

情報量

— 情報理論への招待 —

博士(工学) 山本 宙 [著]

コロナ社

ま え が き

本書は情報理論におけるおもに情報量に関連する内容について大学の学部低学年を対象に 100 分の授業 14 回程度で学習できる内容を選択し、構成したものである。

大学で学ぶべき内容として、最新の技術を取り上げることは当然であるが、情報理論のような基礎学問の重要さもまた大きくなっている。若い学習者は最新の技術に興味をもち、それに必要なものだけを学習したいと考える傾向があるが、個別の最新技術に特化しすぎた学習を行っているとき技術の進歩があまりに速いために後にその知識が役に立たなくなる危険性がある。情報理論などの基礎理論は古い結果であるにもかかわらず、その知識はいまも変わらず利用され続けている。技術がどの方向に発展してゆくかわからないとき、自分で新しいものを考えなければならなくなったときに拠り所となるのは普遍的で基礎的な学問である。情報理論のような基礎学問は、ただちにににかの役に立つ、というよりはいろんな使い道があるのでまずは学習しておく、という専門教養というべき科目である。専門教養的な科目は進歩の速い分野でこそ重要なのである。

情報理論は情報を扱う分野の代表的な専門教養科目である。教科書もすでに多く出版されており、名著といわれる本も多い。それにもかかわらず本書を執筆したのは短い時間で完結する授業が要求されるようになった最近の環境の変化によるものである。かつては情報理論の授業は十分に時間を掛けて学習する科目として設定されていることが多く、既存の名著はこのような授業スタイルの時代に適したものであった。近年、学生の科目選択の自由度を高める目的で、週 1 コマ、半年間という最小限の期間で完結する授業が求められるようになった。既存の教科書で扱う範囲は広く深く、とてもこのコンパクトな授業スタイルではカバーできない。

専門教養を有意義に学ぶため、本書では定理の説明では結果と使用法を説明するだけでなく、証明の内容を厳密に、かつわかりやすく説明することに重きをおいた。短い授業時間でこれに対応するためには扱う内容の大胆な取捨選択が必要である。本書でかけた大きな制限としては以下のものがある。

- 情報源符号化定理をおもなテーマとし、通信路符号化定理は扱わない。
- 情報源を記憶のない情報源に限定する。
- 通信路の符号アルファベットを 2 元に限定する。

なお、本文は確率過程や微分積分に関する基本的な数学の知識があれば理解できるように記述した。

情報理論には情報源符号化定理と通信路符号化定理と呼ばれる二つの重要な定理がある。情報理論の教科書は両方を扱うことが一般的であるが、本書は情報源符号化定理のみを平易に説明することに特化した教科書として執筆した。扱うトピックは一意に復号可能な符号と瞬時に復号可能な符号、符号語長とクラフトの不等式、平均符号語長、コンパクト符号、ハフマンの符号化、シャノンの情報源符号化定理、エントロピーなどである。結果として情報量の概念を理解することを目的とする内容となったため、書名を“情報量 — 情報理論への招待 —”とした。本書で情報理論に興味をもった学生は参考文献で示した本格的な情報理論の教科書を手に取り、本書でおいた制限を取り払った一般的な理論を学習してほしい。本書がその助けになることを期待している。

最後に、情報理論の授業を履修してくれた学生諸君に深く感謝する。挿入されている図や例は普段の授業でのやり取りで培われたものである。「無料版」と称した本書の β 版の誤りも多数発見してくれた。これらは「有料版 (本書)」の完成度の向上に大いに貢献している。また、本書の刊行の機会を与えてくださり、構成について助言をいただいたコロナ社の皆様に心から感謝する。

2019 年 1 月

山 本 宙

目 次

序 章 情報理論とは

1. 問題はなにか

1.1 起きた事柄と伝える手段	4
1.2 通信用の文字に変換—符号化—	6
1.3 通信コストを下げるために—通信に必要な符号語長—	8
章 末 問 題	11

2. 符号の種類

2.1 天気と通信をモデル化—情報源と符号—	12
2.2 必ず元に戻せる符号—一意に復号可能—	15
2.3 遅延なく復号できる符号—瞬時に復号可能—	20
章 末 問 題	25

3. 符号語長の制約

3.1 符号語長に課せられる制限—クラフトの不等式—	26
3.2 瞬時に復号可能な符号が存在するための十分条件	34
3.3 一意に復号可能であるための必要条件	37

章 末 問 題	42
---------	----

4. 平均符号語長とハフマンの符号化

4.1 「最適」な符号——平均符号語長とコンパクト符号——	43
4.2 瞬時に復号可能なコンパクト符号を得るハフマンの符号化	46
4.3 ハフマンの符号化の結果がコンパクト符号であることの証明	51
章 末 問 題	60

5. エントロピーと情報量

5.1 平均符号語長の下界とエントロピー	61
5.2 平均符号語長の下界を達成できる条件	68
5.3 まとめて符号化して通信効率を改善——情報源の拡大——	70
5.4 拡大情報源の平均符号語長の下界	75
5.5 特定の符号化についての平均符号語長の上界	77
5.6 情報源符号化定理	80
5.7 情 報 量	81
章 末 問 題	84

付録 A 数 学 の 準 備

A.1 式の書き方の約束事	86
A.2 入力から出力への対応——写像——	87
A.3 生起確率が与えられた場合の平均値——期待値——	89
A.4 これだけは知っておいてほしい, 必要条件, 十分条件と証明	91
A.5 指数, 対数の考え方	95

A.5.1	よく使う指数の公式	95
A.5.2	対数の定義	96
A.5.3	よく使う対数の公式と証明	96

付録 **B** 情報量の尺度が対数関数に限られる理由

B.1	情報の量の尺度に求められる性質	99
B.2	情報量の定義	100
引用・参考文献		104
章末問題解答		105
索 引		111

Coffee Break ☕

話は短いほうがいい	2
大盛り一丁!	10
電話番号は瞬時に復号可能	23
一周回って一気に証明	33
鳩の巣原理	41
理論と現実	58
現実との対応	81
通信路が100%信用できないときは?	83
人はなぜくじを買う	91
定義はどうやって決める	94

本書の使い方

- 本書では定義、定理、命題、補題、系、条件、図、表には章ごとに独立した通し番号を付している。
 - **定義**は用語の意味を厳密に示したものである。同じ用語に文書によって異なる定義が与えられていることがあるので注意してほしい。
 - **定理**、**命題**、**補題**、**系**はいずれも定義から論理的に証明された事柄をいう。主要な結果として強調したいものを定理、定理というほどではないものを命題、定理や命題を証明するために証明されるものを補題、それらから直接導き出せるものを系と呼んでいるが区分は厳密なものではない。証明の終わりに□を付した。
 - **条件**は状況によって真にも偽にもなり得る事柄をいう。
- 説明をわかりやすくするために例、例題、章末問題を記述した。これらも章ごとに独立した通し番号を付している。
 - 抽象的な定義や証明の一部を理解しやすくするために例として具体例を挿入した。
 - **例題**は理解を助けるために本文中に挿入される問題である。解答を直後に記述し、解答の終わりに◇を付した。
 - より理解を深めるために各章章末にその章で学習したことを総合的に使って解く**章末問題**を配置した。問題が小問に分かれている場合は順番に解くことで自然に全体が解けるように配慮した。解答は章末にまとめて記述した。
- 重要な用語は太字にし、巻末の索引に主要な説明のあるページ番号を記した。一部の用語には英語での表記を付した。

- 数学を学習する機会が少なかった学習者を対象に，指数，対数や和の式の書き方の慣習と，写像，期待値，必要条件，十分条件と証明方法に関する解説，さらに，よく用いられる指数と対数の公式の解説を付録 A に記した。必要に応じて参照してほしい。

本書で用いる変数と関数

記号	意味
a	符号の符号語長の最大値
$H(S)$	情報源 S のエントロピー
$I(s_i)$	情報源シンボル s_i の事象の情報量
L	平均符号語長
l_i	符号語 X_i の系列長
L_n	n 次拡大の平均符号語長
n_i	符号語長が i の符号語の個数
$P(E)$	事象 E が起こる確率
P_i	情報源シンボル s_i の生起確率
q	情報源シンボルの個数
S	情報源シンボルの集合または情報源
S^*	S の要素の長さ 0 以上の系列
S_i	情報源 S を i 回縮約したもの
s_i	情報源シンボル
S^n	情報源 S の n 次拡大
X	符号シンボルの集合
X^*	X の要素の長さ 0 以上の系列
X^+	X の要素の長さ 1 以上の系列
X_i	情報源シンボル s_i の符号語
σ_i	拡大情報源の要素

序章 情報理論とは

従来型の PC に代わってスマートフォンやタブレット端末が情報ツールの主役となり、一般ユーザーへの普及と便利なサービスの広がりとはどまるところを知らない。スマートフォンはコンピュータの基本的な機能を備える「小さな PC」と考えられていたが、近年の利用の広がりを見るともはや別のものと考えるほうがよいだろう。スマートフォンが従来型の PC と大きく異なる点はインターネットへの常時接続を前提とし、必要なデータを必要なときにネットワークから得るといった設計思想にある（図 0.1）。

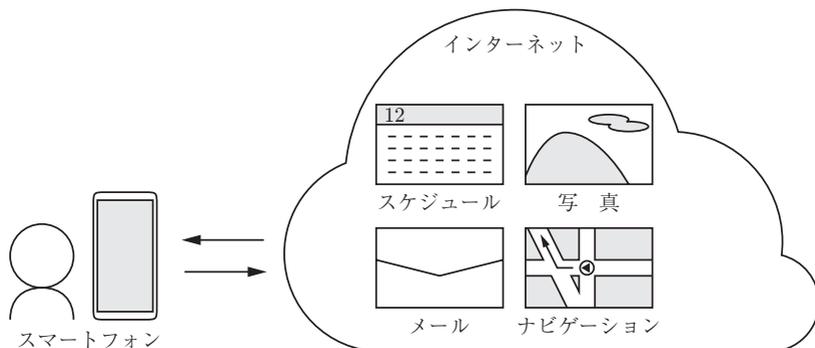


図 0.1 スマートフォンとインターネット

このアイデア自体は何十年も前からある、技術者の夢ともいえるものであったが近年になって情報の通信が飛躍的に高速化、高信頼化されたことでようやく現実のものとなった。通信の高速化、高信頼化をもたらした技術は 1940 年代に始まる情報理論 (information theory) という学問が基礎となっている。

2 序章 情報理論とは

情報理論という学問の名前は誤解を生みやすいところがある。例えば、ある都市の毎日の天気がある確率に従って決まるとする。このような確率に従って決まるものを**確率変数** (random variable) という。毎日の天気を連続して観測することを考える。この場合、日付という時間ごとの確率変数の値、すなわち**離散確率過程** (discrete stochastic process) として毎日の天気の発生をモデル化することができる (図 0.2)。

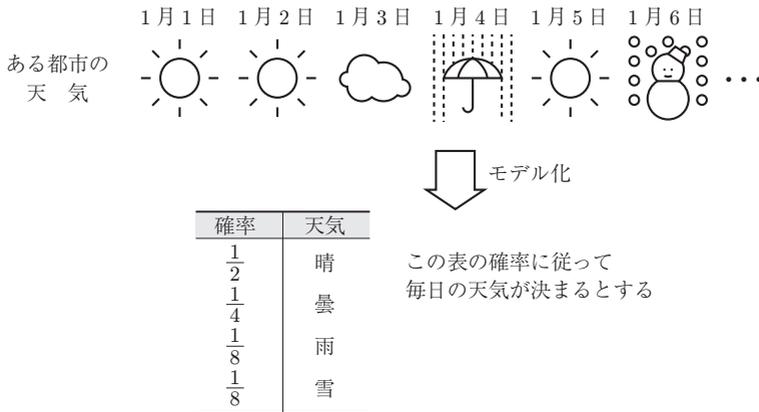


図 0.2 天気のモデル化

情報理論とは情報の発生を離散確率過程としてとらえ、それを受信者に伝えるという枠組みの中で通信の効率などに関する数学的性質をあきらかにする学

Coffee Break ☕ 話は短いほうがいい

本書では通信するデータの量は小さいほうがいいという前提で議論が進みます。通信するのはなにかを伝えるのが目的で、通信文を送るのはそのためのコストだという立場です。

もちろん実世界ではそうでない場合もあります。人気漫才師のライブに行ったのに二言三言で終わってしまったら悲しいですよね。言葉の楽しさ自体に価値がある場合もあるのです。内容が伝わるなら文字は少ないほうがいいというのならこのコラムだって無駄ということになってしまいます。

問といえる。ある都市の天気をほかの都市に通信する場合、通信するデータの量は小さいことが望ましいがどのように伝えれば通信するデータの量が小さくなるのか、その限界はどこまでか、といった議論を行う。つまり、情報理論と呼ばれる学問は、名前から期待されるような情報の内容に関する理論ではなく、離散確率過程の出力を通信する場合の数学的理論である。情報理論は 1948 年にシャノン (C. E. Shannon) により発表された論文『A Mathematical Theory of Communication』¹⁾† により始まる。このタイトル (通信の数学的理論) がまさにこの学問の本来の内容を示している。

情報理論の発展により、デジタルデータの圧縮、デジタル通信の高速化、ノイズのある環境での高信頼性通信などさまざまなことが可能となった。情報理論は現代のデジタル通信網が成り立つうえで最も重要な基礎学問の一つとなっている。

† 肩付き数字は巻末の引用・参考文献を表す。

1. 問題はなにか

情報を通信する、という言葉は、ある場所で起きた事柄をそれを直接観測できない別の場所に伝えるという意味で使われる。「大阪の今日の天気を東京に伝える」というのは情報を通信している例であり、情報理論はこれを繰り返し行うときの問題を議論する学問である。ここでは厳密な定義を行う前に情報を通信するときの問題として平均の通信量の削減を目的とすることをいくつかの要素に分けて説明する。

1.1 起きた事柄と伝える手段

「起きた事柄」として、大阪の天気を例に取り上げる。ここでは簡単のために天気は晴、曇、雨、雪の4種類だけであるとしよう。ある日の大阪の天気はこの4種類のどれかであるが、どれであるかを東京に伝えなければならない(図1.1)。



図 1.1 大阪の天気を東京に伝える

「伝える手段」として、図 1.2 のような装置と、ほかと絶縁されたペアの導線が大阪-東京間に設置されているとする。大阪側にあるスイッチを開いた状態（図 1.3(a)）ではこの導線のペア間の電位差は 0V に、閉じた状態（図 1.3(b)）では 5V に設定される。導線のペアの東京側では現在の導線の電位差が 0V なのか 5V なのかを観測する。ここでは途中で混入するノイズ、測定器の誤差の問題は無視し、東京では 2 通りの電位差のどちらが大阪で設定されているか完全にわかるものとする。ここで、大阪でスイッチが開いていることを、東京で導線のペアの電位差が 0V であることを観測して知ることを「大阪から東京へ 0 を送る」と表現し、大阪でスイッチが閉じていることを、東京で導線のペアの電位差が 5V であることを観測して知ることを「大阪から東京へ 1 を送る」と表現することで抽象化を行うことができる。5V という具体的な電圧値には意味がなく、2 種類の状態の区別を大阪から東京に伝えることができることが本質なので、送られる記号については「アを送る」と「イを送る」でもよいのだが、慣例で 0 と 1 を使用する。

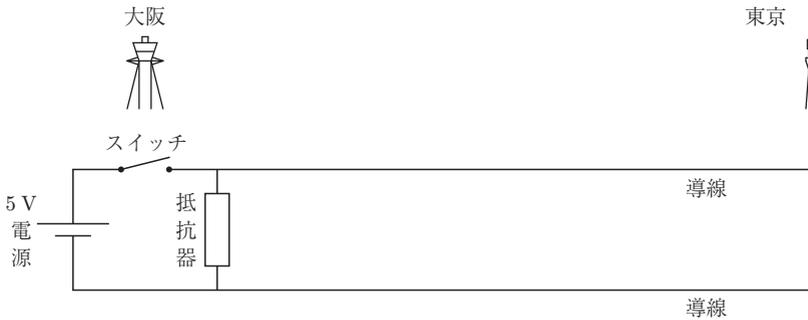


図 1.2 通信装置

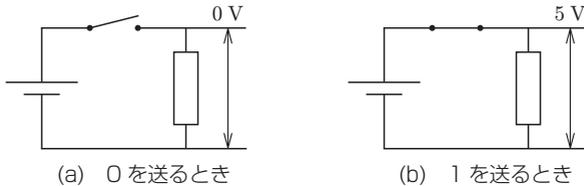


図 1.3 通信装置の大阪側の操作

索引

【い】	
一意に復号可能	17
一意に復号可能な符号	46
【え】	
エントロピー	67
【お】	
親	51
【か】	
確率変数	2
仮定	92
可変長符号	8
【き】	
期待値	89
狭義に特異	16
【く】	
クラフトの不等式	28
【け】	
系列	16
結論	93
【こ】	
子	51
固定長符号	8
語頭	23
語頭条件	22, 23
コンパクト符号	45
根本が十分条件	92
コンマ符号	20

【さ】	
先が必要条件	92
【し】	
事象の情報量	82
写像	87
十分	92
十分条件	92
縮約	46
瞬時に復号可能	22
情報源	12
——の n 次拡大	71
情報源シンボル	12
情報源符号化定理	80
情報量	82
情報理論	1
【た】	
単写	88
【て】	
定義	92, 93
【な】	
内部節点	51
【に】	
2元系列	14
2元ブロック符号	14
2分木	51
【ね】	
根	51

【は】	
葉	51
ハフマンの符号化	46
【ひ】	
ビット	103
必要	92
必要十分条件	93
必要条件	92
【ふ】	
復号	6
符号	6
符号化	6
符号木	52
符号語	6, 14
符号語長	6
符号語長並び	32
符号語長分布	34
符号シンボル	14
符号表	6
【へ】	
平均符号語長	44
【む】	
無記憶情報源	13
【り】	
離散確率過程	2
【れ】	
レベル	54

— 著者略歴 —

1991年 大阪大学基礎工学部情報工学科卒業
1993年 大阪大学大学院基礎工学研究科博士前期課程修了 (物理系情報工学分野)
1996年 大阪大学大学院基礎工学研究科博士後期課程修了 (物理系情報工学分野)
博士 (工学)
1996年 大阪大学助手
2000年 東海大学講師
2004年 東海大学助教授
2007年 東海大学准教授
現在に至る

情報量 — 情報理論への招待 —

Information Entropy

— An Introduction to Information Theory —

© Hiroshi Yamamoto 2019

2019年3月1日 初版第1刷発行



検印省略

著者 ^{やま}山 ^{もと}本 ^{ひろし}宙
発行者 株式会社 コロナ社
代表者 牛来真也
印刷所 三美印刷株式会社
製本所 有限会社 愛千製本所

112-0011 東京都文京区千石 4-46-10

発行所 株式会社 コロナ社
CORONA PUBLISHING CO., LTD.

Tokyo Japan

振替 00140-8-14844 · 電話 (03) 3941-3131(代)

ホームページ <http://www.coronasha.co.jp>

ISBN 978-4-339-02890-4 C3055 Printed in Japan

(松岡)



 < 出版者著作権管理機構 委託出版物 >

本書の無断複製は著作権法上での例外を除き禁じられています。複製される場合は、そのつと事前に、出版者著作権管理機構 (電話 03-5244-5088, FAX 03-5244-5089, e-mail: info@jcopy.or.jp) の許諾を得てください。

本書のコピー、スキャン、デジタル化等の無断複製・転載は著作権法上での例外を除き禁じられています。購入者以外の第三者による本書の電子データ化及び電子書籍化は、いかなる場合も認めていません。落丁・乱丁はお取替えいたします。