Limit Laws for Terminal Nodes in Random Circuits with Restricted Fan-Out: A Family of Graphs Generalizing Binary Search Trees

Hosam Mahmoud[1] **Tatsuie Tsukiji**[2]

Abstract. We introduce a family of graphs $C(n, i, s, a)$ that generalizes the binary search tree. The graphs represent logic circuits with fan-in i, restricted fan-out s, and arising by n progressive additions of random gates to a starting circuit of a isolated nodes. We show via martingales that a suitably normalized version of the number of terminal nodes in binary circuits converges in distribution to a normal random variate.

Keywords: random graph, combinatorial probability.

1 Introduction

The well-known binary search tree has many applications as a data structure (see Mahmoud (1992)), as a model underlying searching and sorting applications (see Knuth 1998) and Mahmoud (2000)), and as a model for formal languages and computer algebra (see Kemp (1984)). We introduce a family of acyclic directed graphs $C(n, i, s, a)$ that generalizes the binary search tree.

The graphs represent logic circuits with indegree (fan-in) i, restricted outdegree (fan-out) s, arising by n progressive additions of random gates to a starting circuit of a isolated nodes. The initial a nodes represent the initial input lines. The model has applications in neurosciences (see Valiant (1994)) and electrosciences (see Hutton, Rose, Grossman and Cornell (1998)).

The family of graphs $C(n, i, s, a)$ we introduce is a hierarchy of combinatorial structures that includes the binary search tree. We shall refer to a circuit with fan in i and restricted fan-out s as the i-ary circuit with fanout s. For specified parameters i, s, a, we refer to the structure simply as the *circuit*. Nodes with outdegree s are considered saturated. At each stage, i insertion positions are chosen from unsaturated nodes as the parents of a new child. The outdegree of each of these nodes increases by the number of insertion positions taken from it. In circuit interpretation, i output lines of a previous stage are taken as input lines into a new i-ary gate, which can have up to s output lines drawn from it. The consumed lines are no longer viable inputs at later stages.

Let $L_{n,i,s,a}$ denote the number of terminal nodes in the circuit having a initial input nodes. The main result of this paper is to demonstrate the convergence in distribution of a suitably normed version of $L_{n,i,s,a}$ to a normal variate for binary circuits. The algebra gets quite unwieldy, if we keep all the parameters. For clarity of the exposition, we shall illustrate the result on the subfamily $C(n, 2, 3, 1)$ of binary fan-in, fan-out 3, and growing out of a single node. Throughout all insertion stages, the underlying undirected graph remains one connected component; the fan-out is restricted to 3. We shall return at the end to the general binary case and state the necessary adjustments in the proof to get a general result for the family $C(n, 2, s, a)$ and make some conjectures about what to expect in the class $C(n, i, s, a)$.

Throughout, we shall use the following notation. We shall denote the normally distributed random variate with mean 0 and variance σ^2 by $\mathcal{N}(0, \sigma^2)$. We shall use the symbols $\xrightarrow{\mathcal{D}}$ and $\xrightarrow{P}$ for convergence in distribution and in probability, respectively. The notation $O_{\mathcal{L}_1}(g(n))$ will stand for a random variable that is $O(g(n))$ in the $\mathcal{L}_1$ norm.

Let the notation L_n be reserved for $L_{n,2,3,1}$. One main result of this investigation is the central limit tendency:

$$\frac{L_n - \frac{1}{7}n}{\sqrt{n}} \xrightarrow{\mathcal{D}} \mathcal{N}\Big(0, \frac{24}{637}\Big),$$

and its extension to a similar central limit theorem for $L_{n,2,s,a}$.

Section 2 gives a precise definition for the circuits. In Section **??** we derive the exact first two moments for the number of terminal nodes for $C(n, 2, 3, 1)$. The central limit theorem for $C(n, 2, 3, 1)$ is derived in Section **??** via a martingale transform. The scope of the result and proof technique is extended in Section **??** to the binary circuits family $C(n, 2, s, a)$, where a conjecture for $C(n, i, s, a)$ is presented, too.

[1] Department of Statistics, The George Washington University, Washington, D.C. 20052, U.S.A. *E-mail address*: hosam@gwu.edu

[2] Department of Information Sciences, Tokyo Denki University Hatoyama-cho, Saitama-ken 350–0394, Japan. *E-mail address*: tatsuie@ybb.ne.jp

2 The growth of a random circuit

An *i-ary circuit with fan-out s* is a directed graph that starts out with a isolated nodes of indegree and outdegree 0. Each node can have up to s positional children (output lines in the language of circuits), with a distinguished left-to-right order. A childless node is a *terminal node*. A nonterminal node with k children has $s-k$ *positional insertion places*, which are the positions not taken by its positional children in the left-to-right order. The circuit evolves in stages as follows. After $n-1$ stages, a circuit $C(n-1,i,s,a)$ has grown. At the nth stage, i positional insertion places from unsaturated nodes are chosen from $C(n-1,i,s,a)$ as parents for a new node (or gate in the circuits). The new node is adjoined to the circuit with edges directed from the i positional insertion places to it, and is given 0 outdegree.

With regard to the interpretation of $C(n,i,s,a)$ as a circuit, the number of terminal nodes is a parameter of interest. In boolean circuits they stand for how many "answers" can be derived from a given inputs. The problem is trivial for $s=1$. The case $i=1$ is not very challenging. The least i that makes the problem significant is 2. The case $C(n,2,2,a)$ is not too interesting, as the number of terminal nodes remains finite and at most a throughout all the stages. The smallest significant member of the family $C(n,2,s,a)$ for the study is $C(n,2,3,a)$. For simplicity of the exposition we take $a=1$ to reduce the number of parameters. We shall focus on the binary circuit (fan-in $i=2$), with fan-out 3, growing out of a single input. We develop results for $C(n,2,3,1)$ in this section. In Section **??** we shall sketch the extension of these results to cover the entire $C(n,2,s,a)$ family.

The graphs $C(n,i,s,a)$ generalize binary search trees (which are $C(n,1,2,1)$). This runs in a parallel vein to the generalization of recursive trees into a hierarchy of random graphs called recursive circuits (Tsukiji and Mahmoud (2001)). For a broad review of definition and uses of recursive trees the reader is referred to the survey in Smythe and Mahmoud (1996).

For modeling and analysis purposes, the random graphs $C(n,i,s,a)$ are *extended* by supplying each node with a sufficient number of special nodes (called *external nodes*) to universally make the outdegree of all the circuit nodes (now viewed as *internal*) equal to s.

Many models of randomness can be imposed on i-ary circuits with fan-out s. A natural probability model is one in which all pairs of external nodes in $C(n-1,i,s,a)$ are equally likely candidate inputs for the nth entrant. Note that under this model, the various circuits of one size are not equally likely.

References

[1] Hall, P. and Heyde, C. (1980). *Martingale Limit Theory and Its Applications*. Academic Press, New York.

[2] Hutton, M., Rose, J., Grossman, J. and Cornell D. (1998). Characterization and Parameterized Generation of Synthetic Combinational Benchmark Circuits. *IEEE Trans. on CAD*, **17**, 985–996.

[3] Kemp, R. (1984). *Fundamentals of the Average Case Analysis of Particular Algorithms*. Wiley-Teubner Series in Computer Science, Wiley, New York.

[4] Knuth, D. (1998). *The Art of Computer Programming*, Vol. 3: *Sorting and Searching*, 2nd ed. Addison-Wesley, Reading, Massachusetts.

[5] Mahmoud, H. (1992). *Evolution of Random Search Trees*. Wiley, New York.

[6] Mahmoud, H. (2000). *Sorting: A distribution Theory*. Wiley, New York.

[7] Smythe, R. and Mahmoud, H. (1996). A survey of recursive trees. *Theory of Probability and Mathematical Statistics*, **51, 1–29**.

[8] Tsukiji, T. and Mahmoud, H. (2001). A limit law for outputs in random circuits. *Algorithmica*, **31**, 403–412.

[9] Valiant, L. (1994). *Circuits of the Mind*. Oxford University Press, Oxford, UK.