第二章课后习题

【2.1】设有 12 枚同值硬币，其中有一枚为假币。只知道假币的重量与真币的重量不同，但不知究竟是重还是轻。现用比较天平左右两边轻重的方法来测量。为了在天平上称出哪一枚是假币，试问至少必须称多少次？

解：从信息论的角度看，

“12 枚硬币中，某一枚为假币”该事件发生的概率为 $P=\frac{1}{12}$；

“假币的重量比真的轻，或重”该事件发生的概率为 $P=\frac{1}{2}$；

为确定哪一枚是假币，即要消除上述两事件的联合不确定性，由于二者是独立的，因此有

$$I=\log 12+\log 2=\log 24 \text{ 比特}$$

而用天平称时，有三种可能性：重、轻、相等，三者是等概率的，均为 $P=\frac{1}{3}$，因此天平每一次消除的不确定性为 $I=\log 3$ 比特

因此，必须称的次数为

$$\frac{I_1}{I_2}=\frac{\log 24}{\log 3}\approx 2.9 \text{ 次}$$

因此，至少需称 3 次。

【延伸】如何测量？分 3 堆，每堆 4 枚，经过 3 次测量能否测出哪一枚为假币。

【2.2】同时扔一对均匀的骰子，当得知“两骰子面朝上点数之和为 2”或“面朝上点数之和为 8”或“两骰子面朝上点数是 3 和 4”时，试问这三种情况分别获得多少信息量？

解：

“两骰子总点数之和为 2”有一种可能，即两骰子的点数各为 1，由于二者是独立的，因此该种情况发生的概率为 $P=\frac{1}{6}\times\frac{1}{6}=\frac{1}{36}$，该事件的信息量为：

$$I = \log 36 \approx 5.17 \text{ 比特}$$

“两骰子总点数之和为 8”共有如下可能：2 和 6、3 和 5、4 和 4、5 和 3、6 和 2，概率为 $P = \frac{1}{6} \times \frac{1}{6} \times 5 = \frac{5}{36}$，因此该事件的信息量为：

$$I = \log \frac{36}{5} \approx 2.85 \text{ 比特}$$

“两骰子面朝上点数是 3 和 4”的可能性有两种：3 和 4、4 和 3，概率为 $P = \frac{1}{6} \times \frac{1}{6} \times 2 = \frac{1}{18}$，因此该事件的信息量为：

$$I = \log 18 \approx 4.17 \text{ 比特}$$

【2.3】如果你在不知道今天是星期几的情况下问你的朋友“明天星期几？”则答案中含有多少信息量？如果你在已知今天是星期四的情况下提出同样的问题，则答案中你能获得多少信息量（假设已知星期一至星期日的顺序）？

解：

如果不知今天星期几时间的话，答案可能有七种可能性，每一种都是等概率的，均为 $P = \frac{1}{7}$，因此此时从答案中获得的信息量为

$$I = \log 7 = 2.807 \text{ 比特}$$

而当已知今天星期几时间同样的问题，其可能性只有一种，即发生的概率为 1，此时获得的信息量为 0 比特。

【2.4】居住某地区的女孩中有 25%是大学生，在女大学生中有 75%是身高 1.6 米以上的，而女孩中身高 1.6 米以上的占总数一半。假如我们得知“身高 1.6 米以上的某女孩是大学生”的消息，问获得多少信息量？

解：

设 A 表示女孩是大学生，$P(A) = 0.25$；

B 表示女孩身高 1.6 米以上，$P(B \mid A) = 0.75$，$P(B) = 0.5$

“身高 1.6 米以上的某女孩是大学生”的发生概率为

$$P(A|B)=\frac{P(AB)}{P(B)}=\frac{P(A)P(B|A)}{P(B)}=\frac{0.25\times0.75}{0.5}=0.375$$

已知该事件所能获得的信息量为

$$I=\log\frac{1}{0.375}\approx1.415\text{ 比特}$$

【 】设离散无记忆信源 $\begin{bmatrix} X \\ P(x) \end{bmatrix}=\begin{bmatrix} a_1=0 & a_2=1 & a_3=2 & a_4=3 \\ 3/8 & 1/4 & 1/4 & 1/8 \end{bmatrix}$，其发出的消息为

（202120130213001203210110321010021032011223210），求

（1） 此消息的自信息是多少?

（2） 在此消息中平均每个符号携带的信息量是多少?

解:

信源是无记忆的，因此，发出的各消息之间是互相独立的，此时发出的消息的自信息即为各消息的自信息之和。根据已知条件，发出各消息所包含的信息量分别为:

$$I(a_0=0)=\log\frac{8}{3}=1.415\text{ 比特}$$

$$I(a_1=1)=\log 4=2\text{ 比特}$$

$$I(a_2=2)=\log 4=2\text{ 比特}$$

$$I(a_3=3)=\log 8=3\text{ 比特}$$

在发出的消息中，共有 14 个“0”符号，13 个“1”符号，12 个“2”符号，6 个“3”符号，则得到消息的自信息为:

$$I=14\times1.415+13\times2+12\times2+6\times3\approx87.81\text{ 比特}$$

45 个符号共携带 87.81 比特的信息量，平均每个符号携带的信息量为

$$I=\frac{87.81}{45}=1.95\text{ 比特/符号}$$

注意：消息中平均每个符号携带的信息量有别于离散平均无记忆信源平均每个符号携带的信息量，后者是信息熵，可计算得

$$H(X)=-\sum P(x)\log P(x)=1.91\text{ 比特/符号}$$

【2.6】如有 6 行 8 列的棋型方格，若有二个质点 A 和 B，分别以等概率落入任一方格内，且它们的坐标分别为（X_A,Y_A）和（X_B,Y_B），但 A 和 B 不能落入同一方格内。

（1） 若仅有质点 A，求 A 落入任一个格的平均自信息量是多少?

（2） 若已知 A 已落入，求 B 落入的平均自信息量。

（3） 若 A、B 是可分辨的，求 A、B 同都落入的平均自信息量。

解:

（1）求质点 A 落入任一格的平均自信息量，即求信息熵，首先得出质点 A 落入任一格的概率空间为:

$$\begin{bmatrix} X \\ P \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & a_3 & \cdots & a_{48} \\ \frac{1}{48} & \frac{1}{48} & \frac{1}{48} & \cdots & \frac{1}{48} \end{bmatrix}$$

平均自信息量为

$$H(A) = \log 48 = 5.58 \text{ 比特/符号}$$

（2）已知质点 A 已落入，求 B 落入的平均自信息量，即求 $H(B|A)$。

A 已落入，B 落入的格可能有 47 个，条件概率 $P(b_j|a_i)$ 均为 $\frac{1}{47}$。平均自信息量为

$$H(B|A) = -\sum_{i=1}^{48}\sum_{j=1}^{47} P(a_i)P(b_j|a_i)\log P(b_j|a_i) = \log 47 = 5.55 \text{ 比特/符号}$$

（3）质点 A 和 B 同时落入的平均自信息量为

$$H(AB) = H(A) + H(B|A) = 11.13 \text{ 比特/符号}$$

【2.7】从大量统计资料知道，男性中红绿色盲的发病率为 7%，女性发病率为 0.5%，如果你问一位男同志:“你是否是红绿色盲？”，他的回答可能是“是”，也可能是“否”，问这两个回答中各含有多少信息量？平均每个回答中含有多少信息量？如果你问一位女同志，则答案中含有的平均自信息量是多少?

解:

男同志红绿色盲的概率空间为：

$$\begin{bmatrix} X \\ P \end{bmatrix} = \begin{bmatrix} a_1 & a_2 \\ 0.07 & 0.93 \end{bmatrix}$$

问男同志回答“是”所获昨的信息量为：

$$I = \log \frac{1}{0.07} \approx 3.836 \text{ 比特/符号}$$

问男同志回答“否”所获得的信息量为：

$$I = \log \frac{1}{0.93} \approx 0.105 \text{ 比特/符号}$$

男同志平均每个回答中含有的信息量为

$$H(X) = -\sum P(x) \log P(x) = 0.366 \text{ 比特/符号}$$

同样，女同志红绿色盲的概率空间为

$$\begin{bmatrix} \mathrm{Y} \\ P \end{bmatrix} = \begin{bmatrix} b_1 & b_2 \\ 0.005 & 0.995 \end{bmatrix}$$

问女同志回答“是”所获昨的信息量为：

$$I = \log \frac{1}{0.005} \approx 7.64 \text{ 比特/符号}$$

问女同志回答“否”所获昨的信息量为：

$$I = \log \frac{1}{0.995} \approx 7.23 \times 10^{-3} \text{ 比特/符号}$$

女同志平均每个回答中含有的信息量为

$$H(Y) = -\sum P(x) \log P(x) = 0.045 \text{ 比特/符号}$$

【2.8】设信源 $\begin{bmatrix} X \\ P(x) \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & a_3 & a_4 & a_5 & a_6 \\ 0.2 & 0.19 & 0.18 & 0.17 & 0.16 & 0.17 \end{bmatrix}$，求此信源的熵，并解释为什么 $H(X) > \log 6$，不满足信源熵的极值性。

解：

$$H(X) = -\sum P(x) \log P(x) = 2.65 > \log 6$$

原因是给定的信源空间不满足概率空间的完备集这一特性，因此不满足极值条件。

【2.9】设离散无记忆信源 S 其符号集 $A=\{a_1,a_2,\ldots,a_q\}$，知其相应的概率分别为 $(P_1,P_2,\ldots,P_q)$。设另一离散无记忆信源 S'，其符号集为 S 信源符号集的两倍，$A'=\{a_i,i=1,2,\ldots,2q\}$，并且各符号的概率分布满足

$$\begin{aligned} &P_i'=(1-e)P_i \quad && i=1,2,\ldots,q \\ &P_i'=eP_i && i=q+1,q+2,\ldots,2q \end{aligned}$$

试写出信源 S' 的信息熵与信源 S 的信息熵的关系。

解:

$$\begin{aligned} H(S') &= -\sum P(x)\log P(x) \\ &= -\sum (1-e)P_i\log(1-e)P_i - \sum eP_i\log eP_i \\ &= -(1-e)\sum P_i\log(1-e) - (1-e)\sum P_i\log P_i - e\sum P_i\log e - e\sum P_i\log P_i \\ &= -(1-e)\log(1-e) - e\log e + H(S) \\ &= H(S) + H(e,1-e) \end{aligned}$$

【2.10】设有一概率空间，其概率分布为 $\{p_1,p_2,\ldots,p_q\}$，并有 $p_1>p_2$。若取 $p_1'=p_1-e$，$p_2'=p_2+e$，其中 $0<2e\le p_1-p_2$，而其他概率值不变。试证明由此所得新的概率空间的熵是增加的，并用熵的物理意义加以解释。

解:

设新的信源为 X'，新信源的熵为:

$$H(X')=-\sum p_i\log p_i=-(p_1-e)\log(p_1-e)-(p_2+e)\log(p_2+e)-\cdots-p_q\log p_q$$

原信源的熵

$$H(X)=-\sum p_i\log p_i=-p_1\log p_1-p_2\log p_2-\cdots-p_q\log p_q$$

因此有，

$$H(X)-H(X')=(p_1-e)\log(p_1-e)+(p_2+e)\log(p_2+e)-p_1\log p_1-p_2\log p_2$$

令 $f(x)=(p_1-x)\log(p_1-x)+(p_2+x)\log(p_2+x)$，$x\in\left(0,\dfrac{p_1-p_2}{2}\right]$，则

$$f'(x)=\log\frac{p_2+x}{p_1-x}\le 0$$

即函数 $f(x)$ 为减函数，因此有 $f(0) \geq f(e)$，即

$$(p_1 - e)\log(p_1 - e) + (p_2 + e)\log(p_2 + e) \leq p_1 \log p_1 + p_2 \log p_2$$

因此 $H(X) \leq H(X')$ 成立。

【解释】

当信源符号的概率趋向等概率分布时，不确定性增加，即信息熵是增加的。

【2.11】试证明：若 $\sum_{i=1}^{L} p_i = 1$，$\sum_{j=1}^{m} q_j = p_L$，则

$$H(p_1, p_2, \mathbf{K}, p_{L-1}, q_1, q_2, \mathbf{K}, q_m) = H(p_1, p_2, \mathbf{K}, p_{L-1}, p_L) + p_L H(\frac{q_1}{p_L}, \frac{q_2}{p_L}, \mathbf{K}, \frac{q_m}{p_L})$$

并说明等式的物理意义。

解：

$$\begin{aligned}
&H(p_1, p_2, \mathbf{K}, p_{L-1}, q_1, q_2, \mathbf{K}, q_m) \\
&= -p_1 \log p_1 - p_2 \log p_2 - \mathbf{K} - p_{L-1} \log p_{L-1} - q_1 \log q_1 - q_2 \log q_2 - \mathbf{K} - q_m \log q_m \\
&= -p_1 \log p_1 - p_2 \log p_2 - \mathbf{K} - p_{L-1} \log p_{L-1} - p_L \log p_L + p_L \log p_L \\
&\quad - q_1 \log q_1 - q_2 \log q_2 - \mathbf{K} - q_m \log q_m \\
&= -p_1 \log p_1 - p_2 \log p_2 - \mathbf{K} - p_{L-1} \log p_{L-1} - p_L \log p_L + (q_1 + q_2 + q_3 + \mathbf{L} + q_m) \log p_L \\
&\quad - q_1 \log q_1 - q_2 \log q_2 - \mathbf{K} - q_m \log q_m \\
&= -p_1 \log p_1 - p_2 \log p_2 - \mathbf{K} - p_{L-1} \log p_{L-1} - p_L \log p_L \\
&\quad - q_1 \log \frac{q_1}{p_L} - q_2 \log \frac{q_2}{p_L} - \mathbf{K} - q_m \log \frac{q_m}{p_L} \\
&= -p_1 \log p_1 - p_2 \log p_2 - \mathbf{K} - p_{L-1} \log p_{L-1} - p_L \log p_L \\
&\quad + p_L (-\frac{q_1}{p_L} \log \frac{q_1}{p_L} - \frac{q_2}{p_L} \log \frac{q_2}{p_L} - \mathbf{K} - \frac{q_m}{p_L} \log \frac{q_m}{p_L}) \\
&= H(p_1, p_2, \mathbf{K}, p_{L-1}, p_L) + p_L H_m(\frac{q_1}{p_L}, \frac{q_2}{p_L}, \mathbf{K}, \frac{q_m}{p_L})
\end{aligned}$$

【意义】

将原信源中某一信源符号进行分割，而分割后的符号概率之和等于被分割的原符号的概率，则新信源的信息熵增加，熵所增加的一项就是由于分割而产生的不确定性量。

【2.12】(1) 为了使电视图像获得良好的清晰度和规定的适当的对比度，需要用 5×10^5 个

像素和 10 个不同亮度电平，求传递此图像所需的信息率（比特/秒）。并设每秒要传送 30 帧图像，所有像素是独立变化的，且所有亮度电平等概率出现。

（2）设某彩电系统，除了满足对于黑白电视系统的上述要求外，还必须有 30 个不同的色彩度，试证明传输这彩色系统的信息率要比黑白系统的信息率约大 2.5 倍。

解:

每个像素的电平取自 10 个不同的电平，每一个像素形成的概率空间为:

$$\begin{bmatrix} X \\ P \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & \cdots & a_{10} \\ \frac{1}{10} & \frac{1}{10} & \cdots & \frac{1}{10} \end{bmatrix}$$

这样，平均每个像素携带的信息量为:

$$H(X) = \log 10 = 3.32 \text{ 比特/像素}$$

现在所有的像素点之间独立变化的，因此，每帧图像含有的信息量为:

$$H(X^N) = NH(X) = 5\times 10^5 \times \log 10 = 1.66\times 10^6 \text{ 比特/帧}$$

按每秒传输 30 帧计算，每秒需要传输的比特数，即信息传输率为:

$$30\times H(X^N) = 4.98\times 10^7 \text{ 比特/秒}$$

除满足黑白电视系统的要求外，还需 30 个不同的色彩度，不妨设每个色彩度等概率出现，则其概率空间为:

$$\begin{bmatrix} Y \\ P \end{bmatrix} = \begin{bmatrix} b_1 & b_2 & \cdots & b_{30} \\ \frac{1}{30} & \frac{1}{30} & \cdots & \frac{1}{30} \end{bmatrix}$$

其熵为 $\log 30$ 比特/符号，由于电平与色彩是互相独立的，因此有

$$H(XY) = H(X) + H(Y) = \log 300$$

这样，彩色电视系统的信息率与黑白电视系统信息率的比值为

$$\frac{H(XY)}{H(X)} = \frac{\log 300}{\log 10} \approx 2.5$$

【2.13】每帧电视图像可以认为是由 3×10^5 个像素组成，所以像素均是独立变化，且每一像素又取 128 个不同的亮度电平，并设亮度电平等概率出现。问每帧图像含有多少信息量？若现有一广播员在约 10000 个汉字的字汇中选 1000 个来口述此电视图像，试问广播员描述此图像所广播的信息量是多少（假设汉字是等概率分布，并且彼此无依赖）？若要恰当地描述此图像，广播员在口述中至少需用多少汉字？

解：

每个像素的电平亮度形成了一个概率空间，如下：

$$\begin{bmatrix} X \\ P \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & \cdots & a_{128} \\ \frac{1}{128} & \frac{1}{128} & \cdots & \frac{1}{128} \end{bmatrix}$$

平均每个像素携带的信息量为：

$$H(X) = \log 128 = 7 \text{ 比特/像素}$$

每帧图像由 3×10^5 个像素组成，且像素间是独立的，因此每帧图像含有的信息量为：

$$H(X^N) = NH(X) = 2.1\times10^6 \text{ 比特/帧}$$

如果用汉字来描述此图像，平均每个汉字携带的信息量为 $H(Y) = \log 10000 = 13.29$ 比特/汉字，选择 1000 字来描述，携带的信息量为

$$H(Y^N) = NH(Y) = 1.329\times10^4 \text{ 比特}$$

如果要恰当的描述此图像，即信息不丢失，在上述假设不变的前提下，需要的汉字个数为：

$$\frac{H(X^N)}{H(Y)} = \frac{2.1106}{13.29} \approx 1.58\times10^5 \text{ 字}$$

【2.14】为了传输一个由字母 A、B、C 和 D 组成的符号集，把每个字母编码成两个二元码脉冲序列，以 00 代表 A，01 代表 B，10 代表 C，11 代表 D。每个二元码脉冲宽度为 5ms。

（1） 不同字母等概率出现时，计算传输的平均信息速率？

（2） 若每个字母出现的概率分别为 $p_A = \frac{1}{5}$，$p_B = \frac{1}{4}$，$p_C = \frac{1}{4}$，$p_D = \frac{3}{10}$，试计算传输的平均速率?

解:

假设不同字母等概率出现时，平均每个符号携带的信息量为

$$H(X) = \log 4 = 2 \text{ 比特}$$

每个二元码宽度为 5ms，每个字母需要 2 个二元码，则其传输时间为 10ms，每秒传送 $n = 100$ 个，因此信息传输速率为:

$$R = nH(X) = 100 \times 2 = 200 \text{ 比特/秒}$$

当不同字母概率不同时，平均传输每个字母携带的信息量为

$$H(X) = \frac{1}{5}\log 5 + \frac{1}{4}\log 4 + \frac{1}{4}\log 4 + \frac{3}{10}\log\frac{10}{3} = 1.985 \text{ 比特/符号}$$

此时传输的平均信息速度为

$$R = nH(X) = 1.985 \times 10^2 \text{ 比特/秒}$$

【2.15】证明离散平稳信源有 $H(X_3 | X_1X_2) \le H(X_2 | X_1)$，试说明等式成立的条件。

解:

$$\begin{aligned} H(X_3 | X_1X_2) &= -\sum\sum\sum P(x_1x_2x_3)\log P(x_3 | x_1x_2) \\ &= -\sum_{X_1}\sum_{X_2} P(x_1x_2)\sum_{X_3} P(x_3 | x_1x_2)\log P(x_3 | x_1x_2) \\ &\le -\sum_{X_1}\sum_{X_2} P(x_1x_2)\sum_{X_3} P(x_3 | x_1x_2)\log P(x_3 | x_2) \\ &= H(X_3 | X_2) \end{aligned}$$

根据信源的平稳性，有 $H(X_3 | X_2) = H(X_2 | X_1)$，因此有 $H(X_3 | X_1X_2) \le H(X_2 | X_1)$。

等式成立的条件是 $P(x_3 | x_1x_2) = P(x_3 | x_2)$。

【2.16】证明离散信源有 $H(X_1X_2 \cdots X_N) \le H(X_1) + H(X_2) + \cdots + H(X_N)$，并说明等式成立的条件。

证明:

$$H(X_1X_2\cdots X_N)=H(X_1)+H(X_2|X_1)+\cdots+H(X_N|X_1X_2\cdots X_{N-1})$$

而

$$\begin{aligned}
&H(X_N|X_1X_2\cdots X_{N-1})\\
&=-\sum_{X_1}\sum_{X_2}\cdots\sum_{X_N}P(x_1x_2\cdots x_N)\log P(x_N|x_1x_2\cdots x_{N-1})\\
&=-\sum_{X_1}\sum_{X_2}\cdots\sum_{X_{N-1}}P(x_1x_2\cdots x_{N-1})\sum_{X_N}P(x_N|x_1x_2\cdots x_{N-1})\log P(x_N|x_1x_2\cdots x_{N-1})\\
&\le-\sum_{X_1}\sum_{X_2}\cdots\sum_{X_{N-1}}P(x_1x_2\cdots x_{N-1})\sum_{X_N}P(x_N|x_1x_2\cdots x_{N-1})\log P(x_N)\\
&=H(X_N)
\end{aligned}$$

即

$$H(X_2|X_1)\le H(X_2)$$

$$H(X_3|X_1X_2)\le H(X_3)$$

……

代入上述不等式，有

$$H(X_1X_2\cdots X_N)\le H(X_1)+H(X_2)+\cdots+H(X_N)$$

等号成立的条件是:

$$P(x_N|x_1x_2\cdots x_{N-1})=P(x_N)$$

$$P(x_{N-1}|x_1x_2\cdots x_{N-2})=P(x_{N-1})$$

……

$$P(x_2|x_1)=P(x_2)$$

即离散平稳信源输出的 N 长的随机序列之间彼此统计无依赖时，等式成立。

【2.17】设有一个信源，它产生 0、1 序列的消息。它在任意时间而且不论以前发生过什么符号，均按 $P(0)=0.4$ ， $P(1)=0.6$ 的概率发出符号。

(1) 试问这个信源是否是平稳的?

(2) 试计算 $H(X^2)$、 $H(X_3|X_1X_2)$ 及 $\lim_{N\to\infty}H_N(X)$ 。

(3) 试计算 $H(X^4)$ 并写出 X^4 信源中可能有的所有符号。

解:

该信源任一时刻发出 0 和 1 的概率与时间无关，因此是平稳的，即该信源是离散平稳信源。其信息熵为

$$H(X) = -\sum P(x)\log P(x) = 0.971 \text{ 比特/符号}$$

信源是平稳无记忆信源，输出的序列之间无依赖，所以

$$H(X^2) = 2H(X) = 1.942 \text{ 比特/符号}$$

$$H(X_3 \mid X_1X_2) = H(X) = 0.971 \text{ 比特/符号}$$

$$\lim_{N\to\infty} H_N(X) = \lim_{N\to\infty} \frac{1}{N} H(X_1X_2 \cdots X_N) = H(X) = 0.971 \text{ 比特/符号}$$

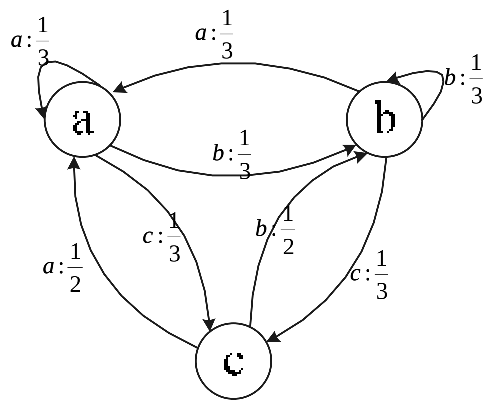
$$H(X^4) = 4H(X) = 3.884 \text{ 比特/符号}$$

X^4 信源中可能的符号是所有 4 位二进制数的排序，即从 0000~1111 共 16 种符号。

【2.18】设有一信源，它在开始时以 $P(a)=0.6$，$P(b)=0.3$，$P(c)=0.1$ 的概率发出 X_1。如果 X_1 为 a 时，则 X_2 为 a、b、c 的概率为 $\frac{1}{3}$；如果为 b 时，则 X_2 为 a、b、c 的概率为 $\frac{1}{3}$；如果 X_1 为 c 时，则 X_2 为 a、b 的概率为 $\frac{1}{2}$，为 c 的概率为 0。而且后面发出 X_i 的概率只与 X_{i-1} 有关，又当 $i \ge 3$ 时，$P(X_i \mid X_{i-1}) = P(X_2 \mid X_1)$。试用马尔克夫信源的图示法画出状态转移图，并计算此信源的熵 H_∞。

解:

信源为一阶马尔克夫信源，其状态转换图如下所示。



根据上述状态转换图，设状态极限概率分别为 $P(a)$、$P(b)$ 和 $P(c)$，根据切普曼—柯尔莫哥洛夫方程有

$$\begin{cases} Q(a)=\frac{1}{3}Q(a)+\frac{1}{3}Q(b)+\frac{1}{2}Q(c) \\ Q(b)=\frac{1}{3}Q(a)+\frac{1}{3}Q(b)+\frac{1}{2}Q(c) \\ Q(c)=\frac{1}{3}Q(a)+\frac{1}{3}Q(b) \\ Q(a)+Q(b)+Q(c)=1 \end{cases}$$

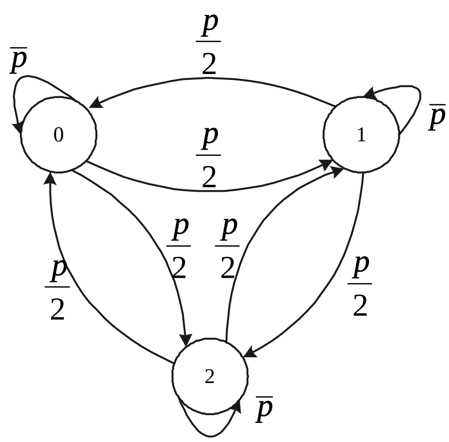
解得:

$$Q(a)=Q(b)=\frac{3}{8}\text{，}\ Q(c)=\frac{1}{4}$$

得此一阶马尔克夫的信息熵为:

$$H_\infty=\sum Q(E_i)H(X\mid E_i)=1.439\ \text{比特/符号}$$

【2.19】一阶马尔克夫信源的状态图如右图所示，信源 X 的符号集为 $\{0,1,2\}$ 并定义 $\overline{p}=1-p$。



（1） 求信源平稳后的概率分布 $P(0)$、$P(1)$ 和 $P(2)$；

（2） 求此信源的熵 H_∞；

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

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

解:

根据切普曼—柯尔莫哥洛夫方程，可得

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

解得：$P(0)=P(1)=P(2)=\frac{1}{3}$

该一阶马尔克夫信源的信息熵为：

$$H_\infty=\sum Q(E_i)H(X\mid E_i)=-\bar{p}\log\bar{p}-p\log p+p \text{ 比特/符号}$$

当信源为无记忆信源，符号的概率分布等于平稳分布，此时信源的概率空间为：

$$\begin{bmatrix}X\\P\end{bmatrix}=\begin{bmatrix}0 & 1 & 2\\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3}\end{bmatrix}$$

此时信源的信息熵为 $H(X)=\log 3=1.585$ 比特/符号

由上述计算结果可知：$H(X)\geq H(\infty)$。

求一阶马尔克夫信源熵 H_∞ 的最大值，$H_\infty=-\bar{p}\log\bar{p}-p\log p+p$，有

$$\frac{dH_\infty}{dp}=\log\frac{2(1-p)}{p}$$

可得，当 $p=\frac{2}{3}$ 时，H_∞ 达到最大值，此时最大值为 $\log 3=1.585$ 比特/符号。

当 $p=0$ 时，$H_\infty=0$ 比特/符号；$p=1$ 时，$H_\infty=1$ 比特/符号

【2.20】黑白气象传真图的消息只有黑色和白色两种，即信源 $X=\{\text{黑},\text{白}\}$，设黑色出现的概率为 $P(\text{黑})=0.3$，白色出现的概率为 $P(\text{白})=0.7$。

（1） 假设图上黑白消息出现前后没有关联，求熵 $H(X)$；

（2） 假设消息前后有关联，其依赖关系为 $P(\text{白}\mid\text{白})=0.9$，$P(\text{黑}\mid\text{白})=0.1$，$P(\text{白}\mid\text{黑})=0.2$，$P(\text{黑}\mid\text{黑})=0.8$，求此一阶马尔克夫信源的熵 H_2。

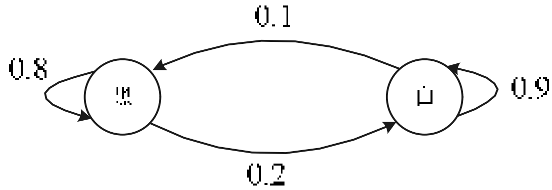
（3） 分别求上述两种信源的冗余度，并比较 $H(X)$ 和 H_2 的大小，并说明其物理意义。

解：

如果出现黑白消息前后没有关联，信息熵为：

$$H(X)=-\sum p_i\log p_i=0.881 \text{ 比特/符号}$$

当消息前后有关联时，首先画出其状态转移图，如下所示。



设黑白两个状态的极限概率为$Q(黑)$和$Q(白)$，根据切普曼—柯尔莫哥洛夫方程可得：

$$\begin{cases} Q(黑)=0.8Q(黑)+0.1Q(白) \\ Q(白)=0.2Q(黑)+0.9Q(白) \\ Q(黑)+Q(白)=1 \end{cases}$$

解得：

$$Q(黑)=\frac{1}{3}，\ Q(白)=\frac{2}{3}$$

此信源的信息熵为：

$$\mathrm{H}_\infty=\sum Q(E_i)H(X\mid E_i)=0.553\text{比特/符号}$$

两信源的冗余度分别为:

$$g_1=1-\frac{H(X)}{\log 2}=0.119$$

$$g_1=1-\frac{H_\infty}{\log 2}=0.447$$

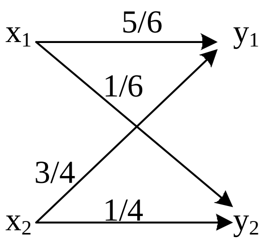
结果表明：当信源的消息之间有依赖时，信源输出消息的不确定性减弱。就本题而言，当有依赖时前面已是白色消息，后面绝大多数可能是出现白色消息；前面是黑色消息，后面基本可猜测是黑色消息。这时信源的平均不确定性减弱，所以信源消息之间有依赖时信源熵小于信源消息之间无依赖时的信源熵，这表明信源熵正是反映信源的平均不确定的大小。而信源剩余度正是反映信源消息依赖关系的强弱，剩余度越大，信源消息之间的依赖关系就越大。

第三章课后习题

【3.1】 设信源

$$\begin{bmatrix} X \\ P(x) \end{bmatrix} = \begin{bmatrix} x_1 & x_2 \\ 0.6 & 0.4 \end{bmatrix}$$

通过一干扰信道，接收符号为$Y=[y_1,y_2]$，信道传递概率如下图所示，求



（1）信源 X 中事件 x_1 和 x_2 分别含有的自信息；

（2）收到消息 $y_j(j=1,2)$ 后，获得的关于 $x_i(i=1,2)$ 的信息量；

（3）信源 X 和信源 Y 的信息熵；

（4）信道疑义度 $H(X|Y)$ 和噪声熵 $H(Y|X)$；

（5）接收到消息 Y 后获得的平均互信息。

解：

（1）信源 X 中事件 x_1 和 x_2 分别含有的自信息分别为：

$$I(x_1)=\log\frac{1}{P(x_1)}=-\log 0.6=0.737 \text{ 比特}$$

$$I(x_2)=\log\frac{1}{P(x_2)}=-\log 0.4=1.32 \text{ 比特}$$

（2）根据给定的信道以及输入概率分布，可得

$$P(y_1)=\sum_X P(x_i)P(y_1|x_i)=0.8$$
$$P(y_2)=\sum_X P(x_i)P(y_2|x_i)=0.2$$

所求的互信息量分别为：

$$I(x_1;y_1)=\log\frac{P(y_1|x_1)}{P(y_1)}=\log\frac{5/6}{0.8}=\log\frac{25}{24}=0.059 \text{ 比特}$$

$$I(x_2;y_1)=\log\frac{P(y_1|x_2)}{P(y_1)}=\log\frac{3/4}{0.8}=\log\frac{15}{16}=-0.093\text{比特}$$

$$I(x_1;y_2)=\log\frac{P(y_2|x_1)}{P(y_2)}=\log\frac{1/6}{0.2}=\log\frac{5}{6}=-0.263\text{比特}$$

$$I(x_2;y_2)=\log\frac{P(y_2|x_2)}{P(y_2)}=\log\frac{1/4}{0.2}=\log\frac{5}{4}=0.322\text{ 比特}$$

（3）信源 X 以及 Y 的熵为：

$$H(X)=-\sum_X P(x)\log P(x)=-0.6\log 0.6-0.4\log 0.4=0.971\text{比特/符号}$$

$$H(Y)=-\sum_Y P(y)\log P(y)=-0.8\log 0.8-0.2\log 0.2=0.722\text{ 比特/符号}$$

（4）信道疑义度 $H(X|Y)=-\sum_X P(x)\sum_Y P(y|x)\log P(x|y)$

而相关条件概率 $P(x|y)$ 计算如下：

$$P(x_1|y_1)=\frac{P(x_1,y_1)}{P(y_1)}=\frac{P(y_1|x_1)P(x_1)}{P(y_1)}=\frac{0.5}{0.8}=\frac{5}{8}$$

$$P(x_2|y_1)=\frac{3}{8}$$

$$P(x_1|y_2)=\frac{P(x_1,y_2)}{P(y_2)}=\frac{P(y_2|x_1)P(x_1)}{P(y_2)}=\frac{0.6/6}{0.2}=\frac{1}{2}$$

$$P(x_2|y_2)=\frac{1}{2}$$

由此计算出信道疑义度为：

$$H(X|Y)=-0.6\left[\frac{5}{6}\log\frac{5}{8}+\frac{1}{6}\log\frac{1}{2}\right]-0.4\left[\frac{3}{4}\log\frac{3}{8}+\frac{1}{4}\log\frac{1}{2}\right]=0.9635\text{ 比特/符号}$$

噪声熵为：

$$\begin{aligned}&H(Y|X)=-\sum P(x)P(y|x)\log P(y|x)\\&=-0.6\left[\frac{5}{6}\log\frac{5}{6}+\frac{1}{6}\log\frac{1}{6}\right]-0.4\left[\frac{3}{4}\log\frac{3}{4}+\frac{1}{4}\log\frac{1}{4}\right]\\&=0.7145\text{比特/符号}\end{aligned}$$

（5）接收到信息 Y 后获得的平均互信息为：

$$I(X;Y)=H(X)-H(X|Y)=0.0075\text{ 比特/符号}$$

【3.2】 设 8 个等概率分布的消息通过传递概率为 p 的 BSC 进行传送，8 个消息相应编成下述码字:

M_1=0000，M_2=0101，M_3=0110，M_4=0011

M_5=1001，M_6=1010，M_7=1100，M_8=1111

试问:

（1）接收到第一个数字 0 与 M_1 之间的互信息;

（2）接收到第二个数字也是 0 时，得到多少关于 M_1 的附加互信息;

（3）接收到第三个数字仍为 0 时，又增加了多少关于 M_1 的互信息;

（4）接收到第四个数字还是 0 时，再增加了多少关于 M_1 的互信息。

解:

各个符号的先验概率均为 $\frac{1}{8}$

（1）根据已知条件，有

$$P(y_1=0|M_1)=P(y_1=0|0000)=P(y_1=0|x_1=0)=\bar{p}$$

$$P(y_1=0)=\sum_{M_i}P(M_i)P(0|M_i)=\frac{1}{2}$$

因此接收到第一个数字 0 与 M_1 之间的互信息为:

$$I(M_1;y_1=0)=\log\frac{P(y_1=0|M_1)}{P(y_1=0)}=\log\frac{\bar{p}}{1/2}=1+\log\bar{p}\text{ 比特}$$

（2）根据已知条件，有

$$P(y_1y_2=00|M_1)=P(y_1y_2=00|0000)=\bar{p}^2$$

$$P(y_1y_2=00)=\sum_{M_i}P(M_i)P(00|M_i)=\frac{1}{8}\left[2\bar{p}^2+4p\bar{p}+2p^2\right]=\frac{1}{4}$$

因此接收到第二个数字也是 0 时，得到多少关于 M_1 的互信息为:

$$I(M_1;y_1y_2=00)=\log\frac{P(y_1y_2=00\,|\,M_1)}{P(y_1y_2=00)}=\log\frac{\overline{p}^2}{1/4}=2+2\log\overline{p}\ \text{比特/符号}$$

得到的附加信息为:

$$I(M_1;y_1y_2=00)-I(M_1;y_1=0)=1+\log\overline{p}\ \text{比特/符号}$$

（3）根据已知条件，有

$$P(y_1y_2y_3=000\,|\,M_1)=P(y_1y_2y_3=000\,|\,000)=\overline{p}^3$$

$$P(y_1y_2y_3=000)=\sum_{M_i}P(M_i)P(000\,|\,M_i)=\frac{1}{8}\left[\overline{p}^3+3p\overline{p}^2+3p^2\overline{p}+p^3\right]=\frac{1}{8}$$

因此接收到第三个数字也是 0 时，得到多少关于 M_1 的互信息为:

$$I(M_1;y_1y_2y_3=000)=\log\frac{P(y_1y_2y_3=000\,|\,M_1)}{P(y_1y_2y_3=000)}=\log\frac{\overline{p}^3}{1/8}=3+3\log\overline{p}$$

此时得到的附加信息为:

$$I(M_1;y_1y_2y_3=000)-I(M_1;y_1y_2=00)=1+\log\overline{p}\ \text{比特/符号}$$

（4）根据已知条件，有

$$P(y_1y_2y_3y_4=0000\,|\,M_1)=P(y_1y_2y_3y_4=0000\,|\,0000)=\overline{p}^4$$

$$P(y_1y_2y_3y_4=0000)=\sum_{M_i}P(M_i)P(0000\,|\,M_i)=\frac{1}{8}\left[\overline{p}^4+6p^2\overline{p}^2+p^4\right]$$

因此接收到第四个符号为 0 时，得到的关于 M_1 的互信息为

$$\begin{aligned}I(M_1;y_1y_2y_3=0000)\quad&=\log\frac{P(y_1y_2y_3y_4=0000\,|\,M_1)}{P(y_1y_2y_3y_4=0000)}\\&=\log\frac{\overline{p}^4}{\frac{1}{8}\left(\overline{p}^4+6\overline{p}^2p^2+p^4\right)}\\&=3+4\log\overline{p}-\log\left(\overline{p}^4+6\overline{p}^2p^2+p^4\right)\end{aligned}$$

此时得到的附加信息为

$$I(M_1;y_1y_2y_3y_4=000)-I(M_1;y_1y_2y_3=000)=\log\overline{p}-\log\left(\overline{p}^4+6\overline{p}^2p^2+p^4\right)$$

【】设二元对称信道的传递矩阵为

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

（1）若 P(0)=3/4，P(1)=1/4，求 $H(X)$， $H(X\,|\,Y)$， $H(Y\,|\,X)$ 和 $I(X;Y)$；

（2）求该信道的信道容量及其达到信道容量时的输入概率分布。

解:

（1）根据已知条件，有

$$\begin{aligned} H(X) &= -\sum_X P(x_i)\log P(x_i) \\ &= -\frac{3}{4}\log\frac{3}{4} - \frac{1}{4}\log\frac{1}{4} \\ &= 0.811 \text{比特/符号} \end{aligned}$$

$$P(y=0) = \sum_X P(x)P(y=0\,|\,x) = \frac{3}{4}\times\frac{2}{3} + \frac{1}{4}\times\frac{1}{3} = \frac{7}{12}$$

$$P(y=1) = \sum_X P(x)P(y=1\,|\,x) = \frac{5}{12}$$

$$P(x=0\,|\,y=0) = \frac{P(x=0)P(y=0\,|\,x=0)}{P(y=0)} = \frac{\frac{3}{4}\times\frac{2}{3}}{7/12} = \frac{6}{7}$$

$$P(x=1\,|\,y=0) = \frac{1}{7}$$

$$P(x=0\,|\,y=1) = \frac{P(x=0)P(y=1\,|\,x=0)}{P(y=1)} = \frac{\frac{3}{4}\times\frac{1}{3}}{5/12} = \frac{3}{5}$$

$$P(x=1\,|\,y=1) = \frac{2}{5}$$

$$\begin{aligned} H(Y\,|\,X) &= -\sum_X P(x)\sum_Y P(y\,|\,x)\log P(y\,|\,x) \\ &= -\frac{3}{4}\left(\frac{2}{3}\log\frac{2}{3} + \frac{1}{3}\log\frac{1}{3}\right) - \frac{1}{4}\left(\frac{1}{3}\log\frac{1}{3} + \frac{2}{3}\log\frac{2}{3}\right) \\ &= 0.918 \text{比特/符号} \end{aligned}$$

$$\begin{aligned} H(X|Y) &= -\sum_X P(x)\sum_Y P(y|x)\log P(x|y) \\ &= -\frac{3}{4}\left(\frac{2}{3}\log\frac{6}{7}+\frac{1}{3}\log\frac{3}{5}\right)-\frac{1}{4}\left(\frac{1}{3}\log\frac{1}{7}+\frac{2}{3}\log\frac{2}{5}\right) \\ &= 0.749\text{比特/符号} \end{aligned}$$

$I(X;Y)=H(X)-H(X|Y)=0.062$ 比特/符号

（2）此信道是对称信道，因此其信道容量为：

$$C=1-H(p)=1-H(\frac{2}{3},\frac{1}{3})=0.082\text{ 比特/符号}$$

根据对称信道的性质可知，当 $P(0)=P(1)=\frac{1}{2}$ 时，信道的传输率 $I(X;Y)$ 达到信道容量。

【3.4】设有一批电阻，按阻值分 70%是 2kΩ，30%是 5kΩ；按功耗分 64%是 1/8W，其余是 1/4W。现已知 2kΩ 阻值的电阻中 80%是 1/8W。问通过测量阻值可以平均得到的关于瓦数的信息量是多少？

解：

根据已知条件，设电阻的阻值为事件 X，电阻的功耗为事件 Y，则两事件的概率空间为：

$$\begin{bmatrix} X \\ P \end{bmatrix}=\begin{bmatrix} x_1=2k\Omega & x_2=5k\Omega \\ 0.7 & 0.3 \end{bmatrix},\quad \begin{bmatrix} Y \\ P \end{bmatrix}=\begin{bmatrix} y_1=1/8W & y_2=1/4W \\ 0.64 & 0.36 \end{bmatrix}$$

给定条件为 $P(y_1|x_1)=0.8$， $P(y_2|x_1)=0.2$，而

$$0.64=P(y_1)=P(x_1)P(y_1|x_1)+P(x_2)P(y_1|x_2)=0.7*0.8+0.3*P(y_1|x_2)$$

$$0.36=P(y_2)=P(x_1)P(y_2|x_1)+P(x_2)P(y_2|x_2)=0.7*0.2+0.3*P(y_2|x_2)$$

解得：

$$P(y_1|x_2)=\frac{4}{15},\quad P(y_2|x_2)=\frac{11}{15}$$

$$H(Y|X)=-0.7*(0.8\log 0.8+0.2\log 0.2)-0.3*\left(\frac{4}{15}\log\frac{4}{15}+\frac{11}{15}\log\frac{11}{15}\right)=0.7567$$

$$I(X;Y)=H(Y)-H(Y|X)=0.186\text{ 比特/符号}$$

【3.5】 若 X 、 Y 和 Z 是三个随机变量，试证明：

(1) $I(X;YZ)=I(X;Y)+I(X;Z|Y)=I(X;Z)+I(X;Y|Z)$

(2) $I(X;Y|Z)=I(Y;X|Z)=H(X|Z)-H(X|YZ)$

(3) $I(X;Y|Z)\geq 0$ 当且仅当 (X,Z,Y) 是马氏链时等式成立。

证明：

(1)

$$
\begin{aligned}
I(X;YZ) &= \sum_{X,Y,Z} P(x,y,z)\log\frac{P(x|yz)}{P(x)} \\
&= \sum_{X,Y,Z} P(x,y,z)\log\left(\frac{P(x|yz)}{P(x|y)}*\frac{P(x|y)}{P(x)}\right) \\
&= \sum_{X,Y,Z} P(x,y,z)\log\frac{P(x|yz)}{P(x|y)}+\sum_{X,Y,Z} P(x,y,z)\log\frac{P(x|y)}{P(x)} \\
&= I(X;Z|Y)+I(X;Y)
\end{aligned}
$$

同理， $I(X;YZ)=I(X;Z)+I(X;Y|Z)$

(2)

$$
\begin{aligned}
I(X;Y|Z) &= \sum_{X,Y,Z} P(x,y,z)\log\frac{P(x|yz)}{P(x|z)} \\
&= \sum_{X,Y,Z} P(x,y,z)\log\frac{P(xyz)P(z)}{P(xz)P(yz)} \\
&= \sum_{X,Y,Z} P(x,y,z)\log\frac{P(y|xz)}{P(y|z)} \\
&= I(Y;X|Z)
\end{aligned}
$$

$$
\begin{aligned}
I(X;Y|Z) &= \sum_{X,Y,Z} P(x,y,z)\log\frac{P(x|yz)}{P(x|z)} \\
&= -\sum_{X,Y,Z} P(x,y,z)\log P(x|z)+\sum_{X,Y,Z} P(x,y,z)\log P(x|yz) \\
&= H(X|Z)-H(X|YZ)
\end{aligned}
$$

(3)

$$
\begin{aligned}
-I(X;Y\mid Z) &= \sum_{X,Y,Z} P(x,y,z)\log\frac{P(x\mid z)}{P(x\mid yz)} \\
&\le \log\sum_{X,Y,Z} P(x,y,z)\frac{P(x\mid z)}{P(x\mid yz)} \\
&= \log\sum_{X,Y,Z}\frac{P(xz)P(yz)}{P(z)} \\
&= 0
\end{aligned}
$$

等号成立当且仅当 $\frac{P(x\mid z)}{P(x\mid yz)}=1=\frac{P(xz)P(yz)}{P(xyz)P(z)}=\frac{P(y\mid z)}{P(y\mid xz)}$，即 $P(y\mid z)=P(y\mid xz)$，即 (X,Z,Y) 是马氏链。

【3.6】若有三个离散随机变量，有如下关系：$X+Y=Z$，其中 X 和 Y 相互统计独立，试证明：

（1） $H(X)\le H(Z)$，当且仅当 Y 是常量时等式成立；

（2） $H(Y)\le H(Z)$，当且仅当 X 为常量时等式成立；

（3） $H(Z)\le H(XY)\le H(X)+H(Y)$，当且仅当 X，Y 中任意一个为常量时等式成立；

（4） $I(X;Z)=H(Z)-H(Y)$；

（5） $I(XY;Z)=H(Z)$；

（6） $I(X;YZ)=H(X)$；

（7） $I(Y;Z\mid X)=H(Y)$；

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

证明：

当 $X+Y=Z$ 时，有 $P(z\mid xy)=\begin{cases}0 & z\ne x+y\\ 1 & z=x+y\end{cases}$，即 $H(Z\mid XY)=0$，而 $H(Z\mid XY)=H(Z)-I(XY;Z)$，因此 $I(XY;Z)=H(Z)$。

$$\begin{aligned} H(Z|X) &= -\sum P(x,z)\log P(z|x) \\ &= -\sum P(x,z)\log\frac{P(x,z)}{P(x)} \\ &= -\sum P(x,y)\log\frac{P(x,y)}{P(x)} \\ &= H(Y) \end{aligned}$$

而$I(X;Z)=H(Z)-H(Z|X)$，因此$I(X;Z)=H(Z)-H(Y)$。

根据互信息的性质，有$I(X;Z)\ge 0$，因此$H(Z)\ge H(Y)$成立，而当X为常量时，Z和X的概率分布相同，因此上述不等式中的等号成立。

同理，$H(Z)\ge H(X)$成立。

由于$I(XY;Z)=H(Z)-H(Z|XY)=H(XY)-H(XY|Z)=H(Z)$，而$H(XY|Z)\ge 0$，因此$H(Z)\le H(XY)$成立。

根据条件，有$P(x|yz)=\begin{cases}0 & z\ne x+y\\ 1 & z=x+y\end{cases}$，因此$H(X|YZ)=0$，而$I(X;YZ)=H(X)-H(X|YZ)$，因此$I(X;YZ)=H(X)$。

$$I(Y;Z|X)=H(Y|X)-H(Y|XZ)=H(Y|X)=H(Y)$$

$$\begin{aligned} I(X;Y|Z) &= H(X|Z)-H(X|YZ) \\ &= H(X|Z) \\ &= I(Y;X|Z) \\ &= H(Y|Z)-H(Y|XZ) \\ &= H(Y|Z) \end{aligned}$$

【3.7】 设X，Y是两个相互统计独立的二元随机变量，其取“0”或“1”的概率为等概率分布。定义另一个二元随机变量Z，而且$Z=XY$（一般乘积），试计算:

（1） $H(X)$，$H(Y)$，$H(Z)$；

（2） $H(XY)$，$H(XZ)$，$H(YZ)$，$H(XYZ)$；

（3） $H(X|Y)$，$H(X|Z)$，$H(Y|Z)$，$H(Z|X)$，$H(Z|Y)$；

（4） $H(X|YZ)$，$H(Y|XZ)$，$H(Z|XY)$；

（5） $I(X;Y)$，$I(X;Z)$，$I(Y;Z)$；

（6） $I(X;Y|Z)$，$I(Y;X|Z)$，$I(Z;X|Y)$，$I(Z;Y|X)$；

（7） $I(XY;Z)$，$I(X;YZ)$，$I(Y;XZ)$；

解:

由于 X 和 Y 是相互独立的等概率分布的随机变量，因此有

$$H(X)=H(Y)=1\text{比特/符号}$$

而符号 Z 的概率空间为: $\begin{bmatrix} Z \\ P \end{bmatrix}=\begin{bmatrix} 0 & 1 \\ \frac{3}{4} & \frac{1}{4} \end{bmatrix}$，因此

$$H(Z)=H(\frac{3}{4},\frac{1}{4})=0.811\text{比特/符号}$$

$H(XY)=H(X)+H(Y)=2$ 比特/符号

根据已知条件可得

$P(x=0,z=0)=P(x=0)=\frac{1}{2}$，$P(x=0,z=1)=0$

$P(x=1,z=0)=P(x=1,y=0)=\frac{1}{4}$，$P(x=1,z=1)=P(x=1,y=1)=\frac{1}{4}$

$P(z=0|x=0)=\frac{P(z=0,x=0)}{P(x=0)}=1$，$P(z=1|x=0)=\frac{P(z=1,x=0)}{P(x=0)}=0$

$P(z=0|x=1)=\frac{P(z=0,x=1)}{P(x=1)}=\frac{1}{2}$，$P(z=1|x=1)=\frac{P(z=1,x=1)}{P(x=1)}=\frac{1}{2}$

$H(Z|X)=-\sum P(x,z)\log P(z|x)=-\frac{1}{2}\log 1-\frac{1}{4}\log\frac{1}{2}-\frac{1}{4}\log\frac{1}{2}=0.5$ 比特/符号

$H(XZ)=H(X)+H(Z|X)=1.5$ 比特/符号

同理，$H(Z|Y)=0.5$ 比特/符号，$H(YZ)=H(Y)+H(Z|Y)=1.5$ 比特/符号

由于 $P(z|xy)=\begin{cases} 1 & z=xy \\ 0 & z\neq xy \end{cases}$，因此 $H(Z|XY)=0$ 比特/符号

$$H(XYZ)=H(XY)+H(Z|XY)=2\text{ 比特/符号}$$

$$H(X|Y)=H(XY)-H(Y)=1\text{比特/符号}$$

$$H(X|Z)=H(XZ)-H(Z)=0.689 \text{ 比特/符号}$$

$$H(Y|Z)=H(YZ)-H(Z)=0.689 \text{ 比特/符号}$$

$$H(X|YZ)=H(XYZ)-H(YZ)=2-1.5=0.5 \text{ 比特/符号}$$

同理，$H(Y|XZ)=0.5$ 比特/符号

$$I(X;Y)=H(X)-H(X|Y)=0 \text{ 比特/符号}$$

$$I(X;Z)=H(X)-H(X|Z)=0.311 \text{ 比特/符号}$$

$$I(Y;Z)=H(Y)-H(Y|Z)=0.311 \text{ 比特/符号}$$

$$I(X;Y|Z)=H(X|Z)-H(X|YZ)=0.689-0.5=0.189 \text{ 比特/符号}$$

$$I(Y;X|Z)=I(X;Y|Z)=0.189 \text{ 比特/符号}$$

$$I(Z;X|Y)=H(Z|Y)-H(Z|XY)=0.5 \text{ 比特/符号}$$

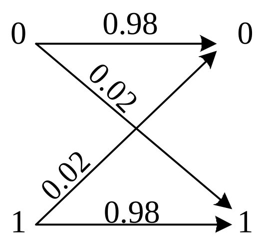
$$I(Z;Y|X)=H(Z|X)-H(Z|XY)=0.5 \text{ 比特/符号}$$

$$I(XY;Z)=H(Z)-H(Z|XY)=0.811 \text{ 比特/符号}$$

$$I(X;YZ)=H(X)-H(X|YZ)=1-0.5=0.5 \text{ 比特/符号}$$

$$I(Y;XZ)=H(Y)-H(Y|XZ)=0.5 \text{ 比特/符号}$$

【3.8】 有一个二元信道，其信道如右图所示。设该信道以 1500 个二元符号/秒的速度传输输入符号，现有一消息序列共有 14000 个二元符号，并设在这消息中 $P(0)=P(1)=\frac{1}{2}$。问从信息传输的角度来考虑，10 秒内能否将这消息序列无失真地传送完。



解:

该信道的信道矩阵为$\begin{bmatrix} 0.98 & 0.02 \\ 0.02 & 0.98 \end{bmatrix}$，信道容量为:

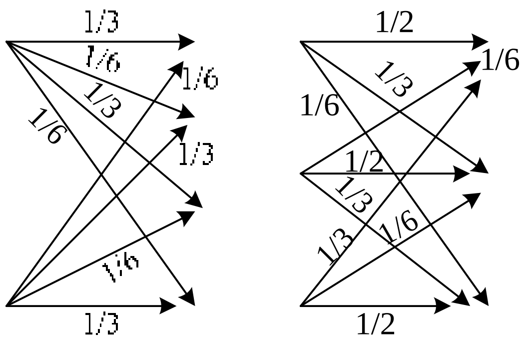
$$C = \log 2 - H(0.98,0.02) = 0.8586 \text{ 比特/符号}$$

10 秒内可以传输的最大信息量为:

$$1500 \times 0.8586 \times 10 = 1.288 \times 10^4 \text{ 比特}$$

而 14000 个符号中所含有的信息量为：14000 比特，因此从信息的角度来考虑，10 秒钟内不可能把上述 14000 个符号传输完。

【3.9】 求下图中信道的信道容量及其最佳的输入概率分布。



解：两个信道的信道矩阵分别如下:

$$\begin{bmatrix} \frac{1}{3} & \frac{1}{6} & \frac{1}{3} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{3} & \frac{1}{6} & \frac{1}{3} \end{bmatrix}, \begin{bmatrix} \frac{1}{2} & \frac{1}{3} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{6} & \frac{1}{2} \end{bmatrix}$$

可见，两个信道均是对称信道，信道容量分别为:

$$C_1 = \log 4 - H(\frac{1}{3},\frac{1}{6},\frac{1}{3},\frac{1}{6}) = 0.0817 \text{ 比特/符号}$$

$$C_2 = \log 3 - H(\frac{1}{2},\frac{1}{6},\frac{1}{3}) = 0.126 \text{ 比特/符号}$$

输入的最佳分布是等概率分布。

【3.10】 求下列两个信道的信道容量，并加以比较

$$(1)\begin{bmatrix} \overline{p}-e & p-e & 2e \\ p-e & \overline{p}-e & 2e \end{bmatrix} \qquad (2)\begin{bmatrix} \overline{p}-e & p-e & 2e & 0 \\ p-e & \overline{p}-e & 0 & 2e \end{bmatrix}$$

解：这两个信道均是准对称信道，当输入符号等概率时，平均互信息达到信道容量，具体如下：

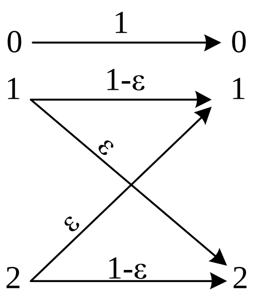
（1）该准对称信道的信道容量为：

$$\begin{aligned} C_1 &= \max\{H(Y)\} - H(\overline{p}-e, p-e, 2e) \\ &= -\frac{1-2e}{2}\log\frac{1-2e}{2} - \frac{1-2e}{2}\log\frac{1-2e}{2} - 2e\log 2e - H(\overline{p}-e, p-e, 2e) \\ &= (1-2e)\log\frac{2}{1-2e} + (p-e)\log(p-e) + (\overline{p}-e)\log(\overline{p}-e) \end{aligned}$$

（2）该准对称信道的信道容量为：

$$\begin{aligned} C_2 &= \max\{H(Y)\} - H(\overline{p}-e, p-e, 2e) \\ &= -\frac{1-2e}{2}\log\frac{1-2e}{2} - \frac{1-2e}{2}\log\frac{1-2e}{2} - e\log e - e\log e - H(\overline{p}-e, p-e, 2e) \\ &= (1-2e)\log\frac{2}{1-2e} + (p-e)\log(p-e) + (\overline{p}-e)\log(\overline{p}-e) + 2e \\ &= C_1 + 2e \end{aligned}$$

【3.11】 求下图中信道的信道容量及其最佳的输入概率分布，并求出 $e=0$ 和 $e=\frac{1}{2}$ 时的信道容量 C。



解：

该信道的信道矩阵如下：

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1-e & e \\ 0 & e & 1-e \end{bmatrix}$$

该信道既非对称信道，也非准对称信道，因此根据一般信道容量的计算公式，有

$$\sum P(b_j \mid a_i) b_j = \sum P(b_j \mid a_i) \log P(b_j \mid a_i)$$

即

$$\begin{cases} b_1 = 0 \\ (1-e)b_2 + eb_3 = (1-e)\log(1-e) + e\log e \\ eb_2 + (1-e)b_3 == (1-e)\log(1-e) + e\log e \end{cases}$$

解得:

$$b_1 = 0\text{，}\quad b_2 = b_3 = (1-e)\log(1-e) + e\log e$$

而信道容量

$$\begin{aligned} C \quad &= \log\sum 2^{b_j} \\ &= \log\left(1 + 2(1-e)^{1-e} e^{e}\right) \end{aligned}$$

信道的输出符号概率为：

$$P(b_1) = 2^{b_1 - C} = \frac{1}{1 + 2(1-e)^{1-e} e^{e}}$$

$$P(b_2) = 2^{b_2 - C} = \frac{(1-e)^{1-e} e^{e}}{1 + 2(1-e)^{1-e} e^{e}}$$

$$P(b_3) = 2^{b_3 - C} = \frac{(1-e)^{1-e} e^{e}}{1 + 2(1-e)^{1-e} e^{e}}$$

而

$$P(b_1) = P(a_1)$$

$$P(b_2) = (1-e)P(a_2) + eP(a_3)$$

$$P(b_3) = eP(a_2) + (1-e)P(a_3)$$

可得：

$$P(a_1) = \frac{1}{1 + 2(1-e)^{1-e} e^{e}}$$

$$P(a_2) = \frac{(1-e)^{1-e} e^{e}}{1 + 2(1-e)^{1-e} e^{e}}$$

$$P(a_3)=\frac{(1-e)^{1-e}e^{e}}{1+2(1-e)^{1-e}e^{e}}$$

当$e=0$时，$C=\log\left(1+2(1-e)^{1-e}e^{e}\right)=\log 3$，信道为一一对应信道；

当$e=\dfrac{1}{2}$时，$C=\log\left(1+2\left(\dfrac{1}{2}\right)\right)=\log 2$。

【3.12】 试证明$H(X)$是输入概率分布$P(x)$的上凸函数。

证明:

$$H(X)=-\sum_{X}P(x)\log P(x)$$

设存在两个概率分布$P_1(x)$和$P_2(x)$，目标是要证明

$$qH(P_1(x))+\bar{q}H(P_2(x))\le H(qP_1(x)+\bar{q}P_2(x))$$

证明过程如下:

$$\begin{aligned}
&qH(P_1(x))+\bar{q}H(P_2(x))-H(qP_1(x)+\bar{q}P_2(x))\\
&=-q\sum P_1(x)\log P_1(x)-\bar{q}\sum P_2(x)\log P_2(x)+\sum\left(qP_1(x)+\bar{q}P_2(x)\right)\log P(x)\\
&=q\sum P_1(x)\log\frac{P(x)}{P_1(x)}+\bar{q}\sum P_2(x)\log\frac{P(x)}{P_2(x)}\\
&\le q\log e\sum P_1(x)\left(\frac{P(x)}{P_1(x)}-1\right)+\bar{q}\log e\sum P_2(x)\left(\frac{P(x)}{P_2(x)}-1\right)\\
&=0
\end{aligned}$$

【3.13】 从平均互信息的表达式证明，当信道和信源都是无记忆时，有

$$I(X^N;Y^N)=NI(X;Y)$$

证明：设$a_k=\left(a_{k_1}a_{k_2}\cdots a_{k_N}\right),b_h=\left(b_{h_1}b_{h_2}\cdots b_{h_N}\right)$，按照给定信道和信源均是无记忆，

有

$$P(a_k)=P(a_{k_1}a_{k_2}\cdots a_{k_N})=P(a_{k_1})P(a_{k_2})\cdots P(a_{k_N})$$

$$P(b_h\mid a_k)=P(b_{h_1}b_{h_2}\cdots b_{h_N}\mid a_{k_1}a_{k_2}\cdots a_{k_N})=P(b_{h_1}\mid a_{k_1})P(b_{h_2}\mid a_{k_2})\cdots P(b_{h_N}\mid a_{k_N})$$

$$\begin{aligned}
P(b_h)\quad&=P(a_k)P(b_h\mid a_k)\\
&=P(a_{k_1})P(a_{k_2})\cdots P(a_{k_N})P(b_{h_1}\mid a_{k_1})P(b_{h_2}\mid a_{k_2})\cdots P(_{h_N}\mid a_{k_N})\\
&=P(b_{h_1})P(b_{h_2})\cdots P(_{h_N})
\end{aligned}$$

$$
\begin{aligned}
&I(X^N;Y^N)=H(Y^N)-H(Y^N\mid X^N)\\
&=-\sum P(b_j)\log P(b_j)+\sum P(a_ib_j)\log P(b_j\mid a_i)\\
&=\sum P(a_ib_j)\log\frac{P(b_j\mid a_i)}{P(b_j)}\\
&=\sum P(a_ib_j)\log\frac{P(b_{h_1}\mid a_{k_1})P(b_{h_2}\mid a_{k_2})\cdots P(b_{h_N}\mid a_{k_N})}{P(b_{h_1})P(b_{h_2})\cdots P(_{h_N})}\\
&=\sum P(a_ib_j)\log\frac{P(b_{h_1}\mid a_{k_1})}{P(b_{h_1})}+\cdots+\sum P(a_ib_j)\log\frac{P(b_{h_N}\mid a_{k_N})}{P(b_{h_N})}\\
&=I(X_1;Y_1)+I(X_2;Y_2)+\cdots+I(X_N;Y_N)
\end{aligned}
$$

【3.14】 证明：若(X,Y,Z)是马氏链，则(Z,Y,X)也是马氏链。

证明：

如果(X,Y,Z)是马氏链，则有$P(z\mid xy)=P(z\mid y)$，即

$$\frac{P(xyz)}{P(xy)}=\frac{P(yz)}{P(y)}$$

因此有$\frac{P(xyz)}{P(yz)}=\frac{P(xy)}{P(y)}$，即$P(x\mid yz)=P(x\mid y)$，即$(Z,Y,X)$也是马氏链。

【3.15】把n个二元对称信道串接起来,每个二元对称信道的错误传递概率为p。证明这n个串接信道可以等效于一个二元对称信道，其错误传递概率为$\frac{1}{2}[1-(1-2p)^n]$，并证明$\lim_{n\to\infty}I(X_0;X_n)=0$，设$p\neq 0$或 1，信道的串接如下图所示。

$$X_0\to\boxed{\text{BSC1}}\to X_1\to\boxed{\text{BSC2}}\to X_2\to\cdots\cdots\to\boxed{\text{BSCn}}\to X_n$$

证明：

当$n=1$时，错误概率$p=\frac{1}{2}(1-(1-2p))$成立；

假设$n=k$成立，即k个串接信道的错误概率为$\frac{1}{2}[1-(1-2p)^k]$；

当$n=k+1$时，其错误概率为：

$$
\begin{aligned}
&\bar{p}\frac{1}{2}[1-(1-2p)^k]+p\left(1-\frac{1}{2}[1-(1-2p)^k]\right)\\
&=\frac{\bar{p}}{2}-\frac{\bar{p}}{2}(1-2p)^k+p-\frac{p}{2}[1-(1-2p)^k]\\
&=\frac{1}{2}-\frac{\bar{p}}{2}(1-2p)^k+\frac{p}{2}(1-2p)^k\\
&=\frac{1}{2}-\frac{1}{2}(1-2p)^k+p(1-2p)^k\\
&=\frac{1}{2}[1-(1-2p)^{k+1}]
\end{aligned}
$$

当 $n\to\infty$ 时，错误概率近似为 $\frac{1}{2}$，总信道矩阵为 $\begin{bmatrix}\frac{1}{2} & \frac{1}{2}\\ \frac{1}{2} & \frac{1}{2}\end{bmatrix}$，此时不论输入为何分布，输出均为等概率分布。其互信息为:

$$\lim_{n\to\infty} I(X_0;X_n)=H(X_n)-H(X_n\mid X_0)=1-H(X_n\mid X_0)=0\text{比特/符号}$$

【3.16】 若有两个串接的离散信道，它们的信道矩阵都是

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

并设第一个信道的输入符号 $X\in\{a_1,a_2,a_3,a_4\}$ 是等概率分布，求 $I(X;Z)$ 和 $I(X;Y)$ 并加以比较。

解:

串接后的信道矩阵为:

$$\begin{bmatrix}0 & 0 & 0 & 1\\ 0 & 0 & 0 & 1\\ \frac{1}{2} & \frac{1}{2} & 0 & 0\\ 0 & 0 & 1 & 0\end{bmatrix}\begin{bmatrix}0 & 0 & 0 & 1\\ 0 & 0 & 0 & 1\\ \frac{1}{2} & \frac{1}{2} & 0 & 0\\ 0 & 0 & 1 & 0\end{bmatrix}=\begin{bmatrix}0 & 0 & 1 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\\ \frac{1}{2} & \frac{1}{2} & 0 & 0\end{bmatrix}$$

$$\begin{bmatrix}P(b_1) & P(b_2) & P(b_3) & P(b_4)\end{bmatrix} = \begin{bmatrix}P(a_1) & P(a_2) & P(a_3) & P(a_4)\end{bmatrix}\begin{bmatrix}0 & 0 & 0 & 1\\ 0 & 0 & 0 & 1\\ \frac{1}{2} & \frac{1}{2} & 0 & 0\\ 0 & 0 & 1 & 0\end{bmatrix}$$

$$= \begin{bmatrix}\frac{1}{8} & \frac{1}{8} & \frac{1}{4} & \frac{1}{2}\end{bmatrix}$$

$$\begin{aligned}I(X;Y) &= H(Y) - H(Y|X)\\ &= -\frac{1}{8}\log\frac{1}{8} - \frac{1}{8}\log\frac{1}{8} - \frac{1}{4}\log\frac{1}{4} - \frac{1}{2}\log\frac{1}{2} - \frac{1}{4}\times\frac{1}{2}\log\frac{1}{2} - \frac{1}{4}\times\frac{1}{2}\log\frac{1}{2}\\ &= 1.5\text{比特/符号}\end{aligned}$$

$$\begin{bmatrix}P(c_1) & P(c_2) & P(c_3) & P(c_4)\end{bmatrix} = \begin{bmatrix}P(a_1) & P(a_2) & P(a_3) & P(a_4)\end{bmatrix}\begin{bmatrix}0 & 0 & 1 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\\ \frac{1}{2} & \frac{1}{2} & 0 & 0\end{bmatrix}$$

$$= \begin{bmatrix}\frac{1}{8} & \frac{1}{8} & \frac{1}{2} & \frac{1}{4}\end{bmatrix}$$

$$\begin{aligned}I(X;Z) &= H(Z) - H(Z|X)\\ &= -\frac{1}{8}\log\frac{1}{8} - \frac{1}{8}\log\frac{1}{8} - \frac{1}{4}\log\frac{1}{4} - \frac{1}{2}\log\frac{1}{2} - \frac{1}{4}\times\frac{1}{2}\log\frac{1}{2} - \frac{1}{4}\times\frac{1}{2}\log\frac{1}{2}\\ &= 1.5\text{比特/符号}\end{aligned}$$

可见， $I(X;Z) = I(X;Y)$。

第四章课后习题

【4.1】 设有一连续随机变量，其概率密度函数为

$$p(x)=\begin{cases} A\cos x & |x|\le \frac{p}{2} \\ 0 & \text{其他值} \end{cases}$$

又有$\int_{-\frac{p}{2}}^{\frac{p}{2}} p(x)dx=1$，试求这随机变量的熵。

解:

$$\begin{aligned} h(X) &= -\int p(x)\log p(x)dx \\ &= -\int A\cos x\log Adx-\int A\cos x\log\cos xdx \\ &= -A\log A\sin x\Big|_{-\frac{p}{2}}^{\frac{p}{2}}-\int A\cos x\log\cos xdx \\ &= -2A\log A-\int A\cos x\log\cos xdx \end{aligned}$$

而

$$\begin{aligned} \int\cos x\log\cos xdx &= \log e\int\ln\sqrt{1-\sin^2 x}d\sin x \\ &= \frac{1}{2}\log e\int\ln(1+\sin x)+\ln(1-\sin x)d\sin x \\ &= \frac{1}{2}\log e\int\ln(1+\sin x)d\sin x+\frac{1}{2}\log e\int\ln(1-\sin x)d\sin x \end{aligned}$$

$$\begin{aligned} \int\ln(1+\sin x)d\sin x &= (1+\sin x)\ln(1+\sin x)\Big|_{-\frac{p}{2}}^{\frac{p}{2}}-\int\frac{1+\sin x}{1+\sin x}d\sin x \\ &= 2\ln 2-2 \end{aligned}$$

$$\begin{aligned} \int\ln(1-\sin x)d\sin x &= -\int\ln(1-\sin x)d(1-\sin x) \\ &= -(1-\sin x)\ln(1-\sin x)\Big|_{-\frac{p}{2}}^{\frac{p}{2}}-\int\frac{1-\sin x}{1-\sin x}d\sin x \\ &= 2\ln 2-2 \end{aligned}$$

因此有

$$
\begin{aligned}
h(X) &= -2A\log A-\frac{A}{2}\log e(2\ln 2-2+2\ln 2-2)\\
&= -2A\log A+2A\log e-2A\log e\ln 2\\
&= -2A\log A+2A\log e-2A
\end{aligned}
$$

而$\int_{-\frac{p}{2}}^{\frac{p}{2}} p(x)dx=1$，即 $A=\frac{1}{2}$，因此

$$
h(X)=-\log\frac{1}{2}+\log e-1=1+\log e-1=\log e
$$

【4.2】计算连续随机变量 X 的差熵

（1） 指数概率密度函数 $p(x)=l e^{-lx}$， $x\ge 0, l>0$

（2） 拉普拉斯概率密度函数， $p(x)=\frac{1}{2}l e^{-l|x|}$， $-\infty<x<\infty, l>0$

解：

（1）

$$
\begin{aligned}
h(X) &= -\int p(x)\log p(x)dx\\
&= -\int l e^{-lx}\log l e^{-lx}dx\\
&= -\int l e^{-lx}\log l\,dx-\int l e^{-lx}\log e^{-lx}dx\\
&= -\log l e^{-lx}\Big|_0^{\infty}+\log e\int \ln e^{-lx}de^{-lx}\\
&= -\log l+\log e t\ln t\Big|_1^0-\log e\int dt\\
&= -\log l+\log e\\
&= \log\frac{e}{l}
\end{aligned}
$$

(2)

$$
\begin{aligned}
h(X) &= -\int p(x)\log p(x)dx\\
&= -\int_{-\infty}^{\infty}\frac{1}{2}l e^{-l|x|}\log\frac{1}{2}l e^{-l|x|}dx\\
&= -\int_0^{\infty} l e^{-lx}\log\frac{1}{2}l e^{-lx}dx\\
&= -\int_0^{\infty} l e^{-lx}\log\frac{1}{2}dx-\int_0^{\infty} l e^{-lx}\log l e^{-lx}dx\\
&= \log 2+\log\frac{e}{l}\\
&= \log\frac{2e}{l}
\end{aligned}
$$

注：（2）题直接借用了（1）的结论。

【4.3】设有一连续随机变量，其概率密度函数为：

$$p(x)=\begin{cases} bx^2 & 0\le x\le a \\ 0 & \text{其他值} \end{cases}$$

试求这随机变量的熵。又若 $Y_1=X+K(K>0)$，$Y_2=2X$，试分别求出 Y_1 和 Y_2 的熵 $h(Y_1)$ 和 $h(Y_2)$。

解：

$$\begin{aligned} h(X) &= -\int p(x)\log p(x)dx \\ &= -\int_0^a bx^2\log bx^2 dx \\ &= -\log b-2b\log e\int_0^a x^2\ln x dx \\ &= \frac{2}{9}a^3 b\log e-\frac{2}{3}a^3 b\log a-\log b \end{aligned}$$

由于 $\int p(x)dx=1$，因此 $a^3b=3$，因此

$$h(X)=\frac{2}{3}\log e+\log a-\log 3$$

当 $Y_1=X+K(K>0)$ 时，$\left|\frac{\partial X}{\partial Y_1}\right|=1$，因此

$$h(Y_1)=h(X)-E[\log 1]=h(X)=\frac{2}{3}\log e+\log a-\log 3$$

当 $Y_2=2X$ 时，$\left|\frac{\partial X}{\partial Y_1}\right|=\frac{1}{2}$，因此

$$h(Y_1)=h(X)-E[\log\frac{1}{2}]=h(X)=\frac{2}{3}\log e+\log a\log\frac{3}{2}$$

【4.4】设给定两随机变量 X_1 和 X_2，它们的联合概率密度为

$$p(x_1x_2)=\frac{1}{2p}e^{-\frac{x_1^2+x_2^2}{2}} \qquad -\infty<x_1,x_2<\infty$$

求随机变量 $Y_1=X_1+X_2$ 的概率密度函数，并计算变量 Y 的熵 $h(Y)$。

解：

$$p(x_1x_2)=\frac{1}{2p}e^{-\frac{x_1^2+x_2^2}{2}}=\frac{1}{\sqrt{2p}}e^{-\frac{x_1^2}{2}}\frac{1}{\sqrt{2p}}e^{-\frac{x_2^2}{2}}=p(x_1)p(x_2)$$

因此 $Y_1=X_1+X_2$ 也是一个高斯分布的随机变量，其均值为 0，方差为 2，即

$$p(x_1 x_2) = \frac{1}{\sqrt{2p}} e^{-\frac{y^2}{4}}$$

因此其差熵为

$$h(Y) = \frac{1}{2}\log 2pes_y^2 = \frac{1}{2}\log 4pe$$

【4.5】设一连续消息通过某放大器，该放大器输出的最大瞬时电压b，最小瞬时电压为a。若消息从放大器中输出，问放大器输出消息在每个自由度上的最大熵是多少？又放大器的带宽为F，问单位时间内输出最大信息量是多少？

解：

该问题等价于取值受限的随机变量的最大熵，根据差熵的极值性，当等概率分布时其差熵最大，即

$$h(Y) = \log(b-a)$$

如果放大器的带宽为F，则取样率为$2F$，单位时间内输出的最大信息量为

$$2F\log(b-a) \text{ 比特/秒}$$

【4.6】有一信源发出恒定宽度，但不同幅度的脉冲，幅度值处在a_1和a_2之间，此信源连至某信道，信道接收端接收脉冲的幅度y处在b_1和b_2之间。已知随机变量X和Y的联合概率密度函数

$$p(xy) = \frac{1}{(a_2-a_1)(b_2-b_1)}$$

试计算$h(X)$，$h(Y)$，$h(XY)$和$I(X;Y)$。

解：

$$\begin{aligned} p(x) &= \int p(x,y)dy \\ &= \int \frac{1}{(a_2-a_1)(b_2-b_1)}dy \\ &= \frac{1}{a_2-a_1} \end{aligned}$$

同理，$p(y)=\dfrac{1}{b_2-b_1}$。

因此

$$h(X)=-\int p(x)\log p(x)dx=\log(a_2-a_1)$$

$$h(Y)=-\int p(y)\log p(y)dy=\log(b_2-b_1)$$

$$h(XY)=-\int p(x,y)\log p(x,y)dxdy=\log(a_2-a_1)+\log(b_2-b_1)$$

$$I(X;Y)=h(X)+h(Y)-h(XY)=0$$

【4.7】在连续信源中，根据差熵、条件差熵和联合差熵的定义，证明

（1） $h(X\mid Y)\le h(X)$，当且仅当 X 和 Y 统计独立时等号成立；

（2） $h(X_1X_2\cdots X_N)\le h(X_1)+h(X_2)+\cdots+h(X_N)$，当且仅当 $X_1X_2\cdots X_N$ 彼此统计独立时等式成立。

证明：

（1）

$$\begin{aligned}h(XY)\quad &=-\int p(y)dy\int p(x\mid y)\log p(x\mid y)dx\\ &\le-\int p(y)dy\int p(x\mid y)\log p(x)dx\\ &=-\int p(x,y)\log p(x)dxdy\\ &=h(X)\end{aligned}$$

等号成立当且仅当 $p(x\mid y)=p(x)$，即 $p(x,y)=p(x)p(y)$，因此仅当 X 和 Y 统计独立时等号成立。

（2）根据条件概率密度的相关公式，有

$$h(X_1X_2X_N)=h(X_1)+h(X_2\mid X_1)+h(X_3\mid X_1X_2)+\cdots+h(X_N\mid X_1X_2X_{N-1})$$

根据（1）的结论，条件差熵小于差熵，因此有

$$h(X_1X_2\cdots X_N)\le h(X_1)+h(X_2)+\cdots+h(X_N)$$

等号成立当且仅当

$$p(x_2 \mid x_1) = p(x_2)$$

$$p(x_3 \mid x_1 x_2) = p(x_3)$$

……

$$p(x_N \mid x_1 x_2 \cdots x_{N-1}) = p(x_N)$$

即

$$p(x_1 x_2) = p(x_1) p(x_2)$$

$$p(x_1 x_2 x_3) = p(x_1) p(x_2) p(x_3)$$

……

$$p(x_1 x_2 \cdots x_N) = p(x_1) p(x_2) \cdots p(x_N)$$

【4.8】设连续随机变量 X，已知 $X \geq 0$，其平均值受限，即数学期望为 A，试求在此条件下获得的最大熵的最佳分布，并求出最大熵。

解:

给定条件如下:

$$\int p(x)dx = 1$$

$$\int xp(x)dx = A$$

目标：求 $-\int p(x)\log p(x)dx$ 的最大值。

构造函数

$$\begin{aligned} F(p(x)) &= -\int p(x)\log p(x)dx + l\int p(x)dx + m\int xp(x)dx \\ &= \int \left(-p(x)\log p(x) + l p(x) + m x p(x)\right)dx \end{aligned}$$

欲使 $\dfrac{dF(p(x))}{dp(x)} = 0$，只需 $\dfrac{d\left(-p(x)\log p(x) + l p(x) + mxp(x)\right)}{dp(x)} = 0$ 即可，因此有

$$-\log p(x) - \log e + l + mx = 0$$

$$p(x) = 2^{l + mx - \log e}$$

根据$\int p(x)dx = 1$，$\int xp(x)dx = A$，可得

$$\int 2^{l+mx-\log e} dx = 1 \Rightarrow m = -2^{l-\log e}$$

$$\int xp(x)dx = A \Rightarrow m = -\frac{1}{A}(\log e)^2$$

因此 $p(x) = \frac{1}{A}(\log e)^2 2^{-\frac{x}{A}(\log e)^2}$，此时

$$\begin{aligned} h(X) &= -\int p(x)\log p(x)dx \\ &= -\log\left(\frac{1}{A}(\log e)^2\right) + (\log e)^2 \end{aligned}$$

【4.9】N 维连续型随机序列 $X_1X_2 \mathbf{L}\, X_N$，有概率密度 $p(X_1X_2 \mathbf{L}\, X_N)$ 以及 $E[(X_i = m_i)] = s_i^2$。证明：当随机序列的分量各自达到正态分布并彼此统计独立时熵最大。最大熵为

$$\frac{N}{2}\log 2pe(s_1^2 s_2^2 \mathbf{L}\, s_N^2)^{1/N}$$

证明：

$$h(X_1X_2 \mathbf{L}\, X_N) \le h(X_1) + h(X_2) + \mathbf{L} + h(X_N)$$

等号成立当且仅当各分量统计独立。

而对于任何一个分量而言，当 $E[(X_i = m_i)] = s_i^2$ 时，高斯分布的差熵最大，为

$$h(X_i) = \frac{1}{2}\log 2pes_i^2$$

因此原序列差熵的最大值为：

$$\begin{aligned} h(X_1X_2 \mathbf{L}\, X_N) &= \frac{1}{2}\log 2pes_1^2 + \frac{1}{2}\log 2pes_2^2 + \mathbf{L} + \frac{1}{2}\log 2pes_N^2 \\ &= \frac{N}{2}\log\left[2pe\left(s_1^2 s_2^2 \mathbf{L}\, s_N^2\right)^{\frac{1}{N}}\right] \end{aligned}$$

【4.10】N 维连续型随机序列 $X_1X_2 \mathbf{L}\, X_N$，其各分量幅度分别受限为$[a_i, b_i]$。证明：当随机序列的分量各自达到均匀分布并彼此统计独立时熵最大。最大熵为

$$\log\prod_{i=1}^{N}(b_i - a_i)$$

证明:

$$h(X_1X_2\cdots X_N)\le h(X_1)+h(X_2)+\cdots+h(X_N)$$

等号成立当且仅当各分量统计独立。

而对于任何一个分量而言，当幅度分别受限为$[a_i,b_i]$时，均匀分布的差熵最大，为

$$h(X_i)=\log\left(b_i-a_i\right)$$

因此原序列差熵的最大值为:

$$\begin{aligned}h(X_1X_2\cdots X_N)\quad &=\log\left(b_1-a_1\right)+\log\left(b_2-a_2\right)+\cdots+\log\left(b_N-a_N\right)\\ &=\log\prod_{i=1}^{N}\left(b_i-a_i\right)\end{aligned}$$

【例1】设$X_1X_2\cdots X_N$都是互相独立的正态分布的随机变量，其方差分别为σ_1^2，σ_2^2，…，σ_N^2，均值分别为$\mu_1,\mu_2,\cdots,\mu_N$。试证明$Y=X_1+X_2+\cdots+X_N$仍是正态随机变量，其均值为$\mu=\sum\mu_i$，方差$\sigma^2=\sum\sigma_i^2$。

证明:

设X_1和X_2是相互独立的正态分布的随机变量，其均值为μ_i，方差为σ_i^2。设$Y_2=X_1+X_2$，根据已知条件，有

$$\begin{aligned}X_1&=X_1\\ Y_2&=X_1+X_2\end{aligned}$$

因此有

$$\begin{aligned}p(x_1,y_2)\quad &=p(x_1,x_2)\begin{vmatrix}\dfrac{\partial X_1}{\partial X_1} & \dfrac{\partial Y_2}{\partial X_1}\\ \dfrac{\partial X_1}{\partial X_2} & \dfrac{\partial Y_2}{\partial X_2}\end{vmatrix}\\ &=p(x_1,x_2)\begin{vmatrix}1&1\\0&1\end{vmatrix}\\ &=p(x_1,x_2)=p(x_1)p(x_2)\\ &=p(x_1)p(y_2-x_1)\end{aligned}$$

因此有

$$
\begin{aligned}
p(y_2) &= \int p(x_1 y_2) dx_1 \\
&= \int p(x_1) p(y_2 - x_1) dx_1 \\
&= \int \frac{1}{\sqrt{2p}\, s_1} \exp\left\{-\frac{(x_1 - m_1)^2}{2s_1^2}\right\} \frac{1}{\sqrt{2p}\, s_2} \exp\left\{-\frac{(y_2 - x_1 - m_2)^2}{2s_2^2}\right\} dx_1 \\
&= \int \frac{1}{2p s_1 s_2} \exp\left\{-\frac{(x_1 - m_1)^2}{2s_1^2} - \frac{(y_2 - x_1 - m_2)^2}{2s_2^2}\right\} dx_1
\end{aligned}
$$

而

$$
\begin{aligned}
& -\frac{(x_1 - m_1)^2}{2s_1^2} - \frac{(y_2 - x_1 - m_2)^2}{2s_2^2} \\
&= -\frac{1}{2}\left\{\frac{(x_1 - m_1)^2}{s_1^2} + \frac{(y_2 - x_1 - m_2)^2}{s_2^2}\right\} \\
&= -\frac{1}{2s_1^2 s_2^2}\left\{s_2^2 (x_1 - m_1)^2 + s_1^2 (y_2 - x_1 - m_2)^2\right\} \\
&= -\frac{1}{2s_1^2 s_2^2}\left\{\left(s_1^2 + s_2^2\right) x_1^2 + 2x_1\left(s_1^2 m_2 - s_1^2 y_2 - m_1 s_2^2\right) + \left(s_2^2 m_1^2 + s_1^2 m_2^2 + s_1^2 y_2^2 - 2s_1^2 y_2 m_2\right)\right\} \\
&= -\frac{1}{2}\left(\frac{1}{s_1^2} + \frac{1}{s_2^2}\right)\left\{x_1^2 + 2x_1\left(\frac{s_1^2 m_2 - s_1^2 y_2 - m_1 s_2^2}{s_1^2 + s_2^2}\right) + \left(\frac{s_2^2 m_1^2 + s_1^2 y_2^2 - 2s_1^2 y_2 m_2}{s_1^2 + s_2^2}\right)\right\} \\
&= -\frac{1}{2}\left(\frac{1}{s_1^2} + \frac{1}{s_2^2}\right)\left\{\left(x_1 + \frac{s_1^2 m_2 - s_1^2 y_2 - m_1 s_2^2}{s_1^2 + s_2^2}\right)^2 + \frac{s_1^2 s_2^2}{\left(s_1^2 + s_2^2\right)^2}(y_2 - m_1 - m_2)^2\right\} \\
&= -\frac{1}{2}\left(\frac{1}{s_1^2} + \frac{1}{s_2^2}\right)\left(x_1 + \frac{s_1^2 m_2 - s_1^2 y_2 - m_1 s_2^2}{s_1^2 + s_2^2}\right)^2 - \frac{1}{2(s_1^2 + s_2^2)}(y_2 - m_1 - m_2)^2
\end{aligned}
$$

所以

$$
\begin{aligned}
p(y_2) &= \int \frac{1}{2p s_1 s_2} \exp\left\{-\frac{(x_1-m_1)^2}{2s_1^2} - \frac{(y_2-x_1-m_2)^2}{2s_2^2}\right\} dx_1 \\
&= \frac{1}{2p s_1 s_2} \exp\left\{-\frac{1}{2(s_1^2+s_2^2)}(y_2-m_1-m_2)^2\right\} \\
&\quad \int \exp\left\{-\frac{1}{2}\left(\frac{1}{s_1^2}+\frac{1}{s_2^2}\right)\left(x_1+\frac{s_1^2 m_2 - s_1^2 y_2 - m_1 s_2^2}{s_1^2+s_2^2}\right)^2\right\} dx_1 \\
&= \frac{1}{2p s_1 s_2} \exp\left\{-\frac{1}{2(s_1^2+s_2^2)}(y_2-m_1-m_2)^2\right\} \sqrt{\frac{p}{\frac{1}{2}\left(\frac{1}{s_1^2}+\frac{1}{s_2^2}\right)}} \\
&= \frac{1}{\sqrt{2p\left(s_1^2+s_2^2\right)}} \exp\left\{-\frac{(y_2-m_1-m_2)^2}{2(s_1^2+s_2^2)}\right\}
\end{aligned}
$$

因此，$Y_2 = X_1 + X_2$ 是均值为 $m_1 + m_2$，方差为 $s_1^2 + s_2^2$ 的高斯分布，同理，$Y_3 = X_1 + X_2 + X_3$，……，$Y = X_1 + X_2 + \mathbf{L} + X_N$ 均为高斯分布，因此

$Y = X_1 + X_2 + \mathbf{L} + X_N$ 是正态随机变量，其均值为 $m = \sum m_i$，方差 $s^2 = \sum s_i^2$

【4.12】设某连续信道，其特性如下：

$$p(y\,|\,x) = \frac{1}{a\sqrt{3p}} e^{-(y-\frac{1}{2}x)^2/3a^2}$$

而且输入变量 X 的概率密度函数为

$$p(x) = \frac{1}{2a\sqrt{p}} e^{-x^2/4a^2}$$

试计算：

（1） 信源的熵 $h(X)$；

（2） 平均互信息 $I(X;Y)$。

解：

$$
\begin{aligned}
p(x) &= \frac{1}{2a\sqrt{p}} e^{-x^2/4a^2} \\
&= \frac{1}{\sqrt{2p \bullet 2a^2}} e^{-x^2/2\bullet 2a^2}
\end{aligned}
$$

可见，X 为均值为 0，方差为 $2a^2$ 的正态分布，其差熵为

$$h(X)=\frac{1}{2}\log 2\pi e 2a^2=\frac{1}{2}\log 4\pi e a^2$$

$$\begin{aligned}
p(x,y) &= p(x)p(y\mid x)\\
&=\frac{1}{2\sqrt{3}\pi a^2}\exp\left\{-\frac{x^2}{4a^2}-\frac{\left(y-\frac{1}{2}x\right)^2}{3a^2}\right\}\\
&=\frac{1}{2\sqrt{3}\pi a^2}\exp\left\{-\frac{4x^2+4y^2-4xy}{12a^2}\right\}\\
&=\frac{1}{2\sqrt{3}\pi a^2}\exp\left\{-\frac{\left(x-\frac{1}{2}y\right)^2+\frac{3}{4}y^2}{3a^2}\right\}\\
&=\frac{1}{2\sqrt{3}\pi a^2}\exp\left\{-\frac{\left(x-\frac{1}{2}y\right)^2}{3a^2}-\frac{y^2}{4a^2}\right\}
\end{aligned}$$

而

$$\begin{aligned}
p(y) &=\int p(x,y)dx=\int\frac{1}{2\sqrt{3}\pi a^2}\exp\left\{-\frac{\left(x-\frac{1}{2}y\right)^2}{3a^2}-\frac{y^2}{4a^2}\right\}dx\\
&=\frac{1}{2\sqrt{3}\pi a^2}e^{-\frac{y^2}{4a^2}}\int e^{-\left(\frac{x-\frac{1}{2}y}{\sqrt{3}a}\right)^2}dx=\frac{1}{2\pi a}e^{-\frac{y^2}{4a^2}}\int e^{-\left(\frac{x-\frac{1}{2}y}{\sqrt{3}a}\right)^2}d\left(\frac{x-\frac{1}{2}y}{\sqrt{3}a}\right)\\
&=\frac{1}{2\pi a}e^{-\frac{y^2}{4a^2}}\int e^{-t^2}dt\\
&=\frac{1}{2a\sqrt{\pi}}e^{-\frac{y^2}{4a^2}}
\end{aligned}$$

因此 Y 是均值为 0，方差为 $2a^2$ 的高斯分布，其差熵为

$$h(Y)=\frac{1}{2}\log 2\pi e\times 2a^2=\frac{1}{2}\log 4\pi e a^2$$

而条件熵为

$$
\begin{aligned}
h(Y\mid X) &= -\int p(x)p(y\mid x)\log p(y\mid x)dxdy \\
&= -\int p(x,y)\log\left(\frac{1}{a\sqrt{3p}}e^{-\frac{\left(y-\frac{1}{2}x\right)^2}{3a^2}}\right)dxdy \\
&= -\int p(x,y)\log\frac{1}{a\sqrt{3p}}dxdy - \int p(x,y)\log e^{-\frac{\left(y-\frac{1}{2}x\right)^2}{3a^2}}dxdy \\
&= \log\sqrt{3p}\,a + \log e\int p(x,y)\frac{\left(y-\frac{1}{2}x\right)^2}{3a^2}dxdy
\end{aligned}
$$

而

$$
\int p(x,y)\frac{\left(y-\frac{1}{2}x\right)^2}{3a^2}dxdy = \int\frac{\left(y-\frac{1}{2}x\right)^2}{3a^2}\frac{1}{2\sqrt{3}p\,a^2}e^{-\frac{\left(x-\frac{1}{2}y\right)^2}{3a^2}-\frac{y^2}{4a^2}}dxdy
$$

$$
=\int\frac{1}{6\sqrt{3}p\,a^4}\left(y-\frac{1}{2}x\right)^2e^{-\frac{y^2-xy+x^2}{3a^2}}dxdy = \int\frac{1}{6\sqrt{3}p\,a^4}\left(y-\frac{1}{2}x\right)^2e^{-\frac{\left(y-\frac{1}{2}x\right)^2+\frac{3}{4}x^2}{3a^2}}dxdy
$$

$$
=\int\frac{1}{6\sqrt{3}p\,a^4}\left(y-\frac{1}{2}x\right)^2e^{-\frac{\left(y-\frac{1}{2}x\right)^2}{3a^2}-\frac{x^2}{4a^2}}dxdy
$$

$$
=\frac{1}{6\sqrt{3}p\,a^4}\int e^{-\frac{x^2}{4a^2}}dx\int\left(y-\frac{1}{2}x\right)^2e^{-\frac{\left(y-\frac{1}{2}x\right)^2}{3a^2}}dy
$$

而

$$
\int\left(y-\frac{1}{2}x\right)^2e^{-\frac{\left(y-\frac{1}{2}x\right)^2}{3a^2}}dy = 3a^2\sqrt{3}a\int\left(\frac{y-\frac{1}{2}x}{\sqrt{3}a}\right)^2e^{-\left(\frac{y-\frac{1}{2}x}{\sqrt{3}a}\right)^2}d\left(\frac{y-\frac{1}{2}x}{\sqrt{3}a}\right)
$$

$$
\begin{aligned}
&= 3\sqrt{3}a^3\int t^2e^{-t^2}dt \\
&= \frac{3}{2}\sqrt{3p}\,a^3
\end{aligned}
$$

因此

$$\int p(x,y)\frac{\left(y-\frac{1}{2}x\right)^2}{3a^2}dxdy=\frac{1}{6\sqrt{3}pa^4}\int e^{-\frac{x^2}{4a^2}}dx\int\left(y-\frac{1}{2}x\right)^2 e^{-\frac{\left(y-\frac{1}{2}x\right)^2}{3a^2}}dy$$
$$=\frac{1}{6\sqrt{3}pa^4}\int e^{-\frac{x^2}{4a^2}}dx\bullet\frac{3}{2}\sqrt{3p}\,a^3=\frac{1}{4a\sqrt{p}}\int e^{-\frac{x^2}{4a^2}}dx$$
$$=\frac{1}{4a\sqrt{p}}\bullet 2a\sqrt{p}$$
$$=\frac{1}{2}$$

$$\begin{aligned}h(Y|X)&=\log\sqrt{3p}\,a+\log e\int p(x,y)\frac{\left(y-\frac{1}{2}x\right)^2}{3a^2}dxdy\\&=\log\sqrt{3p}\,a+\frac{1}{2}\log e\\&=\frac{1}{2}\log 3pea^2\end{aligned}$$

因此平均互信息为:

$$\begin{aligned}I(X;Y)&=H(Y)-H(Y|X)\\&=\frac{1}{2}\log 4pea^2-\frac{1}{2}\log 3pea^2\\&=\frac{1}{2}\log\frac{4}{3}\\&=0.21\end{aligned}$$

注：该题推导过程中引用的相关积分公式:

(1) $\int_{-\infty}^{\infty}e^{-\frac{1}{2}t^2}dt=\sqrt{2p}$

(2) $\int_{-\infty}^{\infty}t^2e^{-t^2}dt=\frac{\sqrt{p}}{2}$

【4.13】试证明两连续随机变量之间的平均互信息 $I(X;Y)$ 是输入随机变量 X 的概率密度函数 $p(x)$ 的 I 型凸函数。

证明:

$$\begin{aligned} I(X;Y) &= \int p(x)p(y\,|\,x)\log\frac{p(y\,|\,x)}{p(y)}dxdy \\ &= \int p(x)p(y\,|\,x)\log\frac{p(y\,|\,x)}{\int p(x)p(y\,|\,x)dx}dxdy \end{aligned}$$

设存在 X 的两个概率密度 $p_1(x)$ 和 $p_2(x)$，参数 $0 \le q \le 1$，目标证明：

$$I(qp_1(x)+\bar{q}p_2(x)) \ge qI(p_1(x))++\bar{q}I(p_2(x))$$

过程如下：

$$\begin{aligned} & qI(p_1(x))++\bar{q}I(p_2(x))-I(qp_1(x)+\bar{q}p_2(x)) \\ &= q\int p_1(x)p(y\,|\,x)\log\frac{p(y\,|\,x)}{p_1(y)}dxdy+\bar{q}\int p_2(x)p(y\,|\,x)\log\frac{p(y\,|\,x)}{p_2(y)}dxdy \\ &-\int p(x)p(y\,|\,x)\log\frac{p(y\,|\,x)}{p(y)}dxdy \\ &= q\int p_1(x)p(y\,|\,x)\log\frac{p(y)}{p_1(y)}dxdy+\bar{q}\int p_2(x)p(y\,|\,x)\log\frac{p(y)}{p_2(y)}dxdy \end{aligned}$$

而

$$\begin{aligned} \int p_1(x)p(y\,|\,x)\log\frac{p(y)}{p_1(y)}dxdy &= \int p_1(x,y)\log\frac{p(y)}{p_1(y)}dxdy \\ &\le \log\int p_1(x,y)\frac{p(y)}{p_1(y)}dxdy \\ &= \log\int p_1(x\,|\,y)p(y)dxdy \\ &= 0 \end{aligned}$$

同理，$\int p_2(x)p(y\,|\,x)\log\dfrac{p(y)}{p_2(y)}dxdy \le 0$，因此有

$$I(qp_1(x)+\bar{q}p_2(x)) \ge qI(p_1(x))++\bar{q}I(p_2(x))$$

【4.14】试证明多维连续无记忆信道的充要条件为

$$p(y\,|\,x)=\prod_{i=1}^{N}p(y_i\,|\,x_i)$$

证明：

（1）充分性。

$$\begin{aligned} & p(y_1y_2\cdots y_N\,|\,x_1x_2\cdots x_N) \\ &= p(y_1\,|\,x_1x_2\cdots x_N)p(y_2\,|\,x_1x_2\cdots x_Ny_1)\cdots p(y_N\,|\,x_1x_2\cdots x_Ny_1y_2\cdots y_{N-1}) \end{aligned}$$

而

$$
\begin{aligned}
p(y_N \mid x_1x_2\cdots x_N y_1y_2\cdots y_{N-1}) &= \frac{p(x_1x_2\cdots x_N y_1y_2\cdots y_{N-1}y_N)}{p(x_1x_2\cdots x_N y_1y_2\cdots y_{N-1})} \\
&= \frac{p(y_1y_2\cdots y_{N-1}y_N \mid x_1x_2\cdots x_N)}{p(y_1y_2\cdots y_{N-1} \mid x_1x_2\cdots x_N)} \\
&= \frac{\prod_{i=1}^{N} p(y_i \mid x_i)}{\int p(y_1y_2\cdots y_{N-1}y_N \mid x_1x_2\cdots x_N)dy_N} \\
&= \frac{\prod_{i=1}^{N} p(y_i \mid x_i)}{\int \prod_{i=1}^{N} p(y_i \mid x_i)dy_N} = \frac{\prod_{i=1}^{N} p(y_i \mid x_i)}{\prod_{i=1}^{N-1} p(y_i \mid x_i)} \\
&= p(y_N \mid x_N)
\end{aligned}
$$

同理

$$p(y_{N-1} \mid x_1x_2\cdots x_N y_1y_2\cdots y_{N-2}) = p(y_{N-1} \mid x_{N-1})$$

$$\cdots\cdots$$

$$p(y_2 \mid x_1x_2\cdots x_N y_1) = p(y_2 \mid x_2)$$

$$p(y_1 \mid x_1x_2\cdots x_N) = p(y_1 \mid x_1)$$

因此该信道是无记信道。

（2）必要性。

根据无记信道的性质，有

$$p(y_{N-1} \mid x_1x_2\cdots x_N y_1y_2\cdots y_{N-2}) = p(y_{N-1} \mid x_{N-1})$$

$$\cdots\cdots$$

$$p(y_2 \mid x_1x_2\cdots x_N y_1) = p(y_2 \mid x_2)$$

$$p(y_1 \mid x_1x_2\cdots x_N) = p(y_1 \mid x_1)$$

而

$$
\begin{aligned}
& p(y_1y_2\cdots y_N \mid x_1x_2\cdots x_N) \\
&= p(y_1 \mid x_1x_2\cdots x_N)p(y_2 \mid x_1x_2\cdots x_N y_1)\cdots p(y_N \mid x_1x_2\cdots x_N y_1y_2\cdots y_{N-1})
\end{aligned}
$$

因此有

$$p(y|x)=\prod_{i=1}^{N}p(y_i|x_i)$$

【4.15】试证明连续信源 X 的相对熵 $h(X)$ 是概率密度 $p(x)$ 的 I 型凸函数。

证明:

设存在 X 的两个概率密度 $p_1(x)$ 和 $p_2(x)$，参数 $0\le q\le 1$，目标证明:

$$h(qp_1(x)+\bar{q}p_2(x))\ge qh(p_1(x))++\bar{q}h(p_2(x))$$

过程如下:

$$\begin{aligned}
&qh(p_1(x))++\bar{q}h(p_2(x))-h(qp_1(x)+\bar{q}p_2(x))\\
&=-q\int p_1(x)\log p_1(x)dx-\bar{q}\int p_2(x)\log p_2(x)dx+\int\left(qp_1(x)+\bar{q}p_2(x)\right)\log p(x)dx\\
&=q\int p_1(x)\log\frac{p(x)}{p_1(x)}dx+\bar{q}\int p_2(x)\log\frac{p(x)}{p_2(x)}dx
\end{aligned}$$

而

$$\begin{aligned}
\int p_1(x)\log\frac{p(x)}{p_1(x)}dx\quad&\le\log\int p_1(x)\frac{p(x)}{p_1(x)}dx\\
&=\log 1\\
&=0
\end{aligned}$$

同理，$\int p_2(x)\log\frac{p(x)}{p_2(x)}dx\le 0$

因此

$$qh(p_1(x))++\bar{q}h(p_2(x))-h(qp_1(x)+\bar{q}p_2(x))\le 0$$

【4.16】设信道输入是连续型随机序列 $X_1X_2\cdots X_N$，输出也是连续型随机序列 $Y_1Y_2\cdots Y_N$，信道传递概率密度为 $p(y|x)$。试证明:

（1）当信源是无记忆时，有

$$I(X_1X_2\cdots X_N;Y_1Y_2\cdots Y_N)\ge\sum I(X_i;Y_i)$$

（2）当信道是无记忆时，有

$$I(X_1X_2\cdots X_N;Y_1Y_2\cdots Y_N)\le\sum I(X_i;Y_i)$$

证明：

$$
\begin{aligned}
& I(X_1X_2\cdots X_N;Y_1Y_2\cdots Y_N)\\
&=\int p(x_1x_2\cdots x_N,y_1y_2\cdots y_N)\log\frac{p(x_1x_2\cdots x_N\mid y_1y_2\cdots y_N)}{p(x_1x_2\cdots x_N)}\\
&=\int p(x_1x_2\cdots x_N,y_1y_2\cdots y_N)\log\frac{p(y_1y_2\cdots y_N\mid x_1x_2\cdots x_N)}{p(y_1y_2\cdots y_N)}
\end{aligned}
$$

$$
\begin{aligned}
& \sum I(X_i;Y_i)\\
&=\int p(x_i,y_j)\log\frac{p(x_i\mid y_j)}{p(x_i)}dx_1dx_2\cdots dx_Ndy_1dy_2\cdots dy_N\\
&=\int p(x_1x_2\cdots x_N,y_1y_2\cdots y_N)\log\frac{p(x_1\mid y_1)p(x_2\mid y_2)\cdots p(x_N\mid y_N)}{p(x_1)p(x_2)\cdots p(x_N)}dx_1dx_2\cdots dx_Ndy_1dy_2\cdots dy_N\\
&=\int p(x_1x_2\cdots x_N,y_1y_2\cdots y_N)\log\frac{p(y_1\mid x_1)p(y_2\mid x_2)\cdots p(y_N\mid x_N)}{p(y_1)p(y_2)\cdots p(y_N)}dx_1dx_2\cdots dx_Ndy_1dy_2\cdots dy_N
\end{aligned}
$$

（1）当信源无记忆时，即 $p(x_1x_2\cdots x_N)=p(x_1)p(x_2)\cdots p(x_N)$

$$
\begin{aligned}
& \sum I(X_i;Y_i)-I(X_1X_2\cdots X_N;Y_1Y_2\cdots Y_N)\\
&=\int p(x_1x_2\cdots x_N,y_1y_2\cdots y_N)\left(\log\frac{p(x_1\mid y_1)p(x_2\mid y_2)\cdots p(x_N\mid y_N)}{p(x_1x_2\cdots x_N\mid y_1y_2\cdots y_N)}\right)dx_1\cdots dx_Ndy_1\cdots dy_N\\
&\le\log\int p(x_1x_2\cdots x_N,y_1y_2\cdots y_N)\frac{p(x_1\mid y_1)p(x_2\mid y_2)\cdots p(x_N\mid y_N)}{p(x_1x_2\cdots x_N\mid y_1y_2\cdots y_N)}dx_1\cdots dx_Ndy_1\cdots dy_N\\
&=\log\int p(x_1\cdots x_N,y_1\cdots y_N)\frac{p(x_1\mid y_1)\cdots p(x_N\mid y_N)p(y_1y_2\cdots y_N)}{p(x_1\cdots x_N,y_1\cdots y_N)}dx_1\cdots dx_Ndy_1\cdots dy_N\\
&=\log\int p(x_1\mid y_1)\cdots p(x_N\mid y_N)p(y_1y_2\cdots y_N)dx_1\cdots dx_Ndy_1\cdots dy_N\\
&=1
\end{aligned}
$$

等号成立当且仅当 $p(x_1\mid y_1)p(x_2\mid y_2)\cdots p(x_N\mid y_N)=p(x_1x_2\cdots x_N\mid y_1y_2\cdots y_N)$。

当 $p(y_1\cdots y_N\mid x_1\cdots x_N)=p(y_1\mid x_1)\cdots p(y_N\mid x_N)$ 时，根据信源的无记忆性，即

$$p(y_1\cdots y_N\mid x_1\cdots x_N)p(x_1\cdots x_N)=p(y_1\mid x_1)\cdots p(y_N\mid x_N)p(x_1)\cdots p(x_N)$$

即

$$p(x_1\cdots x_N,y_1\cdots y_N)=p(x_1,y_1)\cdots p(x_N,y_N) \qquad (1)$$

两边对各自由度积分得

$$p(y_1\cdots y_N)=p(y_1)\cdots p(y_N) \qquad (2)$$

（1）式两边除以（2）式两边得

$$p(x_1 | y_1) p(x_2 | y_2) \cdots p(x_N | y_N) = p(x_1 x_2 \cdots x_N | y_1 y_2 \cdots y_N)$$

因此等号成立当且仅当连续信道无记忆。

（2）当信道无记忆时，即 $p(y_1 \cdots y_N | x_1 \cdots x_N) = p(y_1 | x_1) \cdots p(y_N | x_N)$ 时，

$$\begin{aligned}
& I(X_1 X_2 \cdots X_N; Y_1 Y_2 \cdots Y_N) - \sum I(X_i; Y_i) \\
& = \int p(x_1 x_2 \cdots x_N, y_1 y_2 \cdots y_N) \log \frac{p(y_1) p(y_2) \cdots p(y_N)}{p(y_1 y_2 \cdots y_N)} dx_1 \cdots dx_N dy_1 \cdots dy_N \\
& \le \log \int p(x_1 x_2 \cdots x_N, y_1 y_2 \cdots y_N) \frac{p(y_1) p(y_2) \cdots p(y_N)}{p(y_1 y_2 \cdots y_N)} dx_1 \cdots dx_N dy_1 \cdots dy_N \\
& = 0
\end{aligned}$$

等号成立当且仅当 $p(y_1 \cdots y_N) = p(y_1) \cdots p(y_N)$。

当 $p(x_1 x_2 \cdots x_N) = p(x_1) p(x_2) \cdots p(x_N)$，根据信道的无记忆性，有

$$p(y_1 \cdots y_N | x_1 \cdots x_N) = p(y_1 | x_1) \cdots p(y_N | x_N)$$

因此有

$$p(x_1 \cdots x_N, y_1 \cdots y_N) = p(x_1, y_1) \cdots p(x_N, y_N)$$

两边对自由度 $x_1 \cdots x_N$ 求和得

$$p(y_1 \cdots y_N) = p(y_1) \cdots p(y_N)$$

即（2）式等号成立的条件是信源为无记忆信源。

【4.17】在图片传输中，每帧约 2.25×10^6 个像素，为了能很好地重现图像，需分 16 个亮度电平，并假设亮度电平等概率分布。试计算每秒钟传送 30 帧图片所需信道的带宽（信噪功率比为 30dB）。

解:

每秒需要传输的信息量为:

$$2.25\times10^6 \times 30 \times \log 16 = 2.7\times10^8 \text{ 比特/秒}$$

信道的信噪比为 30dB，即$10\log_{10}\frac{P_S}{P_N}=30$，即$\frac{P_S}{P_N}=1000$

因此信道容量为

$$\begin{aligned} C &= W\log\left(1+\frac{P_S}{P_N}\right) \\ &= 2.7\times10^8 \end{aligned}$$

可推得带宽为

$$W=\frac{2.7\times10^8}{\log1001}=2.71\times10^7\text{ Hz}$$

【4.18】设在平均功率受限高斯加性波形信道中，信道带宽为 3kHz，又设（信号功率+噪声功率）/噪声功率=10dB。

（1） 试计算该信道传送的最大信息率（单位时间）；

（2） 若功率信噪比降为 5dB，要达到相同的最大信息传输率，信道带宽应为多少？

解：

（1） 根据已知条件有，

$$10\log_{10}\frac{P_S+P_N}{P_N}=10$$

因此$\frac{P_S}{P_N}=9$

$$\begin{aligned} C &= W\log\left(1+\frac{P_S}{P_N}\right) \\ &= 3\times10^3\log10 \\ &= 9.96\times10^3 \end{aligned}$$

即最大信息率为9.96×10^3比特/秒。

（2）如果功率信噪比降为 5dB，即$10\log_{10}\frac{P_S+P_N}{P_N}=5$，即$\frac{P_S+P_N}{P_N}=\sqrt{10}$，因此

$$
\begin{aligned}
C &= W\log\left(1+\frac{P_S}{P_N}\right) \\
&= W\log\sqrt{10} \\
&= \frac{1}{2}W\log 10
\end{aligned}
$$

为达到以前的信息传输率，因此带宽应为原来的 2 倍，即

$$W = 6\times10^{3}\,\mathrm{Hz}$$

第五章课后习题

【5.1】某信源按 $P(0)=\frac{3}{4}$， $P(1)=\frac{1}{4}$ 的概率产生统计独立的二元序列。

（1）试求 N_0，使当 $N>N_0$ 时有

$$P\left\{\left|\frac{I(a_i)}{N}-H(S)\right|\geq 0.05\right\}\leq 0.01$$

式中， $H(S)$ 是信源的熵。

（2）试求当 $N=N_0$ 时典型序列集 G_{eN} 中含有的信源序列个数。

解:

（1）该信源的信源熵为

$$H(S)=-\sum p(s_i)\log p(s_i)=0.811 \text{ 比特/符号}$$

自信息的方差为

$$\begin{aligned}D[I(s_i)] &= E[I^2(s_i)]-H^2(S)\\ &=\frac{3}{4}\log^2\frac{4}{3}+\frac{1}{4}\log^2 4-0.811^2\\ &=0.4715\end{aligned}$$

根据等长码编码定理，我们知道

$$P\left\{\left|\frac{I(a_i)}{N}-H(S)\right|\geq e\right\}\leq 1-d$$

根据给定条件可知， $e=0.05$， $d=0.99$。而

$$d=\frac{D[I(s_i)]}{Ne^2}$$

因此

$$N_0\geq\frac{D[I(s_i)]}{de^2}=\frac{0.4715}{0.05^2 * 0.99}=190.5$$

取 $N_0=191$。

（2） e 典型序列中信源序列个数取值范围为：

$$(1-d)2^{N[H(S)-e]} < \|G_{eN}\| < 2^{N[H(S)+e]}$$

代入上述数值得

$$0.01\times 2^{145.351} < \|G_{eN}\| < 2^{164.451}$$

【5.2】有一信源，它有六个可能的输出，其概率分布如下表所示，表中给出了对应的码 A、B、C、D、E 和 F。

表 5.2

消息	$P(a_i)$	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

（1） 求这些码中哪些是惟一可译码；

（2） 求哪些码是非延长码（即时码）；

（3） 求对所有惟一可译码求出其平均码长 $\overline{L}$ 。

解：

（1）上述码字中，A 为等长码，且为非奇异码，因此码 A 为惟一可译码；码 B 中，根据惟一可译码的判断方法，可求得其尾随后缀集合为 {1,11,111,1111,11111}，且其中任何后缀均不为码字，因此码 B 是惟一可译码。码 C 为逗点码，因此码 C 为惟一可译码；码 D 不是惟一可译码，因为其尾随后缀

集合中包含 0，而 0 又是码字；码 E 的尾随后缀集合为空集，因此码 E 是惟一可译码；码 F 不是惟一可译码，因为其尾随后缀集合中包含 0，而 0 又是码字，因此 F 不是惟一可译码。

（2）码 A、C、E 是即时码（非延长码）

（3）码 A 的平均码长为 3；码 B 的平均码长为 2.125；码 C 的平均码长为 2.125；码 F 的平均码长为 2。

【5.3】证明定理 5.6，若存在一个码长为 $l_1, l_2, \ldots, l_q$ 的惟一可译码，则一定存在具有相同码长的即时码。

证明：

如果存在码长为 $l_1, l_2, \ldots, l_q$ 的惟一可译码，则 $l_1, l_2, \ldots, l_q$ 必定满足如下不等式

$$\sum_{i=1}^{q} r^{-l_i} \le 1$$

而如果码长 $l_1, l_2, \ldots, l_q$ 满足上述不等式，根据 Kraft 不等式构造即时码的方法，可以构造出码长为 $l_1, l_2, \ldots, l_q$ 的即时码，具体构造过程略，参照课本相关定理。

【5.4】设信源

$$\begin{bmatrix} S \\ P(s) \end{bmatrix} = \begin{bmatrix} s_1 & s_2 & \cdots & s_6 \\ p_1 & p_2 & \cdots & p_6 \end{bmatrix}$$

将此信源编码为 r 元惟一可译变长码（即码符号集 $X = \{1, 2, \ldots, r\}$），其对应的码长为 $(l_1, l_2, \ldots, l_6) = (1,1,2,3,2,3)$，求 r 值的下限。

解：

如果要构造出惟一可译变长码，则相关码长必须满足 $\sum_{i=1}^{q} r^{-l_i} \le 1$，代入上式有

$$r^{-1} + r^{-2} + r^{-3} \le \frac{1}{2}$$

当 $r = 2$ 时，上述不等式不成立；当 $r = 3$ 时，成立。因此 r 值的下限为 3。

【5.5】若有一信源

$$\begin{bmatrix} S \\ P(s) \end{bmatrix} = \begin{bmatrix} s_1 & s_2 \\ 0.8 & 0.2 \end{bmatrix}$$

每秒钟发出 2.66 个信源符号。将此信源的输出符号送入某一个二元信道中进行传输（假设信道是无噪无损的），而信道每秒钟只传递两个二元符号。试问信源不通过编码能否直接与信道连接？若通过适当编码能否中在信道中进行无失真传输？若能连接，试说明如何编码并说明原因。

解：

如果不通过编码，即信道的两个码符号对应两个信源符号，而信道传输码符号的速度小于信源发出信源符号的速度，因此势必会造成信源符号的堆积，因此不通过编码是无法将信源与信道直接连接。

信源平均每秒发出的信息量为

$$2.66 * H(S) = -2.66 * \sum P(s) \log P(s) = 1.921 \text{ 比特/秒}$$

而该信道的信道容量为 1 比特/符号，平均每秒能够传输的最大信息量为 2 比特，因此通过编码可以实现二者的连接。

若要连接，需要对扩展信源的信源符号进行编码，目的是使送入信道的信息量小于信道每秒能接收的最大信息量（或使每秒钟编码后送入信道的码符号个数必须小于信道所能接受的最大码符号个数），具体编码方法将在第八章进行。

【5.6】设某无记忆二元信源，概率 $p_1 = P(1) = 0.1$，$p_0 = P(0) = 0.9$，采用下述游程编码方案：第一步，根据 0 的游程长度编成 8 个码字，第二步，将 8 个码字变换成二元变长码，如下表所示。

表 5.6

信源符号序列	中间码	二元码字
1	s_0	1000
01	s_1	1001
001	s_2	1010
0001	s_3	1011
00001	s_4	1100
000001	s_5	1101
0000001	s_6	1110
00000001	s_7	1111
00000000	s_8	0

（1） 试问最后的二元变长码是否是否是惟一可译码；

（2） 试求中间码对应的信源序列的平均长度$\overline{L}_1$；

（3） 试求中间码对应的二元变长码码字的平均长度$\overline{L}_2$；

（4） 计算比值$\overline{L}_2/\overline{L}_1$，解释它的意义，并计算这种游程编码的编码效率；

解:

（1）该码是非延长码，因此肯定是惟一可译码；

（2）由于信源本身是无记忆的，因此各信源符号的概率如下表所示。

信源符号序列	概率	中间码	二元码字
1	0.1	s_0	1000
01	0.09	s_1	1001
001	0.081	s_2	1010
0001	0.0729	s_3	1011
00001	0.06561	s_4	1100
000001	0.059049	s_5	1101
0000001	0.0531441	s_6	1110
00000001	0.04782969	s_7	1111
00000000	0.43046721	s_8	0

因此信源序列的平均长度为

$$\overline{L}_1 = \sum p(s_i) l_i = 5.695328$$

（3）中间码对应的二元变长码码长为

$$\overline{L}_2 = \sum p(s_i) l_i = 2.708598$$

（4）$\frac{\overline{L}_2}{\overline{L}_1}=0.4756$，反应了每个信源符号需要的二元码符号数。

平均每个信源符号的信息量为

$$H(S)=-0.9\log 0.9-0.1\log 0.1=0.469$$

编码效率为

$$h=\frac{H(S)}{\overline{L}_2/\overline{L}_1}=0.986$$

第六章课后习题

【6.1】设有一离散信道，其信道传递矩阵为

$$\begin{bmatrix} \frac{1}{2} & \frac{1}{3} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{6} & \frac{1}{2} \end{bmatrix}$$

并设 $P(x_1)=\frac{1}{2}$，$P(x_2)=P(x_3)=\frac{1}{4}$，试分别按最小错误概率准则与最大似然译码准则确定译码规则，并计算相应的平均错误概率。

解:

假设输入符号集和输出符号集分别为 $\{x_1,x_2,x_3\}$ 和 $\{y_1,y_2,y_3\}$。

按照最大似然译码规则，选择如下:

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

平均错误概率为:

$$\begin{aligned} P_E &= \sum_X \sum_{Y-x^* \text{对应的} y_j} P(x_i)P(y_j \mid x_i) \\ &= \frac{1}{2}\left(\frac{1}{3}+\frac{1}{6}\right)+\frac{1}{2}\left(\frac{1}{3}+\frac{1}{6}\right) \\ &= \frac{1}{2} \end{aligned}$$

如果需要按照最小错误概率译码，需要首先求出其联合概率矩阵:

$$\begin{bmatrix} \frac{1}{4} & \frac{1}{6} & \frac{1}{12} \\ \frac{1}{24} & \frac{1}{8} & \frac{1}{12} \\ \frac{1}{12} & \frac{1}{24} & \frac{1}{8} \end{bmatrix}$$

最小错误概率译码规则应如下：

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

此时错误概率为：

$$
\begin{aligned}
P_E \quad &= \sum_{X} \sum_{Y-x^*\text{对应的}y_j} P(x_i)P(y_j \mid x_i) \\
&= \sum_{X} \sum_{Y-x^*\text{对应的}y_j} P(x_i, y_j) \\
&= \frac{1}{12} + \frac{1}{24} + \frac{1}{8} + \frac{1}{12} + \frac{1}{12} + \frac{1}{24} \\
&= \frac{11}{24}
\end{aligned}
$$

【6.2】计算码长 $n=5$ 的二元重复码的译码错误概率。假设无记忆二元对称信道中正确传递概率为 $\overline{p}$，错误传递概率为 $p=1-\overline{p}$。此码能检测出多少错误？又能纠正多少错误。若 $p=0.01$，译码错误概率是多大？

解：

将 0 和 1 编成 00000 和 11111 后，当传输过程中产生 1 位至 4 位错误时均可检测出，但产生 5 位错误却无法检测出，同时，如果输入等概率分布，当传输过程中存在 1 位错误以及 2 位错误时，可以自动纠正。此时的错误概率为：

错 3 位的概率为：$C_5^3 p^3 \overline{p}^2$；

错 4 位的概率为：$C_5^4 p^4 \overline{p}$；

错 5 位的概率为：$C_5^5 p^5$

因此，译码错误概率为：

$$C_5^3 p^3 \overline{p}^2 + C_5^4 p^4 \overline{p} + C_5^5 p^5 = 1.02961 \times 10^{-5}$$

【6.3】设某二元码为 $C = \{11100, 01001, 10010, 00111\}$

（1） 计算此码的最小距离 $d_{\min}$；

（2） 计算此码的码率 R，假设码字等概率分布；

（3） 采用最小距离译码准则，试问接收序列 10000，01100 和 00100 应译成什么码字?

（4） 此码能纠正几位码元的错误?

解:

（1） 此码字的最小距离 $d_{\min} = 3$；

（2） 此码字的码率 $R = \dfrac{\log M}{n} = \dfrac{2}{5}$ 比特/码符号；

（3） 采用最小距离译码，10000 应译成 10010；01100 应译成 11100；00100 译成 11100、00111 均可；

（4） 由于 $d_{\min} = 3 = 2\times 1 + 1$，因此此码能纠正 1 位码元的错误。

【6.4】设无记忆二元对称信道的正确传递概率为 $\overline{p}$，错误传递概率为 $p = 1 - \overline{p} << \overline{p}$。令长度为 n 的 M 个二元码字 $a_i = (a_{i_1} a_{i_2} \mathbf{L}\, a_{i_n})$，其中 $a_{i_k} \in \{0,1\}$（码字为等概率分布），接收的二元序列为 $b_j = (b_{j_1} b_{j_2} \mathbf{L}\, b_{j_n})$，$b_{j_k} \in \{0,1\}$。试证明：采用最小距离译码准则可使平均译码错误概率 P_E 达到最小，并且

$$P_E = 1 - \frac{1}{M} \sum_i p^{D_{*j}} \overline{p}^{\,n - D_{*j}}$$

证明:

构造函数 $f(x) = p^x \overline{p}^{\,n-x}$，有

$$f'(x) = p^x \overline{p}^{\,n-x} \ln p - p^x \overline{p}^{\,n-x} \ln \overline{p} < 0$$

因此该函数为减函数，即当 $D_{*j} \le D_{ij}$ 时，有 $p^{D_{*j}} \overline{p}^{\,n-D_{*j}} \ge p^{D_{ij}} \overline{p}^{\,n-Dij}$。因此按照最小距离选择的译码规则 $F(b_j) = a^*$，使 $D_{*j} \le D_{ij}$ 成立，必然会有 $p^{D_{*j}} \overline{p}^{\,n-D_{*j}} \ge p^{D_{ij}} \overline{p}^{\,n-Dij}$，即

$$\overline{P}_E \ge \overline{P}_E'$$

其中 $\overline{P}_E$ 为最小距离译码得到的正确概率，而 $\overline{P}_E'$ 则是其他译码得到的正确概率，

进一步可以得到运用最小距离译码得到的错误概率为最小。

【6.5】对于离散无记忆强对称信道，信道矩阵为：

$$P=\begin{bmatrix} 1-p & \frac{p}{r-1} & \frac{p}{r-1} & \cdots & \frac{p}{r-1} \\ \frac{p}{r-1} & 1-p & \frac{p}{r-1} & \cdots & \frac{p}{r-1} \\ \vdots & \vdots & \vdots & & \vdots \\ \frac{p}{r-1} & \frac{p}{r-1} & \frac{p}{r-1} & \cdots & 1-p \end{bmatrix}$$

试证明对于此信道，最小距离译码准则等价于最大似然译码准则。

证明：

当信道发送符号序列为 $a_i=\left(a_{i_1}a_{i_2}\cdots a_{i_n}\right)$，接收符号序列为 $b_j=\left(b_{j_1}b_{j_2}\cdots b_{j_n}\right)$ 时，信道矩阵中的条件概率为：

$$\begin{aligned} P(b_j \mid a_i) &= P\left(b_{j_1}b_{j_2}\cdots b_{j_n} \mid a_{i_1}a_{i_2}\cdots a_{i_n}\right) \\ &= P\left(b_{j_1} \mid a_{i_1}\right)P\left(b_{j_2} \mid a_{i_2}\right)\cdots P\left(b_{j_n} \mid a_{i_n}\right) \\ &= \left(\frac{p}{r-1}\right)^{D(a_i,b_j)}\left(1-p\right)^{n-D(a_i,b_j)} \end{aligned}$$

按照最大似然译码规则，选择 $F(b_j)=a^*$，使 $P(b_j \mid a^*)>P(b_j \mid a_i)$ 成立，即

$$\left(\frac{p}{r-1}\right)^{D(a^*,b_j)}\left(1-p\right)^{n-D(a^*,b_j)}>\left(\frac{p}{r-1}\right)^{D(a_i,b_j)}\left(1-p\right)^{n-D(a_i,b_j)}$$

因此有

$$\left(\frac{p}{r-1}\right)^{D(a^*,b_j)-D(a_i,b_j)}>\left(1-p\right)^{D(a^*,b_j)-D(a_i,b_j)}$$

一般情况下，$\bar{p}=1-p>\frac{1}{2}$，因此有 $D(a^*,b_j)\le D(a_i,b_j)$ 成立，而这即为最小距离译码规则。

【6.6】某一信道，其输入 X 的符号集为 $\{0,\frac{1}{2},1\}$，输出 Y 的符号集为 $\{0,1\}$，信道矩阵为

$$P=\begin{bmatrix}1 & 0\\ \frac{1}{2} & \frac{1}{2}\\ 0 & 1\end{bmatrix}$$

现有四个消息的信源通过这信道传输（消息等概率出现）。若对信源进行编码，我们选这样一种码

$$C:\{(x_1,x_2,\frac{1}{2},\frac{1}{2})\}，\quad x_i\in 0或1\quad（i=1,2）$$

其码长为 $n=4$，并选取这样的译码规则

$$f(y_1,y_2,y_3,y_4)=(y_1,y_2,\frac{1}{2},\frac{1}{2})$$

(1) 这样编码后信道的信息传输率等于多少?

(2) 证明在选用的译码规则下，对所有码字 $P_E=0$。

解:

输入码字不同的个数共有 4 个，因此编码后信道的信息传输率为

$$R=\frac{\log 4}{4}=\frac{1}{2}\text{比特/码符号}$$

	$00\frac{1}{2}\frac{1}{2}$	$01\frac{1}{2}\frac{1}{2}$	$10\frac{1}{2}\frac{1}{2}$	$11\frac{1}{2}\frac{1}{2}$
0000	1/4	0	0	0
0001	1/4	0	0	0
0010	1/4	0	0	0
0011	1/4	0	0	0
0100	0	1/4	0	0
0101	0	1/4	0	0
0110	0	1/4	0	0
0111	0	1/4	0	0
1000	0	0	1/4	0
1001	0	0	1/4	0
1010	0	0	1/4	0
1011	0	0	1/4	0
1100	0	0	0	1/4
1101	0	0	0	1/4
1110	0	0	0	1/4
1111	0	0	0	1/4

可以看出，通过上述译码，对每一个输出都有一个输入与其对应，通过计算，可得其平均错误概率 $P_E = 0$ 。

【6.7】考虑一个码长为 4 的二元码，其码字为 $W_1 = 0000$ ，$W_2 = 0011$ ，$W_3 = 1100$ ，$W_4 = 1111$ 。假设码字送入一个二元对称信道（其单符号错误概率为 p ，且 $p < 0.01$），而码字输入是不等概率的，其概率为

$$P(W_1) = \frac{1}{2}，\ P(W_2) = \frac{1}{8}，\ P(W_3) = \frac{1}{8}，\ P(W_4) = \frac{1}{4}$$

试找出一种译码规则使平均错误概率 P_E 最小。

解:

设接收码字为 V_i ，则一共可能有 16 种不同的码字序列，而

$$P(V_j \mid W_i) = \overline{p}^{\,n-D(V_j,W_i)} p^{D(V_j,W_i)}$$

列出所有的输出，如下表所示。

接收码字 V_j	W_i				目标序列
	0000	0011	1100	1111	
	$\frac{1}{2}P(V_j\|W_1)$	$\frac{1}{8}P(V_j\|W_2)$	$\frac{1}{8}P(V_j\|W_3)$	$\frac{1}{4}P(V_j\|W_4)$	
0000	$\frac{1}{2}\bar{p}^4$	$\frac{1}{8}\bar{p}^2p^2$	$\frac{1}{8}\bar{p}^2p^2$	$\frac{1}{4}p^4$	0000
0001	$\frac{1}{2}\bar{p}^3p$	$\frac{1}{8}\bar{p}^3p$	$\frac{1}{8}\bar{p}p^3$	$\frac{1}{4}\bar{p}p^3$	0000
0010	$\frac{1}{2}\bar{p}^3p$	$\frac{1}{8}\bar{p}^3p$	$\frac{1}{8}\bar{p}p^3$	$\frac{1}{4}\bar{p}p^3$	0000
0011	$\frac{1}{2}\bar{p}^2p^2$	$\frac{1}{8}\bar{p}^4$	$\frac{1}{8}p^4$	$\frac{1}{4}\bar{p}^2p^2$	0011
0100	$\frac{1}{2}\bar{p}^3p$	$\frac{1}{8}\bar{p}p^3$	$\frac{1}{8}\bar{p}^3p$	$\frac{1}{4}\bar{p}p^3$	0000
0101	$\frac{1}{2}\bar{p}^2p^2$	$\frac{1}{8}\bar{p}^2p^2$	$\frac{1}{8}\bar{p}^2p^2$	$\frac{1}{4}\bar{p}^2p^2$	0000
0110	$\frac{1}{2}\bar{p}^2p^2$	$\frac{1}{8}\bar{p}^2p^2$	$\frac{1}{8}\bar{p}^2p^2$	$\frac{1}{4}\bar{p}^2p^2$	0000
0111	$\frac{1}{2}\bar{p}p^3$	$\frac{1}{8}\bar{p}^3p$	$\frac{1}{8}\bar{p}p^3$	$\frac{1}{4}\bar{p}^3p$	1111
1000	$\frac{1}{2}\bar{p}^3p$	$\frac{1}{8}\bar{p}p^3$	$\frac{1}{8}\bar{p}^3p$	$\frac{1}{4}\bar{p}p^3$	0000
1001	$\frac{1}{2}\bar{p}^2p^2$	$\frac{1}{8}\bar{p}^2p^2$	$\frac{1}{8}\bar{p}^2p^2$	$\frac{1}{4}\bar{p}^2p^2$	0000
1010	$\frac{1}{2}\bar{p}^2p^2$	$\frac{1}{8}\bar{p}^2p^2$	$\frac{1}{8}\bar{p}^2p^2$	$\frac{1}{4}\bar{p}^2p^2$	0000
1011	$\frac{1}{2}\bar{p}p^3$	$\frac{1}{8}\bar{p}^3p$	$\frac{1}{8}\bar{p}p^3$	$\frac{1}{4}\bar{p}^3p$	1111
1100	$\frac{1}{2}\bar{p}^2p^2$	$\frac{1}{8}p^4$	$\frac{1}{8}\bar{p}^4$	$\frac{1}{4}\bar{p}^2p^2$	1100
1101	$\frac{1}{2}\bar{p}p^3$	$\frac{1}{8}\bar{p}p^3$	$\frac{1}{8}\bar{p}^3p$	$\frac{1}{4}\bar{p}^3p$	1111
1110	$\frac{1}{2}\bar{p}p^3$	$\frac{1}{8}\bar{p}p^3$	$\frac{1}{8}\bar{p}^3p$	$\frac{1}{4}\bar{p}^3p$	1111
1111	$\frac{1}{2}p^4$	$\frac{1}{8}\bar{p}^2p^2$	$\frac{1}{8}\bar{p}^2p^2$	$\frac{1}{4}\bar{p}^4$	1111

【6.8】设一种离散无记忆信道，其信道矩阵为

$$P=\begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} \end{bmatrix}$$

（1） 计算信道容量 C；

（2） 找出一个码长为 2 的重复码，其信息传输率为 $\frac{1}{2}\log 5$（即 5 个码字）。如果按最大似然译码准则设计译码器，求译码器输出端的平均错误概率 P_E（输入码字等概率）。

（3） 有无可能存在一个码长为 2 的码，使 $P_e^{(i)}=0$（$i=1,2,3,4,5$），即 $P_E=0$，如存在的话请找出来。

解:

（1）观察该信道，其每一行数据都是第一行数据的置换，每一列数据都是第一列数据的置换，因此该信道是对称信道，其信道容量为:

$$C=\log r-H(\frac{1}{2},\frac{1}{2},0,0,0)=\log 5-1=1.322 \text{ 比特/码符号}$$

(2) 假设信道中的输入符号集和输出符号集为$\{0,1,2,3,4\}$，进行二次重复码，编得

00、11、22、33、44

其码率为 $R=\dfrac{H(S)}{n}=\dfrac{1}{2}\log 5$ 比特/码符号。

此时，输出方可能有 25 种可能性，进行最大似然译码，如下表所示。

错误概率为

$$P_E = \frac{1}{5}\left(\frac{1}{4}+\frac{1}{4}+\frac{1}{4}+\frac{1}{4}+\frac{1}{4}\right)=\frac{1}{4}$$

输出序列	输入序列					译码
	00	11	22	33	44	
00	1/4	0	0	0	1/4	00/44
01	1/4	0	0	0	0	00
02	0	0	0	0	0	均可
03	0	0	0	0	0	均可
04	0	0	0	0	1/4	44
10	1/4	0	0	0	0	00
11	1/4	1/4	0	0	0	00/11
12	0	1/4	0	0	0	11
13	0	0	0	0	0	均可
14	0	0	0	0	0	均可
20	0	0	0	0	0	均可
21	0	1/4	0	0	0	11
22	0	1/4	1/4	0	0	11/22
23	0	0	1/4	0	0	22
24	0	0	0	0	0	均可
30	0	0	0	0	0	均可
31	0	0	0	0	0	均可
32	0	0	1/4	0	0	22
33	0	0	1/4	1/4	0	22/33
34	0	0	0	1/4	0	33
40	0	0	0	0	1/4	44
41	0	0	0	0	0	均可
42	0	0	0	0	0	均可
43	0	0	0	1/4	0	33
44	0	0	0	1/4	1/4	33/44

（3）将 25 个长度为 2 的序列排成一方块图，如下所示：

00	01	02	03	04	00
10	11	12	13	14	10
20	21	22	23	24	20
30	31	32	33	34	30
40	41	42	43	44	40
00	01	02	03	04	00

为使平均错误概率为 0，须充分利用译码规则，假如选择了一个码字，则在

接收序列中与这个码字对应的条件概率不为零的肯定不能出现在所选的码组中，对应上图，即是如果一个码字被选定了，则以它为左上角的正方形中的四个码字均不能出现在这个码组中，因此可以选择{00，12，24，31，43}，可以保证平均错误概率为 0。

	00	01	02	03	04	10	11	12	13	14	20	21	22	23	24	30	31	32	33	34	40	41	42	43	44
00	1/4	1/4				1/4	1/4																		
01		1/4	1/4				1/4	1/4																	
02			1/4	1/4				1/4	1/4																
03				1/4	1/4				1/4	1/4															
04	1/4				1/4	1/4				1/4															
10						1/4	1/4				1/4	1/4													
11							1/4	1/4				1/4	1/4												
12								1/4	1/4				1/4	1/4											
13									1/4	1/4				1/4	1/4										
14						1/4				1/4	1/4				1/4										
20											1/4	1/4				1/4	1/4								
21												1/4	1/4				1/4	1/4							
22													1/4	1/4				1/4	1/4						
23														1/4	1/4				1/4	1/4					
24											1/4				1/4	1/4				1/4					
30																1/4	1/4				1/4	1/4			
31																	1/4	1/4				1/4	1/4		
32																		1/4	1/4				1/4	1/4	
33																			1/4	1/4				1/4	1/4
34																1/4				1/4	1/4				1/4
40	1/4	1/4																			1/4	1/4			
41		1/4	1/4																			1/4	1/4		
42			1/4	1/4																			1/4	1/4	
43				1/4	1/4																			1/4	1/4
44	1/4				1/4															1/4					1/4

【6.9】证明二元$(2n+1,1)$重复码当采用最大似然译码准则时，译码的平均错误概率为

$$P_E=\sum_{k=n+1}^{2n+1}C_{2n+1}^{k}p^{k}\bar{p}^{2n+1-k}$$

式中，p为二元对称信道的错误传输率，并计算当$n=5,7,9,11$时P_E的近似值。

证明:

对于二元$(2n+1,1)$重复码，采用最大似然译码规则时，收到的码字序列可能有2^{2n+1}个，此时可以自动纠正1位至n位错误，因此平均错误概率为:

$$P_E=C_{2n+1}^{n+1}p^{n+1}\bar{p}^{n}+\mathbf{L}+C_{2n+1}^{2n+1}p^{2n+1}=\sum_{k=n+1}^{2n+1}C_{2n+1}^{k}p^{k}\bar{p}^{2n+1-k}$$

当 $n=2$ 时，$P_E=1\times10^{-5}$；

当 $n=3$ 时，$P_E=3.42\times10^{-7}$；

当 $n=4$ 时，$P_E=1.22\times10^{-8}$；

当 $n=5$ 时，$P_E=4.43\times10^{-10}$。

【6.10】对二元 $(2n,1)$ 重复码，设计一种合适的译码规则，并求出它的译码平均错误概率 P_E。

解：此题在关键在于对于错了 n 位时的处理方法。

显然，对于错误小于 n 位时，我们可以自动纠正其错误，而对于错 n 位时，只能人为指定一个。因此错误概率为

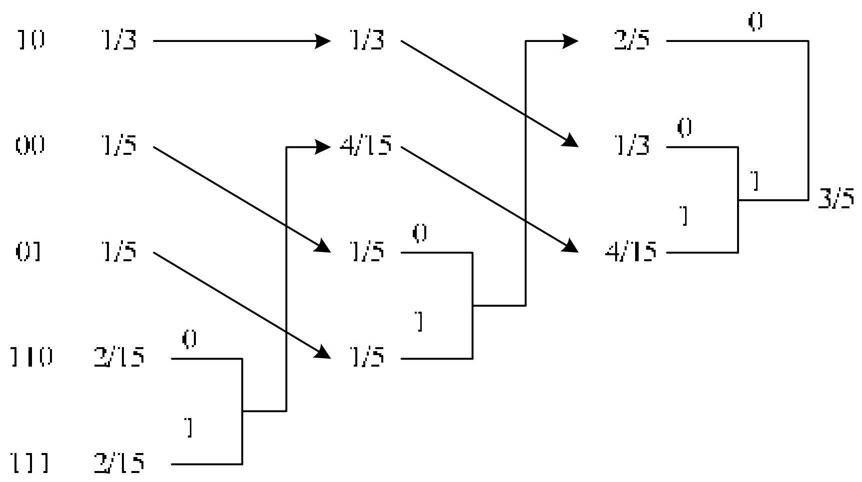
$$
\begin{aligned}
P_E &= \frac{1}{2}\left(\sum_{k=n+1}^{2n} C_{2n}^k p^k \overline{p}^{2n-k} + C_{2n}^n p^n \overline{p}^n + \sum_{k=n+1}^{2n} C_{2n}^k p^k \overline{p}^{2n-k}\right) \\
&= \sum_{k=n+1}^{2n} C_{2n}^k p^k \overline{p}^{2n-k} + \frac{1}{2} C_{2n}^n p^n \overline{p}^n
\end{aligned}
$$

第八章课后习题

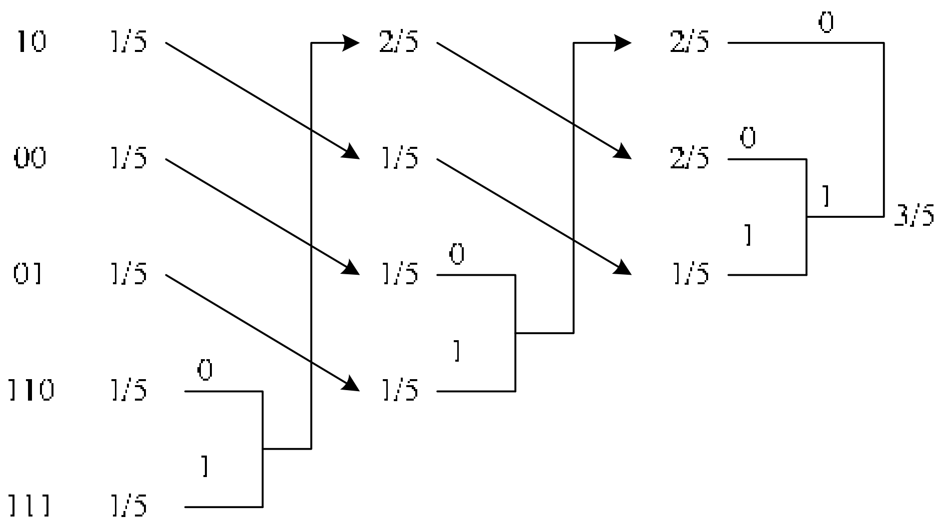
【8.1】求概率分布为(1/3,1/5,1/5,2/15,2/15)信源的二元霍夫曼码。讨论此码对于概率分布为(1/5,1/5,1/5,1/5,1/5)的信源也是最佳二元码。

解:

概率分布为(1/3,1/5,1/5,2/15,2/15)信源二元霍夫曼编码过程如下:



同样，对于概率分布为(1/5,1/5,1/5,1/5,1/5)的信源，编码过程如下:

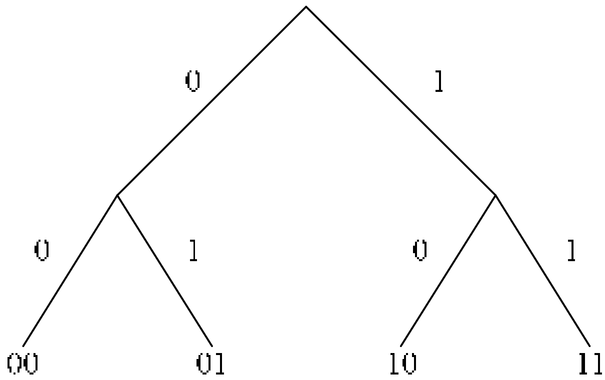


可见，二者的码字完全相同。

【8.2】设二元霍夫曼码为(00,01,10,11)和(0,10,110,111)，求出可以编得这样霍夫曼码的信源的所有概率分布。

解:

二元霍夫曼编码的过程必定是信源缩减的过程，编码为(00,01,10,11)的信源，其码树如下图所示。



假设四个信源符号的概率分别是 p_1, p_2, p_3, p_4，假设 $p_1 \ge p_2 \ge p_3 \ge p_4$，则必定有如下条件成立

$$p_1 \le p_3 + p_4$$

$$p_2 \le p_3 + p_4$$

又 $p_1 + p_2 + p_3 + p_4 = 1$，即 $p_3 + p_4 \ge \frac{1}{3}$，因此要构造上述编码，必定要满足

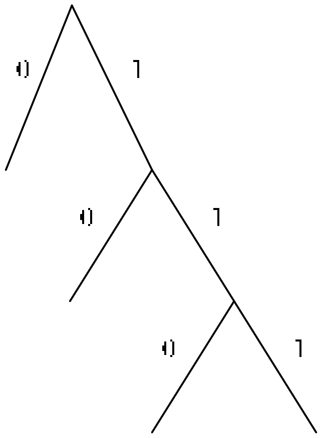
$$p_3 + p_4 \ge \frac{1}{3}$$

$$p_1 \le p_3 + p_4$$

$$p_2 \le p_3 + p_4$$

而编码为(0,10,110,111)的码树如下图所示:

解:



如果按上述情况进行编码，必定要满足 $p_2 \le p_1$， $p_3 + p_4 < p_1$，根据 $p_1 + p_2 + p_3 + p_4 = 1$ 可得， $p_1 > \frac{1}{3}$。因此完成上述编码的概率分布为:

$$p_1 > \frac{1}{3}$$

$$p_2 \le p_1$$

$$p_3 + p_4 < p_1$$

【8.3】设信源符号集

$$\begin{bmatrix} S \\ P(s) \end{bmatrix} = \begin{bmatrix} s_1 & s_2 \\ 0.1 & 0.9 \end{bmatrix}$$

(1) 求 $H(S)$ 和信源冗余度;

(2) 设码符号为 $X = \{0,1\}$，编出 S 的紧致码，并求 S 的紧致码的平均码长 $\overline{L}$;

(3) 把信源的 N 次无记忆扩展信源 S^N 编成紧致码，试求出 $N = 2,3,4,\infty$ 时的平均码长 $\frac{\overline{L}_N}{N}$;

(4) 计算上述 $N = 1,2,3,4$ 这四种码的编码效率和码冗余度。

解:

(1) 信源熵为

$$H(S) = -\sum P(x) \log P(x) = 0.469 \text{ 比特/符号}$$

因此得信源冗余度为

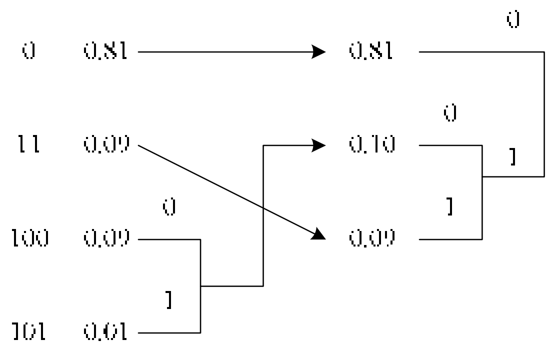
$$g = 1 - \frac{H(S)}{\log r} = 0.531$$

（2）对其进行紧致码编码，二个信源符号一个编码为 0，一个编码为 1，因此平均码长为 1 码符号/信源符号；

（3）对原码字进行二次扩展，其信源空间为：

$$\begin{bmatrix} S^2 \\ P(a_i) \end{bmatrix} = \begin{bmatrix} s_1s_1 & s_1s_2 & s_2s_1 & s_2s_2 \\ 0.01 & 0.09 & 0.09 & 0.81 \end{bmatrix}$$

进行 Huffman 编码，得码字如下：



平均码长为：

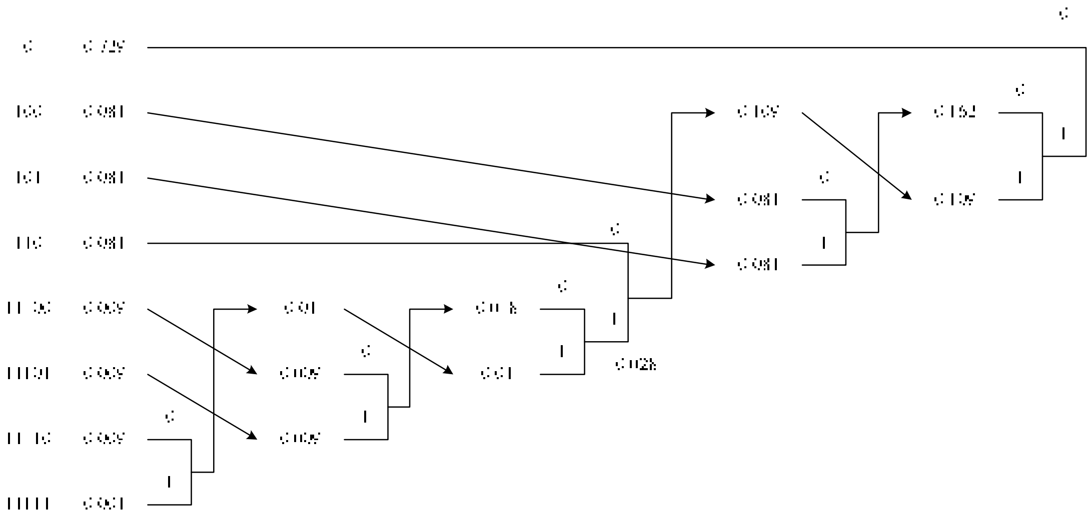
$$\overline{L}_2 = 0.81 + 0.18 + 0.27 + 0.03 = 1.29$$

$$\frac{\overline{L}_2}{N} = 0.645 \text{码符号/信源符号}$$

对原信源进行三次扩展，得扩展信源空间为

$$\begin{bmatrix} s_1s_1s_1 & s_1s_1s_2 & s_1s_2s_1 & s_1s_2s_2 & s_2s_1s_1 & s_2s_1s_2 & s_2s_2s_1 & s_2s_2s_2 \\ 0.729 & 0.081 & 0.081 & 0.1^2 0.9 & 0.081 & 0.1^2 0.9 & 0.1^2 0.9 & 0.1^3 \end{bmatrix}$$

进行 Huffman 编码，码字如下：



扩展信源的平均码长为：

$$\overline{L}_3 = 0.729 + 0.081*9 + 0.009*15 + 0.005 = 1.598$$

$$\frac{\overline{L}_3}{N} = 0.532667 \text{ 码符号/信源符号}$$

四次扩展信源略；

当 $N \to \infty$ 时，根据香农第一定理，平均码长为：

$$\frac{\overline{L}_N}{N} = \frac{H(S)}{\log r} = 0.469 \text{ 码符号/信源符号}$$

（5） 编码效率为：$h = \dfrac{H(S)}{\dfrac{\overline{L}_N}{N}\log r} = \dfrac{H(S)}{\overline{L}}$

因此有

不进行信源扩展时，编码效率为：$h = \dfrac{H(S)}{\overline{L}} = 0.469$

进行一次扩展时，编码效率为：$h = \dfrac{H(S)}{\overline{L}} = 0.727$

进行二次扩展时，编码效率为：$h = \dfrac{H(S)}{\overline{L}} = 0.880476$

当 $N \to \infty$ 时，编码效率趋向于 1。

因此，从本题结论可看出，对于变长紧致码，扩展信源的次数不需很大时就可以达到高效的无失真编码，这一点与等长码有很大的不同。

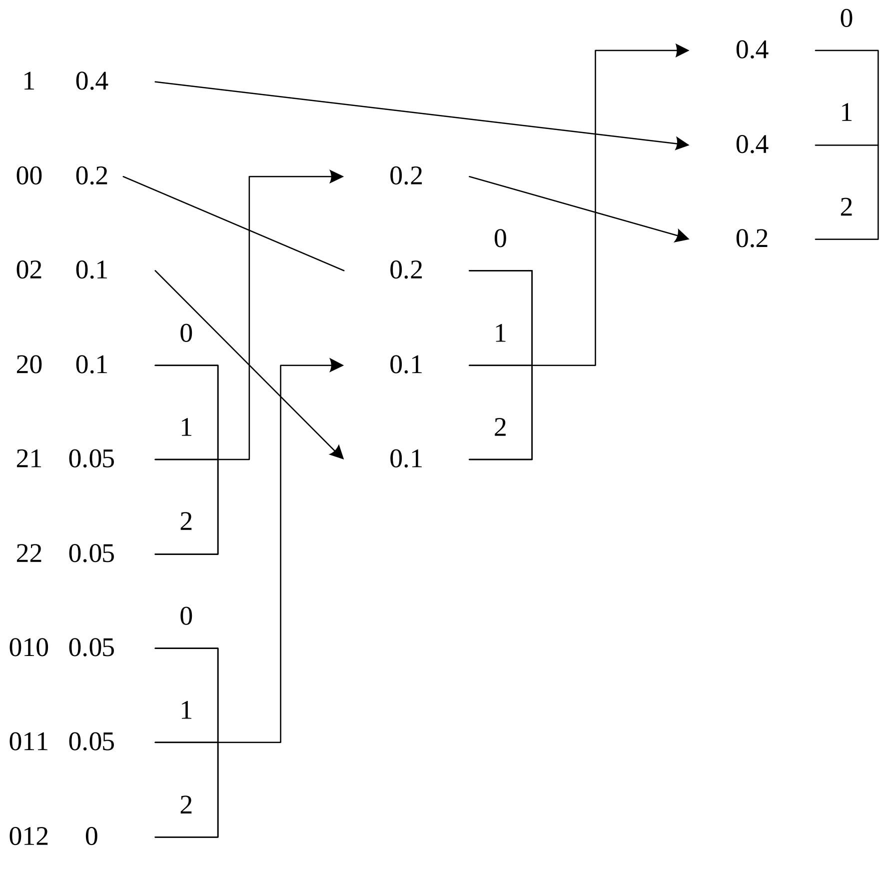
【8.4】信源空间为

$$\begin{bmatrix} S \\ P(s) \end{bmatrix} = \begin{bmatrix} s_1 & s_2 & s_3 & s_4 & s_5 & s_6 & s_7 & s_8 \\ 0.4 & 0.2 & 0.1 & 0.1 & 0.05 & 0.05 & 0.05 & 0.05 \end{bmatrix}$$

码符号为 $X=\{0,1,2\}$，试构造一种三元的紧致码。

解:

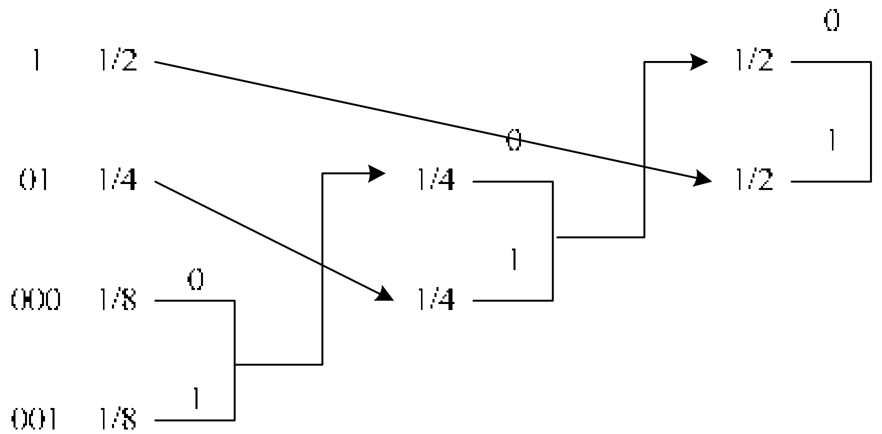
原信源有 8 个信源符号，为了有效利用短码，需对原信源进行扩展，添加 1 个概率为 0 的信源符号，使其满足 9=2*3+3 成立。编码过程如下:



【8.5】某气象员报告气象状态，有四种可能的消息：晴、去、雨和雾。若每个消息是等概率的，那么发送每个消息最少所需的二元脉冲数是多少？又若四个消息出现的概率分别为$\frac{1}{4}$、$\frac{1}{8}$、$\frac{1}{8}$和$\frac{1}{2}$，问在此情况下消息所需的二元脉冲数是多少？如何编码？

解:

平均每个消息携带的信息量为 2 比特，因此发送每个消息最少需要的二元脉冲数为 2。如果四个消息非等概率分布，采用紧致码编码，可使得所需要的二元脉冲数最少，编码过程如下：



平均码长为：

$$\overline{L}=\sum P(s_i)l_i=1.75\text{ 二元码符号/信源符号}$$

即在此情况下消息所需的二元脉冲数为 1.75 个。

【8.6】若某一信源有 N 个符号，并且每个符号等概率出现，对这信源用最佳霍夫曼码进行二元编码，问当 $N=2^i$ 和 $N=2^i+1$（i 是正整数）时，每个码字的长度等于多少？平均码长是多少？

解：

（1）当 $N=2^i$ 时，对每个符号编码所得的码字为等长码，其码长为 i，平均码长 $\overline{L}=i$；

（2）当 $N=2^i+1$ 时，每次进行编码时，总会剩余一个，因此，我们可以先将两个信源符号进行缩减，使得缩减后的信源符号个数恰好为 2^i 个，而这 2^i 个信源符号进行编码时是等长的，其码长为 i，其中包含原信源符号为 2^i-1 个，同时，另外两个信源符号的码长为 $i+1$，因此，原信源符号的平均码长为

$$\overline{L} = \frac{i(2^i - 1) + 2(i+1)}{2^i + 1} = \frac{i2^i + i + 2}{2^i + 1} = i + \frac{2}{2^i + 1}$$

【8.7】设信源

$$S: \begin{bmatrix} s_1 & s_2 & \cdots & s_{M-1} & s_M \\ p_1 & p_2 & \cdots & p_{M-1} & p_M \end{bmatrix}$$

并满足$\sum_{i=1}^{M} p_i = 1$和$p_1 \geq p_2 \geq \cdots \geq p_M$。试证明对于信源$S$一定存在这样的二元最佳码。其码字$W_{M-1}$和$W_M$具有相同的长度，且$W_{M-1}$和$W_M$只有最后一位码符号不同（分别为 1 和 0）。

解:

请参照课本 264 页最佳码三个性质的证明过程，此处略。

【8.8】设码C是均匀概率分布$P = (1/n, 1/n, \ldots, 1/n)$信源的二元霍夫曼码。并设码$C$中各码字的码长为$l_i$，又设$n = a2^k$，而$1 \leq a \leq 2$。

（1） 证明对于此等概率信源的所有即时码中，码C的总码长$T = \sum_i l_i$最短；

（2） 证明码C满足$\sum_i r^{-l_i} = 1$；

（3） 证明码C中，所有i满足$l_i = L$或$l_i = L - 1$，其中$L = \max\{l_i\}$；

（4） 设u是码长为$L-1$的码字个数，v是码长为L的码字个数，根据a, k确定u、v和L。

证明:

（1）根据二元霍夫曼编码的性质，我们知道，它一定是最优即时码，即在所有即时码中，它的平均码长最短。设C'是其他码字，一定有$\overline{L}_{C'} > L_C$，而平均码长$\overline{L}_C = \frac{T}{n}$，$\overline{L}_{C'} = \frac{T'}{n}$，可得

$$T \leq T'$$

（2）当$a = 1$时，所有码字的码长相等，均为$l_i = k$；

当 $a=2$ 时，所有码字的码长相等，均为 $l_i=k+1$；

而当 $1<a<2$ 时，信源符号个数是位于 2^k 至 2^{k+1} 之间的某正整数，反应在码树上，应是一些长度为 k 的节点上生出分枝，作为长度为 $k+1$ 的码字（第（3）小题中所说的码长要么是 $L-1=k$，要么是 $L=k+1$）。设码长为 k 的节点数为 x，则码长为 $k+1$ 的节点的分枝数一定为 $\left(2^k-x\right)\times 2$，而码字的总个数为：

$$x+\left(2^k-x\right)\times 2=2^{k+1}-x=a2^k$$

因此有

$$\begin{aligned}\sum_i r^{-l_i} &= \sum_i 2^{-l_i} \\ &= x2^{-k}+\left(2^k-x\right)\times 2\times 2^{-k-1} \\ &= x2^{-k}+\left(2^k-x\right)\times 2^{-k} \\ &= 1\end{aligned}$$

（3）已在第（2）小题中加以说明，此处略。

（4）根据第（2）小题中的分析可知 $L=k+1$，且有

$$u+v=a2^k$$

$$v=2^{k+1}-2u$$

解得

$$u=2^{k+1}-a2^k$$

$$v=2a2^k-2^{k+1}=(a-1)2^{k+1}$$

编码后的平均码长为

$$\begin{aligned}\overline{L} &= \frac{1}{a2^k}\left[k\left(2^{k+1}-a2^k\right)+(k+1)(a-1)2^{k+1}\right] \\ &= \frac{1}{a}\left[ka+2a-2\right] \\ &= k+2-\frac{2}{a}\end{aligned}$$

【8.9】现有一幅已离散量化后的图像，图像的灰度量化分成 8 级，见下表。表中数字为相应像素上的灰度级。

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

另有一无损无噪二元信道，单位时间（秒）内传输 100 个二元符号。

（1） 现将图像通过给定的信道传输，不考虑图像的任何统计特性，并采用二元等长码，问需要多长时间才能传完这幅图像?

（2） 若考虑图像的统计特性（不考虑图像的像素之间的依赖性），求此图像的信源熵 $H(S)$，并对灰度级进行霍夫曼最佳二元编码，问平均每个像素需用多少二元码符号来表示？这时需多少时间才能传送完这幅图像?

（3） 从理论上简要说明这幅图像还可以压缩，而且平均每个像素所需的二元码符号数可以小于 $H(S)$ 比特。

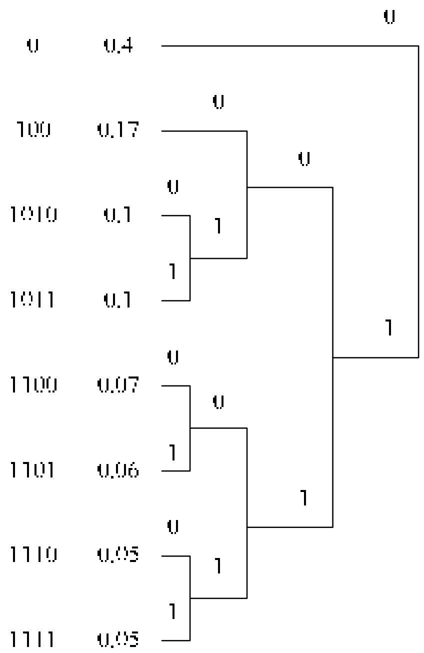
解:

（1）采用二元等长码，不考虑信源符号的统计特性，平均每个灰度需要 3 位二进制表示，在 10*10 的图像上，共需 300 位二进制表示，以每秒传输 100 位计算，共需 3 秒钟传输。

（2）统计图像中各灰度级的出现次数:

1	2	3	4	5	6	7	8
40	17	10	10	7	6	5	5

如果考虑信源符号的统计特性，对上述灰度级进行编码，如下图所示。



得如下码字:

1	2	3	4	5	6	7	8
40	17	10	10	7	6	5	5
0	100	1010	1011	1100	1101	1110	1111
1	3	4	4	4	4	4	4

平均码长为:

$$\overline{L}=0.4+0.51+0.43*4=2.63$$

在 10*10 的图像上，共需 263 位二进制表示，以每秒传输 100 位计算，共需 2.63 秒钟传输完。

（3）在计算（2）小题时，并未考虑像素灰度级之间的依赖关系，例如，灰度 1 后只能跟灰度 1 和 2，并未有其他灰度值。这样得到的信源熵 $H_\infty < H(S)$，而根据香农第一定理，当对信源进行扩展时，其平均码长 $\overline{L}$ 逼近于 $H_\infty < H(S)$，因此，该图像可以进一步压缩，平均每个像素所需要的二元码符号数 $\overline{L} \to H_\infty < H(S)$。

【8.10】有一个含有 8 个消息的无记忆信源，其概率各自为 0.2, 0.15, 0.15, 0.1, 0.1, 0.1, 0.1, 0.1。试编成两种三元非延长码，使它们的平均码长相同，但具有不同的

码长的方差，并计算平均码长和方差，说明哪一种码更实用些。

解：

进行三元编码，需增补一个概率为 0 的信源符号，两种编码方法如下所示。

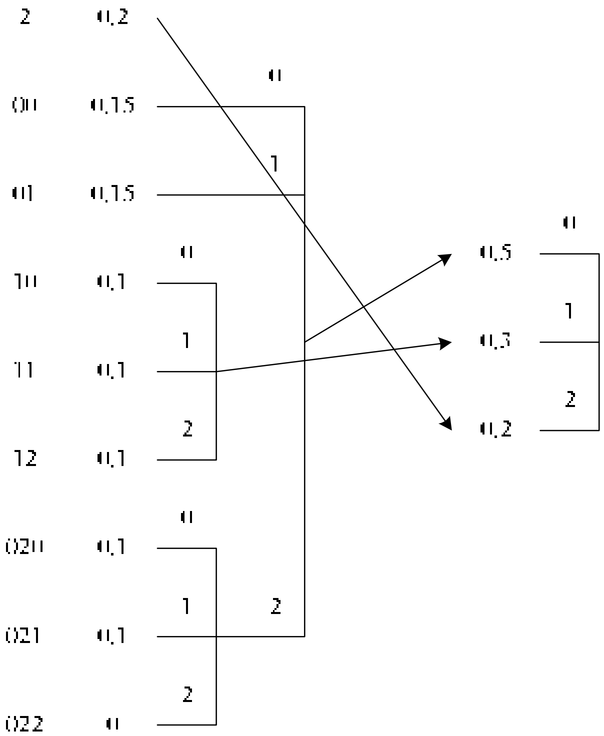


图 1

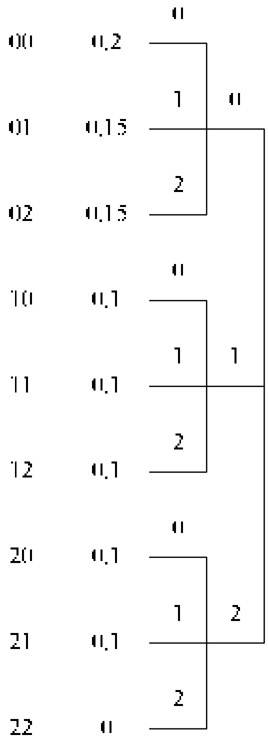


图 2

平均码长为：

$$\overline{L} = 0.2 + 0.3 + 0.3 + 0.6 + 0.6 = 2$$

编码 1 的码方差为：

$$s_1^2 = E[(l_i - \overline{L})^2] = 0.2 + 0.1 + 0.1 = 0.4$$

$$s_2^2 = E[(l_i - \overline{L})^2] = 0$$

虽然两种编码的平均码长相同，但由于编码 2 的码方差较小，因此编码 2 更实用一些。

【8.11】设有两个信源 X 和 Y 如下：

$$\begin{bmatrix} X \\ P(x) \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 \\ 0.2 & 0.19 & 0.18 & 0.17 & 0.15 & 0.1 & 0.01 \end{bmatrix}$$

$$\begin{bmatrix} Y \\ P(y) \end{bmatrix} = \begin{bmatrix} y_1 & y_2 & y_3 & y_4 & y_5 & y_6 & y_7 & y_8 & y_9 \\ 0.49 & 0.14 & 0.14 & 0.07 & 0.07 & 0.04 & 0.02 & 0.02 & 0.01 \end{bmatrix}$$

（1） 分别用霍夫曼码编成二元变长惟一可译码，并计算其编码效率；

（2） 分别用香农编码法编成二元变长惟一可译码，并计算编码效率；

（3） 分别用费诺编码方法编成二元变长惟一可译码，并计算编码效率；

（4） 从 X 、Y 两种不同信源来比较这三种编码方法的优缺点。

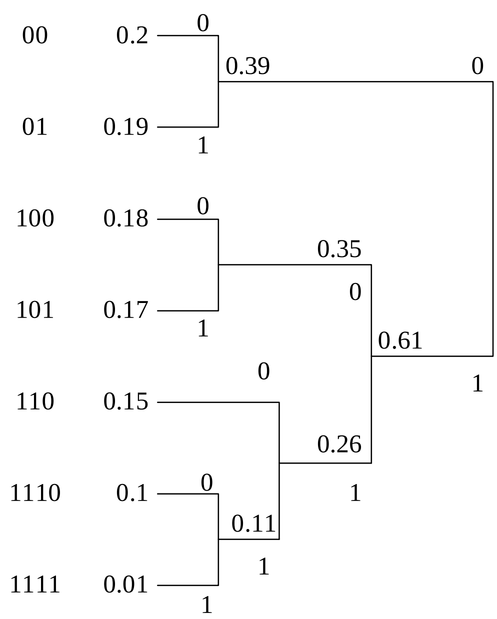
解:

第一个信源:

信源熵为:

$$H(X) = -\sum P(x)\log P(x) = 2.60868 \text{ 比特}$$

对其进行二元霍夫曼编码，如下所示:



平均码长为:

$$\overline{L} = 0.39*2 + 0.55*3 + 0.11*4 = 2.72 \text{ 码符号/信源符号}$$

编码效率为:

$$h=\frac{H(X)}{\overline{L}}=0.95907$$

对其进行费诺编码，如下所示:

<table>
<tr><td>00</td><td>0.2</td><td rowspan="3">0</td><td>0</td><td></td><td></td></tr>
<tr><td>010</td><td>0.19</td><td rowspan="2">1</td><td>0</td><td></td></tr>
<tr><td>011</td><td>0.18</td><td>1</td><td></td></tr>
<tr><td>10</td><td>0.17</td><td rowspan="4">1</td><td>0</td><td></td><td></td></tr>
<tr><td>110</td><td>0.15</td><td rowspan="3">1</td><td>0</td><td></td></tr>
<tr><td>1110</td><td>0.1</td><td rowspan="2">1</td><td>0</td></tr>
<tr><td>1111</td><td>0.01</td><td>1</td></tr>
</table>

平均码长为:

$$\overline{L}=0.4+0.37*3+0.34+0.45+0.44=2.74\text{码符号/信源符号}$$

编码效率为:

$$h=\frac{H(X)}{\overline{L}}=0.95207$$

对其进行香农编码，如下所示。

概率	码长	累积概率分布	二进制小数	码字
0.2	3	0	0.0	000
0.19	3	0.2	0.001100110011	001
0.18	3	0.39	0.011000111101	011
0.17	3	0.57	0.100100011110	100
0.15	3	0.74	0.101111010111	101
0.1	4	0.89	0.111000111101	1110
0.01	7	0.99	0.111111010111	1111111

其平均码长为:

$$\overline{L}=0.89*3+0.4+0.07=3.14\text{码符号/信源符号}$$

编码效率为:

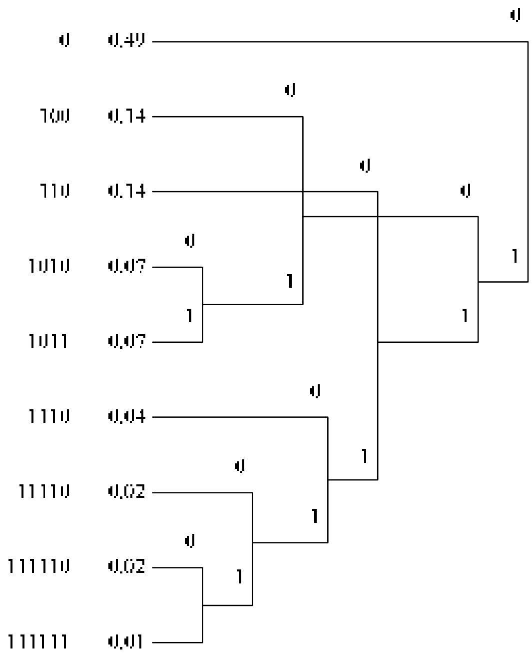
$$h=\frac{H(X)}{\overline{L}}=0.83079$$

第二个信源:

信源熵为:

$$H(Y)=-\sum P(y)\log P(y)=2.31356\text{比特}$$

对其进行二元霍夫曼编码，如下所示。



平均码长为：

$$\overline{L}=0.49+0.28*3+0.18*4+0.02*5+0.03*6=2.33\text{码符号/信源符号}$$

编码效率为：

$$h=\frac{H(Y)}{\overline{L}}=0.99294$$

对其进行费诺编码，如下所示。

<table>
<tr><td>0</td><td>0.49</td><td>0</td></tr>
<tr><td>100</td><td>0.14</td><td rowspan="8">1</td><td rowspan="2">0</td><td>0</td></tr>
<tr><td>101</td><td>0.14</td><td>1</td></tr>
<tr><td>1100</td><td>0.07</td><td rowspan="6">1</td><td rowspan="2">0</td><td>0</td></tr>
<tr><td>1101</td><td>0.07</td><td>1</td></tr>
<tr><td>1110</td><td>0.04</td><td rowspan="4">1</td><td>0</td></tr>
<tr><td>11110</td><td>0.02</td><td rowspan="3">1</td><td>0</td></tr>
<tr><td>111110</td><td>0.02</td><td rowspan="2">1</td><td>0</td></tr>
<tr><td>111111</td><td>0.01</td><td>1</td></tr>
</table>

平均码长为：

$$\overline{L}=0.49+0.28*3+0.18*4+0.02*5+0.03*6=2.33\text{码符号/信源符号}$$

编码效率为：

$$h = \frac{H(Y)}{\overline{L}} = 0.99294$$

对其进行香农编码，如下所示：

概率	码长	累积概率分布	二进制小数	码字
0.49	2	0	0.0	00
0.14	3	0.49	0.011111010111	011
0.14	3	0.63	0.101000010100	101
0.07	4	0.77	0.110001010001	1100
0.07	4	0.84	0.110101110000	1101
0.04	5	0.91	0.111010001111	11101
0.02	6	0.95	0.111100110011	111100
0.02	6	0.97	0.111110000101	111110
0.01	7	0.99	0.111111010111	1111110

平均码长为：

$$\overline{L} = 0.49*2 + 0.28*3 + 0.14*4 + 0.04*5 + 0.04*6 + 0.01*7 = 2.89$$

编码效率为：

$$h = \frac{H(Y)}{\overline{L}} = 0.80054$$

从上述编码中可以看出，霍夫曼编码的平均码长最短，编码效率最高，香农编码的平均码长最长，其编码效率最低，费诺编码居中。

【8.12】信源符号集 $A = \{1, 2, \mathbf{K}, q\}$，其概率分布为 $p_1, p_2, \mathbf{K}, p_q$。采用以下方法对信源符号进行编码。首先将概率分布按大小顺序排列 $p_1 \ge p_2 \ge \mathbf{K} \ge p_q$，定义累积分布函数

$$F_i \stackrel{\Delta}{=} \sum_{k=1}^{i-1} p_k$$

F_i 是所有小于 i 符号的概率和。于是符号 i 的码字是取 F_i 的二进制数的小数 l_i 位，若有尾数就进位到第 l_i 位，其中 $l_i = \left\lceil \log \frac{1}{p_i} \right\rceil$。

（1） 证明这样编得的码是即时码；

（2） 证明：$H(S) \le \overline{L} < H(S)+1$；

（3） 对于概率分布为(0.5,0.25,0.125,0.125)的信源进行编码，求其各码字；

（4） （3）中所得的码是否与此信源的霍夫曼码正巧完全一致？试说明这种完全一致的一般原理。

解：

（1）按照所取的累积分布函数，我们得知

$$C - F(s_i) \ge 0$$

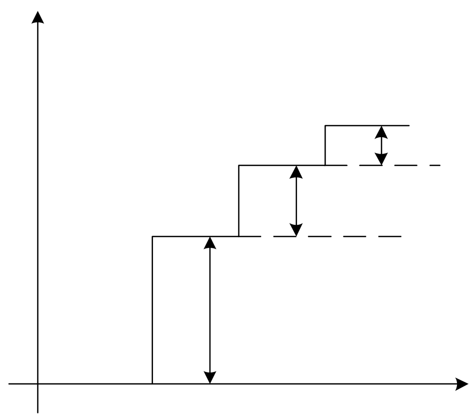
且二者的差必然小于第 l_i 上的权值，因此有

$$0 \le C - F(s_i) < 2^{-l_i}$$

而 $l_i = \left\lceil \log \frac{1}{p_i} \right\rceil$，得 $\log \frac{1}{p_i} \le l_i < \log \frac{1}{p_i} + 1$，即 $2^{-l_i} \le P(s_i)$，可得

$$F(s_i) \le C < F(s_i) + 2^{-l_i}$$

画出累积分布函数的图像，如下图所示。



可以发现，最后所得的码字取值区间并无重叠，根据二进制小数的特性，可得不重叠的区间其二进制小数的前缀部分是不同的，所以，这样得到的码一定满足变长码的前缀条件，因此是即时码。

（2）根据 $l_i = \left\lceil \log \frac{1}{p_i} \right\rceil$，得 $\log \frac{1}{p_i} \le l_i < \log \frac{1}{p_i} + 1$，因此有

$$p_i \log\frac{1}{p_i} \le p_i l_i < p_i \log\frac{1}{p_i} + p_i$$

两边进行统计平均得

$$H(S) \le \overline{L} < H(S)+1$$

（3）对概率分布为(0.5,0.25,0.125,0.125)的信源进行编码，过程如下：

码字	概率	码长	累积分布	二进制小数
0	0.5	1	0	0.0
10	0.25	2	0.5	0.1
110	0.125	3	0.75	0.11
111	0.125	3	0.875	0.111

（4）该码字与此信源的霍夫曼编码完全一致，即它本身构造成了该信源的紧致码，其原因是其码长恰好为 $\log\frac{1}{P(s_i)}$，信源的先验概率恰好为 2 的幂次方分布，构成了香农第一定理下限成立的条件，因此必然可以构造出紧致码。

【8.14】已知二元信源 $\{0,1\}$，其 $p_0=\frac{1}{8}$，$p_1=\frac{7}{8}$，试对其进行算术编码，并计算此序列的平均码长

11111110111110

解：

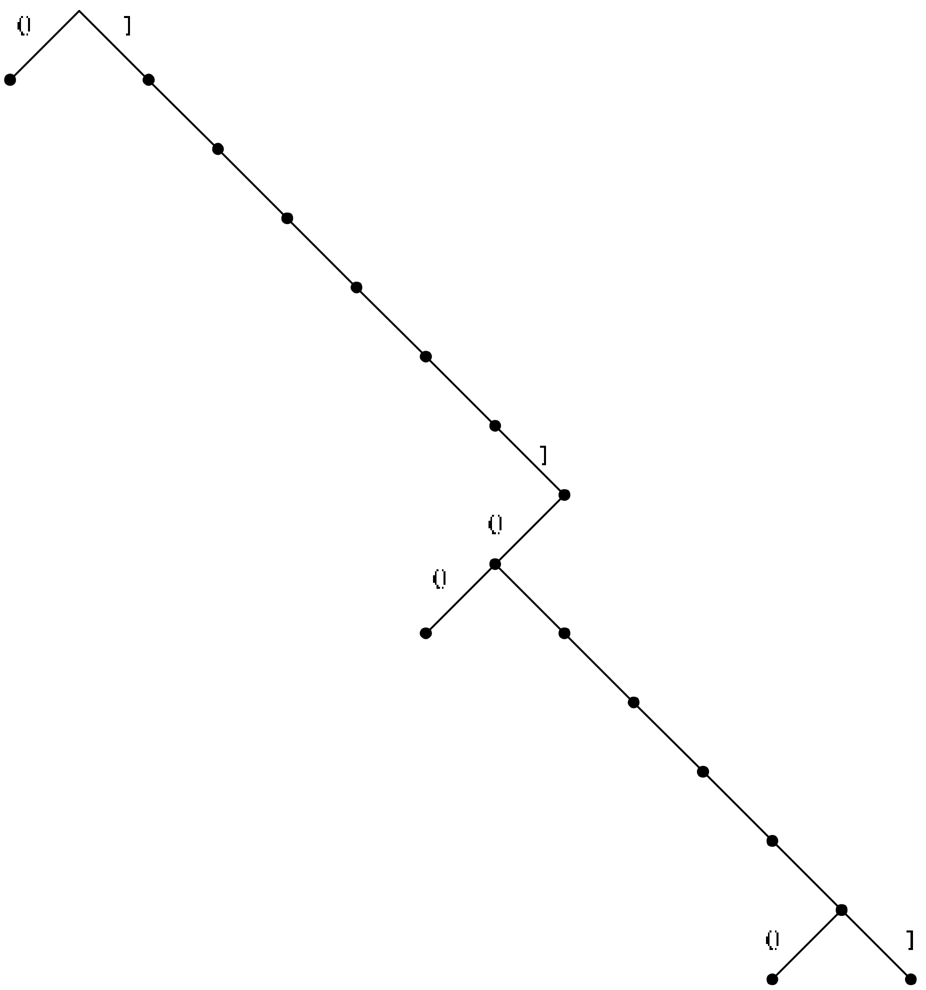
序列的发生概率为：

$$P(s=11111110111110)=p_1^{12}p_0^2=\left(\frac{7}{8}\right)^{12}\left(\frac{1}{8}\right)^2$$

在编码时应对该序列的码长值为：

$$l_i=\left\lceil \log\frac{1}{P(s)} \right\rceil=\left\lceil 12*\log\frac{8}{7}+2*3 \right\rceil=\lceil 8.311 \rceil=9$$

码字在码树上的相对位置如下图所示：



因此可得码序列的累积分布函数为:

$$
\begin{aligned}
F(S) &= 1 - P(11111111) - P(1111111011111) \\
&= 1 - \left(\frac{7}{8}\right)^{8} - \left(\frac{7}{8}\right)^{12} \times \frac{1}{8} \\
&= 0.631213929
\end{aligned}
$$

将累积分布函数变为二进制小数，得 0.101000011001，取小数点后 9 位，并进行进位处理，得上述序列的编码为 101000100。

平均码长为

$$\bar{L} = \frac{9}{14} = 0.6429 \text{码符号/信源符号}$$

编码效率为:

$$h = \frac{H(S)}{\bar{L}} = \frac{0.54356}{0.6429} = 84.555\%$$

【8.15】对输入数据流 000010110011100001001101111 分别用 LZ-77 算法、LZ-78 算法，LZW 算法、K-Y 算法进行编码，并计算各种方法的压缩率。

解:

采用 LZ-77 编码，所得的编码序列为:

(0,0,0)

(1,3,1)

(0,0,1)

(2,2,1)

(6,3,1)

(5,3,0)

(13,3,0)

(9,3,1)

(14,3,eof)

采用 LZ-78 编码，其分段顺序为: 0, 00,01,011,001,1,10,000,100,11,0111,11

字典的建立过程如下:

字典编号	存储条目	发送序列
1	0	(0,0)
2	00	(1,0)
3	01	(1,1)
4	011	(3,1)
5	001	(2,1)
6	1	(0,1)
7	10	(7,0)
8	000	(2,0)
9	100	(7,0)
10	11	(6,1)
11	0111	(4,1)
		(10,eof)

采用 LZW 编码，过程如下：

基本字典	1 2	0 1	
	3	00	1
	4	000	3
	5	01	1
	6	10	2
	7	011	5
	8	100	6
	9	0111	7
	10	1000	8
	11	001	3
	12	1001	8
	13	11	2
	14	101	6
	15	111	13
		11	13,eof

K-Y 算法：

令 D0=0，D1=1，先读入 4 个，0000，前后相等，因此暂时编码 D_2D_2

其中 $D_2=D_0D_0$

继续输入至 D_2D_2 $D_1D_0D_1D_0$，出现重复情况，

令 $D_1D_0=D_3=10$，则该序列变为：D_2D_2 D_3D_3，继续输入，得

$$D_2D_2\ D_3D_3D_0D_1D_1D_1D_2D_2$$

令 $D_4=D_2D_2=0000$，上序列变为：

$$D_4D_3D_3D_0D_1D_1D_1D_4D_1D_0D_0D_1$$

令 $D_5=D_0D_1$，上序列变为：

$$D_4D_3D_3D_5D_1D_1D_4D_1D_0D_5D1$$

令 $D_6=D_5D_1$ 上序列变为：

$$D_4D_3D_3D_6D_1D_4D_1D_0D_6D_0D_1D_1D_1D_1$$

令 $D_7=D_1D_1=11$，上序列变为

$$D_4D_3D_3D_6D_1D_4D_1D_0D_6D_0D_7D_7$$

其中 $D_0=0$，$D_1=1$，$D_2=00$，$D_3=10$，$D_4=0000$，$D_5=01$，D6=011，$D_7=11$

将上述序列以及字典发送至接收方即可。