第 2 章

情報量とエントロピー

2.1　情報量

2.1.1　情報の量

情報の量を計るにはどうすればよいのか, そもそも情報とはいかなるものか, ということを最初に考えてみる.

• 情報とはわれわれに何事かを教えてくれるものであり, われわれのあいまいな知識を確実にしてくれるものである.

このことから, 情報の量を次のような方針で定義してみよう.

• 情報の量は, その情報をもらったことによって知識の不確実さ, あるいはあいまいさがどのくらい減少したかで計ればよい.

n 個の事柄 $A_1, A_2, \cdots, A_n$ があり, そのうちどれか 1 つが起こったものとする. $A_1, A_2, \cdots, A_n$ はどれも等確率で起こりやすいとし, われわれにはどれが起こったかわからないとき, どれが起こったかを教えてくれる通報には, どのくらいの量の情報が含まれているだろうか?

【例題 2-1】 サイコロ: $n = 6$ で, $A_1, A_2, \cdots, A_6$ はそれぞれ"1 の目が出た", "2 の目がでた", $\cdots$ となる.

硬貨なげ: $n = 2$ で, A_1 は"表が出た", A_2 は"裏が出た".

上の例において, どの A_i が起こったかを知らせる情報の量は, 起こり得る事柄

の数 n に関係することが容易に予想される．硬貨なげの方がサイコロの目を当てるより，かなり当てやすいことは確率的に理解できる．つまり，n が大きくなれば，どれが起こるのか予想しにくくなるので，情報の量は n が大きい程多くなる．そこで，この情報の量を $f(n)$ と書くことにする．

関数 $f(n)$ を決定するため，次のような状況を考える．n は 2 つの整数 m と k との積

$$n = mk \tag{2.1}$$

であるとし，n 個の事柄 $A_1, A_2, \cdots, A_n$ は k 個ずつの m グループに分けられているとする．

【例題 2-2】 サイコロで 6 種類の目を偶数と奇数に分けると，$\{A_2,\ A_4,\ A_6\}$ と $\{A_1,\ A_3,\ A_5\}$ の 2 つに分割[*1]される．

どの A_i が起こったかを，直接教える情報の量は $f(n)$ である．情報を段階的に教えるとき，つまり m 個のグループのどのグループに入っているかを教える情報の量は $f(m)$，そのグループの k 個の中のどれかを教える情報の量は $f(k)$ である．情報を直接知らせても，段階的に知らせてもその量は変わらないと考えられるので，

$$f(mk) = f(m) + f(k) \tag{2.2}$$

という式が成り立つと考えることができる．前出のサイコロの場合だと，$f(6) = f(2) + f(3)$ である．これは，段階的に小出しにした情報の量を加え合わせれば，全体の情報の量になるという，**情報の加法性**を示す式である．

加法法則 (2.2) を満足する関数 $f(x)$ は，適当な連続性を仮定すると一意に定まってしまう．実数 x の関数 $f(x)$ を考え，これは微分可能でしかも

$$f(xy) = f(x) + f(y) \tag{2.3}$$

を満たすとする．すると，

$$f(x + \varepsilon x) = f\{(1+\varepsilon)x\} = f(1+\varepsilon) + f(x) \tag{2.4}$$

より，

$$\frac{f(x+\varepsilon x) - f(x)}{\varepsilon x} = \frac{1}{x}\frac{f(1+\varepsilon)}{\varepsilon} \tag{2.5}$$

[*1] 集合 C を部分集合 A と B に分けるとき，A ∩ B = ϕ (空集合) のとき，これを分割とよぶ．

ここで, $\varepsilon \to 0$ の極限をとると, 左辺は明らかに $f(x)$ の導関数 $f'(x)$ となる. したがって,

$$\lim_{\varepsilon \to 0} \frac{f(1+\varepsilon)}{\varepsilon} = c \tag{2.6}$$

とおけば,

$$f'(x) = \frac{c}{x} \tag{2.7}$$

という微分方程式が得られる. これを積分すれば,

$$f(x) = c \log_e x + d \tag{2.8}$$

となる. d は積分定数である.

ここで c と d を決定しよう. 例えば, 起こり得る事柄が 1 つしかないとき ($n = 1$), これが起こっても, 得られる情報は 0 である. 従って,

$$f(1) = 0. \tag{2.9}$$

これより,

$$d = 0. \tag{2.10}$$

定数 c は情報量の尺度に関連している. 情報の最もシンプルで基本的な形は, 2 つの事柄のうち 1 つを選択し教えてくれるものであり, これは yes か no や, "ある"か"ない"のような二者択一の情報の単位である. これを 1 ビット (bit：binary digit) と呼ぶ. これより,

$$f(2) = 1 \quad \text{ビット} \tag{2.11}$$

であり[*2], 従って,

$$c = \log_2 e \tag{2.12}$$

となる[*3]. 結局 $f(x)$ は,

$$f(x) = \log_2 x \quad \text{ビット} \tag{2.13}$$

である[*4]. 以後, **対数の底は, 特に断らないかぎり 2 をとる**ものとする.

[*2] $f(2) = c \log_e 2 = 1$

[*3] $c = \frac{\log_2 e}{\log_2 2} = \log_2 e$

[*4] $f(x) = c \log_e x = \log_2 e \cdot \log_e x = \log_2 e \cdot \frac{\log_2 x}{\log_2 e} = \log_2 x$

いま, 2^m 個の事柄のうち 1 つが起こったとする．どれが起こったかは全体を半分, 半分と分けていって, 起こった事柄がどちらの側に入っているかを右か左で m 回聞けばわかる．すなわち m ビットの情報で定まり, 実際,

$$f(2^m) = \log_2 2^m = m \tag{2.14}$$

である．

【例題 2-3】 8 人の容疑者がいる。それぞれに 1 番から 8 番までの番号がついている．真犯人は 3 番であるとき，まず，事件の目撃者に 1 番から 4 番の中に犯人がいるかを聞く．次に 1 番から 2 番の中に犯人がいるか聞く．すると Yes-No と答えが返ってくる．最後に 3 番かと聞くと Yes で，真犯人が確定する．これは，$\log_2 8 = \log_2 2^3 = 3$ で，3 回の 2 択の答えとなる．

ここで, ある事柄 A があり, これが起こる確率は p であるとする．いま, これが実際に起こったということを教えてくれる情報があったとき, この情報の量はどのくらいだろうか.

p を有理数であるとして,

$$p = \frac{k}{n} \tag{2.15}$$

とおく．ここで n 個の等確率で起こる事柄を考え, このうちの k 個をひとまとめにしたものが A だとしよう．A が起こるとは, この k 個のうちのどれか 1 個が起こることだとすれば, A の起こる確率は $p = k/n$ である．

さて, n 個のうちのどれが起こったかを知るための情報量は $\log_2 n$ であった．いま, A が起こったことを教えてくれる情報の量を I とする．A が起こったあと, A の中の k 個のうち, どの 1 個が起こったかを知るためにはさらに $\log_2 k$ の情報が必要である．従って,

$$\log_2 n = I + \log_2 k \tag{2.16}$$

すなわち

$$I = -log_2 \frac{k}{n} = -\log_2 p \quad [bit] \tag{2.17}$$

が求まる．前に出てきた例では, $A_1, A_2, \cdots, A_n$ はすべて $1/n$ の確率で起こったので, このうちのどれかが起こったかを教える情報の量は,

$$-\log_2 \frac{1}{n} = \log_2 n \tag{2.18}$$

となって, (2.13) 式と一致する.

定義 2.1 確率 p の事象が実際に生起したことを知らせる情報に含まれている情報を

$$-\log_2 p \quad [bit] \tag{2.19}$$

と定義し，これを**自己情報量**とよぶ.

$-\log_2 p$ は p の小さいほど大きい. これは，めったに起こらない事柄が起こることが情報量が多いことに相当するからである.

【例題 2-4】 サイコロを 1 回振って 1 の目が出る場合の自己情報量は,
$p_1 = -\log_2 \frac{1}{6} = \log_2 6 = \frac{\log_{10} 6}{\log_{10} 2} = \frac{0.778}{0.3} = 2.59$ [bit]
硬貨を投げて表が出る場合の自己情報量は,
$p_2 = -\log_2 \frac{1}{2} = \log_2 2 = 1$ [bit]
従って，予想しやすい硬貨投げの自己情報量はサイコロ振りより小さい.

2.1.2　平均情報量とエントロピー

前節では, 確率 p の事象 A が起こったときは, この情報量は $-\log p$ であることを論じた（自己情報量）.

ところで, 事象 A が起こったかどうかの情報量はどうなるだろうか. A が起こった場合の情報量は, 上に述べたように, $-\log p$ である. 逆に, A が起こらなかった場合の情報量は, $-\log(1-p)$ となる.

一般に $A_1, A_2, \cdots, A_n$ の n 個の事象があって, それぞれ $p_1, p_2, \cdots, p_n$ の確率で生ずる場合を考える. そうすると,

$$\sum_{i=1}^{n} p_i = 1 \tag{2.20}$$

が成り立つ. これを完全事象系といった. ここで, どの事象が起こったかを知ったとき, 得られる情報の量は, どの A が生じたかで異なる. すなわち, A_i が起これば $-\log p_i$ という情報が得られる. ところで, A_i の起こる確率は p_i であるので, 得ら

れる情報の量の**期待値**は $-\log_2 p_i$ を確率 p_i で平均したものであり,

$$H = -\sum_{i=1}^{n} p_i \log_2 p_i \tag{2.21}$$

となる．これが, どの A_i が起こったかを聞くときに得られるであろう情報の量 (平均情報量) と考えてよい.

一般にわれわれが得られるのは情報量の期待値であり, この値は不確定な状況を確定するのに要する**平均情報量**だといえる.

状況の不確定度を表す量 (平均情報量) を**エントロピー**とよび, 以下のように定義する.

定義 2.2 n 個の事象がそれぞれ確率 $p_1, p_2, \cdots, p_n$ で生起するとき, どれが生起したのかの不確定度は,

$$H(p_1, p_2, \cdots, p_n) = -\sum_{i=1}^{n} p_i \log p_i \tag{2.22}$$

と定義できる.

エントロピーには次のような性質がある.

性質 1. エントロピー H は非負

$$H \geq 0 \tag{2.23}$$

であり,

$$H = 0 \tag{2.24}$$

が成立するのは, どれか 1 つの p_i が 1 で他はすべて 0 のときであり, そのときに限る．(エントロピーが 0)

性質 2. n 個の事象の表すエントロピーの最大値 $H(n)$ は,

$$H(n) = \log_2 n \tag{2.25}$$

で, これはすべての事象が等しい確率

$$p_i = \frac{1}{n} \tag{2.26}$$

で起こるときである．このことについては次の例題の後に詳述する.

【例題 2-5】 琉球大学のある冬の天気予報が，晴れ 25%, 曇り 50%, 雨 12.5%, あられ 6.25%, みぞれ 6.25% であったとする．このとき天気予報によって得られるエントロピー H を求めよ．

$$\begin{aligned} H &= -\frac{1}{4}\log_2\frac{1}{4} - \frac{1}{2}\log_2\frac{1}{2} - \frac{1}{8}\log_2 18 - 2\cdot\frac{1}{16}\log_2\frac{1}{16} \\ &= \frac{1}{4}\cdot 2 + \frac{1}{2}\cdot 1 + \frac{1}{8}\cdot 3 + \frac{1}{16}\cdot 4\cdot 2 \\ &= 1 + \frac{3}{8} + \frac{1}{2} = 1 + 0.375 + 0.5 \\ &= 1.875[bit] \end{aligned}$$

2.1.3 最大エントロピー

二つの事象 a_1, a_2 からなる事象系（2 元事象系）を考える．a_1, a_2 のそれぞれの生起確率が p_1, p_2 であるとき，これらを並べて次のように表記できる．

$$\mathbf{A} = \begin{pmatrix} a_1, & a_2 \\ p_1, & p_2 \end{pmatrix}$$

ここでエントロピーは，

$$H = -p_1 \log_2 p_1 - p_2 \log_2 p_2 \tag{2.27}$$

である．ここで，p_1, p_2 が排反事象であれば，

$$p_1 + p_2 = 1$$

である．H が最大になるのは H を p_1 で微分してこれを 0 となるところを求めればよい．

$$\begin{aligned} \frac{dH}{dp_1} &= -\log_2 p_1 - p_1\frac{1}{p_1} - \frac{dp_2}{dp_1}\log_2 p_2 - p_2\frac{1}{p_2}\frac{dp_2}{dp_1} \\ &= -\log_2 p_1 - 1 - \frac{dp_2}{dp_1}\log_2 p_2 - \frac{dp_2}{dp_1} \end{aligned}$$

ここで，$p_1 + p_2 = 1$ を p_1 で微分すると，$1 + \frac{dp_2}{dp_1} = 0$ より $\frac{dp_2}{dp_1} = -1$ これより，

$$\frac{dH}{dp_1} = -\log_2 p_1 - 1 + \log_2 p_2 + 1 = 0$$

従って,

$$\log_2 p_1 = \log_2 p_2$$

より,

$$p_1 = p_1$$

よって，エントロピー H が最大になるのは p_1, p_2 が等確率のときである．このときのエントロピーは,

$$H_{max} = -\log_2 \frac{1}{2} = 1 \quad [bit]$$

2.1.4 結合エントロピー

2 種類の事象, つまり 2 つの事象系が組合わさって起こる場合を考える．これを**複合事象**と呼ぶ．複合事象 (A_i, B_j) の生じる確率を $p(A_i, B_j)$ と書くことにすると, その不確定度を示すエントロピー $H(A, B)$ は,

$$H(A,B) = -\sum_{i,j} p(A_i, B_j) \log p(A_i, B_j) \tag{2.28}$$

と書ける．ここで, $i = 1 \sim m$, $j = 1 \sim n$ とする．これを，**結合エントロピー**という．ところで, A_i や B_j が起こる確率は, それぞれ,

$$p(A_i) = \sum_{j=1}^{n} p(A_i, B_j) \tag{2.29}$$

$$p(B_j) = \sum_{i=1}^{m} p(A_i, B_j) \tag{2.30}$$

であり, 各々のエントロピーは,

$$H(A) = -\sum_{i=1}^{m} p(A_i) \log p(A_i) \tag{2.31}$$

$$H(B) = -\sum_{j=1}^{n} p(B_j) \log p(B_j) \tag{2.32}$$

で表される.

【例題 2-6】 ある医者が患者の風邪を診断するのに, まず熱があるかどうかを調べ,

その後でもっと精密な検査をすることにしている．では, 熱を調べただけで, どれくらいの情報が得られるか.

A_1：熱がある
A_2：熱がない
B_1：風邪をひいている
B_2：風邪をひいていない

	B_1	B_2	$p(A_i)$
A_1	0.55	0.05	0.60
A_2	0.10	0.30	0.40
$p(B_j)$	0.65	0.35	1.00

$$
\begin{aligned}
H(A) &= -\sum_{i=1}^{2} p(A_i)\log p(A_i) && (2.33)\\
&= -0.6\log 0.6 - 0.4\log 0.4 = 0.97\ [bit] && (2.34)\\
H(B) &= -\sum_{j=1}^{2} p(B_j)\log p(B_j) && (2.35)\\
&= -0.65\log 0.65 - 0.35\log 0.35 = 0.93\ [bit] && (2.36)\\
H(A,B) &= -\sum_{i,j}^{2} p(A_i,B_j)\log p(A_i,B_j) && (2.37)\\
&= -0.55\log 0.55 - 0.1\log 0.1 - 0.05\log 0.05\\
&\quad -0.3\log 0.3 = 1.54\ [bit] && (2.38)
\end{aligned}
$$

また，上述の事より，

$$H(A), H(B) \quad \leq \quad H(A,B) \quad \leq \quad H(A) + H(B) \tag{2.39}$$

が成立することに注意を要する.

2.1.5　条件付エントロピー

結合エントロピー $H(A_i, B_j)$ において，A_i と B_j が同時に起こることを，$A_i \cap B_j$ と表すと，

$$
\begin{aligned}
H(A_i, B_j) &= -\sum_i \sum_j p(A_i \cap B_j) log_2 p(A_i \cap B_j) \\
&= -\sum_i \sum_j p(A_i)p(B_j|A_i) log_2 p(A_i)p(B_j|A_i) \\
&= -\sum_i \sum_j p(A_i)p(B_j|A_i)\{log_2 p(A_i) + \log_2 p(b_j|A_i)\} \\
&= -\sum_i \sum_j p(A_i)p(B_j|A_i) \log_2 p(A_i) \\
&\qquad -\sum_i \sum_j p(A_i)p(B_j|A_i) \log_2 p(B_j|A_i) \\
&= -\sum_i p(A_i) \log_2 p(A_i) \sum_j p(B_j|A_i) \\
&\qquad -\sum_i p(A_i) \sum_j p(B_j|A_i) \log_2 p(B_j|A_i)
\end{aligned} \tag{2.40}
$$

$\sum_j p(B_j|A_i) = 1$ なので，(2.40) 式の第 1 項は $H(A)$ である．また，第 2 項の $\sum_j p(B_j|A_i) \log_2 p(B_j|A_i)$ は条件 A_i の下での B_j のエントロピーである．したがって第 2 項全体はその A_i に関する平均値であることから $H(B|A)$ を意味しているので，

$$H(B|A) = -\sum_i p(A_i) \sum_j p(B_j|A_i) \log_2 p(B_j|A_i) \tag{2.41}$$

であり，この $H(B|A)$ を**条件付きエントロピー** (conditional entropy) という．これらのことから (2.40) 式は，

$$H(A, B) = H(A) + H(B|A) \tag{2.42}$$

となる．同様に，

$$H(B, A) = H(B) + H(A|B) \tag{2.43}$$

を導くことができるので，これらから，

$$H(A, B) = H(B, A) \tag{2.44}$$

といえる．

【例題 2-7】 前の例において，熱の有無がわかったという条件付の病気のエントロピー (検温後にもなお残る不確定度) を求める．Bayes の公式より，

$$\begin{aligned} p(B_1|A_1) &= 0.55/0.6 = 0.92 \\ p(B_2|A_1) &= 0.05/0.6 = 0.08 \\ p(B_1|A_2) &= 0.1/0.4 = 0.25 \\ p(B_2|A_2) &= 0.3/0.4 = 0.75 \end{aligned}$$

である．従って，

$$\begin{aligned} H(B|A_1) &= -0.92\log 0.92 - 0.08\log 0.08 = 0.402 \\ H(B|A_2) &= -0.25\log 0.25 - 0.75\log 0.75 = 0.811 \end{aligned}$$

これより，

$$H(B|A) = 0.6 \times 0.402 + 0.4 \times 0.811 = 0.5656 \simeq 0.57$$

検温前の B についてのエントロピー $H(B)$ は 0.93 であった．検温をすませれば，それだけ不確定度は減少する．検温によって得られる風邪についての情報量の平均値は，エントロピーの減少分

$$I = H(B) - H(B|A) = 0.93 - 0.57 = 0.36 \text{ bit}$$

である．

コラム — 条件付きエントロピー

A と B が独立でないなら, A が何であるかを知ってしまえば, B に関する不確定度は減少する．一方の事柄が何であるかわかっているという条件のもとでの他方の事柄の不確定度を**条件付エントロピー**というのであった.
条件付確率 $p(B_j|A_i)$ とは A に関する事象が A_i であったときの B_j の確率である．これらの間には次の関係が成立する.

$$\sum_{i,j} p(A_i, B_j) = 1 \tag{2.45}$$

$$\sum_i p(A_i) = \sum_j p(B_j) = 1 \tag{2.46}$$

$$\sum_j p(B_j|A_i) = \sum_i p(A_i|B_j) = 1 \tag{2.47}$$

$$p(A_i) = \sum_j p(A_i, B_j) \tag{2.48}$$

$$p(B_j) = \sum_i p(A_i, B_j) \tag{2.49}$$

$$p(A_i, B_j) = p(A_i)p(B_j|A_i) = p(B_j)p(A_i|B_j) \tag{2.50}$$

$$p(B_j|A_i) = \frac{p(B_j)p(A_i|B_j)}{p(A_i)} = \frac{p(A_i, B_j)}{p(A_i)} \tag{2.51}$$

$$p(A_i|B_j) = \frac{p(A_i)p(B_j|A_i)}{p(B_j)} = \frac{p(A_i, B_j)}{p(B_j)} \tag{2.52}$$

式 (2.49) は, A_i と B_j の同時確率が, A_i の確率に A_i が起こるという条件下での B_j の起こる確率を掛けたものであることを示している.
式 (2.50), (2.51) は **Bayes の公式**と呼ばれ, (2.49) 式から導出できる.

A と B が確率的に独立な場合には,

$$p(A_i, B_j) = p(A_i)p(B_j) \tag{2.53}$$

$$p(B_j|A_i) = p(B_j), \quad p(A_i|B_j) = p(A_i) \tag{2.54}$$

2 つの事象 A, B があり, A では A_i が起こったことがわかったとする．このとき, B として $B_1, B_2, \cdots, B_m$ が起こる確率は, それぞれ条件付確率 $p(B_1|A_i)$, $p(B_2|A_i)$, $\cdots$, $p(B_m|A_i)$, であった．それゆえ, このときの (不確定度を表す) エントロピーは,

$$H(B|A_i) = -\sum_{j=1}^{m} p(B_j|A_i) \log p(B_j|A_i) \tag{2.55}$$

となる．さらに, A_i の起こる確率は $p(A_i)$ である．だから, A が何であるかを知ったあとの B についての不確定度を表すエントロピーは, $H(B|A_i)$ をすべての A_i について平均して, 次式のようになる．

$$\begin{aligned} H(B|A) &= \sum_{i=1}^{n} p(A_i) H(B|A_i) \\ &= -\sum_{i,j} p(A_i) p(B_j|A_i) \log p(B_j|A_i) \end{aligned} \tag{2.56}$$

このエントロピーの平均値 $H_A(B)$ のことを条件付エントロピーという．

以下に条件付エントロピーに関する性質をまとめる．

$$H(A,B) = H(A) + H(B|A) \tag{2.57}$$
$$= H(B) + H(A|B), \tag{2.58}$$
$$H(B|A) = H(A,B) - H(A) \tag{2.59}$$

$$H(A) + H(B) \geq H(A,B) \tag{2.60}$$
$$H(A) \geq H(A|B), \tag{2.61}$$
$$H(B) \geq H(B|A) \tag{2.62}$$
$$H(A,B) \geq H(A), H(B) \tag{2.63}$$

2.1.6　相互情報量

事象系 A と B が多少でも関係があれば, A が何であるかを知ると B についてもある程度の情報を得ることになる．この情報量は, A を知ることによってもたらされる B のエントロピーの減少分, すなわち B のエントロピー $H(B)$ から, A を知った後になお残る B のエントロピー $H(B|A)$ を引いたものである．これを $I(A,B)$ と書くと,

$$I(A,B) = H(A) - H(B|A) \tag{2.64}$$

$H(B|A)$ は A が A_i であることを知った後のエントロピー $H(B|A_i)$ の平均値であるから, $I(A,B)$ は A が A_i であることを聞いて得られる情報量 $H(B)-H(B|A_i)$ の A_i についての平均値である.

$I(A,B)$ を A と B との**相互情報量**と呼ぶ.

相互情報量の性質

$$I(A,B) = I(B,A) \tag{2.65}$$
$$I(A,B) \leq H(A),\ H(B) \tag{2.66}$$
$$I(A,B) \geq 0 \tag{2.67}$$

3 つの事象系 A,B,C の間の情報関係について考える. このため条件付相互情報量 $I(A,B|C)$ を

$$I(A,B|C) = \sum_{i,j,k} p(C_k)p(A_i,B_j|C_k)\log\frac{p(A_i,B_j|C_k)}{p(A_i|C_k)p(B_j|C_k)} \tag{2.68}$$

で定義する. これは C が C_k であることがわかっているときの A と B との相互情報量の平均値である.

これらの性質を以下に示す.

$$I(A,B|C) \geq 0 \tag{2.69}$$
$$I(AB,C) = I(A,C) + I(B,C|A) \tag{2.70}$$
$$I(AB,C) = I(A,C) \tag{2.71}$$
$$I(A,B,C) = I(B,C,A) = \cdots \tag{2.72}$$

情報量のまとめ:

1. 自己情報量 (Amount of self-information) : $I(A)$, $I(B)$
2. 相互情報量 (Amount of mutual information) : $I(A,B)$
3. 条件付自己情報量 : $I(B|A)$, $I(A|B)$
4. 条件付相互情報量 : $I(A,B|C)$

章末問題

問 2.1　一個のサイコロを振って 5 の目がでたことを知らされたとき，この情報の情報量は何ビットか.

問 2.2　一組のトランプからカードを一枚抜いて，それがハートの 2 であることがわかったとき，得られた情報量は何ビットか.

問 2.3　一組 52 枚のトランプのカードの中から一枚のカードを抜いて，それがハートであることを知らされたとき，得られた情報量はいくらか．また，抜いたカードがエースであることを知らされたときの情報量はいくらか．抜いたカードがハートであることを知らされ，さらにエースであることを知らされたときの情報量は何ビットか.

問 2.4　二種類の文字 A, B よりなる言語において，A の出現する確率は 1/3, B の出現する確率は 2/3 であることが知られているものとする．この言語による情報の平均値（エントロピー）を求めよ.

問 2.5　四種類の文字 A, B, C, D よりなる言語において，これらの文字の現れる確率は各々 1/3, 1/4, 1/4, 1/6 である．この言語による平均情報量（エントロピー）を求めよ．但し，$\log_{10} 3 \neq 0.477, \log_{10} 6 \neq 0.778$ である.

問 2.6　つぼの中に白玉 5 個，赤球 15 個の計 20 個が入っている．このつぼから一つの玉を取り出すことにより得られる平均情報量を求めよ.

問 2.7　20 個ずつ玉の入っているつぼが 2 個ある．第一のつぼには白玉 10 個，黒玉 5 個，赤玉 5 個が入っており，第二のつぼには白玉 8 個，黒玉 8 個，赤球 4 個が入っている．この二つのつぼから各々一個だけ玉を取り出すことによって得られる平均情報量を求めよ．また，試行の結果はどちらのつぼの方が平均として予測しにくいか.

問 2.8　今，甲は N(整数) 以下のある一つの正の数を考えている．乙は甲の考えている数を当てようとする．乙のどんな質問にも甲は yes/no で答えるものとすれば，乙は何回質問すれば甲の考えている数を当てることができるか.