数理解析研究所講究録
第522巻 1984年 155-171

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、

$$\lim_{n\to\infty}\frac{1}{n}\sum_{l=0}^{n}{}_nC_l\,p^l q^{n-l}\ln\left({}_nC_l\,p^l q^{n-l}\right)=0,$$
$$q\equiv 1-p,$$

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、必要欠くべからざるものとなろう。このような型の通信は、近い将来、レーザーやグラスファイバーの技術を用いて、現実化すると思われる。‘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（§1）では、簡集合とn項鎖の復習から始める。次に、（§2）情報量を確率変数として表わそう。§3では、この情報量の確率収束を述べる。§4では、エントロピーを確率収束の極限値として考える。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)二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)二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、などを与える。

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として、0と1の二つの要素からなる集合を考える。

$$A = \{0, 1\} \tag{1-1}$$

0又は1の無限列をxとする。

$$x \equiv (x_i) \equiv (\cdots, x_{-1}, x_0, x_1, x_2, \cdots), \quad x_i \in A, \quad i \in \mathbb{Z} \equiv \{\cdots, -2, -1, 0, 1, 2, \cdots\} \tag{1-2}$$

xの形のすべての要素の集合をA^Iと書こう。

$$A^I \equiv \{x;\ x_i \in A,\ i \in \mathbb{Z}\} \tag{1-3}$$

あらかじめ与えられた$x_i \in A$, $i = 1, 2, \cdots, n$で定まる点列$[x_1, x_2, \cdots, x_n]$はn項鎖と呼ばれる。一方、A^Iのn次の筒集合$C(x_1, x_2, \cdots, x_n)$とは、

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

で定義される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。

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

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

$$\mathbb{C}_n \equiv \{ [x_1, x_2, \cdots, x_n] ;\ x_i \in A,\ i = 1, 2, \cdots, n \}, \quad n = 0, 1, 2, \cdots \tag{1-6}$$

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。

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

§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。$A_\ell^{(n)}$を次式で定義する。

$$A_\ell^{(n)} \equiv \{[x_1, x_2, \cdots, x_n] \in \mathbb{C}_n ; \sum_{i=1}^{n} x_i = \ell\}, \quad (2-1)$$
$$\ell = 0, 1, \cdots, n, \qquad n \in \mathbb{N}$$

このとき、次の直和分解が得られる。

$$\mathbb{C}_n = \sum_{\ell=0}^{n} \oplus A_\ell^{(n)}, \qquad n \in \mathbb{N} \quad (2-2)$$

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

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

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

$$P(X_n^{-1}(\ell)) = P(X_n(x) = \ell) = P([x_1, x_2, \cdots, x_n] \in A_\ell^{(n)})$$
$$= \sum_{[x_1, x_2, \cdots, x_n] \in A_\ell^{(n)}} P([x_1, x_2, \cdots, x_n]) = (\#A_\ell^{(n)}) \cdot P([x_1, x_2, \cdots, x_n])$$
$$= {}_nC_\ell \, P([x_1, x_2, \cdots, x_n]), \quad \ell = 0, 1, \cdots, n \quad (2-4)$$

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。

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

即ち、ℓは、x_i, $i=1,2,\cdots,n$の中の1の数, $n-\ell$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っている。このとき、上のX_nの分布は、二項分布、

$$P(X_n^{-1}(\ell)) = {}_nC_\ell\, p^{\ell} q^{n-\ell}, \quad \ell = 0, 1, \cdots, n \qquad (2-6)$$

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

X_nの値域$\mathcal{R}_n$上で定義された関数列 $\mu_n : \mathcal{R}_n \equiv X_n(A^{\mathrm{I}}) (\subset \mathcal{B}_1) \to$ 閉区間$[0,1] \subset \mathbb{R}^1$, $n \in \mathbb{N}$, を

$$\mu_n(\Lambda) \equiv (P \circ X_n^{-1})(\Lambda) = P(X_n \in \Lambda) = \sum_{\ell \in \Lambda} {}_nC_\ell\, p^{\ell} q^{n-\ell},$$
$$\Lambda \in \mathcal{P}(\{0,1,2,\cdots,n\}) \subset \mathcal{B}_1, \qquad (2-7)$$
($\mathcal{B}_1$はBorel集合体, $\mathcal{P}$は巾集合)

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、$X_n : A^{\mathrm{I}} \to \mathbb{R}^1$により、$(A^{\mathrm{I}}, \mathfrak{F}_A, P) \to (\mathbb{R}^1, \mathcal{B}_1, \mu_n)$, $n \in \mathbb{N}$なる、確率空間から確率空間への対応が生ずる。$(\mathbb{R}^1, \mathcal{B}_1, \mu_n)$は、ここでは有限確率空間$(\{0,1,2,\cdots,n\}, \mu_n)$に他ならない。

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

$$F_n(x) \equiv f_n(X_n(x)) \equiv -\frac{1}{n}\ln\mu_n(\{\ell\}) \text{ if } X_n(x)=\ell,\ \ell=0,1,\cdots,n,$$

$$= -\frac{1}{n}\ln P(X_n\in\{\ell\}) \text{ if } \text{〃} \quad = -\frac{1}{n}\ln\left({}_nC_\ell p^\ell q^{n-\ell}\right) \text{ if } \text{〃} \tag{2-8}$$

この式の右辺は, $X_n(x)=\ell$ 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 $(\ell=0,1,\cdots,n)$ を表すことが知られている。[5, 13]

次の結果を得る。[14]

$$E|f_n(X_n)| \equiv \int_{A^I} |f_n(X_n(x))|\cdot P(dx)$$

$$= -\frac{1}{n}\sum_{\ell=0}^{n} {}_nC_\ell p^\ell q^{n-\ell}\ln\left({}_nC_\ell p^\ell q^{n-\ell}\right) < \infty \tag{2-9}$$

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

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

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、

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

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、$n_0 \in \mathbb{N}$ が存在して、$\forall n (\geq n_0) \in \mathbb{N}$ に対して、

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

とできる。

この結果は次のように言いかえることが出来る。(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 $[x_1, x_2, \cdots, x_n]$ に対して、それが $A_q^{(n)}$,

$(\ell = 0, 1, \cdots, n)$ に入る確率 $P([x_1, x_2, \cdots, x_n] \in A_\ell^{(n)})$，即ち，$X_n(x)$ の分布 μ_n は，

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

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，

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

を満足し，その鎖の確率の総和はとより小さくなる。

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. 確率収束の極限値としてのエントロピー。

さて，我々の n 項鎖に対する確率 $P([x_1, x_2, \cdots, x_n])$ 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] $-\frac{1}{n} \ln P([x_1, x_2, \cdots, x_n])$ は $n \to \infty$ 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、$\forall n (\geqq n_0) \in \mathbb{N}$ に対して、

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

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、

$$X_n - Y_n \longrightarrow 0 \quad in\ pr. \quad (n \to \infty), \qquad (4-2)$$

(確率収束)

かつ、

$$Y_n \to Y \quad in\ pr. \quad (n \to \infty) \qquad (4-3)$$

ならば、

$$X_n \to Y \quad in\ pr. \quad (n \to \infty) \qquad (4-4)$$

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] $\frac{1}{n}\ln {}_nC_{X_n}$ は $n \to \infty$ 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 $P([x_1, x_2, \cdots, x_n])$ が、nと $l \equiv \sum_{i=1}^{n} x_i$ にのみ依存して決まるなら、(2-6)式の代りに、(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、

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

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 $P([x_1, x_2, \cdots, x_n])$ 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、$\frac{1}{n}\ln_n C_{x_n}$ は $n \to \infty$ 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、$P([x_1, x_2, \cdots, x_n])$ 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$n\to\infty$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]が成立つ可能性がある。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ろう。この時、(=項分布の)エントロピーそれで割った'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$n_0(\varepsilon,\delta)$が、実際上、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係した極限公式.

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

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

[系5-2]
$$\lim_{n\to\infty}\frac{1}{n}\sum_{\ell=0}^{n}{}_nC_\ell\, p^\ell q^{n-\ell}\ln {}_nC_\ell = H_1(p)$$
$$\equiv -(p\ln p + q\ln q)$$

[系5-3]
$$\lim_{n\to\infty}\sqrt[n]{\prod_{\ell=0}^{n}{}_nC_\ell^{\,{}_nC_\ell}\, p^\ell q^{n-\ell}} = \frac{1}{p^p\cdot q^q}$$

[系5-4]
$$\lim_{n\to\infty}\sqrt[n\cdot 2^n]{\prod_{\ell=0}^{n}{}_nC_\ell^{\,{}_nC_\ell}} = 2$$

[系5-5]
$$\lim_{n\to\infty}\sqrt[n]{\prod_{\ell=1}^{n-1}(n-\ell)^{\frac{{}_nC_\ell}{2^{n-1}}}} = 1$$

[系5-6] $x\in(0,\infty)$ に対し.
$$\lim_{n\to\infty}\frac{1}{(x+1)^{n-1}\cdot n}\sum_{\ell=0}^{n}x^\ell\cdot{}_nC_\ell\cdot\ln {}_nC_\ell = \ln\frac{(x+1)^{x+1}}{x^x}$$

[定理5-7] (Bernsteinの応用. cf. [9, 14])
$$\lim_{n\to\infty}\frac{1}{n}\left[\exp<\ln\{\ell^\ell\cdot(n-\ell)^{n-\ell}\}>\right]^{\frac{1}{n}} = x^x\cdot(1-x)^{1-x}$$

ここで, $<\ln\{\ell^\ell\cdot(n-\ell)^{n-\ell}\}> \equiv \sum_{\ell=0}^{n}{}_nC_\ell\, x^\ell(1-x)^{n-\ell}\ln\{\ell^\ell\cdot(n-\ell)^{n-\ell}\}$

は、nと$x\in[0,1]$の関数である。又、上の収束は一様である。

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$[x_1,\cdots,x_n]$に対し、X_nの値は、$\left|\frac{1}{n}\ln {}_nC_{X_n}-H_1(p)\right|\leqq\delta$を満し、(確率の大きい組)
ii)第二の組の鎖の確率の総和はεより小になる。(確率の小さい組)

第一の組について述べる。その特徴は次式である。
$nH_1(p)\approx\ln {}_nC_{X_n}$。このような$X_n$の値を$l$と書こう。

[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)

$$H_1(p)=H_1(1-p)\approx\frac{1}{n}\ln {}_nC_l=\frac{1}{n}\ln {}_nC_{n-l}$$

$A_l^{(n)}$の要素が、与えられた順序にlコの$\overset{\text{良い}}{\underset{\text{悪い}}{}}0$を取る確率$P_0(n,l)$が$1/{}_nC_l$であることから、$\hat{H}\equiv k_B nH_1(p)\approx k_B\ln {}_nC_l$
$=-k_B\ln P_0(n,l)$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。${}_nC_l$は配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] これにヒントを得て、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。

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

等号は $p=0$, 又は、1の時に限る。

[系5-11] 次の不等式が成立つ。

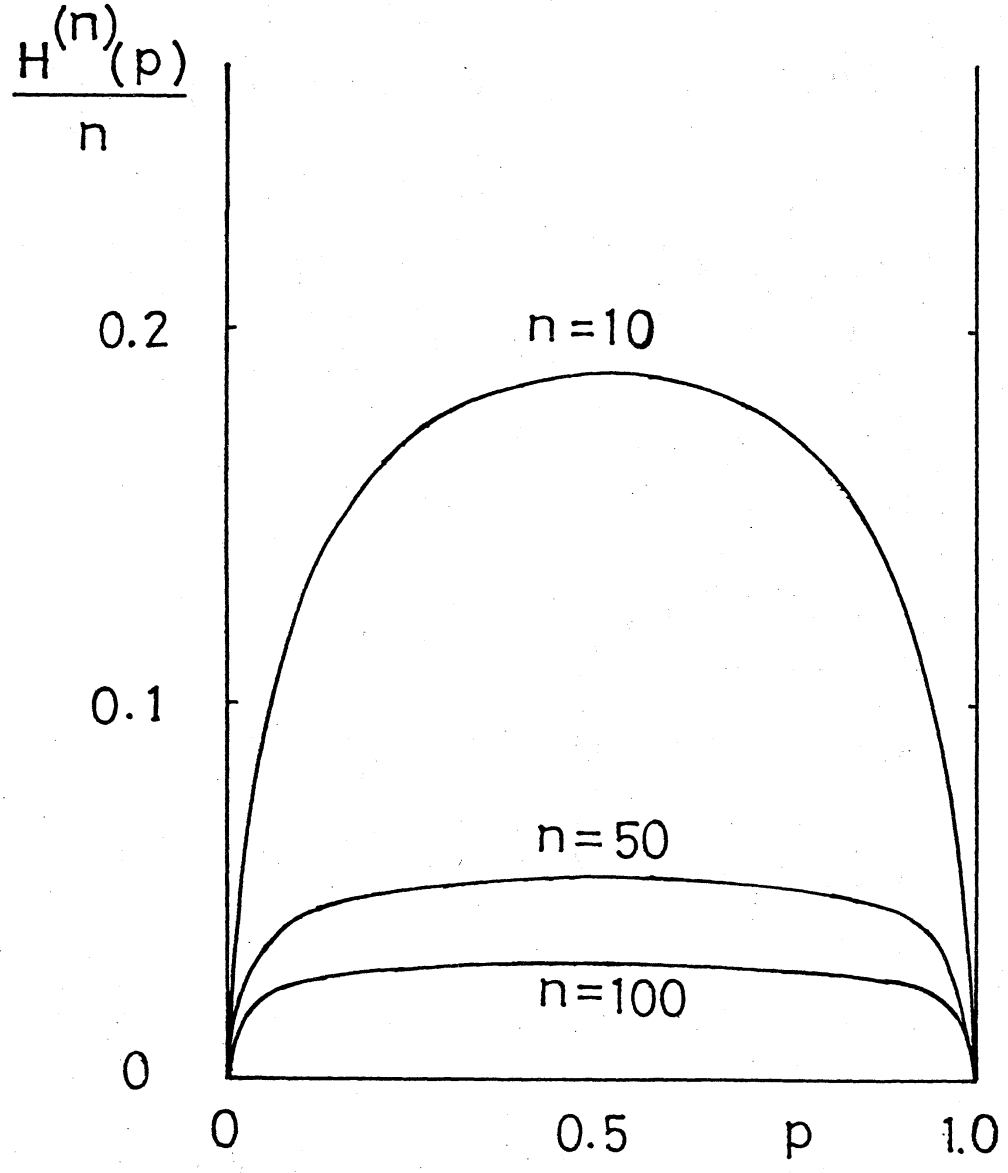
$$\sum_{l=2}^{n+1} {}_nC_{l-1}\left(p^l q^{n+1-l} + q^l p^{n+1-l}\right) \ln l$$

$$\leqq \ln(n+1) \leqq -(p\ln p + q\ln q)$$

$$+ \sum_{l=2}^{n+1} {}_nC_{l-1}\left(p^l q^{n+1-l} + q^l p^{n+1-l}\right) \ln l$$

等号は $p=0$、又は、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。


H(n)(p)
n
0.2
0.1
0
n=10
n=50
n=100
0
0.5
p
1.0

文献

[1] Shannon, C.E. (1948) Bell System. Tech. J. 27, 379−423, 623−656.

[2] McMillan, B. (1953) Ann. Math. Stat. 24, 196-219.

[3] Feinstein, A. (1954) Trans. I.R.E., PGIT-4, 2−22.

[4] Khinchin, A.I. (1958) Mathematical foundations of information theory. Dover.

[5] 国沢清典, 梅垣寿春 (1965) 情報理論の進歩. 岩波.

[6] Aczél, J. and Daróczy, Z. (1975) On measures of information and their characterization. Academic Press.

[7] Ohkuro, S. (1983) VIIth Internat. Congr. on Math. Phys., Univ. of Colorado, USA.

[8] 大沢文夫, 寺本英 (1972) 生命の物理(岩波講座). 岩波.

[9] Ohkuro, S. (1983) Math. Process. No.2. (Private commu.)

[10] 大黒茂 (1984) 日本物理学会(春期)講演(物基・統力).

[11] 大黒茂 (1984) 日本数学会(春期)講演(応数). その1.

[12] 〃 〃 〃 〃 〃 〃 その2.

[13] 佐藤洋 (1973) 情報理論. 裳華房.

[14] 西尾真喜子 (1978) 確率論. 実教出版.

[15] 有本卓 (1982) 確率・情報・エントロピー. 森北出版.

(以上)