情報メディア創成学類『情報理論』演習問題プリント

担当：　システム情報工学研究科ＣＳ専攻　工藤博幸

確率基礎

(1) あるつぼに白色と赤色の玉が合わせて 10 個はいっている．つぼから玉を 1 個取り出して色を確かめてからもとにもどし，よくかきまぜる．次に，また玉を 1 個取り出して同じことをを繰り返す．このような実験を n 回行ったとき，取り出された玉はいずれも白色ばかりであった．このとき，つぼに白色のみがはいっている確率はいくらか．

(答) $1/\sum_{k=1}^{10}(k/10)^n$

(2) 整数の集合 $\{1,2,3,4\}$ の中から任意に 1 個選んだ数を x_1 とする．次に，整数の集合 $\{1,2,\ldots,x_1\}$ の中から任意に 1 個選んだ数を x_2 とするとき，次の諸確率を求めよ．

(a) $P(x_2=1 \mid x_1=k)$，(b) $P(x_2=1)$，(c) $P(x_1=2 \mid x_2=1)$

(答) (a) $1/k$，(b) $25/48$，(c) $6/25$

第 1 章　『情報とその表現』

(1-1) 離散パラメータ-離散的通報，離散パラメータ-連続的通報，連続パラメータ-離散的通報，連続パラメータ-連続的通報の例を各 3 つ以上あげよ．

(1-2) 情報源のモデルとしてマルコフ過程が適していることを，(教科書に示されているもの以外の) 具体的な例をあげて説明せよ．

(1-3) シャノンの情報伝送系のモデルを図示し，各構成要素の役割を説明せよ．

(1-4) 情報源符号化と通信路符号化の相違について説明せよ．

第 2 章　『情報量』

(2-1) 情報源 A の基本記号を a_1,a_2,a_3 とし，それぞれの確率を $p_1=1/2, p_2=p_3=1/4$ とする．エントロピー $H(A)$ を計算せよ．次に，情報源 A からの相続く記号は互いに独立に生起するとし，A から発生した 2 つの記号 (x_1,x_2) の組を 1 つの通報と考える新しい情報源 A^2 を考える．エントロピー $H(A^2)$ を求めよ．

(答) $H(A)=1.5$ ビット，$H(A^2)=3$ ビット

(2-2) 通報 u_0,u_1,u_2,u_3 がそれぞれ確率 $1/2, 1/4, 1/8, 1/8$ で生起し，それらが 4 つの符号語 $0, 10, 110, 111$ で表されて伝送されるものとする．このとき，相続く通報が互いに独立に生起するならば，得られる符号語の系列において，数字 0 と数字 1 はそれぞれどんな確率で生起するか．また，1 通報ごとのエントロピーは何ビットになるか．

(答) $P(0)=P(1)=1/2$，1.75 ビット

(2-3) 印刷された英文について，各文字の使用頻度がそれぞれ次のような値であるとき，1 文字あたりのエントロピー $H(X)$ を求めよ．但し，隣り合う文字間の相関はないものとする．

E,T　$\rightarrow$ 各 1/8
A,I,N,O,S,H,R,D　$\rightarrow$ 各 1/16
L,U,C,M　$\rightarrow$ 各 1/32
F,W,Y,G,P,B　$\rightarrow$ 各 1/64

V,K　→ 各 1/128

Q,J,X,Z　→ 各 1/256

(答) 4.17 ビット

(2-4) 2個のサイコロを考える．第1のサイコロの目を確率変数 X，第2のサイコロの目を確率変数 Y，第1のサイコロと第2のサイコロの目の和を確率変数 Z とするとき，X のエントロピー $H(X)$，Z のエントロピー $H(Z)$，X と Z の平均相互情報量 $\overline{I}(X;Z)$ を求めよ．

(答) $H(X)$ =2.585 ビット，$H(Z)$ =3.277 ビット，$\overline{I}(X;Z) = 0.692$ ビット

(2-5) 2つの確率変数 Y, X の同時確率が下表で与えられているとする．但し，$X = \{x_1, x_2, x_3, x_4, x_5\}, Y = \{y_1, y_2, y_3, y_4\}$ である．エントロピー $H(X,Y), H(X), H(Y), H(Y \mid X), H(X \mid Y)$ と平均相互情報量 $\overline{I}(X;Y)$ を求めよ．

$$
\begin{array}{c}
\\ \mathrm{x1} \\ \mathrm{x2} \\ \mathrm{x3} \\ \mathrm{x4} \\ \mathrm{x5}
\end{array}
\begin{array}{c}
\begin{array}{cccc} \mathrm{y1} & \mathrm{y2} & \mathrm{y3} & \mathrm{y4} \end{array} \\
\begin{bmatrix}
0.25 & 0 & 0 & 0 \\
0.10 & 0.30 & 0 & 0 \\
0 & 0.05 & 0.10 & 0 \\
0 & 0 & 0.05 & 0.10 \\
0 & 0 & 0.05 & 0
\end{bmatrix}
\end{array}
$$

(答) $H(X,Y) = 2.665$ ビット，$H(X) = 2.066$ ビット，$H(Y) = 1.856$ ビット，$H(Y \mid X) = 0.600$ ビット，$H(X \mid Y) = 0.809$ ビット，$\overline{I}(X;Y) = 1.257$ ビット

(2-6) 実際の天気 X と天気予報 Y の結合確率分布が下表のように与えられている．

(a) X と Y の平均相互情報量を求めよ．

(b) 常に晴だと予報する天気予報を Y' とする．X と Y' の平均相互情報量を求めよ．

(c) 天気予報 Y と Y' を適中率と平均相互情報量から比較し論ぜよ．

P(x,y)		Y	
		晴	雨
X	晴	0.50	0.25
	雨	0.10	0.15

(答) (a) 0.0394 ビット，(b) 0，(c) 適中率は Y' の方が高いが Y' は何の情報も与えない

(2-7) 連続パラメータ-連続的情報を離散パラメータ-離散的情報に変換する手順を説明せよ．

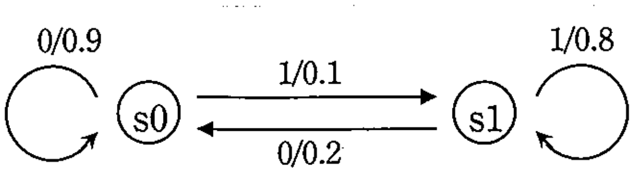
第3章　『情報源と符号化』

(3-1) 下図のマルコフ情報源について次の問いに答えよ．

(a) 定常確率 $P(0), P(1)$ を求めよ．

(b) 1記号あたりのエントロピー $H(X)$ を求めよ．

(c) この情報源をシャノンの第1符号化定理を満たすように符号化する方法を述べ，実際に符号化しシャノンの符号化定理を満たすことを示せ．



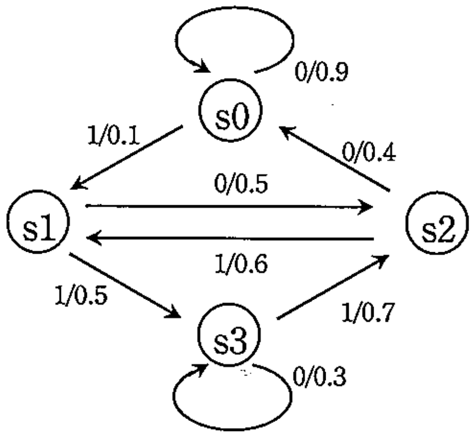
(答) (a) $P(0) = 2/3, P(1) = 1/3$, (b) 0.553 ビット, (b) 情報源から生起する n 個の記号の組を 1 つの通報と考え，$n = 2, 3, \ldots$ としていく

(3-2) 下図のマルコフ情報源について，次の確率を求めよ．但し，情報源は定常的になっているものとする．

(a) 状態 s_0, s_1, s_2, s_3 にいる確率 w_0, w_1, w_2, w_3

(b) 出力が 0 となる確率 $P_X(0)$

(c) 出力に 0 が n 個続いた後に 0 が出る確率 $P(0 \mid 0^n)$（但し，$n \geq 2$ とする）



(答) (a) $w_0 = 28/47, w_1 = 7/47, w_2 = 7/47, w_3 = 5/47$, (b) $P_X(0) = 33/47$, (c) $P(0 \mid 0^n) = (5.32 \times (0.9)^{n-1} + (0.3)^{n+1})/(5.32 \times (0.9)^{n-2} + (0.3)^n)$

(3-3) $A = \{0, 1\}$ の 2 つの基本記号を持つ 2 重マルコフ過程があって，その遷移確率が次のように与えられている．

$$P(0 \mid 00) = P(1 \mid 11) = 0.8,\ P(1 \mid 00) = P(0 \mid 11) = 0.2$$

$$P(0 \mid 01) = P(0 \mid 10) = P(1 \mid 01) = P(1 \mid 10) = 0.5$$

(a) 状態遷移図を描け．

(b) 情報源 A のエントロピーを求めよ．

(答) (a) 略, (b) 0.801 ビット

(3-4) $A = \{a_1, a_2, \ldots, a_6\}$ の 6 つの記号の集合に対する 6 つのブロック符号 $C_1, C_2, C_3, C_4, C_5, C_6$ が下表に示されている．

(a) どの符号が一意復号可能か．

(b) どの符号が瞬時復号可能か．

(c) 一意復号可能な符号系の平均長を求めよ．

記号	確率	C1	C2	C3	C4	C5	C6
a1	1/2	0	1	0	111	1	0
a2	1/4	10	011	10	110	01	01
a3	1/16	110	010	110	101	0011	011
a4	1/16	1110	001	1110	100	0010	0111
a5	1/16	1011	000	11110	011	0001	01111
a6	1/16	1101	110	111110	010	0000	011111

(答) (a) C_3, C_4, C_5, C_6，(b) C_3, C_4, C_5，(c) C_3, C_4, C_5, C_6 の各々について 17/8,3,2,17/8

(3-5) 10 個の基本記号を持つ情報源に対して符号長が 1 のものを 1 個，2 のものを 5 個，3 のものを 4 個持つような 2 元符号系は存在しないことを示せ．

(答) クラフトの不等式を満たさないことを示せばよい

(3-6) シャノンの第 1 符号化定理について説明せよ．

(3-7) 下表のような確率分布を持つ記憶のない情報源を 2 元ハフマン符号化及び 4 元ハフマン符号化せよ．

記号	確率	記号	確率
a0	0.363	a4	0.087
a1	0.174	a5	0.069
a2	0.143	a6	0.045
a3	0.098	a7	0.021

(答) 例えば，$a_0, a_1, \dots, a_7$ を 1,001,010,0000,0001,0110,01110,01111 及び 0,2,3,10,11,12,130,131 と符号化すればよい．

第 4 章　『離散的通信路と情報の信頼性』

(4-1) 入力記号 $A = \{0, 1\}$，出力記号 $B = \{0, 1\}$ の 2 元通信路がある．この通信路の通信路行列は

$$P = \begin{pmatrix} 1/2 & 1/2 \\ 1/4 & 3/4 \end{pmatrix}$$

で与えられる．この通信路の入力に記憶のない情報源をつなぎ，記号 0 を確率 1/3，記号 1 を確率 2/3 で送信するとき，以下のものを求めよ．

(a) 出力記号の確率 $P(B)$ 及び事後確率 $P(A \mid B)$

(b) 入力側情報源のエントロピー $H(A)$

(c) 記号 0 を受信したときの条件付きエントロピー $H(A \mid 0)$，1 を受信したときの条件付きエントロピー $H(A \mid 1)$

(d) あいまい度 $H(A \mid B)$

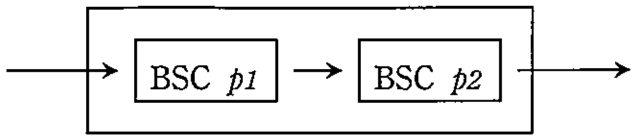
(e) 情報伝送量 $\overline{I}(A; B)$

(答) (a) $P(0) = 1/3, P(1) = 2/3, P(0 \mid 0) = 1/2, P(1 \mid 0) = 1/2, P(0 \mid 1) = 1/4, P(1 \mid 1) = 3/4$，(b) 0.918 ビット，(c) $H(A \mid 0)$ =1 ビット,$H(A \mid 1)$ =0.811 ビット，(d) 0.874 ビット，(e) 0.044 ビット

(4-2) 下図に示す誤り率が p_1, p_2 の２つの２元対称通信路（BSC）を縦続接続した通信路について以下の問いに答えよ．

(a) 通信路行列を求めよ．

(b) 通信容量を求めよ．

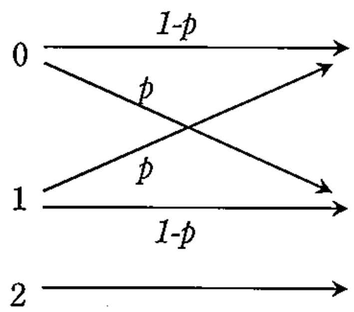


(答) (a) 通信路行列の各成分が $p_{11} = p_{22} = p_1p_2 + (1-p_1)(1-p_2), p_{21} = p_{12} = p_1(1-p_2) + (1-p_1)p_2$,
(b) $1 - H_n(p_1p_2 + (1-p_1)(1-p_2))$（但し，$H_n(x) = -x\log x - (1-x)\log(1-x)$）ビット／記号

(4-3) 下図に示す通信路について問いに答えよ．

(a) $p = 1/2$ のときの通信容量を求めよ．

(b) $p \neq 1/2$ のときの通信容量を p の関数として求め概略を図示せよ．



(答) (a) 1 ビット／記号，(b) $\log(1 + 2^{1+p\log p+(1-p)\log(1-p)})$ ビット／記号

(4-4) 問題 (4-1) の通信路の通信容量を求めよ．

(答) 0.048 ビット／記号（計算たいへん！）

(4-5) シャノンの第２符号化定理について説明せよ．

(4-6) $u_1 = 000, u_2 = 001, u_3 = 010, u_4 = 011, u_5 = 100, u_6 = 101, u_7 = 110, u_8 = 111$ の８つの記号列を，単一誤り訂正が可能なように組織符号化するとき以下の問いに答えよ．

(a) 検査記号の個数はいくつ以上でなくてはならないか．理由も述べよ．

(b) (a) の検査記号数を用いて符号化した例を示せ．

(a) ３つ以上，(b) 略

(4-7) 検査行列が次式で与えられる (15, 11) ハミングの単一誤り訂正符号を，２元対称消失通信路に流したとする．

$$H = \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{pmatrix}$$

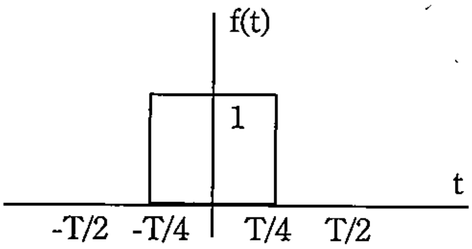
次のような受信語を復号せよ（X は消失を表す）．

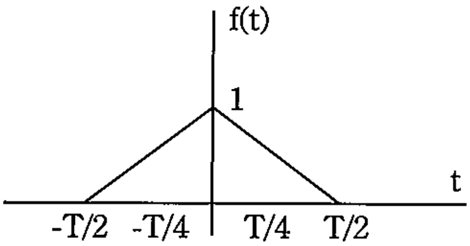
(a) $y = (100001010111011)$
(b) $y = (X01100101001110)$
(c) $y = (0X100X011001111)$

(答) (a) 左から１０番目の記号が誤り，(b) １消失で誤りがあるので訂正不可能，(c) 消失した記号はともに１

第５章　『情報伝送と信号』

(5-1) 下図に示す２つの信号 $f(t)$ がある．
(a) $[-T/2, T/2]$ の区間でフーリエ級数に展開せよ．
(b) $[-T/2, T/2]$ の外で零とみなしてフーリエ変換せよ．
(c) (a) と (b) の関係について論ぜよ．



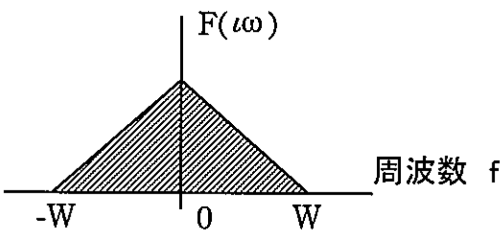


(答) (a) 左 $d_k = \sin(k\pi/2)/(k\pi)$，右 $d_k = 2\sin^2(k\pi/2)/(k\pi)^2$，(b)，左 $F(i\omega) = 2\sin(\omega T/4)/\omega$，右 $F(i\omega) = 8\sin^2(\omega T/4)/(\omega^2 T)$，(c) $d_k = F(i2\pi k/T)/T$ の関係にある

(5-2) 通信路の伝達関数を $H(i\omega)$ で表す．通信路の入力 $f(t)$ と出力 $g(t)$ の間に $g(t) = af(t-b)$ の関係がある（信号が歪まないで伝送される）ためには，$H(i\omega)$ はどのような形でなくてはならないか．

(答) $H(i\omega) = ae^{ib\omega}$

(5-3) 下図に示す周波数スペクトル $F(i\omega)$ を持つ（周波数帯域が $[-W, W]$ に制限された）信号 $f(t)$ がある．
(a) $f(t)$ をサンプリング間隔 Δ で標本化した信号の周波数スペクトルを，$\Delta < 1/(2W)$ の場合と $\Delta > 1/(2W)$ の場合に分けて図示せよ．
(b) $f(t)$ を標本化定理を用いて標本値から復元するためには，サンプリング間隔 Δ は $\Delta < 1/2W$ を満足している必要がある．何故 $\Delta > 1/2W$ だと復元できないかを (a) に関連づけて定性的に説明せよ．



(答) 略

(5-4) 信号変調の概要と目的を説明せよ．振幅変調と周波数変調の原理を説明せよ．

(答) 略

情報理論（情報メディア創成学類）実習：２値画像のエントロピー

担当：システム情報工学研究科ＣＳ専攻　工藤 博幸

１．はじめに

情報理論の講義の第２～３章において，情報源から発生した記号系列を無記憶情報源やマルコフ情報源によってモデル化し，発生する１記号あたりの実質的な情報量であるエントロピーを計算する方法を学んだ．この方法は，テキスト・ＷＥＢ・音楽・画像などの様々な情報メディアに含まれる実質的な情報量や圧縮の容易さ（即ち，情報にどれだけ無駄があるか）を知るのに応用できると考えられる．本実習では，ＦＡＸ画像などの各画素が白（０）と黒（１）の２値のみを取り得る２値画像のエントロピーを計算するプログラムを作成し，講義で学んだエントロピーに関する事項を復習するとともに，情報理論に関連したプログラム作成を体験してより理解を深めることを目的とする．

２．原理

２．１　２値画像と記号系列

まず，ＦＡＸ画像などの２値画像を，どのようにして（情報理論で学んだ）情報源から発生する記号系列とみなすかを説明する．画像の垂直座標を i，水平座標を j で表すと，２値画像 $f(i,j)$ とは各画素 (i,j) の色が白（０）と黒（１）のみを取り得る画像のことである．白・黒と０・１の対応については，（逆の場合もあり得るが）本実習では白を０に黒を１に対応させて表現されているとする．画像は i と j の座標軸が２つ存在するので，情報理論で学んだ情報源から発生する１次元記号系列とみなすには，画素値 $f(i,j)$ を表すデータを一列に並べ換える必要がある．この変換は，ラスター走査と呼ばれる図１に示す方法で行われる．説明のため，画像の画素数は $N \times N$ であると仮定する．まず，画像の１行目 $(i=1)$ を左から右に（$j=1,2,\ldots,N$ の順序で）走査して，画素値 $f(i,j)$ を一列に並べていく．次に，２行目を同じように左から右に（$j=1,2,\ldots,N$ の順序で）走査して，画素値 $f(i,j)$ を一列に並べていく．以下同様に，最後の行（$i=N$ 行目）まで走査を行う．その結果，N^2 個の０・１を並べた１次元記号系列が得られ，これは情報理論で学んだ情報源から発生した記号系列とみなすことができる．画像の右端と左端は座標平面上ではつながっていないが，（本実習では）簡単のためつながっているとみなして良いことにする．

２．２　エントロピーの計算

次に，情報理論の講義の第２～３章において学んだ方法に基づいて，２．１の方法で作成した０・１の記号系列から１記号あたりの実質的な情報量であるエントロピー $H(X)$ を計算する方法を説明する．以降では，N^2 個の０・１の記号系列を $x_1, x_2, \ldots, x_{N^2}$ と表す．まず，エントロピーを計算するには，情報源を無記憶情報源とみなすか，単純マルコフ情報源とみなすか，２次のマルコフ情報源とみなすか，使用する情報源の確率モデルを決定する．情報源の確率モデルとして何を使用するかによって，エントロピー $H(X)$ の値は大きく違ってくる．上述の全てのモデルに適用できるエントロピー $H(X)$ の一般式は，教科書の式 (3.26)（p.60）から

$$H(X) = -\lim_{n\to\infty}\frac{1}{n}\sum_{x_1=0}^{1}\sum_{x_2=0}^{1}\cdots\sum_{x_n=0}^{1}P(x_1,x_2,\ldots,x_n)\log_2 P(x_1,x_2,\ldots,x_n) \tag{1}$$

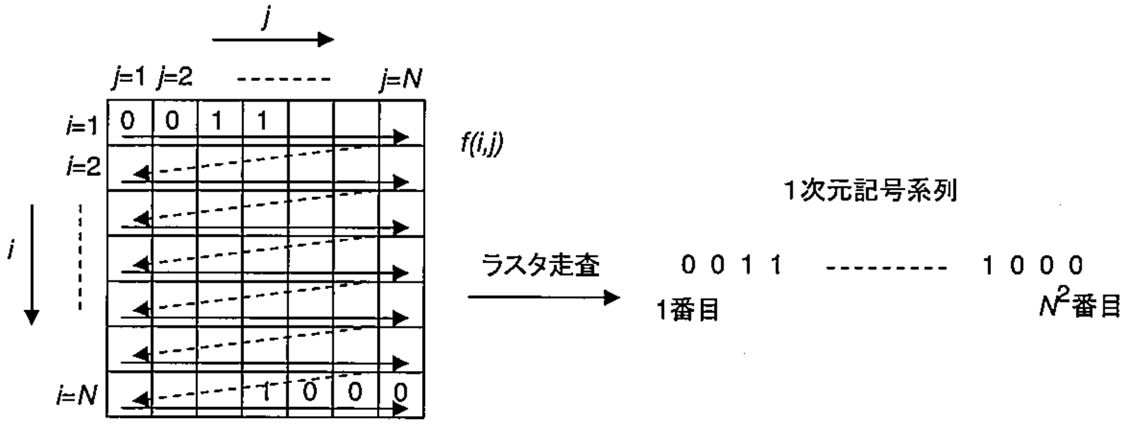


図1　2値画像をラスタ走査により1次元記号系列に変換

image1

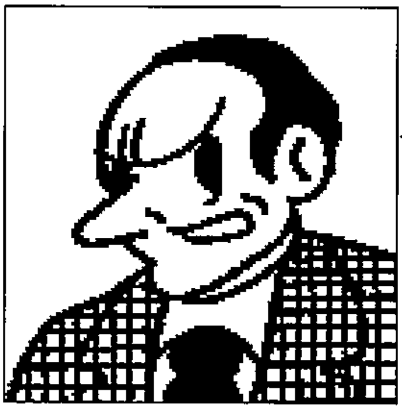


image2

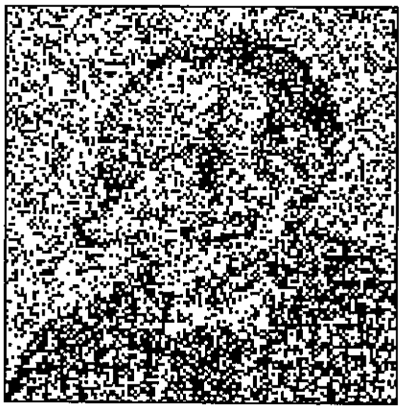


図2　実習に用いる2種類の2値画像

で表される．ただし，$P(x_1,x_2,\ldots,x_n)$ は記号系列 $x_1,x_2,\ldots,x_n$ の結合確率である．式 (1) において n を N^2 より十分小さい自然数で近似すれば，$P(x_1,x_2,\ldots,x_n)$ は与えられた記号系列 $x_1,x_2,\ldots,x_{N^2}$ から計算できるので，式 (1) は全く使えない式ではない．しかし，$x_1,x_2,\ldots,x_n$ が取り得る 0・1 の全ての組み合わせについて確率を計算する必要があり，式 (1) からエントロピー $H(X)$ を求めようと思うと計算は複雑になる．

そこで，教科書の p.60 で説明されている定常確率を用いた方法を利用して，本実習ではもっと上手く簡単に計算する．計算に使用する式は

$$H(X)=\sum_{i=0}^{I-1}P(S_i)H_{S_i}(X) \tag{2}$$

$$H_{S_i}(X)=-\sum_{k=0}^{1}P(k\mid S_i)\log_2 P(k\mid S_i)\ (i=0,1,\ldots,I-1) \tag{3}$$

の2つからなる．ただし，$S_i\ (i=0,1,\ldots,I-1)$ はマルコフ情報源の状態，k は出力記号，$H_{S_i}(X)$ は各状態 S_i に条件付けたエントロピー，$P(S_i)$ は各状態 S_i の定常確率を表す．I はマルコフ情報源の状態数である．式 (2),(3) は教科書の式 (3.24) と同じものであり，意味が理解できない人は教科書のこの部分を再読すると良い．以降では，情報源を無記憶情報源・単純マルコフ情報源・2次のマルコフ情報源とみなす各々の場合について，具体的に式 (2),(3) からエントロピーを計算する方法を説明する．

（1）無記憶情報源の場合

無記憶情報源の場合は状態を考える必要がないので，エントロピー $H(X)$ は単純に1記号だけを考えたエントロピー

$$H(X)=-\sum_{k=0}^{1}P(k)\log_2 P(k) \tag{4}$$

を計算すれば良い．式 (4) を計算するには $P(0),P(1)$ が必要であるが，これは出力記号の確率であるから，$M(0),M(1)$ を与えられた記号系列 $x_1,x_2,\ldots,x_{N^2}$ に含まれる 0・1 の個数として，

$$P(0)=\frac{M(0)}{M(0)+M(1)},\ P(1)=\frac{M(1)}{M(0)+M(1)} \tag{5}$$

により計算できる．必要な確率は $P(0),P(1)$ の2つのみである．

（2）単純マルコフ情報源の場合

単純マルコフ情報源の場合は，(1つ前の出力記号が) 0・1の2つの状態が存在するので，これらを $S_0=(0),S_1=(1)$ で表す ($I=2$)．エントロピー $H(X)$ は式 (2),(3) から計算できるが，その計算には定常確率 $P(S_0),P(S_1)$ と4つの条件付き確率 $P(0\mid S_0),P(1\mid S_0),P(0\mid S_1),P(1\mid S_1)$ が必要になる．これらの確率は，$M(0),M(1)$ を与えられた記号系列 $x_1,x_2,\ldots,x_{N^2}$ に含まれる 0・1 の個数，$M(0\mid 0)$ を 0 の次に 0 が発生した回数，$M(1\mid 0)$ を 0 の次に 1 が発生した回数，$M(0\mid 1)$ を 1 の次に 0 が発生した回数，$M(1\mid 1)$ を 1 の次に 1 が発生した回数として，

$$P(S_0)=\frac{M(0)}{M(0)+M(1)},\ P(S_1)=\frac{M(1)}{M(0)+M(1)} \tag{6}$$

$$P(0 \mid S_0) = \frac{M(0 \mid 0)}{M(0 \mid 0) + M(1 \mid 0)},\ P(1 \mid S_0) = \frac{M(1 \mid 0)}{M(0 \mid 0) + M(1 \mid 0)} \tag{7}$$

$$P(0 \mid S_1) = \frac{M(0 \mid 1)}{M(0 \mid 1) + M(1 \mid 1)},\ P(1 \mid S_1) = \frac{M(1 \mid 1)}{M(0 \mid 1) + M(1 \mid 1)} \tag{8}$$

から求める．即ち，具体的なエントロピー $H(X)$ の計算手順は，

[STEP 1] 与えられた２値画像の記号系列 $x_1, x_2, \ldots, x_{N^2}$ から，個数 $M(0), M(1), M(0 \mid 0), M(1 \mid 0), M(0 \mid 1), M(1 \mid 1)$ をカウントする．

[STEP 2] 式 (6),(7),(8) から必要な確率 $P(S_0), P(S_1), P(0 \mid S_0), P(1 \mid S_0), P(0 \mid S_1), P(1 \mid S_1)$ を求める．

[STEP 3] 式 (2),(3) からエントロピー $H(X)$ を求める．

とまとめられる．必要な確率は合計６個である．

（３） ２次のマルコフ情報源の場合

２次のマルコフ情報源の場合の計算も単純マルコフ情報源の場合とほぼ同じであるが，状態として $S_0 = (00), S_1 = (01), S_2 = (10), S_3 = (11)$ の４つを考える必要がある（$I = 4$）．ただし，$S = (ij)$ は１つ前の記号が j で２つ前の記号が i の状態を表す（i と j の順序を逆に考えてはいけない）．計算の手順を以下にまとめる．まず，$M(00), M(01), M(10), M(11)$ を記号系列に含まれる００・０１・１０・１１の個数，$M(0 \mid 00)$ を００の次に０が発生した回数，$M(1 \mid 00)$ を００の次に１が発生した回数，$M(0 \mid 01)$ を０１の次に０が発生した回数，$M(1 \mid 01)$ を０１の次に１が発生した回数，$M(0 \mid 10)$ を１０の次に０が発生した回数，$M(1 \mid 10)$ を１０の次に１が発生した回数，$M(0 \mid 11)$ を１１の次に０が発生した回数，$M(1 \mid 11)$ を１１の次に１が発生した回数として，与えられた記号系列 $x_1, x_2, \ldots, x_{N^2}$ からこれらの個数をカウントする．次に，計算に必要な合計 12 個の確率を，

$$P(S_0) = \frac{M(00)}{M(00) + M(01) + M(10) + M(11)},\ P(S_1) = \frac{M(01)}{M(00) + M(01) + M(10) + M(11)} \tag{9}$$

$$P(S_2) = \frac{M(10)}{M(00) + M(01) + M(10) + M(11)},\ P(S_3) = \frac{M(11)}{M(00) + M(01) + M(10) + M(11)} \tag{10}$$

$$P(0 \mid S_0) = \frac{M(0 \mid 00)}{M(0 \mid 00) + M(1 \mid 00)},\ P(1 \mid S_0) = \frac{M(1 \mid 00)}{M(0 \mid 00) + M(1 \mid 00)} \tag{11}$$

$$P(0 \mid S_1) = \frac{M(0 \mid 01)}{M(0 \mid 01) + M(1 \mid 01)},\ P(1 \mid S_1) = \frac{M(1 \mid 01)}{M(0 \mid 01) + M(1 \mid 01)} \tag{12}$$

$$P(0 \mid S_2) = \frac{M(0 \mid 10)}{M(0 \mid 10) + M(1 \mid 10)},\ P(1 \mid S_2) = \frac{M(1 \mid 10)}{M(0 \mid 10) + M(1 \mid 10)} \tag{13}$$

$$P(0 \mid S_3) = \frac{M(0 \mid 11)}{M(0 \mid 11) + M(1 \mid 11)},\ P(1 \mid S_3) = \frac{M(1 \mid 11)}{M(0 \mid 11) + M(1 \mid 11)} \tag{14}$$

から求める．最後に，これらの確率を式 (2),(3) に代入してエントロピー $H(X)$ を求める．

２．３　プログラム作成時の注意

プログラム作成の際には以下に注意する．

（1）$0\log_2 0=0$ に特別の配慮が必要である．プログラム中でこの配慮を行わないと，画像によってはエラーになったり誤った結果が求まったりする．

（2）C言語で使用できる対数関数は，自然対数 log(x) と普通対数 log10(x) のみであり，$\log_2(\cdot)$ の計算には対数関数の底の変換を用いる．

3．実習方法

3．1　課題

図２に示す２種類の２値画像をラスタ走査して０・１の記号系列に変換したファイル『image1』・『image2』が準備してある．漫画をスキャナーで読み込んだものである．ただし，両方の画像とも画素数は 128×128 であり，１次元記号系列に直した長さは $N^2 = 128 \times 128 = 16,384$ である．これらの記号系列から，無記憶情報源・単純マルコフ情報源・２次のマルコフ情報源の３つの場合について，１記号あたりのエントロピー $H(X)$ を計算するプログラムを作成せよ．そして，プログラムを実行して，『image1』・『image2』の両方の場合のエントロピーの値を表にまとめよ．ただし，参考データとして計算に使用した確率の値（無記憶情報源の場合２個，単純マルコフ情報源の場合６個，２次のマルコフ情報源の場合 12 個）も出力し，(レポートでは）状態遷移図の中に書き込んで分かりやすく示せ．次に，講義で学んだ理論に基づいて，何故そのような結果が得られたか，得られた結果が何を意味するかなど，実験結果に対する考察を行え．ただし，『image1』は標準的な０・１の発生パターンを持つ２値画像，『image2』は『image1』においてランダムに選んだ 30 ％の画素の０・１の値を反転させたものでランダム性が大きい画像になっている．ランダム性の違いが結果にどのような違いを及ぼすかとその理由，無記憶情報源・単純マルコフ情報源・２次のマルコフ情報源の３つに場合で生じる結果の違いとその理由，を必ず考察する．

無記憶情報源の場合についてエントロピーを計算するサンプルプログラムを，本資料の最終ページに示す．自信がない人は，このプログラムを手直しして使用しても良い．

3．2　レポートについて

説明を加筆したプログラム，実験結果，実験結果に対する考察をまとめてホチキスで閉じ，筆記試験の時間に回収するので持参すること．但し，レポートに原理の項目を書く必要はない．

4．必要ファイルの取得方法

本実習で使用するファイル『image1』と『image2』はＷＥＢサイト

http://www.cs.tsukuba.ac.jp/~kudo/japanese.html

からダウンロードできる．ユーザ認証のＩＤとパスワードは授業時間中に通知する．各自でダウンロードして使用すること．

```c
#include <stdio.h>
#include <stdlib.h>
#include <math.h>
#define N 128

void read_file(unsigned char image[], char name[]);


int main(void) {
  int i;
  unsigned char image[N*N]; //画像の１次元データ
  char name[10]="image1"; //画像データの名前
  int M0=0, M1=0;
  double P[2];
  double H;

  read_file(image, name);

  for(i=0; i<N*N; i++) {
    if(image[i] == 0)
      M0++;
    else
      M1++;
  }
  P[0] = (double)M0/(M0+M1);
  P[1] = (double)M1/(M0+M1);
  printf("P0 = %f,  P1 = %f¥n", P[0], P[1]);

  for(i=0; i<2; i++)
    if(P[i]==0)
      P[i]=1;

  H = (- P[0]*log(P[0]) - P[1]*log(P[1])) / log(2);
  printf("H  = %f¥n", H);

  return 0;
}

void read_file(unsigned char image[], char name[]) {
  FILE *fp;
  char c[N*N+1];
  int i;

  if((fp = fopen(name, "r")) == NULL) {
    printf("error¥n");
    exit(1);
  }

  fgets(c, N*N+1, fp);
  for(i=0; i<N*N; i++)
    image[i] = c[i] - 48;
}
```