Arcsine laws for random walks generated from random permutations with applications to genomics

Preprint · January 2020

7 authors, including:


Susan P Holmes
Stanford University


Adrian Röllin
National University of Singapore


Wenpin Tang
Columbia University

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

XIAO FANG, HAN LIANG GAN, SUSAN HOLMES, HAIYAN HUANG, EROL PEKÖZ, ADRIAN RÖLLIN, AND WENPIN TANG

Abstract. A classical 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.

Key words : Arcsine distribution, Brownian motion, Lévy statistics, limiting distribution, Mallows permutation, random walks, Stein's method, strong embedding, uniform permutation.
AMS 2010 Mathematics Subject Classification: 60C05, 60J65, 05A05.

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 0. 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 $[0, 1]$,
- $\Gamma := \int_0^1 1_{\{B_s>0\}} ds$ be the occupation time of B above zero before time 1.

In [45, 46], Lévy proved the celebrated result that G, $G^{\max}$ and Γ are all arcsine distributed with density

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

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,

- $\Gamma_n := \sum_{k=1}^n \mathrm{I}[S_k > 0]$ the number of times that the walk is strictly positive up to time n, and $N_n := \sum_{k=1}^n \mathrm{I}[S_{k-1} \geq 0,\, S_k \geq 0]$ the number of edges which lie above zero up to time n.

The discrete analog of Lévy's arcsine law was established by Andersen [2], where the limiting distribution (1.1) was computed by Erdös and Kac [27], and Chung and Feller [18]. Feller [29] gave the following refined treatment:

(*i*) If the increments $(X_k;\, k \geq 1)$ of the walk are exchangeable with continuous distribution, then

$$\Gamma_n \stackrel{(d)}{=} G_n^{\max}.$$

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

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

In the Brownian scaling limit, the above identities imply that $\Gamma \stackrel{(d)}{=} G^{\max} \stackrel{(d)}{=} G$. The fact that $G \stackrel{(d)}{=} G^{\max}$ also follows from Lévy's identity $(|B_t|;\, t \geq 0) \stackrel{(d)}{=} (\sup_{s \leq t} B_s - B_t;\, t \geq 0)$. See Williams [68], Karatzas and Shreve [41], Rogers and Williams [55, Section 53], Pitman and Yor [54] for various proofs of Lévy's arcsine law. The arcsine law has further been generalized in several different ways, e.g. Dynkin [25], Getoor and Sharpe [31], and Bertoin and Doney [8] to Lévy processes; Barlow, Pitman and Yor [4], and Bingham and Doney [13] to multidimensional Brownian motion; Akahori [1] and Takács [61] to Brownian motion with drift; Watanabe [67] and Kasahara and Yano [42] to one-dimensional diffusions. See also Pitman [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 rank correlation have been commonly used. 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 $\mathbf{x} = (x_1, \dots, x_n)$ and $\mathbf{y} = (y_1, \dots, y_n)$ respectively, the following statistic, denoted by W_2, was introduced in [65] to consider and aggregate possible local interactions:

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

where $\mathrm{I}[\cdot]$ 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, $\phi(0.5, 1.5, 0.2) = (3, 1, 2)$). 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 then be rewritten as

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

Several variants of W_2 have been studied to detect different types of dependent patterns between $\mathbf{x}$ and $\mathbf{y}$ (see, for example, [65] and [66]).

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 $W^* = \sum_{1\le i_1<i_2\le n} \mathrm{I}[z_{i_1} < z_{i_2}]$. 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 following family of statistics

$$W^*_m = \sum_{i=1}^{m} \mathrm{I}[z_i < z_{i+1}], \qquad 1 \le m \le n-1. \tag{1.3}$$

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 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 a mathematical convenience, considering W^*_m is equivalent to consider

$$T_m = \sum_{i=1}^{m} (2\,\mathrm{I}[z_{i+1} > z_i] - 1), \qquad 1 \le m \le n-1. \tag{1.4}$$

As argued above, exploring the properties of this process-level statistic would be useful to understand a "local" negative relationship between $\mathbf{x}$ and $\mathbf{y}$ 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 by Oshanin and Voituriez [50] in the physics literature, and also appeared in the study of the *zigzag diagrams* by Gnedin and Olshanski [33].

In this article, we consider a more general family of random permutations proposed by Mallows [48], which includes the uniform random permutation. For $0 \le q \le 1$, the one-parameter model

$$\mathbb{P}_q(\pi) = \frac{q^{\operatorname{inv}(\pi)}}{Z_{n,q}} \quad \text{for } \pi \text{ a permutation of } [n], \tag{1.5}$$

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 line of works on this random permutation model; see, for example, Diaconis [22], Gnedin and Olshanski [34], Starr [59], Basu and Bhatnagar [6], Gladkich and Peled [32], and Tang [62].

Question 1.1. For a random walk generated from the Mallows(q) permutation of $[n+1]$, what are the limit distributions of G_n/n, $G_n^{\max}/n$, Γ_n/n, or N_n/n?

For a Mallows(p) 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, 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 MacMahon [47, Vol I], Niven [49], de Bruijn [20], Carlitz [15], Stanley [57], and Viennot [64]. In particular, the number of permutations with a fixed number of descents is known as the *Eulerian number*. See also Stanley [58, Section 7.23], Borodin, Diaconis and Fulman [14, Section 5], and Chatterjee and Diaconis [17] 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. Though 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 sequel, let $(S_t;\, 0 \le t \le n)$ be the linear interpolation of the walk $(S_k;\, 0 \le k \le n)$. That is,

$$S_t = S_{j-1} + (t - j + 1)(S_j - S_{j-1}) \quad \text{for } j - 1 \le t \le j.$$

See Billingsley [11, 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*

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

Then as $n \to \infty$,

$$\left(\frac{S_{nt}}{\sqrt{n}};\, 0 \le t \le 1\right) \xrightarrow{(d)} (\mu t + \sigma B_t;\, 0 \le t \le 1), \tag{2.2}$$

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

Remark 2.2. Given the above theorem, it is a direct consequence (see Remark (3.2)) that by letting $\nu = \mu/\sigma$, $G_n/n \xrightarrow{(d)} G$, $G_n^{\max}/n \xrightarrow{(d)} G^{\max}$, $\Gamma_n/n \xrightarrow{(d)} \Gamma$ and $N_n/n \xrightarrow{(d)} \Gamma$ as $n \to \infty$ with

$$\frac{\mathbb{P}(G \in du)}{du} = \frac{e^{-\frac{\nu^2}{2}}}{\pi\sqrt{u(1-u)}} + \frac{\nu^2}{2}\int_u^1 \frac{e^{-\frac{\nu^2 y}{2}}}{\pi\sqrt{u(y-u)}}dy, \tag{2.3}$$

and

$$\begin{aligned}\frac{\mathbb{P}(G^{\max} \in du)}{du} &= \frac{\mathbb{P}(\Gamma \in du)}{du} \\ &= \frac{1}{\pi\sqrt{u(1-u)}}e^{-\frac{\nu^2}{2}} + \sqrt{\frac{2}{\pi(1-u)}}\nu e^{-\frac{\nu^2(1-u)}{2}}\Phi(\nu\sqrt{u}) \\ &\quad - \sqrt{\frac{2}{\pi u}}\nu e^{-\frac{\nu^2 u}{2}}\Phi(-\nu\sqrt{1-u}) - 2\nu^2\Phi(\nu\sqrt{u})\Phi(-\nu\sqrt{1-u}),\end{aligned} \tag{2.4}$$

where $\Phi(x) := \frac{1}{\sqrt{2\pi}}\int_{-\infty}^x \exp(-y^2/2)dy$ is the cumulative distribution function of the standard normal distribution.

The proof of Theorem 2.1 will be given in Section 3, which makes use of Gnedin-Olshanski's 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 by Tarrago [63, Proposition 9.1] in the framework of zigzag graphs. For this case, we have the following corollary.

Corollary 2.3. *Let $(S_k; 0 \le k \le n)$ be a random walk generated from the uniform permutation of $[n+1]$. Then as $n \to \infty$,*

$$\left(\frac{S_{nt}}{\sqrt{n}}; 0 \le t \le 1\right) \xrightarrow{(d)} \left(\frac{1}{\sqrt{3}} B_t; 0 \le t \le 1\right), \tag{2.5}$$

where $\xrightarrow{(d)}$ 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 2.4. For a random walk generated from the Mallows(q) permutation of $[n+1]$, find error bounds between G_n/n, $G^{\max}/n$, Γ_n/n, N_n/n and their corresponding limits (2.3)-(2.4).

While we cannot answer these questions directly, we were able to prove partial and related results. To state these, we need some notations. 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 m^{th} derivative is Lipschitz continuous. Let $\|h\|_\infty$ be the sup-norm of g, and if the k^{th} derivative of h exists, let

$$|h|_k := \left\|\frac{d^k h}{dx^k}\right\|_\infty \quad \text{and} \quad |h|_{k,1} := \sup_{x,y}\left|\frac{d^k h(x)}{dx^k} - \frac{d^k h(y)}{dy^k}\right| \frac{1}{|x-y|}.$$

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 2.6 below.

Theorem 2.5. *Let $(S_k; 0 \le k \le 2n)$ be a simple symmetric random walk. Then*

$$\mathbb{P}(N_{2n} = 2k) = \alpha_{2k,2n} \quad \textit{for } k \in \{0, \dots, n\}. \tag{2.6}$$

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

$$d_{\mathrm{W}}\left(\frac{N_{2n}}{2n}, Z\right) \le \frac{27}{2n} + \frac{8}{n^2}. \tag{2.7}$$

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

$$\left|\mathbb{E}h\left(\frac{N_{2n}}{2n}\right) - \mathbb{E}h(Z)\right| \le \frac{4|h|_2 + |h|_{2,1}}{64n} + \frac{|h|_{2,1}}{64n^2}. \tag{2.8}$$

Identity (2.6) can be found in [29], the bound (2.7) was proved by [35], and the proof of (2.8) is deferred to Section 4.

Conjecture 2.6. *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], Bernardi, Duplantier and Nadeau 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

$$c_{2k,2n+1} = \binom{2n+1}{2k} b_{2k} b_{2n-2k+1}. \tag{2.9}$$

Note that, for the special cases $k = 0$ and $k = n$, the formula (2.9) agrees with the known results in [7]. The formula (2.9) suggests a bijection between the set of walks generated from permutations of $[2n+1]$ with $2k$ positive edges and the set of pairs of positive walks generated from permutations of $[2k]$ and $[2n-2k+1]$ respectively. A naive idea is to break the walk into positive and negative excursions, and exclude the final visit to the origin before crossing 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 2.6 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 of Skorokhod [56], and Komlós, Major and Tusnády [43, 44] 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 2.7. *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*

$$\beta := \frac{2}{\sigma(1+q)} \quad \text{and} \quad \eta := \frac{2q}{1-q+q^2}. \tag{2.10}$$

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

$$\mathbb{P}\left(\sup_{0\le t\le n}\left|\frac{1}{\sigma}(S_t - \mu t) - B_t\right| > c_1 n^{\frac{1+\varepsilon}{4}} (\log n)^{\frac{1}{2}} \beta\right) \le \frac{c_2(\beta^6 + \eta)}{\beta^2 n^{\varepsilon} \log n}. \tag{2.11}$$

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

Also note that there is a substantial literature studying the relations between random permutations and Brownian motion. Classical results were surveyed in Arratia, Barbour and Tavaré [3], and Pitman [52]. See also Janson [40], Hoffman, Rizzolo and Slivken [36, 37], and Bassino, Bouvel, Féray, Gerin and Pierrot [5] 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, then calculate its moments and use an invariance principle.

3.1. Mallows(q) permutations.

Gnedin and Olshanski [34] provide a nice construction of the Mallows(q) permutation, which is implicit in the original work of Mallows [48]. 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

$$\mathbb{P}(\mathcal{G}_{q,n} = k) = \frac{q^{k-1}(1-q)}{1-q^n} \quad \text{for } k \in [n]. \tag{3.1}$$

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

- $\pi_1 := Y_1$,
- for $k \geq 2$, let $\pi_k := \psi(Y_k)$ where ψ is the increasing bijection from $[n-k+1]$ to $[n] \setminus \{\pi_1, \pi_2, \cdots, \pi_{k-1}\}$.

That is, pick π_1 according to $\mathcal{G}_{q,n}$, and remove π_1 from $[n]$. Then pick π_2 as the $\mathcal{G}_{q,n-1}^{\text{th}}$ smallest element of $[n] \setminus \{\pi_1\}$, and remove π_2 from $[n] \setminus \{\pi_1\}$, and so on. As 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, $\mathbb{P}(X_k = 1) = \mathbb{P}(\mathcal{G}_{q,n+1-k} \leq \mathcal{G}_{q,n-k}) = 1/(1+q)$ 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* hence *one-dependent*; see de Valk [21] for background.

Such construction is also used by Gnedin and Olshanski [34] to construct a random permutation of positive integers, called the *infinite q-shuffle*. The latter is further extended by Pitman and Tang [53] to *p-shifted permutations* as an instance of regenerative permutations, and used by Holroyd, Hutchcroft and Levy [38] 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; that is,

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

where $\mathcal{N}(0,1)$ is standard normal distributed. See Chatterjee and Diaconis [17, Section 3] for a survey of six different approaches to prove 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 (Borodin, Diaconis and Fulman, Proposition 5.2 [14])**.** *Fix $0 < q \leq 1$, let π be the Mallows(q) permutation of $[n]$, and let $\#\mathcal{D}(\pi)$ be the number of descents of π.*

Then

$$\mathbb{E}\#\mathcal{D}(\pi) = \frac{(n-1)q}{1+q} \quad \textit{and} \quad \operatorname{Var}\#\mathcal{D}(\pi) = q\frac{(1-q+q^2)n-1+3q-q^2}{(1+q)^2(1+q+q^2)}. \tag{3.2}$$

Moreover,

$$\frac{1}{\sqrt{n}}\left(\#\mathcal{D}(\pi) - \frac{nq}{1+q}\right) \xrightarrow{(d)} \mathcal{N}\left(0, \frac{q(1-q+q^2)}{(1+q)^2(1+q+q^2)}\right). \tag{3.3}$$

We are now ready to prove Theorem 2.1.

Proof of Theorem 2.1. Since the increments of a permutation generated random walk are 1-dependent, the functional CLT is an immediate consequence of [12, Theorem 5.1] and the moments in Lemma 3.1. □

3.2. **Lévy statistics of Brownian motion with drift.** Let $B_t^{\mu,\sigma} := \mu t + \sigma B_t$ be Brownian motion with drift μ and variance σ^2. For $\mu = 0$, the Lévy statistics G, $G^{\max}$ and Γ are all arcsine distributed. The following remark gives a summary of the distributions of these Lévy statistics of Brownian motion with drift.

Remark 3.2. Let G, $G^{\max}$ and Γ be the Lévy statistics defined for $(B_t^{\mu,\sigma};\, t \ge 0)$. Then by letting $\nu := \mu/\sigma$,

(i) the distribution of G is given by (2.3),

(ii) $G^{\max}$ has the same distribution as Γ, with the distribution given by (2.4).

Part (i) can be derived by Girsanov's change of variables, see Iafrate and Orsingher [39, Theorem 2.1] for details. For part (ii), the fact that $G^{\max} \overset{(d)}{=} \Gamma$ for $B^{\mu,\sigma}$ follows from a path transform of Embrechts, Rogers and Yor [26, (1.b)]. The density formula (2.4) can be read from Akahori [1, Theorem 1.1(i)], see also Takács [61], and Doney and Yor [24] for various proofs.

4. Proof of Theorem 2.5

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 characterising operator.

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

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

for all functions f in a 'rich enough' family of test functions.

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 $h \in \mathrm{BC}^{2,1}$, assume we have a function f that solves

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

Now, replacing x by W in (4.1) and taking expectation, 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 utilising properties of f. Extending the work of Döbler [23], Goldstein and Reinert [35] developed Stein's method for the beta distribution (of which arcsine is special case) and gave an explicit Wasserstein bound between the discrete and the continuous arcsine distributions. We will use the framework from Gan, Röllin and Ross [30] to calculate error bounds for the class of test functions $\mathrm{BC}^{2,1}$.

4.2. **Proof of Theorem 2.5.** To simplify the notation, let

$$W_n := N_{2n}/2n$$

be the fraction of positive edges of a random walk generated from the uniform permutation. 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,

$$\mathbb{E}W_n = \frac{1}{2} \quad \text{and} \quad \mathbb{E}W_n^2 = \frac{3}{8} + \frac{1}{8n}. \tag{4.3}$$

The identity (4.2) can be read from Döbler [23]. The moments are easily derived by plugging in $f(x) = 0$ and $f(x) = x$; see Goldstein and Reinert [35].

Proof of Theorem 2.5. The distribution (2.6) of N_{2n} can be found in Feller [29]. The bound (2.7) follows from the fact that N_{2n} is discrete arcsine distributed, together with Theorem 1.2 of Goldstein and Reinert [35].

We prove the bound (2.8) using the generator method. Recall the Stein equation (4.1) for the arcsine distribution. First set $f = g'$, 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 by

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

and the first term is bounded by

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

where the last equality follows from (4.3). Combining (4.4), (4.5) with Theorem 5 of Gan, Röllin and Ross [30] (for relating the bounds on derivatives g with derivatives of h) yields the desired bound. $\square$

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 $|\mathbb{E}h(W_n) - \mathbb{E}h(Z)| \le \frac{1}{16n} + \frac{1}{64n^2}$.

5. Proof of Theorem 2.7

In this section, we prove Theorem 2.7. 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 $(X_i;\, 1 \le i \le n)$ be a sequence of m-dependent random variables. That is, $(X_1, \ldots, X_j)$ are independent of $(X_{j+m+1}, \ldots, X_n)$ for each $j \in [n-m-1]$. Let $(S_k;\, 0 \le k \le n)$ be a random walk with increments X_i, and $(S_t;\, 0 \le t \le n)$ be the linear interpolation of $(S_k;\, 0 \le k \le n)$. Assume that the random variables X_i are centered and scaled such that

$$\mathbb{E}X_i = 0 \text{ for all } i \in [n] \quad \text{and} \quad \operatorname{Var}(S_n) = n.$$

Let $(B_t;\, t \ge 0)$ be one-dimensional Brownian motion starting at 0. The idea of strong embedding is to couple $(S_t;\, 0 \le t \le n)$ and $(B_t;\, 0 \le t \le n)$ in such a way that

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

for some $b_n = o(n^{\frac{1}{2}})$ and $p_n = o(1)$ as $n \to \infty$.

The study of such embeddings dates back to Skorokhod [56]. When X's are independent and identically distributed, Strassen [60] obtained (5.1) with $b_n = \mathcal{O}(n^{\frac{1}{4}}(\log n)^{\frac{1}{2}}(\log\log n)^{\frac{1}{4}})$. Csörgő and Révész [19] 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 $\varepsilon > 0$. Komlós, Major and Tusnády [43, 44] further obtained $b_n = \mathcal{O}(\log n)$ under a finite moment generating function assumption. See also [10, 16] for recent developments.

We use the argument of Csörgő and Révész [19] 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. Assume that $1 \le m \le n^{\frac{1}{2}}$, $\mathbb{E}X_i = 0$ for each $i \in [n]$, and $\operatorname{Var} S_n = n + \mathcal{O}(1)$. Further assume that $|X_i| \le \beta$ for each $i \in [n]$, where $\beta > 0$ is a constant. Let*

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

For any $\varepsilon \in (0,1)$, if $\eta \le n^{\varepsilon}$, then there exist universal constants $n_0, c_1, c_2 > 0$ 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 with

$$\mathbb{P}\left(\sup_{0 \le t \le n} |S_t - B_t| > c_1 n^{\frac{1+\varepsilon}{4}} (\log n)^{\frac{1}{2}} m^{\frac{1}{2}} \beta\right) \le \frac{c_2(m^4\beta^6 + \eta)}{m\beta^2 n^{\varepsilon} \log n} \tag{5.3}$$

If m and β are constants and $\operatorname{Var}(S_{j+k} - S_j)$ matches k up to 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 $\varepsilon \in (0,1)$.

Proof of Theorem 2.7. 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 3.1,

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

Elementary computation shows that for $k \le n$, $\operatorname{Var} S'_k = k + \eta$, 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. We use C and c to denote positive constants which may differ in expressions. Let

$$d := \lceil n^{\frac{1-\varepsilon}{2}} \rceil, \tag{5.4}$$

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*

$$3m\beta^2 \ge 1 \quad \textit{and} \quad l \ge 6m \log n \tag{5.5}$$

for sufficiently large n.

Proof. Note that

$$n = \operatorname{Var} S_n = \sum_{i=1}^{n} \sum_{j:|j-i|\le m} \mathbb{E} X_i X_j \le n(2m+1)\beta^2 \text{ and } m \ge 1,$$

which implies $3m\beta^2 \ge 1$. The second bound follows from the fact that $m \le n^{\frac{1}{2}}$ and $l \sim n^{\frac{1+\varepsilon}{2}}$. □

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

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

where $\Gamma(\mu, \nu)$ is the space of all probability measures on $\mathbb{R}^2$ with μ and ν as marginals. In the next two lemmas, $\mathcal{N}(\mu, \sigma^2)$ denotes a normal random variable with mean μ and variance σ^2.

Lemma 5.3. *Under the assumptions in Theorem 5.1, we have for n sufficiently large,*

$$d_{W_2}(S_{l-m}, \mathcal{N}(0, \sigma^2)) \le C m^2 \beta^3, \tag{5.6}$$

where $\sigma^2 := \operatorname{Var} S_{l-m}$.

Proof. Specializing Fang [28, Corollary 2.3] to sums of m-dependent and bounded variables, we get

$$d_{W_2}(S_{l-m}, \mathcal{N}(0,\sigma^2)) = \sigma\, d_{W_2}(\sigma^{-1}S_{l-m}, \mathcal{N}(0,1))$$
$$\le \sigma C\left(lm^2\left(\frac{\beta}{\sigma}\right)^3 + \left(lm^3\left(\frac{\beta}{\sigma}\right)^4\right)^{\frac{1}{2}}\right) \le Cm^2\beta^3,$$

where we used $3m\beta^2 \ge 1$ in (5.5), and $\sigma^2 \ge l - m - \eta \ge cl$ for sufficiently large n from $m \le n^{\frac{1}{2}}$, $\eta \le n^{\varepsilon}$, $l \sim n^{\frac{1+\varepsilon}{2}}$ and $\varepsilon \in (0,1)$. □

Lemma 5.4. *There exists a coupling of* $(S_t;\, 0 \le t \le n)$ *and* $(B_t;\, 0 \le t \le n)$ *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 $(e_1, \dots, e_d)$ *are 1-dependent, and*

$$\mathbb{E}e_j^2 \le C(m^4\beta^6 + \eta), \quad \textit{for all } j \in [n].$$

Proof. We use $3m\beta^2 \ge 1$ below implicitly to absorb a few terms into $Cm^4\beta^6$. With σ^2 defined in Lemma 5.3, we have

$$d_{W_2}(\mathcal{N}(0,\sigma^2), \mathcal{N}(0,l)) \le \sqrt{|l-\sigma^2|} \le \sqrt{m+\eta}.$$

Combining (5.6), the above bound and the m-dependence, 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

$$\mathbb{E}[(S_{\lceil jn/d\rceil - m} - S_{\lceil (j-1)n/d\rceil}) - (B_{\lceil jn/d\rceil} - B_{\lceil (j-1)n/d\rceil})]^2 \le C(m^4\beta^6 + \eta).$$

By the m-dependence assumption, we can generate $X_1, \dots, X_n$ from their conditional distribution given $(S_{\lceil jn/d\rceil - m} - S_{\lceil (j-1)n/d\rceil};\, j \in [d])$, thus obtaining $(S_t;\, 0 \le t \le n)$, and generate $(B_t; 0 \le t \le n)$ given $(B_{\lceil jn/d\rceil};\, j \in [d])$. Since

$$\mathbb{E}(S_{\lceil jn/d\rceil} - S_{\lceil jn/d\rceil - m})^2 \le Cm^2\beta^2,$$

we have

$$\mathbb{E}(e_j^2) \le C(m^4\beta^6 + \eta)$$

Finally, the 1-dependence of $(e_1, \dots, e_d)$ follows from the m-dependence assumption. □

Lemma 5.5. *Let* $T_j = \sum_{i=1}^{d} e_i$, $j \in [d]$. *For each* $b > 0$, *we have*

$$\mathbb{P}\left(\max_{j\in[d]} |T_j| > b\right) \le C(m^4\beta^6 + \eta)d/b^2.$$

Proof. Define

$$T_j^{(1)} = \sum_{\substack{i=1,3,5,\dots \\ i\le j}} e_i, \quad T_j^{(2)} = \sum_{\substack{i=2,4,6,\dots \\ i\le j}} e_i.$$

By Lemma 5.4, $T_j^{(1)}$ is a sum of independent random variables with zero mean and finite second moments. By Kolmogorov's maximal inequality,

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

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

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

$\square$

Lemma 5.6. *For any $0<b\le 4l\beta$, we have*

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

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

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

we have for $\theta>0$,

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

for $\theta(2m+1)\beta\le 1$. This implies that $\log h(\theta)\le 3lm\beta^2\theta^2$, and

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

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

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

$\square$

Lemma 5.7. *For each $b>0$, we have*

$$\mathbb{P}\left(\sup_{0\le t\le l}|B_t-tB_l/l|>b\right)\le 2e^{-\frac{b^2}{2l}}.$$

Proof. We have, by symmetry, a reflection argument for Brownian bridge and a normal tail bound,

$$\begin{aligned}\mathbb{P}\left(\sup_{0\le t\le l}|B_t - tB_l/l| > b\right) &\le 2\mathbb{P}\left(\sup_{0\le t\le l}(B_t - tB_l/l) > b\right)\\ &\le 4\mathbb{P}(B_l > b) \le 2e^{-\frac{b^2}{2l}}.\end{aligned}$$

□

Now we proceed to proving Theorem 5.1.

Proof of Theorem 5.1. Let

$$b = (96lm\beta^2 \log n)^{1/2}.$$

It satisfies $b \le 4l\beta$ in Lemma 5.6 since $m \le l/(6\log n)$ by (5.5). Note 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 each subinterval is larger than b. By the union bound and Lemmas 5.5–5.7, we have

$$\mathbb{P}\left(\sup_{0\le t\le n}|S_t - B_t| > 3b\right) \le \frac{C(m^4\beta^6 + \eta)}{m\beta^2 n^{\varepsilon}\log n}.$$

This proves the theorem. □

Acknowledgments

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 the 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 brings 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.

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

References

[1] J. Akahori (1995). Some formulae for a new type of path-dependent option. *Ann. Appl. Probab.* **5**, 383–388.

[2] E. S. Andersen (1953). On sums of symmetrically dependent random variables. *Skand. Aktuarietidskr.* **36**, 123–138.

[3] R. Arratia, A. Barbour and S. Tavaré (2003). *Logarithmic combinatorial structures: a probabilistic approach*, volume 1. European Mathematical Society.

[4] M. Barlow, J. Pitman and M. Yor (1989). Une extension multidimensionnelle de la loi de l'arc sinus. In *Séminaire de Probabilités XXIII*, pages 294–314.

[5] F. Bassino, M. Bouvel, V. Féray, L. Gerin and A. Pierrot (2018). The Brownian limit of separable permutations. *Ann. Probab.* **46**, 2134–2189.

[6] R. Basu and N. Bhatnagar (2017). Limit theorems for longest monotone subsequences in random Mallows permutations. *Ann. Inst. H. Poincaré Probab. Statist.* **53**, 1934–1951.

[7] O. Bernardi, B. Duplantier and P. Nadeau (2010). A bijection between well-labelled positive paths and matchings. *Séminaire Lotharingien de Combinatoire* **63**, B63e.

[8] J. Bertoin and R. Doney (1997). Spitzer's condition for random walks and Lévy processes. *Ann. Inst. H. Poincaré Probab. Statist.* **33**, 167–178.

[9] J. Bertoin (1993). Splitting at the infimum and excursions in half-lines for random walks and Lévy processes. *Stochastic processes and their applications* **47**, 17–35.

[10] C. Bhattacharjee and L. Goldstein (2016). On strong embeddings by Stein's method. *Electronic Journal of Probability* **21**.

[11] P. Billingsley (1999). *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, second edition.

[12] P. Billingsley (1956). The invariance principle for dependent random variables. *Transactions of the American Mathematical Society* **83**, 250–268.

[13] N. Bingham and R. Doney (1988). On higher-dimensional analogues of the arc-sine law. *Journal of Applied Probability* **25**, 120–131.

[14] A. Borodin, P. Diaconis and J. Fulman (2010). On adding a list of numbers (and other one-dependent determinantal processes). *Bull. Amer. Math. Soc. (N.S.)* **47**, 639–670.

[15] L. Carlitz (1973). Permutations with prescribed pattern. *Math. Nachr.* **58**, 31–53.

[16] S. Chatterjee (2012). A new approach to strong embeddings. *Probability Theory and Related Fields* **152**, 231–264.

[17] S. Chatterjee and P. Diaconis (2017). A central limit theorem for a new statistic on permutations. *Indian Journal of Pure and Applied Mathematics* **48**, 561–573.

[18] K. L. Chung and W. Feller (1949). On fluctuations in coin-tossing. *Proceedings of the National Academy of Sciences*pages 605–608.

[19] M. Csörgő and P. Révész (1975). A new method to prove strassen type laws of invariance principle. I. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **31**, 255–259.

[20] N. G. de Bruijn (1970). Permutations with given ups and downs. *Nieuw Arch. Wisk. (3)* **18**, 61–65. ISSN 0028-9825.

[21] V. de Valk (1994). *One-dependent Processes: Two-block-factors and Non-two-block-factors*. CWI Tracts.

[22] P. Diaconis (1988). *Group representations in probability and statistics*, volume 11. Lecture Notes-Monograph Series.

[23] C. Döbler (2012). A rate of convergence for the arcsine law by Stein's method. *arXiv:1207.2401*.

[24] R. Doney and M. Yor (1998). On a formula of Takács for Brownian motion with drift. *Journal of Applied Probability* **35**, 272–280.

[25] E. B. Dynkin (1965). *Markov processes. Vols. I, II*, volume 122 of *Die Grundlehren der Mathematischen Wissenschaften, Bände 121.* Springer-Verlag, Berlin-Göttingen-Heidelberg.

[26] P. Embrechts, L. C. G. Rogers and M. Yor (1995). A proof of Dassios' representation of the α-quantile of Brownian motion with drift. *Ann. Appl. Probab.* **5**, 757–767.

[27] P. Erdös and M. Kac (1947). On the number of positive sums of independent random variables. *Bull. Amer. Math. Soc.* **53**, 1011–1020. ISSN 0002-9904.

[28] X. Fang (2018). Wasserstein-2 bounds in normal approximation under local dependence. *arXiv:1807.05741.*

[29] W. Feller (1968). *An Introduction to Probability Theory and Its Applications. Vol. I.* Second edition. John Wiley & Sons, Inc., New York-London-Sydney.

[30] H. L. Gan, A. Röllin and N. Ross (2017). Dirichlet approximation of equilibrium distributions in Cannings models with mutation. *Advances in Applied Probability* **49**, 927–959.

[31] R. Getoor and M. Sharpe (1994). On the arc-sine laws for Lévy processes. *Journal of Applied Probability* **31**, 76–89.

[32] A. Gladkich and R. Peled (2018). On the cycle structure of Mallows permutations. *Ann. Probab.* **46**, 1114–1169.

[33] A. Gnedin and G. Olshanski (2006). Coherent permutations with descent statistic and the boundary problem for the graph of zigzag diagrams. *Int. Math. Res. Not.* Art. ID 51968, 39.

[34] A. Gnedin and G. Olshanski (2010). q-exchangeability via quasi-invariance. *Ann. Probab.* **38**, 2103–2135.

[35] L. Goldstein and G. Reinert (2013). Stein's method for the beta distribution and the Polya-Eggenberger urn. *Journal of Applied Probability* **50**, 1187–1205.

[36] C. Hoffman, D. Rizzolo and E. Slivken (2017). Pattern-avoiding permutations and Brownian excursion part I: Shapes and fluctuations. *Random Structures & Algorithms* **50**, 394–419.

[37] C. Hoffman, D. Rizzolo and E. Slivken (2017). Pattern-avoiding permutations and Brownian excursion, part II: fixed points. *Probability Theory and Related Fields* **169**, 377–424.

[38] A. Holroyd, T. Hutchcroft and A. Levy (2017). Mallows permutations and finite dependence. *arXiv:1706.09526.*

[39] F. Iafrate and E. Orsingher (2017). The last zero crossing of an iterated Brownian motion with drift. *arXiv:1803.00877.*

[40] S. Janson (2017). Patterns in random permutations avoiding the pattern 132. *Combinatorics, Probability and Computing* **26**, 24–51.

[41] I. Karatzas and S. E. Shreve (1987). A decomposition of the Brownian path. *Statist. Probab. Lett.* **5**, 87–93.

[42] Y. Kasahara and Y. Yano (2005). On a generalized arc-sine law for one-dimensional diffusion processes. *Osaka Journal of Mathematics* **42**, 1–10.

[43] J. Komlós, P. Major and G. Tusnády (1975). An approximation of partial sums of independent RV'-s, and the sample DF. I. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **32**, 111–131.

[44] J. Komlós, P. Major and G. Tusnády (1976). An approximation of partial sums of independent RV's, and the sample DF. II. *Zeitschrift für Wahrscheinlichkeitstheorie*

und verwandte Gebiete **34**, 33–58.
[45] P. Lévy (1939). Sur certains processus stochastiques homogènes. *Compositio Math.* **7**, 283–339. ISSN 0010-437X.
[46] P. Lévy (1965). *Processus stochastiques et mouvement brownien.* Suivi d'une note de M. Loève. Deuxième édition revue et augmentée. Gauthier-Villars & Cie, Paris.
[47] P. A. MacMahon (1960). *Combinatory analysis.* Two volumes (bound as one). Chelsea Publishing Co., New York.
[48] C. L. Mallows (1957). Non-null ranking models. I. *Biometrika* **44**, 114–130.
[49] I. Niven (1968). A combinatorial problem of finite sequences. *Nieuw Arch. Wisk* **16**, 116–123.
[50] G. Oshanin and R. Voituriez (2004). Random walk generated by random permutations of $\{1, 2, 3, \ldots, n{+}1\}$. *Journal of Physics A: Mathematical and General* **37**, 6221.
[51] J. Pitman (2018). Random weighted averages, partition structures and generalized arcsine laws. *arXiv:1804.07896.*
[52] J. Pitman (2006). *Combinatorial stochastic processes*, volume 1875 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin. ISBN 978-3-540-30990-1; 3-540-30990-X.
[53] J. Pitman and W. Tang (2019). Regenerative random permutations of integers. *Ann. Probab.* **47**, 1378–1416.
[54] J. Pitman and M. Yor (1992). Arcsine laws and interval partitions derived from a stable subordinator. *Proc. London Math. Soc. (3)* **65**, 326–356.
[55] L. C. G. Rogers and D. Williams (1987). *Diffusions, Markov processes, and martingales. Vol. 2.* Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York. ISBN 0-471-91482-7.
[56] A. V. Skorokhod (1965). *Studies in the theory of random processes.* Translated from the Russian by Scripta Technica, Inc. Addison-Wesley Publishing Co., Inc., Reading, Mass.
[57] R. Stanley (1976). Binomial posets, möbius inversion, and permutation enumeration. *Journal of Combinatorial Theory, Series A* **20**, 336–356.
[58] R. Stanley (1999). *Enumerative combinatorics. Vol. 2*, volume 62 of *Cambridge Studies in Advanced Mathematics.* Cambridge University Press, Cambridge.
[59] S. Starr (2009). Thermodynamic limit for the Mallows model on S_n. *Journal of mathematical physics* **50**, 095208.
[60] V. Strassen (1967). Almost sure behavior of sums of independent random variables and martingales. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, volume 2.*
[61] L. Takács (1996). On a generalization of the arc-sine law. *Ann. Appl. Probab.* **6**, 1035–1040.
[62] W. Tang (2019). Mallows ranking models: maximum likelihood estimate and regeneration. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 6125–6134.
[63] P. Tarrago (2018). Zigzag diagrams and Martin boundary. *Ann. Probab.* **46**, 2562–2620.
[64] G. Viennot (1979). Permutations ayant une forme donnée. *Discrete Math.* **26**, 279–284.
[65] R. Wang, M. Waterman and H. Huang (2014). Gene coexpression measures in large heterogeneous samples using count statistics. *Proceedings of the National Academy*

of Sciences **111**, 16371–16376.

[66] R. Wang, K. Liu, E. Theusch, J. Rotter, M. Medina, M. Waterman and H. Huang (2017). Generalized correlation measure using count statistics for gene expression data with ordered samples. *Bioinformatics* **34**, 617–624.

[67] S. Watanabe (1995). Generalized arc-sine laws for one-dimensional diffusion processes and random walks. In *Proceedings of Symposia in Pure Mathematics*, volume 57, pages 157–172.

[68] D. Williams (1969). Markov properties of Brownian local time. *Bull. Amer. Math. Soc.* **75**, 1035–1036. ISSN 0002-9904.

Department of Statistics, The Chinese University of Hong Kong.
E-mail address: `xfang@sta.cuhk.edu.hk`

Department of Mathematics, Northwestern University.
E-mail address: `ganhl@math.northwestern.edu`

Department of Statistics, Stanford University.
E-mail address: `susan@stat.stanford.edu`

Department of Statistics, University of California, Berkeley.
E-mail address: `hhuang@stat.berkeley.edu`

Questrom School of Business, Boston University.
E-mail address: `pekoz@bu.edu`

Department of Statistics and Applied Probability, National University of Singapore.
E-mail address: `adrian.roellin@nus.edu.sg`

Department of Industrial Engineering and Operations Research, UC Berkeley.
E-mail address: `wenpintang@stat.berkeley.edu`