情報とは何か
—情報の世界を数学でさぐる—

柳　研二郎
城西大学理学部数学科

1　情報科学の発展の歴史と目的

この講義は 2011 年 10 月 18 日広島県立広高等学校での出前講義および 2015 年 12 月 3 日城西大学での理学部数学科講演会（1 号館 4 階 406 教室）に対して講義したものに基づく.

今日の情報化時代を生み出した理論家のうち, もっとも代表的な 4 人を挙げるとしたら, 常識的には Shannon(シャノン), Wiener(ウィーナー), von Neumann(フォン・ノイマン), Kolmogorov(コルモゴルフ) であろう. この 4 人の研究対象は必ずしも同じでないが, 等しく 1940 年前後に情報科学上の重要な研究をし, 情報量を表わすエントロピーの概念をそれぞれの形において導き出したという共通点がある. また彼らが情報系のモデルとして生物の生命活動に憧れにも似た関心をよせていた点でも共通している. さらにそろって数学の分野にも天才的な才能をもっていたが, 理論の応用面にも意をそそぎ, コンピュータの開発には何らかの形でタッチしてきたのである. 順を追って述べよう.

1.1　Shannon

Shannon の仕事が高く評価される理由に, 彼が「情報とは何か」の問をもっとも明快単純に答えているということである. それは通信工学者としての彼の現象を直視した思考方法にその秘密が隠されていそうである. 少なくとも Shannon の背景に当時の電波工学の発展があったことは否定できない. 1895 年 Marconi が電磁波を通信に初めて利用して以来, 無線通信の歴史の中で Nyquist, Küpfmüller, Hartley, Armstrong, 今堀等の通信方式の理論家を生み出してきた. そして第二次世界大戦中に開発されたレーダーに伴ってパルス技術に対する人々の関心が高まり, ベル研究所においては一連のパルス変調法が完成された. このようなパルス通信の出現は, 情報をデジタル的な考え方にたってみさせるようになる. Shannon は最初に文字の系列

という, 数値とは本質的に異なるものを研究対象にしたために, 情報量の性格を明確な形で捉えることができたといえよう. 今日 Shannon の理論から発生したもの, および関連する密接な理論としてはエントロピー理論, 符号化理論, 情報路の理論などがあげられる. エントロピー理論については後で詳述するが, Kolmogorov が力学系のエントロピーを定義してその同型問題を解決して以来, Sinai, Rohlin, Pinsker などのロシア学派による精力的な一連の仕事によってエントロピーの概念は情報理論から離れて力学系の解析手段としての重要な地位を与えられたのである. このように Shannon のエントロピーは情報理論の基礎概念であると同時に, 数学にも大きな影響を与えた.

1.2 Wiener

Wiener は他の 3 人と比べた場合, 彼らよりずっと饒舌であり, Shannon に数理工学者, von Neumann と Kolmogorov に数学者という役を与えるとすれば, 彼には哲学者という役がふさわしい. Wiener が最も広い意味の情報科学者であった. その理論は数学的にも優れ, また工学の分野への応用も少なくないが, 彼の著書におけるおしゃべりは物理学から工学, 精神病理学から言語学へと進み, 情報化社会のもつ危険性の警告から労働組合幹部との会見記にまで発展するのである. Wiener のサイバネティックスはこのような彼の性向と無関係ではないし, またこれがサイバネティックスを, 専門分野の違う多くの研究者から支持されるようにしたのであろう. Wiener の理論の背景については, すべてを語ることはむずかしいので, 数学的なものについてのみ考えよう. 19 世紀に発見された Fourier 級数論を除けば, Wiener 理論を支える理論はほとんど 20 世紀になって誕生した近代的解析理論である. まず 1902 年に発見された Lebesgue 積分論なくしてその理論を語ることはできないであろう. 1914 年 Hausdorff の「集合論の基礎」は来るべき函数解析の前触れとして意味がある. 1932 年にはバナッハ空間論, von Neumann による抽象バナッハ空間論が誕生した. 同じく 1930 年代の Wiener 自身や G.D.Birkoff, von Neumann, Hoph 等のエルゴート理論, また Kolomogorov の「確率論の基礎」等, 今日でもその重要性を少しでも失っていない数々の理論が明確なあるいは隠れた形で Wiener 理論の背景をなしているといえる. Wiener 周辺から発展した情報科学的分野は特に予測理論, 自動制御理論, 自動機械理論, 学習機械理論, パターン認識論, 自己増殖機会論等である. 前三者は未完成の部分を残しながらも既に現実の問題解決に活躍している. その中の予測理論と自動制御理論は情報科学の一つの根幹をなしながら, 同時にそれぞれが独立した分野になりつつあり, 今も理論の発展が急である. また自動機械論は数学基礎論や工学の順序回路論等と結びついて自動電子計算機となって実現した. コンピュータの発展は従来数学の本流からともすれば異端視されがちであった数値解析学, 近似理論などを重要な情報科学の構成員として迎え入れられたのである. またコ

ンピュータは, 最も応用から遠いと思われていた数学基礎論を情報科学の一分野である language の理論の発展に貢献させるなど, 従来の観念からいうと思いがけなかった役を数学に演じさせようとしている. 学習機械論, パターン認識論, 自己増殖機械論の後三者は今日まさに発展の途上にあり, 将来理論が完成に近づくにつれて, 哲学, 生物学, 医学, 工学などあらゆる分野に影響を与えていくであろう. これらに関連して分子生物学や神経医学における遺伝情報や神経系の研究について述べなければならない. 4 人の巨人が情報科学の分野で活躍しだした 1940 年以降に上記のこれらの分野においても, 電子顕微鏡, 組織培養法, 同位元素による物質代謝の追跡法, X線回析法等の研究手段が発達し, 生命現象の解明に著しい進歩がみられた. 遺伝情報に関しては 1953 年 Watson, Crick による DNA の分子模型の解明がセンセーションを巻き起こした. その後メッセンジャー RNA の塩基列とタンパク質のアミノ酸配列との対応が問題とされ, 1968 年頃には塩基列の Coden と 20 種類のアミノ酸とを結ぶ複合化表が完成した. この分野は今日最も注目を浴びている分野の一つになっている. 遺伝情報やウィルスの研究は根本的には自己増殖系の研究と類似点を多く持つはずであるから, 将来両理論の接触がなされ, この分野において一大発展のあることが期待される. また Wiener の制御理論と関連して神経系における「ニューロン」や「シナップス」の研究は, 今日電子工学者の多くもこれらの語を口にするようになったように, 数理工学的な情報理論の分野でも一層重要性を増しつつある. またそれらの研究は学習機械論やパターン認識論との関連においても重要である.

1.3 von Neumann

von Neumann の情報科学への全般的な貢献は Wiener ほど幅の広いものではないが, 少なくとも数学的立場からみる限りはむしろ Wiener より多種類の解析的手法をこの分野にもたらした. 彼の理論の背景をなすものとしては Wiener の場合と同じく 20 世紀に生まれた近代数学を第一に挙げなくてはならない. ついで物理学, 数理経済学, 生物学等も無視できない. 物理学を背景としたものでは量子統計力学や測定の理論がある. 前者は混合状態のエントロピーが導入されたが, これは Shannon のエントロピーの非可換の場合への拡張になっていた. 後者については函数解析的な手段によって彼の測定の理論を表示すると, まさに Shannon が与えた情報路と同一の数学構造をもつことが知られる. 特筆すべきことはこれらが Shannon の発見よりも十数年先行していたことである. しかしその後の von Neumann の学問的活動の領域は情報理論（エントロピー理論）そのものでなしに, より広く情報科学全般に向って行ったのである. 数理経済学を背景としたものでは言うまでもなく Morgenstern との共著の「ゲームの理論」がある. ここにおける数学的手段は第二次世界大戦中イギリス軍部の作戦研究から生まれ, 早くも北大西洋対独戦で成果をあげたといわれるオペレーションズ・リサーチ（ＯＲ）の手段と共通点が多い. 情報科学につい

て語る場合, 数理経済学や経営学の分野にコンピュータの導入とともに情報革命をもたらしたＯＲについて触れなくてはならない. ある目的を前に, 与えられた情報を分析し最適のシステムを作るためのＯＲは, ＬＰ, ＮＰ, ＤＰ, ＱＰ, 待ち行列理論などの諸解析手段を含み, 情報科学の中にあって一大領域をなしている. ＯＲは最近ジャーナリズムの話題にのぼるようになったシステム工学とも関連して最も現代的な理論であるといえよう.

von Neumann の理論に戻り, 生物学に連なる分野では Wiener の場合にも触れたが, 彼の場合にも高い評価を与えられ, 皮肉まじりに「ノイマン語録」のあだ名がつけられているほどであり, 今後の発展が最も期待される分野である.

1.4 Kolmogorov

以上は主として Shannon, Wiener, von Neumann の 3 天才について述べたのであるが, もう一人の偉大な数学者 Kolmogorov についても述べる必要がある. 彼の名はすでに所々に出ているように, 彼は情報理論の数学への適応面で最も本質的な成果をあげた. またそれより以前に, Wiener とは独立にそれと同一の内容の理論を展開したことには驚嘆に値する. 例えば, 予測理論の一環として線型予測の一般理論, 非線型予測機の数学的分析である. 具体的には可能な信号の雑音を伴う集合体が最もよく推定する問題を解決した.

これら 4 人のそれぞれを中心として実質的にスタートを切って開発されてきた情報科学分野は, 上述の諸分野に加えて離散構造論, 組み合わせ理論, グラフ理論その他多種多様にわたり, 数学の発展と平行してお互いに刺激交流を続けながらますます進展がなされ, 数学に隣接するしかも最も新しい限りない広さと深さをもった分野に成長しつつある. さらに重要なことはこれが時代と社会の要請に応じる有効な学問領域であることであろう.

2 p 進法, 2 進法, 2 進数

我々が数の計算を行う場合 $0, 1, 2, 3, 4, 5, 6, 7, 8, 9$ なる 10 個の数字を組み合わせて数を表示し, 四則演算法則を設定し, それに基づいてすべての事物を数的に処理してきたといえる. これを一口に 10 進法という. たまたま人間の手の指が 10 本であるために, 多くの場合このような数的表示と演算方法がとられてきたと思われる. しかし 10 進法のみが万能ではなかった.

江戸時代の貨幣表示 (4 進法); 近代社会における時刻表示 (60 進法); 近代社会における 10 進法 (これが現代計数文明の基礎である); 現代情報社会の主役である 2 進法; など種々と興味深い数的表示がなされてきた.

このように数の処理はいろいろとなされているが, 数学的には p 進法という形で一般形が与えられている. すなわち, 自然数 $p \geq 2$ に対して, 任意の実数 $x(>0)$ は

$$x = \alpha_n p^n + \alpha_{n-1} p^{n-1} + \cdots + \alpha_1 p + \alpha_0 + \alpha_{-1} p^{-1} + \cdots + \alpha_{-n} p^{-n} + \cdots \qquad (1)$$

$$(\alpha_j = 0, 1, \cdots, p-1)$$

によって展開表示される. この p に論じようとする自然数を当てはめることにより, 所要の進法を得る. 古来, 日常生活で $p = 2, 4, 10, 12, 60$ などが用いられてきたことが知られている.

2.1 10 進法から 2 進法

ここで対象とする数は自然数 $m = 1, 2, 3, \cdots$(i.e.$m \in \mathbb{N}$) とする. 自然数 m を 2 進法 (binary notation) で展開する:

$$m = \sum_{k=0}^{n} \alpha_k 2^k = \alpha_n 2^n + \alpha_{n-1} 2^{n-1} + \cdots + \alpha_1 2 + \alpha_0 \qquad (2)$$

$$(\alpha_k = 0 \text{ or } 1; n \in \mathbb{N}).$$

このとき, 記号上, m の 2 進法表示を

$$m = \alpha_n \alpha_{n-1} \cdots \alpha_2 \alpha_1 \alpha_0$$

とする. いずれかの α_k が 0 でも省略しないでその位置に 0 と書く. 例えば, $\alpha_n = \alpha_{n-1} = 0$, $\alpha_1 = \alpha_0 = 0$ でも $m = 00\alpha_{n-3}\alpha_{n-4} \cdots \alpha_3 \alpha_2 00$ と書く. 具体的には, 例えば自然数 9 と 14 を展開式 (2) の形で表すと,

$$\begin{aligned} 9 &= 2^3 + 1 \\ &= 1 \cdot 2^3 + 0 \cdot 2^2 + 0 \cdot 2^1 + 1 \cdot 2^0 \\ &= 1001, \\ 14 &= 2^3 + 2^2 + 2 + 0 \\ &= 1 \cdot 2^3 + 1 \cdot 2^2 + 1 \cdot 2 + 0 \cdot 2^0 \\ &= 1110, \end{aligned}$$

つまり, 2 進法による表示ができる. このようにして (2) の展開式から 10 進法表示の自然数 $m = 0, 1, 2, 3, \cdots, 50, 60, 100$ を 2 進法表示される. こうして表示されたものを 2 進数 (binary number) という.

2.2　2 進法の加減乗除

自然数の対 $m, m^{'}(\in \mathbb{N})$ に対して 2 進法展開を

$$m = \sum_{k=0}^{n} \alpha_k 2^k, \;\; m^{'} = \sum_{k=0}^{n} \alpha_k^{'} 2^k \; (\alpha_k, \alpha_k^{'} = 0 \text{ or } 1) \tag{3}$$

と表示する. これに対して

和 $m + m^{'} = \sum_{k=0}^{n} (\alpha_k + \alpha_k^{'}) 2^k$

と表わされるが, これが実際に 2 進法表示されているのは全ての係数をつけ直さねばならない. それはつぎのようにすればよい. すなわち $\alpha_0 + \alpha_0^{'} = 0$ か 1 のときは, そのまま係数の位置に 0 か 1 とそれぞれ書き, $\alpha_0 + \alpha_0^{'} = 2$ のときは, 第 2 項に 1 となって繰り上がる. つまり, $\alpha_0 + \alpha_0^{'} = 0$ か 1 のときは第 2 項に影響はないが, $= 2$ のときは, 1 となって繰り上がり第 2 項の係数 $\alpha_1 * \alpha_1^{'}$ は $\alpha_1 + \alpha_1^{'} + 1$ となる. したがって 2 となる場合は第 2 項の係数 $= 0$ で 2 が 1 となって第 3 項の係数に加わり, 3 となる場合は第 2 項の係数 $= 1$ で $3 - 1 = 2$ が 1 となって第 3 項の係数に加わる. 以下帰納的にこれが繰り返されて, $m + m^{'}$ の 2 進法表示が完成する.

差 $m - m^{'} = \sum_{k=0}^{n} (\alpha_k - \alpha_k^{'}) 2^k$ (ただし $m \geq m^{'}$)

この場合 k 項目の係数 $\alpha_k - \alpha_k^{'} = -1$ の処理だけが問題であり, 和の場合が $\alpha_k + \alpha_k^{'} = 2$ が面倒であったとちょうど逆の演算である. 和の場合は繰り上げる操作を行ったが, 差の場合は繰り下げる操作を行う. $m = \sum \alpha_k 2^k$ の k 項目の係数 $\alpha_k = 1$ のときこれを $k - 1$ 項目に繰り下げると 2 の内の 1 を残して 1 だけさらに繰り下げると $k - 2$ 項目に 2 が付け加わる. 簡単な場合から実際の演算をしてみると,

$$\begin{aligned} 100 &= 010 + 010 \\ &= 010 + 001 + 001 \\ 11 &= 011 \\ &= 010 + 001 \end{aligned}$$

故に

$$100 - 11 = 001 = 1$$

この計算を 10 進法の算数演算方式で行えば 100 の第 2, 3 項, つまり 00 の箇所にそれぞれ 1, 2 を割り当てて引き算を実行したこおとになる.

乗除 演算の手段は 10 進数の場合と全く同じである. まず掛け算;

$$100 \times 11 = 1100; \quad 11100 \times 110 = 10101000$$

10 進法では $4 \times 3 = 12;\ \ 28 \times 6 = 168$. 次に割り算;

$$11110 \div 11 = 1010; \quad 1011011 \div 1101 = 111$$

10 進法では $30 \div 3 = 10;\ \ 91 \div 13 = 7$.

2.3 2 進数表示から 10 進数表示

これまでは 10 進法から 2 進法を論じてきたが, ちょうどその逆を, 例を挙げて説明しよう. 基本は 10 進法から 2 進法の変換式 (2) である. 2 進法で表示した数 $\alpha = 01101101110$ を 10 進数で表すことにする. この数の最高位から番号を付してみると,

$$\begin{aligned}
\alpha &= \sum_{k=0}^{n} \alpha_k 2^k \\
&= 01101101110 \\
&= 0 \cdot 2^1 0 + 1 \cdot 2^9 + 1 \cdot 2^8 + 0 \cdot 2^7 + 1 \cdot 2^6 + 1 \cdot 2^5 \\
&\quad +0 \cdot 2^4 + 1 \cdot 2^3 + 1 \cdot 2^2 + 1 \cdot 2^1 + 0 \cdot 2^0 \\
&= 2^9 + 2^8 + 2^6 + 2^5 + 2^3 + 2^2 + 2^1 \\
&= 512 + 256 + 64 + 32 + 8 + 4 + 2 \\
&= 878
\end{aligned}$$

いからる 2 進数に対しても, すべて同様に, 10 進数に表示可能である.

2.4 十干・十二支

まず十干・十二支の説明をしよう.

$$\text{十干} = \{\ \text{甲}, \text{乙}, \text{丙}, \text{丁}, \text{戊}, \text{己}, \text{庚}, \text{辛}, \text{壬}, \text{癸}\ \},$$

ここで読み方は次のようになる.
甲 = コウ, きのえ; 乙 = オツ, きのと; 丙 = ヘイ, ひのえ;
丁 = テイ, ひのと; 戊 = ボ, つちのえ; 己 = キ, つちのと;
庚 = コウ, かのえ; 辛 = シン, かのと; 壬 = ジン, みずのえ;
癸 = キ, みずのと
右の振り仮名は, 木火土金水に 2 つずつ割り当ててその 1 方を兄（え）, 他方を弟（と）とした読み方である.

$$十二支 = \{\ 子, 丑, 寅, 卯, 辰, 巳, 午, 未, 申, 酉, 戌, 亥\ \},$$

ここで読み方は次のようになる.
子 = ネ, し; 丑 = ウシ, ちゅう; 寅 = トラ, いん; 卯 = ウ, ぼう;
辰 = タツ, しん; 巳 = ミ, し; 午 = ウマ, ご; 未 = ヒツジ, び;
申 = サル, しん; 酉 = トリ, ゆう; 戌 = イヌ, じゅう; 亥 = イ, がい
それぞれ左, 右は振り仮名である. この十干・十二支を順に対の組にして, 整列したものを干支という. 中国では千古の昔から暦年の命数に用いられている. 現在では, 西暦 AC である. 任意の西暦年 (AC) を「干支」に換算して表すことができる（その逆も可）. それを実際に換算するためには, 何れか一つの暦年の干支を知ればよい. 例えば

$$AC1993 = 癸酉\ (みずのととり),$$

$$AC1 = AC1993\ の\ 1992\ 年前 = 辛酉\ (かのととり),\ AC4 = 甲子\ (きのえね),$$

$$AC618 = 戊寅\ (つちのえとら),\ AC645 = 乙巳\ (きのとみ),$$

$$AC1603 = 癸卯\ (みずのとう),\ AC1868 = 戊辰\ (つちのえたつ),$$

$$AC1997 = 丁丑\ (ひのとうし),\ AC1998 = 戊寅\ (つちのえとら),$$

$$AC2000 = 庚辰\ (かのえたつ),\ AC2001 = 辛巳\ (かのとみ),$$

$$AC2008 = 戊子\ (つちのえね),\ AC2009 = 己丑\ (つちのとうし),$$

$$AC2010 = 庚寅\ (かのえとら),\ AC2011 = 辛卯\ (かのとう)$$

また 1 干支 = 60 年であることから, 日本では 60 歳を還暦と呼んでおり, お祝いをするならわしがある. さらに 70 歳を古希, 77 歳を喜寿, 80 歳を傘寿, 88 歳を米寿, 90 歳を卒寿, 99 歳を白寿, 100 歳を百寿というように特別にめでたい歳としている.

3　情報量とエントロピー

3.1　情報の量的把握

我々は日々押し寄せる各種多様な情報の波にもまれている. 情報の波は何等かの報道機構によってもたらされる. 情報とは一口で言えば物事をまとめて表現することである. 情報の用語はむやみやたらに出現する. 情報人間, 情報浴, 情報中枢, 情報空間, $\cdots$ ＴＶなどを見ているとだいぶ混乱して情報用語が用いられている. 情報を

論じている評論家が「data を集積することがすなわち情報を得ることである」などと説明したりしている. data を集積しておけば可ということで, これでは情報を論ずることにはならない. data を集めただけでは不十分で, それを整理し, その data を可能な限り利用しやすい形に改変する. つまり, それの計数的変換がなされる必要がある.

このための第 1 段階は data 内の各要素を簡略化し統一的な記号に置き換え, それを事象の集合 A として表示することである. このようなプロセスを「符号化」といい, これは正に初めに行う数理的手段で情報整理の重要なポイントである. 第 2 段階は事象の集合 $A = \{a_1, a_2, \ldots, a_n\}$ を繰り返し試行（実験と観測）することで, このとき A を正確に把握できる「状態」に置くことが可能となる. この「状態」とは A の各事象の出現期待率 p_k を把握することである. この p_k は事象 a_k が生起する「確からしい」(= 確率) を表している. data の一般的性質から 1 回の試行によって A の内のいずれかが常に生起する, すなわち $\{p_k\}$ について

$$\sum_{k=1}^{n} p_k = 1, \;\; 0 \le p_k \le 1, \;\; k = 1, 2, \ldots, n$$

が満たされる (この $\{p_k\}$ を確率分布という). かくして data は扱いやすい情報源として真価が発揮される. このように情報源は data を解析して得る数理形式で完全事象系 (有限確率空間) であり, これを単に系と呼び,

$$\alpha = \begin{pmatrix} a_1 & \cdots & a_n \\ p_1 & \cdots & p_n \end{pmatrix}$$

と書くことにする. 系 α において, 各 p_k は a_k が生起する確からしさを表す数値であるが, 系 α そのものが保有する確からしさが求められないかを, 次のように 6 段階に分けて考えよう.

(1) 完璧な確からしさ,

(2) ほとんど確実,

(3) 多分確実,

(4) 中程度の確からしさ,

(5) 相当不確実,

(6) 全く不明,

これらを 3 つの事象 $\{a, b, c\}$ からなる系 $\alpha_1\alpha_2, \alpha_3, \alpha_4$ を対応させて考えよう.

$\alpha_1 = \begin{pmatrix} a & b & c \\ 0 & 1 & 0 \end{pmatrix}$, 事象 b の生起が確定的;

$\alpha_2 = \begin{pmatrix} a & b & c \\ 0.01 & 0.99 & 0 \end{pmatrix}$, 多分事象 b が生起;

$\alpha_3 = \begin{pmatrix} a & b & c \\ 1/5 & 2/5 & 2/5 \end{pmatrix}$, 事象 b または c のいずれかであろう;

$\alpha_4 = \begin{pmatrix} a & b & c \\ 1/3 & 1/3 & 1/3 \end{pmatrix}$, 事象 a, b, c のいずれかが起こるか全く不明.

各試行によって答えを得るが, 確からしさが不明確である系ほど, 多くの情報を試行者にもたらしたことになる. この「もたらしたもの」は試行の前にあった不確定の解消と同値であると解釈できる. 一般に系 α の試行の結果, 次の関係が得られる.

$$\text{情報の獲得} \Longleftrightarrow \text{不確定さの解消}$$

この「情報の獲得」とはどの位多くのものの獲得か, さらに α に含まれる「情報の量的把握」というように解釈が進展していく. この獲得された量を $S(\alpha)$ と書く. 以上をまとめると

$$\begin{aligned} S(\alpha) &= \alpha\text{に対する情報量} \\ &\equiv \alpha\text{のもつ固有の不確定さを表す量} \\ &\equiv \alpha\text{の不確定度} \\ &\geq 0 \end{aligned}$$

3.2 情報量・エントロピー

情報源を表す系 α に対する情報量 $S(\alpha)$ は Shannon が発見し, 見事な数理的展開がなされた. そこで量 $S(\alpha)$ を α の Shannon エントロピー, 確率論的エントロピーなどと呼ばれるが, 以下, 単に (α の) エントロピーということにする. Shannon 以後も多数の応用数学者 (正確には数理科学者) がより精密にし, 分かりやすい定式化が完成された. 以下それを述べる.

系 α のエントロピーは α に含まれる各事象 a_k そのものではなく, その各々に生起する確率 p_k に従属して定まる. このことは前節の (1) 〜(6) の説明によっても理解されよう. すなわち, 2 つの系 α, α' が

$$\alpha = \begin{pmatrix} a_1 & \cdots & a_n \\ p_1 & \cdots & p_n \end{pmatrix}, \quad \alpha' = \begin{pmatrix} a_1' & \cdots & a_n' \\ p_1' & \cdots & p_n' \end{pmatrix}$$

であれば $S(\alpha) = S(\alpha')$ である. してみると α のエントロピー $S(\alpha)$ は $p = (p_1, \cdots, p_n)$ のみに依存, つまり「p を変数とする函数 $S(\alpha)$」であると考えてよく, また前節に記したように値は非負な実数である. また定義域は次のように表される. n-次元 Euclid 空間 $\mathbb{R}^n$ の部分集合を考える:

$$\Delta_n = \{p = (p_1, \cdots, p_n) \in \mathbb{R}^n;\ \sum_{k=1}^{n} p_k = 1, p_k \geq 0 (k = 1, \cdots, n)\}$$

$$\Delta = \bigcup_{n=1}^{\infty} \Delta_n$$

とおく. Δ_n を n 次元単体という. かくして, エントロピー $S(p)$ は Δ_n 上で定義され, 半直線 $\mathbb{R}_+ = [0, \infty)$ に値をとる函数であることが知られた. この函数はどのような数学的性質を有するかをいくつかの例などを用いて調べよう.

(1) 系 α に対して $S(\alpha) = 0$ であるとは α から得る情報が何もない. つまり不確定さなしということで, これは系 α_1 のような場合.

(2) ベクトル p が Δ_n 上で微小変化すれば $S(\alpha)$ も微小変化する. これは系 α_1 が系 α_2 に変化するような場合.

(3) 系 α_4 のように p が等確率分布の場合, $S(\alpha)$ は Δ_n 上で最大値 > 0. 精巧なサイコロ

$$\begin{pmatrix} a_1 & \cdots & a_6 \\ 1/6 & \cdots & 1/6 \end{pmatrix}$$

の $S(\alpha)$ は単体 Δ_6 上で最大値 $S(1/6, \cdots, 1/6)$ をとる.

(4) $p \in \Delta_n$ に対して $q = (p, 0) \in \Delta_{n+1}$ としても $S(p) = S(q)$, つまり起り得ない不確実さは不変.

(5) 系 α の事象の順列を変えてもエントロピーの値は不変.

(6) 系 α の各事象 α_k が部分事象に分割されたとき, 新たに生じた系のエントロピーは各部分事象のエントロピーの平均値だけ増大する.

3.3 Shannon エントロピーの公理系と定理

エントロピーを表す函数 $S(p)$ がもっている性質を 6 つの命題として列挙したが, これらの命題は純然たる数学的公理系に置き直され, これによる系 α に対する情報量が次に述べるエントロピー定式 [S] によって把握させる論理的根拠が求められる. 目標の定式化と公理系は, 集合 $\Delta = \bigcup \Delta_n$ 上で定義され半直線 $[0, \infty)$ に値をとる函数 $S(\cdot)$ に関して次のように与えられる.

[**S**](Shannon のエントロピー定式)

$$S(p) = S(p_1, \cdots, p_n) = -\sum_{k=1}^{n} p_k \log p_k$$

[**S-K**](Shannon-Khinchin の公理系)

[**SK1**] $S(\cdot)$ は各 Δ_n 上の連続函数であり

$$\max\{S(p); p \in \Delta_n\} = S(1/n, \cdots, 1/n) > 0,$$

[**SK2**] $S(p_1, \cdots, p_n, 0) = S(p_1, \cdots, p_n)$,

[**SK3**] $p_k = \sum_{j=2}^{m_k} q_{k_j} > 0$ $(m_k \geq 2, k = 1, \cdots, n)$ とするとき,

$$\begin{aligned} & S(q_{11}, \cdots, q_{1m_1}, \cdots, q_{n1}, \cdots, q_{nm_n}) \\ = \;& S(p_1, \cdots, p_n) + \sum_{k=1}^{n} p_k S(q_{k1}/p_k, \cdots, q_{km_k}/p_k). \end{aligned}$$

[**F**](Faddeev の公理系)

[**F1**] 函数 $f(p) = S(p, 1-p)$ は $0 \leq p \leq 1$ 上で連続かつ少なくとも 1 点 p_0 で $f(p_0) > 0$,

[**F2**] $(p_1, \cdots, p_n) \in \Delta_n$ の任意の置換 $(p_1', \cdots, p_n')$ に対して

$$, S(p_1, \cdots, p_n) = S(p_1', \cdots, p_n'),$$

[**F3**] $p_n = q + r, q > 0, r > 0$ のとき

$$S(p_1, \cdots, p_{n-1}, q, r) = S(p_1, \cdots, p_n) + p_n S(q/p_n, \cdots, r/p_n).$$

[SK] は Shannon が発見し導入, Khinchin がそれを分かりやすくし, 証明も簡単にしたもので, その後 Faddeev によりさらに簡潔な [F] が示された. この 2 つの公理系は, 命題 (1) 〜(6) で概観したものを定式化したものである. 各対応を調べると, [SK1] $\Longleftrightarrow$ (2) + (3), [SK2] $\Longleftrightarrow$ (4), [F1] $\Longleftrightarrow$ (2) + (3), [F2] $\Longleftrightarrow$ (5), [F3] $\Longleftrightarrow$「(6) で 1 つの事象 a_n のみが二分割された場合」. これらはエントロピーの特徴を表す文章命題を数理定式に変換したものと考えてよい. ここで次の基本定理に到着する.

定理 1 *(エントロピーの特徴化)* 条件 $S(1/2, 1/2) = 1$ の下で *[S],[SK],[F]* は互いに同値である. 即ち,

$$S(p) = -\sum_{k=1}^{n} p_k \log p_k \Longleftrightarrow [SK] \Longleftrightarrow [F]$$

3.4 定理の証明

[S] $\Longrightarrow$ [SK] $\Longrightarrow$ [F] $\Longrightarrow$ [S] の順で行う.

第 1 段： [S] $\Longrightarrow$ [SK]: [SK2] は自明, [SK3] は有限項級数の初等的計算により容易である. [SK1] を示そう. 連続性は明らか. $h(t) = -t\log t (t \geq 0)$ とおくと, $h(t)$ は凹函数であるから,

$$\begin{aligned} S(p) &= S(p_1, \cdots, p_n) = \sum_{k=1}^{n} h(p_k) \\ &= n\frac{\sum_{k=1}^{n} h(p_k)}{n} \leq n \cdot h(\frac{\sum_{k=1}^{n} p_k}{n}) \\ &= n \cdot h(\frac{1}{n}) = n \cdot (-1)\frac{1}{n}\log\frac{1}{n} \\ &= \log n = S(1/n, \cdots, 1/n). \end{aligned}$$

第 2 段： [SK] $\Longrightarrow$ [F] を示そう. [SK1] $\Longrightarrow$ [F1] は自明. [SK1,2,3] $\Longrightarrow$ [F2]: p_k がすべて正の有理数のとき, 各 $p_k = l_k/m (l_k, m$ は整数で $1 \leq l_k \leq m)$ と表せる. これに [SK3] を用いて

$$\begin{aligned} & S(p_1, \cdots, p_n) \\ = \ & S(l_1/m, \cdots, l_n/m) \\ = \ & S(1/m, \cdots, 1/m, \cdots, 1/m, \cdots, 1/m) + \sum_{k=1}^{n} p_k S(1/l_k, \cdots, 1/l_k). \end{aligned}$$

この右辺の第 1 項の各要素は同一の $1/m$ であるから変数の分け方は順列 $(l_1, \cdots, l_n)$ に無関係, また第 2 項の和 $\sum$ も有限和で, そのとり方の順序に無関係であるから $(p_1, \cdots, p_n)$ の置換 $(p_1', \cdots, p_n')$ に対して $p_k' = l_k'/m$ とおくと

$$\begin{aligned} S(p) &= S(1/m, \cdots, 1/m, \cdots, 1/m, \cdots, 1/m, \cdots, 1/m) + \sum_{k=1}^{n} p_k' S(1/l_k', \cdots, 1/l_k') \\ &= S(l_1'/m, \cdots, l_k'/m) \\ &= S(p_1', \cdots, p_n'). \end{aligned}$$

p_k が有理数でないときは, $p = (p_1, \cdots, p_n) \in \Delta_n$ に収束する有理点列 $\{p_j\}_{j=1}^{\infty} \subset \Delta_n$ をとり, $S(\cdot)$ の連続性 ([SK1]) により [F2] を得る. [SK1,2,3] $\Longrightarrow$ [F3]: [SK2,3] と [F2] より

$$\begin{aligned} S(1/2, 1/2) &= S(1/2, 1/2, 0, 0) = S(1/2, 0, 1/2, 0) \\ &= S(1/2, 1/2) + (1/2)S(1, 0) + (1/2)S(1, 0). \end{aligned}$$

故に $S(1,0)$. これを用い結論 [F3] を得る.

$$\begin{aligned} & S(p_1,\cdots,p_{n-1},q,r) \\ = \; & S(p_1,0,p_2,0,\cdots,p_{n-1},0,q,r) \\ = \; & S(p_1,\cdots,p_n)+\sum_{k=1}^{n-1}p_kS(1,0)+p_nS(q/p_n,r/p_n) \\ = \; & S(p_1,\cdots,p_n)+p_nS(q/p_n,r/p_n). \end{aligned}$$

第３段： [F] $\Longrightarrow$ [S]: [F2,3] より, 任意の $p,q\geq 0, r>0, p+q+r=1$ に対して

$$\begin{aligned} & S(p,q,r) \\ = \; & S(p,q+r)+(q+r)S(q/(q+r),r/(q+r)) \\ = \; & S(q,p+r)+(p+r)S(p/(p+r),r/(p+r)). \end{aligned}$$

ここで $f(p)=S(p,1-p)$ とおくと, $q+r=1-p$ および $p+r=1-q$ を用い

$$f(p)+(1-p)f(q/(1-p))=f(q)+(1-q)f(p/(1-q)). \tag{4}$$

この等式は任意の $p,q\in[0,1)$ に対して定義されており, 特に $p=0,q>0$ のときは

$$f(0)+f(q)=f(q)+(1-q)f(0),$$

即ち,

$$f(0)=S(0,1)=S(1,0)=0$$

を得る. ここで (4) の両辺を 0 から $1-p(0\leq p\leq 1)$ まで q に関して積分すると

$$(1-p)f(p)+(1-p)^2\int_0^1 f(t)dt=\int_0^{1-p}f(t)dt+p^3\int_p^1 t^{-3}f(t)dt \tag{5}$$

を得る. [F1] から函数 $f(p)$ は閉区間 $[0,1]$ 上で連続であるから (5) の左辺第 1 項以外はすべて $0<p<1$ で微分可能. したがって $f(p)$ も同じくそうである. (5) の両辺を p で微分する:

$$(1-p)f'(p)-f(p)-2(1-p)\int_0^1 f(t)dt=-f(1-p)+2p\int_p^1 t^{-3}f(t)dt-\frac{f(p)}{p}.$$

ここで $f(p)=f(1-p)$ であるから消去し合い

$$(1-p)f'(p)=2(1-p)\int_0^1 f(t)dt+2p\int_p^1 t^{-3}f(t)dt-\frac{f(p)}{p} \tag{6}$$

を得る. この式においても前述と同じ理由により $f'(p)$ は $0<p<1$ で微分可能であり, (6) を用いると

$$f'(p)=\frac{-2}{p(1-p)}\int_0^1 f(t)dt\ (0<p<1)$$

なる変数分離型方程式を得る. この両辺を不定積分し, 再び $f(p)=f(1-p)$, $f(0)=0$ および $f(1/2)=S(1/2,1/2)=1$ を用いて積分定数を計算すると等式

$$f(p)=-p\log p-(1-p)\log(1-p)$$

が $0\leq p\leq 1$ で成立することがすぐいえる ($\log=\log_2$ に注意). 故に任意の $(p_1,p_2)\in\Delta_2$ に対して

$$S(p_1,p_2)=-(p_1\log p_1+p_2\log p_2)$$

が得られ, 一般の $(p_1,\cdots,p_n)\in\Delta_n$ $(n=2,3,\cdots)$ に対する結論の等式

$$S(p_1,\cdots,p_n)=-\sum_{k=1}^{n}p_k\log p_k$$

は [F3] と数学的帰納法を用いれば容易である. □

4 通信路

4.1 相対エントロピーと相互情報量

n 個の事象からなる事象系 X が 2 つの状態 $p,q\in\Delta_n$ をとる. つまり同一の X に 2 つの完全事象系 $\begin{pmatrix}X\\p\end{pmatrix}$ と $\begin{pmatrix}X\\q\end{pmatrix}$ が対応しているとする. このとき p の q に対する相対的な不確定性を計る量として,

$$S(p|q)=\sum_{k=1}^{n}p_k(\log p_k-\log q_k)\tag{7}$$

が定義される. これを q に関する p の相対エントロピーという. このときエントロピーに関する基本不等式として, 次の定理を得る.

定理 2 任意の $p,q\in\Delta_n(p_kq_k\neq 0,k=1,2,\cdots,n)$ に対し, つねに

$$S(p|q)\geq 0$$

が成立し, 等号 $\Longleftrightarrow p=q$.

証明. 不等式

$$\log t \geq 1 - \frac{1}{t}\ (t > 0) \tag{8}$$

が成り立つから

$$S(p|q) = \sum_k p_k \log \frac{p_k}{q_k} \geq \sum_k p_k (1 - \frac{q_k}{p_k}) = 0 \tag{9}$$

また, 不等式 (8) が $t > 0$ で, 等号成立 $\Longleftrightarrow t = 1$, であるから,

$$\begin{aligned} & S(p|q) = 0 \\ \Longleftrightarrow\ & \sum_k p_k (\log \frac{p_k}{q_k} - 1 + \frac{q_k}{p_k}) = 0 \\ \Longleftrightarrow\ & p_k = q_k (\forall k) \Longleftrightarrow p = q. \end{aligned}$$

□

この定理は次のことを意味している. 同一の事象系を観測し, 試行するとき, 主たる状態でそれを行う場合と, 念のため他の状態 q で行う場合, $S(p|q)$ の値は p の q に対する相対的な不確定さを表している. p_k と q_k の値が近いほどこの相対的な不確定さは小であるということである.

複合事象系とは二つの事象系

$$X = (x_1, x_2, \cdots, x_m),\ \ Y = (y_1, y_2, \cdots, y_n)$$

が重複し合った $m \cdot n$ 個の事象からなる系をいう. 記号上これを次のように 3 通りの書き方で表し, 議論の際の記号上の便宜に応じて, それぞれが用いられる.

$$\begin{aligned} X \times Y &= ((x_1, y_1), (x_1, y_2), \cdots, (x_1, y_n), \\ &\quad (x_2, y_1), (x_2, y_2), \cdots, (x_2, y_n), \\ &\quad (x_3, y_1), (x_3, y_2), \cdots, (x_3, y_n), \\ &\quad \cdots \\ &\quad (x_m, y_1), (x_m, y_2), \cdots, (x_m, y_n)) \\ &= ((x_1, y_1), (x_1, y_2), \cdots, (x_i, y_j), \cdots, (x_m, y_n)) \\ &= ((x_i, y_j); 1 \leq i \leq m, 1 \leq j \leq n). \end{aligned}$$

集合の用語を用いれば, 二つの集合 X, Y の直積集合であるのことである. したがって, X, Y が有限集合でなくても, （すなわち, X, Y が事象の無限集合であっても）同様な記号を用いて論ずることができるが, ここでは”有限” に限定して論ずる. また簡単に複合事象系を二つの事象系 X, Y の直積集合 $X \times Y$ であるといってもよい.

事象系ということは, 前節でも述べたが, 各事象 (x_i, y_j) が何らかの試行とか, 観測などの対象であり, それが生起する確率が伴う. その確率を $p(x_i, y_j)$ で表そう. 単一な事象系 X の場合と同様で条件

$$\sum_{i,j} p(x_i, y_j) = \sum_{i=1}^{m} \sum_{j=1}^{n} p(x_i, y_j) = 1,$$

$$0 \leq p(x_i, y_j) \leq 1, (1 \leq i \leq m, 1 \leq j \leq n)$$

が満たされる. これによって一つの完全事象系が次のように定義される.

$$\begin{pmatrix} X \times Y \\ p \end{pmatrix} = \begin{pmatrix} (x_1, y_1) & \cdots & (x_i, y_j) & \cdots & (x_m, y_n) \\ p(x_1, y_i) & \cdots & p(x_i, y_j) & \cdots & p(x_m, y_n) \end{pmatrix}.$$

これを複合完全事象系（あるいは完全を略して複合事象系）という. このエントロピーを $S(X, Y)$ で表す. 即ち

$$S(X, Y) = -\sum_{i,j} p(x_i, y_j) \log p(x_i, y_j).$$

これを二つの事象 X, Y の同時エントロピーとよぶ.
複合事象系が与えられると, 複合前の各事象 x_i, y_j などの生起確率は次のように定まる：

$$p(x_i) = \sum_{j=1}^{n} p(x_i, y_j) \ (1 \leq i \leq m),$$

$$p(y_j) = \sum_{i=1}^{m} p(x_i, y_j) \ (1 \leq j \leq n).$$

また確率分布を

$$p_X = (p(x_i), \cdots, p(x_m)), \ \ p_Y = (p(y_1), \cdots, p(y_n))$$

とおくと,

$$X = \begin{pmatrix} X \\ p_X \end{pmatrix} = \begin{pmatrix} x_1 & \cdots & x_m \\ p(x_1) & \cdots & p(x_m) \end{pmatrix} = \begin{pmatrix} x \\ \\ p(x) \end{pmatrix} ; x \in X \Bigg),$$

$$Y = \begin{pmatrix} Y \\ p_Y \end{pmatrix} = \begin{pmatrix} y_1 & \cdots & y_n \\ p(y_1) & \cdots & p(y_n) \end{pmatrix} = \begin{pmatrix} y \\ \\ p(y) \end{pmatrix} ; y \in Y \Bigg)$$

は共に完全事象系である.
複合事象系において, $y \in Y$ が生起する (つまり $p(y) > 0$) という条件の下で $x \in X$ が生起する確率は

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

であり, これを y に関する x の条件付確率という. これらの記号を用いると

$$\begin{pmatrix} X \\ p(\cdot|y) \end{pmatrix} = \begin{pmatrix} x & \\ & ; x \in X \\ p(x|y) & \end{pmatrix}$$

は完全事象系であり, これのエントロピー

$$S(X|y) = -\sum_{x \in X} p(x|y) \log p(x|y) \ (\geq 0)$$

が定まる. さらに $S(X|y)$ を (Y, p_Y) について平均したものを

$$S(X|Y) = -\sum_{y \in Y} p(y) S(X|y) \ (\geq 0)$$

で表す. ここで $S(X|y)$ を $y (\in Y)$ に関する X の条件付エントロピー, また $S(X|Y)$ を Y に関する X の条件付エントロピーという. これは次のようにも書ける:

$$\begin{aligned} S(X|Y) &= -\sum_{x \in X, y \in Y} p(y) p(x|y) \log p(x|y) \\ &= -\sum_{x \in X, y \in Y} p(x,y) \log p(x|y), \end{aligned}$$

ここで X, Y は何れも有限事象系であるから重複和 $\sum$ の $x \in X, y \in Y$ の順序は任意である. この重複和は簡単に $\sum_{x,y}$ または自明なときは単に $\sum$ と書く. ここで導入した条件付エントロピーは, Y の観測後においてもなお X に残っている不確定さを表す量であり, これは各事象 $y \in Y$ を観測または試行した後で, X に残る不確定さが $S(X|y)$ によって表され, それを Y の分布 p_Y によって平均したものである.
Y を観測することによって得られる X の情報量は

$$I(X,Y) = S(X) - S(X|Y)$$

で与えられ, これを X と Y の相互情報量または相互エントロピーという.
複合事象系 $X \times Y = \begin{pmatrix} X \times Y \\ p \end{pmatrix}$ において $p(x|y) = p(x)$ のとき x と y は独立であるといい, $x \perp y$ で表す. 明らかに,

$$x \perp y \Longleftrightarrow p(x,y) = p(x)p(y).$$

$p(x)=0$ または $p(y)=0$ ならば $x \perp y$ とする. すべての $x \in X$ とすべての $y \in Y$ が独立のとき, X と Y は互いに独立であるといい, $X \perp Y$ で表すことにする. 定義から直ちに

$$x \perp y\ (\forall x \in X) \Longrightarrow S(X|y) = S(X),$$

$$X \perp Y \Longrightarrow S(X|Y) = S(X).$$

以上の事柄に関して次の定理が得られる.

定理 3 複合事象系 $\begin{pmatrix} X \times Y \\ p \end{pmatrix}$ において, 各エントロピーは次の関係式を満足する:

(1) $S(X|Y) = S(X) + S(Y|X) = S(Y) + S(X|Y)$,

(2) $\max\{S(X), S(Y)\} \leq S(X,Y) \leq S(X) + S(Y)$,

(3) $S(X|Y) \leq S(X),\ S(Y|X) \leq S(Y)$,

(4) $I(X,Y) \geq 0$,

(5) 上記の *(2), (3), (4)* の各不等号 *((2)* の前半の不等号は例外*)* の何れかが等号 $\Longleftrightarrow$ 他の不等号も等号 $\Longleftrightarrow X \perp Y$.

証明： (1) を示そう.

$$\begin{aligned} 0 &\leq S(Y|X) = \sum_x p(x) S(Y|x) \\ &= -\sum_{x,y} p(x)p(y|x) log p(y|x) \\ &= -\sum_{x,y} p(x,y) \log p(x,y) + \sum_{x,y} p(x,y) \log p(x) \\ &= S(X,Y) - S(X). \end{aligned}$$

同様に, $S(X|Y) = S(X,Y) - S(Y)$. 故に (1) が成立.
(2) の前半は (1) から自明. 後半は, 定理 2 を用いて

$$\begin{aligned} S(X,Y) &= -\sum_{x,y} p(x,y) \log p(x,y) \leq -\sum_{xy} p(x,y) \log p(x)p(y) \\ &= -\sum_x p(x) \log p(x) - \sum_y p(y) \log p(y) \\ &= S(X) + S(Y). \end{aligned} \tag{10}$$

(3) は (1) と (2) から自明.
(4) を示そう.

$$\begin{aligned} I(X,Y) &= -\sum_x p(x)\log p(x) + \sum_{x,y} p(x,y)\log p(x|y) \\ &= -\sum_x p(x)\log p(x) + \sum_{x,y} p(x,y)\log p(x,y) - \sum_{x,y} p(x,y)\log p(y) \\ &= -\sum_{x,y} p(x,y)\{\log p(x)p(y) - \log p(x,y)\} \\ &\geq 0. \end{aligned} \tag{11}$$

(5) を示す.

$$S(X,Y) \equiv S(X) + S(Y) \Longleftrightarrow \text{不等式 (10) が等号} \Longleftrightarrow p(x,y) = p(x)p(y).$$

(3) における等号の場合も, これから自明. (4) の $I(X,Y) = 0$ も不等式 (11) に定理 2 の等号の場合を適用すればよい. □

定理 3 の (3) の不等式は, X の不確定性の大きさが Y を試行（観測）し, その結果を見れば自ずから情報理論的意味が知られよう.

4.2 容量

入力アルファベット X と出力アルファベット Y, さらに $x \in X$ を送信したときに $y \in Y$ が出力として観測される確率を表す条件付確率 $p(y|x)$ の三者によって構成される系を通信路または情報路 (channel) といわれる. 出力の分布がその時点の入力の分布のみに従属して, 以前の入力および出力の分布には無関係であるような通信路を無記憶 (memoryless) という. X と Y に関する相互情報量 $I(X,Y)$ を X のすべての分布 $p_X(x)$ に対して最大をとったものを通信路容量 (channel capacity) という. すなわち,

$$C_{inf} = \max_{p_X(x)} I(X,Y)$$

ここで通信路容量 C_{inf} の性質を列挙する.

(1) $C_{inf} \geq 0$,

(2) $C_{inf} \leq \log|X|$, ただし $|X|$ は入力アルファベットをなす X の個数,

(3) $C_{inf} \leq \log|Y|$, ただし $|Y|$ は出力アルファベってをなす Y の個数,

(4) $I(X,Y)$ は $p_X(x)$ の連続函数である.

(5) $I(X,Y)$ は $p_X(x)$ の凹函数である.

それぞれの証明は容易であるので省略する.
次に通信路符号化定理を述べるにあたり, いくつかの定義をする必要がある. $\{1,2,\cdots,M\}$ を index 集合とする. メッセージ W は index 集合からある記号を取り出したもので, それをある符号化函数 f を施して長さ n の入力信号 $f(W)$ に変換する. そしてこの入力信号を通信路を通して送信した結果, 出力として Y^n を得たとする. この出力信号をある復号化函数 g を施すことにより $\hat{W}=g(Y^n)$ を得る. 一般には $\hat{W}$ が W とは異なるときに誤りが生ずることになる. ここで対象とする通信路は離散無記憶通信路 (discrete memoryless channel 略して DMC) で考えることにする.
符号 (M,n) は次の条件を満たすものから構成される.

(1) index 集合 $\{1,2,\cdots,M\}$,

(2) 符号化函数 $f:\{1,2,\cdots,M\}\to X^n$, ここで $f(1),f(2),\cdots,f(M)$ を符号語 (code words) と呼ばれる. またこれらを集めたものは codebook という.

(3) 復号化函数 $g:Y^n\to\{1,2,\cdots,M\}$.

次に誤り確率を定義する.

$$\lambda_i=Pr\{g(Y^n)\neq i|f(i)\}=\sum_{y^n}p(y^n|f(i))I(g(y^n)\neq i),$$

これは i を送信したときに受け取り手に i が届かなかったときの確率である.

$$\lambda^{(n)}=\max_{i\in\{1,2,\cdots,M\}}\lambda_i,$$

これは λ_i の中で最大なもので最大誤り確率という.
(M,n) 符号のレート R は $\dfrac{\log M}{n}$ (bit/1 送信) で定義され, これは 1 送信当たりの情報量に相当する. レート R が到達可能 (achievable rate) であるとは $\lambda^{(n)}\to 0$ $(n\to\infty)$ となる符号の列 $\{([2^{nR}],n))\}$ が存在するときである. この到達可能レートの最大を符号化容量といい, C_{code} と書くことにする. このとき Shannon によって示された通信路符号化定理は次のように表現される.

定理 4 次が成り立つ.

(1) $R<C_{inf}$ ならば R は到達可能である,

(2) R が到達可能ならば $R\leq C_{inf}$ である.

証明は多くの準備が必要であるので, ここでは省略する. この定理の (1) より $C_{inf} \leq C_{code}$ が成り立つことがわかり, 逆に (2) より $C_{code} \leq C_{inf}$ が成り立つことがわかるので合わせれば $C_{inf} = C_{code}$ である. つまり通信路容量と符号化容量が全く同じ量であることを主張しているのである. したがって C_{inf} と C_{code} を区別せずに常に C と書き, それを単に容量ということにする. 情報理論の立場から論ずれば C_{code} が本来の容量と考えられるが, この量を実際に計算しようとすれば困難でどのように導き出してよいかどうかわからない. したがって C_{code} は C_{inf} で求めてもよいことを保証するのが通信路符号化定理であるということができる.

5 結びと総括

この講義では情報とは何かを高等学校 1 年生レベルの知識で分かるように構成されており，情報量の表す意味を簡潔明瞭に説明し理解できることを目指したものである．重さを表す単位はグラム，長さを表す単位はメートル，これと同じように情報の量を表す単位はビットであることが自然に導入されている．この情報量を用いて最近急速に発展を遂げている情報関連機器や通信分野への基礎的な役割を担っていることを主張することができた．

参考文献

[1] R.B.Ash, Information Theory, Dover Publications, Inc., 1965.

[2] T.M.Cover and J.A.Thomas, Elements of Information Theory, John Wiley and Sons, Inc., 1991.

[3] R.G.Gallager, Information theory and reliable communication, John Wiley and Sons, New York, 1968.

[4] 韓太舜, 小林欣吾, 情報と符号化の数理, 岩波書店, 1994.

[5] 韓太舜, 情報理論における情報スペクトル的方法, 倍風館, 1997. Information spectrum and

[6] S.Ihara, Information Theory for Continuous Systems, World Scientific, 1993.

[7] 国沢清典, 梅垣壽春, 情報理論の進歩-エントロピー理論の発展- , 岩波書店, 1965.

[8] 梅垣壽春, 大矢雅則, 確率論的エントロピー-情報理論の函数解析的基礎 1 - , 共立出版, 1983.

[9] 梅垣壽春, 大矢雅則, 量子論的エントロピー-情報理論の函数解析的基礎２- , 共立出版, 1984.

[10] 梅垣壽春, 情報数理の基礎-函数解析的展開- , サイエンス社, 1993.