1. 熵率

定义 1. 随机序列 X_n 的熵率定义为

$$H_\infty(X) = \lim_{n\to\infty} \frac{1}{n} H(X_1, \ldots, X_n) \tag{1}$$

定理 1. 对于离散平稳信源，若 $H(X_1) < \infty$，则 $H_\infty(X)$ 存在且

$$H_\infty(X) = \lim_{n\to\infty} H(X_n|X_{n-1}, X_{n-2}, \ldots, X_1) \tag{2}$$

证明.

$$\begin{aligned} H(X_n|X_{n-1}, X_{n-2}, \ldots, X_1) \leq & H(X_n|X_{n-1}, X_{n-2}, \ldots, X_2) \\ = & H(X_{n-1}|X_{n-2}, X_{n-3}, \ldots, X_1) \end{aligned}$$

所以 $H(X_n|X_{n-1}, X_{n-2}, \ldots, X_1)$ 单调递减。

$$\Rightarrow \frac{1}{n} H(X_1, \ldots, X_n) = \frac{1}{n} \sum_{i=1}^{n} H(X_i|X_{i-1}, X_1)$$

$$\Rightarrow \frac{1}{n} H(X_1, \ldots, X_n) \to \lim_{n\to\infty} H(X_n|X_{n-1}, X_{n-2}, \ldots, X_1) \text{ as } n \to \infty$$

□

以下分别针对三种常见的情形给出熵率的计算公式：

(a) 独立同分布:$H(X_1, \ldots, X_n) = \sum_{i=1}^{n} H(X_i) = nH(X_1) \Rightarrow H_\infty(X) = H(X_1)$ 即在独立同分布的情况下熵率等于熵。

(b) 平稳的马氏链：$H(X_n|X_{n-1}, X_{n-2}, \ldots, X_1) = H(X_n|X_{n-1}) = H(X_2|X_1)$ 设平稳分布为 π,π_i 表示处于状态 i 的概率。$p_{ij} = P(X_2 = j|X_1 = i)$，则

$$\begin{aligned} H(X_2|X_1) = & \sum_{i\in\mathcal{X}} \pi_i H(X_2|X_1 = i) \\ = & -\sum_{i,j\in\mathcal{X}} \pi_i p_{ij} \log p_{ij} \end{aligned}$$

例 1. 考虑一个两状态的马氏链，转移概率矩阵为 $\begin{bmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{bmatrix}$

解下面的方程

$$[\pi_1, \pi_2] = [\pi_1, \pi_2] \begin{bmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{bmatrix}$$

得

$$\pi_1 = \frac{\beta}{\alpha+\beta}$$
$$\pi_2 = \frac{\alpha}{\alpha+\beta}$$

假设两状态分别为 *1* 和 *2*，则 $H(X_2|X_1=1) = -(1-\alpha)\log(1-\alpha) - \alpha\log\alpha = h(\alpha)$ 同理 $H(X_2|X_1=2) = h(\beta)$，因此对两状态的马氏链，熵率为 $\pi_1 h(\alpha) + \pi_2 h(\beta)$

(c) 隐马尔科夫模型

定理 2. X_i 是平稳马氏链，$Y_i = f(X_i)$，则

$$H(Y_n|Y_{n-1},\dots,Y_1,X_1) \le H_\infty(Y) \le H(Y_n|Y_{n-1},\dots,Y_1) \tag{3}$$

$$\lim_{n\to\infty} H(Y_n|Y_{n-1},\dots,Y_1,X_1) = H_\infty(Y) = \lim_{n\to\infty} H(Y_n|Y_{n-1},\dots,Y_1) \tag{4}$$

证明. 上两式右端由平稳性可得,对于左端,首先说明 $H(Y_n|Y_{n-1},\dots,Y_1,X_1)$ 是 n 的增函数，这是因为

$$\begin{aligned} H(Y_n|Y_{n-1},\dots,Y_1,X_1) =& H(Y_n|Y_{n-1},\dots,Y_1,X_1,X_0) \\ =& H(Y_n|Y_{n-1},\dots,Y_1,X_1,X_0,Y_0) \\ \le& H(Y_n|Y_{n-1},\dots,Y_1,Y_0,X_0) \\ =& H(Y_{n+1}|Y_n,\dots,Y_2,Y_1,X_1) \end{aligned}$$

所以 $\lim_{n\to\infty} H(Y_n|Y_{n-1},\dots,Y_1,X_1)$ 存在。下面证明两极限相等，即证明当 $n\to\infty$ 时,$H(Y_n|Y_{n-1},\dots,Y_1) - H(Y_n|Y_{n-1},\dots,Y_1,X_1) = I(X_1;Y_n|Y_{n-1},\dots,Y_1) \to 0$ 把 $I(X_1;Y_n|Y_{n-1},\dots,Y_1)$ 看成某个级数的通项:

$$\sum_{i=1}^{n} I(X_1;Y_i|Y_{i-1},\dots,Y_1) = I(X_1;Y_1,\dots,Y_n) \le H(X_1)$$

所以级数 $\sum_{i=1}^{\infty} I(X_1;Y_i|Y_{i-1},\dots,Y_1)$ 收敛 $\Rightarrow I(X_1;Y_n|Y_{n-1},\dots,Y_1) \to 0$ 即

$$\lim_{n\to\infty} H(Y_n|Y_{n-1},\dots,Y_1,X_1) = \lim_{n\to\infty} H(Y_n|Y_{n-1},\dots,Y_1) = H_\infty(Y)$$

并由 $H(Y_n|Y_{n-1},\dots,Y_1,X_1)$ 递增的特性知(3)式左端成立。 □

2. 典型集

定义 2. 设 $X_1, \dots, X_n \sim p(x)$，*i.i.d.*，则关于 $p(x)$ 的典型集定义为以下序列的集合：

$$A_\epsilon^{(n)} = \{(x_1, \dots, x_n) \in \mathcal{X}^n : 2^{-n[H(x)+\epsilon]} \le p(x_1, \dots, x_n) \le 2^{-n[H(x)-\epsilon]}\} \tag{5}$$

典型集具有以下性质：

(a) 若 $(x_1, x_2, \dots, x_n) \in A_\epsilon^{(n)}$，则

$$H(X) - \epsilon \le -\frac{1}{n} \log p(x_1, \dots, x_n) \le H(X) + \epsilon \tag{6}$$

(b) 任意固定 $\epsilon > 0$，当 n 充分大时，$\Pr\{A_\epsilon^{(n)}\} \ge 1 - \epsilon$

证明. 由性质 (a),

$$\Pr\{A_\epsilon^{(n)}\} = \Pr\{| - \frac{1}{n} \log p(X_1, \dots, X_n) - H(X)| < \epsilon\}$$

由弱大数定律得：

$$-\frac{1}{n} \log p(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^{n} [-\log p(X_i)] \xrightarrow{P} \mathbb{E}_X[-\log p(X)] = H(X)$$

根据依概率收敛的定义得证。 □

(c) $|\cdot|$ 表示集合的元素个数，则 $|A_\epsilon^{(n)}| \le 2^{n[H(X)+\epsilon]}$

证明.

$$\begin{aligned}
|A_\epsilon^{(n)}| &= \sum_{\boldsymbol{x} \in A_\epsilon^{(n)}} 1 \\
&\le \sum_{\boldsymbol{x} \in A_\epsilon^{(n)}} p(x_1, \dots, x_n) 2^{n[H(X)+\epsilon]} \\
&\le 2^{n[H(X)+\epsilon]} \sum_{\boldsymbol{x} \in \mathcal{X}^n} p(x_1, \dots, x_n) \\
&= 2^{n[H(X)+\epsilon]}
\end{aligned}$$

□

(d) $\forall \epsilon > 0, n$ 充分大时，

$$A_\epsilon^{(n)} \geq (1-\epsilon)2^{n[H(X)-\epsilon]}$$

证明. 由 (b) 已知

$$\Pr\{A_\epsilon^{(n)}\} = \sum_{x \in A_\epsilon^{(n)}} p(x_1, \ldots, x_n) \geq 1 - \epsilon$$

$$\begin{aligned} |A_\epsilon^{(n)}| &= \sum_{\boldsymbol{x} \in A_\epsilon^{(n)}} 1 \\ &\geq \sum_{\boldsymbol{x} \in A_\epsilon^{(n)}} p(x_1, \ldots, x_n) 2^{n[H(X)-\epsilon]} \\ &\geq (1-\epsilon)2^{n[H(X)-\epsilon]} \end{aligned}$$

□

定义 3. 如果 $\forall \delta > 0$, 当 n 充分大时, $\Pr\{B_\delta^{(n)}\} \geq 1-\delta$, 则称 $B_\delta^{(n)} \subset \mathcal{X}^n$ 为包含大多数概率的子集（高概率集）。

下面的定理说明 $A_\epsilon^{(n)}$ 在一阶指数意义下是最小的高概率集:

定理 3. 设 $X_1, \ldots, X_n$ *i.i.d.* $\sim p(x)$, 固定 $\delta < \frac{1}{2}$ 及 $B_\delta^{(n)}$, 则对 $\forall \delta' > 0$, 当 n 充分大时,

$$\frac{1}{n} \log |B_\delta^{(n)}| \geq H(X_1) - \delta' \tag{7}$$

证明. 由 $P(A \cap B) > 1 - P(\bar{A}) - P(\bar{B})$ 得到当 n 充分大时，

$$P(B_\delta^{(n)} \cap A_\epsilon^{(n)}) > 1 - \epsilon - \delta$$

$$\begin{aligned} |B_\delta^{(n)}| &\geq |B_\delta^{(n)} \cap A_\epsilon^{(n)}| \\ &= \sum_{\boldsymbol{x} \in B_\delta^{(n)} \cap A_\epsilon^{(n)}} 1 \\ &\geq \sum_{\boldsymbol{x} \in B_\delta^{(n)} \cap A_\epsilon^{(n)}} p(x_1, \ldots, x_n) 2^{n[H(x)-\epsilon]} \\ &\geq (1-\epsilon-\delta)2^{n[H(x)-\epsilon]} \end{aligned}$$

□

3. 变长编码

$x \in \mathcal{X}, C(x)$ 表示对应 x 的码字，$l(x)$ 表示 $C(x)$ 的长度。若信源编码 C 中无任何码字是其他的前缀，则称 C 为即时码。

定理 4 (Kraft 不等式)**.** *D 元字母表上的即时码，设有 m 个码字，码长分别为 $l_1, \ldots, l_m$，则有:*

$$\sum_{i=1}^{m} D^{-l_i} \le 1 \tag{8}$$

证明. 设 $y = (y_1, \ldots, y_{l_i})$ 为码字，$y \leftrightarrow D$ 元小数 $0.y_1y_2\ldots y_{l_i} \triangleq \sum_{j=1}^{l_i} y_j D^{-j} \leftrightarrow$ 小区间 $I_y = (0.y_1y_2\ldots y_{l_i}, 0.y_1y_2\ldots y_{l_i} + D^{-l_i})$。注意到小区间的右端点是在 D 位小数的末位加 1, 如果对于另外一个 D 位小数 $\tilde{y}$ 对应的小区间与 y 对应的小区间相交，不妨设 $0.\tilde{y_1}\ldots\tilde{y_{l_k}} \in I_y$，则可以说明 y 是 $\tilde{y}$ 码字的前缀，这与即时码的定义相矛盾。因此各码字对应的小区间互不相交，其区间总长度为 $\sum_{i=1}^{m} D^{-l_i}$ 小于 $[0,1]$ 区间的长度 1。

□

定理 5. *任给即时码 $C, L(C) = \sum_{x\in\mathcal{X}} p(x)l(x)$ 成为 C 的平均码长，则有 $L(C) \ge H_D(X)$，且等号成立的充要条件是 $D^{-l_i} = p_i$*

注 1. *使得 $L(C)$ 最小的码称为最优码。*

证明. 由定理 4，设 $r = \sum_{x\in\mathcal{X}} D^{-l(x)} \le 1$ 则 $q(x) = \frac{D^{-l(x)}}{r}$ 是一个概率分布。

$$\begin{aligned}
L(C) - H_D(X) &= \sum_{x\in\mathcal{X}} p(x)[l(x) + \log_D(p(x))] \\
&= \sum_{x\in\mathcal{X}} p(x)[\log_D(p(x)) - \log_D D^{-l(x)}] \\
&= \sum_{x\in\mathcal{X}} p(x)\left[\log_D(p(x)) - \log_D \frac{D^{-l(x)}}{r}\right] - \log_D r \\
&= D(p||q) - \log_D r \\
&\ge 0
\end{aligned}$$

上式等号成立当且仅当 $r = 1$ 且 $p = q$，即 $D^{-l_i} = p_i$，从而要求 $-\log_D p_i$ 为整数。 □

定理 6. *设 C 为最优码，则 $L(C) < H_D(X) + 1$*

证明. 只需构造一种编码方式 C' 使得 $L(C') < H_D(X) + 1$。为此，设 $X \sim p(x), p(x_i) = p_i$, 取 $l_i = \lceil -\log_D p_i \rceil$ 因为 $l_i \geq \log_D p_i \Rightarrow D^{-l_i} \leq p_i \Rightarrow \sum_i D^{-l_i} \leq 1$ 所以存在一种即时码 C' 码长分别为 l_i。另一方面 $l_i < -\log_D p_i + 1 \Rightarrow$

$$L(C) = \sum_i p_i l_i < \sum_i p_i(-\log_D p_i + 1) = H_D(X) + 1$$

□

4. Huffmann 码

例 2. *见 3.tex 第 7 题。*

定理 7. *设 C^* 为 Huffman 码，C 为任意编码，则 $L(C^*) \leq L(C)$*

证明. 以二元编码 ($D = 2$) 为例：对 $|\mathcal{X}|$ 使用归纳法，当 $|\mathcal{X}| = 2$ 时显然成立。假设结论对任意给定的 $|\mathcal{X}| \leq m - 1$ 成立。考虑 $\mathcal{X} = \{x_1, \ldots, x_m\}, p_1 \geq p_2 \geq \cdots \geq p_{m-1} \geq p_m$。$C$ 是 $\mathcal{X}$ 上的任意即时码，不妨设 C 满足 $l_1 \leq l_2 \leq \cdots \leq l_{m-1} \leq l_m$。否则通过交换码字由排序不等式可以得到 $l_1 \leq l_2 \leq \cdots \leq l_{m-1} \leq l_m$ 的编码 C' 使得 $L(C') \leq L(C)$。因此，C 对应的概率最小的两个码字是最长的两个码字，进一步设它们有相同的长度，否则由即时码的性质将最长码字的末位去掉可以得到平均码长 $L(C)$ 更小的编码方案。因此 $l_{m-1} = l_m$。

考虑缩减信源 $\mathcal{X}'$, 其中 $p'_i = p_i, i = 1, \ldots, m-2, p'_{m-1} = p_{m-1} + p_m$, 设 C_1^* 为 $\mathcal{X}'$ 的 Huffman 编码，根据归纳假设对于任意 $\mathcal{X}'$ 上的即时码 C_1, 有 $L(C_1^*) \leq L(C_1)$。

另一方面：$L(C) = L(C'_1) + p_{m-1} + p_m$ 其中 C'_1 是 $\mathcal{X}'$ 上的编码方法，其由 C 诱导出，诱导规则为，对于前 $m-2$ 个字元码元不变，码长仍为 $l_1, \ldots, l_{m-2}$，对于第 $m-1$ 个字元，由于 C 是即时码，将原来 x_{m-1} 或 x_m 的码字最后一位去掉可作为第 $m-1$ 个字元的码字，码长为 $l_{m-1} - 1$。

$\Rightarrow L(C) \geq L(C_1^*) + p_{m-1} + p_m$, 不等式右端为常数（给定 $\mathcal{X}$）。

$$
\begin{aligned}
L(C_1^*) + p_{m-1} + p_m &= \sum_{i=1}^{m-2} (p_i' l_i^*) + l_{m-1}^* p_{m-1}' + p_{m-1} + p_m \\
&= \sum_{i=1}^{m-2} (p_i l_i^*) + (l_{m-1}^* + 1) p_{m-1} + (l_{m-1}^* + 1) p_m \\
&= \sum_{i=1}^{m} p_i l_i^*
\end{aligned}
$$

由 Huffmann 编码的构造过程可知，上式等于 $L(C^*)$, 其中 C^* 是 $\mathcal{X}$ 的 Huffman 编码 $\Rightarrow L(C) \geq L(C^*)$。根据归纳法可知对任意有限字母表，均有 $L(C) \geq L(C^*)$。 □

5. Shannon-Fano-Elias 码, 符号约定, $\bar{F}(x)$ 为修正的累积分布函数, $\bar{F}(x) = \sum_{a<x} p(a) + \frac{1}{2} p(x)$ 给定字母表有 5 个字母, 概率分别为 0.25, 0.25, 0.2, 0.15, 0.15 则编码如下表所示:

x	$p(x)$	$\bar{F}(x)$	$l(x) = \lceil \log \frac{1}{p(x)} \rceil + 1$	$\bar{F}(x) l(x)$ 位二进制表示	码字
1	0.25	0.125	3	0.001	001
2	0.25	0.375	3	0.011	011
3	0.2	0.6	4	0.1001	1001
4	0.15	0.775	4	0.1100	1100
5	0.15	0.925	4	0.1110	1110