情報と符号の理論

長田 康敬 （琉球大学 工学部）

2024 年 5 月 17 日

目次

ま え が き

情報理論は, 文字通り情報の数学的基礎を研究する分野である．歴史的には, Shannon が 1948 年に情報理論の分野を提案したと考えられる．その後, 情報理論はコンピュータサイエンスのあらゆる分野で応用されている．現在の通信やマルチメディア技術などの発展においては情報理論が大きな役割を果たしている．したがって, 我々にとって情報理論の基礎を学ぶことは非常に意義があると考えられる．しかし, 情報理論を学習するためには確率論や代数学の知識が不可欠であり, 情報理論の内容を完全に理解するのは必ずしも容易でない．実際, 初心者向けの情報理論のテキストは余りないのが現状である．本書は, 情報理論を効率的に学ぶためのテキストとして執筆されたものである．

本書は, 対数と確率 (1 章), 情報と情報源 (2 章), 通信路と誤り訂正符号の基礎 (3-4 章), 符号 (5 章) の 4 部構成である．

第 1 章は, 序論で対数関数と確率論の復習である．特にベイズの定理は重要である．

第 2 章では情報量について解説する．情報理論の歴史と情報量の性質を説明した後, エントロピー (平均情報量) と相互情報量を扱う．また，情報源, マルコフ情報源, エルゴード情報源などを取り上げる．

第 3 章では, 雑音のない離散通信路を扱う．まず, 通信路の概念を導入し, 通信路容量について考え，ハフマン符号化と情報源符号化定理について解説する．

第 4 章では, 雑音のある場合の離散通信路における情報伝送の詳細を説明する．誤り訂正符号の解説と誤りの検出・訂正の原理を説明する．

第 5 章では誤り訂正符号の例として線形符号, 巡回符号, たたみ込み符号を紹介する．

本書は, 情報系学部の情報理論の半期授業または通年授業のテキストとして使用可能である．また, 自学用の教科書としても有用である．各章末には, 理解度をチェックするためのいくつかの練習問題を用意した．練習問題の解答および解説は巻末に示した．なお, 本書の内容は筆者らが琉球大学で行なった講義に基づく．

2022 年 2 月

長田 康敬

第 1 章

基礎知識

1.1　対数関数の復習

$$a^y = x \tag{1.1}$$

という関係があるとき，その逆関数を，

$$y = \log_a x \tag{1.2}$$

と書き，y を a を**底**とする x の**対数**という．但し[*1]，底 $a \neq 1$，かつ $a > 0$.
さらに，x を y の**真数**といい，上式の関係を対数関数という．

対数関数は指数関数の逆関数なので，ここで指数関数の性質をまとめておく．数値を入れて確かめてみると良い．

1. $a^m \times a^n = a^{m+n}$
2. $a^m \div a^n = a^{m-n}$
3. $(a^m)^n = a^{mn}$
4. $a^{-m} = \frac{1}{a^m}$
5. $a^{\frac{m}{n}} = \sqrt[n]{a^m}$
6. $a^0 = 1$

これより，対数の定義を示しておく．

[*1] $a = 1$ のとき，$1^y = x$ なので y に関係なく $x = 1$ となる．また，$a < 0$ のとき，$a^y = x$ は y の偶奇により振動する．

【定義 1.1】 対数の定義

$$y = log_a x \Longleftrightarrow x = a^y \tag{1.3}$$

$$(a \neq 1,\ \text{かつ}\ a > 0)$$

以下に対数関数の諸性質をまとめる．($a \neq 1$, $b \neq 1$ かつ $a > 0$, $b > 0$)

1. $\log_a(xy) = \log_a x + \log_a y$
2. $\log_a \frac{x}{y} = \log_a x - \log_a y$
3. $\log_a x^y = y \log_a x$
4. $\log_a \sqrt[n]{x} = \frac{1}{n} \log_a x$
5. $-\log_a \frac{1}{x} = \log_a x$
6. $\log_a b = \frac{\log_{10} b}{\log_{10} a} = \frac{\log_e b}{\log_e a}$

底が 10 である対数を**常用対数** (common logarithm) という．また，底がネイピア数 e(=2.718281828$\cdots$) である対数を**自然対数** (natural logarithm) といい，$\ln x$ と書くことがある．

情報理論においては**底が 2 の対数関数**が使用されるので，

$$\log_2 x = \frac{\log_{10} x}{\log_{10} 2} = \frac{\log_e x}{\log_e 2} \tag{1.4}$$

となり，$\log_{10} 2 \simeq 0.3$ や $\log_e 2 \simeq 0.7$ を記憶していると便利である．

1.2　確率と統計の復習

1.2.1　試行と事象

サイコロを振ってどの目がでるか，あるいはコインを投げて裏か表がでるかなど, 結果が確率的な行為を行なうことを**試行** (trial) という．試行を続けて行なった結果を**試行列**あるいは**試行系列**と呼ぼう．サイコロを 2 回振る試行列は $6 \times 6 = 36$ 通りあり，この全ての試行列の集合を**標本空間** (sampling space) という．標本空間の部分集合を**事象** (event) と呼び，特に単一の事象からなる場合を**単純事象**という．

事象には演算が定義できる．A, B を事象とすると，

1. $A \cup B$ 事象の和：A, B の少なくとも一方が起こる事象.

2. $A \cap B$ 事象の積：A と B の両方が同時に起こる事象.
3. A^c 事象の否定：A が生じない事象．**余事象** (complement) という.

$A \cap B = \phi$（空集合）のとき，事象 A, B は**互いに素**であるという.

【例題 1-1】サイコロを 1 回ふる試行では，標本空間 S は,

$$S = \{e_1, e_2, e_3, e_4, e_5, e_6\} \tag{1.5}$$

但し，e_i は目 i がでる事象とする．また，奇数の目がでる事象 S_{odd} と偶数の目がでる事象 S_{even} はそれぞれ,

$$S_{odd} = \{e_1, e_3, e_5\} \tag{1.6}$$
$$S_{even} = \{e_2, e_4, e_6\} \tag{1.7}$$

であり，1 の目がでる事象 $\{e_1\}$ は単純事象である．ここで,

$$S_{odd} \cap S_{even} = \phi \tag{1.8}$$

であり,

$$S_{odd}^c = S - S_{odd} = S_{even} \tag{1.9}$$

である.

1.2.2 確率

標本空間 $S = \{E_1, E_2, \cdots, E_n\}$ において，すべての事象は同じ起こりやすさであるとするとき，事象 E_i の起こる確率 (probability) は,

$$p(E_i) = \frac{1}{n}. \tag{1.10}$$

一般に，ある事象 A が標本空間 S の中の m 個 $(m \leq n)$ の要素からなるとき,

$$p(A) = \frac{m}{n}. \tag{1.11}$$

また，その余事象 A^c の起こる確率は,

$$p(A^c) = \frac{n-m}{n} = 1 - \frac{m}{n} = 1 - p(A) \tag{1.12}$$

ところで，互いに排反な n 個の事象からなる標本空間 $S = \{E_1, E_2, \cdots, E_n\}$ があり，各事象の生起確率を，$p(E_1), p(E_2), \cdots, p(E_n)$ とするとき，

$$\sum_{i=1}^{n} p(E_i) = 1 \tag{1.13}$$

が成立するとき，S を**完全事象系** (complete finite scheme) という．

ここで，ある試行のもとで得られた資料の総数を n, この資料の中で事象 E の起こった回数を r とするとき，$\dfrac{r}{n}$ を事象 E の起こる**相対度数**という．

ここで n が十分大きいとき, 相対度数がある一定の値 p にほぼ等しくなるとき, p を事象 E の起こる**統計的確率**（あるいは経験的確率）という．これに対して上で扱った確率を**数学的確率**（あるいは先験的確率）という．

1.2.3　期待値

起こり得る複数の事象 $e_1, e_2, \cdots, e_n$ のそれぞれの生起確率が $p_1, p_2, \cdots, p_n$ であるとき，これを**確率行列** P で次のように表現できる．

$$P = \begin{pmatrix} e_1 & e_2 & \cdots & e_n \\ p_1 & p_2 & \cdots & p_n \end{pmatrix} \tag{1.14}$$

ここで，この事象における**期待値** (expected value) は，次式で与えられる．

$$E(e_i) = \sum_{i=1}^{n} e_i \cdot p_i \tag{1.15}$$

【例題 1-2】 一個のサイコロを振ったときの期待値は,

$$\begin{aligned} E(X) &= 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} \\ &= \frac{1}{6} \cdot \sum_{i=1}^{6} i = \frac{21}{6} = 3.5 \end{aligned}$$

【例題 1-3】 硬貨を投げて表が出たら 50 ドル，裏が出たら 10 ドルもらえるときの期待値は,

$$E(X) = 50 \cdot \frac{1}{2} + 10 \cdot \frac{1}{2} = 25 + 5 = 30[\text{ドル}] \tag{1.16}$$

1.2.4 分散と標準偏差

期待値を

$$\mu = E(X) \tag{1.17}$$

と書くと，**分散** (variance) $V(X)$ は，

$$V(X) = \sum (x-\mu)^2 p(x) \tag{1.18}$$

と定義される．これは，期待値 μ と事象（あるいはサンプル）x との差の二乗の期待値

$$\begin{aligned} V(X) &= \sum (x-\mu)^2 \cdot p(x) \\ &= \sum (x^2 - 2x\mu + \mu^2) \cdot p(x) \\ &= \sum x^2 \cdot p(x) - 2\mu \underbrace{\sum x \cdot p(x)}_{\mu} + \mu^2 \underbrace{\sum p(x)}_{1} \\ &= \sum x^2 \cdot p(x) - \mu^2 \end{aligned} \tag{1.19}$$

であり，最後の式の第一項目は x^2 の期待値なので分散の式は次のようになる．

$$V(X) = E(X^2) - \{E(X)\}^2 = \sigma^2 \tag{1.20}$$

ここで，σ を**標準偏差** (standard deviation) という．

1.2.5 大数の法則

統計的確率に関連して，以下のような大数の法則が考えられている．

平均が μ，分散が σ^2 である母集団から n 個のサンプル $x_1, x_2, \cdots, x_n$ を取り出し，その平均を $\overline{x}$ とする．

$\overline{x}$ と母集団の平均 μ との差の絶対値 $|\overline{x} - \mu|$ が，ある任意の正の小さな値 ε より小さくなる確率 p_n というものを考える．つまり，

$$p_n = p(|\overline{x} - \mu| < \varepsilon) \tag{1.21}$$

とすると，

$$\lim_{n\to\infty} p_n = \lim_{n\to\infty} p(|\overline{x} - \mu| < \varepsilon) = 1 \quad (\varepsilon > 0) \tag{1.22}$$

これを**大数の (弱) 法則** (law of large numbers) という．確率変数が有限の分散 σ をもつとき，確率変数は分散 $V(\overline{x_n}) = \frac{\sigma^2}{n}$ を用いて，

$$p_n = p(|\overline{x} - \mu| < \varepsilon) \geq 1 - \frac{1}{\varepsilon^2} V(\overline{x_n}) = 1 - \frac{\sigma^2}{n\varepsilon^2} = 1 - \frac{p(1-p)}{n\varepsilon^2} \qquad (1.23)$$

が成立する．ここで，2 項分布の場合，$\sigma^2 = p(1-P)$ である．この式で n を大きくすると p_n は 1 に近づくというのが**大数の法則**である．[*2] [*3]

コラム　大数の法則の例

通常のサイコロを 1 回振るとき，1 の目が出る事象 A の確率 p は，$p = 1/6$ であり，また，$1 - p = 5/6$ である．
$p(|\frac{m}{n} - p| < \varepsilon) \geq 1 - \frac{p(1-p)}{n\varepsilon^2}$ より，
$p(|\frac{m}{n} - 1/6| < \varepsilon) \geq 1 - \frac{5/6 \cdot 1/6}{n\varepsilon^2} = 1 - \frac{5}{36n\varepsilon^2}$
ここで任意の小さな数 ε を $\varepsilon = 1/60$ とすると右辺は，$1 - \frac{5}{36n(1/60)^2} = 1 - \frac{500}{n}$ となる．これより，
(1) n = 6,000 回の試行の場合，
$p(|\frac{m}{6000} - 1/6| < 1/60)$ は，$1 - \frac{500}{6000} \simeq 0.9167$

(2) $n = 60,000$ 回の試行の場合，
$p(|\frac{m}{60000} - 1/6| < 1/60)$ は，$1 - \frac{500}{60000} \simeq 0.9917$

(3) $n = 600,000$ 回の試行の場合，
$p(|\frac{m}{600000} - 1/6| < 1/60)$ は，$1 - \frac{500}{600000} \simeq 0.9992$

(4) $n = 6,000,000$ 回の試行の場合，
$p(\frac{m}{6000000} - 1/6| < 1/60)$ は，$1 - \frac{500}{6000000} \simeq 0.9999$ となる．

n を大きくしていくと 1 の目が出る確率が 1/6 となる精度が 100% に近づいていく．

[*2] あるいは，$p(|\overline{x} - \mu| > \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2}$

[*3] 同様に，$n \to \infty$ のとき $\overline{x_n}$ は μ にほとんど確実に収束するので，$p(\lim_{n\to\infty} \overline{x_n} = \mu) = 1$．これを**大数の強法則**という．

1.2.6 確率の諸性質

事象 A, B が排反事象であるとき，次の**加法定理**が成り立つ．

$$p(A \cup B) = p(A) + p(B) \tag{1.24}$$

また，二つの事象 A, B が独立なとき，A と B が同時に起きる確立は，

$$p(A \cap B) = p(A) \cdot p(B) \tag{1.25}$$

であり，これを**乗法定理**という．

独立でない二つの事象 A, B を順次試行するときの確率を**条件付確率** (conditional probability) という．条件付確率は，

$$p(B|A) \tag{1.26}$$

と書き，その意味は，A **が起こったという条件の下で** B **が起こる確率**であり，$p_A(B)$ と書くこともある．

独立でない事象 A, B がともに起こる確率を**結合確率** (joint probability) といい，

$$p(A \cap B) = p(A) \cdot p(B|A) \tag{1.27}$$
$$= p(B) \cdot p(A|B) \tag{1.28}$$

が成り立つ．この式は大切である．$p(B|A)$ はあくまでも B の確率であり，”B の確率, 但し, A が起こったときの”と読むとよい．

1.2.7 全確率の定理とベイズの定理

n 個の排反事象 $\{A_1, A_2, \cdots, A_n\}$ があって，またこれらとは排反でない別の事象 B があるとする．

ここで，B の，$\{A_1, A_2, \cdots, A_n\}$ の下での条件付き確率が既知の場合，事象 B の起こる確率は，

$$\begin{aligned} p(B) &= p(A_1 \cap B) + p(A_2 \cap B) + \cdots + p(A_n \cap B) \\ &= p(A_1)p(B \mid A_1) + p(A_2)p(B \mid A_2) + \cdots + p(A_n)p(B \mid A_n) \\ &= \sum_{i=1}^{n} p(A_i)p(B \mid A_i) \end{aligned} \tag{1.29}$$

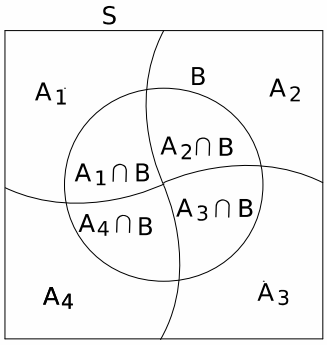

図 1.1　4 つの排反事象 A_1〜A_4 と，これらとは排反でない事象 B

となる．これを**全確率の定理** (theorem of whole probability) と呼ぶ．

また，$p(B \mid A_i)$ を**事前確率**といい，$p(A_i \mid B)$ を**事後確率**という．

ところで，

$$p(A_i \cap B) = p(B)p(A_i \mid B) = p(A_i)p(B \mid A_i) \tag{1.30}$$

より，

$$p(A_i \mid B) = \frac{p(A_i \cap B)}{p(B)} \tag{1.31}$$

$$= \frac{p(A_i)p(B \mid A_i)}{\sum_{i=1}^{n} p(A_i)p(B \mid A_i)} \tag{1.32}$$

となる．これは，事後確率を事前確率で表した式であり，**ベイズの定理** (Bayes' low) と呼ばれている．ベイズの定理と事前確率，事後確率は重要であるので何度も復習しよう．

【例題 1-4】事象 B を 38 度以上の体温の発熱とする．B の原因として次の三つの事象を考える．A_1：カゼ，A_2：胃腸炎，A_3：コロナ．

まず，各事象 A_1, A_2, A_3 の事前確率は次のような意味となる．

$p(A_1)$：カゼをひく確率．

$p(A_2)$：胃腸炎にかかる確率

$p(A_3)$：コロナに感染する確率

次に条件付確率は以下のような意味となる.
$p(B|A_1)$：カゼが原因で 38 度以上の発熱
$p(B|A_2)$：胃腸炎が原因で 38 度以上の発熱
$p(B|A_3)$：コロナに感染して 38 度以上の発熱

このときベイズの定理を用いると，次のような事後確率を求めることができる.
$p(A_1|B)$：38 度以上の発熱があるとき，カゼが原因である確率
$p(A_2|B)$：38 度以上の発熱があるとき，胃腸炎が原因である確率
$p(A_3|B)$：38 度以上の発熱があるとき，コロナが原因である確率

いま，例として以下のような値で事後確率を求めてみよう.
$p(A_1) = 50\%$, $p(A_2) = 30\%$, $p(A_3) = 20\%$
$p(B|A_1) = 30\%$, $p(B|A_2) = 20\%$, $p(B|A_3) = 50\%$
事後確率 $p(A_1|B)$ はベイズの定理より,

$$\begin{aligned} p(A_1|B) &= \frac{p(A_1) \cdot p(B|A_i)}{\Sigma_{i=1}^{3} p(A_i) \cdot p(B|A_i)} \\ &= \frac{0.5 \times 0.3}{0.5 \times 0.3 + 0.3 \times 0.2 + 0.2 \times 0.5} \\ &= \frac{0.15}{0.31} \simeq 0.484 = 48.4\% \end{aligned}$$

同様に,

$$\begin{aligned} p(A_2|B) &= \frac{0.3 \times 0.2}{0.31} \simeq 0.194 = 19.4\% \\ p(A_3|B) &= \frac{0.2 \times 0.5}{0.31} \simeq 0.323 = 32.3\% \end{aligned}$$

分母は共通であり，1 度計算すればよい.

章末問題

問 1.1　$y = 2^x$ の概略図を描け.

問 1.2　2 を底とする対数関数 $y = \log_2 x$ の概略図を描け.

問 1.3　$\log_2 4$ を求めよ.

問 1.4　$\log_2 8$ を求めよ.

問 1.5　$y = -\log_2 \frac{1}{16}$ を計算せよ.

問 1.6　$\log_2 3$ を求めよ．但し，$\log_{10} 3 \simeq 0.477$

問 1.7　2 つのサイコロを同時に振るとき，出た目の合計が 4 になる確率を求めよ.

問 1.8　1,000 本のくじの中に，1 等 1,000,000 円が 1 本，2 等 50,000 円が 10 本，3 等 10,000 円が 100 本含まれているとする（他はすべて空くじ）．このときくじを 1 本引くときの期待値を求めよ.

問 1.9　コンピュータが正しく起動しなくなった事象を B, その原因として,
A_1: コンピュータのメインメモリの不具合（発生確率 65%）
A_2: 電源部の故障（発生確率 25%）
A_3: CPU の熱暴走（発生確率 10%）
という事象があるとする．このとき，条件付き確率（事前確率）を,
$p(B \mid A_1) = 30\%$
$p(B \mid A_2) = 60\%$
$p(B \mid A_3) = 10\%$
とする．ベイズの定理を用いてこれらの事後確率を求めよ.