C2 情報理論

目　次

凡　　例

1、本文を章・節に分け，$\boxed{1}$（1章）$\boxed{1.1}$（1章1節）のように番号をつける．

2、引用文献：例えば Kolmogorov[86] は末尾文献表で［86］番の Kolmogorov の文献を指す．

3、文中の（アンダーライン）のあるものは，その項又はその場所で定義或は基本的説明が述べられていることを示し，大部分は索引に出ている．

4、主な記号

$A \equiv B$：AをBで定義する

X，Y，Z：確率変数

$(\Omega, \mathbb{B}, P)$：基礎の確率空間

$(K_X, \mathbb{B}_X, P_X)$，$(A, \mathbb{B}_A, P_X), \cdots$：確率変数$X$によって導入される確率空間

$P_{X\times Y} \equiv P_X \times P_Y$

$P_{Y|X}$：Yの条件(X)つき確率

$\bar{P}_{Y\times Z|X}(D\times E\times F) \equiv \int_D P_{Y|X}(E|x)P_{Z|X}(F|x)P_X(dx)$：$\mathbb{B}_X\times\mathbb{B}_Y\times\mathbb{B}_Z$ 上の確率測度

$L^2(\Omega)=L^2(\mathbb{B})=L^2(P)\equiv\{X(\omega),\ \omega\in\Omega;\ \mathbb{B}\text{-可測},\ \int_\Omega|X(\omega)|^2P(d\omega)<\infty\}$（Hilbert空間）

$L^{*2}(\Omega)=L^2(\Omega)\ominus 1$：$1$と直交する$L^2(\Omega)$の部分空間

$M(X)$：$\{X(t); -\infty<t<\infty\}$の張る$L^{*2}(\Omega)$の線型部分空間

$X(s,t] \equiv \{X(\tau); s<\tau\le t\}$，$X\{N\}\equiv\{X(\tau); \tau\in N\}$

$\alpha,\beta,\cdots,\xi,\eta,\zeta,\cdots$：$\Omega$の（可測集合への）分割

ε：各点への分割　　$\nu=\{\Omega\}$：trivialな分割

$\mathbb{Z}$：エントロピー有限な分割の全体

$\alpha\le\beta$：βはαの細分

$\alpha\vee\beta$：共通の細分，$\alpha\wedge\beta$：α，βより粗い分割の中で最も細かい分割

$\mathbb{B}(\zeta)$：可測なζ-setsの全体（σ-algebra）

$H(\alpha)$，$H(X)$（$H(\alpha|\beta)$，$H(X|Y)$）：エントロピー（条件つきエントロピー）

$\bar{H}(X)(\bar{H}(Y|X))$：情報源のエントロピー（条件つきエントロピー）又はエントロピー生成速度（条件つき――）（p.14及びp.25）

$H^{\circ}(p) = H^{\circ}(X)$: differential entropy （pp. 9－10）

$H^{\circ}(\alpha;\omega)$ $(H_0(\alpha|\zeta;\omega))$: エントロピー密度（条件つきエントロピー密度）

$I(\xi,\eta)$, $I(X,Y)$: 情報量

$i(x,y) = i_{X,Y}(x,y)$: 情報密度 （p. 18）

$\bar{I}(X,Y)$, $\vec{I}(X,Y)$, $\tilde{I}(X,Y)$, $I^{(g)}(X,Y)$: 情報生成速度 （p. 24, 25）

$[W]$: 通信 , $\{W\}$: 再生精度 （p. 39）

$H(W) \equiv \inf_{P_{X,\tilde{X}} \in W} I(X,\tilde{X})$: 通信のエントロピー （p. 40）

$\{Q,V\}$: 伝達機構 （p. 41）

$C(Q,V) \equiv \sup I(Y,\tilde{Y})$: 通過能力 （p. 41）

T_{ξ} : factor automorphism

T_C : automorphismの成分

$h(T)$: automorphism のエントロピー

$P(\cdot|\zeta;C)$: 測度の標準系

C2. 情報理論

0 エントロピー

0.0 沿革

0.0.1. 今日，情報理論の名で呼ばれている分野に明確な範疇を与えることはできない。それ程目覚ましく発展しつゝある若々しい部門であり，応用分野を着々と開拓しつゝ各方面から研究されている理論である。いま範囲を数学だけに限ってみても，広く確率論に，また解析学に，或は統計学に，その研究手段と観点とを与えることによって大きく貢献している。

このように重要な地位を築きつゝある情報理論の濫觴は遡って1947−49年に亘って発表された Shannon の pioneer work にみることができる。彼は通信理論を数学の理論として定式化することを創め，通信のもっている情報の尺度として，エントロピーという概念を導入した。それは，偶然現象のもっている不確定さを測定する尺度として，極めて重要でありまた適切な量でもあることが情報理論の研究の進歩と共に立証された。Shannon は，通報を確率過程として把え，通報を発する情報源や通話路などの通信機構を数学的に記述し，エントロピーを用いて各種の特性量を定め，通報のもっている情報の量が送られていく様子を定式化して，それを通信理論の数学的な研究の出発点とした。当時 Shannon 及びその協力者達の初期の目標が実際面への応用にあったため，数学的に十分な厳密さをもったものでなかったのは已むを得ないことであろう。

［注］ エントロピーなる量は，古くから熱力学の中に見出される。我々は，いま Shannon の定義したエントロピーとの関係は形式的な類似以上のことは述べ得ないが，例えば，Boltzmann の H-定理については P. and T. Ehrenfest [34]， K. Yosida [197] 等の文献をあげておきたい。

0.0.2. Shannon の研究の概略を述べる前に，それと殆んど同時代に行なわれた Wiener の研究にふれておく。Wiener [186] は ergodic theorem を背景として定常過程の1個の見本過程より確率過程のもつ情報の量をなるべく多く把えようとする観点からの研究を進めていた。例えば雑音の加わっ

った信号を filter を通すことによって原信号をよりよく再生しようといった試みがそれである。しかしそこでは，エントロピーのような量を直接定義したり取扱ったりしてはいない。しかし，その後 [185] においては，Shannon の意味での情報量に相当するものが量的にも取扱われている。簡単な例によって，その思想をみると，例えば

成功の確率が $\frac{1}{2}$ の Bernoulli 試行を考え，その結果を二進小数で

$$(0.1) \qquad 0.\, a_1 a_2 \cdots a_n \cdots$$

とあらわす a_n が 0 か 1 かをきめるのは無限回あり，試行は毎回同じ状況で独立に行なわれるので，(0.1) のような数を知るためには無限個の事象についての情報が必要である。しかし実際の観測では誤差が伴う。その誤差が巾 2ε の間で一様分布するとし，

$$2\varepsilon = 0.\, b_1 b_2 \cdots b_m \cdots, \quad b_1 = \cdots = b_{k-1} = 0, \quad b_k = 1$$

とする。そのとき (0.1) をきめるのに a_k より先のものをみることは無意味となる。従って，a_i，$i \leqq k$ をきめるための観測回数 $(=k)$ は略 $-\log_2 0.b_1 b_2 \cdots$ となる。この例から事前に $[0,1]$ の中に（一様分布で）あることだけを知っていて，観測後 $[a,b]$ の中にあることが知られたとき事後の知識から得る情報の量は

$$-\log_2([a,b] \text{ の測度}) - [-\log_2([0,1] \text{ の測度})] = -\log_2 \frac{[a,b]\text{ の測度}}{[0,1]\text{ の測度}}$$

としてよかろう。より一般に事前確率を P_1，観測結果できまる事後確率を P_2 とするとき，この観測の提供する情報の量は $-\log \frac{P_2}{P_1}$ とみることにしようというのが Wiener の思想である。彼のこの考え方は Cybernetics [185] の中で大きくとりあげられている。

0.0.3. さて一方 Shannon が情報の量としてエントロピーを導入した根拠についてみてみよう。いま n 個の排反事象があって，その中のどれかが起こるとする。n 個の中のどれが実際に起こるかを知る以前では，我々は生起についての曖昧さとか不確定さをもっている。それはどの事象が起こったかを知った瞬間に解消してしまうもので，その結果我々は情報を得たとするものである。そのとき得た情報を数量的に表わそうとすればどのような量が適当であるかが次の問題になる。

n 個の事象の起こる確率をそれぞれ $p_1, p_2, \cdots, p_n$，$(\sum_{i=1}^{n} p_i = 1)$ とし，問題にしている情報の量を $H(p_1, p_2, \cdots, p_n)$ とかくとき H のみたすべき

望ましい性質として，次のようなものが考えられる。

1) H は $p_1, p_2, \cdots, p_n$ について連続な函数で，各 p_i が $\frac{1}{n}$ のとき最大値をとる。

2) $H(q_{11}, q_{12}, \cdots q_{1m_1}, q_{21}, \cdots, q_{2m_2}, \cdots, q_{nm_n})$

$$= H(p_1, p_2, \cdots, p_n) + \sum_{i=1}^{n} p_i H\left(\frac{q_{i1}}{p_i}, \cdots, \frac{q_{im_i}}{p_i}\right), \quad p_i = \sum_{k=1}^{m_i} q_{ik}.$$

3) $H(p_1, p_2, \cdots, p_n, 0) = H(p_1, p_2, \cdots, p_n)$.

2) について，簡単な場合にその意味を説明しよう。例えば $m_1 = m_2 = \cdots$ $m_{n-1} = 1$，$m_n = 2$，$p_n = q_1 + q_2$ のときを考えると，2) は

$$H(p_1, p_2, \cdots, p_{n-1}, q_1, q_2) = H(p_1, p_2, \cdots, p_n) + p_n H\left(\frac{q_1}{p_n}, \frac{q_2}{p_n}\right)$$

となる。その意味は最初にそれぞれ確率 $p_1, p_2, \cdots, p_n$ なる n 個の事象が考えられ，もし n 番目のものが起こったらそれは更に2つの事象（各々は確率 q_1, q_2）に分かれてその1つが起こるものとする。まず n 個のうちのどれが起こったかを知ることによって得る情報の量が $H(p_1, \cdots, p_n)$ で，若し，起こったのが n 番目の事象であったら（その確率は p_n）その条件つき確率空間で考えれば，2つの事象のうちどちらが起こるかを知ることによって $H\left(\frac{q_1}{p_n}, \frac{q_2}{p_n}\right)$ だけの情報量を得ることになる。そのような場合は確率 p_n で実現するため，全体としてみれば当初の $H(p_1, \cdots, p_n)$ に $p_n H\left(\frac{q_1}{p_n}, \frac{q_2}{p_n}\right)$ を加えたものが，総情報量と考えられる。ところが見方をかえると始めから $n+1$ 個の事象を考えていることとしてもよく，それは定義から $H(p_1, p_2, \cdots, p_{n-1}, q_1, q_2)$ に他ならない。これが 2) の要請の意味である。1) の前半については当然で，後半はどれもが同じ確率で起こるときが一番不確定度が大きいとする要請で，3) は確率0でしか起こらないような事象は情報量には影響しないことを意味する。

上に要求した条件1)，2) 及び3) をみたすような $H(p_1, \cdots p_n)$ は幸にも存在して，一意的に

$$(0.2) \qquad H(p_1, p_2, \cdots, p_n) = -\lambda \sum_{i=1}^{n} p_i \log p_i, \qquad \lambda \text{は正の定数},$$

と表わされることが知られる（定理1）。これをエントロピーと呼ぶ。

0.0.4. これまでに偶然現象のもつ不確定さを表わす量として，エントロピー H が適したものであることを知ったが，次に Shannon の original な仕事である通信理論の数学的取扱いの中にこの量がどのように活用されたかをみ

ることにする．通信文が送られてくる以前には，通常どの文字がどの程度の頻度（確率）で送られてくるかということのみが知られているに過ぎない。受信されたのをみて始めて，どの文字であるかが確定する。すなわち事前には不確定さがあったが受信後それが解消する．このようにみると，通報を出す情報源の一文字当りの情報の量 H を上の立場で定義し得ることがわかる。ところで通信は通話路を通って相手側に送られるので，途中雑音などの妨害で当初の情報とは異なったものが受信される．そこで途中で失われるものをなるべく少なくすることが望ましい．そのため通話路のもつ容量とか，情報の伝達される割合（transmission rate）のような量が登場する。

Shannon はこのような概念を駆使して，通信理論における多くの問題を取扱ったが，前述の如く，目標が応用を主としたため大まかな推論であったり，証明を見易くするため研究目的の制限が行なわれたと思われる所が多い。

その後，情報理論の数学的基礎づけに成功したのは Khintchin [81]，[82]である。始め，McMillan は离散型の情報源や通話路等の基本概念を始めて厳密に定義し，情報源のもつエントロピーの定義について Shannon の結果をより一般な定常過程の場合にまで拡張した。さらに Feinstein は McMillan の方法を踏襲し，雑音のある通話路に関する Shannon の定理の厳密な叙述を行なうと共に，新しい概念（判別可能な chain）を用いてそれを証明した。Khinchin の仕事[82]はこれらを集大成したものといってよく，离散型の場合のこの方向の問題を解明し，体系づけたといえよう。

Khinchine に続いて，情報理論は数学の中でも急速にその研究分野を広めていった。一つは Kolmogorov，Dobrushin 等による通信理論の基礎づけで，Shannon-Khintchin の研究と直接につながる方向である。この方向は応用面との接触も益々活発になっている。数学的な部分については本書の第1，2章で述べるし，また発生的な過程での紹介は後節（$\boxed{0.3}$）に譲る。

もう一つの方向は，エントロピーの概念が確率過程論や，解析学のいくつかの分野の研究に取り入れられて，固有な発展をしていることがあげられる。その中での典型的なものについてみてみよう。

0.0.5. まず第一にあげられるのは flow の研究への応用である．測度空間の上の automorphism や flow に対してもエントロピーはそれのもつ本来の意味を保ち乍ら定義することができて，それが新しい metrical invariant になっていることが Kolmogorov によって示された（1958年.[87]).

よく知られたように古典力学系の metrical theory の重要な部分は，抽象的に測度空間 $(\Omega, \mathbb{B}, P)$ 上の flow $\{T_t\}$ の理論として表わされる．P は正規化しておくことにより確率測度とみなしてよい。flow $\{T_t\}$ とは，1 対 1 保測変換 T_t，$-\infty < t < \infty$，の作る 1 助変数群を意味する．Ω，Ω' における flow $\{T_t\}$ と $\{T_t'\}$ とは，1 つの同型写像（isomorphism）S によって

$$T_t = S^{-1} T_t' S \quad , \quad -\infty < t < \infty$$

となるとき 同値 或は同じ metrical type をもつという。

$\{T_t\}$ から

$$U_t f(\omega) \equiv f(T_t \omega) \quad , \quad f \in L^2(\Omega)$$

により $L^2(\Omega)$ 上のユニタリ作用素の 1 助変数群 $\{U_t\}$ が定義される．$\{U_t\}$ の spectral type を flow $\{T_t\}$ の spectral type という。metrical 及び spectral type は flow の分類の上で重要な役割を果している。多くの transitive な flow が $L^2(\Omega)$ の 1 と直交する部分空間で σ-Lebesgue スペクトルをもつが，“2つの σ-Lebesgue スペクトルをもつ flow は互に同値であるか？”というのが open problem であった．Kolmogorov[ ] は，エントロピーが 1 つの metrical invariant になっていることを示し，かつ，任意の h に対して，エントロピーが h であり σ-Lebesgue スペクトルをもつような flow の存在を示すことにより，σ-Lebesgue スペクトルをもつ flow の中に $\aleph$ 個の metrical type が存在することを証明した。それは上の問題に対する否定的解答である．

Kolmogorov は亦今日 Kolmogorov flow と呼ばれる特別な flow をとりあげている：

$\mathbb{B}$ の sub-σ-algebra $\mathbb{B}_0$ が存在して，$T_t \mathbb{B}_0 \equiv \mathbb{B}_t$ とするとき

$$t < t' \text{ なら } \mathbb{B}_t \subseteq \mathbb{B}_{t'} , \quad \bigvee_t \mathbb{B}_t = \mathbb{B} , \quad \bigwedge_t \mathbb{B}_t = \{\phi, \Omega\} \ (\mathrm{mod.}\, 0).$$

このようなものは flow の中でも特に重要な位置を占めて居り，それは正のエントロピーをもち かつ σ-Lebesgue スペクトルをもつことが知られている。なお，本書では述べ得なかったけれども，flow の理論をめぐるこの方面の研究は，負の曲率をもつ manifold 上の geodesic flow の詳しい研究にまで発展していることに注意したい。

この他 解析学では，系統的な発展をみている ε-entropy を用いた函数空間の研究があげられる。Shannon は情報源である確率過程の path の作る函数空間が，時間と共に増加する様子を記述するのに dimension rate なる

量を提起している。それは恰も1つの path を誤差εで指定するのに必要な単位時間当りの次元数の如く考えられる。これと同様な思想の下に Kolmogorov や Tikhomirov は，函数空間の集合に対して ε-entropy を定義し，各種のクラスの集合の大きさを測る手段として有効にそれを用いている。Mityagin [103] の結果は興味あるものである。

さらに情報量の概念は数理統計学の中にも見られる。

以上，これらは情報理論の応用というよりは寧ろそれ自身の理論の深遠さを物語っているというべきであろう。

0.1 測度のエントロピー． 離散型の場合

0.1.1. 生起する事象の不確定の度合いを示す量 エントロピーの定義は，離散型の場合と連続型の場合とでは多少事情を異にするので別々に扱う。本節では離散型測度の場合を考える。

$(\Omega, \mathcal{B}, P)$ を確率空間とする．αは有限個の排反事象の系 $\{A_i\}$ で

$$\bigcup_{i=1}^{n} A_i = \Omega, \qquad A_i \in \mathcal{B},$$

即ち Ω の分割とし，$P(A_i) = p_i$，$(\sum_{i=1}^{n} p_i = 1)$ とする。このとき α のもつエントロピー H は次式により定義される．

$$H \equiv -\sum_{i=1}^{n} p_i \log p_i \ . \tag{0.3}$$

この H を $H(\alpha)$，又は $H(p_1, p_2, \cdots, p_n)$ とかく。但し上式の和で $p_i = 0$ のときは $p_i \log p_i = 0$，と約束する。なお，対数の底は2又はeがよく用いられる。以後，指定する必要のあるときのみその都度注意する．

αの不確定さを表わす量として（0.3）のような定義が妥当であることは次の定理による。

定理1.（Shannon [161], Khintchin [81]）
任意のnと任意の $\sum_{i=1}^{n} p_i = 1$，$p_i \geqq 0$，をみたす $\{p_i\}$ に対して定義された函数 $H(p_1, p_2, \cdots, p_n)$ が条件，

1) nを固定すれば，$H(p_1, p_2, \cdots, p_n)$ は $p_1, p_2, \cdots, p_n$ について連続であり，かつすべての p_i が等しい値 $\frac{1}{n}$ のときその最大値をとる.
2) $p_i = q_{i_1} + q_{i_2} + \cdots + q_{im_i} > 0$，$\sum_{i=1}^{n} p_i = 1$ とするとき

$$H(q_{11}, q_{12}, \cdots, q_{1,m_1}, q_{21}, \cdots, \cdots, q_{nm_n})$$
$$= H(p_1, p_2, \cdots, p_n) + \sum_{i=1}^{n} p_i H\left(\frac{q_{i1}}{p_i}, \cdots, \frac{q_{im_i}}{p_i}\right), \quad p_i = \sum_{k=1}^{m_i} q_{ik} .$$

3) $H(p_1, p_2, \cdots, p_n, 0) = H(p_1, p_2, \cdots, p_n)$.

をみたすとき，適当な正の定数 λ が存在して，H は

$$(0.4) \qquad H(p_1, p_2, \cdots, p_n) = -\lambda \sum_{i=1}^{n} p_i \log p_i$$

と表わされる。

［注］ i）定理は各 n 毎に領域 $\sum_{i=1}^{n} p_i = 1$，$p_i \geqq 0$，で H が定まること，すなわち函数の system がきまることを主張している。

ii）(0.4) で λ の値は本質的でない。$\lambda = 1$ として (0.3) を得る。

定理の証明は，$H(\frac{1}{n}, \cdots, \frac{1}{n}) = L(n)$ とおくとき，

$$L(ml) = L(m) + L(l)$$

が導かれ，$L(n) = \lambda \log n$ がでる。これと 2) を用いて各 p_i が $\frac{n_i}{n}$（n_i は整数，$\sum_i n_i = n$）とかけるとき，(0.4) が証明される。一般の場合については 1) の連続性を用いればよい。

<u>0.1.2.</u> H の定義には函数 $-x \log x$ が用いられているが，この函数が（狭義の）凸函数であることを用いてエントロピーに関する多くの基本的な関係式が導かれる。

<u>エントロピーの性質</u>

a) 任意の $p_1, p_2, \cdots, p_n$（$p_i \geqq 0$，$\sum_{i=1}^{n} p_i = 1$）に対し

$$H(p_1, p_2, \cdots, p_n) \geqq 0$$

であり，$H(p_1, p_2, \cdots, p_n) = 0$ なるための必要かつ十分な条件はすべての p_i が 0 か 1（1 になるのは唯一つ）となることである。

b) 各 p_i が $\frac{1}{n}$ に近づくにつれて $H(p_1, p_2, \cdots, p_n)$ は増大する，即ち

$$\sum_{i=1}^{n} a_{ij} = \sum_{j=1}^{n} a_{ij} = 1, \quad a_{ij} \geqq 0, \quad i, j = 1, 2, \cdots, n.$$

なる $\{a_{ij}\}$ により $\{p_i\}$ の平均化：

$$p_i' = \sum_{j=1}^{n} a_{ij} p_j ,$$

をすれば，

$$H(p_1, p_2, \cdots, p_n) \leqq H(p_1', p_2', \cdots, p_n'). \tag{0.5}$$

特に等号が成立するのは $\{p_i'\}$ が $\{p_i\}$ の並べかえになっている場合に限る。

$\alpha=\{A_1, A_2, \cdots, A_m\}$, $\beta=\{B_1, B_2, \cdots, B_n\}$ をそれぞれ Ω の分割とし

$$P(A_i)=p_i, \qquad P(B_j)=q_j$$

とする。$\alpha \vee \beta$ で両分割の細分 $\{A_i \cap B_j;\ i=1,2,\cdots,m,\ j=1,2,\cdots,n\}$ を表わし，

$$P(A_i \cap B_j)=\pi_{ij}, \qquad P(B_j \mid A_i)=p_{ij}$$

とかくとき，

$$H(\alpha \vee \beta)=H(\pi_{11}, \cdots, \pi_{1n}, \pi_{21}, \cdots, \pi_{mn})$$

を<u>同時エントロピー</u>という。また<u>条件(α)つきエントロピー</u> $H(\beta|\alpha)$ を

$$H(\beta|\alpha)=-\sum_{i=1}^{m} p_i \sum_{j=1}^{n} p_{ij} \log p_{ij}=-\sum_{ij} \pi_{ij} \log p_{ij}$$

で定義する．簡単な計算で次の c), d), e) がでる．

c) $H(\alpha \vee \beta)=H(\alpha)+H(\beta|\alpha)$

d) $H(\beta|\alpha) \leqq H(\beta)$

e) α と β が独立（$\pi_{ij}=p_i \cdot q_j$ がすべての i, j について成立つ）のとき，しかもそのときに限り

$$H(\beta|\alpha)=H(\beta)$$
$$H(\alpha \vee \beta)=H(\alpha)+H(\beta).$$

$\alpha_1, \alpha_2, \cdots, \alpha_n$ を Ω の分割とするとき，$H(\alpha_1 \vee \alpha_2 \vee \cdots \vee \alpha_n)$ について上の c), d), e) の拡張に当る次の性質が容易に導かれる。

c') $H(\alpha_1 \vee \alpha_2 \vee \cdots \vee \alpha_n)=H(\alpha_1)+H(\alpha_2|\alpha_1)+\cdots+H(\alpha_n|\alpha_1 \vee \cdots \vee \alpha_{n-1}) \leqq \sum_{i=1}^{n} H(\alpha_i)$

d') $H(\alpha_n|\alpha_1 \vee \cdots \vee \alpha_{n-1}) \leqq H(\alpha_n|\alpha_2 \vee \cdots \vee \alpha_{n-1})$.

e') d') で等号が成立するのは，

$$P(A_n \cap A_1/A_2 \cap \cdots \cap A_{n-1})=P(A_n|A_2 \cap \cdots \cap A_{n-1})P(A_1|A_2 \cap \cdots \cap A_{n-1})$$

がすべての A_i（α_i の元），$i=1,2,\cdots,n$，に対して成立するときで，しかもそのときに限る。

<u>0.1.3.</u>　確率変数のもつエントロピー

$X(\omega)$ を $(\Omega, \mathbb{B}, P)$ 上の確率変数で，有限個の値 $a_1, a_2, \cdots, a_n$ のみを

とるものとする。

$$A_i=\{\omega;\ X(\omega)=a_i\},\qquad \alpha=\{A_1, A_2, \cdots, A_n\},$$
$$P(A_i)=p_i\ ,\quad i=1,2,\cdots,n,$$

として $H(X)$ を

$$H(X)\equiv H(\alpha)=H(p_1, p_2, \cdots, p_n)$$

で定義して確率変数 X のもつエントロピーという.

同じく $Y(\omega)$ も有限個の値のみをとる確率変数で, 対応する Ω の分割を β とするとき, 同時エントロピーおよび条件 (X) つきエントロピー $H(Y|X)$ は, それぞれ

$$H(X,Y)=H(\alpha\vee\beta),\qquad H(Y|X)=H(\beta|\alpha)$$

で定義される. c), d, e) はすべて上のような記号のおきかえで X, Y に対するエントロピーの性質とみなせる.

有限個の値をとる確率ベクトル $(X_1(\omega), X_2(\omega), \cdots, X_n(\omega))$ に対しても, α_i を X_i に対応する Ω の分割とし,

$$H(X_1, X_2, \cdots, X_n)=H(\alpha_1\vee\alpha_2\vee\cdots\vee\alpha_n)$$
$$H(X_n|X_1, X_2, \cdots X_{n-1})=H(\alpha_n|\alpha_1\vee\alpha_2\vee\cdots\vee\alpha_{n-1})$$

等とすることにより, 同様な言いかえができて, c'), d'), e') がそのまま確率ベクトルのエントロピーの性質になる.

[注] これまでエントロピーは, Ω の有限分割や有限個の値をとる確率変数(ベクトル)に対するもののみを扱ってきたが, この概念は離散型分布をもつものであれば容易に無限個の場合に拡張されることを注意する. 例えば X が幾何分布 $\{pq^{n-1}\}\quad n=1,2,\cdots$ に従うときは

$$H(X)=-\sum_{n=1}^{\infty}q^{n-1}p(\log q^{n-1}+\log p)=\frac{1}{p}\left\{p\log\frac{1}{p}+q\log\frac{1}{q}\right\}.$$

0.2 測度のエントロピー. 連続型の場合

0.2.1. 分布 (R^1 の) が絶対連続で密度函数 $p(x)$ をもつ場合, その分布がもつ情報量を定義する. 離散型の場合の (0.3) の形式的な類似

$$H^\circ=H^\circ(p)\equiv-\int_{-\infty}^{\infty}p(x)\log p(x)\,dx \tag{0.6}$$

がその分布のエントロピーである.

また，確率変数Xの分布函数が密度函数 $p(x)$ をもつとき，Xのエントロピーを上の $H^\circ(p)$ で定義し $H^\circ(X)$ とかく。

［注］ $H^\circ(p)$ について注意することは（0.6）は離散型の場合のように，Ωの有限個の事象への分割を考え，それに対して各事象の生起する不確定さの尺度として定義するのではなくて，直接密度函数を用いて類似の定義をしたに過ぎないということである。離散型と連続型の場合との著しい差異は，例えば前者では $\varphi(x)$ を1対1写像とするとき，$H(X)=H(\varphi(X))$ であるが，後者ではφによって密度函数が変換され，一般にはその等式は成立しない点にある （後出 f））． 両者を区別するため H° を differential entropy と呼ぶことがある。

多次元分布の（確率ベクトルの）エントロピーも同様に定義される。すなわち（X_1, X_2, …, X_n） の分布が絶対連続で密度函数 $p(x_1,\cdots,x_n)$ をもつ場合は，その（differential）エントロピーは

$$H^\circ(X_1,\cdots,X_n)=H^\circ(p)=-\int\cdots\int p(x_1,\cdots,x_n)\log p(x_1,\cdots,x_n)\,dx_1\cdots dx_n$$

である。また条件つきエントロピーも離散型の場合と同様で，$p(x,y)$ を (X,Y) の，$p(x)$ をXの分布の密度とするとき

$$(0.7)\qquad H^\circ(Y|X)\equiv-\iint p(x,y)\log\frac{p(x,y)}{p(x)}\,dx\,dy$$

で与えられる。但し積分は $p(x)>0$ なる範囲で考える。この場合 X 又はY，或は両者共確率ベクトルとしてもよい。

連続型の場合にもエントロピーの性質は $-x\log x$ が凸函数であるということから導かれるものが多い。

以後 $p(x)\log p(x)$ が Riemann 積分の意味でも，Lebesgue 積分の意味でも，積分可能であるような $p(x)$ のみ（或はそのような分布に従う確率変数のみ）を扱うことにする。

0.2.2. $H^\circ(p)$ の性質

a） 分布が有限区間 $[a,b]$ 内に制限されている場合は一様の分布が最大のエントロピー $\log(b-a)$ をもつ。

この性質から直ちに，区間を制限しないときはいくらでも大きいエントロピーをもつ分布の存在がわかる。また離散型のときは H は常に $\geqq 0$ であったが，連続型のときは負のエントロピーをもつ分布（例えば $b-a<1$ なる区間にお

ける一様分布），及びいくらでも小さいエントロピーをもつ分布，が存在することが知られる。

b) 分布に平均操作を施すとエントロピーは増大する。すなわち与えられたpに対して

$$\int a(x,y)\,dx = \int a(x,y)\,dy = 1 , \qquad a(x,y) \geqq 0$$

なる $a(x,y)$ により平均したものを $p'(x)$ とする：

$$p'(x) = \int a(x,y)\, p(y)\, dy .$$

そのとき次の不等式がなりたつ：

$$H^\circ(p) \leqq H^\circ(p') .$$

c) $H^\circ(X,Y) = H^\circ(X) + H^\circ(Y|X)$.

d) $H^\circ(Y|X) \leqq H^\circ(Y)$.

e) 特にXとYが独立のとき，しかもそのときに限り

$$H^\circ(X,Y) = H^\circ(X) + H^\circ(Y), \qquad H^\circ(Y|X) = H^\circ(Y) .$$

また上の各式で X, Y を確率ベクトルとしても同様である．

前に注意したように $H^\circ(p)$ は座標系のとり方，すなわち scale に依存する量である。

f) $p(x_1, \cdots, x_n)$ が座標変換：

$$x = (x_1, x_2, \cdots, x_n) \longrightarrow y = (y_1, y_2, \cdots, y_n)$$

（但し，y_i はxの滑らかな函数で Jacobian $J\left(\frac{x}{y}\right) = det\left(\frac{\partial x_i}{\partial y_j}\right) \neq 0$）によって，$q(y_1, y_2, \cdots, y_n)$ に移ったとすれば

(0.8) $$H^\circ(q) = H^\circ(p) - \int\cdots\int p(x_1, x_2, \cdots, x_n)\, log\left|J\left(\frac{x}{y}\right)\right| dx_1 \cdots dx_n .$$

この系として，

f.1) 一次変換 $y_i = \sum_{j=1}^{n} a_{ij} x_j$, $i = 1, 2, \cdots, n$, のとき

$$H^\circ(q) = H^\circ(p) + log\,|\,det\,(a_{ij})\,| .$$

特に (a_{ij}) が回転を与えるときエントロピーは不変である．

f.2) Xを確率変数，aを定数とするとき

$$H^\circ(aX) = H^\circ(X) + log\,|a| .$$

f.3) $H^\circ(p)$ は座標の平行移動に関して不変である。

0.2.3. 具体的な例について H° の計算をしておく．対数の底は e である。

(i) 正規分布： $p(x_1,\cdots,x_n)=\dfrac{\sqrt{\det(a_{ij})}}{(\sqrt{2\pi})^n}\, e^{-\frac{1}{2}\sum_{i,j=1}^{n} a_{ij}x_ix_j}$ のとき

$$H^\circ(p)=\log(2\pi e)^{\frac{n}{2}}\,|\det(a_{ij})|^{-\frac{1}{2}}$$

特に一次元（$n=1$）のときは

$$H^\circ(p)=\log\sqrt{2\pi e}\,\sigma \qquad (\sigma^2 \text{ は分散}). \tag{0.9}$$

(ii) 指数分布： $p(x)=\frac{1}{m}e^{-\frac{x}{m}}$, $x\geqq 0$, のとき

$$H^\circ(p)=\log me \ . \tag{0.10}$$

0.2.4. $H^\circ(p)$ は p を動かしたとき，上にも下にも有界ではないが，p を適当なクラスに制限したとき，$H^\circ(p)$ が有界になり，その中で最大なエントロピーをもつ分布が決定される場合がある。 a）はその1例である．以下その他の典型的と考えられる例をあげるが，これは別な観点からすれば，エントロピーによる分布の特徴づけともみることができる。

前提となっている $p(x)\log p(x)$ が可積分という条件をみたし，かつ分散が σ^2 であるような p の全体を $\mathfrak{P}(\sigma^2)$ であらわす。

g) p が $\mathfrak{P}(\sigma^2)$ を動くとき $H^\circ(p)$ は上に有界である．詳しくいえば

$$\sup_{p\in\mathfrak{P}(\sigma^2)} H^\circ(p)\leqq c_1+c_2\log\sigma \ , \qquad c_1, c_2 \text{ は正の定数}$$

（下には有界でない）。

さらに強く，上の上限に等しいエントロピーをもつ分布の存在が証明される。そのため Jensen の不等式を用いて証明される次の補題を準備する。

補題 $p(x)\log p(x)$ 及び $p(x)\log q(x)$ が共に可積分なら

$$H^\circ(p)=-\int p(x)\log p(x)\,dx\leqq-\int p(x)\log q(x)\,dx\ .$$

この補題で $p\in\mathfrak{P}(\sigma^2)$, q を正規分布にとることにより次の主張が証明される。

h) $\mathfrak{P}(\sigma^2)$ の中では正規分布が最大のエントロピー $\log\sqrt{2\pi e}\,\sigma$ をもつ。

（Shannon [161] にこの証明があるが必ずしも厳密でない）。

同じく $p(x)\log p(x)$ が可積分で $p(x)=0$, $x<0$, かつ平均が存在して

それが m であるような p の全体を $\mathcal{P}(m)$ とかく。そのとき

i) $\mathcal{P}(m)$ の中で指数分布が最大のエントロピー $\log me$ をもつ。

証明はやはり補題で q を指数分布, $p \in \mathcal{P}(m)$ として容易にできる。

この他 確率ベクトルで，共分散が指定された場合等，いろいろな条件の下でエントロピーを最大にするものを求める問題がある。そのような場合補題が重要な役割を果すことが多い。

$\boxed{0.3}$ 情報量

0.3.1. ここでは情報の量として，エントロピーの他に，所謂 *information quantity* を考えたい。

形式的な定義をする前に，そのような量が考えられるようになった端緒を与えた通信機構を考えてみよう。それには Shannon による次のような模型を考えると便利である（Shannon [161]）。

$$
\begin{tikzcd}
\boxed{\text{情報源}} \arrow[r] & \boxed{\text{送信器}} \arrow[r] & \boxed{\text{通話路}} \arrow[r] & \boxed{\text{受信器}} \arrow[r] & \boxed{\text{目的地}} \\
 & & \boxed{\text{雑　音}} \arrow[u] & &
\end{tikzcd}
$$

以下必ずしも工学的な事実に忠実ではないかもしれないが，確率論の問題として取扱い易い形で上に出た用語を了解し，又一方通信の用語を借りて議論の内容の直観的な理解を深める一助ともしたい。

まず 情報源は伝達すべき通報を発する源である．例えば電信やテレタイプなどの場合なら通信文の集合が情報源で，それが送信器によって相手方の受信器に到着する。送信器では通報をより有効に送信できるよう，また機構上の必要からも，適当な符号に変え（これを *coding* という）た後送信する。送信器から受信器までの信号伝達の媒介となるものをすべて通話路（*channel*）の名で総称する。この通話路を介して伝達される途上で，送信された信号はいろいろな妨害を受けたりして，幾分歪んだものとなって受信器に入る。この際信号を歪ませる原因となるものをすべて雑音（*noise*）と呼ぶことにする．特別な場合として，送信されたものがそのままの形で受信されるような通話路は *noiseless* であると呼ばれる。

通信路を通った信号は受信器に入り，そこで情報源の発したものと同種の信号

に復元される．この操作を <u>*decoding*</u> という．*decode* された後通報として受信者に受け取られる．

情報源の発する通報にはいろいろな種類のものが考えられるが，代表的なものとして離散的な量の場合と連続的なものとに分けられる．本節では前者を中心にして述べる．離散的な中でも特に有限個の場合，詳しくいえば有限個の種類の符号のみが予め準備されていて，そのうち1個が適当な方法で逐次選ばれて信号となって送信器から発信されていくような場合はかなり詳しいことが知られている。

0.3.2.　以上の諸概念を確率論の言葉に直すことを試みる．以下簡単のため *coding* を必要としない場合を考える。まず よく知られている場合，すなわち情報源が有限 *state* の定常過程と考えられる場合についてみよう．

発信に用いられる符号の集合を $A=\{E_1, \cdots, E_a\}$ とする。Aを *input alphabet* と呼ぶことがある。これらが時刻nと共に *stochastically* に送られていき，しかも定常的であるとしよう。そのような情況の記述は次のような方法でよい。信号は符号の列

$$x=(\cdots, x_{-1}, x_0, x_1, \cdots, x_n, \cdots), \tag{0.11}$$

である．各 x_n はAの値をとる。xの全体を Ω_X とし，その *cylinder sets* から生成される σ-*algebra* を $\mathbb{B}_X$ とする．各信号xの使用される確率を定めることは，Ω_X上の確率測度 P_X を定めることであり，定常性は $(\Omega_X, \mathbb{B}_X, P_X)$ が定常過程になっていることである．すなわち，xが (0.11) で与えられるとき *shift* T:

$$T:\ (Tx)_n = x_{n+1}$$

が Ω_X の *automorphism* になっていることである．そこで我々は情報源として定常過程（確率変数は X_n，$\{X_n\} \equiv X$ とかく）が与えられたとして出発する．情報源の情報量を定義するため $\boxed{0.1}$ の意味に戻ってまず次のような和を考える．

$$H_n = -\sum_i P(c_i)\ \log P(c_i).$$

但し c_i は $x_0, x_1, \cdots, x_{n-1}$ が指定されたAの値をとるようなxの全体（Ω_X の *cylinder set*）で，和はそのような a^n 個の *cylinder sets* のすべてを亘る。（$\{c_i\}$は Ω_X の有限分割，H_n はそのエントロピー）。このとき *McMillan* の定理（$\boxed{3}$）により，次の極限が存在する．

$$\lim_{n\to\infty} \frac{H_n}{n} = \overline{H}(X). \tag{0.12}$$

これを情報源のエントロピーという．H_n の定義から，$\bar{H}(X)$ は単位時間（単位符号）当りの，情報源の情報量と考えてよいであろう．

0.3.3． 次に通話路についていえば，それをきめる要素として，1）送信号の state A 2）受信々号（output alphabet）の state B および 3）相互の関係を規定する条件つき確率 $P_{Y|X}(\cdot|x)$ がある．

x が送られ $y=(\cdots, y_{-1}, y_0, y_1, \cdots, y_n, \cdots)$，$\{y\}=\Omega_Y$ が受信されたとする．（こゝに $y_n\in B$，$n=0, \pm 1, \pm 2, \cdots$）．$\Omega_Y$ の cylinder sets から生成される σ-algebra $\mathbb{B}_Y$ の任意の元 N に対して，条件つき確率 $P_{Y|X}(N|x)$ が input と output との関係を定めている。従って，我々は $[A, P_{Y|X}(\cdot|x), B]$ を通話路と呼ぶことにする．

情報源（Ω_X，$\mathbb{B}_X$，P_X）と通話路 $[A, P_{Y|X}(\cdot|x), B]$ とが与えられたとき 確率空間（$\Omega_X\times\Omega_Y$，$\mathbb{B}_X\times\mathbb{B}_Y$，$P_{X,Y}$）が次のように定まる。

(0.13) $S=M\times N$, $M\in\mathbb{B}_X$, $N\in\mathbb{B}_Y$ ならば $P_{X,Y}(S)=\int_M P_{Y|X}(N|x)\,dP_X(x)$.

$P_{X,Y}$ は (X, Y) の同時分布を定めている。こゝで (X, Y) の同時エントロピーを考えたいが，そのためには(0.12)のような極限の存在を保障する条件として，定常性を要求しなければならない。$y\in\Omega_Y$ の shift も同じ記号 T で表わすと通話路が定常であるとは

(0.14) 任意の x, $N\in\mathbb{B}_Y$ に対し $P_{Y|X}(TN|Tx)=P_{Y|X}(N|x)$

がなりたつときをいう。通話路が定常で，情報源も定常過程ならば，T は $\Omega_X\times\Omega_Y$ 上の automorphism とも考えうる：

$$T(M\times N)\equiv TM\times TN\ ,\quad M\in\mathbb{B}_X,\quad N\in\mathbb{B}_Y,$$
$$P_{X,Y}(T(M\times N)) = P_{X,Y}(TM\times TN)$$
$$=\int_M P_{Y|X}(N|T^{-1}x)\,dP_X(T^{-1}x) = P_{X,Y}(M\times N).$$

これから (X, Y) も Y も定常過程になることがわかる．そこで $H_n(Y)$，$H_n(X, Y)$ は前と同じ意味に用いて

$$\lim_{n\to\infty}\frac{1}{n}H_n(Y)=\bar{H}(Y)$$
$$\lim_{n\to\infty}\frac{1}{n}H_n(X, Y)=\bar{H}(X, Y)$$

が存在する（A，Bは有限）。従って

$$\lim_{n\to\infty}\frac{1}{n}(H_n(X,Y)-H_n(X))=\bar{H}(X,Y)-\bar{H}(X)$$

$$\lim_{n\to\infty}\frac{1}{n}(H_n(X,Y)-H_n(Y))=\bar{H}(X,Y)-\bar{H}(Y)$$

も存在する．何れも条件つきエントロピーと呼び，それぞれ $\bar{H}(Y|X)$，$\bar{H}(X|Y)$ とかく。その直観的な意味を言えば，例えば $\bar{H}(X|Y)$ は上の極限をとる前の式をみるとき，yが受信されたことを知ったとき どの x が送信されたかをきめるのにどの程度の曖昧さがあるかを表わす量（平均的な量）と考えられよう。雑音の無いときは勿論 $\bar{H}(X|Y)=0$ である。

［注］ $\bar{H}(X|Y)$ は *equivocation*，$\bar{H}(Y|X)$ は *dissemination* とも呼ばれる。

$\bar{H}(X|Y)$ を前述のような意味に解するとき

(0.15) $$R(X,Y)=\bar{H}(X)-\bar{H}(X|Y)$$

は，当初にもっていた情報量 $\bar{H}(X)$ から送り得なかった情報量 $\bar{H}(X|Y)$ を減じたもので，真に送られた情報量と考えられる。その意味でRはYに含まれているXの情報の量とみることができ *rate of transmission* と呼ばれる．Rはまた $\bar{H}(X)+\bar{H}(Y)-\bar{H}(X,Y)$とかけば X と Y について対称である．従って，Rは X に含まれる Y の情報量とも考えることができる。すなわち，互に他に含まれる情報量といってよい。

Rを決めるものは情報源と通話路である。そこで

(0.16) $$C=\sup_{X} R(X,Y),$$

（但し 上限は *state* がAの *ergodic* な定常過程すべてについてとる），は通話路のみに関係した量で，通話路の容量（*capacity* 又は *ergodic capacity*）と呼ばれる．それは，この通話路を通って送り得る最大の情報量に他ならない。

［注］ *noiseless* のとき *Shannon* は別な容量の定義を与えているが同じ意味をもつものである。

0.3.4. *coding* の方法も併せて考えたとき，*Shannon* は情報源のエントロピーが$\bar{H}$，通話路の容量がCで，$\bar{H}<C$ ならば，いくらでも少ない割合の誤りで，その通話路を通って *output* が得られるような *coding* の方法があることを主張している．この事実は *Khintchin* ［８２］に厳密な形に述べられ，予報しない，かつ有限記憶をもつ場合に証明されている。

容量Cはそのように通話路を特徴づける重要な量であるが，Cが定義される前の$R(X,Y)$に戻ってみよう。Rは一方が他方についてもっている情報の量を表わしていた。この立場から二つの曖昧さをもったsystem，すなわち二種類のΩの分割ξ,ηがあったとき，一方がもっている他方についての情報の量といった概念に到達する。これをξがηについて（ηがξについて）もっている情報量$I(\xi,\eta)$が考えられる。

$I(\xi,\eta)$をGelfand, Kolmogorov, Yaglom等に従って定義する。
ξ,ηがΩの有限分割，$\xi=(A_1,\cdots,A_m)$，$\eta=(B_1,\cdots,B_n)$のとき

$$(0.17)\qquad I(\xi,\eta)\equiv\sum_{i,j}P(A_i\cap B_j)\log\frac{P(A_i\cap B_j)}{P(A_i)P(B_j)}$$

とする。これは極限をとらない前のRを定義する式と同じである。
次にξ,ηが一般の分割のとき

$$(0.18)\qquad I(\xi,\eta)\equiv\sup_{\substack{\xi_1\le\xi,\ \eta_1\le\eta\\ \xi_1,\eta_1\text{は有限分割}}}I(\xi_1,\eta_1)$$

と定義する。

0.3.5、 情報量$I(\xi,\eta)$の簡単な性質を述べれば

a) $0\le I(\xi,\eta)\leqq+\infty$で

$$I(\xi,\eta)=I(\eta,\xi),$$
$$I(\xi,\xi)\equiv H(\xi)$$

b) "$I(\xi,\eta)=0$"，と"ξとηが独立"，とは同等である。

c) $\xi_1\vee\eta_1$と$\xi_2\vee\eta_2$とが独立ならば，

$$I([\xi_1\vee\xi_2],[\eta_1\vee\eta_2])=I(\xi_1,\eta_1)+I(\xi_2,\eta_2).$$

d) $\xi_1\le\xi_2$ ならば 任意のηに対して，

$$I(\xi_1,\eta)\le I(\xi_2,\eta).$$

e) $\xi_1\le\xi_2\le\cdots\leqslant\xi_n\leqslant\cdots$，$\vee\xi_n=\xi$ ならば

$$\lim_{n\to\infty}I(\xi_n,\eta)=I(\xi,\eta)$$

等が定義から容易に証明される。

次に確率変数X,Yについて$I(X,Y)$を定義する。
X,Yが共に有限個の値をとる確率変数（確率ベクトル）のとき，X,Yのとる値をそれぞれ$x_1,\cdots,x_m$，$y_1,\cdots,y_n$，とし

$$(X=x_i)\equiv E_i,\qquad P(E_i)=p_i$$
$$(Y=y_j)\equiv F_j,\qquad P(F_j)=q_j,\qquad P(E_i\cap F_j)=p_{ij}$$

とすれば

$$I(X,Y) \equiv \sum_{ij} p_{ij} \log \frac{p_{ij}}{p_i q_j} \quad .$$

これは $\xi=\{E_1,\cdots,E_m\}$, $\eta=\{F_1,\cdots,F_n\}$ としたときの $I(\xi,\eta)$ である。

一般の確率変数 X,Y に対しては，それらの値域の各点が可測集合であることを仮定する。そうすれば $\boxed{0.1}$ の場合のように X,Y に対応する Ω の分割 ξ, η が得られる。それを用いて

(0.19) $\quad I(X,Y) \equiv I(\xi,\eta)$

とする。$I(X,X)$ は $\boxed{0.1}$ の意味でのエントロピー $H(X)$ である。

定理2 $I(X,Y)$ が有限であるための必要かつ十分な条件は，(X,Y) の分布 $P_{X,Y}$ が X 及び Y の分布の直積 $P_X\cdot P_Y$ に関し絶対連続となることである。そのとき，$I(X,Y)$ は次式で与えられる。

(0.20) $\quad I(X,Y)=\iint \alpha(x,y)\log\alpha(x,y)\,dP_X(x)\,dP_Y(y)$,

但し $\alpha(x,y)$ は $P_{X,Y}$ の $P_X\cdot P_Y$ に関する密度函数 $\left(\alpha(x,y)=\dfrac{dP_{XY}}{dP_X dP_Y}\right)$ である。$\log\alpha(x,y)\equiv i(x,y)$ を情報密度とよぶ。

特に $X=(X_1,\cdots,X_n)$, $Y=(Y_1,\cdots,Y_n)$ (n次元確率ベクトル), (X,Y) の分布が絶対連続（Lebesgue 測度に関し）ならば，密度函数を夫々 p_X, p_Y, p_{XY} とすると

(0.21) $\quad I(X,Y)=\int\cdots\int p_{XY}(x_1,\cdots,x_n;y_1\cdots y_n)\log\dfrac{p_{XY}(x_1\cdots x_n\, y_1\cdots y_n)}{p_X(x_1\cdots x_n)p_Y(y_1\cdots y_n)}\,dx_1\cdots dx_n\,dy_1\cdots dy_n$

と表わされる。

0.3.6. $I(\xi,\eta)$ に関する性質から容易に次のことがわかる。

a) $0\leqq I(X,Y)\leqq\infty$ で

$$I(X,Y)=I(Y,X)\leqq H(X)$$

b) $I(X,Y)=0$ と，XとYとが独立，とは同等である。

c) (X_1,Y_1) と (X_2,Y_2) とが独立ならば

$$I((X_1,X_2),(Y_1,Y_2))=I(X_1,Y_1)+I(X_2,Y_2)$$

d) X_1 がXの可測函数ならば，任意のYに対して

$$I(X_1,Y)\leqq I(X,Y)$$

e) $X_1,X_2,\cdots,X_n,\cdots$ があって，Xは $\widetilde{X}=(X_1,X_2,\cdots,X_n,\cdots)$ と同等，

すなわちXと$\tilde{X}$を可測にする最小のσ-$algebra$が同値ならば,

$$\lim_{n\to\infty} I((X_1, \cdots, X_n), Y) = I(X, Y).$$

f) 情報密度 $i(x,y)$をもつとき，次の不等式をみたす$\Gamma > 0$が存在する:

$$J(X,Y) \equiv \iint \alpha(x,y)\,|i(x,y)|\,dP_X(x)\,dP_Y(y) \leqq I(X,Y) + \Gamma\sqrt{I(X,Y)}.$$

エントロピーについては 0.2 の(0.9)でみたように変数変換に際して，$\log|J|$ の平均だけの変化が現われるが，情報量Iは，$scale$ の変換に対し不変な量であり，それだけに$differential\ entropy$よりも使い易い量になっている。

$\boxed{1}$ 情報量と情報的安定性

$\boxed{1.0}$ 序

この章では $\boxed{0.3}$ で定義した情報量 $I(X,Y)$，その概念を発展させた条件付情報量 $MI(Y,Z|X)$ や情報生成速度 $\overline{I}(X,Y)$，$I^{(g)}(X,Y)$ 等，及び情報量に関する大数の法則ともいうべき情報的安定について述べる。

$\boxed{2}$ で述べるように，Shannon-Khintchin の通信理論の新しい定式化が Dobrushin [30] によってなされたが，そこで導入された情報的安定性は，McMillan の定理を敷衍した基本的な概念であり，通信理論の研究に於て非常に重要な役割りを果すものである．

情報量や情報生成速度も $\boxed{0.3}$ で述べたように通信理論から生まれた概念であるが，Kolmogorov [86]，Gelfand-Yaglom [49]，Pinsker [128] 等により確率論的にそれ自体の研究が進められ，定常過程の regularity や 2 つの過程相互の従属性等，定常過程の構造と密接な関係をもつことが示された。情報的安定性と共に情報量や情報生成速度も，今後の定常過程の研究にとって興味ある対象であるように思われる。

本章の証明については Pinsker [128] を参照されたい。

$\boxed{1.1}$ 情報量 MI

1.1.1.　$(\Omega, \mathbb{B}, P)$ を基礎の確率空間，$X \equiv X(\omega)$，$(X,Y) \equiv (X(\omega), Y(\omega))$ によって導入される確率空間をそれぞれ $(K_X, \mathbb{B}_X, P_X)$，$(K_X \times K_Y, \mathbb{B}_X \times \mathbb{B}_Y, P_{X,Y})$ と書く．

P_1，P_2 を可測空間 $(K, \mathbb{B})$ 上の 2 つの確率測度，$\alpha \equiv \{C\}$ で K の有限分割を表わすとき，$\boxed{0.3}$ の $I(X,Y)$ の定義に準じて，P_2 に関する P_1 のエントロピー

$$H_{P_2}(P_1) \equiv \sup_{\alpha} \sum_{C} P_1(C) \log \frac{P_1(C)}{P_2(C)}$$

を定義する．$P_1 \ll P_2$（絶対連続）ならば $\boxed{0.3}$ 定理 2 に相当することがこの場合にも成り立ち，エントロピー密度 $h(x) \equiv h_{P_2}(x; P_1) \equiv \log \alpha_{P_2}(x; P_1)$ が定義できる．こゝに $\alpha_{P_2}(x; P_1) \equiv P_1(dx)/P_2(dx)$ である．

$\boxed{0.3}$ で定義した $I(X,Y)$, $i(x,y)$ はそれぞれ $H_{P_X\times Y}(P_{X,Y})$, $h_{P_X\times Y}(x,y;P_{X,Y})$ である。

1.1.2. 確率変数 X, Y, Z に対して,

(1.1) $\quad \bar{P}_{Y\times Z|X}(D\times E\times F)\equiv\int_D P_{Y|X}(E|x)P_{Z|X}(F|x)P_X(dx)$, $D\in\mathbb{B}_X$, $E\in\mathbb{B}_Y$, $F\in\mathbb{B}_Z$,

で $\mathbb{B}_X\times\mathbb{B}_Y\times\mathbb{B}_Z$ 上の確率測度を定義し, $\alpha\equiv\{C\}$ で $K_X\times K_Y\times K_Z$ の有限分割を表わすとき,

(1.2) $\quad MI(Y,Z|X)\equiv H_{\bar{P}_{Y\times Z|X}}(P_{X,Y,Z})=\sup_\alpha\sum_C P_{X,Y,Z}(C)\log\dfrac{P_{X,Y,Z}(C)}{\bar{P}_{Y\times Z|X}(C)}$

を (Y,Z) の条件 (X) つき情報量という。

$P_{X,Y,Z}\ll\bar{P}_{Y\times Z|X}$ ならば $\boxed{0.3}$ 定理2に相当することが成り立ち,

(1.3) $\quad MI(Y,Z|X)=\int_{K_X\times K_Y\times K_Z}\bar{i}_{Y,Z|X}(x,y,z)P_{X,Y,Z}(dxdydz)=E_{X,Y,Z}\bar{i}(Y,Z|X)$.

但し $\bar{i}_{Y,Z|X}(x,y,z)\equiv\log\bar{\alpha}_{Y,Z|X}(x,y,z)$, $\bar{\alpha}_{Y,Z|X}(x,y,z)\equiv P_{X,Y,Z}(dxdydz)/\bar{P}_{Y\times Z|X}(dxdydz)$, $\bar{i}(Y,Z|X)\equiv\bar{i}_{Y,Z|X}(X,Y,Z)$ である。$\bar{i}_{Y,Z|X}(x,y,z)$ を条件付情報密度という。

注意 1°) (Y,Z) の条件 $(X=x)$ つき分布 $P_{Y,Z|X}(N|x)$ が, 1) $\forall x\in K_X$ を固定すれば $N\in\mathbb{B}_Y\times\mathbb{B}_Z$ の函数として確率測度, 2) $\forall N$ を固定すれば x の函数として $\mathbb{B}_X$-可測函数と $a.e.(P_X)\,x$ で一致する, 3) $P_{Y,Z|X}(\cdot|x)\ll P_{Y\times Z|X}(\cdot|x)$ $a.e.(P_X)\,x$, をみたせば, (Y,Z) の条件 $(X=x)$ つき情報量 $I(Y,Z|x)\equiv H_{P_{Y\times Z|X}(\cdot|x)}(P_{Y,Z|X}(\cdot|x))$ が $a.e.(P_X)\,x$ について定義できて,

(1.4) $\quad I(Y,Z|x)=\int_{K_Y\times K_Z}\bar{i}_{Y,Z|X}(x,y,z)P_{Y,Z|X}(dy,dz|x)$ $a.e.(P_X)\,x$,

(1.5) $\quad MI(Y,Z|X)=E_X I(Y,Z|X)$

が成り立つ。こゝに $P_{Y\times Z|X}(\cdot|x)$ は $P_{Y\times Z|X}(E\times F|x)\equiv P_{Y|X}(E|x)P_{Z|X}(F|x)$, $E\in\mathbb{B}_Y$, $F\in\mathbb{B}_Z$, で定義される $\mathbb{B}_Y\times\mathbb{B}_Z$ 上の確率測度である。

2°) 密度 $p_X(x)$, $p_{X,Y}(x,y)$, $p_{X,Y,Z}(x,y,z)$ が存在すれば, 1°) の条件はすべてみたされるので (1.4), (1.5) が成り立ち, 条件つき情報密度は

(1.6) $\quad \bar{i}_{Y,Z|X}(x,y,z)=\log\dfrac{p_{X,Y,Z}(x,y,z)}{p_{Y|X}(y|x)p_{Z|X}(z|x)p_X(x)}=\log\dfrac{p_{Y,Z|X}(y,z|x)}{p_{Y|X}(y|x)p_{Z|X}(z|x)}$

である。

3°) Y, Z がそれぞれ有限個の値 $y_1,\cdots,y_\ell$, $z_1,\cdots,z_m$ をとる場合も, 1°) の条件はすべてみたされるので (1.4), (1.5) が成り立つ。このとき

$$(1.7) \qquad \bar{i}_{Y,Z|X}(x, y_i, z_j) = \log \frac{P_{Y,Z|X}(y_i, z_j|x)}{P_{Y|X}(y_i|x)P_{Z|X}(z_j|x)}$$

である。

条件つき情報量の性質としては

a) $0 \leq MI(Y,Z|X) = MI(Z,Y|X) \leq \infty$.

b) X_1, X_2, X_3 が Markov chain $\Leftrightarrow MI(X_1, X_3|X_2) = 0$.

c) (i) $I((X,Y),Z) = I(X,Z) + MI(Y,Z|X)$.

(ii) X が定数のとき $MI(Y,Z|X) = I(Y,Z)$.

(iii) 特に $P_{X,Y,Z} \ll P_{(X,Y)\times Z}$ ならば

$$i((X,Y),Z) = i(X,Z) + \bar{i}(Y,Z|X) \qquad a.e.(P).$$

c′) (i) $MI((X,Y),Z|U) = MI(X,Z|U) + MI(Y,Z|(U,X))$.

(ii) 特に $P_{X,Y,Z,U} \ll P_{(X,Y)\times Z|U}$ ならば

$$\bar{i}((X,Y),Z|U) = \bar{i}(X,Z|U) + \bar{i}(Y,Z|(U,X)) \qquad a.e.(P).$$

d) $MI(Y,Z|X) \leqq MH(Y|X)$.

確率空間 $(K, \mathbb{B}, P)$ と $\mathbb{B}$ の sub-σ-algebra $\widetilde{\mathbb{B}}$ があって, $\forall E \in \mathbb{B}$ に対して $P(E \ominus \widetilde{E}) = 0$ なる $\widetilde{E} \in \widetilde{\mathbb{B}}$ が存在するとき $\widetilde{\mathbb{B}}$ は P に関し<u>$\mathbb{B}$で到る処稠密</u>という。確率変数 $Y = f(X)$ (可測函数) は, σ-algebra$\{E;\ E \equiv f^{-1}(F),\ F \in \mathbb{B}_Y\}$ が P_X に関し $\mathbb{B}_X$ で到る処稠密のとき, <u>Xで到る処稠密</u>という。更に X が (X,Y) で到る処稠密のとき Y は X で <u>dominated</u> といい, $Y \prec X$ と書く。$Y = f(X)$ ならば $Y \prec X$ であり, 又 $Y = f(X)$ が X で到る処稠密ならば $Y \precsim X$ である.

e) (i) $\widetilde{Y} \prec Y$ ならば $MI(\widetilde{Y},Z|X) \leqq MI(Y,Z|X)$

$$= MI(\widetilde{Y},Z|X) + MI(Y,Z|(X,\widetilde{Y})).$$

(ii) $\widetilde{Y} \precsim Y$ ならば

$MI(\widetilde{Y},Z|X) = MI(Y,Z|X)$ かつ $\bar{i}(\widetilde{Y},Z|X) = \bar{i}(Y,Z|X)$ $a.e.(P)$.

f) $\widetilde{X} = f(X)$ が X で到る処稠密ならば

$MI(Y,Z|\widetilde{X}) = MI(Y,Z|X)$ かつ $\bar{i}(Y,Z|\widetilde{X}) = \bar{i}(Y,Z|X)$ $a.e.(P)$.

$X \equiv (X_1, X_2, \cdots)$, $Y \equiv (Y_1, Y_2, \cdots)$ のとき

g) (i) $\lim_{n\to\infty} MI((Y_1, \cdots, Y_n), Z|X) = MI(Y,Z|X)$.

(ii) $\lim_{n\to\infty} E|\bar{i}((Y_1, \cdots, Y_n), Z|X) - \bar{i}(Y,Z|X)| = 0$.

h) (i) $\underline{\lim}_{n\to\infty} MI(Y,Z|(X_1, \cdots, X_n)) \geq MI(Y,Z|X)$.

(ii) 特に $\lim_{n\to\infty} MI(Y,Z|(X_1, \cdots, X_n)) = MI(Y,Z|X)$ が成り立つとき

$$\lim_{n\to\infty} E\,|\,\bar{i}(Y,Z\,|\,(X_1,\cdots,X_n)) - \bar{i}(Y,Z\,|\,X)\,| = 0 .$$

YヌはZが有限個の値をとる確率変数ならば(ii)の条件がみたされる.

1.2 情報的安定性

正整数又は実数 $t \geq 0$ に関係する確率変数の族 $\{(X_t, Y_t, Z_t)\}$ を考える。

$\{(X_t, Y_t)\}$ が次の 1), 2), 3) の何れかをみたすとき情報的に安定であるという。

1) (i) $\forall t > T$ に対して　$0 < I(X_t, Y_t) < \infty$, かつ

(ii) $\forall \varepsilon > 0$ に対して　$\lim_{t\to\infty} P\left\{\left|\frac{i(X_t, Y_t)}{I(X_t, Y_t)} - 1\right| > \varepsilon\right\} = 0$。

2) $\forall t > T$ に対して P_{X_t, Y_t} が $P_{X_t \times Y_t}$ に関し特異である。

3) 情報密度 $i(X_t, Y_t)$ をもつ列 $\{(X_t, Y_t)\,;\, t \in N_1\}$ ともたない列 $\{(X_t, Y_t)\,;\, t \in N_2\}$ とに分け，それぞれが（片方のもの集合が有界のときは一方のみが）1) 又は 2) の何れかをみたす。

$\{(Y_t, Z_t)\}$ が次の 1)′, 2)′, 3)′ の何れかをみたすとき $\{X_t\}$ に関して情報的に相対安定であるという。

1)′ (i) $\forall t > T$ に対して　$0 < MI(Y_t, Z_t\,|\,X_t) < \infty$, かつ

(ii) $\forall \varepsilon > 0$ に対して　$\lim_{t\to\infty} P\left\{\left|\frac{\bar{i}(Y_t, Z_t\,|\,X_t)}{MI(Y_t, Z_t\,|\,X_t)} - 1\right| > \varepsilon\right\} = 0$.

2)′ $\forall t > T$ に対して　P_{X_t, Y_t, Z_t} が $P_{Y_t \times Z_t | X_t}$ に関し特異である。

3)′ 条件付情報密度 $\bar{i}(Y_t, Z_t\,|\,X_t)$ をもつ列 $\{(Y_t, Z_t)\,;\, t \in N_1\}$, $\{X_t\,;\, t \in N_1\}$ ともたない列 $\{(Y_t, Z_t)\,;\, t \in N_2\}$, $\{X_t\,;\, t \in N_2\}$ とに分け，それぞれが（片方のもの集合が有界のときは一方のみが） 1)′ 又は 2)′ の何れかをみたす。

a) $\forall t > T$ に対して　$0 < d < I(Y_t, Z_t) < \infty$ なる d が存在し, $\forall t > T$ に対して　$\tilde{Y}_t \prec Y_t$ かつ

$$\lim_{t\to\infty} \frac{I(\tilde{Y}_t, Z_t)}{I(Y_t, Z_t)} = 1$$

ならば $\{(Y_t, Z_t)\}$ と $\{(\tilde{Y}_t, Z_t)\}$ は一方が情報的に安定のとき他方もそうである。

b) $\forall t > T$ に対して $I(Y_t, Z_t) = \infty$ かつ $\tilde{Y}_t \prec Y_t$ で, $\{(\tilde{Y}_t, Z_t)\}$

が情報的に安定ならば $\{(Y_t, Z_t)\}$ もそうである。

情報的に相対安定の場合にも，a)， b) と全く同様なことがいえる。

1.3 定常過程の情報生成速度

確率過程 ($X \equiv \{X(t)\} \equiv \{X(t); -\infty < t < \infty\}$) の情報量は多くの場合 $I(X,Y) = \infty$, $MI(Y,Z|X) = \infty$ となるので，"単位時間当りの情報量" ともいうべきものを導入する必要がある。この節では（強）定常過程だけを取り扱い，時間径数 t を制限するときは $X(s,t] \equiv \{X(\tau); s < \tau \leq t\}$, $X\{N\} \equiv \{X(\tau); \tau \in N\}$ ($N \subset R^1$) と書く。定常超過程についても情報量や情報生成速度が定義されて，以下述べることは殆んど成り立つが，この手引きでは触れない (Gelfand-Yaglom [49] 又は Pinsker [128] を参照)。

$X \equiv \{X(t)\}$, $Y \equiv \{Y(t)\}$ を定常過程とする。

$$(1.8)\quad \begin{cases} \bar{I}(X,Y) \equiv \lim_{t\to\infty} \frac{1}{t} I(X(0,t], Y(0,t]) \\ \vec{I}(X,Y) \equiv \lim_{t\to\infty} \frac{1}{t} MI(X(0,t], Y \mid X(-\infty,0]) \\ \tilde{I}(X,Y) \equiv \lim_{t\to\infty} \frac{1}{t} I(X(0,t], Y) \end{cases}$$

の右辺の極限が（$+\infty$ も含めて）形式的に存在するとき，これらの量を Y に含まれる X の（各種の）情報生成速度という。但し $I(X(0,t], Y(0,t])$ 等は，$X(0,t]$, $Y(0,t]$ 等をそれぞれ確率変数と考えたときの情報量又は条件つき情報量である。

有限個の値をとる離散時間径数定常過程の場合は，(1.8) の諸量がすべて定義可能で

$$(1.9)\qquad 0 \leq \bar{I}(X,Y) = \vec{I}(X,Y) = \tilde{I}(X,Y) < \infty\ .$$

次に，$X \equiv \{X(t)\}$ が定常過程のとき，$\alpha(t) \equiv \{C_j(t)\}$ を時刻 t における $K_{X(t)}$（$X(t)$ の値域）の有限分割で $\forall t$ に対し $C_j(t) = C_j(0) \equiv C_j$ をみたすもの，$X^* \equiv \{X^*(t)\}$ を $X(t) \in C_j$ のとき $X^*(t) \equiv C_j$ と定義すれば，X^* は有限個の値をとる定常過程であるが，X に対するこのような X^* の族を $\mathcal{F}_X$ と書く。更に $X^{*(h)} \equiv \{X^{*(h)}(n)\} \equiv \{X^*(nh); n = 0, \pm 1, \pm 2, \cdots\}$ は有限個の値をとる離散時間径数定常過程となり，従って (1.9) より $\sup \frac{1}{h}\bar{I}(X^{*(h)},$

$Y^{*(h)})$, $\sup \frac{1}{h} \vec{I}(X^{*(h)}, Y^{*(h)})$, $\sup \frac{1}{h} \tilde{I}(X^{*(h)}, Y^{*(h)})$ は（∞ も含めて）すべて存在して等しい。こゝに sup はすべての $X^* \in \mathcal{F}_X$, $Y^* \in \mathcal{F}_Y$, $0 < h < \infty$ についてとる．(1.8) の諸量に加えて，

(1.8)′ $$I^{(g)}(X, Y) \equiv \sup \frac{1}{h} \bar{I}(X^{*(h)}, Y^{*(h)})$$

も1つの情報生成速度である。

情報生成速度の定義式 (1.8)，(1.8)′ で，特に $Y = X$ の場合をXの<u>エントロピー生成速度</u>という．<u>条件つき情報生成速度</u>も上に準じて定義される．

各種の情報生成速度の間には次の関係がある。

a) $\vec{I}(X, Y)$ は常に定義可能で，$\forall t > 0$ に対し

(1.9) $$0 \leq MI(X(0,1], Y | X(-\infty, 0]) = \frac{1}{t} MI(X(0,t], Y | X(-\infty, 0]) = \vec{I}(X, Y) \leq \infty .$$

b) それぞれ定義可能な場合は，

$$0 \leq \vec{I}(X, Y) \leq I^{(g)}(X, Y) \leq \bar{I}(X, Y) \leq \tilde{I}(X, Y) \leq \infty .$$

c) 或る $s > 0$ と或る集合 $N \subset (0, s)$ が存在して，

(1.10) $$\lim_{t \to \infty} \frac{1}{t} MI(X(-\infty, 0], X(s, t] | X\{N\}) = 0$$

をみたすとき，

(i) $I(X\{N\}, Y) < \infty$ ならば $\tilde{I}(X, Y)$, $\bar{I}(X, Y)$ が定義できて

$$\tilde{I}(X, Y) = \bar{I}(X, Y) = I^{(g)}(X, Y) = \vec{I}(X, Y) .$$

(ii) $\lim_{t \to \infty} \frac{1}{t} I(X\{N\}, Y(0, t]) = 0$ ならば $\bar{I}(X, Y)$ が定義できて

$$\bar{I}(X, Y) = I^{(g)}(X, Y) = \vec{I}(X, Y) .$$

c)′ (1.10) が成り立つ十分条件を3つ挙げる。

1) $I(X(-\infty, 0], X(s, \infty)) < \infty$.

2) $\forall t > 0$ に対して $I(X(-\infty, 0], X(0, t]) < \infty$.

3) $\forall t > s$ に対して $MI(X(-\infty, 0], X(s, t] | X(0, s]) < \infty$.

1), 2) の場合には，条件 (1.10) は $N = \phi$ として成り立つ．即ち

(1.10)′ $$\lim_{t \to \infty} \frac{1}{t} I(X(-\infty, 0], X(s, t]) = 0$$

となり，(i) は（従って (ii) も）追加条件 $I(X\{N\}, Y) < \infty$ ($\lim_{t \to \infty} \frac{1}{t} I(X\{N\}, Y(0, t]) = 0$) なしに成り立つ。

情報量 I，MI の性質及び情報生成速度の定義より，次に列挙する性質が容易にわかる。

d) (X,Y)と(Z,U)が独立ならば $\bar{I}((X,Z),(Y,U))=\bar{I}(X,Y)+\bar{I}(Z,U)$
（他の３つについても同様）.

e) X,Y,Z が Markov chain であれば $\vec{I}(X,(Y,Z))=\vec{I}(X,Y)$.
（$\tilde{I}$, $I^{(g)}$ についても同様）.

f) (i) $Z \prec X$ ならば $I^{(g)}(Z,Y) \leq I^{(g)}(X,Y)$.

(ii) 或る $T \geq 0$ があって，$\forall t$ に対して $Y(t) \prec X(-\infty, t+T]$ ならば
$\vec{I}(Z,Y) \leq \vec{I}(X,Y)$.

(iii) 或る $T>0$ があって，$\forall t$ に対し $Y(t) \prec X(t-T, t+T]$ ならば
$\bar{I}(Z,Y) \leq \bar{I}(X,Y)$ （$\tilde{I}$ についても同様）。

$\mathbb{B}_X(-\infty, t]$ を $\mathbb{B}_X$ の $-\infty < s \leq t$ における cylinder sets から生成される σ-algebra とする．$\bigcap_{-\infty<t<\infty} \mathbb{B}_X(-\infty, t]$ が，測度0か1の集合だけから成るとき定常過程 $X \equiv \{X(t)\}$ は regular といい，$\mathbb{B}_X$ で到る処稠密のとき singular であるという．Vinokurov [181] によれば $X \equiv \{X(t)\}$ が regular なことと，有限個の値をとる任意の確率変数 U に対して

$$\lim_{t \to -\infty} V(P_{X(-\infty,t],U},\ P_{X(-\infty,t] \times U}) = 0$$

が成り立つこととは同値である。こゝに $V(P_1, P_2)$ は確率測度 P_1, P_2 の variation である.[1] 更に条件

$$\lim_{t-s \to \infty} V(P_{X(-\infty,s],X(t,\infty)},\ P_{X(-\infty,s] \times X(t,\infty)}) = 0$$

をみたすとき completely regular といい，又 $\forall X^* \in \mathfrak{F}_X$, $\forall h>0$ に対して，$X^{*(h)} \equiv \{X^{*(h)}(n)\}$ が regular のとき weakly regular, $X^{*(h)}$ が singular のとき completely singular という．明らかに $X \equiv \{X(t)\}$ が regular ならば weakly regular であり，completely singular ならば singular である。

g) (i) $X \equiv \{X(t)\}$ が singular ならば $\vec{I}(X,Y)=0$.

(ii) X が regular のとき $\vec{I}(X;Y)=0 \Leftrightarrow X \perp\!\!\!\perp Y$.

---

1) $(K, \mathbb{B})$ 上の確率測度 P_1, P_2 があるとき，$\alpha \equiv \{C\}$ で K の有限分割を表わすとき

$$V(P_1, P_2) \equiv \sup_{\alpha} \sum_{C} |P_1(C) - P_2(C)|$$

を P_1 と P_2 の variation という.

g)' (i) X が completely regular ならば $I^{(g)}(X,Y)=0$。

(ii) X が weakly regular のとき

$I^{(g)}(X,Y)=0 \Longleftrightarrow X \perp\!\!\!\perp Y$ 又は Y が completely sigular.

1.4 定常過程の情報的安定性

定常過程 $X\equiv\{X(t)\}$, $Y\equiv\{Y_t\}$ に於て, $\bar{I}(X,Y)=0$ か, 又は $\{(X(0,t], Y(0,t]);\ t>0\}$ が情報的に安定のとき, 即ち

1) (i) $\forall t>T$ に対して $0<I(X(0,t],\ Y(0,t])<\infty$, かつ

(ii) $\forall\varepsilon>0$ に対して $\displaystyle\lim_{t\to\infty}P\left\{\left|\frac{i(X(0,t],\,Y(0,t])}{I(X(0,t],\,Y(0,t])}-1\right|>\varepsilon\right\}=0$

が成り立つか又は

2) $\forall t>T$ に対して $P_{X(0,t],Y(0,t]}$ が $P_{X(0,t]\times Y(0,t]}$ に関し特異であるとき[2], 定常過程 $(X,Y)\equiv\{(X(t),\ Y(t))\}$ は情報的に安定であるという.

$\vec{I}(X,Y)=0$ か, 又は $\{(X(-\infty,t],\ Y)\}$ が $\{X(-\infty,0]\}\equiv\{Z_t\}$ に対して情報的に相対安定のとき[2], 即ち

1)' (i) $\forall t>T$ に対して $0<MI(X(-\infty,t],\ Y|X(-\infty,0])<\infty$, かつ

(ii) $\forall\varepsilon>0$ に対して $\displaystyle\lim_{t\to\infty}P\left\{\left|\frac{\bar{i}(X(-\infty,t],\,Y|X(-\infty,0])}{MI(X(-\infty,t],\,Y|X(-\infty,0])}-1\right|>\varepsilon\right\}=0$

が成り立つか又は

2)' $\forall t>T$ に対して $P_{X(-\infty,t],Y}$ が $\bar{P}_{X(-\infty,t]\times Y|X(-\infty,0]}$ に関し特異であるとき[2], 定常過程 (X,Y) は情報的に相対安定であるという。

$(X,Y)\equiv\{(X(t),\ Y(t))\}$ が有限個の値をとる離散時間径数定常過程で, (X,Y) がエルゴード的ならば (X,Y) は情報的に安定且つ相対安定であることは, [3] に述べる エントロピーに関する McMillan の定理により容易に証明できる。この結果は次のように拡張できる。

定理 1 $(X,Y)\equiv\{(X(t),\ Y(t))\}$ が定常過程のとき, 1) (X,Y) がエルゴード的で $0\leq I^{(g)}(X,Y)=\bar{I}(X,Y)<\infty$ が成り立つか又は 2) $P_{X(0,t],\,Y(0,t]}$ が $P_{X(0,t]\times Y(0,t]}$ に関し特異となるような $t>0$ が存在するならば, (X,Y) は情報的に安定である。

---

2) [1.2] の定義における 3) 又は 3)' は起こりえない。

証明の概略を示す。(X,Y) が 1) の条件をみたす場合については $X^*\in\mathcal{F}_X$, $Y^*\in\mathcal{F}_X$ を用いて, $\forall\delta>0$ に対して

$$\frac{1}{h}\bar{I}(X^{*(h)},\ Y^{*(h)})/\bar{I}(X,Y)>1-\delta$$

をみたすように, $(X,Y)=\{(X(t),\ Y(t))\}$ を有限個の値をとる離散時間径数定常過程 $(X^{*(h)},\ Y^{*(h)})\equiv\{(X^*(nh),\ Y^*(nh))\}$ で近似し, 前述の結果及び $\boxed{1.2}$ a) を用いて証明できる。 2) の条件をみたす場合には, $\forall s\geq t$ に対し $X(0,t]$, $Y(0,t]$ は $X(0,s]$, $Y(0,s]$ の可測函数だから, 時刻 t で $P_{X(0,t],Y(0,t]}$ が $P_{X(0,t]\times Y(0,t]}$ に関し特異なら, $P_{X(0,s]\,Y(0,s]}$ が $P_{X(0,s]\times Y(0,s]}$ に関し特異であることがいえるので明らかである。

情報的相対安定性は定理1よりも弱い条件で得られる。即ち 1)′ (X,Y) がエルゴード的で $0\leq\vec{I}(X,Y)<\infty$ が成り立つか又は 2)′ $P_{X(-\infty,t],Y}$ が $\bar{P}_{X(-\infty,t]\times Y|X(-\infty,0]}$ に関し特異となるような $t>0$ が存在するならば, $(X,Y)=\{(X(t),\ Y(t))\}$ は情報的に相対安定である。

$\boxed{1.5}$ 正規系の情報量

1.5.1. 正規系の場合には情報量 I, MI や情報生成速度等を具体的に計算することができる。先ず基礎となる有限次元正規確率変数系の場合の $I(X,Y)$ の計算から始める。

$X\equiv(X_1,\cdots,X_n)$, $Y\equiv(Y_1,\cdots,Y_m)$, $(X,Y)\equiv(X_1,\cdots,X_n,Y_1,\cdots,Y_m)$ は正規系で共分散行列をそれぞれ $\Gamma_X\equiv(\gamma_{X_i,X_j})_{i,j=1,\cdots,n}$, Γ_Y, $\Gamma_{(X,Y)}$ とするとき

a) $\det\Gamma_{(X,Y)}\neq0$ ならば $\det\Gamma_X\neq0$, $\det\Gamma_Y\neq0$ で

$$I(X,Y)=\frac{1}{2}\log\frac{\det\Gamma_X\cdot\det\Gamma_Y}{\det\Gamma_{(X,Y)}}\quad. \tag{1.11}$$

b) $\det\Gamma_{(X,Y)}=0$, $\det\Gamma_X\neq0$, $\det\Gamma_Y\neq0$ ならば

$$I(X,Y)=\infty\quad. \tag{1.12}$$

c) $\det\Gamma_X=\det\Gamma_Y=0$ ならば, $\det\Gamma_X$, $\det\Gamma_Y$ の最大次数の0でない principal minor を与える $X\equiv(X_1,\cdots,X_n)$, $Y\equiv(Y_1,\cdots,Y_m)$ の部分列をそれぞれ $\tilde{X}\equiv(X_{i_1},\cdots,X_{i_s})$, $\tilde{Y}\equiv(Y_{j_1},\cdots,Y_{j_r})$ とするとき

$$(1.13)\qquad I(X,Y)=\begin{cases}\frac{1}{2}\log\dfrac{\det\Gamma_{\tilde{X}}\cdot\det\Gamma_{\tilde{Y}}}{\det\Gamma_{(\tilde{X},\tilde{Y})}} & \det\Gamma_{(\tilde{X},\tilde{Y})}\neq 0 \text{ のとき},\\ \infty & \det\Gamma_{(\tilde{X},\tilde{Y})}=0 \text{ のとき}.\end{cases}$$

d) 特に $n=m=1$ のときは

$$(1.14)\qquad I(X,Y)=-\frac{1}{2}\log(1-r_{X,Y}^2)\,;$$

こゝに $r_{X,Y}$ はXとYの相関係数である。

a) の場合は定理2を用いて $I(X,Y)$ を直接計算することにより容易にえられる。

b) の場合は $P_{X,Y}$ が $P_{X\times Y}$ に関し特異であるから明らか。

c), d) は a), b) のことから明らかである。

1.5.2. 一般の正規系について情報量を計算するための準備をする。

基礎の確率空間を $(\Omega,\mathbb{B},P)$, $X\equiv\{X_t\,;\,t\in N\}$ を $EX_t=0$, $DX_t<\infty$ なる $(\Omega,\mathbb{B},P)$ 上の（複素数値）確率変数の族とする。$L^{*2}(\Omega)$ を $EZ=0$, $DZ<\infty$ なるすべての確率変数の作る Hilbert 空間 $((Y,Z)\equiv EY\bar{Z})$, $M(X)$ を $X\equiv\{X_t\}$ の張る $L^{*2}(\Omega)$ の線型部分空間, π_X, $\pi_X^\perp$ をそれぞれ $M(X)$, $M(X)^\perp$ への射影作用素とする。$Z\in M(X)$ のとき $Z\prec X$ (linear) (linearly dominated) という。

$X\equiv\{X_t\,;\,t\in N\}$, $Y\equiv\{Y_s\,;\,s\in Q\}$ を上のような確率変数の族とする。作用素

$$\Pi_{XYX}\equiv\pi_X\pi_Y\pi_X\,,\qquad \Pi_{YXY}\equiv\pi_Y\pi_X\pi_X$$

を定義すれば,

1°) Π_{XYX}, Π_{YXY} は共に自己共役である。

2°) 一方が完全連続ならば他方も完全連続である。

3°) 固有値を共有し, 0でない固有値に対するそれぞれの固有ベクトルは, 互に他の固有ベクトルの射影（の定数倍）になるようにとれる。

4°) 固有値 $\lambda_j\neq 0$ に対する Π_{XYX}, Π_{YXY} の（3°のような）固有ベクトルをそれぞれ U_j, V_j とすれば

$$(1.15)\qquad 0<\lambda_j=|r_{U_j,V_j}|^2=|EU_j\cdot \bar{V}_j|^2/(E|U_j|^2\cdot E|V_j|^2)\leq 1\,.$$

3°) と 4°) の証明の概略を示す（1°), 2°) は定義より明らか）。

$\lambda_j V_j=\Pi_{XYX}U_j$ の両辺に π_Y を作用させると

$$\lambda_j \pi_Y U_j = \pi_Y \Pi_{XYX} U_j = \pi_Y \pi_X \pi_Y \pi_X U_j = \pi_Y \pi_X \pi_Y U_j$$
$$= \pi_Y \pi_X \pi_Y \pi_Y U_j = \Pi_{YXY} \ (\pi_Y U_j)$$

とかけるから 3°) が証明できる。

上と同様な計算により，$\pi_Y U_j = V_j$，$\pi_X V_j = \lambda U_j$ とできることが示され，

$$(U_j, V_j) = (\pi_X U_j, V_j) = (U_j, \pi_X V_j) = (U_j, \lambda_j U_j) = \lambda_j \| U_j \|^2,$$
$$(U_j, V_j) = (U_j, \pi_Y V_j) = (\pi_Y U_j, V_j) = (V_j, V_j) = \| V_j \|^2 .$$

これより $|(U_j, V_j)|^2 = \lambda_j \| U_j \|^2 \cdot \| V_j \|^2$ がえられるが，これは (1.15) 式に他ならない。

この結果を用いて，正規系の情報量の計算に基本的な次の定理が証明できる。

定理 2　$X \equiv \{X_t ; t \in N\}$，$Y \equiv \{Y_s ; s \in Q\}$，$(X, Y) \equiv \{X_t, Y_s\}$ が正規系で各 X_t，Y_s は実数値をとるとする。このとき

(A)　$I(X, Y) < \infty$ が成り立つ必要十分条件は，Π_{XYX}（又は Π_{YXY}）が完全連続で，有限 $trace$ をもつことである。

(B)　$I(X, Y) < \infty$ のとき，Π_{XYX} と Π_{YXY} の 0 でない固有値を λ_j $(j = 1, 2, \cdots)$，λ_j に対するそれぞれの固有ベクトルを U_j，V_j $(j = 1, 2, \cdots)$ とすれば，

(1.16)　$$I(X, Y) = \sum_j I(U_j, V_j) = -\frac{1}{2} \sum_j \log (1 - r^2_{U_j, V_j}) = -\frac{1}{2} \sum_j \log (1 - \lambda_j),$$

(1.17)　$$i(X, Y) = \sum_j i(U_j, V_j) \qquad a.e. (P).$$

先ず (A) の条件を仮定して (B) の証明の概略を述べる。

$U \equiv \{U_j\}$，$V \equiv \{V_j\}$，$\check{X} \equiv \pi_U X$，$\hat{X} \equiv \pi_U^{\perp} X$，$\check{Y} \equiv \pi_V Y$，$\hat{Y} \equiv \pi_V^{\perp} Y$ とおけば，$\check{X} \prec U \prec X \prec (\hat{X}, \check{X})$，$\check{Y} \prec V \prec Y \prec (\hat{Y}, \check{Y})$ だから（$\boxed{0.3}$ d）より

$$I(\check{X}, \check{Y}) \leq I(U, V) \leq I(X, Y) \leq I((\hat{X}, \check{X}), (\hat{Y}, \check{Y})) .$$

(X, Y) は正規系だから $(\hat{X}, \hat{Y}) \perp\!\!\!\perp (\check{X}, \check{Y})$．故に $\boxed{0.3}$ c）より

$$I((\hat{X}, \check{X}), (\hat{Y}, \check{Y})) = I(\hat{X}, \hat{Y}) + I(\check{X}, \check{Y}) .$$

しかるに $\hat{X}$，$\hat{Y}$ はそれぞれ固有値 0 の Π_{XYX}，Π_{YXY} の固有空間に属するから，$I(\hat{X}, \hat{Y}) = 0$ となることが証明できて，結局

$$I(\check{X}, \check{Y}) = I(U, V) = I(X, Y)$$

となることがわかる。

一方 $(U, V) = \{(U_1, V_1), (U_2, V_2), \cdots\}$，$(U_i, V_i) \perp\!\!\!\perp (U_j, V_j)$ $(i \neq j)$ だから $\boxed{0.3}$ c）より

$$I(U,V)=I((U_1,U_2,\cdots),\ (V_1,V_2,\cdots))=\sum_j I(U_j,V_j)$$
$$=\sum_j -\frac{1}{2}\log(1-|r^2_{U_j,V_j}|)=-\frac{1}{2}\sum_j \log(1-\lambda_j).$$

$\pi_{X,Y,X}$ の trace は有限だから
$$I(X,Y)=I(U,V)=-\frac{1}{2}\sum_j \log(1-\lambda_j)<\infty$$
が証明できた。

この証明で (B) と (A) の十分性は証明できたことになる。(A) の必要性の証明は π_{XYX} の定義と上の証明の一部分を用いて，$I(X,Y)<\infty$ から $\sum\lambda_j<\infty$ (trace 有限) がいえ，$\sum\lambda_j^2<\infty$ 即ち π_{XYX} の完全連続性が出る。

定理２の系として，次のことがいえる。

$(X,Y)\equiv\{X_t,Y_s;\ t\in N,\ s\in Q\}$ が正規系のとき，$I(X,Y)<\infty$ となる必要十分条件は，次の条件をみたす高々可算個の正規確率変数列 $\tilde{X}\equiv(\tilde{X}_1,\tilde{X}_2,\cdots)$, $\tilde{Y}\equiv(\tilde{Y}_1,\tilde{Y}_2,\cdots)$ が存在することである。

1) $\tilde{X}_j\in M(X)$, $\tilde{Y}_j\in M(Y)$, $j=1,2,\cdots$, かつ $j\neq k$ ならば $(\tilde{X}_j,\tilde{Y}_j)\perp\!\!\!\perp(\tilde{X}_k,\tilde{Y}_k)$.

2) $X,\tilde{X},\tilde{Y};Y$ は Markov chain をなす。

1.5.3. 条件つき情報量は，$(X,Y,Z)\equiv\{X_t,Y_s,Z_r;\ t\in N,\ s\in Q,\ r\in R\}$ が正規系のとき，$\hat{Y}\equiv\pi_X^\perp Y$, $\hat{Z}\equiv\pi_X^\perp Z$ とおけば

$$(1.18)\qquad MI(Y,Z|X)=I(\hat{Y},\hat{Z})$$

が成り立つことが容易にわかるので，普通の情報量 I の計算に帰着できる。

1.6 正規定常過程の情報生成速度

$(X,Y)\equiv\{(X(t),\ Y(t))\}$ が正規定常過程のときの情報生成速度についてしらべよう。

先ず $X\equiv\{X(t)\}$, $Y\equiv\{Y(t)\}$ がそれぞれ１次元正規定常過程の場合を取り扱う。(X,Y) のスペクトル密度函数（スペクトル函数の絶対連続部分の密度函数）を $f_{X,X}(\lambda)$, $f_{Y,Y}(\lambda)$, $f_{X,Y}(\lambda)$,

$$|R_{X,Y}(\lambda)|^2\equiv\begin{cases}\dfrac{|f_{X,Y}(\lambda)|^2}{f_{X,X}(\lambda)f_{Y,Y}(\lambda)} & f_{X,Y}(\lambda)\neq 0 \quad のとき, \\ 0 & f_{X,Y}(\lambda)=0 \quad のとき,\end{cases}$$

$$(1.19)\qquad L_{X,Y} \equiv -\frac{1}{4\pi}\int \log(1-|R_{X,Y}(\lambda)|^2)\,d\lambda$$

とおく．但し（1.19）の積分区間は，離散時間径数のときは $(-\pi,\pi)$，連続時間径数のときは $(-\infty,\infty)$ である．このとき次の結果がえられる．

a) (i) $X\equiv\{X(t)\}$ が singular でないならば $\vec{I}(X,Y)=L_{X,Y}$．

(ii) X が singular ならば $\vec{I}(X,Y)=0$．

b) 常に $I^{(g)}(X,Y)=L_{X,Y}$ が成り立つ。

c) 次の各条件の下で $\tilde{I}(X,Y)=\bar{I}(X,Y)=L_{X,Y}$ が成り立つ。[3]

1) X, Y が離散時間径数で X が singular でないとき。

2) スペクトル密度函数 $f_{XX}(\lambda)$ が

$$(1.20)\qquad f_{X,X}(\lambda)=\frac{Q(\lambda)}{P(\lambda)}|\psi(\lambda)|^2$$

と表わされるとき。但し $Q(\lambda)$, $P(\lambda)$ は多項式で，$\psi(\lambda)$ は $|\psi(\lambda)|\le 1$ 且つ $\int|\log|\psi(\lambda)||\,d\lambda<\infty$ をみたす．

d) スペクトル密度函数 $f_{X,X}(\lambda)$，$f_{Y,Y}(\lambda)$，$f_{X,Y}(\lambda)$ が有理函数で，$L_{X,Y}=\infty$ のときは，$\forall T>0$ に対して

$$(1.21)\qquad I(X(0,T],\ Y(0,T])=\infty\ .$$

e) $L_{X,Y}=\infty$ が成り立つとき，$Y(t)$ が

$$Y(t)=X(t)+Z(t),\quad X\equiv\{X(t)\}\perp\!\!\!\perp Z\equiv\{Z(t)\}$$

と分解できて，Z のスペクトル密度函数が（1.20）の形で書けるならば，（1.21）が成り立つ．

$X\equiv\{(X_1(t),\cdots,X_n(t))\}$，$Y\equiv\{(Y_1(t),\cdots,Y_m(t))\}$ がそれぞれ n 次元及び m 次元正規定常過程のときも同様な結果がえられる．即ち X，Y，(X,Y) のスペクトル密度行列（スペクトル函数の絶対連続部分の密度函数よりなる行列）をそれぞれ $\Delta_X\equiv(f_{X_i,X_j}(\lambda))$，$\Delta_Y\equiv(f_{Y_i,Y_j}(\lambda))$，$\Delta_{(X,Y)}$，とし，

$$(1.22)\qquad L_{X,Y}\equiv\frac{1}{4\pi}\int\log\frac{\det\Delta_{\tilde{X}}(\lambda)\cdot\det\Delta_{\tilde{Y}}(\lambda)}{\det\Delta_{(\tilde{X},\tilde{Y})}(\lambda)}\,d\lambda\ .$$

とおく．ここに $\det\Delta_{\tilde{X}}(\lambda)$（$\det\Delta_{\tilde{Y}}(\lambda)$ も同様）は $\det\Delta_X(\lambda)$ の最大次数の 0 でない principal minor を与える $\tilde{X}\equiv\{(X_{i_1}(t),\cdots,X_{i_s}(t))\}$ のスペクトル密度行列で，$\Delta_{(\tilde{X},\tilde{Y})}$ は $(\tilde{X},\tilde{Y})$ のスペクトル密度行列である．このと

---

3) 一般には $\tilde{I}(X,Y)\ge\bar{I}(X,Y)\ge I^{(g)}(X,Y)=L_{X,Y}$ （$\boxed{1.3}$ b)）

き次の結果がえられる．

a)′ (i) 一般に $\vec{I}(X,Y) \leq L_{X,Y}$ ．

(ii) Xが regular ならば $\vec{I}(X,Y) = L_{X,Y}$ ．

(iii) Xが singular ならば $\vec{I}(X,Y) = 0$ ．

b)′ 常に $I^{(g)}(X,Y) = L_{X,Y}$ が成り立つ．

c)′ 次の各条件の下で $\widehat{I}(X,Y) = \overline{I}(X,Y) = L_{X,Y}$ が成り立つ．

1) X, Y が離散時間径数で，$\det \Delta_X(\lambda) \neq 0$ のとき．

2) スペクトル密度函数 $f_{X_i,X_j}(\lambda)$ が有理函数のとき．

d)′ (X,Y)のスペクトル密度函数（$\Delta_{(X,Y)}$ の各元）が有理函数で，$L_{X,Y} = \infty$ が成り立つときは，$^\forall T>0$ に対して (1.21) $I(X(0,T],\ Y(0,T]) = \infty$ が成り立つ．

1.7 正規定常過程の情報的安定性

1.7.1. 正規系の場合には，1.5.2．定理2と次のa)，b)，c)を使って情報的安定性に関するくわしい結果がえられる．

(X,Y)を平均値0の2次元正規確率変数とする．

a)

(1.22) $$I^{(n)}_{X,Y} \equiv E(i(X,Y)-I(X,Y))^2 = \begin{cases} g(n)\, r^{n}_{X,Y} & n=2k, \\ 0 & n=2k-1, \end{cases} \quad k=1,2,\cdots .$$

特に $n=2$ のときは

(1.23) $$I^{2}_{X,Y} \doteq D_i(X,Y) \leq 2I(X,Y) .$$

但し

(1.24) $$g(n) = \frac{1}{2^n}\sum_{k=0}^{n}\left\{(-1)^k \cdot 1\cdot 3\cdots(2k-1)^* \cdot 1\cdot 3\cdots[2(n-k)-1]^* \binom{n}{k}\right\},$$

$$(2k-1)^* \equiv \begin{cases} 2k-1 & k \geq 1 , \\ 0 & k=0 . \end{cases}$$

b) $i(X,Y)$ の分布は次の確率変数Zの分布と同じである．

(1.25) $$Z \equiv \frac{1}{2} r_{X,Y}(U^2 - V^2) - \frac{1}{2}\log(1 - r^2_{X,Y}) .$$

ここにUとVは互に独立な標準正規確率変数である．

b)の証明の概略を示す．X, Yは標準正規確率変数として一般性を失わない．

変換

$$\tilde{X} = X\cos\frac{\pi}{4} + Y\sin\frac{\pi}{4}, \quad \tilde{Y} = X\sin\frac{\pi}{4} - Y\cos\frac{\pi}{4}$$

をほどこせば，$(\tilde{X}, \tilde{Y})$ も２次元正規確率変数で，$E\tilde{X} = E\tilde{Y} = 0$，$D\tilde{X} = 1 + r_{X,Y}$，$D\tilde{Y} = 1 - r_{X,Y}$，$E\tilde{X}\tilde{Y} = 0$ をみたす．従って密度函数の関係は

$$p_{X,Y}(x,y) = p_{\tilde{X},\tilde{Y}}(\tilde{x},\tilde{y}) = \frac{1}{\sqrt{2\pi}\sqrt{1-r_{X,Y}^2}}\ exp\left(-\frac{\tilde{x}^2}{2(1+r_{X,Y})} - \frac{\tilde{y}^2}{2(1-r_{X,Y})}\right)$$

$$p_X(x)p_Y(y) = \frac{1}{2\pi}exp\left(-\frac{x^2}{2} - \frac{y^2}{2}\right) = \frac{1}{2\pi}\ exp\left(-\frac{\tilde{x}^2}{2} - \frac{\tilde{y}^2}{2}\right)$$

となるから，情報密度は定義より

$$i_{X,Y}(x,y) = \log\frac{p_{X,Y}(x,y)}{p_X(x)p_Y(y)} = -\frac{1}{2}\log(1-r_{X,Y}^2) + \frac{r_{X,Y}\,\tilde{x}^2}{2(1+r_{X,Y})} - \frac{r_{X,Y}\,\tilde{y}^2}{2(1-r_{X,Y})}$$

となる．こゝで $U \equiv \tilde{X}/\sqrt{1+r_{X,Y}}$，$V \equiv \tilde{Y}/\sqrt{1-r_{X,Y}}$ とおけば (1.25) がえられる．

a) は Z の moment を計算することにより直ちにわかる。

これにより次の結果が証明できる．

c) $(X,Y) \equiv \{X_t, Y_s;\ t \in N,\ s \in Q\}$ が正規系のとき，

$$I(X,Y) = \infty \iff P_{X,Y} \text{ が } P_{X\times Y} \text{ に関し特異}$$

という結果を証明することができる。（$\boxed{0.3}$定理２より一般には右から左が出るがその逆はわからない）

こゝで正規系の情報的安定性に関する最も基本的な定理を示す．

定理３　(A)　t に関係する正規系 (X_t, Y_t) の族 $\{(X_t, Y_t);\ 0 < t < \infty\}$（$t$ は実数又は整数）が情報的に安定であるための必要十分条件は

$$(1.26)\qquad \lim_{t\to\infty} I(X_t, Y_t) = \infty$$

が成り立つことである．

(B)　$^\forall t$ に対して

$$(1.27)\qquad I(X_t, Y_t) < \infty \quad \text{且つ} \quad \lim_{t\to\infty} D_i(X_t, Y_t) = \infty$$

が成り立つとき

$$(1.28)\qquad \lim_{t\to\infty} P\left\{a < \frac{i(X_t, Y_t) - I(X_t, Y_t)}{\sqrt{Di(X_t, Y_t)}} < b\right\} = \frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{x^2}{2}}dx.$$

証明の概略を示す．

(A) の十分性．　$N_1 \equiv \{t;\ I(X_t, Y_t) < \infty\}$，$N_2 \equiv \{t;\ I(X_t, Y_t) = \infty\}$ とする．　この頁の c) により $I(X_t, Y_t) = \infty$ ならば P_{X_t, Y_t} は $P_{X_t \times Y_t}$ に関

し特異であるから $\{(X_t, Y_t);\ t\in N_2\}$ は（若し N_2 が有界集合でないなら）情報的に安定である。$\{(X_t, Y_t);\ t\in N_1\}$ については（若し N_1 が有界集合でないなら），$t\in N_1$，$t\to\infty$ のとき $\forall\varepsilon>0$ に対して

$$P\left\{\left|\frac{i(X_t,Y_t)}{I(X_t,Y_t)}-1\right|>\varepsilon\right\}\le\frac{D\left(\frac{i(X_t,Y_t)}{I(X_t,Y_t)}\right)}{\varepsilon^2}\le\frac{2I(X_t,Y_t)}{\varepsilon^2 I^2(X_t,Y_t)}=\frac{2}{\varepsilon^2 I(X_t,Y_t)}\to 0$$

となることから，情報的に安定であることがわかる。

(A)の必要性．“$I(X_t, Y_t)\not\to\infty$ ならば $\{(X_t, Y_t)\}$ は情報的に安定でない”ことを示す。$I(X_t, Y_t)\not\to\infty$ より次のような $t_1, t_2, \cdots$ と $C>0$ が存在する：$\lim_{k\to\infty} t_k=\infty$ 且つ

(1.29) $\qquad I=(X_t, Y_t)<c\ ,\qquad t=t_1, t_2, \cdots .$

X_t，Y_t に対して 1.5.2、定理２により $\Pi_{X_t\ Y_t\ X_t}$，$\Pi_{Y_t\ X_t\ Y_t}$ の固有値及び固有函数を $\lambda_1, \lambda_2, \cdots$，$(X_t)^*_1$，$(X_t)^*_2$，$\cdots$，$(Y_t)^*_1$，$(Y_t)^*_2$，$\cdots$ とすれば

(1.30) $\qquad i(X_t,Y_t)=\sum_j i((X_t)^*_j, (Y_t)^*_j)$，$i((X_t)^*_j,(Y_t)^*_j)\perp\!\!\!\perp i((X_t)^*_k,(Y_t)^*_k)\quad(i\neq k)$.

ここで

$$1^\circ)\quad \frac{Di((X_t)^*_j,(Y_t)^*_j)}{I(X_t,Y_t)}=\frac{r^2_{(X_t)^*_j,(Y_t)^*_j}}{-\frac{1}{2}\sum_k(1-r^2_{(X_t)^*_k,(Y_t)^*_k})}\to 0,\qquad t=t_1, t_2, \cdots\to\infty,$$

が成り立つ場合と

2°）或る j に対して

$$\frac{Di((X_s)^*_j,(Y_s)^*_j)}{I(X_s,Y_s)}>\delta_1>0\ ,\qquad s=s_1, s_2, \cdots$$

なる $\{t_1, t_2, \cdots\}$ の部分集合 $\{s_1, s_2, \cdots\}$ と δ_1 が存在する場合に分けて考える。

1°）の場合は $r^2_{(X_t)^*_j,(Y_t)^*_j}\to 0$，$t=t_1, t_2, \cdots\to\infty$ だから結局

$$\frac{Di(X_t,Y_t)}{I^2(X_t,Y_t)}=D\left(\frac{i(X_t,Y_t)}{I(X_t,Y_t)}\right)>\frac{2}{c}>0\qquad t=t_1, t_2, \cdots\to\infty$$

となって $i(X_t, Y_t)/I(X_t, Y_t)\not\to 1$ が言える。

2°）の場合には 1.7.1、b) を用いて，$\forall\varepsilon>0$ に対して

$$P\left\{\left|\frac{i(X_s,Y_s)}{I(X_s,Y_s)}-1\right|>\varepsilon\right\}\ge\delta\ ,\qquad s=s_1, s_2, \cdots\to\infty$$

をみたす $\delta>0$ が存在することが証明できる。

従って(A)の必要性が証明される。

(B)の証明． (1.30) の右辺の確率変数列 $i((X_t)_1^*, (Y_t)_1^*)$, $i((X_t)_2^*, (Y_t)_2^*)$, …… は互に独立で，a) を用いて計算すれば Liapounov の条件をみたすことがわかるから (1.28) が成り立つ。

［註］ 情報的相対安定性についても，全く同様な定理がえられる．

1.7.2. 正規定常過程の情報的安定性については，非常に明快な次の定理が成り立つ。

定理4 $X \equiv \{X(t)\} \equiv \{(X_1(t), \cdots, X_n(t))\}$, $Y \equiv \{Y(t)\} \equiv \{(Y_1(t), \cdots, Y_m(t))\}$, $(X, Y) \equiv \{(X(t), Y(t))\}$ がそれぞれ n 次元，m 次元，$n+m$ 次元正規定常過程のときは常に情報的安定且つ相対安定である。

実際 $\lim_{T\to\infty} I(X(0,T], Y(0,T]) = I(X(0,\infty), Y(0,\infty)) = \infty$ の場合と $\lim_{T\to\infty} I(X(0,T], Y(0,T]) = I(X(0,\infty), Y(0,\infty)) < \infty$ の場合とに分けて考えれば，前者は定理3より，後者は

$$\bar{I}(X,Y) = \lim_{T\to\infty} \frac{1}{T} I(X(0,T], Y(0,T]) \leq \lim_{T\to\infty} \frac{1}{T} I(X(0,\infty), Y(0,\infty)) = 0$$

より,情報的に安定となることがわかる。

情報的相対安定性は，「(X, Y) の情報的相対安定性と $(\hat{X}, \hat{Y})$ の安定性とは同値」((1.18) 参照) より明らか．ここに $\hat{X}_j(t) \equiv \pi^{\perp}_{X(-\infty,0]} X_j(t)$, $\hat{Y}_j(t) \equiv \pi^{\perp}_{X(-\infty,0]} Y_j(t)$, $\hat{X} \equiv \{(\hat{X}_1(t), \cdots, \hat{X}_n(t))\}$, $\hat{Y} \equiv \{(\hat{Y}_1(t), \cdots, \hat{Y}_m(t))\}$ である．

中心極限定理ともいうべき，定理3,(B) に対応する結果は次のようになる。

a)′ $(X, Y) \equiv \{(X(t), Y(t))\}$ が2次元正規定常過程で $\bar{I}(X, Y) \geq L_{X,Y}$ ならば

$$(1.31) \qquad \lim_{T\to\infty} P\left\{a < \frac{i(X(0,T], Y(0,T]) - I(X(0,T], Y(0,T])}{\sqrt{D i(X(0,T], Y(0,T])}} < b\right\} = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx.$$

a)″ $(X, Y) \equiv \{(X(t), Y(t)\}$ が多次元正規定常過程で，$0 < \bar{I}(X, Y) = L_{X,Y} < \infty$ ならば (1.31) が成り立つ．

$i(X, Y)$ の平均値 $I(X, Y)$ のまわりの moment については，$X \equiv \{(X_1(t), \cdots, X_n(t))\}$, $Y \equiv \{(Y_1(t), \cdots, Y_m(t))\}$ に対して，$U \equiv \{(U_1(t), \cdots, U_n(t))\}$, $V \equiv \{(V_1(t), \cdots, V_m(t))\}$ を 1) $U \precsim X$, $V \precsim Y$ (linear), 2) $(M(U_j), M(V_j))$, $j = 1, 2, \cdots, k \leq \min(n, m)$, が互に直交するように作り，

$$Q_{X,Y}^{(q)} \equiv Q_{U,V}^{(q)} \equiv \sum_{j=1}^{k} Q_{U_j,V_j}^{(q)} \equiv \frac{1}{\pi}\sum_{j=1}^{k} \int |R_{U_j,V_j}(\lambda)|^q \, d\lambda ,$$

(但し $|R_{U_j,V_j}(\lambda)|^2 \equiv |f_{U_j,V_j}(\lambda)|^2/(f_{U_j,V_j}(\lambda)\cdot f_{V_j,V_j}(\lambda))$) とすれば,

b) $X \equiv \{X(t)\}$, $Y \equiv \{Y(t)\}$, $(X,Y) \equiv \{(X(t),Y(t))\}$ が多次元正規定常過程で $\bar{I}(X,Y) = L_{X,Y}$ ならば

(1.32) $$\bar{I}_{X,Y}^{(q)} \equiv \lim_{T\to\infty} \frac{1}{T} I_{X(0,T],Y(0,T]}^{(q)} = \begin{cases} g(q)\, Q_{X,Y}^{(q)} & q = 2k , \\ 0 & q = 2k-1 ; \end{cases} \quad k = 1, 2, \cdots .$$

[註] b) で $\bar{I}(X,Y) = L_{X,Y}$ の条件がなかったら，一般には (1.32) の"="を"≧"でおきかえた関係がある.

c) (X,Y) が多次元正規定常過程で $\vec{I}(X,Y) = L_{X,Y}$ ならば

(1.33) $$\vec{I}_{X,Y}^{(q)} \equiv \lim_{T\to\infty} \frac{1}{T} I_{\hat{X}(-\delta,T],\hat{Y}}^{(q)} = \begin{cases} g(q)\, Q_{X,Y}^{(q)} & q = 2k , \\ 0 & q = 2k-1 , \end{cases} \quad k = 1, 2, \cdots .$$

但し $\hat{X}(-\delta, T] \equiv \pi^{\perp}_{X(-\infty,0]} X(-\delta, T]$, $\hat{Y} \equiv \pi^{\perp}_{X(-\infty,0]} Y$ で, $\delta > 0$ は任意のの正数, $I_{\hat{X}(-\delta,T],\hat{Y}}^{(q)} \equiv E(\bar{i}(X(-\delta,T], Y | X(-\infty,0]) - I(X(-\delta,T], Y | X(-\infty,0])^q$ である.

2 通信理論の基礎づけ

2.0 序

通信の確率論的定式化では，種々の情報源から生みだされる通信文は，ある確率過程として考えられる。この確率過程のもつ情報を，どれ程確実にキャッチし得るか，或いは漸近的にキャッチし得る可能性があるかを述べたものが，本章で最終的に証明しようとする *Shannon* の定理である．現実においては，情報源から生みだされる確率過程は，人為的に設けられた観測の手段である伝達機構（通話路）を通して，受信側にキャッチされるので，情報源のもつ情報の複雑さ，（内容の豊富さ）と伝達機構の機能的限度とのかねあいが通信の可能性を決定する。この章では記述を簡単にするために，確率過程の時間パラメーターは離散的場合のみを取扱う。今，かりに，相当の長時間内に情報源から送り出される言葉の数を L 個，その時間間隔内に通話路内を送られる言葉の数を K 個とすれば，この情報源の構造を充分よく知る得るためには，先ず $K>L$ なることが必要であろう。従って L によって情報源の構造を表現するのに，なるべく饒舌をさけて，本質的に内容のある言葉だけを採用して L を小さくする事につとめ，K については，信号の送り方を上手に工夫して，通話路の能力を充分発揮できるような送信を組み立てる事，即ち，K を大きくすることが問題点となる。前者，L についての1つの主張が後章でのべる *McMillan* の定理であり，後者，K についての命題が *Feinstein* の *Lemma* であって，これ等の組み合わせによって，通信の成立と可能性を主張するのが *Shannon* の定理である。こういう立場での通信理論の基礎づけ，及び定式化は，*Shannon* 以来，*McMillan*，*Feinstein* の *idea*，により発展させられ，*Kolmogorov*，*Khintchin* に至って数学的厳密化に成功している。更に最近，*Dobrushin* は，情報密度の情報的安定性の概念を使って，より広い立場からの定式化に成功した。彼は，命題の叙述や証明にあたって，通信理論の予備的知識を仮定せず，最小限必要な言葉の定義は，すべて測度論的語彙の範疇で定義することに努めている。従って読者は，従来あった通信理論特有の不慣れな諸量に惑わされずに理解できるようになった。この章に於ても，*Dobrushin* の定式化に負うところが大きい。

2.1 諸種の定義と例

2.1.1. 通信[W]とは"直積可測空間（$A\times\widetilde{A}$, $\mathbb{B}_A\times\mathbb{B}_{\widetilde{A}}$）上の確率測度のある集合$W$"のことをいう。又，同時分布が$W$に入るような任意の確率変数の組（$X,\widetilde{X}$）は再生精度{W}の条件をみたすという。

特に $W\ni P_{X\widetilde{X}}(\cdot)$ から $P_X(\cdot)\equiv P(\cdot\times\widetilde{A})$ によって導かれる $P_X(\cdot)$ を送信の分布という。

再生精度{W}は，伝達されるものXと，受信された量$\widetilde{X}$との同時分布についての制限であり，その制限のおき方は，M個の $\mathbb{B}_A\times\mathbb{B}_{\widetilde{A}}$-可測な函数 $\rho_k(x,\widetilde{x})$ $k=1,2,\cdots,M$，と M次元の集合$\overline{W}$ に対して

$$(2.1)\qquad (E\{\rho_1(X,\widetilde{X})\},\ E\{\rho_2(X,\widetilde{X})\},\cdots,\ E\{\rho_M(X,\widetilde{X})\})\in\overline{W}$$

であることとして与える。

（例1）

$A\equiv\{x=(z_1,z_2,\cdots,z_n);\ z_i=1$ 又は 0, $i=1,2,\cdots,n\}$, $\widetilde{A}\equiv\{\widetilde{x}=(\widetilde{z}_1,\widetilde{z}_2,\cdots,\widetilde{z}_n);\ \widetilde{z}_i=1$ 又は 0, $i=1,2,\cdots,n\}$, $\rho_k(x,\widetilde{x})\equiv|z_k-\widetilde{z}_k|$, $\overline{W}\equiv\{0\le\rho_n\le\alpha,\ k=1,2,\cdots,M,\ M\le n\}$ を与え，A, $\widetilde{A}$を値域とする確率変数をそれぞれ $X=(X_1,X_2,\cdots,X_n)$, $\widetilde{X}=(\widetilde{X}_1,\widetilde{X}_2,\cdots,\widetilde{X}_n)$とすれば，制限式(2.1)は

$$(2.2)\qquad E\{\rho_k(X,\widetilde{X})\}=P\{X_k\ne\widetilde{X}_k\}\le\alpha,\quad k=1,2,\cdots,M,$$

となる。従って，こゝで通信[W]とは(2.2) を満足する $P_{X\widetilde{X}}(\cdot)$ の集合である。

（例2）

Aは$\widetilde{A}$と一致し，A上の距離を $\bar{\rho}(x,\widetilde{x})$, $M=1$, $\overline{W}=[0,a]$, $\alpha\ge 0$, $\rho_1(x,\widetilde{x})=\bar{\rho}(x,\widetilde{x})$ とすると(2.1) は $E\{\rho(X,\widetilde{X})\}\le\alpha$ となる。これは送信と受信の平均偏差が与えられた定数αをこえないということである。

（例3） （例2）で

$$\rho_1(x,\widetilde{x})=\begin{cases}0, & \bar{\rho}(x,\widetilde{x})\le b \text{ のとき}\\ 1, & \bar{\rho}(x,\widetilde{x})>b \text{ のとき}\end{cases}$$

とおけば(2.1)は $P\{\bar{\rho}(X,\widetilde{X})>b\}\le\alpha$ となる。又（例2）で $\alpha=0$ のとき，即ち

$$\rho_1(x,\tilde{x}) = \begin{cases} 0, & x \neq \tilde{x} \text{ のとき}, \\ 1, & x = \tilde{x} \text{ のとき} \end{cases}$$

であれば (2.1) は $P(X \neq \tilde{X}) = 0$ となる．これは送信と受信の完全一致を要請していることになり，従ってWは $P_{XX}(\cdot)$ という一点からなる．(P_Xを与えたら1点の意味．以後，その意味に使う)．

2.1.2. 通信[W]のエントロピーと通信の列[W^t]の情報的安定性．通信[W]のエントロピー $H(W)$を $H(\overline{W}) \equiv \inf_{P_{X\tilde{X}} \in W} I(X,\tilde{X})$ で定義する．Wが空のとき は $H(W) = \infty$ とおく。又通信の列[W^t] ($t=1,2,\cdots$)を考え，次の条件をみたす変数の組の列 $\{(X^t,\tilde{X}^t)\}$ ($t=1,2,\cdots$)が存在するとき，通信の列[W^t]は情報的に安定であるという．

1) $(X^t,\tilde{X}^t)$が再生精度の条件$\{W^t\}$をみたす。

2) $\lim_{t\to\infty} \dfrac{I(X^t,\tilde{X}^t)}{H(W^t)} = 1$ ，　（但し $\dfrac{+\infty}{+\infty} = 1$）．

3) $\{(X^t,\tilde{X}^t)\}$ は情報的に安定．

（例1）最も簡単な場合．各tでW^tが1点からなるときは，$(X^t,\tilde{X}^t)$の分布が唯一つ定まるので，通信の情報的安定は組$(X^t,\tilde{X}^t)$の列自身が情報的に安定のとき，及びそのときに限り成立する．そこで，有限個の値をとるエルゴード的な離散時間定常過程 $X^t \equiv \{X(\tau);\ 0 \leq \tau \leq t\}$ で，$\tilde{X}^t = X^t$ であれば，McMillan の定理(3.5)から，通信の情報的安定が成立することがわかる。

（例2）通信[W]，即ち可測空間$(C\times\tilde{C},\ \mathbb{B}_C\times\mathbb{B}_{\tilde{C}})$，可測函数$\rho(z,\tilde{z})$，$z\in C$，$\tilde{z}\in\tilde{C}$，制限域$\overline{W}$ が与えられたとする．更に送信Zの分布 P_Z を考える。このとき[W]のt重積ともいうべき，次のような通信を独立な要素[W]をもつ通信という．

(i) 可測空間；$(A^t\times\tilde{A}^t,\ \mathbb{B}_A^t\times\mathbb{B}_{\tilde{A}}^t) \equiv (C\times\cdots\times C,\ \tilde{C}\times\cdots\times\tilde{C},\ \mathbb{B}_C\times\cdots\times\mathbb{B}_C,\ \mathbb{B}_{\tilde{C}}\times\cdots\times\mathbb{B}_{\tilde{C}})$．

(ii) 可測函数の組；$\rho_k(x,\tilde{x}) \equiv \rho(z_k,\tilde{z}_k)$　$k=1,2,\cdots,t$．　こゝに $x=(z_1,\cdots,z_t)\in A^t$，　$\tilde{x}=(\tilde{z}_1,\cdots,\tilde{z}_t)\in\tilde{A}^t$．

(iii) 制限域：$\overline{W}^t \equiv \overline{W}\times\cdots\times\overline{W}$

(iv) 送信$X=(Z_1,\cdots,Z_t)$の分布はt-直積 $P_X(\cdot)\equiv P_Z\times\cdots\times P_Z(\cdot)$の形に限る．

独立な要素をもつ通信の列$\{[W^t]\}$については，$H(W^1)$が有限であれば，

$H(W^t) = tH(W^1)$ が成り立ち，従って通信の列[W^t] ($t=1,2,\cdots$)は情報的に安定である．通信の多くの場合が，この場合に帰着できるので重要で

2.1.3. 伝達機構 $\{Q, V\}$

i) $Q(y, \tilde{E})$ は可測空間 $(\mathbb{B}, \mathbb{B}_B)$ から可測空間 $(\tilde{\mathbb{B}}, \mathbb{B}_{\tilde{B}})$ 上への遷移確率.

ii) V は直積空間 $(B \times \tilde{B},\ \mathbb{B}_B \times \mathbb{B}_{\tilde{B}})$ 上の確率分布のある集合.

のとき $\{Q, V\}$ を伝達機構という。又，$B \times \tilde{B}$ の値をとる確率変数の組 $(Y, \tilde{Y})$ の分布が V に属し，${}^{\forall}\tilde{E} \in \mathbb{B}_{\tilde{B}}$ に対し，

$$P\{\tilde{Y} \in \tilde{E}/Y\} = Q(Y, \tilde{E}) \qquad a.e(Y)$$

のとき $(Y, \tilde{Y})$ は $\{Q, V\}$ によって結ばれている という.

なお，V の与え方は，通信の場合の W と同様，$B \times \tilde{B}$ 上に N 個の実可測函数 $\pi_i(y, \tilde{y})$ $(i = 1, 2, \cdots, N)$ を与え，更に N 次元集合 $\overline{V}$ をあたえて，

(2.3) $$(E\{\pi(Y, \tilde{Y})\},\ \cdots,\ E\{\pi_N(Y, \tilde{Y})\}) \in \overline{V}$$

により，V を制限する．これは $\boxed{0}$ でいうところの $\boxed{\text{送信器}}$ — $\boxed{\text{通話路}}$ — $\boxed{\text{受信器}}$ の部分の役割をはたすものに他ならない。

2.1.4. 伝達機構 $\{Q, V\}$ の通過能力 $C(Q, V)$ と伝達機構の列の情報的安定性。

$C(Q, V) \equiv \sup I(Y, \tilde{Y})$ を通過能力という。ここで $\sup$ は $\{Q, V\}$ で結ばれた確率変数 $(Y, \tilde{Y})$ のすべての組についてとる。

又，次ぎの条件をみたす変数の組の列 $\{(Y^t, \tilde{Y}^t)\}$ $(t = 1, 2, \cdots)$ が存在するとき，伝達機構の列 $\{Q^t, V^t\}$ $(t = 1, 2, \cdots)$ は情報的に安定であるという.

1) $(Y^t, \tilde{Y}^t)$ は伝達機構 $\{Q^t, V^t\}$ で結ばれている.

2) $$\lim_{t \to \infty} \frac{I(Y^t, \tilde{Y}^t)}{C(Q^t, V^t)} = 1$$

3) $\{(Y^t, \tilde{Y}^t)\}$, $(t = 1, 2, \cdots)$ は情報的に安定。

(例 1)

B^t は n^t 個の点からなる有限集合，$\tilde{B}^t = B^t$, $Q^t(\cdot, \cdot)$ は ${}^{\forall}y^t \in B^t$ に対し $Q^t(y^t, y^t) = 1$ とする．この場合 $\{Q^t, V^t\}$ で結ばれているエルゴード的離散時間定常過程の segment[1] $Y^t, \tilde{Y}^t$ の組 $(Y^t, \tilde{Y}^t)$ に対し $P\{Y^t = \tilde{Y}^t\} = 1$. 制限条件 V^t は $B^t \times \tilde{B}^t$ 上のすべての分布からなるとする．(このような集合 V^t は π_i^t を使っては $\overline{V}^t$ として，全空間をえらべばよ

---

[1] 定常過程 $X(t)$ に対し，$X^t \equiv \{X(\tau),\ 0 < \tau \leq t\}$ を $X(t)$ の segment と呼ぶ。

い）。このような伝達機構の列は情報的に安定である．実際，ここで，通過能力 $C(Q^t, V^t)$ は B^t の値をとる確率変数のエントロピーの上限に等しい． $\boxed{1}$ でもわかるように，この確率変数の列は情報的に安定であり，従って，$\{Q^t, V^t\}$ の情報的安定が出る．

（例２） 記憶のない伝達機構．

独立な要素をもつ通信の場合と同様の，t 重積の可測空間を $state\ space$ とする．

$$(B^t, \mathbb{B}_{B^t}) \equiv (C\times\cdots\times C,\ \mathbb{B}_C\times\cdots\times\mathbb{B}_C),\ (\widetilde{B}^t, \mathbb{B}_{\widetilde{B}^t}) \equiv (\widetilde{C}\times\cdots\times\widetilde{C},\ \mathbb{B}_{\widetilde{C}}\times\cdots\times\mathbb{B}_{\widetilde{C}})$$

$$Y^t \equiv (Z_1^t, \cdots, Z_t^t), \qquad \widetilde{Y}^t \equiv (\widetilde{Z}_1^t, \cdots, \widetilde{Z}_t^t)$$

$$\widetilde{E}^t \equiv \widetilde{E}_1\times\cdots\times\widetilde{E}_t, \quad \widetilde{E}_k \in \mathbb{B}_{\widetilde{C}} \quad (k=1,\cdots,t),$$

$(C, \mathbb{B}_C)$ から $(\widetilde{C}, \mathbb{B}_{\widetilde{C}})$ への遷移確率を $Q_Z^t(z, \widetilde{E})$ とするとき

$$Q^t(y^t, \widetilde{E}^t) \equiv Q_Z^t(z_1, \widetilde{E}_1)\cdot Q_Z^t(z_2, \widetilde{E}_2)\cdots Q_Z^t(z_t, \widetilde{E}_t)$$

と定義すれば，$Q^t(y^t, \widetilde{E}^t)$ は任意の $\widetilde{E}^t\in\mathbb{B}_{\widetilde{B}^t}$ まで拡張される。制限条件の函数 $\pi_k^t(y^t, \widetilde{y}^t)$ は

$$\pi_k^t(y^t, \widetilde{y}^t) = \pi^t(z_k^t, \widetilde{z}_k^t)$$

と定義し，t 次元集合 $\overline{V}^t$ は t 個の見本集合 $\overline{V}_C$ の直積とする．この方法であたえられる伝達機構を記憶のない伝達機構と呼ぶ．このとき $\{Q^t, V^t\}$ は $C(Q^1, V^1)$ が有限であれば，任意の t に対して，$C(Q^t, V^t) = tC(Q^1, V^1)$ であり，従って情報的安定となることがわかる．

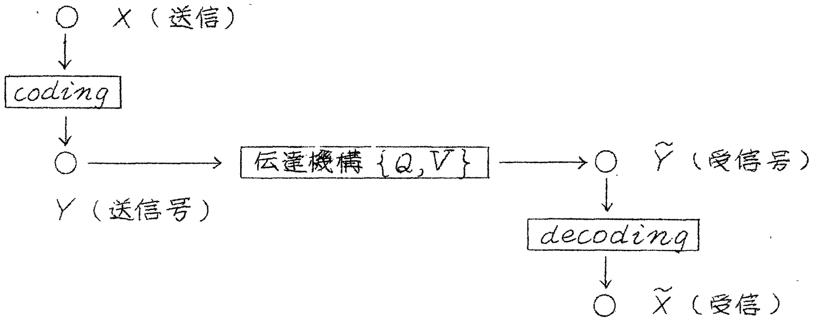
2.1.5. 通信伝達 $\{W, Q, V\}$

通信 $[W]$ と伝達機構 $\{Q, V\}$ が与えられ，次のような４つの確率変数 X, Y, $\widetilde{Y}$, $\widetilde{X}$ が存在すれば，通信 $[W]$ は伝達機構 $\{Q, V\}$ によって伝達され得る という．

1) X, Y, $\widetilde{Y}$, $\widetilde{X}$ からなる列は $Markov\ chain$ に従う。
2) $(X, \widetilde{X})$ は再生精度の条件 $\{W\}$ をみたす。
3) $(Y, \widetilde{Y})$ は $\{Q, V\}$ によって結ばれている．

以上の定義を直観的に説明すれば，条件つき確率分布 $P\{Y\in E / X=x\}$ は通信文 x を符号化（$coding$）して送るときの送信号の確率分布であり，$P\{\widetilde{X}\in\widetilde{E}/\widetilde{Y}=\widetilde{y}\}$ は信号 $\widetilde{y}$ を受けたとき，逆符号化（$decoding$ 即ち符号を解くこと）して得られる受信の確率分布である。符号化や逆符号化の操作は一般的な取扱いにおいては当然 $deterministic$ な操作ではなく，$random$ な可能性を含んでいる．列 X, Y, $\widetilde{Y}$, $\widetilde{X}$ が $Markov\ chain$ という要請は，次の事を意味する．即

ち　送信号を固定したとき，伝達機構の受信号の確率分布は符号化して　その信号になる通信文（送信）に関係しない．更に受信号を固定したとき，逆符号化によってその信号から生成される通信文（受信）の確率分布は送信号にも，符号化される通信文（送信）にも関係しない。



2.1.6.　「通信$[W]$が $\{Q,V\}$ によって，確率εの事象までの精度で伝達される」ということ等の意味を説明しておく。その前に，n次元ユークリッド空間の部分集合Uとその中の点 $\bar{x}=(x_1,\cdots,x_n)$，　$\bar{x}'=(x_1',\cdots,x_n')$ との間の距離を $r(\bar{x},\bar{x}')$ として

$$r(\bar{x},\bar{x}')=\max_{i=1,2,\cdots,n}|x_i-x_i'|$$

をとる．

$$[U]_\varepsilon \equiv \begin{cases} \{\bar{x}\in R^n;\ \exists\bar{\bar{x}}\in U:\ r(\bar{x},\bar{\bar{x}})\leq\varepsilon\}, & \varepsilon\geq 0 \text{ のとき} \\ [[U^c]_{-\varepsilon}]^c, & \varepsilon<0 \text{ のとき} \end{cases}$$

と定義しておく。

通信$[W]$の定義で制限域 $\overline{W}$ の代りに $[\overline{W}]_\varepsilon$ を使って与えられる通信を通信$[W\varepsilon]$　と名づける．同様に，集合 $\overline{V}$ を $[\overline{V}]_\varepsilon$にかえた伝達機構を伝達機構$\{Q,V_\varepsilon\}$ と名づける。　そして，通信$[W]$が $\{Q,V\}$ によって確率εの事象までの精度で伝達されうる　というのは，4つの確率変数 $X,Y,\tilde{Y},\tilde{X}$ と空間$(\hat{A},\mathbb{B}_{\hat{A}})$ の値をとる5番目の確率変数$\tilde{X}'$があって

1）4つの確率変数の列X，Y，$\tilde{Y}$，$\tilde{X}$は Markov chain をなす．

2）変数 $X,\tilde{X}'$ の組は再生精度の条件 $\{W\}$ をみたす．

3）変数 $(Y,\tilde{Y})$ は $\{Q,V\}$ によって結ばれている．

4）　$P\{\tilde{X}\neq\tilde{X}'\}\leq\varepsilon$

となることである．条件 1)～4)の直観的な意味は，受信を確率εで変えて送信

と受信の組が再生精度の条件をみたすようにできることである。

2.2 基本定理（Shannon の定理）の内容の説明

2.2.1. 定理 1（Dobrushin） 次の如き通信 $[W^t]$ と伝達機構 $\{Q^t, V^t\}$ の列が与えられたとする。

I) $\lim_{t\to\infty} H(W^t)=\infty$

II) $\lim_{t\to\infty} \dfrac{H(W^t)}{C(Q^t, V^t)} < 1$

III) 通信の定義に入って来る函数 $\rho_i^t(x^t, \tilde{x}^t)$ の数 M^t と 伝達機構の定義に入って来る函数 $\pi_i^t(y^t, \tilde{y}^t)$ の数 N^t が任意の $a>0$ に対し

$$M^t = o\left(2^{aH(W^t)}\right), \quad N^t = o\left(2^{aC(Q^t, V^t)}\right) \tag{2.4}$$

IV) 伝達機構 $\{Q^t, V^t\}$ で結ばれた確率変数の組 $(Y^t, \tilde{Y}^t)$ の列が存在し，情報的に安定で，かつ

(i) $\lim_{t\to\infty} \dfrac{I(Y^t, \tilde{Y}^t)}{C(Q^t, V^t)} = 1$

（即ち伝達機構の列は情報的に安定）．

(ii) 或る $\bar{b}>0$，と

$$\bar{C}^t \equiv \max_{k=1,2,\cdots,N^t} E\{|\pi_k^t(Y^t, \tilde{Y}^t) - E\pi_k^t(Y^t, \tilde{Y}^t)|^{1+\bar{b}}\}$$

に対し，任意の $a>0$ のとき

$\bar{C}^t = o\left(2^{aC(Q^t, V^t)}\right)$．

V) 再生精度の条件 $\{W^t\}$ をみたす確率変数の組 $(X^t, \tilde{X}^t)$ の列が存在し，情報的に安定，かつ

(i) $\lim_{t\to\infty} \dfrac{I(X^t, \tilde{X}^t)}{H(W^t)} = 1$，

（即ち通信の列 $[W^t]$ は情報的に安定）

(ii) 或る $b>0$ と

$$C^t \equiv \max_{k=1,2,\cdots,M^t} E\{|\rho_k^t(X_1^t, X^t) - E\rho_k^t(X^t, \tilde{X}^t)|^{1+b}\}$$

に対して任意の $a>0$ のとき

$$C^t = o(2^{aH(W^t)}).$$

以上 (I) ～ (V) の条件のもとで，任意の $\varepsilon > 0$ に対し，充分大きい T があって，$t \geqq T$ に対し，通信 $[W_\varepsilon^t]$ は伝達機構 $\{Q^t, V_\varepsilon^t\}$ によって確率 ε の事象までの精度で伝達されうる。

これが Shannon の定理を新しく定式化した Dobrushin の定理の 1 つである。この定理の内容の説明に入る前に，Shannon 以来，McMillan，Feinstein を経て Khintchin により定式化された Shannon の定理を紹介しよう。

2.2.2. Shannon の第 1 定理 (Khinchin)

1) 定常で予報せず，ergodic capacity C と有限の記憶 m をもつ通話路 $[A, P_{\tilde{Y}|Y}, B]$ と，[2)]

2) エントロピー $H < C$ のエルゴード的な情報源 $(\Omega, \mathbb{B}, P)$ とが与えられている。

その時，$\varepsilon > 0$ に対して n が充分大きければ，情報源 $(\Omega, \mathbb{B}, P)$ から出る通信文をアルファベット A で符号化し，通信文 ω の文字の各 n 項 chain α_i がアルファベット A の文字の $(n+m)$ 項 chain u_i にうつり (符号化され)，与えられた通話路によって chain u_i を送ったとき，通話路の受信側で得られた chain から $1-\varepsilon$ より大きい確率で送られた chain α_i を正しく求めることが出来る。

Shannon の第 2 定理　第 1 定理と同条件のもとで transmission rate (伝送速度) ($\boxed{0}$ 参照) が H にいくらでも近くなるような符号化をえらぶことができる。

先ず，Shannon の定理の内容の検討から始めよう。

Shannon の第 1 定理の条件　1) の定常性と，2) のエルゴード性　は，エントロピーに関する McMillan の定理 ($\boxed{3}$) を保証し，従って，実質的に情報をもつ chain の個数 L が大体 $2^{n(H-\varepsilon)}$ 個の order で，その全体の測度が $1-\varepsilon$ より大きいこと (定理の終結の後半) を保証する。又

---

2) $Y = (\cdots Y_{-1}, Y_0, Y_1, \cdots)$，$\tilde{Y} = (\cdots, \tilde{Y}_{-1}, \tilde{Y}_0, \tilde{Y}_1, \cdots)$，で $P_{\tilde{Y}|Y}(\tilde{Y}_0 \in E|y) = P(\tilde{Y}_0 \in E | Y_{-m} = y_{-m}, \cdots Y_{-1} = y_{-1})$ が成り立つとき，予報せず，有限の記憶 m をもつ通話路という。

条件 1) は Feinstein の lemma "定常で予報せず，且つ有限な記憶をもつ通話路が与えられたとき，$\forall \varepsilon > 0$ に対し，nを充分大きくすれば，$m+n$ 項 chain u_i の ε-判別可能な組$\{u_i\}$[3] $(1 \leq i \leq N)$ で，その個数Kが

$$K > 2^{n(C-\varepsilon)}$$

なるものが存在する．但しCは通話路の ergodic capacity である" の成立条件である。

かくて McMillan の定理と Feinstein の lemma は $H<C$ の時 $L<K$を成立させ，Shannon の定理の結論を成立させる。

ひるがえって，本章に於ける定理 1（Dobrushin）と比較した場合，定理の仮定条件に於て，Khintchin の定式化と Dobrushin のそれとは，一見非常に異なるように見えるが，本質的には同じ事実の数学的 model であり，従って同じ存在の妥当性をもつものなのである。つまり，一見の相異は次のようなことから来ている。Khinchin の定式化では，エルゴード的な情報源の概念から，HとかCとかいう量が命題の叙述の中心概念となっている。ということは，時間の無限区間の間に於ける情報伝達が基調となっている。一方，Dobrushin はいわば，これと反対の立場，即ち，大ざっぱにいえば，時間の大きいが，しかし或る有限な区間の segment の間に於ける実際の情報伝達，及びその漸近性を基調とする（情報的安定！）。それと同時に，Dobrushin の定式化は定理の内容においても Khintchin のそれよりも広い。例えば $\boxed{1}$ にも見る如く，McMillan の定理の成立は情報的安定を包含し，又，Khintchin は離散時間の場合であるが，Dobrushin の結果は連続時間の場合にも適用し得る。

そこで，Dobrushin の定理の内容の検討に入ろう。定理の条件 I)，II) の意味は今まで述べて来た議論より推察される．条件 III) は函数 ρ_k，π_k の数を上から制限している。勿論（応用の大部分においてはそうであるが），M^t，N^t がtに無関係（しばしば単に1に等しい．$\boxed{2.1}$（例２），（例３））ならば，明らかにそれは成り立つ．しかし，M^t，N^t が無限に増大する重要な場合を示すことも容易である．例えば，2.1.2.（例２）の $t(=M^t)$ 個の独立な要素をもつ通

---

3) $\{u_i\}$ に対して，受信側の chain の集合の組 $\{V_i\}$ が存在して，各 V_i は，いくつかの chain の和集合で i) $i \neq k$ ならば，V_i と V_k は共通の chain を含まず，ii) $P_{\tilde{Y}|Y}(V_i | u_i) > 1-\varepsilon$ $(1 \leq i \leq N)$ のとき，$\{u_i\}$ を ε-判別可能な組という。

信がある．そこでは，$H(W^t)=tH(W^1)$ であるから（$H(W^t)\neq 0$ であれば）定理の条件は成立する。又，t 個からなる　もっと一般の場合（必ずしも独立な要素ではない）は $H(W^t)$ の order が t の order だという仮定で充分であることがわかる。N^t についても類似のことがいえる．特に記憶のない伝達機構（2.1.4.（例2））に対しては，独立な要素をもった場合の M^t と全く類似している。総じて条件 Ⅲ）は大きい制限条件にはなっていないといってよい程，具体例に於いては　満足されている。

次ぎに条件 Ⅳ）。もし $\pi_k^t(y,\tilde{y})$ が k と t に関し一様有界（具体例では，しばしばそうである）なら Ⅳ）は伝達機構の列の情報的安定の条件と同値である。条件 Ⅴ）についても同様。

次ぎに定理の結論について考えてみる。先に述べた Shannon の才1定理（Khintchin）と本定理の結論を比較して容易に分るように，確かに前者の仮定は後者の条件を満足し，従って，本定理を適用し得るが，結論は後者の方が，前者程細かくはない。これは Khintchin が process の有限な segment の coding（有限項 chain!）の問題のみを取り扱っておるのに反し，Dobrushin は時間に関して無限なすべての process の coding の問題をも含めて処理しようとしていることに起因している．従って彼は本定理の条件 Ⅰ）〜Ⅴ）の他に制限函数のモーメントにもっと強い条件を附加することにより，“充分大きい T があって，すべての $t\geq T$ に対し，通信 $[W^t]$ が伝達機構 $\{Q^t,V^t\}$ によって確率 ε の事象までの精度で伝達され得る” ことも証明している。これ等いずれの場合にも，基本定理の証明には Feinstein の lemma の idea を必要とするのである。

2.2.3.　伝達機構に関する Feinstein の Lemma

［条件］　次のような伝達機構の列 $\{Q^t,V^t\}$ が与えられたとする。

(i) 十分大きいすべての t に対し $C(Q^t,V^t)<\infty$，かつ $\lim_{t\to\infty} C^t(Q^t,V^t)=\infty$.

(ii) 条件 (2.4) と定理1の Ⅳ）が満たされている。

(iii) $\forall\varepsilon>0$ に対し $L_\varepsilon^t \equiv [2^{(1-\varepsilon)C(Q^t,V^t)}]$（[　] は Gauss 記号）とするとき，次の条件をみたす L_ε^t 個の確率の組（$P_1^t,\cdots,P_{L_\varepsilon^t}^t$）がある。

$$\sum_{i=1}^{L_\varepsilon^t} P_i^t=1,\quad \max_{i=1,\cdots,L_\varepsilon^t} P_i^t\leq\frac{2}{L_\varepsilon^t}\quad .$$

［結論］　十分大きい T がとれて，$\forall t>T$ に対し送信号の空間 $(B^t,\mathbb{B}_{B^t})$ の中に L_ε^t 個の点 $y_1^t,\cdots,y_{L_\varepsilon^t}^t$，が得られて，これ等の点 y_i^t の各々に対して，次

のような受信号空間 $(\widetilde{B}^t, \mathbb{B}_{\widetilde{B}^t})$ の部分集合 $E_i^t \in \mathbb{B}_{\widetilde{B}^t}$ $(i=1, \cdots, L_\varepsilon^t)$ を対応させることができる：

1) $E_1^t, \cdots, E_{L_\varepsilon^t}^t$ は固定した t に対し，互に素．

2) $\displaystyle\sum_{i=1}^{L_\varepsilon^t} P_i^t Q^t(y_i^t, E_i^t) \geq 1-\varepsilon$,　$\forall t \geq T$.

3) $\forall t \geq T$ に対して N^t 次元ベクトル

$$\Big(\sum_{i=1}^{L_\varepsilon^t} P_i^t \int_{\widetilde{B}^t} \pi_1^t(y_i^t, \tilde{y}^t) Q^t(y_i^t, d\tilde{y}^t),$$

$$\cdots, \sum_{i=1}^{L_\varepsilon^t} P_i^t \int_{\widetilde{B}^t} \pi_{N^t}^t(y_i^t, \tilde{y}^t) Q^t(y_i^t, d\tilde{y}^t) \Big) \in [\overline{V}^t]_\varepsilon .$$

さて，この一般化された *Feinstein* の *lemma* の内容を検討しよう．2.2.2. で述べたように，基本定理の *Khintchin* による定式化は，*McMillan* の定理と *Feinstein* の *lemma* の併用によって *Shannon* の定理を証明したものである。*Khintchin* の場合と全く同様であって，［条件］の (i)，(ii) を置くことにより *McMillan* の定理に類似のこと（情報的安定性）の成立を予見し，［結論］で，送信空間の点 $y_1^t, \cdots, y_{L_\varepsilon^t}^t$ とそれに対する受信空間の組 $\{E_1^t, \cdots, E_{L_\varepsilon^t}^t\}$ の存在を主張しているのは，*Khintchin* の場合でいう判別可能な組の存在に他ならない。ただ，2.2.2. にも述べたように，*Khintchin* と *Dobrushin* の定式化の相異（*Dobrushin* は無限区間での符号化を含ませようとしている）から［条件］の (iii) と［結論］3）の如き述べ方が出て来ている。しかし，*lemma* の結論と基本定理の結論の間には いささかの距離があるようである．基本定理の結論では通信か伝達機構 $\{Q^t, V_\varepsilon^t\}$ によって送られるをことを要請している．即ち送信 Y，受信 $\tilde{Y}$ としたとき，(2.3) より $(E\{\pi_1^t(Y, \tilde{Y})\}, \cdots, E\{\pi_{N^t}^t(Y, \tilde{Y})\}) \in [\overline{V}^t]_\varepsilon$ でなければならない。これは，明らかに *lemma* の結論 3）とは異なっているが，前者を後者より導き出す *idea* はこうである．初めに L_ε^t 個の点 $Y_1^t, \cdots, Y_{L_\varepsilon^t}^t$ を互に独立に，かつ Y^t と同分布をもつようにとれば，$\{\sum_{i=1}^{L_\varepsilon^t} P_i^t \int_{\widetilde{B}^t} \pi_k^t(y_i^t, \tilde{y}^t) Q^t(y_i^t, d\tilde{y}^t)\}$ $(k=1, 2, \cdots, N^t)$ の各々は，独立確率変数 $Y_1^t, \cdots, Y_{L_\varepsilon^t}^t$ に関する大数の法則が適用できて（大数の法則の精密化），t を充分大きくすれば，$E\{\int_{\widetilde{B}^t} \pi_k^t(Y^t, \tilde{y}^t) Q^t(Y^t, d\tilde{y}^t)\} = E\{\pi_k(Y^t, \tilde{Y}^t)\}$，$(k=1, 2, \cdots, N^t)$ に収束するというわけである．従って，ここで使用される精密化された大数の法則を述べておくことは，一般化された *lemma* の［条件］を理解する上からも無駄で

はあるまい．

精密化された大数の法則[30]：　$Y_1, \cdots, Y_L$ を独立な同分布をもつ確率変数とし，ある $\bar{c} < \infty$ と $\bar{b} > 0$ に対し $E|Y_i - E(Y_i)|^{1+\bar{b}} \leq \bar{c}$ とする．更に $\sum_{i=1}^{L} P_i = 1$ かつ $\max_{i=1,\cdots,L} P_i \leq \frac{2}{L}$ なる L 個の確率の組 $P_1, \cdots, P_L$，が与えられたとする。このとき 任意の $\varepsilon > 0$ に対して十分大きな N がとれて，$L > N \cdot (\bar{c})^{\frac{1}{\bar{b}}}$，かつ

$$P\{|\sum_{i=1}^{L} P_i Y_i - E(Y_1)| > \varepsilon\} \leq \frac{D(\bar{c}+1)}{L^{\min(\bar{b},1)}}$$

となる。但し定数 D と N は $\bar{b}$ と ε にだけ関係し，L，$\bar{c}$ や Y_i の分布には関係しない。

2.2.4.　通信に関する Feinstein-Dobrushin の lemma．

2.2.3.で述べた Feinstein の lemma の結論は，雑にいえば“通過能力 $C(Q^t, V^t)$ をもった伝達機構に関し，（非常に小さい間違った受信の確率で）$L_\eta^t \doteqdot 2^{C(Q^t, V^t)}$ 個の異なった信号を送ることができる”ということであるが，ここで述べる通信に関する lemma の内容は，“エントロピー $H(W^t)$ をもった通信を 2^H 個の異なった通信で置きかえることができる”ということである。

通信に関する Feinstein-Dobrushin の Lemma．　情報的に安定な通信の列 $\{[W^t]\}$ が与えられ，$H(W^t) < \infty$，かつ　$H(W^t) \to \infty$（$t \to \infty$）とする。定理 1 の条件 V）がみたされ，更に制限函数の個数 M^t は定理 1 の条件 III）をみたしていることを仮定する（従って情報密度 $i(x^t, \tilde{x}^t) (\equiv i_{X^t, \tilde{X}^t}(x^t, \tilde{x}^t)$ $\boxed{0.3}$ 参照）の存在を仮定してよい）．$\varepsilon > 0$ に対して

$$(2.6) \qquad F_\varepsilon^t \equiv \{(x^t, \tilde{x}^t); \left|\frac{i(x^t, \tilde{x}^t)}{H(W^t)} - 1\right| \leq \varepsilon/4\}$$

$$r_\varepsilon^t(x) \equiv P\{(X^t, \tilde{X}^t) \notin F_\varepsilon^t \mid X^t = x_t\}$$

と定義する．

このとき任意の $\varepsilon > 0$ に対し十分大きい T があって，$\forall t \geq T$ に対して $K_\varepsilon^t \equiv [2^{(1+\varepsilon)H(W^t)}]$ 個の点 $\tilde{x}_1^t, \cdots, \tilde{x}_{K_\varepsilon}^t$ を受信空間 $\tilde{A}^t$ の上にとり，この点の各々に B_{A^t}-可測な函数 $q_i^t(x)$ が対応して，次の条件をみたすようにできる：

1）すべての $\tilde{x}_i^t \in \tilde{A}^t$（$i = 1, \cdots, K_\varepsilon^t$）と $t \geq T$ に対して

$$0 \leq q_i^t(x) \leq 1 \ , \quad \text{かつ} \quad Q^t(x) \equiv \sum_{i=1}^{K_\varepsilon^t} q_i^t(x) + r_\varepsilon^t(x) \leq 1 \ .$$

2）ある固定した $\varkappa > 0$ とすべての $t \geq T$ に対して

$$Q^t \equiv \int_{A^t} Q^t(x)\, P_{X^t}(dx) \geq 1-2^{-\varkappa H(W^t)} \quad .$$

3) $\forall t \geq T$ に対し,

$$(S_1^t, \cdots, S_{M^t}^t) \in [\overline{W}^t]_\varepsilon \quad .$$

但し $S_k^t \equiv \sum_{i=1}^{K_\varepsilon^t} \int_{A^t} \rho_k^t(x, \tilde{x}_i)\, q_i^t(x)\, P_{X^t}(dx) + \iint_{A^t \times \tilde{A}^t - F_\varepsilon^t} \rho_k^t(x, \tilde{x})\, P_{X^t, \tilde{X}^t}(dx, d\tilde{x})$

$$+ [1-Q^t]\, E\{\rho_k^t(X^t, \hat{X}^t)\} \quad .$$

証明は2.2.3.の伝達に関する lemma と殆んど同様の idea と方法により組み立てられる。

2.3 基本定理の証明

紙数の制限上完全な証明をなし得ないので，基本定理の成立の経緯を理解する助けとなることを目標として，道筋の概略のみを述べる。

基本定理の内容をこまかい条件を抜きにして，極く簡単にいえば，「情報的に安定な通信の列$[W^t]$と伝達機構の列$\{Q^t, V^t\}$が与えられているとき，$[W_\varepsilon^t]$が$\{Q^t, V_\varepsilon^t\}$によって，確率εの精度で伝達され得る。」ということである。従って，定理の証明は，そういうことを可能にする伝達方法の構成である。数学的に換言すれば，それぞれ$(A^t, \mathbb{B}_{A^t})$,$(B^t, \mathbb{B}_{B^t})$,$(\tilde{B}^t, \mathbb{B}_{\tilde{B}^t})$,$(\tilde{A}^t, \mathbb{B}_{\tilde{A}^t})$の値をとり，Markov chain をなす確率変数の組$X^t, Y^t, \tilde{Y}^t, \tilde{X}^t$と5番目の確率変数$\tilde{X}'^t$を2.1.6.の定義を満足するように構成すればよい，ということである。所で Markov chain の構成は，初期分布と遷移確率をあたえればよいので，初期分布$P_{X^t}(\cdot)$としては，あたえられた送信の分布$P_{X^t}(\cdot)$と同分布としておいて，遷移確率を構成し，最後に与件を満足する$\tilde{X}'^t$を構成する。そのためにまず，遷移確率を与えるのに重要な役割を果たす$\mathbb{B}_{A^t}$-可測函数の族$\{S_{ij}^t(x)\}$を定義する。

基本定理の条件から2.2.4.の Feinstein-Dobrushin の lemma (2.2.4) が適用できるので，そこでの諸量，記号をそのまま使うことにする。

1° $L_\delta^t \equiv [2^{(1-\delta)C(Q^t, V^t)}]$, $\delta \leq \frac{\varepsilon}{3}$, として，$\{1, 2, \cdots, L_\delta^t\}$を$(K_\delta^t+3)$個の部分集合に分ける。

$$R_i^t \equiv \left[\frac{q_i^t L_\delta^t}{2}\right] + 1 \ , \qquad (i=-1, 0, 1, \cdots, K_\delta^t) \ ,$$

$$R_{-2}^t \equiv L_\delta^t - \sum_{i=-1}^{K_\delta^t} R_i^t$$

と定義すれば $\sum_{i=-2}^{K_\delta^t} R_i^t = L_\delta^t$. 但し

$$\bar{q}_i^t \equiv \int_{A^t} q_i^t(x)\, P_{X^t}(dx) \qquad (i=1,\cdots,K_\delta^t),$$

$$\bar{q}_0^t \equiv \int_{A^t} r_\delta^t(x)\, P_{X^t}(dx), \quad \bar{q}_{-1}^t \equiv 1-Q^t .$$

2° $S_{ij}^t(x) \equiv \frac{1}{R_i^t} q_i^t(x), \quad (i=1,\cdots,K_\delta^t \ ; \ j=1,\cdots,R_i^t)$

$$S_{0j}^t(x) \equiv \frac{1}{R_0^t} r_\delta^t(x), \quad (j=1,\cdots,R_0^t),$$

$$S_{-1j}^t(x) \equiv \frac{1}{R_{-1}^t}[1-Q^t(x)] \qquad (j=1,\cdots,R_{-1}^t)$$

$$S_{-2j}^t(x) \equiv 0 \qquad (j=1,\cdots,R_{-2}^t)$$

と定義すれば $\sum_{j=1}^{R_i^t} \int_{A^t} S_{ij}^t(x)\, P_{X^t}(dx) = \bar{q}_i^t \quad (i=-1,0,\cdots,K_\delta^t) = 0 \ (i=-2)$ となる. そこで, $\int_{A^t} S_{ij}^t(x)\, P_{X^t}(dx) \equiv p_{ij}^t$ とおけば, $p_{ij}^t = \frac{1}{R_i^t}\bar{q}_i^t \quad (i=-1,\cdots,K_\delta^t), = 0 \quad (i=-2)$, かつ $\sum_{i=-2}^{K_\delta^t}\sum_{j=1}^{R_i^t} p_{ij}^t = \sum_{i=-2}^{K_\delta^t} \bar{q}_i^t = 1$ となり, 又 $\{p_{ij}^t\}$ の各々は $\frac{2}{L_\delta^t}$ を越えないので, 2.2.3.の

Feinstein の *lemma* の確率の組の条件を満足している.

従って, 基本定理の条件 Ⅲ), Ⅳ) とあわせて, *Feinstein* の *lemma* が適用できて, t が十分大きければ,

$$(2.7) \qquad \Big\{ \sum_{i=-2}^{K_\delta^t}\sum_{j=1}^{R_i^t} p_{ij}^t \int_{\tilde{B}^t} \pi_1^t(y_{ij}^t, \tilde{y}^t)\, Q(y_{ij}^t, d\tilde{y}^t),$$

$$\cdots, \sum_{i=-2}^{K_\delta^t}\sum_{j=1}^{R_i^t} p_{ij}^t \int_{\tilde{B}^t} \pi_{N^t}(y_{ij}^t, \tilde{y}^t)\, Q(y_{ij}^t, d\tilde{y}^t) \Big\} \in [\bar{V}^t]_\delta$$

かつ

$$(2.8) \qquad \sum_{i=-2}^{K_\delta^t}\sum_{j=1}^{R_i^t} p_{ij}^t\, Q(y_{ij}^t, \tilde{E}_{ij}^t) \geq 1-\delta$$

を満足するような B_t 上の点の組 $\{y_{ij}^t\}$ とそれに対応する $\tilde{B}_t$ の素な部分集合の組 $\{\tilde{E}_{ij}^t\}$ が得られる。これらの $\{y_{ij}^t\}$ 及び $\{\tilde{E}_{ij}^t\}$ を使って, 遷移確率を構成してゆく.

まず,

(2.9) $\qquad \widetilde{P}_{Y^t|X^t}(x,E) \equiv \sum_{y_{ij}^t \in E} S_{ij}^t(x),\quad x\in A^t,\quad E\in \mathbb{B}_{B^t}$ と定義する.[4)]

この函数のxに関する可測性は $S_{ij}^t(x)$ の可測性より,Eに関して確率測度になっていることは, $\sum_{i=-2}^{K_\delta^t}\sum_{j=1}^{R_i^t} S_{ij}^t(x)=\sum_{i=1}^{K_\delta^t} q_i(x)+r_\delta^t(x)+[1-Q^t(x)]=1$ より出る。

次ぎに, $P_{\widetilde{Y}^t|Y^t}(y,\widetilde{E})$ は, 与えられた伝達機構 $\{Q^t, V^t\}$ により $P_{\widetilde{Y}^t|Y^t}(y,\widetilde{E}) \equiv Q^t(y,\widetilde{E})$ と定義しておく. さらに $P_{\widetilde{X}^t|\widetilde{Y}^t}(\widetilde{y},\widetilde{E})$ を定義する. まず, 勝手な方法で1つの点 $\widetilde{x}_+^t\in\widetilde{A}^t$ をえらぶ. 通信に関する Feinstein-Dobrushin の lemma の $\widetilde{A}^t$ 上の点 $\widetilde{x}_i^t$ と A^t 上の函数 $q_i^t(x)$ との対応関係を補足し, $S_{0j}^t(x)$, $S_{-1j}^t(x)$, $S_{-2j}^t(x)$ には, $\widetilde{A}^t$ 上の $(\widetilde{x}_1, \widetilde{x}_2, \cdots, \widetilde{x}_{K_\delta^t}$ 以外の) 1点 $\widetilde{x}_+^t$ を対応させることとする. その時,

$$P_{\widetilde{X}^t|\widetilde{Y}^t}(\widetilde{y},\widetilde{E}) \equiv \begin{cases} \widetilde{y}\in \widetilde{E}_{i_0j}^t & \text{かつ}\ \widetilde{E}\ \text{が}\ \widetilde{x}_{i_0}^t\ \text{を含まねば}\ 0 \\ \widetilde{y}\in \widetilde{E}_{i_0j}^t & \text{かつ}\ \widetilde{E}\ \text{が}\ \widetilde{x}_{i_0}^t\ \text{を含めば}\ 1 \\ \widetilde{y}\notin \bigcup_{i,j} \widetilde{E}_{ij}^t & \text{かつ}\ \widetilde{E}\ \text{が}\ \widetilde{x}_+^t\ \text{を含まねば}\ 0 \\ \widetilde{y}\notin \bigcup_{i,j} \widetilde{E}_{ij}^t & \text{かつ}\ \widetilde{E}\ \text{が}\ \widetilde{x}_+^t\ \text{を含めば}\ 1 \end{cases} \quad \begin{pmatrix} i_0=-2,-1,\cdots,K_\delta^t \\ j=1,\cdots,R_i^t \end{pmatrix}$$

と定義する(遷移確率になることは明らか)。初期分布と3つの遷移確率が決定したので, 残るのは $\widetilde{X}'^t$ の構成と構成された $X^t, Y^t, \widetilde{Y}^t, \widetilde{X}^t, \widetilde{X}'^t$ が定理の要求を充足しているかどうかの検証である。$\widetilde{X}'^t$ の構成には, $X^t, Y^t, \widetilde{Y}^t, \widetilde{X}^t, \widetilde{X}'^t$ の同時分布を与えればよい。しかし, 前4つの同時分布はすでに与えてあるので, 条件つき確率

$$\widetilde{P}\{\widetilde{X}'^t\in\widetilde{E}/X^t=x^t, Y^t=y^t,\ \widetilde{Y}^t=\widetilde{y}^t,\ \widetilde{X}^t=\widetilde{x}^t\}$$

をあたえれば十分である。Y^t は離散的な値 $\{y_{ij}^t\}$ をとることから

(i) $\quad \widetilde{P}\{\widetilde{X}'^t\in\widetilde{E}/X^t=x^t,\ Y^t=y_{ij}^t,\ \widetilde{Y}^t=\widetilde{y}^t,\ \widetilde{X}^t=\widetilde{x}^t\}$

$$\equiv \begin{cases} \widetilde{x}_i^t\in\widetilde{E}\ \text{のとき}\ 1 \\ \widetilde{x}_i^t\notin\widetilde{E}\ \text{のとき}\ 0 \end{cases} \qquad \begin{pmatrix} i=1,2,\cdots,K_\delta^t \\ j=1,2,\cdots,R_i^t \end{pmatrix}$$

---

4) 定理で最初に存在する確率変数の確率空間を $(\Omega,\mathbb{B},P)$ とすれば X, Y, etc は, version の意味で構成するのであるから, その基本空間は $(\widetilde{\Omega},\widetilde{\mathbb{B}},\widetilde{P})$ として区別しておく.

(ii) $\widetilde{P}\{\widetilde{X}'^t \in \widetilde{E} / X^t = x^t,\ Y^t = y_{oj},\ \widetilde{Y}^t = \widetilde{y}^t,\ \widetilde{X}^t = \widetilde{x}^t\}$

$\qquad \equiv P\{\widetilde{X}^t \in \widetilde{E} / (X^t, \widetilde{X}^t) \notin F^t_\delta,\ X^t = x^t\} \qquad (j = 1, \cdots, R^t_o)$

(iii) $\widetilde{P}\{\widetilde{X}'^t \in \widetilde{E} \mid X^t = x^t,\ Y^t = y^t_{-1,j},\ \widetilde{Y}^t = \widetilde{y}^t,\ \widetilde{X}^t = \widetilde{x}^t\}$

$\qquad \equiv P\{\widetilde{X}^t \in \widetilde{E} \mid X^t = x^t\} \qquad (j = 1, \cdots, R^t_{-1})$

と定義しておく。これが 2.1.6.の条件 4) 即ち

1° $P\{\widetilde{X}^t \neq \widetilde{X}'^t\} \leq \varepsilon$ なることを確かめる。

先ず，上の定義 (ii) と，遷移確率の定義から

$$P\{\widetilde{X}^t \neq \widetilde{X}'^t / Y = y^t_{ij}\} \leq 1 - Q^t\{y^t_{ij},\ \widetilde{E}^t_{i,j}\} \quad i = 1, \cdots, K^t_\delta,\ \ j = 1, \cdots, R^t_i .$$

故に

$$P\{\widetilde{X}^t \neq \widetilde{X}'^t\} \leq \sum_{i=1}^{K^t_\delta} \sum_{j=1}^{R^t_i} P\{Y^t = y^t_{ij}\}[1 - Q^t(y^t_{ij},\ \widetilde{E}^t_{i,j})]$$
$$+ \sum_{j=1}^{R^t_o} P\{Y^t = y^t_{oj}\} + \sum_{j=1}^{R^t_{-1}} P\{Y^t = y^t_{-1,j}\}$$
$$\leq \sum_{i=1}^{K^t_\delta} \sum_{j=1}^{R^t_i} p^t_{ij}[1 - Q^t(y^t_{ij},\ \widetilde{E}^t_{i,j})] + \bar{q}^t_o + \bar{q}^t_{-1}$$

然るに十分大きい t に対し右辺第 1 項は (2, 8) より，$\bar{q}^t_o$ は $(X^t, \widetilde{X}^t)$ の情報的安定より $(\bar{q}^t_o = 1 - P_{X^t\widetilde{X}^t}(F^t_\delta))$，及び $\bar{q}^t_{-1} = 1 - Q^t$ より，十分小さくなることが出る.

2° $(X^t,\ \widetilde{X}'^t)$ が十分大きい t に対し再生精度の条件 $\{W^t_\delta\}$ を満たすこと.

その為に $E\{\rho^t_k(X^t, \widetilde{X}'^t)\}$ を計算する.

$$E\{\rho^t_k(X^t, \widetilde{X}'^t)\} = \sum_{i=-2}^{K^t_\delta} \sum_{j=1}^{R^t_i} \int_{\{\widetilde{\omega};\, Y^t(\widetilde{\omega}) = y^t_{ij}\}} \rho^t_k(X^t,\ \widetilde{X}'^t)\ \widetilde{P}(d\widetilde{\omega}) .$$

然るに，

$$\int_{\{Y^t = y^t_{ij}\}} \rho_k(X^t, \widetilde{X}'^t)\ \widetilde{P}(d\widetilde{\omega}) = \int_{A^t} \rho_k(x^t, \widetilde{x}^t_i)\ S^t_{ij}(x^t) P_{X^t}(dx^t),$$

$$\sum_{j=1}^{R^t_o} \int_{\{Y^t = y^t_{oj}\}} \rho_k(X^t,\ \widetilde{X}'^t)\ \widetilde{P}(d\widetilde{\omega}) = \iint_{A^t \times \widetilde{A}^t - F^t_\delta} \rho_k(x^t, \widetilde{x}^t) P_{X^t\widetilde{X}^t}(dx^t, d\widetilde{x}^t),$$

$$\int_{\{Y^t = y^t_{-1,j}\}} \rho_k(X^t,\ \widetilde{X}'^t)\ \widetilde{P}(d\widetilde{\omega}) = \int_{A^t \times \widetilde{A}^t} S^t_{-1,j}(x^t)\, \rho_k(x^t, \widetilde{x}^t) P_{X^t, \widetilde{X}^t}(dx^t, d\widetilde{x}^t),$$

$$(j = 1, \cdots, R^t_{-1})$$

$i = -2$ に対応する所は 0.

これ等の関係を使うと，

$$E\{\rho_k(X^t, \tilde{X}'^t)\} = \sum_{i=1}^{K_\delta^t} \int_{A^t} q_i^t(x^t)\, \rho_k(x^t, \tilde{x}_i^t)\, P_{X^t}(dx^t)$$

$$+ \iint_{A^t \times \tilde{A}^t - F_\delta^t} \rho_k(x^t, \tilde{x}^t)\, P_{X^t, \tilde{X}^t}(dx^t, d\tilde{x}^t) + \iint_{A^t \times \tilde{A}^t} [1 - Q^t(x)]\, \rho_k(x^t, \tilde{x}^t)\, P_{X^t \tilde{X}^t}(dx^t, d\tilde{x}^t)$$

が分かる．Feinstein-Dobrushin の Lemma の結論より $(S_1^t, \cdots, S_{N^t}^t) \in [\overline{W}^t]_\varepsilon$ はいえているので，上式と，S_k^t との比較をすればよい．即ち両者の右辺における第３項の比較である．

$$\left| \iint_{A^t \times \tilde{A}^t} [1 - Q^t(x^t)]\, \rho_k(x^t, \tilde{x}^t)\, P_{X^t \tilde{X}^t}(dx^t, d\tilde{x}^t) - [1 - Q^t]\, E\{\rho_k(X^t \tilde{X}^t)\} \right|$$

$$= \left| \iint_{A^t \times \tilde{A}^t} [1 - Q^t(x^t)]\, [\rho_k(x^t, \tilde{x}^t) - E\{\rho_k(X^t, \tilde{X}^t)\}\, P_{X^t \tilde{X}^t}(dx^t, d\tilde{x}^t) \right|$$

$$\leq \iint_{A^t \times \tilde{A}^t} |\rho_k(x^t, \tilde{x}^t) - E\{\rho_k\}|^{1+b}\, P_{X^t, X^t}(dx^t, d\tilde{x}^t)]^{\frac{1}{1+b}}$$

$$\times \left[\iint_{A^t \times \tilde{A}^t} [1 - Q^t(x^t)]\, P_{X^t, \tilde{X}^t}(dx^t, d\tilde{x}^t) \right]^{\frac{b}{1+b}}$$

$$\leq [C^t]^{\frac{1}{1+b}}\, [1 - Q^t]^{\frac{b}{1+b}} \longrightarrow 0\,.$$

［註］　第１の不等式は Q^t の定義より，第２の不等式は次の Lemma を使う。
可測函数 $\varphi(x)$ と $0 \leq f(x) \leq 1$，$b > 0$ に対して，

$$\left[\int_A |\varphi(x)|\, f(x)\, P(dx) \right]^{1+b} \leq \int_A |\varphi(x)|^{1+b} P(dx) \left[\int_A f(x)\, P(dx) \right]^b$$

3° 以上の 1°, 2° によって $X^t, Y^t, \tilde{Y}^t, \tilde{X}^t, \tilde{X}'^t$ が再生精度の条件 $\{W_\varepsilon^t\}$ を満足しながら確率 ε の精度で通信を成立させることがわかる。

最後に 2.1.6.の条件 3），即ち $(Y^t, \tilde{Y}^t)$ が，$\{Q^t, V_\varepsilon^t\}$ によって結ばれていること，

$$(2.10) \qquad \left(E\{\pi_1(Y^t, \tilde{Y}^t)\}, \cdots, E\{\pi_{N^t}(Y^t, \tilde{Y}^t)\} \right) \in [\overline{V}^t]_\varepsilon$$

をいえば，基本定理の証明は終る。それには Y^t が確率 1 で y_{ij}^t の値しかとらないこと（2.9）と $\hat{P}(Y^t = y_{ij}^t) = p_{ij}^t$ ということから任意の k に対して，

$$E\{\pi_k(Y^t, \tilde{Y}^t)\} = \sum_{i=-2}^{K_\delta^t} \sum_{j=1}^{R_i^t} p_{ij}^t \int_{\tilde{B}^t} \pi_k(y_{ij}^t, \tilde{y}^t)\, Q^t(y_{ij}^t, d\tilde{y}^t)$$

となる。従って，（2.10）は（2.7）の系として得られ証明が終了する。

$\boxed{2.4}$ 通信のエントロピーと ε-entropy.

送信及び受信 $X, \tilde{X}$ は距離 ρ をもつ同一の距離空間の値をとるとする。再生精度の条件 $\{W\}$ を $E\{\rho^2(X,\tilde{X})\} \leq \varepsilon^2$ として，通信 $[W]$ を考える．若し X が n 次元ユークリッド空間の値をとり，その分布が十分滑かな密度函数をもつときは，

$$(2.11) \qquad H(W) = n \log \frac{1}{\varepsilon} + (H^0(x) - n \log \sqrt{2\pi e}) + o(1)$$

が知られている（ここで，H^0 は differential entropy）。さらに，$X = (X_1, X_2, \cdots, X_n)$，$\{X_k\}$ は独立で正規分布 $N(0, \sigma_k^2)$ に従うときは，与えられた ε に対し，$\varepsilon^2 = \sum_k \min(\theta^2, \sigma_k^2)$ なる θ をとり，

$$(2.12) \qquad H(W) = \frac{1}{2} \sum_{\sigma_k^2 > \theta^2} \log \frac{\sigma_k^2}{\theta^2}$$

と表わされる。

$X = \{X(t)\}$，$\tilde{X} = \{\tilde{X}(t)\}$，$(X, \tilde{X})$ がすべて定常過程で，$\bar{I}(X, \tilde{X})$ が存在するときを考える（$\boxed{1.3}$ を参照）。再生精度の条件として $\alpha^2 = E\{|X(t) - \tilde{X}(t)|^2\} \leq \varepsilon^2$ をとり，

$$\bar{H}(W) \equiv \inf_{P_{X,\tilde{X}} \in \{W\}} \bar{I}(X, \tilde{X})$$

と定義する．例えば，X が正規系で，そのスペクトル密度函数を $f(\lambda)$ とするとき (2.12) に対応するものとして $\varepsilon^2 = \int \min(\theta^2, f(\lambda))\, d\lambda$ で θ をきめると

$$(2.13) \qquad \bar{H}(W) = \frac{1}{2} \int_{f(\lambda) > \theta^2} \log \frac{f(\lambda)}{\theta^2}\, d\lambda$$

が知られている。

特に $f(\lambda)$ が $\varphi(\lambda) = a^2$ $(A_0 \leq |\lambda| \leq A_0 + \omega)$，$= 0$（然らざるとき），で十分近似できるときは，$\theta^2 \sim \frac{\varepsilon^2}{2\omega}$ となり，

$$(2.14) \qquad \bar{H}(W) \sim \omega \log \frac{2\omega a^2}{\varepsilon^2} = 2\omega\left(\log \frac{1}{\varepsilon} + \log \sqrt{2\omega}\, a^2\right)$$

である。そこで (2.11) と (2.14) の比較と $H(W)$，$\bar{H}(W)$ の定義から (2.14) は単位時間当り，略 2ω 次元のものが生成されていることを表わしている (Kolmogorov [85])。

Shannon の dimension rate

(2.15) $$\lambda = \lim_{\delta \downarrow 0} \lim_{\varepsilon \downarrow 0} \lim_{T \to \infty} \frac{\log N(\varepsilon, \delta, T)}{T \log \varepsilon^{-1}}$$

はこの事実に対応するもので，λ で測ると上例では，やはり $\lambda = 2\omega$ となる。但し (2.15) で $N(\varepsilon, \delta, T)$ は $[0, T]$ 区間の path 集合から確率 δ の分だけ除いた残りの ε-covering を考えたとき，その元の最小数を表わす。

Kolmogorov-Tichomirov [89] では $\{W\}$ として $P(\rho(X, \tilde{X}) \leq \varepsilon) = 1$ を考え，ε-entropy ($\boxed{4.2}$参照）との関係が説明されている。

$\boxed{3}$ *Flow*

$\boxed{3.0}$ 序

エントロピー概念の *flow* の研究へのかゝわりは，1958 年に Kolmogorov [87] によって，分類の問題の一般的な解決の手段としてエントロピーが用いられたこと（$\boxed{3.8}$ 参照）に端を発する．以来それは *flow* の研究に新しい観点と有効な方法を提供している．本章の目的はエントロピーを中心とした *flow* の研究の解説にある．

準備として，$\boxed{3.1}$ ～ $\boxed{3.4}$ で定義と基本的な事項を述べる．通信理論における Shannon の定理の Khintchin による定式化の基礎を与える McMillan の定理を $\boxed{3.5}$ で説明する．$\boxed{3.6}$ 以後が *flow* のエントロピーに関する記述である．

なお紙数の関係で混合性に関すること（例えば K-*flow* はすべての位数の混合性をもつなど），やカテゴリ理論（エントロピー 0 の automorphism は到る所稠密な G_δ 集合をなすなど）は省略する．

$\boxed{3.1}$ *flow* の定義とエルゴード性

3.1.1. 確率空間 $(\Omega, \mathbb{B}, P)$ から確率空間 $(\Omega', \mathbb{B}', P')$ への 1 価点変換 T が (i) $T^{-1}\mathbb{B}' \subset \mathbb{B}$ （可測性），(ii) $P(T^{-1}B') = P'(B')$, $\forall B' \in \mathbb{B}'$ （保測性），をみたすとき <u>homomorphism</u> という。更に T が 1 対 1 で T^{-1} も homomorphism のとき <u>isomorphism</u> という。特に $\Omega = \Omega'$ [1] のとき homomorphism を <u>endomorphism</u>，isomorphism を <u>automorphism</u> と呼ぶ。automorphism の 1 助変数群 $\{T_t;\ t \in N\}$ （N は整数又は実数全体）：

（F.1） 各 T_t は automorphism

（F.2） $T_t T_s = T_{t+s}$

（F.3） $T_0 = I$ （恒等写像）

を *flow* という．N が整数のときは *flow* を単に automorphism $T (\equiv T_1)$

---

1) $(\Omega, \mathbb{B}, P)$ を Ω と略記する．

と呼ぶことがある。以後更に可測性：

（F.4）　写像 $\Omega\times N\ni(\omega,t)\to T_t\omega\in\Omega$ が可測

を仮定する。

isomorphism で結ばれる二つの確率空間は互に *isomorphic* であるという。$\{T_t\}$, $\{T'_t\}$ がそれぞれ Ω, Ω' の *flow* のとき，Ω から Ω' への *isomorphism* S があって，$T'_t=ST_tS^{-1}$　$\forall t\in N$，であれば $\{T_t\}$ と $\{T'_t\}$ は同値（又は同じ *metrical type* をもつ）という．測度 0 の或る集合を除いて成り立つ命題には *mod* 0 と添書する．

典型的な例題をあげておく。

（例1）　$\Omega=[0,1)^{2)}$; $T_\theta x=x+\theta\ (\mathrm{mod}\ 1)$，は θ が有理数のときに限って周期的となる。θ が無理数のとき *Weyl-automorphism* という．

（例2）　*Bernoulli-automorphism*（*B-automorphism*）：確率空間 K の直積測度空間 K^N（N は整数全体）を Ω とし，*shift* $(T\omega)_n=\omega_{n-1}$ で定義される *automorphism* を *B-automorphism* という．特に K が n 個の等確率の点から成るとき *n-shift* ともいう。

（例3）　定常過程の *flow*：強定常過程（$X(t,\omega)$）を *path* 函数の空間の測度と考える．$\Omega=R^N$，$\mathcal{B}$ は *cylinder sets* から生成される σ-*algebra* とする。*process* $X(t,\omega)=\omega_t$ が定常であることは *shift* $(T_t\omega)_s=\omega_{s-t}$ が *flow* であることと同等である．また *process* が可測であれば *flow* も可測である．

（例4）　加法過程の *flow*：定常増分の加法過程では，$\mathcal{B}$ として $\{\Delta X(I)\equiv X(t)-X(s)\,;\ I=(s,t]\subset N\}$ を可測にする最小の σ-*algebra* をとって，*shift* が *flow* になる．*Brown* 運動の場合に，その *flow* を *white noise* という（[71]，[177]参照）．

B-automorphism は加法過程の *flow* の特別な場合（*discrete time*）である．

Kolmogorov の $0-1$ 法則により，加法過程の *flow* は *K-flow*（$\boxed{0}$ 参照，*Kolmogorov-flow* を *K-flow* と略称する）であることがわかる。

3.1.2.　集合 $A\in\mathcal{B}$（$\mathcal{B}$-可測函数 f）が，任意の t に対し $P(T_tA\ominus A)=0$（$f(T_t\omega)=f(\omega)$　*a.e.*）のとき，不変集合（不変函数）という．任意の

2）特に断わらない限り，区間では *Lebesgue* 測度 dx を考える。

不変集合の測度が0又は1のとき，flowはエルゴード的であるという。それは不変函数は定数（mod 0）に限ること，即ち $\lambda=0$ が単純な固有値であることと同等である．更にそれは，すべての固有値が単純であることと同等でもある。実際固有函数の絶対値は定数であり，f と g が同じ固有値に対する固有函数であれば f/g は不変函数となり，従って f は g の定数倍となる．

任意の $A,\ B\in\mathbb{B}$ に対し

$$\lim_{t\to\infty} P(A\cap T_tB)=P(A)P(B)$$

となるとき，flow $\{T_t\}$ は強混合，幾分弱く

$$\lim_{t\to\infty}\frac{1}{T}\int_0^T|P(A\cap T_tB)-P(A)P(B)|\,dt=0$$

のとき弱混合であるという．明らかに強混合 ⇒ 弱混合 ⇒ エルゴード的，である。弱混合は $\lambda=0$ が唯一の固有値であることによって特徴づけられる（[71]参照）．

Weyl-automorphism はエルゴード的であり，弱混合ではない。K-flow は強混合である．

次のエルゴード定理はよく知られている。

Birkhoff の個別エルゴード定理： 任意の $f\in L^p$ $(1\leq p<\infty)$ に対し，極限

$$(3.1)\qquad \lim_{R\to\infty}\frac{1}{R}\int_0^R f(T_t\omega)\,dt=\bar{f}(\omega),\qquad a.e.$$

が存在し，$\bar{f}$ は不変で $\bar{f}\in L^p$ である。更に任意の不変集合 A に対し

$$\int_A\bar{f}(\omega)\,dP=\int_A f(\omega)\,dP.$$

$\{T_t\}$ がエルゴード的であれば，$\bar{f}=E\{f\}$ mod 0 となる。

von Neumann の平均エルゴード定理： （3.1)式の収束は L^p-収束でもある．

3.2 スペクトル

3.2.1. flow $\{T_t\}$ はヒルベルト空間 $H=L^2(\mathbb{B})^{3)}$ にユニタリ作用素の1助変数群 $\{U_t\}$ を定義する．flowの可測性により $\{U_t\}$ は強連続となり，

Stone の定理によってスペクトル分解ができる:

$$U_t = \int e^{2\pi i t\lambda}\, dE(\lambda).$$

単位分解 $E=\{E(\lambda)\}$ に対応する Hellinger-Hahn の分解を行なえば,

$$H \simeq \sum_n \oplus L^2(\Lambda_n,\, d\rho)$$

こゝに $\Lambda_1 \supset \Lambda_2 \supset \cdots$ は $\mathrm{mod}\, 0(d\rho)$ で定まる R^1 (又は $[0,1]$) の部分集合である。

$$H \ni f \longleftrightarrow \tilde{f} = \{f_n\} \qquad f_n \in L^2(\Lambda_n,\, d\rho)$$

であれば

$$U_t f \longleftrightarrow \{e^{2\pi i t\lambda} f_n(\lambda)\}$$

$$E(\lambda) f \longleftrightarrow \{\tilde{E}(\lambda) f_n\}, \quad (\tilde{E}(\lambda)f_n)(\mu) = f(\mu) \ (\mu \le \lambda), = 0 \ (\mu > \lambda)$$

である。$m(\lambda) \equiv \max\{n;\ \lambda \in \Lambda_n\}$ を λ の multiplicity という。$d\rho$ は互に絶対連続なものを除き1意に定まる。

一般にそれぞれ H_1, H_2 上の作用素の族 $\{A_l^1\}$, $\{A_l^2\}$ がユニタリ同値であるとは, H_1 から H_2 上への isomorphism (即ち1対1等距離変換) V があって, すべての l で $A_l^2 = V A_l^1 V^{-1}$ となることをいう。次の三命題は同等である:

(i) $\{U_t^1\}$, $\{U_t^2\}$ がユニタリ同値,

(ii) 対応する単位分解 $\{E^1(\lambda)\}$, $\{E^2(\lambda)\}$ がユニタリ同値,

(iii) 対応する $d\rho^1$ と $d\rho^2$ が互に絶対連続で, $m^1(\lambda) = m^2(\lambda)\ \mathrm{mod}\, 0\,(d\rho^i)$.

これらが成り立つとき, $\{U_t^1\}$ と $\{U_t^2\}$ とはスペクトル同値 (又は同じ spectral type を持つ) という。$m(\lambda) \equiv \varkappa$ のとき multiplicity $\varkappa$ の一様スペクトル, 特に $m(\lambda) \equiv 1$ のとき単純スペクトル, $d\rho(\lambda) \sim d\lambda$ のとき Lebesgue スペクトル, multiplicity ∞ の一様 Lebesgue スペクトルを単に σ-Lebesgue スペクトルという。

3.2.2. 一様 Lebesgue スペクトルは, $L^2(d\lambda)$ を Fourier 変換 (又は展開) することによって, 次の様にいい換えられる:

$$\begin{cases} H \sim \sum_{n=1}^{\varkappa} \oplus L_n, \quad L_n = L^2(R^1,\, dx) \quad (\text{又は } \ell^2) \\ U_t f \longleftrightarrow \{f_n(x-t)\}. \end{cases}$$

---

3) $\mathcal{B}$ が Halmos [60] の意味で可分であれば $L^2(\Omega)$ は可分になる. 以後これを仮定する.

これは更に，次の3条件をみたす部分空間 H_0 が存在することと同等である：

(i) $H_t \equiv U_t H_0 \supset H_s$, $t > s$,

(ii) $\bigvee_{-\infty}^{\infty} H_t = H$

(iii) $\bigwedge_{-\infty}^{\infty} H_t = \{0\}$.

flow $\{T_t\}$ のスペクトルとは，対応する $\{U_t\}$ の（厳密には部分空間 $H \ominus 1$ における）スペクトルのことをいう．K-flow は一様 Lebesgue スペクトルを持つことが直ちにわかる。加法過程の flow はすべて σ-Lebesgue スペクトルを持つことが，重複 Wiener 積分を使って Ito によって示された([72], [107])。

3.3 Lebesgue 空間（[136]）

3.3.1. 確率空間 $(\Omega, \mathbb{B}, P)$ において，可算系 $\Gamma = \{G_n\} \subset \mathbb{B}$ があって，

(S.1) $\forall A \in \mathbb{B}$, $\exists B \in \mathcal{B}(\Gamma)$[4]: $B \supset A$, $P(B-A) = 0$,

(S.2) $\forall x, y \in \Omega$, $x \neq y$, $\exists G_n \in \Gamma$: $x \in G_n$, $y \in G_n^c$, 又は $x \in G_n^c$, $y \in G_n$,

をみたすとき，$(\Omega, \mathbb{B}, P)$ は可分であるといい，Γ を Ω の base という。(S.1) は $\mathbb{B} = \overline{\mathcal{B}(\Gamma)}$ (completion)を意味する。可分な (Ω, Γ)[5] ($\Gamma = \{G_n\}$ は base) において，$F_n = G_n$ 又は G_n^c とするとき，すべての組 $\{F_n\}$ に対し $\bigcap_n F_n \neq \phi$ であれば，Ω は完備であるという。

可分な (Ω, Γ, P)[5] に対し，その完備拡大 $(\tilde{\Omega}, \tilde{\Gamma}, \tilde{P})$:

(i) $(\tilde{\Omega}, \tilde{\Gamma}, \tilde{P})$ は可分，完備，

(ii) Ω の $\tilde{P}$-外測度は 1 ,

(iii) $\Gamma = \tilde{\Gamma} \cap \Omega$,

が常に存在することは容易にわかる．

可分な測度空間 $(\Omega, \mathbb{B}, P)$ で，完備拡大 $(\tilde{\Omega}, \tilde{\mathbb{B}}, \tilde{P})$ において Ω が可測（即ち Ω が mod 0 で完備）になるものを Lebesgue空間という．

3.3.2. atom のない Lebesgue 空間は区間 $[0,1]$ と isomorphic mod 0 である．測度代数としての isomorphic は Halmos [60] p.173 に示されているが，この場合 (S.2) によってそれが点変換としての isomor-

---

4) $\mathcal{B}(\Gamma)$ は Γ を含む最小の σ-algebra.

5) $\mathbb{B}$ の代りに base Γ を書くことあり。

phism になるのである。

Lebesgue 空间の例をあげる。

(例1) $\Omega=\{\omega_n ; n=1,2,\cdots\}$, $\sum_n P(\omega_n)=1$

(例2) $\Omega=R^n$, P:確率分布, $\mathbb{B}$: Borel 集合全体の completion. 特に ([0,1], dx) は Lebesgue 空间である。

(例3) Ω:完備可分(距离空间の意味で)な距离空间, P:確率測度, $\mathbb{B}$:開集合から生成される σ-algebra の completion. 特に $\Omega=R^{\aleph_0}$ の場合を含む。

3.3.3. Ω の分割 ζ が可測というのは, 可算系 $\Sigma=\{S_n\}\subset\mathbb{B}$ があって,

(M.P.) $\forall C_1, C_2\in\zeta$, $C_1\neq C_2$, $\exists S_n\in\Sigma$; $C_1\subset S_n$, $C_2\subset S_n^c$, 又は $C_1\subset S_n^c, C_2\subset S_n$,

となることである。Σ を ζ の base という。

ζ の元の任意の和集合を ζ-set という。可測な ζ-set の全体は明らかに σ-algebra をなすが, それを $\mathbb{B}(\zeta)$ で表わす。ζ の元を点と考える空间を Ω/ζ, Ω から Ω/ζ への自然な写像を Φ_ζ とするとき, $\mathbb{B}_\zeta\equiv\{Z ; \Phi_\zeta^{-1}Z\in\mathbb{B}\}=\Phi_\zeta(\mathbb{B}(\zeta))$, $P_\zeta(Z)\equiv P(\Phi_\zeta^{-1}Z)$ によって定義される確率空间 $(\Omega/\zeta, \mathbb{B}_\zeta, P_\zeta)$ を ζ による商空间 (factor-space) という。Φ_ζ は $(\Omega, \mathbb{B}(\zeta), P)$ から $(\Omega/\zeta, \mathbb{B}_\zeta, P_\zeta)$ の上への homomorphism になる。両者を混同し, P_ζ を P と書くことがある。Lebesgue 空间の可測分割による商空间はまた Lebesgue空间である。

Ω の任意の部分集合族 $\Sigma=\{S_\lambda ; \lambda\in\Lambda\}$ に対し, $R_\lambda=S_\lambda$ 又は S_λ^c のとき, $\zeta(\Sigma)\equiv\{C ; C=\bigcap_{\lambda\in\Lambda}R_\lambda\}$ は Ω の分割である。特に Σ が可測集合の可算系であれば $\zeta(\Sigma)$ は可測分割, Σ はその base になる。確率変数, 或いは homomorphism も分割を生成する。例えば, ξ を Ω 上の実確率変数とすれば, ξ は Ω の可測分割 $\xi\equiv\{D_a ; D_a=\xi^{-1}(a), a\in R^1\}$ を定義する。このとき Ω/ξ と $\xi(\Omega)$ とは isomorphic である。

3.3.4. 可測分割 $\zeta=\{C\}$[6] に対し, 次のような測度の標準系 $\{P(\cdot|\zeta;C)\}$ が一意に定まる:

(C.1) a.e. $C(P_\zeta)$ に対し $P(\cdot|\zeta;C)$ は C 上の測度であり, $(C, P(\cdot|\zeta;C))$ は Lebesgue 空间,

(C.2) $\mathbb{B}\ni\forall A$ に対し

---

6) 分割 ζ の元を代表して C であらわし, $\zeta=\{C\}$ とかく。

a) $A\cap C$ は a.e. $C(P_\zeta)$ に対し $P(\cdot|\zeta;C)$-可測,

b) $P(A\cap C|\zeta;C)$ は C の函数として $\mathbb{B}_\zeta$-可測,

c)
$$P(A)=\int_{\Omega/\zeta}P(A\cap C|\zeta;C)\,dP_\zeta(C).$$

$P(C^c|\zeta;C)=0$ として $P(\cdot|\zeta;C)$ を Ω 上の測度に拡大しておく。測度の標準系は条件づき確率法則の精密化であり,

$$P(\cdot|\mathbb{B}(\zeta);\omega)\equiv P(\cdot|\zeta;C),\qquad \omega\in C$$

は条件づき確率法則に他ならない.

明らかに,任意の $f\in L^1(P)$, $Z\in\mathbb{B}(\zeta)$ に対し,

$$\int_Z f(\omega)dP(\omega)=\int_Z\left(\int_C f(\omega)dP(\omega|\zeta;C)\right)dP,\quad (\text{Fubini の定理})$$

$$(E\{f;Z\}=E\{E\{f|\zeta;C\};Z\})$$

が成り立つ。その他普通の条件つき平均値に関して成り立つことは,測度の標準系による積分に対しても当然成り立つ。

3.3.5. 可測分割 ζ' が可測分割 ζ の細分 mod 0 のとき $\zeta\le\zeta'$ と書く。mod 0 で等しい可測分割のクラスの全体 [M.P.] はこの順序で完備束をなす。任意の族 $\{\zeta_\lambda;\lambda\in\Lambda\}\subset$[M.P.] に対し,可測な ζ_λ-set の全体($\lambda\in\Lambda$ も動かす)の距離 $\rho(A,B)\equiv P(A\ominus B)$ に関する稠密な可算部分集合 Σ の生成する分割 $\zeta\equiv\zeta(\Sigma)$ は,明らかに可測であり,次の性質を持つ:

1) $\zeta\ge\zeta_\lambda$, $\forall\lambda\in\Lambda$

2) $\zeta'\ge\zeta_\lambda$, $\forall\lambda\in\Lambda$, ζ':可測 $\Rightarrow$ $\zeta'\ge\zeta$.

この ζ を $\{\zeta_\lambda\}$ の上限と呼び, $\zeta=\bigvee_{\lambda\in\Lambda}\zeta_\lambda$ と表わす。下限 $\bigwedge_{\lambda\in\Lambda}\zeta_\lambda$ は,すべての λ に対し $\zeta\le\zeta_\lambda$ となる $\{\zeta\}$ の上限として定まり,次の性質をもつ:

1) $\wedge\zeta_\lambda\le\zeta_\lambda$, $\forall\lambda\in\Lambda$,

2) $\zeta'\le\zeta_\lambda$, $\forall\lambda\in\Lambda$, ζ':可測 $\Rightarrow$ $\zeta'\le\wedge\zeta_\lambda$.

[M.P.] の 0 元を ν, 単位元を ε とすれば($\nu=\{\Omega\}$, ε: 各点への分割),

$$\mathbb{B}(\nu)=\{\Omega,\phi\},\qquad \mathbb{B}(\varepsilon)=\mathbb{B}$$

であり,更に

$$\mathbb{B}(\zeta)\subset\mathbb{B}(\zeta')\Leftrightarrow\zeta\le\zeta'$$

$$\mathbb{B}(\bigvee_\lambda\zeta_\lambda)=\bigvee_\lambda\mathbb{B}(\zeta_\lambda),\qquad \mathbb{B}(\bigwedge_\lambda\zeta_\lambda)=\bigwedge_\lambda\mathbb{B}(\zeta_\lambda)$$

となる.[7]

---

7) 以上の関係はすべて mod 0 で成り立つ。

3.4 Lebesgue空間上のflow ([137])

3.4.1. TをΩ上のautomorphismとし，T-不変な可測分割$\zeta=\{C\}$（即ち$TC=C$, $\forall C\in\zeta$）に対し，（a.e.Cに対し）$T_C\equiv T|C$は$(C, P(\cdot|\zeta;C))$上のautomorphismになる．T_Cをζによって定まるTの成分という．各成分がエルゴード的であるような分解を既約分解という．不変な可測集合の全体でρ-稠密な可算個の不変（mod 0でなくきっちりと）集合を$\Sigma=\{S_n\}$とすれば$\zeta(\Sigma)$が既約分解を与えることがわかる．

既約分解は一意である．即ちζ, ζ'が共に既約分解を与えるとすれば，Ω/ζとΩ/ζ'はisomorphicであり，更に対応するCとC'においてT_Cと$T_{C'}$とは同値である．

flowに対しても同様の既約分解が可能である．

3.4.2. 測度空間$(\Omega^*, \mathcal{B}^*, P^*)$を$P^*$が必ずしも正規化されていないLebesgue空間とし，SをΩ^*上のautomorphismとする．fはΩ^*上の実可測函数で$f(\omega^*)\geq\tau>0$, $\int_{\Omega^*}f(\omega^*)\,dP^*=1$とする．直積測度空間$(\Omega^*, P^*)\times(R^1, du)$を集合$\Omega\equiv\{(\omega^*, u);\ \omega^*\in\Omega^*,\ 0\leq u<f(\omega^*)\}$に制限した確率空間$(\Omega, \mathcal{B}, P)$はLebesgue空間になる．次のように定義されるΩ上のflow $\{S_t\}$を，基礎空間Ω^*上のautomorphism Sと天井函数（ceiling function）fによって構成されるAmbrose-flow（単にA-flow）と呼ぼう．

$$S_t(\omega^*, u)\equiv\begin{cases}(\omega^*, u+t), & -u\leq t<-u+f(\omega^*)\ \text{のとき}\\ (S^n\omega^*,\ u+t-f(\omega^*)-\cdots-f(S^{n-1}\omega^*)), & \\ \qquad -u+\sum_{k=0}^{n-1}f(S^k\omega^*)\leq t<-u+\sum_{k=0}^{n}f(S^k\omega^*) & \text{のとき}\\ (S^{-n}\omega^*,\ u+t+f(S^{-1}\omega^*)+\cdots+f(S^{-n}\omega^*)), & \\ \qquad -u-\sum_{k=1}^{n}f(S^{-k}\omega^*)\leq t<-u-\sum_{k=1}^{n-1}f(S^{-k}\omega^*) & \text{のとき}\\ (n=1,2,\cdots) & \end{cases}$$

不動点を持たない可測flowはA-flowに同値(mod 0)な可算個の成分に分解される（才一表現定理）．

周期的な点のない（特にエルゴード的な）可測flowはA-flowによって表現できる（即ち或るA-flowとmod 0で同値）（才二表現定理）．

3.4.3. 可測分割ζがすべてのtで$T_t\zeta=\zeta$であれば，商空間Ω/ζにflow $(T_t)_\zeta\equiv\Phi_\zeta T_t\Phi_\zeta^{-1}$が導かれる．こゝに$\Phi_\zeta$は$\Omega$から$\Omega/\zeta$への自

然な写像である．$\{(T_t)_\zeta\}$ を <u>factor-flow</u> という．

2つの Lebesgue 空間 Ω_1, Ω_2 上にそれぞれ flow $\{T_t^1\}$, $\{T_t^2\}$ があり，Ω_1 から Ω_2 上への homomorphism H があって

$$HT_t^1 = T_t^2 H, \qquad \forall t$$

であるとき，$\{T_t^2\}$ は $\{T_t^1\}$ の <u>homomorphic image</u> という．homomorphism H から生成される分割 ζ_H による $\{T_t^1\}$ の factor-flow $\{(T_t^1)_{\zeta_H}\}$ と，$\{T_t^2\}$ は同値である．

一般に確率空間 $(\Omega, \mathbb{B}, P)$ 上の automorphism T を考える．

$$P\left(\bigcup_{n=0}^{\infty} T^n A\right) = 1 \tag{3.2}$$

をみたす部分集合 A に測度 $P_A(B) \equiv P(B)/P(A)$, $A \supset B \in \mathbb{B}$ を定義し，(A, P_A) 上の automorphism T' を次ぎのように定義する：

$$\sigma(\omega) \equiv \min\{n > 0 ; \ T^n\omega \in A\}$$
$$T'\omega \equiv T^{\sigma(\omega)}\omega, \qquad \omega \in A.$$

T' を T が導く <u>derived automorphism</u> という．T がエルゴード的であれば，$P(A) > 0$ なる任意の A に対し (3.2) が成り立ち，この場合 T' もエルゴード的になる．更に

$$\int_A \sigma(\omega)\, dP_A(\omega) = \frac{1}{P(A)}$$

が成り立つ．

$\boxed{3.5}$ <u>McMillan の定理</u>

以後 $(\Omega, \mathbb{B}, P)$ は Lebesgue 空間，分割はすべて可測分割とする．[8)]

<u>3.5.1.</u> 分割 $\alpha = \{A\}$ に対し，その不確定性の測度として，<u>エントロピー密度</u>：

$$H_0(\alpha ; \omega) \equiv -\log P(A), \qquad \omega \in A$$
$$= -\sum_{A \in \alpha} \chi_A(\omega) \log P(A)$$

を定義する．こゝに χ_A は集合 A の定義函数である．分割 $\zeta = \{C\}$ に対応する

---

8) McMillanの定理自身は一般の確率空間上の homomorphism に対して成り立つ．測度の標準系の代りに条件づき確率を使えばよい．

測度の標準系を $\{P(\cdot|\zeta;C)\}$ とするとき，同様に分割 $\alpha=\{A\}$ の条件 $\zeta=\{C\}$ つきエントロピー密度：

$$H_0(\alpha|\zeta;\omega)\equiv -\log P(A\cap C|\zeta;C), \qquad \omega\in A\cap C$$

を定義する。これは $\mathbb{B}(\alpha)\vee\mathbb{B}(\zeta)$-可測である。特に $\zeta=\nu$ のとき $H_0(\alpha|\nu)=H_0(\alpha)$ である。定義より明らかに次の性質をもつ。

(a) $H_0(\alpha|\zeta;\omega)\geq 0$

$H_0(\alpha|\zeta;\omega)=0$, a.e. $\omega \iff \alpha\leq\zeta$

(b) $\alpha\leq\zeta\Rightarrow H_0(\alpha\vee\beta|\zeta)=H_0(\beta|\zeta)$

(c) $\alpha\leq\beta\Rightarrow H_0(\alpha|\zeta)\leq H_0(\beta|\zeta)$

(d) $H_0(\alpha\vee\beta|\zeta)=H_0(\alpha|\zeta)+H_0(\beta|\zeta\vee\alpha)$.

最後の式は関係 $P(A\cap B\cap C|\zeta;C)=P(A\cap C|\zeta;C)P(B\cap C\cap A|\zeta\vee\alpha;C\cap A)$ による。

エントロピー密度の平均値：

$$H(\alpha)\equiv E\{H_0(\alpha;\omega)\}$$
$$H(\alpha|\zeta)\equiv E\{H_0(\alpha|\zeta;\omega)\}$$

をそれぞれ分割 α のエントロピー，分割 α の条件 ζ つきエントロピーと呼ぶ。$H(\alpha)$ が有限であるような分割 α の全体を $\mathbb{Z}$ と書く。$\alpha\in\mathbb{Z}$ であるためには α が可算分割（mod 0）であることが必要であり，このとき

$$H(\alpha)=-\sum_{A\in\alpha}P(A)\log P(A)$$
$$H(\alpha|\zeta)=-\sum_{A\in\alpha}E\{P(A|\zeta;C)\log P(A|\zeta;C)\}$$

となる.

3.5.2. σ-algebra の列 $\{\mathbb{B}_n\}$ が単調増大で $\mathbb{B}_\infty=\overset{\infty}{\vee}\mathbb{B}_n$ であれば

$$\lim_{n\to\infty}P(A|\mathbb{B}_n;\omega)=P(A|\mathbb{B}_\infty;\omega), \qquad \text{a.e.}\ \omega$$

（Doob [33] VII，定理 4.3）。α が可算分割だから，

(e) $\zeta_n\uparrow$, $\zeta=\overset{\infty}{\vee}\zeta_n \Rightarrow \lim_{n\to\infty}H_0(\alpha|\zeta_n)=H_0(\alpha|\zeta)$, a.e. ω.

更にこの場合

$$E\{\sup_n H_0(\alpha|\zeta_n)\}<+\infty \tag{3.3}$$

がわかる。実際，α の元に $P(A_j)\downarrow$ であるように番号をつけて $\alpha=\{A_j\}$ としておけば，

$$E\{\sup_n H_0(\alpha|\zeta_n)\}$$

$$
\begin{aligned}
&= \sum_{r=0}^{\infty} E\{\sup_n H_0(\alpha|\zeta_n);\ r \le \sup_n H_0(\alpha|\zeta_n) < r+1\} \\
&< \sum_{r=0}^{\infty} (r+1)\, P(r \le \sup_n H_0(\alpha|\zeta_n) < r+1) \\
&= \sum_{r=0}^{\infty} P(\sup_n H_0(\alpha|\zeta_n) \ge r) \\
&= \sum_{r=0}^{\infty} \sum_{k=1}^{\infty} P(E^r(k))
\end{aligned}
$$

こゝに

$$E^r(k) \equiv \{\omega;\ \max_{1\le \ell \le k-1} H_0(\alpha|\zeta_\ell;\omega) < r,\ H_0(\alpha|\zeta_k;\omega) \ge r\}\ .$$

更に

$$E_j^r(k) \equiv \{\omega;\ \max_{1\le \ell \le k-1} \{-\log P(A_j|\zeta_\ell;\omega)\} < r,\ -\log P(A_j|\zeta_k;\omega) \ge r\}$$

とおけば，$A_j \cap E^r(k) = A_j \cap E_j^r(k)$，$E_j^r(k) \in \mathbb{B}(\zeta_k)$ だから

$$P(E^r(k) \cap A_j) = \int_{E_j^r(k)} P(A_j|\zeta_k)\,dP \le e^{-r} P \le e^{-r} P(E_j^r(k))\ .$$

従って $f(r) = e^r(r+1)^{-2}$ として，

$$
\begin{aligned}
&\sum_{k=1}^{\infty} P(E^r(k)) = \sum_k \sum_j P(E^r(k) \cap A_j) \\
&\le e^{-r} \sum_{j \le f(r)} \sum_k P(E_j^r(k)) + \sum_{j > f(r)} P(A_j) \\
&\le \frac{1}{(r+1)^2} + \sum_{j > f(r)} P(A_j)\ .
\end{aligned}
$$

故に

$$E\{\sup_n H_0(\alpha|\zeta_n)\} \le \sum_{r=0}^{\infty} \frac{1}{(r+1)^2} + \sum_{j=1}^{\infty} f^{-1}(j)\, P(A_j)\ .$$

$P(A_j)$ の単調性より $jP(A_j) \le 1$ となり，またすべての j に対し $f^{-1}(j) \le a \log j + b$ となる定数 a, b が存在するので，ｵ2項 $\le a H(\alpha) + b$ となって（3.3）が示された。

（3.3）により $\{H_0(\alpha|\zeta_n)\}$ は一様可積分であることがわかり，(e) とあわせると，

(f) $H(\alpha) < +\infty$, $\zeta_n \uparrow$, $\zeta = \bigvee^{\infty} \zeta_n \Rightarrow \lim_{n\to\infty} H_0(\alpha|\zeta_n) = H_0(\alpha|\zeta)$, L^1-収束.

<u>3.5.3.</u> $\alpha \in \mathbb{Z}$ に対し (e) (f) により，$H_0(\alpha|\bigvee_1^{\infty} T^{-k}\alpha)$ に概収束及び L^1-収束する．h_0 は可積分だからエルゴード定理により極限

$$\bar{h}_0(\omega) \equiv \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} h_0(T^k\omega), \qquad a.e.\ \omega,\ L^1\text{-収束}$$

が存在する。

$$h(T,\alpha) \equiv E\{h_0(\omega)\} = H(\alpha \mid \bigvee_1^\infty T^{-k}\alpha)$$

とおく。

<u>McMillan の定理</u>　分割 α が $H(\alpha) < +\infty$ であれば

$$\lim_{n\to\infty} \frac{1}{n} H_0(\bigvee_{k=0}^{n-1} T^{-k}\alpha) = \bar{h}_0, \qquad a.e.\ \text{かつ}\ L^1\text{-収束}$$

が成り立つ。従ってまた

$$\lim_{n\to\infty} \frac{1}{n} H(\bigvee_{k=0}^{n-1} T^{-k}\alpha) = h(T,\alpha). \tag{3.4}$$

特に T がエルゴード的であれば

$$\bar{h}_0(\omega) = h(T,\alpha) \qquad a.e.\ \omega.$$

証明の概略を記す。$g_0(\omega) \equiv H_0(\alpha;\omega)$, $g_k(\omega) \equiv H_0(\alpha \mid \bigvee_1^k T^{-j}\alpha;\omega)$ とおけば関係 $H_0(\alpha \mid \beta; T\omega) = H_0(T^{-1}\alpha \mid T^{-1}\beta;\omega)$ と (d) により

$$\frac{1}{n} H_0(\bigvee_{k=0}^{n-1} T^{-k}\alpha;\omega) = \frac{1}{n} \sum_{k=0}^{n-1} g_k(T^{n-k-1}\omega)$$

$$= \frac{1}{n} \sum_{k=0}^{n-1} h_0(T^k\omega) + \frac{1}{n} \sum_{k=0}^{n-1} \{g_k(T^{n-k-1}\omega) - h_0(T^{n-k-1}\omega)\}.$$

第1項は $\bar{h}_0$ に概収束及び L^1-収束するので，第2項がそれぞれの意味で 0 に収束することを示せばよい．先ず (f) により

$$E\{|\text{第2項}|\} \le \frac{1}{n} \sum_{k=0}^{n-1} E\{|g_k(\omega) - h_0(\omega)|\} \longrightarrow 0.$$

(3.3) により $G_N(\omega) \equiv \sup_{k \ge N} |g_k(\omega) - g(\omega)| \in L^1$ だから

$$\overline{\lim_{n\to\infty}} |\text{第2項}| \le \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} G_N(T^k\omega) = \bar{G}_N(\omega), \qquad a.e.$$

$G_N \downarrow 0$ だから $E\bar{G}_N = EG_N \downarrow 0$ $(N\to\infty)$，従って $\bar{G}_N(\omega) \to 0$, a.e. となり，第2項は 0 に概収束する．残りの部分は明らかである。

$h(T,\alpha)$ は T の α に関するエントロピーというが，(3.4) に注目して<u>単位時間当りのエントロピー</u>とも呼ばれる ([82])。

3.6 flowのエントロピー

3.6.1. H_0 の性質から直ちにエントロピーの次の性質を得る.

(a) $H(\xi|\zeta) \geq 0$, "$=$" $\Leftrightarrow$ $\xi \leq \zeta$

$H(\xi|\nu) = H(\xi)$

(b) $\xi \leq \zeta \Rightarrow H(\xi \vee \eta|\zeta) = H(\eta|\zeta)$

(c) $\xi \leq \eta \Rightarrow H(\xi|\zeta) \leq H(\eta|\zeta)$

(d) $H(\xi \vee \eta|\zeta) = H(\xi|\zeta) + H(\eta|\zeta \vee \xi)$

(e) $\alpha \in \mathbb{Z}$, $\zeta_n \uparrow$, $\zeta = \overset{\infty}{\vee} \zeta_n \Rightarrow \lim_{n\to\infty} H(\alpha|\zeta_n) = H(\alpha|\zeta)$.

更に

(f) $\xi_n \uparrow$, $\xi = \overset{\infty}{\vee} \xi_n \Rightarrow \lim_{n\to\infty} H(\xi_n|\zeta) = H(\xi|\zeta)$

がわかる。

まず, Lを $H(\xi) = \infty$ のときは任意に大きい数, $H(\xi) < \infty$ のときは任意に小さい $\varepsilon > 0$ に対し $L \equiv H(\xi) - \varepsilon$ とするとき

(3.5) $$L \leq H(\alpha)$$

をみたす有限分割 $\alpha \leq \xi$ が存在することを示そう. 実際, $\xi = \{D_j\}$が可算無限分割であれば, Nがあって

$$L \leq -\sum_{j=1}^{N} P(D_j) \log P(D_j)$$

となる. 従って $\alpha \equiv \{D_0 \equiv \bigcup_{j\geq N+1} D_j, D_1, D_2, \cdots, D_N\}$ が求めるものである. ξが可算分割でなければ, 正測度の ξ-set A があって, Aに含まれるξの元の測度は0となる。従つて, 任意のnに対し, Aの ξ-sets による有限分割 $\{A_j^{(n)};\ 1 \leq j \leq k_n\}$があって

$$P(A_j^{(n)}) < \frac{1}{n}, \quad 1 \leq j \leq k_n$$

をみたす。$\alpha_n \equiv \{A_1^{(n)}, \cdots, A_{k_n}^{(n)}, A_0^{(n)} \equiv A^c\}$ とおけば

$$H(\alpha_n) \geq P(A) \log n$$

となって (3.5) をみたす $\alpha (\leq \xi)$ の存在がわかつた。

$\zeta = \nu$ のときに (f) を示す. 左辺の極限の存在と不等号$\leq$は (c) より明らかである. 上のαに対し (e) と (a) により

$$\lim_{n\to\infty} H(\alpha|\xi_n) = H(\alpha|\xi) = 0 .$$

従って (c), (d) より

$$\lim_{n\to\infty} H(\xi_n) = \lim_{n\to\infty} H(\xi_n \vee \alpha) \geq H(\alpha) \geq K .$$

一般の $\zeta=\{C\}$ に対しては，空間 $(C, P(\cdot|\zeta; C))$ で $H_1(\xi|\zeta; C) \equiv E\{H_0(\xi|\zeta;\omega)|\zeta; C\}$ を考えると (Ω, P) での $H(\xi)$ と全く同じ性質を持つので，上のことより

$$\lim_{n\to\infty} H_1(\xi_n|\zeta; C) = H_1(\xi|\zeta; C) , \qquad a.e. C\in\zeta .$$

これは単調増大極限だから，積分して (f) が出る。

分割 ξ の可測性により (f) の条件をみたす有限分割の増加列が存在するので，$H(\xi|\zeta)$ は有限分割の（条件つき）エントロピーの極限として求まる．応用として例えば，

(g) $\quad \eta\leq\zeta \Rightarrow H(\xi|\eta) \geq H(\xi|\zeta)$

が示せる．ξ が有限分割であれば，この関係は函数 $-x \log x$ が凸なることより明らかで，一般の場合はその極限として求まる。(d) (g) により，

(h) $\quad H(\xi\vee\eta|\zeta) \leq H(\xi|\zeta) + H(\eta|\zeta)$

$\xi \perp\!\!\!\perp \eta$ [9] $\Rightarrow H(\xi\vee\eta) = H(\xi) + H(\eta)$

が成り立つ。

$\mathbb{Z}$ において

$$\sigma(\alpha|\beta) \equiv H(\alpha|\beta) + H(\beta|\alpha)$$

を定義すれば，(a) と関係

$$\begin{aligned} H(\alpha|\beta) &\leq H(\alpha\vee\gamma|\beta) \\ &= H(\alpha|\beta\vee\gamma) + H(\gamma|\beta) \\ &\leq H(\alpha|\gamma) + H(\gamma|\beta) \end{aligned}$$

により，σ は $\mathbb{Z}$ の距離になる．(e), (f) により有限分割は $\mathbb{Z}$ の中で σ-稠密であることがわかる。

3.6.2. automorphism T と分割 α に対し

$$\alpha_T^n \equiv \bigvee_{k=0}^{n-1} T^k\alpha , \quad \alpha_T^- \equiv \bigvee_{k=1}^{\infty} T^{-k}\alpha , \quad \alpha_T \equiv \bigvee_{-\infty}^{\infty} T^k\alpha$$

$$h(T,\alpha) \equiv H(\alpha|\alpha_T^-) , \qquad h(T) \equiv \sup_{\alpha\in\mathbb{Z}} h(T,\alpha)$$

---

9) ξ と η が独立，即ち $\mathbb{B}(\xi)$ と $\mathbb{B}(\eta)$ が独立。

とおく．$h(T)$ を automorphism T のエントロピーという．(3.4) より

1°) $\alpha \in \mathbb{Z} \Rightarrow h(T,\alpha) = \lim_{n\to\infty} \frac{1}{n} H(\alpha_T^n)$.

$H(\alpha_T^n) = H(\alpha_{T^{-1}}^n)$ だから，$\alpha \in \mathbb{Z}$ に対し $h(T,\alpha) = h(T^{-1},\alpha)$，従って

2°) $h(T) = h(T^{-1})$.

函数 $h(T,\alpha)$ は $\alpha \in \mathbb{Z}$ に関し σ-連続，即ち

3°) $|h(T,\alpha) - h(T,\beta)| \leq \sigma(\alpha,\beta)$, $\alpha,\beta \in \mathbb{Z}$.

実際，関係

$$H(\alpha_T^n) - H(\beta_T^n) = H(\alpha_T^n | \beta_T^n) - H(\beta_T^n | \alpha_T^n)$$

$$H(\alpha_T^n | \beta_T^n) \leq \sum_{k=0}^{n-1} H(T^k\alpha | \beta_T^n) \leq \sum_{k=0}^{n-1} H(T^k\alpha | T^k\beta)$$
$$= nH(\alpha|\beta)$$

が成り立つから，従って $h(T)$ の定義において，有限分割だけに関する上限としてもよいことがわかる．

4°) $h(T,\alpha) \leq h(T,\beta) + H(\alpha|\beta_T) \leq h(T,\beta) + H(\alpha|\beta)$, $\alpha,\beta \in \mathbb{Z}$.

何故ならば，

$$\begin{aligned} H(\alpha_T^n) &\leq H(\alpha_T^n \vee T^{-k}\beta_T^{2k+n}) \\ &= H(T^{-k}\beta_T^{2k+n}) + H(\alpha_T^n | T^{-k}\beta_T^{2k+n}) \\ &\leq H(\beta_T^{2k+n}) + nH(\alpha | T^{-k}\beta \vee \cdots \vee T^k\beta) \end{aligned}$$

だから，最初の不等式より，

5°) $\alpha \in \mathbb{Z}$, $\alpha_T = \varepsilon \Rightarrow h(T,\alpha) = h(T)$.

第2の不等式より

6°) $\alpha_n \uparrow$, $\{\alpha_n\} \subset \mathbb{Z}$, $\bigvee_n \alpha_n = \varepsilon$

$\Rightarrow$ 任意の automorphism T に対し，$h(T) = \lim_{n\to\infty} h(T,\alpha_n)$

が出る．実際，任意の $\beta \in \mathbb{Z}$ に対し

$$h(T,\beta) \leq h(T,\alpha_n) + H(\beta|\alpha_n), \qquad \forall n$$

(c) と 1°) により $h(T,\alpha_n)$ は単調増大であり，$H(\beta|\varepsilon) = 0$ だから (e) により 6°) が出る．

ζ を $T\zeta = \zeta$ なる分割とすれば，Ω/ζ の有限分割の Φ_ζ による逆像は Ω の有限分割だから，

7°) T_1 が T_2 の factor-automorphism (或いは homomorphic image) であれば，

$$h(T_1) \leq h(T_2).$$

従って

8°) T_1 と T_2 が同じ *metrical type* を持てば,

$$h(T_1) = h(T_2).$$

即ちエントロピーは *metrical invariant* である。

9°) $\zeta_n \uparrow$, $\bigvee_n \zeta_n = \varepsilon$, $T\zeta_n = \zeta_n \quad \forall n$

$$\Rightarrow \qquad \lim_{n\to\infty} h(T_{\zeta_n}) = h(T).$$

実際 $\{\zeta_n\} \subset \mathbb{Z}$ であれば, 5°) より $h(T_{\zeta_n}) = h(T, \zeta_n)$ だから 6°) の特別な場合である。一般の場合は, 有限分割の増大列 $\{\alpha_k\}$ で, 各 k に対し $\alpha_k \leq \zeta_{n_k}$ となる n_k があり, 更に $\bigvee \alpha_k = \varepsilon$ となるものが存在するので,

$$h(T) \geq h(T_{\zeta_{n_k}}) \geq h(T, \alpha_k).$$

7°) より $h(T_{\zeta_n})$ は単調増大だから, 6°) により 9°) が出る。 9°) の特別な場合として,

10°) $\alpha_n \uparrow$, $\{\alpha_n\} \subset \mathbb{Z}$, $\bigvee_n (\alpha_n)_T = \varepsilon$

$$\Rightarrow \qquad h(T) = \lim_{n\to\infty} h(T, \alpha_n)$$

が成り立つ。実際 $\zeta_n \equiv (\alpha_n)_T$ は 9°) の仮定をみたし, 5°) より

$$h(T, \alpha_n) = h(T_{\zeta_n})$$

だから,

直積測度空間 $\Omega_1 \times \Omega_2$ に *automorphism* の直積 $T_1 \times T_2$: $T_1 \times T_2(\omega_1, \omega_2) \equiv (T_1\omega_1, T_2\omega_2)$ が定義されていれば, 明らかに,

11°) $$h(T_1 \times T_2) = h(T_1) + h(T_2)$$

が成り立つ。

<u>3.6.3.</u> 任意の *automorphism* T と任意の整数 n に対し

12°) $$h(T^n) = |n|\, h(T)$$

が成り立つ。 2°) により $n > 0$ のとき示せばよい。任意の $\alpha \in \mathbb{Z}$ に対し

$$\begin{aligned} h(T^n, \alpha) &= \lim_{k\to\infty} \frac{1}{k} H(\alpha \vee T^n\alpha \vee \cdots \vee T^{n(k-1)}\alpha) \\ &\leq \lim_{k\to\infty} \frac{n}{nk} H(\alpha \vee T\alpha \vee \cdots \vee T^{nk-1}\alpha) \\ &= n h(T, \alpha). \end{aligned}$$

逆に任意の $\delta > 0$ に対し $\alpha \in \mathbb{Z}$ があって

$$h(T, \alpha) > h(T) - \delta.$$

$\beta \equiv \alpha_T^n \in \mathbb{Z}$ とおけば $\beta_{T^n}^k = \alpha_T^{nk}$ だから，

$$h(T^n, \beta) = \lim_{k\to\infty} \frac{1}{k} H(\beta_{T^n}^k) = \lim_{k\to\infty} \frac{n}{nk} H(\alpha_T^{nk}) = nh(T, \alpha).$$

従って，$h(T^n) \geq nh(T, \alpha) > n(h(T) - \delta)$.

注意. 以上3.6.1.～3.6.3.の諸性質は一般の確率空間でも成り立つ。ただし，$\boxed{3.5}$ の冒頭の脚註に述べた様に，測度の標準系の代りに条件づき確率を使い，可算でない可測分割ξに対しては $H(\xi|\zeta) = \sup\{H(\alpha|\zeta);\ \xi \geq \alpha$：有限分割$\}$ と定義しなければならない。

3.6.4. $\zeta = \{C\}$ が T-不変な分割であれば，成分 T_C のエントロピーについて次ぎの関係が成り立つ。

13°) $$h(T) = \int_{\Omega/\zeta} h(T_C)\, dP_\zeta(C).$$

実際，$\{\alpha_n\}$を有限分割の単調増大列で $\bigvee_n \alpha_n = \varepsilon$ となるものとすれば，各$C \in \zeta$ に対し $\{\alpha_n \cap C\}$ はCにおいて同じ性質をもつ。従って 6°) により，単調増大極限として

$$\lim_{n\to\infty} h(T, \alpha_n) = h(T)$$

$$\lim_{n\to\infty} h(T_C, \alpha_n \cap C) = h(T_C) \qquad \forall C \in \zeta$$

を得る。一方定義により

$$\begin{aligned} h(T_C, \alpha_n \cap C) &= E\{H_0(\alpha_n \cap C | (\alpha_n \cap C)_{T_C}^-) | \zeta; C\} \\ &= E\{H_0(\alpha_n | (\alpha_n)_T^-) | \zeta; C\} \end{aligned}$$

$$\begin{aligned} &\int_{\Omega/\zeta} h(T_C, \alpha_n \cap C)\, dP_\zeta \\ &= E\{E\{H_0(\alpha_n | (\alpha_n)_T^-) | \zeta; C\}\} \\ &= h(T, \alpha_n). \end{aligned}$$

T' を集合A上にTが導く derived automorphism とすると，

14°) $$h(T')\,P(A) = h(T)$$ [10]

が成り立つ。

先ずTがエルゴード的であることを仮定しよう。T' もエルゴード的である。A上の確率変数 $\sigma(\omega) = \min\{n > 0;\ T^n\omega \in A\}$ によって定まるAの可算

---

10) エントロピーhやHは測度を正規化して定義する。

分割を α' とすると，$E\{\sigma\} = 1/P(A) < \infty$ により $H(\alpha') < \infty$ である。ξ' を $\xi' \geq \alpha'$，$H(\xi') < \infty$ であるような A の分割とし，$\xi \equiv \{\xi', A^c\}$ を対応する Ω の分割とする。$\varphi_n(\omega) \equiv 1 + \sigma(T'^{-1}\omega) + \cdots + \sigma(T'^{-n+1}\omega)$ に対し，個別エルゴード定理により

$$\lim_{n\to\infty} \frac{1}{n}\varphi_n(\omega) = \int_A \sigma(\omega)\, dP_A = \frac{1}{P(A)}, \qquad a.e.\ \omega \in A\ (P_A).$$

他方一般に $\zeta = \{C_j\} \in \mathbb{Z}$，$B \in \mathbb{B}$ に対し

$$H(\zeta \cap A) \equiv -\sum_j P(C_j \cap A) \log P(C_j \cap A)$$

とおくとき，任意の ξ_T^n-set B に対し $H(\xi_T^n \cap B) = E\{H_0(\xi_T^n); B\}$ だから，$McMillan$ の定理により

$$\lim_{n\to\infty} \frac{1}{n} H(\xi_T^n \cap B) = h(T, \xi)\, P(B).$$

上の２つの注意と $Egorov$ の定理により，任意の $\delta > 0$ に対し A_δ と n_0 があって，$P(A - A_\delta) < \delta$，すべての $n > n_0$ に対し次ぎの (i) (ii) (iii) が成り立つ。

(i) $\quad \left|\frac{1}{n}\varphi_n(\omega) - \frac{1}{P(A)}\right| < \delta, \qquad \forall \omega \in A_\delta$

(ii) $\quad \left|\frac{1}{n} H(\xi_T^n \cap B) - h(T, \xi)\, P(B)\right| < \delta, \quad \forall B : \xi_T^n\text{-set}$

(iii) $\quad \left|\frac{1}{n} H(\xi'^{\,n}_{T'} \cap B') - h(T', \xi')\, P_A(B')\right| < \delta, \quad \forall B' : \xi'^{\,n}_{T'}\text{-set}.$

（勿論 (iii) は空間 (A, P_A) で考えている）．更に任意の $E \subset A - A_\delta \equiv D$ に対し，$H(\xi_T^n \cap E) \leq H(\xi_T^n \cap D)$ であり，$\eta \equiv \{D, D^c\}$ とおけば

$$\overline{\lim_{n\to\infty}} \frac{1}{n} H(\xi_T^n \cap E) \leq \overline{\lim_{n\to\infty}} \frac{1}{n} H(\xi_T^n \cap D)$$

$$= h(T, \xi \vee \eta)\, P(D) \leq (H(\xi) + H(\eta))\, P(D).$$

こゝで $H(\eta) \leq \log 2 < 2$，$P(D) < \delta$ だから，$n > n_0$ のとき

(iv) $\quad \frac{1}{n} H(\xi_T^n \cap E) < \delta\,(H(\xi) + 2), \qquad \forall E \subset D$

が成り立つとしてよい。$\xi'^{\,n}_{T'}$ の元の上で各 $\sigma(T'^{-k}\omega)$ は，従って $\varphi_n(\omega)$ は，定数となる。従って $A^{(n)} \cap A_\delta \neq \phi$ であるような $A^{(n)} \in \xi'^{\,n}_{T'}$ の和集合 E_n 上で (i) が成り立つ。$\xi'^{\,n}_{T'} \ni A^{(n)}$ 上で $\varphi_n(\omega) = m$ であれば $A^{(n)} \in \xi_T^m$ となることが，$T'^{-k}\omega = T^{-\varphi_k(\omega)}\omega$ を使って示されるので，$m_1 \equiv min\{\varphi_n(\omega);$

$\omega \in E_n\}$, $m_2 \equiv max\{\varphi_n(\omega); \omega \in E_n\}$ に対し $\xi_T^{m_1} \cap E_n \leq \xi_{T'}^{\prime n} \cap E_n \leq \xi_T^{m_2} \cap E_n$ となる。従って

$$H(\xi_T^{m_1} \cap E_n) \leq H(\xi_{T'}^{\prime n} \cap E_n) P(A) \leq H(\xi_T^{m_2} \cap E_n).$$

$\zeta \equiv \{E_n, E_n^c\}$とおけば，$i = 1, 2$ のとき

$$H(\xi_T^{m_i} \cap A) \leq H((\xi_T^{m_i} \vee \zeta) \cap A) = H(\xi_T^{m_i} \cap E_n) + H(\xi_T^{m_i} \cap (A - E_n))$$
$$\leq H(\xi_T^{m_i} \cap A) + 2.$$

$A - E_n \subset D$ だから(iv)より

$$\frac{1}{m_i} H(\xi_T^{m_i} \cap E_n) = \frac{1}{m_i} H(\xi_T^{m_i} \cap A) + O(\delta), \qquad i = 1, 2 .$$

(ii)より

$$\frac{1}{m_i} H(\xi_T^{m_i} \cap A) = h(T, \xi) P(A) + O(\delta)$$

だから，$m_i/n = 1/P(A) + O(\delta)$，$(i = 1, 2)$ を考慮して

$$\frac{P(A)}{n} H(\xi_{T'}^{\prime n} \cap E_n) = h(T, \xi) + O(\delta).$$

一方，$P(A) - P(E_n) < \delta$ だから，(iii)より

$$\frac{P(A)}{n} H(\xi_{T'}^{\prime n} \cap E_n) = h(T', \xi') P(A) + O(\delta)$$

となり，結局，

$$h(T, \xi) = h(T', \xi') P(A)$$

を得た。$\{\xi_n'\} \subset \mathbb{Z}$ $(\xi_n' \geq \alpha')$ をAの分割の増大列で，$\bigvee_n \xi_n' = \varepsilon(A)$ となるものとし，$\xi_n \equiv \{\xi_n', A^c\}$ とおけば，$\{\xi_n\} \subset \mathbb{Z}$，$\bigvee_n (\xi_n)_T = \varepsilon$ (Tのエルゴード性による)である。上のことより，

$$h(T, \xi_n) = h(T', \xi_n') P(A).$$

$n \to \infty$ として，6°)，10°) により，14°)を得る。

Tがエルゴード的でないときは，既約分解をして，13°)を使えばよい。

<u>3.6.5.</u> $(\Omega^*, \mathcal{B}^*, P^*)$ を Lebesgue 空間，$U = [0, 1)$ を普通の Lebesgue 測度 du を持つ区間とする。$\varphi: \Omega^* \to U$ は可測函数，S は Ω^* 上の automorphism とする。直積測度空間 $\Omega^* \times U$ に，automorphism

$$S'(\omega^*, u) \equiv (S\omega^*, u + \varphi(\omega^*)), \ (mod\, 1), \ \omega^* \in \Omega^*, \ u \in U$$

を定義する。このとき

(3.6) $$h(S') = h(S)$$

が成り立つ。$h(S') \geq h(S)$ は明らかだから，逆の不等式を示す．$\{G_n\}$ を Ω^* の base，$G_1, G_2, \cdots, G_n$ から出来る分割を α_n，$\xi_n \equiv \alpha_n \times U$，$\xi \equiv \bigvee_n \xi_n = \{\omega^* \times U;\ \omega^* \in \Omega^*\}$，$U$ の k 等分を $\beta_k = \{\Delta_k^1, \cdots, \Delta_k^k\}$，$\zeta_k \equiv \Omega^* \times \beta_k = \{\Omega^* \times \Delta_k^j;\ 1 \leq j \leq k\}$ とする．分割 $(\zeta_k)_{S'}^n$ は各元 $\omega^* \times U \in \xi$ 上に高々 nk 個の線への分割を導くので，

$$\lim_{m\to\infty} H((\zeta_k)_{S'}^n \mid \xi_m) = H((\zeta_k)_{S'}^n \mid \xi) \leq \log nk .$$

任意の $\delta > 0$ に対し n_k を $\frac{1}{n_k} \log n_k k < \frac{\varepsilon}{2}$ （$k = 1, 2, \cdots$ ）となる正整数とする。上のことより

$$H((\zeta_k)_{S'}^{n_k} \mid \xi_m) < \log n_k k + \frac{\delta}{2}$$

となる m が存在する．このような m の最小値を m_k' とし，$m_0 \equiv 1$，$m_k \equiv \max\{m_{k-1}, m_k', k\}$ （$k = 1, 2, \cdots$ ）とおけば，すべての k, ℓ に対し

$$\frac{1}{\ell n_k} H((\zeta_k)_{S'}^{\ell n_k} \mid (\xi_{m_k})_{S'}^{\ell n_k}) \leq \frac{1}{\ell n_k} \sum_{j=0}^{\ell-1} H(S'^{j n_k} (\zeta_k)_{S'}^{n_k} \mid (\xi_{m_k})_{S'}^{\ell n_k})$$
$$\leq \frac{1}{n_k} H((\zeta_k)_{S'}^{n_k} \mid \xi_{m_k}) < \delta .$$

$H((\xi_m)_{S'}^n) = H((\alpha_m)_S^n)$ だから，式

$$H((\xi_{m_k} \vee \zeta_k)_{S'}^{\ell n_k}) = H((\xi_{m_k})_{S'}^{\ell n_k}) + H((\zeta_k)_{S'}^{\ell n_k} \mid (\xi_{m_k})_{S'}^{\ell n_k})$$

の両辺を ℓn_k で割って $\ell \to \infty$ とすれば，

$$h(S', \xi_{m_k} \vee \zeta_k) \leq h(S, \alpha_{m_k}) + \delta .$$

$k \to \infty$ として，6°) により

$$h(S') \leq h(S) + \delta .$$

$\{S_t\}$ を Ω^* 上の automorphism S と函数 $f(\omega^*) > \tau > 0$ によって，$\Omega = \{(\omega^*, u);\ \omega^* \in \Omega^*,\ 0 \leq u < f(\omega^*)\}$ 上に構成される A-flow とする．このとき，

15°) $$h(S_t) = |t|\, h(S)\, P^*(\Omega^*)$$

が成り立つ。実際，$0 < t \leq \tau$ に対し

$$\varphi_t(\omega^*) \equiv t - f(\omega^*) + t\left[\frac{1}{t} f(\omega^*)\right]^{11)}$$

とおけば，$0 \leq \varphi_t(\omega^*) < t$ であり，各 t に対し

$$S'_t(\omega^*, u) \equiv (S\omega^*, u + \varphi_t(\omega^*)), \quad (mod\ t), \quad 0 \leq u < t$$

は $\Omega_t \equiv \Omega^* \times [0, t)$ 上の automorphism を定義する。(3.6) より

$$h(S'_t) = h(S).$$

他方 S'_t は S_t から Ω_t 上に導かれた derived automorphism になっているので，14°) により

$$h(S'_t)\, t P^*(\Omega^*) = h(S_t).$$

即ち $0 < t \leq \tau$ に対し 15°) が示された。任意の $t \neq 0$ に対しては，$0 < \frac{t}{n} \leq \tau$ となる整数 n をとり 12°) により

$$h(S_t) = |n|\, h(S_{\frac{t}{n}}) = |t|\, h(S)\, P^*(\Omega^*).$$

3.6.6. 一般の可測 flow $\{T_t\}$ に対し

16°) $$h(T_t) = |t|\, h(T_1), \qquad \forall t$$

が成り立つ。不動点の全体 Ω_0 は可測な不変集合であり，Ω_0^c で $\{T_t\}$ は A-flow と同値な可算個の成分に分解される。対応する分割を $\{\Omega_j\}$ とすれば，各 t に対し 13°) により

$$h(T_t) = \sum_{j \geq 0} h((T_t)_{\Omega_j})\, P(\Omega_j) .$$

明らかに $h((T_t)_{\Omega_0}) = 0$ であり，$j \geq 1$ に対し，15°) より

$$h((T_t)_{\Omega_j}) = |t|\, h((T_1)_{\Omega_j}) .$$

故に再び 13°) により，

$$h(T_t) = |t| \sum_{j \geq 0} h((T_1)_{\Omega_j})\, P(\Omega_j) = |t|\, h(T_1).$$

$h(T_1)$ を flow $\{T_t\}$ のエントロピーという。

3.6.7. 若干の典型的な場合にエントロピーを計算しておく。先ず一般に

17°) $$\exists \alpha \in \mathbb{Z}, \quad \bigvee_0^{\infty} T^{-k}\alpha = \varepsilon \Rightarrow h(T) = 0.$$

実際，$\alpha \leq \varepsilon = T^{-1}\varepsilon = \alpha_T^-$ だから 5°) により，$h(T) = h(T, \alpha) = H(\alpha | \alpha_T^-) = 0$ である。

T が周期 p の周期的 automorphism であれば，T^p は恒等変換で，恒等変換のエントロピーは明らかに 0 であるから

$$h(T) = p^{-1} h(T^p) = 0 .$$

---

11) [ ] は Gauss 記号

特に3.1.1.の例1の T_θ（θ：有理数）のエントロピーは 0 である。

Ω が atom から成るときは，すべての automorphism のエントロピーは 0 である。

Weyl-automorphism は任意の有限分割に対し 17°) の仮定が成り立つから，エントロピー 0 である。

B-automorphism を考えよう。$K=\{a_j\}$，$p_j \equiv P(\omega_0=a_j)$，$\Omega$ の分割 $\alpha \equiv \{\{\omega;\ \omega_0=a_j\};\ j=1,2,\cdots\}$ とすれば，$\{T^n\alpha\}$ は互に独立で $\alpha_T=\varepsilon$ だから，(h) と 1°)，5°) より

$$h(T)=h(T,\alpha)=H(\alpha)=-\sum p_j \log p_j \ .$$

特に n-shift のエントロピーは $\log n$ である。

定常マルコフ連鎖のエントロピーも容易に計算できる。状態空間を $K=\{a_j\}$，せん移確率を $p_{ij} \equiv P\{\omega_n=a_j \mid \omega_{n-1}=a_i\}$，不変測度を $\pi_j \equiv P\{\omega_n = a_j\}$ とすれば，分割 α を前例と同じとし

$$\begin{aligned} h(T)&=h(T,\alpha)=H(\alpha\mid\alpha_T^-)=H(\alpha\mid T^{-1}\alpha) \\ &=-\sum_i\sum_j \pi_i\, p_{ij} \log p_{ij} \ . \end{aligned}$$

$\boxed{3.7}$ Kolmogorov-automorphism

3.7.1. K-flow の定義は $\boxed{0}$ で述べたが，分割の言葉で次ぎのように定義しなおす。

定義. 次ぎの 3 条件をみたす可測分割 ζ_0 が存在するとき T を K-automorphism という：

(K.1) $\zeta_n \equiv T^n\zeta_0 \geq \zeta_m\ ,\qquad n>m$

(K.2) $\bigvee_{\infty} \zeta_n = \varepsilon$

(K.3) $\bigwedge_{-\infty} \zeta_n = \nu$.

(K.3) により (K.1) はもっと強く

(K.1′) $\zeta_n > \zeta_m\ ,\qquad n>m$

となる。(K.1)，(K.2) をみたす可測分割は ε に限るとき，T は特異であるという。

まず容易にわかることは，

a) T が K-automorphism で ζ_0 が (K.1)～(K.3) をみたす可測分割であれば，Ω/ζ_0 には atom がない。

実際もし $C\in\zeta_0$ が $P(C)>0$ であるとすれば，任意の $n\geq 0$ に対し $T^{-n}C$ は ζ_0-set で，$P(T^{-n}C)=P(C)$ だから，$T^{-n_0}C=C$ となる n_0 がある。従って $C=T^{kn_0}C\in\zeta_{kn_0}$ がすべての k に対して成りたつことになって矛盾である。

一般の automorphism T を考え，有限分割 α に対し，商空間 Ω/α_T 上に導かれる factor-automorphism T_{α_T} を単に $T(\alpha)$ と書くことにしよう。これは $\boxed{1.3}$ に定義された X^* に相当する。$h(T)\geq h(T,\alpha)=h(T(\alpha))$ だから，

b) 次ぎの3命題は同等である：

(1°) $h(T)>0$，

(2°) 有限分割 α があって，$h(T(\alpha))>0$，

(3°) (K.1′) をみたす可測分割 ζ_0 が存在する．

換言すれば

c) 次ぎの4命題は同等である：

(1°) $h(T)=0$，

(2°) すべての有限分割 α に対し，$h(T(\alpha))=0$，

(3°) T は特異である，

(4°) すべての有限分割 α に対し，factor-automorphism $T(\alpha)$ は特異である．

non-trivial なすべての factor-automorphism のエントロピーが正であるとき，完全正 (completely positive) のエントロピーを持つという。明らかに，

d) 完全正のエントロピー（或いは0エントロピー）をもつ automorphism の non-trivial なすべての factor-automorphism は完全正のエントロピー（或いは0エントロピー）をもつ。

b), c) に対応して

e) 次ぎの3命題は同等である ([127])：

(1°) 完全正のエントロピーを持つ，

(2°) non-trivial なすべての有限分割 α に対し，$h(T(\alpha))>0$，

(3°) non-trivial なすべての有限分割 α に対し，

$T(\alpha)$ は K-automorphism.

K-automorphism に対しては上の (2°) 又は (3°) が容易に示せるので，

f) K-automorphism は完全正のエントロピーを持つ。

有限分割αに対し，$\alpha_{-\infty} \equiv \bigwedge_{-\infty} \bigvee_{k \le n} T^k \alpha$ とおく。Pinsker [127] は，分割 $\pi = \pi(T) \equiv \bigvee \{\alpha_{-\infty};\ \alpha:$ 有限分割$\}$ に対し，

g) factor-automorphism $T(\pi)$ はエントロピー 0 の最大の factor automorphism である，

ことを主張している。つまり $T(\pi)$ 自身エントロピー 0 であり，また任意のエントロピー 0 の factor-automorphism は $T(\pi)$ の factor-automorphism になる。従って

h) 完全正のエントロピーを持つことと $\pi(T) = \nu$ とは同等である。

Pinsker は次ぎのような問題を提起している。完全正のエントロピーをもつ最大の factor-automorphism は存在するか？　またエルゴード的な automorphism は，完全正及び 0 エントロピーの独立な factor-automorphism の直積に分解できないか？

Rohlin-Sinai [145] は f) の逆を示した．即ち，

i) automorphism T に対し次ぎの4性質を持つ可測分割ζが存在する：
(i) $T\zeta \geq \zeta$，(ii) $\bigvee T^k\zeta = \varepsilon$，(iii) $\bigwedge T^k\zeta = \pi(T)$，
(iv) $H(T\zeta \mid \zeta) = h(T)$．

この系として直ちに，

j) 完全正のエントロピーをもつ automorphism は K-automorphism である．従ってそれらのクラスは一致する。

k) T が K-automorphism であれば，T^{-1} も K-automorphism である。

d) と j) より，

l) K-automorphism の non-trivial な factor-automorphism（従ってまた homomorphic image）は K-automorphism である。

<u>3.7.2.</u>　automorphism T が $L^2(\mathbb{B})$ に定めるユニタリ作用素を U：$(Uf)(\omega) \equiv f(T^{-1}\omega)$，とする。

m) K-automorphism は $L^2(\mathbb{B}) \ominus 1$ で σ-Lebesgue スペクトルを持つ．

定義にある分割 ζ_n に対し，$H_n \equiv L^2(\mathbb{B}(\zeta_n)) \ominus 1$ とおけば，

(i) $H_n = U^n H_0 \supsetneqq H_m\ (n > m)$，(ii) $\overset{\infty}{\bigvee} H_n = L^2(\mathbb{B}(\zeta_n)) \ominus 1$，(iii) $\bigwedge_{-\infty} H_n = \{0\}$，

が成り立つ．$H_1 \ominus H_0$の完全正規直交系 $\{h_j ; j=1, 2, \cdots, \varkappa\}$ をとって来れば，$n \neq 0$ のとき $U^n h_j \in H_{n+1} \ominus H_n \perp H_1 \ominus H_0$ だから，

$$(U^n h_j, U^m h_i) = (U^{n-m} h_j, h_i) = 0, \quad i \neq j \text{ 又は, } i=j \text{ かつ } n \neq m.$$

従って，$i \neq j$ のとき $L(h_j) \equiv [\{U^n h_j ; -\infty < n < \infty\}$の張る線型閉部分空間$]$ $\perp L(h_i)$ であり，他方 $d\rho_j(\lambda) \equiv \|dE(\lambda) h_j\|^2$ に対し[12]，$n \neq 0$のとき，

$$\int e^{2\pi i n\lambda} d\rho_j(\lambda) = \int e^{2\pi i n\lambda} (dE(\lambda) h_j, h_j) = (U^n h_j, h_j) = 0 .$$

従って（Paley-Wienerの定理），$d\rho_j(\lambda) \sim d\lambda$，即ち各$L(h_j)$上で$U$は単純Lebesgueスペクトルを持つことがわかった。明らかに，

$$\sum_{j=1}^{\varkappa} \oplus L(h_j) \supset H_1 \ominus H_0$$

であり，$H_n = (H_n \ominus H_{n-1}) \oplus (H_{n-1} \ominus H_{n-2}) \oplus \cdots = U^{n-1}(H_1 \ominus H_0) \oplus U^{n-2}(H_1 \ominus H_0) \oplus \cdots$ だから，すべてのnに対し $\sum_{j=1}^{\varkappa} \oplus L(h_j) \supset H_n$，故に

$$\sum_{j=1}^{\varkappa} \oplus L(h_j) = L^2(\mathbb{B}) \ominus 1 .$$

即ちUは $L^2(\mathbb{B}) \ominus 1$ で multiplicity $\varkappa$ の一様Lebesgueスペクトルを持つ。

次ぎに $\varkappa = \infty$ を示そう。$\zeta_1 > \zeta_0$ だから $A \in \mathbb{B}(\zeta_0)$，$P(A) > 0$ があって，Aに含まれる $C \in \zeta_0$ は ζ_1 で本当に細分される（即ち $B \in \mathbb{B}(\zeta_1)$ があって $0 < P(B \cap C \mid \zeta_0 ; C) < 1$）。$\chi(\omega)$ を次ぎの3条件をみたす函数とする（存在は明らかであろう）．

(i) $\mathbb{B}(\zeta_1)$-可測，

(ii) a.e. $C \in \zeta_0$ に対し，$E\{\chi(\omega) \mid \zeta_0 ; C\} = 0$，

(iii) Aに含まれる $C \in \zeta_0$ に対し，$E\{|\chi(\omega)|^2 \mid \zeta_0 ; C\} = 1$.

α）により，A^c で0であるような H_0 の正規直交系 $\{\psi_j\}_{j=1}^{\infty}$ が存在する。

$$\varphi_j(\omega) \equiv \chi(\omega) \psi_j(\omega) , \qquad j = 1, 2, \cdots$$

とおけば，φ_j は $\mathbb{B}(\zeta_1)$-可測であり，

$$(\varphi_i, \varphi_j) = E\{\psi_i \chi \overline{\psi}_j \overline{\chi}\} = \{\psi_i \overline{\psi}_j |\chi|^2 ; A\}$$
$$= E\{\psi_i \overline{\psi}_j E\{|\chi|^2 \mid \zeta_0 ; C\} ; A\} = (\psi_i, \psi_j) = \delta_{ij} .$$

任意の $\varphi \in H_0$ に対し，

---

12) $\{E(\lambda)\}$ は U に対応する単位の分解．

$$(\varphi_j, \varphi) = E\{\psi_j \chi \overline{\varphi}\} = E\{\psi_j \overline{\varphi} E\{\chi | \zeta_0\}\} = 0.$$

故に $\{\varphi_j\}_{j=1}^{\infty}$ は $H_1 \ominus H_0$ の正規直交系であり，$n = \infty$ がわかった。

系として次ぎのことが直ちにわかる．

n) B-automorphism は $L^2(\mathbb{B}) \ominus 1$ で σ-Lebesgue スペクトルをもつ。

同様にして

o) $h(T) > 0$ であれば，T は $L^2(\mathbb{B}) \ominus L^2(\mathbb{B}(\pi))$ において σ-Lebesgue スペクトルをもつ。

p) 純粋離散スペクトル，特異スペクトル，有限 multiplicity スペクトルを持つ automorphism のエントロピーは 0 である。

$\boxed{3.8}$ metrical invariant としてのエントロピー；分類の問題

3.8.1. 　2つの flow が同値であることよりそれらのスペクトル同値が出ることは，確率空間の isomorphism が，対応する L^2-空間の isomorphism を導くことから明らかである．逆の問題，即ち "spectral type が等しければ metrical type も等しいか" という問題は，離散スペクトルの場合に肯定的に解かれた（[71]参照）だけで，最近まで未解決であった．

1958 年に Kolmogorov [87]は，新しい metrical invariant としてエントロピーを導入することによって，この問題の答えは一般的には否定的であることを示した。B-automorphism はすべて σ-Lebesgue スペクトルを持つが，エントロピーの異なる，従って metrical type の違う B-automorphism が連続濃度存在する．一般的には metrical type による分類が spectral type による分類よりも細かいわけである．

しかしながら，例えば既に解かれている離散スペクトルの場合のように，σ-Lebesgue 以外の特定のスペクトルの型に限れば問題は未解決である．更にそれは，spectral type にどのような metrical invariant をつけ加えれば metrical type を決定できるか（即ち metrical invariant の完全系をなすか）という問題に関連する．

スペクトル同値であり，更にエントロピーが等しければ同値であるか（Fomin の予想），という問題は，3.8.2 に述べるように Adler によって，一般的には否定された。むしろ興味は，如何なるクラスに限れば Fomin の予想が肯

定されるか，という点にあるのであろう。Meshalkin は 3.8.3 で述べるように B-automorphism の特定なクラスに対して肯定されることを示した。これに関連して，Sinai [169] は弱同値という概念を導入し，B-automorphism に対してエントロピーが弱同値の invariant の完全系をなすことを示した（3.8.1 参照）。弱同値は同値とスペクトル同値の中間の概念である。

metrical invariant としては他に，一般化された固有値及び固有函数（quasi-proper value と quasi-proper function ともいう）が提案され，いくらかの結果が得られている（[61], [5] 参照）が，本書の目的からやゝはずれると思われるので省略する。その他混合性の位数が提案されている（[140], [95]）。

flow の同値の問題は，分類の問題それ自体として重要であるばかりでなく，更に定常過程の表現の問題とも関連して，metrical type の等しい，より典型的な flow を見出すことや，（それよりも更に重要と思われるが）同値を与える isomorphism を具体的に求めることなどがある。

3.8.2. Fomin の予想の反例について，Adler [7] に従って述べる。普通の Lebesgue 測度をもつ区間 $[0, 1)$ を I で表わす。次ぎのような I 及び直積測度空間 $I\times I$，$I\times I\times I$ 上の automorphism を考えよう：

$$T_\theta(x) \equiv x+\theta$$
$$T_{\theta,\varphi}(x,y) \equiv (x+\theta,\ y+\varphi(x))$$
$$T_{\theta,\varphi,S}(x,y,z) \equiv (x+\theta,\ y+\varphi(x),\ Sz)$$

こゝに θ は実数，φ は可測函数，S は I 上の automorphism とし，＋はすべて mod 1 で考える。θ が無理数で $\varphi(x)=\varphi_n(x)\equiv nx$ であれば，T_{θ,φ_n} はエルゴード的になる。$|n|\neq|m|$ のとき，T_{θ,φ_n} と T_{θ,φ_m} はスペクトル同値であるが同値ではない（[13]）。所が（3.6）式と 3.6.7 に述べたことより，$h(T_{\theta,\varphi_n})=h(T_\theta)=0$ となって，これは反例になる。

エントロピー正の反例が $T_{\theta,\varphi,S}$ によって与えられる。S を 2-shift の I への isomorphic な像とする。$T_{\theta,\varphi,S}=T_{\theta,\varphi}\times S$ だから，$\boxed{3.6}$ 11°) と 3.6.7 により，すべての n に対し

$$h(T_{\theta,\varphi_n,S})=h(T_{\theta,\varphi_n})+h(S)=h(S)=\log 2$$

であり，更に $T_{\theta,\varphi_n,S}$ と $T_{\theta,\varphi_m,S}$ はスペクトル同値となる。$T_{\theta,\varphi_n,S}$ と $T_{\theta,\varphi_m,S}$ が $|n|\neq|m|$ のとき同値でないことを示す。そのためにそれらが同値であると仮定して矛盾を出す。$RT_{\theta,\varphi_m,S}R^{-1}=T_{\theta,\varphi_n,S}$ となる automorphism:

$$R(x,y,z)\equiv(f(x,y,z),\ g(x,y,z),\ h(x,y,z))$$

があるとする．次ぎの関係を得る，

(i) $f(x,y,z)+\theta=f(x+\theta,\ y+mx,\ Sz)$,

(ii) $g(x,y,z)+nf(x,y,z)=g(x+\theta,\ y+mx,\ Sz)$

(i)により，$e^{2\pi if}$ は $T_{\theta,\varphi_m,S}$ の固有値θに対応する固有函数であることがわかる。所が $F(x,y,z)=e^{2\pi ix}$ も同様である．$T_{\theta,\varphi_m,S}$ はエルゴード的(θ:無理数)従って固有値は単純であるから，

$$f(x,y,z)=x+u$$

となる $u\in I$ が存在する。

函数gを定めるためにFourier級数に展開する。

$$e^{2\pi ig(x,y,z)}\sim\sum_{p,q}g_{p,q}(z)e^{2\pi i(px+qy)}$$

$$g_{p,q}(z)=\iint_{II}e^{2\pi i[g(x,y,z)-px-qy]}dxdy$$

$$g_{p,q}(Sz)=\iint_{II}e^{2\pi i[g(x,y,Sz)-px-qy]}dxdy$$

において T_{θ,φ_m} で変数変換すれば

$$g_{p,q}(Sz)=\iint_{II}e^{2\pi i[g(x+\theta,\ y+mx,\ Sz)-p(x+\theta)-q(y+mx)]}dxdy$$

$$=e^{2\pi i(nu-p\theta)}\iint_{II}e^{2\pi i[g(x,y,z)+(n-p-mq)x-qy]}dxdy$$

$$=e^{2\pi i(nu-p\theta)}g_{p-(n-mq),q}(z).$$

故にすべてのkに対し

$$(3.7)\qquad |g_{p-k(n-mq),q}(z)|=|g_{p,q}(S^kz)|.$$

もし $g_{p,q}(z)=0$ (a.e)でないとすれば，$\delta>0$ があって $A_{p,q}\equiv\{z;|g_{p,q}(z)|\geq\delta\}$，$P(A_{p,q})>0$ となる。エルゴード性により，ほとんどすべての $z\in A_{p,q}$ に対し無限個のkで $|g_{p,q}(S^kz)|\geq\delta$ となり，(3.7)により $n\neq mq$ であれば Fourier係数の無限個が絶対値δ以上となる．故に $q\neq n/m$ であれば $g_{p,q}(z)=0$ (a.e.) となる．

$q=\frac{n}{m}$ に対し

$$g_{p,\frac{n}{m}}(Sz)=e^{2\pi i(nu-p\theta)}g_{p,\frac{n}{m}}(z).$$

即ち $g_{p,\frac{n}{m}}$ はSの固有函数である．2-shift は強混合で，従って固有函数は定数だけだから，

$$g_{p,\frac{n}{m}}(z) \equiv g_p : \text{定数}(a.e.).$$

Fourier 展開は

$$e^{2\pi i g(x,y,z)} \sim \sum_p g_p e^{2\pi i(px+\frac{n}{m}y)}$$
$$\sim e^{2\pi i(\psi(x)+\frac{n}{m}y)}$$

となる．こゝに $\psi(x)$ は或る可測函数である．

Fourier 展開の一意性により結局，

$$g(x,y,z) = \psi(x) + \frac{n}{m}y$$

となる．Rが保測であるためには，$n/m = \pm 1$ でなければならない．故に $|n| \neq |m|$ のとき，$T_{\theta,\varphi_n,S}$ と $T_{\theta,\varphi_m,S}$ とは同値ではない。

3.8.3. B-automorphism で，state $K=\{a_j\}$ が有限集合，しかも $p_j \equiv P_r(a_j) = P\{\omega;\ \omega_0 = a_j\}$ があるpに対し $p_j = p^{-n_j}$ $(j=1,2,\cdots)$ の形のものだけを考える．Meshalkin [102]は，このような２つのB-automorphism が，もし(i) 同じpをもつ (ii) エントロピーが等しい（即ち，$\sum_j n_j p^{-n_j}$ が等しい）ならば（state の個数は異なってもよい），それらは同値であることを示した。例えば $\{\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\}$ で与えられる B-automorphism と，$\{\frac{1}{2}, 1/8, 1/8, 1/8, 1/8\}$で与えられるそれとは同値である．

3.9 最近の研究

この節では目につく最近の結果を２つほど紹介する．3.9.1 は Sinai [169]，3.9.2 は Rohlin [143] による。

3.9.1. automorphism T_1 と T_2 が互に他の homomorphic image であるとき，それらの弱同値であるという．

T_i $(i=1,2)$ のスペクトル測度を $d\rho_i(\lambda)$，multiplicity函数を $m_i(\lambda)$ とするとき，T_2 が T_1 の homomorphic image であれば，$d\rho_2$ は $d\rho_1$ に関し絶対連続であり，更に $m_2(\lambda) \leq m_1(\lambda)$ a.e.λ $(d\rho_1)$である。従って

a) T_1 とT_2 が弱同値であれば，それらは同じ spectral type を持つ．

更にこのときエントロピーが等しい，$h(T_1)=h(T_2)$．

エントロピーの部分は 3.6.2，7°）による。

以下エントロピーが有限な automorphism のみを考える。Sinai の基本的な結果は次ぎの主張である．

b）Tを任意のエルゴード的な automorphism とする。$h(T)\geq h(T_1)$ をみたす任意の B-automorphism T_1 は T の或る factor-automorphism に同値である．

この系として次ぎの諸結果が得られる。

c）エントロピーの等しい B-automorphisms は弱同値である。

d）正のエントロピーをもつエルゴード的な automorphism は，同じエントロピーをもつ K-automorphism であるような factor-automorphism をもつ。

e）任意のエルゴード的な automorphism に対し，分割αが存在して，$H(\alpha)=h(T,\alpha)=h(T)$。

3.9.2. 可測分割ζ，$T\zeta=\zeta$ に対し，factor-automorphism T_ζ が principal というのは，すべての可測分割 $\eta\geq\zeta$，$T\eta\geqq\eta$ が $T\eta=\eta$ となることである．

a）もし T_ζ が principle factor-automorphism であれば $h(T_\zeta)=h(T)$．$h(T)<+\infty$ であれば逆も成り立つ．

Rohlin はエントロピーを公理的に定義しようと，エントロピーの次ぎの4性質を公理に選んでいる：

(1°) SがTの homomorphic image であれば，$h(S)\leq h(T)$．

(2°) $h(S\times T)=h(S)\times h(T)$．

(3°) Tが 2-shift であれば，$h(T)=\log 2$．

(4°) SがTの principal factor-automorphism であれば $h(S)=h(T)$。

エントロピーが上記の諸性質を持つことは，今迄に述べた。

b）エルゴード的 automorphism に対し，エントロピーhは性質(1°)～(4°)をもつ唯一の非負函数である。

証明のために，h' を (1°)～(4°) をみたす任意の非負函数として $h'=h$ を示す。エントロピー有限な B-automorphism の全体を $\mathcal{B}$ とすれば，3.9.1, c) と (1°) により，

$$S, T \in \mathcal{L}, \quad h(S) = h(T) \Rightarrow h'(S) = h'(T).$$

従って $\mathcal{L}$ においては h' は h の函数である．h は $\mathcal{L}$ を区間 $[0, \infty)$ にうつすので，$[0, \infty)$ 上の非負函数 φ があって

$$h'(T) = \varphi(h(T)), \qquad T \in \mathcal{L}.$$

3.9.1，b) と (1°) により φ は増大函数であることがわかる．$s, t \in [0, \infty)$，$S, T \in \mathcal{L}$，$h(S) = s$, $h(T) = t$ とすれば，$S \times T \in \mathcal{L}$ だから 3.6.2，11°) と公理 (2°) より，

$$\varphi(s+t) = \varphi(h(S)+h(T)) = \varphi(h(S\times T)) = h'(S\times T)$$
$$= h'(S) + h'(T) = \varphi(h(S)) + \varphi(h(T)) = \varphi(s) + \varphi(t).$$

T が 2-$shift$ であれば $h(T) = \log 2$ だから，

$$\varphi(\log 2) = \varphi(h(T)) = h'(T) = \log 2.$$

函数 φ は線型，単調増大で $\varphi(\log 2) = \log 2$ だから $\varphi(t) = t$ となる。従って $\mathcal{L}$ 上では $h' = h$ である．

Tがエントロピー有限なエルゴード的 $automorphism$ であれば，3.8.1，b) により $factor$-$automorphism$ $S \in \mathcal{L}$ があって $h(S) = h(T)$，a) により S は $principal$ だから公理 (4°) によって

$$h'(T) = h'(S) = h(S) = h(T).$$

$h(T) = \infty$ のときは，エントロピーがどれだけでも大きいエルゴード的な $factor$-$automorphism$ の存在を使って $h'(T) = \infty$ が示せる．結局 $h' = h$ がわかった。

エルゴード的という制限を除き，すべての $automorphism$ に適するエントロピーの公理としては，上記 (1°) ～ (4°) に次ぎの公理をつけ加えればよい。

(5°) もし分割 $\zeta = \{C\}$ が $automorphism$ T を成分 T_C に分解するならば，

$$h(T) = \int_{\Omega/\zeta} h(T_C)\, dP_\zeta .$$

4 その他

4.1 中心極限定理への応用

確率論における極限定理を，エントロピーが増大する方向への移行として把握するのは，興味あることと思われる．Lindebergの条件の下での中心極限定理の情報理論的な証明を，Linnik[96]に従って紹介する．

一次元確率変数 X の分布が次ぎの条件をみたす連続な密度函数 $p(x)$ をもつとする．

$$(4.1)\qquad \sup_x p(x) < \infty,\quad EX = \int_{-\infty}^{\infty} x\,p(x) = 0,\quad DX = \int_{-\infty}^{\infty} x^2 p(x)\,dx < \infty .$$

このような X のエントロピーとして，こゝでは次ぎの量を考える：

$$\begin{aligned} H^1(X) &\equiv H^0(X) - \frac{1}{2}\log DX \\ &= -\left\{\int_{-\infty}^{\infty} p(x)\log p(x)\,dx + \frac{1}{2}\log\int_{-\infty}^{\infty} x^2 p(x)\,dx\right\} . \end{aligned}$$

H^0 は 0.2 で考えた differential entropy であり，X が $N(0,\sigma^2)$ に従うときは $H^0(X) = \log(\sigma\sqrt{2\pi e})$ であった。従って上の $H^1(X)$ は，X の differential entropy と，X と同じ分散の正規変数のそれとの差（絶対定数を除いて）である．

問題は「$X_1, X_2, \cdots, X_m$ がそれぞれ分布 $F_j(x)$ をもつ独立確率変数で，$EX_j = 0$，$0 < DX_j \equiv b_j < \infty$，$B_m \equiv b_1 + \cdots + b_m$ とする．このとき，任意の $\theta > 0$，$\delta > 0$ に対し $m_0(\theta, \delta)$ があって

$$\frac{1}{B_m}\sum_{j=1}^{m}\int_{|x|>\theta|B_m|} x^2\,dF_j(x) < \delta ,\qquad m > m_0(\theta,\delta)$$

（Lindebergの条件）であれば，任意の $\varepsilon > 0$ に対し

$$|P(S_m/\sqrt{B_m} < x) - G(x)| < \varepsilon ,\qquad m > m_1(\varepsilon)$$

となる $m_1(\varepsilon)$ が存在する．たゞし，$S_m \equiv X_1 + \cdots + X_m$，$G(x)$ は $N(0,1)$ の分布函数.」を示すことである．H^1 が定義できるように X_j を変形する．(1°) truncate する： 大きい T_0 を固定して，

$$X_j' \equiv \begin{cases} X_j , & |X_j| \le T_0\sqrt{B_j} , \\ 0 , & |X_j| > T_0\sqrt{B_j} . \end{cases}$$

(2°) Y_j' を正規変数で $EY_j' = -EX_j'$，分散 $\delta_0^2 b_j'$（δ_0 は小さい数，$b_j' = DX_j'$）

なるものとし，$X_j'' \equiv X_j' + Y_j'$ とおく。X_j'' は (4.1) をみたし，更に $m\to\infty$，$T_0\to\infty$，$\delta_0\to 0$ のとき $S_m''/\sqrt{B_m''}$ の分布は $S_m/\sqrt{B_m}$ の分布に近づくので，以下 X_j'' を X_j として考える．*Lindeberg* の条件を変形しておく．$\delta_1 = \delta_1(\delta)$ を与えたとき

$$\int_{|x|>\theta\sqrt{B_n}} x^2 dF_j(x) < \delta_1 b_j ,$$

が成り立てば，X_j は $j\le n$ について個別に *Lindeberg* の条件をみたすという。

定理の証明の概略は次ぎの通りである。

$$H^1(S_{m+1}) - H^1(S_m) = \frac{\sigma_{m+1}^2}{2}\left(\int_{-\infty}^{\infty}\left(\frac{p_m'}{p_m}\right)^2 p_m\,dx - 1\right) + B\mu\sigma_{m+1}^2$$

こゝに $\sigma_{m+1} = \sqrt{b_{m+1}/B_m}$，$p_m$ は $S_m/\sqrt{B_m}$ の分布の密度，B は定数，X_{m+1} に対し個別に *Lindeberg* 条件が成り立っていれば $\mu=\delta_1$，成り立っていなければ $\mu=1$．こゝで

$$F_m \equiv \int_{-\infty}^{\infty}\left(\frac{p_m'}{p_m}\right)^2 p_m\,dx - 1$$

とおくと，$F_m\ge 0$ で，$F_m\to 0$ となれば，$p_m(x)\to(2\pi)^{-\frac{1}{2}}e^{-\frac{x^2}{2}}$ （一様）が証明されている。個別に *Lindeberg* 条件が成り立たないような X_j は無視可能であり，他方個別に *Lindeberg* 条件が成り立てば

$$H^1(S_{m+1}) - H^1(S_m) = F_m\frac{\sigma_{m+1}^4}{2} + \gamma\varepsilon_0\sigma_{m+1}^4 , \quad |\gamma|\le 1, \quad \varepsilon_0 : 小$$

となる。従って ① $H^1(S_{m+1}) - H^1(S_m) > 0$，即ちエントロピー増加か，或いは ② $F_m < 2\varepsilon_0$ かである。②の場合は上記のことより正規分布に収束する。①の場合にも，$\bar H^1 \equiv \sup\{H^1(S_m);\ m>M_0\}$，$G\equiv\log\sqrt{2\pi e}$ ($N(0,1)$のエントロピー) に対し，$|\bar H^1 - G| < \varepsilon_1(\varepsilon_0)$，($\varepsilon_1$ は 小) がいえて，結局 $N(0,1)$ に収束することがわかる。

$\boxed{4.2}$ <u>ε-entropy</u> ([89])

ε-*entropy* の定義．A を距離 ρ をもつ距離空間，E をその部分集合とするとき，$V(\subset A)$ が E の <u>ε-*net*</u> というのは，任意の $x\in E$ に対し，$\rho(x, y)\le\varepsilon$ をみたす $y\in V$ が存在することである．$N_\varepsilon(E)$ を E の ε-*cover*-

ing の集合の元の最小数，$N_\varepsilon^A(E)$ を E の ε-*net* のそれとし，$H_\varepsilon(E) \equiv \log N_\varepsilon(E)$ を E の ε-*entropy*，$H_\varepsilon^A(E) \equiv \log N_\varepsilon^A(E)$ を E の A に関する ε-*entropy* という。A の任意の全有界集合 E に対し，次ぎの不等式が成り立つ。

$$N_\varepsilon(E) \leq N_\varepsilon^A(E) \leq N_\varepsilon^E(E) \leq N_{\frac{\varepsilon}{2}}(E),$$

$$H_\varepsilon(E) \leq H_\varepsilon^A(E) \leq H_\varepsilon^E(E) \leq H_{\frac{\varepsilon}{2}}(E).$$

ε-*entropy* は解析学において有効に用いられている([89], [103], [182] etc 参照)。

伝達機構 $\{Q, V\}$ において，$\{Q, V\}$ によって結ばれる変数 $\{Y, \tilde{Y}\}$ は $A \times A$ の値をとるとし，$\{V\}$ を $P_{Y,\tilde{Y}}\{\rho(y, \tilde{y}) \leq \frac{\varepsilon}{2}\} = 1$ (2.1 参照)，E を A の任意の全有界部分集合とすれば，伝達機構の通過能力に関し

$$C(Q, V) \geq H_\varepsilon^E(E)$$

が成り立つ．通信 $[W]$ については次式が成りたつ。

$$H(W) \leq H_\varepsilon(E).$$

$\boxed{4.3}$ 統計量と情報量

4.3.1.　$(A, \mathbb{B}_A, \mu)$ は測度空間，確率変数 X は A の値をとるとする。X の分布を P_X であらわし，P_X としては μ に関して絶対連続なものだけを考える。その密度函数を $p(x)$，$P_X(E) = \int_E p(x)\,\mu(dx)$，であらわす．今仮説 H_i を $P_X = P^{(i)}$ $(i = 1, 2)$ とする．X を観測することによって得られる " H_1 を H_2 に対して判別する" 情報の量を，

$$I(p_1 : p_2 ; X) = \int \log \frac{p_1(x)}{p_2(x)}\, dP^{(1)}(x) = \int p_1(x) \log \frac{p_1(x)}{p_2(x)}\, \mu(dx)$$

で定義し，*Kullback-Leibler* の平均情報量という。

$Y = T(X)$ を X の統計量とし，変換 $y = T(x)$ によって，$(A, \mathbb{B}_A, P^{(i)})$ が $(B, \mathbb{B}_B, Q^{(i)})$ に移されるとする。但し，$Q^{(i)}(G) = P^{(i)}(T^{-1}(G))$ で，$\nu(G) = \mu(T^{-1}(G))$，$Q^{(i)}(G) = \int_G q_i(y)\,\nu(dy)$ とする．このとき

$$I(p_1 : p_2 ; X) \geq I(q_1 : q_2 ; Y)$$

が成り立ち，特に等号が成立するのは，μ-測度 0 を除いて

$$\frac{p_1(x)}{p_2(x)} = \frac{q_1(T(x))}{q_2(T(x))}$$

の時に限る．即ち T が十分統計量の時である（[199]）．

4.3.2. $p(x,\theta)$ は k 次元パラメーター $\theta=(\theta_1,\theta_2,\cdots,\theta_k)$ を持つ確率密度（μ に関する）とする．今 1 つの値 θ とその近傍（開凸集合）の点 $\theta+\Delta\theta$ を考え，$p_1(x)=p(x,\theta)$，$p_2(x)=p(x,\theta+\Delta\theta)$ とすれば，

$$I(p_1:p_2;X) = -\int p(x,\theta)\,\Delta\log p(x,\theta)\,\mu(dx)$$

但し，$\Delta\log p(x,\theta) \equiv \log p(x,\theta+\Delta\theta) - \log p(x,\theta)$．こゝで，$p$ に次ぎの正則条件を仮定する：

1° μ-測度 0 を除いて，$\dfrac{\partial\log p}{\partial\theta_\alpha}$，$\dfrac{\partial^2\log p}{\partial\theta_\alpha\partial\theta_\beta}$，$\dfrac{\partial^3\log p}{\partial\theta_\alpha\partial\theta_\beta\partial\theta_\gamma}$ がすべての $\alpha,\beta,\gamma\,(=1,2,\cdots,k)$ と，すべての $\theta'=(\theta_1',\cdots,\theta_k')\in\Theta\equiv\{\bar\theta=(\bar\theta_1,\cdots,\bar\theta_k);\ \theta_\alpha<\bar\theta_\alpha<\theta_\alpha+\Delta\theta_\alpha,\ \alpha=1,\cdots,k\}$ に対して存在する．

2° 任意の $\theta'\in\Theta$ に対して，$\left|\dfrac{\partial p}{\partial\theta_\alpha}\right|<F(x)$，$\left|\dfrac{\partial^2 p}{\partial\theta_\alpha\partial\theta_\beta}\right|<G(x)$，$\left|\dfrac{\partial^3\log p}{\partial\theta_\alpha\partial\theta_\beta\partial\theta_\gamma}\right|<H(x)$ がすべての α,β,γ に対して成り立つ．但し $F,G\in L^1(\mu)$，$\int H(x)p(x,\theta)\,\mu(dx)<M<+\infty$（$M$ は θ に無関係）．

3° $\displaystyle\int\frac{\partial p}{\partial\theta_\alpha}\mu(dx)=0$，$\displaystyle\int\frac{\partial^2 p}{\partial\theta_\alpha\partial\theta_\beta}\mu(dx)=0$，$\alpha,\beta=1,2,\cdots,k$．

これら 1°，2°，3° がみたされているとき，$\Delta\theta$ の 2 次の項までの近似として

$$I(p_1;p_2;X) \approx \frac{1}{2}\sum_{\alpha=1}^{k}\sum_{\beta=1}^{k} g_{\alpha\beta}\,\Delta\theta_\alpha\,\Delta\theta_\beta$$

を得る．但し

$$g_{\alpha\beta} = \int p(x,\theta)\left(\frac{1}{p(x,\theta)}\cdot\frac{\partial p(x,\theta)}{\partial\theta_\alpha}\right)\left(\frac{1}{p(x,\theta)}\frac{\partial p(x,\theta)}{\partial\theta_\beta}\right)\mu(dx).$$

$G\equiv(g_{\alpha\beta})$ が Fisher の情報行列と呼ばれるものである．

4.3.3. $I(p_1:p_2;X)$ を最小にする p_1 とその最小値を求める．但し，この p_1 は，或る統計量 $T(X)$ に対して

$$\int T(x)\,p_1(x)\,\mu(dx) = \theta$$

となるものの中で求めることにする．$I(p_1 : p_2 ; X)$ を最小にする $p_1 = p^*$ は次式で与えられる．

$$p^*(x) = e^{\tau T(x)} p_2(x) / M(\tau), \qquad a.e.(\mu),$$

こゝに $M(\tau)$ は正規化の定数 $(M(\tau) = \int e^{\tau T(x)} p_2(x) \mu(dx))$, τ は $M(\tau)$ が存在するような実数である．そしてこの時，最小値は

$$I(p^* : p_2 ; X) = \theta\tau - \log M(\tau)$$

である（[200]）．

あ　と　が　き

情報理論は，誕生は新しく，その後の発展が余りに急速な為，「手引き」を書く著者達自身，謂わば「手引かれる？」立場にあった。そのため，応用と理論に，多岐にわたっている情報理論ではあるが，私達はもっぱら，確率論として内容のある，乃至興味ある部分のみの紹介にとどまった。それも，例えば具体的な *coding* や *Channel*，或は古典的力学系との問題等，紙数の制限もあったが，筆が及ばなかった点の多々あることをわびねばならない。

おもうに，*Khintchin* が，「今日までエントロピー理論の多少ともまとまった説明は，情報伝達の諸問題に関連ある専門的論文や *monograph* でしか見ることができないが，事実上はすでに確率論の重要な，内容豊富な１章にまで発展している。にもかゝわらず一般理論の見通しのもとでの説明は，今日までなされていない」，と述べ，その最初の試みとして発表した１９５３年の論文以来１０年，この小著の作成を終るに当って，今や情報理論は確率論として，*Khintchin* のいう完全なる一般理論の見通しの上に立っている，ということができる。

「手引き」の作成にあたったのは，押川，小野山（編集責任者），久保，佐藤（担），佐藤（由），瀬口，十時，飛田，藤井であるが，文責は，飛田，瀬口，小野山，十時にあることをお断りしておく。

索　　　引

文　　献

〔1〕 Abramov, L. M.; The entropy of the derived automorphism. Dokl. Akad. Nauk, 128 (1959), 647-650.

〔2〕 ――――――――; On the entropy of a flow. Dokl. Akad. Nauk, 128 (1959), 873-876.

〔3〕 ――――――――; The entropy of a automorphism of a solenoidal group. Theory Prob. Appl., IV, 3 (1959), 249-254.

〔4〕 ――――――――; Some metrical theory of dynamical systems. MGY, 1960.

〔5〕 ――――――――; Metrical automorphisms with quasi-discrete spectra. Izv. Akad. Nauk, 26 (1962), 513-530.

〔6〕 ―――――――― and V. A. Rohlin; Entropy of the skew product of transformtions with invariant measure. Vestnik LGY, No7, 2 (1962).

〔7〕 Adler, R. L.; On the conjecture of Fomin. Proc. Amer. Math. Soc., 13, 3 (1962), 433-436.

〔8〕 Alekseev, V. G.; Conditions for the orthogonlity and equivalence of Gaussian measures in functional space. Dokl. Akad. Nauk, 147 (1962), 751-754. Soviet Math. Dokl. 3, 6 (1962), 1680-1684.

〔9〕 Ambrose, W.; Representation of ergodic flow. Ann. Math., 42 (1941), 723-739.

〔10〕 ―――――――― and S. Kakutani; Structure and continuity of measurable flow. Duke Math. J., 9 (1942), 25-42.

〔11〕 ――――――――, P. R. Halmos and S. Kakutani; The decomposition of measures II. Duke Math. J., 9 (1942), 43-47.

[12] Anosov, D. V.; Roughness of geodesic flows on compact Riemannian manifolds of negative curvature. Soviet Math. Dokl., 3, 4 (1962), 1068-1070.

[13] Anzai, H.; Ergodic skew product transformations on the torus. Osaka Math. J., 3 (1951), 83-99.

[14] Basharin, G. P.; On a statistical estimate for the entropy of a sequence of independent random variables. Theory Prob. Appl., 4 (1959), 361-364.

[15] Billingsley, P.; Hausdorff dimension in probability theory. Ill. J. of Math., 4, 2 (1959), 187-209.

[16] Birch, J. J.; Approximations for the entropy for functions of Markov chains. Ann. Math. Stat., 33, 3 (1962), 930-938.

[17] Blackwell, D. L., L. Breiman and A. J. Thomasian; Proof of shannon's transmission theorem for finite state indecomposable channels. Ann. Math Stat., 29, 4 (1958), 1209-1220.

[18] ——————, —————— and ——————; The capacity of a class of channels. Ann. Math. Stat., 30, 4 (1959), 1229-1241.

[19] Breiman, L.; The individual ergodic theorem of information theory. Ann. Math. Stat., 31(1960), 809-810. Correction, Ann. Math. Stat., 31(1960), 809-810.

[20] ——————; On achieving channel capacity in finite memory-channels. to appear in Ill. J. Math.

[21] Brillouin, L.; Science and information theory.

Academic press, Inc., New York, 1956.

[22] Brown, T. A.; Entropy and conjugacy. Ann. Math. Stat., 34, 1(1963), 226-232.

[23] Chaing Tse-Pei; A remark on the definition of the quantit, of information. Teor. Veroyat. Primenen., 3(1958), 99-103.

[24] Chung, K. L.; A note on the ergodic theorem of information theory. Ann. Math. Stat., 32, 2 (1961), 612-614.

[25] Dobrushin, R. L.; On formulation Shannon's fundamental theorem. Theory Prob. Appl., 2 (1957), 480-482.

[26] ——————; A statistical problem arising in the theory of detection of signals in the presence of noise in a multi-channel system and leading to stable distribution laws. Theory Prob. Appl., 3(1958), 161-173.

[27] ——————; Transmission of information in channels with feedback. Theory Prob. Appl., 3(1958), 395-412.

[28] ——————; A simplified method of experimentally evaluating the entropy of a stationary sequence. Theory Prob. Appl., 3(1958), 428-430.

[29] ——————; General statement of Shannon's main theorem in the information theory. Dokl. Akad. Nauk, 126(1959), 474-477.

[30] ——————; A general formulation of Shannon's fundamental theorem in the information theory. Uspehi Mat. Nauk, 14, 6(1959), 3-104. Arbeiten zur Inform. IV, VEB Deutscher

Verlag der Wissenschaften, Berlin, 1963.

[31] Dobrushin, R. L.; The asymptotic behavior of the transmission through a channel without memory with a symmetric matrix of transition probabilities. Dokl. Akad. Nauk, 133(1960), 265-268.

[32] ————————; Limit approach under the signs of information and entropy. Theory Prob. Appl., 5 (1960), 29-37.

[33] Doob, J. L.; Stochastic processes. Wiley New York, 1953.

[34] Ehrenfest, P. und T. Ehrenfest; Über zwei bekannte Einwände gegen das Boltzmannsche H-theorem. Z. Physik, 8(1906), 311-314.

[35] Elias, P.; Coding for noisy channels. IRE Convention Record, 4(1955), 37-44.

[36] ————————; Coding for two noisy channels. Proc. of London Symp. on Inform. Theory, London 1955.

[37] Erohin, V.; ε-entropy of a discrete random variable. Teor. Veryat. Primenen., 3, 1(1958), 103-106.

[38] Fadiev, D. A.; On the notion of entropy of a finite probability space. Uspehi Mat. Nauk, 11, 1(67), (1956), 227-231. Arbeiten zur Inform. I, Berlin, 1957.

[39] Fano, R. M.; Statistical theory of communication. Notes on a course given at M. I. T., 1952, 1954.

[40] Feinstein, A.; A new basic thorem of information theory. Trans. IRE, Sect. Inform. Theory, PGIT-4 (1954), 2-22.

[41] Feinstein, A.; Error bounds in noisy channels without memory. Trans. IRE, PGIT, (1955), 13-14.

[42] ——————; Foundations of information theory. McGraw-Hill, 1958.

[43] ——————; On the coding theorem and its converse for finitememory channels. Inform. and Control, 2, 1 (1959), 25-44.

[44] Feldman, J.; Equivalence and perpendicularity of Gaussian processes. Pacific J. Math., 8 (1958), 699-708.

[45] Fortet, R.; Hypothesis testing and estimation for Laplacian functions. Proc. 4th Berkeley Symp., I (1962), 289-305.

[46] Gelfand, I. M. and S. V. Fomin; Geodesic flows on the manifold with a constant negative curvature. Uspehi Mat. Nauk, 7, 1 (1952), 118-137.

[47] ——————, A. N. Kolmogorov and A. M. Yaglom; The quantity of information and entropy for continuous distributions. Trudy III-20 Vsesoyozn. Mat. Sbezda, (1956, 1958), 300-320.

[48] ——————, —————— and ——————; On the general definition of amount of information. Dokl. Akad. Nauk, 111 (1956), 745-748. Arbeiten zur Inform. II, Berlin, 1958.

[49] —————— and A. M. Yaglom; Calculation of the amount of information about a random function contained in another such function. Uspehi Mat. Nauk, 12, 1 (1957), 3-52. Amer. Math. Soc. Transl. (2) 12, 199-246. Arbeiten zur Inform. II, Berlin, 1958.

[50] Genis, A. L.; Metric properties of the endomorphisms of the n-dimensional torus. Soviet Math. Dokl., 2, 3 (1961), 750-752.

[51] Girsanov, I. V.; On the spectrum of dynamical systems which is generated by stationary Gaussian processes. Dokl. Akad. Nauk, 119, 5 (1958), 851-853.

[52] Goldman, S.; Information theory. Princeton-Hall, New-York, 1953.

[53] Gurevic, B. M.; The entropy of a horocycle flow. Soviet Math. Dokl., 2, 1 (1961), 124-126.

[54] Hájek, J.; Über eine Eigenschaft von Normalverteilungen zufälliger Prozesse, Dokl. Akad. Nauk, 112 (1957), 16-19.

[55] Halmos, P. R.; The decomposition of measures. Duke Math. J., 8 (1941), 386-392.

[56] ——————; Square roots of measure preserving transformations. Amer, J. Math., 64 (1942), 153.

[57] ——————; On automorphisms of compact groups. Bull. Amer. Math. Soc., 49, 8 (1943), 619-624.

[58] ——————; Approximation theories for measure preserving transformations. Trans. Amer. Math. Soc., 55, 1 (1944).

[59] ——————; Measurable transformations. Bull. Amer. Math. Soc., 55, 11 (1949), 1015-1034.

[60] ——————; Measure theory. D. van Nostrand Comp. Inc., Toronto-New York-London, 1951 (2-nd printing).

[61] ——————; Lectures on ergodic theory. Math. Soc. of Japan, Tokyo, 1956.

[62] Halmos, P. R.; Entropy in ergodic theory Univ. of Chicago, 1956.

[63] ——————; Recent progress in ergodic theory. Bull. Amer. Math. Soc., 67 (1961), 70-80.

[64] —————— and J. von Neumann; Operator methods in classical mechanics, II. Ann. Math., 43 (1942), 332-350.

[65] Hamming, R. W.; Error detecting and error correcting codes. Bell System Tech. J., 29 (1950), 147-160.

[66] Heldlund, G. A.; The dynamics of geodesic flows, Bull. Amer. Math. Soc., 45 (1939), 241-246.

[67] Hopf, E.; Complete transitivity and the ergodic principle. Proc. Nat. Acad. Sci. USA, 18 (1932), 204.

[68] ————; Ergodentheorie. Berlin, 1937.

[69] ————; Statistik der geodätischen Linien in Mannigfaltigkeiten negativer Krummung. Ber. Verh. Sachs. Akad. Wiss. Leipzig, 91 (1939), 261-304.

[70] Hu Kuo - Ting; On information quantities. Teor. Veroyat. Primenen., 7, 4 (1962), 447-455.

[71] 池田信行，飛田武幸，吉沢尚明； Flowの理論 (上).. Seminar on Prob., Vol. 12, (1962).

[72] Ito, K.; Spectral type of the shift transformation of differential processes with stationary increments. Trans. Amer. Math. Soc.. 81(1956), 253-263.

[73] Jacobs, K.; Die Übertragung des diskreter Informationen durch periodische und fast periodische Kanäle. Math. Ann., 137(1959), 125-135.

〔74〕 Jacobs, K.; Neuere Mathoden und Ergebnisse der Ergodenthorie, Springer, Berlin, 1960.

〔75〕 確率論セミナー; 確率論の手引|Vol.1 B1 Brown運動(上). (1962).

〔76〕 Kakutani, S.; Representation of measurable flows in Euclidean 3-space. Proc. Nat. Acad. Sci. USA, 23(1942), 16.

〔77〕 ————————; Determination of the spectrum of the flow of Brownian motion. Proc. Nat. Acad. Sci. USA, 36(1950), 319-323.

〔78〕 ————————; On ergodic theorem. Proc. Inter. Cong. Math. 2(1950), 128-142.

〔79〕 ————————; Spectral analysis of stationary Gaussian processes. Proc. 4-th Berkeley Symp. II (1961).

〔80〕 河田敬義; 確率論, 共立出版, 1948.

〔81〕 Khintchin, A. Ya.; The concept of entropy in the theory of probability. Uspehi Mat. Nauk, 8, 3 (1953), 3-20. Arbeiten znr Inform. I, Berlin, 1957. Math. Found. Inform. Theory, Dover Inc., New York, 1958.

〔82〕 ————————; On the fundamental theorems of information theory. Uspehi Mat. Nauk, II, 1 (1956), 17-75. Arbeiten zur Inform. I, Berlin, 1957. Math. Found. Inform. Theory, Dover Inc., New York, 1958.

〔83〕 喜安善市, 室賀三郎; 情報理論. 岩波応用数学講座.

〔84〕 Kolmogorov, A. N.; General theory of dynamical systems and classical mechanics. Proc. Inter. Cong. Math. (1954), 315-333.

〔85〕 ————————; To the shannon theory of information transmission in the continuous signal case. IRE. Trans. PGIT, Dec. (1956), 102-108.

[86] Kolmogorov, A. N.; Theorie der Nachrichtenübermittlung. Arbeiten zur Inform. I. Berlin. (1957).

[87] ——————————; A new metrical invariant of transient dynamical systems and automorphisms in Lebesgue spaces. Dokl. Akad. Nauk, 119, 5(1958). 861-864.

[88] ——————————; On the entropy to the unit time, as a metrical invariant of automorphisms. Dokl. Akad. Nauk. 124. 4(1959), 754-755.

[89] —————————— und W. M. Tichomirov; ε-Entropy und ε-Kapazität von Mengen in Funktionalräumen. Arbeiten zur Inform. III, Berlin, 1960. Uspehi Mat. Nauk, 14(1956), 3-86.

[90] Kotz, S.; Exponential bounds for the probability of error in discrete memoryless channels. Unpublished thesis. Cornell Univ., 1960.

[91] Kraft, L. K.; A device for quantizing grouping, and coding amplitude modulated pulses. M.S. Thesis, Elect. Engin. Depart. M.I.T., (1949).

[92] Kullback, S.; Information theory and statistics. Wiley, New York, 1959.

[93] —————— and R. A. Leibler; On information and Sufficiency. Ann. Math. Stat., 22(1951), 79-86.

[94] Lebin, B. R. and V. S. Rozanov; Investigation of capacity of multi-channel system accounting the statistical structure of the source. Trudy VI Vsesoyozn. Sobesh. Teor. Veroyat. Math. Stat., (1962), 215-222.

[95] Leonov, V. P; The use of the Characterisic func-

tional and semi-invariants on the ergodic theory of stationary processes. Soviet Math. Dokl. 1(1960), 78-...

[96] Linnik, Yu. V.; An information theoretic proof of the central limit theorem on Lindeberg conditions. Theory Prob. Appl., 4(1959). 288-299.

[97] ——————; On some relation to the information quantity of C. Shannon and R. Fisher about the theory of summation of random vectors. Trans. 2-nd Prague Conf., (1959), 313-327.

[98] Lynbich. Yu. I.; A remark on the capacity of discrete communication channel free from noises. Uspehi Mat. Nauk, 17, 1(1962), 191-198.

[99] Markov, Al. A.; Alphabet coding. Soviet Math. Dokl., 1(1960), 521-523.

[100] McMillan, B.; The basic theorems of information theory. Ann. Math. Stat., 24(1953), 196- 219.

[101] ——————; Two inequalities implied by unique decipherability. IRE. Trans. PGIT, Dec. (1956), 115-116.

[102] Meshalkin, L. D.; One case of the isomorphism of Bernoulli's schemes. Dokl. Akad. Nauk, 128, 1 (1959). 41-44.

[103] Mityagin, B. S.; Relationship between the ε-entropy, approximation rate and nuclearity of compactum in a linear space. Dokl. Akad. Nauk, 134(1960). 765-768.

[104] Muroga , S.; On the capacity of a noisy continuons channel. IRE. Trans. Sect. Inform. Theory, 3, 1(1957), 44-51.

[105] Nedoma, J.; The correlation of disecrete channel.

Trans. 1-st Prague Conf., (1957), 143-182

[106] Neumann, J. von; Zur Operatoren Methode in der klassischen Mechanik. Ann. Math., 33(1932), 587-642.

[107] 西尾真喜子; Wiener 積分と強定常過程の表現。 Seminar on Prob., Vol. 10, 1961.

[108] Obukhov, A. M.; Normal correlation of vectors, Izv. Akad. Nauk, 3(1938), 339-370.

[109] ——————; Theory of correlation of vectors, Moscov Gos. Univ. Uc. Zap. Mat., 45(1940), 73-92.

[110] Oxtoby, J. C.; Ergodic sets. Bull. Amer. Math. Soc., 58(1952), 116-136.

[111] —————— and S. M. Ulam; Measure-preserving homeomorphisms and metrical transitivity. Ann. of Math., (2), 42(1941), 874-920.

[112] Parthasaraty, K. R.; On the category of ergodic measures. Ill. J. Math., 5, 4(1961), 648-656.

[113] Pérez, A.; Notions generalisees d'incertitude, d'entropie et d'information de point de vue de la theorie de martingales. Trans. 1-st Prague Conf., (1967), 183-208.

[114] ——————; Sur la théorie de l'information dans le cas d'un alphabet abstrait. Trans. 1-st Prague Conf., (1957), 209-244.

[115] ——————; Sur la convergence des incertituds, entropies et informations échantillon (sample) vers leurs valeus vraies. Trans. 1-st Prague Conf., (1957), 245-252.

[116] ——————; Information theory with an abstract alphabet. Theory Prob. Appl., 4(1959), 105-109.

[117] ——————; Sur la théorie de l'information et

la discernabilité dans les problemes de décision statistique. Trans. 2-nd Prague Conf., (1959), 413-497.

[118] Pinsker, M. S.; The quantity of information about a Gaussian random stationary process, contained in a second process connected with it in a stationary manner. Dokl. Akad. Nauk, 99(1954), 213-216.

[119] ——————; The computation of the debit of a stationary channel and of the velocity with which communication is created for stationary random processes. Dokl. Akad. Nauk, 111, 4(1956), 753-756.

[120] ——————; The quantity of information of one stationary process, contained in another stationary process. Trudy III Vsesoyozn. Mat. Sbezda, I(1956), 125.

[121] ——————; The extrapolation of homogeneous random fields and the quantity of information of a Gaussian random field, contained in another Gaussian random field. Dokl. Akad. Nauk, 112(1957), 815-818.

[122] ——————; Berechung und Abschatzung der Information, der Durchlasskapazitat eines Kanals und der Entstehungsgeschwindigkeit von Nachrichten mittels zweiter Momente der Verteilungen. Dissertation, Moskau, (1957).

[123] ——————; The extrapolation of random vector processes and the quantity of information contained in a stationary random vector process relative to another one stationarily connected with it. Dokl. Akad. Nauk, 121(1958),

49-51.

[124] Pinsker, M. S.; The concept of the regularity of random processes. Theory Prob. Appl., 4, 4 (1959), 475-476.

[125] ——————; The information stability of Gaussian random variables and processes. Dokl. Akad. Nauk, 133(1960), 28-30.

[126] ——————; The entropy, the rate of establishment of entropy, and entropic stability of Gaussian random variables and processes. Dokl. Akad. Nauk, 133(1960),531-534.

[127] ——————; Dynamical systems with completely positive or zero entropy. Dokl. Akad. Nauk, 133(1960), 1025-1026.

[128] ——————; Information und Informationsstabillität zufälliger Grössen und Prozesse. Arbeiten zur Inform. V, Berlin, 1963.

[129] Plesner, A. I. and V. A. Rohlin; Spectral theory of linear operators. Uspehi Mat. Nauk, 1, 1 (1946), 71-191.

[130] Pugachev, V. S.; On the change of entropy of random functions for a linear transformation of random functions. Trudy III Vsesoyozn. Mat. Sbezda, I(1956), 125-127.

[131] Quastler, H.; Information theory and biology. Univ. of Ill. Press,1953.

[132] ——————; Information theory in psychology. Free Press, Illinois, 1956.

[133] Rényi, A.; On measures of entropy and information. Proc. 4-th Berkeley Symp., I(1962), 547-561.

[134] Rohlin, V. A.; On the classification of measur-

able partitions. Dokl. Akad. Nauk, 58, 1(1947), 29-32.

[135] Rohlin, V. A.; On the decomposition of a dynamical system to the ergodic components. Mat. Sb., 25(1949).

[136] ——————; On the fundamental ideas of measure theory. A.M.S. Transl. Ser. 1, Mat. Sb., 25(1949), 107-150.

[137] ——————; Selected problems of the metrical theory of dynamical system. Uspehi Mat. Nauk, 4, 2(1949), 57-128.

[138] ——————; On the endomorphism of compact commutative groups. Izv. Akad. Nauk, 13(1949), 324-340.

[139] ——————; On the entropy of metrical automorphism. Dokl. Akad. Nauk, 124, 5(1959), 980-983.

[140] ——————; New progress in the theory of transformations with invariant measure. Uspehi Mat. Nauk, 15, 4(1960), 3-26.

[141] ——————; Exact endomorphisms of Lebesgue spaces. Izv. Akad. Nauk, 25(1961), 499-530.

[142] ——————; On the entropy of an automorphism of a compact commutative group. Theory Prob. Appl., 6, 3(1961), 322-323.

[143] ——————; Axiomatic determination of the entropy in a transformation with invariant measure. Dokl. Akad. Nauk, 148, 4(1963), 779-781.

[144] —————— and S. V. Fomin; The spectral theory of dynamical systems. Trudy III Vsesoyozn. Mat. Sbezda, 3(1958), 284-292.

[145] Rohlin, V. A. and Ya. G. Sinai; Construction and properties of invariant measurable partitions. Dokl. Akad. Nauk, 141, 5(1961), 1034-1041.

[146] Rosenblat-Rot, M.; The notion of entropy in the theory of probability and its application for the theory of transmission with channel of communications. Trudy III Vsesoyozn. Mat. Sbezda, II(1956), 132-133.

[147] ————————; On the entropy of stochastic processes. Dokl. Akad. Nauk, 112(1957), 16-19.

[148] ————————; Theory of transmission of information through stochastic communication channels. Dokl. Akad. Nauk, 112, 2(1957), 202.

[149] ————————; Normed ε-entropy of sets and transmission of information from continuons sources through continuons communication channels. Dokl. Akad. Nauk, 130(1960), 256-268.

[150] Rozanov, Yu. A.; On a density of one Gaussian distribution with respect to another. Teor. Veroyat. Primenen. 7, 1(1962), 84-89.

[151] Sardinas, A. A. and G. W. Patterson; A neccessary and sufficient condition for unique decomposition of coded messages. IRE convention Record, 8(1953),104-108.

[152] Shannon, C. E.; A mathematical theory of communication. Bell System Tech. J., 27(1948), 379-423, 623-656.

[153] ————————; Communication in the presence of noise. Proc. IRE, 37(1949), 10-21.

[154] ————————; Prediction and entropy of printed English. Bell System Tech. J., 30(1951), 50-64.

[155] Shannon, C. E.; On the zero-error capacity of noisy channel. IRE, Trans. PGIT, (1956), 8-19.

[156] ——————; Geometrische Deutung einiger Ergebnisse bei der Berechnung der Kanalkapazität. Nachrtech, Z., 10(1957).

[157] ——————; Certain results in coding theory for noisy channels. Inform. and Control, 1 (1957), 6-25.

[158] ——————; Channels with side information at the transmitter. IBM J. Research and Development, 2, 4(1958), 289-293.

[159] ——————; Probability of error for optimal codes in a Gaussian channel. Bell System Tech. J., 38, 3(1959), 611-656.

[160] ——————; Tow-way communication channels. Proc. 4-th Berkeley Symp, I (1962), 611-644.

[161] —————— and W. Weaver; The mathematical theory of communication. Univ. of Ill. Press, 1949.

[162] Sinai, Ya. G.; On the concept of the entropy of a dynamical system. Dokl. Akad. Nauk, 124, 4 (1959), 768-771.

[163] ——————; On flows with finite entropy. Dokl. Akad. Nauk, 125, 1(1959), 1200-1202.

[164] ——————; Dynamical systems and stationary Markov processes. Theory Prob. Appl., 5, 3 (1960), 335-338.

[165] ——————; Geodesic flows on the manifold with a constant negative curvature. Soviet Math. Dokl., 1(1960), 335-339.

[166] ——————; The central limit theorem for geodesic flows on manifolds of constant nega-

tive curvature. Soviet Math. Dokl., 1(1960), 983-987.

[167] Sinai, Ya. G.; Geodesic flows on compact surfaces of negative curvature. Soviet Math. Dokl. 2, 1(1961), 106-109.

[168] ————————; Dynamical systems with countably multiple Lebesgue spectrum. I. Izv. Akad. Nauk, 25(1961), 899-924.

[169] ————————; Weak isomorphism of transformations with an invariant measure. Soviet Math. Dokl., 3, 6(1962), 1725-1729.

[170] Skpian, D.; Some comments on the detections of Ganssian signals in Ganssian noise. Trans. IRE on Inform. Theory, It-4, 2(1958), 65-68.

[171] Slepian, D.; A Note on tow binary signalling alphabets. IRE, Trans. PGIT, June (1956), 84-86.

[172] ————————; A class of binary signalling alphabets. Bell System Tech. J., 35(1956), 203-234.

[173] ————————; Comments on the detection of Gaussian signal in Gaussian noise. Trans. IRE. Sect. Inform. Theory 4, Nr. 2(1958), 165.

[174] Tempel'man, A. A.; An ergodic theorem for random fields homogeneous in the wide sense. Soviet Math. Dokl., 3, 3(1962), 817-820.

[175] Thomasian, A. J.; An elementary proof of the AEP of information theory. Ann. Math. Stat., 31(1960), 452-546.

[176] Tikhomirov, V. M.; On ε-entropy of certain classes of periodic functions. Uspehi Mat. Nauk, 17, 6(1962), 163-170.

[177] 統計数学分科会　定常過程，昭和37年度 秋季学会シンポジウム

[178] Tsaregradsky, I. P.; On the capacity of a stationary channel with finite memory. Theory Prob. Appl., 3, 1(1958), 84-96.

[179] Veršik, A. M.; On the theory of normal dynamic systems. Soviet Math. Dokl., 3, 3(1962), 625-628.

[180] ――――――――; Spectral and metric isomorphism of some normal dynamical systems. Soviet Math. Dokl., 3, 3(1962), 693-695.

[181] Vinokurov, V. G.; A condition of the regularity of a stochastic process. Dokl. Akad. Nauk, 113, 5(1957), 959-961.

[182] Vitushkin, A. G.; Theory of the transmission and processing of information. Pergamon Press, 1961.

[183] Weiss, L.; On the strong converse of the coding theorem for symmetric channels without memory. Quart. J. Appl. Math., 18, 3(1960), 209-214.

[184] Wiener, N.; The ergodic theorem. Duke Math. J., 5(1939), 1-18.

[185] ――――――; Cybernetics. Wiley, New York, 1948.

[186] ――――――; Extrapolation, interpolation, and smoothing of stationary time series. New York, 1949.

[187] Wolfowitz, J.; The coding of messanges subject to chance errors. Ill. J. Math., 1(1957), 591-606.

[188] ――――――――; The maximum achievable length of an error correcting code. Ill. J. Math., 2 (1958), 454-458.

[189] ――――――――; Stro[illegible] converse of the coding

theorem for semi-continous channels. Ill. J. Math., 3, 4(1959), 477-489.

[190] Wolfowitz, J.; Simultaneous channels. Arch. Rat. Mech. Anal., 4, 4(1960), 371-386.

[191] ——————; Strong converse of the coding theorem for the general discrete finite-memory channel. Inform. and Control, 3, 1(1960), 89-93.

[192] ——————; On coding theorems for general simultaneous channels. Trans. IRE, Prof. Group on Ciruit Theory, CT-7, 4(1960), 513-516.

[193] Yaglom, A. M.; Correlation theory of processes with random stationary n-th increments. Mat. Sb., 37(1955), 141-196. Amer. Math. Soc. Tramsl. (2), 8(1958), 87-141.

[194] ——————; Extrapolation, Interpolation and filtering of stationary random processes with rational spectral density. Trudy Moskov Mat. Obsc., 4(1955), 333-374.

[195] ——————; Einfuhrung in die theorie der stationaren znfalligen Funktionen. Heft 6 der Scheiftenreihe des Forschungsinstitus fur Mathematik bei der Deutschen Akademie der Wissenschaften zu Berlin, Akademie-Verlag, Berlin, 1959.

[196] ——————; 正規確率過程の理論における extrapolation, filtration 及び情報量の計算に対する explicit formulas. Trans. 2-nd Prague Conf., (1959), 251-262.

[197] 吉田耕作　Markov chain と H定理．位相数学，2，1 (1939)．

[198] Zaregradski, I. P.; Eine bewerkung uber die Purchlabkapazitat eines stationaren Kanals wit endlichem Gedachtnis. Arbeiten Zur Inform. II, Berlin, 1958.

[199] Halmos, P. R. and L. J. Savage; Applications of the Radon-Nikodym theorem to the theory of sufficient statistics. Ann. Math. Stat., 20(1949) 225-241.

[200] Kullback, S.: Certain inequalities in information theory and the Cramér-Rao inequality. Ann. Math. Stat., 25(1954), 745-751.