

6. 記憶のある情報源

今回は自己の過去の振る舞いに基づいて情報源記号生成確率分布が変わる記憶のある情報源のモデル化を取り上げて、その性質を調べる。

6.1 記憶のある情報源のモデル化の例

まず、記憶のある情報源のモデル化をどのように行うか、例題を用いて説明する。

【例題】{直球, 変化球}を投げ分けられる投手がいたとしよう。どの球種も同じく $1/2$ の確率で投げることがわかっていたとすると、次に来るボールがどれであるかは等しく予想しがたいが、直球、変化球と続いた後に必ずと言っていいくらい、変化球を投げるといふ癖があるとわかっていると、直球、変化球の後には変化球に山を張ればよいということになる。

この投手を各時刻で{直球, 変化球}の球種を発生する情報源としてモデル化したとき、先行する投球内容を考慮に入れない場合が無記憶情報源モデル、考慮に入れる場合が記憶のある情報源モデルとなる。

この投手の行動の確率論に基づくモデル化をしてみよう。投球するのは時刻 $0, 1, 2, \dots$ とする。時刻 t の投球を確率変数 X_t で表す。普通の変数とは異なり、確率変数 X_0, X_1, X_2, \dots の値は定められた確率により試行（この例では「試合」に相当するとでも思っておけばよい）ごとに決まるものとして定式化されている。試行は第1回、第2回、...と繰り返されていく。その都度、確率変数の値が決まる。例えば、第1回目の試行では、 X_0 の値は直球、 X_1 の値は直球、...；試行2では、 X_0 の値は変化球、 X_1 の値は直球、...；と続くイメージだ。概して、たくさん回の試行を行うと、各確率変数の値を決める確率が浮き上がってくる。

確率変数を用いた議論をするときは、こうした試行を引き起こす原因となる確率モデルを想定し、次の1回の試行を考察の対象とする。

記憶のない情報源モデルを使うならば、各試行において各回の投球は独立に決まるものとみなされる。投手の行動は、

直球を投げる確率=0.42

変化球を投げる確率=0.58

といった具合にモデル化される。これに従って次の試行が行われ、その結果、

X_0 の値は変化球、 X_1 の値は直球、...

というように、その試行における確率変数の値が決まる。

これに対して、記憶のある情報源モデルを使う場合は、各回の球種の確率は、先行する（あらかじめ定められた回数、例えば3回の）有限回の投球内容、あるいは、投手の「心の状態」

(技術的には内部状態という)によって決まるものとしてモデル化が行われる。後で示すように、次の球種が(有限個の)内部状態によって決まるというモデル化(一般マルコフモデルによるモデル化)は、次の球種の確率が先行する投球内容で決まるというモデル化(多重マルコフモデルによるモデル化)より強力であり、より複雑な球種確率の決定法を記述できる。

まず、次の球種が先行する有限回の投球内容によって決まる、というケースのモデル化を示そう。例えば、「直球、変化球の後には、確率0.8で変化球が投げられる」といったモデル化を行う。これを例えば

$$P_{X_{17}|X_{15},X_{16}}(\text{変}|直,変) = 0.8$$

と記述する。これは、与えられた試行において、確率変数 X_{15} の値が「直」(直球を表す)、確率変数 X_{16} の値が「変」(変化球を表す)となったとき、確率変数 X_{17} の値が変化球になる確率が0.8であると言っている。|は条件付き確率であることを示し、その右側に並べられた確率変数の値が決まった場合に、|の左側に置かれた確率変数の確率を問題にすることを示唆している。また、|の右側の確率変数の並べ方は時刻順としている(ここは、教科書と逆になっている)。

さて、上では陽に述べてこなかったが、 X_i の i は $-\infty$ 、ないしは、0あるいは1から始まり、果てしなく続くものとして抽象化している。また、さも当たり前のように、「直球、変化球の後には、変化球が投げられる確率はいつでも同じ」であるかのように、つまり、

$$\begin{aligned} & \dots \\ & = P_{X_{16}|X_{14},X_{15}}(\text{変}|直,変) \\ & = P_{X_{17}|X_{15},X_{16}}(\text{変}|直,変) \\ & = P_{X_{18}|X_{16},X_{17}}(\text{変}|直,変) \\ & \dots \\ & = 0.8 \end{aligned}$$

としたがそうでないこともあるかもしれない。「直球、変化球の後には、変化球が投げられる確率はいつでも同じ」というのは(実際はそうでないかもしれないが、単純化によって問題を考えやすくするために導入した)一つの仮定である。このような仮定が導入された情報源を定常情報源という。定常情報源は、必ずしも記憶のない情報源を意味しないので注意しよう。上の例にあるように、「...がいつも同じ」の部分が定常に相当する。「直球、変化球の後には」のところが「記憶のある」に相当する。

定常情報源であるという仮定を導入したとたん、情報源を完全に記述することが容易になる。所与の情報源が定常情報源であり、 $P_{X_2|X_0,X_1}(\text{変}|直,変) = 0.8$ であるとモデルに記述したとたん、それはすべての i について $P_{X_{i+2}|X_i,X_{i+1}}(\text{変}|直,変) = 0.8$ であることを宣言したことになる。仮に、その投手の投球する球種の確率をその前に2回投球した内容に従って決定されるというモデルを採用したとすると、上と同様に、 $P_{X_2|X_0,X_1}(x_2|x_0,x_1) = 0.8$ の値を x_0,x_1,x_2

表 1. 情報源記号{直球,変化球}に対する条件確率 $P_{X_2|X_0,X_1}(x_2|x_0,x_1)$ の例

x_0	x_0	x_2	$P_{X_2 X_0,X_1}(x_2 x_0,x_1)$
直	直	直	0.6
直	直	変	0.4
直	変	直	0.2
直	変	変	0.8
変	直	直	0.5
変	直	変	0.5
変	変	直	0.6
変	変	変	0.4

の取りえるすべての値の組み合わせ, $2^3 = 8$ 通り与えればこのモデルの核心部分を定めたことになる.

一般に, X_n の値の出現確率が, 先行する m 個の連続する確率変数 $X_{n-1}, X_{n-2}, \dots, X_{n-m}$ がどう具体化されたかに依存して決まるものと規定した確率モデルを m 重マルコフモデルという.

この章の初めに取り上げていた例題は, 例えば, 表 1 のように規定されることになる. m 重マルコフモデルを完全に規定するためにはもう少し情報を与える必要がある. 上記のように条件付き確率を与えただけではこの投手の繰り出す投球の確率過程を完全に規定できないからである. 例えば, 上に与えた情報だけからは, $P_{X_{10}}$ (変)の値がどうなるか計算できない. m 重マルコフモデルの場合, 確率過程を完全に規定するためには, さらに表 2 のように, $P_{X_0}(x_0), P_{X_1|X_0}(x_1|x_0), \dots, P_{X_m|X_0, \dots, X_{m-1}}(x_m|x_0, \dots, x_{m-1})$ を規定するか, 結合確率

$$P_{X_0, \dots, X_{m-1}, X_m}(x_0, \dots, x_{m-1}, x_m)$$

を X_i の可能な具体化である x_i の取りえるすべての値について規定しなければならない.

表 2. 情報源記号{直球,変化球}に対する結合確率 $P_{X_0,X_1,X_2}(x_0,x_1,x_2)$ の例

x_0	x_1	x_2	$P_{X_0,X_1,X_2}(x_0,x_1,x_2)$
直	直	直	0.168
直	直	変	0.112
直	変	直	0.024
直	変	変	0.096
変	直	直	0.24
変	直	変	0.24
変	変	直	0.072
変	変	変	0.048

結合確率を表2のように与えると、この2重マルコフ情報源の性質を完全に与えたことになる。このモデルが所与の投手の妥当なモデルとすれば、この投手が直球、変化球のあとに変化球を投げる確率は、0.8であることになる。

上において $P_{X_0, X_1, X_2}(x_0, x_1, x_2)$ の値を8個指定したが、それらの合計は1でなければならないので、実質的に指定した値は7個である。これに対して、先に規定した条件付き確率による指定では、 $P_{X_2|X_0, X_1}(\text{直}|x_0, x_1) + P_{X_2|X_0, X_1}(\text{変}|x_0, x_1) = 1$ でなければならないので、実質的に指定した値は4個である。この場合はマルコフ情報源の性質を完全に規定するためには、さらに、3つの確率： $P_{X_0}(\cdot)$, $P_{X_1|X_0}(\cdot|\text{直})$, $P_{X_1|X_0}(\cdot|\text{変})$ （ここで、 \cdot は直か変のいずれか一方）を指定する必要があるので、結局、結合確率の場合と同じく実質的に7個の値を指定する必要がある。

ここからいろいろな確率を計算することができる。

$$\begin{aligned} P_{X_2|X_0, X_1}(\text{変}|\text{直}, \text{変}) &= \frac{P_{X_0, X_1, X_2}(\text{直}, \text{変}, \text{変})}{P_{X_0, X_1}(\text{直}, \text{変})} \\ &= \frac{P_{X_0, X_1, X_2}(\text{直}, \text{変}, \text{変})}{P_{X_0, X_1, X_2}(\text{直}, \text{変}, \text{直}) + P_{X_0, X_1, X_2}(\text{直}, \text{変}, \text{変})} \\ &= \frac{0.096}{0.024 + 0.096} \\ &= 0.8 \end{aligned}$$

1球目、2球目、3球目が変化球である確率はそれぞれ、

$$\begin{aligned} P_{X_0}(\text{変}) &= \sum_{x_1} \sum_{x_2} P_{X_0, X_1, X_2}(\text{変}, x_1, x_2) \\ &= P_{X_0, X_1, X_2}(\text{変}, \text{直}, \text{直}) + P_{X_0, X_1, X_2}(\text{変}, \text{直}, \text{変}) \\ &\quad + P_{X_0, X_1, X_2}(\text{変}, \text{変}, \text{直}) + P_{X_0, X_1, X_2}(\text{変}, \text{変}, \text{変}) \\ &= 0.24 + 0.24 + 0.072 + 0.048 \\ &= 0.6 \end{aligned}$$

$$\begin{aligned} P_{X_1}(\text{変}) &= \sum_{x_0} \sum_{x_2} P_{X_0, X_1, X_2}(x_0, \text{変}, x_2) \\ &= P_{X_0, X_1, X_2}(\text{直}, \text{変}, \text{直}) + P_{X_0, X_1, X_2}(\text{直}, \text{変}, \text{変}) \\ &\quad + P_{X_0, X_1, X_2}(\text{変}, \text{変}, \text{直}) + P_{X_0, X_1, X_2}(\text{変}, \text{変}, \text{変}) \\ &= 0.024 + 0.096 + 0.072 + 0.048 \\ &= 0.24 \end{aligned}$$

$$\begin{aligned}
P_{X_2}(\text{変}) &= \sum_{x_0} \sum_{x_1} P_{X_0, X_1, X_2}(x_0, x_1, \text{変}) \\
&= P_{X_0, X_1, X_2}(\text{直}, \text{直}, \text{変}) + P_{X_0, X_1, X_2}(\text{直}, \text{変}, \text{変}) \\
&\quad + P_{X_0, X_1, X_2}(\text{変}, \text{直}, \text{変}) + P_{X_0, X_1, X_2}(\text{変}, \text{変}, \text{変}) \\
&= 0.112 + 0.096 + 0.24 + 0.048 \\
&= 0.496
\end{aligned}$$

上記の値は、先行する球種のパターンを押しなべて計算した値である。3球目以降からは、先行する球種を考慮に入れた予測が可能であり、場合によってはより、予測しやすくなることがある。たとえば、1球目が直球、2球目が変化球であったという前提の下では、3球目に変化球である確率は、 $P_{X_2|X_0, X_1}(\text{変}|\text{直}, \text{変}) = 0.8$ となり、さらに、3球目に変化球であった場合は、次の球が変化球である確率は、

$$P_{X_3|X_1, X_2}(\text{変}|\text{変}, \text{変}) = P_{X_2|X_0, X_1}(\text{変}|\text{変}, \text{変}) = \frac{0.048}{0.048 + 0.072} = 0.4$$

となる（最後のステップでは定常性も利用している）。

今度は内部状態に基づく方法について紹介する。この方法では（有限個の）設けられている。投手は各時刻で $\{S_1, \dots, S_n\}$ のうちの一つの内部状態（いわば「心の状態」）にあり、状態ごとに定められた確率で投球球種と次の時点の心の状態が決められる。内部状態は、心の状態は顔色や表情・しぐさを通した外部から推測できるかもしれないし、そうでないかもしれない。今取り上げている例題でどのような内部状態を設けたらいいだろうか？

もし投手が直前とその前の時刻に投げた球種で次の球種の確率を変えているとしたら、直前とその前の時刻に投げた球種をそのまま内部状態として採用することができる。このアイデアによると、先に示した条件付き確率を用いた2重マルコフモデルは、先行する2球が直球-直球、直球-変化球、変化球-直球、変化球-変化球の場合に相当する4状態を設ければよい。そしてその間に状態遷移規則を導入する（図1）。

例えば、「直球、変化球の次に、確率0.8で変化球が投げられる」という部分に対応して、 $S_{\text{直変}}$ から確率0.8で「変化球」が投げられ、その結果、先行する2つの投球内容が変化球、変化球である状態にそうとする状態 $S_{\text{変変}}$ に遷移する。

6.2 記憶のある情報源モデル

次に、例題で用いたいくつかの概念の定式化を行う。定常情報源概念を与えた後、記憶の有無に応じた情報源記号の発生確率の算出法を与える。さらに、記憶のある情報源をモデル化する手法である、 m 重マルコフ情報源と、一般化されたマルコフ情報源を導入する。

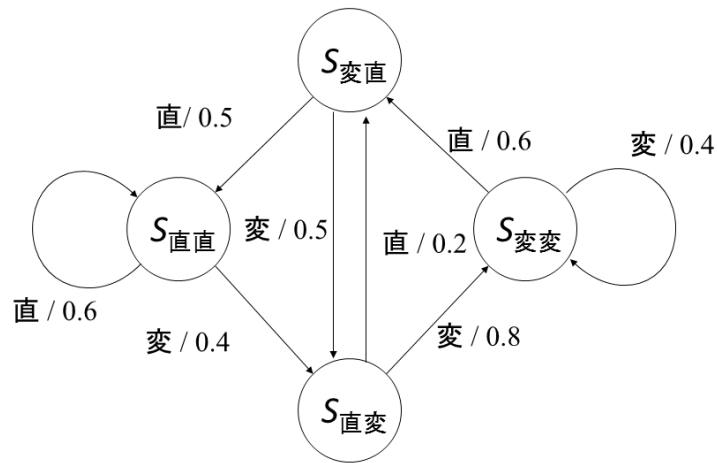


図 1. 2重マルコフモデルの例

(1) 定常情報源の概念の導入

情報源記号 $A = \{a_1, a_2, \dots, a_M\}$ をもつ情報源 S に対して、時間の経過に伴う S からの出力を X_0, \dots, X_{n-1} と表す (X_i は $\{a_1, a_2, \dots, a_M\}$ のどれかを取り得る確率変数)。

定常情報源：各時点における情報源記号の発生が同一の確率分布に従う。時間をずらしても統計的性質は変わらない。つまり、任意の正整数 n と i 、情報源記号の任意の元 x_0, x_1, \dots, x_{n-1} に対して、 $X_0 = x_0, X_1 = x_1, \dots, X_{n-1} = x_{n-1}$ となる確率と $X_i = x_0, X_{i+1} = x_1, \dots, X_{i+n-1} = x_{n-1}$ となる確率が等しい。すなわち、

$$P_{X_0, X_1, \dots, X_{n-1}}(x_0, x_1, \dots, x_{n-1}) = P_{X_i, X_{i+1}, \dots, X_{i+n-1}}(x_0, x_1, \dots, x_{n-1})$$

次に、記憶の有無に応じて情報源記号の発生確率の算出法を与える。

(a) 記憶のない定常情報源

記憶のない定常情報源の出力の結合確率分布は、各時点における情報源記号の発生が、他の時点とは独立である。

$$P_{X_0, \dots, X_{n-1}}(x_0, \dots, x_{n-1}) = \prod_{i=0}^{n-1} P_{X_i}(x_i)$$

(b) 記憶のある定常情報源モデル

X_0, \dots, X_{n-1} の統計的性質は、結合確率分布：

$$P_{X_0, \dots, X_{n-1}}(x_0, \dots, x_{n-1}) = [X_0 = x_0, X_1 = x_1, \dots, X_{n-1} = x_{n-1} \text{ となる確率}]$$

を与えれば完全に規定することができる。

結合確率として明示的に与えられていない確率はベイズの原理で求める。例えば、

$$P(x_2|x_0, x_1) = \frac{P(x_0, x_1, x_2)}{P(x_0, x_1)} = \frac{P(x_0, x_1, x_2)}{P(x_0, x_1, A) + P(x_0, x_1, B)}$$

(2) マルコフ情報源

記憶のある情報源で最も基本的なモデルである。 $m \leq n$ なる全ての n に対して、

$$P_{X_i|X_{i-1}, \dots, X_{i-n}}(x_i|x_{i-1}, \dots, x_{i-n}) = P_{X_i|X_{i-1}, \dots, X_{i-m}}(x_i|x_{i-1}, \dots, x_{i-m})$$

が成り立つとき、 m 重マルコフ情報源と呼ばれる。 先行する m 個の出力を記憶し、それに基づいて次に出力する情報源アルファベットの生起確率が決まる。

(3) 一般化されたマルコフ情報源

状態図によって規定される。 m 重マルコフ情報源は一般化されたマルコフ情報源としてモデル化できる。

2元{A, B}の3重マルコフ情報源に対する状態図は一般に図2のようになる。

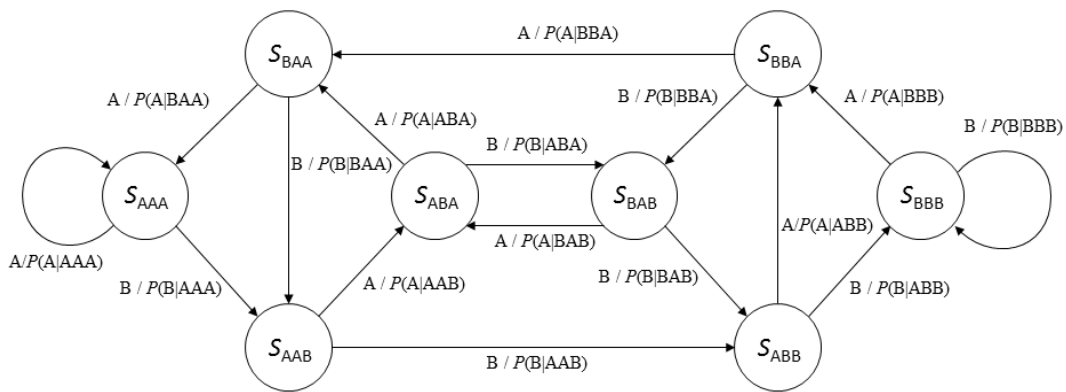


図2. 2元{A, B}の3重マルコフ情報源に対する状態図

一方、一般化されたマルコフ情報源は必ずしも m 重マルコフ情報源としてモデル化できない。 図3の例を見てみよう。 直前にAだけからなる出力系列が続く場合、 $S_0 \leftrightarrow S_1$ によるものか、 $S_2 \leftrightarrow S_2$ によるものか区別できない。 そのため、 $P(A|A, \dots, A)$ が0.7なのか、 0.1なのか、 0.2なのか、 決められない。

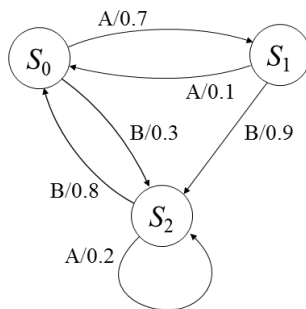


図3. m 重マルコフ情報源としてモデル化できない一般化されたマルコフ情報源の例

6.3 エルゴード性

次に、マルコフ情報源の性質を調べるための観点としてエルゴード性を導入する。

定常情報源のなかで、十分長い任意の出力系列にその情報源の統計的性質が完全に現れるものをエルゴード的な情報源という。換言すれば、エルゴード情報源では、その情報源と同一の情報源を無数に用意したときに、ある1時点の出力を同時に調べたものと、任意の一つの情報源を選んでその出力を永遠に観測したときの統計が一致する。

エルゴード情報源では、次の2つの量が一致する。第一は、集合平均：

$$\overline{f(x)} = \sum_{x \in A} f(x) P_X(x)$$

ここで、 $f(x)$: ある時刻における x の値、 $P_X(x)$: ある時刻において $X = x$ になる確率。

第二は、時間平均：

$$\langle f(x) \rangle = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} f(x_i)$$

【例】 $\{A, B\}$ を確率 $1/2$ で発生する記憶のない定常情報源はエルゴード情報源である。

これに対してAかBが確率 $1/2$ で選ばれた後はそれを出し続ける情報源はエルゴード情報源ではない (図4)。

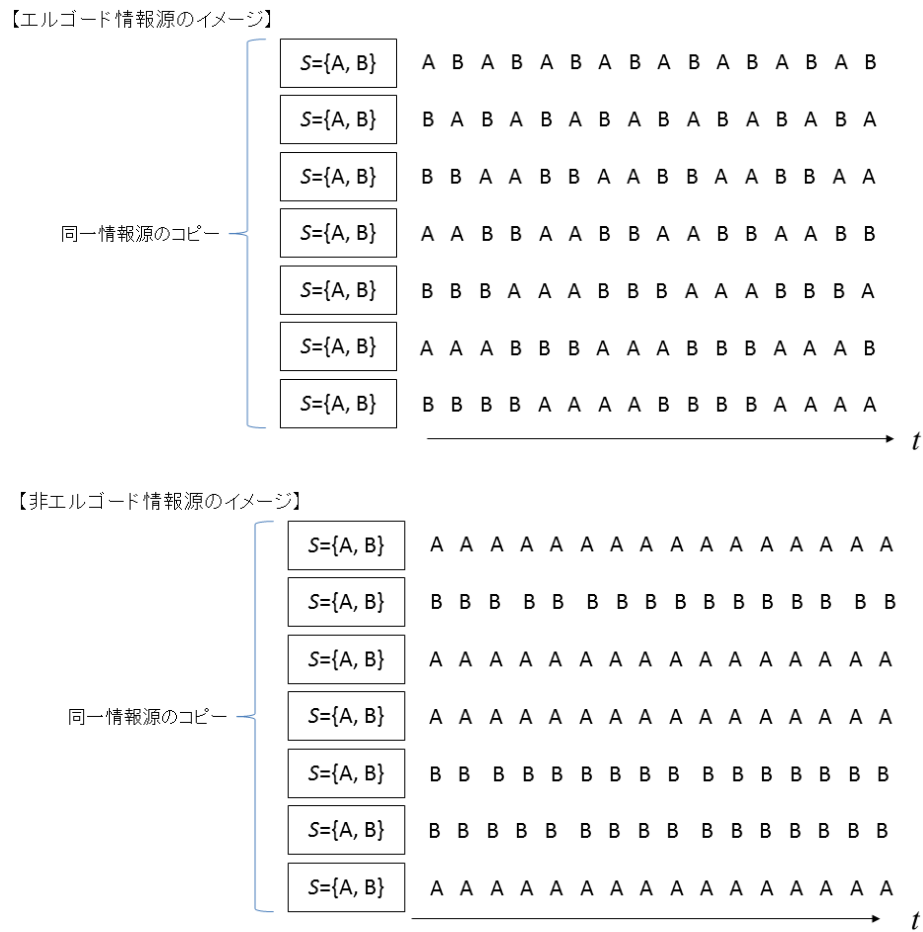


図 4. エルゴード情報源 vs 非エルゴード情報源

6.4 状態の分類

一般化されたマルコフ情報源の状態の分類を与える．正規マルコフ情報源という概念も導入する．図 5 の状態図で表された一般化されたマルコフ状態源について考えてみよう．

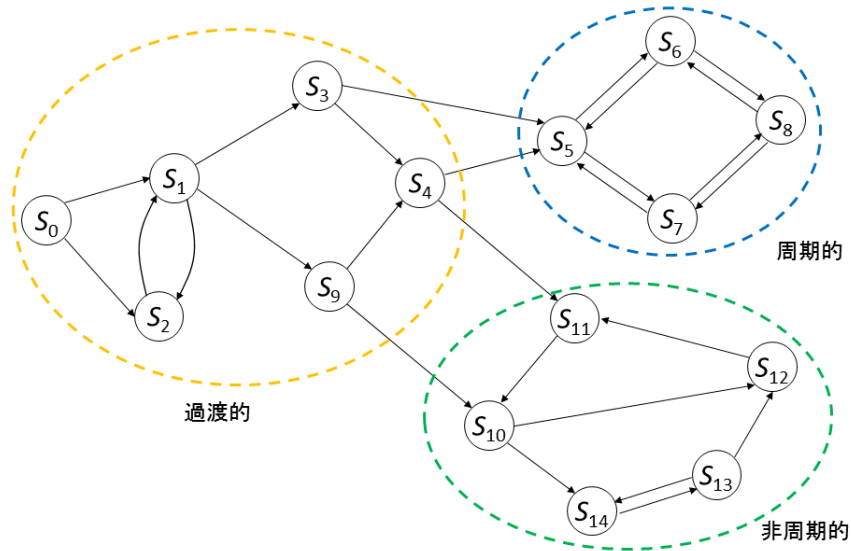


図 5. マルコフ情報源の例

状態は次のように分類される．

- 過渡状態
- 閉じた状態：いったんこの状態に入ると抜け出すことはない。
 - 非周期的状態集合
 - 周期的状態集合

一つの閉じた状態集合だけからなるマルコフ情報源を規約マルコフ情報源という．非周期的な規約マルコフ情報源（「正規マルコフ情報源」と呼ばれる）は，エルゴード的である．図 6 に例を示す．周期的な規約マルコフ情報源はある条件の下でしかエルゴード的にならない（図 7）．

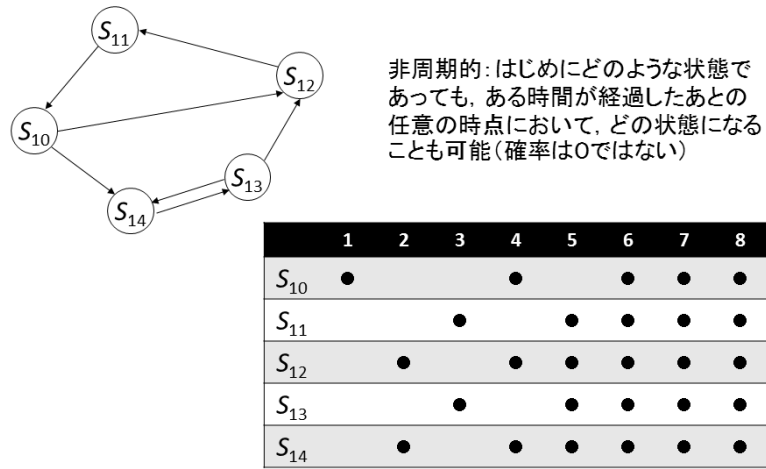


図 6. 規約マルコフ情報源の例

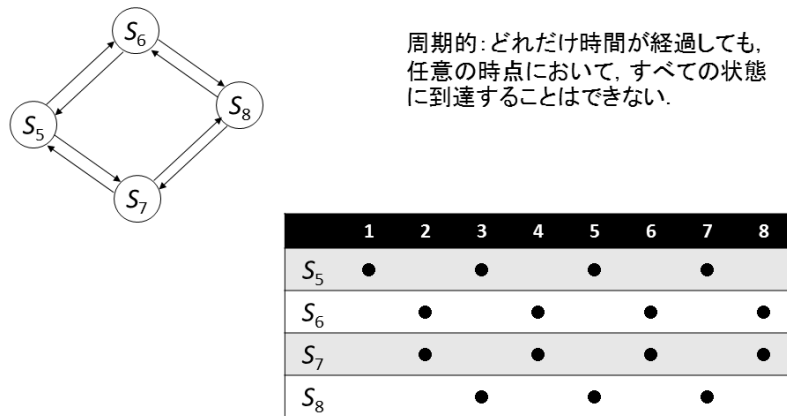


図 7. 非規約マルコフ情報源の例

6.5 マルコフ情報源の遷移確率行列表現

状態 s_i の次に状態 s_j に移る確率 $p_{ij} = P(s_j|s_i)$ を遷移確率と呼ぶ。 N 個の状態 $s_0 \sim s_{N-1}$ に対して遷移確率を定義し、 $N \times N$ の正方行列として表したもの：

$$\Pi = \begin{bmatrix} p_{0,0} & \cdots & p_{0,N-1} \\ \vdots & \ddots & \vdots \\ p_{N-1,0} & \cdots & p_{N-1,N-1} \end{bmatrix}$$

を遷移確率行列と呼ぶ。

【例】 投球の2重マルコフモデルから構成された一般化マルコフ情報源モデル（図8）：
 については、 遷移確率行列は次のようになる。

$$\Pi = \begin{bmatrix} p_{直直,直直} & p_{直変,直直} & p_{変直,直直} & p_{変変,直直} \\ p_{直直,直変} & p_{直変,直変} & p_{変直,直変} & p_{変変,直変} \\ p_{直直,変直} & p_{直変,変直} & p_{変直,変直} & p_{変変,変直} \\ p_{直直,変変} & p_{直変,変変} & p_{変直,変変} & p_{変変,変変} \end{bmatrix} = \begin{bmatrix} 0.6 & 0.4 & 0 & 0 \\ 0 & 0 & 0.2 & 0.8 \\ 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.6 & 0.4 \end{bmatrix}$$

(2) $p_{ij}^{(t)}$ ：ちょうど t 回目の遷移で s_i から s_j に至る確率とすれば、

$$\begin{cases} p_{ij}^{(1)} = p_{ij} \\ p_{ij}^{(t)} = \sum_{k=0}^{N-1} p_{ik}^{(t-1)} p_{kj} \end{cases}$$

$p_{ij}^{(t)}$ で構成される行列 $\Pi^{(t)}$ の i, j 要素は、 遷移確率行列を t 乗して得られる行列の i, j 要素に等しい。つまり、

$$\Pi^{(t)} = \begin{bmatrix} p_{0,0}^{(t)} & \cdots & p_{0,N-1}^{(t)} \\ \vdots & \ddots & \vdots \\ p_{N-1,0}^{(t)} & \cdots & p_{N-1,N-1}^{(t)} \end{bmatrix} = \Pi^t$$

となる。

例えば、図8の場合、ちょうど3回の遷移で状態 $S_{変直}$ から状態 $S_{直変}$ に遷移するのは、 $S_{変直} \rightarrow S_{直直} \rightarrow S_{直直} \rightarrow S_{直変}$ という遷移と $S_{変直} \rightarrow S_{直変} \rightarrow S_{変直} \rightarrow S_{直変}$ という遷移だけである。これに対応して、

$$\begin{aligned} p_{変直,直変}^{(3)} &= p_{変直,直直}^{(2)} \cdot p_{直直,直変} + p_{変直,変直}^{(2)} \cdot p_{変直,直変} \\ &= p_{変直,直直}^{(1)} \cdot p_{直直,直直} \cdot p_{直直,直変} + p_{変直,直変}^{(1)} \cdot p_{直変,変直} \cdot p_{変直,直変} \\ &= 0.5 \times 0.6 \times 0.4 + 0.5 \times 0.2 \times 0.5 \\ &= 0.17 \end{aligned}$$

一方、行列演算では、

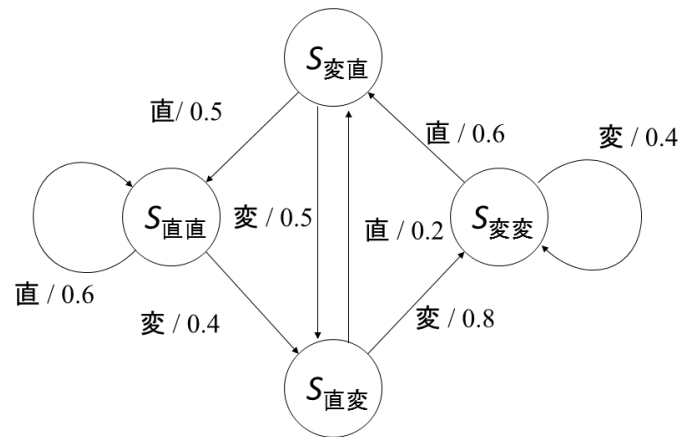


図 8. 投球の2重マルコフモデルから構成された一般化マルコフ情報源モデル

$$\Pi^3 = \begin{bmatrix} 0.6 & 0.4 & 0 & 0 \\ 0 & 0 & 0.2 & 0.8 \\ 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.6 & 0.4 \end{bmatrix}^3 = \begin{bmatrix} 0.256 & 0.184 & 0.24 & 0.32 \\ 0.3 & 0.28 & 0.212 & 0.208 \\ 0.23 & 0.17 & 0.28 & 0.32 \\ 0.3 & 0.24 & 0.156 & 0.304 \end{bmatrix}$$

となる. この行列の3行2列の要素が $p_{\text{変直,直変}}^{(3)}$ を表している.

6.6 正規マルコフ情報源の漸近的な性質

正規マルコフ情報源は十分な時間が経過した後は, どの状態からどの状態へも遷移し得る. 換言すれば, 正規マルコフ情報源に対する遷移確率行列を

$$\Pi = \begin{bmatrix} p_{0,0} & \cdots & p_{0,N-1} \\ \vdots & \ddots & \vdots \\ p_{N-1,0} & \cdots & p_{N-1,N-1} \end{bmatrix}$$

とすれば, ある t_0 が存在し, 任意の $t \geq t_0, i, j > 0$ に対して, $p_{ij}^{(t_0)} > 0$ となる.

【例】図 8 の一般化マルコフ情報源モデルについては, 2回以上の遷移でどの状態からどの状態にも到達し得る. 実際,

$$\Pi^2 = \begin{bmatrix} 0.6 & 0.4 & 0 & 0 \\ 0 & 0 & 0.2 & 0.8 \\ 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.6 & 0.4 \end{bmatrix}^2 = \begin{bmatrix} 0.36 & 0.24 & 0.08 & 0.32 \\ 0.1 & 0.1 & 0.48 & 0.32 \\ 0.3 & 0.2 & 0.1 & 0.4 \\ 0.3 & 0.3 & 0.24 & 0.16 \end{bmatrix}$$

となっている.

正規マルコフ情報源では, $t \rightarrow \infty$ のとき, $p_{ij}^{(t)}$ は i に無関係な値に収束する. すなわち,

$$\lim_{t \rightarrow \infty} p_{ij}^{(t)} = u_j$$

つまり,

$$\lim_{t \rightarrow \infty} \Pi^t = U = \begin{bmatrix} u_0 & \cdots & u_{N-1} \\ \vdots & \ddots & \vdots \\ u_0 & \cdots & u_{N-1} \end{bmatrix}$$

となる.

【例】これまでの例題で取り上げてきた

$$\Pi = \begin{bmatrix} 0.6 & 0.4 & 0 & 0 \\ 0 & 0 & 0.2 & 0.8 \\ 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.6 & 0.4 \end{bmatrix}$$

については, Π^n は次のようになる.

```
In[8]:= MatrixForm[p]
Out[8]/MatrixForm=

$$\begin{pmatrix} 0.6 & 0.4 & 0 & 0 \\ 0 & 0 & 0.2 & 0.8 \\ 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.6 & 0.4 \end{pmatrix}$$

In[13]:= MatrixForm[p.p.p.p.p]
Out[13]/MatrixForm=

$$\begin{pmatrix} 0.27856 & 0.22384 & 0.2096 & 0.288 \\ 0.262 & 0.2048 & 0.22952 & 0.30368 \\ 0.2798 & 0.2242 & 0.2048 & 0.2912 \\ 0.27 & 0.2184 & 0.22776 & 0.28384 \end{pmatrix}$$

In[14]:= MatrixForm[p.p.p.p.p.p]
Out[14]/MatrixForm=

$$\begin{pmatrix} 0.271936 & 0.216224 & 0.217568 & 0.294272 \\ 0.27196 & 0.21956 & 0.223168 & 0.285312 \\ 0.27028 & 0.21432 & 0.21956 & 0.29584 \\ 0.27588 & 0.22188 & 0.213984 & 0.288256 \end{pmatrix}$$

In[16]:= MatrixForm[p.p.p.p.p.p.p.p.p.p.p.p.p.p.p.p]
Out[16]/MatrixForm=

$$\begin{pmatrix} 0.272728 & 0.218183 & 0.218181 & 0.290908 \\ 0.272726 & 0.218179 & 0.218182 & 0.290913 \\ 0.272729 & 0.218184 & 0.218179 & 0.290908 \\ 0.272726 & 0.218181 & 0.218185 & 0.290909 \end{pmatrix}$$

```

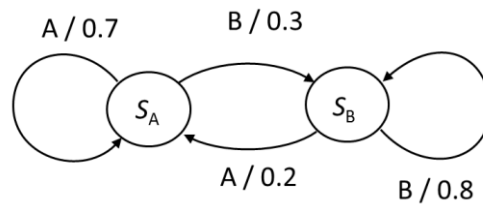


図 9. 正規マルコフ情報源モデルの例

図 9 の正規マルコフ情報源の場合は、次のようになる。

$$\begin{aligned} \Pi &= \begin{bmatrix} 0.7 & 0.3 \\ 0.2 & 0.8 \end{bmatrix} \\ \Pi^2 &= \begin{bmatrix} 0.55 & 0.45 \\ 0.3 & 0.7 \end{bmatrix} \\ \Pi^3 &= \begin{bmatrix} 0.475 & 0.525 \\ 0.35 & 0.65 \end{bmatrix} \\ \Pi^4 &= \begin{bmatrix} 0.4375 & 0.5625 \\ 0.375 & 0.625 \end{bmatrix} \\ \Pi^5 &= \begin{bmatrix} 0.41875 & 0.58125 \\ 0.3875 & 0.6125 \end{bmatrix} \\ \Pi^6 &= \begin{bmatrix} 0.409375 & 0.590625 \\ 0.39375 & 0.60625 \end{bmatrix} \\ \Pi^7 &= \begin{bmatrix} 0.404688 & 0.595313 \\ 0.396875 & 0.603125 \end{bmatrix} \\ \Pi^8 &= \begin{bmatrix} 0.402344 & 0.597656 \\ 0.38438 & 0.601563 \end{bmatrix} \\ \Pi^9 &= \begin{bmatrix} 0.401172 & 0.598828 \\ 0.399219 & 0.600781 \end{bmatrix} \\ \Pi^{10} &= \begin{bmatrix} 0.400586 & 0.599414 \\ 0.399609 & 0.600391 \end{bmatrix} \\ &\dots \end{aligned}$$

なぜこのような性質があるのか？ $N = 2$ の場合について説明する。

$$\Pi = \begin{bmatrix} p & 1-p \\ q & 1-q \end{bmatrix} = \begin{bmatrix} p & \bar{p} \\ q & \bar{q} \end{bmatrix}$$

とする。

$$\Pi^n = \begin{bmatrix} X^{(n)} & \bar{X}^{(n)} \\ Y^{(n)} & \bar{Y}^{(n)} \end{bmatrix} \quad (X^{(n)} + \bar{X}^{(n)} = 1, Y^{(n)} + \bar{Y}^{(n)} = 1)$$

と表記しよう。すると、 $\Pi^{n+1} = \Pi \cdot \Pi^n$ であるので、

$$\begin{aligned} \Pi^{n+1} &= \begin{bmatrix} X^{(n+1)} & \bar{X}^{(n+1)} \\ Y^{(n+1)} & \bar{Y}^{(n+1)} \end{bmatrix} \\ &= \begin{bmatrix} p & \bar{p} \\ q & \bar{q} \end{bmatrix} \cdot \begin{bmatrix} X^{(n)} & \bar{X}^{(n)} \\ Y^{(n)} & \bar{Y}^{(n)} \end{bmatrix} \\ &= \begin{bmatrix} pX^{(n)} + \bar{p}Y^{(n)} & p\bar{X}^{(n)} + \bar{p}\bar{Y}^{(n)} \\ qX^{(n)} + \bar{q}Y^{(n)} & q\bar{X}^{(n)} + \bar{q}\bar{Y}^{(n)} \end{bmatrix} \end{aligned}$$

となる. 図 10 のように, $X^{(n+1)}$ と $Y^{(n+1)}$ は, $X^{(n)} \sim Y^{(n)}$ をそれぞれ $\bar{p}:p, \bar{q}:q$ に内分した値になっている. 従って, $n \rightarrow \infty$ に従って, 相互に近づいていく. $\bar{X}^{(n+1)}$ と $\bar{Y}^{(n+1)}$ についても同様. また, $X^{(n+1)} + \bar{X}^{(n+1)} = 1, Y^{(n+1)} + \bar{Y}^{(n+1)} = 1$ であることも確認できる. ■

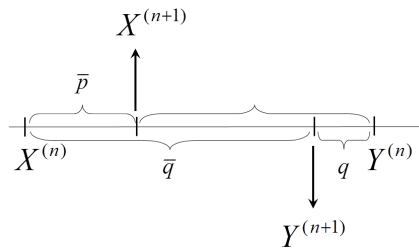


図 10. $X^{(n+1)}$ と $Y^{(n+1)}$ は, $X^{(n)} \sim Y^{(n)}$ をそれぞれ $\bar{p}:p, \bar{q}:q$ に内分した値になっている

6.7 正規マルコフ情報源の定常分布の計算法

定常分布を $w = (w_0, w_1, \dots, w_{N-1})$ とする. ある時点での状態分布が定常的で w であれば, 「定常分布」の定義により, 次の状態分布も w であるはずだ. この性質を利用して, (確かにそれが存在するならば) 定常分布がどうなるか, 次の連立方程式を解くことによって計算できる.

$$\begin{cases} w_0 + w_1 + \dots + w_{N-1} = 1 \\ w\Pi = w \end{cases}$$

正規マルコフ情報源にはこの式を満たす定常分布が存在し, 極限分布と一致する.

【例】これまで取り上げてきた図 8 の例については,

$$\begin{cases} w_{直直} + w_{直変} + w_{変直} + w_{変変} = 1 \\ 0.6w_{直直} + 0.5w_{変直} = w_{直直} \\ 0.4w_{直直} + 0.5w_{変直} = w_{直変} \\ 0.2w_{直変} + 0.6w_{変変} = w_{変直} \\ 0.8w_{直変} + 0.4w_{変変} = w_{変変} \end{cases}$$

ないしは,

$$\begin{cases} w_{直直} + w_{直変} + w_{変直} + w_{変変} \\ (w_{直直}, w_{直変}, w_{変直}, w_{変変}) \begin{bmatrix} 0.6 & 0.4 & 0 & 0 \\ 0 & 0 & 0.2 & 0.8 \\ 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.6 & 0.4 \end{bmatrix} \end{cases} = (w_{直直}, w_{直変}, w_{変直}, w_{変変})$$

を解いて、 $w_{直直} \approx 0.273, w_{直変} \approx 0.218, w_{変直} \approx 0.218, w_{変変} \approx 0.291$ を得る（上で、1個の等式は冗長）。

6.8 一般マルコフ情報源のエントロピー計算法

- 情報源アルファベット $\{a_1, \dots, a_M\}$
- 状態 $\{S_0, \dots, S_{N-1}\}$
- その定常確率分布 $\{w_0, \dots, w_{N-1}\}$
- 状態 S_i にあるときの情報源記号 a_i を発生する確率 $P(a_i|S_i)$

とすれば、マルコフ情報源 S のエントロピーは、次のようになる。

$$H(S) = \sum_{i=0}^{N-1} w_i \left(- \sum_{j=1}^M P(a_j|S_i) \log_2 P(a_j|S_i) \right)$$

計算の手順は、次のようになる。

- ・ 定常分布を求める。
- ・ 各状態におけるエントロピーを求める。
- ・ それをブレンドする。

【例】 図 11 のマルコフ情報源は次の遷移確率

$$\begin{cases} P_{X_i|X_{i-1}}(A|A) = 0.9 \\ P_{X_i|X_{i-1}}(B|A) = 0.1 \\ P_{X_i|X_{i-1}}(A|B) = 0.4 \\ P_{X_i|X_{i-1}}(B|B) = 0.6 \end{cases}$$

で規定される。定常的な状態分布を (w_A, w_B) 、状態が S_A である確率を w_A 、 S_B である確率を w_B とすると

$$\begin{cases} w_A + w_B = 1 \\ (w_A, w_B) \begin{bmatrix} 0.9 & 0.1 \\ 0.4 & 0.6 \end{bmatrix} = (w_A, w_B) \end{cases}$$

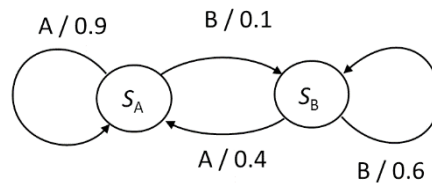


図 11. マルコフ情報源の例

を解いて、 $w_A = 0.8, w_B = 0.2$ を得る.

状態が S_A であるときのエントロピーは、 $H_{S_A}(S) = \mathcal{H}(0.1) \approx 0.469$ である.

状態が S_B であるときのエントロピーは、 $H_{S_B}(S) = \mathcal{H}(0.4) \approx 0.971$ である.

それぞれの状態の確率に応じてブレンドすると、 $H(S) = 0.8H_{S_A}(S) + 0.2H_{S_B}(S) \approx 0.5694$ が得られる.

6.9 マルコフ情報源の n 次エントロピー

直観を得るために、図 11 のマルコフ情報源 S の n 次エントロピーを求めてみよう.

(1) S の1次エントロピー $H_1(S)$ は次のように求められる.

$$\begin{aligned}
 P(A) &= P(A, S_A) + P(A, S_B) \\
 &= P(S_A)P(A|S_A) + P(S_B)P(A|S_B) \\
 &= 0.8 \times 0.9 + 0.2 \times 0.4 \\
 &= 0.8 \\
 P(B) &= P(B, S_A) + P(B, S_B) \\
 &= P(S_A)P(B|S_A) + P(S_B)P(B|S_B) \\
 &= 0.8 \times 0.1 + 0.2 \times 0.6 \\
 &= 0.2
 \end{aligned}$$

よって、 S の1次エントロピー $H_1(S)$ は、 $H_1(S) = \mathcal{H}(0.8) \approx 0.722$ である.

(2) S の(1情報源記号あたりの)2次エントロピー $H_2(S) = \frac{H_1(S^2)}{2}$ は次のように求められる.

$$\begin{aligned}
P(AA) &= P(AA, S_A) + P(AA, S_B) \\
&= P(S_A)P(AA|S_A) + P(S_B)P(AA|S_B) \\
&= 0.8 \times 0.9 \times 0.9 + 0.2 \times 0.4 \times 0.9 \\
&= 0.72 \\
P(AB) &= P(AB, S_A) + P(AB, S_B) \\
&= P(S_A)P(AB|S_A) + P(S_B)P(AB|S_B) \\
&= 0.8 \times 0.9 \times 0.1 + 0.2 \times 0.4 \times 0.1 \\
&= 0.08 \\
P(BA) &= P(BA, S_A) + P(BA, S_B) \\
&= P(S_A)P(BA|S_A) + P(S_B)P(BA|S_B) \\
&= 0.8 \times 0.1 \times 0.4 + 0.2 \times 0.6 \times 0.4 \\
&= 0.08 \\
P(BB) &= P(BB, S_A) + P(BB, S_B) \\
&= P(S_A)P(BB|S_A) + P(S_B)P(BB|S_B) \\
&= 0.8 \times 0.1 \times 0.6 + 0.2 \times 0.6 \times 0.6 \\
&= 0.12
\end{aligned}$$

よって、 S の2次拡大情報源 S^2 の1次エントロピー $H_1(S^2)$ は、

$$\begin{aligned}
H_1(S^2) &= -(0.72 \times \log_2 0.72 + 0.08 \times \log_2 0.08 + 0.08 \times \log_2 0.08 + 0.12 \times \log_2 0.12) \\
&\approx 1.29
\end{aligned}$$

よって、 S の(1情報源記号あたりの)2次エントロピーは、

$$H_2(S) = \frac{H_1(S^2)}{2} \approx 0.646$$

となる.

(3) S の(1情報源記号あたりの)3次エントロピー $H_3(S) = \frac{H_1(S^3)}{3}$ は次のように求められる.

$$\begin{aligned}
P(AAA) &= P(AAA, S_A) + P(AAA, S_B) \\
&= 0.8 \times 0.9 \times 0.9 \times 0.9 + 0.2 \times 0.4 \times 0.9 \times 0.9 \\
&= 0.648 \\
P(AAB) &= P(AAB, S_A) + P(AAB, S_B) \\
&= 0.8 \times 0.9 \times 0.9 \times 0.1 + 0.2 \times 0.4 \times 0.9 \times 0.1 \\
&= 0.072 \\
P(ABA) &= P(ABA, S_A) + P(ABA, S_B) \\
&= 0.8 \times 0.9 \times 0.1 \times 0.4 + 0.2 \times 0.4 \times 0.1 \times 0.4 \\
&= 0.032 \\
P(ABB) &= P(ABB, S_A) + P(ABB, S_B) \\
&= 0.8 \times 0.9 \times 0.1 \times 0.6 + 0.2 \times 0.4 \times 0.1 \times 0.6 \\
&= 0.048 \\
P(BAA) &= P(BAA, S_A) + P(BAA, S_B) \\
&= 0.8 \times 0.1 \times 0.4 \times 0.9 + 0.2 \times 0.6 \times 0.4 \times 0.9 \\
&= 0.072 \\
P(BAB) &= P(BAB, S_A) + P(BAB, S_B) \\
&= 0.8 \times 0.1 \times 0.4 \times 0.1 + 0.2 \times 0.6 \times 0.4 \times 0.1 \\
&= 0.008 \\
P(BBA) &= P(BBA, S_A) + P(BBA, S_B) \\
&= 0.8 \times 0.1 \times 0.6 \times 0.4 + 0.2 \times 0.6 \times 0.6 \times 0.4 \\
&= 0.048
\end{aligned}$$

$$\begin{aligned}
 P(\text{BBB}) &= P(\text{BBB}, S_A) + P(\text{BBB}, S_B) \\
 &= 0.8 \times 0.1 \times 0.6 \times 0.6 + 0.2 \times 0.6 \times 0.6 \times 0.4 \\
 &= 0.072
 \end{aligned}$$

よって、 S の3次拡大情報源の1次エントロピー $H_1(S^3)$ は、

$$\begin{aligned}
 H_1(S^3) &= -(0.648 \times \log_2 0.648 + 0.072 \times \log_2 0.072 \\
 &\quad + 0.032 \times \log_2 0.032 + 0.048 \times \log_2 0.048 \\
 &\quad + 0.072 \times \log_2 0.072 + 0.008 \times \log_2 0.008 \\
 &\quad + 0.048 \times \log_2 0.048 + 0.072 \times \log_2 0.072) \\
 &\approx 1.8607
 \end{aligned}$$

S の(1情報源記号あたりの)3次エントロピーは、

$$H_3(S) = \frac{H_1(S^3)}{3} \approx 0.620$$

このように、 $H_n(S)$ の1次エントロピーは $H_1(S) = \mathcal{H}(0.8) \approx 0.722$ であるが、 n の増加とともに $H_n(S)$ の値は少しずつ減少し、 $H_\infty(S) = 0.8\mathcal{H}(0.9) + 0.2\mathcal{H}(0.4) \approx 0.569$ に漸近していく。この様子を図12に示す。

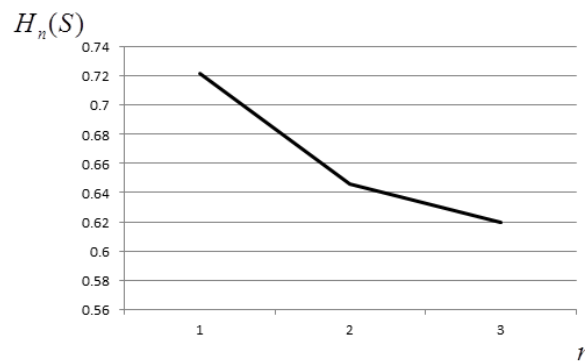


図12. n の増加とともに $H_n(S)$ の値は少しずつ減少し、 $H_\infty(S) \approx 0.569$ に漸近していく

n が増加するときの n 次エントロピーの値の変化の背後にある構造は図13のようにとらえることができる。

$H_1(S)$ は、記憶のあるマルコフ情報源を状況に関わりなく0.8の確率で0を発生する無記憶情報源として近似するものであり、記憶を無視する分だけエントロピーは大きい(次に発生する情報源記号が予測しがたい)。しかし、 n 個の連続した情報源記号を考慮に入れて予測し難さを評価する n 次エントロピーでは、先行する情報源記号の系列を考慮に入れられるので、予測しやすさが少しずつ増していき、それに伴ってエントロピー値が減少する。

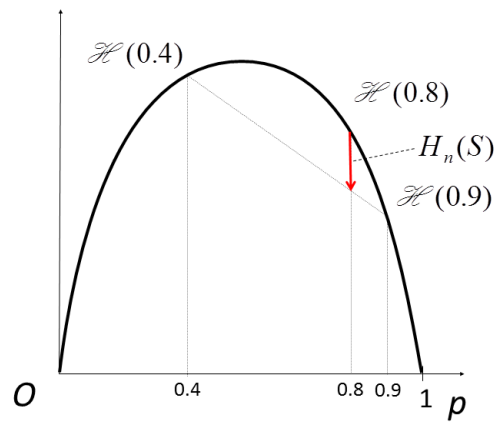


図 13. n 次エントロピーの値の変化の背後にある構造