PU. M. A. Vol. **22** (2011), No.2, pp. 99 – 105

Non-overlapping permutation patterns

Miklós Bóna
Department of Mathematics
University of Florida
358 Little Hall, PO Box 118105
Gainesville, FL 32611–8105 (USA)
e-mail: bona@ufl.edu

(Received: February 28, 2011, and in revised form May 1, 2011)

Abstract. We show a way to compute, to a high level of precision, the probability that a randomly selected permutation of length n is non-overlapping. As a byproduct, we find some combinatorial identities that are routine to prove using generating functions, but difficult to prove bijectively.

Mathematics Subject Classification(2010). 05A16, 05A05, 05A10.
Keywords: permutation, pattern, overlap, asymptotics.

To Doron Zeilberger, for his Sixtieth Birthday

1 Introduction

Let us say that the permutation $p = p_1p_2\cdots p_n$ *tightly* contains the permutation $q = q_1q_2\cdots q_k$ if there exists an index $0 \leq i \leq n-k$ so that $p_{i+j} < p_{i+r}$ if and only if $q_i < q_j$. In other words, p tightly contains q if there is a string of k entries in p in *consecutive* positions which relate to each other as the entries of q do.

If p does not tightly contain q, then we say that p *tightly avoids* q. Let $T_n(q)$ denote the number of n-permutations that tightly avoid q. For instance, 1436725 tightly contains 123 (consider the third, fourth, and fifth entries), but tightly avoids 321 and 4231. An intriguing conjecture of Elizalde and Noy [3] from 2001 is the following.

Conjecture 1.1 Let q be any pattern of length k. Then $T_n(q) \leq T_n(12\cdots k)$, and equality holds only if $q = 12\cdots k$ or $q = k(k-1)\cdots 1$.

A permutation $q = q_1q_2\cdots q_k$ is called *non-overlapping* if there is no permutation $p = p_1p_2\cdots p_n$ so that both $p_1p_2\cdots p_k$ and $p_{n-k+1}p_{n-k+2}\cdots p_n$ form a q-pattern, and k satisfies $k < n < 2k-1$. For instance, $q = 132$ is non-overlapping, but $q' = 2143$ is not since $p = 214365$ has the property that both its first four entries and its last four entries form a 2143-pattern. In other words, a permutation is called non-overlapping if it is impossible for two of its copies to overlap in more than one entry. Equivalently, q is non-overlapping if there is no j so that $2 \leq j \leq k-1$ and the pattern of the first j entries of q is identical to the pattern of the last j entries of q.

Non-overlapping patterns have recently been the subject of vigorous research. See [2] for an overview of these results. In particular, both numerical evidence and intuition suggests that non-overlapping patterns should be the ones for which Conjecture 1.1 is the easiest to prove. Indeed, the *total* number of tight copies of a pattern q of length k in all $n!$ permutations of length n is

$$\binom{n}{k}^2 (n-k)! = n!\frac{n!}{k!}.$$

Crucially, this number does not depend on q. In other words, no matter what q is (as long as its length is k), the set of all $n!$ permutations of length n must contain the same total number of tight copies of q. If q is non-overlapping, then it should be difficult to pack many tight copies of q into one permutation, so there should be many permutations that contain *some* tight copies of q, and hence, there should be not so many permutations that tightly avoid q. So $T_n(q)$ should be small for non-overlapping patterns.

This motivates the *enumeration* of non-overlapping patterns. If we can prove Conjecture 1.1 for such patterns, for how large a portion of all patterns will the conjeture be proved?

2 A basic lower bound

Even a rather crude argument shows that a reasonably high portion of all permutations is non-overlapping. Indeed, if p is overlapping, then for some $i \geq 2$, the pattern of the first i entries and the pattern of the last i entries is identical. Let F_i be the event that this happens. Clearly $P(F_i) = \frac{1}{i!}$, since there are $i!$ favorable outcomes and $i!^2$ possible outcomes as far as the pattern of the first i entries and the pattern of the last i entries is concerned.

Let n be an even positive integer. Then the probability that a randomly selected permutation p of length n is overlapping is

$$P\left(\bigcup_{i\geq 2}^{n/2} F_i\right) \leq \sum_{i\geq 2} P(F_i) = \sum_{i\geq 2} \frac{1}{i!} = e-2 \approx 0.718.$$

So the probability that p is non-overlapping is at least $1-(e-2) = 3-e \approx 0.282$.

3 Monotonicity

For $n \geq 2$, let a_n be the probability that a randomly selected n-permutation is non-overlapping. The simple argument of the previous section shows that $a_n \geq 3-e$ for all n. In this section we prove that the sequence $a_2, a_3, \cdots$ of positive real numbers is strictly monotone decreasing, hence it has a limit.

It is routine to verify that $a_2 = 1$, $a_3 = 2/3$, $a_4 = 1/2$, $a_5 = 2/5$, and $a_6 = 7/18$. Furthermore, for *even* values of n, the following simple recurrence relation holds.

Lemma 3.1 *Let n be an even positive integer. Then we have*

$$a_n = 1 - \sum_{j=2}^{n/2} \frac{a_j}{j!}. \tag{1}$$

Proof. If p is overlapping, then there is a unique smallest index j so that $2 \le j \le n/2$ and the pattern q of the first j entries of p agrees with the pattern q' of the last j entries of p. Note that because of the minimality of j, the pattern q, and hence the pattern q' are non-overlapping. Indeed, if for some $i \in [2, j-1]$, the pattern r of the first i entries of q agreed with the pattern of the last q entries of q (and hence, of q'), then the patterns of the first and last i entries of p would both be r, contradicting the minimality of j.

Moreover, the minimal index j discussed in the last paragraph cannot be more than $n/2$, since then q and q' would intersect in $h \ge 2$ entries, meaning that the patterns of the first h and last h entries of q were identical, contradicting the minimality of j.

For a fixed index j, the probability that the pattern q of the first j entries of a a random permutation p of length n is the same non-overlapping pattern as the pattern q' of the last j entries of p is $\frac{a_j j!}{j!^2} = \frac{a_j}{j!}$. Indeed, there are $j!$ possible outcomes for each of q and q', and $a_j j!$ of them are favorable. □

Recalling that the values of a_m are easy to obtain by hand for $m \le 6$, Formula (1) allows us to compute the values of a_n if $n \le 12$ is an even number. We get $a_8 = \frac{53}{144}$, $a_{10} = \frac{1313}{3600}$, and $a_{12} = \frac{23599}{64800}$.

For *odd* values of n, the situation is more complicated since there are permutations of length $n = 2k+1$ that are overlapping because the pattern of their first $k+1$ entries and the pattern of their last $k+1$ entries are identical, while the pattern of their first j entries and last j entries is not identical for any j satisfying $1 < j < k+1$. Let us call such permutations *barely overlapping*. An example is the permutation $p = 13254$. The first three and the last three entries of this permutation both form a 132-pattern, but the first two form a 12-pattern, and the last two form a 21-pattern.

For odd n, let b_n be the probability that a randomly selected permutation of length n is barely overlapping. It is easy to verify that $b_3 = 1/3$, and $b_5 = 1/10$. We then have the following recurrence relation.

Corollary 3.2 *Let $n > 1$ be an odd integer. Then we have*

$$a_n = 1 - b_n - \sum_{j=2}^{n/2} \frac{a_j}{j!} = a_{n-1} - b_n. \tag{2}$$

With a little work, one can compute by hand that $b_7 = 88/7! = 11/630$, so (2) yields

$$a_7 = a_6 - b_7 = \frac{7}{18} - \frac{11}{630} = \frac{13}{35}.$$

This allows the computation of the exact values of a_{14} and a_{16}.

Comparing Lemma 3.1 and Corollary 3.2, it is obvious that $a_{2k+1} \le a_{2k}$, and in fact it is straightforward to prove that the inequality is strict, since $b_n > 0$ for $n \ge 3$.

However, it is not obvious that $a_{2k-1} \le a_{2k}$ also holds for all k. It follows Lemma 3.1 and Corollary 3.2 that this inequality is equivalent to

$$b_{2k-1} \le \frac{a_k}{k!}. \tag{3}$$

Inequality (3) is not obvious since neither the numbers a_n nor the numbers b_n are easy to determine. In fact, even if we disregard the requirements related to the non-overlapping property, the equality corresponding to (3) is not a trivial one. The question then becomes the following. What is more likely, that the patterns of the first k and last k entries of a permutation of length $2k$ are identical, or that the patterns of the first k and last k entries of a permutation of length $2k-1$ are identical? The former clearly has probability $1/k!$, but the probability of the latter takes some work to obtain. This is the content of the next lemma.

LEMMA 3.3 *Let d_k be the number of permutations of length $2k-1$ in which the pattern of the first k entries is identical to the pattern of the last k entries. Then for $k \geq 2$, we have*

$$d_k = (k-2)! \cdot \left((2k-1)\binom{2k-2}{k-1} - 4^{k-1} \right).$$

Proof. Let p be a permutation counted by d_k. Let the first k entries of p be called *front entries*, and let the last k entries of p be called *back entries*. The kth entry of p, which is both a front and back entry, is also called the *middle entry*.

Clearly, if we know the set of front entries of p, and the middle entry m of p, then we also know the set of back entries of p, and we then have $(k-2)!$ possible candidates for p itself. Indeed, if m is the ith smallest front entry, then the rightmost entry of p is the mth smallest back entry. Similarly, if m is the ith smallest back entry, then the leftmost entry of p is the ith smallest front entry. There are $(k-2)!$ ways to permute the remaining $k-2$ front entries, and then the pattern of the $k-2$ remaining back entries is uniquely determined.

Therefore, the claim of the Lemma will be proved if we can show that there are $(2k-1)\binom{2k-2}{k-1} - 4^{k-1}$ ways to select the set of F front entries of p and the middle element m of p. There are clearly $2k-1$ ways to select an entry from the set $[2k-1] = \{1, 2, \cdots, 2k-1\}$ for the role of m, and then there are $\binom{2k-2}{k-1}$ ways to select the remaining $2k-2$ front entries. This leads to $(2k-1)\binom{2k-2}{k-1}$ choices for the ordered pair (m, F), but some of these choices are *invalid*, that is, they will never occur as the middle entry and the set of front entries for a permutation p counted by d_k.

Indeed, note the following. Given m and F, the relative rank of m in F determines the relative rank of m among the *back entries* as well. Let us say that m is the ith smallest front entry and the jth smallest back entry. We have explained two paragraphs earlier how this determines the leftmost and rightmost entries of p. However, that argument breaks down if $i = j$. Indeed, that would mean that the pattern q of the front entries (equivalently, back entries) of p would both start and end with its ith smallest entry, which is obviously impossible.

Observe that if m is an even number, then there are an odd number of entries of p that are less than m, so m cannot simultaneously be the ith smallest front entry and the ith smallest back entry. So if m is even, then no pair (m, F) is invalid. However, when $m = 2i+1$, then there are $\binom{2i}{i}\binom{2(k-1-i)}{k-1-i}$ invalid choices for F. Indeed, there are $\binom{2i}{i}$ ways to split the set of entries less than m evenly between the front and the back of p, and then there are $\binom{2(k-1-i)}{k-1-i}$ ways to split the set of entries larger than m evenly between the front and back of p. The pairs (m, F) obtained this way are precisely the invalid pairs.

Summing over $i = 0, 1, \cdots, k-1$, we get that the total number of choices for the ordered pair (m, F) that result in an invalid pair is

$$\sum_{0 \leq i \leq k-1} \binom{2i}{i}\binom{2(k-1-i)}{k-1-i} = 4^{k-1}. \tag{4}$$

Note that the fact that the left-hand side of (4) is equal to the closed expression of the right-hand side is not easy to prove combinatorially. On the other hand, a proof using generating functions is immediate, since both sides are the equal to the coefficient of x^{k-1} in

$$\frac{1}{\sqrt{1-4x}} \cdot \frac{1}{\sqrt{1-4x}} = \frac{1}{1-4x}.$$

The interested reader should consult Exercise 2.c. of [5], where the history of the combinatorial proofs of (4) is explained. $\square$

As an example, the formula of Lemma 3.3 says that $d_3 = 1 \cdot (3 \cdot 2 - 4) = 2$, and indeed, there are two permutations of length three in which the pattern of the first two entries is the same as the pattern of the last two entries, namely 123 and 321.

LEMMA 3.4 *For all $k \geq 2$, the inequality*

$$b_{2k-1} < \frac{a_k}{k!}$$

holds.

Proof. It follows directly from the definitions that $b_{2k-1} \leq \frac{d_k}{(2k-1)!}$, since the set enumerated by $(2k-1)!b_{2k-1}$ is a subset of the set enumerated by d_k as the latter has no non-overlapping requirements.

Therefore, it suffices to show that $\frac{d_k}{(2k-1)!} < \frac{a_k}{k!}$. For $k = 2$, we have $d_k = 2$ and $a_2 = 1$, so the inequality holds.

If $k \geq 3$, then note that Lemma 3.3 provides an exact formula for d_k, and the basic lower bound proved in Section 2 implies that $a_k > 1/4$. Therefore, it suffices to show that

$$\frac{(k-2)!}{(2k-1)!} \cdot \left((2k-1)\binom{2k-2}{k-1} - 4^{k-1} \right) < \frac{1}{4},$$

or, equivalently,

$$\frac{1}{(k-1)!(k-1)} - \frac{4^{k-1} \cdot (k-2)!}{(2k-1)!} < \frac{1}{4}.$$

The last displayed inequality is clearly true if $k \geq 3$, since in that case the first term of the left-hand side is at most $1/4$. $\square$

It is clear that $b_{2k-1} > 0$ for $k \geq 2$. Therefore, Lemma 3.1, Corollary 3.2, and Lemma 3.4 together immediately imply the main result of this section.

THEOREM 3.5 *The sequence $a_2, a_3, \cdots$ is strictly monotone decreasing.*

4 Bounds

Theorem 3.5 shows that the sequence $a_2, a_3, \cdots$ is strictly monotone decreasing. As it is a sequence of positive real numbers, it follows that it has a limit L. We did not succeed in giving an explicit and exact formula for this L. However, even simple methods result in a good approximation of L.

First, as the sequence of the a_i is strictly monotone decreasing, $L < a_n$ for all $n \geq 2$. In particular, setting $n = 2j$, this and Lemma 3.1 imply that

$$L < a_{2j} = 1 - \sum_{i=2}^{j} \frac{a_i}{i!}.$$

For instance, setting $j = 8$, we get that

$$L < 1 - \frac{1}{2} - \frac{2}{18} - \frac{1}{48} - \frac{2}{600} - \frac{7}{12960} - \frac{13}{35 \cdot 7!} - \frac{53}{144 \cdot 8!} = 0.3640992743.$$

On the other hand, note that

$$\begin{aligned} L &= \lim_{m\to\infty} a_{2m} \\ &= 1 - \lim_{m\to\infty} \sum_{j=2}^{m} \frac{a_j}{j!} \\ &= 1 - \sum_{j=2}^{\infty} \frac{a_j}{j!} \end{aligned}$$

The infinite sum of the last line can be bounded from above by replacing a_j by a_v for all $j > v$. of j. In the practice, this means that we leave a_j unchanged for all values of j for which a_j is known, and change it to a_v for all other values. As the infinite sum occurs with a negative sign, this yields the lower bound

$$L > 1 - \sum_{j=2}^{v} \frac{a_j}{j!} - a_v \sum_{j\geq v+1} \frac{1}{j!} = a_{2v} - a_v \left(e - \sum_{j=0}^{v} \frac{1}{j!} \right). \tag{5}$$

It goes without saying that the larger v is, the more precise the lower bound of (5) is. For instance, for $v = 8$, formula (5) yields

$$L > a_{16} - a_8 \cdot \left(e - \sum_{j=0}^{8} \frac{1}{i!} \right) = 0.364098149.$$

So even our very simple methods of estimation determine the first five digits after the decimal point in L. This level of precision is enough to verify that L is not in the very extensive database of mathematical constants given in [4].

5 An interesting fact about the numbers $d_k/(k-2)!$

As we have seen in the proof of Lemma 3.3, the numbers $d_k/(k-2)!$ count ordered pairs (m, F), where $m \in \{1, 2, \cdots, 2k-1\} = [2k-1]$, while F is a k-element subset of $[2k-1]$ so that F contains m, and the relative rank of m in F is *not* equal to the relative rank of m in $([2k-1] \setminus F) \cup \{m\}$.

Starting with $k = 2$, the first few numbers $h_k = d_k/(k-2)!$ are, 2, 14, 76, 374. This is sequence A172060 of the On-Line Encyclopedia of Integer Sequences [6] (shifted by one). The interpretation given to this sequence in [6] is equivalent to the following. Let g_k be the number of ordered pairs (b, p), where p is a lattice path starting at $(0, 0)$ and using $2(k-1)$ steps, each of which is $(1, -1)$ or $(1, 1)$, while b is an intersection point of p and the horizontal axis that is different from the origin.

It is straightforward to prove that $g_2 = 2$, and $g_k = 4g_{k-1} + \binom{2k-2}{k-1}$. Solving this recurrence relation using ordinary generating functions, we get that indeed, $g_k = (2k-1) \cdot \binom{2k-2}{k-1} - 4^{k-1} = h_k$ as claimed.

This raises the question whether we can prove the identity $g_k = h_k$ *combinatorially*. This is equivalent to asking for a direct bijective proof for the formula $g_k = (2k-1) \cdot \binom{2k-2}{k-1} - 4^{k-1}$. That, in turn, is equivalent to the following question.

QUESTION 5.1 Is there a simple bijective proof for the identity

$$\sum_{i=0}^{n} \binom{2i}{i} 4^{n-i} = (2n+1)\binom{2n}{n}? \tag{6}$$

There are several easy ways to interpret identity (6) combinatorially, using, for instance, lattice paths. In terms of generating functions, the left-hand side is the coefficient of x^n in the power series $\frac{1}{\sqrt{1-4x}} \cdot \frac{1}{1-4x}$, while the right-hand side is the coefficient of x^n in the obviously identical power series $(1-4x)^{-3/2}$, as computed by the Binomial theorem.

If we rewrite the factor 4^{n-i} using formula (4), we are led to the following intriguing question.

QUESTION 5.2 Is there a simple bijective proof for the identity

$$\sum_{i+j+k=n} \binom{2i}{i}\binom{2j}{j}\binom{2k}{k} = (2n+1)\binom{2n}{n}?$$

The sum is taken over all ordered triples (i, j, k) of non-negative integers satisfying $i + j + k = n$.

We hope to answer Questions 5.1, 5.2, and perhaps some of their generalizations, in a subsequent paper.

Added in proof. Question 5.2 has recently been answered in the affirmative [1].

References

[1] M. BÓNA, *A Bijective Proof of an Identity Extending a Classic Result of Hajos*, preprint, available at arXiv:1203.3264.

[2] A. DUANE AND J. REMMEL, *Minimal Overlapping Patterns in Colored Permutations*, submitted, 2011.

[3] S. ELIZALDE AND M. NOY, *Consecutive patterns in permutations*, Adv. in Appl. Math., 30 (2003) 110–125.

[4] S. R. FINCH, *Mathematical Constants*, Cambridge University Press, 2005.

[5] R. STANLEY, *Enumerative Combinatorics, Volume 1*, second edition, Cambridge University Press, 1999.

[6] *The On-Line Encyclopedia of Integer Sequences*, internet database, available at `http://oeis.org`.