信息论作业 5

史泽宇

2020 年 4 月 19 日

题目 1

1. 为了满足题目排序条件，对字母进行重排

表 1: 二元码

字母	a_8	a_7	a_6	a_5	a_4	a_3	a_2	a_1
概率	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{4}$
cdf	0	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{3}{16}$	$\frac{4}{16}$	$\frac{6}{16}$	$\frac{8}{16}$	$\frac{12}{16}$
I	4	4	4	4	3	3	2	2
n	4	4	4	4	3	3	2	2
编码	0000	0001	0010	0011	010	011	10	11

2. (a) 首先证明该编码是异字头码。$\forall a_k$，有 $n_k = \lceil \log_2 \frac{1}{P(a_k)} \rceil \geq \log_2 \frac{1}{P(a_k)}$，所以有 $2^{-n_k} \leq P(a_k) = cdf(a_{k+1}) - cdf(a_k) \leq cdf(a_{k+n}) - cdf(a_k), where\ n \geq 1$。即 a_{k+n} 与 a_k 的编码至少有一位不同

   (b) $\bar{n} = \sum_{a \in U} P(a) \lceil \log_2 \frac{1}{P(a)} \rceil \geq \sum_{a \in U} P(a) \log_2 \frac{1}{P(a)} = H(U) bits$
   $\bar{n} = \sum_{a \in U} P(a) \lceil \log_2 \frac{1}{P(a)} \rceil < \sum_{a \in U} P(a) \log_2 \frac{1}{P(a)} + 1 = (H(U) + 1) bits$

题目 2

1. 由于字母出现的概率相等，所以当 $\alpha = 1$ 时，对应的编码树为满二叉树，编码长度为 j。当 $\alpha = 2$ 时，编码长度为 $j + 1$。当 $1 < \alpha < 2$ 时，由于取整编码长度向上取证，所以只存在 $j, j + 1$ 长度的编码

2. 设长为 j 的码字个数为 N_j，长度为 $j + 1$ 的码字数目为 N_{j+1}，根据二元 Huffman 编码思想（必定占满整个码树），即 $N_j + N_{j+1} = K = \alpha 2^j, N_j 2^j + N_{j+1} 2^{-(j+1)} = 1$。从而 $N_j = (2 - \alpha) 2^j, N_{j+1} = (\alpha - 1) 2^{j+1}$

3. $l = \frac{jN_j + (j+1)N_{j+1}}{K} = \frac{(2-\alpha)j + 2(\alpha-1)(j+1)}{\alpha} = \frac{2j - j\alpha + 2j\alpha + 2\alpha - 2j - 2}{\alpha} = \frac{j\alpha + 2\alpha - 2}{\alpha} = j + 2 - \frac{2}{\alpha}$

题目 3

表 2: 算术编码

步骤	输入字母	编码区间	间隔
1	1	[0.25, 1)	0.75
2	0	[0.25, 0.4375)	0.1875
3	1	[0.296875, 0.4375)	0.140625
4	1	[0.300390625, 0.4375)	0.10546875
5	0	[0.33203125, 0.3583984375)	0.0263671875
6	1	[0.338623046875, 0.3583984375)	0.019775390625
7	1	[0.34356689453125, 0.3583984375)	0.01483154296875
8	1	[0.3472747802734375, 0.3583984375)	0.0111236572265625
9	1	[0.3500556945800781, 0.3583984375)	0.008342742919921875
10	0	[0.3500556945800781, 0.3521413803100586)	0.002085685729980469
11	1	[0.35057711601257324, 0.3521413803100586)	0.0015642642974853516
12	1	[0.3509681820869446, 0.3521413803100586)	0.0011731982231140137
13	0	[0.3509681820869446, 0.3512614816427231)	0.0002932995557785034
14	1	[0.3510415069758892, 0.3512614816427231)	0.00021997466683387756
15	1	[0.3510965006425977, 0.3512614816427231)	0.00016498100012540817
16	1	[0.351137745892629, 0.3512614816427231)	0.00012373575009405613

1. 编码长度为 $n = \lceil -\log_2 0.00012373575009405613 \rceil = 13$，而 $0.351137745892629 \approx (0.0101100111100100)_2$，所以编码结果为 0101100111101。编码效率为 $\frac{H(U)}{\frac{13}{16}} = \frac{\frac{1}{4}\log_2 4+\frac{3}{4}\log_2\frac{4}{3}}{\frac{13}{16}} \approx 0.9985$

表 3: LZ 编码

段号	前一段号	最后一位编码	段编码
1	0	1	0001
2	0	0	0000
3	1	1	0011
4	2	1	0101
5	3	1	0111
6	4	1	1001
7	6	1	1101

2. 所以最终编码为 0001000000110101011110011101，编码效率为 $\frac{H(U)}{\frac{28}{16}} \approx 0.4636$

题目 4 第二版的作业中弃用

题目似乎有点问题，如果按照题目计算的话，结果为

设信源输出长度为 l，则编码长度为 $\bar{l} = 1\frac{l}{2} + 2\frac{l}{4} + 3\frac{l}{8} + 4\frac{l}{8} = \frac{15}{8}l$，0 的个数为 $\bar{l}_0 = \frac{l}{2} + \frac{l}{4} + \frac{l}{8} + \frac{l}{8} = l, \bar{l}_1 = \frac{l}{4} + 2\frac{l}{8} + 3\frac{l}{8} = \frac{7}{8}l, P(0) = \frac{8}{15}, P(1) = \frac{7}{15}$

教材上课后题目中，最后一个字母的编码为 111，使用教材上的编码计算结果为

设信源输出长度为 l，则编码长度为 $\bar{l}=1\frac{l}{2}+2\frac{l}{4}+3\frac{l}{8}+3\frac{l}{8}=\frac{7}{4}l$，0 的个数为 $\bar{l}_0=\frac{l}{2}+\frac{l}{4}+\frac{l}{8}=\frac{7}{8}l,\bar{l}_1=\frac{l}{4}+2\frac{l}{8}+3\frac{l}{8}=\frac{7}{8}l,P(0)=\frac{1}{2},P(1)=\frac{1}{2}$

表 4: 信源概率分布

字母	概率	字母长度
1	0.1	1
01	$(0.9)(0.1)$	2
001	$(0.9)^2(0.1)$	3
0001	$(0.9)^3(0.1)$	4
00001	$(0.9)^4(0.1)$	5
000001	$(0.9)^5(0.1)$	6
0000001	$(0.9)^6(0.1)$	7
00000001	$(0.9)^7(0.1)$	8
00000000	0.9^8	8

题目 5

1. 由信源的概率分布可计算 $H(U)\approx 0.469bits$
2. 由表 4 中的概率分布与字母长度可计算 $\bar{n}_1\approx 5.6953$
3. 由表 4 中的概率分布可计算 $\bar{n}_2\approx 2.7086$
4. 观察编码结果发现，任意编码都不是其他编码的前缀，所以是异字头码

题目 6

1. 显然是三元对称信道，输入等概时达到信道容量 $C=\log_2 3-H(P)bps$
2. 显然是输入为二元输出为四元的对称信道,输入等概时达到信道容量 $C=\log_2 4+(1-p)\log_2\frac{(1-p)}{2}+p\log_2\frac{p}{2}=1-H(P)bps$
3. 可以拆分成两个和信道的组合, $C_1=1-H(P),C_2=0,C=\log_2(2^{C_1}+2^{C_2})=\log_2(2^{1-H(P)}+1)bps$

题目 7

1. $$\begin{bmatrix}\frac{3}{4} & \frac{1}{4} & 0\\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3}\\ 0 & \frac{1}{4} & \frac{3}{4}\end{bmatrix} \tag{1}$$

   可以简化为如下的二元纯删除信道

   $$\begin{bmatrix}\frac{3}{4} & \frac{1}{4} & 0\\ 0 & \frac{1}{4} & \frac{3}{4}\end{bmatrix} \tag{2}$$

   显然达到信道容量的分布为等概分布，信道容量为 $C=\frac{3}{4}bps$。

2.

$$\begin{bmatrix} \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} \\ 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \end{bmatrix} \tag{3}$$

显然为准对称信道，输入等概时达到信道容量，可以拆分为以下两个弱对称信道[1]

$$\begin{bmatrix} \frac{1}{3} & \frac{1}{3} & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & 0 & \frac{1}{3} \end{bmatrix} \begin{bmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{bmatrix} \tag{4}$$

$C_1 = \log_2 3 - H(\frac{\frac{1}{3}}{\frac{2}{3}}, \frac{\frac{1}{3}}{\frac{2}{3}}, \frac{0}{\frac{2}{3}}) = \log_2 \frac{3}{2}, C_2 = \log_2 1 - H(\frac{\frac{1}{3}}{\frac{1}{3}}) = 0, C = \frac{2}{3}C_1 + \frac{1}{3}C_2 = \frac{2}{3}\log_2 \frac{3}{2} bps$

题目 8 信道容量为 $C = W\log_2(1+\frac{S}{N}) = 3000\log_2(1+1000) \approx 30kbps$，三十分钟传送的信息为 $3 * 60 * C \approx 5.3823mb$

题目 9

1. 最大似然译码

$$\begin{bmatrix} P(y_i|x_j) & y_1 & y_2 & y_3 \\ x_1 & \frac{1}{2} & \frac{1}{3} & \frac{1}{6} \\ x_2 & \frac{1}{6} & \frac{1}{2} & \frac{1}{3} \\ x_3 & \frac{1}{3} & \frac{1}{6} & \frac{1}{2} \end{bmatrix} \tag{5}$$

由矩阵中的最大转移概率可得译码规则为

$$\begin{aligned} F(y_1) &= x_1 \\ F(y_2) &= x_2 \\ F(y_3) &= x_3 \end{aligned} \tag{6}$$

误码率为 $P_e = (\frac{1}{2}+\frac{1}{4}+\frac{1}{4})(\frac{1}{6}+\frac{1}{3}) = \frac{1}{2}$

2. 最大后验译码，由贝叶斯公式可得

$$\begin{bmatrix} P(x_i|y_j) & y_1 & y_2 & y_3 \\ x_1 & \frac{2}{3} & \frac{1}{2} & \frac{2}{7} \\ x_2 & \frac{1}{9} & \frac{3}{8} & \frac{2}{7} \\ x_3 & \frac{2}{9} & \frac{1}{8} & \frac{3}{7} \end{bmatrix} \tag{7}$$

由矩阵中的最大后验概率可得译码规则为

$$\begin{aligned} F(y_1) &= x_1 \\ F(y_2) &= x_1 \\ F(y_3) &= x_3 \end{aligned} \tag{8}$$

误码率为 $P_e = (\frac{1}{2})(\frac{1}{6}) + \frac{1}{4} + \frac{1}{4}(\frac{1}{3}+\frac{1}{6}) = \frac{11}{24}$

[1] Lemma 4.18