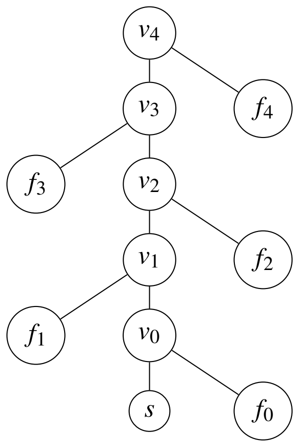
2.4 - Afitant Huffman

Tenim un alfabet Σ on per a cada símbol $a \in \Sigma$, p_a es la probabilitat que aparegui el caràcter a. Demostreu que, per a qualsevol símbol $a \in \Sigma$, la seva profunditat en un arbre prefix que produeix un codi de Huffman òptim és $O(\lg \frac{1}{p_a})$. (Ajuts: en un arbre prefix que s'utilitzi per a dissenyar el codi Huffman, la probabilitat d'un nus és la suma de les probabilitats dels fills. La probabilitat de l'arrel és, doncs, 1.)

Solució:

Fixem un símbol qualsevol a. Anomenem T a l'arbre amb el codi prefix de Huffman de l'alfabet. Localitzem el full s que conté el símbol a i considerem el camí $P = s, v_0, v_1, \ldots, v_k$ format per la seqüència de nodes al camí que va de s fins a l'arrel del arbre. $k+1$ és la profunditat on apareix a. Calculem una fita superior a k. Per això analitzarem el cas pitjor.
Anomenem f_i, $1 \leq i \leq k$ a l'altre fill de v_i. Aquests nodes poden aparèixer a la dreta o a l'esquerra, depenent de l'algoritme usat. Un possible arbre per a $k = 4$ és:



Tenint en compte la construcció de l'arbre sabem que:

- $p(v_0) \geq p(s) = p_a$
- Per $i \geq 1$, $p(v_i) = p(f_i) + p(v_{i-1})$
- Per $i \geq 2$, $p(f_i) \geq p(v_{i-2})$ y $p(f_i) \geq p(f_{i-1})$, si no fos així hauríem seleccionat abans f_i i no seria germà del seu pare.

De la darrera propietat tenim

$$2p(f_i) \geq p(v_{i-2}) + p(f_{i-1}) = p(v_{i-1}).$$

D'altra banda

$$p(v_i) = p(f_i) + p(v_{i-1}) \geq \frac{p(v_{i-1})}{2} + p(v_{i-1}) \geq \frac{3}{2} p(v_{i-1}).$$

Deduïm que

$$p(v_k) \geq \left(\frac{3}{2}\right)^k p(v_0) \geq (1.5)^k p_a.$$

Com que v_k és l'arrel de l'arbre $p(v_k) = 1$, tenim l'equació $(1.5)^k p_a \leq 1$. Així

$$k \log 1.5 \leq \log \frac{1}{p(a)},$$

i deduïm

$$k = O(\log \frac{1}{p(a)}).$$