

4.9 正規マルコフ情報源

◆単純マルコフ情報源を対象とする.

情報源アルファベット: $S = \{s_1, s_2, \dots, s_n\}$

状態集合: $Q = S = \{s_1, s_2, \dots, s_n\}$

状態遷移行列: $P = \{P_{kl}\}, P_{kl}: s_k \rightarrow s_l$

$$P^t = \{P_{kl}^{(t)}\}, P_{kl}^{(t)}: s_k \rightarrow t\text{ステップ} \rightarrow s_l$$

◆理想的情報源モデルの条件

[条件1] 全ての記号が万遍なく発生する.

[条件2] 周期性を持たない.

[条件3] 初期状態によらない.

正規(正則)マルコフ情報源

正規(正則)マルコフ情報源

$$\forall k, l, \quad \exists t \quad P_{kl}^{(t)} > 0$$

任意の k, l に対して, $P_{kl}^{(t)} > 0$ となるある t が存在するマルコフ情報源が正規(正則)マルコフ情報源である.

あるステップにおいて P^t の全ての成分が正となる.

初期値が何であろうと, t ステップ後には全ての情報源記号が出現確率を持つ.

初期値によらず全ての情報源記号が万遍なく出現する.

P^t の例題

$$\mathbf{P} = \begin{bmatrix} 0.2 & 0 & 0.8 \\ 0.5 & 0.5 & 0 \\ 0 & 0.9 & 0.1 \end{bmatrix} \rightarrow \mathbf{P}^2 = \begin{bmatrix} 0.04 & 0.72 & 0.24 \\ 0.35 & 0.25 & 0.4 \\ 0.45 & 0.54 & 0.01 \end{bmatrix}$$

正規マルコフ情報源の性質

<性質4. 1>

(i) $\lim_{t \rightarrow \infty} \mathbf{P}^t = \mathbf{A}$ 収束値行列に収束する

(ii) $\mathbf{A} = \boldsymbol{\theta} \mathbf{w}$

$$\boldsymbol{\theta} = (1, 1, \dots, 1)^T, \quad \mathbf{w} = (w_1, w_2, \dots, w_n)$$

確率ベクトル

(iii) $\forall k \quad w_k > 0, \quad \sum_{k=1}^n w_k = 1$
全ての情報源記号が出現する

(iv) $\exists_1 \mathbf{z} \quad \mathbf{z} = \mathbf{zP}, \quad \mathbf{z} = \mathbf{w}$

定常分布 $\mathbf{z}$ はただ1つ存在し, 極限分布 $\mathbf{w}$ と一致する

収束値行列 A の形式

$$A = \theta \mathbf{w} = \begin{bmatrix} 1 \\ 1 \\ \cdot \\ \cdot \\ 1 \end{bmatrix} (w_1, w_2, \dots, w_n) = \begin{bmatrix} w_1 & \dots & w_n \\ w_1 & \dots & w_n \\ \cdot & \dots & \cdot \\ \cdot & \dots & \cdot \\ w_1 & \dots & w_n \end{bmatrix} = \begin{bmatrix} \mathbf{w} \\ \mathbf{w} \\ \cdot \\ \cdot \\ \mathbf{w} \end{bmatrix}$$

収束値行列 A の各行は同じ確率ベクトル $\mathbf{w}$ となる.

$\mathbf{w}$: 収束値行列(極限行列)の行を構成する $\rightarrow$ 極限分布

定常分布 z の例

$$P = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

$$(z_1, z_2) = (z_1, z_2) \begin{bmatrix} \frac{1}{3} & \frac{2}{3} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

$$z_1 = \frac{1}{3}z_1 + \frac{1}{2}z_2, \quad z_2 = \frac{2}{3}z_1 + \frac{1}{2}z_2$$

$$4z_1 = 3z_2, \quad z_1 + z_2 = 1$$

$$z_1 = \frac{3}{7}, \quad z_2 = \frac{4}{7}$$

$$z = (z_1, z_2) = \left(\frac{3}{7}, \frac{4}{7} \right)$$

第5章 情報源符号化

情報源記号($a, b, c, \dots$)系列を符号(0, 1)系列に変換し、
実際の伝送や保存を可能とする.

5.1 符号化

情報源記号を符号に変換

情報源アルファベット $S = \{s_1, \dots, s_n\}$ に対して
符号 $C = \{c_1, c_2, \dots, c_n\}$ (符号語, 簡単に符号と呼ぶ) を
構成する $\rightarrow$ 符号化

$$\phi_k: s_k \rightarrow c_k$$

符号を構成する記号集合

$$c_k = c_1^k c_2^k \dots c_{l_k}^k \quad l_k: \text{符号語長}$$

$$c_i^k (i = 1, 2, \dots, l_k) \in X = \{x_1, x_2, \dots, x_r\} \quad r\text{元符号}$$

Cの平均符号長

$$L = \sum_{k=1}^n l_k P(s_k)$$

符号化の例

$$S = \left\{ \begin{array}{cccc} a & b & c & d \\ 0.4 & 0.3 & 0.2 & 0.1 \end{array} \right\}$$

$$a \rightarrow 0, \quad b \rightarrow 10, \quad c \rightarrow 110, \quad d \rightarrow 1110$$

$$l_a = 1, \quad l_b = 2, \quad l_c = 3, \quad l_d = 4$$

平均符号長

$$\begin{aligned} L &= l_a P(a) + l_b P(b) + l_c P(c) + l_d P(d) \\ &= 1 \times 0.4 + 2 \times 0.3 + 3 \times 0.2 + 4 \times 0.1 = 2 \end{aligned}$$

等長符号: 任意の k に対して l_k 一定

可変長符号: 任意の k に対して l_k 可変

5.2 符号のクラス

復号化(復号) 符号 $\rightarrow$ 元の記号 $\phi_k^{-1}: \mathbf{c}_k \rightarrow s_k$

符号のクラス分け(復号化の観点から)

[I]一意に復号化不可能な符号 (例5.2)

特異符号 (例5.1)

[II]一意に復号化可能な符号

[II - 1]瞬時符号 (例5.3)

[II - 2]非瞬時符号 (例5.4)

符号クラスの問題

例5.1 $a \rightarrow 0, b \rightarrow 10, c \rightarrow 10, d \rightarrow 11$
 $10 \rightarrow b \text{ or } c$? 特異符号

例5.2 $a \rightarrow 0, b \rightarrow 01, c \rightarrow 10, d \rightarrow 11$
 $0110 \rightarrow ada \text{ or } bc$? 一意に復号化不可能

例5.3 $a \rightarrow 0, b \rightarrow 10, c \rightarrow 110, d \rightarrow 1110$ 瞬時符号
 $0 \rightarrow a \quad 10 \rightarrow b \quad 110 \rightarrow c \quad 1110 \rightarrow d$

例5.4 $a \rightarrow 0, b \rightarrow 01, c \rightarrow 011, d \rightarrow 0111$ 非瞬時符号
 $0 \rightarrow ? \quad 00 \rightarrow a0 \quad 01 \rightarrow ? \quad 010 \rightarrow b0$
 $011 \rightarrow ? \quad 0110 \rightarrow c0$

5.3 瞬時符号

符号木による解析

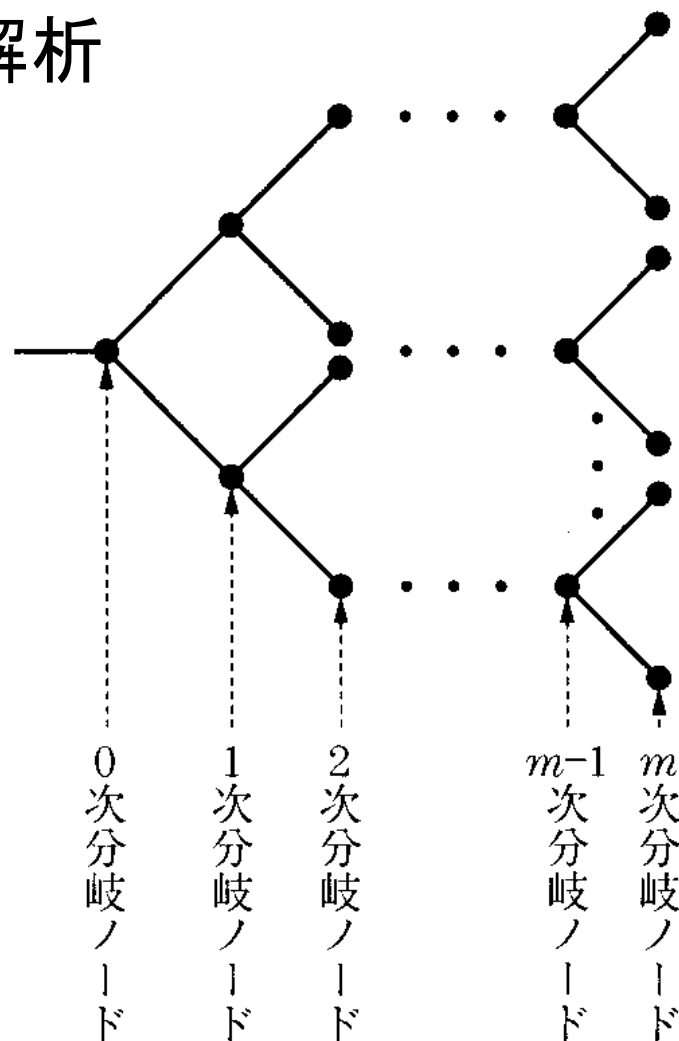
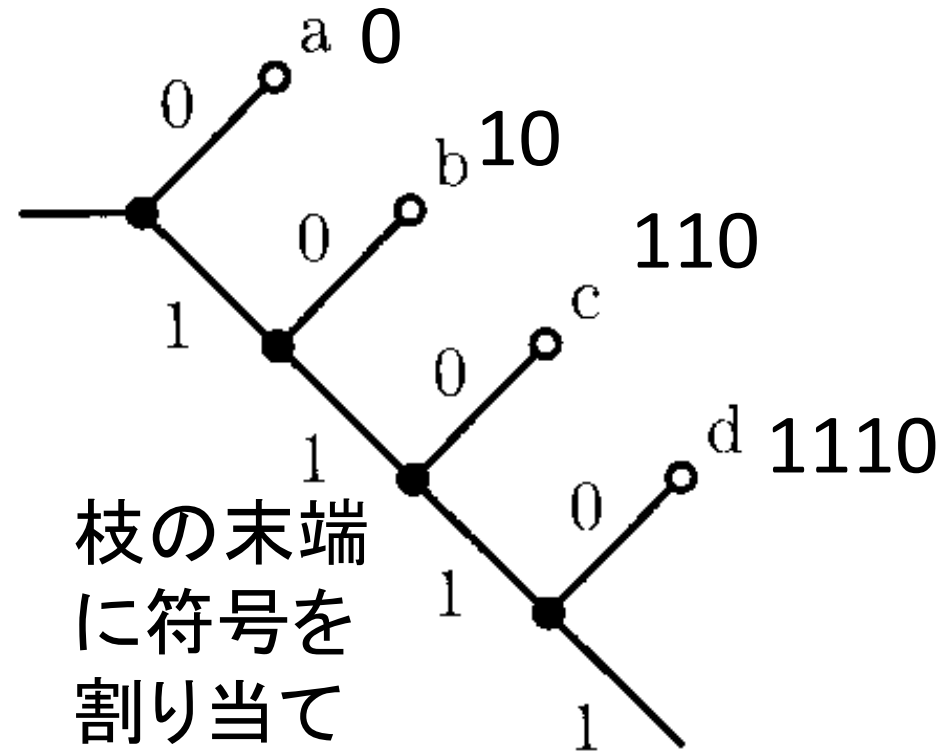
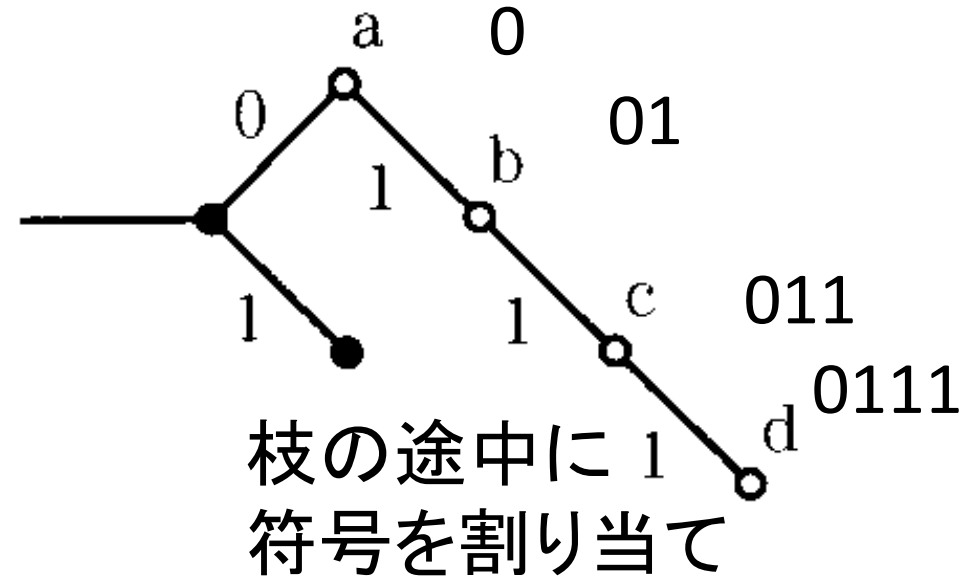


図 5.3 符号木 (2 元符号の場合)

符号木の比較



(1) 【例 5.3】 瞬時符号



(2) 【例 5.4】 非瞬時符号

図 5.4 符号木の比較

瞬時符号の特徴

性質5.1

- (1) 瞬時符号である.
- (2) 全ての符号語が符号木の枝の末端に割り当てられている.
- (3) 各符号語が他の符号の語頭(プレフィックス)になっていない.

性質5.2

等長符号は特異符号でなければ瞬時符号である.

5.4 クラフトの不等式 — 符号語長への制約 —

情報源符号化： 短い平均符号長を持つ符号を実現

[性質5.3]

符号語長 $\{l_1, l_2, \dots, l_n\}$ を持つ r 元瞬時符号が存在するための必要十分条件(クラフトの不等式)

$$\sum_{k=1}^n r^{-l_k} \leq 1$$

クラフトの不等式の例題

[例5.3], [例5.4]に対するクラフトの不等式

$$\sum_{k=1}^4 2^{-l_k} = \frac{1}{2^1} + \frac{1}{2^2} + \frac{1}{2^3} + \frac{1}{2^4} = \frac{15}{16} \leq 1$$

同じ符号長で瞬時符号と非瞬時符号が構成できる.

クラフトの不等式を満足する符号長で工夫すれば瞬時符号が構成できることを示している.

瞬時符号の性質

性質5.4

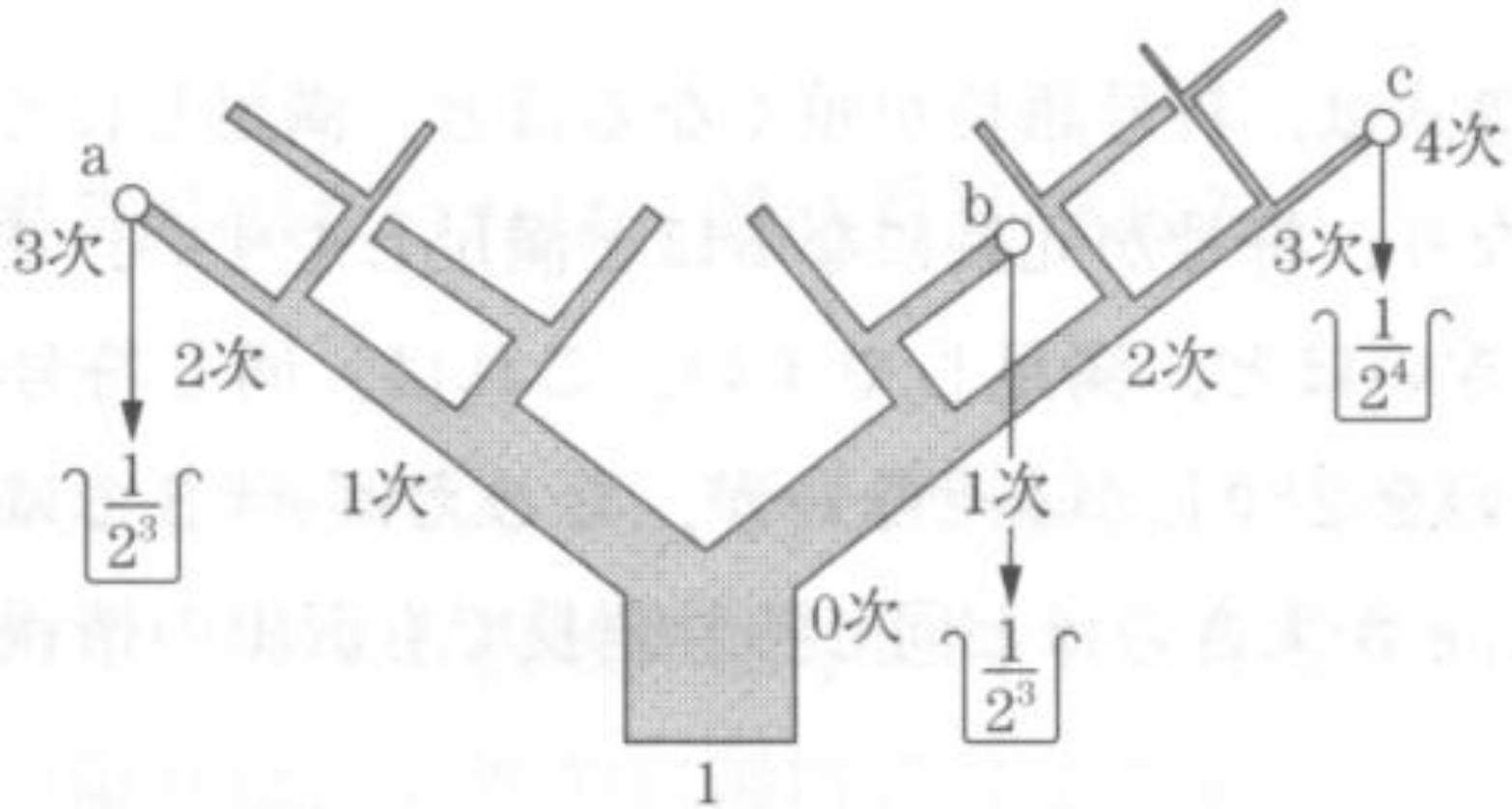
瞬時符号であればクラフトの不等式を満たす.
その逆は成り立たない.

性質5.5

符号語長 $\{l_1, l_2, \dots, l_n\}$ を持つ2元瞬時符号が存在するための必要十分条件(クラフトの不等式)

$$\sum_{k=1}^n 2^{-l_k} \leq 1$$

2元符号木の模式図



途中の枝に符号を割り当てる→水量をダブルカウント

→水量の総和> 1

枝の末端に符号を割り当てる→ダブルカウントしない

→水量の総和 $\leq$ 1

全ての末端に符号を割り当てる→水量の総和=1

5.5 拡大情報源

N次拡大情報源

情報源 S の記号から重複を許して N 個をとる順列を作り、それらを全て情報源記号(マクロな記号)とする情報源.

N: 記号列(記号ブロック)の長さ, マクロな記号の長さ(ブロック長)

$$S^N = \{s_1^N, s_2^N, \dots, s_{nN}^N\}$$
$$\forall k \quad s_k^N = s_{k_1} s_{k_2} \dots s_{k_N}, \quad \forall k, l \quad s_{k_l} \in S$$

2次拡大情報源の例

例5.5

$$S = \begin{pmatrix} a & b \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix}$$

2次拡大情報源

$$S^2 = \{aa, ab, ba, bb\}$$

$$P(aa) = P(a)P(a) = \frac{1}{3} \times \frac{1}{3} = \frac{1}{9}$$

$$P(ab) = P(a)P(b) = \frac{1}{3} \times \frac{2}{3} = \frac{2}{9}$$

$$P(ba) = P(b)P(a) = \frac{2}{3} \times \frac{1}{3} = \frac{2}{9}$$

$$P(bb) = P(b)P(b) = \frac{2}{3} \times \frac{2}{3} = \frac{4}{9}$$

$$S^2 = \begin{pmatrix} aa & ab & ba & bb \\ \frac{1}{9} & \frac{2}{9} & \frac{2}{9} & \frac{4}{9} \end{pmatrix}$$

5.6 情報源符号化定理 — 平均符号長の下限 —

[性質5.6] 無記憶情報源 S に対して, 次の条件を満足する平均符号長 L を持つ r 元瞬時符号が構成できる.

$$\frac{H(S)}{\log r} \leq L < \frac{H(s)}{\log r} + 1$$

[性質5.7] 無記憶情報源 S に対して, 次の条件を満足する平均符号長 L を持つ2元瞬時符号を構成できる.

$$H(S) \leq L < H(S) + 1$$

[定理5.1]情報源符号化定理(シャノンの第1基本定理)

無記憶情報源 S の N 次拡大情報源 S^N に対して, 次の条件を満足する1情報源記号当たりの平均符号長 L を持つ r 元瞬時符号が構成できる.

$$\forall \varepsilon \quad \frac{H(S)}{\log r} \leq L < \frac{H(S)}{\log r} + \varepsilon$$

5.7 符号の効率と冗長度

2元符号

$$L \geq H(S)$$

効率 e と冗長度 r の関係

$$e = \frac{H(S)}{L} \quad 0 \leq e \leq 1$$

$$r = 1 - e = 1 - \frac{H(S)}{L} = \frac{L - H(S)}{L} \quad 0 \leq r \leq 1$$

r 元符号: $H(S) \rightarrow H(S)/\log r$

5.8 コンパクト符号

情報源に対して最短な符号

コンパクト符号であつても効率 $e = 1$ とは限らない.

拡大次数 N を大きくする

→平均符号長 L は下限に近づく.

2元符号 $L \rightarrow H(S)$

r 元符号 $L \rightarrow H(S)/\log r$

→効率 e は1に近づく.