連載

世界をORする視線 (18) 第I部　通信・デジタル技術の発展 (3) コンピュータの発展：コンピュータ科学の数学的基礎（続き5）

住田　潮

（註：本稿は前回からの続きであるので，文献リストは継続し，新たに必要となる分を追加する）

1.　通信システムの一般モデル

通信の本質は，ある地点で用意されたメッセージをコード化したうえで信号に変換し，それを送信することにより，別の地点で元のメッセージを復元することにある．メッセージ自体は，当然，意味を伴うが，たとえば受信された文章，音楽や画像に対して受信者がどのような解釈を与えるかは千差万別である．通信システムを工学的に捉えようとするならば，情報の価値や量を定義するに際し，そのような意味的側面を引き剥がす必要がある．情報科学のパイオニアたちは，この必要性を直感的に把握し，厳密な表現を与えることに苦労した．連載第16回で論じたように，ナイキストは通信に内在する情報量を，単位時間内に送信できる文字数の対数値と通信スピード（1秒間に送信できる単位時間の数）で表わすことを提案し，これを「Intelligence（インテリジェンス）」という曖昧な概念で表わした．ハートレーは「Information（インフォーメーション）」という言葉を考案してIntelligenceという概念の曖昧さを払拭し，数学的にはナイキストの式を一般化したが，心理に影響される意味的内容を捨て去ることの重要性を説明するのに，さまざまな事例を駆使して苦労した．

これらの先駆的研究を受けて，1948年，シャノンは情報と通信に関する記念碑的論文「The Mathematical Theory of Communication（通信の数学的理論）」を発表，翌年，書籍版を出版した．この論文の偉大さは，

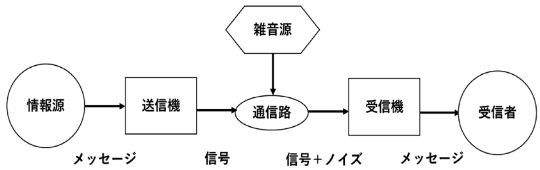


図1　シャノンの通信一般モデル

どのような通信システムをも網羅することのできる一般モデルを確立，通信に伴う量と価値を厳密に定義したうえで，さまざまな課題を構造的に分解・把握し，各課題に対して数学的な回答を与えた点にある．この体系の中で，通信理論を確立するためには，通信内容の意味的側面を無視すべきであるという視点の重要さも，自ずから明らかとなった．シャノンは，「情報の価値はメッセージを用意する際の選択の自由度に基づいて計量化されるべきである」という普遍的本質を見抜き，確率論を主要な解析の道具として用い，情報理論を全く新しい科学分野として確立したのである．

シャノンが構想した通信システムの一般モデルは，図1に示すようにシンプルなものである [12].

① 情報源でメッセージが作成される
② 送信機は，メッセージを符号化したうえで，通信可能な信号へと変換する
③ 通信路は，送信機から受信機へと信号が通過する媒体である
④ 通信路を通過する信号に対して，雑音源は歪みや乱れを発生させる
⑤ 受信機は，送信機とは逆の操作を行い，メッセージを復元する
⑥ 受信者は，復元されたメッセージを受け取る

この単純な構造モデルの長所は，アナログ通信，デジタル通信，DNA-RNAが媒介する人間の生体システムに

すみた　うしお
筑波大学名誉教授
〒305–8573 茨城県つくば市天王台1–1–1

表 1

目の和	2	3	4	5	6	7	8	9	10	11	12
確率	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

表 2

目の和	2	3	4	5	6	7	8	9	10	11	12
確率｜A	1/6	2/6	3/6	0	0	0	0	0	0	0	0

目の和	2	3	4	5	6	7	8	9	10	11	12
確率｜B	0	1/18	1/18	2/18	2/18	3/18	3/18	2/18	2/18	1/18	1/18

表 3

目の和	2	3	4	5	6	7	8	9	10	11	12
確率｜BC	0	1/9	0	2/9	0	3/9	0	2/9	0	1/9	0

おける情報伝播，暗号システムなど，あらゆる通信システムを網羅できる点にある．シャノンは，六つのブロックで構成されるモデルの各段階に対応して解明すべき課題を明らかにし，それらを数学的に解析したうえで統合化することにより，異なる通信システムを共通の視座から眺めることを可能にする「情報理論」という全く新しい科学分野を切り拓いた．

2. 情報と情報量

シャノンは，前節で述べた通信システムの一般モデルを数学的に解析するためには，情報源で作成されるメッセージの意味を問わずに「メッセージの価値」を定量的に把握する方法論が必要であると考えた．すなわち，メッセージを情報として捉え，意味内容に依存しない形で，その情報の量と価値を定める基本単位を決めることを出発点としたのである．そうした体系を構築するに際しては，どのような理論構築においてもそうであるように，

1) 常識的な直観と整合性をもつこと

2) 計算上，扱いやすいこと

という普遍的な課題を克服する必要があった．

「私は，今日，象を見た」(文章①)，「私は，今日，友人と一緒に動物園で，巨大な象を見た」(文章②)，という二つの文章を比較して見たとき，誰しも，文章②の方が文章①よりも情報量は大きいと思うに違いない．それは，文章①の提供する内容に加えて，文章②では，「友人と一緒に」「動物園で」「巨大な象」という新たな情報が付加されているからである．このような情報量に関する経験的理解を「意味を引き剥がして」捉えるべく，シャノンは「メッセージとは選択である」という観点を着想した．上述した二つの文章の例を，この観点から整理してみよう．

日本語を用いて生成される文章全体の集合を Ω とし，文章①と内容的に矛盾しない Ω の部分集合を F_1，文章②に関するそれを F_2 とする．「私は，今日，独りで町中を歩いていると，痩せた象を見た」という文章は，文章①の内容と矛盾しないので F_1 に属するが，文章②の内容とは矛盾するので F_2 の要素ではない．すなわち，限定することは選択肢を狭めるのであり，$F_1 \supset F_2$ が成立する．有限集合 A の要素の数を $|A|$ で表わすと，$P_{F_1} = |F_1| \,/\, |\Omega| > |F_2| \,/\, |\Omega| = P_{F_2}$ となり，ここから，「確率が小さく，より珍しい内容を教えてくれる方が，情報量は大きい」という原則を，意味を引き剥がす形で導くことができる．

もう少し話をわかりやすくするために，公平なサイコロを 2 回振った結果の和を当てるゲームを考えてみる．起こり得る場合の集合は $N = \{2, 3, \cdots, 12\}$ であり，それぞれの生起確率は，表 1 で与えられる．ここで，二つの情報が与えられたとしよう．

A：和は 4 以下である

B：1 回目に出た目は偶数である

この情報の下で，表 1 は表 2 のように変わる．常識的な直観に拠れば，明らかに A の方が情報量の大きい情報であり，確率で見ると，$P(A) = 1/6, P(B) = 1/2$ である．この例に隠される普遍性を見抜き，シャノンは情報量の満たすべき最初の性質として，次の規則を採用した．

規則 2.1：発生する確率が低い（珍しい）内容を伝える情報の方が，高い内容を伝える情報よりも，情報量は大きい

ここで，B に対して，さらに次の追加情報 C が与えられたとしよう．

C：2 回目に出た目は奇数である

すると，表は表 3 のように変わる．

規則 2.1 に照らすと，情報 BC は明らかに情報 C よりも情報量は大きい．確率で見ると，

$$P\,(BC) = 1/2 \times 1/2 = 1/4 < 1/2 = P\,(C)$$

である．この例に見られるように，独立な二つの事象の確率はそれぞれの事象の確率の積となる．人間は，重さや光量の強度を表す物理量が 2 乗倍されると 2 倍に感じる感覚をもっており，情報量の増減に対しても同様の感覚をもつと考えることが自然である．そこで，次の第 2 の規則が採用されることになる．

規則 2.2：確率の積で決まる情報量の増減を，人間は和・差として知覚する

規則 2.1 と規則 2.2 に基づき，シャノンは自己情報量を以下のように定義した．

定義 2.1　自己情報量

確率 $P\,(A)$ で生起する事象 A を報せる情報に対し，その情報のもつ自己情報量を $i\,(A) = -\log_2 P(A)$ と定める．

マイナスの符号があるのは，1 より小さい数の対数が負となるので，情報量を正の数として定義するためである．対数を取ることにより，定義 2.1 は「確実に生起する事象を報せる情報の情報量は 0 である」という直感と矛盾しない．対数の底として 2 を採用する理由は，不確実性に関して最も単純な形である「等確率をもつ二者択一の問題」に対し，「その解答のもつ自己情報量を 1 とする」ということにある．たとえば，等確率で表 (Head) か裏 (Tail) が出るコインを投げることを考え，表が出る事象を H とすると，その自己情報量は，$i\,(H) = -\log_2 P\,(H) = -\log_2 1/2 = -\log_2 2^{-1} = \log_2 2 = 1$ となる．これは，コンピュータによる 2 進法演算に適した表記法となっている．

シャノンは，論文執筆に際して周囲の協力を求めることをほとんどしなかったが，自己情報量の基本単位の命名に関しては例外であった．ベル研究所の食堂のテーブルに集まった同僚に，何か覚えやすい名前はないかと相談し，「ビニット」や「ビジット」という候補が検討の末に却下された後，最後に，当時ベル研究所に勤務しその後プリンストン大学へ移ったジョン・ワイルダー・テューキー (John Wilder Tukey) が提案した「binary digit（2 進数）を短縮してビット (bit) と呼ぶ案」が採用された．

前出のサイコロの目の和の例題における情報 B，情報 C と情報 BC に定義 2.1 を適用すると，それぞれの自己情報量は，

$$\begin{aligned} i\,(B) &= i\,(C) = -\log_2 1/2 = \log_2 2 = 1 \\ i\,(BC) &= -\log_2 P(BC) = -\log_2\{P(B)\times P(C)\} \\ &= -\log_2 P(B) - \log_2 P(C) = i\,(B) + i\,(C) = 2 \end{aligned}$$

となり，自己情報量は $i(B) = i(C) = 1$ から $i\,(BC) = i\,(B) + i\,(C) = 2$ へ和として増大している．

3.　情報エントロピー (entropy) と情報の価値

シャノンは，集合とその補集合の確率の和は常に 1 となるので，情報を議論するのに，単一の事象を取り上げて議論するだけでは不十分であることも看破していた．ある事象に追加情報が与えられると，特定の度合によってその事象の自己情報量は増加する．しかし，同時に，その補集合に対応する事象は，自己情報量を減少させることになる．したがって，「全体的な不確実性」を指標で捉えるためには，自己情報量の平均を考える必要があり，シャノンはこの概念を情報エントロピーとして定義した．

全体集合 Ω の部分集合族 $X = \{A_i \subset \Omega\colon i = 1,\cdots,K\}$ が，$A_i \neq \varnothing$，$A_i \cap A_j = \varnothing$ $(i \neq j)$，$\Omega = \cup_{j=1}^{K} A_j$ を満たすとき，X を Ω の直和分割と呼ぶ．このとき，それぞれの部分集合の確率の和は $\sum_{j=1}^{K} P\,(A_j) = 1$ となる．

定義 3.1　情報エントロピー

Ω の直和分割 $X = \{A_i \subset \Omega\colon i = 1,\cdots,K\}$ に対し，A_i の自己情報量の平均を X の情報エントロピーと呼び，

$$I_E\,(X) = -\sum_{j=1}^{K} P\,(A_j) \log_2 P\,(A_j) \qquad (3.1)$$

と表わす．

情報エントロピーは自己情報量の平均なので，その基本単位は自己情報量と同じ「ビット (bit)」であることを注意しておく．

直和分割 X の情報エントロピーは，直和分割を構成する事象群のどれが実際に起こるのかに関する「わからなさ」や「不確実性」を表していると考えられる．公平なコインを 1 回投げ，表が出るか裏が出るかを当てる最も単純なゲームに戻ると，結果の集合は $X = \{H, T\}$ で

あり，それぞれの基本事象が直和分割を構成するので，

$$I_E(X)=-\frac{1}{2}\log_2\frac{1}{2}-\left(1-\frac{1}{2}\right)\log_2\left(1-\frac{1}{2}\right)=1 \quad (3.2)$$

となり，この $X=\{H,T\}, P(H)=P(T)=1/2$ の場合が情報エントロピーの基本単位 1 bit を代表している．

問題をもう少し複雑にして，公平なサイコロを 1 回振った結果を当てるゲームを考える．起こり得る場合の集合は $N=\{1,2,\cdots,6\}$ であり，直和分割を構成するそれぞれの基本事象の生起確率は 1/6 である．このとき，N の情報エントロピーは，

$$I_E(N)=-\sum_{j=1}^{6}\frac{1}{6}\log_2\frac{1}{6}=\log_2 6=2.585 \quad (3.3)$$

となる．ここで，「A：出た目は偶数である」という情報が与えられたとしよう．このとき，起こり得る場合の集合は $N_{|A}=\{2,4,6\}$，その情報エントロピーは

$$I_E\left(N_{|A}\right)=-\sum_{j=1}^{3}\frac{1}{3}\log_2\frac{1}{3}=\log_2 3=1.585 \quad (3.4)$$

で与えられる．追加情報により，情報エントロピーは $I_E(N)=2.585$ から $I_E\left(N_{|A}\right)=1.585$ へと減少した．すなわち，結果に対する不確実性が，情報 A によって低まったということになる．この差をもたらした情報 A の価値が，情報エントロピーの差として把握されることになる．より一般的に，Ω に関する二つの直和分割 X_1, X_2 が与えられたとき，X_1 に対する X_2 のもつ情報の相対的価値 $VD(X_1,X_2)$ を，以下のように定義する．

定義 3.2　情報の価値
Ω の直和分割を X_1, X_2 したとき，X_1 のもつ情報に対する X_2 のもつ情報の価値を

$$VD(X_1,X_2)=I_E(X_1)-I_E(X_2) \quad (3.5)$$

と表わす．

4.　エントロピーの歴史的背景

ここで，エントロピーという用語に関する歴史的背景について，簡単に触れておく．1824 年，フランスの物理学者ニコラ・レオナール・サディ・カルノー (Nicolas Léonard Sadi Carnot) [57] は，熱量は保存され，熱が高温から低温へと移動するときに仕事が発生するという理論を発表した．熱源から仕事を生み出す熱機関の最大効率を生み出すためには可逆的な過程が必要であると考え，「①等温膨張 ⇒ ②断熱膨張 ⇒ ③等温圧縮 ⇒ ④断熱圧縮 ⇒ ①等温膨張」という可逆過程を考案した．この可逆過程はカルノーサイクルと呼ばれ，その後，熱力学の中心テーマとして重要な役割を果たした．

カルノーサイクルが発表されると，多くの研究者がその熱力学的解明に取り組むようになった．電気エネルギーが熱エネルギーに変わることに注目し，1840 年，ジュールの法則を確立したイギリスの物理学者ジェームズ・プレスコット・ジュール (James Prescott Joule) [58] や，熱と仕事が相互に変換可能であることを発見し，1842 年，エネルギー保存則を確立したドイツの物理学者ユリウス・ロベルト・フォン・マイヤー (Julius Robert von Mayer) [59]，ジュールが行ってきた熱の仕事当量に関する実験をもとに，1847 年，熱力学第一法則を導き出したドイツの物理学者で生理学者でもあったヘルマン・ルートヴィヒ・フェルディナント・フォン・ヘルムホルツ (Hermann Ludwig Ferdinand von Helmholtz) [60] など，カルノーサイクルを巡って活発な研究が行われた．

これらの先行研究を受け，熱力学の分野で大きな足跡を残したのがドイツの物理学者ルドルフ・ユリウス・エマヌエル・クラウジウス (Rudolf Julius Emmanuel Clausius) [61] である．カルノーサイクルにおける熱の出入りを研究し，1865 年，内部エネルギーの変化量 dU は，内部的に成される仕事の変化量 dQ と外部に成される仕事の変化量 dw に分解できることを示し，エネルギー保存則である熱力学第一法則を

$$dU=dQ-dw \quad (4.1)$$

という式で定式化することに成功した．さらに，「熱は常に温度差をなくす傾向を示し，したがって常に高温物体から低温物体へと移動する」という熱力学第二法則を導出した．特に，カルノーサイクルのような可逆過程においては，単位温度当りの仕事量変化はサイクルが閉じた際にゼロとなる，すなわち

$$\oint\frac{dQ}{T}=0 \quad (4.2)$$

という式によって熱力学第二法則を定式化した．

1865 年の論文で，クラウジウスは，可逆性を有するカルノーサイクルを 1 周した時点でその積分値が 0 になる dQ/T という量に着目し，

$$dS=\frac{dQ}{T} \quad (4.3)$$

という定義を導入，その積分値 S をエントロピーと名付けた．この時点で原子論に基づく統計熱力学は確立されておらず，エントロピーの変化量 dS が熱機関の可逆性の指標として導入されたことになる．エントロピーの概念は，当時の多くの研究者により反論されたが，電磁気学の最も偉大な研究者の一人とされたイギリスの理論物理学者ジェームズ・クラーク・マクスウェル (James Clerk Maxwell) [62] の強力な支持によって，学界に場所を占めることができた．

エントロピーが「乱雑さ」を表わす物理量であることを証明したのは，統計力学の確立に大きな役割を果たしたオーストリアの物理学者ルートヴィッヒ・エードゥアルト・ボルツマン (Ludwig Eduard Boltzmann) [63] である．クラウジウスの研究成果を受け，1872 年，まず，熱力学第二法則からエントロピー増大の法則を導いた [64]．簡単のため，温度 T_1 の吸熱源から Q_1 の熱を得て，温度 T_2 の排熱源に Q_2 の熱を捨てる熱機関（サイクル）を考える．この熱機関が外部に行う仕事は式 (4.1) のエネルギー保存則から $W = Q_1 - Q_2$ であり，熱機関の熱効率 η は

$$\eta = \frac{W}{Q_1} = 1 - \frac{Q_2}{Q_1} \tag{4.4}$$

によって与えられる．カルノーの定理 [65] によれば，可逆でない熱機関の熱効率には上限が存在し，

$$\eta < \eta_{max} = 1 - \frac{T_2}{T_1} \tag{4.5}$$

が成立する．式 (4.4) と (4.5) から，

$$\frac{Q_1}{T_1} < \frac{Q_2}{T_2} \tag{4.6}$$

を得る．すなわち，可逆でない過程において高熱源で熱を得た後，低熱源でその熱を捨てると，エントロピーは増大する．上記の議論を有限個の熱源に拡張し，さらに熱源の個数を無限大にする極限操作を通して積分表現に変換すると，一般的なエントロピー増大の法則を証明することができる．

統計力学において，系の巨視的な状態は，その系のもつエネルギー，体積，物質量などの巨視的な物理量の組によって決定される．しかし，それらの巨視的な物理量を定めたとしても，系の微視的状態は完全には定まらず，いくつかの状態を取り得る．巨視的な拘束条件の下で，実現可能な微視的状態の数を状態数と呼ぶ．1877 年，状態数を W としたとき，ボルツマンはエントロピー S が次式で与えられることを証明した．

$$S = k \log W \tag{4.7}$$

式 (4.7) はボルツマンの関係式と呼ばれ，エントロピーと系の取り得る状態数の対数値を関係付ける比例定数 k は，ボルツマン定数と呼ばれている．

ボルツマンの関係式が意味するのは，微視的状態の可能な数が増えるに連れてエントロピーは対数関数にしたがって増大し，逆に，微視的状態が確定的 $(W = 1)$ であれば，エントロピーは $S = 0$ となるということである．巨視的な情報しか知り得ないとき，微視的状態の可能な数が増えるということは，それだけ微視的世界に関する情報が欠如していると理解することができる．この意味で，ボルツマンの関係式は，エントロピーが「微視的世界の乱雑さ」を表す指標としての意味をもっていることを示している．

微視的状態 $w_i, i = 1, \cdots, N$ が確率 $P(w_i)$ で実現する場合，ボルツマンの関係式を一般化し，改めてエントロピーをその平均値として定義したのは，アメリカの数学者であり理論物理化学者でもあるジョサイア・ウィラード・ギブズ (Josiah Willard Gibbs) [66] である．具体的には，エントロピーを

$$S = -k \sum_{i=1}^{N} P(w_i) \log_e P(w_i) \tag{4.8}$$

と定義した．この式はギブズエントロピーとも呼ばれ，式 (3.1) で与えられた情報エントロピーと定数倍の違いを除いて一致する．

シャノンは，本人の言によれば統計熱力学におけるギブズエントロピーの概念を知らなかったが，「状態あるいは情報のもつ不確実性を計量化する」という目的意識は同じだったのであり，全く別の経路からこれだけ同等の表現に到達している事実は，「自然界における真理の有り様」や「人間が理解できることの本質」について，深く考えさせられる．

5. 情報エントロピーの基本的性質

情報エントロピーの性質を理解するため，まず，式 (3.1) の各項を構成する関数

$$f(p) = -p \times \log_2 p \tag{5.1}$$

を考える．$f(1) = 0$ であり，$f(0)$ は定義されないが，ロピタルの規則により，$p \to 0$ で $f(p) \to 0$ より，$f(0) = 0$ と定義する．

ネピア数を $e = 2.718\cdots$ とすると，$0 < p < 1$ で $f(p) > 0$ であり，$f'(p) = -\log_2 p - (1/\log_e 2)$ より，

$$f'(p) = -\log_2 (p \cdot e) \tag{5.2}$$

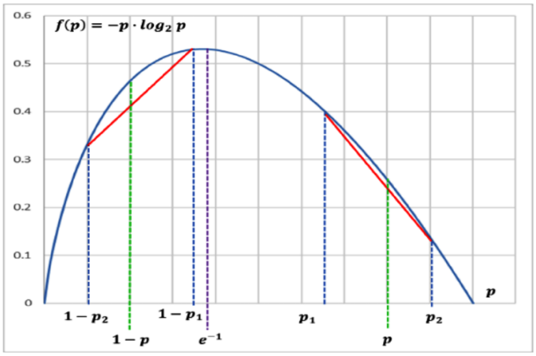


図 2　$f(p) = -p \cdot \log_2 p$ のグラフ

を得る．明らかに $p = e^{-1}$ で $f'(p) = 0$ となる．さらに，式 (5.2) より

$$f''(p) = -(1/p) \cdot \log_e 2 < 0 \tag{5.3}$$

であり，$f(p)$ は $p = e^{-1}$ で最大値を取る狭義の凹関数となることがわかる．

狭義の凹関数の性質として，$0 < p_1 < p_2 < 1$, $\alpha, \beta > 0$, $\alpha + \beta = 1$, $p = \alpha p_1 + \beta p_2$ であれば，

$$\alpha f(p_1) + \beta f(p_2) < f(p)$$
$$\alpha f(1-p_1) + \beta f(1-p_2) < f(1-p)$$

が成立する．これは，図 2 に示す $f(p)$ のグラフにおいて，上の二つの式の左辺を示す斜線よりも関数 $f(p)$ が上側にあることを意味する．

次の定理で，情報エントロピーの重要な性質の一つである，均等確率最大化の法則を示す．

定理 5.1

定義 3.1 で，$p_j = P(A_j)$，$j = 1, \cdots, K$ とすると，X の情報エントロピー $I_E(X)$ は $\boldsymbol{p} = [p_1, \cdots, p_K]^T$ に関して狭義の凹関数であり，制約条件 $0 < p_j < 1$，$j = 1, \cdots, K$，$\sum_{j=1}^{K} p_j = 1$，の下では，$p_j^* = K^{-1}$，$j = 1, \cdots, K$ で最大値 $\log_2 K$ を取る．

[証明]

ラグランジュ関数 $L(\boldsymbol{p}, \lambda) = I_E(X) - \lambda\left(\sum_{j=1}^{K} p_j - 1\right)$ に対し，その停留点を求める．式 (3.1), (5.1), (5.2) より，$j = 1, \cdots, K$ に対して

$$\frac{\partial}{\partial p_j} L(\boldsymbol{p}, \lambda) = -\log_2 p_j - \log_2 e - \lambda = 0$$
$$\frac{\partial}{\partial \lambda} L(\boldsymbol{p}, \lambda) = -\left(\sum_{j=1}^{K} p_j - 1\right) = 0$$

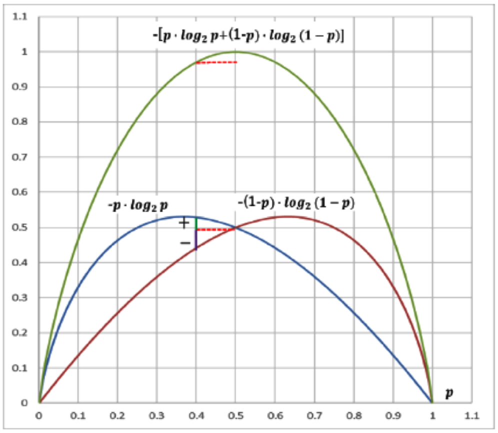


図 3　$K = 2$ の場合の情報エントロピー

が成立する．これを解くと，$p_j^* = K^{-1}$，$j = 1, \cdots, K$，$\lambda^* = \log_2 K - \log_2 e$ を得る．また，式 (5.3) より，

$$\frac{\partial^2}{\partial p_i \partial p_j} L(\boldsymbol{p}, \lambda) = \begin{cases} -\frac{1}{p_j} \cdot \log_e 2 < 0 & i = j \\ 0 & i \neq j \end{cases}$$

が成立し，$L(\boldsymbol{p}, \lambda)$ のヘッセ行列は対角行列で，その対角要素がすべて負となることがわかる．よって，$L(\boldsymbol{p}, \lambda)$ は $\boldsymbol{p}$ に関して狭義の凹関数であり，その停留点が制約条件を満たすことから，そこで最大値 $\log_2 K$ を取ることが証明された．□

定理 5.1 は，複数の事象が等確率をもつとき，その平均的な不確実性は最大になり，そこから，一つの事象の確率を減少させると，その事象のもつ情報量は増加するが，その増分はほかの事象の確率和の増加による平均情報量の減少を下回ることを意味する．この関係を，$K = 2$ で確率を $p = 0.5$ から $p = 0.4$ へ減少させた場合について，図 3 に示す．

次に，事象を分割することによって情報エントロピーは減少することを証明する．

定理 5.2

全体集合 $X = \{x_m : m = 1, \cdots, N\}$ に対し，ある条件を満たす n 個の要素からなる部分集合 A_1 とそれ以外の要素の集合 A_2 で構成される直和分割 $X = A_1 \bigcup A_2$ を考える．ここで，新たな直和分割 $X = X_1 \cup X_2$ を導入し，条件を満たす要素も，それぞれに異なる割合で分割されるとする．分割前の条件成立の有無に関する情報エントロピーを $I_E(X)$，分割後のそれを $I_E(X_1, X_2)$

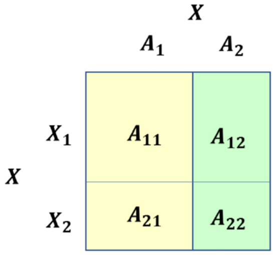


図 4　$\boldsymbol{X}$ の 2 通りの直和分割

とすると，$I_E(X) > I_E(X_1, X_2)$ が成立する．

[証明]

A_1 の純度を $p = n/N$ とすると，X の条件成立の有無に関する情報エントロピーは，

$$I_E(X) = -[p \log_2 p + (1-p) \log_2 (1-p)] \quad (5.4)$$

で与えられる．$p = 0$ or 1 の場合，式 (5.4) は数学的に定義されないが，前回の議論と同様，極限操作によってその値を 0 と定義する．

新たに，$X = X_1 \cup X_2$ と直和分割されたとき，$A_{ij} = X_i \cap A_j$，$i, j = 1, 2$ と定義すると，図 4 に示すように，$X_i = A_{i1} \cup A_{i2}$，$i = 1, 2$ も直和分割となる．このとき，$|X_i| = N_i$，$|A_{i1}| = n_i$，$p_i = n_i/N_i$，$i = 1, 2$ とおくと，X_1 と X_2 の情報エントロピーは，

$$I_E(X_1) = -[p_1 \log_2 p_1 + (1-p_1) \log_2 (1-p_1)] \quad (5.5)$$

$$I_E(X_2) = -[p_2 \log_2 p_2 + (1-p_2) \log_2 (1-p_2)] \quad (5.6)$$

となる．さらに，$\alpha = N_1/N$，$\beta = N_2/N$ とおくと，分割後の全体の情報エントロピーは，

$$I_E(X_1, X_2) = \alpha \times I_E(X_1) + \beta \times I_E(X_2) \quad (5.7)$$

となる．

異なる割合で分割されているという前提から，一般性を失うことなく $p_1 < p_2$ と仮定する．すると実数の不等式より，

$$p_1 = \frac{n_1}{N_1} < \frac{n_1 + n_2}{N_1 + N_2} = \frac{n}{N} = p < \frac{n_2}{N_2} = p_2 \quad (5.8)$$

すなわち，$p_1 < p < p_2$ が成立する．これから，ただちに，$1 - p_1 > 1 - p > 1 - p_2$ も結論される．明らかに，$p = \alpha \times p_1 + \beta \times p_2$ であるから，図 2 に示したグラフで赤線部分より関数値が上側になる狭義の凹関数の性質より，

$$I_E(X) > I_E(X_1, X_2) \quad (5.9)$$

が成立し，定理が証明された．□

情報エントロピー $I_E(X)$ は，「ある要素が与えられた条件を満たしている」という命題の「わからなさ」あるいは「不確実性」を表していると考えられる．定理 5.2 は，分割によってこの「不確実性」が減少することを意味する．式 (5.9) の差を，

$$I_G(X, X_1, X_2) = I_E(X) - I_E(X_1, X_2) > 0 \quad (5.10)$$

と定義するとき，$I_G(X, X_1, X_2)$ は「X を X_1, X_2 に分割することによって得られる平均情報量の増分 (information gain)」と呼ばれる．この性質が，この連載の後半で議論する，判別分析の一つである決定木分析において大きな役割を果たすことになる．

6. 情報源の符号化

情報の量と価値を計る単位として情報エントロピーの概念を確立したシャノンが次に考えたのは，情報源からの出力系列を効率的に表現することを目的とする情報源符号化に関する理論体系の確立であった．ここでは，植松 [67] に沿って，その概略を紹介する．

定義 3.1 では Ω の直和分割 X に対して情報エントロピーを定義したが，簡便のため Ω 上の確率変数 X に対する定義に書き改めたうえで単にエントロピーと呼び，$H(X)$ で表わすことにする．

定義 6.1　確率変数 X のエントロピー

Ω 上の確率変数 X が確率分布 $P[X = x] = P(x), x \in \Omega$ をもつとき，X のエントロピーを，

$$H(X) = -\sum_{x \in \Omega} P(x) \log_2 P(x) \quad (6.1)$$

と表わす．

メッセージを符号化するためには，まず，メッセージを表わす記号の集合を考える必要がある．そうした記号の集合を $\mathcal{S}$ としよう．一般的には，$\mathcal{S}$ の要素となる記号はアルファベットや数字であるが，その他のシンボルを含んでもよく，$\mathcal{S}$ は有限集合の場合も無限集合の場合もある．重要なことは，ある環境で生成される可能性のあるメッセージのすべてを，$\mathcal{S}$ の要素を並べることによって表現できるということである．

情報源の符号化は，メッセージを通信に適した信号へ変換する前段階の手続きであり，何らかの仕組みで $\mathcal{S}$ の要素に有限な 0-1 の列を割り当てることに相当す

る．$B=\{0,1\}$ の要素を k 個並べてできる系列の集合は，B の k 回の直積集合 $B^k=B\times\cdots\times B$ として表わされる．$B^\infty=\bigcup_{k=1}^\infty B^k$ とすると，情報源の符号化は，記号集合 $\mathcal{S}$ から有限 0-1 系列集合 B^∞ への写像と考えられる．

定義 6.2　情報源の符号化と符号語長

(1) 写像 $C\colon\mathcal{S}\to B^\infty$ を情報符号（以下，単に符号）と呼ぶ．

(2) $C(x)$ を記号 x の符号語，その長さを符号語長と呼び，$l_C(x)=|C(x)|$ で表わす．

(3) C の逆像 $C^{-1}\colon B^\infty\to\mathcal{S}$ が存在するとき，$y=C(x)$ に対して $x=C^{-1}(y)$ を y の x への復元と呼ぶ．

記号 $x\in\mathcal{S}$ に対して，$l_C(x)=k\Longrightarrow C(x)\in B^k$ であることを注意しておく．

定義 6.3　平均符号長

$\mathcal{S}$ 上の確率変数 X に対する符号 C の平均符号語長 $L_C(X)$ を

$$L_C(X)=\sum_{x\in\mathcal{S}}P(x)\,l_C(x)\tag{6.2}$$

と定める．

情報源の生成するメッセージとは，$\mathcal{S}$ に属する記号を有限個並べて生成される系列である．B^∞ の場合と同様に，$\mathcal{S}^k=\mathcal{S}\times\cdots\times\mathcal{S}$ の和集合を $\mathcal{S}^\infty=\bigcup_{k=1}^\infty\mathcal{S}^k$ と定義する．

定義 6.4　メッセージとその符号化

(1) メッセージの集合を $\mathcal{M}=\{\boldsymbol{x}\colon\boldsymbol{x}\in\mathcal{S}^\infty\}$ で表わす．

(2) $\boldsymbol{x}=[x_1,\cdots,x_n]\in\mathcal{S}^n$ に対応する符号列を $C(\boldsymbol{x})=C(x_1)\cdots C(x_n)\in B^\infty$ と定義する．

定義 6.4 の $C(\boldsymbol{x})$ は，個別記号を符号化する $C(x_j)$ の拡張写像であり，隣接する符号間の境界を示す記号をもたないことに注意する必要がある．

メッセージの復元可能性は，符号写像の満たす条件によって定まる．

定義 6.5　メッセージの復元可能性

(1) 異なる記号には異なる符号値が対応する符号を非特異 (non-singular) 符号と呼ぶ．

(2) 符号の拡張 $C(\boldsymbol{x})$ が非特異である符号を一意復元可能符号と呼ぶ．

(3) どの符号値も，ほかの符号値の先頭部分と一致しない符号を語頭符号と呼ぶ．

表 4　符号の例

記号	符号 C_1	符号 C_2	符号 C_3	符号 C_4	符号 C_5
a	0	0	1	00	0
b	0	1	10	01	10
c	1	00	100	10	111
d	01	10	1000	11	110

表 4 に，$\mathcal{S}=\{a,b,c,d\}$ に対する符号の例を示す．

符号 C_1 は，記号 a,b に対する符号値がともに 0 であり，0 が受信されたとき a か b かの判定ができず，非特異符号ではない．

符号 C_2 はすべての符号値が異なっているので非特異符号であるが，

$$C_2(ba)=C_2(b)\,C_2(a)=10=C_2(d)$$

であり，一意に復元可能ではない．

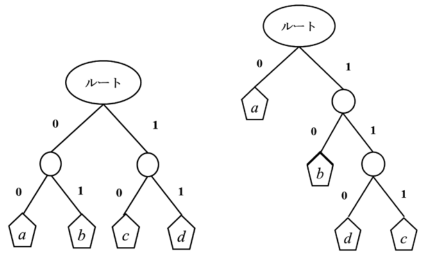
C_3 はすべての符号値の先頭が 1 であり，次に現れる 1 を読み込んだ段階で符号列を区切ることができるので，一意復元可能符号である．たとえば，

$$10100=C_3(b)?$$

とすると，受信者は 101 を読み込んだ時点で $C_3(b)=10$ を復元することができる．次の符号値は，メッセージがここで終了か次に 1 が現れれば $100=C_3(c)$，次に 0 が現れれば $1000=C_3(d)$ を判読し，以下，同様に継続していくことになる．しかし，記号 a の符号値 1 はその他の記号の先頭部分に含まれており，同様に，記号 b の符号値 10 は記号 c,d の符号値の先頭部分に，記号 c の符号値 100 は記号 d の記号値の先頭部分に，それぞれ含まれており，C_3 は語頭符号ではない．

符号 C_4，C_5 では，どの符号値もほかの符号値の先頭部分に含まれておらず，ともに語頭符号である．

語頭符号は，ルート・ノードから出発し，0-1 の分岐を繰り返すことによって，すべての符号をリーフ・ノードにもつ木構造として表現することが可能である．どの符号値も，ほかの符号値の先頭部分と一致しないという定義から，ある記号の符号値が中間ノードとして現れることはなく，すべての符号値がリーフ・ノードとして表現されることになる．これにより，次の手順を通して，語頭符号が一意復元可能符号であること



(a) C_4　　**(b)** C_5

図 5　語頭符号の木構造

表 5　C_4，C_5 の P_1，P_2，P_3 に対する平均符号長

記号	確率 P_1	確率 P_2	確率 P_3
a	1/4	1/2	1/8
b	1/4	1/4	1/8
c	1/4	1/8	1/4
d	1/4	1/8	1/2
L_{C_4}	2.000	2.000	2.000
L_{C_5}	2.250	1.750	2.625

がわかる．まず，ルート・ノードから出発し，与えられた符号値の順に木構造を辿り，リーフ・ノードに到達したら，そこにある符号値を復元してそれを記録し，ルート・ノードに戻ってメッセージが終了するまでこの操作を継続する．図 5(a) (b) に，表 4 の符号 C_4，C_5 を木構造で示しておく．

以上の議論より，符号化に際しては語頭符号を採用することが望ましいことがわかる．それでは，たとえば表 4 の符号 C_4，C_5 のように，複数の語頭符号の中から一つを選択する基準は何であろうか？ 連載第 2 回で，モールス信号に関し，アルフレッド・ルイス・ヴェイル (Alfred Lewis Vail) [68] が文字の使用頻度と符号の組み合わせに基づいて文字符号を決定し，通信効率を向上させたことを論じた．シャノンも，同様の基準を採用することが自然であると考えた．$\mathcal{S}$ 上で定義される語頭符号の集合を $HC(\mathcal{S})$ とし，定義 6.3 の平均符号長を用いると，この選択問題は以下のように定式化される．

$$C_X^* = \arg\min_{C \in HC(\mathcal{S})} L_C(X) = \sum_{x \in \mathcal{S}} P(x)\, l_C(x) \quad (6.3)$$

表 5 で，表 4 の符号 C_4，C_5 について，三つの確率分布 P_1，P_2，P_3 に対するそれぞれの平均符号長を示しておく．確率分布 P_1，P_3 については C_4 の平均符号長が C_5 のそれより小さいが，P_2 は出現頻度の高い記号に短い符号を対応させているので，その関係は逆転する．

シャノンは，平均符号長をエントロピーと関連付け，第一基本定理（情報源符号化定理）として，「符号化という工学的操作の効率性の上限が，情報源の確率分布から定まるエントロピーという量で与えられる」ことを示した．言葉を換えると，ある環境でメッセージを生成する際の記号出現の確率分布が与えられたとき，符号化を工夫することで通信量を小さくしようとする努力は，その確率分布のエントロピーで定まる限界に直面するということである．次回は，この興味深い定理を解説する．（この項，なお続く）

参考文献

[1] H. Goldstine, The Computer from Pascal to von Neumann, Princeton University Press, 1972.（末包良太，米口肇，犬伏茂之訳，『復刊　計算機の歴史―パスカルからノイマンまで―』，共立出版，2016.）

[2] S. McCartney, The Triumphs and Tragedies of the World's First Computer, Walker, 1999.（日暮雅通訳，『エニアック―世界最初のコンピュータ開発秘話―』，パーソナルメディア，2001.）

[3] 坂村健，『痛快！　コンピュータ学』，集英社，1999（文庫版 2002）．

[4] 竹内伸，『実物でたどるコンピュータの歴史―石ころからリンゴへ―』，東京理科大学出版センター（編），東京書籍，2012.

[5] 小田徹，『コンピュータ開発のはてしない物語―起源から驚きの近未来まで―』，技術評論社，2016.

[6] Wikipedia, Francois Viéte, https://enwikipedia.org/wiki/Francois_Viéte（2021 年 12 月 14 日閲覧）

[7] E. T. Bell, *Men of Mathematics Volume 1*, Simon & Schuster, 1937.（田中勇・銀林浩訳，『数学をつくった人びと上』，東京図書，1976.）

[8] Wikipedia, René Descartes, https://en.wikipedia.org/wiki/René_Descartes（2021 年 12 月 21 日閲覧）

[9] E. T. Bell, *Men of Mathematics Volume 2*, Simon & Schuster, 1937.（田中勇・銀林浩訳，『数学をつくった人びと下』，東京図書，1976.）

[10] Wikipedia, George Boole, https://en.wikipedia.org/wiki/George_Boole（2021 年 12 月 14 日閲覧）

[11] P. J. Nahin, *The Logician and the Engineer: How George Boole and Claude Shannon Created the Information Age*, Princeton University Press, 2012,（松浦俊輔訳，『0 と 1 の話―ブール代数とシャノン理論―』，青土社，2013.）

[12] J. Soni and R. Goodman, *A Mind at Play: How Claude Shannon Invented the Information Age*, Simon & Schuster, 2017,（小坂恵理訳，『クロード・シャノン―情報時代を発明した男―』，筑摩書房，2019.）

[13] Wikipedia, Claude Shannon, https://en.wikipedia.org/wiki/Claude_Shannon（2021 年 12 月 20 日閲覧）

[14] Wikipedia, Alan Turing, https://en.wikipedia.org/wiki/Alan_Turing（2021 年 12 月 20 日閲覧）

[15] B. J. Copeland,*Turing: Pioneer of the Information Age*, Oxford University Press, 2012,（服部桂訳，『チューリング―情報時代のパイオニア―』，NTT 出版, 2013.）

[16] A. Hodges, *Alan Turing: The Enigma*, Princeton University Press, 2014,（土屋俊・土屋希和子訳，『エニグマ―アラン・チューリング伝―』，勁草書房，2015.）

[17] 高岡詠子，『チューリングの計算理論入門―チューリン

グ・マシンからコンピュータへ―』，講談社，2014.
[18] Wikipedia, Galileo Galilei, https://en.wikipedia.org/wiki/Galileo_Galilei（2021 年 12 月 21 日閲覧）
[19] Wikipedia, Nicolaus Copernicus, https://en.wikipedia.org/wiki/Nicolaus_Copernicus（2021 年 12 月 21 日閲覧）
[20] Wikipedia, Marin Mersenne, https://en.wikipedia.org/wiki/Marin_Mersenne（2021 年 12 月 21 日閲覧）
[21] Wikipedia, Isaac Beeckman, https://en.wikipedia.org/wiki/Isaac_Beeckman（2022 年 1 月 2 日閲覧）
[22] Wikipedia, Adrien Baillet, https://en.wikipedia.org/wiki/Adrien_Baillet（2022 年 1 月 2 日閲覧）
[23] Wikipedia, Elisabeth of the Palatinate, https://en.wikipedia.org/wiki/Elisabeth_of_the_Palatinate（2022 年 1 月 2 日閲覧）
[24] Wikipedia，エリーザベト・フォン・デア・プファルツ (1618-1680), https://ja.wikipedia.org/wiki/エリーザベト・フォン・デア・プファルツ (1618-1680)（2022 年 1 月 2 日閲覧）
[25] 有賀暢迪，"合理力学の一例としての衝突理論　1720–1730 年"，科学哲学科学史研究，**6**, pp. 17–37, 2012.
[26] Wikipedia，ソディの 6 球連鎖，https://ja.wikipedia.org/wiki/ソディの 6 球連鎖（2022 年 1 月 4 日閲覧）
[27] Wikipedia, Thorold Gosset, https://en.wikipedia.org/wiki/Thorold_Gosset（2022 年 1 月 4 日閲覧）
[28] 寒川町ガイド，https://samukawaguide.blogspot.com/2019/12/6.html（2022 年 1 月 4 日閲覧）
[29] Wikipedia, Gottfried Wilhelm Leibniz, https://en.wikipedia.org/wiki/Gottf_ried_Wilhelm_Leibniz（2022 年 1 月 4 日閲覧）
[30] Wikipedia, Christina, Queen of Sweden, https://en.wikipedia.org/wiki/Christina,_Queen_of_Sweden（2022 年 1 月 4 日閲覧）
[31] Wikipedia, Isaac Newton, https://en.wikipedia.org/wiki/Isaac_Newton（2022 年 1 月 4 日閲覧）
[32] 向井茂，"不変式の話，" 数学セミナー連載，2005 年 12 月号，2006 年 1，2，4 月号.
[33] 日本医学会ホームページ，https://jams.med.or.jp/news/013.html（2022 年 2 月 4 日閲覧）
[34] Wikipedia, Vannevar Bush, https://en.wikipedia.org/wiki/Vannevar_Bush（2022 年 2 月 25 日閲覧）
[35] Britanica, William-Thomson-Baron-Kelvin, https://www.britannica.com/biography/William-Thomson!-Baron-Kelvin（2022 年 3 月 6 日閲覧）
[36] Wikipedia, Hannibal Ford, https://en.wikipedia.org/wiki/Hannibal_Ford（2022 年 3 月 6 日閲覧）
[37] Wikipedia, Joseph Fourier, https://en.wikipedia.org/wiki/Joseph_Fourier（2022 年 3 月 6 日閲覧）
[38] Wikipedia，ユトランド沖海戦，https://ja.wikipedia.org/wiki/ユトランド沖海戦（2022 年 3 月 6 日閲覧）
[39] Wikipedia, Mark I Fire Control Computer, https://en.wikipedia.org/wiki/Mark_I_Fire_Control_Computer（2022 年 3 月 7 日閲覧）
[40] Wikipedia, Bell Labs, https://en.wikipedia.org/wiki/Bell_Labs（2022 年 4 月 7 日閲覧）
[41] Wikipedia, Thornton Carle Fry, https://en.wikipedia.org/wiki/Thornton_Carle_Fry（2022 年 4 月 7 日閲覧）
[42] Wikipedia, Schön scandal, https://en.wikipedia.org/wiki/Schön_scandal（2022 年 4 月 7 日閲覧）
[43] Wikipedia，ヘンドリック・シェーン，https://ja.wikipedia.org/wiki/ヘンドリック・シェーン（2022 年 4 月 7 日閲覧）
[44] Wikipedia，ジョン・フォン・ノイマン，https://ja.wikipedia.org/wiki/ジョン・フォン・ノイマン（2022 年 4 月 29 日閲覧）
[45] Wikipedia，ヘルマン・ワイル，https://ja.wikipedia.org/wiki/ヘルマン・ワイル（2022 年 4 月 29 日閲覧）
[46] Wikipedia，第二次世界大戦，https://ja.wikipedia.org/wiki/第二次世界大戦（2022 年 5 月 31 日閲覧）
[47] Wikipedia，フランクリン・ルーズベルト，https://ja.wikipedia.org/wiki/フランクリン・ルーズベルト（2022 年 4 月 30 日閲覧）
[48] Wikipedia，ウォーレン・ウィーバー，https://ja.wikipedia.org/wiki/ウォーレン・ウィーバー（2022 年 5 月 3 日閲覧）
[49] Wikipedia，ジェイムス・コナント，https://ja.wikipedia.org/wiki/ジェイムス・コナント（2022 年 5 月 3 日閲覧）
[50] Wikipedia，ロバート・オッペンハイマー，https://ja.wikipedia.org/wiki/ロバート・オッペンハイマー（2022 年 5 月 3 日閲覧）
[51] Wikipedia, Homer Dudley, https://en.wikipedia.org/wiki/Homer_Dudley（2022 年 4 月 7 日閲覧）
[52] Wikipedia, SIGSALY, https://ja.wikipedia.org/wiki/SIGSALY（2022 年 5 月 31 日閲覧）
[53] Wikipedia，ワンタイムパッド，https://ja.wikipedia.org/wiki/ワンタイムパッド（2022 年 5 月 3 日閲覧）
[54] 釜賀一夫，藤原邦樹，吉村昭，"座談会日本陸軍暗号はなぜ破られなかったか，" 歴史と人物―太平洋戦争シリーズ―，昭和 60 年冬号，1985.
[55] Wikipedia, Harry Nyquist, https://en.wikipedia.org/wiki/Harry_Nyquist（2022 年 5 月 7 日閲覧）
[56] Wikipedia, Ralph Hartley, https://en.wikipedia.org/wiki/Ralph_Hartley（2022 年 5 月 7 日閲覧）
[57] Wikipedia，ニコラ・レオナール・サディ・カルノー，https://ja.wikipedia.org/wiki/ニコラ・レオナール・サディ・カルノー（2022 年 6 月 6 日閲覧）
[58] Wikipedia，ジェームズ・プレスコット・ジュール，https://ja.wikipedia.org/wiki/ジェームズ・プレスコット・ジュール（2022 年 6 月 6 日閲覧）
[59] Wikipedia，ユリウス・ロベルト・フォン・マイヤー，https://ja.wikipedia.org/wiki/ユリウス・ロベルト・フォン・マイヤー（2022 年 6 月 6 日閲覧）
[60] Wikipedia，ヘルマン・フォン・ヘルムホルツ，https://ja.wikipedia.org/wiki/ヘルマン・フォン・ヘルムホルツ（2022 年 6 月 6 日閲覧）
[61] Wikipedia，ルドルフ・クラウジウス，https://ja.wikipedia.org/wiki/ルドルフ・クラウジウス（2022 年 6 月 6 日閲覧）
[62] Wikipedia，ジェームズ・クラーク・マクスウェル，https://ja.wikipedia.org/wiki/ジェームズ・クラーク・マクスウェル（2022 年 6 月 6 日閲覧）
[63] Wikipedia，ルートヴィッヒ・ボルツマン，https://ja.wikipedia.org/wiki/ルートヴィッヒ・ボルツマン（2022 年 6 月 6 日閲覧）
[64] Wikipedia，エントロピー，https://ja.wikipedia.org/wiki/エントロピー（2022 年 6 月 6 日閲覧）
[65] Wikipedia，カルノーの定理_（熱力学），https://ja.wikipedia.org/wiki/カルノーの定理_(熱力学)（2022 年 6 月 7 日閲覧）
[66] Wikipedia，ウィラード・ギブズ，https://ja.wikipedia.org/wiki/ウィラード・ギブズ（2022 年 6 月 7 日閲覧）
[67] 植松友彦，『イラストで学ぶ情報理論の考え方』，講談社 (2012)
[68] Wikipedia，アルフレッド・ヴェイル，https://ja.wikipedia.org/wiki/アルフレッド・ヴェイル（2022 年 6 月 7 日閲覧）

