J. Appl. Prob. **58,** 851–867 (2021)
*doi:*10.1017/jpr.2021.14

ARCSINE LAWS FOR RANDOM WALKS GENERATED FROM RANDOM PERMUTATIONS WITH APPLICATIONS TO GENOMICS

XIAO FANG,* *The Chinese University of Hong Kong*
HAN L. GAN,** *Northwestern University*
SUSAN HOLMES,*** *Stanford University*
HAIYAN HUANG,**** *University of California, Berkeley*
EROL PEKÖZ,***** *Boston University*
ADRIAN RÖLLIN,****** *National University of Singapore*
WENPIN TANG,******* *Columbia University*

Abstract

Aclassical result for the simple symmetric random walk with $2n$ steps is that the number of steps above the origin, the time of the last visit to the origin, and the time of the maximum height all have exactly the same distribution and converge when scaled to the arcsine law. Motivated by applications in genomics, we study the distributions of these statistics for the non-Markovian random walk generated from the ascents and descents of a uniform random permutation and a Mallows(q) permutation and show that they have the same asymptotic distributions as for the simple random walk. We also give an unexpected conjecture, along with numerical evidence and a partial proof in special cases, for the result that the number of steps above the origin by step 2n for the uniform permutation generated walk has exactly the same discrete arcsine distribution as for the simple random walk, even though the other statistics for these walks have very different laws. We also give explicit error bounds to the limit theorems using Stein's method for the arcsine distribution, as well as functional central limit theorems and a strong embedding of the Mallows(q) permutation which is of independent interest.

Keywords: Brownian motion; Lévy statistics; Mallows permutation; Stein's method; strong embedding; uniform permutation

2010 Mathematics Subject Classification: Primary 60C05
Secondary 60J65; 05A05

Received 4 February 2020; revision received 15 February 2021.

* Postal address: Department of Statistics, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong. Email: xfang@sta.cuhk.edu.hk
** Postal address: University of Waikato, Private Bag 3105, Hamilton 3240, New Zealand. Email: han.gan@waikato.ac.nz
*** Postal address: Department of Statistics, 390 Jane Stanford Way, Stanford University, Stanford, CA 94305-4020. Email: susan@stat.stanford.edu
**** Postal address: Department of Statistics, University of California, Berkeley, 367 Evans Hall, Berkeley, CA 94720-3860. Email: hhuang@stat.berkeley.edu
***** Postal address: Boston University, Questrom School of Business, Rafik B. Hariri Building, 595 Commonwealth Avenue, Boston, MA 02215. Email: pekoz@bu.edu
****** Postal address: Department of Statistics and Applied Probability, National University of Singapore, 6 Science Drive 2, Singapore 117546. Email: adrian.roellin@nus.edu.sg
******* Postal address: Department of Industrial Engineering and Operations Research, Columbia University, 500 W. 120th Street #315, New York, NY 10027. Email: wt2319@columbia.edu

1. Introduction

The arcsine distribution appears surprisingly in the study of random walks and Brownian motion. Let $B := (B_t; t \geq 0)$ be one-dimensional Brownian motion starting at zero. Let $G := \sup\{0 \leq s \leq 1 : B_s = 0\}$ be the last exit time of B from zero before time 1, $G^{\max} := \inf\{0 \leq s \leq 1 : B_s = \max_{u\in[0,1]} B_u\}$ be the first time at which B achieves its maximum on [0,1], and $\Gamma := \int_0^1 1_{\{B_s>0\}}\, \mathrm{d}s$ be the occupation time of B above zero before time 1. In [43, 44], Lévy proved the celebrated result that G, $G^{\max}$, and Γ are all arcsine distributed with density

$$f(x) = \frac{1}{\pi\sqrt{x(1-x)}} \qquad \text{for } 0 < x < 1. \tag{1.1}$$

For a random walk $S_n := \sum_{k=1}^n X_k$ with increments $(X_k; k \geq 1)$ starting at $S_0 := 0$, the counterparts of G, $G^{\max}$, and Γ are given by $G_n := \max\{0 \leq k \leq n : S_k = 0\}$, the index at which the walk last hits zero before time n, $G_n^{\max} := \min\{0 \leq k \leq n : S_k = \max_{0\leq k\leq n} S_k\}$, the index at which the walk first attains its maximum value before time n, $\Gamma_n := \sum_{k=1}^n \mathbf{1}[S_k > 0]$, the number of times that the walk is strictly positive up to time n, and $N_n := \sum_{k=1}^n \mathbf{1}[S_{k-1} \geq 0,\ S_k \geq 0]$, the number of edges which lie above zero up to time n. The discrete analog of Lévy's arcsine law was established in [2], where the limiting distribution (1.1) was computed in [19, 26]. Feller [28] gave the following refined treatment:

(i) If the increments $(X_k; k \geq 1)$ of the walk are exchangeable with continuous distribution, then $\Gamma_n \overset{(d)}{=} G_n^{\max}$.

(ii) For a simple random walk with $\mathbb{P}(X_k = \pm 1) = 1/2$ we have $N_{2n} \overset{(d)}{=} G_{2n}$, which follows the discrete arcsine law given by

$$\alpha_{2n,2k} := \frac{1}{2^{2n}}\binom{2k}{k}\binom{2n-2k}{n-k} \qquad \text{for } k \in \{0, \ldots, n\}. \tag{1.2}$$

In the Brownian scaling limit, the above identities imply that $\Gamma \overset{(d)}{=} G^{\max} \overset{(d)}{=} G$. The fact that $G \overset{(d)}{=} G^{\max}$ also follows from Lévy's identity $(|B_t|; t \geq 0) \overset{(d)}{=} (\sup_{s\leq t} B_s - B_t; t \geq 0)$. See [39, 53, 69] and [54, Section 53] for various proofs of Lévy's arcsine law. The arcsine law has been further generalized in several different ways, e.g. in [8, 25, 30] to Lévy processes; [4, 13] to multidimensional Brownian motion; [1, 62] to Brownian motion with drift; and [40, 68] to one-dimensional diffusions. See also [51] for a survey of arcsine laws arising from random discrete structures.

In this paper we are concerned with the limiting distribution of the *Lévy statistics* G_n, $G_n^{\max}$, Γ_n, and N_n of a random walk generated from a class of random permutations. Our motivation comes from a statistical problem in genomics.

1.1. Motivation from genomics

Understanding the relationship between genes is an important goal of systems biology. Systematically measuring the co-expression relationships between genes requires appropriate measures of the statistical association between bivariate data. Since gene expression data routinely require normalization, rank correlations such as Spearman's rank correlation [45, p. 221] have been commonly used; see, for example, [55]. Compared to many other measures, although some information may be lost in the process of converting numerical values to ranks, rank correlations are usually advantageous in terms of being invariant to monotonic transformation, and also robust and less sensitive to outliers. In genomics studies, however, these correlation-based and other kinds of global measures have a practical limitation: they measure a stationary dependent relationship between genes across all samples. It is very likely that the patterns of gene association may change or only exist in a subset of the samples, especially when the samples are pooled from heterogeneous biological conditions. In response to this consideration, several recent efforts have considered statistics that are based on counting local patterns of gene expression ranks to take into account the potentially diverse nature of gene interactions. For instance, denoting the expression profiles for genes X and Y over n conditions (or n samples) by $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{y} = (y_1, \ldots, y_n)$ respectively, the following statistic, denoted by W_2, was introduced in [66] to consider and aggregate possible local interactions:

$$W_2 = \sum_{1 \le i_1 < \cdots < i_k \le n} \big(\mathbf{1}[\phi(x_{i_1}, \ldots, x_{i_k}) = \phi(y_{i_1}, \ldots, y_{i_k})] + \mathbf{1}[\phi(x_{i_1}, \ldots, x_{i_k}) = \phi(-y_{i_1}, \ldots, -y_{i_k})]\big),$$

where $\mathbf{1}[\cdot]$ denotes the indicator function and ϕ is the rank function that returns the indices of elements in a vector after they have been sorted in an increasing order (for example, consider the values (0.5, 1.5, 0.2), which after ordering become (0.2, 0.5, 1.5), described by the permutation $1 \mapsto 2$, $2 \mapsto 3$, $3 \mapsto 1$; applying the *same* permutation to the vector (1,2,3), we thus obtain $\phi(0.5, 1.5, 0.2) = (3, 1, 2)$, which is just the sequence of positions that, after ordering, indicate where the values were *before* they were ordered). The statistic W_2 aggregates the interactions across all subsamples of size $k \le n$; indeed, W_2 is equal to the total number of increasing and decreasing subsequences of length k in a suitably permuted sequence. To see this, suppose σ is a permutation that sorts the elements of $\mathbf{y}$ in a decreasing order. Let $\mathbf{z} = \sigma(\mathbf{x}) = (z_1, \ldots, z_n)$ be that permutation applied to $\mathbf{x}$; then W_2 can be rewritten as

$$W_2 = \sum_{1 \le i_1 < \cdots < i_k \le n} \big(\mathbf{1}[z_{i_1} < \cdots < z_{i_k}] + \mathbf{1}[z_{i_1} > \ldots > z_{i_k}]\big).$$

Several variants of W_2 have been studied to detect different types of dependent patterns between $\mathbf{x}$ and $\mathbf{y}$ (see, for example, [66, 67]).

One variant, for example, is to have $k = 2$ and consider only increasing patterns in $\mathbf{z}$ to assess a negative dependent relationship between $\mathbf{x}$ and $\mathbf{y}$. Denoted by W^*, this variant can be simply expressed as $W^* = \sum_{1 \le i_1 < i_2 \le n} \mathbf{1}[z_{i_1} < z_{i_2}]$. If a more specific negative dependent structure is concerned, say gene Y is an active repressor of gene X when the expression level of gene Y is above a certain value, then we would expect a negative dependent relationship between $\mathbf{x}$ and $\mathbf{y}$, but with that dependence happening only locally among some vector elements. More specifically, this situation suggests that for a condition/sample, the expression of gene X is expected to be low when the expression of gene Y is sufficiently high, or equivalently, this dependence presents between a pair of elements (with each from $\mathbf{x}$ and $\mathbf{y}$ respectively) only when the associated element in $\mathbf{y}$ is above a certain value. To detect this type of dependent relationship, naturally we may consider the family of statistics $W_m^* = \sum_{i=1}^m \mathbf{1}[z_i < z_{i+1}]$, $1 \le m \le n-1$. Note that the elements in $\mathbf{y}$ are ordered in a decreasing order. Thus, in this situation that gene Y is an active repressor of gene X when the expression of gene Y is above a certain level, there should exist a change point m_0 such that W_m^* is significantly high (in comparison to the null case that $\mathbf{x}$ and $\mathbf{y}$ are independent) when $m < m_0$, and the significance would become gradually weakened or disappear as m grows from m_0 to n. For mathematical convenience, considering W_m^* is equivalent to considering $T_m = \sum_{i=1}^m (2\mathbf{1}[z_{i+1} > z_i] - 1)$,

$1 \le m \le n-1$. As argued above, exploring the properties of this process-level statistic would be useful to understand a 'local' negative relationship between **x** and **y** that happens only among a subset of vector elements, as well as for detecting when such relationships would likely occur. To the best of our knowledge, the family of statistics $(T_m; 1 \le m \le n-1)$ has not been theoretically studied in the literature. This statistic provides a motivation for studying the related problem of the permutation generated random walk.

1.2. Permutation generated random walk

Let $\pi := (\pi_1, \ldots, \pi_{n+1})$ be a permutation of $[n+1] := \{1, \ldots, n+1\}$. Let

$$X_k := \begin{cases} +1 & \text{if } \pi_k < \pi_{k+1}, \\ -1 & \text{if } \pi_k > \pi_{k+1}, \end{cases}$$

and denote by $S_n := \sum_{k=1}^n X_k$, $S_0 := 0$, the corresponding walk generated by π. That is, the walk moves to the right at time k if the permutation has a *rise* at position k, and the walk moves to the left at time k if the permutation has a *descent* at position k. An obvious candidate for π is the uniform permutation of $[n+1]$. This random walk model was first studied in [49] in the physics literature, and also appeared in the study of *zigzag diagrams* in [32].

In this article, we consider a more general family of random permutations proposed by Mallows [47], which includes the uniform random permutation. For $0 \le q \le 1$, the one-parameter model

$$\mathbb{P}_q(\pi) = \frac{q^{\operatorname{inv}(\pi)}}{Z_{n,q}} \qquad \text{for } \pi \text{ a permutation of } [n]$$

is referred to as the *Mallows(q) permutation* of $[n]$, where $\operatorname{inv}(\pi) := \#\{(i, j) \in [n] : i < j \text{ and } \pi_i > \pi_j\}$ is the number of inversions of π, and where

$$Z_{n,q} := \sum_{\pi} q^{\operatorname{inv}(\pi)} = \prod_{j=1}^{n} \sum_{i=1}^{j} q^{i-1} = (1-q)^{-n} \prod_{j=1}^{n} (1-q^j)$$

is known as the *q-factorial*. For $q = 1$, the Mallows(1) permutation is the uniform permutation of $[n]$. There have been a number of works on this random permutation model; see, for example, [6, 23, 31, 33, 60, 63].

Question 1.1. *For a random walk generated from the Mallows(q) permutation of* $[n+1]$, *what are the limit laws of the statistics defined at the beginning of Section 1?*

For a Mallows(q) permutation of $[n+1]$, the increments $(X_k; 1 \le k \le n)$ are not independent or even exchangeable. Moreover, the associated walk $(S_k; 0 \le k \le n)$ is not Markov, and as a result the Andersen–Feller machine does not apply. Indeed, when $q = 1$ this random walk has a tendency to change directions more often than a simple symmetric random walk, and thus tends to cross the origin more frequently. Note that the distribution of the walk $(S_k; 0 \le k \le n)$ is completely determined by the up–down sequence or, equivalently, by the descent set $\mathcal{D}(\pi) := \{k \in [n] : \pi_k > \pi_{k+1}\}$ of the permutation π. The number of permutations given the up–down sequence can be expressed either as a determinant, or as a sum of multinomial coefficients; see see [46, Vol. I] and [15, 21, 48, 58, 65]. In particular, the number of permutations with a fixed number of descents is known as the *Eulerian number*. See also [59, Section 7.23], [14, Section 5], and [18] for the descent theory of permutations. None of these results give a simple expression for the limiting distributions of G_n/n, $G_n^{\max}/n$, Γ_n/n, and N_n/n of a random walk generated from the uniform permutation.

2. Main results

To answer Question 1.1, we prove a functional central limit theorem for the walk generated from the Mallows(q) permutation. Although for each $n > 0$ the associated walk $(S_k; 0 \le k \le n)$ is not Markov, the scaling limit is Brownian motion with drift. As a consequence, we derive the limiting distributions of the Lévy statistics, which can be regarded as generalized arcsine laws. In the following, let $(S_t; 0 \le t \le n)$ be the linear interpolation of the walk $(S_k; 0 \le k \le n)$. That is, $S_t = S_{j-1} + (t - j + 1)(S_j - S_{j-1})$ for $j - 1 \le t \le j$. See [12, Chapter 2] for background on the weak convergence in the space $C[0,1]$. The result is stated as follows.

Theorem 2.1 *Fix $0 < q \le 1$, and let $(S_k; 0 \le k \le n)$ be a random walk generated from the Mallows(q) permutation of $[n+1]$. Let*

$$\mu := \frac{1-q}{1+q}, \qquad \sigma := \sqrt{\frac{4q(1-q+q^2)}{(1+q)^2(1+q+q^2)}}. \tag{2.1}$$

Then, as $n \to \infty$,

$$\left(\frac{S_{nt} - \mu n t}{\sigma\sqrt{n}}; 0 \le t \le 1\right) \xrightarrow{(\mathrm{d})} (B_t; 0 \le t \le 1),$$

where $\xrightarrow{(\mathrm{d})}$ denotes the weak convergence in C[0,1] equipped with the sup-norm topology.

Given the above theorem, it is natural to consider the *dragged-down walk* $S_k^q := S_k - \mu k$, $0 \le k \le n$. Let G_n^q, $G_n^{q,\max}$, Γ_n^q, and N_n^q be the Lévy statistics corresponding to the dragged-down walk. As a direct consequence of Theorem 2.1, the random variables G_n^q/n, $G_n^{q,\max}/n$, Γ_n^q/n, and N_n^q/n all converge to the arcsine distribution whose density is given by (1.1).

The proof of Theorem 2.1 is given in Section 3, and makes use of the Gnedin–Olshanski construction of the Mallows(q) permutation. By letting $q = 1$, we get the scaling limit of a random walk generated from the uniform permutation, which has recently been proved in the framework of zigzag graphs [64, Proposition 9.1]. For this case, we have the following corollary.

Corollary 2.2 *Let $(S_k; 0 \le k \le n)$ be a random walk generated from the uniform permutation of $[n+1]$. Then, as $n \to \infty$,*

$$\left(\frac{S_{nt}}{\sqrt{n}}; 0 \le t \le 1\right) \xrightarrow{(\mathrm{d})} \left(\frac{1}{\sqrt{3}} B_t; 0 \le t \le 1\right),$$

where $\xrightarrow{(\mathrm{d})}$ denotes the weak convergence in C[0,1] equipped with the sup-norm topology. Consequently, as $n \to \infty$, the random variables G_n/n, $G_n^{\max}/n$, and Γ_n/n converge in distribution to the arcsine law given by the density (1.1).

Now that the limiting process has been established, we can ask the following question.

Question 2.3. *For a random walk generated from the Mallows(q) permutation of $[n+1]$, what are the error bounds between G_n^q/n, $G^{q,\max}/n$, Γ_n^q/n, N_n^q/n, and their arcsine limit?*

While we cannot answer these questions directly, we were able to prove partial and related results. To state these, we need some notation. For two random variables X and Y, we define the Wasserstein distance as $d_{\mathrm{W}}(X, Y) := \sup_{h\in \mathrm{Lip}(1)} |\mathbb{E}h(X) - \mathbb{E}h(Y)|$, where $\mathrm{Lip}(1) := \{h : |h(x) - h(y)| \le |x - y|\}$ is the class of Lipschitz-continuous functions with Lipschitz constant 1. For $m \ge 1$, let $\mathrm{BC}^{m,1}$ be the class of bounded functions that have m bounded and continuous derivatives and whose mth derivative is Lipschitz continuous. Let $\|h\|_\infty$ be the sup-norm of h, and if the kth derivative of h exists, let

$$|h|_k := \left\| \frac{\mathrm{d}^k h}{\mathrm{d}x^k} \right\|_\infty, \qquad |h|_{k,1} := \sup_{x,y} \left| \frac{\mathrm{d}^k h(x)}{\mathrm{d}x^k} - \frac{\mathrm{d}^k h(y)}{\mathrm{d}y^k} \right| \frac{1}{|x-y|}.$$

The following results hold true for a simple random walk. However, we have strong numerical evidence that they are also true for the permutation generated random walk; see Conjecture 2.5.

Theorem 2.4 *Let $(S_k; 0 \le k \le 2n)$ be a simple symmetric random walk. Then*

$$\mathbb{P}(N_{2n} = 2k) = \alpha_{2k,2n} \qquad \text{for } k \in \{0, \ldots, n\}. \tag{2.2}$$

Moreover, let Z be an arcsine distributed random variable; then

$$d_{\mathrm{W}}\left(\frac{N_{2n}}{2n}, Z\right) \le \frac{27}{2n} + \frac{8}{n^2}. \tag{2.3}$$

Furthermore, for any $h \in \mathrm{BC}^{2,1}$,

$$\left| \mathbb{E}h\left(\frac{N_{2n}}{2n}\right) - \mathbb{E}h(Z) \right| \le \frac{4|h|_2 + |h|_{2,1}}{64n} + \frac{|h|_{2,1}}{64n^2}. \tag{2.4}$$

Identity (2.2) can be found in [28], the bound (2.3) was proved by [34], and the proof of (2.4) is deferred to Section 4.

Conjecture 2.5 *For a uniform random permutation generated random walk of length $2n+1$, the probability that there are $2k$ edges above the origin equals $\alpha_{2n,2k}$, which is the same as that of a simple random walk (see (1.2)).*

For a walk generated from a permutation of $[n+1]$, call it a *positive walk* if $N_n = n$, and a *negative walk* if $N_n = 0$. In [7] it was proved that the number of positive walks b_n generated from permutations of $[n]$ is $n!!\,(n-2)!!$ if n is odd, and $[(n-1)!!]^2$ if n is even. Computer enumerations suggest that $c_{2k,2n+1}$, the number of walks generated from permutations of $[2n+1]$ with $2k$ edges above the origin, satisfies

$$c_{2k,2n+1} = \binom{2n+1}{2k} b_{2k} b_{2n-2k+1}. \tag{2.5}$$

Note that, for the special cases $k = 0$ and $k = n$, the formula (2.5) agrees with the known results in [7]. The formula (2.5) suggests a bijection between permutations of $[2n+1]$ with $2k$ positive edges and pairs of permutations of disjoint subsets of $2n+1$ of respective cardinality $2k$ and $2n+1-2k$ whose associated descent walks are positive. A naive idea is to break the walk into positive and negative excursions, and exclude the final visit to the origin before crossing to the other side of the origin in each excursion [2, 9]. However, this approach does not work since not all pairs of positive walks are obtainable. For example, for $n = 3$, the pair (1,2,3) and (7,6,5,4) cannot be obtained. If Conjecture 2.5 holds, we get the arcsine law as the limiting distribution of $N_{2n}/2n$ with error bounds.

While we are not able to say much about G_n, $G_n^{\max}$, and Γ_n with respect to a random walk generated from the uniform permutation for finite n, we can prove that the limiting distributions of these Lévy statistics are still arcsine; this is a consequence of the fact that the scaled random walks converge to Brownian motion.

Classical results in [41, 42, 57] provide strong embeddings of a random walk with independent increments into Brownian motion. In view of Theorem 2.1, it is also interesting to understand the strong embedding of a random walk generated from the Mallows(q) permutation. We have the following result.

Theorem 2.6 *Fix $0 < q \le 1$, and let $(S_k; 0 \le k \le n)$ be a random walk generated from the Mallows(q) permutation of $[n+1]$. Let μ and σ be defined by (2.1), and let*

$$\beta := \frac{2}{\sigma(1+q)}, \qquad \eta := \frac{2q}{1-q+q^2}.$$

Then, there exist universal constants $n_0, c_1, c_2 > 0$ such that, for any $\varepsilon \in (0, 1)$ and $n \ge n_0$, we can construct $(S_t; 0 \le t \le n)$ and $(B_t; 0 \le t \le n)$ on the same probability space such that

$$\mathbb{P}\left(\sup_{0\le t\le n} \left|\frac{1}{\sigma}(S_t - \mu t) - B_t\right| > c_1 n^{\frac{1+\varepsilon}{4}} (\log n)^{\frac{1}{2}} \beta\right) \le \frac{c_2(\beta^6 + \eta)}{\beta^2 n^{\varepsilon} \log n}.$$

In fact, a much more general result, namely a strong embedding for m-dependent random walks, is proved in Section 5.

Also note that there is substantial literature studying the relations between random permutations and Brownian motion. Classical results were surveyed in [3, 50]; see also [5, 35, 36, 38] for recent progress on the Brownian limit of pattern-avoiding permutations.

3. Proof of Theorem 2.1

In this section we prove Theorem 2.1. To establish the result, we first show that the Mallows(q) permutation can be constructed from *one-dependent* increments $(X_1, \dots, X_n)$; that is, $(X_1, \dots, X_j)$ are independent of $(X_{j+2}, \dots, X_n)$ for each $j \in [n-2]$. Then we calculate its moments and use an invariance principle.

Gnedin and Olshanski [33] provide a nice construction of the Mallows(q) permutation, which is implicit in the original work [47]. This representation of the Mallows(q) permutation plays an important role in the proof of Theorem 2.1.

For $n > 0$ and $0 < q < 1$, let $\mathcal{G}_{q,n}$ be a *truncated geometric* random variable on $[n]$ whose probability distribution is given by

$$\mathbb{P}(\mathcal{G}_{q,n} = k) = \frac{q^{k-1}(1-q)}{1-q^n} \qquad \text{for } k \in [n].$$

Since $\mathbb{P}(\mathcal{G}_{q,n} = k) \to n^{-1}$ if $q \to 1$, we can extend the definition of $\mathcal{G}_{q,n}$ to $q = 1$, which is just the uniform distribution on $[n]$. The Mallows(q) permutation π of $[n]$ is constructed as follows. Let $(Y_k; k \in [n])$ be a sequence of independent random variables, where Y_k is distributed as $\mathcal{G}_{n+1-k}$. Set $\pi_1 := Y_1$ and, for $k \ge 2$, let $\pi_k := \psi(Y_k)$ where ψ is the increasing bijection from

$[n-k+1]$ to $[n]\setminus\{\pi_1, \pi_2, \ldots, \pi_{k-1}\}$. That is, pick π_1 according to $\mathcal{G}_{q,n}$, and remove π_1 from $[n]$. Then pick π_2 as the $\mathcal{G}_{q,n-1}$th smallest element of $[n]\setminus\{\pi_1\}$, and remove π_2 from $[n]\setminus\{\pi_1\}$; and so on. As an immediate consequence of this construction we have that, for the increments $(X_k; k\in[n])$ of a random walk generated from the Mallows(q) permutation of $[n+1]$:

- for each k, $\mathbb{P}(X_k=1)=\mathbb{P}(\mathcal{G}_{q,n+1-k}\le\mathcal{G}_{q,n-k})=1/(1+q)$, which is independent of k and n; thus, $\mathbb{E}X_k=(1-q)/(1+q)$ and $\operatorname{var}X_k=4q/(1+q)^2$;
- the sequence of increments $(X_k; k\in[n])$, though not independent, is *two-block factor* and hence *one-dependent*; see [22] for background.

Such a construction is also used in [33] to construct a random permutation of positive integers, called the *infinite q-shuffle*. The latter is further extended in [52] to *p-shifted permutations* as an instance of regenerative permutations, and used in [37] to construction symmetric k-dependent q-coloring of positive integers.

If π is a uniform permutation of $[n]$, the central limit theorem of the number of descents $\#\mathcal{D}(\pi)$ is well known:

$$\frac{1}{\sqrt{n}}\left(\#\mathcal{D}(\pi)-\frac{n}{2}\right)\xrightarrow{(\mathrm{d})}\frac{1}{\sqrt{12}}\mathcal{N}(0,1),$$

where $\mathcal{N}(0,1)$ is a standard normal distribution. See See [18, Section 3] for a survey of six different approaches to proving this fact. The central limit theorem of the number of descents of the Mallows(q) permutation is known, and is as follows.

Lemma 3.1. (Proposition 5.2 of [14].) *Fix* $0<q\le 1$*, let* π *be the Mallows(q) permutation of [n], and let* $\#\mathcal{D}(\pi)$ *be the number of descents of* π*. Then*

$$\mathbb{E}\#\mathcal{D}(\pi)=\frac{(n-1)q}{1+q},\qquad \operatorname{var}\#\mathcal{D}(\pi)=q\frac{(1-q+q^2)n-1+3q-q^2}{(1+q)^2(1+q+q^2)}.$$

Moreover,

$$\frac{1}{\sqrt{n}}\left(\#\mathcal{D}(\pi)-\frac{nq}{1+q}\right)\xrightarrow{(\mathrm{d})}\mathcal{N}\left(0,\frac{q(1-q+q^2)}{(1+q)^2(1+q+q^2)}\right).$$

We are now ready to prove Theorem 2.1.

Proof of Theorem 2.1. We first recall [11, Theorem 5.1]: Let $X_1, X_2, \ldots$ be an m-dependent sequence, and let $s_n^2=\sum_{i=1}^n\mathbb{E}X_i^2$. If $\mathbb{E}X_n=0$ for all $n\ge 1$, if $\limsup_{n\to\infty}\mathbb{E}X_n^2<\infty$, if $|s_n^2-n\sigma^2|=\mathrm{O}(1)$ for some $\sigma^2>0$, and if

$$\lim_{n\to\infty}s_n^{-2-\delta}\sum_{i=1}^n\mathbb{E}X_i^{2+\delta}=0$$

for some $\delta>0$, then the invariance principle holds for the sequence $X_1, X_2, \ldots$ with normalizing factor $\sigma n^{1/2}$; that is, the sequence of processes $S_n(t)$, $0\le t\le 1$, defined by $S_n(k/n)=\sigma^{-1}n^{-1/2}\sum_{i=1}^k X_i$ for $0\le k\le n$ and linearly interpolated otherwise, converges weakly to a standard Brownian motion on the unit interval with respect to the Borel sigma-algebra generated by the topology of the supremum norm on the space of continuous functions on the unit interval.

Since the increments of a permutation generated random walk are one-dependent, the functional central limit theorem is an immediate consequence of [11, Theorem 5.1] and the moments in Lemma 3.1. □

4. Proof of Theorem 2.4

4.1. Stein's method for the arcsine distribution

It is well known that for a simple symmetric walk, G_{2n} and N_{2n} are discrete arcsine distributed, thus converging to the arcsine distribution. To apply Stein's method for arcsine approximation we first need a characterizing operator.

Lemma 4.1 *A random variable Z is arcsine distributed if and only if*

$$\mathbb{E}[Z(1-Z)f'(Z) + (1/2 - Z)f(Z)] = 0$$

for all functions $f \in \mathrm{BC}^{2,1}[0, 1]$.

To apply Stein's method, we proceed as follows. Let Z be an arcsine distributed random variable. Then, for any $h \in \mathrm{Lip}(1)$ or $h \in \mathrm{BC}^{2,1}[0, 1]$, assume we have a function f that solves

$$x(1-x)f'(x) + (1/2 - x)f(x) = h(x) - \mathbb{E}h(Z). \tag{4.1}$$

For an arbitrary random variable W, replace x with W in (4.1) and, by taking expectations, this yields an expression for $\mathbb{E}h(W) - \mathbb{E}h(Z)$ in terms of just W and f. Our goal is therefore to bound the expectation of the left-hand side of (4.1) by utilizing properties of f. Extending [24, 34] developed Stein's method for the beta distribution (of which arcsine is a special case) and gave an explicit Wasserstein bound between the discrete and the continuous arcsine distributions. We will use the framework from [29] to calculate error bounds for the class of test functions $\mathrm{BC}^{2,1}$.

4.2. Proof of Theorem 2.4

To simplify the notation, let $W_n := N_{2n}/2n$ be the fraction of positive edges of a simple symmetric random walk. Let $\Delta_y f(x) := f(x+y) - f(x)$. We will use the following known facts for the discrete arcsine distribution. For any function $f \in \mathrm{BC}^{m,1}[0, 1]$,

$$\mathbb{E}\left[nW_n\left(1 - W_n + \frac{1}{2n}\right)\Delta_{1/n}f\left(W_n - \frac{1}{n}\right) + \left(\frac{1}{2} - W_n\right)f(W_n)\right] = 0. \tag{4.2}$$

Moreover,

$$\mathbb{E}W_n = \frac{1}{2}, \qquad \mathbb{E}W_n^2 = \frac{3}{8} + \frac{1}{8n}. \tag{4.3}$$

The identity (4.2) can be read from [24, Lemma 2.9] and [34, Proof of Theorem 1.1]. The moments are easily derived by plugging in $f(x) = 1$ and $f(x) = x$.

Proof of Theorem 2.4. The distribution (2.2) of N_{2n} can be found in [28]. The bound (2.3) follows from the fact that N_{2n} is discrete arcsine distributed, together with [34, Theorem 1.2].

We prove the bound (2.4) using the generator method. Assume $h \in \mathrm{BC}^{2,1}([0, 1])$, and recall the Stein equation (4.1) for the arcsine distribution. It follows from [29, Theorem 5] that there exists $g \in \mathrm{BC}^{2,1}([0, 1])$ such that $x(1-x)g''(x) + (1/2 - x)g'(x) = h(x) - \mathbb{E}h(Z)$, which is just (4.1) with $f = g'$. We are therefore required to bound the absolute value of

$$\mathbb{E}h(W_n) - \mathbb{E}h(Z) = \mathbb{E}\left[W_n(1-W_n)g''(W_n) - \left(\frac{1}{2} - W_n\right)g'(W_n)\right].$$

Applying (4.2) with f being replaced by g', we obtain

$$\begin{aligned}
&\mathbb{E}h(W_n) - \mathbb{E}h(Z)\\
&= \mathbb{E}\left[W_n(1-W_n)g''(W_n) - nW_n\left(1 - W_n + \frac{1}{2n}\right)\Delta_{1/n}g'\left(W_n - \frac{1}{n}\right)\right]\\
&= \mathbb{E}\left[W_n(1-W_n)\left(g''(W_n) - n\Delta_{1/n}g'\left(W_n - \frac{1}{n}\right)\right) - \frac{W_n}{2}\Delta_{1/n}g'\left(W_n - \frac{1}{n}\right)\right].
\end{aligned}$$

The second term in the expectation is bounded as

$$\left|\mathbb{E}\left[\frac{W_n}{2}\Delta_{1/n}g'\left(W_n - \frac{1}{n}\right)\right]\right| \leq \frac{\mathbb{E}W_n}{2}\cdot\frac{|g|_2}{n} = \frac{|g|_2}{4n}, \tag{4.4}$$

and the first term can be bounded as

$$\begin{aligned}
&\left|\mathbb{E}\left[nW_n(1-W_n)\int_{W_n-\frac{1}{n}}^{W_n} g''(W_n) - g''(x)\,\mathrm{d}x\right]\right|\\
&\quad\leq \left|\mathbb{E}\left[nW_n(1-W_n)|g|_{2,1}\int_{W_n-\frac{1}{n}}^{W_n}|W_n - x|dx\right]\right|\\
&\quad= |g|_{2,1}n\mathbb{E}\left[W_n(1-W_n)\int_0^{\frac{1}{n}} s\,\mathrm{d}s\right] = \frac{|g|_{2,1}}{16}\left(\frac{1}{n} + \frac{1}{n^2}\right),
\end{aligned} \tag{4.5}$$

where the last equality follows from (4.3). Combining (4.4), (4.5), and [29, Theorem 5] (for relating the bounds on derivatives g with derivatives of h) yields the desired bound. □

Remark 4.2 The above bound is essentially sharp. Take $h(x) = \frac{x^2}{2}$, $\mathbb{E}h(W_n) - \mathbb{E}h(Z) = -\frac{1}{16n}$, and the above bound gives $|\mathbb{E}h(W_n) - \mathbb{E}h(Z)| \leq \frac{1}{16n} + \frac{1}{64n^2}$.

5. Proof of Theorem 2.6

In this section we prove Theorem 2.6. To this end, we prove a general result for strong embeddings of a random walk with finitely dependent increments.

5.1. Strong embeddings of m-dependent walks

Let n, m be positive integers. Let $(X_i; i \in [n])$ be a sequence of m-dependent random variables. That is, $\{X_1, \ldots, X_j\}$ is independent of $\{X_{j+m+1}, \ldots, X_n\}$ for each $j \in [n-m-1]$. Let $(S_k; k \in \{0, 1, \ldots, n\})$ be a random walk with increments X_i, and $(S_t; 0 \leq t \leq n)$ be the linear interpolation of $(S_k; k \in \{0, 1, \ldots, n\})$. Assume that the random variables X_i are centered and scaled such that $\mathbb{E}X_i = 0$ for all $i \in [n]$ and $\mathrm{var}(S_n) = n$. Let $(B_t; t \geq 0)$ be a one-dimensional standard Brownian motion. The idea of strong embedding is to couple $(S_t; 0 \leq t \leq n)$ and $(B_t; 0 \leq t \leq n)$ in such a way that

$$\mathbb{P}\left(\sup_{0\leq t\leq n}|S_t - B_t| > b_n\right) = p_n \tag{5.1}$$

for some $b_n = o(n^{\frac{1}{2}})$ and $p_n = o(1)$ as $n \to \infty$ (note that the typical fluctuation of B_n is $O(n^{\frac{1}{2}})$).

The study of such embeddings dates back to [57]. When the Xs are independent and identically distributed, [61] obtained (5.1) with $b_n = \mathcal{O}(n^{\frac{1}{4}}(\log n)^{\frac{1}{2}}(\log\log n)^{\frac{1}{4}})$; [20] used a novel approach to prove that under the additional conditions $\mathbb{E}X_i^3 = 0$ and $\mathbb{E}X_i^8 < \infty$ we get $b_n = \mathcal{O}(n^{\frac{1}{6}+\varepsilon})$ for any $\varepsilon > 0$; and [41, 42] further obtained $b_n = \mathcal{O}(\log n)$ under a finite moment-generating function assumption. See also [10, 17] for recent developments.

We use the argument from [20] to obtain the following result for m-dependent random variables.

Theorem 5.1 *Let $(S_t; 0 \le t \le n)$ be the linear interpolation of partial sums of m-dependent random variables $(X_i; i \in [n])$. Assume that $1 \le m \le n^{\frac{1}{5}}$ and $\mathbb{E}X_i = 0$ for each $i \in [n]$. Further assume that $|X_i| \le \beta$ for each $i \in [n]$, where $\beta > 0$ is a constant. Let*

$$\eta := \max_{\substack{k\in[n],\\ j\in\{0,\ldots,n-k\}}} |\mathrm{var}(S_{j+k} - S_j) - k|. \tag{5.2}$$

For any $\varepsilon \in (0, 1)$, if $\eta \le n^{\varepsilon}$ then there exist positive constants n_0, c_1, and c_2 depending only on ε such that, for any $n \ge n_0$, we can define $(S_t; 0 \le t \le n)$ and $(B_t; 0 \le t \le n)$ on the same probability space such that

$$\mathbb{P}\left(\sup_{0\le t\le n} |S_t - B_t| > c_1 n^{\frac{1+\varepsilon}{4}}(\log n)^{\frac{1}{2}} m^{\frac{1}{2}} \beta\right) \le \frac{c_2(m^4\beta^6 + \eta)}{m\beta^2 n^{\varepsilon}\log n}.$$

If m and β are absolute constants and $\mathrm{var}(S_{j+k} - S_j)$ matches k up to an absolute constant, from Theorem 5.1 we get (5.1) with $b_n = \mathcal{O}(n^{\frac{1+\varepsilon}{4}}(\log n)^{\frac{1}{2}})$ and $p_n = \mathcal{O}(1/(n^{\varepsilon}\log n))$ for any fixed $\varepsilon \in (0, 1)$.

Proof of Theorem 2.6. We apply Theorem 5.1 with $m = 1$, and a suitable choice of β and η. By centering and scaling, we consider the walk $(S_t^{'}; 0 \le t \le n)$ with increments $X_i^{'} = \frac{1}{\sigma}(X_i - \mu)$. It is easy to see that $|X_i^{'}| \le \frac{1}{\sigma}\max(1 - \mu, 1 + \mu) = \beta$. According to the result in Section 3,

$$\mathbb{P}(X_k = X_{k+1} = 1) = \mathbb{P}(\mathcal{G}_{q,n+1-k} \le \mathcal{G}_{q,n-k} \le \mathcal{G}_{q,n-k-1}) = \frac{1}{(1+q)(1+q+q^2)},$$
$$\mathbb{P}(X_k = -1, X_{k+1} = 1) = \mathbb{P}(X_{k+1} = 1) - \mathbb{P}(X_k = X_{k+1} = 1) = \frac{q}{1+q+q^2},$$
$$\mathbb{P}(X_k = X_{k+1} = -1) = \mathbb{P}(X_k = -1) - \mathbb{P}(X_k = -1, X_{k+1} = 1) = \frac{q^3}{(1+q)(1+q+q^2)}.$$

By the one-dependence property, elementary computation shows that, for $k \le n$, $\mathrm{var} S_k^{'} = k + \eta$, which leads to the desired result. □

5.2. Proof of Theorem 5.1

The proof of Theorem 5.1 boils down to a series of lemmas. In the following, 'sufficiently large n' means $n \ge n_0$ for some n_0 depending only on ε. We use C and c to denote positive constants depending only on ε and may differ in different expressions. Let $d := \lceil n^{\frac{1-\varepsilon}{2}} \rceil$, where $\lceil x \rceil$ is the least integer greater than or equal to x. We divide the interval $[0,n]$ into d subintervals by points $\lceil jn/d\rceil$, $j\in[d]$, each with length $l=\lceil n/d\rceil$ or $l=\lceil n/d\rceil-1$. The following results hold for both values of l.

Lemma 5.2 *Under the assumptions in Theorem 5.1, we have, for sufficiently large n,*

$$4m\beta^2\geq 1, \qquad l\geq 6m\log n. \tag{5.3}$$

Proof. By the definition of η in (5.2), the m-dependence assumption, and the upper bounds on η and $|X_i|$, we have

$$n-n^{\varepsilon}\leq n-\eta\leq \mathrm{var}S_n=\sum_{i=1}^n\sum_{j:|j-i|\leq m}\mathbb{E}X_iX_j\leq n(2m+1)\beta^2, \qquad m\geq 1,$$

which implies $4m\beta^2\geq 1$ for sufficiently large n. The second bound in (5.3) follows from the fact that $m\leq n^{\frac{1}{5}}$ and $l\sim n^{\frac{1+\varepsilon}{2}}$. □

Given two probability measures μ and ν on $\mathbb{R}$, define their Wasserstein-2 distance by

$$d_{W_2}(\mu,\nu)=\left(\inf_{\pi\in\Gamma(\mu,\nu)}\int|x-y|^2\,\mathrm{d}\pi(x,y)\right)^{\frac{1}{2}},$$

where $\Gamma(\mu,\nu)$ is the space of all probability measures on $\mathbb{R}^2$ with μ and ν as marginals. We will use the following Wasserstein-2 bound from [27]. We use $\mathcal{N}(\mu,\sigma^2)$ to denote the normal distribution with mean μ and variance σ^2.

Lemma 5.3. (Corollary 2.3 of [27].) *Let* $W=\sum_{i=1}^n\xi_i$ *be a sum of m-dependent random variables with* $\mathbb{E}\xi_i=0$ *and* $\mathbb{E}W^2=1$*. We have*

$$d_{W_2}(\mathcal{L}(W),\mathcal{N}(0,1))\leq C_0\left\{m^2\sum_{i=1}^n\mathbb{E}|\xi_i|^3+m^{3/2}\left(\sum_{i=1}^n\mathbb{E}\xi_i^4\right)^{1/2}\right\}, \tag{5.4}$$

where C_0 *is an absolute constant.*

Specializing the above lemma to bounded random variables, we obtain the following result.

Lemma 5.4 *Under the assumptions in Theorem 5.1, we have, for sufficiently large n,*

$$d_{W_2}(\mathcal{L}(S_{l-m}),\ \mathcal{N}(0,\sigma^2))\leq Cm^2\beta^3, \tag{5.5}$$

where $\sigma^2:=\mathrm{var}S_{l-m}$.

Proof. Applying (5.4) to $\sigma^{-1}S_{l-m}$ and using $|X_i|\leq\beta$, we obtain

$$\begin{aligned}d_{W_2}(\mathcal{L}(S_{l-m}),\ \mathcal{N}(0,\sigma^2))&=\sigma\, d_{W_2}(\sigma^{-1}S_{l-m},\mathcal{N}(0,1))\\&\leq\sigma C_0\left(lm^2\left(\frac{\beta}{\sigma}\right)^3+\left(lm^3\left(\frac{\beta}{\sigma}\right)^4\right)^{\frac{1}{2}}\right)\leq Cm^2\beta^3,\end{aligned}$$

where we used $4m\beta^2\geq 1$ from (5.3), and $\sigma^2\geq l-m-\eta\geq cl$ for sufficiently large n from (5.2), $m\leq n^{\frac{1}{5}}$, $\eta\leq n^{\varepsilon}$, $l\sim n^{\frac{1+\varepsilon}{2}}$, and $\varepsilon\in(0,1)$. □

Lemma 5.5 *For sufficiently large n, there exists a coupling of* $(S_t;0\leq t\leq n)$ *and* $(B_t;0\leq t\leq n)$ *such that, with* $e_j:=(S_{\lceil jn/d\rceil}-S_{\lceil(j-1)n/d\rceil})-(B_{\lceil jn/d\rceil}-B_{\lceil(j-1)n/d\rceil})$*, the sequence* $(e_1,\ldots,e_d)$ *is one-dependent, and* $\mathbb{E}e_j^2\leq C(m^4\beta^6+\eta)$ *for all* $j\in[n]$.

Proof. We use $4m\beta^2 \geq 1$ implicitly below to absorb a few terms into $Cm^4\beta^6$. With σ^2 defined in Lemma 5.4, we have $d_{W_2}(\mathcal{N}(0, \sigma^2), \mathcal{N}(0, l)) \leq \sqrt{|l-\sigma^2|} \leq \sqrt{m+\eta}$. Combining (5.5), the above bound, and the m-dependence assumption, we can couple $S_{\lceil jn/d\rceil - m} - S_{\lceil (j-1)n/d\rceil}$ and $B_{\lceil jn/d\rceil} - B_{\lceil (j-1)n/d\rceil}$ for each $j \in [d]$ independently with $\mathbb{E}[(S_{\lceil jn/d\rceil - m} - S_{\lceil (j-1)n/d\rceil}) - (B_{\lceil jn/d\rceil} - B_{\lceil (j-1)n/d\rceil})]^2 \leq C(m^4\beta^6 + \eta)$. By the m-dependence assumption, we can generate $X_1, \ldots, X_n$ from their conditional distribution given $(S_{\lceil jn/d\rceil - m} - S_{\lceil (j-1)n/d\rceil}; j \in [d])$, thus obtaining $(S_t; 0 \leq t \leq n)$, and generate $(B_t; 0 \leq t \leq n)$ given $(B_{\lceil jn/d\rceil}; j \in [d])$. Since $\mathbb{E}(S_{\lceil jn/d\rceil} - S_{\lceil jn/d\rceil - m})^2 \leq Cm^2\beta^2$, we have $\mathbb{E}(e_j^2) \leq C(m^4\beta^6 + \eta)$. Finally, the one-dependence of $(e_1, \ldots, e_d)$ follows from the m-dependence assumption. □

Lemma 5.6 *Let* $T_j = \sum_{i=1}^{j} e_i$, $j \in [d]$. *For any* $b > 0$ *and sufficiently large* n, $\mathbb{P}(\max_{j\in[d]} |T_j| > b) \leq C(m^4\beta^6 + \eta)d/b^2$.

Proof. Define $T_j^{(1)} = \sum_{\substack{i=1,3,5,\ldots\\ i\leq j}} e_i$ and $T_j^{(2)} = \sum_{\substack{i=2,4,6,\ldots\\ i\leq j}} e_i$. By Lemma 5.5, $T_j^{(1)}$ is a sum of independent random variables with zero mean and finite second moments. By Kolmogorov's maximal inequality,

$$\mathbb{P}\left(\max_{1\leq j\leq d} \left|T_j^{(1)}\right| > \frac{b}{2}\right) \leq \frac{C(m^4\beta^6 + \eta)d}{b^2}.$$

The same bound holds for $T_j^{(2)}$. The lemma is proved by the union bound

$$\mathbb{P}\left(\max_{j\in[d]} |T_j| > b\right) \leq \mathbb{P}\left(\max_{1\leq j\leq d} |T_j^{(1)}| > \frac{b}{2}\right) + \mathbb{P}\left(\max_{1\leq j\leq d} |T_j^{(2)}| > \frac{b}{2}\right). \qquad \square$$

Lemma 5.7 *For any* $0 < b \leq 4l\beta$, *we have*

$$\mathbb{P}\left(\max_{j\in[l]} |S_j - jS_l/l| > b\right) \leq 2l \exp\left(-\frac{b^2}{48lm\beta^2}\right).$$

Proof. We first prove a concentration inequality for S_j, $j \in [l]$, then use the union bound. Let $h(\theta) = \mathbb{E}\mathrm{e}^{\theta S_j}$, with $h(0) = 1$. Let $S_j^{(i)} = S_j - \sum_{k\in[j]:|k-i|\leq m} X_k$. Using $\mathbb{E}X_i = 0$, $|X_i| \leq \beta$, the m-dependence assumption, and the inequality (cf. [16, Eq. (7)])

$$\left|\frac{\mathrm{e}^x - \mathrm{e}^y}{x-y}\right| \leq \frac{1}{2}(\mathrm{e}^x + \mathrm{e}^y),$$

we have, for $\theta > 0$ and $\theta(2m+1)\beta \leq 1$,

$$\begin{aligned} h'(\theta) = \mathbb{E}(S_j\mathrm{e}^{\theta S_j}) &= \sum_{i=1}^{j} \mathbb{E}X_i(\mathrm{e}^{\theta S_j} - \mathrm{e}^{\theta S_j^{(i)}}) \leq \frac{\theta}{2}\sum_{i=1}^{j} \mathbb{E}|X_i|\left|S_j - S_j^{(i)}\right|\left(\mathrm{e}^{\theta S_j} + \mathrm{e}^{\theta S_j^{(i)}}\right) \\ &\leq \left(m + \frac{1}{2}\right)\theta l\beta^2\mathbb{E}\mathrm{e}^{\theta S_j}(1 + \mathrm{e}^{\theta(2m+1)\beta}) \leq 6\theta lm\beta^2 h(\theta). \end{aligned}$$

This implies that $\log h(\theta) \leq 3lm\beta^2\theta^2$, and

$$\mathbb{P}(S_j > b/2) \leq \mathrm{e}^{-\theta b/2}\mathbb{E}\mathrm{e}^{\theta S_j} \leq \exp\left(-\frac{b^2}{48lm\beta^2}\right)$$

by choosing $\theta = b/(12lm\beta^2)$, provided that $b \le 4l\beta$. The same bound holds for $-S_j$. Consequently,

$$\mathbb{P}\left(\max_{j\in[l]} |S_j - jS_l/l| > b\right) \le \mathbb{P}\left(\max_{j\in[l-1]} |S_j| > b/2\right) + \mathbb{P}\left(|S_l| > b/2\right) \le 2l \exp\left(-\frac{b^2}{48lm\beta^2}\right). \qquad \square$$

Lemma 5.8 *For any $b > 0$, we have $\mathbb{P}(\sup_{0\le t\le l} |B_t - tB_l/l| > b) \le 2\mathrm{e}^{-\frac{2b^2}{l}}$.*

Proof. We have

$$\mathbb{P}\left(\sup_{0\le t\le l} |B_t - tB_l/l| > b\right) \le \mathbb{P}\left(\sup_{0\le t\le l} (B_t - tB_l/l|) > b\right) + \mathbb{P}\left(\inf_{0\le t\le l} (B_t - tB_l/l) < -b\right) = 2\mathbb{P}\left(\sup_{0\le t\le l} (B_t - tB_l/l) > b\right) = 2\mathrm{e}^{-\frac{2b^2}{l}},$$

where the last equality is the well-known distribution of the maximum of the Brownian bridge (cf. [56, p. 34]). $\square$

Now we proceed to proving Theorem 5.1.

Proof of Theorem 5.1. Let $b_l = (96lm\beta^2 \log n)^{1/2}$ and $b := b_{\lceil \frac{n}{d} \rceil}$. Note that, since $m \le l/(6\log n)$ from (5.3), b_l satisfies the condition $b_l \le 4l\beta$ in Lemma 5.7 for sufficiently large n. Note also that if $\sup_{0\le t\le n} |S_t - B_t| > 3b$ then either $\max_{j\in[d]} |T_j| > b$ or the fluctuation of either S_t or B_t within some subinterval of length l is larger than b_l. By the union bound and Lemmas Lemma 5.6–Lemma 5.8, we have

$$\begin{aligned}
&\mathbb{P}\left(\sup_{0\le t\le n} |S_t - B_t| > 3b\right)\\
&\quad\le \mathbb{P}\left(\max_{j\in[d]} |T_j| > b\right) + d\mathbb{P}\left(\max_{j\in[l]} |S_j - jS_l/l| > b_l\right) + d\mathbb{P}\left(\sup_{0\le t\le l} |B_t - tB_l/l| > b_l\right)\\
&\quad\le \frac{C(m^4\beta^6 + \eta)d}{b^2} + 2dl\exp\left(-\frac{b_l^2}{48lm\beta^2}\right) + 2d\exp\left(-\frac{2b_l^2}{l}\right)\\
&\quad\le \frac{C(m^4\beta^6 + \eta)}{n^\varepsilon m\beta^2 \log n} + \frac{C}{n} \le \frac{C(m^4\beta^6 + \eta)}{m\beta^2 n^\varepsilon \log n},
\end{aligned}$$

where we used $4m\beta^2 \ge 1$ for sufficiently large n. This proves the theorem. $\square$

Acknowledgements

The initial portion of this work was conducted at the meeting 'Stein's method and applications in high-dimensional statistics' held at the American Institute of Mathematics in August 2018. We are indebted to Bhaswar Bhattacharya, Sourav Chatterjee, Persi Diaconis, and Jon Wellner for helpful discussions throughout the project. We would also like to express our gratitude to John Fry and staff Estelle Basor, Brian Conrey, and Harpreet Kaur at the American

Institute of Mathematics for their generosity and excellent hospitality in hosting this meeting at the Fry's Electronics corporate headquarters in San Jose, CA, and Jay Bartroff, Larry Goldstein, Stanislav Minsker, and Gesine Reinert for organizing such a stimulating meeting.

WT thanks Yuting Ye for communicating the problem of the limiting distribution of the sojourn time of a random walk generated from the uniform permutation, which brought him to this work, and Jim Pitman for helpful discussions. SH acknowledges support from the NSF DMS grant 1501767. XF acknowledges support from Hong Kong RGC ECS 24301617, GRF 14302418, 14304917. We also thank the anonymous referees for their careful reading and for pointing out some errors in the first draft of the manuscript.

Finally, we thank the Institute for Mathematical Sciences at the National University of Singapore, where part of this work was undertaken, for their kind support.

References

[1] Akahori, J. (1995). Some formulae for a new type of path-dependent option. *Ann. Appl. Prob.* **5**, 383–388.
[2] Andersen, E. S. (1953). On sums of symmetrically dependent random variables. *Skand. Aktuarietidskr.* **36**, 123–138.
[3] Arratia, R., Barbour, A. and Tavaré, S. (2003). *Logarithmic Combinatorial Structures: A Probabilistic Approach* (EMS Monogr. Math. 1). EMS Publishing House, Zurich.
[4] Barlow, M., Pitman, J. and Yor, M. (1989). Une extension multidimensionnelle de la loi de l'arc sinus. In *Séminaire de Probabilités* (Lect. Notes Math. 23). Springer, Berlin, pp. 294–314.
[5] Bassino, F., Bouvel, M., Féray, V., Gerin, L. and Pierrot, A. (2018). The Brownian limit of separable permutations. *Ann. Prob.* **46**, 2134–2189.
[6] Basu, R. and Bhatnagar, N. (2017). Limit theorems for longest monotone subsequences in random Mallows permutations. *Ann. Inst. H. Poincaré Prob. Statist.* **53**, 1934–1951.
[7] Bernardi, O., Duplantier, B. and Nadeau, P. (2010). A bijection between well-labelled positive paths and matchings. *Séminaire Lotharingien de Combinatoire* **63**, B63e.
[8] Bertoin, J. and Doney, R. (1997). Spitzer's condition for random walks and Lévy processes. *Ann. Inst. H. Poincaré Prob. Statist.* **33**, 167–178.
[9] Bertoin, J. (1993). Splitting at the infimum and excursions in half-lines for random walks and Lévy processes. *Stoch. Process. Appl.* **47**, 17–35.
[10] Bhattacharjee, C. and Goldstein, L. (2016). On strong embeddings by Stein's method. *Electron. J. Prob.* **21**, 1–30.
[11] Billingsley, P. (1956). The invariance principle for dependent random variables. *Trans. Amer. Math. Soc.* **83**, 250–268.
[12] Billingsley, P. (1999). *Convergence of Probability Measures*, 2nd ed. John Wiley, New York.
[13] Bingham, N. and Doney, R. (1988). On higher-dimensional analogues of the arc-sine law. *J. Appl. Prob.* **25**, 120–131.
[14] Borodin, A., Diaconis, P. and Fulman, J. (2010). On adding a list of numbers (and other one-dependent determinantal processes). *Bull. Amer. Math. Soc.* **47**, 639–670.
[15] Carlitz, L. (1973). Permutations with prescribed pattern. *Math. Nachr.* **58**, 31–53.
[16] Chatterjee, S. (2007). Stein's method for concentration inequalities. *Prob. Theory Relat. Fields* **138**, 305–321.
[17] Chatterjee, S. (2012). A new approach to strong embeddings. *Prob. Theory Relat. Fields* **152**, 231–264.
[18] Chatterjee, S. and Diaconis, P. (2017). A central limit theorem for a new statistic on permutations. *Indian J. Pure Appl. Math.* **48**, 561–573.
[19] Chung, K. L. and Feller, W. (1949). On fluctuations in coin-tossing. *Proc. Nat. Acad. Sci.* 35, 605–608.
[20] Csörgö, M. and Révész, P. (1975). A new method to prove Strassen type laws of invariance principle. I. *Z. Wahrscheinlichkeitsth.* 31, 255–259.
[21] de Bruijn, N. G. (1970). Permutations with given ups and downs. *Nieuw Arch. Wiskd.* **18**, 61–65.
[22] de Valk, V. (1994). *One-Dependent Processes: Two-Block-Factors and Non-Two-Block-Factors* (CWI tracts 85). Centrum voor Wiskunde en Informatica, Amsterdam.
[23] Diaconis, P. (1988). *Group Representations in Probability and Statistics* (Lect. Notes Monogr. 11). Institute of Mathematics and Statistics, Hayward, CA.
[24] Döbler, C. (2012). A rate of convergence for the arcsine law by Stein's method. Preprint, arXiv:1207.2401.
[25] Dynkin, E. B. (1965). *Markov Processes* Vols. I, II (Grundlehren der Mathematischen Wissenschaften 121, 122). Springer, Berlin.

[26] ERDÖS, P. AND KAC, M. (1947). On the number of positive sums of independent random variables. *Bull. Amer. Math. Soc.* **53**, 1011–1020.
[27] FANG, X. (2019). Wasserstein-2 bounds in normal approximation under local dependence. *Electron. J. Prob.* **24**, 1–14.
[28] FELLER, W. (1968). *An Introduction to Probability Theory and Its Applications*, 2nd ed. Vol. **I**. John Wiley, New York.
[29] GAN, H. L., RÖLLIN, A. AND ROSS, N. (2017). Dirichlet approximation of equilibrium distributions in Cannings models with mutation. *Adv. Appl. Prob.* **49**, 927–959.
[30] GETOOR, R. AND SHARPE, M. (1994). On the arc-sine laws for Lévy processes. *J. Appl. Prob.* **31**, 76–89.
[31] GLADKICH, A. AND PELED, R. (2018). On the cycle structure of Mallows permutations. *Ann. Prob.* **46**, 1114–1169.
[32] GNEDIN, A. AND OLSHANSKI, G. (2006). Coherent permutations with descent statistic and the boundary problem for the graph of zigzag diagrams. *Int. Math. Res. Not.* **2006**, 51968.
[33] GNEDIN, A. AND OLSHANSKI, G. (2010). q-exchangeability via quasi-invariance. *Ann. Prob.* **38**, 2103–2135.
[34] GOLDSTEIN, L. AND REINERT, G. (2013). Stein's method for the beta distribution and the Polya–Eggenberger urn. *J. Appl. Prob.* **50**, 1187–1205.
[35] HOFFMAN, C., RIZZOLO, D. AND SLIVKEN, E. (2017a). Pattern-avoiding permutations and Brownian excursion part I: Shapes and fluctuations. *Random Structures Algorithms* **50**, 394–419.
[36] HOFFMAN, C., RIZZOLO, D. AND SLIVKEN, E. (2017b). Pattern-avoiding permutations and Brownian excursion, part II: Fixed points. *Prob. Theory Relat. Fields* 169, 377–424.
[37] HOLROYD, A., HUTCHCROFT, T. AND LEVY, A. (2020). Mallows permutations and finite dependence. *Ann. Prob.* **48**, 343–379.
[38] JANSON, S. (2017). Patterns in random permutations avoiding the pattern 132. *Combinatorics Prob. Comput.* **26**, 24–51.
[39] KARATZAS, I. AND SHREVE, S. E. (1987). A decomposition of the Brownian path. *Statist. Prob. Lett.* **5**, 87–93.
[40] KASAHARA, Y. AND YANO, Y. (2005). On a generalized arc-sine law for one-dimensional diffusion processes. *Osaka J. Math.* **42**, 1–10.
[41] KOMLÓS, J., MAJOR, P. AND TUSNÁDY, G. (1975). An approximation of partial sums of independent RVs, and the sample DF. I. *Z. Wahrscheinlichkeitsth.* 32, 111–131.
[42] KOMLÓS, J., MAJOR, P. AND TUSNÁDY, G. (1976). An approximation of partial sums of independent RVs, and the sample DF. II. *Z. Wahrscheinlichkeitsth.* 34, 33–58.
[43] LÉVY, P. (1939). Sur certains processus stochastiques homogènes. *Compositio Math.* **7**, 283–339.
[44] LÉVY, P. (1965). *Processus stochastiques et mouvement brownien*. Suivi d'une note de M. Loève. Deuxième édition revue et augmentée. Gauthier-Villars & Cie, Paris.
[45] MCDONALD, J. H. (2009). *Handbook of Biological Statistics*. Sparky House Publishing, Baltimore, MD.
[46] MACMAHON, P. A. (1960). *Combinatory Analysis*. Chelsea Publishing Co., New York.
[47] MALLOWS, C. L. (1957). Non-null ranking models. I. *Biometrika* 44, 114–130.
[48] NIVEN, I. (1968). A combinatorial problem of finite sequences. *Nieuw Arch. Wisk* **16**, 116–123.
[49] OSHANIN, G. AND VOITURIEZ, R. (2004). Random walk generated by random permutations of $\{1, 2, 3, \ldots, n+1\}$. *J. Phys. A* **37**, 6221.
[50] PITMAN, J. (2006). *Combinatorial Stochastic Processes* (Lect. Notes Math. 1875). Springer, Berlin.
[51] PITMAN, J. (2018). Random weighted averages, partition structures and generalized arcsine laws. Preprint, arXiv:1804.07896.
[52] PITMAN, J. AND TANG, W. (2019). Regenerative random permutations of integers. *Ann. Prob.* **47**, 1378–1416.
[53] PITMAN, J. AND YOR, M. (1992). Arcsine laws and interval partitions derived from a stable subordinator. *Proc. London Math. Soc.* 65, 326–356.
[54] ROGERS, L. C. G. AND WILLIAMS, D. (1987). *Diffusions, Markov Processes, and Martingales*, Vol. **2**. John Wiley, New York.
[55] SALARI, K., TIBSHIRANI, R. AND POLLACK, J. R. (2010). DR-Integrator: A new analytic tool for integrating DNA copy number and gene expression data. *Bioinformatics* **26**, 414–416.
[56] SHORACK, G. R. AND WELLNER, J. A. (1986). *Empirical Processes with Applications to Statistics*. John Wiley, New York.
[57] SKOROKHOD, A. V. (1965). *Studies in the Theory of Random Processes*. Addison-Wesley Publishing Co., Inc., Reading, MA.
[58] STANLEY, R. (1976). Binomial posets, Möbius inversion, and permutation enumeration. *J. Combinatorial Theory A* **20**, 336–356.
[59] STANLEY, R. (1999). *Enumerative Combinatorics*, Vol. **2** (Camb. Studies Adv. Math. 62). Cambridge University Press.
[60] STARR, S. (2009). Thermodynamic limit for the Mallows model on S_n. *J. Math. Phys.* **50**, 095208.

[61] STRASSEN, V. (1967). Almost sure behavior of sums of independent random variables and martingales. In *Proc. Fifth Berkeley Symp. Math. Statist. Prob.*, Vol. 2.
[62] Takács, L. (1996). On a generalization of the arc-sine law. *Ann. Appl. Prob.* **6**, 1035–1040.
[63] TANG, W. (2019). Mallows ranking models: Maximum likelihood estimate and regeneration. *Proc. Mach. Learn. Res.* 97, 6125–6134.
[64] TARRAGO, P. (2018). Zigzag diagrams and Martin boundary. *Ann. Prob.* **46**, 2562–2620.
[65] VIENNOT, G. (1979). Permutations ayant une forme donnée. *Discrete Math.* **26**, 279–284.
[66] WANG, R., WATERMAN, M. AND HUANG, H. (2014). Gene coexpression measures in large heterogeneous samples using count statistics. *Proc. Nat. Acad. Sci.* 111, 16371–16376.
[67] WANG, R., LIU, K., THEUSCH, E., ROTTER, J., MEDINA, M., WATERMAN, M. AND HUANG, H. (2017). Generalized correlation measure using count statistics for gene expression data with ordered samples. *Bioinformatics* **34**, 617–624.
[68] WATANABE, S. (1995). Generalized arc-sine laws for one-dimensional diffusion processes and random walks. In *Proc. Symp. Pure Math.*, Vol. 57, pp. 157–172. American Mathematical Society, Providence, RI.
[69] WILLIAMS, D. (1969). Markov properties of Brownian local time. *Bull. Amer. Math. Soc.* **75**, 1035–1036.