記憶のない通信路の容量の計算法*

有 本 卓**

Abstract

The capacity of discrete, memoryless channels is evaluated only in certain special circumstances by a straightforward calculation or Muroga's method which directly maximizes the transinformation employing inversion of channel matrices. Unfortunately, Muroga's method cannot be adopted to the case when dealing nonsquare channel matrices and even square channel matrices for which the method leads to negative probabilities of input signal distribution.

This paper presents a systematic method of computing capacity of arbitrary, discrete, memoryless channels. The algorithm of this method which approximates the capacity iteratively is very simple and involves only calculations of logarithm and exponential besides elementary arithmetical operations. It has also the property of monotone convergence for the capacity, and it is shown that, in general, the order of convergence of the error is at least inversely proportionate to the number of iterations and, in a certain case, it is exponential. Finally, a few inequalities which give an upper or lower bound to the capacity are derived.

1. まえがき

記憶のない離散的な（あるいは半連続な）通信路の容量は，ある特殊な場合に限って，求められている，すなわち，通信路行列が正方形で，かつ入力に関しても出力に関しても一様であるような通信路の容量は，よく知られた閉じた形の公式[1]によって簡単に求められる．また，Muroga[2]は，相互情報量を直接最大化することによって，通信路行列の逆行列を利用し，正方形の通信路行列の容量を求める公式を示した．しかし，この方法では，行列が正方形であって，かつ行列式がゼロでないことが前提となっており，しかも，求めた入力信号の確率が負になる場合は適用できない．

こうして，現在に至るも，非正方形の通信路行列に関しては，容量を求めることができる一般公式は，何も発見されておらず，わずかに容量のいくつかの性質，とくに，容量の上限と下限を与える式を導く研究が少しずつ進められ[3)~5)]，一方で「concave programming」の方法による逐次近似の計算法[6]が考えられている程度である．

* A Computation Algorithm of the Capacity of Arbitrary Memoryless Channels, by Suguru Arimoto (Faculty of Engineering Science, Osaka University)

** 大阪大学・基礎工学部

この論文は，任意に与えられた長方形の行列によって特徴づけられる通信路に関して，その容量を組織的に逐次近似する計算法を与える．この方法の基本的な構造は，情報理論でいう「あいまい度」よりは，もう少し拡張された意味の「あいまい度」[7),8)]を用いることにより成り立っている．

主要な結果は，つぎのようにまとめられる．

（1） 逐次計算の過程は非常に簡単であって，しかも，誤差は計算回数が増大するに従って単調に減少し，一般には，N回の逐次計算回数のとき誤差は $1/N$ のオーダ以下であることが保証される．

（2） ある場合には，誤差は計算回数 N の指数関数のオーダで減少することが保証される．

なお，ここで与えた方法は，容量を逐次近似計算すると同時に，最大の相互情報量（それが容量なのだが）を実現する入力確率分布の一つを逐次近似計算することにもなっている．

最後に，容量の持つある性質と，上界・下界についての不等式関係を与えるが，これらの結果の一部は，すでに Shannon や Helgert によって得られているものもある．しかし，ここでは，これらの性質が一般化された「あいまい度」の性質から直接導かれることを

指摘するとともに，ある場合には，もっと精密な近似公式をも直接的に導くことができることを述べる．

2. 相互情報量とあいまい度

記憶のない n 入力で m 出力の離散的通信路は，つぎのような通信路行列（$m\times n$ 行列である）

$$P=\{p(i/j)\}$$

によって特徴づけられる．ただし，$p(i/j)$ はつぎの条件を満足している．

$$p(i/j)\geq 0 \quad (i=1,\cdots,m,\ \ j=1,\cdots,n)$$

$$\sum_{i=1}^{m} p(i/j)=1 \quad (j=1,\cdots,n)$$

断わるまでもないのだが念のためすべての $i=1,\cdots,m$ について，$\sum_{j=1}^{n} p(i/j)>0$ であると仮定しておく．入力信号の確率分布はベクトル形式で

$$p=(p_1, p_2, \cdots, p_n)$$

$$p_j\geq 0, \quad \sum_{j=1}^{n} p_j=1$$

と表わされるが，このようなベクトル全体の集合を D^n と表わそう．同じような意味で D^m を考えることもある．たとえば，任意に固定された j に対して，ベクトル

$$p(*/j)=(p(1/j), p(2/j), \cdots, p(m/j))$$

は D^m に属する．

さて，与えられた入力信号の確率分布 p と通信路 P に対して，相互情報量は

$$I(P:p)=H(p)-H(P:p) \tag{2.1}$$

と定義されている．ここに $H(p)$ はエントロピー関数であって，つぎのように表わされる．

$$H(p)=-\sum_{j=1}^{n} p_j \log p_j \tag{2.2}$$

ここで，対数は便宜上自然対数としておく（議論の本質には無関係である）．また，量 $H(P:p)$ はあいまい度と呼ばれ，つぎのように表わされる．

$$H(P:p)=-\sum_{i=1}^{m}\sum_{j=1}^{n} p(i/j)p_j \log \frac{p(i/j)p_j}{\sum_{k=1}^{n} p(i/k)p_k} \tag{2.3}$$

通信路 P の容量は，よく知られているように，つぎのように定義される．

$$C(P)=\max_{p\epsilon D^n} I(P:p) \tag{2.4}$$

つぎに，次節で計算手法を提案するときの便宜を考えて，あいまい度 $H(P:p)$ をもっと一般化しておく[7]．いま，$f(x)$ をつぎの二つの条件を満足し，かつ区間 $0<x\leq 1$ で連続で，しかも，連続な導関数を持つ関数とする．

（1） $f(x)>0$ となるような点 x が，少なくとも一つ区間 $0<x\leq 1$ に存在する．

（2） $f(x)$ は単調非増加で，かつ $f(1)=0$．

そして，つぎのような定義を導入する．

定義 任意の組 (i, j) $(i=1,\cdots,m, j=1,\cdots,n)$ に対して，定義された非負の数値をとる関数 $\varphi(j/i)$ について，もし，すべての $i=1,\cdots,m$ に対してベクトル

$$\varphi(*/i)=(\varphi(1/i), \varphi(2/i), \cdots, \varphi(n/i))$$

が D^n に属するならば，m 個の $\varphi(*/i)$ 全体を決定法と呼ぶ．

すなわち，一つの決定法は，一つの $n\times m$ 行列とみなすことができ，その行列の第 i 列 $\varphi(*/i)$ が D^n にはいっているものと考えられる．そして，このような決定法 φ の全体の集合を Φ で表わしておく．

筆者は，前の論文で[7),8)]，上記のような関数 $f(x)$ を用いて，上で定義された決定法に関する量として

$$F(P:p,\varphi)=\sum_{i=1}^{m}\sum_{j=1}^{n} p(i/j)p_j f(\varphi(j/i)) \tag{2.5}$$

を定義し，これを「一般化されたあいまい度」と名づけた．とくに，$f(x)=-\log x$ であるとき

$$J(P:p,\varphi)=-\sum_{i=1}^{m}\sum_{j=1}^{n} p(i/j)p_j \log \varphi(j/i) \tag{2.6}$$

と書くと，$J(P:p,\varphi)$ を最小にする φ はベイズ決定法のとき，すなわち

$$\varphi(j/i)=\frac{p(i/j)p_j}{\sum_{k=1}^{n} p(i/k)p_k} \tag{2.7}$$

であることを示した．このことは，(2.7)を(2.6)に代入することによって

$$\min_{\varphi\epsilon\Phi} J(P:p,\varphi)=H(P:p) \tag{2.8}$$

であることを意味する．こうして，通信路 P の容量は

$$C(P)=\max_{p\epsilon D^n}\max_{\varphi\epsilon\Phi}[H(p)-J(P:p,\varphi)] \tag{2.9}$$

と書くこともできる．

最後に，次節以降においてしばしば用いられるので，よく知られたつぎの補題を述べておく（たとえば，Fano[9] あるいは Jelinek[10] の著書を参照）．

補題 1 もし，$p=p^0\epsilon D^n$ が相互情報量 $I(P:p)$ を最大にするならば，$p_j^0 \neq 0$ なる j に対して

$$C(P)=I(P:p^0)=\sum_{i=1}^{m} p(i/j)\log\frac{p(i/j)}{\sum_{k=1}^{n} p(i/k)p_k^0} \tag{2.10}$$

が成立している．また，l 個の入力確率分布 $p^0, \cdots, p^{l-1}\epsilon D^n$ がすべて $I(P:p)$ を最大にしているならば，その線形結合

$$p^l=\alpha_0p^0+\alpha_1p^1+\cdots+\alpha_{l-1}p^{l-1}$$
$$\alpha=(\alpha_0,\alpha_1,\cdots,\alpha_{l-1})\epsilon D^l$$

も $I(P:p)$ を最大にしている．すなわち

$$C(P)=I(P:p^0)=\cdots=I(P:p^{l-1})$$

ならば $C(P)=I(P:p^l)$ である．また

$$Pp^0=Pp^1=\cdots=Pp^{l-1}$$

も成立している．ここに Pp^0 はベクトル p^0 の行列 P による写像であって，入力確率分布が p^0 のときの出力確率分布である．

3. 逐次近似計算法と収束性

前節で述べた(2.9)に基づいて，$C(P)$ を逐次的に近似計算する方法を与えよう．その算法の手順は，つぎのように構成される．

1) まず，初期入力確率分布 $p^1\epsilon D^n$ を D^n の内点の中で任意にとる（実際問題としては，等確率分布 $p_j^1=1/n$ とするのが望ましいことが多い）．

2) ついで，つぎの計算を行なう．

$$C(1,1)=\max_{\varphi\epsilon\Phi}[H(p^1)-J(P:p^1,\varphi)] \tag{3.1}$$

このとき，前節においてすでに述べたように，$C(1,1)$ はベイズ決定

$$\varphi^1(j/i)=\frac{p(i/j)p_j^1}{\sum_{k=1}^{n} p(i/k)p_k^1} \tag{3.2}$$

を用いて

$$C(1,1)=H(p^1)-J(P:p^1,\varphi^1)$$

と表わされる．

3) ついで，φ^1 を固定した上で $H(p)-J(P:p,\varphi^1)$ を最大にする p を D^n で求め，それを $p^2\epsilon D^n$ とする．すなわち

$$C(2,1)=H(p^2)-J(P:p^2,\varphi^1)$$
$$=\max_{p\epsilon D^n}[H(p)-J(P:p,\varphi^1)] \tag{3.4}$$

このとき，p^2 は具体的にはつぎのように求まる（詳しくは定理1で述べる）．

$$p_j^2=\lambda_j^1\Big/\left(\sum_{k=1}^{n}\lambda_k^1\right)$$
$$\lambda_j^1=\exp\left[\sum_{i=1}^{m} p(i/j)\log\varphi^1(j/i)\right] \tag{3.5}$$

4) この p^2 を用いて再び 2) の計算を繰り返す．

5) こうして，一般に，つぎの逐次計算を行なえばよいのである．

$$\begin{cases} \varphi^N(j/i)=\dfrac{p(i/j)p_j^N}{\sum_{k=1}^{n} p(i/k)p_k^N} \\ \lambda_j^N=\exp\left[\sum_{i=1}^{m} p(i/j)\log\varphi^N(j/i)\right] \\ p_j^{N+1}=\lambda_j^N\Big/\left(\sum_{k=1}^{n}\lambda_k^N\right) \end{cases} \tag{3.6}$$

このようにして定められた $C(N,N)$，あるいは $C(N+1,N)$ が $N\to\infty$ のとき容量 $C(P)$ に収束するかどうか，また，その収束の速度の問題などについて論じていこう．前記の手順 3) において，既知の p^1 あるいは φ^1 から p^2 を求める計算式を示したが，まず，その根拠をつぎの定理で明らかにしておこう．

定理 1 任意に固定された $\varphi\epsilon\Phi$ に対して関係式

$$\max_{p\epsilon D^n}[H(p)-J(P:p,\varphi)]$$
$$=H(p^*)-J(P:p^*,\varphi)$$
$$=\log\left[\sum_{j=1}^{n}\exp\left\{\sum_{i=1}^{m} p(i/j)\log\varphi(j/i)\right\}\right]$$
$$\leq C(P) \tag{3.7}$$

が成立する．ここに p^* はつぎのようにして与えられる D^n のベクトルである．

$$\begin{cases} p_j^*=\lambda_j\Big/\left(\sum_{k=1}^{n}\lambda_k\right) \\ \lambda_j=\exp\left[\sum_{i=1}^{m} p(i/j)\log\varphi(j/i)\right] \end{cases} \tag{3.8}$$

証明 Lagrange の未定乗数法を用いるために

$$G(p)=H(p)-J(P:p,\varphi)$$
$$+\lambda\left(1-\sum_{j=1}^{n}p_j\right) \tag{3.9}$$

とおこう．そのとき，非線形計画法のよく知られた定理から，$G(p)$ を最大にする $p=p^*$ は

$$\frac{\partial G}{\partial p_j}\Big/_{p_j=p_j^*}$$
$$=-1-\log p_j^*+\sum_{i=1}^{m} p(i/j)\log\varphi(j/i)-\lambda$$
$$=0 \tag{3.10}$$

を満足しなければならないことがわかる．これよりすぐにつぎのことがわかる．

$$p_j^* = e^{-(1+\lambda)} \cdot \exp\left[\sum_{i=1}^{m} p(i/j) \log \varphi(j/i)\right] \tag{3.11}$$

これを $j=1$ から n まで加えて λ を求め，(3.10)に p_j^* を乗じて $j=1$ から n まで加え合わせたものと $J(P:p,\varphi)$ の定義とを比較して，つぎの式が得られる．

$$\begin{aligned}C(P) &\geq H(p^*) - J(P:p^*,\varphi)\\ &= \log\left[\sum_{j=1}^{n} \exp\left\{\sum_{i=1}^{m} p(i/j) \log \varphi(j/i)\right\}\right]\end{aligned}$$

逆に，(3.10)を満足する p^* は $H(p)-J(P:p,\varphi)$ を最大にするものでなければならないことは，$G(p)$ が p に関して concave であることより，よく知られた concave programming の定理によって，明らかである．こうして，(3.10)と等価な(3.11)より，(3.8)が得られる（証明終わり）．

つぎに，前述の計算法 1)〜5) について，一般に $C(N,N)$ あるいは $C(N+1,N)$ が $N\to\infty$ と共に $C(P)$ に単調に収束することを示そう．

定理 2 p^1 が D^n の内点であるとする．すなわち，すべての $j=1,\cdots,n$ について $p_j^1>0$ とする．そのとき $C(N,N)$ は単調に $C(P)$ に収束する．

証明 まず，(2.9) と 1)〜4) の計算手順から直接

$$C(1,1) \leq C(2,1) \leq C(2,2) \leq \cdots \leq C(N,N) \leq C(N+1,N) \leq \cdots \leq C(P) \tag{3.12}$$

となっていることがわかる．さて，相互情報量 $I(P:p)$ を最大にする入力確率分布の一つを $p^0 \in D^n$ としよう．すなわち，$C(P)=I(P:p^0)$ となっているものと仮定する．そのとき，つぎの式

$$\sum_{k=1}^{n} p_k^0 \log \frac{p_k^{N+1}}{p_k^N} = \sum_{i=1}^{m} q_i^0 \log \frac{q_i^0}{q_i^N} + C(p) - C(N+1,N) \tag{3.13}$$

が成立することを示そう．ここに，$q^0, q^N \in D^m$ はつぎのように定義されたベクトルである．

$$q^0 = Pp^0, \qquad q^N = Pp^N \tag{3.14}$$

そのために，まず

$$\alpha_j^N = \exp\left[\sum_{i=1}^{m} p(i/j) \log \frac{p(i/j)}{q_i^N}\right] \tag{3.15}$$

とおこう．(3.6) と (3.15) を比較することによって，すぐに

$$\lambda_j^N = p_j^N \alpha_j^N \tag{3.16}$$

となっていることがわかる．さらに，定理 1 から

$$\log\left(\sum_{j=1}^{n} \lambda_j^N\right) = C(N+1,N) \tag{3.17}$$

となっていることに気がつこう．こうして得られた (3.15)，(3.16)，(3.17)と補題 1 を利用して，(3.13) がつぎのようにして導かれる．

$$\begin{aligned}\sum_{k=1}^{n} p_k^0 \log \frac{p_k^{N+1}}{p_k^N} &= \sum_{k=1}^{n} p_k^0 \log\left(\frac{1}{p_k^N} \cdot \frac{\lambda_k^N}{\sum_{j=1}^{n} \lambda_j^N}\right)\\ &= -C(N+1,N) + \sum_{k=1}^{n} p_k^0 \log \alpha_k^N\\ &= -C(N+1,N) + \sum_{k=1}^{n} p_k^0 \left(\sum_{i=1}^{m} p(i/k) \log \frac{p(i/k)}{q_i^N}\right)\\ &= -C(N+1,N) + \sum_{k=1}^{n} p_k^0 \left[\sum_{i=1}^{m} p(i/k) \log\left(\frac{p(i/k)}{q_i^0} \cdot \frac{q_i^0}{q_i^N}\right)\right]\\ &= C(P) - C(N+1,N) + \sum_{i=1}^{m} q_i^0 \log \frac{q_i^0}{q_i^N}\end{aligned}$$

(3.13)の右辺の第 1 項は一般に非負であるから(3.13)より不等式

$$C(P) - C(N+1,N) \leq \sum_{k=1}^{n} p_k^0 \log \frac{p_k^{N+1}}{p_k^N} \tag{3.18}$$

が得られる．この式は任意の $N=1,2,\cdots$ について成立するから，一般に

$$\sum_{l=1}^{N} \left[C(P) - C(l+1,l)\right] \leq \sum_{k=1}^{n} p_k^0 \log \frac{p_k^{N+1}}{p_k^1} \leq \sum_{k=1}^{n} p_k^0 \log \frac{p_k^0}{p_k^1} \tag{3.19}$$

が成立する．(3.19)の右辺は，$p_k^1 \neq 0$ であるから，ある有限な値で限定され，また，左辺のカッコの各項はすべて非負であるから

$$\lim_{l\to\infty} C(l+1,l) = C(P)$$

でなければならない（証明終わり）．

この定理で示されたように，初期入力確率分布 p^1 は D^n の内点でありさえすれば何を選んできても，収束性は保証された．しかし，収束の速さには p^1 の選び方が大きく関与してくるであろう．そして，一般には，最大入力確率分布 p^0 にある意味で近いように p^1 を選べばよいのだろうが，実際問題として p^0 は未知なのだから，p^1 を選ぶ決め手はないといってよい．

そこで，実用的には，p^1 を等確率分布として選ぶことが，ミニマックスの意味で無難になってくる．

4. 逐次近似の収束の速度

まず最初に，逐次近似の誤差 $C(P)-C(N+1,N)$ は，一般的にいって，高々 $1/N$ のオーダであることを定理2の証明に基づいて示そう．

系 1 とくに，入力確率分布の第1次近似として，等確率分布 $p_j{}^1=1/n$ を選べば，誤差は少なくともつぎのように限定される．

$$C(P)-C(N+1,N)\leq\frac{\log n-H(p^0)}{N} \tag{4.1}$$

ここに $p^0\epsilon D^n$ は $I(P:p^0)=C(P)$ を満足する任意の確率分布である．

証明 $C(l+1,l)$ は正整数 l に関して単調非減少であるから，(3.19)を考慮して

$$\begin{aligned}&N[C(P)-C(N+1,N)]\\&\leq\sum_{l=1}^{N}\Big[C(P)-C(l+1,l)\Big]\\&\leq\sum_{k=1}^{n}p_k{}^0\log\frac{p_k{}^0}{p_k{}^1}=\log n-H(p^0)\end{aligned}$$

が得られる．これはまさに(4.1)と等価である（証明終わり）．

さて，議論をスムースにするために，相互情報量 $I(P:p)$ を最大にする入力信号の確率分布 p^0 の全体を $\mathcal{P}$ と表わしておこう．すなわち

$$\mathcal{P}=\{p^0/I(P:p^0)=C(P),\,p^0\epsilon D^n\}$$

ところで，補題1によれば，任意の $p^1,p^2\epsilon\mathcal{P}$ に対して $Pp^1=Pp^2$ となっている．P を $D^n\to D^m$ の写像と考えると，このことは $\mathcal{P}$ の P による像 $P(\mathcal{P})$ が D^m の中のただ一点であることを意味している．

系 2 逐次計算の過程において，ある有限な回数 N で $q^N=Pp^N\epsilon P(\mathcal{P})$ となれば，$p^{N+1}=p^N\epsilon\mathcal{P}$ であって，かつ $C(P)=C(N+1,N)$ である．

証明 すぐ上で述べたことに注意し，かつ補題1の(2.10)と(3.15)を比較することにより，$\alpha_j{}^N$ は j に無関係に定数 $\exp\{C(P)\}$ となる†．したがって，(3.16)と(3.6)より $p^{N+1}=p^N$ となり，これと(3.13)より $C(P)-C(N+1,N)=0$ が結論される（証明終わり）．

系 3 逐次近似計算の過程について

$$\lim_{N\to\infty}q^N=q^0 \tag{4.2}$$

が成立する．ここに $q^N=Pp^N$，$q^0=Pp^0$ であって，p^0 は $\mathcal{P}$ の任意の元である．

証明 Kullback の定理 3.1（参考文献 11）の p. 14）によれば，q^N が q^0 に収束することと量

$$\omega(N)=\sum_{i=1}^{m}q_i{}^0\log\frac{q_i{}^0}{q_i{}^N}$$

が $N\to\infty$ のときゼロに収束することとは同値である．さて，(3.13)より，$C(P)-C(N+1,N)\geq0$ であるから，一般に不等式

$$\begin{aligned}&\sum_{t=1}^{N}\left(\sum_{i=1}^{m}q_i{}^0\log\frac{q_i{}^0}{q_i{}^t}\right)\\&\leq\sum_{t=1}^{N}\left(\sum_{k=1}^{n}p_k{}^0\log\frac{p_k{}^{t+1}}{p_k{}^t}\right)\\&=\sum_{k=1}^{n}p_k{}^0\log\frac{p_k{}^{N+1}}{p_k{}^1}\leq\sum_{k=1}^{n}p_k{}^0\log\frac{p_k{}^0}{p_k{}^1}\end{aligned} \tag{4.3}$$

が成立する．(4.3)の右側の最後の項は，p^1 を D^n の内点にとって固定しているのだから，有界である．しかも，(4.3)の左辺の和の各項である $\omega(t)$ は非負なのだから，$N\to\infty$ のとき $\omega(N)\to0$ でなければならない（証明終わり）．

このことから $\mathcal{P}$ がただ一点から成る場合，すなわち，$I(P:p)$ を最大にする入力確率分布 p^0 が D^n においてただ一つである場合には，$N\to\infty$ のとき必ず $p^N\to p^0$ となることがわかる．

定理 3 $\mathcal{P}$ がただ一点 p^0 から成る場合，逐次計算過程において $N\to\infty$ のとき $p^N\to p^0$ となる，

証明 (3.13)をつぎのように書きなおす．

$$\begin{aligned}&\left[\sum_{k=1}^{n}p_k{}^0\log\frac{p_k{}^0}{p_k{}^{t+1}}\right]\\&=\left[\sum_{k=1}^{n}p_k{}^0\log\frac{p_k{}^0}{p_k{}^t}\right]-[C(P)-C(t+1,t)]\\&\quad-\left[\sum_{i=1}^{m}q_i{}^0\log\frac{q_i{}^0}{q_i{}^t}\right]\end{aligned} \tag{4.4}$$

この結果，(4.4)の左辺あるいは右辺の第一項は，t に関して単調減少であることがわかる．そこで，いま(4.4)の左辺が $t\to\infty$ のとき，ゼロに収束しないと仮定しよう．すなわち，すべての正整数 t に対して，つぎの不等式を満足する正数 $\delta>0$ が存在すると仮定する．

$$\sum_{k=1}^{n}p_k{}^0\log\frac{p_k{}^0}{p_k{}^t}\leq\delta>0 \tag{4.5}$$

さて，数列 $\{p^t\}$ を考えると，p^t の属する点集合 D^n は n 次元ユークリッド空間 R^n の中の有界閉集合であるから，$\{p^t\}$ は少なくとも一つの集積点 $\bar{p}$ を持つ．ところが系3より $Pp^t\to q^0$ であるから $P\bar{p}=q^0$ とな

† p^1 を D^n の内点にとっておいたから，逐次計算の途中で p^N が D^n の境界点に到達することはあり得ないことに注意する必要がある．

っている，このことは，補題1から，$\bar{p}$ も相互情報量 $I(P:p)$ を最大にしていることを意味する．こうして $\bar{p}\epsilon\mathscr{P}$ であることがわかったが，一方(4.5)より

$$\sum_{k=1}^{n} p_k{}^0 \log \frac{p_k{}^0}{\bar{p}_k} \geq \delta$$

となっている．これは $p^0 \neq \bar{p}$ を意味するが，このことは $\mathscr{P}$ が異なった二点 $p^0, \bar{p}$ を含むことになって，最初の仮定に矛盾する．こうして $t\to\infty$ のとき $p^t \to p^0$ であることが示された（証明終わり）．

つぎに，近似の誤差 $C(P)-C(N+1,N)$ は，定理2で示されたオーダー N^{-1} よりもある場合にはもっと小さいオーダーであることを示そう．

定理 4 $I(P:p)$ を最大にする入力確率分布 p^0 がただ一つであって，かつ p^0 が D^n の内点である場合，ある正整数 N と定数 $0<K\leq 1$ が存在して，$t\geq N$ なるすべての t に対して不等式

$$\left[\sum_{k=1}^{n} p_k{}^0 \log \frac{p_k{}^0}{p_k{}^t}\right] \leq (1-K)^{t-N}\left[\sum_{k=1}^{n} p_k{}^0 \log \frac{p_k{}^0}{p_k{}^N}\right] \tag{4.6}$$

が成立する．すなわち，p^t は $t\to\infty$ のとき，指数関数のオーダーで p^0 に収束する．

証明 定理3から，任意の $\delta>0$ に対して

$$\sum_{k=1}^{n} p_k{}^0 \log \frac{p_k{}^0}{p_k{}^N} < \delta$$

となるような正整数 $N=N(\delta)$ が存在する．いま，δ を十分小さくとると，p^N は p^0 に十分近くなる．そこで，p^0 と p^N の差のベクトル $\varepsilon=p^0-p^N$ について，つぎの式が成立することに注目する．

$$\begin{aligned}
&\sum_{k=1}^{n} p_k{}^0 \log \frac{p_k{}^0}{p_k{}^N} \\
&\quad = \sum_{k=1}^{n} -p_k{}^0 \log\left[1-\frac{\varepsilon_k}{p_k{}^0}\right] \\
&\quad = \frac{1}{2}\sum_{k=1}^{n} \frac{\varepsilon_k{}^2}{p_k{}^0} + \sum_{k=1}^{n} O(\varepsilon_k{}^3)
\end{aligned}$$

$$\begin{aligned}
&\sum_{j=1}^{m} q_j{}^0 \log \frac{q_j{}^0}{q_j{}^N} \\
&\quad = -\sum_{j=1}^{n} q_j{}^0 \log\left[1+\sum_{i=1}^{n} \frac{P(j/i)}{q_j{}^0}\varepsilon_i\right] \\
&\quad = \frac{1}{2}(\varepsilon, P'Q_0{}^{-1}P\varepsilon) + \sum_{k=1}^{n} O(\varepsilon_k{}^3)
\end{aligned}$$

ここに，P' は P の転置行列，ベクトル x, y についての記法 (x, y) は内積を意味するものとする．また，つぎの記号も用いる．

$$P_0=\begin{pmatrix} p_1{}^0 & & 0 \\ & \ddots & \\ 0 & & p_n{}^0 \end{pmatrix}, \qquad Q_0=\begin{pmatrix} q_1{}^0 & & 0 \\ & \ddots & \\ 0 & & q_m{}^0 \end{pmatrix}$$

そこで，δ を十分小さくとると，この二つの式から，つぎの二つの不等式が同時に成立するような N を選ぶことができることがわかる．

$$\sum_{k=1}^{n} p_k{}^0 \log \frac{p_k{}^0}{p_k{}^N} \leq \frac{2}{3}(\varepsilon, P_0{}^{-1}\varepsilon)$$

$$\sum_{j=1}^{m} q_j{}^0 \log \frac{q_j{}^0}{q_j{}^N} \geq \frac{1}{3}(\varepsilon, P'Q_0{}^{-1}P\varepsilon)$$

さて，いま $\mathscr{P}$ がただ一点 p^0 から成っていることより，補題1から，$\sum_{i=1}^{n}\varepsilon_i=0$ である任意のベクトル ε に対して，もし，$\varepsilon\neq 0$ ならば $P\varepsilon\neq 0$ となっていなければならない．このことより $P'Q_0{}^{-1}P$ が正定値行列であることを示そう．そのために，いま $P\xi=0$ となるような，ゼロでないベクトル $\xi\epsilon R^n$ が存在したとしよう．そのとき ξ は

$$\xi=\alpha e+\beta\varepsilon$$

と表わされる．ここに e は各座標がすべて1であるような n 次元ベクトル，ε は e に直交するベクトルで，α, β はスカラ量である．このときつぎの計算が成立する．

$$\begin{aligned}
0 &= \sum_{j=1}^{m}\sum_{i=1}^{n} p(j/i)\xi_i = \sum_{j=1}^{m}\sum_{i=1}^{n} p(j/i)(\alpha e_i+\beta\varepsilon_i) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m} p(j/i)[\alpha e_i+\beta\varepsilon_i] \\
&= \sum_{i=1}^{u}(\alpha e_i+\beta\varepsilon_i)=n\alpha
\end{aligned}$$

こうして $\alpha=0$ であることが導かれたが，このことは $0=P\xi=\beta P\varepsilon$ となり，$\beta=0$ か $\varepsilon=0$ を意味する．すなわち，$\xi=0$ となってもともとの仮定に矛盾する．このようにして $P'Q_0{}^{-1}P$ は正定値であることがわかったが，このことから任意の $\varepsilon\epsilon R^n$ にに対して不等式

$$\frac{1}{3}(\varepsilon, P'Q_0{}^{-1}P\varepsilon) \geq K\cdot\frac{2}{3}(\varepsilon, P_0{}^{-1}\varepsilon)$$

を満足する正数 $K>0$ が存在することがわかる．これより，N より大きい任意の t に対して

$$\sum_{j=1}^{m} q_j{}^0 \log \frac{q_j{}^0}{q_j{}^t} \geq K\sum_{i=1}^{n} p_i{}^0 \log \frac{p_i{}^0}{p_i{}^t}$$

が成立する．これを(4.4)に代代入すると不等式

$$\left[\sum_{i=1}^{n} p_i{}^0 \log \frac{p_i{}^0}{p_i{}^{t+1}}\right] \leq (1-K)\left[\sum_{i=1}^{n} p_i{}^0 \log \frac{p_i{}^0}{p_i{}^t}\right]$$

が得られるが，これより(4.6)は明らかである（証明終わり）．

5. 容量の上界・下界

この節では，いままでの考察から簡単に導出できる

容量の性質について，また，その上界・下界を与える式について述べておこう．

まず

$$C(P,\varphi)=\max_{p\in D^n}[H(p)-J(P:p,\varphi)]$$

とおく．定理1からすぐわかるように，この量は具体的に

$$C(P,\varphi)=\log\left[\sum_{j=1}^{n}\exp\left\{\sum_{i=1}^{m}p(i/j)\log\varphi(j/i)\right\}\right] \tag{5.1}$$

と与えられる．また，(3.7)から一般に

$$C(P,\varphi)\leq C(P) \tag{5.2}$$

であるが，とくに

$$\max_{\varphi\in\Phi}C(P,\varphi)=C(P) \tag{5.3}$$

であることが示される（証明はわざわざ述べなくとも，いままでの議論からほとんど了解されよう）．さらにつぎの性質を示そう．

定理 5 P_1, P_2 をともに $m\times n$ の通信路行列とする．また，α を $0\leq\alpha\leq 1$ なる任意の数とする．そのとき，つぎの不等式が成立する．

$$C(\alpha P_1+(1-\alpha)P_2,\varphi)\leq\alpha C(P_1,\varphi)+(1-\alpha)C(P_2,\varphi) \tag{5.4}$$

証明 いま

$$x_j=\exp\left\{\sum_{i=1}^{m}p_1(i/j)\log\varphi(j/i)\right\}$$

$$y_j=\exp\left\{\sum_{i=1}^{m}p_2(i/j)\log\varphi(j/i)\right\}$$

とおくと

$$C(\alpha P_1+(1-\alpha)P_2,\varphi)=\log\left[\sum_{j=1}^{n}x_j^{\alpha}y_j^{1-\alpha}\right]$$

と表わされる．ところが，Hölder の不等式より

$$\begin{aligned}&\log\left[\sum_{j=1}^{n}x_j^{\alpha}y_j^{1-\alpha}\right]\\&\leq\log\left[\left(\sum_{j=1}^{n}x_j\right)^{\alpha}\left(\sum_{j=1}^{n}y_j\right)^{1-\alpha}\right]\\&=\alpha\log\left[\sum_{j=1}^{n}x_j\right]+(1-\alpha)\log\left[\sum_{j=1}^{n}y_j\right]\\&=\alpha C(P_1,\varphi)+(1-\alpha)C(P_2,\varphi)\end{aligned}$$

となり，(5.4)が示された（証明終わり）．

この定理からつぎの Shannon の不等式[3]が簡単に示される．

系 $C(\alpha P_1+(1-\alpha)P_2)\leq\alpha C(P_1)+(1-\alpha)C(P_2)$ (5.5)

証明

$$\begin{aligned}&C(\alpha P_1+(1-\alpha)P_2)\\&=\max_{\varphi\in\Phi}C(\alpha P_1+(1-\alpha)P_2,\varphi)\\&\leq\max_{\varphi\in\Phi}[\alpha C(P_1,\varphi)+(1-\alpha)C(P_2,\varphi)]\\&\leq\alpha\max_{\varphi\in\Phi}C(P_1,\varphi)+(1-\alpha)\max_{\varphi\in\Phi}C(P_2,\varphi)\\&=\alpha C(P_1)+(1-\alpha)C(P_2)\end{aligned}$$

（証明終わり）

つぎに $C(P)$ の下界を与える式を二三導びこう．$m\times n$ 行列 P の各 i 行のベクトル $p(i/*)$ について，それの n 個の要素の中から任意に一つ $p(i/j_i)$ を選ぶ．ここに j_i は各 i について1から n までの勝手に指定された番号である．そして

$$\Delta=\sum_{i=1}^{m}p(i/j_i) \tag{5.6}$$

とおく．また $0\leq e\leq 1$ なる数 e を任意にとる．そこで，つぎのような入力確率分布 p と決定法 φ を考えよう．

$$\begin{cases}p_j=1/n, & j=1,\cdots,n\\ \varphi(j/i)=\begin{cases}1-e & \text{if } j=j_i\\ e/(n-1) & \text{if } j\neq j_i\end{cases}\end{cases}$$

そのとき，つぎの不等式が導びかれる．

$$\begin{aligned}C(P)&\geq H(p)-J(P:p,\varphi)\\&=\log n+\sum_{i=1}^{m}\frac{1}{n}p(i/j_i)\log(1-e)\\&\quad+\sum_{i=1}^{m}\sum_{j\neq j_i}\frac{1}{n}p(i/j)\log\frac{e}{n-1}\\&=\log n+\frac{\Delta}{n}\log(1-e)+\left(1-\frac{\Delta}{n}\right)\log e\\&\quad-\left(1-\frac{\Delta}{n}\right)\log(n-1)\end{aligned} \tag{5.7}$$

ここで e は $0\leq e\leq 1$ で任意であったことを考慮して，(5.7)の右辺の最大値をとることにより，$C(P)$ の下界の一つを与える不等式

$$C(P)\geq\log n-\left(1-\frac{\Delta}{n}\right)\log(n-1)-H\left(\frac{\Delta}{n},1-\frac{\Delta}{n}\right) \tag{5.8}$$

が得られる．これは Helgert[5] の Theorem 1 を拡張したものになっている．

今度は入力確率分布 p と決定法 φ をつぎのようにとろう．

$$\begin{cases}p_j=1/n, & j=1,\cdots,n\\ \varphi(j/i)=p(i/j)\Big/\left[\sum_{k=1}^{n}p(i/k)\right]\end{cases}$$

これに定理1を応用すると，つぎの二つの不等式が得られる．

$$C(P)\geq\log\left[\sum_{j=1}^{n}\exp\left\{\sum_{i=1}^{m}p(i/j)\log\frac{p(i/j)}{\sum_{k=1}^{n}p(i/k)}\right\}\right] \tag{5.9}$$

$$C(P)\geq\log n-\frac{1}{n}\sum_{j=1}^{n}H(p(*/j)) -\frac{1}{n}\sum_{i=1}^{m}\left(\sum_{j=1}^{n}p(i/j)\right)\log\left(\sum_{j=1}^{n}p(i/j)\right) \tag{5.10}$$

つぎに $C(P)$ の上界の一つを与える式を導く．そのために，補題1を用いると，つぎの式が得られる．

$$\begin{aligned}C(P)&=\sum_{i=1}^{m}\sum_{j=1}^{n}p(i/j)p_j{}^0\log p(i/j)/q_i{}^0\\&=-\sum_{j=1}^{n}q_i{}^0\log q_i{}^0/q_i\\&\quad+\sum_{i=1}^{m}\sum_{j=1}^{n}p(i/j)p_j{}^0\log p(i/j)/q_i\\&\leq\max_{j=1\sim n}\sum_{i=1}^{m}p(i/j)\log p(i/j)/q_i\end{aligned} \tag{5.11}$$

ここに p^0 は $C(P)=I(P:p^0)$ を実現する入力確率分，p は D^n の任意のベクトル，q^0, q はそれぞれ $q^0=Pp^0$，$q=Pp$ である．p は任意であるから，とくに $p_j=1/n$ と選べば，(5.11)から

$$C(P)\leq\log n+\max_{j=1\sim n}\sum_{i=1}^{m}p(i/j)\log\frac{p(i/j)}{\sum_{k=1}^{n}p(i/k)} \tag{5.12}$$

が得られる．

最後に，いままで述べてきたいくつかの不等式による計算例を示しておく．対象とした通信路は Helgert の論文[5)]から採用したつぎの 3×3 の通信路行列である．

$$P=\begin{pmatrix}0.6&0.7&0.5\\0.3&0.1&0.05\\0.1&0.2&0.45\end{pmatrix} \tag{5.13}$$

第1表 (5.13)で与えられた通信路 P の容量についての上界・下界の計算例

	(5.8)	(5.9)	(5.10)	Th. 1
下界	0.0691	0.1257	0.1238	0.0691

	(5.12)	Th. 2	Th. 3	Th. 4
上界	0.1708	1.0161	0.5985	0.5547

第1表の計算値はビット単位で表わされている．また，表中の (5.8)～(5.12) はそれぞれ本文中の不等式 (5.8)～(5.12)に従って計算したもので，Th. 1～Th. 4 は Helgert の論文[5)]の Theorem 1～4 の不等式に従って計算した値である．なお(5.8)において $\Delta=0.7+0.3+0.45=1.45$ を用いた．

第1表からわかるように，この場合 $C(P)$ は

$$0.1257\leq C(P)\leq 0.1708$$

であるが，念のため，第2節で述べた逐次計算法によって求めた $C(P)$ の値と，そのときの最大入力確率分布 p^0 の値を**第2表**にまとめておく．

第2表 (5.13)で与えられた通信路 P の容量 $C(P)$ と，それを実現する最大入力確率分布 p^0 の値

$C(P)$	0.161628
$p_1{}^0$	0.501723
$p_2{}^0$	0.0
$p_3{}^0$	0.498277

なお，(5.13)の P については，Muroga の方法[2)]は直接用いることができないことを注意しておく．Muroga の方法で計算を進めると，最大入力確率分布 p^0 の第2座標 $p_2{}^0$ が負になる．したがって，p^0 は D^3 の境界上になければならないことがわかるが，このとき P はすでに 3×2 の通信路行列とみなされ，正方形でなくなって Muroga の方法は使えないのである．

6. あとがき

通信路の容量を計算するための必ず収束することが保証された逐次近似法を示した．ここで述べた方法は，その算法の構造から，半連続通信路にまで拡張できることは明らかであろう．

初期入力確率分布 p^1 をどのように選ぶかという問題は未解決である．定理4で述べたように，最大入力確率分布 p^0 が D^n の内点であれば，p^1 が D^n の内点でありさえすれば，指数関数のオーダーによる収束が保証される．p^0 が D^n の境界にあるとき，やはり指数関数のオーダーで収束することが保証されるように p^1 を選ぶには，p^0 のどの座標がゼロであるかをまず知って，そして，p^0 のゼロの座標に応ずる p^1 の座標もゼロにしておけば十分である．しかし，現在までの研究では，p^0 が D^n の境界にあるとき，どの座標がゼロになっているかを知る一般的な方法は知られていないようである．がともかく，ここで述べた方法によれば，p^0 が D^n のどんな境界にあっても，少なく

とも計算回数に反比例して，誤差がゼロに収束することだけは保証されるのである.

参考文献

1) C. E. Shannon: "A Mathematical Theory of Communication", Bell Spstem Tech. J., Vol. 27, pp. 379～423 and 623～656, 1948.
2) S. Muroga: "On the Capacity of a Discrete Channel I", J. Phys. Soc. Japan, Vol. 8, pp. 484～494, 1953.
3) C. E. Shannon: "Geometrishe Deutung Einiger Ergebnisse bei der Berechnung der Kanalkapazität", Nachrichtentech. Z., Vol. 1, pp. 1～4, 1957.
4) H. J. Helgert: "On a Bound for Channel Capacity", IEEE Trans., IT-13, pp. 124～126, 1967.
5) H. J. Helgert: "A Comparison of Arbitrary and Symmetric Channels on the Basis of Capacity", IEEE Trans., IT-13, pp. 467～470, 1967.
6) B. Meister and W. Oettli: On the Capacity of a Discrete, Constant Channel", Information and Control, Vol. 11, pp. 341～351, 1967.
7) 有本 卓: "逐次推定過程の情報理論的基礎づけ", 情報処理, Vol. 10, No. 1, pp. 61～67, 1969.
8) 有本 卓: "ベイズ決定法とあいまい度", 電子通信学会論文誌(C), Vol. 53-C, 1970 (印刷中).
9) R. M. Fano: "Transmission of Information", The M. I. T. Press, Cambridge, Mass., 1961.
10) F. Jelinek: "Probabilistic Information Theory", McGraw-Hill Book, New York, 1968.
11) S. Kullback: "Information Theory and Statistics", Dover Publications, New York, 1968.

(昭和 44 年 9 月 10 日受付)