

情報理論入門

川端

平成32年5月7日

第1章 はじめに

1.1 デジタル通信システムと情報理論の概要

通信の文字通りの意味は、信（真実、約束、誠実さ）を通じ（伝達）させることである。信は、音・言葉・ジェスチャー・或るいは何かの気などのなんらかの手段で伝達される。この意味では動物同士の間でも立派に通信が成り立っている。しかし通信を高度化したのは人間であろう。例えば文字などない時代から敵の来襲を味方に伝えるのにのろしを用いていたがこれは符号を用いた遠隔通信である。文字が発明されると、手紙や書物として豊富なメッセージの通信が可能となった。さらに時代が下り近現代になると、電気信号をメッセージの媒体として用いることができるようになり、有線通信や無線通信のシステムにより豊富さに加えて高速な通信が可能となったことはいうまでもない。

さて、もっとも簡単な通信のモデルはいかなるものであろうか。通信は二者の間で行われる。通信は双方向で行われることが多いが通信の本質を考えるためには、方向を分離しもっとも簡単な一方向の場合だけを考えればよい。通信にとって、**情報源 (source)** すなわち通信の対象であるメッセージを出力する源、および**受信者 (destination)** すなわち信号を受け取る者（所）、この2つは最低限不可欠である。一方、情報源の出力メッセージが直接受信者に伝わる場合は自明な通信というべきであり、特になにも考える必要がない。自明でない通信では、情報源と受信者の間に**通信路 (channel)** すなわち入出力をもつ物理媒体、が存在していることが本質である。つまり情報源、通信路、受信者の3者が通信に不可欠な基本要素であるということになる。また通信の目的は情報源の信号に含まれるメッセージを通信路を介して受信者に伝えることであると規定することができる。

さて、通信の標準的な手段である電気信号は音波や電磁界をセンシングして得られる連続時間連続値信号（アナログ信号）である。この抽象化により上述の3者にはいずれもシステム論的な定義をすることができる。すなわち、情報源とは信号を出力するシステムであり、通信路とは入力信号と出力信号をもつシステムであり、受信者は、対応する destination（目的者）という言葉からもわかるとおり、入力信号だけをもつシステムである。

さて通信の3要素に加えて、通信を可能とする仕組みである**送信機 (transmitter)** と**受信機 (receiver)** が必要となる。送信機とは情報源出力信号を

通信路入力信号へ変換するシステムであり、受信機とは通信路出力信号を受信者への入力信号に変換するシステムである。

以上の5者が通信のモデルを構成する。以下ではさらに、**情報通信**の基本モデルであるシャノン (Shannon)-ファノ (Fano) のモデルについて説明する。通信が情報通信と呼ばれるときの新たな本質は送信機と受信機の内部において共通の bit 系列を再現すべくシステムを設計することにある。情報通信では送信機は、**広義の情報源符号化**と**広義の通信路符号化**の接続であり、両者は、bit 系列 (0 と 1 の列) のみにより結ばれる。すなわち、送信機の入力信号は広義の情報源符号化により bit 系列に変換され、それが広義の通信路符号化に入力され出力として送信機の出力信号を与える。同様に受信機は**広義の通信路復号化**と**広義の情報源復号化**の接続である。すなわち受信機の入力信号は広義の通信路復号化に入力され、送信機内において出現した bit 系列と同じものがその出力として再現される。さらにその bit 系列は広義の情報源復号化に入力され受信機の出力信号として出力される。

共通の bit 系列を再現するという制約を設けることにより、広義の情報源符号化/復号化の設計問題は通信路とは独立になり、情報源と受信者へのみ依存する設計問題になる。すなわち広義通信路符号化/復号化は別な路や機能で置き換えてもかまわなくなる。また広義の通信路符号化/復号化の設計問題は通信路のみに依存し情報源と受信者とは独立な問題になる。その場合には通信路符号器はどのような入力ビット系列に対してもそれを再現する符号化でなければならない。

以上のモデルのもうひとつの意味合いは、受信者が受け取る信号に含まれる情報はそれ以前のすべての段階、特に共通のビット列の中に含まれていなければならないということである。

この問題をすこし簡単化して考えてみよう。広義情報源符号化における典型的な処理はサンプリング・量子化と**狭義の情報源符号化**を接続することである。前者はアナログ信号を離散時間離散データに変換する操作である。対応して広義情報源復号化では**狭義の情報源復号化**と平滑化を接続することである。この場合、離散時間離散アルファベットのデータ列は両過程で同じものが再現されることを要請する。狭義の情報源符号化/復号化を本稿では**情報源符号化/復号化**と呼ぶことにする。情報源符号化と情報源復号化は互いに逆変換を与えることになる。情報源符号化/復号化の性能は情報源の出力記号あたりの共通の bit 系列の長さであり、この比を R (単位 bit/記号, これを情報圧縮率と呼ぶ) と表そう。 R が最小となる情報源符号化/復号化はどのようなものであろうか。実は情報源に対して後に述べる**確率的性質**とそのひとつの性質である**定常性**を仮定できるならば R の下限が存在することを論じることができる。この下限は定常な離散時間離散アルファベット情報源の特性量でありエントロピー H と呼ばれる。さて、 R が最小化されたとき、すなわちデジタル情報の圧縮が最良

に行われた場合、その出力 bit 列は実は公平なコイン投げを用いた乱数列に近い性質を持つようになる。

さて、情報源に確率性と定常性が仮定できる状況下で、今度は広義の通信路符号化の問題を考えてみる。ここで通信路は情報源とは独立な確率的な入出力関係であると仮定することにする。また通信路にも定常性を仮定することにする。上述したように広義の通信路符号化は任意のビット列を通信の対象とする。ビット列を再生することができる確率が十分 1 に近いという条件を満足するように広義の通信路符号化/復号化をビット列のレートが最大化されるように設計した場合、そのレートが先ほどの H/Ts を上回るならば離散時間離散アルファベットのデータ列に関しては通信が成立していることになる。

さて広義の通信路符号化/復号化には通信を可能とする最大のレートが存在するかもしれない。このレートは通信路容量と呼ばれ bit 毎秒の単位で表される。以上の考察から H/Ts がこの通信路容量を下回ることが通信ができることの必要十分な条件であるとみてとれる。

上述の通信路容量を求めることは一般にむづかしい。とくにアナログ通信路の場合ごく特殊な通信路に対してしかそれを計算することができていない。一方広義の通信路符号化においては、変復調器が用いられることが多い。変復調器にはシンボル周期 T_c とシンボル配置というパラメータがある。今これを固定することにしてみよう。そうすると広義通信路符号化は離散時間離散アルファベットの通信路を対象とするものになる。

離散時間離散アルファベット通信路に定常性が仮定できる場合その通信路容量が計算できる場合が増えてくる。今それを C (bit/通信路入力記号) と表すことにしよう。時間当たりでは C/T_c (bit/sec) となる。したがってサンプリング/平滑化+量子化と、変復調器を固定した場合の通信可能性条件は $H/Ts < C/T_c$ と書くことができる。右辺は変復調器を変えて最大化が可能でありその結果がアナログ通信路の通信路容量である。さて、通信路では入力信号の電力を大きくとることができれば高い C/T_c を得ることができる。そこで電力を制限することにしよう。 $0 < C/T_c - H/Ts < \text{微少量}$ の不等式が成り立つような電力の通信路を考える。このとき通信は可能であり、再現されているビット列は先ほどの考察により公平なコイン投げの系列であるということができる。

そこで、情報通信システムにおいては広義情報源符号化/復号化と広義通信路符号化/復号化の両システムの間で純粋情報 (bit 系列) が再現されることを目標とすることを設計原理とするようになったのである。

1.2 解題

もう少し簡単な言葉で言い換えることができる。データが変換を施しても再生できるということは、変換をしてもデータのもつ情報が不変に保

たれているということができる。変換により見た目の情報量が変化するがその見た目の情報量には最小値があるであろう。これがデータの持っている情報量であると考えべきである。事実この情報量こそ変換に関するデータの不変量である。

また通信路はランダムな入出力システムである。これを介して保たれているメッセージの情報量こそ、この通信路を介して伝送可能な情報量である。メッセージは入力の中に符号化されて存在していなければならない。一方出力の中にも存在していなければならない。すなわちメッセージは入出力の共通の情報に含まなければならない。一方、入力を選択できるわけであるから、この共通情報を最大化するよう入力を選択することができる。すなわち伝送可能なメッセージの情報量 C/T_c とは、通信路の入出力の共通情報 (毎秒) を入力の特性を変えて得られる最大値であるといえる。

第2章 Huffmanの符号化と情報源 符号化/復号化

2.1 Huffmanの符号化

1番から M 番まで番号がふられたカードがある。Aさんはこれらのなかからカードによって決まっている確率で一枚をでたために選ぶものとする。BさんはAさんが選ぶ各カードの確率は知っているが、Aさんがどのカードを選ぶかは最初はいっさい知らないものとする。ただ、BさんはAさんに yes か no かで答えられる質問を何回か行うことができ、その回答によりBさんが選んだカードを知ることができるものとする。問題は、Bさんはどのような質問をしてゆけば平均的に一番早くAさんが選んだカードを知ることができるだろうか？である。

まず yes no でこたえることができるような質問とはつまるところカードの部分集合に対して選んだカードはそこにあるか？と問うことに他ならないことがわかる。例えば選んだカードは偶数かどうかとか、選んだカードを3倍して1を足したら偶数になるかなどはみな部分集合を指定する形に記述できる。またこの部分集合はこれまでに得られた質問の答えに当然依存して構わない。

そこで集合の言葉を使ってすこし厳密に論じよう。 $S_\lambda = \{1, \dots, M\}$ をカードの集合としよう。Aさんが選ぶカードはBさんには未知なので X とあらわすことにしよう。言い替えるとAさんには X は既知であるがBさんは $X \in S_\lambda$ であることだけが分かっているのである。さて、答え yes を1, no を0であらわすことにしよう。

まずBさんは S_λ の部分集合 S_1 を決め、 $X \in S_1$? と問うものとしよう。 $S_0 = S_\lambda - S_1$ と書くことにすると答えが1であれば、 $X \in S_1$ 、0であれば、 $X \in S_0$ に限定されることが分かる。後はこの繰り返しである。

それまでの答えの列を左から順に書き二進列 c としてあらわす。(実は λ というのは空列の意味である。) c によって限定される X を含む集合を S_c と書く。すなわちBさんには $X \in S_c$ が既知であるとする。 S_c の要素数が1ならそれが知りたい X なので、2以上であるとする。次に意味のある質問は S_c の空でない真部分集合 S_{c1} をとって $X \in S_{c1}$? と問うことである。(あるいは、 S_c から S_{c1} を切り取るような Q で $X \in Q$? を問うてもよい。) $S_{c0} = S_c - S_{c1}$ とすれば、答えにより $X \in S_{c1}$ または $X \in S_{c0}$ が分かれるはずであり、そのうち X が唯一つの要素として決定する。こ

の意味で X はある答えの系列 c によって表現される．そこでこの系列を $c(X)$ とあらわしてもよい．また，その系列の長さを $l(X)$ とあらわすことにしよう．

さて A, B にとってあらかじめ $X = x$ となる確率 $p(x)$ がすべての $x \in S_\lambda$ について既知であると仮定しよう．この仮定のもとに質問を設計してもよいものとしよう．

期待質問回数は

$$\sum_{x \in S_\lambda} l(x)p(x)$$

とかける．さて，この期待質問回数を最小にする質問系を見出すことができるであろうか？

以下ではそれを説明しよう．まず最初に，和の各項は $p(x)$ と $l(x)$ の積なので，最適な質問系は $p(x)$ が大きいほど $l(x)$ が小さいものでなければならないことが分かるであろう．（もし $p(x_1) > p(x_2)$ でありながら $l(x_1) > l(x_2)$ となっている質問系 $c: \mathcal{X} \rightarrow \{0,1\}^*$ があるとする．と $c(x_1)$ と $c(x_2)$ を交換したのも質問系であるが，この交換によって期待質問回数は減少する．）これはかなり役立つ知識である．

実際このことを知っていると次のことが分かる． $p(1)$ が仮に一番小さな確率で， $p(2)$ が二番目に小さいとする．今最適な質問の系があるとする． $l(1)$ は最大の長さをもつので最後に 1 と分離する相手は単一要素 (y とよぼう) でなければならない．すなわち $l(y) = l(1)$ である． y が 2 でないとすれば長さの単調性により $l(2)$ も同じ長さでなければならない． y と 2 を入れ替えた質問の系でも最適性は失われない．すなわち，一般性を失わず 1 と 2 を最後に分離する質問系が最適な系の一つであるといえる．これは cool だ．どうしてか？ 最後に 1 と 2 の分離を質問するまでは 1 と 2 は同じものだと考えてよいことになる．1 か 2 が出る確率は $p(1) + p(2)$ である． M 枚に対する最適な平均質問回数はこの対を一枚のカードだと思って組んだ $M - 1$ 枚に対する最適平均質問回数追加平均質問回数 $p(1) + p(2)$ を加えたものになる．要は M 枚のカードに対する設計問題が $M - 1$ 枚のカードに対する設計問題に帰着されたわけである．これは再帰的に一枚ずつ設計問題の枚数を減らしてゆくと最後はカードの枚数が 2 枚になる．この場合の最適な質問系は自明であるので結局最初の M 枚のカードに対する問題の答えが得られたことになる．具体的には以下のようにする．質問による分離を二分木の分岐で表現するとよい．

最初の確率 $\{p(1), \dots, p(M)\}$ は将来の二分木の葉に付随するとする．確率は大きさの順にソートしておく．以下を繰り返せばよい．「最小の 2 確率の共通の親を生成しその接点に 2 確率の和を付随させる．この親からの 2 本の枝に 0,1 のラベルをつける．2 確率を除き和を加えた確率リストができるがこれをソートしなおす．（この場合 2 確率の和がソートされたリストにマージするだけになる．）」この繰り返しの前後で確率リストが 2

本できるがこれを隣り合った列で表現すれば、次々と縮小する $M - 1$ 本の列ができて最後が2確率になる。

この方法を、それを最初に考えた人物の名前をとって、Huffman 符号化と呼ぶ。符号化というのは、各 x を $c(x)$ で符号化しているからである。

一つの注意をする。上では議論を簡単にするため無駄な質問はしないことにした。無駄な質問を許す質問系の平均符号長は無駄な質問は0回と数えるような平均符号長で下界されるが、これこそ無駄な質問を許さない質問系の平均符号長にほかならない。

2.2 Kraft の等式

カードの集まりを $\mathcal{X}$ と書こう。集合 $\{c(x), x \in \mathcal{X}\}$ を符号とよび、各要素 $c(x)$ を符号語とよぶ。この区別は対応関係あるいは集合を表す言葉と個々の要素を表す言葉を区別するためである。符号について以下の式 (Kraft の等式) が成り立つ。

$$\sum_{x \in \mathcal{X}} 2^{-l(x)} = 1$$

証明 S_λ には単位実数区間 $[0, 1)$ が付随するものとする。同様に質問系の S_c に $[0, 1)$ の連続部分区間 $[a, b)$ が付随するものとしよう。もし c においてさらに質問が必要であり S_{c0} と S_{c1} に分割されるならばこれら各々には $[a, b)$ を二分割した $[a, (a+b)/2)$ と $[(a+b)/2, b)$ を付随させる。また、 c が実際に符号語であるなら、 S_c に付随した区間はそれ以上分割しないことにする。こうすれば区間 $[0, 1)$ は単一要素からなる集合 $S_{c(x)}$ に付随する互いに素な部分区間の和である。また $l(x)$ とは x を決定するために二分割した回数であるから、 $S_{c(x)}$ に付随する区間の長さは $2^{-l(x)}$ である。Kraft の等式はこれら二つの事実を総合したものである。証明終

例えば、符号が $\{0, 10, 11\}$ であるならば、即ち $l(0) = 1$, $l(10) = 2$, $l(11) = 2$ であるならば、 $S_\lambda \leftrightarrow [0, 1)$, $S_0 \leftrightarrow [0, 1/2)$, $S_1 \leftrightarrow [1/2, 1)$, $S_{10} \leftrightarrow [1/2, 3/4)$, $S_{11} \leftrightarrow [3/4, 1)$ となり、3つの符号語に付随する区間の長さは各々 $1/2$, $1/4$, $1/4$ である。

クラフトの等式を満足する $\{l(x), x \in \mathcal{X}\}$ について、これを質問回数とする質問系が存在することを示すことができる。

2.3 エントロピーの概念

$p(x)$ は $\mathcal{X}$ 上の確率変数 X が x に等しくなる確率であると考え。この意味で関数 $p(x), x \in \mathcal{X}$ を X の確率分布という。 X あるいはその確率分布 $p(x), x \in \mathcal{X}$ について、

$$H(X) = \sum_{x \in \mathcal{X}} p(x) \log \frac{1}{p(x)}$$

のように定義される量をエントロピーという。確率分布の関数であることを明示するために、 $H(X) = H(p(\cdot))$ と書くときもある。この量は確率変数 $\log(1/p(X))$ の期待値でもある。 $\hat{l}(x)$ を自然数として $p(x) = 2^{-\hat{l}(x)}$ であるような確率分布は diadic であるという。diadic な確率分布では最小の確率は偶数個ある。 $p(x) = 2^{-\hat{l}(x)}$ を二進小数として表現して加え合わせたとき 1 にならないからである。diadic な確率分布についてハフマンの符号化を行うと（最小の 2 確率を統合する）各手続きにおいて生ずる確率はやはり 2 の負の冪乗である。 $\log \frac{1}{p(x)}$ 回の統合の後に確率が 1 になるので、各符号語の長さは $\hat{l}(x)$ となることも分かる。従って、最適な平均符号長は上で定義したエントロピーに等しいことが分かる。

では diadic でない一般の確率分布について、最適平均符号長とエントロピーはどのような関係にあるであろうか？両者の差をとると、

$$\sum_{x \in \mathcal{X}} p(x) \{l(x) + \log p(x)\} \quad (2.1)$$

$$= \sum_{x \in \mathcal{X}} p(x) \log \frac{p(x)}{2^{-l(x)}} \quad (2.2)$$

$$\geq \sum_{x \in \mathcal{X}} p(x) \frac{1}{\ln 2} \left\{1 - \frac{2^{-l(x)}}{p(x)}\right\} \quad (2.3)$$

$$= \frac{1}{\ln 2} \left\{ \sum_{x \in \mathcal{X}} p(x) - \sum_{x \in \mathcal{X}} 2^{-l(x)} \right\} \quad (2.4)$$

$$= 0 \quad (2.5)$$

ただし、不等号を得るには基本的な関係式 $\ln z \leq (z - 1), z > 0$ の変形

$$\ln \frac{1}{z} \geq (1 - z)$$

を用いた。等号が成り立つのは確率分布が diadic であるときのみであるが、これが上で述べた Huffman 符号によって達成されたものに他ならない。まとめると、エントロピーはあらゆる質問系に関する平均質問回数の下界である。

2.4 平均質問回数の上界

さて一般の確率分布 $p(x), x \in \mathcal{X}$ について、最適平均質問回数とエントロピーはどの程度離れているであろうか？これを考えるため、

$$l(x) = \lceil \log \frac{1}{p(x)} \rceil$$

と定義すると、

$$\sum_{x \in \mathcal{X}} 2^{-l(x)} \leq \sum_{x \in \mathcal{X}} p(x) = 1$$

である． $\hat{p}(x) = 2^{-l(x)}$, $x \in \mathcal{X}$ と定義し，最大符号長 $l_{max} := \max_{x \in \mathcal{X}} l(x)$ に対応した確率 $2^{-l_{max}}$ をもつカードを複数個 $\mathcal{X}$ に追加し，有限集合 $\hat{\mathcal{X}}$ とし，その上に diadic 確率分布 $\hat{p}(x)$, $x \in \hat{\mathcal{X}}$ を構成することができる．この確率分布に対する Huffman 符号を構成しよう．その結果得られる質問系は無駄な質問をいくつか含むかもしれないが， $\hat{\mathcal{X}}$ を区別する． $p(x)$, $x \in \mathcal{X}$ について平均質問回数長を計算すると，

$$\sum_{x \in \mathcal{X}} p(x)l(x) \leq \sum_{x \in \mathcal{X}} p(x) \left\{ \log \frac{1}{p(x)} + 1 \right\} \leq H(X) + 1$$

がなりたつ．即ち，2.1 節の最後に注意したように無駄な質問を許容する質問系の平均質問回数は許さない質問系の平均質問回数で常に下界することができるので，上の不等式は Huffman 符号に対する上界を与えることになる．

以上の議論から， $H(X)$ が 1 に比べて相対的に大きければ，最適な期待符号長と $H(X)$ の比は 1 に漸近することがわかる．

2.5 長いデータ列の符号化

系列の先頭部分のことを**語頭 (prefix)** という．例えば系列列 x_1^∞ についてその先頭部分 x_1^l は長さ l の語頭である．一般に，符号化 $c: \mathcal{X} \rightarrow \{0,1\}^*$ が一対一写像のとき c は**正則**であるという．正則な符号の性質はほとんどその像 $c(\mathcal{X})$ により決まってしまう．Huffman の符号化 c_{Huff} は正則である． $c_{Huff}(\mathcal{X})$ については，加えて，「任意の無限列 x_1^∞ は $c_{Huff}(\mathcal{X})$ の中にただひとつの語頭をもつ」という著しい性質がある．一般に，正則な符号 c について「任意の無限列 x_1^∞ が $c(\mathcal{X})$ の中に高々ひとつの語頭をもつとき」 c は**語頭符号**であるという．語頭符号に関しては，符号化を任意の回数行ってその結果を接続させたものが提示されたとき，復号器は先読みをすることなく瞬時に各符号語の終わりを見出して分解しながら復号することができる．この意味で語頭符号は**瞬時符号**とも呼ばれる．語頭符号の上記の条件において「高々ひとつ」を「必ず 1 つ」と言い換えることができるとき符号はさらに**完全語頭符号**であるという．完全語頭符号では任意の列をただひとつの方法で瞬時に $c(\mathcal{X})$ の列に分解してゆくことができる．この事実は集合 $c(\mathcal{X})$ の要素を二進木の節点として表現したときある部分二進木の外部に張り付いた葉の集合を構成していることから明確に理解できる．

メッセージ集合 $\mathcal{X}$ の n 次拡大 $\mathcal{X}^n$ を考える． $\mathcal{X}^n$ の上に確率分布 $\{p(x^n), x^n \in \mathcal{X}^n\}$ が与えられているときこの確率分布を利用してその上に Huffman の符号化を構成することができるであろう．この符号は $\mathcal{X}^n$ の上の完全語頭符号の中でもっとも期待符号長の短い符号化である．

さて確率分布が時間をシフトしても不変に保たれるようなデータ（これを定常なデータと呼ぶことにする）を対象に情報源の符号化の問題を考え

る．加えてこのデータがエルゴート性，すなわち時間平均と期待値が長期的に一致するような性質をもつと仮定しよう．

各 m について $X_{(m-1)n+1}^{mn} \in \mathcal{X}^n$ をその確率 $\{p(x^n), x^n \in \mathcal{X}^n\}$ に関する Huffman 符号により符号化し，符号語を $m = 1, 2, \dots$ の順に接続することにより情報源を符号化してゆくことにする．Huffman 符号は完全語頭ふごうであるから，接続した結果から瞬時に一意的に復号可能である．またその文字当りの符号化性能は含まれるブロック数が大きくなると

$$(1/n) \sum_{x^n \in \mathcal{X}^n} p(x^n) l(x^n)$$

に収束することが知られている．一方前節の結果からこの式は区間

$$[(1/n)H(X^n), (1/n)(H(X^n) + 1)]$$

に含まれる．一方定常性だけから $(1/n)H(X^n)$ は極限值 H をもつことが既知である．結局文字当りの符号化性能は H に収束することになる．

この事実は第1節において，情報源符号化/復号化の性質として述べられたことの理由であり，定常情報源に対する情報源符号化定理として知られている講義前半の主要な結果である．

第3章 確率論とエントロピー

3.1 確率

確率論では、この世の中のすべてのランダムネスを神様が支配しており、神様は無限個の標本 (sample) の集まりの中からランダムに一つを選んでその結果により出来事の生起、非生起を決定するという数学的モデルを用いる。標本の全体を Ω と表しこれを標本空間 (sample space) といい、各標本を小文字 ω で表す。例としては、標本空間は無限回の公平なコイン投げのあらゆる結果からなるものとしよう。即ち各標本 ω はコインの表裏の無限列とする。

標本の集合 A を事象 (event) という。標本 ω が選ばれたときにそれが A に含まれるならば、すなわち $\omega \in A$ ならば、事象 A が起こる (occur) と定義する。従って逆に $\omega \notin A$ ならば、事象 A が起こらなかったということになる。事象という用語は哲学用語なのでわかりにくいだが event という英語からもわかるとおり出来事という意味である。

確率を定義する前に、事象についてまとめておこう。上で述べたことから分かるように事象とは標本空間の部分集合のことであり、事象について成り立つことは集合論を用いて言い換えることができる。以下に定義する基本的用語は集合論の用語に置き換えると分かりやすいであろう。

事象 A に属さない標本の全体は一つの事象であるがこれを A の余事象 (complement) といい、 A^c と書く。即ち A^c は全体集合 Ω に関する A の補集合である。事象 A, B について、 $A \cup B$ を和事象 (union) $A \cap B$ を積事象 (intersection) という。 $A \cap B = \phi$ のとき事象 A と B は互いに素 (disjoint) であるという。複数の事象はそれらのうちどの2つの事象をとっても互いに素であるならばやはり互いに素であるという。 Ω 自身は事象でありこれを全事象 (total event) という。全事象の余事象 ϕ は空事象 (empty event) とよぶ。1 標本 $\omega \in \Omega$ からなる事象 $\{\omega\}$ を基本事象 (simple event) という。

事象は集合であるがその全体を定義しておくとも便利である。

定義 1 (σ 加法族) Ω の部分集合の族 $\mathcal{B}$ が以下の3つの条件を満たすとき、 σ 加法族であるという。

1. $\Omega \in \mathcal{B}$
2. $A \in \mathcal{B}$ ならば $A^c \in \mathcal{B}$

3. $A_1, A_2, A_3, \dots \in \mathcal{B}$ ならば $\bigcup_{l=1}^{\infty} A_l \in \mathcal{B}$

実は、以上の定義は σ 加法族は加算回のブール演算について閉じた集合族であるということを述べている。これこそが事象の族として望ましい性質を現しており、以降事象族とは σ 加法族であるものとする。またこの事象族を $\mathcal{B}$ とあらわす。

以下では事象 A の確率を $P(A)$ と表わしたい。すなわち、確率とは事象族 $\mathcal{B}$ の上で定義される（実数値）関数であるとしたい。

定義 2 (確率の公理) 関数 $P : \mathcal{B} \rightarrow \mathbb{R}$ が以下の3つの条件を満たすとき確率であるということにする。

非負性 $(\forall A \in \mathcal{B})(P(A) \geq 0)$

全確率性 $P(\Omega) = 1$

加算加法性 $A_1, A_2, \dots$ が互いに素な $\mathcal{B}$ の可算無限列であるとする。このとき

$$P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$$

この定義から導かれる基本的性質を列挙しよう。

空事象の確率 $P(\phi) = 0$

有限加法性 加法性は無限集合列を有限集合列に置き換えても成立する

相補性 $P(A) + P(A^c) = 1$

事象には事象の特性関数が付随する。逆に標本に関する命題にはその真値を与える事象が対応する。この事象の確率は、命題が真である確率に他ならない。この確率を $P(\text{命題})$ と書く。

3.2 確率変数と確率分布

標本空間から集合 $\mathcal{X}$ への関数

$$X : \Omega \rightarrow \mathcal{X}$$

は、 $\mathcal{X}$ 上の確率変数と呼ばれる。また $\mathcal{X}$ は $X(\omega)$ が分布する集合なので、 X の分布空間であるという。

分布空間の例には、離散集合（有限集合、可算無限集合）、連続集合やそれらの直積集合である。分布空間の部分集合には集合の大きさが定められていることが多い。集合の大きさとは集合に非負の数値を対応させる関数で、確率と同様加法性をもつ集合の関数である。例えば離散集合の場合

には集合の大きさとは集合の要素数である．また， n 次元連続集合の場合その n 次元体積である．これは $n = 1$ では長さ， $n = 2$ では面積， $n = 3$ では体積である．集合 A の大きさを $\mu(A)$ と記述する．

確率変数 X に付随して関数

$$p_X : \mathcal{X} \rightarrow [0, 1]$$

を

$$p_X(x) = \lim_{A \downarrow \{x\}} \frac{P(X \in A)}{\mu(A)}$$

のように定める．ただし $\mu(A)$ は集合 A の体積である．

この関数を X の確率分布（あるいは確率密度）という．引数として確率変数の小文字を用いる場合添え字を省略し単に $p(x)$ と略記する．上の定義において，特に離散集合の場合には $p_X(x) = P(X = x)$ となる．

3.3 条件付確率

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

事象 A を全事象とみたときの事象 $A \cap B$ の相対確率を表す．これを事象 B の事象 A に関する条件付き確率 (**conditional probability**) と定義する．また $P(A) = 0$ の場合 $P(B|A) = 0$ であり比は未定義であるが， $[0, 1]$ に属する任意数とする．

$$P(B|A)P(A) = P(A \cap B)$$

が成り立つ．特に

$$P(A \cap B) = P(B)P(A)$$

のとき，事象 A と B は独立である (**independent**) という．

確率変数 $X : \Omega \rightarrow \mathcal{X}$ と $Y : \Omega \rightarrow \mathcal{Y}$ の同時確率分布 (**joint distribution**) を

$$p_{XY}(x, y) = \lim_{A \downarrow \{(x, y)\}} \frac{P((X, Y) \in A)}{\mu(A)}$$

と定義する．これは確率変数 $(X, Y) : \Omega \rightarrow \mathcal{X} \times \mathcal{Y}$ の確率分布 $p_{(X, Y)}((x, y))$ の定義において，余計な括弧とカンマを省いたものになっている．引数から関数形が理解できるならば添え字を省き $p(x, y)$ と記述する．特に，離散確率変数の場合には， $p(x, y) = P(X = x, Y = y)$ である．

X の Y に関する条件付確率分布 (**conditional distribution**) を

$$p_{X|Y}(x|y) = \frac{p_{XY}(x, y)}{p_Y(y)}$$

と定義する。これは離散の場合,

$$\frac{P(X = x, Y = y)}{P(Y = y)}$$

に他ならない。 $p_{X|Y}(x|y)$ も $p(x|y)$ と略記する。 $\mathcal{X}$ が離散集合の場合, 同時確率分布 $p(x, y)$ について, $p(x) = \sum_y p(x, y)$ を X に関する周辺分布 (**marginal distribution**) という。 Y に関する周辺分布 $p(y)$ も同様に定義される。 $p(x, y) = p(x)p(y)$ のとき, 確率変数 X と Y は独立であるという。

3.4 エントロピー, ダイバージェンス, 相互情報量

本節以下では, $\mathcal{X}$ は離散集合とする。有限集合であればことわる。確率変数 (X, Y) のエントロピー $H((X, Y))$ を $H(X, Y)$ と略記し確率変数 X と Y の同時エントロピー (**joint entropy**) と定義する。 $H(X, Y) = H(Y, X)$ が成り立つ。 $H(X, Y) - H(X)$ を $H(Y|X)$ と書いて, X に関する Y の条件付きエントロピー (**conditional entropy**) いう。 $I(X, Y)$ を $H(Y) - H(Y|X)$ と定義する。これを X と Y の相互情報量 (**mutual information**) という。これは $I(X, Y) = H(Y) + H(X) - H(X, Y)$ と書ける。即ち X と Y について対称な量である。 $\mathcal{X}$ 上の確率分布 p と q について

$$D(p||q) := \sum_{x \in \mathcal{X}} p(x) \log \frac{p(x)}{q(x)}$$

を q に関する p の **Kullback-Leiber** のダイバージェンス (**divergence**) という。

これについて以下の定理が成り立つ。

定理 1

$$D(p||q) \geq 0,$$

ただし, 等号は $p = q$ のときのみ成立。

証明:

$$\begin{aligned} D(p||q) &= \sum_{x \in \mathcal{X}} p(x) \log \frac{p(x)}{q(x)} \\ &\geq \sum_{x: p(x) > 0} p(x) \frac{1}{\ln 2} \left(1 - \frac{q(x)}{p(x)}\right) \\ &= \frac{1}{\ln 2} \left(\sum_{x: p(x) > 0} p(x) - \sum_{x: p(x) > 0} q(x) \right) \\ &\geq 0 \end{aligned}$$

等号成立条件は容易にチェックできる．証明終わり．

分布空間を $\mathcal{X} \times \mathcal{Y}$ にもつ二つの確率分布 $p(x, y), q(x, y) := p(x)p(y)$ について,

$$\begin{aligned} D(p||q) &= \sum_{x,y \in \mathcal{X} \times \mathcal{Y}} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} \\ &= H(X) + H(Y) - H(X, Y) \\ &= I(X; Y) \end{aligned}$$

が成り立つ．従って, $I(X; Y) \geq 0$ が成り立つ．また等号成立条件は $p(x, y) = p(x)p(y)$, 即ち X, Y が独立な確率変数のときそのときのみである．

$\mathcal{X}$ 上の $D(p||q)$ について, $q(x) = (1/\#\mathcal{X})$ と置くと, $D(p||q) = \log \#\mathcal{X} - H(X)$ となる．従って, $\log \#\mathcal{X} - H(X) \geq 0$ が成り立つ．さらに等号は $p(x) = (1/\#\mathcal{X})$ の場合だけである．

さて,

$$\begin{aligned} H(Y|X) &= H(X, Y) - H(X) \\ &= \sum_{x,y} p(x, y) \log \frac{p(x, y)}{p(x, y)} \\ &= \sum_x p(x) \sum_y p(y|x) \log \frac{1}{p(y|x)} \end{aligned} \tag{3.1}$$

であるが, 内側の和はエントロピーであるから非負かつ上から $\log \#\mathcal{Y}$ で上界されている．従ってその平均も同様．まとめると

$$0 \leq H(Y|X) \leq \log \#\mathcal{Y}$$

3.5 条件付相互情報量

2つの確率変数 Y, Z の確率変数 X に関する条件付同時分布を考える．今 $X = x$ の条件で Y と Z は相互情報量を有するが, これを $I(Y; Z|X = x)$ と書くことにしよう．別な表現は

$$D(p_{YZ|X}(\cdot, \cdot|X = x) || p_{Y|X}(\cdot|X = x)p_{Z|X}(\cdot|X = x))$$

である．

定義 3

$$I(Y; Z|X) = \sum_x p_X(x) I(Y; Z|X = x)$$

を X に関する Y と Z の条件付相互情報量という．

これは明らかに非負の量である.

定理 2

$$I(Y; Z|X) \geq 0 \quad (3.2)$$

次のような変形がある.

定理 3

$$\begin{aligned} I(Y; Z|X) &= E \log \frac{p(Y, Z|X)}{p(Y|X)p(Z|X)} \\ &= -H(Y, Z|X) + H(Y|X) + H(Z|X) \\ &= H(Y|X) - H(Y|Z, X) \\ &= H(Z|X) - H(Z|Y, X) \end{aligned}$$

3.6 情報源と定常性

同一分布空間 $\mathcal{X}$ 上の確率変数列 $\{X_l\}_{l \in \mathcal{N}}$ を情報源 (source) あるいは確率過程 (stochastic process) という.

$$\begin{aligned} p_{X_1 X_2 \dots X_k}(x_1, \dots, x_k) \\ &= p_{X_1}(x_1) p_{X_2|X_1}(x_2|x_1) \dots \\ &\quad p_{X_k|X_1 X_2 \dots X_{k-1}}(x_k|x_1 \dots x_{k-1}) \end{aligned}$$

が成り立つ. 以前の例からも類推できるがこれも

$$p(x_1, \dots, x_k) = p(x_1)p(x_2|x_1) \dots p(x_k|x_1 \dots x_{k-1})$$

と簡略して記述できる.

あらゆる $k \geq 1$ について, 組 $\{X_l\}_{1 \leq l \leq k}$ と組 $\{X_l\}_{2 \leq l \leq k+1}$ が同じ同時確率分布を持つとき, 情報源は定常 (stationary) であるという. 定常情報源ではあらゆる $m \geq 1, k \geq 1$ について組 $\{X_l\}_{1 \leq l \leq k}$ と組 $\{X_l\}_{m \leq l \leq k+m-1}$ が同じ同時確率分布を持つことが帰結される. またこのことから情報源からとった任意のサンプルがなす確率変数の同時確率分布は任意時間シフトしても同じ同時確率分布を持つことが分かる.

通信では情報源の確率分布は情報源からのサンプルにより推定可能でなければならない. 即ち確率分布の推定にとって一サンプル列が統計量として十分になっている必要がある. この条件を満足する情報源は定常エルゴート (stationary and ergodic) 情報源とよばれる. 定常情報源は定常エルゴート成分の混合 (確率的重ね合わせ) として表現できる. 通信装置の処理は区分的に定常エルゴートの情報源を仮定して処理が行われることが多い.

一方、定常エルゴート情報源も表現能力が高すぎる。例えば必要なサンプルパスの長さは決まらない。従ってやはりモデルを小さく限定する必要がある。定常マルコフ連鎖は定常情報源の近似を行うときに必要になる工学的に重要な情報源である。

3.7 独立情報源とマルコフ連鎖

情報源 $\{X_l\}_{l \in \mathcal{N}}$ の同時確率分布があらゆる $k \geq 1$ について

$$p_{X_1 X_2 \dots X_k}(x_1, \dots, x_k) = p_{X_1}(x_1) p_{X_2}(x_2) \dots p_{X_k}(x_k)$$

を満たすとき、独立であるという。この関係も

$$p(x_1, \dots, x_k) = p(x_1) p(x_2) \dots p(x_k)$$

と略記できる。

情報源 $\{X_l\}_{1 \leq l \leq k}$ の同時確率分布が

$$\begin{aligned} p_{X_1 X_2 \dots X_k}(x_1, \dots, x_k) \\ = p_{X_1}(x_1) p_{X_2|X_1}(x_2|x_1) \dots p_{X_k|X_{k-1}}(x_k|x_{k-1}) \end{aligned}$$

を満たすとき、 $X_1, X_2, \dots, X_k$ はこの順にマルコフ連鎖 (Markov chain) をなすといい、 $X_1 - X_2 - \dots - X_k$ と略記する。 $\mathcal{X}$ はマルコフ連鎖の状態空間 (state space) といいその要素は状態 (state) と呼ぶ。

$\mathcal{Y}$ 上の情報源 $\{Y_l\}$ と状態空間 $\mathcal{X}$ 上のマルコフ連鎖 $\{X_l\}$ があり、関数 $f: \mathcal{X} \rightarrow \mathcal{Y}$ を用いて $Y_l = f(X_l)$ と書けると、 $\{Y_l\}$ はマルコフ情報源という。マルコフ情報源は初期状態が与えられたとき情報源系列から状態の系列が復元できるとき、ユニフィラ（糸手繰り可能）と呼ばれる。ユニフィラでないマルコフ情報源は隠れマルコフ情報源であると呼ばれる。

3.8 マルコフ連鎖とデータ処理不等式

補題 1 $X_0 \rightarrow X_1 \rightarrow \dots \rightarrow X_n \Leftrightarrow X_n \rightarrow X_{n-1} \rightarrow X_{n-2} \rightarrow \dots \rightarrow X_0$

証明： 3つの確率変数 $X \rightarrow Y \rightarrow Z$ の場合についてのみ証明する。

$$\begin{aligned} p(x, y, z) &= p(x) p(y|x) p(z|y) \\ &= p(x) \frac{p(y) p(x|y)}{p(x)} \cdot \frac{p(z) p(y|z)}{p(y)} \\ &\quad (\text{Bayes の定理を 2 回利用}) \\ &= p(z) p(y|z) p(x|y) \end{aligned}$$

例： 確率変数 $X \rightarrow Y \rightarrow Z$ について，

- i) $Z = g(Y)$ なら， $X - Y - Z$
- ii) $X \rightarrow Y \rightarrow Z$ の定義で X の確率分布を用いていないことに注意する．実際 X は確率変数でなくてもよく，一般のパラメータ x であってもよい．
- iii) $X - Y - Z \Rightarrow p(x, y, z) = p(y)p(x|y)p(z|y)$ つまり $Y = y$ が与えられたとき， X と Z は条件付に独立であり， $I(X; Z|Y) = 0$ が成立する．

定理 4 (データ処理不等式) $X - Y - Z \Rightarrow I(X; Y) \geq I(X; Z)$

証明： 相互情報量 $I(X; (Y, Z))$ を省略して $I(X; Y, Z)$ とかこう．この情報量は次の2通りに変形できる．

$$\begin{aligned} I(X; Y, Z) &= I(X; Y) + I(X; Z|Y) \\ &= I(X; Z) + I(X, Y|Z) \end{aligned}$$

ここで，最初の等号の右辺の第二項は零であり，第二の等号の右辺の第二項は非負である．従って，定理が成り立つ．この変形の副産物として次の系もなり立つ．

系 1 $X - Y - Z$ ならば $I(X; Y|Z) \leq I(X; Y)$ ，すなわち， $I(X; Y; Z) \geq 0$ が成り立つ．

3.9 定常独立情報源と定常マルコフ連鎖

定常独立情報源は別名ベルヌーイ情報源，あるいは i.i.d (独立同一分布) 情報源とよばれる．パラメータ数が $\mathcal{X}$ のサイズマイナス 1 であり，取り扱いやすく基本的である．マルコフ連鎖は状態遷移確率 $p_{X_l|X_{l-1}}(b|a)$ が時刻 l によらないならば，定常であるという．マルコフ連鎖の確率分布は初期状態の分布 $p_{X_1}(a)$ とともに決まる． $\mathcal{X} = \mathcal{Y}$ とみなすとマルコフ連鎖はマルコフ情報源となるが，この情報源が定常であるためには，初期状態の分布に制約が生じる．この必要な制約条件は例えば $p_{X_1}(a) = p_{X_2}(a), a \in \mathcal{X}$ である．実は，この方程式は必要かつ十分な条件を与えることが簡単に見て取れる．その意味で，この方程式を状態方程式 (state equation) という．またそれを満たす初期状態の確率分布を定常状態分布 (stationary distribution) という．少し具体的に，方程式を記述し解いてみよう．状態集合 $\{0, 1, 2\}$ 上に定常マルコフ連鎖が状態遷移確率

$$\begin{pmatrix} p(0|0) & p(1|0) & p(2|0) \\ p(0|1) & p(1|1) & p(2|1) \\ p(0|2) & p(1|2) & p(2|2) \end{pmatrix} = \begin{pmatrix} \bar{\alpha} & \alpha & 0 \\ \beta & 0 & \bar{\beta} \\ 1 & 0 & 0 \end{pmatrix}$$

で与えられている。ただし $0 < \alpha, \beta < 1$ とする。この行列を P と表す。定常状態の確率分布を行ベクトル $q = (q_0, q_1, q_2)$ で表すと、状態方程式は

$$qP = q$$

と書ける。最後の 2 要素と、全確率性に対応する方程式は具体的に

$$\begin{aligned} q_1 &= \alpha q_0 \\ q_2 &= \bar{\beta} q_1 \\ q_0 + q_1 + q_2 &= 1 \end{aligned}$$

となる。従って、

$$(q_0, q_1, q_2) = \frac{1}{1 + \alpha + \alpha\bar{\beta}} \times (1, \alpha, \alpha\bar{\beta})$$

となる。

3.10 定常情報源のエントロピーレート

定義 4 極限

$$\lim_{n \rightarrow \infty} \frac{1}{n} H(X_1, \dots, X_n)$$

が存在するときこれを情報源 $\{X_l\}_{l \in \mathcal{N}}$ のエントロピーあるいはエントロピーレート (**entropy rate**) という。この値を H_∞ とかく。

次の定理が成り立つ。

定理 5 $\{X_l\}_{l \in \mathcal{Z}}$ が定常であり、 $H(X_1) < \infty$ のとき、 H_∞ が存在しその値は

$$H_\infty = \lim_{n \rightarrow \infty} H(X_n | X_{n-1}, X_{n-2}, \dots, X_1)$$

で与えられる。

証明:

補題 2 有限な極限が存在する数列に対しては、その算術平均のなす数列も同じ極限を有する。

証明: 数列 $\{a_n\}$ が $\lim_n a_n = a < \infty$ とするとき、 $\forall \varepsilon > 0, \exists n_0(\varepsilon), \forall n \geq n_0(\varepsilon)$ に対して、

$$\begin{aligned} \left| \frac{1}{n} \sum_{l=1}^n a_l - a \right| &\leq \frac{1}{n} \sum_{l=1}^n |a_l - a| \\ &\leq \frac{1}{n} \sum_{l=1}^{n_0(\varepsilon)} |a_l - a| + \frac{n - n_0(\varepsilon)}{n} \varepsilon. \end{aligned}$$

ここで、 n を十分大きくすると、右辺は、 2ε で上から上界される。

補題 3 定常情報源 $\{X_n\}$ について, $H(X_1) < \infty$ ならば, $H_\infty(X)$ が存在し,

$$H_\infty(X) = \lim_{n \rightarrow \infty} H(X_n | X_1, \dots, X_{n-1})$$

がなりたつ. また,

$$H_\infty(X) \leq H(X_n | X_1, \dots, X_{n-1})$$

が成り立つ.

証明: エントロピーに関するチェインルールより,

$$H(X^n) = \sum_{l=1}^n H(X_l | X_1, \dots, X_{l-1})$$

が成り立つ.

$$\begin{aligned} a_l &:= H(X_l | X_1, \dots, X_{l-1}) \\ &= H(X_{l+1} | X_2, \dots, X_l) \quad (\text{定常性}) \\ &\geq H(X_{l+1} | X_1, \dots, X_l) \quad (\text{条件の追加}) \\ &= a_{l+1} \end{aligned}$$

より, 数列 $\{a_l\}$ は, 初項 $a_1 = H(X_1)$ が有限であり, 単調減少, かつ非負である. したがって, $\lim_{n \rightarrow \infty} a_n$ が有限の値として存在する. 定理は補題をこの場合に解釈することにより得られる.

3.11 情報源符号化定理

第2.1節ではカードの集合に対する質問の構成を行う最適な方法考えた. 情報源 $\{X_l\}_{l \in \mathcal{Z}}$ からの固定の長さのブロック $\{X_l\}_{1 \leq l \leq n}$ の全てををカードの集合とみて, これに対する質問の構成を考えてみよう. 平均質問回数の下限は $H(X_1, \dots, X_n)$ で与えられる. また, 上限は2.4節で示されたように $H(X_1, \dots, X_n) + 1$ である. 従って, 次の事実にいたる.

定理 6 情報源が定常ならば, 情報源 1 記号あたりの平均質問回数の最適値は以下の不等式を満足する. 即ち, 任意の正数 ε に対して, 十分大きなすべての n について, 確率変数 $(X_1, \dots, X_n)$ に対する質問系

$$c: \mathcal{X}^n \rightarrow \{0, 1\}^*$$

が存在しその一記号当たりの平均符号長 $\frac{1}{n} El(X_1, \dots, X_n)$ の最適値は

$$H_\infty \leq \frac{1}{n} El(X_1, \dots, X_n) \leq H_\infty + \varepsilon$$

を満足する.

3.12 定常マルコフ情報源に対するエントロピーレート

前節で情報源のエントロピーレートの重要性が理解できたので定常マルコフ連鎖 $\{X_l\}_{l \in \mathcal{N}}$ についてそれを求めてみよう。 X_1 は定常分布に従うとする。

$$\begin{aligned}
 H(X_1, \dots, X_n) &= \\
 &= H(X_1) + \sum_{l=2}^n H(X_l | X_1, \dots, X_{l-1}) \\
 &= H(X_1) + \sum_{l=2}^n H(X_l | X_{l-1}) \\
 &= H(X_1) + (n-1)H(X_2 | X_1)
 \end{aligned}$$

が成り立つ。第一の等式はチェイン規則と呼ばれるものである。第二の等式はマルコフ性による。第3の等式は定常性による。即ち、 $H_\infty = H(X_2 | X_1)$ となる。3.9 節の例では、

$$\begin{aligned}
 H_\infty &= H(X_2 | X_1) \\
 &= \sum_{i=0}^2 q(i) H(p(\cdot | i)) \\
 &= q_0 h(\alpha) + q_1 h(\beta) \\
 &= q_0 h(\alpha) + q_1 h(\beta) \\
 &= \frac{h(\alpha) + \alpha h(\beta)}{1 + \alpha + \alpha \bar{\beta}}
 \end{aligned} \tag{3.3}$$

ただし、上式では二進エントロピー関数

$$h(r) = -r \log r - \bar{r} \log \bar{r}$$

を用いて表している。

3.13 AEP 性 (漸近等分割性)

情報理論で最も重要な概念が情報源（すなわち確率過程）の AEP 性（漸近等分割性）である。その意味を厳密に示すため無記憶情報源を例にとる。

補題 4 (チェビシェフの不等式) 期待値 $\mu = EX$, 分散 $\sigma^2 = V(X) = E(X - \mu)^2$ をもつ確率変数 X について、チェビシェフの不等式：

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

が成り立つ。

証明：

$$\begin{aligned}
 \text{左辺} &= E1\{|X - \mu| \geq k\sigma\} \\
 &= E \frac{(X - \mu)^2}{k^2\sigma^2} \\
 &= \frac{1}{k^2}.
 \end{aligned} \tag{3.4}$$

定理 7 無記憶情報源 $\{Y_l\}$ において、 Y_1 の期待値、分散が有限のとき、その平均値 $\frac{1}{n} \sum_{l=1}^n Y_l$ は期待値に確率収束 (**convergent in probability**) する。正確に言い換えると、あらゆる $\varepsilon > 0$ において、

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{1}{n} \sum_{l=1}^n Y_l - EY_1\right| \geq \varepsilon\right) = 0$$

が成り立つ。

証明：チェビシェフの不等式を用い、 $\varepsilon = k\sqrt{\frac{1}{n}V(Y)}$ とおくと、

$$P\left(\left|\frac{1}{n} \sum_{l=1}^n Y_l - EY_1\right| \geq \varepsilon\right) = \frac{V(Y)}{n\varepsilon^2} \tag{3.5}$$

が成り立つ。この右辺は $n \rightarrow \infty$ のとき零に収束する。

定理 8 $\mathcal{X}$ 上の離散無記憶情報源 $\{X_l\}$ において、 $H(X) < \infty$, $V(-\log p(X)) < \infty$ のとき、 $-(1/n) \log p(X^n)$ は $H(X)$ に確率収束する。

注意 1 同じ条件下で、もっと強い収束：

$$P\left(\lim_{n \rightarrow \infty} -\frac{1}{n} \log p(X^n) = H(p())\right) = 1 \tag{3.6}$$

がいえる。確率変数列のこの収束を概収束といい、

$$\lim_{n \rightarrow \infty} -\frac{1}{n} \log p(X^n) = H(p()) \quad (a.s.)$$

のようにかく。節 3.6 において、定常情報源ではエントロピーが存在することを示した。実は、上の定理 8 が成り立つためには定常性に加えて情報源がエルゴートのこと（無記憶情報源では成立）が十分である。エルゴートのとは一サンプル列を用いて確率分布（したがってエントロピー）を推定できることであると述べたが、*AEP* 性がその事実を示していると解釈できる。（ブロック長 l の同時分布の 1 記号あたりのエントロピーは $l \rightarrow \infty$ のときエントロピーレートに収束するが、同時分布は 1 サンプル列から推定できる。）

定義 5 確率分布 $p: \mathcal{X} \rightarrow \mathcal{R}$ により定義される, 無記憶情報源について, 集合:

$$\{x^n: \left| -\frac{1}{n} \log p(x^n) - H(p()) \right| \leq \varepsilon\}$$

をこの確率分布に関する ε -標準集合 (**typical set**) とよび, $A_\varepsilon^{(n)}$ とかくまたこれに属する系列 x^n を ε 標準系列 (**typical sequence**) という.

次の定理がなり立つ.

定理 9 (AEP 性) $p: \mathcal{X} \rightarrow \mathcal{R}$ に関する, $A_\varepsilon^{(n)}$ について以下が成り立つ.

1.

$$\begin{aligned} x^n \in A_\varepsilon^{(n)} \leftrightarrow 2^{-n(H(X^n)+\varepsilon)} &\leq p(x^n) \\ &\leq 2^{-n(H(X^n)-\varepsilon)} \end{aligned}$$

2. $\exists n_0(\varepsilon), \forall n \geq n_0(\varepsilon),$

$$P\left(X^n \in A_\varepsilon^{(n)}\right) > 1 - \varepsilon$$

3.

$$\#A_\varepsilon^{(n)} \leq 2^{n(H(X)+\varepsilon)}$$

4. $\exists n_0(\varepsilon), \forall n \geq n_0(\varepsilon),$

$$\#A_\varepsilon^{(n)} \geq (1 - \varepsilon)2^{n(H(X)-\varepsilon)}.$$

注意 2 AEP 性は *Asymptotic Equipartition Property* の略である. $\mathcal{X}^n$ はほぼ確率 1 の標準集合と確率的に無視可能な部分に分かれるが, 標準集合において, 系列は「一様分布」している (あるいは「漸近的に等分割」している).

証明: 1 は標準系列の定義のいい替えである. 2 は, 定理 8 そのものである. 3. は不等式

$$\begin{aligned} 1 &= \sum_{x^n \in \mathcal{X}^n} p(x^n) \\ &\geq \sum_{x^n \in A_\varepsilon^{(n)}} p(x^n) \\ &\geq \sum_{x^n \in A_\varepsilon^{(n)}} 2^{-n(H(X^n)+\varepsilon)} \quad (1.より) \\ &= \#A_\varepsilon^{(n)} 2^{-n(H(X^n)+\varepsilon)} \end{aligned}$$

からいえる．4. については，2 を成り立たせる $n \geq n_0(\varepsilon)$ について，

$$\begin{aligned}
 1 &= \sum_{x^n \in \mathcal{X}^n} p(x^n) \\
 &= \sum_{x^n \in A_\varepsilon^{(n)}} p(x^n) + \sum_{x^n \notin A_\varepsilon^{(n)}} p(x^n) \\
 &\leq 2^{-n(H(X^n) - \varepsilon)} \left(\sum_{x^n \in A_\varepsilon^{(n)}} 1 \right) + P(X^n \notin A_\varepsilon^{(n)}) \\
 &\quad (1.より) \\
 &\leq 2^{-n(H(X^n) - \varepsilon)} \#A_\varepsilon^{(n)} + \varepsilon \quad (2.より)
 \end{aligned}$$

がなりたつが，この不等式を変形することにより得られる．さて，情報理論では多出力情報源や，ネットワーク型通信路の符号化の問題を考えることができるが，AEP 性は複雑な問題の本質をえぐりだすのに有用である．例えば，AEP 性を用いると情報源符号化の順定理が簡単に得られる．

定理 10 有限のアルファベットをもつ定常無記憶情報源 $\{X_l\}$ を考える．これに対して，情報源ブロックサイズ n の一意復号可能な可変長符号の列が存在し， $n \rightarrow \infty$ においてその期待平均符号長は情報源エントロピーに収束する．

証明： $x^n \in \mathcal{X}^n$ にたいして一意復号可能な符号 φ_n を以下のように定義する： $x^n \in A_\varepsilon^{(n)}$ のとき，フラグビット 1 に x^n の $A_\varepsilon^{(n)}$ における固定符号長 $\lceil \log \#A_\varepsilon^{(n)} \rceil$ の数え上げ符号を接続する． $x^n \notin A_\varepsilon^{(n)}$ のとき，フラグビット 0 に x^n の $\mathcal{X}^n$ における固定符号長 $\lceil \log \#\mathcal{X}^n \rceil$ の数え上げ符号を接続する．任意の $\varepsilon > 0$ にたいして， $n_0(\varepsilon/2)$ を定理のようにとる． $n \geq n_0(\varepsilon/2)$ にたいして，期待平均符号長は，

$$\begin{aligned}
 E\varphi(x^n) \text{ の長さ} &\leq P(X^n \in A_\varepsilon^{(n)}) (2 + \log \#A_\varepsilon^{(n)}) + \\
 &\quad P(X^n \notin A_\varepsilon^{(n)}) (2 + \log \#\mathcal{X}^n) \\
 &\leq 2 + \log \#A_\varepsilon^{(n)} + \varepsilon/2 (2 + \log \#\mathcal{X}^n) \\
 &\leq 2 + n(H(X) + \varepsilon/2) + \varepsilon/2 (2 + n \log \#\mathcal{X})
 \end{aligned}$$

のように上界できるが，両辺を n で割り，適当な $n_1(\varepsilon) \geq n_0(\varepsilon/2)$ をとるならば， $n \geq n_1(\varepsilon)$ にたいして，右辺は $H(X) + \varepsilon$ で上界できる．一方一意復号可能な符号であるから，平均期待符号長は $H(X)$ で下界される．

第4章 通信路符号化

4.1 通信システムのモデル

送信者は送信したい情報に対応した 入力信号を 通信路に入力する。入力された信号は通信路を伝播するうちにランダムな雑音を受け変化した通信路の 出力信号として 受信者により受信される。送信者が送りたい情報の総数を M 個とするとき対応する信号を $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_M$ とかく。信号は一定間隔 T だけ続くものとしよう。例えば、連続時間信号の場合 T は時間であると考え、従って、最初の情報 1 に対応した信号は $x_1(t)$ であり、 t は送信を開始した時刻を基点として T 秒続くことになる。また離散時間信号の場合、信号は T 個のサンプルの系列であるものと考え、従って、最初の情報 1 に対応した信号は $x_1(n), 0 < n \leq T$ であり送信を開始した時刻を基点として T 個続くことになる。

一般に、入力信号を $\mathbf{x}$ であらわす。入力信号 $\mathbf{x}$ は通信路の伝播をへてランダムに変化し受信側で $\mathbf{y}$ として受け取られる。 $\mathbf{y}$ は送信開始時点から通信路による伝播時間だけの遅れを開始時点とする T 秒続く信号 $y(t)$ のことである。通信路の特性はこの間の確率的な関係を表す 条件付確率 $p(\mathbf{y}|\mathbf{x})$ としてあらわされる。 $\sum_{\mathbf{y}} p(\mathbf{y}|\mathbf{x}) = 1$ が成り立つ (Σ は和あるいは積分を表す)。通信システムを設計する問題とは、この条件付き確率に基づき $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_M$ ならびにそれらに対応する受信信号の 受容域 $D_1, D_2, \dots, D_M$ を互いに素な部分集合として定めることである。例えば、 $\mathbf{y}$ が D_1 に含まれるならば $\mathbf{x}_1$ が送信されたと判断する。 $D_1, D_2, \dots, D_M$ のいずれにも含まれない受信信号の集合を $\bar{D}$ と書く。これは送信誤りと判断する領域である。

あらゆる番号の入力信号にたいして出力信号がその番号に対応する受容域に含まれる確率がすべて $1 - \lambda$ 以上であるような通信システムを考えたい。すなわち、この通信システムでは情報が正しく受信者に伝わらない確率は λ 以下であるということになる。 λ はやがて小さくすることを求められる基準である。例えば、 $\lambda = 10^{-9}$ であれば、かなりよい通信システムであろう。いずれにせよ、任意の λ に対し、このような通信システムを $\lambda - M$ システムということにする。

また、この通信システムは T 秒に $\log M$ ビットの通信を行っている。 T を 1 マイクロ (10^{-6}) 秒とし M を 2^{20} とするならば、1 秒あたりでは 2000 万個の二進ビットが送信される。一般に、一秒あたりの送信ビット

数 $(1/T) \log M$ を 伝送レートと呼ぶ.

受信者が $\mathbf{y}$ をうる確率 $p(\mathbf{y})$ は条件付確率 $p(\mathbf{y}|\mathbf{x})$ の $\mathbf{x}$ の確率変動に対する期待値である. $\mathbf{x}$ がどのような確率法則に従うかはわからないので, 任意の確率分布 $p(\mathbf{x})$ を仮定することにしよう. 受信者からみた $\mathbf{y}$ の確率は, $p(\mathbf{y}) = \sum_{\mathbf{x}} p(\mathbf{y}|\mathbf{x})p(\mathbf{x})$ と表わすことができる. さて, 入力信号の組を所与として, 受容域を決めるとは, あらゆる $\mathbf{y}$ についてそれがどの D_l に含まれるのかを決めることに他ならない. 実は, $\mathbf{y}$ に対しては $p(\mathbf{y}|\mathbf{x})/p(\mathbf{y})$ の値が大きい $\mathbf{x}$ ほど入力である可能性が大きい. より正確にいうと,

定理 11 M 個の入力信号が等しい確率で選ばれると仮定する. このとき $\mathbf{y}$ は $p(\mathbf{y}|\mathbf{x}_l)/p(\mathbf{y})$ の値が最大になる入力信号 $\mathbf{x}_l$ の受容域に含める場合に平均誤り確率が最小になる.

(従って, $p(\mathbf{y})$ には関係はない). この受容域の決定方法は最尤復号法といわれる. 証明は以下では必要ないので省くが, 我々が考えたい $\lambda - M$ システムを設計する上でこの事実に対する知識が役にたつ.

さて, 信号の持続時間 T を伸ばせば M を大きくしても誤り確率を小さくすることができよう. では, このようにして達成可能な $\lambda - M$ システムの伝送レートの上限はどれだけか, この問題を考えてみよう.

4.2 通信路符号化順定理

まず第一に, 次のような考察をすると, 達成可能なひとつの伝送レートが見出せる. いま, $p(\mathbf{y}|\mathbf{x})/p(\mathbf{y})$ が 2^a 以上の $(\mathbf{x}, \mathbf{y})$ の対を高確率対とよびその集合を A と書こう. 同時分布 $p(\mathbf{X})p(\mathbf{Y}|\mathbf{X})$ に関して, A が生じる確率を考えることができる. 次の事実が成り立つ.

定理 12 任意の $p(\mathbf{x})$ について, $M > 2^a(P(A) - 1 + \lambda)$ を満たす $\lambda - M$ システムが存在する.

$p(\mathbf{y}|\mathbf{x})$ と $p(\mathbf{x})$ に付随して

$$\sum_{\mathbf{x}, \mathbf{y}} p(\mathbf{x})p(\mathbf{y}|\mathbf{x}) \log \frac{p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})}$$

なる値が定まる. これを $\mathbf{X}$ と $\mathbf{Y}$ の相互情報量とよび, $I(\mathbf{X}; \mathbf{Y})$ と記述する. (確率論ではランダムな変数は大文字を用い, その特定の実現値を小文字で表す習慣がある. そこで以降, この記述法を用いる) 相互情報量は非負であることが証明されている.

もし $p(\mathbf{Y}|\mathbf{X})$ と $p(\mathbf{X})$ に定常性を仮定すると, 若干の数学的条件を加えるならば $I(\mathbf{X}; \mathbf{Y})$ は時間に比例して増大することがいえる. さらにエルゴート性を仮定すると速度は同じであると考えてよい. ただし定常性とエルゴート性というのは通信システムや入力信号の統計的な性質のこと

である．例えば $p(\mathbf{x})$ を規定するのは， $x(t)$ に対する確率的な法則であるが，この確率法則が時間不変であるとき，信号 $X(t)$ は定常であるという．また，定常な信号が任意の長さの有限窓に基づく任意の関数として観測されるとき，その観測値の長時間平均が，観測値の期待値と一致するならば，信号 $X(t)$ はエルゴートのであるという．このような性質はマルコフ過程などでは広く成り立つことが議論できる．

さて， $\delta > 0$ を任意に定めて $a := I(\mathbf{X}; \mathbf{Y}) - \delta T$ と定義する．このとき，「高確率集合はほとんど確率 1 に近くなり，その違いは T に対して指数的となる」ことが適切な仮定の元に証明される．さて，いまこの事実が成り立つことを認め，伝送レートの達成可能値を求めてみよう．

入力信号全体を $\mathcal{X}^T$ と書く．また， $A_{\mathbf{x}}$ で A の $\mathbf{x}$ における断面，すなわち $\mathbf{x}$ との対として A に含まれる受信信号の集合をあらわす．いまあらゆる l に対して $\mathbf{x}_l$ の受容域はすべて $A_{\mathbf{x}_l}$ に含まれること（すなわち $D_l \subset A_{\mathbf{x}_l}$ ）を要求することにしよう．この条件を満たす $\lambda - M$ システムで M が最大となるものが存在するはずである．これをひとつ仮定しよう．そうすると，どんな $\mathbf{x}$ についても $P(\bar{D} \cap A_{\mathbf{x}} | \mathbf{x}) < 1 - \lambda$ が成立する．もし逆なら，もう一つ $\bar{D}$ の中に $A_{\mathbf{x}}$ に含まれる D_{M+1} を見つけることができるからである．

両辺を $\mathbf{x}$ について期待値をとると，

$$\begin{aligned}
 1 - \lambda &> \sum_{\mathbf{x}} p(\mathbf{x}) P(\bar{D} \cap A_{\mathbf{x}} | \mathbf{x}) \\
 &= \sum_{\mathbf{x}} \sum_{\mathbf{y}} p(\mathbf{x}) p(\mathbf{y} | \mathbf{x}) 1\{\mathbf{y} \in \bar{D}, (\mathbf{x}, \mathbf{y}) \in A\} \\
 &= \sum_{\mathbf{x}} \sum_{\mathbf{y}} p(\mathbf{x}) p(\mathbf{y} | \mathbf{x}) 1\{(\mathbf{x}, \mathbf{y}) \in A\} \\
 &\quad - \sum_{\mathbf{x}} \sum_{l=1}^M \sum_{\mathbf{y} \in D_l} p(\mathbf{x}) p(\mathbf{y} | \mathbf{x}) 1\{(\mathbf{x}, \mathbf{y}) \in A\} \\
 &\geq P(A) - \sum_{l=1}^M \sum_{\mathbf{y} \in D_l} p(\mathbf{y}) \\
 &= P(A) - \sum_{l=1}^M \sum_{\mathbf{y} \in D_l \cap A_{\mathbf{x}_l}} p(\mathbf{y})
 \end{aligned}$$

一方，右辺第 2 項に現れる $p(\mathbf{y})$, $\mathbf{y} \in D_l \cap A_{\mathbf{x}_l}$ については，条件より $p(\mathbf{y}) \leq p(\mathbf{y} | \mathbf{x}_l) 2^{-a}$ なる不等式が成り立っている．一方 $\sum_{\mathbf{y} \in D_l \cap A_{\mathbf{x}_l}} p(\mathbf{y} | \mathbf{x}_l) \leq 1$ であるから，結局

$1 - \lambda > P(A) - M 2^{-a}$ あるいは， $M > 2^a (P(A) - 1 + \lambda)$ なる下界が成立する．

以上は，次のような正方形の地図を描像すると理解しやすい． $\mathcal{X}^T$ は横軸で実長は 1，座標 $\mathbf{x}$ は $p(\mathbf{x})$ に逆比例した密度で不等間隔に描かれてい

る. $\mathcal{Y}^T$ は縦軸で実長はやはり 1, 座標 $\mathbf{y}$ は $p(\mathbf{y})$ に逆比例した密度で不等間隔に描かれている. そこには対角線方向にほとんど対角線と同形なほどシャープな山脈が走っている. 山の高さは $p(\mathbf{x}, \mathbf{y})/p(\mathbf{x})p(\mathbf{y})$ である. 従って山の体積要素は $p(\mathbf{x}, \mathbf{y})d\mathbf{x}d\mathbf{y}$ となり, 正方形内での積分は 1 となる. A は高さ 2^a の等高線で囲まれる細長い領域である. 各 $\mathbf{x}_l$ に対して断面 $A_{\mathbf{x}_l}$ を描くのは容易であろう. $\{\mathbf{x}_l\} \times D_l$ はこの部分区間である. この部分区間の山脈の断面の測度は $P(D_l|\mathbf{x}_l) \geq 1 - \lambda$ を満たす. 全ての $\mathbf{x}_1, \dots, \mathbf{x}_M$ に対して, $D_1, \dots, D_M$ に重なりはない. $\bar{D}$ は $\cup_l D_l$ の補集合であるが, どの $\mathbf{x}$ をとっても $P(\bar{D} \cap A_{\mathbf{x}}|\mathbf{x})$ の測度は $1 - \lambda$ 未満になっている.

この描像以外にもうひとつの、描像があると便利である. 上と同様に直積集合 $\mathcal{X}^T \times \mathcal{Y}^T$ を考える. 各軸の座標は上と同様に不当間隔に描かれているが, 順序は指定した例えば辞書順などを用い独立に与えられている. 上と同様周辺分布は各軸上一様に分布しているが $(\mathbf{X}, \mathbf{Y})$ は正方形の内部に一様に発生するはずである. この場合の高確率集合はどこにあるのかというと、正方形の内部に一様に分布して存在する. この描像は統計的結晶とよばれる. 両方の描像のもとに事実を整理してみることも有益であろう.

$P(A)$ は仮定した事実により, T が十分大のとき, 1 に漸近する. 従って, λ が正であるならば,

$$\frac{1}{T} \log M \geq \left(\frac{1}{T} I(\mathbf{X}; \mathbf{Y}) - \delta' \right)$$

なる M がとれることになる. ただし, δ' は $\delta \downarrow 0$ のときに $\delta' \downarrow 0$ となる正数である.

さて, 上で仮定した事実「高確率集合はほとんど確率 1 に近くなり, その違いは T に対して指数的となる」について説明しよう.

まず, 通信路の入出力が離散時間信号であり, 通信路は独立通信路であると仮定しよう. このときに $(1/T) \sup_{p(\mathbf{x})} I(\mathbf{X}; \mathbf{Y})$ を計算してみよう. 離散時間信号であるから, $\mathbf{X} = (X_1, \dots, X_n), \mathbf{Y} = (Y_1, \dots, Y_n)$ とおこう. 次の定理が成り立つ.

定理 13

$$I(\mathbf{X}; \mathbf{Y}) \leq \sum_{l=1}^n I(X_l; Y_l)$$

ただし, 等号は $(Y_1, \dots, Y_n)$ が互いに独立であるときに成立する.

Proof: $Y_1, \dots, Y_n$ の各々の周辺分布 $p(Y_1), \dots, p(Y_n)$ の積 $\prod_{l=1}^n p(Y_l)$ を $\tilde{p}(\mathbf{Y})$ とかく.

$$\begin{aligned}
 I(\mathbf{X}; \mathbf{Y}) &= E_{\mathbf{XY}} \log \frac{p(\mathbf{Y}|\mathbf{X})}{p(\mathbf{Y})} \\
 &= E_{\mathbf{XY}} \log \left\{ \frac{\prod_{l=1}^n p(Y_l|X_l) \tilde{p}(\mathbf{Y})}{\prod_{l=1}^n p(Y_l) p(\mathbf{Y})} \right\} \\
 &= \sum_{l=1}^n E_{X_l, Y_l} \log \frac{p(Y_l|X_l)}{p(Y_l)} - E_p \log \frac{p(\mathbf{Y})}{\tilde{p}(\mathbf{Y})} \\
 &= \sum_{l=1}^n I(X_l; Y_l) - D(p||\tilde{p}) \\
 &\leq \sum_{l=1}^n I(X_l; Y_l)
 \end{aligned}$$

等号は $p = \tilde{p}$ のとき, すなわち $(Y_1, \dots, Y_n)$ が互いに独立であるときに成立する. 次に, この定理より, 次の定理が成り立つ.

定理 14 離散時間独立通信路 $p(\mathbf{Y}|\mathbf{X})$ について, 以下の式が成り立つ.

$$\sup_{p_{\mathbf{X}}} I(\mathbf{X}; \mathbf{Y}) = \sum_{l=1}^n \sup_{p_{X_l}} I(X_l; Y_l)$$

ただし, 左辺の $\sup$ は $(X_1, \dots, X_n)$ が互いに独立であるときに達成する.

Proof: $p_{\mathbf{X}}$ に関する $I(\mathbf{X}; \mathbf{Y})$ の最大化をまず $p_{\mathbf{X}}$ の n 個の周辺分布を固定して行くと, 前定理により $\sum_{l=1}^n I(X_l; Y_l)$ に等しくなる. さらに周辺分布を動かして最大化すると, 定理の不等式になる. 証明終わり

離散時間情報源は独立でかつ定常なとき無記憶であるという. 同様に, 離散時間通信路は独立でかつ定常なとき無記憶であるという. 無記憶通信路の場合に上の不等式は以下の定理の形で述べられる. $C = \sup_{p(x)} I(X; Y)$ を無記憶通信路の通信路容量と定義する.

定理 15 離散時間無記憶通信路について

$$\frac{1}{n} \sup_{p_{\mathbf{X}}} I(\mathbf{X}; \mathbf{Y}) = C,$$

ただし, 左辺は通信路容量を達成する確率分布 $p(x)$ が定める無記憶情報源の場合に達成される.

さて, 離散時間無記憶通信路を仮定する. また, 以下では通信路容量は有限であると仮定する. これは $\log p(Y|X)/p(Y)$ が有限の期待値をもつことを意味している. 追加して, $\log p(Y|X)/p(Y)$ は有限の分散を持つものとしよう. たいていの確率分布は $\log$ をとると性質のよい関数になるので広範囲な通信路について成り立つ仮定である.

定理 16 $\mathbf{X}$ は $I(\mathbf{X}; \mathbf{Y}) = nC$ を達成する確率分布とする．任意の正数 δ を固定する． $P(A)$ は $a = I(\mathbf{X}; \mathbf{Y}) - n\delta$ により定義されているとする．十分大きな n について

$$P(A) \geq 1 - \delta$$

が成り立つ．

Proof:

$$Z_l := I(X_l; Y_l) - \log \frac{p(Y_l|X_l)}{p(Y_l)}$$

とおく．これは独立同一分布に従い， $EZ = 0$ である．また仮定より， Z の分散は有限である．従って，大数の法則により，十分大きな n について，

$$P(A) = 1 - P\left(\frac{1}{n} \sum_l Z_l < -\delta\right) \geq 1 - \delta$$

が成り立つ．証明終わり

さて，任意の λ に対して， $\delta = \lambda/2$ とおく．定理は $M \geq 2^{n(C-\lambda/2)}(\lambda/2)$ なる $\lambda - M$ システム，すなわち伝送率が十分 C に近い $\lambda - M$ システムの存在を主張している．以上をまとめると，

定理 17 離散時間通信路を考える． $\mathbf{X}, \mathbf{Y}$ を通信路の入出力 n 対， $(X_1, \dots, X_n), (Y_1, \dots, Y_n)$ とみる．通信路が定常無記憶通信路の場合，任意の正の λ に対して， n を十分大きくすると，伝送率が十分 $C = \max_{p(x)} I(X_1; Y_1)$ に近い $\lambda - M$ システムが存在する．

さて，以上説明した信号系ならびに受容域の設計の要点は，

- 高確率集合にのみ着目し，それ以外の入出力対が生じる可能性を無視する．
- 高確率集合に属していても，受容確率が $1 - \lambda$ に至らない対については新たに信号を用意することはしない．

である．また，通信路容量が達成されるためには，通信路容量に対応した入力確率 $p(X)$ 従って入力信号が発生する必要がある．

4.3 ランダム符号化による通信路符号化順定理

さて前節の定理は数学的に符号を構成しているがこれを実際に遂行することは不可能といってよい．本節では，もう少し直感的にやっていることが分かる符号化順定理を述べる．要点を先にのべると通信路容量を達成する入力確率分布に従って符号語 $\mathbf{x}$ をデタラメに生成しそれを符号とすればよい．対応する通信路出力 $\mathbf{y}$ との対 $(\mathbf{x}, \mathbf{y})$ が高確率であることが期待できるからである．ただしすぐに分かるが高確率でない対も生成され

うる．この点だけを最後に補正すれば前節と同じ符号化定理が得られるのである．

さて，確率分布 $p : (\mathcal{X} \times \mathcal{Y})^n \rightarrow \mathcal{R}$ により定義される，情報源について定常性が成り立ち，1時刻におけるエントロピー $H(X_1, Y_1)$ が有限であるならばエントロピーレート $H_\infty(X, Y)$ が存在して $H(X_1, Y_1)$ 以下の値をとることを我々は既に知っている．さらにエルゴート性も仮定できるならば，以下の議論はエントロピーをエントロピーレートに置き換えた形ですべて成り立つ．しかし，本小節から本節の最後まで議論の目的は，直感的な符号構成を理解することにあるので，本質を失わず厳密な議論が可能なる無記憶情報源 $p(x^n, y^n) = \prod_{l=1}^n p_{XY}(x_l, y_l)$ に話を限定する．なお，直感的な符号構成に関心がない場合には，次節に飛んで構わない．

4.3.1 同時標準系列および AEP 性

定義 6 無記憶情報源集合：

$$\left\{ (x^n, y^n) : \begin{cases} \left| -\frac{1}{n} \log p(x^n) - H(p_X()) \right| \leq \varepsilon, & (E^x) \\ \left| -\frac{1}{n} \log p(y^n) - H(p_Y()) \right| \leq \varepsilon, & (E^y) \\ \left| -\frac{1}{n} \log p(x^n, y^n) - H(p()) \right| \leq \varepsilon, & (E^{xy}) \end{cases} \right\}$$

をこの確率分布に関する ε -同時標準集合 (**typical set**) とよび， $A_\varepsilon^{(n)}$ とかくまたこれに属する系列 (x^n, y^n) を ε 同時標準系列 (**typical sequence**) という．

次の定理がなり立つ．

定理 18 (同時 AEP 性) $p : \mathcal{X} \times \mathcal{Y} \rightarrow \mathcal{R}$ に関する， $A_\varepsilon^{(n)}$ について以下が成り立つ．

1.

$$(x^n, y^n) \in A_\varepsilon^{(n)} \Leftrightarrow \begin{aligned} 2^{-n(H(X,Y)+\varepsilon)} &\leq p(x^n) \\ &\leq 2^{-n(H(X,Y)-\varepsilon)} \end{aligned}$$

2. $\exists n_0(\varepsilon), \forall n \geq n_0(\varepsilon),$

$$P\left((X^n, Y^n) \in A_\varepsilon^{(n)}\right) > 1 - \varepsilon$$

3.

- $\#A_\varepsilon^{(n)} \leq 2^{n(H(X,Y)+\varepsilon)}$
- $\exists n_0(\varepsilon), \forall n \geq n_0(\varepsilon), \#A_\varepsilon^{(n)} \geq (1 - \varepsilon)2^{n(H(X,Y)-\varepsilon)}.$

4. $\tilde{X}^n \sim p_{X^n}$ と $\tilde{Y}^n \sim p_{Y^n}$ は (X^n, Y^n) と同じ周辺分布をもつ独立な確率変数列の対とする. このとき

- $P\left((X^n, Y^n) \in A_\varepsilon^{(n)}\right) \leq 2^{-n(I(X;Y)-3\varepsilon)}$
- $\exists n_0(\varepsilon), \forall n \geq n_0(\varepsilon),$

$$P\left((X^n, Y^n) \in A_\varepsilon^{(n)}\right) \geq 2^{-n(I(X;Y)+3\varepsilon)}$$

証明: 1 は標準系列の定義のいい替えである. 2 は,

$$P\left((X^n, Y^n) \notin A_\varepsilon^{(n)}\right) < \varepsilon$$

とも書けるが, この左辺はドモルガンの定理により

$$P\left((X^n, Y^n) \notin A_\varepsilon^{(n)}\right) = P\left((E^x)^c \cup (E^y)^c \cup (E^{xy})^c\right)$$

と書ける. 合併上界式より

$$P\left((E^x)^c \cup (E^y)^c \cup (E^{xy})^c\right) \leq P((E^x)^c) + P((E^y)^c) + P((E^{xy})^c)$$

と上界できる. 一方この右辺の各項は定理8を利用して十分大きな n に関しては任意に小さな正数で上界できる. 3. の一つ目は不等式

$$\begin{aligned} 1 &= \sum_{(x^n, y^n) \in \mathcal{X}^n} p(x^n, y^n) \\ &\geq \sum_{(x^n, y^n) \in A_\varepsilon^{(n)}} p(x^n, y^n) \\ &\geq \sum_{(x^n, y^n) \in A_\varepsilon^{(n)}} 2^{-n(H(X,Y)+\varepsilon)} \quad (1.より) \\ &= \#A_\varepsilon^{(n)} 2^{-n(H(X,Y)+\varepsilon)} \end{aligned}$$

からいえる. これは一変数の場合の性質の証明と同じである. 二つ目については, やはり一変数の場合の証明と同様に示すことができる. 4. について示そう. まず

$$P\left((X^n, Y^n) \in A_\varepsilon^{(n)}\right) = \sum_{(x^n, y^n) \in A_\varepsilon^{(n)}} p(x^n)p(y^n) \quad (4.1)$$

に注意する. 一つ目は

$$(4.1) \leq 2^{-n(H(X)-\varepsilon)} 2^{-n(H(Y)-\varepsilon)} \#A_\varepsilon^{(n)} \quad (4.2)$$

$$\leq 2^{-n(H(X)-\varepsilon)} 2^{-n(H(Y)-\varepsilon)} 2^{n(H(Y)+\varepsilon)} \quad (4.3)$$

$$\leq 2^{-n(I(X;Y)-3\varepsilon)} \quad (4.4)$$

から、二つ目は条件を満たす n について、

$$(4.1) \geq 2^{-n(H(X)+\varepsilon)} 2^{-n(H(Y)+\varepsilon)} \sharp A_\varepsilon^{(n)} \quad (4.5)$$

$$\geq 2^{-n(H(X)+\varepsilon)} 2^{-n(H(Y)+\varepsilon)} (1-\varepsilon) 2^{n(H(X,Y)-\varepsilon)} \quad (4.6)$$

$$\geq (1-\varepsilon) 2^{-n(I(X;Y)-3\varepsilon)} \quad (4.7)$$

のように得られる．証明終わり．

4.3.2 通信路符号化の枠組み

入力アルファベット $\mathcal{X}$, $\sharp \mathcal{X} < \infty$, 出力アルファベット $\mathcal{Y}$, $\sharp \mathcal{Y} < \infty$ の通信路を用いた n ブロック符号の符号器 φ , 復号器 ϕ を以下の関数として表す．

$$\varphi : \{1, \dots, M\} \rightarrow \mathcal{X}^n \quad (4.8)$$

$$\psi : \mathcal{Y}^n \rightarrow \{1, \dots, M\} \quad (4.9)$$

ここで $R = (1/n) \log M$ は通信路 1 記号当たりの伝送レートである．

離散時間無記憶通信路 $p(y^n | x^n) = \prod_{l=1}^n p(y_l | x_l)$ とする．離散時間無記憶通信路は行列 $(p(y | x))_{x \in \mathcal{X}, y \in \mathcal{Y}}$ により決まるのでこれを通信路行列という．無記憶通信路の通信路容量 C はこの通信路行列だけから（以前述べたように、 $\mathcal{X}$ 上の確率分布 $p_X(x)$, $x \in \mathcal{X}$ に関する最適値） $\max_{p_X} I(X; Y)$ と表現できる．

この符号器・復号器の平均（ブロック）誤り確率を

$$P_e^{av} = \frac{1}{M} \sum_{m=1}^M P(\psi(Y^n) \neq \varphi(m))$$

と定義する．ここで $X^n := \varphi(m)$ とおけば和をとる項は無記憶通信路の確率によりすなわち通信路行列により表現できることになる．

以下に証明するのは次の定理である．

定理 19 （通信路符号化順定理）未満の任意の正の伝送レートに対して符号長を一定の長さ以上にとれば平均誤り確率が任意に小さくなるようなブロック符号器及びブロック復号器が存在する．

より厳密には

$$(0 < \forall R < C)(\forall \lambda > 0)(\exists n_0(\lambda, R))(\forall n \geq n_0) \\ (\exists \varphi : \{1, \dots, 2^{\lceil nR \rceil}\} \rightarrow \mathcal{X}^n)(\exists \psi : \mathcal{Y}^n \rightarrow \{1, \dots, 2^{\lceil nR \rceil}\})(P_e^{av} < \lambda)$$

とあらわせる．

定理 20 （通信路符号化逆定理）通信路容量より大きな任意の伝送率 R が与えられたときどんなブロック長 n をとり伝送率 $\lceil nR \rceil$ のどんな n -ブロック符号器、 n -ブロック復号器を用いたとしてもその平均誤り確率は一定の正の値より小さくすることはできない．

この内容を表現する表現の一つは

$$(C < \forall R)(\exists \lambda > 0)(\exists n_0(\lambda, R))(\forall n \geq n_0) \\ (\forall \varphi : \{1, \dots, 2^{\lceil nR \rceil}\} \rightarrow \mathcal{X}^n)(\forall \psi : \mathcal{Y}^n \rightarrow \{1, \dots, 2^{\lceil nR \rceil}\})(P_e^{av} > \lambda)$$

より厳密に表現した形は

$$(C < \forall R)(\exists \lambda > 0)(\forall n > 0) \\ (\forall \varphi : \{1, \dots, 2^{\lceil nR \rceil}\} \rightarrow \mathcal{X}^n)(\forall \psi : \mathcal{Y}^n \rightarrow \{1, \dots, 2^{\lceil nR \rceil}\})(P_e^{av} > \lambda)$$

である。両方とも正しい。

4.3.3 通信路符号化順定理の証明

任意の正数 λ を固定する。 ε を $\lambda/2$ 以下の正数に固定する。まず符号器 $\varphi : \{1, \dots, 2^{\lceil nR \rceil}\} \rightarrow \mathcal{X}^n$ が所与であるとき、本証明で用いる標準系列復号法を定義する。さて証明ではブロック長は n に固定している。従って x^n, y^n などの n ブロック記号列はベクトル記号 $\mathbf{x}, \mathbf{y}$ として表記する。

まずあらゆる $\mathbf{y} \in \mathcal{Y}^n$ に付随する集合 $D(\mathbf{y}) := \{m : (\varphi(m), \mathbf{y}) \in A_\varepsilon^{(n)}\}$ を定める。標準系列復号法 $\psi : \mathcal{Y}^n \rightarrow \{1, \dots, 2^{\lceil nR \rceil}$ を

$$\psi(\mathbf{y}) := \begin{cases} 1 & \text{if } \#D(\mathbf{y}) \neq 1 & \text{Case 1} \\ m' & \text{if } D(\mathbf{y}) = \{m'\} & \text{Case 2} \end{cases} \quad (4.10)$$

と定める。Case 2 は $D(\mathbf{y})$ が任意の単一メッセージからなる集合であることを意味する。いま $\psi(\mathbf{y}) \neq m$ であると仮定しよう。必然的に $D(\mathbf{y}) \neq \{m\}$ である。(対偶は $D(\mathbf{y}) = \{m\} \Rightarrow \psi(\mathbf{y}) = m$ だから。) このとき以下のいずれかが成り立つ。

- $m \notin D(\mathbf{y})$
- $m \in D(\mathbf{y})$

$D(\mathbf{y}) \neq \{m\}$ であるから後者は $D(\mathbf{y})$ には m 以外の他のメッセージを含むことを意味する。すなわち $\cup_{\tilde{m} \neq m} \{\tilde{m} \in D(\mathbf{y})\}$ が成り立つ。以上の論理から、平均誤り確率の解析に活躍する次の関数関係がえられる。

$$1\{\psi(\mathbf{y}) \neq m\} \leq 1\{m \notin D(\mathbf{y})\} + \sum_{\tilde{m} \neq m} 1\{\tilde{m} \in D(\mathbf{y})\} \quad (4.11)$$

$$= 1\{(\varphi(m), \mathbf{y}) \notin A_\varepsilon^{(n)}\} + \sum_{\tilde{m} \neq m} 1\{(\varphi(\tilde{m}), \mathbf{y}) \notin A_\varepsilon^{(n)}\} \quad (4.12)$$

さていよいよ符号器の構成である。符号器はメッセージに付随する符号語の集まり(すなわち符号)として表すことができる。4.3節で述べたように符号は通信路容量 C を達成する入力分布 $p_{\mathbf{X}}$ に従って独立にランダムに生成する。符号(通信路容量ではない) $\mathcal{C} := \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_M\}$ は確率分布

$p_C(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_M) = \prod_{m=1}^M p_{\mathbf{X}}(\mathbf{x}_m)$ に従って生成される．平均誤り確率 P_e^{av} は

$$\begin{aligned} P_e^{av} &= (1/M) \sum_{m=1}^M E 1\{\psi(\mathbf{Y}) \neq m \mid \varphi(m)\} \\ &= (1/M) \sum_{m=1}^M \sum_{\mathbf{y}} p(\mathbf{y} \mid \mathbf{x}_m) 1\{\psi(\mathbf{y}) \neq m\} \end{aligned}$$

と書ける．従って $E_C P_e^{av}$ は

$$\begin{aligned} E_C P_e^{av} &= \frac{1}{M} \sum_m \sum_{\mathbf{y}} E_C p(\mathbf{y} \mid \mathbf{x}_m) 1\{\psi(\mathbf{y}) \neq m\} \\ &\leq \frac{1}{M} \sum_m \sum_{\mathbf{y}} E_C p(\mathbf{y} \mid \mathbf{x}_m) \left\{ 1\{(\mathbf{x}_m, \mathbf{y}) \notin A_\varepsilon^{(n)}\} + \sum_{\tilde{m} \neq m} 1\{(\mathbf{x}_{\tilde{m}}, \mathbf{y}) \in A_\varepsilon^{(n)}\} \right\} \\ &= \sum_{\mathbf{y}} E_C p(\mathbf{y} \mid \mathbf{x}_1) 1\{(\mathbf{x}_1, \mathbf{y}) \notin A_\varepsilon^{(n)}\} + \sum_{\tilde{m} \neq 1} \sum_{\mathbf{y}} E_C p(\mathbf{y} \mid \mathbf{x}_1) 1\{(\mathbf{x}_{\tilde{m}}, \mathbf{y}) \in A_\varepsilon^{(n)}\} \\ &=: (\text{式 1}) + (\text{式 2}) \end{aligned}$$

ここで、 E_C をとる項は m に対して対称であり、一般性を失わず $m = 1$ と置いている．

$$(\text{式 1}) \leq \varepsilon$$

ここで同時 AEP 性の第 2 の性質を用いている．

$$\begin{aligned} (\text{式 2}) &= (M-1) \sum_{\mathbf{y}} \sum_{\mathbf{x}_1} \sum_{\mathbf{x}_2} p(\mathbf{x}_1) p(\mathbf{y} \mid \mathbf{x}_1) p(\mathbf{x}_2) 1\{(\mathbf{x}_2, \mathbf{y}) \in A_\varepsilon^{(n)}\} \\ &= (M-1) \sum_{\mathbf{y}} \sum_{\mathbf{x}_2} p(\mathbf{y}) p(\mathbf{x}_2) 1\{(\mathbf{x}_2, \mathbf{y}) \in A_\varepsilon^{(n)}\} \\ &\leq (M-1) 2^{-n(C-3\varepsilon)} \\ &< 2^{-n(C-R-3\varepsilon)} \end{aligned}$$

ここで、第 1 の等号は E_C をとる項が $\tilde{m} \neq 1$ なる $\tilde{m}$ について対称なので、一般性を失わず $m = 2$ と置いている．最初の不等式は同時 AEP 性の第 4 の性質の 1 つ目を用いている．最後の評価式は n が大きくなると順定理の条件より ε より小さくなる．従って

$$E_C P_e^{av} \leq 2\varepsilon \quad (4.13)$$

が成り立つ． P_e^{av} の符号に関する期待値が定数以下なら実際ある符号が存在して P_e^{av} はその定数以下になるはずである． ε を $\lambda/2$ 以下の正数に固定していたのでこの符号に関して $P_e^{av} < \lambda$ が成り立つ．

前節の $\lambda - M$ システムは符号の最大誤り確率

$$P_e^{max} := \max_{1 \leq m \leq M} P(\psi(\mathbf{Y}) \neq m \mid \varphi(m) := \mathbf{X}) \quad (4.14)$$

を λ で上界するシステムであった。従って、本小節の平均誤り確率に関する結果は弱い結果であるといえるが、実はほぼ同じ強さの結果である。これを説明しよう。まず、 $\varepsilon = \lambda/4$ と設定し、メッセージ数が $2M$ の符号が $P_e^{av} < 2\varepsilon$ を満たすように得られているとしよう。(メッセージ数が2倍になったくらいでは R は漸近的に変化しないことにも注意がいる。) こうして得られた $2M$ 個の符号語を $P(\psi(\mathbf{Y}) \neq m \mid \varphi(m))$ の昇順に並べた符号語のリストを考える。その最初の M 個は $P(\psi(\mathbf{Y}) \neq m \mid \varphi(m)) \leq 4\varepsilon$ を満足する。(もしそうでないなら後半の M 個はすべて 4ε より大きな誤り確率を持つ。すると前半の M 個の誤り確率が0であったとしても後半だけで $2M$ 個全体の平均誤り確率を 2ε より大きくしてしまう。これは矛盾である。) すなわち前半の M 個の符号は $\lambda - M$ システムを成している。このようにある単純な方法で要素を沢山用意しておいてその中から特性の悪い要素をふるいにかけるようにしてす抜く (expurgate) ことは有用な方法である。

4.4 通信路符号化逆定理

さて前々節の最後で述べた制限はどの程度の制約になっているであろうか。実は、制約にはなっていないことが証明できる。別な言葉でいうと、

定理 21 通信路容量より大きな伝送レートを有する $\lambda - M$ システムについて、 λ には M には無関係な正の下限がある。

これを以下に示そう。まず、ある $\lambda - M$ システムを仮定しよう。 $R := (1/T) \log M$ と定義しよう。 $\mathbf{y}$ により入力信号を推定する関数を $\hat{\mathbf{X}}(\mathbf{y})$ とかく。 E を $\mathbf{X} \neq \hat{\mathbf{X}}(\mathbf{Y})$ のとき1そうでないとき0を表す確率変数とする。信号が等確率で選ばれるときの消失誤りも含めて推定誤り確率は $P_e := P(E = 1)$ と定義される。 $\lambda - M$ システムでは当然 $\lambda \geq P_e$ になっている。 $R > C$ ならば P_e には、従って λ にはどのような方法を用いても下回れない正の下限があることを以下で情報理論的な方法を用いて示す。まず、受信側が、入力に対して持つ条件付エントロピー $H(\mathbf{X}|\mathbf{Y})$ には、以下のような下界がある。

$$\begin{aligned} H(\mathbf{X}|\mathbf{Y}) &= H(\mathbf{X}) - I(\mathbf{X}; \mathbf{Y}) \\ &= TR - I(\mathbf{X}; \mathbf{Y}) \\ &\geq T(R - C) \end{aligned}$$

一方,

$$\begin{aligned}
H(\mathbf{X}|\mathbf{Y}) &= H(\mathbf{X}, E|\mathbf{Y}) && (\mathbf{Y} \text{ の条件下で } E \text{ は } \mathbf{X} \text{ の関数}) \\
&= H(E|\mathbf{Y}) + H(\mathbf{X}|\mathbf{Y}, E) && (\text{チェイン規則}) \\
&\leq H(P_e) + H(\mathbf{X}|\mathbf{Y}, E) && (H(P_e) \text{ は二進エントロピー関数}) \\
&= H(P_e) + P(E=1)H(\mathbf{X}|\mathbf{Y}, E=1) + P(E=0)H(\mathbf{X}|\mathbf{Y}, E=0) \\
&= H(P_e) + P_e H(\mathbf{X}|\mathbf{Y}, E=1) && (\mathbf{X} \text{ は } E=0 \text{ のもとで } \mathbf{Y} \text{ の関数}) \\
&\leq H(P_e) + P_e \log M && (4.15) \\
&\leq 1 + P_e \log M && (H(P_e) \leq 1) \\
&= 1 + P_e TR
\end{aligned}$$

なる上界が成り立つ. (4.15) 式では $E=1$ のもとで $\mathbf{X}$ は $\hat{\mathbf{X}}(\mathbf{y})$ 以外の M 個の値をとりうることを用いた. (4.15) 式と左辺の関係をファノの不等式という. まとめると,

$$P_e \geq (R - C)/R - 1/(TR)$$

従って, M を大きくして P_e を任意に小さくできるためには, $(R - C)/R$ は漸近的に非正でなければならない.

左辺の P_e は個別メッセージに対する復号誤り確率の平均, 即ち平均誤り確率であるから, 右辺は $\lambda - M$ の λ すなわち最大誤り確率 $\lambda - M$ に対する下限を与えていることになる.

第5章 通信路容量（離散無記憶通信路）

本章では前節で基本的な役割を演じた定常通信路の通信路容量を特に取り扱いの容易な離散無記憶通信路に限って様々な通信路に関して計算する。

入力アルファベット $\mathcal{X}$ と出力アルファベット $\mathcal{Y}$ をもつ離散無記憶通信路 (Discrete Memoryless Channel, 略して DMC) は通信路行列 $(p(y | x))_{x \in \mathcal{X}, y \in \mathcal{Y}}$ により特徴づけられることを述べた。本節では $\mathcal{X}$ と $\mathcal{Y}$ は何れも有限サイズであることを仮定する。通信路行列は入力を行に出力を列にとることとする。例えば $\mathcal{X} = \{a_1, a_2\}, \mathcal{Y} = \{b_1, b_2, b_3\}$ すなわち 2×3 行列の場合には

$x \backslash y$	b_1	b_2	b_3
a_1	$p(b_1 a_1)$	$p(b_2 a_1)$	$p(b_3 a_1)$
a_2	$p(b_1 a_2)$	$p(b_2 a_2)$	$p(b_3 a_2)$

となる。

DMC に対して通信路容量は入力アルファベット上の確率分布 $\{p_{\mathbf{X}}(x)\}_{x \in \mathcal{X}}$ に関する相互情報量 $I(X; Y)$ の最大値

$$C = \max_{p_{\mathbf{X}}} I(X; Y)$$

と定義されていたことを思い出そう。これを計算するには $I(X; Y) = H(Y) - H(Y | X) = H(X) - H(X | Y)$ 等の様々な表現を利用できる。

通信路容量はアルファベットの要素名とは無関係であるから行列

$$\begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \end{bmatrix}$$

と表現してもよい。

5.1 二元対称通信路 (BSC)

つぎの通信路行列で定義される離散無記憶通信路を二元対称通信路 (Binary Symmetric Channel, 略して BMC) という。

$$\begin{bmatrix} \bar{\varepsilon} & \varepsilon \\ \varepsilon & \bar{\varepsilon} \end{bmatrix}$$

ε ($0 < \varepsilon < 1$) はクロスオーバー確率と呼ばれる． $\bar{\varepsilon}$ は $1 - \varepsilon$ を表す便利な記号であり今後用いることにしよう．二元位相変調を用いた通信路の出力を二元にしていれば通信路は BSC となる．

通信路容量を求めよう．

$$\begin{aligned}
 I(X; Y) &= H(Y) - H(Y | X) \\
 &= H(Y) - \sum_x p(x) H(Y | X = x) \\
 &= H(Y) - \sum_x p(x) h(\varepsilon) \\
 &= H(Y) - h(\varepsilon) \\
 &\leq 1 - h(\varepsilon)
 \end{aligned}$$

ここで二元エントロピー関数 $h(\varepsilon) = \varepsilon \log(1/\varepsilon) + \bar{\varepsilon} \log(1/\bar{\varepsilon})$ を用いている．等号が達成されるには p_Y が一様分布 $\{1/2, 1/2\}$ となればよいがそのためには p_X が一様分布になれば十分である．すなわち通信路容量 $C(\varepsilon)$ は $1 - h(\varepsilon)$ となる． $h(\varepsilon) = h(\bar{\varepsilon})$ であり $C(\varepsilon)$ は $\varepsilon = 1/2$ に関して左右対称である． $\varepsilon = 0, 1$ のときに 1, $\varepsilon = 1/2$ のときに 0 となる．

5.2 二元消失通信路

つぎの通信路行列で定義される離散無記憶通信路を二元消失通信路 (Binary Erasure Channel, 略して BEC) という．

$$\begin{bmatrix} \bar{\varepsilon} & \varepsilon & 0 \\ 0 & \varepsilon & \bar{\varepsilon} \end{bmatrix}$$

入力アルファベットは順番に $\{0, 1\}$ ，出力アルファベットは順番に $\{0, e, 1\}$ と解釈する．出力記号の e は入力が出力で消失してしまうことを意味し消失シンボルと呼ばれる．これは，前小節と同様に二元位相変調により得られる通信路のモデルであり，非常に有用である．

通信路容量を求めよう．

$$\begin{aligned}
 I(X; Y) &= H(Y) - H(Y | X) \\
 &= H(Y) - h(\varepsilon)
 \end{aligned}$$

$H(Y)$ を計算するためまず， $\{\text{非消失}, \text{消失}\}$ 上の確率分布を Y_1 とすると Y_1 は Y の関数である．

$$\begin{aligned}
 H(Y) &= H((Y_1, Y)) \\
 &= H(Y_1) + H(Y | Y_1) \\
 &= h(\varepsilon) + \bar{\varepsilon} H(Y | Y_1 = \text{非消失}) + \varepsilon H(Y | Y_1 = \text{消失}) \\
 &= h(\varepsilon) + \bar{\varepsilon} H(Y | Y_1 = \text{非消失})
 \end{aligned}$$

すなわち $I(X; Y) = \bar{\varepsilon} H(Y | Y_1 = \text{非消失})$ である。ここで

$$H(Y | Y_1 = \text{非消失}) \leq 1,$$

を用いると, $C(\varepsilon) = 1 - \varepsilon$, が得られる。ただし達成は用いた不等式の等号条件, すなわち $p_{\mathbf{X}}$ が一様分布の場合である。この BEC は ε の広い領域で高い通信路容量を与える優れた通信路である。

5.3 対称通信路

対称通信路は, 二元対称通信路, 二元消失通信路を一般化した通信路である。 $\mathcal{Y}$ が $\bigsqcup_{k \in \mathcal{K}} \mathcal{Y}_k$ のように直和に分解されており, 各 k について $\mathcal{X} \times \mathcal{Y}_k$ 上で通信路行列は行置換, 列置換に関して対称であるとき, 通信路は対象通信路であるという。例えば

$$\begin{bmatrix} \mu\bar{\varepsilon} & \varepsilon & \bar{\mu}\bar{\varepsilon} \\ \bar{\mu}\bar{\varepsilon} & \varepsilon & \mu\bar{\varepsilon} \end{bmatrix}$$

は $\mathcal{Y} = \{0, 1\} \sqcup \{e\}$ の分解において上記の性質を満たしている。定義から各 $k \in \mathcal{K}$ について $p_k = P(Y \in \mathcal{Y}_k | X = x)$ は x の選択によらない。また, $k \in \mathcal{K}$ について $\mathcal{Y}_k$ 上に X と相関をもって分布する確率変数 Y_k を考えることができる。

$$\begin{aligned} I(X; Y) &= H(Y) - H(Y | X) \\ &= (H(\{p_k\}) + \sum_k p_k H(Y_k)) - \sum_x p(x) (H(\{p_k\}) - H(Y | X = x)) \\ &= \sum_k p_k (H(Y_k) - H(Y_k | X = x_0)) \\ &\leq \sum_k p_k (\log \#\mathcal{Y}_k - H(Y_k | X = x_0)) \end{aligned}$$

二番目の等号では x は任意であり x_0 に固定している。また最後の不等式は $p(x) = 1/\#\mathcal{X}$ と置いて得ている。つまりこの入力確率分布において右辺が通信路容量を与えている。

例で与えた 2×3 の通信路の容量は $C = \bar{\varepsilon}(1 - h(\mu))$ となる。

5.4 アルファベット分離型通信路

これは入力アルファベットおよび出力アルファベットが $\mathcal{X} = \mathcal{X}_1 \sqcup \mathcal{X}_2, \mathcal{Y} = \mathcal{Y}_1 \sqcup \mathcal{Y}_2$ のように2つに分離し $\mathcal{X}_1$ の出力は $\mathcal{Y}_1$ にしか届かず $\mathcal{X}_2$ の出力は $\mathcal{Y}_2$ にしか届かない, すなわち, 通信路が2つの互いに行き来のない通信路に分離している場合をアルファベット分離型通信路という。各々の通信路容量を C_1 と C_2 であると仮定する。

定理 22 アルファベット分離型の通信路の容量 C は

$$C = \log(2^{C_1} + 2^{C_2})$$

で与えられる．容量を達成するのは分離された通信路 1 と 2 に配分される確率が 2_1^C 対 2_2^C の比であり，各々の分離通信路内では各々の通信路容量を達成する（条件付き）確率の配分になっている場合である．

証明：

$$\begin{aligned} I(X; Y) &= H(Y) - H(Y | X) \\ &= h(\lambda) + \lambda H(Y_1) + \bar{\lambda} H(Y_2) \\ &\quad - \lambda H(Y_1 | X_1) - \bar{\lambda} H(Y_2 | X_2) \\ &= h(\lambda) + \lambda I(X_1; Y_1) + \bar{\lambda} I(X_2; Y_2) \\ &\leq h(\lambda) + \lambda C_1 - \bar{\lambda} C_2 \\ &= h(\lambda) + \lambda \log 2^{C_1} - \bar{\lambda} \log 2^{C_2} \\ &= \lambda \log \frac{2^{C_1}}{\lambda} + \bar{\lambda} \log \frac{2^{C_2}}{\bar{\lambda}} \\ &\leq (\lambda + \bar{\lambda}) \log \frac{2^{C_1} + 2^{C_2}}{\lambda + \bar{\lambda}} \\ &= \log(2^{C_1} + 2^{C_2}) \end{aligned}$$

第二不等式は，対数和不等式である．通信路に配分される確率の比はこの等号条件から決まる．

5.5 二元非対称通信路

二元非対称通信路の通信路行列は

$$\begin{bmatrix} \bar{\varepsilon} & \varepsilon \\ \bar{\mu} & \mu \end{bmatrix}$$

である．一般性を失うことなく $\varepsilon < \bar{\mu}$ を仮定する．入力確率分布を $[\bar{\lambda}, \lambda]$ ，出力確率分布を $[\bar{q}, q]$ と置くと必然的に $q = \bar{\lambda}\varepsilon + \lambda\bar{\mu}$ ，つまり q は確率 ε と確率 $\bar{\mu}$ を λ 対 $\bar{\lambda}$ に内分する確率である．さて

$$\begin{aligned} I(X; Y) &= H(Y) - H(Y; X) \\ &= h(q) - \lambda h(\bar{\mu}) - \bar{\lambda} h(\varepsilon) \\ &= h(\bar{\lambda}\varepsilon + \lambda\bar{\mu}) - \lambda h(\bar{\mu}) - \bar{\lambda} h(\varepsilon) \end{aligned} \tag{5.1}$$

と表現しておく．今二元エントロピー関数 $h(q)$ は q 軸を出力確率にとった上に凸な関数でありそのグラフは点 $(q, h(q))$ において λ を変化させた

軌跡としてみる．同様にして $((q :=) \bar{\lambda}\varepsilon + \lambda\bar{\mu}, \lambda\bar{\mu}) - \lambda h(\bar{\mu}) - \bar{\lambda}h(\varepsilon)$ は q 軸上点 $(\varepsilon, h(\varepsilon))$ と $(\bar{\mu}, h(\bar{\mu}))$ を結ぶ線分 $l(q)$ のグラフである．従って式 (5.1) はエントロピー関数と線分の縦軸方向の差分である．この差分を q (あるいは λ) を変化させたときの最大値が通信路容量を与えることになる．今この最大値を与える q を q^* とするとそこでの $h(q)$ の接線 $\tilde{l}(q) := h(q^*) + (q - q^*)h'(q^*)$ は線分 $l(q)$ と平行である．また $\tilde{l}(q)$ と $h(q)$ の差分は非負でありエントロピー関数の接触変換すなわち $D(q \parallel q^*)$ である． $\tilde{l}(q)$ と $l(q)$ のグラフとしての差 (C) は $D(q \parallel q^*)$ の $q = \varepsilon$ における値, および $q = \bar{\mu}$ における値にも等しくなっている．言い換えると $q = q^*$ において

$$C = D(\varepsilon \parallel q) = D(\bar{\mu} \parallel q)$$

が成立する．逆に言えばこの方程式の解が q^* および C を与えることになる．

5.6 有本-Blahut アルゴリズム [Arimoto:72][Brahut:72]

通信路行列 $P(y \mid x)$ と入力確率 $\lambda(x)$ を仮定すると入出力の同時確率分布が定まるがこれから必然的に周辺分布としての出力確率 $q(y)$ も定まることになる．一方 $\lambda(x)P(y \mid x) = q(y)Q(x \mid y)$ を満たす出力に関する入力の条件付き確率分布 $Q(x \mid y)$ が定まる．これは後ろ向き通信路を定義していると解釈できる．

相互情報量 $I(X; Y)$ は $\lambda(x)$ と $P(y \mid x)$ の汎関数であり $I(\lambda, P)$ と記述しても良い．これは通常の変換を用いて書くこともできるが, 後ろ向き通信路 $Q(x \mid y)$ に関する最大化問題としても定式化できる：

$$I(\lambda, P) = \sum_{x,y} \lambda(x)p(y \mid x) \log \frac{p(y \mid x)}{q(y)} = \max_Q \sum_{x,y} \lambda(x)p(y \mid x) \log \frac{Q(x \mid y)}{\lambda(x)},$$

これを利用すると通信路容量は

$$C = \max_{\lambda} \max_Q \sum_{x,y} \lambda(x)p(y \mid x) \log \frac{Q(x \mid y)}{\lambda(x)}$$

のように凸関数の $\lambda(x)$ と $Q(x \mid y)$ に関する 2 重最大化問題として定式化できる．これを交互最大化問題として解くアルゴリズムが有本-Blahut 法であり多くの場合安定である程度効率的に通信路容量を計算してゆくことができる．

交互最大化は以下の 2 つのステップの繰り返しとなる．

- $Q(x \mid y) \propto \lambda(x)p(y \mid x)$ となるように後ろ向き通信路 $Q(x \mid y), x \in \mathcal{X}, y \in \mathcal{Y}$ を定める．
- $\log \lambda(x) \propto \sum_y p(y \mid x) \log Q(x \mid y)$ となるように入力確率分布 $\lambda(x), x \in \mathcal{X}$ を定める．

第6章 通信路容量（加法的白色ガウス通信路）

連続時間加法的白色ガウス通信路は通信理論の標準モデルでありその理論的取り扱いには情報理論の重要な領域である。取り扱いには確率過程論の高度な知識を必要とする。そこで情報理論としての厳密さは保持しつつ、本書のレベルを超える部分については証明は直感的ものに置き換えて本質を伝えるよう努める。

6.1 微分エントロピー

通信路符号化の対象としたいのは、最終的に入出力が連続時間連続値信号となる通信路、いわゆるアナログ通信路である。我々は連続信号はサンプリングにより離散時間連続値信号に変換しうることを、また離散時間連続値信号は適切な平滑化によりそれをサンプル値列とする連続時間信号に逆変換できることを仮定する。従って当対象は離散時間連続値信号である。これまで考えてきたのは離散値の情報源、あるいは通信路であった。本節では、連続値の情報源（したがって通信路）においてシャノンのエントロピーに代わる役割をもつ微分エントロピーについて述べる。

6.1.1 微分エントロピーの定義

実数値確率変数 $X : \Omega \rightarrow \mathcal{X}$ について、累積分布関数 $F(x) = P(X \leq x)$ が微分 $f(x) = F'(x)$ すなわち確率密度関数をもつと仮定する。 $f(x)$ の台を S で表す。

定義 7 (微分エントロピー): 確率密度関数 $f(x)$ をもつ実数値確率変数 X について

$$\begin{aligned} h(X) &:= - \int_S f(x) \log f(x) dx \\ &= E\left[\log \frac{1}{f(X)}\right] \end{aligned}$$

を X の微分エントロピーとよぶ。

2つの例を挙げる。

例 1 (一様分布)

$$\begin{aligned} f(x) &= \frac{1}{a}, 0 \leq x \leq a \\ h(X) &= E \log a = \log a \end{aligned} \quad (6.1)$$

例 2 (正規分布)

$$\begin{aligned} f(x) &= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} \\ h(X) &= E[\log \sqrt{2\pi}\sigma + \log e^{\frac{X^2}{2\sigma^2}}] \\ &= \frac{1}{2} \log(2\pi e\sigma^2) \end{aligned}$$

確率変数 X を幅 $\Delta > 0$ で量子化した離散確率変数 X^Δ を以下のように定義する. 平均値の定理より

$$f(x_i)\Delta := \int_{i\Delta}^{(i+1)\Delta} f(x)dx = P(i\Delta \leq X \leq (i+1)\Delta)$$

を満たす x_i が区間 $[i\Delta, (i+1)\Delta]$ に存在する. そこで X^Δ を

$$X^\Delta = x_i \quad \text{if} \quad i\Delta \leq X \leq (i+1)\Delta$$

のように定義する. $p_i = P(X^\Delta = x_i)$ は $f(x_i)\Delta$ に等しいことに注意する.

定理 23 (微分エントロピーとシャノンエントロピーの関係)

$$H(X^\Delta) \rightarrow h(X) - \log \Delta \quad \text{as} \quad \Delta \rightarrow 0.$$

証明:

$$\begin{aligned} H(X^\Delta) &= \sum_i p_i \log \frac{1}{p_i} \\ &= E \log \frac{1}{f(X^\Delta)\Delta} \\ &= E \log \frac{1}{f(X^\Delta)} - E \log \Delta \\ &\rightarrow h(X) - \log \Delta \end{aligned}$$

証明終

6.1.2 同時微分エントロピー，条件付き微分エントロピー

2つの実数値確率変数 (X, Y) の同時確率密度関数についても

$$f(x, y) := \frac{\partial^2}{\partial x \partial y} P(X \leq x, Y \leq y)$$

が定義される．

定義 8 同時微分エントロピーは以下のように定義される．

$$h(X, Y) := E \log \frac{1}{f(X, Y)} \quad (6.2)$$

X に関する Y の条件付き確率密度は

$$f(y|x) := \frac{f(x, y)}{f(x)}$$

で与えられる．

定義 9 X に関する Y の条件付き微分エントロピーを

$$h(Y|X) = E \log \frac{1}{f(Y|X)}$$

と定義する．

定理 24

$$h(Y|X) = h(X, Y) - h(X) \quad (6.3)$$

すなわち，微分エントロピーに関するチェインルール

$$h(X, Y) = h(X) + h(Y|X) \quad (6.4)$$

が成り立つ．

以上の定義は，さらに多変数についても一般化される．例えば， $X_1, X_2, X_3, \dots, X_n$ について同時確率密度関数 $f(x_1, x_2, x_3, \dots, x_n)$ から同時微分エントロピー $h(X_1, X_2, X_3, \dots, X_n)$ が定義される．これらについて， $X^n, \mathbf{X} = X^n = X_1^n, f(x^n), f(\mathbf{x}) = f(x^n) = f(x_1^n)$ 等の記法の利用が可能となる．

例 3 n 次元確率ベクトル $\mathbf{X}$ の多次元確率密度関数がベクトル $\mu = (\mu_i)$ ，および $n \times n$ 正定値行列 K を用いて

$$f(\mathbf{x}) = \frac{1}{\sqrt{2\pi}^n |K|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \mu)K^{-1}(\mathbf{x} - \mu)^T\right\}$$

と表されるとき， $\mathbf{X}$ は平均ベクトル μ ，分散共分散行列 K をもつ多次元正規分布に従うという． $K = (K_{ij})$ の (i, j) 要素は $K_{ij} = E(X_i - \mu_i)(X_j - \mu_j)$ で与えられる．このとき， $\mathbf{X} \sim N(\mu, K)$ と略記する．

多変数の確率密度についても微分エントロピーが定義できる． n 変数の場合， $h(\mathbf{X}) = h(X^n) = h(X_1^n)$ 等のような記法で表すことにする．

例 4 n 次元正規分布に従う $\mathbf{X}$ の微分エントロピーは次のように計算できる．

$$\begin{aligned}
 h(\mathbf{X}) &= h(X_1, X_2, \dots, X_n) = E \log \frac{1}{f(\mathbf{X})} \\
 &= \log \sqrt{2\pi}^n |K|^{1/2} + \frac{1}{2} \{E(\mathbf{x} - \mu)K^{-1}(\mathbf{x} - \mu)^T\} \log e \\
 &= \frac{1}{2} \log(2\pi)^n |K| + \frac{1}{2} \text{tr}(K^{-1}K) \log e \\
 &= \frac{1}{2} \log(2\pi e)^n |K|
 \end{aligned} \tag{6.5}$$

また，チェインルールは以下のように一般化できる．

例 5

$$h(X^n) = \sum_{i=1}^n h(X_i | X^{i-1})$$

すなわち，シャノン理論と記号の上でも対応した微分エントロピーの情報理論が成り立つことが分かる．

6.1.3 KL ダイバージェンス，相互情報量

次に，シャノン情報量の理論における鍵概念であった KL ダイバージェンスについて考えよう．

定義 10 多次元実空間 $\mathcal{X}$ 上の 2 つの連続な確率密度を $f(x), g(x)$ とする． $g(x)$ に関する $f(x)$ の KL ダイバージェンスを

$$D(f||g) = \int f(x) \log \frac{f(x)}{g(x)} dx$$

と定義する．

定理 25

$$D(f||g) \geq 0$$

等号は $f(\cdot) = g(\cdot)$ のときのみ．

証明:

$$\begin{aligned}
 D(f||g) &= \lim_{\Delta \downarrow 0} E_f \log \frac{f(X_f^\Delta) \Delta}{g(X_g^\Delta) \Delta} \\
 &= \lim_{\Delta \downarrow 0} D(\{f(X_f^\Delta) \Delta\} || \{g(X_g^\Delta) \Delta\}) \\
 &\geq 0
 \end{aligned} \tag{6.6}$$

最後から2番目の式は離散分布に対するKLダイバージェンスであり、最後の不等式にはその非負性を用いている。従って等号に関する事実もこのことと密度の連続性から分かる。証明終

- 同じ核文字が用いられるのは同じ値を取る理由による。
- 直接 Jensen の不等式を用いて

$$\begin{aligned} -E_f \log \frac{g(X)}{f(X)} &\geq -\log E_f \frac{g(X)}{f(X)} \\ &\geq 0 \end{aligned}$$

と証明しても良い。

定義 11 連続な確率密度をもつ確率変数 X と Y の相互情報量は

$$I(X; Y) = D(f_{XY}(x, y) \parallel f_X(x)f_Y(y))$$

と定義される。

- $I(X; Y) \geq 0$ であり、等号は X と Y が独立のときだけである。
-

$$\begin{aligned} I(X; Y) &= h(X) + h(Y) - h(X, Y) \\ &= h(X) - h(X | Y) \\ &= h(Y) - h(Y | X) \end{aligned}$$

例 6 1次元密度をもつ X について

$$h(aX) = h(X) + \log |a|$$

証明: $Y = aX$ とおく。 $f_Y(ax) = f_X(x) \left| \frac{dx}{dy} \right| = f_X(x) \frac{1}{|a|}$ に注意すると、

$$\begin{aligned} h(aX) &= E \log \frac{1}{f(aX)} \\ &= E \log \frac{1}{f_X(X) \left| \frac{dx}{dy} \right|} \\ &= h(X) + \log |a| \end{aligned}$$

証明終

定理 26

$$h(A\mathbf{X}) = h(\mathbf{X}) + \log |\det A|$$

定理 27 (最大エントロピー分布の正規性) $\mathbf{X}$ は平均が $E\mathbf{X} = \mathbf{0}$ を満たし、分散共分散行列 $E\mathbf{X}^T\mathbf{X}$ が K で与えられる n 次元確率変数とする。このとき、

$$h(bX) \leq \frac{1}{2} \log(2\pi e)^n | \det K |$$

であり、等号は多次元正規分布のときに成り立つ。

証明: $\mathbf{X}$ の密度関数を f とおく。 f とおなじ平均、分散共分散行列をもつ多次元正規分布の密度関数を ϕ すなわち

$$\phi(\mathbf{x}) = \frac{1}{\sqrt{2\pi}^n |K|^{1/2}} \exp\left\{-\frac{1}{2} \mathbf{x} K^{-1} \mathbf{x}^T\right\}$$

とおく。この関数の対数は $\mathbf{X}$ についての二次形式であるから、 f でも ϕ でも同じ期待値をもつ、すなわち $E_f - \log \phi(\mathbf{X}) = E_\phi - \log \phi(\mathbf{X})$ 。これより

$$0 \leq D(f \parallel \phi) = h(\phi) - h(f)$$

証明終

注意 3 微分エントロピーは、シャノンのエントロピーと似ており、確率密度 $p(X)$ の負対数 $-\log p(X)$ の X に関する期待値であるが確率変数の値域の変数変換に関して不変ではないことが今までの議論からも分かるであろう。この点が離散確率分布に基づくシャノンのエントロピーとの大きな違いである。一方、 KL ダイバージェンスは値域の変数変換に関して不変である。したがって相互情報量も X, Y の各々独立な変数変換について不変である。この理由で両者に関しては離散の場合と同じ記号を用いているのである。

6.2 離散時間加法的白色ガウス通信路の通信路容量

離散時間連続無記憶通信路は任意の時刻 n について入力系列が X_1^n 与えられた条件下での出力系列 Y_1^n の確率密度が

$$p(y_1^n | x_1^n) = \prod_{l=1}^n p(y_l | x_l)$$

で与えられるような通信路である。加法的白色ガウス雑音 (Additive White Gaussian Noise, 略して AWGN) 通信路とは各時刻における入出力関係が加法的すなわち $Y_l = X_l + N_l$ であって雑音 $N_l = Y_l - X_l$ は独立同一のガウス密度に従う場合である。すなわち各時刻の確率密度 $p(y | x)$ は

$$p(y | x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{y-x}{\sigma}\right)^2\right\}$$

の形を取る．ここで σ^2 は雑音の分散である．

物理系では雑音は独立性のある様々な雑音の重ね合わせとして現れると考えられる．例えば電気・電子回路では電気信号が，無線通信路では電磁波が，伝搬する際に，電磁波の散乱により，ショット雑音の無数の和が加わる．一方，独立性のある多くの雑音の和はガウス分布に従うことが知られている（中心極限定理）．またショット雑音はインパルス性が高いため時間的相関の影響は殆どない．そのため雑音には通常の周波数でサンプルする限り時間相関が殆どない．そのために上記のモデルが非常によくフィットする．

さてこれまで学んだように連続通信路であっても入出力の間で共有できる相互情報量は定義され，その入力確率分布に関する最大値が符号化率の達成可能な下限となる．相互情報量は例えば

$$I(X; Y) = h(Y) - h(Y | X) \quad (6.7)$$

のように表現される．そこで加法的白色ガウス雑音通信路について通信路容量を考えてみよう．ただし，入力信号には利用可能な電力に実際上の制限があるので，仮ではあるが $EX^2 \leq Q$ という制限下で求めてみよう．

$$\begin{aligned} C(Q) &= \max_{EX^2 \leq Q} I(X; Y) \\ &= \max_{EX^2 \leq Q} (h(Y) - h(Y | X)) \\ &= \max_{EX^2 \leq Q} h(Y) - \frac{1}{2} \log(2\pi e \sigma^2) \\ &= \max_{EY^2 \leq Q + \sigma^2} h(Y) - \frac{1}{2} \log(2\pi e \sigma^2) \\ &\leq \frac{1}{2} \log(2\pi e (Q + \sigma^2)) - \frac{1}{2} \log(2\pi e \sigma^2) \\ &= \frac{1}{2} \log\left(1 + \frac{Q}{\sigma^2}\right) \end{aligned} \quad (6.8)$$

ここで等号を達成するのは最大エントロピー分布の正規性により $Y \sim N(0, Q + \sigma^2)$ のとき，すなわち $X \sim N(0, Q)$ の時である．

さて各時刻における入力電力の制限は制御可能であり実際の電力制限は平均的な制約である．とすれば本来の最大化は

$$(1/n) \max_{Q_1 + Q_2 + \dots + Q_n \leq nQ} \max_{EX_l^2 \leq Q_l} I(X_1^n; Y_1^n)$$

のように行われるべきである．ところが通信路の無記憶性のため

$$I(X_1^n; Y_1^n) \leq \sum_l I(X_l; Y_l)$$

が成り立つ．従って上記の最大化問題は

$$\begin{aligned}
 C^{(n)}(Q) &\leq (1/n) \max_{Q_1+Q_2+\dots+Q_n \leq nQ} \sum_{l=1}^n C(Q_l) \\
 &= \max_{(1/n)(Q_1+Q_2+\dots+Q_n) \leq Q} (1/n) \sum_{l=1}^n C(Q_l) \\
 &= C(Q)
 \end{aligned} \tag{6.9}$$

なる上界を持つ．結局 $C(Q)$ が平均電力 Q により制限された通信路容量（単位 bit/sample ）であることになる．このことからさらに入力信号は電力 Q の白色ガウス雑音であることも分かる．

6.3 連続時間帯域制限加法的白色ガウス通信路の通信路容量

本節では連続時間信号を入出力とする通信路の中で最も基本的である加法的白色ガウス雑音通信路（雑音電力スペクトル密度は $N_0/2$ （単位は Watt/Hz ））を考える．また利用する周波数帯域が $|f| \leq W$ のように制限されている場合を考え，通信路には入力の平均電力が最大 P (Watt) の制約をもつ場合を考える．

1. 通信路の入力を $x(t)$ ，出力を $y(t)$ ，加法的白色雑音を $z(t)$ とするとき $y(t) = x(t) + z(t)$ が成り立っている．ただし加法的白色ガウス雑音とは $z(t)$ は期待値が 0 の定常確率過程であって以下の 2 つが成り立つものをいう．

- 任意個の任意時点におけるサンプル値 $(z(t_1), z(t_2), \dots, z(t_k))$ は k 次元ガウス分布に従う．
- 自己相関関数 $r(\tau) := E z(t) z(t + \tau)$ は一定値の Fourier スペクトル密度：

$$\int_{-\infty}^{\infty} r(\tau) e^{-2\pi f \tau} d\tau = \frac{N_0}{2}$$

をもつ．

2. 入力 $x(t)$ が帯域制限を受けているとはあるスペクトル密度 $X(f)$ を用いて

$$x(t) = \int_{-W}^W X(f) e^{2\pi f t} df$$

のように表現できることをいう．

このとき $x(t)$ はその標本点 $\{x(l/(2W))\}_{l \in \mathbf{Z}}$ から以下のように再現される（サンプリング定理）また，この表現は一意でもある．

$$x(t) = \sum_{l \in \mathbf{Z}} x\left(\frac{l}{2W}\right) \text{sinc}(2Wt - l)$$

ただし $\text{sinc}(u) := \frac{\sin(\pi u)}{\pi u}$.

証明: 周期 $2W$ をもつ周波数関数 $X(f)$ は, フーリエ係数

$$x_n = \frac{1}{2W} \int_{-W}^W X(f) e^{-2\pi j n \frac{f}{2W}} df. \quad (6.10)$$

を用いて

$$X(f) = \sum_{n=-\infty}^{\infty} x_n e^{2\pi j n \frac{f}{2W}} \quad (6.11)$$

のようなフーリエ級数として表される. さて, 時間信号 $x(t)$ が帯域 $(-W, W)$ に制限されたスペクトル密度をもつとすると, 周期 $2W$ の周波数スペクトル密度 $X(f)$ を用いて

$$x(t) = \int_{-W}^W X(f) e^{-2\pi j f t} df. \quad (6.12)$$

のように表される.

ここで式 (6.12) に, $t = n/(2W)$ を代入すると,

$$x\left(\frac{n}{2W}\right) = 2W x_n \quad (6.13)$$

が成り立つ. すなわち数列 $\{2W x_n\}$ とは $x(t)$ の時間周期 $1/(2W)$ によるサンプル値である. この事実 (6.13) を用い, さらに式 (6.12) に式 (6.11) を代入すると,

$$x(t) = \sum_n \int_{-W}^W x\left(\frac{n}{2W}\right) \frac{1}{2W} e^{-2\pi j f (t - \frac{n}{2W})} df \quad (6.14)$$

$$= \sum_n x\left(\frac{n}{2W}\right) \int_{-1/2}^{1/2} e^{-2\pi j f' (2Wt - n)} df' \quad (6.15)$$

という表現が得られる.

式 (6.15) の和の中に現れる積分は次のよう計算できる. $u = 2Wt - n$ と置こう.

$$\begin{aligned} \int_{-1/2}^{1/2} e^{2\pi j f' u} df' &= \frac{e^{2\pi j (1/2)u} - e^{2\pi j (-1/2)u}}{2\pi j u} \\ &= \frac{\sin \pi u}{\pi u} \end{aligned}$$

即ち,

$$x(t) = \sum_n x\left(\frac{n}{2W}\right) \frac{\sin \pi (2Wt - n)}{\pi (2Wt - n)} \quad (6.16)$$

なる表現が得られた。ここで

$$\text{sinc}(u) = \frac{\sin \pi u}{\pi u}$$

と名付けたことに注意しよう。sinc(u) は振動する関数であるが、 y 軸に関して左右対称で包絡線 $1/|\pi u|$ をもつ。また、

$$\text{sinc}(n) = \begin{cases} 1 & \text{if } n = 0 \\ 0 & \text{if } n \neq 0 \end{cases} \quad (6.17)$$

を満足する。証明終

3. 等価離散時間 AWGN 通信路

$$y_W(t) := x(t) + z_W(t), t \in \mathcal{R}$$

ただし $z_W(t)$ は $z(t)$ に $z_W(t) := z(t) * 2W \text{sinc}(2Wt)$ のようにローパスフィルタを掛けた信号である。（すなわち周波数領域において $Z_W(f) = Z(f)1\{|f| \leq W\}$ のように帯域制限をかけた信号である。）このようにするとサンプリング定理により

$$y_W\left(\frac{l}{2W}\right) = x\left(\frac{l}{2W}\right) + z_W\left(\frac{l}{2W}\right), l \in \mathbf{Z}$$

が成り立つ。すなわち、制限された帯域内においてもとのアナログ信号は等価な離散時間 AWGN 通信路に変換されたことになる。ここで、

$$\{z_W\left(\frac{l}{2W}\right)\}_{l \in \mathbf{Z}}, \text{ i.i.d., in } N(0, N_0 W),$$

が成り立っている。

証明: $z_W(t)$ は $z(t)$ の線形関数であるから平均 0 のガウス分布に従う。独立性を示そう。まず sinc 関数の直交性

$$\int du \text{sinc}(u - k) \text{sinc}(u - l) = \delta_{kl}$$

を証明しておく。sinc(t) は関数 $1\{|f| \leq 1/2\}$ の Fourier 逆変換であること、およびパーセバルの等式を用いると

$$\int_{-\infty}^{\infty} \text{sinc}(t - n) \text{sinc}(t - m) dt \quad (6.18)$$

$$\begin{aligned} &= \int_{-\infty}^{\infty} df 1\left\{f \in \left(-\frac{1}{2}, \frac{1}{2}\right)\right\} e^{-2\pi j f k} \cdot 1\left\{f \in \left(-\frac{1}{2}, \frac{1}{2}\right)\right\} e^{2\pi j f l} \\ &= \int_{-\infty}^{\infty} df 1\left\{f \in \left(-\frac{1}{2}, \frac{1}{2}\right)\right\} e^{-2\pi j f (k-l)} \end{aligned} \quad (6.19)$$

$$= \text{sinc}(k - l) \quad (6.20)$$

$$= \delta_{kl}. \quad (6.21)$$

となる。最後の δ_{lk} はクロネッカーのデルタ $\delta_{lk} = 1\{k = l\}$ である。この直交性から即

$$\int du \operatorname{sinc}(2Wu - k) \operatorname{sinc}(2Wu - l) = \frac{1}{2W}$$

がいえる。次に sinc 関数が対称性であることから

$$z_W\left(\frac{k}{2W}\right) = \int du z(u) 2W \operatorname{sinc}(2Wu - k)$$

がいえることに注意しておく。

以上の準備のもとに目的とする事実を示そう。白色ガウス雑音の性質 $Ez(u)z(v) = (N_0/2)\delta(u - v)$ に注意すると

$$\begin{aligned} Ez_W\left(\frac{k}{2W}\right)z_W\left(\frac{l}{2W}\right) &= \int \int dudv Ez(u)z(v) 2W \operatorname{sinc}(2Wu - k) 2W \operatorname{sinc}(2Wv - l) \\ &= ((N_0/2)(2W)^2) \int du \operatorname{sinc}(2Wu - k) \operatorname{sinc}(2Wu - l) \\ &= (N_0W) \delta_{kl} \end{aligned}$$

となる。証明終

4. 次に電力制約を考慮する。適切な長時間 T において平均電力制約は

$$E \frac{1}{T} \int_0^T x(t)^2 dt \leq P$$

と表現されるが、等価離散時間 AWGN 通信路では

$$E \frac{1}{2WT} \sum_{l=1}^{2WT} x\left(\frac{l}{2W}\right)^2 \leq P$$

と表現される。

5. 前節の結果より平均電力が P で制限された連続時間帯域制限加法的白色ガウス通信路の通信路容量 $C(\text{bit/sec})$ は

$$\begin{aligned} C &= \frac{1}{2} \log\left(1 + \frac{P}{N_0W}\right) \times (2W) \\ &= W \log\left(1 + \frac{P}{N_0W}\right) \end{aligned}$$

となる。