

電子・情報工学講座

17

情報工学基礎論

田村進一 著

培風館

電子・情報工学講座

刊行にあたって

「現代の科学技術は、電子工学と情報工学で支えられている」と言っても過言ではない。最近、大学は大講座制に移行する傾向があるが、その大講座制の名称を見ても、工学系には電子・情報工学の名称のない学部は見あたらない。

この時代の要請にしたがって、この分野の大学の教科書を目的とした「電子・情報工学講座」のシリーズの出版計画を、培風館から持ちかけられた。それは1984年の春であった。これらの分野の単行本はこれまで優れたものが数多く出版されているので、シリーズとして何か特徴がなければ意味がないと考えた。大学の講義と言うものは、70%前後理解されるように講義するのがよいと言われている。30%程度理解できない部分を含んだ講義でないと、学生が退屈してしまうと言われている。しかし本来は100%以上理解される講義の方がさらに良いのではないかと考えている。できればこの程度のレベルで、かつ内容的には、最新の情報が得られ、かつ基本的事項を十分理解できるような内容のシリーズ物にしたいと考えた。しかし、最終的には、「著書の善し悪しは執筆者の筆で決まる」と言うごく当り前の結論に到達してしまった。

そこで執筆者としては、現在この分野の第一線で活躍されている第一級の研究者の方々にお願いすることにした。それは第一級の教育者は第一級の研究者でなければならないとの編者のフィロソフィーに基づいている。しかしここで大きな難問に出くわしてしまった。それは、第一級の研究者の方々は、当然国際会議などの出席が多く、海外出張などきわめて多忙であり、とても執筆をお願いしてもお引受け頂けないのではないか、と言うことであった。そこでまた考えを改めて別の執筆者の先生方をお願いした方がよいのではないかと考えた

㊦ 〈日本複写権センター委託出版物・特別扱い〉

本書の無断複写は、著作権法上での例外を除き、禁じられています。
本書は、日本複写権センターへの特別委託出版物ですので、包括許諾の対象になっていません。
本書を複写される場合は、日本複写権センター(03-3401-2382)を通してその都度当社の許諾を得てください。

が、「物事を願うするには、一番忙しい方をお願いしろ」と言われている結論に到達してしまった次第である。この二つの結論が今回のシリーズの全体像である。この全体像を描くまでに1年を要し、1985年春に執筆者の方々に執筆をお願いし、今回このシリーズの出版の運びになった次第である。

本シリーズは、基礎、回路、物性、デバイス、通信、情報基礎、情報のハードウェア、ソフトウェア、及び最新の分子電子工学に至るまで26巻から成っており、重要な分野をすべて網羅し、「電子・情報工学」の基礎事項は勿論、最新の学問がすべて凝縮されているものと確信している。

本シリーズが、この分野を専攻する学生はもちろん、研究者に対しても100%以上の理解がえられることを期待したい。また本書から、この分野の第一級の研究者・教育者の、研究並びに教育に対するフィロソフィーをも学びとっていただければ幸いである。

きわめて多忙な中、本シリーズの執筆をお引受け頂いた執筆者の方々には、厚く御礼申し上げる次第である。また本シリーズの出版に関しては、全面的なバックアップをいただいた培風館の松本和宣氏に深く感謝する。

平成元年師走

東京工業大学
大岡山キャンパスにて
監修者 高橋 清

はじめに

本書は、電気、電子、情報専攻の学生のみならず、広く一般の学生を対象とし、情報工学の基礎理論を学ぶことを目的として書かれたものである。

情報工学の基礎理論としては、まず情報理論があげられる。情報伝達および蓄積の効率化、信頼性向上に関してそのバックボーンを与えるものである。この分野は近年、ファクシミリ、CD、コードレス電話の符号や暗号などとして日常生活の身近な機器にまでその成果が利用されるようになってきている。このように、情報の処理・蓄積・通信が普遍的になるにつれて、その適切な取扱いが必要になるケースが種々発生している。そのようなことから、情報源と通信路のモデル、符号化の手法と限界などを正しく理解する必要性が増えてきている。またこのような直接的必要性から離れて、情報とは何かを静かに思索するとき、情報理論は統計的な視点からの明確な道標を与えてくれる。そこで本書では、情報理論を主要な内容とし、2章～8章をこれにあてている。

デジタル時代を反映させて、これらはデジタル情報を基本として構成されており、その基礎のもとにアナログ的な信号の取扱いや周波数空間については最後の7、8章で学ぶ。これらの章ではデジタルの世界とアナログの世界を結び付ける標本化定理やフーリエ変換、アナログ情報源、およびアナログ通信路などについて学ぶ。

以上の内容は大阪大学基礎工学部情報工学科、および神戸大学工学部システム工学科における情報理論関連の講義ノートをもとに整理したものである。

上記のような情報理論関連の内容にくわえて、情報工学分野では最近、人工知能、知識処理、神経回路網計算など、知能工学ともいうべき分野の研究が広く行われるようになってきている。そこで、本書ではまず、音声や画像など知的情報の取扱いに必要となってくるパターン認識の基礎理論を9章で学ぶ。これは情報理論と比較的関連の深い分野である。つぎに10, 11章において人工知能分野の基礎理論である命題論理と述語論理について学ぶ。さらに、柔軟な論理に従って情報処理を行うコンピュータであるニューラルネットワーク情報処理の基本原則について最後の12章において学ぶ。これらは情報の意味内容に立ち入って議論する点で前半の情報理論とは傾向の異なる部分を含むのでこれらの章は前半と独立に理解できるようになっている。この部分が他の類書と大きく異なり、本書の特徴となっているところである。

以上のように、本書は情報工学、知能工学の基礎理論を述べるものである。とくに本書では、これらの基本的なことがらを記述することに重点を置いた。そのため、簡単なことがらでもできるだけ例を入れるようにし、理論の本筋の理解に必ずしも必要でない記述は付録にまわした。また、本書の内容とレベルに合わせて演習問題をつくってある。巻末にできるだけ丁寧にその解答を付しておいた。したがって、演習は本書の内容の理解の確認に役立つであろうと思っている。例題や演習に必要なだけでなく本書の理解に役立つ数表とフーリエ変換対表、また、もっと深く内容を理解するための参考書のリストなども巻末に付した。

1991年12月

田村 進一

目 次

1. 序 論	1
1.1 情報理論とは	1
1.2 通信系モデル	2
1.3 情報源符号化	3
1.4 通信路符号化	3
1.5 その他の符号化	4
1.6 アナログ情報の標本化	4
1.7 アナログ情報の量子化と情報量	5
1.8 パターン認識	6
1.9 命題論理と述語論理	7
1.10 ニューラルネットワーク	8
2. 情報量	9
2.1 事象と確率	9
2.2 条件付き確率	10
2.3 完全事象系	12
3. 情報源	15
3.1 情報源の統計的表現	15
3.2 記憶のない定常情報源	15
3.3 エルゴード情報源	16

3.4	記憶のある情報源	18
3.5	一般化されたマルコフ情報源	19
3.6	状態の分類	21
3.7	マルコフ情報源の定常分布	22

4. 情報源符号化

25

4.1	離散的情報源の符号化	25
4.2	瞬時符号であるための必要十分条件	27
4.3	符号の木	28
4.4	クラフトの不等式	29
4.5	平均符号長の限界	31
4.6	ハフマン符号化	34
4.7	ブロック符号化	38
4.8	情報源符号化定理	39
4.9	非等長情報源記号系列に対するハフマン符号化	42

5. 離散的通信路

45

5.1	通信路モデル	45
5.2	記憶のない定常通信路	46
5.3	相互情報量	48
5.4	通信路容量	50
5.5	記憶のない一様通信路の通信路容量	51
5.6	情報伝送速度	53
5.7	通信路符号化定理	54

6. 通信路符号化法

58

6.1	組織符号	58
6.2	単一パリティ検査符号	59
6.3	水平垂直パリティ検査符号	60
6.4	線形符号	61
6.5	ハミング距離	61
6.6	線形符号の加法性	62
6.7	最小距離・最小重みと誤り訂正能力	62

6.8	パリティ検査方程式	64
6.9	単一誤り訂正ハミング符号	66
6.10	バースト誤り	67
6.11	巡回符号	68
6.12	巡回ハミング符号	75

7. 標本化定理

78

7.1	簡単な例	78
7.2	フーリエ級数	80
7.3	フーリエ変換	83
7.4	標本化定理の導出	84
7.5	デルタ関数	86
7.6	インパルス応答	88
7.7	自己相関関数とパワースペクトル	90
7.8	標本値列からの復元	90

8. アナログ通信路

96

8.1	確率密度関数	96
8.2	アナログ情報源のエントロピー	98
8.3	最大エントロピー定理	100
8.4	通信路容量	101
8.5	量子化雑音	103
8.6	ベクトル量子化	104

9. 統計的パターン認識

107

9.1	多次元ガウス分布	107
9.2	統計的決定法	109

10. 命題論理

115

10.1	論理式	115
10.2	論理式の構成法	117
10.3	論理式の解釈	117
10.4	意味の木	123

10.5 命題論理式の標準形	124
11. 述語論理	127
11.1 論理式の構成法	127
11.2 恒真性と充足性	129
11.3 節 形	130
11.4 導 出 節	132
11.5 単 一 化	134
11.6 導出原理	136
12. ニューラルネットワーク計算	141
12.1 ニューラルネットワークの形態	141
12.2 階層型ネットワークと誤差逆伝搬学習アルゴリズム	142
12.3 相互結合型ネットワーク	146
付 録	155
A 1 デイビス・パットナムのアルゴリズム	155
A 2 命題論理の公理系	157
A 3 エルブラン空間	159
文 献	163
演習問題解答	165
付 表	183
索 引	187

1

序 論

本章では本書の内容の概略について述べる。

1.7 節までは本書の8章までの内容であり、主として情報理論と信号理論の話である。情報理論は、情報伝達および蓄積の効率化、信頼性向上の基礎理論となるものである。これらは情報の内容や意味については立ち入らず、主として統計的性質に着目して、情報を発生・伝送・変換するときの定量的な諸性質を議論するものである。1.8 節以降は9章以降の内容であり、知的情報処理の基礎理論を述べている。これらはパターン認識、知識表現と推論、およびニューラルネットワークに関するものであり、情報をその意味内容にまで立ち入って議論する。これらは人工知能や知能情報の基礎理論となるものである。

1.1 情報理論とは

情報理論は情報伝達の効率化、高信頼化のための符号化の理論である。現代の情報理論の中心はシャノン理論である。シャノン理論は情報の量的表示および符号化の限界に関する理論である。

日常生活において、ある日の株価を知るとはAさんにとっては非常に値打ちのあることであろうし、Bさんにとってはまったく意味がないものであるかも知れない。このように、情報の価値は受け手により大きく変わってくるといえるであろう。このような情報の価値まで考慮した定量化は、重要ではあっても難しい。それに対して、シャノンは対象の統計的な性質に着目し、その定量化を行った。たとえば通信線を通して英文が送られてくる状況を考えてみよう。ここでは、英文の内容には立ち入らずにその文字や単語の出現頻度のみに

注目するものとする。文章を次々と受け取っている場合には、次に受け取るであろう文字や文章の察しがおおよそつくこともあるであろうし、まったくわからない場合もある。すなわち、受信者の受信文に対する予見的確率分布は状況によって変化する。つぎに、文章を実際に受信するとその確率は1(または1に近い)となり、それ以外の文章の確率は0となるであろう。このように、通信によって情報の受け手の確率分布が変化する。この変化を情報の伝達ととらえるのである。これにより、情報の客観的な定量化が可能となる。またその性質を調べることで通信の限界も明らかとなるのである。さらに、このような定式化により得られた結果は、通信系のみならず、物理現象の観測系などにも適用できる部分がある。

1.2 通信系モデル

情報理論は情報伝達の効率化を図るための理論であることはすでに述べた。

図 1.1 は、通信系のモデルを示す。

情報源は、たとえばキーボードからの入力文字や、ファクシミリにおいては入力図形の白黒パターンなどである。

情報源符号化は、情報源から発生する情報源記号系列から冗長性を取り除き、それに続く通信路符号化に適した2元符号などに変換することである[†]。得られる符号はできるだけ短いものがよい。ただし、あまり短くすると、復号時に誤りが生じたとき、誤りの伝搬が生じることがある。

通信路符号化の目的は、情報源符号系列に対して何らかの冗長性を付加することにより、誤りの検出・訂正を可能とする符号を得ることである^{††}。

受信側ではこれらの符号を受け取りそれらをもとに戻すための復号を行う。

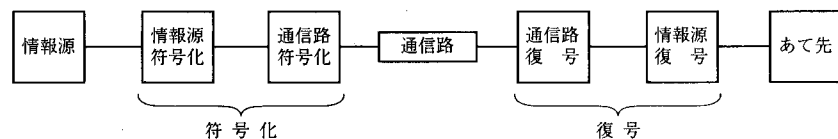


図 1.1 通信系モデル。

[†] 情報源符号化の例としては、ファクシミリにおける MR 符号や、画像伝達・蓄積のための離散コサイン変換などがある。

^{††} 通信路符号化の例としては、パケット通信における CRC 符号や、計算機の記憶装置、光ディスクにおける信頼性向上のための ECC 符号がある。

1.3 情報源符号化

いま、表 1.1 に示すような情報源記号 $\{A, B, C, D\}$ をそれぞれ確率 0.6, 0.2, 0.15, 0.05 で発生する記憶のない情報源 S を考える。情報源に記憶がないということは、記号の発生確率が過去にどのような記号が発生したかによらないということである。

いま、 C_1 なる符号を用いると、その符号長の平均(平均符号長)は

$$L=2$$

である。それに対して、符号 C_2 を用いると平均符号長は

$$L=1 \times 0.6 + 2 \times 0.2 + 3 \times 0.15 + 4 \times 0.05 \\ = 1.65$$

となり、 C_1 より効率がよい。このように、記号の生起確率に応じてその符号長を変えると平均符号長が短くなり、それだけ効率のよい情報源符号化が行える。この身近な例は電話の市外局番であり、利用者の多い大都市は短く、少ない地方は長い番号が与えられている。

表 1.1 情報源符号化の例

情報源記号	生起確率	符号 C_1	符号 C_2
A	0.6	0 0	0
B	0.2	0 1	1 0
C	0.15	1 0	1 1 0
D	0.05	1 1	1 1 1 0

1.4 通信路符号化

ランダムに確率 10^{-3} で誤りを起こす通信路があったとする。通信路符号化では単に各入力記号を3度づつ繰り返して通信路に送り出すものとする。このとき入力記号に対する誤り率は、通信路記号が3個中2個誤りを起こす確率でほぼ決まる。なぜなら、3個中3個とも誤る確率 P_e はこれに比べてかなり小さいからである。したがって、それは

$$P_e = 3 \times (10^{-3})^2 = 3 \times 10^{-6}$$

で与えられる。これは単純な例であるが、一般に通信路符号化により、冗長性

付加, すなわち表面的な効率低下と引き換えに, 通信の信頼性を大きく向上させることができる。

1.5 その他の符号化

情報源記号の生起確率が未知, または時間的に変動するような場合など, 通信系のモデル化が十分に行えないときに, そのときの状況に応じた符号化により効率のよい通信を行う方式として**適応的符号化**や, さらに送受間の共通の知識を積極的に用いて通信の効率化を図る方式として**知的符号化**が提案・研究されている。後者の知的符号化の例として, Aさんの笑顔を送ろうとすると, そのときの顔画像そのものを送るのではなく, すでにAさんの顔画像が受信側にあれば, 顔の筋肉の動きのパラメータのみ送る, または笑っているという情報のみ送る。受信側では, 手元にあるAさんの顔画像と受け取った情報をもとにAさんの笑顔を復元する。これは, 送受両者が持っているAさんの普通の顔という共通の知識を利用して, もっと効率のよい通信を行おうとするものである。ただし, これらの符号化については本書ではふれない。

通信路符号化では通信の信頼性ということに重点を置いたが, 現実世界ではデータの保全性ということに重点を置いて符号化を行うことがある。この場合は特に**暗号**と呼ばれるが, これについても本書ではふれない。

1.6 アナログ情報の標本化

自分のおおよその体重を知りたいとき, 毎日測定してもあまり変化がなく, 一週間に一度測定すれば十分で, 日々の体重はおおよそ推定ができるであろう。世の中に自然に存在する情報は, たとえば音声情報や, また上の体重の例のように, 時間および振幅に関して連続的であることが多い。時間的に変化する振幅(たとえば体重)をある時間を決めて測定することを**標本化**という(図1.2)。それらの量の時間的な変化速度は通常限られている。そのような場合, 時間的にあまり細かく標本化しても変化が少なく情報を与えないであろう。

標本化定理は, 音声情報のように最高周波数が制限されている連続的信号を“どのような時間間隔で標本化すればもとの情報を損なわずに標本化できるか”に解答を与えてくれる。すなわち, 最高周波数の2倍以上の速度で標本化すればよい, というのがその結論である。たとえば, 電話音声はせいぜい4kHz

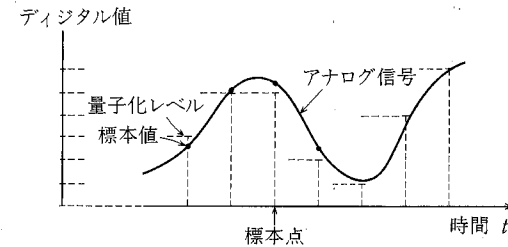


図 1.2 アナログ信号の標本化と量子化

までの周波数しか含まないので, 通常その2倍の8kHzで標本化される。また, ヒトの耳にはおよそ20kHzまでの音しか聞こえないので, CDやデジタル録音では40数kHzの標本化周波数が使われる。

一方, このような条件が満たされないとエイリアシング(折り返し歪)と呼ばれる縞模様などが発生する。日常生活でもこのような現象を見かけることがある。たとえば規則的な縞模様の格子が二つあったとき, それらの間にモアレ縞が観察される。また, 細かい格子縞の印刷写真にも格子の周波数と印刷写真の網点の差の周波数の縞模様がしばしば見られる。もともとなかった縞模様(歪)が差の周波数により生じたのである。標本化という操作は乗算的操作であり, その結果, 和と差の周波数が発生する。それらが見えなければ問題がないが, 信号の帯域と重なってしまうとエイリアシングが発生して問題となる。それをクリアする条件が“2倍”である。このようにして標本化された標本値列からは, もとの連続的信号が完全に復元できる。

このように, 標本化定理はアナログ(連続的)情報とデジタル(離散的)情報をつなぐ役目をするものである。

1.7 アナログ情報の量子化と情報量

連続値をとるアナログ信号の標本値が与えられたとき, それは振幅に関してはまだアナログ値のままである。これをデジタル値に変換するためにはある範囲内の連続値を同一のデジタル値に変換する量子化操作を行う(図1.2)。この操作はA/D変換(アナログ-デジタル変換)と呼ばれる。このとき量子化誤差が発生する。

アナログ情報は任意の細かさのレベルを使用できるとすると無限の情報を担

い得る。たとえば有理数のレベルだけ使用するとしてもそれは無限個あり、したがって、無限レベル数のデジタル値と同じとなり、任意のそれを表現するためには無限長の2進データがいることになる。しかしながら実際には必ず雑音が存在するために、雑音レベル以下のレベル差は識別できず、識別できるレベル数は有限となる。通信路においても同様の議論が成り立つ。そこでは、送られた信号は途中で雑音や歪のため乱される。たとえば、高速に送ろうとすると歪が増し、受信信号の識別が困難となり、誤りが増えてかえって情報の実効伝達速度を低下させることになる。たとえていうならば、道路には時間あたり通過できる車の量に限界がある。それ以上に車が入り込んでくれば渋滞となり、かえって通過できる車の量は減少するであろう。同様に通信路にも通信路容量が存在し、その容量以上の速度で情報を伝達することは不可能であることが証明できる。アナログ通信路においてはそれは帯域幅とS/N比の単調増加関数である。通信路容量と道路の容量の類似性を図1.3に示す。

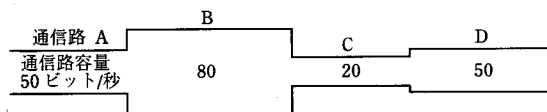


図 1.3 縦続接続された通信路の容量はいちばん狭いところの容量(20)となる。

1.8 パターン認識

ある時刻において雑音を含む信号中にパルスが存在するかどうかを判定するためには、そのときの信号電圧を測定し、この標本値が本来あるべきパルスの電圧の1/2以下であればパルスなし、それ以上であればパルスあり、と判定を行う(図1.4)ことは合理的であろう。

もっと一般的には信号をいくつかの点で標本化したり、画像から幾何学的ないくつかの特徴を取り出すと、それらは多次元的な特徴ベクトルで表現される。このような特徴ベクトルは雑音を含むことが多い。その変動を確率現象とみなし、統計的な見地より理論的にいちばん良い決定を行う手法が統計的パターン認識である。これらは文字認識や信号検出等の基礎となる。

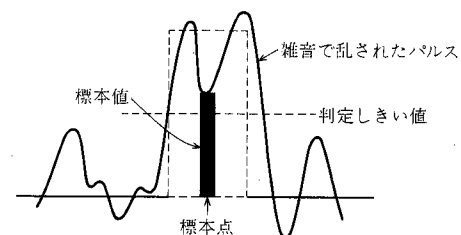


図 1.4 標本値からのパルス有無の判定

1.9 命題論理と述語論理

アリストテレスの時代には、論理学は文章による論証の形式的な正しさを論ずる学問であった。これを記号を用いて厳密な学問の体系に仕上げたのが**記号論理学**である。記号論理学では記号を用いて物事の真偽を論理的に導き(推論し)、決定する。

人間の日常的な知的活動の主要な部分もやはり、与えられた外界の状況と自分がもっている知識とから新しい知識を生み出す“**推論**”機能であろう。この推論の代表的方式は**三段論法**である。三段論法の例は、

例 1	仮定	今日は雨が降っている
	仮定	雨が降っているなら傘がいる
例 2	結論	今日は傘がいる
	結論	ソクラテスは人である
	仮定	人であれば足をもつ
	結論	ソクラテスは足をもつ

であり、仮定から三段論法により新しい結論を導いている。これらは命題論理や述語論理でうまく表現できる。述語論理は命題論理よりも多様な対象を取り扱うことができる。特に述語論理において、このような操作を機械的に効率よく行うための方式に**導出原理**がある。これはまた、論理型プログラミング言語 Prolog、蓄積された知識をもとに推論を行い、質問に対してうまく合致する答を探し導き出してくる質問応答システムや、専門家の知識をもったエキスパートシステムの構成原理にもなっている。

1.10 ニューラルネットワーク

ニューラルネットワークは生体の神経回路を模した回路で各種の情報処理を行うとするものである。論理学の体系が堅いのに対して、ニューラルネットワークの研究は人間のような柔らかな論理に従って柔軟に処理・行動を行う機械を作ることをめざしている。

ニューラルネットワークは多数の神経素子と、それらの間を結ぶ重み付き配線からなっている。それらの重みを変えることにより、目的の処理に適した回路を構成することができる。すなわち、人間のように学習することが可能であることがニューラルネットワークの大きな特徴である。これにより、プログラミングなしにさまざまな情報処理を行うことが可能になるであろうとの期待がかけられている。

ニューラルネットワークの代表的形態である階層型ネットワークでは、入力パターンから出力パターンへの任意の変換が可能であることが知られている。したがってこれは、画像の輪郭を抽出したり、ボケを取り除いたりするフィルタリングやパターン変換器に使うことができ、またその特殊な場合であるパターン分類やパターン認識に使うことができる。このネットワークは、多くのことを覚えすぎると新しいことを覚えられないなど人間の脳に似た性質をもつ。

もう一つのニューラルネットワークの代表的形態である相互結合型ネットワークはその基本的機能としてパターンの連想機能をもつ。たとえば覚え込んだ人の鼻写真を見て、顔全体を出力する機能である。これを利用すると、巡回セールスマン問題など、計算量的に問題の多い組合せ問題も人間のように近似解を速く求めることができる。

2 情報量

本章では、まず事象間の演算、およびその上で定義される確率の基本的な性質について述べる。これらの準備のもとに、事象系のもつ情報量、およびあいまいさを表わすエントロピーを定義する。

2.1 事象と確率

世の中には「今日は晴れている」や「硬貨を投げて裏が出る」のように、さまざまな事象(event; 出来事)が起こっている。

一般に人が感じる情報とは、自分や周囲の人に利害をもたらしたり、関心のある事柄が重要な情報となる。たとえば、まったく知らない人が宝くじに当たったという情報よりも、自分の家族が宝くじに当たったという情報のほうがはるかに大きい情報をもたらす。また、興味のない分野のニュースよりも関心の大きな分野のニュースのほうがより大きな情報をもたらす。しかしながら、このような受け手の状態や主観に依存する測り方では客観的な議論はできない。

シャノン(Shannon)に始まる“情報理論”では、確率的に定義することのできる事象の集合の上で客観的な情報の量の測り方を定義し、それに基づく情報伝達(通信)の理論を与えている。

任意の二つの事象を E, F とする。これらの間に次のような集合論的演算を定義することができる。

- (1) “ E と F の少なくとも一方が起こる”という事象を E と F の和事象といい、 $E \cup F$ で表わす。

(2) “ E と F がともに起こる” という事象を E と F の積事象といい、 $E \cap F$ で表わす。

(3) “ E が起こらない” という事象を E の余事象といい、 $\bar{E}$ で表わす。

基本的な事象(根元事象)から上のような積和、および余事象演算を何度か行つて、合成事象を導くことができる。さらに、つぎのような集合論的包含関係を定義できる。

(4) “ E が起これば必ず F も起こる” とき、それを $E \subset F$ で表わす。

【例 2.1】 A, B, C を事象とすると、 $\overline{(A \cup B)} \cap C$ も事象である。

根元事象を一個の要素からなる集合とみたとき、上の合成演算によって得られる合成事象は集合となる。したがって、事象には確率が定義できる。確率 $P(E)$ は事象 E の起こりやすさを示す値であり、つぎの(i)~(iii)の性質をもつ。なお、このような性質(コルモゴロフ(Kolmogorov)の公理)をもつ量は、より一般には測度(measure)と呼ばれる。

E, F を全事象 U に含まれる任意の事象、 ϕ を空事象とする。

- (i) $0 \leq P(E) \leq 1, \quad P(\phi) = 0, \quad P(U) = 1$
- (ii) $E \subset F$ ならば、 $P(E) \leq P(F)$ 。
- (iii) $E \cap F = \phi$ ならば、 $P(E \cup F) = P(E) + P(F)$ 。

【例 2.2】 サイコロを投げたとき、出た目の数が4またはそれ以下になる確率を求める。

この場合の全事象は $U = \{1, 2, 3, 4, 5, 6\}$ である。4またはそれ以下になる事象は $E = \{1\} \cup \{2\} \cup \{3\} \cup \{4\} = \{1, 2, 3, 4\}$ であるから、 $P(E) = 2/3$ である。

2.2 条件付き確率

いま、100 人の人(U)が自動車運転免許の試験を受けたとする。このうち 40 人(Y)が合格し、60 人(N)が不合格であった。合格者のうちの 32 人、および不合格者のうちの 40 人は自動車学校の卒業生(G)であった。自動車学校の卒業生の合格率はいくらであろうか。

集合 E の要素の数を $n(E)$ で表わすと、図 2.1 のベン図 (Venn diagram) より、答は

$$P(Y|G) = \frac{n(Y \cap G)}{n(G)} = \frac{P(Y \cap G)}{P(G)} = \frac{32}{72} = \frac{4}{9}$$

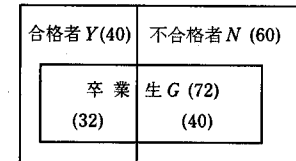


図 2.1 ベン図。()内は人数

となる。

なお一般に、

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad (2.1)$$

は、事象 B が起こったとき、事象 A が起こる確率であり、条件付き確率と呼ばれる。また、 $P(A \cap B)$ は A と B の積事象、すなわち A と B が同時に起こる確率であり、 A と B の結合確率と呼ばれる。同様に

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \quad (2.2)$$

である。式(2.1)の分子に式(2.2)の分子を代入すれば、一般に

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (2.3)$$

が成り立つ。これは、ベイズの定理(Bayes' theorem)と呼ばれる。事象 A が起こる確率は事前には $P(A)$ であるが、事象 B が起こったことを知った後は、その確率は $P(A|B)$ に変化する。そのような意味で $P(A)$ は事前確率、 $P(A|B)$ は事後確率と呼ばれることもある。それらの間の変換を与えるのがベイズの定理である。

また、 U が互いに素な n 個の事象 $A_1, A_2, \dots, A_n$ からなるとき、すなわち、 $U = A_1 \cup A_2 \cup \dots \cup A_n = \sum_{i=1}^n A_i$ のとき、

$$P(B) = \sum_{i=1}^n P(A_i \cap B) = \sum_{i=1}^n P(B|A_i)P(A_i)$$

であるから、ベイズの定理は

$$P(A_j|B) = \frac{P(B|A_j)P(A_j)}{\sum_{i=1}^n P(B|A_i)P(A_i)} \quad (2.4)$$

のように表現できる。なお、この式の右辺の分母は $\sum_{j=1}^n P(A_j|B) = 1$ とするための正規化係数とみなすことができる。

【例 2.3】 三つの壺がある。これらを U_1, U_2, U_3 とする。 U_1 には黒玉 1 個と白玉 1 個、 U_2 には黒玉 3 個と白玉 2 個、 U_3 には黒玉 2 個と白玉 1 個がある。いま、運まかせに一つの壺を選び、その中から玉を 1 個取り出す。その玉が黒であったとき、それが U_1 から取り出された確率を求めよ。

解 壺 U_i から玉が取り出されるという事象を U_i 、取り出された玉が黒、白であるという事象をそれぞれ黒、白で表わすと、

$$P(U_1|\text{黒}) = \frac{P(\text{黒}|U_1)P(U_1)}{\sum_{i=1}^3 P(\text{黒}|U_i)P(U_i)} = \frac{\frac{1}{2} \times \frac{1}{3}}{\left(\frac{1}{2} + \frac{3}{5} + \frac{2}{3}\right) \times \frac{1}{3}} = \frac{15}{53}$$

2.3 完全事象系

いま標本空間 U が、図 2.2 のように、確率 P_k で生起する互いに排他的な n 個の事象 E_k に分割できるとする。このとき、このような事象の集合は**完全事象系**と呼ばれ、

$$\left| \frac{E}{P} \right| = \left| \begin{array}{cccc} E_1 & E_2 & \cdots & E_n \\ P_1 & P_2 & \cdots & P_n \end{array} \right| \quad (2.5)$$

のように表現される。ただし、 $\sum_{i=1}^n E_i = U$ (全事象)、 $\sum_{i=1}^n P_i = 1$ である。たとえば、サイコロを振ったとき i の目がでる事象を E_i で表わすと、

$$\left| \frac{E}{P} \right| = \left| \begin{array}{cccc} E_1 & E_2 & \cdots & E_6 \\ \frac{1}{6} & \frac{1}{6} & \cdots & \frac{1}{6} \end{array} \right|$$

のように表現される。

いま、ある事象の起こったことを観測した場合、何らかの情報を得たと考えることができる。確率 P で起こる事象を観測した場合、それによって得られる情報の量(これを自己情報量という)を

$$I = -a \log_k P \quad (2.6)$$

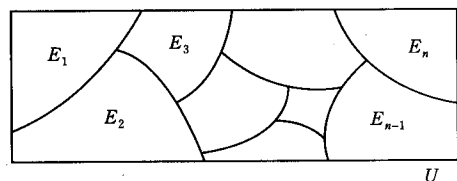


図 2.2 完全事象系

で定義するとつごうがよい。対数を用いることにより、二つの独立な事象を観測した場合、結合事象の生起確率はそれぞれの生起確率の積で表わされるから、自己情報量はそれぞれの自己情報量の和になる。さらに、定数は一般に $a=1, k=2$ にとられる。後で述べるように、定数をこのように決めることにより、デジタル情報や記憶の単位であるビット(bit)との整合性がとれるのである。このとき、自己情報量の単位も**ビット**となる。すなわち、

$$\begin{aligned} I &= -\log_2 P \quad [\text{ビット}] \\ &= \log_2 \frac{1}{P} \quad [\text{ビット}] \end{aligned} \quad (2.7)$$

したがって、非常にめずらしい事象を観測したとき、 P は非常に小さいから I は大きくなる。たとえば、1000 分の 1 の確率で当たるくじに当たったという情報は $I = \log_2 1000 \cong 10$ [ビット] の自己情報量をもっている。逆にそのくじにはずれたという情報はめずらしくない情報であり、自己情報量はほとんど 0 である。

さらに、完全事象系 E の事象が一つ起こったことを知ったときに平均的に得られる情報量を、それから得られる自己情報量の期待値で定義する。これを**平均情報量**、または単に**情報量**という。すなわち、得られる情報量 $\bar{I}$ は

$$\bar{I} = \sum_{i=1}^n (-P_i \log P_i) \quad [\text{ビット}] \quad (2.8)$$

で与えられる。

一方、完全事象系 E は最初 $H(E)$ なるあいまいさをもっていたとする。事象の観測後はこのあいまいさはなくなり、0 となる。その差が観測により得られる情報量であるとする。このあいまいさ $H(E)$ を完全事象系 E の**エントロピー**と呼ぶ。すなわち、エントロピーは情報量に等しく、

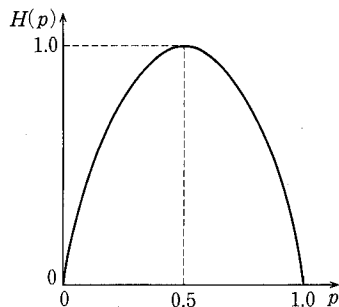
$$H(E) = \sum_{i=1}^n (-P_i \log P_i) \quad [\text{ビット}] \quad (2.9)$$

で定義される。特に断わらない限り、対数の底は 2 である。ときには自然対数が使われることもあるが、そのときにはエントロピーの単位は**ナット(nat)**となる。

【例 2.4】 いま、1, 0 の 2 種類の記号をそれぞれ確率 p および $(1-p)$ で発生する記憶のない情報源を考える。この情報源のエントロピーは

$$H(p) = -p \log p - (1-p) \log (1-p) \quad (2.10)$$

となる。これを p の関数として図示すると図 2.3 のようになる。この図より、一方

図 2.3 2元情報源のエントロピー $H(p)$

の生起確率が0または1なら、エントロピーは0であることがわかる。すなわち、その情報源からは常に0(または1)が発生し、あいまいさは0である。また、 $p=0.5$ 、すなわち0と1の両者が等確率で生じるときエントロピーは最大値1をとる。一般に、 n 個の事象があるとき、それらの生起確率がすべて等しい、すなわち $P_1=P_2=\dots=P_n=1/n$ のとき、エントロピーは最大値 $H=\log n$ をとる。エントロピー関数の値は巻末の付表1に示す。

上の例で述べたように、1, 0の2種類の記号のどちらかをとり二者択一の状態は最大1ビットの情報量をもつ。これはちょうど2元系列の1記号もしくは2進数の1桁や、それらのための記憶単位のビットに対応している。しかしながら、エントロピーは確率に関係した量であり、後者と区別するためエントロピーの単位をシャノンということもあるが、一般には普及していない。

なお、4.5, 4.8節では、エントロピーは情報源を一意的に復号可能なように符号化するときの2元符号の符号長の下限として自然な形で誘導される。このことから、エントロピーが情報量・あいまいさの測度として実質的な意味をもっていることがわかる。また詳しくは述べないが、このエントロピーは熱力学における分子などの無秩序さを表わすエントロピーと同じ概念である。

演習問題 2.

- 2.1 ある大学の自宅通学の学生の割合は2/3であり、その3/4がクラブ活動をしている。また、下宿している学生の割合は1/3であり、その1/2がクラブ活動をしている。クラブ活動をしている学生のうち、自宅通学の学生の割合はいくらか。
- 2.2 1/6の確率で起こる事象を観測したとき、平均的に得られる情報量はいくらか。ただし、 $\log_2 3=1.58$, $\log_2 5=2.32$ である。

3

情報源

情報発生源には各種のものがあるが、本章ではそれらを取扱いやすい形にモデル化する。特に、出力がいつでも同じ統計的性質をもつ定常情報源や、一定以上過去には影響されないマルコフ情報源が重要であり、それらの性質について述べる。

3.1 情報源の統計的表現

情報源記号の集合 $A=\{a_1, a_2, \dots, a_M\}$ のうちからどれか一つずつ確率的に記号を順次発生・出力する離散的 M 元情報源を考える。時刻 i ($i=0, 1, 2, \dots$) において、情報源記号は A のうちのどれかをとり、それを確率変数 X_i で表わす。 A は情報源アルファベットと呼ばれる。時刻 $i=0, 1, 2, \dots, n-1$ において、出力がそれぞれ $x_0, x_1, x_2, \dots, x_{n-1}$ である結合確率を

$$P_{X_0, X_1, X_2, \dots, X_{n-1}}(x_0, x_1, x_2, \dots, x_{n-1}) \quad (3.1)$$

で表わす。確率変数名を省略し、

$$P(x_0, x_1, x_2, \dots, x_{n-1})$$

と略記することもある。

3.2 記憶のない定常情報源

各時刻における情報源記号の発生が他の時刻の発生記号とは独立であるとき、この情報源を記憶のない情報源という。定常ということは、出力記号の発生確率が、いつそれを観測するかによらないで同じであるということである。

記憶のない定常情報源においては

$$P(x_0, x_1, \dots, x_{n-1}) = \prod_{i=0}^{n-1} P(x_i) \quad (3.2)$$

である。サイコロを振る場合がこの例である。

3.3 エルゴード情報源

定常情報源に加えてさらに重要な性質として、エルゴード性(ergodicity)がある。情報源がエルゴード的(ergodic)とは、十分長い情報源系列においては、その情報源の統計的性質が完全に現れているものをいう。

確率変数 $X^\dagger$ の任意の関数を $f(X)$ とする。 X は確率的に変動するから、 $f(X)$ も確率変数である。この平均値は

$$\langle f(X) \rangle = \sum_x f(x) P(x) \quad (\text{集合平均}) \quad (3.3)$$

である。これは、無数の同じ性質をもった情報源についての平均値であり、集合平均(ensemble average)と呼ばれる。一方、経時的に特定の一つの情報源出力 $x_0, x_1, x_2, \dots$ を観測し、それらの平均を求めたのが時間平均(time average)である。それは

$$\overline{f(X)} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} f(x_k) \quad (\text{時間平均}) \quad (3.4)$$

で与えられる。集合平均と時間平均は一致するとは限らないが、エルゴード情報源ではそれらが一致する。すなわち、

$$\overline{f(X)} = \langle f(X) \rangle$$

が成立する。この様子を図 3.1 に示す。エルゴード情報源では、多数の同じ性質をもつ情報源のある時刻における出力^{††}の統計的性質と、特定の一つの情報源からの出力^{†††}の統計的性質が等しい。

$f(x)$ は x の任意の関数であるが、いちばん簡単な $f(x) = x$ の場合を考えると $\overline{f(X)} = \bar{X}$ は X の時間平均となり、 $\langle f(X) \rangle = \langle X \rangle$ は X の集合平均となる。また、 $f(x) = x^2$ とおくと、 $\overline{f(X)} = \bar{X}^2$ 、すなわち二乗時間平均、…等となる。

† 本節においては、大文字の X は確率変数を表わし、集合的・確率的な意味をもつが、小文字の x はそのような性質をもたない単なる変数を表わす。

†† 位相的集合(phase ensemble)、または単に集合(ensemble)と呼ばれる。

††† 時間的集合(time ensemble)と呼ばれる。

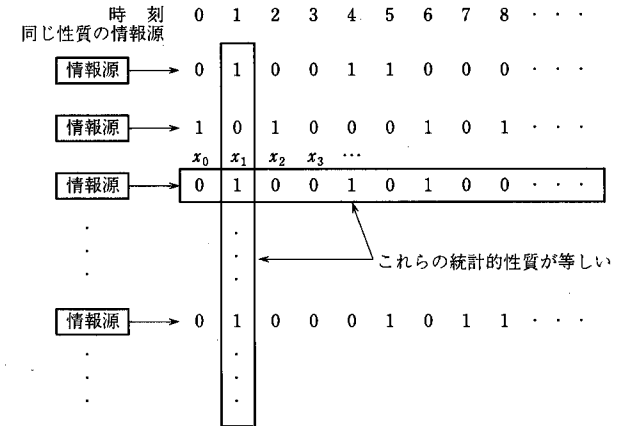


図 3.1 エルゴード情報源

一般に、 $\overline{f(X)}$ は情報源を一つもってきてそれを観測すればよいので比較的簡単に求まるが、集合平均はそのような情報源を多数集めてこななければならない、たいへんである。両者が一致するエルゴード情報源では時間平均だけで話がすむのでこの性質はたいへん有用な性質である。

例としてサイコロを振る場合を考えてみよう。 X を出る目の数とし、 $f(x) = x$ としておく。 X の集合平均は、多数のサイコロを振ってそれらの平均を求めたものであり、

$$\langle X \rangle = \frac{1+2+3+4+5+6}{6} = 3.5$$

となる。一方、自分の持っているただ一つのサイコロを何度も振った場合に出る目の数の平均(時間平均)はやはり

$$\bar{X} = 3.5$$

となる。すなわち、この情報源はエルゴード的である。この例が示すように、記憶のない定常情報源はエルゴード的である。

これに対して、エルゴード性を満たさない情報源もある。たとえば、確率 $1/2$ ずつで 0 だけからなる系列と、 1 だけからなる系列のどちらかを発生する情報源を考えてみよう。この系列の集合平均は $1/2$ である。ところが、時間平均は 0 または 1 のどちらかである。したがって、この情報源はエルゴード的ではない。

3.4 記憶のある情報源

本節では、記憶のある情報源の代表例として情報源記号のそれぞれが各時刻で独立に生起するのではなく、その時刻をさかのぼる m 個の記号が何であったかにより依存し、それ以前にどのような記号が発生したかによらない情報源を考える。このような情報源を m 重マルコフ (Markov) 情報源という。特に、 $m=1$ の場合を単純マルコフ情報源という。また、記憶のない情報源は 0 重マルコフ情報源と考えることができる。

m 重マルコフ情報源において情報源記号の数を N とすると、任意の時刻で出力記号の生起確率に影響する過去の記号系列の数は N^m 個ある。これらを状態 (state) としてとらえ、情報源の確率的な性質を各状態において各情報源記号の発生する条件付き確率によって表わすことを考える。

【例 3.1】 情報源記号 $\{0, 1\}$ からなる単純マルコフ情報源を考える。ある時刻における情報源出力が与えられたとき、その次の時刻における情報源記号発生確率を

$$\begin{aligned} P(0|0) &= 0.4, & P(1|0) &= 0.6, \\ P(0|1) &= 0.2, & P(1|1) &= 0.8 \end{aligned}$$

とする。

このようなマルコフ情報源の動作は、状態図 (state diagram) または遷移図 (transition diagram) と呼ばれる図 3.2 のような図で表わすとわかりやすい。図において、丸印は各状態を表わし、矢印は可能な状態遷移と方向を表わす。また、矢印につけられている記号と数字は状態遷移にともなって発生する記号とその確率を表わす。

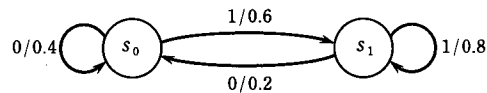


図 3.2 例 3.1 の単純マルコフ情報源の状態図

【例 3.2】 情報源記号 $\{0, 1\}$ からなる 2 重マルコフ情報源を考える。相続く二つの情報源記号の列が与えられたとき、その次の時刻における情報源記号発生確率を

$$\begin{aligned} P(0|00) &= 0.4, & P(1|00) &= 0.6, \\ P(0|01) &= 0.2, & P(1|01) &= 0.8, \\ P(0|10) &= 0.3, & P(1|10) &= 0.7, \\ P(0|11) &= 0.9, & P(1|11) &= 0.1 \end{aligned}$$

とし、相続いて発生した情報源記号 00, 01, 10, 11 をそれぞれ状態 s_0, s_1, s_2, s_3 で表わす。このとき、たとえば状態 $s_1=01$ にあるとき、情報源がその次に記号 0 を発生

3.5 一般化されたマルコフ情報源

すれば状態は $s_2=10$ に遷移し、また 1 を発生すれば状態は $s_3=11$ に遷移する。

このような 2 重マルコフ情報源の動作は図 3.3 に示すような状態図で表わされる。

状態間の遷移だけに注目すると、状態遷移はある時刻の状態が s_i であったときに次の状態が s_j になる確率

$$p_{ij} = P(s_j | s_i) \quad (3.5)$$

を要素とする遷移確率行列 T で完全に記述される。上の例では

$$T = \begin{bmatrix} 0.4 & 0.6 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.2 & 0.8 \\ 0.3 & 0.7 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.9 & 0.1 \end{bmatrix} \quad (3.6)$$

となる。

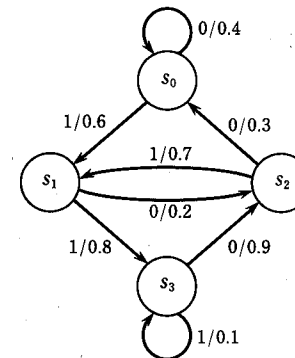


図 3.3 例 3.2 の 2 重マルコフ情報源の状態図

同様に、もっと高次のマルコフ情報源においても、次にどのような状態をとるかの確率は、現在の状態が与えられると完全に定まる。これがマルコフ情報源の大きな特徴である。

マルコフ情報源の遷移確率行列の各行の和は 1 である。また、当然のことであるが各要素は非負である。このような性質をもつ行列は一般にマルコフ行列と呼ばれる。

3.5 一般化されたマルコフ情報源

上に述べたマルコフ情報源では、各状態は直前の m 個の出力に対応していた。したがって、状態間の遷移は制限されたものであった。このような対応をなくした情報源も定義することができる。それらは一般化されたマルコフ情報源、あるいは区別せずに単にマルコフ情報源と呼ばれる。一般化されたマルコ

情報源においても、次にどのような状態をとるかの確率は、現在の状態が与えられると完全に定まる。状態遷移のみに注目した場合、このモデルはマルコフ連鎖と呼ばれることがある。

一般化されたマルコフ情報源はつぎのように定義される。

(1) 情報源は、状態の集合 $S = \{\sigma_1, \sigma_2, \dots, \sigma_N\}$ 、および情報源アルファベット $A = \{a_1, a_2, \dots, a_M\}$ をもち、各時刻においてそのときの状態に応じて出力記号を1個ずつ発生し、つぎに新しい状態に遷移する。

(2) 時刻 k における状態を s_k とするとき、任意の状態 σ_i, σ_j について、つぎの状態遷移確率

$$p_{ij} = P(s_{k+1} = \sigma_j | s_k = \sigma_i) \quad (3.7)$$

が時刻 k に関係なく与えられ、かつこの確率は現在の状態のみに依存するものとする。すなわち、

$$P(s_{k+1} | s_k, s_{k-1}, s_{k-2}, \dots) = P(s_{k+1} | s_k). \quad (3.8)$$

マルコフ連鎖においては出力記号を考えないので、以上の(1)、(2)の定義で十分である。一般のマルコフ情報源においては、状態遷移を行うとき現在の状態と次の状態に応じて出力記号の確率が定まるが、本書ではこれ以上詳しくは述べない。

一般化されたマルコフ情報源は必ずしも m 重マルコフ情報源になっているとは限らない。たとえば、図3.4のマルコフ情報源においては、 m をどんなに大きくとっても、直前の m 個の出力だけではそれ以前の出力と無関係に現在の出力について確率を定められない場合が生じる。たとえば、出力に0が続いた場合、それだけでは現在の状態が s_1, s_2, s_3 のいずれであるか定められない。したがって、次の出力が0である確率は0.3であるか、0.6であるか、0.8であるかは決まらない。

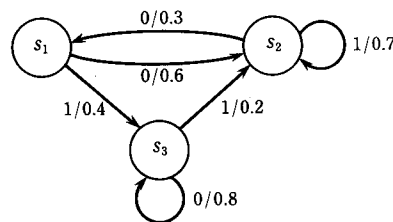


図 3.4 一般化されたマルコフ情報源

しかしながら、このような一般化されたマルコフ情報源も m 重マルコフ情報源と同じように扱える場合が多い。また逆に遷移の拘束がない分だけ見通しが良くなる面もある。そこで、以下では一般化されたマルコフ情報源、マルコフ連鎖について述べることにする。

3.6 状態の分類

マルコフ連鎖には状態図で表わしたとき、ばらばらの図形に分かれるものも考えられる。しかし、本書ではお互いにつながった状態図で表わされるもののみを考える。これを分解不可能なマルコフ連鎖と呼ぶ。この分解不可能なマルコフ連鎖において、ある状態から矢印に沿って出発し、再びもとの状態の帰ってくるのできる場合と、それができない場合がある。前者のような性質の状態だけを集めた部分を閉じた状態部分、また後者のような性質の状態のみを集めた部分を過渡状態部分と呼ぶ。

図3.5において、状態 $\{s_0, s_1\}$ は過渡状態部分、状態 $\{s_2, s_3, s_4, s_5\}$ および $\{s_6, s_7, s_8\}$ はともに閉じた状態部分である。しかしながら、閉じた状態部分 $\{s_2, s_3, s_4, s_5\}$ においては、たとえば時刻 k において状態 s_4 にあったとすると、状態 s_5 は時刻 $k+1, k+3, k+5, \dots$ においてのみ現れ、状態 s_4 は時刻 $k+2, k+4, k+6, \dots$ においてのみ現れる。このように、閉じた状態部分のある部分

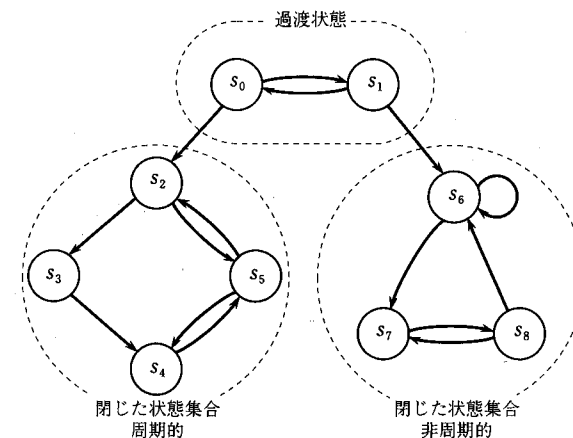


図 3.5 マルコフ情報源の状態の分類

集合が、ある周期的な時刻においてのみ現れうるという場合、この状態部分を**周期的状態部分**という。一方、閉じた状態部分 $\{s_6, s_7, s_8\}$ においては、はじめにどのような状態にあっても、ある時間が経った後にはどの状態にあることも可能となる。このような状態部分を**非周期的状態部分**という。

閉じた状態部分がいくつかある場合、それらは別々に扱えばよい。なぜなら一度閉じた状態部分に入ればそこから出ることはないからである。したがって、単一の閉じた状態部分からなるマルコフ情報源が基本となる。このようなマルコフ情報源を**既約マルコフ情報源**という。さらに、非周期的な既約マルコフ情報源を**正規マルコフ情報源**という。次節で詳しく説明するが、正規マルコフ情報源は十分時間が経過すればエルゴード情報源とみなすことができる。

以上の分類を図に示せば図 3.6 のようになる。

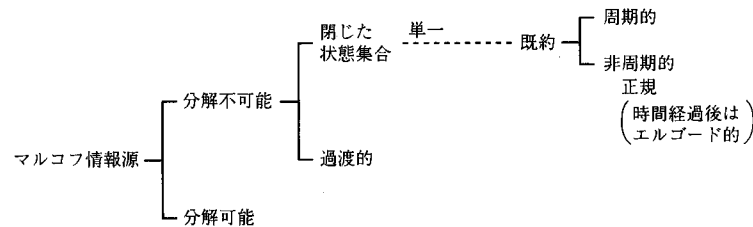


図 3.6 マルコフ情報源の形態および状態の分類

3.7 マルコフ情報源の定常分布

状態図の遷移確率行列が

$$P = |p_{ij}| \quad (n \times n) \quad (3.9)$$

で与えられるとする。いま、状態 s_i から出発し、 t 時刻後に状態 s_j に到達する確率を p_{ij}^t とすると、明らかに

$$p_{ij}^1 = p_{ij}, \quad p_{ij}^2 = \sum_{k=1}^n p_{ik}^1 \cdot p_{kj}, \quad p_{ij}^3 = \sum_{k=1}^n p_{ik}^2 \cdot p_{kj}, \quad \dots$$

$$p_{ij}^t = \sum_{k=1}^n p_{ik}^{t-1} \cdot p_{kj}$$

である。これらを行列表記にすると次のようになる：

$$w^t = w^0 P^t \quad (3.10)$$

3.7 マルコフ情報源の定常分布

ここに

$$P^t = |p_{ij}^t| \quad (3.11)$$

は行列 P の t 乗である。また

$$w^t = (w_1^t, w_2^t, \dots, w_n^t) \quad (3.12)$$

は時刻 t においてマルコフ情報源が各状態にある確率を与える**状態確率分布**である。状態確率分布のことを単に**状態分布**ともいう。 w^0 は時刻 0 の初期状態確率分布である。

さて、マルコフ情報源において十分時間が経ち、各状態にある確率が初期状態に無関係にそれぞれ一定値になったとき、そのマルコフ情報源は**定常状態**に達したという。一般のマルコフ情報源では定常状態に到達可能とは限らないが、正規マルコフ情報源では、どのような初期状態確率分布からでも十分時間が経てば、定常状態に達する。

いま、定常状態確率分布を

$$w = (w_1, w_2, \dots, w_n) \quad (3.13)$$

とすると、定常状態では次の時刻の状態確率分布も w でなければならないから、

$$w = wP \quad (3.14)$$

の関係を満たさなければならない。実際、正規マルコフ情報源ではこの式を満たす状態確率分布がただ一つ存在し、十分時間が経った後の定常分布に一致することが示される。さらに、正規マルコフ情報源はエルゴード的であることも示される。

〔例 3.3〕 図 3.7 に示す正規マルコフ情報源を考える。遷移確率行列 T は、

$$T = \begin{bmatrix} 0.4 & 0.6 & 0.0 \\ 0.2 & 0.4 & 0.4 \\ 0.2 & 0.8 & 0.0 \end{bmatrix} \quad (3.15)$$

で与えられる。この定常状態分布を $w = (x, y, z)$ とすると、

$$w = wT \quad (3.16)$$

および

$$x + y + z = 1 \quad (3.17)$$

より、

$$x = \frac{1}{4}, \quad y = \frac{15}{28}, \quad z = \frac{3}{14}$$

となる。したがって、定常状態分布は $w = (1/4, 15/28, 3/14) = (0.25, 0.536, 0.214)$ である。

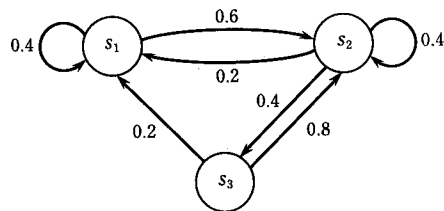


図 3.7 例 3.3 の正規マルコフ情報源

いま、初期状態分布を $w^0 = (1, 0, 0)$ とする。このとき、時刻 t における状態分布 w^t は

$$\begin{aligned} w^1 &= (0.4, 0.6, 0.0), \\ w^2 &= (0.28, 0.48, 0.24), \\ w^3 &= (0.256, 0.552, 0.192), \\ &\vdots \end{aligned}$$

となり、順次定常状態分布 w に近づくことがわかる。

演習問題 3.

3.1 (a)~(d)の各事象を次の

- (1) 記憶のある情報源
- (2) 非定常情報源
- (3) 記憶のない定常情報源
- (4) 非エルゴード情報源

のどれかに一つずつ必ず分類するとする。最もよく説明できる対応を示せ。

- (a) 人口が急激に増えているある新興都市の毎年1月1日の人口。
- (b) サイコロをつづけて振ったときに次々とする目の数。
- (c) 東京の毎日正午の気温。
- (d) ある病院にその日入院した患者の血糖値の1時間ごとの値。

3.2 遷移確率行列が

$$T = \begin{bmatrix} 0.6 & 0.4 & 0.0 \\ 0.3 & 0.0 & 0.7 \\ 0.2 & 0.0 & 0.8 \end{bmatrix}$$

で与えられるマルコフ情報源の定常状態分布を求めよ。

4

情報源符号化

情報源符号化とは、情報源記号を通信路に適した符号に変換することである。その際、できるだけ短い符号で効率よく情報を伝送できるようにする必要がある。情報源からは英文字や画像の白黒パターンなどが発生する。これらは次にくる文字や信号を、それまでの状態からある程度予測できることが多く、記憶をもった情報源とみなすことができる。この出力には、予測できる部分とできない部分がある。予測できる部分は新しい情報ではなく、予測不可能の部分が純粋に新しい情報(エントロピー)である。本章では、一意の復号可能な符号の平均長は情報源のエントロピーにいくらでも近づき得るが、それより短くできないことを示す。これにより、2章で半ば直観的に定義した情報量が符号の長さとして明確な意味をもってくる。

4.1 離散的信息源の符号化

通信路を通して情報を伝送するには、まず情報源から発生する情報源記号列を情報源符号により、通信路に適した符号系列に変換しなければならない。

たとえば、キーボードから打ち込まれた英文は、各文字7ビットのASCII(アスキー)符号で符号化される。実際にMODEM(モデム)を使って電話回線上に送り出されるときには、通信路における誤りを検出するため、この7ビット符号にさらに1ビットのパリティ検査ビットを付けて8ビットにして送り出されることが多い。しかしながら、検査ビットを付けることは後で述べる通信路符号化に属することであり、本章では考えないことにする。

表 4.1 情報源符号化

情報源記号	確 率	瞬 時 符 号			非瞬時符号
		等長符号 C_1	コンマ符号 C_2	C_3	C_4
A	0.6	0 0	0	0	0
B	0.25	0 1	1 0	1 0	0 1
C	0.1	1 0	1 1 0	1 1 0	0 1 1
D	0.05	1 1	1 1 1 0	1 1 1	1 1 1
平均符号長		2	1.60	1.55	1.55

上に述べた ASCII 符号による符号化は、単純に英文の各文字に 7 ビットの符号語を割り当てている。したがって、たとえば

I am a boy.

という文章は、空白、句読点を含めて 11 文字であるから、基本的には合計 77 ビットのビット列に符号化され、通信線に送り出される。受信側では、受け取った 77 ビットのビット列を 7 ビットづつに区切って復号を行い、もとの英文に戻す。

効率の点からこのような符号化法よりもっとよい符号化法はないであろうか。本章ではこの点に関し、主としてできるだけ短い符号長の観点から、符号化方式を考察するとともに、その理論的な限界を明らかにする。

いま簡単化のために、情報源アルファベットは $\{A, B, C, D\}$ の 4 記号からなり、各記号は表 4.1 に示された確率で出現するとする。これらは C_1, C_2, C_3, C_4 の符号化法により符号化される。得られた具体的な符号表現、たとえば C_2 における 110 は符号語と呼ばれる。まず、情報源符号化方式の基本的に必要な条件として、もとの記号系列が一意的に復号可能であることが望ましいことは論をまたないであろう。 C_1, C_2, C_3, C_4 はいずれもこの条件を満たしている。

符号 C_1 は各符号語の長さが等しいので、等長符号と呼ばれる。さきに述べた ASCII 符号の例がそれにあたる。これに対して、 C_2, C_3, C_4 のように出現確率の多い記号に対しては短い符号語を割り当て、出現確率の小さい記号に対しては長い符号語を割り当てる符号化方式を用いると、1 記号当りの平均符号長を短くすることができる。東京や大阪のような大都市には短い市外局番を与え、地方には長い市外局番を与えることにより、平均として市外局番のためのダイヤル回数を減らすことができるのと同じ原理である。このような符号は非

等長符号と呼ばれる。

C_2 は符号語の終わりが必ず 0 となっているので、符号語の切れ目は簡単にわかり、復号が容易である。この場合、0 がコンマの役割をしているので、コンマ符号と呼ばれる。一方、 C_2 において符号語 1110 における最後の 0 は識別のためには必ずしも必要ではないので、それを省いたのが C_3 である。この結果、平均符号長が少し小さくなっている。一方、 C_4 は C_3 における 0 の位置を逆にしたものである。平均符号長は変わらないが、復号時に問題が生じる。たとえば、01111110... という符号系列を受けたとき、

0 1 1 1 1 1 1 0 ...

A D D ...

と一意的に復号できるのであるが、2 番目の 0 を受けるまでは

0 1 1 1 1 1 1 0 1 1 1 1 1 1

B D D C D

という可能性も残されている。したがって、復号には時間遅れが伴う。これに対して、 C_1, C_2, C_3 では符号語の区切りがくれば直ちに迷うことなく復号できる。このような意味で、 C_1, C_2, C_3 は瞬時符号、また、 C_4 は非瞬時符号と呼ばれる。

これらのことから、情報源符号化方式として好ましい条件をまとめると、

- (1) 一意的に復号可能、かつ瞬時符号である。
- (2) 1 情報源記号当たりの平均符号長ができるだけ短い。
- (3) 装置、特に復号装置があまり複雑とならない。

最後の条件(3)は実用上重要になってくる。

4.2 瞬時符号であるための必要十分条件

ある符号が瞬時的であるかどうかを見分けるための条件としては、つぎに示す接頭語条件が有用である。

いま、ある符号語を $X = x_1 x_2 \cdots x_n$ としたとき、この符号語の先頭部分 $Y = x_1 x_2 \cdots x_m$ ($m \leq n$) を X の接頭語(prefix)という。

定理 4.1 (接頭語条件)

ある符号が瞬時的であるための必要十分条件は、その符号のどの符号語も、他の符号語の接頭語になっていないことである。

証明 十分性： もし、これから判定すべき受信符号系列の先頭に符号語が見い出されたら、仮定によりそれは他の符号語の接頭語ではない。したがって、直ちに正しく復号できる。このようにして、時間遅れなく逐語的に復号できる。

必要性： いま符号系のある語 X が、他の符号語 Y の接頭語になっていると仮定する。もし、これから判定すべき受信符号系列の先頭に符号語 X が見い出されたら、それは符号語 X そのものか、あるいは符号語 Y の初めの部分のどちらかである。このいずれであるかは受信符号系列をもっと先まで調べなければ決定できない。したがって、この符号語は瞬時的ではない。つまり、瞬時的であるためには接頭語条件を満足することが必要である。

図 4.1 は、瞬時符号が一意的復号可能な符号の一部になっていることを示している。

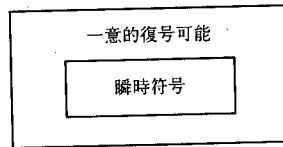


図 4.1 瞬時符号

4.3 符号の木

符号の木を用いると瞬時符号をよく理解できる。4.1 節の C_1, C_2, C_3, C_4 に対する符号の木を図 4.2 に示す。左端の節点は根 (root)、節点から出ている各線分は枝 (branch)、節点のうち、枝が一本も出ていない節点を葉 (leaf) と呼ぶ。また、根から 1 本の枝を通して達する節点を 1 次の節点という。根は 0 次の節点である。

各節点から出る枝には、それぞれ符号アルファベットの互いに異なる記号が一つずつ対応づけられている。図 4.2 は一般的な $\{0, 1\}$ からなる 2 元符号アルファベットの場合を示しており、各枝には 0 と 1 が対応づけられている。したがって、根から出発して 1 本の枝を経由して到達した節点には長さ 1 の系列が対応づけられている。

さて、図 4.2 からわかるように、瞬時符号 (C_1, C_2, C_3) においては符号語はすべて葉に対応づけられるのに対し、非瞬時符号 (C_4) では葉以外の節点にも符号語が対応づけられる。葉以外の節点は他の符号語の接頭語である。したがって、符号の木では接頭語条件は符号語がすべて葉に対応づけられるという形で表現される。すなわち、瞬時符号となるための必要十分条件は、符号語がすべて葉に対応づけられることである。

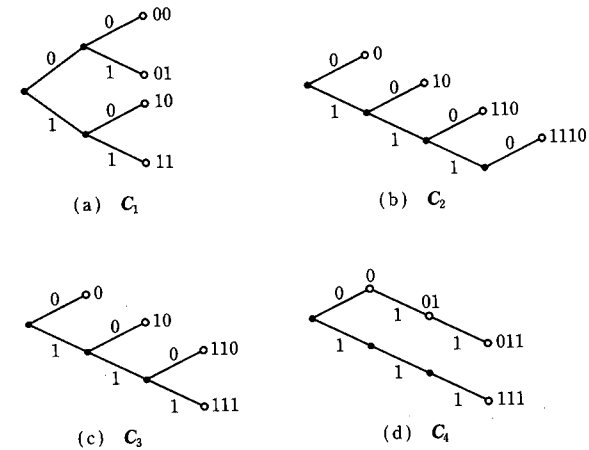


図 4.2 符号 C_1, C_2, C_3, C_4 に対する符号の木。
 C_1, C_2, C_3 は瞬時符号, C_4 は非瞬時符号

4.4 クラフトの不等式

上に述べたいいくつかの符号系のうち、瞬時符号では C_3 が平均符号長がいちばん短かった。平均符号長がもっと短い符号系はつくれないのであろうか。そのような可能性を調べる方法がクラフトの不等式である。

いま、図 4.3 に示すように、2 元符号の符号の木を考える。根から 1 の養分を流し込む。葉でない各節点では枝に流れる養分は $1/2$ ずつに分かれる。ある葉の次数を l としたとき、その葉には 2^{-l} の養分が到達する。瞬時符号においては、すべての符号語は葉に存在する。したがって、長さが $l_1, l_2, \dots, l_M$ なる M 個の符号語からなる符号が、瞬時符号となるためには、

$$2^{-l_1} + 2^{-l_2} + \dots + 2^{-l_M} \leq 1 \quad (4.1)$$

補助定理 (シャノンの補助定理)

 $r_1, r_2, \dots, r_M$ を

$$r_1 + r_2 + \dots + r_M \leq 1 \quad (4.8)$$

を満たす任意の非負の数とする。ただし、 $p_i \neq 0$ のときは $r_i \neq 0$ とする。
このとき、不等式

$$-\sum_{i=1}^M p_i \log r_i \geq -\sum_{i=1}^M p_i \log p_i = H(S) \quad (4.9)$$

が成立する。等号は $p_i = r_i$ ($i=1, 2, \dots, M$) のとき、かつそのときに限り成立する。

証明 図 4.4 のように $y = \log_e x$ の $x=1$ における接線を考えると、

$$\log_e x \leq x - 1.$$

ただし、等号は $x=1$ においてのみ成立する。ここで、 $x = r_i/p_i$ を代入すると

$$\log_e \left(\frac{r_i}{p_i} \right) \leq \frac{r_i}{p_i} - 1$$

両辺に p_i をかけて、 i について加えれば、

$$\begin{aligned} \sum_{i=1}^M p_i \log_e \left(\frac{r_i}{p_i} \right) &\leq \sum_{i=1}^M (r_i - p_i) \\ &= \sum_{i=1}^M r_i - 1 \leq 0. \end{aligned} \quad (4.10)$$

したがって、

$$\sum_{i=1}^M p_i \log_e r_i - \sum_{i=1}^M p_i \log_e p_i \leq 0$$

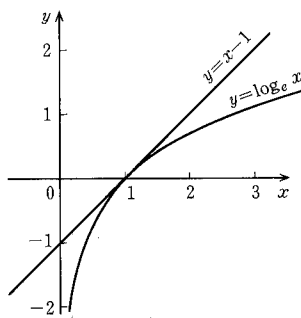


図 4.4 $\log_e x \leq x - 1$ の説明

が成立する。これらは、底を 2 としても係数が異なるだけであるから、同様の式を得る。これより、補助定理の式(4.9)を得る。なお、等号はすべての i について $r_i/p_i=1$ 、すなわち $r_i=p_i$ のとき、かつそのときに限り成立する。■

この補助定理を使って、定理 4.3 を証明する。

証明 まず、

$$r_i = 2^{-l_i} \quad (i=1, 2, \dots, M) \quad (4.11)$$

とおく。明らかに $r_i > 0$ であり、また一意的復号可能性より $\{r_i | i=1, 2, \dots, M\}$ はクラフトの不等式(4.2)を満たす。したがって、補助定理の条件(4.8)を満たす。ここで、 $l_i = -\log_2 r_i$ に注意すると、補助定理より

$$L = \sum_{i=1}^M p_i l_i \geq H(S) \quad (4.12)$$

を得る。なお、等号はすべての i について、 $p_i = r_i$ 、すなわち

$$p_i = 2^{-l_i} \quad (4.13)$$

のとき、かつそのときに限り成立する。

つぎに、式(4.7)を満たす瞬時符号がつくれることを示そう。平均符号長が定理で与えられる限界を達成する符号は一般には、

$$l_i = -\log_2 p_i \quad (4.14)$$

が整数でないため、そのままの形では実現できない。そこで、次善の策として、

$$-\log_2 p_i \leq l_i < -\log_2 p_i + 1 \quad (4.15)$$

なるよう、整数 l_i を定める。このとき、

$$2^{-l_i} \leq p_i$$

であるから

$$\sum_{i=1}^M 2^{-l_i} \leq \sum_{i=1}^M p_i = 1 \quad (4.16)$$

となり、式(4.15)を満たす $l_1, l_2, \dots, l_M$ はクラフトの不等式を満たす。したがって、符号語の長さが $l_1, l_2, \dots, l_M$ となる瞬時符号をつくることができる。このとき、式(4.15)の各辺に p_i をかけて総和をとると

$$H(S) \leq L < H(S) + 1 \quad (4.17)$$

を得る。以上のように、式(4.15)を満たすように l_i を選ぶと、平均符号長が式(4.7)を満たすような瞬時符号を構成できることが導けた。■

なお、 S が記憶をもつ場合でも、 $H(S)$ を後の 4.8 節で述べる 1 次エントロピーと呼び換えるだけでまったく同じ定理が成立する。

【例 4.1】 表 4.1 の情報源につき、平均符号長の下限を求めてみよう。情報源のエントロピーは

$$H(S) = -0.6 \log 0.6 - 0.25 \log 0.25 - 0.1 \log 0.1 - 0.05 \log 0.05 \\ = 1.490 \dots \quad (4.18)$$

である。つぎに、式(4.15)に従って記号 A に対する符号長 l_A を求めると

$$-\log 0.6 = 0.737 \leq l_A < -\log 0.6 + 1 = 1.737$$

より、 $l_A = 1$ となる。同様に $l_B = 2$ 、 $l_C = 4$ 、 $l_D = 5$ が導かれる。このとき、平均符号長は 1.75 となり、 C_2 、 C_3 よりも長くなってしまふ。すなわち、式(4.15)をそのまま用いて符号を構成しても、効率の点では必ずしもよいものではない。しかしながら、それから得られた定理の式(4.7)は、拡大情報源を用いたときに平均符号長が情報源エントロピーにいくらでも近い符号が構成できることを示すときに重要な式となる。

では、情報源記号を一つずつ逐次的に符号化する方法で平均符号長を最小にする手法はないのであろうか。それを与えるのが次に述べるハフマン符号化である。

4.6 ハフマン符号化

ある与えられた情報源 S に対し、情報源記号を一つずつ逐次的に一意的復号可能な符号に符号化するとき、平均符号長を最小にする符号をコンパクト符号(compact code)、または最短符号という。また、そのような符号はコンパクトであるという。ハフマンは具体的に最短符号の構成法を示した。これがハフマン(Huffman)符号である。ハフマン符号は符号の木を葉のほうからつくっていくことにより構成される。以下では、2 元の場合についてハフマン符号のつくりかたを述べる。

【2 元ハフマン符号化法の手順】

(1) M 個の情報源記号に対応する葉をつくる。それぞれの葉には生起確率を記しておく。

(2) 確率の最も小さい 2 個の葉をまとめる新しい節点をつくり、その節点と 2 つの葉を枝で結ぶ。2 本の枝の一方には符号アルファベットの記号 0 を、もう一方には 1 を割り当てる。新しい節点には二つの葉の確率の和を記してお

4.6 ハフマン符号化

く。今度はこの節点を一つの葉とみなし、そこから出ていた枝はなくなったと考える。

(3) 葉が 1 枚しか残っていなければ符号構成は完了している。そうでなければ、手順(2)に戻る。

上記手順では、すべての情報源記号は葉に対応づけられる。したがって、このようにしてつくられたハフマン符号は瞬時符号である。なお、確率の 2 番目に小さい葉が二つ以上あると、その選択に任意性が生じる。また、枝への記号の割り当てかたにもどちらの枝を 0 とするかで 2 通りの任意性がある。したがって、前者だけ考えても、得られるハフマン符号は 1 種類とは限らない。しかしながら、いずれの場合でもその平均符号長は同じである。

【例 4.2】 さきの表 4.1 の例の場合を考える。すなわち、情報源アルファベットを $\{A, B, C, D\}$ 、それらの生起確率をそれぞれ 0.6, 0.25, 0.1, 0.05 とする。このとき、図 4.5 のようなハフマン符号が得られる。これは、表 4.1 の C_3 のものであり、その平均符号長は 1.55 である。

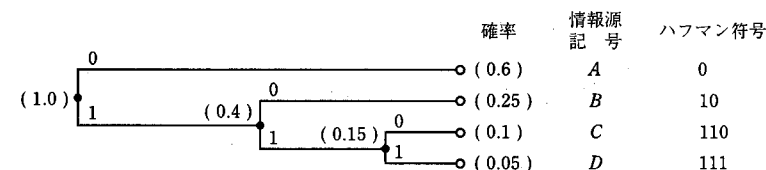


図 4.5 ハフマン符号化

【例 4.3】 情報源アルファベットを $\{A, B, C, D, E, F\}$ 、それらの生起確率をそれぞれ 0.3, 0.25, 0.2, 0.1, 0.1, 0.05 とする。このとき、図 4.6 のようなハフマン符号が得られる。

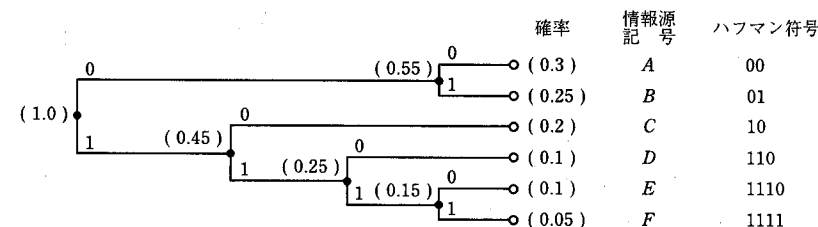


図 4.6 例 4.3 のハフマン符号の構成

補助定理 コンパクトな瞬時符号の木においては、葉以外の節点は必ず分岐している。また、根から最も遠く離れた枝の先(すなわち、最も長い符号語)に確率最小の葉 α があり、それは直前の節点 N で分岐したものである。さらに、 N から出ているもう一つの枝の先には確率が2番目に小さい葉 β を割り当てることができる(図4.7)。

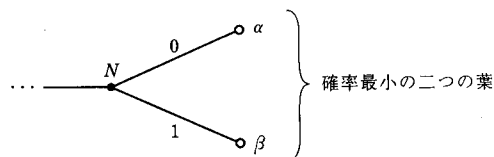


図 4.7 コンパクト符号の最高次の葉

証明 もし、葉以外の節点が分岐していなければ、すなわちその節点から枝が一つしか出ていなければ、その節点を省くことにより、平均符号長がより短い符号をつくることができる。これはコンパクト性の仮定に反する。したがって、葉以外の節点は必ず分岐している。また、ある葉に対して、根からの距離がより短い場所にもっと確率の小さい葉があれば、それと入れ換えることにより、平均符号長がより短い符号をつくることができる。これはコンパクト性の仮定に反する。したがって、確率最小の葉 α は最高次数の葉になっている。なお、 α はいちばん遠いことから、その直前の節点 N から分岐している枝の先には、 α と同じ長さ(次数)の葉 β' がある。 β は確率が2番目に小さな節点 β と異なるかも知れないが、もしそうだとすると上と同様の理由により β は最高次数の葉のうちのどれかである。これらを交換しても平均符号長は変わらないから、 N から出ているもう一つの枝の先に確率が2番目に小さい葉 β を割り当てることができる。■

この補助定理を述べるところにより、図4.8はコンパクトでないことがわかる。また、図4.9に対しては、葉を入れ換えることにより、平均符号長がそれと等しい図4.10のようなコンパクトな瞬時符号をつくることができる。したがって、コンパクトな瞬時符号としては、図4.10のような確率最小の二つの葉が同じ節点から分岐している符号の木のみ考えておけば十分である。

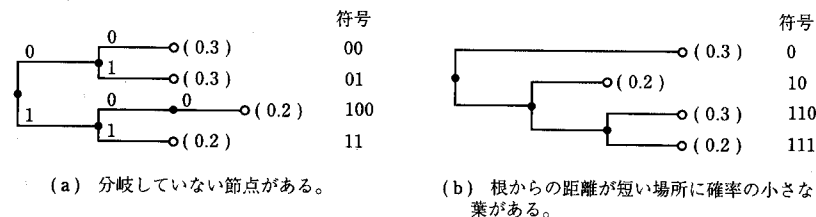


図 4.8 コンパクトでない符号。

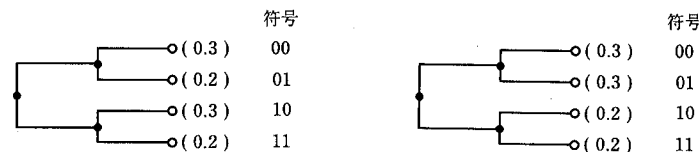


図 4.9 コンパクトであるが、ハフマン符号でない例。

図 4.10 平均符号長が図4.9に等しいハフマン符号化。

定理 4.4 (ハフマン符号のコンパクト性)

ハフマン符号はコンパクト符号である。

証明 数学的帰納法による。対象とする情報源記号の数が各段階で一つずつ減少していることに注目する。もとの情報源を S_0 とし、第 k 番目の段階における情報源 S_k に対する符号の木を T_k としたとき、 T_{k+1} に対する符号 C_{k+1} がコンパクトなら、 T_k に対する符号 C_k もコンパクトであることを証明する。

T_k に対する符号 C_k においては最も生起確率の小さい二つの情報源記号に関して C_{k+1} より1ビット長い。したがって、 T_{k+1} に対する平均符号長 L_{k+1} は

$$L_{k+1} = L_k - p_{k0} - p_{k1}$$

である。ここに、 p_{k0} および p_{k1} はそれぞれ S_k における最も生起確率の小さい二つの情報源記号 s_{k0} および s_{k1} の生起確率である。

いま、 C_{k+1} がコンパクトであったと仮定する。このとき、 C_k もコンパクトであることを証明する。そのため、 C_k がコンパクトでないとし、これと異なるコンパクト符号 C'_k が存在するとする。すなわち、 T_k と同じ葉(確率を含めて)をもち、平均符号長がそれより短い C'_k が存在する。この C'_k の平均符号

長を L_k' とすると

$$L_k' < L_k$$

である。このとき補助定理より、 T_k' には直前で分岐している最も確率の小さな2枚の葉 s_{k0}' および s_{k1}' が存在する。そこで、この2枚の葉をまとめて節点 N' を葉とする新たな符号の木をつくる。この木を T_{k+1}' とする。 T_k' で選ばれた2枚の葉 s_{k0}' および s_{k1}' は、 T_k において選ばれた確率最小の2枚の葉 s_{k0} および s_{k1} と異なるかも知れないが、それらの確率は、一般性を失うことなく $P(s_{k0}) = P(s_{k0}') = p_{k0}$ 、および $P(s_{k1}) = P(s_{k1}') = p_{k1}$ であるとしてとることができる。したがって、 T_{k+1} と T_{k+1}' の葉は(それらの確率を含めて)集合として同じである。このとき、 T_{k+1}' のつくる符号 C_{k+1}' の平均符号長は

$$L_{k+1}' = L_k' - p_{k0} - p_{k1} < L_{k+1}$$

となり、 C_{k+1} がコンパクトであるという仮定に反する。したがって、 C_k はコンパクトである。これで、 C_{k+1} がコンパクトであるならば、 C_k もコンパクトであることが証明された。

さて、最終段階における符号の木は、2枚の葉をもち、それらに0と1が割り当てられ、それ以外の割当てかたはないから、この符号は明らかにコンパクトである。したがって、数学的帰納法により、与えられたもとの情報源に対するハフマン符号の木はコンパクトである。

4.7 ブロック符号化

これまでの、情報源から出力される記号を一つずつ符号化することを考えてきた。しかしながら、このような方法では効率を上げられないことがある。たとえば、 α と β をそれぞれ確率0.9, 0.1で発生する2元情報源を考える。このとき、情報源記号 α, β に対して、ハフマン符号0, 1がそれぞれ割り当てられる。この場合、平均符号長は明らかに1となり、符号化の効果がまったくない。このとき、いくつかの情報源記号をひとまとめ(ブロック)にして符号化することを考える。たとえば、情報源記号を二つまとめて符号化を行う。一般に、情報源 S から出力される情報源記号二つをまとめたものを新しい情報源記号とみなすとき、それを出力とする情報源を2次拡大情報源といい、 S^2 で表わす(図4.11)。このとき、図4.12のようなハフマン符号が得られる。なお、確率は情報源 S に記憶がないとしている。この場合、2次拡大情報源記号1個に対する平均符号長は $1 \times 0.81 + 2 \times 0.09 + 3 \times 0.09 + 3 \times 0.01 = 1.29$ で

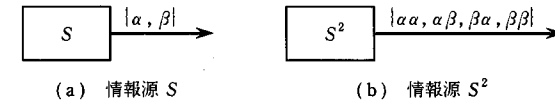


図 4.11 2次拡大情報源 S^2

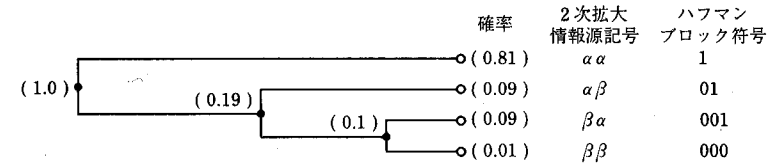


図 4.12 ハフマンブロック符号化

あるから、情報源記号1個に対する平均符号長は $1.29/2 = 0.645$ となり、情報源記号を一つずつ符号化する方法に比べて、効率がおおよそ35%向上する。一般に、情報源 S を n 個まとめたものを n 次拡大情報源とよび、 S^n で表わす。また、 n 次拡大情報源に対して符号化を行う方法をブロック符号化と呼び、そのとき得られる符号をブロック符号という。また、 n 次拡大情報源に対してハフマン符号化を行う方法をハフマンブロック符号化という。

4.8 情報源符号化定理

n 次拡大情報源に対する符号化を用いると、定理4.3により

$$H_1(S^n) \leq L_n < H_1(S^n) + 1 \quad (4.19)$$

を満たす平均符号長 L_n の2元瞬時符号をつくることはできるが、この式の左辺より小さな平均符号長の2元符号は存在しない。ここで、 $H_1(S^n)$ は n 次拡大情報源 S^n の1次エントロピーであり、

$$H_1(S^n) = - \sum_{x_0} \sum_{x_1} \cdots \sum_{x_{n-1}} p(x_0, \dots, x_{n-1}) \log p(x_0, \dots, x_{n-1}) \quad (4.20)$$

で与えられる。ここで、 x_i は情報源 S の情報源記号であり、 x_i に関する和はそのアルファベットすべてについてとるものとする。したがってこのとき、 $(x_0, \dots, x_{n-1})$ に関する和全体は S^n の元のすべてについてとられることになる。なお、情報源の記憶を無視して従来のやりかたで求めたエントロピーが1次エントロピーである。したがって、これまでの記憶をもたない情報源におけるエ

ントロピーは求めかたからすれば1次エントロピーに対応する。また、情報源記号を一つずつ符号化する方法ではそのような1次エントロピーのみ考えておけば十分であった。本節では記憶をもつ情報源のエントロピーは、それらの極限で定義できることを示す。

n 次拡大情報源 S^n のアルファベットは S のアルファベットの n 字の組である。したがって、 S の情報源記号1個当たりの平均符号長は

$$L = \frac{L_n}{n}$$

で与えられる。それゆえ、式(4.19)は

$$H_n(S) \leq L < H_n(S) + \frac{1}{n} \quad (4.21)$$

となる。ここに、

$$H_n(S) = \frac{H_1(S^n)}{n} \quad (4.22)$$

は S の n 次エントロピーと呼ばれ、 S から出力される長さ n の系列を切り出して考えたときの、1情報源記号当たりの平均エントロピーを表わしている。

以上のことより、 S の情報源記号を n 個ずつまとめて符号化することにより、1情報源記号当たりの平均符号長を $H_n(S)$ にいくらでも近づけることはできるが、 $H_n(S)$ より小さくはできないことがわかる。なお、証明は略すが $H_n(S)$ は n に関して単調減少であり、その極限として

$$H(S) = \lim_{n \rightarrow \infty} H_n(S) \quad (4.23)$$

が存在する。この $H(S)$ を情報源が記憶をもつ場合も含めて改めて、情報源 S のエントロピー (entropy) と呼ぶ。したがって、つぎの定理を導くことができる。

定理 4.5 (情報源符号化定理)

情報源 S は、任意の正数 ϵ に対して、1情報源記号当たりの平均符号長 L が

$$H(S) \leq L < H(S) + \epsilon \quad (4.24)$$

となるような2元瞬時符号に符号化できる。しかし、どのような一意復号可能な2元符号を用いても、その平均符号長を $H(S)$ より小さくはできない。なお、この $H(S)$ は式(4.23)で定義される。

情報源符号化定理の教えるところは、記憶をもつ一般の情報源を符号化しようとするとき、2元符号の平均符号長は情報源のエントロピーより小さくすることはできない、ということである。これにより、情報量としてのエントロピーの性質が2元符号の長さという明確な尺度で表現される。また逆に、平均符号長 L を情報源エントロピー $H(S)$ に近づけるためには、十分長い n 個の情報源記号系列をまとめて符号化しなければならない。

【例 4.4】 図4.13のようなマルコフ情報源 S において、定常状態にあるときの、1次エントロピー $H_1(S)$ 、2次エントロピー $H_2(S)$ 、エントロピー $H(S)$ を求めよ。

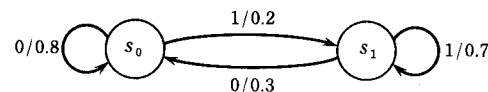


図 4.13 マルコフ情報源 S

解 [1次エントロピーの計算] S の状態の定常分布 (w_0, w_1) は、

$$(w_0, w_1) \begin{bmatrix} 0.8 & 0.2 \\ 0.3 & 0.7 \end{bmatrix} = (w_0, w_1).$$

これから、 $0.3w_1 = 0.2w_0$ 。また、 $w_0 + w_1 = 1$ であるから、

$$w_0 = \frac{3}{5}, \quad w_1 = \frac{2}{5}.$$

したがって、定常状態にある時刻 n での情報源の出力 x_n が 0 である確率は

$$P(x_n = 0) = 0.8w_0 + 0.3w_1 = \frac{3}{5}.$$

ゆえに、

$$H_1(S) = -\frac{3}{5} \log_2 \frac{3}{5} - \frac{2}{5} \log_2 \frac{2}{5} = 0.971.$$

[2次エントロピーの計算]

$$P(x_n = 0, x_{n-1} = 0) = (0.8)^2 w_0 + 0.8 \times 0.3 w_1 = 0.48$$

$$P(x_n = 0, x_{n-1} = 1) = 0.2 \times 0.3 w_0 + 0.7 \times 0.3 w_1 = 0.12$$

$$P(x_n = 1, x_{n-1} = 0) = 0.8 \times 0.2 w_0 + 0.3 \times 0.2 w_1 = 0.12$$

$$P(x_n = 1, x_{n-1} = 1) = 0.2 \times 0.7 w_0 + (0.7)^2 w_1 = 0.28$$

それゆえ、

$$\begin{aligned} H_1(S^2) &= -\sum P(x_n, x_{n-1}) \log_2 P(x_n, x_{n-1}) \\ &= -0.48 \log_2 0.48 - 2 \times 0.12 \log_2 0.12 - 0.28 \log_2 0.28 \\ &= 1.76, \end{aligned}$$

$$H_2(S) = \frac{H_1(S^2)}{2} = 0.878$$

〔エントロピーの計算〕 いま、 S が状態 s_0 にあるときだけに注目すると、この情報源 S は、1, 0 をそれぞれ確率 0.2, 0.8 で発生する記憶のない情報源とみなせる。そのエントロピー $H_{s_0}(S)$ は、

$$H_{s_0}(S) = 0.722.$$

同様に、

$$H_{s_1}(S) = 0.881$$

状態 s_0 にいる確率 $w_0 = 0.6$, s_1 にいる確率 $w_1 = 0.4$ であるから、

$$H(S) = 0.6 \times 0.722 + 0.4 \times 0.881 = 0.786$$

4.9 非等長情報源記号系列に対するハフマン符号化

前節では、相続く情報源記号 n 個をまとめた n 次拡大情報源を考え、そこから出てくる n 次拡大情報源記号に対して符号化を行った。この符号化の問題点は、 $H(S)$ が小さいとき平均符号長 L を $H_n(S)$ に十分近づけるためには、 n を大きくとらねばならないことである。 n が大きいと、符号化回路が複雑となりたいへんである。

〔例 4.5〕 $P(0) = 0.1$, $P(1) = 0.9$ なる記憶のない情報源 S を考える。

$$H(S) = H(0.1) = -0.1 \log_2 0.1 - 0.9 \log_2 0.9 = 0.469$$

である。 L をこの値の 10% 増以内に抑えようとするとき、

$$\frac{1}{n} = 0.0469$$

すなわち、 $n = 21$ 以上に設定する必要がある可能性がある。符号化回路は 2^n 種の情報源記号に対して符号化を行う必要があるが、このような大きな n では符号割当てのための符号表が大きくなりすぎ、実用上不可能である。すなわち、この例のように、情報源記号の生起に片寄りがある場合にはエントロピーが小さく、まとめて符号化を行う長さ n が大きくなりすぎる問題がある。

これまでの n 次拡大情報源では必ず決まった長さ n 個の情報源記号に対して符号語を割り当てていたが、情報源記号の数の増大を抑えるため、必ずしも長さが一定でない非等長情報源記号系列の符号化を考える。この場合、拡大された情報源記号の数はあまり増やさず、平均的に長い情報源記号系列をつくり、その符号化を行うこととする。

4.9 非等長情報源記号系列に対するハフマン符号化

〔例 4.6〕 0, 1 をそれぞれ確率 0.1 および 0.9 で発生する記憶のない情報源 S を考える。図 4.14 のように、情報源記号系列を節点に対応させ、確率のいちばん大きな節点を順次伸ばしていく。葉が 4 枚になるまでの展開の様子を図 4.14(a) から (c) に示す。このとき最後の (c) では情報源記号系列は $\{0, 10, 110, 111\}$ の 4 種類に分類されている。この 4 種の不等長系列により、任意の情報源記号系列は一意的に分解される。たとえば、10110100111 は 10, 110, 10, 0, 111 のように分解される。

いま、上記 4 種の系列の平均長 $\bar{n}$ は

$$\bar{n} = 1 \times 0.1 + 2 \times 0.09 + 3 \times 0.081 + 3 \times 0.729 = 2.71$$

である。上記の構成法では、最も確率の大きな葉を伸ばしていった。これにより、任意の情報源系列を一意的に分解することのできる 4 種の系列としては平均符号長をいちばん長くできた。このことは、確率の大きな枝を順次伸ばしていったことから理解できるであろう。

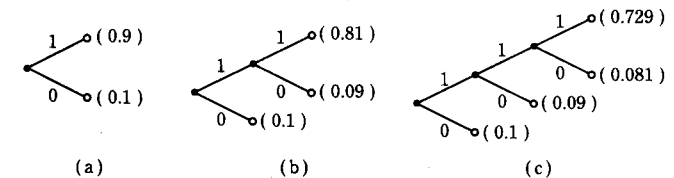


図 4.14 例 4.6 の平均的に長い非等長情報源記号系列

さて、この四つの非等長系列に対してハフマン符号を構成すると、図 4.15 のようになる。このとき、平均符号長は 1.442 となる。したがって、1 情報源記号当たりの平均符号長 L は、

$$L = \frac{1.442}{\bar{n}} = \frac{1.442}{2.71} = 0.532$$

となる。これは、4.7 節の 2 次拡大情報源において長さ 2 一定の 4 種の系列に対して行ったハフマンブロック符号の平均符号長 0.645 より効率が良くなっている。また、長さ 3 のハフマンブロック符号の平均符号長 0.532 (演習問題 4.1) とほぼ等しい。

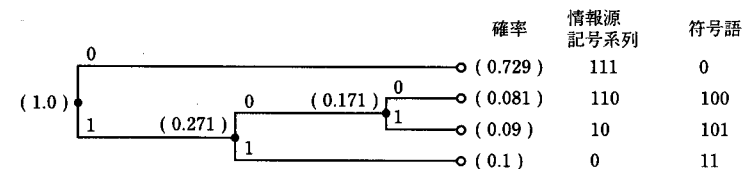
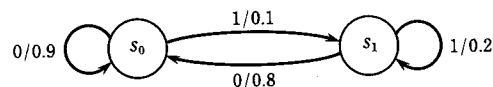


図 4.15 四つの非等長系列に対するハフマン符号化

演習問題 4.

- 4.1 $P(0)=0.1$, $P(1)=0.9$ なる記憶のない情報源 S を考える。長さ 3 のハフマンブロック符号を求めよ。
- 4.2 図 4.16 のような定常マルコフ情報源の出力記号を n 個ずつまとめてハフマン符号化を行う。
- (1) $n=1$ および 2 の場合について、情報源記号当たりの平均符号長 $L(n)$ を求めよ。
- (2) n を大きくしたとき、 $L(n)$ はどのような値に収束するか。

図 4.16 定常マルコフ情報源 S

- 4.3 確率 $P(\alpha)=0.7$, $P(\beta)=0.3$ で 2 元記号 $\{\alpha, \beta\}$ を発生する記憶のない 2 元情報源 S がある。非等長情報源記号系列に対するハフマン符号化の例を示し、その 1 情報源記号当たりの平均符号長を求めよ。

5

離散的通信路

本章ではまず、離散的通信路を通信路行列によりモデル化し、通信路を通して送られる情報量、すなわち相互情報量を定義する。これを用いて通信路にはそれを通して送ることのできる最大の情報量、すなわち通信路容量が存在することを示す。また、その容量以下の速度で情報を送る場合であれば誤り率をいくらでも小さくできる通信路符号化法が存在することを示す。

5.1 通信路モデル

情報の送り手と受け手の間には、通信路が存在する。そこにはしばしば雑音が存在し、送った情報が必ずしもそのままの形で到達しない。

通信路モデルでは、各時刻において一つの記号が入力され、それに応じて一つの記号が出力される。出力記号は入力記号から確率的に決定される。入力記号は入力アルファベットと呼ばれる記号の集合の元であり、出力記号は出力アルファベットと呼ばれる記号の集合の元である。入力記号および出力記号のことを、それぞれ送信記号および受信記号と呼ぶこともある。入力アルファベットと出力アルファベットは一致することもあるし、異なることもある。一致する場合、集合の元の数が r であればこの通信路を r 元通信路 (r -ary channel) と呼ぶ。

一般的には通信路の統計的な性質は、任意の長さ n の入力記号の系列 $x_1, x_2, \dots, x_n$ が与えられたとき、それに対応する出力記号の系列 $y_1, y_2, \dots, y_n$ の確率が与えられれば完全に定まる。しかしながら、あまりにも一般的な場合を取

り扱っても意味のある結果は出てこない。そこで、以下ではいちばん簡単な記憶のない定常通信路について述べる。

5.2 記憶のない定常通信路

各時刻における出力記号の出現確率はその時刻の入力記号のみに依存して定まるとき、その通信路を**記憶のない通信路**という。また、通信路の統計的な性質が、時刻によらずに一定であるとき、その通信路を**定常通信路**という。

記憶のない定常通信路は、図 5.1 のようにモデル化できるであろう。すなわち、ある入力記号が与えられたとき、各出力記号は図に示された条件付き確率で出現する。ここで、入力アルファベットは $A = \{a_1, a_2, \dots, a_r\}$ 、出力アルファベットは $B = \{b_1, b_2, \dots, b_s\}$ であり、 a_i が入力されたとき b_j が出力される確率は

$$p_{ij} = P(b_j | a_i) \quad (5.1)$$

である。このとき、 p_{ij} を要素とする行列

$$T = |p_{ij}| = \begin{bmatrix} p_{11} & p_{12} & \dots & p_{1s} \\ p_{21} & p_{22} & \dots & p_{2s} \\ \vdots & \vdots & & \vdots \\ p_{r1} & p_{r2} & \dots & p_{rs} \end{bmatrix} \quad (5.2)$$

を**通信路行列**(channel matrix)と呼ぶ。通信路行列の各行は各入力記号に対応し、各列は各出力記号に対応する。通信路行列では各入力記号に対してどれかの出力記号が必ず出力されるとすると、その出現確率の和は 1 でなければならないから、各入力記号 i ($i=1, 2, \dots, r$) について

$$p_{i1} + p_{i2} + \dots + p_{is} = 1 \quad (5.3)$$

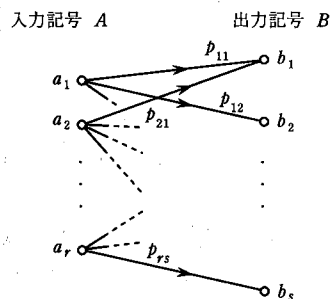


図 5.1 通信路線図

が成立する。通信路線図ではこれは、各入力記号から出ていく矢印の確率の和は 1 となっていることを表わしている。

[例 5.1] 2 元対称通信路

図 5.2 のような通信路線図をもつ通信路を考えよう。ここでは、入力アルファベット、出力アルファベットともに $\{0, 1\}$ の 2 元であり、その通信路行列は

$$T = \begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix}$$

である。このような通信路は**2 元対称通信路**(BSC; binary symmetric channel)と呼ばれる。

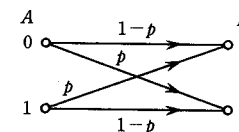


図 5.2 2 元対称通信路

[例 5.2] 2 元対称消失通信路

図 5.3 のような通信路線図をもつ通信路を考えよう。ここでは、入力アルファベットは上の例と同じく $\{0, 1\}$ であるが、出力アルファベットは $\{0, 1\}$ に加えて、 X という**消失**(erasure)記号が加わっている。消失は 0 と 1 とも判定しかねることを表わす。この通信路行列は

$$T = \begin{bmatrix} 1-p_x-p & p_x & p \\ p & p_x & 1-p_x-p \end{bmatrix} \quad (5.4)$$

である。このような通信路を**2 元対称消失通信路**という。

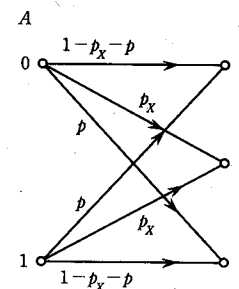


図 5.3 2 元対称消失通信路

上の例のような対称通信路では、どの行も出力記号を入れ換えれば同じ要素をもっていることがわかる。すなわち、各行は同じ要素の順序を入れ換えたものになっている。このような性質をもつ通信路を入力に関して一様な通信路と呼ぶ。さらに、図5.4では出力からみて一様になっている。このような通信路を出力に関して一様な通信路と呼ぶ。また、入力に関しても出力に関しても一様な通信路を2重に一様な通信路と呼ぶ。

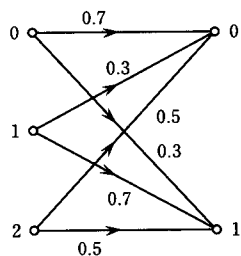


図 5.4 出力に関して一様な通信路

5.3 相互情報量

与えられた通信路を通して、どれだけの情報量を送ることができるかについて考えてみよう。まず、相互情報量を定義する。

ふたたび、図5.1の通信路を考える。入力記号は $A = \{a_1, a_2, \dots, a_r\}$ 、それらの生起確率は $P_A = \{P(a_1), P(a_2), \dots, P(a_r)\}$ とする。また、通信路行列は

$$P = [p_{ij}] \quad (i=1, 2, \dots, r; j=1, 2, \dots, s) \quad (5.5)$$

とする。このとき、入力記号のもつエントロピーは

$$H(A) = - \sum_{i=1}^r P(a_i) \log P(a_i) \quad (5.6)$$

である。これは、 A に関する事前のあいまいさを表わしている。一方、通信路においては雑音などにより乱されるから、出力記号 b_j を知っても、入力記号 a_i に関しては確率 $P(a_i|b_j)$ だけ散布していて定まらない。すなわち、通信路においては出力記号を完全に知っても、入力記号に関して平均的に

$$\begin{aligned} H(A|B) &= - \sum_{i=1}^r \sum_{j=1}^s P(b_j) P(a_i|b_j) \log P(a_i|b_j) \\ &= - \sum_{i=1}^r \sum_{j=1}^s P(a_i, b_j) \log P(a_i|b_j) \end{aligned} \quad (5.7)$$

なるあいまいさが残る。この $H(A|B)$ を条件付きエントロピーまたは事後エントロピーという。これに対して、 $H(A)$ 、 $H(B)$ などは事前エントロピーと呼ばれることがある。したがって、通信によって

$$I(A; B) = H(A) - H(A|B) \quad (5.8)$$

だけ入力記号に関するあいまいさが減少したといえることができる。 $I(A; B)$ を A と B の間の相互情報量と呼ぶ。またこれは、通信によって送られた情報量と解釈できる。なお、

$$P(a_i|b_j) = \frac{P(b_j|a_i) P(a_i)}{\sum_{i=1}^r P(b_j|a_i) P(a_i)} \quad (5.9)$$

は出力を知ったときの入力の条件付き確率(事後確率)である。式(5.8)を書き直すと、

$$\begin{aligned} I(A; B) &= H(A) - H(A|B) \\ &= - \sum_{i=1}^r P(a_i) \log P(a_i) + \sum_{i=1}^r \sum_{j=1}^s P(b_j) P(a_i|b_j) \log P(a_i|b_j) \\ &= \sum_{i=1}^r \sum_{j=1}^s \{ -P(a_i, b_j) \log P(a_i) + P(b_j) P(a_i|b_j) \log P(a_i|b_j) \} \\ &= \sum_{i=1}^r \sum_{j=1}^s P(a_i, b_j) \{ -\log P(a_i) + \log P(a_i|b_j) \} \\ &= \sum_{i=1}^r \sum_{j=1}^s P(a_i, b_j) \log \frac{P(a_i|b_j)}{P(a_i)} \\ &= \sum_{i=1}^r \sum_{j=1}^s P(a_i, b_j) \log \frac{P(a_i, b_j)}{P(a_i) P(b_j)}. \end{aligned} \quad (5.10)$$

同様に、 $H(B) - H(B|A)$ および $H(A) + H(B) - H(A, B)$ も式(5.10)に等しい(演習問題5.1)。事実、式(5.10)は a_i と b_j に関して、対称な形をしている。すなわち、

$$\begin{aligned} I(A; B) &= H(A) - H(A|B) \\ &= H(B) - H(B|A) \\ &= H(A) + H(B) - H(A, B) \geq 0. \end{aligned} \quad (5.11)$$

ここで、出力記号のエントロピーは

$$H(B) = - \sum_{j=1}^s P(b_j) \log P(b_j)$$

である。また、

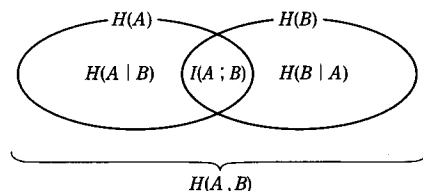
$$H(B|A) = - \sum_{i=1}^r \sum_{j=1}^s P(a_i) P(b_j|a_i) \log P(b_j|a_i),$$

$$H(A, B) = - \sum_{i=1}^r \sum_{j=1}^s P(a_i, b_j) \log P(a_i, b_j)$$

である。さらに、式(5.11)より

$$\begin{aligned} H(A, B) &= H(A) + H(B|A) \\ &= H(B) + H(A|B) \end{aligned} \quad (5.12)$$

が得られる。 $H(A, B)$ は $P(a_i, b_j)$ なる結合確率に関するエントロピーであり、**結合エントロピー**と呼ばれる。これは、たとえば A の事前のあいまいさと、 A を知ったときの B のあいまいさの和に等しい。これら図示すると図5.5のようになる。



$H(A), H(B)$: A および B のエントロピー
 $H(A|B)$: B を知った場合の A に関するエントロピー
 $H(B|A)$: A を知った場合の B に関するエントロピー
 $H(A, B)$: 結合事象 $A \times B$ のエントロピー
 $I(A; B)$: A と B の間の相互情報量

図 5.5 相互情報量とエントロピーの関係

5.4 通信路容量

相互情報量の定義ができたところで、通信路容量を定義する。通信路容量は

$$C = \max_{\{P(a_i)\}} I(A; B) \quad (5.13)$$

で定義される。すなわち、情報源記号の生起確率をいろいろ変えたときに、与

5.5 記憶のない一様通信路の通信路容量

えられた通信路を通して送られる最大の情報量を**通信路容量**と定義する。今の場合、単位は[ビット/記号]である。

通信路に記憶がある場合には、記憶のある情報源の場合と同じく、拡大の手法を用いればよい。すなわち、長さ n の入力および出力記号系列の集合 A_n , B_n , および A_n の生起確率 $\{P(a_i^{(n)})\}$ について考え、

$$C = \lim_{n \rightarrow \infty} \left[\max_{\{P(a_i^{(n)})\}} \frac{1}{n} I(A_n; B_n) \right] \quad (5.14)$$

で通信路容量が定義できる。

5.5 記憶のない一様通信路の通信路容量

まず、入力について一様な記憶のない通信路の通信路容量を求める。この場合、通信路行列の各行 $\{P(b_1|a_i), P(b_2|a_i), \dots, P(b_s|a_i)\}$ は、同一の要素の並びの順序を入れ換えたものになっている。このとき $f[\]$ を任意の関数として、

$$\sum_{j=1}^s f[P(b_j|a_i)] = \sum_{j=1}^s f[P(b_j|a_1)]$$

は i によらない値となる。したがって、

$$\begin{aligned} H(B|A) &= - \sum_{i=1}^r P(a_i) \sum_{j=1}^s P(b_j|a_i) \log P(b_j|a_i) \\ &= - \sum_{j=1}^s P(b_j|a_1) \log P(b_j|a_1) \\ &= - \sum_{j=1}^s p_{1j} \log p_{1j}. \end{aligned} \quad (5.15)$$

このとき、

$$\begin{aligned} C &= \max_{\{P(a_i)\}} I(A; B) \\ &= \max_{\{P(a_i)\}} [H(B) - H(B|A)] \\ &= \left[\max_{\{P(a_i)\}} H(B) \right] + \sum_{j=1}^s p_{1j} \log p_{1j} \end{aligned} \quad (5.16)$$

となる。

ここでさらに、出力についても一様な通信路、すなわち、2重に一様な通信路について考えよう。このとき入力の確率分布を、 $P(a_1) = P(a_2) = \dots =$

$P(a_r)=1/r$ とすると, 出力の確率分布は $P(b_1)=P(b_2)=\dots=P(b_s)=1/s$ となり, $H(B)$ はその最大値 $\log s$ をとる。したがって, 2 重に一樣な通信路においては

$$C = \log s + \sum_{j=1}^s p_{ij} \log p_{ij} \quad (5.17)$$

である。

【例 5.3】 2 元対称通信路の通信路容量

図 5.2 の 2 元対称通信路の通信路容量を求めてみよう。通信路行列は

$$T = \begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix} \quad (5.18)$$

である。2 重に一樣であるから, 式(5.17)より

$$C = 1 + p \log p + (1-p) \log (1-p) \\ = 1 - H(p) \quad (5.19)$$

となる。これを図示すると, 図 5.6 のようになる。

$p=0$ のときには誤りはまったく発生しないから, $C=1$ [ビット/記号] となる。 $p=0.5$ となると, 出力記号は入力記号と独立となり, $C=0$ となって入力側の情報はまったく伝わらなくなる。さらに, $p=1$ となるとこれは必ず誤る, すなわち受信記号を反対にすれば正しい送信記号がわかるから, $p=0$ のときと同じく $C=1$ となる。

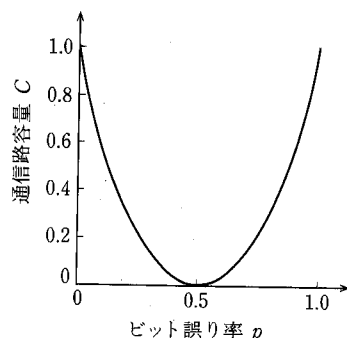


図 5.6 2 元対称通信路の通信路容量

【例 5.4】 2 元対称消失通信路の通信路容量

2 元対称消失通信路の通信路行列を

$$T = \begin{bmatrix} 0 & X & 1 \\ 1-p_x-p & p_x & p \\ p & p_x & 1-p_x-p \end{bmatrix} \quad (5.20)$$

とする。これは, 入力について一樣であるから, 式(5.16)より

$$C = \left[\max_{\{P(a_i)\}} H(B) \right] + (1-p_x-p) \log (1-p_x-p) + p_x \log p_x + p \log p \quad (5.21)$$

となる。出力については一樣ではないが, $H(B)$ は $P(a_1)=P(a_2)=1/2$ のとき最大になることが容易に確かめられる。このとき,

$$C = (1-p_x) \left[1 - H\left(\frac{p}{1-p_x}\right) \right] \quad (5.22)$$

となる。すなわち, 消失を除外した誤り率 $p/(1-p_x)$ をもつ通信路の通信路容量の $1-p_x$ 倍となる。

5.6 情報伝送速度

通信路符号化では, 冗長度を増やすことにより, 雑音のため符号語が変化をうけても, 受信語からできるだけ元の符号語が復元できるようにすることを考えた。たとえば, 符号語の長さを 3, 符号を $\{000, 111\}$ とした場合を考える。長さ 3 の可能な符号は 8 個あるが, いまの場合, そのうちの 2 個を符号語として採用したことになる。

一般に, 長さ n の M 個の 2 元符号をそれぞれ等確率で使用した場合, 情報が送り出される速度 R は

$$R = \frac{\log_2 M}{n} \quad [\text{ビット/記号}] \quad (5.23)$$

となる。これを, 符号の情報伝送速度, または情報速度という。これは情報を送り出す側から考えた速度であり, 途中での誤りを考慮した実効的に情報が伝達される速度(相互情報量)ではないので, 情報送出速度といったほうが実態をよく表わしているが, この用語は一般的には用いられない。

式(5.23)より, R と M には,

$$M = 2^{nR} \quad (5.24)$$

なる関係がある。一方, 長さ n の 2 元系列は 2^n 種ある。もし, これらをすべて符号語として使用した場合, 情報伝送速度は最大値

$$R_{\max} = \frac{\log_2 2^n}{n} = 1 \quad [\text{ビット/記号}] \quad (5.25)$$

となる。しかしながら, このようにすると通信路で雑音が加わると直ちに誤りとなり, 誤りの訂正も検出もできない。

符号の効率を

$$\eta = \frac{R}{R_{\max}} (=R) \quad (5.26)$$

で定義する。これはまた符号化率とも呼ばれる。一般に,

$$0 < \eta < 1$$

である。また、

$$\rho = 1 - \eta \quad (5.27)$$

を、符号の冗長度という。冗長度を付加することにより、誤りの検出や訂正ができるようになり、通信の信頼性が向上するのである。

5.7 通信路符号化定理

前節では、効率を下げて冗長度を付加することにより、通信の信頼性を高めることができることを述べた。では、いくら冗長度をつけ加えればよいのだろうか。冗長度が多すぎると効率が悪くなり、情報伝送速度そのものが低下する。また、冗長度が少なすぎると、誤りが多くなり、あいまいさが増えて、実効的に情報が送られる速度である相互情報量は大きくならないであろう。

本節では、通信路容量以下の情報伝送速度であれば、いくらでも誤り率を小さく、すなわち事実上 0 にすることができることを示そう。

5.7.1 典型的系列

情報源記号 $a_1, a_2, \dots, a_M$ をそれぞれ確率 $p_1, p_2, \dots, p_M$ で発生する記憶のない情報源 S を考え、その出力系列から十分長い長さ n の系列を切り出して、その統計的性質について調べよう。

長さ n の系列中に含まれる a_i の個数を n_i とする。 n は十分長いから、ほとんどの系列では系列中に記号 a_i はほぼ $n_i = np_i$ 個含まれるとしてよいであろう。ここで、長さ n の典型的系列を λ とし、その生起確率を求めてみよう。 λ 中には a_i がほぼ n_i 個含まれているから、

$$\begin{aligned} P(\lambda) &= \prod_{i=1}^M p_i^{n_i} \\ &= \prod_{i=1}^M p_i^{np_i} \\ &= \prod_{i=1}^M 2^{np_i \log p_i} \\ &= 2^{\sum_{i=1}^M np_i \log p_i} \\ &= 2^{-nH(S)} \end{aligned} \quad (5.28)$$

となる。すなわち、典型的系列はどれも一定の確率 $2^{-nH(S)}$ で出現するのであ

る。したがって、典型的系列の数は

$$\frac{1}{P(\lambda)} = 2^{nH(S)} \quad (5.29)$$

個である。典型的系列でない系列が出現する確率はほとんど 0 である。

5.7.2 ランダム符号化

いま、図 5.7 のような記憶のない通信路と、その通信路容量 C を達成するような記憶のない適合情報源 S_0 を考える。 S_0 から発生する長さ n の典型的系列は $2^{nH(X)}$ 個であるが、そのうち、 $M = 2^{nR}$ 個の系列をランダムに選び、それを符号 $C_R = \{w_1, w_2, \dots, w_M\}$ とする。なお、 $2^{nH(X)} \geq 2^{nR}$ であるから、 $H(X) \geq R$ である。 C_R はランダムに選んでつくったので、同じ w_i が 2 度以上選ばれている可能性もある。以下の議論は、ランダムに選ばれたある特定の符号 C_R に対してではなく、そのような符号全体についての平均の話である。

S_0 は通信路容量 C を達成する適合情報源であるから、

$$C = H(X) - H(X|Y) \quad (5.30)$$

である。ただし、 S_0 は符号語を選ぶために用いられる情報源であり、 S_0 の情報を通信路に通して送るのが目的ではない。図 5.8 のように、 S_0 からある長

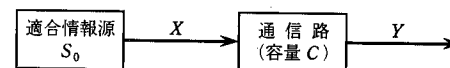


図 5.7 通信路容量 C を達成する情報源と通信路

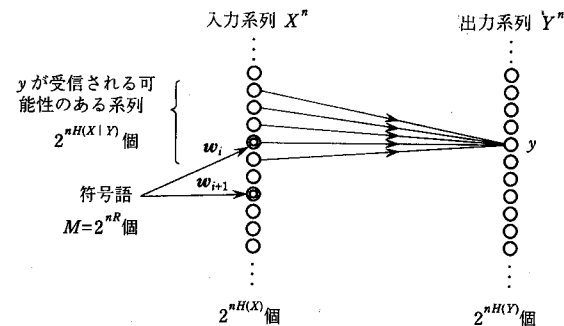


図 5.8 ランダム符号化における入力系列と出力系列

さ n の系列が発生した場合、それに近い C_R の符号語 w_i が送られる。出力側では、 y を受信した場合、入力系列に対して $H(X|Y)$ だけのあいまいさがあるので、送られた可能性のある系列は $2^{nH(X|Y)}$ 個あることになる。

もし、このとき、送られた可能性のある $2^{nH(X|Y)}$ 個の入力系列中に C_R の符号語が w_i 以外に一つもなければ、送られた符号語は w_i に確定し、誤りは生じない。

ところで、任意の典型的入力系列が符号語に選ばれる確率は $2^{nR}/2^{nH(X)}$ であるから、 $2^{nH(X|Y)}$ 個の送信系列中に w_i 以外に一つも符号語に選ばれない確率 P_e は、 n が十分大きいとき

$$\begin{aligned} P_e &= \left(1 - \frac{2^{nR}}{2^{nH(X)}}\right)^{2^{nH(X|Y)}} \\ &\cong 1 - \frac{2^{nH(X|Y)} 2^{nR}}{2^{nH(X)}} \\ &= 1 - 2^{-n(C-R)} \end{aligned} \quad (5.31)$$

となる。また、復号誤り率 P_e は

$$P_e = 2^{-n(C-R)} \quad (5.32)$$

これより、 $C > R$ であれば、 $n \rightarrow \infty$ のとき $P_e \rightarrow 0$ なることがわかる。この結果は、ランダム符号化法の平均についてのものである。したがって、なかにはこれより誤り率が小さい符号も存在する。このことから、つぎの定理を得る。

定理 5.1 (通信路符号化定理)

与えられた通信路の通信路容量を C とする。このとき、 $R < C$ であれば、任意の正数 ϵ に対して、誤り率 P_e が

$$P_e < \epsilon \quad (5.33)$$

なる情報伝送速度 R の符号が存在する。しかし、 $R > C$ であればそのような符号は存在しない。

証明 定理の前半はすでに証明した。後半は、通信路容量の定義より明らかである。

この定理は通信路容量に明確な意味を与える。すなわち、通信路容量以下の情報伝送速度であれば、適切な符号化により任意に小さい誤り率で情報を伝送できる。また、通信路容量は任意に小さい誤り率で伝送できる情報伝送速度の上限であるということもできる。

上の定理では、そのような符号が存在することの証明は与えているが、証明に用いた符号はランダム符号であった。したがって、そのような符号を試行によっていくつかつくり、その中から効率のよいものを選べば理論的には任意に小さい誤り率で伝送可能となるはずである。しかしながら、これでは確率的な要素が大きく、符号の長さや数が一定せず、符号化回路や機構の複雑さの点から実際的ではない。たとえば、符号長が長いということは、入力系列を一時的に蓄えておく記憶回路が大きくなり、符号化のための遅延時間も長くなって、実用上支障がでてくる。

演習問題 5.

- 5.1 式(5.11)を示せ。すなわち、 $I(A; B) = H(A) - H(A|B)$ は、(1) $H(B) - H(B|A)$ 、(2) $H(A) + H(B) - H(A, B)$ に等しいことを示せ。(3) また、これらは非負であることを示せ。
- 5.2 m 個の白玉と n 個の黒玉が入った壺がある ($m, n \geq 1$)。その壺から続けて2個の玉を取り出すとき、最初の玉を取り出す試行を α 、2個目の玉を取り出す試行を β とする。試行 α と β のエントロピーおよびこれらの条件付きエントロピーを求めよ。また、 $m=3$ 、 $n=1$ のとき、これらの値を付表1を用いて求めよ。
- 5.3 ある地方では一年を通して、10日のうち2日雨(r)が降り、残り8日は晴(f)である。一方、この地方の天気予報(F)の適中率は晴雨にかかわらず70%である。毎日晴を予想する人(H)の適中率は80%である。 F と H の W (明日の天気)に関する相互情報量はそれぞれいくらか。
- 5.4 相続く誤りをバースト誤りという。バースト誤りを発生する2元通信路(図5.9)がある。この通信路のビット誤り率は 10^{-1} 、バースト誤りの長さの平均値は5である。この通信路の誤り源が図5.10のようにモデル化できるとき、通信路容量を求めよ。

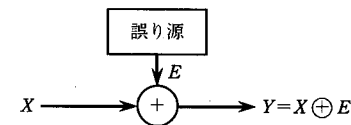


図 5.9 加法的 2 元通信路

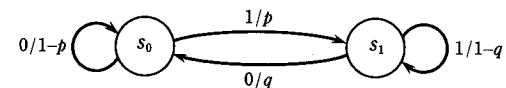


図 5.10 誤り源モデル

6

通信路符号化法

前章では、通信路符号化の限界について学んだが、実際の符号化法についてはふれなかった。本章では、実際に用いられている通信路符号化法の理論の基礎について述べる。これらは、**符号理論**と呼ばれているものである。シャノンの理論が確率・統計を基礎としているのに対して、符号理論は代数学に基礎をおいている。

6.1 組織符号

一般に、ある長さの情報源記号系列にそれから一定の方法で算出される**検査記号**系列を付加することにより得られる符号を**組織符号**という。情報記号の個数が k 個で検査記号の個数が $n-k$ 個である長さ n の組織符号を (n, k) 符号という(図 6.1)。

(n, k) 符号の効率

$$\eta = \frac{k}{n} \quad (6.1)$$

である。組織符号のなかでもとくに、情報記号の線形式で検査記号が与えられる**線形符号**が理論的にも、実用的にも重要である。

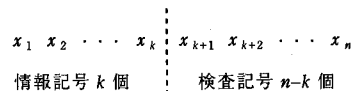


図 6.1 (n, k) 符号

6.2 単一パリティ検査符号

線形符号の一番簡単なものは、単一パリティ検査符号である。いま、 $\{0, 1\}$ の 2 元よりなる情報ビット列を $x_1, x_2, \dots, x_k$ とし、検査ビット c を

$$c = x_1 + x_2 + \dots + x_k \quad (6.2)$$

で求める。ただし、本章では $+$ は排他的論理和、すなわち、 $\text{mod } 2$ の演算を表わすことにする。この演算では、 $0+0=0$, $1+0=0+1=1$, $1+1=0$ であり、また、 $-0=0$, $-1=1$ である。式(6.2)は、 $x_1, x_2, \dots, x_k$ に 1 が偶数個含まれていれば $c=0$ 、またそれが奇数個であれば $c=1$ となる。このようにして得られた c を情報ビット列につけ加えた

$$w = x_1 x_2 \dots x_k c \quad (6.3)$$

を**単一パリティ検査符号**という。単一パリティ検査符号においては、 $n=k+1$ である。なお、式(6.3)を

$$w = (x_1, x_2, \dots, x_k, c) \quad (6.4)$$

または

$$w = (x_1 \ x_2 \ \dots \ x_k \ c) \quad (6.4)'$$

と表現することもある。実用上は、このように 1 の個数を偶数にする偶数パリティ検査以外に、それを奇数にする奇数パリティ検査があるが、後者は検査ビットを反転して考えれば前者と同じである。

いま、単一パリティ検査符号の符号語を

$$w = (w_1, w_2, \dots, w_n) \quad (6.5)$$

とするとき、

$$w_n = w_1 + w_2 + \dots + w_{n-1} \quad (6.6)$$

であるから

$$w_1 + w_2 + \dots + w_n = 0 \quad (6.7)$$

となる。逆に、式(6.7)を満たす w は単一パリティ検査符号の符号語である。したがって、式(6.7)は単一パリティ検査符号の符号語であるための必要十分条件になっている。一般に、符号中のビットの線形和が 0 となる(6.7)のような式(もしくは式の組)は、**パリティ検査方程式**と呼ばれる。

単一パリティ検査符号では 1 個の誤りが検出可能である。受信時には 1 の個数が偶数であれば誤りなし、またそれが奇数であれば 1 個(または奇数個)の誤りを含むとみなす。ただし、1 の個数が偶数であっても、偶数個の誤りを含んでいる可能性はある。

6.3 水平垂直パリティ検査符号

図 6.2 のように、 $k \times m$ 個の情報ビットがあるとき、各行各列にそれぞれ単一パリティ検査ビットを付加し、さらにそれらの検査ビットに 1 個の検査ビットを付加する。すなわち、

$$c_i = x_{1i} + x_{2i} + \cdots + x_{ki} \quad (i=1, 2, \dots, m), \quad (6.8)$$

$$c'_j = x_{j1} + x_{j2} + \cdots + x_{jk} \quad (j=1, 2, \dots, k), \quad (6.9)$$

$$c = c_1 + c_2 + \cdots + c_m = c'_1 + c'_2 + \cdots + c'_k. \quad (6.10)$$

このとき、 $k+m+1$ 個の検査ビットが付加され、 $(k+1) \times (m+1)$ ビットの符号語となる。これらを適当な順序で並べて 1 次元系列にすれば $((k+1)(m+1), km)$ 符号となる。図 6.3 からわかるように、水平垂直パリティ検査符号では 1 ビットの誤りは必ず訂正できる（もちろん検出もできる）。しかしながら、2 ビットの誤りは検出可能であるが、訂正はできない。

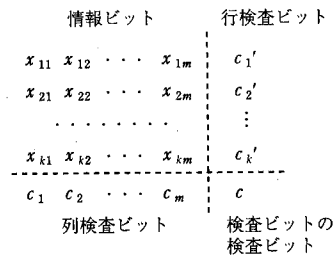


図 6.2 水平垂直パリティ検査符号

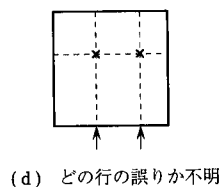
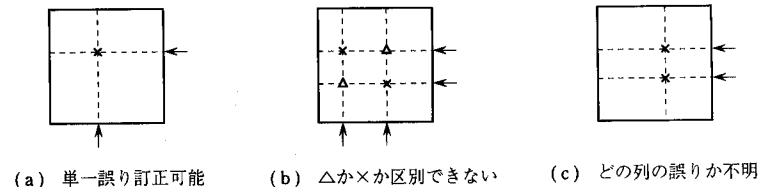


図 6.3 水平垂直パリティ検査符号による、
(a) 単一誤りの訂正と、(b)~(d) 2 個の誤りの検出。矢印が誤り検出の位置。

6.4 線形符号

単一パリティ検査符号では単一誤りを検出できた。また、水平垂直パリティ検査符号では、単一誤りを訂正できた。このような考え方を一般化した線形符号について述べる。線形符号はつぎの式で与えられる符号語 w の集合である：

$$w = xG \quad (6.11)$$

ここで、 x は長さ k の任意の 2 元情報ビット列を表わす行ベクトルであり、 G は生成行列と呼ばれる大きさ $k \times n$ の 2 元行列である。

[例 6.1]

$$x = (x_1, x_2, x_3) \quad (6.12)$$

$$G = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 1 & 1 \end{bmatrix} \quad (6.13)$$

とする。このとき、生成される符号をすべて書き出すと、表 6.1 のようになる。これを符号 C とする。すなわち、

$$C = \{w_1, w_2, \dots, w_8\} \quad (6.14)$$

である。

表 6.1 例 6.1 で生成される符号 C

	情報ビット	検査ビット
w_1	0 0 0	0 0 0 0
w_2	0 0 1	1 0 1 1
w_3	0 1 0	0 1 1 1
w_4	0 1 1	1 1 0 0
w_5	1 0 0	1 1 0 1
w_6	1 0 1	0 1 1 0
w_7	1 1 0	1 0 1 0
w_8	1 1 1	0 0 0 1

6.5 ハミング距離

$\{0, 1\}$ を元とする二つの n 次元ベクトル $u = (u_1, u_2, \dots, u_n)$ と $v = (v_1, v_2, \dots, v_n)$ の間のハミング (Hamming) 距離 d_H を次のように定義する：

$$d_H(u, v) = \sum_{k=1}^n d(u_k, v_k). \quad (6.15)$$

ここに,

$$d(u, v) = \begin{cases} 0 & (u=v \text{ のとき}) \\ 1 & (u \neq v \text{ のとき}) \end{cases} \quad (6.16)$$

である。すなわち, 異なる要素の数で定義する。この Σ は通常のとおりである。このハミング距離 d_H は, つぎの距離の三公理を満たす。

- (i) $d_H(u, v) \geq 0$ 等号が成立するのは $u=v$ のときに限る。
- (ii) $d_H(u, v) = d_H(v, u)$
- (iii) $d_H(u, v) + d_H(v, w) \geq d_H(u, w)$ (三角不等式)

つぎに, n 次元ベクトル w のハミング重み $w_H(u)$ を, 0 でない要素の数とする。すなわち,

$$w_H(u) = d_H(u, 0) \quad (6.17)$$

ここに, 0 は n 次元ベクトル

$$0 = (0, 0, \dots, 0)$$

である。ハミング重みを単に重みともいう。また逆に,

$$d_H(u, v) = w_H(u-v) = w_H(u+v) \quad (6.18)$$

なる関係がある。

6.6 線形符号の加法性

情報ビット列 x_1 および x_2 から生成される線形符号の符号語 w_1 および w_2 を考えよう。式(6.11)より,

$$\begin{aligned} w_1 &= x_1 G \\ w_2 &= x_2 G \end{aligned} \quad (6.19)$$

とし, これらの和を w とすると

$$w = w_1 + w_2 = (x_1 + x_2)G \quad (6.20)$$

である。ここで, $(x_1 + x_2)$ も情報ビット列であるから, これに対する符号語 w も符号中に存在する。すなわち, 一般に線形符号はこのような意味で, 加法に関して閉じている。和と差は同じであるから, 差に関して閉じている。

6.7 最小距離・最小重みと誤り訂正能力

上に述べたように, mod 2 の加法では, $+$ と $-$ は同一である。したがって, 二つの符号語 w_1 と w_2 の差

$$\begin{aligned} w_1 - w_2 &= (x_1 + x_2)G \\ &= w \end{aligned} \quad (6.21)$$

はやはり符号語 w である。すなわち, w_1 と w_2 のハミング距離は別の符号語 w の重みに等しい。通信路において, 誤りが各ビットにおいて等しく独立に起こるとすれば, ハミング距離に近いほどそれらの間で誤りが起こりやすい。したがって, 異なる符号語間のハミング距離はできるだけ大きいほうがよい。それは, 符号語の重みをみればわかる。例 6.1 では, 最小重みは 4 である。最小ハミング距離は同じく $4 = d_H(w_1, w_2) = d_H(w_1, w_3) = \dots$ である。ただしこの例の場合は, 符号語の重みが w_1 を除いてすべて 4 であるから, どの符号語間の距離も最小重みに等しく 4 である。

一般に符号 C が与えられたとき, 符号の最小距離 $d_{\min}$ とは, C に含まれるあらゆる符号語間のハミング距離の最小値をいう。すなわち,

$$d_{\min} = \min_{\substack{w_i \neq w_j \\ w_i, w_j \in C}} \{d_H(w_i, w_j)\} \quad (6.22)$$

であり, 線形符号の場合, これは符号の最小重みに等しい。

長さ n の 2 元系列の総数は 2^n 個である。その中から $w_1, w_2, \dots, w_m$ なる m 個の系列が符号語として選出される。通信路においては, 誤り e が加わり, y が受信される。よく使われる限界距離復号法においては, 図 6.4 に示すように, 各符号語 w_i のまわりに半径 t_0 の球を設定し, 受信語がその中に落ちればそれは w_i として復号する。すなわち, 球の内部は誤り訂正領域である。それ以外の灰色の領域に落ちれば誤りの検出はするが, 訂正はしない。このような領域は誤り検出領域である。

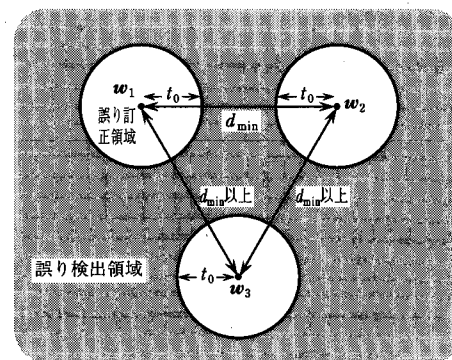


図 6.4 限界距離復号法。

もし w_i を送信したとき、誤りが多くて別の符号 w_j の訂正領域に入ってしまったとすると、それは w_j として復号され、復号誤りが生じる。誤り訂正領域の半径 t_0 の球がお互いに重ならないとすると、 t_0 個までの誤りは訂正可能である。なお、

$$t_0 \leq \left\lfloor \frac{d_{\min} - 1}{2} \right\rfloor \quad (6.23)$$

である。ここに、 $\lfloor x \rfloor$ は x を越えない最大の整数を表わす。右辺は t_0 の最大値であり、この符号の誤り訂正能力という。さらに、 $(d_{\min} - t_0 - 1)$ 個までの誤りは検出可能である。これ以上の誤りに対しては、検出できる場合もあるが、誤って復号されることもあり、必ず検出できるとは限らない。これらの例を表 6.2 に示す。

表 6.2 限界距離復号法における誤り検出・訂正能力

$d_{\min}$	t_0	訂正可能 誤り個数	検出可能 誤り個数
3	0	0	2
3	1	1	1
4	0	0	3
4	1	1	2
5	0	0	4
5	1	1	3
5	2	2	2

6.8 パリティ検査方程式

例 6.1 の (7, 3) 符号のパリティ検査行列は

$$H = \begin{bmatrix} 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (6.24)$$

で与えられる。このときパリティ検査方程式は

$$wH^T = 0 \quad (6.25)$$

である。一般に、 (n, k) 符号の $k \times n$ 生成行列を $G = [I | G_0]$ としたとき、パリティ検査行列 H は $(n-k) \times n$ 行列であり、

$$H = [G_0^T | I] \quad (6.26)$$

で与えられる。このとき、

$$\begin{aligned} wH^T &= xGH^T = x[I | G_0][G_0^T | I]^T \\ &= x[G_0 + G_0] = x[2G_0] = x0 = 0 \end{aligned} \quad (6.27)$$

である。符号 w を送ったとき、通信路で誤り e が加わると、受信語 y は

$$y = w + e \quad (6.28)$$

となる。このとき、パリティ検査方程式は

$$s = yH^T = (w + e)H^T = wH^T + eH^T = eH^T \quad (6.29)$$

となる。これを、誤りの症候という意味からシンドロームパターン、または単にシンドロームと呼ぶ。

上の例において、単一誤りの場合にシンドロームがどのようなになるかを表 6.3 に示す。

表 6.3 において、パリティ検査行列の各列が単一誤りに対するシンドロームとして現れている。それらは互いに異なるから、シンドロームをみれば誤り位置が特定できる。また、2 個の誤りに対してはシンドロームはそれぞれの誤りに対するシンドロームの和になる。誤りがない場合を除きシンドロームは重みが 1 または 3 であり、かつお互いにハミング距離が 2 以上離れている。任意の二つのシンドロームの和は別のシンドロームになっていない。このことから、この符号は 2 個の誤りの検出が可能であることがわかる。

表 6.3 単一誤りに対するシンドローム

誤りパターン	シンドローム
0 0 0 0 0 0 0	0 0 0 0
1 0 0 0 0 0 0	1 1 0 1
0 1 0 0 0 0 0	0 1 1 1
0 0 1 0 0 0 0	1 0 1 1
0 0 0 1 0 0 0	1 0 0 0
0 0 0 0 1 0 0	0 1 0 0
0 0 0 0 0 1 0	0 0 1 0
0 0 0 0 0 0 1	0 0 0 1

6.9 単一誤り訂正ハミング符号

線形符号において単一誤りを訂正できるためには、そのシンドローム、すなわちパリティ検査行列の各列がすべて異なればよい。ただし、0は誤りなしの場合と区別できないので除く。これにより、誤りの位置を特定できる。たとえば、検査ビット数 m の場合、異なるシンドロームの数は $2^m - 1$ である。したがって、このとき符号長 n は

$$n \leq 2^m - 1 \quad (6.30)$$

でなければならない。この関係から、符号語数 $M = 2^k = 2^{n-m}$ に関しては

$$M \leq \frac{2^n}{n+1} \quad (6.31)$$

なる限界が存在する。

符号長を式(6.30)で与えられる最大値 $n = 2^m - 1$ とした場合、情報ビット数は $k = 2^m - 1 - m$ となる。このとき、式(6.31)は等号が成立する。このような符号を単一誤り訂正ハミング符号、または単にハミング符号という。

ハミング符号においては、パリティ検査行列の各列はすべて異なり、かつ

符号長: $n = 2^m - 1$

情報ビット数: $k = 2^m - 1 - m$

検査ビット数: $m = n - k$

なるように選ばれる。ハミング符号は誤り訂正符号の研究の端著となったものであり、これを契機として、符号理論の研究が盛んに行われるようになった。

【例 6.2】 検査ビット数 $m=3$ の場合、 $n=7$ 、 $k=4$ となり、(7,4)ハミング符号と呼ばれる符号が得られる。この検査行列はたとえば

$$H = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}$$

であり、このとき表 6.4 のような符号が得られる。また、 $m=4$ のとき $n=15$ 、 $k=11$ であり、検査行列はたとえば

$$H = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}$$

となる。

表 6.4 (7,4)ハミング符号

情報ビット				検査ビット		
x_1	x_2	x_3	x_4	x_5	x_6	x_7
0	0	0	0	0	0	0
0	0	0	1	0	1	1
0	0	1	0	1	1	0
0	0	1	1	1	0	1
0	1	0	0	1	1	1
0	1	0	1	1	0	0
0	1	1	0	0	0	1
0	1	1	1	0	1	0
1	0	0	0	1	0	1
1	0	0	1	1	1	0
1	0	1	0	0	1	1
1	0	1	1	0	0	0
1	1	0	0	0	1	0
1	1	0	1	0	0	1
1	1	1	0	1	0	0
1	1	1	1	1	1	1

6.10 バースト誤り

通信路における誤りは定常雑音などによるランダムな誤りと、回線の瞬断などによるバースト誤り(密集誤り)がある。これまでは、通信路には記憶がなく、したがって誤りはランダムに発生するものとして取り扱ってきた。しかしながら、実際の通信路においては記憶のあるものが多く、とくに連続して誤るバースト誤りに対する考察が重要である。その誤りパターンは図 6.5 に示されるように、ある長さ l の期間に集中して誤りが起こるのである。誤り期間中のビットデータは誤ることもあるが正しく受信されることもある。なお、必ず訂正(検出)可能なバースト誤りの最大の長さを、その符号 C のバースト誤り訂正(検出)能力という。

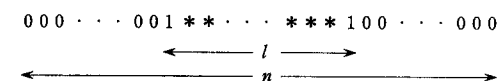


図 6.5 長さ l のバースト誤りの誤りパターン(*は 0 または 1 任意である)

符号 C のバースト誤り検出能力が l_0 であるための必要十分条件は、 C の任意の符号語 w_1 に、長さ l_0 以下の任意のバースト誤りパターン e を加えた $w_1 + e$ が、 C 中の別の符号語 w_2 にならないことである。線形符号の場合、この条件は長さ l_0 以下の任意のバースト誤りパターン $e (\neq 0)$ が、符号語にならないことであると言い換えることができる。

また、符号 C のバースト誤り訂正能力が l_0 であるための必要十分条件は、 C の任意の符号語 w_1 に、長さ l_0 以下の任意のバースト誤りパターン e_1 を加えた $w_1 + e_1$ が、 C 中の別の符号語 w_2 に長さ l_0 以下の任意のバースト誤りパターン e_2 を加えた $w_2 + e_2$ に一致しないことである。線形符号の場合、この条件は長さ l_0 以下の任意の二つのバースト誤りパターンの和 $e_1 + e_2 (\neq 0)$ が、符号語にならないことであると言い換えることができる。

単一のバースト誤りに対しては以下に述べる巡回符号を含め、比較的強い符号が存在するが、複数のバースト誤りが重なれば、多くの場合、誤り検出・訂正は困難となる。

6.11 巡回符号

国際的な CCITT 勧告に従い、実際によく使用されている誤り検査方式は水平・垂直パリティ検査方式と、CRC 方式である。CRC 方式は、巡回符号を用いる。巡回符号はその特殊な場合としてパリティ検査方式を含む。巡回符号は冗長度が小さいわりに、信頼性が高い符号である。すなわち、巡回符号はその構造が数的に整然としており、見通しよく高い誤り訂正能力をもたせることができる。また符号の操作が回路で簡単に実現できることから、実用性の高い符号になっている。

6.11.1 2元系列の多項式表現

0, 1 を要素とする n 次元 2 元ベクトルもしくは 2 元系列

$$(a_{n-1}, a_{n-2}, \dots, a_0)$$

を多項式

$$a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + \dots + a_1x + a_0 \quad (6.32)$$

で表わすと便利ことが多い。たとえば (1101001) は、 $x^6 + x^5 + x^3 + 1$ と表現される。以下、多項式の係数は 0 または 1 とする。多項式演算において、加算は mod 2、すなわち排他的論理和で計算される。すなわち $x + x = 0$ 、 $x +$

6.11 巡回符号

$0 = x$ 、 $2x = 0$ 、 $-x = x$ である。また、 x を掛けることは、系列を左に一つシフトすることに対応する。右端には 0 が入る。逆に x で割ることは系列を右に一つシフトすることに対応する。右端にあった系列要素は余りとしてはき出される。なお、このような演算はシフトレジスタと呼ばれる回路で容易に実現できる。

多項式表現を用いると、長さ n の符号 C は $n-1$ 次以下の多項式の集合として表わされる。このとき、各符号語に対応する多項式は符号多項式と呼ばれる。

6.11.2 生成多項式と巡回符号

巡回符号(cyclic code)は、定数項が 1 の m 次の生成多項式

$$G(x) = x^m + g_{m-1}x^{m-1} + \dots + g_1x + 1 \quad (6.33)$$

の $n-1$ 次以下の倍多項式、すなわち

$$W(x) = Q(x)G(x) \quad (6.34)$$

を符号多項式とする符号として定義される。ここに、 $Q(x)$ は、 $n-m-1$ 次以下の 0 または 1 を係数とする任意の多項式である。

式(6.34)で定義される巡回符号は線形符号である。なぜなら、二つの符号多項式 $W_1(x) = Q_1(x)G(x)$ と $W_2(x) = Q_2(x)G(x)$ の和

$$\begin{aligned} W(x) &= W_1(x) + W_2(x) \\ &= [Q_1(x) + Q_2(x)]G(x) \end{aligned} \quad (6.35)$$

は、やはり式(6.34)を満たす符号多項式になっているからである。

[例 6.3] 生成多項式を $G(x) = x^3 + x + 1$ とする。この多項式の次数、すなわち検査ビット数は $m=3$ である。いま、 $n=6$ としよう。このとき、生成される巡回符号は表 6.5 のようになる。

表 6.5 $G(x) = x^3 + x + 1$ により生成される巡回符号 ($n=6$)

$Q(x)$	$W(x)$	w
0	0	0 0 0 0 0 0
1	$x^3 + x + 1$	0 0 1 0 1 1
x	$x^4 + x^2 + x$	0 1 0 1 1 0
$x + 1$	$x^4 + x^3 + x^2 + 1$	0 1 1 1 0 1
x^2	$x^5 + x^3 + x^2$	1 0 1 1 0 0
$x^2 + 1$	$x^5 + x^2 + x + 1$	1 0 0 1 1 1
$x^2 + x$	$x^5 + x^4 + x^3 + x$	1 1 1 0 1 0
$x^2 + x + 1$	$x^5 + x^4 + 1$	1 1 0 0 0 1

つぎに、情報ビット $a_{n-m-1}, a_{n-m-2}, \dots, a_1, a_0$ が与えられたとき、符号多項式を求める方法について述べよう。この情報ビット多項式表現を $A(x)$ とする。このとき、 $x^m A(x)$ を $G(x)$ で割った余り(剰余多項式)を $R(x)$ とする。すなわち、

$$x^m A(x) = Q(x) G(x) + R(x) \quad (6.36)$$

であり、 $R(x)$ は $m-1$ 次以下の多項式である。また、 $Q(x)$ は商多項式である。これらを用いて符号多項式を

$$W(x) = x^m A(x) + R(x) \quad (6.37)$$

とする。mod 2 の演算では $+$ と $-$ は同じであることに注意すると、式(6.36)および式(6.37)より、 $W(x) = Q(x) G(x)$ となり、式(6.37)により得られる符号多項式は $G(x)$ の倍多項式となっている。したがって、このようにして得られる符号は巡回符号になっていることがわかる。またそれは、長さ n 、検査ビット数 m である $(n, n-m)$ 符号である。

【例 6.4】 $G(x) = x^3 + x + 1$ とする。 $m=3$ である。また、情報ビットを 11010 とする。このとき、 $n=5+m=8$ となる。また、 $A(x) = x^4 + x^3 + x$ である。この $A(x)$ に $x^m = x^3$ を掛ける。これは、後ろに $m=3$ 個の検査ビットをつけるため、情報ビットを三つ左へシフトしていることになる。 $x^3 A(x) = x^7 + x^6 + x^4$ を $G(x)$ で割ると、図 6.6(a) となり、その余りは x である。これを $m-1$ 次多項式($m=3$)として 2 元系列に直すと 010 となる。したがって、情報ビットの後に検査ビット 010 をつけ、符号語は 11010010 となる。

上の割算は系列のままでも同様に、図 6.6(b) のように行うことができる。

$$\begin{array}{r} x^4 + x^3 + x^2 + x \quad \dots \text{商} \\ x^3 + x + 1 \overline{) x^7 + x^6 + x^4} \\ \underline{x^7 + x^5 + x^4} \\ x^6 + x^5 \\ \underline{x^6 + x^4 + x^3} \\ x^5 + x^4 + x^3 \\ \underline{x^5 + x^3 + x^2} \\ x^4 + x^2 \\ \underline{x^4 + x^2 + x} \\ x \quad \dots \text{余り} \end{array}$$

(a)

$$\begin{array}{r} 11110 \\ 1011 \overline{) 11010000} \\ \underline{1011} \\ 1100 \\ \underline{1011} \\ 1110 \\ \underline{1011} \\ 1010 \\ \underline{1011} \\ 0010 \\ \underline{0000} \\ 010 \end{array}$$

(b)

図 6.6 例 6.4 における剰余の求め方。(a) 多項式のまま求める方法、(b) 系列のまま求める方法。

6.11.3 巡回符号の名前の由来

生成多項式を $G(x)$ とする長さ n の巡回符号を考える。いま $G(x)$ が $x^n - 1$ を割り切ると仮定する。ある符号語の符号多項式を

$$W(x) = w_{n-1}x^{n-1} + w_{n-2}x^{n-2} + \dots + w_1x + w_0 \quad (6.38)$$

としよう。このとき、この符号語を左へ一つ巡回的にシフトした符号語の符号多項式は

$$\begin{aligned} W_L(x) &= w_{n-2}x^{n-1} + w_{n-3}x^{n-2} + \dots + w_0x + w_{n-1} \\ &= W(x)x - w_{n-1}x^n + w_{n-1} \\ &= W(x)x - w_{n-1}(x^n - 1) \end{aligned} \quad (6.39)$$

となる。 $W(x)$ は $G(x)$ で割り切れ、また仮定より $(x^n - 1)$ も $G(x)$ で割り切れるから、 $W_L(x)$ もやはり $G(x)$ で割り切れることがわかる。したがって、ある符号語を巡回的にシフトしたものはすべて符号語になっている。これが巡回符号の名前の由来である。

一般の生成多項式は $x^n - 1$ を割り切るとは限らないが、そのようなときでも符号の性質はほぼ同じで同様に扱えるので、この場合も含めて一般的に生成多項式の倍多項式で与えられる符号を巡回符号と呼ぶ。

6.11.4 符号の周期

一般に、多項式 $G(x)$ が割り切ることのできる多項式 $x^p - 1$ の最小の p を、 $G(x)$ の周期という。

生成多項式 $G(x)$ から生成される巡回符号 W の符号長 n は一般に、 $n \leq p$ にする必要がある。なぜなら、もし $n > p$ にしたとすると、多項式 $x^p - 1$ も $n-1$ 次以下であり、かつ $G(x)$ で割り切れるので符号多項式となる。このとき、巡回符号 W は重み 2 の符号語をもつことになり、符号語間の最小距離は 2 以下となる。したがって、誤りの訂正を行うためには、 $n \leq p$ としなければならない。すなわち、符号長は生成多項式の周期以下にしておく必要がある。

【例 6.5】 $G(x) = x^3 + x + 1$ は $x^7 - 1 = x^7 + 1$ を割り切るが、 $x^s - 1$ ($s=1, 2, \dots, 6$) を割り切らない。したがって、この $G(x)$ の周期は 7 である。この $G(x)$ により生成される符号長 $n=7$ の符号語は表 6.6 のように 16 個あり、同じ重みの符号語ごとに巡回的になっている。これは例 6.2 の (7, 4) ハミング符号そのものである。

表 6.6 例 6.5 の $G(x)$ で生成される符号語

$A(x)$	$W(x)$	w
0	0	0 0 0 0 0 0 0
1	$x^3 + x + 1$	0 0 0 1 0 1 1
x	$x^4 + x^2 + x$	0 0 1 0 1 1 0
$x + 1$	$x^4 + x^3 + x^2 + 1$	0 0 1 1 1 0 1
x^2	$x^5 + x^3 + x^2$	0 1 0 1 1 0 0
$x^2 + 1$	$x^5 + x^2 + x + 1$	0 1 0 0 1 1 1
$x^2 + x$	$x^5 + x^4 + x^3 + x$	0 1 1 1 0 1 0
$x^2 + x + 1$	$x^5 + x^4 + 1$	0 1 1 0 0 0 1
x^3	$x^6 + x^4 + x^3$	1 0 1 1 0 0 0
$x^3 + 1$	$x^6 + x^4 + x + 1$	1 0 1 0 0 1 1
$x^3 + x$	$x^6 + x^3 + x^2 + x$	1 0 0 1 1 1 0
$x^3 + x + 1$	$x^6 + x^2 + 1$	1 0 0 0 1 0 1
$x^3 + x^2$	$x^6 + x^5 + x^4 + x^2$	1 1 1 0 1 0 0
$x^3 + x^2 + 1$	$x^6 + x^5 + x^4 + x^3 + x^2 + x + 1$	1 1 1 1 1 1 1
$x^3 + x^2 + x$	$x^6 + x^5 + x$	1 1 0 0 0 1 0
$x^3 + x^2 + x + 1$	$x^6 + x^5 + x^3 + 1$	1 1 0 1 0 0 1

6.11.5 誤り検出および訂正

いま、誤りを含んだ受信系列を

$$Y(x) = W(x) + E(x) \quad (6.40)$$

とする。ここに、 $W(x)$ は符号多項式、 $E(x)$ は誤りパターンの多項式である。誤りの検出および訂正はつぎの手順で行う。

- (1) $Y(x)$ を $G(x)$ で割り、余りを求める。余りが0であれば誤りなし。
- (2) 余りに対応する誤りパターン $E(x)$ を表あるいは計算により求める。
- (3) $Y(x)$ から $E(x)$ を引き、正しい $W(x)$ を求める。

以上のように、巡回符号による誤り検出は簡単な操作で行うことができる。このような巡回符号による誤り検出方式は **CRC**(cyclic redundancy check)方式と呼ばれ、広く実用に供されている。この方式の誤り検出能力についていくつかの定理を示そう。

定理 6.1 $G(x)$ が2項またはそれ以上の項数を有するとき、単一誤りは必ず検出される。

証明 単一誤りの多項式は x^i と書くことができる。 $G(x)$ に2項、またはそれ以上の項数があれば x^i は $G(x)$ の倍多項式になることはない。

定理 6.1 の一般化として、バースト誤りに関し、つぎの定理がある。

定理 6.2 m 次の生成多項式をもつ巡回符号は、長さ m またはそれ以下のバースト誤りを必ず検出する。

証明 バースト誤りの位置を x^i とし、バースト誤りの部分のみ取り出した多項式を $B(x)$ とすると、誤りパターンは

$$E(x) = x^i B(x) \quad (6.41)$$

となる。 $G(x)$ は一般に定数項1を含むように決められる。それゆえ、 $G(x)$ は x を因数としてもたず、 x^i とは互いに素である。さらに、 $B(x)$ の次数は与えられた条件より高々 $m-1$ 次で、 $G(x)$ の次数より低く、 $B(x)$ は $G(x)$ で割り切られることはない。したがって、 $E(x)$ は $G(x)$ で割り切れない。

さらに、長さ $(m+1)$ を越えるバースト誤りの見逃し率は 2^{-m} であることが証明できる。これは、余りに対応する $m-1$ 次の多項式が偶然0になる確率を考えれば察しがつく。たとえば CCITT 勧告の生成多項式 $G(x) = x^{16} + x^{12} + x^5 + 1$ を使用した場合、その見逃し率は 2^{-16} となる。

つぎに、パリティ検査方式も CRC 方式の特殊な場合であることを示そう。

定理 6.3 (パリティ検査)

$G(x) = (x+1)P(x)$ なる生成多項式で構成される巡回符号は偶数個の1を含む。

証明 式(6.34)より、いまの場合

$$\begin{aligned} W(x) &= x^a + x^b + x^c + \dots \\ &= Q(x) \cdot (x+1)P(x) \end{aligned} \quad (6.42)$$

となる。 $x=1$ をこれに代入すると

$$\begin{aligned} W(1) &= 1+1+1+\dots \\ &= Q(1) \cdot (1+1)P(1) = 0 \end{aligned}$$

したがって、多項式 $W(x)$ の項の数は偶数個であることがわかる。

上の定理の特殊な場合として $P(x)=1$ 、すなわち $G(x)=(x+1)$ の場合を考えると検査ビットの数は1ビットであり、通常の(偶数)パリティ検査方式と一致することがわかる。このとき奇数個の誤りは検出される。

生成多項式 $G(x) = x^{16} + x^{12} + x^5 + 1$ は

$$G(x)=(x+1)(x^{15}+x^{14}+x^{13}+x^{12}+x^4+x^3+x^2+x+1) \quad (6.43)$$

のように既約多項式に因数分解されるので、このような巡回符号は偶数個の1を含んでおり、したがって奇数個の誤りは必ず検出できることがわかる。

$x+1$ の周期は明らかに1である。また、 $x^{15}+x^{14}+x^{13}+x^{12}+x^4+x^3+x^2+x+1$ の周期は $2^{15}-1=32767$ であることが確かめられている。異なる既約多項式(それ以上積の形に分解できない多項式)の積の周期は、それぞれの周期の最小公倍数となる(演習問題6.5)。したがって、 $G(x)=x^{16}+x^{12}+x^5+1$ の周期は $p=32767$ である。

定理 6.4 $G(x)=(x+1)A(x)$ で生成される巡回符号は、符号長 n が $A(x)$ の周期に等しければ、長さ 2 以下の任意の位置に生じた二つのバースト誤りを検出できる。また、長さ 2 以下の単一バースト誤りを訂正できる。

証明 前半の長さ 2 以下の任意の位置に生じた二つのバースト誤りが検出できたとする。このとき、6.10 節に述べたことより、後半の長さ 2 以下の単一バースト誤りを訂正できることがわかる。前半を証明する。長さ 2 以下の任意の位置に生じた二つのバースト誤りのパターンは

- (1) $E(x)=x^j+x^i$
- (2) $E(x)=x^j+(x^{i+1}+x^i)$
- (3) $E(x)=(x^{j+1}+x^j)+x^i$
- (4) $E(x)=(x^{j+1}+x^j)+(x^{i+1}+x^i)$

の4通りである。ただし、 $i < j$ としておく。このうち、(2)と(3)は奇数個の誤りであるから、定理6.3より検出可能である。(1)は $E(x)=x^i(x^{j-i}+1)$ であり、 $j-i < n$ であるから、周期 n の $A(x)$ で割り切れない。(4)は $E(x)=(x+1)(x^j+x^i)$ であり、 $(x+1)$ は $G(x)$ 中の $(x+1)$ と打ち消しあうが、残りの (x^j+x^i) は(1)と同じく $A(x)$ で割り切れない。したがって、(1)~(4)はいずれも検出可能である。

つぎに $n \leq p$ の場合に、この $G(x)$ によって生成される巡回符号の最小距離 $d_{\min}$ を求めてみよう。いま、最小距離 $d_{\min}$ が2であったとする。このとき、

$$x^j+x^i=x^i(x^{j-i}+1) \quad (0 \leq i < j < n)$$

なる符号語が存在するはずである。これは $G(x)$ で割り切れなければならない。 $G(x)$ は定数項1をもつように選ばれるから、 x を因数としてもたない。

したがって、 $(x^{j-i}+1)$ が $G(x)$ で割り切れなければならない。 $G(x)$ の周期 p の定義より $p \leq j-i < n$ となり、矛盾が生じる。それゆえ、 $d_{\min} \neq 2$ である。同様にして $d_{\min} \neq 1$ も示すことができる。したがって、 $d_{\min} \geq 3$ である。さらに、 $G(x)$ は $(x+1)$ を因数として含むから、定理6.3より符号語の重みは偶数でなければならない。したがって、 $d_{\min} \neq 3$ である。よって、 $d_{\min} \geq 4$ である。

生成多項式 $G(x)=x^{16}+x^{12}+x^5+1$ の場合についていうと、 $G(x)$ 自身も符号語であり、その重みは4である。それゆえ、 $d_{\min}=4$ であることがわかる。したがって、この場合、3までの誤りは必ず検出できる。

CCITT 勧告の生成多項式 $G(x)=x^{16}+x^{12}+x^5+1$ による巡回符号の誤り検出能力をまとめると、符号長が $2^{15}-1$ を越えない場合、(1) 3個までの誤り、(2) 奇数個の誤り、(3) 長さ16までのバースト誤り、(4) 長さ2までの2重のバースト誤り、はすべて検出でき、(5) 長さ17のバースト誤りの99.997%、(6) それ以上の長さのバースト誤りの99.9985%、を検出する。

6.12 巡回ハミング符号

0,1を係数とする m 次多項式の周期は最大 2^m-1 である。周期に関する重要な定理を示そう(定理の証明は省略する)。

定理 6.5 すべての m に対して、周期 $n=2^m-1$ の m 次多項式が必ず存在する。また、どんな m 次多項式も x を因数として含まなければ、その周期は $n=2^m-1$ 、またはそれ以下である。既約 m 次多項式の周期は 2^m-1 の約数である。

周期がちょうど 2^m-1 となる m 次の多項式を原始多項式 (primitive polynomial) という。表6.7には20次までの各次数につき、代表例として非零項の数がいちばん少ない原始多項式を示す。

ここで、 m 次の原始多項式を生成多項式とする符号長 $n=2^m-1$ の符号について考えよう。符号長 n が周期と一致するから、前節の議論より $d_{\min} \geq 3$ であることがいえる。また、ちょうど $d_{\min}=3$ であることも簡単に証明できる。したがって、この巡回符号は符号長 $n=2^m-1$ 、情報ビット数 $k=2^m-1-m$ 、検査ビット数 m の、単一誤り訂正符号である。これは実はハミング符号である。このようなハミング符号を巡回ハミング符号という。

表 6.7 20 次までの原始多項式の例

次数	原始多項式	次数	原始多項式
1	$x+1$	11	$x^{11}+x^2+1$
2	x^2+x+1	12	$x^{12}+x^6+x^4+x+1$
3	x^3+x+1	13	$x^{13}+x^4+x^3+x+1$
4	x^4+x+1	14	$x^{14}+x^{10}+x^6+x+1$
5	x^5+x^2+1	15	$x^{15}+x+1$
6	x^6+x+1	16	$x^{16}+x^{12}+x^3+x+1$
7	x^7+x+1	17	$x^{17}+x^3+1$
8	$x^8+x^4+x^3+x^2+1$	18	$x^{18}+x^7+1$
9	x^9+x^4+1	19	$x^{19}+x^5+x^2+x+1$
10	$x^{10}+x^3+1$	20	$x^{20}+x^3+1$

【例 6.6】 多項式 $G(x)=x^3+x+1$ は 3 次の原始多項式 ($m=3$) であり, これから生成される長さ $n=2^m-1=7$ の巡回符号は巡回ハミング符号である。この符号の生成行列 G および検査行列 H を求めてみよう。長さ $k=7-3=4$ の情報ビット列を x とするとき, 符号語は

$$w=xG$$

で与えられる。したがって, 情報ビットのうち右から第 i ビットのみが 1 で, 残りは 0 であるような情報ビット列 x_i ($i=1, 2, 3, 4$) (多項式表現は x^{i-1}) に対する符号語 w_i が G の第 $k+1-i$ 行である。この符号語は多項式 $x^m \cdot x^{i-1} = x^{i+2}$ に対する剰余 $R(x)$ から求めることができる。表 6.8 の符号多項式より, 生成行列 G は

$$G = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix} \quad (6.44)$$

で与えられる。また, 検査行列 H は

$$H = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{bmatrix} \quad (6.45)$$

である。この符号は, 例 6.2 と同じものである。

表 6.8 x^{i+2} ($i=1, 2, 3, 4$) に対する剰余多項式 $R(x)$ と符号多項式

i	剰余多項式 $R(x)$	符号多項式
1	$x+1$	$x^3 + x + 1$
2	x^2+x	$x^4 + x^2 + x$
3	x^2+x+1	$x^5 + x^2 + x + 1$
4	$x^2 + 1$	$x^6 + x^2 + 1$

巡回符号のもっと広いクラスで, ハミング符号を含み, かつそれを拡張したものとして, BCH (Bose-Chaudhuri-Hocquenghem) 符号がある。これは, 広い範囲の符号長と情報ビット数に対して高い誤り訂正能力をもったものであるが, 本書では取り扱わない。

演習問題 6.

6.1 生成多項式を $G(x)=x^3+x+1$ とする巡回符号を考える。

- (1) 情報ビットを 0110 とする符号語をつくれ。
- (2) 受信語 11001 は 1 ビットの誤りを含む。誤りを訂正せよ。

$$w_1=1011111$$

$$w_2=0110101$$

$$w_3=0000000$$

$$w_4=1101010$$

$$w_5=0010011$$

$$w_6=1111001$$

$$w_7=0100110$$

$$w_8=1001100$$

6.2 (1) 図 6.7 の巡回符号 C の生成多項式 $G(x)$ を求めよ。

- (2) 何ビットまでの誤りであれば必ず検出できるか。ただし, 誤り訂正は行わない。

(3) C の生成行列 G を求めよ。

図 6.7 演習問題 6.2 の巡回符号

6.3 (1) 多項式 $g(x)=x^4+x+1$ を生成多項式とする符号長 15 の巡回ハミング符号の生成行列, および検査行列を求めよ。

- (2) この符号を使って情報ビット列 $x=(01101000111)$ を送りたい。対応する符号語 w をつくれ。

(3) 受信語 $y=(101101101001101)$ は 1 ビットの誤りを含む。もとの符号語は何か。

6.4 生成多項式 $G(x)$ の周期を p とする。 $G(x)$ が多項式 $A(x)$ を割り切ることを $G(x) \mid A(x)$

と記すことにする。このとき,

$$G(x) \mid x^n - 1$$

であるため必要十分条件は, n が p の倍数, すなわち $n=ap$ ($a=1, 2, \dots$) であることを証明せよ。

6.5 生成多項式が $G(x)=G_1(x)G_2(x)\cdots G_s(x)$ のように互いに異なる既約多項式 $G_1(x), G_2(x), \dots, G_s(x)$ の積に因数分解されたとする。このとき, $G(x)$ の周期 p は各既約多項式 $G_1(x), G_2(x), \dots, G_s(x)$ の周期 $p_1, p_2, \dots, p_s$ の最小公倍数 $p=\text{LCM}(p_1, p_2, \dots, p_s)$ であることを示せ。

7

標本化定理

アナログ信号波をデジタル化するためには、一定の間隔でサンプリング(標本化)し、それらによって得られたアナログ標本値を **A/D 変換** により、デジタル値にすればよい。このとき、どのような間隔で標本化すればよいかを与えるのが標本化定理である。これはもとの信号の最高周波数の2倍の速度で標本化を行えばよい、表現を変えると、信号の最高周波数の1周期間に二つの標本点をとればよいという定理である。これを満たさないときにはスペクトルの折返し歪み(エイリアシング現象)が生じる。

7.1 簡単な例

本論に入る前に簡単な計算で標本化定理をみてみよう。

図 7.1 に示すように、周波数 f [Hz] をもつ三角関数波 $x(t)$ があり、それを f_s [Hz] の速度で等間隔標本化(サンプリング)することを考える。 $x(t)$ は余弦波

$$x(t) = \cos 2\pi f t \quad (7.1)$$

で与えられるとする。サンプリングパルスは周期をもつので、それは基本波とその整数倍の周波数をもつ高調波に分解できる。その基本波を、

$$g(t) = \cos 2\pi f_s t \quad (7.2)$$

とする。 $x(t)$ をサンプリングするということは、 $x(t)$ とサンプリングパルスを乗算するということである。サンプリングパルスは周波数 f_s [Hz] の基本波とその整数倍の周波数 nf_s [Hz] ($n=2, 3, \dots$) の高調波に分解できるが、基本波

7.1 簡単な例

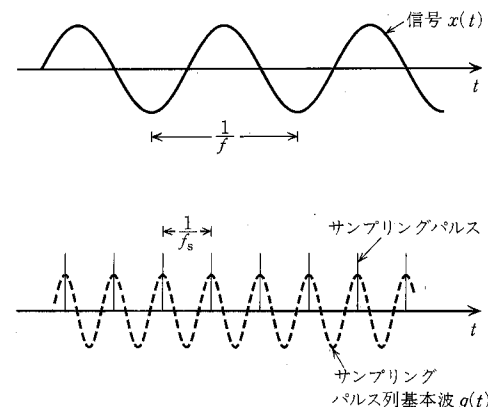


図 7.1 三角関数波の標本化

に関していえばサンプリングにより、

$$\begin{aligned} x(t)g(t) &= \cos 2\pi f t \cos 2\pi f_s t \\ &= \frac{\cos 2\pi (f_s + f) t + \cos 2\pi (f_s - f) t}{2} \end{aligned} \quad (7.3)$$

なる成分が生ずる。この式は、一般にサンプリングなどの乗算や非線形特性により二つの波の和と差の周波数成分が出現することを示している。式(7.3)の分子の第1項は高い周波数であるから問題はないが、問題は第2項である。 $f_s - f$ なる周波数は小さな値となるが、これがもとの信号付近にあれば原信号と区別がつかなくなってくる。したがって少なくとも、

$$f_s - f > f$$

すなわち、

$$f_s > 2f \quad (7.4)$$

であることが必要である。サンプリングパルスの高調波に関しては周波数が高いので、このような干渉の問題は起こらない。結局、式(7.4)、すなわち、「サンプリング周波数は原信号の最高周波数の2倍以上必要である」ことがわかる。これが標本化定理の主要な結論である。

実際にこれを満たさないことによる影響を二、三みてみよう。図 7.2 は、二つの縞の周波数が近いことによるモアレ縞であるが、これは一方を信号とし、もう一方をサンプリングパルスとみれば上記の議論にあてはまる。両者を重ねて置いただけであるが、サンプリングのような乗算的操作が行われた結果、低

周波成分がモアレ縞として出現しているのである。以下にくわしい標本化定理の導出を示そう。



図 7.2 モアレ縞。二つのパターンの相互作用によるエイリアシングの一種。二つのパターンの周波数が近い場合信号と区別できない干渉縞が現われている。

7.2 フーリエ級数

一般に T を周期とする偶関数 $x(t)$ は、

$$x(t) = \sum_{k=0}^{\infty} a_k \cos \frac{2\pi kt}{T}, \quad (7.5)$$

また $x(t)$ が T を周期とする奇関数であれば、

$$x(t) = \sum_{k=1}^{\infty} \beta_k \sin \frac{2\pi kt}{T} \quad (7.6)$$

のように級数展開できる。 $x(t)$ が偶関数の場合、奇関数の場合、およびそれ以外の場合を含めてもっと一般に $x(t)$ は、

$$x(t) = \frac{a_0}{2} + \sum_{k=1}^{\infty} \left(a_k \cos \frac{2\pi kt}{T} + \beta_k \sin \frac{2\pi kt}{T} \right) \quad (\text{フーリエ級数}) \quad (7.7)$$

のように級数展開できる。この式(7.7)はフーリエ級数(Fourier series)と呼ばれる。これは直交関数による $x(t)$ の級数展開の一種である。ここに、 a_k および β_k はその展開係数であり、次式で与えられる：

$$a_k = \frac{2}{T} \int_{-T/2}^{T/2} x(t) \cos \frac{2\pi kt}{T} dt \quad (k=0, 1, 2, \dots) \quad (7.8)$$

$$\beta_k = \frac{2}{T} \int_{-T/2}^{T/2} x(t) \sin \frac{2\pi kt}{T} dt \quad (k=1, 2, \dots) \quad (7.9)$$

これらは、 $\sin$ 関数および $\cos$ 関数の直交性

$$\frac{2}{T} \int_{-T/2}^{T/2} \cos \frac{2\pi mt}{T} \cdot \cos \frac{2\pi nt}{T} dt = \begin{cases} 1 & (m=n \text{ のとき}) \\ 0 & (m \neq n \text{ のとき}) \end{cases} \quad (7.10)$$

$$\frac{2}{T} \int_{-T/2}^{T/2} \sin \frac{2\pi mt}{T} \cdot \sin \frac{2\pi nt}{T} dt = \begin{cases} 1 & (m=n \text{ のとき}) \\ 0 & (m \neq n \text{ のとき}) \end{cases} \quad (7.11)$$

$$\frac{2}{T} \int_{-T/2}^{T/2} \sin \frac{2\pi mt}{T} \cdot \cos \frac{2\pi nt}{T} dt = 0 \quad (7.12)$$

から得られる。式(7.8), および式(7.9)の a_k , β_k はフーリエ係数 (Fourier coefficient) と呼ばれる。

【例 7.1】 図 7.3 は区間 $[-T/2, T/2]$ † において定義された矩形波 $x(t)$ のフーリエ級数展開の結果を表わしている。区間 $[-T/2, T/2]$ において、 $x(t)$ は次式で与えられたとする：

$$x(t) = \begin{cases} 1 & (-T/4 \leq t < T/4) \\ 0 & (\text{その他}) \end{cases} \quad (7.13)$$

全区間においては $x(t)$ は周期 T をもつ周期関数となっているが、その一周期に対応する一つの区間だけに注目している場合には周期性は関係がない。いま、 $x(t)$ を区間 $[-T/2, T/2]$ においてフーリエ級数で表わすとき、その展開係数は以下のように与えられる：

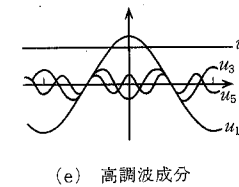
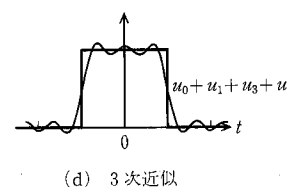
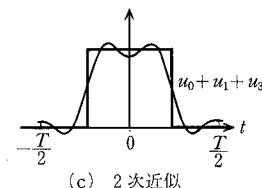
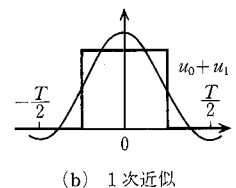
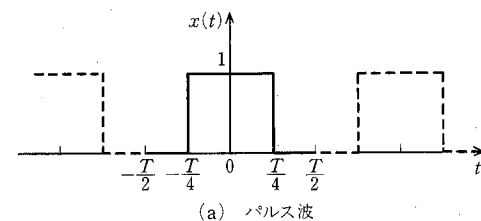


図 7.3 パルス波の正弦波による合成

† $[a, b)$ は下限 a を含み、上限 b を含まない区間を表わす。

$$\alpha_k = \frac{2}{T} \int_{-T/4}^{T/4} \cos \frac{2\pi kt}{T} dt = \begin{cases} 1 & (k=0 \text{ のとき}) \\ 0 & (k=2m \text{ のとき}) \\ (-1)^m \frac{2}{\pi k} & (k=2m+1 \text{ のとき}) \end{cases} \quad (7.14)$$

$$\beta_k = 0 \quad (7.15)$$

これらの係数(振幅)を用いて

$$u_k = \begin{cases} \frac{\alpha_0}{2} & (k=0) \\ \alpha_k \cos \frac{2\pi kt}{T} & (k=1, 2, \dots) \end{cases} \quad (7.16)$$

とすると、 u_k の和で矩形波がよりよく近似されていく様子が図 7.3 に示されている。なお、このフーリエ級数表現

$$x(t) = \frac{\alpha_0}{2} + \sum_{k=1}^{\infty} \left(\alpha_k \cos \frac{2\pi kt}{T} + \beta_k \sin \frac{2\pi kt}{T} \right) \quad (7.17)$$

は定義域外では図 7.3(a) の点線で示すように、周期 T をもつ周期関数となっている。

フーリエ級数は複素数を用いると表現が簡単になる。これを以下に示す。なお、電気関係の文献では電流の記号 i と区別するため虚数単位を $j (= \sqrt{-1})$ と表現することが多いので、本書はこれに従う。

式(7.17)を

$$\cos \theta = \frac{e^{j\theta} + e^{-j\theta}}{2}, \quad \sin \theta = \frac{e^{j\theta} - e^{-j\theta}}{2j} \quad (7.18)$$

を使って書き直すと、

$$\begin{aligned} x(t) &= \frac{\alpha_0}{2} + \sum_{k=1}^{\infty} \left(\alpha_k \frac{e^{j2\pi kt/T} + e^{-j2\pi kt/T}}{2} + \beta_k \frac{e^{j2\pi kt/T} - e^{-j2\pi kt/T}}{2j} \right) \\ &= \sum_{k=-\infty}^{\infty} a_k \cdot e^{j2\pi kt/T} \quad (\text{複素フーリエ級数}) \end{aligned} \quad (7.19)$$

となる。ここに

$$a_k = \begin{cases} \frac{1}{2}(\alpha_k - j\beta_k) & (k=0, 1, 2, \dots) \\ \frac{1}{2}(\alpha_{-k} + j\beta_{-k}) & (k=-1, -2, \dots) \end{cases} \quad (7.20)$$

である。また、 $\beta_0 = 0$ である。式(7.19)を複素フーリエ級数という。係数 a_k は複素フーリエ係数と呼ばれ、式(7.8)、(7.9)、(7.20)より

$$a_k = \frac{1}{T} \int_{-T/2}^{T/2} x(t) e^{-j2\pi kt/T} dt \quad (k=\dots, -2, -1, 0, 1, 2, \dots) \quad (\text{複素フーリエ係数}) \quad (7.21)$$

で与えられる。なお、 a_k と a_{-k} はお互いに複素共役の関係にある。

式(7.7)は実数のみを用いて級数表現をしているのに対して、複素フーリエ級数では複素数を用いている。複素数を用いた場合、項の形が一種類で簡単になる利点がある。その際、 $e^{-j2\pi kt/T} = \cos(-2\pi kt/T) + j \sin(-2\pi kt/T)$ の形の負の周波数成分がでてくる。負の周波数の存在や正の周波数でもそれらが複素数であることは奇妙なことであるが、実際には負の周波数単独で観測されることはなく、その複素共役である正の周波数と対になって現れる。すなわち、

$$x(t) = \sum_{k=-\infty}^{\infty} a_k \cdot e^{j2\pi kt/T} \quad (7.22)$$

において、 $k = -K$ と $k = K$ の項を加えたものは

$$\begin{aligned} & a_{-K} \cdot e^{-j2\pi Kt/T} + a_K \cdot e^{j2\pi Kt/T} \\ &= \frac{1}{2}(\alpha_K + j\beta_K) \left(\cos \frac{2\pi Kt}{T} - j \sin \frac{2\pi Kt}{T} \right) \\ & \quad + \frac{1}{2}(\alpha_K - j\beta_K) \left(\cos \frac{2\pi Kt}{T} + j \sin \frac{2\pi Kt}{T} \right) \\ &= \alpha_K \cos \frac{2\pi Kt}{T} + \beta_K \sin \frac{2\pi Kt}{T} \end{aligned} \quad (7.23)$$

のように実数である。周波数成分は振幅と位相の二つの属性をもつが、複素数を用いると、その両方が同時に一つの複素数で表現でき、表現が簡単になるのである。

7.3 フーリエ変換

フーリエ級数は基本的には周期関数に対して適用できるものである。それに対して周期性を必ずしももたない関数に対して適用されるのがフーリエ変換である。いま、式(7.19)において $T \rightarrow \infty$ の極限を考える。これは周期が無限大、すなわち周期性がなくなった場合に相当する。このとき、 $(1/T) = \Delta f$ 、 $f = k\Delta f = k/T$ 、 $a_k = X(f)\Delta f$ とおくと、式(7.21)は

$$X(f) = \int_{-\infty}^{\infty} x(t) e^{-j2\pi ft} dt \quad (\text{フーリエ変換}) \quad (7.24)$$

また、式(7.22)は

$$x(t) = \int_{-\infty}^{\infty} X(f) e^{j2\pi ft} df \quad (\text{フーリエ逆変換}) \quad (7.25)$$

となる。式(7.24)はフーリエ変換と呼ばれる。 $X(f)$ は複素振幅スペクトル密度(または単にスペクトル)と呼ばれ、 $x(t)$ 中の周波数 f 付近の波の成分の単

位周波数当たりの密度を表わす。これは大文字で表現することが多い。また、式(7.25)はフーリエ逆変換と呼ばれ、 $X(f)$ から $x(t)$ を復元する式である。これらのいくつかの例を付表2に示す。

7.4 標本化定理の導出

周波数帯域が W [Hz] 以下に制限された時間関数 $x(t)$ を考えよう。この関数の周波数スペクトルは、次のフーリエ変換で与えられる：

$$X(f) = \int_{-\infty}^{\infty} x(t) e^{-j2\pi ft} dt. \quad (7.26)$$

周波数帯域は W [Hz] 以下に制限されているので、 $X(f)$ は図 7.4(a) のように W [Hz] 以上で値は 0 である。 $X(f)$ を使うと $x(t)$ はフーリエ逆変換

$$x(t) = \int_{-\infty}^{\infty} X(f) e^{j2\pi ft} df = \int_{-W}^W X(f) e^{j2\pi ft} df \quad (7.27)$$

で表わされる。式(7.27)の積分は $-W \sim +W$ の範囲で計算されるから、 $X(f)$ は図 7.4(b) のように $X(f)$ を周期関数化した $X'(f)$ で置き換えてもよい。すなわち、

$$x(t) = \int_{-W}^W X'(f) e^{j2\pi ft} df. \quad (7.28)$$

ここで、

$$X'(f) = \sum_{k=-\infty}^{\infty} X(f-2kW). \quad (7.29)$$

式(7.29)は周期 $2W$ の周期関数であるから、フーリエ級数で表わすことが

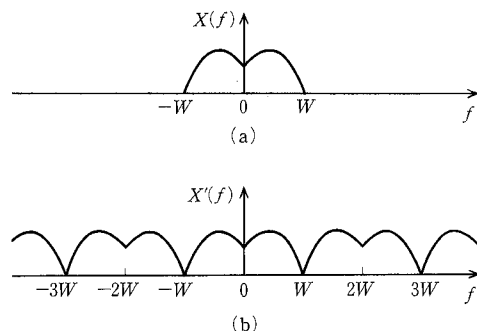


図 7.4 帯域制限スペクトル(a)と、その周期関数化(b)

できる。すなわち、

$$X'(f) = \sum_{k=-\infty}^{\infty} a_k e^{-j\pi kf/W}. \quad (7.30)$$

この式を(7.28)に代入するとつぎのようになる：

$$\begin{aligned} x(t) &= \int_{-W}^W \sum_{k=-\infty}^{\infty} a_k e^{-j\pi kf/W} e^{j2\pi ft} df \\ &= \sum_{k=-\infty}^{\infty} a_k \int_{-W}^W e^{j(2\pi t - \pi k/W)f} df \\ &= \sum_{k=-\infty}^{\infty} W a_k \frac{e^{j(2\pi Wt - k\pi)} - e^{-j(2\pi Wt - k\pi)}}{j(2\pi Wt - k\pi)} \\ &= \sum_{k=-\infty}^{\infty} 2W a_k \frac{\sin(2\pi Wt - k\pi)}{2\pi Wt - k\pi}. \end{aligned} \quad (7.31)$$

a_k を求めるため、 $t = l/2W$ とおくと式(7.31)は、

$$\begin{aligned} x\left(\frac{l}{2W}\right) &= \sum_{k=-\infty}^{\infty} 2W a_k \frac{\sin(l-k)\pi}{(l-k)\pi} \\ &= 2W a_l. \end{aligned} \quad (7.32)$$

なぜなら、

$$\lim_{x \rightarrow 0} \frac{\sin \pi x}{\pi x} = \text{sinc}(0) = 1 \quad (7.33)$$

ここに、

$$\text{sinc}(x) \triangleq \frac{\sin \pi x}{\pi x}$$

である。したがって、式(7.32)を用いると式(7.31)は、

$$x(t) = \sum_{k=-\infty}^{\infty} x\left(\frac{k}{2W}\right) \frac{\sin 2\pi W(t - k/2W)}{2\pi W(t - k/2W)} \quad (\text{標本化定理}) \quad (7.34)$$

と表わされる。ここで、

$$\frac{\sin 2\pi W(t - k/2W)}{2\pi W(t - k/2W)} = \text{sinc}\left[2W\left(t - \frac{k}{2W}\right)\right] \quad (\text{標本化関数}) \quad (7.35)$$

は図 7.5 に示すように、 $\text{sinc}(2Wt)$ を $k/2W$ だけ右にシフトした関数である。

式(7.34)の意味するところをみてみよう。右辺の $x(k/2W)$ は $x(t)$ の $k/2W$ における値、すなわち $1/2W$ ごとの標本値を表わす。標本化関数と呼ばれる $\text{sinc}[2W(t - k/2W)]$ は既知の関数形状をもつ。 $x(t)$ は任意の t についての関数値である。すなわち、任意の時刻の関数値 $x(t)$ が、 $1/2W$ ごとの標本点における標本値 $x(k/2W)$ ($k=0, \pm 1, \pm 2, \dots$) のみから完全に定まること

を式(7.34)は述べている。図 7.6 を用いて説明する。 $x(t)$ は t の付近の標本点

$$\dots, \frac{0}{2W}, \frac{1}{2W}, \frac{2}{2W}, \frac{3}{2W}, \frac{4}{2W}, \frac{5}{2W}, \dots$$

における標本値

$$\dots, x\left(\frac{0}{2W}\right), x\left(\frac{1}{2W}\right), \dots, x\left(\frac{5}{2W}\right), \dots$$

を高さとする標本化関数(点線)の和として完全に定まる。

なお、標本化関数は異なる k の値に関して直交関数系をなしており、式(7.34)は、 $f(x)$ の標本化関数による直交関数展開となっている。

$$\text{sinc}(2Wt) = \frac{\sin 2\pi Wt}{2\pi Wt} \quad \text{sinc}\left[2W\left(t - \frac{k}{2W}\right)\right] = \frac{\sin 2\pi W\left(t - \frac{k}{2W}\right)}{2\pi W\left(t - \frac{k}{2W}\right)}$$

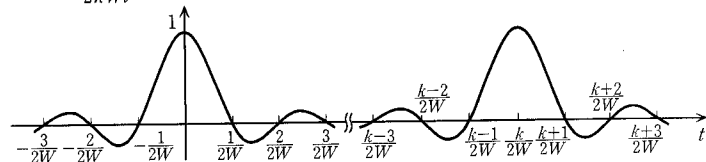


図 7.5 標本化関数

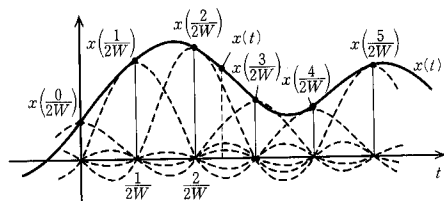
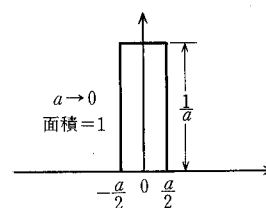


図 7.6 標本化定理。任意の点の値 $x(t)$ は $1/2W$ ごとの周辺の標本点を中心とする標本化関数(点線)の総和で与えられる。

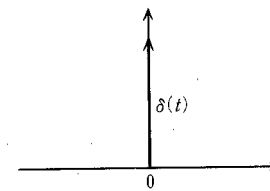
7.5 デルタ関数

図 7.7 のような特殊な関数を考える。すなわち、(a) のように面積を一定 (=1) に保ったまま幅が $\rightarrow 0$ 、高さが $\rightarrow \infty$ になった矩形パルスの極限を考える。このような極限で定義される関数を δ 関数または単位インパルス関数といい、 $\delta(t)$ で表す。この関数の性質を次に記す。

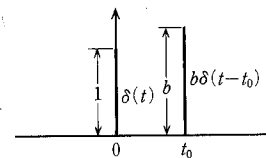
7.5 デルタ関数



(a) δ 関数は $a \rightarrow 0$ の極限として定義される。



(b) δ 関数の表示法 1



(c) δ 関数の表示法 2

図 7.7 δ 関数

- (1) $t \neq 0$ のとき $\delta(t) = 0$
- (2) $t = 0$ のとき $\delta(0)$ は定義されないが、 $t = 0$ 付近の積分値は、

$$\lim_{\epsilon \rightarrow 0} \int_{-\epsilon}^{\epsilon} \delta(t) dt = 1 \quad (\epsilon > 0) \quad (7.36)$$

となる。また、一般の関数 $f(x)$ に対して、

$$\int_{-\infty}^{\infty} f(t) \delta(t) dt = f(0) \quad (7.37)$$

なる性質をもつ。すなわち、 δ 関数は $f(t)$ の $t = 0$ の部分を切り取り出してくる。

また、図 7.8 に示すような三角形を $a \rightarrow 0$ とした極限としても δ 関数が定義できる。同様に sinc 関数によっても定義することができる。

δ 関数は通常関数の範囲を超えた関数であるので超関数といわれる。

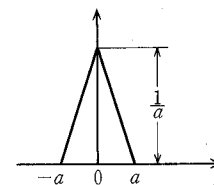


図 7.8 三角波による δ 関数の定義

7.6 インパルス応答

入力端子1個と出力端子1個をもち、入力の大きさが2倍になれば出力の大きさも2倍になるような記憶をもつ線形システムを考える。このような系にデルタ関数で表わされるインパルスを時刻 $t=0$ で加えたとする(図7.9)。このときの出力関数 $h(t)$ をこの系の**インパルス応答**という。

$h(t)$ の形は、インパルスを加える時刻によらないとする。すなわち、入力が $\delta(t-\tau)$ のとき出力は $h(t-\tau)$ となる。いまこの系に $x(t)$ なる入力を加えたとする。 $x(t)$ は図7.10に示すように、幅の狭い矩形波の集まりで構成されていると考えることができる。矩形波は極限としてデルタ関数になるから、このときの出力関数 $y(t)$ は

$$\begin{aligned} y(t) &= \lim_{\Delta t \rightarrow 0} \sum_{k=-\infty}^{\infty} x(k\Delta t) \Delta t \cdot h(t-k\Delta t) \\ &= \int_{-\infty}^{\infty} x(\tau) h(t-\tau) d\tau \end{aligned} \quad (7.38)$$

となる。式(7.38)の積分範囲は**因果律**(インパルス入力が入った後に出力がでる)を満足する場合には $\tau=t$ までで十分であるが、理論的な考察を行う場合には ∞ までとしたほうが便利である。式(7.38)の積分において変数 $t-\tau$ を改めて τ とおくと

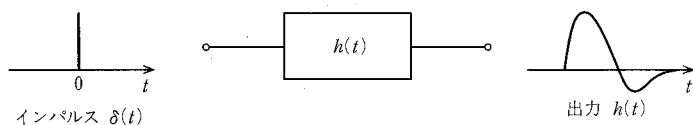


図 7.9 インパルス応答

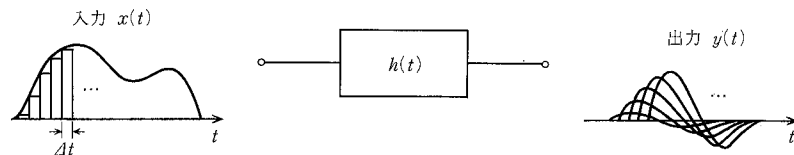


図 7.10 一般の入力関数に対する出力応答。出力は入力中の各インパルスに対する応答の和である。

$$y(t) = \int_{-\infty}^{\infty} x(t-\tau) h(\tau) d\tau \quad (7.39)$$

と表わせる。式(7.38)や式(7.39)は**畳み込み積分**と呼ばれ、

$$\begin{aligned} y(t) &= x(t) * h(t) \\ &= h(t) * x(t) \end{aligned} \quad (7.40)$$

と記される。

ここで、式(7.39)の両辺をフーリエ変換すると

$$\begin{aligned} Y(f) &= \int_{-\infty}^{\infty} e^{-j2\pi ft} \left[\int_{-\infty}^{\infty} x(\tau) h(t-\tau) d\tau \right] dt \\ &= \int_{-\infty}^{\infty} x(\tau) \left[\int_{-\infty}^{\infty} e^{-j2\pi ft} h(t-\tau) dt \right] d\tau \quad (\text{積分順序交換}) \\ &= \int_{-\infty}^{\infty} x(\tau) e^{-j2\pi f\tau} \left[\int_{-\infty}^{\infty} e^{-j2\pi f\lambda} h(\lambda) d\lambda \right] d\tau \\ &\quad (\text{積分変数変換}) \\ &= X(f) H(f). \end{aligned} \quad (7.41)$$

ここに $H(f)$ はインパルス応答 $h(t)$ のフーリエ変換で**伝達関数**と呼ばれる。

したがって、時間関数の畳み込み積分は、周波数領域においては入力と伝達関数の積で表わされる。この様子を図7.11に示す。

いま、 $x(t) = \delta(t)$ とすると

$$\begin{aligned} X(f) &= \int_{-\infty}^{\infty} \delta(t) e^{-j2\pi ft} dt \\ &= 1 \end{aligned} \quad (7.42)$$

したがって、このときの出力は

$$Y(f) = H(f) \quad (7.43)$$

となる。すなわち、インパルス応答が出力されることが確かめられる。

$$\begin{array}{ccc} x(t) \text{ 入力} & \boxed{\begin{array}{c} h(t) \\ H(f) \end{array}} & \text{出力 } y(t) = \int_{-\infty}^{\infty} x(\tau) h(t-\tau) d\tau \\ X(f) & & Y(f) = X(f) H(f) \end{array}$$

図 7.11 インパルス応答と伝達関数

7.7 自己相関関数とパワースペクトル

エネルギー有限な関数 $g(t)$ の自己相関関数を

$$\begin{aligned}\phi(t) &= \int_{-\infty}^{\infty} g(\tau+t) g^*(\tau) d\tau \\ &= \int_{-\infty}^{\infty} g(\tau) g^*(\tau-t) d\tau\end{aligned}\quad (7.44)$$

で定義する。ここで $*$ は複素共役を表す。このとき $\phi(t)$ のフーリエ変換は

$$\begin{aligned}\Phi(f) &= \int_{-\infty}^{\infty} \phi(t) e^{-j2\pi ft} dt \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(\tau+t) g^*(\tau) d\tau e^{-j2\pi ft} dt \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(\tau+t) e^{-j2\pi ft} dt g^*(\tau) d\tau \\ &= G(f) \int_{-\infty}^{\infty} e^{j2\pi f\tau} g^*(\tau) d\tau \\ &= |G(f)|^2\end{aligned}\quad (7.45)$$

で与えられる。この $|G(f)|^2$ は $g(t)$ のパワースペクトルと呼ばれる。すなわち、自己相関関数のフーリエ変換がパワースペクトルである。パワースペクトルは実数値をもつ。

7.8 標本値列からの復元

前節で定義されたデルタ関数を用いると、標本化速度 f_s [Hz] の標本化パルス列は

$$s_p(t) = \sum_{k=-\infty}^{\infty} \delta\left(t - \frac{k}{f_s}\right) \quad (7.46)$$

で表わすことができる。これは $T=1/f_s$ を周期とする周期関数である。したがって、 $[-T/2, T/2]$ で複素フーリエ級数展開することができる。すなわち、

$$\begin{aligned}s_p(t) &= \sum_{k=-\infty}^{\infty} \delta\left(t - \frac{k}{f_s}\right) \\ &= \sum_{n=-\infty}^{\infty} a_n e^{j2\pi n f_s t}.\end{aligned}\quad (7.47)$$

ここで、

7.8 標本値列からの復元

$$\begin{aligned}a_n &= \frac{1}{T} \int_{-T/2}^{T/2} \left[\sum_{k=-\infty}^{\infty} \delta(t-kT) \right] e^{-j2\pi nt/T} dt \\ &= \frac{1}{T} \int_{-T/2}^{T/2} \delta(t) e^{-j2\pi nt/T} dt \\ &= \frac{1}{T} = f_s.\end{aligned}\quad (7.48)$$

したがって、

$$s_p(t) = f_s \sum_{n=-\infty}^{\infty} e^{j2\pi n f_s t}.\quad (7.49)$$

すなわち、標本化パルス列は 7.2 節のフーリエ級数の項でも少しふれたように、周波数 f_s [Hz] の基本波とその整数倍の周波数 $m f_s$ [Hz] ($m=2, 3, \dots$) の高調波で構成されている。標本化は時間関数 $x(t)$ にこの標本化パルス列を掛けることにより行われる。したがってこの標本化パルス列により、時間関数 $x(t)$ は標本化され、

$$x(t) s_p(t) = \sum_{k=-\infty}^{\infty} x\left(\frac{k}{f_s}\right) \delta\left(t - \frac{k}{f_s}\right) \quad (7.50)$$

なる標本値列となる。これは $x(t)$ と $s_p(t)$ の積である。時間関数の積のフーリエ変換は、個々の関数のフーリエ変換の畳み込み積分で表わされることが、式(7.41)と同様に証明できる。したがって、式(7.50)のフーリエ変換は

$$X(f) * S_p(f) \quad (7.51)$$

で与えられる。ここに $S_p(f)$ は $s_p(t)$ のフーリエ変換であり、

$$\begin{aligned}S_p(f) &= \int_{-\infty}^{\infty} s_p(t) e^{-j2\pi ft} dt \\ &= f_s \sum_{n=-\infty}^{\infty} \int_{-\infty}^{\infty} e^{j2\pi n f_s t} \cdot e^{-j2\pi ft} dt \\ &= f_s \sum_{n=-\infty}^{\infty} \delta(f - n f_s).\end{aligned}\quad (7.52)$$

ここでは公式

$$\int_{-\infty}^{\infty} e^{-j2\pi ft} dt = \delta(f) \quad (7.53)$$

を用いた。これは回転する複素数を積分すると 0 であるが、 $f=0$ のときだけは回転が止まることから積分が $\delta(f)$ になることが理解できるであろう。

式(7.52)を用いると式(7.51)は

$$\begin{aligned}
 X(f) * S_p(f) &= f_s \sum_{n=-\infty}^{\infty} \int_{-\infty}^{\infty} X(f-\lambda) \delta(\lambda - nf_s) d\lambda \\
 &= f_s \sum_{n=-\infty}^{\infty} X(f - nf_s)
 \end{aligned} \quad (7.54)$$

となる。

ここで、標本化操作をスペクトルの面から説明する。一般に時間関数の乗算操作により、和と差の周波数成分が生じることは7.1節で述べた。 $x(t)$ の周波数が W [Hz] 以下に帯域制限されているとする。これを図7.12(a)に示す。このとき、もし $2W < f_s$ ならば上の議論より標本値列(7.54)のスペクトルは図7.12(b)のようになるであろう。すなわち、標本化パルス列中の直流成分との積によって得られた第0次スペクトル $X(f)$ の成分、および基本波、高調波と

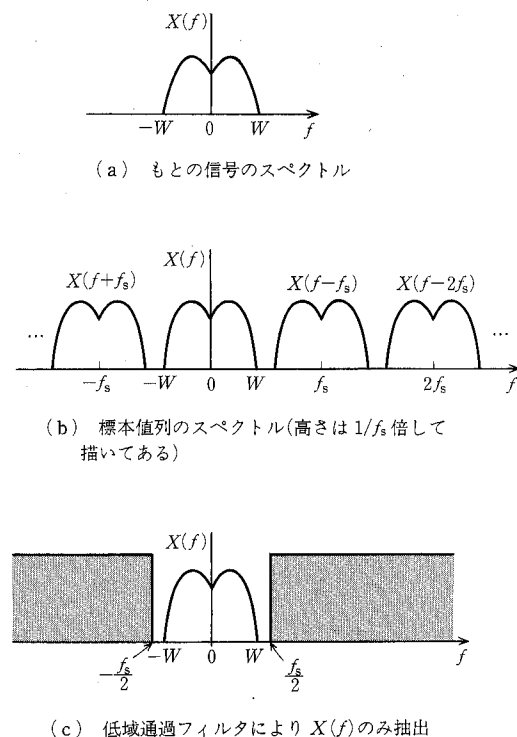


図 7.12 標本値列からの原信号 $x(t)$ の復元(標本化定理を満している場合)

の積により $\pm mf_s$ の周波数シフトを受けた第 m 次スペクトル $X(f \pm mf_s)$ ($m=1, 2, \dots$)の成分からなる。したがって、この標本値列(7.54)を図7.13に示すように遮断周波数 $f_s/2$ 、振幅伝達比 $1/f_s$ の低域通過フィルタに通すと、もとの信号(第0次スペクトル) $X(f)$ 、すなわち $x(t)$ が復元される。

これをインパルス応答の立場から説明しよう。簡単化のため、標本化速度はちょうど $f_s = 2W$ とする。遮断周波数 W 、伝達係数 $(1/2W)$ の低域通過フィルタはフーリエ逆変換を用いるとインパルス応答

$$\begin{aligned}
 h(t) &= \int_{-W}^W \frac{1}{2W} e^{j2\pi ft} df \\
 &= \frac{\sin 2\pi Wt}{2\pi Wt}
 \end{aligned} \quad (7.55)$$

をもつことがわかる。これは標本化関数の形であり、対応する標本点の高さをもち式(7.55)のインパルス応答関数を加え合わせると、もとの信号の復元ができることは式(7.34)で示されている。

標本化定理では、元信号を再現するには $1/2W$ ごとの標本値を必要としたが、 $1/2W$ より広い標本化を行ったときにはどうなるかをみてみよう。図7.14に、ちょうど W [Hz]の波を標本化定理に従って $1/2W$ より狭い間隔で標本化を行った場合(a)と、それに反して $1/2W$ より広い間隔で標本化を行った場合(b)を示す。(a)では各標本値を通る波は一つしか存在しないが(b)ではそれ以外に低い周波数の波が存在しうる。このことより、元信号を忠実に再現しうるための標本化間隔は $1/2W$ 以下であることが理解できるであろう。もしこの条件が満たされないと、低い周波数の波がエイリアシングとして現れてくる。

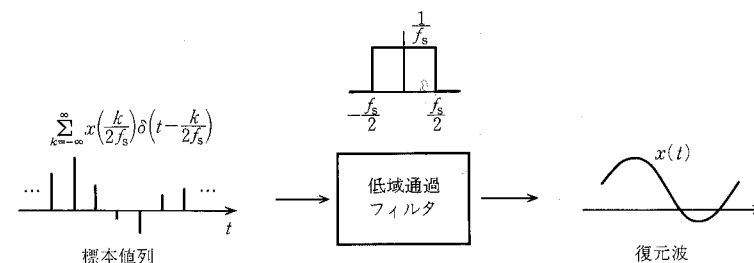
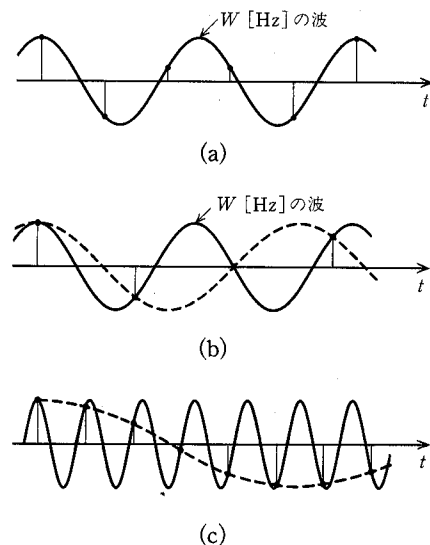
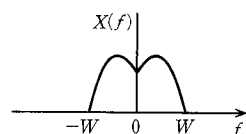
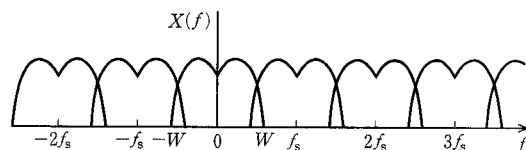


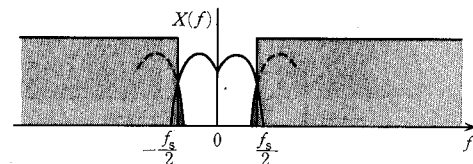
図 7.13 低域通過フィルタによる帯域制限波の標本値列からの復元。

図 7.14 標本間隔 $1/2W$ の必要性

(a) もとの信号のスペクトル



(b) 標本値列のスペクトル

(c) 遮断周波数 $f_s/2$ の低域通過フィルタにより $f_s/2$ 以上の成分が折り返した形で出力される。図 7.15 標本化定理を満たさない標本化 ($2W > f_s$)

極端な場合を(c)に示すが、これは W が $f_s/2$ よりかなり大きな場合であり、この場合は標本化パルス列の高調波成分と入力信号との差の周波数成分(点線)がエイリアシング(折り返し歪)として生じている。

このように、標本化定理の条件を満たさない $2W > f_s$ の場合について、標本値列(7.50)のスペクトル式(7.54)を図7.15に示す。このとき、この標本値列を遮断周波数 $f_s/2$ の低域通過フィルタに通すと、 $f_s/2$ 以上の周波数成分が折り返した形で出力され、エイリアシングを生じることがわかる。

演習問題 7.

7.1 $W=5\text{kHz}$ までの周波数を含む $T=10$ 秒間の信号を、標本化定理に従って標本化する。各標本値は8ビットのデジタルデータに変換される。合計何ビットのデータとなるか。

7.2 $W[\text{Hz}]$ まで含む信号 $g(t)$ を $1/2f_s$ [秒]ごとに標本化し、遮断周波数 f_s の理想低域フィルタに通すと f_s 以上の成分は折り返した形で出力されることが知られている。ただし、その部分の位相は反転する。また、 $W/2 < f_s < W$ とする。

いま、 $g(t) = \sin(2\pi Wt)$ 、 $f_s = (3/4)W$ としてこのことを確かめよ。

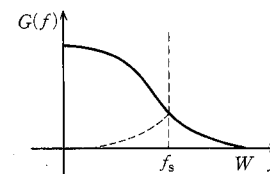


図 7.16

$$\left[\text{ヒント: フィルタ出力} = \sum_k \sin\left(2\pi W \frac{k}{2f_s}\right) \frac{\sin(2\pi f_s t - k\pi)}{2\pi f_s t - k\pi} \right]$$

これが、 $-\sin(\pi Wt)$ となることを示せばよい。

7.3 下記のインパルス列 $g(t)$ を区間 $[-3T/2, 3T/2]$ で複素フーリエ級数で表わせ。

$$g(t) = \sum_{k=-\infty}^{\infty} \delta(t - kT)$$

7.4 エネルギー有限関数 $g(t)$ のパワースペクトルを $|G(f)|^2$ とする。時間をずらした関数 $g(t - t_0)$ のパワースペクトルも $|G(f)|^2$ に等しいことを示せ(パワースペクトルの位置不変性)。

7.5 伝達関数 $H(f)$ の線形回路に $X(f)$ を入力したときの出力のパワースペクトルを求めよ。

7.6 関数 $g(t)$ のフーリエ変換を $G(f)$ とする。このとき、これらのエネルギーが等しい、すなわち

$$\int_{-\infty}^{\infty} |g(t)|^2 dt = \int_{-\infty}^{\infty} |G(f)|^2 df \quad (\text{パーセバルの定理})$$

を証明せよ。ただし

$$\int_{-\infty}^{\infty} e^{j2\pi ft} df = \delta(t)$$

である。

8

アナログ通信路

前章で、帯域制限された時間波形は標本化定理により標本値の組で完全に表わすことができることが示された。すなわち、 W [Hz]までの周波数を含む T [秒]間の信号は $1/2W$ [秒]ごとにとられた合計 $2TW$ 個の標本値 $\{x_1, x_2, \dots, x_{2TW}\}$ で完全に表現できる。したがって、帯域制限時間波形はこれらの標本値を時系列的に発生する情報源からの出力とみなすことができる。このような情報源はアナログ値(連続値)を一定時間ごとに発生する**アナログ情報源(連続的情報源)**でモデル化できる。これと区別する必要がある場合には、離散値(ディジタル値)を出力する従来の情報源は離散的情報源と呼ばれる。

本章では、アナログ情報源から発生するアナログ情報をアナログ通信路を通して伝送するときの問題について述べる。

8.1 確率密度関数

n 個の確率変数 $X_0, X_1, \dots, X_{n-1}$ は n 次元確率密度関数

$$p_{X_0, X_1, \dots, X_{n-1}}(x_0, x_1, \dots, x_{n-1}) \quad (8.1)$$

で特徴づけられる。アナログ量の場合、ディジタル量と異なるのは、確率が確率そのものではなく、確率密度(8.1)の形で与えられることである。 n 次元確率密度関数(8.1)はまた

$$p(x_0, x_1, \dots, x_{n-1}) \quad (8.2)$$

と略記されることもある。確率密度関数は n 次元空間における全確率が1になる。すなわち、

8.1 確率密度関数

97

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} p(x_0, x_1, \dots, x_{n-1}) dx_0 dx_1 \dots dx_{n-1} = 1. \quad (8.3)$$

アナログ値を出力するアナログ情報源の統計的な諸性質、たとえば定常情報源やエルゴード性は離散的な場合と同様に成り立つ。これを1次元確率密度関数について示すと次のようになる。ただし、確率変数 X の確率密度関数を $p(x)$ 、 x の任意の関数を $f(x)$ とする。

[平均]

$$\text{集合平均: } \langle f(X) \rangle = \int_{-\infty}^{\infty} f(x) p(x) dx \quad (8.4)$$

$$\text{時間平均: } \overline{f(X)} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T f(x(t)) dt \quad (8.5)$$

エルゴード情報源ではこれらが等しく

$$\langle f(X) \rangle = \overline{f(X)} \quad (8.6)$$

となる。以下では、エルゴード性は成り立つものと仮定する。したがって、集合平均と時間平均は区別しない。いま、特に $f(x) = x$ とおくと、 X の平均値が

$$\mu \triangleq \overline{X} = \int_{-\infty}^{\infty} x p(x) dx \quad (8.7)$$

で与えられる。また、 $f(x) = (x - \mu)^2$ とおくと分散が

$$\sigma^2 \triangleq \overline{(X - \mu)^2} = \int_{-\infty}^{\infty} (x - \mu)^2 p(x) dx \quad (8.8)$$

で与えられる。

[条件付き確率とベイズの定理]

$$p(x|y) = \frac{p(x, y)}{p(y)} \quad (\text{条件付き確率}) \quad (8.9)$$

$$= \frac{p(y|x)p(x)}{p(y)} \quad (\text{ベイズの定理}) \quad (8.10)$$

[例 8.1] 正規分布

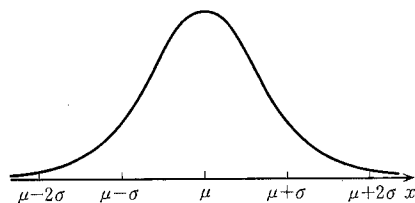
確率密度関数が

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (8.11)$$

で与えられる情報源を考える。この確率密度関数は図8.1のようになる。これを**正規分布(normal distribution)**、または**ガウス分布(Gaussian distribution)**という。1次元正規分布(8.11)の平均は

$$\overline{X} = \int_{-\infty}^{\infty} x p(x) dx = \mu \quad (8.12)$$

また、分散は

図 8.1 1次元正規分布の確率密度関数 $p(x)$

$$\overline{(X-\mu)^2} = \int_{-\infty}^{\infty} (x-\mu)^2 p(x) dx = \sigma^2 \quad (8.13)$$

である。

記憶のない定常情報源からの出力 $X_0, X_1, \dots, X_{n-1}$ はそれぞれ独立で同じ分布 $p(x)$ に従うから、それらの結合確率密度関数は

$$p(x_0, x_1, \dots, x_{n-1}) = \prod_{i=0}^{n-1} p(x_i) \quad (8.14)$$

で与えられる。特に $p(x)$ がガウス分布のとき、この情報源はガウス情報源と呼ばれる。さらに、 $p(x)$ の平均値が 0 のとき、この情報源は白色ガウス情報源と呼ばれる。白色とは、対象とする周波数帯域内でスペクトルが一様(どの周波数成分も均等に含まれている)であることから、白色光にちなみこのように呼ばれる。さらに、先に述べたようにこのような記憶のない定常情報源はエルゴード情報源である。このとき、このような情報源からの出力を電圧もしくは電流とみなし、これを 1Ω の抵抗に印加したときに消費される電力は $x^2(t)$ の時間平均となる。これは $x^2(t)$ の集合平均にも等しく、分散値 σ^2 であるから、 σ^2 を情報源の平均電力と呼ぶ。

8.2 アナログ情報源のエントロピー

量子化とは、アナログ値をいくつかの区間に分け、その区間内のアナログ値はその中の代表値(デジタル値)で表わすことである。

いま図 8.2 に示すように、アナログ情報源 S の標本値 x を量子化ステップ Δx で量子化することを考える。このとき、区間 $(i\Delta x - \frac{1}{2}\Delta x, i\Delta x + \frac{1}{2}\Delta x)$ 内のアナログ値 x はデジタル値 i に変換される。変換後の $\{i\}$ を離散の情報源 $S_{\Delta x}$ の出力とみなす。いま、標本値 x の確率密度関数を $p(x)$ とする。この

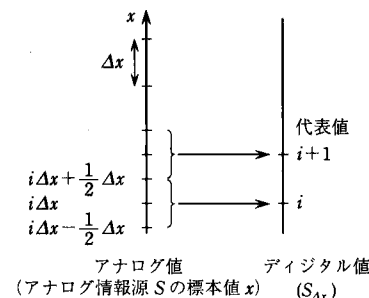


図 8.2 アナログ情報のデジタル化

とき、 x をデジタル化した情報源 $S_{\Delta x}$ の 1 次エントロピーは

$$H_1(S_{\Delta x}) = - \sum_i p_i \log_2 p_i \quad (8.15)$$

で与えられる。ここに、

$$p_i = \int_{i\Delta x - \frac{1}{2}\Delta x}^{i\Delta x + \frac{1}{2}\Delta x} p(x) dx \approx p(i\Delta x) \Delta x. \quad (8.16)$$

したがって、

$$\begin{aligned} H_1(S_{\Delta x}) &\approx - \sum_i p(i\Delta x) \Delta x \log_2 [p(i\Delta x) \Delta x] \\ &= - \sum_i [p(i\Delta x) \log_2 p(i\Delta x)] \Delta x - \log_2 \Delta x \end{aligned} \quad (8.17)$$

である。 $\Delta x \rightarrow 0$ としたとき、 $S_{\Delta x}$ はいくらでももとの S に近づくであろう。そこで、アナログ情報源 S のエントロピーを $H_1(S_{\Delta x})$ の極限で定義したとしよう。すなわち、アナログ情報源 S の 1 次エントロピーを

$$H_1(S) = - \int_{-\infty}^{\infty} p(x) \log_2 p(x) dx - \lim_{\Delta x \rightarrow 0} \log_2 \Delta x \quad (8.18)$$

で定義する。このとき、この第 2 項は無限大となり、発散してしまう。しかしながら、二つのアナログ情報源 S_1 および S_2 があるときのエントロピー差は、 Δx を両方で同じにすると

$$H_1(S_1) - H_1(S_2) = - \int_{-\infty}^{\infty} p_1(x) \log_2 p_1(x) dx + \int_{-\infty}^{\infty} p_2(x) \log_2 p_2(x) dx \quad (8.19)$$

となる。そこで、アナログ情報源 S の 1 次エントロピーを

$$H_1(S) = - \int_{-\infty}^{\infty} p(x) \log_2 p(x) dx \quad (8.20)$$

で定義し直す。実際の使用にあたっては、最終的に相互情報量のような差の形式で使用すれば問題ない。また、記憶のあるアナログ情報源 S の 1 標本値当たりの n 次エントロピーを同様に

$$H_n(S) = - \frac{1}{n} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} p(x_0, x_1, \dots, x_{n-1}) \cdot \log_2 p(x_0, x_1, \dots, x_{n-1}) dx_0 dx_1 \cdots dx_{n-1} \quad (8.21)$$

で定義する。これをもとに、1 標本値当たりのエントロピーを

$$H(S) = \lim_{n \rightarrow \infty} H_n(S) \quad [\text{ビット/標本値}] \quad (8.22)$$

で定義する。

記憶のない定常アナログ情報源 S では、離散系の場合と同じく 1 次エントロピー $H_1(S)$ とエントロピー $H(S)$ は等しい。したがって、そのときにはこれらを区別する必要はない。また、 $H(S)$ を S の標本値 X を用いて $H(X)$ と表わすこともある。

8.3 最大エントロピー定理

電力制限のあるアナログ情報源 S の最大エントロピーを求めてみよう。 S の標本点における出力を X とし、その確率密度関数を $p(x)$ とする。 $p(x)$ は確率密度関数であるから、

$$\int_{-\infty}^{\infty} p(x) dx = 1 \quad (8.23)$$

が成り立つ。また、 X は σ^2 に電力制限され、

$$\int_{-\infty}^{\infty} x^2 p(x) dx = \sigma^2 \quad (8.24)$$

が成り立つとする。これらの条件のもとに、エントロピー

$$H(X) = - \int_{-\infty}^{\infty} p(x) \log_e p(x) dx \quad (8.25)$$

を最大にする $p(x)$ を求めることを考える。式(8.25)の対数の底は解析のつごう上 e となっているが、2 としても係数が異なるだけである。

式(8.23)、(8.24)の条件のもとに、式(8.25)を最大にする問題は変分法で解くことができる。

いま、 λ, μ をラグランジュ (Lagrange) の未定乗数とし、関数

$$\begin{aligned} \phi(X) &= H(X) + \lambda \int_{-\infty}^{\infty} p(x) dx + \mu \int_{-\infty}^{\infty} x^2 p(x) dx \\ &= \int_{-\infty}^{\infty} -p(x) \log_e p(x) + \lambda p(x) + \mu x^2 p(x) dx \end{aligned} \quad (8.26)$$

を導入する。もし、 $p(x)$ が式(8.23)および式(8.24)の拘束条件のもとで $H(X)$ を最大にするなら、 $p(x)$ を少し変動させても $\phi(X)$ の値は変わらないであろう。そこで、式(8.26)の被積分関数を $p(x)$ について偏微分し、0 とおくと

$$-\log_e p(x) - 1 + \lambda + \mu x^2 = 0 \quad (8.27)$$

を得る。このとき、

$$p(x) = e^{\mu x^2 + \lambda - 1} \quad (8.28)$$

これを式(8.23)、(8.24)に代入し、

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} \quad (\text{ガウス分布}) \quad (8.29)$$

を得る。すなわち、 X の分布がガウス分布のとき、エントロピーが最大となる。このとき、最大エントロピーは

$$\begin{aligned} H(X) &= \log_e \sqrt{2\pi e \sigma^2} \quad [\text{ナット/標本値}] \\ &= \log_2 \sqrt{2\pi e \sigma^2} \quad [\text{ビット/標本値}] \end{aligned} \quad (8.30)$$

で与えられる。なお、式(8.30)は σ の単調増加関数であるから、式(8.24)の代わりに

$$\int_{-\infty}^{\infty} x^2 p(x) dx \leq \sigma^2 \quad (8.31)$$

の条件で $H(X)$ の最大値を求めても同じ結果を得ることは明らかであろう。

8.4 通信路容量

図 8.3 のような、白色ガウス雑音を加法的に加わるアナログ通信路を考える。このような通信路は加法的白色ガウス通信路と呼ばれる。 Z は平均 0、分散 N のガウス分布に従うとすると、入力 X が与えられたときの出力 Y の分布の確率密度関数は

$$p(y|x) = \frac{1}{\sqrt{2\pi N}} e^{-\frac{(y-x)^2}{2N}} \quad (8.32)$$

で与えられる。

通信路容量は離散系(ディジタル)の場合と同様に

$$C = \max_{P(x)} \{I(X; Y)\} \quad (8.33)$$

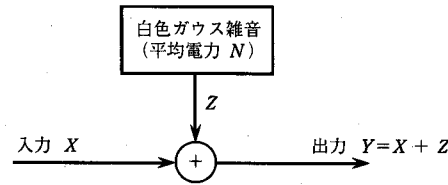


図 8.3 加法的白色ガウス通信路

で定義される。ここに、 $I(X; Y)$ は相互情報量であり、

$$I(X; Y) = H(Y) - H(Y|X) \quad (8.34)$$

また、 $H(Y|X)$ は条件付きエントロピー

$$H(Y|X) = - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x) p(y|x) \log_2 p(y|x) dx dy \quad (8.35)$$

である。式(8.32)を式(8.35)に代入すると

$$H(Y|X) = \log_2 \sqrt{2\pi e N} \quad (8.36)$$

となる。これは X の確率密度関数 $p(x)$ に無関係である。したがって、 $I(H|Y)$ を最大にするには $H(Y)$ を最大にすればよい。

いま、 X と Z が独立とすると

$$\bar{Y}^2 = \bar{X}^2 + \bar{Z}^2 = \bar{X}^2 + N. \quad (8.37)$$

さらに、信号の電力が S で抑えられているとすると

$$\bar{X}^2 \leq S, \quad (8.38)$$

$$\bar{Y}^2 \leq S + N. \quad (8.39)$$

この条件のもとで $H(Y)$ を最大にするには、 Y の分布を分散が $S+N$ のガウス分布とすればよい。実際、 X の分布がガウス分布のとき、ガウス分布の再生性より Y の分布もガウス分布となる(演習問題 8.1)。このとき、

$$H(Y) = \log_2 \sqrt{2\pi e (S+N)}. \quad (8.40)$$

したがって、通信路容量は式(8.34)、(8.36)、(8.40)より

$$C = \frac{1}{2} \log_2 \left(1 + \frac{S}{N} \right) \quad [\text{ビット/標本値}], \quad (8.41)$$

または

$$C = W \log_2 \left(1 + \frac{S}{N} \right) \quad [\text{ビット/秒}] \quad (8.42)$$

が得られる。式(8.42)をシャノン・ハートレー(Shannon-Hartley)の定理という。

[例 8.2] 実効帯域幅 2.4 [kHz], S/N 比 = 1000 $\approx 2^{10}$ (電力比) なる電話回線の通信路容量 C [ビット/秒] を求めよう。

$$\begin{aligned} C &\approx 2.4 \times 10^3 \log_2 (1 + 2^{10}) \\ &= 2.4 \times 10^4 \quad [\text{ビット/秒}] \end{aligned}$$

8.5 量子化雑音

量子化においてはアナログ値がいくつかの区間に分けられ、各区間内のアナログ値はその区間の代表値(ディジタル値)で表わされることを述べた。そのとき、区間の取り方を直線的(等間隔)にするか、非直線的(不等間隔)にするかにより、直線量子化と非直線量子化に分けられる。非直線量子化の例としては、対数量子化がある。

アナログ値 X がディジタル値に変換されるとき、誤差が生じる。これを量子化誤差または量子化雑音という。

いま、区間 $[0, N\Delta x]$ において、アナログ値 X は一様分布し、区間 $\left[i\Delta x - \frac{1}{2}\Delta x, i\Delta x + \frac{1}{2}\Delta x \right)$ の X はディジタル値 i に変換されるとする。このとき、直線量子化による量子化誤差の分散は

$$e = N \int_{-\frac{1}{2}\Delta x}^{\frac{1}{2}\Delta x} \frac{1}{N\Delta x} x^2 dx \quad (8.43)$$

$$= \frac{1}{12} (\Delta x)^2 \quad (8.44)$$

で与えられる。このような例を図 8.4(a) に示す。

いま、もし X の分布が一様ではなく、図 8.4(b) のように偏りがあると、直線量子化よりも、非直線量子化のほうが量子化雑音が少なくなる。図 8.4(b) の例では直線量子化 ① の量子化雑音電力は同図(a)と同じく、 $(1/12)(\Delta x)^2$ となる。それに対して、非直線量子化 ② において実線 } のように量子化した場合、量子化雑音電力は

$$4 \cdot \frac{1}{24} \cdot \frac{1}{12} \cdot 4(\Delta x)^2 + 2 \cdot \frac{5}{12} \cdot \frac{1}{12} \cdot \frac{1}{4} (\Delta x)^2 = \frac{37}{48} \cdot \frac{1}{12} (\Delta x)^2$$

となり、直線量子化よりも 18.75% 小さくなる。なお、点線 } で示すように、高密度領域に接する低密度領域の一部をディジタル値 2 および 5 に量子化すると、量子化雑音はさらに小さくなる。

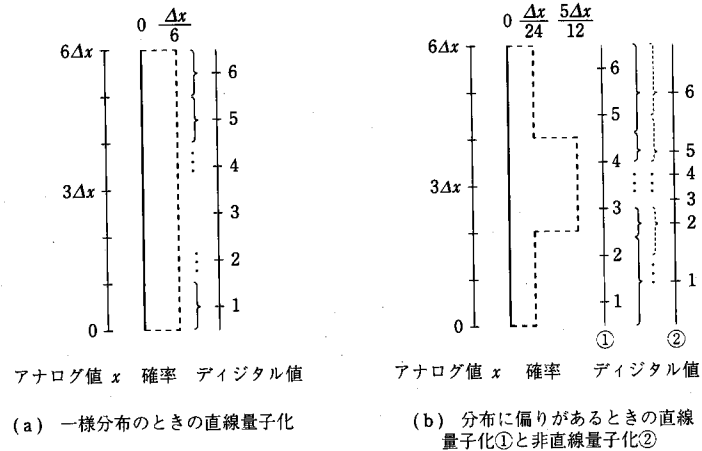


図 8.4 直線量子化(等間隔量子化)と非直線量子化(不均等間隔量子化)

8.6 ベクトル量子化

これまでの量子化では、各アナログ値に対して逐一ひとつのディジタル値を対応させた。これに対して、複数のアナログ値をまとめてディジタル値に変換すると量子化誤差を減少させることができる。この方式をベクトル量子化という。

[例 8.3] 六角形ベクトル量子化

図 8.5 のように、一様分布からの独立な二つの標本値をまとめて量子化することを考える。図 8.5 では代表値は六角格子状に配置されている。それに対していままで述べてきた量子化は図 8.6 に示されるように、代表値が多次元標本値空間では正方格子状に配置されたスカラー量子化で表現される。図 8.6 の格子点間隔は Δx であるとする。これに対して、図 8.5 の標本値空間における格子点(代表点)の密度が図 8.6 に等しいとすると、ある代表点に投影される標本値空間(正六角形)の面積は $(\Delta x)^2$ に等しくなる。いま、格子点間隔を $2d$ とすると、正六角形の 1 辺は $2d/\sqrt{3}$ となる。このような正六角形の面積が $(\Delta x)^2$ であるとする、

$$d = \sqrt{\frac{\sqrt{3}}{6}} \Delta x$$

となる。まず、図 8.7 のような六角形の $(1/12)$ の領域について量子化雑音電力を求める。正六角形全体ではそれを 12 倍すればよい。格子点の総数を N^2 とすると、正

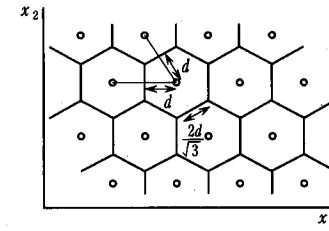


図 8.5 2次元ベクトル量子化

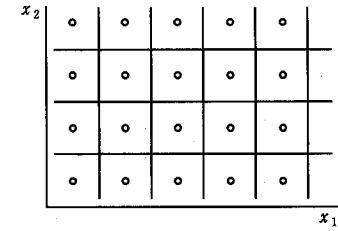


図 8.6 2次元標本値空間で表わしたスカラー量子化

六角形内における標本値の分布密度は $1/(N^2(\Delta x)^2)$ で与えられる。したがって、量子化雑音電力は

$$N^2 \cdot 12 \int_0^d \int_0^{\frac{x}{\sqrt{3}}} (x^2 + y^2) \frac{1}{N^2(\Delta x)^2} dy dx = \frac{5}{18\sqrt{3}} (\Delta x)^2 \quad (8.45)$$

となる。これより、1 標本値当たりの量子化雑音電力は

$$\frac{5}{36\sqrt{3}} (\Delta x)^2 \approx 0.082 (\Delta x)^2$$

であることがわかる。これは、スカラー量子化の場合の量子化雑音電力 $1/12 \approx 0.0833$ に比べて約 3.7% 小さい。

ちなみに、図 8.6 の量子化雑音電力を求めてみると、図 8.8 より、

$$\int_0^{\frac{1}{2}} \int_0^x (x^2 + y^2) \frac{1}{(\Delta x)^2} dy dx = \frac{1}{6} (\Delta x)^2 \quad (8.46)$$

したがって、1 標本値当たりの量子化雑音電力は $(1/12)(\Delta x)^2$ となり、先の結果と一致する。

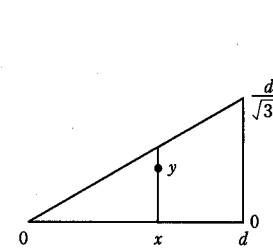


図 8.7 正六角形ベクトル量子化雑音電力

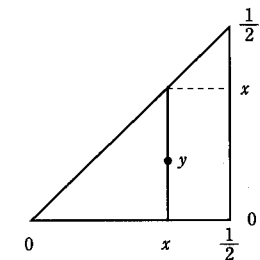


図 8.8 2次元標本値空間におけるスカラー量子化の量子化雑音電力

演習問題 8.

- 8.1 いま、確率変数 X および Y は、それぞれ独立なガウス分布 $\mathcal{N}(0, \sigma_x^2)$ 、および $\mathcal{N}(0, \sigma_y^2)$ に従うものとする。このとき、それらの和である確率変数 $Z = X + Y$ はガウス分布 $\mathcal{N}(0, \sigma_x^2 + \sigma_y^2)$ に従うことを示せ。
- 8.2 (1) 単位帯域幅当たりの相加雑音電力 N_0 [ワット/Hz] をもつ帯域幅 W [Hz] の通信路を通して、信号電力 S [ワット] の信号を送る。このとき、通信路容量 C [ビット/秒] はいくらとなるか。
- (2) 通信路容量 C は W の単調増加関数である。すなわち、広い帯域を使用したほうが通信容量は大きくなる。 $W \rightarrow \infty$ の極限において C はいくらとなるか。
- [ヒント: $h \rightarrow \infty$ のとき、 $(1 + (1/h))^h \rightarrow e$]
- 8.3 連続的通信路において通信路容量 C 、情報伝送速度 R とするとき、 $C > R$ ならば誤り率を任意に小さくできる(そのような符号が存在する)。この符号の電力を S 、単位帯域幅当たりの雑音電力を N_0 、符号の語数を M 、その伝送時間を T とするとき、ビット当たりのエネルギー比には

$$\frac{ST_b}{N_0} > \log_e 2$$

の関係があることを示せ。ただし

$$T_b = \frac{T}{\log_2 M} = 1 \text{ ビット当たりの伝送時間}$$

とする。

$$\left[\begin{array}{l} \text{ヒント: } C = W \log_2 \left(1 + \frac{S}{N} \right), \quad N = WN_0, \quad W = \text{帯域幅} \\ \log_2 \left(1 + \frac{1}{n} \right)^n < \log_2 e \quad (n > 0) \end{array} \right]$$

- 8.4 変数 X の範囲が $-x_0 \leq X \leq x_0$ に制限されているとき、エントロピー $H(X)$ を最大にする X の確率密度関数 $p(x)$ を求めよ。

9

統計的パターン認識

帯域制限を受けた信号波形は、標準化や直交関数展開により、ベクトル表現できることは7章ですでに述べた。また、文字や画像の認識においては、画像から各種の特徴が計算される。したがって、信号や図形・画像の認識を含めて、一般にパターン認識問題においては特徴ベクトルをもとに、パターンの識別・クラス分けが行われる。これらのパターンはしばしば変形を受けたり、雑音で乱されたりする。パターンの変形も特徴ベクトル上では雑音加わったとみなせることが多い。そこで本章では、ガウス雑音加わった標本値が与えられたときに、統計的な見地からパターンの認識・分類の問題を考える。

9.1 多次元ガウス分布

正規分布は多くの連続的確率事象によくあてはまる分布である。複数の標本値が独立でそれらが同じガウス分布に従うときの確率密度関数については8.1節ですでに述べた。これらの場合も含め、一般の n 次元ガウス分布の確率密度関数は次式で与えられる。

$$p(x_1, x_2, \dots, x_n) = \frac{1}{\sqrt{|2\pi\Sigma|}} \exp \left\{ -\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right\} \quad (9.1)$$

ここに、

$$x = (x_1, x_2, \dots, x_n),$$

$$\mu = (\mu_1, \mu_2, \dots, \mu_n); \quad \text{平均値,}$$

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n} \\ \vdots & \vdots & & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn} \end{bmatrix}; \quad \text{共分散行列}(n \times n \text{ 対称行列}),$$

$$\sigma_{ij} = \overline{(x_i - \mu_i)(x_j - \mu_j)}^\dagger,$$

T : 転置,

定常情報源では各標本値の確率密度関数は時刻によらず同一であるから,

$$\begin{aligned} \mu_i &= \mu \quad (i=1, 2, \dots, n) \\ \sigma_{i-l, j-l} &= \sigma_{ij} \quad (l=1, 2, \dots) \end{aligned} \quad (9.2)$$

となる。いちばん簡単な $n=2$ で、かつ平均値が $\mathbf{0}$ の場合についてこれを述べる。この2次元定常ガウス情報源では

$$\mu_1 = \mu_2 = 0$$

$$\sigma_{11} = \sigma_{22} = \sigma^2$$

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} = \begin{bmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{bmatrix}$$

である。このとき

$$\begin{aligned} |2\pi\Sigma| &= |2\pi I\Sigma| \\ &= |2\pi I| |\Sigma| \\ &= (2\pi)^n (\sigma^4 - \rho^2\sigma^4) \\ &= (2\pi)^2 \sigma^4 (1 - \rho^2), \end{aligned} \quad (9.3)$$

$$\Sigma^{-1} = \frac{1}{\sigma^4(1-\rho^2)} \begin{bmatrix} \sigma^2 & -\rho\sigma^2 \\ -\rho\sigma^2 & \sigma^2 \end{bmatrix}, \quad (9.4)$$

$$\begin{aligned} \mathbf{x}\Sigma^{-1}\mathbf{x}^T &= \frac{1}{\sigma^4(1-\rho^2)} [x_1 \ x_2] \begin{bmatrix} \sigma^2 & -\rho\sigma^2 \\ -\rho\sigma^2 & \sigma^2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\ &= \frac{1}{\sigma^2(1-\rho^2)} (x_1^2 - 2\rho x_1 x_2 + x_2^2). \end{aligned} \quad (9.5)$$

したがって,

$$p(x_1, x_2) = \frac{1}{2\pi\sigma^2\sqrt{1-\rho^2}} \exp\left[-\frac{x_1^2 - 2\rho x_1 x_2 + x_2^2}{2\sigma^2(1-\rho^2)}\right]. \quad (9.6)$$

なお,

$$\rho = \frac{\sigma_{12}}{\sigma^2} = \frac{\overline{x_1 x_2}}{\overline{x_1^2}}; \quad \text{相関係数}$$

[†] 本章では表記の簡単化のため、確率変数と単なる変数は同じ記号を用いる。

一般的には

$$\rho = \frac{\sigma_{12}}{\sqrt{\sigma_{11}\sigma_{22}}} = \frac{\overline{(x_1 - \mu_1)(x_2 - \mu_2)}}{\sqrt{\overline{(x_1 - \mu_1)^2} \overline{(x_2 - \mu_2)^2}}}$$

である。

2次元定常ガウス分布を $\rho=0$ および 0.9 の場合につき確率密度関数の等高線で表示すると図9.1のようになる。ガウス分布の場合、 $\rho=0$ は x_1 と x_2 が独立であることを示している。また、 $\rho=0.9$ は x_1 と x_2 が高い相関をもっており、たとえば、 x_1 が大きな値のときには x_2 も大きな値をとることが多いことを示している。

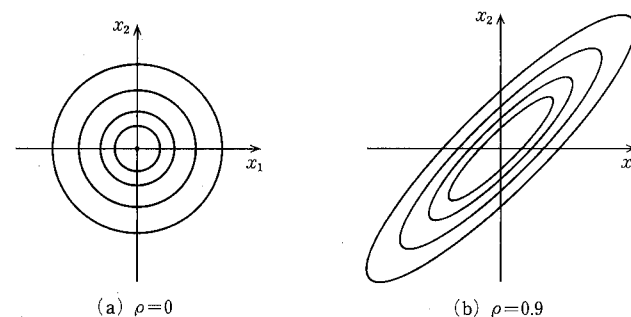


図 9.1 2次元ガウス分布の等高線

9.2 統計的決定法

いま、パターン認識の代表例として、図9.2のような通信系モデルを考える。受信信号は

$$\mathbf{Y} = \mathbf{S}_\theta + \mathbf{N} \quad (9.7)$$

で与えられる。受信機では $\mathbf{Y}$ を観測して $\mathbf{S}_0$ が送られたか、 $\mathbf{S}_1$ が送られたかの識別を行う (θ の推定値 $\hat{\theta}$ を得る)。なお $\mathbf{S}_0$ および $\mathbf{S}_1$ は既知であり、それらの生起確率はそれぞれ $1/2$ であることがわかっているとする。また通信路雑音 $\mathbf{N}$ は平均 $\mathbf{0}$ 、分散 Σ の既知のガウス分布 $\mathcal{N}(\mathbf{0}, \Sigma)$ に従う。これらのベクトルはいずれも n 次元行ベクトル、また行列は $n \times n$ 行列としておく。

$\mathbf{Y}$ を観測したとき、 θ に関する事後確率(条件付き確率) $P(\theta|\mathbf{Y})$ を最大にする決定を行うことは一つの妥当なやり方であろう。以下 $P(\cdot)$ は確率、 $p(\cdot)$ は確率密度関数を表わすものとする。いま、パターンには $\theta=0$ および $\theta=1$

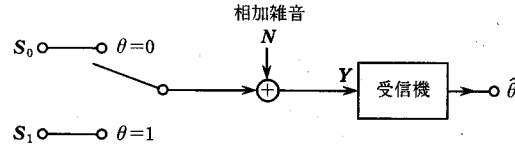


図 9.2 通信系モデル

の二つのクラスしかないので、この決定法に従えば、比

$$\begin{aligned}\lambda(Y) &= \frac{P(\theta=1|Y)}{P(\theta=0|Y)} \\ &= \frac{\frac{P(\theta=1)p(Y|\theta=1)}{p(Y)}}{\frac{P(\theta=0)p(Y|\theta=0)}{p(Y)}} \\ &= \frac{p(Y|\theta=1)}{p(Y|\theta=0)}\end{aligned}\quad (9.8)$$

を求め、

$$\begin{cases} \lambda \geq 1 \text{ ならば } \hat{\theta}=1, \\ \lambda < 1 \text{ ならば } \hat{\theta}=0 \end{cases}\quad (9.9)$$

と決定することになる。式(9.8)の

$$\lambda(Y) = \frac{p(Y|\theta=1)}{p(Y|\theta=0)} \quad (\text{尤度比}) \quad (9.8)'$$

は各クラスを仮定したときの Y の出現確率(密度)の比であり、尤度比と呼ばれる。したがって、式(9.9)に従って識別を行う決定方式は、尤度比を用いて決定を行うことから**尤度比検定**と呼ばれるものになっている。

ここで尤度比 $\lambda(Y)$ を求めてみよう。雑音 N の分布はガウス確率密度関数

$$p(N) = \frac{1}{\sqrt{2\pi\Sigma}} \exp\left\{-\frac{1}{2} N\Sigma^{-1}N^T\right\} \quad (9.10)$$

で与えられる。ここに $|\cdot|$ は行列式、 Σ^{-1} は Σ の逆行列を表わす。 $Y=S_\theta+N$ より、 Y の条件付き確率密度関数は

$$p(Y|\theta) = |2\pi\Sigma|^{-1/2} \exp\left\{-\frac{1}{2} (Y-S_\theta)\Sigma^{-1}(Y-S_\theta)^T\right\}$$

で与えられる。このとき、尤度比は

$$\begin{aligned}\lambda(Y) &= \frac{p(Y|\theta=1)}{p(Y|\theta=0)} \\ &= \exp\left\{-\frac{1}{2} [(Y-S_1)\Sigma^{-1}(Y-S_1)^T - (Y-S_0)\Sigma^{-1}(Y-S_0)^T]\right\}\end{aligned}\quad (9.11)$$

で計算できる。

ここで簡単な場合について式(9.8)の決定規則を考えてみよう。ON-OFF 通信系、すなわち $S_1=S$ および $S_0=0$ を仮定する。また、通信路雑音は無相関(uncorrelated)、すなわち、 $\Sigma=\sigma^2 I$ とする。このとき、式(9.11)は

$$\lambda(Y) = \exp\left[-\frac{1}{2\sigma^2}\{(Y-S)(Y-S)^T - YY^T\}\right] \quad (9.12)$$

となる。したがって、式(9.8)において λ を 1 と比較することは、上の $\exp$ の中身が正か負かを調べることである。すなわち、 $(Y-S)(Y-S)^T$ は Y と S の間の距離の二乗であり、 YY^T は Y と 0 の間の距離の二乗であるから、結局いま観測された Y が S と 0 のどちらにより近いかに観測値の識別を行っていることがわかる。この決定規則を改めて書くと次のようになる。

$$\begin{cases} YY^T \geq (Y-S)(Y-S)^T \text{ ならば } \hat{\theta}=1, \\ YY^T < (Y-S)(Y-S)^T \text{ ならば } \hat{\theta}=0, \end{cases}\quad (9.13)$$

また、別の見方も可能である。式(9.12)の $\exp$ の中は

$$\begin{aligned}-\frac{1}{2}[(Y-S)(Y-S)^T - YY^T] &= -\frac{1}{2}(YY^T - YS^T - SY^T + SS^T - YY^T) \\ &= YS^T - \frac{1}{2} SS^T\end{aligned}\quad (9.14)$$

である。ここで、内積 YS^T および SY^T はスカラー量であり、これらは

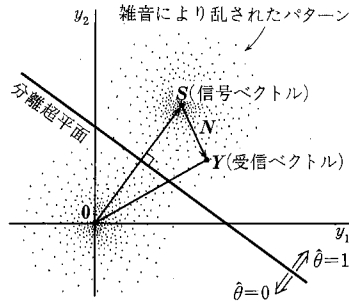
$$YS^T = SY^T = \sum_{i=1}^n y_i s_i \quad (9.15)$$

となって等しいことを用いた。したがって、式(9.14)の正負を調べて識別を行う上の決定規則は

$$\begin{cases} YS^T \geq \frac{1}{2} SS^T \text{ ならば } \hat{\theta}=1, \\ YS^T < \frac{1}{2} SS^T \text{ ならば } \hat{\theta}=0 \end{cases}\quad (9.16)$$

となるので、 Y と S の内積により識別を行っていることがわかる。このような方式は、信号の形をしたフィルタを用いていることから、**整合フィルタ**(matched filter)または**整合機械**(template machine)と呼ばれる。雑音が無相関でない場合には、式(9.16)の左辺は $Y\Sigma^{-1}S^T$ のように雑音の分散で正規化(白色化)した形となる(演習問題9.2)。このような一般形をした方式も整合機械と呼ばれる。

これらを各ベクトルが2次元の場合について図示すれば図9.3のようになる。すなわち、 S と 0 を結ぶ直線を垂直二等分する分離超平面で空間を二等

図 9.3 分離超平面 ($n=2$)

分し、この分離超平面より S に近い側に Y がくれば $\hat{\theta}=1$ と決定し、 0 に近い側に Y がくれば $\hat{\theta}=0$ と決定する。図 9.3 は $n=2$ の場合で直線により受信ベクトル平面が分割されているが、さらにもっと多次元の場合も含めて、この直線を超平面と呼んでいる。なお、一般に Y の線形関数を用いて信号クラスの識別を行なう方式を線形識別方式という。整合機械はこの線形識別方式の代表的なものである。

ここで、線形識別方式の誤り率 P_e を求めてみよう。いま線形識別規則を

$$\begin{cases} YA^T \geq \delta \text{ ならば } \hat{\theta}=1 \\ YA^T < \delta \text{ ならば } \hat{\theta}=0 \end{cases} \quad (9.17)$$

とする。式(9.17)の左辺を z とおく。すなわち

$$z = YA^T \quad (9.18)$$

とおくと、 z はガウス分布をする Y の各要素の荷重和である。したがって、 z もガウス分布に従う。その平均および分散はそれぞれつぎようになる：

$$\begin{aligned} \bar{z} &= \overline{YA^T} \\ &= \bar{Y}A^T \\ &= \theta SA^T, \\ \overline{(z - \bar{z})^2} &= \overline{(NA^T)^2} \\ &= \overline{(AN^T)(AN^T)} \\ &= \overline{AN^TNA^T} \\ &= \sigma^2 AA^T \\ &\triangleq \sigma^2. \end{aligned} \quad (9.19) \quad (9.20)$$

以上より、線形識別関数 $z = YA^T$ は平均 θSA^T 、分散 σ^2 の1次元ガウス分布に従うことがわかる。この様子を図 9.4 に示す。いま、いき値は δ である

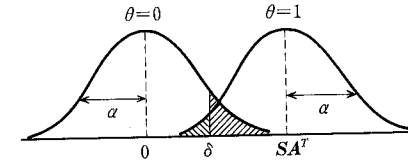


図 9.4 線形識別関数の分布

から図の斜線の部分が誤りとなる。標準ガウス分布関数を

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt \quad (9.21)$$

とすると、これらの誤り率はそれぞれ

$$\begin{aligned} P(\hat{\theta}=1 | \theta=0) &= 1 - \Phi\left(\frac{\delta}{\alpha}\right) \\ P(\hat{\theta}=0 | \theta=1) &= \Phi\left(\frac{\delta - SA^T}{\alpha}\right) \end{aligned} \quad (9.22)$$

となる。また、平均誤り率は

$$P_e = P(\theta=0)P(\hat{\theta}=1 | \theta=0) + P(\theta=1)P(\hat{\theta}=0 | \theta=1) \quad (9.23)$$

で与えられる。

いま、線形識別の重みが式(9.16)で与えられる整合フィルタの場合の誤り率をもう少し具体的に求めてみる。このとき、

$$A = S, \quad \delta = \frac{1}{2} SS^T \quad (9.24)$$

であるから、式(9.22)において

$$\begin{aligned} \frac{\delta}{\alpha} &= \frac{\frac{1}{2} SS^T}{\sqrt{\sigma^2 SS^T}} \\ &= \frac{1}{2} \sqrt{R} \end{aligned} \quad (9.25)$$

となる。また

$$\begin{aligned} \frac{\delta - SA^T}{\alpha} &= \frac{\frac{1}{2} SS^T - SS^T}{\sqrt{\sigma^2 SS^T}} \\ &= -\frac{1}{2} \sqrt{R}. \end{aligned} \quad (9.26)$$

ここに

$$R = \frac{SS^T}{\sigma^2}; \quad \text{S/N 比}$$

これらを用いて式(9.22)は

$$\begin{aligned} P(\hat{\theta}=1|\theta=0) &= 1 - \Phi\left(\frac{1}{2}\sqrt{R}\right) \\ &= \Phi\left(-\frac{1}{2}\sqrt{R}\right), \end{aligned} \quad (9.27)$$

$$P(\hat{\theta}=0|\theta=1) = \Phi\left(-\frac{1}{2}\sqrt{R}\right) \quad (9.28)$$

となる。これら両者は等しいから平均誤り率も

$$P_e = \Phi\left(-\frac{1}{2}\sqrt{R}\right) \quad (9.29)$$

となり、これらは信号のエネルギーと雑音電力の比 R のみで決まることがわかる。なお、標準ガウス分布関数 $\Phi(\cdot)$ の値は付表3に示されている。

これらの誤り率を図示すると図9.5のようになる。

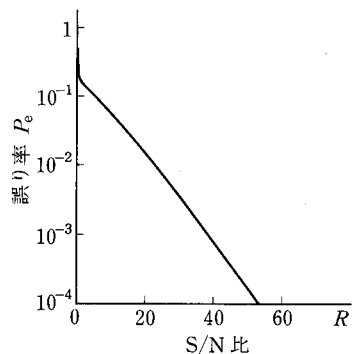


図 9.5 S/N 比と整合機械の誤り率

演習問題 9.

9.1 x_1, x_2 の結合確率密度関数を

$$p(x_1, x_2) = \frac{1}{2\pi\sigma^2\sqrt{1-\rho^2}} \exp\left\{-\frac{x_1^2 - 2\rho x_1 x_2 + x_2^2}{2\sigma^2(1-\rho^2)}\right\}$$

とする。このとき、 $x_1 x_2$ および x_1^2 の期待値をそれぞれ求めよ。

9.2 (1) 雑音の共分散行列が一般の Σ のとき、式(9.16)の決定規則はどのようなになるか。

(2) また、

$$S = [1 \ 2], \quad \Sigma = \begin{bmatrix} \frac{1}{3} & 0 \\ 0 & 1 \end{bmatrix}$$

のとき、分離超平面はどのようなになるか、図示せよ。

(3) このときの誤り率を求めよ。

10

命題論理

人間の知的活動の主要なものは、既知の事実から必要な推論を行い、与えられた問題の解答を得たり、判断を行うことである。

計算機にこのようなことを行わせるためには、推論過程の定式化と理論的な考察が必要となる。本章ではその基本となる命題論理について述べる。

命題論理では真偽を与えることのできる命題(基本論理式)をもとに、論理記号を用いてそれらの複合命題(論理式)をつくることができる。これらの命題相互間の真偽の関係を論じるのが命題論理である。これらの命題はその構成法から人工的な言語とみなすことができる。

10.1 論理式

命題論理が対象とする命題は、たとえば次のような文である。

P (今日は風が強い)

Q (今日は雨だ)

R (1たす1は3である)

これらの命題は論理式の基本となるので基本論理式または基本命題と呼ぶ。基本論理式はこれ以上分解することのできない論理式である。基本論理式は現在対象とする世界において真(T)または偽(F)のいずれかの値をとる。それを真理値と呼ぶ。

表 10.1 に示す論理記号を用いてこれらの基本論理式より論理式をつくることができる。なお、次章で述べる述語論理と区別する必要がある場合にはこれ

表 10.1 論理記号

名 称	英 語 名	記号表現
否 定	not	$\sim$
論理和	or	$\vee$
論理積	and	$\wedge$
含 意	implies	$\rightarrow$
同 値	iff	$\leftrightarrow$

らの論理式を**命題論理式**ということもある。たとえば、

$\sim P$ (今日は風が強くない),

$P \vee Q$ (今日は風が強いか, または今日は雨だ),

$P \wedge Q$ (今日は風が強く, かつ今日は雨だ),

$P \rightarrow Q$ (今日は風が強いならば, 今日は雨だ),

$P \leftrightarrow Q$ (今日は風が強いならば, かつそのときに限り今日は雨だ)。

これらの論理式はまた基本論理式の真偽に応じて真偽の値をとる。

基本論理式の代りに別の論理式を入れても論理式ができる。そこで、基本論理式を**論理変項**ともいう。これらの基本的な論理式の真理値表を表 10.2 に示す。たとえば、 $P \vee Q$ が真となるのは P も Q もともに真のときだけである。また、含意 $P \rightarrow Q$ において、 P が真の場合は Q の真偽に応じて論理式 $P \rightarrow Q$ はそれぞれ真偽をとる。一方、 P が偽の場合には Q の真偽によらず $P \rightarrow Q$ は真となる。たとえば

$$P \rightarrow Q \begin{cases} \text{(犬が西を向いているならば, 尾は東を向いている)} \\ \text{(1 たす 1 が 3 であるならば, 今日は雨だ)} \end{cases}$$

を考えると、特に下段の例で P が偽ならば論理式 $P \rightarrow Q$ は Q の真偽によらず真となる。

表 10.2 真理値表

P	Q	$\sim P$	$P \vee Q$	$P \wedge Q$	$P \rightarrow Q$	$P \leftrightarrow Q$
F	F	T	F	F	T	T
F	T	T	T	F	T	F
T	F	F	T	F	F	F
T	T	F	T	T	T	T

10.2 論理式の構成法

10.1 節で述べたような論理式の一般的な構成法を以下に示そう。

- (1) 基本論理式は論理式である。
- (2) 任意の論理式を P, Q とすると、 $\sim P, (P \vee Q), (P \wedge Q), (P \rightarrow Q), (P \leftrightarrow Q)$ も論理式である。括弧は省略してもよい。
- (3) 以上の規則を有限回適用して得られたものだけが論理式である。■

以上の構成規則より、 $\sim P, (P \wedge Q), (P \rightarrow Q), ((P \wedge Q) \rightarrow R), (\sim P \rightarrow (P \wedge Q)), (\sim (P \vee Q) \rightarrow (P \wedge Q))$ などは論理式であり、一方、 $P \wedge \wedge Q, (P \rightarrow \vee Q)$, などは論理式でないことがわかる。このように、記号の集合と構成規則から**論理式(wff; well-formed formula)**の集合 W が生成される。このような構成法は句構造文法による文の生成と同じ様式であり、 W を**命題論理言語**と呼ぶ。

() が省略されたときには、論理記号 $\sim, \wedge, \vee, \rightarrow$ の間の

$$\boxed{\sim} > \boxed{\wedge} > \boxed{\vee} > \boxed{\rightarrow}$$

なる優先順位に従って構成されているものとする。

なお、後(10.5 節)で示すように、

$$P \wedge Q = \sim (P \rightarrow \sim Q)$$

$$P \vee Q = \sim P \rightarrow Q$$

$$P \leftrightarrow Q = (P \rightarrow Q) \wedge (Q \rightarrow P)$$

なる関係があるので、論理記号 $\{\wedge, \vee, \leftrightarrow\}$ を使わずに $\{\sim, \rightarrow\}$ だけ使って論理式を定義することも可能である。このような意味で論理記号の組 $\{\sim, \rightarrow\}$ は関数的に完全であるといわれる。

10.3 論理式の解釈

本書では命題論理式の真偽をそれに含まれる基本論理式の真偽から求める。この手法はモデル理論と呼ばれるものである。これに対して公理と推論規則から論理式の性質を論じる方法もあるが、両者は等価であり後者については付録に示すに止める。

10.3.1 解釈とモデル

論理式は、各論理変項の真理値(真偽の値; **F** または **T**)に応じてその真理値が定まる。基本的な論理式の真理値の決め方は 10.1 節に記した。もっと複雑な論理式の真理値も同様に求めることができる。

定義 10.1 命題論理式の解釈 I とは、世界を一つ定めて、それに応じて各論理変項に真理値を与えることである。

P を論理式とする。このとき、解釈 I のもとでの P の真理値を $(P)_I$ で表わすことにする。 I および $()$ は省略することもある。

【例 10.1】 $(P)_I = \mathbf{F}$, $(Q)_I = \mathbf{T}$, $(R)_I = \mathbf{T}$ のとき、論理式

$$S \triangleq ((P \wedge Q) \rightarrow R) \quad (10.1)$$

を評価してみよう。表 10.2 より、 $(P \wedge Q)_I = \mathbf{F}$ となるから、式(10.1)は $(S)_I = \mathbf{T}$ と評価される。

定義 10.2 論理式 α が解釈 I のもとで真と評価されるとき、解釈 I は α を充足するという。これを

$$\models_I \alpha \quad (10.2)$$

と書く。このとき解釈 I は α のモデルであるという。また、 α がモデルをもてば、 α は充足可能であるという。また、 α がモデルをもたないときは、 α は充足不可能であるという。

【例 10.2】 $P \rightarrow Q$ の一つのモデルは $P = \mathbf{F}$, $Q = \mathbf{F}$ である。このモデル(解釈)を I とすると

$$\models_I P \rightarrow Q \quad (10.3)$$

これは、 $P \rightarrow Q$ が解釈 I のもとで真となること、あるいは I は $P \rightarrow Q$ が真となりうる解釈であることを示している。

別のモデルは $P = \mathbf{T}$, $Q = \mathbf{T}$ および $P = \mathbf{F}$, $Q = \mathbf{T}$ である。 $P \rightarrow Q$ のモデルはこの三つである。

【例 10.3】 $P \wedge Q$ は $(P)_I = \mathbf{T}$, $(Q)_I = \mathbf{T}$ なる解釈 I により充足されるので、充足可能である。また、 $P \wedge \sim P$ はどのような解釈のもとでも偽と評価されるので充足不可能である。

定義 10.3 論理式 α が任意の解釈 I で常に充足されるとき、 α は恒真であるといい、

$$\models \alpha \quad (10.4)$$

と書く。

10.3 論理式の解釈

単に $\alpha \rightarrow \beta$ と書いた場合、これが真なのかどうかについては述べていない。解釈 I のもとでこれが真となることを示すためには

$$\models_I \alpha \rightarrow \beta \quad (10.5)$$

と書く。また、常に真であることを示すためには

$$\models \alpha \rightarrow \beta \quad (10.6)$$

と書くのである。

【例 10.4】 P の真理値が **F** でも **T** でも $(P \vee \sim P)$ は **T** となる。したがって

$$\models (P \vee \sim P) \quad (10.7)$$

である。

【例 10.5】 表 10.2 より任意の論理式 β に対して

$$\models (\mathbf{F} \rightarrow \beta) \quad (10.8)$$

である。

【例 10.6】 P の真理値が **F** でも **T** でも $(P \wedge \sim P)$ の真理値は **F** である。したがって、

$$\models (P \wedge \sim P) \rightarrow \beta \quad (10.9)$$

である。

10.3.2 論理的帰結

定義 10.4 Γ を論理式または論理式の集合とし、 α を論理式とする。 Γ 内のすべての論理式を充足する任意の解釈 I がまた α を充足するとき、 α は Γ の論理的帰結である、または α は Γ に論理的に従属するといい、これを

$$\Gamma \models \alpha \quad (10.10)$$

と書く。このときの Γ のモデルと α のモデルの包含関係を図 10.1 に示す。

【例 10.7】 $\Gamma = \{P \wedge Q, (P \vee Q) \wedge R\}$, $\alpha = P \wedge Q \wedge R$, $\beta = P \wedge Q$ とすると $\Gamma \models \alpha$, $\Gamma \models \beta$ である。なぜなら、表 10.3 に示すように Γ をすべて充足するモデル(P, Q, R がすべて **T**)に対して α が真となり、また β も真となるからである。

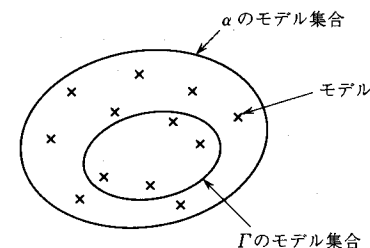


図 10.1 論理的帰結 $\Gamma \models \alpha$. $\times$ 印は個々のモデルを示す。

表 10.3 真理値表

P	Q	R	$P \wedge Q$	$(P \vee Q) \wedge R$	Γ をすべて充足するか?	$\alpha = P \wedge Q \wedge R$	$\beta = P \wedge Q$
F	F	F	F	F	×	F	F
F	F	T	F	F	×	F	F
F	T	F	F	F	×	F	F
F	T	T	F	T	×	F	F
T	F	F	F	F	×	F	F
T	F	T	F	T	×	F	F
T	T	F	T	F	×	F	T
T	T	T	T	T	○	T	T

定義 10.5 二つの論理式 α および β について、 $\alpha \models \beta$ かつ $\beta \models \alpha$ のとき、 α と β は同値であるといい

$$\alpha \models \beta, \text{ または単に } \alpha = \beta \quad (10.11)$$

と記す。

$\alpha = \beta$ のとき、 α と β のモデルは等しく、したがって、 $\alpha \leftrightarrow \beta$ は恒真となる。すなわち

$$\models \alpha \leftrightarrow \beta. \quad (10.12)$$

逆に $\alpha \leftrightarrow \beta$ が恒真ならば $\alpha = \beta$ である。

【例 10.8】 $(P \rightarrow Q)$ および $(\sim P \vee Q)$ においては、あらゆる解釈 I (P および Q が F, T をとる 4 通り) につき真理値を調べることにより、 $(P \rightarrow Q) \models (\sim P \vee Q)$ および $(\sim P \vee Q) \models (P \rightarrow Q)$ であることがわかる。したがって、 $(P \rightarrow Q) = (\sim P \vee Q)$ である。

論理的帰結と含意には次のような関係がある。

定理 10.1 α が $\Gamma = \{P_1, P_2, \dots, P_n\}$ の論理的帰結ならば、すなわち

$$\{P_1, P_2, \dots, P_n\} \models \alpha \quad (10.13)$$

ならば

$$P_1 \wedge P_2 \wedge \dots \wedge P_n \rightarrow \alpha \quad (10.14)$$

は恒真である。すなわち、

$$\models P_1 \wedge P_2 \wedge \dots \wedge P_n \rightarrow \alpha \quad (10.15)$$

また、その逆も成り立つ。

証明 まず、 $\Gamma \models \alpha$ ならば式(10.14)が恒真であることを示す。 Γ の論理式 $P_1, P_2, \dots, P_n$ をすべて真とする任意の解釈 I について論理的帰結の定義より α は真となる ($T \rightarrow T$)。また、 $P_1, P_2, \dots, P_n$ のうち少なくとも一つを偽とする解釈 I について $P_1 \wedge P_2 \wedge \dots \wedge P_n$ は偽となるから、そのとき式(10.14)は真となる ($F \rightarrow \alpha$)。したがって、表 10.2 より任意の解釈 I について式(10.14)は真となり、恒真である。逆に、式(10.14)が恒真であるならば Γ 内のすべての論理式を充足する任意の解釈 I について α は真でなければならない。すなわち、これは α が Γ の論理的帰結であることを示している。

【例 10.9】

$$\{(P \vee Q), R\} \models R \quad (10.16)$$

であるから、

$$\models (P \vee Q) \wedge R \rightarrow R \quad (10.17)$$

すなわち、式(10.16)の左辺 $(P \vee Q), R$ をともに T とする任意の解釈はまた R を T とするから、 $(P \vee Q) \wedge R \rightarrow R$ は恒真である。すなわち、式(10.17)が成り立つ。

10.3.3 モデルとベン図

命題論理言語 W のすべての解釈の集合を U 、論理式 α のすべてのモデルの集合を M_α と記すことにする。このとき、結合子で結ばれた論理式のモデルは表 10.2 より、つぎのようになることが容易に確かめられる。

- (1) $M_{\sim \alpha} = \sim M_\alpha$
- (2) $M_{\alpha \wedge \beta} = M_\alpha \cap M_\beta$
- (3) $M_{\alpha \vee \beta} = M_\alpha \cup M_\beta$
- (4) $M_{\alpha \rightarrow \beta} = \sim M_\alpha \cup M_\beta$
- (5) $M_{\alpha \leftrightarrow \beta} = (\sim M_\alpha \cup M_\beta) \cap (M_\alpha \cup \sim M_\beta)$

これをベン図で示すと、図 10.2 のようになる。

10.3.4 三段論法

$$\{(P \rightarrow Q) \wedge (Q \rightarrow R)\} \rightarrow (P \rightarrow R) \quad (10.18)$$

の真理値表は表 10.4 のようになる。

したがって、式(10.18)は恒真であることがわかる。すなわち、

$$\models \{(P \rightarrow Q) \wedge (Q \rightarrow R)\} \rightarrow (P \rightarrow R) \quad (10.19)$$

である。また、定理 10.1 より

$$\{P \rightarrow Q, Q \rightarrow R\} \models P \rightarrow R \quad (10.20)$$

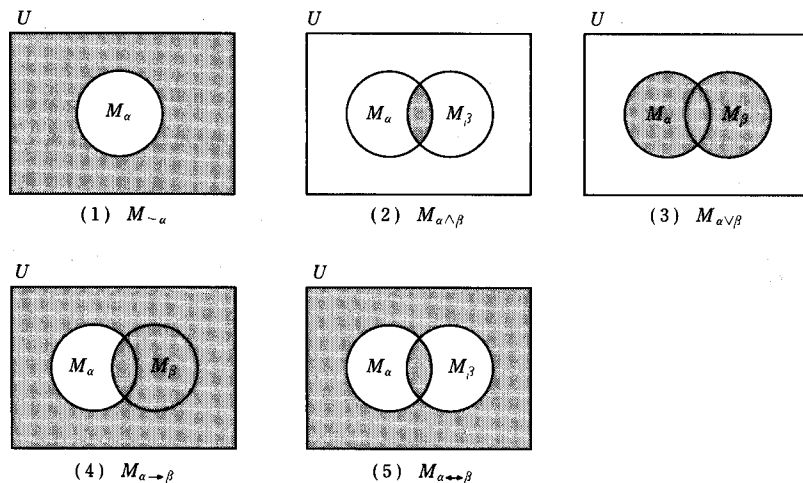


図 10.2 モデルとベン図

表 10.4 三段論法の真理値表

P	Q	R	$P \rightarrow Q$	$Q \rightarrow R$	$P \rightarrow R$	$\{(P \rightarrow Q) \wedge (Q \rightarrow R)\} \rightarrow (P \rightarrow R)$
F	F	F	T	T	T	T
F	F	T	T	T	T	T
F	T	F	T	F	T	T
F	T	T	T	T	T	T
T	F	F	F	T	F	T
T	F	T	F	T	T	T
T	T	F	T	F	F	T
T	T	T	T	T	T	T

である。すなわち、 $P \rightarrow R$ は $\{P \rightarrow Q, Q \rightarrow R\}$ の論理的帰結であるということが
ができる。

[例 10.10] 三段論法

- 前提 1 切符を忘れたならば(P)、飛行機に乗れない(Q).
 前提 2 飛行機に乗れなければ(Q)、アメリカへ行けない(R).
 結 論 切符を忘れたならば(P)、アメリカへ行けない(R).

10.4 意味の木

これまでは命題論理式の充足性やモデルを調べるため、真理値表を用いてきた。命題論理式には有限個の論理変項しか含まれていないので、それらの真偽の組合せは有限である。しかしながら、命題論理式の恒真性(または恒偽性)を調べるとなると、 n 論理変項に対して 2^n 個の行をもつ表を調べることになる、たいへんである。

真理値表を用いなくて論理式の恒真性を調べる方法の一つに、意味の木による方法がある。これは、命題論理式を各論理変項 x_i ($i=1, 2, \dots, n$) のパラメータ表現、すなわち

$$P(x_1, x_2, \dots, x_i, \dots, x_n)$$

としたとき、これは

$$\begin{aligned} &P(x_1, x_2, \dots, x_i, \dots, x_n) \\ &= P(x_1, x_2, \dots, \mathbf{T}, \dots, x_n) \wedge x_i \vee P(x_1, x_2, \dots, \mathbf{F}, \dots, x_n) \wedge \sim x_i \end{aligned} \quad (10.21)$$

と展開されることを利用する。式(10.21)は、 x_i が $\mathbf{T}$ のとき $P(x_1, x_2, \dots, x_i, \dots, x_n)$ は $P(x_1, x_2, \dots, \mathbf{T}, \dots, x_n)$ の真理値に等しく、 x_i が $\mathbf{F}$ のときには $P(x_1, x_2, \dots, \mathbf{F}, \dots, x_n)$ の真理値に等しくなることを表わしている。

式(10.21)の展開を木で表わすと、図 10.3 のようになる。展開されるべきノード(上の○印)から二本の枝が出ている。左の枝についている x_i は $x_i = \mathbf{T}$ を表わしている。また、右の枝の $\sim x_i$ は $x_i = \mathbf{F}$ を表わしている。枝の先にはそれらの真理値が代入された葉ノードがつく。展開された葉ノードでまだ残っている論理変項についてさらに展開を進めていくと、この木は有限の大きさで終了する。このとき、末端の各葉ノードの命題論理式で各論理変項にはすべて $\mathbf{F}$ または $\mathbf{T}$ の真理値が与えられているので、命題論理式の真理値は確定している。

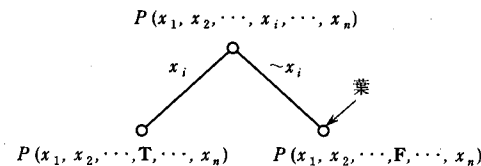


図 10.3 意味の木

【例 10.11】 命題論理式 $\sim(\alpha \vee \beta \vee \sim \gamma) \wedge (\alpha \vee \sim \beta) \wedge \sim \gamma$ を意味の木により展開すると、図 10.4 のようになる。葉ノードはすべて **F** となったので、この命題論理式は恒偽式であることがわかる。ただし、意味の木の形はどの論理変項を用いて展開していくかにより異なってくる。

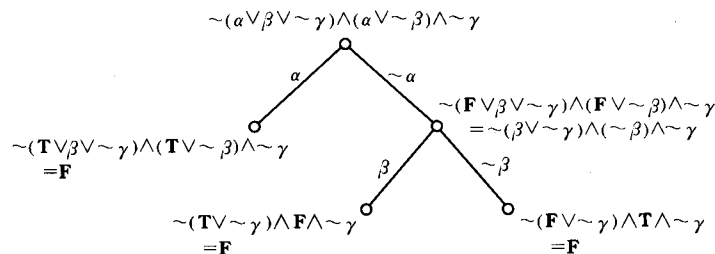


図 10.4 意味の木による命題論理式の展開

意味の木をつくらずに、論理式を一定のアルゴリズムに従って機械的に変形・簡単化していき、論理式の恒真性・恒偽性を調べる方法にデイビス・パットナム (Davis-Putnam) の方法がある。この方法は論理式の連言標準形から出発してその恒偽性(与えられた論理式の否定から出発すれば恒真性)を調べるものである。述語論理の場合と異なり、命題論理ではこのような方法を用いるまでもなく、次節の表 10.5 などの同値な関係を使うなどして簡単化を行うことによりその恒偽性を調べることは容易である。したがって、本書ではデイビス・パットナムの方法は付録に示すに止めておく。

10.5 命題論理式の標準形

命題論理式はさまざまな形に書くことができるが、デイビス・パットナムのアルゴリズム(付録 A 1)や意味の木を用いて論理式の恒偽性を導く場合には、その標準形を定めておくことと便利である。標準形には連言標準形(conjunctive normal form)および選言標準形(disjunctive normal form)がある。

基本論理式を p とするとき、 p および $\sim p$ はリテラル(literal)と呼ばれる。いま P_i をリテラルの選言からなるとする、すなわち p_{i1} をリテラルとして

$$P_i = p_{i1} \vee p_{i2} \vee \cdots \vee p_{im}$$

なるとき、 P_i は節(clause)と呼ばれる。このとき、節の連言からなる論理式

$$P = P_1 \wedge P_2 \wedge \cdots \wedge P_n$$

表 10.5 よく使われる同値な関係。■は恒真(T), □は恒偽(F)

(1) 巾等則	(a) $P \wedge P = P$	(b) $P \vee P = P$
(2) 可換則	(a) $P \wedge Q = Q \wedge P$	(b) $P \vee Q = Q \vee P$
(3) 結合則	(a) $(P \wedge Q) \wedge R = P \wedge (Q \wedge R)$	(b) $(P \vee Q) \vee R = P \vee (Q \vee R)$
(4) 分配則	(a) $P \wedge (Q \vee R) = (P \wedge Q) \vee (P \wedge R)$	(b) $P \vee (Q \wedge R) = (P \vee Q) \wedge (P \vee R)$
(5) 吸収則	(a) $P \wedge (P \vee Q) = P$	(b) $P \vee (P \wedge Q) = P$
(6) 排中則	(a) $P \wedge \sim P = \square$	(b) $P \vee \sim P = \blacksquare$
(7) 二重否定	$\sim(\sim P) = P$	
(8) ド・モルガンの法則	(a) $\sim(P \wedge Q) = \sim P \vee \sim Q$	(b) $\sim(P \vee Q) = \sim P \wedge \sim Q$
(9) その他	$(P \rightarrow Q) = (\sim Q \rightarrow \sim P) = (\sim P \vee Q)$ $(P \leftrightarrow Q) = (P \rightarrow Q) \wedge (Q \rightarrow P)$ $((P \rightarrow Q) \rightarrow R) = ((P \wedge Q) \rightarrow R)$ $P \wedge \blacksquare = P, \quad P \vee \blacksquare = \blacksquare, \quad P \wedge \square = \square, \quad P \vee \square = P$	

は連言標準形と呼ばれる。同様に q_{ij} をリテラルとしたとき

$$Q = Q_1 \vee Q_2 \vee \cdots \vee Q_n$$

ここに

$$Q_i = q_{i1} \wedge q_{i2} \wedge \cdots \wedge q_{im}$$

なる論理式 Q は選言標準形と呼ばれる。

【標準形への変換】

- (1) 論理記号 $\rightarrow$ および $\leftrightarrow$ はつぎのようにして取り除くことができる。

$$P \leftrightarrow Q = (P \rightarrow Q) \wedge (Q \rightarrow P)$$

$$P \rightarrow Q = \sim P \vee Q$$

- (2) 二重否定およびド・モルガン (de Morgan) の法則を用いてつぎのように否定記号を各基本論理式の直前にもってくる。

$$\sim(\sim P) = P \quad (\text{二重否定})$$

$$\left. \begin{aligned} \sim(P \vee Q) &= \sim P \wedge \sim Q \\ \sim(P \wedge Q) &= \sim P \vee \sim Q \end{aligned} \right\} \quad (\text{ド・モルガンの法則})$$

- (3) つぎの分配則を用いて、連言標準形、選言標準形をつくる。

$$P \vee (Q \wedge R) = (P \vee Q) \wedge (P \vee R)$$

$$P \wedge (Q \vee R) = (P \wedge Q) \vee (P \wedge R)$$

以上のプロセスや途中で表 10.5 の同値な関係を適宜用いて変形を行う。

演習問題 10.

- 10.1 論理式 $W = (P \rightarrow Q) \rightarrow R$ のモデルを求めよ。
- 10.2 つぎの各論理式の恒真性または恒偽性を意味の木を用いて求めよ。
- (1) $P \rightarrow (Q \rightarrow P)$
 - (2) $(P \rightarrow Q) \rightarrow (\sim Q \rightarrow \sim P)$
 - (3) $(P \rightarrow (Q \rightarrow R)) \rightarrow ((P \rightarrow Q) \rightarrow (P \rightarrow R))$
 - (4) $(P \vee Q \vee R) \wedge (P \vee \sim Q) \wedge \sim P \wedge \sim R$
 - (5) $(P \rightarrow Q) \wedge (Q \rightarrow R) \rightarrow (P \rightarrow R)$ (三段論法)
- 10.3 論理式 $W_1 = (P \rightarrow Q) \vee (Q \wedge R)$, および $W_2 = (P \leftrightarrow Q) \vee (\sim Q \wedge P \wedge R)$ を連言標準形にせよ。

11

述語論理

命題論理より強力な論理体系が述語論理である。命題論理では完全に閉じた文のみを扱った。その文の中で何が述べられているか関知せず、文の真偽のみに関心があった。述語論理は命題論理を拡張したものであり、文が述べている対象の間の関係まで取り扱うことができる。

一変数述語の例を示すと

$\text{man}(\text{socrates})$: ソクラテスは人である。

また、述語(基本論理式)と論理記号から合成される論理式(wff)の例は

$\text{man}(x) \rightarrow \text{die}(x)$: x が人ならば x は死ぬ。

である。命題論理と同じく、これらの述語や論理式は真または偽の値をとる。

11.1 論理式の構成法

まず、項(term)について述べる。項は、上の例でいうと socrates や x であり、述語が述べる対象である。これには、 socrates のように特定の対象を指す個体定項と、 x のようにそれが変化しうる個体変項がある。前者を $a, b, c, \dots$, 後者を $x, y, \dots$ で表わす。これらは定義域 D の要素である。項を $t_1, t_2, \dots, t_n$ とするとき、 n 変数関数 $f(t_1, t_2, \dots, t_n)$ も項であり、これは D の要素の n 組を D の一要素に写像する。述語中の項は引数ということもある。

いま、

$$f(x, y) = x + y$$

とする。 D を自然数の集合 $\{1, 2, \dots\}$ とすると、たとえば、 $x=1, y=2$ のとき

$f(1, 2) = 3$ となる。このとき、 $x, y, f(x, y), 1, 2, \dots, f(1, 2)$ などは項である。

項を $t_1, t_2, \dots, t_n$ 、また n 変数述語記号を P とするとき、述語 $P(t_1, t_2, \dots, t_n)$ は真または偽の値(真理値)をとる。たとえば、 $P(t_1, t_2)$ を「 t_1 は t_2 より小さいかまたは等しい」という述語とし、 $f(x, y) = x + y$ とすると、 $P(1, 2)$ 、 $P(f(2, 3), 10)$ は真、 $P(5, 4)$ は偽となる。

述語を基本論理式とする。基本論理式と論理記号から論理式が合成される。その真理性は命題論理の場合と同様にして求めることができる。ただし、述語論理では特に論理記号として、項の値の限量を行う $\forall$ および $\exists$ が用いられる。論理記号 $\forall$ を全称記号、 $\exists$ を存在記号と呼び、両者を限量記号または限定作用素と呼ぶ。 $\forall x \dots$ は「すべての x について…」、また、 $\exists y \dots$ は「…となるような y が存在する」という意味を表わす。

ここで述語論理式の定義を行う。

- (1) $t_1, t_2, \dots, t_n$ が項であり、 n 変数述語記号を P とするとき、 $P(t_1, t_2, \dots, t_n)$ は基本論理式である。
- (2) P, Q が論理式であれば、 $P \wedge Q, P \vee Q, \sim P, P \rightarrow Q, P \leftrightarrow Q$ も論理式である。
- (3) P が論理式であり x が個体変項であるとき、 $\forall x P$ は論理式である。
- (4) P が論理式であり x が個体変項であるとき、 $\exists x P$ は論理式である。

$\forall x, \exists y$ で束縛される変項 x, y を束縛変項(bound variable)という。束縛変項でない変項を自由変項(free variable)という。 $\forall x(\dots), \exists y(\dots)$ のような x, y の束縛の及ぶ範囲を $\forall x, \exists y$ の作用範囲(scope)という。束縛変項はその作用範囲外には何らの影響も及ぼさないから、 $\forall x P(x)$ は $\forall y P(y)$ のように束縛変項を置き換えてもよい。なお、

$$(\exists x)(\sim P(x)) = \sim(\forall x)P(x) \quad (11.1)$$

なる関係がある。

全称記号と存在記号が組み合わされている場合、存在記号はその前にある全称記号に依存する。たとえば、 $P(x, y)$ を「整数 x は y より大きい」という述語としたとき、

$$(\forall x)(\exists y)P(x, y) \quad (11.2)$$

は「すべての整数 x に対して、それより大きな整数 y が存在する」という言

明になる。もしこの論理式が真ならば、 y は明らかに x に依存する。

なお、上のように述語は変数化しない述語論理を第一階述語論理という。これに対して、述語も変数化させる述語論理を高階述語論理という。たとえば

$$(\forall x)(\exists P)(P(x) \rightarrow Q(x)) \quad (11.3)$$

が高階述語論理の例であるが、本書では第一階述語論理のみを対象とし、高階述語論理についてはふれない。

11.2 恒真性と充足性

変項を含まない述語論理式(wff)の真理値は、命題論理式と同様にそれに含まれない基本論理式の真理値から求めることができる。述語論理式が自由変項を含む場合には、それにどのような具体的な個体定項があてはめられるか決まっていないので、その真偽は一般には定まらない。しかしながら、論理式の恒真性を論じる場合には自由変項は全称記号で束縛されていると考えてよい。

述語論理式の場合どのような世界のもとでも(任意の解釈のもとで)、その論理式が常に真となるという“恒真性”が成立するか否かを判定することが必ずできるとは限らない。定義域の要素が無限個のとき、そのすべてについて論理式の値を決定できる一般的な方法を見つけることは不可能であると証明されているからである。したがって、述語論理式の値の決定は決定不能である。これは一般的には決定不能であるということを述べているだけであり、個々の論理式については真偽を決定できる場合がある。たとえば、 $(\forall x)(P(x) \rightarrow P(x))$ は恒真である。

また、ある論理式が恒真であるならば、その恒真性を実際に証明する手順が存在することが証明されている。もし恒真でない論理式にその手順を適用した場合、その手順は決して終了しない。この意味で論理式が恒真か否か一般には判定できないのである。論理式の否定を考えることにより、同様に恒偽性を判定することができる。ある論理式が恒偽のとき、その論理式は充足不可能であるという。

【例 11.1】 (i) 論理式 $Q = \forall x P(x)$ は、 $P(x) \equiv \mathbf{F}$ なる述語に対しては $Q = \mathbf{F}$ であるが、 $P(x) \equiv \mathbf{T}$ なる述語に対しては $Q = \mathbf{T}$ となる。したがって、 Q は充足可能である。ただし、恒真でも恒偽でもない。

(ii) 論理式 $R = \forall x \exists y [P(x) \rightarrow P(y)]$ は、どのような $P(x)$ を与えても、すなわち任意の解釈のもとで真となるから、 R は恒真である。

(iii) 論理式 $S = \forall x[P(x) \rightarrow \sim P(x)]$ は、どのような $P(x)$ を与えても、すなわち任意の解釈のもとで偽となるから、 S は恒偽である。

論理式 Q が論理式の集合 P に論理的に従属しているとは、 P を充足するあらゆる解釈がまた Q をも充足することである。このとき、 Q は P の論理的帰結であるともいわれる。 Q が P に論理的に従属している場合、 P を充足するあらゆる解釈は $\{\sim Q\}$ を充足しないことに注意しよう。したがって、集合 $P \cap \{\sim Q\}$ は充足不可能となる。逆に $P \cap \{\sim Q\}$ が充足不可能であれば、 Q は P に従属することがいえる。このことを利用すれば定理証明を行うことができる。これを効率よく行う方法が導出原理(resolution principle)である。これについて述べるため、いくつかの準備をしておく。

11.3 節 形

基本論理式またはその否定をリテラル(literal)と呼ぶ。リテラルの論理和(選言)のみで構成された論理式を節(clause)と呼ぶ。本節では、与えられた任意の論理式を節の集合に変換する手順について述べる。

(1) 含意記号の除去

含意 $A \rightarrow B$ は $\sim A \vee B$ を代入することにより除去できる。たとえば、

$$\forall x[\forall y(A(f(x, y)) \wedge B(x)) \rightarrow \forall yC(y)] \quad (11.4)$$

は

$$\forall x[\sim \forall y(A(f(x, y)) \wedge B(x)) \vee \forall yC(y)] \quad (11.5)$$

となる。

(2) 否定記号の個別化

否定記号 $\sim$ はただ一つの述語記号のみに作用するように、以下の置換・代入を行う。

$\sim(A \wedge B)$ は $\sim A \vee \sim B$ に置き換える

$\sim(A \vee B)$ は $\sim A \wedge \sim B$ に置き換える

$\sim \sim A$ は A に置き換える

$\sim(\forall x)A$ は $(\exists x)\sim A$ に置き換える

$\sim(\exists x)A$ は $(\forall x)(\sim A)$ に置き換える

これらの置換により、上の式(11.5)は

$$\forall x[(\exists y)(\sim A(f(x, y)) \vee \sim B(x)) \vee \forall yC(y)] \quad (11.6)$$

となる。

(3) 変数記号の個別化

束縛変数は限量記号の作用範囲外には何らの影響を及ぼさないから、それらは別物と考えてよい。これを明確化し、後での混乱を避けるため、それらに別々の変数名をつける。たとえば、式(11.6)はつぎのようになる：

$$\forall x[(\exists y)(\sim A(f(x, y)) \vee \sim B(x)) \vee \forall zC(z)] \quad (11.7)$$

(4) 存在記号の除去

$\forall x \exists y A(x, y)$ なる論理式を考える。存在記号 $\exists y$ は全称記号 $\forall x$ の作用範囲内にあり、各 x に対して $A(x, y)$ を真とする y が存在すると解釈できる。この y は x に依存するからそれを陽に示し、存在記号は除去する。依存関係を示す新しくつくられた関数 $y = g(x)$ はスコーム関数(Skolem function)と呼ばれる。これを用いると、 $\forall x \exists y A(x, y)$ は $\forall x A(x, g(x))$ と表記できる。式(11.7)は

$$\forall x[(\sim A(f(x, g(x))) \vee \sim B(x)) \vee \forall zC(z)] \quad (11.8)$$

となる。

全称記号に支配されない存在記号を含む場合、たとえば $\exists x P(x)$ は $P(a)$ が真となるある個体定項 a を用いて $P(a)$ と表記される。

(5) 前置形

この段階では存在記号はなく、変項はすべて異なった名称で全称化されている。したがって、全称記号をすべていちばん前にもってきて、その作用範囲をそれに続く論理式全体に及ぶとしてもよい。式(11.8)の例でこれを行うと

$$\forall x \forall z[(\sim A(f(x, g(x))) \vee \sim B(x)) \vee C(z)] \quad (11.9)$$

となる。

(6) 積標準形

論理式から前置全称記号を除いた部分を積標準形(連言標準形)に変換する。これには

$$A \vee (B \wedge C) \text{ を } (A \vee B) \wedge (A \vee C) \text{ と置き換える}$$

という規則を用いる。

たとえば、 $A(x, y) \wedge B(x) \vee C(x, y) \wedge D(x) \wedge E(y)$ は以下のように変形される。

$$\begin{aligned} & A \wedge B \vee C \wedge D \wedge E \\ &= ((A \wedge B) \vee C) \wedge ((A \wedge B) \vee (D \wedge E)) \\ &= ((A \wedge B) \vee C) \wedge ((A \wedge B) \vee D) \wedge ((A \wedge B) \vee E) \end{aligned}$$

$$= (A \vee C) \wedge (B \vee C) \wedge (A \vee D) \wedge (B \vee D) \wedge (A \vee E) \wedge (A \vee B) \quad (11.10)$$

また,

$$\{A(x, y) \wedge B(x)\} \vee \{C(x, y) \wedge D(x)\} \vee E(y) \quad (11.11)$$

は以下のように変換されていく。

$$\begin{aligned} &= \{(A \wedge B) \vee C\} \wedge \{(A \wedge B) \vee D\} \vee E \\ &= \{(A \vee C) \wedge (B \vee C)\} \wedge \{(A \vee D) \wedge (B \vee D)\} \vee E \\ &= [\{(A \vee C) \wedge (B \vee C)\} \vee E] \wedge [\{(A \vee D) \wedge (B \vee D)\} \vee E] \\ &= [(A \vee C \vee E) \wedge (B \vee C \vee E)] \wedge [(A \vee D \vee E) \wedge (B \vee D \vee E)] \\ &= (A \vee C \vee E) \wedge (B \vee C \vee E) \wedge (A \vee D \vee E) \wedge (B \vee D \vee E) \quad (11.12) \end{aligned}$$

(7) 全称記号の除去

この段階ではすべての変項は全称化されているので、全称記号を論理式の表現から省く。全称記号がなくても、変項はすべて全称化されているとみなす。

(8) 論理積記号の除去

論理積 $A \wedge B$ は二つの論理式 A, B に置き換える。この置き換えを続けて行くと、論理式の有限集合が得られる。たとえば、式(11.12)からは節の集合

$$\{A \vee C \vee E, B \vee C \vee E, A \vee D \vee E, B \vee D \vee E\} \quad (11.13)$$

が得られる。これらの各要素、たとえば $A \vee C \vee E$ は節である。基本論理式またはその否定がリテラルであり、リテラルを論理和記号で結んだものが節である。

11.4 導出節

述語論理の真偽も、原理的には真理値表によって確かめることができる。しかしながら、述語論理においては引数(項)がさまざまな値をとるため、それらすべての場合について真偽を調べ上げることは一般的には不可能である。したがって、真理値表による方法では恒偽性、恒真性や定理の証明は不可能である。充足不可能性(恒偽性)の証明は、実際にはそのようなことをしなくても、ある集合(エルブラン空間)の要素についてのみ意味の木を調べれば可能であることを付録A3に示す。本節ではさらに効率よく、導出原理に従って論理式の集合から論理的に空節(偽)を導くことにより、恒偽性(ひいては定理証明)の判定を行う方式について述べる。導出原理は代入と三段論法を基本に推論を進めていく方式である。これは論理型言語 Prolog の基礎となっている。

導出原理では与えられた節集合に対して直接論理演算を行い、それらのいずれかの節が空節(**F**)となるか、あるいはそれらの節の組合せからそれらと同値な節集合を導き、その中に**F**となる節がでてくるようにする。その際、基本となる方法が命題論理のところでもふれたデビス・パットナムの方法である。ここではその基本的事項のみを述べるが、詳しくは付録A1を参照されたい。

デビス・パットナムの方法の基本は二つの節 $L \vee A$ および $\sim L \vee B$ から $A \vee B$ を導くことである。これは、 $(L \vee A) = (\sim A \rightarrow L)$ および $(\sim L \vee B) = (L \rightarrow B)$ から、三段論法により $(A \vee B) = (\sim A \rightarrow B)$ を導くことと基本的には同じである。すなわち、

$$\{L \vee A, \sim L \vee B\} \vdash A \vee B. \quad (11.14)$$

これは右の論理式は左の論理式の集合に従属しており、左から右が導かれることを示している。このとき、左辺の二つの節 $L \vee A$ と $\sim L \vee B$ は導出(resolve)するといひ、右辺の節 $A \vee B$ は $\{L \vee A, \sim L \vee B\}$ の導出節(resolvent)と呼ばれる。特殊な場合として、

$$\{L, \sim L\} \vdash \square \quad (11.15)$$

が適用できるときには空節 $\square$ が導かれる。このとき、左辺の論理式集合は偽であることがわかる。

ただし、この導出節を得る方法は、命題論理のときには簡単に行えたが、述語論理においては述語の引数に出てくる項が種々異なった形をしており、それらを一致させる手法(単一化)が必要となってくる。

【例 11.2】 単一化と導出節

(1) 節集合 $\{P(x), \sim P(y)\}$ からは $x=y$ とおくことにより空節 $\square$ が導かれる。

(2) 節集合 $\{P(x) \vee Q(x), \sim P(a)\}$ からは、 $x=a$ とおくことによって、 $Q(a)$ が導かれる。

(3) $\{P(x, f(x)) \vee Q(x), \sim P(a, f(y)), \sim Q(g(z))\}$ からは、 $x=y=a=g(z)$ とおくことにより、

$$\{P(a, f(a)) \vee Q(a), \sim P(a, f(a)), \sim Q(a)\}$$

が得られ、さらに式(11.14)および式(11.15)を適用することにより、導出節として空節が導かれる。

導出のプロセスは三段論法が基本であるが、上の例の、特に(2)において、その際特定の個体定項 a に対して成立している論理式と任意の変項 x に対して成立する論理式とから、単一化により前者のより狭い特定の個体定項 a に対して成立する論理式を得ていることがわかるであろう。

11.5 単一化

本節では、導出を行うための単一化について述べる。

置換は

$$\lambda = \{t_1 \leftarrow v_1, t_2 \leftarrow v_2, \dots, t_n \leftarrow v_n\}$$

で表現される。ここに、 $v_1, v_2, \dots, v_n$ はすべて異なる個体変項であり、 t_i は項である。この置換がリテラル(基本論理式、述語およびその否定)集合に適用されると、各リテラル中の v_i はすべて同時に t_i に置き換えられる。

単一化(unification)とは、同じ述語からなるリテラルの集合において、引数を置換により等しくすることである。このとき、引数の異なっていたリテラルは同一化され、単一要素の集合となることからそう呼ばれる。

たとえば、リテラル集合 $\{P(a, f(x)), P(y, f(x))\}$ は置換 $\lambda = \{a \leftarrow y, b \leftarrow x\}$ により、 $\{P(a, f(b))\}$ となり、単一化される。ただし、単一化するだけなら x を置換する必要はなく、置換 $\lambda' = \{a \leftarrow y\}$ で十分である。このとき、単一化により $\{P(a, f(x))\}$ が得られる。 λ' はこの場合のいちばん単純な単一化のための置換であり、**最も一般的な単一化置換(most general unifier)**と呼ばれる。

なお、リテラル L_i の論理和よりなる節 c はリテラル L_i よりなる集合とみなし、 $\{L_i\}$ または単に $\{L\}$ と表現する。たとえば、

$$\text{節 } c = L_1 \vee L_2 \vee L_3 = \{L_1, L_2, L_3\} = \{L_i \mid i=1, 2, 3\} = \{L_i\} = \{L\}.$$

【例 11.3】 リテラル集合 $\{P(x, y, g(a, b)), P(u, h(v), g(a, w))\}$ に対して、置換 $\rho = \{u \leftarrow x, h(v) \leftarrow y, b \leftarrow w\}$ を行くと、単一化された集合 $\{P(u, h(v), g(a, b))\}$ が得られる。この置換は、最も一般的な単一化置換である。置換 $\rho' = \{x \leftarrow u, h(v) \leftarrow y, b \leftarrow w\}$ も変項名が異なるだけで、同じく最も一般的な単一化置換である。

置換の合成 置換 $\rho = \{s_1 \leftarrow u_1, s_2 \leftarrow u_2, \dots, s_m \leftarrow u_m\}$

に引き続き、

$$\text{置換 } \lambda = \{t_1 \leftarrow v_1, t_2 \leftarrow v_2, \dots, t_n \leftarrow v_n\}$$

を行うときには、合成された置換 $\rho \cdot \lambda$ は

$$\{s_1 \lambda \leftarrow u_1, s_2 \lambda \leftarrow u_2, \dots, s_m \lambda \leftarrow u_m, t_1 \leftarrow v_1, t_2 \leftarrow v_2, \dots, t_n \leftarrow v_n\}$$

から、 $s_i \lambda = u_i$ となるような $(s_i \lambda \leftarrow u_i)$ 、および $v_j = u_k$ となるような $(t_j \leftarrow v_j)$ を取り除けば得られる。

11.5 単一化

【例 11.4】 置換 $\rho = \{u \leftarrow v, h(v) \leftarrow y\}$ と $\lambda = \{v \leftarrow u, a \leftarrow y, b \leftarrow v\}$ の合成置換 $\rho \cdot \lambda$ は

$$\{v \leftarrow v, h(b) \leftarrow y, v \leftarrow u, a \leftarrow y, b \leftarrow v\}$$

から $(v \leftarrow v)$ 、および $(a \leftarrow y)$ 、 $(b \leftarrow v)$ を取り除き、

$$\rho \cdot \lambda = \{h(b) \leftarrow y, v \leftarrow u\} \quad (11.16)$$

となる。これを実際、個々に $\{P(v, y, g(c, u))\}$ に適用すると

$$\{P(v, y, g(c, u))\} \rho = \{P(u, h(v), g(c, u))\},$$

$$\{P(u, h(v), g(c, u))\} \lambda = \{P(v, h(b), g(c, v))\}$$

となり、直接、合成置換 $\rho \cdot \lambda$ を適用するのと同じになる。

置換の合成には結合則

$$(\alpha\beta)\gamma = \alpha(\beta\gamma) \quad (11.17)$$

が成り立つ。

さらに、最も一般的な単一化置換は変項名の違いを除いて一意的である。そして、 $\{L_i\}$ の最も一般的な単一化置換を λ とするとき、任意の単一化置換 α には $\{L_i\} \alpha = \{L_i\} \lambda \beta$ となるような置換 β が存在する。

不一致集合 同一述語で構成される基本論理式もしくはリテラルの集合において、各基本論理式を文字列とみなし、左から見比べていく。共通の文字でない位置から始まる部分項を各基本論理式から取り出して集めたものを不一致集合という。

【例 11.5】

$$\{P(x, f(a, b), c), P(x, f(g(x, y), b), y), P(x, f(g(h(z), a), h(v)))\}$$

に対する不一致集合は

$$\{a, g(x, y), g(h(z), a)\}$$

である。

単一化アルゴリズム 単一化可能なリテラルの集合に対して最も一般的な単一化置換をつくりだし、もしその集合が単一化可能でない場合にはそれを知らせるアルゴリズムである。

いま対象とするリテラル集合を

$$\{L_1, L_2, \dots, L_n\}$$

とする。

[ステップ 0] 空なる置換 λ_0 から出発する。

[ステップ $i+1$] ステップ i ($i \geq 0$) で得られた置換を λ_i とする。

(1) 置換を受けた集合

$$\{L_1, L_2, \dots, L_n\} \lambda_i = \{L_1 \lambda_i, L_2 \lambda_i, \dots, L_n \lambda_i\}$$

の要素がすべて同じになるとき、この λ_i が最も一般的な単一化置換である。そしてアルゴリズムは終了する。なお、単一化された集合要素 $L_1 \lambda_i$ をこの集合の因子(factor)という。

(2) そうでなければ

$$\{L_1 \lambda_i, L_2 \lambda_i, \dots, L_n \lambda_i\}$$

の不一致集合 D_i をつくる。 D_i の中に個体変項 v_i および項 t_i があり、 t_i が v_i を含まないとき、置換 $(t_i \leftarrow v_i)$ をつくり、それを λ_i に加えたものを λ_{i+1} とする。すなわち、

$$\lambda_{i+1} = \lambda_i \cdot (t_i \leftarrow v_i)$$

とする。もしこのような $(t_i \leftarrow v_i)$ がなければ、対象とする集合に単一化置換はないとしてアルゴリズムを終了する。

[例 11.6] つぎの最も一般的な単一化置換を求める。

(1) $\{P(a, f(x, y)), P(u, f(x, b))\}$

最初の不一致集合は $\{a, u\}$ であり、対応する置換 $(a \leftarrow u)$ を行った後の不一致集合は $\{y, b\}$ である。したがって、置換 $\{a \leftarrow u, b \leftarrow y\}$ により、

$$P(a, f(x, b))$$

に単一化される。

(2) $\{P(a, f(x, y)), P(u, f(u, b))\}$

最初の不一致集合は(1)と同じく $\{a, u\}$ である。これを単一化する置換 $(a \leftarrow u)$ により、与えられたリテラル集合は

$$\{P(a, f(x, y)), P(a, f(a, b))\}$$

となる。この不一致集合は $\{x, a\}$ であり、それに対応する置換を行った後の次の不一致集合は $\{y, b\}$ である。したがって、置換 $\{a \leftarrow u, b \leftarrow y, a \leftarrow x\}$ により、

$$P(a, f(a, b))$$

に単一化される。

11.6 導出原理

節集合において、あるリテラル l を含む節とその補リテラル $\sim l$ を含む節を取り出し、それから導出節を求め、節集合に加える。これによりリテラル l を節集合から取り除くことができる。このような操作を続けていくことにより最終的に空節を導くことができればその節集合は充足不可能である(確実性)こと

が証明できる。また、充足不可能であれば有限回の操作で必ず空節を導くことができる(完全性)ことも証明できる。これが導出原理である。本節では、この導出原理について、概略を述べる。

節 A および B があれば、それぞれの個体変項は異なっているとす。もし同じ変項名のものであれば名前をつけかえることにより、それが A と B で重ならないようにすることができる。

いま、節 $\{A\}$ の一部に

$$\{l\} = P(s_{11}, s_{12}, \dots, s_{1n}) \vee P(s_{21}, s_{22}, \dots, s_{2n}) \vee \dots \vee P(s_{L1}, s_{L2}, \dots, s_{Ln})$$

なるリテラルの論理和があり、また別の節 $\{B\}$ の一部に

$$\{m\} = \sim P(t_{11}, t_{12}, \dots, t_{1n}) \vee \sim P(t_{21}, t_{22}, \dots, t_{2n}) \vee \dots \vee \sim P(t_{M1}, t_{M2}, \dots, t_{Mn})$$

なるリテラルの論理和があったとする。さらに、これらのリテラルの集合を単一化する最も一般的な単一化置換 λ があったとする。このとき、導出節

$$r = [\{A\} \lambda - \{l\} \lambda] \vee [\{B\} \lambda - \{m\} \lambda]$$

が得られる。

[例 11.7] 二つの節を

$$\{A\} = P(x) \vee P(f(y)) \vee Q(g(y)),$$

$$\{B\} = \sim P(f(g(a))) \vee Q(f(z))$$

とする。このとき、集合 $\{P(x), P(f(y)), P(f(g(a)))\}$ に対する最も一般的な単一化置換

$$\lambda = \{f(g(a)) \leftarrow x, g(a) \leftarrow y\}$$

により、導出節

$$r = Q(g(g(a))) \vee Q(f(z)) \quad (11.18)$$

が得られる。

導出原理と探索 与えられた述語論理式が恒真であることを証明するためには、その否定が充足不可能であることを証明すればよい。この証明は導出原理により機械的に効率よく行うことができる。

[例 11.8] $\forall x P(x) \rightarrow Q(x)$ および $P(a)$ から $Q(a)$ を証明しよう。

まず、 $\forall x P(x) \rightarrow Q(x)$ を節形式 $\sim P(x) \vee Q(x)$ に変える。題意の否定

$$(\sim P(x) \vee Q(x)) \wedge P(a) \wedge \sim Q(a)$$

を証明する。導出木は図 11.1 のようになり、空節が導かれる。したがって、与えられた問題が証明された。

導出を行うとき、次にどのリテラルに関してそれを行うかにより導出の筋道が異なってくる。このやりかたによっては証明の効率などが異なる。

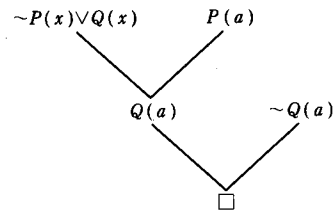


図 11.1 節集合 $\{\sim P(x) \vee Q(x), P(a), \sim Q(a)\}$ の導出木。

【例 11.9】 節集合 $\{P(x) \vee Q(x), \sim P(f(y)) \vee R(y), \sim Q(f(z)), \sim R(v)\}$ の導出木の二つの例は図 11.2 のようになる。

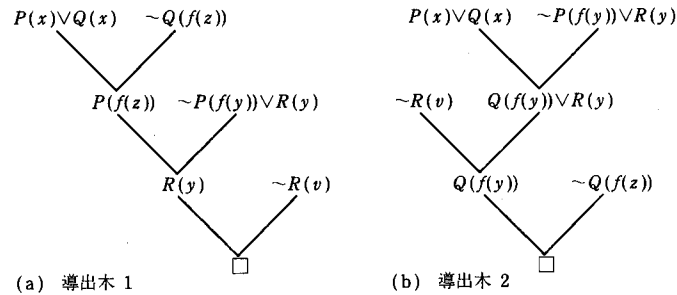
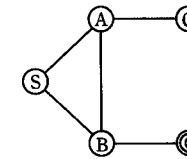


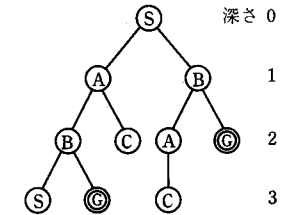
図 11.2 充足不可能な節集合 $\{P(x) \vee Q(x), \sim P(f(y)) \vee R(y), \sim Q(f(z)), \sim R(v)\}$ の導出木例。

基本的には節集合内すべての対について導出の可能性を調べる。空節が導かれればこの手順は終了する。これは探索プロセスである。基本的な探索には図 11.3 に示すように深さ優先探索と幅優先探索がある。深さ優先探索は、一つの枝の先をどこまでも深く探索する方法であり、幅優先探索はできるだけ多くの可能性を同時に探索していく方法である。図 11.3 は同図(a)のグラフを例として、S(スタート)点から G(ゴール)に至る経路を探索する場合のやり方を示している。同図(b)はもとの場所に戻るような経路を除外したすべての経路を木の形で表示している。同図(c)は S から出発して深さ優先探索を行った結果である。この例では G に至る一つの経路が少ない手間で見つかるが、その経路は一般に最短ではない。一方、同図(d)は幅優先探索の結果であり、この例では探索の手間が大きくかかったが、一般に最短経路が見つかる。



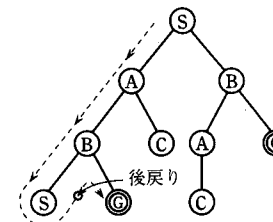
S から G に至る経路を求めたい。

(a) グラフ

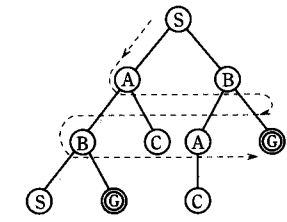


それ以上行けなくなったり、もとに戻ったりすれば探索を打ち切る。

(b) 探索の木



(c) 深さ優先探索



(d) 幅優先探索

図 11.3 グラフの探索

導出原理と質問応答 導出原理はまた質問に対して答を与える求値機構として使うことができる。つぎの例が示すように単一化のプロセスが答を与えてくれるのである。

【例 11.10】 つぎのような事実があるとする。

- (1) 人間ならば死ぬ運命にある。
- (2) 神ならば死なない。
- (3) ソクラテスは人間である。

このとき、質問：

- (4) 神でないものは何か？

上の(1)~(4)を論理式で表現するとつぎのようになる。

- (1)' $\forall x \text{ human}(x) \rightarrow \text{mortal}(x)$
- (2)' $\forall y \text{ god}(y) \rightarrow \sim \text{mortal}(y)$
- (3)' $\rightarrow \text{human}(\text{socrates})$

質問は

- (4)' $\rightarrow \sim \text{god}(z)$

である。

(1)'~(3)' から(4)'を証明するため、(1)'~(3)'の節形に(4)'を否定した節を加えた節集合
 $\{\sim\text{human}(x) \vee \text{mortal}(x), \sim\text{god}(y) \vee \sim\text{mortal}(y), \text{human}(\text{socrates}), \text{god}(z)\}$
 をつくる。これから空節が導出される導出木の例を念のため図 11.4 に二つ示すが、いずれも解 $z=\text{socrates}$ を与えている。

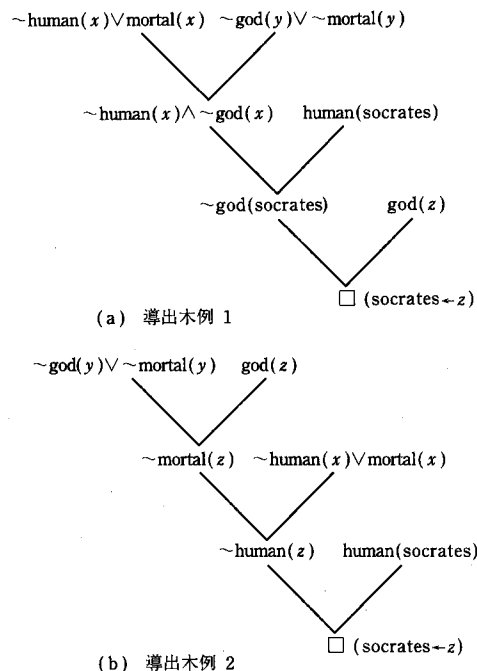


図 11.4 導出原理による質問応答。(a), (b)は導出木は異なるが、いずれにしろ同じ結果を与える。

演習問題 11.

11.1 (1) $\{Q(x, f(x, y)), Q(u, v)\}$

および,

(2) $\{Q(x, f(y, z)), Q(u, v), Q(a, f(h(b), z))\}$

をそれぞれ最も一般的な単一化置換により単一化せよ。

11.2 つぎの節集合(1), (2)の導出木を書け。

(1) $\{P(x) \vee P(f(y)) \vee Q(x), \sim P(x) \vee R(f(x)), \sim R(f(z)) \vee \sim Q(z)\}$

(2) $\{P(x), \sim P(x) \vee Q(x), \sim Q(x) \vee R(f(x)), \sim Q(f(g(y))) \vee S(y), \sim R(f(a)) \vee \sim S(z)\}$

12

ニューラルネットワーク計算

ニューラルネットワークは多数の神経素子と、それらの間を結ぶ重み付き配線からなっている。これらの重みは可変であり、学習によって処理対象に適した重みを得ることも可能である。すなわち、人間のように学習が可能であることが、ニューラルネットワークの大きな特徴である。また、いろいろな処理を同時に行う並列処理が可能であることも大きな特徴である。

ネットワーク表現は前章のような記号表現に比し、より①低レベル、②均一的である。したがって、ニューラルネットワークはパターン認識、画像処理など、あいまいな情報の処理に向いているといえよう。

本章では、ニューラルネットワークの基本原理について述べる。以下、12.1節ではニューラルネットワークの形態について、また12.2節および12.3節ではその代表例である階層型、および相互結合型ネットワークについてそれぞれ述べる。

12.1 ニューラルネットワークの形態

ニューラルネットワークは図 12.1(a)に示すような神経素子を多数接続したものである。個々の素子(神経素子)は入力加重和を求め、それをシグモイド関数(S字型単調増加関数; sigmoid function)に通したものを出力する。シグモイド関数の代表例がロジスティック(logistic)関数

$$f(x) = \frac{1}{1 + e^{-x}} \quad (12.1)$$

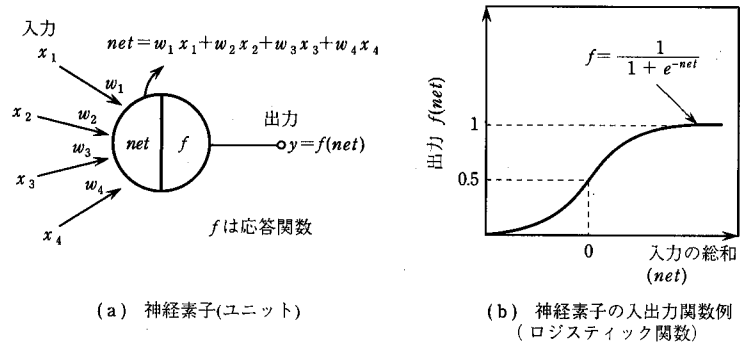


図 12.1 ニューラルネットワークで使われる素子と入出力関係。

である(図 12.1(b))。実際の神経細胞はパルスの頻度で信号を伝えるが、人工のニューラルネットワークでは、パルス頻度をアナログ値(神経の活性値)で置き換えて表現する。ニューラルネットワークの代表的な形態には図 12.2 に示すような階層型ネットワークと相互結合型ネットワークがある。

12.2 階層型ネットワークと誤差逆伝搬学習アルゴリズム

階層型ネットワークにおいては、各神経素子は入力層、中間層、出力層に分かれている。入力層から入ったパターンは、ネットワークで変換を受け、出力層から出力される。同一層内での結合はないとする。階層型ネットワークは基本的にはパターン変換を行うネットワークであり、その特別な場合がパターン認識・識別・分類である。

階層型ネットワークの学習は誤差逆伝搬学習アルゴリズムにより行われることが多い。この学習アルゴリズムにおいては、教師の教えと出力の誤差の責任量が、中間層を通して各結線の重みへフィードバックされる。これにより、3層、4層など多層のネットワークを教育することができる。この学習機能を利用して、手書き文字認識、発音記号の学習、音声認識など多数の応用システムが開発されている。

誤差逆伝搬学習アルゴリズム 誤差逆伝搬学習(error back-propagation learning)アルゴリズムは、教師付き学習アルゴリズムであり、階層型ネット

ワーク(別名 BPN; Back-Propagation Network)における出力素子の出力値ができるだけ教師の教えに近づくよう、重みの調整(学習)が行われる。

図 12.2(a)に示す階層型モデルについて説明する。最上層は出力層であり、出力層を構成する各素子 l の出力 o_l のとるべき値 t_l が教師から与えられる。入力を入力層の各素子に与えられる。中間層も含め、各素子は一つ下の層の素子からの入力の重み付き線形和(net)に対応する出力 $f(net)$ を出す。これらを式で表現すると次のようになる。

出力層の素子 l の出力誤差を次式で定義する：

$$E = \frac{1}{2}(t_l - o_l)^2 \quad (12.2)$$

素子 l の出力 o_l は、

$$o_l = f_l(net_l) \quad (12.3)$$

で与えられる。ここに、

$$net_l = \sum_k w_{kl} o_k \quad (12.4)$$

この o_k は一つ下の層の素子 k の出力である。なお、添字 l, k などは対応する層(l 層および k 層など)内の素子についてとるものとする。以下、同じである。また同様に、

$$o_k = f_k(net_k) \quad (12.5)$$

$$net_k = \sum_j w_{jk} o_j \quad (12.6)$$

⋮

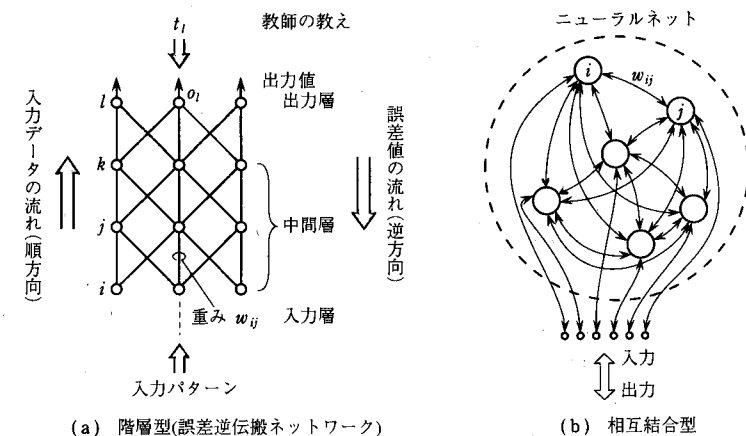


図 12.2 ニューラルネットワークの基本形態

ここで、出力誤差 E を素子 i から j への重み w_{ij} で偏微分する。これは w_{ij} に関する E の感度を表わしており、これが大きいことは、 w_{ij} を変化させることによる E の減少が大きいことを示している。したがって、重み w_{ij} の調整量は $-\partial E / \partial w_{ij}$ に比例した大きさとすればよい。すなわち、

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial net_j} \frac{\partial net_j}{\partial w_{ij}} \quad (12.7)$$

$$\begin{aligned} &= \frac{\partial E}{\partial o_j} \frac{\partial o_j}{\partial net_j} \frac{\partial net_j}{\partial w_{ij}} \\ &= \left[\sum_k \frac{\partial E}{\partial net_k} \frac{\partial net_k}{\partial o_j} \right] \frac{\partial o_j}{\partial net_j} \frac{\partial net_j}{\partial w_{ij}}. \end{aligned} \quad (12.8)$$

いま、

$$\delta_k \triangleq -\frac{\partial E}{\partial net_k}$$

とおく。

$$\frac{\partial net_k}{\partial o_j} = w_{jk}, \quad \frac{\partial net_j}{\partial w_{ij}} = o_i, \quad \frac{\partial o_j}{\partial net_j} = f'_j(net_j) \quad (12.9)$$

であるから、中間層では、

$$\begin{aligned} -\frac{\partial E}{\partial w_{ij}} &= \delta_j o_i \\ &= \left[\sum_k \delta_k w_{jk} \right] f'_j(net_j) o_i \end{aligned} \quad (12.10)$$

なる関係が得られる。ここに、 δ_j は誤差伝搬量(誤差に対する責任の大きさ)であり、次の逆方向伝搬関係式で出力層側から順に求められる。

$$\delta_j = \left[\sum_k \delta_k w_{jk} \right] f'_j(net_j) \quad (12.11)$$

なお、出力層においては δ_i は、

$$\begin{aligned} \delta_i &= -\frac{\partial E}{\partial net_i} \\ &= -\frac{\partial E}{\partial o_i} \frac{\partial o_i}{\partial net_i} = (t_i - o_i) f'_i(net_i) \end{aligned} \quad (12.12)$$

で与えられる。

出力における δ_i は式(12.12)から出発して、順次逆向きに式(12.11)を適用していくと、各素子における誤差伝搬量 δ_j が求められる。これらをもとに、式(12.10)から結合重み w_{ij} の E に関する責任量である $\partial E / \partial w_{ij}$ が求められる。これから、たとえば時刻 n における w_{ij} の増加量を、適当な正係数 η を用いて、

$$\Delta w_{ij}^n = -\eta \frac{\partial E^n}{\partial w_{ij}} \quad (12.13)$$

とする。ここに E^n は時刻 n における出力誤差量である。この方法は、 E のいちばん小さくなる方向に向ってパラメータを調整することから最急降下法(steepest descent method)と呼ばれる。あるいは、雑音に惑わされずに安定に降下するため過去の重み変化に対する慣性項 α ($0 \leq \alpha < 1$) を導入して、

$$\Delta w_{ij}^n = -\eta \frac{\partial E^n}{\partial w_{ij}} + \alpha \Delta w_{ij}^{n-1} \quad (12.14)$$

とする(モーメント法)。

【例 12.1】 図 12.3 のようなネットワークがある。○印は単純ないき値素子であり、そのいき値が○内に記されている。入力 x_1, x_2 の総和がいき値以下のとき、出力は 0、それ以外のときには出力は 1 となる。線上的数値は結合の重みである。 x_1, x_2 は入力、 z_1, z_2 は中間層の出力、 y はネットワークの出力である。

(1) x_1, x_2 が $\{0, 1\}$ の値のみをとるとき、真理値表は表 12.1 のようになる。このとき、出力 y は x_1, x_2 の排他的論理和(XOR)になっている。XOR のような非線形識別は単独の神経素子では困難であっても、多層にすると可能となる。なお、問題の簡単化のためにいき値素子を使用したがる、いき値素子の入出力特性は微分可能ではない。そのため誤差逆伝搬学習は不可能である。しかしながら、パターン変換・識別は可能である。

(2) x_1, x_2 が連続値をとるとき、出力 y は図 12.4 のようになる。ここでは線形識別関数を二つ並べた非線形識別関数でパターンの識別が行われていることがわかる。

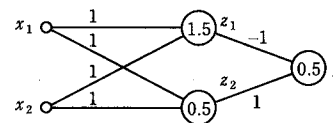


図 12.3 例 12.1 のネットワーク

表 12.1 図 12.3 のネットワークの真理値表

$x_1 \backslash x_2$	0	1
0	$z_1 = 0$ $z_2 = 0$ $\therefore y = 0$	$z_1 = 0$ $z_2 = 1$ $\therefore y = 1$
1	$z_1 = 0$ $z_2 = 1$ $\therefore y = 1$	$z_1 = 1$ $z_2 = 1$ $\therefore y = 0$

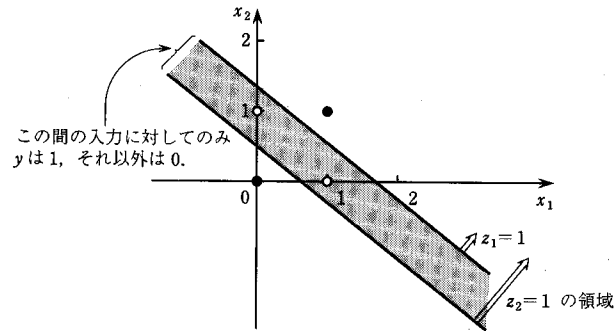


図 12.4 連続値入力に対する図 12.3 のネットワークの出力。

12.3 相互結合型ネットワーク

相互結合型ネットワーク (Hopfield ネットワーク) は、素子間に入出力の区別がなく、それらの間に相互の結合を許すものである。基本的な動作として、複数のパターンを結合の重みとしてネットワークに格納したのち、あるパターンの一部を入力すると、それに最も近いパターンがしばらくするとネットワーク上に素子の活性状態として想起・連想される。たとえば、多数の顔画像を格納し、ある人の鼻の部分だけ入力すると、その人の顔全体が出力される。また、人間の意識・無意識活動のように、多くのあいまいな要因が絡んだ相互作用のある系における問題解決やシミュレーションには相互結合型ネットワークが適している。

相互結合型ネットワークによる処理において使われる評価基準はネットワーク全体のエネルギー最小化である。問題に応じて適切な重みを選択する必要があるが、これにより、巡回セールスマン問題のような組合せ問題も解ける。個々の素子の活性状態を局所的状態に応じて変化させると全体のエネルギーが最小値に向い、解が得られる。好ましくない極小値に引っかかるのを避けるため、金属の焼きなましとのアナロジーから、確率的に素子の活性度を変化させる模擬焼きなまし (simulated annealing) の手法なども提案されているが本書ではふれない。

12.3.1 エネルギー最小化と連想

図 12.2(b) のようなニューラルネットワークにおいて、 w_{ij} を素子 i と素子 j の間の重みであるとする。素子 i の活性状態を u_i ($=0$ または 1) で表わす。すなわち、 $u_i=0$ のときには素子 i は非活性状態であることを表わし、 $u_i=1$ のときには活性状態であることを表わす。さらに、重み w_{ij} は i, j に関して対称、すなわち

$$w_{ij} = w_{ji} \quad (12.15)$$

であるとする。また、

$$w_{ii} = 0 \quad (12.16)$$

すなわち、自分自身への結合はないとする。

ここで、ネットワークのエネルギーを

$$E = -\frac{1}{2} \sum_i \sum_j w_{ij} u_i u_j - \sum_i (s_i - \theta_i) u_i \quad (12.17)$$

で定義する。ここで、 s_i および θ_i は、それぞれ素子 i の外部入力およびいき値である。このとき簡単にいえば、大きな重み w_{ij} で結ばれた素子は同時に興奮し ($u_i=1$ かつ $u_j=1$)、小さな重み w_{ij} で結ばれた素子どうしは同時に興奮することがない ($u_i=0$ または $u_j=0$) とき、第 1 項は小さく (エネルギー状態が低く) なる。また第 2 項は、外部入力 s_i が大きな素子はできるだけ活性化されたほうがネットワーク全体のエネルギーは小さくなることを表わしている。

いま、素子 i の値 (活性値) が 0 から 1 に変わったとする。これによる E の変化は

$$\Delta_i E = - \sum_j w_{ij} u_j - (s_i - \theta_i) \quad (12.18)$$

である。また、素子 i の値が 1 から 0 に変わった場合には、これによる E の変化は

$$\Delta_i E = \sum_j w_{ij} u_j + (s_i - \theta_i) \quad (12.19)$$

である。素子 i が以下の規則にしたがって状態変化をするとする。

$$\begin{cases} \sum_j w_{ij} u_j + s_i \geq \theta_i \text{ ならば, } u_i = 1 \\ \sum_j w_{ij} u_j + s_i < \theta_i \text{ ならば, } u_i = 0 \end{cases} \quad (12.20)$$

すなわち、他の素子からの入力と外部入力 s_i の総和がいき値より大きいとき、素子は 1 となり、小さいとき 0 となる。このとき、 $\Delta_i E$ は負または零であり、

正になることはない。したがって、上のような状態変化規則に従うとき、系全体のエネルギーは順次減少していくことになる。ただし、ネットワークのエネルギーはどこまでも小さくなれるわけではなく、最小値が存在する。なぜなら各 u_i は有界で 0 と 1 の間の値(実際上は 0 と 1 のみ考えておけば十分)しかとれないからである。最小値を与える u_i の組 $\{u_i\}$ は 1 個とは限らない。また、最小値以外に極小値があるとそこに引っかかって抜け出れないことがある。

このように、ある初期状態から出発したネットワークは状態変化規則に従って状態変化を繰り返すうちに、最小値または極小値に落ち込むことになる。そこで、エネルギーが最小または極小になるこれらの状態をある意味での望ましい状態としておくと、状態遷移によりそれらの状態が誘導されて出てくことになる。すなわち、望ましい状態にある程度近いパターンを入力すると、望ましい状態が連想出力されるのである。最小値を含め、エネルギーの極小値は一般に複数個存在するから、連想される状態もそれに対応して一般には複数個存在する。

図 12.5、図 12.6 には三つのパターン C, F, 4 を覚え込ませたネットワークに、ある初期状態を入力するとそれに近いパターンが連想出力されてくる様子を示す。連想したいパターンをネットワークエネルギー最小または極小状態に対応させる方法については次の項に示す。

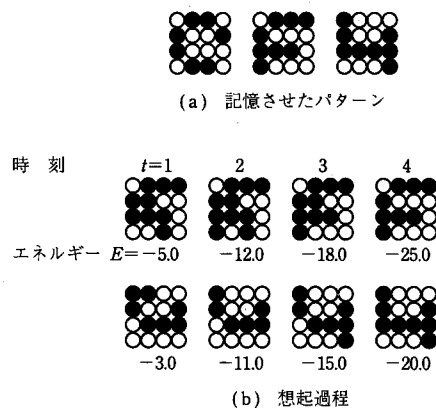


図 12.5 自己想起型文字連想過程におけるネットワークの状態変化とエネルギーの減少

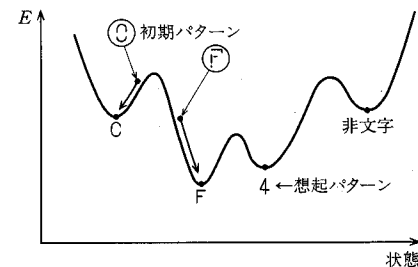


図 12.6 連想過程とネットワークのエネルギー変化

[例 12.2] (連想ネットワーク)

図 12.7 のようなネットワークがある。各素子の入出力関数 f は図 12.8 に示すように、区間 $[0.0, 1.0]$ 内の入力はそのまま出力し、負入力に対しては最小値 0.0、また 1.0 を越える入力に対しては最大値 1.0 をそれぞれ出力する。

(1) 最初(時刻 $t=1$)に a(肉食)だけを 0.1 の活性値で刺激したとき、ネットワークの活性状態はどのように変化していくであろうか。刺激は $t=2$ 以降では取り去るものとする。

図 12.7 より、ネットワークの重み行列は

$$W = \begin{bmatrix} a & b & c & d & e \\ 0 & -1 & 1 & 1 & -1 \\ -1 & 0 & 1 & -1 & 1 \\ 1 & 1 & 0 & 1 & 1 \\ 1 & -1 & 1 & 0 & -1 \\ -1 & 1 & 1 & -1 & 0 \end{bmatrix} \begin{matrix} a \\ b \\ c \\ d \\ e \end{matrix}$$

である。

時刻 $t=1$ における初期活性値ベクトルは

$$u_1 = (0.1, 0, 0, 0, 0)$$

である。時刻 $t=2$, およびそれ以降の活性値ベクトルは

$$u_2 = f(u_1 W)$$

$$= (0, 0, 0.1, 0.1, 0)$$

$$u_3 = f(u_2 W)$$

$$= (0.2, 0, 0.1, 0.1, 0)$$

$$u_4 = (0.2, 0, 0.3, 0.3, 0)$$

$$u_5 = (0.6, 0, 0.5, 0.5, 0)$$

$$u_6 = (1.0, 0, 1.0, 1.0, 0)$$

$$u_7 = (1.0, 0, 1.0, 1.0, 0)$$

⋮

となる。したがって、a(肉食), c(動物), d(ライオン)の三つが最終的に活性化されることがわかる。これは a(肉食)からこれらが連想されたことを表わしている。

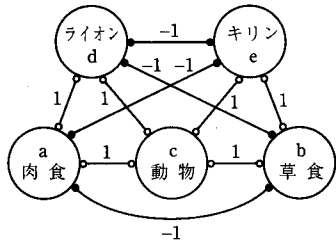
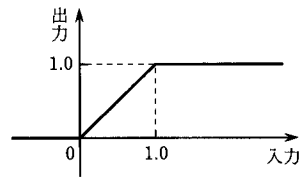


図 12.7 動物連想ネットワーク

図 12.8 入出力関数 f

(2) 時刻 t におけるネットワークのエネルギーを

$$\begin{aligned} E(t) &= -\frac{1}{2} \sum_{i,j} w_{ij} u_i u_j \\ &= -\frac{1}{2} \mathbf{u}_t^T \mathbf{W} \mathbf{u}_t \end{aligned} \quad (12.21)$$

で定義する。 u_i は素子 i の活性値, w_{ij} は素子 i, j 間の重みである。このとき, 各時刻におけるエネルギーは

$$\begin{aligned} E(1) &= -\frac{1}{2} \mathbf{u}_1^T \mathbf{W} \mathbf{u}_1 = 0, \\ E(2) &= -\frac{1}{2} \mathbf{u}_2^T \mathbf{W} \mathbf{u}_2 = -0.01, \\ E(3) &= -0.05, \\ E(4) &= -0.21, \\ E(5) &= -0.85, \\ E(6) &= E(7) = \dots = -3.0. \end{aligned}$$

時刻 $t=6$ 以降では, $E(t)$ は最小値に落ち込んでいる。

この例で

12.3.2 パターンの記憶

ここでは K 個の n 次元パターンベクトル $\mathbf{u}^s = (u_1^s, u_2^s, \dots, u_n^s)$ ($s=1, 2, \dots, K$) をネットワークに記憶させることを考える。そこで, 重み w_{ij} を

$$\begin{aligned} w_{ij} &= \sum_s (2u_i^s - 1)(2u_j^s - 1) \\ w_{ii} &= 0 \end{aligned} \quad (12.22)$$

としよう。ここで, $s_i=0, \theta_i=0$ としておく。上の w_{ij} は, 同時に興奮 ($u_i^s = u_j^s = 1$) する, または同時に抑制される ($u_i^s = u_j^s = 0$) ような素子間の重みは大きくなることを示している。

いまネットワークの状態が記憶された状態 $s=L$ にあるとする。このとき, 素子 i への入力 I_i の総和は

$$I_i = \sum_j w_{ij} u_j^L = \sum_s (2u_i^s - 1) \left[\sum_j u_j^L (2u_j^s - 1) \right] \quad (12.23)$$

となる。 $[\]$ 内の値は s により変化する。 u_j^L は 0, 1 をとり, $(2u_j^s - 1)$ は ± 1 をとる。パターン L と s が独立であれば $[\]$ 内の値はほぼ 0 となる。実際, 記憶パターンにそのような条件をつけることは厳しいことではない。したがって, $[\]$ は $s=L$ の場合のみ正の値 C をもち, それ以外の場合には 0 とすると,

$$I_i = (2u_i^L - 1) C \quad (12.24)$$

となる。いま仮定により, ネットワークは状態 L にあるから, 素子 i の値は $u_i = u_i^L$ である。つぎの時刻における u_i は

$$\begin{cases} u_i^L = 1 \text{ ならば, } I_i > 0. \text{ したがって, 次の時刻における } u_i \text{ も } 1, \\ u_i^L = 0 \text{ ならば, } I_i < 0. \text{ したがって, 次の時刻における } u_i \text{ も } 0 \end{cases}$$

となる。すなわち, ネットワークの状態は変化しない。これはネットワークが平衡状態にあることを示している。したがって, 状態が記憶パターン $\mathbf{u}^L$ の近くにあれば, 引き込まれて状態は $\mathbf{u}^L$ となり, そこで落ちつく。すなわち, 初期状態に近い記憶パターンが連想されることになる。

【例 12.3】巡回セールスマン問題

巡回セールスマン問題とは, いくつかの都市を一度づつ通り, もとに戻ってくる問題で, そのときたとえば費用や所要時間などのコストを最小にするにはどういう順番で回ればよいかを求める問題である。これは人にとっては近似解がすぐにみえてくるが, 従来法で計算機にやらせようとする, 都市の数 n とともに急激に計算時間が増大する (n の多項式オーダーの時間では解けない) やっかいな問題として知られている。

これに対して, 図 12.9(a) に示すように, 都市名と訪問の順番の組を一つの素子に割り当てる。このとき, ネットワークのエネルギー E が問題の拘束と巡回のコストの加重和になるようにする。すなわち, (1) 同一都市は一度しか訪れないから, 同一都市の異なる訪問順の素子間には抑制性重みをつける, また, (2) 同じ訪問順番のときには一つの都市しか訪問できないから, 同一訪問順番の異なる都市間には抑制性重みをつける, (3) それ以外の素子間には, 都市間移動のコストに応じた重みをつける。

これらの各拘束 (1) ~ (3) はいくつの素子間結合を与えるか求めてみよう。

(1) 神経素子配列 (図 12.9(a)) の各行においてはすべての組合せ nC_2 個の結合に相互抑制性重みがつけられる。行の数は n である。したがって, 総計

$$nC_2 \cdot n = \frac{1}{2} n^2 (n-1)$$

個の重みを与える。

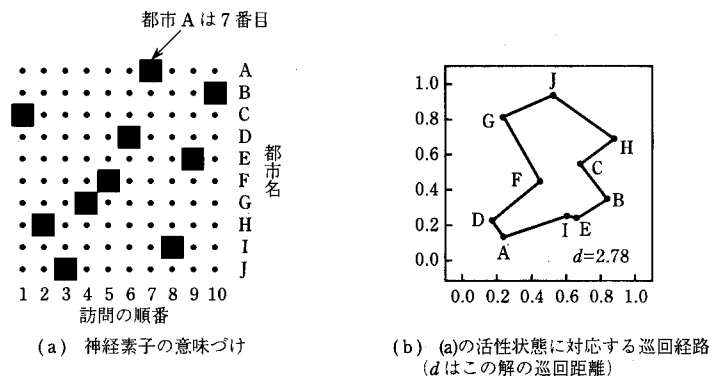


図 12.9 巡回セールスマン問題

(2) 同様にして

$${}_nC_2 \cdot n = \frac{1}{2} n^2 (n-1).$$

(3) 図 12.9(a)の各神経素子からはすぐ右隣の列の異なる都市の神経素子へ抑制性重みがつけられる。右端の列から左端の列へ同様に付けられる。すなわち、各神経素子からは $(n-1)$ 個の相互抑制性結合が出ていく。神経素子は n^2 個あるから、総計

$$n^2(n-1)$$

個の結合を与える。

なお、上の(1)~(3)の結合の場所は重ならないから、(1)~(3)合せて $2n^2(n-1)$ 個の結合がつくられる。これらの結合の重みは通常抑制性のみであり、このままではすべての神経素子の活性値が0になったときネットワークのエネルギーが最小になってしまう。そこで実際には、たとえば n^2 個の神経素子の活性値の総和と n との差の2乗に比例した量をエネルギーにつけて加えておく。

適当な初期状態から出発し、各素子は局所的な状況に従って状態を変化させていくと、ネットワークのエネルギーは順次低下していき、問題の拘束と巡回のコストを両方バランス良く満足させる最小値へ到達することが期待できる。そのときの各素子の活性値が解を与える(図 12.9(b))。しかしながら、しばしば最小値に到達する前に極小値に引っかかる。このときには、最良解ではなく近似解となる。

演習問題 12.

12.1 最急降下法

図 12.10 の 2 層ネットワークにおいて、出力素子の入出力特性を $f(net) = (net)^k$ とする。ここに $net = xW^T$, $x = (x_1, x_2, x_3)$ はネットワークの入力, $W = (w_1, w_2, w_3)$ は重みベクトルである。なお, $net > 0$ の範囲のみ考える。ネットワークの出力と教師の教え t の 2 乗誤差 E が小さくなる方向に修正する最急降下法に従って、重み W を修正せよ。

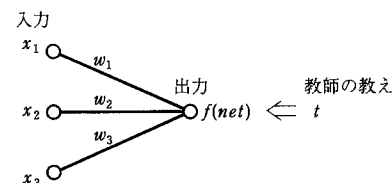


図 12.10 2層ネットワーク

12.2 例 12.2 において、c (動物) だけを 1.0 の活性値で刺激したとき、ネットワークの状態および $E(t)$ は振動する可能性があるが、この状態は不安定で、素子のばらつきなどにより容易に最小エネルギーの安定状態におちつくことを説明せよ。

12.3 A/D 変換器

相互結合型ネットワークにより A/D 変換器を作りたい。入力電圧を V_{in} , LSB (最下位ビット) に相当する最小量子化ステップ電圧を v_0 , n ビットの変換結果を $x_{n-1}, x_{n-2}, \dots, x_k, \dots, x_0$ ($x_k=0$ または 1) とする。3 ビットの A/D 変換器を実現する相互結合型ネットワークは図 12.11 で与えられることを示せ。

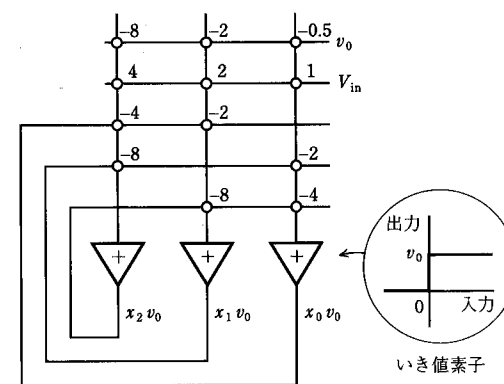


図 12.11 3 ビット A/D 変換ネットワーク (結線交点
○印の数値は結合重み)

付 録

A1 デイビス・パットナムのアルゴリズム

ある論理式 P の恒偽性を調べるため、まず P を連言標準形

$$P = P_1 \wedge P_2 \wedge \cdots \wedge P_n \quad (\text{A1.1})$$

になおす。 P が恒偽であることを証明するためには、ある節 P_i が偽であるか、またはいくつかの節の連言 $P_i \wedge P_j \wedge \cdots$ からつくられる新しい節が偽であることを証明すればよい。これを次のような操作の繰返しにより行う。

- (1) l をリテラルとする。 $P_i = l$ かつ $P_j = \sim l$ であるとき、 P は恒偽である。
- (2) l をリテラルとする。 $P_i = l \vee \sim l \vee \cdots$ の形をした節があるならば、それを P から取り除いた論理式をつくる。それを P' としたとき、 P' が空、すなわち何も残らなくなれば P は充足可能であり、恒偽ではない。
- (3) P_i がただ一つのリテラル l からなる場合には P から P_i を取り除き、それを P' とする。 P' が空の場合には P は恒偽とはならない。空でないとき、 P' 中のリテラル $\sim l$ はすべて取り除き、それを P'' とする。たとえば、 $P = l \wedge (\sim l \vee A)$ のとき $P' = \sim l \vee A$ であり、 $P'' = A$ である。また、 $l \wedge (\sim l \vee A) \wedge (\sim l \vee B)$ のときには $P'' = A \wedge B$ である。
- (4) リテラル l は P 中にあるが、 $\sim l$ が P 中になく、 l を含む節 P_i はすべて P から除去する。このようにして得られた論理式 P' が空であれば P は充足可能となり、偽ではない。
- (5) あるリテラル l に関し、 P が

$$P = (l \vee A_1) \wedge \cdots \wedge (l \vee A_m) \wedge (\sim l \vee B_1) \wedge \cdots \wedge (\sim l \vee B_n) \wedge C \quad (\text{A1.2})$$

なる形をしている場合には

$$P = (A_1 \wedge \cdots \wedge A_m \wedge C) \vee (B_1 \wedge \cdots \wedge B_n \wedge C) \quad (\text{A1.3})$$

と置き換えることができる。ただし、 A_i , B_j および C は l , $\sim l$ を含んでいないとする。以降は (A1.3) の二つの連言標準形がともに恒偽であるかどうかを調べる。

〔説明〕 上の(1), (2)は明らかである。

(3)の場合, P' が空になれば $I=T$ のとき $P=T$ となるから P は充足可能である。空でないときは, 意味の木が図 A.1 のようになることから, $I \wedge (\sim I \vee A)$ の恒偽性を調べることは A の恒偽性を調べることに同値である。

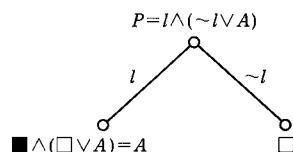


図 A.1 ステップ(3)の意味の木。

■は真, □は偽を表わす。

(4)の場合, $P = (I \vee A) \wedge B$ とすると意味の木が図 A.2 のようになることから, P が充足不可能であることを示すためには B および $A \wedge B$ がともに充足不可能, すなわち B が充足不可能であることを示すことが必要十分である。

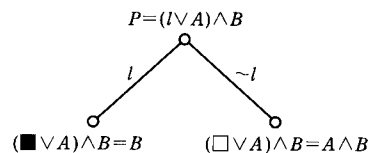


図 A.2 ステップ(4)の意味の木

(5)の場合, 意味の木は図 A.3 のようになる。したがって, (A 1.2) が充足不可能であることを示すためには両方の枝の論理式の選言, すなわち (A 1.3) が充足不可能であることを示せばよい。

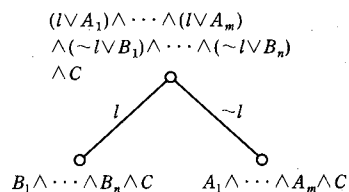


図 A.3 ステップ(5)の意味の木

以上(1)～(5)のステップにより連言標準形の論理式 P を簡単化していき, いずれかの節が偽(□)となれば, もとの論理式 P が充足不可能, すなわち恒偽であることが証明される。アルゴリズムの1回の繰返しでステップ(3)～(5)の少なくとも一つは適用され, 論理変項は必ず一つは減る。したがって, このアルゴリズムは必ず終了する。

(5)の基本的な操作は二つの節 $L \vee A$ および $\sim L \vee B$ から $A \vee B$ を導くことであり, この場合, 特に節 $A \vee B$ を節 $L \vee A$ および節 $\sim L \vee B$ の導出節(resolvent)と呼ぶ。これは, $(L \vee A) = (\sim A \rightarrow L)$ および $(\sim L \vee B) = (L \rightarrow B)$ を用いて三段論法により $(\sim A \rightarrow B) = (A \vee B)$ を得るのと同じである。

〔例 A 1.1〕 $P = (A \vee B) \wedge (\sim A \vee C) \wedge (\sim B) \wedge (\sim C)$ が恒偽であるかどうかを調べよう。図 A.4 のように導出節をつくっていくことにより空節(□)が導かれる。したがって, P は恒偽である。

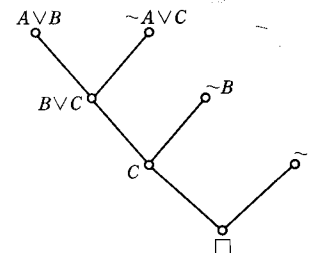


図 A.4 例 A 1.1 の導出木

A2 命題論理の公理系

これまでは, 命題論理式の真偽をそれに含まれる各基本論理式の真偽から求める方法について述べてきた。基本的には真理値表を用いて論理式の真偽を決めることができた。このような方法は, どのような解釈・モデルのもとで与えられた論理式が真となるかという議論に基づいており, モデル理論と呼ばれている。この方法はまた, 個々の基本論理式の真偽と論理式全体の真偽の関係を論ずるという立場から意味論とも呼ばれる。

一方, これとは別に, 論理式の形を公理系と推論規則により変えていくことにより, 論理式の性質を論じる方法がある。これは論理式の形や構造を変形していくという形式から構文論的方法ともいわれる。以下ではこれについて述べる。論理式の形式はさきに述べたもの(wff)とまったく同じである。

命題論理の公理系としてここでは, 以下の公理系を用いる。

(A 1): $P \rightarrow (Q \rightarrow P)$

(A 2): $(P \rightarrow (Q \rightarrow R)) \rightarrow ((P \rightarrow Q) \rightarrow (P \rightarrow R))$

(A 3): $(P \rightarrow Q) \rightarrow (\sim Q \rightarrow \sim P)$

公理系としては, これ以外に種々のものが考えられているが, それらは本質的に同じものである。

また, 推論規則は次のように与えられる。

(I1): P と $P \rightarrow Q$ が定理式ならば, Q は定理式である。

この推論規則は **modus ponens** といわれるが, これを三段論法という人もいる(本来の意味の三段論法は $P \rightarrow Q$ および $Q \rightarrow R$ から $P \rightarrow R$ を導くものである)。このとき, Q は P および $P \rightarrow Q$ から直接導かれるという。公理式は定理式であるので, (I1)を繰返し適用することにより種々の定理式を導くことができる。

A 2.1 命題論理の形式的証明

定理式の形式的証明とは、公理系(A 1), (A 2), (A 3)と推論規則とから定理式を求めていく過程をいう。

公理 A_1 から始まる論理式の有限列

$$A_1, A_2, \dots, A_n$$

があり、 $k=2, 3, \dots, n$ なる A_k は

- (1) 公理である、または
- (2) $i, j < k$ なる A_i , および A_j から推論規則(I 1)によって直接導かれたものである。

という場合、列 $A_1, A_2, \dots, A_n$ を A_n の形式的証明という。

$A_1, A_2, \dots, A_{n-1}$ は A_n を導き出すため直接的・間接的な役割を果たしている。論理式 A の形式的証明があるとき、 A は証明可能である、または A は定理であるといい、

$$\vdash A$$

と書く。

[例 A 2.1] $\vdash (A \rightarrow A)$ を証明する。

公理(A 1)において $P=A$, $Q=(A \rightarrow A)$ とおくと

- (1) $A \rightarrow ((A \rightarrow A) \rightarrow A)$

を得る。さらに公理(A 2)において $R=A$ とおくと

- (2) $(A \rightarrow ((A \rightarrow A) \rightarrow A)) \rightarrow ((A \rightarrow (A \rightarrow A)) \rightarrow (A \rightarrow A))$

を得る。(1)と(2)に推論規則 modus ponens を適用して

- (3) $((A \rightarrow (A \rightarrow A)) \rightarrow (A \rightarrow A))$

を得る。いま、公理(A 1)において $P=A$, $Q=A$ とおくと

- (4) $A \rightarrow (A \rightarrow A)$

を得る。ここで再び(3), (4)に推論規則を適用すると

- (5) $A \rightarrow A$

を得る。すなわち、公理式(1)から出発して推論規則を適用しつつ(5)を得たから

$$\vdash (A \rightarrow A)$$

である。 $A \rightarrow A$ は定理であり、その形式的証明は列(1), (2), ..., (5)である。

A 2.2 公理系の無矛盾性と独立性

いかなる論理式 P をとっても、 P と $\sim P$ の両方が証明可能とならないことを、公理系は無矛盾であるという。 P と $\sim P$ の両方が証明可能ならば、その公理系は矛盾しているという。 P と $\sim P$ が証明可能であれば、 $P \wedge \sim P$ も証明可能(定理)となる。このとき、恒真式 $(P \wedge \sim P) \rightarrow Q$ から任意の Q は証明可能となる。すなわち、公理系が矛盾していれば、いかなる論理式も証明可能となる。さきに述べた命題論理の公理系(A 1), (A 2), (A 3)は無矛盾である。

公理系のある公理が、他の公理と推論規則とから証明可能であれば、その公理は必要のない冗長な公理であるとみなすことができる。すべての公理が他の公理と推論規則から証明可能でないとき、各公理は独立であるという。さきの公理系(A 1), (A 2), (A 3)は互いに独立であることが知られている。

A 2.3 完全性

本項では上で述べた二つのシステムの間の関係について述べる。結論をいうと、モデル理論における恒真式は公理系から導かれる定理式となっており、また逆に公理系から導かれる定理式はモデル理論における恒真式となっている。したがって、これら二つのシステムはまったく同値なものである。これを完全性定理という。

定理 A 2.1 公理系から導かれる定理式は恒真である。

証明 公理系の式(A 1), (A 2), (A 3) が恒真であることは真理値表より容易にわかる。一般に P および $P \rightarrow Q$ が真であるならば Q も真となることも真理値表よりわかる。すなわち、推論規則は恒真性を保持する。したがって、恒真な公理に推論規則を適用して得られた定理式はすべて恒真である。

定理 A 2.2 恒真式は定理式である。

この証明は省略する。

A 3 エルブラン空間

11.3 節に述べた手順で節集合 S が得られる。これらは全称記号に由来する変項を含むため、充足不可能性をそれらのすべての場合について調べることは一見不可能のようにみえる。しかしながら、以下に述べる高々可算無限個のエルブラン空間、あるいはそれから求められるエルブラン基底についてのみ S の充足不可能性を調べればよいことがエルブラン(Herbrand)により示された。 S が充足不可能の場合、この手続は有限回で終了する。

この原理を簡単な例について示そう。

いま、 $\{P(x), \sim P(a)\}$ なる節集合があったとする。矛盾が露呈するのは $x=a$, すなわち、 $\{P(a), \sim P(a)\}$ となった場合であり、この場合二つの節の“干渉”がいちばん強くなって充足不可能性が明確化したのである。

また、 $\{P(x) \vee Q(y), \sim P(z), \sim Q(u)\}$ なる節集合の場合、いちばん“干渉”が強くなり矛盾が露呈しやすいのは $x=y=z=u$ のときである。これ以外の場合、たとえばどれか一つの変項が他と異なるときには矛盾がみえない。そこで、この節集合の充足不可能性を調べるには $x=y=z=u=a$ (a は仮につくった定項)とおいた $\{P(a) \vee Q(a), \sim P(a), \sim Q(a)\}$ の充足不可能性を調べればよいことがわかる。変項をすべての範囲にわたって動かして節集合の充足不可能性を調べるのではなく、高々可算無限個のエルブラン空間の要素についてのみ調べればよいのである。上の例ではエルブラン空間は $\{a\}$ である。以下ではエルブラン空間の正式な定義を記す。

定義 節集合の S のエルブラン空間(Herbrand universe) H_S はつぎのように再帰的に定義される。

- (1) S 内のすべての定項は H_S の要素である。 S が定項を含まなければ、ある任

意の一つの定項 a が H_s に含まれるとする。

(2) S 内の n 項関数を f とし、項 $t_1, t_2, \dots, t_n \in H_s$ とする。このとき、 $f(t_1, t_2, \dots, t_n) \in H_s$ である。

(3) 他のいかなる項も H_s に存在しない。

【例 A 3.1】 次の節集合 S を考える。

$$S = \{P(a, x), Q(y)\} \quad (\text{A 3.1})$$

このときエルブラン空間は

$$H_s = \{a\} \quad (\text{A 3.2})$$

となる。

【例 A 3.2】 次の節集合 T を考える。

$$T = \{P(x, y), Q(x, f(x))\} \quad (\text{A 3.3})$$

このとき、エルブラン空間 H_T はつぎのようになる：

$$H_T = \{a, f(a), f(f(a)), f(f(f(a))), \dots\}. \quad (\text{A 3.4})$$

【例 A 3.3】 次の節集合を考える。

$$U = \{P(x) \vee \sim P(f(x)) \vee Q(a), P(g(x, y)) \vee \sim Q(b)\} \quad (\text{A 3.5})$$

この U の定項は $\{a, b\}$ であり、関数は $\{f, g\}$ である。このとき、エルブラン空間は

$$H_U = \{a, b, f(a), f(b), g(a, a), g(a, b), g(b, a), g(b, b), f(f(a)), f(f(b)), g(a, f(a)), g(a, f(b)), \dots\} \quad (\text{A 3.6})$$

となる。

これらの例を比較してみればわかるように、節集合が関数記号を含むときエルブラン空間は無限個の要素を含み、関数記号を含まないときには有限となる。無限の場合でも、それらは順に番号をふっていける。したがって、一般には、エルブラン空間は高々可算無限の要素からなる。

節集合 S に含まれる述語 $P(t_1, t_2, \dots, t_m)$ の各項にエルブラン空間 H_s の要素を入れたものを基礎例 (ground instance) という。たとえば、例 A 3.2 の基礎例は

$$P(a, a), Q(a, a), P(a, f(f(a))), Q(f(a), f(a))$$

などである。基礎例のすべての集合をエルブラン基底 (Herbrand base) B_s という。

例 A 3.1 のエルブラン基底は

$$B_s = \{P(a, a), Q(a)\} \quad (\text{A 3.7})$$

また例 A 3.2 では

$$B_T = \{P(a, a), Q(a, a), P(a, f(a)), Q(a, f(a)), P(f(a), a), \dots\} \quad (\text{A 3.8})$$

となる。

エルブラン空間が可算個の要素をもつことから、エルブラン基底も可算個の要素からなる。それをいま

$$B = \{B_1, B_2, \dots, B_n, \dots\} \quad (\text{A 3.9})$$

とする。ここで、エルブラン基底の各要素に **T** または **F** を割り当てるため、

$$h_n = B_n \quad (B_n \text{ に } \mathbf{T} \text{ を割り当てたとき})$$

$$h_n = \sim B_n \quad (B_n \text{ に } \mathbf{F} \text{ を割り当てたとき})$$

とし、

$$HI = \{h_1, h_2, \dots, h_n, \dots\} \quad (\text{A 3.10})$$

をつくる。これをエルブラン解釈 (Herbrand interpretation) という。エルブラン解釈とは、節集合 S 中の各々の述語が取りうる可能性のある真偽値を表現したものである。ただし、述語名が同じ場合にはそれらの間の拘束は考慮して求める。たとえば、

$$S = \{P(x) \vee Q(x), R(f(y))\} \quad (\text{A 3.11})$$

のとき、

$$HI_1 = \{P(a), Q(a), \sim R(a), \sim P(f(a)), Q(f(a)), R(f(a)), \dots\},$$

$$HI_2 = \{\sim P(a), \sim Q(a), R(a), \sim P(f(a)), Q(f(a)), R(f(a)), \dots\}$$

などである。また、

$$T = \{P(x, f(x)) \vee Q(a)\} \quad (\text{A 3.12})$$

のとき、エルブラン解釈の例は

$$HI_1 = \{P(a, a), \sim Q(a), P(f(a), a), \sim P(a, f(a)), Q(f(a)), \sim P(f(a), f(a)), \dots\},$$

$$HI_2 = \{\sim P(a, a), Q(a), P(f(a), a), P(a, f(a)), \sim Q(f(a)), P(f(a), f(a)), \dots\}$$

である。

エルブラン解釈はエルブラン基底集合の各要素に **T**, **F** を任意に与えたものである。エルブラン解釈は可算個であるから、基本的にはエルブラン基底集合の各要素の **T**, **F** に従って意味の木を作成すれば、節集合 S が充足不可能であるとき木の葉がどこかで必ず偽となる閉じた木ができる。

たとえば、節集合を

$$S = \{P(x) \vee Q(x), \sim P(f(y)), \sim Q(f(y))\}$$

とする。このエルブラン基底集合は

$$B_s = \{P(a), Q(a), P(f(a)), Q(f(a)), P(f(f(a))), Q(f(f(a))), \dots\}$$

である。意味の木は図 A 5 のようになる。意味の木においては根より出発して、エルブラン基底の各要素に真理値を割り当てていく。真理値 **T** を与えたときには節点から出る左の枝を、また **F** を与えたときには右の枝をたどる。たとえば、根のすぐ下の枝において左側は $P(a)$ に **T** を与えたことを、また右側の $\sim P(a)$ は $P(a)$ に **F** を与えたことを示している。根節点から最終節点 (葉) までの経路をたどることは一つの解釈に対応する。たとえば、図 A 5 において根節点 r から葉 h_1 に至る経路は

$$M_1 = \{P(a), \sim Q(a), \sim P(f(a)), Q(f(a))\}$$

なる解釈に対応する。この解釈は $\sim Q(f(y))$ を偽とするため、 S を充足しない。与えられた節集合内のどれかの節を充足しない解釈に対応する節点はそれ以上展開しても無意味であるので葉節点とし、□で示してある。これを破綻節点という。この図の意味の木は□で閉じている。与えられた節点集合が充足不可能であれば、たとえエルブラン基底が無限であっても、意味の木は必ず閉じる。すなわち、すべての枝は無限に続くことはなく、必ずどこかで□で示された破綻節点で止まる。なぜなら、もしある解釈が破綻節点でカットオフされないなら、この経路を永久にたどる

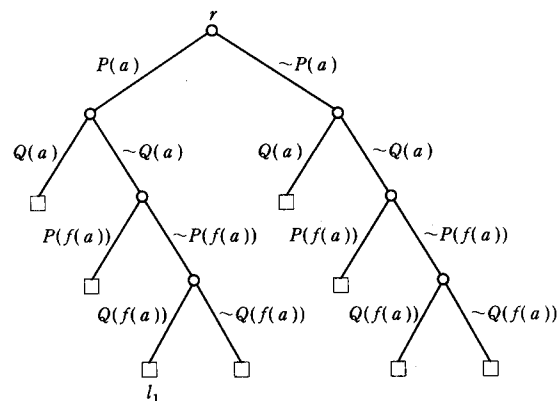


図 A.5 節集合 $S = \{P(x) \vee Q(x), \sim P(f(y)), \sim Q(f(y))\}$ の意味の木。

ことができ、与えられた節集合を充足する解釈が存在することになるからである。逆に、与えられた節集合が充足可能ならば、一般に有限時間内ではアルゴリズムは停止しない。

実際には意味の木を使って充足不可能性を調べることはあまり効率がよいとはいえない。意味の木はいわば、網羅的に充足不可能性をチェックする方式であるので、効率に限界があるのである。そこで、節集合から論理的な操作により直接空節を導く導出原理に基づく方法が、より効率的な方法として使われている。

文 献

本書を書くにあたって以下の図書を参考にさせていただいた。項目別、出版年代順に記す。

1. 情報理論関係(1章~8章)

現在の情報理論の出発点となったのは、

- [1] C. E. Shannon: A mathematical theory of communication, Bell System Technical Journal, vol. 27, pp. 379-423, pp. 623-656, 1948.

である。1949年にはウィーバの解説をつけて単行本として出版された。その和訳が

- [2] C. E. Shannon, W. Weaver (長谷川淳, 井上光洋 共訳): コミュニケーションの数学的理論, 明治図書, 1969.

として出版されている。

- [3] 細野敏夫: 情報工学の基礎, コロナ社, 1967.
 [4] 田中幸吉: 情報工学, 朝倉書店, 1969.
 [5] 藤田広一: 基礎情報理論, 昭晃堂, 1969.
 [6] N. Abramson (宮川洋 訳): 情報理論入門, 好学社, 1969.
 [7] 福村晃夫: 情報理論, コロナ社, 1970.
 [8] 小沢一雅: 情報理論の基礎, 国民科学社, 1980.
 [9] 橋本清: 情報・符号理論入門, 森北出版, 1984.
 [10] 今井秀樹: 情報理論, 昭晃堂, 1984.
 [11] 笠原正雄, 田崎三郎, 小倉久直: 情報理論—基礎と応用—, 昭晃堂, 1985.
 [12] 嵩忠雄: 情報と符号の理論入門, 昭晃堂, 1989.

2. 統計的パターン認識関係(9章)

- [13] 田村進一：データ伝送と計算機ネットワーク，綜文館，1979.

3. 命題論理，述語論理関係(10, 11章)

- [14] Nils J. Nilsson: "Problem-Solving Methods in Artificial Intelligence", McGraw-Hill Book Co Inc., 1971(合田周平，増田一比古 共訳：人工知能—問題解決のシステム論—，コロナ社，1973).
- [15] 長尾真，淵一博：理論と意味，岩波書店，1983.
- [16] 田村進一，柳原圭雄，唐沢博：人工知能の世界，技術評論社，1985.
- [17] 野口正一，滝沢誠：知識工学基礎論，オーム社，1986.
- [18] 上野晴樹，石塚満 編：知識の表現と利用，オーム社，1987.

4. ニューラルネットワーク関係(12章)

- [19] McClelland JL, Rumelhart DE, *et al* : Parallel distributed processing Vol. 1, 2, MIT Press, Cambridge, MA, 1986(甘利俊一 監訳：PDP モデル—認知科学とニューロン回路網の探索，産業図書，1990).
- [20] 麻生英樹：ニューラルネットワーク情報処理，産業図書，1988.
- [21] J. L. McClelland, D. E. Rumelhart: Explorations in parallel distributed processing, MIT Press, Cambridge, MA, 1988. (ソフト付き)

演習問題解答

2章

2.1 求める割合は

$$P(\text{自宅通学} | \text{クラブ活動}) = P(\text{自宅通学}) \frac{P(\text{クラブ活動} | \text{自宅通学})}{P(\text{クラブ活動})}$$

$$= \frac{(2/3)(3/4)}{(2/3)(3/4) + (1/3)(1/2)} = \frac{3}{4}$$

$$2.2 \quad H(1/6) = -(1/6)\log_2(1/6) - (5/6)\log_2(5/6)$$

$$= (1/6)(\log_2 2 + \log_2 3) + (5/6)(\log_2 2 + \log_2 3 - \log_2 5)$$

$$= (1/6) \times 2.58 + (5/6) \times 0.26$$

$$= 0.647$$

3章

$$3.1 \quad (a)-(2), (b)-(3), (c)-(1), (d)-(4)$$

3.2 定常分布を (x, y, z) とする。

$$(x, y, z) = (x, y, z) T$$

より，

$$x = 0.6x + 0.3y + 0.2z$$

$$y = 0.4x$$

$$z = 0.7y + 0.8z$$

これらより，

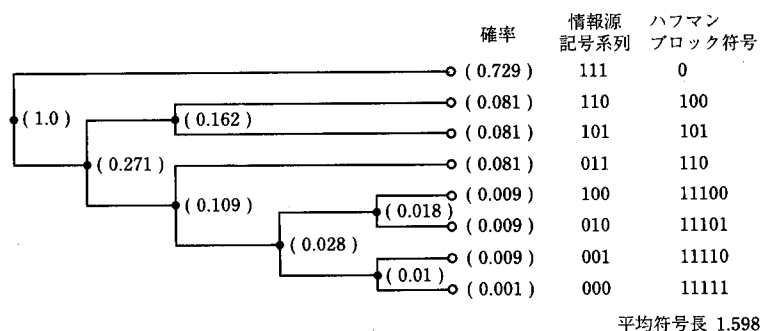
$$x = \frac{5}{14}, \quad y = \frac{1}{7}, \quad z = \frac{1}{2}$$

したがって，定常状態分布は

$$\left(\frac{5}{14}, \frac{1}{7}, \frac{1}{2} \right).$$

4章

4.1 この1情報源記号当たりの平均符号長は0.532(=1.598/3)である。



4.2 定常分布 (w_0, w_1) は

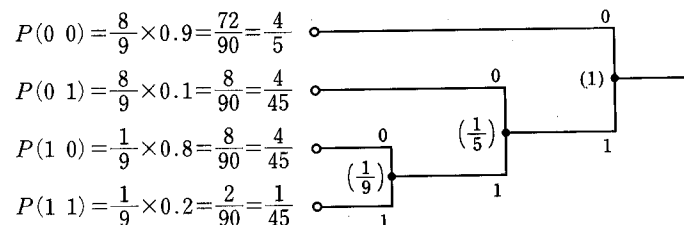
$$(w_0, w_1) \begin{bmatrix} 0.9 & 0.1 \\ 0.8 & 0.2 \end{bmatrix} = (w_0, w_1) \quad \text{および} \quad w_0 + w_1 = 1$$

より,

$$w_0 = \frac{8}{9}, \quad w_1 = \frac{1}{9}.$$

(1) $n=1$ のとき, 明らかに $L(1)=1$.

$n=2$ のとき



より,

$$\begin{aligned} 2L(2) &= 3 \times \frac{1}{45} + 3 \times \frac{4}{45} + 2 \times \frac{4}{45} + 1 \times \frac{36}{45} \\ &= \frac{3+12+8+36}{45} = \frac{59}{45} \end{aligned}$$

したがって,

$$L(2) = \frac{59}{90} = 0.656$$

(2) 状態 s_0 および s_1 にあるときのエントロピーは, 付表1より, それぞれ,

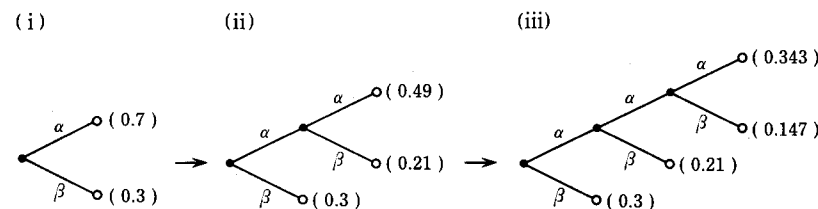
$$\begin{aligned} H_{s_0}(S) &= H(0.1) = 0.469, \\ H_{s_1}(S) &= H(0.2) = 0.722. \end{aligned}$$

これらより, 定常分布 (w_0, w_1) を用いて, 情報源のエントロピーは,

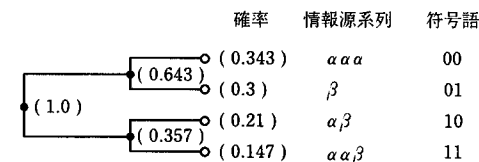
$$H(S) = w_0 H_{s_0}(S) + w_1 H_{s_1}(S) = 0.497.$$

n が大きくなったとき, 情報源符号化定理より $L(n)$ はこの $H(S)=0.497$ に収束する。

4.3



(iii) に対しては



平均情報源記号系列長は

$$\bar{n} = 3 \times 0.343 + 3 \times 0.147 + 2 \times 0.21 + 1 \times 0.3 = 2.19.$$

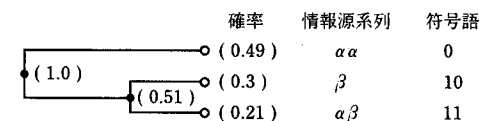
平均符号長は

$$L_{\bar{n}} = 2 \times 0.343 + 2 \times 0.3 + 2 \times 0.21 + 2 \times 0.147 = 2,$$

したがって, 1 情報源記号当たりの平均符号長は

$$L = \frac{2}{2.19} = 0.913.$$

一方, (ii) に対しては



このとき,

$$\bar{n} = 2 \times 0.49 + 2 \times 0.21 + 1 \times 0.3 = 1.7,$$

$$L_{\bar{n}} = 1 \times 0.49 + 2 \times 0.3 + 2 \times 0.21 = 1.51,$$

$$L = \frac{1.51}{1.7} = 0.888.$$

したがって, この例では (iii) より (ii) のほうが平均符号長の短い符号が得られている。

5章

5.1 (1) 式(5.10)より,

$$I(A; B) = H(A) - H(A|B) = \sum_{i=1}^r \sum_{j=1}^s P(a_i, b_j) \log \frac{P(a_i, b_j)}{P(a_i)P(b_j)}$$

ここで, A と B を入れ換えると, 同様に

$$\begin{aligned} H(B) - H(B|A) &= \sum_{i=1}^r \sum_{j=1}^s P(a_i, b_j) \log \frac{P(a_i, b_j)}{P(a_i)P(b_j)} \\ &= H(A) - H(A|B). \end{aligned}$$

(2) $H(A) + H(B) - H(A, B)$

$$\begin{aligned} &= -\sum_{i=1}^r P(a_i) \log P(a_i) - \sum_{j=1}^s P(b_j) \log P(b_j) \\ &\quad - \sum_{i=1}^r \sum_{j=1}^s P(a_i, b_j) \log P(a_i, b_j) \\ &= -\sum_{i=1}^r P(a_i) \log P(a_i) - \sum_{i=1}^r \sum_{j=1}^s P(a_i, b_j) \log P(b_j) \\ &\quad + \sum_{i=1}^r \sum_{j=1}^s P(a_i, b_j) \log P(a_i, b_j) \\ &= \sum_{i=1}^r \sum_{j=1}^s P(a_i, b_j) [-\log P(a_i) - \log P(b_j) + \log P(a_i, b_j)] \\ &= \sum_{i=1}^r \sum_{j=1}^s P(a_i, b_j) \log \frac{P(a_i, b_j)}{P(a_i)P(b_j)} \\ &= H(A) - H(A|B) \end{aligned}$$

(3) (2)より

$$\begin{aligned} H(A) - H(A|B) &= \sum_{i=1}^r \sum_{j=1}^s P(a_i, b_j) \log \frac{P(a_i, b_j)}{P(a_i)P(b_j)} \\ &= \sum_{i=1}^r \sum_{j=1}^s P(a_i, b_j) \log P(a_i, b_j) \\ &\quad - \sum_{i=1}^r \sum_{j=1}^s P(a_i, b_j) \log [P(a_i)P(b_j)] \end{aligned}$$

この式に対しては, シャノンの補助定理が適用でき, その結果

$$H(A) - H(A|B) \geq 0.$$

5.2

$$\begin{aligned} H(\alpha) &= -\frac{m}{m+n} \log \left(\frac{m}{m+n} \right) - \frac{n}{m+n} \log \left(\frac{n}{m+n} \right) \\ &= H\left(\frac{m}{m+n}\right) \\ &= H\left(\frac{3}{4}\right) = H\left(\frac{1}{4}\right) = 0.811 \end{aligned}$$

一方, β が白玉である確率は

$$\begin{aligned} P(\beta = \text{白}) &= \sum_{\alpha} P(\alpha) P(\beta = \text{白} | \alpha) \\ &= P(\alpha = \text{白}) P(\beta = \text{白} | \alpha = \text{白}) + P(\alpha = \text{黒}) P(\beta = \text{白} | \alpha = \text{黒}) \\ &= \left(\frac{m}{m+n}\right) \left(\frac{m-1}{m+n-1}\right) + \left(\frac{n}{m+n}\right) \left(\frac{m}{m+n-1}\right) \end{aligned}$$

$$= \frac{m}{m+n}$$

これは $P(\alpha = \text{白})$ に等しい。このことは, たとえていうと, 「くじに当たる確率は最初に引く人も2回目に引く人も事前には同じである」ことを表わしている。したがって, β のエントロピーは

$$H(\beta) = H(\alpha) = 0.811.$$

また, α を知った後の β の条件付き確率(事後確率)は, それぞれの事象から

$$\begin{aligned} H(\beta | \alpha) &= -P(\beta = \text{白}, \alpha = \text{白}) \log P(\beta = \text{白} | \alpha = \text{白}) \\ &\quad - P(\beta = \text{黒}, \alpha = \text{白}) \log P(\beta = \text{黒} | \alpha = \text{白}) \\ &\quad - P(\beta = \text{白}, \alpha = \text{黒}) \log P(\beta = \text{白} | \alpha = \text{黒}) \\ &\quad - P(\beta = \text{黒}, \alpha = \text{黒}) \log P(\beta = \text{黒} | \alpha = \text{黒}) \\ &= P(\alpha = \text{白}) H(\beta | \alpha = \text{白}) + P(\alpha = \text{黒}) H(\beta | \alpha = \text{黒}) \\ &= \left(\frac{m}{m+n}\right) H\left(\frac{m-1}{m+n-1}\right) + \left(\frac{n}{m+n}\right) H\left(\frac{n-1}{m+n-1}\right) \\ &= \frac{3}{4} H\left(\frac{2}{3}\right) + \frac{1}{4} H(0) \\ &= \frac{3}{4} H\left(\frac{1}{3}\right) = 0.75 \times 0.917 = 0.688. \end{aligned}$$

いま,

$$I(\alpha; \beta) = H(\alpha) - H(\alpha | \beta) = H(\beta) - H(\beta | \alpha)$$

において, $H(\alpha) = H(\beta)$ のとき $H(\alpha | \beta) = H(\beta | \alpha)$ であるから, β を知った後の α の条件付き確率はやはり

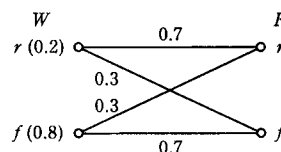
$$H(\alpha | \beta) = H(\beta | \alpha) = 0.688.$$

このことは, やはり「くじに当たる確率は引く順番によらない」ことを表わしている。

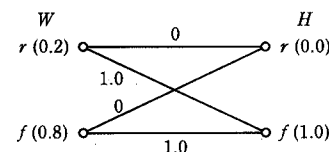
5.3 (1) 天気予報について図1(a)のような通信路を考える。

$$\begin{aligned} I(W; F) &= H(F) - H(F|W) \\ &= H(0.2 \times 0.7 + 0.8 \times 0.3) - [0.2 \times H(0.3) + 0.8 \times H(0.7)] \\ &= H(0.38) - H(0.3) \\ &= 0.958 - 0.881 = 0.077 \end{aligned}$$

(適中率が80%, および90%に改善されれば, この値はそれぞれ0.182, および0.358に改善される。なお, 100%適中するとすれば, この値は0.722である。)



(a) 天気予報(F)の通信路



(b) 毎日晴を予想する人(H)の通信路

図 1

(2) 毎日晴を予想する人の通信路は図1(b)に示されるが、相互情報量 $I(W; H)$ は明らかに0である。すなわち、この人の予想を聞いても何も情報が伝わってこないのである。これは天気予報の適中率が50%になった場合(このとき、 $I(W; F)=0$)と同じである。

5.4 バーストを1つ取り出したとき、その長さが l である確率を $P_B(l)$ とする。これは

$$P_B(l) = (1-q)^{l-1}q$$

で与えられる。このとき、バーストの平均長は

$$\bar{l} = \sum_{l=1}^{\infty} l P_B(l) = \sum_{l=1}^{\infty} l (1-q)^{l-1} q = \frac{1}{q} = 5$$

より、 $q = \frac{1}{5}$ 。

一方、状態 s_0, s_1 の定常確率分布を (w_0, w_1) とすると、

$$w_0 + w_1 = 1 \quad \text{および} \quad (w_0, w_1) \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix} = (w_0, w_1)$$

より

$$w_0 = \frac{q}{p+q}, \quad w_1 = \frac{p}{p+q}.$$

誤り源の状態が s_1 になるとき誤るから、誤り率は

$$w_0 p + w_1 (1-q) = 10^{-1},$$

すなわち、

$$\frac{pq}{p+q} + \frac{p(1-q)}{p+q} = 10^{-1}.$$

このとき

$$p = \frac{1}{45}, \quad w_0 = \frac{9}{10}, \quad w_1 = \frac{1}{10}.$$

誤り源のエントロピーは

$$\begin{aligned} H(E) &= w_0 H\left(\frac{1}{45}\right) + w_1 H\left(\frac{1}{5}\right) \\ &= 0.9 \times 0.154 + 0.1 \times 0.722 = 0.211. \end{aligned}$$

誤り源は入力 X と相関がない。また入力がランダムに確率 $1/2$ ずつで0および1の値をとれば、通信路の出力 Y もそのようになり、そのとき $H(Y)$ は最大値1をとることができる。すなわち、通信路容量は、

$$\begin{aligned} C &= \max[H(Y) - H(Y|X)] \\ &= \max[H(Y) - H(X \oplus E|X)] \\ &= \max[H(Y) - H(E)] \\ &= \max H(Y) - H(E) \\ &= 1 - 0.211 = 0.789. \end{aligned}$$

6章

6.1 (1) 情報ビットの多項式表現は

$$D(x) = x^2 + x.$$

このとき

$$x^3 D(x) = (x^2 + x + 1) G(x) + 1$$

すなわち、余りは1であるから、符号語は

$$0110001$$

(2) 情報ビット数は2、検査ビット数は3である。このとき、表1となる。したがって、ハミング距離の最も近い11101(*印)が復号される。

表1

情報ビット	多項式表現 $Q(x)$	$Q(x)G(x)$	符号語
0 0	0	0	0 0 0 0 0
0 1	1	$x^3 + x + 1$	0 1 0 1 1
1 0	x	$x^4 + x^2 + x$	1 0 1 1 0
1 1	$x + 1$	$x^4 + x^3 + x^2 + 1$	1 1 1 0 1*

6.2 (1) すべての0の w_3 を除く符号語のうち、最低次数の多項式に対応するのは w_5 であり、その符号多項式は生成多項式 $G(x)$ に1を掛けたものである。したがって、生成多項式は

$$G(x) = x^4 + x + 1$$

である。

(2) 最小重みは3である。したがって、2ビットまでの誤りであれば必ず検出できる。

(3) 情報ビット(先頭の3ビット)の重みが1の符号語 w_5, w_7, w_8 が生成行列の各行を表わしているから、生成行列は

$$G = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 \end{bmatrix}$$

である。ただし、この線形符号は、 $m=4$ であるが、符号長 $n=7 < 2^m - 1$ であるから、符号長はハミング符号より短い。

6.3 (1) 検査行列の求め方

x^i を $g(x)$ で割ったときの余りを $R_i(x)$ とすると、

$$\begin{aligned}
R_0(x) &= 1 \\
R_1(x) &= x \\
R_2(x) &= x^2 \\
R_3(x) &= x^3 \\
R_4(x) &= x+1 \\
R_5(x) &= x^2+x \\
R_6(x) &= x^3+x^2 \\
R_7(x) &= x^3+x+1 \\
R_8(x) &= x^2+1 \\
R_9(x) &= x^3+x \\
R_{10}(x) &= x^2+x+1 \\
R_{11}(x) &= x^3+x^2+x \\
R_{12}(x) &= x^3+x^2+x+1 \\
R_{13}(x) &= x^3+x^2+1 \\
R_{14}(x) &= x^3+1
\end{aligned}$$

したがって、検査行列 H は、

$$H = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}$$

また、生成行列 G は

$$G = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 \end{bmatrix}$$

(2) 符号語 w は情報ビット列 $x=(01101000111)$ を用いて

$$w = xG = (011010001111100)$$

(3) シンドロームは $yH^T = (0110)$ である。これは検査行列 H の 10 列目のパターンに等しい。したがって、 y の 10 ビット目が誤っており、正しい符号語は

$$w = (101101101101101)$$

である。

6.4 十分性: $n=ap$ ($a=1, 2, \dots$) とする。

一般に、 $x^t-1=(x-1)(x^{t-1}+x^{t-2}+\dots+1)$ ($t=1, 2, \dots$) であるから、

$$x^n-1=(x^p)^a-1=(x^p-1)((x^p)^{a-1}+(x^p)^{a-2}+\dots+1).$$

すなわち、 $(x^p-1)|(x^n-1)$ 。一方、周期の定義より、 $G(x)|(x^p-1)$ 。したがって、 $G(x)|(x^n-1)$ 。

必要性: $n=ap+b$ ($a=1, 2, \dots; b=1, 2, \dots, p-1$) とする。

$$x^n-1=x^{ap+b}-1=x^b(x^{ap}-1)+x^b-1$$

であり、 $(x^p-1)|(x^{ap}-1)$ 、かつ $G(x)|(x^p-1)$ であるから、 $G(x)|(x^n-1)$ であるためには $G(x)|(x^b-1)$ でなければならない。ここで、 $G(x)$ の周期の定義より $b=0$ でなければならない。

6.5 いま、ある n について、 $G(x)|(x^n-1)$ とする。このとき、 $G_1(x), G_2(x), \dots, G_s(x)$ が互いに素であるから、ある商多項式 $Q(x)$ について、 $(x^n-1)=Q(x)G_1(x)G_2(x)\dots G_s(x)$ であることが必要十分である。すなわち、問 6.4 の結果より、すべての $i=1, 2, \dots, s$ について、 n は p_i の倍数であることが必要十分である。周期 p は、このような性質をもつ n のうち最小のものであるから、 $p=\text{LCM}(p_1, p_2, \dots, p_s)$ である。

7 章

7.1 標本値の数 $2TW=2 \cdot 10 \cdot 5k=100k$ [点] ,

各標本値は 8 ビットで表わされるから、答は $800k$ [ビット]。

7.2 フィルタ出力 $= \sum_k \sin\left(2\pi W \frac{k}{(3/2)W}\right) \frac{\sin(2\pi f_s t - k\pi)}{2\pi f_s t - k\pi}$

$$\begin{aligned}
&= \sum_k \sin\left(\frac{4}{3}\pi k\right) \frac{\sin(2\pi f_s t - k\pi)}{2\pi f_s t - k\pi} \\
&= -\sum_k \sin\left(\frac{2}{3}\pi k\right) \frac{\sin(2\pi f_s t - k\pi)}{2\pi f_s t - k\pi} \quad \dots\dots ①
\end{aligned}$$

一方、 $\sin(\pi Wt)$ は $1/2f_s$ [秒] ($<1/W$) ごとに標本化すれば標本化定理の条件を満たす。このとき

$$\begin{aligned}
\sin(\pi Wt) &= \sum_k \sin\left(\pi W \frac{k}{2f_s}\right) \frac{\sin(2\pi f_s t - k\pi)}{2\pi f_s t - k\pi} \\
&= \sum_k \sin\left(\frac{2}{3}\pi k\right) \frac{\sin(2\pi f_s t - k\pi)}{2\pi f_s t - k\pi} \quad \dots\dots ②
\end{aligned}$$

①と②より、

$$\text{フィルタ出力} = -\sin(\pi Wt).$$

[説明] 一般に折返しがあると歪みとなる(図 2)。

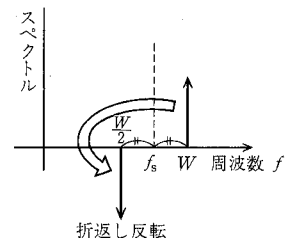


図 2

7.3 展開区間長は T ではなくて $3T$ であるから(図 3)、式(7.19)より

$$g(t) = \sum_{k=-\infty}^{\infty} a_k e^{j2\pi kt/3T}.$$

ここに、展開係数は式(7.21)より

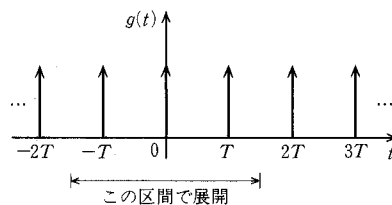


図 3

$$\begin{aligned}
 a_k &= \frac{1}{3T} \int_{-\frac{3}{2}T}^{\frac{3}{2}T} g(t) e^{-j2\pi kt/3T} dt \\
 &= \frac{1}{3T} \int_{-\frac{3}{2}T}^{\frac{3}{2}T} [\delta(t+T) + \delta(t) + \delta(t-T)] e^{-j2\pi kt/3T} dt \\
 &= \frac{1}{3T} (e^{j2\pi k/3} + 1 + e^{-j2\pi k/3}) \\
 &= \frac{1}{3T} \left(1 + 2\cos \frac{2}{3}\pi k \right).
 \end{aligned}$$

したがって,

$$\begin{cases} a_k = a_{3m} = \frac{1}{T} & (k=3m \text{ (} m=0, \pm 1, \pm 2, \dots \text{) のとき}) \\ a_k = 0 & (\text{その他}). \end{cases}$$

7.4 $g(t-t_0)$ のスペクトルは

$$\begin{aligned}
 \int_{-\infty}^{\infty} g(t-t_0) e^{-j2\pi ft} dt &= \int_{-\infty}^{\infty} g(t-t_0) e^{-j2\pi f(t-t_0)} e^{-j2\pi ft_0} dt \\
 &= e^{-j2\pi ft_0} G(f).
 \end{aligned}$$

したがって, $g(t-t_0)$ のパワースペクトルは

$$|e^{-j2\pi ft_0} \cdot G(f)|^2 = |G(f)|^2.$$

7.5 出力スペクトルは

$$Y(f) = H(f) X(f).$$

したがって, 出力のパワースペクトルは

$$|Y(f)|^2 = |H(f)|^2 |X(f)|^2.$$

7.6 $G(f) = \int_{-\infty}^{\infty} g(t) e^{-j2\pi ft} dt$ より,

$$\begin{aligned}
 |G(f)|^2 &= G(f) G^*(f) \\
 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(t) e^{-j2\pi ft} g^*(t') e^{j2\pi ft'} dt' dt \\
 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(t) g^*(t') e^{-j2\pi f(t-t')} dt' dt.
 \end{aligned}$$

したがって,

$$\begin{aligned}
 \int_{-\infty}^{\infty} |G(f)|^2 df &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(t) g^*(t') e^{-j2\pi f(t-t')} df dt' dt \\
 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(t) g^*(t') \delta(t-t') dt' dt
 \end{aligned}$$

$$\begin{aligned}
 &= \int_{-\infty}^{\infty} g(t) g^*(t) dt \\
 &= \int_{-\infty}^{\infty} |g(t)|^2 dt.
 \end{aligned}$$

8 章

$$\begin{aligned}
 8.1 \quad p(z) &= \int_{-\infty}^{\infty} p(z, x) dx \\
 &= \int_{-\infty}^{\infty} p(z|x) p(x) dx \\
 &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma_y^2}} e^{-\frac{(z-x)^2}{2\sigma_y^2}} \frac{1}{\sqrt{2\pi\sigma_x^2}} e^{-\frac{x^2}{2\sigma_x^2}} dx \\
 &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi \frac{\sigma_x^2 \sigma_y^2}{\sigma_x^2 + \sigma_y^2} 2\pi(\sigma_x^2 + \sigma_y^2)}} e^{-\left[\frac{(x-\sigma_x^2 z / (\sigma_x^2 + \sigma_y^2))^2}{2\sigma_x^2 \sigma_y^2 / (\sigma_x^2 + \sigma_y^2)} + \frac{z^2}{2(\sigma_x^2 + \sigma_y^2)} \right]} dx \\
 &= \frac{1}{\sqrt{2\pi(\sigma_x^2 + \sigma_y^2)}} e^{-\frac{z^2}{2(\sigma_x^2 + \sigma_y^2)}}
 \end{aligned}$$

8.2 (1) この通信路の雑音電力は WN_0 [ワット] である。したがって, シャン・ハートレーの定理よりただちに

$$C = W \log_2 \left(1 + \frac{S}{WN_0} \right) \quad [\text{ビット/秒}].$$

(2) $h = \frac{WN_0}{S}$ とおくと

$$C = \frac{S}{N_0} \frac{WN_0}{S} \log_2 \left(1 + \frac{S}{WN_0} \right) = \frac{S}{N_0} \log_2 \left(1 + \frac{1}{h} \right).$$

したがって,

$$\lim_{W \rightarrow \infty} C = \lim_{h \rightarrow 0} C = \frac{S \log_2 e}{N_0} \quad [\text{ビット/秒}].$$

$$8.3 \quad \frac{S}{N_0} \log_2 e > \frac{S}{N_0} \log_2 \left(1 + \frac{S}{N_0 W} \right)^{\frac{N_0 W}{S}} = C$$

$$> R = \frac{\log_2 M}{T} = \frac{1}{T_B}$$

$$\therefore \frac{ST_B}{N_0} > \log_e 2$$

8.4 $p(x)$ は確率密度であるから,

$$\int_{-x_0}^{x_0} p(x) dx = 1.$$

いま, λ をラグランジュの未定乗数とし, 関数

$$\begin{aligned}
 \phi(X) &= H(X) + \lambda \int_{-x_0}^{x_0} p(x) dx \\
 &= \int_{-x_0}^{x_0} -p(x) \log_e p(x) + \lambda p(x) dx
 \end{aligned}$$

を定義し, $H(X)$ を最大とするため, この被積分関数を $p(x)$ について偏微分して, 0 とおく。このとき,

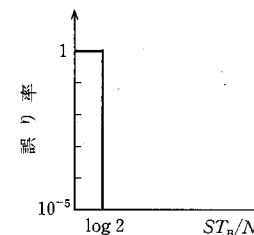


図 4

$$-\log_e p(x) - 1 + \lambda = 0.$$

したがって

$$p(x) = e^{\lambda-1}$$

となる。この右辺は x によらない。すなわち $p(x)$ は一様分布

$$p(x) = \begin{cases} \frac{1}{2x_0} & (-x_0 \leq x \leq x_0) \\ 0 & (\text{その他}) \end{cases}$$

のとき、 $H(X)$ を最大にする。このとき、最大値 $H(X)$ は

$$H(X) = \log_e(2x_0) \quad [\text{ナット}]$$

$$= \log_2(2x_0) \quad [\text{ビット}]$$

である。

9章

9.1 x_1x_2 の期待値は

$$\begin{aligned} \overline{x_1x_2} &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1x_2 p(x_1, x_2) dx_1 dx_2 \\ &= \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{\infty} x_1 \frac{1}{\sqrt{2\pi\sigma^2(1-\rho^2)}} \exp\left[-\frac{(x_1-\rho x_2)^2 + x_2^2(1-\rho^2)}{2\sigma^2(1-\rho^2)}\right] dx_1 \right\} \\ &\quad \cdot \frac{1}{\sqrt{2\pi\sigma^2}} x_2 \cdot dx_2 \end{aligned}$$

$$= \int_{-\infty}^{\infty} \rho x_2^2 \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{x_2^2}{2\sigma^2}\right] dx_2$$

$$= \rho\sigma^2.$$

x_1^2 の期待値は

$$\begin{aligned} \overline{x_1^2} &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1^2 p(x_1, x_2) dx_1 dx_2 \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1^2 \frac{1}{\sqrt{2\pi\sigma^2(1-\rho^2)}} \exp\left[-\frac{(x_1-\rho x_2)^2}{2\sigma^2(1-\rho^2)}\right] dx_1 \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{x_2^2}{2\sigma^2}\right] dx_2 \\ &= \int_{-\infty}^{\infty} [\sigma^2(1-\rho^2) + \rho^2 x_2^2] \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{x_2^2}{2\sigma^2}\right] dx_2 \\ &= \sigma^2(1-\rho^2) + \rho^2\sigma^2 = \sigma^2. \end{aligned}$$

なお、ここではよく知られた公式

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx = 1,$$

$$\int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx = \mu,$$

$$\int_{-\infty}^{\infty} x^2 \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx = \sigma^2 + \mu^2$$

を用いた。

9.2 (1) 式(9.11)より

$$\begin{aligned} \lambda(Y) &= \exp\left[-\frac{1}{2}\{(Y-S)\Sigma^{-1}(Y-S)^T - Y\Sigma^{-1}Y^T\}\right] \\ &= \exp\left[-\frac{1}{2}(Y\Sigma^{-1}Y^T - 2Y\Sigma^{-1}S^T + S\Sigma^{-1}S^T - Y\Sigma^{-1}Y)\right] \\ &= \exp\left(Y\Sigma^{-1}S^T - \frac{1}{2}S\Sigma^{-1}S^T\right) \end{aligned}$$

これと 1 との大小を比較することから、決定規則はつぎのようになる。

$$\begin{cases} Y\Sigma^{-1}S^T \geq \frac{1}{2}S\Sigma^{-1}S^T \text{ ならば, } \hat{\theta}=1, \\ Y\Sigma^{-1}S^T < \frac{1}{2}S\Sigma^{-1}S^T \text{ ならば, } \hat{\theta}=0. \end{cases}$$

$$(2) \Sigma^{-1}S^T = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}, \quad S\Sigma^{-1}S^T = \begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} 3 \\ 2 \end{bmatrix} = 7$$

ここで、 $Y = [y_1 \ y_2]$ とおくと、識別関数は

$$[y_1 \ y_2] \begin{bmatrix} 3 \\ 2 \end{bmatrix} - \frac{7}{2} = 3y_1 + 2y_2 - \frac{7}{2}.$$

これを 0 に等しいとおいた分離超平面(いまの場合は直線)を図示すると、図 5 のようになる。

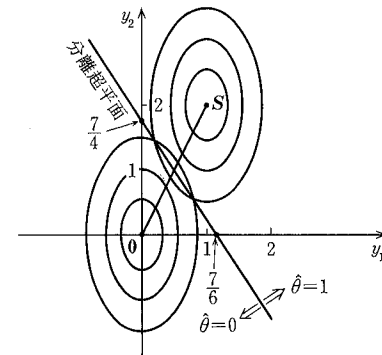


図 5

(3) 式(9.17)~(9.20)において、 $A^T = \Sigma^{-1}S^T$ 、 $\delta = \frac{1}{2}S\Sigma^{-1}S^T$ であるから、

$$\begin{aligned} \bar{z} &= \theta S\Sigma^{-1}S^T, \\ \frac{(\bar{z}-z)^2}{(\bar{z}-\bar{z})^2} &= A\Sigma A^T = S\Sigma^{-1}S^T = a^2. \end{aligned}$$

このとき、S/N 比は

$$R = S\Sigma^{-1}S^T = 7.$$

したがって、誤り率は式(9.29)および付表 3 を用いて、

$$\begin{aligned} P_e &= \Phi\left(-\frac{1}{2}\sqrt{R}\right) = \Phi(-1.323) \\ &= 1 - \Phi(1.323) = 1 - 0.907 = 0.093 \end{aligned}$$

10 章

10.1 真理値表は表2のようにになる。したがって、表の○印の P, Q, R の真理値がモデル(5個)である。

表2 $W=(P \rightarrow Q) \rightarrow R$ の真理値表

P	Q	R	W	モデル
F	F	F	F	×
F	F	T	T	○
F	T	F	F	×
F	T	T	T	○
T	F	F	T	○
T	F	T	T	○
T	T	F	F	×
T	T	T	T	○

- 10.2 (1) 図6より恒真である。 (2) 図7より恒真である。
 (3) 図8より恒真である。 (4) 図9より恒偽である。
 (5) 図10より恒真である。

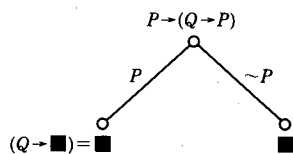


図6

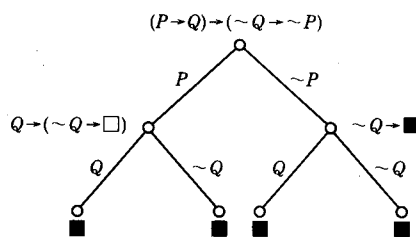


図7

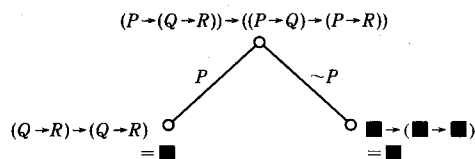


図8

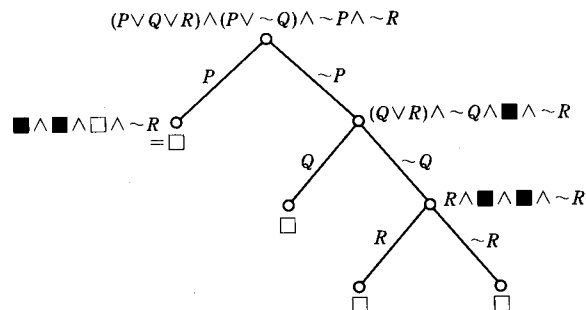


図9

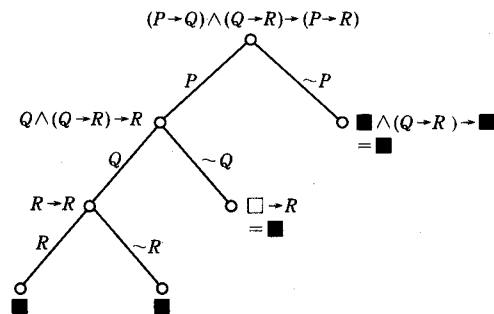


図10

$$\begin{aligned}
 10.3 \quad W_1 &= (\sim P \vee Q) \vee (Q \wedge R) \\
 &= (\sim P \vee Q \vee Q) \wedge (\sim P \vee Q \vee R) \\
 &= (\sim P \vee Q) \wedge (\sim P \vee Q \vee R) \\
 W_2 &= ((\sim P \vee Q) \wedge (P \vee \sim Q)) \vee (\sim Q \wedge P \wedge R) \\
 &= ((\sim P \vee Q) \vee (\sim Q \wedge P \wedge R)) \wedge ((P \vee \sim Q) \vee (\sim Q \wedge P \wedge R)) \\
 &= ((\sim P \vee Q \vee \sim Q) \wedge (\sim P \vee Q \vee P) \wedge (\sim P \vee Q \vee R)) \\
 &\quad \wedge ((P \vee \sim Q) \vee \sim Q) \wedge (P \vee \sim Q \vee P) \wedge (P \vee \sim Q \vee R) \\
 &= \blacksquare \wedge \blacksquare \wedge (\sim P \vee Q \vee R) \wedge (P \vee \sim Q) \wedge (P \vee \sim Q) \wedge (P \vee \sim Q \vee R) \\
 &= (\sim P \vee Q \vee R) \wedge (P \vee \sim Q) \wedge (P \vee \sim Q \vee R)
 \end{aligned}$$

11 章

- 11.1 (1) 置換 $\{x \leftarrow u, f(x, y) \leftarrow v\}$ により $Q(x, f(x, y))$ に単一化される。
 (2) 置換 $\{a \leftarrow x, a \leftarrow u, h(b) \leftarrow y, f(h(b), z) \leftarrow v\}$ により $Q(a, f(h(b), z))$ に単一化される。

11.2 (1) 図 11 (2) 図 12

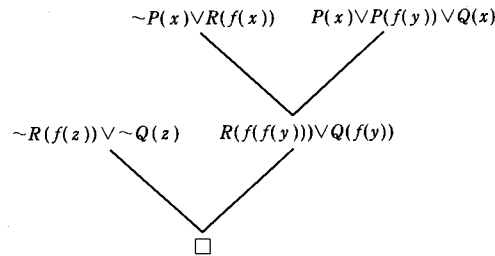


図 11 節集合 $\{P(x) \vee P(f(y)) \vee Q(x), \sim P(x) \vee R(f(x)), \sim R(f(z)) \vee \sim Q(z)\}$ の導出木。

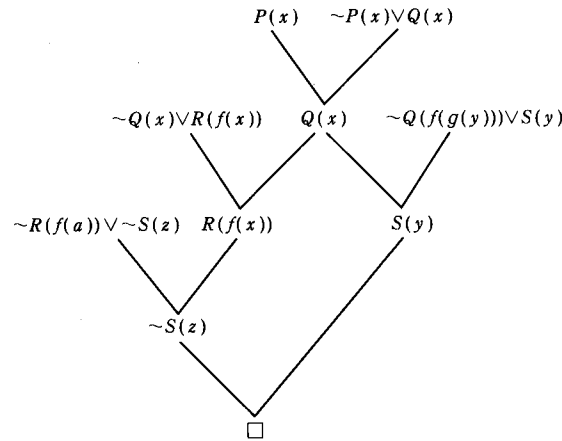


図 12 充足不可能な節集合 $\{P(x), \sim P(x) \vee Q(x), \sim Q(x) \vee R(f(x)), \sim Q(f(g(y))) \vee S(y), \sim R(f(a)) \vee \sim S(z)\}$ の導出木。

12 章

12.1 $E = [t - (xW^T)^k]^2$

最急降下法に従うと重みの修正量は、 η を正定数として

$$\Delta w_i = -\eta \frac{\partial E}{\partial w_i} = 2\eta [t - (xW^T)^k] k(xW^T)^{k-1} \cdot x_i$$

で与えられる。

12.2 (1) ネットワークの状態変化： 例 12.2 と同様に

$$\begin{aligned} s_1 &= (0, 0, 1, 0, 0) \\ s_2 &= (1, 1, 0, 1, 1) \\ s_3 &= (0, 0, 1, 0, 0) \\ s_4 &= (1, 1, 0, 1, 1) \\ s_5 &= (0, 0, 1, 0, 0) \\ s_6 &= (1, 1, 0, 1, 1) \\ &\vdots \end{aligned}$$

したがって、c(動物)とそれ以外が交互に興奮している。ただし後で示すように、このサイクルは不安定であり、素子の反応速度のばらつきなどにより容易に例 12.2 のような安定状態におちつく。

(2) ネットワークのエネルギー変化： この場合のエネルギーは

$$E(1)=0, E(2)=2, E(3)=0, E(4)=2, \dots$$

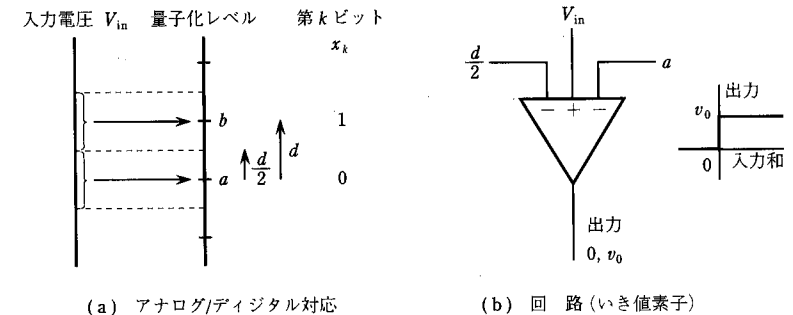
のように振動する。このことは、12.3.1 項で述べたエネルギーの単調減少性に反しているようにみえる。これは、12.3.1 項では非同期的に各素子を独立に動作させているのに対して、この解では同期的に動作させたことに原因がある。

この問題のネットワークは左右対称につくられており、重みも ± 1 の 2 種類しかない。ネットワークに非対称性、重みや素子の応答時間のばらつきなどがあれば、このサイクルから抜け出る。たとえば $t=1$ から $t=2$ へ遷移するとき、素子 e の興奮が他より少し早く起こったとする。そうすると a と d の活性化は抑えられ、b の活性化は加速されるから、e および b が興奮し、 $t=3$ 以降で b, c, e が興奮し、他は抑制された状態となり遅かれ早かれ安定化する。この安定状態のエネルギーは例 12.2(2) と同じく

$$E(t) = -3.0$$

である。

12.3 第 k ビット ($k=0, 1, \dots, n-1$) の変換について考える。他のビットはすべて確定しているとする。このとき、第 k ビットの値 x_k は図 13(a) にしたがって、つぎのように得られる。



(a) アナログ/デジタル対応

(b) 回路 (いき値素子)

図 13 第 k ビットの変換

$$x_k = \begin{cases} 0 & \left(a + \frac{1}{2}d \geq V_{in} \text{ のとき} \right) \\ 1 & \text{(その他)} \end{cases}$$

ここに、 a および b は第 k ビットをそれぞれ 0, 1 としたときの量子化レベル、 d は第 k ビットの量子化ステップ幅であり、つぎで与えられる：

$$a = (x_{n-1} \cdot 2^{n-1} + \dots + 0 \cdot 2^k + \dots + x_0 \cdot 2^0) v_0,$$

$$b = (x_{n-1} \cdot 2^{n-1} + \dots + 1 \cdot 2^k + \dots + x_0 \cdot 2^0) v_0,$$

$$d = 2^k v_0.$$

これを実現する回路は図 13 (b) で与えられる。この回路をそのまま接続すると図 14 のようになる。相互結合型ネットワークの重みは対称である。各素子はいき値素子であるから、入力を正の定数倍しても特性は変らない。そこで、第 k ビットの素子の入力重みを 2^k 倍すると、図 12.11 が得られる。

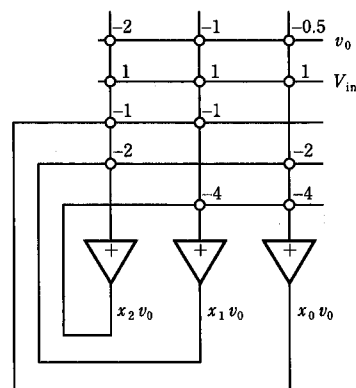


図 14 3ビット A/D 変換器の直接実現法

付 表 1

(a) 2 を底とする対数の表

n	$\log n$	n	$\log n$	n	$\log n$	n	$\log n$
1	0.000000	26	4.700439	51	5.672425	76	6.247927
2	1.000000	27	4.754887	52	5.700439	77	6.266786
3	1.584962	28	4.807355	53	5.727920	78	6.285402
4	2.000000	29	4.857981	54	5.754887	79	6.303780
5	2.321928	30	4.906890	55	5.781359	80	6.321928
6	2.584962	31	4.954196	56	5.807355	81	6.339850
7	2.807355	32	5.000000	57	5.832890	82	6.357552
8	3.000000	33	5.044394	58	5.857981	83	6.375039
9	3.169925	34	5.087463	59	5.882643	84	6.392317
10	3.321928	35	5.129283	60	5.906890	85	6.409391
11	3.459431	36	5.169925	61	5.930737	86	6.426264
12	3.584962	37	5.209453	62	5.954196	87	6.442943
13	3.700440	38	5.247927	63	5.977280	88	6.459431
14	3.807355	39	5.285402	64	6.000000	89	6.475733
15	3.906890	40	5.321928	65	6.022367	90	6.491853
16	4.000000	41	5.357552	66	6.044394	91	6.507794
17	4.087463	42	5.392317	67	6.066089	92	6.523562
18	4.169925	43	5.426264	68	6.087462	93	6.539158
19	4.247927	44	5.459431	69	6.108524	94	6.554583
20	4.321928	45	5.491853	70	6.129283	95	6.569855
21	4.392317	46	5.523562	71	6.149747	96	6.584962
22	4.459431	47	5.554589	72	6.169925	97	6.599912
23	4.523562	48	5.584962	73	6.189824	98	6.614709
24	4.584962	49	5.614710	74	6.209453	99	6.629356
25	4.643856	50	5.643856	75	6.228818	100	6.643856

(b) 2元アルファベットのエントロピーの表

$$H(p) = -p \log_2 p - (1-p) \log_2 (1-p)$$

p	$H(p)$	p	$H(p)$	p	$H(p)$	p	$H(p)$
0.005	0.045415	0.130	0.557438	0.255	0.819107	0.380	0.958042
0.010	0.080793	0.135	0.570993	0.260	0.826746	0.385	0.961497
0.015	0.112364	0.140	0.584239	0.265	0.834198	0.390	0.964800
0.020	0.141441	0.145	0.597185	0.270	0.841465	0.395	0.967951
0.025	0.168661	0.150	0.609840	0.275	0.848548	0.400	0.970951
0.030	0.194392	0.155	0.622213	0.280	0.855451	0.405	0.973800
0.035	0.218878	0.160	0.634310	0.285	0.862175	0.410	0.976550
0.040	0.242292	0.165	0.646138	0.290	0.868721	0.415	0.979051
0.045	0.264765	0.170	0.657705	0.295	0.875093	0.420	0.981454
0.050	0.286397	0.175	0.669016	0.300	0.881291	0.425	0.983708
0.055	0.307268	0.180	0.680077	0.305	0.887317	0.430	0.985815
0.060	0.327445	0.185	0.690894	0.310	0.893173	0.435	0.987775
0.065	0.346981	0.190	0.701471	0.315	0.898861	0.440	0.989588
0.070	0.365924	0.195	0.711815	0.320	0.904381	0.445	0.991254
0.075	0.384312	0.200	0.721928	0.325	0.909736	0.450	0.992774
0.080	0.402179	0.205	0.731816	0.330	0.914925	0.455	0.994149
0.085	0.419556	0.210	0.741483	0.335	0.919953	0.460	0.995378
0.090	0.436470	0.215	0.750932	0.340	0.924819	0.465	0.996462
0.095	0.452943	0.220	0.760167	0.345	0.929523	0.470	0.997402
0.100	0.468996	0.225	0.769193	0.350	0.934068	0.475	0.998196
0.105	0.484648	0.230	0.778011	0.355	0.938454	0.480	0.998846
0.110	0.499916	0.235	0.786626	0.360	0.942683	0.485	0.999351
0.115	0.514816	0.240	0.795040	0.365	0.946755	0.490	0.999711
0.120	0.529361	0.245	0.803257	0.370	0.950672	0.495	0.999928
0.125	0.543564	0.250	0.811278	0.375	0.954434	0.500	1.000000

索引

あ 行

アナログ情報 4, 5
 アナログ情報源 96
 誤り検出 72
 誤り検出能力 75
 誤り検出領域 63
 誤り訂正能力 64
 誤り訂正領域 63
 誤りの検出・訂正 2
 誤りパターン 72
 誤り率 112, 114
 r 元通信路 45
 r 元符号 30
 暗号 4
 いき値素子 145
 ECC 符号 2
 一意的復号可能 26, 28, 40
 1 次エントロピー 39, 41, 99
 位置不変性 95
 パワースペクトルの—— 95
 一様な通信路 48
 出力に関して—— 48
 2 重に—— 48, 51
 入力に関して—— 48
 意味の木 123

意味論 157
 因果律 88
 因子 136
 インパルス応答 88
 エイリアシング 5, 78, 93
 エキスパートシステム 7
 XOR 145
 枝 28
 A/D 変換 5, 78
 A/D 変換器 153
 (n, k) 符号 58
 n 次エントロピー 40, 100
 n 次拡大情報源 39, 40
 エネルギー 147
 MR 符号 2
 m 重マルコフ情報源 18
 LSB 153
 エルゴード情報源 16
 エルゴード性 16, 97
 エルゴード的 16
 エルブラン解釈 161
 エルブラン基底 160
 エルブラン空間 132, 159
 エントロピー 13, 31, 40, 42, 99, 100
 結合—— 50

重み 62
 折り返し歪 5, 95

か行

解釈 118

階層型ネットワーク 8, 142

ガウス分布 97, 107

学習 8, 143

拡大の手法 51

確率 9, 10

活性値 142

過渡状態部分 21

加法性 62
 線形符号の—— 62

加法的白色ガウス通信路 101

含意 120

含意記号
 ——の除去 130

慣性項 145

完全
 関数的に—— 117

完全事象系 12

完全性 159

完全性定理 159

偽(F) 115

記憶がない
 情報源に—— 3

記憶のある情報源 18

記憶のない情報源 15

記憶のない通信路 46

記号論理学 7

奇数パリティ検査 59

基礎例 160

基本命題 115

基本論理式 115, 127

既約多項式 74

既約マルコフ情報源 22

教師付き学習 142

共分散行列 108

距離の三公理 62

偶数パリティ検査 59

空節 132, 133

組合せ問題 8

クラフトの不等式 29, 30

形式的証明 158

結合エントロピー 50

結合確率密度関数 98

結合則 135

決定規則 111

決定不能 129

限界距離復号法 63

検査記号 58

検査ビット 60

原始多項式 75

限定作用素 128

限量記号 128

項 127

高階述語論理 129

恒偽性 124, 132

恒真 118

恒真性 123, 124, 129

構文論的方法 157

公理系
 命題論理の—— 157

効率 53, 58

効率符号の—— 53

誤差逆伝搬学習アルゴリズム 142

誤差伝搬量 144

個体定項 127

個体変項 127

根元事象 9

コンパクト 34, 36

コンパクト符号 34, 37

コンマ符号 27

さ行

最急降下法 145, 153

最小重み 62, 63

最小距離 63, 74

最小ハミング距離 63

最大エントロピー定理 100

最短符号 34

作用範囲 128

三角不等式 62

三段論法 7, 121, 126, 133, 157

サンプリング 78

CRC 68, 72

CRC 符号 2

時間平均 16, 97

シグモイド関数 141

事後エントロピー 49

事後確率 11, 109

自己情報量 12

自己相関関数 90

CCITT 68, 75

事象 9

事前エントロピー 49

事前確率 11

質問応答 139

質問応答システム 7

シャノンの補助定理 32

シャノン・ハートレーの定理 102

シャノン理論 1

周期 71

周期的状態部分 21

集合平均 16, 97

集合論的演算 9

充足可能 118

充足する 118

充足性 129

充足不可能 118, 129, 132, 136

周波数成分 83

自由変項 128

受信記号 45

述語 127

述語論理 7, 127

述語論理式 128, 129

出力アルファベット 45

出力記号 45

出力誤差 143

出力層 142

巡回セールスマン問題 151

巡回ハミング符号 75

巡回符号 68, 69, 71

瞬時符号 27, 28, 29, 31

条件付きエントロピー 49

条件付き確率 109, 10, 11, 97

消失 47

状態 18, 21

状態確率分布 23

状態図 18

状態分布 23

商多項式 70

冗長度 54
 符号の—— 54

情報源 2

情報源アルファベット 15

情報源記号系列
 平均的に長い—— 42

情報源符号化 2, 3, 27

情報源符号化定理 39, 40

情報速度 53

情報伝送速度 53, 56

情報ビット 59

情報量 5, 9, 13

情報理論 1, 9

証明可能 158

剰余多項式 70

真(T) 115
 常に—— 119

真偽 116

神経回路 8

シンドローム 65

振幅スペクトル密度 83

真理値 115, 129

水平垂直パリティ検査符号 60
 水平垂直パリティ検査方式 68
 推 論 7, 115
 推論規則 117, 152
 スコーレム関数 131
 スペクトル 83

 正規分布 97, 107
 正規マルコフ情報源 22, 23
 整合機械 111
 整合フィルタ 111
 生成行列 61
 生成多項式 75
 世 界 129
 積事象 10
 積標準形 131
 節 124, 130, 132
 節 点 28
 接頭語 27
 接頭語条件 27
 遷移確率行列 19
 遷移図 18
 線形識別関数 145
 線形識別方式 112
 線形符号 58, 61
 選 言 124
 選言標準形 124, 125
 全称記号 128
 —の除去 132
 前置形 131

 相関係数 108
 相互結合型ネットワーク 8, 146
 相互情報量 48, 49, 54
 送信記号 45
 測 度 10
 束縛変項 128
 組織符号 58
 存在記号 128
 —の除去 131

た 行
 帯域制限 96
 第一階述語論理 129
 代 入 132
 多項式 68
 畳み込み積分 89
 wff 117, 127
 単位インパルス関数 86
 単一誤り 72
 単一誤り訂正ハミング符号 66
 単一化 133, 134
 単一化アルゴリズム 135
 単一化置換
 最も一般的な—— 134
 単一パリティ検査符号 59
 探 索 137
 単純マルコフ情報源 18

 置 換 134
 —の合成 134
 知的符号化 4
 中間層 142
 超関数 87
 超平面 112
 直線量子化 103
 直交関数展開 86

 通信系モデル 2, 109
 通信路 45
 通信路行列 46, 48
 通信路線図 47
 通信路符号化 2, 3, 25
 通信路符号化定理 56
 通信路モデル 45
 通信路容量 102, 51, 56

 低域通過フィルタ 93
 定常状態 23
 定常情報源 15, 97
 定常通信路 46

デイビス・パットナムのアルゴリズム 155
 デイビス・パットナムの方法 124, 133
 定 理 158
 適応的符号化 4
 δ 関数 86
 典型的系列 54
 伝達関数 89

 統計的パターン認識 6
 導 出 133
 導出原理 7, 130, 132, 137
 導出節 133, 156
 同 値 120
 等長符号 26
 特徴ベクトル 107
 閉じた状態部分 21
 閉じている 62
 加法に関して—— 62
 ド・モルガンの法則 125

 な 行
 ナット 13

 2元対称消失通信路 52
 2元対称通信路 47, 52
 2次元エントロピー 41
 2次拡大情報源 38
 二重否定 125
 ニューラルネットワーク 8, 141
 —の形態 141
 入力アルファベット 45
 入力記号 45
 入力層 142

 根 28

 は 行
 葉 28
 バースト誤り 67, 73, 74
 バースト誤り訂正(検出)能力 67

パーセバルの定理 95
 倍多項式 69
 排他的論理和 59, 68, 145
 白色ガウス情報源 98
 パターン認識 6, 8, 107, 109, 142
 パターンの記憶 150
 破綻節点 161
 幅優先探索 138
 ハフマン符号 34, 42
 ハフマンブロック符号 39, 43
 ハミング重み 62
 ハミング距離 61
 ハミング符号 66
 パラメータ表現
 命題論理式の—— 123
 パリティ検査 73
 パリティ検査行列 64
 パリティ検査ビット 25
 パリティ検査方程式 59, 64
 パワースペクトル 90

 BSC 47
 引 数 127
 BCH 符号 77
 非周期的状態部分 22
 非瞬時符号 27
 非直線量子化 103
 ビット 13
 否定記号
 —の個別化 130
 非等長情報源記号系列 42
 非等長符号 26
 BPN 143
 標準ガウス分布関数 112
 標準形 124
 命題論理式の—— 124
 標本化 4, 78
 標本化関数 85
 標本化定理 4, 78, 85
 標本化パルス 90

田 村 進 一 略歴
た むら しん いち

- 1944 年 兵庫県に生まれる
1966 年 大阪大学基礎工学部電気工学科
卒業
1971 年 大阪大学大学院基礎工学研究科
物理系博士課程修了 (工学博士)
大阪大学基礎工学部情報工学科
助手, 助教授を経て
1988 年 大阪大学医学部附属バイオメデ
ィカル教育研究センター教授

© 田村進一 1992

1992 年 1 月 15 日 初 版 発 行
1999 年 1 月 30 日 初版第 4 刷発行

電子・情報工学講座 17
情報工学基礎論

著 者 田村進一
発行者 山 本 格

発 行 所 株 式 会 社 培 風 館
東京都千代田区九段南4-3-12・郵便番号102-8260
電話 (03) 3262-5256 (代表)・振替 00140-7-44725

中央印刷・三水舎製本

PRINTED IN JAPAN

ISBN4-563-03347-2 C3355