Journal of Automata, Languages and Combinatorics 6 (2001) 3, 313–343

ON THE EXPECTED NUMBER OF LEFTIST NODES IN SIMPLY GENERATED TREES

RAINER KEMP
Fachbereich Biologie und Informatik, Johann Wolfgang Goethe-Universität
D-60054 Frankfurt am Main, Germany
e-mail: kemp@sads.informatik.uni-frankfurt.de

ABSTRACT

A node x appearing in an ordered tree is said to be a leftist node if in the subtree with root x, the leaf nearest to the root is the leftmost leaf of that subtree. Assuming that all trees belonging to a family of simply generated trees with a specified number of nodes and leaves are equally likely, we present a general approach to the computation of the average number of leftist nodes appearing in such a tree. If the number of nodes in a simply generated tree are specified only, that approach yields exact asymptotical formulae for the average number of leftist nodes, for the average number of internal leftist nodes and for the variances of the corresponding random variables. We illustrate these general results by applying them to various families of simply generated trees, such as t-ary trees, binary trees, unary–binary trees, unbalanced 2,3-trees, ordered trees, ordered trees without unary nodes and ordered trees with even node degrees.

Keywords: simply generated trees, leftist nodes, average-case analysis.

1. Introduction and Preliminaries

We recall that a family $\mathcal{F}$ of rooted trees is said to be *simply generated* if the generating function $E(z) := \sum_{n\geq 1} t(n)\, z^n$ of the number $t(n)$ of all trees $T \in \mathcal{F}$ with n nodes satisfies a functional equation of the form $E(z) = z\,\Theta(E(z))$, where $\Theta(y) := \sum_{\lambda\geq 0} c_\lambda\, y^\lambda$ is a regular function when $|y| < R < \infty$ with $c_0 = 1$, $c_\lambda \geq 0$ for $\lambda \in \mathbb{N}$, and $c_\lambda > 0$ for some $\lambda \in \mathbb{N}\backslash\{1\}$ [14]. This definition obviously includes the most common classes of trees such as *t-ary trees* ($\Theta(y) := 1 + y^t$, $t \in \mathbb{N}\backslash\{1\}$), *extended binary trees* ($\Theta(y) := 1 + y^2$), *binary trees* ($\Theta(y) := (1 + y)^2$), *unary-binary trees* ($\Theta(y) := 1 + y + y^2$), *unbalanced* 2-3-*trees* ($\Theta(y) := 1 + y^2 + y^3$) and *ordered trees* ($\Theta(y) := (1 - y)^{-1}$). Given the regular function Θ, the corresponding simply generated family of trees $\mathcal{F}(\Theta)$ is completely characterized: the elements of the set $\mathcal{DEG}(\Theta) := \{\lambda \in \mathbb{N}_0 \mid \langle y^\lambda; \Theta(y)\rangle \neq 0\}$[1] are the allowed node degrees in the trees appearing in $\mathcal{F}(\Theta)$ and $c_\lambda = \langle y^\lambda; \Theta(y)\rangle$, $\lambda \in \mathbb{N}_0$, reflects whether different orderings of the λ edges of a node of degree λ are taken into account in distinguishing between the trees in $\mathcal{F}(\Theta)$. Here, the *degree* $\deg(x)$ of a node x is the number of its sons.

[1]The abbreviation $\langle z_1^{n_1} \ldots z_m^{n_m}; f(z_1, \ldots, z_m)\rangle$ denotes the coefficient of $z_1^{n_1} \ldots z_m^{n_m}$ in the expansion of $f(z_1, \ldots, z_m)$ at $(z_1, \ldots, z_m) = (0, \ldots, 0)$.