INSTITUTUL
DE
MATEMATICA

INSTITUTUL NATIONAL
PENTRU CREATIE
STIINTIFICA SI TEHNICA

ISSN 0250 3638

THE ENTROPY, AS A MINIMAL COST OF A TREE

by

Gh. ŞTEFANESCU

PREPRINT SEREIS IN MATHEMATICS
No.11/1982

BUCURESTI

THE ENTROPY, AS A MINIMAL COST OF A TREE

by

Gh. ŞTEFĂNESCU*)

February 1982

*) Department of Mathematics, National Institute for Scientific and Technical Creation, Bd. Pacii 220, 79622 Bucharest, Romania.

THE ENTROPY, AS A MINIMAL COST OF A TREE

by

Gh. ŞTEFĂNESCU

0. Abstract

In this paper we try to give an answer to the question: "Is optimal, in any sense, the entropy". The affirmative answer is obtained by leaving from the observation ([?], Cap.1.5) that for $p_i > 0$, $\sum_{i=1}^{n} p_i = 1$ the optimal binary tree with n-outs by p_i probability has the average cost between $H(p_1, \ldots, p_n)$ and $H(p_1, \ldots, p_n)+1$, where $H(p_1, \ldots, p_n) = -\sum_{i=1}^{n} p_i \log_2 p_i$. We construct continuous tree and show that the minimal cost, in this context, is exact $H(p_1, \ldots, p_n)$.

1. Definitions, notations

a) The power space

Let $(X, \mathcal{X}, \mu)$ and $(\mathbf{2}, \mathcal{P}(\mathbf{2}), \nu)$ be two measure spaces ($\mathcal{X}$ algebra, μ not necessary σ-additive), where $\mathbf{2} = \{1, 2\}$ and $\nu(\{1\}) = \nu(\{2\}) = 1$. The power space is:

$$Y = \{f : X \multimap \mathbf{2} \mid f \text{ measurable, partial function}\}$$

Let denote:

1) $f \in Y \Rightarrow \text{dom } f = f^{-1}(\mathbb{2}) \in \mathfrak{X}$; $M \subseteq Y \Rightarrow \text{dom } M = \bigcup_{f \in M} \text{dom } f$

2) $f \in Y$, $A \subseteq X \Rightarrow f|_A \in Y$ is the restriction of f to A.

3) $f, g \in Y$ with $\text{dom } f \cap \text{dom } g = \emptyset \Rightarrow f+g \in Y$ is the function:

$$(f+g)(x) = \begin{cases} f(x), & \text{if } x \in \text{dom } f \\ g(x), & \text{if } x \in \text{dom } g \\ \text{undefined}, & \text{elsewhere} \end{cases}$$

Consider also the natural extension: $M, N \subseteq Y$ with $\text{dom } M \cap \text{dom } N = \emptyset \Rightarrow$

$M+N = \{f+g \mid f \in M,\ g \in N\}$.

4) $f \in Y$, $A \in \mathfrak{X} \Rightarrow lr_f(A) = \mu(f^{-1}(\{1\}) \cap A) - \mu(f^{-1}(\{2\}) \cap A)$

(lr: left-right)

5) $A \in \mathfrak{X} \Rightarrow \mathbb{2}^A = \{f : A \to \mathbb{2} \,/\, f \text{ measurable, total function}\} \subseteq Y$.

6) $f \in Y$ and $A \in \mathfrak{X}$ with $A \subseteq \text{dom } f$:

$$f\uparrow A = \{f|_{X \setminus A}\} + \{g \in \mathbb{2}^A \,/\, lr_f(A) \geq 0 \iff lr_g(A) \geq 0\}$$

$$f\downarrow A = \{f|_{X \setminus A}\} + \{g \in \mathbb{2}^A \,/\, lr_f(A) \geq 0 \iff lr_g(A) < 0\} .$$

The natural extension is: $M \subseteq Y$ and $A \in \mathfrak{X}$ with $\forall f \in M$, $A \subseteq \text{dom } f$

$\Rightarrow M\uparrow A = \bigcup_{f \in M} f\uparrow A$ and $M\downarrow A = \bigcup_{f \in M} f\downarrow A$.

b) <u>Syntactic representation</u>

Now we give, in some words, the basic notions in the theory of (ordinary) labeled trees (in ADJ's notations: see, for example, [2]). We work only with binary trees. If $\{1,2\}^*$ is the free monoid over $\{1,2\}$ (the set of finite words with 1 or 2, Ω the void word and as operation, concatenation) and Σ a label set, then a <u>Σ-tree</u> is a partial function $T : \{1,2\}^* \to \Sigma$

for which: $\forall\, w_1 \ldots w_n \in \operatorname{dom} T \Rightarrow w_1 \ldots w_{n-1} \in \operatorname{dom} T$. (if $n=1$ the word $w_1 \ldots w_{n-1}$ means Ω).

Every $w_1 \ldots w_n \in \operatorname{dom} T$ is a <u>nod</u> and $T(w_1 \ldots w_n)$ is its label.

A Σ-tree T is <u>total</u> if: $\forall\, w_1 \ldots w_n \in \operatorname{dom} T \Rightarrow w_1 \ldots w_{n-1} \overline{w}_n \in \operatorname{dom} T$ (where $\overline{1}=2$ and $\overline{2}=1$) and <u>finite</u> if $\operatorname{dom} T$ is a finite set.

A <u>leaf</u> of T is a nod $w_1 \ldots w_n \in \operatorname{dom} T$ with $w_1 \ldots w_n i \notin \operatorname{dom} T$ for any $i \in \{1,2\}$. A <u>path</u> in T from root Ω to a nod $w_1 \ldots w_n \in \operatorname{dom} T$ is the sequence of nods: $(\Omega, w_1, w_1 w_2, \ldots, w_1 \ldots w_n)$

In this notation a path is unique defined by its last nod.

A <u>terminal path</u> is a path $(\Omega, \ldots, w_1 \ldots w_n)$ with $w_1 \ldots w_n$ a leaf.

The <u>length</u> of T is $\lg(T) = \max\{i \mid \exists\, w_1 \ldots w_i \in \operatorname{dom} T\}$.

Another useful (recursive) definition of Σ-trees is:

i) for any $\sigma \in \Sigma \Rightarrow \cdot\sigma$ is a Σ-tree.

ii) if T_1, T_2 are Σ-trees and $\sigma \in \Sigma$ then $\begin{array}{c}\sigma\\ \swarrow\ \searrow\\ T_1\quad T_2\end{array}$ is a Σ-tree.

In what follows we use only finite, total (binary) X-trees T with the property (decisional trees):

(DT): $\forall\, \{w_1 \ldots w_k,\ w_1 \ldots w_k w_{k+1} \ldots w_n\} \subseteq \operatorname{dom} T$ with $k<n \Rightarrow$

$$T(w_1 \ldots w_k) \cap T(w_1 \ldots w_n) = \emptyset$$

Let $\mathrm{Tr}(X)$ denote this set.

c) <u>Trees in power space</u>

We define a function $\mathbf{C}: \mathrm{Tr}(X) \to Y$ using the recursive definition of $\mathrm{Tr}(X)$:

1) $C(.A)=2^A$

2) $C\left(\begin{array}{c} A \\ / \ \backslash \\ T_1 \quad T_2 \end{array}\right)=(1_A\uparrow A+C(T_1))\cup(2_A\uparrow A+C(T_2))$.

Correctness: The condition (DT) show that in 2) $A\cap \mathrm{dom}\, C(T_i)=\emptyset$ $i=1,2$ and so, the operatorion + is defined.

We can define and a measure function $\lambda:\mathcal{T}r(\mathcal{X})\longrightarrow R_+$:

1) $\lambda(.A)=2^{\mu(A)}$

2) $\lambda\left(\begin{array}{c} A \\ / \ \backslash \\ T_1 \quad T_2 \end{array}\right)=\frac{1}{2}2^{\mu(A)}\lambda(T_1)+\frac{1}{2}2^{\mu(A)}\lambda(T_2)$.

Let denote $\mathcal{T}=C(\mathcal{T}r(\mathcal{X}))$. We say that $M\in\mathcal{T}$ is a (continuous) binary tree and $C^{-1}(M)$ is the set of its representations.

2. λ is a measure in $\mathcal{T}$

The purpose of this section is to show that $\lambda(T_1)=\lambda(T_2)$ if $C(T_1)=C(T_2)$, when μ has the property:

(*) $\forall A\in\mathcal{X}\Rightarrow\{\mu(A')/A'\subseteq A,\ A'\in\mathcal{X}\}$ is a dense set in real interval $[0,\mu(A)]$.

Lemma 2.1: If $l_1,\dots,l_m$ are all terminal paths in $T\in\mathcal{T}r(\mathcal{X})$ and for $i=1,\dots,m$, $l_i=w_1^i\dots w_{n_i}^i$ and $A_k^i=T(w_1^i\dots w_k^i)$, $k=0,\dots,n_i$ then:

$$\lambda(T)=\sum_{i=1}^{m}\frac{1}{2^{n_i}}\,2^{\sum_{k=0}^{n_i}\mu(A_k^i)}$$

Proof: Directly follows from definitions and means that if we apply a transformation S to a tree $T\in\mathcal{T}r(\mathcal{X})$ which do not change the number of terminal paths and their sets of labels, then $\lambda(S(T))=\lambda(T)$. $\square$

<u>Lemma 2.2</u>. For any $A\in\mathfrak{X}$, $\mu(A)>0$, $\varepsilon>\varepsilon'>\cancel{\mu(A)}$, $\varepsilon'<\mu(A)$ and $\alpha\in\{1,2\}$ there exists $f\in\mathbb{Z}^A$ for which $\varepsilon>\mu(f=\alpha)-\mu(f=\bar{\alpha})>\varepsilon'$.

<u>Proof:</u> By the condition $(*)$ about μ, we can find a subset $B\in\mathfrak{X}$ of A with: $\max\left\{0,\frac{1}{2}(\mu(A)-\varepsilon)\right\}<\mu(B)<\frac{1}{2}(\mu(A)-\varepsilon')$.
Let be :

$$f(x)=\begin{cases}\bar{\alpha} & \text{, if } x\in B\\ \alpha & \text{, if } x\in A\\ \text{undefined, otherwise}\end{cases}$$

Of course $\varepsilon>\mu(f=\alpha)-\mu(f=\bar{\alpha})=\mu(A\setminus B)-\mu(B)>\varepsilon'$. $\square$

A notation: if T is $\mathfrak{X}$-tree and $A\in\mathfrak{X}$ then $T\circ A$ is the tree:

$$(T\circ A)(w_1\ldots w_n)=\begin{cases}T(w_1\ldots w_n)\cup A & \text{if } w_1\ldots w_n \text{ is leaf in } T\\ T(w_1\ldots w_n) & \text{elsewhere}\end{cases}$$

<u>Observation 2.3</u>. If $T\in\mathrm{Tr}(\mathfrak{X})$ and $A\cap\mathrm{dom}\,C(T)=\emptyset$ then $T\circ A\in\mathrm{Tr}(\mathfrak{X})$. $\square$

<u>Lemma 2.4</u>. If $T=\begin{array}{c}A\\ \diagup\ \diagdown\\ T_1\quad T_2\end{array}$ has $C(T_1)=C(T_2)$, then:

1) $C(T)=C(T_1\circ A)=C(T_2\circ A)$

2) If $\lambda(T_1)=\lambda(T_2)$ then $\lambda(T)=\lambda(T_1\circ A)=\lambda(T_2\circ A)$.

<u>Proof:</u> 1) Because $C(T_1)=C(T_2)$ we have:

$$C(T)=(1_A{\uparrow}A+C(T_1))\cup(1_A{\downarrow}A+C(T_2))=(1_A{\uparrow}A\cup 1_A{\downarrow}A)+C(T_1)=\mathbb{Z}^A+C(T_1).$$

On the other hand, we can inductively show that:

$$\forall\ T\in \mathcal{T}r(\mathcal{X}),\ A\cap \mathrm{dom}\ C(T)=\emptyset\ \Rightarrow\ C(T\circ A)=2^{A}+C(T)\ .$$

So $C(T)=C(T_1\circ A)=C(T_2\circ A)$.

2) In the same inductive way it is shown that:

$$\forall\ T\in \mathcal{T}r(\mathcal{X}),\ A\cap \mathrm{dom}\ C(T)=\emptyset \Rightarrow \lambda(T\circ A)=2^{\mu(A)}\lambda(T)\ .$$

Follows that: $\lambda(T_1)=\lambda(T_2)\Rightarrow \lambda(T)=\frac{1}{2}(2^{\mu(A)}\lambda(T_1)+2^{\mu(A)}\lambda(T_2))=$

$=2^{\mu(A)}\lambda(T_1)=\lambda(T_1\circ A)$ and identicaly $\lambda(T)=\lambda(T_2\circ A)$. $\square$

Let be $f\in Y$ and $T\in \mathcal{T}r(\mathcal{X})$. The <u>behaviour</u> of f in T is the longest word over $\{1,2\}$, $w_1\dots w_n$ such that: $\forall\ 0\leqslant i<n \Rightarrow w_1\dots w_i\in$ $\in \mathrm{dom}\ T$, $T(w_1\dots w_i)\subseteq \mathrm{dom}\ f$ and $(\mathrm{lr}_f T(w_1\dots w_i)\geqslant 0 \Leftrightarrow w_{i+1}=1)$. We use the notation: $\varphi(f,T)$. Of course, if $T(\Omega)\notin \mathrm{dom}\ f$ then $\varphi(f,T)$ is undefined.

<u>Lemma 2.5</u>. Let be n trees $T_1,\dots,T_n\in \mathcal{T}r(\mathcal{X})$ and $f\in Y$. Then there exists $\bar{f}\in Y$, $\mathrm{dom}\ \bar{f}=\mathrm{dom}\ f$ with the same behaviour like f in every $T_i, i=1,\dots,n$ and if $\varphi(\bar{f},T_i)=w_1^i\dots w_{n_i}^i$ then for any $k\in\{0,\dots,n_i-1\}$: $\mu(T_i(w_1^i\dots w_k^i))>0 \Rightarrow \mathrm{lr}_{\bar{f}}(T_i(w_1^i\dots w_k^i))\neq 0$.

<u>Proof</u>. We may limit to those T_i for which $\varphi(f,T_i)$ is defined. Let they be, with a possible permutation, $T_1,\dots,T_m$. We shall denote $\varphi(f,T_i)=w_1^i\dots w_{n_i}^i$, $n_i\geqslant 1$ and $A_k^i=T_i(w_1^i\dots w_k^i)$, $k=0,\dots n_i-1$ and complete A_k^i, $k=0,\dots,n_i-1$ with $A_{n_i}^i$ to dom f.

Let be $A_{i_1\dots i_m}=A^1_{i_1}\cap\dots\cap A^m_{i_m}$. The matrix $M=(m_{i_1\dots i_m})$ with $m_{i_1\dots i_m}=lr_f(A_{i_1\dots i_m})$, $i_k\in\{0,\dots,n_k\}$, $k=1,\dots,m$, give us the behaviour of f: $w^k_{i_k+1}=1 \Leftrightarrow \Sigma^k_{i_k}(M)=\sum_{j\neq k}\sum_{i_j=0}^{n_j} m_{i_1\dots i_m}\geq 0$ for $i_k=0,\dots,n_k-1$ and $k=1,\dots,m$.

Step 1. By changing same $m_{i_1\dots i_m}$ we can get a matrix $\overline{M}$ with: $\left(\left(\Sigma^k_{i_k}(\overline{M})\geq 0 \Leftrightarrow \Sigma^k_{i_k}(M)\geq 0\right) \text{ and } \Sigma^k_{i_k}(\overline{M})\neq 0 \text{ if } \mu(A^k_{i_k})>0\right)$ for any k, i_k. If one $\Sigma^k_{i_k}(M)=0$ and $\mu(A^k_{i_k})>0$ then there exists $A_{i_1\dots i_m}$ with $\mu(A_{i_1\dots i_m})>0$. Decause $A_{i_1\dots i_m}$ intervens only in $\Sigma^1_{i_1}(M),\dots,\Sigma^m_{i_m}(M)$ if we choose $0<r<\min\left(\{1\}\cup\{|\Sigma^j_{i_j}|/\Sigma^j_{i_j}<0\}\right)$ and increase $m_{i_1\dots i_m}$ with r in the new matrix M^* we have: $\Sigma^j_{i_j}(M^*)\geq 0 \Leftrightarrow \Sigma^j_{i_j}(M)\geq 0$ for any j, i_j and $\Sigma^k_{i_k}(M^*)\neq 0$. By repeating we can get $\overline{M}$.

Step 2. There is $\varepsilon>0$ such that the matrix $\varepsilon\overline{M}$ has $\varepsilon\cdot m_{i_1\dots i_m}<\mu(A'_{i_1\dots i_m})$ for any $i_1,\dots,i_m$ with $\mu(A_{i_1\dots i_m})>0$ (clear). More than: because the functions $\Sigma^k_{i_k}(M)$ are continuous in $m_{i_1\dots i_m}$, for any $i_1,\dots,i_m$ there is an interval $I_{i_1\dots i_m}\subseteq[-a,a]$ $[a_r=\mu(A_{i_1\dots i_m})]$ (which is not a point if $\mu(A_{i_1\dots i_m})>0$) such that for any $M'=(m'_{i_1\dots i_m})$ with $m'_{i_1\dots i_m}\in I_{i_1\dots i_m}$ we have $\left(\left(\Sigma^k_{i_k}(M')\geq 0\Leftrightarrow\Sigma^k_{i_k}(M)\geq 0\right)\text{ and } \Sigma^k_{i_k}(M')\neq 0 \text{ if } \mu(A^k_{i_k})>0\right)$,

$\forall\, k,\ i_k$.

Step 3: f construction: By Lemma 2.2 we can find $\bar{f}$ with $lr_{\bar{f}}(A_{i_1 \ldots i_m}) \in I_{i_1 \ldots i_m}$. So this $\bar{f}$ has the same behaviour like f in every T_k and $lr_{\bar{f}}(A^k_{i_k}) \neq 0$ if $\mu(A^k_{i_k}) > 0$, $\forall\, i_k \in \{0, \ldots, n_k - 1\}$ and $k = 1, \ldots, m$. $\square$

Proposition 2.6. If $T, T' \in \mathfrak{Tr}(\mathfrak{X})$ have $C(T) = C(T')$ then $\lambda(T) = \lambda(T')$.

Proof: Let be $T, T' \in \mathfrak{Tr}(\mathfrak{X})$ with $C(T) = C(T')$. We give a proof by induction after $n = \max\{lg(T),\ lg(T')\}$.

V e r i f i c a t i o n: If n=0 then $T = \cdot A$, $T' = \cdot B$ and because $2^A = C(T) = C(T') = 2^B$ is necessary $A = B$. So $\lambda(T) = 2^{\mu(A)} = 2^{\mu(B)} = \lambda(T')$.

I n d u c t i v e s t e p: A "subinduction" is necessary after the sum $m = lg(T) + lg(T')$. If $m = n$, say $lg(T) = 0$ then $T = \cdot A$ and $T' = \begin{array}{c} A' \\ \diagup \quad \diagdown \\ S_1 \qquad S_2 \end{array}$. Of course $A' \subseteq A$. For any $f \in C(S_1)$ results $1_{A'} + f \in C(T') = C(T) \Rightarrow 2_{A'} + f \in C(T) = C(T') \Rightarrow f \in C(S_1)$. So $C(S_1) = C(S_2)$ and, by induction hypothesis, $\lambda(S_1) = \lambda(S_2)$. By Lemma 2.4 we have $\lambda(T') = 2^{\mu(A')}\lambda(S_1)$. On the other hand $C(S_1) = C(S_2) = C(\cdot A \setminus A')$ and by applying more one induction hypothesis we have $\lambda(S_1) = 2^{\mu(A \setminus A')}$. So $\lambda(T') = 2^{\mu(A')} 2^{\mu(A \setminus A')} = 2^{\mu(A)} = \lambda(T)$.

In the general case ($m > n$) let be $T = \begin{array}{c} A \\ \diagup \quad \diagdown \\ T_1 \qquad T_2 \end{array}$. For any terminal path $l = w_1 \ldots w_n$ in T' let be $p(l) = \min\{i \mid \mu(A_i \cap A) > 0\}$

the first essential intersection with one $A_i = T'(w_1 \ldots w_i)$, $i=0,\ldots,n$. If for one l the subtree with the root $A_{p(l)}$: $V = \begin{smallmatrix} & A_{p(l)} & \\ V_1 & & V_2 \end{smallmatrix}$

has $C(V_1)=C(V_2)$ then $\max\{lg(V_1), lg(V_2)\} < n$ and, by induction hypothesis, $\lambda(V_1)=\lambda(V_2)$. By Lemma 2.4 $\lambda(V)=\lambda(V_1 \circ A_{p(l)})$ and $C(V)=C(V_1 \circ A_{p(l)})$ such that we can replace in T' the subtree V with $V_1 \circ A_{p(l)}$ and obtain T_1 with $C(T_1)=C(T')$ and $\lambda(T_1)=\lambda(T')$. In the same time $lg(T_1) < lg(T')$. By successive application of this transformation to $T_1, \ldots$ we get a tree $\overline{T}$ with the same contain and measure like T', with no greater length and in which on every terminal path l the first nod with an essential intersection with A or is a leaf, or go on with two subtrees with different contains.

I. We shall show that for any terminal path $l = w_1 \ldots w_n$ $A \subseteq A_{p(l)}$ a.e. [almost everywhere: $\mu(A \setminus A_{p(l)})=0$].

Suppose that for one l, $\mu(A \setminus A_{p(l)}) > 0$ and $\overline{T}|_{w_1 \ldots w_{p(l)}} = V^* = \begin{smallmatrix} & A_{p(l)} & \\ V_1 & & V_2 \end{smallmatrix}$ with $C(V_1) \neq C(V_2)$ (if there are not V_1, V_2 then $A_{p(l)}$ is leaf and because $p(l)$ is the first with $\mu(A \cap A_{p(l)}) > 0$ and $\forall f \in C(\overline{T}) \Rightarrow A \subseteq \operatorname{dom} f$, we have $\mu(A \setminus A_{p(l)})=0$). Then there exists $\overline{f} \in C(V_1) \setminus C(V_2)$. [identically may be treat the case $\overline{f} \in C(V_2) \setminus C(V_1)$]

Let be g with $\varphi(g,\overline{T}) = w_1 \ldots w_{p(l)}$ and $\operatorname{dom} g = \bigcup_{k=0}^{p(l)-1} \overline{T}(w_1 \ldots w_k)$. $(p(l)=0 \Rightarrow \operatorname{dom} g = \emptyset)$.

With Lemma 2.5 for $\overline{f}$, $\{V_1, V_2, A \setminus A_{p(l)}\}$ we can choose $f \in C(V_1) \setminus C(V_2)$ with $s = lr_f(A \setminus A_{p(l)}) \neq 0$. There exist (Lemma 2.2)

$h_1 \in 1_{A_{p(1)} \cap A} \uparrow (A_{p(1)} \cap A)$, $h_2 \in 2^{A_{p(1)} \setminus A}$ such that:

$$\begin{cases} h_1 + h_2 \in 1_{A_{p(1)}} \uparrow A_{p(1)} \\ \bar{h}_1 + h_2 \in 1_{A_{p(1)}} \downarrow A_{p(1)} \end{cases} \quad \text{and } \left| \mathrm{lr}_{h_1}(A_{p(1)} \cap A) \right| < |s| \, .$$

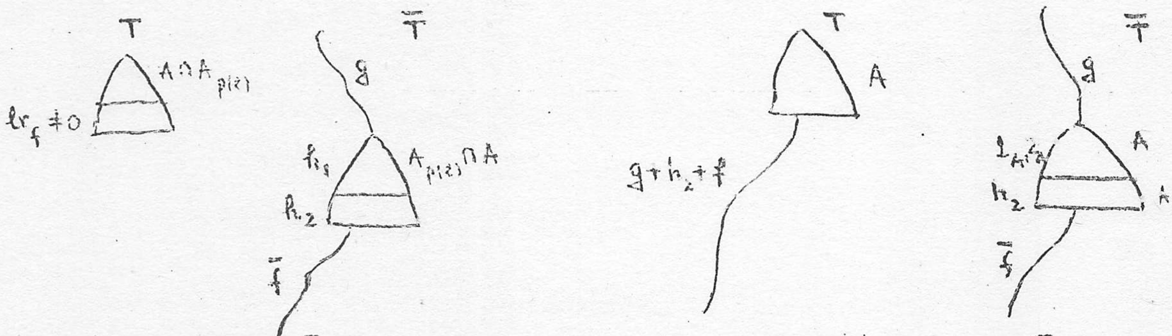
$(\bar{h}_1(x)=1 \Leftrightarrow h(x)=2$ and $\mathrm{dom}\, \bar{h}_1 = \mathrm{dom}\, h_1)$.

Then $g+\bar{h}_1+h_2+f \notin C(\bar{T})$ and, on the other hand ,

$g+h_1+h_2+f \in C(\bar{T}) = C(T)$ and $\mathrm{lr}_{g+h_1+h_2+f}(A) < 0 \Leftrightarrow \mathrm{lr}_{g+\bar{h}_1+h_2+f} < 0$

so $g+\bar{h}_1+h_2+f \in C(T)$. Contradiction !

The conclusion is: for any 1, terminal path in $\bar{T}$, $A \subseteq A_{p(1)}$ a.e., and we can easy do even $A \subseteq A_{p(1)}$ (movind all non--essential intersections of A to $A_{p(1)}$).



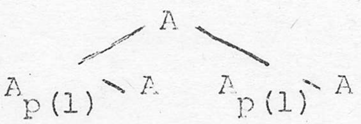
Case I. Case II.

II. For any terminal path $w_1 \ldots w_n$ in T for which $A_{p(1)}$ is not leaf, $A \supseteq A_{p(1)}$ a.e.

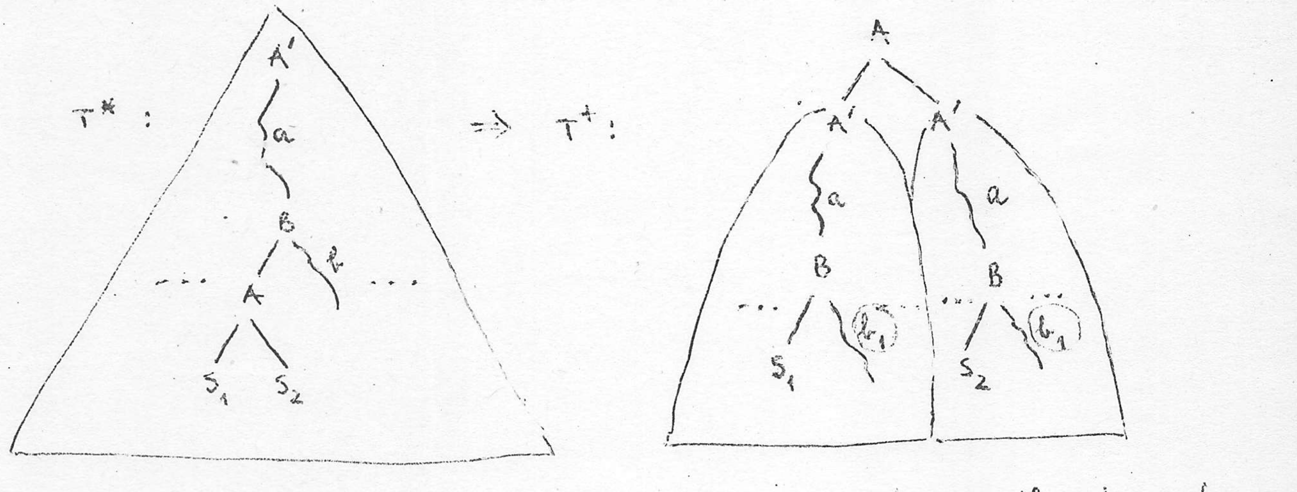
Suppose that for one 1 with $A_{p(1)}$ not leaf, $\mu(A_{p(1)} \setminus A) > 0$. With the precedence functions g and $\bar{f}$ and $h_2 \in 2^{A_{p(1)} \setminus A}$ with $0 \leq \mathrm{lr}_{h_2}(A_{p(1)} \setminus A) < \mu(A)$ we have: $g+1_A+h_2+\bar{f} \in C(\bar{T})$ and $g+2_A+h_2+\bar{f} \notin C(\bar{T})$ therefore $\bar{r} = g+h_2+\bar{f} \in C(T_1) \setminus C(T_2)$. Let be (Lemma 2.5:

$\bar{r}, \{T_1, T_2, \bar{T}, \cdot A_{p(l)} \setminus A\}) r$ with the same behaviour like $\bar{r}$ in T_1, T_2 and $\bar{T}$, dom r=dom $\bar{r}$ and $lr_r(A_{p(l)} \setminus A) > 0$. If we choose $h \in l_A \uparrow A$ with $0 < lr_h(A) < lr_r(A_{p(l)} \setminus A)$ then $h+r \in l_A \uparrow A + r \subseteq C(T)$ and $\bar{h}+r \notin C(T)$ and, on the other hand, because $lr_{\bar{h}+r}(A_{p(l)}) > 0$, $\bar{h}+r \in C(\bar{T})$. Contradiction !

The conclusion is: for any l, terminal path in T for which $A_{p(l)}$ is not leaf, $A \geq A_{p(l)}$ a.e. and, like before, we can do even $A = A_{p(l)}$. We can extend this conclusion to all l so: if for same l, $A_{p(l)}$ is leaf we can replace $A_{p(l)}$ with



. So we have a tree T^* with the same contain and measure like T', with the length no greater than n+1 and A label in every terminal path. We can reorder T^* without change contains or measure (Lema 2.1) to a tree with the root A: We can complete the tree T^* with ϕ label nods such that every A let be not leaf. Then the transformation is:



We show only one step. In the same time the transformation is applied to the all internal nods with label A and so b_1 is not b in which A appears.

Let be this tree $T^+ = \begin{array}{c} A \\ \diagup \ \diagdown \\ U_1 \quad U_2 \end{array}$. Then $C(T_1)=C(U_1)$, $C(T_2)=C(U_2)$

and by induction hypothesis $[\lg(T_1)+\lg(U_1) < \lg(T)+\lg(T')$ and $\max\{\lg(T_1),\lg(U_1)\} \leq n$ and the same for $T_2, U_2]$ $\lambda(T_1)=\lambda(U_1)$, $\lambda(T_2)=\lambda(U_2)$. Therefore $\lambda(T)=\lambda(T^+)$ and because $\lambda(T^+)=\lambda(T')$ we have $\lambda(T)=\lambda(T')$. □

<u>Observations</u>: 2.7) The condition (*) about μ is more then a technical one. For example if $X=(N, \mathcal{P}(N), \mu)$ is the measure space of natural numbers and μ generated by $\mu(\{i\})=1$, $\forall\, i \in N$ then for $A=\{1,2\}$ we have: $1_A \uparrow A = \left\{ \begin{array}{ccc} 1 & \to & 1 \\ 2 & \nearrow & 2 \end{array} , \begin{array}{ccc} 1 & \times & 1 \\ 2 & & 2 \end{array} , \begin{array}{ccc} 1 & \to & 1 \\ 2 & \to & 2 \end{array} \right\}$

and $2_A \uparrow A = \left\{ \begin{array}{ccc} 1 & \searrow & 1 \\ 2 & \to & 2 \end{array} \right\}$. The condition (*) provides that the

set of undecidable functions: $\{ f \in \mathbb{Z}^A / \mu(f=1) = \mu(f=2) \}$ has not weight. □

2.8) The proof of Proposition 2.6 say more than its enunciation. For $M \in C(\mathcal{T}r(\mathcal{X}))$ its representation $T \in \mathcal{T}r(\mathcal{X})$ with $C(T)=M$ is unique leaving the transformation: $A \cup B \leftrightarrow \begin{array}{c} A \\ \diagup \ \diagdown \\ B \quad B \end{array}$, reordering of trees and zero sets, aside. □

2.9) Let be for $T \in \mathcal{T}r(\mathcal{X})$ and 1 terminal path in T, $C(1)$ and $\lambda(1)$ the restriction of C and λ to 1. If $1=w_1 \ldots w_n$ let denote $\mu(1)=\sum_{k=0}^{n} \mu(T(w_1 \ldots w_k))$. [We do not write and the variable T.] By using the same technique (but more simpl) may be shown that λ is right defined in $\mathcal{Y} = \{ C(1_1) \cup \ldots \cup C(1_m) / 1_1, \ldots, 1_m$ terminal paths in one $T \in \mathcal{T}r(\mathcal{X}) \}$.

Of course $\mathcal{T} \subseteq \mathcal{Y}$. Because $\mathcal{Y}$, in general, is not close to intersection also it's not close to general union. It is close to arbitrary (finite) application of operation + . When $\cup$ and + are defined (in $\mathcal{Y}$) we have:

1) $\lambda(N\cup M)=\lambda(N)+\lambda(M)$, $M\cap N=\emptyset$.

2) $\lambda(N+M)=\lambda(N)\cdot\lambda(M)$. $\square$

3. A geometrical property of $(Y,\mathcal{T})$

In the case when μ is contably additive ($\mathcal{X}$ is σ-algebra and μ is σ-measure) the condition (*) become:

(*) $\forall A\in\mathcal{X},\ \{\mu(A')\ /\ A'\subseteq A,\ A'\in\mathcal{X}\}=[0,\ \mu(A)]$.

[The reason is: if $c\in[0,\ \mu(A)]$ by using (*) there exists a decreasing sequence $A\supseteq A_1\supseteq A_2$ ... with $\mu(A_n)\in(c,c+\frac{1}{n})$. This means: $\mu(\bigcap_{n\geq 1} A_n)=c$]. Also we suppose that:

(**) $\mu(x)=+\infty$.

Is useful, also, a "volumic" measure. If $M\in\mathcal{Y}$ then:

$$m(M)=\int_M \mu(\text{dom } f)d\,\lambda(f)$$

(In fact, the integral is here a finite sum: $m(M)=\sum_{\ell \text{ terminal path in } T}\lambda(\ell)\mu(\ell)$ with $C(T)=M$).

Obs. 3.1: $m_1, m_2, d>0 \Rightarrow (m_1+m_2)\log_2\dfrac{m_1+m_2}{m_1+m_2 2^{-d}}<d\,m_2$.

Proof: Let put $p=\dfrac{m_1}{m_1+m_2}$, $q=\dfrac{m_2}{m_1+m_2}$.

The inequality is:

$\log_2 \frac{1}{p+q2^{-d}} < dq$. It's clear its equivalence with $\frac{1}{p+q2^{-d}} < 2^{dq} \Leftrightarrow$

$\Leftrightarrow (1+(2^d-1))^p < 1+p(2^d-1)$, which is well known Bernoulli inequality. □

Obs. 3.2: $m_1, m_2 > 0$, $0<c<c+d$ and ε given by $m_1 2^{\varepsilon} + m_2 2^{-(d-\varepsilon)} = m_1+m_2$ then: $(c+\varepsilon)m_1 2^{\varepsilon} + (c+d-(d-\varepsilon))m_2 2^{-(d-\varepsilon)} < cm_1 + (c+d)m_2$.

Proof: How $\varepsilon = \log_2 \frac{m_1+m_2}{m_1+m_2 2^{-d}}$ the inequality is:

$(c+\varepsilon)(m_1+m_2) < cm_1+(c+d)m_2$, in fact the precedence: $(m_1+m_2)\varepsilon < dm_2$. □

Obs.3.3. Let be $T \in Tr(X)$ and $0 \le a \le \mu(l_i) \le b$ for all l_i terminal path in T. Then: $\lambda(T) \in [2^a, 2^b]$.

Proof: by induction after $n=lg(T)$.

If $n=0$ then $T = \cdot A$ and $\mu(A) \in [a,b]$. Follows that: $\lambda(T) = 2^{\mu(A)} \in [2^a, 2^b]$. In the general case, if $T = \begin{array}{c} A \\ \diagup \; \diagdown \\ T_1 \quad T_2 \end{array}$ then by induction hypothesis $\lambda(T_1), \lambda(T_2) \in [2^{a-\mu(A)}, 2^{b-\mu(A)}]$ and so $\lambda(T) \in [2^a, 2^b]$. □

Obs. 3.4. For every $T \in Tr(X)$ and $r>0$ there is $T_r \in Tr(X)$ with the same measures: $\lambda(T_r) = \lambda(T)$, $m(T_r) = m(T)$ and such that:

$$\sum_{w_1 \dots w_n \in \mathrm{dom}\, T_r} \mu(T_r(w_1 \dots w_n)) < r .$$

and is not leaf

Proof: The transformation $\begin{array}{c} A \\ \diagup \; \diagdown \\ S_1 \quad S_2 \end{array} \rightsquigarrow \begin{array}{c} A_1 \\ \diagup \; \diagdown \\ S_1 \cup A_2 \quad S_2 \cup A_2 \end{array}$

where $A_1 \cup A_2 = A$, $A_1 \cap A_2 = \emptyset$ and $\mu(A_1) > 0$ if $\mu(A) > 0$, preserves the measures λ and m. By applying this transformation to all interval nods is possibly to get a tree with total measure of internal nods less than r. $\square$

<u>Proposition 3.5.</u> ("The sphere has minimal volum for a given surface")

$$\inf_{\substack{M\in\mathcal{T}\\ \lambda(M)=a}} m(M)=a\cdot\log_2 a \quad , \text{ where } \quad a\geqslant 1.$$

<u>Proof.</u> There is $A\in\mathfrak{X}$ (**+*) with $\mu(A)=\log_2 a$. For this $m(2^A)=a\cdot\log_2 a$. Optimalness: Suppose that $T\in\mathfrak{Tr}(\mathfrak{X})$ with $\lambda(T)=a$, has two terminal paths l_1, l_2 with $\mu(l_1)<\mu(l_2)$. Let be $m_1=\lambda(l_1)$, $m_2=\lambda(l_2)$, $c=\mu(l_1)$, $c+d=\mu(l_2)$ in Obs. 3.2. Then there is $\varepsilon>0$ such that $m_1 2^{\varepsilon}+m_2 2^{-(d-\varepsilon)}=m_1+m_2$ and $(c+\varepsilon)m_1 2^{\varepsilon}+(c+d-(d-\varepsilon))m_2 2^{-(d-\varepsilon)}<$ $< m_1 c+m_2(c+d)$. Let be (by 3.4). T_r with $r<c$. Replace the final labels from l_1 , say A and from l_2 , say B by two labels A^+_{ε} , $B^-_{d-\varepsilon}$ (by (**) there exist A^+_{ε} , $B^-_{d-\varepsilon}$ that preserve the property DT) with $\mu(A^+_{\varepsilon})=\mu(A)+\varepsilon$. $\mu(B^-_{d-\varepsilon})=\mu(B)-(d-\varepsilon)$ (it's positive: $\mu(B)>\mu(l_2)-r>(c+d)-c=d$), the new tree $\overline{T}_r$ has: $\lambda(\overline{T}_r)=a$ and (by 3.2) $m(\overline{T}_r)<m(T_r)=m(T)$. The observation that $\overline{T}_r$ has $\mu(l_1)=\mu(l_2)$ is sufficient to show that after a (finite) number of this transformations we get a balanced tree T^* (all terminal paths have the same measure μ). This has (by $3.3+\lambda(T^*)=a$) $\mu(l_i)=\log_2 a$ for same l_i=terminal path in T^*, and so $m(C(T^*))=$ $=a\cdot\log_2 a<m(T)$. $\square$

4. Connection with the (discrete) entropy

Let be $p_1,\dots,p_n > 0$ with $\sum_1^n p_i = 1$ and $H(p_1,\dots,p_n) =$

$$= -\sum_1^n p_i \log_2 p_i \; .$$

Lemma 4.1. For any $A \in \mathfrak{X}$, $0 \le r \le 2^{\mu(A)}$ and $\varepsilon > 0$ there exists $B^\varepsilon \subseteq A$, $|\mu(B^\varepsilon) - \mu(A)| < \varepsilon$ such that $2^{B^\varepsilon} = M_1 \cup M_2$ with $\lambda(M_1) = r$.

Proof: Every $\mathcal{Y} \ni M \subseteq 2^C$ has the measure (see and 2.1)

$$\lambda(M) = \frac{k}{2^n}\, 2^{\mu(C)} \quad \text{with } n \in N \text{ and } k \in \{1,2,3\dots,2^n\}.$$

Because $\left\{ r\frac{2^n}{k} \,/\, n \in N,\ k \in \{1,2,3,\dots,2^n\} \right\}$ is a dense set in $[r,+\infty)$ then $\left\{ \log_2 r\frac{2^n}{k} \ \ n,k\dots \right\}$ is dense in $[\log_2 r, +\infty) \ni \mu(A)$

Therefore there are $n_o, k_o \in \{1,\dots,2^{n_o}\}$ with $\left| \log_2 r\frac{2^{n_o}}{k_o} - \mu(A) \right| < \varepsilon$.

Let be $B^\varepsilon \subseteq A$ with $\mu(B^\varepsilon) = \log_2 \frac{r2^{n_o}}{k_o}$. Then $\lambda(2^{B^\varepsilon}) = r\frac{2^{n_o}}{k_o}$ and, by the initial observation, admits a subset $M_1 \subseteq 2^{B^\varepsilon}$ with

$$\lambda(M_1) = \frac{k_o}{2^{n_o}}\left(r\frac{2^{n_o}}{k_o}\right) = r \; . \quad \Box$$

Theorem 4.2. Let be for $M = M_1 \cup \dots \cup M_n$ (disjoint union)

$\varphi : M \to \mathbb{R}$ given by: $\varphi(f) = p_i \iff f \in M_i$.

$$H(p_1,\dots,p_n) = \inf_{\substack{M = M_1 \cup \dots \cup M_n \in \mathcal{T} \\ \mathcal{Y} \ni M_i \text{ disjoint sets} \\ \lambda(M_i) = 1}} \int_M \varphi(f)\, \mu(\mathrm{dom}\, f)\, d\lambda f .$$

Proof. Let be k with $\log_2 k > \max\{-\log_2 p_i / i=1\ldots n\}$ and $M=M_1\cup\ldots\cup M_n \in \mathfrak{T}$ with M_i disjoint sets and $\lambda(M_i)=1$. Then the tree $\overline{M}=(M_1+2^{A_1})\cup\ldots\cup(M_n+2^{A_n})$ with $\mu(A_i)=\log_2 k+\log_2 p_i > 0$ and so that $\overline{M}\in\mathfrak{T}$, has: $\lambda(\overline{M})=\sum_{i=1}^{n} 1\cdot 2^{\log_2 k+\log_2 p_i}=k$. By 3.5 we have

$m(\overline{M})=\sum_{i=1}^{n}\int_{M_i}\mu(\mathrm{dom}\ g)d\lambda(g)\geq k\cdot\log_2 k$. Suppose that $T\in\mathfrak{Tr}(\mathfrak{X})$ has $C(T)=M$ and every $M_i=\bigcup_{k=0}^{n_i} C(1_k^i)$, 1_k^i terminal paths in T. The union of according A_i to leaves give us a tree $\overline{T}$: $C(\overline{T})=\overline{M}$ and $\overline{M}_i=\bigcup_{k=0}^{n_i} C(\overline{1}_k^{\,i})$ ($\overline{1}_k^{\,i}$ has the same form like 1_k^i but in other tree).

Let be $f\in C(1_k^i)$, $\overline{f}\in C(\overline{1}_k^{\,i})$. Then $\int_{C(\overline{1}_k^{\,i})}\mu(\mathrm{dom}\ g)d\lambda(g)=\lambda(\overline{1}_k^{\,i})\mu(\mathrm{dom}\ \overline{f})=$

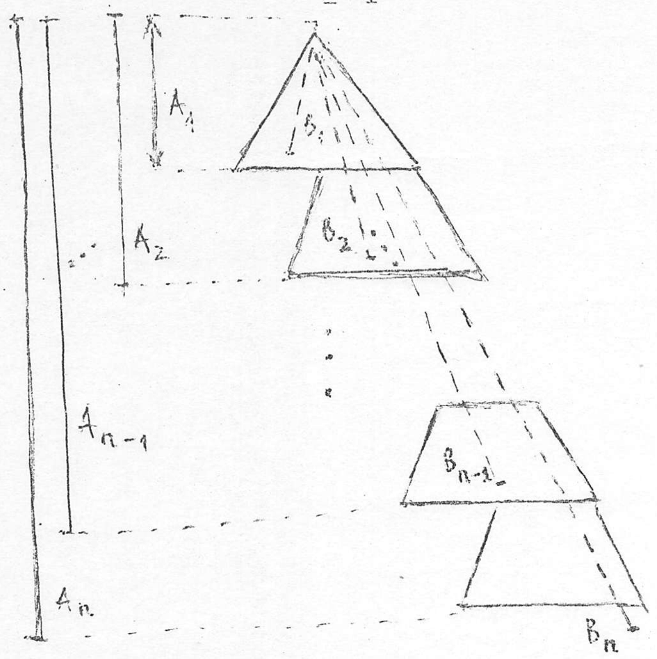
$=kp_i\lambda(1_k^i)[\mu(\mathrm{dom}\ f)+\log_2 k+\log_2 p_i]=k\int_{C(1_k^i)}p_i\mu(\mathrm{dom}\ g)d\lambda(g)+\lambda(1_k^i)kp_i\cdot$ $\cdot\log_2 kp_i$. So $m(M_i)=k\int_{M_i}p_i\mu(\mathrm{dom}\ g)d\lambda(g)+kp_i\log_2 kp_i$ and $m(\overline{M})=$

$$=k\int_M \varphi(g)\mu(\mathrm{dom}\ g)d\lambda(g)+\sum_{i=1}^{n}kp_i\log_2 k+\sum_{i=1}^{n}kp_i\log_2 p_i\geq k\log_2 k \ .$$

Therefore: $\int_M \varphi(g)\mu(\mathrm{dom}\ g)d\lambda(g)\geq -\sum_{i=1}^{n}p_i\log_2 p_i=H(p_1,\ldots,p_n)$.

The bound is effective. Let be $-\log_2 p_1\leq\ldots\leq-\log_2 p_n$ and $A_1\subseteq\ldots\subseteq A_n$ with $\mu(A_i)=-\log_2 p_i$. By 4.1. for any small $\varepsilon>0$ there are $B_i^\varepsilon\subseteq A_i$, $|\mu(A_i)-\mu(B_i^\varepsilon)|<\varepsilon$, $i=1\ldots n-1$ and B_n^ε , which give a tree $M^\varepsilon=M_1^\varepsilon\cup\ldots\cup M_n^\varepsilon$ with $\lambda(M_i^\varepsilon)=1$, $\forall\ i=1\ldots n$.
When $\varepsilon\to 0$ of course $\mu(B_i^\varepsilon)\to\mu(A_i)$, $\forall\ i=1,\ldots,n-1$ but more: and $\mu(B_n^\varepsilon)\to\mu(A_n)$. So :

$$\int_{M^\varepsilon} \varphi(f)\,\mu(\mathrm{dom}\ f)\,d\lambda(f)=\sum_{i=1}^{n} p_i\,\mu(B_i^{\varepsilon})\cdot 1 \xrightarrow{\varepsilon\to 0} \sum_{i=1}^{n} p_i\,\mu(A_i) = -\sum_{i=1}^{n} p_i \log_2 p_i \ .$$



$\square$

Corollary 4.3. (Gibb's theorem): $H(p_1,\dots,p_n) \leq \sum_{i=1}^{n} p_i(-\log_2 q_i)$ when $q_1,q_2,\dots,q_n > 0$ and $\sum_{i=1}^{n} q_i = 1$.

Proof: One other tree has the value $\int_M \varphi(\dots$ less than the minimal value . $\square$

Better understanding list

1 P.R.Halmos: Measure theory, Springer-Verlag, 1974.

2 J.W.Thatcher, E.G.Wagner, J.B.Wright: Notes on algebraic fundamentals for theoretical computer science, Math. Centre Thacts 109(1979), 83-163.

3 S.Watanabe: Knowing and guessing, John-Wiley and Sons, 1969.