Problem 3. Mar 6th, 8:00PM chaoqian & silao_xu

12.

Claim: Using Shannon-Fano algorithm, if a symbol σ ends up at depth d then its probability must be at most $2/2^d$.

Proof:

Step 1: claim: for any internal node v, define the function $f(v)$ to be the total of all the probabilities of all the descendents of v except the smallest probability one, if p is a parent of v, then $f(p) \geq 2f(v)$

Proof: ① If v is the right child of p, then, let s be the smallest probability symbol of all the descendents of p, according to the definition of f:

$f(p)$ = Total probility of left child + ~~f(v)~~ Total probability of right child $- s$

because the probabilities have been sorted in descending order, so the smallest probility of p is also the smallest probability of v (right child), so:

$f(p)$ = Total probability of left child $+ f(v)$

~~because~~ according to step 3 of the algorithm, the left part's total probability must be greater than (Total probability of parent $- s)/2$, so:

Total probability of left child $\geq \frac{1}{2} * f(p)$

Thus: $f(v) = f(p) -$ Total probability of left child $\leq f(p) - \frac{1}{2}f(p) = \frac{1}{2}f(p)$

$f(p) \geq 2f(v)$

② If v is the left child of p, let s be the smallest probability of all descend of p, s' be the smallest probability of all descendents of v, and T is the total probability function, then:

("p→right" means the right child of p)

$f(p) = T(v) + T(p\to right) - s$

$f(v) = T(v) - s'$

Contradiction Hyperthesis:

Assume: ~~若~~ $f(v) > \frac{1}{2} f(p)$ then: $2f(v) > T(v) + T(p\to right) - S$

because $T(v) = f(v) + s'$ and $f(p\to right) = T(p\to right) - S$

so $f(v) > f(p\to right) + s'$ (#)

The difference of left part and right part is:

$$D_1 = \left| T(v) - T(p\to right) \right| = \left| f(v) + s' - f(p\to right) - s \right|$$

because $f(v) > f(p\to right) + s'$, so $D_1 = f(v) + s' - f(p\to right) - s$;

and we change the ~~stp~~ split location of P, letting s' removed from descents of v and ~~intersed~~ inserted into p's right child's descendents.

In this case, $T(v') = f(v)$ $T(p'\to right) = f(p\to right) + s' + s$, $T(p') = f(p) + s$

because $f(v) > f(p\to right) + s'$ so: $T(v') = f(v) > \cancel{T(p') - T(v') - s}\ f(p\to right) + s'$

$$= T(p'\to right) - S$$
$$= T(p') - T(v') - S$$

$T(v') > \frac{T(p') - S}{2}$, this satisfies one condition in step 3 of the algorithm

so to make sure We do not ~~stip~~ split P in this way, we have the make sure the difference of the new 2 parts is larger than D_1, thus:

$$\left| f(v) - s' - f(p\to right) - s \right| > D_1 = f(v) + s' - f(p\to right) - s \quad (*)$$

as ~~s'~~ $s' > 0$, so to satisfy (*), ~~the~~

$$-f(v) + s' + f(p\to right) + s > f(v) + s' - f(p\to right) - s$$

So: $f(v) - f(p\to right) - s < 0$

$f(v) < f(p\to right) + S$

according to (#): $f(v) > f(p\to right) + s'$

So: $S' < S$, ~~t~~ this is in contradiction with the fact that the probabilities are sorted in descending order, so contradiction hyperthesis does not stand.

$f(v) \leq \frac{1}{2} f(p)$, $f(p) \geq 2f(v)$

according to ① and ②., we can conclude that $f(p) \geq 2f(v)$
step 1 complete.

Step2: Claim: if a symbol σ ends up at depth d then its probability must be at most $2/2^d$.

Proof: Induction:

base: $f(root) = 1 - s \leq 1 = 1/2^0$ with root in depth 0

Induction Hypothesis: if p is in depth d-2, and $f(p) \leq 1/2^{d-2}$

Then, according to step 1, $f(p) \geq 2f(v)$

so $f(v) \leq \frac{1}{2} f(p) \leq 1/2^{d-1}$, with v in depth $(d-1)$.

Because $f(v) \leq 1/2^{d-1}$ with internal node v in depth $d-1$,

and $f(v) =$ Total probility of leaf nodes in depth d - smallest one,

so for a symbol σ with depth d

$p(\sigma) \leq f(\sigma.parent) \leq 1/2^{d-1} = 2/2^d$ (σ.parent is the parent node of σ, with depth $d-1$)

proof complete.