信息与编码
总结与复习

（2007.12.25.）

《信息与编码》的主要内容

一个理论和三个编码：

理论--------香农信息论

编码--------信源编码

信道编码

保密编码

第一部分、信息论基础

1.1 信源的信息理论：

1、信息的定义：

（1）自信息 $I = \log(1/p) = -\log p$

（2）信息量=通信所消除掉的不确定度

=通信前的不确定度-通信后的不确定度

（3）信息的单位：对数的底取2时，自信息的单位叫比特(*bit*)。

2、信息熵的定义：

（1）离散信源

$$H(X)=\sum_{i=1}^{m}p_iI_i=\sum_{i=1}^{m}p_i\log\frac{1}{p_i}=-\sum_{i=1}^{m}p_i\log p_i$$

（2）连续信源

$$h(x)=-\int_{-\infty}^{\infty}p(x)\log\,p(x)dx$$

3、信息熵的特点

（1）非负性：$H(X) \geqslant 0$

（2）对称性：$H(p_1 p_2 \cdots\cdots) = H(p_2 p_1 \cdots\cdots)$

（3）极值性：

《1》离散信源各符号等概率时出现极大值：

$$H_0 = \log m$$

《2》连续信源信号幅度受限时均匀分布出现极大值：$h_{max}(X) = \log (b-a)$；

《3》连续信源信号方差有限时高斯分布出现极大值：$h_{\max}(X) = \frac{1}{2}\log(2\pi e\sigma^2)$

4、离散序列的信息熵

（1）无记忆信源的联合熵与单符号熵：

$$H(X_1X_2\cdots\cdots X_N)$$

$$= H(X_1)+H(X_2)+H(X_3)+\cdots\cdots+H(X_N) = NH(X_1)$$

（2）有记忆信源的联合熵与条件熵：

$$H(X_1X_2\cdots\cdots X_N)=H(X_1) + H(X_2|X_1)$$

$$+ H(X_3|X_1X_2) + \cdots\cdots+H(X_N|X_1X_2\cdots\cdots X_{N-1})$$

（3）平均符号熵：

$$H_N =H(X_1X_2\cdots\cdots X_N) / N$$

（4）序列信息熵的性质：

《1》条件熵不大于无条件熵，强条件熵不大于弱条件熵：$H(X_1) \geqslant H(X_2|X_1) \geqslant H(X_3|X_1X_2) \geqslant \cdots$

$\cdots\cdots \geqslant H(X_N|X_1X_2\cdots\cdots X_{N-1})$

《2》条件熵不大于同阶的平均符号熵：

$$H_N \geqslant H(X_N|X_1X_2\cdots\cdots X_{N-1})$$

《3》序列越长，平均每个符号的信息熵就越小：

$$H_1 \geqslant H_2 \geqslant H_3 \geqslant \cdots\cdots \geqslant H_N$$

总之：$H_0 > H_1 \geqslant H_2 \geqslant H_3 \geqslant \cdots\cdots \geqslant H_N \geqslant H_\infty$

（无记忆可作为有记忆的特例统一处理。）

5、马尔科夫信源的信息熵

（1）马尔科夫信源的数学模型和定义：

*N*阶马尔科夫信源的关联长度是*N*+1，*N*+2以外不关联。

（2）状态、状态转移与稳态概率：

状态、状态转移、状态转移图、稳定状态、稳态方程

$$p(a_k)=\sum_{i=1} Q(E_i)p(a_k\mid E_i)$$

（3）稳态符号概率：

（4）稳态信息熵：

$$H=\sum_{i=1}^{L}\sum_{k=1}^{m}Q(E_i)H(a_k\mid E_i)=-\sum_{i=1}^{L}Q(E_i)\sum_{k=1}^{m}p(a_k\mid E_i)\log p(a_k\mid E_i)$$

结论：*N*阶马氏信源稳态信息熵(即极限熵)等于*N*+1阶条件熵。

$$H_\infty=\lim_{N\to\infty}\frac{1}{N}H(X_1X_2\cdots X_{N-1}X_N)=H(X_{N+1}\mid X_1X_2\cdots X_N)$$

[例1] 已知二阶马尔科夫信源的条件概率：

$p(0|00)=p(1|11)=0.8$；$p(0|01)=p(1|10)=0.6$；

求稳态概率、稳态符号概率和稳态信息熵。

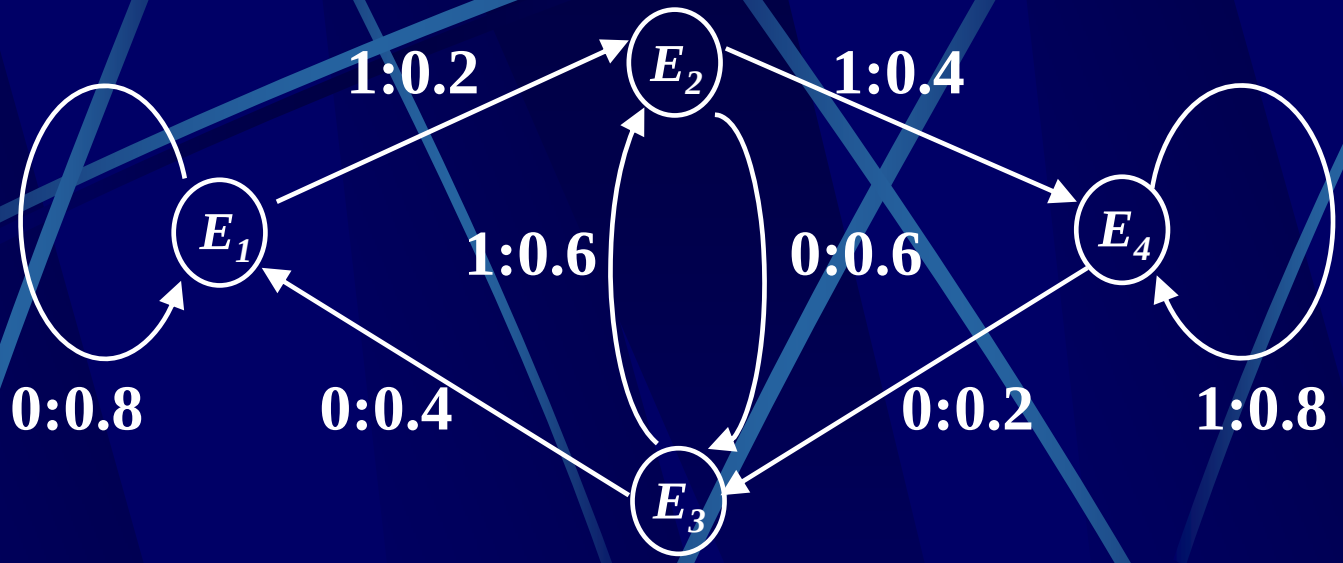
解：二阶马氏信源关联长度=3，状态由2符号组成，共有4个状态，分别为：$E_1=00$；$E_2=01$；$E_3=10$；$E_4=11$；

已知的条件概率即是：

$p(0|E_1)=p(1|E_4)=0.8$；$p(0|E_2)=p(1|E_3)=0.6$；

根据归一化条件可求出另外4个状态符号依赖关系为：

$p(1|E_1)=p(0|E_4)=0.2$；$p(1|E_2)=p(0|E_3)=0.4$；



稳 态 方 程 组 是:

$$\begin{cases} Q(E_1) = 0.8Q(E_1) + 0.4Q(E_3) \\ Q(E_2) = 0.2Q(E_1) + 0.6Q(E_3) \\ Q(E_3) = 0.6Q(E_2) + 0.2Q(E_4) \\ Q(E_4) = 0.4Q(E_2) + 0.8Q(E_4) \\ Q(E_1) + Q(E_2) + Q(E_3) + Q(E_4) = 1 \end{cases}$$

可　解得：

$$\begin{cases} Q(E_1)=Q(E_4)=\dfrac{1}{3} \\ Q(E_2)=Q(E_3)=\dfrac{1}{6} \end{cases}$$

稳态符号概率为：

$$\begin{cases} p(0)=\sum_{i=1}^{4}Q(E_i)p(0\mid E_i)=\dfrac{1}{3}\times 0.8+\dfrac{1}{6}\times 0.6+\dfrac{1}{6}\times 0.4+\dfrac{1}{3}\times 0.2=\dfrac{1}{2} \\ p(1)=\sum_{i=1}^{4}Q(E_i)p(1\mid E_i)=\dfrac{1}{3}\times 0.2+\dfrac{1}{6}\times 0.4+\dfrac{1}{6}\times 0.6+\dfrac{1}{3}\times 0.8=\dfrac{1}{2} \end{cases}$$

稳态信息熵为：

$$H=-\frac{1}{3}\times[0.8\log 0.8+0.2\log 0.2]\times 2-\frac{1}{6}\times[0.6\log 0.6+0.4\log 0.4]\times 2=$$

$=0.895$ **bit**/符号

1.2 信道的信息理论：

1、信道的数学模型：

进入广义信道的符号为$a_i \in A$；从广义信道出来的符号$b_j \in B$；其前向概率为 $p_{ij}=p(b_j|a_i)$。

传输矩阵：

$$P=\begin{pmatrix} p_{11} & p_{12} & \cdots\cdots & p_{1s} \\ p_{21} & p_{22} & \cdots\cdots & p_{2s} \\ \cdots & \cdots & \cdots\cdots & \cdots \\ p_{m1} & p_{m2} & \cdots\cdots & p_{ms} \end{pmatrix}$$

2、信道的分类：

（1）无噪无损信道：a_i与b_j是一一对应的，$p(b_j|a_i)=\delta_{ij}$，传输矩阵为单位方阵。

（2）有噪有损信道：a_i与b_j多-多对应的，传输矩阵中所有的矩阵元都有可能不为零。特例是BSC信道BEC信道。

（3）有噪无损信道分组一对多，传输矩阵应具有一行多列的分块对角化形式。

（4）无噪有损信道：分组多对一，其传输矩阵应具有多行一列的分块对角化形式。

（5）对称信道：传输矩阵的各行都是一些相同元素的重排，各列也是一些相同元素的重排。

3、信道有关的信息熵：

（1）信源熵 (先验熵)：$H(X)=-\sum_{i=1}^{m}p(a_{i)}\log p(a_i)$

（2）噪声熵 (散布度)：

$$H(Y\mid X)=-\sum_{i=1}^{m}\sum_{j=1}^{s}p(a_ib_j)\log p(b_j\mid a_i)$$

（3）联合熵：$H(XY)=-\sum_{i=1}^{m}\sum_{j=1}^{s}p(a_ib_j)\log p(a_ib_j)$

（4）接收符号熵：$H(Y)=-\sum_{i=1}^{m}p(b_{j)}\log p(b_j)$

（5）损失熵(后验熵)：

$$H(X\mid Y)=-\sum_{i=1}^{m}\sum_{j=1}^{s}p(a_ib_j)\log p(a_i\mid b_j)$$

4. 平均互信息

（1）平均互信息的定义：

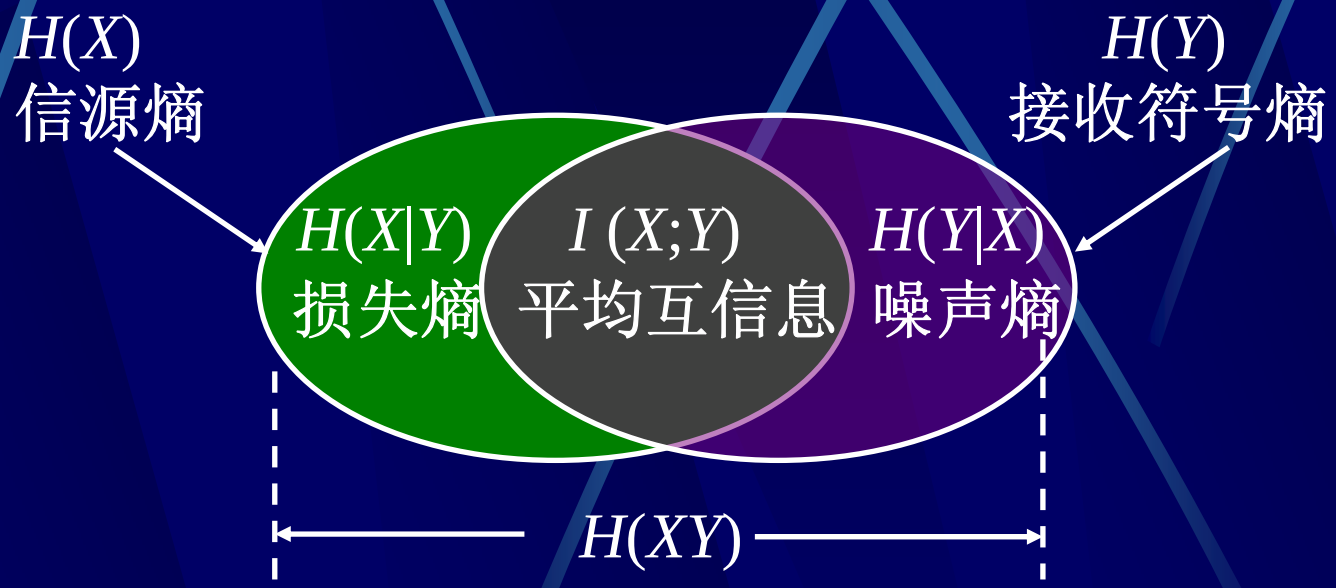
◆系统完成一个符号从发到收的通信过程后，关于符号a_i的不确定度的变化为：

$$I(a_i;b_j)=[-\log p(a_i)]-[-\log p(a_i\mid b_j)]=\log\frac{p(a_i\mid b_j)}{p(a_i)}$$

◆统计平均而言，平均每收发一对符号信宿所获得的信息量为：

$$I(X;Y)=E[I(a_i;b_j)]=\sum_{i=1}^{m}\sum_{j=1}^{s}p(a_ib_j)\log\frac{p(a_i\mid b_j)}{p(a_i)}$$

（2）平均互信息与信息熵之间的关系：



计算公式：

$$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) = H(X) + H(Y) - H(XY)$$

[例2]已知信源先验概率$p(x)=\{0.7, 0.3\}$，信道传输矩阵$P=\begin{pmatrix}0.3 & 0.2 & 0.5\\ 0.4 & 0.3 & 0.3\end{pmatrix}$；试计算各信息熵和互信息。

解:（1）先验熵：　$H(X)= -0.7\log_2 0.7 -0.3\log_2 0.3$

$= (-0.7\lg 0.7 -0.3\lg 0.3)/\lg 2 = 0.881$ bit/符号

（2)联合熵：

$$P(XY)=\begin{pmatrix}0.7\times0.3 & 0.7\times0.2 & 0.7\times0.5\\ 0.3\times0.4 & 0.3\times0.3 & 0.3\times0.3\end{pmatrix}=\begin{pmatrix}0.21 & 0.14 & 0.35\\ 0.12 & 0.09 & 0.09\end{pmatrix}$$

$H(XY)= -0.21\log 0.21 -0.14\log 0.14 -0.35\log 0.35$

$-0.12\log 0.12 -0.09\log 0.09 -0.09\log 0.09$

$=2.3924$ bit/符号

(3)噪声熵：

由 $P=\begin{pmatrix}0.3 & 0.2 & 0.5\\ 0.4 & 0.3 & 0.3\end{pmatrix}$ 和 $P(XY)=\begin{pmatrix}0.21 & 0.14 & 0.35\\ 0.12 & 0.09 & 0.09\end{pmatrix}$

$$H(Y\mid X)= -0.21\log 0.3 -0.14\log 0.2 -0.35\log 0.5$$
$$-0.12\log 0.4 -0.09\log 0.3 -0.09\log 0.3$$
$$= 1.5114 \text{ bit/符号}$$

(4)接收符号熵： 由 $p(y_j)=\sum_{i=1}^{m} p(x_i y_j)$

$$P(Y)=（0.21+0.12，\ 0.14+0.09，\ 0.35+0.09）$$
$$=（0.33,\ 0.23,\ 0.44）$$
$$H(Y)= -0.33\log 0.33 -0.23\log 0.23 -0.44\log 0.44$$
$$=1.5366 \text{ bit/符号}$$

(5)损失熵：

$$p(x_i \mid y_j) = \frac{p(a_i b_j)}{p(b_j)} \qquad P(XY) = \begin{pmatrix} 0.21 & 0.14 & 0.35 \\ 0.12 & 0.09 & 0.09 \end{pmatrix}$$

$$P(X \mid Y) = \begin{pmatrix} \frac{0.21}{0.33} & \frac{0.14}{0.23} & \frac{0.35}{0.44} \\ \frac{0.12}{0.33} & \frac{0.09}{0.23} & \frac{0.09}{0.44} \end{pmatrix} = \begin{pmatrix} \frac{7}{11} & \frac{14}{23} & \frac{35}{44} \\ \frac{4}{11} & \frac{9}{23} & \frac{9}{44} \end{pmatrix}$$

$H(X|Y)$= -0.21log(7/11) - ……0.09log(9/44)=0.8558 bit/符号

或：$H(X|Y)=H(XY)-H(Y)$=2.3924-1.5266=0.8558 bit/符号

(6)平均互信息：

$I(X;Y)=H(X)-H(X|Y)$= 0.881 –0.8558=0.0252 bit/符号

5. 信道容量

（1）传码率与传信率：

单位时间信道传输的码元数量叫做传码率，用R_B（每秒符号数）表示，也叫波特率。

单位时间信道传输的净信息量值叫做传信率，用R_b（每秒信息量）表示，也叫比特率。显然：

$$R_b = I(X;Y)R_B$$

（2）信道容量的定义：

对于给定的信道，既然总存在一个信源能使互信息取极大值，就可以把这个极大值定义为该信道的信道容量：　$C = \underset{\{p(x)\}}{Max} I(X;Y)$

有时，也把单位时间的最大传信率定义为信道容量，记做$C_t = C\ R_B$；

信道容量反映了一个信道最大所能传输的平均互信息，是给定信道的属性。

（3）信道容量的计算：

对于简单信道要求能计算信道容量：

1）无损信道：$C = \max\{I(X;Y)\} = \max\{H(X)\} = \log m$；

2）无噪信道：$C = \max\{I(X;Y)\} = \max\{H(Y)\} = \log S$；

3）对称信道：$C = \max\{I(X;Y)\} = \log S - H(p1, p2, \ldots\ldots, pS)$；

[例3]求对称信道 $P = \begin{pmatrix} 0.2 & 0.3 & 0.2 & 0.3 \\ 0.3 & 0.2 & 0.3 & 0.2 \end{pmatrix}$ 的信道容量。

解：$C = \log 4 - H(0.2, 0.3, 0.2, 0.3)$

$= 2 + (0.2\log 0.2 + 0.3\log 0.3) \times 2 = 0.03$ bit/符号；

（4）波形信道的信道容量：

- 发送连续信号的信源是连续信源，传输连续信号的信道是波形信道。
- 高斯加性噪声波形信道的容量由Shannon公式给出：

$$C = B\log(1+P_X/P_n)$$

- 香农公式给出了带宽B、信道容量和信噪比P_X/P_n三者之间的制约关系。
- 信道容量不变 时带宽与信噪比有互换关系。

- 香农公式在C=常数条件下给出的是出了带宽B与信噪比P_X/P_n的等效搭配关系：较大的带宽搭配较小的信噪比与较大的带宽搭配较小的信噪比均能得到同样的信道容量，达到相同的通信效果。
- 决不要以为这个公式给出的是带宽B与信噪比P_X/P_n之间的因果关系：误认为当带宽较大时信噪就会比较小，尤其是不能得出“较大的带宽得到了较小的输出信噪比”这样的错误结论。

[例4]某图片含**2.25×10^6**个像素，采用**12**级量化电平传输。假定各电平等概出现，信道中信噪比为**30dB**，若要求**3**分钟完成传输，需要多大的带宽？

解：

传信率：$R=\dfrac{2.25\times10^6\times\log12}{3\times60}=4.48\times10^4\ bit/s\le C$

信噪比：$10\log(P_x/P_n)=30\text{dB}$；即$P_x/P_n=10^3$

根据香农公式：

$$B=\frac{C}{\log_2(1+P_x/P_n)}=\frac{4.48\times10^4}{\log_2 1001}=4.49kHz$$

第二部分、无失真信源编码

1.1 信源编码理论：

1、信源的相对信息率和冗余度：

（1）实际信源由于非等概，使$H(X)<H_0=\log m$

（2）实际信源由于有记忆，使$H_\infty< H_N< H(X)$

（3）信源每个符号最大可以荷载的信息量是 H_0

（4）平均每个符号的实际信息荷载量是H_∞

（5）信源普遍存在着信息含量不饱满的现象。

- 定义相对信息率：　$\mu = H_\infty / H_0$
- 信源冗余度（或剩余度）：　$\gamma = 1 - \mu$
- 怎样将这些冗余压缩掉？
- 寻找一种更短的代码序列，在不损失信息的前提下，替代原来的符号序列。
- 应当尽量使所找的编码序列各个码元相互独立且等概，就会使单位符号信息含量更多，代码就比原来更短。

2、变长码编码原理：

（1）概率匹配原则：信息量大(不常出现)的符号用长码，信息量小(经常出现)的符号用短码。

$$l_i = \log_r \frac{1}{p_i} = I_r(a_i)$$

（2）平均码长：$\bar{l} = \sum_{i=1}^{m} p_i l_i$

（3）**Shannon**变长码编码定理：$\frac{H_N}{\log r} \le \bar{l} \le \frac{H_N}{\log r} + \frac{1}{N}$

（4）极限码长：$\bar{l} \to (H_\infty / \log r)$;

（5）编码效率：$\eta = \frac{H_\infty}{\bar{l} \log r}$

3、唯一可译性与即时性：

（1）断字问题：分组编码的变长码被连接起来发送，接收端如何才能将它们分开进行译码呢？

（2）唯一可译码的判断

（3）即时码的判断

（4）构造码树的方法

（5）克拉夫特不等式：唯一可译码的必要条件。

[例5] 以下哪些编码一定不是惟一可译码？写出每种编码克拉夫特不等式的计算结果。

- 码A：0，10，11，101；
- 码B：1，01，001，0001；
- 码C：0，10，110，1110，111110；

解：码A：1/2+1/4+1/4+1/8>1，不能唯一可译；

码B：1/2+1/4+1/8+1/16<1,有可能唯一可译；

码C：1/2+1/4+1/8+1/16+1/64<1，有可能唯一可译；

1.2 编码方法：

1、Huffman编码：

（1）信源符号按概率大小排队。

（2）合并概率最小的两个符合为一个节点。

（3）节点参与排队放在与自己概率相等符号前面。

（4）重复这个过程直到合并完全部符号。

（5）标记每个分支的的0与1。

（6）从根到叶的路径就给出了相应符号的码字。

（7）计算平均码长与编码效率。

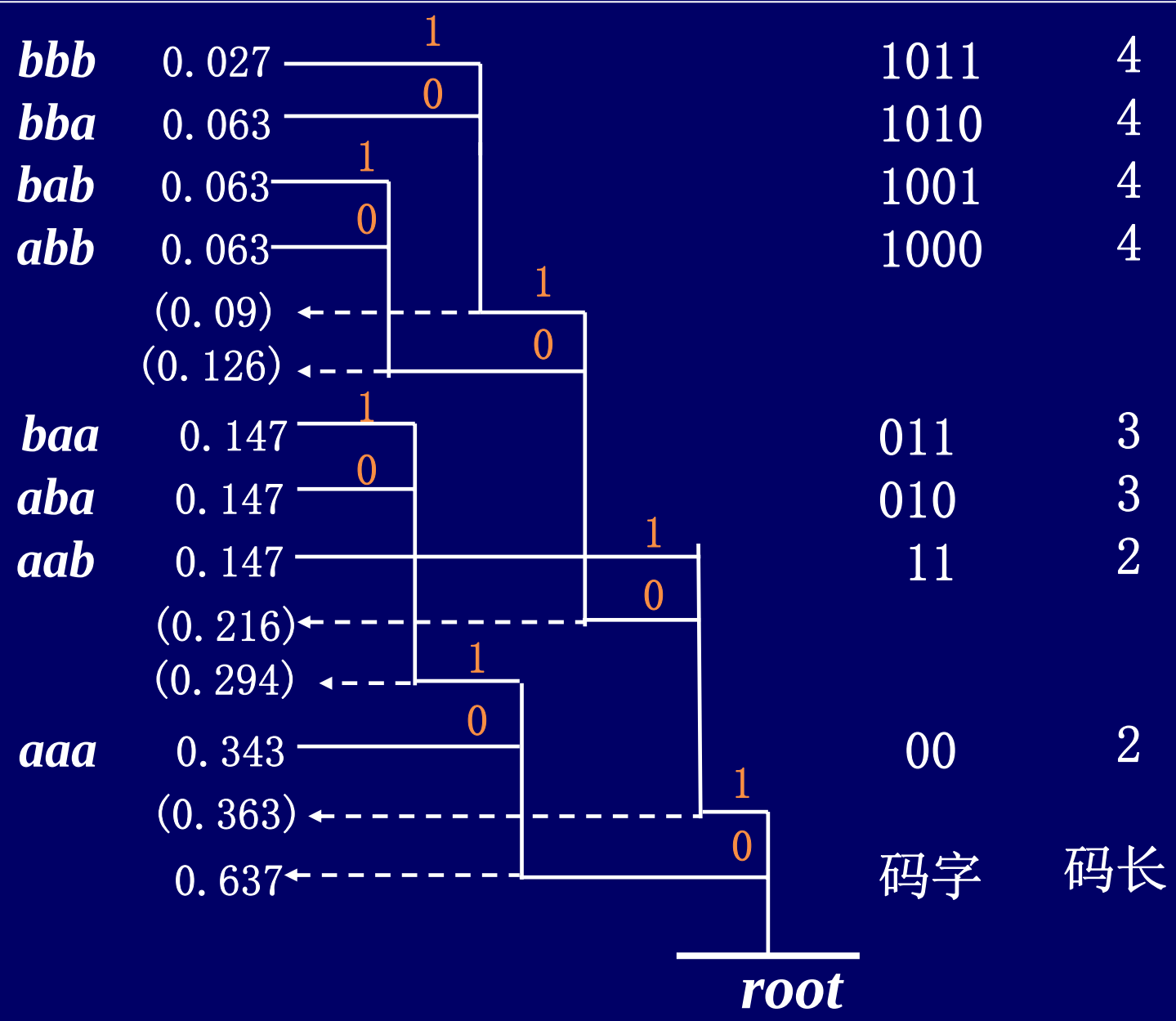
2、Huffman编码的推广：

[例6]二元无记忆信源发出*a*、*b*两个符号，概率分别为0.7和0.3，试用三次扩展信源进行编码。

解：根据 $p(x_1x_2x_3)=p(x_1)p(x_2)p(x_3)$ 不难求出三次扩展信源的概率空间为：

$$\begin{pmatrix} X^3 \\ p(X^3) \end{pmatrix}=\begin{pmatrix} aaa & aab & aba & baa & abb & bab & bba & bbb \\ 0.343 & 0.147 & 0.147 & 0.147 & 0.063 & 0.063 & 0.063 & 0.027 \end{pmatrix}$$

按8个符号的概率排队，进行赫付曼编码


bbb
0.027
bba
0.063
bab
0.063
abb
0.063
(0.09)
(0.126)
baa
0.147
aba
0.147
aab
0.147
(0.216)
(0.294)
aaa
0.343
(0.363)
0.637
root
1
0
1011
1010
1001
1000
011
010
11
00
码字
码长
4
4
4
4
3
3
2
2

$$\begin{pmatrix} X^3 \\ p(X^3) \end{pmatrix} = \begin{pmatrix} aaa & aab & aba & baa & abb & bab & bba & bbb \\ 0.343 & 0.147 & 0.147 & 0.147 & 0.063 & 0.063 & 0.063 & 0.027 \end{pmatrix}$$

码字	00	11	010	011	1000	1001	1010	1011
码长	2	2	3	3	4	4	4	4

码字平均长度： L_N=2. 726；

信源符号平均编码长度：$\overline{l}$ =2.726/3=0.909

编码效率： η=0. 881/0. 909=96. 9%

3、游程编码

适用于连0连1情况。不改变信息熵，但将二元码变成了多元码。为使用Huffman编码创造了条件。

4、算术编码

属于序列编码。通过计算非等概信源的序列概率的积累概率直接找到相应的等概序列。

5、词典编码

不依赖于概率的通用编码。建立初始小词典后边输入边查词典边补充新词条，以词条序号为编码。

第三部分、信道编码

3.1 信道编码理论：

1、检错与纠错原理：

（1）检错原理：添加冗余避免码字非此即彼；

（2）纠错原理：添加冗余拉大码字汉明间距；

（3）汉明距离：两码字不同码元的个数；

（4）检错能力：$d_0 \geqslant e+1$

纠错能力：$d_0 \geqslant 2t+1$

纠检同时：$d_0 \geqslant e+t+1$ 且 $e>t$

2、译码规则与错误概率：

（1）最小错误准则：选联合概率矩阵每列最大元素

（2）最大似然准则：选传输概率矩阵每列最大元素

（3）错误概率计算：未选中元素的总概率

（4）差错率计算：采用信道编码与译码后仍然不能纠正的错误所具有的概率。也就是（3）的结果。

（5）漏检率计算：使用信道编码与译码后仍然不能发现的错误具有的概率。使用反馈重发方式时的差错率就等于漏检率。

[例7] 已知对称信道 $P=\begin{pmatrix} \frac{5}{9} & \frac{3}{9} & \frac{1}{9} \\ \frac{1}{9} & \frac{5}{9} & \frac{3}{9} \\ \frac{3}{9} & \frac{1}{9} & \frac{5}{9} \end{pmatrix}$ (1)求信道容量和最佳信源。

(2)按最大似然译码准则确定其译码准则，并计算错误概率。

解：（1）最佳输入为等概率分布。信道容量为：$-\frac{5}{9}\log\frac{5}{9}-\frac{3}{9}\log\frac{3}{9}-\frac{1}{9}\log\frac{1}{9}$

$C=\log S-H(p_1,p_2,p_3)=\log 3-$

$=1.5850-1.3516=0.2334$ *bit*/符号

（2）译码规则: $F(b_1)=a_1$; $F(b_2)=a_2$; $F(b_3)=a_3$;

错误概率： $P_E=3(3/9+1/9)/3=4/9=0.44$

3.2 线性分组码:

码长为n，信息位为k，记作（n, k）；

监督位$r = n-k$

1、编码

$C = K \cdot G$

生成矩阵G是$k \times n$的矩阵。

- 左半边是$k \times k$单位信息位方阵I_k
- 右半边是$k \times r$的监督位矩阵Q

2、纠错和译码

- ***H*—致校验矩阵**
  - 左半边是$r \times k$矩阵P
  - 右半边是$r \times r$单位方阵I_r;
  - P与生成矩阵中的Q互为转置关系：$P=Q^T$
- 监督方程也可表示为：　$C \cdot H^T = 0$;
- 满足此方程的均为正确的许用码字，否则，便是误码。

- **N 维错误格式矢量 E**

  发送码字为C，接收码字为**R**，三者的关系是：

  $E = C \oplus R$；　$R = C \oplus E$；　$C = R \oplus E$；

- **伴随子向量 S = $R \cdot H^T$**

  $S = R \cdot H^T = (C \oplus E) \cdot H^T = C \cdot H^T \oplus E \cdot H^T = E \cdot H^T$；

  - 若$R = C$，E为全零向量，则S=0；
  - 反之，若$R \neq C$，则$E \neq 0$，导致$S \neq 0$；因此由伴随子向量S是否为零就能检查出接收码R是否有误。

- 纠错：
  - R错一位的情况：S与H^T的哪一行相同，就表明错在哪一位。
  - R错两位以上：查表法， 查R-C对照来译码。

3、纠错能力不等式：

$$2^r \geqslant C_n^0 + C_n^1 + C_n^2 + \cdots\cdots + C_n^t$$

- 完备码：上式取等号的情况。
- 汉明码：纠1位错的完备码， $2^r = 1 + n$

[例8] 已知二元无记忆对称信道的单符号传输的错误概率为0.01；试讨论（7, 4）汉明码对减小差错率的作用。

（1）不进行信道编码, 每位信息差错率为: $p=0.01$

（2）通过编码：(7, 4）码能纠正1位错。

7位码元中1位错6位对的概率是：$7p(1-p)^6$;

7位全对的概率是$(1-p)^7$，所以每个码字的差错率为：$1-(1-p)^7-7p(1-p)^6$

$$=1-0.99^7-7\times0.01\times0.99^6=0.002$$

平均每位信息的差错率为0.002/4=0.0005

3.3、循环码

1. 码多项式

2. 生成多项式——码多项式中那个次数最低的非零多项式$g(x)$

- $n-k=r$ 次； 常数项为1。
- 任意码多项式都是生成多项式 $g(x)$ 的倍式。
- $g(x)$ 是x^n-1的一个因式。

3. 编码

- 确定编码的n、k、r 值；
- 写出给定信息位多项式：$K(x)$；
- 左移r位：$x^r \cdot k(x)$
- 计算监督位多项式：

$$r(x) = x^r \cdot K(x) \bmod g(x)$$ ；

- 写出码多项式：$C(x) = x^r \cdot K(x) + r(x)$
- 写出码字：C

4. 纠错、译码

- 接收码多项式$R(x)$;
- 伴随子多项式$S(x) = R(x) \mod g(x)$;
  - 若 $S(x) = 0$，则表明接收码无误。
  - 若 $S(x) \neq 0$，表明接收码有误。
- $S(x) =[C(x)+E(x)] \mod g(x) = E(x) \mod g(x)$;
- 列$S(x)$—$E(x)$ 对照表，由$S(x)$查出$E(x)$
- $C(x) = R(x) + E(x)$

[例9] 求(7, 4)循环码生成多项式且为信息位（0100）编码。

若接收到一个码字R=(0110010)，试进行检、纠错。

解：1）分解x^7-1得到3次的生成多项式$g(x) = x^3+x+1$；

2） $k(x) = x^2$，r = 3，$x^r \cdot k(x) = x^5$

∵ $r(x) = x^5 \bmod x^3+x+1 = x^2+x+1$；

∴ $C(x) = x^5 + x^2 + x + 1$；

∴ $C = (0100111)$；

3） $R = (0110010)$，$R(x) = x^5+x^4+x$；

$S(x)=（x^5+x^4+x）\bmod（x^3+x+1）= x+1$；

当$E(x)=x^3$时$S(x)=x^3 \bmod（x^3+x+1）= x+1$；

∴ $C(x) =R(x)+E(x)= x^5 + x^4 + x^3 + x$；

∴ $C = (0111010)$；

3.4、循环码的扩展

1. 截短的循环码

由(n,k)循环码截短生成的$(n-i, k-i)$码。如**CRC**码

2. 本原BCH码

本原**BCH**码取码长为$n = 2^m - 1$，生成多项式为：

$$g(x) = LCM[m_1(x)\, m_3(x)\, m_5(x) \cdots\cdots m_{2t-1}(x)]$$

它是能纠正多位错的循环码。式中 t 是纠错位数，$m_i(x)$是x^n-1的因式。

3.5、卷积码

1. (n, k, m)的含义
2. 输出对输入的依赖关系
   - 监督方程
   - 冲激响应
   - 转移函数矩阵
3. 状态转移图与格图：状态指寄存器的值。
4. 编码与译码

[例10]已知(2, 1, 2)卷积码的冲击响应为：

g_1=(1, 1, 1)，　g_2=(1, 0, 1)；

（1）写出监督方程和转移函数矩阵。

（2）画出编码器的电路框图。

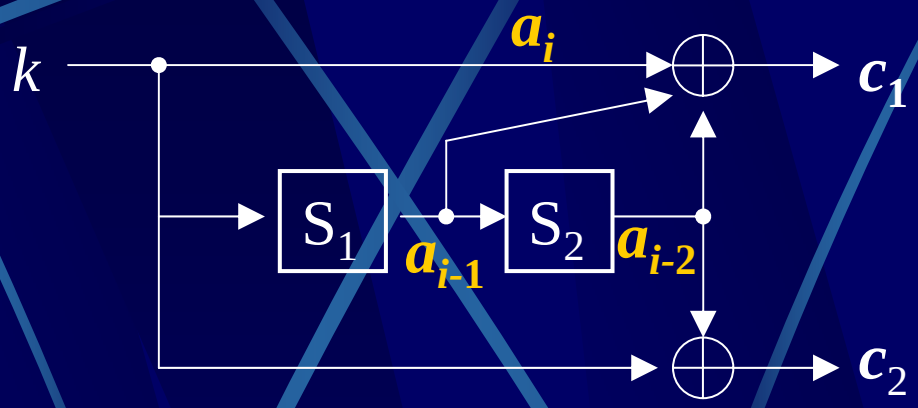
（3）画出状态转移图与格图。

（4）求输入序列为u=(10111)时的输出编码序列。

解：（1）转移函数矩阵 $\boldsymbol{G}(\boldsymbol{D})=(1+\mathbf{D}+\boldsymbol{D}^2,\ \ 1+\boldsymbol{D}^2)$

监督方程：
$$\begin{cases} c_1 = a_i + a_{i-1} + a_{i-2} \\ c_2 = a_i + a_{i-2} \end{cases}$$

（2）电路图



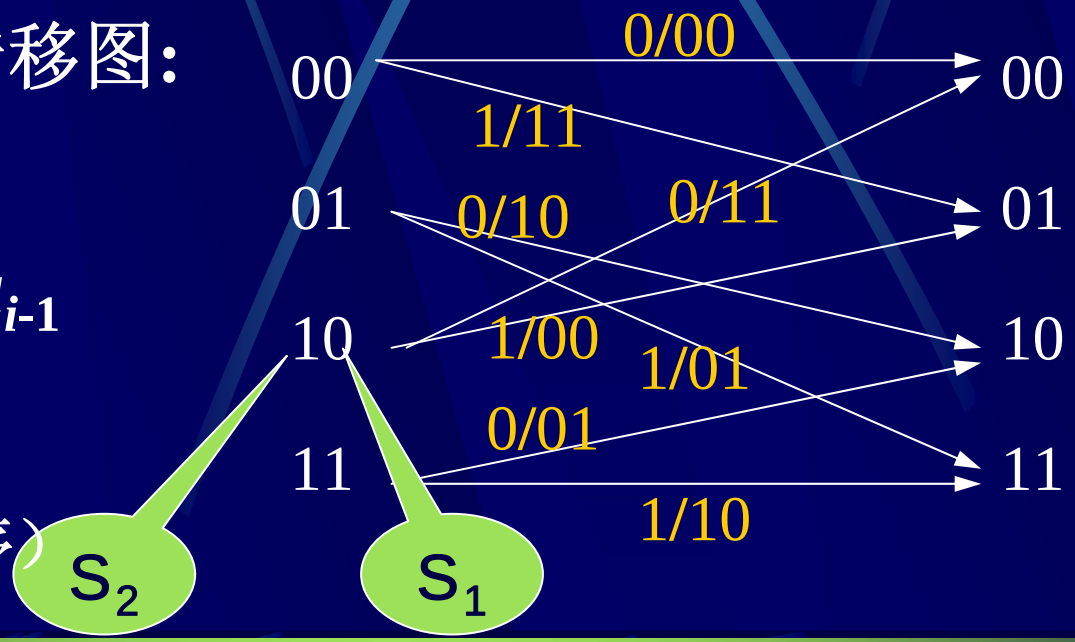
（3）状态转移图：

注意！

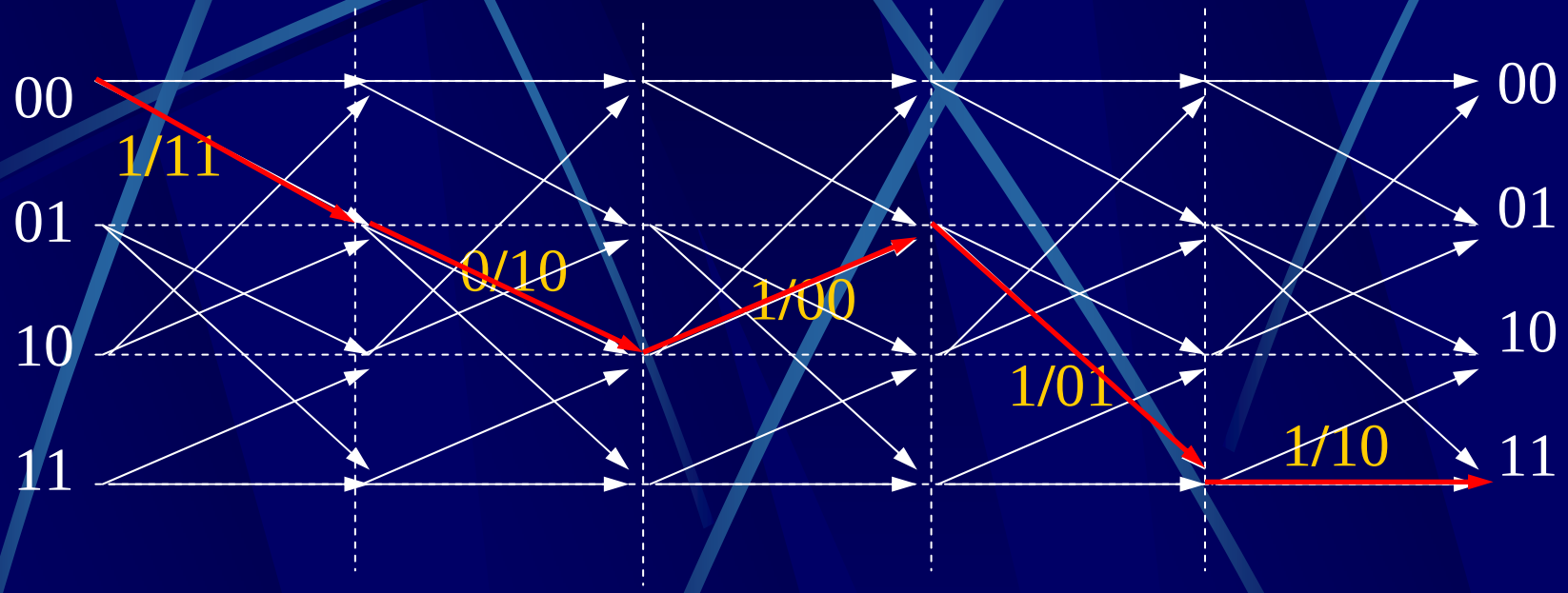
a_i的状态是$a_{i-2}a_{i-1}$

即寄存器S_2S_1的值。

（注意先后次序）



（3）格图



（4）当输入为**10111**时，沿红色路径前进，

得到的输出编码是：**11 10 00 01 10**

第四部分、密码

4.1 密码有关概念：

1、密码的功能：保密与认证

2、密码系统的分类：

（1）体制：对称密钥体系与公开密钥体系；

（2）结构：序列密码与分组密码

（3）学科分支：密码编码学与密码分析学

（4）发展阶段：传统密码学与现代密码学

3、对称密钥体制的缺憾：

4.2 序列密码：

1、**加、解密流程**：保密与认证

2、对序列密钥的要求：

3、线性反馈移位寄存器与*m*序列

（1）*m*序列的周期：$m=2^n-1$

（2） *m*序列的数目：$\lambda(n)=\Phi(m)/n$

（3）**LFSR**的结构：特征多项式=本原多项式

4、*m*序列的非线性组合

4.3 公开密钥密码：

1、特点：

2、如何实现加密与认证功能：

3、RSA体制：

（1）系统构建：$n=pq$，$ed=1 \bmod (p-1)(q-1)$

（2）加解密方法：$c=m^e \bmod n$，$m=c^d \bmod n$

（3）安全性：销毁p, q后，由(n,e)得不到d

4、公开密钥的设计理念

典型题目讲解

一．填空：每题2分，共28分

1. 码长为17可纠正2位错的线性分组码应表示为(17,9) 码；

2. 某线性分组码的全部码字为：C1＝000000；C2＝100011；C3＝010101；C4＝001111；C5＝110110；C6＝011010；C7＝101100；C8＝111001。该码为（ 6 ， 3 ）分组码，最小汉明距离为 3 ，能够纠正 1 位错。

3. 信源编码的主要目的是压缩代码长度；信道编码的主要目的是检查纠正错误；保密编码的主要目的是增强抗攻击性。

4. 某二元无记忆信源发出100个二元符号，其中有m个“1”，若$P(0)=1/4$，则总自信息为$100-1.585m$。

第五部分、典型题目

5. 信道误码率为$p=10^{-3}$，采用三连重复码编码传输时的差错率为 3×10^{-6} 。

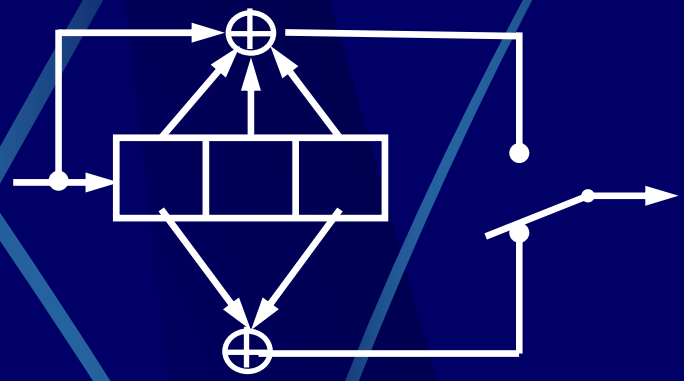
6. 用户A的公钥为(n_1, e_1),私钥为d_1，用户B的公钥为(n_2, e_2),私钥为d_2。为了对消息内容保密，要求用户A利用RSA密码体系把明文M加密后发送给B，则加密算法为 $C=M^{e_2} \bmod n_2$, B收到密文C后, 解密算法为$M=C^{d_2} \bmod n_2$ 。

7. 利用6级线性反馈移位寄存器可以产生周期为 63 的 m 序列。

8.香农信源编码定理指出，无失真压缩的平均码长不能小于信源信息熵 ，香农信道编码定理指出，无差错传输的传信率不能小于 信道容量 。

9. 已知$p(0)=0.25$，$p(1)=0.75$，则序列$S=11101$的序列概率$p(S)$为 0.0791，积累概率$F(S)$为0.6045 ，算术编码 1010 为_____。

第五部分、典型题目



10. 右图电路所示卷积码是(2, 1, 3)码。当输入代码序列为1011011时，输出码流为 10 11 00 10 01 11 10 。(设初态为000）

11. 二元对称信道的信道容量为 $1+p\log p+(1-p)\log(1-p)$ 。

12.某理想通信系统，信噪比为3dB，为使功率节省一半又不损失信息量，带宽应增加到原来的 $\log_2 3=1.585$ 倍。

13.连续信号的概率密度函数在($0\leqslant x\leqslant 2$) 范围 $p(x)=0.5\,x$，其它范围恒为0；则相对信息熵为 $2-\log_2 e=0.5573$ ；

14.英文字母a—z的序号分别对应0—25，按公式C=5M+12 mod 26将明文M=security进行加密，密文 YGWITADC C= _____ 。

$P(X_2 \mid X_1)=\begin{pmatrix}0.9 & 0.1\\ 0.2 & 0.8\end{pmatrix}$

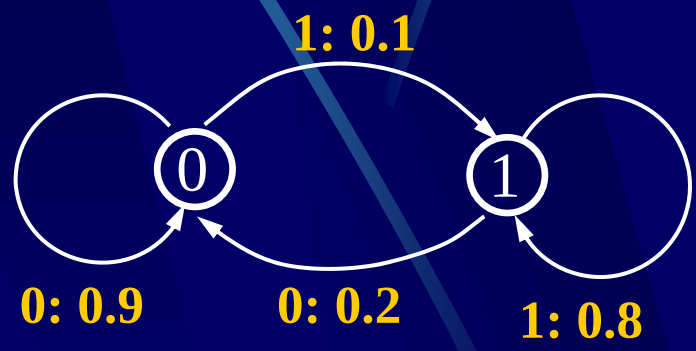
第五部分、典型题目

二．计算题：每题12分，共72分

1． 一阶马尔科夫信源发出二元序列；前后符号之间的条件概率为：$P(X_2 \mid X_1)=\begin{pmatrix}0.9 & 0.1\\ 0.2 & 0.8\end{pmatrix}$ 计算无失真压缩编码的极限码长。

解：由状态转移图得到稳态方程：

$$\begin{cases}Q(0)=0.9Q(0)+0.2Q(1)\\ Q(0)+Q(1)=1\end{cases}$$



解得：**Q(0)=1/3，Q(1)=2/3；**

极限熵：$H=H(x|Q(0))+H(x|Q(1))$

$=-(0.9\log 0.9+0.1\log 0.1)/3-2(0.8\log 0.8+0.2\log 0.2)/3=0.5533$；

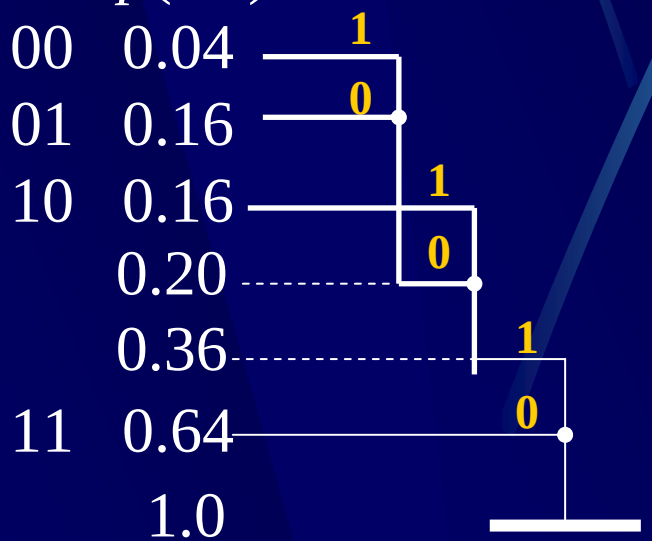
答：无失真压缩编码的极限码长是**0.5533**。

第五部分、典型题目

2．信源发出二元序列，符号概率$p(0)=0.2$，$p(1)=0.8$；请设计一个最佳二元压缩编码方案并进行编码，使编码效率在90%以上。写出所编码字，并算出编码前后信息率。（如果编码前后信源码率不变，问发信率提高多少？）

解：采用二次扩展信源。概率空间为：

X^2	$p(X^2)$	码字：	码长
00	0.04	101	3
01	0.16	100	3
10	0.16	11	2
11	0.64	0	1



平均码长：0.78

信息熵：0.7219

效率：0.9255 >0.9

编码前信息率0.7219

编码后信息率0.9255

发信率提高0.2036

第五部分、典型题目

3. 二元信源每秒发出**100**个符号，若信源概率为**0.6**和**0.4**；信道传输矩阵为：$P=\begin{pmatrix}\frac{3}{4} & \frac{1}{4}\\ \frac{1}{4} & \frac{3}{4}\end{pmatrix}$；求发信率和传信率。

解：发信率就是信源符号熵乘以码率：

$H(X) = -0.6\log 0.6-0.4\log 0.4= 0.9710$ *bit*/符号

$R_b=R_B H(X)= 97.1$ *bit/s*

传信率就是平均互信息乘以码率：

$\because p(xy)=\begin{pmatrix}0.45 & 0.15\\ 0.10 & 0.30\end{pmatrix}$；　$p(y)=(0.55 \quad 0.45)$

联合熵：$H(XY)=1.7822$;

接收符号熵：$H(Y)=0.9928$;

平均互信息：$I(X;Y)=H(X)+H(Y)-H(XY)=0.1816$ *bit*/符号

传信率：$R_t=R_B I(X;Y)= 18.16$ *bit/s*

第五部分、典型题目

4.（1）求(7, 3)循环码生成矩阵与一致校验矩阵。

（2）为信息111编码；（3）为R=1101011 译码。

解：$n=7, k=3, r=4, 2^r-1=15$, 可构造(15,11) 汉明码，而(7, 3)循环码可以由(15,11) 截短得到（15-8，11-8）

(15,11)码的生成多项式为：$g(x)=x^4+x+1$

对应的(7, 3)码的码字是：0010011，

循环移位得到：0100110 与 1001100

因此生成矩阵是：

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

一致监督矩阵是：

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

（2）信息111编码为：

$$C = K \cdot G = (1\ 1\ 1) \cdot \begin{pmatrix} 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 \end{pmatrix} = (1\,1\,1\,1\,0\,0\,1)$$

（3）当收到R=1101011 时：

$$S = R \cdot H^{T} = (1\,1\,0\,1\,0\,1\,1) \cdot \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = (0\ 0\ 0\ 1)$$

S与H^T最下面一行相同，错误格式为E=(0000001)，纠错得到：$C=R \oplus E$=(1101010)

第五部分、典型题目

5.（15,7）循环码的生成多项式为$g(x)=x^8+x^7+x^6+x^4+1$；请为信息位（0111010）编码；若接收到一个码字$R=(111010110010001)$，问是否有错？

解：（1）编码过程：

信息位多项式 $K(x)=x^5+x^4+x^3+x$

监督位多项式 $r(x)=x^rK(x) \mod g(x)=$

$=(x^{13}+x^{12}+x^{11}+x^9) \mod (x^8+x^7+x^6+x^4+1)=x^5$

码多项式：$C(x)= x^{13}+x^{12}+x^{11}+x^9+x^5$

所以码字是(0111010000100000)

（2）译码过程：

$R(x)= x^{14}+x^{13}+x^{12}+x^{10}+x^8+x^7+x^4+1$

因为 $R(x) \mod g(x) = 0$；　　所以此码正确。

第五部分、典型题目

6. 构造码长17且能纠正2位错的非本原BCH码的多项式。

解： 由$n=17, t=2,$ 不难求出监督位$r=8$，这是因为：

$$2^8-1=255>C_{17}^1+C_{17}^2=17+136=153$$

$r=8$时循环码的码长本是255，但因为255/15=17，表明$(x^{255}-1)$有一个非本原因式$m_{15}(x)$能够除尽$(x^{17}-1)$

由此可以构造码长17的非本原BCH码。

查194页表4，查$r=8$阶 $i=15$类，是$(727)_8=(111010111)_2$

$m_{15}(x)= x^8+x^7+x^6+x^4+x^2+x+1$

因此 $g(x)=x^8+x^7+x^6+x^4+x^2+x+1$为(17,9)码的生成多项式。

全面复习

重点掌握

重视概念

细心运算

动手动脑

不存侥幸

祝大家取得满意的考试成绩！