確率論とエントロピー

会田 茂樹

1 Introduction

エントロピーとは何か? エントロピーは考えている系の不確実性さを表すものとして統計物理学や情報理論に登場した概念である. 情報理論の創始者 Shannon に従って, 確率分布のエントロピーを導入しよう.

定義 1.1 (Shannon)**.** 有限集合 $\Omega = \{\omega_1, \ldots, \omega_N\}$ を考える. Ω 上の非負値関数 p で $\sum_{i=1}^N p(\omega_i) = 1$ を満たすものを考える. 各 ω_i を根元事象または標本点と言う. Ω の部分集合 A を事象といい, $P(A) = \sum_{\omega \in A} p(\omega)$ と A の確率を定める [*]. この Ω 上の確率 (または確率分布) P に対し, エントロピーを次で定義する [†].

$$H(P) = -\sum_{i=1}^{N} p(\omega_i) \log p(\omega_i). \tag{1.1}$$

注意 1.2. 上記で $0 \log 0 = 0$ とする. 対数 log は, 特に何も書かなければ, 自然対数を考えることにする.

例 1.3. (1) **コイン投げ**
$\Omega = \{$ 表, 裏 $\}$ and $P_1(\{$ 表 $\}) = P_1(\{$ 裏 $\}) = 1/2$. $H(P_1) = \log 2$ となる.
(2) **サイコロ投げ** $\Omega = \{1, 2, 3, 4, 5, 6\}$. $P_2(\{i\}) = 1/6$ $(1 \le i \le 6)$. $H(P_2) = \log 6$.
(3) **偏りのあるサイコロ投げ** $\Omega = \{1, 2, 3, 4, 5, 6\}$. $P_3(\{1\}) = 9/10, P_3(\{i\}) = 1/50$ $(2 \le i \le 6)$.

$$H(P_3) = \log \left[\left(\frac{10}{9} \right)^{9/10} (50)^{1/10} \right] \le \log \left(\frac{10}{9} \cdot \frac{3}{2} \right) < \log 2 = H(P_1) \tag{1.2}$$

可能性が 6 通りあるサイコロ投げの方が可能性が 2 通りしか無い硬貨投げより不確実性さが大きいと考えられるが実際, $H(P_2) > H(P_1)$ とエントロピーの大小関係に現れている. しかし, (3) のように偏りのあるサイコロ投げのエントロピーは硬貨投げのエントロピーより小さく, この意味で不確実性さが小さいと言える.

問題 1. 次のように出る目にかたよりのあるサイコロ投げを考える. $\Omega = \{1, 2, 3, 4, 5, 6\}$ で確率が $P_4(\{1\}) = 8/10$, $P_4(\{2\}) = 1/10$, $P_4(\{i\}) = 1/40$ $(i = 3, 4, 5, 6)$. この場合のエントロピー $H(P_4)$ を計算せよ. また, $H(P_4)$ は $H(P_1)$ より大きいか?

上記の例 (1), (2) では, エントロピーは log(起こりえるすべての場合の数) と同じである. というのは, 根元事象がすべて同様に確からしく起こりえるからである. エントロピーという概念は情報理論より先に, 熱力学, 統計物理学の分野でも現れていることを注意しておくが, ここではそれらには深入りしない. エントロピーの基本性質をあげる.

[*] $P(\{\omega_i\}) = p(\omega_i)$ $(\omega_i \in \Omega)$ に注意しよう.
[†] 今の場合, P は p から決まるので p のエントロピーと言ってもよい.

定理 1.4. $\sharp\Omega = N$ (すなわち, Ω の要素の数は N). このとき, Ω 上の任意の確率分布 P に対して,

$$0 \leq H(P) \leq \log N. \tag{1.3}$$

$H(P)$ の最小値 0 は, ある $\omega \in \Omega$ があって $P(\{\omega\}) = 1$ となる場合に達成される. また, 最大値 $\log N$ は, Ω 上の一様分布 P_u, すなわち, 任意の $\omega \in \Omega$ に対して, $P_u(\{\omega\}) = 1/N$ $(1 \leq i \leq N)$ の場合にのみ達成される.

証明は後で行う. エントロピーは次のような問題を解くのにも用いられる.

問題 8 枚の金貨と天秤ばかりがある. 8 枚のうち, 1 枚は贋金でわずかに軽い. 何回, 天秤ばかりを使えば、どれが贋金かわかるか？

解 情報理論において, エントロピーは情報の量を表す. 上記の問題で, コイン 8 枚のうちどれか一枚が偽造であるという可能性がある. したがってエントロピーは $\log 8$. 天秤を 1 回使うことによって, 次の 3 つの可能性がある. 1. 天秤がつりあう, 2. 左側の皿が軽くてあがる, 3. 右側の皿が軽くてあがる. したがって, $\log 3$ の情報量がある. よって, k 回, 天秤を使うことにより, $k \log 3$ の情報量を得る. $k \log 3 \geq \log 8$ ならば, 必要な情報が得られる. よって $k \geq 2$. 実際, 2 回で十分なことも簡単にわかる. コインの枚数 N が $3^{n-1} < N \leq 3^n$ を満たしていれば, n 回で十分である. 詳細は [16] を見よ.

問題 2. コインの枚数が 27 枚の時, 3 回で贋金を見つけ出す具体的な手順を説明せよ.

この講義では以下のように話を進める.

2. 確率論の基礎概念
確率論の基本的な概念を説明する.

3. エントロピーと大数の法則
まず, 凸関数に対する基本的な定理を用いて定理 1.4 を示す. 定理 1.4 にあるように, エントロピーが大きいということは, 多様性があるということであり, エントロピーが 0 ということは決定的で多様性は無いということに相当する. ここでは, エントロピーが小さいと情報の損失が少なく情報を圧縮することが可能であることを示す Shannon-McMillan の定理を紹介する. また, この定理は確率論における基本的な定理 (大数の弱法則) が関係することを注意する.

4. エントロピーと中心極限定理
ある状態空間 E 上を運動する粒子があり, その位置を x_t と書くことにする. $\lim_{t\to\infty} x_t = x_*$ を言いたいとしよう. このとき, E 上の連続関数 V で $V(x_*) = 0$, $V(x) > 0$ $(x \neq x_*)$ で $\lim_{t\to\infty} V(x_t) = 0$ となるものあるとしよう. さらに, $\lim_{t\to\infty} x_t$ の収束が示せれば V の性質から $\lim_{t\to\infty} x_t = x_*$ が示されることになる. このような関数をリヤプノフ関数 (Lyapunov function) という. 中心極限定理はある確率分布の時間発展 (したがって状態空間は確率分布からなる空間) が標準正規分布 (エントロピーが最大の分布) に収束するという命題だが, エントロピーをリヤプノフ関数として証明する考え方がある (最大なので, 大小を入れ替えて見る必要があるが). この話題を紹介し, エントロピー, フィッシャー情報量に関係する他の話題にも触れる.

5. Boltzmann's H-theorem, マルコフ連鎖, エントロピー

この講義では, 統計力学には深入りしないが, エントロピー増大を示している Boltzmann's H-theorem を紹介する. このエントロピーの増大法則の証明は簡単ではないがマルコフ連鎖という確率的な時間発展で確率分布のエントロピーが増大することは 3 節の結果を用いて簡単に示せる. このことを用いて (すなわちエントロピーをリヤプノフ関数として) マルコフ連鎖の分布が一様分布 (エントロピー最大の分布) に収束することを示す.

2 確率論の基礎概念

前の節で, 根元事象の数が有限個の離散型の確率空間を考えたが, 確率空間は, 全測度 1 の測度空間として定義される. 例えば, [15], [14], [12] を参照せよ.

定義 2.1. 3 つ組 $(\Omega, \mathcal{F}, P)$ が確率空間とは, 次が成立する時に言う.

(1) Ω は集合である. $\mathcal{F}$ は Ω の部分集合の族 (すべての部分集合からなる集合 2^Ω というわけではない) であり, 次を満たしている.

  (a) $A_1, A_2, \ldots, A_i, \ldots \in \mathcal{F}$ ならば $\cup_{i=1}^\infty A_i \in \mathcal{F}$.

  (b) $A \in F$ ならば, $A^c \in F$.

  (c) $\Omega, \emptyset \in F$.

(2) $A \in \mathcal{F}$ に対して, 非負値の数 $P(A)$ が定まり次を満たす.

  (a) すべての $A \in \mathcal{F}$ に対し, $0 \leq P(A) \leq 1$.

  (b) $P(\Omega) = 1$.

  (c) (完全加法性) $A_1, A_2, \ldots, A_i, \ldots \in \mathcal{F}$, $A_i \cap A_j = \emptyset$ $(i \neq j)$ ならば

$$P\left(\cup_{i=1}^\infty A_i\right) = \sum_{i=1}^\infty P(A_i).$$

  関数 $P : \mathcal{F} \to [0, 1]$ は $(\Omega, \mathcal{F})$ 上の確率 (測度) という. $A \in \mathcal{F}$ は事象と呼ばれ, $P(A)$ は, 事象 A の (起こる) 確率を表す.

次に「確率変数」という概念を説明する. X が確率変数とは,

1. X は実数値を取る変数,
2. I を $\mathbb{R}$ の区間 $((a, b), [a, b], (a, b], (-\infty, b)$ 等々) とすると X が I の値を取る確率 $P(X \in I)$ が定まっている

ような変数を言う. 例えば,

1. 1 から 6 のどの目も出るのが同様に確からしいサイコロを考える. 1 回投げて出る目の数 X は確率変数で. $P(X = i) = \frac{1}{6}$ $(1 \leq i \leq 6)$. サイコロを 2 回投げて最初の目と 2 回めの目の数を足した物を Y とすると Y は 2 から 12 までの整数を取る確率変数で $P(Y = 2) = P(Y = 12) = \frac{1}{36}$.

2. 区間 $[0,1]$ からでたらめに一つ数を選びそれを X とする. これも確率変数で

$$P(X \in [a,b]) = b - a \qquad 0 \le a \le b \le 1,$$

となる.

また, 確率変数 X が与えられた時, 各区間 I に対して確率 $P_X(I) := P(X \in I)$ が定まる. これにより, $\mathbb{R}$ の部分集合に確率が与えられることになる. この P_X を X の (確率) 分布と言う. 1回のサイコロ投げの X の確率分布 P_X は $P_X(\{i\}) = \frac{1}{6}$ $(1 \le i \le 6)$ のように, $\{1,2,3,4,5,6\}$ という有限集合に分布している.

上で説明した確率変数の定義は, 直感的には理解できるが, 数学における確率変数は「確率空間上の可測関数」として定義される.

定義 2.2. $(\Omega, \mathcal{F}, P)$ を確率空間とする.
(1) Ω 上の関数 $X : \Omega \to \mathbb{R}$ が確率変数とは, 任意の区間 $I \subset \mathbb{R}$ に対して

$$X^{-1}(I) := \{\omega \in \Omega \mid X(\omega) \in I\} \in \mathcal{F}$$

である (このことを $\mathcal{F}$ 可測という) ときに言う. また, X が I の値を取る確率 $P(X \in I)$ は $P(X^{-1}(I))$ で与えられる.
(2) 確率変数 X に対して, $P_X(I) = P(X \in I)$ と定めることにより得られる $\mathbb{R}$ 上の確率を X の確率分布という.

問題 3. $(\Omega, \mathcal{F}, P)$ を確率空間とする. $A_i \in \mathcal{F}$ $(i = 1, 2, \ldots)$ ならば $\cap_{i=1}^{\infty} A_i \in \mathcal{F}$ となることを定義から示せ.

問題 4. $(\Omega, \mathcal{F}, P)$ を確率空間とする. $A, B \in \mathcal{F}$ が $A \subset B$ を満たす時, $P(A) \le P(B)$ を示せ.

この講義では次のような (1) または (2) のタイプの確率変数のみを考えることにする.

定義 2.3. (1) 確率変数 X の取る値が可算無限個 $\{a_i\}_{i=1}^{\infty}$ (もちろん有限個の場合も考える) のとき, X は離散型確率変数という. $p_i = P(X = a_i)$ とおくと $p_i > 0$ かつ $\sum_{i=1}^{\infty} p_i = 1$ となることがわかる. ($P(X = a_i) = 0$ となる状況を考える場合もあるが, 確率 0 の事象なので, ここではそういう場合を排除して考える).
(2) 確率変数 X の分布が確率密度関数 $f(x)$ を持つ場合. すなわち,

$$f(x) \ge 0, \quad \int_{\mathbb{R}} f(x)dx = 1$$

を満たす $f(x)$ を用いて, 任意の区間 I に対して,

$$P(X \in I) = \int_I f(x)dx$$

と確率分布が与えられる場合.

確率変数 X の平均 (期待値 (expectation) とも言う) $E[X]$, 分散 (variance) $V[X]$ を定義する.

定義 2.4. (1) X が離散型のとき
$P(X = a_i) = p_i$ のとき,

$$E[X] = \sum_{i=1}^{\infty} a_i p_i, \tag{2.1}$$

$$V[X] = E[(X-m)^2] = \sum_{i=1}^{\infty} (a_i - m)^2 p_i, \quad \text{ここで, } m = E[X]. \tag{2.2}$$

$\Omega = \{\omega_k \mid k \in \mathbb{N}\}$ のように根元事象が可算無限個 (有限個でもよいが) のときは,

$$E[X] = \sum_{k=1}^{\infty} X(\omega_k) P(\{\omega_k\}) \tag{2.3}$$

とも書ける.
(2) X の分布が確率密度関数 $f(x)$ を持つ時

$$E[X] = \int_{\mathbb{R}} x f(x) dx, \tag{2.4}$$

$$V[X] = E[(X-m)^2] = \int_{\mathbb{R}} (x-m)^2 f(x) dx, \quad \text{ここで } m = E[X]. \tag{2.5}$$

注意 2.5. (1) 正確には, 無限級数, 積分が収束する場合 (このとき, 可積分という) のみ $E[X]$, $V[X]$ は定義される.
(2) 上記の定義でいずれの場合も

$$V[X] = E[X^2] - E[X]^2 \tag{2.6}$$

が成立する.

サイコロを N 回投げる時 ($N \geq 1$), i 回目に出た目を X_i とする. $i \neq j$ のとき, i 回目の試行と j 回目の試行は独立と考えられるので, X_i, X_j の確率変数も独立であると考えられる. 一般に確率変数 $X_1, \ldots, X_N$ の独立性を次のように定義する.

定義 2.6 (確率変数の独立性)**.** $\{X_i\}_{i=1}^N$ を確率空間 $(\Omega, \mathcal{F}, P)$ 上の確率変数とする. $\{X_i\}_{i=1}^N$ が独立とは, 任意の区間 $I_1, \ldots, I_N$ に対して

$$P(X_1 \in I_1, \cdots, X_N \in I_N) = P(X_1 \in I_1) \cdots P(X_N \in I_N) \tag{2.7}$$

となる時に言う.

N 回のサイコロ投げではなく, 無限回のサイコロ投げを考えることもできる. その場合, 無限個の確率変数を考えることになる.

注意 2.7. (1) $\{X_i\}_{i=1}^N$ が独立の時, この中の任意有限個の確率変数 $\{X_{i_k}\}_{k=1}^n$ も独立である.
(2) 可算無限個の確率変数 $\{X_i\}_{i=1}^\infty$ が独立とは, 任意の N に対して, $\{X_i\}_{i=1}^N$ が独立であると定義する.

定義 2.8. 確率変数 $\{X_i\}_{i=1}^{\infty}$ が独立で各 X_i の分布がすべて同じ時, $\{X_i\}_{i=1}^{\infty}$ は独立同分布に従う確率変数という. 英語では, independent and identically distributed random variables (略して, i.i.d. random variables) という.

期待値に関する基本的な結果と例をあげる.

定理 2.9. 確率空間 $(\Omega, \mathcal{F}, P)$ 上の確率変数 X, Y について次が成立する.
(1) [線形性] $E[aX + bY] = aE[X] + bE[Y] \quad (a, b \in \mathbb{R})$.
(2) X, Y が独立で $E[X]$, $E[Y]$ が定まれば $E[XY]$ も収束し, $E[XY] = E[X]E[Y]$.

例 2.10 (確率分布の例)**.** (1) $\{X_i\}_{i=1}^{n}$ は $P(X_i = 1) = p$, $P(X_i = 0) = 1 - p$ $(0 < p < 1)$ となる独立確率変数とする. これは例えば, 表が出る確率が p, 裏が出る確率が $1 - p$ の硬貨投げを独立に n 回行うことにより得られる確率変数である. $S_n = \sum_{i=1}^{n} X_i$ とおくと S_n は表の出た回数を表す確率変数になる.

$$P(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad 0 \le k \le n.$$

ここで, $\binom{n}{k} = {}_nC_k = \dfrac{n!}{k!(n-k)!}$.

(2) 離散型確率変数 X の分布がパラメータ λ $(\lambda > 0)$ のポアソン分布に従うとは,

$$P_X(\{k\}) = \frac{\lambda^k}{k!} e^{-\lambda} \qquad k = 0, 1, 2, \ldots$$

のときに言う. 簡単な計算で $E[X] = \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = \lambda$, $V[X] = \lambda$ が示せる. ポアソン分布は, 起こることが稀だが, 長期間経過するとある程度起こるような現象で現れる確率分布である.

(3) X の分布が平均 m, 分散 σ^2 の正規分布に従うとは,

$$P_X(I) = \int_I \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}} dx$$

となる時にいう. この分布を記号 $N(m, \sigma^2)$ と書いて表す. このとき, 実際に $E[X] = m$, $V[X] = \sigma^2$ が計算でわかる. $m = 0$, $\sigma^2 = 1$ のとき, 標準正規分布という.

問題 5. 上記の例 (1) について, 考える. $E[X_i] = p$, $E[S_n] = np$, $V[S_n] = np(1-p)$ を示せ. また, $np = \lambda$ のとき, $\lim_{n\to\infty} P(S_n = k) = \frac{\lambda^k}{k!} e^{-\lambda}$ $(k = 0, 1, 2, \ldots)$ を示せ.

再びエントロピーの話題に戻る.

3 エントロピーと大数の法則 (Shannon and McMillan's theorem)

ここでは, 情報理論における通信文の圧縮に関する基本定理 (Shannon-McMillan の定理) を紹介し, 確率論における大数の弱法則とエントロピーの関係を見る.

自然数の有限部分集合 $A = \{1, \ldots, N\} \subset \mathbb{N}$ を考える. ただし, $N \ge 2$. A の各元を 文字 (letter), 集合 A をアルファベット (alphabet) と呼ぶことにする. 有限列 $(\omega_1, \omega_2, \ldots, \omega_n)$ $(\omega_i \in A)$ を長さ

n の文 (sentence) という. 長さ n の文全体は, A の直積空間 $A^n := \{(\omega_1, \ldots, \omega_n) \mid \omega_i \in A\}$ と同一視できる. P を A 上の確率分布とする. $P(\{i\}) = p_i$ とおく. この節では, P のエントロピーを $\log_N$ を用いて定義する:

$$H(P) = -\sum_{i=1}^{N} p_i \log_N p_i. \tag{3.1}$$

凸関数の性質を用いて次の定理を証明しよう. これは, 定理 1.4 と同じ内容である.

定理 3.1. (1) A 上の任意の確率 P に対し, $0 \leq H(P) \leq 1$ となる.

(2) $H(P) = 0 \iff$ ある $i \in A$ に対して $p_i = 1$.

(3) $H(P) = 1 \iff P$ は A 上の一様分布である. すなわち, $p_i = 1/N$ $(1 \leq i \leq N)$.

補題 3.2. $-\infty \leq a < b \leq \infty$ とする. $\varphi : (a, b) \to \mathbb{R}$ は C^2 級で $\varphi''(x) > 0$ $(a < x < b)$ とする. N を自然数とする. $m_i \geq 0$ および $\sum_{i=1}^{N} m_i = 1$ を満たす任意の $\{m_i\}_{i=1}^{N}$ と x_i $(a < x_i < b, 1 \leq i \leq N)$ に対して

$$\varphi\left(\sum_{i=1}^{N} m_i x_i\right) \leq \sum_{i=1}^{N} m_i \varphi(x_i). \tag{3.2}$$

さらに, $m_i > 0$ for all i の時, (3.2) で等号が成立するための条件は $x_1 = \cdots = x_N$ となる.

φ の例としては, $(0, \infty)$ における $x \log x, -\log x$ などがある.

問題 6 (). $f(x)$ を $\mathbb{R}$ 上の確率密度関数とする. $g(x)$ が正値連続関数で $\int_{\mathbb{R}} g(x) f(x) dx = 1$ を満たすとする.

$$\int_{\mathbb{R}} g(x) \left(\log g(x)\right) f(x) dx \geq 0$$

を示せ. また, 等号が成立するのは, $g(x) \equiv 1$ のときであることを示せ.

補題 3.2 の証明. 帰納法で示す. $N = 2$ の時: $\varphi(tx + (1-t)y) \leq t\varphi(x) + (1-t)\varphi(y)$ が示すべき式になる. $z = tx + (1-t)y$ とおくと Taylor の定理を用いて x, z の間の値 c, y, z の間の値 c' が存在して

$$\begin{aligned}
&t\varphi(x) + (1-t)\varphi(y) - \varphi(tx + (1-t)y) \\
&= t\left(\varphi(x) - \varphi(z)\right) + (1-t)\left(\varphi(y) - \varphi(z)\right) \\
&= t\left(\varphi'(z)(x-z) + \frac{\varphi''(c)}{2}(x-z)^2\right) + (1-t)\left(\varphi'(z)(y-z) + \frac{\varphi''(c')}{2}(y-z)^2\right) \\
&= t\frac{\varphi''(c)}{2}(x-z)^2 + (1-t)\frac{\varphi''(c')}{2}(y-z)^2 \geq 0.
\end{aligned}$$

$\varphi''(x) > 0$ なので, 等号成立は $t > 0, 1-t > 0$ の場合なら $x = y$ の場合に限る.

$N-1$ の時に成立しているとし, N の場合を示す. まず不等式を示す. $m_1 \neq 1$ と仮定して一般性は失われない. $y = \frac{1}{1-m_1}\sum_{i=2}^{N} m_i x_i$ とおくと $\sum_{i=2}^{N} \frac{m_i}{1-m_1} = 1$ と $N=2, N-1$ の結果を用いて

$$\begin{aligned}
\varphi\left(\sum_{i=1}^{N} m_i x_i\right) &= \varphi(m_1 x_1 + (1-m_1)y) \\
&\leq m_1\varphi(x_1) + (1-m_1)\varphi(y) \\
&\leq m_1\varphi(x_1) + (1-m_1)\sum_{i=2}^{N} \frac{m_i}{1-m_1}\varphi(x_i) \\
&\leq \sum_{i=1}^{N} m_i \varphi(x_i).
\end{aligned}$$

また, $m_i > 0$ $(1 \leq i \leq N)$ の時の等号成立条件を求める. 上記の式変形ですべて不等号が等号になる必要がある. したがって, 帰納法の仮定より, $x_1 = y$ かつ $x_2 = \cdots = x_N$. よって, $x_1 = \cdots = x_N$ が条件となる. □

定理 3.1 の証明. $H(P) \geq 0$ を示す.

$$\begin{aligned}
H(P) &= -\sum_{\{i \,|\, p_i > 0\}} p_i \log_N p_i \\
&= \sum_{\{i \,|\, p_i > 0\}} p_i \log_N \left(\frac{1}{p_i}\right) \geq 0.
\end{aligned}$$

また, この式より, $H(P) = 0$ となるのは, $p_i > 0$ となる i について $\log(1/p_i) = 0$ となることが必要. したがって, $H(P) = 0$ となるのは, ある i について $p_i = 1$ の時である. $H(P) \leq 1$ と (3) を示す. (3.2) を $m_i = 1/N, x_i = p_i$ and $\varphi(x) = x\log x$ の場合に使うと任意の確率分布 $\{p_i\}$ に対して,

$$\varphi\left(\frac{1}{N}\sum_{i=1}^{N} p_i\right) \leq \frac{1}{N}\sum_{i=1}^{N} \varphi(p_i). \tag{3.3}$$

$\sum_{i=1}^{N} p_i = 1$ なので,

$$\frac{1}{N}\log\left(\frac{1}{N}\right) \leq \frac{1}{N}\sum_{i=1}^{N} p_i \log p_i.$$

したがって, $-\sum_{i=1}^{N} p_i \log p_i \leq \log N$. ゆえに $-\sum_{i=1}^{N} p_i \log_N p_i \leq 1$. $H(P) = 1$ は (3.3) で等号成立の時のみなので, 補題 3.2 から, $p_i = 1/N$ $(1 \leq i \leq N)$ の時に限ることもわかる. □

次のような状況を考えよう. ここに, 無記憶情報源があり, ある確率分布 P に従って独立に, アルファベット $A = \{1, \ldots, N\}$ $(N \geq 2)$ の中の文字を一定時間間隔で送り出しているとする. 数学的には, A に値を取る確率分布 P に従う独立確率変数列 (i.i.d.) $\{X_i\}_{i=1}^{\infty}$ を考えていることになる. このとき, 情報伝達のため, 長さ n の文 $\{X_1, \ldots, X_n\}$ を短い文に圧縮することを考えたい.

基本的な Observation:

(1) 確率分布 P のエントロピーが 0 の場合を考えよう. 補題 3.2 からある $1 \leq i \leq N$ があって $P(\{i\}) = 1$, $P(\{j\}) = 0$ $(j \neq i)$ となる. このとき, 独立確率変数列 $\{X_i\}$ は実際は, 一意に決まり $\{i, \ldots, i, \ldots\}$ と同じ文字が並ぶことになる. 従って, 文の多様性はありえず, エントロピーが 0 とわかっていれば, 最初の文字を見た瞬間, すべての文章 (と言っても一つしか無いが) が復元できることになる. すなわち, 長さ n の文章はいくら長くても長さ 1 の文に圧縮しても何ら情報の損失は無い. では, エントロピーが 0 で無い場合は何が言えるだろうか?

(2) $N \geq 3$ とし, $P(\{1\}) = P(\{2\}) = 1/2$, $P(\{i\}) = 0$ for $3 \leq i \leq N$ となる確率分布を考える. すなわち, $i \geq 3$ は発信されないことになる. このとき, 長さ n の文の場合の数は, 2^n となる. A のアルファベットを使って長さ k の文で表現できるものは N^k 通りある. したがって, k が $N^k \geq 2^n (\Longleftrightarrow \frac{k}{n} \geq \log_N 2 = H(P))$ を満たせば, 確率分布 P に従って送付される長さ n の文は A をすべて使って表される長さ k の文字列に圧縮し, 情報の損失が無い (すなわち, 復元が可能) とわかる.

(2) で言えたことをまとめると次のようになる:
$\frac{k}{n} \geq H(P)$ ならば符号器 (encoder) $\Phi : A^n \to A^k$ と復号器 (decoder) $\Psi : A^k \to A^n$ が存在して

$$P\Big(\Psi(\Phi(X_1, \ldots, X_n)) \neq (X_1, \ldots, X_n)\Big) = 0. \tag{3.4}$$

確率 $P\Big(\Psi(\Phi(X_1, \ldots, X_n)) \neq (X_1, \ldots, X_n)\Big)$ は誤り確率 (error probability) と呼ばれている. 一般には, 誤り確率は 0 にできず, 次の結果が証明できる. 関連する事は [7] を参照.

定理 3.3 (Shannon and McMillan)**.** P を A 上の確率分布とし, $H(P) < 1$ とする. 正数 $R > H(P)$ を一つ取る. 任意の $\varepsilon > 0$ に対して, $M \in \mathbb{N}$ が存在して, $n \geq M$ と $\frac{k}{n} \geq R$ を満たす n, k に対して, 写像 $\Phi : A^n \to A^k$, $\Psi : A^k \to A^n$ が存在し,

$$P\Big(\Psi(\Phi(X_1, \ldots, X_n)) \neq (X_1, \ldots, X_n)\Big) \leq \varepsilon. \tag{3.5}$$

R は符号レートと呼ばれる. この定理の証明には次の評価が必要である.

補題 3.4. $\{Z_i\}_{i=1}^{\infty}$ を i.i.d. とする. $E[|Z_i|] < \infty$, $E[|Z_i|^2] < \infty$ を仮定する. このとき,

$$P\left(\left|\frac{Z_1 + \cdots Z_n}{n} - m\right| \geq \delta\right) \leq \frac{\sigma^2}{n\delta^2}, \tag{3.6}$$

ここで $m = E[Z_i], \sigma^2 = E[(Z_i - m)^2]$.

問題 7. 補題 3.4 をチェビシェフの不等式を用いて証明せよ. チェビシェフの不等式とは $E[|X|^2] < \infty$ を満たす確率変数に対する不等式:

$$\text{任意の } \delta > 0 \text{ に対して} \quad P(|X| \geq \delta) \leq \frac{E[|X|^2]}{\delta^2} \tag{3.7}$$

を言う.

補題 3.4 から $\frac{Z_1+\cdots+Z_n}{n}$ は次の意味で平均 m に収束することがわかる.

定理 3.5. [大数の弱法則] $\{Z_i\}$ が補題 3.4 の仮定を満たすとすると任意の $\delta > 0$ に対して

$$\lim_{n\to\infty} P\left(\left|\frac{Z_1 + \cdots Z_n}{n} - m\right| \geq \delta\right) = 0. \tag{3.8}$$

注意 3.6. 実際は同じ仮定の下で, 次の大数の強法則が証明できる.

$$P\left(\left\{\omega \in \Omega \;\middle|\; \lim_{n\to\infty} \frac{Z_1(\omega) + \cdots + Z_n(\omega)}{n} = m\right\}\right) = 1. \tag{3.9}$$

定理 3.3 の証明. $n \in \mathbb{N}$ とする. $A = \{1, \ldots, N\}$ 上の確率 P を用いて,

$$A^n := \{\boldsymbol{\omega} = (\omega_1, \ldots, \omega_n) \mid \omega_i \in A, 1 \leq i \leq n\}$$

上に次のように確率 P_n を入れる.

$$P_n(\{\boldsymbol{\omega}\}) = \prod_{i=1}^{n} P(\{\omega_i\}) \qquad \boldsymbol{\omega} = (\omega_1, \ldots, \omega_n),\ \omega_i \in A \quad (1 \leq i \leq n). \tag{3.10}$$

$\boldsymbol{\omega}$ は文であり, ω_i は $\boldsymbol{\omega}$ の i 番目の文字であることに注意せよ. $X_i(\boldsymbol{\omega}) = \omega_i$ と定めると $\{X_i\}_{i=1}^n$ は各 X_i の分布が P となる i.i.d. となり, 無記憶情報源から発信される文字列となる.
[証明の方針] n が大きい時, $C_n \subset A^n$ で次の性質を満たすものの存在を言えばよい.

(1) $P_n(C_n) \geq 1 - \varepsilon$.

(2) $\sharp C_n \leq N^{nR}$. ただし, $\sharp C_n$ は C_n の要素の数を表す.

この評価が示せれば $k \geq nR$ ならば単射な写像 $\Phi : C_n \to A^k$ と写像 $\Psi : A^k \to C_n$ で

$$\text{すべての } (\omega_1, \ldots, \omega_n) \in C_n \text{ について}, \qquad \Psi(\Phi(\omega_1, \ldots, \omega_n)) = (\omega_1, \ldots, \omega_n)$$

とできる. 写像 $\Phi : A^n \to A^k$ を C_n で定義されている写像を拡張して (C_n 以外では何でもよい) 定めることにする. $X_i(\boldsymbol{\omega}) = \omega_i$ ($\omega \in A^n$) だったので

$$P_n\Big(\Psi(\Phi(X_1, \ldots, X_n)) = (X_1, \ldots, X_n)\Big) \geq P_n(C_n) \geq 1 - \varepsilon \tag{3.11}$$

となり証明が完了する.
上記の C_n の存在を補題 3.4 を用いて示そう.
(A^n, P_n) 上の確率変数

$$Z_i(\boldsymbol{\omega}) = -\log_N P(\{\omega_i\}) \qquad (1 \leq i \leq n).$$

を考えると $\{Z_i\}_{i=1}^n$ は i.i.d. で期待値, 分散は次のようになる.

$$m = E[Z_i] = -\sum_{k=1}^{N} P(\{k\}) \log_N P(\{k\}) = H(P)$$

$$\sigma^2 = E[(Z_i - E[Z_i])^2] = \sum_{k=1}^{N} (\log_N P(\{k\}))^2 P(\{k\}) - H(P)^2. \tag{3.12}$$

$\delta > 0$ を $R > H(P) + \delta$ となるように小さく取る. 補題 3.4 より,

$$P_n\left(\frac{1}{n}\sum_{i=1}^{n}(-\log_N P(\{\omega_i\})) \geq H(P) + \delta\right) \leq \frac{\sigma^2}{n\delta^2}. \tag{3.13}$$

したがって, 任意の $\varepsilon > 0$ に対して, $M \in \mathbb{N}$ が存在して

$$P_n\left(\frac{1}{n}\sum_{i=1}^{n}(-\log_N P(\{\omega_i\})) \geq H(P) + \delta\right) \leq \varepsilon \qquad \text{for all } n \geq M. \tag{3.14}$$

$$\begin{aligned}
&\left\{(\omega_1, \ldots, \omega_n) \;\middle|\; \frac{1}{n}\sum_{i=1}^{n}(-\log_N P(\{\omega_i\})) < H(P) + \delta\right\} \\
&= \left\{(\omega_1, \ldots, \omega_n) \;\middle|\; \prod_{i=1}^{n} P(\{\omega_i\}) > N^{-n(H(P)+\delta)}\right\} \\
&\subset \left\{(\omega_1, \ldots, \omega_n) \;\middle|\; \prod_{i=1}^{n} P(\{\omega_i\}) \geq N^{-nR}\right\} \\
&= \left\{\boldsymbol{\omega} \in A^n \;\middle|\; P_n(\{\boldsymbol{\omega}\}) \geq N^{-nR}\right\} =: C_n
\end{aligned} \tag{3.15}$$

なので, (3.14) より, $n \geq M$ のとき,

$$P_n(C_n) \geq 1 - \varepsilon \tag{3.16}$$

C_n の定義から

$$P_n(C_n) = \sum_{\boldsymbol{\omega} \in C_n} P_n(\{\boldsymbol{\omega}\}) \geq \sum_{\boldsymbol{\omega} \in C_n} N^{-nR} = \sharp C_n \cdot N^{-nR}. \tag{3.17}$$

$P_n(C_n) \leq 1$ だから $\sharp C_n \cdot N^{-nR} \leq 1$, すなわち,

$$\sharp C_n \leq N^{nR}.$$

これで, (1),(2) が示された. □

4 エントロピーと中心極限定理

$\{X_i\}_{i=1}^{\infty}$ を $E[X_i] = 0$ and $E[X_i^2] = 1$ を満たす i.i.d. とする.

$$T_n = \frac{X_1 + \cdots X_n}{\sqrt{n}}$$

と定めると $E[T_n] = 0$, $V[T_n] = 1$. 大数の法則によれば, $\frac{T_n}{\sqrt{n}}$ はある意味で 0 に収束するが T_n は $\sqrt{n}$ 倍されているため, 確率変数としては 0 に収束せず以下の意味で収束することがわかる.

定理 4.1 (中心極限定理 (Central limit theorem))**.** 任意の $-\infty < a < b < \infty$ に対して

$$\lim_{n\to\infty} P\left(T_n \in [a,b]\right) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx. \tag{4.1}$$

すでに説明したように $G(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$ は確率密度関数であり, G を密度関数とする確率分布は標準正規分布と呼ばれる.

1. 中心極限定理の標準的な証明は T_n の特性関数 $\varphi(t) = E[e^{\sqrt{-1}tT_n}]$ $(t \in \mathbb{R})$ (確率分布のフーリエ変換) を用いるものである.

2. $P(X_i = 1) = p$, $P(X_i = 0) = 1 - p$ を満たす確率変数の時, $S_n = \sum_{i=1}^n X_i$ とおくと二項分布 $P(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k}$ を得る. Stirling の公式 $\lim_{n\to\infty} \frac{n!}{\sqrt{2\pi n} e^{-n} n^n} = 1$ と二項分布の形から

$$\lim_{n\to\infty} P\left(a \le \frac{S_n - np}{\sqrt{np(1-p)}} \le b\right) = \int_a^b G(x)dx \tag{4.2}$$

   を示す方法 (De Moivre-Laplace の証明) もある.

ここでは, エントロピーを用いた証明を紹介する. 参考文献としては [8, 2, 3] がある. X_i の確率分布は確率密度関数 f を持つ. すなわち,

$$P\left(X_i \in [a,b]\right) = \int_a^b f(x)dx.$$

の場合, $T_n = \frac{\sum_{i=1}^n X_i}{\sqrt{n}}$ の分布も確率密度関数を持つことがわかる (下の補題 4.4 (1) より). 密度関数を $f_n(x)$ とおく. (4.1) は次と同値になる.

$$\lim_{n\to\infty} \int_a^b f_n(x)dx = \int_a^b G(x)dx. \tag{4.3}$$

f に付加的な条件を置くと, より強い結果が証明できる:

$$\lim_{n\to\infty} \int_{\mathbb{R}} |f_n(x) - G(x)|dx = 0. \tag{4.4}$$

ここで確率分布に対するエントロピーを定義する.

定義 4.2. f を密度関数とする確率分布 P に対して, エントロピー $H(P)$, フィッシャー情報量 (Fisher information) $I(P)$ を次のように定める.

$$H(P) = -\int_{\mathbb{R}} f(x)\log f(x)dx, \tag{4.5}$$

$$I(P) = \int_{\mathbb{R}} \frac{f'(x)^2}{f(x)}dx. \tag{4.6}$$

また, 確率変数 X の分布が P でその密度が f のとき, $H(P)$, $I(P)$ を $H(f)$, $I(f)$, $H(X)$, $I(X)$ などとも書くことにする.

問題 8. 集合 $\{H(f_t) \mid t > 0\}$ は実数全体の集合と一致することを示せ. ただし, $f_t(x) = \frac{1}{\sqrt{2\pi t}}e^{-\frac{x^2}{2t}}$. また, $H(f) = -\infty$ となる 確率密度関数 f の例をあげよ.

定理 4.3. [[2, 8] 確率分布 P の密度関数は C^1 級で $I(f) < \infty$ を満たすとする. このとき, すべての n について, f_n は連続関数であり, 次の意味で G に収束する.

$$\lim_{n\to\infty} f_n(x) = G(x) \qquad \mathbb{R} \text{ 上広義一様} \tag{4.7}$$

$$\lim_{n\to\infty} H(f_n) = H(G), \tag{4.8}$$

$$\lim_{n\to\infty} \int_{\mathbb{R}} |f_n(x) - G(x)|dx = 0. \tag{4.9}$$

なお, $H(G) = \log\sqrt{2\pi} + \frac{1}{2}$ である. この定理を証明するために必要なことを以下にまとめる.

補題 4.4. (1) 確率変数 X, Y が独立でそれぞれ密度関数 f, g を持つとすると $a(X+Y)$ $(a > 0)$ の分布も密度関数 $h(x) = \dfrac{1}{a}\displaystyle\int_{\mathbb{R}} f\left(\frac{x}{a} - y\right)g(y)dy.$ を持つ.
(2) (Gibbs の補題) $f(x)$ を平均 0, 分散 1 の確率分布の連続な密度関数とする. すなわち

$$\int_{\mathbb{R}} xf(x)dx = 0, \tag{4.10}$$

$$\int_{\mathbb{R}} x^2 f(x) = 1. \tag{4.11}$$

このとき,

$$H(f) \le H(G). \tag{4.12}$$

等号成立は $f(x) = G(x)$ $(x \in \mathbb{R})$ の場合に限る.
(3) (Shannon-Stam's inequality) X, Y を独立な確率変数で密度関数が (4.10), (4.11) を満たすとする. このとき, $0 < a < 1$ を満たす実数について, に対し,

$$aH(X) + (1-a)H(Y) \le H(\sqrt{a}X + \sqrt{1-a}Y). \tag{4.13}$$

等号が成立するのは, X, Y の分布が $N(0,1)$ の場合に限る.
(4) (Blachman-Stam's inequality) X, Y を独立な確率変数で密度関数が (4.10), (4.11) を満たすとする. このとき, $0 < a < 1$ を満たす 実数に対し,

$$I(\sqrt{a}X + \sqrt{1-a}Y) \le aI(X) + (1-a)I(Y). \tag{4.14}$$

等号が成立するのは, X, Y の分布が $N(0,1)$ の場合に限る.
(5) (Csiszár-Kullback-Pinsker) 確率密度関数 f が (4.10), (4.11) を満たすとすると

$$\left(\int_{\mathbb{R}} |f(x) - G(x)|dx\right)^2 \leq 2\left(H(G) - H(f)\right). \tag{4.15}$$

補題 4.4 (1) により, T_n は 密度関数 f_n を持つことがわかる. また, $H(T_{2^n})$ が単調増加, $I(T_{2^n})$ が単調減少であることもわかる. このことを用いて, $H(T_n)$ が $H(G)$ に収束することもわかる.

定理 4.3 の証明のスケッチ. Shannon-Stam の不等式により, (4.8) を示すことができる. これと Csiszár-Kullback-Pinsker の不等式により (4.9) がわかる. また, Blachman-Stam の不等式により $I(f_n) \leq I(f)$ が得られる. これと (4.9) を用いて (4.7) を得る. 詳しくは [2], [8] とその中の参考文献を見よ. □

$I(f)$ と $H(f)$ の間には次のような関係がある.

補題 4.5. (1) (Stam's inequality) C^1 級の確率密度関数 f に対し,

$$e^{-2H(f)} \leq \frac{1}{2\pi e} I(f). \tag{4.16}$$

(2) (Gross's inequality) C^1 級の確率密度関数 f に対し,

$$-2H(f) \leq I(f) - \log\left(2\pi e^2\right). \tag{4.17}$$

(1) と (2) は同値な不等式である.

問題 9. (1) $\log x \leq x - 1$ $(x > 0)$ を示せ.
(2) (1) の不等式と Stam の不等式を用いて Gross の不等式を示せ.

問題 10. Gross の不等式を用いて Stam の不等式を以下の議論を用いて示せ.
(1) $t > 0$ とする. $f_t(x) = \sqrt{t} f(\sqrt{t}x)$ は確率密度関数であることを示せ. $f_t(x)$ を Gross の不等式に代入し, $-2H(f)$ の上からの評価を得よ.
(2) (1) で得られた評価を t を動かして最良評価を得ることにより Stam の不等式を示せ.

問題 11. Gross の不等式は次の Gross の対数ソボレフ不等式と同値であることを示せ.
$\int_{\mathbb{R}} u(x)^2 d\mu(x) = 1$ を満たす, C^1 級関数について

$$\int_{\mathbb{R}} u(x)^2 \log u(x)^2 d\mu(x) \leq 2\int_{\mathbb{R}} |u'(x)|^2 d\mu(x), \tag{4.18}$$

ここで $d\mu(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$.

Stam の不等式と Gross の (対数) ソボレフ不等式に関する歴史的経緯を少しまとめておく.

注意 4.6. $\mathcal{N}(f) = e^{2H(f)}$ は Shannon's entropy power functional と呼ばれている量である. Stam [10] は彼の不等式を 1959 年に証明した. これは, Fisher information と Shannon のエントロピーの関係を明らかにしたと言う意味で重要な業績である. これとは独立に, Leonard Gross [6] は彼の名前のついた対数ソボレフ不等式を 1975 年に証明した. (実は, 同じ内容は P. Federbush も証明している). 重要な事は, Gross は, さらに, 対数ソボレフ不等式が Ornstein-Uhlenbeck 半群 T_t の超縮小性 (hypercontractivity) と同値であると示したことである. T_t とその超縮小性を説明する. $\mathbb{R}$ 上の実数値有界連続関数 f に対して

$$(T_t f)(x) = \int_{\mathbb{R}} f\left(e^{-t}x + \sqrt{1-e^{-2t}}y\right)\mu(dy), \quad \mu(dy) := \frac{e^{-\frac{y^2}{2}}}{\sqrt{2\pi}}dy$$

とおく. $p \geq 2, t \geq \dfrac{\log(p-1)}{2}$ のとき,

$$\|T_t f\|_{L^p(\mu)} \leq \|f\|_{L^2(\mu)}. \tag{4.19}$$

この超縮小性は場の量子論に現れるハミルトニアンの解析や, その後明らかになったが, 一般の拡散半群の評価に有効に用いられるなど様々な応用があることがわかっている. 一方, Stam の業績は孤立した結果であり, しばらく忘れられていたが, [2, 4] などの仕事の中でも取り上げられ現在では Stam-Gross の対数ソボレフ不等式と呼ばれることもある ([11]).

問題 12. Gibbs の補題を問題 6 の不等式を用いて示せ.

問題 13. (1) $f(x)$ を $[0,1]$ 上の確率密度関数とする.

$$\left(\int_0^1 |f(x)-1|dx\right)^2 \leq 2\int_0^1 f(x)\log f(x)dx \tag{4.20}$$

を示せ.
(2) (1) の不等式を用いて, Csiszár-Kullback-Pinsker の不等式を示せ.

5 Boltzmann's H-theorem, マルコフ連鎖, エントロピー

(希薄) 気体分子運動論を思い出そう. N 個の気体分子を考え, $(v_x^i(t), v_y^i(t), v_z^i(t))$ を i 番目の分子 $(1 \leq i \leq N)$ の時刻 t での速度とする. 速度 $v^i(t) = (v_x^1(t), v_y^i(t), v_z^i(t))$ はニュートンの運動方程式を満たすが N が巨大な数のため, 個々の分子の挙動を追うのは意味が無い. その代わり, Boltzmann は統計的に考えて微小領域 $dv_x dv_y dv_z$ に気体分子の速度が見出される確率 $f_t(v_x, v_y, v_z)dv_x dv_y dv_z$ を考え (したがって, $f_t(v_x, v_y, v_z)$ は確率密度関数である) 次の H-定理を導出した:

定理 5.1 (Boltzmann)**.**

$$H(t) = -\int_{\mathbb{R}^3} f_t(v_x, v_y, v_z)\log f_t(v_x, v_y, v_z)dv_x dv_y dv_z$$

とおくと $\dfrac{d}{dt}H(t) \geq 0$.

注意 5.2. (1) 統計力学におけるエントロピーは$kH(t)$ と表される. ここで$k(=1.38\times10^{-23}J\cdot K^{-1})$ は Boltzmann 定数である. したがって, これはエントロピーの増大を示している.
(2) H 定理については, 何人かの人が疑問の声を挙げた. それをいくつかあげる.
(**i**) Newton の運動方程式は, 質点 (気体分子) $\boldsymbol{x}(t)=(\boldsymbol{x_i}(t))_{i=1}^N$ がポテンシャル関数 U からの力を受けて運動している場合, 次のような微分方程式となる.

$$m_i\ddot{\boldsymbol{x}}_{\boldsymbol{i}}(t)=-\frac{\partial U}{\partial \boldsymbol{x_i}}(\boldsymbol{x}(t)) \tag{5.1}$$

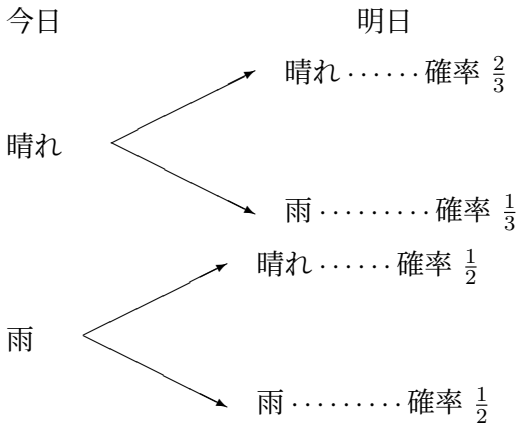
$$\boldsymbol{x_i}(0)=\boldsymbol{x_{i,0}}, \tag{5.2}$$

$$\dot{\boldsymbol{x}}_{\boldsymbol{i}}(0)=\boldsymbol{v_i}, \tag{5.3}$$

ここで $(\boldsymbol{x_{i,0}}),(\boldsymbol{v_i})$ は初期値と初速ベクトルを表す. また m_i は, 質点 $\boldsymbol{x_i}$ の質量を表す. 時間反転した解 $\boldsymbol{x}(-t)$ は, (5.1) の初期速度を逆向きに反転した $-\boldsymbol{v_i}$ ものの解であることがわかる. 気体分子の状態, エントロピーなどは 解 $\boldsymbol{x}$ で決まるはずであるが, $\boldsymbol{x}(t)$, $\boldsymbol{x}(-t)$ の両方のエントロピーが増大するのはありえない. これは矛盾である.
(**ii**) 2 番目の疑義は Poincaré の再帰定理に基づいている:
• ある時間 t_0 があって $\boldsymbol{x}(t_0)$ は初期状態 $\boldsymbol{x}(0)$ にいくらでも近くなる.
(このポアンカレの再帰定理は, ポテンシャル関数 U がある性質を満たしていれば数学的に証明できることである.) したがって, 時刻 t_0 におけるエントロピーは, 初期状態のエントロピーとそんなに違わないはずである. したがって, エントロピーが単調に増大するのはおかしい.
上記のパラドックスが発生するのは気体分子全体の統計的扱いにあるのだが、それはここでは説明しない (できない). この話題に関係する確率モデルについては [5] を参照せよ.

これから, マルコフ連鎖 (Markov chain) と呼ばれる確率的な時間発展 (確率過程) を考え, そのエントロピーが確かに増大することを示し, その応用として, 定常状態への収束を示そう.
次のような例を考える.

例 5.3. ある町の天気の変化に対する次のような統計データがあるとする. (状況を単純にするため曇りなどは考えない)



今日が (=0 日) 晴れとして, n 日目が晴れである確率はいくらか?

解 p_n を n 日目に晴れになる確率とする. すると n 日目に雨になる確率 q_n は $q_n = 1-p_n$. (p_n, q_n) は前日の晴れ, 雨の確率 (p_{n-1}, q_{n-1}) からの推移を考慮して次のように定まる.

$$(p_n, q_n) = (p_{n-1}, q_{n-1}) \begin{pmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}. \tag{5.4}$$

したがって

$$(p_n, q_n) = (1, 0) \begin{pmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}^n.$$

問題 14. 上記の例題で, $\lim_{n\to\infty} p_n, \lim_{n\to\infty} q_n$ を求めよ.

定義 5.4 (有限状態のマルコフ連鎖)**.** N 個の状態からなる状態空間 $E = \{1, \dots, N\}$ を考える. 時刻 1 経過すると状態が次の規則で確率的に変化する E 上の確率過程 (マルコフ連鎖) を考えよう. 非負の数の組 $\{p_{ij}\}_{i,j\in E}$ で次の式を満たすものを考える.

$$\sum_{j=1}^{N} p_{ij} = 1 \qquad \text{for all } 1 \le i \le N. \tag{5.5}$$

p_{ij} を (i, j) 成分とする N 次正方行列 $P = (p_{ij})$ は確率行列と呼ばれる行列である. p_{ij} は, 時刻 1 経過すると状態 i から j へ推移する確率を表している. 時刻 n に状態 i にいる確率が π_i $(1 \le i \le N)$ のとき, 時刻 $n+1$ に 状態 j に存在する確率は $\sum_{i=1}^{N} \pi_i p_{ij}$ であると考えられる. すなわち, 初期状態の確率分布が $\pi(0) = (\pi_1, \dots, \pi_N)$ の時, 時刻 n に状態 i に存在する確率を $\pi_i(n)$ $(1 \le i \le N)$ とし $\pi(n) = (\pi_1(n), \dots, \pi_N(n))$ とおくと

$$\pi(n) = \pi(0)P^n \tag{5.6}$$

となる確率過程がマルコフ連鎖である. P^n は行列の n 乗であり, 以下, $p_{ij}^{(n)}$ は P^n の (i, j) 成分を表す.

次の定理を証明しよう.

定理 5.5. 次を仮定する.

(A1) すべての $1 \le j \le N$ に対して $\sum_{i=1}^{N} p_{ij} = 1$.

(A2) (マルコフ連鎖のエルゴード性) ある時刻 $n_0 \in \mathbb{N}$ が存在して, 任意の $i, j \in \{1, \dots, N\}$ に対し $p_{ij}^{(n_0)} > 0$.

このとき, 任意の初期分布 $\pi = (\pi_1, \dots, \pi_N)$ に対して,

$$\lim_{n\to\infty} \pi(n)_i = \frac{1}{N}. \qquad \text{for all } 1 \le i \le N. \tag{5.7}$$

注意 5.6. (1) $(\pi(n)_i)_{i=1}^N = \pi(n) = \pi P^n$.

(2) (A1) が成立していなくても (A2) があれば, $\lim_{n\to\infty} \pi_i(n)$ は存在し, すべての i について極限は正と言える.

(3) (A1) は

$$(1, \dots, 1) = (1, \dots, 1)P$$

と同値. また, 任意の $1 \le i, j \le N$ に対して, $p_{ij} = p_{ji}$ ならば (A1) が成立する.

この定理をエントロピーと補題 3.2 を用いて証明しよう.

$$H(\pi) = -\sum_{i=1}^{N} \pi_i \log \pi_i$$

と定義したことを思い出そう.

補題 5.7. N 次正方行列 $Q = (q_{ij})$ が次の (a), (b) を満たすとする.

(a) Q は確率行列である. すなわち任意の $1 \leq i, j \leq N$ に対して, $q_{ij} \geq 0$ であり, すべての i について $\sum_{j=1}^{N} q_{ij} = 1$.

(b) すべての $1 \leq j \leq N$ について, $\sum_{i=1}^{N} q_{ij} = 1$.

このとき, 任意の E 上の確率分布 π に対し, $H(\pi Q) \geq H(\pi)$. さらに, すべての $1 \leq i, j \leq N$ について, $q_{ij} > 0$ ならば, $\pi \neq (1/N, \ldots, 1/N)$ のとき, $H(\pi Q) > H(\pi)$.

Proof.

$$\begin{aligned} H(\pi Q) &= -\sum_{i=1}^{N} (\pi Q)_i \log (\pi Q)_i \\ &= -\sum_{i=1}^{N} \left(\sum_{k=1}^{N} \pi_k q_{ki} \right) \log \left(\sum_{k=1}^{N} \pi_k q_{ki} \right). \end{aligned} \tag{5.8}$$

$\sum_{k=1}^{N} q_{ki} = 1$ に注意する. 補題 3.2 を $g(x) = x \log x$, $m_k = q_{ki}, x_k = \pi_k$ の場合に適用し,

$$\left(\sum_{k=1}^{N} \pi_k q_{ki} \right) \log \left(\sum_{k=1}^{N} \pi_k q_{ki} \right) \leq \sum_{k=1}^{N} q_{ki} \pi_k \log \pi_k. \tag{5.9}$$

(5.9) で両辺を i で和を取り, $\sum_{i=1}^{N} q_{ki} = 1$ より $H(\pi Q) \geq H(\pi)$. 次の等号成立条件の主張はやはり, 補題 3.2 から従う. □

この補題を用い, 定理 5.5 を証明しよう.

定理 5.5 の証明. 補題 5.7 の最初の命題からエントロピー $H(\pi(n))$ は単調非減少であることに注意する. さて, $\pi(n)$ は $\mathbb{R}^N$ の有界閉集合上を動くので, 集積点がある. すなわち, 部分列 $\{\pi(n(k))\}_{k=1}^{\infty}$ と $x = (x_1, \ldots, x_N) \in \mathbb{R}^N$ が存在して $\lim_{k\to\infty} \pi(n(k)) = x$. $x = (1/N, \ldots, 1/N)$ を示せばよい. $x \neq (1/N, \ldots, 1/N)$ として矛盾を導く. x は E 上の確率分布であり, $H(x) = \lim_{k\to\infty} H(\pi(n(k)))$ が成立する. P^{n_0} は補題 5.7 の (a), (b) を満たし, P^{n_0} の成分はすべて正. したがって 補題 5.7 より, $x \neq (1/N, \ldots, 1/N)$ ならば, $H(xP^{n_0}) > H(x)$. しかし

(i) $xP^{n_0} = \lim_{k\to\infty} \pi P^{n(k)+n_0}$,

(ii) $H(xP^{n_0}) = \lim_{k\to\infty} H(\pi P^{n(k)+n_0})$,

(iii) 任意の k に対して, $k' > k$ が存在して $n(k) + n_0 < n(k')$. したがって,

$$H(\pi P^{n(k)+n_0}) \leq H(\pi P^{n(k')})$$

より $H(xP^{n_0}) \leq H(x)$. これは矛盾である. したがって $x = (1/N, \ldots, 1/N)$. □

参考文献

[1] S. Aida, Tunneling for spatially cut-off $P(\phi)_2$-Hamiltonians. J. Funct. Anal. 263 (2012), no. 9, 2689–2753.

[2] A. Barron, Entropy and the central limit theorem. Ann. Probab. 14 (1986), no. 1, 336–342.

[3] A. Barron and J. Oliver, Fisher information inequalities and the central limit theorem. Probab. Theory Related Fields 129 (2004), no. 3, 391–409.

[4] E. Carlen, Super additivity of Fisher's Information and logarithmic Sobolev inequalities, Journal of Functional Analysis, **101** (1991), 194–211.

[5] 内山耕平, 舟木直久著, ミクロからマクロへ, 2. 格子気体の流体力学極限, シュプリンガーフェアラーク東京, 2002 年

[6] L. Gross, Logarithmic Sobolev inequalities. Amer. J. Math. 97 (1975), no. 4, 1061–1083.

[7] A.I. Khinchin, Mathematical foundations of information theory, Dover books on advanced mathematics, 1957.

[8] P.L. Lions and G. Toscani, A strengthened central limit theorem for smooth densities, Journal of Functional Analysis, **129** (1995), 148–167.

[9] C.E. Shannon and W. Weaver, The Mathematical Theory of Communication. Univ. of Illinois Press, Urbana, Ill., 1949.

[10] A.J. Stam, Some inequalities satisfied by the quantities of information of Fisher and Shannon, Information and Control **2** (1959), 101–112.

[11] C. Villani, Topics in optimal transportation. Graduate Studies in Mathematics, 58. American Mathematical Society, Providence, RI, 2003.

[12] D. Williams, Probability with martingales, Cambridge Mathematical Textbooks, 1991.

[13] K. Yoshida, Markoff process with a stable distribution, Proc. Imp. Acad. Tokyo, **16** (1940), 43–48.

[14] 熊谷 隆, 確率論, 共立出版株式会社.

[15] 西尾真喜子, 確率論, 実教出版.

[16] ヤグロム, 情報理論入門, みすず書房.(邦訳)