第 7 回
学部 3 年後期ゼミナール発表資料

青山和樹

2024 年 12 月 2 日

発表の目的

テキスト [1] の

- 4.1 Markov chains
- 4.2 Entropy rate

について発表する．

目次

チャプター 3 の漸近等分割性は独立同分布（i.i.d.）なランダムな $X_1, X_2, \ldots, X_n$ を記述するのに, 平均して $nH(X)$ ビットあれば十分なことを確立した. ここでは確率変数が互いに依存し, 特に定常過程（stationary process）を形成している場合を考える.

1 Markov chains

確率過程（stochastic process）$\{X_i\}$ は確率変数の添え字付き列である. 一般に確率変数間には任意の依存関係が存在し得る. またこの過程は, 同時確率関数 $\mathbb{P}[(X_1, X_2, \ldots, X_n) = (x_1, x_2, \ldots, x_n)] = p(x_1, x_2, \ldots, x_n), (x_1, x_2, \ldots, x_n) \in \mathcal{X}^n$ for $n = 1, 2, \ldots$ によって, 特徴づけられる.

定義 1.1 (定常な確率過程)

任意の n と任意のシフト l, 全ての $x_1, x_2, \ldots, x_n \in X$ について

$$\mathbb{P}[X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n] = \mathbb{P}[X_{l+1} = x_1, X_{l+2} = x_2, \ldots, X_{l+n} = x_n] \tag{1}$$

が成り立つとき, 定常であると言う. つまり, 確率過程は確率変数列の任意の部分集合の同時分布がそれぞれ時間インデックスのシフトに関して不変であるときである.

特に, 各確率変数が直前の確率変数にのみ依存し, それ以前の他のすべての確率変数から条件付きで独立している場合, このような過程をマルコフ（Markov）であると言う.

定義 1.2 (マルコフ連鎖)

離散確率過程（discrete stochastic process）$X_1, X_2, \ldots$ が全ての $x_1, x_2, \ldots, x_n \in X$ for $(n = 1, 2, \ldots)$ について,

$$\mathbb{P}[X_{n+1} = x_{n+1} | X_n = x_n, X_{n-1} = x_{n-1} \ldots, X_1 = x_1] = \mathbb{P}[X_{n+1} = x_{n+1} | X_n - x_n] \tag{2}$$

が成り立つとき, マルコフ連鎖（Markov cahin）またはマルコフ過程（Markov process）であると言う.

この場合, 確率変数の同時確率関数は

$$p(x_1, x_2, \ldots, x_n) = p(x_1)p(x_2|x_1)p(x_3|x_2) \cdots p(x_n|x_{n-1}) \tag{3}$$

と書ける.

定義 1.3 (時不変なマルコフ連鎖)

マルコフ連鎖において, すべての $n = 1, 2, \ldots$ について

$$\mathbb{P}[X_{n+1} = b | X_n = a] = \mathbb{P}[X_2 = b, X_1 = a], \forall a, \forall b \in X \tag{4}$$

が成り立つとき, 時不変であるという. つまり, マルコフ連鎖が時間に依存しない場合, すなわち条件付確率 $p(x_{n+1}|x_n)$ が n に依存しない場合である.

注意 1.4 今後, マルコフ連鎖が特に明記されていない場合, 時不変であると仮定する.

もし $\{X_i\}$ がマルコフ連鎖である場合, X_n は時刻 n における「状態」と呼ばれる. 時不変のマルコフ連鎖は, 初期状態と確率遷移行列（probablility transition matrix）$P = [P_{ij}], i, j \in \{1, 2, \ldots, m\}$, where $P_{ij} = \mathbb{P}[X_{n+1} = j | X_n = i]$ によって特徴づけられる.

解説 1.5 マルコフ連鎖において, 任意の状態から他の任意の状態へ有限のステップで正の確率で到達可能な場合, マルコフ連鎖は「既約（irreducible）」であると言う. また, ある状態からその状態自身へ戻る経路の長さの最大公約数が 1 である場合, そのマルコフ連鎖は非周期的（aperiodic）であると言う.

時刻 n における確率変数の確率関数を $p(x)$ とすると, 時刻 $n+1$ における確率関数は次のように表される. [*1]

$$p(x_{n+1}) = \sum_{x_n} p(x_n) P_{x_n x_{n+1}} \tag{5}$$

例 1.6 例えば, さいころの目が 1, 2, 3, 4, 5, 6 の 6 つの状態を持つとして

- x_n：時刻 n に出たサイコロの目.
- x_{n+1}：時刻 $n+1$ に出るサイコロの目.
- $p(x_n)$：時刻 n に x_n が出る確率.
- $P_{x_n x_{n+1}}$：目 x_n が出た後に目 x_{n+1} が出るか確率（遷移確率）.

時刻 $n+1$ に x_{n+1} が出る確率は, 時刻 n に x_n が出て, それが X_{n+1} に遷移するすべての可能性を考慮して計算する.

注意 1.7 時刻 $n+1$ における分布が時刻 n と同じであるような状態の分布を, 定常分布（stationary distribution）と呼ぶ.

定常分布と呼ばれる理由は, マルコフ連鎖の初期状態が定常分布に従って選ばれる場合, そのマルコフ連鎖が定常過程を形成するためである. 有限状態のマルコフ連鎖が既約かつ非周期的である場合, 定常分布は一意であり, 任意の初期分布から始めても, 時刻 n における分布 X_n は $n \to \infty$ の時, 定常分布に収束する.

[*1]

- 式の意味
  この式は, 次の時刻 $n+1$ における状態 x_{n+1} の確率を計算するために, 現在の時刻 n における全ての状態 x_n の寄与を足し合わせる, という操作をしています. 式を分解して考えると
  1. $p(x_n)$：時刻 n に状態 x_n にいる確率.
  2. $P_{x_n x_{n+1}}$：状態 x_n から x_{n+1} に遷移する確率.
  3. $p(x_n) \cdot P_{x_n x_{n+1}}$：時刻 $n+1$ に x_{n+1} にいるために, 時刻 n に x_n にいる確率関数の貢献.
  4. $\sum_{x_n}$：すべての x_n について足し合わせる.

  時刻 $n+1$ に x_{n+1} にいる確率を求めるためには, 全ての可能な状態 x_n に対して足し合わせる必要があります.

例 1.8 以下に示す遷移確率行列を持つ図 1 のような 2 状態マルコフ連鎖を考える．

$$P = \begin{bmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{bmatrix} \tag{6}$$

定常分布を，それぞれ状態 1 と状態 2 の定常確率を成分とするベクトル μ で表すとする．定常確率は，方程式 $\mu P = \mu$ を解くか，より単純に確率のバランスをとることによって求めることができる．定常分布の場合，状態遷移グラフのどのカット・セットでも，純確率流（net probablility flow）はゼロである．これを図 1 に当てはめると，次のようになる．

$$\mu_1 \alpha = \mu_2 \beta \tag{7}$$

また確率の総和は 1 であるため

$$\mu_1 = \frac{\beta}{\alpha+\beta},\ \mu_2 = \frac{\alpha}{\alpha+\beta} \tag{8}$$

もしマルコフ連鎖の初期状態がこの定常分布に従って選ばれる場合，結果として得られる過程は定常的になる．

解説 1.9 時刻 n における状態 X_n のエントロピー $H(X_n)$ は以下の式で与えられる．

$$H(X_n) = H\left(\frac{\beta}{\alpha+\beta}, \frac{\alpha}{\alpha+\beta}\right) \tag{9}$$

これは過程全体のエントロピー増加率 $H(X_1, X_2, \ldots, X_n)$ を表すものではない．確率変数 X 同士の依存関係が影響を与え増加の割合を減らすためである．

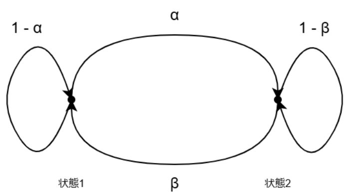

図 1　2 状態マルコフ連鎖

2　Entropy rate

確率変数の列が n 個ある場合，その列のエントロピーが n に伴ってどのように増加するかを考える．

定義 2.1 (エントロピーレート)

確率過程 $\{X_i\}$ のエントロピーレート（entropy rate）は次のように定義される．

$$H(\mathcal{X}) = \lim_{n\to\infty} \frac{1}{n} H(X_1, X_2, \ldots, X_n) \tag{10}$$

ただし，極限が存在する時

例 2.2 タイプライター

m 個の等確率の文字を出力する場合を考える．このタイプライターは, 長さ n の m^n 通りの系列を生成でき, それらはすべて等確率である．よって, $H(X_1, X_2, \ldots, X_n) = \log m^n$ となり, エントロピーレートは $H(\mathcal{X}) = \log m$ (一文字あたりのビット数) となる．

解説 2.3 タイプライターが m 個の文字を等確率で生成することから, 各文字のエントロピーは

$$
\begin{aligned}
H(X) &= -\sum_{i=1}^{m} p(x_i) \log p(x_i) && (11)\\
&= -\sum_{i=1}^{m} \frac{1}{m} \log \frac{1}{m} && (12)\\
&= \log m && (13)
\end{aligned}
$$

独立同分布であるから長さ n の文字列全体のエントロピーは

$$H(X_1, X_2, \ldots, X_n) = n \cdot H(X) = n \log m \tag{14}$$

よって, エントロピーレートは

$$H(\mathcal{X}) = \lim_{n \to \infty} \frac{1}{n} H(X_1, X_2, \ldots, X_n) = \log m \tag{15}$$

例 2.4 独立同分布の確率変数

$X_1, X_2, \ldots$ が i.i.d である場合, 同時エントロピーは

$$H(X_1, X_2, \ldots, X_n) = nH(X) \tag{16}$$

と表せ, エントロピーレートは

$$H(\mathcal{X}) = \lim \frac{H(X_1, X_2, \ldots, X_n)}{n} = \lim \frac{nH(X_1)}{n} = H(X_1) \tag{17}$$

と表せる．これは, シンボル当たりのエントロピーレートとして期待させる．

例 2.5 独立だが同分布ではない確率変数

$X_1, X_2, \ldots$ が独立であるが, 同一分布ではない場合, 同時エントロピーは

$$H(X_1, X_2, \ldots, X_n) = \sum_{i=1}^{n} H(X_i) \tag{18}$$

と表せる．

しかし, $H(X_i)$ がすべて等しくはなく, $\frac{1}{n} \sum H(X_i)$ の極限が存在しないように, $X_1, X_2, \ldots$ 上の分布の列を選ぶこともできる．例として, $p_i = \mathbb{P}[X_i = 1]$ が定数ではなく i の関数であるランダムな 2 進数列を考え, 極

限が存在しないように調整する．例えば $k = 0, 1, \ldots$ として

$$p_i = \begin{cases} 0.5 & 2^k < \log\log i \leq 2^k + 1 \text{ のとき}, \\ 0 & 2^k + 1 < \log\log i \leq 2^k + 2 \text{ のとき}. \end{cases} \tag{19}$$

のように定義する．すると，指数的に長い $H(X_i) = 0$ の後に続いて，$H(X_i) = 1$ となるような，任意に長い広がりが生じる．したがって，$H(X_i)$ は 0 と 1 の間を振動し，極限を持たない．そのため，この確率過程で $H(\mathcal{X})$ は定義されない．

定義 2.7 (関連するエントロピーレート)

$$H'(\mathcal{X}) = \lim_{n\to\infty} H(X_n|X_{n-1}, X_{n-2}, \ldots X_1) \tag{20}$$

ただし，極限が存在する時

解説 2.8 2 つの量 $H(\mathcal{X}), H'(\mathcal{X})$ はエントロピーレートに関する 2 つの異なる概念に対応する．$H(\mathcal{X})$ は n 個の確率変数に対するシンボル当たりのエントロピーを表し，$H'(\mathcal{X})$ は過去が与えられた場合の最後の確率変数に関する条件付エントロピーを表す．

定理 2.9 (チェザロ平均（cesáro mean）)

もし $a_n \to a$ かつ $b_n = \frac{1}{n}\sum_{i=1}^{n} a_i$ ならば $b_n \to a$

証明 2.10 2.9 の証明（概要）

列 $\{a_n\}$ のほとんどの項が最終的に a に近くなるため，b_n（最初の n 項の平均）は最終的に a 近づく．

証明 2.11 2.9 の形式的な証明

任意の $\epsilon > 0$ として，$a_n \to a$ であるから，ある数 $N(\epsilon)$ が存在し，すべての $n \geq N(\epsilon)$ に対して

$$|a_n - a| \leq \epsilon \tag{21}$$

を満たす．したがって，

$$
\begin{aligned}
|b_n - a| &= |\frac{1}{n}\sum_{i=1}^{n}(a_i - a)| && (22)\\
&\le \frac{1}{n}\sum_{i=1}^{n}|a_n - a| && (23)\\
&= \frac{1}{n}\sum_{i=1}^{N(\epsilon)}|a_n - a| + \frac{1}{n}\sum_{i=N(\epsilon)+1}^{n}|a_n - a| && (24)\\
&\le \frac{1}{n}\sum_{i=1}^{N(\epsilon)}|a_n - a| + \frac{1}{n}\sum_{i=N(\epsilon)+1}^{n}\epsilon && (25)\\
&= \frac{1}{n}\sum_{i=1}^{N(\epsilon)}|a_n - a| + \frac{n-N(\epsilon)}{n}\epsilon && (26)\\
&\le \frac{1}{n}\sum_{i=1}^{N(\epsilon)}|a_n - a| + \epsilon && (27)
\end{aligned}
$$

がすべての $n \ge N(\epsilon)$ に対して成り立つ．最初の項は $n \to \infty$ に従って，0 に近づくため，n を十分大きくとることによって，$|b_n - a| \le 2\epsilon$ とすることができる．よって，$n \to \infty$ に対して $b_n \to a$ となる． □

定理 2.12

定常過程において，条件付エントロピー $H(X_n|X_{n-1}, X_{n-2}, \ldots, X_1)$ は n に対して単調非増加であり，極限 $H'(\mathcal{X})$ を持つ．

証明 2.13 条件付エントロピーの定義から条件付けはエントロピーを減らし，定常性から時間インデックスに関するシフトに対して不変であるから，

$$
\begin{aligned}
H(X_{n+1}|X_n, X_{n-1}, \ldots, X_1) &\le H(X_{n+1}|X_n, X_{n-1}, \ldots, X_2) && (28)\\
&= H(X_n|X_{n-1}, X_{n-2}, \ldots, X_1) && (29)
\end{aligned}
$$

が成り立つ．よって $H(X_n|X_{n-1}, X_{n-2}, \ldots, X_1)$ は非負の減少列であるため，極限値 $H(\mathcal{X})$ を持つ．

定理 2.14

定常過程に対して，式（10）および式（20）の極限は等しい．

$$
H(\mathcal{X}) = H'(X). \tag{30}
$$

証明 2.15 チェイン則によって

$$
\frac{H(X_1, X_2, \ldots, X_n)}{n} = \frac{1}{n}\sum_{i=1}^{n} H(X_i|X_{i-1}, \ldots, X_1), \tag{31}
$$

つまり, エントロピーレートは条件付きエントロピーの時間平均となる. 条件付きエントロピーが極限値 H' に収束することから, 定理（2.9）を用いることによって, それら全体にわたる平均は極限値を持ち, それは各項の極限値 H' に等しい. よって定理（2.12）より

$$H(\mathcal{X}) = \lim_{n\to\infty} \frac{1}{n} H(X_1, X_2, \ldots, X_n) = \lim_{n\to\infty} H(X_n|X_{n-1}, \ldots, X_1) \tag{32}$$

$$= H'(\mathcal{X}) \quad \square \tag{33}$$

解説 2.16 確率過程のエントロピーレートの重要性は定常エルゴード過程（stationary ergodic process）に対する漸近等分割性に由来する. テキスト [1]16.8 にて任意の定常エルゴード過程に対して

$$\frac{1}{n} \log p(X_1, X_2, \ldots, X_n) \to H(\mathcal{X}) \; (n \to \infty) \tag{34}$$

が確率 1 で成り立つことを証明する. これは第 3 章の定理を定常エルゴード過程に拡張することを容易にする.

第 3 章で i.i.d. の場合に行ったように, 典型集合を同様の方法で定義でき, 以下が示される.

- 典型集合の確率は 1 に近い
- 長さ n の典型系列は訳 $2^{nH(\mathcal{X})}$ 個である.
- 各典型系列の確率は $2^{-nH(\mathcal{X})}$ に近い.

したがって, 長さ n の典型系列はほぼ $nH(\mathcal{X})$ ビットで表現できる. これはエントロピーレートが定常エルゴード過程における平均記述長であることを示している.

エントロピーレートはすべての定常過程に対して明確に定義されており, 特に, マルコフ連鎖においてはエントロピーレートの計算が非常に容易である.

解説 2.17 マルコフ連鎖のエントロピーレート

定常マルコフ連鎖におけるエントロピーレートは

$$H(\mathcal{X}) = H'(\mathcal{X}) = \lim_{n\to\infty} H(X_n|X_{n-1}, \ldots, X_1) = H(X_n|X_{n-1}) \tag{35}$$

$$= H(X_2|X_1) \tag{36}$$

で与えられる. ここで, 条件付きエントリーは与えられた定常分布によって計算され, 定常分布 μ は

$$\mu_j = \sum_i \mu_{ij} P_{ij} \text{（全ての状態 j に対して）} \tag{37}$$

の解として与えられる.

定理 2.18

$\{X_i\}$ を定常分布が μ で確率遷移行列が P であるような定常マルコフ連鎖を考える．もし $X_1 \sim \mu$ であるなら, エントロピーレートは

$$H(\mathcal{X}) = -\sum_{i,j} \mu_i P_{ij} \log P_{ij} \tag{38}$$

で表せる

証明 2.19

$$\begin{aligned} H(\mathcal{X}) &= H(X_2|X_1) && (39)\\ &= \sum_i \mathbb{P}[X_1 = i] \cdot H(X_2|X_1 = i) && (40)\\ &= \sum_i \mathbb{P}[X_1 = i] \cdot \left(\sum_j -P_{ij} \log P_{ij}\right) && (41)\\ &= \sum_i \mu_i \left(\sum_j -P_{ij} \log P_{ij}\right) \quad \square && (42) \end{aligned}$$

証明 2.20 2 状態のマルコフ連鎖

図 1 のエントロピーレートは

$$H(\mathcal{X}) = H(X_2|X_1) = \frac{\beta}{\alpha+\beta} H(\alpha) + \frac{\alpha}{\alpha+\beta} H(\beta) \tag{43}$$

である.

■備考　もし, マルコフ連鎖が既約かつ非周期的であるなら,

1. 状態上の定常分布は一意である.
2. 任意の初期分布は $n \to \infty$ で定常分布に収束する.

初期分布が定常分布でなくても, 長期的な挙動に基づくエントロピーレートは式（35）および式（38）に定義されるよう, $H(\mathcal{X})$ である.

参考文献

[1] T.M.Cover and J.A.Thomas, Elements of Information Theory, Second Edition, John Wiley & Sons, 2006.