第二章作业题

1．一副充分洗乱了的牌（含 52 张牌），试问

(1) 任一特定排列所给出的信息量是多少?

(2) 若从中抽取 13 张牌，所给出的点数都不相同能得到多少信息量?

解： (1) 52 张牌共有 52！种排列方式，假设每种排列方式出现是等概率的，则任一特定排列所给出的信息量是：

$$p(x_i)=\frac{1}{52!};\quad I(x_i)=-\log p(x_i)=\log 52!=225.581\ bit$$

(2) 52 张牌共有 4 种花色、13 种点数，抽取 13 张点数不同的牌的概率如下：

$$p(x_i)=\frac{4^{13}}{C_{52}^{13}}$$

$$I(x_i)=-\log p(x_i)=-\log\frac{4^{13}}{C_{52}^{13}}=13.208\ bit$$

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（11/66）=1/36 该事件的信息量为： I= log 36 ≈ 5.17 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 P5/36，因此该事件的信息量为：I= log≈ 2.85 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=1/18 因此该事件的信息量为：I= log18≈ 4.17 比特

3．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%，如果你问一位男士："你是否是色盲？"他的回答可能是"是"，可能是"否"，问这两个回答中各含多少信息量，平均每个回答中含有多少信息量？如果问一位女士，则答案中含有的平均自信息量是多少？

解： 男士：

$$p(x_Y)=7\%$$

$$I(x_Y)=-\log p(x_Y)=-\log 0.07=3.837\ bit$$

$$p(x_N)=93\%$$

$$I(x_N)=-\log p(x_N)=-\log 0.93=0.105\ bit$$

$$\begin{aligned}H(X)&=-\sum_{i}^{2}p(x_i)\log p(x_i)\\&=-(0.07\log 0.07+0.93\log 0.93)\\&=0.366\ bit/symbol\end{aligned}$$

女士：

$$\begin{aligned}H(X)&=-\sum_{i}^{2}p(x_i)\log p(x_i)\\&=-(0.005\log 0.005+0.995\log 0.995)\\&=0.045\ \ bit/symbol\end{aligned}$$

4. 居住某地区的女孩子有 25%是大学生，在女大学生中有 75%是身高 160 厘米以上的，而女孩子中身高 160 厘米以上的占总数的一半。假如我们得知“身高 160 厘米以上的某女孩是大学生”的消息，问获得多少信息量？

解：

设随机变量 X 代表女孩子学历

X	x_1（是大学生）	x_2（不是大学生）
P(X)	0.25	0.75

设随机变量 Y 代表女孩子身高

Y	y_1（身高>160cm）	y_2（身高<160cm）
P(Y)	0.5	0.5

已知：在女大学生中有 75%是身高 160 厘米以上的

即：$p(y_1/x_1)=0.75$

求：身高 160 厘米以上的某女孩是大学生的信息量

$$I(x_1/y_1)=-\log p(x_1/y_1)$$

即：
$$\begin{aligned}&=-\log\frac{p(x_1)p(y_1/x_1)}{p(y_1)}\\&=-\log\frac{0.25\times 0.75}{0.5}=1.415\ \ bit\end{aligned}$$

5. 设有一个信源，它产生 0，1 序列的信息。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 H_∞；

(3) 试计算 $H(X^4)$并写出 X^4 信源中可能有的所有符号。

解：

(1) 这个信源是平稳无记忆信源。因为有这些词语：“它在任意时间而且不论以前发生过什么符号……”

(2) $H(X^2)=2H(X)=-2\times(0.4\log 0.4+0.6\log 0.6)=1.942\ \ bit/symbol$

$$\begin{aligned}H_\infty&=\lim_{N->\infty}H(X_N/X_1X_2...X_{N-1})\\&=H(X_N)=0.971\ \ bit/symbol\end{aligned}$$

$$H(X_3/X_1X_2)=H(X_3)=-\sum_i p(x_i)\log p(x_i)$$
$$=-(0.4\log 0.4+0.6\log 0.6)$$
$$=0.971\ \ bit/symbol$$

(3)

$$H(X^4)=4H(X)$$
$$=-4\times(0.4\log 0.4+0.6\log 0.6)$$
$$=3.884\ \ bit/symbol$$

X^4中所有可能符号：

0000	0001	0010	0011
0100	0101	0110	0111
1000	1001	1010	1011
1100	1101	1110	1111

6. 根据信息论信源熵性质，对于一个离散平稳无记忆信源，证明 $H(\mathbf{X})=H(X^N)=H(X_1X_2\cdots X_N)=NH(X)$

7. 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 个象素和 8 个不同亮度电平，并设每秒要传送 30 帧图像，所有象素是独立变化的，且所有亮度电平等概率出现。(1)求传递此图像所需的信息率（比特/秒）。(2) 设某彩色电视系统，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，试计算传输该彩色系统的信息率为传输黑白系统的信息率的多少倍？

问答案

电视机的每个像素亮度作为信源，则信源空间为：

$$\begin{bmatrix} X \\ P(X) \end{bmatrix}=\begin{bmatrix} a_1 & a_2 & \cdots & a_8 \\ 1/8 & 1/8 & \cdots & 1/8 \end{bmatrix} \text{其中}\sum_{i=1}^{8} p(a_i)=1$$

每个像素亮度含有的信息量为

$$H(X)=\log 8=3bit/\ pixel$$

每帧图像数就是每一个离散亮度信源的无记忆N次扩展信源

因此每帧图像含有的信息量为

$$H(X^N)=NH(X)=5\times10^5\times\log 8=1.5\times10^6 bit/\ frame$$

电视机每秒30帧图像，因此每秒传递的信息率为

$$R=30\,frame/\ s\times1.5\times10^6 bit/\ frame=4.5\times10^7 bit/\ s$$

（2）问答案

H(X) = log8*30=log240bit/pixel

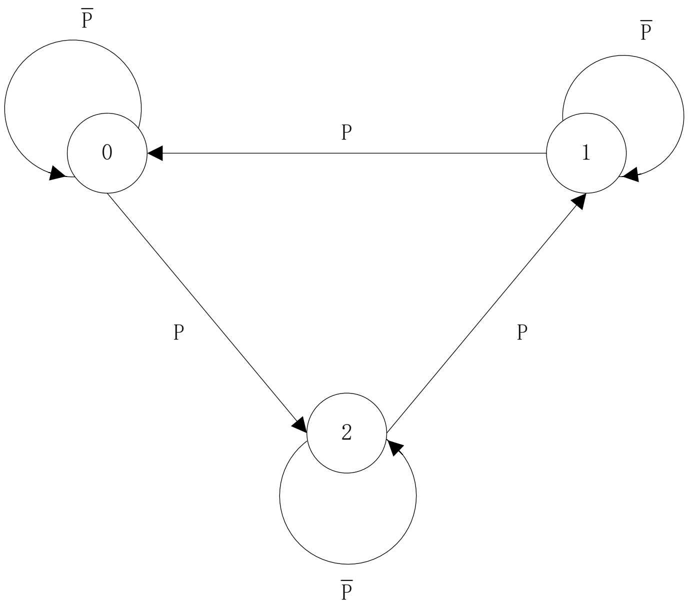
因此传输该彩色系统的信息率为传输黑白系统的信息率的：log(240)/log(8) = 2.636 倍

8.设有 12 个体积、颜色均相同的小球，其中一个与其它球不同（或者轻或者重），

现采用一个无砝码的天平来测量，为了在天平上称出哪一个球与其它球不同，且判断其重量是比其它球轻还是重，请问至少必须称多少次？

第一问：从信息论来看，12 个球一个重量异常，出现概率 1/12；该球质量可能轻也可能重，那么出现概率为 1/2。
那么要得到结果所需信息量为（log2(2)+log2(12)）bit。
称一次可能有轻、重、相等三种结果，可以得到信息量为 log2(3)bit。
log2(24)/log2(3)<3，因此三次应该能称出来。
第二问：这样的话，可以直接给出 120 个球问题的称量次数为 log240/log3<5,
5 次应该得到。

9. 一阶马尔可夫信源的状态图如下图所示。信源 X 的符号集为{0, 1, 2}。
(1) 求平稳后信源的概率分布；
(2) 求信源的熵 H_∞。



解：(1) 设信源的稳态分布概率为W_0,W_1,W_2，
由题意可知信源的状态转移矩阵 P 为:

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

则有:

$$\begin{cases} W_0\bar{p}+W_1 p \quad\quad\quad = W_0 \\ \quad\quad\quad W_1\bar{p}+W_2 p = W_1 \\ W_0 p \quad\quad\quad + W_2\bar{p} = W_2 \\ W_0+W_1+W_2=1 \\ p+\bar{p}=1 \end{cases}$$

解得: $W_0=W_1=W_2=\dfrac{1}{3}$

(2) 根据题中给出的状态转移图，可判断该马尔可夫信源具有不可约性和非周期性，因此，该马尔可夫信源具有遍历性，存在极限熵。

信源在 $S_i(i=0,1,2)$ 状态下信源的平均信息量:

$$H(X/S_0)=-(p\log p+\bar{p}\log\bar{p})$$
$$H(X/S_1)=-(p\log p+\bar{p}\log\bar{p})$$
$$H(X/S_2)=-(p\log p+\bar{p}\log\bar{p})$$

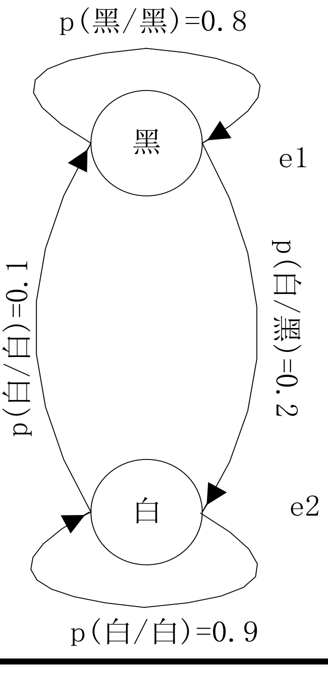
$$\begin{aligned} H_\infty &= \sum_{i=0}^{2} W_i H(X/S_i) \\ &= -3\cdot\frac{1}{3}(p\log p+\bar{p}\log\bar{p}) \\ &= -(p\log p+\bar{p}\log\bar{p}) \ \ bit/symbol \end{aligned}$$

10. 黑白气象传真图的消息只有黑色和白色两种，即信源 *X*={黑，白}。设黑色出现的概率为 *P(黑)* = 0.3，白色出现的概率为 *P(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(白/白)* = 0.9，*P(黑/白)* = 0.1，*P(白/黑)* = 0.2，*P(黑/黑)* = 0.8，求此一阶马尔可夫信源的熵 $H_2(X)$;

(3) 分别求上述两种信源的剩余度，比较 *H(X)*和 $H_2(X)$的大小，并说明其物理含义。



解：(1) 黑白消息出现前后没有关联时的信源熵:

$$
\begin{aligned}
H(X) &= -\sum_i p(x_i)\log p(x_i) \\
&= -(0.3\log 0.3 + 0.7\log 0.7) \\
&= 0.881 \ \ bit/symbol
\end{aligned}
$$

(2) 根据题中给出的状态转移图，可判断该马尔可夫信源具有不可约性和非周期性，因此，该马尔可夫信源具有遍历性，存在极限熵。设信源的稳态分布概率 W_1, W_2 分别对应黑和白两种状态，由题意可知信源的状态转移矩阵 P 为:

$$P = \begin{bmatrix} 0.8 & 0.2 \\ 0.1 & 0.9 \end{bmatrix}$$

则有:

$$
\begin{cases}
0.8W_1 + 0.1W_2 = W_1 \\
0.2W_1 + 0.9W_2 = W_2 \\
W_1 + W_2 = 1
\end{cases}
$$

解得

$$W_1 = \frac{1}{3}, \quad W_2 = \frac{2}{3}$$

(2) 信源在 $S_i (i = 1, 2)$ 状态下信源的平均信息量:

$$
\begin{aligned}
H(X/S_1) &= -(0.8\log 0.8 + 0.2\log 0.2) \\
&= 0.7219 \ \ bit/symbol \\
H(X/S_2) &= -(0.9\log 0.9 + 0.1\log 0.1) \\
&= 0.4690 \ \ bit/symbol
\end{aligned}
$$

$$
\begin{aligned}
H_2(X) &= \sum_{i=1}^{2} W_i H(X/S_i) \\
&= 0.5533 \ \ bit/symbol
\end{aligned}
$$

(3)

$$\eta_1 = \frac{H_0 - H(X)}{H_0} = \frac{\log 2 - 0.881}{\log 2} = 11.9\%$$

$$\eta_2 = \frac{H_0 - H_2(X)}{H_0} = \frac{\log 2 - 0.553}{\log 2} = 44.7\%$$

$H(X) > H_2(X)$

物理含义：无记忆信源的不确定度大于有记忆信源的不确定度，有记忆信源的结构化信息较多，能够进行较大程度的压缩。