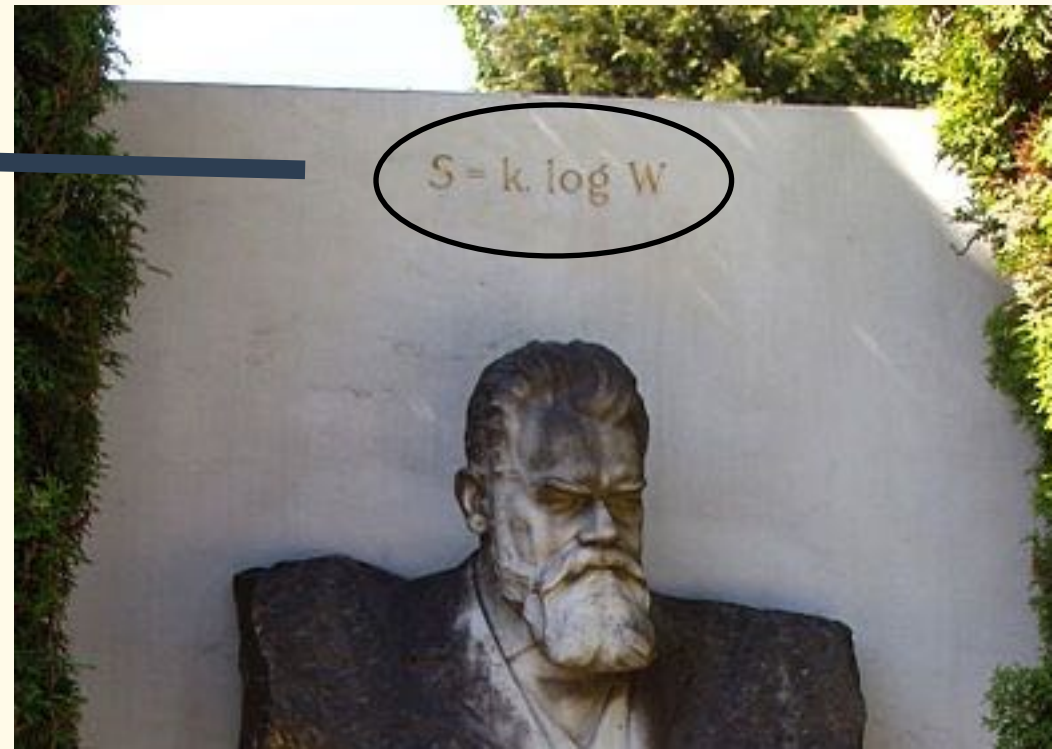
ボルツマン(1844-1906)

- ・ 1844年ウィーン生まれ
- ・ 1877年ボルツマンの関係式発表
 - ・ エントロピーは状態の数と関係がある



$$S = k \log W$$



A photograph of a stone bust of Boltzmann on a wall. Above the bust, the formula $S = k \cdot \log W$ is inscribed on the wall and circled in black.
$$S = k \cdot \log W$$

wikipediaより

統計力学

- ・ 原子の存在を前提としている。
 - ・ 還元主義的
- ・ 原子の状態がばらばらだとエントロピーが増える。
 - ・ 現実社会に当てはめると乱雑さといえるかも。
- ・ 状態数 W の対数がエントロピー

$$S = k \log W$$

ボルツマンの関係式

シャノン(1916-2001)

- ・ 1916年ミシガン州出身
- ・ 1948年通信の数学理論発表
 - ・ 情報を定量化
 - ・ 情報源符号化
 - ・ 通信路符号化



wikipediaより

まとめ

- ・ 情報量

$$I(p) = -\log_2 p \quad \text{ビット}$$

- ・ エントロピー

$$H(p_1, p_2, \dots, p_n) = -\sum_{i=1}^n p_i \log_2 p_i$$

次回

- ・ 結合エントロピー
- ・ 相互エントロピー