信息论与编码作业

（第三次）

1．一信源产生概率为 $P(1)=0.005$, $P(0)=0.995$ 的统计独立二进制数符。这些数符组成长度为 100 的数符组。我们为每一个含有 3 个或少于 3 个“1”的源数符组提供一个二进制码字，所有码字的长度相等。

(a) 求出为所规定的所有源符组都提供码字所需的最小码长。

(b) 求信源发出一数符组，而编码器无相应码字的概率。

(c) 用契比雪夫不等式求编码器无相应码字的概率上界，并将结果与(b)的结果比较。

2．设信源 $\begin{pmatrix} X \\ P(X) \end{pmatrix} = \begin{Bmatrix} a_1 & a_2 \\ \frac{3}{4} & \frac{1}{4} \end{Bmatrix}$ 。对该信源进行等长二元编码，要求编码损失（差错率） $P_e \le 10^{-5}$，编码效率 $\eta = 0.96$，求符合要求的最低联合编码长度。、

3．某信源有六种可能的输出，概率分布如下表所示。表中给出了六种编码方式。

符号	$P(a_i)$	A	B	C	D	E	F
a_1	1/2	000	0	0	0	0	0
a_2	1/4	001	01	10	10	10	100
a_3	1/16	010	011	110	110	1100	101
a_4	1/16	011	0111	1110	1110	1101	110
a_5	1/16	100	01111	11110	1011	1110	111
a_6	1/16	101	011111	111110	1101	1111	011

(a) 这些码中哪些是唯一可译码？

(b) 这些码中哪些是即刻码？

(c) 对所有唯一可译码求出其平均码长和编码效率。

4．设某城市有 805 门公务电话和 60 000 门居民电话。作为系统工程师，你需要为这些用户分配电话号码。所有号码均是十进制数，且不考虑电话系统中 0、1 不可用在号码首位的限制。（提示：用异前缀码概念）

(a) 如果要求所有公务电话号码为 3 位长，所有居民电话号码等长，求居民

号码长度 L_1 的最小值。

(b) 设城市分为 A、B 两个区，其中 A 区有 9 000 门电话，B 区有 51 000 门电话。现进一步要求 A 区的电话号码比 B 区的短 1 位，试问 A 区号码长度 L_2 的最小值。

5．设有信源

$$\begin{pmatrix} X \\ P(X) \end{pmatrix} = \begin{pmatrix} a_1 & a_2 & a_3 & a_4 & a_5 & a_6 & a_7 & a_8 \\ 0.2 & 0.15 & 0.15 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 \end{pmatrix}$$

分别对该信源进行二进制和三进制 Huffman 编码，并求出平均码长和编码效率。

6. 某离散无记忆信源两个符号的概率分布为 $P=(3/4,1/4)$，分别对其进行 $L=2,3$ 的联合二进制 Huffman 编码，并求出平均码长和编码效率。

7.

根据定长编码理论，可以通过只对典型序列编码，而不给出现概率非常小的非典型序列提供对应的码字序列，从而实现了对信源的压缩。

我们知道，对于抛硬币实验，多次抛硬币所得的结果正面和反面出现的次数近似相等，很难看到几乎全是正面或反面的情况出现，因此我们可以不给这样的序列编码。（比如，如果对抛 100 次硬币的结果进行编码，可以不对出现 100 次正面或反面这样的序列编码）。

然而，根据信源编码定理，如果二进制信源的熵 $H(U)=1\text{bit}$，若对其进行二进制编码，则无法实现压缩。因为

$$N\log D \ge L\left[H(U)+\delta\right]$$

当 $D=2$ 时，$N \ge L+L\delta$。即只有当 $N \ge L+L\delta$ 时，才存在使 $P_e \to 0$ 的编码，当不满足该条件时，一定不存在 $P_e \to 0$ 的编码方法。

请对上述看似矛盾的结论给出你的解释。