《信息、控制与计算》作业习题解答

民间版

作者：SuperbRain

组织：浙江大学信息与电子工程学院

时间：Janurary, 2024

版本：1.0

“考前勤奋，考后乐观。”

目录

第二章 信息度量和性质

2.1 A 村有一半人说真话，3/10 人总说假话，2/10 人拒绝回答；B 村有 3/10 人诚实，一半人说谎，2/10 人拒绝回答。现随机地从 A 村和 B 村抽取人，p 为抽到 A 村人的概率，$1-p$ 为抽到 B 村人的概率，问通过测试某人说话的状态平均能获得多少关于该人属于哪个村的信息？通过改变 p，求出该信息的最大值。

解 记随机变量 X, Y 分别表示说话的状态和属于某村的情况。二者的状态空间分别为：

$$\mathcal{X} = \{x_1, x_2, x_3\}, \mathcal{Y} = \{y_1, y_2\}$$

其中 x_1, x_2, x_3 分别表示说真话、拒绝回答、说假话，y_1, y_2 分别表示属于 A 村，属于 B 村。因此有

$$\begin{aligned}
&p(X = x_1) = 0.5p + 0.3(1-p) = 0.2p + 0.3 \\
&p(X = x_2) = 0.2 + 0.2(1-p) = 0.2 \\
&p(X = x_3) = 0.3p + 0.5(1-p) = -0.2p + 0.5 \\
&p(X = x_1 \mid Y = y_1) = 0.5 \qquad p(X = x_1 \mid Y = y_2) = 0.3 \\
&p(X = x_2 \mid Y = y_1) = 0.2 \qquad p(X = x_2 \mid Y = y_2) = 0.2 \\
&p(X = x_3 \mid Y = y_1) = 0.3 \qquad p(X = x_3 \mid Y = y_2) = 0.5
\end{aligned}$$

因此可以分别求出 $H(X)$ 和 $H(X \mid Y)$：

$$\begin{aligned}
H(X) &= -[(0.2p + 0.3)\log_2(0.2p + 0.3) + 0.2\log_2 0.2 + (-0.2p + 0.5)\log_2(-0.2p + 0.5)] \\
H(X \mid Y) &= p(Y = y_1)H(X \mid Y = y_1) + p(Y = y_2)H(X \mid Y = y_2) \\
&= -p(0.5\log_2 0.5 + 0.2\log_2 0.2 + 0.3\log_2 0.3) - (1-p)(0.5\log_2 0.5 + 0.2\log_2 0.2 + 0.3\log_2 0.3) \\
&= 1.485\text{bit}
\end{aligned}$$

因此可以求得平均互信息量 $I(X;Y)$ 为：

$$\begin{aligned}
I(X;Y) &= H(X) - H(X \mid Y) \\
&= \{-[(0.2p + 0.3)\log_2(0.2p + 0.3) + (-0.2p + 0.5)\log_2(-0.2p + 0.5)] - 1.02\}\text{bit}
\end{aligned}$$

易知 $I(X;Y)$ 是关于 p 和 $1-p$ 对称的，即

$$I(X;Y) = f(p) + f(1-p)$$

则有

$$\frac{\mathrm{d}I(X;Y)}{\mathrm{d}p} = f'(p) - f'(1-p)$$

因此当 $p = 0.5$ 时，$\dfrac{\mathrm{d}I(X;Y)}{\mathrm{d}p} = 0$。即 $p = 0.5$ 时，$I(X;Y)$ 最大，为

$$I(X;Y)\Big|_{p=0.5} = 0.037\text{bit}$$

2.2 一个无偏骰子，抛掷一次，如果出现 1，2，3，4 点，则把一枚均匀硬币投掷一次，如果骰子出现 5，6 点，则硬币投掷 2 次，求硬币投掷中正面出现次数对于骰子出现点数所提供的信息。

解 记 X 为硬币正面出现次数，Y 为骰子点数。并定义如下两个集合：

$$A = \{1, 2, 3, 4\} \qquad B = \{5, 6\}$$

因此有

$$p(X=0)=\frac{2}{3}\times\frac{1}{2}+\frac{1}{3}\times\frac{1}{2}\times\frac{1}{2}=\frac{5}{12}$$
$$p(X=1)=\frac{2}{3}\times\frac{1}{2}+\frac{1}{3}\times C_2^1\times\frac{1}{2}\times\frac{1}{2}=\frac{1}{2}$$
$$p(X=2)=\frac{1}{3}\times\frac{1}{2}\times\frac{1}{2}=\frac{1}{12}$$
$$p(X=0\mid Y\in A)=\frac{1}{2}\qquad p(X=0\mid Y\in B)=\frac{1}{2}\times\frac{1}{2}=\frac{1}{4}$$
$$p(X=1\mid Y\in A)=\frac{1}{2}\qquad p(X=1\mid Y\in B)=C_2^1\times\frac{1}{2}\times\frac{1}{2}=\frac{1}{2}$$
$$p(X=2\mid Y\in A)=0\qquad p(X=2\mid Y\in B)=\frac{1}{2}\times\frac{1}{2}=\frac{1}{4}$$

因此可以分别求出 $H(X)$ 和 $H(X\mid Y)$:

$$H(X)=-\sum_{i=0}^{2}p(X=i)\log_2 p(X=i)=1.325\text{bit}$$

$$\begin{aligned}H(X\mid Y)&=p(Y\in A)H(X\mid Y\in A)+p(Y\in B)H(X\mid Y\in B)\\&=\frac{2}{3}\left[-\sum_{i=0}^{2}p(X=i\mid Y\in A)\log_2 p(X=i\mid Y\in A)\right]+\frac{1}{3}\left[-\sum_{i=0}^{2}p(X=i\mid Y\in B)\log_2 p(X=i\mid Y\in B)\right]\\&=1.167\text{bit}\end{aligned}$$

因此可求得互信息量为

$$I(X;Y)=H(X)-H(X\mid Y)=0.158\text{bit}$$

2.4 随机掷 3 颗骰子，以 X 表示第一颗骰子抛掷的结果，以 Y 表示第一颗和第二颗骰子抛掷之和，以 Z 表示 3 颗骰子的点数之和，试求 $H(X\mid Y),H(Y\mid X),H(Z\mid X,Y),H(X,Z\mid Y)$ 和 $H(Z\mid X)$。

解 设第一颗骰子结果为 X_1，第二颗骰子结果为 X_2，第三颗骰子结果为 X_3，则 X_1,X_2,X_3 为独立同分布的随机变量，且有

$$X=X_1$$
$$Y=X_1+X_2$$
$$Z=X_1+X_2+X_3$$

进而可以写出 X 和 Y 的分布:

X	1	2	3	4	5	6
$p(x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Y	2	3	4	5	6	7	8	9	10	11	12
$p(y)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

因此

$$H(X)=H(X_1)=H(X_2)=H(X_3)=(\log_2 6)\text{bit}=2.585\text{bit}$$
$$H(Y)=-\sum_{y_i}p(y_i)\log_2 p(y_i)=3.274\text{bit}$$

$$
\begin{aligned}
H(X \mid Y) &= H(X) + H(Y \mid X) - H(Y) \\
&= H(X) + H(X_2) - H(Y) \\
&= 2H(X) - H(Y) \\
&= 1.896\text{bit} \\
H(Y \mid X) &= H(X_2) = 2.585\text{bit} \\
H(Z \mid X, Y) &= H(X_3) = 2.585\text{bit} \\
H(X, Z \mid Y) &= H(X, Z, Y) - H(Y) \\
&= H(Z \mid X, Y) + H(X, Y) - H(Y) \\
&= H(Z \mid X, Y) + H(X \mid Y) \\
&= 4.481\text{bit} \\
H(Z \mid X) &= H(X_2 + X_3) = H(X_1 + X_2) = H(Y) = 3.274\text{bit}
\end{aligned}
$$

✍ **2.5** 设一个系统传送 10 个数字：$0, 1, 2, \cdots, 9$，奇数在传送时以 0.5 概率等可能地错成另外的奇数，而其他数字总能正确接收。试求收到一个数字后平均得到的信息量。

解 设随机变量 X 表示发送数字，Y 表示接收数字。那么 (X, Y) 的取值 (x, y) 可以分成如下三类：

$$
S_1 = \{(0,0), (2,2), (4,4), (6,6), (8,8)\}
$$

$$
S_2 = \{(1,1), (3,3), (5,5), (7,7), (9,9)\}
$$

$$
S_3 = \left\{ \begin{array}{l} (1,3), (1,5), (1,7), (1,9), \\ (3,1), (3,5), (3,7), (3,9), \\ (5,1), (5,3), (5,7), (5,9), \\ (7,1), (7,3), (7,5), (7,9), \\ (9,1), (9,3), (9,5), (9,7) \end{array} \right\}
$$

三种情况下的条件概率分别为

$$
p_1(Y = y \mid X = x) = 1, \quad p_2(Y = y \mid X = x) = \frac{1}{2}, \quad p_3(Y = y \mid X = x) = \frac{1}{8}
$$

而对于 X 任意取值 x，有

$$
p(X = x) = 0.1
$$

至此可以分三部分求得 $I(X;Y)$：

$$
\begin{aligned}
I(X;Y) &= \sum_x \sum_y p(x,y) \log_2 \frac{p(x \mid y)}{q(x)} \\
&= 5p(X=0, Y=0) \log_2 \frac{p(X=0 \mid Y=0)}{p(X=0)} + 5p(X=1, Y=1) \log_2 \frac{p(X=1 \mid Y=1)}{p(X=1)} \\
&\quad + 20p(X=1, Y=3) \log_2 \frac{p(X=1 \mid Y=3)}{p(X=1)} \\
&= 2.3214\text{bit}
\end{aligned}
$$

✍ **2.6** 对任意概率分布的随机变量，证明下述三角不等式成立

$$
\text{(a) } H(X \mid Y) + H(Y \mid Z) \geq H(X \mid Z)
$$

$$
\text{(b) } \frac{H(X \mid Y)}{H(X,Y)} + \frac{H(Y \mid Z)}{H(Y,Z)} \geq \frac{H(X \mid Z)}{H(X,Z)}
$$

[提示]

定理

对任意概率分布的随机变量 X,Y,Z，有

$$\frac{H(X\mid Y)}{H(X,Y)}+\frac{H(Y\mid Z)}{H(Y,Z)}\geq\frac{H(X\mid Y)+H(Y\mid Z)}{H(X\mid Y)+H(Y\mid Z)+H(Z)}$$

证明

(a) 在两端同时加 $H(Z)$ 有：

$$H(X\mid Y)+H(Y\mid Z)+H(Z)=H(X\mid Y)+H(Y,Z)\geq H(X\mid Y,Z)+H(Y,Z)=H(X,Y,Z)$$
$$H(X\mid Z)+H(Z)=H(X,Z)$$

由于 $H(X,Y,Z)\geq H(X,Z)$，因此有

$$H(X\mid Y)+H(Y\mid Z)+H(Z)\geq H(X,Y,Z)\geq H(X,Z)=H(X\mid Z)+H(Z)$$

也即

$$H(X\mid Y)+H(Y\mid Z)\geq H(X\mid Z)$$

(b) 由 [提示] 得

$$\begin{aligned}\frac{H(X\mid Y)}{H(X,Y)}+\frac{H(Y\mid Z)}{H(Y,Z)}&\geq\frac{H(X\mid Y)+H(Y\mid Z)}{H(X\mid Y)+H(Y\mid Z)+H(Z)}\\&=1-\frac{H(Z)}{H(X\mid Y)+H(Y\mid Z)+H(Z)}\\&\geq 1-\frac{H(Z)}{H(X\mid Z)+H(Z)}\\&=\frac{H(X\mid Z)}{H(X,Z)}\end{aligned}$$

2.9 若 3 个随机变量 X,Y,Z，有 $X+Y=Z$ 成立，其中 X 和 Y 独立，试证：

(a) $H(X)\leq H(Z)$

(b) $H(Y)\leq H(Z)$

(c) $H(X,Y)\geq H(Z)$

(d) $I(X;Z)=H(Z)-H(Y)$

(e) $I(X,Y;Z)=H(Z)$

(f) $I(X;Y,Z)=H(X)$

(g) $I(Y;Z\mid X)=H(Y)$

(h) $I(X;Y\mid Z)=H(X\mid Z)=H(Y\mid Z)$

证明 (a) 由于 X 和 Y 独立，因此有

$$H(X,Y)=H(X)+H(Y)$$

由于

$$I(X,Y;Z)=H(Z)-H(Z\mid X,Y)=H(X,Y)-H(X,Y\mid Z)$$

而 $H(Z \mid X,Y)=0$, 因此有

$$\begin{aligned} H(Z) &= H(X,Y) - H(X,Y \mid Z) \\ &= H(X) + H(Y) - H(X,Y \mid Z) \\ &= H(X) + H(Y) - [H(Y \mid Z) + H(X \mid Y,Z)] \\ &= H(X) + H(Y) - H(Y \mid Z) \\ &= H(X) + I(Y;Z) \\ &\geq H(X) \end{aligned}$$

(b) 与 (a) 同理，只需换成 $H(Z)=H(Y)+I(X;Z)$ 即可。
(c) 由 (a) 可知

$$H(X,Y) - H(Z) = H(X,Y \mid Z) \geq 0$$

因此 $H(X,Y) \geq H(Z)$.
(d) 由 (b) 可知

$$H(Z) = H(Y) + I(X;Z)$$

因此有 $I(X;Z)=H(Z)-H(Y)$.
(e) 由于 $H(Z \mid X,Y)=0$，因此有

$$I(X,Y;Z) = H(Z) - H(Z \mid X,Y) = H(Z)$$

(f) 与 (e) 同理，由于 $H(X \mid Y,Z)=0$, 因此有

$$I(X;Y,Z) = H(X) - H(X \mid Y,Z) = H(X)$$

(g) 由于

$$I(Y;X,Z) = I(Y;X) + I(Y;Z \mid X)$$

因此

$$H(Y) - H(Y \mid X,Z) = H(Y) - H(Y \mid X) + I(Y;Z \mid X)$$

也即

$$\begin{aligned} I(Y;Z \mid X) &= H(Y \mid X) - H(Y \mid X,Z) \\ &= H(Y) - 0 \\ &= H(Y) \end{aligned}$$

(h) 与 (g) 同理，有

$$I(X;Y \mid Z) = H(X \mid Z) - H(X \mid Y,Z) = H(X \mid Z) = H(Y \mid Z)$$

2.10 令 X 是离散随机变量，$Y=g(X)$ 是 X 的任意函数，求证 $H(X) \geq H(Y)$.
证明 由于

$$I(X;Y) = H(X) - H(X \mid Y) = H(Y) - H(Y \mid X)$$

因此

$$\begin{aligned} H(X) &= H(Y) + H(X \mid Y) - H(Y \mid X) \\ &= H(Y) + H(X \mid Y) - 0 \\ &\geq H(Y) \end{aligned}$$

✍ **2.12** 证明下列关系

(a) $H(Y,Z \mid X) \le H(Y \mid X) + H(Z \mid X)$ 等号成立的充要条件为对一切 i,j,k 有 $p(y_j,z_k \mid x_i) = p(y_j \mid x_i)p(z_k \mid x_i)$ 成立。

(b) $H(Y,Z \mid X) = H(Y \mid X) + H(Z \mid X,Y)$.

(c) $H(Z \mid X,Y) \le H(Z \mid Y)$ 等号成立的充要条件为对一切 i,j,k 有 $p(x_i,z_k \mid y_j) = p(x_i \mid y_j)p(z_k \mid y_j)$ 成立。

证明 先证 (b) 的结论：

(b)

$$\begin{aligned}
H(Y,Z \mid X) &= -\sum_x \sum_y \sum_z p(x,y,z) \log p(y \mid z,x) \\
&= -\sum_x \sum_y \sum_z p(x,y,z) \log [p(z \mid x,y)p(y \mid x)] \\
&= -\sum_x \sum_y \sum_z p(x,y,z) [\log p(z \mid x,y) + \log p(y \mid x)] \\
&= H(Y \mid X) + H(Z \mid X,Y)
\end{aligned}$$

(a)

$$\begin{aligned}
H(Y,Z \mid X) &= H(Y \mid X) + H(Z \mid X,Y) \\
&\le H(Y \mid X) + H(Z \mid X)
\end{aligned}$$

当且仅当 $H(Z \mid X,Y) = H(Z \mid X)$ 时等号成立。也即对任意的 i,j,k，有

$$p(z_k \mid x_i,y_j) = \frac{p(z_k,y_j \mid x_i)}{p(y_j \mid x_i)} = p(z_k \mid x_i)$$

即有

$$p(y_j,z_k \mid x_i) = p(y_j \mid x_i)p(z_k \mid x_i)$$

(c) 与 (a) 同理

$$\begin{aligned}
H(Z \mid X,Y) &= H(X,Z \mid Y) - H(X \mid Y) \\
&\le H(X \mid Y) + H(Z \mid Y) - H(X \mid Y) \\
&= H(Z \mid Y)
\end{aligned}$$

因此当且仅当对一切 i,j,k 有 $p(x_i,z_k \mid y_j) = p(x_i \mid y_j)p(z_k \mid y_j)$ 成立时，等号成立。

✍ **2.13** 令 X_1 和 X_2 为分别定义在字符表 $\mathscr{X}_1 = \{1,2,\cdots,m\}$ 和 $\mathscr{X}_2 = \{m+1,\cdots,n\}$ 上的 2 个离散随机变量，它们的分布函数为 $p_1(\cdot)$ 和 $p_2(\cdot)$。现构造随机变量

$$X = \begin{cases} X_1 & \text{以概率}\alpha \\ X_2 & \text{以概率}1-\alpha \end{cases}$$

(a) 试用 $H(X_1), H(X_2)$ 和 α 来表示 X 的熵 $H(X)$.

(b) 在 α 上求 $H(X)$ 的极大值，证明 $2^{H(X)} \le 2^{H(X_1)} + 2^{H(X_2)}$。

解 (a)

$$
\begin{aligned}
H(X) &= -\sum_{x\in\mathscr{X}_1\cup\mathscr{X}_2}[\alpha p_1(x)+(1-\alpha)p_2(x)]\log[\alpha p_1(x)+(1-\alpha)p_2(x)] \\
&= -\alpha\sum_{x\in\mathscr{X}_1}p_1(x)\log[\alpha p_1(x)]-(1-\alpha)\sum_{x\in\mathscr{X}_2}p_2(x)\log[(1-\alpha)p_2(x)] \\
&= -\alpha\sum_{x\in\mathscr{X}_1}p_1(x)\log p_1(x)-\alpha\log\alpha\sum_{x\in\mathscr{X}_1}p_1(x)-(1-\alpha)\sum_{x\in\mathscr{X}_2}p_2(x)\log p_2(x)-(1-\alpha)\log(1-\alpha)\sum_{x\in\mathscr{X}_2}p_2(x) \\
&= \alpha H(X_1)+(1-\alpha)H(X_2)-\alpha\log\alpha-(1-\alpha)\log(1-\alpha)
\end{aligned}
$$

(b)$H(X)$ 对 α 求一阶导数并令其为 0 有

$$\frac{\mathrm{d}H(X)}{\mathrm{d}\alpha}=H(X_1)-H(X_2)-\log\alpha+\log(1-\alpha)=0$$

因此 $a=\dfrac{2^{H(X_1)}}{2^{H(X_1)}+2^{H(X_2)}}$ 时 $H(X)$ 取极大值，此时

$$H(X)_{\max}=\log[2^{H(X_1)}+2^{H(X_2)}]$$

因此有

$$2^{H(X)}\le 2^{H(X_1)}+2^{H(X_2)}$$

2.17 有 n 枚硬币，其中有一枚是次品，它可能比正品重或轻。今用一个无砝码的天平对这 n 个硬币称重，检测出这个次品硬币，并确定它比正品“重”或“轻”。

(a) 通过 k 次称重，可以检测出次品硬币，并正确判定它比次品“重”或“轻”的硬币数 n 最多为多大?

(b) 对 $k=3, n=12$ 这种情况，试设计一种称重的方案。

解 (a) 称重一次可能出现的情况为“轻”、“平”、“重”三种，因此在三种情况等可能时获得的信息量最大，为

$$I_0=\left(-\frac{1}{3}\log_2\frac{1}{3}\right)\times 3\text{bit}=(\log_2 3)\text{bit}$$

称重 n 次可能出现的情况有 $2n$ 种，即每一枚硬币都可能出现“轻假币”和“重假币”两种情况，当这 $2n$ 种情况等可能时获得的信息量最大，为

$$I_n=\log_2(2n)\text{bit}$$

因此若想正确判定，必须有 $kI_0\ge I_n$，即

$$k\log_2 3\ge\log_2(2n)$$

因此 $n\le\dfrac{3^k}{2}$，即 n 最多为 $\left\lfloor\dfrac{3^k}{2}\right\rfloor$.

(b) 见下表

第一次称重	第二次称重	第三次称重
若 $A_1+A_2+A_3+A_4=A_5+A_6+A_7+A_8$，则 A_9,A_{10},A_{11},A_{12} 中有假币	若 $A_1+A_9=A_{10}+A_{11}$，则 A_{12} 为假币	若 $A_1>A_{12}$，则 A_{12} 是轻假币
		若 $A_1<A_{12}$，则 A_{12} 是重假币
	若 $A_1+A_9>A_{10}+A_{11}$，则 A_9 是重假币或 A_{10},A_{11} 中有轻假币	若 $A_{10}=A_{11}$，则 A_9 是重假币
		若 $A_{10}>A_{11}$，则 A_{11} 是轻假币
		若 $A_{10}<A_{11}$，则 A_{10} 是轻假币
	若 $A_1+A_9<A_{10}+A_{11}$，则 A_9 是轻假币或 A_{10},A_{11} 中有重假币	若 $A_{10}=A_{11}$，则 A_9 是轻假币
		若 $A_{10}>A_{11}$，则 A_{10} 是重假币
		若 $A_{10}<A_{11}$，则 A_{11} 是重假币
若 $A_1+A_2+A_3+A_4>A_5+A_6+A_7+A_8$，则 A_1,A_2,A_3,A_4 中有重假币或 A_5,A_6,A_7,A_8 中有轻假币	若 $A_1+A_2+A_6=A_5+A_3+A_4$，则 A_7,A_8 中有轻假币	若 $A_1=A_7$，则 A_8 为轻假币
		若 $A_1\neq A_7$，则 A_7 为轻假币
	若 $A_1+A_2+A_6>A_5+A_3+A_4$，则 A_1,A_2 中有重假币或 A_5 为轻假币	若 $A_1=A_2$，则 A_5 为轻假币
		若 $A_1>A_2$，则 A_1 为重假币
		若 $A_1<A_2$，则 A_2 为重假币
	若 $A_1+A_2+A_6<A_5+A_3+A_4$，则 A_3,A_4 中有重假币或 A_6 为轻假币	若 $A_3=A_4$，则 A_6 为轻假币
		若 $A_3>A_4$，则 A_3 为重假币
		若 $A_3<A_4$，则 A_4 为重假币
若 $A_1+A_2+A_3+A_4<A_5+A_6+A_7+A_8$，则 A_1,A_2,A_3,A_4 中有轻假币或 A_5,A_6,A_7,A_8 中有重假币	若 $A_1+A_2+A_6=A_5+A_3+A_4$，则 A_7,A_8 中有重假币	若 $A_1=A_7$，则 A_8 为重假币
		若 $A_1\neq A_7$，则 A_7 为重假币
	若 $A_1+A_2+A_6<A_5+A_3+A_4$，则 A_1,A_2 中有轻假币或 A_5 为重假币	若 $A_1=A_2$，则 A_5 为重假币
		若 $A_1>A_2$，则 A_2 为轻假币
		若 $A_1<A_2$，则 A_1 为轻假币
	若 $A_1+A_2+A_6>A_5+A_3+A_4$，则 A_3,A_4 中有轻假币或 A_6 为重假币	若 $A_3=A_4$，则 A_6 为重假币
		若 $A_3>A_4$，则 A_4 为轻假币
		若 $A_3<A_4$，则 A_3 为轻假币

2.20 令 X 是取值 ±1 的二元随机变量，概率分布为 $p(x=1)=p(x=-1)=0.5$，令 Y 是连续随机变量，已知条件概率密度为

$$p(y\mid x)=\begin{cases}\frac{1}{4} & -2<y-x\leq 2\\ 0 & \text{其他}\end{cases}$$

试求

(a) Y 的概率密度。

(b) $I(X;Y)$。

(c) 若对 Y 做硬判决

$$V=\begin{cases}1 & y>1\\ 0 & -1<y\leq 1\\ -1 & y\leq -1\end{cases}$$

求 $I(V;X)$，并对结果加以解释。

解 (a)

$$
\begin{aligned}
p(y) &= \sum_{x\in\{-1,1\}} p(y\mid x)p(x)\\
&= p(X=-1)p(y\mid X=-1)+p(X=1)p(y\mid X=1)\\
&= \begin{cases} \frac{1}{8} & -3<y\le -1\text{或}1<y\le 3\\ \frac{1}{4} & -1<y\le 1\\ 0 & \text{其他}\end{cases}
\end{aligned}
$$

(b) 记 X 和 Y 的样本空间分别为 $\mathscr{X}$ 和 $\mathscr{Y}$，则有

$$
\begin{aligned}
I(X;Y) &= \sum_{x\in\mathscr{X}}\int_{\mathscr{Y}} p(x,y)\log\frac{p(y\mid x)}{p(y)}\mathrm{d}y\\
&= \frac{1}{2}\int_{-3}^{-1}\frac{1}{4}\log\frac{\frac{1}{4}}{\frac{1}{8}}\mathrm{d}y+\frac{1}{2}\int_{-1}^{1}\frac{1}{4}\log\frac{\frac{1}{4}}{\frac{1}{4}}\mathrm{d}y+\frac{1}{2}\int_{-1}^{1}\frac{1}{4}\log\frac{\frac{1}{4}}{\frac{1}{4}}\mathrm{d}y+\frac{1}{2}\int_{1}^{3}\frac{1}{4}\log\frac{\frac{1}{4}}{\frac{1}{8}}\mathrm{d}y\\
&= 0.5\text{bit}
\end{aligned}
$$

(c) 可以求出如下的概率：

$$
\begin{aligned}
&p(V=1\mid X=1)=\int_{1}^{3}\frac{1}{4}\mathrm{d}y=\frac{1}{2} && p(V=1\mid X=-1)=0\\
&p(V=0\mid X=1)=\int_{-1}^{1}\frac{1}{4}\mathrm{d}y=\frac{1}{2} && p(V=0\mid X=-1)=\int_{-1}^{1}\frac{1}{4}\mathrm{d}y=\frac{1}{2}\\
&p(V=-1\mid X=1)=0 && p(V=-1\mid X=-1)=\int_{-3}^{-1}\frac{1}{4}\mathrm{d}y=\frac{1}{2}
\end{aligned}
$$

因此可求出 V 的分布

$$
\begin{aligned}
p(V=1)&=\frac{1}{2}\times\frac{1}{2}+\frac{1}{2}\times 0=\frac{1}{4}\\
p(V=0)&=\frac{1}{2}\times\frac{1}{2}+\frac{1}{2}\times\frac{1}{2}=\frac{1}{2}\\
p(V=-1)&=\frac{1}{2}\times\frac{1}{2}+\frac{1}{2}\times 0=\frac{1}{4}
\end{aligned}
$$

因此有

$$
\begin{aligned}
I(V;X) &= \sum_{x\in\{-1,1\}}\sum_{v\in\{-1,0,1\}} p(v,x)\log\frac{p(v\mid x)}{p(v)}\\
&= 0.5\text{bit}
\end{aligned}
$$

可以看出 $I(V;X)=I(Y;X)$，因此对 Y 做硬判决 V 是不损失信息的。

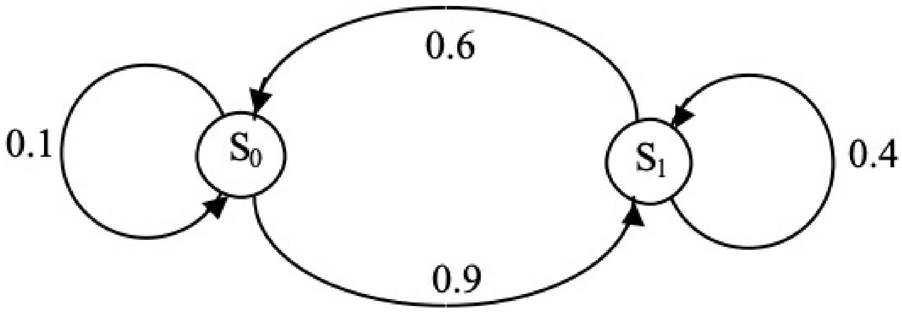
✍ **2.21** 1 个二状态平稳马尔可夫信源，输出为 $\cdots,X_0,X_1,X_2,\cdots$，信源状态转移概率矩阵为 $\boldsymbol{P}=\begin{pmatrix}\frac{1}{10} & \frac{9}{10}\\ \frac{3}{5} & \frac{2}{5}\end{pmatrix}$，对任何 n 求 $\frac{1}{n}H(X_1X_2\cdots X_n)$ 和 $H(X_n\mid X_1X_2\cdots X_{n-1})$。

解 由于信源是一阶平稳马尔可夫过程，因此有

$$
\frac{1}{n}H(X_1X_2\cdots X_n)=\frac{1}{n}[H(X_1)+H(X_2\mid X_1)+\cdots+H(X_n\mid X_{n-1})]=\frac{1}{n}H(X_1)+\frac{n-1}{n}H(X_2\mid X_1)
$$

$$
H(X_n\mid X_1X_2\cdots X_{n-1})=\begin{cases} H(X_1) & n=1\\ H(X_2\mid X_1) & n\ge 2\end{cases}
$$

由转移概率矩阵可以得到状态转移图如下：



其稳定状态下概率满足方程组

$$\begin{cases} 0.1p(S_0) + 0.6p(S_1) = p(S_0) \\ 0.9p(S_0) + 0.4p(S_1) = p(S_1) \\ p(S_0) + p(S_1) = 1 \end{cases}$$

解得平稳分布为 $p(S_0) = 0.4, p(S_1) = 0.6$。同时可以求得

$$H(X \mid S_0) = -0.1\log 0.1 - 0.9\log 0.9 = 0.469\text{bit}$$

$$H(X \mid S_1) = -0.4\log 0.4 - 0.6\log 0.6 = 0.971\text{bit}$$

因此

$$H(X_2 \mid X_1) = H_\infty = p(S_0)H(X \mid S_0) + p(S_1)H(X \mid S_1) = 0.77\text{bit}$$

$$H(X_1) = -0.4\log 0.4 - 0.6\log 0.6 = 0.97\text{bit}$$

综上，

$$\frac{1}{n}H(X_1X_2\cdots X_n) = \left(\frac{0.97}{n} + 0.77\cdot\frac{n-1}{n}\right)\text{bit}$$

$$H(X_n \mid X_1X_2\cdots X_{n-1}) = \begin{cases} 0.97\text{bit} & n = 1 \\ 0.77\text{bit} & n \geq 2 \end{cases}$$

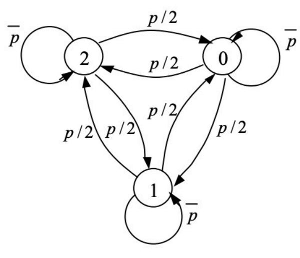
2.23 一阶马尔可夫信源的状态图如习题 2.23 图所示，信源符号集为 $\{0,1,2\}$，并定义 $\bar{p} = 1 - p$。

(a) 求信源平稳概率分布 $p(0), p(1), p(2)$。

(b) 求此信源的熵。

(c) 近似认为此信源为无记忆时，符号概率分布等于平稳分布，求此近似信源的熵 $H(X)$，并与 H_∞ 进行比较。

(d) 对一阶马尔可夫信源，p 取何值时 H_∞ 取最大值？又当 $p = 0$ 和 $p = 1$ 时结果如何？



习题 2.23 图

解 (a) 由图可列出方程组

$$\begin{cases}\bar{p}\cdot p(0)+\frac{p}{2}\cdot p(1)+\frac{p}{2}\cdot p(2)=p(0)\\ \frac{p}{2}\cdot p(0)+\bar{p}\cdot p(1)+\frac{p}{2}\cdot p(2)=p(1)\\ \frac{p}{2}\cdot p(0)+\frac{p}{2}\cdot p(1)+\bar{p}\cdot p(2)=p(2)\\ p(0)+p(1)+p(2)=1\end{cases}$$

解得 $p(0)=p(1)=p(2)=\dfrac{1}{3}$.

(b)$\forall i\in\{0,1,2\}$，有

$$H(X\mid i)=-\bar{p}\log\bar{p}-\frac{p}{2}\log\frac{p}{2}-\frac{p}{2}\log\frac{p}{2}=-\bar{p}\log\bar{p}-p\log\frac{p}{2}$$

因此信源的熵

$$H_\infty=\sum_{i=0}^{2}p(i)H(X\mid i)=\bar{p}\log\bar{p}-p\log\frac{p}{2}$$

(c)

$$p(X=i)=\sum_{j=0}^{2}p(j)p(X=i\mid j)=\frac{1}{3}\sum_{j=0}^{2}p_{ij}=\frac{1}{3}$$

因此

$$H(X)=(\log_2 3)\mathrm{bit}\geq H_\infty$$

(d) 令 $\dfrac{\mathrm{d}H_\infty}{\mathrm{d}p}=0$ 有 $p=\dfrac{2}{3}$，此时 H_∞ 取极大值 $(\log_2 3)\mathrm{bit}$。
$p=0$ 时，$H_\infty=0\mathrm{bit}$，$p=1$ 时，$H_\infty=1\mathrm{bit}$。

2.25 (a) 求状态转移矩阵为

$$\boldsymbol{P}=\begin{pmatrix}1-p_{01} & p_{01}\\ p_{10} & 1-p_{10}\end{pmatrix}$$

的 2 状态马尔可夫信源的熵率。

(b) 使 (a) 中熵率最大的 p_{01},p_{10} 为多少?

(c) 如状态转移矩阵为

$$\boldsymbol{P}=\begin{pmatrix}1-p & p\\ 1 & 0\end{pmatrix}$$

则相应的 2 状态马尔可夫信源熵率为多少?

(d) 寻找最大化 (c) 中熵率的 p 值。

(e) 令 $N(t)$ 表示 (c) 中马尔可夫信源输出的长度为 t 的可允许状态序列的数目。求 $N(t)$ 并计算 $H=\lim\limits_{t\to\infty}\dfrac{1}{t}\log N(t)$。

解 (a) 可列出方程组

$$\begin{cases}(1-p_{01})p(S_0)+p_{10}p(S_1)=p(S_0)\\ p_{01}p(S_0)+(1-p_{10})p(S_1)=p(S_1)\\ p(S_0)+p(S_1)=1\end{cases}$$

解得 $p(S_0)=\dfrac{p_{10}}{p_{01}+p_{10}},p(S_1)=\dfrac{p_{01}}{p_{01}+p_{10}}$ 因此熵率为

$$\begin{aligned}H_\infty(p_{01},p_{10})&=p(S_0)H(X\mid S_0)+p(S_1)H(X\mid S_1)\\&=\frac{p_{10}}{p_{01}+p_{10}}[-p_{01}\log p_{01}-(1-p_{01})\log(1-p_{01})]+\frac{p_{01}}{p_{01}+p_{10}}[-p_{10}\log p_{10}-(1-p_{10})\log(1-p_{10})]\end{aligned}$$

(b) 记 $\boldsymbol{p}=[p_{01},p_{10}]^{\mathrm{T}}$，因此熵率的梯度为 $\boldsymbol{0}$ 对应熵率最大的点。记 $m(p)=-p\log p-(1-p)\log(1-p)$, 注意到

$m'(0.5)=0$。构造两个向量函数：

$$\boldsymbol{f}(\boldsymbol{x})=\left[\frac{x_2}{x_1+x_2},\frac{x_1}{x_1+x_2}\right]^{\mathrm{T}}$$

$$\boldsymbol{g}(\boldsymbol{x})=[m(x_1),m(x_2)]^{\mathrm{T}}$$

因此很容易求得两个向量函数的梯度矩阵：

$$\nabla_{\boldsymbol{x}}\boldsymbol{f}(\boldsymbol{x})=\begin{bmatrix}-\frac{x_2}{(x_1+x_2)^2} & \frac{x_2}{(x_1+x_2)^2}\\ \frac{x_1}{(x_1+x_2)^2} & -\frac{x_1}{(x_1+x_2)^2}\end{bmatrix}\qquad \nabla_{\boldsymbol{x}}\boldsymbol{g}(\boldsymbol{x})=\begin{bmatrix}m'(x_1) & 0\\ 0 & m'(x_2)\end{bmatrix}$$

注意到

$$H_\infty(\boldsymbol{p})=(\boldsymbol{f}(\boldsymbol{p}))^{\mathrm{T}}\boldsymbol{g}(\boldsymbol{p})$$

因此可求梯度向量：

$$\nabla_{\boldsymbol{p}}H_\infty(\boldsymbol{p})=\nabla_{\boldsymbol{p}}\boldsymbol{f}(\boldsymbol{p})\cdot\boldsymbol{g}(\boldsymbol{x})+\nabla_{\boldsymbol{p}}\boldsymbol{g}(\boldsymbol{p})\cdot\boldsymbol{f}(\boldsymbol{x})=\begin{bmatrix}-\frac{p_{10}}{(p_{01}+p_{10})^2} & \frac{p_{10}}{(p_{01}+p_{10})^2}\\ \frac{p_{01}}{(p_{01}+p_{10})^2} & -\frac{p_{01}}{(p_{01}+p_{10})^2}\end{bmatrix}\begin{bmatrix}m(p_{01})\\ m(p_{10})\end{bmatrix}+\begin{bmatrix}m'(p_{01}) & 0\\ 0 & m'(p_{10})\end{bmatrix}\begin{bmatrix}\frac{p_{10}}{p_{01}+p_{10}}\\ \frac{p_{01}}{p_{01}+p_{10}}\end{bmatrix}$$

注意到 $\boldsymbol{p}=\left[\frac{1}{2},\frac{1}{2}\right]^{\mathrm{T}}$ 时，上式的两项均为 $\mathbf{0}$。即此时熵率最大，因此使 (a) 中熵率最大的 $p_{10}=p_{01}=\frac{1}{2}$.

(c) 取 $\boldsymbol{p}=[p,1]$, 则有

$$H_\infty(p)=-\frac{1}{1+p}[p\log p+(1-p)\log(1-p)]$$

(d) 注意到 $H_\infty(p)=\frac{m(p)}{1+p}$，因此对 p 求一阶导数有

$$\frac{\mathrm{d}H_\infty(p)}{\mathrm{d}p}=\frac{m'(p)(1+p)-m(p)}{(1+p)^2}$$

因此若一阶导数为 0 必有

$$m'(p)(1+p)=m(p)$$

进而得到方程

$$p=(1-p)^2$$

解该方程可得 $p=\frac{3-\sqrt{5}}{2}$。

(e)$N(t)$ 在正常情况下显然为 2^t，然而在 (c) 中，状态 2 的下一个状态只能是状态 1，因此可以将二者看成一个整体，因此只存在状态 1 直接变成状态 1 以及状态 1 先变成状态 2 再变成状态 1 两种情况（类比爬楼梯问题）。因此可以得到递推关系：

$$N(t)=N(t-1)+N(t-2)$$

其中 $N(1)=2,N(2)=3$。因此可以得到 $N(t)$ 为：

$$N(t)=\frac{1}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{2}\right)^{t+2}-\left(\frac{1-\sqrt{5}}{2}\right)^{t+2}\right]$$

进一步地有

$$
\begin{aligned}
H = \lim_{t\to\infty} \frac{\log N(t)}{t} &= \lim_{t\to\infty} \frac{\log \frac{1}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{2}\right)^{t+2} - \left(\frac{1-\sqrt{5}}{2}\right)^{t+2}\right]}{t} \\
&= \lim_{t\to\infty} \frac{\log \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^{t+2} \cdot \left[1-\left(\frac{1-\sqrt{5}}{1+\sqrt{5}}\right)^{t+2}\right]}{t} \\
&= \lim_{t\to\infty} \frac{\log \left(\frac{1+\sqrt{5}}{2}\right)^{t+2}}{t} + \lim_{t\to\infty} \frac{\log \left[1-\left(\frac{1-\sqrt{5}}{1+\sqrt{5}}\right)^{t+2}\right]}{t} + \lim_{t\to\infty} \frac{\log \frac{1}{\sqrt{5}}}{t} \\
&= \log \left(\frac{1+\sqrt{5}}{2}\right)
\end{aligned}
$$

第三章 信息感知与压缩

3.1 试证明最长长度不大于 N 的 D 元不等长码至多有 $D(D^N-1)/(D-1)$ 个码字。

证明 最长长度不大于 N 的 D 元不等长码的最多码字数情况为每种长度的所有码字都取到，即

$$\sum_{l=1}^{N-1} D^l = \frac{D(D^N-1)}{D-1}$$

证毕。

3.2 设一个 DMS,$U = \begin{pmatrix} a_1 & a_2 \\ p(a_1)=0.25 & p(a_2)=0.75 \end{pmatrix}$，其熵为 $H(U)$。考察长度为 L 的输出序列，当 $L \geq L_0$ 时满足

$$P_r\left\{\left|\frac{I(u^L)}{L} - H(U)\right| > \delta\right\} \leq \varepsilon$$

(a) 在 $\delta = 0.05, \varepsilon = 0.1$ 时求 L_0 值。

(b) 在 $\delta = 10^{-3}, \varepsilon = 10^{-6}$ 时求 L_0 值。

(c)

$$A = \left\{u^L : \left|\frac{I(u^L)}{L} - H(U)\right| < \delta\right\}$$

求在 (a)、(b) 给定的 $L = L_0$ 情况下 A 中元素数目的上、下限。

解 根据 Chebyshev 不等式，有

$$P_r\left\{\left|\frac{I(u^L)}{L} - H(U)\right| > \delta\right\} \leq \frac{\sigma_I^2}{L\delta^2}$$

因此若在 $l \geq L_0$ 时满足题意，只需

$$\frac{\sigma_I^2}{L_0\delta^2} \leq \varepsilon$$

因此取

$$L_0 = \left\lceil \frac{\sigma_I^2}{\varepsilon\delta^2} \right\rceil$$

下面求 $H(U)$ 和 σ_I^2：

$$H(U) = \sum_{i=1}^{2} p(a_i)I(a_i) = -\sum_{i=1}^{2} p(a_i)\log p(a_i) = 0.811\text{bit}$$

$$\sigma_I^2 = \sum_{i=1}^{2} p(a_i)[I(a_i) - H(U)]^2 = 0.471\text{bit}$$

(a) 在 $\delta = 0.05, \varepsilon = 0.1$ 时，

$$L_0 = \left\lceil \frac{\sigma_I^2}{\varepsilon\delta^2} \right\rceil = 1884$$

(b) 在 $\delta = 10^{-3}, \varepsilon = 10^{-6}$ 时，

$$L_0 = \left\lceil \frac{\sigma_I^2}{\varepsilon\delta^2} \right\rceil = 4.71 \times 10^{11}$$

(c) 由推论 3.1.1 知，对一切 $L \geq L_0$ 及 $u^L \in A$ 有

$$2^{-L[H(U)+\delta]} \leq p(u^L) \leq 2^{-L[H(U)-\delta]}$$

因此有

$$1-\varepsilon \le p\{u^L \in A\} = \sum_{u^L \in A} p(u^L) \le |A| \cdot 2^{-L[H(U)-\delta]}$$

$$1 = \sum_{u^L \in U^L} p(u^L) \ge \sum_{u^L \in A} p(u^L) \ge |A| \cdot 2^{-L[H(U)+\delta]}$$

即

$$(1-\varepsilon)2^{L[H(U)-\delta]} \le |A| \le 2^{L[H(U)+\delta]}$$

因此在 (a) 的情况下，$|A|$ 的下界为 0.9×2^{1434}，上界为 2^{1622}；
在 (b) 的情况下，$|A|$ 的下界为 $(1-10^{-6})2^{3.815\times 10^{11}}$，上界为 $2^{3.825\times 10^{11}}$。

3.3 下面哪些码是唯一可译的？

(a)$\{0, 10, 11\}$ (b)$\{0, 01, 11\}$ (c)$\{0, 01, 10\}$ (d)$\{0, 01\}$
(e)$\{00, 01, 10, 11\}$ (f)$\{110, 11, 10\}$ (g)$\{110, 100, 00, 10\}$

解 (a) $\{0, 10, 11\}$ 是前缀码，因此唯一可译；
(b) $\{0, 01, 11\}$ 的后缀分解集为

S_0	S_1	S_2
0	1	1
01		
11		

故后缀分解集中不含码字，因此唯一可译；
(c) $\{0, 01, 10\}$ 的后缀分解集为

S_0	S_1	S_2
0	1	0
01		
10		

后缀分解集 S_2 中包含码字 0，因此不是唯一可译；
(d) $\{0, 01\}$ 的后缀分解集为

S_0	S_1
0	1
01	

后缀分解集中不含码字，因此唯一可译；
(e) $\{00, 01, 10, 11\}$ 是前缀码，因此唯一可译；
(f) $\{110, 11, 10\}$ 的后缀分解集为

S_0	S_1	S_2
110	0	
11		
10		

后缀分解集中不含码字，因此唯一可译；

(g) $\{110, 100, 00, 10\}$ 的后缀分解集为

S_0	S_1	S_2	S_3
110	0	0	0
100			
00			
10			

后缀分解集中不含码字，因此唯一可译。

3.4 确定下面码是否唯一可译，若不是则构造一个模糊序列。

(a)

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
010	0001	0110	1100	00011	00110	11110	101011

(b)

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
abc	abcd	e	dba	bace	ceac	ceab	eabd

解 (a) 后缀分解为

S_0	S_1	S_2	S_3	S_4	S_5	S_6
010	1	100	11	00	01	0
0001		1110		110	011	10
0110		01011			110	001
1100					0	110
00011						0011
00110						0110
11110						
101011						

故后缀分解集 S_6 中包含码字 0110，因此不是唯一可译。
例如序列”000110101111000110” 可译为”0001,101011,1100,0110” 和”00011,010,11110,00110” 两种译法。
(b) 后缀分解为

S_0	S_1	S_2	S_3	S_4	S_5	S_6	S_7
abc	d	ba	ce	ac	c	eac	d
abcd	abd			ab	cd	eab	ac
e							ab
dba							
bace							
ceac							
ceab							
eabd							

后缀分解集中不含码字，因此唯一可译。

3.6 令 DMS 为

$$U = \begin{pmatrix} a_1 & a_2 & a_3 & a_4 & a_5 & a_6 & a_7 & a_8 & a_9 & a_{10} \\ 0.16 & 0.14 & 0.13 & 0.12 & 0.1 & 0.09 & 0.08 & 0.07 & 0.06 & 0.05 \end{pmatrix}$$

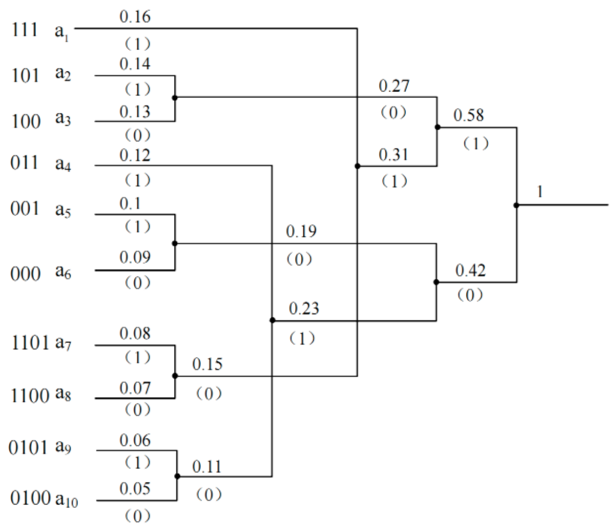
(a) 求二元 Huffman 码，计算 $\bar{n}$ 和 η。

(b) 求三元 Huffman 码，计算 $\bar{n}$ 和 η。

解 由信源概率分布可知

$$H(U) = -\sum_{i=1}^{10} p_i \log p_i = 3.234\text{bit}$$

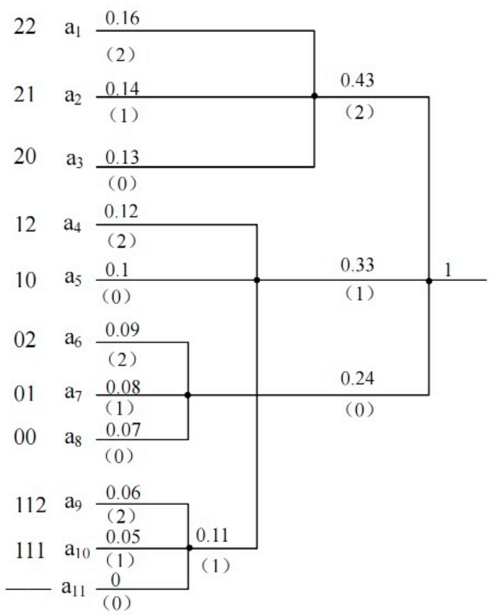
(a) 二元 Huffman 编码过程如下图所示：



$$\bar{n} = \sum_{i=1}^{10} p_i n_i = 3.26$$

$$\eta = \frac{H(U)}{\bar{n}\log D} = 99.2\%$$

(b) 三元 Huffman 编码过程如下图所示：



$$\bar{n} = \sum_{i=1}^{10} p_i n_i = 2.11$$

$$\eta = \frac{H(U)}{\bar{n}\log D} = 96.7\%$$

✍ **3.7** 下面三个码中，哪些对任何概率分布都不可能成为 Huffman 码?

(a)$\{0, 10, 11\}$

(b)$\{00, 01, 10, 110\}$

(c)$\{01, 10\}$

解 (a) 显然 $\{0, 10, 11\}$ 是 $U = \begin{pmatrix} a_1 & a_2 & a_3 \\ 0.5 & 0.25 & 0.25 \end{pmatrix}$ 的二元 Huffman 码。

(b)110 可以被 11 代替，此时仍为前缀码，因此不是 Huffman 码。

(c)$\{0, 1\}$ 是码长更短的前缀码，因此不是 Huffman 码。

✍ **3.9** 对于一个码，如果没有一个码字是任何其他码字的后缀，则称该码为后缀码，试证后缀码是唯一可译的；对于任何随机变量，最短后缀码的平均码长等于相应 Huffman 码的平均码长。

证明 (1) 如果一个码是后缀码，即没有一个码字是任何其他码字的后缀，则这些码的后缀分解集当中必然不包含这些码字，故后缀码是唯一可译的。

(2)Huffman 码是使平均码长最短的编码方式，因此对于后缀码，其最短平均码长必定为相应 Huffman 码的平均码长。

✍ **3.10** 考虑一个概率分布为 $\begin{pmatrix} x_1 & x_2 & x_3 & x_4 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{4} & \frac{1}{12} \end{pmatrix}$ 的随机变量。

(a) 对该随机变量构成一个 Huffman 码。

(b) 证明存在 2 个不同的最佳码长集合，即证明码长集合为 $(1, 2, 3, 3)$ 和 $(2, 2, 2, 2)$ 都是最佳的。

(c) 证明某些符号的最佳码字长度可能超出 Shannon 编码码长 $\left\lceil \log \frac{1}{p(x)} \right\rceil$。

解 (a) 显然构成一个 Huffman 码为 $\{00, 01, 10, 11\}$。

(b) 由于 $\{00, 01, 10, 11\}$ 为一个 Huffman 码，因此其码长集合为最佳的，此时的平均码长为

$$\bar{n}_1 = \sum_{i=1}^{4} p_i n_{1i} = 2$$

而对于码长集合 $(1, 2, 3, 3)$，容易看出 $\{0, 11, 101, 110\}$ 是一个 Huffman 码，此时的平均码长为

$$\bar{n}_2 = \sum_{i=1}^{4} p_i n_{2i} = 2$$

因此码长集合为 $(1, 2, 3, 3)$ 和 $(2, 2, 2, 2)$ 都是最佳的。

(c) 对于给定信源计算 Shannon 编码的码长集合为 $(2, 2, 2, 4)$，因此对于最佳编码的码长集合 $(1, 2, 3, 3)$，其 x_3 的编码是超出了 Shannon 编码码长的，因此某些符号的最佳码字长度可能超出 Shannon 编码码长。

✍ **3.12** 令 DMS 为

$$U = \begin{pmatrix} a_1 & a_2 & a_3 & \cdots & a_K \\ p(a_1) \geq & p(a_2) \geq & p(a_3) \geq & \cdots & \geq p(a_K) \end{pmatrix}$$

定义

$$\begin{cases} Q_i = \sum_{k=1}^{i-1} p(a_k) \quad i > 1 \\ Q_1 = 0 \end{cases}$$

按如下方法进行二元编码，消息 a_k 用 Q_k 的二进制小数展开的截短来表示（例如 1/2 的二进制小数展开为 $0.100\cdots$，1/4 为 $0.01000\cdots$）截短的长度为保留小数点后 l_k 位，$l_k = \left\lceil \log \frac{1}{p(a_k)} \right\rceil$，$\lceil x \rceil$ 表示大于、等于 x 的最小整数。

(a) 对 DMS

$$U = \begin{pmatrix} a_1 & a_2 & a_3 & a_4 & a_5 & a_6 & a_7 & a_8 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{8} & \frac{1}{8} & \frac{1}{16} & \frac{1}{16} & \frac{1}{16} & \frac{1}{16} \end{pmatrix}$$

进行上述编码。

(b) 证明上述编码满足异字头条件，且平均码长 $\bar{n}$ 满足 $H(U) \le \bar{n} \le H(U) + 1$。

解 (a) 计算 Q 得

$$Q = \left\{0, \frac{1}{4}, \frac{1}{2}, \frac{5}{8}, \frac{3}{4}, \frac{13}{16}, \frac{7}{8}, \frac{15}{16}\right\}$$

因此可得各个码的码长

$$l = \{2, 2, 3, 3, 4, 4, 4, 4\}$$

因此编码为

$$C = \{00, 01, 100, 101, 1100, 1101, 1110, 1111\}$$

(b) 证明：1) 假设上述编码不满足异字头条件，则总存在 a_i 是 a_j 的前缀码，因此有 $i < j, l_i < l_j$ 且

$$a_j = a_i d_{l_i+1} \cdots d_{l_j}$$

因此有

$$Q_j - Q_i = 0.0 \cdots 0 d_{l_i+1} \cdots d_{l_j} \cdots < 2^{-l_i} \le p_i$$

但同时由已知得

$$Q_j - Q_i = \sum_{k=1}^{j-1} p(a_k) - \sum_{k=1}^{i-1} p(a_k) = \sum_{k=i}^{j-1} p_k > p_i$$

因此相互矛盾，故上述编码满足异字头条件。

2)

$$\begin{aligned} \bar{n} - H(U) &= \sum_{k=1}^{K} p_k l_k + \sum_{k=1}^{K} p_k \log p_k \\ &= \sum_{k=1}^{K} \left(l_k - \log \frac{1}{p_k} \right) \\ &= \sum_{k=1}^{K} p_k \left(\left\lceil \log \frac{1}{p_k} \right\rceil - \log \frac{1}{p_k} \right) \end{aligned}$$

由于 $\sum_{k=1}^{K} p_k = 1$，因此有 $0 \le \bar{n} - H(U) < 1$，即

$$H(U) \le \bar{n} < H(U) + 1$$

✍ **3.14** 假定 DMS 为

$$X=\begin{pmatrix}a_1 & a_2 & \cdots & a_m\\ p_1 & p_2 & \cdots & p_m\end{pmatrix}$$

令 l_i 表示对应于消息 a_i 的二元码字的长度，C_i 表示消息 a_i 重要性的加权，于是这个码的平均代价为

$$C=\sum_{i=1}^{m}p_il_i\cdot C_i$$

(a) 在 $\sum_{i=1}^{m}2^{-l_i}\le 1$ 的约束下最小化 C，求出最小化 C 的值 C^* 和响应的 $l_i^*, i=1,2,\cdots,m$(这里忽略对于 l_i 是整数的限制)。

(b) 如何利用 Huffman 编码方法对唯一可译码进行最小化 C，这个最小化 C 记为 C_{Huffman}。

(c) 证明

$$C^*\le C_{\text{Huffman}}\le C^*+\sum_{i=1}^{m}p_ic_i$$

解 (a) 已知最小值必定在等式约束下取到，否则总可以找到

$$l_i'=l_i-\log\frac{1}{\sum_{i=1}^{m}2^{-l_i}}$$

使得 $\sum_{i=1}^{m}2^{-l_i'}=1$，而

$$C'-C=\sum_{i=1}^{m}p_iC_i(l_i'-l_i)<0$$

因此最小值必定在等式约束下取到，故构造 Lagrangian 函数为

$$J(\boldsymbol{l},\lambda)=\sum_{i=1}^{m}p_iC_il_i+\lambda\left(1-\sum_{i=1}^{m}2^{-l_i}\right)$$

令其梯度为 0 有

$$\frac{\partial J(\boldsymbol{l},\lambda)}{\partial l_i}=p_ic_i-\lambda 2^{-l_i}\ln 2=0$$

$$\frac{\partial J(\boldsymbol{l},\lambda)}{\partial\lambda}=1-\sum_{i=1}^{m}2^{-l_i}=0$$

因此有

$$\lambda=\frac{1}{\ln 2}\sum_{i=1}^{m}p_iC_i$$

故

$$l_i^*=\log\frac{\sum_{i=1}^{m}p_iC_i}{p_iC_i}$$

$$C^*=\sum_{i=1}^{m}p_iC_i\log\frac{\sum_{k=1}^{m}p_kC_k}{p_iC_i}$$

(b) 记

$$q_i=\frac{p_iC_i}{\sum_{k=1}^{m}p_kC_k}$$

则 $U=\{q_i\}_{i=1}^m$ 构成一个概率分布，根据此概率分布构造 Huffman 编码即可，此时的平均码长为

$$\bar{n}=\sum_{i=1}^m q_i l_i=\frac{\sum_{i=1}^m p_i C_i l_i}{\sum_{k=1}^m p_k C_k}$$

因此有

$$C_{\text{Huffman}}=\sum_{i=1}^m p_i C_i l_i=\bar{n}\sum_{k=1}^m p_k C_k$$

(c) 对于 Huffman 编码，有 $H(U)\le\bar{n}<H(U)+1$，其中

$$H(U)=\sum_{i=1}^m \frac{p_i C_i}{\sum_{k=1}^m p_k C_k}\log\frac{\sum_{k=1}^m p_k C_k}{p_i C_i}$$

因此

$$C^*\le C_{\text{Huffman}}==\sum_{i=1}^m p_i C_i l_i=\bar{n}\sum_{k=1}^m p_k C_k<C^*+\sum_{i=1}^m p_i C_i$$

3.15 已知二元信源 $\{0,1\}, p_0=\frac{1}{8}, p_1=\frac{7}{8}$，试用算术码对以下序列进行编码

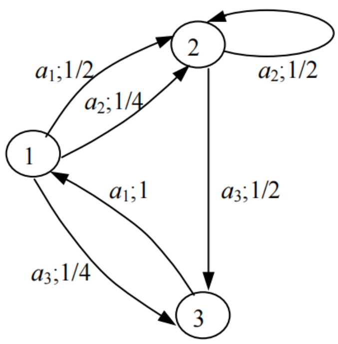
11111110111110

解 信源的算术编码为 即将 0 编码为 0001,1 编码为 10，因此序列的编码为

x	$p(x)$	$\bar{F}(x)$	$l(X)$	码字
0	0.125	$\frac{1}{16}$	4	0001
1	0.875	$\frac{9}{16}$	2	10

101010101010100001101010101000001

3.16 设一个马尔科夫信源，其状态图如下图所示。



(a) 求稳态下各状态概率 $q(i)$，以及各字母 a_i 的出现概率，$i=1,2,3$。

(b) 求 $H(U\mid s_i), i=1,2,3$。

(c) 求 $H_\infty(U)$。

(d) 对各状态 S_j 求最佳二元码。

(e) 计算平均码长。

解 (a) 状态转移概率满足

$$\begin{cases} q(1) = q(3) \\ q(2) = \frac{3}{4}q(1) + \frac{1}{2}q(2) \\ q(1) + q(2) + q(3) = 1 \end{cases}$$

解得 $q(1) = q(3) = \frac{2}{7}, q(2) = \frac{3}{7}$。因此

$$p(a_1) = q(1)p(a_1 \mid s_1) + q(3)p(a_1 \mid s_3) = \frac{3}{7}$$
$$p(a_2) = q(1)p(a_2 \mid s_1) + q(2)p(a_2 \mid s_2) = \frac{2}{7}$$
$$p(a_3) = q(1)p(a_3 \mid s_1) + q(2)p(a_3 \mid s_2) = \frac{2}{7}$$

(b)

$$H(U \mid s_i) = \sum_{j=1}^{3} p(a_j \mid s_i) \log \frac{1}{p(a_j \mid s_i)}$$

因此可求得 $H(U \mid s_1) = 1.5\text{bit}, H(U \mid s_2) = 1\text{bit}, H(U \mid s_3) = 0$。

(c)

$$H_\infty(U) = \sum_{i=1}^{3} q(i) H(U \mid s_i) = \frac{6}{7}\text{bit}$$

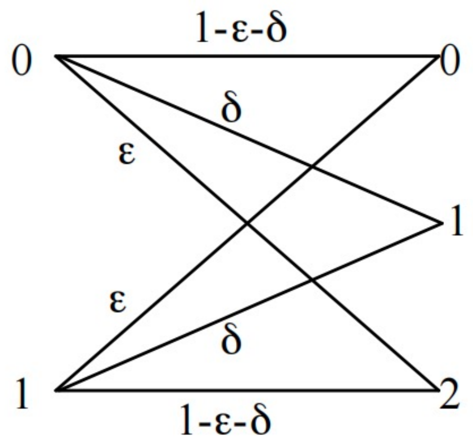
(d) 在 s_1 状态的 Huffman 码为 $\{1, 01, 00\}$,s_2 状态的 Huffman 码为 $\{1, 0\}$，s_3 状态无需编码。

(e)$\bar{n}_1 = 1.5, \bar{n}_2 = 1, \bar{n}_3 = 0$，因此

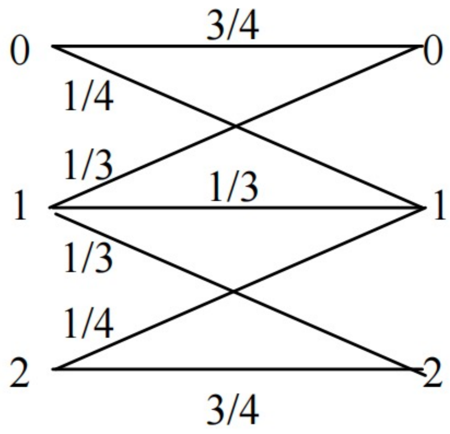
$$\bar{n} = \sum_{i=1}^{3} q(i)\bar{n}_i = \frac{6}{7}$$

第四章 信息传递-通信理论

✍ **4.1** 计算如图所示离散无记忆信道的容量。



(a)



(d)

解 (a) 信道的转移概率矩阵为

$$\boldsymbol{P}=\begin{bmatrix}1-\varepsilon-\delta & \delta & \varepsilon\\ \varepsilon & \delta & 1-\varepsilon-\delta\end{bmatrix}$$

可知这是一个准对称信道，其中

$$\mathscr{Y}_1=\{0,2\},\mathscr{Y}_2=\{1\}$$

因此当信道等概即 $P(X=0)=P(X=1)=0.5$ 时达到信道容量 C，此时

$$\begin{bmatrix}P(Y=0)\\P(Y=1)\\P(Y=2)\end{bmatrix}=\begin{bmatrix}1-\varepsilon-\delta & \varepsilon\\ \delta & \delta\\ \varepsilon & 1-\varepsilon-\delta\end{bmatrix}\begin{bmatrix}0.5\\0.5\end{bmatrix}=\begin{bmatrix}0.5-0.5\delta\\ \delta\\ 0.5-0.5\delta\end{bmatrix}$$

相应的信道容量为

$$\begin{aligned}C&=I(X=0;Y)=I(X=1;Y)\\&=\sum_{j=0}^{2}p(j\mid 0)\log\frac{p(j\mid 0)}{q(j)}\\&=(1-\varepsilon-\delta)\log\frac{1-\varepsilon-\delta}{(1-\delta)/2}+\delta\log\frac{\delta}{\delta}+\varepsilon\log\frac{\varepsilon}{(1-\delta)/2}\\&=(1-\varepsilon-\delta)\log(1-\varepsilon-\delta)+\varepsilon\log\varepsilon-(1-\delta)\log\frac{1-\delta}{2}\end{aligned}$$

(d) 信道的转移概率矩阵为

$$\boldsymbol{P}=\begin{bmatrix}\frac{3}{4} & \frac{1}{4} & 0\\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3}\\ 0 & \frac{1}{4} & \frac{3}{4}\end{bmatrix}$$

注意到当 $P(X=0)=P(X=2)=0.5, P(X=1)=0$ 时，有

$$p(\boldsymbol{y})=\boldsymbol{P}^{\mathrm{T}}p(\boldsymbol{x})=\left[\frac{3}{8},\frac{1}{4},\frac{3}{8}\right]^{\mathrm{T}}$$

此时有

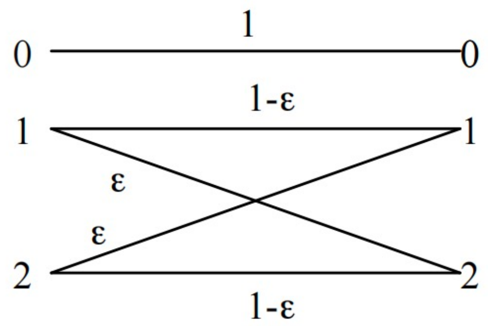
$$I(X=0;Y)=I(X=2;Y)=\sum_{j=0}^{2}p(j\mid 0)\log\frac{p(j\mid 0)}{q(j)}=0.75\mathrm{bit}$$

$$I(X=1;Y)=\sum_{j=0}^{2}p(j\mid 1)\log\frac{p(j\mid 1)}{q(j)}=\left(\frac{2}{3}\log\frac{8}{9}+\frac{1}{3}\log\frac{4}{3}\right)\mathrm{bit}<0.75\mathrm{bit}$$

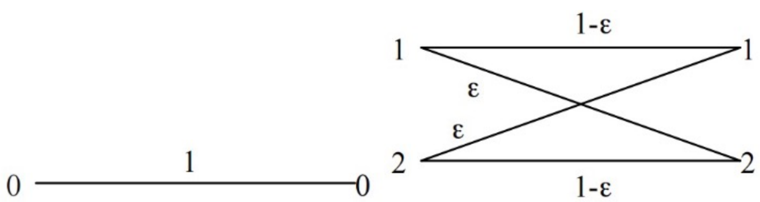
因此由定理 4.2.2 可知

$$C=I(X=0;Y)=I(X=2;Y)=0.75\mathrm{bit}$$

4.2 计算如图所示的信道的容量。



解 可以将信道看成如下图所示的 C_1 和 C_2 的和信道：



(a) C_1 (b) C_2

显然有信道容量 $C_1=0$，而 C_2 是一个二元对称信道，因此在 $P(X=1)=P(X=2)=0.5$ 时信道容量最大，此时有 $P(Y=1)=P(Y=2)=0.5(1-\varepsilon)+0.5\varepsilon=0.5$，因此

$$\begin{aligned}C_2&=(1-\varepsilon)\log\frac{1-\varepsilon}{0.5}+\varepsilon\log\frac{\varepsilon}{0.5}\\&=(1-\varepsilon)+(1-\varepsilon)\log(1-\varepsilon)+\varepsilon+\varepsilon\log\varepsilon\\&=1-H(\varepsilon)\end{aligned}$$

由和信道的性质，有

$$C=\log\left(2^{C_1}+2^{C_2}\right)=\log\left[1+2^{1-H(\varepsilon)}\right]$$

达到信道容量的输入分布为

$$P(X=0)=2^{C_1-C}=\frac{1}{1+2^{1-H(\varepsilon)}}$$

$$P(X=1)=P(X=2)=0.5\cdot 2^{C_2-C}=\frac{2^{-H(\varepsilon)}}{1+2^{1-H(\varepsilon)}}$$

4.3 考虑离散无记忆信道 $Y = (X + Z)(\mod 11)$，其中

$$Z = \begin{bmatrix} 1 & 2 & 3 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}, X \in \{0, 1, \cdots, 10\}$$

假定 Z 和 X 独立，求

(a) 信道容量。

(b) 达到信道容量的输入分布 $\{P^*(x)\}$。

解 易知信道为对称信道，因此当输入分布满足

$$P(X = i) = \frac{1}{11}, i \in \{0, 1, \cdots, 10\}$$

时达到信道容量，信道容量为

$$C = \log 11 - \log 3 = \log \frac{11}{3} \text{bit/次}$$

4.4 用差错率为 ε 的二元对称信道 BSC 构成如下复合信道：

(a) 长度为 N 的级联信道，如图 (a) 所示。求该级联信道的容量 C_N，并证明 $\lim\limits_{N\to\infty} C_N = 0$。

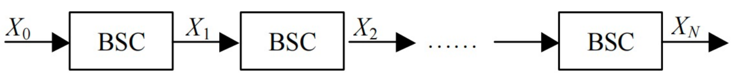


图 (a)

(b) 并联输入信道，把输入 X 并联接到各信道如图(b)所示，输出是矢量，当 $N \to \infty$ 时并联输入信道容量趋于 1。

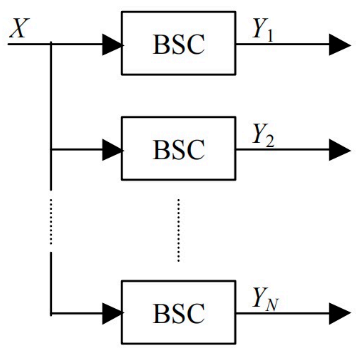


图 (b)

(c)N 个相同的 BSC 的积信道，如图(c)所示，求这时积信道容量 C_N，并证明 $\lim\limits_{N\to\infty} C_N = \infty$。

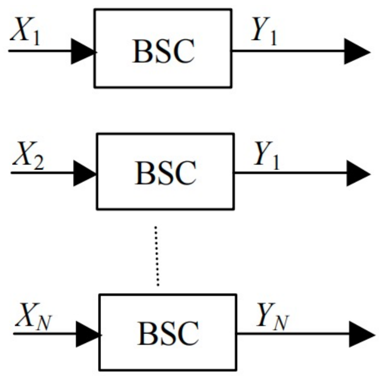


图 (c)

解 (a) 每个二元对称信道的转移概率矩阵为

$$P_0 = \begin{bmatrix} 1-\varepsilon & \varepsilon \\ \varepsilon & 1-\varepsilon \end{bmatrix}$$

则 N 个相同的二元对称信道的转移概率矩阵为

$$P = P_0^N$$

设 P_0 的 EVD 为 $P_0 = U^{\mathrm{T}} \Sigma U$，容易求得

$$U = \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix}, \Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1-2\varepsilon \end{bmatrix}$$

因此有

$$P = P_0^N = U^{\mathrm{T}} \Sigma^N U = U^{\mathrm{T}} \begin{bmatrix} 1 & 0 \\ 0 & (1-2\varepsilon)^N \end{bmatrix} U = \begin{bmatrix} \frac{1+(1-2\varepsilon)^N}{2} & \frac{1-(1-2\varepsilon)^N}{2} \\ \frac{1-(1-2\varepsilon)^N}{2} & \frac{1+(1-2\varepsilon)^N}{2} \end{bmatrix}$$

注意到这是一个差错率为 $\frac{1-(1-2\varepsilon)^N}{2}$ 的二元对称信道，于是级联信道的容量为

$$C_N = 1 - H\left(\frac{1-(1-2\varepsilon)^N}{2}\right)$$

因此

$$\lim_{N\to\infty} C_N = 1 - H\left(\frac{1}{2}\right) = 0$$

(b) 记随机向量 $Y = [Y_1, Y_2, \cdots, Y_N]^{\mathrm{T}} \in \mathbb{R}^N$，则 Y 的取值共有 2^N 种情况。记 $k = (Y_1 Y_2 \cdots Y_N)_2$，$k = [Y_1, Y_2, \cdots, Y_N]^{\mathrm{T}}_{k=(Y_1Y_2\cdots Y_N)_2}$, l_k 为 k 中 1 的个数，$k \in \{0, 1, \cdots 2^N - 1\}$，那么则有

$$P(Y = k \mid X = 0) = (1-\varepsilon)^{N-l_k} \varepsilon^{l_k}$$

$$P(Y = k \mid X = 1) = \varepsilon^{N-l_k} (1-\varepsilon)^{l_k}$$

若把 $\mathscr{Y}$ 看成 $\mathscr{Y}_i = \{k_i, k_{\bar{i}}\}, i \in \{0, 1, \cdots, 2^{N-1} - 1\}$ 之并，其中 $\bar{i}$ 是 i 的二进制取反运算，则显然每一个 $\mathscr{Y}_i$ 都是对称的，因此 $\mathscr{Y}$ 满足准对称条件，则当输入为等概分布即 $P(X = 0) = P(X = 1) = 0.5$ 时达到信道容量，此时有

$$P(Y = k) = 0.5(1-\varepsilon)^{N-l_k} \varepsilon^{l_k} + 0.5\varepsilon^{N-l_k} (1-\varepsilon)^{l_k}$$

因此可求信道容量

$$\begin{aligned} C_N = I(X = 0; Y) = I(X = 1; Y) &= \sum_{k=0}^{2^N-1} P(Y = k \mid X = 0) \log \frac{P(Y = k \mid X = 0)}{P(Y = k)} \\ &= \sum_{k=0}^{2^N-1} (1-\varepsilon)^{N-l_k} \varepsilon^{l_k} \log \frac{(1-\varepsilon)^{N-l_k} \varepsilon^{l_k}}{0.5(1-\varepsilon)^{N-l_k} \varepsilon^{l_k} + 0.5\varepsilon^{N-l_k} (1-\varepsilon)^{l_k}} \\ &= \sum_{k=0}^{2^N-1} (1-\varepsilon)^{N-l_k} \varepsilon^{l_k} + \sum_{k=0}^{2^N-1} (1-\varepsilon)^{N-l_k} \varepsilon^{l_k} \log \frac{(1-\varepsilon)^{N-l_k} \varepsilon^{l_k}}{(1-\varepsilon)^{N-l_k} \varepsilon^{l_k} + \varepsilon^{N-l_k} (1-\varepsilon)^{l_k}} \\ &= 1 + \sum_{k=0}^{2^N-1} (1-\varepsilon)^{N-l_k} \varepsilon^{l_k} \log \frac{(1-\varepsilon)^{N-l_k} \varepsilon^{l_k}}{(1-\varepsilon)^{N-l_k} \varepsilon^{l_k} + \varepsilon^{N-l_k} (1-\varepsilon)^{l_k}} \end{aligned}$$

注意到

$$\lim_{N\to\infty} (1-\varepsilon)^{N-l_k} \varepsilon^{l_k} = 0$$

因此

$$\lim_{N\to\infty} C_N = 1$$

(c) 由积信道容量可知

$$C_N = NC_0 = N[1 - H(\varepsilon)]$$

因此

$$\lim_{N\to\infty} C_N = \infty$$

4.9 计算下述信道容量。

$$\boldsymbol{P} = \begin{bmatrix} \bar{p} & p & 0 & 0 \\ p & \bar{p} & 0 & 0 \\ 0 & 0 & \bar{p} & p \\ 0 & 0 & p & \bar{p} \end{bmatrix}$$

解 可以看成两个差错率为 p 的二元对称信道的和信道，每个信道的容量为

$$C_1 = C_2 = 1 - H(p)$$

因此和信道的信道容量为

$$C = \log\left(2^{C_1} + 2^{C_2}\right) = 2 - H(p)$$

当输入为等概分布即 $P(X=0) = P(X=1) = P(X=2) = P(X=3) = 0.25$ 时达到信道容量。

4.12 有一个叠加噪声的信道，输入 X 是离散随机变量，取值 $\{+1, -1\}$，噪声 N 的概率密度为

$$P_N(n) = \begin{cases} \frac{1}{4} & |n| \le 2 \\ 0 & |n| > 2 \end{cases}$$

信道输出 $Y = X + N$ 是连续随机变量，求

(a) 该半连续信道容量 $C = \max\limits_{\{p(x)\}} I(X;Y)$

(b) 若在信道输出上接一个检测器作为这个信道的一部分，检测器输出变量为

$$Z = \begin{cases} 1 & Y > 1 \\ 0 & -1 \le Y \le 1 \\ -1 & Y < -1 \end{cases}$$

这样构成离散信道，求它的容量。

解 (a) 显然信道对称，因此当输入分布为 $P(X=-1) = P(X=1) = 0.5$ 时达到信道容量，此时的输出的概率密度为

$$p(y) = \begin{cases} \frac{1}{8} & 1 \le y \le 3 \\ \frac{1}{4} & -1 \le y \le 1 \\ \frac{1}{8} & -3 \le y \le -1 \\ 0 & \text{otherwise} \end{cases}$$

在 $X = 1$ 的条件下输出的概率密度为

$$p(y \mid 1) = \begin{cases} \frac{1}{4} & -1 \le y \le 3 \\ 0 & \text{otherwise} \end{cases}$$

因此信道容量为

$$C=I(Y;X=1)=H(Y)-H(Y\mid X=1)=\int_1^3\frac{1}{8}\log 8\mathrm{d}x+\int_{-1}^1\frac{1}{4}\log 4\mathrm{d}x+\int_{-3}^{-1}\frac{1}{8}\log 8\mathrm{d}x-\int_{-1}^3\frac{1}{4}\log 4\mathrm{d}x=0.5\mathrm{bit}$$

(b) 仍然在输入等概分布时达到信道容量，此时 Z 的分布为 $P(Z=1)=\frac{1}{4},P(Z=0)=\frac{1}{2},P(Z=-1)=\frac{1}{4}$，且在 $X=1$ 时的条件概率分布为 $P(Z=0\mid X=1)=P(Z=1\mid X=1)=0.5$, 因此有

$$C=I(Z;X=1)=H(Z)-H(Z\mid X=1)=\frac{1}{4}\log 4+\frac{1}{2}\log 2+\frac{1}{4}\log 4-\frac{1}{2}\log 2-\frac{1}{2}\log 2=0.5\mathrm{bit}$$

4.13 设一个连续信道，传送相位信息 $X\in(-\pi,\pi)$，信道受到加性噪声 Z 的干扰，而 Z 的概率密度为

$$p(z)=\begin{cases}\dfrac{1}{2a} & |z|\le a,a>0\\ 0 & |z|>a\end{cases}$$

输出 $Y=(X+Z)\mod 2\pi$，试求

(a)$a>\pi$ 时信道容量。

(b)$a\le\pi$ 时信道容量。

解 令 $Y'=X+Z$，则 $I(X;Y)$ 和 $I(X;Y')$ 在相同的输入分布下取最大值，而 Y' 的输出范围为 $(-\pi-a,\pi+a)$，因此有

$$I(X;Y')=H(Y')-H(Y'\mid X)=H(Y')-H(Z)$$

因此在 Y' 也为均匀分布时，$I(X;Y')$ 取最大值，此时

$$p(y')=\begin{cases}\dfrac{1}{2(\pi+a)} & |y'|\le\pi+a\\ 0 & |y'|>\pi+a\end{cases}$$

(a)$a>\pi$ 时，Y 的概率密度为

$$p(y)=\begin{cases}\dfrac{3}{2(\pi+a)} & 0<y<a-\pi,3\pi-a<y<2\pi\\ \dfrac{1}{\pi+a} & a-\pi<y<3\pi-a\\ 0 & \text{otherwise}\end{cases}$$

因此信道容量为

$$\begin{aligned}C&=\max\left[H(Y)-H(Y\mid X)\right]=\max\left[H(Y)-H(Z)\right]\\&=\int_0^{a-\pi}\frac{3}{2(a+\pi)}\log\frac{2(\pi+a)}{3}\mathrm{d}x+\int_{3\pi-a}^{2\pi}\frac{3}{2(a+\pi)}\log\frac{2(\pi+a)}{3}\mathrm{d}x+\int_{a-\pi}^{3\pi-a}\frac{1}{\pi+a}\log(\pi+a)\mathrm{d}x-\int_{-a}^{a}\frac{1}{2a}\log(2a)\mathrm{d}x\\&=\frac{3(a-\pi)}{a+\pi}\log\frac{2(a+\pi)}{3}+\frac{2(2\pi-a)}{a+\pi}\log(a+\pi)-\log(2a)\end{aligned}$$

(b)$a\le\pi$ 时，Y 的概率密度为

$$p(y)=\begin{cases}\dfrac{1}{2(\pi+a)} & 0<y<\pi-a,a+\pi<y<2\pi\\ \dfrac{1}{\pi+a} & \pi-a<y<\pi+a\\ 0 & \text{otherwise}\end{cases}$$

因此信道容量为

$$\begin{aligned}C&=\max\left[H(Y)-H(Y\mid X)\right]=\max\left[H(Y)-H(Z)\right]\\&=\int_0^{\pi-a}\frac{1}{2(a+\pi)}\log 2(\pi+a)\mathrm{d}x+\int_{a+\pi}^{2\pi}\frac{1}{2(a+\pi)}\log 2(\pi+a)\mathrm{d}x+\int_{a-\pi}^{a+\pi}\frac{1}{\pi+a}\log(\pi+a)\mathrm{d}x-\int_{-a}^{a}\frac{1}{2a}\log(2a)\mathrm{d}x\\&=\frac{\pi-a}{\pi+a}\log 2(a+\pi)+\frac{2a}{a+\pi}\log(a+\pi)-\log(2a)\end{aligned}$$

✍ **4.14** 给定系统带宽为 W，噪声功率谱密度为 N_0，试证明传送 1bit 信息所需最小能量为 $0.693N_0$(W)。如果要求 $\dfrac{P}{WN_0} > 4000$，其中 P 为信号功率，试证明所需的信号能量至少为此最小值的 482 倍。

证明 在频带效率 η 下，每传 1bit 信息所需能量 $E_b(\eta)$ 必须满足

$$\frac{E_b(\eta)}{N_0} \geq \frac{2^\eta - 1}{\eta}$$

因此有

$$\lim_{\eta \to 0} E_b(\eta) \geq N_0 \lim_{\eta \to 0} \frac{2^\eta - 1}{\eta} = N_0 \ln 2 = 0.693N_0$$

因此传送 1bit 信息所需最小能量为 $0.693N_0$(W)。

设传送 1bit 信息需要时间为 T，则 T 时间的信道容量为

$$C_T = WT \log\left\{1 + \frac{P}{N_0 W}\right\} = 1\text{bit}$$

又 $E_b = PT$，则有

$$\frac{E_b}{WTN_0} > 4000$$

因此

$$E_b > 4000WTN_0 = \frac{4000N_0}{\log\left\{1 + \dfrac{P}{N_0 W}\right\}}$$

故

$$\frac{E_b}{E_{\min}} > \frac{4000}{\ln 2 \log 4001} = 482$$

即所需的信号能量至少为此最小值的 482 倍。

第五章 信息的计算

✍ **1** 令 x, y 为任意长度的 $0, 1$ 序列，证明

$$K(x, y) \leq K(x) + K(y) + c$$

证明 假设有两个序列 x, y 分别具有 Kolmogorov 复杂度 $k(x), K(y)$，则一定存在长度分别为 $K(x)$ 和 $K(y)$ 的程序 p_x 和 p_y，分别输出 x, y。那么可以形成如下的程序：

运行以下两个程序，不要在第一个程序之后停止；
运行程序 p_x，结束后跳转到下一步；
运行程序 p_y；

这个程序的长度即为 $K(x) + K(y) + c$，且输出 x, y。因此

$$K(x, y) \leq K(x) + K(y) + c$$

✍ **2** 整数和的复杂度

a) 证明：$K(n) \leq \log n + 2 \log \log n + c$

b) 证明：$K(n_1 + n_2) \leq K(n_1) + K(n_2) + c$

证明 a) 为了描述一个整数 n，则要将 n 的长度和二进制位输入。因此程序将是自定界的。为了表示 n 的长度，即 $\log n$，可以将 $\log n$ 的每个位重复两次，并以 10 结束。这种表示法需要 $2 \log \log n + 2$ 位。它需要 $\log n$ 位来表示 n 的位，因此程序的总长度是 $\log n + 2 \log \log n + c$，这是 n 的复杂度上界。因此有

$$K(n) \leq \log n + 2 \log \log n + c$$

b) 给定两个程序输出 n_1 和 n_2，添加一条指令，将这两个数相加并输出。这个程序的长度为 $K(n_1) + K(n_2) + c$，因此有

$$K(n_1 + n_2) \leq K(n_1) + K(n_2) + c$$

✍ **3** 给定两个正态分布 $p(x \mid C_1) \sim \mathcal{N}(\mu_1, \sigma_1^2), p(x \mid C_2) \sim \mathcal{N}(\mu_2, \sigma_2^2), p(C_1)$ 和 $p(C_2)$，计算贝叶斯判别点。

解 贝叶斯判别点应满足条件

$$p(x \mid C_1) p(C_1) = p(x \mid C_2) p(C_2)$$

注意到 $p(x \mid C_i), i \in \{1, 2\}$ 的概率密度满足

$$p(x \mid C_i) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left[-\frac{(x - \mu_i)^2}{2\sigma_i^2}\right]$$

因此对等式取对数以便计算

$$\ln p(x \mid C_1) + \ln p(C_1) = \ln p(x \mid C_2) + \ln p(C_2)$$

因此有

$$-\frac{1}{2} \ln 2\pi - \ln \sigma_1 - \frac{(x - \mu_1)^2}{2\sigma_1^2} + \ln p(C_1) = -\frac{1}{2} \ln 2\pi - \ln \sigma_2 - \frac{(x - \mu_2)^2}{2\sigma_2^2} + \ln p(C_2)$$

化简得

$$(\sigma_1^2 - \sigma_2^2)x^2 - 2(\mu_2\sigma_1^2 - \mu_1\sigma_2^2)x + \sigma_1^2\mu_2^2 - \sigma_2^2\mu_1^2 - 2\sigma_1^2\sigma_2^2 \ln \frac{p(C_2)\sigma_1}{p(C_1)\sigma_2} = 0$$

所求的贝叶斯判别点即为该方程的根。

✍ **4** 设计一个两输入的感知器来实现布尔函数 $A \wedge \neg B$。

解 由真值表可知，只有在输入向量 $\boldsymbol{x}=[x_1,x_2]^{\mathrm{T}}=[1,0]^{\mathrm{T}}$ 时输出 1，否则输出 0，因此可构造如下形式的感知器：

$$y=\frac{1+\operatorname{sgn}(\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}-b)}{2}$$

那么感知器需要满足条件：

$$\begin{cases}\operatorname{sgn}(\omega_1-b)=1\\ \operatorname{sgn}(\omega_2-b)=-1\\ \operatorname{sgn}(-b)=-1\\ \operatorname{sgn}(\omega_1+\omega_2-b)=-1\end{cases}$$

因此有 $\omega_1>b,\omega_2<b,b>0,\omega_1+\omega_2-b<0$。例如取 $\omega_1=0.6,\omega_2=-0.4,b=0.4$ 即满足条件，因此可构造感知器为

$$y=\frac{1+\operatorname{sgn}(0.6x_1-0.4x_2-0.4)}{2}$$

✍ **5** 设计一个两层的感知器网络来实现异或布尔函数 $A\mathrm{XOR}B$。提示：$A\mathrm{XOR}B=(A\wedge\neg B)\vee(\neg A\wedge B)$

解 在上题中已经将第一层感知器设计完成记为 $y_1(x_1,x_2)$，因此只需要设计第二层感知器，即实现布尔函数 $A\vee B$。与上题同理，可构造该感知器为

$$y_2(x_1,x_2)=\frac{1+\operatorname{sgn}(0.5x_1+0.5x_2-0.5)}{2}$$

因此该二层感知器网络可构造为

$$y(x_1,x_2)=y_2(y_1(x_1,x_2),y_1(x_2,x_1))$$

✍ **6** 推导输出为 o 的单个单元的梯度下降训练法则，其中

$$o=\omega_0+\omega_1x_1+\omega_1x_1^2+\cdots+\omega_nx_n+\omega_nx_n^2$$

解 记 $\boldsymbol{\omega}=[\omega_1,\omega_2,\cdots,\omega_n]^{\mathrm{T}},\boldsymbol{x}=[x_1,x_2,\cdots,x_n]^{\mathrm{T}},\boldsymbol{W}=\operatorname{diag}(\omega_1,\omega_2,\cdots,\omega_n)$ 将输出写成向量的形式有

$$o=\omega_0+\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}+\boldsymbol{x}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{x}$$

因此可求 o 的梯度为

$$\nabla o=\boldsymbol{\omega}+2\boldsymbol{W}\boldsymbol{x}$$

因此训练法则为

$$\boldsymbol{x}_{k+1}=\boldsymbol{x}_k-\mu(\boldsymbol{\omega}+2\boldsymbol{W}\boldsymbol{x}_k)$$

第六章 信息控制理论

✍ **1** 图(a)为新型自平衡的站立骑行车 (HTV)，参照闭环框图，描述一个闭环反馈控制系统，协助 HTV 的骑手在车上保持平衡和机动。

图 (a)

解 如图(b)所示

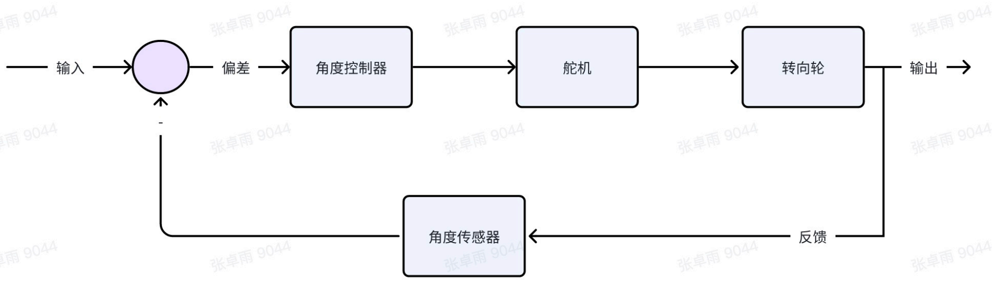


图 (b)

✍ **2** 某单输入-输出系统的状态空间模型为:

$$\dot{\boldsymbol{x}} = \begin{bmatrix} 0 & 1 \\ -1 & -2 \end{bmatrix} \boldsymbol{x} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \boldsymbol{u}, y = \begin{bmatrix} 1 & 1 \end{bmatrix} \boldsymbol{x}$$

1) 试求解该系统的传递函数 $G(s) = Y(s)/U(s)$?

2) 判断系统是否能控、能观?

解 1) 拉普拉斯变换得

$$s\boldsymbol{X}(s) = \begin{bmatrix} 0 & 1 \\ -1 & -2 \end{bmatrix} \boldsymbol{X}(s) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \boldsymbol{U}(s)$$

因此有

$$G(s)=\frac{Y(s)}{U(s)}=\begin{bmatrix}1 & 1\end{bmatrix}\begin{bmatrix}s & -1\\ 1 & s+2\end{bmatrix}^{-1}\begin{bmatrix}0\\1\end{bmatrix}=\frac{1}{s+1}$$

2) 记

$$\boldsymbol{A}=\begin{bmatrix}0 & 1\\ -1 & -2\end{bmatrix},\boldsymbol{B}=\begin{bmatrix}0\\1\end{bmatrix}$$

易知 $\operatorname{rank}([\boldsymbol{B},\boldsymbol{AB}])=2$，因此是能控的。

$$\operatorname{rank}\left(\begin{bmatrix}\boldsymbol{C}\\ \boldsymbol{CA}\end{bmatrix}\right)=1$$

✍ **3** 利用劳斯判据，判断如下特征多项式对应的系统是否稳定：

$$p_1(s)=s^2+10s+5=0$$
$$p_2(s)=s^4+s^3+5s^2+20s+10=0$$

解 (1)Routh 表如下：

$$\begin{array}{c|cc} s^2 & 1 & 5\\ s^1 & 10 & \\ s^0 & 5 & \end{array}$$

第一列均大于 0，因此系统稳定。

(2)Routh 表如下：

$$\begin{array}{c|ccc} s^4 & 1 & 5 & 10\\ s^3 & 1 & 20 & \\ s^2 & -15 & 0 & \\ s^1 & 20 & & \\ s^0 & 10 & & \end{array}$$

第一列出现小于 0 的元素，因此系统不稳定。

✍ **4** 利用李雅普诺夫第二方法判断下列线性系统的稳定性（假设 $V(\boldsymbol{x})=\dfrac{x_1^2+x_2^2}{2}$）

$$(1)\begin{bmatrix}\dot{x}_1\\ \dot{x}_2\end{bmatrix}=\begin{bmatrix}0 & 1\\ -1 & -1\end{bmatrix}\begin{bmatrix}x_1\\ x_2\end{bmatrix};\quad (2)\begin{bmatrix}\dot{x}_1\\ \dot{x}_2\end{bmatrix}=\begin{bmatrix}-1 & 1\\ 2 & -3\end{bmatrix}\begin{bmatrix}x_1\\ x_2\end{bmatrix}$$

解 由题可知 $V(\boldsymbol{x})$ 正定，$\dot{V}(\boldsymbol{x})=x_1\dot{x}_1+x_2\dot{x}_2$，因此有

(1)$V(\boldsymbol{x})=-x_2^2$ 负定，因此系统稳定；

(2)$V(\boldsymbol{x})=-x_1^2+3x_1x_2-3x_2^2$ 不定，因此系统不稳定。