On the covariance of the level sizes in random recursive trees

Remco van der Hofstad* Gerard Hooghiemstra*
Piet Van Mieghem*

October 3, 2001

Abstract

In this paper we study the covariance structure of the number of nodes k and l steps away from the root in random recursive trees. We give an analytic expression valid for all k, l and tree sizes N. The fraction of nodes k steps away from the root is a random probability distribution in k. The expression for the covariances allows us to show that the total variation distance between this (random) probability distribution and its mean converges in probability to zero.

Key words and phrases: Recursive trees, covariance of level sizes, convergence in probability, convergence in total variation.

AMS 1991 *Subject classifications:* Primary 05C05, 60C05, Secondary 60F25, 60F99, 05C80.

1 Introduction

A random recursive tree of size N is a random tree that starts from node 1 (the root) and where at each stage a new node is attached at random to one of the existing nodes until the total number of nodes is equal to N. The depth $D_{1,N}$ is the number of links between the root 1 and a randomly chosen destination (chosen uniformly from all nodes $\{1, 2, \ldots, N\}$). The research of random recursive trees started in the seventies of last century by Moon [7], Meir and Moon [5], and was continued by others, see [8] and the references therein. In [8], the authors use the definition $d_{1,n}$ for the depth of the n^{th} node in the tree, which is related to our definition $D_{1,N}$ as $D_{1,N} = d_{1,N+1} - 1$. For our purpose, $D_{1,N}$ is more appropriate as will be clarified below. We define the distribution of $D_{1,N}$ by

$$\mathbb{P}(D_{1,N} = k) = p_N^{(k)}, \quad 0 \leq k \leq N - 1, \tag{1}$$

*Department of Information Technology and Systems, Delft University of Technology, Mekelweg 4, 2628 CD Delft, the Netherlands, e-mail: R.W.vanderHofstad@its.tudelft.nl, G.Hooghiemstra@its.tudelft.nl, p.vanmieghem@its.tudelft.nl

with corresponding generating function

$$\varphi_N(x) = \sum_{k=0}^{N-1} p_N^{(k)} x^k. \tag{2}$$

Szymański [9] was the first to identify the law of $D_{1,N}$ as

$$p_N^{(k)} = \frac{(-1)^{N-1-k} S_N^{(k+1)}}{N!}, \quad 0 \le k \le N-1, \tag{3}$$

where $S_N^{(k)}$ denote the (signed) Stirling numbers of the first kind (cf. [1, 24.1.3]). Therefore,

$$\varphi_N(x) = \frac{\Gamma(N+x)}{\Gamma(N+1)\Gamma(x+1)}. \tag{4}$$

For large N, the distribution of $D_{1,N}$ is close to the Poisson distribution with mean $\log N$ and hence also to the normal distribution with identical mean and variance given by $\log N$. In fact, Dobrow and Smythe [3] show that the total variation distance between the law of $D_{1,N}$ and the Poisson distribution with mean $\lambda_N = \mathbb{E}[D_{1,N}]$, is at most $C/\log N$.

We define the level k-set to be the set of all nodes k steps from the root and $X_N^{(k)}$ to be its size, with the understanding that $X_N^{(k)} = 0$, $k \notin \{0, 1, \ldots, N-1\}$. Then

$$p_N^{(k)} = \frac{\mathbb{E}\left[X_N^{(k)}\right]}{N}. \tag{5}$$

In this paper, we focus on the covariance between the two level sizes $X_N^{(k)}$ and $X_N^{(l)}$. More precisely, we study

$$s_N[k,l] = \mathbb{E}\left[X_N^{(k)} X_N^{(l)}\right], \quad 0 \le k, l \le N-1. \tag{6}$$

In the first part of Section 2, we will prove Lemma 2.1, which shows that $X_N^{(k)}$ is equal in law to $X_{N_1}^{(k-1)} + X_{N-N_1}^{(k)}$, where N_1 is uniform on $\{1, 2, \ldots, N-1\}$ and $X_{N_1}^{(k-1)}$ and $X_{N-N_1}^{(k)}$ are independent given N_1. In fact, the equality in law is even true simultaneously for all k. From this we obtain the following recurrence formula for $s_N[k,l]$. For $k, l \ge 0$,

$$\begin{aligned} s_N[k,l] &= \frac{1}{N-1} \sum_{m=1}^{N-1} \{\mathbb{E}\,[X_m^{(k-1)} X_m^{(l-1)}] + \mathbb{E}\left[X_{N-m}^{(k)} X_{N-m}^{(l)}\right]\} \\ &\quad + \frac{1}{N-1} \sum_{m=1}^{N-1} \{\mathbb{E}\,[X_m^{(k-1)}]\,\mathbb{E}\left[X_{N-m}^{(l)}\right] + \mathbb{E}\left[X_{N-m}^{(k)}\right]\mathbb{E}\,[X_m^{(l-1)}]\} \\ &= \frac{1}{N-1} \sum_{m=1}^{N-1} \{s_m[k-1, l-1] + s_{N-m}[k,l]\} + a_N[k,l], \end{aligned} \tag{7}$$

where

$$\begin{aligned} a_N[k,l] &= \frac{1}{N-1} \sum_{m=1}^{N-1} \left(\mathbb{E}\,[X_m^{(k-1)}]\,\mathbb{E}\left[X_{N-m}^{(l)}\right] + \mathbb{E}\,[X_m^{(k)}]\,\mathbb{E}\left[X_{N-m}^{(l-1)}\right]\right) \\ &= \frac{(-1)^{N-k-l-1}}{(N-1)!} \sum_{m=1}^{N-1} \binom{N-2}{m-1} \left(S_m^{(k)} S_{N-m}^{(l+1)} + S_m^{(k+1)} S_{N-m}^{(l)}\right). \end{aligned} \tag{8}$$

Equation (7) is the starting point of our analysis. In Section 2, we will solve the recurrence relation, and obtain a formula for $s_N[k,l]$ valid for all k,l,N. This formula reads, for $k \le l$,

$$\mathbb{E}\left[X_N^{(k)} X_N^{(l)}\right] = \frac{1}{(N-1)!} \sum_{j=0}^{k} (-1)^{N-j-l+1} \binom{2j+l-k}{j+l-k} S_N^{(j+l+1)}. \tag{9}$$

Putting $k = l$, we obtain the second moment

$$\mathbb{E}\left[\left(X_N^{(k)}\right)^2\right] = \frac{1}{(N-1)!} \sum_{j=0}^{k} (-1)^{N-j-k+1} \binom{2j}{j} S_N^{(j+k+1)}. \tag{10}$$

Using (3) and (5), we see that (9) is equivalent to

$$\mathbb{E}\left[X_N^{(k)} X_N^{(l)}\right] = \sum_{j=0}^{k} \binom{2j+l-k}{j+l-k} \mathbb{E}\left[X_N^{(j+l)}\right]. \tag{11}$$

It would be of interest to find a probabilistic interpretation of (11). We will prove (9) in Section 2 by computing and inverting the moment generating function $\sum_{k,l \ge 0} s_N[k,l] x^k y^l$.

Recursive trees are used to model all sorts of phenomena such as spread of epidemics, pyramid schemes, etc. (see e.g. [8] for an extensive list of examples). Our interest in recursive trees was triggered by the hopcount problem in the Internet: “What is the distribution of the number of hops (or traversed routers) along the shortest path between two arbitrary nodes in the Internet?” (cf. [6]). As initial model, we concentrated on the random graph of the class $G_p(N)$ (see e.g. [2]) consisting of all graphs with N nodes in which the edges (links) are chosen independently and with probability p. The links are further specified by independent exponentially distributed random weights with mean 1 and the shortest path between two arbitrary nodes minimizes the sum of the weights of the path between these two nodes. We have shown that the shortest path in a complete graph ($p = 1$) is equal to the depth $D_{1,N}$ from the source to an arbitrary node in a recursive tree of size N.[1] In [4], we have extended this result asymptotically for large N to $G_p(N)$ for $p = p_N < 1$: the law of the hopcount of the shortest path in $G_p(N)$ with exponentially distributed link weights is close (as $N \to \infty$) to the law of $D_{1,N}$. We have proved this fact in [4] under the condition that $Np_N/(\log N)^3 \to \infty$. Computer simulations confirm the limit law even when $Np_N \to \infty$ at a slower rate. In the simulations we have used $\hat{p}_N^{(k)} = X_N^{(k)}/N, \quad 0 \le k \le N-1$, as an unbiased estimator for the probability vector $p_N^{(k)}$ and we noticed that this estimator was close to $p_N^{(k)}$ (for all the realizations in the simulation). This suggests that $\hat{p}_N^{(k)}$ is close to $p_N^{(k)}$ *almost surely*. See also Figure 1 in Section 4.

In Section 3, we set forth to prove convergence of $\hat{p}_N^{(k)}$ to $p_N^{(k)}$. We show that the estimator $\hat{p}_N^{(k)} = X_N^{(k)}/N$, $0 \le k \le N-1$, is consistent in the sense that the total variation distance

$$d_{\mathrm{TV}}(\hat{p}_N, p_N) = \frac{1}{2} \sum_{k=0}^{N-1} |\hat{p}_N^{(k)} - p_N^{(k)}|, \tag{12}$$

[1] This is the reason why we study $D_{1,N}$ rather than $d_{1,N}$.

converges to 0 in probability. This result is similar to the above quoted result of Dobrow and Smythe, who showed that the total variation distance of p_N and a Poisson random variable with mean λ_N is small. Consequently, our result also implies that

$$\frac{1}{2}\sum_{k=0}^{\infty}\left|\hat{p}_N^{(k)} - e^{-\lambda_N}\frac{\lambda_N^k}{k!}\right|, \tag{13}$$

converges to 0 in probability. However, note that our total variation distance is *random*, since the estimator $\hat{p}_N = X_N^{(k)}/N$, $0 \le k \le N-1$, is actually a random measure. The proof of (12) uses the second moment of $X_N^{(k)}$ in (10).

In Section 4, we use (9) to compute the variance of the moment generating function of $\hat{p}_N$. Remarkably, this variance is *not* small compared to the square of its first moment, as one could expect when $\hat{p}_N$ is close to p_N. Moreover, there is an interesting change of behavior when the parameter substituted in the moment generating function exceeds 2, where the variance divided by the first moment squared tends to infinity. This indicates that even when $\hat{p}_N$ is close to p_N, the tails of $\hat{p}_N$ tend to be heavier than those of p_N.

Some technical lemmas concerning Stirling numbers are proved in Appendix I, and moderate deviation results concerning the probability distribution p_N in Appendix II.

2 The solution of the recursion relation for $s_N[k,l]$.

Before presenting Lemma 2.1, the key ingredient for the recursionrelation, some properties of recursive trees are reviewed.

When $N = 2$, the unique recursive tree of size 2 consists of a single edge between the root 1 and the second node, denoted here by 2. The larger recursive trees are now generated by attaching nodes to this tree recursively. Recursive trees are determined by the unlabeled tree, together with the labeling of the nodes of the unlabeled tree, indicating the order in which each of the nodes is attached. Therefore, recursive trees are a subset of all labeled trees. Not all labeled trees can arise as a recursive tree though, since the vertices that are further away from the root must have larger labels. This follows from the way a recursive tree is obtained by adding the numbered vertices recursively, and explains why there are only $(N-1)!$ recursive trees, whereas there are N^{N-1} labeled trees. Uniform recursive trees are now obtained by choosing one of the recursive trees uniformly.

We will use the following identity in law for the sequence $\{X_N^{(k)}\}_{k\ge 0}$. In the statement, we denote by $\stackrel{d}{=}$ equality in distribution.

Lemma 2.1 *Let $\{Y_N^{(k)}\}_{k,N\ge 0}$ and $\{Z_N^{(k)}\}_{k,N\ge 0}$ be two independent copies of the vector of level sets of two sequences of independent recursive trees. Then*

$$\{X_N^{(k)}\}_{k\ge 0} \stackrel{d}{=} \{Y_{N_1}^{(k-1)} + Z_{N-N_1}^{(k)}\}_{k\ge 0}, \tag{14}$$

where on the right-hand side the random variable N_1 is uniformly distributed over the set $\{1, 2, \dots, N-1\}$.

Proof: The recursive tree is divided into two subtrees, namely, the tree of nodes which are in graph distance closest to node 1, and the ones that are closer to node 2. We will

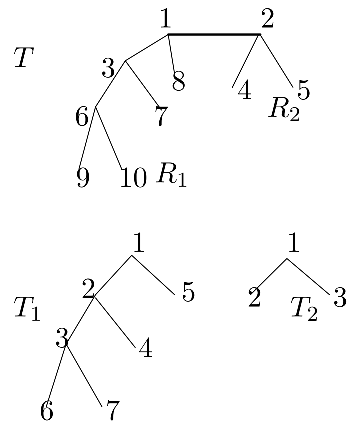

Figure 1: Decomposition of the tree T into the two subtrees R_1, R_2 and the necessary relabeling to obtain the recursive trees T_1, T_2 for $N = 10$.

call these trees the subtrees rooted at 1, respectively 2, and will denote these trees by R_1 and R_2. Clearly, the labels of R_1 and R_2 together are $1, \dots, N$. Denote the size of R_2 by N_1. Then the size of R_1 is $N - N_1$. We now relabel each of the vertices of the trees R_1 and R_2 to obtain new trees T_1 and T_2. This relabeling is done in such a way that T_1 obtains labels $1, \dots, N - N_1$ and T_2 obtains labels $1, \dots, N_1$. Moreover, the order of the labels the vertices of T_i is preserved compared to R_i. Consequently, as illustrated in Figure 1, T_1 and T_2 are recursive trees.

We will now prove that T_1 and T_2 are independent given N_1, and that N_1 is uniform on the set $\{1, 2, \dots, N-1\}$. That proves Lemma 2.1.

Let t_1 be a recursive tree of size of size m, and t_2 a recursive tree of size $N - m$. We note that since T is a *uniform* recursive tree, we have

$$\mathbb{P}(T_1 = t_1, T_2 = t_2, N_1 = m) = \frac{1}{(N-1)!} \#\{T : T_1 = t_1, T_2 = t_2\}, \tag{15}$$

i.e., the probability that $T_1 = t_1, T_2 = t_2, N_1 = m$ is just $\frac{1}{(N-1)!}$ times the number of recursive trees T for which the relabeling of the trees rooted at 1 and 2 are t_1 and t_2 respectively. Moreover,

$$\#\{T : T_1 = t_1, T_2 = t_2\} = \frac{(N-2)!}{(m-1)!(N-1-m)!}, \tag{16}$$

since this is just the number of orders in which we can attach $m-1$ vertices to the tree rooted at 1 and $N-1-m$ vertices to the tree rooted at 2. Indeed, since t_1 and t_2 contain the information of the labeling of each vertex of t_1 and t_2, we can reconstuct R_1 and R_2, and hence T, from t_1, t_2 and the order in which each of the vertices of t_1 and t_2 are attached to the tree rooted at 1, respectively, 2.

Therefore,

$$\mathbb{P}(T_1 = t_1, T_2 = t_2, N_1 = m) = \frac{1}{N-1} \frac{1}{(N-m-1)!} \frac{1}{(m-1)!}. \tag{17}$$

This proves the claim, as we can identify the right hand side as

$$\mathbb{P}(N_1 = m)\mathbb{P}(T_1 = t_1 | N_1 = m)\mathbb{P}(T_2 = t_2 | N_1 = m), \tag{18}$$

so that indeed N_1 is uniform over the set $\{1, 2, \ldots, N-1\}$, and the trees T_1 and T_2 are conditionally independent given N_1. This in particular implies that the level sets of the trees T_1 and T_2 are independent.

$\square$

We now come to our main result identifying the cross moments of the level sizes.

Theorem 2.2 *For $k \leq l$, the solution of (7), with initial conditions $s_N[k,0] = (-1)^{N-k-1}\frac{S_N^{(k+1)}}{(N-1)!}$, is*

$$\mathbb{E}[X_N^{(k)}X_N^{(l)}] = s_N[k,l] = \frac{1}{(N-1)!}\sum_{j=0}^{k}(-1)^{N-j-l+1}\binom{2j+l-k}{j+l-k}S_N^{(j+l+1)}. \tag{19}$$

Proof: Define the generating functions

$$\chi_N(x,y) = \sum_{k,l=0}^{N-1} s_N[k,l]x^k y^l, \quad A_N(x,y) = \sum_{k,l=0}^{N-1} a_N[k,l]x^k y^l.$$

Multiply both sides of the recursion (7) by $x^k y^l$ and sum over $k, l \in \{0, 1, \ldots, N-1\}$, to get

$$(N-1)\chi_N(x,y) = (1+xy)\sum_{m=1}^{N-1}\chi_m(x,y) + (N-1)A_N(x,y).$$

Subtraction of $N\chi_{N+1}(x,y) - (N-1)\chi_N(x,y)$, yields

$$N\chi_{N+1}(x,y) - (N+xy)\chi_N(x,y) = NA_{N+1}(x,y) - (N-1)A_N(x,y). \tag{20}$$

We denote $F_N(x,y) = NA_{N+1}(x,y) - (N-1)A_N(x,y)$ and rewrite (20) as

$$\chi_N(x,y) = \frac{N-1+xy}{N-1}\chi_{N-1}(x,y) + \frac{1}{N-1}F_{N-1}(x,y). \tag{21}$$

We use $\chi_1(x,y) = 1$, to obtain, by iteration,

$$\chi_N(x,y) = \frac{\Gamma(xy+N)}{(N-1)!\Gamma(xy+1)} + \frac{\Gamma(xy+N)}{(N-1)!}\sum_{q=1}^{N-1}\frac{F_q(x,y)(q-1)!}{\Gamma(xy+q+1)}. \tag{22}$$

Substituting (50) into (22), we end up with

$$\begin{aligned}\chi_N(x,y) &= \frac{\Gamma(xy+N)}{(N-1)!\Gamma(xy+1)} + \frac{\Gamma(xy+N)}{(N-1)!}\sum_{q=1}^{N-1}\frac{\Gamma(x+y+q)}{\Gamma(x+y)\Gamma(xy+q+1)} \\ &= \frac{\Gamma(xy+N)}{(N-1)!\Gamma(x+y)}\sum_{q=0}^{N-1}\frac{\Gamma(x+y+q)}{\Gamma(xy+q+1)}. \end{aligned} \tag{23}$$

Now, using Lemma 5.4 with $a = x+y$ and $b = xy+1$,

$$\begin{aligned}\chi_N(x,y) &= \frac{\Gamma(xy+N)}{(N-1)!\Gamma(x+y)}\frac{1}{(x+y-xy)}\left(\frac{\Gamma(x+y+N)}{\Gamma(xy+N)} - \frac{\Gamma(x+y)}{\Gamma(xy)}\right) \\ &= \frac{\Gamma(x+y+N)}{(N-1)!(x+y-xy)\Gamma(x+y)} - \frac{\Gamma(xy+N)}{(N-1)!(x+y-xy)\Gamma(xy)}, \end{aligned} \tag{24}$$

and finally from $\frac{\Gamma(z+n)}{\Gamma(z)} = \sum_{j=0}^{n}(-1)^{n-j} S_n^{(j)} z^j$ (cf. [1, 1.24.3]),

$$\begin{aligned}\chi_N(x,y) &= \sum_{j=0}^{N} \frac{(-1)^{N-j} S_N^{(j)}}{(N-1)!} \frac{(x+y)^j}{(x+y-xy)} - \sum_{j=0}^{N} \frac{(-1)^{N-j} S_N^{(j)}}{(N-1)!} \frac{(xy)^j}{(x+y-xy)} \\ &= \sum_{j=0}^{N} \frac{(-1)^{N-j} S_N^{(j)}}{(N-1)!} \left(\frac{(x+y)^j - (xy)^j}{x+y-xy} \right). \end{aligned} \tag{25}$$

To obtain the coefficients of χ_N, write

$$\begin{aligned}\frac{(x+y)^j - (xy)^j}{x+y-xy} &= (xy)^{j-1} \frac{\left(\frac{x+y}{xy}\right)^j - 1}{\frac{x+y}{xy} - 1} = (xy)^{j-1} \sum_{i=0}^{j-1} \left(\frac{x+y}{xy}\right)^i \\ &= \sum_{i=0}^{j-1} \sum_{m=0}^{i} \binom{i}{m} x^{j-m-1} y^{m-i+j-1},\end{aligned}$$

so that

$$\sum_{0\le k,l \le N-1} s_N[k,l] x^k y^l = \frac{1}{(N-1)!} \sum_{j=0}^{N} (-1)^{N-j} S_N^{(j)} \sum_{i=0}^{j-1} \sum_{m=0}^{i} \binom{i}{m} x^{j-m-1} y^{m-i+j-1}.$$

Taking $k = j-m-1$ and $l = m-i+j-1$, or $m = j-k-1$ and $i = 2j-k-l-2$, we obtain that $0 \le m \le i$ is equivalent to $j \ge \max(k+1, l+1)$ and $0 \le i \le j-1$ is equivalent to $(k+l)/2 + 1 \le j \le k+l+1$, implying for $k \le l$,

$$\begin{aligned}\mathbb{E}\left[X_N^{(k)} X_N^{(l)}\right] = s_N[k,l] &= \frac{1}{(N-1)!} \sum_{j=\max(k,l)+1}^{k+l+1} S_N^{(j)} (-1)^{N-j} \binom{2j-k-l-2}{j-k-1} \\ &= \frac{1}{(N-1)!} \sum_{j=0}^{k} (-1)^{N-j-l-1} S_N^{(j+l+1)} \binom{2j-k+l}{j-k+l}.\end{aligned}$$

□

3 Convergence in probability

In this section, we will show the following theorem.

Theorem 3.1 *As $N \to \infty$,*

$$\mathbb{E}\left[\sum_{k=0}^{N-1} (\hat{p}_N^{(k)} - p_N^{(k)})^2\right] = \frac{2-\pi^2/6}{4\sqrt{\pi(\log N)^3}} + \mathcal{O}((\log N)^{-5/2}). \tag{26}$$

Consequently, for all $\varepsilon > 0$,

$$\lim_{N\to\infty} \mathbb{P}(d_{\mathrm{TV}}(\hat{p}_N, p_N) > \varepsilon) = 0. \tag{27}$$

In words, (27) shows that the total variation distance between the random probability vector $(\hat{p}_N^{(k)})_{k=0}^{N-1}$ and its expectation $(p_N^{(k)})_{k=0}^{N-1}$ converges to zero in probability. Equation (26) will be used to prove (27). However, it is interesting on its own, as it shows that $\hat{p}_N$ converges to a normal density in L^2-sense. Indeed, define

$$\hat{q}_N^{(y)} = \sqrt{\log N}\hat{p}_N^{(\lceil \log N + y\sqrt{\log N}\rceil)},$$

and let $q_N^{(y)} = \mathbb{E}\left[\hat{q}_N^{(y)}\right]$. Then (26) implies that with $\|\cdot\|_2$ denoting the L^2-norm, we have

$$\mathbb{E}\left[\|\hat{q}_N - q_N\|_2^2\right] = \frac{2-\pi^2/6}{4\sqrt{\pi}\log N} + \mathcal{O}((\log N)^{-2}). \tag{28}$$

To see (28), we write

$$\mathbb{E}\left[\|\hat{q}_N - q_N\|_2^2\right] = \mathbb{E}\left[\int_{-\infty}^{\infty}(\hat{q}_N^{(y)} - q_N^{(y)})^2 dy\right] = (\log N)^{1/2}\mathbb{E}\left[\sum_{k=0}^{N-1}(\hat{p}_N^{(k)} - p_N^{(k)})^2\right]. \tag{29}$$

Furthermore, by the triangle inequality, we have that with

$$q(y) = \frac{e^{-y^2/2}}{\sqrt{2\pi}}, \qquad y \in \mathbb{R}, \tag{30}$$

denoting the standard Gaussian density,

$$\mathbb{E}\left[\|\hat{q}_N - q\|_2^2\right]^{1/2} \le \mathbb{E}\left[\|\hat{q}_N - q_N\|_2^2\right]^{1/2} + \|q_N - q\|_2. \tag{31}$$

In [3] it is shown that the total variation distance between p_N and the Poisson with mean λ_N tends to zero. This result implies that $\|q_N - q\|_2^2 \to 0$. Therefore, $\hat{q}_N$ converges in L^2 to q.

Proof: We first show that (27) follows from (26) together with Lemma 6.1 proved in Appendix II.

Invoking the Markov inequality we get

$$\mathbb{P}(d_{\mathrm{TV}}(\hat{p}_N, p_N) > \varepsilon) \le \frac{1}{2\varepsilon}\mathbb{E}\left[\sum_{k=0}^{N-1}|\hat{p}_N^{(k)} - p_N^{(k)}|\right].$$

Hence for the convergence in probability it suffices to show that

$$\lim_{N\to\infty}\mathbb{E}\left[\sum_{k=0}^{N-1}|\hat{p}_N^{(k)} - p_N^{(k)}|\right] = 0. \tag{32}$$

We will split the sum over k into two parts, depending on whether $|k-\log N| \le c_N\sqrt{\log N}$, for some sequence $c_N \to \infty$, or not. Denote

$$K_N = \{k : |k - \log N| \le c_N\sqrt{\log N}\}.$$

The first part, where $|k - \log N| > c_N\sqrt{\log N}$, we handle as follows. We first bound

$$\mathbb{E}\Big[\sum_{k\in K_N^c}|\hat{p}_N^{(k)} - p_N^{(k)}|\Big] \le \mathbb{E}\Big[\sum_{k\in K_N^c}\hat{p}_N^{(k)} + p_N^{(k)}\Big] = 2\sum_{k\in K_N^c}p_N^{(k)}.$$

In Lemma 6.1 in Appendix II, it is proved that for $c_N = o(\sqrt{\log N})$,

$$\sum_{k \in K_N^c} p_N^{(k)} \leq 8 \exp\left\{-\frac{c_N^2}{2}\left(1 - \frac{c_N}{\sqrt{\log N}}\right)\right\}.$$

This deals with the first part ($k \in K_N^c$). For the second part, we use Cauchy-Schwarz, together with $|K_N| \leq 2c_N\sqrt{\log N} + 1$, to bound

$$\mathbb{E}\left[\sum_{k \in K_N} |\hat{p}_N^{(k)} - p_N^{(k)}|\right] \leq \left[\mathbb{E}\sum_{k \in K_N} (\hat{p}_N^{(k)} - p_N^{(k)})^2\right]^{1/2} (2c_N\sqrt{\log N} + 1)^{1/2}.$$

Therefore, (27) follows once we prove (26). We next prove (26).

Since $\mathbb{E}\left[\hat{p}_N^{(k)}\right] = p_N^{(k)}$, we have

$$\mathbb{E}\left[\sum_{k=0}^{N-1} (\hat{p}_N^{(k)} - p_N^{(k)})^2\right] = \sum_{k=0}^{N-1} \mathbb{E}\left[\left(\hat{p}_N^{(k)}\right)^2\right] - \sum_{k=0}^{N-1} \left(p_N^{(k)}\right)^2. \tag{33}$$

We will start with the second sum on the right hand side of (33).

According to the Parseval relation (cf. [11]) and the generating function φ_N in (2) we find

$$\sum_{k=0}^{N-1} \left(p_N^{(k)}\right)^2 = \frac{1}{2\pi\Gamma^2(N+1)} \int_{-\pi}^{\pi} \left|\frac{\Gamma(N + e^{i\theta})}{\Gamma(1 + e^{i\theta})}\right|^2 d\theta. \tag{34}$$

We first use that

$$\left|\frac{\Gamma(N + e^{i\theta})}{\Gamma(N + \cos\theta)}\right|^2 = 1 + \mathcal{O}(N^{-1}) \tag{35}$$

(cf. [1, 6.1.47]). As a first approximation we again use [1, 6.1.47] to obtain, for large N and all $\theta \in (-\pi, \pi)$,

$$\frac{\Gamma(N + \cos\theta)}{\Gamma(N+1)} = N^{\cos\theta - 1}(1 + \mathcal{O}(N^{-1})). \tag{36}$$

Denote $\varepsilon_N = \varepsilon(\log N)^{-1/4}$. An upper bound for the integral outside the interval $(-\varepsilon_N, \varepsilon_N)$, where $\varepsilon > 0$ is small enough, is obtained by combining (35) with

$$\frac{1}{2\pi} \int_{(-\pi,\pi)\setminus(-\varepsilon_N,\varepsilon_N)} \frac{\Gamma^2(N + 1 + (\cos\theta - 1))}{\Gamma^2(N+1)} \, d\theta \leq N^{2(\cos\varepsilon_N - 1)}(1 + \mathcal{O}(N^{-1})) \leq e^{-\frac{\varepsilon^2}{2}\sqrt{\log N}}. \tag{37}$$

When $\theta \in [-\varepsilon_N, \varepsilon_N]$, we obtain from (36) and the expansion of $\theta \mapsto \cos\theta$,

$$\frac{\Gamma^2(N + \cos\theta)}{\Gamma^2(N+1)} = e^{2(\cos\theta - 1)\log N}(1 + \mathcal{O}(N^{-1})) = e^{-\theta^2 \log N}\left(1 + \frac{\theta^4 \log N}{12} + \mathcal{O}(\theta^6 \log N)\right).$$

Now we calculate the integral on the right-hand side of (34), where we again use that $\left|\frac{\Gamma(N+e^{i\theta})}{\Gamma(N+\cos\theta)}\right|^2 = 1+\mathcal{O}(N^{-1})$ and $|\Gamma(1+e^{i\theta})| = |\Gamma(e^{i\theta})|$,

$$\begin{aligned}
\frac{1}{2\pi}\int_{-\pi}^{\pi}\left|\frac{\Gamma(N+e^{i\theta})}{\Gamma(N+1)\Gamma(1+e^{i\theta})}\right|^2 d\theta &= \frac{1}{2\pi}\int_{-\pi}^{\pi}\left|\frac{\Gamma(N+\cos\theta)}{\Gamma(N+1)}\right|^2\cdot\left|\frac{\Gamma(N+e^{i\theta})}{\Gamma(N+\cos\theta)\Gamma(1+e^{i\theta})}\right|^2 d\theta \\
&= \frac{1}{2\pi}\int_{-\varepsilon_N}^{\varepsilon_N}\left|\frac{\Gamma(N+e^{i\theta})}{\Gamma(N+\cos\theta)}\right|^2\cdot\left|\frac{\Gamma(N+\cos\theta)}{\Gamma(N+1)\Gamma(1+e^{i\theta})}\right|^2 d\theta + \mathcal{O}(e^{-\frac{\varepsilon^2}{2}\sqrt{\log N}}) \\
&= \frac{1+\mathcal{O}(N^{-1})}{2\pi}\int_{-\varepsilon_N}^{\varepsilon_N}\left|\frac{\Gamma(N+\cos\theta)}{\Gamma(N+1)\Gamma(1+e^{i\theta})}\right|^2 d\theta + \mathcal{O}(e^{-\frac{\varepsilon^2}{2}\sqrt{\log N}}) \\
&= \frac{1}{2\pi}\int_{-\varepsilon_N}^{\varepsilon_N}\frac{e^{-\theta^2\log N}}{|\Gamma(e^{i\theta})|^2}\left(1+\frac{\theta^4\log N}{12}+\mathcal{O}(\theta^6\log N)\right) d\theta + \mathcal{O}(e^{-\frac{\varepsilon^2}{2}\sqrt{\log N}}). \qquad (38)
\end{aligned}$$

In the above, we have absorbed error terms bounded by $\mathcal{O}(N^{-1})$ into the error term $\mathcal{O}(e^{-\frac{\varepsilon^2}{2}\sqrt{\log N}})$. On the right hand side we will can go to the full integral over $(-\infty,\infty)$, since

$$\int_{(-\infty,\infty)\setminus(-\varepsilon_N,\varepsilon_N)}\frac{e^{-\theta^2\log N}}{|\Gamma(e^{i\theta})|^2}\left(1+\frac{\theta^4\log N}{12}+\mathcal{O}(\theta^6\log N)\right) d\theta = \mathcal{O}(e^{-\frac{\varepsilon^2}{2}\sqrt{\log N}}).$$

Therefore, we end up with

$$\begin{aligned}
\frac{1}{2\pi}\int_{-\pi}^{\pi}\left|\frac{\Gamma(N+e^{i\theta})}{\Gamma(N+1)\Gamma(1+e^{i\theta})}\right|^2 d\theta &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{e^{-z^2}dz}{\sqrt{\log N}} + \frac{a}{2\pi}\int_{-\infty}^{\infty}\frac{z^2e^{-z^2}dz}{(\log N)^{3/2}} \qquad (39)\\
&\quad + \frac{1/12}{2\pi}\int_{-\infty}^{\infty}\frac{z^4e^{-z^2}dz}{(\log N)^{3/2}} + \mathcal{O}\left(\frac{1}{(\log N)^{5/2}}\right),
\end{aligned}$$

where a equals

$$a = \frac{1}{2}\frac{d^2}{d\theta^2}\frac{1}{|\Gamma(e^{i\theta})|^2}\bigg|_{\theta=0} = \frac{\pi^2}{6}-\gamma,$$

and γ is Euler's constant. Hence,

$$\sum_{k=0}^{N-1}\left(p_N^{(k)}\right)^2 = \frac{1}{2\sqrt{\pi\log N}} + \frac{(\frac{\pi^2}{6}-\gamma+\frac{1}{8})}{4\sqrt{\pi(\log N)^3}} + \mathcal{O}\left(\frac{1}{(\log N)^{5/2}}\right). \qquad (40)$$

We next derive the asymptotics of the first sum. We first use (10) to write it as

$$\sum_{k=0}^{N-1}\mathbb{E}\left[\left(\hat{p}_N^{(k)}\right)^2\right] = \frac{1}{N}\sum_{k=0}^{N-1}\sum_{j=0}^{k}\binom{2j}{j}p_N^{(k+j)}. \qquad (41)$$

We then rewrite

$$\begin{aligned}
\sum_{k=0}^{N-1}\sum_{j=0}^{k}\binom{2j}{j}p_N^{(k+j)} &= \sum_{k=0}^{\infty}\sum_{j=0}^{k}\binom{2j}{j}p_N^{(k+j)} = \sum_{l=0}^{\infty}\sum_{2j\le l}\binom{2j}{j}p_N^{(l)} \\
&= \sum_{k=0}^{\infty}[p_N^{(2k)}+p_N^{(2k+1)}]\sum_{j=0}^{k}\binom{2j}{j},
\end{aligned}$$

where we use the change of variables $l = k + j$.

We next use Wallis' formula (cf. [1, 6.1.49]),

$$\binom{2j}{j} = \frac{2^{2j+1}}{\pi} \int_0^{\frac{\pi}{2}} \cos^{2j}(\theta)\, d\theta.$$

We split the integral into two parts, corresponding to $0 \le \theta \le \frac{\pi}{3}$ and $\frac{\pi}{3} \le \theta \le \frac{\pi}{2}$. We start with the latter, which is an error term. We use that $\cos^2\theta \le \frac{1}{2}$, so that this contribution is bounded as

$$\frac{1}{N}\sum_{k=0}^{\infty}[p_N^{(2k)} + p_N^{(2k+1)}]\sum_{j=0}^{k}\frac{2^{2j+1}}{\pi}\int_{\frac{\pi}{3}}^{\frac{\pi}{2}} 2^{-j}\, d\theta = \frac{1}{6N}\sum_{k=0}^{\infty}(2^{k+1}-1)[p_N^{(2k)} + p_N^{(2k+1)}]. \tag{42}$$

We further bound this term, using (4), by

$$\frac{1}{N}\sum_{l=0}^{\infty} 2^{\frac{l}{2}} p_N^{(l)} \le \frac{1}{N}\varphi_N(\sqrt{2}) = \mathcal{O}(N^{\sqrt{2}-2}).$$

We proceed with the main term, due to $0 \le \theta \le \frac{\pi}{3}$. We will use that

$$\sum_{k=0}^{\infty} z^k p_N^{(2k)} = \frac{1}{2}[\varphi_N(\sqrt{z}) + \varphi_N(-\sqrt{z})], \qquad \sum_{k=0}^{\infty} z^k p_N^{(2k+1)} = \frac{1}{2\sqrt{z}}[\varphi_N(\sqrt{z}) - \varphi_N(-\sqrt{z})]. \tag{43}$$

This yields:

$$\begin{aligned}
\frac{1}{N}\sum_{k=0}^{N-1}\sum_{j=0}^{k}\binom{2j}{j}p_N^{(k+j)} &= \frac{2}{\pi N}\int_0^{\frac{\pi}{3}}\sum_{k=0}^{N-1}[p_N^{(2k)} + p_N^{(2k+1)}]\frac{(4\cos^2\theta)^{k+1}-1}{4\cos^2\theta - 1}\, d\theta + \mathcal{O}(N^{\sqrt{2}-2}) \qquad (44)\\
&= \frac{1}{\pi N}\int_0^{\frac{\pi}{3}}\frac{4\cos^2\theta[\varphi_N(2\cos\theta) + \varphi_N(-2\cos\theta)]}{4\cos^2\theta - 1}\, d\theta\\
&\quad + \frac{1}{\pi N}\int_0^{\frac{\pi}{3}}\frac{2\cos\theta[\varphi_N(2\cos\theta) - \varphi_N(-2\cos\theta)] - 2}{4\cos^2\theta - 1}\, d\theta + \mathcal{O}(N^{\sqrt{2}-2})\\
&= \frac{1}{\pi N}\int_0^{\frac{\pi}{3}}\frac{(2\cos\theta + 4\cos^2\theta)\varphi_N(2\cos\theta)}{4\cos^2\theta - 1}\, d\theta\\
&\quad - \frac{1}{\pi N}\int_0^{\frac{\pi}{3}}\frac{(2\cos\theta - 4\cos^2\theta)\varphi_N(-2\cos\theta) + 2}{4\cos^2\theta - 1}\, d\theta + \mathcal{O}(N^{\sqrt{2}-2}).
\end{aligned}$$

Since $\frac{(2\cos\theta - 4\cos^2\theta)\varphi_N(-2\cos\theta)+2}{4\cos^2\theta - 1}$ is bounded on $[0, \frac{\pi}{3}]$, the second integral on the right hand side of (44) is $\mathcal{O}(N^{-1})$.

As before,

$$\varphi_N(2\cos\theta) = \frac{\Gamma(N + 2\cos\theta)}{\Gamma(N+1)}\frac{1}{\Gamma(1 + 2\cos\theta)}.$$

From [1, 16.1.47], and using Maple for large N and $\theta \to 0$,

$$\begin{aligned}
&\frac{\Gamma(N + 2\cos\theta)}{N\,\Gamma(N+1)} = N^{2\cos\theta - 2}(1 + \mathcal{O}(N^{-1})) = N^{-\theta^2 + \frac{\theta^4}{12}}(1 + \mathcal{O}(N^{-1})),\\
&(\Gamma(1 + 2\cos\theta))^{-1} = \frac{1}{2} + \left(\frac{3}{4} - \frac{\gamma}{2}\right)\theta^2 + \mathcal{O}(\theta^4),\\
&\frac{2\cos\theta + 4\cos^2\theta}{4\cos^2\theta - 1} = 2 + \theta^2 + \mathcal{O}(\theta^4).
\end{aligned}$$

Hence,

$$\begin{aligned}
&\frac{1}{\pi N}\int_0^{\frac{\pi}{3}} \frac{(2\cos\theta + 4\cos^2\theta)\varphi_N(2\cos\theta)}{4\cos^2\theta - 1}\, d\theta \\
&\quad = \frac{1}{\pi}\int_0^{\frac{\pi}{3}} N^{-\theta^2+\frac{\theta^4}{12}}\left(\frac{1}{2}+\left(\frac{3}{4}-\frac{\gamma}{2}\right)\theta^2\right)\left(2+\theta^2\right)\, d\theta + \mathcal{O}((\log N)^{-5/2}) \\
&\quad = \frac{1}{\pi}\int_0^{\frac{\pi}{3}} e^{-\theta^2\log N}\left(1+(2-\gamma)\,\theta^2+\frac{\theta^4\log N}{12}\right)\, d\theta + \mathcal{O}((\log N)^{-5/2}) \\
&\quad = \frac{1}{2\sqrt{\pi\log N}} + \frac{2-\gamma+\frac{1}{8}}{4\sqrt{\pi(\log N)^3}} + \mathcal{O}((\log N)^{-5/2}).
\end{aligned}$$

Together with (40) this shows (26). □

4 The variance of the moment generating function

Define the *random* generating function of $\hat{p}_N$ to be

$$\hat{\varphi}_N(r) = \sum_{k=0}^{N-1} r^k \hat{p}_N^{(k)}.$$

Clearly, $\hat{\varphi}_N(0) = \hat{p}_N^{(0)} = 1/N$, and $\hat{\varphi}_N(1) = 1$.

We compute the variance of the above random generating function, using (24),

$$\begin{aligned}
\mathrm{Var}(\hat{\varphi}_N(r)) &= \frac{1}{N^2}\mathrm{cov}\left(\sum_{k=0}^{N-1} r^k X_N^{(k)}, \sum_{l=0}^{N-1} r^l X_N^{(l)}\right) \\
&= \frac{1}{N^2}\sum_{k=0}^{N-1}\sum_{l=0}^{N-1} r^{k+l}\mathrm{cov}\left(X_N^{(k)}, X_N^{(l)}\right) \\
&= \frac{\chi_N(r,r)}{N^2} - \left(\frac{\Gamma(N+r)}{\Gamma(N+1)\Gamma(r+1)}\right)^2 \\
&= \frac{\left(\frac{\Gamma(2r+N)}{\Gamma(2r)} - \frac{\Gamma(r^2+N)}{\Gamma(r^2)}\right)}{N\Gamma(N+1)(2r-r^2)} - \left(\frac{\Gamma(N+r)}{\Gamma(N+1)\Gamma(r+1)}\right)^2.
\end{aligned} \tag{45}$$

We know that

$$\frac{\Gamma(N+r)}{\Gamma(N+1)} = N^{r-1}\left(1+\mathcal{O}\left(\frac{1}{N}\right)\right).$$

Therefore, we find that for $r < 2$,

$$\frac{\mathrm{Var}(\hat{\varphi}_N(r))}{\varphi_N(r)^2} = \left(\frac{\Gamma^2(r+1)}{(2r-r^2)\Gamma(2r)} - 1\right)\left(1+\mathcal{O}\left(\frac{1}{N^{r^2-2r}}\right)\right). \tag{46}$$

The function on the right hand side is strictly positive for $0 < r < 2$, except at $r = 1$. This means that the variance of $\hat{\varphi}_N(r)$ is of the same order as its expectation squared in this regime of r. Moreover, as $r \uparrow 2$, we have that the ratio in (46) tends to infinity as $N \to \infty$.

Therefore, $\hat{\varphi}_N(r)$ does not converge to $\varphi_N(r)$ in L^2, even though it is the generating function of a random distribution that *does* converge to its mean (in probability) in total variation and in L^2. This indicates that $\hat{p}_N^{(k)}$ has generally fatter tails than $p_N^{(k)}$. We can also see this in Figure 1, where 25 realizations of $\hat{p}_N$ are drawn, together with p_N, for $N = 100,000$.

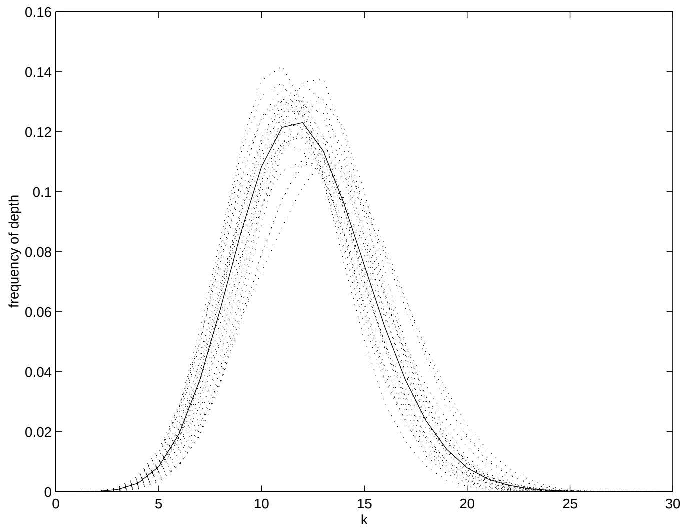

Figure 2: The probability distribution p_N (solid line) and 25 realizations of $\hat{p}_N$ for $N = 100,000$ (dashed lines).

Acknowledgement: We thank both referees for their comments which improved the manuscript at various places.

References

[1] M. Abramowitz and I.A. Stegun, *Handbook of Mathematical Functions*, Dover (1968).

[2] B. Bollobas, *Random Graphs*, Academic Press, New York (1985).

[3] R.P. Dobrow and R.T. Smythe, *Poisson approximations for functionals of random trees*, Random Struct. and Alg. **9** (1996), 79-92.

[4] R. van der Hofstad, G. Hooghiemstra and P. Van Mieghem, *First passage percolation on the random graph,* Prob Engin. and Inform. Sciences, **15** (2001), 225-237.

[5] A. Meir and J.W. Moon, *On the altitude of nodes in random trees*, Canadian J. Math. **30** (1978), 997-1015.

[6] P. Van Mieghem, G. Hooghiemstra and R. van der Hofstad, *A scaling law for the hopcount*, Delft report 2000125, http://wwwtvs.et.tudelft.nl/people/piet/telconference.html

[7] J.W. Moon, *The distance between nodes in recursive trees*, London Math. Soc. Lecture Notes Ser., **13** (1974), Cambridge University Press, London, 125-132.

[8] R.T. Smythe and H.M, Mahmoud, *A survey of recursive trees,* Theor. Probability and Math. Statist. **51** (1995), 1-27.

[9] J. Szymański, *On the maximum degree and the height of a random recursive tree,* Random Graphs '87 (Poznań, 1987, M. Karoński and A. Ruciński, eds.) Wiley New York, (1990), 313-324.

[10] J. Riordan, *Combinatorical Identities,* John Wiley, New York (1968).

[11] E.C. Titchmarsh, *Introduction to the Theory of Fourier Integrals,* Oxford University Press, Ely House, London (1949).

5 Appendix I.

In this appendix we prove the formula for $F_q(x,y)$ used in Section 2 and the summation formula for the Gamma function used to prove (24). We start with two preparatory lemmas containing identities for Stirling numbers.

Lemma 5.1 *For $N > 1$ and $0 \leq k, l \leq N-1$,*

$$a_N[k,l] = \frac{(-1)^{N-1}}{(N-1)!}\binom{k+l}{k}\sum_{q=k+l}^{N-1}(-1)^q\binom{q-1}{k+l-1}S_{N-1}^{(q)}. \tag{47}$$

Proof: From the identity (cf. [10, Chapter 1]),

$$\binom{a+b-2}{k-2} = \sum_{q=1}^{k-2}\binom{a-1}{q-1}\binom{b-1}{k-q-1},$$

and the generating function of the Stirling numbers (cf. [1, 24.1.3]), we obtain

$$\binom{a+b-2}{N-2} = \frac{1}{(N-2)!}\sum_{m=1}^{N-2}\sum_{j=0}^{m}\sum_{s=0}^{N-m}\binom{N-2}{m-1}S_m^{(j)}S_{N-m}^{(s)}a^{j-1}b^{s-1}.$$

Reversing summation orders, gives

$$ab(N-2)!\binom{a+b-2}{N-2} = \sum_{j=0}^{N-2}a^j\sum_{s=0}^{N-j}b^s\sum_{m=j}^{N-l}\binom{N-2}{m-1}S_m^{(j)}S_{N-m}^{(s)}.$$

From this identity one finds

$$\frac{(N-2)!}{k!l!}\frac{\partial^{k+l}}{\partial^l b\partial^k a}\left[ab\binom{a+b-2}{N-2}\right]\Bigg|_{a=b=0} = \sum_{m=k}^{N-l}\binom{N-2}{m-1}S_m^{(k)}S_{N-m}^{(l)}.$$

Using the generating function [1, 24.1.3] again, we find for $N > 1$,

$$\begin{aligned} ab\binom{a+b-2}{N-2} &= \sum_{q=1}^{N-1} \frac{S_{N-1}^{(q)}}{(N-2)!} ab(a+b-1)^{q-1} \\ &= \sum_{q=1}^{N-1} \frac{S_{N-1}^{(q)}}{(N-2)!} \sum_{j=0}^{q-1} \binom{q-1}{j} a^{j+1} b(b-1)^{q-1-j}, \end{aligned}$$

so that on the other hand

$$\frac{\partial^{k+l}}{\partial^l b \partial^k a}\left[ab\binom{a+b-2}{N-2}\right]\Bigg|_{a=b=0} = \frac{(-1)^{k+l-1}k!l!}{(N-2)!(k-1)!(l-1)!} \sum_{q=k+l-1}^{N-1} (-1)^q \frac{(q-1)!}{(q-k-l+1)!} S_{N-1}^{(q)}.$$

Equating both expressions for the $(k+l)^{\text{th}}$ partial derivative yields

$$\sum_{m=k}^{N-l} \binom{N-2}{m-1} S_m^{(k)} S_{N-m}^{(l)} = (-1)^{k+l-1}\binom{k+l-2}{k-1} \sum_{q=k+l-1}^{N-1} (-1)^q \binom{q-1}{k+l-2} S_{N-1}^{(q)}, \tag{48}$$

and hence (47).

The above expression shows in particular that $a_N[k,l] = 0$, when $k+l \geq N$. $\square$

Lemma 5.2 *For $N \geq k+l+1$,*

$$\sum_{j=k+l}^{N-2} (-1)^j \binom{j-1}{k+l-1} \frac{S_{N-2}^{(j-1)}}{(N-2)!} = \frac{(-1)^{k+l} S_{N-1}^{(k+l)}}{(N-2)!} + \frac{(-1)^N}{(N-k-l-1)!(k+l-1)!}. \tag{49}$$

Proof: Multiply both sides by x^N and sum over $N \geq k+l+1$ to obtain after some manipulations on both sides

$$(-1)^{k+l} \frac{x^2}{(k+l-1)!} \left\{ \frac{\log^{k+l-1}(1+x)}{1+x} - x^{k+l-1} e^{-x} \right\}. \qquad \square$$

Lemmas 5.1 and 5.2 allow us to prove the following identity for $F_q(x,y)$.

Lemma 5.3 *For $q \geq 1$,*

$$F_q(x,y) = \frac{1}{(q-1)!} \frac{\Gamma(x+y+q)}{\Gamma(x+y)}. \tag{50}$$

Proof: We compute

$$\begin{aligned} F_q(x,y) &= qA_{q+1}(x,y) - (q-1)A_q(x,y) \\ &= \sum_{k,l=0}^{q} \left(qa_{q+1}[k,l] - (q-1)a_q[k,l]\right) x^k y^l + (q-1) \sum_{k,l:\max(k,l)=q}^{q} a_q[k,l] x^k y^l. \end{aligned}$$

From Lemma 5.1, $a_q[q,l] = a_q[k,q] = 0$. Thus, we have that $F_q(x,y) = \sum_{k,l} f_q[k,l] x^k y^l$, with $f_q[k,l] = qa_{q+1}[k,l] - (q-1)a_q[k,l]$. Using (47), we obtain

$$\begin{aligned} f_q[k,l] &= \frac{(-1)^q}{(q-2)!}\binom{k+l}{k} \sum_{j=k+l}^{q-1} (-1)^j \binom{j-1}{k+l-1} \left(\frac{S_q^{(j)}}{q-1} + S_{q-1}^{(j)}\right) \\ &\quad + \frac{1}{(q-1)!}\binom{k+l}{k}\binom{q-1}{k+l-1}, \end{aligned} \tag{51}$$

and with $\frac{S_q^{(j)}}{q-1} + S_{q-1}^{(j)} = \frac{S_{q-1}^{(j-1)}}{q-1}$ (see [1, 24.1.3.II.A]),

$$f_q[k,l] = \frac{(-1)^q}{(q-1)!}\binom{k+l}{k}\sum_{j=k+l}^{q-1}(-1)^j\binom{j-1}{k+l-1}S_{q-1}^{(j-1)} + \frac{1}{(q-1)!}\binom{k+l}{k}\binom{q-1}{k+l-1}.$$

Invoking relation (49), rewritten as

$$\sum_{j=k+l}^{q-1}(-1)^j\binom{j-1}{k+l-1}\frac{S_{q-1}^{(j-1)}}{(q-1)!} = \frac{(-1)^{k+l}S_q^{(k+l)}}{(q-1)!} + \frac{(-1)^{q+1}}{(q-1)!}\binom{q-1}{k+l-1},$$

gives

$$f_q[k,l] = \frac{(-1)^{q-k-l}}{(q-1)!}\binom{k+l}{k}S_q^{(k+l)}. \tag{52}$$

Hence, using that $S_q^{(m)} = 0$ for all $m > q$

$$\begin{aligned} F_q(x,y) &= \sum_{k,l=0}^{q} f_q[k,l]x^k y^l = \frac{(-1)^q}{(q-1)!}\sum_{k,l=0}^{q}\binom{k+l}{k}S_q^{(k+l)}(-x)^k(-y)^l \\ &= \frac{(-1)^q}{(q-1)!}\sum_{m=0}^{q}\sum_{k=0}^{m}(-1)^m S_q^{(m)}\binom{m}{k}x^k y^{m-k} \\ &= \frac{1}{(q-1)!}\sum_{m=0}^{q}(-1)^{q-m}S_q^{(m)}(x+y)^m = \frac{1}{(q-1)!}\frac{\Gamma(x+y+q)}{\Gamma(x+y)}. \end{aligned}$$

□

Lemma 5.4 *For all a, b and $N \in \mathbb{N}$,*

$$R(a,b) = \sum_{q=0}^{N-1}\frac{\Gamma(a+q)}{\Gamma(b+q)} = \frac{1}{(1+a-b)}\left(\frac{\Gamma(a+N)}{\Gamma(b+N-1)} - \frac{\Gamma(a)}{\Gamma(b-1)}\right). \tag{53}$$

Proof: Verify that

$$R(a,b-1) = \frac{\Gamma(a)}{\Gamma(b-1)} + \sum_{q=1}^{N-1}\frac{\Gamma(a+q)}{\Gamma(b-1+q)} = (b-1)R(a,b) + \sum_{q=1}^{N-1}\frac{q\Gamma(a+q)}{\Gamma(b+q)},$$

and alternatively

$$\begin{aligned} R(a,b-1) &= \frac{\Gamma(a)}{\Gamma(b-1)} + \sum_{q=0}^{N-2}\frac{\Gamma(a+q+1)}{\Gamma(b+q)} \\ &= aR(a,b) + \frac{\Gamma(a)}{\Gamma(b-1)} - a\frac{\Gamma(a+N-1)}{\Gamma(b+N-1)} + \sum_{q=1}^{N-2}\frac{q\Gamma(a+q)}{\Gamma(b+q)}. \end{aligned}$$

Equating both sides yields the desired result. □

6 Appendix II

In this appendix, we prove the moderate deviation bounds, used in Section 3 to prove Theorem 3.1.

Lemma 6.1 *For $c_N = o\left(\sqrt{\log N}\right)$, as $N \to \infty$,*

$$\sum_{|k-\log N| \geq c_N\sqrt{\log N}} p_N^{(k)} \leq 8\exp\left\{-\frac{c_N^2}{2}\left(1 - \frac{c_N}{\sqrt{\log N}}\right)\right\}.$$

Proof: Using the generating function (4) and the Markov inequality, we find for each $x > 0$,

$$\begin{aligned}\sum_{k \geq \log N + c_N\sqrt{\log N}} p_N^{(k)} &= \mathbb{P}(D_{1,N} \geq \log N + c_N\sqrt{\log N}) = \mathbb{P}\left(x^{D_{1,N}} \geq x^{\log N + c_N\sqrt{\log N}}\right)\\ &\leq x^{-\log N - c_N\sqrt{\log N}}\varphi_N(x)\\ &= \left(x^{-\log N - c_N\sqrt{\log N}}\right)\frac{e^{(x-1)\log N}}{\Gamma(x+1)}\left(1 + \mathcal{O}\left(\frac{1}{N}\right)\right).\end{aligned}$$

Hence, for each $x > 0$, and for N large enough, using that $\Gamma(x+1) > \frac{1}{2}$ (cf. [1, 6.3.19]),

$$\sum_{k \geq \log N + c_N\sqrt{\log N}} p_N^{(k)} \leq 4\exp\{-(\log N + c_N\sqrt{\log N})\log x + (x-1)\log N\}.$$

Take $x = 1 + \frac{c_N}{\sqrt{\log N}}$, to get for N sufficiently large

$$\sum_{k \geq \log N + c_N\sqrt{\log N}} p_N^{(k)} \leq 4\exp\left\{-\frac{c_N^2}{2}\left(1 - \frac{c_N}{\sqrt{\log N}}\right)\right\}, \tag{54}$$

using $\log(1+x) \geq x - \frac{x^2}{2}$, $x \downarrow 0$. Similarly we obtain

$$\begin{aligned}\mathbb{P}(D_{1,N} \leq \log N - c_N\sqrt{\log N}) &= \mathbb{P}\left(x^{-D_{1,N}} \geq x^{-\left(\log N - c_N\sqrt{\log N}\right)}\right)\\ &\leq x^{(\log N - c_N\sqrt{\log N})}\varphi_N(x^{-1})\\ &= \left(x^{\log N - c_N\sqrt{\log N}}\right)\frac{e^{(x^{-1}-1)\log N}}{\Gamma(x^{-1}+1)}\left(1 + \mathcal{O}\left(\frac{1}{N}\right)\right),\end{aligned}$$

so that for each $x > 0$, using again that $\Gamma(x^{-1}+1) > \frac{1}{2}$, we end up with

$$\sum_{k \leq \log N - c_N\sqrt{\log N}} p_N^{(k)} \leq 4\exp\{(\log N - c_N\sqrt{\log N})\log x + (x^{-1}-1)\log N\}.$$

Take $x^{-1} = 1 - \frac{c_N}{\sqrt{\log N}}$ and in exactly the same way as above we obtain

$$\sum_{k \leq \log N - c_N\sqrt{\log N}} p_N^{(k)} \leq 4\exp\left\{-\frac{c_N^2}{2}\left(1 - \frac{c_N}{\sqrt{\log N}}\right)\right\}. \tag{55}$$

This completes the proof. $\square$