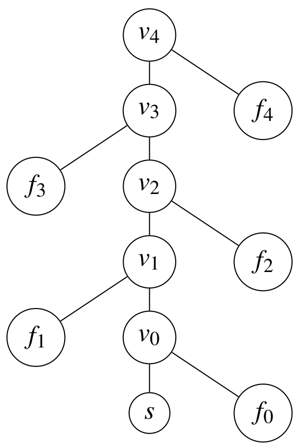
2.9 - Huffman té profunditat òptima

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 de prof. M. Serna]

Fijamos un símbolo cualquiera a. Llamemos T al árbol con el código prefijo de Huffman del alfabeto. Localizamos la hoja s que contiene el símbolo a y consideramos el camino $P = s, v_0, v_1, \ldots, v_k$ formado por la secuencia de nodos en el camino que va de s hasta la raíz del árbol. $k+1$ es la profundidad en la que aparece a. Vamos a calcular una cota superior a k. Para ello analizaremos el caso peor.
Llamemos f_i, $1 \leq i \leq k$ al otro hijo de v_i. Estos nodos pueden aparecer a la derecha o a la izquierda, dependiendo del algoritmo usado un posible árbol, para $k = 4$, es:



Teniendo en cuenta la construcción del árbol sabemos que:

- $p(v_0) \geq p(s) = p_a$
- Para $i \geq 1$, $p(v_i) = p(f_i) + p(v_{i-1})$
- Para $i \geq 2$, $p(f_i) \geq p(v_{i-2})$ y $p(f_i) \geq p(f_{i-1})$, si no fuese así habríamos seleccionado antes f_i y no sería hermano de su padre.

De la última propiedad tenemos

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

Por otra parte

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

Deducimos que

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

Como v_k es la raíz del árbol $p(v_k) = 1$, tenemos la ecuación $(1.5)^k p_a \leq 1$. Así

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

y deducimos

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