

An analogue of McMillan's theorem for the counting entropies

東北大. 工. 大黒 茂 (Shigeru Ohkuro)

§0. 序.

Shannon の有名な論文 [1] が世に出してから、通信の理論は数学の一つの分野となった。今日、それは情報理論、又はエントロピー理論として知られている。Shannon 以後、この分野への重要な寄与が、McMillan [2] 及び Feinstein [3] によってなされた。現在、我々は、これらの理論を Khinchin [4] の数学的にすぐれた仕事を通して、より容易に知ることができる。（[5] 及び [6] も参照せよ。）

ここでは、二項分布のエントロピーに対して、

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{l=0}^n C_{nl} p^l q^{n-l} \ln(C_{nl} p^l q^{n-l}) = 0, \\ q = 1 - p,$$

が、各点 $p \in [0, 1]$ で成立つこと [7] の応用として、有名な McMillan の定理 [2, 4, 5] を用いて、'counting entropy'

と二項係数の間にも、McMillanの定理と同様な形の（確率収束に関する）関係式が成立することを示す。上の形のエントロピーは統計物理学においても重要である。又、その研究は、‘pulse counting’を用いた通信（例えば、Geiger counter など）及び、‘photon（又は、electron）counting’を用いた通信などに対する新しい型の情報理論において、必要不可欠なものとなる。このような型の通信は、近い将来、レーザーやガラスファイバーの技術を用いて、現実化されると思われる。‘counting entropy’の概念は、生物の情報過程、特に、人間の知能の本質、例えば、情報の脳における発生（これは、情報の伝送とは相補的な概念である）を研究する上でも有用であると思われる。（cf. [8]）

その重要性にもかかわらず、筆者の知る限り、‘counting entropy’に関する数学的な仕事は、これまで、皆無のようである。ここでは、出来るだけ、統計学より一層厳密な、確率空間を基礎とした、確率論の言葉、特に、都合の良い確率変数の導入により、定式化してみよう。

次節(3.1)では、筒集合と n 項鎖の復習から始める。次に、(3.2)情報量を確率変数として表わそう。3.3では、この情報量の確率収束を述べる。3.4では、エントロピーを確率収束の極限值として考える。統計的独立性を仮定しない

場合についても触れる。付録(§5)では、筆者がこれまでに得た[7, 9, 10, 11, 12]、1) = 項分布のエントロピーに
関係したいくつかの極限公式、2) Boltzmann のエントロピー
公式との関係、3) = 項分布のエントロピーのいくつかの性質、
及び、数値計算の例、などを示す。

以下、§1 以外では、(確率)系列の定常性を常に仮定している。
又、簡単のため、主として、(確率)系列の統計的独立性を仮定した場合に
つき述べることにする。

§1. 筒集合と n 項鎖. (cf. [5])

今、アルファベット A とし、0 と 1 の二つの要素からなる
集合を考える。

$$A = \{0, 1\} \quad (1-1)$$

0 または 1 の無限列を x とする。

$$x \equiv (x_i) \equiv (\dots, x_{-1}, x_0, x_1, x_2, \dots), \quad (1-2)$$

$$x_i \in A, \quad i \in \mathbb{Z} \equiv \{\dots, -2, -1, 0, 1, 2, \dots\}$$

x の形のすべての要素の集合を $A^{\mathbb{Z}}$ と書く。

$$A^{\mathbb{Z}} \equiv \{x; x_i \in A, i \in \mathbb{Z}\} \quad (1-3)$$

あらかじめ与えられた $x_i \in A, i = 1, 2, \dots, n$ で定まる点列
 $[x_1, x_2, \dots, x_n]$ は n 項鎖と呼ばれる。一オ、 $A^{\mathbb{Z}}$ の n 次の筒
 集合 $\mathcal{C}(x_1, x_2, \dots, x_n)$ とは、

$$C(x_1, x_2, \dots, x_n) \equiv \{y \in A^I; y_1 = x_1, y_2 = x_2, \dots, y_n = x_n\} \quad (1-4)$$

が定義される A^I の部分集合のことである。 n 項鎖と n 次の筒集合とは同一視できる。

$$[x_1, x_2, \dots, x_n] \sim C(x_1, x_2, \dots, x_n), \\ x_i \in A, \quad i = 1, 2, \dots, n \quad (1-5)$$

n 次の筒集合全体を $\mathcal{C}_n$ とする。 ($\mathcal{C}_0$ の要素は A^I のみ。)

$$\mathcal{C}_n \equiv \{[x_1, x_2, \dots, x_n]; x_i \in A, i = 1, 2, \dots, n\}, \\ n = 0, 1, 2, \dots \quad (1-6)$$

A^I のすべての筒集合からなる集合族から生成された σ -加法族を $\mathcal{F}_A$ とする。 A^I のすべての筒集合に対して確率 P が与えられるものと仮定すると、良く知られているように、 P は $\mathcal{F}_A$ の上に一意的に拡張される。(Carathéodory の拡張定理) 即ち、 $\forall S \in \mathcal{F}_A$ に対して、その確率 $P(S)$ が定まる。このようにして、確率空間 $(A^I, \mathcal{F}_A, P)$ が導入され、 n 項鎖は $\mathcal{F}_A$ の要素とみなされることになる。

$$[x_1, x_2, \dots, x_n] = C(x_1, x_2, \dots, x_n) \in \mathcal{F}_A, \quad (1-7) \\ x_i \in A, \quad i = 1, 2, \dots, n, \quad n \in \mathbb{N}$$

§2. 確率変数としての情報量.

以下では、確率系列の定常性を仮定する。 $A_l^{(n)}$ を次式で定義する。

$$A_l^{(n)} \equiv \left\{ [x_1, x_2, \dots, x_n] \in \mathcal{C}_n ; \sum_{i=1}^n x_i = l \right\}, \quad (2-1)$$

$$l = 0, 1, \dots, n, \quad n \in \mathcal{N}$$

このとき、次の直和分解が得られる。

$$\mathcal{C}_n = \sum_{l=0}^n \oplus A_l^{(n)}, \quad n \in \mathcal{N} \quad (2-2)$$

確率変数 $X_n: A^{\mathcal{I}} \rightarrow \{0, 1, 2, \dots, n\} \subset \mathbb{R}^1$, $n \in \mathcal{N}$ を次式で定義する。

$$X_n(x) = \sum_{i=1}^n x_i, \quad x \in A^{\mathcal{I}} \quad (2-3)$$

X_n の分布は、もし、 $[x_1, x_2, \dots, x_n]$ の確率 $P([x_1, x_2, \dots, x_n])$ が、 n と $l \equiv \sum_{i=1}^n x_i$ (即ち、 l は $x_1, \dots, x_n$ の中に 1 が現れる回数である) により依存して決まるなら、

$$\begin{aligned} P(X_n^{-1}(l)) &= P(X_n(x) = l) = P([x_1, x_2, \dots, x_n] \in A_l^{(n)}) \\ &= \sum_{[x_1, x_2, \dots, x_n] \in A_l^{(n)}} P([x_1, x_2, \dots, x_n]) = (\#A_l^{(n)}) \cdot P([x_1, x_2, \dots, x_n]) \\ &= {}_n C_l P([x_1, x_2, \dots, x_n]), \quad l = 0, 1, \dots, n \end{aligned} \quad (2-4)$$

で与えられる。

我々は、この論文で、主に二項分布 $B(n, p)$ について考える。 n 項鎖の確率を具体的に次のようにおく。

$$P([x_1, x_2, \dots, x_n]) = p^l q^{n-l}, \quad q \equiv 1-p, \quad 0 \leq p \leq 1, \\ l \equiv \sum_{i=1}^n x_i \quad (2-5)$$

即ち、 l は、 $x_i, i=1, 2, \dots, n$ の中の 1 の数、 $n-l$ は 0 の数である。ここで、各カラムの統計的独立性を仮定したことになる。このとき、上の X_n の分布は、二項分布、

$$P(X_n^{-1}(l)) = {}_n C_l p^l q^{n-l}, \quad l=0, 1, \dots, n \quad (2-6)$$

となる。以下では、二項分布の列、 $B(n, p), n=1, 2, \dots$ 、を考える代りに、確率空間 $(A^{\mathbb{I}}, \mathcal{F}_A, P)$ 上の確率変数列 $X_n, n=1, 2, \dots$ 、を考えよう。

X_n の値域 $\mathcal{R}_n$ 上で定義された関数列 $\mu_n: \mathcal{R}_n \equiv X_n(A^{\mathbb{I}}) (\subset B_1) \rightarrow$ 閉区間 $[0, 1] \subset \mathbb{R}^1, n \in \mathbb{N}$ 、を

$$\mu_n(\Lambda) \equiv (P \circ X_n^{-1})(\Lambda) = P(X_n \in \Lambda) = \sum_{l \in \Lambda} {}_n C_l p^l q^{n-l}, \\ \Lambda \in \mathcal{P}(\{0, 1, 2, \dots, n\}) \subset B_1, \quad (2-7)$$

(B_1 は Borel 集合体、 $\mathcal{P}$ は中集合)

で定義すると、 μ_n は確率になり、 X_n の分布と呼ばれる。即ち、 $X_n: A^{\mathbb{I}} \rightarrow \mathbb{R}^1$ により、 $(A^{\mathbb{I}}, \mathcal{F}_A, P) \rightarrow (\mathbb{R}^1, B_1, \mu_n), n \in \mathbb{N}$ なる、確率空間から確率空間への対応が生ずる。 $(\mathbb{R}^1, B_1, \mu_n)$ は、ここでは有限確率空間 $(\{0, 1, 2, \dots, n\}, \mu_n)$ に他ならない。

さらに次のような確率変数 $F_n: A^{\mathbb{I}} \rightarrow \mathbb{R}^1$, 及び,
 $f_n: \{0, 1, \dots, n\} \rightarrow \mathbb{R}^1$, ($n=1, 2, \dots$) を定義しよう。

$$\begin{aligned} F_n(x) &\equiv f_n(X_n(x)) \equiv -\frac{1}{n} \ln \mu_n(\{l\}) \text{ if } X_n(x) = l, l=0, 1, \dots, n, \\ &= -\frac{1}{n} \ln P(X_n \in \{l\}) \text{ if } " = -\frac{1}{n} \ln (C_n p^l q^{n-l}) \text{ if } " \end{aligned} \quad (2-8)$$

この式の右辺は、 $X_n(x) = l$ の条件の下での、受取ったアルファベット一個当りの情報量 ($l=0, 1, \dots, n$) を表すことが知られている。[5, 13]

次の結果を得る。[14]

$$\begin{aligned} E|f_n(X_n)| &\equiv \int_{A^{\mathbb{I}}} |f_n(X_n(x))| \cdot P(dx) \\ &= -\frac{1}{n} \sum_{l=0}^n C_n p^l q^{n-l} \ln(C_n p^l q^{n-l}) < \infty \end{aligned} \quad (2-9)$$

不等号の理由は、一般に有限分布 p_l , $l=0, 1, 2, \dots, n$ に対して、Shannon のエントロピーは、 $-\sum_{l=0}^n p_l \ln p_l \leq \ln(n+1)$ となるからである。(cf. [7]) 上式より、

$$E(F_n) = E f_n(X_n) = \int_{A^{\mathbb{I}}} f_n(X_n(x)) \cdot P(dx) = \frac{1}{n} H_n(p) \quad (2-10)$$

を得る。ここで、 $H_n(p)$ は二項分布 $B(n, p)$ のエントロピーである。かくして、 $H_n(p)/n$ なる量が n に無関係な確率 p

に関する平均値 $E(F_n)$ で表現することが出来た。

§3. 情報量の確率収束.

次の結果を得る。

[定理3-1] $F_n(x)$ は0に平均収束する。即ち、

$$\lim_{n \rightarrow \infty} E|F_n(x)| = 0 \quad (3-1)$$

が成立つ。

平均収束から確率収束が言えるので次の系を得る。

[系3-2] $F_n(x)$ は0に確率収束する。即ち、 $\forall \varepsilon > 0$, $\forall \delta > 0$ に対して、 $n_0 \in \mathcal{N}$ が存在して、 $\forall n (\geq n_0) \in \mathcal{N}$ に対して、

$$P(|F_n(x)| > \delta) < \varepsilon \quad (3-2)$$

とできる。

この結果は次のように言い換えることが出来る。(cf. [5])

[系3-3] $\varepsilon > 0$ 及び $\delta > 0$ がどんなに小さくても、 n を十分大きく取れば、すべての鎖は、次の性質を持つ二つの組に分けることができる。

1) 第一の組の任意の鎖 $[x_1, x_2, \dots, x_n]$ に対して、それが $A_\varepsilon^{(n)}$,

($l=0, 1, \dots, n$) に属する確率 $P([x_1, x_2, \dots, x_n] \in A_l^{(n)})$, 即ち, $X_n(x)$ の分布 μ_n は,

$$\left| \frac{\ln \mu_n}{n} \right| \leq \delta \quad (3-3)$$

を満足し,

2) 第二の組の鎖に対しては, $X_n(x)$ の分布 μ_n は,

$$\left| \frac{\ln \mu_n}{n} \right| > \delta \quad (3-4)$$

を満足し, その鎖の確率の総和は ε より小さくなる。

第一の組を「確率の大きい」組, 第二の組を「確率の小さい」組と呼ぶことにする。

§4. 確率収束の極限值としてのエントロピー.

さて, 我々の n 項鎖に対する確率 $P([x_1, x_2, \dots, x_n])$ は, 情報理論で言う「無記憶情報源」に相当する。[15] 従って, 次の結果が知られている。

[補助定理4-1] $-\frac{1}{n} \ln P([x_1, x_2, \dots, x_n])$ は $n \rightarrow \infty$ のとき, エントロピー $H_1(p)$ に確率収束する。即ち, $\forall \varepsilon > 0$, $\forall \delta > 0$ に対して, $n_0 \in \mathbb{N}$ が存在して, $\forall n (\geq n_0) \in \mathbb{N}$ に対して,

$$P\left(\left| -\frac{1}{n} \ln P([x_1, x_2, \dots, x_n]) - H_1(p) \right| > \delta\right) < \varepsilon \quad (4-1)$$

と出来る。

この定理は、有名な McMillan の定理 [5] の特殊な場合である。次の結果も容易に示せる。

[補助定理 4-2] 一般に、二つの確率変数列 $\{X_n(x)\}$, $\{Y_n(x)\}$ と一つの確率変数 $Y(x)$ に対し、

$$X_n - Y_n \rightarrow 0 \text{ in pr. } (n \rightarrow \infty), \quad (4-2)$$

(確率収束)

かつ、

$$Y_n \rightarrow Y \text{ in pr. } (n \rightarrow \infty) \quad (4-3)$$

ならば、

$$X_n \rightarrow Y \text{ in pr. } (n \rightarrow \infty) \quad (4-4)$$

が成立つ。

補助定理 4-2 を用いると、系 3-2 と補助定理 4-1 とから、次の主定理を得る。

[定理 4-3] $\frac{1}{n} \ln_n C_{X_n}$ は $n \rightarrow \infty$ の時、 $H_1(p)$ に確率収束する。

ここで、 n 項鎖の各カラムの統計的独立性を仮定しない場合について簡単に触れておこう。この場合、§2で述べたように、もし、確率 $P([x_1, x_2, \dots, x_n])$ が、 n と $l \equiv \sum_{i=1}^n x_i$ のみ依存して決まるなら、(2-6)式の代わりに、(2-4)式を使えば、定理3-1の形の結果がそのまま成立つ。従って、系3-2が、この場合にも成立つ。但し、ここで、確率変数 $f_n(x)$ の定義式(2-8)において、もはや、二項分布の形は仮定していない。関係式、

$$-\frac{1}{n} \ln P(X_n^{-1}(l)) = -\frac{1}{n} \ln C_n l - \frac{1}{n} \ln P([x_1, x_2, \dots, x_n]), \quad (4-5)$$

を用いると、補助定理4-1の代り、一般のMcMillanの定理[5, 15]により、定理4-3と同様の論法で次の結果を得る。

[定理4-4] 独立でない一般の $P([x_1, x_2, \dots, x_n])$ の時でも、 n と l のみで決まり、かつ、エルゴード的な場合は、 $\frac{1}{n} \ln C_n l$ は $n \rightarrow \infty$ の時、情報源のエントロピー H に確率収束する。

ここで、エルゴード的情報源のエントロピーについては[5, 15]を参照されたい。この定理4-4に関係して、 $P([x_1, x_2, \dots, x_n])$ が n と l のみで決まることと、エルゴード性との関係が問題として生ずるが、将来の課題としたい。

次の二つの事を注意しておく。1) (3-1)式で示したよう

に、countingによる情報の伝送は、一文字当りのエントロピー（これは通常 $n \rightarrow \infty$ で定義される）が0なので、情報の伝送速度から見れば、効率が悪いことになる。しかし、 n 項鎖に対する演算の前後で、エントロピーの保存則 [8] が成立つ可能性もある。この点から見ると、counting は情報の生成、消滅の研究に有用であろう。2) 現実には、情報は有限の長さの鎖で送ることになる。この時、(二項分布の) エントロピーを n で割った、'現実的な一文字当りの counting entropy' は、 n の実際的な値に対して、'情報理論で言う一文字当りのエントロピー' より十分大きい値を持つ可能性もある。さらに、系 3-2 及び、定理 4-3 等において、収束を保証する $n_0(\epsilon, \delta)$ が、實際上、あまり大きな数でなく、それ故、我々の結果が実用上でも、重要である可能性も十分あると思われる。

§5. 付録. [7, 9, 10, 11, 12]

1) 二項分布のエントロピーに関する極限公式、

[定理 5-1] 任意の有限分布 $p_i, i=0, 1, 2, \dots, n$ に対してそのエントロピーを $H(p_0, p_1, \dots, p_n)$ とすると、次式が成立つ。

$$\lim_{n \rightarrow \infty} \frac{1}{n} H(p_0, p_1, \dots, p_n) = 0$$

$$[\text{系} 5-2] \quad \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{l=0}^n C_n^l p^l q^{n-l} \ln C_n^l = H_1(p) \\ \equiv -(p \ln p + q \ln q)$$

[系 5-3]

$$\lim_{n \rightarrow \infty} n \sqrt{\frac{n}{\prod_{l=0}^n C_n^l} p^l q^{n-l}} = \frac{1}{p^p \cdot q^q}$$

[系 5-4]

$$\lim_{n \rightarrow \infty} n \cdot 2^n \sqrt{\frac{n}{\prod_{l=0}^n C_n^l}} = 2$$

[系 5-5]

$$\lim_{n \rightarrow \infty} n \sqrt{\frac{n-1}{\prod_{l=1}^n (n-l)} \frac{n-l}{2^{n-1}}} = 1$$

[系 5-6] $x \in (0, \infty)$ に対して.

$$\lim_{n \rightarrow \infty} \frac{1}{(x+1)^{n-1} \cdot n} \sum_{l=0}^n x^l \cdot C_n^l \cdot \ln C_n^l = \ln \frac{(x+1)^{x+1}}{x^x}$$

[定理 5-7] (Bernstein の応用. cf. [9, 14])

$$\lim_{n \rightarrow \infty} \frac{1}{n} \left[\exp \left\{ \ln \{ l^l \cdot (n-l)^{n-l} \} \right\} \right]^{\frac{1}{n}} = x^x \cdot (1-x)^{1-x}$$

$$\therefore \therefore \quad \langle \ln \{ l^l \cdot (n-l)^{n-l} \} \rangle \equiv \sum_{l=0}^n C_n^l x^l (1-x)^{n-l} \ln \{ l^l \cdot (n-l)^{n-l} \}$$

は、 n と $x \in [0, 1]$ の関数である。又、上の収束は一樣である。

2) Boltzmann のエントロピー公式について。

定理 4-3 から次の結果を得る。

[系 5-8] n が大きいとき、すべての鎖は、つぎの二つの組に分けることができる。i) この組の鎖 $[x_1, \dots, x_n]$ に対し、 X_n の値は、 $\left| \frac{1}{n} \ln_n C_{X_n} - H_1(p) \right| \leq \delta$ を満し、(確率の大きい組)
ii) ϕ の組の鎖の確率の総和はとより小になる。(確率の小さい組)

ϕ の組について述べる。その特徴は次式である。

$n H_1(p) \approx \ln_n C_{X_n}$. このような X_n の値を l と書く。

[系 5-9] 確率の大きい組は、次式を満すような $A_l^{(n)}$ と $A_{n-l}^{(n)}$ の要素から成っている。(n : 大)

$$H_1(p) = H_1(1-p) \approx \frac{1}{n} \ln_n C_l = \frac{1}{n} \ln_n C_{n-l}$$

$A_l^{(n)}$ の要素が、与えられた順序に l の $1/n$ を取る確率 $p_0(n, l)$ が $1/n C_l$ であることから、 $\hat{H} \equiv k_B n H_1(p) \approx k_B \ln_n C_l$
 $= -k_B \ln p_0(n, l)$ である。ここで、 $\hat{H}$, k_B は、夫々 Boltzmann のエントロピー及び定数であり、 $p_0(n, l)$ は $A_l^{(n)}$ の元について等しく、等確率である。 C_l は組合数と呼ばれている。[13]
 確率収束の意味から、“ n が小さければ、上の公式の成立つ確率は小さくなり得る。”

3) 二項分布のエントロピーの性質と数値計算の例.

情報理論においては、鎖の間に結合が導入され、 m 項鎖と n 項鎖から一つの $(m+n)$ 項鎖が作られる。[5] にこれヒントを得て、二つの「二項分割 (the binomial decomposition)」（有限分割の一種）の間に新しい積を導入することにより次の結果を得る。[7, 9, 11]

[定理 5-10] 二項分布のエントロピー $H_n(p)$ に対し、次の不等式が成立つ。

$$H_n(p) \leq H_{m+n}(p) \leq H_m(p) + H_n(p), \\ m, n \in \mathbb{N}, \quad p \in [0, 1]$$

等号は $p=0$, または、 1 の時に限る。

[系 5-11] 次の不等式が成立つ。

$$\sum_{l=2}^{n+1} n C_{l-1} (p^l q^{n+1-l} + q^l p^{n+1-l}) \ln l \\ \leq \ln(n+1) \leq -(p \ln p + q \ln q) \\ + \sum_{l=2}^{n+1} n C_{l-1} (p^l q^{n+1-l} + q^l p^{n+1-l}) \ln l$$

等号は $p=0$, または、 1 の時に限る。

$n=10, 50, 100$ に対する $H^{(n)}(p)/n$ の数値計算の結果を示しておこう。



