北海道大学
Hokkaido University

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5回

第3章 情報源のモデル[後半]

3.5 情報源のエントロピー

[復習]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 $A=\{a_1,a_2,\ldots,a_M\}$ 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 ($i = 0, 1, 2, \cdots$)で表す

確率変数

- 時点 n-1 まで（長さ n）の情報源系列 $X_0 X_1 X_2 \ldots X_{n-1}$ を考える

情報源系列 $X_0\, X_1\, X_2\, X_3 \;\cdots\; X_{n-2} X_{n-1}$

0 1 2 3 … n-2 n-1 時点

離散的M元情報源
$A=\{a_1,a_2,\ldots,a_M\}$

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

$X_0, X_1, \ldots, X_{n-1}$の結合確率分布

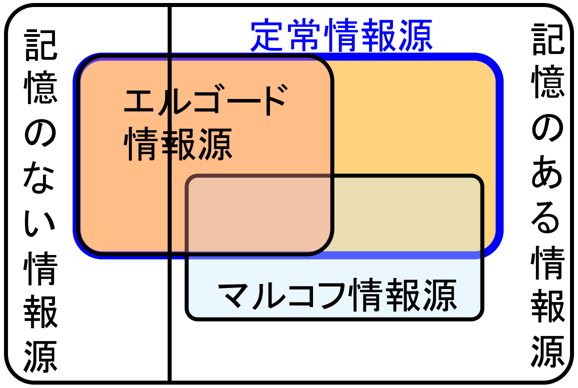
$P_{X_0, X_1 \ldots X_{n-1}}(x_0, x_1, \ldots, x_{n-1})=$〔$X_0 = x_0, X_1 = x_1, \ldots, X_{n-1} = x_{n-1}$ となる確率〕

[復習]扱いやすい情報源の性質

- 定常情報源
  - 時間をずらしても、統計的性質の変わらない情報源
  - 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 $x_0, x_1, \ldots, x_{n\text{-}1}$ に対し、以下が成立。

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

- エルゴード情報源
  - 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
  - 集合平均と時間平均が一致する。

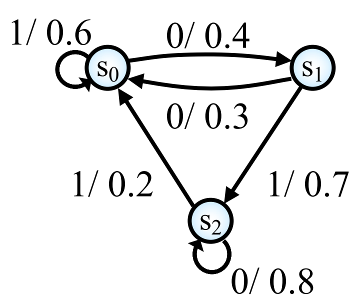


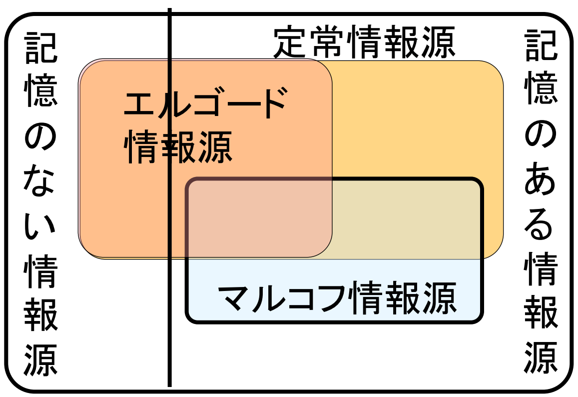
[復習]よく用いられる情報源

- 記憶のない定常情報源
  - 各時点における情報源記号の発生が、他の時点と独立で、各時点における情報源記号の発生が同一の確率分布 $P_X(x)$ に従う情報源

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

- 正規マルコフ情報源
  - 記憶のある情報源。
  - 閉じた状態集合からなる非周期的なマルコフ情報源
  - 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。



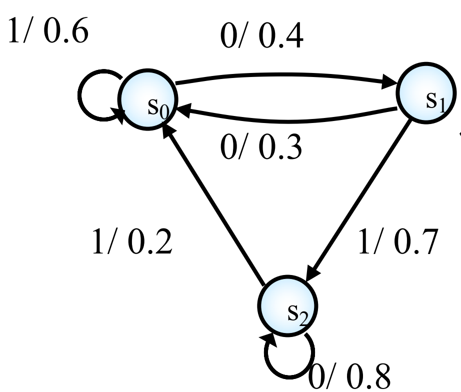


[復習]遷移確率行列

- N個の状態$\mathrm{s}_0, \mathrm{s}_1, \ldots, \mathrm{s}_{N\text{-}1}$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 $p_{ij}=P(\mathrm{s}_j \mid \mathrm{s}_i)$ 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

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

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なる。

$$\mathit{\Pi} = \begin{array}{c} \begin{bmatrix} 0.6 & 0.4 & 0 \\ 0.3 & 0 & 0.7 \\ 0.2 & 0 & 0.8 \end{bmatrix} \\ \begin{array}{ccc} \mathrm{s}_0 & \mathrm{s}_1 & \mathrm{s}_2 \end{array} \;\text{へ} \end{array} \begin{array}{l} \leftarrow \mathrm{s}_0 \\ \leftarrow \mathrm{s}_1 \\ \leftarrow \mathrm{s}_2 \end{array} \left.\vphantom{\begin{array}{l} a \\ b \\ c \end{array}}\right\} \text{から}$$

[復習]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 $\boldsymbol{w} = (w_0, w_1, w_2)$ を求める。
式⑤ $\boldsymbol{w}\Pi = \boldsymbol{w}$ より、

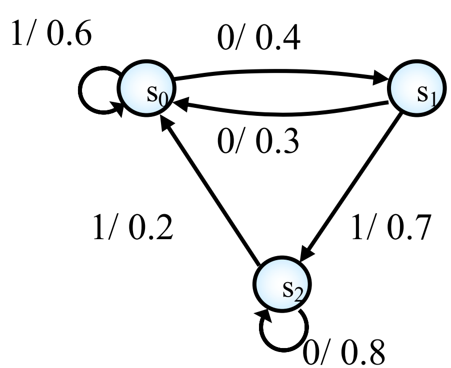
$$\left.\begin{array}{l} 0.6\,w_0 + 0.3\,w_1 + 0.2\,w_2 = w_0 \\ 0.4\,w_0 = w_1 \\ 0.7\,w_1 + 0.8\,w_2 = w_2 \end{array}\right\}$$

$\boldsymbol{w}^{\mathrm{T}}$は$\Pi^{\mathrm{T}}$の固有値1の固有ベクトル

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。



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

よくある間違い

遷移確率行列がΠのマルコフ情報源の定常分布を求めるとき、状態の定常分布を行ベクトル$\boldsymbol{w}$とすると定常性(一時刻後の分布は変わらない)を表す式は

$$\boldsymbol{w}\Pi = \boldsymbol{w}$$

これを列ベクトル$\boldsymbol{w}^{\mathrm{T}}$で右からかけて

$$\Pi\boldsymbol{w}^{\mathrm{T}} = \boldsymbol{w}^{\mathrm{T}}$$

と計算する人がいる。これを行うと必ず解が一様分布になる。

Πの行ベクトルの要素を足すと1なので一様分布が解になることは容易に確かめられる。従って、一様分布が解として出た時は、このミスをしている可能性があるので注意が必要。

[復習]確率変数のエントロピー

確率変数Xの取りうる値を$a_1, a_2, \cdots, a_M$とし、Xがそれぞれの値をとる確率が$p_1, p_2, \cdots, p_M$ （ただし、$p_1+p_2+\cdots+p_M=1$)であるとき、**確率変数XのエントロピーH(X)**は、

$$\mathrm{H}(X) = -\sum_{i=1}^{M} p_i \log_2 p_i$$

と定義される。

定常情報源の１次エントロピー

情報源記号列$X_0,X_1,\ldots$を出力する情報源Sのエントロピーはどのように定義すればよいか？

定常情報源であれば、$P_{X_i}(x)$は時点iによらず同じであるから、X_iのエントロピー$H(X_i)$は時点iによらず一定

定常情報源Sのエントロピーを$H(X_i)$で定義してはどうか。

各時点に情報源記号$a_1,a_2,\ldots,a_M$を出力する確率がそれぞれ$p_1,p_2,\ldots,p_M$であるようなM元定常情報源Sの**１次エントロピー**$H_1(S)$を

$$H_1(S) = -\sum_{i=1}^{M} p_i \log_2 p_i$$

と定義する。

１次エントロピーで情報源の曖昧さは測れるか？

どの時点でも出力情報源記号Xの確率分布$P_X(x)$が同じであれば
その情報源から出力される記号列X_0X_1･･･の曖昧さは同じ？

無記憶定常情報源S

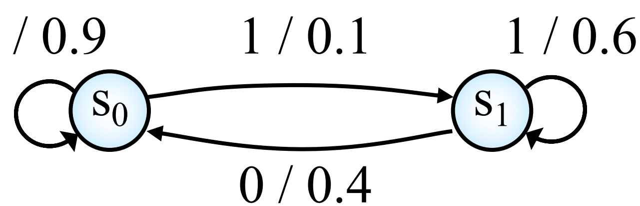
$P_X(0)=0.8$
$P_X(1)=0.2$

ともに各時点において0, 1が出力される確率はそれぞれ0,8と0.2

どちらも1次エントロピーは同じ

$H_1(S)=-0.8\log_2 0.8-0.2\log_2 0.2=0.7219$

マルコフ情報源S



Sの定常分布(w_0, w_1)は、

$$(w_0, w_1)\begin{pmatrix} 0.9 & 0.1 \\ 0.4 & 0.6 \end{pmatrix} = (w_0, w_1)$$

および $w_0+w_1=1$ から、$w_0=0.8$、$w_1=0.2$

$P_X(0)=0.8\times0.9+0.2\times0.4=0.8$

$P_X(1)=0.8\times0.1+0.2\times0.6=0.2$

もう少し長い系列の分布でみたら

定常情報源であればX_iX_{i+1}の結合確率分布$\mathrm{P}_{X_iX_{i+1}}(x,y)$も時点$i$によらず一定

X_0とX_1の結合エントロピーに情報源の曖昧さの差が現れるのでは？

無記憶定常情報源S

$\mathrm{P_X}(0)=0.8$

$\mathrm{P_X}(1)=0.2$

$\mathrm{P}_{X0X1}(0,0)=0.64$

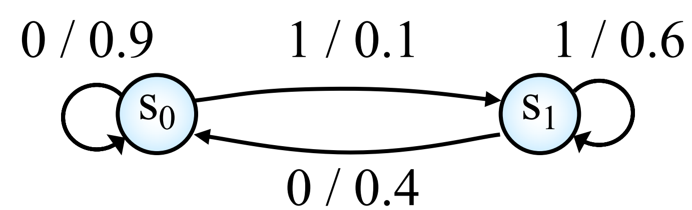
$\mathrm{P}_{X0X1}(0,1)=0.16$

$\mathrm{P}_{X0X1}(1,0)=0.16$

$\mathrm{P}_{X0X1}(1,1)=0.04$

$\mathrm{H}(X_0,X_1)=1.444$

マルコフ情報源S



$\boldsymbol{S}$の定常分布:$(\boldsymbol{w}_0, \boldsymbol{w}_1)=(0.8,0.2)$

$\mathrm{P_X}(0)=0.8$　$\mathrm{P_X}(1)=0.2$

$\mathrm{P}_{X0X1}(0,0)=0.8\times0.9\times0.9+0.2\times0.4\times0.9=0.72$

$\mathrm{P}_{X0X1}(0,1)=0.8\times0.9\times0.1+0.2\times0.4\times0.1=0.08$

$\mathrm{P}_{X0X1}(1,0)=0.8\times0.1\times0.4+0.2\times0.6\times0.4=0.08$

$\mathrm{P}_{X0X1}(1,1)=0.8\times0.1\times0.6+0.2\times0.6\times0,6=0.12$

$\mathrm{H}(X_0,X_1)=1.291$

無記憶定常情報源の方が曖昧さが大きい！

n次エントロピー

M元情報源Sのn次拡大情報源S^n

Sが出力する連続するn個の情報源記号をまとめて1つの情報源記号とみたとき、それを発生するM^n元情報源

n次拡大情報源S^nの1次エントロピー$H_1(S^n)$

$H_1(S^n)=H(X_0,X_1,\ldots,X_{n-1})$ ← n確率変数の結合エントロピー

情報源Sのn次エントロピー$H_n(S)$

$H_n(S)=H_1(S^n)/n=H(X_0,X_1,\ldots,X_{n-1})/n$

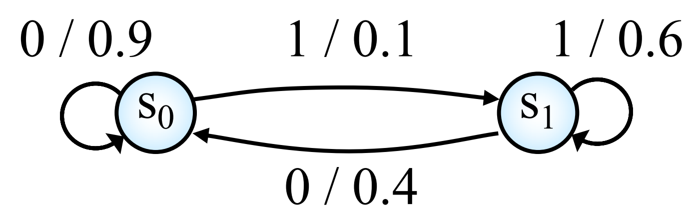
無記憶定常情報源S

$P_X(0)=0.8$

$P_X(1)=0.2$

$H_2(S)=H(X_0,X_1)/2=1.444/2=0.722$

マルコフ情報源S



$H_2(S)=H(X_0,X_1)/2=1.291/2=0.646$

情報源のエントロピー

時刻iに情報源記号X_iを出力する定常情報源SのエントロピーH(S)は

$$H(S)=\lim_{n\to\infty}H_n(S)=\lim_{n\to\infty}H_1(S^n)/n=\lim_{n\to\infty}H(X_0,X_1,\ldots,X_n)/n$$

で定義される。

ただし、$H(X_0,X_1,\ldots,X_{n-1})$は$X_0,X_1,\ldots,X_{n-1}$の結合エントロピーであり、

$$H(X_0,X_1,\ldots,X_{n-1})=-\sum_{x_0}\sum_{x_1}\cdots\sum_{x_{n-1}}P(x_0,x_1,\ldots,x_{n-1})\log_2 P(x_0,x_1,\ldots,x_{n-1})$$

で計算される。

H(*S*)=情報源Sから得られる1記号あたりの平均情報量

無記憶定常情報源のエントロピー

無記憶 定常情報源Sに対し$X_0, X_1, \ldots, X_{n-1}$は互いに独立で、同じ分布$P(x)$に従うので

$$H(X_0, X_1, \ldots, X_{n-1}) = -\sum_{x_0, x_1, \ldots, x_{n-1}} P(x_0, x_1, \ldots, x_{n-1}) \log P(x_0, x_1, \ldots, x_{n-1})$$

$$= -\sum_{x_0, x_1, \ldots, x_{n-2}} \sum_{x_{n-1}} P(x_0, x_1, \ldots, x_{n-2}) P(x_{n-1}) \log P(x_0, x_1, \ldots, x_{n-2}) P(x_{n-1})$$

$$= -\sum_{x_0, x_1, \ldots, x_{n-2}} P(x_0, x_1, \ldots, x_{n-2}) \log P(x_0, x_1, \ldots, x_{n-2}) - \sum_{x_{n-1}} P(x_{n-1}) \log P(x_{n-1})$$

$$= H(X_0, X_1, \ldots, X_{n-2}) + H(X_{n-1}) = \cdots = \sum_{i=0}^{n-1} H(X_i) = nH(X_0)$$

したがって、SのエントロピーH(S)は

$$\boldsymbol{H(S)} = \lim_{n\to\infty} H(X_0, X_1, \ldots, X_{n-1})/n = \lim_{n\to\infty} \mathrm{H}(X_0) = \mathrm{H}(X_0) = \boldsymbol{-\sum P(x) \log P(x)}$$

マルコフ情報源のエントロピー

- 図のマルコフ情報源S のエントロピーを考える

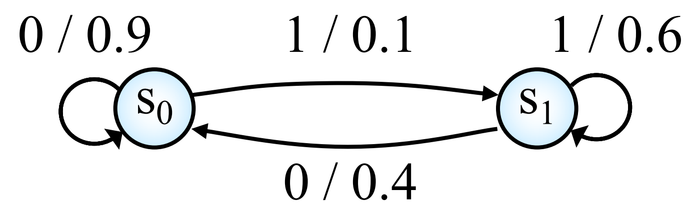


図: マルコフ情報源S

Sの定常分布(w_0, w_1)は、(0.8,0.2)。

いまS が状態 s_0 にあるときだけに注目すると、この情報源は
1, 0 を 0.1, 0.9 の確率で発生する無記憶情報源とみなせる。
その場合のエントロピーを $Hs_0(S)$ と書くと

$$Hs_0(S) = \mathcal{H}(0.1) = 0.4690$$

$$\mathcal{H}(x) = -x\log x - (1-x)\log(1-x)$$

同様に、Sが状態 s_1 にあるときだけを注目すれば

$$Hs_1(S) = \mathcal{H}(0.6) = 0.9710$$

(定常分布では)s_0にいる確率が0.8、s_1にいる確率が0.2だから、$H(S)$は

$$H(S) = 0.8 \times 0.4690 + 0.2 \times 0.9710 = 0.5694$$

になると考えられる。

一般のマルコフ情報源のエントロピー

一般のマルコフ情報源のエントロピー

情報源アルファベット $\{a_1, a_2, \cdots, a_M\}$
状態 $s_0, s_1, \cdots, s_{N-1}$
定常分布 $(w_0, w_1, \cdots, w_{N-1})$
状態s_i にあるときに記号 a_j を発生する確率 $P(a_j \mid s_i)$
この情報源に対するエントロピー$H(S)$は

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

マルコフ情報源のn次エントロピー

nを大きくしていくと、図1のマルコフ情報源Sのn次エントロピー$H_n(S)=H_1(S^n)/n$は図2のように0.5694に収束していく。

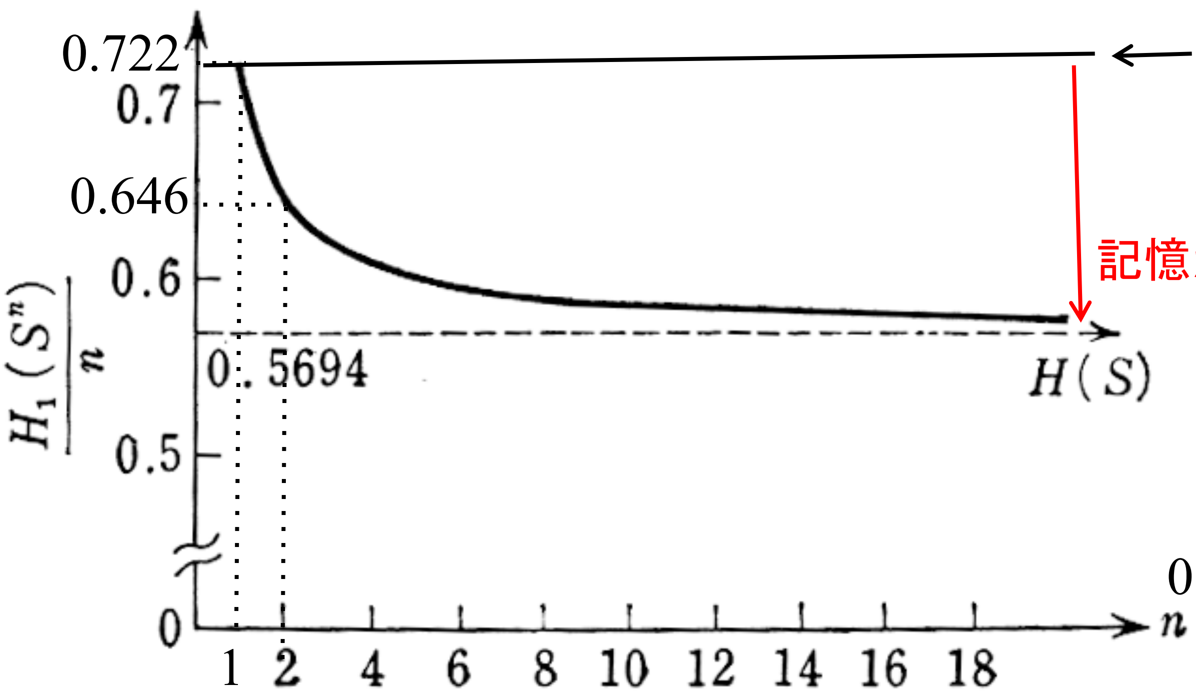


図2: マルコフ情報源のエントロピー

予想しやすくなっている
（圧縮しやすくなっている）

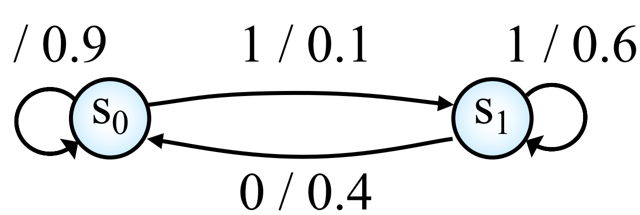


図1: マルコフ情報源S