Input n:

Print "$\lfloor \frac{n}{15} \rfloor$ porters"

$n := n - 15\lfloor \frac{n}{15} \rfloor$

Print "$\lfloor \frac{n}{9} \rfloor$" miners"

$n := n \bmod 9$

Print "$\lfloor \frac{n}{4} \rfloor$ sickles"

$n := n \bmod 4$

Print "n pennies"

27, 12, 18, 19, 31, ...

Huffman coding is optimal

Define cost of a coding tree T

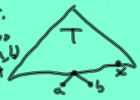
$$B(T) = \sum_{a \in C} a.f \times a.d$$

where a.d is depth of the leaf labeled by a in the tree, and a.f is the frequency of a.

<u>Lemma</u>: For any alphabet C with letter frequencies, ($|C| \geq 2$) there exists an optimal tree for C where the two lowest frequency letters are siblings on the deepest level.

<u>Proof</u>: Let T be some encoding tree for C. Let x and y be the two lowest freq chars in C. WOLOG, $x.f \leq y.f$.

There exist two siblings (say a, b) on the deepest level of T.



If $\{a,b\} = \{x,y\}$ done.
So assume $\{a,b\} \neq \{x,y\}$
Assume $x \notin \{a,b\}$ and $a \notin \{x,y\}$. (Other combo's are similar)

Let T′ be the tree obtained from T by swapping the labels a with x. Consider

$B(T) - B(T')$

<u>Notation</u>: for any $c \in C$, let
$d_T(c)$ be the depth of c in T
$d_{T'}(c)$ " " " " T′

$\underline{B(T)} - B(T') = d_T(x)\cdot(x.f)$
$+ d_T(a)\cdot(a.f) - d_{T'}(x)\cdot(x.f)$
$- d_{T'}(a)\cdot(a.f)$
$= a.f\left(d_T(a) - d_{T'}(a)\right)$
$+ x.f\left(d_T(x) - d_{T'}(x)\right)$

$\begin{bmatrix} d_T(a) = d_{T'}(x) \\ d_T(x) = d_{T'}(a) \end{bmatrix}$

$= a.f\left(d_T(a) - d_T(x)\right)$
$+ x.f\left(d_T(x) - d_T(a)\right)$

$= \underbrace{(a.f - x.f)}_{\geq 0}\underbrace{(d_T(a) - d_T(x))}_{\geq 0}$

≥ 0

$\therefore B(T) - B(T') \geq 0.$

$\therefore B(T) \geq B(T')$

Swap again if necessary to get a T″ having x, y in place of a, b in T and $B(T'') \leq B(T)$.

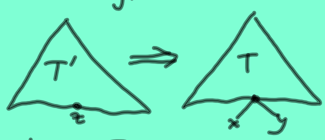
Picking T optimal means that T″ is also optimal ▨

Huffman(C):

Choose $x, y \in C$ with min freq
remove x, y from C
Form new symbol $z \notin C$ with $z.f := x.f + y.f$
Let $C' := (C - \{x,y\}) \cup \{z\}$
Let T′ be the tree returned by Huffman(C′)

Oct 27-12:50 PM

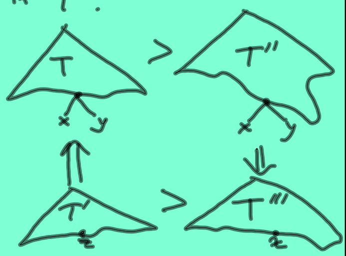
Let T be the result of replacing leaf z with an internal node with children x and y.



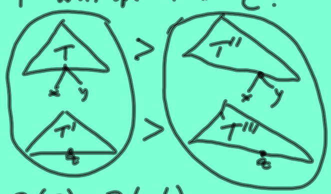
return T.

<u>Claim</u>: If T' is optimal for C', then T is optimal for C.
(apply Claim inductively shows Huffman(C) is optimal for all C)

<u>Proof of Claim</u>: For contradiction, assume T' is optimal but T is <u>not</u> optimal. So there exists an (optimal) tree T'' for C such that $B(T'') < B(T)$.
By the lemma, we can assume that x & y are siblings in T''.



Let T''' result from T'' by replacing the parent of x and y with leaf z with freq
$$z.f := x.f + y.f.$$
T''' is an encoding tree for C'
We show that $B(T''') < B(T')$ contradicting the assumption that T' was optimal for C'.



$$B(T) - B(T')$$
$$= d_T(x)\cdot(x.f) + d_T(y)\cdot(y.f) - d_{T'}(z)\cdot(z.f)$$
$$= d_T(x)\underbrace{(x.f + y.f)}_{z.f} - d_{T'}(z)\cdot(z.f)$$
$$= (z.f)\left(d_T(x) - d_{T'}(z)\right) = z.f$$

Similarly, $B(T'') - B(T''') = z.f$
$$= B(T) - B(T')$$
$\therefore$
$$B(T''') = \underbrace{B(T'') + B(T') - B(T)}_{<0}$$
$$< B(T')$$
$\therefore$ T' is not optimal $\Rightarrow\!\Leftarrow$

Oct 27-1:23 PM