Forest volume decompositions and Abel-Cayley-Hurwitz multinomial expansions*

Jim Pitman

Technical Report No. 591

Department of Statistics
University of California
367 Evans Hall # 3860
Berkeley, CA 94720-3860

March 29, 2001; revised October 24, 2001

Abstract

This paper presents a systematic approach to the discovery, interpretation and verification of various extensions of Hurwitz's multinomial identities, involving polynomials defined by sums over all subsets of a finite set. The identities are interpreted as decompositions of *forest volumes* defined by the enumerator polynomials of sets of rooted labeled forests. These decompositions involve the following basic *forest volume formula*, which is a refinement of Cayley's multinomial expansion: for $R \subseteq S$ the polynomial enumerating out-degrees of vertices of rooted forests labeled by S whose set of roots is R, with edges directed away from the roots, is

$$(\textstyle\sum_{r\in R} x_r)(\sum_{s\in S} x_s)^{|S|-|R|-1}.$$

*Supported in part by N.S.F. Grants DMS 97-03961 and DMS-0071448

1 Introduction

Hurwitz [21] discovered a number of remarkable identities of polynomials in a finite number of variables $x_s, s \in S$, involving sums of products over all $2^{|\Omega|}$ subsets A of some fixed set $\Omega \subset S$, with each product formed from the subset sums

$$x_A := \sum_{s \in A} x_s \text{ and } x_{\bar{A}} = x_\Omega - x_A, \text{ where } \bar{A} := \Omega - A. \tag{1}$$

For instance, for $S - \Omega = \{0, 1\}$, with x substituted for x_0 and y for x_1 to ease the notation, there are the following:

Hurwitz identities [21, II', III, IV]

$$\sum_{A \subseteq \Omega} x(x + x_A)^{|A|-1}(y + x_{\bar{A}})^{|\bar{A}|} = (x + y + x_\Omega)^{|\Omega|} \tag{2}$$

$$\sum_{A \subseteq |\Omega|} x(x + x_A)^{|A|-1} y(y + x_{\bar{A}})^{|\bar{A}|-1} = (x + y)(x + y + x_\Omega)^{|\Omega|-1} \tag{3}$$

$$\sum_{A \subseteq |\Omega|} (x + x_A)^{|A|}(y + x_{\bar{A}})^{|\bar{A}|} = \sum_{B \subseteq |\Omega|} (x + y + x_\Omega)^{|\bar{B}|} \, |B|! \prod_{b \in B} x_b \tag{4}$$

For $|\Omega| = n$ and $x_s \equiv 1$, by summing first over A with $|A| = k$, these *Hurwitz sums* reduce to corresponding *Abel sums* [1, 48]

$$A_n^{\gamma,\delta}(x, y) := \sum_{k=0}^{n} \binom{n}{k} (x + k)^{k+\gamma}(y + n - k)^{n-k+\delta} \tag{5}$$

for particular integers γ and δ. Especially, (2) with $x_s \equiv w$ and $y = z - nw$ reduces to *Abel's binomial theorem* [1]

$$\sum_{k=0}^{n} \binom{n}{k} x(x + kw)^{k-1}(z - kw)^{n-k} = (x + z)^n$$

and (4) for $x_s \equiv w$ reduces to a classical identity due to Cauchy. Strehl [56] explains how Hurwitz was led to such identities via the combinatorial problem, which arose in the theory of Riemann surfaces [20], of counting the number of ways a given permutation can be written as a product of a minimal number of transpositions which generate the full symmetric group.

This paper develops a systematic approach to the discovery, interpretation and verification of identities of Hurwitz type such as (2)-(4). This method is related to, but not the same, as Knuth's method in [32] and [31, Ex. 2.3.4.30]. There Hurwitz's binomial theorem (3) was proved by interpreting both sides for positive integer x, y and $x_s, s \in \Omega$ as a number of rooted forests with $x + y + x_\Omega + |\Omega|$ vertices, subject to constraints depending on the x_s. The present method is closer to the approach of Françon[18], who derived (2) and (3) using a form of Cayley's multinomial expansion over trees, expressed in terms of enumerator polynomials for mappings from S to S. Section 2 introduces the notion of *forest volumes*, defined by the enumerator polynomials of sets of rooted labeled forests, with emphasis on a generalization of Cayley's multinomial expansion over trees, called here the *forest volume formula.* This formula relates Hurwitz type identities to decompositions of forest volumes. Probabilistic interpretations of forest volumes were developed in [40, 41, 42, 43]. This paper, written mostly in combinatorial rather than probabilistic language, is a condensed version of [40]. A combined version of [40] and [41] appears in the companion paper [44]. Section 3 offers two different extensions of the forest volume formula. Then Section 4 presents a number of Hurwitz type identities, with proofs by forest volume decompositions. Finally, Section 5 points out some specializations of these Hurwitz identities which give combinatorial interpretations of Abel sums.

2 Forest volumes

Rooted labeled forests. Françon's approach to Hurwitz identities is simplified by working exclusively with various subsets of the set

$$\boldsymbol{F}_S := \{\text{all rooted forests } F \text{ labeled by } S\}$$

whose enumerative combinatorics has been extensively studied [37, 2.4 and 3.5], [54, 5.3]. Each $F \in \boldsymbol{F}_S$ is a *directed graph with vertex set* S, that is a subset of $S \times S$, each of whose components is a tree with some root $r \in S$, with the convention in this paper that the edges of F are directed away from the roots of its trees. The set of vertices of the tree component of F rooted at r is the set of all $s \in S$ such that there is a *directed path from* r *to* s *in* F, denoted $r \overset{F}{\rightsquigarrow} s$, meaning either $r = s$ or there is a sequence of one or more edges $r \overset{F}{\to} \cdots \overset{F}{\to} s$, where $s_1 \overset{F}{\to} s_2$ means $(s_1, s_2) \in F$. Let $F_s := \{x : s \overset{F}{\to} x\}$

denote the *set of children of* s in F. So $|F_s|$ is the *out-degree* of vertex s in the forest F. Note that the F_s are possibly empty disjoint sets with $\cup_{s\in S} F_s = S - \mathrm{roots}(F)$, where $\mathrm{roots}(F)$ is the set of root vertices F. So the total number of edges of F is $\sum_{s\in S} |F_s| = |S| - |\mathrm{roots}(F)|$ and $|\mathrm{roots}(F)|$ is the number of tree components of F.

Forest volumes. For $B \subseteq \boldsymbol{F}_S$, the enumerator polynomial in variables $x_s, s \in S$

$$V_S[F \in B] := V_S[F \in B](x_s, s \in S) := \sum_{F\in B} \prod_{s\in S} x_s^{|F_s|} \tag{6}$$

is called here the *volume of* B, to emphasise that $B \to V_S[F \in B]$ is a measure on subsets B of $\boldsymbol{F}_S$, for each fixed choice of $(x_s, s \in S)$ with $x_s \geq 0$. As explained later in this section, this notion of forest volumes includes both the probabilistic interpretations developed in [43, 41, 42, 40], and the forest volume of a graph defined by Kelmans [29, 26, 27].

Consider the volume of all forests $F \in \boldsymbol{F}_S$ with a given set of roots R. This volume decomposes according to the set $A = \cup_{r\in R} F_r$ of children of all root vertices of F. With $\Omega := S - R$, this gives the recursive volume decomposition

$$V_{R\cup\Omega}[\mathrm{roots}(F) = R] = \sum_{A\subseteq\Omega} x_R^{|A|} V_\Omega[\mathrm{roots}(F) = A] \tag{7}$$

where each side is a polynomial in variables $x_s, s \in S$, and $S = \Omega \cup R$ with $R \cap \Omega = \emptyset$. Consideration of (7) for small $|S|$ leads quickly to the following generalization of Cayley's multinomial expansion over trees [13, 47]:

Theorem 1 (The forest volume formula) [13, 45, 47, 18, 11, 40]
For $R \subseteq S$, the volume of forests labeled by S whose set of roots is R is

$$V_S[\mathrm{roots}(F) = R] = x_R\, x_S^{|S|-|R|-1}. \tag{8}$$

Proof. Observe first that (8) transforms (7) into the following Hurwitz type identity of polynomials in variables x_R and $x_s, s \in \Omega$:

$$x_R(x_R + x_\Omega)^{|\Omega|-1} = \sum_{A\subseteq\Omega} x_R^{|A|}\, x_A\, x_\Omega^{|\Omega|-|A|-1}. \tag{9}$$

But (9) is very easily verified directly, and (8) follows from (7) and (9) by induction on $|S|$. □

History and alternative proofs of the forest volume formula. Cayley [13] formulated the special case of (8) with $R = \{r\}$, call it the *tree volume formula*, along with the special case of (8) with general R and $x_s \equiv 1, s \in S$, that is the enumeration

$$|\{F \in \boldsymbol{F}_S \text{ with } \mathrm{roots}(F) = R\}| = |R|\,|S|^{|S|-|R|-1} \tag{10}$$

which for $|R| = 1$ yields the Borchardt-Cayley formula n^{n-2} for the number of unrooted trees labeled by a set of n vertices. These special cases of the forest volume formula are among the best known results in enumerative combinatorics. See for instance [45, 47, 43, 42, 54, 7] for various proofs of the tree volume formula and [19, 36, 37, 57, 51, 3] for (10). The preceding proof of the forest volume formula parallels a proof of (10) by induction, using the consequence of (7) that the number in (10) is $\#(|S|, |R|)$ with the recursion

$$\#(k+n, k) = \sum_{a=1}^{n} \binom{n}{a} k^a \#(n, a). \tag{11}$$

Moon [37, p. 33] attributes this proof of (10) to Göbel [19]. The forest volume formula can also be derived by the method of Prüfer codes [45], which has been applied to obtain a host of other results in the same vein [30],[37, Chapter 2]. Another approach to the forest volume formula is to combine any of the proofs of the tree volume formula cited above with the following:

Reduction of forest volumes to tree volumes. Fix some arbitrary $r \in R$, and decompose the volume $V_S[\mathrm{roots}(F) = R]$ according to the tree $T = \mathrm{tree}(F)$ derived by identifying all root vertices of F with r. So T is labeled by $\{r\} \cup (S - R)$, and the sets of children of T are $T_r := \cup_{s \in R} F_s$ and $T_s := F_s$ for $s \in S - R$. For each possible tree T, it is easily seen that

$$V_S[\mathrm{roots}(F) = R, \mathrm{tree}(F) = T] = x_R^{|T_r|} \prod_{s \in S-R} x_s^{|T_s|}.$$

Now sum over all T and apply the tree volume formula to deduce the forest volume formula.

Reformulation in terms of mappings. The forest volume formula is easily recast as a formula for the enumerator of mappings from S to S whose set of cyclic points is a set of fixed points equal to R. This mapping enumerator was derived by Françon [18, Prop. 3.1 and proof of Prop. 3.5] from the Foata-Fuchs encoding of mappings [17] for $|R| \leq 2$. The corresponding enumerator for general R can be inferred from the discussion in [18, p. 337] and derived the same way. As indicated in [40, 7], a probabilistic form of the forest volume formula, expressed in terms of random mappings, can be read from Burtin [11]. The general form (8) of the forest volume formula was given in [40, Th. 1] with a roundabout proof via random mappings.

Variations of the forest volume formula. For each vector $(c_s, s \in S)$ of non-negative integers with $\sum_s c_s = |S| - |R|$, the identity of coefficients of $\Pi_s x_s^{c_s}$ in (8) reads

$$|\{F \in \boldsymbol{F}_S : |F_s| = c_s \text{ for all } s \in S\}| = \frac{c_R\,(|S| - |R| - 1)!}{\prod_{s \in S} c_s!}. \tag{12}$$

In the tree case $|R| = 1$, this is easily verified by a direct argument [43, §5.1]. The same argument gives the identity of coefficients of $\Pi_s x_s^{c_s}$ in the formula for the volume of all forests with exactly k tree components [42, §2]:

$$V_S[F \text{ with } k \text{ trees}] = \binom{|S| - 1}{k - 1} x_S^{|S|-k}. \tag{13}$$

This in turn is the identity of coefficients of x_0^k in the tree volume formula applied to $S \cup \{0\}$ instead of S for some $0 \notin S$. See Stanley [54, Th. 5.3.4] for two other proofs of (13). It is obvious that (8) implies (13) by summation over R with $|R| = k$, but less obvious how to recover (8) from (13). Neither does it seem easy to check the coefficient identities (12) directly for general R. By summation of (13) over k, the total volume of all forests labeled by S is

$$V_S[F \in \boldsymbol{F}_S] = (1 + x_S)^{|S|-1}. \tag{14}$$

Hurwitz's multinomial formula. An argument suggested by Françon[18, p. 337] in terms of mapping enumerators can be simplified as follows. The forest volume $V_S(\text{roots}(F) = R)$ can be decomposed by classifying F according to the sets B_r of non-root vertices of the tree components of F, as r ranges over R. For given $(B_r, r \in R)$, there is an obvious factorization of the forest volume over disjoint tree components, and the tree volume formula can be applied within each component. Therefore

$$V_S[\mathrm{roots}(F) = R] = \sum_{(B_r)} \prod_{r \in R} x_r (x_r + x_{B_r})^{|B_r|-1} \tag{15}$$

where the sum is over the $|R|^{|S|-|R|}$ possible choices of disjoint, possibly empty sets $(B_r, r \in R)$ with $\cup_{r \in R} B_r = S - R$. The equality of right hand sides of (8) and (15) is *Hurwitz's multinomial formula* [21, VI]:

$$x_R (x_R + x_\Omega)^{|\Omega|-1} = \sum_{(B_r)} \prod_{r \in R} x_r (x_r + x_{B_r})^{|B_r|-1} \tag{16}$$

whose binomial case $|R| = 2$ is (3). As a check on this circle of results, (16) is easily derived by induction on $|R|$ from its binomial case, which was interpreted combinatorially by Knuth [32] and [31, Ex. 2.3.4.30]. The forest volume formula can then be read from the consequence (15) of the tree volume formula.

Probabilistic interpretations. Suppose in this paragraph that $(x_s, s \in S)$ is a *probability distribution on* S, meaning $x_s \geq 0$ for all $s \in S$ and $x_S = 1$. Let F_1^* denote a random tree distributed according to the forest volume distribution restricted to forests with a single tree component. That is to say, for each tree T labeled by S, the probability that F_1^* equals T is

$$P(F_1^* = T) = \prod_{s \in S} x_s^{|T_s|}$$

where the probabilities sum to 1 over all trees T by (13) for $k = 1$. For $0 \leq p \leq 1$ let F_p^* denote the random forest obtained by retaining each edge e of F_1^* with probability p, and deleting it with probability $1-p$, independently as e ranges over the set of $|S| - 1$ edges of F_1^*. According to [43, Theorem 11 and (43)] the distribution of F_p^* on $\boldsymbol{F}_S$ is given by the formula

$$P(F_p^* \in B) = p^{|S|-1} V_{S,(1-p)/p}[F \in B] \tag{17}$$

where

$$V_{S,z}[F \in B] := z^{-1} \sum_{k=1}^{n} z^k V_S[F \text{ has } k \text{ trees and } F \in B] \tag{18}$$

is a polynomial in variables z and $(x_s, s \in S)$. For for $z = x_0$ with $0 \notin S$, this polynomial is x_0^{-1} times the volume of trees labeled by $\{0\} \cup S$ whose restriction to S is a forest in B. The probabilistic interpretation of forest volumes (17) makes an important connection between the theory of measure-valued and partition-valued coalescent processes and the asymptotic structure of large random trees [6, 4, 12, 15].

Forest volumes of graphs. In the particular case when B is the set of all forests contained in some directed graph G with vertex set S, Kelmans, Pak and Postnikov [26, 27] call the polynomial (18) the *forest volume of* G. They obtain a formula for the forest volume of a graph G built up by a composition operation from simpler graphs, and apply this formula to obtain some Hurwitz type identities by the same method of forest volume decompositions used in this paper. In [26, 27] the basic building block for explicit formulae is taken to be the tree volume formula([26, 12.1] and [27, 3.2]) rather than the forest volume formula. But the forest volume formula yields Hurwitz type sums more easily, as can be seen by comparing the previous discussion around (15)-(16) with the treatment of the same result in [26, §13] and [27, 4.2]. There seems to be little overlap between the variations and extensions of Hurwitz's identities found in Section 4 of this paper, by consideration of volumes $V_S(F \in B)$ for various subsets B of $\boldsymbol{F}_S$, and the family of Hurwitz type identities found in [27] by consideration of forest volumes of graphs. The basic method of forest volume decompositions is doubtless capable of generating still more identities in the same vein.

Other related work. Abramson [2] derived the particular case of Hurwitz's multinomial formula (16) with $x_s \equiv 1, s \in \Omega$ by Knuth's method, along with more complicated multinomial identities associated with a directed graph. Presumably the results in this paper could be also recovered by Knuth's method, but the present approach seems much simpler. See also Stam [52] for another kind of extension of Hurwitz's identities, involving polynomials of binomial type. Some other papers which treat aspects of Abel and Hurwitz identities and their connections to rooted labeled trees are [10, 22, 49, 50, 55, 58].

3 Extensions of the forest volume formula

Theorems 2 and 3 in this section present two different extensions of the forest volume formula. Only the first of these extensions is needed for the applications to Hurwitz identities in Section 4.

An oriented percolation volume. The following formula was discovered in connection with the problem, solved by Corollary 5 in the next section, of finding the probability of the event $r \stackrel{F}{\rightsquigarrow} s$ for a random forest F distributed according to the forest volume distribution conditioned on forests of k trees.

Theorem 2 *For each $R \subseteq S$ and each fixed choice of an $r \in R$ and an $s \in S - R$,*

$$V_S[\text{roots}(F) = R \textit{ and } r \stackrel{F}{\rightsquigarrow} s] = x_r \, x_S^{|S|-|R|-1}. \tag{19}$$

Proof. Let $\boldsymbol{F}_{S,R}$ denote the set of $F \in \boldsymbol{F}_S$ with $\text{roots}(F) = R$. For $r, r' \in R$ there is a bijection between

$$\{F \in \boldsymbol{F}_{S,R} \text{ with } r \stackrel{F}{\rightsquigarrow} s\} \text{ and } \{F' \in \boldsymbol{F}_{S,R} \text{ with } r' \stackrel{F'}{\rightsquigarrow} s\}$$

whereby F' is derived from F by deleting the first edge, $r \to t$ say, on the path from r to s in F, and adding the edge $r' \to t$. Denote the the volume on the left side of (19) by $V_{S,R}(r \rightsquigarrow s)$. The bijection gives

$$\frac{x_{r'}}{x_r} V_{S,R}(r \rightsquigarrow s) = V_{S,R}(r' \rightsquigarrow s).$$

Sum this over $r' \in R$ and use the forest volume formula (8) to get (19). □

Probabilistic applications. To simplify the next two displays (20) and (21), let

$$P_{S,k}[\cdots] := \frac{V_S[\cdots \text{ and } F \text{ has } k \text{ trees}]}{V_S[F \text{ has } k \text{ trees}]}$$

where the denominator is given explicitly by (13). Assuming that $x_s \geq 0$ for all s, this ratio can be interpreted as the probability of the event $[\cdots]$ for a random forest F distributed according to the normalized volume measure on forests of k trees labeled by S, as considered in [43, 41, 40]. Sum (19) over appropriate R to obtain for each choice of distinct $r, s \in S$, and each $1 \leq k \leq n := |S|$,

$$P_{S,k}[r \in \mathrm{roots}(F), r \overset{F}{\rightsquigarrow} s] = \frac{(n-k)}{(n-1)} \frac{x_r}{x_S} \tag{20}$$

and

$$P_{S,k}[r \in \mathrm{roots}(F), r \overset{F}{\not\rightsquigarrow} s] = \frac{k-1}{n-1}. \tag{21}$$

According to (21), for each choice of non-negative integers $c_v, v \in S$ with $\sum_{v \in S} c_v = n - k$, among all forests F of k trees labeled by S such that v has c_v children in F for every $v \in S$, the fraction of F such that both $r \in \mathrm{roots}(F)$ and $r \overset{F}{\not\rightsquigarrow} s$ equals $(k-1)/(n-1)$. This rather surprising result can also be checked by direct enumeration as in [42, §2].

Adding the right sides of (20) and (21) gives a simple formula for the probability $P_{S,k}[r \in \mathrm{roots}(F)]$. This formula can be checked by summation of the forest volume formula over all R containing r with $|R| = k$. Obvious variations of this argument yield formulae for $P_{S,k}[\mathrm{roots}(F) \supseteq A]$ and $P_{S,k}[\mathrm{roots}(F) \subseteq A]$ for $A \subseteq S$.

The volume of forests containing a given subforest. Another generalization of the forest volume formula with interesting probabilistic applications [41] can be formulated as follows:

Theorem 3 *For each $G \in \boldsymbol{F}_S$ with g tree components $T_r(G), r \in \mathrm{roots}(G)$, and each $R \subseteq \mathrm{roots}(G)$,*

$$V_S[\mathrm{roots}(F) = R \text{ and } F \supseteq G] = \left(\sum_{r \in R} x_{T_r(G)} \right) x_S^{g-|R|-1} \prod_{s \in S} x_s^{|G_s|}. \tag{22}$$

Proof. This is similar to the condensation argument used in Section 2 to reduce the forest volume formula to the tree volume formula. For $F \in \boldsymbol{F}_S$ with $F \supseteq G$, define a *G-condensed forest* $\tilde{F}$, with vertex set roots(G), by the following adaptation of Moon's construction in [34, 37]: collapse each tree component $T_r(G)$ onto its root r, and link these components according to F. That is, G-condensed$(F) = \tilde{F}$ where for $s, t \in \mathrm{roots}(G)$,

$$s \overset{\tilde{F}}{\to} t \Leftrightarrow s' \overset{F}{\to} t \text{ for some } s' \in T_s(G).$$

It is easily checked that $\tilde{F} \in \boldsymbol{F}_{\text{roots}(G)}$ and $\text{roots}(\tilde{F}) = R$. Moreover, for each such $\tilde{F}$ the volume of $F \in \boldsymbol{F}_S$ with $\text{roots}(F) = R$, $F \supseteq G$ and $G\text{-condensed}(F) = \tilde{F}$ is easily found to be

$$\left(\prod_{r \in \text{roots}(G)} x_{T_r(G)}^{|\tilde{F}_r|} \right) \prod_{s \in S} x_s^{|G_s|}$$

and (22) follows by summation over all $\tilde{F} \in \boldsymbol{F}_{\text{roots}(G)}$ with $\text{roots}(\tilde{F}) = R$, using the forest volume formula and

$$\sum_{r \in \text{roots}(G)} x_{T_r(G)} = x_S. \tag{23}$$

□

By summation of (22) over $R \subseteq \text{roots}(G)$, using (23), for each $G \in \boldsymbol{F}_S$ with g tree components

$$V_S(F \text{ has } k \text{ trees and } F \supseteq G) = \binom{g-1}{k-1} x_S^{g-k} \prod_{s \in S} x_s^{|G_s|} \tag{24}$$

which was found by a different method in [41]. The case of (24) with $x_s \equiv 1$ is a result of Stanley [53, Ex. 2.11.a], which goes back to Moon [34] and [37, Th. 6.1] for $k = 1$. See also Pemantle [39, Th. 4.2], where a determinantal formula was found for the probability $P(T \supseteq G)$ for T a uniform random spanning tree of a graph. Other related papers are [38, 28, 14]. In contrast to (24), there is no simple general formula for $V_S[F \text{ has } k \text{ trees and } F \subseteq G]$ for a general directed graph G. But see [26] for expressions of the generating function (18) of these volumes for graphs with special structure. Such expressions, which can be related to (24) by the method of inclusion and exclusion, extend classical results on the enumeration of the numbers of spanning trees and spanning forests of G, discussed in Moon [37, Chapter 6] and Stanley [53, Ex. 2.11.a].

4 Hurwitz identities

This section presents five extensions and variations of Hurwitz's identities (2)-(4) as corollaries of the forest volume formulae provided by Theorems 1 and 2. Each identity is derived from its interpretation as a decomposition of forest volumes. Throughout the section, Ω is a finite set disjoint from $\{0,1\}$, and the notation (1) is used. First, an interpretation of Hurwitz's identity (2) as a forest volume decomposition:

Proof of (2). Let $x = x_0, y = x_1$, and multiply both sides of (2) by x_1, so that (2) can be rewritten

$$\sum_{A \subseteq \Omega} x_0(x_0 + x_A)^{|A|-1} x_1(x_1 + x_{\bar{A}})^{|\bar{A}|} = x_1(x_0 + x_1 + x_\Omega)^{|\Omega|}. \tag{25}$$

By the tree volume formula, the right side of (25) is the volume of trees T with root 1 labeled by $\{0,1\} \cup \Omega$. The left side classifies these trees T according to the set $A = \{v \in \Omega : 0 \overset{T}{\rightsquigarrow} v\}$ of non-root vertices of the fringe subtree of T rooted at 0. By applications of the tree volume formula, the factor $x_0(x_0 + x_A)^{|A|-1}$ is the volume of trees rooted at 0 labeled by $\{0\} \cup A$, while the factor $x_1(x_1 + x_{\bar{A}})^{|\bar{A}|}$ is the volume of trees rooted at 1 labeled by $\{0,1\} \cup \bar{A}$ in which 0 is a leaf vertex (and the variable x_0 is set equal to 0 to achieve this constraint). □

The particular case $m = 0$ of the next result gives another interpretation of the same Hurwitz identity (2). See also [40] for discussion of further interpretations of (2) in terms of random mappings, including those of Françon [18, p. 339], and Jaworski [23, Theorem 3], whose equivalence with the interpretations given here can be explained using Joyal's bijection [24] between marked rooted trees and mappings [9].

Corollary 4 *For* $0 \le m \le |\Omega|$

$$\sum_{A \subseteq \Omega} x_0(x_0 + x_A)^{|A|-1} \binom{|\bar{A}|}{m} (x_1 + x_{\bar{A}})^{|\bar{A}|-m} = \binom{|\Omega|}{m} (x_0 + x_1 + x_\Omega)^{|\Omega|-m} \tag{26}$$

$$= x_0^{-1} V_{\{0,1\}\cup\Omega}[F \textit{ with } m+1 \textit{ trees} : 0 \in \mathrm{roots}(F) \textit{ and } 0 \overset{F}{\rightsquigarrow} 1]. \tag{27}$$

Proof. The right side of (26) equals (27) by summation of (19) with $r = 0$ over $R \subseteq \{0\} \cup \Omega$ with $0 \in R$ and $|R| = m + 1$. So does the left side, by decomposing the volume in (27) according to the set A of all $s \in \Omega$ such that there is a directed path from 0 to s in F which does not pass via 1, using (13). □

As a check, (26) can also be deduced from its special case $m = 0$ by replacing x_1 by $x_1 + \theta$ and equating equating coefficients of θ^m. The same remark applies to the next identity, whose special case $m = 0$ is Hurwitz's identity (4).

Corollary 5 *For* $0 \le m \le |\Omega|$

$$\sum_{A\subseteq\Omega}(x_0+x_A)^{|A|}\binom{|\bar{A}|}{m}(x_1+x_{\bar{A}})^{|\bar{A}|-m} = \sum_{B\subseteq\Omega}(x_0+x_1+x_\Omega)^{|\bar{B}|-m}\binom{|\bar{B}|}{m}|B|!\prod_{s\in B}x_s$$

$$= x_0^{-1}V_{\{0,1\}\cup\Omega}[F \textit{ with } m+1 \textit{ trees} : 0 \overset{F}{\leadsto} 1]. \tag{28}$$

Proof. On the left side F is classified by the set A of $s \in \Omega$ such that $0 \overset{F}{\leadsto} s$ and 1 does not lie on the path from 0 to s, and the volume is evaluated with the help of (19). On the right side F is classified by the set B of $s \in \Omega$ such that s lies on the path in F which joins 0 to the root of its tree component in F, and the volume for given B is computed by consideration of the forest obtained from F by cutting the edges along this path. □

The last proof used the idea, adapted from Meir and Moon [33] and Joyal [24], of generating a forest by cutting the edges along some path in a tree. The same idea yields the following identity:

Corollary 6

$$\sum_{A\subseteq\Omega}(x+x_A)(x+x_\Omega)^{|\bar{A}|-1}|A|!\prod_{s\in A}x_s = (x+x_\Omega)^{|\Omega|}. \tag{29}$$

Proof. For $x = x_0 + x_1$ the right side is $x_0^{-1}V_{\{0,1\}\cup\Omega}$[trees F with root 0], by the tree volume formula. On the left side the trees are classified by the set A of all vertices that lie on the path in the tree from 0 to 1. Cutting the edges along this path makes a forest labeled by $\{0,1\}\cup\Omega$ whose set of roots is $\{0,1\}\cup A$. So the required volume is easily found using the forest volume formula. □

As a final example of Hurwitz type sums over subsets, (13) yields easily:

Corollary 7 *For* $1 \leq m \leq |\Omega|$

$$\sum_{A \subseteq \Omega} \binom{|\bar{A}| - 1}{m - 1} (x_0 + x_A)^{|A|} (x_{\bar{A}})^{|\bar{A}|-m} = \binom{|\Omega|}{m} (x_0 + x_\Omega)^{|\Omega|-m} \tag{30}$$

which is the the volume of forests F *of* $m+1$ *rooted trees labeled by* $\{0\} \cup \Omega$*, with* F *classified on the left side by the set* A *of vertices other than* 0 *in the tree component of* F *that contains* 0*.*

More exotic Hurwitz type identities, involving sums over partitions, can be obtained in a similar way. For instance:

Corollary 8 *The volume of all trees labeled by* $\{0\} \cup \Omega$ *and rooted at* 0 *is*

$$\sum_{h=1}^{|\Omega|} \sum_{(L_1,\dots,L_h)} x_0^{|L_1|} \prod_{j=2}^{h} x_{L_{j-1}}^{|L_j|} = x_0(x_0 + x_\Omega)^{|\Omega|-1} \tag{31}$$

where the inner sum is over all ordered partitions of Ω *into* h *non-empty subsets* $L_1, \dots, L_h$*, and the trees are classified according to the set* L_j *of all vertices of all vertices at level* j *of the tree (meaning at distance* j *from the root* 0*) with* h *representing the maximum height of all vertices of the tree.*

Proof. This follows easily from the tree volume formula by iteration of the argument leading to (7). □

The instance of (31), with $x_s \equiv 1$ for all $s \in \Omega$ and x_0 a positive integer, was discovered by Katz [25], who used it to show that the number $C(n,k)$ of mappings from an n element set to itself whose digraph is connected with exactly k cyclic points is

$$C(n,k) = \binom{n}{k} (k-1)!(kn^{n-k-1}). \tag{32}$$

As remarked by Rényi [46], the transparent combinatorial meaning of the factors in (32) allows either of the formulas (10) and (32) to be derived immediately from the other. See Moon [37, 3.6] for further discussion, and [23, 40] for the extension of this argument which gives the distribution of the number of cyclic points in the digraph of a random mapping $s \to M_s$ when the M_s are independent with a common probability distribution which might not be uniform.

Remarks. The identities in this section were first obtained in [40] by finding the probability of some event determined by a random forest in two different ways. More identities of this kind can be gleaned from [40] or derived by the same method. The technique of conditioning on a suitable subtree yields Hurwitz type sums for many other probabilities of interest. A polynomial sum of 2^n terms is of course hard to compute for large n, but not as hard as the typical sum of n^{n-1} terms which defines the probability, and large n asymptotics of many Hurwitz sums can be handled as in [5, 4, 8, 12].

5 Abel identities

For $x_s \equiv 1$ the Hurwitz type identities and their interpretations described in the previous section reduce to corresponding results for Abel sums. For instance, the Abel type identity derived from (30) is

$$\sum_{k=0}^{n-m} \binom{n}{k} \binom{n-k-1}{m-1} (x+k)^k (n-k)^{n-k-m} = \binom{n}{m} (x+n)^{n-m} \tag{33}$$

for $1 \leq m \leq n$. The Abel type identity derived from (29) is the case $b = 0$ of the telescoping sum

$$\sum_{k=b}^{n} (n)_k (x+k)(x+n)^{n-k-1} = (n)_b (x+n)^{n-b} \quad (0 \leq b \leq n). \tag{34}$$

Probabilistic interpretation of an Abel sum. Corollary 5 specializes for $m = 0$ to give the following probabilistic interpretation of the Abel sum $A_n^{0,0}(x,y)$ defined by (5), with an asymptotic expression obtained by a straightforward integral approximation using the local normal approximation to the binomial distribution [16]:

Corollary 9 *For T a random tree distributed according to the forest volume distribution conditioned on the set of all trees labeled by $S := \{0, 1, \ldots, n+1\}$, with $x_0 = x, x_1 = y$ and $x_s \equiv 1$ for $1 < s \leq n+1$, the probability that there is a directed path from 0 to 1 in T is*

$$P(0 \stackrel{T}{\rightsquigarrow} 1) = \frac{x A_n^{0,0}(x,y)}{(n+x+y)^{n+1}} \sim \sqrt{\frac{\pi}{2}} \frac{x}{\sqrt{n}} \text{ as } n \to \infty. \tag{35}$$

In the particular case when $x = y = 1$, the distribution of T is uniform on the $(n+2)^{n+1}$ rooted trees labeled by S, so (35) implies that for all distinct $u, v \in S := \{0, 1, \ldots, n+1\}$

$$P(u \overset{T}{\rightsquigarrow} v) = A_n^{0,0}(1,1)/(n+2)^{n+1}. \tag{36}$$

So $A_n^{0,0}(1,1)$ is the number of rooted trees labeled by S in which there is a directed path from u to v. It is easy to deduce from (36) the result of Moon [35, Theorem 1], that the conditional expectation of the size $|\{v : 0 \overset{T}{\rightsquigarrow} v\}|$ of the fringe subtree of T with root 0, given that T has some root other than 0, is $A_n^{0,0}(1,1)/(n+2)^n$. The asymptotic evaluation in (35) agrees with Moon's asymptotic formula $\sqrt{\pi n/2}$ for this conditional expectation. As observed by Moon, this conditional expectation is also the expected distance in T between any two distinct vertices $u, v \in S$. See [8] for a study of the asymptotic behaviour for large n of the distribution of the size of the fringe tree $\{v : 0 \overset{T}{\rightsquigarrow} v\}$ for T distributed according to a non-uniform volume distribution on trees labeled by S, and [4, 12] for further study of the asymptotics of large random trees of this kind.

Acknowledgment

Thanks to Jürgen Bennies, Ira Gessel, Richard Stanley, Volker Strehl and some anonymous referees for helpful suggestions and remarks on earlier versions of this paper.

References

[1] N.H. Abel. Beweis eines Ausdrucks von welchem die Binomial-Formel einer einzelner Fall ist. *Crelle's J. Reine Angew. Math.*, 1:159–160, 1826.

[2] H. D. Abramson. A note on a result of Knuth or identities grow on trees. *J. Combinatorial Theory*, 7:371–373, 1969.

[3] M. Aigner and G. M. Ziegler. *Proofs from The Book*. Springer-Verlag, Berlin, second edition, 2001.

[4] D. Aldous and J. Pitman. A family of random trees with random edge lengths. *Random Structures and Algorithms*, 15:176–195, 1999.

[5] D. Aldous and J. Pitman. Inhomogeneous continuum random trees and the entrance boundary of the additive coalescent. *Probab. Th. Rel. Fields*, 118:455–482, 2000.

[6] D.J. Aldous and J. Pitman. The standard additive coalescent. *Ann. Probab.*, 26:1703–1726, 1998.

[7] S. Anoulova, J. Bennies, J. Lenhard, D. Metzler, Y. Sung, and A. Weber. Six ways of looking at Burtin's lemma. *Amer. Math. Monthly*, 106(4):345–351, 1999.

[8] J. Bennies and J. Pitman. Asymptotics of the Hurwitz binomial distribution related to mixed Poisson Galton-Watson trees. *Combinatorics, Probability and Computing*, 10:203–211, 2001.

[9] J. Bennies and A. Weber. A bijection between marked trees and random mappings. Preprint. Dept. Math., Univ. of Frankfurt., 1999.

[10] A.Z. Broder. A general expression for Abelian identities. In L.J. Cummings, editor, *Combinatorics on words*, pages 229–245. Academic Press, New York, 1983.

[11] Y. D. Burtin. On a simple formula for random mappings and its applications. *J. Appl. Probab.*, 17:403 – 414, 1980.

[12] M. Camarri and J. Pitman. Limit distributions and random trees derived from the birthday problem with unequal probabilities. *Electron. J. Probab.*, 5:Paper 2, 1–18, 2000.

[13] A. Cayley. A theorem on trees. *Quarterly Journal of Pure and Applied Mathematics*, 23:376–378, 1889. (Also in *The Collected Mathematical Papers of Arthur Cayley. Vol XIII*, 26-28, Cambridge University Press, 1897).

[14] N. Dean, A. K. Kelmans, K.-W. Lih, W. A. Massey, and P. Winkler. The spanning tree enumeration problem for digraphs. In *Graph theory, combinatorics, and algorithms, Vol. 1, 2 (Kalamazoo, MI, 1992)*, pages 277–287. Wiley, New York, 1995.

[15] S.N. Evans and J. Pitman. Construction of Markovian coalescents. *Ann. Inst. Henri Poincaré*, 34:339–383, 1998.

[16] W. Feller. *An Introduction to Probability Theory and its Applications,* Vol 1, 3rd ed. Wiley, New York, 1968.

[17] D. Foata and A. Fuchs. Réarrangements de fonctions et dénombrement. *J. Comb. Theory*, 8:361–375, 1970.

[18] J. Françon. Preuves combinatoires des identités d'Abel. *Discrete Mathematics*, 8:331–343, 1974.

[19] F. Göbel. Mutual Choices. Technical Report S 317 (VP 22), Mathematisch Centrum Amsterdam, Afdeling Mathematische Statistik, 1963.

[20] A. Hurwitz. Über Riemann'sche Flächen mit gegebenen Verzweigungspunkten. *Math. Annalen*, 39:1–66, 1891.

[21] A. Hurwitz. Über Abel's Verallgemeinerung der binomischen Formel. *Acta Math.*, 26:199–203, 1902.

[22] J.L.W. Jensen. Sur une identité d'abel et sur autres formules analogues. *Acta Math.*, 26:307–318, 1902.

[23] J.Jaworski. On a random mapping (T, P_j). *J. Appl. Probab.*, 21:186 – 191, 1984.

[24] A. Joyal. Une théorie combinatoire des séries formelles. *Adv. in Math.*, 42:1–82, 1981.

[25] L. Katz. Probability of indecomposability of a random mapping function. *Ann. Math. Statist.*, 26:512–517, 1955.

[26] A. Kelmans, I. Pak, and A. Postnikov. Tree and forest volumes of graphs. Technical Report 47-99, RUTCOR, 1999.

[27] A. Kelmans and A. Postnikov. On generalizations of Abel's and Hurwitz's identities. Technical Report 48-99, RUTCOR, 1999.

[28] A. K. Kelmans. The number of trees of a graph that contains a given forest. *Acta Math. Acad. Sci. Hungar.*, 27(1-2):89–95, 1976.

[29] A. K. Kelmans. Spanning trees of extended graphs. *Combinatorica*, 12(1):45–51, 1992.

[30] D. E. Knuth. Another enumeration of trees. *Can. J. Math.*, 20:1077–1086, 1968.

[31] D.E. Knuth. *The Art of Computer Programming*, volume 1. Addison-Wesley, 1968.

[32] D.E. Knuth. Discussion on Mr. Riordan's paper. In R.C. Bose and T.A. Dowling, editors, *Combinatorial Mathematics and its Applications*, pages 91–97. Univ. of North Carolina Press, Chapel Hill, 1969.

[33] A. Meir and J.W. Moon. The distance between points in random trees. *J. Comb. Theory*, 8:99–103, 1970.

[34] J. W. Moon. Enumerating labelled trees. In F. Harary, editor, *Graph Theory and Theoretical Physics*, pages 261–272. Academic Press, 1967.

[35] J. W. Moon. Two problems on random trees. In *Graph theory and applications (Proc. Conf., Western Michigan Univ., Kalamazoo, Mich., 1972; dedicated to the memory of J. W. T. Youngs)*, pages 197–206. Lecture Notes in Math., Vol. 303. Springer, Berlin, 1972.

[36] J.W. Moon. Various proofs of Cayley's formula for counting trees. In F. Harary, editor, *A Seminar on Graph Theory*, pages 70–78. Holt, Rineharte and Winston, New York, 1967.

[37] J.W. Moon. *Counting Labelled Trees*. Canadian Mathematical Congress, 1970. Canadian Mathematical Monographs No. 1.

[38] H. S. Na and A. Rapoport. A formula for the probability of obtaining a tree from a graph constructed randomly except for an "exogamous bias". *Ann. Math. Statist.*, 38:226–241, 1967.

[39] R. Pemantle. Uniform random spanning trees. In J. L. Snell, editor, *Topics in Contemporary Probability*, pages 1–54, Boca Raton, FL, 1995. CRC Press.

[40] J. Pitman. Abel-Cayley-Hurwitz multinomial expansions associated with random mappings, forests and subsets. Technical Report 498, Dept. Statistics, U.C. Berkeley, 1997.

[41] J. Pitman. The multinomial distribution on rooted labeled forests. Technical Report 499, Dept. Statistics, U.C. Berkeley, 1997.

[42] J. Pitman. Enumerations of trees and forests related to branching processes and random walks. In D. Aldous and J. Propp, editors, *Microsurveys in Discrete Probability*, number 41 in DIMACS Ser. Discrete Math. Theoret. Comp. Sci, pages 163–180, Providence RI, 1998. Amer. Math. Soc.

[43] J. Pitman. Coalescent random forests. *J. Comb. Theory A.*, 85:165–193, 1999.

[44] J. Pitman. Random mappings, forests and subsets associated with Abel-Cayley-Hurwitz multinomial expansions. Technical Report 593, Dept. Statistics, U.C. Berkeley, 2001.

[45] H. Prüfer. Neuer Beweis eines Satzes über Permutationen. *Archiv für Mathematik und Physik*, 27:142–144, 1918.

[46] A. Rényi. On connected graphs. I. *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, 4:385–388, 1959.

[47] A. Rényi. On the enumeration of trees. In R. Guy, H. Hanani, N. Sauer, and J. Schonheim, editors, *Combinatorial Structures and their Applications*, pages 355–360. Gordon and Breach, New York, 1970.

[48] J. Riordan. *Combinatorial Identities.* Wiley, New York, 1968.

[49] M. M. Robertson. A generalisation of a result of Abel with an application to tree enumerations. *Proc. Edinburgh Math. Soc. (2)*, 14:239–241, 1964/1965.

[50] L.W. Shapiro. Voting blocks, reluctant functions, and a formula of Hurwitz. *Discrete Mathematics*, 87:319–322, 1991.

[51] P.W. Shor. A new proof of Cayley's formula for counting labelled trees. *J. Combinatorial Theory A.*, 71:154–158, 1995.

[52] A.J. Stam. Two identities in the theory of polynomials of binomial type. *J. Math. Anal.*, 122:439–443, 1987.

[53] R. Stanley. *Enumerative Combinatorics, Vol. 1.* Cambridge University Press, 1997.

[54] R. Stanley. *Enumerative Combinatorics, Vol. 2*. Cambridge University Press, 1999.

[55] V. Strehl. Identities of the Rothe-Abel-Schläfli-Hurwitz-type. *Discrete Math.*, 99:321–340, 1992.

[56] V. Strehl. Minimal transitive products of transpositions: the reconstruction of a proof of A. Hurwitz. *Séminaire Lotharingien de Combinatoire*, B37c: 12 pp., 1996. Available at http://cartan.u-strasbg.fr/∼slc.

[57] L. Takács. Counting forests. *Discrete Mathematics*, 84:323–326, 1990.

[58] J. Zeng. Multinomial convolution polynomials. *Discrete Math.*, 160:219–228, 1996.