sciendo
PU. M. A. Vol. **30** (2022), No.4, pp. 1 – 10

The expected number of distinct consecutive patterns in a random permutation

Austin Allen
The Ohio State University
email: allen.2678@buckeyemail.osu.edu

and

Dylan Cruz Fonseca
University of Puerto Rico, Río Piedras
email: dylan.cruz1@upr.edu

and

Veronica Dobbs
Marquette University
email: vonniedobbs@gmail.com

and

Egypt Downs
Virginia State University
email: downs.egypt@yahoo.com

and

Evelyn Fokuoh, Anant Godbole
East Tennessee State University
email: evefoks@gmail.com, godbolea@etsu.edu

and

Sebastián Papanikolaou Costa
Universidad Ana G.Méndez
email: spapanikolaou1@email.uagm.edu

and

Christopher Soto
Queens College, CUNY
email: csotonyc@gmail.com

and

Lino Yoshikawa
University of Hawaii at Hilo
email: lino.yoshikawa@gmail.com

(Received: November 24, 2020, and in revised form October 20, 2021.)

Abstract. Let π_n be a uniformly chosen random permutation on $[n]$. Using an analysis of the probability that two overlapping consecutive k-permutations are order isomorphic, we show that the expected number of distinct consecutive patterns of all lengths $k \in \{1, 2, \ldots, n\}$ in π_n is $\frac{n^2}{2}(1 - o(1))$ as $n \to \infty$. This exhibits the fact that random permutations pack consecutive patterns near-perfectly.

Mathematics Subject Classification(2020). 05D40, 60C05, 05A05.

Keywords: random permutation, permutation pattern, consecutive pattern.

DOI: 10.2478/puma-2022-0031

1 Introduction

Let $\pi = \pi_n$ be a permutation on $[n]$. The one-line notation will be used for permutations in this paper; e.g., (2134) is shorthand for $\pi(1) = 2; \pi(2) = 1; \pi(3) = 3; \pi(4) = 4$. We say that π_n *contains* a pattern $\mu = \mu_k$ of length k if there are k indices $n_1 < n_2 < \ldots < n_k$ such that $(\pi(n_1), \pi(n_2), \ldots, \pi(n_k))$ are in the same relative order as $(\mu(1), \mu(2), \ldots, \mu(k))$. We say that π_n *consecutively contains* the pattern μ_k if there are k consecutive indices $(m, m+1, \ldots, m+k-1)$ such that $(\pi(m), \pi(m+1), \ldots, \pi(m+k-1))$ are in the same relative order as $(\mu(1), \mu(2), \ldots, \mu(k))$. Let $\phi(\pi_n)$ be the number of distinct *consecutive* patterns of *all* lengths $k; 1 \le k \le n$, contained in π_n. We focus on the case where π_n is a *uniformly chosen random permutation* on $[n]$, denote the random value of $\phi(\pi_n)$ by X, and study, in this paper, its expected value $\mathbb{E}(X)$.

Throughout the paper we will employ the o, O notation, defined as follows for non-negative sequences a_n and b_n and constants K, L:

$$a_n = o(b_n) \text{ if } \frac{a_n}{b_n} \to 0 \ (n \to \infty);$$

$$a_n = O(b_n) \text{ if } \frac{a_n}{b_n} \le K \ (n \to \infty);$$

$$a_n = \omega(b_n) \text{ if } \frac{b_n}{a_n} \to 0 \ (n \to \infty);$$

$$a_n = \Omega(b_n) \text{ if } \frac{a_n}{b_n} \ge L \ (n \to \infty); \text{ and}$$

$$a_n = \Theta(b_n) \text{ if } a_n = O(b_n) \text{ and } a_n = \Omega(b_n).$$

1.1 Distinct subsequences and non-consecutive patterns

For context, first we summarize results on the extremal values of $\psi(\pi_n)$, where $\psi(\pi_n)$ is the number of distinct (*and not necessarily consecutive*) patterns of all lengths contained in π_n. The identity permutation reveals that

$$\min_{\pi_n \in S_n} \psi(\pi_n) = \min_{\pi_n \in S_n} \phi(\pi_n) = n + 1,$$

since the embedded patterns are $\emptyset, 1, 12, \ldots, (12\ldots n)$. On the other hand, motivated by a question posed by Herb Wilf at the inaugural Permutation Patterns meeting, held in Dunedin in 2003 (PP2003), several authors have studied the maximum value of $\psi(\pi_n)$. First we have the trivial pigeonhole bound

$$\max_{\pi_n \in S_n} \psi(\pi_n) \le \sum_{k=1}^{n} \min\left(\binom{n}{k}, k!\right) \sim 2^n, \tag{1}$$

which was mirrored soon after PP2003 by Coleman [6]:

$$\max_{\pi_n \in S_n} \psi(\pi_n) \ge 2^{n-2\sqrt{n}+1} \ \ (n = 2^k); \tag{2}$$

which led to

$$\left(\max_{\pi_n \in S_n} \psi(\pi_n)\right)^{1/n} \to 2.$$

A team of researchers began to see if this (surprising) bound could be improved. This led to the result in [1] that

$$\max_{\pi_n \in S_n} \psi(\pi_n) \ge 2^n \left(1 - 6\sqrt{n}2^{-\sqrt{n}/2}\right), \tag{3}$$

and thus to the conclusion that $\max_{\pi_n \in S_n} \psi(\pi_n) \sim 2^n$. Alison Miller improved both the upper and lower bounds (2) and (3), showing in [10] that

$$2^n - O(n^2 2^{n-\sqrt{2n}}) \leq \max_{\pi_n \in S_n} \psi(\pi_n) \leq 2^n - \Theta(n 2^{n-\sqrt{2n}}). \quad (4)$$

By extracting the constants in (4) and conducting an asymptotic analysis, Fokuoh showed in [9] that the trivial upper bound actually performs better than the one in (4) for small and not-too-small values of n, though, of course (4) does better asymptotically.

Turning to words, in [2] the authors studied the expected number $\mathbb{E}(\xi(W))$ of distinct subsequences of all lengths contained in the word $W = W_n$ obtained when n letters $s_1, \ldots, s_n$ are independently generated from a d-letter alphabet – with the ith letter being "typed" with probability α_i (they also covered the two-state Markov case). In the simplest case, when $d = 2$, it was shown in [2] (with $\alpha_1 := \alpha$) that asymptotically

$$\mathbb{E}(\xi(W)) \sim k\big(1 + \sqrt{\alpha(1-\alpha)}\big)^n,$$

which contains the earlier result from [8] that in the equiprobable case, $\mathbb{E}(\xi(W)) \sim k(\frac{3}{2})^n$ for a constant k.

The fact that $\mathbb{E}(\xi(W)) \sim A^n$ for $A < 2$ might suggest that the same is true for $\mathbb{E}(\psi(\pi_n))$. But consider the following argument. Since

$$k! \gg \binom{n}{k}$$

for large k, it would seem reasonable, via a heuristic "balls in boxes" argument that most or all of the patterns of large size contained in π_n would be distinct. It was accordingly conjectured in [9] that

$$\mathbb{E}(\psi(\pi_n)) \sim 2^n. \quad (5)$$

While we are unable to prove that (5) holds, we show in this paper that the following is true for the number X of distinct consecutive patterns contained in a random permutation:

Main Theorem

$$\mathbb{E}(X) = \max_{\pi_n \in S_n} (\phi(\pi_n))(1 - o(1)) = \frac{n^2}{2}(1 - o(1)).$$

2 Proof of main theorem

Lemma 2.1 *For any $\pi_n \in S_n$,*

$$\phi(\pi_n) \leq \sum_{k=1}^{n} \min\{(n-k+1), k!\} \leq \frac{n^2}{2}(1 + o(1)).$$

Proof. There are $k!$ permutation patterns of length k. However, not all of these can be present unless the number of consecutive positions of length k, namely $(n - k + 1)$, provide "enough room" for this to occur, i.e., if $(n - k + 1) \geq k!$ This proves the first inequality. Next note that

$$\phi(\pi_n) \leq \sum_{k=1}^{n} \min\{(n-k+1), k!\} \leq \sum_{k=1}^{n} (n-k+1) = \sum_{k=1}^{n} k \leq \frac{n^2}{2}(1 + o(1)). \quad (6)$$

This completes the proof. □

Lemma 2.2

$$\sum_{k=1}^{n} \min\{(n-k+1), k!\} \geq \frac{n^2}{2}(1 - o(1)).$$

Proof. The equation $(n-k+1)=k!$, by Stirling's approximation, holds if

$$\sqrt{2\pi k}\left(\frac{k}{e}\right)^k(1+o(1))+k=n+1,$$

which is true if and only if

$$(\exp\{k\log k-k+(1/2)(\log k+\log(2\pi))+o(1)\})\,(1+o(1))=\exp\{\log n(1+o(1))\}.$$

A good approximation to the above is the solution to $k\log k=\log n$, namely

$$k=\frac{\log n}{\log\log n}(1+o(1)),$$

which yields, with

$$a_n=\left\lceil\frac{\log n}{\log\log n}(1+o(1))\right\rceil,$$

$$\sum_{k=1}^{n}\min\{(n-k+1),k!\}\geq\sum_{k=a_n}^{n}(n-k+1)\sim\frac{(n-a_n)^2}{2}=\frac{n^2}{2}(1-o(1)),$$

as asserted. □

We mention that the evidence in support of the Main Theorem is strong, as evidenced by the following data for small n (the evidence in support of (5) is not as strong; see [9])

n	$\sum_{k=1}^{n}\min(n-k+1,k!)$	Bound attained (Y/N)	$\mathbb{E}(X)$
3	4	Yes	3.67
4	6	Yes	5.83
5	9	Yes	8.7
6	13	Yes	12.33
7	18	Yes	16.78
8	24	Yes	22.08

For example, the permutation 14325 contains the 9=$\sum_{k=1}^{5}\min\{k!,6-k\}$ patterns 1, 12, 21, 132, 321, 213, 1432, 3214, and 14325.

A study of patterns that occur in consecutive positions in a permutation is not new. For example, the so called vincular patterns partially follow this scheme. More relevant to this paper, however, is the work of [3] and [7] on non-consecutive permutations that touches on some of the aspects of this paper. Of far greater relevance, however, are the works of Borga and Penaguiao [4], [5], who study the *feasible region* for consecutive patterns, showing that this is the cycle polytope of the *overlap graph*. They also make deep comparisons between classical and consecutive occurrences of patterns.

2.1 Auxiliary random variables

First recall that a_n is the integer k that first causes $\min\{n-k+1,k!\}=(n-k+1)$. To ease the analysis, we could lower bound the expected number of distinct patterns of all lengths by the expected number of distinct patterns of length $\geq a_n$. We will see later that it facilitates the analysis even more if we count the expected number of distinct patterns of length b_n or more, where $b_n\geq a_n$ will be chosen later. We will attack our problem using several auxiliary random variables. We wish to analyze the behavior of

$$X=\sum_{k=1}^{n}X_k,$$

where X_k is the number of distinct patterns of length k. Next, let us list the sets of $n-k+1$ consecutive positions of length k lexicographically starting from $k=1$ and going till $k=n$, thus getting the list

$$1, 2, \ldots, n, (1,2), (2,3), \ldots, (n-1,n), (1,2,3), (2,3,4), \ldots, (n-2,n-1,n), \ldots$$

$$(1, 2, \ldots n)$$

The number of patterns of length $\geq b_n$ equal those contained in the portion of the above list starting with $(1, 2, \ldots, b_n)$. The number of distinct patterns of length $\geq b_n$ thus equals

$$\sum_{k=b_n}^{n} (n-k+1) - Y_k,$$

where Y_k is the number of patterns which are "repeat" patterns of length k, i.e., those that have appeared lexicographically before.

Also, for $1 \leq k \leq n$, let Z_k be the number of *pairs of consecutive positions (overlapping or not) that yield the same pattern of length* k*, and let* $Z = \sum_{k=1}^{n} Z_k$.

We give an example to fix these ideas: For $n = 8$, consider the permutation

14573268.

It is easy to check that $a_8 = 3$ and we lexicographically list the patterns of length 3 and higher, getting the list 123, 123, 231, 321, 213, 123, 1234, 2341, 3421, 4213, 2134, 13452, 34521, 35214, 42135, 145632, 346215, 462156, 1457326, 3462157, and 14573268. The only pattern that repeats is the pattern 123 of length 3, and it repeats twice. Thus the number of distinct patterns of length 3 or more is

$$[(8-3+1)-2] + (8-4+1) + (8-5+1) + (8-6+1) + (8-7+1) + (8-8+1)$$

$$= 21 - 2 = 19.$$

Note moreover that we will later be replacing a_n by b_n.

In the above example, the pattern 123 appears lexicographically in the first, second, and sixth sets of consecutive positions, and so there are $\binom{3}{2} = 3$ sets of offending pairs of consecutive positions. In general, if a pattern such as 123 appears r times, then its contribution to Y_3 is $r-1$, while its contribution to Z_3 is the larger quantity $\binom{r}{2}$. Moreover $Y_k = 0$ iff $Z_k = 0$ since there are no repeated patterns if and only if there are no pairs of equal patterns. If $Y_k = 1$ then there is a k-pattern that repeats once and hence there is one pair of equal patterns. Thus $Y_k = 1$ iff $Z_k = 1$, and, in general, we will see that $Y_k \leq Z_k$.

Formalizing the above example discussion and example we see that

$$\mathbb{E}(X) = \sum_{k=1}^{n} \mathbb{E}(X_k) \geq \sum_{k=b_n}^{n} \mathbb{E}(X_k) = \sum_{k=b_n}^{n} \left((n-k+1) - \mathbb{E}(Y_k)\right), \tag{7}$$

where Y_k denotes the number of k patterns that are consecutively contained in the random n-permutation as lexicographical repeats.

We introduced the variables Z_k above since Y_k is difficult to work with directly. Since Z_k counts pairs of patterns of length k, we organize them according to

(i) the magnitude of the overlap $l; 0 \leq l \leq k-1$ between the pairs; and

(ii) the starting position j of the first pattern in the pair;

Thus

$$Z_k = \sum_{j=1}^{n-k+1} \sum_{l=1}^{k-1} I_{j,l,k} + \sum_{j=1}^{n-k+1} \sum_{r \leq t_j} I_{j,0,k,r} \tag{8}$$

where for $l \geq 1$, the indicator variable $I_{j,l,k}$ equals one iff $(\pi(j), \ldots, \pi(j+k-1))$ and $(\pi(j+k-l)\ldots, \pi(j+2k-l-1))$ are order isomorphic ($I_{j,l,k}$ equals zero otherwise.) For $l = 0$, $I_{j,0,k,r} = 1$ if $(\pi(j), \ldots, \pi(j+k-1))$ and $(\pi(j+k-1+r), \ldots, \pi(j+2k-2+r))$ are order isomorphic. Note that for $l = 0$, there are several consecutive positions disjoint from $(j, j+1, \ldots, j+k-1)$, where we loosely bound the number of these positions by an unspecified $t_j \leq n$.

Observe also that for some j's towards the end, not all $l \geq 1$ (or r's in the case of $l = 0$), are feasible and in this case we have $\mathbb{P}(I_{j,l,k} = 1) = 0$ or $\mathbb{P}(I_{j,0,k,r} = 0)$. We will see several spots where a non-zero upper bound is used for expressions such as these instead of 0. We have seen above that for any k,

$$Y_k = 0 \text{ (resp. 1)} \Leftrightarrow Z_k = 0 \text{ (resp. 1)}.$$

Note that for any k

$$\begin{aligned}
\mathbb{E}(Z_k) &= \sum_{j=1}^{n-k+1} \sum_{l=1}^{k-1} \mathbb{E}(I_{j,l,k}) + \sum_{j=1}^{n-k+1} \sum_{r \leq t_j} \mathbb{E}(I_{j,0,k,r}) \\
&= \sum_{j=1}^{n-k+1} \sum_{l=1}^{k-1} \mathbb{P}(I_{j,l,k} = 1) + \sum_{j=1}^{n-k+1} \sum_{r \leq t_j} \mathbb{P}(I_{j,0,k,r} = 1).
\end{aligned}$$

The next result formally proves that $\mathbb{E}(Y_k) \leq \mathbb{E}(Z_k)$.

PROPOSITION 2.3 *For each k, $Y_k \leq Z_k$, i.e., the random variable (r.v.) Y_k is majorized by the r.v. Z_k and thus $\mathbb{E}(Y_k) \leq \mathbb{E}(Z_k)$.*

Proof. Suppose that

$$Y_k = r = \sum_m r_m,$$

where r_m is the number of repeats of the mth pattern-type. Then there are $\binom{r_m+1}{2}$ pairs of isomorphic patterns of the mth pattern type, and since for each m, $r_m \leq \binom{r_m+1}{2}$, the proposition is proved. □

LEMMA 2.4 *For $l = 0$ and any feasible j, r, $\mathbb{P}(I_{j,0,k,r} = 1) = \frac{1}{k!}$.*

Proof. Consider any set of k consecutive positions disjoint from (and lexicographically greater than) $\{j, j+1, \ldots, j+k-1\}$. The probability that these positions contains a pattern isomorphic to the one in $\{j, j+1, \ldots, j+k-1\}$ is $\frac{k! \cdot 1}{k!^2} = \frac{1}{k!}$, since given the pattern in the first set of k positions, the pattern in the second set must be the same. This proves the result. □

LEMMA 2.5 *For $l = k-1$ and any k, $\mathbb{P}(I_{j,k-1,k} = 1) = \frac{2}{(k+1)!}$.*

Proof. Consider the $k+1$ spots spanned by the two sets of positions. If the numbers in these spots are monotone increasing or monotone decreasing then we have that $I_{j,k-1,k} = 1$, so $\mathbb{P}(I_{j,k-1,k} = 1) \geq \frac{2}{(k+1)!}$. If, on the other hand the numbers in positions $(j, j+1, \ldots, j+k-1)$ are not monotone increasing or decreasing, then (in the increasing case), there exist indices $i, i+1, i+2$ such that $\pi(i+1) > \max\{\pi(i), \pi(i+2)\}$. But then, for the pattern in the k positions shifted to the right by 1, $\pi(i) > \pi(i+1)$, so that we cannot have $I_{j,k-1,k} = 1$. Thus $\mathbb{P}(I_{j,k-1,k} = 1) = \frac{2}{(k+1)!}$. □

LEMMA 2.6 *For $1 \leq l \leq \lfloor \frac{k}{2} \rfloor$,*

$$\mathbb{P}(I_{j,l,k} = 1) \leq \frac{3^k}{k!}. \tag{9}$$

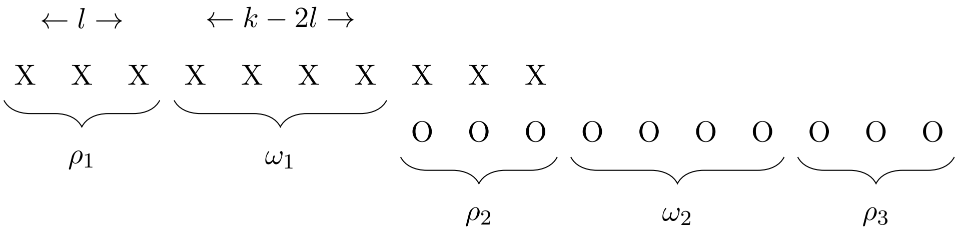

Figure 1: Consecutive overlapping patterns, $l \leq \lfloor \frac{k}{2} \rfloor$

Proof. Denote the two k-patterns in question, overlapping in l positions, by η_1 and η_2. For any patterns ξ_1, ξ_2, we write $\xi_1 \simeq \xi_2$ if ξ_1 and ξ_2 are order isomorphic. We must have the following situation shown in Figure 1:

$$\rho_1 \simeq \rho_2 \simeq \rho_3,$$

and

$$\omega_1 \simeq \omega_2$$

in order for η_1 and η_2 to be *consistent with being order isomorphic.* Thus

$$\begin{aligned} \mathbb{P}(\eta_1 \simeq \eta_2) &\leq \mathbb{P}(\rho_1 \simeq \rho_2 \simeq \rho_3, w_1 \simeq w_2) \\ &\leq \frac{1}{l!^2} \frac{1}{(k-2l)!}. \end{aligned} \tag{10}$$

Now if $k = 3m$ for some integer m then it is easy to show that the trinomial coefficient

$$\frac{k!}{l!^2(k-2l)!}$$

is maximized on setting $l = \frac{k}{3}$. This may be done by setting $\beta(l) = \frac{k!}{l!^2(k-2l)!}$ and solving the quadratic that arises on setting $\beta(l+1)/\beta(l) \geq 1$, to identify the non-decreasing regions of β. Thus for some constant K, by Stirling's approximation we get

$$\frac{k!}{(l!^2)(k-2l)!} \leq \frac{k!}{(k/3)!^3} \sim K \cdot \frac{3^k}{k},$$

and thus (10) yields

$$\mathbb{P}(\eta_1 \simeq \eta_2) \leq K \cdot \frac{3^k}{k! \cdot k} \leq \frac{3^k}{k!}. \tag{11}$$

Moreover if $k = 3m+1$ or $k = 3m+2$, then the trinomial coefficient is maximized by setting l and $k - 2l$ to be as equal as possible and it is easy to verify that the bound in (9) holds as well. Note that the bound in (9) is uniform, i.e., independent of the value of $l \leq \lfloor k/2 \rfloor$. Also, any analysis that tries to be more exact gets very complicated rapidly and so we will settle for the upper bound in (9). □

The same kind of analysis, in which we use consistency with order isomorphism as a driving method, is used for $l > \lceil \frac{k}{2} \rceil$, which we turn to next.

Lemma 2.7 *For* $k - 2 \geq l > \lceil \frac{k}{2} \rceil$,

$$\mathbb{P}(I_{j,l,k} = 1) \leq \left(\frac{1}{(k-l)!} \right)^{\frac{k}{k-l} - 1}. \tag{12}$$

η_1:	A	B	A	B	A	B	A	B			
			η_2:	A	B	A	B	A	B	A	B

Figure 2: Consecutive overlapping patterns, $l \geq \lceil \frac{k}{2} \rceil$

Proof. We first illustrate the idea of the proof for $k = 8; l = 6$; see Figure 2. If the first two elements of η_1 form the pattern AB, then so must the first two elements of η_2, which are also the third and fourth elements of η_1 – forcing the third and fourth elements of η_2 to form an AB pattern too. This repetition of the AB pattern persists till we reach the end of η_2, for a total of four induced AB patterns caused by the first two elements of η_1. Thus

$$\mathbb{P}(I_{j,l,k} = 1) \leq \left(\frac{1}{2!}\right)^4.$$

In general the pattern in the first $k - l$ positions of η_1 is repeated $\lfloor \frac{k}{k-l} \rfloor \geq \frac{k}{k-l} - 1$ times, each of which has a probability $\frac{1}{(k-l)!}$. This completes the proof. □

2.2 Putting it all together

For $k \geq b_n$ (b_n is still to be specified), we seek to find $\sum_{l=0}^{k-1} \mathbb{P}(I_{j,l,k} = 1)$. We address the case of $l = 0$, $l = k - 1$ first. By Lemmas 2.4 and 2.5, we have

$$\sum_{r \leq t_j} \mathbb{P}(I_{j,0,k,r} = 1) \leq \frac{n}{k!}, \tag{13}$$

and

$$\mathbb{P}(I_{j,k-1,k} = 1) \leq \frac{2}{(k+1)!}. \tag{14}$$

Hence

$$\sum_{k=b_n}^{n} \sum_{j=1}^{n-k+1} \sum_{r \leq t_j} \mathbb{P}(I_{j,0,k,r} = 1) \leq \frac{n^3}{k!}, \tag{15}$$

and

$$\sum_{k=b_n}^{n} \sum_{j=1}^{n-k+1} \mathbb{P}(I_{j,k-1,k} = 1) \leq \frac{2n^2}{(k+1)!}. \tag{16}$$

For the case of small overlaps, Lemma 2.6 gives

$$\sum_{k=b_n}^{n} \sum_{j=1}^{n-k+1} \sum_{l=1}^{\lfloor k/2 \rfloor} \mathbb{P}(I_{j,k,l} = 1) \leq \frac{n^2 k 3^k}{k!}. \tag{17}$$

Finally, for the large overlap case, we have

$$\sum_{k=b_n}^{n} \sum_{j=1}^{n-k+1} \sum_{l=\lceil k/2 \rceil}^{k-2} \mathbb{P}(I_{j,k,l} = 1)$$
$$\leq \quad n^2 \sum_{l} \left(\frac{1}{(k-l)!}\right)^{\frac{k}{k-l}-1} = n^2 \sum_{l} \left(\frac{1}{(k-l)!}\right)^{\frac{l}{k-l}}. \tag{18}$$

In (18), the $l = k-2$ term is $2(1/2)^{k/2}$, which we treat separately. For the other terms we use the inequality $r! \geq \sqrt{2\pi r}(r/e)^r$ to provide the estimate

$$\left(\frac{1}{(k-l)!}\right)^{\frac{l}{k-l}} \leq \left\{\left(\frac{1}{(k-l)!}\right)^{\frac{1}{k-l}}\right\}^{k/2} \leq \left(\frac{e(1+o(1))}{k-l}\right)^{k/2} \leq (0.96)^k, \tag{19}$$

where we plug in $l = k-3$ at the last step. The total contribution of the large overlap case is thus

$$2n^2\left(\frac{1}{2}\right)^{k/2} + n^2k(0.96)^k \tag{20}$$

Proposition (2.3) together with (15), (16), (17), and (20) yield

$$\begin{aligned}\sum_{k=b_n}^{n}\mathbb{E}(Y_k) &\leq \sum_{k=b_n}^{n}\mathbb{E}(Z_k) \leq \frac{n^3}{k!} + \frac{2n^2}{(k+1)!} + \frac{n^2 3^k}{(k-1)!}\\ &+ 2n^2\left(\frac{1}{2}\right)^{k/2} + n^2k(0.96)^k\\ &= T_1 + T_2 + T_3 + T_4 + T_5 \text{ say.}\end{aligned} \tag{21}$$

Clearly, $T_1 \geq T_2$ and $T_5 \geq T_4$. Moreover $T_5 \geq T_1$ if $k \cdot k!(0.96)^k \geq n$, or, since $k! \geq (k/e)^k$ if $\left(\frac{k}{3}\right)^k \geq n$, which certainly holds if $k \geq \ln n$. Finally, $T_5 \geq T_3$ if $(0.32)^k \geq 1/k!$, which holds if $k \geq 7$ or if $n \geq e^7$. Thus the dominant term in (21) is $n^2k(0.96)^k$ provided that n is large enough and $k \geq \ln n$. In this case, $\sum_{k=b_n}^{n}\mathbb{E}(Z_k) \leq 5n^2k(0.96)^k \leq n^4(0.96)^k \leq 1$ if $(e^{-0.0408\ldots})^k \leq e^{-4\ln n}$, or if $k \geq 100\ln n$. With the above discussion in mind, we let

$$b_n = \lceil 100\ln n\rceil,$$

which yields

$$\begin{aligned}\mathbb{E}(X) &\geq \left(\sum_{k=\lceil 100\ln n\rceil}^{n}(n-k+1)\right) - 1\\ &= \frac{(n-\lceil 100\ln n\rceil)^2}{2}(1-o(1)) - 1\\ &= \frac{n^2}{2}(1-o(1)),\end{aligned}$$

proving the main theorem.

Acknowledgments. This research was supported by NSF Grant 1852171. We thank the referee for a careful and incisive reading of the paper. Implementing the comments in the referee report improved the paper significantly, both in form and in content.

References

[1] M. Albert, M. Coleman, R. Flynn and I. Leader, *Permutations containing many patterns*, Ann. Comb., 11 (2007) 265—270.

[2] Y. Biers-Ariel, A. Godbole and E. Kelley, *Expected number of distinct subsequences in randomly generated binary strings*, Discrete Math. Theor. Comput. Sci., 19 (2018) #10.

[3] M. Bóna, *On non-overlapping patterns of permutations*, Pure Math. Appl. (PU.M.A.), 22 (2011) 99–105.

[4] J. Borga and R. Penaguiao, *The feasible region for consecutive patterns of permutations is a cycle polytope*, available at https://arxiv.org/pdf/1910.02233.pdf.

[5] J. Borga and R. Penaguiao, *The feasible regions for consecutive patterns of pattern-avoiding permutations*, available at https://arxiv.org/pdf/2010.06273.pdf.

[6] M. Coleman, *An answer to a question by Wilf on packing distinct patterns in a permutation*, Electron. J. Combin., 11 (2004) #N8.

[7] A. Duane and J. Remmel, *Minimal overlapping patterns in colored permutations*, Electron. J. Combin., 18 (2011) #P25.

[8] A. Flaxman, A. Harrow and G. Sorkin, *Strings with maximally many distinct subsequences and substrings*, Electron. J. Combin., 11 (2004) #R8.

[9] E. Fokuoh, *The expected number of patterns in a randomly generated permutation on* $[n] = \{1, 2, ..., n\}$, M.Sc. Thesis, East Tennessee State University, 2018, available at https://dc.etsu.edu/cgi/viewcontent.cgi?article=4916&context=etd.

[10] A. Miller, *Asymptotic bounds for permutations containing many different patterns*, J. Combin. Theory Ser. A, 116 (2009) 92–108.