On the internal structure of random recursive circuits

Hosam Mahmoud\textsuperscript{a}*, Tatsuie Tsukiji\textsuperscript{b}

\textsuperscript{a}Department of Statistics, The George Washington University, 2201 G Street NW #315, Washington, DC 20052, USA
\textsuperscript{b}School of Informatics and Sciences, Nagoya University, Nagoya 464-8601, Japan

Received 12 September 2000; received in revised form 13 April 2001

Abstract

We study the joint probability distribution of the number of nodes of fan-out \(k\) in random recursive circuits. For suitable norming we obtain a limiting multivariate normal distribution for the numbers of node of fan-out at most \(k\), where we compute explicitly the limiting covariance matrix by solving a recurrence satisfied among its entries. © 2002 Elsevier Science B.V. All rights reserved.

\textit{MSC:} 60C05; 69F05; 05C80

\textit{Keywords:} Random structure; Multivariate central limit theorem; Fan-out; Joint probability distribution

1. Introduction

We study a class of random graphs with fixed in-degree and unlimited out-degree. Toward a logic circuit interpretation, we call the in-degree \textit{fan-in} and the out-degree \textit{fan-out}. Nodes with fan-in 0 in the graph are \textit{inputs}, whereas nodes of fan-out 0 are \textit{outputs}.

A recursive circuit of size \(n\) is a directed acyclic graph, the nodes of which are labeled with \(\{1, \ldots, n\}\) in such a way that labels increase along every input–output path.

We study the joint asymptotic distribution of the number of nodes of fan-out \(i\), for \(i = 0, \ldots, k\), in a binary recursive circuit, where the in-degree of all non-input nodes is 2. (The number \(k\) is arbitrary, but fixed, as \(n \to \infty\).)

Stratification of a random structure into a profile of node degrees has recently been a popular subject. Several classes of trees have been studied from this point of view. For example, Meir and Moon [14] calculated the probability that a recursive tree has no node of out-degree 1, Mahmoud

* Corresponding author. Fax: +1-202-994-6917.

\textit{E-mail addresses:} hosam@gwu.edu (H. Mahmoud), tsukiji@info.human.nagoya-u.ac.jp (T. Tsukiji).

0377-0427/02 - see front matter © 2002 Elsevier Science B.V. All rights reserved.

PH: S0377-0427(01)00466-6
and Smythe [10,11] Lew and Mahmoud [8], Mahmoud et al. [12], Mahmoud [9], Kemp [7], and Prodinger [15] have studied joint distributions of node degrees in various classes of random trees.

The binary recursive circuit can be viewed as the second member in a hierarchy of combinatorial structures, with the recursive forest being the first in the hierarchy. A recursive tree is of course the unit of a forest. (See the survey by Smythe and Mahmoud [16].) Circuits, however, are harder to analyze than randomly grown trees because of stronger types of dependency that appear in random circuits and have no counterpart in random trees. For instance, forests enjoy a decomposition property, admitting easy recurrence formulation, that has no analog in circuits—when a forest node is deleted, the result is a forest consisting of trees and subtrees of the original forest, whereas deletion of a node in a binary circuit does not necessarily result in binary circuits and subcircuits. A characterizing property of the complexity of a member in this hierarchy is the fan-in \( f \). Recursive forests are the case \( f = 1 \), and binary circuits are the case \( f = 2 \).

The present work thus shows how results on joint distributions in trees, especially recursive trees, can be extended to more complex members of the hierarchy. We present the result for binary circuits; extension for \( f \geq 2 \) does not seem to add new difficulty; for example, the recent manuscript of Johnson et al. [6] handles exact distributions in the univariate case studying only outputs for arbitrary \( f \geq 2 \). Recurrences in the case \( f = 2 \) extend to more elaborate forms leading to hypergeometrics in the solution, but one still obtains linear means and variances.

We shall denote the \((k+1)\)-variate normally distributed random row vector \( (X_0, \ldots, X_k) \) with mean \( M_k = (E[X_0], \ldots, E[X_k]) \) and covariance

\[
\Sigma_k = (\text{Cov}[X_i, X_j])_{0 \leq i, j \leq k},
\]

by \( \mathcal{N}_{k+1}(M_k, \Sigma_k) \). We shall use the symbols \( \overset{D}{\to} \) and \( \overset{P}{\to} \) for convergence in distribution and in probability, respectively. The notation \( o_P(1) \) will denote an asymptotically negligible random variable that converges to 0 in probability.

As we shall see in Section 2, the random circuit grows in stages. Let \( L_{n}^{(i)} \) be the number of nodes of fan-out \( i \) in a random binary circuit at its \( n \)th stage of growth. Let \( L_{n}^{(k)} \) be the row vector \( (L_{n}^{(0)}, \ldots, L_{n}^{(k)}) \). The main result of this paper is the multivariate central limit theorem:

\[
\frac{L_{n}^{(k)} - M_k n}{\sqrt{n}} \overset{D}{\to} \mathcal{N}_{k+1}(0_k, \Sigma_k),
\]

where \( 0_k \) represents a row vector of \( k+1 \) zeros; the \((k+1)\)-component row vector \( M_k \) is ultimately found to be given by a simple formula, and the covariance matrix \( \Sigma_k \) will be computed explicitly from a recurrence among its entries.

Our technical setup will also allow us to compute the rates of convergence of the (scaled) sequences of means and covariances to their limits in the following form:

\[
E[L_{n}^{(k)}] = \frac{2^k}{3^{k+1}} n + O(1),
\]

and, for explicitly computed constants \( D_{ij} \),

\[
\text{Cov}[L_{n}^{(i)}, L_{n}^{(j)}] = D_{ij} n + O(1).
\]

The paper is organized as follows. In Section 2 the growth of a random circuit and the associated probability model are discussed both as a stochastic process in discrete time and as a combinatorial
sample space. In Section 3 both the expectation of the number of nodes of fan-out \( k \), as well as covariances between the number of nodes of fan-out \( i \) and the number of those of fan-out \( j \) are discussed. The limiting covariance matrix is obtained by solving a recurrence satisfied by its entries. In Section 4, arbitrary linear combinations of the numbers of nodes of fan-out up to \( k \) are shown to be a part of a martingale, the other part of which behaves in a similar way, a result sufficient for a multivariate central limit theorem. Section 5 concludes the paper with some remarks.

2. The growth of a random circuit

A recursive circuit with fan-in \( f \) grows randomly as follows. The circuit starts out with \( a \geq f \) isolated inputs, labeled 1, \ldots, \( a \), and evolves in stages. After \( n - 1 \) stages, a circuit \( RC_{n-1} \) has grown. At the \( n \)th stage, \( f \) distinct nodes are chosen from \( RC_{n-1} \) as parents for a new entrant labeled \( n + a \). The new node is joined to the circuit with edges directed from the \( f \) parents to it, and is given 0 out-degree forming the circuit \( RC_n \). A recursive forest corresponds to the case \( f = 1 \) (see Balińska et al. [1]). The building block of a recursive forest is the recursive tree, which grows out of a single node. The recursive tree has been a popular topic in both probability and computer science (see Smythe and Mahmoud [16] and the many references therein).

Fig. 1 below shows all the possible binary circuits after two insertion steps into an initial graph of two isolated nodes of fan-out 0. (We shall focus on binary circuits, but anticipate no additional difficulty for higher fan-in.) We impose a probability distribution induced by a growth process that chooses two distinct parents uniformly at random from all existing nodes. It can be easily argued that the growth after \( n \) insertions according to this stochastic view is equivalent to a sample space of all recursive circuits of size \( n + a \), where each circuit is equally likely. The stochastic growth viewpoint allows us to come up with a recursive formulation amenable to a probabilistic approach. On the other hand, we speculate that the uniform probability space viewpoint may admit a generating function formulation amenable to the methods of analytic combinatorics; see Flajolet and Soria [4] for work in random graphs along this alternative methodology.

In the circuit \( RC_n \), let the set of nodes of fan-out \( i \) be denoted by \( L_n^{(i)} \), and so \( L_n^{(i)} \) is its cardinality. It is convenient for our purpose to study the number of nodes of fan-out \( up \ to \ and \ including \) \( k \), for which we use the notation \( L_n^{(\leq k)} = \sum_{i=0}^{k} L_n^{(i)} \). The set \( L_n^{(k)} = L_n^{(\leq k)} - L_n^{(\leq k-1)} \) plays an important role in our recurrence formulation; let us call this set the boundary. Nodes of fan-out \( k \) are on the boundary; nodes of fan-out lower (higher) than \( k \) fall in sets below (above) the boundary.

![Fig. 1. All binary circuits of size 4 grown from two inputs. Square nodes are outputs.](image-url)
3. Expectation and covariance of fan-outs

Fix $k \geq 0$ arbitrarily. For the $n$th stage, we calculate the net gain in the number of nodes of fan-out $k$ or less. We say that an insertion at the $n$th stage causes a node to cross the boundary, if the node is on or below the boundary of $RC_{n-1}$, but lies above the boundary of $RC_n$. At the $n$th stage choosing two parents from the nodes of $RC_{n-1}$ can occur in one of three ways:

(i) Both parents belong to the boundary of $RC_{n-1}$. Let $I_n^{(k)}$ be the indicator of this event. In this case, the circuit acquires a new output (of fan-out 0 $\leq k$ and falling on or below the boundary of $RC_n$), while letting the two parent nodes in $RC_{n-1}$ cross the boundary, a net gain of $-1$ in $L_n^{(\leq k)}$. That is,

$$L_n^{(\leq k)} = I_{n-1}^{(\leq k)} - 1.$$  

(ii) Both parents do not belong to the boundary. Let $J_n^{(k)}$ be the indicator of this event. We have three possibilities: the two parents are below the boundary, both are above, or one is below, the other is above. In any of these cases, one new output appears without letting any parent cross the boundary, and so the net gain in $L_n^{(\leq k)}$ is $+1$ and we have

$$L_n^{(\leq k)} = I_{n-1}^{(\leq k)} + 1.$$  

(iii) One parent belongs to the boundary, the other does not. The evolving circuit has only one boundary-crossing parent and gains one new output, a net gain of $-1 + 1 = 0$ in $L_n^{(\leq k)}$, implying

$$L_n^{(\leq k)} = I_{n-1}^{(\leq k)}.$$  

These three relations can be expressed in terms of the associated indicators:

$$L_n^{(\leq k)} = I_{n-1}^{(\leq k)} + J_n^{(k)} - I_n^{(k)}. \tag{1}$$

The gain machinery in (i) and (ii) allows us to calculate the conditional expectations of $I_n^{(k)}$ and $J_n^{(k)}$, respectively, given the previous $n-1$ steps. Let $F_n$ be the sigma field generated by the first $n$ insertions. Supposing that an event in $F_{n-1}$ has happened, we have

$$E[I_n^{(k)}|F_{n-1}] = \frac{L_{n-1}^{(\leq k)} - L_{n-1}^{(\leq k-1)}}{2}, \tag{2}$$

$$E[J_n^{(k)}|F_{n-1}] = \frac{n + a - 1 - L_{n-1}^{(\leq k)} + L_{n-1}^{(\leq k-1)}}{2}. \tag{3}$$

We also take the conditional expectation of (1) and obtain

$$E[L_n^{(\leq k)}|F_{n-1}] = I_{n-1}^{(\leq k)} + E[J_n^{(k)}|F_{n-1}] - E[I_n^{(k)}|F_{n-1}]. \tag{4}$$
Substituting (2) and (3) into (4) yields
\[ E[L_n^{(\leq k)} | \mathcal{F}_{n-1}] = \frac{n + a - 3}{n + a - 1} L_n^{(\leq k)} + \frac{2}{n + a - 1} L_n^{(\leq k-1)} + 1. \] (5)

**Theorem 1.** For fixed \( k \),
\[ E[L_n^{(k)}] = \frac{2^k}{3^{k+1}} (n + a) + O \left( \frac{1}{n} \right), \]
as \( n \to \infty \).

**Proof.** It suffices for our purpose to show that
\[ \left[ 1 - \left( \frac{2}{3} \right)^{k+1} \right] (n + a) + \frac{\rho_1(k)}{n} \leq E[L_n^{(\leq k)}] \leq \left[ 1 - \left( \frac{2}{3} \right)^{k+1} \right] (n + a) + \frac{\rho_2(k)}{n}, \]
for any functions \( \rho_1 \) and \( \rho_2 \) that depend only on \( k \) (which is held fixed, while \( n \to \infty \)). The expectation in the theorem will follow after taking the difference \( E[L_n^{(\leq k)}] - E[L_n^{(\leq k-1)}] \).

Taking expectation of Eq. (5) gives
\[ E[L_n^{(\leq k)}] = \frac{n + a - 3}{n + a - 1} E[L_n^{(\leq k)}] + \frac{2}{n + a - 1} E[L_n^{(\leq k-1)}] + 1. \]

The lower bound is easy to check because in fact \( \rho_1(k) \equiv 0 \) will do the job. That is, we shall show that
\[ E[L_n^{(\leq k)}] \geq \left[ 1 - \left( \frac{2}{3} \right)^{k+1} \right] (n + a). \]

We prove this lower bound by a double induction. The basis of this induction is the initial condition (in \( k \))
\[ E[L_n^{(\leq \ell)}] = a \geq \left[ 1 - \left( \frac{2}{3} \right)^{\ell+1} \right] a \quad \text{for } \ell = 0, \ldots, k, \]
and the initial condition (in \( n \))
\[ E[L_n^{(\leq 0)}] \geq \frac{n + a}{3}, \]
the latter condition can be established from an exact expression for the average number of outputs in [13].

Assume the assertion holds for all \( \ell \leq k \) up to \( n - 1 \), and further at \( n \) it holds for \( \ell = 0, \ldots, k - 1 \). It then follows that
\[ E[L_n^{(\leq k)}] \geq \frac{n + a - 3}{n + a - 1} \left[ 1 - \left( \frac{2}{3} \right)^{k+1} \right] (n + a - 1) + \frac{2}{n + a - 1} \left[ 1 - \left( \frac{2}{3} \right)^{k} \right] (n + a - 1) + 1 \]
\[ = \left[ 1 - \left( \frac{2}{3} \right)^{k+1} \right] (n + a). \]
A similar argument holds for an upper bound with, say, $\rho_2(k) = 2^k a(a - 1)$. □

**Remark.** The case $E[L_n^{(\leq 0)}] = E[L_n^{(0)}] \sim \frac{1}{n}$ is the average number of outputs and, of course, agrees with the calculation in [13], and with Dondajewski and Szymański [3], who arrived at this result by a different approach. (The latter reference considers only average-case analysis.)

We now turn to evaluating the covariance between $L_n^{(\leq k)}$ and $L_n^{(\leq k')}$ for arbitrarily non-equal fixed $k$ and $k'$. We have derived $E[I_n^{(\ell)}|\mathcal{F}_{n-1}]$ and $E[J_n^{(\ell)}|\mathcal{F}_{n-1}]$ for $\ell \in \{k, k'\}$ as the respective conditional probabilities of the events (i) and (ii) happening at the $n$th insertion, given a previous history in $\mathcal{F}_{n-1}$. The covariance calculation will require us to further evaluate the conditional expectation of the four quadratic quantities, $E[A_n^{(k)} B_n^{(k')}|\mathcal{F}_{n-1}]$ for each $A, B \in \{I, J\}$. Since the event (i) happens exclusively for different $k$ and $k'$, $I_n^{(k)} I_n^{(k')} = 0$ always holds. Consequently,

$$E[I_n^{(k)} I_n^{(k')}|\mathcal{F}_{n-1}] = 0. \quad (6)$$

By contrast, $I_n^{(k)} = 1$ implies $J_n^{(k')} = 1$, so that $I_n^{(k)} J_n^{(k')} = I_n^{(k)}$. Consequently,

$$E[I_n^{(k)} J_n^{(k')}|\mathcal{F}_{n-1}] = E[I_n^{(k)}|\mathcal{F}_{n-1}]. \quad (7)$$

Symmetrically,

$$E[J_n^{(k)} I_n^{(k')}|\mathcal{F}_{n-1}] = E[I_n^{(k')}|\mathcal{F}_{n-1}]. \quad (8)$$

Finally, $J_n^{(k)} J_n^{(k')} = 1$ holds if and only if the two parents of the $n$th insertion belong to neither $\mathcal{Q}_n^{(k)}$ nor $\mathcal{Q}_n^{(k')}$, thus

$$E[J_n^{(k)} J_n^{(k')}|\mathcal{F}_{n-1}] = \frac{\left( n + a - 1 - L_n^{(\leq k)} + L_n^{(\leq k-1)} - L_n^{(\leq k')} + L_n^{(\leq k'-1)} \right)}{2}.$$

(9)

Conditional expectation of the product $L_n^{(\leq k)} L_n^{(\leq k')}$ can be developed starting from (1) for nonequal $k$ and $k'$:

$$E[L_n^{(\leq k)} L_n^{(\leq k')}|\mathcal{F}_{n-1}] = E[(L_n^{(\leq k)} + J_n^{(k)} - I_n^{(k)})(L_n^{(\leq k')} + J_n^{(k')} - I_n^{(k')})|\mathcal{F}_{n-1}].$$

Simplifying cross-products according to (6)–(9), then taking expectation over $\mathcal{F}_{n-1}$ gives

$$E[L_n^{(\leq k)} L_n^{(\leq k')}] = \frac{(n + a - 3)(n + a - 4)}{(n + a - 1)(n + a - 2)} E[L_n^{(\leq k)}] E[L_n^{(\leq k')}]
+ \frac{2(n + a - 3)}{(n + a - 1)(n + a - 2)} E[L_n^{(\leq k-1)}] E[L_n^{(\leq k')}]
+ \frac{2(n + a - 3)}{(n + a - 1)(n + a - 2)} E[L_n^{(\leq k)}] E[L_n^{(\leq k'-1)}]
+ \frac{2}{(n + a - 1)(n + a - 2)} E[L_n^{(\leq k-1)}] E[L_n^{(\leq k'-1)}].$$
Both systems are to be solved under the initial condition

\[ \text{recurrence in the case} \]

\[ \text{and} \]

We only sketch a double inductive proof on

\[ \text{where} \]

\[ C_{kk'} \]

If \( k = k' \), some cross-products are different. For example, while \( I_k I_k' \) vanishes in the case \( k \neq k' \), this product is \( I_k \) in the case \( k = k' \), which is not necessarily 0. A few terms in the recurrence are therefore slightly different and an additional term in the solution for the case \( k = k' \) will appear. The recurrence in the case \( k = k' \) is

\[
\mathbb{E}[(L_n^{(\leq k)})^2] = \frac{(n+a-3)(n+a-4)}{(n+a-1)(n+a-2)} \mathbb{E}[(L_n^{(\leq k)})^2]
\]

\[ + \frac{4(n+a-3)}{(n+a-1)(n+a-2)} \mathbb{E}[L_n^{(\leq k-1)}L_n^{(\leq k)}]
\]

\[ + \frac{2}{(n+a-1)(n+a-2)} \mathbb{E}[(L_n^{(\leq k-1)})^2]
\]

\[ + \frac{2(n+a-3)}{n+a-2} \mathbb{E}[L_n^{(\leq k-1)}] + \frac{2}{n+a-2} \mathbb{E}[L_n^{(\leq k-1)}]
\]

\[ + 1. \quad (10) \]

Both systems are to be solved under the initial condition \( \mathbb{E}[L_0^{(k)}] = a^2 \); for \( k + k' = 0 \), both \( k \) and \( k' \) are 0 providing the boundary condition \( \mathbb{E}[L_n^{(\leq 0)}L_n^{(\leq 0)}] = \mathbb{E}[(L_n^{(0)})^2] \), which again can be found from an exact variance expression for outputs in [13].

**Proposition 1.** Fix an arbitrary constant \( k_0 \geq 0 \). For each \( 0 \leq k, k' \leq k_0 \),

\[
\mathbb{E}[L_n^{(\leq k)}L_n^{(\leq k')} = \left[ 1 - \left( \frac{2}{3} \right)^{k+1} \right] \left[ 1 - \left( \frac{2}{3} \right)^{k'+1} \right] (n+a)^2 + C_{kk'}n + O(1), \quad (11)
\]

where \( C_{kk'} \) is given by the solution to the recurrence equations

\[
C_{kk'} = \begin{cases} 
\frac{2}{5}(C_{k-1,k' + 1} + C_{k-1,k'}) - \frac{2}{45}(\frac{3}{2})^{k+k'} & \text{if} \ k \neq k' , \\
\frac{4}{5}C_{k-1,k'} - \frac{2}{45}(\frac{3}{2})^{2k} + \frac{2}{15}(\frac{3}{2})^{k} & \text{if} \ k = k' 
\end{cases}
\]

with boundary conditions: \( C_{k-1} = C_{-1,k} = 0 \), for all \( k \leq k_0 \).

**Proof.** We only sketch a double inductive proof on \( k + k' \) and \( n \), as it runs in the same vein as the proof of Theorem 1. So, we prove an upper and lower bound in the form of (11), with the \( O(1) \) term substituted for by suitable constants. Let us take up the lower bound. After checking trivial initial conditions, suppose that, for a suitable constant \( K \),

\[
\mathbb{E}[L_n^{(\leq \ell)}L_n^{(\leq \ell')} \geq \left[ 1 - \left( \frac{2}{3} \right)^{\ell+1} \right] \left[ 1 - \left( \frac{2}{3} \right)^{\ell'+1} \right] n^2 + C_{\ell \ell'}n + K(\ell + \ell')
\]
is satisfied for every \( \ell, \ell' \leq k_0 \), with \( \ell + \ell' \leq k + k' \) up to \( n - 1 \), and for \( n \) it is satisfied up to \( \ell + \ell' = k + k' - 1 \).

If \( k \neq k' \), all the terms on the right-hand side of (10) are either in the induction hypothesis or asymptotically known from Theorem 1, giving

\[
\mathbf{E}[L_n^{(\leq k)}L_n^{(\leq k')}] \geq \left[ 1 - \left( \frac{2}{3} \right)^{k+1} \right] \left[ 1 - \left( \frac{2}{3} \right)^{k'+1} \right] (n + a)^2 + C_{kk'} n

+ K(k + k') - 5C_{kk'} + 2(C_{k,k'-1} + C_{k-1,k'}) + \frac{2}{9} \left( \frac{2}{3} \right)^{k+k'}.
\]

The induction is completed upon choosing the constants \( C_{ij} \) as in the theorem. A parallel argument for the upper bound holds.

Else one has \( k = k' \); the argument for the second moment is essentially similar. \( \Box \)

The recurrence in Proposition 1 can be solved exactly via generating functions. Let \( F(x, y) \) be the generating function of the sequence \( C_{kk'} \); that is

\[
F(x, y) = \sum_{0 \leq k, k' \leq \infty} C_{kk'} x^k y^{k'}.
\]

The double-decker recurrence in Proposition 1 can be concisely written in one line; the second line has one additional term that appears only in the case \( k = k' \):

\[
C_{kk'} = \frac{2}{5} (C_{k,k'-1} + C_{k-1,k'}) - \frac{2}{45} \left( \frac{2}{3} \right)^{k+k'} + \frac{2}{15} \left( \frac{2}{3} \right)^k I_{\{k=k'\}},
\]

where \( I_{\mathcal{B}} \) is the indicator of the condition \( \mathcal{B} \) that assumes the value 1 if \( \mathcal{B} \) holds and assumes the value 0 otherwise.

Multiply both sides of the recurrence by \( x^k y^{k'} \) and rearrange in the form

\[
F(x, y) = \frac{2}{5} \left[ y \sum_{k=0}^{\infty} \sum_{k'=1}^{\infty} C_{k,k'-1} x^k y^{k'-1} + x \sum_{k=1}^{\infty} \sum_{k'=0}^{\infty} C_{k-1,k} x^{k-1} y^{k'} \right]

- \frac{2}{45} \sum_{0 \leq k, k \leq \infty} \left( \frac{2}{3} \right)^k \left( \frac{2}{3} \right)^{k'} + \frac{2}{15} \sum_{j=0}^{\infty} \left( \frac{2}{3} x y \right)^j

= \frac{2}{5} \left[ yF(x, y) + xF(x, y) \right] - \frac{2}{45(1 - \frac{2}{3} x)(1 - \frac{2}{3} y)} + \frac{2}{15(1 - \frac{2}{3} xy)}.
\]

Collecting terms with \( F(x, y) \) on one side, we get an explicit equation for this generating function

\[
F(x, y) = \frac{2}{15(1 - \frac{2}{3} xy)(1 - \frac{2}{3}(x + y))} - \frac{2}{45(1 - \frac{2}{3} x)(1 - \frac{2}{3} y)(1 - \frac{2}{3}(x + y))}.
\]
Lemma 1. For each $0 \leq k, k' \leq k_0,$

$$C_{kk'} = \frac{2}{15} \left(\frac{2}{3}\right)^{(k+k')/2} \left[1_{\{k \equiv k' \mod 2\}} \sum_{m=0}^{k+k'} \left(\frac{2}{3}\right)^{-(1/2)m} \left(\frac{2}{5}\right)^m \left(\frac{1}{2}(k-k'+m)\right)\right]$$

$$+ 1_{\{k \equiv k' - 1 \mod 2\}} \sum_{m=1}^{k+k'} \left(\frac{2}{3}\right)^{-(1/2)m} \left(\frac{2}{5}\right)^m \left(\frac{1}{2}(k-k'+m)\right)$$

$$- \frac{2}{45} \left(\frac{2}{3}\right)^{k+k'} \sum_{m=0}^{k+k'} \left(\frac{3}{5}\right)^m \sum_{r=\max\{0,m-k'\}}^{\min\{m,k\}} \binom{m}{r}.$$ 

Proof. This lemma is proved by routine recovery of coefficients from a generating function. The negative term in $F(x, y)$ is

$$- \frac{2}{45(1 - \frac{2}{3}x)(1 - \frac{2}{3}y)(1 - \frac{2}{3}(x+y))}$$

$$= - \frac{2}{45} \left[\sum_{i=0}^{\infty} \left(\frac{2}{3}x\right)^i\right] \left[\sum_{j=0}^{\infty} \left(\frac{2}{3}y\right)^j\right] \left[\sum_{m=0}^{\infty} \left(\frac{2}{5}\right)^m \sum_{r=0}^{m} x^r y^{m-r} \binom{m}{r}\right].$$

To obtain the coefficient of $x^k y^{k'}$, we take the coefficient of $x^r y^{m-r}$ in the innermost sum and match it with $(\frac{2}{3})^{k'-r}$ from the first sum, and $(\frac{2}{3})^{k'-(m-r)}$ from the middle sum, for feasible $m$ and $r$. For each feasible $m$, $r$ must meet the constraints

$$0 \leq r, \quad r \leq m, \quad r \leq k, \quad m - k' \leq r.$$ 

Each feasible $m$ contributes

$$\sum_{r=\max\{0,m-k'\}}^{\min\{m,k\}} \left(\frac{2}{3}\right)^{k+k'-m} \left(\frac{2}{5}\right)^m \binom{m}{r}. $$

For $m$ to be feasible, it has to fall in the range $[0, k + k']$. 

The positive term of the generating function is

$$\frac{2}{15(1 - \frac{2}{3}xy)(1 - \frac{2}{3}(x+y))} = \frac{2}{15} \sum_{j=0}^{\infty} \left(\frac{2}{3}xy\right)^j \sum_{m=0}^{j} \left(\frac{2}{5}\right)^m \sum_{r=0}^{m} x^r y^{m-r} \binom{m}{r}. $$

To obtain the coefficient of $x^k y^{k'}$ contributed by this part, we take the coefficient of $x^r y^{m-r}$ in the innermost sum and match it with $(2/3)^j$ from the first sum for feasible indexes. For feasible
there is only one feasible pair \((j, r)\) that solves the equations

\[
\begin{align*}
j + r &= k, \\
j + m - r &= k'.
\end{align*}
\]

Observing that the parity of feasible \(m\) follows that of \(k + k'\), the result follows upon rearranging. \(\square\)

**Corollary 1.** For fixed \(k, k' \geq 0\),

\[
\text{Cov}[L^{(\leq k)}_n, L^{(\leq k')}_n] = C_{kk'} n + O(1),
\]

as \(n \to \infty\).

**Proof.** The leading term in \(E[L^{(\leq k)}_n]E[L^{(\leq k')}_n]\) matches exactly that in the product \(E[L^{(\leq k)}_n]E[L^{(\leq k')}_n]\). Cancellation of quadratic terms occur leaving behind linear leading terms. \(\square\)

**Corollary 2.** For fixed \(k, k' \geq 0\),

\[
\text{Cov}[L^{(k)}_n, L^{(k')}_n] = D_{kk'} n + O(1),
\]

as \(n \to \infty\), where

\[
D_{kk'} = C_{kk'} - \sum_{0 \leq \ell \leq k} \sum_{0 \leq \ell' \leq k'} D_{\ell \ell'}.
\]

**Proof.** The covariance of interest can be extracted from \(\text{Cov}[L^{(\leq \ell)}_n, L^{(\leq \ell')}_n]\) by an iterative procedure. For each \(\ell \geq 0\), \(L^{(\leq \ell)}_n\) is the sum \(L^{(0)}_n + L^{(1)}_n + \cdots + L^{(\ell)}_n\). For \(k' \geq k\), take the covariance between \(L^{(\leq k)}_n\) and \(L^{(\leq k')}_n\), yielding

\[
\text{Cov}[L^{(\leq k)}_n, L^{(\leq k')}_n] = \sum_{0 \leq \ell \leq k} \sum_{0 \leq \ell' \leq k'} \text{Cov}[L^{(\ell)}_n, L^{(\ell')}_n].
\]

Having computed inductively the covariances for all pairs of indexes, except \((k, k')\), the ingredients on the right-hand side of this formula are known except \(\text{Cov}[L^{(k)}_n, L^{(k')}_n]\). The left-hand side is what has been computed in Corollary 1. \(\square\)

As an illustration, we give here the covariance matrix of the vector \(L^{(5)}_n = (L^{(0)}_n, \ldots, L^{(5)}_n)\) in exact rational form

\[
\text{Cov}[L^{(5)}_n] = \Sigma_5 n + O(1),
\]
where

\[ \Sigma_5 = \begin{pmatrix}
4 & -56 & -236 & -416 & 7504 & 145024 \\
45 & 675 & 10125 & 151875 & 2278125 & 34171875 \\
-56 & 675 & 10125 & 151875 & 2278125 & 34171875 \\
-236 & -6248 & 257524 & 6834375 & 102515625 & 145024 \\
-416 & -44984 & 8105576 & 1537734375 & 23066015625 & 3516302576 \\
-151875 & -2278125 & -6834375 & -102515625 & -1729951171875 & -688671875 \\
-7504 & -54976 & 1389568 & -21453952 & -14642160896 & -688671875 \\
2278125 & 6834375 & 102515625 & -1537734375 & -23066015625 & 34171875 \\
54171875 & 512578124 & 7688611875 & 11530078125 & 1729951171875 & 25949267578125
\end{pmatrix}, \]

the notation \( O(1) \) stands for a matrix, the coefficients of which are all \( O(1) \), as \( n \to \infty \). In particular, the variances \( \text{Var}[L_n^{(i)}] \) are asymptotically linear in \( n \). In the illustrating example we have

\[ \text{Var}[L_n^{(5)}] = \frac{958356569248}{25949267578125} n + O(1). \]

4. Joint distribution of fan-outs

We are poised to consider the asymptotic joint distribution of the components of \( L_n^{(k)} \), for arbitrarily fixed \( k \), as the circuit size grows to infinity. We shall work with recurrences for \( L_n^{(i)} \), for \( i = 0, \ldots, k \). These components all \( (i \geq 1) \) satisfy the recurrence

\[ \mathbb{E}[L_n^{(i)} \mid \mathcal{F}_{n-1}] = \frac{n + a - 3}{n + a - 1} L_n^{(i)} + \frac{2}{n + a - 1} L_n^{(i-1)}, \quad (12) \]

except the boundary case \( L_n^{(0)} \). This latter component satisfies the recurrence

\[ \mathbb{E}[L_n^{(0)} \mid \mathcal{F}_{n-1}] = \frac{n + a - 3}{n + a - 1} L_n^{(0)} + 1. \quad (13) \]

These recurrences can be obtained by taking differences of the recurrences on \( \mathbb{E}[L_n^{(\leq i)} \mid \mathcal{F}_{n-1}] \), for two successive values of the superscript \( i \geq 1 \) (cf. (5)). The case \( i = 0 \) has already been considered in [13].

We next describe the streamlined path that we shall follow. Toward a multivariate central limit theorem, we shall study linear combinations of \( L_n^{(i)} \). It will be shown that, for any arbitrary fixed constants \( \lambda_0, \ldots, \lambda_k \), the univariate random variable

\[ \lambda_0 L_n^{(0)} + \cdots + \lambda_k L_n^{(k)}, \]

when suitably normed, converges in distribution to a centered normal distribution. The multivariate result will then follow from the Cramér–Wold device (see [2, p. 44]).

To prove the normality of a normed version of the linear combination, we shall appeal to the martingale central limit theorem. Let us look at a centered version of the linear combination:

\[ W_n \overset{\text{def}}{=} \sum_{i=0}^{k} \lambda_i L_n^{(i)} - \sum_{i=0}^{k} \lambda_i \mathbb{E}[L_n^{(i)}]. \]
In what follows the symbol $\nabla$ stands for the backward difference operator, that is for any function $h$, $\nabla h_n = h_n - h_{n-1}$. Using recurrences (12) and (13), we can reorganize the conditional expectation of $W_n$ in the form
\[
E[W_n|\mathcal{F}_{n-1}] = \frac{n + a - 3}{n + a - 1} \hat{\lambda}_0(L^{(0)}_{n-1} - E[L^{(0)}_{n-1}]) + \frac{n + a - 3}{n + a - 1} \sum_{i=1}^{k} \hat{\lambda}_i(L^{(i)}_{n-1} - E[L^{(i)}_{n-1}]) \\
+ \frac{2}{n + a - 1} \sum_{i=1}^{k} \lambda_i(L^{(i-1)}_{n-1} - E[L^{(i-1)}_{n-1}])
\]
\[
= \frac{n + a - 3}{n + a - 1} W_{n-1} + \frac{2}{n + a - 1} \sum_{i=1}^{k} (\lambda_i - \lambda_{i-1} + \lambda_{i-1})(L^{(i-1)}_{n-1} - E[L^{(i-1)}_{n-1}])
\]
\[
= \frac{n + a - 3}{n + a - 1} W_{n-1} + \frac{2}{n + a - 1} \sum_{i=1}^{k} \lambda_{i-1}(L^{(i-1)}_{n-1} - E[L^{(i-1)}_{n-1}])
\]
\[
+ \frac{2}{n + a - 1} \sum_{i=1}^{k} (L^{(i-1)}_{n-1} - E[L^{(i-1)}_{n-1}]) \nabla \lambda_i
\]
\[
= \frac{n + a - 3}{n + a - 1} W_{n-1} + \frac{2}{n + a - 1}(W_{n-1} - \lambda_k(L^{(k)}_{n-1} - E[L^{(k)}_{n-1}]))
\]
\[
+ \frac{2}{n + a - 1} \sum_{i=1}^{k} (L^{(i-1)}_{n-1} - E[L^{(i-1)}_{n-1}]) \nabla \lambda_i
\]
\[
= W_{n-1} + A_n,
\]
where
\[
A_n = \frac{2}{n + a - 1} \sum_{i=1}^{k+1} (L^{(i-1)}_{n-1} - E[L^{(i-1)}_{n-1}]) \nabla \lambda_i
\]
($\lambda_{k+1}$ is assumed 0 in this definition). Thus, $W_n$ is almost a martingale, but not quite so owing to the presence of the nuance term $A_n$.

A few words about $A_n$ are in order. Firstly, note that $A_n$ depends entirely on random variables of the circuit $RC_{n-1}$. So, $A_n$ is $\mathcal{F}_{n-1}$-measurable. Secondly, $A_n$ is a linear combinations of differences of the form $L^{(i)}_{n-1} - E[L^{(i)}_{n-1}]$. Normed by $n$, each such difference converges to 0 in probability (this follows from the Chebyshev’s inequality and the order of the variance in Corollary 2).

To turn our almost-martingale into a true one, we introduce a correction or a martingale transform. Let
\[
M_n = W_n + B_n
\]
be a martingale. So
\[
E[M_n | \mathcal{F}_{n-1}] = E[W_n | \mathcal{F}_{n-1}] + E[B_n | \mathcal{F}_{n-1}]
\]
\[
= W_{n-1} + A_n + B_n.
\]
We want this to be \( M_{n-1} = W_{n-1} + B_{n-1} \). This is possible if we choose 
\[ B_n = B_{n-1} - A_n \]
for each \( n \), or more explicitly,
\[ B_n = -\sum_{j=1}^{n} A_j. \]

We shall show in the following two technical lemmas that the martingale difference \( n^{-1/2} \nabla M_j \) satisfies Lindeberg’s conditional condition and the conditional variance condition, the two building blocks in the martingale central limit theorem. The form of these conditions for our specific context is in the text below; the reader can refer to Hall and Heyde [5] for a broader representation.

**Lemma 2.** The martingale difference \( n^{-1/2} \nabla M_j \) satisfies Lindeberg’s conditional condition.

**Proof.** The difference \( \nabla M_j \) can be written as
\[ \nabla M_j = \nabla B_j + \nabla W_j = -A_j + \sum_{i=0}^{k} \lambda_i (\nabla L_j^{(i)} - \nabla E[L_j^{(i)}]). \]

Of course, after the \( n \)th insertion, the number of nodes of any particular fan-out is at most \( n + a \), and the absolute change in the number of nodes of any particular fan-out is at most \( 2 \). Therefore, the sum in \( \nabla M_j \) is \( O(1) \). The quantity \( A_j \) is also \( O(1) \). For large \( n \), the sets \( \{RC_n : |\nabla M_j| > \varepsilon \sqrt{n} \} \) are all empty, for all \( j \leq n \). Thus, now with \( 1_\mathcal{C} \) being the indicator of a condition \( \mathcal{C} \),
\[ \sum_{j=1}^{n} E \left[ \left\{ \frac{\nabla M_j}{\sqrt{n}} \mid 1_{\{ |\nabla M_j| > \varepsilon \sqrt{n} \}} \right\} \mid \mathcal{F}_{j-1} \right] \to 0, \]
conditional Lindeberg’s condition has been verified. \( \Box \)

**Lemma 3.** Let \( \Lambda_k' = (5\lambda_0 - 4\lambda_1, \ldots, 5\lambda_{k+1} - 4\lambda_k) \) and let \( \Lambda_k^T \) be its transpose. The martingale difference \( n^{-1/2} \nabla M_j \) satisfies the conditional variance condition
\[ \sum_{j=1}^{n} E \left[ \left( \frac{\nabla M_j}{\sqrt{n}} \right)^2 \right] \mathcal{F}_{j-1} \xrightarrow{p} \Lambda_k' \Sigma_k \Lambda_k^T, \]
where \( \Sigma_k \) is a \((k+1) \times (k+1)\) matrix of constant coefficients given by Corollary 2.

**Proof.** Start with the conditional expectation
\[ E \left[ \left( \frac{\nabla M_j}{\sqrt{n}} \right)^2 \right] \mathcal{F}_{j-1} = \frac{1}{n} E[ (\nabla W_j - A_j)^2 \mid \mathcal{F}_{j-1}] = \frac{1}{n} E[ (\nabla W_j)^2 + A_j^2 - 2A_j \nabla W_j] \mathcal{F}_{j-1}. \]
We have argued in the proof of Lemma 2 that the difference $\nabla W_j$ is unconditionally $O(1)$, consequently $\mathbb{E}[(\nabla W_j)^2|\mathcal{F}_{j-1}]$ and $\mathbb{E}[\nabla W_j|\mathcal{F}_{j-1}]$ are $O(1)$. The $\mathcal{F}_{n-1}$-measurable correction $A_j$ is itself $O(1)$. Hence the squared martingale difference is $O(1)$ and

$$\sum_{j=1}^{n} \mathbb{E} \left[ \left( \frac{\nabla M_j}{\sqrt{n}} \right)^2 \bigg| \mathcal{F}_{j-1} \right] = \sum_{j=1}^{n} \frac{O(1)}{n} = O(1).$$

The last $O(1)$ quantity is of course a function of $\Lambda'_k$; a more detailed argument (not shown) can be given to specify the constant in it as stated. The conditional variance condition has been verified. \qed

In view of Lemmas 2 and 3, the martingale difference $n^{-1/2}\nabla M_j$, satisfies the requirements of the martingale central limit theorem (see Corollary 3.1 in [5]). It follows that

$$\sum_{j=1}^{n} \nabla M_j = \frac{M_n}{\sqrt{n}} \xrightarrow{\mathcal{D}} N(0, \Lambda'_k \Sigma_k \Lambda'^T_k).$$

The ergodic theorem assures its probabilistic convergence with respect to Kolmogorov’s topology over the infinite sequences of randomly growing circuits.

In what follows we prove the convergence of $1/\sqrt{n}(L^{(i)}_n - \mathbb{E}[L^{(i)}_n])$ by induction on $i$, assuming the convergence of $M_n/\sqrt{n}$. Set $\lambda_j = 1$, if $j=i$, and 0 otherwise. So, $W_n$ is reduced to $L^{(i)}_n - \mathbb{E}[L^{(i)}_n]$. The case $i=0$ has been proved in Mahmoud and Tsukiji [13]. For $i > 0$,

$$W_n = M_n - B_n$$

$$= M_n - \sum_{j=1}^{n} \frac{2}{j+a-1}(L^{(i)}_{j-1} - \mathbb{E}[L^{(i)}_{j-1}]) + \sum_{j=1}^{n} \frac{2}{j+a-1}(L^{(i-1)}_{j-1} - \mathbb{E}[L^{(i-1)}_{j-1}]).$$

By differencing we have

$$W_n - \frac{n+a-3}{n+a-1}W_{n-1} = \nabla M_n + \frac{2}{n+a-1}(L^{(i-1)}_{n-1} - \mathbb{E}[L^{(i-1)}_{n-1}]).$$

To get a widely recurrence, multiply both sides by $(n+a-1)(n+a-2)$ to rewrite the recurrence in the form

$$\nabla(j+a-1)(j+a-2)W_j = (j+a-1)(j+a-2)\nabla M_j + 2(j+a-2)(L^{(i-1)}_{j-1} - \mathbb{E}[L^{(i-1)}_{j-1}]).$$

In what follows, all the asymptotically negligible $o_p$ terms are in fact negligible in $L_1$, too; the details of the univariate case of outputs is in [13], the calculation carries over by induction to the multivariate case. Unwinding the last recurrence relation (by adding up from $j=1, \ldots, n$) we derive

$$(n+a-1)(n+a-2)W_n = \sum_{j=1}^{n} (j+a-1)(j+a-2) \left( \sqrt{j} M_j - \sqrt{j-1} \frac{M_{j-1}}{\sqrt{j-1}} \right)$$

$$+ 2\sum_{j=1}^{n} (j+a-2) \sqrt{j} \frac{(L^{(i-1)}_{j-1} - \mathbb{E}[L^{(i-1)}_{j-1}])}{\sqrt{j}}$$

$$= O(1) + \left[ o_p(1) + \lim_{j \to \infty} \frac{M_j}{\sqrt{j}} \right]$$
\[ \sum_{j=1}^{n} (j + a - 1)(j + a - 2)(\sqrt{j} - \sqrt{j-1}) \]

\[ + \left[ o_p(1) + 2 \lim_{j \to \infty} \frac{1}{\sqrt{j}} (L_{j-1}^{(i-1)} - E[L_{j-1}^{(i-1)}]) \right] \sum_{j=1}^{n} (j + a - 2)\sqrt{j}. \]

Divide throughout by \( n^{5/2} \) to obtain the probabilistic limit

\[ \frac{L_{n}^{(i)} - E[L_{n}^{(i)}]}{\sqrt{n}} \xrightarrow{p} \frac{1}{5} \lim_{n \to \infty} \frac{M_n}{\sqrt{n}} + \frac{4}{5} \lim_{n \to \infty} \frac{1}{\sqrt{n}} (L_{n}^{(i-1)} - E[L_{n}^{(i-1)}]). \]

A probabilistic limit for \( M_n/\sqrt{n} \) has been shown by the martingale central limit theorem, and a probabilistic limit for \( 1/\sqrt{n}(L_{n}^{(i-1)} - E[L_{n}^{(i-1)}]) \) is assumed by hypothesis, completing the induction.

We now translate the result back to the full-fledged \( W_n \), where we revert to the case that all \( \lambda_i \) are arbitrary. We have shown that \( (W_n + B_n)/\sqrt{n} \xrightarrow{p} \mathcal{N}(0, \Lambda_0' \Sigma_k \Lambda_k^T) \). The behavior of \( B_n \) is essentially similar to \( W_n \), they are both linear combinations of \( L_j^{(i)} \), for \( j = 1, \ldots, n \). For arbitrary \( \lambda_0, \ldots, \lambda_k \), the sum \( \sum_{i=0}^{k} \lambda_i \left( \frac{L_{n}^{(i)} - E[L_{n}^{(i)}]}{\sqrt{n}} \right) \) converges in probability to a limit, as has been demonstrated; the term \( B_n/\sqrt{n} \) behaves like

\[ -\frac{2}{\sqrt{n}} \sum_{j=1}^{n} \sum_{i=0}^{k} \left( \frac{L_{j-1}^{(i)} - E[L_{j-1}^{(i)}]}{j + a - 1} \right) \nabla \lambda_{i+1} \]

\[ = O \left( \frac{1}{\sqrt{n}} \right) - 2 \left[ o_p(1) + \lim_{j \to \infty} \sum_{i=0}^{k} \left( \frac{L_{j-1}^{(i)} - E[L_{j-1}^{(i)}]}{\sqrt{j}} \right) \nabla \lambda_{i+1} \right] \frac{1}{\sqrt{n}} \sum_{j=1}^{n} (j + a - 1) \]

\[ \xrightarrow{p} - 4 \lim_{j \to \infty} \sum_{i=0}^{k} \left( \frac{L_{j-1}^{(i)} - E[L_{j-1}^{(i)}]}{\sqrt{j}} \right) \nabla \lambda_{i+1}. \]

Combined, the two terms in \( (W_n + B_n)/\sqrt{n} \) behave like the probabilistic limit

\[ \lim_{n \to \infty} \sum_{i=0}^{k} \lambda_i' \left( \frac{L_{n}^{(i)} - E[L_{n}^{(i)}]}{\sqrt{n}} \right), \]

where \( \lambda_i' = \lambda_i - 4 \nabla \lambda_{i+1} = 5 \lambda_i - 4 \lambda_{i+1} \). Of course, \( \lambda_i' \), for \( i = 0, \ldots, k \), are still arbitrary constants.

Let \( W_n' = \sum_{i=0}^{k} \lambda_i' \left( L_{n}^{(i)} - E[L_{n}^{(i)}] \right) \). A main result follows:

\[ \frac{W_n'}{\sqrt{n}} \xrightarrow{p} \mathcal{N}(0, \Lambda_k' \Sigma_k \Lambda_k^T). \]

We have shown that the centered linear combination \( W_n' \), when normed by \( \sqrt{n} \), converges to a centered normal distribution, for any arbitrary choice of the coefficients \( \lambda_0', \ldots, \lambda_k' \). The Cramér–Wold mechanism converts such a result back into a multivariate central limit result.
Theorem 2. For arbitrary $k \geq 0$,
\[
\frac{L_n^{(k)} - M_k n}{\sqrt{n}} \Rightarrow \mathcal{N}_{k+1}(0_k, \Sigma_k),
\]
where
\[
M_k = \left( \frac{1}{3^k}, \frac{2}{3^k}, \ldots, \frac{2^k}{3^k} \right),
\]
and $\Sigma_k$ is a $(k + 1) \times (k + 1)$ matrix of constant coefficients given by Corollary 2.

5. Concluding remarks

We investigated the number of nodes of outdegree $k$ in a random graph evolving in a fashion that emulates a random binary circuit of gates. The outdegree of a node in the random graph, interpreted as the fan-out of a gate in the random circuit, has engineering implications such as the amount of electric current flowing in the circuit. We have shown that both the mean and variance of the number of nodes of fan-out $k$ grow linearly with the circuit size, and that the fan-outs up to $k$ together satisfy a multivariate central limit theorem. The investigation was done under the assumption that $k$ is fixed as the size of the circuit grows to infinity. If $k$ is not fixed, some of the results remain valid with minor modification in the asymptotic terms. For example, Theorem 1 remains valid under the less stringent assumption that $k = o(\ln n)$, if we replace $O(1/n)$ with the less informative $o(1)$.

References


