Algorithm 1 SearchTreeSampler(\(F, k, M\))

**Input:** Formula \(F\) with \(n\) vars. Parameters \(k, M = 2^\ell\)

**Output:** A set \(S\) of solutions of \(F\)

if \(F\) is not satisfiable then
    return \(\emptyset\)
else
    Let \(\Phi_0 = \{\top\}\) // True, empty variable assignment
    Let \(L = \lceil \frac{2}{\ell} \rceil\) // number of levels
    for \(i = 1, \cdots, L\) do
        \(\Phi_i \leftarrow \text{BlackBoxSampler}(\Phi_{i-1}, k, \ell)\)
    end for
    return \(\Phi_L\)
end if

\[
\begin{align*}
\text{Algorithm 2 BlackBoxSampler}(\Phi, k, \ell) & \\
\text{Input:} & \text{A set } \Phi \text{ of uniformly sampled pseudosolutions of level } i \text{. Parameters } k, \ell \\
\text{Output:} & \text{A set } S \text{ of pseudosolutions of level } i + \ell \text{ (approximately uniformly sampled) with } 2^\ell |\Phi| \geq |S| \geq |\Phi| \\
& \text{for } j = 1, \cdots, \min\{k, |\Phi|\} \text{ do} \\
& \hspace{1cm} \text{Sample } s_j \text{ from } \Phi \text{ without replacement} \\
& \hspace{1cm} \text{Generate } D(s_j), \text{ the set of all pseudosolutions of level } i + \ell \text{ with } s_j \text{ as ancestor (using a complete SAT solver)} \\
& \hspace{1cm} S = S \cup D(s_j) \\
& \text{end for} \\
\end{align*}
\]

*Proof.* Suppose \(k \leq |\Phi|\) (otherwise, it means that \(\Phi\) contains all pseudosolutions of level \(i\), hence by definition \(S = S_{i+\ell}\) and therefore \(P(s) = P(s')\)), and that \(\Phi \subseteq S_i\) is a set of uniformly sampled pseudosolutions of level \(i\).

We can think of each pseudosolution \(s \in \Phi\) as an urn, that contains a certain number \(1 \leq |D(s)| \leq M = 2^\ell\) of pseudosolutions at a lower level \(i + \ell\) (its descendants). Let \(|S_i| = N \geq |\Phi|\) be the total number of pseudosolutions of level \(i\).

Since \(\Phi\) contains uniform samples, \(s_1, \cdots, s_k\) in Algorithm 2 are also uniform samples of pseudosolutions of level \(i\). Let \(S\) as in Algorithm 2 (the union of the contents of the \(k\) urns selected). Let \(s \in S_{i+\ell}\) be a pseudosolution of level \(i + \ell\), and let \(a(s) \in S_i\) be its unique ancestor at level \(i\). Clearly, the probability that \(s \in S\) is \(P(s \in S) = \frac{\binom{N-1}{k-1}}{\binom{N}{k}}\)

that is equal to the probability of selecting the ancestor \(a(s)\) of \(s\) on level \(i\). Let \(e\) be a randomly selected element of \(S\). Ideally, we would like the probability \(P(s) \equiv P[e = s]\) to be a constant independent of \(s\) (uniform sampling). However, intuitively this is not exactly constant because there is a bias towards elements such that \(|D(a(s))|\) is small.

Specifically,

\[
P(s) = \frac{1}{\left(\binom{N}{k}\right)} \left[ \sum_{t=1}^{k} \frac{1}{|D(a(s)) \cup D(s_{t_1}) \cdots \cup D(s_{t_{k-1}})|} \right].
\]

Since the sets are disjoint,

\[
P[e = s] = P(s) = \frac{\binom{N-1}{k-1}}{\binom{N}{k}} \sum_{t=1}^{k} \frac{1}{\binom{N-1}{k-1}} \left[ \frac{1}{|D(a(s))| + z_t} \right],
\]

where \(z_t = |D(s_{t_1}) \cup \cdots \cup D(s_{t_{k-1}})| \geq k - 1\). Let \(s' \in S_{i+\ell}\) be another pseudosolution of level \(i + \ell\). We rewrite \(P(s)\) as

\[
\frac{1}{\left(\binom{N}{k}\right)} \left[ \sum_{t=1}^{k} \frac{1}{|D(a(s)) + a_t + |D(a(s'))|} + \sum_{t=1}^{k} \frac{1}{|D(a(s))| + b_t} \right].
\]

Figure 1: Representation of how Algorithm 1 explores the search tree.

The parameter \(k > 1\) controls the uniformity of the sampling. This is formalized by the following Theorem:

*Theorem 1.* Let \(S\) be the output of Algorithm 2 with input \(\Phi\) such that \(|\Phi| \geq \min\{k, |S_i|\}\), and \(s, s' \in S_{i+\ell}\) be any two pseudosolutions of level \(i + \ell\). We have

\[
\frac{k}{M + k - 1} \leq \frac{P(s)}{P(s')} \leq \frac{M + k - 1}{k},
\]

where \(M = 2^\ell\) and \(P(s)\) is the probability that a uniformly sampled element from \(S\) is equal to \(s\).