第5章

信源编码

本章学习重点：

- 信源编码的目的和码的定义。
- 无失真信源编码定理。
- 无失真信源编码方法。
- 常用的无失真信源编码方法。
- 限失真信源编码定理。
- 常用的限失真信源编码方法。

5.1 知识点

5.1.1 信源编码

1. 信源编码

分组码：信源输出符号序列 $\boldsymbol{X}=(X_1,X_2,\cdots,X_L)$，序列长度为 L，$X_1\in\{a_1,a_2,\cdots,a_n\}$；每个符号序列 $\boldsymbol{X}_i$ 依照码表映射成一个码字 $\boldsymbol{Y}_i$，这种码也称为分组码或块码，$\boldsymbol{Y}=(Y_1,Y_2,\cdots,Y_K)$，$Y_k\in\{b_1,b_2,\cdots,b_m\}$。

非分组码：不存在码表。

2. N 次扩展码

分组码的 N 次扩展码：设信源符号 $s_i,i=1,2,\cdots,q$，映射为一个固定的码字 $W_i,i=1,2,\cdots,q$，则码 $\alpha_j=(s_{j1},s_{j2},\cdots,s_{jN})$ 映射为 $W_j=(W_{j1},W_{j2},\cdots,W_{jN})$ 的分组码称为原分组码的 N 次扩展。

3. 分组码的性质

1）非奇异性

若一种分组码中的所有码字都互不相同，则称此分组码为非奇异码，否则称为奇异码。

分组码是非奇异码只是正确译码的必要条件。

2）唯一可译性

一个分组码若对于任意有限的整数 N，其 N 阶扩展码均为非奇异的，则称为唯一可译码。

分组码是唯一可译的，这是分组码正确译码的充要条件。

3）即时码

无须考虑后续的码符号即可以从码符号序列中译出码字，这样的唯一可译码称为即时码。

即时码的条件：设 $W_1=W_{i1}W_{i2}\cdots W_{il}$ 为一个码字，对于任意的 $1\leqslant j\leqslant l$，称码符号序列的前 j 个元素 $W_{i1}W_{i2}\cdots W_{ij}$ 为码字 W_i 的前缀。

唯一可译码成为即时码的充要条件是码组中任何一个码字都不是其他码字的前缀。

5.1.2　无失真信源编码

1. 定长编码

离散无记忆信源 $\begin{bmatrix}X\\P\end{bmatrix}=\begin{bmatrix}x_1 & x_2 & \cdots & x_n\\ p(x_1) & p(x_2) & \cdots & p(x_n)\end{bmatrix}$ 的熵为 $H(X)$，由 L 个符号组成的平稳序列$(X_1,X_2,\cdots,X_L)$，可用 K 个符号$(Y_1,Y_2,\cdots,Y_K)$，$y_k\in\{y_1,y_2,\cdots,y_m\}$进行定长编码。对于$\forall\varepsilon>0,\delta>0$，只要满足$\dfrac{K}{L}\log m\geqslant H_L(\boldsymbol{X})+\varepsilon$，则当 L 足够大时，必可使译码差错小于 δ。

反之，若$\dfrac{K}{L}\log m\leqslant H_L(\boldsymbol{X})-2\varepsilon$，译码差错一定是有限值，且当 L 足够大时，译码错误概率趋于 1。

2. 变长编码

为了能够即时译码，变长码必须是即时码，且变长码一般也要为唯一可译码，所以需要先判断唯一可译码的充要条件。

克拉夫特(Kraft)不等式：设信源符号集 $S=(s_1,s_2,\cdots,s_q)$，码符号集 $X=(x_1,x_2,\cdots,x_r)$，对信源进行编码，相应的码字 $W=(W_1,W_2,\cdots,W_q)$，分别对应的码长为 $l_1,l_2,\cdots,l_q$，则唯一可译码存在的充要条件是 $\sum_{i=1}^{q}r^{-l_i}\leqslant 1$。

单个符号变长编码定理：若离散无记忆信源的符号熵为 $H(X)$，每个信源符号用 m 进制码元进行变长编码，一定存在一种无失真编码方法，其码字平均长度 $\overline{K}$ 满足下列不等式

$$\frac{H(X)}{\log m}\leqslant\overline{K}\leqslant\frac{H(X)}{\log m}+1$$

离散平稳无记忆序列变长编码定理：对于平均符号熵为 $H_L(\boldsymbol{X})$的离散平稳无记忆信源，必存在一种无失真编码方法，使平均信息率 $\overline{K}$ 满足不等式

$$H_L(\boldsymbol{X})\leqslant\overline{K}\leqslant H_L(\boldsymbol{X})+\varepsilon$$

式中，ε 为任意小的正数。

5.1.3　无失真信源编码方法

1. 香农编码

香农编码是一种常见的可变长编码，其理论基础是符号的码字长度完全由该符号出现

的概率来决定。

2. 哈夫曼编码

哈夫曼编码也是一种常见的可变长编码，依据各符号出现的概率来构造码字，其基本原理是基于二叉树的编码思想，所有可能的输入符号在哈夫曼树上对应为一个节点，节点的位置就是该符号的哈夫曼编码。

3. 算术编码

算术编码是非分组码的编码方法，它从全信源序列出发，考虑符号之间的依赖关系直接对信源符号序列进行编码。

5.1.4 限失真信源编码

限失真信源编码定理：设离散无记忆信源 X 的信息率失真函数为 $R(D)$，则当信息传输率 $R>R(D)$ 时，只要信源序列长度 L 足够长，一定存在一种编码方法，其译码失真小于或等于 $D+\varepsilon$，ε 为任意小的正数；反之，若 $R<R(D)$，则无论采用什么样的编码方法，其译码失真必大于 D。

该定理也称香农第三极限定理。在允许一定失真的情况下，信源的信息率失真函数 $R(D)$ 可以作为衡量各种压缩编码方法性能优劣的一种尺度。

5.1.5 限失真信源编码方法

1. 矢量量化编码

矢量量化编码的原理是将输入的 k 维矢量 $\boldsymbol{X}_i$ 通过一个矢量量化器 $Q(\boldsymbol{X})$，映射成对应的 k 维输出矢量 $\boldsymbol{Y}_i$，$\boldsymbol{Y}=\{\boldsymbol{Y}_1,\boldsymbol{Y}_2,\cdots,\boldsymbol{Y}_N\}$，共有 N 种矢量组成的集合 $\boldsymbol{Y}$ 称为码书。矢量量化编码的设计关键是码书设计和码字搜索。

2. 预测编码

预测编码是利用信源的相关性来压缩码率，它从已收到的符号中提取关于未收到符号的信息，从而预测最可能的值作为预测值，并对预测值和实际值之差进行编码。

3. 变换编码

变换编码通过变换来解除或减弱信源符号间的相关性，再将变换后的样值进行标量量化，或采用对于独立信源符号的编码方法，以达到压缩码率的目的。

5.2 习题详解

5.2.1 选择题

1. 下列组合中不属于即时码的是(　　)。

A. {0,01,011}　　B. {0,10,110}

C. {00,10,11}　　D. {1,01,00}

解答：A。A 中的 0 码字是 01 和 011 的前缀，01 又是 011 的前缀，所以不是即时码。

2. 为提高通信系统传输消息的有效性，信源编码采用的方法是(　　)。

A. 压缩信源的冗余度　　B. 在信息比特中适当加入冗余比特

C. 设计码的生成矩阵　　D. 对多组信息进行交织处理

解答：A。

3. 关于克拉夫特不等式，正确的是(　　)。

A. 如果编出的码字长度满足克拉夫特不等式，则该码字是唯一可译码

B. 克拉夫特不等式是唯一可译码的充要条件

C. 定长码一定满足克拉夫特不等式

D. 即时码有时不满足克拉夫特不等式

解答：D。

4. 根据树图法构成规则，正确的是(　　)。

A. 在树根上安排码字　　B. 在树枝上安排码字

C. 在中间节点上安排码字　　D. 在终端节点上安排码字

解答：D。

5. 下列说法正确的是(　　)。

A. 奇异码是唯一可译码　　B. 非奇异码是唯一可译码

C. 非奇异码不一定是唯一可译码　　D. 非奇异码不是唯一可译码

解答：C。

6. 有关哈夫曼编码，以下(　　)不是哈夫曼编码需要进一步研究的问题。

A. 误差扩散　　B. 速率匹配　　C. 概率匹配　　D. 方差匹配

解答：D。

7. 信源编码中，编码效率与(　　)无关。

A. 平均码长　　B. $H(X)$

C. 信源符号的概率分布　　D. 码长的方差

解答：D。

8. 以下关于哈夫曼编码方法的说法中，错误的是(　　)。

A. 哈夫曼编码需要精确测定信源的概率特性

B. 对于二元信源进行哈夫曼编码时，需要多个符号合起来编码

C. 由哈夫曼编码方法得到的码字集合是唯一的

D. 哈夫曼编码是用概率匹配方法进行信源编码

解答：C。哈夫曼编码方法得到的码字集合并不唯一。

5.2.2　名词解释

1. 最佳变长码

解答：凡是能载荷一定的信息量，且码字的平均长度最短，可分离的变长码的码字集合称为最佳变长码。

2. 算术编码

解答：从全序列出发，将各信源序列的概率映射到[0,1]区间，使每个序列对应到区间上的一点，这些点将区间分成若干小段，每段的长度对应某一序列的概率。再在段内取一个二进制小数，其长度与该序列的概率匹配，达到高效率编码的目的。

3. 变换编码

解答：变换编码是一种限失真信源编码方法，经变换后的信号的样值能更有效地编码，

即通过变换解除或减弱信源符号间的相关性，再将变换后的样值进行标量量化，从而达到压缩码率的目的。

4. 非奇异码

解答：非奇异码是信源编码中信源符号和编出的码字是一一对应的一类码。

5.2.3 判断题

1. 异前置码一定是唯一可译码。

解答：对。

2. 信源编码通常是通过压缩信源的冗余度来实现的。

解答：对。

3. 香农编码、费诺编码和哈夫曼编码中，编码方法唯一的是香农编码。

解答：对。

4. 非奇异码一定是唯一可译码，唯一可译码不一定是非奇异码。

解答：错。非奇异码并不一定是唯一可译码，唯一可译码一定是非奇异码。

5. 哈夫曼编码是用概率匹配方法进行信源编码。

解答：对。

6. 一个唯一可译码成为即时码的充要条件是其中任何一个码字都不是其他码字的前缀。

解答：对。

7. 对于独立信源，不可能进行预测编码。

解答：对。

8. 对某 7 符号信源进行编码，编出的一组码组为{a，c，bad，ad，abb，deb，bbcde}，这组码是唯一可译码。

解答：错。码组中的 ad 在该码的后缀集合中，所以不是唯一可译码。

9. 哈夫曼编码在编码时充分考虑了信源符号分布的不均匀性和符号之间的相关性。

解答：错。哈夫曼编码完全依据符号出现的概率，即信源符号的概率分布来进行编码。

10. 存在长度为(1,2,3,3,3,4,4)的二进制唯一可译码。

解答：错。由克拉夫特不等式得 $\sum_{i=1}^{7}2^{-K_i}=2^{-1}+2^{-2}+3\times2^{-3}+2\times2^{-4}=\frac{5}{4}>1$，则不存在长度为(1,2,3,3,3,4,4) 的二进制唯一可译码。

11. 变长编码的核心问题是寻找紧致码(最佳码)，而哈夫曼编码方法构造的是最佳码。

解答：对。

12. 算术编码是一种无失真的分组信源编码，其基本思想是将一定精度的数值作为序列的编码，是以另外一种形式实现的最佳统计匹配编码。

解答：错。算术编码不是统计匹配编码。

13. 即时码的任意一个码字都不是其他码字的前缀部分。

解答：对。

14. 游程序列的熵(“0”游程序列的熵与“1”游程序列的熵的和)大于或等于原二元序列的熵。

解答：错。游程序列的熵小于或等于原二元序列的熵。

15. 克拉夫特不等式是唯一可译码的充要条件。

解答：错。克拉夫特不等式是唯一可译码存在的充要条件。

5.2.4　填空题

1. 游程编码比较适用于_____序列。

解答：二元。

2. 即时码又称为_____，有时也称为_____。

解答：非延长码；异前缀码。

3. 唯一可译码可分为两大类，_____码和_____码。

解答：即时码；非即时码。

4. 无失真的信源中，信源输出由_____来度量，在有失真的信源中，信源输出由_____来度量。

解答：$H(X)$；$R(D)$。

5. 无失真信源编码的中心任务是编码后的信息传输率压缩接近_____；限失真压缩的中心任务是在给定的失真度条件下，信息传输率压缩接近到_____。

解答：$H(X)$；$H(X)-\varepsilon$。

6. 若一离散无记忆信源的信息熵 $H(X)$ 等于 2.5bit/符号，对信源进行等长的无失真二进制编码，则编码的长度至少为_____ bit。

解答：3。

7. 变长无失真信源编码定理，又称为_____定理，定理说明了只要平均码长_____信源熵，就可以实现唯一可译码。

解答：香农第一；不小于。

5.2.5　问答题

1. 简述保真度准则下的信源编码定理及其物理意义。

解答：保真度准则下的信源编码定理：设离散无记忆信源 X 的信息率失真函数为 $R(D)$，当信息传输率 $R>R(D)$ 时，只要信源序列长度 L 足够长，一定存在一种编码方法，其译码失真小于或等于 $D+\varepsilon$，ε 为任意小的正数；反之，若 $R<R(D)$，则无论采用什么样的编码方法，其译码失真必大于 D。

2. 试写出香农的三个编码定理的主要内容。

解答：香农第一定理：设离散无记忆信源 S 的信源熵为 $H(S)$，它的 N 次扩展信源 $S^N=\{s_1,s_2,\cdots,s_m\}$，$m=q^N$，其熵为 $H(S^N)$。并用码符号 $X=\{x_1,x_2,\cdots,x_r\}$ 对信源 S^N 进行编码，总可以找到一种唯一可译码，使信源 S 中每个符号所需的平均码长满足 $\frac{H(S)}{\log r}\leqslant\frac{\overline{L_N}}{N}\leqslant\frac{H(S)}{\log r}+\frac{1}{N}$。

香农第二定理：设有一个离散无记忆信源平稳信道，其信道容量为 C。当信息传输率 $R<C$ 时，只要码长足够长，则总存在一种编码，可以使平均译码错误概率任意小。

香农第三定理：设 $R(D)$ 是离散无记忆信源的信息率失真函数且为有限值，对于任意的允许失真度 $D\geqslant 0$ 和任意小的正数 ε，当信源序列长度 N 足够长时，一定存在一种编码 C_k，其码字个数 $M\leqslant \exp\{N[R(D)+\varepsilon]\}$，而编码后的平均失真度 $\overline{D}\leqslant D+\varepsilon$。

3. 简述信源编码的主要作用。

解答：信源编码的主要作用有：(1)符号变换，使信源的输出符号与信道的输入符号相匹配；(2)冗余度压缩，编码之后的新信源概率均匀化，信息含量效率等于或接近 100%。

4. 简述信源译码的错误扩展现象。

解答：变长码经由信道传送时，由于信道的干扰作用，造成了一定的错误，这些错误在译码时又造成了更多的错误，这就是信源译码的错误扩展现象。

5. 在哈夫曼编码过程中，对缩减信源符号按概率由大到小的顺序重新排列时，应将合并后的新符号排在同概率大小信源符号的前面还是后面？并说明原因。

解答：合并后的新符号应排在同概率大小信源符号的前面，这样可以使合并的信源符号位于缩减信源序列尽可能高的位置上，减少了概率的合并次数，充分利用了短码。这样编出的码字长度的方差最小，从而提高哈夫曼编码的质量。

6. 预测编码的原理是什么？要实现预测编码，需要考虑哪几个方面的问题？

解答：预测编码是通过解除信源符号在时域上的相关性来实现信源匹配编码的方法。实现预测编码，还要考虑 3 个方面的问题：(1)预测误差准则的选取，决定预测质量标准；(2)预测函数的选取，决定预测质量好坏；(3)预测器的输入数据的选取，决定预测质量好坏。

5.2.6 计算题

1. 将某六进制信源进行二进制编码如表 5.1 所示。(1)这些码中哪些是唯一可译码？(2)哪些码是非延长码(即时码)？(3)对所有唯一可译码求出它们的平均码长和编码效率。

表 5.1 题 1 的表

消息	概率	C_1	C_2	C_3	C_4	C_5	C_6
u_1	1/2	000	0	0	0	1	01
u_2	1/4	001	01	10	10	000	001
u_3	1/16	010	011	110	1101	001	100
u_4	1/16	011	0111	1110	1100	010	101
u_5	1/16	100	01111	11110	1001	110	110
u_6	1/16	101	011111	111110	1111	110	111

解答：(1) 唯一可译码有 C_1、C_2、C_3、C_6。

(2) 非延长码(即时码)有 C_1、C_3、C_6。

(3) $H_L(\boldsymbol{X})=\sum -p_i\log p_i=2\text{bit/符号}$

C_1 码组是定长码，平均码长即为编码长度，$\overline{K}_1=3\text{bit/符号}$，编码效率 $\eta_1=\dfrac{H_L(\boldsymbol{X})}{\overline{K_1}}=\dfrac{2}{3}=66.7\%$。

C_2 码组是变长码，$\overline{K}_2=1\times\frac{1}{2}+2\times\frac{1}{4}+\frac{1}{16}\times(3+4+5+6)=2.125$bit/符号，编码效率 $\eta_2=\frac{H_L(\boldsymbol{X})}{\overline{K}_2}=\frac{2}{2.125}=94.1\%$。

C_3 码组是变长码，平均码长 $\overline{K}_2=\overline{K}_3=2.125$bit/符号，编码效率也为 94.1%。

C_6 码组是变长码，平均码长 $\overline{K}_6=2\times\frac{1}{2}+3\times\frac{1}{4}+3\times\frac{4}{16}=2.5$bit/符号，编码效率 $\eta_6=\frac{H_L(\boldsymbol{X})}{\overline{K_6}}=\frac{2}{2.5}=80\%$。

注：本题考查的知识点是 5.1 节编码的概念，题解：(1)应用克拉夫特不等式和唯一可译码的判别方法判断唯一可译码；(2)根据平均码长的定义计算平均码长。

2. 已知信源的字母消息集为 $X\in\{A,B,C,D\}$，现用二进制码元对消息字母作信源编码，A→00，B→01，C→10，D→11，每个二进制码元的长度为 5ms。(1)若各个字母以等概率出现，计算在无扰离散信道上的平均信息传输速率。(2)若各个字母消息的出现概率分别为 $p(A)=\frac{1}{5}$，$p(B)=\frac{1}{4}$，$p(C)=\frac{1}{4}$，$p(D)=\frac{3}{10}$，再计算在无扰离散信道上的平均信息传输速率。(3)若字母消息改用四进制码元作信源编码，码元幅度分别为 0V，1V，2V，3V，码元长度为 10ms。重新计算(1)和(2)两种情况下的平均信息传输速率。

解答：(1) 4 个字母等概率出现时，$H(X)=\log_2 4=2$bit/字母，平均信息传输速率 $v_1=\frac{H(X)}{t}=\frac{2}{2\times5\times10^{-3}}=200$bit/s。

(2) 各字母不等概率出现时，$H(X)=H\left(\frac{1}{5},\frac{1}{4},\frac{1}{4},\frac{3}{10}\right)=1.985$bit/字母，平均信息传输速率 $v_2=\frac{H(X)}{t}=\frac{1.985}{2\times5\times10^{-3}}=198.5$bit/s。

(3) 当字母消息改用四进制码元时，各字母消息等概率出现时，$H(X)=\log_2 4=2$ 四进制码元/字母，码元长度为 10ms，则平均信息传输速率

$$v_1'=\frac{H(X)}{t}=\frac{2}{2\times10\times10^{-3}}=100\text{ 四进制码元 /s}=200\text{bit/s}$$

同理，各字母不等概率出现时，平均信息传输速率仍为 198.5bit/s。

注：本题考查的知识点是 5.1 节编码的概念。题解：(1)平均信息传输速率的概念；(2)平均信息传输速率与编码进制的关系。

3. 设信道的基本符号集合 $A=\{a_1,a_2,a_3,a_4,a_5\}$，它们的时间长度分别为 $t_1=1$，$t_2=2$，$t_3=3$，$t_4=4$，$t_5=5$(个码元时间)。用这样的信道基本符号编成消息序列，且不能出现 a_1a_1，a_2a_2，a_1a_2，a_2a_1 这四种符号相连的情况。(1)若信源的消息集合 $X=\{x_1,x_2,\cdots,x_7\}$，它们的出现概率分别为 $p(x_1)=\frac{1}{2}$，$p(x_2)=\frac{1}{4}$，$p(x_3)=\frac{1}{8}$，$p(x_4)=\frac{1}{16}$，$p(x_5)=\frac{1}{32}$，$p(x_6)=p(x_7)=\frac{1}{64}$。试按最佳编码原则利用上述信道来传输这些消息时的信息传输速率；(2)求上述信源编码的编码效率。

解答：(1) 因为规定 a_1 和 a_2 不能连用，故不能用 a_1 和 a_2 单独作为消息的代码组，且在用这 2 个符号时，a_1 和 a_2 要么都在开头，要么都在结尾，以避免头尾相连时出现上述情况。这样，可以编出如表 5.2 所示(这只是符合要求的一种)的码字来传输消息。

表 5.2　题 3 的编码表

消息 x_i	代码组 a_i	码元数 k_i	码元时长 t_i	概率 $p(x_i)$
x_1	a_3	1	3	1/2
x_2	a_4	1	4	1/4
x_3	a_1a_3	2	4	1/8
x_4	a_5	1	5	1/16
x_5	a_1a_4	2	5	1/32
x_6	a_2a_3	2	5	1/64
x_7	a_2a_4	2	6	1/64

信源熵 $H(X)=H\left(\frac{1}{2},\frac{1}{4},\frac{1}{8},\frac{1}{16},\frac{1}{32},\frac{1}{64},\frac{1}{64}\right)=1.969\text{bit/消息}$

代码组的平均时长 $\overline{T}=\sum_i p(x_i)\cdot t_i=3.641$ 码元时间/消息

此时，信息传输速率 $R=\frac{H(X)}{\overline{T}}=\frac{1.969}{3.641}=0.541\text{bit/码元时间}$。

(2) 编码后的平均码长 $\overline{K}=\sum_i p(x_i)\cdot k_i=1.1875$ 码元/消息

则编码效率 $\eta=\frac{H(X)}{\overline{K}\log_2 5}=\frac{1.969}{1.1875\times\log_2 5}=71.4\%$。

注：本题考查的知识点是 5.1 节编码的概念。题解：(1)编码的基本方法；(2)平均码长的定义；(3)考虑编码时的特殊要求，如某些符号的特殊位置等，编码赋码元时要注意。

4. 若消息符号、对应概率分布和二进制编码如表 5.3 所示。试求：(1)消息符号熵；(2)每个消息符号所需的平均二进制码个数；(3)若各消息符号间相互独立，求编码后对应的二进制码序列中出现“0”和“1”的无条件概率 p_0 和 p_1，并求相邻码间的条件概率 $p(0|0)$，$p(0|1)$，$p(1|0)$和 $p(1|1)$。

表 5.3　题 4 的表

消息符号	概　率	编　码
u_0	1/2	0
u_1	1/4	10
u_2	1/8	110
u_3	1/8	111

解答：(1) 该信源的熵 $H(U)=H\left(\frac{1}{2},\frac{1}{4},\frac{1}{8},\frac{1}{8}\right)=\frac{7}{4}\text{bit/符号}$

(2) 平均码长 $\overline{K}=\frac{1}{2}+\frac{2}{4}+\frac{3}{8}+\frac{3}{8}=\frac{7}{4}$二进制码元/符号

(3) 若各消息符号间相互独立，编码后对应的二进制码序列中出现“0”和“1”的无条件

概率为

$$p_0=\left(\frac{1}{2}+\frac{1}{4}+\frac{1}{8}\right)\div\frac{7}{4}=\frac{1}{2},p_1=\left(\frac{1}{4}+\frac{2}{8}+\frac{3}{8}\right)\div\frac{7}{4}=\frac{1}{2}$$

$$p(0\mid 0)=p(u_0\mid u_0)=p(u_0)=\frac{1}{2}$$

$$p(0\mid 1)=p(u_0\mid u_3)=p(u_0)=\frac{1}{2}$$

$$p(1\mid 0)=p(u_1\mid u_0)+p(u_2\mid u_0)+p(u_3\mid u_0)=p(u_1)+p(u_2)+p(u_3)$$
$$=\frac{1}{4}+\frac{1}{8}+\frac{1}{8}=\frac{1}{2}$$

$$p(1\mid 1)=p(u_1\mid u_3)+p(u_2\mid u_3)+p(u_3\mid u_3)=p(u_1)+p(u_2)+p(u_3)=\frac{1}{2}$$

注：本题考查的知识点是5.1节编码的概念。题解：(1)编码的基本方法；(2)有特殊要求的编码。

5. 设信源U可能发出的数字有1、2、3、4、5、6、7，对应的概率分别为$p(1)=p(2)=\frac{1}{3}$，$p(3)=p(4)=\frac{1}{9}$，$p(5)=p(6)=p(7)=\frac{1}{27}$。在二进制或三进制无噪信道中传输，二进制信道中传输一个码字需要1.8元，三进制信道中传输一个码字需要2.7元。试：(1)编出二进制符号的哈夫曼码，求其编码效率；(2)编出三进制符号的费诺码，求其编码效率；(3)根据(1)和(2)的结果，确定在哪种信道中传输的花费较小？

解答：(1) 二进制符号的哈夫曼编码过程如图5.1所示。

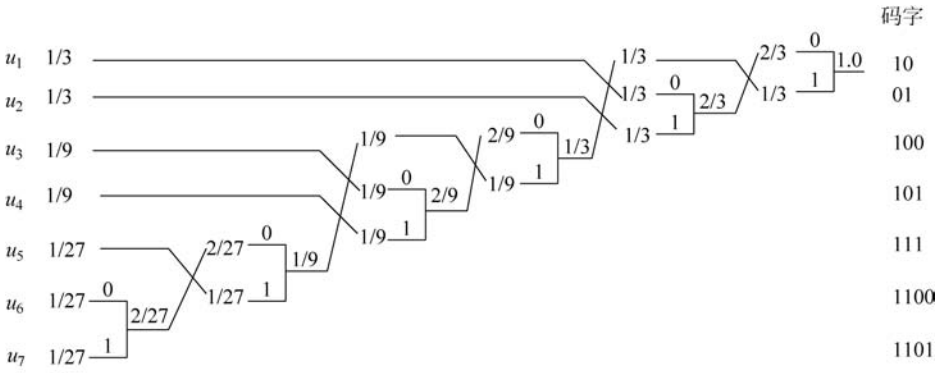

图5.1　题5的哈夫曼编码

信源熵$H(U)=\frac{2}{3}\log_2 3-\frac{2}{9}\log_2 9-\frac{3}{27}\log_2 27=2.29\text{bit/符号}$

平均码长$\overline{K}_H=\frac{1}{3}\times 2\times 2+\frac{1}{9}\times 2\times 3+\frac{1}{27}\times 3+\frac{1}{27}\times 2\times 4=\frac{65}{27}=2.41$二进制码元/符号

编码效率$\eta_H=\frac{H(U)}{\overline{K}_H}=\frac{2.29}{2.41}=0.95$

(2) 如用三进制码元(0,1,2)来进行费诺编码,三次分组及编出的码字如表 5.4 所示。

表 5.4 题 5 的三进费诺编码

符 号	概 率	第一次分组	第二次分组	第三次分组	三进制码字
u_1	1/3	0			0
u_2	1/3	1			1
u_3	1/9		0		20
u_4	1/9		1		21
u_5	1/27	2		0	220
u_6	1/27		2	1	221
u_7	1/27			2	222

平均码长 $K_F=\frac{1}{3}\times(1+1)+\frac{1}{9}\times(2+2)+\frac{1}{27}\times(3+3+3)=1.44$ 三进制码元/符号编码效率 $\eta_F=\frac{H(U)}{K_F}=\frac{2.29}{1.44\times\log_2 3}=1$

(3) 二进制哈夫曼码的花费为 $\frac{65}{27}\times1.8=4.33$ 元/码字,三进制费诺码的花费为 $\frac{13}{9}\times2.7=3.9$ 元/码字。由此可见,三进制费诺码的花费较小。

注:本题考查的知识点是 5.2.2 节变长编码、5.4.1 节哈夫曼编码。题解:(1)二进制最佳编码方法进行编码,并计算其编码效率;(2)多进制编码方法在二进制编码上变化一下。

6. 有一 9 个符号的信源,概率分别为 1/4、1/4、1/8、1/8、1/16、1/16、1/16、1/32、1/32,用三进制符号(a、b、c)编码。试:(1)编出三进制哈夫曼码,并求编码效率;(2)若要求符号 c 后不能紧跟另一个 c,编出一种有效码,其编码效率是多少?

解答:(1) 分别用两种方法对信源进行三进制符号哈夫曼编码,编出的码字如图 5.2 和图 5.3 所示。从图中所示的编码过程可以看出,图 5.2 编出的码字充分利用了短码。两种方法编出的码字平均码长一样,但图 5.2 编出码字的方差较小。

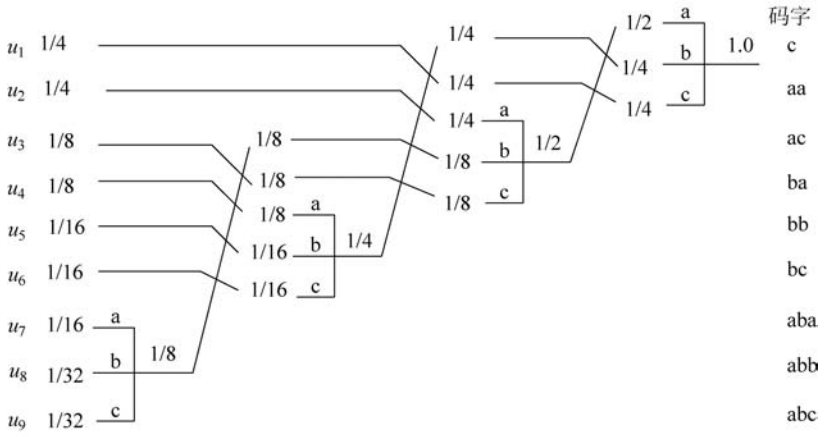

图 5.2 题 6 的三进制哈夫曼编码 1

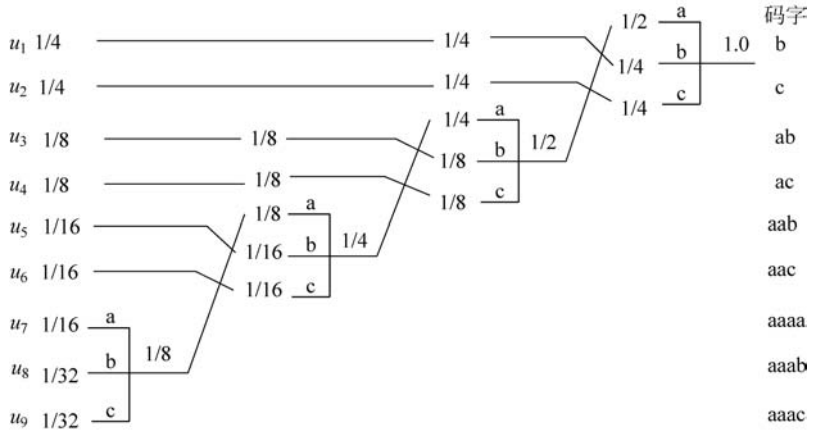

图 5.3　题 6 的三进制哈夫曼编码 2

信源熵

$$H(U)=H\left(\frac{1}{4},\frac{1}{4},\frac{1}{8},\frac{1}{8},\frac{1}{16},\frac{1}{16},\frac{1}{16},\frac{1}{32},\frac{1}{32}\right)=2.81\text{bit/符号}$$

三进制哈夫曼编码的平均码长

$$\overline{K}=\sum_{i=1}^{9}p(u_i)K_i=\frac{15}{8}=1.875\text{ 三进制码元/符号}$$

编码效率 $\eta=\dfrac{2.81}{1.875\times\log_2 3}=0.94$

(2) 若要求符号 c 后不跟 c,则有两种情况必须禁止:

- 在一个码字中出现 cc 组合;
- 若有一个或多个码字是以符号 c 开始的,则码字结尾出现 c。

在如图 5.2 和图 5.3 的编码过程中,避免这两种情况就可以得到满足条件的编码,表 5.5 中给出了满足条件的其中一种可能编码。

表 5.5　题 6 的编码

符　　号	概　　率	三进制码字
u_1	1/4	a
u_2	1/4	ba
u_3	1/8	bb
u_4	1/8	caa
u_5	1/16	cab
u_6	1/16	cba
u_7	1/16	cbb
u_8	1/32	cbca
u_9	1/32	cbcb

该码的平均码长

$$\overline{K}=\frac{1}{4}+2\times\left(\frac{1}{4}+\frac{1}{8}\right)+3\times\left(\frac{1}{8}+\frac{3}{16}\right)+4\times\left(\frac{1}{32}+\frac{1}{32}\right)=2.1875\text{ 三进制码元/符号}$$

此时的编码效率 $\eta=\dfrac{H(U)}{\overline{K}\times\log_2 3}=\dfrac{2.81}{2.1875\times\log_2 3}=0.81$。

还有两种可能的编码如表 5.6 所示。

表 5.6　题 6 的三进制哈夫曼编码

符　　号	概　　率	三进制码字 1	三进制码字 2
u_1	1/4	a	c
u_2	1/4	ca	ba
u_3	1/8	cb	bb
u_4	1/8	baa	aa
u_5	1/16	bab	aba
u_6	1/16	bba	abb
u_7	1/16	bbb	aca
u_8	1/32	bca	acba
u_9	1/32	bcb	acbb

码字 1 的平均码长 $\overline{K}_1=2.125$ 三进制码元/符号

编码效率 $\eta_1=\dfrac{H(U)}{\overline{K}_1\times\log_2 3}=\dfrac{2.82}{2.125\times\log_2 3}=0.84$

码字 2 的平均码长 $\overline{K}_2=2.0625$ 三进制码元/符号

编码效率 $\eta_2=\dfrac{H(U)}{\overline{K}_2\times\log_2 3}=\dfrac{2.82}{2.0625\times\log_2 3}=0.87$

注：本题考查的知识点是 5.4.1 节哈夫曼编码。题解：(1)多进制哈夫曼编码方法；(2)多进制编码的编码效率计算；(3)特殊要求的哈夫曼编码方法。

7. 某信源有 8 个符号$\{u_1,u_2,\cdots,u_8\}$，概率分别为 1/2、1/4、1/8、1/16、1/32、1/64、1/128、1/128，编成这样的码：000,001,010,011,100,101,110,111。试：(1)求信源的符号熵 $H(U)$；(2)求出现一个“1”或一个“0”的概率；(3)求这种码的编码效率；(4)如采用香农编码方法，求编出的香农码及编码效率。

解答：(1) 信源熵为 $H(U)=H\left(\dfrac{1}{2},\dfrac{1}{4},\dfrac{1}{8},\dfrac{1}{16},\dfrac{1}{32},\dfrac{1}{64},\dfrac{1}{128},\dfrac{1}{128}\right)=1.98\text{bit/符号}$

(2) 由题意，编码为定长码，$L=3$。出现一个“0”的概率

$$p(0)=\frac{\sum_i p(u_i)c_{i0}}{L}=\frac{\frac{1}{2}\times3+\frac{1}{4}\times2+\frac{1}{8}\times2+\frac{1}{16}\times1+\frac{1}{32}\times2+\frac{1}{64}\times1+\frac{1}{128}\times1}{3}$$

$$=0.8$$

出现一个“1”的概率 $p(1)=1-p(0)=1-0.8=0.2$。

(3) 该定长码的编码效率 $\eta=\dfrac{H(U)}{L}=\dfrac{1.98}{3}=0.66$。

(4) 根据所给的概率，编出的香农码如表 5.7 所示。

表 5.7　题 7 的香农码

符　号	符号概率	累加概率	$-\log p(x_i)$	码字长度	码　字
u_1	1/2	0	1	1	0
u_2	1/4	1/2	2	2	10
u_3	1/8	3/4	3	3	110
u_4	1/16	7/8	4	4	1110
u_5	1/32	15/16	5	5	11110
u_6	1/64	31/32	6	6	111110
u_7	1/128	63/64	7	7	1111110
u_8	1/128	127/128	7	7	1111111

平均码长

$$\overline{K}=\frac{1}{2}\times 1+\frac{1}{4}\times 2+\frac{1}{8}\times 3+\frac{4}{16}+\frac{5}{32}+\frac{6}{64}+2\times\frac{7}{128}=1.98 \text{ 二进制码元/符号}$$

编码效率 $\eta=\dfrac{H(U)}{\overline{K}}=\dfrac{1.98}{1.98}=1.00$。

注：本题考查的知识点是 5.5.1 节定长编码、5.2.2 节变长编码。题解：(1)应用香农编码的方法进行编码；(2)计算其编码效率。

8. 设无记忆二元信源，概率为 $p_0=0.005$，$p_1=0.995$。信源输出 $N=100$ 的二元序列。在长为 $N=100$ 的信源序列中只对含有 3 个或小于 3 个“0”的各信源序列构成一一对应的一组定长码。(1)求码字所需的最小长度。(2)考虑没有给予编码的信源序列出现的概率，该定长码引起的错误概率 P_e 是多少？

解答：(1) 由题意，含有 3 个或小于 3 个“0”的信源序列共有 $\begin{bmatrix}100\\0\end{bmatrix}+\begin{bmatrix}100\\1\end{bmatrix}+\begin{bmatrix}100\\2\end{bmatrix}+\begin{bmatrix}100\\3\end{bmatrix}=166\,751$ 种，若用二进制码元构成定长码，则需最小长度为 $\log_2 166\,751=17.347\text{bit}$，取整后定长码所需最小长度为 18bit。

(2) 由题意，只对含有 3 个或小于 3 个“0”的信源序列编出了一一对应的一组定长码。因为是定长码，所以这些序列一一对应编码后，对应于正确概率。因此，剩下没有给予编码的信源序列出现的概率，是该定长码引起的错误概率。

$$P_e=1-\begin{bmatrix}100\\0\end{bmatrix}\times 0.995^{100}-\begin{bmatrix}100\\1\end{bmatrix}\times 0.995^{99}\times 0.005-\begin{bmatrix}100\\2\end{bmatrix}\times 0.995^{98}\times 0.005^2-\begin{bmatrix}100\\3\end{bmatrix}\times 0.995^{97}\times 0.005^3=0.0016$$

注：本题考查的知识点是 5.2.1 节定长编码。题解：(1)计算所需最小定长编码的长度；(2)错误概率＝1－正确概率。

9. 已知符号集合$\{x_1,x_2,x_3,\cdots,x_i,\cdots\}$为无限离散消息集合，它们的出现概率分别为 $p(x_1)=\dfrac{1}{2}$，$p(x_2)=\dfrac{1}{2^2}$，$p(x_3)=\dfrac{1}{2^3}$，$\cdots$，$p(x_i)=\dfrac{1}{2^i}$，$\cdots$。试：(1)用香农编码方法编出各个符号消息的码字；(2)计算码字的平均信息传输速率；(3)计算信源编码效率。

解答：(1) 由香农编码方法可得各个符号消息的码字如表 5.8 所示。

表 5.8 题 9 的香农编码

信源符号	符号概率	累加概率 P_i	$-\log p(x_i)$	码字长度	码字
x_1	1/2	0	1	1	1
x_2	1/4	0.5	2	2	01
x_3	1/8	0.75	3	3	001
⋮	⋮	⋮	⋮	⋮	⋮
x_i	$1/2^i$	$1-1/2^{i-1}$	i	i	0000…1

所以，码字为 1，01，001，0001，…，0…01($i-1$ 个“0”和 1 个“1”)，…

(2) 码字的平均信息传输速率

$$\overline{K}=\sum_i p(x_i)K_i=\frac{1}{2}+\frac{2}{4}+\frac{3}{8}+\cdots+\frac{i}{2^i}+\cdots=2 \text{ 二进制码元/符号}$$

(3) 由于 $H(X)=-\sum_i p(x_i)\log p(x_i)=2\text{bit/符号}$，所以信源编码效率 $\eta=\dfrac{H(X)}{\overline{K}}=\dfrac{2}{2}=1.00$。

注：本题考查的知识点是 5.2.2 节变长编码。题解：(1)应用香农编码的方法进行编码；(2)计算其编码效率。

10. 若某一信源有 N 个符号，并且每个符号均以等概率出现，对此信源用最佳哈夫曼二元编码，问当 $N=2^i$ 和 $N=2^i+1$(i 为正整数)时，每个码字的长度等于多少？平均码长是多少？

解答：当信源具有 $N=2^i$ 个符号时，每个符号等概率，采用哈夫曼二元编码时，为每个符号编出的码字长度相等，且均为 i 个二进制码元。此时，平均码长即为 i 二进制码元/符号。

当信源具有 $N=2^i+1$ 个符号时，其中 2^i-1 个符号的码字长度为 i 二进制码元，2 个符号的码字长度为($i+1$)二进制码元，平均码长为 $\left(i+\dfrac{2}{2^i+1}\right)$ 二进制码元/符号。

注：本题考查的知识点是 5.4.1 节哈夫曼编码。

11. 设有离散无记忆信源 $P(X)=\{0.37,0.25,0.18,0.10,0.07,0.03\}$。(1)求该信源符号的熵 $H(X)$。(2)用哈夫曼编码编成二元变长码，计算其编码效率。(3)要求译码错误小于 10^{-3}，采用定长二元码要达到(2)中哈夫曼编码的效率，问需要多少个信源符号一起编码？

解答：(1) 该信源符号熵 $H(X)=H(0.37,0.25,0.18,0.10,0.07,0.03)=2.23\text{bit/符号}$。

(2) 哈夫曼编码编成的二元变长码如表 5.9 所示。

平均码长 $\overline{K}=\sum_{i=1}^{6} p(x_i)K_i=2.3$ 二进制码元/符号

编码效率 $\eta=\dfrac{H(X)}{\overline{K}}=\dfrac{2.23}{2.3}=0.97$

表 5.9　题 11 的哈夫曼编码

符　　号	概　　率	二进制码字
x_1	0.37	00
x_2	0.25	01
x_3	0.18	11
x_4	0.01	100
x_5	0.07	1010
x_6	0.03	1011

(3) 要求译码错误小于 10^{-3}，采用定长二元码，此时

$$\sigma^2(x)=E\{[I(x_i)-H(X)]^2\}=0.7916$$

由编码效率 $\eta=\dfrac{H(X)}{H(X)+\varepsilon}=0.97$ 求得 $\varepsilon=0.069$，由 $L\geqslant\dfrac{\sigma^2(x)}{\varepsilon^2\delta}$ 可得 $L\geqslant1.66\times10^5$。

注：本题考查的知识点是 5.2.1 节定长编码、5.4.1 节哈夫曼编码。题解：(1)应用哈夫曼编码的方法进行编码，并计算编码效率；(2)比较定长编码和变长编码的效率。

12. 信源符号 X 有 6 种字母，概率分布为{0.32，0.22，0.18，0.16，0.08，0.04}。试：(1)求符号熵 $H(X)$。(2)用香农编码编成二进制变长码，计算其编码效率。(3)用哈夫曼编码编成二进制变长码，计算其编码效率。(4)用哈夫曼编码编成三进制变长码，计算其编码效率。(5)若用逐个信源符号来编定长二进制码，要求不出差错译码，求所需要的每个符号的平均信息传输率和编码效率。(6)当译码差错小于 10^{-3} 的定长二进制码要达到(3)中哈夫曼编码的效率时，估计要多少个信源符号一起编才能办到？

解答：(1) 符号熵 $H(X)=H(0.32,0.22,0.18,0.16,0.08,0.04)=2.35$bit/符号。

(2) 香农编码编成二进制变长码，如表 5.10 所示。

表 5.10　题 12 的香农编码

符号概率 $p(x_i)$	累加概率 P_i	$-\log p(x_i)$	码字长度 K_i	码　　字
0.32	0	1.644	2	00
0.22	0.32	2.184	3	010
0.18	0.54	2.474	3	100
0.16	0.72	2.644	3	101
0.08	0.88	3.644	4	1110
0.04	0.96	4.644	5	11110

平均码长 $\overline{K_1}=\sum_{i=1}^{6}p(x_i)K_i=2.84$ 二进制码元/符号

编码效率 $\eta_1=\dfrac{H(X)}{\overline{K_1}}=\dfrac{2.35}{2.84}=0.83$

(3) 哈夫曼编码编成二进制变长码，编出的码字如表 5.11 所示。

表 5.11 题 11 的二进制哈夫曼编码

符　　号	概　　率	码字(二进制)
x_1	0.32	10
x_2	0.22	00
x_3	0.18	01
x_4	0.16	110
x_5	0.08	1110
x_6	0.04	1111

平均码长 $\overline{K_2}=2.4$ 二进制码元/符号,则编码效率 $\eta_2=\dfrac{H(X)}{\overline{K}_2}=0.98$。

(4) 哈夫曼编码编成三进制变长码,编出的码字如表 5.12 所示。

表 5.12 题 11 的三进制哈夫曼编码

符　　号	概　　率	码字(三进制)
x_1	0.32	1
x_2	0.22	2
x_3	0.18	00
x_4	0.16	01
x_5	0.08	021
x_6	0.04	022

平均码长 $\overline{K_3}=\sum_{i=1}^{6}p(x_i)K_i=1.58$ 三进制码元/符号

编码效率 $\eta_3=\dfrac{H(X)}{\overline{K_3}\log_2 3}=0.94$

(5) 若用逐个信源符号来编定长二进制码,且要求不出差错译码,所需要的每个符号的平均信息传输率为 3bit/符号,此时编码效率为 78.3%。

(6) 当译码差错小于 10^{-3} 的定长二进制码要达到(3)中哈夫曼编码的效率时,此时方差 $\sigma^2(x)=E\{[I(x_i)-H(X)]^2\}=0.535$,由 $\eta=\dfrac{H(X)}{H(X)+\varepsilon}=0.98$,得 $\varepsilon=0.05$,则 $L\geqslant\dfrac{\sigma^2(x)}{\varepsilon^2\delta}=2.1\times10^5$,即需要 210 000 个符号一起编码才能达到 98%的编码效率。

注:本题考查的知识点是 5.2.1 节定长编码、5.2.2 节变长编码、5.4.1 节哈夫曼编码。题解:(1)熟悉编码方法及其编码效率的计算,注意多进制编码的效率计算;(2)比较定长编码和变长编码。

13. 有二元独立序列,已知 $p_0=0.9$,$p_1=0.1$,求这序列的符号熵。当用哈夫曼编码时,以三个二元符号合成一个新符号,求这种符号的平均代码长度和编码效率。设输入二元符号的速率为每秒 100 个,要求 3min 内溢出和取空的概率均小于 0.01,求所需的信道码率(bit/s)和存储器容量(比特数)。若信道码率已规定为 50bit/s,存储器容量将如何选择?

解答:符号熵 $H(X)=H(0.9,0.1)=0.47$bit/符号。

三个二元符号合成一个新符号，记为 U，各符号的概率（按从大到小的顺序排列）和哈夫曼编码的结果如表 5.13 所示。

表 5.13　题 13 的新符号概率和哈夫曼编码

新　符　号	三个二元符号	概　　率	哈夫曼编码
u_1	000	0.729	0
u_2	001	0.081	100
u_3	010	0.081	101
u_4	100	0.081	110
u_5	011	0.009	11100
u_6	101	0.009	11101
u_7	110	0.009	11110
u_8	111	0.001	11111

新符号 U 经哈夫曼编码后的平均码长

$$\overline{K}_U = 0.729\times 1 + 0.081\times 3\times 3 + 0.009\times 5\times 3 + 0.001\times 5 = 1.598\text{bit/ 三个二元符号}$$

则平均到每个二元符号的平均码长 $\overline{K} = \dfrac{1.598}{3} = 0.533\text{bit/符号}$，编码效率 $\eta = \dfrac{H(X)}{\overline{K}} = \dfrac{0.47}{0.533} = 0.88$。

若输入二元符号的速率为每秒 100 个，则所需信道码率为 53.3bit/s。

若要求 3min 内存储器溢出和取空的概率均小于 0.01，查误差函数表，可得 $A = 2.332$，存储器容量应为 $M > 2A\sqrt{N\sigma^2}$，式中，N 为 3min 内信源符号输出个数；σ^2 为码长的方差。

$$N = \frac{100}{3}\times 60\times 3 = 6000$$

由编出的哈夫曼编码可得

$$\begin{aligned}\sigma^2 &= E[K_i^2] = \sum_{j=1}^{m} p(u_j)K_j^2 - \overline{K}_U^2 \\ &= (0.729\times 1^2 + 0.081\times 3^2\times 3 + 0.009\times 5^2\times 3 + 0.001\times 5^2) - 1.598^2 \\ &= 1.06\end{aligned}$$

则存储器容量 $M = 2\times 2.332\times\sqrt{6000\times 1.06} = 372\text{bit}$

存储器半满为 186bit。若信道码率为 50bit/s，小于输入信道符号的速率，则 3min 后存储器中会增加 594bit，因而开始时存储器不应到半满，存储器的容量可略小于 (372＋594) bit＝966bit。

注：本题考查的知识点是 5.4.1 节哈夫曼编码。题解：(1)哈夫曼编码方法；(2)变长码编码简单的代价是需要存储设备来缓冲码字长度的差异。

14. 若已知一离散信源符号集为：(1) 若 $\begin{bmatrix} U \\ P \end{bmatrix} = \begin{bmatrix} a & b & c & d \\ 0.5 & 0.3 & 0.15 & 0.05 \end{bmatrix}$，试对它进行

哈夫曼编码并求编码效率？（2）若该信源的状态转移概率矩阵 $\boldsymbol{P}=\begin{bmatrix}0.3 & 0.4 & 0.3 & 0\\ 0.5 & 0.5 & 0 & 0\\ 0.5 & 0 & 0 & 0.5\\ 0 & 0.5 & 0.5 & 0\end{bmatrix}$，求稳态时符号的出现概率并求信源熵；（3）试设计一个哈夫曼编码，每次对两个信源符号编码，求每个信源符号平均比特数以及编码效率。

解答：（1）哈夫曼编码结果如表 5.14 所示。

表 5.14　题 14 的哈夫曼编码

符　号	概　率	码　字
u_1	0.5	1
u_2	0.3	00
u_3	0.15	010
u_4	0.05	011

信源熵 $H(U)=H(0.5,0.3,0.15,0.05)=1.65\text{bit/符号}$

平均码长 $\overline{K}=0.5\times1+0.3\times2+0.15\times3+0.05\times3=1.7\text{bit/符号}$

编码效率 $\eta=\dfrac{H(U)}{\overline{K}}=\dfrac{1.65}{1.7}=0.97$

（2）该信源的状态转移矩阵 $\boldsymbol{P}=\begin{bmatrix}0.3 & 0.4 & 0.3 & 0\\ 0.5 & 0.5 & 0 & 0\\ 0.5 & 0 & 0 & 0.5\\ 0 & 0.5 & 0.5 & 0\end{bmatrix}$，设状态的稳定分布 $\boldsymbol{W}=(w_1,w_2,w_3,w_4)$，则求解方程组 $\begin{cases}\boldsymbol{W}\cdot\boldsymbol{P}=\boldsymbol{W}\\ \sum\limits_{i=1}^{4}w_i=1,\end{cases}$ 可得状态的稳定概率为 $w_1=w_2=\dfrac{5}{13},w_3=\dfrac{2}{13},w_4=\dfrac{1}{13}$。

平稳时符号的出现概率为 $p_1=\dfrac{5}{13},p_2=\dfrac{5}{13},p_3=\dfrac{2}{13},p_4=\dfrac{1}{13}$。

该信源的信息熵

$$H(U)=H\left(\frac{5}{13},\frac{5}{13},\frac{2}{13},\frac{1}{13}\right)=1.76\text{bit/符号}$$

（3）对两个符号一起进行哈夫曼编码后的结果如表 5.15 所示。

表 5.15　题 14 两个符号一起的哈夫曼编码结果

状　态	信源消息	概　率	编　码
s_1	u_1u_1	$0.5\times0.5=0.25$	10
s_2	u_1u_2	$0.5\times0.3=0.15$	001
s_3	u_2u_1	$0.3\times0.5=0.15$	010
s_4	u_2u_2	$0.3\times0.3=0.09$	111

续表

状　态	信源消息	概　率	编　码
s_5	u_1u_3	$0.5\times0.15=0.075$	0000
s_6	u_3u_1	$0.15\times0.5=0.075$	0001
s_7	u_2u_3	$0.3\times0.15=0.045$	1100
s_8	u_3u_2	$0.15\times0.3=0.045$	1101
s_9	u_1u_4	$0.5\times0.05=0.025$	01110
s_{10}	u_4u_1	$0.05\times0.5=0.025$	01111
s_{11}	u_3u_3	$0.15\times0.15=0.0225$	011000
s_{12}	u_2u_4	$0.3\times0.05=0.015$	011010
s_{13}	u_4u_2	$0.05\times0.3=0.015$	011011
s_{14}	u_3u_4	$0.15\times0.05=0.075$	0110011
s_{15}	u_4u_3	$0.05\times0.15=0.075$	01100100
s_{16}	u_4u_4	$0.05\times0.05=0.0025$	01100101

$$\bar{L}=2\times0.25+3\times0.15+3\times0.09+4\times(0.075+0.075+0.045+0.045)+5\times(0.025+0.025)+6\times(0.0225+0.015+0.015)+7\times0.0075+8\times(0.0075+0.0025)=3.3275\text{bit/ 二个二元符号}$$

所以 $\bar{K}=\dfrac{\bar{L}}{2}=1.66375\text{bit/符号}$，编码效率 $\eta=\dfrac{H(U)}{\bar{K}}=\dfrac{1.6474}{1.66375}=0.99$。

注：本题考查的知识点是 5.4.1 节哈夫曼编码。题解：(1)哈夫曼编码方法进行编码；(2)两个信源符号的哈夫曼编码。

15. 设有一阶马尔可夫信源，信源集合 $U=\begin{pmatrix}u_j\\q_j\end{pmatrix}$，$j=1,2,3$，状态集合 $S=\begin{pmatrix}s_i\\p_i\end{pmatrix}$，$i=1,2,3$，其状态转移概率矩阵 $\boldsymbol{P}=\begin{bmatrix}0&\frac{3}{4}&\frac{1}{4}\\0&\frac{1}{2}&\frac{1}{2}\\1&0&0\end{bmatrix}$，试求：(1) $H(U\mid S_1)$、$H(U\mid S_2)$、$H(U\mid S_3)$；(2)对各种状态，分别将 U 的符号编成最佳变长哈夫曼二进码；(3)求 $H_\infty(U)$，并证明上述编码方法的平均码长等于 $H_\infty(U)$。

解答：(1) 该马尔可夫信源的状态稳定概率为可由方程组 $\begin{cases}p_1+p_2+p_3=1\\p_1=p_3\\p_2=\frac{3}{4}p_1+\frac{1}{2}p_2\\p_3=\frac{1}{4}p_1+\frac{1}{2}p_2\end{cases}$ 求得，$p_1=\dfrac{2}{7}$，$p_2=\dfrac{3}{7}$，$p_3=\dfrac{2}{7}$，即 $P(S=s_1)=\dfrac{2}{7}$，$P(S=s_2)=\dfrac{3}{7}$，$P(S=s_3)=\dfrac{2}{7}$。

$$H(U\mid S_j)=-\sum_{i=1}^{3}p(U=u_i,S=s_j)\log p(U=u_i,S=s_j)$$

$$=-\sum_{i=1}^{3}p(U=u_i\mid S=s_j)p(S=s_j)\log p(U=u_i,S=s_j)$$

由状态转移矩阵可得

$$p(U=u_1\mid S=s_1)=\frac{1}{2},p(U=u_2\mid S=s_1)=\frac{1}{4},p(U=u_3\mid S=s_1)=\frac{1}{4}$$

$$p(U=u_1\mid S=s_2)=0,p(U=u_2\mid S=s_2)=\frac{1}{2},p(U=u_3\mid S=s_2)=\frac{1}{2}$$

$$p(U=u_1\mid S=s_3)=1,p(U=u_2\mid S=s_3)=0,p(U=u_3\mid S=s_3)=0$$

所以，

$$H(U\mid S_1)=-\frac{1}{2}\times\frac{2}{7}\times\log\frac{1}{2}-\frac{1}{4}\times\frac{2}{7}\times\log\frac{1}{4}-\frac{1}{4}\times\frac{2}{7}\times\log\frac{1}{4}=0.428\,57$$

$$H(U\mid S_2)=0-\frac{1}{2}\times\frac{3}{7}\times\log\frac{1}{2}-\frac{1}{2}\times\frac{3}{7}\times\log\frac{1}{2}=0.428\,57$$

$$H(U\mid S_3)=-1\times\frac{2}{7}\times\log 1=0$$

(2) 各种状态的信源概率分布及相应的哈夫曼编码结果如表 5.16 所示。其中，

$$P(U=u_1)=\sum_{i=1}^{3}P(S=s_i)P(U=u_1\mid S=s_i)=\frac{2}{7}\times\frac{1}{2}+\frac{3}{7}\times 0+\frac{2}{7}\times 1=\frac{3}{7}$$

$$P(U=u_2)=\sum_{i=1}^{3}P(S=s_i)P(U=u_2\mid S=s_i)=\frac{2}{7}\times\frac{1}{4}+\frac{3}{7}\times\frac{1}{2}+\frac{2}{7}\times 0=\frac{2}{7}$$

$$P(U=u_3)=\sum_{i=1}^{3}P(S=s_i)P(U=u_3\mid S=s_i)=\frac{2}{7}\times\frac{1}{4}+\frac{3}{7}\times\frac{1}{2}+\frac{2}{7}\times 0=\frac{2}{7}$$

表 5.16　题 15 的哈夫曼编码

符　　号	概　　率	编出的码字
u_1	3/7	1
u_2	2/7	00
u_3	2/7	01

(3) 极限熵 $H_\infty(U)=-\left(\frac{3}{7}\log\frac{3}{7}+\frac{2}{7}\log\frac{2}{7}+\frac{2}{7}\log\frac{2}{7}\right)=1.56\text{bit/符号}$

平均码长 $\overline{K}=\frac{3}{7}\times 1+\frac{2}{7}\times 2+\frac{2}{7}\times 2=\frac{11}{7}=1.57\text{bit/符号}$

编码效率 $\eta=\frac{H_\infty(U)}{\overline{L}}=\frac{1.56}{1.57}=99\%$

可以看出，平均码长逼近极限熵 $H_\infty(U)$。

注：本题考查的知识点是 2.1.3 节马尔可夫信源、5.4.1 节哈夫曼编码。题解：(1)这是一道综合题，结合了马尔可夫信源和哈夫曼编码；(2)计算马尔可夫信源的稳态概率；(3)采用哈夫曼编码方法进行编码。

16. 离散无记忆信源发出 A、B、C 三种符号，概率分布分别为 5/9、1/3、1/9，应用算术编

码方法对序列 CABA 进行编码，并对结果进行解码。

解答：先计算三种符号的累积概率，写成二进制后，结果如表 5.17 所示。其中，小写的 p 是符号概率，大写的 P 是累积概率。

表 5.17　题 16 的累积概率

符　　号	符号概率 p	符号概率的二进制表示	符号累积概率 P（二进制表示）
A	5/9	0.1	0
B	1/3	0.01	0.1
C	1/9	0.01	0.11

按照主教材 5.4.2 节中关于算术编码的过程描述进行编码，用累积概率 $P(S)$ 表示码字 $C(S)$，符号概率 $p(S)$ 表示状态区间 $A(S)$，用 φ 表示起始状态为空序列。具体过程如下：

假设起始状态 $A(\varphi)=1, C(\varphi)=0$；

“C”：$C(\varphi\mathrm{C})=C(\varphi)+A(\varphi)P_{\mathrm{C}}=0+1\times0.11=0.11$

$A(\varphi\mathrm{C})=A(\varphi)p_{\mathrm{C}}=1\times0.01=0.01$

“CA”：$C(\mathrm{CA})=C(\mathrm{C})+A(\mathrm{C})P_{\mathrm{A}}=0.11+0.01\times0=0.11$

$A(\mathrm{CA})=A(\mathrm{C})p_{\mathrm{A}}=0.01\times0.1=0.001$

“CAB”：$C(\mathrm{CAB})=C(\mathrm{CA})+A(\mathrm{CA})P_{\mathrm{B}}=0.11+0.001\times0.1=0.1101$

$A(\mathrm{CAB})=A(\mathrm{CA})p_{\mathrm{B}}=0.001\times0.01=0.000\,01$

“CABA”：$C(\mathrm{CABA})=C(\mathrm{CAB})+A(\mathrm{CAB})P_{\mathrm{A}}=0.1101+0.000\,01\times0=0.1101$

$A(\mathrm{CABA})=A(\mathrm{CAB})p_{\mathrm{A}}=0.000\,01\times0.1=0.000\,001$

所以，编出的码字为“1101”。

解码过程：

(1) 0.1101 位于区间[0.11,1]，所以第一个码字是“C”；

(2) 减去累积概率 P_{C}，0.1101−0.11=0.0001，乘上加权 0.0001×100=0.01，位于区间[0,0.1]，所以第 2 个码字是“A”；

(3) 0.01×10=0.1，位于区间[0.1,0.11]，所以第三个码字是“B”；

(4) 0.1−0.1=0，位于区间[0,0.1]，所以第 4 个码字是“A”。

所以，解码的码序列是“CABA”。

注：本题考查的知识点是 5.4.2 节算术编码。题解：算术码的编码和解码。

17. 一离散无记忆信源{0,1}，符号出现概率分别为：$p_0=0.3, p_1=0.7$。已知信源序列为 1101110011…，试：(1)对此序列进行算术编码；(2)若将 0,1 符号的概率近似取为 $p_0=0.25, p_1=0.75$，进行算术编码。

解答：(1) 将概率化为二进制得 $p_1=0.101, p_0=0.011$，累积概率为 $P_1=0.000, P_0=0.101$。设起始状态为空序列 φ，令 $A(\varphi)=1, C(\varphi)=0$。由主教材上关于算术编码的过程描述进行编码，具体过程如下：

“1”：$C(\varphi1)=C(\varphi)+A(\varphi)P_1=0+1\times0=0$

$A(\varphi1)=A(\varphi)p_1=1\times0.101=0.101$

“11”：$C(11)=C(1)+A(1)P_1=0+0.101\times0=0$

$$A(11)=A(1)p_1=0.101\times0.101=0.011001$$

“110”：$C(110)=C(11)+A(11)P_0=0+0.011001\times0.101=0.001111101$

$$A(110)=A(11)p_0=0.011001\times0.011=0.001001011$$

“1101”：$C(1101)=C(110)+A(110)P_1=0.001111101$

$$A(1101)=A(110)p_0=0.000101110111$$

“11011”：$C(11011)=C(1101)+A(1101)P_1=0.001111101$

$$A(11011)=A(1101)p_1=0.000011101010011$$

“110111”：$C(110111)=C(11011)+A(11011)P_1=0.001111101$

$$A(110111)=A(11011)p_1=0.000010010010011111$$

后面的几位可以以此推算出来。

(2) 将概率化为二进制为 $p_1=0.11$，$p_0=0.01$，累积概率为 $P_1=0.00$，$P_0=0.11$。

“1”：$C(\varphi1)=C(\varphi)+A(\varphi)P_1=0+1\times0=0$

$$A(\varphi1)=A(\varphi)p_1=1\times0.11=0.11$$

“11”：$C(11)=C(1)+A(1)P_1=0$

$$A(11)=A(1)p_1=0.11\times0.11=0.1001$$

“110”：$C(110)=C(11)+A(11)P_0=0+0.1001\times0.11=0.011011$

$$A(110)=A(11)p_0=0.1001\times0.01=0.001001$$

“1101”：$C(1101)=C(110)+A(110)P_1=0.011011$

$$A(1101)=A(110)p_0=0.00011011$$

“11011”：$C(11011)=C(1101)+A(1101)P_1=0.011011$

$$A(11011)=A(1101)p_1=0.0001010001$$

“110111”：$C(110111)=C(11011)+A(11011)P_1=0.011011$

$$A(110111)=A(11011)p_1=0.000011110011$$

以此类推，后面各位可以计算出来。

注：本题考查的知识点是 5.4.2 节算术编码。题解：(1)符号概率和累积概率的意义；(2)算术编码的编解码过程。

18. 对题 17 中的第二种编码结果，利用比较法写出算术码的译码过程。

解答：算术码的译码可通过比较编码后的数值大小来进行，即判断码字落在哪个区间就可以得出一个相应的符号序列。

因为 $C(110111)=0.0011011<0.11$，所以译出第一个符号为“1”；

去掉被乘概率因子(加权)：$0.0011011\times10=0.011011<0.1$，译出第二个符号为“1”；

再去掉被乘概率因子(加权)：$0.0011011\times100=0.11011$，在区间$[0.11,1]$内，所以译出第三个符号是“0”；

去掉累积概率：即 $0.11011-0.11=0.00011$，再去掉被乘概率因子(加权)：$0.00011\times1000=0.11$，属于区间$[0.11,1]$，所以译出第四个符号为“0”；

去掉累积概率：$0.11-0.11=0.0$，$0.0\times8=0.0<0.11$，所以译出第五个符号为“1”；

同理，译得第六、七个符号为 11；

所以，译出的序列 $S^*=1100111=S$。

注：本题考查的知识点是 5.4.2 节算术编码。题解：(1)算术码的解码过程；(2)算术

码的译码所需参数较少，且不像哈夫曼编码那样需要一个很大的码表。

19. 一阶马尔可夫信源符号集 $X\in\{a_1,a_2,a_3\}$，状态集 $S\in\{s_1,s_2,s_3\}$，状态转移如图 5.4 所示，某状态下发出符号的概率标在相应的线段旁。试：(1)求 $H(X|s_1)$、$H(X|s_2)$和 $H(X|s_3)$；(2)对各种状态，分别把 X 的符号编成最佳变长二进制码；(3)求 $H(X)$。

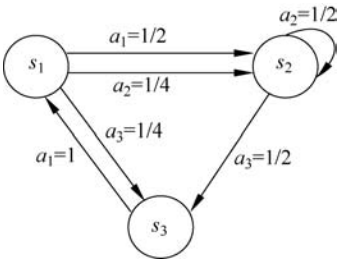

图 5.4　题 19 的状态转移图

解答：(1) 由图 5.1 可得状态转移矩阵 $\boldsymbol{P}=[P(s_j|s_i)]=\begin{bmatrix}0 & \frac{3}{4} & \frac{1}{4}\\ 0 & \frac{1}{2} & \frac{1}{2}\\ 1 & 0 & 0\end{bmatrix}$，符号条件概率矩阵 $[P(a_j|s_i)]=\begin{bmatrix}\frac{1}{2} & \frac{1}{4} & \frac{1}{4}\\ 0 & \frac{1}{2} & \frac{1}{2}\\ 1 & 0 & 0\end{bmatrix}$。

各状态下的条件熵为

$$H(X|s_1)=\sum_i p(a_i|s_1)\log p(a_i|s_1)=\frac{3}{2}\text{bit/符号}$$

$$H(X|s_2)=\sum_i p(a_i|s_2)\log p(a_i|s_2)=1\text{bit/符号}$$

$$H(X|s_3)=\sum_i p(a_i|s_3)\log p(a_i|s_3)=0\text{bit/符号}$$

(2) 用哈夫曼编码对各种状态进行编码。对状态 s_1 编出的哈夫曼编码如表 5.18 所示。

表 5.18　题 19 的哈夫曼编码

符　号	概　率	编出的码字
a_1	1/2	0
a_2	1/4	10
a_3	1/4	11

对状态 s_2 编出的哈夫曼编码如表 5.19 所示。

表 5.19　题 19 的哈夫曼编码

符　号	概　率	编出的码字
a_1	1/2	0
a_2	1/2	1
a_3	0	不编码

对状态 s_3 编出的哈夫曼编码如表 5.20 所示。

表 5.20　题 19 的哈夫曼编码

符　号	概　率	编出的码字
a_1	1	0
a_2	0	不编码
a_3	0	不编码

(3) 设各状态的稳定概率满足方程组 $\begin{cases} p(s_1)=p(s_3) \\ p(s_2)=\frac{3}{4}p(s_1)+\frac{1}{2}p(s_2) \\ p(s_3)=\frac{1}{4}p(s_1)+\frac{1}{2}p(s_2) \\ \sum_{i=1}^{3}p(s_i)=1 \end{cases}$，求解该方程组可得 $p(s_1)=p(s_3)=\frac{2}{7}$，$p(s_2)=\frac{3}{7}$，则信源的极限熵

$$H_\infty(X)=\sum_i p(s_i)H(X\mid s_i)=\frac{2}{7}\times\frac{3}{2}+\frac{3}{7}\times 1+\frac{2}{7}\times 0=\frac{6}{7}\text{bit/符号}$$

注：本题考查的知识点是2.1.3节马尔可夫信源、2.3.2节离散有记忆信源的序列熵以及5.4.1节哈夫曼编码，是一道综合题。题解：(1)由状态转移图得到状态转移矩阵和符号转移矩阵；(2)马尔可夫信源的极限熵计算方法；(3)用哈夫曼编码方法进行编码。

20. 在电视信号中，亮度信号的黑色电平为0，白色电平为L。用均匀分割来量化其样值，要求峰功率信扰比大于50dB，求每样值所需的量化比特数。

解答：设电视信号的电平在[0，L]区间均匀分布，即 $p(x)=\frac{1}{L}$，$0<x<L$。

当量化级数为n，量化级差为 $\Delta=\frac{L}{n}$，可将$(0,L)$均匀分割成n个小区间，每个小区间的中点作为量化值，即量化值 $y_i=\frac{(2i-1)L}{2n}$。

当失真函数是均方失真函数时，量化的平均失真

$$D=\sum_{i=1}^{n}\int_{\frac{L(i-1)}{n}}^{\frac{Li}{n}}[x-y_i]^2p(x)\mathrm{d}x=\sum_{i=1}^{n}\int_{\frac{L(i-1)}{n}}^{\frac{Li}{n}}\left[x-\frac{L(2i-1)}{2n}\right]^2\frac{\mathrm{d}x}{L}=\frac{1}{12}\left(\frac{L}{n}\right)^2=\frac{\Delta^2}{12}$$

由峰功率信扰比$=\frac{\text{信号峰功率}}{\text{量化噪声功率}}$，信号峰功率等于$L^2$，量化噪声功率 $W=\sum_{i=1}^{n}Dp_i=\frac{\Delta^2}{12}\int_0^L p(x)\mathrm{d}x=\frac{\Delta^2}{12}$，即要求 $10\log_{10}\frac{L^2}{W}>50$。将 $W=\frac{\Delta^2}{12}$和 $\Delta=\frac{L}{n}$代入，得 $\log_{10}12n^2>5$。

设量化比特数为b，则 $2^b=n$，要求 $\log_{10}12\times 2^{2b}>5$，可得 $b>6.5$。所以，每样值所需的量化比特数为7bit。

注：本题考查的知识点是5.4.5节矢量量化编码。题解：由峰功率信扰比＝信号峰功率/量化噪声功率，计算得到量化级数，而后得到量化比特数。

21. 设信源为 $\begin{bmatrix} S \\ p(s) \end{bmatrix}=\begin{bmatrix} s_1 & s_2 & \cdots & s_6 \\ p_1 & p_2 & \cdots & p_6 \end{bmatrix}$，将此信源编码作为$r$元唯一可译变长码，要求对应的码长为$(K_1,K_2,K_3,K_4,K_5,K_6)=(1,1,2,3,2,3)$，求$r$的最佳下限值。

解答：要将此信源编码成为r元的唯一可译变长码，其码字对应的码长$(K_1,K_2,K_3,K_4,K_5,K_6)=(1,1,2,3,2,3)$必须满足克拉夫特不等式：

$$\sum_{i=1}^{6}r^{-K_i}=r^{-1}+r^{-1}+r^{-2}+r^{-3}+r^{-2}+r^{-3}\leqslant 1$$

即$\frac{2}{r}+\frac{2}{r^2}+\frac{2}{r^3}\leqslant 1$，其中，$r$ 是大于或等于 1 的正整数。

当 $r=1$ 和 2 时，$\sum_{i=1}^{6} r^{-K_i} > 1$，不能满足克拉夫特不等式。

当 $r=3$ 时，$\sum_{i=1}^{6} r^{-K_i} = \frac{26}{27} < 1$，能满足克拉夫特不等式。

所以，r 的最佳下限值为 $r=3$。

注：本题考查的知识点是 5.1 节编码的概念。题解：满足克拉夫特不等式是唯一可译码的必要条件。

22. 某通信系统使用文字字符共 10000 个，据长期统计，使用频率占 80％的共有 500 个，占 90％的有 1000 个，占 99％的有 4000 个，占 99.9％的 7000 个。试：(1)求该系统使用的文字字符的熵；(2)请给出该系统的一种信源编码方法并作简要评价。

解答：(1) 10000 个文字字符中，使用频率占 80％的有 500 个，则这 500 个文字字符共出现了 8000 次，平均每个字符出现了 $\frac{8000}{500}=16$ 次，即这 500 个字符每个字符出现的概率 $p_1=\frac{16}{10000}=1.6\times10^{-4}$。

同理，使用频率 90％的有 1000 个，减去使用频率 80％的 500 个，有 500 个字符出现频率为 10％，在 10000 个字符中出现的次数为 1000 次，平均每个字符出现了 $\frac{1000}{500}=2$ 次，这 500 个字符中每个字符出现的概率 $p_2=\frac{2}{10000}=2\times10^{-4}$。

类似地，可以得到余下 90％～99％的 3000 个字符每个字符出现的概率 $p_3=\frac{900}{3000\times10000}=3\times10^{-5}$；99％～99.9％的 3000 个字符每个字符出现的概率 $p_4=\frac{90}{3000\times10000}=3\times10^{-6}$；99.9％～100％的 3000 个字符每个字符出现的概率 $p_5=\frac{10}{3000\times10000}=\frac{1}{3}\times10^{-6}$。

根据熵的定义，这 10000 个文字字符的熵

$$H(\boldsymbol{X})=-500(p_1\log_2 p_1+p_2\log_2 p_2)-3000(p_3\log_2 p_3+p_4\log_2 p_4+p_5\log_2 p_5)$$

$$=0.8\log\frac{0.8}{500}-0.1\log\frac{0.1}{500}-0.09\log\frac{0.09}{3000}-0.009\log\frac{0.009}{3000}-$$

$$0.001\log\frac{0.001}{3000}=10.2\text{bit/字符}$$

(2) 可以使用哈夫曼编码的方法对该系统的信源进行编码，为使压缩效果理想，可以使用扩展信源的方法。

注：本题考查的知识点是 2.2 节离散信源熵和互信息、5.4.1 节哈夫曼编码。题解：(1)本题不需要考虑字符间的关系，只需紧扣熵的定义计算文字字符熵；(2)有了统计概率，哈夫曼编码是较好的选择。

23. 信源符号 X 的概率空间为 $\begin{bmatrix}X\\P\end{bmatrix}=\begin{bmatrix}x_1 & x_2\\0.1 & 0.9\end{bmatrix}$，如每次两个符号一起编码，试写出哈夫曼编码，并求平均码长 $\overline{K}$ 和编码效率。

解答：哈夫曼编码的参考编码结果为

0,11,100,101(哈夫曼编码并不唯一)

平均码长 $\overline{K}=0.81+0.09\times2+0.1\times3=1.29$bit/两个符号$=0.645$bit/符号。

编码效率 $\eta=\dfrac{H(X)}{\overline{K}}=\dfrac{H(0.1,0.9)}{0.645}=\dfrac{0.47}{0.645}=0.73$。

注：本题考查的知识点是 5.4.1 节哈夫曼编码。题解：(1)得到 X 的二次扩展信源；(2)利用哈夫曼编码方法进行编码。

24. 一通信系统传送的符号只有 3 个，使用概率分别为 0.2、0.3 和 0.5，但传送时总是以 3 个符号为一个字，故该系统的信源编码以字为基础并采用二进制哈夫曼编码。根据字的概率大小，编码结果为：概率在(0,0.020]区间，采用 6bit；在(0.020,0.045]区间，采用 5bit；在(0.045,0.100]区间，采用 4bit；在 0.100 以上的区间，采用 3bit。求该种信源编码的效率。

解答：假设该系统传输的三个符号分别为 a,b 和 c，且 $p_a=0.2$，$p_b=0.3$，$p_c=0.5$。下面对每个字，即 3 个符号可能出现的情况加以讨论，结果如表 5.21 所示。

表 5.21　题 24 的符号和编码

字(3 个符号)	概　　率	概率所在区间	编 码 位 数	种　　数
aaa	$0.2^3=0.008$	(0,0.020]	6	1
bbb	$0.3^3=0.027$	(0.020,0.045]	5	1
ccc	$0.5^3=0.125$	(0.100,1.00]	3	1
2 个 a、1 个 b	$0.2^2\times0.3=0.012$	(0,0.020]	6	3
2 个 a、1 个 c	$0.2^2\times0.5=0.02$	(0.020,0.045]	6	3
2 个 b、1 个 a	$0.3^2\times0.2=0.018$	(0,0.020]	6	3
2 个 b,1 个 c	$0.3^2\times0.5=0.045$	(0.020,0.045]	5	3
2 个 c,1 个 a	$0.5^2\times0.2=0.05$	(0.045,0.100]	4	3
2 个 c,1 个 b	$0.5^2\times0.3=0.075$	(0.045,0.100]	4	3
1 个 a,1 个 b,1 个 c	$0.2\times0.3\times0.5=0.03$	(0.020,0.045]	5	6

平均码长

$$\begin{aligned}\overline{K}&=0.008\times6+0.027\times5+0.125\times3+(0.012\times6+0.02\times6+0.018\times6+\\&\quad 0.045\times5+0.05\times4+0.075\times4)\times3+0.03\times5\times6\\&=4.533\text{bit/字}\end{aligned}$$

$$\begin{aligned}H(\boldsymbol{X})&=\Big(-\sum_{i=1}^{s}p_i\log p_i\Big)\times3=(-0.2\log_20.2-0.3\log_20.3-0.5\log_20.5)\times3\\&=4.455\text{bit/字}\end{aligned}$$

该信源编码的效率 $\eta=\dfrac{H(\boldsymbol{X})}{\overline{K}}=\dfrac{4.455}{4.533}=0.983$。

注：本题考查的知识点是 5.4.1 节哈夫曼编码。题解：(1)变长码的平均码长计算；

(2)哈夫曼编码效率的计算。

25. 设有一个无记忆信源 X 发出符号 A 和 B,已知 $p(\mathrm{A})=\frac{1}{4}$,$p(\mathrm{B})=\frac{3}{4}$。试:(1)计算该信源熵;(2)设该信源改为发三重序列消息的信源,采用哈夫曼编码方法对其进行编码,求平均信息传输速率。

解答:(1) 该离散无记忆信源的熵为 $H(X)=H\left(\frac{1}{4},\frac{3}{4}\right)=0.811\text{bit/符号}$。

(2) 三重序列消息信源的哈夫曼编码结果如表 5.22 所示。

表 5.22　题 25 的哈夫曼编码结果

三重序列消息	概　　率	编　　码
BBB	27/64	0
BAA	9/64	100
BAB	9/64	101
ABB	9/64	110
AAB	3/64	11100
ABA	3/64	11101
BAA	3/64	11110
AAA	1/64	11111

编码的平均长度 $\overline{K_2}=\frac{27}{64}+\frac{9}{64}\times3\times3+\frac{3}{64}\times5\times3+\frac{1}{64}\times5=2.46875$ 码元/三重符号。

因为信源是无记忆的,所以三重符号的熵 $H(\boldsymbol{X})=3H(X)$。

平均传输速率 $R_2=\frac{H(\boldsymbol{X})}{\overline{K}_2}=0.9855\text{bit/时间}$。

注:本题考查的知识点是 5.4.1 节哈夫曼编码。题解:(1)三重序列消息信源;(2)哈夫曼编码及其编码效率的计算。

26. 已知离散无记忆信源 $\begin{bmatrix}S\\P(s)\end{bmatrix}=\begin{bmatrix}s_1 & s_2 & s_3 & s_4 & s_5 & s_6\\0.3 & 0.2 & 0.15 & 0.15 & 0.1 & 0.1\end{bmatrix}$,试:(1)求信源熵 $H(S)$;(2)进行哈夫曼编码,并求平均码长和编码效率;(3)哈夫曼编码是否唯一?如果不唯一,哪种编码方法更佳?

解答:(1) 该信源的熵 $H(S)=H(0.3,0.2,0.15,0.15,0.1,0.1)=2.47\text{bit/符号}$。

(2) 哈夫曼编码结果如表 5.23 所示。

表 5.23　题 26 的哈夫曼编码结果

符　　号	概　　率	编　　码
s_1	0.3	00
s_2	0.2	01
s_3	0.15	010
s_4	0.15	011
s_5	0.1	110
s_6	0.1	111

平均码长 $\overline{K}=0.3\times2+0.2\times2+0.15\times3+0.1\times3+0.1\times3=2.5$ 二进制码元/符号。

编码效率 $\eta=\dfrac{H(S)}{\overline{K}}=\dfrac{2.47}{2.5}=0.988$。

(3) 由于按照概率大小顺序排队方法的不唯一,所以哈夫曼编码不唯一。将新合并的等概率消息排列到上支路(即大概率位置),能充分利用短码,缩短码长的方差,即编出的码更接近于等长码,这种编码方法更佳。

注:本题考查的知识点是5.4.1节哈夫曼编码,题解:二进制哈夫曼编码及其编码效率的计算。

27. 已知一个信源 X 包含8个符号消息,其概率分布为

$$\begin{bmatrix} X \\ P(x) \end{bmatrix}=\begin{bmatrix} A & B & C & D & E & F & G & H \\ 0.1 & 0.18 & 0.4 & 0.05 & 0.06 & 0.1 & 0.07 & 0.04 \end{bmatrix}。$$

(1) 信源每秒发出一个符号,求该信源的熵及信息传输速率;

(2) 对这8个符号作二进制码元的哈夫曼编码,写出各个代码组,并求出编码效率。

解答:(1) 该信源的熵

$$H(X)=H(0.1,0.18,0.4,0.05,0.06,0.1,0.07,0.04)=2.55\text{bit/符号}$$

信源每秒发出一个符号,则信息传输速率为 $R=2.55\text{bit/s}$。

(2) 哈夫曼编码结果如表5.24所示。

表5.24 题27的哈夫曼编码结果

符　号	概　率	编　码
C	0.4	0
B	0.18	110
A	0.1	100
F	0.1	1110
G	0.07	1010
E	0.06	1011
D	0.05	1110
H	0.04	11111

平均码长

$\overline{K}=0.4+0.18\times3+0.1\times3+0.1\times4+0.07\times4+0.06\times4+0.05\times5+0.04\times5$

$=2.61$ 码元/符号。

编码效率 $\eta=\dfrac{H(X)}{\overline{K}}=\dfrac{2.55}{2.61}=0.978$。

注:本题考查的知识点是5.4.1节哈夫曼编码。题解:二进制哈夫曼编码及其编码效率的计算。

28. 信源符号 X 有6种字母,概率分别为(0.32,0.22,0.18,0.16,0.08,0.04),试:(1)求符号熵 $H(X)$;(2)用哈夫曼编码编成二进制变长码,计算其编码效率;(3)当译码差错小于 10^{-3} 的定长二进制码要达到(2)中的效率时,估计要多少信源符号一起编码才能达到?

解答：(1) $H(X)=H(0.32,0.22,0.18,0.16,0.08,0.04)=2.352$bit/符号。

(2) 二进制哈夫曼编码的结果如表 5.25 所示。

表 5.25　题 28 的哈夫曼编码结果

符　　号	概　　率	编　　码
x_1	0.32	00
x_2	0.22	10
x_3	0.18	11
x_4	0.16	010
x_5	0.08	0110
x_6	0.04	0111

平均码长 $\overline{K}=(0.32+0.22+0.18)\times2+0.16\times3+(0.08+0.04)\times4=2.4$ 码元/符号

编码效率 $\eta=\dfrac{H(X)}{\overline{K}}=\dfrac{2.352}{2.4}=0.98$。

(3) 由编码效率 $\eta=\dfrac{H(X)}{H(X)+\varepsilon}=\dfrac{2.352}{2.352+\varepsilon}=0.98$ 可得 $\varepsilon=0.048$，代入 $L\geqslant\dfrac{\sigma^2(x)}{\varepsilon^2\delta}$，知如采用定长编码，需要 $L\geqslant\dfrac{\sigma^2(x)}{\varepsilon^2\delta}=2.29\times10^5$ 位信源符号一起编码才能达到 0.98 的编码效率。

注：本题考查的知识点是 5.4.1 节哈夫曼编码。题解：(1)二进制哈夫曼编码及其编码效率的计算；(2)定长编码与变长编码的比较。

29. 现有一幅已经离散化后的图像，图像的灰度量化分成 8 级，如表 5.26 所示。表中数字为相应像素上的灰度级。另有一个无噪无损二元信道，单位时间内传输 100 个二元符号。试回答以下问题：(1)统计此图像的像素点数。(2)不考虑图像的统计特性，并采用二元定长码，计算每个像素需要的码元数；将此图像通过给定的信道传输，计算传送完这幅图像所需的时间。(3)若考虑图像的统计特性，将像素的灰度值作为信源，写出信源的概率分布函数，计算出信源熵。(4)对此灰度级进行哈夫曼最佳二元编码，列出编码码表，计算每个像素需要的码元数、编码效率及其冗余度；若将此图像通过给定的信道传输，计算传送完这幅图像所需的时间。(5)这幅图像是否可以进一步压缩到小于信源熵 $H(X)$，说明理由。

表 5.26　题 29 的表

1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1
2	2	2	2	2	2	2	2	2	2
2	2	2	2	2	2	2	3	3	3
3	3	3	3	3	3	3	4	4	4
4	4	4	4	4	4	4	5	5	5
5	5	5	5	6	6	6	6	6	6
7	7	7	7	7	8	8	8	8	8

解答：(1) 该图像的像素点数共有 $10\times10=100$ 个。

(2) 若不考虑统计特性，因为有 8 个灰度级，采用二元定长码，故每个像素需要 3 个二元码，100 个像素需要 300 个二元码，单位时间内传输 100 个二元符号，需要 3s 传完。

(3) 若考虑图像的统计特性，这 100 个像素点灰度的概率分布为

$$\begin{bmatrix} X \\ P(X) \end{bmatrix}=\begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ \frac{40}{100} & \frac{17}{100} & \frac{10}{100} & \frac{10}{100} & \frac{7}{100} & \frac{6}{100} & \frac{5}{100} & \frac{5}{100} \end{bmatrix}$$

则信源熵 $H(X)=-\sum_{i=1}^{8}p(x_i)\log p(x_i)=2.572\text{bit/灰度级}$。

(4) 对灰度级进行哈夫曼编码，结果如表 5.27 所示。

表 5.27　题 29 的哈夫曼编码

灰　度　级	概　　率	编　　码
1	0.4	1
2	0.17	001
3	0.1	0000
4	0.1	0001
5	0.07	0100
6	0.06	0101
7	0.05	0110
8	0.05	0111

平均码长 $\overline{K}=2.63$ 二元码/灰度级。

编码效率 $\eta=\dfrac{H(X)}{\overline{K}}=\dfrac{2.572}{2.63}=0.978$。

冗余度 $r=1-0.978=0.022$。

此图像经过给定信道传输，需要 2.63s。

(5) 这幅图像可以进一步压缩。因为前面的编码是按无记忆编码的，没有考虑像素之间的记忆性，实际上像素之间是有依赖性的，可以看作 m 阶马尔可夫信源，从而得到极限熵 $H_\infty(X)$，是所有熵中最小的。

注：本题是一道综合题，考查 2.2.2 节离散信源熵、5.2.1 节定长编码和 5.4.1 节哈夫曼编码。题解：(1)离散信源熵的计算；(2)信源编码方法；(3)信息传输的概念。

30. 已知信源的协方差矩阵 $\boldsymbol{\Phi}_u=\begin{bmatrix}1&0&1\\0&1&0\\1&0&1\end{bmatrix}$，试求：最佳正交变换矩阵 $\boldsymbol{A}$ 和变换后的协方差矩阵 $\boldsymbol{\Phi}_x$。

解答：先计算信源的协方差矩阵 $\boldsymbol{\Phi}_u$ 的特征值。特征方程为

$$|\boldsymbol{\Phi}_u-\lambda\boldsymbol{I}|=\begin{vmatrix}1-\lambda&0&1\\0&1-\lambda&0\\1&0&1-\lambda\end{vmatrix}=(1-\lambda)(-\lambda)(2-\lambda)=0$$

可得 $\boldsymbol{\Phi}_u$ 的三个特征值为 $\lambda_1=2,\lambda_2=1,\lambda_3=0$。

最佳正交变换后的协方差矩阵为$\boldsymbol{\Phi}_x=\begin{bmatrix}2&0&0\\0&1&0\\0&0&0\end{bmatrix}$，设此时的最佳正交变换矩阵为$\boldsymbol{A}=[\xi_1\quad\xi_2\quad\xi_3]$，则

$$\boldsymbol{\Phi}_u\xi_1=\lambda_1\boldsymbol{I}\Rightarrow\begin{bmatrix}1&0&1\\0&1&0\\1&0&1\end{bmatrix}\xi_1=2\xi_1\Rightarrow\xi_1=\begin{pmatrix}\frac{1}{\sqrt{2}}\\0\\\frac{1}{\sqrt{2}}\end{pmatrix}$$

$$\boldsymbol{\Phi}_u\xi_2=\lambda_2\boldsymbol{I}\Rightarrow\begin{bmatrix}1&0&1\\0&1&0\\1&0&1\end{bmatrix}\xi_2=\xi_2\Rightarrow\xi_2=\begin{pmatrix}0\\1\\0\end{pmatrix}$$

$$\boldsymbol{\Phi}_u\xi_3=\lambda_3\boldsymbol{I}\Rightarrow\begin{bmatrix}1&0&1\\0&1&0\\1&0&1\end{bmatrix}\xi_3=0\Rightarrow\xi_3=\begin{pmatrix}\frac{1}{\sqrt{2}}\\0\\-\frac{1}{\sqrt{2}}\end{pmatrix}$$

所以，最佳正交变换矩阵$\boldsymbol{A}=\begin{bmatrix}\frac{1}{\sqrt{2}}&0&\frac{1}{\sqrt{2}}\\0&1&0\\\frac{1}{\sqrt{2}}&0&-\frac{1}{\sqrt{2}}\end{bmatrix}$。

注：本题考查的知识点是5.4.7节变换编码。题解：(1)由变换前信号的协方差矩阵的特征根得变换后的协方差矩阵；(2)利用线性代数求正交变换矩阵。

31. 设信源协方差矩阵$\boldsymbol{\Phi}_u=\begin{bmatrix}a&b&b&0\\b&a&0&b\\b&0&a&b\\0&b&b&a\end{bmatrix}$，试用DFT变换计算$\boldsymbol{\Phi}_x$。

解答：由DFT的变换公式得

$$\begin{aligned}\boldsymbol{\Phi}_x&=\boldsymbol{A}_{\mathrm{DF}}\cdot\boldsymbol{\Phi}_u\cdot\boldsymbol{A}_{\mathrm{DF}}^{*\mathrm{T}}\\&=\frac{1}{2}\begin{bmatrix}1&1&1&1\\1&-i&-1&+i\\1&-1&1&-1\\1&+i&-1&-i\end{bmatrix}\begin{bmatrix}a&b&b&0\\b&a&0&b\\b&0&a&b\\0&b&b&a\end{bmatrix}\times\frac{1}{2}\begin{bmatrix}1&1&1&1\\1&+i&-1&-i\\1&-1&1&-1\\1&-i&-1&+i\end{bmatrix}\\&=\begin{bmatrix}a+2b&0&0&0\\0&(a-b)&0&0\\0&0&a&0\\0&0&0&a-b\end{bmatrix}\end{aligned}$$

注：本题考查的知识点是5.4.7节变换编码。题解：(1)DFT变换矩阵；(2)变换前后

的协方差矩阵。

32. 若仍应用 31 题中的协方差矩阵$\boldsymbol{\Phi}_u$值，变换采用沃尔什-哈达马变换，试求$\boldsymbol{\Phi}_x$，并与 31 题的结果相比较。

解答：由沃尔什-哈达马变换公式得

$$\boldsymbol{\Phi}_x = \boldsymbol{A}_{\mathrm{WH}}(4)\,\boldsymbol{\Phi}_u \boldsymbol{A}_{\mathrm{WH}}(4)^{\mathrm{T}}$$

$$= \frac{1}{2}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix} \begin{bmatrix} a & b & b & 0 \\ b & a & 0 & b \\ b & 0 & a & b \\ 0 & b & b & a \end{bmatrix} \times \frac{1}{2}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} a+2b & 0 & 0 & 0 \\ 0 & a & 0 & 0 \\ 0 & 0 & a & 0 \\ 0 & 0 & 0 & a-2b \end{bmatrix}$$

可以看出，与 31 题的结果作比较，两矩阵在形式上都是对角阵，只是对角线上的元素不同。

注：本题考查的知识点是 5.4.7 节变换编码。题解：(1)哈达马变换矩阵；(2)变换前后的协方差矩阵。

33. (1)若用 DCT 对第 31 题中的信源进行变换，试求变换后的协方差矩阵；(2)若信源的协方差矩阵$\boldsymbol{\Phi}_u = \begin{bmatrix} a & b & 0 & b \\ b & a & b & 0 \\ 0 & b & a & b \\ b & 0 & b & a \end{bmatrix}$，试求变换后的协方差矩阵。

解答：(1) DCT 变换矩阵 $\boldsymbol{A}_{\mathrm{DCT}}(4) = \frac{1}{2}\begin{bmatrix} 1 & 1 & 1 & 1 \\ c & d & -d & -c \\ 1 & -1 & -1 & 1 \\ d & -c & c & -d \end{bmatrix}$，其转置 $\boldsymbol{A}_{\mathrm{DCT}}^{\mathrm{T}}(4) = \frac{1}{2}\begin{bmatrix} 1 & c & 1 & d \\ 1 & d & -1 & -c \\ 1 & -d & -1 & c \\ 1 & -c & 1 & -d \end{bmatrix}$，其中，$c = \sqrt{2}\cos\frac{\pi}{8}$，$d = \sqrt{2}\cos\frac{3\pi}{8}$。

$$\boldsymbol{\Phi}_x = \boldsymbol{A}_{\mathrm{DCT}}\boldsymbol{\Phi}_u\boldsymbol{A}_{\mathrm{DCT}}^{\mathrm{T}} = \frac{1}{2}\begin{bmatrix} 1 & 1 & 1 & 1 \\ c & d & -d & -c \\ 1 & -1 & -1 & 1 \\ d & -c & c & -d \end{bmatrix} \cdot \begin{bmatrix} a & b & b & 0 \\ b & a & 0 & b \\ b & 0 & a & b \\ 0 & b & b & a \end{bmatrix} \times \frac{1}{2}\begin{bmatrix} 1 & c & 1 & d \\ 1 & d & -1 & -c \\ 1 & -d & -1 & c \\ 1 & -c & 1 & -d \end{bmatrix}$$

$$= \begin{bmatrix} a+2b & 0 & 0 & 0 \\ 0 & \frac{1}{2}a(c^2+d^2) & 0 & 0 \\ 0 & 0 & a-2b & 0 \\ 0 & 0 & 0 & \frac{1}{2}a(c^2+d^2) \end{bmatrix}$$

将 c 和 d 的值代入后得：$\boldsymbol{\Phi}_x=\begin{bmatrix}a+2b & 0 & 0 & 0\\ 0 & a & 0 & 0\\ 0 & 0 & a-2b & 0\\ 0 & 0 & 0 & a\end{bmatrix}$。

(2) 由题意，协方差矩阵变为$\boldsymbol{\Phi}_u=\begin{bmatrix}a & b & 0 & b\\ b & a & b & 0\\ 0 & b & a & b\\ b & 0 & b & a\end{bmatrix}$，则

$$\boldsymbol{\Phi}_x=\boldsymbol{A}_{\mathrm{DCT}}\boldsymbol{\Phi}_u\boldsymbol{A}_{\mathrm{DCT}}^{\mathrm{T}}=\frac{1}{2}\begin{bmatrix}1 & 1 & 1 & 1\\ c & d & -d & -c\\ 1 & -1 & -1 & 1\\ d & -c & c & -d\end{bmatrix}\begin{bmatrix}a & b & 0 & b\\ b & a & b & 0\\ 0 & b & a & b\\ b & 0 & b & a\end{bmatrix}\times\frac{1}{2}\begin{bmatrix}1 & c & 1 & d\\ 1 & d & -1 & -c\\ 1 & -d & -1 & c\\ 1 & -c & 1 & -d\end{bmatrix}$$

$$=\frac{1}{4}\begin{bmatrix}4a+8b & 0 & 0 & 0\\ 0 & 2a(c^2+d^2)-2b(c^2+d^2)+4bcd & 0 & 2b(d^2-c^2)\\ 0 & 0 & 4a & 0\\ 0 & 2b(d^2-c^2) & 0 & 2a(c^2+d^2)-2b(c^2+d^2)+4bcd\end{bmatrix}$$

将 c 和 d 代入，得$\boldsymbol{\Phi}_x=\begin{bmatrix}a+2b & 0 & 0 & 0\\ 0 & (a-b)+0.707b & 0 & -0.708b\\ 0 & 0 & a & 0\\ 0 & -0.708b & 0 & a-b-0.707b\end{bmatrix}$。

注：本题考查的知识点是 5.4.7 节变换编码。题解：(1)DCT 变换矩阵；(2)变换后，信号的协方差矩阵的对角性。

34. 若已知信源的协方差矩阵$\boldsymbol{\Phi}_u=\begin{bmatrix}a & b & b & b\\ b & a & b & b\\ b & b & a & b\\ b & b & b & a\end{bmatrix}$，试求：(1)由 K-L 变换所得输出协方差矩阵$\boldsymbol{\Phi}_x$；(2)用哈尔变换求输出的协方差矩阵$\boldsymbol{\Phi}_x$。

解答：(1) 设 K-L 变换的变换矩阵 $\boldsymbol{A}=\begin{bmatrix}1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\end{bmatrix}$，因为 $\boldsymbol{A}$ 是正交的，所以有

$a_i\boldsymbol{\Phi}_u a_i^{\mathrm{T}}=\lambda_i$，其中，$\lambda_1=\begin{pmatrix}1 & 0 & 0 & 0\end{pmatrix}\boldsymbol{\Phi}_u\begin{pmatrix}1\\0\\0\\0\end{pmatrix}=a$，$\lambda_2=\begin{pmatrix}0 & 1 & 0 & 0\end{pmatrix}\boldsymbol{\Phi}_u\begin{pmatrix}0\\1\\0\\0\end{pmatrix}=a$，

$\lambda_3=\begin{pmatrix}0 & 0 & 1 & 0\end{pmatrix}\boldsymbol{\Phi}_u\begin{pmatrix}0\\0\\1\\0\end{pmatrix}=a$，$\lambda_4=\begin{pmatrix}0 & 0 & 0 & 1\end{pmatrix}\boldsymbol{\Phi}_u\begin{pmatrix}0\\0\\0\\1\end{pmatrix}=a$。

此时，$\boldsymbol{\Phi}_x=\boldsymbol{A}\boldsymbol{\Phi}_u\boldsymbol{A}^{\mathrm{T}}=\begin{bmatrix} a & & & \\ & a & & \\ & & a & \\ & & & a \end{bmatrix}$。

（2）设哈尔变换的变换矩阵 $\boldsymbol{A}_{\mathrm{Hr}}(4)=\frac{1}{2}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ \sqrt{2} & -\sqrt{2} & 0 & 0 \\ 0 & 0 & \sqrt{2} & -\sqrt{2} \end{bmatrix}$，则

$$\begin{aligned}\boldsymbol{\Phi}_x&=\boldsymbol{A}_{\mathrm{Hr}}\boldsymbol{\Phi}_u\boldsymbol{A}_{\mathrm{Hr}}^{\mathrm{T}}\\&=\frac{1}{4}\begin{bmatrix} a+3b & a+3b & a+3b & a+3b \\ a-b & a-b & b-a & b-a \\ \sqrt{2}(a-b) & \sqrt{2}(b-a) & 0 & 0 \\ 0 & 0 & \sqrt{2}(a-b) & \sqrt{2}(b-a) \end{bmatrix}\cdot\begin{bmatrix} 1 & 1 & \sqrt{2} & 0 \\ 1 & 1 & -\sqrt{2} & 0 \\ 1 & -1 & 0 & \sqrt{2} \\ 1 & -1 & 0 & -\sqrt{2} \end{bmatrix}\\&=\begin{bmatrix} a+3b & 0 & 0 & 0 \\ 0 & a-b & 0 & 0 \\ 0 & 0 & a-b & 0 \\ 0 & 0 & 0 & a-b \end{bmatrix}\end{aligned}$$

注：本题考查的知识点是 5.4.7 节变换编码。题解：K-L 变换。