*Séminaire Lotharingien de Combinatoire* **61A** (2009), Article B61Ab

# PROBABILISTIC PROOFS OF HOOK LENGTH FORMULAS INVOLVING TREES

Bruce E. Sagan[1]

Department of Mathematics, Michigan State University,
East Lansing, MI 48824-1027, USA

*This paper is dedicated to the memory of Pierre Leroux. In fact, reference* [8] *was written while both Sagan and Yeh were visiting the Université de Québex à Montréal and partaking of Leroux's legendary hospitality.*

Abstract. Recently, Han discovered two formulas involving binary trees which have the interesting property that hooklengths appear as exponents. The purpose of this note is to give a probabilistic proof of one of Han's formulas. Yang has generalized Han's results to ordered trees. We show how the probabilistic approach can also be used in Yang's setting, as well as for a generalization of Han's formula in terms of certain infinite trees.

## 1. Introduction and definitions

Frame, Robinson, and Thrall [1] discovered the hook formula for standard Young tableaux. Greene, Nijenhuis, and Wilf [2] then gave a probabilistic proof of this result where the hook lengths appeared in a very natural way. The same trio also used probabilistic methods to prove the sum of squares formula for the symmetric group [3]. Sagan [7] and Sagan and Yeh [8] gave probabilistic algorithms for proving hook formulas for shifted Young tableaux and trees, respectively.

Recently, inspired by an identity of Postnikov [6], Han [4] proved two identities involving binary trees which have the interesting property that hooklengths appear as exponents. (Han [5] also discovered an identity with this same property which generalizes Postnikov's.) Han's demonstration was by algebraic manipulation of recursions. Yang [9] generalized Han's identities to weighted ordered trees. Again, the proofs were algebraic in nature, this time using generating functions.

The purpose of this note is to give a probabilistic proof of Han's first formula which is similar in some ways to the second algorithm of Greene, Nijenhuis, and

2000 *Mathematics Subject Classification.* Primary 05A19; Secondary 60C05.

*Key words and phrases.* algorithm, hook length, probability, tree.

[1] This material is based on research done while working at the National Science Foundation (NSF). The views expressed do not necessarily reflect those of the NSF..

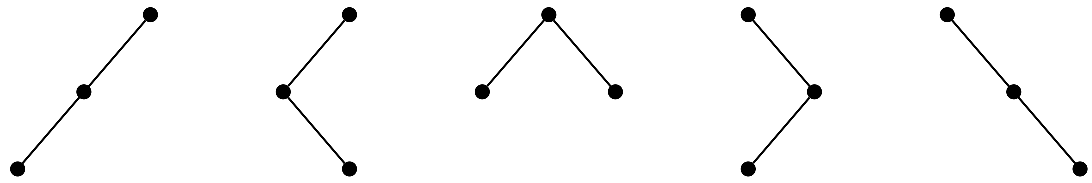

FIGURE 1. The trees in $\mathcal{B}(3)$

Wilf. A weighted version of the algorithm proves the analogous identity of Yang. A second generalization of Han's original formula to certain infinite trees is also demonstrated by this method. The rest of this section is devoted to the necessary definitions to state the identities to be proved. Section 2 gives the probabilistic algorithm and proofs. The final section is devoted to indicating how Han's second formula might be demonstrated probabilistically.

For any tree, $T$, we denote the vertex set of $T$ by $V(T)$. If no confusion will result, we will often write $v \in T$ and $|T|$ in place of the more cumbersome $v \in V(T)$ and $|V(T)|$, where $|\cdot|$ denotes cardinality. If $T$ is rooted and $v \in T$, then the set of children of $v$ will be denoted $C_v$, and we let $c_v = |C_v|$. The *hook of* $v$, $H_v$, is the set of descendents of $v$ (including $v$ itself) with corresponding *hook length* $h_v = |H_v|$.

A *binary tree*, $T$, is a rooted tree where every vertex has either no children, a left child, a right child, or both children. Let $\mathcal{B}$ denote the family of all binary trees and let

$$\mathcal{B}(n) = \{T \in \mathcal{B} \;:\; |T| = n\}.$$

For example, the trees in $\mathcal{B}(3)$ are displayed in Figure 1. In what follows, we will use similar notation for other families of trees. The formula of Han which we will prove is as follows.

**Theorem 1.1** (Han). *For each positive integer $n$ we have*

$$\sum_{T\in\mathcal{B}(n)} \prod_{v\in T} \frac{1}{h_v 2^{h_v-1}} = \frac{1}{n!}. \tag{1}$$

Now consider finite ordered trees weighted by

$$w(T) = \prod_{v\in T} \binom{m}{c_v},$$

where $m$ is a variable. Let $\mathcal{O}$ denote the family of weighted ordered trees. The trees in $\mathcal{O}(4)$ along with their weights are shown in Figure 2. Then the identity of Yang we are considering is equivalent to:

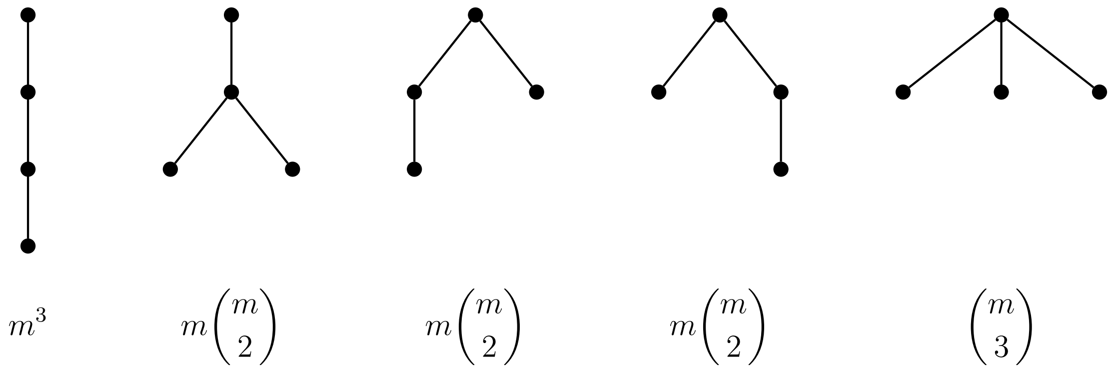


FIGURE 2. The trees in $\mathcal{O}(4)$ and their weights

**Theorem 1.2** (Yang). *For each positive integer $n$ we have*

$$\sum_{T\in\mathcal{O}(n)} w(T)\prod_{v\in T}\frac{1}{h_v m^{h_v-1}} = \frac{1}{n!}. \tag{2}$$

Some comments about this result are in order. First of all, it is remarkable because the right-hand side of the equation does not depend on $m$. Secondly, it implies Han's formula by letting $m = 2$, since then $w(T)$ is just the number of ways of turning an ordered tree into a binary tree. Finally, Yang's weighting was actually

$$w(T) = \prod_{v\in T}\binom{m}{c_v}s^{c_v} = s^{|T|-1}\prod_{v\in T}\binom{m}{c_v},$$

where $s$ is another parameter. So one can recover Yang's equation by multiplying both sides of (2) by $s^{n-1}$. Also, Yang assumes that $m$ and $s$ are constants satisfying certain conditions, but it is clearly not necessary to do so.

For our second generalization of equation (1), consider a fixed, infinite, ordered tree $\overline{T}$ such that $0 < \overline{c}_v < \infty$ for all $v \in \overline{T}$. We are using $\overline{c}_v$ for the number of children of $v$ to emphasize that this is being calculated in $\overline{T}$. Let $\overline{\mathcal{T}}$ be the family of all subtrees of $\overline{T}$ which contain the root of $\overline{T}$. Since $\overline{T}$ is ordered, its vertices are distinguishable, i.e., $V(\overline{T})$ is a set rather than a multiset. So we consider two subtrees $T, T'$ to be equal if and only if $V(T) = V(T')$. For example, $\overline{\mathcal{T}} = \mathcal{B}$ if we let $\overline{T}$ be the tree with $\overline{c}_v = 2$ for all $v \in \overline{T}$. As another illustration, Figure 3 shows part of one possible $\overline{T}$ and the corresponding trees in $\overline{\mathcal{T}}(3)$.

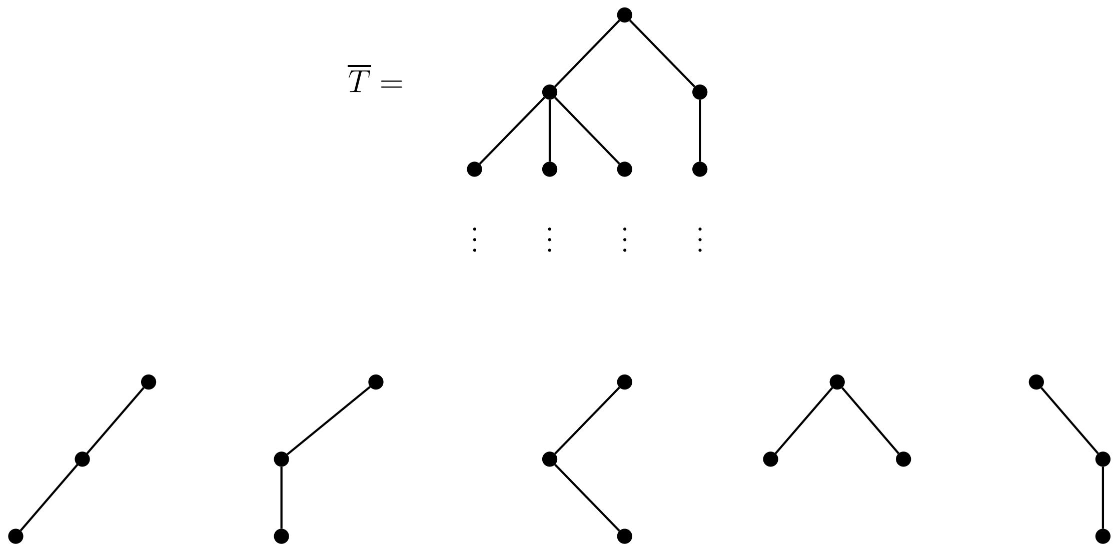


FIGURE 3. The subtrees in $\overline{\mathcal{T}}(3)$ for the given $\overline{T}$

**Theorem 1.3.** *For each positive integer $n$ and each tree $\overline{T}$ satisfying the above restrictions, we have*

$$\sum_{T\in\overline{\mathcal{T}}(n)} \prod_{v\in T} \frac{1}{h_v \overline{c}_v^{\,h_v-1}} = \frac{1}{n!}. \tag{3}$$

## 2. The algorithm

For any rooted tree $T$, an *increasing labeling of $T$* is a bijection

$$\ell : T \to \{1, 2, \ldots, |T|\}$$

such that for any $v \in T$ and any $w \in C_v$ we have $\ell(v) < \ell(w)$. We will often let $L = L(T)$ stand for an increasing labeling of $T$ viewed as $T$ with the labels attached to its vertices. Similarly, we will write $L(\mathcal{F})$ for the family of all increasing labelings of trees in the family $\mathcal{F}$. Let $f^T$ be the number of increasing labelings $L(T)$ where $T$ has distinguishable vertices. The following hook length formula for $f^T$ is well known and easy to prove

$$f^T = \frac{n!}{\prod_{v\in T} h_v}. \tag{4}$$

So if we multiply any of the three identities from the previous section by $n!$, we obtain a sum of the form

$$\sum_{T\in\mathcal{F}(n)} f^T\pi(T) = 1$$

where $\mathcal{F}$ is a family of trees and $\pi(T)$ is a product. We wish to interpret $\pi(T)$ as the probability of obtaining an increasing labeling $L$ of $T$ for an appropriate probability distribution on $L(\mathcal{F}(n))$. The identity will then follow since

$$1 = \sum_{L\in L(\mathcal{F}(n))} \mathrm{Prob}(L) = \sum_{T\in\mathcal{F}(n)} \sum_{L=L(T)} \pi(T) = \sum_{T\in\mathcal{F}(n)} f^T\pi(T).$$

Note that $\mathrm{Prob}(L)$ will depend on $T$ where $L = L(T)$, but not on the specific labeling of $T$.

The probability distribution will be obtained by specifying the parameters in the following basic algorithm which takes as input a positive integer $n$ and a family of trees $\mathcal{F}$ and outputs a labeling $L$ of some $T \in \mathcal{F}(n)$.

(1) Let $L$ consist of a single root labeled 1.
(2) While $|L| < n$, consider all possible leaves $v$ one could add to $L$ and still stay in $L(\mathcal{F})$. Pick one such leaf, label it $|L| + 1$, and add it to $L$ with probability $\mathrm{Prob}(v, L)$.
(3) Output $L$.

It will be convenient to also use the notation $\mathrm{Prob}(v, L)$ when $v \in L$. In that case, it should be interpreted as $\mathrm{Prob}(v, L')$ where $L'$ is subtree of $L$ induced by those vertices with labels less than $\ell(v)$.

To finish the proofs, we just need to specify for each of the three families what the probabilities $\mathrm{Prob}(v, L)$ are, and prove that they describe a probality distribution such that all increasing labelings $L$ of a given tree are equally likely with the common value being

$$\mathrm{Prob}(L) = \prod_{v\in L} \mathrm{Prob}(v, L) = \pi(T).$$

*Proof of* (1). Given a tree $T$ rooted at $r$ and $v \in T$ we let $P_v$ be the unique $r$–$v$ path. The *depth of* $v$, $d_v$, is the length of $P_v$. In the algorithm, let

$$\mathrm{Prob}(v, L) = \frac{1}{2^{d_v}}.$$

For example, Figure 4 shows one of the trees $T$ in $\mathcal{B}(3)$ along with these probabilities for each possible leaf $v$ which could be added to $T$. To further distinguish such leaves from the nodes of $T$, the corresponding edges are dashed.

We first need a lemma which will be used in all three proofs. So we will state it in general terms.

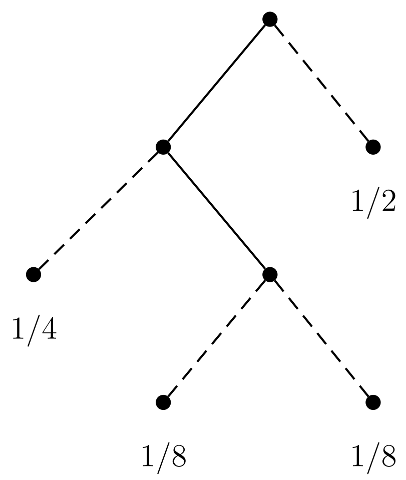


FIGURE 4. A tree in $\mathcal{B}(3)$ and the probabilities of its additional leaves

**Lemma 2.1.** *For each of the three families $\mathcal{F}$ under consideration and for each $L \in L(\mathcal{F})$ we have*

$$\sum_v \operatorname{Prob}(v, L) = 1 \tag{5}$$

*where the sum is over all leaves $v$ which could be added to $L$.*

*Proof of the lemma for $\mathcal{F} = \mathcal{B}$.* We induct on $|L|$ where the base case is easy to do. Given $L$, let $w$ be the leaf of $L$ such that $\ell(w) = |L|$ and let $L' = L - w$. Then the terms in the sum for $L'$ are the same as those in the sum for $L$ except that the summand $1/2^{d_w}$ in the former has been replaced by $1/2^{d_w+1} + 1/2^{d_w+1}$. Since these two expressions are equal, so are the sums, and we are done by induction. □

Next we need to verify that for $L = L(T)$ we have $\operatorname{Prob}(L) = \pi(T)$, i.e.,

$$\operatorname{Prob}(L) = \prod_{v \in T} \frac{1}{2^{h_v - 1}}$$

Again, let $\ell(w) = |L|$ and $L' = L - w$. Then the hook lengths in $L$ are the same as those in $L'$ except that the $d_w$ vertices on the path $P_w - w$ have all been increased by one. Note also that $w$ itself does not contribute to the product above since $1/2^{h_w - 1} = 1$. So, by induction,

$$\operatorname{Prob}(L) = \operatorname{Prob}(w, L') \operatorname{Prob}(L') = \frac{1}{2^{d_w}} \prod_{v' \in T'} \frac{1}{2^{h_{v'} - 1}} = \prod_{v \in T} \frac{1}{2^{h_v - 1}}. \tag{6}$$

There remains to show that $\mathrm{Prob}(L)$ defines a probability distribution. But using the Lemma and induction as well as our usual notation:

$$\begin{aligned}
\sum_{L\in L(\mathcal{B}(n))} \mathrm{Prob}(L) &= \sum_{L\in L(\mathcal{B}(n))} \mathrm{Prob}(w,L')\,\mathrm{Prob}(L') \\
&= \sum_{L'\in L(\mathcal{B}(n-1))} \mathrm{Prob}(L') \sum_{w} \mathrm{Prob}(w,L') \\
&= \sum_{L'\in L(\mathcal{B}(n-1))} \mathrm{Prob}(L') \\
&= 1.
\end{aligned}$$

This finishes the proof of (1). $\square$

Note that the proof that $\mathrm{Prob}(L)$ forms a probability distribution only depends on Lemma 2.1. So in the next two proofs, we will skip this step.

*Proof of* (2). Note that the left-hand side of (2) is a rational function of $m$, so clearing denominators gives a polynomial equation. Thus it suffices to prove that this identity holds for infinitely many values of $m$. We will provide a proof for all real numbers $m \geq M$ where $M$ is chosen sufficiently large such that $0 \leq \mathrm{Prob}(v,L) \leq 1$ for all $v \in L \in L(\mathcal{O}(n))$. This will be possible because $\mathrm{Prob}(v,L)$ will be asymptotic to, but smaller than or equal to, $1/m^{d_v-1}$. Specifically, let

$$\mathrm{Prob}(v,L) = \frac{m-c_p}{(c_p+1)m^{d_v}}$$

where $p$ is the parent of $v$. Remember that, according to our convention following the description of the algorithm, $c_p$ is calculated in the subtree of $L$ induced by those vertices with labels less that $\ell(v)$. In particular, $c_p$ does not count $v$ itself. Figure 5 displays a tree of $\mathcal{O}(4)$ and the probabilities of the leaves which can be added to it.

Our first order of business will be to prove Lemma 2.1 in this setting.

*Proof of the Lemma for $\mathcal{F} = \mathcal{O}$.* As before, we induct on $L$, keeping the notation the same as the first proof. We also let $p$ be the parent of $w$ and write $p'$ if we are considering $p$ as a vertex of $L'$. So $c_p = c_{p'} + 1$ and the terms in the sum for $L'$ corresponding to the $c_{p'} + 1$ possible children which could be added to $p'$ give a total of

$$(c_{p'}+1)\frac{m-c_{p'}}{(c_{p'}+1)m^{d_w}} = \frac{m-c_p+1}{m^{d_w}}.$$

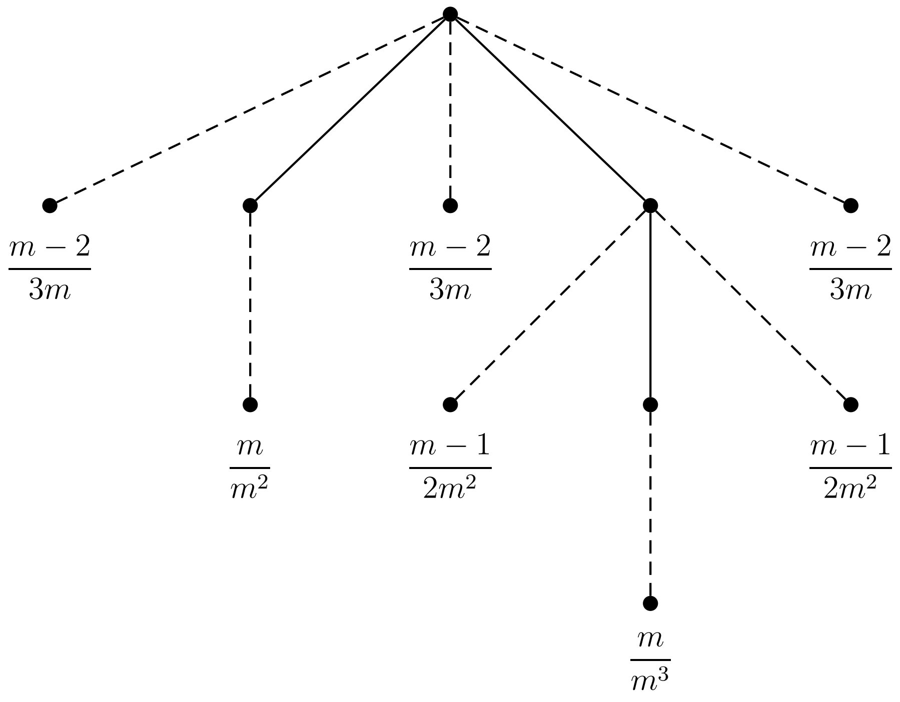


FIGURE 5. A tree in $\mathcal{O}(4)$ and the probabilities of its additional leaves

In the sum for $L$, these terms are replaced by $c_p + 1$ summands for children of $p$ and one for a child of $w$, giving

$$(c_p+1)\frac{m-c_p}{(c_p+1)m^{d_w}} + \frac{m}{m^{d_w+1}} = \frac{m-c_p+1}{m^{d_w}}.$$

Since these two contributions are the same and all other terms in two sums match up, we are done. □

We next need to show that $\mathrm{Prob}(L) = \pi(T)$ for $\mathcal{F} = \mathcal{O}$. Keeping our usual notation we have $\mathrm{Prob}(L)/\mathrm{Prob}(L') = \mathrm{Prob}(w, L')$. So the desired equality will follow by induction, the reasoning applied to obtain (6), and the computation

$$\frac{\pi(T)}{\pi(T')} = \frac{\prod_{v\in T}\binom{m}{c_v}/m^{h_v-1}}{\prod_{v'\in T'}\binom{m}{c_{v'}}/m^{h_{v'}-1}} = \frac{\binom{m}{c_p}\binom{m}{c_w}}{\binom{m}{c_{p'}}m^{d_w}} = \frac{m-c_{p'}}{(c_{p'}+1)m^{d_w}} = \mathrm{Prob}(v, L').$$

This completes the proof for $\mathcal{O}$. □

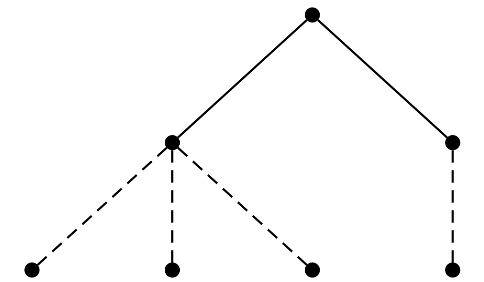


FIGURE 6. A subtree in $\overline{\mathcal{T}}(3)$ and additional leaves

*Proof of* (3). For this case, we proceed as usual, but letting

$$\mathrm{Prob}(v, L) = \prod_{x \in P_v - v} \frac{1}{\overline{c}_x}.$$

Figure 6 gives an example using a tree from $\overline{\mathcal{T}}(3)$ where $\overline{\mathcal{T}}$ is as in Figure 3.

*Proof of the Lemma for* $\mathcal{F} = \overline{\mathcal{T}}$. Now in passing from the sum for $L'$ to the sum for $L$, a single term $\prod_{x \in P_w - w} 1/\overline{c}_x$ has been replaced by $\overline{c}_w$ terms all equal to $\prod_{x \in P_w} 1/\overline{c}_x$. Clearly this does not change the sum. □

The proof that $\mathrm{Prob}(L) = \pi(t)$ is just like the one for $\mathcal{B}$ except that the hook length powers of 2 are replaced by powers of $\overline{c}_x$. So we are done with the case of $\overline{\mathcal{T}}$. □

## 3. An open problem

As remarked in the introduction, Han actually proved two formulas in [4], both having hook lengths as exponents. We have unable to give a probabilistic proof of the second one. But will indicate how one might go in the hopes that someone else may be able to push it through.

Call a binary tree *complete* if every vertex has 0 or 2 children. Given a binary tree $T$ on $n$ nodes it has *completion* $\hat{T}$ which is the complete binary tree obtained from $T$ by adding all $n+1$ possible leaves. If $T$ is the tree with the solid edges in Figure 4 then $\hat{T}$ is the tree which also includes the dashed edges. It is not hard to show using (4) that

$$f^{\hat{T}} = \frac{(2n+1)!}{\prod_{v \in T}(2h_v + 1)}. \qquad (7)$$

Han's second formula is

$$\sum_{T\in\mathcal{B}(n)}\prod_{v\in T}\frac{1}{(2h_v+1)2^{2h_v-1}}=\frac{1}{(2n+1)!}.$$

Using (7), this can be rewritten as

$$\sum_{T\in\mathcal{B}(n)} f^{\hat{T}}\prod_{v\in T}\frac{1}{2^{2h_v-1}}=1.$$

It would be very interesting to find a probability distribution on increasing labelings of complete trees $\hat{T}$ whose probabilities are given by $\prod_{v\in T} 1/2^{2h_v-1}$. Once this is done, similar ideas should prove the generalization to $\mathcal{O}$ due to Yang [9]. It is not clear how to generalize Han's formula to the $\overline{\mathcal{T}}$ case, but would be interesting to do if possible.

## References


[1] Frame, J. S., Robinson, G. d. B., and Thrall, R. M. The hook graphs of the symmetric group. *Canad. J. Math. 6* (1954), 316–325.

[2] Greene, C., Nijenhuis, A., and Wilf, H. S. A probabilistic proof of a formula for the number of Young tableaux of a given shape. *Adv. in Math. 31*, 1 (1979), 104–109.

[3] Greene, C., Nijenhuis, A., and Wilf, H. S. Another probabilistic method in the theory of Young tableaux. *J. Combin. Theory Ser. A 37*, 2 (1984), 127–135.

[4] Han, G.-N. New hook length formulas for binary trees. *Combinatorica*, to appear, preprint ar$\chi$iv:0804.3638.

[5] Han, G.-N. Yet another generalization of Postnikov's hook formula for binary trees. *SIAM J. Discrete Math.*, 23 (2009), 661–664.

[6] Postnikov, A. Permutohedra, associahedra, and beyond. *Internat. Math. Res. Not.*, 2009, Article rnn153v1-rnn153.

[7] Sagan, B. On selecting a random shifted Young tableau. *J. Algorithms 1*, 3 (1980), 213–234.

[8] Sagan, B. E., and Yeh, Y. N. Probabilistic algorithms for trees. *Fibonacci Quart. 27*, 3 (1989), 201–208.

[9] Yang, L. L. M. Generalizations of Han's hook length identities. Preprint ar$\chi$iv:0805.0109.