

STOCHASTIC ENUMERATION METHOD FOR COUNTING TREES*

RADISLAV VAISMAN AND DIRK P. KROESE
THE UNIVERSITY OF QUEENSLAND, BRISBANE 4072, AUSTRALIA
R.VAISMAN@UQ.EDU.AU, KROESE@MATHS.UQ.EDU.AU

Abstract. The problem of estimating the size of a backtrack tree is an important but hard problem in computational sciences. An efficient solution of this problem can have a major impact on the hierarchy of complexity classes. The first randomized procedure, which repeatedly generates random paths through the tree, was introduced by Knuth. Unfortunately, as was noted by Knuth and a few other researchers, the estimator can introduce a large variance and become ineffective in the sense that it underestimates the cost of the tree. Recently, a new sequential algorithm called Stochastic Enumeration (SE) method was proposed by Rubinstein et al. The authors showed numerically that this simple algorithm can be very efficient for handling different counting problems, such as counting the number of satisfiability assignments and enumerating the number of perfect matchings in bipartite graphs. In this paper we introduce a rigorous analysis of SE and show that it results in significant variance reduction as compared to Knuth’s estimator. Moreover, we establish that for almost all random trees the SE algorithm is a *fully polynomial time randomized approximation scheme* (FPRAS) for the estimation of the overall tree size.

Key words. Randomized Algorithms, Monte Carlo sampling, Multilevel Splitting.

AMS subject classifications. 68W20, 68W40.

1. Introduction. The problem of estimating the total cost of a backtrack tree is an important but hard problem from both a theoretical and practical point of view. Consider a tree with node set \mathcal{V} , where a cost $c(v)$ is associated with each node $v \in \mathcal{V}$. In many practical situations it is important to find the total cost of the tree, $\sum_{v \in \mathcal{V}} c(v)$.

For large trees the direct computation of this cost using tree traversal is expensive in terms of computation effort. Hence, a Monte Carlo approach can be beneficial. Knuth [17] introduces a randomized algorithm that gives an unbiased estimator for this problem. The proposed algorithm repeatedly generates a random walk from the tree root to a leaf and counts the encountered node degrees (a more detailed explanation will be provided in the next section).

The main disadvantage of Knuth’s approach is a large variance of the estimator that can occur for some tree instances. A few attempts were made to improve Knuth’s algorithm. For example, Pudrom’s partial backtracking algorithm [22] allows more than one child to be explored by the random walk. This multiple children exploration is achieved by returning to the same node and choosing a different path to continue. Pudrom’s method, which employs an importance sampling approach [27], can offer an improved efficiency. Nevertheless, Chen [6] showed that in this case it is also hard to achieve a significant variance reduction without considerable computational effort. Chen [6] introduced another approach to handle the tree cost estimation problem by using a stratified sampling method. Chen’s algorithm has proven performance guaranty for some randomly generated instances.

The general problem of estimating the cost of a tree belongs to the complexity class called $\#P$ [30]. This complexity class consists of the set of *counting* problems that are associated with a decision problem in NP (non-deterministic polynomial time), e.g., how many solutions does a propositional formula have ($\#SAT$)? The $\#P$ -complete complexity class is a sub-class of $\#P$ consisting of those problems in $\#P$ to which any other problem in $\#P$ can be reduced via a polynomial reduction. $\#SAT$, for example, is $\#P$ -complete. Interestingly, various $\#P$ -complete problems are associated with an easy decision problem, i.e., the corresponding decision prob-

*This work is supported by the Australian Research Council, under grant CE140100049.

lem is in P (polynomial time), such as satisfiability of propositional formulas in *Disjunctive Normal Form* (DNF).

For some #P-complete problems there are known efficient approximations. For example, Karp and Luby [16] introduced a FPRAS for counting the solutions of DNF satisfiability formulas. Similar results were obtained for the knapsack and permanent counting problems by Dyer and Jerrum et al. [8, 14].

Unfortunately, there are also many negative results. For example, [9, 29] showed that counting the number of vertex covers remains hard, even when restricted to planar bipartite graphs of bounded degree or regular graphs of constant degree. The theoretical importance of this counting problem follows from the fact that finding an efficient algorithm (for example FPRAS) to some specific problems will result in the collapse of polynomial hierarchy. For example, it was showed in [18] that there is no efficient approximation algorithm capable to count vertex covers if some vertex can appear in 6 clauses unless $NP=RP$, where RP stands for the *Randomized Polynomial Time* complexity class [19]. The findings of Liu and Lu are not very encouraging, since there is a direct correspondence between counting combinatorial objects and their associated backtrack trees. For example, we can define $c(v) = 1$ for every $v \in \mathcal{V}$ that corresponds to a valid vertex cover. Thus, the estimation of a backtrack tree is at least as hard as the estimation of the number of vertex covers.

There are two main approaches to tackle such difficult counting problems. The first one is Markov Chain Monte Carlo (MCMC) and the second is sequential importance sampling (SIS). Both approaches exploit the finding of Jerrum et al. [15] that counting is equivalent to uniform sampling over a suitably restricted set. MCMC methods sample from such restricted regions by constructing an ergodic Markov chain with limiting distribution equal to the desired uniform distribution. A number of MCMC approaches with good empirical performance have been proposed [5, 13, 24, 26]. There are also many examples of successful SIS implementations on various counting problems; see, for example, [4, 7, 16, 25]. More recent advances and background material can be found in [3].

In this article, we develop an adaptation of the *Stochastic Enumeration* (SE) method, originally proposed in [25], for backtrack tree estimation. The SE algorithm belongs to the sequential importance sampling family of algorithms, but the main difference between general SIS procedures and SE is that the latter employs polynomial oracles during the execution and runs multiple trajectories in parallel instead of repeatedly running single trajectories. It was shown numerically in [28] that SE may introduce a significant variance reduction. The SE algorithm has a budget parameter that limits the number of parallel random walks. We show that the SE algorithm is an extension of Knuth’s estimator, in sense that if we set the budget equal to one, the SE becomes equivalent to Knuth’s estimator. The reader might think that the SE extension is just Knuth’s estimator that runs some trajectories in parallel (and therefore brings little added value), but careful consideration will reveal an additional and much more important property. The SE algorithm does not operate on the original tree but rather on the associated “hyper tree”. We show that this property has a crucial impact on the SE performance. In particular, it turns the SE into a *splitting* algorithm. It was shown that such splitting mechanisms can introduce a significant variance reduction [28]. For a background on the splitting methods, see [5, 10, 11]. For additional detailed explanation about the general SE method, we refer to [28].

The rest the paper is organized as follows. In Section 2 we introduce the SE algorithm for counting trees. In Sections 3 and 4 we develop important theoretical properties of the SE estimator. In particular, we prove that the estimator is unbiased, and that it has a recursive expression for the variance that is similar to the one given in [17]. Moreover, we develop an upper bond for the variance and show that for almost all random trees, the SE algorithm is a FPRAS for the estimation

of the overall tree size. In Section 5 we provide additional numerical evidence for the accuracy of our method. Finally, in Section 6 we summarize our findings and discuss possible directions for future research.

2. Estimating the total cost of a tree. The SE extension is heavily based on the Knuth's estimator [17], so we feel that a short review of the latter is beneficial. Our setting is as follows. Consider a rooted tree $T = (\mathcal{V}, \mathcal{E})$ with node set \mathcal{V} and edge set \mathcal{E} (so that $|\mathcal{E}| = |\mathcal{V}| - 1$). We denote the root of the tree by v_0 , and for any $v \in \mathcal{V}$ the subtree rooted at v is denoted by T_v . With each node v is associated a non-negative cost $c(v)$. The main quantity of interest is the total cost of the tree,

$$\text{Cost}(T) = \sum_{v \in \mathcal{V}} c(v)$$

or, more generally, the total cost of a subtree T_v — denoted by $\text{Cost}(T_v)$. For each node v we denote the set of successors of v by $S(v)$. Knuth's Algorithm 2.1, [17], outputs an unbiased estimator of the total cost of a subtree T_v rooted at v .

Algorithm 2.1: Knuth's algorithm for estimating the cost of a backtrack tree

Input: A tree T_v of height h , rooted at v .

Output: An unbiased estimator C of the total cost of tree T_v .

- 1 **(Initialization):** Set $k \leftarrow 0$, $D \leftarrow 1$, $X_0 = v$ and $C \leftarrow c(X_0)$. Here D is the product of all node degrees encountered in the tree.
 - 2 **(Compute the successors):** Let $S(X_k)$ be the set of all successors of X_k and let D_k be the number of elements of $S(X_k)$. If $k = h$ or when $S(X_k)$ is empty, set $D_k = 0$.
 - 3 **(Terminal position?):** If $D_k = 0$, the algorithm stops, returning C as an estimator of $\text{Cost}(T_v)$.
 - 4 **(Advance):** Choose an element $X_{k+1} \in S(X_k)$ at random, each element being equally likely. (Thus, each choice occurs with probability $1/D_k$.) Set $D \leftarrow D_k D$, then set $C \leftarrow C + c(X_{k+1})D$. Increase k by 1 and return to Step 2.
-

It was shown in [17] that for any sub-tree T_v of height h this algorithm has the following properties.

- It creates a random walk $\{X_0, X_1, \dots, X_\tau\}$ on the tree (from root $X_0 = v$ to a leaf at level $\tau \leq h$) and returns the estimator

$$C = c(X_0) + D_0 c(X_1) + D_0 D_1 c(X_2) + D_0 D_1 D_2 c(X_3) + \dots + \left(\prod_{0 \leq j \leq \tau-1} D_j \right) c(X_\tau),$$

where X_j is the random variable that represents the tree node that was encountered at step j of the random walk, and random variable D_j is the corresponding degree of X_j .

- This estimator is unbiased:

$$\mathbb{E}(C(T_v)) = \text{Cost}(T_v).$$

- The variance of the estimator is given by:

$$(2.1) \quad \text{Var}(C(T_v)) = d \sum_{1 \leq j \leq d} \text{Var}(C(T_{w_j})) + \sum_{1 \leq i < j \leq d} (\text{Cost}(T_{w_i}) - \text{Cost}(T_{w_j}))^2,$$

where w_1, \dots, w_d are the successors of v .

REMARK 2.1 (Knuth's Importance Sampling). Instead of choosing one of v 's successors from $S(v) = \{w_1, \dots, w_d\}$ with equal probability, one can choose w_j with probability p_j for $j = 1, \dots, d$, provided that $\sum_{1 \leq j \leq d} p_j = 1$, where one should set $D \leftarrow (1/p_j)D$ in Step 4 (instead of $D \leftarrow D_0D$), when w_j is selected. Knuth shows that in this case the estimator remains unbiased, and that the variance is given by:

$$(2.2) \quad \text{Var}(C(T_v)) = \sum_{1 \leq j \leq d} \frac{\text{Var}(C(T_{w_j}))}{p_j} + \sum_{1 \leq i < j \leq d} p_i p_j \left(\frac{\text{Cost}(T_{w_i})}{p_i} - \frac{\text{Cost}(T_{w_j})}{p_j} \right)^2.$$

In fact, if during the algorithm execution one is able to “guess” the costs of the subtrees, a zero-variance estimator can be obtained.

Clearly, the last assumption is not practical, but some algorithms have tried to make improvements by estimating those probabilities in a better manner [2]. Unfortunately, this strategy may even damage the algorithm performance, as explained in the following example.

EXAMPLE 2.1 (Choosing nonuniform probabilities). To understand why estimating the importance sampling parameters (the probabilities of choosing the successors in Remark 2.1) is probably hopeless in the general case, consider the simple example in Figure 1. Suppose that the cost of every node is 1. The tree root v_0 has two successors (v_1 and v_2). The node v_1 has a single successor v_3 that is connected to a complete tree with $M - 1$ nodes, such that $M \gg k$, and the node v_2 is connected to $k - 1$ leaves. In this context, the complete tree is a tree in which every level, except possibly the deepest, is entirely filled.

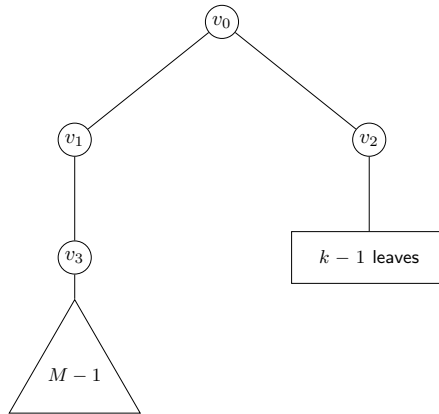


FIG. 1. An example tree.

We will use the recursive formula (2.2) to calculate the variance of the Knuth's estimator. Note, that in this formula, each successor j is chosen with probability p_j

subject to $\sum_{1 \leq j \leq d} p_j = 1$. It is not very hard to see that Knuth's estimator delivers a zero-variance estimator on any complete tree, so $\text{Var}(T_{v_1}) = 0$. It is also easy to verify from (2.2) that $\text{Var}(T_{v_2}) = 0$, because the node v_2 is a father of $k - 1$ leaves. We next consider the following possibilities.

1. Suppose we choose v_1 and v_2 with equal probabilities, so $p_{v_1} = p_{v_2} = 0.5$. Then,

$$\sigma_1^2 = \text{Var}(C) = 0 + (M - k)^2.$$

2. Suppose that one chooses to explore the tree from node v_0 (level zero) to level 2; that is, until node v_3 in the left subtree and the $k - 1$ leaves in the right subtree. Then, reasonable choices for p_{v_1} and p_{v_2} are $1/k$ and $(k - 1)/k$ respectively. In this case,

$$\begin{aligned} \sigma_2^2 = \text{Var}(C) &= 0 + \frac{k-1}{k^2} \left(\frac{M}{\frac{1}{k}} - \frac{k}{\frac{k-1}{k}} \right)^2 \\ &= (k-1) \left(\frac{\frac{M}{\frac{1}{k}} - \frac{k}{\frac{k-1}{k}}}{k} \right)^2 = (k-1) \left(M - \frac{k}{k-1} \right)^2. \end{aligned}$$

Consider now the benefit one can get from this particular change of measure. Clearly,

$$\frac{\sigma_2^2}{\sigma_1^2} = \frac{(k-1) \left(M - \frac{k}{k-1} \right)^2}{(M-k)^2},$$

which approaches $k-1$ as $M \rightarrow \infty$. Hence, for large M the performance is $k-1$ times worse when applying the change of measure. So, it would be better to choose the successors with equal probabilities. The latter will not only save on computational effort needed to compute the new probabilities (p_{v_1} and p_{v_2}) but will also reduce the estimator's variance.

The SE philosophy is that such estimation of importance weights is probably hopeless and potentially harmful. Instead, SE will try to improve the estimation by running a number of parallel random walks. We will show that this simple strategy can be very beneficial in sense that it turns the SE into a *splitting* algorithm. To start with, we will need a few definitions.

DEFINITION 2.1 (Hyper nodes and forests). *Let $\{v_1, \dots, v_r\} \subseteq \mathcal{V}$ be tree nodes.*

1. *We call a collection $\mathbf{v} = \{v_1, \dots, v_r\}$ of distinct nodes in the same level of the tree a hyper node of cardinality $|\mathbf{v}| = r$.*
2. *Let \mathbf{v} be a hyper node. Generalizing the tree node cost, we define the cost of the hyper node as*

$$c(\mathbf{v}) = \sum_{v \in \mathbf{v}} c(v).$$

3. *Let \mathbf{v} be a hyper node. Define the set of successors of \mathbf{v} as*

$$S(\mathbf{v}) = \bigcup_{v \in \mathbf{v}} S(v).$$

4. *Let \mathbf{v} be a hyper node and let $B \in \mathbb{N}$, $B \geq 1$. Define:*

$$H(\mathbf{v}) = \begin{cases} \{S(\mathbf{v})\} & \text{if } |S(\mathbf{v})| \leq B \\ \{\mathbf{w} \mid \mathbf{w} \subseteq S(\mathbf{v}), |\mathbf{w}| = B\} & \text{if } |S(\mathbf{v})| > B. \end{cases}$$

5. For each hyper node \mathbf{v} let

$$T_{\mathbf{v}} = \bigcup_{v \in \mathbf{v}} T_v$$

be the forest of trees rooted at \mathbf{v} . See Figure 2 for an example of hyper node $\mathbf{v} = \{v_1, v_2, v_3, v_4\}$ and its corresponding forest $T_{\mathbf{v}} = \{T_{v_1}, T_{v_2}, T_{v_3}, T_{v_4}\}$.

6. Let $T_{\mathbf{v}}$ be the forest of trees rooted at \mathbf{v} . Define

$$S^{(m)}(\mathbf{v}) = \underbrace{S(S(S \cdots (\mathbf{v})))}_{m \text{ times}} \cdots$$

to be the set of vertices at level m . (Note that $S^{(0)}(\mathbf{v}) = \mathbf{v}$.)

7. For each forest rooted at hypernode \mathbf{v} , define its total cost as

$$(2.3) \quad \text{Cost}(T_{\mathbf{v}}) = \sum_{v \in \mathbf{v}} \text{Cost}(T_v).$$

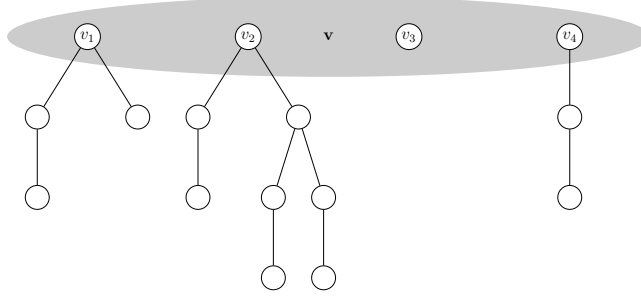


FIG. 2. Hyper node example.

We are ready to state the main SE algorithm.

Algorithm 2.2: Stochastic Enumeration algorithm for estimating the cost of a backtrack tree

Input: A forest $T_{\mathbf{v}}$ of height h rooted at a hypernode \mathbf{v} , and a budget $B \geq 1$.

Output: An unbiased estimator $|\mathbf{v}|C_{\text{SE}}$ of the total cost of forest $T_{\mathbf{v}}$.

- 1 **(Initialization):** Set $k \leftarrow 0$, $D \leftarrow 1$, $\mathbf{X}_0 = \mathbf{v}$ and $C_{\text{SE}} \leftarrow c(\mathbf{X}_0)/|\mathbf{X}_0|$.
 - 2 **(Compute the successors):** Let $S(\mathbf{X}_k)$ be the set of all successors of \mathbf{X}_k .
 - 3 **(Terminal position?):** If $|S(\mathbf{X}_k)| = 0$, the algorithm stops, returning $|\mathbf{v}|C_{\text{SE}}$ as an estimator of $\text{Cost}(T_{\mathbf{v}})$.
 - 4 **(Advance):** Choose hyper node $\mathbf{X}_{k+1} \in H(\mathbf{X}_k)$ at random, each choice being equally likely. (Thus, each choice occurs with probability $1/|H(\mathbf{X}_k)|$.) Set $D_k = \frac{|S(\mathbf{X}_k)|}{|\mathbf{X}_k|}$ and $D \leftarrow D_k D$, then set $C_{\text{SE}} \leftarrow C_{\text{SE}} + \left(\frac{c(\mathbf{X}_{k+1})}{|\mathbf{X}_{k+1}|} \right) D$. Increase k by 1 and return to Step 2.
-

Let us have a closer look at Algorithm 2.2 and compare it with Knuth's Algorithm 2.1. Note that in Step 1, Knuth's algorithm initializes X_0 to be v while SE sets $\mathbf{X}_0 = \mathbf{v}$. In general, Algorithm 2.1 will operate on a single node at each step, while SE will maintain a collection of such nodes (hyper nodes).

The SE algorithm continues to move down the tree to $\mathbf{X}_1, \mathbf{X}_2, \dots$ by examining the hyper nodes from the sets $H(S(\mathbf{X}_0)), H(S(\mathbf{X}_1)), \dots$. This procedure is

performing a random walk on the hyper nodes of the original tree T . It follows that SE is actually a generalization of Knuth’s algorithm. Namely, consider the SE algorithm with budget $B = 1$. In this case, all the hyper nodes have cardinality of 1 and we get the Knuth’s estimator. Similar to Knuth’s Algorithm 2.1, the output of Algorithm 2.2 is a random variable

$$C_{\text{SE}} = \frac{c(\mathbf{X}_0)}{|\mathbf{X}_0|} + \frac{|S(\mathbf{X}_0)|}{|\mathbf{X}_0|} \frac{c(\mathbf{X}_1)}{|\mathbf{X}_1|} + \frac{|S(\mathbf{X}_0)|}{|\mathbf{X}_0|} \frac{|S(\mathbf{X}_1)|}{|\mathbf{X}_1|} \frac{c(\mathbf{X}_2)}{|\mathbf{X}_2|} \\ + \cdots + \left(\prod_{0 \leq j \leq \tau-1} \frac{|S(\mathbf{X}_j)|}{|\mathbf{X}_j|} \right) \frac{c(\mathbf{X}_\tau)}{|\mathbf{X}_\tau|},$$

where $\tau \leq h$.

At first glance, one might think that the SE extension is just Knuth’s estimator that runs B parallel trajectories while estimating average node costs and degrees at each tree level. We will show in the following sections that this is not so. Before we proceed with a rigorous analysis, consider the following example. Despite its simplicity, it provides a good demonstration of SE’s built-in splitting mechanism and its corresponding benefits.

EXAMPLE 2.2. Consider the “hair brush” tree T in Figure 3 and suppose that the costs of all vertices are zero except for v_{n+1} , which has a cost of unity. Our goal is to estimate the cost of this tree, which obviously satisfies $\text{Cost}(T) = 1$. It will become clear from the following discussion that the budget parameter B is controlling the splitting capability of SE Algorithm 2.2. We will consider two cases. In particular we examine the behavior of the Algorithm 2.2 with budgets $B = 1$ and $B = 2$ respectively.

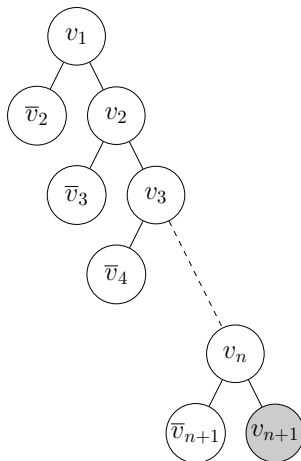


FIG. 3. *The hair brush tree.*

- If we set $B = 1$, the SE Algorithm 2.2 essentially adopts the behavior of Knuth’s estimator, [17]. Note that in this case the algorithm reaches the vertex of interest, v_{n+1} , with probability $1/2^n$ and with $D = 2^n$. In all other cases, the algorithm terminates with some D' and a zero cost node \bar{v}_i , $i = 2, \dots, n + 1$. It follows that the expectation and variance of the corresponding SE estimator are

$$\mathbb{E}(C_{\text{SE}}) = \frac{1}{2^n} \cdot 2^n \cdot 1 + \frac{2^n - 1}{2^n} \cdot D' \cdot 0 = 1,$$

and

$$\begin{aligned}\mathbb{E}(C_{\text{SE}}^2) &= \frac{1}{2^n} \cdot (2^n \cdot 1)^2 + \frac{2^n - 1}{2^n} \cdot (D' \cdot 0)^2 = 2^n \Rightarrow \\ &\Rightarrow \text{Var}(C_{\text{SE}}) = \mathbb{E}(C_{\text{SE}}^2) - \mathbb{E}(C_{\text{SE}})^2 = 2^n - 1.\end{aligned}$$

- On the other hand, setting $B = 2$ will force Algorithm 2.2 to reach v_{n+1} with probability 1. To see this, note that at each tree level ($m = 2, \dots, n$), $\mathbf{X}_m = \{v_m, \bar{v}_m\}$. The latter is true because \bar{v}_m has no successors at all and v_m has exactly two successors respectively. Following the execution steps of the Algorithm 2.2 one can verify that at the final iteration the cost of the hyper node $\mathbf{X}_{n+1} = \{\bar{v}_{n+1}, v_{n+1}\}$ is $0 + 1 = 1$, so

$$\left(\frac{c(\mathbf{X}_{n+1})}{|\mathbf{X}_{n+1}|} \right) = \frac{1}{2}.$$

In addition, the final value of D is

$$D = 2 \cdot \underbrace{1 \cdots 1}_{n-1 \text{ times}} = 2.$$

It follows that the expectation and variance of the corresponding SE estimator are

$$\mathbb{E}(C_{\text{SE}}) = 1 \cdot 2 \cdot \frac{1}{2} = 1,$$

and

$$\begin{aligned}\mathbb{E}(C_{\text{SE}}^2) &= 1 \cdot \left(2 \cdot \frac{1}{2} \right)^2 = 1 \Rightarrow \\ &\Rightarrow \text{Var}(C_{\text{SE}}) = \mathbb{E}(C_{\text{SE}}^2) - \mathbb{E}(C_{\text{SE}})^2 = 1 - 1 = 0.\end{aligned}$$

By increasing the budget B from 1 to 2 we managed to achieve remarkable variance reduction, from $2^n - 1$ to zero. Clearly, we presented an artificial example but it is also illustrative enough for our purposes. Generally speaking, by increasing the budget we cannot expect to obtain a zero variance estimator for hard approximation problems, but we do hope to achieve a significant variance reduction. We next proceed to the analysis.

3. Analysis. Recall from Definition 2.1 that for any hypernode \mathbf{v} , $T_{\mathbf{v}}$ denotes the forest rooted at \mathbf{v} , and that its cost is $\text{Cost}(T_{\mathbf{v}})$. Let $|\mathbf{v}| C_{\text{SE}}(T_{\mathbf{v}})$ be the corresponding estimator, as returned by Algorithm 2.2. The following theorem shows that this estimator is unbiased.

THEOREM 3.1 (Unbiased Estimator). *Let \mathbf{v} be a hyper node and let $H(S(\mathbf{v})) = \{\mathbf{w}_1, \dots, \mathbf{w}_d\}$ be its set of hyper children. Then,*

$$(3.1) \quad \mathbb{E}(C_{\text{SE}}(T_{\mathbf{v}})) = \frac{\text{Cost}(T_{\mathbf{v}})}{|\mathbf{v}|}.$$

Proof. By the recursive structure of Algorithm 2.2, we have

$$(3.2) \quad C_{\text{SE}}(T_{\mathbf{v}}) = \frac{c(\mathbf{v})}{|\mathbf{v}|} + \frac{|S(\mathbf{v})|}{|\mathbf{v}|} C_{\text{SE}}(T_{\mathbf{w}}),$$

where \mathbf{W} is a hyperchild of \mathbf{v} selected uniformly at random from $H(S(\mathbf{v}))$. To show (3.1) we proceed by induction on the tree height.

- $h = 0$:

$$\begin{aligned}\mathbb{E}(C_{\text{SE}}(T_{\mathbf{v}})) &= \mathbb{E}\left(\frac{c(\mathbf{v})}{|\mathbf{v}|} + \frac{|S(\mathbf{v})|}{|\mathbf{v}|} \cdot 0\right) \\ &= \frac{c(\mathbf{v})}{|\mathbf{v}|} = \frac{\sum_{v \in \mathbf{v}} c(v)}{|\mathbf{v}|} = \frac{\text{Cost}(T_{\mathbf{v}})}{|\mathbf{v}|}.\end{aligned}$$

- Suppose that the proposition is correct for heights $0, \dots, h-1$. Combining this with (3.2) we get

$$\begin{aligned}\mathbb{E}(C_{\text{SE}}(T_{\mathbf{v}})) &= \mathbb{E}\left(\frac{c(\mathbf{v})}{|\mathbf{v}|} + \frac{|S(\mathbf{v})|}{|\mathbf{v}|} C_{\text{SE}}(T_{\mathbf{W}})\right) \\ &= \frac{c(\mathbf{v})}{|\mathbf{v}|} + \frac{|S(\mathbf{v})|}{|\mathbf{v}|} \left(\frac{1}{d} \sum_{1 \leq j \leq d} \mathbb{E}(C_{\text{SE}}(T_{\mathbf{w}_j}))\right) \\ &\stackrel{\text{hypothesis}}{=} \frac{c(\mathbf{v})}{|\mathbf{v}|} + \frac{|S(\mathbf{v})|}{|\mathbf{v}|} \left(\frac{1}{d} \sum_{1 \leq j \leq d} \frac{\text{Cost}(T_{\mathbf{w}_j})}{|\mathbf{w}_j|}\right).\end{aligned}$$

Consider now the following two cases.

1. $|S(\mathbf{v})| \leq B$. Hence, $H(S(\mathbf{v})) = \{\mathbf{w}_1\}$, $|S(\mathbf{v})| = |\mathbf{w}_1|$, and $d = 1$, so that

$$\begin{aligned}\frac{c(\mathbf{v})}{|\mathbf{v}|} + \frac{|S(\mathbf{v})|}{|\mathbf{v}|} \left(\frac{1}{d} \sum_{1 \leq j \leq d} \frac{\text{Cost}(T_{\mathbf{w}_j})}{|\mathbf{w}_j|}\right) &= \frac{c(\mathbf{v})}{|\mathbf{v}|} + \frac{|S(\mathbf{v})|}{|\mathbf{v}|} \frac{\text{Cost}(T_{\mathbf{w}_1})}{|\mathbf{w}_1|} \\ &= \frac{c(\mathbf{v}) + \text{Cost}(T_{\mathbf{w}_1})}{|\mathbf{v}|} = \frac{\text{Cost}(T_{\mathbf{v}})}{|\mathbf{v}|}.\end{aligned}$$

2. $|S(\mathbf{v})| > B$. In this case, there is a set of possible hyper nodes that will be chosen uniformly at random from $H(S(\mathbf{v}))$. So,

$$H(S(\mathbf{v})) = \{\mathbf{w}_1, \dots, \mathbf{w}_d\}, \quad d = |H(S(\mathbf{v}))| = \binom{|S(\mathbf{v})|}{B} > 1$$

and $|\mathbf{w}_j| = B$ for all $j = 1, \dots, d$. We continue with

$$\begin{aligned}&\frac{c(\mathbf{v})}{|\mathbf{v}|} + \frac{|S(\mathbf{v})|}{|\mathbf{v}|} \left(\frac{1}{d} \sum_{1 \leq j \leq d} \frac{\text{Cost}(T_{\mathbf{w}_j})}{|\mathbf{w}_j|}\right) \\ &= \frac{c(\mathbf{v})}{|\mathbf{v}|} + \frac{|S(\mathbf{v})|}{|\mathbf{v}|} \left(\binom{|S(\mathbf{v})|}{B}^{-1} \sum_{1 \leq j \leq d} \frac{\text{Cost}(T_{\mathbf{w}_j})}{B}\right) \\ &\stackrel{(2.3)}{=} \frac{c(\mathbf{v})}{|\mathbf{v}|} + \frac{|S(\mathbf{v})|}{|\mathbf{v}|} \left(\binom{|S(\mathbf{v})|}{B}^{-1} \sum_{1 \leq j \leq d} \frac{\sum_{w \in \mathbf{w}_j} \text{Cost}(T_w)}{B}\right) \\ &\stackrel{(*)}{=} \frac{c(\mathbf{v})}{|\mathbf{v}|} + \frac{|S(\mathbf{v})|}{|\mathbf{v}|} \left(\binom{|S(\mathbf{v})|}{B}^{-1} \frac{\sum_{w \in S(\mathbf{v})} \text{Cost}(T_w)}{|S(\mathbf{v})|}\right) \\ &= \frac{c(\mathbf{v})}{|\mathbf{v}|} + \frac{|S(\mathbf{v})|}{|\mathbf{v}|} \left(\frac{1}{|S(\mathbf{v})|} \sum_{w \in S(\mathbf{v})} \text{Cost}(T_w)\right) \\ &\stackrel{(2.3)}{=} \frac{c(\mathbf{v})}{|\mathbf{v}|} + \frac{|S(\mathbf{v})|}{|\mathbf{v}|} \left(\frac{\text{Cost}(T_{S(\mathbf{v})})}{|S(\mathbf{v})|}\right) = \frac{c(\mathbf{v}) + \text{Cost}(T_{S(\mathbf{v})})}{|\mathbf{v}|} = \frac{\text{Cost}(T_{\mathbf{v}})}{|\mathbf{v}|},\end{aligned}$$

where (*) follows from Lemma A.1. To see this, substitute

$$|S(\mathbf{v})| = n, \text{Cost}(T_{w_j}) = r_j \text{ for } 1 \leq j \leq n, u = B, \text{ and } d = \begin{pmatrix} |S(\mathbf{v})| \\ |\mathbf{v}| \end{pmatrix}.$$

□

If for the original tree T with root v_0 we define $\mathbf{v}_0 = \{v_0\}$, then the forest $T_{\mathbf{v}_0}$ is identical to T and, with $|\mathbf{v}_0| = 1$, Theorem 3.1 yields the following corollary.

COROLLARY 3.2 (Unbiased tree estimator). *Let T be a tree rooted at v_0 . Then, SE returns unbiased estimator for the total tree cost; that is,*

$$\mathbb{E}(C_{\text{SE}}(T_{\mathbf{v}_0})) = \text{Cost}(T).$$

Next we can also derive the recursive expression for the variance.

THEOREM 3.3 (Stochastic Enumeration Algorithm Variance). *Let \mathbf{v} be a hyper node and let $H(S(\mathbf{v})) = \{\mathbf{w}_1, \dots, \mathbf{w}_d\}$ be its set of hyper children. Then,*

$$(3.3) \quad \text{Var}(C_{\text{SE}}(T_{\mathbf{v}})) = \frac{\left(\frac{|S(\mathbf{v})|}{|\mathbf{v}|}\right)^2}{d} \sum_{1 \leq j \leq d} \text{Var}(C_{\text{SE}}(T_{\mathbf{w}_j})) \\ + \underbrace{\frac{\left(\frac{|S(\mathbf{v})|}{|\mathbf{v}|}\right)^2}{d^2} \sum_{1 \leq i < j \leq d} \left(\frac{\text{Cost}(T_{\mathbf{w}_i})}{|\mathbf{w}_i|} - \frac{\text{Cost}(T_{\mathbf{w}_j})}{|\mathbf{w}_j|} \right)^2}_{(*)}$$

Proof. Using (3.2) and the law of total variance, we have

$$\text{Var}(C_{\text{SE}}(T_{\mathbf{v}})) = \text{Var}\left(\frac{c(\mathbf{v})}{|\mathbf{v}|} + \frac{|S(\mathbf{v})|}{|\mathbf{v}|} C_{\text{SE}}(T_{\mathbf{W}})\right) = \left(\frac{|S(\mathbf{v})|}{|\mathbf{v}|}\right)^2 \text{Var}(C_{\text{SE}}(T_{\mathbf{W}})) \\ = \left(\frac{|S(\mathbf{v})|}{|\mathbf{v}|}\right)^2 \left(\mathbb{E}(\text{Var}(C_{\text{SE}}(T_{\mathbf{W}}) | \mathbf{W})) + \text{Var}(\mathbb{E}(C_{\text{SE}}(T_{\mathbf{W}}) | \mathbf{W}))\right),$$

where \mathbf{W} is a hyperchild of \mathbf{v} selected uniformly at random from $H(S(\mathbf{v}))$. We complete the proof by noting that

$$\mathbb{E}(\text{Var}(C_{\text{SE}}(T_{\mathbf{W}}) | \mathbf{W})) = \frac{1}{d} \sum_{1 \leq j \leq d} \text{Var}(C_{\text{SE}}(T_{\mathbf{w}_j}))$$

and that

$$\text{Var}(\mathbb{E}(C_{\text{SE}}(T_{\mathbf{W}}) | \mathbf{W})) \stackrel{\text{Theorem 3.1}}{=} \text{Var}\left(\frac{\text{Cost}(T_{\mathbf{W}})}{|\mathbf{W}|}\right) = \\ \stackrel{(*)}{=} \frac{1}{d^2} \sum_{1 \leq i \leq d} \sum_{1 \leq j \leq d} \frac{1}{2} \left(\frac{\text{Cost}(T_{\mathbf{w}_i})}{|\mathbf{w}_i|} - \frac{\text{Cost}(T_{\mathbf{w}_j})}{|\mathbf{w}_j|} \right)^2 \\ = \frac{1}{d^2} \sum_{1 \leq i < j \leq d} \left(\frac{\text{Cost}(T_{\mathbf{w}_i})}{|\mathbf{w}_i|} - \frac{\text{Cost}(T_{\mathbf{w}_j})}{|\mathbf{w}_j|} \right)^2,$$

where (*) follows from the expression for variance of a set of d equally likely values that can be expressed, without directly referring to the mean, in terms of squared deviations of all points from each other [31]. □

REMARK 3.1. Note that under the settings of Theorem 3.3 and provided that we set the budget of the SE algorithm to be equal to 1, the variance of SE Algorithm

2.2 given in (3.3) becomes equal to the variance of Knuth's Algorithm 2.1 given in (2.1). This follows from the fact that in this case, $|S(\mathbf{v})| = d$ and all hyper nodes have the cardinality of 1.

Recall that in Example 2.2 when we set the budget B equal to the maximum of the number of nodes in each tree level ($B = 2$), the hyper tree collapses and SE returns a zero variance estimator. This true more generally, as summarized in the following corollary.

COROLLARY 3.4 (Zero Variance Estimator). *Let \mathbf{v} be a hyper root of the forest $T_{\mathbf{v}}$ and let $S^{(m)}(\mathbf{v})$ be the set of nodes at the same tree level m , $m = 0, \dots, h$. Then, the SE algorithm with budget*

$$B = \max_{0 \leq m \leq h} \{|S^{(m)}(\mathbf{v})|\}$$

results in a zero variance estimator.

Proof. The proof is an immediate consequence of Theorem 3.3 combined with the fact that the resulting hyper tree has a single hyper node $\mathbf{v}^{(m)}$ at each tree level $0 \leq m \leq h$ and its corresponding degree is equal to one or zero depending whether it is a leaf hyper node or not. More formally, let $\mathbf{v}^{(0)}$ be the hyper root of the tree and

$$\begin{aligned} H(S(\mathbf{v}^{(0)})) &= \{\mathbf{v}^{(1)}\}, H(S(\mathbf{v}^{(1)})) = \{\mathbf{v}^{(2)}\}, \dots, \\ &\dots, H(S(\mathbf{v}^{(h-1)})) = \{\mathbf{v}^{(h)}\}, H(S(\mathbf{v}^{(h)})) = \{\emptyset\} \end{aligned}$$

be the sets of hyper children at tree levels $0, \dots, h$. Then, from Theorem 3.3,

$$\text{Var}(C_{\text{SE}}(T_{\mathbf{v}^{(0)}})) \underbrace{=}_{d=1} \left(\frac{|S(\mathbf{v}^{(0)})|}{|\mathbf{v}^{(0)}|} \right)^2 \text{Var}(C_{\text{SE}}(T_{\mathbf{v}^{(1)}})) + 0.$$

Iterating the above equation results in

$$\text{Var}(C_{\text{SE}}(T_{\mathbf{v}^{(0)}})) = \prod_{0 \leq m \leq h-1} \left(\frac{|S(\mathbf{v}^{(m)})|}{|\mathbf{v}^{(m)}|} \right)^2 \text{Var}(C_{\text{SE}}(T_{\mathbf{v}^{(h)}})) = 0,$$

because

$$\text{Var}(C_{\text{SE}}(T_{\mathbf{v}^{(h)}})) = 0,$$

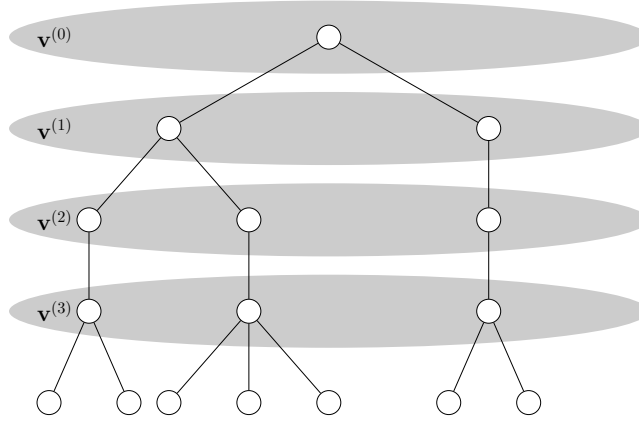
thus completing the proof. \square

Clearly, the most interesting behavior of SE occurs when the number of nodes in the tree levels is greater than the predefined budget. However, in practice, the first few levels are usually "fully enumerated" by the SE algorithm. Consider as an example the tree in Figure 4. Note that for budget $B = 3$ the last level that still can be fully enumerated is level $m = 3$ and the first split of the hyper nodes will occur at level 4. In this example, the degree of $\mathbf{v}^{(0)}$, $\mathbf{v}^{(1)}$ and $\mathbf{v}^{(2)}$ is 1. The degree of $\mathbf{v}^{(3)}$ is

$$\binom{|S(\mathbf{v}^{(3)})|}{|\mathbf{v}^{(3)}|} = \binom{7}{3}$$

respectively. The following lemma summarizes this behavior.

LEMMA 3.5 (SE full enumeration). *Let \mathbf{v} be a hyper root of the forest $T_{\mathbf{v}}$ and let $S^{(m)}(\mathbf{v})$ be the set of nodes at the same tree level m , $m = 0, \dots, h$. Let $1 \leq m^* \leq h - 1$ be the first tree level such that $|S^{(m^*)}(\mathbf{v})| > B$, where B is the*

FIG. 4. SE full enumeration with budget $B = 3$.

budget of the SE algorithm. Suppose that $\mathbf{v}^{(0)}, \dots, \mathbf{v}^{(m^*-1)}$ are the hyper nodes at levels $0, \dots, m^* - 1$ encountered by SE. Then,

$$\text{Var}(C_{\text{SE}}(T_{\mathbf{v}^{(0)}})) = \left(\frac{|\mathbf{v}^{(m^*-1)}|}{|\mathbf{v}^{(0)}|} \right)^2 \text{Var}(C_{\text{SE}}(T_{\mathbf{v}^{(m^*-1)}})).$$

Proof. Similar to the proof of Corollary 3.4, the variance is given by

$$\text{Var}(C_{\text{SE}}(T_{\mathbf{v}^{(0)}})) = \prod_{0 \leq m \leq m^*-2} \left(\frac{|S(\mathbf{v}^{(m)})|}{|\mathbf{v}^{(m)}|} \right)^2 \text{Var}(C_{\text{SE}}(T_{\mathbf{v}^{(m^*-1)}})).$$

We complete the proof by noting that

$$|S(\mathbf{v}^{(m)})| = |\mathbf{v}^{(m+1)}| \text{ for } m = 0, \dots, m^* - 2$$

and thus

$$\prod_{0 \leq m \leq m^*-2} \left(\frac{|S(\mathbf{v}^{(m)})|}{|\mathbf{v}^{(m)}|} \right)^2 = \left(\frac{|S(\mathbf{v}^{(m^*-2)})|}{|\mathbf{v}^{(0)}|} \right)^2 = \left(\frac{|\mathbf{v}^{(m^*-1)}|}{|\mathbf{v}^{(0)}|} \right)^2$$

is a telescopic product. \square

Recall that the *coefficient of variation* (CV) of a random variable Z is defined as

$$\text{CV} = \sqrt{\frac{\text{Var}(Z)}{\mathbb{E}(Z)^2}}.$$

Let Z be an unbiased estimator of ℓ and consider the Monte Carlo algorithm that outputs the average of N independent copies Z_1, \dots, Z_N of Z :

$$\hat{\ell} = \frac{1}{N} \sum_{1 \leq j \leq N} Z_j.$$

The CV is an important measure of efficiency of such a Monte Carlo algorithm because the *relative error* (RE) of $\hat{\ell}$ is

$$(3.4) \quad \text{RE}(\hat{\ell}) = \sqrt{\frac{\text{Var}(Z)}{N\mathbb{E}(Z)^2}} = \frac{\text{CV}}{\sqrt{N}}.$$

Clearly, the computational effort that is needed to achieve a small RE depends on the CV, so if the latter is large, the number of experiments required to achieve a reasonable relative error will also be unmanageable. Having in mind the above discussion, we would like to obtain the upper bound on the variance of SE.

3.1. Upper bound for the variance. For the rest of this section our setting is as follows. Let \mathbf{v} be a hyper root of the forest $T_{\mathbf{v}}$ and let $S^{(m)}(\mathbf{v})$ be the set of nodes at the same tree level m , $m = 0, \dots, h$. The set of hyper nodes at level m is defined by $H(S^{(m)}(\mathbf{v}))$. Let

$$H(S^{(m)}(\mathbf{v})) = \{\mathbf{w}_1^{(m)}, \dots, \mathbf{w}_d^{(m)}\}.$$

Consider a random variable

$$\frac{\text{Cost}(T_{\mathbf{W}^{(m)}})}{|\mathbf{W}^{(m)}|},$$

where $\mathbf{W}^{(m)}$ is taken uniformly from all hypernodes at level m . Define:

$$\gamma^{(m)} = \text{CV}^2 \left(\frac{\text{Cost}(T_{\mathbf{W}^{(m)}})}{|\mathbf{W}^{(m)}|} \right) = \frac{\text{Var} \left(\frac{\text{Cost}(T_{\mathbf{W}^{(m)}})}{|\mathbf{W}^{(m)}|} \right)}{\left(\mathbb{E} \left(\frac{\text{Cost}(T_{\mathbf{W}^{(m)}})}{|\mathbf{W}^{(m)}|} \right) \right)^2}.$$

LEMMA 3.6 (Upper bound of $(*)$ in (3.3)). *For a forest rooted at \mathbf{v} and for any level $m = 0, \dots, h - 1$, it holds that*

$$\frac{\left(\frac{|S(\mathbf{v}^{(m+1)})|}{|\mathbf{v}^{(m+1)}|} \right)^2}{d^2} \sum_{1 \leq i < j \leq d} \left(\frac{\text{Cost}(T_{\mathbf{w}_i^{(m)}})}{|\mathbf{w}_i^{(m)}|} - \frac{\text{Cost}(T_{\mathbf{w}_j^{(m)}})}{|\mathbf{w}_j^{(m)}|} \right)^2 \leq \gamma^{(m)} \left[\frac{\text{Cost}(T_{\mathbf{v}})}{|\mathbf{v}|} \right]^2.$$

Proof.

$$\begin{aligned} & \frac{\left(\frac{|S(\mathbf{v}^{(m+1)})|}{|\mathbf{v}^{(m+1)}|} \right)^2}{d^2} \sum_{1 \leq i < j \leq d} \left(\frac{\text{Cost}(T_{\mathbf{w}_i^{(m)}})}{|\mathbf{w}_i^{(m)}|} - \frac{\text{Cost}(T_{\mathbf{w}_j^{(m)}})}{|\mathbf{w}_j^{(m)}|} \right)^2 \\ &= \left(\frac{|S(\mathbf{v}^{(m+1)})|}{|\mathbf{v}^{(m+1)}|} \right)^2 \text{Var} \left(\frac{\text{Cost}(T_{\mathbf{W}^{(m)}})}{|\mathbf{W}^{(m)}|} \right) \\ &= \gamma^{(m)} \left(\frac{|S(\mathbf{v}^{(m+1)})|}{|\mathbf{v}^{(m+1)}|} \right)^2 \left(\mathbb{E} \left(\frac{\text{Cost}(T_{\mathbf{W}^{(m)}})}{|\mathbf{W}^{(m)}|} \right) \right)^2 \\ &= \gamma^{(m)} \left(\frac{|S(\mathbf{v}^{(m+1)})|}{|\mathbf{v}^{(m+1)}|} \right)^2 \left(\frac{1}{d} \sum_{1 \leq j \leq d} \frac{\sum_{w \in \mathbf{w}_j^{(m)}} \text{Cost}(T_w)}{|\mathbf{w}_j^{(m)}|} \right)^2 \\ &\stackrel{\text{Lemma A.1}}{\leq} \gamma^{(m)} \left(\frac{|S(\mathbf{v}^{(m+1)})|}{|\mathbf{v}^{(m+1)}|} \right)^2 \left(\frac{\sum_{w \in S(\mathbf{v}^{(m+1)})} \text{Cost}(T_w)}{|S(\mathbf{v}^{(m+1)})|} \right)^2 \\ &\leq \gamma^{(m)} \left(\frac{1}{|\mathbf{v}^{(m+1)}|} \sum_{w \in S(\mathbf{v}^{(m+1)})} \text{Cost}(T_w) \right)^2 \leq \gamma^{(m)} \left[\frac{\text{Cost}(T_{\mathbf{v}})}{|\mathbf{v}|} \right]^2. \end{aligned}$$

□

With the above lemma, we are ready to prove the bound for the variance.

LEMMA 3.7 (Upper bound for variance in (3.3)). *Let $T_{\mathbf{v}}$ be a forest of height h and let \mathcal{P}_h be the power set of $\{0, \dots, h\}$ excluding the empty set. Then,*

$$\text{Var}(C_{\text{SE}}(T_{\mathbf{v}})) \leq \left(\sum_{J \in \mathcal{P}_{h-1}} \prod_{j \in J} \gamma^{(j)} \right) \left[\frac{\text{Cost}(T_{\mathbf{v}})}{|\mathbf{v}|} \right]^2.$$

Proof. The proof is by induction on the forest height combined with the recursive variance formula (3.3). For $h = 1$,

$$\begin{aligned} \text{Var}(C_{\text{SE}}(T_{\mathbf{v}})) &= 0 + \frac{\left(\frac{|S(\mathbf{v})|}{|\mathbf{v}|}\right)^2}{d^2} \sum_{1 \leq i < j \leq d} \left(\frac{\text{Cost}(T_{\mathbf{w}_i^{(0)}})}{|\mathbf{w}_i^{(0)}|} - \frac{\text{Cost}(T_{\mathbf{w}_j^{(0)}})}{|\mathbf{w}_j^{(0)}|} \right)^2 \\ &\stackrel{\text{Lemma 3.6}}{\leq} \underbrace{\gamma^{(0)}}_{\text{Lemma 3.6}} \left[\frac{\text{Cost}(T_{\mathbf{v}})}{|\mathbf{v}|} \right]^2 = \left(\sum_{J \in \mathcal{P}_0} \prod_{j \in J} \gamma^{(j)} \right) \left[\frac{\text{Cost}(T_{\mathbf{v}})}{|\mathbf{v}|} \right]^2. \end{aligned}$$

Suppose that the lemma is true for heights less than or equal to $h - 1$ and let us examine the estimator variance at height h . Then, by combining the induction hypothesis and Lemma 3.6, we have

$$\begin{aligned} \text{Var}(C_{\text{SE}}(T_{\mathbf{v}})) &\leq \frac{\left(\frac{|S(\mathbf{v})|}{|\mathbf{v}|}\right)^2}{d} \left(\sum_{1 \leq j \leq d} \left(\sum_{J \in \mathcal{P}_{h-2}} \prod_{j \in J} \gamma^{(j)} \right) \left[\frac{\text{Cost}(T_{\mathbf{w}_j^{(h-1)}})}{|\mathbf{w}_j^{(h-1)}|} \right]^2 \right) \\ &\quad + \gamma^{(h-1)} \left[\frac{\text{Cost}(T_{\mathbf{v}})}{|\mathbf{v}|} \right]^2 \\ &\leq \left(\frac{|S(\mathbf{v})|}{|\mathbf{v}|} \right)^2 \left(\sum_{J \in \mathcal{P}_{h-2}} \prod_{j \in J} \gamma^{(j)} \right) \left(\sum_{1 \leq j \leq d} \left[\frac{\text{Cost}(T_{\mathbf{w}_j^{(h-1)}})}{|\mathbf{w}_j^{(h-1)}|} \right]^2 \frac{1}{d} \right) \\ &\quad + \gamma^{(h-1)} \left[\frac{\text{Cost}(T_{\mathbf{v}})}{|\mathbf{v}|} \right]^2 \\ &\leq \left(\frac{|S(\mathbf{v})|}{|\mathbf{v}|} \right)^2 \left(\sum_{J \in \mathcal{P}_{h-2}} \prod_{j \in J} \gamma^{(j)} \right) (\gamma^{(h-1)} + 1) \left(\sum_{1 \leq j \leq d} \left[\frac{\text{Cost}(T_{\mathbf{w}_j^{(h-1)}})}{|\mathbf{w}_j^{(h-1)}|} \right]^2 \right) \\ &\quad + \gamma^{(h-1)} \left[\frac{\text{Cost}(T_{\mathbf{v}})}{|\mathbf{v}|} \right]^2 \\ &\leq \left(\frac{|S(\mathbf{v})|}{|\mathbf{v}|} \right)^2 \left(\sum_{J \in \mathcal{P}_{h-2}} \prod_{j \in J} \gamma^{(j)} \right) (\gamma^{(h-1)} + 1) \left(\frac{d \sum_{w \in S(\mathbf{v})} \text{Cost}(T_w)}{|S(\mathbf{v})|} \right)^2 \\ &\quad + \gamma^{(h-1)} \left[\frac{\text{Cost}(T_{\mathbf{v}})}{|\mathbf{v}|} \right]^2 \leq \left(\sum_{J \in \mathcal{P}_{h-2}} \prod_{j \in J} \gamma^{(j)} \right) (\gamma^{(h-1)} + 1) \left[\frac{\text{Cost}(T_{\mathbf{v}})}{|\mathbf{v}|} \right]^2 \end{aligned}$$

$$\begin{aligned}
& + \gamma^{(h-1)} \left[\frac{\text{Cost}(T_{\mathbf{v}})}{|\mathbf{v}|} \right]^2 \leq \left(\left(\sum_{J \in \mathcal{P}_{h-2}} \prod_{j \in J} \gamma^{(j)} \right) (\gamma^{(h-1)} + 1) + \gamma^{(h-1)} \right) \left[\frac{\text{Cost}(T_{\mathbf{v}})}{|\mathbf{v}|} \right]^2 \\
& \leq \left(\sum_{J \in \mathcal{P}_{h-1}} \prod_{j \in J} \gamma^{(j)} \right) \left[\frac{\text{Cost}(T_{\mathbf{v}})}{|\mathbf{v}|} \right]^2.
\end{aligned}$$

□

As an almost immediate consequence of Lemma 3.7 the following corollaries can be derived.

COROLLARY 3.8 (First upper bound for variance in (3.3)). *Let*

$$\gamma = \max_{0 \leq m \leq h-1} \left\{ \gamma^{(m)} \right\}.$$

Then, the upper bound on the variance is given by

$$\text{Var}(C_{\text{SE}}(T_{\mathbf{v}})) \leq ((\gamma + 1)^h - 1) \left[\frac{\text{Cost}(T_{\mathbf{v}})}{|\mathbf{v}|} \right]^2.$$

Proof. The proof is by straightforward application of the binomial theorem (BT). That is,

$$\sum_{J \in \mathcal{P}_{h-1}} \prod_{j \in J} \gamma^{(j)} \leq \sum_{J \in \mathcal{P}_{h-1}} \gamma^{|J|} = \sum_{0 \leq m \leq h-1} \binom{h-1}{m} \gamma^m - 1 \stackrel{\text{(BT)}}{=} (\gamma + 1)^h - 1.$$

□

Clearly, γ is generally not available, but if we know something about the subtree costs, the following bound can be derived similarly to [17].

COROLLARY 3.9 (Second upper bound for variance in (3.3)). *Let*

$$\gamma = \max_{0 \leq m \leq h-1} \left\{ \gamma^{(m)} \right\}$$

and suppose that γ is maximized at level m^* . Suppose without loss of generality that

$$H(S^{(m^*)}(\mathbf{v})) = \{\mathbf{w}_1, \dots, \mathbf{w}_d\}$$

and there exists constant a such that

$$\frac{\text{Cost}(T_{\mathbf{w}_1})}{|\mathbf{w}_1|} \leq \frac{\text{Cost}(T_{\mathbf{w}_2})}{|\mathbf{w}_2|} \leq \dots \leq \frac{\text{Cost}(T_{\mathbf{w}_d})}{|\mathbf{w}_d|} \leq a \frac{\text{Cost}(T_{\mathbf{w}_1})}{|\mathbf{w}_1|}.$$

Then, the variance of SE estimator satisfies

$$\text{Var}(C_{\text{SE}}(T_{\mathbf{v}})) \leq (\beta^h - 1) \left[\frac{\text{Cost}(T_{\mathbf{v}})}{|\mathbf{v}|} \right]^2,$$

where $\beta = \left(\frac{a^2 + 2a + 1}{4a} \right)$.

Proof. Before we begin, let us state a technical result from Lemma A.2: Let r_1, \dots, r_n be non-negative scalars and suppose without loss of generality that $r_1 \leq r_2 \leq \dots \leq r_n \leq a r_1$ for a positive scalar $a \geq 1$. Then,

$$(3.5) \quad \sum_{1 \leq j \leq n} \frac{r_j^2}{\frac{1}{n}} \leq \left(\frac{a^2 + 2a + 1}{4a} \right) \left(\sum_{1 \leq j \leq n} r_j \right)^2.$$

Using the definition of γ we continue with

$$\begin{aligned} \gamma &= \text{CV}^2 \left(\frac{\text{Cost}(T_{\mathbf{W}})}{|\mathbf{W}|} \right) = \frac{\text{Var} \left(\frac{\text{Cost}(T_{\mathbf{W}})}{|\mathbf{W}|} \right)}{\left(\mathbb{E} \left(\frac{\text{Cost}(T_{\mathbf{W}})}{|\mathbf{W}|} \right) \right)^2} = \frac{\mathbb{E} \left(\frac{\text{Cost}(T_{\mathbf{W}})}{|\mathbf{W}|} \right)^2}{\left(\mathbb{E} \left(\frac{\text{Cost}(T_{\mathbf{W}})}{|\mathbf{W}|} \right) \right)^2} - 1 \\ &= \frac{\sum_{1 \leq j \leq d} \left(\frac{\text{Cost}(T_{\mathbf{w}_j})}{|\mathbf{w}_j|} \right)^2 \frac{1}{d}}{\left(\frac{1}{d} \sum_{1 \leq j \leq d} \frac{\text{Cost}(T_{\mathbf{w}_j})}{|\mathbf{w}_j|} \right)^2} - 1 = \frac{\sum_{1 \leq j \leq d} \left(\frac{\text{Cost}(T_{\mathbf{w}_j})}{|\mathbf{w}_j|} \right)^2}{d^{-1} \left(\sum_{1 \leq j \leq d} \frac{\text{Cost}(T_{\mathbf{w}_j})}{|\mathbf{w}_j|} \right)^2} - 1 \stackrel{(3.5)}{\leq} \beta - 1. \end{aligned}$$

Combining this with Corollary 3.8, we complete the proof with

$$\text{Var}(C_{\text{SE}}(T_{\mathbf{v}})) \leq ((\gamma + 1)^h - 1) \left[\frac{\text{Cost}(T_{\mathbf{v}})}{|\mathbf{v}|} \right]^2 \stackrel{\gamma \leq \beta - 1}{\leq} (\beta^h - 1) \left[\frac{\text{Cost}(T_{\mathbf{v}})}{|\mathbf{v}|} \right]^2.$$

□

As the direct consequence of Theorem 3.1 and Corollaries 3.8 and 3.9 we can conclude that the CV of the SE algorithm, when applied to the tree T_{v_0} , is

$$\text{CV}_{\text{SE}} = \sqrt{\frac{\text{Var}(C_{\text{SE}})}{\mathbb{E}(C_{\text{SE}})^2}} \leq \sqrt{\frac{((\gamma + 1)^h - 1) [\text{Cost}(T_{v_0})]^2}{[\text{Cost}(T_{v_0})]^2}} = \sqrt{(\gamma + 1)^h - 1} \leq \sqrt{\beta^h - 1}.$$

Similarly to the bound in [17], it means that if β is not too large (and as a consequence, $\beta^h - 1$ has manageable size too), SE will be able to deliver a sufficiently acceptable estimator (similar to Knuth's Algorithm 2.1 [17]). Next, we consider a special case of a random tree counting problem; that is, a counting the number of vertices in a random tree.

4. Random trees. In this section we will work with families of random trees.

DEFINITION 4.1 (Family of random trees). *Consider a probability vector $\mathbf{p} = (p_0, \dots, p_k)$ that corresponds to the probability of a vertex to have $0, \dots, k$ successors respectively. Define a family of random trees $\mathcal{F}_{\mathbf{p}}^h$ as all possible trees of height at most h that are generated using \mathbf{p} up to the level h . (Note that if $p_0 > 0$, there exists a non-zero probability of extinction before generation h , and therefore the tree can have a height that is smaller than h .)*

The family $\mathcal{F}_{\mathbf{p}}^h$ is fully characterized by the probability vector \mathbf{p} and the parameter h . Moreover, the tree generation corresponds to a branching process [1]. Let $T = (\mathcal{V}, \mathcal{E})$ be a random tree from $\mathcal{F}_{\mathbf{p}}^h$. By assigning the cost $c(v) = 1$ for all $v \in \mathcal{V}$, the cost of the tree — $\text{Cost}(T)$ is equal to $|\mathcal{V}|$. Our objective is to analyse the behavior of Knuth's and SE's estimators under this setting. In particular, we show the following.

- In Lemma 4.3 and Corollary 4.4 we develop the expected variance and the lower bound of the Knuth's Algorithm 2.1.
- Using Lemmas 4.6, 4.7 and 4.8 we obtain an upper bound for the variance of the SE Algorithm 2.2. This result is summarized in Theorem 4.5.
- Combining the results of Corollary 4.4 and Theorem 4.5, we get the lower bound for the variance reduction introduced by the SE as compared to the Knuth's estimator. The latter is summarized in Corollary 4.9.

Consider now a random tree rooted at v_0 and let R_m be the total number of children (population size) at level (generation) m . Define

$$\mu = \mathbb{E}(R_1) = \sum_{0 \leq j \leq k} jp_j$$

and

$$\sigma^2 = \text{Var}(R_1) = \left(\sum_{0 \leq j \leq k} j^2 p_j \right) - \mu^2.$$

The tree counting problem — (that is, counting the overall number of vertices in the tree) becomes hard when the number of nodes in a tree is large, so we are naturally interested in the super-critical branching case ($\mu > 1$). Well-known results on the distribution of such trees are available from the general theory of the branching processes. Denote by M_m the total progeny at generation m , that is $M_n = 1 + R_1 + \dots + R_m$. Then, from [20] we have

$$\mathbb{E}(R_m) = \mu^m,$$

$$(4.1) \quad \nu_m = \mathbb{E}(M_m) = \mathbb{E} \left(1 + \sum_{1 \leq j \leq m} R_j \right) = \frac{1 - \mu^{m+1}}{1 - \mu},$$

$$\text{Var}(R_m) = \sigma^2 \frac{1 - \mu^m}{1 - \mu},$$

and

$$(4.2) \quad \zeta_m^2 = \text{Var}(M_m) = \frac{\sigma^2}{(1 - \mu)^2} \left[\frac{1 - \mu^{2m+1}}{1 - \mu} - (2m + 1)\mu^m \right].$$

Before proceeding to the performance analysis of the basic Knuth's and the SE's estimators, we will require the following technical lemma, the proof of which is given in appendix.

LEMMA 4.2 (Sum of squared differences). *Let $n \in \mathbb{N}$, and Z_1, \dots, Z_n be independent and identically distributed random variables with expectation and variance equal to μ and σ^2 respectively. Then,*

$$(4.3) \quad \mathbb{E} \left(\sum_{1 \leq i < j \leq n} (Z_i - Z_j)^2 \right) = n(n - 1)\sigma^2.$$

Also for $m \in \mathbb{N}$, let $Z_1^{(i)}, \dots, Z_m^{(i)}$, $i = 1, \dots, \binom{n}{m}$ be the unique subsets of $\{Z_1, \dots, Z_n\}$, and define $U_i = \frac{1}{m} \sum_{1 \leq j \leq m} Z_j^{(i)}$. Then,

$$(4.4) \quad \mathbb{E} \left(\sum_{1 \leq i < j \leq \binom{n}{m}} (U_i - U_j)^2 \right) \leq \binom{n}{m} \left(\binom{n}{m} - 1 \right) \frac{\sigma^2}{m}.$$

We next consider the expected variances of Knuth's and the SE's estimators under the random tree model. It is important to note that in this settings and for a random tree $T \in \mathcal{F}_{\mathbf{p}}^h$

$$\text{Var}(C(T) | T) \text{ and } \text{Var}(C_{\text{SE}}(T) | T)$$

are random variables.

LEMMA 4.3 (Knuth's expected variance). *Let $T^{(h)}$ be a random tree of height h . Then,*

$$(4.5) \quad \mathbb{E} \left(\text{Var} \left(C \left(T^{(h)} \right) \mid T^{(h)} \right) \right) = (\sigma^2 + \mu^2) \mathbb{E} \left(\text{Var} \left(C \left(T^{(h-1)} \right) \mid T^{(h-1)} \right) \right) + (\sigma^2 + \mu^2 - \mu) \zeta_{h-1}^2.$$

Proof. Suppose that the random tree $T^{(h)}$ is rooted at node v and let $S(v) = \{w_1, \dots, w_D\}$ be the set of its children. Recall that we are working under the random tree model and each vertex can be a parent of $0, \dots, k$ children with probability $\mathbf{p} = (p_0, \dots, p_k)$ respectively. From the variance formula (2.1) we have

$$(4.6) \quad \begin{aligned} \mathbb{E} \left(\text{Var} \left(C \left(T^{(h)} \right) \mid T^{(h)} \right) \right) &= \mathbb{E} \left(\text{Var} \left(C \left(T_v \right) \mid T_v \right) \right) = \\ &= \mathbb{E} \left(D \sum_{1 \leq j \leq D} \text{Var} \left(C \left(T_{w_j} \right) \mid T_{w_j} \right) \right. \\ &\quad \left. + \sum_{1 \leq i < j \leq D} \left(\text{Cost} \left(T_{w_i} \right) - \text{Cost} \left(T_{w_j} \right) \right)^2 \right) \\ &= \underbrace{\sum_{0 \leq d \leq k} \mathbb{E} \left(D^2 \frac{1}{D} \sum_{1 \leq j \leq D} \text{Var} \left(C \left(T_{w_j} \right) \mid T_{w_j} \right) \mid D = d \right) \mathbb{P}(D = d)}_{\text{Part (a)}} \\ &\quad + \underbrace{\sum_{0 \leq d \leq k} \mathbb{E} \left(\sum_{1 \leq i < j \leq D} \left(\text{Cost} \left(T_{w_i} \right) - \text{Cost} \left(T_{w_j} \right) \right)^2 \mid D = d \right) \mathbb{P}(D = d)}_{\text{Part (b)}}. \end{aligned}$$

Having in mind that D , the degree of v , is a random variable such that $\mathbb{E}(D) = \mu$ and $\text{Var}(D) = \sigma^2$, it follows that $\mathbb{E}(D^2) = \sigma^2 + \mu^2$, and by linearity of expectation we have the following.

- Part (a):

$$(4.7) \quad \begin{aligned} &\sum_{0 \leq d \leq k} \mathbb{E} \left(D^2 \frac{1}{D} \sum_{1 \leq j \leq D} \text{Var} \left(C \left(T_{w_j} \right) \mid T_{w_j} \right) \mid D = d \right) \mathbb{P}(D = d) \\ &\stackrel{(*)}{=} \sum_{0 \leq d \leq k} \mathbb{E} \left(D^2 \text{Var} \left(C \left(T_{w_1} \right) \mid T_{w_1} \right) \mid D = d \right) \mathbb{P}(D = d) \\ &= \mathbb{E} \left(D^2 \right) \mathbb{E} \left(\text{Var} \left(C \left(T_{w_1} \right) \mid T_{w_1} \right) \right) \\ &= (\sigma^2 + \mu^2) \mathbb{E} \left(\text{Var} \left(C \left(T^{(h-1)} \right) \mid T^{(h-1)} \right) \right), \end{aligned}$$

where “(*)” follows from the fact that w_1, \dots, w_d are roots of independent and identically distributed random trees, and as a consequence, the random variables

$$\text{Var} \left(C \left(T_{w_j} \right) \mid T_{w_j} \right), \quad j = 1, \dots, d$$

have the same distribution for any $d \in \mathbb{N}$, so,

$$\begin{aligned}
& \mathbb{E} \left(\frac{1}{D} \sum_{1 \leq j \leq D} \text{Var} (C (T_{w_j}) \mid T_{w_j}) \right) \\
&= \sum_{0 \leq d \leq k} \mathbb{E} \left(\frac{1}{D} \sum_{1 \leq j \leq D} \text{Var} (C (T_{w_j}) \mid T_{w_j}) \mid D = d \right) \mathbb{P}(D = d) \\
&= \sum_{0 \leq d \leq k} \mathbb{E} (\text{Var} (C (T_{w_1}) \mid T_{w_1}) \mid D = d) \mathbb{P}(D = d) = \mathbb{E} (\text{Var} (C (T_{w_1}) \mid T_{w_1})).
\end{aligned}$$

- For Part (b) we continue with

$$\begin{aligned}
(4.8) \quad & \sum_{0 \leq d \leq k} \mathbb{E} \left(\sum_{1 \leq i < j \leq D} (\text{Cost} (T_{w_i}) - \text{Cost} (T_{w_j}))^2 \mid D = d \right) \mathbb{P}(D = d) \\
& \underbrace{=}_{(**)} \sum_{0 \leq d \leq k} \mathbb{E} (D(D-1) \zeta_{h-1}^2 \mid D = d) \mathbb{P}(D = d) \\
& = \mathbb{E} (D(D-1)) \zeta_{h-1}^2 = (\sigma^2 + \mu^2 - \mu) \zeta_{h-1}^2,
\end{aligned}$$

where (**) follows from (4.2) and from the fact that $\text{Cost} (T_{w_1}), \dots, \text{Cost} (T_{w_d})$ are independent and identically distributed random variables with expectation and variance equal to ν_{h-1} and ζ_{h-1}^2 respectively. So, by Lemma 4.2 (4.3),

$$\mathbb{E} \left(\sum_{1 \leq i < j \leq d} (\text{Cost} (T_{w_i}) - \text{Cost} (T_{w_j}))^2 \right) = d(d-1) \zeta_{h-1}^2.$$

We complete the proof of the lemma by combining (4.6), (4.7), and (4.8). \square

COROLLARY 4.4 (Lower bound on Knuth's expected variance). *Under the settings of the random tree model,*

$$(4.9) \quad \mathbb{E} \left(\text{Var} \left(C \left(T^{(h)} \right) \mid T^{(h)} \right) \right) \geq (\sigma^2 + \mu^2 - \mu) \frac{1 - (\sigma^2 + \mu^2)^h}{1 - (\sigma^2 + \mu^2)}.$$

Proof.

$$\begin{aligned}
& \mathbb{E} \left(\text{Var} \left(C \left(T^{(h)} \right) \mid T^{(h)} \right) \right) \\
&= (\sigma^2 + \mu^2) \mathbb{E} \left(\text{Var} \left(C \left(T^{(h-1)} \right) \mid T^{(h-1)} \right) \right) + (\sigma^2 + \mu^2 - \mu) \zeta_{h-1}^2 \\
& \underbrace{\geq}_{\zeta_h^2 \geq 1} (\sigma^2 + \mu^2) \mathbb{E} \left(\text{Var} \left(C \left(T^{(h-1)} \right) \mid T^{(h-1)} \right) \right) + (\sigma^2 + \mu^2 - \mu) 1 \\
& \underbrace{=}_{(*)} (\sigma^2 + \mu^2 - \mu) \frac{1 - (\sigma^2 + \mu^2)^h}{1 - (\sigma^2 + \mu^2)},
\end{aligned}$$

where (*) follows from the fact that the recursion of type

$$F(n) = \alpha F(n-1) + \beta, \quad F(0) = 0,$$

has the solution of the form

$$F(n) = \beta \frac{1 - \alpha^n}{1 - \alpha}.$$

(See Lemma A.3). \square

Next, we proceed to the analysis of $\mathbb{E}(\text{Var}(C_{\text{SE}}(T^{(h)}) \mid T^{(h)}))$.

THEOREM 4.5 (Upper bound on the SE variance under the random tree model).
Let $T^{(h)}$ be a random tree of height h . Let

$$B' = \max \left\{ \left\lceil \frac{hk^2 \ln \left(2h(\sigma^2 + \mu^2) \frac{\sigma^2 \mu}{(\mu-1)^3} \right)}{2(\mu-1)^2} \right\rceil, \left\lceil \frac{h\sigma^2}{\mu^2} \right\rceil \right\},$$

and let m^* be the first level of $T^{(h)}$ for which the number of nodes is greater or equal to B' ; that is

$$m^* = \min_m \left\{ m : S^{(m)} \geq B' \right\}.$$

Then, for the budget B that satisfies

$$B = |S^{(m^*)}| \leq kB',$$

$$\mathbb{E} \left(\text{Var} \left(C_{\text{SE}} \left(T^{(h)} \right) \mid T^{(h)} \right) \right) \leq B^2 h e \mu^{2h} \left(\frac{\sigma^2 \mu}{(\mu-1)^3} + 1 \right).$$

Similar to Lemma 4.3, the idea is to upper-bound the variance formula (2.1). However, in the SE analysis, the cardinality of hyper nodes during the SE algorithm execution is playing an important role and must be taken into consideration. Our objective is to show that starting from some budget size, the hyper node cardinality will not vary much, which will cause a significant variance minimization. The proof is completed in few steps using Lemmas 4.6, 4.7 and 4.8 and Corollary 4.9.

LEMMA 4.6 (Random forest variance - Part 1). Consider a random forest $T_{\mathbf{v}}^{(h)}$ of height h where \mathbf{v} is the root hyper node. Let $S(\mathbf{v}) = \{\mathbf{w}_1, \dots, \mathbf{w}_D\}$ be the set of \mathbf{v} 's children. Then,

$$\begin{aligned} \mathbb{E} \left(\text{Var} \left(C_{\text{SE}} \left(T_{\mathbf{v}}^{(h)} \right) \mid T_{\mathbf{v}}^{(h)} \right) \right) &\leq \mathbb{E} \left(\left(\frac{|S(\mathbf{v})|}{|\mathbf{v}|} \right)^2 \text{Var} \left(C_{\text{SE}} \left(T_{\mathbf{w}_1}^{(h-1)} \right) \mid T_{\mathbf{w}_1}^{(h-1)} \right) \right) \\ &\quad + \mathbb{E} \left(\left(\frac{|S(\mathbf{v})|}{|\mathbf{v}|} \right)^2 \frac{\sigma^2 \mu}{(\mu-1)^3} \left(\frac{\sigma^2}{B} + \mu^2 \right)^{h-1} \right). \end{aligned}$$

REMARK 4.1. Note that $T_{\mathbf{w}_i}^{(h-1)}$, $i = 1, \dots, D$ are forests and that $\mathbf{w}_1, \dots, \mathbf{w}_D$ are just hyper nodes that contains roots of random trees so with out loss of generality we write \mathbf{w}_1 in the result of the above lemma.

Proof. Recall that we are working under the random tree model and each vertex $v \in \mathbf{v}$ can be a parent of $0, \dots, k$ children with probability $\mathbf{p} = (p_0, \dots, p_k)$ respectively. From the variance formula (3.3) we have

$$\begin{aligned} (4.10) \quad \mathbb{E} \left(\text{Var} \left(C_{\text{SE}} \left(T_{\mathbf{v}}^{(h)} \right) \mid T_{\mathbf{v}}^{(h)} \right) \right) &= \mathbb{E} \left(\text{Var} \left(C_{\text{SE}} \left(T_{\mathbf{v}} \right) \mid T_{\mathbf{v}} \right) \right) = \\ &= \underbrace{\mathbb{E} \left(\frac{\left(\frac{|S(\mathbf{v})|}{|\mathbf{v}|} \right)^2}{D} \sum_{1 \leq j \leq D} \text{Var} \left(C_{\text{SE}} \left(T_{\mathbf{w}_j} \right) \mid T_{\mathbf{w}_j} \right) \right)}_{\text{Part (a)}} \end{aligned}$$

$$+ \underbrace{\mathbb{E} \left(\frac{\left(\frac{|S(\mathbf{v})|}{|\mathbf{v}|} \right)^2}{D^2} \sum_{1 \leq i < j \leq D} \left(\frac{\text{Cost}(T_{\mathbf{w}_i})}{|\mathbf{w}_i|} - \frac{\text{Cost}(T_{\mathbf{w}_j})}{|\mathbf{w}_j|} \right)^2 \right)}_{\text{Part (b)}}.$$

- Part (a):

Note that $\mathbf{w}_1, \dots, \mathbf{w}_D$ are roots of identically distributed random forests of height $h - 1$ and consider the random variables

$$\text{Var}(C_{\text{SE}}(T_{\mathbf{w}_j}) \mid T_{\mathbf{w}_j}), \quad j = 1, \dots, D.$$

Despite the fact that those variables are clearly dependent (because \mathbf{w}_i and \mathbf{w}_j , $1 \leq i < j \leq D$ can share a subset of nodes), they still have the same expected value, so,

$$\begin{aligned} (4.11) \quad & \mathbb{E} \left(\left(\frac{|S(\mathbf{v})|}{|\mathbf{v}|} \right)^2 \frac{1}{D} \sum_{1 \leq j \leq D} \text{Var}(C_{\text{SE}}(T_{\mathbf{w}_j}) \mid T_{\mathbf{w}_j}) \right) \\ &= \sum_{1 \leq d \leq |\mathbf{v}|k} \mathbb{E} \left(\frac{\left(\frac{|S(\mathbf{v})|}{|\mathbf{v}|} \right)^2}{D} \sum_{1 \leq j \leq D} \text{Var}(C_{\text{SE}}(T_{\mathbf{w}_j}) \mid T_{\mathbf{w}_j}) \mid D \right) \mathbb{P}_d \\ &= \sum_{1 \leq d \leq |\mathbf{v}|k} \mathbb{E} \left(\left(\frac{|S(\mathbf{v})|}{|\mathbf{v}|} \right)^2 \text{Var}(C_{\text{SE}}(T_{\mathbf{w}_1}) \mid T_{\mathbf{w}_1}) \mid D \right) \mathbb{P}_d \\ &= \mathbb{E} \left(\left(\frac{|S(\mathbf{v})|}{|\mathbf{v}|} \right)^2 \text{Var}(C_{\text{SE}}(T_{\mathbf{w}_1}^{(h-1)}) \mid T_{\mathbf{w}_1}^{(h-1)}) \right), \end{aligned}$$

where $\mathbb{P}_d = \mathbb{P}(D = d)$.

- Part (b):

$$\begin{aligned} (4.12) \quad & \mathbb{E} \left(\frac{\left(\frac{|S(\mathbf{v})|}{|\mathbf{v}|} \right)^2}{D^2} \sum_{1 \leq i < j \leq D} \left(\frac{\text{Cost}(T_{\mathbf{w}_i})}{|\mathbf{w}_i|} - \frac{\text{Cost}(T_{\mathbf{w}_j})}{|\mathbf{w}_j|} \right)^2 \right) \\ &= \sum_{0 \leq d \leq |\mathbf{v}|k} \left[\mathbb{E} \left(\frac{\left(\frac{|S(\mathbf{v})|}{|\mathbf{v}|} \right)^2}{D^2} \sum_{1 \leq i < j \leq D} \left(\frac{\text{Cost}(T_{\mathbf{w}_i})}{|\mathbf{w}_i|} - \frac{\text{Cost}(T_{\mathbf{w}_j})}{|\mathbf{w}_j|} \right)^2 \mid D \right) \right] \mathbb{P}_d \\ &= \sum_{0 \leq d \leq |\mathbf{v}|k} \left[\frac{\left(\frac{|S(\mathbf{v})|}{|\mathbf{v}|} \right)^2}{D^2} \mathbb{E} \left(\sum_{1 \leq i < j \leq D} \left(\frac{\text{Cost}(T_{\mathbf{w}_i})}{|\mathbf{w}_i|} - \frac{\text{Cost}(T_{\mathbf{w}_j})}{|\mathbf{w}_j|} \right)^2 \mid D \right) \right] \mathbb{P}_d \\ &\stackrel{\text{Lemma 4.2 (4.4)}}{\leq} \sum_{0 \leq d \leq |\mathbf{v}|k} \left[\frac{\left(\frac{|S(\mathbf{v})|}{|\mathbf{v}|} \right)^2}{D^2} \mathbb{E} \left(D(D-1) \frac{\zeta_{h-1}^2}{|\mathbf{w}_1|} \mid D \right) \right] \mathbb{P}_d \\ &= \sum_{0 \leq d \leq |\mathbf{v}|k} \left[\mathbb{E} \left(\left(\frac{|S(\mathbf{v})|}{|\mathbf{v}|} \right)^2 \frac{D(D-1) \zeta_{h-1}^2}{D^2 |\mathbf{w}_1|} \mid D \right) \right] \mathbb{P}_d \\ &\stackrel{\text{w}_1| \geq 1}{\leq} \mathbb{E} \left(\left(\frac{|S(\mathbf{v})|}{|\mathbf{v}|} \right)^2 \frac{\zeta_{h-1}^2}{1} \right) \end{aligned}$$

$$\begin{aligned}
& \underbrace{\leq}_{(4.2)} \mathbb{E} \left(\left(\frac{|S(\mathbf{v})|}{|\mathbf{v}|} \right)^2 \right) \frac{\sigma^2}{(\mu-1)^2} \left(\frac{1-\mu^{2h-1}}{\mu-1} - (2h-1)\mu^{h-1} \right) \\
& \leq \mathbb{E} \left(\left(\frac{|S(\mathbf{v})|}{|\mathbf{v}|} \right)^2 \right) \frac{\sigma^2}{(\mu-1)^3} \mu^{2h-1} \leq \mathbb{E} \left(\left(\frac{|S(\mathbf{v})|}{|\mathbf{v}|} \right)^2 \right) \frac{\sigma^2 \mu}{(\mu-1)^3} (\mu^2)^{h-1} \\
& \underbrace{\leq}_{\frac{\sigma^2}{B} \geq 0} \mathbb{E} \left(\left(\frac{|S(\mathbf{v})|}{|\mathbf{v}|} \right)^2 \right) \frac{\sigma^2 \mu}{(\mu-1)^3} \left(\frac{\sigma^2}{B} + \mu^2 \right)^{h-1}.
\end{aligned}$$

We complete the proof of the lemma by combining (4.10), (4.11), and (4.12). \square

We are interested in having all the tree hyper nodes to retain cardinality B . This will allow us to control the value of the $(\sigma^2/B + \mu^2)$ term and as a consequence to introduce a significant variance reduction as will be shown in the following lemmas. In Lemma 4.7 we assume that the root has the desired cardinality and consider the following events:

$$A = \{|S(\mathbf{v})| \geq B\}, \quad \bar{A} = \{|S(\mathbf{v})| < B\}.$$

LEMMA 4.7 (Random forest variance - Part 2). *Consider a random forest $T_{\mathbf{v}}^{(h)}$ of height h where \mathbf{v} is the root hyper node such that $|\mathbf{v}| = B$. Then,*

$$\begin{aligned}
& \mathbb{E} \left(\text{Var} \left(C_{\text{SE}} \left(T_{\mathbf{v}}^{(h)} \right) \mid T_{\mathbf{v}}^{(h)} \right) \right) \leq \\
& \leq \mathbb{E} \left(\left(\frac{\sigma^2}{B} + \mu^2 \right) \text{Var} \left(C_{\text{SE}} \left(T_{\mathbf{w}_1}^{(h-1)} \right) \mid T_{\mathbf{w}_1}^{(h-1)} \right) + \frac{\sigma^2 \mu}{(\mu-1)^3} \left(\frac{\sigma^2}{B} + \mu^2 \right)^h \mid A \right) \mathbb{P}(A) \\
& + h (\sigma^2 + \mu^2)^h \frac{\sigma^2 \mu}{(\mu-1)^3} \mathbb{P}(\bar{A}).
\end{aligned}$$

Proof. It is not hard to prove by induction that recursive equations of the type

$$(4.13) \quad F(n) = \alpha F(n-1) + \beta \alpha^n + \gamma, \quad F(0) = 0, \alpha > 1, \gamma > 0,$$

are bounded by

$$(4.14) \quad F(n) \leq \alpha^n n(\beta + \gamma).$$

See Lemma A.3 for a proof. In addition, under the random tree model,

$$\mathbb{E} \left(\frac{|S(\mathbf{v})|}{|\mathbf{v}|} \right)^2 = \text{Var} \left(\frac{|S(\mathbf{v})|}{|\mathbf{v}|} \right) + \mu^2 = \frac{\sigma^2}{|\mathbf{v}|} + \mu^2.$$

See Lemma A.4 for the proof. Having in mind that by the law of total expectation

$$\mathbb{E} \left(\left(\frac{|S(\mathbf{v})|}{|\mathbf{v}|} \right)^2 \mid A \right) \mathbb{P}(A) + \mathbb{E} \left(\left(\frac{|S(\mathbf{v})|}{|\mathbf{v}|} \right)^2 \mid \bar{A} \right) \mathbb{P}(\bar{A}) = \frac{\sigma^2}{B} + \mu^2,$$

we can conclude that

$$\mathbb{E} \left(\left(\frac{|S(\mathbf{v})|}{|\mathbf{v}|} \right)^2 \mid A \right) \mathbb{P}(A) \leq \frac{\sigma^2}{B} + \mu^2, \quad \mathbb{E} \left(\left(\frac{|S(\mathbf{v})|}{|\mathbf{v}|} \right)^2 \mid \bar{A} \right) \mathbb{P}(\bar{A}) \leq \sigma^2 + \mu^2.$$

We complete the proof by applying the law of total expectation on the result of Lemma 4.6.

$$\begin{aligned}
& \mathbb{E} \left(\text{Var} \left(C_{\text{SE}} \left(T_{\mathbf{v}}^{(h)} \right) \mid T_{\mathbf{v}}^{(h)} \right) \right) \\
& \leq \mathbb{E} \left(\left(\frac{\sigma^2}{B} + \mu^2 \right) \text{Var} \left(C_{\text{SE}} \left(T_{\mathbf{w}_1}^{(h-1)} \right) \mid T_{\mathbf{w}_1}^{(h-1)} \right) + \frac{\sigma^2 \mu}{(\mu-1)^3} \left(\frac{\sigma^2}{B} + \mu^2 \right)^h \mid A \right) \mathbb{P}(A) \\
& + \mathbb{E} \left((\sigma^2 + \mu^2) \text{Var} \left(C_{\text{SE}} \left(T_{\mathbf{w}_1}^{(h-1)} \right) \mid T_{\mathbf{w}_1}^{(h-1)} \right) + \frac{\sigma^2 \mu}{(\mu-1)^3} (\sigma^2 + \mu^2)^h \mid \bar{A} \right) \mathbb{P}(\bar{A}) \\
& \leq \mathbb{E} \left(\left(\frac{\sigma^2}{B} + \mu^2 \right) \text{Var} \left(C_{\text{SE}} \left(T_{\mathbf{w}_1}^{(h-1)} \right) \mid T_{\mathbf{w}_1}^{(h-1)} \right) + \frac{\sigma^2 \mu}{(\mu-1)^3} \left(\frac{\sigma^2}{B} + \mu^2 \right)^h \mid A \right) \mathbb{P}(A) \\
& + h (\sigma^2 + \mu^2)^h \frac{\sigma^2 \mu}{(\mu-1)^3} \mathbb{P}(\bar{A}).
\end{aligned}$$

□

LEMMA 4.8 (Random forest variance - Part 3). *Consider a random forest $T_{\mathbf{v}}^{(h)}$ of height h where \mathbf{v} is the root hyper node such that $|\mathbf{v}| = B$. Then, for the budget B that satisfies*

$$B \geq \max \left\{ \left\lceil \frac{hk^2 \ln \left(2h(\sigma^2 + \mu^2) \frac{\sigma^2 \mu}{(\mu-1)^3} \right)}{2(\mu-1)^2} \right\rceil, \left\lceil \frac{h\sigma^2}{\mu^2} \right\rceil \right\},$$

it holds that

$$\mathbb{E} \left(\text{Var} \left(C_{\text{SE}} \left(T_{\mathbf{v}}^{(h)} \right) \mid T_{\mathbf{v}}^{(h)} \right) \right) \leq h e \mu^{2h} \left(\frac{\sigma^2 \mu}{(\mu-1)^3} + 1 \right).$$

Proof. Recall that $\mathbb{P}(\bar{A})$ is the probability that $S(|\mathbf{v}|) < B$. Hoeffdings inequality yields

$$\mathbb{P}(\bar{A}) \leq 2e^{-\frac{2|\mathbf{v}|(\mu-1)^2}{k^2}},$$

See Lemma A.5 for the proof. By requiring

$$(4.15) \quad h (\sigma^2 + \mu^2)^h \frac{\sigma^2 \mu}{(\mu-1)^3} \mathbb{P}(\bar{A}) \underbrace{\leq}_{|\mathbf{v}|=B} h (\sigma^2 + \mu^2)^h \frac{\sigma^2 \mu}{(\mu-1)^3} 2e^{-\frac{2B(\mu-1)^2}{k^2}} \leq 1,$$

we arrive at

$$B \geq \left\lceil \frac{hk^2 \ln \left(2h(\sigma^2 + \mu^2) \frac{\sigma^2 \mu}{(\mu-1)^3} \right)}{2(\mu-1)^2} \right\rceil.$$

With this budget and from Lemma 4.7 we have

$$\begin{aligned}
& \mathbb{E} \left(\text{Var} \left(C_{\text{SE}} \left(T_{\mathbf{v}}^{(h)} \right) \mid T_{\mathbf{v}}^{(h)} \right) \right) \leq \\
& \leq \mathbb{E} \left(\left(\frac{\sigma^2}{B} + \mu^2 \right) \text{Var} \left(C_{\text{SE}} \left(T_{\mathbf{w}_1}^{(h-1)} \right) \mid T_{\mathbf{w}_1}^{(h-1)} \right) + \frac{\sigma^2 \mu}{(\mu-1)^3} \left(\frac{\sigma^2}{B} + \mu^2 \right)^h \mid A \right) \mathbb{P}(A) \\
& + \underbrace{h (\sigma^2 + \mu^2)^h \frac{\sigma^2 \mu}{(\mu-1)^3} \mathbb{P}(\bar{A})}_{\leq 1 \text{ by (4.15)}} \leq
\end{aligned}$$

$$\underbrace{\leq}_{\mathbb{P}(A) \leq 1} \mathbb{E} \left(\left(\frac{\sigma^2}{B} + \mu^2 \right) \text{Var} \left(C_{\text{SE}} \left(T_{\mathbf{w}}^{(h-1)} \right) \mid T_{\mathbf{w}}^{(h-1)} \right) + \frac{\sigma^2 \mu}{(\mu-1)^3} \left(\frac{\sigma^2}{B} + \mu^2 \right)^h \right) + 1,$$

where $|\mathbf{w}| = B$. We complete the proof by applying the recursive bound (4.13) and (4.14) on the above equation and arrive at

$$\begin{aligned} \mathbb{E} \left(\text{Var} \left(C_{\text{SE}} \left(T_{\mathbf{v}}^{(h)} \right) \mid T_{\mathbf{v}}^{(h)} \right) \right) &\leq h \left(\frac{\sigma^2}{B} + \mu^2 \right)^h \left(\frac{\sigma^2 \mu}{(\mu-1)^3} + 1 \right) \\ &\underbrace{\leq}_{B \geq \lceil \frac{h\sigma^2}{\mu^2} \rceil} h \mu^{2h} \left(\frac{1}{h} + 1 \right)^h \left(\frac{\sigma^2 \mu}{(\mu-1)^3} + 1 \right) \leq h e \mu^{2h} \left(\frac{\sigma^2 \mu}{(\mu-1)^3} + 1 \right). \end{aligned}$$

□

We are ready to complete the proof of Theorem 4.5.

Proof. The proof is almost an immediate consequence of Lemmas 3.5 and 4.8. Let \mathbf{v} be the hyper node at level m and note that by definition of the Lemma, $|\mathbf{v}| = |S^{(m^*)}|^2 = B^2$. Then, the variance of SE Algorithm 2.2 satisfies:

$$\begin{aligned} \mathbb{E} \left(\text{Var} \left(C_{\text{SE}} \left(T^{(h)} \right) \mid T^{(h)} \right) \right) &\underbrace{\leq}_{\text{Lemma 3.5}} |\mathbf{v}|^2 \mathbb{E} \left(\text{Var} \left(C_{\text{SE}} \left(T_{\mathbf{v}}^{(h-m^*)} \right) \mid T_{\mathbf{v}}^{(h-m^*)} \right) \right) \\ &\underbrace{\leq}_{\text{Lemma 4.8}} B^2 h e \mu^{2h} \left(\frac{\sigma^2 \mu}{(\mu-1)^3} + 1 \right). \end{aligned}$$

Having in mind that each vertex has at most k successors and combining this with the fact that $|S^{(m^*-1)}| < B'$ we conclude that $|S^{(m^*)}| \leq kB'$, thus completing the proof of Theorem 4.5. □

The following corollary reveals the true strength of the SE Algorithm 2.2.

COROLLARY 4.9 (Average variance of SE and Knuth). *For the problem of counting the number of nodes in a random tree $T \in \mathcal{F}_{\mathbf{p}}^h$ and provided that the budget B is chosen according to the conditions of Theorem 4.5, the SE Algorithm 2.2 introduces an expected variance reduction that is greater or equal to*

$$\frac{1}{\text{P}(h; B, k, \sigma, \mu)} \left(\left(1 + \frac{\sigma^2}{\mu^2} \right)^h - \frac{1}{\mu^{2h}} \right),$$

where $\text{P}(h; B, \sigma, \mu)$ is a polynomial function of h, B, k, σ and μ .

REMARK 4.2. Note that if $(1 + \sigma^2/\mu^2)^h$ is an exponentially fast growing function, the variance reduction becomes significant. This happens for example when μ and σ^2 do not depend on the parameter h .

Proof. Combining Corollary 4.4 and Theorem 4.5 results in

$$\begin{aligned} \frac{\mathbb{E} \left(\text{Var} \left(C \left(T^{(h)} \right) \mid T^{(h)} \right) \right)}{\mathbb{E} \left(\text{Var} \left(C_{\text{SE}} \left(T^{(h)} \right) \mid T^{(h)} \right) \right)} &\geq \frac{(\sigma^2 + \mu^2 - \mu) \frac{1 - (\sigma^2 + \mu^2)^h}{1 - (\sigma^2 + \mu^2)}}{B^2 h e \mu^{2h} \left(\frac{\sigma^2 \mu}{(\mu-1)^3} + 1 \right)} \\ &= \frac{1}{\underbrace{\frac{(\sigma^2 + \mu^2 - \mu)/(1 - \sigma^2 + \mu^2)}{B^2 h e \left(\frac{\sigma^2 \mu}{(\mu-1)^3} + 1 \right)}}_{\text{P}(h; B, k, \sigma, \mu)}} \cdot \frac{(\sigma^2 + \mu^2)^h - 1}{\mu^{2h}} = \frac{1}{\text{P}(h; B, k, \sigma, \mu)} \left[\frac{\left(\mu^2 \left(1 + \frac{\sigma^2}{\mu^2} \right) \right)^h}{\mu^{2h}} - \frac{1}{\mu^{2h}} \right] \\ &= \frac{1}{\text{P}(h; B, k, \sigma, \mu)} \left(\left(1 + \frac{\sigma^2}{\mu^2} \right)^h - \frac{1}{\mu^{2h}} \right). \end{aligned}$$

□

4.1. FPRAS for special families of trees. Following [23], we first formalize the notation of “computationally efficient approximation algorithm”. A *randomized approximation scheme* for $\text{Cost}(T)$ is a non-deterministic algorithm which, when given an input tree T and a real number $\varepsilon \in (0, 1)$, outputs a random variable \mathcal{K} such that

$$\mathbb{P}((1 - \varepsilon)\text{Cost}(T) \leq \mathcal{K} \leq (1 + \varepsilon)\text{Cost}(T)) \geq \frac{3}{4}.$$

It was shown in [15] that the success probability $3/4$ can be boosted to any $1 - \delta$ by running the algorithm $O(\log(\delta^{-1}))$ times and taking the median of the obtained results. Such a scheme is said to be fully polynomial if its execution time is bounded by some polynomial in the tree height (that is assumed to be polynomial in our case) and ε^{-1} . If this condition holds, such algorithm is a *fully polynomial randomized approximation scheme* or FPRAS. A FPRAS is considered a very good algorithm, especially when a polynomial time exact algorithm does not exist (unless $P = NP$), so, FPRAS for such problems is essentially the best approximation one can hope to achieve [14]. We will construct a FPRAS from the unbiased SE estimator \mathcal{K} for $\text{Cost}(T)$ — the number of vertices in T .

Those settings allow us to run a number of independent copies of the algorithm on the same input, and output the average of the results. Actually, for any random variable U with expectation μ' and variance σ'^2 , Chebyshev’s inequality yield

$$\mathbb{P}(|U - \mu'| > r\sigma') \leq \frac{1}{r^2}.$$

Combining this with the independence of repetitive calculations of the random variable \mathcal{K} and provided that

$$\text{Cost}(T) = \mathbb{E}(\mathcal{K}) = \mu' \text{ and } \text{Var}(\mathcal{K}) = \sigma'^2,$$

one can show that for a random variable $\bar{\mathcal{K}}$ defined by

$$\bar{\mathcal{K}} = \frac{1}{N} \sum_{i=1}^N \mathcal{K}_i,$$

we have

$$\mathbb{P}(|\bar{\mathcal{K}} - \mu'| > \varepsilon\mu') \leq \frac{\text{Var}(\bar{\mathcal{K}})}{\varepsilon^2\mu'^2} = \frac{\frac{\sigma'^2}{N}}{\varepsilon^2\mu'^2}.$$

So, if we demand the final expression to be less than $1/4$, one can immediately conclude that the use of N samples such that

$$\frac{\frac{\sigma'^2}{N}}{\varepsilon^2\mu'^2} \leq \frac{1}{4} \Rightarrow N \geq \frac{4}{\varepsilon^2} \frac{\sigma'^2}{\mu'^2}$$

will satisfy

$$\mathbb{P}(|\bar{\mathcal{K}} - \mu'| > \varepsilon\mu') \leq \frac{1}{4} \Rightarrow \mathbb{P}(|\bar{\mathcal{K}} - \mu'| \leq \varepsilon\mu') \geq \frac{3}{4}.$$

Moreover, from this, it readily follows that the critical factor of the algorithm is the variance of the unbiased estimator divided by the square of its expectation — the CV^2 . We conclude the above discussion with an important observation. If the coefficient of variation is bounded above by a polynomial in the size of the tree height, then we have constructed a FPRAS.

As it was already discussed in Section 1, we probably cannot hope to develop FPRAS for the general tree counting problem, because the latter is considered to be a very hard task. Still, we show that the SE algorithm is efficient for counting the number of vertices in random trees. More specifically, following the work of Rasmussen [23], we show that for *almost all* trees that were generated under the random tree model and provided that the number of vertices is not much smaller than their expected number (that is given by equation (4.1)), one can construct FPRAS.

THEOREM 4.10 (Almost sure FPRAS). *Let $\mathcal{F}_{\mathbf{p}'}^h$ be a family of random trees such that for $T \in \mathcal{F}_{\mathbf{p}'}^h$,*

$$(4.16) \quad \lim_{h \rightarrow \infty} \mathbb{P} \left(\text{Cost}(T) < \frac{1}{P(h)} \nu_h \right) = 0,$$

where $P(h) > 0$ is some polynomial function in h and $\nu_h = \frac{1-\mu^{h+1}}{1-\mu}$ is the expected number of nodes. In other words, for most instances, (almost surely), the actual number of nodes is not much smaller than the expectation. Then, under the above condition, and provided that

$$(4.17) \quad \mu > 1 + \varepsilon \quad \text{for any } \varepsilon > 0, \quad (\text{see Remark 4.3}),$$

the SE algorithm is FPRAS for most of the instances in $T \in \mathcal{F}_{\mathbf{p}'}^h$, that is,

$$\text{CV}^2 = \frac{\text{Var}(C_{\text{SE}}(T) \mid T)}{(\mathbb{E}(C_{\text{SE}}(T) \mid T))^2}$$

is bounded by a polynomial in h with high probability.

REMARK 4.3. The condition (4.17) is technical in sense that we demand μ to be not very close to 1.

Proof. Let $T \in \mathcal{F}_{\mathbf{p}'}^h$ be a random tree that was generated by the corresponding branching process. Recall that both $\text{Var}(C_{\text{SE}}(T) \mid T)$ and $\mathbb{E}(C_{\text{SE}}(T) \mid T)$ are random variable, so, by the Markov inequality,

$$\mathbb{P}(\text{Var}(C_{\text{SE}}(T) \mid T) > a) \leq \frac{\mathbb{E}(\text{Var}(C_{\text{SE}}(T) \mid T))}{a}.$$

By setting

$$a = h \mathbb{E}(\text{Var}(C_{\text{SE}}(T) \mid T)),$$

we get

$$(4.18) \quad \mathbb{P}(\text{Var}(C_{\text{SE}}(T) \mid T) > h \mathbb{E}(\text{Var}(C_{\text{SE}}(T) \mid T))) \leq \frac{1}{h} \rightarrow 0 \quad \text{as } h \rightarrow \infty.$$

Note also that by Theorem 3.1, for any tree T ,

$$\mathbb{E}(C_{\text{SE}}(T) \mid T) = \text{Cost}(T),$$

so, under the condition (4.16),

$$(4.19) \quad \mathbb{P} \left(\mathbb{E}(C_{\text{SE}}(T) \mid T) = \text{Cost}(T) < \frac{1}{P(h)} \frac{1-\mu^{h+1}}{1-\mu} \right) \rightarrow 0 \quad \text{as } h \rightarrow \infty.$$

And now, by combining (4.18) and (4.19), we conclude that for $T \in \mathcal{F}_{\mathbf{p}'}^h$, and as $h \rightarrow \infty$,

$$\mathbb{P} \left(\frac{\text{Var}(C_{\text{SE}}(T) | T)}{(\mathbb{E}(C_{\text{SE}}(T) | T))^2} = \frac{\text{Var}(C_{\text{SE}}(T) | T)}{(\text{Cost}(T))^2} \leq \frac{h\mathbb{E}(\text{Var}(C_{\text{SE}}(T) | T))}{\left(\frac{1}{P(h)}\mathbb{E}(\text{Cost}(T))\right)^2} \right. \\ \left. \stackrel{\text{Corollary 4.9}}{\leq} h(P(h))^2 \frac{B^2 h e (\sigma^2 \mu + (\mu - 1)^3)}{\mu^2 (\mu - 1)} \right) \rightarrow 1 \quad \text{as } h \rightarrow \infty.$$

We complete the proof by noting that under (4.17),

$$h(P(h))^2 \frac{B^2 h e (\sigma^2 \mu + (\mu - 1)^3)}{\mu^2 (\mu - 1)} \stackrel{\text{Remark 4.3}}{\leq} h\varepsilon(P(h))^2 \frac{B^2 h e (\sigma^2 \mu + (\mu - 1)^3)}{\mu^2}$$

is bounded by a polynomial, while Theorem 4.5 provides the upper bound on B^2 ; that is

$$B^2 \leq [k(B')]^2 \leq \left[k \left(\max \left\{ \left\lceil \frac{hk^2 \ln \left(2h(\sigma^2 + \mu^2) \frac{\sigma^2 \mu}{(\mu - 1)^3} \right)}{2(\mu - 1)^2} \right\rceil, \left\lceil \frac{h\sigma^2}{\mu^2} \right\rceil \right\} \right) \right]^2.$$

As a consequence,

$$\frac{\text{Var}(C_{\text{SE}}(T) | T)}{(\mathbb{E}(C_{\text{SE}}(T)))^2}$$

is bounded by a polynomial in h for most of random tree instances. \square

We complete this section by showing a *specific* family of trees $T \in \mathcal{F}_{\mathbf{p}^*}^h$ for which SE is FPRAS for counting the number of nodes in $\mathcal{F}_{\mathbf{p}^*}^h$.

Let $\mathcal{F}_{\mathbf{p}^*}^h$ be a family of random trees defined by a fixed height $h \in \mathbb{N}$ and probability vector

$$\mathbf{p}^* = (p_0, \dots, p_{h[\log(h)]-1}, p_{h[\log(h)]}, \dots, p_{h[\log(h)]+h}),$$

where

$$p_i = 0 \text{ for } i = 0, \dots, h[\log(h)] - 1 \text{ and } p_i = \frac{1}{h} \text{ for } i = h[\log(h)], \dots, h[\log(h)] + h.$$

In other words, each vertex can have

$$a = h[\log(h)], \dots, h[\log(h)] + h = b$$

successors with equal probability $1/h$. Suppose that such trees are generated up to height h . Note that this is always possible since $p_0 = 0$. Under the above settings, we have a discrete uniform distribution between a and b over h values. Such a distribution satisfies

$$(4.20) \quad \mu = \frac{a+b}{2} = \frac{2h[\log(h)] + h}{2} \text{ and } \sigma^2 = \frac{h^2 - 1}{12},$$

and as a consequence,

$$(4.21) \quad \frac{\sigma^2}{\mu^2} \rightarrow \frac{(h^2 - 1)/12}{(2h[\log(h)] + h)^2/2^2} \rightarrow \frac{1 - 1/h^2}{3 \left(4[\log^2(h)] + \frac{4[\log(h)]}{h} + 1 \right)} \rightarrow \\ \rightarrow \frac{1}{12[\log^2(h)]}, \quad h \rightarrow \infty.$$

THEOREM 4.11 (FPRAS). *For the random tree family $\mathcal{F}_{\mathbf{p}^*}^h$ the following holds.*

1. *Knuth's estimator average variance is growing at an exponential rate compared to the SE average variance.*
2. *For most of the trees in $\mathcal{F}_{\mathbf{p}^*}^h$, SE is FPRAS.*

Proof.

1. Follows immediately from Corollary 4.9 and the fact that

$$\begin{aligned} \left(1 + \frac{\sigma^2}{\mu^2}\right)^h &\rightarrow \left(1 + \frac{1}{12\lceil\log^2(h)\rceil}\right)^h \underset{\log(1+x) \geq \frac{2x}{2+x}}{\geq} e^{h \frac{\frac{2}{12\lceil\log^2(h)\rceil}}{2 + \frac{2}{12\lceil\log^2(h)\rceil}}} \\ &= e^{\frac{2h}{12\lceil\log^2(h)\rceil} \frac{12\lceil\log^2(h)\rceil}{24\lceil\log^2(h)\rceil+2}} = e^{\frac{h}{12\lceil\log^2(h)\rceil+1}}, \quad h \rightarrow \infty. \end{aligned}$$

2. To complete the second part of this theorem, we will show that

$$\mathbb{P}\left(\text{Cost}(T) < \frac{1}{\mathbb{P}(h)} \frac{1 - \mu^{h+1}}{1 - \mu}\right) \rightarrow 0, \quad h \rightarrow \infty.$$

thus satisfying the condition of Theorem 4.10. From (4.1) and (4.2), we have

$$\lim_{h \rightarrow \infty} \frac{\zeta_h^2}{\nu_h^2} = \lim_{h \rightarrow \infty} \frac{\frac{\sigma^2}{(1-\mu)^2} \left[\frac{1-\mu^{2h+1}}{1-\mu} - (2h+1)\mu^h \right]}{\left(\frac{1-\mu^{h+1}}{1-\mu} \right)^2} = \frac{\sigma^2}{\mu^2} \underset{(4.21)}{=} \frac{1}{12\lceil\log^2(h)\rceil}.$$

And now, by choosing

$$\mathbb{P}(h) = \frac{\sqrt{\lceil\log(h)\rceil} - 1}{\sqrt{\lceil\log(h)\rceil}},$$

we arrive at

$$\begin{aligned} &\lim_{h \rightarrow \infty} \mathbb{P}\left(\text{Cost}(T) < \frac{1}{\mathbb{P}(h)} \frac{1 - \mu^{h+1}}{1 - \mu}\right) \\ &= \lim_{h \rightarrow \infty} \mathbb{P}\left(\text{Cost}(T) < \left(1 - \frac{1}{\sqrt{\lceil\log(h)\rceil}}\right) \nu_h\right) \\ &\leq \underbrace{\frac{\zeta_h^2}{\zeta_h^2 + \frac{\nu_h^2}{\lceil\log(h)\rceil}}}_{(*)} = \frac{\zeta_h^2/\nu_h^2}{\zeta_h^2/\nu_h^2 + \frac{1}{\lceil\log(h)\rceil}} \\ &= \frac{\frac{1}{12\lceil\log^2(h)\rceil}}{\frac{1}{12\lceil\log^2(h)\rceil} + \frac{1}{\lceil\log(h)\rceil}} = \frac{1}{1 + 12\lceil\log(h)\rceil} = 0, \end{aligned}$$

where (*) follows from the one-sided Chebyshev's bound

$$\mathbb{P}(X < \nu_h - a) \leq \frac{\zeta_h^2}{\zeta_h^2 + a^2}.$$

We complete the proof by noting that (4.17) holds, because

$$\begin{aligned} \lim_{h \rightarrow \infty} 1 + \frac{1}{\mathbb{P}(h)} &= \lim_{h \rightarrow \infty} \frac{2\lceil\log(h)\rceil - 1}{\lceil\log(h)\rceil - 1} = 2 \\ &< \lim_{h \rightarrow \infty} \mu = \lim_{h \rightarrow \infty} \frac{2h\lceil\log(h)\rceil + h}{2} = h\lceil\log(h)\rceil + h/2. \end{aligned}$$

□

5. Numerical Results. In this section we present some numerical results obtained by the SE Algorithm 2.2. The results were obtained using an Intel Core i7 machine with 16 GB of RAM. The tables should be interpreted as follows.

- Run: independent runs of the algorithm.
- \widehat{C} , \widehat{C}_{SE} : estimators returned by Knuth's Algorithm 2.1 and SE Algorithm 2.2 respectively.
- RE: estimated relative error of the unbiased estimator (3.4).

Random Trees: As a motivating example for further investigation of the SE method, we consider a random tree that was generated in the following manner. Starting from the tree root and for each node recursively, we connect 0, 1, 2 or 3 children with probabilities 0.3, 0.4, 0.1 and 0.2 respectively. In the following examples, we consider trees of specific heights. Hence, we stop the tree generation at the predetermined height.

Model 1: The generated tree is of height 60 and the number of nodes is 1976527. Tables 1 and 2 present the comparison between the Knuth's estimator and the SE algorithm. Note that if SE uses $B = 1$ we get Knuth's algorithm. We introduce the parameter R — the number of independent replications of unbiased estimator — and deliberately set the simulation effort to be the same ($B \cdot R$) for both algorithms.

TABLE 1
Performance of 10 independent runs of Knuth's Algorithm 2.1 for a randomly generated tree (Model 1) with $B = 1$ and $R = 20000$.

Run	\widehat{C}	\widehat{RE}
1	3.06×10^3	6.09×10^{-1}
2	1.44×10^4	9.52×10^{-1}
3	1.05×10^3	2.40×10^{-1}
4	7.08×10^3	7.93×10^{-1}
5	3.01×10^3	6.08×10^{-1}
6	4.36×10^4	8.41×10^{-1}
7	3.26×10^3	5.36×10^{-1}
8	3.01×10^3	4.14×10^{-1}
9	1.51×10^3	2.65×10^{-1}
10	1.06×10^3	3.06×10^{-1}
Average	8.10×10^3	5.56×10^{-1}

TABLE 2
Performance of 10 independent runs of SE Algorithm 2.2 for a randomly generated tree (Model 1) with $B = 20$ and $R = 1000$.

Run	\widehat{C}_{SE}	\widehat{RE}
1	2.04×10^6	5.03×10^{-2}
2	1.83×10^6	5.56×10^{-2}
3	1.99×10^6	7.18×10^{-2}
4	2.02×10^6	5.80×10^{-2}
5	1.90×10^6	5.97×10^{-2}
6	1.95×10^6	5.70×10^{-2}
7	2.03×10^6	6.38×10^{-2}
8	1.83×10^6	5.25×10^{-2}
9	2.14×10^6	6.88×10^{-2}
10	1.97×10^6	5.97×10^{-2}
Average	1.97×10^6	5.97×10^{-2}

Model 2: We consider another model with $\mathbf{p} = (0.5, 0.1, 0.2, 0.2, 0.1)$ the tree was generated until height $h = 30$. The number of vertices was 551. Tables 3 and 4 summarize the results.

Model 3: Finally, we consider a model with $\mathbf{p} = (0.0, 0.7, 0.2, 0.1)$, where the tree was generated until height $h = 30$. The number of vertices was 25723. Tables 5 and 6 summarize the results.

It is interesting to note that for the first two models, Knuth's estimator performs poorly but this is not the case for the last model. The key to understand this phenomenon lies in the analysis of the variance of corresponding child distributions. Recall the result that was obtained in Corollary 4.9. The variance reduction of SE is governed by the $(1 + \sigma^2/\mu^2)^h$ term that clearly governs the exponential growth.

TABLE 3

Performance of 10 independent runs of Knuth's Algorithm 2.1 for a randomly generated tree (Model 2) with $B = 1$ and $R = 20000$.

Run	\widehat{C}	\widehat{RE}
1	1.38×10^2	1.71×10^{-1}
2	1.53×10^2	4.05×10^{-1}
3	1.68×10^2	4.63×10^{-1}
4	1.85×10^2	4.28×10^{-1}
5	1.15×10^2	1.11×10^{-1}
6	1.89×10^2	2.90×10^{-1}
7	1.25×10^2	1.16×10^{-1}
8	2.06×10^2	4.84×10^{-1}
9	2.06×10^2	4.25×10^{-1}
10	2.38×10^2	8.09×10^{-1}
Average	1.72×10^2	3.70×10^{-1}

TABLE 5

Performance of 10 independent runs of Knuth's Algorithm 2.1 for a randomly generated tree (Model 3) with $B = 1$ and $R = 20000$.

Run	\widehat{C}	\widehat{RE}
1	2.52×10^4	2.98×10^{-2}
2	2.60×10^4	3.50×10^{-2}
3	2.73×10^4	3.82×10^{-2}
4	2.63×10^4	3.31×10^{-2}
5	2.55×10^4	3.05×10^{-2}
6	2.56×10^4	3.13×10^{-2}
7	2.63×10^4	3.87×10^{-2}
8	2.51×10^4	3.17×10^{-2}
9	2.55×10^4	3.67×10^{-2}
10	2.50×10^4	3.40×10^{-2}
Average	2.58×10^4	3.39×10^{-2}

TABLE 4

Performance of 10 independent runs of SE Algorithm 2.2 for a randomly generated tree (Model 2) with $B = 20$ and $R = 1000$.

Run	\widehat{C}_{SE}	\widehat{RE}
1	5.51×10^2	5.97×10^{-3}
2	5.52×10^2	6.24×10^{-3}
3	5.45×10^2	6.18×10^{-3}
4	5.47×10^2	6.00×10^{-3}
5	5.49×10^2	6.31×10^{-3}
6	5.51×10^2	6.35×10^{-3}
7	5.46×10^2	6.08×10^{-3}
8	5.53×10^2	5.97×10^{-3}
9	5.56×10^2	6.01×10^{-3}
10	5.58×10^2	6.05×10^{-3}
Average	5.51×10^2	6.12×10^{-3}

TABLE 6

Performance of 10 independent runs of SE Algorithm 2.2 for a randomly generated tree (Model 3) with $B = 20$ and $R = 1000$.

Run	\widehat{C}_{SE}	\widehat{RE}
1	2.55×10^4	1.05×10^{-2}
2	2.58×10^4	1.07×10^{-2}
3	2.59×10^4	1.10×10^{-2}
4	2.55×10^4	1.11×10^{-2}
5	2.59×10^4	1.08×10^{-2}
6	2.56×10^4	1.08×10^{-2}
7	2.60×10^4	1.08×10^{-2}
8	2.61×10^4	1.08×10^{-2}
9	2.54×10^4	1.09×10^{-2}
10	2.59×10^4	1.04×10^{-2}
Average	2.57×10^4	1.08×10^{-2}

By examining the parameters of the above models, we get a desired intuition for different performance rates.

- For Model 1 we have

$$\begin{aligned} \mathbf{p} = (0.3, 0.4, 0.1, 0.2) &\Rightarrow \mu = 1.2, \sigma^2 = 2.6 \\ &\Rightarrow \left(1 + \frac{2.6}{1.2^2}\right)^{60} \approx 7.61 \times 10^{26}. \end{aligned}$$

- For Model 2 we have

$$\begin{aligned} \mathbf{p} = (0.5, 0.1, 0.2, 0.2, 0.1) &\Rightarrow \mu = 1.5, \sigma^2 = 2.05 \\ &\Rightarrow \left(1 + \frac{2.05}{1.5^2}\right)^{30} \approx 2.75 \times 10^8. \end{aligned}$$

- For Model 3 we have

$$\begin{aligned} \mathbf{p} = (0.0, 0.7, 0.2, 0.1) &\Rightarrow \mu = 1.4, \sigma^2 = 0.44 \\ &\Rightarrow \left(1 + \frac{0.44}{1.4^2}\right)^{30} \approx 435. \end{aligned}$$

Non-random Model: Numerous experiments we performed with the SE algorithm indicate that it can introduce a good performance for different problems and models, not necessarily random ones. Unfortunately, it is generally not easy to develop rigorous theoretical guarantees similar to those introduced in Section 3. But the numerical performance can readily be examined. As an additional example, consider the following, very structured tree.

The root has 3 children. The leftmost child becomes the root of full binary tree and the rest of the children will continue the root behavior recursively. Let $T_{\text{bft}}(h)$ be the number of vertices in the big fat tree (BFT) of height h . Then,

$$T_{\text{bft}}(h) = \underbrace{1}_{\text{root}} + \underbrace{2^{h-1} - 1}_{\text{left successor full binary tree}} + \underbrace{2T_{\text{bft}}(h-1)}_{\text{rightmost and middle successors}}.$$

Solving the recursion gives

$$(5.1) \quad T_{\text{bft}}(h) = (h + 1)2^h \text{ for } h \geq 0.$$

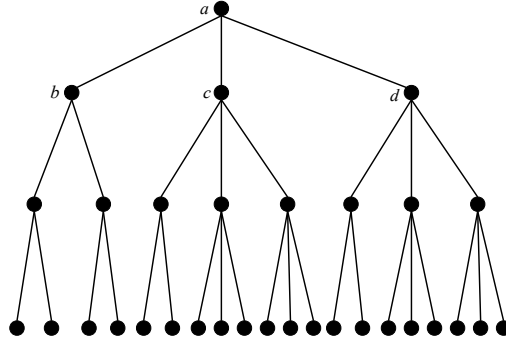


FIG. 5. BFT tree of height 3 with 32 nodes.

Let $C(T_v)$ be Knuth’s estimator for the BFT tree of height h rooted at v and let

$$S(v) = \{w_1, w_2, w_3\},$$

where w_1 is a root of a full binary tree of height $h - 1$ and w_2, w_3 are the roots of BFT trees of height $h - 1$ respectively. Then, having in mind that for a full k -ary tree Knuth’s estimator has zero variance [17], and, combining with (2.1), we have

$$\begin{aligned} (5.2) \quad \text{Var}(C(T_v)) &= \\ &= 3 \sum_{1 \leq j \leq 3} \text{Var}(C(T_{w_j})) + \sum_{1 \leq i < j \leq 3} (\text{Cost}(T_{w_i}) - \text{Cost}(T_{w_j}))^2 \\ &= 3(\text{Var}(C(T_{w_1})) + 2\text{Var}(C(T_{w_2}))) + \sum_{1 \leq i < j \leq 3} (\text{Cost}(T_{w_i}) - \text{Cost}(T_{w_j}))^2 \end{aligned}$$

$$\begin{aligned}
&= 6\text{Var}(C(T_{w_2})) + 2(T_{\text{bft}}(h-1) - \text{Cost}(T_{w_1}))^2 \\
&= 6\text{Var}(C(T_{w_2})) + 2(h2^{h-1} - (2^h - 1))^2.
\end{aligned}$$

Using the equation above, we show by induction that

$$(5.3) \quad \text{Var}(C(T_v)) \geq 1.4^{h-1} (2^{h-2})^2.$$

For $h = 2$,

$$\text{Var}(C(T_v)) = 0 + 2(2-3)^2 = 2 \geq 1.4^{2-1} (2^{2-2})^2.$$

Suppose that (5.3) holds for heights $1, \dots, h-1$ and consider the variance at level h .

$$\begin{aligned}
\text{Var}(C(T_v)) &= 6\text{Var}(C(T_{w_2})) + 2(h2^{h-1} - (2^h - 1))^2 \\
&\geq 6 \left(1.4^{h-2} (2^{h-3})^2\right) + 2(h2^{h-1} - (2^h - 1))^2 \\
&\geq \frac{1.5}{1.4} (1.4 \cdot 2 \cdot 2) \left(1.4^{h-2} (2^{h-3})^2\right) \\
&\geq \frac{1.5}{1.4} \left(1.4 \cdot 1.4^{h-2} 4 \cdot (2^{h-3})^2\right) \geq 1.4^{h-1} (2^{h-2})^2.
\end{aligned}$$

Combining now (5.1) and (5.3), we can derive a lower bound on Knuth's estimator's coefficient of variation:

$$\begin{aligned}
\text{CV}^2 &= \frac{\text{Var}(C(T_v))}{(\mathbb{E}(C(T_v)))^2} \geq \frac{1.4^{h-1} (2^{h-2})^2}{((h+1)2^h)^2} \\
&\geq \frac{1.4^{h-1}}{(h+1)^2} \frac{(2^{h-2})^2}{16(2^{h-2})^2} \geq \frac{1.4^{h-1}}{16(h+1)^2}.
\end{aligned}$$

Clearly, this bound implies that the efficiency of Knuth's estimator deteriorates at an exponential rating for BFT trees.

In Figure 6, we plotted the numerically obtained coefficients of variations for the BFT trees of different heights. In the left panel, the numerical CV of the Knuth's estimator is plotted against the exact CV value, which was calculated using the recurrence formula (5.2). In the right panel, we plotted the numerical coefficient of variation of SE. Unfortunately, we do not possess a good upper bound for the CV of SE. Still, we can clearly observe the superiority of the SE algorithm. The CV of SE Algorithm 2.2 seems to grow linearly while the CV of Knuth's Algorithm 2.1 has a clear exponential rate of growth. Moreover, while we could easily estimate the CV of SE Algorithm 2.2 on the trees of height 100, we could not do the same for Algorithm 2.1 because of the large relative error. As a consequence, we only report the CV until $h = 30$ for the Knuth's method.

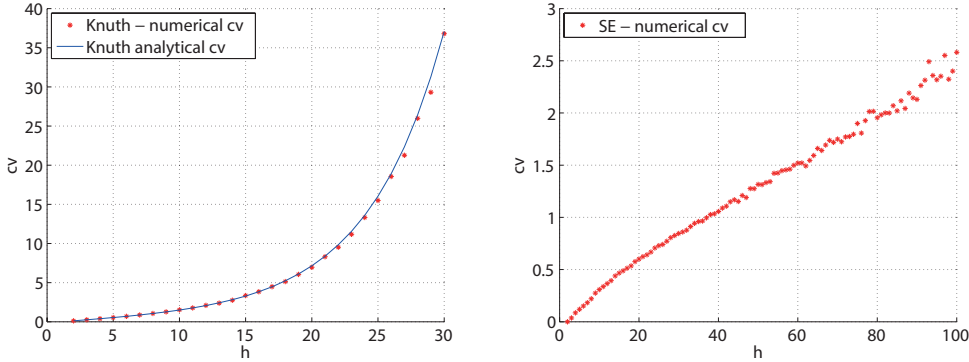


FIG. 6. The performance of Knuth’s Algorithm 2.1 and the SE Algorithm 2.2 on counting BFT trees of different heights. Left panel: Knuth. Right panel: SE.

6. Concluding remarks. We introduced an adaptation of the SE algorithm for counting backtrack trees. We showed that this estimator is unbiased, developed the upper bound for the coefficient of variation and established that for almost all random trees the SE algorithm is a FPRAS for estimation of the overall number of vertices. The proposed algorithm is easy to implement and the numerical results introduce an excellent practical performance. Of interest for future research is to theoretically investigate specific families of trees that can be efficiently enumerated using the proposed SE algorithm. Additionally, it will be interesting to rigorously study further adaptations of SE to different hard counting problems.

Appendix.

LEMMA A.1 (Sum of k -subsets). Let $A = \{r_1, \dots, r_n\}$ be a set of scalars and define the set J to be

$$J = \{R \mid R \subseteq A, |R| = u\}, \quad u = 1, \dots, n.$$

Then,

$$\sum_{R \in J} \frac{\sum_{r \in R} r}{u} = \frac{\binom{n}{u} \sum_{1 \leq j \leq n} r_j}{n}.$$

Proof. Note that there are exactly $\binom{n-1}{u-1}$ subsets in which each r_j ($j = 1, \dots, n$) appears. We conclude the proof with

$$\begin{aligned} \sum_{R \in J} \frac{\sum_{r \in R} r}{u} &= \frac{\binom{n-1}{u-1}}{u} \sum_{1 \leq j \leq n} r_j = \frac{\binom{n-1}{u-1}}{u} \sum_{1 \leq j \leq n} r_j \\ &= \frac{\frac{n!}{u!(n-u)!}}{n} \sum_{1 \leq j \leq n} r_j = \frac{\binom{n}{u} \sum_{1 \leq j \leq n} r_j}{n}. \end{aligned}$$

□

LEMMA A.2 (Sums of squares bound - from [17]). Let r_1, \dots, r_n be non-negative scalars and suppose without loss of generality that $r_1 \leq r_2 \leq \dots \leq r_n \leq a r_1$ for a positive scalar $a \geq 1$. Then,

$$\sum_{1 \leq j \leq n} \frac{r_j^2}{\frac{1}{n}} \leq \left(\frac{a^2 + 2a + 1}{4a} \right) \left(\sum_{1 \leq j \leq n} r_j \right)^2.$$

Proof. The proof is an immediate consequence from the following basic calculus inequality [21] (pages 71,253 – 254). Let l, L, b, B be positive numbers such that

$$0 < l \leq l_1 \leq l_2 \leq \dots \leq l_n \leq L \text{ and } 0 < b \leq b_1 \leq b_2 \leq \dots \leq b_n \leq B.$$

Then,

$$\frac{\left(\sum_{1 \leq j \leq n} l_j^2\right) \left(\sum_{1 \leq j \leq n} b_j^2\right)}{\left(\sum_{1 \leq j \leq n} l_j b_j\right)^2} \leq \frac{1}{4} \left(\sqrt{\frac{LB}{lb}} + \sqrt{\frac{lb}{LB}} \right)$$

holds. We continue the development by assigning

$$l_i = 1 \text{ and } b_i = r_i \quad \forall i = 1, \dots, n.$$

Having in mind that $r_n \leq ar_1$, we get

$$l = L = 1 \text{ and } b = r_1, B = r_n \leq ar_1,$$

and the lemma follows immediately from

$$\begin{aligned} \sum_{1 \leq j \leq n} \frac{r_j^2}{n} &= n \sum_{1 \leq j \leq n} r_j^2 \leq \frac{1}{4} \left(\sqrt{\frac{LB}{lb}} + \sqrt{\frac{lb}{LB}} \right) \left(\sum_{1 \leq j \leq n} r_j \right)^2 \\ &\leq \frac{1}{4} \left(\sqrt{\frac{ar_1}{r_1}} + \sqrt{\frac{r_1}{ar_1}} \right) \left(\sum_{1 \leq j \leq n} r_j \right)^2 \leq \\ &\leq \frac{1}{4} \left(\sqrt{a} + \sqrt{\frac{1}{a}} \right) \left(\sum_{1 \leq j \leq n} r_j \right)^2 \leq \frac{1}{4} \frac{(a+1)^2}{a} \left(\sum_{1 \leq j \leq n} r_j \right)^2. \end{aligned}$$

□

Proof. [of Lemma 4.2]

$$\begin{aligned} \mathbb{E} \left(\sum_{1 \leq i < j \leq n} (Z_i - Z_j)^2 \right) &= \mathbb{E} \left((n-1) \sum_{1 \leq j \leq n} Z_j^2 - 2 \sum_{1 \leq i < j \leq n} Z_i Z_j \right) \\ &= (n-1) \sum_{1 \leq j \leq n} \mathbb{E}(Z_j^2) - 2 \sum_{1 \leq i < j \leq n} \mathbb{E}(Z_i Z_j) \\ &\stackrel{Z_i, Z_j \text{ i.i.d.}}{=} (n-1) \sum_{1 \leq j \leq n} \mathbb{E}(Z_j^2) - 2 \sum_{1 \leq i < j \leq n} \mathbb{E}(Z_i) \mathbb{E}(Z_j) \\ &= n(n-1) (\mathbb{E}(Z_1^2)) - 2 \frac{n(n-1)}{2} (\mathbb{E}(Z_1))^2 \\ &= n(n-1) (\mathbb{E}(Z_1^2)) - (\mathbb{E}(Z_1))^2 = n(n-1) \sigma^2. \end{aligned}$$

Consider now

$$U_a = \frac{1}{m} \sum_{1 \leq j \leq m} Z_j^{(a)} \text{ and } U_b = \frac{1}{m} \sum_{1 \leq j \leq m} Z_j^{(b)},$$

and suppose that

$$\left| \{Z_1^{(a)}, \dots, Z_m^{(a)}\} \cap \{Z_1^{(b)}, \dots, Z_m^{(b)}\} \right| = k \quad (k = 0, \dots, m)$$

where k is the number of the same Z variables. Suppose also without loss of generality that

$$\{Z_1^{(a)}, \dots, Z_m^{(a)}\} \cap \{Z_1^{(b)}, \dots, Z_m^{(b)}\} = \{Z_1^{(a)}, \dots, Z_k^{(a)}\} = \{Z_1^{(b)}, \dots, Z_k^{(b)}\}.$$

Then,

$$\begin{aligned} \mathbb{E}(U_a - U_b)^2 &= \mathbb{E}\left(\frac{1}{m} \sum_{1 \leq j \leq m} Z_j^{(a)} - \frac{1}{m} \sum_{1 \leq j \leq m} Z_j^{(b)}\right)^2 \\ &= \mathbb{E}\left(\frac{1}{m} \sum_{k+1 \leq j \leq m} Z_j^{(a)} - \frac{1}{m} \sum_{k+1 \leq j \leq m} Z_j^{(b)}\right)^2. \end{aligned}$$

Having in mind that in this case,

$$U'_a = \frac{1}{m} \sum_{k+1 \leq j \leq m} Z_j^{(a)} \text{ and } U'_b = \frac{1}{m} \sum_{k+1 \leq j \leq m} Z_j^{(b)}$$

are independent random variables that satisfies

$$\mathbb{E}(U'_a) = \frac{m-k}{m} \mathbb{E}(U_a), \quad \mathbb{E}(U'_b) = \frac{m-k}{m} \mathbb{E}(U_b), \quad \mathbb{E}(U'_a U'_b) = \mathbb{E}(U'_a) \mathbb{E}(U'_b),$$

and

$$\text{Var}(U'_a) = \text{Var}(U'_b) = \text{Var}\left(\frac{1}{m} \sum_{k+1 \leq j \leq m} Z_j\right) = \frac{1}{m^2} (m-k) \sigma^2,$$

we can write,

$$\begin{aligned} \mathbb{E}(U_a - U_b)^2 &= \mathbb{E}(U'_a - U'_b)^2 = \mathbb{E}(U'^2_a - 2U'_a U'_b + U'^2_b) = \\ &= 2\left(\mathbb{E}(U'^2_a) - (\mathbb{E}(U'_a))^2\right) = 2\text{Var}(U'_a) = 2\frac{1}{m^2} (m-k) \sigma^2 \leq 2\frac{\sigma^2}{m}. \end{aligned}$$

By the previous equation combined with the linearity of expectation, we complete the proof with

$$\begin{aligned} \mathbb{E}\left(\sum_{1 \leq i < j \leq \binom{n}{m}} (U_i - U_j)^2\right) &\leq \binom{\binom{n}{m}}{2} \mathbb{E}(U_a - U_b)^2 \\ &\leq \frac{\binom{n}{m} \left(\binom{n}{m} - 1\right)}{2} \mathbb{E}(U_a - U_b)^2 \\ &\leq \binom{n}{m} \left(\binom{n}{m} - 1\right) \frac{\sigma^2}{m}. \end{aligned}$$

□

LEMMA A.3 (Recursions).

1. The recursive formulas of the type

$$F(n) = \alpha F(n-1) + \beta, \quad F(0) = 0,$$

has the solution of the form

$$F(n) = \beta \frac{1 - \alpha^n}{1 - \alpha}.$$

2. The recursive formulas of the type

$$F(n) = \alpha F(n-1) + \beta \alpha^n + \gamma, \quad F(0) = 0, \alpha > 1, \gamma > 0,$$

is bounded by

$$F(n) \leq \alpha^n n(\beta + \gamma).$$

Proof. For both equations, the proof is by induction on n .

1. For

$$F(n) = \alpha F(n-1) + \beta, \quad F(0) = 0.$$

- For $n = 1$,

$$F(1) = 0 + \beta = \beta \frac{1 - \alpha^1}{1 - \alpha}.$$

- Suppose that the proposition holds for $n - 1$, so,

$$F(n) = \alpha \beta \frac{1 - \alpha^{n-1}}{1 - \alpha} + \beta = \frac{\beta \alpha - \beta \alpha^n + \beta - \beta \alpha}{1 - \alpha} = \beta \frac{1 - \alpha^n}{1 - \alpha}.$$

2. For,

$$F(n) = \alpha F(n-1) + \beta \alpha^n + \gamma, \quad F(0) = 0, \alpha > 1.$$

- For $n = 1$,

$$F(1) = \alpha \cdot 0 + \beta \alpha + \gamma \underset{\alpha > 1}{\leq} 1 \cdot \alpha(\beta + \gamma) = \beta \alpha + \gamma \alpha.$$

- Suppose that the proposition holds for $n - 1$, so,

$$\begin{aligned} F(n) &\leq \alpha \left((n-1) \alpha^{n-1} (\beta + \gamma) \right) + \beta \alpha^n + \gamma \leq \\ &\leq \underbrace{\alpha}_{\alpha > 1} \left((n-1) \alpha^{n-1} (\beta + \gamma) \right) + \beta \alpha^n + \alpha^n \gamma \\ &\leq \alpha^n (n-1) (\beta + \gamma) + \beta \alpha^n + \gamma \alpha^n = n \alpha^n (\beta + \gamma). \end{aligned}$$

□

LEMMA A.4 (Hyper node child distribution). *Consider a random tree model where each node have $0, \dots, k$ successors with the corresponding probabilities $\mathbf{p} = (p_0, \dots, p_k)$ such that*

$$\mu = \sum_{0 \leq j \leq k} j p_j, \quad \sigma^2 = \sum_{0 \leq j \leq k} j^2 p_j - \mu^2,$$

and consider a hyper node \mathbf{v} . Then,

$$\mathbb{E} \left(\frac{|S(\mathbf{v})|}{|\mathbf{v}|} \right)^2 = \frac{\sigma^2}{|\mathbf{v}|} + \mu^2.$$

Proof. Having a hyper node $\mathbf{v} = \{v_1, \dots, v_{|\mathbf{v}|}\}$ in hand, the number of children of each $v_j \in \mathbf{v}$ is distributed with $\mathbf{p} = (p_0, \dots, p_k)$. Under the above settings,

$$|S(\mathbf{v})| = \left| \sum_{1 \leq j \leq |\mathbf{v}|} S(v_j) \right|,$$

where $S(v_j)$ are independent and identically distributed random variables such that $0 \leq S(v_j) \leq k$, $\mathbb{E}(S(v_j)) = \mu$ and $\text{Var}(S(v_j)) = \sigma^2$. Note that

$$\text{Var} \left(\frac{|S(\mathbf{v})|}{|\mathbf{v}|} \right) = \text{Var} \left(\frac{\sum_{1 \leq j \leq |\mathbf{v}|} |S(v_j)|}{|\mathbf{v}|} \right) = \frac{\sigma^2}{|\mathbf{v}|},$$

and as a consequence,

$$\mathbb{E} \left(\frac{|S(\mathbf{v})|}{|\mathbf{v}|} \right)^2 = \text{Var} \left(\frac{|S(\mathbf{v})|}{|\mathbf{v}|} \right) + \mu^2 = \frac{\sigma^2}{|\mathbf{v}|} + \mu^2.$$

□

LEMMA A.5 (Hyper node child concentration). *Consider a random tree model where each node have $0, \dots, k$ successors with the corresponding probabilities $\mathbf{p} = (p_0, \dots, p_k)$ such that*

$$\mu = \sum_{0 \leq j \leq k} j p_j,$$

and consider a hyper node \mathbf{v} . Then,

$$\mathbb{P}(|S(\mathbf{v})| < |\mathbf{v}|) \leq 2e^{-\frac{2|\mathbf{v}|(\mu-1)^2}{k^2}}.$$

Proof. By Hoeffding's inequality [12], let

$$Y_1, \dots, Y_n$$

be independent random variables. If Y_j are almost surely bounded; that is,

$$\mathbb{P}(Y_j \in [a_j, b_j]) = 1,$$

then

$$(A.1) \quad \mathbb{P} \left(\left| \sum_{0 \leq j \leq n} Y_j - \mathbb{E} \left(\sum_{0 \leq j \leq n} Y_j \right) \right| \geq t \right) \leq 2e^{-\frac{2t^2}{\sum_{0 \leq j \leq n} (b_j - a_j)^2}}.$$

With a hyper node $\mathbf{v} = \{v_1, \dots, v_{|\mathbf{v}|}\}$, the number of successors of each $v_j \in \mathbf{v}$ is distributed with $\mathbf{p} = (p_0, \dots, p_k)$ for $0, \dots, k$ respectively. Under the above settings,

$$|S(\mathbf{v})| = \sum_{1 \leq j \leq |\mathbf{v}|} |S(v_j)|,$$

where $S(v_j)$ are independent and identically distributed random variables such that $0 \leq S(v_j) \leq k$ and $\mathbb{E}(S(v_j)) = \mu$. Combining this with (A.1) for $t = \mu|\mathbf{v}| - |\mathbf{v}|$ we immediately arrive to

$$\begin{aligned} \mathbb{P}(|S(\mathbf{v})| < |\mathbf{v}|) &\leq \mathbb{P}(|S(\mathbf{v})| - \mu|\mathbf{v}| \geq \mu|\mathbf{v}| - |\mathbf{v}|) \leq 2e^{-\frac{2(\mu|\mathbf{v}| - |\mathbf{v}|)^2}{\sum_{0 \leq j \leq |\mathbf{v}|} (k-0)^2}} \leq \\ &\leq 2e^{-\frac{2|\mathbf{v}|^2(\mu-1)^2}{|\mathbf{v}|k^2}} = 2e^{-\frac{2|\mathbf{v}|(\mu-1)^2}{k^2}}. \end{aligned}$$

□

REFERENCES

- [1] SØREN ASMUSSEN AND HEINRICH HERING, *Branching processes*, Progress in probability and statistics, Birkhäuser, Switzerland, 1983.
- [2] DAVID AVIS AND LUC DEVROYE, *Estimating the number of vertices of a polyhedron.*, Inf. Process. Lett., 73 (2000), pp. 137–143.
- [3] JOSE BLANCHET AND DANIEL RUDOY, *Rare event simulation and counting problems*, in Rare Event Simulation Using Monte Carlo Methods, Wiley, 2009.
- [4] JOSEPH K. BLITZSTEIN AND PERSI DIACONIS, *A sequential importance sampling algorithm for generating random graphs with prescribed degrees*, Internet Mathematics, 6 (2011), pp. 489–522.
- [5] ZDRAVKO I. BOTEV AND DIRK P. KROESE, *Efficient monte carlo simulation via the generalized splitting method*, Statistics and Computing, 22 (2012), pp. 1–16.
- [6] PANG C. CHEN, *Heuristic sampling: A method for predicting the performance of tree searching programs*, SIAM J. Comput., 21 (1992), pp. 295–315.
- [7] YUGUO CHEN, PERSI DIACONIS, SUSAN P. HOLMES, AND JUN S. LIU, *Sequential Monte Carlo methods for statistical analysis of tables*, Journal of the American Statistical Association, 100 (2005), pp. 109–120.
- [8] MARTIN DYER, *Approximate counting by dynamic programming*, in Proceedings of the 35th ACM Symposium on Theory of Computing, 2003, pp. 693–699.
- [9] MARTIN DYER, ALAN FRIEZE, AND MARK JERRUM, *On counting independent sets in sparse graphs*, in In 40th Annual Symposium on Foundations of Computer Science, 1999, pp. 210–217.
- [10] MARNIX J. J. GARVELS, *The splitting method in rare event simulation*, PhD thesis, Enschede, October 2000.
- [11] PAUL GLASSERMAN, PHILIP HEIDELBERGER, PERWEZ SHAHABUDDIN, AND TIM ZAJIC, *Splitting for rare event simulation: Analysis of simple cases.*, in Winter Simulation Conference, 1996, pp. 302–308.
- [12] WASSILY Hoeffding, *Probability inequalities for sums of bounded random variables*, Journal of the American Statistical Association, (1962).
- [13] MARK JERRUM AND ALISTAIR SINCLAIR, *The Markov chain Monte Carlo method: An approach to approximate counting and integration*, in Approximation Algorithms for NP-hard Problems, D. Hochbaum, ed., PWS Publishing, 1996, pp. 482–520.
- [14] MARK JERRUM, ALISTAIR SINCLAIR, AND ERIC VIGODA, *A polynomial-time approximation algorithm for the permanent of a matrix with non-negative entries*, Journal of the ACM, (2004), pp. 671–697.
- [15] MARK JERRUM, LESLIE G. VALIANT, AND VIJAY V. VAZIRANI, *Random generation of combinatorial structures from a uniform distribution*, Theor. Comput. Sci., 43 (1986), pp. 169–188.
- [16] RICHARD M. KARP AND MICHAEL LUBY, *Monte-Carlo algorithms for enumeration and reliability problems*, in Proceedings of the 24th Annual Symposium on Foundations of Computer Science, SFCS '83, Washington, DC, USA, 1983, IEEE Computer Society, pp. 56–64.
- [17] DONALD E. KNUTH, *Estimating the efficiency of backtrack programs*, Math. Comp., 29 (1975).
- [18] JINGCHENG LIU AND PINYAN LU, *FPTAS for counting monotone CNF*, CoRR, abs/1311.3728 (2013).
- [19] RAJEEV MOTWANI AND PRABHAKAR RAGHAVAN, *Randomized Algorithms*, Cambridge University Press, New York, NY, USA, 1995.
- [20] ANTHONY G. PAKES, *Some limit theorems for the total progeny of a branching process*, Advances in Applied Probability, 3 (1971), pp. 176–192.
- [21] GEORGE POLYA AND GABOR SZEGÖ, *Problems and Theorems in Analysis I*, Classics in Mathematics, Springer Berlin, 1998.
- [22] PAUL W. PURDOM, *Tree size by partial backtracking.*, SIAM J. Comput., 7 (1978), pp. 481–491.
- [23] LARS E. RASMUSSEN, *Approximately counting cliques*, Random Struct. Algorithms, 11 (1997), pp. 395–411.
- [24] REUVEN Y. RUBINSTEIN, *The Gibbs cloner for combinatorial optimization, counting and sampling*, Methodology and Computing in Applied Probability, 11 (2009), pp. 491–549.
- [25] ———, *Stochastic enumeration method for counting NP-hard problems*, Methodology and Computing in Applied Probability, (2012), pp. 1–42.
- [26] REUVEN Y. RUBINSTEIN, ANDREY DOLGIN, AND RADISLAV VAISMAN, *The splitting method for decision making*, Communications in Statistics - Simulation and Computation, 41 (2012), pp. 905–921.
- [27] REUVEN Y. RUBINSTEIN AND DIRK P. KROESE, *Simulation and the Monte Carlo Method, Second Edition*, John Wiley and Sons, New York, 2008.
- [28] REUVEN Y. RUBINSTEIN, AD RIDDER, AND RADISLAV VAISMAN, *Fast Sequential Monte Carlo Methods for Counting and Optimization*, John Wiley and Sons, New York, 2013.

- [29] SALIL P. VADHAN, *The complexity of counting in sparse, regular, and planar graphs*, SIAM Journal on Computing, 31 (1997), pp. 398–427.
- [30] LESLIE G. VALIANT, *The complexity of enumeration and reliability problems*, SIAM J. Comput., 8 (1979), pp. 410–421.
- [31] YULI ZHANG, HUAIYU WU, AND LEI CHENG, *Some new deformation formulas about variance and covariance*, Proceedings of 4th International Conference on Modelling, Identification and Control, (2012).