INFORMATIQUE THÉORIQUE ET APPLICATIONS

U. TRIER

On the average number of registers needed to evaluate a special class of backtrack trees

Informatique théorique et applications, tome 27, nº 6 (1993), p. 541-554

<http://www.numdam.org/item?id=ITA_1993__27_6_541_0>

Informatique théorique et Applications/Theoretical Informatics and Applications
(vol. 27, n° 6, 1993, p. 541 à 554)

ON THE AVERAGE NUMBER OF REGISTERS NEEDED TO EVALUATE A SPECIAL CLASS OF BACKTRACK TREES (*)

by U. TRIER (1)

Communicated by J. BERSTEL

Abstract. – *We derive a lower bound for the average number of registers* $R_p(h)$ *needed to evaluate the family* $\mathcal{F}_p(h)$ *of non-uniformly distributed binary trees introduced by P. W. Purdom. This family consists of binary trees of height less than or equal to h. Based on a parameter* $p \in [0, 1]$, *the probability of a particular tree* $T \in \mathcal{F}_p(h)$ *is given by a recursively defined function. We show that* $R_p(h)$ *is smaller than 2, for* $0 \leq p \leq 1/2$, *and that, for* $1/2 < p < 1$, *it grows up to at least* $O(\log(h))$. *Near* $p = 1$, $R_p(h)$ *jumps to* $h+1$.

Résumé. – *Nous décrivons une borne inférieure sur le nombre moyen* $R_p(h)$ *de registres nécessaires pour évaluer la famille* $\mathcal{F}_p(h)$ *d'arbres binaires, distribués de façon non uniforme, introduite par P. W. Purdom. Cette famille est constituée d'arbres binaires de hauteur au plus h. Pour un paramètre* $p \in [0, 1]$ *donné, la probabilité d'un arbre* $T \in \mathcal{F}_p(h)$ *est donnée par une fonction définie récursivement. Nous montrons que* $R_p(h)$ *est inférieur à 2 pour* $0 \leq p \leq 1/2$, *et que, pour* $1/2 < p < 1$, *elle dépasse* $O(\log h)$. *Près de* $p = 1$, $R_p(h)$ *saute à* $h+1$.

1. INTRODUCTION AND BASIC DEFINITIONS

Let $T = (I, L, r)$ be an extended binary tree [7, p. 399] with the set of internal nodes I (nodes of degree 2), the nonempty set of leaves L (nodes of degree 0) and the root r. The one node tree is denoted by "$\square$". For any two nodes $u, v \in I \cup L$, let $d(u, v)$ be the "distance" from u to v, which is defined as the length of the shortest path from u to v (= number of nodes on the path minus 1). A node $w \in I \cup L$ with $d(r, w) = l$ is said to be at level l. We say that the tree T has height h, if the maximum level of a node in the tree is equal to h.

(*) Received March 1992, accepted may 1993.

(1) Johann Wolfgang Goethe-Universität Fachbereich Informatik (20) D-60054 Frankfurt/Main.

In [10], a family of binary backtrack trees $\mathscr{F}_p(h)$, $p \in [0, 1]$, $h \in \mathbb{N}_0 := \mathbb{N} \cup \{0\}$, has been introduced, which consists of all extended binary trees of height less than or equal to h, and where each tree $T \in \mathscr{F}_p(h)$ is associated with a nonnegative real number $\varphi_{p,h}(T)$, recursively defined by:

a) $\varphi_{p,h}(\square) := p\,\delta_{h,0} + 1 - p$, $h \geqq 0$;

b) if the root $r(T)$ has the two subtrees T_1, $T_2 \in \mathscr{F}_p(h-1)$, then

$$\varphi_{p,h}(T) := p\,\varphi_{p,h-1}(T_1)\,\varphi_{p,h-1}(T_2),\ h \geqq 1.$$

The numbers $\varphi_{p,h}(T)$, $T \in \mathscr{F}_p(h)$, define a probability distribution on $\mathscr{F}_p(h)$ ([5, Lemma 1]) In [5], the behaviour of additive weights ([4]) defined on $\mathscr{F}_p(h)$ has been investigated. In [11], these results have been generalized to simply generated trees ([8]) of bounded arity. Lower and upper bounds for the average stacksize over the family $\mathscr{F}_p(h)$ are derived in [6].

In this paper, we deal with the number of registers needed to evaluate a tree $T \in \mathscr{F}_p(h)$ optimally. In [1, 3], this register-function hs been investigated for the family of extended binary trees with n leaves, in which all trees are equally likely. For a given extended binary tree T, the register-function $R(T)$ is recursively defined as follows:

$$R(T) := \ \text{if } T = \text{“}\square\text{''} \text{ then } 1 \ \text{else} \ \text{if } R(T_1) = R(T_2) \text{ then } R(T_1) + 1 \ \text{else } \mathrm{MAX}(R(T_1), R(T_2)),$$

where T_1 and T_2 are the two subtrees of the root of T.

Figure 1 shows all trees $T \in \mathscr{F}_p(h)$, $h \leqq 2$, together with their probabilities $\varphi_{p,h}(T)$ and their (encircled) number of required registers $R(T)$.
The computation of the average number of registers $\underline{R}_p(h)$ over the family $\mathscr{F}_p(h)$ leads to a nonlinear double-recursive recurrence (Lemma 1). We are not able to derive its exact solution, but it is possible to find a nontrivial lower bound for $\underline{R}_p(h)$, nontrivial in the sense that it is not a constant. This will be done in the next section. In the sequel we shall omit the subscript p to simplify the notation whenever possible.

2. THE AVERAGE NUMBER OF REGISTERS

Let $\mathscr{F}^{(h)}(r) := \{T \in \mathscr{F}(h) \mid R(T) \leqq r\}$ be the set of all trees $T \in \mathscr{F}(h)$ that need r or less registers. The probability $g_{r,h}(p)$ (or short $g_{r,h}$) that a tree

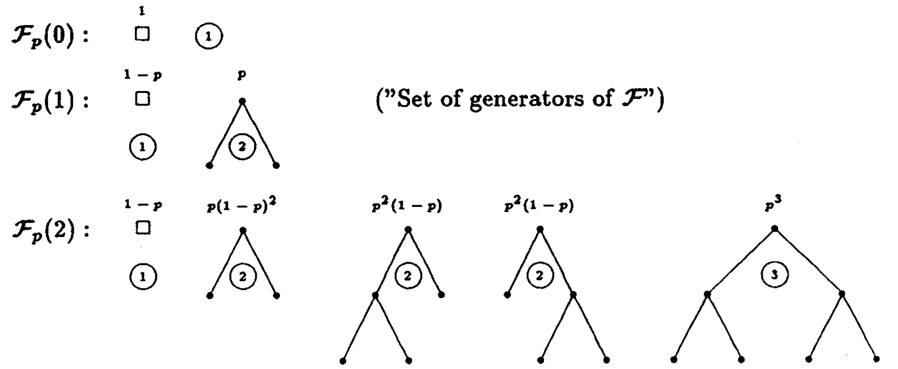

Figure 1. — **All trees $T \in \mathcal{F}_p(h)$ for $h \leqq 2$ together with $\varphi_{p,h}(T)$ and $R(T)$ (encircled).**

$T \in \mathcal{F}(h)$ requires at most r registers is defined by:

$$g_{r,h} := \text{prob}[T \in \mathcal{F}^{(h)}(r)] = \sum_{T \in \mathcal{F}^{(h)}(r)} \varphi_h(T).$$

The average number of registers over the family $\mathcal{F}(h)$ is then given by:

$$\underline{R}(h) = \sum_{1 \leqq r \leqq h+1} r[g_{r,h} - g_{r-1,h}] = 1 + h - \sum_{1 \leqq r \leqq h} g_{r,h}, \quad h \geqq 0, \tag{1}$$

with $g_{0,h} = 0$ and $g_{h+1,h} = 1$, $h \geqq 0$, because a binary tree of height h requires at least 1 and at most $h+1$ registers. Hence, an upper (lower) bound for $g_{r,h}$ leads to a lower (upper) bound for $\underline{R}(h)$.

LEMMA 1: *Let* $(r, h) \in \mathbb{N} \times \mathbb{N}_0$. *Then*

a) $g_{r,0} = 1$, $r \in \mathbb{N}$;

b) $g_{1,h} = p\,\delta_{h,0} + 1 - p$, $h \geqq 0$;

c) $g_{r,h} = 1 - p + 2p\,g_{r,h-1}\,g_{r-1,h-1} - p\,g_{r-1,h-1}^2$, $r \geqq 2$, $h \geqq 1$.

Proof: The lemma follows from the observation that a tree $T \in \mathcal{F}^{(h)}(r)$ is either the one node tree "□", or it has two subtrees T_1, $T_2 \in \mathcal{F}(h-1)$, both of them needing at most $r-1$ registers or one of them needing exactly r registers while the other requires at most $r-1$ registers. □

Some basic observations about $g_{r,h}$ are summarized in the following lemma.

LEMMA 2: *The probability* $g_{r,h}$, *that a tree* $T \in \mathcal{F}(h)$ *requires at most r registers has the following properties:*

a) $g_{r,h}=1,\ r \geqq h+1$; *b)* $g_{h,h}=1-p^{2^h-1}, h \in \mathbb{N}$;

c) $g_{r,h} \leqq g_{r+1,h},\ (r, h) \in \mathbb{N} \times \mathbb{N}_0$; *d)* $g_{r,h} \geqq g_{r,h+1},\ (r, h) \in \mathbb{N} \times \mathbb{N}_0$;

e) $g_{r,h} \leqq g_{r+1,h+1},\ (r, h) \in \mathbb{N} \times \mathbb{N}_0$;

Proof: Part *a)* of the lemma follows from the fact, that a binary tree of height less than or equal to h requires at most $h+1$ registers. Part *b)* is true, because the only tree of height less than or equal to h, which needs exactly $h+1$ registers is the complete binary tree of height h. It has the probability p^{2^h-1}. Part *c)* is trivial. The parts *d)* and *e)* of the lemma can be proved simultaneously by induction. For this purpose let $\Delta_{r,h}:=g_{r,h}-g_{r,h+1}$ and $\nabla_{r,h}:=g_{r+1,h+1}-g_{r,h}$. We have $\Delta_{1,0}=p$, $\nabla_{1,0}=0$, $\Delta_{1,h+1}=0$, $h \geqq 1$ and $\nabla_{1,h}=\nabla_{1,h+1}=0$, $h \geqq 1$. Furthermore $\Delta_{2,0}=0$ and $\nabla_{2,0}=0$. Now assume that $\Delta_{s,l} \geqq 0$ and $\nabla_{s,l} \geqq 0$, $(s, l) \in ([1:r-1] \times \mathbb{N}_0) \cup (\{r\} \times [0:h-1])$. By application of the rules of finite calculus [2, p. 55] we find

$$\begin{aligned}\Delta_{r,h}=2p\,(g_{r-1,h-1}(\Delta_{r,h-1}-\Delta_{r-1,h-1})+(g_{r,h}-g_{r-1,h})\,\Delta_{r-1,h-1})\\ +p\,(g_{r-1,h-1}\,\Delta_{r-1,h-1}+g_{r-1,h}\,\Delta_{r-1,h-1}),\quad (2)\end{aligned}$$

$$\begin{aligned}\nabla_{r,h}=2p\,(g_{r-1,h-1}(\nabla_{r,h-1}-\nabla_{r-1,h-1})+(g_{r+1,h}-g_{r,h})\,\nabla_{r-1,h-1})\\ +p\,(g_{r-1,h-1}\,\nabla_{r-1,h-1}+g_{r,h}\,\nabla_{r-1,h-1}).\quad (3).\end{aligned}$$

Multiplying out the right hand side of (2) we obtain

$$\begin{aligned}\Delta_{r,h}&=-pg_{r-1,h}\,\Delta_{r-1,h-1}-p\underbrace{g_{r-1,h-1}}_{\leqq g_{r,h}}\Delta_{r-1,h-1}\\ &\quad+2pg_{r-1,h-1}\,\Delta_{r,h-1}+2pg_{r,h}\,\Delta_{r-1,h-1}\\ &\geqq -pg_{r-1,h}\,\Delta_{r-1,h-1}+pg_{r,h}\,\Delta_{r-1,h-1}+2pg_{r-1,h-1}\,\Delta_{r,h-1}\\ &=p\,\Delta_{r-1,h-1}(g_{r,h}-g_{r-1,h})+2pg_{r-1,h-1}\,\Delta_{r,h-1} \geqq 0.\end{aligned}$$

By an analogous computation we find for (3)

$$\begin{aligned}\nabla_{r,h}&=-pg_{r,h}\,\nabla_{r-1,h-1}-p\underbrace{g_{r-1,h-1}}_{\leqq g_{r,h}}\nabla_{r-1,h-1}\\ &\quad+2pg_{r-1,h-1}\,\nabla_{r,h-1}+2pg_{r+1,h}\,\nabla_{r-1,h-1}\\ &\geqq -2pg_{r,h}\,\nabla_{r-1,h-1}+2pg_{r+1,h}\,\nabla_{r-1,h-1}+2pg_{r-1,h-1}\,\Delta_{r,h-1}\\ &=2\mathrm{p}\,(g_{r+1,h}-g_{r,h})\,\nabla_{r-1,h-1}+2pg_{r-1,h-1}\,\nabla_{r,h-1} \geqq 0. \quad \square\end{aligned}$$

Based on the Lemmata 1 and 2 it is a simple matter to derive two additional relations, that are worth to be stated, because one of them shows a relation to the stacksize problem ([6]) of our family $\mathscr{F}(h)$, and the other serves as an introduction to Lemma 4.

LEMMA 3: *The probability* $g_{r,h}$, *that a tree* $T \in \mathscr{F}(h)$ *requires at most* r *registers, satisfies the following relations:*

a) $g_{r,h} \leqq 1-p+pg_{r,h-1}^2$, $(r, h) \in \mathbb{N} \times \mathbb{N}$; (4)

b) $g_{r,h} \geqq 1-p+pg_{r,h}g_{r-1,h-1}$, $(r, h) \in \mathbb{N} \times \mathbb{N}$. □

Unfortunately both relations given in Lemma 3 are note very helpful at the first glance, because there is no hope to solve the recurrence corresponding to the first relation (4), and because the second relation leads to the recurrence

$$g_{r,h} \geqq F_r(p) := \frac{1-p}{1-pF_{r-1}(p)}, \quad r \geqq 2,$$

where $F_1(p) := 1-p$. This recurrence can be found in [6, p. 6], where it appears in the computation of the lower bound for the corresponding stack-size probability. It has the following solution:

if $p \in [0, 1] \setminus \{1/2\}$ then $g_{r,h} \geqq (1-p)(p^r-(1-p)^r)/(p^{r+1}-(1-p)^{r+1})$, $(r, h) \in \mathbb{N} \times \mathbb{N}_0$;

if $p = \{1/2\}$ then $g_{r,h} \geqq (r/(r+1))$, $(r, h) \in \mathbb{N} \times \mathbb{N}_0$.

The corresponding upper bound for the average number of registers $\underline{R}(h)$ over the family $\mathscr{F}(h)$ satisfies

$$\underline{R}(h) \begin{cases} < \dfrac{1-2p}{1-p} \mathrm{C}\left(\dfrac{p}{1-p}\right), & \text{if } 0 \leqq p < 1/2, \\ \leqq \mathrm{H}_h, & \text{if } p = 1/2, \\ < \dfrac{2p-1}{p} h + \dfrac{2p-1}{p}\left[1 + \dfrac{1-p}{p} \mathrm{C}\left(\dfrac{1-p}{p}\right)\right], & \text{if } 1/2 < p \leqq 1, \end{cases}$$

where $\mathrm{H}_n := \sum_{1 \leqq i \leqq n} 1/i$ denotes the n-th harmonic number, and $\mathrm{C}(x) := \sum_{r \geqq 0} d(r+1)x^r$, $|x| < 1$, where $d(n)$ denotes the number of positive divisors of $n \in \mathbb{N}$ (*see* [6, Theorem 2]).

This is a trivial upper bound, because the number of registers required is always smaller than or equal to the stacksize. However it is possible to improve this upper bound for $p = 1/2$.

THEOREM 1: *Let* $p=1/2$. *Then* $g_{r,h} \geqq C_r := 1-(1/2)^r$, $(r, h) \in \mathbb{N} \times \mathbb{N}_0$, *and the average number of registers* $\underline{R}(h)$ *over the family* $\mathscr{F}(h)$ *is smaller than* 2.

Proof: Let $p=1/2$. We have $1=g_{1,0}>g_{1,h}=1/2 \geqq C_1$, $h \in \mathbb{N}$, and $g_{2,0}=g_{2,1} \geqq 1-1/4=C_2$. By Lemma 1, and by our induction basis, we obtain

$$1-g_{r,h} \leqq p-2pg_{r,h-1}C_{r-1}+pC_{r-1}^2=\frac{1}{2}(1-C_{r-1})^2+C_{r-1}(1-g_{r,h-1}).$$

Introducing $M_{r,h-r} := 1-g_{r,h}$, we have to solve

$$M_{r,h-r}=\frac{1}{2}(1-C_{r-1})^2+C_{r-1}M_{r,h-r-1},\ r \geqq 1,\ h \geqq r.$$

If we now define the generating function $M_r(z) := \sum_{n \geqq 0} M_{r,n} z^n$, with $M_{r,0}=1-g_{r,r}=p^{2^r-1}$, we find

$$M_r(z)=p^{2^r-1}+p(1-C_{r-1})^2 \sum_{n \geqq 1} z^n+2pC_{r-1}zM_r(z),$$

and this leads to

$$M_r(z)=p^{2^r-1} \sum_{n \geqq 0} (2pC_{r-1})^n z^n+p(1-C_{r-1})^2 \sum_{n \geqq 1} z^n \sum_{0 \leqq k \leqq n-1} (2pC_{r-1})^k,$$

from which follows

$$M_{r,n}=\frac{1}{2}(1-C_{r-1})+C_{r-1}^n\left[\left(\frac{1}{2}\right)^{2^r-1}-\frac{1}{2}(1-C_{r-1})\right] \leqq \left(\frac{1}{2}\right)^r,\ (r, n) \in \mathbb{N} \times \mathbb{N}_0.$$

The theorem follows by plugging this formula into formula (1). □

Now let us return to the lower bound for the average number of registers over he family $\mathscr{F}(h)$. As already mentioned, there is only a little chance to solve the inhomogeneous recurrence related to (4), although it looks very simple. However, as we shall see, we are able to derive an easier and also better relation than the one defined in (4).

LEMMA 4: *Let* $l_{r,h}(p)$, $(r, h) \in \mathbb{N} \times \mathbb{N}_0$, (*or short* $l_{r,h}$), *be defined as follows:*

$$l_{r,r} := 1-p^{2^r-1},\ l_{r,r+1} := 1-p+pl_{r,r}^2=1-p+p(1-p^{2^r-1})^2, \tag{5}$$

$$l_{r,h} := \frac{1-p}{1-pl_{r,h-2}},\ r \geqq 2,\ h \geqq r+2.$$

Then the probability $g_{r,h,}$ *that a tree* $T \in \mathcal{F}(h)$ *requires at most r registers satisfies the relation* $g_{r,h} \leqq l_{r,h}$, $r, h \geqq 2$.

Proof: First we have to note, that there is a relationship between the following two functional equations:

$$f(p) = 1 - p + pf^2(p) \qquad \text{and} \qquad f(p) = \frac{1-p}{1-pf(p)}.$$

Both equations have the same (nontrivial) solution $f(p) = 1/p - 1$. Now let $p_r \in ((1/2), 1)$ be the (unique) solution of the equation $g_{r,r} = 1 - p^{2^r - 1} = f(p)$. For $0 \leqq p \leqq p_r$, the relation $g_{r,r} \leqq f(p)$ holds, hence, the recurrences related to (4) and (5) both yield monotonically increasing sequences of functions. Therefore, Lemma 4 is true for $0 \leqq p \leqq p_r$, because, as we know from Lemma 2 *d*, $g_{r,h} \leqq g_{r,h-1}$, $(r, h) \in \mathbb{N} \times \mathbb{N}$, and we may concentrate on the case $p_r \leqq p \leqq 1$. In this case, the recurrences related to (4) and (5) both yield monotonically decreasing sequences of functions with $f(p)$ as the limit. Because of this, and by Lemma 3 *a*, $g_{r,r+1} \leqq l_{r,r+1} \leqq l_{r,r}$, $r \geqq 2$. Furthermore, because the recurrence (5) decreases faster than that related to (4), $l_{r,r+2} \leqq l_{r,r+1} \leqq l_{r,r}$, $r \geqq 2$.

We now show, that for $p_r \leqq p \leqq 1$ the following relation holds:

$$g_{r,h} \leqq 1 - p + pg_{r,h}g_{r,h-2}, \qquad r \geqq 2,\ h \geqq r+2. \tag{6}$$

To show (6), let us recall that $g_{r,h} = 1 - p + pg_{r,h-1}^2 - p(g_{r,h-1} - g_{r-1,h-1})^2$, $(r, h) \in \mathbb{N} \times \mathbb{N}$.

Now let us assume that $\Delta_{r,h-1} \leqq \Delta_{r,h-2}$, $r \geqq 2$. Then

$$\begin{aligned} g_{r,h} &\leqq 1 - \mathrm{p} + pg_{r,h-1}^2 - p\Delta_{r,h-1}^2 = 1 - p + pg_{r,h-1}^2 - p(g_{r,h-1} - g_{r,h})^2 \\ &= 1 - p + 2pg_{r,h-1}g_{r,h} - pg_{r,h}^2 = 1 - p + pg_{r,h}(g_{r,h-1} + \Delta_{r,h-1}) \\ &\leqq 1 - p + pg_{r,h}g_{r,h-2}. \end{aligned}$$

Next let us assume that $\Delta_{r,h-1} > \Delta_{r,h-2}$. In this case, let $\eta_{r,h-1}$ be defined as $\eta_{r,h-1} := g_{r,h-1} - (1 - p + pg_{r,h-1}^2 - p\Delta_{r,h-2}^2)$. By a short computation, we find that $\eta_{r,h-1} \leqq \Delta_{r,h-2}$. This implies, that there is a $\gamma_{r,h}$, $g_{r,h} \leqq \gamma_{r,h} \leqq g_{r,h-1}$, such that $\gamma_{r,h} = 1 - p + pg_{r,h-1}^2 - p(g_{r,h-1} - \gamma_{r,h})^2$ and

$$g_{r,h-1} - \gamma_{r,h} \leqq \eta_{r,h-1} \leqq \Delta_{r,h-2}.$$

Hence, the value $\gamma_{r,h}$ satisfies $\gamma_{r,h} \leqq 1 - p + p\gamma_{r,h}g_{r,h-2}$, and this implies that $g_{r,h} \leqq 1 - p + pg_{r,h}g_{r,h-2}$. □

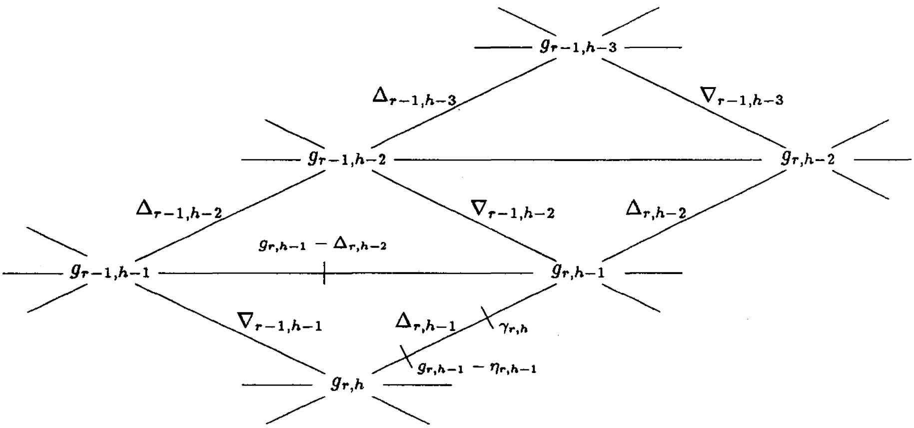

Figure 2. — An illustration of Lemma 2 together with some quantities defined in the proof of Lemma 4.

Figure 2 shows an illustration of the results of Lemma 2 together with the quantities $\eta_{r,h-1}$ and $\gamma_{r,h}$ defined in the proof of Lemma 4. The proof shows, that the upper bound for $g_{r,h}$ is "better", when $\Delta_{r,h-1} \leqq \Delta_{r,h-2}$, because we do not loose as much information as in the case $\Delta_{r,h-1} > \Delta_{r,h-2}$.

Figure 3 shows some examples for $r=4$. Here, $\Delta_{4,4}(\phi) = \Delta_{4,5}(\phi)$. Note that the function $g_{4,5} - \eta_{4,5}$ is larger than $g_{4,6}$, if $\Delta_{4,5} > \Delta_{4,4}$. Near $p=1$, the difference $\Delta_{4,5}$ is smaller than $\Delta_{4,4}$. It makes no sense to define $g_{4,5} - \eta_{4,5}$ there, because, as the graph of it shows, it falls below $g_{4,6}$ there.

Remark: With equation (2) it is a simple matter to show that $\Delta_{2,h} \leqq \Delta_{2,h-1}$, $h \geqq 3$. The values p_r, $r \geqq 2$, can be computed by numerical methods. For example, $p_2 = .543689\ldots$, $p_3 = .502017\ldots$, $p_4 = .5000076\ldots$

With Lemma 4 we are now able to derive a lower bound for the average number of registers over the family $\mathscr{F}(h)$, because the "interleaved" recurrence (5) can be solved by standard methods. We obtain

THEOREM 2: *The probability $g_{r,h}$, that a tree $T \in \mathscr{F}(h)$ requires at most r registers, satisfies the following relations:*

$$a) \quad g_{r,h} \leqq (1-p) \frac{p^{((h-r-2)/2)} p^{2^r} - (1-p)^{((h-r-2)/2)} (1-2p+p^{2^r})}{p^{((h-r)/2)} p^{2^r} - (1-p)^{((h-r)/2)} (1-2p+p^{2^r})},$$

$r \in \mathbb{N}, \quad h \geqq r, \quad h-r$ *even,*

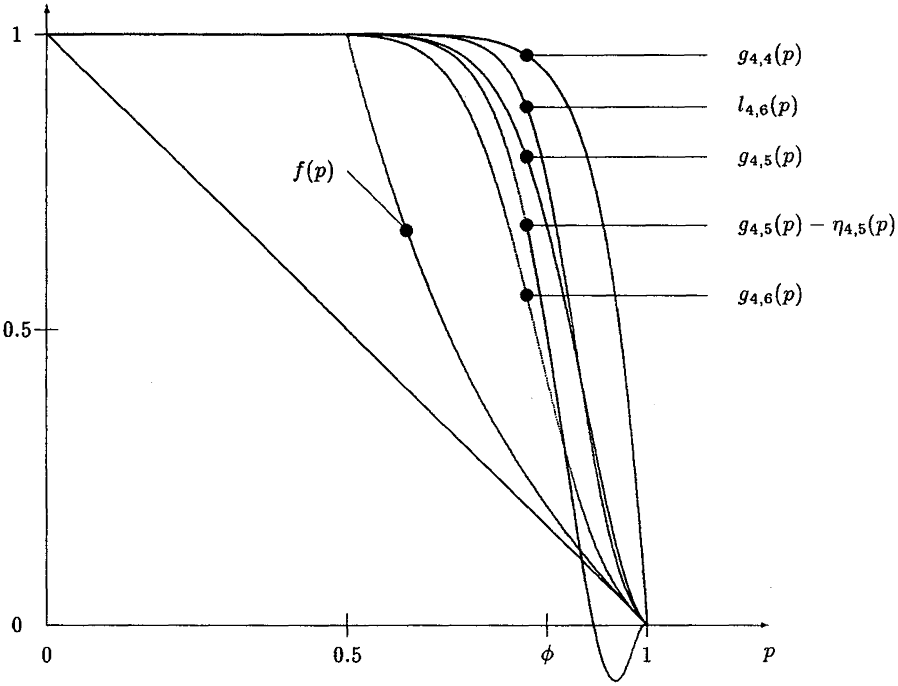

Figure 3. – The functions $g_{4,4}$, $g_{4,5}$, $g_{4,6}$, $g_{4,5} - \eta_{4,5}$ and $l_{4,6}$ together with $1-p$ and $f(p)$.

b) $g_{r,h} \leqq (1-p)$

$$\times \frac{p^{((h-r-3)/2)} p^{2^r+1}(2-p^{2^r-1}) - (1-p)^{((h-r-3)/2)}(1-2p+p^{2^r+1}(2-p^{2^r-1}))}{p^{((h-r-1)/2)} p^{2^r+1}(2-p^{2^r-1}) - (1-p)^{((h-r-1)/2)}(1-2p+p^{2^r+1}(2-p^{2^r-1}))},$$

$r \in \mathbb{N}$, $h \geqq r$, *$h-r$ odd.*

Proof: We only depict the proof for $h-r$ even. Let $a_0 := 1-p^{2^r-1}$, let $b_0 := 1$, and let $R_0 := a_0/b_0$. Then, for $n \geqq 1$, the recurrence $R_n := (a_n/b_n) := (1-p)/(1-p^{(a_{n-1}/b_{n-1})})$ can be translated into the following system of recurrences for a_n and b_n:

$$a_n := (1-p)\, b_{n-1}$$
$$b_n := b_{n-1} - p(1-p)\, b_{n-2}$$

The recurrence for b_n can now be solved using the method of characteristic polynomials. This solution can be used for a similar solution of the recurrence for a_n. □

We can now plug the upper bounds for $g_{r,h}$ into formula (1) and obtain for even h:

$$\underline{R}(h) \geqq 1+\sum_{\mathrm{I}}+\sum_{\mathrm{II}}$$

$$:=1+\sum_{1 \leqq r \leqq h/2} \frac{(2p-1)\,p^{2^{2r}+(h/2)-r-1}}{p^{(h/2)-r}\,p^{2^{2r}}-(1-p)^{(h/2)-r}\,(1-2p+p^{2^{2r}})} \tag{7}$$

$$+\sum_{1 \leqq r \leqq h/2} \frac{(2p-1)\,(2p-p^{2^{2r-1}})\,p^{2^{2r-1}+(h/2)-r-1}}{p^{(h/2)-r}\,p^{2^{2r-1}}\,(2-p^{2^{2r-1}-1})-(1-p)^{(h/2)-r}\,(1-2p+p^{2^{2r-1}+1}\,(2-p^{2^{2r-1}-1}))}.$$

We shall investigate the behaviour of sum $\sum_{\mathrm{I}}$. The investigation of the sum $\sum_{\mathrm{II}}$ can be omitted, because its terms are interleaved with the terms of sum $\sum_{\mathrm{I}}$. Hence, sum $\sum_{\mathrm{II}}$ behaves similarly. Sum $\sum_{\mathrm{I}}$ can be simplified to

$$\sum_{\mathrm{I}}=\sum_{0 \leqq r<h/2} \frac{p^{r}\,(2p-1)\,p^{-1}}{p^{r}-(1-p)^{r}+(1/p^{2^{h}-2^{r}})\,(2p-1)\,(1-p)^{r}}.$$

Unfortunately there does not seem to be a simple closed expression for an asymptotic equivalent of $\sum_{\mathrm{I}}$. Euler's summation formula, for example, is not very helpful, because it requires integration with respect to r. Clearly, each of the terms of $\sum_{\mathrm{I}}$ is 0, if $p=0$, and 1, if $p=1$. Hence, we have to study, *how* they grow up from 0 to 1. We find that the behaviour of each term of $\sum_{\mathrm{I}}$ is determined by its denominator. If p is small, the term $(1/p^{2h-2r})$ $(2p-1)\,(1-p)^{r}$ is very large, hence, the other term $p^{r}-(1-p)^{r}$ has only little influence on the total behaviour. However, if p is large (p near 1), the term $(1/p^{2h-2r})\,(2p-1)\,(1-p)^{r}$ is very small, and the total behaviour is determined by the term $p^{r}-(1-p)^{r}$. In this case, the complete term grows up to 1 with slope about 1. Rougly speaking, we can summarize these observations as follows:

The terms of $\sum_{\mathrm{I}}$ are small for small values of p. If p gets larger, than for each term of $\sum_{\mathrm{I}}$ it reaches a value, where the term jumps up and then, as p increases towards $p=1$, also grows up to 1 with slope about 1.

Given a real number p_ε somewhere near 1, we shall study, *how many* of the terms of $\sum_{\mathrm{I}}$ jump at $p<p_\varepsilon$, because they give a "contribution" to the lower bound of the average number of registers $\underline{R}(h)$ at the "left hand side" of p_ε. For instance, if $r=0$, the corresponding term of $\sum_{\mathrm{I}}$ is equal to p^{2^h-1}, which means, that the jumping point is near 1. If r gets larger, the point moves down, until there is not more a real jumping point (when $p=(h/2)-1$). Let us follow this intention and formalize what we mean, when we talk about a jumping point. For this purpose let us define the jumping point to be the point of intersection of the terms $p^r-(1-p)^r$ and $(1/p^{2^h-2r})(2p-1)(1-p)^r$, if such a point exists, that is a value s, $(1/2)<s<1$, such that $s^r-(1-s)^r=(1/s^{2^h-2r})(2s-1)(1-s)^r$. By a simple comparison of the derivatives of both terms at $p=1/2$, we find that for $r=(h/2)-n$, $n\in[1:h/2]$, there is no such point of intersection, iff $h>2n+2^{2^{2n}}$. In this case the term $p^r-(1-p)^r$ is always larger than $(1/p^{2^h-2r})(2p-1)(1-p)^r$, which can be interpreted as a jumping point at the left hand side of *any* value p_ε, $1/2<p_\varepsilon<1$. These terms give a contribution to the lower bound of the average number of registers $\underline{R}(h)$, however, their number is very small compared with h.

Let us proceed in the investigation of the points of intersection. This is simplified by the fact, that the sum $\sum_{\mathrm{I}}$ satisfies the relation

$$\sum_{\mathrm{I}}>p^{2^h-1}+\sum_{1\leqq r<(h/2)}\frac{p^r(2p-1)p^{-1}}{(2p-1)+(1/p^{2^h-2r})(2p-1)(1-p)^r}$$

$$=p^{2^h-1}+\sum_{1\leqq r<(h/2)}\frac{p^{r-1}}{1+((1-p)^r/p^{2^h-2r})}. \quad (8)$$

The right hand side of this relation behaves also "stairlike", however, the slope at the "right hand side" of a jumping point is larger (about $r-1$). For a given h, we now consider an arbitrary p_ε, $(1/2)<p_\varepsilon<1$, such that there is a r_0, $r_0\in[0:(h/2)-1]$, which satisfies $p_\varepsilon^{2^h-2r_0}=(1-p_\varepsilon)^{r_0}$. All terms of the sum in (8) with $r>r_0$ give a "contribution" to the average number of registers $\underline{R}(h)$ for p smaller than p_ε. We now successively consider the families $\mathscr{F}(2h)$, $\mathscr{F}(4h)$, ..., $\mathscr{F}(2^i h)$, ..., $i\in\mathbb{N}$, and determine the corresponding values r_i, $i\in\mathbb{N}$. In order to do this, let $c:=\log_{p_\varepsilon}(1-p_\varepsilon)$ and $ld(x):=\log_2(x)$. Then for each $i\in\mathbb{N}$ we have to solve $c\,r_i=2^{2^i h-2r_i}$. Hence, $ld(r_i)+ld(c)=2^i h-2r_i$. The term $ld(c)$ may be neglected, therefore, we have to solve $2r_i+ld(r_i)=2^i h$. The solution of this equation is of the form $r_i\approx 2^{i-1}h-(i-1)/2$. This means, that whenever we duplicate the maximum height, we obtain a constant number of significant terms *more.* We can summarize all these observations in the following

THEOREM 3: *Let* $p_\varepsilon := 1-\varepsilon$, *for some small* $\varepsilon > 0$. *Then for* $(1/2) < p \leqq 1$ *and for large* h, $(h \to \infty)$, *the average number of registers* $\underline{R}(h)$ *over the family* $\mathscr{F}(h)$ *satisfies the relation* $\underline{R}(h) \geqq r_h(p)$, *where* $r_h(p)$ *has the following properties:*

– $r_h(p)$ *is a monotonically increasing function in the form of stairs;*

– *there are* $O(\log(h))$ *steps at the "left hand side" of* p_ε;

– *at the right hand side of* p_ε *the function* $r_h(p)$ *rapidly grows up to* $h+1$. □

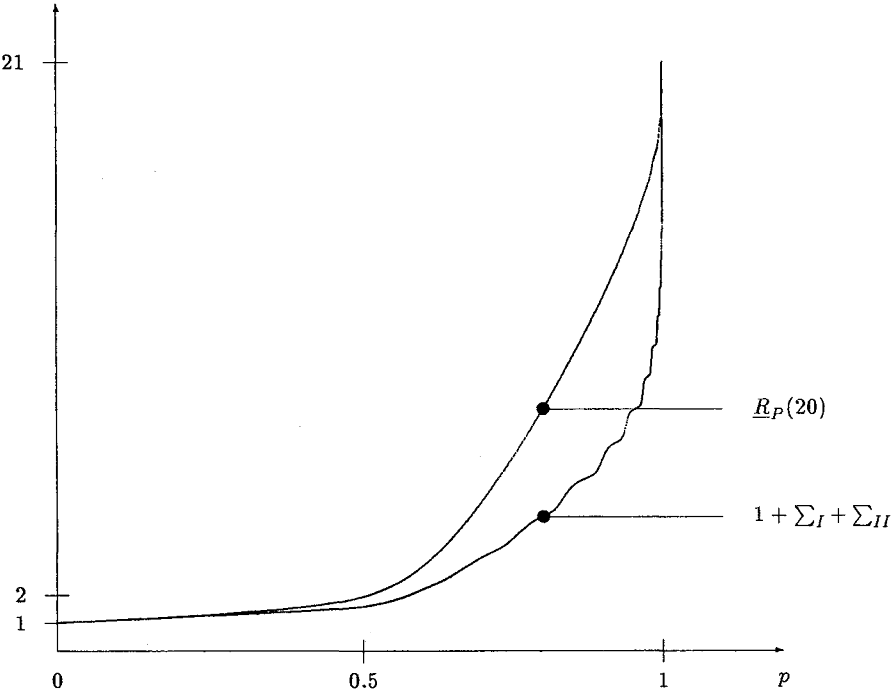

Figure 4. – The average number of register $\underline{R}_p(20)$ together with the lower bound $1+\sum_{\mathrm{I}}+\sum_{\mathrm{II}}$ of (7).

Figure 4 shows the graph of the average number of registers $\underline{R}(20)$ together with the lower bound $1+\sum_{\mathrm{I}}+\sum_{\mathrm{II}}$ given in formula (7). Near $p=1$ the function $\underline{R}(20)$ also *"jumps"* to 21. One reason for this is the term $g_{20,20} = 1-p^{2^{20}-1}$. However, we do not know, how the other terms $g_{..}$ behave near $p=1$. Clearly, as Figure 4 shows, the "stairs" of the lower bound do not have "sharp edges".

3. FINAL REMARKS

In this paper, we have derived a lower bound for the average number of registers $\underline{R}(h)$ over the family $\mathscr{F}(h)$. However, we do not know, how good this lower bound is. Figure 4 shows, that $\underline{R}(20)$ nearly grows linearly from a constant smaller than 2 at $p=(1/2)$ to height 20 near $p=1$. The term $g_{20, 20}$ is responsible for the jump from 20 to 21. There is only little hope to study large values of h empirically, because of the complexity of the involved recurrences and the huge exponents of p, that result from these recurrences. Although we used a relation better than that one in (4), it seems that we loose too much information hidden in the term $p(g_{r, h-1}-g_{r-1, h-1})^2$, which has to be subtracted in the exact recurrence of Lemma 1.

The answer to these questions could be given, if we could find a better upper bound for $\underline{R}(h)$ as that one cited below Lemma 3. A slight improvement could be achieved by starting the related recurrence with $g_2, \ldots$ instead of $g_1, \ldots$ For large h, the probability $g_{2, h}$ is asymptotically equal to $(1-2p+2p^2-p^3)/(1-2p(1-p))$, which can be derived from the exact recurrence by a simple generating function method. However, this improvement doesn't seem to be good enough to obtain a real better knowledge about the probabilities $g_{r, h}$.

Another aspect of this paper is the fact, that the relations of Lemma 2 are the same as the corresponding relations for the stacksize, with the exception of $g_{h, h}$, which is equal to $1-p^{2^h-1}$ instead of $1-p^h$ (compare with [6], Lemma 3). The question arises, if there is a "class of properties", in which these relations hold. Is there another class of properties, in which similar relations hold, but with a finer "granularity"? For instance, in [9], the number of internal nodes has been investigated, that cause the register-function to be incremented. The complete binary tree of height h has 2^h-1 such *critical* nodes, which means, that there are 2^h different probabilities $m_{i, h}$, $i \in [0 : 2^h-1]$, to be considered.

ACKNOWLEDGEMENTS

The author would like to thank an anonymous referee for detailed criticisms and the suggestions for improving the clarity of the presentation.

REFERENCES

1. Ph. Flajolet, J.-C. Raoult and J. Vuillemin, *The Number of Registers Required to Evaluate Arithmetic Expressions*, Theoret. *Comp. Sci.* 1979, *9*, pp. 99-125.
2. R. L. Graham, D. E. Knuth and O. Patashnik, *Concrete Mathematics*, Addison-Wesley, Reading, Mass. 1988.
3. R. Kemp, *The Average Number of Registers Needed to Evaluate a Binary Tree Optimally*, *Acta Inf.*, 1979, *11*, pp. 363-372.
4. R. Kemp, *The Expected Additive Weight of Trees*, *Acta Inf.*, 1989, *26*, pp. 7-740.
5. R. Kemp, *The Analysis of a Special Class of Backtrack Trees*, preprint Johann Wolfgang Goethe-Universität Frankfurt, 1991.
6. R. Kemp, *On the Stacksize of a Class of Backtrack Trees*, preprint Johann Wolfgang Goethe-Universität Frankfurt, 1991.
7. D. E. Knuth, *The Art of Computer Programming*, *Vol.* 1, (*2nd ed.*), Addison-Wesley, Reading, Mass., 1973.
8. A. Meir and J. W. Moon, *On the Altitude of Nodes in Random Trees*, *Can. J. Math.*, 1978, *30* pp. 997-1015.
9. H. Prodinger, *Détermination de certains paramètres d'arbres binaires à l'aide de méthodes analytiques*, Laboratoire de Recherche en Informatique, Université Paris 11, 91405 Orsay cedex-France, Rapport de Recherche N° 177, 1984.
10. P. W. Purdom, *Tree Size by Partial Backtracking*, *SIAM J. Comput.*, 1978, *7* (4) pp. 481-491.
11. U. Trier, *Additive Weights of a Special Class of Nonuniformly Distributed Backtrack Trees*, *Inform. Proc. Letters*, 1992, *42*, pp. 67-76.