



情報エレクトロニクス学科共通科目・2年次・夏ターム〔必修科目〕

講義「情報理論」第4回

第3章 情報源のモデル[前半]

- 3.1 情報源の統計的表現
- 3.2 情報源の基本的なモデル
- 3.3 マルコフ情報源
- 3.4 マルコフ情報源の確率分布



これから学ぶこと

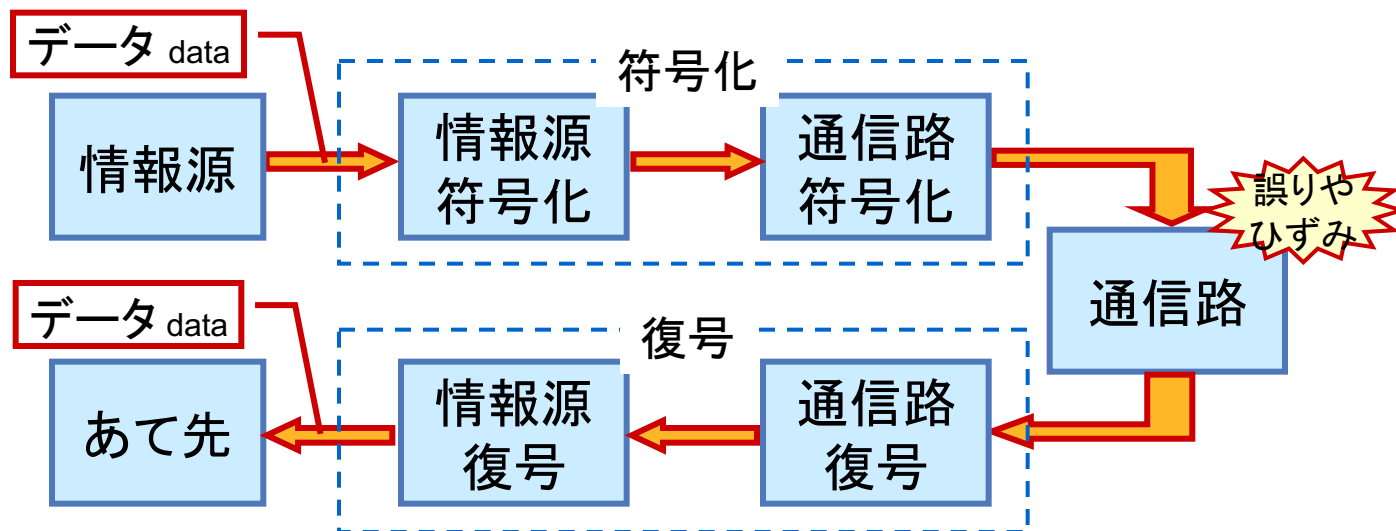
情報理論 = 情報の伝達をいかに効率よく、そして信頼性高く
行うかに関する理論

受け手の知識の変化

= 何らかの統計的知識に基づいて
受け手が与えている確率分布の変化

情報源
符号化

通信路
符号化



良い符号化法・復号法とは？ 効率性・信頼性の限界は？

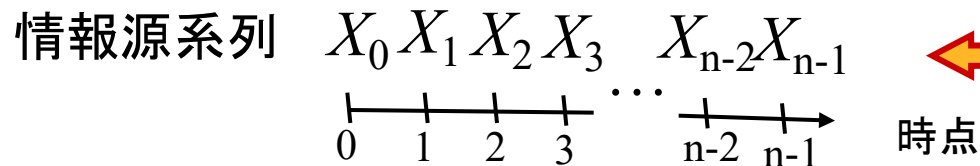


情報源の統計的表現

■ 離散的 M 元情報源

- M 個の元を持つ情報源アルファベット $A = \{a_1, a_2, \dots, a_M\}$ から
- 時点0 より情報源記号を発生する(時点は整数値)
- 時点 i の情報源の出力を X_i ($i = 0, 1, 2, \dots$) で表す
- 時点 $n-1$ まで(長さ n)の情報源系列 $X_0 X_1 X_2 \dots X_{n-1}$ を考える

確率変数



離散的 M 元情報源
 $A = \{a_1, a_2, \dots, a_M\}$

この情報系列の統計的性質は
 $X_0 X_1 X_2 \dots X_{n-1}$ の結合確率分布で完全に定まる。

X_0, X_1, \dots, X_{n-1} の結合確率分布

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



結合確率分布により統計的性質が完全に定まるとは？

$P_{X_0, X_1 \dots X_{n-1}}(x_0, x_1, \dots, x_{n-1})$ から $X_0, X_1 \dots X_{n-1}$ に関する
どんな確率分布も計算できる。

↑例えば

- $X_0, X_1 \dots X_{n-1}$ のうちの $m(<n)$ 個の結合確率分布
- $X_0, X_1 \dots X_{n-1}$ のうちの k 個の変数の値に条件を付けた m 個 ($k+m \leq n$) の条件付結合確率分布

添え字を省略している
 $P_{X_0, X_1, X_2}(x_0, x_1, x_2)$ の意

(例) 情報源アルファベット $A = \{0, 1\}$

X_0, X_1, X_2 の結合確率分布は右図

$X_0=0$ となる確率 $P_{X_0}(0)$ は？

$$\begin{aligned}
 P_{X_0}(0) &= \sum_{x_1=0}^1 \sum_{x_2=0}^1 P(0, x_1, x_2) \\
 &= 0.648 + 0.072 + 0.032 + 0.048 \\
 &= 0.8
 \end{aligned}$$

x_0	x_1	x_2	$P(x_0, x_1, x_2)$
0	0	0	0.648
0	0	1	0.072
0	1	0	0.032
0	1	1	0.048
1	0	0	0.072
1	0	1	0.008
1	1	0	0.048
1	1	1	0.072



例の続き: 全体の結合確率分布からの条件付確率分布の求め方

- **条件付確率分布** $P_{X_1|X_0}(x_1 | x_0) = P_{X_0X_1}(x_0, x_1) / P_{X_0}(x_0)$

$X_0 = x_0$ という条件の下での $X_1 = x_1$ となる確率

例) $X_0 = 0$ という条件の下での $X_1 = 0$ となる条件付確率 $P_{X_1|X_0}(0|0)$

$$\begin{aligned} P_{X_1|X_0}(0|0) &= \frac{P_{X_0X_1}(0, 0)}{P_{X_0}(0)} \\ &= \frac{0.72}{0.8} \\ &= 0.9 \end{aligned}$$

x_0	x_1	x_2	$P(x_0, x_1, x_2)$
0	0	0	0.648
0	0	1	0.072
0	1	0	0.032
0	1	1	0.048
1	0	0	0.072
1	0	1	0.008
1	1	0	0.048
1	1	1	0.072

さらに、 X_0 と X_1 で条件を付けた X_2 の条件付確率分布

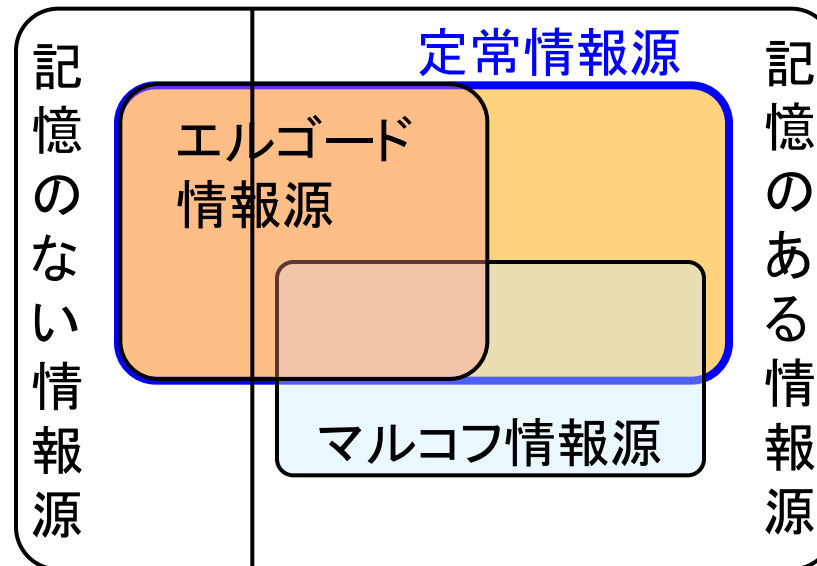
$$P(x_2 | x_0, x_1) = \frac{P(x_0, x_1, x_2)}{P(x_0, x_1)}$$

演習1



情報源の各種モデル

- どんな大きい n についても、 X_0, X_1, \dots, X_{n-1} の結合確率分布を与える(M^n 個の確率を与える)ことができれば、この情報源の統計的性質は完全に記述できる。しかし、一般には困難。
- 議論を進めるためには、情報源に何らかの扱いやすい性質を仮定する必要がある。



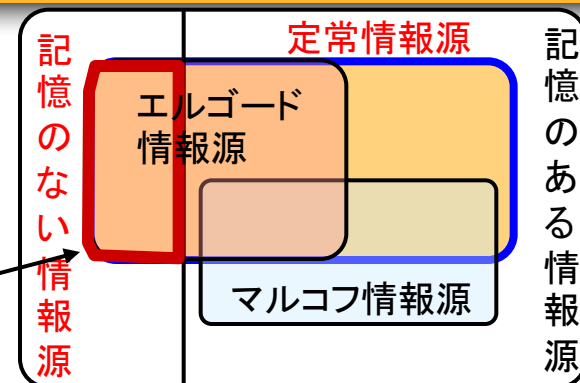


記憶のない定常情報源(最も簡単な性質をもつ情報源)

■ 記憶のない情報源 (memoryless information source)

- 各時点 i における情報源記号 X_i が、他の時点 j の情報源記号 X_j と独立である情報源

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



■ 記憶のない定常情報源

- 記憶のない情報源において、各時点 i における情報源記号 X_i が同一の確率分布 $P_X(x)$ に従う情報源

記憶のない定常情報源における長さ n の系列の結合確率分布

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

例) さいころの1の目の出方を調べる。目が $1 \rightarrow 1$ 、それ以外0。

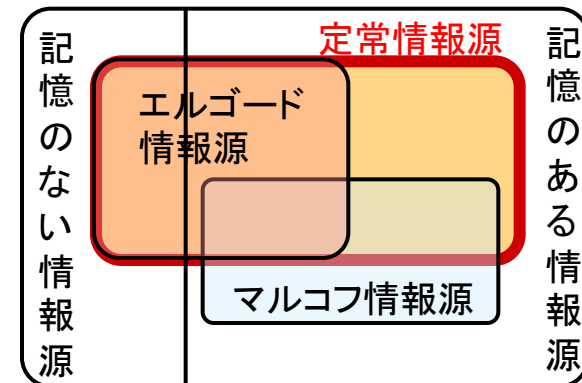
$\Rightarrow \{0, 1\}$ の2元情報源で、かつ1、0の発生確率が $1/6$ 、 $5/6$ となる記憶のない定常情報源になる。



(一般の)定常情報源の定義

■ 定常情報源 (stationary information source)

- 時間をずらしても、統計的性質の変わらない情報源
- すなわち、任意の正整数 n と i および情報源アルファベットの任意の元 x_0, x_1, \dots, x_{n-1} に対し、



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

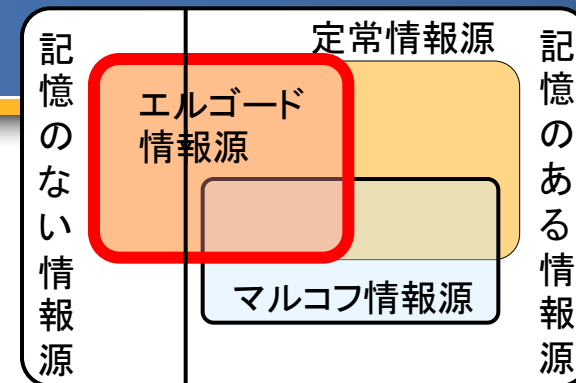
が成立するとき、この情報源を**定常**という。

- 定常情報源の出力は、各時点において同一の確率分布に従う。
この確率分布を**定常分布**と呼ぶ。

式①において、 $n=1$ とおくと、 $P_{X_0}(x_0) = P_{X_i}(x_0)$ となる

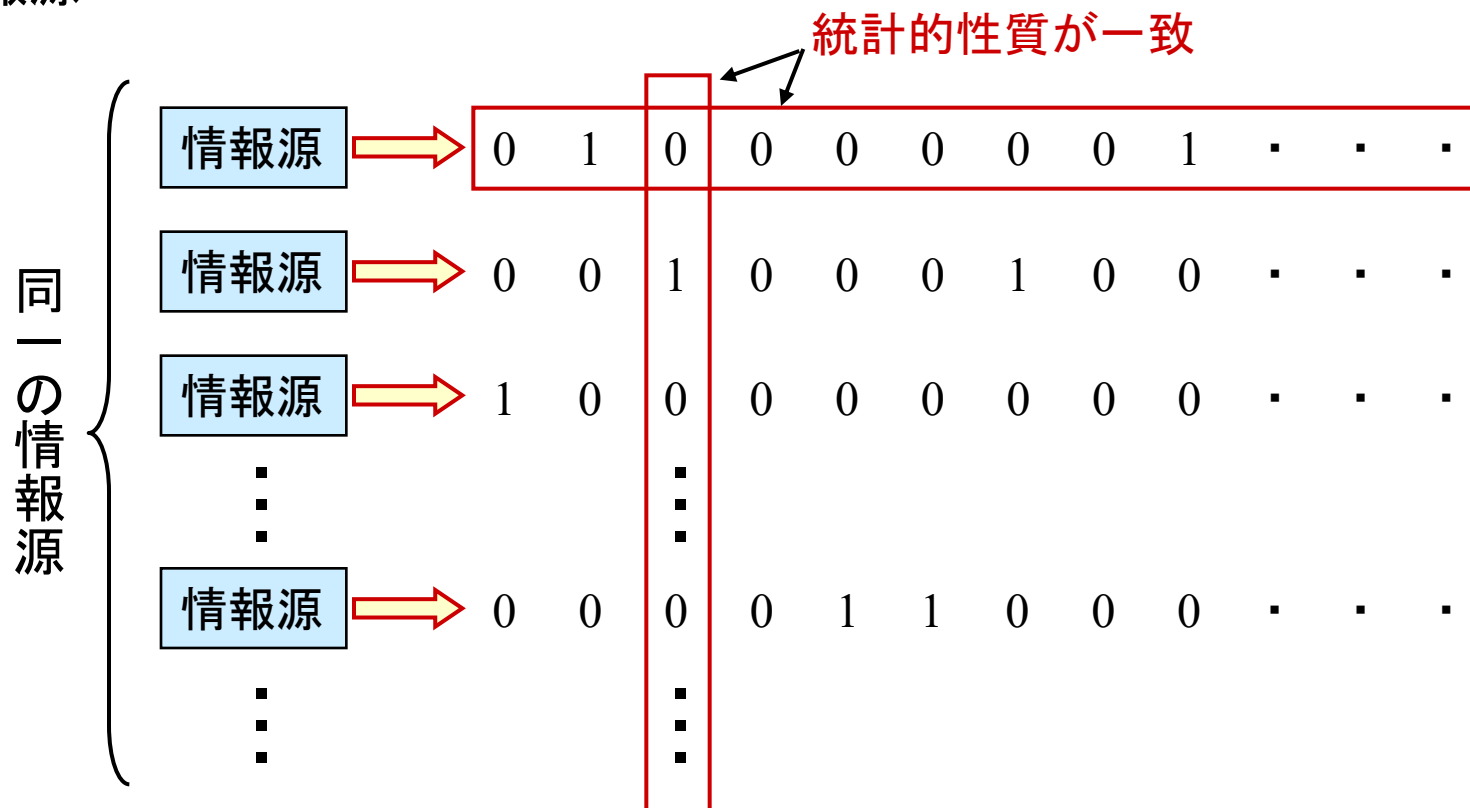


エルゴード情報源



エルゴード情報源 (ergodic information source)

- それが発生する十分長い任意の系列に、その情報源の統計的性質が完全に現れている定常情報源





エルゴード情報源の性質

- エルゴード情報源では、**集合平均と時間平均が一致する**
 - 集合平均を求めるのは難しい。時間平均を観測するのは容易。

定常分布が $P_X(x)$ である定常情報源の出力 X を変数とする任意の関数 $f(X)$ を考える。ただし $f(X)$ は実数値をとるものとする。

$f(X)$ の**集合平均** (ensemble average)

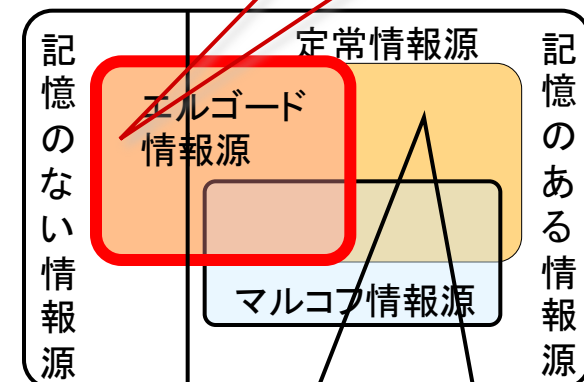
$$\overline{f(X)} = \sum_{x \in A} f(x) P_X(x)$$

$f(X)$ の**時間平均** (time average)

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

エルゴード情報源においては

$$\overline{f(X)} = \langle f(X) \rangle$$



定常情報源であってもエルゴード的でないものはいくらでもある。

例) 確率1/2で0だけからなる系列か1だけからなる系列を発生する情報源



マルコフ情報源

■ マルコフ情報源

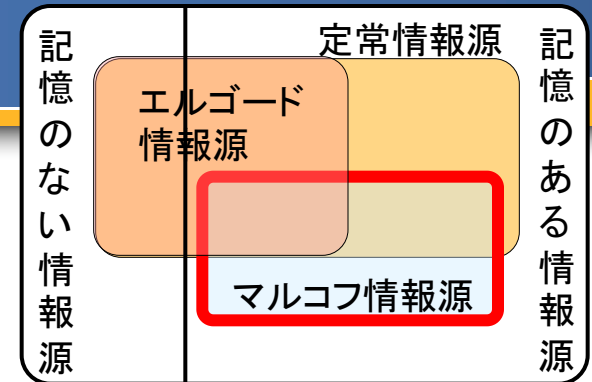
– 記憶のある情報源の基本的なモデル

■ m 重マルコフ情報源 (m -th order Markov information source)

– n を m 以上の任意の整数とすると、任意の時点 i について、その直前の n 個の出力 $X_{i-1}, X_{i-2}, \dots, X_{i-n}$ で条件を付けた X_i の条件付確率分布が、直前の m 個の出力 $X_{i-1}, X_{i-2}, \dots, X_{i-m}$ だけで条件を付けた X_i の条件付確率と一致するとき、すなわち

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

となるとき、この情報源を m 重マルコフ情報源という。



系列中の過去 m 文字の並びにしたがって、次の文字の生起確率が決まるモデル



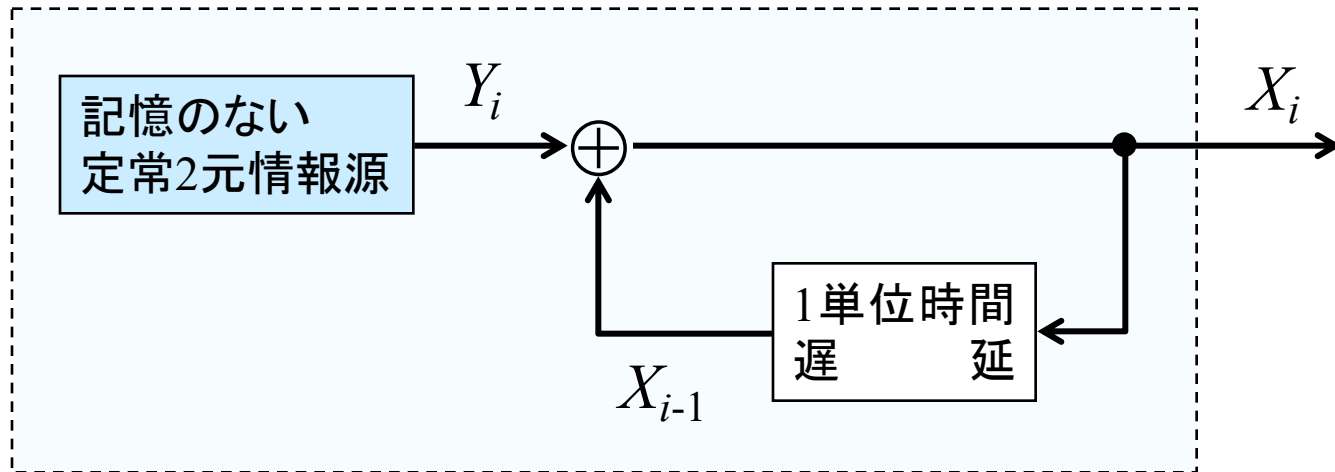
1重マルコフ情報源(単純マルコフ情報源)の例

[例] 1, 0をそれぞれ確率 p , $1-p$ で発生する記憶のない2元情報源の出力 Y_i と、1時点前のこの情報源の出力 X_{i-1} とから

$$X_i = X_{i-1} \oplus Y_i$$

により、現時点の出力 X_i が定まる2元情報源。

\oplus は、排他的論理和



$$\begin{aligned} P_{X_i | X_{i-1} X_{i-2}}(0 | 1\mathbf{0}) &= P_{X_i | X_{i-1} X_{i-2}}(0 | 1\mathbf{1}) \\ &= P_{X_i | X_{i-1}}(0 | 1) = p \end{aligned}$$

2つ前の出力には左右されない



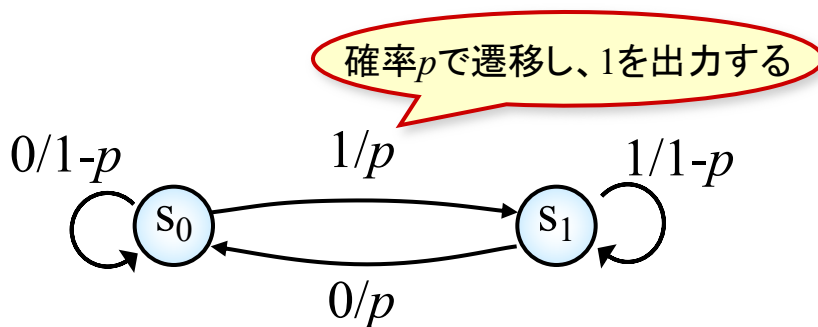
状態(遷移)図

■ q 元 m 重マルコフ情報源を考える

- 直前 m 個の出力によって、出力の確率分布が定まる。
- このことは情報源が、 m 個の出力のとり値に対応した q^m 個の状態を持ち、それぞれの状態において出力の確率分布が定まるとみなせる。
- 1つの記号を出力するたびに、情報源の状態は q^m 個の状態の中を遷移していく。

⇒状態間の遷移関係を表した図(状態遷移図)として表現できる！

[例]の単純マルコフ情報源の状態(遷移)図



$P_{X_i X_{i-1}}(x_i x_{i-1})$		x_i	
		0	1
x_{i-1}	0	$1-p$	p
	1	p	$1-p$

状態の集合= $\{S_0, S_1\}$

S_0 : 直前の出力が0, S_1 : 直前の出力が1

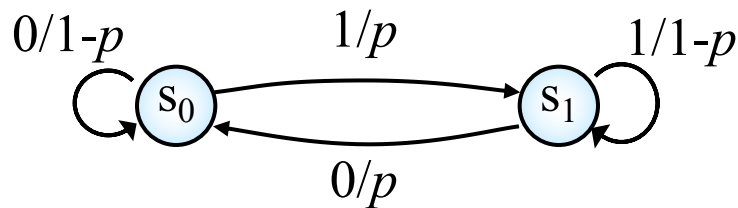


一般化されたマルコフ情報源

■ 一般化された q 元マルコフ情報源

- N 個の状態 S_0, S_1, \dots, S_N をもつ。
- それぞれの状態 S_i に依存して出力の確率分布が定まる。
- 1つの記号を出力するたびに、情報源の状態は N 個の状態の中を遷移していく。

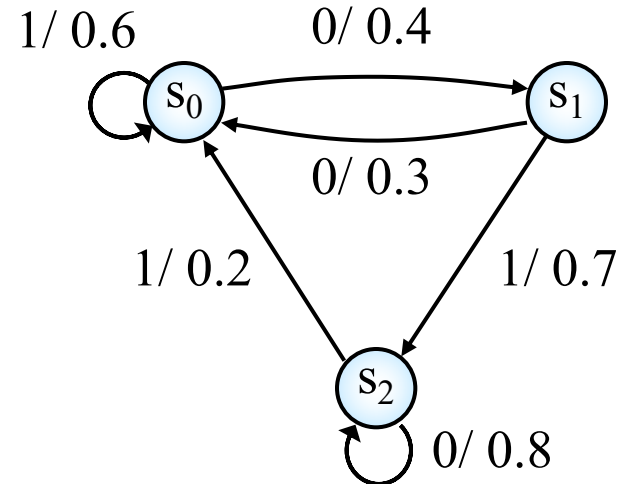
単純マルコフ情報源の状態図



現在の状態は直前の出力により決まる。

直前の出力が i なら現在の状態は S_i

一般化されたマルコフ情報源の状態図



現在の状態は直前の出力により決まらない。

直前の出力が1なら現在の状態は S_0 か S_2

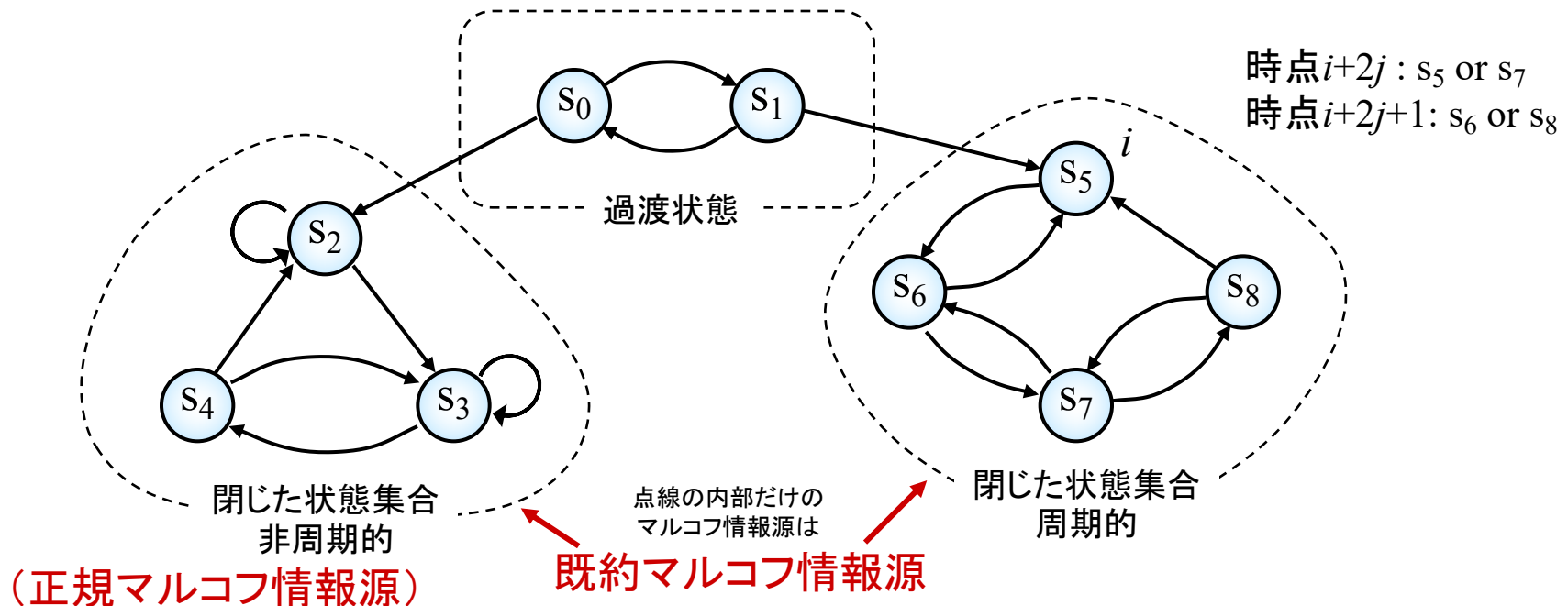


一般化



状態の分類

- **過渡状態**: 十分時間が経過すれば抜け出てしまい、戻ることのない状態
- **閉じた状態集合**: いったん入ると抜け出すことがない状態集合で、任意の状態から任意の状態へ矢印をたどっていくことのできる状態の集合
- **非周期的状態集合**: 閉じた状態集合で、ある時間が経過した後の任意の時点において、どの状態にある確率も0でないようなもの
- **周期的状態集合**: 閉じた状態集合がいくつかの部分集合に分割され、そのおのおのが周期的な時点においてのみ現れえるもの



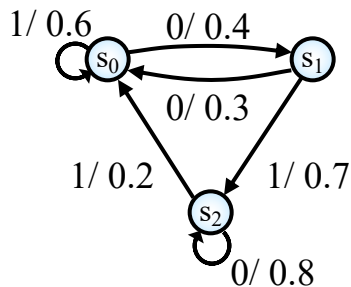


遷移確率行列

- N 個の状態 s_0, s_1, \dots, s_{N-1} を持つ、一般のマルコフ情報源を考える
 - 状態遷移の仕方は、状態 s_i にあるとき、次の時点で状態 s_j に遷移する確率 $p_{ij} = P(s_j | s_i)$ により決まる。これを遷移確率という。
 - 遷移確率 p_{ij} を (i, j) 要素とする $N \times N$ 行列

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

を遷移確率行列と呼ぶ。



例) 図のマルコフ情報源の遷移確率行列は次式のようにになる。

$$\Pi = \begin{pmatrix} 0.6 & 0.4 & 0 \\ 0.3 & 0 & 0.7 \\ 0.2 & 0 & 0.8 \end{pmatrix} \begin{matrix} \leftarrow s_0 \\ \leftarrow s_1 \\ \leftarrow s_2 \end{matrix} \left. \vphantom{\begin{pmatrix} 0.6 & 0.4 & 0 \\ 0.3 & 0 & 0.7 \\ 0.2 & 0 & 0.8 \end{pmatrix}} \right\} \text{から}$$

$s_0 \quad s_1 \quad s_2 \quad \wedge$



遷移確率行列による t 時点後の遷移確率

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

- ここで、状態 s_i から出発し、 t 時点後に s_j に到達する確率を $p_{ij}^{(t)}$ とする。明らかに

$$p_{ij}^{(1)} = p_{ij}$$

であり、 $p_{ij}^{(t)}$ は $p_{ij}^{(t-1)}$ から

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

Π^{t-1} と Π の積の (i,j) 成分の計算式

により計算できる。この式から

$$p_{ij}^{(t)} = \Pi^t \text{ の } (i,j) \text{ 要素}$$

行列の掛算

$$\underbrace{\Pi \cdot \Pi \cdots \Pi}_{t \text{ 回}}$$

t 回

となることが直ちに導ける。



正規マルコフ情報源の極限分布

- 正規マルコフ情報源の定義より、 $t > t_0$ となる任意の t について

$$p_{ij}^{(t)} > 0 \quad (i, j = 0, 1, \dots, N-1)$$

が成り立つようなある正定数 t_0 が存在する。

- 正規マルコフ情報源では、 $t \rightarrow \infty$ とするとき、 $p_{ij}^{(t)}$ は i には無関係な値に収束する。[証明は省略]

すなわち、

$$\lim_{t \rightarrow \infty} p_{ij}^{(t)} = u_j \quad (j = 0, 1, \dots, N-1)$$

となる u_j が存在する。

- これを行列の形に書き直すと

$$\lim_{t \rightarrow \infty} \Pi^t = U \quad \text{-----} \quad \textcircled{2}$$

となる。ここで、 U はすべての行が

$$\mathbf{u} = (u_0, u_1, \dots, u_{N-1})$$

となる $N \times N$ 行列である。

十分時間が経てば、どの状態にもあり得るような情報源だから



正規マルコフ情報源の極限分布(つづき)

- 時点 t において状態 s_j にいる確率を $w_j^{(t)}$ で表し、

$$\mathbf{w}_t = (w_0^{(t)}, w_1^{(t)}, \dots, w_{N-1}^{(t)})$$

という N 次元ベクトルを定義する。 \mathbf{w}_t は時点 t における状態の確率分布を表すものであるから、**状態確率分布ベクトル**または簡単に**状態分布**と呼ぶ。

- 時点 $t-1$ で状態 s_i にいる確率が $w_i^{(t-1)}$ であり、状態 s_i にいるときの次の時点で状態 s_j に遷移する確率が p_{ij} であるから、

$$w_j^{(t)} = \sum_{i=0}^{N-1} w_i^{(t-1)} p_{ij}$$

\mathbf{w}_{t-1} と Π の j 列の積と同じ

となる。したがって、

$$\mathbf{w}_t = \mathbf{w}_{t-1} \Pi$$

であり、これを繰り返せば次式を得る。

$$\mathbf{w}_t = \mathbf{w}_0 \Pi^t \quad \text{-----} \quad \textcircled{3}$$

ここに、 \mathbf{w}_0 は時点0における状態分布である。これを**初期分布**と呼ぶ。

- \mathbf{w}_t の $t \rightarrow \infty$ とした極限を**極限分布**と呼ぶ。

– 極限分布は一般に存在するとは限らない

– 正規マルコフ情報源では、②と③より

$$\lim_{t \rightarrow \infty} \mathbf{w}_t = \mathbf{w}_0 \lim_{t \rightarrow \infty} \Pi^t = \mathbf{w}_0 \mathbf{U} = \mathbf{u}$$

すなわち、極限分布は \mathbf{u} となる。



正規マルコフ情報源の極限分布(つづき)

- 十分時間が経過すれば、初期分布がどうであれ、状態分布は定常的な確率分布、すなわち定常分布に落ち着く。
- 正規マルコフ情報源が落ち着く定常分布を

$$w = (w_0, w_1, \dots, w_{N-1})$$

とする。もちろん、

$$w_0 + w_1 + \dots + w_{N-1} = 1 \quad \text{-----} \quad \textcircled{4}$$

である。

- ある時点の状態分布が定常的で w であるとするれば、次の時点の状態分布も w でなければならないので、 w は

$$w\Pi = w \quad \text{-----} \quad \textcircled{5}$$

を満たさなければならない。

正規マルコフ情報源の遷移確率行列 Π に対しては、式⑤を満たす w が唯一存在し、極限分布と一致する。

- 十分時間が経過した後はエルゴード的であることも示されている。[証明は省略]

大事なのは、式④と式⑤！



正規マルコフ情報源の定常分布

- 例) 下図のマルコフ情報源の定常分布 $w = (w_0, w_1, w_2)$ を求める。

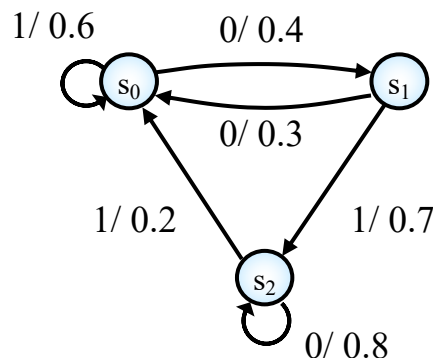
式⑤ $w\Pi = w$ より、

$$\begin{cases} 0.6 w_0 + 0.3 w_1 + 0.2 w_2 = w_0 \\ 0.4 w_0 & = w_1 \\ 0.7 w_1 + 0.8 w_2 = w_2 \end{cases}$$

さらに、式④より $w_0 + w_1 + w_2 = 1$ を満たさなければならない。
これらの式から

$$w_0 = 0.3571, \quad w_1 = 0.1429, \quad w_2 = 0.5$$

が求まる。



$$\Pi = \begin{bmatrix} 0.6 & 0.4 & 0 \\ 0.3 & 0 & 0.7 \\ 0.2 & 0 & 0.8 \end{bmatrix}$$

演習2



マルコフ情報源の定常分布

- (一般の)マルコフ情報源には、定常分布は1つ以上存在
- 初期分布として定常分布を与えれば定常情報源となる
- 既約情報源の場合

定常分布は一意に存在。初期分布として定常分布を与えればエルゴード情報源となる。

– 非周期的 (正規マルコフ情報源)

初期分布としてどんな分布を与えても十分時間がたてば定常情報源でありエルゴード情報源とみなせる。

– 周期的

初期分布として定常分布以外を与えると十分時間がたったあとも定常分布にならない場合がある。

