Europ. J. Combinatorics (1990) **11**, 35–40

The Cycle Lemma and Some Applications

NACHUM DERSHOWITZ AND SHMUEL ZAKS

Two proofs of a frequently rediscovered combinatorial lemma are presented. Using the lemma, a combinatorial proof is given that the average height of an ordered (plane-planted) tree is approximately twice the average node (vertex) level.

1. THE CYCLE LEMMA

A sequence $p_1 p_2 \cdots p_l$ of boxes and circles is called *k-dominating* (for positive integer k) if for every position i, $1 \leqslant i \leqslant l$, the number of boxes in $p_1 p_2 \cdots p_i$ is more than k times the number of circles. For example, the sequence □□□□□○○□□○ is 2-dominating; the sequence □□□□○○□□○ is 1-dominating (or just *dominating*) but not 2-dominating; the sequences ○□□□□○○□□ and □□○○□□○□□ are not even 1-dominating.

The following lemma has been rediscovered many times. Although not difficult to prove, it is a powerful tool in enumeration arguments.

CYCLE LEMMA (Dvoretzky and Motzkin [9]). *For any sequence $p_1 p_2 \cdots p_{m+n}$ of m boxes and n circles, $m \geqslant kn$, there exist exactly $m - kn$ (out of $m + n$) cyclic permutations $p_j p_{j+1} \cdots p_{m+n} p_1 \cdots p_{j-1}$, $1 \leqslant j \leqslant m + n$, that are k-dominating.*

For example, of the nine cyclic permutations of the sequence ○□□□□○○□□ of six boxes and three circles, only three are dominating; □□○□□□□○○, □□□□○○□□○ and □□□○○□□○□. None are 2-dominating. As a special case of this lemma, if $m = n + 1$, then there is a unique dominating permutation.

In Section 2, we will present two applications of this lemma, one from each of the points of view taken in the following two proofs. Our first proof is a generalization of proofs appearing in Silberger [23], Bergman [2], and Singmaster [24]; our second proof follows Grossman [12], Raney [21], and Yaglom and Yaglom [28].

1.1. FIRST PROOF. For the first proof of the lemma, arrange the $m + n$ figures on a cycle. Removing a subsequence of k boxes followed by one circle from the cycle does not change the number of k-dominating permutations, since the $k + 1$ figures have no net effect and no k-dominating permutation could have begun with any of the deleted figures. By the 'pigeon-hole principle,' as long as $m \geqslant kn > 0$, there must be such a subsequence on the cycle; these subsequences may be removed one by one until only boxes remain. The remaining $m - kn$ boxes yield $m - kn$ k-dominating cyclic permutations.

Example. Consider the sequence ○□□□□○○□□, with $k = 1$. Placing them on a cycle and removing three pairs, leaves three boxes, corresponding to the three dominating cyclic permutations □□□□○○□□○, □□□○○□□○□, and □□○○□□○□□ (see Figure 1).

Note that not all cyclic permutations are necessarily distinct; rather, there are duplicates to the extent that there is periodicity in the cycle. Still, the proportion of

0195-6698/90/010035 + 06 $02.00/0

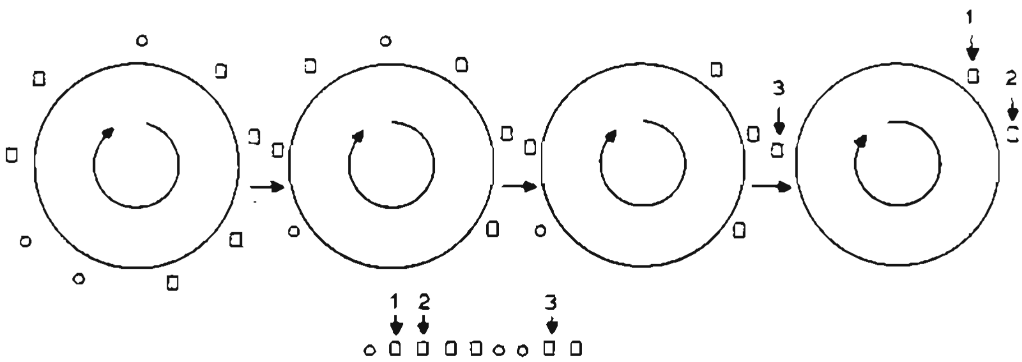

FIGURE 1. First proof ($k = 1$).

distinct k-dominating cyclic permutations among all the distinct cyclic permutations is the same as the proportion of k-dominating ones among all of them. For example, one fourth of the 16 cyclic permutations of □□□○○□□○□□□○○□□○ are dominating, as are one fourth of the eight distinct ones.

1.2. SECOND PROOF. Another simple proof is the following. Given a sequence of figures, construct a 'mountain range' (lattice path). Begin to the left at 'sea level.' For each box, draw a straight line to a point one unit to the right and one unit upwards; for each circle, slope downwards k units and right one unit. The resultant graph extends $m + n$ units to the right and ends $m - kn$ units above sea level.

By cyclically permuting the sequence, the origin is moved to a different point along the range. A k-dominating sequence corresponds to a range that is completely (all but the origin) above sea level. Choose as a new, valid origin any point:

(1) for which there is no equally low (or lower) valley to its right (otherwise, that valley would end up at (or below) sea level); and
(2) which is less than $m - kn$ units above the deepest valley (otherwise, that valley would descend to (or past) sea level).

Any point that was to the right of the new origin was higher and is therefore above sea level now; any point that was to its left was less than $m - kn$ units lower and is therefore above sea level now.

Clearly, there are exactly $m - kn$ such points to choose a valid origin from.

Example. Consider the sequence ○□□□□□○○□□□, with $k = 2$. Constructing the corresponding 'mountain range,' shows two possible starting points, corresponding to the two 2-dominating cyclic permutations □□□□□○○□□□○ and □□□○□□□□□○○ (see Figure 2).

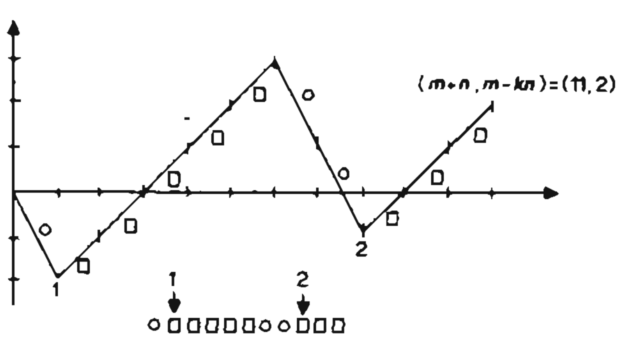

FIGURE 2. Second proof ($k = 2$).

1.3. Other Proofs. The Cycle Lemma is the combinatorial analogue of the Lagrange inversion formula; see Raney [21], Cori [4] and Gessel [11]. Other proofs of varying degree of generality may be found in Dvoretzky and Motzkin [9] (discussed in Grossman [12]), Motzkin [19] (two proofs), Hall [14], Raney [21], Yaglom and Yaglom [28], Takács [26], Silberger [23], Bergman [2] (three proofs), Sands [22] and Singmaster [24]. (The first paper [9] is not credited by the other authors, but is referenced in Barton and Mallows [1] and Mohanty [18].) Dvoretzky and Motzkin, Motzkin, and Yaglom and Yaglom give the lemma in its general form; the other papers prove only the case $k = 1$ or $m - kn = 1$. Generalizations of the Cycle Lemma to non-integer k and sequences of reals may be found in Dvoretzky and Motzkin [9] and Spitzer [25], respectively.

2. Applications

We demonstrate the power of the Cycle Lemma with two applications. The first is an enumeration of forests of trees with nodes of fixed degree; the second is an approximation of the height of trees with nodes of arbitrary degree.

2.1. Forests. The number of (ordered) forests containing s trees with n *internal* nodes of (out-) degree t and $tn + s - n$ *leaves* of degree 0 (*t-ary trees*) is

$$\frac{s}{tn+s}\binom{tn+s}{n}.$$

To see this, note the correspondence between forests of t-ary trees and $(t-1)$-dominating (postfix Polish) sequences obtained by traversing the trees in *postorder*, i.e. first each subtree from left to right is traversed and then the node connecting them, and recording a circle for each internal node encountered and a box for each leaf encountered. By the Cycle Lemma, $s/(tn+s)$ of the cyclic permutations of the $\binom{tn+s}{n}$ sequences of n circles and $tn + s - n$ boxes are $(t-1)$-dominating.

Limiting the forest to one tree ($s = 1$), gives

$$\frac{1}{tn+1}\binom{tn+1}{n},$$

the total number of t-ary trees with n internal nodes (see Klarner [15] and Knuth [16]; Grunert [13] gives the analogous result for polygons). In particular, the number of binary trees ($t = 2$) is

$$\frac{1}{2n+1}\binom{2n+1}{n} = \frac{1}{n+1}\binom{2n}{n},$$

the well known Catalan numbers (see Cayley [3]; Silberger [23], Sands [22] and Singmaster [24] also derive the Catalan sequence using the Cycle Lemma).

2.2. Trees. In an *ordered* (a.k.a. *plane-planted*) tree the order in which subtrees of a node are arranged is significant, but the number of outgoing edges is not fixed. Every ordered tree with n edges corresponds to a mountain range of $n + 1$ upward-sloping steps and n downward-sloping ones, obtained by slicing each edge of the tree lengthwise and pulling the tree apart at the root and adding one extra upward step at the start (see Figures 3(a) and 3(b)). Thus, every tree corresponds to a mountain range starting at sea level and ending $2n + 1$ units to the right (of the origin) and one unit up.

The *level* of a node in an ordered tree is the length of the path from the root to the node; the *height* of an ordered tree is the length of the longest path from the root to a

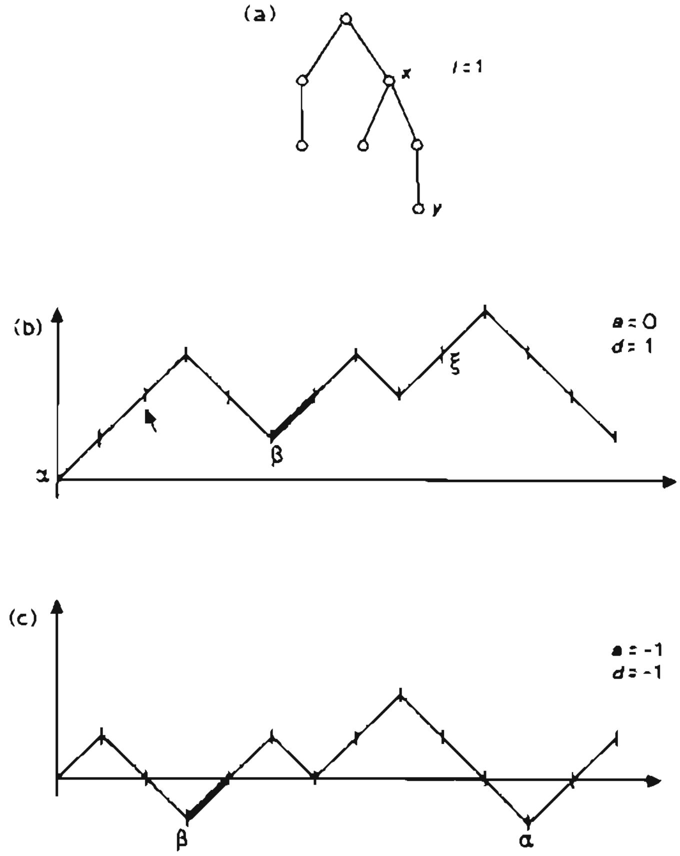

FIGURE 3. (a) A tree T. (b) A mountain range r corresponding to T. (c) A cyclic shift of r.

leaf. The level of a node in a tree corresponds to the elevation (measured from sea level) of the corresponding step in the range, measured at the bottom of the up step corresponding to the incoming edge; the height corresponds to the maximum elevation of a step in the range. Let $\bar{l}$ denote the expected level of a node in an n-edge ordered tree, and let $\bar{h}$ denote the expected height, with all ordered trees equiprobable. Applying the lemma to the representation of trees as mountain ranges, we can show that

$$2\bar{l} - 2 \leq \bar{h} \leq 2\bar{l} + 1. \qquad (*)$$

Using the fact that

$$\bar{l} = \frac{2^{2n-1}}{\binom{2n}{n}} - \frac{1}{2} \approx \frac{\sqrt{\pi n}}{2} - \frac{1}{2}$$

(Volosin [27], Meir and Moon [17] and Dasarathy and Yang [5]; see Dershowitz and Zaks [7] for a proof using the Cycle Lemma), it follows that

$$\bar{h} \approx \sqrt{\pi n}.$$

deBruijn, Knuth and Rice [6] give an analytic proof of the asymptotic value of $\bar{h}$; a more general analytic proof may be found in Flajolet and Odlyzko [10].

To prove (*), we make use of the Cycle Lemma in order to estimate the height of a tree, using all the cyclic ranges corresponding to that tree. We first note that, by the Cycle Lemma, every tree can be said to correspond to $2n + 1$ (distinct) mountain ranges, one for each of the (distinct) cyclic permutations of the range. (An upward step corresponds to a box and a downward one to a circle, as in the second proof of the lemma.)

Let x be a node at level l of a given tree T of height h, and let r be any of the $2n+1$ ranges corresponding to T. The level l of x satisfies

$$d-a \leqslant l \leqslant d-a+1, \tag{**}$$

where a is the minimum (signed) elevation of a step along r, and d is the elevation of the step corresponding to x in r.

Similarly, the height h of T satisfies

$$z-a \leqslant h \leqslant z-a+1, \tag{***}$$

where z is the maximum elevation of a step along r.

This is because a cyclic permutation can bring the step in question one level closer to the lowest step, on account of the disparity between the starting and ending elevations of the range.

By considerations of symmetry (the reflection—with respect to the sea—of any range must also be among the $\binom{2n+1}{n}$ ranges), the average values $\bar{z}$ and $-\bar{a}$ (over all ranges) of z and $-a$ are the same, and the average value of d (ranging over all steps in all ranges) is 0. Combining all the above, we obtain

$$\bar{z}=0-\bar{a} \leqslant \bar{l} \leqslant 0-\bar{a}+1 \leqslant \bar{z}+1$$

and

$$2\bar{z}=\bar{z}-\bar{a} \leqslant \bar{h} \leqslant \bar{z}-\bar{a}+1=2\bar{z}+1$$

from which (*) follows immediately.

Example. A tree T is depicted in Figure 3(a); x is a node in T. The range r corresponding to T is shown in Figure 3(b). The step in r corresponding to x is marked with a heavy line. The starting point of r is marked α, and the bottom of the step corresponding to x is marked β. $d-a=l$. The equality holds in every cyclic shift in which α is to the left of β. On the other hand if, in a cyclic shift of r, α is to the right of β, then $d-a$ is decreased by one, and therefore we have $l=d-a+1$. (See Figures 3(b) and 3(c); the point where the cyclic shift starts is marked with an arrow in Figure 3(b).) The same argument is used for (***), using ξ instead of β, where ξ is the bottom of the step corresponding to the lowest node y in T (see Figures 3(a) and 3(b)).

Acknowledgment

This research was supported in part by the National Science Foundation under Grant DCR 85-13417.

References

1. D. E. Barton and C. L. Mallows, Some aspects of the random sequence, *Ann. Math. Stat.* **36** (1965), 236–260.
2. G. M. Bergman, Terms and cyclic permutations, *Algebra Universalis* **8** (1978), pp. 129–130.
3. A. Cayley, On analytical forms called trees, *Phil. Mag.* **20** (1859), 374–378; also in *Collected Math. Papers* **4** (1859), 112–115.
4. R. Cori, in: *Combinatorics on Words* (M. Lothaire, ed.), Encyclopedia of Mathematics and Its Applications, vol. 17, Addison-Wesley, Reading, Massachusetts, 1983.
5. B. Dasarathy and C. Yang, A transformation on ordered trees, *Comput. J.* **23**(2) (1980), 161–164.
6. N. deBruijn, D. E. Knuth and 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.
7. N. Dershowitz and S. Zaks, Applied tree enumerations, in: *Proc. 6th Colloq. on Trees in Algebra and Programming*, Genoa, Italy, Lecture Notes in Computer Science, vol. 112, Springer-Verlag, Berlin, (1981) pp. 180–193.

8. N. Dershowitz, and S. Zaks, Patterns in trees, *Discr. Appl. Math.*, to appear.
9. A. Dvoretzky and Th. Motzkin, A problem of arrangements, *Duke Math. J.* **14** (1947), 305–313.
10. Ph. Flajolet and A. Odlyzko, the average height of binary trees and other simple trees, *J. Comput. System Sci.*, **25** (1980), 171–213.
11. I. M. Gessel, A combinatorial proof of the multivariate Lagrange inversion formula, *J. Combin. Theory Ser. A* **45** (1987), 178–195.
12. H. D. Grossman, Fun with lattice points—21, *Scripta Math.* **16** (1950), 120–124.
13. J. A. Grunert, Über die Bestimmung de Anzahl der verschiedenen Arten, auf welche sich ein n-eck durch Diagonalen in lauter m-ecke zerlegen lässt, *Arch. Math. Phys.* **1** (1841), 193–203.
14. P. Hall, Some word problems, *J. London Math. Soc.* **33** (1958), 482–496.
15. D. A. Klarner, Correspondences between plane trees and binary sequences, *J. Combin. Theory* **9** (1970), 401–411.
16. D. E. Knuth, The Art of Computer Programming 1: Fundamental Algorithms (Addison-Wesley, Reading, MA, 2nd ed., 1973).
17. A. Meir and J. Moon, On the altitude of nodes in random trees, *Can. J. Math.* **30**(5) (1978), 997–1015.
18. S. G. Mohanty, *Lattice Path Counting and Applications*, Academic Press, New York, 1979.
19. Th. Motzkin, Relations between hypersurface cross ratios and a combinatorial formula for partitions of a polygon, for permanent preponderance, and for non-associative products, *Bull. Am. Math. Soc.* **54** (1948), 352–360.
20. T. V. Narayana, A partial order and its application to probability, *Sankhya*, **21** (1959), 91–98.
21. G. M. Raney, Functional composition patterns and power series reversion, *Trans. Am. Math. Soc.* **94** (1960), 441–451.
22. A. D. Sands, On generalized Catalan numbers, *Discr. Math.* **21**(2) (1978), 219–221.
23. D. M. Silberger, Occurrences of the integer $(2n-2)!/n!(n-1)!$, *Roczniki Polskiego Towarzystwa Math. I* **13** (1969), 91–96.
24. D. Singmaster, An elementary evaluation of the Catalan numbers, Am. Math. Monthly **85** (1978), 366–368.
25. F. Spitzer, A combinatorial lemma and its applications to probability theory, *Trans. Am. Math. Soc.* **82** (1956), 323–339.
26. L. Takács, *Combinatorial Methods in the Theory of Stochastic Processes*, John Wiley, New York, 1967.
27. Ju.M. Volosin, Enumeration of the terms of object domains according to the depth of embedding, *Sov. Math. Dokl.* **15** (1974), 1777–1782.
28. A. M. Yaglom and I. M. Yaglom, *Challenging Mathematical Problems with Elementary Solutions*, vol. 1: *Combinatorial Analysis and Probability Theory*, Holden Day, San Francisco, 1964.

Received 25 October 1984 and accepted in revised form 7 June 1989

NACHUM DERSHOWITZ
Department of Computer Science,
University of Illinois at Urbana-Champaign,
Urbana, Illinois 61801, U.S.A.
and
SHMUEL ZAKS
Department of Computer Science,
Technion, Haifa 32000, Israel