ENTROPY OF TERMINAL DISTRIBUTIONS AND THE FIBONACCI TREES

YASUICHI HORIBE
Shizuoka University, Hamamatsu, 432, Japan

(Submitted March 1986)

Continuing the previous papers (see [1] and [2]), several new properties of binary trees, especially Fibonacci trees, have been found and will be shown in this note. For this, we shall occasionally need to refer to some of the notations, definitions, and results given in those papers.

1. BINARY TREES WITH BRANCH COST

Consider a binary tree with $n - 1$ internal nodes 1, 2, ..., $n - 1$ and n terminal nodes (leaves) 1, 2, ..., n. An internal node has two sons, while a terminal node has none. A node is at level ℓ if the path from the root to this node has ℓ branches. When, as in [1] and [2], unit cost 1 is assigned to each left branch and cost $c > 0$ to each right, we say the tree is "$(1, c)$-assigned." The cost of a node is then defined as follows: The cost of the root node is 0, and the cost of the left [right] son of a node of cost b is $b + 1$ [$b + c$]. Denoting by a_i [b_j] the cost of terminal node i [internal node j], we have the relation:

$$\sum_{i=1}^{n} a_i = \sum_{j=1}^{n-1} b_j + (n - 1)(1 + c). \tag{1}$$

This is proved easily by induction on n (see [1]).

The sum on the left-hand side of (1) is called the *total cost* of the tree. Let us say that a binary tree is *c-minimal* (or *c-optimal* [2]) if, when $(1, c)$-assigned, it has the minimum total cost of all the $(1, c)$-assigned binary trees having the same number of terminal nodes.

2. BINARY TREES WITH BRANCH PROBABILITY

We may also assign, instead of cost, probability p $(0 < p < 1)$ to each left branch and $\bar{p} = 1 - p$ to each right. We then say the tree is "$(p, \bar{p})$-assigned." The probability of a node is defined as follows: The probability of the root is 1 and the probability of the left [right] son of a node of probability q is pq [$\bar{p}q$]. Let p_i [q_j] be the probability of terminal node i [internal node j].

The $(p, \overline{p})$-assignment may be interpreted as a transportation of "nourishment" of unit amount, along paths from the root to leaves, with rates p and $\overline{p}$ to the left and right branches at each internal node. The probabilities p_i of terminal nodes, whose sum is of course 1, show the distribution of the nourishment among leaves, and will be called the *terminal distribution*.

We are especially interested in such trees that have terminal distribution as uniform as possible, given p and the number of terminal nodes. For fixed n, the uniformity of a probability distribution $p_1, \ldots, p_n$ can be measured appropriately by the entropy function

$$H(p_1, \ldots, p_n) = -\sum p_i \log p_i \quad \text{(log-base = 2)},$$

as will be seen in the following sections.

A binary tree is called *p-maximal* if, when $(p, \overline{p})$-assigned, it has the maximum entropy of all $(p, \overline{p})$-assigned binary trees having the same number of terminal nodes.

3. ENTROPY FUNCTION

The entropy function measures the uniformity or the uncertainty of the probability distribution (see [3], and also [1]). It is well known that $H(p_1, \ldots, p_n)$ attains its maximum value $\log n$ only in the case of the complete uniformness:

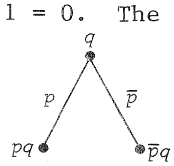
$$p_1 = \cdots = p_n = 1/n.$$

The following lemma is a variant of the so-called "branching property" of the entropy.

Lemma 1: Given a $(p, \overline{p})$-assigned tree, the entropy of the terminal distribution is given by

$$H(p_1, \ldots, p_n) = H(p, \overline{p}) \sum_{j=1}^{n-1} q_j. \tag{2}$$

Proof: Our binary tree can be viewed as grown by $n - 1$ successive branchings, starting with the branching of the root node. The entropy is initially zero: $H(1) = -1 \log 1 = 0$. The entropy increase due to the branching of a node of probability q:



is readily seen to be

$$-(pq)\log(pq) - (\overline{p}q)\log(\overline{p}q) - (-q \log q) = (-p \log p - \overline{p} \log \overline{p})q, \tag{3}$$

hence completing the proof by induction. ■

It should be noted that the sum of the internal node probabilities is equal to the *average path length* for the terminal nodes:

$$\sum_{j=1}^{n-1} q_j = \sum_{i=1}^{n} p_i \ell_i .$$

Here, ℓ_i is the level where terminal node i exists. This equality holds because each p_i contributes to both sides exactly ℓ_i times.

Lemma 1 can, therefore, be interpreted as follows: "A terminal node can be reached from the root with $\sum q_j$ branchings on the average, and the uncertainty produced per branching is $H(p, \overline{p})$, so the uncertainty of the terminal distribution should be $H(p, \overline{p})\sum q_j$."

Let us digress here to consider the following question: Suppose, conversely, that the following functional equation in the same form as (3) is given for some nonnegative function $f(t)$ defined on $0 < t \leqslant 1$:

$$f(pq) + f(\overline{p}q) - f(q) = (f(p) + f(\overline{p}))q, \quad 0 < p < 1, \quad 0 < q \leqslant 1. \tag{4}$$

Then, how well will f be characterized?

Theorem 0: If $f(t)$ is defined on $0 < t \leqslant 1$ and satisfies (4), then

$f(t) = -ct \log t$ for some constant $c \geqslant 0$.

Proof: Take $q = 1$. Then $f(1) = 0$. Let us put $g(t) = f(t)/t$. We have $g(1) = 0$, and (4) becomes

$$pg(pq) + \overline{p}g(\overline{p}q) - g(q) = pg(p) + \overline{p}g(\overline{p}). \tag{5}$$

Taking $p = 2^{-1}$ gives $g(2^{-1}q) = g(2^{-1}) + g(q)$. Repeating this gives $g(2^{-N}q) = Ng(2^{-1}) + g(q)$; hence,

$$2^{-N}g(2^{-N}q) \to 0, \quad N \to \infty. \tag{6}$$

Rearrange terms in (5) to obtain:

$$p^2 \frac{g(p) - g(pq)}{p - pq} + \overline{p}^2 \frac{g(\overline{p}) - g(\overline{p}q)}{\overline{p} - \overline{p}q} = \frac{f(1) - f(q)}{1 - q} \cdot \frac{1}{q}, \text{ for } q < 1. \tag{7}$$

Letting $q \to 1$ in (7) yields

$$p^2 g'(p) = \overline{p}^2 g'(\overline{p}) = -c_1 \text{ (constant)}. \tag{8}$$

Next, we take the integral $\int_{2^{-N}}^{1} dq$ to both sides of (5). Then,

$$\int_{2^{-N}p}^{p} g(t)dt + \int_{2^{-N}\overline{p}}^{\overline{p}} g(t)dt - \int_{2^{-N}}^{1} g(t)dt = (pg(p) + \overline{p}g(\overline{p}))(1 - 2^{-N}). \tag{9}$$

Differentiating (9) with respect to p, and then letting N go to infinity, we have, using (6),

$$pg'(p) = \overline{p}g'(\overline{p}). \tag{10}$$

From (8) and (10), we have $pg'(p) = -c_1$. Hence, $g(p) = -c_1 \ln p + d$. We must have $d = 0$ and $c_1 \geqslant 0$, because $g(1) = 0$ and $g(p) \geqslant 0$. Consequently,

$$g(t) = -c \log t \text{ on } 0 < t \leqslant 1, \text{ for some constant } c \geqslant 0. \blacksquare$$

(For a derivation of the entropy function under a more general condition, see [4].)

4. DUALITY

In this section, we present and prove the following theorem.

Theorem 1: Let $c > 0$ and $0 < p < 1$ satisfy $p^c = \overline{p}$. Then a binary tree is c-minimal if and only if it is p-maximal.

Proof: Consider the infinite complete binary tree T_∞. Because of (1), a c-minimal tree having n terminal nodes can be found in T_∞ by picking the $n - 1$ cheapest nodes $1, 2, \ldots, n - 1$ to be internal, if the nodes of the $(1, c)$-assigned T_∞ are numbered $1, 2, \ldots$ such that

$$b_1 \leqslant b_2 \leqslant \cdots . \tag{11}$$

(Also see [1] in this respect.) The ordering (11) is equivalent to the ordering

$$p^{b_1} \geqslant p^{b_2} \geqslant \cdots . \tag{12}$$

If node j is reached from the root by r left branches and s right branches, we have $b_j = r + sc$. Now, from the assumption $p^c = \overline{p}$, we have $q_j = p^r(\overline{p})^s = p^{b_j}$ in the $(p, \overline{p})$-assigned T_∞. Hence, because of Lemma 1, the tree thus found must be p-maximal. $\blacksquare$

A most interesting c, p satisfying $p^c = \overline{p}$ is $c = 2$,

$$p = \psi = (\sqrt{5} - 1)/2 \quad (\overline{\psi} = \psi^2).$$

5. FIBONACCI TREES

We can now apply Theorem 1 to the Fibonacci trees (see [2]). The Fibonacci tree of order k, denoted by T_k, is a binary tree having $n = F_k$ terminal nodes, and defined inductively as follows: T_1 and T_2 are simply the root nodes only. The left subtree of T_k $(k \geqslant 3)$ is T_{k-1} and the right is T_{k-2}. Here, F_k is the k^{th} Fibonacci number: $F_0 = 0$, $F_1 = 1$, $F_k = F_{k-1} + F_{k-2}$.

It was shown in [1] that T_k is 2-minimal for every k. Hence, by Theorem 1, T_k is ψ-maximal for every k.

The following theorem was proved in [2].

Theorem 2: When $1 \leqslant c < 2$, T_k $(k \geqslant 3)$ is c-minimal if and only if

$$k \leqslant 2\left\lfloor \frac{1}{2 - c} \right\rfloor + 3.$$

When $c > 2$, T_k $(k \geqslant 3)$ is c-minimal if and only if

$$k \leqslant 2\left\lfloor \frac{1}{c-2} \right\rfloor + 4.$$

($\lfloor x \rfloor$ is the largest integer $\leqslant x$.)

Translating this into its dual form by using Theorem 1, we have

Theorem 3: For even $k \geqslant 6$, T_k is p-maximal if and only if

$$p^{2\left(1+\frac{1}{k-4}\right)} \leqslant \overline{p} \leqslant p^{2\left(1-\frac{1}{k-2}\right)}.$$

For odd $k \geqslant 5$, T_k is p-maximal if and only if

$$p^{2\left(1+\frac{1}{k-3}\right)} \leqslant \overline{p} \leqslant p^{2\left(1-\frac{1}{k-3}\right)}.$$

In [1] it was shown that the (1, 2)-assigned T_k has F_{k-1} terminal nodes (called α-nodes in [2]) of cost $k - 2$ and F_{k-2} terminal nodes (β-nodes) of cost $k - 1$. Since each α-node [β-node] has probability ψ^{k-2} [ψ^{k-1}] in the $(\psi, \overline{\psi})$-assigned T_k, we have the following terminal distribution:

$$\underbrace{\psi^{k-2}, \ldots, \psi^{k-2}}_{F_{k-1}}, \underbrace{\psi^{k-1}, \ldots, \psi^{k-1}}_{F_{k-2}}.$$

Hence, we have

$$F_{k-1}\psi^{k-2} + F_{k-2}\psi^{k-1} = \frac{1}{\sqrt{5}}\psi^{-1} + \frac{1}{\sqrt{5}}\psi = 1.$$

$$\left(F_{k-1}\psi^{k-2} \sim \frac{1}{\sqrt{5}}\psi^{-1} = 0.724,\ F_{k-2}\psi^{k-1} \sim \frac{1}{\sqrt{5}}\psi = 0.276.\right)$$

The entropy of the above terminal distribution is computed as

$$\begin{aligned} &-F_{k-1}\psi^{k-2}\log\psi^{k-2} - F_{k-2}\psi^{k-1}\log\psi^{k-1} \\ &= (-\log\psi)\{(k-2) + F_{k-2}\psi^{k-1}\}. \end{aligned} \tag{13}$$

By a numerical computation, the ratio of this entropy and the entropy $\log F_k$ of the completely uniform distribution is approximately $1 - (0.05)/(k - 1.67)$.

Finally, let us compute the entropy of the terminal distribution of the $(p, \overline{p})$-assigned T_k. Denote the entropy by H_k for simplicity. Then, trivially, $H_1 = H_2 = 0$ and $H_3 = H(p, \overline{p})$. By Lemma 1 and by the recursive structure of the Fibonacci tree, the sum of the internal node probabilities of T_k is given by

$$1 + p\frac{H_{k-1}}{H_3} + \overline{p}\frac{H_{k-2}}{H_3} \quad (k \geqslant 3).$$

Hence, we have the "Fibonacci branching of the entropy":

$$H_k = H_3 + pH_{k-1} + \overline{p}H_{k-2}.$$

Putting $\Delta H_k = H_k - H_{k-1}$, we have $\Delta H_k = -\overline{p}\Delta H_{k-1} + H_3$; therefore,

$$\Delta H_k = \frac{H_3}{2 - p}\{1 - (-\overline{p})^{k-2}\}, \; k \geqslant 3,$$

$$H_k = \sum_{m=3}^{k} \Delta H_m = \frac{H(p, \overline{p})}{2 - p}\left\{(k - 2) + \frac{\overline{p} + (-\overline{p})^{k-1}}{1 + \overline{p}}\right\}, \; k \geqslant 3.$$

When $p = \psi$, H_k becomes (13), as can be checked. The p that maximizes H_k approaches ψ as k becomes large. This is because the maximization then almost becomes the maximization of the function

$$F(p) = \frac{H(p, \overline{p})}{2 - p},$$

and the maximum (= $-\log \psi$) of $F(p)$ is attained only when $p = \psi$. The maximization of $F(p)$ has already appeared in [1] in a closely related context.

REFERENCES

1. Y. Horibe. "An Entropy View of Fibonacci Trees." *The Fibonacci Quarterly* *20*, no. 2 (1982):168-178.
2. Y. Horibe. "Notes on Fibonacci Trees and Their Optimality." *The Fibonacci Quarterly* *21*, no. 2 (1983):118-128.
3. A. Rényi. *Tagebuch über die Informationstheorie*. (German translation from Hungarian). Berlin: VED Deutscher Verlag der Wissenschaften, 1982. [English translation: *A Diary on Information Theory*. Budapest: Akadémiai Kiadó, 1984.]
4. H. Tverberg. "A New Derivation of the Information Function." *Math. Scand.* *6* (1958):297-298.

◆◇◆◇◆