Europ. J. Combinatorics (1984) **5**, 373–376

Two Short Proofs of Kemp's Identity for Rooted Plane Trees

VOLKER STREHL

R. Kemp has recently presented a nice identity relating different sets of rooted plane trees numerically. Two short proofs of this result are given here, a 'bijective' one, and one making use of continued fraction generating functions—both avoiding explicit expressions for the numbers involved.

1. INTRODUCTION

In a recent talk, R. Kemp [5] presented the following result:

For positive integers n, k, p let
$b_{n,k,p}$:= the number of rooted plane trees with n nodes, height $\leqslant k$, and root-degree $= p$,
$q_{n,k,p}$:= the number of rooted plane trees with n nodes, height $= k$, and exactly p nodes of maximal height.
Then the following identity holds:

$$q_{n,k,p} = b_{n+1,k,p+1} - b_{n+1,k,p} + b_{n,k,p-1}.$$

Kemp derived this identity from rather explicit summation formulae for these numerical quantities. Since the identity is by no means obvious if one looks at the sets of trees involved, even for small values of the parameters, he asked for a direct combinatorial proof of it—without any use of numerical considerations.

It is the purpose of this note to present two proofs of this kind of Kemp's result. Though completely different in character, these two proofs enjoy a common feature: they are based on the combinatorial model of Dyck-words (equiv. properly parenthesized strings, ballot sequences, histories, ...), well-known to be a 'bijective' equivalent to rooted plane trees (see e.g. [1], [2], [4]). In this model Kemp's identity may be proved by a rather simple 'factorization and rearrangement' argument—this will be the first approach presented below. As an alternative, one may use the ideas developed and exploited by Flajolet ([2], [3]) to obtain generating functions for the numbers $b_{n,k,p}$ ($q_{n,k,p}$ resp.) in their continued fraction version. The close connection between the two (finite) continued fractions thus obtained again leads to a short proof of Kemp's identity—without determining the ordinary generating functions explicitly.

2. NOTATION

Let $A = \{a, b\}$, and let A^* denote the free monoid over A. For $w \in A^*$ $|w|$ denotes the length of w, $h(w)$ denotes the height of w (= the number of as in w minus the number of bs in w). For u, $w \in A^*$ let u/w denote the fact that u is a left factor of w of positive length. Define

$\bar{h}(w) := \max\{h(u); u/w\}$,

$d_0(w) := |\{u \in A^*; u/w \text{ and } h(u) = 0\}|$,

$d_m(w) := |\{u \in A^*; u/w \text{ and } h(u) = \bar{h}(w)\}|$.

0195-6698/84/040373+04 $02.00/0

The Dyck-language over A may then be written as

$$D^* = \{w \in A^*;\ u/w \Rightarrow h(u) \geq 0,\ h(w) = 0\},$$

which is the submonoid of A^* freely generated by $D = aD^*b$, so that $d_0(w)$ is just the number of D-factors of $w \in D^*$. The bijection which associates to each rooted plane tree t with $n+1$ nodes a word $w_t \in D^*$ of length $2n$ (via preorder traversal) is well-known and needs no restatement here. In view of this fact Kemp's identity is equivalent to:

PROPOSITION. *For positive integers n, k, p let*

$$B(n, k, p) := \{w \in D^*;\ |w| = 2n,\ \bar{h}(w) \leq k,\ d_0(w) = p\},$$

$$Q(n, k, p) := \{w \in D^*;\ |w| = 2n,\ \bar{h}(w) = k,\ d_m(w) = p\}.$$

Then $|Q(n, k, p)| = |B(n+1, k, p+1)| - |B(n+1, k, p)| + |B(n, k, p-1)|$.

3. First Proof of the Proposition

For $w \in D^*$ of positive length let w_0 denote the unique $u \in D$ s.th. u/w. For fixed n, k, p we introduce four auxiliary sets:

$$B_< := \{w \in B(n+1, k, p+1);\ \bar{h}(w_0) < k\},$$

$$B_= := \{w \in B(n+1, k, p+1);\ \bar{h}(w_0) = k\},$$

$$B_{aa} := \{w \in B(n+1, k, p);\ aa/w\},$$

$$B_{ab} := \{w \in B(n+1, k, p);\ ab/w\}.$$

The assertion then follows from the fact that we may construct three bijective mappings:

$$f_1 : B_< \to B_{aa},$$

$$f_2 : B_{ab} \to B(n, k, p-1),$$

$$f_3 : B_= \to Q(n, k, p).$$

f_1: If $w \in B_<$, then $w \neq w_0$ by definition; hence w may be written as $w = w_0 a v$. Define $f_1(w) := a w_0 v$, then $f_1(w)$ belongs to B_{aa}, and f_1 is easily seen to be a bijection.

f_2: For $w \in B_{ab}$ we write $w = abv$, and we define $f_2(w) := v$, which obviously meets our requirements.

f_3: Any $w \in B_=$ will be written as $w = auvb$, where $h(au) = k$, and where au is the shortest left factor of w with this property. Note that au/w_0 by definition of $B_=$. We now put $f_3(w) := u\tilde{v}$, where $\tilde{v}$ denotes the reverse of v.

Now consider the left factors $r/f_3(w)$:

(a) if r/u, then obviously $0 \leq h(r) \leq k-1$;

(b) if $r = us$ with $s/\tilde{v}$, then $0 \geq h(\tilde{s}b) \geq -k$ since $\tilde{s}b$ is a right factor of w; hence $1 \geq h(s) = h(s) \geq 1-k$, and $k \geq h(us) = h(u) + h(s) \geq 0$ follows from $h(u) = k-1$.

In particular: the left factors $r/f_3(w)$ with $h(r) = k$ correspond to the left factors $s/\tilde{v}$ with $h(b\tilde{s}) = 0$, i.e. to the $t/\tilde{w}$ with $h(t) = 0$ and $t \neq \tilde{w}$. Thus $d_0(w) - 1 = d_m(f_3(w))$.

This shows that f_3 maps $B_=$ into $Q(n, k, p)$, and the injectivity of this map should be evident from its definition. The surjectivity is easily checked: take any $w \in Q(n, k, p)$ and write it as $w = uv$, where u is the shortest z/w such that $h(z) = k-1$; then $au\tilde{v}b$ is the f_3-preimage of w in $B_=$.

As a concluding remark: the map f_3, which is easily visualized by drawing a diagram (see Figure 1), does not translate back into a 'nice' map of trees (via the standard

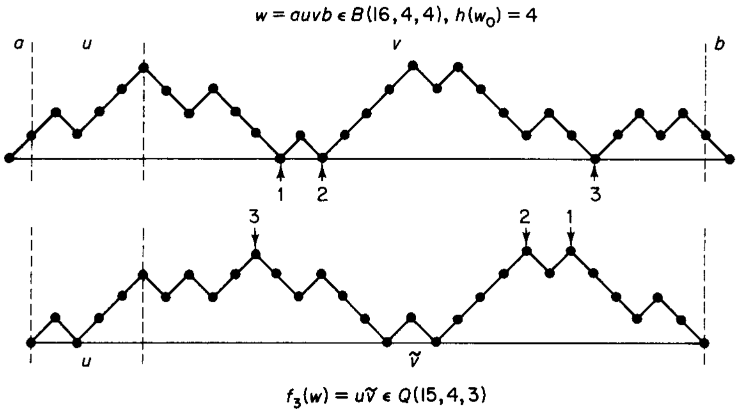

FIGURE 1.

bijections). Thus one may have difficulties to find a 'bijective' proof of Kemp's identity by looking at trees only.

4. Second Proof of the Proposition

We now proceed to give a second proof of Kemp's result, again avoiding any (more or less) explicit expression for the numbers $b_{n,k,p}$ and $q_{n,k,p}$. Consider the generating functions

$$b_k(x,t)=\sum_{n,p} b_{n,k,p}x^n t^p, \qquad q_k(x,t)=\sum_{n,p} q_{n,k,p}x^n t^p.$$

From the combinatorial model it follows (by the method of Flajolet) that these generating functions can be written as continued fractions:

$$b_k(x,t)=\left.\cfrac{1}{1-\cfrac{xt}{1-\cfrac{x}{1-\cfrac{x}{\begin{matrix}\vdots\\ \overline{1-x}\end{matrix}}}}}\right\}k, \qquad q_k(x,t)=\left.\cfrac{1}{1-\cfrac{x}{1-\cfrac{x}{1-\cfrac{x}{\begin{matrix}\vdots\\ \cfrac{1-x}{1-xt}\end{matrix}}}}}\right\}k.$$

Thus $q_k(x,t)$ and $b_k(x,t)$ are rational functions in x, which can be written as

$$b_k(x,t)=\frac{a_{k-1}(x,1)}{a_k(x,t)}, \qquad q_k(x,t)=\frac{a_{k-1}(x,t)}{a_k(x,t)},$$

where the sequence $(a_k(x,t))_{k\geq 0}$ is recursively defined by

$$a_0(x,t)=1, \tag{1}$$

$$a_1(x,t)=1-xt, \tag{2}$$

$$a_k(x,t)=a_{k-1}(x,t)-x\cdot a_{k-2}(x,t), \qquad k\geq 2. \tag{3}$$

For $k=1,2$ it can be immediately checked that

$$a_k(x,t)-t\cdot a_{k+1}(x,t)=(1-t+xt^2)\cdot a_{k-1}(x,1), \tag{4}$$

thus (4) holds for all $k \geq 1$, since $(a_k(x, t))_{k\geq 0}$ and $(a_k(x, 1))_{k\geq 0}$ satisfy the same recurrence (3), the only difference in their definition scheme being initial condition (2). But from (3) we have:

$$a_k(x, t) - t \cdot a_{k+1}(x, t) = a_k(x, t) - t \cdot (a_k(x, t) - x \cdot a_{k-1}(x, t))$$
$$= (1 - t) \cdot a_k(x, t) + xt \cdot a_{k-1}(x, t),$$

and thus

$$(1 - t) \cdot a_k(x, t) + xt \cdot a_{k-1}(x, t) = (1 - t + xt^2) \cdot a_{k-1}(x, 1).$$

Division by $a_k(x, t)$ leads to

$$1 - t + xt \cdot q_k(x, t) = (1 - t + xt^2) \cdot b_k(x, t),$$

from which Kemp's identity follows by comparing coefficients.

REMARK. The calculation of explicit expressions for the $b_{n,k,p}$ and $q_{n,k,p}$ could now be undertaken, following e.g. the guidelines of deBruijn, Knuth and Rice [1], and using

$$\sum_k a_k(x, t) z^k = \frac{1 - txz}{1 - z(1 - xz)}.$$

REFERENCES

1. N. G. de Bruijn, D. E. Knuth and S. O. Rice, The average height of planted plane trees, in *Graph Theory and Computing* (R. C. Read, ed.), Academic Press, New York, 1972, pp. 15–22.
2. P. Flajolet, Analyse d'algorithmes de manipulation d'arbres et de fichiers, Thése, Paris-Orsay, 1979; published as *Cahiers du Bureau Universitaire de Recherche Operationnelle* **38–39** (1981), 1-209.
3. P. Flajolet, Combinatorial aspects of continued fractions, *Discrete Math.* **32** (1980), 125–161.
4. J. Goldman, Formal languages and enumeration, *J. Combin. Theory, Ser. A* **24** (1978), 318–338.
5. R. Kemp, Erwartungswerte charakteristischer Grössen in geordneten Bäumen, Vortrag auf der Jahrestagung der DMV, Bayreuth, September, 1982.

Received 25 April 1983

VOLKER STREHL
Institut für Mathematische Maschinen und Datenverarbeitung der Universität Erlangen-Nürnberg, Martensstrasse 3, D-8520 Erlangen, F.R.G.