目次

★3　パターン情報の表現 (2) — データ圧縮と情報量（承前）

「承前」って？ ⇒ 辞書引きましょう

★ 3.6　情報量の概念

★ 3.6.1　情報量の定義

ある事象が起こったことを知らせる際に伝達される「情報の量」を考えてみよう.「1 億枚中 1 枚しか当たりのないくじではずれをひいた」という知らせよりも「そのくじで当たりをひいた」という知らせの方が情報が多い気がしないだろうか (☆ 1).「犬が人を噛んだ」というニュースよりも「人が犬を噛んだ」という方が情報が多い気がしないだろうか．実は，このような素朴な議論から出発して，**情報量**というものを次のように定義するとよいことが知られている.

☆ 1)「びっくり度」が高い，と言ってもよいかも.

ある事象 E の生起確率が $P(E)$ であるとき，E が起こったことを知らせる際に伝達される情報量 $I(E)$ を次式で定義する：

$$I(E) = \log\left(\frac{1}{P(E)}\right) = -\log P(E) \tag{1}$$

対数の底は任意に定めればよい (☆ 2) が，bit との対応づけができるため，2 を選ぶことが多い．その場合，情報量の単位は [bit] となる.

☆ 2) ある二つのことの情報量に注目すると，底の選び方を変えてもそれらの大小関係は変化しない．ただし底の選び方によって情報量の値そのものは変化する.

例： 確率 $\frac{1}{2}$ で赤旗か白旗のいずれかが上がるのを観測する場合,「赤（白）の旗が揚がった」という知らせの情報量は $-\log_2 \frac{1}{2} = 1$[bit] となる (☆ 3).

☆ 3)「赤旗が揚がった」と「白旗が〜」の 2 つに 2 進数を対応づけるためには 1 桁 (bit) の 2 進数が必要であるということに対応している．たとえば，0 が「赤旗が〜」で 1 が「白旗が〜」.

例： 天気が「晴れ」か「雨」のどちらかにしかならない地域があったとする．晴れの確率は 0.9 だという．この場合，この地域の天気が晴れだということ，および雨だということを知らせる情報の情報量は次のようになる.

$$I(\text{晴れ}) = -\log_2 0.9 = -\frac{\log 0.9}{\log 2} = \text{約 } 0.152\text{[bit]}$$

$$I(\text{雨}) = -\log_2 0.1 = -\frac{\log 0.1}{\log 2} = \text{約 } 3.32\text{[bit]}$$

例： 4 つの事象 A, B, C, D のどれかが起こるとして，それらの生起確率が等しい（つまり $\frac{1}{4}$ である）場合，どの事象が起こったかを知らせる情報の情報量は 2[bit] となる (☆ 4).

☆ 4) A, B, C, D に対応させるには 2 桁 (bit) の 2 進数が必要.

Q1. 20 通りの事象のいずれかが起こるとする．それらの生起確率が等しい場合，どの事象が起こったかを知らせる情報の情報量は何 bit か．必要ならば $\log_2 5 = 2.32$ を用いたらよい.

Q2. 「晴れ」,「曇り」,「雨」,「雪」の 4 通りの天気のいずれかが起こる地域があったとする．この地域の雪の確率が $\frac{1}{512}$ だとすると，この地域の天気が雪だということを知らせる情報の情報量は何 bit になるか.

★ 3.6.2 情報量の性質

例： $4 \times 13 = 52$ 枚のカードから成るトランプから無作為に 1 枚引くなら

- 「それが ♡ だった」という知らせの情報量は $-\log \frac{1}{4} = \log 4$
- 「エース (1) だった」という知らせの情報量は $-\log \frac{1}{13} = \log 13$
- 「♡ のエースだった」という知らせの情報量は $-\log \frac{1}{52} = \log 52$

となる．つまり，$I(♡) + I(\text{エース}) = I(♡ \text{ かつエース})$ である (☆ 5)．

ここから予想できるように，情報量には次のように**加法性**が成り立つ (☆ 6)：

事象 A の情報量が $I(A)$, 事象 B の情報量が $I(B)$ であるとき，A, B が独立ならば，事象 $A \cap B$ の情報量（A も B も起こったことを知らせる）は

$$I(A \cap B) = I(A) + I(B) \tag{2}$$

となる (☆ 7)．

Q3. さいころを振って出た目が「3 の倍数だった」,「偶数だった」,「6 だった」ということを知らされることで得られる情報量をそれぞれ求めなさい．ただし，単位は bit とすること．これらの間にはどのような関係があるだろう（ヒント：「3 の倍数」かつ「偶数だった」というのはどんな場合？）．

☆ 5) ここでは対数の底を e としているが，他の値にしても結論は変わらない．

☆ 6) 実際の理論の成り立ちは逆で,「素朴に考えると情報量というのは『生起確率の単調減少関数』でありかつ『加法性が成り立つ』ものでなければならない」というのを出発点にして，そのような性質を満たすのは式 (1) のような対数の形でなければならないことが導出されている．

☆ 7) 「ハートだった」を A,「赤いのだった」を B としてみると，どうなるだろう．

★ 3.6.3 平均情報量（エントロピー）

n 個の独立な事象 $E_1, E_2, \ldots, E_n$ がそれぞれ確率 $p_1, p_2, \ldots, p_n$ で生起する場合を考える．$\sum_{i=1}^{n} p_i = 1$ とする．このとき，事象 E_i が「起こった」という知らせの情報量 $I(E_i)$ は $I(E_i) = -\log p_i$ である．それでは，「$E_1, E_2, \ldots, E_n$ のどれが起こったかまだ知らない状況でどれが起こったかを知らせてもらう場合，その知らせは平均としてどれだけの情報量をもつと期待できるか」を考えてみよう．その値を H とおくと，H は得られる情報量の期待値であるから，

$$H = \sum_{i=1}^{n} p_i I(E_i) = -\sum_{i=1}^{n} p_i \log p_i \tag{3}$$

ただし，$\lim_{p\to+0} p\log p = 0$ より，$p = 0$ の場合の $p\log p$ の値は 0 として扱う．

となる．これは，「どれが起こったか知らない不確定な状況を確定させることで得られる情報量の平均値」を表しており，**平均情報量**あるいは**エントロピー**と呼ばれる．「知ろうとしている状況の不確かさの度合い」を表していると考えてもよい．

エントロピー: entropy. 熱力学でででてくるエントロピーと同じようなものと考えられる．

エントロピーについてはいろいろ興味深い話があるが，この授業では割愛する．

例: 確率 p で表が，確率 $1-p$ で裏が出るコインがある．このコインを投げたときに得られるエントロピーは，(底を 2 とすると)

$$H = p \times I(\text{表}) + (1-p) \times I(\text{裏}) = -p\log_2 p - (1-p)\log_2(1-p) \quad [\text{bit}] \tag{4}$$

となる．右図からわかるように，このエントロピーは $p = 0$ または $p = 1$ のときに 0（投げる前から結果がわかっている）となり，$p = \frac{1}{2}$ のときに最大 (1[bit]) となる（最も不確かな状況）．これは，「確率に偏りがある ⇔ エントロピーが小さい ⇔ 不確かさが小さい」ということを意味している．

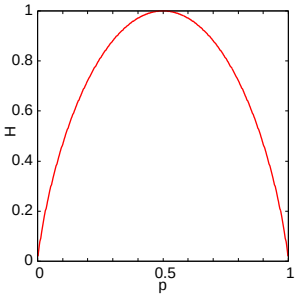

図: コイン投げのエントロピー

Q4. 「ほげおくんが 'H' という文字を書いた」という事象を E_1 と表すことにする．同様に，'O','G','E' を書いたという事象をそれぞれ E_2, E_3, E_4 と表記する．これら 4 つの事象 H,O,G,E がそれぞれ独立に確率 $\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8}$ で生起する場合を考える (☆ 8)．次の問に答えなさい．

(1) 事象 E_i が起こったという知らせの情報量 $I(E_i)$ を求めなさい $(i = 1, 2, 3, 4)$．
(2) 平均情報量を求めなさい．
(3) これら 4 つの事象をビットパターンに対応させて区別したい．すべて同じ bit 数で表すなら，1 つの事象を何 bit のビットパターンに対応させればよいか．
(4) ほげおくんが 'H' ばかり書くようになった（(E_1 の生起確率) $\to 1$）ら，平均情報量はどうなるか．

☆ 8) $\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{8} = 1$ ですね．ほげおくんが他の文字を書くことはないらしい...

★ 3.7　エントロピー符号化

前節の Q の条件で，ほげおくんが書いた文字の並び（‘H’,‘O’,‘G’,‘E’ の 4 種のみでできている）を単純に 1 文字あたり 2bit で符号化するなら，この文字の並びのデータ量は当然 1 文字あたり 2bit となる．しかし，各文字の出現頻度に偏りがあるせいで，彼の書いた文字の並びの平均情報量（エントロピー）は 2bit よりも小さくなっていた．何か工夫することで，1 文字あたり 2bit より少ないデータ量で符号化できる方法があるんではないだろうか？

実は，データ圧縮の手法として，エントロピーの概念に基づく**エントロピー符号化**という符号化法が知られている．

★ 3.7.1　エントロピー符号化の考え方

表に示す 6 つの記号が，**それらを並べた記号列中の並び方や位置に無関係に** (☆ 9)，表に示す確率で出現するとする．この記号列を 0,1 で符号化したい．

☆ 9) H のあとに O が出てきやすいとか，最初の方は H が出てきやすい，というようなことがない，という意味．

記号	D	E	F	G	H	O
確率	$\frac{10}{100}$	$\frac{2}{100}$	$\frac{15}{100}$	$\frac{13}{100}$	$\frac{20}{100}$	$\frac{40}{100}$

最も単純な方法は，各記号を 3bit の符号で表現する（記号 D に 000，E に 001，等），というものである．この場合，たとえば `HOGEHOHOOO` という 10 文字の並びは，$10 \times 3 = 30$bit のビット列となる．この例では，どの記号にも 3bit を割り当てることになるから，「1 つの記号を平均何 bit で表せるか」を表す**平均符号語長**を考えると，その値は 3bit となる．

しかし，これらの記号の出現頻度には偏りがあるので，平均情報量はもっと小さいと考えられる (☆ 10)．この例の場合に実際に計算してみると，

$$H = -\frac{10}{100}\log_2\frac{10}{100} - \frac{2}{100}\log_2\frac{2}{100} - \cdots - \frac{40}{100}\log_2\frac{40}{100} = \text{約 } 2.23[\text{bit}] \quad (5)$$

となる．このことは，符号の割り当て方を工夫すれば，平均符号語長をここまで小さくできる可能性があることを意味している．たとえば，右表のように出現確率の高いものに短い符号を割り当てるようにしてやると，平均符号語長を

$$\frac{10}{100}\times 4 + \frac{2}{100}\times 4 + \frac{15}{100}\times 3 + \frac{13}{100}\times 3 + \frac{20}{100}\times 3 + \frac{40}{100}\times 1 = 2.32[\text{bit}] \quad (6)$$

にすることができる．実際に上記の 10 文字の並びの例でやってみると，

$$11101011000011101110000 \quad (7)$$

となり，30bit だったものを 21bit に減らせている．

☆ 10) ちなみに，6 通りの記号が等確率で出現する場合でも平均情報量は $-6\times\frac{1}{6}\log_2\frac{1}{6} = \log_2 6 = $ 約 2.58[bit] である．

記号	符号
D	1001
E	1000
F	110
G	101
H	111
O	0

このように，エントロピー符号化とは，記号の出現頻度（確率）に偏りがある，すなわち平均情報量が小さい場合に，その偏りを利用して効率的な（平均符号語長の短い）符号化を実現しようとするものである．上記の例では平均情報量が約 2.23bit であるから，理想的なエントロピー符号化ができれば平均符号長をそこまで減らせることを意味している．

Q5. 式 (7) の符号を表に従って先頭から順に復号していくとちゃんと元通りになることを確認してみよう．