Logical limit laws for Mallows random permutations

Tobias Müller[*] Fiona Skerman[†] Teun W. Verstraaten[‡]

May 26, 2024

Abstract

A random permutation Π_n of $\{1,\dots,n\}$ follows the Mallows(n,q) distribution with parameter $q>0$ if $\mathbb{P}(\Pi_n=\pi)$ is proportional to $q^{\mathrm{inv}(\pi)}$ for all π. Here $\mathrm{inv}(\pi) := |\{i<j : \pi(i)>\pi(j)\}|$ denotes the number of inversions of π. We consider properties of permutations that can be expressed by the sentences of two different logical languages. Namely, *the theory of one bijection* (TOOB), which describes permutations via a single binary relation, and *the theory of two orders* (TOTO), where we describe permutations by two total orders. We say that the *convergence law* holds with respect to one of these languages if, for every sentence φ in the language, the probability $\mathbb{P}(\Pi_n$ satisfies $\varphi)$ converges to a limit as $n\to\infty$. If moreover that limit is $\in\{0,1\}$ for all sentences, then the *zero–one law* holds.

We will show that with respect to TOOB the Mallows(n,q) distribution satisfies the zero–one law when $0<q<1$ is fixed, and for fixed $q>1$ the convergence law fails. (In the case when $q=1$ Compton [15] has shown the convergence law holds but not the zero–one law.)

We will prove that with respect to TOTO the Mallows(n,q) distribution satisfies the convergence law but not the zero–one law for any fixed $q\neq 1$, and that if $q=q(n)$ satisfies $1-1/\log^* n < q < 1+1/\log^* n$ then Mallows(n,q) fails the convergence law. Here $\log^*$ denotes the discrete inverse of the tower function.

1 Introduction

Throughout the paper, we denote by $[n] := \{1,\dots,n\}$ the first n positive integers and by S_n the set of permutations on $[n]$. A pair $i,j\in[n]$ is an inversion of the permutation $\pi\in S_n$ if $i<j$ and $\pi(i)>\pi(j)$. We denote by $\mathrm{inv}(\pi)$ the number of inversions of a permutation π.

For $n\in\mathbb{N}$ and $q>0$, the Mallows distribution Mallows(n,q) samples a random element Π_n of S_n such that for all $\pi\in S_n$ we have

$$\mathbb{P}(\Pi_n=\pi)=\frac{q^{\mathrm{inv}(\pi)}}{\sum_{\sigma\in S_n} q^{\mathrm{inv}(\sigma)}}. \tag{1}$$

In particular, if $q=1$ then the Mallows distribution is simply the uniform distribution on S_n.

The Mallows distribution was first introduced by C.L. Mallows [44] in 1957 in the context of statistical ranking theory. It has since been studied in connection with a diverse range of topics, including mixing times of Markov chains [7, 18], finitely dependent colorings of the integers [34], stable matchings [5], random binary search trees [1], learning theory [12, 58], q-analogs of exchangeability [25, 26], determinantal point processes [10], statistical physics [56, 57], genomics [22] and random graphs with tunable properties [20].

[*]Bernoulli Institute for Mathematics, CS and AI, Groningen University, `tobias.muller@rug.nl`. Supported in part by the Netherlands Organisation for Scientific Research (NWO) under project nos 612.001.409 and 639.032.529.

[†]Department of Mathematics, Uppsala University, `fiona.skerman@math.uu.se`. Supported by AI4Research at Uppsala University and by the Wallenberg AI, Autonomous Systems and Software Program (WASP).

[‡]Bernoulli Institute for Mathematics, CS and AI, Groningen University, `t.w.verstraaten@rug.nl`.

Properties of the Mallows distribution that have been investigated include pattern avoidance [16, 17, 49], the number of descents [30], the longest monotone subsequence [6, 8, 46], the longest common subsequence of two Mallows permutations [37] and the cycle structure [24, 31].

In the present paper we will study the Mallows distribution from the perspective of first order logic. Given a sequence of random permutations $(\Pi_n)_{n\geq 1}$, we say that the *convergence law* holds with respect to some fixed logical language describing permutations if the limit $\lim_{n\to\infty} \mathbb{P}\left(\Pi_n \models \varphi\right)$ exists, for every sentence φ in the language. Here and in the rest of the paper the notation $\pi \models \varphi$ denotes that the sentence φ holds for the permutation π. If this limit is either 0 or 1 for all such φ then we say that the *zero–one law* holds. Following [2], we will consider two different logical languages for permutations: the Theory of One Bijection (TOOB) and the Theory of Two Orderings (TOTO). Here we give an informal overview of them. More precise definitions follow in Section 3.3.

In TOOB, we can express a property of a permutation using variables representing the elements of the domain of the permutation, the usual quantifiers $\exists, \forall$, the usual logical connectives $\vee, \wedge, \neg, \dots$, parantheses and two binary relations $=, R$. Here $=$ has the usual meaning ($x = y$ denotes that the variables x, y represent the same element of the domain of the permutation) and $R(x, y)$ holds if and only if $\pi(x) = y$. We are for instance able to query in TOOB if a permutation has a fixed point by

$$\exists x : R(x, x).$$

As is shown [2] (Proposition 3 and the comment following it), we cannot express by a TOOB sentence whether or not a permutation contains the pattern 231. (A permutation π contains the pattern 231 if there exist $i < j < k$ such that $\pi(k) < \pi(i) < \pi(j)$.)

The logical language TOTO is constructed similarly to TOOB. Instead of the relation R there now are two relations $<_1, <_2$. The relation $<_1$ represents the usual linear order on the domain $[n]$ of $\pi \in S_n$ and the relation $<_2$ represents the usual linear order of the images $\pi(1), \dots, \pi(n)$. That is, $x <_1 y$ if and only if $x < y$, while $x <_2 y$ if and only if $\pi(x) < \pi(y)$. In TOTO we can for instance express that a permutation contains the pattern 231, via

$$\exists x, y, z : (x <_1 y) \wedge (y <_1 z) \wedge (z <_2 x) \wedge (x <_2 y).$$

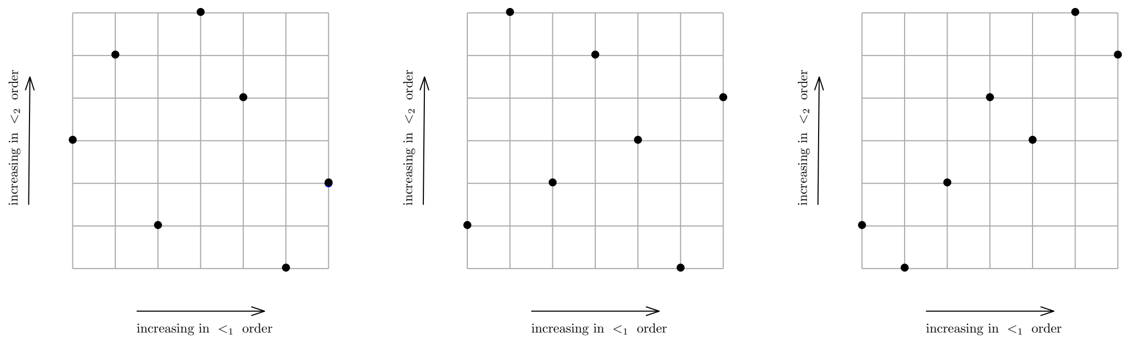
See Figure 1.1 for an illustration. We can also express that $\pi(1) < \pi(2)$ by:

$$\exists x, y : (x <_1 y) \wedge (x <_2 y) \wedge (\forall z : (z <_1 y) \to (z = x)).$$

See Section 3 of [2] for generalizations of pattern containment that can be expressed in TOTO. On the other hand, in TOTO we cannot express whether or not a permutation has a fixed point, as shown in Corollary 27 of [2].

It is not hard to see that for a uniform random permutation the probability that $\pi(1) < \pi(2)$ equals $1/2$. So in particular, for uniform permutations TOTO does not satisfy a zero-one law. See also the remark following Question 1 in [2]. What is more, in TOTO uniform permutations do not even satisfy the convergence law as was first shown by Foy and Woods [23]. Let us also mention two very recent results on logical limit laws for random permutations. In [3] it is shown the uniform distribution on the set of permutations in S_n that avoid 231 admits a convergence law in TOTO, and in [11] it is shown the same result holds for the uniform distribution on the class of layered permutations.

Both TOOB and TOTO fall under the umbrella of *first order* logical languages, as they only allow quantification over the elements of the domain. In contrast, second order logic also allows us to quantify over relations. Second order logic is much more powerful than first order logic. It is for instance possible to express in second order logic the property that the domain has an even number of elements. In particular, the convergence law will fail for second order logic, for trivial reasons.



(a) the initial three elements induce the pattern 231.

(b) contains 231 but not on the initial three elements.

(c) does not contain 231.

Figure 1.1

The study of first order properties of random permutations is part of a larger theory of first order properties of random discrete structures. See for instance the monograph [54]. Examples of structures for which the first order properties have been studied include the Erdős-Rényi random graph (see e.g. [36],[54]), Galton-Watson trees [50], bounded degree graphs [39], random graphs from minor-closed classes [32], random perfect graphs [47], random regular graphs [29] and the very recent result on bounded degree uniform attachment graphs [45].

Let us also mention some work on *random orders* that is closely related to the topic of the present paper. The logic of random orders was introduced in [59] and has a single binary relation $<$. To sample a "k-dimensional" random order on $[n]$ pick k random orderings $<_1, \ldots, <_k$ on $[n]$ and set $x < y$ if $x <_i y$ in each of the orders $i = 1, \ldots, k$.

Non-convergence was proven for two-dimensional random orders in [53] by constructing a first–order sentence φ for which no limit probability exists. Replacing every occurrence of $<$ in this sentence by $(x <_1 y) \wedge (x <_2 y)$ yields an alternative proof of non-convergence also in TOTO for uniformly random permutations. Many properties of random orders are known [4, 9], see [13] and [14] for surveys of different models of random partial orders and also their relation to theories of spacetime [14].

An appeal of TOTO is that it allows one to express pattern containment. Permutation patterns arise naturally in statistics. Let (X, Y) be drawn from a continuous distribution on $\mathbb{R}^2$ and suppose we have n random samples (X_i, Y_i). Many statistical tests rely only on the relative ordering in the two dimensions, i.e. on the permutation induced by the n points. For example the Kendall rank correlation coefficient is $\tau = 1 - 2\,\mathrm{inv}(\pi)/\binom{n}{2}$ and indeed there is a test for independence of X and Y using only counts of permutations of length 4 [40, 60]. See [21] for a combinatorial account of the use of permutation patterns in statistics.

1.1 Main results

The main results of this paper are collected in the following two theorems:

Theorem 1.1. *In* TOOB *the following holds:*

(i) **[Compton,[15]]** *For* $q = 1$ *the* $\mathrm{Mallows}(n, q)$ *distribution satisfies the convergence law but not the zero–one law;*

(ii) *If* $q < 1$ *is fixed then* $\mathrm{Mallows}(n, q)$ *satisfies the zero–one law;*

(iii) *If* $q > 1$ *is fixed then* $\mathrm{Mallows}(n, q)$ *does not satisfy the convergence law.*

(Part **(i)** of Theorem 1.1 was already shown by Compton [15] in 1989, who in fact proved a much stronger statement.)

The function $\log^* n$ equals the number of times we need to iterate the base two logarithm to reach a number below one (starting from $n \in \mathbb{N}$).

Theorem 1.2. *In* TOTO *the following holds:*

(i) *If $q \neq 1$ is fixed, then the convergence law holds for* Mallows(n, q) *but not the zero–one law;*

(ii) *If $q = q(n)$ satisfies $1 - \frac{1}{\log^* n} < q < 1 + \frac{1}{\log^* n}$, then the* Mallows$(n, q)$ *distribution does not satisfy the convergence law.*

Part **(ii)** of Theorem 1.2 extends the previously mentioned result of Foy and Woods [23].

The $\log^* n$ term is a very slowly growing function. As will be clear from the proof of Theorem 1.2 Part **(ii)**, it is certainly not best possible and can be replaced by even more slowly growing functions, such as $\log^{**} n$ defined in Section 3.4, with little effort.

We will in fact show that there exists a single TOTO formula φ such that for all sequences $q = q(n)$ satisfying $1 - \frac{1}{\log^* n} < q < 1 + \frac{1}{\log^* n}$, the quantity $\mathbb{P}(\Pi_n \models \varphi)$ does not have a limit as $n \to \infty$.

1.2 Organization of the paper

To aid the reader, in the next section we give an informal, intuitive synopsis of the proofs of our main results. In Section 3 we collect definitions and results from the literature that we will rely on in our proofs. We also provide some short proofs of relatively elementary observations that we need but are not readily available in the literature. In Section 4 we give the full proof of Theorem 1.1. In Section 5 we prove Part **(i)** of Theorem 1.2. As a preparation for the full proof of Part **(ii)** of Theorem 1.2, we first give a proof in the special case when $q = 1$, in Section 6. In Section 7, we then extend the result to the full range of q stated in Part **(ii)** of Theorem 1.2.

Finally, in Section 8, we offer some additional discussion and some suggestions for further work.

2 Overview of proof methods

Proof Sketch of Theorem 1.1: The proof of Theorem 1.1 will rely on the following observations: Firstly, given any sentence $\varphi \in$ TOOB and $\pi \in S_n$ the vector of cycle counts $(C_1(\pi_n), \ldots, C_n(\pi_n))$ completely determines whether or not $\pi_n \models \varphi$ (where $C_i(\pi_n)$ denotes the number of i–cycles in π_n). Secondly, it follows from the Hanf-Locality Theorem for bounded degree structures (stated as Theorem 3.18 below) that for any fixed $\varphi \in$ TOOB there exists an $h \in \mathbb{N}$ such that for any $t \in \mathbb{N}$ and $n, m \geq h$, the sentence φ cannot distinguish between the disjoint union of n cycles of length t and the disjoint union of m cycles of length t. Moreover, this h can be selected such that φ additionally cannot distinguish between two cycles both of length at least h.

For $\Pi_n \sim$ Mallows(n, q) with $0 < q < 1$, we use a result given in [31] by Jimmy He together with the first and last author of the current paper. They show that there are positive constants $m_1, m_2, \ldots$ depending on q such that $\frac{1}{\sqrt{n}} (C_1(\Pi_n) - m_1 n, \ldots, C_\ell(\Pi_n) - m_\ell n)$ converges in distribution to a multivariate normal with zero mean. This implies in particular that Π_n will contain more than h cycles of each of the lengths $1, \ldots, h$ with probability tending to 1, giving the zero–one law for $0 < q < 1$.

For fixed $q > 1$, another result (stated as Theorem 3.11 below) from [31] implies the existence of a value $x = x(q)$ such that for $\Pi_n \sim$ Mallows(n, q) the quantity $\mathbb{P}(C_1(\Pi_n) \geq x)$ does not have a limit as $n \to \infty$. This property can be queried by a TOOB sentence, establishing that Mallows(n, q) does not satisfy the convergence law w.r.t. TOOB for fixed $q > 1$.

Proof Sketch of Theorem 1.2 Part (i): The proof of Theorem 1.2 Part **(i)** is inspired by the approach taken by Lynch in [43] to show a convergence law for random strings having certain letter distributions. We consider some TOTO sentence φ having quantifier depth d (quantifier depth will be defined in the next section). Writing $\pi \equiv_d \sigma$ if π and σ agree on all TOTO sentences of depth at most d, we note that $\equiv_d$ is an equivalence relation with only finitely many equivalence classes (Theorem 3.17). We now rely on a sampling procedure for generating a sequence $\Pi_1, \Pi_2, \dots$ of Mallows(n, q) distributed permutations from a sequence $Z_1, Z_2, \dots$ of independent Geo$(1-q)$ random variables (The details of the construction will be given in Section 3.2.) Taking a dynamic viewpoint, we can imagine exposing the values $Z_1, Z_2, \dots$ one by one and following the sequence of permutations that arises. Foregoing complicating details, we can define in this manner a Markov chain on a countably infinite state space that follows the equivalence class under $\equiv_d$ of this sequence of permutations. We show that we can the state space have finitely many recurrent classes, each of them irreducible, aperiodic and positive recurrent, and that the chain will hit such a class a.s. Standard results on the convergence of Markov chains will give the convergence of the quantity $\mathbb{P}(\Pi_n \models \varphi)$ as $n \to \infty$.

Proof Sketch of Theorem 1.2 Part (ii): The proof of Theorem 1.2 Part **(ii)** proceeds in several steps. We start by determining a sentence $\varphi \in \mathsf{TOTO}$ such that for $\Pi_n \sim$ Mallows$(n, 1)$ we have

$$\mathbb{P}(\Pi_n \models \varphi) = \begin{cases} 1 - O(n^{-100}) & \text{if } \log^{**}\log^{**} n \text{ is even,} \\ O(n^{-100}) & \text{if } \log^{**}\log^{**} n \text{ is odd.} \end{cases}, \tag{2}$$

for n in a certain set that contains both an infinite number of odd and an infinite number of even values. We do not mention the additional technical conditions on the values of n in this sketch.

We stress that for the $q = 1$ case both Foy and Woods [23] and Spencer [53] have constructed first order sentences whose probability of holding does not converge with n. We have however not found a way to use their proofs "off the shelf" as an ingredient to prove our Theorem 1.2 Part **(ii)**.

We will show (2) by associating pairs of intervals in $\Pi_n \sim$ Mallows$(n, 1)$ to directed graph structures having $N = \Theta(\log\log n)$ vertices. See Figure 6.2 for an example permutation and pair of intervals corresponding to a directed cherry. The directed graph corresponding to the permutation is random, and the number of pairs of intervals is sufficiently high so that with high probability we will be likely to find any directed graph on N vertices. We then adapt the method developed by Spencer and Shelah in [52] of using graphs to model arithmetic on sets to find a sentence about these graphs that oscillates between being true and false depending on N, which will provide the dependency of $\mathbb{P}(\Pi_n \models \varphi)$ on the parity of $\log^{**}\log^{**} n$.

We then show that if $q = 1 \pm O(n^{-4})$ then the total variation distance between Mallows(n, q) and Mallows$(n, 1)$ is $O(n^{-2})$. This allows us to extend the previous result from $q = 1$ to $q = 1 - O(n^{-4})$, replacing the $O(n^{-100})$ terms in (2) by $O(n^{-2})$. We extend the result several more times but now we will have to work a little harder each time.

Next, we consider the range $q = 1 \pm O(1/n)$. We are able to define (in TOTO) a random variable Z_n such that $n^{.1} \le Z_n \le n^{.2}$ with an appropriately high probability. Then, we restrict attention to the permutation induced by Π_n (using the rank function defined in the next section) on $\{1, \dots, Z_n\}$. Rewriting q in terms of Z_n, we have $q = 1 \pm o(|Z_n|^{-4})$, which will allow us to exploit the previous case.

We subsequently extend again, now to the (asymmetric) range $1 - 1/\log^* n < q \le 1 + O(1/n)$. Here we consider the permutation induced on $\{1, \dots, \Pi_n^{-1}(1) - 1\}$. It will turn out that the quantity $N := \Pi_n^{-1}(1) - 1$ is of the order $\Theta(\min(n, 1/|1-q|)$, in an appropriate probabilistic sense. Rewriting q in terms of N tells us $q = 1 \pm O(1/N)$, which will allow us to exploit the previous case.

Finally, we will extend to the full range. Here the key will be that the (TOTO-expressible) property that $\Pi_n(1) < \Pi_n(n)$ is able to distinguish, in an appropriate probabilistic sense, between the cases $q < 1 - \Omega(1/n)$ and $q > 1 + \Omega(1/n)$.

3 Notation and preliminaries

We use the notation $\mathbb{N} = \{1, 2, \ldots\}$ and $\mathbb{Z}_{\geq 0} = \{0, 1, \ldots\}$.

3.1 Requisites from probability theory

We collect some results from probability theory that we will use in the sequel.

Theorem 3.1 (Chernoff's inequality)**.** *Let $X_1, \ldots, X_n$ be independent such that $0 \leq X_i \leq 1$ for $i \in [n]$. Define $S_n \coloneqq X_1 + \ldots + X_n$ and $\mu \coloneqq \mathbb{E}S_n$. Let $0 < \varepsilon < 1$, then*

$$\mathbb{P}\left(S_n \geq (1+\varepsilon)\mu\right) \leq \exp\left\{-\frac{\mu\varepsilon^2}{3}\right\},$$

$$\mathbb{P}\left(S_n \leq (1-\varepsilon)\mu\right) \leq \exp\left\{-\frac{\mu\varepsilon^2}{2}\right\}.$$

We denote by $\mathrm{Bi}(n, p)$ the binomial distribution. We will need the following crude bound:

Lemma 3.2. *Let $X \sim \mathrm{Bi}(n, k/n)$ for $0 \leq k \leq n$. Then $\mathbb{P}\left(X = k\right) \geq 1/(n+1)$.*

Proof. If $k = n$ then the result is clear, so assume that $k \neq n$. Setting $p = k/n$ we have for all $1 \leq j \leq n$ that

$$\frac{\mathbb{P}\left(X = j\right)}{\mathbb{P}\left(X = j-1\right)} = \frac{\binom{n}{j}(1-p)^{n-j}p^j}{\binom{n}{j-1}(1-p)^{n-j+1}p^{j-1}} = \frac{(n-j+1)p}{j(1-p)}.$$

The right–hand side above is less than one if and only if

$$j - jp > (n-j+1)p \quad \iff \quad j > \frac{(n+1)k}{n} \quad \iff \quad j \geq k+1,$$

the last equivalence holding due to j being an integer and $k \neq n$. Thus $\mathbb{P}\left(X = j\right)$ is maximized at $j = k$. As there are only $n+1$ possible outcomes for X we therefore must have $\mathbb{P}\left(X = k\right) \geq 1/(n+1)$. ■

Given two discrete probability distributions μ_1 and μ_2 on a countable set Ω, their total variation distance is defined as

$$d_{\mathrm{TV}}(\mu_1, \mu_2) = \max_{A \subseteq \Omega} |\mu_1(A) - \mu_2(A)|.$$

This can be expressed alternatively as

$$d_{\mathrm{TV}}(\mu_1, \mu_2) = \frac{1}{2}\sum_{x \in \Omega} |\mu_1(x) - \mu_2(x)| = \sum_{x : \mu_1(x) > \mu_2(x)} \mu_1(x) - \mu_2(x). \tag{3}$$

(See for instance Proposition 4.2 in [41] for a proof.) As is common, we will interchangeably use the notation $d_{\mathrm{TV}}(X, Y) \coloneqq d_{\mathrm{TV}}(\mu, \nu)$ if $X \sim \mu$ and $Y \sim \nu$.

A *coupling* of two probability measures μ, ν is a joint probability measure for a pair of random variables (X, Y) satisfying $X \stackrel{d}{=} \mu, Y \stackrel{d}{=} \nu$. We will also speak of a coupling of X, Y as being a probability space for (X', Y') with $X' \stackrel{d}{=} X, Y' \stackrel{d}{=} Y$.

Lemma 3.3. *Let μ and ν be two probability distributions on the same countable set Ω. Then*

$$d_{\mathrm{TV}}(\mu, \nu) = \min\{\mathbb{P}\left(X \neq Y\right) : (X, Y) \textit{ is a coupling of } \mu \textit{ and } \nu\}.$$

(See for instance [41], Proposition 4.7 and Remark 4.8 for a proof.)

For $n \in \mathbb{N}$ and $p \in (0,1)$ we define the TruncGeo(n,p) distribution where $Y \sim$ TruncGeo(n,p) means

$$\mathbb{P}(Y = k) = \frac{p(1-p)^{k-1}}{1-(1-p)^n}, \qquad k \in \{1, \dots, n\}. \tag{4}$$

Observe that, setting $X \sim$ Geo(p), the probability mass in (4) is exactly equal to $\mathbb{P}(X = k \mid X \le n)$. Therefore we shall refer to the TruncGeo(n,p) as the truncated geometric distribution.

3.1.1 Markov chains

We give a brief overview of some of the concepts related to Markov chains which will be used in Section 5. The definitions and results that follow are widely known and can be found in most books on Markov chains or more general stochastic processes, see for instance [41], [28] or [48].

Let X_n be a Markov chain with state space $\mathcal{S}$. Given $i, j \in \mathcal{S}$, if $\mathbb{P}(X_n = j \mid X_0 = i) > 0$ for some $n \in \mathbb{Z}_{\ge 0}$ we say that *j is reachable from i*. We denote this by $i \to j$. If $i \to j$ and $j \to i$ then we say that *the states i and j communicate* and we write $i \leftrightarrow j$. The relation $\leftrightarrow$ is an equivalence relation, the equivalence classes under $\leftrightarrow$ will be called *communicating classes*. For every state $i \in \mathcal{S}$ we have $i \leftrightarrow i$. In the case that the entire state space is a single communicating class we call the Markov chain *irreducible*. The evolution of a Markov chain depends on the initial distribution over the states. We use the notations $\mathbb{P}_i(\cdot)$ and $\mathbb{E}_i(\cdot)$ for probabilities and expectations with respect to the Markov chain started in state $X_0 = i$. If X_0 is distributed according to some distribution μ_0 over $\mathcal{S}$ then we write $\mathbb{P}_{\mu_0}$ and $\mathbb{E}_{\mu_0}$.

For any $i \in \mathcal{S}$ we define the *hitting time* $\tau_i := \inf\{n \ge 0 : X_n = i\}$ to be the first time that the Markov chain is in state i, and we define the *first visit time* $\tau_i^+ := \inf\{n \ge 1 : X_n = i\}$ to be the first time after zero that the Markov chain is in state i. If $\mathbb{P}_i(\tau_i^+ < \infty) = 1$ then we say that the state i is *recurrent*, otherwise we say that it is *transient*. If additionally $\mathbb{E}_i \tau_i^+ < \infty$ then we say that i is *positive recurrent*. If $\mathcal{S}$ is finite then recurrence implies positive recurrence. If i is transient, then all j in the communicating class containing i are transient, and similarly if i is (positive) recurrent. Thus we will talk about positive recurrent classes, etc. If X_n is in a recurrent class for some n, then X_{n+k} is contained in this class for all $k \ge 0$. The *period* of a state is the greatest common divisor of the set $\{n \ge 1 : \mathbb{P}_i(X_n = i) > 0\}$. A state is *aperiodic* if the period equals one. A communicating class or chain is called aperiodic if all of its states are. Proofs for the following two results are given in Theorem 4 and Lemma 5, respectively, in Section 6.3 of [28].

Theorem 3.4. *The state space $\mathcal{S}$ of a Markov chain can be partitioned uniquely as*

$$\mathcal{S} = T \cup C_1 \cup C_2 \cup \dots$$

where T is the set of transient states, and the C_i are recurrent communicating classes.

Lemma 3.5. *If $\mathcal{S}$ is finite, then at least one state is recurrent and all recurrent states are positive recurrent.*

Lemma 3.6. *If $\mathcal{S}$ is finite then with probability 1 the chain X_n will be in a recurrent state for some $n \ge 0$.*

Proof. Let $\mathcal{S} = T \cup C_1 \cup C_2 \cup \dots$ be as in Theorem 3.4. For each of the finitely many $i \in T$ there is a positive probability that the chain moves from i to a state in $\mathcal{S} \setminus T$ in at most $|T|$ steps. Then $\mathbb{P}(X_n \in T) \le Kc^n$ for some constants $K > 0$ and $0 < c < 1$. By the Borel–Cantelli Lemma, X_n is in T for only finitely many n with probability one, so that it must visit a recurrent state. ∎

The following theorem is the Markov chain convergence theorem, see e.g. [19] Theorem 7.6.4 for a proof (note that in [19], what we call irreducibility is called strong irreducibility).

Theorem 3.7 (Markov chain convergence theorem)**.** *Consider an irreducible, aperiodic and positive recurrent Markov chain on a countable state space $\mathcal{S}$. There exists a unique invariant probability measure π over $\mathcal{S}$ such that for every probability measure μ_0 over $\mathcal{S}$:*

$$d_{\mathrm{TV}}(X_n, \pi) \xrightarrow[n\to\infty]{} 0$$

where $X_0 \overset{d}{=} \mu_0$ and $(X_n)_{n\geq 0}$ denotes the evolving sequence of states of the Markov chain.

We will need the following result in Section 5. It is alluded to in many texts on Markov chains when discussing the convergence behavior of a Markov chains that are not irreducible. We have not found an appropriate reference however. So we present a short proof.

Lemma 3.8. *Let X_n be a Markov chain on a countable state space $\mathcal{S}$, and let $\mathcal{M} \subseteq \mathcal{S}$ be an aperiodic and positively recurrent communicating class. Then the limit*

$$\lim_{n\to\infty} \mathbb{P}\left(X_n \in A\right)$$

exists for all $A \subseteq \mathcal{M}$.

Proof. We define $\tau_{\mathcal{M}} := \inf\{n \geq 0 : X_n \in \mathcal{M}\}$. We have

$$\mathbb{P}\left(X_n \in A\right) = \sum_{j=0}^{\infty} \mathbb{P}\left(X_n \in A \,|\, \tau_{\mathcal{M}} = j\right) \mathbb{P}\left(\tau_{\mathcal{M}} = j\right).$$

(We use that $X_n \in A$ can only happen if the chain enters $\mathcal{M}$ at some point.) Conditioning on $\tau_{\mathcal{M}} = j$ induces a probability distribution ν_j over the states in $\mathcal{M}$ where for $B \subseteq \mathcal{M}$ we have $\nu_j(B) = \mathbb{P}\left(X_j \in B \,|\, \tau_{\mathcal{M}} = j\right)$. Consider the Markov chain we get by discarding all states outside of $\mathcal{M}$ and leaving the transition probabilities inside $\mathcal{M}$ as is. This is a valid chain, since states in $\mathcal{M}$ can only reach states inside $\mathcal{M}$, as $\mathcal{M}$ is recurrent. This "reduced chain" is clearly aperiodic, irreducible and positive recurrent. So by Theorem 3.7 there exists a unique distribution π over $\mathcal{M}$ such that for all $j \in \mathbb{Z}_{\geq 0}$

$$\lim_{n\to\infty} \mathbb{P}\left(X_n \in A \,|\, \tau_{\mathcal{M}} = j\right) = \lim_{n\to\infty} \mathbb{P}_{\nu_j}\left(X_{n-j} \in A\right) = \pi(A).$$

Now

$$\left|\mathbb{P}\left(X_n \in A\right) - \pi(A) \cdot \mathbb{P}(\tau_{\mathcal{M}} < \infty)\right| \leq \sum_{j=0}^{\infty} \left|\mathbb{P}\left(X_n \in A \,|\, \tau_{\mathcal{M}} = j\right) - \pi(A)\right| \cdot \mathbb{P}(\tau_{\mathcal{M}} = j) \xrightarrow[n\to\infty]{} 0,$$

where the limit follows by dominated convergence (each term of the summand tends to zero separately, and the sum is bounded by $\sum_j \mathbb{P}(\tau_{\mathcal{M}} = j) = \mathbb{P}(\tau_{\mathcal{M}} < \infty) \leq 1$). ■

3.2 Mallows permutations

It is a standard result in enumerative combinatorics (see Corollary 1.3.13 in [55]) that for $q \neq 1$ the denominator in the definition of the Mallows distribution (1) satisfies

$$\sum_{\sigma \in S_n} q^{\mathrm{inv}(\sigma)} = \prod_{i=1}^{n} \frac{1-q^i}{1-q}.$$

We define for $n \geq 1$ the permutation $r_n : i \mapsto n - i + 1$. For every $\pi \in S_n$ we have

$$\begin{aligned} \mathrm{inv}(r_n \circ \pi) = \mathrm{inv}(\pi \circ r_n) &= \binom{n}{2} - \mathrm{inv}(\pi), \\ \mathrm{inv}(\pi^{-1}) &= \mathrm{inv}(\pi). \end{aligned}$$

From this it can be seen that, if $\Pi_n \sim \text{Mallows}(n,q)$, then

$$\Pi_n^{-1} \overset{d}{=} \Pi_n \quad \text{ and } \quad r_n \circ \Pi_n \overset{d}{=} \Pi_n \circ r_n \overset{d}{=} \text{Mallows}(n, 1/q).$$

For $X = \{x_1, \dots, x_n\}$ a set of distinct numbers and $x \in X$ we define the *rank* $\text{rk}(x, X)$ of x in X as the unique i such that x is the i-th smallest element of X. For $x = (x_1, \dots, x_n)$ a sequence of distinct numbers, let us write

$$\text{rk}(x_1, \dots, x_n) := (\text{rk}(x_1, \{x_1, \dots, x_n\}), \dots, \text{rk}(x_n, \{x_1, \dots, x_n\})) \,. \tag{5}$$

With some abuse of notation we will sometimes write $\pi = \text{rk}(x_1, \dots, x_n)$ to mean that $\pi \in S_n$ is the permutation satisfying $\pi(i) = \text{rk}(x_i, \{x_1, \dots, x_n\})$ for all $i \in [n]$.

Lemma 3.9 ([8], Corollary 2.7)**.** *If $\Pi_n \sim \text{Mallows}(n,q)$ and $1 \leq i < j \leq n$ then*

$$\text{rk}(\Pi_n(i), \Pi_n(i+1), \dots, \Pi_n(j-1), \Pi_n(j)) \overset{d}{=} \text{Mallows}(j-i+1, q).$$

In [26] Gnedin and Olshansky introduced a random bijection $\Sigma : \mathbb{Z} \to \mathbb{Z}$ which in a suitable sense is an extension of the finite $\text{Mallows}(n,q)$ model with $0 < q < 1$. We denote this distribution by $\text{Mallows}(\mathbb{Z}, q)$ where $0 < q < 1$. We will not need the details of the construction and refer the reader to the original paper for a detailed description.

In [31], Jimmy He together with the first and last author of the current paper studied the limit behavior of the cycles counts of $\text{Mallows}(n,q)$ distributed random permutations for fixed $q \neq 1$. Given a permutation $\pi \in S_n$ we define $C_i(\pi)$ to be the number of i–cycles of π for $i \geq 1$. We will use the following results:

Theorem 3.10 ([31], Theorem 1.1)**.** *Fix $0 < q < 1$ and let $\Pi_n \sim \text{Mallows}(n,q)$. There exist positive constants $m_1, m_2, \dots$ and an infinite matrix $P \in \mathbb{R}^{\mathbb{N}\times\mathbb{N}}$ such that for all $\ell \geq 1$ we have*

$$\frac{1}{\sqrt{n}} \left(C_1(\Pi_n) - m_1 n, \dots, C_\ell(\Pi_n) - m_\ell n\right) \quad \xrightarrow[n\to\infty]{d} \quad \mathcal{N}_\ell(\underline{0}, P_\ell),$$

where $\mathcal{N}_\ell(\cdot,\cdot)$ denotes the ℓ–dimensional multivariate normal distribution and P_ℓ is the submatrix of P on the indices $[\ell] \times [\ell]$.

We define the bijections $r, \rho : \mathbb{Z} \to \mathbb{Z}$ by $r(i) = -i$ and $\rho(i) = 1 - i$.

Theorem 3.11 ([31], Theorems 1.3 and 1.7)**.** *Let $q > 1$ and $\Pi_n \sim \text{Mallows}(n,q)$ and $\Sigma \sim \text{Mallows}(\mathbb{Z}, 1/q)$. We have*

$$(C_1(\Pi_{2n+1}), C_3(\Pi_{2n+1}), \dots) \xrightarrow[n\to\infty]{d} \quad (C_1(r\circ\Sigma), C_3(r\circ\Sigma), \dots)$$

and

$$(C_1(\Pi_{2n}), C_3(\Pi_{2n}), \dots) \xrightarrow[n\to\infty]{d} \quad (C_1(\rho\circ\Sigma), C_3(\rho\circ\Sigma), \dots).$$

Moreover, the two limit distributions above are distinct for all $q > 1$. In particular, there exists a value $x = x(q)$ such that $\mathbb{P}(C_1(\rho\circ\Sigma) \geq x) \neq \mathbb{P}(C_1(r\circ\Sigma) \geq x)$.

The following result is due to Bhatnagar and Peled.

Theorem 3.12 ([8], Theorem 1.1)**.** *For all $0 < q < 1$, and integer $1 \leq i \leq n$, if $\Pi_n \sim \text{Mallows}(n,q)$ then*

$$\mathbb{E}|\Pi_n(i) - i| \leq \min\left(\frac{2q}{1-q}, n-1\right).$$

The following method of sampling a Mallows(n,q) distributed permutation for $0 < q < 1$ goes back to the work of Mallows [44]. Let $Z_1, \dots, Z_n$ be independent with $Z_i \sim \text{TruncGeo}(n-i+1, 1-q)$. We now set

$$\begin{aligned} \Pi_n(1) &:= Z_1, \\ \Pi_n(i) &:= \text{the } Z_i\text{-th smallest number in } [n] \setminus \{\Pi_n(1), \dots, \Pi_n(i-1)\}, \qquad \text{for } 1 < i \le n. \end{aligned} \tag{6}$$

Then Π_n follows a Mallows(n,q) distribution. If in this construction we replace the Z_i with independent uniform random variables on the the set $[n-i+1]$, then the above procedure samples a uniformly distributed permutation from S_n.

A natural adaptation of the sampling procedure generates a random bijection $\Pi : \mathbb{N} \to \mathbb{N}$, as follows. We let $Z_1, Z_2, \dots$ be i.i.d. Geo$(1-q)$ and set:

$$\begin{aligned} \Pi(1) &:= Z_1, \\ \Pi(i) &:= \text{the } Z_i\text{-th smallest number in } \mathbb{N} \setminus \{\Pi(1), \dots, \Pi(i-1)\}, \qquad \text{for } i > 1. \end{aligned}$$

The distribution on permutations of $\mathbb{N}$ generated by this procedure will denoted by Mallows$(\mathbb{N}, q)$. It was first introduced by Gnedin and Olshanki in [25].

To avoid confusing the reader we stress that, while the Mallows$(\mathbb{Z}, q)$ distribution that features in Theorem 3.11 above was also introduced by Gnedin and Olshanski, it is very much distinct from the Mallows$(\mathbb{N}, q)$ distribution. (In particular, one is a random permutation of $\mathbb{N}$ and the other a permutation of $\mathbb{Z}$.) At least as far as we are aware, there is no simple iterative procedure for generating $\Sigma \sim$ Mallows$(\mathbb{Z}, q)$ available in the literature.

Remark 3.13. For any finite sequence $i_1, \dots, i_n$ of distinct positive integers we can uniquely associate a sequence $z_1, \dots, z_n$ such that if $\Pi \sim$ Mallows$(\mathbb{N}, q)$ is constructed from an i.i.d. sequence $Z_1, Z_2, \dots$ of Geo$(1-q)$ random variables as described, then $\{\Pi(1) = i_1, \dots, \Pi(n) = i_n\} = \{Z_1 = z_1, \dots, Z_n = z_n\}$.

The following result was obtained by Basu and Bhatnagar ([6], Lemma W) and independently by Crane and DeSalvo ([16], Lemma W):

Lemma 3.14. *For $q < 1$, if $\Pi \sim$ Mallows$(\mathbb{N}, q)$ and $\Pi_n := \text{rk}(\Pi(1), \dots, \Pi(n))$ then $\Pi_n \stackrel{d}{=}$ Mallows(n, q).*

This gives us a natural coupling of $\Pi \sim$ Mallows$(\mathbb{N}, q)$ and a sequence $\Pi_1, \Pi_2, \dots$ with $\Pi_n \sim$ Mallows(n, q).

Given $\Pi \sim$ Mallows$(\mathbb{N}, q)$ we can define a sequence of *regeneration times* $T_0 < T_1 < T_2 < \dots$ as follows:

$$\begin{aligned} T_0 &:= 0, \\ T_i &:= \inf\{j > T_{i-1} \text{ s.t. } \Pi([j]) = [j]\} \qquad (i = 1, 2, \dots),. \end{aligned}$$

We have

Lemma 3.15. *[[31]] $\mathbb{E}T_1^k < \infty$ for every $k \in \mathbb{N}$.*

Although this last result is not spelled out explicitly in [31] – or anywhere else in the literature as far as we are aware – it readily follows by applying Lemma 2.11 in [31] to the (∞, q)-arc chain as defined there. See also the proof of Lemma 4.3 in [31]. Basu and Bhatnagar [6] had previously shown that T_1 has finite second moment.

We also define the *interarrival times*

$$X_i := T_i - T_{i-1},$$

Looking at the sampling procedure generating Π, it is not difficult to see that conditional on the event $T_1 = t$, the bijection $i \mapsto \Pi(i+t) - t$ is distributed like Π. It follows that the interarrival times $X_1, X_2, \dots$ are i.i.d. Moreover, writing $\mathcal{X}_i := \{T_{i-1}+1, \dots, T_i\}$ we see that Π maps $\mathcal{X}_i$ bijectively onto $\mathcal{X}_i$, and in fact the permutations $\tilde{\Pi}_1 : [X_1] \to [X_1], \tilde{\Pi}_2 : [X_2] \to [X_2], \dots$ given by

$$\tilde{\Pi}_i(j) := \Pi(T_{i-1}+j) - T_{i-1} \quad \text{for } j = 1, \dots, X_i,$$

are i.i.d. as well.

For permutations $\pi \in S_n$ and $\sigma \in S_m$ we define $\pi \oplus \sigma$ to be the permutation $\tau \in S_{n+m}$ satisfying

$$\tau(i) = \begin{cases} \pi(i) & \text{if } i \leq n, \\ \sigma(i) + n & \text{if } i > n. \end{cases}$$

With this definition, we can write

$$\Pi = \tilde{\Pi}_1 \oplus \tilde{\Pi}_2 \oplus \dots.$$

For $n \in \mathbb{N}$ let us write:

$$N(n) := \max\{i : T_i \leq n\}.$$

We have the identity

$$\Pi_n = \begin{cases} \tilde{\Pi}_1 \oplus \dots \oplus \tilde{\Pi}_{N(n)} & \text{if } T_{N(n)} = n, \\ \tilde{\Pi}_1 \oplus \dots \oplus \tilde{\Pi}_{N(n)} \oplus \mathrm{rk}(\tilde{\Pi}_{N(n)+1}(1), \dots, \tilde{\Pi}_{N(n)+1}(n - T_{N(n)})) & \text{otherwise.} \end{cases} \tag{7}$$

3.3 First–order logic

Here we briefly cover some of the concepts from logic and model theory that we will need in our proofs. For a more complete descriptions, see for instance Chapter 1 of [35].

Given a set of relation symbols $R_1, \dots, R_k$ with associated arities $a_1, \dots, a_k$, we define $\mathcal{L}_{R_1,\dots,R_k}$ as the first–order language consisting of all first–order formulas built from the usual Boolean connectives together with $R_1, \dots, R_k$. As an example, for a single binary relation R, we may consider the formula $\varphi_0(x) \in \mathcal{L}_R$ defined as

$$\varphi_0(x) := R(x,x) \wedge \neg\exists y : (\neg(x = y) \wedge R(y,y)).$$

A *free variable* of a formula φ is a variable that does not have a quantifier. A formula with no free variables is called a *sentence*. Given $R_1, \dots, R_k$ with arities $a_1, \dots, a_k$, a $(R_1, \dots, R_k)$–structure $\mathfrak{A}$ is a tuple $(A, R_1^{\mathfrak{A}}, \dots, R_k^{\mathfrak{A}})$ where A is a set, called the domain of $\mathfrak{A}$, and each $R_j^{\mathfrak{A}}$ is a relation of arity a_j over A. We will routinely write simply R_j for $R_j^{\mathfrak{A}}$ and talk about structures instead of $(R_1, \dots, R_k)$–structures. If a structure $\mathfrak{A}$ satisfies a sentence φ then we write $\mathfrak{A} \models \varphi$. Similarly, if $i_1, \dots, i_k$ are elements in the domain of $\mathfrak{A}$ and $\varphi(x_1, \dots, x_k)$ is a formula whose free variables are $x_1, \dots, x_k$ then we write $(\mathfrak{A}, i_1, \dots, i_k) \models \varphi$ if φ is satisfied by $\mathfrak{A}$ under the assignment $x_j \mapsto i_j$. We have for instance that $(\mathfrak{A}, i) \models \varphi_0$ precisely when the element x is the unique element satisfying $R(x,x)$ in $\mathfrak{A}$. An *atomic formula* is a formula without quantifiers and Boolean connectives.

Two $(R_1, \dots, R_k)$–structures $\mathfrak{A}, \mathfrak{B}$ with domain A and B respectively are called *isomorphic* if there is a bijection $f : A \to B$ such that

$$(i_1, \dots, i_k) \in R_j^{\mathfrak{A}} \iff (f(i_1), \dots, f(i_k)) \in R_j^{\mathfrak{B}} \qquad \text{for all } i_1, \dots, i_k \in A \text{ and all } R_j.$$

The following is a well–known result on isomorphic structures, see e.g. Theorem W of [51] for a proof.

Lemma 3.16. *If $\mathfrak{A}$ and $\mathfrak{B}$ are isomorphic, then for all first–order sentences φ we have $\mathfrak{A} \models \varphi$ if and only if $\mathfrak{B} \models \varphi$.*

We recursively define the quantifier depth of a first–order sentence φ, denoted $D(\varphi)$, as follows:

- For any atomic formula φ we set $D(\varphi) = 0$;
- For any formula φ we set $D(\neg\varphi) = D(\varphi)$;
- For any two formulas φ, ψ we set $D(\varphi \wedge \psi) = D(\varphi \vee \psi) = \max\{D(\varphi), D(\psi)\}$;
- For a formula $\psi(x)$ with a free variable x we set $D(\exists x : \psi(x)) = D(\forall x : \psi(x)) = D(\psi(x)) + 1$.

We say that two structures $\mathfrak{A}$ and $\mathfrak{B}$ are *d–equivalent* (notation : $\mathfrak{A} \equiv_d \mathfrak{B}$) if $\mathfrak{A}$ and $\mathfrak{B}$ agree on all sentences φ with $D(\varphi) \leq d$. The relation $\equiv_d$ is an equivalence relation. The following lemma will be crucial in the proof of Theorem 1.2 Part **(i)**. A proof can be found for instance following Corollary W in [42].

Lemma 3.17. *The equivalence relation $\equiv_d$ has only finitely many equivalence classes.*

For a structure $\mathfrak{A}$ over domain A we define the *Gaifman Graph of* $\mathfrak{A}$, denoted $G_{\mathfrak{A}}$, as the graph $(A, E^{\mathfrak{A}})$ where $E^{\mathfrak{A}}$ is the collection of all pairs $\{a, b\} \subseteq A$ such that a and b occur together in an element of at least one relation $R^{\mathfrak{A}}$. For an element $x \in A$ we define its r–neighborhood $N(x, r)$ as

$$N(x, r) = \{y \in A : d_{G_{\mathfrak{A}}}(x, y) \leq r\}.$$

We say that a structure $\mathfrak{A}$ has *degree d* if the maximum degree of $G_{\mathfrak{A}}$ is d.

For a structure $\mathfrak{A}$ and $X \subseteq A$ we define $\mathfrak{A} \upharpoonright X$ to be the restriction of $\mathfrak{A}$ to X. That is $\mathfrak{A} \upharpoonright X$ has domain X and for every relation symbol R_i, $R_i^{\mathfrak{A} \upharpoonright X} = R_i^{\mathfrak{A}} \cap X^{a_i}$, where a_i is the arity of R_i.

For $r \geq 0$ and a structure $\mathfrak{A}$, the *r–type* of an element $x \in A$ is the isomorphism class of $\mathfrak{A} \upharpoonright N(x, r)$. Two structures $\mathfrak{A}$ and $\mathfrak{B}$ are said to be *(r, s)–equivalent* if for every r–type, $\mathfrak{A}$ and $\mathfrak{B}$ either have the same number of elements of this type, or they both have more than s elements of this type. We will use the following powerful result in the proof of Theorem 1.1. A proof can for instance be found in [35] (Theorem W).

Theorem 3.18 (Bounded–Degree Hanf Theorem)**.** *Let k and d be fixed. Then there is an integer s such that for all structures $\mathfrak{A}$ and $\mathfrak{B}$ of degree at most k, if $\mathfrak{A}$ and $\mathfrak{B}$ are $(2^d, s)$ equivalent, then $\mathfrak{A} \equiv_d \mathfrak{B}$.*

We also briefly mention *second–order logic*. In second–order logic, formulas can now also take arguments that are relations, and we can quantify over relations. An example of such a formula is

$$\xi(R) \coloneqq \forall x \, (\neg R(x, x) \wedge (\exists! y : R(x, y) \wedge R(y, x))),$$

where we use the shorthand $\exists! x$ to mean that there exists a unique such x, i.e., $\exists! x : \psi(x) \coloneqq \exists x : (\psi(x) \wedge \forall y \, (\psi(y) \to y = x))$. The formula $\xi(R)$ is satisfied by a binary relation R if and only if R encodes a perfect matching. Moreover, in second–order logic we can quantify over relations. So we may write the sentence $\varphi \coloneqq \exists R : \xi(R)$. A finite structure satisfies φ if and only if its domain has even cardinality.

3.3.1 TOOB and TOTO

Given a single binary relation R, we define $\mathsf{TOOB} := \mathcal{L}_R$, the first–order language obtained from the single binary relation R. A permutation π can be encoded as a structure $([n], R^\pi)$ by stipulating that $R^\pi(i,j)$ precisely when $\pi(i) = j$.

We define $\mathsf{TOTO} = \mathcal{L}_{<_1,<_2}$ where $<_1$ and $<_2$ are two binary relations. A permutation π is then encoded as $([n], <_1^\pi, <_2^\pi)$ where $i <_1^\pi j$ if and only $i < j$ in the usual ordering of $[n]$, and $i <_2^\pi j$ if and only if $\pi(i) < \pi(j)$. So $<_1$ and $<_2$ (recall that we often leave out the superscripts) are both total orders on the domain $[n]$.

We remark that an R–structure or a $(<_1, <_2)$–structure does not necessarily correspond to a permutation. In an R-structure $\mathfrak{A}$ the relation $R^{\mathfrak{A}}$ might not correspond to a bijection and similarly in a $(<_1, <_2)$-structure one or both of the relations $<_1^{\mathfrak{A}}, <_2^{\mathfrak{A}}$ might not be a total order on the domain. One way to fix this is to identify sets of axioms that structures in TOOB, respectively TOTO, should satisfy. In our setting we will be sampling from a distribution over the set of permutations and regarding them as either R–structures or $(<_1, <_2)$–structures, so there is no need for such axioms.

We define the TOTO formulas

$$\mathsf{succ}_1(x,y) := (x <_1 y) \wedge \neg\exists w((x <_1 w) \wedge (w <_1 y)),$$
$$\mathsf{succ}_2(x,y) := (x <_2 y) \wedge \neg\exists w((x <_2 w) \wedge (w <_2 y)).$$

These formulas are such that $(\pi, i, j) \models \mathsf{succ}_1(x, y)$ if and only if j is the successor of i under $<_1$, and $(\pi, i, j) \models \mathsf{succ}_2(x, y)$ if and only if j is the successor of i under $<_2$. We also recursively define $\mathsf{succ}_1^{(1)} := \mathsf{succ}_1$ and for $k \geq 1$

$$\mathsf{succ}_1^{(k+1)}(x,y) := \exists w(\mathsf{succ}_1(w,y) \wedge \mathsf{succ}_1^{(k)}(x,w)),$$

and similarly $\mathsf{succ}_2^{(k)}$. These functions are such that for $k \geq 1$

$$(\pi, i, j) \models \mathsf{succ}_1^{(k)}(x,y) \iff i + k = j,$$
$$(\pi, i, j) \models \mathsf{succ}_2^{(k)}(x,y) \iff \pi(i) + k = \pi(j).$$

Recall the definition of $\mathrm{rk}(\underline{x})$ for a sequence $\underline{x} = (x_1, \ldots, x_n)$ of distinct numbers as given in (5). The following lemma is essentially a special case of Theorem 4.2.1 in [33]. For completeness we provide a short proof.

Lemma 3.19. *Let $\varphi(\underline{x}) \in \mathsf{TOTO}$ be a formula with $k \geq 0$ free variables, and let $\psi(y) \in \mathsf{TOTO}$ be a formula with one free variable. There exists a formula $\varphi^\psi(\underline{x}) \in \mathsf{TOTO}$ such that for all n, all $\pi \in S_n$ and all $\underline{i} = (i_1, \ldots, i_k) in [n]^k$ we have*

$$(\pi, \underline{i}) \models \varphi^\psi(\underline{x}) \quad \text{if and only if} \quad \pi \models \psi(i_1) \wedge \cdots \wedge \psi(i_k) \text{ and } (\pi^*, \underline{i}) \models \varphi(\underline{x}),$$

where $\pi^* = \mathrm{rk}(\pi(a_1), \ldots, \pi(a_m))$ with $a_1 < \cdots < a_m$ and $\{a_1, \ldots, a_m\} = \{a \in [n] : (\pi, a) \models \psi(y)\}$.

Proof. The proof is by induction on the structure of the formula φ, starting with the case when it is an atomic formula. Every atomic formula is of the form $R(x_1, x_2)$, for $R \in \{=, <_1, <_2\}$. In this case we define $\varphi^\psi(\underline{x})$ simply as $\psi(x_1) \wedge \psi(x_2) \wedge R(x_1, x_2)$. This clearly does the trick for atomic formulas.

For formulas constructed from smaller formulas using one of the logical connectives, we define $(\neg\varphi)^\psi$ as $\neg(\varphi^\psi)$ and $(\varphi \vee \psi)^\psi$ as $\varphi^\psi \vee \psi^\psi$ and similarly for $\wedge, \rightarrow, \leftrightarrow$. And, if φ starts with a quantifier then we set

$$(\exists x : \varphi(x))^\psi := \exists x : \psi(x) \wedge \varphi^\psi(x), \quad (\forall x : \varphi(x))^\psi := \forall x : \psi(x) \rightarrow \varphi^\psi(x).$$

As the reader can readily check the obtained formula has the behaviour described by the lemma statement. ■

The formula φ^{ψ} provided by Lemma 3.19 is called the *relativization of φ to ψ*.

Lemma 3.20. *Let φ be a* TOTO *formula with $k \geq 0$ free variables. There exists a* TOTO *formula φ^{reverse} such that for all $n \geq 1$, all $\pi \in S_n$ and all $\underline{i} \in [n]^k$ we have*

$$(\pi, \underline{i}) \models \varphi \quad \text{if and only if} \quad (r_n \circ \pi, \underline{i}) \models \varphi^{\text{reverse}}.$$

Proof. As in the proof Lemma 3.19, we use induction on the structure of formulas. For atomic formulas we set

$$(x = y)^{\text{reverse}} := (x = y), \qquad (x <_1 y)^{\text{reverse}} := (x <_1 y), \qquad (x <_2 y)^{\text{reverse}} := (y <_2 x).$$

Now, $\pi(i) < \pi(j)$ if and only if $r_n \circ \pi(i) > r_n \circ \pi(j)$, so that

$$(\pi, i, j) \models x <_2 y \quad \iff \quad (r_n \circ \pi, i, j) \models y <_2 x.$$

The other two relations are straightforward to check and the result now easily follows by induction on the structure of the formula. ∎

3.4 Two rapidly growing functions

The *tower function* is defined as

$$T(n) = 2^{2^{\cdot^{\cdot^{2}}}}$$

where the tower of 2's has height n. It may also be defined recursively by $T(0) = 1$ and $T(i) = 2^{T(i-1)}$ for $i \geq 1$.

Similarly, we define the *wowzer function* $W(n)$ by

$$W(0) = 1, \qquad \text{and} \qquad W(n) = T(W(n-1)) \quad \text{for } n \geq 1.$$

The function $W(\cdot)$ escalates rapidly. We have $W(1) = T(1) = 2$, $W(2) = T(2) = 4$, $W(3) = T(4) = 65536$, and $W(4) = T(65536)$ is a tower of 2's of height 65536.

The previous two functions are part of a larger sequence of functions called the hyperoperation sequence, see e.g. [27]. The tower function is sometimes known under the name tetration and the wowzer function under the name pentation. Knuth also describes these functions in [38] as part of a larger hierarchy in terms of arrow notation.

The *log–star* function $\log^* n$ is the "discrete inverse" of the Tower function. It can be defined by

$$\log^* n := \min\{k \in \mathbb{Z}_{\geq 0} : T(k) \geq n\}.$$

We have $\log^*(T(n)) = n$ for all $n \geq 0$. Also note that, phrased differently, $\log^* n$ is the number of times we need to iterate the base two logarithm, starting from n to reach a number less than 1.

We define the $\log^{**}$ function as

$$\log^{**} n := \min\{k \in \mathbb{Z}_{\geq 0} : W(k) \geq n\}. \tag{8}$$

The $\log^{**} n$ function grows incredibly slowly. Although the wowzer function is part of a larger hierarchy of functions as mentioned, we have not found the function $\log^{**}$ used anywhere in the literature. We emphasize that the notation $\log^{**} n$ is our own and may not be standard.

Although the $\log^{**}$ function is not strictly increasing, we do have the following simple monotonicity principle.
(Here and in the rest of the paper, for f a function and i an integer,

$$f^{(i)} = \underbrace{f \circ \cdots \circ f}_{i\times},$$

denotes the i-fold iteration of f.)

Lemma 3.21. *For all $i \in \mathbb{Z}_{\geq 0}$ and integers x, y, if $W^{(i)}(x) < y$ then $x < (\log^{**})^{(i)}(y)$.*

Proof. We use induction. The base case $i = 0$ holds trivially. So suppose that $W^{(i+1)}(x) = W(W^{(i)}(x)) < y$. From (8) we immediately obtain that $W^{(i)}(x) < \log^{**} y$, and the result follows by induction. ■

Lemma 3.22. *For all $m > 2$ we have*

$$T^{(3)} \circ W^{(2)}(m) < W^{(2)}(m+1).$$

Proof. By the definition of W we have $T^{(3)}(W(W(m))) = W(W(m) + 3)$, so it is enough to check that $W(W(m) + 3) < W(W(m + 1))$. This reduces to checking that

$$W(m) + 3 < W(m+1) = T(W(m)). \tag{9}$$

But $T(x) > x + 3$ for all $x > 2$. As $W(m) > 2$ for $m > 2$, (9) holds. ■

4 Proof of Theorem 1.1

Here prove Theorem 1.1. We are in particular considering TOOB and all mention of (r, k)-equivalence classes and all uses of $\equiv_d$ in this section are thus wrt. TOOB. The bounded–degree Hanf theorem (Theorem 3.18) gives the following basic corollary. For completeness we spell out a proof.

Corollary 4.1. *For every $d \geq 1$ there is a $h = h(d)$ such that the following holds. If π, σ are two permutations such that $C_i(\pi), C_i(\sigma) \geq h$ for $1 \leq i \leq h$ then $\pi \equiv_d \sigma$.*

Proof. The Gaifman graph (wrt. TOOB) of a permutation has maximum degree at most 2. Let s be as provided by the bounded–degree Hanf theorem (with $k = 2$), and set $h := 1000 \cdot \max(2^d, s)$. We point out that all points on a cycle of length $\geq h$ in any permutation have the same 2^d–type. Both π and σ have at least s such points. For each $1 \leq i < h$ both π and σ have at least s points on i-cycles. (And for each i all points on i-cycles have the same 2^d–type.)

The result thus follows from the bounded–degree Hanf theorem. ■

We now have the necessary tools to prove Theorem 1.1. Part **(i)** has already been shown by Compton [15], so we will only supply proofs for the other two parts.

Proof of Theorem 1.1 Part **(ii)**. Fix an arbitrary sentence $\varphi \in$ TOTO, let d be its quantifier depth and let $h = h(d)$ be as supplied by Corollary 4.1. Fix a permutation π satisfying $C_1(\pi) = h$ for $i = 1, \ldots, h$, and let $\Pi_n \sim \text{Mallows}(n, q)$ as usual. By Theorem 3.10 we have

$$\mathbb{P}(C_1(\Pi_n), \ldots, C_h(\Pi_n) \geq h) \xrightarrow[n\to\infty]{} 1.$$

By the previous corollary we thus have that

$$\lim_{n\to\infty} \mathbb{P}\left(\Pi_n \models \varphi\right) = 1_{\{\pi \models \varphi\}}.$$

In particular the limiting probability is $\in \{0, 1\}$, which is what needed to be shown. ■

Proof of Theorem 1.1 Part **(iii)**. Theorem 3.11 tells that for $q > 1$ there exists a value x such that, with $\Pi_n \sim \text{Mallows}(n, q)$ and $\Sigma \sim \text{Mallows}(\mathbb{Z}, 1/q)$, we have

$$\lim_{n\to\infty} \mathbb{P}(C_1(\Pi_{2n+1}) \geq x) = \mathbb{P}(C_1(r \circ \Sigma) \geq x) \neq \mathbb{P}(C_1(\rho \circ \Sigma) \geq x) = \lim_{n\to\infty} \mathbb{P}(C_1(\Pi_{2n}) \geq x).$$

The event that $C_1(\Pi_n) \geq x$ can clearly be queried by a sentence in TOOB. ■

5 Proof of Theorem 1.2 Part (i)

The approach we will take to prove Theorem 1.2 Part **(i)** is inspired by the approach taken by Lynch in [43] where a convergence law for random strings over the alphabet $\{0, 1\}$ is proven for various distributions over the letters. Lynch defines a Markov chain on a finite state space that follows the equivalence class of such a random string as its length increases, and then uses standard convergence results for Markov chains to conclude the argument. The Markov chain that we will define is more complicated to analyze than the one used in [43], mainly because it is defined on a countably infinite state space.

Throughout this section we will consider permutations in the first–order language TOTO, that is, using the two total orders $<_1$ and $<_2$ as described in Section 3.3.1. So $\pi \equiv_d \sigma$ now means that π and σ agree on all TOTO sentences of depth at most d. As per Lemma 3.17, there are finitely many equivalence classes for the equivalence relation $\equiv_d$. Given some fixed d, we denote these classes by $\mathcal{E}_1, \dots, \mathcal{E}_\ell$, where $\ell = \ell(d)$. For a permutation π we let $[\pi]_d$ be the equivalence class containing π. It is convenient to also allow the "permutation of length zero", denoted π_{empty}. In terms of TOTO this is simply the $(<_1, <_2)$-structure with domain $S = \varnothing$. Recall the definition of $\oplus$ from Section 3.2. It is convenient, and makes sense, to set $\pi_{\text{empty}} \oplus \sigma = \sigma \oplus \pi_{\text{empty}} = \sigma$ for every permutation σ.

We denote by $\text{id}_n \in S_n$ the identity on $[n]$. Albert et al. [2] have shown the following useful facts:

Proposition 5.1 ([2], Proposition 26)**.** *Let n, m and d be positive integers with $n, m \geq 2^d - 1$. Then we have* $\text{id}_m \equiv_d \text{id}_n$.

Proposition 5.2 ([2], Proposition 28)**.** *Suppose the permutations $\pi_1, \dots, \pi_n, \sigma_1, \dots, \sigma_n$ satisfy $\sigma_i \equiv_d \pi_i$ for $i = 1, \dots, n$. Then $\sigma_1 \oplus \dots \oplus \sigma_n \equiv_d \pi_1 \oplus \dots \oplus \pi_n$.*

Until further notice we let $0 < q < 1$ be fixed. We couple $\Pi \sim \text{Mallows}(\mathbb{N}, q)$ and $\Pi_1, \Pi_2, \dots$ with $\Pi_i \sim \text{Mallows}(n, q)$ via the coupling described in Section 3.2. That is, $\Pi_n = \text{rk}(\Pi(1), \dots, \Pi(n))$. As usual,we denote by $Z_1, Z_2, \dots$ the sequence of i.i.d. $\text{Geo}(1-q)$–variables that generates Π. We let $T_1, T_2, \dots$ and $X_1, X_2, \dots$ and $\tilde{\Pi}_1, \tilde{\Pi}_2, \dots$ and $N(n)$ be as defined in Section 3.2.

Let us say a vector $(y_1, \dots, y_k) \in \mathbb{N}^k$ is *allowable* if $y_1, \dots, y_k$ are distinct and $\{y_1, \dots, y_j\} \neq [j]$ for all $1 \leq j \leq k$. Let us denote by $\mathcal{A}$ the set of all allowable sequences, of any length. Here we also consider the sequence of length zero allowable. We will denote this "empty sequence" by s_{empty}.

We define a Markov chain $(M_n)_n$ on state space

$$\mathcal{S} := \{\mathcal{E}_1, \dots, \mathcal{E}_\ell\} \times \mathcal{A}.$$

Given that $M_n = ([\pi]_d, (y_1, \dots, y_k))$ for some permutation π and $(y_1, \dots, y_k) \in \mathcal{A}$, we determine M_{n+1} by drawing $Z \sim \text{Geo}(1-q)$ independently of the previous history, letting Y be the Z-th smallest element of $\mathbb{N} \setminus \{y_1, \dots, y_k\}$ and setting

$$M_{n+1} = \begin{cases} ([\pi]_d, (y_1, \dots, y_k, Y)) & \text{if } (y_1, \dots, y_k, Y) \in \mathcal{A}, \\ ([\pi \oplus (y_1, \dots, y_k, Y)]_d, s_{\text{empty}}) & \text{otherwise.} \end{cases}$$

(Here we also allow $k = 0$ in which case $(y_1, \dots, y_k)$ should be interpreted as the empty sequence s_{empty}.)

Having another look at (7) and Lemma 5.2, we see that M_n describes the evolution of

$$\Big([\Pi_{T_{N(n)}}]_d, (\Pi(T_{N(n)} + 1) - T_{N(n)}, \dots, \Pi(n) - T_{N(n)})\Big), \tag{10}$$

where the second argument is interpreted as the empty sequence if $T_{N(n)} = n$ and where as the starting state we take $M_0 = ([\pi_{\text{empty}}]_d, s_{\text{empty}})$.

Recall that for states i, j of some Markov chain, the notation $i \to j$ means that j is reachable from i.

Lemma 5.3. *For every state $(\mathcal{E}, s) \in \mathcal{S}$ there exists an $\equiv_d$–equivalence class $\mathcal{E}'$ such that*

$$(\mathcal{E}, s) \to (\mathcal{E}', s_{\textit{empty}}).$$

Proof. Suppose $M_n = (\mathcal{E}, s)$ with $s = (y_1, \dots, y_k)$ and $k > 0$. In every subsequent step, with probability $1 - q$ we have $Z = 1$, which means that Y is the first element of $\mathbb{N} \setminus \{y_1, \dots, y_k\}$. Therefore, we can reach a state of the form $(\mathcal{E}', s_{\text{empty}})$ in $\max(y_1, \dots, y_k) - k$ steps.

In case when $M_n = (\mathcal{E}, s)$ with $s = s_{\text{empty}}$ then we can reach the state $(\mathcal{E}', s_{\text{empty}})$ in one step where $\mathcal{E}' = [\pi \oplus \text{id}_1]_d$ with $\pi \in \mathcal{E}$ arbitrary. ■

The next lemma contains all the information about the chain M_n we will need.

Lemma 5.4. *The following hold for the Markov chain $(M_n)_n$:*

- **(i)** *Every recurrent communicating class of M_n contains an element of the form $(\mathcal{E}, s_{\textit{empty}})$;*
- **(ii)** *There are only finitely many recurrent communicating classes of M_n;*
- **(iii)** *Every recurrent communicating class of M_n is positive recurrent;*
- **(iv)** *Every recurrent communicating class of M_n is aperiodic;*
- **(v)** *With probability one, the chain M_n is in a positive recurrent state for some $n \geq 1$.*

Proof. For the first two claims, suppose that $(\mathcal{E}, s)$ is contained in some recurrent communicating class. Then Lemma 5.3 implies that this communicating class contains an element of the form $(\mathcal{E}', s_{\text{empty}})$. There are only finitely many such elements, proving the first two claims of the lemma.

For the third claim, consider a recurrent class of M_n. It contains an element α of the form $\alpha = (\mathcal{E}, s_{\text{empty}})$. We define the "reduced chain" $(L_i)_i$ by

$$L_i := M_{T_i},$$

so that L_i takes values in the finite set $\{\mathcal{E}_1, \dots, \mathcal{E}_\ell\} \times \{s_{\text{empty}}\}$. It is easily seen that L is also a Markov chain. The state α is recurrent for L and since L takes only finitely many states, the state α is in fact positively recurrent for L.

We denote by τ_α^+ the first visit time of M_n to α, and by $\tilde{\tau}_\alpha^+$ the first visit time of L_i to α. That α is positively recurrent for L means that $\mathbb{E}_\alpha \tilde{\tau}_\alpha^+ < \infty$. We need to deduce that $\mathbb{E}_\alpha \tilde{\tau}_\alpha^+ < \infty$ as well.

Let $\delta > 0$ be small but fixed, to be determined in the course of the proof. We have

$$\mathbb{P}_\alpha(\tau_\alpha^+ > n) \leq \mathbb{P}_\alpha(\tilde{\tau}_\alpha^+ > \delta n) + \mathbb{P}_\alpha(T_{\lfloor \delta n \rfloor} > n).$$

Note that the second probability in the RHS does not depend on the starting state. For convenience we'll write $m := \lfloor \delta n \rfloor$. Recall that $T_m = X_1 + \cdots + X_m$ where $X_1, \dots, X_m$ are i.i.d. and distributed like T_1. We see that

$$
\begin{aligned}
\mathbb{P}(T_m > n) &= \mathbb{P}(T_m - \mathbb{E}T_m > n - \mathbb{E}T_m) \\
&= \mathbb{P}(T_m - m\mathbb{E}T_1 > n - m\mathbb{E}T_1) \\
&\le \mathbb{P}(T_m - m\mathbb{E}T_1 > n/2) \\
&\le \mathbb{P}\left((T_m - m\mathbb{E}T_1)^4 > n^4/16\right) \\
&\le \frac{16\mathbb{E}(T_m - m\mathbb{E}T_1)^4}{n^4}.
\end{aligned}
$$

where the third line holds since $\mathbb{E}T_1 < \infty$ by Lemma 3.15 and assuming the constant δ is chosen sufficiently small. The last line uses Markov's inequality. Now note that

$$
\begin{aligned}
\mathbb{E}\,(T_m - m\mathbb{E}T_1)^4 &= \mathbb{E}\,((X_1 - \mathbb{E}T_1) + \cdots + (X_m - \mathbb{E}T_1))^4 \\
&= \sum_{i=1}^m \sum_{j=1}^m \sum_{k=1}^m \sum_{\ell=1}^m \mathbb{E}\,((X_i - \mathbb{E}T_1)\cdot(X_j - \mathbb{E}T_1)\cdot(X_k - \mathbb{E}T_1)\cdot(X_\ell - \mathbb{E}T_1)) \\
&= m\mathbb{E}(X_1 - \mathbb{E}T_1)^4 + 3m(m-1)\left(\mathbb{E}(X_1 - \mathbb{E}T_1)^2\right)^2 \\
&= O(n^2),
\end{aligned}
$$

again using that all moments of T_1 are finite (and that the X_i are i.i.d. distributed like T_1). It follows that

$$
\mathbb{P}(T_m > n) = O(n^{-2}).
$$

This gives

$$
\begin{aligned}
\mathbb{E}_\alpha \tau_\alpha^+ &= \textstyle\sum_n \mathbb{P}_\alpha(\tau_\alpha^+ > n) \\
&\le \textstyle\sum_n \mathbb{P}_\alpha(\tilde{\tau}_\alpha^+ > \delta n) + \sum_n O(n^{-2}) \\
&\le \lceil 1/\delta \rceil \cdot \textstyle\sum_k \mathbb{P}_\alpha(\tilde{\tau}_\alpha^+ > k) + O(1) \\
&= \lceil 1/\delta \rceil \cdot \mathbb{E}_\alpha \tilde{\tau}_\alpha^+ + O(1) \\
&< \infty,
\end{aligned}
$$

using a standard formula for the expectation of non-negative integer valued random variables. We've established that α is positively recurrent. Therefore, so are all states in the same communicating class.

For the fourth claim, consider a recurrent class containing $([\pi]_d, s_{\text{empty}})$ for some permutation π. For all $n, t \ge 1$ we have

$$
\mathbb{P}\left(M_{n+t} = ([\pi \oplus \mathrm{id}_t]_d, s_{\text{empty}}) \,|\, M_n = ([\pi]_d, s_{\text{empty}})\right) \ge (1-q)^t > 0.
$$

In particular the recurrent class contains both $([\pi \oplus \mathrm{id}_{2^d-1}]_d, s_{\text{empty}})$ and $([\pi \oplus \mathrm{id}_{2^d}]_d, s_{\text{empty}})$, where M_n may transition from the former to the latter in one step with positive probability. By Propositions 5.1 and 5.2 we have $[\pi \oplus \mathrm{id}_{2^d-1}]_d = [\pi \oplus \mathrm{id}_{2^d}]_d$ so that the class is indeed aperiodic.

For the fifth claim, note the probability that M_n is never equal to an element $(\mathcal{E}, s_{\text{empty}})$ of a recurrent class is exactly the probability that L_i is never equal to a recurrent state $(\mathcal{E}, s_{\text{empty}})$. This probability is zero by Lemma 3.6. ■

Proof of Theorem 1.2 Part **(i)**. Fix some $\varphi \in \mathsf{TOTO}$ with quantifier depth d. Let the recurrent classes of M_n be denoted $\mathcal{M}_1, \ldots, \mathcal{M}_k$. (There are finitely many of them by Lemma 5.4 Part **(ii)**.) Define

$$
\tau_{\mathcal{M}_i} = \inf\{n \ge 0 : M_n \in \mathcal{M}_i\}.
$$

By (7) and Proposition 5.2, the equivalence class under $\equiv_d$ of Π_n can be recovered from M_n. In particular there exist $\mathcal{M}_1^\varphi \subseteq \mathcal{M}_1, \ldots, \mathcal{M}_k^\varphi \subseteq \mathcal{M}_k$ such that, for all n, $\Pi_n \models \varphi$ if $M_n \in \mathcal{M}_i^\varphi$

and $\Pi_n \models \neg\varphi$ if $M_n \in \mathcal{M}_i \setminus \mathcal{M}_i^\varphi$. (Here and in the remainder of the proof we implicitly take M_n as given via (10).) This gives

$$\sum_{i=1}^k \mathbb{P}\left(M_n \in \mathcal{M}_i^\varphi\right) \leq \mathbb{P}\left(\Pi_n \models \varphi\right) \leq \sum_{i=1}^k \mathbb{P}\left(M_n \in \mathcal{M}_i^\varphi\right) + \mathbb{P}(\tau_{\mathcal{M}_1}, \dots, \tau_{\mathcal{M}_k} > n).$$

By Lemma 3.8 each of the terms in the LHS has a limit and by Lemma 5.4 Part **(v)** we have

$$\lim_{n\to\infty} \mathbb{P}(\tau_{\mathcal{M}_1}, \dots, \tau_{\mathcal{M}_k} > n) = 0.$$

This establishes that $\lim_n \mathbb{P}(\Pi_n \models \varphi)$ exists.

To see that this limit is not restricted to being 0 or 1, let φ be the TOTO sentence $\exists x(\neg\exists y(y <_1 x \vee y <_2 x))$ expressing that $\Pi_n(1) = 1$. Then (still considering the case $q < 1$ fixed), we have

$$\mathbb{P}\left(\Pi_n \models \varphi\right) = \mathbb{P}\left(\mathrm{TruncGeo}(n, 1-q) = 1\right) \xrightarrow[n\to\infty]{} 1 - q \notin \{0,1\}.$$

It remains to handle the (fixed) $q > 1$ case. In this case we have for all TOTO sentences φ that

$$\lim_{n\to\infty} \mathbb{P}\left(\Pi_n \models \varphi\right) = \lim_{n\to\infty} \mathbb{P}\left(r_n \circ \Pi_n \models \varphi^{\text{reverse}}\right),$$

where φ^{reverse} is as in Lemma 3.20. As $r_n \circ \Pi_n \overset{d}{=} \mathrm{Mallows}(n, 1/q)$ we can use the result for $q < 1$ to conclude that this limit exists and that this limit is not in $\{0, 1\}$ for the TOTO sentence expressing that $\Pi_n(1) = n$. ■

6 Non–convergence in TOTO for uniformly random permutations

In this section we state and prove a proposition that will be used later on in the proof of Theorem 1.2 Part **(ii)**.

Recall the definition of $W(\cdot)$ and $\log^{**}(\cdot)$ given in Section 3.4. We will prove the following result:

Proposition 6.1. *There exists a $\varphi \in \mathsf{TOTO}$ such that for $\Pi_n \sim \mathrm{Mallows}(n, 1)$ and n satisfying $W^{(2)}(\log^{**}\log^{**} n - 1) < \log\log n$, we have*

$$\mathbb{P}(\Pi_n \models \varphi) = \begin{cases} 1 - O(n^{-100}) & \text{if } \log^{**}\log^{**} n \text{ is even,} \\ O(n^{-100}) & \text{if } \log^{**}\log^{**} n \text{ is odd.} \end{cases}$$

(We remark that by Lemma 3.22 the condition $W^{(2)}(\log^{**}\log^{**} n - 1) < \log\log n$ is for instance satisfied – with plenty of room to spare – if n is of the form $n = W^{(2)}(m)$ with $m > 3$.)

Proposition 6.1 is an explicit version of a result given in [23] by Foy and Woods who showed that there is a TOTO sentence ψ such that $\mathbb{P}\left(\mathrm{Mallows}(n, 1) \models \psi\right)$ does not have a limit. Our plan of attack will follow in broad lines the proof given by Shelah and Spencer in [52] to show non–convergence in first–order logic for graphs in the Erdős–Rényi random graph $G(n, p)$ with p near $n^{-1/7}$. See also Chapter 8 of the excellent monograph [54] by Spencer for a similar argument for p near $n^{-1/3}$.

6.1 Arithmetic on sets in second–order logic

We move away from permutations for the moment, and consider a finite set S equipped with a total order $<$. Our aim will be to determine a second–order sentence Parity such that $S \models \mathsf{Parity}$ if and only if $\log^{**}\log^{**}|S|$ is even.

Given that there is a total order $<$, we can identity S with $[n]$ in the obvious way, where of course $n := |S|$. Note that we can express that $i = 1$ in a (first order) logical formula via $\neg(\exists x : x < i)$. Similarly we can express that $i = n$, and that $i = k$ or $i = n - k$ for any fixed k. In Section 3.3.1 we defined the formulas $\mathsf{succ}_1^{(k)}$ using the $<_1$ relation in TOTO. We will use $\mathsf{succ}^{(j)}$ to denote analogue of the formulas $\mathsf{succ}_1^{(j)}$ with $<_1$ replaced by $<$. For clarity, we point out that $\mathsf{succ}^{(j)}(x, y)$ expresses that $y = x + j$.

Our sentence Parity will begin with the existential quantification over four binary relations R_D, R_E, R_T and R_W. Here the subscript D stands for "double", E for (base two) "exponential", T for "tower", and W for "wowser". We want to demand that these relations are such that:

$$\begin{aligned} R_D(i,j) &\iff j = 2i, \\ R_E(i,j) &\iff j = 2^i, \\ R_T(i,j) &\iff j = T(i), \\ R_W(i,j) &\iff j = W(i), \end{aligned}$$

where $T(i)$ and $W(i)$ denote the tower and wowser functions defined in Section 3.4. For R_D, the following two formulas axiomatize that R_D is as required:

$$\begin{aligned} &\forall\, i,j : (\neg\exists x : (x < i)) \leftrightarrow (R_D(i,j) \leftrightarrow \mathsf{succ}(i,j)) \\ &\forall\, i,j : (\exists x : x < i) \leftrightarrow \left(R_D(i,j) \leftrightarrow \left(\exists i', j' : \mathsf{succ}(i', i) \wedge \mathsf{succ}^{(2)}(j', j) \wedge R_D(i', j')\right)\right). \end{aligned}$$

(Translated into more intuitive/informal terms, the first line says that if i is the first element of S, then $R_D(i, j)$ if and only if $j = i + 1$. The second line says that if i is not the first element of S then $R_D(i, j)$ if and only if $R_D(i - 1, j - 2)$.)

For R_E, we can write:

$$\begin{aligned} &\forall\, i,j : (\neg\exists x : (x < i)) \leftrightarrow (R_E(i,j) \leftrightarrow R_D(i,j)) \\ &\forall\, i,j : (\exists x : x < i) \leftrightarrow (R_E(i,j) \leftrightarrow (\exists i', j' : \mathsf{succ}(i', i) \wedge R_D(j', j) \wedge R_E(i', j'))). \end{aligned} \tag{11}$$

(The first line says that if i is the first element of S then $R_E(i, j)$ if and only if $j = 2i$. The second line says that if i is not the first element then $R_E(i, j)$ if and only if there are i', j' such that $i = i' + 1$ and $j = 2j'$ and $R_E(i', j')$.)

It is easily seen that, to express that the relation R_T is as desired, we can substitute in (11) R_D and R_E by R_E and R_T, respectively. Similarly, to express that R_W is as desired, we can substitute R_D and R_E by R_T and R_W, respectively.

The conjunction of all the axioms we've obtained gives a formula $\mathsf{Arith}(R_D, R_E, R_T, R_W)$ that is satisfied if and only if the relations R_D, R_E, R_T, R_W correctly encode the arithmetic relations.

Given such relations we now want to express that $\log^{**}\log^{**} n$ is even, where $n := |S|$. Suppose that there is an element $x \in S$ for which there exist $y, z \in S$ such that

$$x = W(y) \quad \text{and} \quad y = W(z).$$

Then $\log^{**}\log^{**} x = z$. Now let x be the largest element for which such y and z exist, certainly $x \leq n$. Now, if $x = n$, which we can check using $<$, then $z = \log^{**}\log^{**} n$ and we can query whether or not this is even by $\exists w : R_D(w, z)$. If instead $x < n$, then $\log^{**}(\log^{**} n) = z + 1$, and we can again query whether this is even by $\neg\exists w : R_D(w, z)$. All of this can be formalized in

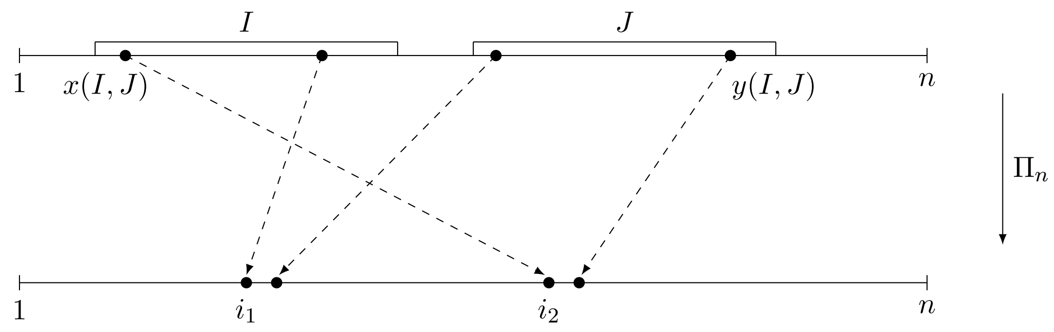


Figure 6.1: An illustration of $S(I,J)$, $x(I,J)$ and $y(I,J)$ in the case that I and J are disjoint intervals. The elements i_1, i_2 are both contained in $S(I,J)$, as $\Pi_n(\{i_1,i_2\})+1 \subseteq \Pi_n(J)$. The element $x(I,J)$ is the smallest element in I such that $\Pi_n(x(I,J)) \in S(I,J)$, and $y(I,J)$ is such that $\Pi_n(y(I,J)) - 1 = \Pi_n(x(I,J))$.

a second–order formula $\mathsf{EvenSize}(R_D, R_E, R_T, R_W)$ in a straightforward manner. So a second–order sentence Parity of the form

$$\mathsf{Parity} := \exists R_D, R_E, R_T, R_W : \mathsf{Arith}(R_D, R_E, R_T, R_W) \wedge \mathsf{EvenSize}(R_D, R_E, R_T, R_W)$$

exists such that $S \models \mathsf{Parity}$ if and only if $\log^{**}\log^{**}|S|$ is even. Here Arith expresses that the relations define the correct arithmetic operations, and $\mathsf{EvenSize}$ uses these relations to express that $\log^{**}\log^{**}|S|$ is even.

In broad lines we will do the following in the next few sections: on a random permutation $\Pi_n \sim \mathrm{Mallows}(n,1)$ we will define directed graphs using TOTO formulas, whose vertices are intervals $\{i, i+1, \dots, i+j\} \subseteq [n]$. These directed graphs will be so abundant that, with probability $1 - o_n(1)$, we can find four such graphs, all defined on the same set S of vertices/intervals of cardinality roughly $|S| \approx \log\log n$, that encode the correct relations R_D, R_E, R_T and R_W on this set. We will then be able to exploit the fact that $\log^{**}(\log^{**}(\log(\log n)))$ oscillates between being even and odd indefinitely, giving the desired non–convergence.

6.2 Defining directed graphs on permutations in TOTO

We now exhibit our definition of directed graphs on Π_n in TOTO. In what follows all directed graphs will be simple, i.e., contain no self–loops or multiple arcs. Recall that we write $i <_1 j$ if and only if $i < j$ and $i <_2 j$ if and only if $\Pi_n(i) < \Pi_n(j)$. Inclusion of i in an interval $\{a, \dots, b\}$ can be checked by $(a = i) \vee (b = i) \vee ((a <_1 i) \wedge (i <_1 b))$.

Let $\Pi_n \sim \mathrm{Mallows}(n,1)$. That is, Π_n is selected uniformly at random from S_n. For disjoint sets $I, J \subseteq [n]$ we define the (random) set

$$\begin{aligned} S(I,J) &:= \{i : i \in \Pi_n(I), i+1 \in \Pi_n(J)\} \\ &= \{\Pi_n(x) : x \in I, \Pi_n(x)+1 \in \Pi_n(J)\}. \end{aligned}$$

For $I, J \subseteq [n]$ with $S(I,J) \neq \varnothing$, we define

$$\begin{aligned} x(I,J) &:= \min\{x : \Pi_n(x) \in S(I,J)\} \\ &= \min\{x \in I : \text{there is a } y \in J \text{ such that } \Pi_n(y) = \Pi_n(x)+1\}, \\ y(I,J) &:= \Pi_n^{-1}(\Pi_n(x(I,J))+1). \end{aligned}$$

Thus $y(I,J)$ is the element of J that gets mapped to $\Pi_n(x(I,J))+1$ by Π_n. If $S(I,J) = \varnothing$, then $x(I,J)$ and $y(I,J)$ are *undefined*. See Figure 6.1 for an illustration of these definitions. Membership of x in I or J can be queried in TOTO if I and J are intervals using the $<_1$

ordering, and $\Pi_n(j) = \Pi_n(i) + 1$ implies simply that j is the successor of i in the $<_2$ ordering. So $x(I, J)$ and $y(I, J)$ can be determined in TOTO when I and J are intervals.

For a sequence $\mathcal{I} = (I_1, \ldots, I_N)$ of subsets of $[n]$ and $J \subseteq [n]$ a single subset, we define the ordered pair $e(\mathcal{I}; J) \in \{I_1, \ldots, I_N\} \times \{I_1, \ldots, I_N\}$ as follows.

$$e(\mathcal{I}; J) = (I_i, I_j) \quad \text{if} \quad \begin{cases} y(I_i, J) = \min\limits_{\ell=1,\ldots,N} y(I_\ell, J), \text{ and} \\ y(I_j, J) = \max\limits_{\ell=1,\ldots,N} y(I_\ell, J). \end{cases}$$

(If $S(I_\ell, J) = \varnothing$ for some $1 \leq \ell \leq N$ then $e(\mathcal{I}; J)$ is undefined.)

For two sequences $\mathcal{I} = (I_1, \ldots, I_N), \mathcal{J} = (J_1, \ldots, J_M)$ of subsets of $[n]$ we now define the directed graph $H(\mathcal{I}; \mathcal{J})$ by setting

$$\begin{aligned} V\left(H(\mathcal{I}; \mathcal{J})\right) &:= \{I_1, \ldots, I_N\}, \\ E\left(H(\mathcal{I}; \mathcal{J})\right) &:= \{e(\mathcal{I}; J_\ell) : \ell = 1, \ldots, M\}. \end{aligned}$$

(If $e(\mathcal{I}; J_\ell)$ is undefined for some J_ℓ, then J_ℓ simply does not contribute to $H(\mathcal{I}; \mathcal{J})$. In particular $H(\mathcal{I}; \mathcal{J})$ is the empty graph on N vertices if all $e(\mathcal{I}; J_\ell)$ are undefined.) Note that it is entirely possible that $e(\mathcal{I}; J_\ell)$ and $e(\mathcal{I}; J_{\ell'})$ code the same arc for some $\ell \neq \ell'$. This will not be an issue.

For $A \subseteq [n]$ we define

$$\begin{aligned} W_k(A) &:= \{i : \{\Pi_n(i), \Pi_n(i) + 1, \ldots, \Pi_n(i) + k - 1\} \subseteq \Pi_n[A]\}, \\ w_k(A) &:= |W_k(A)|. \end{aligned}$$

We can express that i is an element of $W_k(J)$ for an interval J in the following manner: For i, j we can determine whether $\Pi_n(i) = \Pi_n(j)$ by the formula $i = j$, as Π_n is a permutation. For i, j we can determine whether or not $\Pi_n(j) = \Pi_n(i) + 1$ by saying that j is the successor of i under $<_2$. Then $i \in W_k(J)$ for an interval $J = \{a, \ldots, b\}$ if and only if $i \in J$ and there is no c satisfying both $c \notin J$ and $\mathsf{succ}_2^{(j)}(i, c)$ for some $1 \leq j \leq k - 1$, where $\mathsf{succ}_2^{(j)}$ is as defined in Section 3.3.1. This may be expressed in TOTO as

$$\neg \exists c \left(((c <_1 a) \vee (b <_1 c)) \wedge \bigvee_{j=1}^{k-1} \mathsf{succ}_2^{(j)}(i, c) \right), \tag{12}$$

Then we can also express for instance that $w_k(J) \neq 0$.

For $J \subseteq [n]$ an interval with $W_k(J) = \{i_1, \ldots, i_L\}$, where $i_1 < \cdots < i_L$, the sequence $\mathcal{I}_k(J)$ of the *minimal intervals* between points of $W_k(J)$ is defined as:

$$\mathcal{I}_k(J) := (\{i_1 + 1, \ldots, i_2 - 1\}, \ldots, \{i_{L-1} + 1, \ldots, i_L - 1\}).$$

(If $|W_k(J)| \leq 1$ then $\mathcal{I}_k(J)$ is the "empty sequence".) We will sometimes abuse notation and consider $\mathcal{I}_k$ to be a set instead of a sequence.

The induced graph structures that we will consider will be of the form

$$H(\mathcal{I}_k(I); \mathcal{I}_k(J)) \qquad \text{or} \qquad H((\mathcal{I}_k(I_1), \mathcal{I}_k(I_2)); \mathcal{I}_k(J)),$$

where I, I_1, I_2 and J are intervals and $(\mathcal{I}_k(I_1), \mathcal{I}_k(I_2))$ denotes the concatenation of $\mathcal{I}_k(I_1)$ and $\mathcal{I}_k(I_2)$. In Figure 6.2 we give an example such that the induced graph structure is a directed cherry.

As noted, the vertices of these graphs are elements of $\mathcal{I}_k(I)$ and $(\mathcal{I}_k(I_1), \mathcal{I}_k(I_2))$, respectively. Now, these intervals are of the form $\{i_k + 1, \ldots, i_{k+1} - 1\}$ where i_k and i_{k+1} are both elements of $W_k(I')$ for I' equal to one of I, I_1 or I_2. Membership in $W_k(I')$ can be determined via (12). So an interval $\{a, \ldots, b\}$ is an element of for instance $\mathcal{I}_k(I)$ if and only if $a - 1$ and $b + 1$ are two

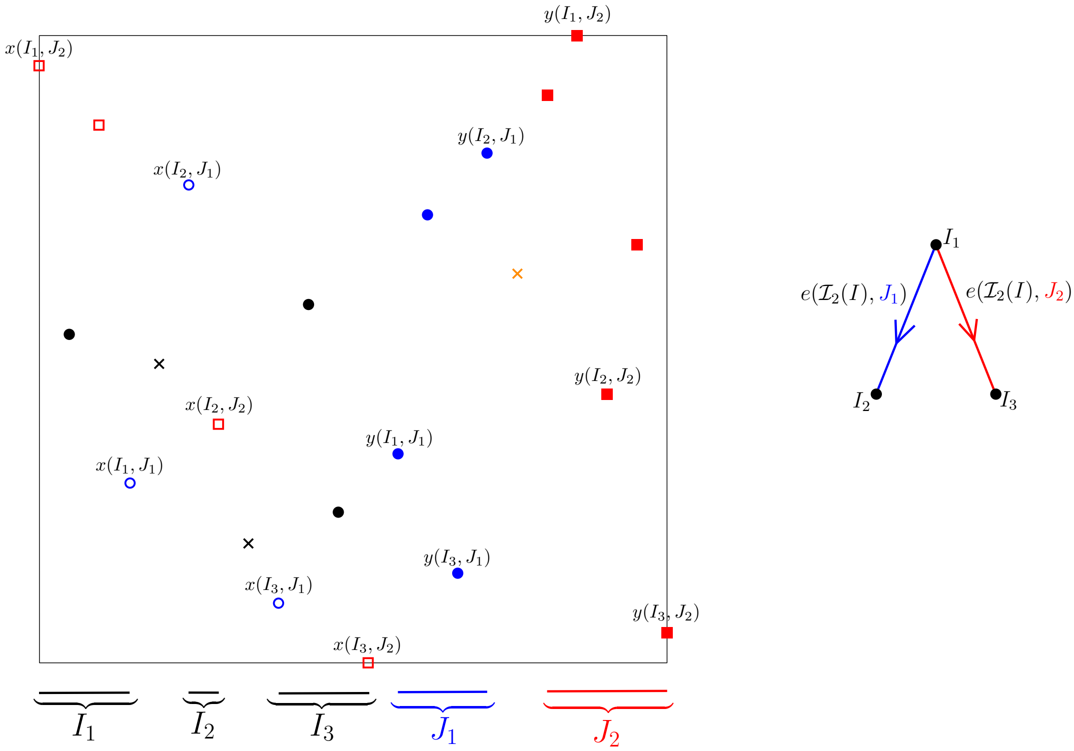


Figure 6.2: In the figure we give an example of a permutation π and intervals I and J such that $H(\mathcal{I}_2(I); \mathcal{I}_2(J))$ is a directed cherry. Take $n = 22$, $I = \{1, \dots, 12\}$, $J = \{13, \dots, 22\}$ and permutation $\pi = 21, 12, 19, 7, 11, 17, 9, 5, 3, 13, 6, 1, 8, 16, 4, 18, 14, 20, 22, 10, 15, 2$. One may check that $\mathcal{I}_2(I) = \{I_1, I_2, I_3\}$ and that $\mathcal{I}_2(J) = \{J_1, J_2\}$ as shown. For each i the point $(i, \pi(i))$ is plotted with a blue dot if $i \in J_1$, a blue circle if $i \in I$ and $\pi(i) + 1 \in \pi(J_1)$, a filled red square if $i \in J_2$, an empty red square if $i \in I$ and $\pi(i) + 1 \in \pi(J_2)$, a black cross if $i \in W_2(I)$ and an orange cross if $i \in W_2(J)$.

consecutive elements of $W_k(I)$. Then if we want to check whether $I' \in \mathcal{I}_k(I)$ and $I'' \in \mathcal{I}_k(I)$ are connected by an arc, we simply need to check whether there is an interval $J' \in \mathcal{I}_k(J)$ for which $y(I', J') = \min_{\ell=1,\dots,|\mathcal{I}_k(I)|} y(I_\ell, J')$ and $y(I'', J') = \max_{\ell=1,\dots,|\mathcal{I}_k(I)|} y(I_\ell, J')$. The elements $y(I_\ell, J')$ can be determined in TOTO as we have already seen, and we can compare $y(I_\ell, J')$ and $y(I_{\ell'}, J')$ by the total ordering $<_1$.

We need one more ingredient in order to describe the final sentence φ whose probability does not converge in Proposition 6.1. Given two disjoint sets A, B we say that a directed graph G with vertex set $A \uplus B$ is a *directed matching from A to B* if every arc of G connects a vertex in A to a vertex in B. Given disjoint intervals I, I' we associate to them the intervals $\mathcal{I}_k(I) = (I_1, \dots, I_N)$, respectively $\mathcal{I}_k(I') = (I'_1, \dots, I'_{N'})$. We will need to determine if $N > N'$. We construct a formula $\mathsf{Bigger}(I, I')$ satisfied with high probability by two intervals I, I' whenever $N' < \log\log n < N < 2\log\log n$. This formula may be written in TOTO as $\exists a, b$ such that, writing $J := (a, b)$, the following holds in $H((\mathcal{I}_k(I), \mathcal{I}_k(I')); \mathcal{I}_k(J))$:

- All vertices in $\mathcal{I}_k(I')$ have outdegree 1;
- All vertices in $\mathcal{I}_k(I)$ have indegree at most 1;
- There is a vertex in $\mathcal{I}_k(I)$ with indegree 0;
- All arcs start at a vertex in $\mathcal{I}_k(I')$ and end at a vertex in $\mathcal{I}_k(I)$.

This directed graph is then necessarily a directed matching from $\mathcal{I}_k(I')$ to $\mathcal{I}_k(I)$ that saturates the vertices in $\mathcal{I}_k(I')$ but not those in $\mathcal{I}_k(I)$. Such a matching exists if and only if $|\mathcal{I}_k(I)| > |\mathcal{I}_k(I')|$.

The following two lemmas represent the "meat" of the proof of Proposition 6.1. They will be proved in Section 6.4.

Lemma 6.2. *For every fixed $k \geq 3$, with probability $1 - e^{-n^{\Omega(1)}}$ there is an interval I with $N := |\mathcal{I}_k(I)|$ satisfying $\log\log n < N < 2\log\log n$ such that for every directed graph G with vertex set $\mathcal{I}_k(I)$ there exists an interval $J \subseteq [n]$ such that*

$$H(\mathcal{I}_k(I); \mathcal{I}_k(J)) = G,$$

and moreover $w_{k+1}(I') = 0$ and $w_{k-1}(I') > 0$ for all $I' \in \mathcal{I}_k(I)$.

Lemma 6.3. *For k sufficiently large but not depending on n, with probability $1 - O(n^{-998})$, for every two disjoint intervals $I_1, I_2 \subseteq [n]$ satisfying*

- $0 < |\mathcal{I}_k(I_1)| < \log\log n < |\mathcal{I}_k(I_2)| < 2\log\log n$*;*
- $w_{k+1}(I_1) = w_{k+1}(I_2) = 0$*;*
- $w_{k-1}(I') > 0$ *for all* $I \in \mathcal{I}_k(I_1) \cup \mathcal{I}_k(I_2)$*,*

there exists an interval $J \subseteq [n]$ such that $H((\mathcal{I}_k(I_1), \mathcal{I}_k(I_2)); \mathcal{I}_2(J))$ is a directed matching between $\mathcal{I}_k(I_1)$ and $\mathcal{I}_k(I_2)$ that saturates $\mathcal{I}_k(I_1)$.

6.3 Proof of Proposition 6.1 assuming Lemmas 6.2 and 6.3

Given an interval I we have seen in the previous section how to define a directed graph on $\mathcal{I}_k(I) = (I_1, \dots, I_N)$ by selecting an interval J. Note that the I_i are disjoint intervals, so we inherit a total ordering on them from the relation $<_1$ on the domain of Π_n. This allows us to distinguish for instance I_1, I_2 and I_N in TOTO.

We will apply the ideas developed in Section 6.1 to sets of the form $\mathcal{I}_k(I)$. Such sets are subsets of the collection of all intervals in $[n]$, and (given I) we can determine membership of an interval $\{a, \dots, b\}$ in $\mathcal{I}_k(I)$ by a TOTO formula as described in the previous section. Any four intervals J_D, J_E, J_T, J_W induce four directed graphs on $\mathcal{I}_k(I)$, and we may check whether these graphs satisfy Arith as given in Section 6.1. If so, then $\mathsf{EvenSize}(J_D, J_E, J_T, J_W)$ holds if and only if $\log^{**}\log^{**}|\mathcal{I}_k(I)|$ is even. Lemma 6.2 will allow us to deduce that with high probability there are J_D, J_E, J_T, J_W satisfying Arith so that we can then check the parity of $\log^{**}\log^{**}|\mathcal{I}_k(I)|$.

We now describe the sentence φ appearing in Proposition 6.1. It will be of the form

$$\varphi = \exists x_1, y_1, \dots, x_5, y_5\, (\varphi_0 \wedge \forall x_6, y_6, \dots, x_{10}, y_{10}(\varphi_1 \to \varphi_2))\,.$$

We use the ten pairs of elements to define intervals $I, J_D, J_E, J_T, J_W, I', J'_D, J'_E, J'_T, J'_W$ by $I = (x_1, y_1), \dots, J'_W = (x_{10}, y_{10})$. We define $\varphi_0(x_1, y_1, \dots, x_5, y_5)$ as the TOTO formula such that $(\Pi_n, i_1, j_1, \dots, i_5, j_5) \models \varphi_0$ if and only if

- $w_{k+1}(I) = 0$ and $w_{k-1}(\tilde{I}) > 0$ for all $\tilde{I} \in \mathcal{I}_k(I)$, and;
- Arith is satisfied wrt. the set $S := \mathcal{I}_k(I)$ and the relations $R_D := \mathcal{H}(\mathcal{I}_k(I); \mathcal{I}_k(J_D)), R_E := \mathcal{H}(\mathcal{I}_k(I); \mathcal{I}_k(J_E)), R_T := \mathcal{H}(\mathcal{I}_k(I); \mathcal{I}_k(J_T)), R_W := \mathcal{H}(\mathcal{I}_k(I); \mathcal{I}_k(J_W))$, and;
- $\mathsf{EvenSize}$ is satisfied wrt. the same set and relations.

We define the formula $\varphi_1(x_6, y_6, \dots, x_{10}, y_{10})$ that holds if and only if the first two conditions above are satisfied wrt $I', J'_D, J'_E, J'_T, J'_W$, but $\mathsf{EvenSize}$ is *not* satisfied wrt. $I', J'_D, J'_E, J'_T, J'_W$. Finally, we let $\varphi_2(x_1, y_1, x_6, y_6)$ express that Bigger holds wrt. $\mathcal{I}_k(I), \mathcal{I}_k(I')$.

To recap, in informal terms the formula φ asks for an interval I such that (some intervals/relations certify that) $\log^{**}\log^{**}|\mathcal{I}_k(I)|$ is even, and moreover there is no interval I' such that (some intervals/relations certify that) both $\log^{**}\log^{**}|\mathcal{I}_k(I')|$ is odd and $|\mathcal{I}_k(I)| < |\mathcal{I}_k(I')|$.

Proof of Proposition 6.1. We pick an n satisfying $W^{(2)}(\log^{**}\log^{**} n - 1) < \log\log n$.
By $W^{(2)}(\log^{**}\log^{**} n - 1) < \log\log n$ and Lemma 3.21 we have

$$\log^{**}\log^{**} n - 1 < \log^{**}\log^{**}\lceil \log\log n\rceil \le \log^{**}\log^{**} n.$$

As the three quantities above are all integers, the second inequality is in fact an equality. Thus, for any integer m satisfying $\log\log n \le m \le n$ we have $\log^{**}\log^{**} m = \log^{**}\log^{**} n$.

We let k be such that the conclusions of Lemmas 6.2 and 6.3 hold.

Consider first the case $\log^{**}\log^{**} n$ even. By Lemma 6.2 and $\log^{**}\log^{**} m$ being even for all $\log\log n \le m \le n$, with probability $1 - e^{-n^{\Omega(1)}}$ there exist intervals I, J_D, J_E, J_T and J_W such that $\varphi_0(I, J_D, J_E, J_T, J_W)$ holds and such that $\log\log n < |\mathcal{I}_k(I)| < 2\log\log n$. Now let I', J'_D, J'_E, J'_T and J'_W be five intervals such that $\varphi_1(I', J'_D, J'_E, J'_T, J'_W)$ holds (if such intervals do not exist then $\Pi_n \models \varphi$ and we are done). Then $N' := |\mathcal{I}_k(I')|$ must be smaller than $\log\log n$, as otherwise $\log^{**}\log^{**} N'$ would be even. Thus indeed $N' < \log\log n < N < 2\log\log n$. By Lemma 6.3, with probability $1 - O(n^{-998})$ there exists an interval J such that $H((\mathcal{I}_k(I), \mathcal{I}_k(I')); J)$ is a directed matching between $\mathcal{I}_k(I)$ and $\mathcal{I}_k(I')$ that saturates the vertices of $\mathcal{I}_k(I')$ but not the vertices of $\mathcal{I}_k(I)$. That is, $\mathsf{Bigger}(I, I')$ holds with at least this probability. So if n is even $\mathbb{P}(\Pi_n \models \varphi) = 1 - O(n^{-998})$.

We now consider the case $\log^{**}\log^{**} n$ odd. Assume that there are intervals I, J_D, J_E, J_T, J_W such that $\varphi_0(I, J_D, J_E, J_T, J_W)$ holds, otherwise we immediately have $\Pi_n \not\models \varphi$. We must then have $|\mathcal{I}_k(I)| < \log\log n$. As before, with probability $1 - e^{-n^{\Omega(1)}}$ there are I', J'_D, J'_E, J'_T and J'_W such that $\varphi_1(I', J'_D, J'_E, J'_T, J'_W)$ holds and $\log\log n < |\mathcal{I}_k(I')| < 2\log\log n$. But then $|\mathcal{I}_k(I')| > |\mathcal{I}_k(I)|$, so $\mathsf{Bigger}(I, I')$ cannot be satisfied. ■

6.4 The probabilistic component

For $B \subseteq [n]$ and $\tau : B \to [n]$ an injection, we will write

$$\mathbb{P}_\tau(.) := \mathbb{P}(. \mid \Pi_n(i) = \tau(i) \text{ for all } i \in B),$$
$$\mathbb{E}_\tau(.) := \mathbb{E}(. \mid \Pi_n(i) = \tau(i) \text{ for all } i \in B).$$

Recall that we defined

$$w_k(A) := |\{i : \{\Pi_n(i), \Pi_n(i)+1, \dots, \Pi_n(i)+k-1\} \subseteq \Pi_n(A)\}|.$$

We define

$$Y_i(A) := \mathbf{1}_{\{\{i,\dots,i+k-1\} \subseteq \Pi_n(A)\}}.$$

We note that

$$w_k(A) = Y_1(A) + \ldots + Y_{n-k+1}(A).$$

Lemma 6.4. *If k is fixed, and $A, B \subseteq [n]$ are disjoint sets of cardinality $1 \ll |A| \ll n$ and $|B| \leq n/\log n$, and $\tau : B \to [n]$ is an arbitrary injection, then*

$$\mathbb{E}_\tau w_k(A) = (1+o(1))n^{1-k}|A|^k,$$

$$\mathbb{E}_\tau w_k(A)^2 \leq (\mathbb{E}_\tau w_k(A))^2 + (2k-1) \cdot \mathbb{E}_\tau w_k(A).$$

(where the $o(.)$ is uniform over all such B and τ.)

Proof. Let us denote by J the set of all $1 \leq j \leq n-k+1$ for which $\{j, \dots, j+k-1\} \cap \mathrm{Im}(\tau) = \varnothing$. For notational convenience, we will write Y_j instead of $Y_j(A)$. We note that

$$\mathbb{E}_\tau Y_j = \begin{cases} \frac{|A|(|A|-1)\dots(|A|-k+1)}{(n-|B|)\dots(n-|B|-k+1)} & \text{if } j \in J, \\ 0 & \text{otherwise.} \end{cases}.$$

Hence

$$\mathbb{E}_\tau w_k(A) = |J| \cdot \frac{|A|(|A|-1)\dots(|A|-k+1)}{(n-|B|)\dots(n-|B|-k+1)} = (1+o(1)) \cdot n^{1-k}|A|^k,$$

using that $|B| \leq n/\log n$ and $(n - k(|B|+1)) \leq |J| \leq n$. The $o(1)$ term above is uniform for all B by the uniform bound $B \leq n/\log n$.

For the second moment, we remark that

$$\mathbb{E}_\tau Y_i Y_j = \begin{cases} \frac{|A|(|A|-1)\dots(|A|-2k+1)}{(n-|B|)\dots(n-|B|-2k+1)} & \text{if } i, j \in J \text{ and } |i-j| > k, \\ 0 & \text{if } i \notin J \text{ or } j \notin J, \\ \leq \mathbb{E}_\tau Y_i & \text{if } |i-j| \leq k. \end{cases}$$

and we also remark that

$$\frac{|A|(|A|-1)\dots(|A|-2k+1)}{(n-|B|)\dots(n-|B|-2k+1)} \leq \left(\frac{|A|(|A|-1)\dots(|A|-k+1)}{(n-|B|)\dots(n-|B|-k+1)}\right)^2,$$

since $1 \ll |A|, |B| \ll n$ so that $(|A|-x)/(n-|B|-x) \leq (|A|-x-1)/(n-|B|-x-1)$ for all $x = 1, \dots, 2k-1$.

We see that

$$\begin{aligned} \mathbb{E}_\tau w_k(A)^2 &= \textstyle\sum_i \sum_j \mathbb{E}_\tau Y_i Y_j \\ &\leq |J|^2 \tfrac{|A|(|A|-1)\dots(|A|-2k+1)}{(n-|B|)\dots(n-|B|-2k+1)} + (2k-1)\textstyle\sum_i \mathbb{E}_\tau Y_i \\ &\leq (\mathbb{E}_\tau w_k(A))^2 + (2k-1) \cdot \mathbb{E}_\tau w_k(A). \end{aligned}$$

■

The following two lemmas establish that if $w_k(I) > 0$ then I is has to be relatively large and if $w_k(I) = 0$ then I has to be relatively small.

Lemma 6.5. *For each fixed $k \geq 2$, with probability $1 - O(n^{-998})$, for every interval $I \subseteq [n]$ of length at most $n^{1-1000/k}$ we have $w_k(I) = 0$.*

Proof. It suffices to show the result for all intervals I of length $\lfloor n^{1-1000/k} \rfloor$ as $w_k(I) = 0$ implies $w_k(I') = 0$ for all $I' \subseteq I$. By Markov's inequality and Lemma 6.4 with $B = \varnothing$ we have for any given interval $I \subseteq [n]$ with $|I| = \lfloor n^{1-1000/k} \rfloor$ that

$$\mathbb{P}\left(w_k(I) > 0\right) \leq \mathbb{E} w_k(I) = (1+o(1))n^{1-k}|I|^k \leq (1+o(1))n^{1-k}n^{k-1000} = O\left(n^{-999}\right).$$

There are $O(n)$ intervals in $[n]$ of length $\lfloor n^{1-1000/k} \rfloor$, so the union bound gives the desired result. ■

Lemma 6.6. *For each fixed k and $0 < \varepsilon < 1/k$, with probability at least $1 - e^{-\Omega(n^{1-k\varepsilon})}$, every interval $I \subseteq [n]$ of length at least $\lceil n^{1-\varepsilon} \rceil$ satisfies $w_k(I) > 0$.*

Proof. Since there are $O(n^2)$ possible intervals, it is enough to show that for any fixed interval I of length $\lceil n^{1-\varepsilon} \rceil$ we have $\mathbb{P}(w_k(I) = 0) \leq e^{-\Omega(n^{1-k\varepsilon})}$.

So let $I \subseteq [n]$ be an interval of length $|I| \geq \lceil n^{1-\varepsilon} \rceil$. Let $J \subseteq [n]$ be a random subset where we independently include i in J with probability

$$\mathbb{P}\left(i \in J\right) = \frac{|I|}{n}, \quad \text{for all } i \in [n].$$

Denote by E the event $\{|J| = |I|\}$. By Lemma 3.2 we have $\mathbb{P}\left(E\right) \geq 1/(n+1)$. Define for $1 \leq j \leq n-k+1$ also the random variable

$$X_j := \mathbf{1}_{\{\{j,\ldots,j+k-1\} \subseteq J\}}.$$

If E holds then $J \overset{d}{=} \Pi_n(I)$. That is, conditional on the event E both I and J are uniformly selected random subsets of $[n]$ of size exactly $\lceil n^{1-\varepsilon} \rceil$. We have

$$\begin{aligned}\mathbb{P}\left(w_k(I) = 0\right) &= \mathbb{P}\left(Y_j(I) = 0 \text{ for all } j \in [n-k+1]\right) = \mathbb{P}\left(X_j = 0 \text{ for all } j \in [n-k+1] \middle| E\right) \\ &\leq \frac{\mathbb{P}\left(X_j = 0 \text{ for all } j \in [n-k+1]\right)}{\mathbb{P}\left(E\right)} \leq (n+1)\,\mathbb{P}\left(X_j = 0 \text{ for all } j \in [n-k+1]\right).\end{aligned}$$

Let $S = \{ki : i \in \mathbb{N}\} \cap [n-k+1]$. The X_j are independent for $j \in S$. Therefore

$$\begin{aligned}\mathbb{P}\left(w_k(I) = 0\right) &\leq (n+1)\mathbb{P}(X_j = 0 \text{ for all } j \in S) = (n+1)\left(1 - \frac{|I|^k}{n^k}\right)^{|S|} \\ &\leq (n+1)\exp\left\{-|S|\frac{|I|^k}{n^k}\right\}.\end{aligned}$$

By $|S| = (1+o(1))n/k$ and $|I|/n \geq n^{-\varepsilon}$ we conclude that

$$\mathbb{P}\left(w_k(I) = 0\right) \leq \exp\left\{-(1+o(1))\frac{n^{1-k\varepsilon}}{k}\right\}.$$

■

Lemma 6.7. *For $k \geq 3$ and $\varepsilon > 0$, let $I \subseteq [n]$ be an interval of length $\lfloor n^{1-1/k} \cdot (\log\log n)^{3/k} \rfloor$, and let $B \subseteq [n]$ be disjoint from I with $|B| \leq n/\log n$ and let $\tau : B \to [n]$ be an arbitrary injection. Let E denote the event that*

(i) $(1-\varepsilon)(\log\log n)^3 < w_k(I) < (1+\varepsilon)(\log\log n)^3$, *and;*

(ii) $w_{k+1}(I) = 0$, *and;*

(iii) $w_{k-1}(I') > 0$ *for each* $I' \in \mathcal{I}_k(I)$, *and;*

(iv) $|I'| > n^{.51}$ *for all* $I' \in \mathcal{I}_k(I)$.

Then $\mathbb{P}_\tau(E) = 1 - o(1)$, *where the* $o(1)$ *term is uniform over the choice of* B *and* τ.

Proof. By Chebyshev's inequality and Lemma 6.4 we have

$$\mathbb{P}_\tau\left(|w_k(I) - \mathbb{E}_\tau w_k(I)| > \frac{\varepsilon}{2}(\log\log n)^3\right) \le \frac{(2k-1)\mathbb{E}_\tau w_k(I)}{\left(\frac{\varepsilon}{2}(\log\log n)^3\right)^2},$$

where

$$\mathbb{E}_\tau w_k(I) = (1+o(1))n^{1-k}|I|^k = (1+o(1))(\log\log n)^3.$$

Thus

$$\mathbb{P}_\tau\left(\left|w_k(I) - (1+o(1))(\log\log n)^3\right| > \frac{\varepsilon}{2}(\log\log n)^3\right) < \frac{8k}{\varepsilon^2(\log\log n)^3} = o(1),$$

showing Part **(i)**.

A simple application of Markov's inequality together with Lemma 6.4 gives

$$\mathbb{P}_\tau\left(w_{k+1}(I) > 0\right) \le \mathbb{E}_\tau w_{k+1}(I) = (1+o(1))n^{-k}\left(n^{1-\frac{1}{k}}(\log\log n)^{\frac{3}{k}}\right)^{k+1} = o(1),$$

giving Part **(ii)**.

We now look at Parts **(iii)** and **(iv)**. We will show that for any fixed $\delta > 1/k$ we have

$$\mathbb{P}_\tau\left(|I'| \ge n^{1-\delta} \text{ for all } I' \in \mathcal{I}_k(I)\right) = 1 - o(1). \tag{13}$$

This will imply Part **(iii)** with an application Lemma 6.6 with e.g. $\delta = \frac{1}{2}(\frac{1}{k} + \frac{1}{k-1})$ satisfying $\frac{1}{k} < \delta < \frac{1}{k-1}$. Part **(iv)** will follow by setting $\delta = .49$.

Note that $w_k(I)$ depends only on the image $\Pi_n[I]$. Conditioning on $\Pi_n[I] = I'$ for some $I' \subseteq [n] \setminus \tau[B]$ (in addition to conditioning on $\Pi_n(i) = \tau(i)$ for all $i \in B$), the restriction $\Pi_n \restriction I$ is distributed uniformly over all bijections $I \to I'$. From this it can be seen that, conditional on the value of $w_k(I)$, $W_k(I)$ is a subset of I of size $w_k(I)$ selected uniformly at random. Let F be the event $\{w_k(I) < 2(\log\log n)^3\}$. It follows that, conditional on F, the probability that $|I'| < n^{1-\delta}$ for some $I' \in \mathcal{I}_k(I)$ is at most the probability that out of $\lfloor 2(\log\log n)^3 \rfloor$ points selected uniformly at random from I, there is a pair i, j with $|i-j| \le n^{1-\delta}$. There are at most $|I| \cdot n^{1-\delta}$ such pairs in I. The probability that two points chosen uniformly at random from I (without replacement) have distance $\le n^{1-\delta}$ is thus at most

$$\frac{|I| \cdot n^{1-\delta}}{\binom{|I|}{2}} = \frac{2n^{1-\delta}}{|I|-1} = n^{1/k-\delta+o(1)}.$$

Let

$$X := \left|\left\{\{i,j\} : i, j \in W_k(I), i \neq j, |i-j| \le n^{1-\delta}\right\}\right|,$$

denote the number of pairs of elements of $W_k(I)$ that are at distance $\le n^{1-\delta}$. It follows that

$$\begin{aligned}
\mathbb{P}_\tau(X > 0|F) &\le \mathbb{E}_\tau(X|F) \\
&\le \tbinom{\lfloor 2(\log\log n)^3 \rfloor}{2} \cdot n^{1/k-\delta+o(1)} \\
&= n^{1/k-\delta+o(1)} \\
&= o(1),
\end{aligned}$$

where the last inequality holds by choice of $\delta > 1/k$. By Part **(i)**, $\mathbb{P}_\tau(F) = 1 - o(1)$, which gives that

$$\mathbb{P}_\tau(X > 0) = \mathbb{P}_\tau(X > 0|F) \cdot \mathbb{P}_\tau(F) + \mathbb{P}_\tau(X > 0|\overline{F}) \cdot \mathbb{P}_\tau(\overline{F}) \leq \mathbb{P}_\tau(X > 0|F) + \mathbb{P}_\tau(\overline{F}) = o(1),$$

proving (13) and hence Parts **(iii)** and **(iv)**. ■

Corollary 6.8. *Let $k \geq 3$ be arbitrary, $0 < \varepsilon < 1/k$ and let $I_1, \ldots, I_N$ be disjoint intervals of length $\lfloor n^{1-1/k} \cdot (\log\log n)^{3/k} \rfloor$ where $N = \lceil n^\varepsilon \rceil$. With probability $1 - e^{-\Omega(n^\varepsilon)}$, at least half of the I_i satisfy Properties* **(i)**-**(iv)** *of Lemma 6.7.*

Proof. Let E_i denote the event that I_i satisfies properties **(i)**-**(iv)** of Lemma 6.7. Let $S = \{s_1, \ldots, s_K\} \subseteq [N]$ be a subset of cardinality $K > N/2$. We can assume without loss of generality that n is sufficiently large for the $o(1)$ term (uniform over the choice of B, τ) in Lemma 6.7 to be $< 1/100$.

We note that for $i = 1, \ldots, K$:

$$\mathbb{P}(\overline{E_{s_i}}|\overline{E_{s_1}} \cap \cdots \cap \overline{E_{s_{i-1}}}) < 1/100.$$

This follows from Lemma 6.7, noting that whether or not the event $\overline{E_{s_1}} \cap \cdots \cap \overline{E_{s_{i-1}}}$ holds is determined completely by the restriction of Π_n to $I_{s_1} \cup \cdots \cup I_{s_{i-1}}$, a set of cardinality at most

$$|I_1| + \cdots + |I_N| = N \cdot \lfloor n^{1-1/k} \cdot (\log\log n)^{3/k} \rfloor = n^{1+\varepsilon-1/k+o(1)} \leq \frac{n}{\log n},$$

where the last inequality holds for sufficiently large n (and by choice of ε). It follows that

$$\mathbb{P}(\overline{E_{s_1}} \cap \cdots \cap \overline{E_{s_K}}) = \mathbb{P}(\overline{E_{s_1}}) \cdot \mathbb{P}(\overline{E_{s_2}}|\overline{E_{s_1}}) \cdot \cdots \cdot \mathbb{P}(\overline{E_{s_K}}|\overline{E_{s_1}} \cap \cdots \cap \overline{E_{s_{K-1}}}) \leq (1/3)^{N/2},$$

for each set $S \subseteq [N]$ of cardinality $> N/2$. Hence, if F denotes the event that at least $N/2$ of the events $E_1, \ldots, E_N$ hold then

$$\mathbb{P}(\overline{F}) \leq 2^N \cdot (1/100)^{N/2} = (1/5)^N = \exp[-\Omega(n^\varepsilon)].$$

(Here 2^N is a crude upper bound on the number of subsets of $[N]$ of cardinality $> N/2$.) ■

Lemma 6.9. *Let I be an interval satisfying Conditions* **(ii)**-**(iv)** *of Lemma 6.7. Let J be obtained from I by deleting an arbitrary element. Then J satisfies Conditions* **(ii)**-**(iv)** *of Lemma 6.7 and*

$$|\mathcal{I}_k(I)| - k \leq |\mathcal{I}_k(J)| \leq |\mathcal{I}_k(I)|. \tag{14}$$

Proof. As $J \subseteq I$ we have $W_t(J) \subseteq W_t(I)$ for all $t \geq 1$, by definition of $W_t(.)$. IN particular $w_{k+1}(I) = 0$ implies $w_{k+1}(J) = 0$. In other words, Part **(iv)** holds for J if it holds for I. Similarly, any interval of $J' \in \mathcal{I}_k(J)$ must contain some interval $I' \in \mathcal{I}_k(I)$, showing that Part **(iv)** holds for J if it holds for I.

To see that (14), suppose $J := I \setminus \{j\}$ for some j. Then there are at most k elements $i \in I$ such that $\Pi_n(j) \in \{\Pi_n(i), \ldots, \Pi_n(i) + k - 1\}$. Thus, by definition of $W_k(.)$, we have $|W_k(I) \setminus W_k(J)| \leq k$. ■

Recall that we defined

$$S(I, J) = \{i : i \in \Pi_n(I), i + 1 \in \Pi_n(J)\}.$$

Lemma 6.10. *Let $\varepsilon, \delta > 0$ be arbitrary (but fixed). With probability at least $1 - e^{-n^{\Omega(1)}}$, for every two disjoint intervals $I_1, I_2 \subseteq [n]$ of length at least $n^{\frac{1}{2}+\varepsilon}$, we have*

$$(1-\delta) \cdot \frac{|I_1| \cdot |I_2|}{n} \leq |S(I_1, I_2)| \leq (1+\delta) \cdot \frac{|I_1| \cdot |I_2|}{n}.$$

Proof. Fix disjoint intervals I_1, I_2 of the stated sizes. Note $J_1 := \Pi_n(I_1), J_2 := \Pi_n(I_2)$ are disjoint, random sets of cardinalities exactly $k_1 := |I_1|$, resp. $k_2 := |I_2|$, every pair of disjoint sets of the correct cardinalities being equally likely.

We approximate these by sets with a random number of elements, generated as follows. Write $p_i := k_i/n$ for $i = 1, 2$. Let $Z_1, \ldots, Z_n$ be i.i.d. with common distribution given by

$$\mathbb{P}(Z_j = x) = \begin{cases} 1 - (p_1 + p_2) & \text{if } x = 0, \\ p_1 & \text{if } x = 1, \\ p_2 & \text{if } x = 2. \end{cases}$$

We now set $J_1' := \{j : Z_j = 1\}, J_2' := \{j : Z_j = 2\}$. Conditional on the event

$$E := \{|J_1'| = k_1, |J_2'| = k_2\},$$

the pair (J_1', J_2') has the same distribution as (J_1, J_2). Namely, the uniform distribution over all pairs of disjoint sets of the correct cardinalities k_1, k_2. Lemma 3.2 gives the crude bound

$$\begin{aligned}
\mathbb{P}(E) &= \mathbb{P}\left(|J_1'| = k_1\right) \mathbb{P}\left(|J_2'| = k_2 \middle| |J_1'| = k_2\right) \\
&= \mathbb{P}(\mathrm{Bi}(n, p_1) = k_1) \cdot \mathbb{P}\left(\mathrm{Bi}\left(n - k_1, p_2/(1-p_1)\right) = k_2\right) \\
&= \mathbb{P}\left(\mathrm{Bi}\left(n, k_1/n\right) = k_1\right) \cdot \mathbb{P}\left(\mathrm{Bi}\left(n - k_1, k_2/(n-k_1)\right) = k_2\right) \\
&\geq \frac{1}{(n+1)^2}.
\end{aligned} \tag{15}$$

Next, we consider the size of

$$S' := \{j : j \in J_1', j+1 \in J_2'\}.$$

Write $X_j := \mathbf{1}_{\{j \in S'\}}$, so that $|S'| = X_1 + \cdots + X_{n-1}$. We have $\mathbb{E}X_j = p_1 p_2 =: p$ and

$$\mathbb{E}|S'| = (n-1)p = \frac{|I_1| \cdot |I_2|}{n} - p. \tag{16}$$

In particular $\mathbb{E}|S'| = (1 + o(1))n^{2\varepsilon}$ by our assumption on the sizes $|I_1|, |I_2| \geq n^{1/2+\varepsilon}$. Now $|S'|$ is not a binomial random variable as the X_j are (mildly) dependent – it is for instance not possible to have two consecutive integers in S'. Note that X_j is however (mutually) independent from the collection of all X_i with $|i - j| \geq 2$. So, writing

$$X_{\text{even}} := \sum_{j \text{ even}} X_j, \quad X_{\text{odd}} := \sum_{j \text{ odd}} X_j,$$

we see that $|S'| = X_{\text{even}} + X_{\text{odd}}$ and $X_{\text{even}} \stackrel{d}{=} \mathrm{Bi}(\lfloor (n-1)/2 \rfloor, p), X_{\text{odd}} \stackrel{d}{=} \mathrm{Bi}(\lceil (n-1)/2 \rceil, p)$. By the Chernoff bound we have

$$\mathbb{P}\left(|X_{\text{even}} - \mathbb{E}X_{\text{even}}| \geq \frac{\delta}{100} \cdot \mathbb{E}X_{\text{even}}\right) \leq e^{-\Omega(\mathbb{E}X_{\text{even}})} = e^{-\Omega(np)} = e^{-n^{\Omega(1)}},$$

and analogously $\mathbb{P}(|X_{\text{odd}} - \mathbb{E}X_{\text{odd}}| \geq \frac{\delta}{100} \cdot \mathbb{E}X_{\text{odd}}) = e^{-n^{\Omega(1)}}$. It follows that

$$\mathbb{P}\left(\left||S'| - \mathbb{E}|S'|\right| \geq \frac{\delta}{100}\mathbb{E}|S'|\right) \leq e^{-n^{\Omega(1)}}. \tag{17}$$

We see that

$$\begin{aligned}
\mathbb{P}\left(\left|S - \tfrac{|I_1|\cdot|I_2|}{n}\right| > \delta\tfrac{|I_1|\cdot|I_2|}{n}\right) &\leq \mathbb{P}\left(|S - \mathbb{E}S'| > \tfrac{\delta}{100}\mathbb{E}|S'|\right) \\
&= \mathbb{P}\left(|S' - \mathbb{E}S'| > \tfrac{\delta}{100}\mathbb{E}|S'|\middle|E\right) \\
&\leq \mathbb{P}\left(|S' - \mathbb{E}S'| > \tfrac{\delta}{100}\mathbb{E}|S'|\right)/\mathbb{P}(E) \qquad (18)\\
&\leq (n+1)^2 \cdot e^{-n^{\Omega(1)}} \\
&= e^{-n^{\Omega(1)}}.
\end{aligned}$$

(Here the first inequality holds for n sufficiently large and uses (16). In the second line we use the earlier observation that $(S'|E)\stackrel{d}{=}S$ and in the fifth line we use (15) and (17).)

It is easily seen that the bound (18) holds uniformly over the choice of all pairs of disjoint intervals $I_1, I_2 \subseteq [n]$ of sizes $|I_1|, |I_2| \geq n^{1/2+\varepsilon}$. (Looking more carefully at the Chernoff bound as stated in Theorem 3.1, the error bound $e^{-n^{\Omega(1)}}$ is bounded above by an expression of the form $\exp[-c \cdot n^{2\varepsilon}]$, where $c = c(\delta)$ depends only on δ.) Since there are $O(n^4)$ choices of such pairs of intervals, the lemma follows by the union bound. ■

Lemma 6.11. *For every fixed $\varepsilon > 0$ and $k \geq 3$, with probability $1 - e^{-n^{\Omega(1)}}$, for every sequence $\mathcal{I} = (I_1, \ldots, I_N)$ of at most $10 \log\log n$-many disjoint intervals satisfying*

$$n^{\frac{1}{2}+\varepsilon} \leq |I_1|, \ldots, |I_N| \leq n^{1-\varepsilon},$$

and every directed graph G with $V(G) = \{I_1, \ldots, I_N\}$, there exists an interval $J \subseteq [n]$ such that

$$H(\mathcal{I}; \mathcal{I}_k(J)) = G.$$

Proof. There are no more than $10 \log\log n \cdot n^{20\log\log n} = e^{O(\log n \cdot \log\log n)}$ many ways to pick $I_1, \ldots, I_N$, and at most $2^{O((\log\log n)^2)}$ corresponding directed graphs G for each such a sequence. We will show that for each particular choice of $I_1, \ldots, I_N$ and G with $v(G) = N$, the probability that no J exists as desired is at most $e^{-n^{\Omega(1)}}$. This will prove the lemma.

Let us thus fix $I_1, \ldots, I_N$, meeting the conditions of the lemma but otherwise arbitrary, and an arbitrary directed graph G on N vertices. Note that if G is the "empty directed graph" ($e(G) = 0$) then we can simply take $J = \{1\}$. Thus, we can and do assume from now on that G has at least one arc.

Inside the set $[n] \setminus (I_1 \cup \cdots \cup I_N)$ we can find $M = n^{\Omega(1)}$ disjoint intervals $J_1, \ldots, J_M$ each of size $s := \lceil n^{1-1/k} \cdot (\log\log n)^{3/k} \rceil$.

For $\tau : I_1 \cup \cdots \cup I_N \to [n]$ an injection and $L_1, \ldots, L_M \geq 0$ and $A_{11}, \ldots, A_{1L_1} \subseteq J_1, \ldots, A_{M1}, \ldots, A_{ML_M} \subseteq J_M$ disjoint intervals, and $B_{11}, \ldots, B_{1L_1}, \ldots, B_{ML_M} \subseteq [n] \setminus \tau[I_1 \cup \cdots \cup I_N]$ disjoint sets with $|A_{ij}| = |B_{ij}|$ for all i, j, we define the event

$$F_{\tau,\underline{A},\underline{B}} := \left\{ \begin{array}{l} \mathcal{I}_k(J_\ell) = (A_{\ell 1}, \ldots, A_{\ell L_\ell}) \text{ for all } \ell = 1, \ldots, M, \\ \Pi_n[A_{\ell\ell'}] = B_{\ell\ell'} \text{ for all } \ell = 1, \ldots, M \text{ and } \ell' = 1, \ldots, L_\ell, \\ \Pi_n(x) = \tau(x) \text{ for all } x \in I_1 \cup \cdots \cup I_n. \end{array} \right\}$$

We observe that, conditional on $F_{\tau,\underline{A},\underline{B}}$, every bijection $A_{ij} \to B_{ij}$ is equally likely. That is, for every choice of bijections $\tau_{11} : A_{11} \to B_{11}, \ldots, \tau_{ML_M} : A_{ML_M} \to B_{ML_M}$ we have

$$\mathbb{P}\left(\Pi_n \restriction A_{ij} \equiv \tau_{ij} \text{ for all } i, j \middle| F_{\tau,\underline{A},\underline{B}}\right) = \frac{1}{|A_{11}|! \cdot \cdots \cdot |A_{ML_M}|!}.$$

Let us say an index $1 \leq \ell \leq M$ is *good* if $(1-\varepsilon)(\log\log n)^3 \leq L_\ell \leq (1+\varepsilon)(\log\log n)^3$ and $|S(I_i, J')| > 0$ for all $1 \leq i \leq N$ and $J' \in \mathcal{I}_k(J_\ell)$. (So in particular $H(\mathcal{I}, \mathcal{I}_k(J_\ell))$ is defined when ℓ is good.) Note also that whether ℓ is good or not is determined completely by $F_{\tau,\underline{A},\underline{B}}$, because the sets $S(I_i, A_{\ell j})$ depend only on $B_{\ell j}$, not on the choice of the specific bijection $A_{\ell j} \to B_{\ell j}$.

Note that if ℓ is good and we randomly choose bijections $\tau_{i\ell} : A_{i\ell} \to B_{i\ell}$ then the points $y(I_1, A_{i\ell}), \ldots, y(I_N, A_{i\ell})$ are chosen uniformly without replacement from $A_{i\ell}$. In particular, defining $P := \{(I_i, I_j) : i \neq j\}$, for each ordered pair $uv \in P$ the probability that $e(\mathcal{I}, A_{i\ell}) = uv$ is $\frac{1}{|P|} = \frac{1}{N(N-1)}$.

We fix a sequence $e_1, e_2, \ldots, e_{\lceil (1+\varepsilon)(\log\log n)^3 \rceil} \in P$ with the property that

$$E(G) = \{e_1, \ldots, e_{\lceil (1-\varepsilon)(\log\log n)^3 \rceil}\} = \{e_1, \ldots, e_{\lceil (1+\varepsilon)(\log\log n)^3 \rceil}\}.$$

(Since $(1-\varepsilon)(\log\log n)^3 > N^2$, there are necessarily some duplicate arcs. This will not be a problem.) We note that $H(\mathcal{I}, (A_{\ell 1}, \ldots, A_{\ell L_\ell})) = G$ is implied by the event $\{e(\mathcal{I}, A_{\ell i}) = e_i$ for all $1 \leq i \leq L_i\}$.

For notational convenience, let us write

$$E := \{H(\mathcal{I}, \mathcal{I}_k(J_\ell)) = G \text{ for some } 1 \leq \ell \leq M\ \}.$$

It follows from the previous observations that if Z denotes the number of good indices, then the value of Z is completely determined by the event $F_{\tau,\underline{A},\underline{B}}$, and

$$\begin{aligned}
\mathbb{P}\left(E | F_{\tau,\underline{A},\underline{B}}\right) &\geq 1 - \left(1 - \left(\frac{1}{N^2}\right)^{(1+\varepsilon)(\log\log n)^3}\right)^Z \\
&\geq 1 - \exp\left[-Z \cdot \left(\frac{1}{N^2}\right)^{(1+\varepsilon)(\log\log n)^3}\right] \\
&= 1 - \exp\left[-Z \cdot n^{o(1)}\right].
\end{aligned}$$

By Corollary 6.8 and Lemma 6.10 we have that

$$\mathbb{P}(Z < M/2) = e^{-n^{\Omega(1)}}.$$

It follows that

$$\begin{aligned}
\mathbb{P}(E) &\geq \sum_{\tau,\underline{A},\underline{B}} \mathbb{P}\left(E | F_{\tau,\underline{A},\underline{B}}\right) \cdot \mathbb{P}(F_{\tau,\underline{A},\underline{B}}) \\
&\geq 1 - \exp[-(M/2) \cdot n^{o(1)}] - \mathbb{P}(Z < M/2) \\
&= 1 - e^{-n^{\Omega(1)}}.
\end{aligned}$$

This proves the statement of the lemma. ■

Proof of Lemma 6.2. Let $I_1, \ldots, I_N$ be $\lceil n^{1/2k} \rceil$ intervals of length $\lfloor n^{1-1/k} \cdot (\log\log n)^{3/k} \rfloor$. By Corollary 6.8, with probability $1 - e^{-\Omega(n^{1/2k})}$ at least 1 of them satisfies Conditions **(i)**-**(iv)** of Lemma 6.7, by relabeling if necessary, we can assume that I_1 satisfies these conditions. We have $\mathcal{I}_k(I_1) \gg \log\log n$ by Condition **(i)**. By Lemma 6.9 we may shrink I_1 by iteratively removing an element until $\log\log n < |\mathcal{I}_k(I_1)| < 2\log\log n$, ensuring that I_1 still satisfies Conditions **(ii)**-**(iv)** of Lemma 6.7. In particular $|I'| > n^{.51}$ for all $I' \in \mathcal{I}_k(I_1)$, and by $|I_1| \leq n^{1-1/2k}$ we also have $I' \leq n^{1-1/2k}$ for all $I' \in \mathcal{I}_k(I_1)$.

By Lemma 6.11, for all directed graphs G with $V(G) = \mathcal{I}_k(I_1)$ there is an interval J satisfying $H(\mathcal{I}_k(I_1), \mathcal{I}_k(J)) = G$, with probability $1 - e^{-n^{\Omega(1)}}$. ■

Proof of Lemma 6.3. Let I_1, I_2 be any two disjoint intervals such that $0 < |\mathcal{I}_k(I_1)| < \log\log n < |\mathcal{I}_k(I_2)| < 2\log\log n$, and such that $w_{k+1}(I_1) = w_{k+1}(I_2) = 0$ and $w_{k-1}(I') > 0$ for all $I \in \mathcal{I}_k(I_1) \cup \mathcal{I}_k(I_2)$. Each of these $I' \in \mathcal{I}_k(I_1) \cup \mathcal{I}_k(I_2)$ has length at most $\max\{|I_1|, |I_2|\}$. By Lemma 6.6 with $\varepsilon = \frac{1}{2(k+1)}$

$$\mathbb{P}\left(w_{k+1}(I') > 0 \text{ for all intervals } I' \text{ s.t. } |I'| > n^{1-\frac{1}{2(k+1)}}\right) \geq 1 - e^{-\Omega(n^{1/2})}.$$

So with at least this probability, all I_1 and I_2 as described satisfy $|I_1|, |I_2| \leq n^{1-\frac{1}{2(k+1)}}$. By $w_{k-1}(I') > 0$ for all $I' \in \mathcal{I}_k(I_1) \cup \mathcal{I}_k(I_2)$ and Lemma 6.5 we have $|I'| \geq n^{1-1000/(k-1)}$ for all such I_1, I_2 and $I' \in \mathcal{I}_k(I_1) \cup \mathcal{I}_k(I_2)$ with probability $1 - O(n^{-998})$. So for sufficiently large k we have

$$n^{\frac{1}{2}+\frac{1}{2(k+1)}} < n^{1-1000/(k-1)} \leq |I'| < n^{1-\frac{1}{2(k+1)}}.$$

The result then follows by $|\mathcal{I}_k(I_1) \cup \mathcal{I}_k(I_2)| \leq 3 \log\log n$ and Lemma 6.11. ∎

7 Proof of Theorem 1.2 Part (ii)

Here we extend the proof of the case when $q = 1$ to $|1 - q| < 1/\log^* n$. We do so step by step, first extending to $|1 - q| \leq n^{-4}$ then extending to $|1 - q| = O(1/n)$ and finally to the entire range.

7.1 Extending the $q = 1$ case to $q = 1 \pm n^{-4}$

We first extend Proposition 6.1 to the following:

Proposition 7.1. *There exists a $\varphi \in \mathsf{TOTO}$ such that for all sequences $q = q(n)$ satisfying $1 - n^{-4} < q < 1 + n^{-4}$ and all sufficiently large n satisfying $W^{(2)}(\log^{**}\log^{**} n - 1) < \log\log n$ we have*

$$\mathbb{P}(\Pi_n \models \varphi) = \begin{cases} 1 - O(n^{-2}) & \text{if } \log^{**}\log^{**} n \text{ is even,} \\ O(n^{-2}) & \text{if } \log^{**}\log^{**} n \text{ is odd.} \end{cases},$$

where $O(n^{-2})$ terms can be taken uniform over all sequences $q(n)$ considered.

(Again Lemma 3.22 ensures that $W^{(2)}(\log^{**}\log^{**} n - 1) < \log\log n$ is for instance satisfied for all n of the form $n = W^{(2)}(m)$ with $m > 3$.)

Proposition 7.1 will follow by showing that for $1 - n^{-4} < q < 1$, a Mallows(n, q) distribution is in some sense very close to the uniform distribution over S_n.

Lemma 7.2. *Let $q = q(n)$ satisfy $1 - n^{-4} < q < 1$. For $1 \leq i \leq n$ let $Z_i \sim \mathrm{TruncGeo}(n - i + 1, 1 - q)$ and $X_i \sim \mathrm{Unif}([n - i + 1])$. Then*

$$d_{\mathrm{TV}}(Z_i, X_i) \leq 4 \cdot n^{-3},$$

for n sufficiently large and all $1 \leq i \leq n$.

Proof. Let us write $m := n - i + 1$ and $p_j := \mathbb{P}(Z_i = j)$. By (3) we may write

$$d_{\mathrm{TV}}(Z_i, X_i) = \sum_{j:p_j > 1/m} \left(p_j - \frac{1}{m}\right). \tag{19}$$

Now note that

$$\frac{\max_j p_j}{\min_j p_j} = q^{-(m-1)} < \left(\frac{1}{1 - n^{-4}}\right)^{m-1} < \left(\frac{1}{1 - n^{-4}}\right)^{n} \leq \left(1 + 2n^{-4}\right)^n \leq e^{2n^{-3}} \leq 1 + 4n^{-3},$$

where the first inequality holds for n sufficiently large; the third inequality uses $1 + x \leq e^x$, and; the last inequality holds for n sufficiently large, and uses the Taylor expansion for e^x. It follows that

$$p_j \leq (1 + 4n^{-3}) \cdot \frac{1}{m},$$

for all j. Filling this in to (19) gives $d_{\mathrm{TV}}(Z_i, X_i) \leq 4n^{-3}$ as desired. ∎

Lemma 7.3. *Let $q = q(n)$ satisfy $|1 - q| < n^{-4}$, and let $\Pi_n \sim \mathrm{Mallows}(n, q)$ and $\Pi_n^* \sim \mathrm{Mallows}(n, 1)$. There is a coupling of Π_n and Π_n^* such that $\mathbb{P}(\Pi_n \neq \Pi_n^*) \leq 4n^{-2}$, for all sufficiently large n.*

Proof. If $q = 1$, then the result is trivial. We now assume that $1 - n^{-4} < q < 1$.

Let Π_n be generated by a sequence $Z_1, \ldots, Z_n$ of independent random variables with $Z_i \sim \mathrm{TruncGeo}(n-i+1, 1-q)$ for $i \in [n]$ (see Section 3.2). Let Π_n^* be generated in the same manner by a sequence $X_1, \ldots, X_n$ of independent random variables where $X_i \sim \mathrm{Unif}([n-i+1])$ for $i \in [n]$. By Lemma 7.2 and 3.3 we may couple Z_i and X_i for each i such that

$$\mathbb{P}(Z_i \neq X_i) \leq 4n^{-3}, \quad \text{as } n \to \infty.$$

This gives a coupling of Π_n and Π_n^*. Now, $\Pi_n = \Pi_n^*$ if and only if $Z_i = X_i$ for all $i \in [n]$, giving

$$\mathbb{P}(\Pi_n \neq \Pi_n^*) \leq 4n^{-2}.$$

Finally we consider the case when $1 < q < 1 + n^{-4}$. Recall that if $r_n \in S_n$ is the "reverse" map given by $r_n(i) := n + 1 - i$, then $r_n \circ \Pi_n \stackrel{d}{=} \mathrm{Mallows}(n, 1/q)$. Note that $1 \geq 1/q \geq 1 - n^{-4}$, so by the previous case there exists a coupling of Π_n and $\Pi_n^{**} \sim \mathrm{Mallows}(n, 1)$ such that

$$\mathbb{P}(r_n \circ \Pi_n \neq \Pi_n^{**}) \leq 4n^{-2}.$$

Setting $\Pi_n^* = r_n \circ \Pi_n^{**}$, we have

$$\mathbb{P}(\Pi_n \neq \Pi_n^*) = \mathbb{P}(r_n \circ \Pi_n \neq r_n \circ \Pi_n^*) = \mathbb{P}(r_n \circ \Pi_n \neq \Pi_n^{**}) \leq 4n^{-2},$$

where we use that $r_n \circ r_n$ is the identity. Finally we note that if Π_n^{**} is uniformly distributed on S_n then so is $\Pi_n^* = r_n \circ \Pi_n^{**}$. ∎

Proof of Proposition 7.1. Let q satisfy $|1 - q| \leq n^{-4}$ and let $\Pi_n \sim \mathrm{Mallows}(n, q)$ and $\Pi_n^* \sim \mathrm{Mallows}(n, 1)$ be coupled such that

$$\mathbb{P}(\Pi_n \neq \Pi_n^*) = O(n^{-2}).$$

This coupling exists by Lemma 7.3. Let φ be the TOTO sentence from Proposition 6.1. Then Π_n and Π_n^* agree on φ with probability $1 - O(n^{-2})$, giving the desired result. ∎

7.2 Extending the $q = 1 \pm n^{-4}$ case to $q = 1 \pm O(1/n)$

In this subsection we prove the following intermediate result.

Proposition 7.4. *There is a sentence $\psi \in \mathsf{TOTO}$ such that the following holds. Suppose that $q = q(n)$ satisfies $1 - c/n < q < 1 + c/n$ for some constant $c > 0$. For all n satisfying $W^{(2)}(\log^{**}\log^{**} n - 1) < \log\log\log n$ we have*

$$\mathbb{P}(\Pi_n \models \psi) = \begin{cases} 1 - O(n^{-.001}) & \text{if } \log^{**}\log^{**} n \text{ is even,} \\ O(n^{-.001}) & \text{if } \log^{**}\log^{**} n \text{ is odd,} \end{cases}$$

where the $O(n^{-.001})$ terms can be taken uniform over all sequences considered (but may depend on c).

(Again, the condition $W^{(2)}(\log^{**}\log^{**} n - 1) < \log\log\log n$ is satisfied if n is of the form $n = W^{(2)}(m)$ with $m > 3$, by Lemma 3.22.)

For $\pi \in S_n$ a permutation, we define

$$\Psi(\pi) := \min\{1 \leq j \leq n : \text{ there exist } 1 \leq i_1, i_2 \leq j \text{ such that } \pi(i_2) = \pi(i_1) + 1\}.$$

Letting $\Pi_n \sim \text{Mallows}(n, q)$ as usual, we define:

$$X_n := \Psi(\Pi_n).$$

Before studying the behaviour of X_n, we first present a preparatory lemma.

Lemma 7.5. *Suppose that $q = q(n)$ satisfies $1 - c/n < q < 1$. There exist $c_U = c_U(c) > 0$ and $c_L = c_L(c) > 0$ such that for n sufficiently large and all $1 \le m \le n$ and $A \subseteq [m]$ we have*

$$c_L \frac{|A|}{m} \le \mathbb{P}\left(\text{TruncGeo}(m, 1-q) \in A\right) \le c_U \frac{|A|}{m}.$$

Proof. We fix n and some $m \in [n]$. For convenience we again set $p_j := \mathbb{P}(Z_i = j)$. We again have

$$\frac{\max_j p_i}{\min_j p_j} = q^{-(m-1)} \le \left(\frac{1}{1 - cn^{-1}}\right)^n \le (1 + 2cn^{-1})^n \le e^{2c},$$

where the penultimate inequality holds for n sufficiently large. Similarly

$$\frac{\min_j p_j}{\max_j p_j} = q^{m-1} \ge (1 - cn^{-1})^n \ge \frac{1}{2} e^{-c},$$

where the last inequality holds for n sufficiently large. We can conclude that

$$c_L \cdot \frac{1}{m} \le p_j \le c_U \cdot \frac{1}{m},$$

for all $j = 1, \dots, m$, with $c_L := \frac{1}{2} e^{-c}, c_U := e^{2c}$. The result follows. ■

Lemma 7.6. *Let $q = q(n)$ satisfy $1 - cn^{-1} < q < 1 + cn^{-1}$ for some $c > 0$. For every $\varepsilon > 0$ we have*

$$\mathbb{P}\left(n^{1/2-\varepsilon} \le X_n \le n^{1/2+\varepsilon}\right) > 1 - n^{-\varepsilon},$$

provided that $n \ge n_0$ for some $n_0 = n_0(c, \varepsilon)$ depending on ε and c.

Proof. We first assume $q \le 1$. For $i \ge 2$, let E_i denote the event that $\Pi_n(i) \notin \{\Pi_n(1) - 1, \Pi_n(1) + 1, \dots, \Pi_n(i-1) - 1, \Pi_n(i-1) + 1\} =: A_i$. We point out that if $E_1, \dots, E_{i-1}$ hold, then $i - 1 \le |A_i| \le 2i$. Recalling the sampling procedure 6 for the Mallows model described in Section 3.2 and using the previous lemma, we find that for $i \ge 2$:

$$1 - c_U \frac{2i}{n - i + 1} \le \mathbb{P}(E_i | E_2 \cap \dots \cap E_{i-1}) \le 1 - c_L \frac{i-1}{n-i+1}.$$

Writing $x := \lfloor n^{1/2+\varepsilon} \rfloor$, we find that

$$\begin{array}{rcl} \mathbb{P}(X > x) & = & \mathbb{P}(E_2 \cap \dots \cap E_x) \\ & = & \mathbb{P}(E_2) \cdot \mathbb{P}(E_3 | E_2) \cdot \dots \cdot \mathbb{P}(E_x | E_2 \cap \dots \cap E_{x-1}) \\ & \le & \prod_{i=2}^x \left(1 - c_L \frac{i-1}{n-i+1}\right) \le \exp\left[-c_L \sum_{i=2}^x \frac{i-1}{n-i+1}\right] \\ & \le & \exp\left[-c_L \cdot (2/n) \cdot \binom{x}{2}\right] = \exp\left[-\Omega(n^{2\varepsilon})\right], \end{array}$$

where the third inequality holds assuming n is sufficiently large.

Writing $y := \lceil n^{1/2-\varepsilon} \rceil$, we find that

$$\begin{array}{rcl} \mathbb{P}(X < y) & = & \mathbb{P}(\overline{E_2}) + \mathbb{P}(\overline{E_3} | E_2) + \dots + \mathbb{P}(\overline{E_y} | E_2 \cap \dots \cap E_{y-1}) \\ & \le & \sum_{i=2}^y c_U \frac{2i}{n-i+1} \le \frac{c_U}{n} \sum_{i=1}^y i \\ & = & \frac{c_U}{n} \cdot \binom{y+1}{2} = O(n^{-2\varepsilon}). \end{array}$$

Combining these bounds completes the proof for the case when $q \le 1$. For the case when $1 \le q \le 1 + c/n$, we recall that $r_n \circ \Pi_n \stackrel{d}{=} \text{Mallows}(n, 1/q)$. We also note that $1 \ge 1/q \ge 1 - cn^{-1}$

and that $\Psi(\pi) = \Psi(r_n \circ \pi)$ for all $\pi \in S_n$, by definition of Ψ. So the result for $1 \leq q \leq 1 + c/n$ follows immediately from the result for $1 - cn^{-1} \leq q \leq 1$. ■

We next define

$$Y_n := \Psi\left(\mathrm{rk}(\Pi_n(1), \ldots, \Pi_n(X_n))\right),$$

$$Z_n := \Psi\left(\mathrm{rk}(\Pi_n(1), \ldots, \Pi_n(Y_n))\right).$$

The next observation will allow us to make use of Lemma 7.6 also for Y_n and Z_n.

Lemma 7.7. *For every $n \geq 3$, $2 \leq k \leq n-1$ and every $\pi \in S_n$ we have*

$$\Psi(\pi_k) \leq \Psi(\pi_{k+1}),$$

where $\pi_j := \mathrm{rk}(\pi(1), \ldots, \pi(j))$.

Proof. This follows readily from the observation that, for all distinct $i_1, i_2 \leq k$, $\pi_{k+1}(i_2) = \pi_{k+1}(i_1) + 1$ implies that $\pi_k(i_2) = \pi_k(i_1) + 1$. ■

Lemma 7.8. *Suppose that $q = q(n)$ satisfies $1 - c/n < q < 1 + c/n$ for some constant $c > 0$. Then*

$$\mathbb{P}\left(n^{.1} \leq Z_n \leq n^{.2}\right) > 1 - n^{-.001},$$

provided that $n \geq n_0$ for some $n_0 = n_0(c)$ depending on c.

Proof. We fix $\delta > 0$, to be determined during the course of the proof. For notational convenience, let us write $x := \lceil n^{1/2-\delta} \rceil, y := \lceil n^{1/4-\delta} \rceil, z := \lceil n^{1/8-\delta} \rceil$. Recall that by Lemma 3.9, for each fixed j we have $\Pi_j^* := \mathrm{rk}(\Pi_n(1), \ldots, \Pi_n(j)) \stackrel{d}{=} \mathrm{Mallows}(j, q)$. Using also the previous two lemmas and the definitions of X_n, Y_n, Z_n, we find:

$$\begin{array}{rcl} \mathbb{P}(Z < z) & \leq & \mathbb{P}(X_n < x) + \mathbb{P}(\Psi(\Pi_x^*) < y) + \mathbb{P}(\Psi(\Pi_y^*) < z) \\ & = & \mathbb{P}(X_n < x) + \mathbb{P}(\Psi(\Pi_x) < y) + \mathbb{P}(\Psi(\Pi_y) < z) \\ & = & \mathbb{P}(X_n < x) + \mathbb{P}(X_x < y) + \mathbb{P}(X_y < z) \\ & \leq & n^{-\delta} + x^{-\delta} + y^{-\delta} \\ & \leq & n^{-\delta/8}, \end{array}$$

where the fourth line holds for n sufficiently large and we use

$$\begin{array}{c} x^{1/2-\delta} = (1+o(1))n^{(1/2-\delta)^2} > y = (1+o(1))n^{1/4-\delta}, \\ y^{1/2-\delta} = (1+o(1))n^{(1/2-\delta)^3} > z = (1+o(1))n^{1/8-\delta}, \end{array}$$

and the last line holds for n sufficiently large.

Analogously we have, writing $x' := \lfloor n^{1/2+\delta} \rfloor, y' := \lfloor n^{1/4+2\delta} \rfloor, z := \lfloor n^{1/8+4\delta} \rfloor$:

$$\begin{array}{rcl} \mathbb{P}(Z > z') & \leq & \mathbb{P}(X_n > x') + \mathbb{P}(X_{x'} > y') + \mathbb{P}(X_{y'} > z') \\ & \leq & n^{-\delta} + x'^{-\delta} + y'^{-\delta} \\ & \leq & n^{-\delta/8}, \end{array}$$

for n sufficiently large.

Setting $\delta = .025$ and combining the bounds gives the result, with some room to spare. ■

We are almost ready to prove Proposition 7.4. But, we first point out that it is possible to express that $i \leq Z_n$ in TOTO.

Lemma 7.9. *There is a formula $\varphi_Z(x)$ in* TOTO *with one free variable such that for every $n \in \mathbb{N}$, every permutation $\pi \in S_n$ and every $1 \le i \le n$:*

$$(\pi, i) \models \varphi_Z(x) \text{ if and only if } i \le Z_n(\pi).$$

Proof. We first show there is a formula $\varphi_X(x)$ that expresses that $i \le \Psi(\pi)$ if $\pi \in S_n, i \in [n]$ for some $n \in \mathbb{N}$. This formula is given by

$$\forall y, z : ((y <_1 x) \wedge (z <_1 x)) \implies (\neg \mathsf{succ}_2(y,z) \wedge \neg \mathsf{succ}_2(z,y)).$$

("For every pair $y, z < x$ we have $|\pi(y) - \pi(z)| \neq 1$".) Now let $\varphi_Y := \varphi_X^{\varphi_X}$, the relativization of φ_X to φ_X, as provided by Lemma 3.19. Then, invoking that lemma, $(\pi, i) \models \varphi_Y(x)$ if and only if $i \le \Psi(\pi)$ and $(\sigma, i) \models \varphi_X(x)$ where $\sigma := \mathrm{rk}(\pi(1), \dots, \pi(\Psi(\pi)))$. In other words, $(\pi, i) \models \varphi_Y(x)$ if and only if $i \le Y_n(\pi)$.

Similarly, if we now set $\varphi_Z := \varphi_X^{\varphi_Y}$ then $(\pi, i) \models \varphi_Z(x)$ if and only if $i \le Z_n(\pi)$. ■

Proof of Proposition 7.4: Let φ be the sentence provided by Proposition 7.1 and let φ_Z be the formula provided by Lemma 7.9. Then the relativization $\psi := \varphi^{\varphi_Z}$ (as provided by Lemma 3.19) expresses that φ is satisfied by $\mathrm{rk}(\Pi_n(1), \dots, \Pi_n(Z_n))$.

Now let n be such that $W^{(2)}(\log^{**} \log^{**} n - 1) < \log\log\log n$. First assume that $\log^{**}\log^{**} n$ is even. We have, writing $\Pi_j^* := \mathrm{rk}(\Pi_n(1), \dots, \Pi_n(j))$:

$$\begin{aligned}
\mathbb{P}(\Pi_n \models \psi) &\ge \mathbb{P}(\Pi_m^* \models \varphi \text{ for all } n^{.1} \le m \le n^{.2}) - \mathbb{P}(Z_n \notin [n^{.1}, n^{.2}]) \\
&\ge 1 - \sum_{n^{.1} \le m \le n^{.2}} \mathbb{P}(\Pi_m^* \not\models \varphi) - \mathbb{P}(Z_n \notin [n^{.1}, n^{.2}]) \\
&= 1 - \sum_{n^{.1} \le m \le n^{.2}} O(m^{-100}) - O(n^{-.001}) \\
&= 1 - O(n^{-.001}),
\end{aligned}$$

where we use Lemma 3.9 and Proposition 7.1 together with the fact that, for n sufficiently large and all $n^{.1} \le m \le n^{.2}$, we have

$$\log\log m > \log\log\log n > W^{(2)}(\log^{**}\log^{**} n - 1) \ge W^{(2)}(\log^{**}\log^{**} m - 1).$$

Now assume that $\log^{**}\log^{**} n$ is odd. In this case we have

$$\begin{aligned}
\mathbb{P}(\Pi_n \models \psi) &\le \sum_{n^{.1} \le m \le n^{.2}} \mathbb{P}(\Pi_m^* \models \varphi) + \mathbb{P}(Z_n \notin [n^{.1}, n^{.2}]) \\
&\le \sum_{n^{.1} \le m \le n^{.2}} O(m^{-100}) + O(n^{-.001}) \\
&= O(n^{-.001}),
\end{aligned}$$

with the same justifications as previously. ■

7.3 Lifting to $q = 1 \pm 1/\log^* n$

In this section we finalize the proof of Theorem 1.2 Part **(ii)**. Specifically, we will prove the following more detailed version of Theorem 1.2 Part **(ii)**.

Proposition 7.10. *There exists a sentence $\xi \in$ TOTO such that, for all sequences $q = q(n)$ satisfying $1 - 1/\log^* n < q < 1 + 1/\log^* n$ and all n satisfying $W^{(2)}(\log^{**}\log^{**} n - 1) < \log^*\log^*\log^* n$ we have*

$$\mathbb{P}(\Pi_n \models \xi) = \begin{cases} 1 - o(1) & \textit{if } \log^{**}\log^{**} n \textit{ is even,} \\ o(1) & \textit{if } \log^{**}\log^{**} n \textit{ is odd,} \end{cases},$$

where the error terms $o(1)$ can be taken uniform over all sequences $q(n)$ considered.

(Again, we note the condition $W^{(2)}(\log^{**}\log^{**} n - 1) < \log^*\log^*\log^* n$ is for instance satisfied for all n of the form $n = W^{(2)}(m)$ with $m > 3$, by Lemma 3.22.)

The quantity $\Pi_n^{-1}(1)$ will play an important role in the definition of the TOTO property ξ. We start with some preparatory lemmas on its behaviour.

Lemma 7.11. *For every $c, \varepsilon > 0$ there exist constants $c_1 = c_1(c, \varepsilon), c_2 = c_2(c, \varepsilon) > 0$ and $n_0 = n_0(c, \varepsilon)$ such that for $n \geq n_0$ and any sequence $q = q(n)$ satisfying $1 - 1/\log^* n < q < 1 - c/n$, the random permutation $\Pi_n \sim \mathrm{Mallows}(n, q)$ satisfies*

$$\mathbb{P}\left(\frac{c_1}{1-q} \leq \Pi_n^{-1}(1) \leq \frac{c_2}{1-q}\right) > 1 - \varepsilon.$$

Proof. Recall from Section 3.2 that if $\Pi_n \sim \mathrm{Mallows}(n, q)$ then also $\Pi_n^{-1} \stackrel{d}{=} \Pi_n$. In particular $\Pi_n^{-1}(1) \stackrel{d}{=} \Pi_n(1) \stackrel{d}{=} \mathrm{TruncGeo}(n, 1-q)$, where the last distributional equality follows from the the sampling algorithm presented in Section 3.2. Writing $x_1 := \lfloor c_1/(1-q) \rfloor$, we first note that provided the constant $c_1 > 0$ is chosen sufficiently small, we have $x_1 < (c_1/c) \cdot n \leq n$. We can thus write:

$$\mathbb{P}\left(\Pi_n^{-1}(1) \leq \frac{c_1}{1-q}\right) = \sum_{i=1}^{x_1} \frac{q^{i-1}(1-q)}{1-q^n} = \frac{1-q^{x_1}}{1-q^n} < \frac{1-q^{x_1}}{1-(1-c/n)^n} < \frac{1-q^{x_1}}{1-e^{-c}}.$$

Now we note that

$$q^{x_1} = (1-(1-q))^{x_1} \geq 1 - x_1 \cdot (1-q) \geq 1 - c_1.$$

It follows that

$$\mathbb{P}\left(\Pi_n^{-1}(1) \leq \frac{c_1}{1-q}\right) \leq \frac{c_1}{1-e^{-c}} < \varepsilon/2,$$

provided we chose the constant $c_1 > 0$ sufficiently small.

Similarly, writing $x_2 := c_2/(1-q)$ we have

$$\mathbb{P}(\Pi_n^{-1}(1) > x_2) < \frac{1}{1-(1-c/n)^n} \cdot q^{x_2} \leq \frac{1}{1-e^{-c}} \cdot (1-(1-q))^{c_2/(1-q)} \leq \frac{e^{-c_2}}{1-e^{-c}} < \varepsilon/2,$$

where the last inequality holds provided we chose c_2 sufficiently large. The result follows combining the two bounds. ■

Lemma 7.12. *For every $c, \varepsilon > 0$ there exist $\delta = \delta(c, \varepsilon), n_0 = n_0(c, \varepsilon)$ such that for all sequences $q = q(n)$ with $1 - c/n \leq q \leq 1 + c/n$ and all $n \geq n_0$:*

$$\mathbb{P}\left(\delta n \leq \Pi_n^{-1}(1) \leq (1-\delta)n\right) > 1 - \varepsilon.$$

Proof. First note that in the special case when $q = 1$, we have that $\Pi_n^{-1}(1)$ is chosen uniformly at random from $[n]$. So any $\delta < \varepsilon/2$ would work in this case. It therefore suffices to consider $q \neq 1$ so long as we ensure the value of δ we end up selecting in the end is $< \varepsilon/2$. (Which is easily achieved by adjusting the choice of δ in the end.)

Next, we consider the case when $1 - c/n \le q < 1$. Again we will use that $\Pi_n^{-1}(1) \sim \text{TruncGeo}(n, 1-q)$. With $c_U = c_U(c)$ as provided by Lemma 7.5, we can write:

$$\mathbb{P}\left(\delta n \le \Pi_n^{-1}(1) \le (1-\delta)n\right) \ge 1 - c_U \cdot 2\delta > 1 - \varepsilon,$$

provided we chose δ small enough.

It remains to consider $1 < q < 1 + c/n$. We recall that $\Pi_n^{-1} \stackrel{d}{=} \Pi_n \stackrel{d}{=} \text{Mallows}(n, q)$ and hence $r_n \circ \Pi_n^{-1} \stackrel{d}{=} \text{Mallows}(n, 1/q)$. But then $1/(1 + c/n) > 1 - c/n$, together with the previous case gives

$$\begin{aligned}
\mathbb{P}\left(\delta n \le \Pi_n^{-1}(1) \le (1-\delta)n\right) &= \mathbb{P}\left(n+1-(1-\delta)n \le (r_n \circ \Pi_n^{-1})(1) \le n+1-\delta n\right) \\
&\ge \mathbb{P}\left(2\delta \le (r_n \circ \Pi_n^{-1})(1) \le (1-2\delta)n\right) \\
&> 1 - \varepsilon,
\end{aligned}$$

where the first inequality holds for n sufficiently large, and the last inequality by the previous case and adjusting the value of δ. ■

For the remainder of the section, we let Σ_n denote the permutation on $\{1, \dots, \Pi_n^{-1}(1) - 1\}$ induced by Π_n. (If $\Pi_n(1) = 1$, then Σ_n is undefined.) In other words, we set

$$\Sigma_n := \text{rk}(\Pi_n(1), \dots, \Pi_n(\Pi_n^{-1}(1) - 1)).$$

Lemma 7.13. *For $0 < q < 1$, $n \ge 2$ and $2 \le k \le n$ we have*

$$\left(\Sigma_n \middle| \Pi_n^{-1}(1) = k\right) \stackrel{d}{=} \text{Mallows}(k-1, q).$$

Proof. Fix an arbitrary $0 < q < 1, n \ge 2$ and $2 \le k \le n$. Let $Z_1, \dots, Z_n$ be the random variables used in the sampling algorithm from Section 3.2. In particular, they're independent with $Z_i \sim \text{TruncGeo}(n + 1 - i, 1 - q)$. We first observe that

$$\{\Pi_n^{-1}(1) = k\} = \{\Pi_n(k) = 1\} = \{Z_1 > 1, \dots, Z_{k-1} > 1, Z_k = 1\}, \tag{20}$$

We set $W_i := Z_i - 1$ for $i < k$ and $W_i = Z_{i+1}$ for $k \le i \le n-1$. Next, we claim that conditional on the event $\Pi_n^{-1}(1) = k$ the random variables $W_1, \dots, W_{n-1}$ are independent with $W_i \sim \text{TruncGeo}(n - i, 1 - q)$. That they are independent (conditional on $\Pi_n^{-1}(1) = k$) should be obvious from (20) and the independence of $Z_1, \dots, Z_n$. Their distribution is also obvious for $i \ge k$. For $i < k$ the claim follows from

$$\begin{aligned}
\mathbb{P}(W_i = w | \Pi_n^{-1}(1) = k) &= \mathbb{P}(Z_i = w + 1 | Z_i > 1) \\
&= \left(\frac{q^w(1-q)}{1-q^{n+1-i}}\right) / \left(\frac{q - q^{n+1-i}}{1 - q^{n+1-i}}\right) \\
&= \frac{q^{w-1}(1-q)}{1-q^{n-i}}.
\end{aligned}$$

Conditional on $\Pi_n^{-1}(1) = k$, we can define $\Sigma_n^* \in S_{n-1}$ by

$$\Sigma_n^*(i) := \begin{cases} \Pi_n(i) - 1 & \text{if } 1 \le i \le k-1, \\ \Pi_n(i+1) - 1 & \text{if } k \le i \le n-1. \end{cases}$$

Now observe that, still conditional on $\Pi_n^{-1}(1) = k$, we have

$$\Sigma_n^*(i) = \begin{cases} W_1 & \text{if } i = 1, \\ \text{the } W_i\text{-th element of } [n-1] \setminus \{W_1, \dots, W_{i-1}\} & \text{otherwise.} \end{cases},$$

It follows that

$$\left(\Sigma_n^* \middle| \Pi_n^{-1}(1) = k\right) \stackrel{d}{=} \text{Mallows}(n-1, q).$$

This gives

$$\begin{aligned}\left(\Sigma_n \middle| \Pi_n^{-1}(1) = k\right) &= \left(\text{rk}(\Pi_n(1), \dots, \Pi_n(k-1)) \middle| \Pi_n^{-1}(1) = k\right) \\ &= \left(\text{rk}(\Sigma_n^*(1), \dots, \Sigma_n^*(k-1)) \middle| \Pi_n^{-1}(1) = k\right) \\ &\stackrel{d}{=} \text{Mallows}(k-1, q),\end{aligned}$$

applying Lemma 3.9 in the last line. ■

We next establish that $i < \Pi_n^{-1}(1)$ is in fact expressible in TOTO. This is a seemingly innocuous statement. As a side remark, let us mention that it is not possible to express that $i < \Pi_n(1)$ in TOTO (since it is not relevant for the proofs in this paper, we do not offer a formal proof). This is the reason we use $\Pi_n^{-1}(1)$ and not $\Pi_n(1)$ to define Σ_n above.

Lemma 7.14. *There is a* TOTO *formula* χ *with one free variable, such that for ever* $n \in \mathbb{N}$, *every permutation* $\pi \in S_n$ *and every* $1 \le i \le n$ *we have*

$$(\pi, i) \models \chi(x) \textit{ if and only if } i < \pi^{-1}(1).$$

Proof. The formula χ can be written as:

$$\exists y : (x <_1 y) \wedge \neg\left(\exists z : z <_2 y\right).$$

So, the formula χ asks that there exists y with $x < y$ such that there is no z with $\pi(z) < \pi(y)$. This clearly does the trick. ■

We now present a final preparatory TOTO sentence whose probability does not convergence on part of the target range for $q = q(n)$.

Corollary 7.15. *There exists a sentence* $\zeta \in$ TOTO *such that for every fixed* $c > 0$, *all sequences* $q = q(n)$ *satisfying* $1 - 1/\log^* n < q < 1 + c/n$ *and all* n *satisfying* $W^{(2)}(\log^{**}\log^{**} n - 1) < \log^*\log^*\log^* n$ *we have*

$$\mathbb{P}(\Pi_n \models \zeta) = \begin{cases} 1 - o(1) & \textit{if } \log^{**}\log^{**} n \textit{ is even,} \\ o(1) & \textit{if } \log^{**}\log^{**} n \textit{ is odd} \end{cases},$$

where the error terms $o(1)$ *can be taken uniform over all sequences considered.*

Proof. Let ψ is as provided by Proposition 7.4 and let χ be as provided by Lemma 7.14. We let $\zeta := \psi^\chi$ be the relativization as provided by Lemma 3.19. Then ζ formalizes that Σ_n satisfies ψ.

Now let $\varepsilon > 0$ be arbitrary (but fixed) and let n be such that $W^{(2)}(\log^{**}\log^{**} n - 1) < \log^*\log^*\log^* n$. Let $\delta = \delta(c, \varepsilon)$ be as provided by Lemma 7.12 and let $c_1 = c_1(c, \varepsilon), c_2 = c_2(c, \varepsilon)$ be as provided by Lemma 7.11. We define

$$k_1 := \begin{cases} \delta n & \text{if } q > 1 - c/n, \\ \frac{c_1}{1-q} & \text{otherwise.} \end{cases}, \quad k_2 := \begin{cases} n & \text{if } q > 1 - c/n, \\ \frac{c_2}{1-q} & \text{otherwise.} \end{cases}.$$

A crucial step in the current proof is the observation that, provided n is sufficiently large, for all $k_1 \le k \le k_2$ we have

$$1 - \frac{C}{k-1} < q < 1 + \frac{C}{k-1}, \tag{21}$$

where we can take $C := 2c/\delta$ in the case when $q > 1 - c/n$ and $C := 2/c_1$ otherwise.

If $\log^{**}\log^{**} n$ is even then, using Lemma 7.13, we can write

$$\begin{aligned}
\mathbb{P}(\Pi_n \models \zeta) &= \mathbb{P}(\Sigma_n \models \psi) \\
&\geq \sum_{k_1 \leq k \leq k_1} \mathbb{P}(\Sigma_n \models \psi | \Pi_n^{-1}(1) = k) \cdot \mathbb{P}(\Pi_n^{-1}(1) = k) \\
&= \sum_{k_1 \leq k \leq k_2} \mathbb{P}(\text{Mallows}(k-1, q) \models \psi) \cdot \mathbb{P}(\Pi_n^{-1}(1) = k) \\
&\geq \min_{k_1 \leq k \leq k_2} \mathbb{P}(\text{Mallows}(k-1, q) \models \psi) - \mathbb{P}(\Pi_n^{-1}(1) < k_1 \text{ or } \Pi_n(1) > k_2) \\
&> 1 - O\left((\log^* n)^{-.001}\right) - \varepsilon,
\end{aligned}$$

using in the last line Proposition 7.4 together with (21) and that the fact that for all $k \geq k_1$ we have

$$\begin{aligned}
\log\log\log k &\geq \log\log\log(c_1 \cdot \log^* n) \geq \log^* \log^* \log^* n \\
&> W^{(2)}(\log^{**}\log^{**} n - 1) \geq W^{(2)}(\log^{**}\log^{**} k - 1),
\end{aligned}$$

for n sufficiently large. If $\log^{**}\log^{**} n$ is odd then, analogously:

$$\begin{aligned}
\mathbb{P}(\Pi_n \models \zeta) &\leq \max_{k_1 \leq k \leq k_2} \mathbb{P}(\text{Mallows}(k-1, q) \models \psi) + \mathbb{P}(\Pi_n^{-1}(1) < k_1 \text{ or } \Pi_n^{-1}(1) > k_2) \\
&< O\left((\log^* n)^{-.001}\right) + \varepsilon.
\end{aligned}$$

Sending $\varepsilon \searrow 0$ finishes the proof. ■

The sentence ζ provided in the previous corollary works in an asymmetric range of q. In order to remedy this, we will use a construction that allows us to distinguish (in TOTO, with appropriately high probability) between the case when $q < 1 - c/n$ and $q > 1 + c/n$. The following observation is the key to that construction.

Lemma 7.16. *For every $\varepsilon > 0$ there are $c = c(\varepsilon), n_0 = n_0(\varepsilon)$ such the following holds for all $n \geq n_0$;*

- **(i)** *If $0 < q < 1 - c/n$ then $\mathbb{P}(\Pi_n(1) < \Pi_n(n)) > 1 - \varepsilon$, and;*
- **(ii)** *If $q > 1 + c/n$ then $\mathbb{P}(\Pi_n(1) < \Pi_n(n)) < \varepsilon$.*

Proof. We first consider the case $q < 1 - c/n$. By Theorem 3.12 and Markov's inequality:

$$\mathbb{P}(\Pi_n(1) > n/10) \leq \mathbb{P}(|\Pi_n(1) - 1| > n/100) \leq \frac{100 \cdot \mathbb{E}|\Pi_n(1) - 1|}{n} \leq \frac{200q}{n(1-q)} \leq \frac{12}{c} < \varepsilon/2,$$

where the first inequality holds for n sufficiently large and the last inequality holds provided the constant c is chosen sufficiently large. Analogously

$$\mathbb{P}(\Pi_n(n) < 9n/10) \leq \frac{100 \cdot \mathbb{E}|\Pi_n(n) - n|}{n} < \varepsilon/2.$$

Part **(i)** follows.

Assume now $q > 1 + c/n$ for some large constant c to be determined. Recall that $r_n \circ \Pi_n \stackrel{d}{=} \text{Mallows}(n, 1/q)$. We have $1/q < 1/(1 + c/n) < 1 - (c/2)/n$, where the last inequality holds for n sufficiently large. Using the previous part:

$$\mathbb{P}(\Pi_n(1) < \Pi_n(n)) = \mathbb{P}(r_n \circ \Pi_n(1) > r_n \circ \Pi_n(n)) < \varepsilon,$$

for n sufficiently large (and provided we chose c sufficiently large). ■

Proof of Proposition 7.10: Let ζ be as provided by the previous corollary, let ζ^{reverse} be as provided by Lemma 3.20, and let

$$\theta := \exists x \exists y : (x <_2 y) \wedge \neg (\exists z : z <_1 x) \wedge \neg (\exists z : y <_1 z),$$

denote a TOTO sentence that formalizes that $\Pi_n(1) < \Pi_n(n)$. The sentence ξ is now given by

$$\xi := (\theta \wedge \zeta) \vee (\neg\theta \wedge \zeta^{\text{reverse}}).$$

That is, ξ demands that one of **a)** $\Pi_n(1) < \Pi_n(n)$ and ζ holds, or **b)** $\Pi_n(1) > \Pi_n(n)$ and ζ holds for $r_n \circ \Pi_n$.

Let $\varepsilon > 0$ be arbitrary (but fixed) and let $c > 0$ be as provided by Lemma 7.16. We consider an arbitrary sequence $q = q(n)$ with $1 - 1/\log^* n < q < 1 + 1/\log^* n$. Let n be such that $W^{(2)}(\log^{**}\log^{**} n - 1) < \log^* \log^* \log^* n$. First we assume that $\log^{**}\log^{**} n$ is even and $1 - c/n < q < 1 + c/n$. In this case

$$\begin{aligned}
\mathbb{P}(\Pi_n \models \psi) &\geq \mathbb{P}(\Pi_n \models \zeta \text{ and } r_n \circ \Pi_n \models \zeta) \\
&\geq 1 - \mathbb{P}(\Pi_n \not\models \zeta) - \mathbb{P}(r_n \circ \Pi_n \not\models \zeta) \\
&= 1 - o(1),
\end{aligned}$$

If $\log^{**}\log^{**} n$ is odd and $1 - c/n < q < 1 + c/n$ then similarly

$$\mathbb{P}(\Pi_n \models \psi) \leq \mathbb{P}(\Pi_n \models \zeta) + \mathbb{P}(r_n \circ \Pi_n \models \zeta) = o(1).$$

If $\log^{**}\log^{**} n$ is even and $1 - 1/\log^* n < q \leq 1 - c/n$ then we have

$$\mathbb{P}(\Pi_n \models \psi) \geq \mathbb{P}(\Pi_n \models \zeta) - \mathbb{P}(\Pi_n(1) > \Pi_n(n)) > 1 - o(1) - \varepsilon.$$

If $\log^{**}\log^{**} n$ is odd and $1 - 1/\log^* n < q \leq 1 - c/n$ then

$$\mathbb{P}(\Pi_n \models \psi) \leq \mathbb{P}(\Pi_n \models \zeta) + \mathbb{P}(\Pi_n(1) > \Pi_n(n)) < o(1) + \varepsilon.$$

Now suppose that $\log^{**}\log^{**} n$ is even and $1 + c/n \leq q < 1 + 1/\log^* n$. Observe that $1 - 1/\log^* n < 1/q < 1 - (c/2)/n$ for n sufficiently large. Adjusting the value of c if needed, we can again apply Corollary 7.15 to obtain:

$$\mathbb{P}(\Pi_n \models \psi) \geq \mathbb{P}(r_n \circ \Pi_n \models \zeta) - \mathbb{P}(\Pi_n(1) < \Pi_n(n)) > 1 - o(1) - \varepsilon.$$

Similarly, if $\log^{**}\log^{**} n$ is odd and $1 + c/n \leq q < 1 + 1/\log^* n$ then

$$\mathbb{P}(\Pi_n \models \psi) \leq \mathbb{P}(r_n \circ \Pi_n \models \zeta) + \mathbb{P}(\Pi_n(1) < \Pi_n(n)) < o(1) + \varepsilon.$$

The result follows by sending $\varepsilon \searrow 0$. ■

8 Discussion and suggestions for further work

Theorem 1.1 established that for $q < 1$ the zero-one law holds wrt. TOOB, while for $q > 1$ there is some TOOB sentence whose probability of holding does not converge with n. (In the case when $q = 1$ Compton [15] has shown decades ago that the convergence law holds, but not the zero-one law.) The property exhibiting the non-convergence simply asks for at least a certain number of fixed points, exploiting a result from [31] (repeated as Theorem 3.11 in the present paper). Having another look at the result on the limiting distributions for the cycle counts provided by that result, it should be relatively straightforward to show that when $q > 1$ and we consider either the sequence $\text{Mallows}(2n, q)$ of Mallows permutations restricted to even domain sizes or the sequence Mallows$(2n+1, q)$ restricted to odd domain sizes, then in fact the convergence law will hold.

Theorem 1.1 only considers the case of q fixed. The case when $q = q(n)$ depends on n is certainly worth investigating, in particular the "phase transition" regime when $q \to 1$. A natural line of inquiry would be to try and determine a "critical window" inside which the behaviour is essentially the same as $q = 1$ (convergence law but not zero-one law) while below we get the zero-one law and above the convergence law fails. The behaviour of TOOB sentences is essentially determined by the limiting distribution of (short) cycles, so one might expect this to be relatively straightforward. The behaviour of cycles in the regime when $q \to 1$ is however far from well understood at the moment. We point the reader to [31] for some more discussion of the relevant open questions on cycles in the Mallows model.

In Theorem 1.2, Part **(ii)** we did in fact consider a range of q with $q \to 1$. We constructed a sequence $\varphi \in \mathsf{TOTO}$ such that $\mathbb{P}(\text{Mallows}(n, q) \models \varphi)$ does not converge for any sequence $q = q(n)$ with $1 - 1/\log^* n < q < 1 + 1/\log^* n$. For the φ we've constructed, we believe it should be possible to determine a sequence $q = q(n)$ that converges to 1 (very slowly) while $\mathbb{P}(\text{Mallows}(n, q) \models \varphi)$ converges.

On the other hand, as mentioned earlier, the range $(1-1/\log^* n, 1+1/\log^* n)$ is certainly not best possible, as should evident from the proof. It should be possible to achieve a range for q that approaches 1 much more slowly, with a minor variation on our proof. The only constraint on the range we can obtain seems to be what we can axiomatize using the arithmetization construction in Section 6.1. That is, if we can axiomatize some increasing function $F(n)$ using binary relations as in Section 6.1 and $f(n)$ is its discrete inverse (i.e. $f(n) := \max\{m \leq n : F(m) \leq n\}$), then one should be able to construct a φ such that $\mathbb{P}(\text{Mallows}(n, q) \models \varphi)$ does not converge for all $1 - 1/f(n) < q < 1 + 1/f(n)$. It is not quite clear to the authors at present, most likely just for want of a proper background in the relevant subfields of Mathematics, exactly how wide of a range can be obtained in this manner. But, perhaps a different proof technique can yield even wider ranges. In particular, perhaps one can widen the range of q where there is non-convergence even more if one does not insist on finding a single sentence that works in the entire range?

Acknowledgements

We thank Kevin Compton, Valentin Féray, Tomasz Łuczak, Marc Noy and Joel Spencer for helpful discussions and pointers to the literature. We thank Pieter Trapman and Rineke Verbrugge for helpful and very detailed comments on the chapter of TWV's PhD thesis on which this paper is based

References

[1] L. Addario-Berry and B. Corsini. The height of Mallows trees. *Ann. Probab.*, 49(5):2220–2271, 2021.

[2] M. Albert, M. Bouvel, and V. Féray. Two first-order logics of permutations. *J. Combin. Theory Ser. A*, 171:105–158, 46, 2020.

[3] M. Albert, M. Bouvel, V. Féray, and M. Noy. Convergence law for 231-avoiding permutations. *arXiv preprint arXiv:2210.05537*, 2022.

[4] M. H. Albert and A. M. Frieze. Random graph orders. *Order*, 6(1):19–30, 1989.

[5] O. Angel, A. E. Holroyd, T. Hutchcroft, and A. Levy. Mallows permutations as stable matchings. *Canad. J. Math.*, 73(6):1531–1555, 2021.

[6] R. Basu and N. Bhatnagar. Limit theorems for longest monotone subsequences in random Mallows permutations. *Ann. Inst. Henri Poincaré Probab. Stat.*, 53(4):1934–1951, 2017.

[7] I. Benjamini, N. Berger, C. Hoffman, and E. Mossel. Mixing times of the biased card shuffling and the asymmetric exclusion process. *Trans. Amer. Math. Soc.*, 357(8):3013–3029, 2005.

[8] N. Bhatnagar and R. Peled. Lengths of monotone subsequences in a Mallows permutation. *Probab. Theory Related Fields*, 161(3-4):719–780, 2015.

[9] B. Bollobás and P. Winkler. The longest chain among random points in euclidean space. *Proceedings of the American Mathematical Society*, 103(2):347–353, 1988.

[10] A. Borodin, P. Diaconis, and J. Fulman. On adding a list of numbers (and other one-dependent determinantal processes). *Bull. Amer. Math. Soc. (N.S.)*, 47(4):639–670, 2010.

[11] S. Braunfeld and M. Kukla. Logical limit laws for layered permutations and related structures. *arXiv preprint arXiv:2111.06307*, 2021.

[12] M. Braverman and E. Mossel. Sorting from noisy information. Preprint, available from `http://arxiv.org/abs/0910.1191`, 2009.

[13] G. Brightwell. *Models of random partial orders*, volume 187. Cambridge University Press Cambridge, 1993.

[14] G. Brightwell and M. Luczak. The mathematics of causal sets. In *Recent Trends in Combinatorics*, pages 339–365. Springer, 2016.

[15] Kevin J. Compton. A logical approach to asymptotic combinatorics. II. Monadic second-order properties. *J. Combin. Theory Ser. A*, 50(1):110–131, 1989.

[16] H. Crane and S. DeSalvo. Pattern avoidance for random permutations. *Discrete Math. Theor. Comput. Sci.*, 19(2):Paper No. 13, 24, 2017.

[17] H. Crane, S. DeSalvo, and S. Elizalde. The probability of avoiding consecutive patterns in the Mallows distribution. *Random Structures Algorithms*, 53(3):417–447, 2018.

[18] P. Diaconis and A. Ram. Analysis of systematic scan Metropolis algorithms using Iwahori-Hecke algebra techniques. volume 48, pages 157–190. 2000. Dedicated to William Fulton on the occasion of his 60th birthday.

[19] R. Douc, E. Moulines, P. Priouret, and P. Soulier. *Markov chains*. Springer Series in Operations Research and Financial Engineering. Springer, Cham, 2018.

[20] J. Enright, K. Meeks, W. Pettersson, and J. Sylvester. Tangled paths: A random graph model from Mallows permutations. *arXiv preprint arXiv:2108.04786*, 2021.

[21] C. Even-Zohar and C. Leng. Counting small permutation patterns. In *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 2288–2302. SIAM, 2021.

[22] X. Fang, H. L. Gan, S. Holmes, H. Huang, E. Peköz, A. Röllin, and W. Tang. Arcsine laws for random walks generated from random permutations with applications to genomics. *J. Appl. Probab.*, 58(4):851–867, 2021.

[23] J. Foy and A. R. Woods. Probabilities of sentences about two linear orderings. In *Feasible mathematics (Ithaca, NY, 1989)*, volume 9 of *Progr. Comput. Sci. Appl. Logic*, pages 181–193. Birkhäuser Boston, Boston, MA, 1990.

[24] A. Gladkich and R. Peled. On the cycle structure of Mallows permutations. *Ann. Probab.*, 46(2):1114–1169, 2018.

[25] A. Gnedin and G. Olshanski. q-exchangeability via quasi-invariance. *Ann. Probab.*, 38(6):2103–2135, 2010.

[26] A. Gnedin and G. Olshanski. The two-sided infinite extension of the Mallows model for random permutations. *Adv. in Appl. Math.*, 48(5):615–639, 2012.

[27] R. L. Goodstein. Transfinite ordinals in recursive number theory. *J. Symbolic Logic*, 12:123–129, 1947.

[28] G. R. Grimmett and D. R. Stirzaker. *Probability and random processes.* Oxford University Press, New York, third edition, 2001.

[29] S. Haber and M. Krivelevich. The logic of random regular graphs. *Journal of combinatorics*, 1(4):389–440, 2010.

[30] J. He. A central limit theorem for descents of a Mallows permutation and its inverse. *Ann. Inst. Henri Poincaré Probab. Stat.*, 58(2):667–694, 2022.

[31] J. He, T. Müller, and T. W. Verstraaten. Cycles in Mallows random permutations. Preprint, available from `https://arxiv.org/abs/2201.11610`, 2022.

[32] P. Heinig, T. Müller, M. Noy, and A. Taraz. Logical limit laws for minor-closed classes of graphs. *Journal of Combinatorial Theory, Series B*, 130:158–206, 2018.

[33] W. Hodges. *A shorter model theory.* Cambridge University Press, Cambridge, 1997.

[34] A. E. Holroyd, T. Hutchcroft, and A. Levy. Mallows permutations and finite dependence. *Ann. Probab.*, 48(1):343–379, 2020.

[35] N. Immerman. *Descriptive complexity.* Graduate Texts in Computer Science. Springer-Verlag, New York, 1999.

[36] S. Janson, T. Łuczak, and A. Rucinski. *Random graphs.* Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, New York, 2000.

[37] K. Jin. The length of the longest common subsequence of two independent Mallows permutations. *Ann. Appl. Probab.*, 29(3):1311–1355, 2019.

[38] D. E. Knuth. Mathematics and computer science: Coping with finiteness. *Science*, 194(4271):1235–1242, 1976.

[39] V. Koponen. Random graphs with bounded maximum degree: asymptotic structure and a logical limit law. *Discrete Mathematics & Theoretical Computer Science*, 14, 2012.

[40] D. Král and O. Pikhurko. Quasirandom permutations are characterized by 4-point densities. *Geometric and Functional Analysis*, 23(2):570–579, 2013.

[41] D. A. Levin and Y. Peres. *Markov chains and mixing times.* American Mathematical Society, Providence, RI, 2017. Second edition, With contributions by Elizabeth L. Wilmer, With a chapter on “Coupling from the past” by James G. Propp and David B. Wilson.

[42] L. Libkin. *Elements of finite model theory.* Texts in Theoretical Computer Science. An EATCS Series. Springer-Verlag, Berlin, 2004.

[43] J. F. Lynch. Convergence laws for random words. *Australas. J. Combin.*, 7:145–156, 1993.

[44] C. L. Mallows. Non-null ranking models. I. *Biometrika*, 44:114–130, 1957.

[45] Y. Malyshkin and M. Zhukovskii. Logical convergence laws via stochastic approximation and markov processes. *arXiv preprint arXiv:2210.13437*, 2022.

[46] C. Mueller and S. Starr. The length of the longest increasing subsequence of a random Mallows permutation. *J. Theoret. Probab.*, 26(2):514–540, 2013.

[47] T. Müller and M. Noy. The first order convergence law fails for random perfect graphs. *Random Structures & Algorithms*, 53(4):717–727, 2018.

[48] J. R. Norris. *Markov chains.* Cambridge series in statistical and probabilistic mathematics. Cambridge University Press, 1998.

[49] R. G. Pinsky. Permutations avoiding a pattern of length three under Mallows distributions. *Random Structures Algorithms*, 58(4):676–690, 2021.

[50] M. Podder and J. H. Spencer. First order probabilities for galton–watson trees. In *A journey through discrete mathematics*, pages 711–734. Springer, 2017.

[51] W. Rautenberg. *A concise introduction to mathematical logic.* Universitext. Springer, New York, second edition, 2006. With a foreword by Lev Beklemishev.

[52] S. Shelah and J. H. Spencer. Zero-one laws for sparse random graphs. *J. Amer. Math. Soc.*, 1(1):97–115, 1988.

[53] J. H. Spencer. Nonconvergence in the theory of random orders. *Order*, 7(4):341–348, 1990.

[54] J. H. Spencer. *The strange logic of random graphs*, volume 22 of *Algorithms and Combinatorics.* Springer-Verlag, Berlin, 2001.

[55] R. P. Stanley. *Enumerative combinatorics. Volume 1*, volume 49 of *Cambridge Studies in Advanced Mathematics.* Cambridge University Press, Cambridge, second edition, 2012.

[56] S. Starr. Thermodynamic limit for the Mallows model on S_n. *J. Math. Phys.*, 50(9):095208, 15, 2009.

[57] S. Starr and M. Walters. Phase uniqueness for the Mallows measure on permutations. *J. Math. Phys.*, 59(6):063301, 28, 2018.

[58] W. Tang. Mallows ranking models: maximum likelihood estimate and regeneration. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*, volume 97 of *Proceedings of Machine Learning Research*, pages 6125–6134. PMLR, 2019.

[59] P. Winkler. Random orders. *Order*, 1(4):317–331, 1985.

[60] T. Yanagimoto. On measures of association and a related problem. *Annals of the Institute of Statistical Mathematics*, 22(1):57–63, 1970.