「符号の木」と瞬時復号可能符号
およびハフマン符号化

瞬時復号可能な符号かどうかを判定するのに「**符号の木**」を応用するのが便利であるのでその方法を紹介する。
「符号の木」はまた**ハフマン符号化**にも応用される。

1.　**瞬時復号可能な符号**とは、符号語系列を受信した際、符号語の切れ目が次ぎの符号語の先頭部分を受信しなくても分り、次の符号語を受信する前にその符号語を正しく復号できることをいう。
次のように、a b c d　4文字について、3 種類の符号化の方法があるとして、 0101101110という符号語系列を受信したとき瞬時に復号可能かどうかを調べてみよう。

①	a = 00	b = 01	c = 10	d = 11
②	a = 0	b = 10	c = 110	d = 111
③	a = 0	b = 01	c = 011	d = 111

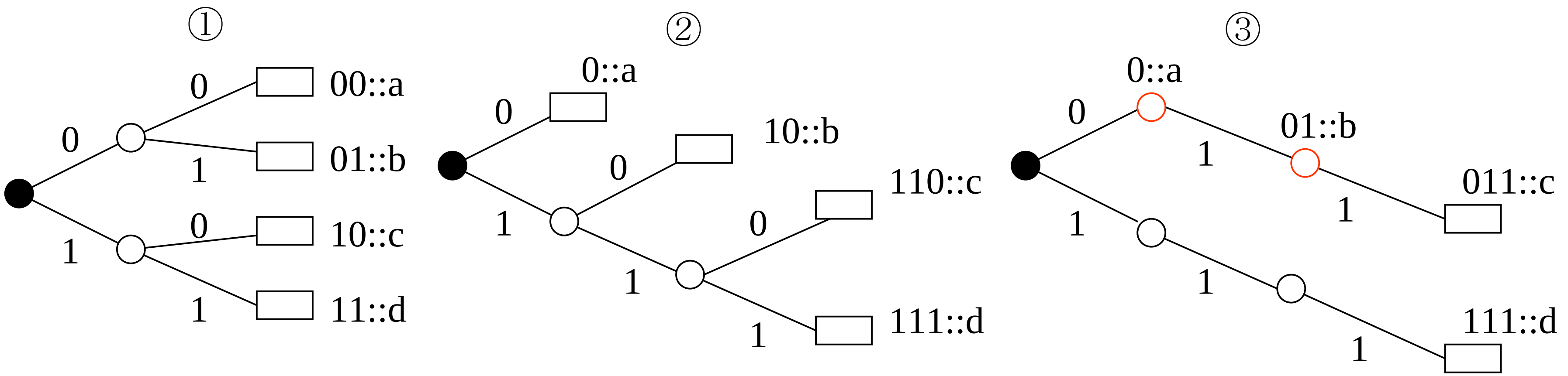
下表に示すように、①②では、1語の区切りが明瞭で次の語を読まなくても直ちに復号可能だが、③では次の語やその次の語まで見ないと復号化が確定できない。

①	符号語系列0101101110	01	01	10	11	10	2bit区切り
	復号化	b	b	c	d	c	
②	符号語系列0101101110	0	10	110	111	0	0で終わるか3ビットで終わり
	復号化	a	b	c	d	a	
③	符号語系列0101101110	0	10				語の区切りが不明で次の語まで読んで初めて確定する
	復号化 試行1	a?	?				
	符号語系列0101101110	01	01	10			
	復号化 試行2	b	b?	?			
	符号語系列0101101110	01	011	0	111	0	
	復号化 試行3	b	c	a	d	a	

前頁の符号化表を再掲する。

①	a = 00	b = 01	c = 10	d = 11
②	a = 0	b = 10	c = 110	d = 111
③	a = 0	b = 01	c = 011	d = 111

上表を次の要領で図化してみる（このような図を「**符号の木**」と呼ぶ）。
1．図の出発点を●で表し、これを**根**と名づける。
2．根の他に○で示す**節**、あるいは□で示す**葉**が存在する。
3．根と節、節と節、節と葉を結ぶ線を**枝**と名づけ、**枝**には０または１を対応させる。
4．**葉**は終端でありその先に**枝**は出ない。
5．根から符号に至る枝の０又は1をその順に連結して**語**の位置に**符号**として表示する。



この3 図を比較すると、①②は**葉で終端**しているのに対し、③はa b のように、**節に対応する符号**が存在することが分る。そして、その節の先に別の符号があり語の区切りが直ちには判定できないことが分る。**→符号の木を描き、すべて葉で終端していれば瞬時復号可能符号**

２．　次に**瞬時復号可能な符号**のうち、符号化すべき語の発生確率の差を利用して、発生確率の高い語には短い符号、発生確率の低い語には長い符号を割り当て、送受信する符号の全長が極力短くなるような符号化方法（データ圧縮技術の1種）を考える。

発生確率A
a: 0.50
b: 0.30
c: 0.12
d: 0.08

①	a = 00	b = 01	c = 10	d = 11
②	a = 0	b = 10	c = 110	d = 111
④	a = 000	b = 001	c = 01	d = 1

発生確率B
a: 0.35
b: 0.25
c: 0.22
d: 0.18

a,b,c,d の出現確率が左上A表又は右上B表の値であるとき、上表のような符号化を行なうと、1 語当りの平均符号長（ビット数）がどうなるかを求めてみる。下表に示すようにAでは②の場合が最短で1.700となることが分る。Bの場合は、①が最短2.000となる。④は発生確率の高い語の符号が長く、当然ながら最長となった。このように、発生確率によって、最短となる符号化は異なるものとなる。

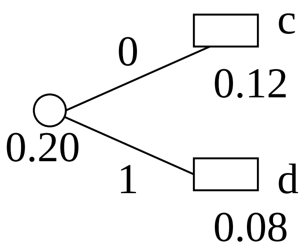
語	発生確率A	符号長			平均符号長の計算			発生確率B	符号長			平均符号長の計算		
		①	②	④	①	②	④		①	②	④	①	②	④
a	0.50	2	1	3	1.000	0.500	1.500	0.35	2	1	3	0.700	0.350	1.050
b	0.30	2	2	3	0.600	0.600	0.900	0.25	2	2	3	0.500	0.500	0.750
c	0.12	2	3	2	0.240	0.360	0.240	0.22	2	3	2	0.440	0.660	0.440
d	0.08	2	3	1	0.160	0.240	0.080	0.18	2	3	1	0.360	0.540	0.180
計	1.00				2.000	1.700	2.720	1.00				2.000	2.050	2.420

前頁の発生確率表Aを再掲する。これを元に符号の木を作り、瞬時復号可能で、かつ、平均符号長が短くなる符号化方法の一つである**ハフマン符号化法**を実行してみる。

発生確率A
a: 0.50
b: 0.30
c: 0.12
d: 0.08

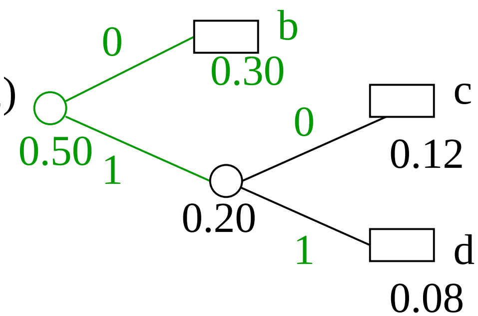
(1)　まず、使用確率（発生確率）の最も低い二つの語（c,d）を選び、1個の節と二つの葉を設け、節と葉をつなぐ枝を描く。節の近傍に確率の合計値を記入する。確率の高い方の枝に0 低い方の枝に1 を割り当てる。

(1)



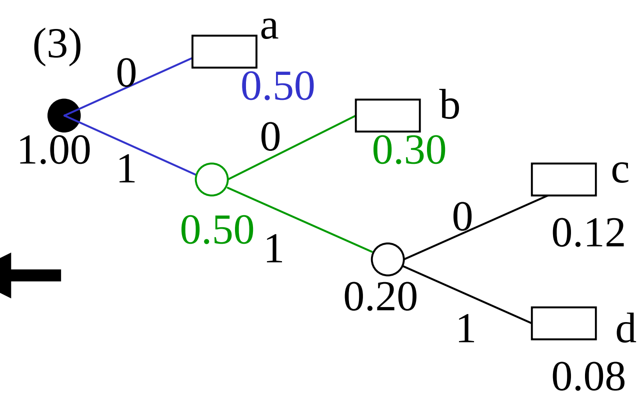
(2)　次に、残る語と(1)で集約したもののうち確率の最も低い二つの語（b 0.30と (c d)0.20）を括る。確率の高い b に 0 、低い cd に1 が来るように新たに節と葉を設け枝でつなぐ。節の近傍に確率の合計値を記入する。

(2)



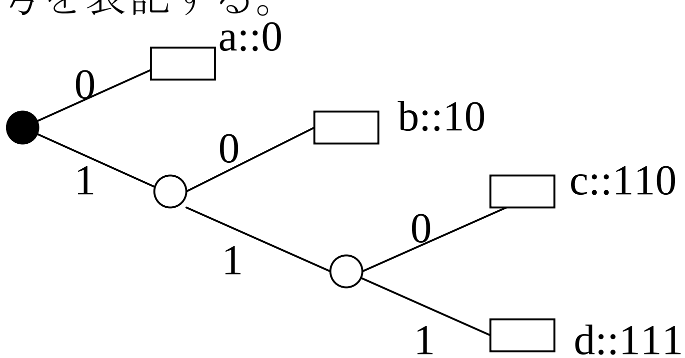
(3)　(2)の結果(b(cd)0.50)と残り(a 0.50) を括る。最後にできた節を**根**とする。
確率は0.5で等しいので枝の0と1の付与方法は任意である。

(3)



(4)　「符号の木」の形ができたら、根側から枝をたどり符号を表記する。

平均符号長	
a: 0.50*1=0.50	
b: 0.30*2=0.60	
c: 0.12*3=0.36	
d: 0.08*3=0.24	
計	1.70

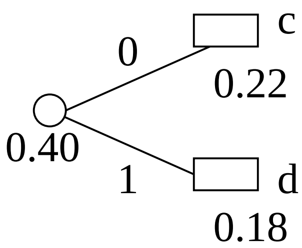


応用問題１． 前例の使用確率表を変更しBを取り上げる。確率最小の語の組合せが前問と異なり木の形が変わる。

発生確率B
a: 0.35
b: 0.25
c: 0.22
d: 0.18

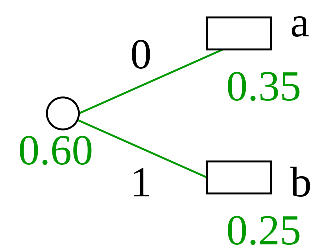
(1)　まず、使用確率（発生確率）の最も低い二つの語（c,d）を選び、1個の節と二つの葉を設け、節とつなぐ枝を描く。節の近傍に確率の合計値を記入する。確率の高い方に0、低い方に1 を割り当てる。

(1)



(2)　次に、(1)の集約値と残る語のうち使用確率（0.40, 0.25, 0.35）の最も低い二つの語（a 0.25と b0.35）を括る。確率の高い方に０、低い方に1 が来るように新たに節と葉を設け枝でつなぐ。
節の近傍に確率の合計値を記入する。

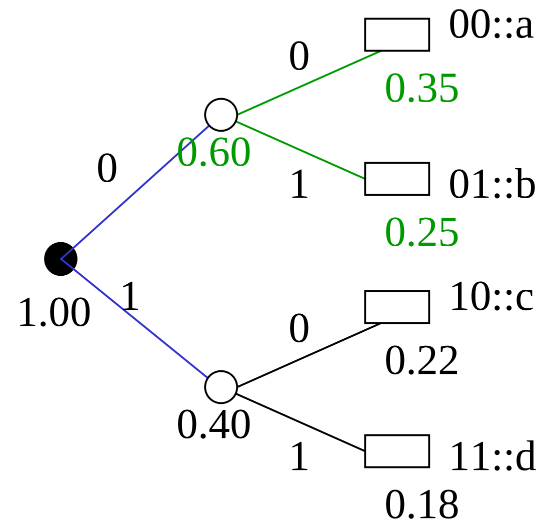
(2)



(3)　(1)の結果と(2)の結果とを括る。最後にできた節を**根**とする。

(4)　「符号の木」の形ができたら、根側から枝をたどり符号を表記する。

(3)



符号	
a	00
b	01
c	10
d	11

すべて2ビットなので、平均符号長は2.000 となる。

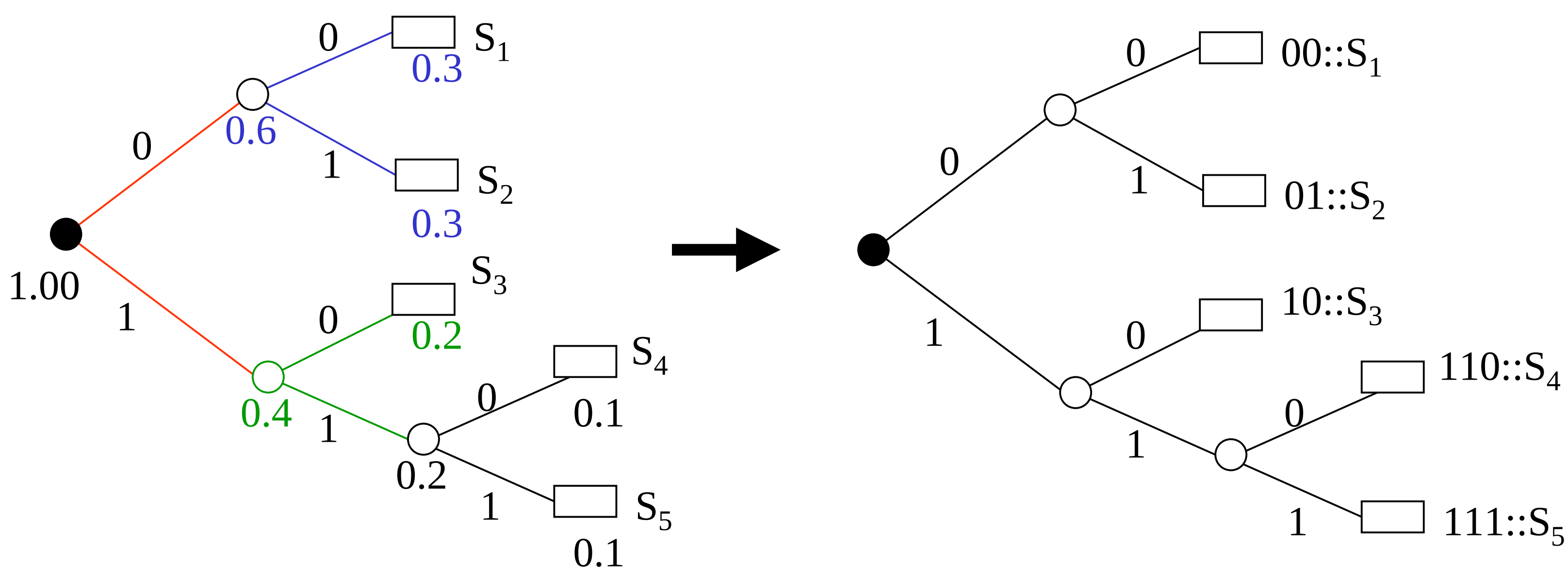
応用問題２　電気電子H20

下左表の発生確率の情報源シンボルをハフマン符号によって二元符号化を行ったときの平均符号長は？

情報源シンボル	発生確率
S_1	0.3
S_2	0.3
S_3	0.2
S_4	0.1
S_5	0.1

ハフマン符号化後

ハフマン符号	平均符号長計算
00	2*0.3 = 0.6
01	2*0.3 = 0.6
10	2*0.2 = 0.4
110	3*0.1 = 0.3
111	3*0.1 = 0.3
	計　2.2

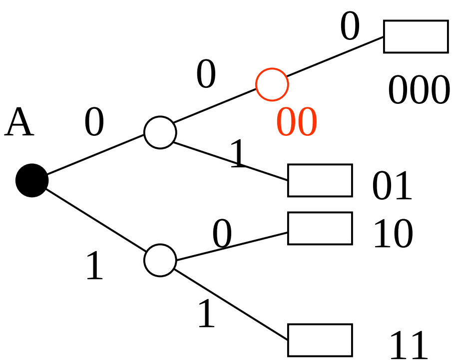


応用問題３　電気電子H21

次のうち下表の符号で瞬時復号可能な符号の組合せで正しいのは、
①BCDE ②BCE ③AD ④BE ⑤CD のどれか。

符号A	符号B	符号C	符号D	符号E
000	1	0	1	0
11	10	10	01	01
10	110	110	001	011
01	1110	1110	0001	0111
00	11110	11110	0000	01111

「符号の木」を描けば下記の通りCDだけがすべての符号が葉で終端しているのに対しABEは節で終わるものがある。これまでの学習から、観察によっても判断できる。瞬時復号不能な場合は、ある符号が別の符号の終端のひとつ手前までの経路の一部分になっている場合で、左の表のA(00 と000), B(1と10,110,,,), E(0と01,01と011,,,,)が挙げられる。
Cは0で区切られ、Dは1で区切られるか又は4桁の0で区切られ瞬時復号化可能。
（⑤が正解）



B