LAW OF LARGE NUMBERS FOR INCREASING SUBSEQUENCES OF RANDOM PERMUTATIONS

Ross G. Pinsky

Department of Mathematics
Technion-Israel Institute of Technology
Haifa, 32000 Israel
email:pinsky@math.technion.ac.il

Abstract. Let the random variable $Z_{n,k}$ denote the number of increasing subsequences of length k in a random permutation from S_n, the symmetric group of permutations of $\{1,...,n\}$. We show that $Var(Z_{n,k_n})=o((EZ_{n,k_n})^2)$ as $n\to\infty$ if and only if $k_n=o(n^{\frac{2}{5}})$. In particular then, the weak law of large numbers holds for Z_{n,k_n} if $k_n=o(n^{\frac{2}{5}})$; that is,

$$\lim_{n\to\infty}\frac{Z_{n,k_n}}{EZ_{n,k_n}}=1, \text{ in probability.}$$

We also show the following approximation result for the uniform measure U_n on S_n. Define the probability measure $\mu_{n;k_n}$ on S_n by

$$\mu_{n;k_n}=\frac{1}{\binom{n}{k_n}}\sum_{x_1<x_2<...<x_{k_n}}U_{n;x_1,x_2,...x_{k_n}},$$

where $U_{n;x_1,x_2,...,x_{k_n}}$ denotes the uniform measure on the subset of permutations which contain the increasing subsequence $\{x_1,x_2,...,x_{k_n}\}$. Then the weak law of large numbers holds for Z_{n,k_n} if and only if

$$(*)\qquad \lim_{n\to\infty}||\mu_{n;k_n}-U_n||=0,$$

where $||\cdot||$ denotes the total variation norm. In particular then, (*) holds if $k_n=o(n^{\frac{2}{5}})$.

In order to evaluate the asymptotic behavior of the second moment, we need to analyze occupation times of certain conditioned two-dimensional random walks.

1991 *Mathematics Subject Classification.* 60C05, 60F05.

Key words and phrases. random permutations, law of large numbers, increasing subsequences in random permutations, conditioned random walks.

This research was supported by the Fund for the Promotion of Research at the Technion and by the V.P.R. Fund.

1. Introduction and Statement of Results. Let S_n denote the symmetric group of permutations of $\{1, ..., n\}$. By introducing the uniform probability measure U_n on S_n, one can consider $\sigma \in S_n$ as a random permutation. Probabilities and expectations according to U_n will frequently be denoted by the generic notation P and E respectively. The problem of analyzing the distribution of the length, L_n, of the longest increasing subsequence in a random permutation from S_n has a long and distinguished history; see [1] and references therein. In particular, the work of Logan and Shepp [6] together with that of Vershik and Kerov [8] show that $EL_n \sim 2n^{\frac{1}{2}}$ and that $\sigma^2(L_n) = o(n)$, as $n \to \infty$. Profound recent work by Baik, Deift and Johansson [2] has shown that $\lim_{n\to\infty} P(\frac{L_n - 2n^{\frac{1}{2}}}{n^{\frac{1}{6}}} \le x) = F(x)$, where F is an explicitly identifiable function.

There doesn't seem to be any literature on the random variable $Z_{n,k} = Z_{n,k}(\sigma)$, which we define to be the number of increasing subsequences of length k in a permutation $\sigma \in S_n$. Thus, for example, if $\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 3 & 4 & 5 & 2 \end{pmatrix}$, then $Z_{5,3}(\sigma) = 4$ because there are four increasing subsequences of length three; namely, 134, 135, 145 and 345. It is useful to represent $Z_{n,k}$ as a sum of indicator random variables. For positive integers $\{x_1, ..., x_k\}$ satisfying $1 \le x_1 < x_2 < ... < x_k \le n$, let $B^n_{x_1,...x_k} \subset S_n$ denote the subset of permutations which contain the increasing subsequence $\{x_1, x_2, ..., x_k\}$. Then we have

$$Z_{n,k} = \sum_{x_1 < x_2 < ... < x_k} 1_{B^n_{x_1,x_2,...,x_k}},$$

where the sum is over the $\binom{n}{k}$ distinct increasing subsequences of length k. Since the probability that a random permutation fixes any particular increasing sequence of length k is $\frac{1}{k!}$, it follows that the expected value of $Z_{n,k}$ is given by

$$EZ_{n,k} = \frac{\binom{n}{k}}{k!}. \tag{1.1}$$

One can consider k to depend on n in which case we write k_n. We are interested in a law of large numbers of the form $\frac{Z_{n,k_n}}{EZ_{n,k_n}} \to 1$ in probability, for appropriate choices of k_n. Of course, in light of the above cited works on the longest increasing subsequence, such a result cannot hold for $k_n \geq cn^{\frac{1}{2}}$ with $c > 2$. A straightforward calculation using Stirling's formula shows that

$$\text{(1.2)} \qquad \begin{aligned} EZ_{n,cn^l} &\sim \frac{1}{2\pi cn^l}\Big[(\frac{e}{c})^2 n^{1-2l}\Big]^{cn^l}, \text{ as } n \to \infty, \text{ for } l \in (0, \frac{1}{2}); \\ EZ_{n,cn^{\frac{1}{2}}} &\sim \frac{\exp(-\frac{c^2}{2})}{2\pi cn^{\frac{1}{2}}}(\frac{e}{c})^{2cn^{\frac{1}{2}}}. \end{aligned}$$

(For the case $k_n = cn^{\frac{1}{2}}$, we have used the fact that $\lim_{n\to\infty}\prod_{j=0}^{cn^{\frac{1}{2}}-1}(1-\frac{j}{n}) = \exp(-\frac{c^2}{2})$, which is proved by taking the logarithm of the above product. Note that the factor $\exp(-\frac{c^2}{2})$ suddenly appears in the formula when $l = \frac{1}{2}$.) In particular then, it follows from (1.2) that $\lim_{n\to\infty} EZ_{n,k_n} = \infty$, if $k_n \leq cn^{\frac{1}{2}}$ with $c < e$, and $\lim_{n\to\infty} EZ_{n,k_n} = 0$, if $k_n \geq en^{\frac{1}{2}}$.

The law of large numbers for Z_{n,k_n} is in fact equivalent to a certain approximation result for the uniform measure, which we now describe. Recall that for probability measures P_1 and P_2 on S_n, the total variation norm is defined by

$$||P_1 - P_2|| \equiv \max_{A\subset S_n}(P_1(A) - P_2(A)) = \frac{1}{2}\sum_{\sigma\in S_n}|P_1(\sigma) - P_2(\sigma)|.$$

For $x_1 < x_2 < \ldots < x_{k_n}$, let $U_{n;x_1,x_2,\ldots,x_{k_n}}$ denote the uniform measure on permutations which have $\{x_1, x_2, \ldots, x_{k_n}\}$ as an increasing sequence; that is $U_{n;x_1,x_2,\ldots,x_{k_n}}$ is uniform on $B^n_{x_1,x_2,\ldots,x_{k_n}}$. Note that $U_{n;x_1,x_2,\ldots,x_{k_n}}$ is defined by $U_{n;x_1,x_2,\ldots,x_{k_n}}(\sigma) = \frac{k_n!}{n!}1_{B^n_{x_1,x_2,\ldots,x_{k_n}}}(\sigma)$. Now define the probability measure $\mu_{n;k_n}$ on S_n by

$$\mu_{n;k_n} = \frac{1}{\binom{n}{k_n}}\sum_{x_1<x_2<\ldots<x_{k_n}} U_{n;x_1,x_2,\ldots x_{k_n}}.$$

Equivalently,

$$\text{(1.3)} \qquad \mu_{n;k_n}(\sigma) = \frac{1}{\binom{n}{k_n}}\frac{k_n!}{n!}Z_{n,k_n}(\sigma), \quad \sigma \in S_n.$$

The measure $\mu_{n;k_n}$ can be realized concretely as follows. Consider n cards, numbered from 1 to n, and laid out on a table from left to right in increasing order. Place a black mark on k_n of the cards, chosen at random. Pick up all the cards without black marks and then randomly insert them between the k_n cards with black marks that remain on the table. The resulting distribution is $\mu_{n;k_n}$.

Proposition 1. *The law of large numbers holds for* Z_{n,k_n}*; that is*

$$\lim_{n\to\infty}\frac{Z_{n,k_n}}{EZ_{n,k_n}}=1 \text{ in probability,}$$

if and only if

$$\lim_{n\to\infty}||\mu_{n;k_n}-U_n||=0. \tag{1.4}$$

The proof of Proposition 1 appears at the end of this section.

The measure $\mu_{n;k_n}$ corresponds to ignoring a set of k_n random cards and randomizing the rest of the cards. How many random cards can one afford to ignore like this and maintain asymptotic randomness? Corollary 2 below shows that one can afford to ignore $k_n=o(n^{\frac{2}{5}})$ cards, while the results cited above on the longest increasing subsequence show that one certainly cannot afford to ignore $cn^{\frac{1}{2}}$ cards for $c>2$.

For the law of large numbers we will use Chebyshev's inequality. The calculation of the second moment is nontrivial because it involves expectations of the form $E1_{B^n_{x_1,\dots,x_{k_n}}}1_{B^n_{y_1,\dots,y_{k_n}}}$, and these expectations depend rather intimately on the relative positions of $\{x_1,x_2,....,x_{k_n}\}$ and $\{y_1,y_2,....,y_{k_n}\}$. We begin with the explicit form of the second moment of $Z_{n,k}$ for any $k\le n$.

Proposition 2.

$$EZ^2_{n,k}=\sum_{j=0}^{k}\binom{n}{2k-j}\frac{1}{(2k-j)!}A(k-j,j),$$

where

$$A(N,j)=\sum_{\substack{\sum_{r=0}^{j}l_r=N\\ \sum_{r=0}^{j}m_r=N}}\prod_{r=0}^{j}\left(\frac{(l_r+m_r)!}{l_r!m_r!}\right)^2. \tag{1.5}$$

In order to evaluate the asymptotic behavior of $\text{Var}(Z_{n,k_n})$, one must be able to adequately evaluate the asymptotic behavior of $A(k_n-j,j)$. In fact, it turns out that we need a good lower bound for $A(k_n-1,1)$ and a good upper bound for

$A(k_n - j, j)$, for all $j = 1, 2, ...k_n$. We were able to interpret $\frac{A(N,j)}{\binom{2N}{N}^2}$ as the sum of certain expected occupation times of the horizontal axis for the standard, simple, symmetric two-dimensional random walk starting from the origin and conditioned on returning to the origin at the $2N$-th step. This characterization was sufficient to obtain the appropriate bounds to prove the following theorem.

Theorem 1.

i. If $k_n = o(n^{\frac{2}{5}})$, *then*

$$\frac{Var(Z_{n,k_n})}{(EZ_{n,k_n})^2} = O(\frac{k_n^{\frac{5}{2}}}{n}), \text{ as } n \to \infty;$$

In particular then, $Var(Z_{n,k_n}) = o((EZ_{n,k_n})^2)$, *as* $n \to \infty$.

ii. If $c_1 n^{\frac{2}{5}} \le k_n \le c_2 n^{\frac{2}{5}}$, *for constants* $c_1, c_2 > 0$, *then*

$$c_3(EZ_{n,k_n})^2 \le Var(Z_{n,k_n}) \le c_4(EZ_{n,k_n})^2,$$

for constants $c_3, c_4 > 0$.

iii. If $\lim_{n\to\infty} n^{-\frac{2}{5}} k_n = \infty$ *and* $\limsup_{n\to\infty} n^{-\frac{1}{2}} k_n < \infty$, *then*

$$\lim_{n\to\infty} \frac{Var(Z_{n,k_n})}{(EZ_{n,k_n})^2} = \infty.$$

Corollary 1. *i. If* $k_n = o(n^{\frac{2}{5}})$, *then*

$$\lim_{n\to\infty} \frac{Z_{n,k_n}}{EZ_{n,k_n}} = 1, \text{ in probability;}$$

ii. If $k_n = O(n^{\frac{2}{5}})$, *then*

$$\liminf_{n\to\infty} P(\frac{Z_{n,k_n}}{EZ_{n,k_n}} > \delta) > 0, \text{ for some } \delta > 0.$$

Part (i) of Corollary 1 follows immediately from Chebyshev's inequality and Theorem 1-i. The proof of part (ii) of Corollary 1 appears below.

Corollary 1 and Proposition 1 yield immediately the following approximation result.

Corollary 2. *If* $k_n = o(n^{\frac{2}{5}})$, *then*

$$\lim_{n\to\infty} ||\mu_{n;k_n} - U_n|| = 0.$$

In light of the above results, we pose the following question:

Open Question: *Presumably there exists a critical exponent* l_c *such that the law of large numbers holds for* Z_{n,n^l} *with* $l < l_c$ *and does not hold for* $l > l_c$. *What is* l_c*?*

In section two we prove Proposition 2 and in section 3 we prove Theorem 1. Lemmas 2 and 3, which appear in section 3 and give the key estimates on $A(N, j)$ used in the proof of Theorem 1, are proved in section four.

The literature on increasing subsequences in random permutations in a context other than that of the largest such subsequence is very scarce. The random variable Z_n, defined as the total number of increasing subsequences of all possible lengths in a random permutation, was studied in [5]. Both EZ_n and $Var(Z_n)$ were calculated explicitly and evaluated asymptotically. It turns out that $Var(Z_n)$ is of a larger order than $(EZ_n)^2$, so it is not possible to apply Chebyshev's inequality and obtain a law of large numbers. However, the authors were able to show that $\frac{\log Z_n}{n^{\frac{1}{2}}}$ converges in probability and in mean to a positive constant. In [3], the random variable $Z_{n,k}$ actually appears in a different guise. Equation (1.1) appears there as well as an upper bound for $EZ_{n,cn^{\frac{1}{2}}}$; however, this random variable is not the object of study in that paper. In [7], *inversions*–which are decreasing subsequences of length 2—are studied, and a central limit theorem is proved.

We conclude this section with the proofs of Corollary 1-ii and Proposition 1.

Proof of Corollary 1-ii. Assume to the contrary that the result is not true. Then there exists a subsequence $\{(n_i, k_{n_i})\}_{i=1}^{\infty}$ of $\{(n, k_n)\}_{n=1}^{\infty}$, such that $\frac{Z_{n_i,k_{n_i}}}{EZ_{n_i,k_{n_i}}}$ goes to 0 in probability. By taking a further subsequence if necessary, we may assume that either $k_{n_i} = o(n_i^{\frac{2}{5}})$ as $i \to \infty$, or $\lim_{i\to\infty} n_i^{-\frac{2}{5}} k_{n_i} = c > 0$. In light of part (i), we obtain a contradiction in the former case. Thus, it remains to consider the latter case. In this case, it follows from Theorem 1-ii that $\frac{(EZ_{n_i,k_{n_i}})^2}{Var(Z_{n_i,k_{n_i}})}$ is bounded away from 0 and ∞. Using this along with the assumption that $\frac{Z_{n_i,k_{n_i}}}{EZ_{n_i,k_{n_i}}} \to 0$ in probability, we conclude that

$$\lim_{n\to\infty} P(\frac{Z_{n_i,k_{n_i}} - EZ_{n_i,k_{n_i}}}{\sqrt{Var(Z_{n_i,k_{n_i}})}} \leq -\rho) = 1, \tag{1.6}$$

for some $\rho > 0$. However, since the second moments of the $\frac{Z_{n_i,k_{n_i}} - EZ_{n_i,k_{n_i}}}{\sqrt{Var(Z_{n_i,k_{n_i}})}}$ are equal to 1, this quotient is uniformly bounded. The uniform boundedness along with (1.6) contradict the fact that the first moment of $\frac{Z_{n_i,k_{n_i}} - EZ_{n_i,k_{n_i}}}{\sqrt{Var(Z_{n_i,k_{n_i}})}}$ is 0. □

Proof of Proposition 1. For $\epsilon \in (0,1)$, define

$$D_{n,\epsilon,k_n} = \{\sigma \in S_n : \frac{\mu_{n;k_n}(\sigma)}{U_n(\sigma)} \in [1-\epsilon, 1+\epsilon]\}.$$

We claim that (1.4) holds if and only if

$$\lim_{n\to\infty} U_n(D^c_{n,\epsilon,k_n}) = 0, \text{ for all } \epsilon > 0. \tag{1.7}$$

We first show the sufficiency of (1.7). Since $\lim_{n\to\infty} U_n(D_{n,\epsilon,k_n}) = 1$, it follows from the definition of D_{n,ϵ,k_n} that $\liminf_{n\to\infty} \mu_{n;k_n}(D_{n,\epsilon,k_n}) \geq 1-\epsilon$, and thus

$$\limsup_{n\to\infty} \mu_{n;k_n}(D^c_{n,\epsilon,k_n}) \leq \epsilon. \tag{1.8}$$

Thus, for any $A_n \subset S_n$, we have

$$\begin{aligned} |U_n(A_n) - \mu_{n;k_n}(A_n)| &\leq |U_n(A_n \cap D_{n,\epsilon,k_n}) - \mu_{n;k_n}(A_n \cap D_{n,\epsilon,k_n})| \\ &+ |U_n(A_n \cap D^c_{n,\epsilon,k_n}) - \mu_{n;k_n}(A_n \cap D^c_{n,\epsilon,k_n})|. \end{aligned} \tag{1.9}$$

By the definition of D_{n,ϵ,k_n}, the first term on the right hand side of (1.9) is no greater than ϵ. By (1.7) and (1.8), the lim sup of the second term on the right hand side of (1.9) is no greater than ϵ. Since $\epsilon > 0$ is arbitrary, we conclude that

$$\lim_{n\to\infty} |U_n(A_n) - \mu_{n;k_n}(A_n)| = 0.$$

Since the sets $\{A_n\}$ are arbitrary, this proves (1.4).

We now show the necessity of (1.7). Let

$$C_{n,\epsilon,k_n} = \{\sigma \in S_n : \frac{\mu_{n;k_n}(\sigma)}{U_n(\sigma)} < 1-\epsilon\}.$$

If (1.7) does not hold, then we may assume without loss of generality that there exists a $\delta > 0$ and an $\epsilon_0 > 0$ such that $U_n(C_{n,\epsilon_0,k_n}) \geq \delta$, for all n. But then, from the definition of C_{n,ϵ_0,k_n}, it follows that $\mu_{n;k_n}(C_{n,\epsilon_0,k_n}) < (1-\epsilon_0)U_n(C_{n,\epsilon_0,k_n})$, and thus $|U_n(C_{n,\epsilon_0,k_n}) - \mu_{n;k_n}(C_{n,\epsilon_0,k_n})| > \epsilon_0\delta$, for all n, which shows that (1.4) does not hold.

To complete the proof of the proposition then, it remains to prove that (1.7) holds if and only if the law of large numbers holds. Using (1.3) for the first equality below, and using (1.1) for the second equality, we have

$$\begin{aligned} U_n(D^c_{n,\epsilon,k_n}) &= P(\frac{1}{\binom{n}{k_n}}\frac{k_n!}{n!}Z_{n,k_n} \notin [\frac{1-\epsilon}{n!}, \frac{1+\epsilon}{n!}]) = \\ &= P(|\frac{Z_{n,k_n}}{EZ_{n,k_n}} - 1| > \epsilon). \end{aligned} \tag{1.10}$$

From (1.10), it follows that (1.7) holds if and only if the law of large numbers holds for Z_{n,k_n}. □

2. Proof of Proposition 2. From the definition of $Z_{n,k}$, it follows that

$$EZ^2_{n,k} = \sum E1_{B^n_{x_1,x_2,\ldots,x_k}}1_{B^n_{y_1,y_2,\ldots,y_k}}, \tag{2.1}$$

where the sum is over the $\binom{n}{k}^2$ pairs $B^n_{x_1,x_2,\ldots,x_k}, B^n_{y_1,y_2,\ldots,y_k}$ with $x_1 < x_2 < ... < x_k$ and $y_1 < y_2 < ... < y_k$. It turns out that $E1_{B^n_{x_1,x_2,\ldots,x_k}}1_{B^n_{y_1,y_2,\ldots,y_k}}$ depends rather intimately on the relative positions of $\{x_1, x_2, ..., x_k\}$ and $\{y_1, y_2, ..., y_k\}$.

Let $j \in \{0, 1, ..., k\}$. For any particular subset $A \subset \{1, 2, ..., n\}$ satisfying $|A| = 2k-j$, there are $\binom{2k-j}{j}\binom{2k-2j}{k-j}$ ordered pairs of sets $B^n_{x_1,x_2,\ldots,x_k}, B^n_{y_1,y_2,\ldots,y_k}$ for which $\{x_1, x_2, ..., x_k\} \cup \{y_1, y_2, ..., y_k\} = A$. Of course, it follows that $|\{x_1, x_2, ..., x_k\} \cap \{y_1, y_2, ..., y_k\}| = j$. We will say that such a pair $B^n_{x_1,x_2,\ldots,x_k}, B^n_{y_1,y_2,\ldots,y_k}$ *corresponds to* A. For any pair $B^n_{x_1,x_2,\ldots,x_k}, B^n_{y_1,y_2,\ldots,y_k}$ corresponding to A, there exist numbers $\{l_r\}_{r=0}^j$ and $\{m_r\}_{r=0}^j$ such that exactly l_0 elements of $\{x_1, x_2, ..., x_k\}$ and m_0 elements of $\{y_1, y_2, ..., y_k\}$ strictly precede the first element that is common to $\{x_1, x_2, ..., x_k\}$ and $\{y_1, y_2, ..., y_k\}$, exactly l_r elements of $\{x_1, x_2, ..., x_k\}$ and m_r elements of $\{y_1, y_2, ..., y_k\}$ fall strictly between the r-th and the $(r+1)$-th element that is common to $\{x_1, x_2, ..., x_k\}$ and $\{y_1, y_2, ..., y_k\}$, for $r = 1, 2, ..., j-1$, and exactly

l_j elements of $\{x_1, x_2, ..., x_k\}$ and m_j elements of $\{y_1, y_2, ..., y_k\}$ strictly follow the j-th and final element that is common to $\{x_1, x_2, ..., x_k\}$ and $\{y_1, y_2, ..., y_k\}$. We will refer to the numbers $\{l_r\}_{r=0}^{j}$ and $\{m_r\}_{r=0}^{j}$ as the "interlacing numbers" for the pair $B^n_{x_1,x_2,...,x_k}, B^n_{y_1,y_2,...,y_k}$.

Lemma 1. *Let the pair* $B^n_{x_1,x_2,...,x_k}, B^n_{y_1,y_2,...,y_k}$ *satisfy*

$$|\{x_1, x_2, ..., x_k\} \cap \{y_1, y_2, ..., y_k\}| = j$$

and let $\{l_r\}_{r=0}^{j}$, $\{m_r\}_{r=0}^{j}$ *be the corresponding interlacing numbers. Then*

$$E1_{B^n_{x_1,x_2,...,x_k}} 1_{B^n_{y_1,y_2,...,y_k}} = \frac{1}{(2k-j)!} \prod_{r=0}^{j} \frac{(l_r+m_r)!}{l_r! m_r!}. \tag{2.2}$$

Proof. Without loss of generality, we may assume that $n = 2k - j$, since only the relative positions of the $2k - j$ distinct points in the set $\{x_1, x_2, ..., x_k, y_1, y_2, ..., y_k\}$ are relevant. Thus, we are considering permutations from S_{2k-j}. For each $r = 0, 1, ..., j$, consider the $l_r + m_r$ positions between common elements (of course, for $r = 0$ and $r = j$, "between" is not the correct word). There are $(l_r + m_r)!$ ways to fill these positions. However, if we require that the l_r positions reserved for the x-chain and the m_r positions reserved for the y-chain be in increasing order, this reduces the number of ways to $\frac{(l_r+m_r)!}{l_r!m_r!}$. Thus, there are $\prod_{r=0}^{j} \frac{(l_r+m_r)!}{l_r!m_r!}$ ways to fill all the positions so that $B^n_{x_1,x_2,...,x_k} \cap B^n_{y_1,y_2,...,y_k}$ will occur, and of course, all together there are $(2k - j)!$ ways to fill the positions with no restrictions. □

We now complete the proof of the proposition. Simple combinatorial considerations show that out of the $\binom{2k-j}{j}\binom{2k-2j}{k-j}$ pairs $B^n_{x_1,x_2,...,x_k}$, $B^n_{y_1,y_2,...,y_k}$ corresponding to a set A satisfying $|A| = 2k - j$, there are $\prod_{r=0}^{j} \binom{l_r+m_r}{l_r} = \prod_{r=0}^{j} \frac{(l_r+m_r)!}{l_r!m_r!}$ of them with the interlacing numbers $\{l_r\}_{r=0}^{j}$, $\{m_r\}_{r=0}^{j}$. Using this fact along with (2.1), (2.2) and the fact that there are $\binom{n}{2k-j}$ distinct subsets $A \subset \{1, 2, ..., n\}$ such that $|A| = 2k - j$, we obtain the formula for $EZ^2_{n,k}$ in Proposition 2. □

3. Proof of Theorem 1. Similar to (2.1), we can write the variance of Z_{n,k_n} in the form

(3.1)

$$Var(Z_{n,k_n}) = \sum E1_{B^n_{x_1,x_2,...,x_{k_n}}} 1_{B^n_{y_1,y_2,...,y_{k_n}}} - \sum E1_{B^n_{x_1,x_2,...,x_{k_n}}} E1_{B^n_{y_1,y_2,...,y_{k_n}}},$$

where the sum is over the $\binom{n}{k_n}^2$ pairs $B^n_{x_1,x_2,...,x_{k_n}}, B^n_{y_1,y_2,...,y_{k_n}}$. By (2.2),

$$E1_{B^n_{x_1,x_2,...,x_{k_n}}}1_{B^n_{y_1,y_2,...,y_{k_n}}} - E1_{B^n_{x_1,x_2,...,x_{k_n}}}E1_{B^n_{y_1,y_2,...,y_{k_n}}} = 0, \text{ if } \tag{3.2}$$

$\{x_1, x_2, ..., x_{k_n}\}$ and $\{y_1, y_2, ..., y_{k_n}\}$ are disjoint.

The number of pairs $\{x_1, x_2, ..., x_{k_n}\}$, $\{y_1, y_2, ..., y_{k_n}\}$ which are not disjoint is equal to $\binom{n}{k_n}^2 - \binom{n}{k_n}\binom{n-k_n}{k_n}$. If $k_n = o(n^{\frac{1}{2}})$, then a simple calculation reveals that $\binom{n}{k_n}^2 - \binom{n}{k_n}\binom{n-k_n}{k_n} = o(\binom{n}{k_n}^2)$. Thus,

$$\sum_{\{x_1,x_2,...,x_{k_n}\}\cap\{y_1,y_2,...,y_{k_n}\}\neq\emptyset} E1_{B^n_{x_1,x_2,...,x_{k_n}}}E1_{B^n_{y_1,y_2,...,y_{k_n}}} \tag{3.3}$$
$$= \left(\binom{n}{k_n}^2 - \binom{n}{k_n}\binom{n-k_n}{k_n}\right)\frac{1}{(k_n!)^2} = o(\frac{\binom{n}{k_n}^2}{(k_n!)^2}) = o((EZ_{n,k_n})^2),$$

where the final equality follows from (1.1). On the other hand, if it is not true that $k_n = o(n^{\frac{1}{2}})$, then the left hand side of (3.3) will be $O((EZ_{n,k_n})^2)$. In light of this last remark along with (3.1)-(3.3), the theorem will be proved once we show that

$$\frac{\sum_{\{x_1,x_2,...,x_{k_n}\}\cap\{y_1,y_2,...,y_{k_n}\}\neq\emptyset} E1_{B^n_{x_1,x_2,...,x_{k_n}}}1_{B^n_{y_1,y_2,...,y_{k_n}}}}{(EZ_{n,k_n})^2} = O(\frac{k_n^{\frac{5}{2}}}{n}), \tag{3.4-a}$$

if k_n is as in part (i);

$$\frac{\sum_{\{x_1,x_2,...,x_{k_n}\}\cap\{y_1,y_2,...,y_{k_n}\}\neq\emptyset} E1_{B^n_{x_1,x_2,...,x_{k_n}}}1_{B^n_{y_1,y_2,...,y_{k_n}}}}{(EZ_{n,k_n})^2} \tag{3.4-b}$$

is bounded from 0 and ∞ if k_n is as in part (ii);

$$\lim_{n\to\infty}\frac{\sum_{\{x_1,x_2,...,x_{k_n}\}\cap\{y_1,y_2,...,y_{k_n}\}\neq\emptyset} E1_{B^n_{x_1,x_2,...,x_{k_n}}}1_{B^n_{y_1,y_2,...,y_{k_n}}}}{(EZ_{n,k_n})^2} = \infty, \tag{3.4-c}$$

if k_n is as in part (iii).

By Proposition 2 and its proof, it follows that

$$\sum_{\{x_1,x_2,...,x_{k_n}\}\cap\{y_1,y_2,...,y_{k_n}\}\neq\emptyset} E1_{B^n_{x_1,x_2,...,x_{k_n}}}1_{B^n_{y_1,y_2,...,y_{k_n}}}$$
$$= \sum_{j=1}^{k_n}\binom{n}{2k_n - j}\frac{1}{(2k_n - j)!}A(k_n - j, j), \text{ where } A(N, j) \text{ is as in (1.5).}$$

Using this with (3.4) and the fact that $EZ_{n,k_n} = \frac{\binom{n}{k_n}}{k_n!}$, the proof will be complete if we show that

(3.5-a)
$$\frac{(k_n!)^2}{\binom{n}{k_n}^2}\sum_{j=1}^{k_n}\binom{n}{2k_n - j}\frac{1}{(2k_n - j)!}A(k_n - j, j) = O(\frac{k_n^{\frac{5}{2}}}{n}), \text{ if } k_n \text{ is as in part (i);}$$

(3.5-b) $$\frac{(k_n!)^2}{\binom{n}{k_n}^2}\sum_{j=1}^{k_n}\binom{n}{2k_n-j}\frac{1}{(2k_n-j)!}A(k_n-j,j) \text{ is bounded from 0 and } \infty$$
if k_n is as in part (ii) ;

(3.5-c)
$$\lim_{n\to\infty}\frac{(k_n!)^2}{\binom{n}{k_n}^2}\sum_{j=1}^{k_n}\binom{n}{2k_n-j}\frac{1}{(2k_n-j)!}A(k_n-j,j)=\infty,\text{ if } k_n \text{ is as in part (iii).}$$

It remains therefore to analyze the left hand side of (3.5). In the next section we will prove the following key estimates:

Lemma 2. *For each $\rho \in (0,\infty)$, there exists a constant $C_\rho > 0$ such that*

$$A(N,j) \le C_\rho^j \frac{j^{\frac{1}{2}}}{\Gamma(\frac{j+1}{2})}(2N)^{\frac{j}{2}}\binom{2N}{N}^2, \text{ for } j,N\ge 1 \text{ and } \frac{j}{N}\le \rho.$$

In particular, since $A(N,j)$ is increasing in N, one has

$$A(k-j,j)\le A(k,j)\le C_1^j\frac{j^{\frac{1}{2}}}{\Gamma(\frac{j+1}{2})}(2k)^{\frac{j}{2}}\binom{2k}{k}^2, \text{ for } j,k\ge 1 \text{ and } j\le k.$$

Lemma 3. *There exists a constant $C>0$ such that*

$$A(k-1,1)\ge C(2k-2)^{\frac{1}{2}}\binom{2k-2}{k-1}^2.$$

We now use Lemma 2 to show that (3.5-a) and the part of (3.5-b) concerning boundedness from ∞ hold. Afterwards, we will use Lemma 3 to show that (3.5-c) and the part of (3.5-b) concerning boundedness from 0 hold.

In light of Lemma 2, it suffices to show that (3.5-a) and the part of (3.5-b) concerning boundedness from ∞ hold with $A(k_n-j,j)$ replaced by $C_1^j\frac{j^{\frac{1}{2}}}{\Gamma(\frac{j+1}{2})}(2k_n)^{\frac{j}{2}}\binom{2k_n}{k_n}^2$. Letting

$$B(n,k_n,j)=\frac{(k_n!)^2}{\binom{n}{k_n}^2}\binom{n}{2k_n-j}\frac{1}{(2k_n-j)!}\binom{2k_n}{k_n}^2(2k_n)^{\frac{j}{2}},$$

it follows that (3.5-a) (respectively the part of (3.5-b) concerning boundedness from ∞) will hold if we show that

$$\sum_{j=1}^{k_n}B(n,k_n,j)\frac{C_1^j j^{\frac{1}{2}}}{\Gamma(\frac{j+1}{2})}$$

is $O(\frac{k_n^{\frac{5}{2}}}{n})$, if k_n is as in part (i) (respectively, bounded if k_n is as in part (ii)). Simplifying and making some cancellations, we have

$$B(n,k_n,j) = \frac{((n-k_n)!)^2}{n!(n-2k_n+j)!}\left(\frac{(2k_n)!}{(2k_n-j)!}\right)^2 (2k_n)^{\frac{j}{2}}. \tag{3.6}$$

We have

$$b_1 n^{-j} \le \frac{((n-k_n)!)^2}{n!(n-2k_n+j)!} \le b_2 n^{-j},\ j=1,...,k_n, \tag{3.7}$$

for positive constants b_1, b_2. (For the lower bound, we have used the fact that k_n is of an order not larger than $n^{\frac{1}{2}}$. The upper bound holds as long as $k_n \le cn$ for some $c<1$.) We also have

$$k_n^j \le \frac{(2k_n)!}{(2k_n-j)!} \le (2k_n)^j,\ j=1,...,k_n. \tag{3.8}$$

From (3.6)-(3.8) we have

$$B(n,k_n,j)\frac{C_1^j j^{\frac{1}{2}}}{\Gamma(\frac{j+1}{2})} \le b_2 n^{-j}(2k_n)^{2j}\ (2k_n)^{\frac{j}{2}}\ \frac{C_1^j j^{\frac{1}{2}}}{\Gamma(\frac{j+1}{2})} \le \frac{j^{\frac{1}{2}}C^j}{\Gamma(\frac{j+1}{2})}(n^{-1}k_n^{\frac{5}{2}})^j, \tag{3.9}$$

for some $C>0$. Since $\sum_{j=1}^{\infty}\frac{j^{\frac{1}{2}}C^j}{\Gamma(\frac{j+1}{2})} < \infty$, it follows from (3.9) that

$$\sum_{j=1}^{k_n} B(n,k_n,j)\frac{C_1^j j^{\frac{1}{2}}}{\Gamma(\frac{j+1}{2})}$$

is $O(\frac{k_n^{\frac{5}{2}}}{n})$ as $n\to\infty$, if k_n is as in part (i), and is bounded if k_n is as in part (ii). This proves (3.5-a) and the part of (3.5-b) concerning boundedness from ∞.

We now turn to (3.5-c) and the part of (3.5-b) concerning boundedness from 0. The term in (3.5-b,c) corresponding to $j=1$ is $\frac{(k_n!)^2}{\binom{n}{k_n}^2}\binom{n}{2k_n-1}\frac{1}{(2k_n-1)!}A(k_n-1,1)$. Define $C(n,k_n) = \frac{(k_n!)^2}{\binom{n}{k_n}^2}\binom{n}{2k_n-1}\frac{1}{(2k_n-1)!}(2k_n-2)^{\frac{1}{2}}\binom{2k_n-2}{k_n-1}^2$. Using the bound on $A(k_n-1,1)$ from Lemma 3, it follows that for the part of (3.5-b) concerning boundedness from 0, it is enough to show that $\liminf_{n\to\infty} C(n,k_n) > 0$, when k_n is as in part (ii), and for (3.5-c) it is enough to show that $\lim_{n\to\infty} C(n,k_n) = \infty$, when k_n is as in part (iii). Simplifying and making some cancellations, we have

$$C(n,k_n) = \frac{((n-k_n)!)^2}{n!(n-2k_n+1)!}k_n^4(2k_n-1)^{-2}(2k_n-2)^{\frac{1}{2}}.$$

Using this with (3.7) gives

$$C(n,k_n) \geq b_1 n^{-1} k_n^4 (2k_n-1)^{-2}(2k_n-2)^{\frac{1}{2}}.$$

Thus, the above stated inequalities indeed hold. □

4. Proofs of Lemmas 2 and 3.

Proof of Lemma 2. The first step of the proof is to develop a probabilistic representation for $A(N,j)$. Fix $j \geq 1$ and $N \geq 1$. Consider two rows each containing $2N$ spaces. Randomly fill each of the two rows with N blue balls and N white balls. Define $X_0 = 0$, and then for $m = 1, 2, ..., 2N$, use the balls in the first row to define X_m as the number of blue balls in the first m spaces minus the number of white balls in the first m spaces. Define Y_m the same way using the balls in the second row. Then $\{X_m\}$ and $\{Y_m\}$ are independent, and as is well known, each one has the distribution of the simple, symmetric one-dimensional random walk, conditioned to return to 0 at the $2N$-th step. Let $U_m = \frac{X_m+Y_m}{2}$ and $V_m = \frac{X_m-Y_m}{2}$. Then (U_m, V_m) has the distribution of the standard, simple, symmetric two-dimensional random walk (jumping one unit in each of the four possible directions with probability $\frac{1}{4}$), starting from the origin and conditioned to return to the origin at the $2N$-th step. To see this, let $\{\mathcal{X}_m\}$ and $\{\mathcal{Y}_m\}$ be independent copies of the unconditioned, simple, symmetric one-dimensional random walk starting from the origin, and let $\mathcal{U}_m = \frac{\mathcal{X}_m+\mathcal{Y}_m}{2}$ and $\mathcal{V}_m = \frac{\mathcal{X}_m-\mathcal{Y}_m}{2}$. Then clearly, $\{\mathcal{U}_m, \mathcal{V}_m\}$ is the unconditioned, simple symmetric two-dimensional random walk starting from the origin. Now $\{U_m, V_m\}$ is equal to $\{\mathcal{U}_m, \mathcal{V}_m\}$ conditioned on $\mathcal{X}_{2N} = \mathcal{Y}_{2N} = 0$, or equivalently, conditioned on $\mathcal{U}_{2N} = \mathcal{V}_{2N} = 0$.

The total number of possible ways of placing N blue balls and N white balls in the first row, and the same number of such balls in the second row is $\binom{2N}{N}^2$. For the moment, fix a set $\{s_r\}_{r=0}^j$ of $j+1$ nonnegative integers satisfying $\sum_{r=0}^j s_r = 2N$. Let $t_r = \sum_{i=0}^r s_i$. Let $D_{s_0,s_1,...,s_j}$ denote the event $\{V_{t_0} = V_{t_1} = ... = V_{t_j} = 0\} = \{X_{t_0} = Y_{t_0}, X_{t_1} = Y_{t_1}, ..., X_{t_j} = Y_{t_j}\}$. Now for any sequence $\{l_r\}_{r=0}^j$ satisfying $l_r \leq s_r$ and $\sum_{r=0}^j l_r = N$, the probability of the event $\{X_{t_r} = \sum_{i=0}^r (l_i - (s_i -$

$l_i)),\ r=0,1,...,j\} = \{X_{t_r} = \sum_{i=0}^{r}(2l_i - s_i),\ r=0,1,...,j\}$ is $\binom{2N}{N}^{-1}\prod_{r=0}^{j}\binom{s_r}{l_r}$, and thus the probability of the event $\{X_{t_r} = Y_{t_r} = \sum_{i=0}^{r}(2l_i - s_i),\ r=0,1,...,j\}$ is $\binom{2N}{N}^{-2}(\prod_{r=0}^{j}\binom{s_r}{l_r})^2$. Summing now over all possible $\{l_r\}_{r=0}^{j}$ as above, it follows that

(4.1)
$$P(D_{s_0,s_1,...,s_j}) = \frac{1}{\binom{2N}{N}^2}\sum_{\substack{\sum_{r=0}^{j} l_r = N\\ l_r\le s_r}}\prod_{r=0}^{j}\binom{s_r}{l_r}^2.$$

Letting $m_r = s_r - l_r \ge 0$, one sees that the term involving the summation on the right hand side of (4.1) can be written as $\sum\prod_{r=0}^{j}\left(\frac{(l_r+m_r)!}{l_r!m_r!}\right)^2$, where the sum is over all $\{l_r\}_{r=0}^{j}$ and $\{m_r\}_{r=0}^{j}$ satisfying $\sum_{r=0}^{j} l_r = \sum_{r=0}^{j} m_r = N$, and $l_r + m_r = s_r$, for $r = 0,1,...,j$. Thus, summing (4.1) over all the possible choices of $\{s_r\}_{r=0}^{j}$, we obtain

(4.2)
$$\sum_{\sum_{r=0}^{j} s_r = 2N} P(D_{s_0,s_1,...,s_j}) = \frac{A(N,j)}{\binom{2N}{N}^2}.$$

The next step of the proof is to estimate $P(D_{s_0,s_1,...,s_j})$. For this we will need several lemmas.

Lemma 4. *Let $\{Z_n\}_{n=0}^{\infty}$ be a one-dimensional random walk which takes jumps of ±1 with probability $\frac{1}{4}$ each, and remains in its place with probability $\frac{1}{2}$. Then there exit constants $C_1, C_2 > 0$ such that*

$$\frac{C_1}{n^{\frac{1}{2}}} \le P(Z_n = 0|Z_0 = 0) \le \frac{C_2}{n^{\frac{1}{2}}},\ \textit{for } n \ge 1.$$

Proof. A direct calculation gives

$$P(Z_n = 0|Z_0 = 0) = \sum_{i=0}^{[\frac{n}{2}]}(\frac{1}{2})^{n-2i}(\frac{1}{4})^{2i}\binom{n}{i}\binom{n-i}{i}.$$

We rewrite this as

(4.3)
$$P(Z_n = 0|Z_0 = 0) = (\frac{1}{2})^n\sum_{i=0}^{[\frac{n}{2}]}(\frac{1}{2})^{2i}\binom{2i}{i}\frac{\binom{n}{i}\binom{n-i}{i}}{\binom{2i}{i}} = (\frac{1}{2})^n\sum_{i=0}^{[\frac{n}{2}]}(\frac{1}{2})^{2i}\binom{2i}{i}\binom{n}{2i}.$$

By Stirling's approximation, there exist positive constants c_1, c_2 such that

$$\frac{c_1}{\sqrt{i+1}} \leq (\frac{1}{2})^{2i}\binom{2i}{i} \leq \frac{c_2}{\sqrt{i+1}}, \ i = 0, 1, ... \tag{4.4}$$

Thus, from (4.3) and (4.4) we have

$$c_1 \sum_{i=0}^{[\frac{n}{2}]} \frac{1}{\sqrt{i+1}}\binom{n}{2i}(\frac{1}{2})^n \leq P(Z_n = 0|Z_0 = 0) \leq c_2 \sum_{i=0}^{[\frac{n}{2}]} \frac{1}{\sqrt{i+1}}\binom{n}{2i}(\frac{1}{2})^n. \tag{4.5}$$

Now let S_n be a random variable distributed according to Binom$(n, \frac{1}{2})$. Then we have

$$\sum_{i=0}^{[\frac{n}{2}]} \frac{1}{\sqrt{i+1}}\binom{n}{2i}(\frac{1}{2})^n = E(\frac{1}{2}S_n + 1)^{-\frac{1}{2}} \ 1_{\{S_n \text{ is even}\}}. \tag{4.6}$$

By standard large deviations estimates, $P(|\frac{S_n}{n} - \frac{1}{2}| > \epsilon)$ decays exponentially in n for each $\epsilon > 0$. Using this along with (4.5) and (4.6) and leaving to the reader the little argument to accommodate the requirement in (4.6) that S_n be even, we conclude that there exist constants $C_1, C_2 > 0$ such that

$$\frac{C_1}{\sqrt{n+1}} \leq P(Z_n = 0|Z_0 = 0) \leq \frac{C_2}{\sqrt{n+1}}.$$

□

Lemma 5. *Let $\{\hat{Z}_n\}_{n=0}^{\infty}$ be a simple, symmetric one-dimensional random walk.*

i. There exists a constant $C_0 > 0$ such that

$$P(\hat{Z}_n = 0|\hat{Z}_0 = a) \leq \frac{C_0}{\sqrt{n}}\exp(-\frac{a^2}{2n}), \ \textit{for all } a \in Z \textit{ and all } n \geq 1.$$

ii. Let $L > 0$. There exists a constant $c_L > 0$ such that for all sufficiently large n,

$$P(\hat{Z}_{2n} = 0|\hat{Z}_0 = 2a) \geq \frac{c_L}{\sqrt{n}}\exp(-\frac{a^2}{n}), \ \textit{for all } a \in Z \textit{ satisfying } |a| \leq Ln^{\frac{1}{2}}.$$

Proof. The lemma follows from the local central limit theorem. It can be proved via a direct calculation, using Stirling's approximation. (See, for example, [4, page 65].) □

Lemma 6. *Let $\{\hat{X}_n, \hat{Y}_n\}_{n=0}^{\infty}$ be a simple, symmetric two-dimensional random walk.*

i. There exist constants $c_1, c_2 > 0$ *such that*

$$P((\hat{X}_n, \hat{Y}_n) = (0,0) | (\hat{X}_0, \hat{Y}_0) = (a,0)) \le \frac{c_1}{n} \exp(-\frac{c_2 a^2}{n}), \text{ for all } a \in Z \text{ and all } n \ge 1.$$

ii. There exists a constant $c_3 > 0$ *such that*

$$P((\hat{X}_{2n}, \hat{Y}_{2n}) = (0,0) | (\hat{X}_0, \hat{Y}_0) = (0,0)) \ge \frac{c_3}{n}, \text{ for all } n \ge 1.$$

Proof. Let H_n and V_n denote respectively the number of horizontal and the number of vertical steps made by the random walk $\{(\hat{X}_\cdot, \hat{Y}_\cdot)\}$ during its first n steps. Then we have

$$\begin{aligned} &P((\hat{X}_n, \hat{Y}_n) = (0,0) | (\hat{X}_0, \hat{Y}_0) = (a,0)) = \\ &\sum_{j+k=n} P(\hat{Z}_j = 0 | \hat{Z}_0 = a) P(\hat{Z}_k = 0 | \hat{Z}_0 = 0) \times P(H_n = j, V_n = k), \end{aligned} \tag{4.7}$$

where $\{\hat{Z}_n\}$ is as in Lemma 5. Since H_n and V_n are each distributed like Binom$(n, \frac{1}{2})$, a standard large deviations estimate gives

$$P(H_n \ge \frac{1}{4} n, V_n \ge \frac{1}{4} n) \ge 1 - \frac{1}{C} \exp(-Cn), \tag{4.8}$$

for some $C > 0$. Since $\frac{1}{\sqrt{j}} \exp(-\frac{a^2}{2j}) \le \frac{2}{\sqrt{n}} \exp(-\frac{a^2}{2n})$, for $\frac{1}{4} n \le j \le n$, it follows from (4.7), (4.8) and Lemma 5-i that

$$P((\hat{X}_n, \hat{Y}_n) = (0,0) | (\hat{X}_0, \hat{Y}_0) = (a,0)) \le \frac{4C_0^2}{n} \exp(-\frac{a^2}{n}) + \frac{1}{C} \exp(-Cn). \tag{4.9}$$

Choosing c_1 sufficiently large and $c_2 > 0$ sufficiently small, part (i) follows from (4.9) along with the fact that we need only consider $|a| \le n$.

For part (ii), note that

$$\begin{aligned} &P((\hat{X}_{2n}, \hat{Y}_{2n}) = (0,0) | (\hat{X}_0, \hat{Y}_0) = (0,0)) = \\ &\sum_{2j+2k=2n} P(\hat{Z}_{2j} = 0 | \hat{Z}_0 = 0) P(\hat{Z}_{2k} = 0 | \hat{Z}_0 = 0) \times P(H_n = 2j, V_n = 2k). \end{aligned}$$

Also, we have $P(H_n$ and V_n are even, $H_n \ge \frac{1}{4} n, V_n \ge \frac{1}{4} n) \ge C$, for some $C > 0$ independent of n. Finally, $P(\hat{Z}_{2j} = 0 | \hat{Z}_0 = 0)$ can be bounded from below as in Lemma 5-ii. Part (ii) follows from these observations. □

We can now estimate $P(D_{s_0,s_1,...,s_j})$.

Lemma 7. *Let* $\hat{s}_r = s_r + 1$. *There exists a constant* $c > 0$ *such that*

$$P(D_{s_0,s_1,...,s_j}) \le (2N+1)^{\frac{1}{2}} c^{j+1} (\hat{s}_0 \hat{s}_1 ... \hat{s}_j)^{-\frac{1}{2}}. \tag{4.10}$$

Proof. Let $\{\hat{X}_n, \hat{Y}_n\}_{n=0}^{\infty}$ be a simple, symmetric two-dimensional random walk. Recalling that $t_j = 2N$, it follows by definition that

$$P(D_{s_0,s_1,...,s_j}) = P(\hat{Y}_{t_0} = \hat{Y}_{t_1} = \cdots = \hat{Y}_{t_{j-1}} = 0 | \hat{X}_0 = \hat{Y}_0 = \hat{X}_{2N} = \hat{Y}_{2N} = 0). \tag{4.11}$$

By the Markov property, we have

$$\begin{aligned} &P(\hat{Y}_{t_0} = \hat{Y}_{t_1} = \cdots = \hat{Y}_{t_{j-1}} = 0 | \hat{X}_0 = \hat{Y}_0 = \hat{X}_{2N} = \hat{Y}_{2N} = 0) = \\ &P(\hat{Y}_{t_0} = 0 | \hat{X}_0 = \hat{Y}_0 = 0) \cdot P(\hat{Y}_{t_1} = 0 | \hat{Y}_{t_0} = \hat{X}_0 = \hat{Y}_0 = 0) \times \cdots \times \\ &P(\hat{Y}_{t_{j-1}} = 0 | \hat{Y}_{t_0} = \cdots = \hat{Y}_{t_{j-2}} = \hat{X}_0 = \hat{Y}_0 = 0) \times \\ &\frac{P(\hat{X}_{2N} = \hat{Y}_{2N} = 0 | \hat{Y}_{t_0} = \cdots \hat{Y}_{t_{j-1}} = \hat{X}_0 = \hat{Y}_0 = 0)}{P(\hat{X}_{2N} = \hat{Y}_{2N} = 0 | \hat{X}_0 = \hat{Y}_0 = 0)}. \end{aligned} \tag{4.12}$$

Note that the process $\{\hat{Y}_n\}$ in isolation is a one-dimensional random walk distributed according to the distribution of $\{Z_n\}$ in Lemma 4. Thus, letting $t_{-1} = 0$, we have from (4.12)

$$\begin{aligned} &P(\hat{Y}_{t_0} = \hat{Y}_{t_1} = \cdots = \hat{Y}_{t_{j-1}} = 0 | \hat{X}_0 = \hat{Y}_0 = \hat{X}_{2N} = \hat{Y}_{2N} = 0) = \\ &\frac{P(\hat{X}_{2N} = \hat{Y}_{2N} = 0 | \hat{Y}_{t_0} = \cdots \hat{Y}_{t_{j-1}} = \hat{X}_0 = \hat{Y}_0 = 0)}{P(\hat{X}_{2N} = \hat{Y}_{2N} = 0 | \hat{X}_0 = \hat{Y}_0 = 0)} \times \prod_{k=0}^{j-1} P(Z_{t_k} = 0 | Z_{t_{k-1}} = 0). \end{aligned} \tag{4.13}$$

Recall that $t_k - t_{k-1} = s_k$ and $s_j = 2N - t_{j-1}$. Let $s'_k = s_k$, if $s_k \ge 1$, and $s'_k = 1$, if $s_k = 0$. Since $P(Z_{t_k} = 0 | Z_{t_{k-1}} = 0) = P(Z_{t_k - t_{k-1}} = 0 | Z_0 = 0)$ it follows from Lemma 4 that $P(Z_{t_k} = 0 | Z_{t_{k-1}} = 0) \le \frac{C_2}{\sqrt{s'_k}}$. From Lemma 6-ii, it follows that $P(\hat{X}_{2N} = \hat{Y}_{2N} = 0 | \hat{X}_0 = \hat{Y}_0 = 0) \ge \frac{c_3}{N}$. Using these facts along with (4.11) and (4.13), it follows that if we show that

$$P(\hat{X}_{2N} = \hat{Y}_{2N} = 0 | \hat{Y}_{t_0} = \cdots \hat{Y}_{t_{j-1}} = \hat{X}_0 = \hat{Y}_0 = 0) \le \frac{C^j}{N^{\frac{1}{2}} (2N - t_{j-1})^{\frac{1}{2}}}, \tag{4.14}$$

for some $C > 0$ and $s_j = 2N - t_{j-1} \geq 1$, then we will obtain (4.10) with $\hat{s}_r$ replaced by s'_r. By increasing the constant c in (4.10), one can always replace s'_r by $\hat{s}_r$. Thus, it remains to prove (4.14).

By the Markov property, we have

$$\begin{aligned} & P(\hat{X}_{2N} = \hat{Y}_{2N} = 0 | \hat{Y}_{t_0} = \cdots \hat{Y}_{t_{j-1}} = \hat{X}_0 = \hat{Y}_0 = 0) = \\ & \sum_{a \in R} P(\hat{X}_{2N-t_{j-1}} = \hat{Y}_{2N-t_{j-1}} = 0 | \hat{X}_0 = a, \hat{Y}_0 = 0) \times \nu(a), \end{aligned} \tag{4.15}$$

where

$$\nu(a) = P(\hat{X}_{t_{j-1}} = a | \hat{Y}_{t_0} = \cdots \hat{Y}_{t_{j-1}} = \hat{X}_0 = \hat{Y}_0 = 0).$$

As in the proof of Lemma 6, let H_n denote the number of horizontal steps taken by the random walk $\{\hat{X}(\cdot), \hat{Y}(\cdot)\}$ during its first n steps. Let

$$\mu(m) = P(H_{t_{j-1}} = m | \hat{Y}_{t_0} = \cdots \hat{Y}_{t_{j-1}} = \hat{X}_0 = \hat{Y}_0 = 0),$$

and let W be distributed like μ. Then ν is distributed like $\hat{Z}_W$, where $\{\hat{Z}_n\}$ is a simple, symmetric one-dimensional random walk, starting from 0 and independent of W. We will show later that for some $\gamma, C > 0$,

$$\begin{aligned} & \mu([0, \gamma t_{j-1}]) = P(H_{t_{j-1}} \leq \gamma t_{j-1} | \hat{Y}_{t_0} = \cdots \hat{Y}_{t_{j-1}} = \hat{X}_0 = \hat{Y}_0 = 0) = \\ & P(W \leq \gamma t_{j-1}) \leq \frac{C^j}{\gamma} \exp(-\gamma t_{j-1}). \end{aligned} \tag{4.16}$$

By Lemma 5-i, it follows that

$$P(\hat{Z}_n = a | \hat{Z}_0 = 0) \leq \frac{C_0}{\sqrt{\gamma t_{j-1}}} \exp(-\frac{a^2}{2t_{j-1}}), \text{ for } \gamma t_{j-1} \leq n \leq t_{j-1}. \tag{4.17}$$

From (4.16) and (4.17) we conclude that

$$\nu(a) = P(\hat{Z}_W = a) \leq \frac{C_0}{\sqrt{\gamma t_{j-1}}} \exp(-\frac{a^2}{2t_{j-1}}) + \frac{C^j}{\gamma} \exp(-\gamma t_{j-1}). \tag{4.18}$$

Since $\nu(a) = 0$, if $a > t_{j-1}$, it follows from (4.18) that

$$\nu(a) \leq \frac{k_1^j}{\sqrt{t_{j-1}}} \exp(-\frac{k_2 a^2}{t_{j-1}}), \tag{4.19}$$

for some $k_1, k_2 > 0$. From (4.15), (4.19) and Lemma 6-i, we obtain

$$\begin{aligned} & P(\hat{X}_{2N} = \hat{Y}_{2N} = 0 | \hat{Y}_{t_0} = \cdots \hat{Y}_{t_{j-1}} = \hat{X}_0 = \hat{Y}_0 = 0) \leq \\ & \sum_{a \in R} \frac{c_1}{2N - t_{j-1}} \exp(-\frac{c_2 a^2}{2N - t_{j-1}}) \frac{k_1^j}{\sqrt{t_{j-1}}} \exp(-\frac{k_2 a^2}{t_{j-1}}). \end{aligned} \tag{4.20}$$

For an appropriate $\hat{C} > 0$, the right hand side of (4.20) can be bounded from above by $\hat{C}^j \int_{-\infty}^{\infty} \frac{1}{(2N-t_{j-1})\sqrt{t_{j-1}}} \exp(-\frac{c_2 x^2}{2N-t_{j-1}}) \exp(-\frac{k_2 x^2}{t_{j-1}}) dx$. Evaluating this integral gives the estimate in (4.14). Thus, to complete the proof of the lemma, it remains to prove (4.16).

We mention that it is intuitive that $P(H_{t_{j-1}} \leq \gamma t_{j-1} | \hat{Y}_{t_0} = \cdots \hat{Y}_{t_{j-1}} = \hat{X}_0 = \hat{Y}_0 = 0) \leq P(H_{t_{j-1}} \leq \gamma t_{j-1})$, and this would then give (4.16). The intuition comes from the fact that the smaller $H_{t_{j-1}}$ is, the more moves $\{\hat{Y}_n\}$ makes, and the more moves $\{\hat{Y}_n\}$ makes, the more difficult it is for it to have the prescribed zeroes. However, a proof of this is rather complicated and quite tedious. It turns out that a rather crude estimate will suffice in order to obtain (4.16). We have

(4.21)

$$\mu([0, \gamma t_{j-1}]) = P(H_{t_{j-1}} \leq \gamma t_{j-1} | \hat{Y}_{t_0} = \cdots \hat{Y}_{t_{j-1}} = \hat{X}_0 = \hat{Y}_0 = 0) = \frac{P(H_{t_{j-1}} \leq \gamma t_{j-1}, \hat{Y}_{t_0} = \cdots \hat{Y}_{t_{j-1}} = \hat{X}_0 = \hat{Y}_0 = 0)}{P(\hat{Y}_{t_0} = \cdots \hat{Y}_{t_{j-1}} = \hat{X}_0 = \hat{Y}_0 = 0)} \leq \frac{P(H_{t_{j-1}} \leq \gamma t_{j-1})}{P(\hat{Y}_{t_0} = \cdots \hat{Y}_{t_{j-1}} = 0 | \hat{Y}_0 = 0)}.$$

By a standard large deviations estimate,

$$P(H_{t_{j-1}} \leq \gamma t_{j-1}) \leq c \exp(-l_\gamma t_{j-1}), \text{ where } \lim_{\gamma \to 0} l_\gamma = \log 2. \tag{4.22}$$

(To see this, note that $P(H_{t_{j-1}} = 0) = (\frac{1}{2})^{t_{j-1}} = \exp(-(\log 2) t_{j-1})$.) By Lemma 4, we have

$$P(\hat{Y}_{t_0} = \cdots \hat{Y}_{t_{j-1}} = 0 | \hat{Y}_0 = 0) = \prod_{k=0}^{j-1} P(\hat{Y}_{t_k} = 0 | \hat{Y}_{t_{k-1}} = 0) \geq \frac{C_1^j}{(s'_0 s'_1 \cdots s'_{j-1})^{\frac{1}{2}}}.$$

Since the $\{s'_k\}$ satisfy $\sum_{k=0}^{j-1} s'_k \leq t_{j-1} + j$, it follows that $\sup_{\{s'_k\}} s'_0 s'_1 \cdots s'_{j-1} \leq (\frac{t_{j-1}+j}{j})^j \leq \exp(t_{j-1})$. Thus,

$$P(\hat{Y}_{t_0} = \cdots \hat{Y}_{t_{j-1}} = 0 | \hat{Y}_0 = 0) \geq C_1^j \exp(-\frac{1}{2} t_{j-1}). \tag{4.23}$$

Now (4.16) follows from (4.21)-(4.23) along with the fact that $\log 2 > \frac{1}{2}$. □

We can now complete the proof of Lemma 2. From (4.2) and (4.10), we have

$$\frac{A(N,j)}{\binom{2N}{N}^2} \leq (2N+1)^{\frac{1}{2}} c^{j+1} \sum_{\sum_{r=0}^{j} s_r = 2N} (\hat{s}_0 \hat{s}_1 \ldots \hat{s}_j)^{-\frac{1}{2}}.$$

Let $\hat{S}_j = \hat{s}_j + j - 1$. Then $(\hat{s}_j)^{-\frac{1}{2}} = (\hat{S}_j)^{-\frac{1}{2}}(\frac{\hat{S}_j}{\hat{s}_j})^{\frac{1}{2}} \leq \frac{j^{\frac{1}{2}}}{(\hat{S}_j)^{\frac{1}{2}}}$. Thus it follows from the above inequality that

$$\frac{A(N,j)}{\binom{2N}{N}^2} \leq (2N+1)^{\frac{1}{2}} c^{j+1} j^{\frac{1}{2}} \sum_{\sum_{r=0}^{j} s_r = 2N} (\hat{s}_0 \hat{s}_1 ... \hat{s}_{j-1} \hat{S}_j)^{-\frac{1}{2}}. \tag{4.24}$$

The replacement of $\hat{s}_j$ by $\hat{S}_j$ was made for technical reasons which will become clear below. Making the substitutions $x_r = \frac{s_r}{2N}$ and $\hat{x}_r = \frac{\hat{s}_r}{2N}$, for $r = 0, ..., j$, and $\hat{X}_j = \frac{\hat{S}_j}{2N}$, we rewrite the right hand side of (4.24) as

(4.25)

$$(2N+1)^{\frac{1}{2}} c^{j+1} j^{\frac{1}{2}} (2N)^{\frac{j-1}{2}} \sum_{\substack{\sum_{r=0}^{j} x_r = 1 \\ (2N)x_r \text{ is a nonnegative integer}}} (\hat{x}_0 \hat{x}_1 ... \hat{x}_{j-1} \hat{X}_j)^{-\frac{1}{2}} (2N)^{-j}.$$

Let $C_{x_0, x_1, ..., x_{j-1}}$ denote the hyper-cube $\prod_{r=0}^{j-1}[x_r, x_r + \frac{1}{2N}] = \prod_{r=0}^{j-1}[x_r, \hat{x}_r]$. Consider $\cup C_{x_0, x_1, ..., x_{j-1}}$, where the union is over all $\{x_0, ..., x_{j-1}\}$ for which $(2N)x_r$ is a nonnegative integer and $\sum_{r=0}^{j-1} x_r \leq 1$. This union is contained in $V_{1+\frac{j}{2N}} \equiv \{(y_0, y_1, ..., y_{j-1}) : y_r \geq 0, \ \sum_{r=0}^{j-1} y_r \leq 1 + \frac{j}{2N}\}$. We have

(4.26)

$$(\hat{x}_0 \hat{x}_1 ... \hat{x}_{j-1} \hat{X}_j)^{-\frac{1}{2}} \leq (y_0 y_1 ... y_{j-1} y_j)^{-\frac{1}{2}}, \text{ for all } (y_0, y_1, ..., y_{j-1}) \in C_{x_0, x_1, ..., x_{j-1}},$$

$$\text{where } y_j = 1 + \frac{j}{2N} - y_0 - y_1 - ... - y_{j-1}.$$

To see that (4.26) holds, note that $\hat{x}_r \geq y_r, \ r = 0, ..., j-1$, for $(y_0, y_1, ..., y_{j-1}) \in C_{x_0, x_1, ..., x_{j-1}}$. Also,

$$\hat{X}_j = \frac{\hat{S}_j}{2N} = \frac{\hat{s}_j + j - 1}{2N} = \frac{s_j + j}{2N} = \frac{2N + j - s_0 - s_1 - ... - s_{j-1}}{2N}$$
$$= 1 + \frac{j}{2N} - x_0 - x_1 - ... - x_{j-1} \geq 1 + \frac{j}{2N} - y_0 - y_1 - ... - y_{j-1},$$

for $(y_0, y_1, ..., y_{j-1}) \in C_{x_0, x_1, ..., x_{j-1}}$.

In light of these facts it follows that the sum on the right hand side of (4.25) is dominated by a certain lower Riemann sum for $\int_{\mathcal{S}_{1+\frac{j}{2N}}} (\prod_{r=0}^{j} y_r)^{-\frac{1}{2}} dy_0 dy_1 ... dy_{j-1}$, where $\mathcal{S}_\lambda = \{(y_0, y_1, ..., y_{j-1}) : y_r \geq 0, \sum_{r=0}^{j-1} y_r \leq \lambda\}$. Replacing the sum in (4.25) with this integral, and substituting the resulting expression into the right hand side of (4.24) gives

$$\frac{A(N,j)}{\binom{2N}{N}^2} \leq (2N+1)^{\frac{1}{2}} c^{j+1} j^{\frac{1}{2}} (2N)^{\frac{j-1}{2}} \int_{\mathcal{S}_{1+\frac{j}{2N}}} (\prod_{r=0}^{j} y_r)^{-\frac{1}{2}} dy_0 dy_1 ... dy_{j-1}. \tag{4.27}$$

A change of variables shows that

(4.28)
$$\int_{\mathcal{S}_{1+\frac{j}{2N}}} (\prod_{r=0}^{j} y_r)^{-\frac{1}{2}} dy_0 dy_1 ... dy_{j-1} = (1+\frac{j}{2N})^{\frac{j-1}{2}} \int_{\mathcal{S}_1} (\prod_{r=0}^{j} y_r)^{-\frac{1}{2}} dy_0 dy_1 ... dy_{j-1}.$$

As is well-known from the theory of Dirichlet distributions,

$$\int_{\mathcal{S}_1} (\prod_{r=0}^{j} y_r)^{-\frac{1}{2}} dy_0 dy_1 ... dy_{j-1} = \frac{\pi^{\frac{j+1}{2}}}{\Gamma(\frac{j+1}{2})}. \tag{4.29}$$

From (4.27)-(4.29) we conclude that

$$\frac{A(N,j)}{\binom{2N}{N}^2} \le (2N+1)^{\frac{1}{2}} c^{j+1} j^{\frac{1}{2}} (2N)^{\frac{j-1}{2}} (1+\frac{j}{2N})^{\frac{j-1}{2}} \frac{\pi^{\frac{j+1}{2}}}{\Gamma(\frac{j+1}{2})}. \tag{4.30}$$

The inequality for $A(N,j)$ in Lemma 2 follows from (4.30). It is trivial to check that $A(N,j)$ is increasing in N; thus, the inequality for $A(k-j,j)$ in Lemma 2 holds as stated. □

Proof of Lemma 3. To prove the lemma we will need the following lemma, which complements Lemma 6.

Lemma 8. *Let $\{\hat{X}_n, \hat{Y}_n\}_{n=0}^{\infty}$ be a simple, symmetric two-dimensional random walk. Let $L > 0$. There exist constants $c_{L,1}, c_{L,2} > 0$ such that for all sufficiently large n*

$$P((\hat{X}_{2n}, \hat{Y}_{2n}) = (0,0) | (\hat{X}_0, \hat{Y}_0) = (2a, 0)) \ge \frac{c_{L,1}}{n} \exp(-\frac{c_{L,2} a^2}{n}),$$
$$\text{for } \ a \in Z \ \text{ satisfying } |a| \le L\sqrt{n}.$$

Proof. Let H_n, V_n be as in the proof of Lemma 6, and let $\{\hat{Z}_n\}$ be as in Lemma 5. We have

$$\begin{aligned} &P((\hat{X}_{2n}, \hat{Y}_{2n}) = (0,0) | (\hat{X}_0, \hat{Y}_0) = (2a,0)) = \\ &\sum_{j+k=n} P(\hat{Z}_{2j} = 0 | \hat{Z}_0 = 2a) P(\hat{Z}_{2k} = 0 | \hat{Z}_0 = 0) \times P(H_n = j, V_n = k). \end{aligned} \tag{4.31}$$

The proof of the lemma follows easily from (4.31), (4.8) and Lemma 5-ii. □

We can now prove Lemma 3. Let $\{(\hat{X}_n, \hat{Y}_n)\}$ denote a simple, symmetric two-dimensional random walk. Using the notation in the proof of Lemma 2, but with $k-1$ in place of N, recall that for $j = 1$,

$$P(D_{s_0,s_1}) = P(\hat{Y}_{s_0} = 0 | \hat{X}_0 = \hat{Y}_0 = \hat{X}_{2k-2} = \hat{Y}_{2k-2} = 0),$$

and thus from (4.2),

$$\frac{A(k-1,1)}{\binom{2k-2}{k-1}^2} = \sum_{l=0}^{2k-2} P(\hat{Y}_l = 0 | \hat{X}_0 = \hat{Y}_0 = \hat{X}_{2k-2} = \hat{Y}_{2k-2} = 0). \tag{4.32}$$

For m satisfying $[\frac{1}{4}k] \leq m \leq [\frac{3}{4}k]$, we have

$$\begin{aligned}
&P(\hat{Y}_{2m} = 0 | \hat{X}_0 = \hat{Y}_0 = \hat{X}_{2k-2} = \hat{Y}_{2k-2} = 0) \geq \\
&\sum_{r=-[\sqrt{k}]}^{[\sqrt{k}]} P(\hat{Y}_{2m} = 0, \hat{X}_{2m} = 2r | \hat{X}_0 = \hat{Y}_0 = 0) \times \\
&\frac{P(\hat{X}_{2k-2} = \hat{Y}_{2k-2} = 0 | \hat{X}_0 = \hat{Y}_0 = \hat{Y}_{2m} = 0, \hat{X}_{2m} = 2r)}{P(\hat{X}_{2k-2} = \hat{Y}_{2k-2} = 0 | \hat{X}_0 = \hat{Y}_0 = 0)}.
\end{aligned} \tag{4.33}$$

We have $P(\hat{X}_{2k-2} = \hat{Y}_{2k-2} = 0 | \hat{X}_0 = \hat{Y}_0 = \hat{Y}_{2m} = 0, \hat{X}_{2m} = 2r) = P(\hat{X}_{2k-2-2m} = \hat{Y}_{2k-2-2m} = 0 | \hat{X}_0 = 2r, \hat{Y}_0 = 0)$. Thus, in light of the above-specified range of m and of r, it follows from Lemma 8 and Lemma 6-i that for sufficiently large k, $\frac{P(\hat{X}_{2k-2} = \hat{Y}_{2k-2} = 0 | \hat{X}_0 = \hat{Y}_0 = \hat{Y}_{2m} = 0, \hat{X}_{2m} = 2r)}{P(\hat{X}_{2k-2} = \hat{Y}_{2k-2} = 0 | \hat{X}_0 = \hat{Y}_0 = 0)}$ is bounded from below by a positive constant. By Lemma 8 and the above-specified bound on m, it also follows that $P(\hat{Y}_{2m} = 0, \hat{X}_{2m} = 2r | \hat{X}_0 = \hat{Y}_0 = 0)$ is bounded from below by $\frac{C}{k}$, for some $C > 0$. Thus, we conclude from (4.33) that for sufficiently large k,

$$\begin{aligned}
&P(\hat{Y}_{2m} = 0 | \hat{X}_0 = \hat{Y}_0 = \hat{X}_{2k-2} = \hat{Y}_{2k-2} = 0) \geq C_1 k^{-\frac{1}{2}}, \\
&\text{for some } C_1 > 0 \text{ and for } \ m \text{ satisfying } [\frac{1}{4}k] \leq m \leq [\frac{3}{4}k].
\end{aligned} \tag{4.34}$$

Now (4.32) and (4.34) give

$$\frac{A(k-1,1)}{\binom{2k-2}{k-1}^2} \geq C_2 k^{\frac{1}{2}},$$

for some $C_2 > 0$ and k sufficiently large. This is clearly equivalent to the lemma.

□

Acknowledgement. The author thanks the referees for their careful reading of the paper. He thanks one of them in particular for pointing out an error in a previous version of the paper.

References

1. Aldous, D. and Diaconis, P., *Longest increasing subsequences: from patience sorting to the Baik-Deift-Johansson theorem*, Bull. Amer. Math. Soc. **36** (1999), 413–432.
2. Baik, J., Deift, P. and Johansson, K., *On the distribution of the length of the longest increasing subsequence of random permutations*, J. Amer. Math. Soc. **12** (1999), 1119–1178.
3. Bollobás, B. and Winkler, P., *The longest chain among random points in Euclidean space.*, Proc. Amer. Math. Soc. **103** (1988), 347–353.
4. Durrett, R., *Probability: Theory and Examples*, Duxbury Press—Wadsworth Pub. Co., 1991.
5. Lifschitz, V. and Pittel, B., *The number of increasing subsequences of the random permutation*, J. Combin. Theory Ser. A **31** (1981), 1–20.
6. Logan, B. F. and Shepp, L. A., *A variational problem for random Young tableaux*, Advances in Math. **26** (1977), 206–222.
7. Sachov, V., *Probabilistic Methods in Combinatorial Analysis, Encyclopedia of Mathematics and its Applications*, vol. 56, Cambridge University Press, 1997.
8. Vershik, A. M. and Kerov, S. V., *Asymptotic behavior of the maximum and generic dimensions of irreducible representations of the symmetric group.*, Functional Anal. Appl. **19** (1985), 21–31.