

2.1 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“两骰子面朝上点数之和为 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？

P19 自信息量： $I(a_k) = \log_r \frac{1}{p(a_k)} = -\log_r p(a_k)$ ，其中 $p(a_k)$ 为事件发生的先验概率。

解：同时扔一对均匀的骰子，可能呈现的状态数有 36 种，各面呈现的概率为 1/6，所以 36 种中任何一种状态出现的概率都是相等，为 1/36。

(1) 设“两骰子面朝上点数之和为 2”为事件 A。在 36 种情况中，只有 1 种情况，即 1+1。

则： $P(A) = \frac{1}{36}$ ， $I(A) = -\log P(A) = \log_2 36 \approx 5.17 \text{ bit}$

(2) 设“两骰子面朝上点数之和为 8”为事件 B。在 36 种情况中，有 5 种情况，即 5+3，3+5，2+6，6+2，4+4。则： $P(B) = \frac{5}{36}$ ， $I(B) = -\log P(B) = \log_2 \frac{36}{5} \approx 2.85 \text{ bit}$

(3) 设“两骰子面朝上点数是 3 和 4”为事件 C。在 36 种情况中，有两种情况，即 3+4 和 4+3。则： $P(C) = \frac{2}{36}$ ， $I(C) = -\log P(C) = \log_2 18 \approx 4.17 \text{ bit}$

2.2 同时掷两个均匀的骰子，也就是各面呈现的概率都是 1/6，求：

(1) 事件“3 和 5 同时出现”的自信息量； $P(A) = \frac{2}{36}$ ， $I(A) \approx 4.17 \text{ bit}$

(2) 事件“两个 1 同时出现”的自信息量； $P(B) = \frac{1}{36}$ ， $I(B) \approx 5.17 \text{ bit}$

(3) 两个点数之和(即 2, 3, ..., 12 构成的子集)的熵；

(4) 事件“两个骰子点数中至少有一个是 1”的自信息量。 $P(C) = \frac{11}{36}$ ， $I(C) \approx 1.71 \text{ bit}$

P21 信息熵： $H_r(X) = E\left(\log_r \frac{1}{p(a_i)}\right) = -\sum_{i=1}^q p(a_i) \log_r p(a_i)$ ，自信息的数学期望，每个符号的平均自信息量。

(3) 列出两个点数之和的概率空间：

$$\begin{aligned} \begin{bmatrix} Z \\ p(z) \end{bmatrix} &= \begin{bmatrix} 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \\ 1/36 & 2/36 & 3/36 & 4/36 & 5/36 & 6/36 & 5/36 & 4/36 & 3/36 & 2/36 & 1/36 \end{bmatrix} \\ H(Z) &= -\sum_Z P(z) \log P(z) \approx 3.27 \text{ bit/符号} \end{aligned}$$

2.3 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(202 120 130 213 001 203 210 110 321 010 021 032 011 223 210)，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：(1) 因为离散信源是无记忆的，所以它发出的消息序列中各个符号是无依赖的，统计独立的。因此，此消息的自信息就等于各个符号的自信息之和。则可得：

$$\begin{aligned} I(a_1 = 0) &= -\log P(a_1) = -\log \frac{3}{8} = \log_2 \frac{8}{3} \approx 1.45 \text{ bit} \\ I(a_2 = 1) &= -\log P(a_2) = -\log \frac{1}{4} = \log_2 4 = 2 \text{ bit} \\ I(a_3 = 2) &= -\log P(a_3) = -\log \frac{1}{4} = \log_2 4 = 2 \text{ bit} \\ I(a_4 = 3) &= -\log P(a_4) = -\log \frac{1}{8} = \log_2 8 = 3 \text{ bit} \end{aligned}$$

此消息中共有 14 个符号“0”，13 个符号“1”，12 个符号“2”和 6 个符号“3”，则此消息的自信息是： $I = 14I(a_1 = 0) + 13I(a_2 = 1) + 12I(a_3 = 2) + 6I(a_4 = 3) = 87.81 \text{ bit}$

(2) 此消息中共有 45 个信源符号，携带了 87.81 比特信息量，因此，此消息中平均每个符号携带的信息量为： $H(X) = \frac{87.81}{45} \approx 1.95 \text{ bit/符号}$

2.4 设有一个二元信源 $\begin{bmatrix} X \\ p(x) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0.9 & 0.1 \end{bmatrix}$, 计算该信源的熵。

$$H(X) = -0.9 \times \log_2 0.9 - 0.1 \times \log_2 0.1 \approx 0.469 \text{ bit/符号} \approx 0.47 \text{ bit/符号}$$

2.5 设信源 $\begin{bmatrix} X \\ p(x) \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & a_3 & a_4 & a_5 & a_6 \\ 0.2 & 0.19 & 0.18 & 0.17 & 0.16 & 0.17 \end{bmatrix}$, 求该信源的熵, 并解释为什么在本题中 $H(X) > \log 6$, 不满足信源熵的极值性。

$$\begin{aligned} H(X) &= - \sum_x P(x) \log P(x) \\ &= -0.2 \log 0.2 - 0.19 \log 0.19 - 0.18 \log 0.18 - 0.17 \log 0.17 - 0.16 \log 0.16 - 0.17 \log 0.17 \\ &\approx 2.657 \text{ bit/符号} \approx 2.66 \text{ bit/符号} \end{aligned}$$

P25 信源熵的极值性是指离散无记忆信源输出 q 个不同的信息符号, 当且仅当各个符号出现的概率相等时, 信息熵最大, 最大值为 $\log q = \log 6 = 2.58 \text{ bit/符号}$ 。现在 $H(X) > \log 6$, 不满足信源熵的极值性, 这是因为, 我们讨论的信源的概率空间应该是一个完备集, 即 $\sum_{i=1}^6 p(a_i) = 1$, 而在本题当中, $\sum_{i=1}^6 p(a_i) = 1.07 \neq 1$, 不是完备集, 所以不满足信源熵的极值性。

2.6 每帧电视图像可以认为是由 3×10^5 个像素组成, 每个像素均是独立变化, 若每个像素可取 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? 若有一广播员在约 1w 个汉字的字汇中选 1k 个字来口述此电视图像, 试问广播员描述此图像所广播的信息量是多少(假设汉字字汇是等概率分布, 并彼此无依赖)? 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: (1) 亮度电平等概出现, 即每个亮度电平含有的平均信息量为:

$$H(X) = \log_2 128 = 7 \text{ bit/符号}$$

一帧图像每个像素均是独立变化的, 则每帧图像信源就是离散亮度信源的无记忆 N 次扩展信源, 得到每帧图像含有的信息量为:

$$H(X^N) = NH(X) = 3 \times 10^5 \times H(X) = 2.1 \times 10^6 \text{ bit/帧}$$

参考 P21 例 2.2.4: $P(A) = (\frac{1}{128})^{3 \times 10^5}$, $H(A) = -\log(\frac{1}{128})^{3 \times 10^5} = 2.1 \times 10^6 \text{ bit/帧}$

(2) 每个汉字含有的信息量为:

$$H(Y) = \log_2 10000 \approx 13.29 \text{ bit/字}$$

广播员口述电视图像是从此汉字字汇信源中独立的选取 1000 个字来描述的, 所以广播员描述此帧图像所广播的信息量为:

$$H(Y^N) = NH(Y) = 1000 \log_2 10^4 \approx 1.329 \times 10^4 \text{ bit/千字}$$

参考 P21 例 2.2.5: $P(B) = (\frac{1}{10000})^{1000}$, $H(B) = -\log(\frac{1}{10000})^{1000} = 1.329 \times 10^4 \text{ bit/千字}$

(3) 若广播员仍从此汉字字汇信源 Y 中独立选取汉字来描述电视图像, 每次口述一次汉字含有的信息量是 $H(Y)$, 每帧电视图像含有的信息量是 $H(X^N)$, 则广播员口述次图像至少需要使用的汉字数为:

$$\frac{H(X^N)}{H(Y)} \approx \frac{2.1 \times 10^6}{13.29} \approx 1.58 \times 10^5 \text{ 字}$$

2.7 为了传输一个由字母 A、B、C、D 组成的符号集, 把每个字母编码成两个二进制脉冲序列, 以 00 代表 A, 01 代表 B, 10 代表 C, 11 代表 D。每个二进制脉冲宽度为 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 $\{1/5, 1/4, 1/4, 3/10\}$ ，试计算传输的平均信息速率？

解：(1) 不同字母等概率出现时，符号集的概率空间为：

$$\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{bmatrix} A & B & C & D \\ 1/4 & 1/4 & 1/4 & 1/4 \end{bmatrix}$$

平均每个字母含有的信息量为： $H(X) = \log_2 4 = 2 \text{ bit/符号}$

现在用两个二元码脉冲代表一个字母，每个二元码脉冲宽度为 $\tau = 5\text{ms}$ ，则每个字母占用 $t = 2\tau = 10\text{ms}$ 。一秒内可以传输的字母个数为： $n = \frac{1}{t} = 100 \text{ 字母/秒}$

则传输的平均信息率为： $R = nH(X) = 200 \text{ bit/s}$

(2) 字母出现概率不同时，每个字母含有的平均自信息量为：

$$H(X) = - \sum_{i=1}^4 P(a_i) \log P(a_i) = \frac{1}{5} \log 5 + \frac{1}{4} \log 4 + \frac{1}{4} \log 4 + \frac{3}{10} \log \frac{10}{3} \approx 1.985 \text{ bit/符号}$$

其传输的平均信息率为： $R = nH(X) \approx 1.985 \times 10^2 \text{ bit/s}$

2.8 试问四进制、八进制脉冲所含的信息量是二进制脉冲的多少倍？

解：二进制脉冲当中，共可以表示两种不同信息。假设两种不同信息等概率分布，则每个二进

制脉冲的信息量为： $I_2 = -\log_2 p(x) = -\log_2 \frac{1}{2} = 1 \text{ bit/脉冲}$

同理，四进制脉冲中： $I_4 = -\log_2 p(x) = -\log_2 \frac{1}{4} = 2 \text{ bit/脉冲}$

八进制脉冲中： $I_8 = -\log_2 p(x) = -\log_2 \frac{1}{8} = 3 \text{ bit/脉冲}$

所以，四进制脉冲所含的信息量是二进制脉冲的 2 倍，八进制脉冲所含的信息量是二进制脉冲的 3 倍。

2.9 国际摩尔斯电码用点和划的序列发送英文字母，“划”用持续三个单位的电流脉冲表示，

“点”用持续一个单位的电流脉冲表示。其中“划”出现的概率是“点”出现概率的 1/3。

计算：

(1) 点和划的信息量；

(2) 电码信源的平均信息量。区别于 2.3(2)

解：(1) 根据题意： $\frac{1}{3} p(x = \text{点}) = p(x = \text{划})$ ， $p(x = \text{点}) + p(x = \text{划}) = 1$

可得： $p(x = \text{点}) = \frac{3}{4}$ ， $p(x = \text{划}) = \frac{1}{4}$

“点”含有的信息量为： $I(\text{点}) = -\log_2 p(x) = -\log_2 \frac{3}{4} = 0.415 \text{ bit}$

“划”含有的信息量为： $I(\text{划}) = -\log_2 p(x) = -\log_2 \frac{1}{4} = 2 \text{ bit}$

(2) 电码信源的平均信息量为：

$$H(X) = - \sum_{i=1}^2 p(x_i) \log_2 p(x_i) = -\frac{1}{4} \log_2 \frac{1}{4} - \frac{3}{4} \log_2 \frac{3}{4} = 0.811 \text{ bit/符号}$$

错误： $H(X) = \frac{I(\text{划}) + I(\text{点})}{4} = \frac{2.415}{4} = 0.6038 \text{ bit/符号}$

平均信息量是有可能事件的信息量的加权平均，权重是这些事件发生的概率。

$$H = P(\text{点}) I(\text{点}) + P(\text{划}) I(\text{划}) = \frac{3}{4} I(\text{点}) + \frac{1}{4} I(\text{划})$$

2.10 某一记忆信源的符号集为 $\{0, 1\}$ ，已知 $p(0)=1/4$ ， $p(1)=3/4$ 。

(1) 求信源的熵。

(2) 由 100 个符号构成的序列，求某一特定的序列(例如有 m 个 0 和 $100-m$ 个 1)的自信

息量的表达式。

(3) 计算(2)中的序列的熵。

解: (1) 信源熵为: $H(X) = -\sum_{i=1}^2 p(x_i) \log_2 p(x_i) = -\frac{1}{4} \log_2 \frac{1}{4} - \frac{3}{4} \log_2 \frac{3}{4} = 0.811 \text{ bit/符号}$

(2) $I(x=0) = -\log_2 p(x=0) = -\log_2 \frac{1}{4} = 2 \text{ bit}$

$I(x=1) = -\log_2 p(x=1) = -\log_2 \frac{3}{4} = 0.415 \text{ bit}$

自信息量表达式为: $I = m \times I(x=0) + (100-m) \times I(x=1) = 41.5 + 1.585m \text{ bit}$

(3) 因为是无记忆信源, 所以序列中 100 个符号相互独立, 为 100 次扩展信源, 其熵为:

$$H(X^{100}) = 100H(X) = 81.1 \text{ bit/符号序列}$$

错误: $H(X) = \frac{I}{100} = \frac{41.5+1.585m}{100}$

P21 信息熵是随机变量不确定性的度量, 它描述的是随机变量所有可能取值的信息量的加权平均。一旦我们知道了 m , 我们就知道了序列中 0 和 1 的确切数量, 这个序列就不再是随机的了。极端一点, 如果 $m=0$, 那么这个序列就是一个全 1 序列, 这个消息是完全确定的, 不再是随机变量, 这样的信源不可能提供任何信息量。 m 不确定, 那么该序列就会有 2^{100} 个可能的符号序列, 是二元无记忆信源的 100 次扩展。

$$H(X) = -m \times \frac{1}{4} \log_2 \frac{1}{4} - (100-m) \times \frac{3}{4} \log_2 \frac{3}{4}$$

2.16 黑白气象传真图的消息只有黑色和白色两种, 即信源 $X=\{\text{黑}, \text{白}\}$, 设黑色的出概率为 $P(\text{黑})=0.3$, 白色的出现概率为 $P(\text{白})=0.7$ 。

(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 $P(\text{白}/\text{白})=0.9$, $P(\text{黑}/\text{白})=0.1$, $P(\text{白}/\text{黑})=0.2$, $P(\text{黑}/\text{黑})=0.8$, 求此平稳离散信源的熵 $H_2(X)$ 。

(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, 并说明其物理意义。

解: (1) 假设黑白气象传真图上黑白消息出现的前后没有关联, 则等效于一个离散无记忆信源, 信源的信息熵为:

$$H(X) = -\sum_{i=1}^2 p(a_i) \log p(a_i) = -0.7 \log 0.7 - 0.3 \log 0.3 \approx 0.881 \text{ bit/符号}$$

(2) 假设黑白气象传真图的消息前后有关联, 从给出的依赖关系可得联合熵为:

$$\text{符号序列 } X_1X_2 \text{ 的联合熵: } H(X_1X_2) = -\sum_{i=1}^q \sum_{j=1}^q p(a_i a_j) \log p(a_i a_j)$$

二维平稳信源的平均符号熵: $H_2(X) = \frac{1}{2} H(X_1X_2)$

联合信源的概率分布: $p(a_i a_j) (i, j = 1, 2, \dots, q)$

条件概率公式: $p(a_j | a_i) = \frac{p(a_i a_j)}{p(a_i)} (i, j = 1, 2, \dots, q)$

条件熵:

$$H(X_2/X_1) = -\sum_{i=1}^q \sum_{j=1}^q p(a_i) p(a_j/a_i) \log p(a_j/a_i) = -\sum_{i=1}^q \sum_{j=1}^q p(a_i a_j) \log p(a_j/a_i)$$

先求信源的联合概率分布:

$$P(\text{白白}) = P(\text{白}/\text{白}) \times P(\text{白}) = 0.9 \times 0.7 = 0.63$$

$$P(\text{白黑}) = P(\text{黑/白}) \times P(\text{白}) = 0.1 \times 0.7 = 0.07$$

$$P(\text{黑白}) = P(\text{白/黑}) \times P(\text{黑}) = 0.2 \times 0.3 = 0.06$$

$$P(\text{黑黑}) = P(\text{黑/黑}) \times P(\text{黑}) = 0.8 \times 0.3 = 0.24$$

$$H_2(X) = -\frac{1}{2} \sum_{x,y} p(xy) \log p(xy) = -\frac{1}{2} (0.63 \log 0.63 + 0.07 \log 0.07 + 0.06 \log 0.06 + 0.24 \log 0.24)$$

$$\approx 0.7125 \text{ bit/符号} \approx 0.713 \text{ bit/符号}$$

P(白黑) 和 P(黑白) 这两个概率不能相加的原因在于它们描述的是两个完全不同的事件序列。P(黑白) 涉及的是从“黑”到“白”的转换，而 P(白黑) 涉及的是从“白”到“黑”的转换。

另一种解法：

用全概率公式求出 $P(\text{白})$ 和 $P(\text{黑})$ ：

$$P(\text{白}) = P(\text{白/白}) \times P(\text{白}) + P(\text{白/黑}) \times P(\text{黑})$$

$$P(\text{黑}) = P(\text{黑/黑}) \times P(\text{黑}) + P(\text{黑/白}) \times P(\text{白})$$

$$P(\text{白}) + P(\text{黑}) = 1$$

$$\text{得：} P(\text{白}) = \frac{2}{3}, P(\text{黑}) = \frac{1}{3}$$

$$P(\text{白白}) = P(\text{白/白}) \times P(\text{白}) = 0.9 \times \frac{2}{3} = \frac{3}{5}$$

$$P(\text{白黑}) = P(\text{黑/白}) \times P(\text{白}) = 0.1 \times \frac{2}{3} = \frac{1}{15}$$

$$P(\text{黑白}) = P(\text{白/黑}) \times P(\text{黑}) = 0.2 \times \frac{1}{3} = \frac{1}{15}$$

$$P(\text{黑黑}) = P(\text{黑/黑}) \times P(\text{黑}) = 0.8 \times \frac{1}{3} = \frac{4}{15}$$

$$H_2(X) = -\frac{1}{2} \sum_{x,y} p(xy) \log p(xy) = -\frac{1}{2} \left(\frac{3}{5} \log \frac{3}{5} + \frac{2}{15} \log \frac{1}{15} + \frac{4}{15} \log \frac{4}{15} \right)$$

$$\approx \frac{1}{2} \times 1.472 = 0.736 \text{ bit/符号}$$

(3) 信源的剩余度：

信源的冗余度： $\gamma = 1 - \eta = 1 - \frac{H_\infty}{H_0}$ ，其中 H_∞ 为实际熵， $H_0 = \log q$ 为最大熵。信源效率 η 表示一个信源实际的信息熵与具有同样符号集的最大熵的比值。

$$\begin{aligned} H_0 &= \log q = \log 2 = 1 \\ \gamma_1 &= 1 - \frac{H(X)}{H_0} \approx 1 - \frac{0.881}{\log 2} \approx 0.119 \approx 0.12 \\ \gamma_2 &= 1 - \frac{H_2(X)}{H_0} \approx 1 - \frac{0.713}{\log 2} \approx 0.2875 \approx 0.287 \end{aligned}$$

结果说明：当信源的消息（符号）之间有依赖时，信源输出信息的不确定性减弱。本题中，当有依赖时前面已经是白色信息，后面绝大多数可能是出现白色消息；前面是黑色消息，后面基本上可猜测是黑色消息。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信源熵。这表明，信源熵正是反映信源的平均不确定性的。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（符号）之间依赖关系就越大。