- Fano 编码：

  设信源为

$$U=\begin{pmatrix} a_1 & a_2 & \cdots & a_n \\ p_1\geq & p_2\geq & \cdots & \geq p_n \end{pmatrix},$$

首先把消息分成两大组，使每组的概率之和尽量相同，即选一个k值，使

$$\left|\sum_{i=1}^{k} p_i-\sum_{i=k+1}^{n} p_i\right|$$

尽量小。这个k使消息分成两大组，给一组指定“0”，另一组指定“1”。然后再重复把每组中消息分成尽可能等概的两部分，再给每个部分指定“0”“1”，……

- 二元加权平衡树： 设信源U按 Fano 算法编码，可以得到一个二元加权平衡树。

  例：

面例子给出了Fano编码的过程。

[例 3.3.3] 一个有6个消息的信源

$$U=\begin{bmatrix} a_1 & a_2 & a_3 & a_4 & a_5 & a_6 \\ 0.3 & 0.25 & 0.2 & 0.15 & 0.05 & 0.05 \end{bmatrix}$$

熵为 $H(U)=2.3282$ bit。按照 Fano 编码算法

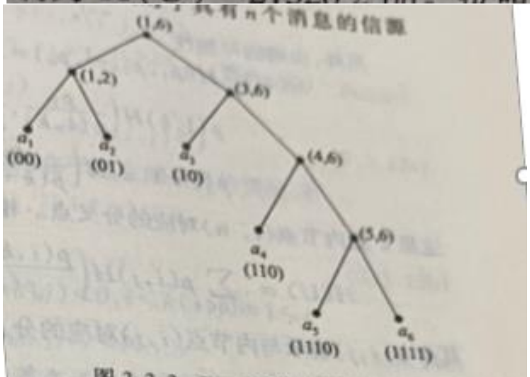

图 3.3.3 Fano编码的编码树

用(i,i)表示端节点 a_i。所有内节点(i,j)，$i<j$ 的全体记为 I。

在编码树的每个节点上定义权重。在节点(i,j)上的权重定义为

$$p(i,j)=\sum_{n=i}^{j} p_n$$

因此，在端节点(i,i)上的权重就是 p_i。这样就得到一个加权的二元判决树。判决树的每个内节点(i,j)，$i<j$，存在一个分叉，它把端节点子集$\{a_i,a_{i+1},\cdots,a_j\}$分成2个小的子集$\{a_i,a_{i+1},\cdots,a_k\}$和$\{a_{k+1},a_{k+2},\cdots,a_j\}$，$k$ 为(i,j)的分叉点。对于 Fano 编码来说，式(3.3.9)要求每个内节点(i,j)对应的分叉点 k 选得使左边和右边子集的权重尽可能相同，即要求

$$|p(i,k)-p(k+1,j)|=\min_{i\leq l<j}|p(i,l)-p(l+1,j)|$$

定理 3.3.3 信源(3.3.8)对应的 Fano 编码的平均码长

$$\bar{L} \leqslant H(U) + 1 - 2p_n \tag{3.3.10}$$

其中，p_n 是信源最小符号概率。

[证明] 设信源(3.3.8)按 Fano 算法编码，得到一个二元加权平衡树。消息按概率大小次序被安排在树的端节点上。每个码字的长度分别为从根节点$(1,\ n)$到对应端节点的路径中分支的数目。由于对内节点$(i,\ j)$

$$p(i,j) = p_i + p_{i+1} + \cdots + p_j \qquad i<j$$

从图 3.3.3 看出平均码长为

$$\bar{L} = \sum_{m=1}^{n} p_m l_m = \sum_{(i,j)\in I} p(i,j)$$

另外，由熵的可加性

$$\begin{aligned} H(U) &\stackrel{\text{def}}{=} H(p_1, p_2, \cdots, p_n) = H(p(1,k), p(k+1,n)) + \\ &p(1,k)H\left(\frac{p_1}{p(1,k)}, \frac{p_2}{p(1,k)}, \cdots, \frac{p_k}{p(1,k)}\right) + \\ &p(k+1,n)H\left(\frac{p_{k+1}}{p(k+1,n)}, \frac{p_{k+2}}{p(k+1,n)}, \cdots, \frac{p_n}{p(k+1,n)}\right) \end{aligned}$$

这里 k 是内节点$(1,\ n)$对应的分叉点。相继应用熵的可加性，最后可以得到

$$H(U) = \sum_{(i,j)\in I} p(i,j) H\left(\frac{p(i,k(i,j))}{p(i,j)}, \frac{p(k(i,j)+1,j)}{p(i,j)}\right)$$

其中 $k(i,j)$表示与内节点(i,j)对应的分叉点。所以

$$\begin{aligned} \Delta &= \bar{L} - H(U) \\ &= \sum_{(i,j)\in I} p(i,j)\left[1 - H\left(\frac{p(i,k(i,j))}{p(i,j)}, \frac{p(k(i,j)+1,j)}{p(i,j)}\right)\right] \end{aligned} \tag{3.3.11}$$

由熵函数的下凸特性(见图 3.3.4)，容易证明

$$\begin{cases} 1 - H(p,q) \leqslant q - p & p \leqslant q \\ 1 - H(p,q) \leqslant p - q & p \geqslant q \end{cases}$$

其中 $p + q = 1$。

所以 $0 \leqslant 1 - H(p,q) \leqslant |p - q|$

因此 $$\Delta \leqslant \sum_{(i,j)\in I} |\, p(i,k(i,j)) - p(k(i,j)+1,j)\,| \tag{3.3.12a}$$

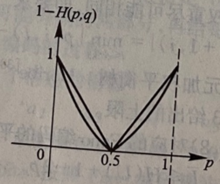

图 3.3.4 熵函数下凸示意图

下面来考察

$$\Delta(i,j) = |\, p(i,k(i,j)) - p(k(i,j)+1,j)\,| \qquad (3.3.$$

如果 $p(i,k(i,j))-p(k(i,j)+1,j)\geqslant 0 \quad i<k(i,j)<j$

由 Fano 编码原则,必定有

$$p(i,k(i,j)-1)-p(k(i,j),j)<0$$

于是
$$\begin{aligned}\Delta(i,j)&\leqslant p(k(i,j),j)-p(i,k(i,j)-1)\\&=[p_{k(i,j)}+p(k(i,j)+1,j)]-[p(i,k(i,j))-p_{k(i,j)}]\\&=2p_{k(i,j)}-\Delta(i,j)\end{aligned}$$

所以
$$\Delta(i,j)\leqslant p_{k(i,j)} \qquad i<k(i,j)<j \tag{3.3.13a}$$

对于 $k(i,j)=i$ 的情况,也就是分岔点落在子集左端边界的情况,有

$$\begin{aligned}\Delta(i,j)&=p_{k(i,j)}-p(k(i,j)+1,j)\leqslant p_{k(i,j)}-p_{\min}\\&=p_{k(i,j)}-p_n\end{aligned} \tag{3.3.13b}$$

同样如果 $p(i,k(i,j))-p(k(i,j)+1,j)<0, i<k(i,j)+1<j$

则
$$\Delta(i,j)\leqslant p_{k(i,j)+1}\leqslant p_{k(i,j)} \tag{3.3.14a}$$

对于 $k(i,j)+1=j$ 的情况,即分岔点落在子集右端边界,则

$$\begin{aligned}\Delta(i,j)&=p(k(i,j)+1,j)-p(i,k(i,j))\\&\leqslant p_{k(i,j)+1}-p_{\min}\\&\leqslant p_{k(i,j)}-p_n\end{aligned} \tag{3.3.14b}$$

由于内节点数目为 $n-1$,以及至少总有一个内点节具有$(i,i+1)$的形式,也就是说至少有一个内节点对应的分岔点落在子集左端或右端边界。所以

$$\begin{aligned}\Delta&\leqslant \sum_{(i,j)\in I}\Delta(i,j)\\&\leqslant \sum_{k=1}^{n-1}p_k-p_n\\&=1-2p_n\end{aligned}$$

所以
$$\bar{L}\leqslant H(U)+1-2p_n$$
Q.E.D

思考题：

- Shannon 编码：

  设信源为

$$\mathrm{U} = \begin{pmatrix} a_1 & a_2 & \cdots & a_n \\ p_1 \geq & p_2 \geq & \cdots & \geq p_n \end{pmatrix}。$$

对于符号a_i，首先确定它对应的码字长$l_i = \left\lceil \log_2 \frac{1}{p_i} \right\rceil$。接着，计算它对应的累加概率$\mathrm{F}_(i) = \sum_{k=1}^{i-1} p_k$。最后，它的码字为$\mathrm{F}_(i)$的二进制表示中小数点后$l_i$位数。

请问能否证明：对于任何离散无记忆信源$\mathbf{U} = \begin{pmatrix} \boldsymbol{a_1} & \boldsymbol{a_2} & \cdots & \boldsymbol{a_n} \\ \boldsymbol{p_1} \geq & \boldsymbol{p_2} \geq & \cdots & \geq \boldsymbol{p_n} \end{pmatrix}$，Fano 编码的信息冗余不会超过 Shannon 编码。