THE REGISTER FUNCTION FOR LATTICE PATHS

GUY LOUCHARD AND HELMUT PRODINGER

Abstract. The register function for binary trees is the minimal number of extra registers required to evaluate the tree. This concept is also known as Horton-Strahler numbers. We extend this definition to lattice paths, built from steps ± 1, without positivity restriction. Exact expressions are derived for appropriate generating functions. A procedure is presented how to get asymptotics of all moments, in an almost automatic way; this is based on an earlier paper of the authors.

1. Introduction

The register function of binary trees was introduced by Ershov [7]; the equivalent notion of (Horton-)Strahler numbers was introduced earlier by hydrogeologists Horton [14] and Strahler [25].

This function is recursively defined by $\mathrm{reg}(\Box) = 0$, and, if a binary tree T has subtrees T_1 and T_2, then $\mathrm{reg}(T) = \max\{\mathrm{reg}(T_1), \mathrm{reg}(T_2)\}$, provided $\mathrm{reg}(T_1) \neq \mathrm{reg}(T_2)$, otherwise it is $1 + \mathrm{reg}(T_1)$.

Assuming all binary trees with n internal nodes to be equally likely, the average value of the register function was found independently and at the same time [12, 16]; compare also [21]. It is $\log_4 n + O(1)$, and more precision is available and involves complicated (fluctuating) terms.

The concept has been extended to unary-binary trees [10]. Various papers about the register function (or Horton-Strahler numbers) have been written; we cite a few here [5, 17, 22, 27, 24].

Auber et al. [1] have introduced a generalisation to general rooted trees, see also [6].

Binary trees are enumerated by *Catalan numbers*, and nonnegative lattice paths, with steps ± 1, returning to level 0, as well. We will describe such lattice paths in the popular notation of well-formed words over the alphabet $\{(,)\}$, with n parantheses of each type. (Well-formed means that, when scanning the word from left to right, there are never more closing than opening brackets.) There is the formal equation

$$\mathcal{D} = (\mathcal{D})\mathcal{D} + \varepsilon$$

for the family of nonnegative paths, with ε denoting the empty word.

This decomposition, according to the first return to the 0-level, can be used to encode binary trees, since binary trees satisfy essentially the same equation. Consequently, the register function can be defined on the set of nonnegative lattice paths (returning to the 0-level), by borrowing the definition from the binary trees.

We will make use of known formulæ: Denote by $R_p(z)$ and $S_p(z)$ the generating function where the coefficient of z^n is the number of trees with n nodes and register

Date: April 8, 2008.

function $= p$ resp. $\geq p$. (Equivalenty, the number of paths of length $2n$ with these properties.) Then

$$R_p(z) = \frac{1-u^2}{u}\frac{u^{2^p}}{1-u^{2^{p+1}}}, \tag{1}$$

$$S_p(z) = \frac{1-u^2}{u}\frac{u^{2^p}}{1-u^{2^p}}, \tag{2}$$

with the substitution $z = \frac{u}{(1+u)^2}$.

Now comes the new contribution of this paper: we also allow the paths to go below the x-axis. There are $\binom{2n}{n}$ such paths of length $2n$.

If we denote the family of these paths by $\mathcal{A}$, then our decomposition extends to

$$\mathcal{A} = (\mathcal{D})\mathcal{A} +)\overline{\mathcal{D}}(\mathcal{A} \ + \ \varepsilon,$$

where $\overline{\mathcal{D}}$ is just a copy of $\mathcal{D}$ with opening and closing parentheses exchanged, in other words, it describes non*positive* paths.

These paths have a straight forward interpretation as marked binary trees: Each left edge, which has as predecessors only right edges, gets a label from $\{\mathsf{pos}, \mathsf{neg}\}$, indicating, whether the first excursion is strictly positive or strictly negative.

For the register function, we just borrow it from this corresponding tree, ignoring the extra labels. We can formulate this directly in terms of paths:

$$\operatorname{reg}\big((w)x\big) = \operatorname{reg}\big(\,)\overline{w}(x\,\big) = \max\{\operatorname{reg}(w), \operatorname{reg}(x)\},$$

if $\operatorname{reg}(w) \neq \operatorname{reg}(x)$, otherwise $1 + \operatorname{reg}(w)$, and $\operatorname{reg}(\varepsilon) = 0$.

In the following sections, we first derive explicit expressions for the probability that a random path of length $2n$ has register function $\geq p$ (the instance of register function $= p$ follows from that by taking differences). We derive in particular an exact expression for the average; exact expressions for higher moments are possible, but become messy.

However, we evaluate the higher moments asymptotically. First, it is shown that the asymptotic distribution in our register instance is the same as in the classical register problem for binary trees (lower order terms are different, though). Thus, the asymptotic analysis of moments applies to both, the classical instance, and the one presented in this paper.

The machinery to achieve an almost automatic computation of the moments was presented in our earlier paper [20]. The analysis of the register instance was not included, because of the restriction on the length of this paper. Now we feel that it is a good opportunity that this treatment can go into print.

There are some related parameters, like the size of the maximal complete subtree in a binary tree. Our methods apply here as well, but we do not include that here, as it would lead us a bit far apart.

Some technical considerations that are in [20] are not repeated here, to allow for a smooth reading.

2. Generating functions

Let $I_p(z)$ and $J_p(z)$ be the generating function of paths (unrestricted), with register function $= p$ resp. $\geq p$. The register random variable related to a path of length $2n$ is denoted by X_n: $\mathbb{P}\{X_n = p\} = [z^n]I_p(z)/\binom{2n}{n}$. This will later be denoted by

$1 - P(p-1)$. We will find a recursion for $J_p(z)$. Of course, the coefficient of z^n refers to paths of length $2n$, and $J_0(z) = A(z) = \frac{1}{\sqrt{1-4z}}$, the generating function of all paths.

The following recursion is known [23]:

$$S_p = zS_{p-1}^2 + z(D - S_{p-1})S_p + zS_p(D - S_{p-1}), \tag{3}$$

where

$$D = D(z) = \frac{1 - \sqrt{1-4z}}{2z} \tag{4}$$

is the generating function of binary trees (and of nonnegative lattice paths). This recursion is an immediate translation of the definition of the register function. The same type of reasoning leads to the following recursion, which we state as a proposition.

Proposition 2.1.

$$J_p = 2zS_{p-1}J_{p-1} + 2z(D - S_{p-1})J_p + 2z(A - J_{p-1})S_p. \tag{5}$$

The solution is stated as a theorem, with the standard substitution $z = u/(1+u)^2$:

Theorem 1.

$$J_p = \frac{(1+u)^2}{u}\frac{2^p u^{2^p}}{(1-u^{2^p})^2} - 2\frac{1+u}{1-u}\frac{u^{2^p}}{1-u^{2^p}}. \tag{6}$$

Proof. Once the formula is known, the verification is routine, and can be done by a computer. In order to find it, however, one has to solve the first-order recursion by iteration. Doing this, the following sum appears:

$$\sum_{k=0}^{p-1} \frac{(1-u^{2^k})^2}{2^k},$$

which, quite amazingly, has the explicit evaluation

$$2(1-u) - \frac{1-u^{2^p}}{2^{p-1}}.$$

□

In order to study the average value of the register function, we must consider $E(z) = \sum_{p\geq 1} J_p$. Note that

$$\sum_{p\geq 1} \frac{u^{2^p}}{1-u^{2^p}} = \sum_{p,k\geq 1} u^{k2^p} = \sum_{n\geq 1} v_2(n)u^n,$$

where $v_2(n)$ is the number of trailing zeroes in the binary representation of n. Similarly,

$$\sum_{p\geq 1} \frac{2^p u^{2^p}}{(1-u^{2^p})^2} = \sum_{p,k\geq 1} k2^p u^{k2^p} = \sum_{n\geq 1} nv_2(n)u^n.$$

To read off coefficients, we notice the formula

$$[z^n]f(z) = [u^n](1-u)(1+u)^{2n-1}f(z(u)), \tag{7}$$

see [4], which is obtained by Cauchy's integral formula and a change of variable. Thus we compute

$$\begin{aligned}[z^n]E(z) &= [u^{n+1}](1-u)(1+u)^{2n+1}\sum_{k\ge1} kv_2(k)u^k - 2[u^n](1+u)^{2n}\sum_{k\ge1} v_2(k)u^k\\ &= \sum_{k\ge1} kv_2(k)[u^{n+1-k}](1-u)(1+u)^{2n+1} - 2\sum_{k\ge1} v_2(k)u^k[u^{n-k}](1+u)^{2n}\\ &= \sum_{k\ge1} kv_2(k)\left[\binom{2n+1}{n+1-k} - \binom{2n+1}{n-k}\right] - 2\sum_{k\ge1} v_2(k)\binom{2n}{n-k}.\end{aligned}$$

Normalising, we get the following explicit result.

Theorem 2. *The average value of the register function, considering all (unrestricted) lattice path of length $2n$ to be equally likely, is given by*

$$\frac{1}{\binom{2n}{n}}\sum_{k\ge1} v_2(k)\left[k\binom{2n+1}{n+1-k} - k\binom{2n+1}{n-k} - 2\binom{2n}{n-k}\right].$$

Alternatively, it can be written as

$$\frac{1}{\binom{2n}{n}}\sum_{k\ge1} v_2(k)\left[k\binom{2n}{n+1-k} - 2\binom{2n}{n-k} - k\binom{2n}{n-1-k}\right].$$

□

Note the similar formula for binary trees:

$$\frac{n+1}{\binom{2n}{n}}\sum_{k\ge1} v_2(k)\left[\binom{2n}{n+1-k} - 2\binom{2n}{n-k} + \binom{2n}{n-1-k}\right].$$

The asymptotic evaluation of this (and higher moments) will be studied in the next sections.

In particular, we will prove

Theorem 3. *The average number of registers is asymptotic to*

$$\log_4 n + \frac{1}{\ln 4}(\ln\pi - 1 - \frac{\gamma}{2}) + \frac{1}{2} + -\frac{1}{\ln 4}\sum_{l\neq0}\Gamma(-\chi_l/2)\zeta(-\chi_l)(\chi_l+1)e^{-l\pi i\log_4 n}$$

with $\chi_l := 2l\pi i/\ln 4$. □

Formulæ for $I_p = J_p - J_{p+1}$ and its coefficients can also be given.

Note that

$$J_p(z) = \frac{(1+u)^2}{u}\sum_{k\ge1} k2^p u^{k2^p} - 2\frac{1+u}{1-u}\sum_{k\ge1} u^{k2^p}.$$

Thus

$$\begin{aligned}[z^n]J_p(z) &= [u^n](1-u)(1+u)^{2n-1}\frac{(1+u)^2}{u}\sum_{k\ge1} k2^p u^{k2^p}\\ &\quad - [u^n](1-u)(1+u)^{2n-1}2\frac{1+u}{1-u}\sum_{k\ge1} u^{k2^p}\end{aligned}$$

$$= \sum_{k\geq 1} k2^p [u^{n+1-k2^p}](1-u)(1+u)^{2n+1} - 2\sum_{k\geq 1}[u^{n-k2^p}](1+u)^{2n}$$
$$= \sum_{k\geq 1} k2^p \left[\binom{2n+1}{n+1-k2^p} - \binom{2n+1}{n-k2^p}\right] - 2\sum_{k\geq 1}\binom{2n}{n-k2^p}.$$

Therefore we find the following formula for the probability that a random path of length $2n$ has register function $\geq p$:

$$\frac{[z^n]J_p(z)}{\binom{2n}{n}} = \sum_{k\geq 1} k2^p \frac{\binom{2n+1}{n+1-k2^p} - \binom{2n+1}{n-k2^p}}{\binom{2n}{n}} - 2\sum_{k\geq 1}\frac{\binom{2n}{n-k2^p}}{\binom{2n}{n}}. \tag{8}$$

This can be approximated by

$$\sum_{k\geq 1}\Big(-2 + \frac{k^2 4^{p+1}}{n}\Big)e^{-k^2 4^p/n}. \tag{9}$$

We just give the following remark about this: The approximation of binomial coefficients viz. $\binom{2n}{n-k}/\binom{2n}{n}$ for $|k| \leq n^{\frac{1}{2}+\varepsilon}$ is well known (it appears in all the earlier papers on the register function); it is just a consequence of Stirling's approximation of the factorials. Outside this range, the quantity is exponentially small. A careful error analysis of similar expressions can be found in [8].

For the classical register problem, the same quantity arises, with differences only in lower order terms that are not displayed. This shows that the register function is quite immune to the input model.

3. Semi-automatic computation of moments: an outline

We will use the following paradigm: to compute the asymptotics of the moments, we use the techniques described in great detail in Louchard and Prodinger [20], which are usually simpler than the ones found in the literature. We encounter extreme-value (Gumbel-like) related distributions functions. The Gumbel distribution function is given by $\exp(-\exp(-x))$. If we compare this approach with other ones that appeared previously, then we can notice the following. Traditionally, one would stay with exact enumerations as long as possible, and only at a late stage move to asymptotics. Doing this, one would, in terms of asymptotics, carry many unimportant contributions around, which makes the computations quite heavy, especially when it comes to higher moments. Here, however, approximations are carried out as early as possible, and this allows for streamlined (and often automatic) computations of the higher moments.

Note that the distributions that we can handle do not converge in the weak sense, they do however converge along subsequences n_m for which the fractional part of $\log n_m$ is constant (the logarithm is to a base that will be specified).

Here are the main steps of our approach.

We set

$$p(j) := \mathbb{P}\{X_n = j\}, \qquad P(j) := \mathbb{P}\{X_n \leq j\}.$$

(Note that $P(j) = 1 - [z^n]S_{j+1}(z)/\binom{2n}{n}$.)

We write $\log n$ for $\log_Q n$; the base Q will be given later. Setting $\eta = j - \log n$, we will first compute f and F such that

$$p(j) \sim f(\eta), \qquad P(j) \sim F(\eta), \qquad n \to \infty,$$

and, of course,

$$f(\eta) = F(\eta) - F(\eta - 1).$$

(That is why we do not use/need the dependency on n in the notation of $p(j)$ and $P(j)$ just introduced; $F(z)$ is of the Gumbel type.)

Asymptotically, the distribution will be a periodic function of the fractional part of $\log n$.

Next, we are going to show that

$$\mathbb{E}\left(X_n^k\right) = \sum_j j^k p(j) \sim \sum_j (\eta + \log n)^k [F(\eta) - F(\eta - 1)]. \tag{10}$$

by computing a suitable rate of convergence (in particular for large and small values of η). This is related to a uniform integrability condition (see Loève [18, Section 11.4].)

So, once we gain enough knowledge about the function $F(x)$, we know the moments. This will be achieved essentially by using the Mellin transform technique. For $F(x)$ that are related to the Gumbel distribution (details are in the following lemma), the desired asymptotic moments come out as the coefficients of a certain generating function, and they can be computed by Maple.

Finally we will use the following result from Hitczenko and Louchard [13], Louchard and Prodinger [20], related to the dominant part of the moments (the '~' sign is related to the moments of the integer-valued random variable X_n).

Lemma 1. *Let a (integer-valued) random variable X_n be such that $\mathbb{P}\{X_n - \log n \le \eta\} \sim F(\eta)$, where $F(\eta)$ is the distribution function of a continuous random variable Z with mean m_1, second moment m_2, variance σ^2 and centered moments μ_k. Assume that $F(\eta)$ is either an extreme-value distribution function or a convergent series of such and that (10) is satisfied. Let*

$$\varphi(\alpha) = \mathbb{E}(e^{\alpha Z}) = 1 + \sum_{k=1}^{\infty} \frac{\alpha^k}{k!} m_k = e^{\alpha m_1} \lambda(\alpha),$$

say, with

$$\lambda(\alpha) = 1 + \frac{\alpha^2}{2}\sigma^2 + \sum_{k=3}^{\infty} \frac{\alpha^k}{k!} \mu_k.$$

Let w, κ's (with or without subscripts) denote periodic functions of $\log n$, with period 1 and with usually small mean and amplitude. Actually, these functions depend on the fractional part of $\log n$, denoted by $\{\log n\}$, as usual.

Then the corresponding moments of X_n are given by

$$\begin{aligned}
\mathbb{E}(X_n - \log n) &\sim \int_{-\infty}^{+\infty} x[F(x) - F(x-1)]dx + w_1 \\
&= \tilde{m_1} + w_1, \qquad \text{with} \qquad \tilde{m_1} = m_1 + \frac{1}{2}, \\
\mathsf{Var}(X_n) &\sim \mathbb{E}(X_n - (\log n + \tilde{m_1} + w_1))^2 \\
&\sim \int_{-\infty}^{+\infty} x^2[F(x) - F(x-1)]dx - \tilde{m_1}^2 + \kappa_2
\end{aligned}$$

$$= m_2 + m_1 + \frac{1}{3} - \tilde{m_1}^2 + \kappa_2 = \tilde{\sigma}^2 + \kappa_2, \quad \textit{with} \quad \tilde{\sigma}^2 = \sigma^2 + \frac{1}{12}.$$

More generally, the centered moments of X_n are asymptotically given by $\tilde{\mu}_i + \kappa_i$, where

$$\Theta(\alpha) := 1 + \sum_{k=2}^{\infty} \frac{\alpha^k}{k!} \tilde{\mu}_k = \frac{2}{\alpha} \sinh\big(\frac{\alpha}{2}\big) \lambda(\alpha).$$

The neglected part is of order $1/n^\beta$ with $0 < \beta < 1$. □

For instance, we derive

$$\begin{aligned}
\tilde{\mu}_2 &= \tilde{\sigma}^2 = \mu_2 + \frac{1}{12}, \\
\tilde{\mu}_3 &= \mu_3, \\
\tilde{\mu}_4 &= \mu_4 + \frac{\sigma^2}{2} + \frac{1}{80}, \\
\tilde{\mu}_5 &= \mu_5 + \frac{5}{6}\mu_3.
\end{aligned}$$

The moments of $X_n - \log n$ are asymptotically given by $\tilde{m}_i + w_i$, where the generating function of $\tilde{m}_i$ is given by

$$\phi(\alpha) := \int_{-\infty}^{\infty} e^{\alpha\eta} f(\eta) d\eta = 1 + \sum_{i=1}^{\infty} \frac{\alpha^i}{i!} \tilde{m}_i = \varphi(\alpha) \frac{e^\alpha - 1}{\alpha}. \tag{11}$$

The convergence domain for α will be studied in Section 4. This leads to

$$\begin{aligned}
\tilde{m}_1 &= m_1 + \frac{1}{2}, \\
\tilde{m}_2 &= m_2 + m_1 + \frac{1}{3}, \\
\tilde{m}_3 &= m_3 + \frac{3m_2}{2} + m_1 + \frac{1}{4};
\end{aligned}$$

w_i and κ_i will be analyzed in the next section.

Note that $\Theta(\alpha) = \phi(\alpha) e^{-\alpha \tilde{m}_1}$; from this we can compute the centered moments:

$$\begin{aligned}
\tilde{\mu}_2 &= \tilde{m}_2 - \tilde{m}_1^2, \\
\tilde{\mu}_3 &= \tilde{m}_3 + 2\tilde{m}_1^3 - 3\tilde{m}_2 \tilde{m}_1.
\end{aligned} \tag{12}$$

Now we turn to the fluctuating components that appear invariably in the asymptotic expansions. We analyze the periodic component w_i to be added to the moments $\tilde{m}_i$. Recall formula (10):

$$\mathbb{E}(X_n - \log n) \sim E^{(1)}(n) = \sum_{j=1}^{\infty} \big(F(j - \log n) - F(j - \log n - 1)\big)(j - \log n). \tag{13}$$

Set $y = Q^{-x}$ and $G(y) = F(x)$. Equation (13) becomes a *harmonic sum*

$$E^{(1)}(n) := \sum_j \big(G(n/Q^j) - G(n/Q^{j+1})\big)\big(-\log(n/Q^j)\big),$$

the Mellin transform of which is (for a good reference on Mellin transforms, see Flajolet et al. [9] or Szpankowski [26])

$$\frac{Q^s}{1-Q^s}\Upsilon_1^*(s), \tag{14}$$

and

$$\begin{aligned}\Upsilon_1^*(s) &= \int_0^\infty y^{s-1}(G(y)-G(y/Q))(-\log y)dy\\ &= \int_{-\infty}^\infty Q^{-sx}(F(x)-F(x-1))xLdx.\end{aligned}$$

From this we see that (we use $L=\ln Q$)

$$\Upsilon_1^*(s) = L\phi'(\alpha)\big|_{\alpha=-Ls}. \tag{15}$$

The fundamental strip of (14) is usually of the form $s\in\langle -C_1,0\rangle$, $C_1>0$. This will be detailed in Section 4.

Set also

$$\Upsilon_0^*(s) = L\phi(\alpha)\big|_{\alpha=-Ls},\ \Upsilon_0^*(0)=L.$$

We assume now that all poles of $\frac{Q^s}{1-Q^s}\Upsilon_1^*(s)$ are simple, which will be the case here, and given by $s=\chi_l$, with $\chi_l := 2l\pi i/L$, $l\in\mathbb{Z}$; usually one has to distinguish the case $l=0$ from the others.

Using Mellin's inversion formula viz.

$$E^{(1)}(n) = \frac{1}{2\pi i}\int_{C_2-i\infty}^{C_2+i\infty}\frac{Q^s}{1-Q^s}\Upsilon_1^*(s)n^{-s}ds, \qquad -C_1<C_2<0,$$

the asymptotic expression of $E^{(1)}(n)$ is obtained by moving the line of integration to the right, for instance to the line $\Re = C_4>0$, taking residues into account (with a negative sign). This gives

$$E^{(1)}(n) = -\operatorname{Res}\Big[\frac{Q^s}{1-Q^s}\Upsilon_1^*(s)n^{-s}\Big]\Big|_{s=0} - \sum_{l\neq 0}\operatorname{Res}\Big[\frac{Q^s}{1-Q^s}\Upsilon_1^*(s)n^{-s}\Big]\Big|_{s=\chi_l} + \mathcal{O}(n^{-C_4}). \tag{16}$$

The residue at $s=0$ gives of course

$$\tilde{m}_1 = \Upsilon_1^*(0)/L = \phi'(0).$$

The other residues lead to

$$w_1 = \frac{1}{L}\sum_{l\neq 0}\Upsilon_1^*(\chi_l)e^{-2l\pi i\log n}. \tag{17}$$

More generally,

$$\mathbb{E}(X_n-\log n)^k \sim \tilde{m}_k + w_k,$$

with

$$w_k = \frac{1}{L}\sum_{l\neq 0}\Upsilon_k^*(\chi_l)e^{-2l\pi i\log n}, \tag{18}$$

and

$$\Upsilon_k^*(s) = L\phi^{(k)}(\alpha)\big|_{\alpha=-Ls}.$$

To compute the periodic component κ_i to be added to the centered moments $\tilde{\mu}_i$, we first set

$$\mathfrak{m}_1 := \tilde{m}_1 + w_1.$$

The variance of $X_n - \log n$ is asymptotically given by

$$\mathbb{E}[(X_n - \log n - \mathfrak{m}_1)^2] \sim \tilde{m}_2 + w_2 - \mathfrak{m}_1^2 = \tilde{\mu}_2 + \kappa_2.$$

The third centered moment is asymptotically given by

$$\mathbb{E}[(X_n - \log n - \mathfrak{m}_1)^3] \sim \tilde{m}_3 + w_3 - 3(\tilde{m}_2 + w_2)\mathfrak{m}_1 + 3\mathfrak{m}_1^3 - \mathfrak{m}_1^3 = \tilde{\mu}_3 + \kappa_3.$$

More generally, we start from

$$\phi(\alpha) := 1 + \sum_{k=1}^{\infty} \frac{\alpha^k}{k!}\tilde{m}_k = \varphi(\alpha)\frac{e^\alpha - 1}{\alpha}.$$

We replace $\tilde{m}_k$ by $\tilde{m}_k + w_k$, leading to

$$\phi_p(\alpha) = \phi(\alpha) + \sum_{k=1}^{\infty} \frac{\alpha^k}{k!} w_k.$$

But since $\phi(2l\pi i) = 0$ for all $l \in \mathbb{Z}$, we have

$$\sum_{l\neq 0} \phi(-L\chi_l) e^{-2l\pi i \log n} = 0,$$

so we obtain

$$\begin{aligned}
\phi_p(\alpha) &= \phi(\alpha) + \sum_{k=0}^{\infty} \sum_{l\neq 0} \phi^{(k)}(\alpha)\Big|_{\alpha=-L\chi_l} e^{-2l\pi i \log n} \frac{\alpha^k}{k!} \\
&= \phi(\alpha) + \sum_{l\neq 0} \phi(\alpha - L\chi_l) e^{-2l\pi i \log n} \qquad (19) \\
&= \sum_{l\in\mathbb{Z}} \phi(\alpha - L\chi_l) e^{-2l\pi i \log n}.
\end{aligned}$$

Finally, we compute

$$\Theta_p(\alpha) = \phi_p(\alpha) e^{-\alpha \mathfrak{m}_1} = 1 + \sum_{k=2}^{\infty} \frac{\alpha^k}{k!}(\tilde{\mu}_k + \kappa_k) = \Theta(\alpha) + \sum_{k=2}^{\infty} \frac{\alpha^k}{k!}\kappa_k, \qquad (20)$$

leading to the (exponential) generating function of κ_k. The first few instances are

$$\begin{aligned}
\kappa_2 &= w_2 - w_1^2 - 2\tilde{m}_1 w_1, \\
\kappa_3 &= 6\tilde{m}_1^2 w_1 + 6\tilde{m}_1 w_1^2 + 2w_1^3 - 3\tilde{m}_2 w_1 - 3\tilde{m}_1 w_2 - 3w_1 w_2 + w_3.
\end{aligned} \qquad (21)$$

All algebraic manipulations in this context are mechanically performed by Maple. We give explicit expressions for $\tilde{\mu}_2$, κ_2, $\tilde{\mu}_3$ and κ_3 for illustration.

It will appear that $\Upsilon_k^*(s)$ are analytic functions (in some domain), depending on classical functions such as Γ, ζ. The justification of (16) is by contour integration, see [20] for details.

It is not always evident that the limiting function $F(\eta)$ is indeed a distribution function. But here we can use the following lemma from Janson [15]. Assume that

Y_n is a sequence of integer-valued random variables that is monotone. If, for every sequence of integers k_n,

$$\mathbb{P}\{Y_n \le k_n\} = F(k_n - a_n) + o(1), \qquad n \to \infty,$$

for a right-continuous function F with $\lim_{x\to-\infty} F(x) = 0$, $\lim_{x\to\infty} F(x) = 1$ and a sequence a_n such that $a_n \to \infty$ and $a_{n+1} - a_n \to 0$ as $n \to \infty$, then F is a distribution function.

These conditions are satisfied in our case. We apply the presented machinery in the next section to the number of registers.

4. The moments for the number of registers

From Flajolet [8], Flajolet and Prodinger [10], Louchard [19] and Section 2 we have

$$P(j) \sim 1 + 2\sum_{k=1}^{\infty}(1 - 2k^2 4^{j+1}/n)e^{-k^2 4^{j+1}/n},$$
$$p(j) \sim 2\sum_{k \text{ odd} >0}^{\infty}(2k^2 4^j/n - 1)e^{-k^2 4^j/n}.$$

Let $\eta = j - \log_4 n$; this leads to

$$\begin{aligned} F(\eta) &= 1 + 2\sum_{k=1}^{\infty}(1 - 8k^2 4^{\eta})e^{-4k^2 4^{\eta}}, \\ f(\eta) &= 2\sum_{k \text{ odd} >0}^{\infty}(2k^2 4^{\eta} - 1)e^{-k^2 4^{\eta}}, \end{aligned} \tag{22}$$

where we recognize the Gumbel distribution (Compare also [12, 16].)

We have no rate of convergence problem here: the convergence of moments has been proved in Flajolet [8].

It is easier in this case to start from $f(\eta)$ given by (22) and to compute

$$\phi(\alpha) = \int_{-\infty}^{+\infty} e^{\alpha\eta} f(\eta) d\eta.$$

Setting $y = 2^{\eta}$ and using Mellin transforms, this gives

$$\phi(\alpha) = \Gamma(\xi/2)(\xi - 1)\zeta(\xi)(1 - 2^{-\xi})/L\big|_{\xi=\alpha/L}, \qquad -\infty < \Re(\alpha) < \infty,$$

as shown in Louchard [19]; compare also Biane and Yor [3] and Biane, Pitman and Yor [2]. It is easy to check that there is no singularity at $\alpha = 0$, $\alpha = L$, $\alpha = -2kL$, $k > 0$.

The fundamental strip for (14) is $\Re(s) \in \langle -\infty, 0\rangle$.

Let us first analyze $X_n^* := X_n - \log_4 n$. After all computations (we must be careful about the presence of $\log_4$), and setting

$$\gamma_k := \lim_{n\to\infty}\left[\sum_{i=1}^{n}\frac{(\ln i)^k}{i} - \frac{(\ln n)^{k+1}}{k+1}\right],$$

also called the Stieltjes constants, we derive

$$\begin{aligned}
\tilde{m}_1 &= \frac{1}{L}(\ln\pi - 1 - \frac{\gamma}{2}) + \frac{1}{2},\\
\tilde{m}_2 &= \frac{1}{L^2}(-2\ln\pi - \gamma\ln\pi + \gamma + \frac{1}{8}\pi^2 - \frac{3}{4}\gamma^2 + \ln^2\pi - 2\gamma_1) + \frac{1}{L}(\ln\pi - 1 - \frac{\gamma}{2}) + \frac{1}{3},\\
\tilde{m}_3 &= \frac{1}{L^3}\Big(-\frac{3}{8}\pi^2 + \frac{9}{4}\gamma^2 - \frac{3}{16}\pi^2\gamma + \frac{7}{4}\zeta(3) - \frac{5}{8}\gamma^3 - 3\gamma_1\gamma + 6\gamma_1 - 3\gamma_2 - 3\ln^2\pi - \frac{3}{2}\gamma\ln^2\pi\\
&\quad + 3\gamma\ln\pi - \frac{9}{4}\gamma^2\ln\pi - 6\gamma_1\ln\pi + \frac{3}{8}\pi^2\ln\pi + \ln^3\pi\Big)\\
&\quad + \frac{1}{L^2}\Big(-\frac{9}{8}\gamma^2 + \frac{3}{2}\gamma - 3\ln\pi + \frac{3}{2}\ln^2\pi - 3\gamma_1 + \frac{3}{16}\pi^2 - \frac{3}{2}\gamma\ln\pi\Big)\\
&\quad + \frac{1}{L}(\ln\pi - 1 - \frac{\gamma}{2}) + \frac{1}{4},\\
\tilde{\mu}_2 &= \frac{1}{L^2}\Big(-1 + \frac{1}{8}\pi^2 - \gamma^2 - 2\gamma_1\Big) + \frac{1}{12},\\
\tilde{\mu}_3 &= \frac{1}{L^3}\Big(-2 - 3\gamma_2 - 6\gamma_1\gamma - 2\gamma^3 + \frac{7}{4}\zeta(3)\Big).
\end{aligned}$$

$\tilde{m}_1$ was computed in Flajolet and Prodinger [10], $\tilde{m}_2$ was computed in Louchard [19].

Let us now turn to the fluctuating components. First of all, (15) and (17) lead to

$$w_1 = -\frac{1}{L}\sum_{l\neq 0}\Gamma(-\chi_l/2)\zeta(-\chi_l)(\chi_l+1)e^{-l\pi i\log n}.$$

Note that, in the exponent of the Fourier component, we have now $-l\pi i\log n$. Equations (19) and (20) lead to

$$\begin{aligned}
\kappa_2 &= -2w_1 - w_1^2 + \frac{1}{L}w_1(-2\ln\pi + \gamma)\\
&\quad - \frac{1}{L^2}\sum_{l\neq 0}\Gamma(-\chi_l/2)\Big(2\zeta'(-\chi_l)(\chi_l+1) + \zeta(-\chi_l)\psi(-\chi_l/2)(\chi_l+1) + 2\zeta(-\chi_l)\chi_l\Big)e^{-l\pi i\log n},\\
\kappa_3 &= \frac{3}{8L^2}w_1\Big(16 + 10\gamma^2 - 8\gamma\ln\pi + 8\ln\pi^2 - 8\gamma L - 16L + 16L\ln\pi\\
&\qquad\qquad + 8\gamma - 16\ln\pi + 8L^2 - \pi^2 + 16\gamma_1\Big)\\
&\quad + \frac{3}{L}w_1^2(2\ln\pi - \gamma - 2 + 2L) + 2w_1^3\\
&\quad + \frac{1}{L^3}\sum_{l\neq 0}\Gamma(-\chi_l/2)\Big(-\frac{3}{4}\zeta(-\chi_l)\psi(-\chi_l/2)^2(\chi_l+1) - 3\zeta'(-\chi_l)\psi(-\chi_l/2)(\chi_l+1)\\
&\qquad + 3\zeta(-\chi_l)\psi(-\chi_l/2) + \frac{3}{2}\zeta(-\chi_l)\psi(-\chi_l/2)(\chi_l+1)(2Lw_1 - \gamma - 2 + 2L + 2\ln\pi)\\
&\qquad - \frac{3}{4}\zeta(-\chi_l)\psi(1,-\chi_l/2)(\chi_l+1) - 3\zeta(-\chi_l)(2\ln\pi - \gamma - 2 + 2L + 2Lw_1)\\
&\qquad + \zeta'(-\chi_l)[6 + 3(2\ln\pi + 2Lw_1 - \gamma + 2L - 2)(\chi_l+1)] - 3\zeta''(-\chi_l)(\chi_l+1)\Big)e^{-l\pi i\log n}.
\end{aligned}$$

The expression for w_1 is identical to the expressions given in Flajolet and Prodinger [10].

5. Conclusion

We plan to investigate other models of paths in the context of the register function in a future paper.

References

[1] D. Auber, J.-P. Domenger, M. Delest, P. Duchon, and J.-M. Fédou. New Strahler numbers for rooted plane trees. In *Mathematics and Computer Science III*, pages 203–215. Birkäuser, Basel, 2004.

[2] P. Biane, J. Pitman, and M. Yor. Probability laws related to the Jacobi theta and Riemann zeta functions, and the Brownian excursions. *Bulletin (New series) of the American Mathematical Society*, 38(4):435–465, 2001.

[3] P. Biane and M. Yor. Valeurs principales associées aux temps locaux browniens. *Bulletin de la Société Mathématique de France*, 2e série,111:23–101, 1987.

[4] N. G. De Bruijn, D. E. Knuth, and S. O. Rice. The average height of planted plane trees. In R. C. Read, editor, *Graph Theory and Computing*, pages 15–22. Academic Press, 1972.

[5] L. Devroye and P. Kruszewski. On the Horton–Strahler number for random tries. *RAIRO Inform. Théor. Appl.*, 30(5):443–456, 1996.

[6] M. Drmota and H. Prodinger. The register function for t-ary trees. *ACM Trans. Algorithms*, 2(3):318–334, 2006.

[7] A. P. Ershov. On programming of arithmetic operations. *Communications of the ACM*, 1:3–6, 1958.

[8] P. Flajolet. *Analyse d'algorithmes de manipulation d'arbres et de fichiers*. Université Pierre et Marie Curie, Paris, 1981.

[9] P. Flajolet, X. Gourdon, and P. Dumas. Mellin transforms and asymptotics: Harmonic sums. *Theoretical Computer Science*, 144:3–58, 1995.

[10] P. Flajolet and H. Prodinger. Register allocation for unary-binary trees. *SIAM J. Comput.*, 15:629–640, 1986.

[11] P. Flajolet and H. Prodinger. Register allocation for unary-binary tress. *SIAM Journal on Computing*, 15(3):629–640, 1986.

[12] P. Flajolet, J.-C. Raoult, and J. Vuillemin. The number of registers required for evaluating arithmetic expressions. *Theoret. Comput. Sci.*, 9(1):99–125, 1979.

[13] P. Hitczenko and G. Louchard. Distinctness of compositions of an integer: a probabilistic analysis. *Random Structures and Algorithms*, 19(3,4):407–437, 2001.

[14] R. E. Horton. Erosioned development of systems and their drainage basins. *Bulletin geological society of America*, 56:275–370, 1945.

[15] S. Janson. Rounding of continuous random variables and oscillatory asymptotics. *Annals of Probability*, 34:5:1807–1826, 2006.

[16] R. Kemp. The average number of registers needed to evaluate a binary tree optimally. *Acta Inform.*, 11(4):363–372, 1978/79.

[17] P. Kruszewski. A note on the Horton–Strahler number for random binary search trees. *Inform. Process. Lett.*, 69(1):47–51, 1999.

[18] M. Loève. *Probability Theory*. D. Van Nostrand, 1963.

[19] G. Louchard. Brownian motion and algorithm complexity. *BIT*, 26:17–34, 1986.

[20] G. Louchard and H. Prodinger. Asymptotics of the moments of extreme-value related distribution functions. *Algorithmica*, 46:431–467, 2006. Long version: http://www.ulb.ac.be/di/mcs/louchard/moml.ps.

[21] A. Meir, J. W. Moon, and J. R. Pounder. On the order of random channel networks. *SIAM J. Algebraic Discrete Methods*, 1(1):25–33, 1980.

[22] M. Nebel. A unified approach to the analysis of Horton-Strahler parameters of binary tree structures. *Random Structures Algorithms*, 21(3-4):252–277, 2002. Random structures and algorithms (Poznan, 2001).

[23] H. Prodinger. Some analytic techniques for the investigation of the asymptotic behaviour of tree parameters. *EATCS Bulletin*, 47:180–199, 1992.
[24] H. Prodinger. On a problem of Yekutieli and Mandelbrot about the bifurcation ratio of binary trees. *Theoretical Computer Science*, 181:181–194, 1997.
[25] A. N. Strahler. Hypsomic analysis of erosional topography. *Bulletin geological society of America*, 63:1117–1142, 1952.
[26] W. Szpankowski. *Average Case Analysis of Algorithms on Sequences*. Wiley, New York, 2001.
[27] I. Yekutieli and B. Mandelbrot. Horton–Strahler ordering of random binary trees. *J. Phys. A*, 27(2):285–293, 1994.

Guy Louchard, Université Libre de Bruxelles, Département d'Informatique, CP 212, Boulevard du Triomphe, B-1050 Bruxelles, Belgium
E-mail address: louchard@ulb.ac.be

Helmut Prodinger, Department of Mathematics, University of Stellenbosch, 7602 Stellenbosch, South Africa
E-mail address: hproding@sun.ac.za