Abstract

A random permutation $\Pi_n$ of $\{1,\ldots,n\}$ follows the Mallows($n,q$) distribution with parameter $q > 0$ if $\Pr(\Pi_n = \pi)$ is proportional to $q^{\inv(\pi)}$ for all $\pi$. Here $\inv(\pi) := |\{i < j : \pi(i) > \pi(j)\}|$ denotes the number of inversions of $\pi$. 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 $\phi$ in the language, the probability $\Pr(\Pi_n \text{ satisfies } \phi)$ 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 [16] 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 $[n] := \{1,\ldots,n\}$ 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 $\inv(\pi)$ the number of inversions of a permutation $\pi$.

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

$$\Pr(\Pi_n = \pi) = \frac{q^{\inv(\pi)}}{\sum_{\sigma \in S_n} q^{\inv(\sigma)}}. \quad (1)$$

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 [48] 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 [8, 19], finitely dependent colorings of the integers [36], stable matchings [5], random binary search trees [1], learning theory [13, 64], $q$-analogs of exchangeability [26, 27], determinantal point processes [11], statistical physics [62, 63], genomics [23] and random graphs with tunable properties [21].
A wide range of properties of the Mallows distribution has been investigated, including pattern avoidance \[17, 18, 54\], the number of descents \[32\], the longest monotone subsequence \[7, 9, 50\], the longest common subsequence of two Mallows permutations \[39\] and the cycle structure \[25, 33\].

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 \textit{convergence law} holds with respect to some fixed logical language describing permutations if the limit \(\lim_{n \to \infty} \mathbb{P} (\Pi_n \models \varphi) \) exists, for every sentence \(\varphi\) in the language. Here and in the rest of the paper the notation \(\pi \models \varphi\) denotes that the sentence \(\varphi\) holds for the permutation \(\pi\). If this limit is \(0\) or \(1\) for all such \(\varphi\) then we say that the \textit{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). We give here 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 \(\land, \lor, \neg, \ldots\), 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 \(\pi\) 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 \(\pi(1), \ldots, \pi(n)\). 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

\[ \exists x, y, z: (x <_1 y) \land (y <_1 z) \land (z <_2 x) \land (x <_2 y). \]

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

\[ \exists x, y: (x <_1 y) \land (x <_2 y) \land (\forall z: (z <_1 y) \rightarrow (z = x)). \]

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 \[24\]. 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 \[12\] 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 \textit{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.

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 \[60\]. Examples of structures for which the first order properties have been studied include the Erdős-Rényi random graph (see e.g. \[38\],\[60\]), Galton-Watson trees \[56\], bounded degree graphs \[42\], random graphs from minor-closed classes \[34\], random perfect graphs \[51\], random regular graphs \[31\] and the very recent result on bounded degree uniform attachment graphs \[49\].

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 \[65\] 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 \(i = 1,\ldots,k\). Notice that every sentence expressible with the single order \(<\) in a two dimensional random order is expressible in TOTO; each of instance of \((x < y)\) we replace by \((x <_1 y) \land (x <_2 y)\). Non-convergence was proven for two dimensional random orders in \[59\] by constructing a first–order sentence \(\varphi\) for which no limit probability exists; and this yields a constructive proof of non-convergence also in random permutations (adding to the earlier non-constructive proof \[24\]). Many properties of random orders are known \[4, 10\], see \[14\] and \[15\] for surveys of different models of random partial orders and also their relation to theories of spacetime \[15\].

An appeal of TOTO is it allows one to express pattern containment (though not counting of patterns above a constant). 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 \(\tau = 1 - \frac{2 \text{inv}(\pi)}{n(n-1)}\) and indeed there is a test for independence of \(X\) and \(Y\) using only counts of permutations of length 4 \[43, 66\]. See \[22\] 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,\[16\]] For \(q = 1\) the Mallows\((n,q)\) distribution satisfies the convergence law but not the zero–one law;

(ii) If \(q < 1\) is fixed then Mallows\((n,q)\) satisfies the zero–one law;
(iii) If $q > 1$ is fixed then Mallows($n, q$) does not satisfy the convergence law.

(Part (i) of Theorem 1.1 was already shown by Compton [16] in 1989, who in fact showed 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 a result of Foy and Woods [24] for uniform permutations (the case $q = 1$).

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 $\varphi$ 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 $n \to \infty$.

2 Overview of proof methods

**Proof Sketch of Theorem 1.1:** Theorem 1.1 concerns TOOB and as such we consider the logic containing a single binary relation symbol $R$. 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 $\pi_n$). Secondly, it follows from the Hanf-Locality Theorem for bounded degree structures (Theorem 4.1) 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 $\varphi$ 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 $\varphi$ 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 [33] 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 $\frac{1}{\sqrt{n}}(C_1(\Pi_n) - m_1 n, \ldots, C_t(\Pi_n) - m_t n)$ converges in distribution to a multivariate normal with zero mean. This implies in particular that $\Pi_n$ will contain more than $h$ cycles of each of the lengths $1, \ldots, h$ giving the zero–one–law for $0 < q < 1$.

For fixed $q > 1$, another result (Theorem 3.12 from [35]) implies the existence of a vector $(c_1, \ldots, c_t)$ such that for $\Pi_n \sim$ Mallows($n, q$) the quantity $\mathbb{P}(C_1(\Pi_n) = c_1, \ldots, C_t(\Pi_n) = c_t)$ 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 [46] to show a convergence law for random strings having certain letter distributions. We consider some TOTO sentence $\varphi$ having quantifier depth $d$.

Writing $\equiv_d$ if $\pi$ and $\sigma$ 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 3.20). We now rely on a sampling procedure for Mallows($n, q$) distributed permutations. The details of the
construction will be given in Section 3.2.2 it suffices to know that from a sequence \( Z_1, Z_2, \ldots Z_n \) of independent Geo(1 – q) random variables we may construct \( \Pi_n \) by first determining the image of 1 under \( \Pi_n \) using \( Z_1 \), then the image of 2 using \( Z_1, Z_2 \) and so on. Taking a dynamic viewpoint, we can imagine exposing the values \( Z_1, Z_2, \ldots \) one by one and following the sequence of permutations that arises. Forgoing 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 divide the state space into finitely many classes, each an irreducible aperiodic and positive recurrent, and that the chain will hit such a class a.a.s. Then standard results on the convergence of Markov chains will give the convergence of the quantify \( P(\Pi_n \models \varphi) \) as \( n \to \infty \).

**Proof Sketch of Theorem 1.2 Part (ii)** The proof of Theorem 1.2 Part (ii) proceeds in several steps. First we determine a sentence \( \varphi \) such that for \( \Pi_n \sim \text{Mallows}(n, 1) \) we have

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

(2)

for infinitely many \( n \). We need some additional technical conditions on the values of \( n \) such that the above is satisfied but we do not mention them here. We note that this is an explicit version of a result by Foy and Woods given in [22] (and which differs from the construction by Spencer in [59]). We show this result by associating pairs of intervals in \( \Pi_n \sim \text{Mallows}(n, 1) \) 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 [58] 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 \( P(\Pi_n \models \varphi) \) on the parity of \( \log^* \log^* n \).

We then show that if \( q = 1 - O(n^{-4}) \) 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 twice more but now we will have to work a little harder each time.

If \( j \in [n] \) and \( \Pi_n \sim \text{Mallows}(n, q) \) then \( \Pi_n \), the permutation of \( [j] \) inheriting the order of \( \Pi_n|_j \) follows a Mallows(\( j, q \)) distribution. We consider now \( q = 1 - O(1/n) \) and define \( J_1(\pi) \) as the smallest \( j \in [n] \) for which there exists an \( i \in [j] \) satisfying \( \pi(i) + 1 \in \pi([j]) \). For a uniform permutation \( \Pi_n \in S_n \) we have \( E[|i : \{\Pi_n(i), \Pi_n(i) + 1 \} \subseteq \Pi_n([j])|] \approx j \cdot j/n \), and it turns out the \( q = 1 - O(1/n) \) case is close enough to the uniform case that we will have \( J_1(\Pi_n) = \Theta(\sqrt{n}) \). Then we define \( K_1(\Pi_n) := J_1(\Pi_n|_j) \) where \( K_1(\Pi_n) = \Theta(n^{1/4}) \) will be shown to hold with probability \( 1 - \varepsilon \) for some \( \varepsilon > 0 \). See Figure 7.1 Now \( \Pi_{K_1(\Pi_n)} \) is roughly distributed as Mallows(\( \Theta(n^{1/4}) \), \( q \)) with \( q = 1 - O(1/n) = 1 - O(K_1^{-1}) \), where we must take care of some dependency between \( K_1 \) and the induced permutation \( \Pi_n|_{K_1(\Pi_n)} \). We then apply the previous result to show non–convergence for the case \( q = 1 - O(1/n) \).

We subsequently extend the result once more, where we now set \( N = \Pi_n^{-1}(1) + 1 \). By a result of Bhatnagar and Peled [9] (Theorem 3.15) we will have \( N = \Theta(1/(1-q)) \) so that also \( q = 1 - \Theta(1/N) \). Moreover, we show that \( \Pi_{N-1} \) follows a Mallows(\( N-1, q \)) distribution. Here we will additionally have to handle the case \( q > 1 \) and \( q < 1 \) separately at first. However, defining the sentence \( \rho \) saying that \( \Pi_n(1) > \Pi_n(n) \) provides us with a mechanism to ‘detect’ when \( q > 1 \), in which case \( \rho \) will likely be satisfied, and when \( q < 1 \) in which case \( \rho \) is likely not satisfied. This allows us to give a single (universal) non–convergent sentence \( \varphi \) such that \( P(\Pi_n \models \varphi) \) does not converge for \( \Pi_n \sim \text{Mallows}(n, q) \) for all \( 1 - 1/\log^* n < q < 1 + 1/\log^* n \).
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 here some mostly well–known results from probability theory.

**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 \( S_n := X_1 + \ldots + X_n \) and \( \mu := \mathbb{E}S_n \). Let \( 0 < \varepsilon < 1 \), then

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

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

Panconesi and Srinivasan in [53] developed a generalization of Chernoff’s bound that omits the condition that the \( X_i \) be independent.

**Theorem 3.2.** Let \( X_1, \ldots, X_n \) be random variables taking values in \( \{0, 1\} \) such that, for some \( 0 \leq \delta \leq 1 \), we have that, for every subset \( S \subseteq [n] \), \( \mathbb{P} ( \bigwedge_{i \in S} X_i = 1 ) \leq \delta^{|S|} \). Then, for any \( 0 \leq \delta \leq \gamma \leq 1 \)

\[
\mathbb{P} \left( \sum_{i=1}^{n} X_i \geq \gamma n \right) \leq e^{-2n(\gamma - \delta)^2}.
\]

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

**Lemma 3.3.** Let \( X \sim \text{Bi}(n, k/n) \) for \( 0 \leq k \leq n \). Then \( \mathbb{P} (X = k) \geq 1/(n+1) \).

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

\[
\frac{\mathbb{P} (X = j)}{\mathbb{P} (X = j - 1)} = \frac{(n)_j (1-p)^{n-j} p^j}{(n-1)_j (1-p)^{n-j+1} p^{j-1}} = \frac{(n-j+1)p}{j(1-p)}.
\]

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

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

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

Given two discrete probability distributions \( \mu_1 \) and \( \mu_2 \) on a countable set \( \Omega \), their total variation distance is defined as

\[
d_{\text{TV}}(\mu_1, \mu_2) = \max_{A \subseteq \Omega} |\mu_1(A) - \mu_2(A)|.
\]

This can be expressed alternatively as

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

(See for instance Proposition 4.2 in [44] for a proof.) As is common, we will interchangeably use the notation \( d_{\text{TV}}(X, Y) := d_{\text{TV}}(\mu, \nu) \) if \( X \sim \mu \) and \( Y \sim \nu \).

A **coupling** of two probability measures \( \mu, \nu \) is a joint probability measure for a pair of random variables \((X, Y)\) satisfying \( X \overset{d}{=} \mu, Y \overset{d}{=} \nu \). We will also speak of a coupling of \( X, Y \) as being a probability space for \((X', Y')\) with \( X' \overset{d}{=} X, Y' \overset{d}{=} Y \).
Lemma 3.4. Let \( \mu \) and \( \nu \) be two probability distributions on the same countable set \( \Omega \). Then
\[
\text{d}_{\text{TV}}(\mu, \nu) = \inf\{\mathbb{P}(X \neq Y) : (X, Y) \text{ is a coupling of } \mu \text{ and } \nu\}.
\]
There is a coupling that attains this infimum.

(See for instance [44], Proposition 4.7 and Remark 4.8 for a proof.)

For \( n \in \mathbb{N} \) and \( p \in (-\infty, 0) \cup (0, 1) \) we define the TruncGeo\((n, p)\) distribution where \( Y \sim \text{TruncGeo}(n, p) \) means
\[
\mathbb{P}(Y = k) = \frac{p(1-p)^{k-1}}{1-(1-p)^n}, \quad k \in \{0, \ldots, n\}.
\]
Observe that if \( p \in (0, 1) \) then, setting \( X \sim \text{Geo}(p) \), the probability mass in (4) is exactly equal to \( \mathbb{P}(X = k \mid X \leq n) \). Therefore we shall refer to the TruncGeo\((n, p)\) as the truncated geometric distribution. But we stress that the TruncGeo\((n, p)\) is a valid probability distribution for any \( p \in (-\infty, 0) \cup (0, 1) \).

3.1.1 Markov chains

We give here a brief overview of some of the concepts related to Markov chains which will be used in Section 5. The following definitions and results are widely known and can be found in most books on Markov chains or more general stochastic processes, see for instance [30], [44], or [52].

Let \( X_n \) be a Markov chain with state space \( S \). Given \( i, j \in S \), if \( \mathbb{P}(X_n = j \mid X_0 = i) > 0 \) for some \( n \in \mathbb{Z}_{\geq 0} \) we say that \( j \) is reachable from \( i \). We denote this also by \( i \rightarrow j \). If \( i \rightarrow j \) and \( j \rightarrow i \) 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 \( i \in S \) 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. If \( X_0 = i \) for some \( i \in S \) then we write \( \mathbb{P}_i(\cdot) \) for the corresponding probability and \( \mathbb{E}_i(\cdot) \) for the expectation. If \( X_0 \) is distributed according to some distribution \( \mu_0 \) over \( S \) then we write \( \mathbb{P}_{\mu_0} \) and \( \mathbb{E}_{\mu_0} \).

For any \( i \in S \) we define \( \tau_i \) to be the first time that the Markov chain is in state \( i \), and we define \( \tau_i^+ > 0 \) to be the first time after zero that the Markov chain is in state \( i \), given that \( X_0 = i \). 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 \( S \) 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 \( k \geq 0 \).

Proofs for the following two results are given in Theorem 4 and Lemma 5, respectively, in Section 6.3 of [30].

Theorem 3.5. The state space \( S \) of a Markov chain can be partitioned uniquely as
\[
S = T \cup C_1 \cup C_2 \cup \ldots
\]
where \( T \) is the set of transient states, and the \( C_i \) are irreducible closed sets of recurrent states.

Lemma 3.6. If \( S \) is finite, then at least one state is recurrent and all recurrent states are positive recurrent.

Lemma 3.7. If \( S \) is finite then with probability 1 the chain \( X_n \) will be in a recurrent state for some \( n \geq 0 \).
Proof. Let $S = T \cup C_1 \cup C_2 \cup \ldots$ be as in Theorem 3.5. For each of the finitely many $i \in T$ there is a positive probability that the chain moves from $i$ to a state in $S \setminus T$ in at most $|T|$ steps. Then $P(X_n \in T) \leq Ke^n$ for some constants $K > 0$ and $0 < c < 1$. By the Borel–Cantelli Lemma, $X_n$ is in $T$ for only finitely many $n$, so that it must visit a recurrent state. ■

The following theorem is the Markov chain convergence theorem, see e.g. [20] Theorem 7.6.4 for a proof (note that in [20], what we call irreducibility is called strong irreducibility).

**Theorem 3.8 (Markov chain convergence theorem).** Let $P$ be a Markov kernel on a discrete (countable) state space $S$. Assume that $P$ is irreducible, aperiodic and positive recurrent. There exists a unique invariant probability measure $\pi$ over $S$ such that for every probability measure $\mu_0$ over $S$

$$\lim_{n \to \infty} d_{TV}(\mu_0P^n, \pi) = 0.$$  

The following result 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 state and prove here a result that we will need in Section 5.

**Lemma 3.9.** Let $X_n$ be a Markov chain on the countable state space $S$, and let $M$ be a communicating class of $X_n$ such that the chain restricted to $M$ is aperiodic, irreducible and positive recurrent. Then the limit

$$\lim_{n \to \infty} P(X_n \in A \mid \tau_M < \infty)$$

exists for all $A \subseteq M$.

Proof. We have

$$P(X_n \in A \mid \tau_M < \infty) = \sum_{j=0}^{\infty} P(X_n \in A \mid \tau_M = j) P(\tau_M = j \mid \tau_M < \infty).$$

Conditioning on $\tau_M = j$ induces a probability distribution $\nu_j$ over the states in $M$ where for $B \subseteq M$ we have $\nu_j(B) = P(X_j \in B \mid \tau_M = j)$. The Markov chain $X_n$ restricted to $M$ is assumed aperiodic, irreducible and positive recurrent. So by Theorem 3.8 there exits a unique distribution $\pi$ over $M$ such that for all $j \in \mathbb{Z} \geq 0$

$$\lim_{n \to \infty} P(X_n \in A \mid \tau_M = j) = \lim_{n \to \infty} P_{\nu_j}(X_{n-j} \in A) = \pi(A).$$  \hspace{1cm} (5)

Now, let $\varepsilon > 0$ be arbitrary, and let $j_0 \in \mathbb{Z} \geq 0$ be such that $P(\tau_M > j_0 \mid \tau_M < \infty) < \varepsilon/2$. Then

$$\left| P(X_n \in A \mid \tau_M < \infty) - \pi(A) \right| \leq \sum_{j=0}^{\infty} \left| P(X_n \in A \mid \tau_M = j) - \pi(A) \right| P(\tau_M = j \mid \tau_M < \infty)$$

$$< \left( 1 - \frac{\varepsilon}{2} \right) \max_{0 \leq j \leq j_0} \left| P(X_n \in A \mid \tau_M = j) - \pi(A) \right| + \frac{\varepsilon}{2}.$$

The maximum above is taken over a finite set, so by (5) there exists an $n_0$ such that $n \geq n_0$ implies that the above is at most $\varepsilon$. Thus,

$$\lim_{n \to \infty} P(X_n \in A \mid \tau_M < \infty) = \pi(A).$$  ■
3.2 Mallows permutations

It is a standard result in enumerative combinatorics (see Corollary 1.3.13 in [61]) that for \( q \neq 1 \) the denominator in the definition of the Mallows distribution (1) satisfies

\[
\sum_{\sigma \in S_n} q^{\text{inv} (\sigma)} = \prod_{i=1}^{n} \frac{1 - q^i}{1 - q}
\]

Moreover, the two limit distributions above are distinct for all \( q > 1 \).

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

\[
\text{inv}(r_n \circ \pi) = \text{inv}(\pi \circ r_n) = \binom{n}{2} - \text{inv}(\pi),
\]

\[
\text{inv}(\pi^{-1}) = \text{inv}(\pi).
\]

That is, if \( \Pi_n \sim \text{Mallows}(n, q) \), then \( r_n \circ \Pi_n \equiv \Pi_n \circ r_n \equiv \text{Mallows}(n, 1/q) \) and \( \Pi_n^{-1} \equiv \text{Mallows}(n, q) \).

For \( X = \{x_1, \ldots, x_n\} \) 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 \( x = (x_1, \ldots, x_n) \) a sequence of distinct numbers, let us write

\[
\text{rk}(x_1, \ldots, x_n) := (\text{rk}(x_1, \{x_1, \ldots, x_n\}), \ldots, \text{rk}(x_n, \{x_1, \ldots, x_n\})).
\]

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

**Lemma 3.10** ([9], Corollary 2.7). Let \((i, i+1, \ldots, i+m-1) \in [n]\) be a sequence of consecutive elements. If \( \Pi_n \sim \text{Mallows}(n, q) \)

\[
\text{rk}(\Pi_n(i), \ldots, \Pi_n(i+m-1)) \equiv \text{Mallows}(m, q).
\]

In [27] Gnedin and Olshansky introduced a random bijection \( \Sigma : \mathbb{Z} \to \mathbb{Z} \) which is a natural extension to the finite Mallows\((n, q)\) model with \( 0 < q < 1 \). We denote this distribution by 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 [33], Jimmy He together with the first and last author of the current paper studied the limit behavior of the cycles counts of 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 \( \pi \) for \( i \geq 1 \). We will use the following results:

**Theorem 3.11** ([33], Theorem 1.1). Fix \( 0 < q < 1 \) and let \( \Pi_n \sim \text{Mallows}(n, q) \). There exist positive constants \( m_1, m_2, \ldots \) and an infinite matrix \( P \in \mathbb{R}^{N \times N} \) such that for all \( \ell \geq 1 \) we have

\[
\frac{1}{\sqrt{n}} (C_1(\Pi_n) - m_1 n, \ldots, C_\ell(\Pi_n) - m_\ell n) \xrightarrow{d} N_\ell(0, P_\ell),
\]

where \( N_\ell(\cdot, \cdot) \) denotes the \( \ell \)-dimensional multivariate normal distribution and \( P_\ell \) 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 3.12** ([33], Theorem 1.3). Let \( q > 1 \) and \( \Pi_n \sim \text{Mallows}(n, q) \) and \( \Sigma \sim \text{Mallows}(\mathbb{Z}, 1/q) \). We have

\[
(C_1(\Pi_{2n+1}), C_3(\Pi_{2n+1}), \ldots) \xrightarrow{d} (C_1(r \circ \Sigma), C_3(r \circ \Sigma), \ldots)
\]

and

\[
(C_1(\Pi_{2n}), C_3(\Pi_{2n}), \ldots) \xrightarrow{d} (C_1(\rho \circ \Sigma), C_3(\rho \circ \Sigma), \ldots).
\]

Moreover, the two limit distributions above are distinct for all \( q > 1 \).
Remark 3.13. The convergence in Theorem 3.12 should be understood to mean that for all sequences \(c_1, c_3, \ldots\) with only finitely many nonzero values we have

\[
\mathbb{P} \left( \left( C_1(\Pi_{2n+1}), C_3(\Pi_{2n+1}), \ldots \right) = (c_1, c_3, \ldots) \right) \xrightarrow[n \to \infty]{} \mathbb{P} \left( (C_1(\sigma \circ \Sigma), C_3(\sigma \circ \Sigma), \ldots) = (c_1, c_3, \ldots) \right).
\]

For sequences \(c_1, c_3, \ldots\) with infinitely many nonzero values the limit of the left hand side above is zero, as is the limit of the right-hand side.

When \(q = 1\) the Mallows distribution simply samples a permutation of \(1, \ldots, n\) uniformly at random. A classical result going back to Gontcharoff [28] and Kolchin [41] states that in this case, for every fixed \(\ell\):

\[
(C_1(\Pi_n), \ldots, C_\ell(\Pi_n)) \xrightarrow{d} (X_1, \ldots, X_\ell),
\]

where \(X_1, \ldots, X_\ell\) are independent random variables with \(X_i \sim \text{Po}(1/i)\).

Arratia and Tavaré extend this result to convergence of the vector \((C_1(\Pi_n), \ldots, C_b(\Pi_n))\) in the case that \(b = b(n)\) is such that \(b/n \to 0\).

**Theorem 3.14** ([6], Theorem 2). Let \(\Pi_n\) be selected uniformly from \(S_n\), and let \(X_1, \ldots, X_n\) be a sequence of independent random variables where \(X_i \sim \text{Po}(1/i)\) for \(i = 1, \ldots, n\). Let \(b = b(n) : \mathbb{N} \to \mathbb{N}\) be a function. Then

\[
d_{\text{TV}}((C_1(\Pi_n), \ldots, C_b(\Pi_n)), (X_1, \ldots, X_b)) \xrightarrow{n \to \infty} 0
\]

if and only if \(b/n \to 0\).

The following result is due to Bhatnagar and Peled.

**Theorem 3.15** ([9], Theorem 1.1). For all \(0 < q < 1\), and integer \(1 \leq i \leq n\), if \(\Pi_n \sim \text{Mallows}(n, q)\) then

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

### 3.2.1 Construction using truncated geometric random variables

The following method of sampling a Mallows\((n, q)\) distributed permutation for \(q < 1\) goes back to the work of Mallows [48]. Let \(Z_1, \ldots, Z_n\) be independent with \(Z_i \sim \text{TruncGeo}(n - i + 1, 1 – q)\). We now set

\[
\Pi_n(1) = Z_1,
\]

\[
\Pi_n(i) = \text{the } Z_i\text{-th smallest number in the set } [n] \setminus \{\Pi_n(1), \ldots, \Pi_n(i-1)\}, \quad \text{for } 1 < i \leq n.
\]

Then \(\Pi_n\) 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 \(S_n\).

### 3.2.2 Construction using geometric random variables

In [55] Pitman and Tang give a general construction for a random bijection \(\Pi : \mathbb{N} \to \mathbb{N}\). For a probability distribution \(p\) on \(\mathbb{N}\) with \(p(1) > 0\) they call a random permutation \(\Pi\) of \(\mathbb{N}\) a \(p\)-shifted random permutation of \(\mathbb{N}\) if for \(Z_1, Z_2, \ldots\) i.i.d. sampled according to \(p\), we set

\[
\Pi(1) := Z_1
\]

\[
\Pi(i) := \text{the } Z_i\text{-th smallest element in the set } \mathbb{N} \setminus \{\Pi(1), \ldots, \Pi(i-1)\}, \quad \text{for } i > 1.
\]
Remark 3.17. in [26]. of variables as described, then \{π\} will be examined in more detail in Section 5.

Proposition 3.16 ([55], Proposition 1.4). For any probability distribution \( p \) on \( \mathbb{N} \) with \( p(1) > 0 \), and \( Π \) a \( p \)-shifted random permutation of \( \mathbb{N} \), if \( \sum_{i} ip(i) < \infty \) then \( Π \) is positive recurrent.

Here a random permutation is defined to be positive recurrent if with probability 1 there are infinitely many \( T \in \mathbb{N} \) such that \( |Π[T]| = |T| \). Such \( T \) are called regeneration times, they will be examined in more detail in Section 5.

If \( p \) is a Geo(1−\( q \)) distribution we denote the distribution of a \( p \)-shifted random permutation of \( \mathbb{N} \) by Mallows(\( \mathbb{N}, q \)). The Mallows(\( \mathbb{N}, q \)) distribution was first studied by Gneden and Olshanki in [26].

Remark 3.17. To any \( π \in S_n \) we can uniquely associate a sequence \( z_1, \ldots, z_n \) such that if \( Π \sim \text{Mallows}(\mathbb{N}, q) \) is constructed from an i.i.d. sequence \( Z_1, Z_2, \ldots \) of Geo(1−\( q \)) random variables as described, then \( \{Π \upharpoonright \{n\} = π\} = \{Z_1 = z_1, \ldots, Z_n = z_n\} \). In particular, for any \( π \in S_n \) we have \( \mathbb{P}(Π \upharpoonright \{n\} = π) > 0 \).

The following result was obtained by Basu and Bhatnagar ([7], Lemma 2.1) and independently by Crane and DeSalvo ([17], Lemma 5.2):

Lemma 3.18. For \( q < 1 \), if \( Π \sim \text{Mallows}(\mathbb{N}, q) \) and \( Π_n := \text{rk}(Π(1), \ldots, Π(n)) \) then \( Π_n \models \text{Mallows}(n, q) \).

3.3 First–order logic

Given a set of relation symbols \( R_1, \ldots, R_k \) with associated arities \( a_1, \ldots, a_k \), we define \( L_{R_1, \ldots, R_k} \) as the first–order language consisting of all first–order formulas built from the usual first–order connectives together with \( R_1, \ldots, R_k \). As an example, for a single binary relation \( R \), we may consider the formula \( ϕ_0(x) \in L_R \) defined as

\[
ϕ_0(x) := R(x, x) \land \neg\exists y : (\neg(x = y) \land R(y, y)).
\]  

A free variable of a formula \( ϕ \) is a variable that occurs outside of the scope of a quantifier. We write \( \text{Free}(ϕ) \) for the set of free variables in \( ϕ \), and say that a formula with no free variables is a sentence. For \( ϕ_0 \) as in (8) we have \( \text{Free}(ϕ_0) = \{x\} \). Given \( R_1, \ldots, R_k \) with arities \( a_1, \ldots, a_k \), a \((R_1, \ldots, R_k)\)-structure \( \mathfrak{A} \) is a tuple \((A, R^A_1, \ldots, R^A_k)\) where \( A \) is a set, called the domain of \( \mathfrak{A} \), and each \( R^A_i \) is a relation of arity \( a_i \) over \( A \). We will routinely write simply \( R_j \) for \( R^A_j \) and talk about structures instead of \((R_1, \ldots, R_k)\)-structures. If a structure \( \mathfrak{A} \) satisfies a sentence \( ϕ \) then we write \( \mathfrak{A} \models ϕ \). Similarly, if \( i_1, \ldots, i_k \) are elements in the domain of \( \mathfrak{A} \) and \( ϕ(x_1, \ldots, x_k) \) is a formula with \( \text{Free}(ϕ) = \{x_1, \ldots, x_k\} \) then we write \((\mathfrak{A}, i_1, \ldots, i_k) \models ϕ \) if \( ϕ \) is satisfied by \( \mathfrak{A} \) under the assignment \( x_j \mapsto i_j \). We have for instance that \((\mathfrak{A}, i) \models ϕ_0 \) precisely when the element \( x \) is the unique element satisfying \( R(x, x) \) in \( \mathfrak{A} \).

For a complete and rigorous definition of formulas, free variables and structures satisfying formulas, see for instance Chapter 1 of [37].

Two \((R_1, \ldots, R_k)\)-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

\[
(i_1, \ldots, i_k) \in R^A_j \iff (f(i_1), \ldots, f(i_k)) \in R^B_j \quad \text{for all } i_1, \ldots, i_k \in A \text{ and all } R_j.
\]

The following is a well–known result on isomorphic structures, see e.g. Theorem 3.4 of [37] for a proof.

Lemma 3.19. If \( \mathfrak{A} \) and \( \mathfrak{B} \) are isomorphic, then for all sentences \( ϕ \) we have \( \mathfrak{A} \models ϕ \) if and only if \( \mathfrak{B} \models ϕ \).

We recursively define the quantifier depth of a first–order sentence \( ϕ \), denoted \( D(ϕ) \), as follows:
• For any atomic formula $\varphi$ we set $D(\varphi) = 0$;
• For any formula $\varphi$ we set $D(\lnot \varphi) = D(\varphi)$;
• For any two formulas $\varphi, \psi$ we set $D(\varphi \land \psi) = D(\varphi \lor \psi) = \max\{D(\varphi), D(\psi)\}$;
• For a formula $\psi$ with at least one free variable we set $D(\exists x(\psi)) = D(\forall x(\psi)) = D(\psi(x)) + 1$.

If two structures $\mathfrak{A}$ and $\mathfrak{B}$ agree on all sentences $\varphi$ with $D(\varphi) \leq d$ we write $\mathfrak{A} \equiv_d \mathfrak{B}$ and say that $\mathfrak{A}$ and $\mathfrak{B}$ are $d$–equivalent. The relation $\equiv_d$ is an equivalence relation. The following lemma will be crucial in the proof of Theorem 5. A proof can be found for instance following Corollary 3.16 in [15].

Lemma 3.20. The equivalence relation $\equiv_d$ has only finitely many equivalence classes.

Given two structures $\mathfrak{A}, \mathfrak{B}$ with disjoint domains $A$ and $B$ we define $\mathfrak{A} \uplus \mathfrak{B}$ as the structure $\mathfrak{C}$ with domain $C = A \uplus B$ such that

$$R^\mathfrak{C} = R^\mathfrak{A} \uplus R^\mathfrak{B},$$

for all relation symbols $R$.

In a more general setting where we allow also constants in our first–order language, we need to be careful with the above as we would have two conflicting candidates in $\mathfrak{C}$ for each constant. We will not have this problem as we do now use any constant symbols.

For a $(R_1, \ldots, R_k)$–structure $\mathfrak{A}$ we define $\text{Th}(\mathfrak{A}) = \{ \varphi \in L_{R_1, \ldots, R_k} : \mathfrak{A} \models \varphi \}$, often referred to as the complete theory of $\mathfrak{A}$. For $d \geq 0$ we define $\text{Th}^d(\mathfrak{A})$ as the set of all sentences $\varphi$ in $\text{Th}(\mathfrak{A})$ with quantifier depth at most $d$. The following theorem is due to Ehrenfeucht, Fraïssé, Feferman and Vaught, the statement as given is Theorem 1.5 in [17] which also includes a proof.

Theorem 3.21. $\text{Th}^d(\mathfrak{A} \uplus \mathfrak{B})$ is uniquely determined by $\text{Th}^d(\mathfrak{A})$ and $\text{Th}^d(\mathfrak{B})$.

Recall that we write $\mathfrak{A} \equiv_d \mathfrak{B}$ when $\mathfrak{A}$ and $\mathfrak{B}$ satisfy exactly the same sentences of quantifier depth at most $d$, that is, when $\text{Th}^d(\mathfrak{A}) = \text{Th}^d(\mathfrak{B})$.

Corollary 3.22. If $\mathfrak{A}_1, \ldots, \mathfrak{A}_\ell$ is a sequence of structures with mutually disjoint domains, and $\mathfrak{B}_1, \ldots, \mathfrak{B}_\ell$ is another such a sequence satisfying $\mathfrak{A}_i \equiv_d \mathfrak{B}_i$ for $i = 1, \ldots, \ell$, then

$$\mathfrak{A}_1 \uplus \ldots \uplus \mathfrak{A}_\ell \equiv_d \mathfrak{B}_1 \uplus \ldots \uplus \mathfrak{B}_\ell.$$

Permutations $\pi \in S_n$ and $\sigma \in S_m$ do not have disjoint domains, so we cannot talk about $\pi \uplus \sigma$ as defined above. We instead define $\pi \oplus \sigma$ to be the permutation $\kappa \in S_{n+m}$ satisfying

$$\kappa(i) = \begin{cases} 
\pi(i) & \text{if } i \leq n, \\
\sigma(i) + n & \text{if } i > n.
\end{cases} \quad (9)$$

The operator $\oplus$ is associative. Given a permutation $\sigma \in S_m$ and $n \in \mathbb{Z}_{\geq 0}$ we define the bijection $\sigma^n$ on $\{n + 1, \ldots, n + m\}$ by $\sigma^n(n + i) = \sigma(i) + n$. Then $\pi \oplus \sigma = \pi \uplus \sigma^n$ and $\sigma^n \equiv_d \sigma$ for all $d \geq 0$. Thus Corollary 3.22 can be stated in terms of permutations.

Corollary 3.23. For $t \geq 1$, if $\pi_1, \ldots, \pi_t, \sigma_1, \ldots, \sigma_t$ are permutations such that $\pi_i \equiv_d \sigma_i$ for $i = 1, \ldots, t$, then

$$\pi_1 \oplus \ldots \oplus \pi_t \equiv_d \sigma_1 \oplus \ldots \oplus \sigma_t.$$

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

$$\xi(R) := \forall x ((\exists y : R(x, y)) \land (\exists ! y : R(y, x))),$$

where we use the shorthand $\exists x$ to mean that there exists a unique such $x$, i.e., $\exists x : \psi(x) := \exists ! x : (\psi(x) \land \gamma y (\psi(y) \rightarrow y = x))$. 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 $\varphi := \exists R : \xi(R)$. A finite structure satisfies $\varphi$ if and only if its domain has even cardinality.
3.3.1 TOOB and TOTO

We now look more closely at TOOB and TOTO mentioned in the introduction.

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

We define $\text{TOTO} = \mathcal{L}_{<1, <2}$ where $<1$ and $<2$ are two binary relations. A permutation $\pi$ is then encoded as $([n], <^\pi_1, <^\pi_2)$ where $i <^\pi_j$ if and only $i < j$ in the usual ordering of $[n]$, and $i <^\pi_2 j$ 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 define a permutation in TOOB, respectively TOTO; indeed the interpretation of for instance $R$ may be such that it does not encode a bijection. Usually we identify a set of axioms that structures should satisfy to ensure such conditions, however 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 we will not need any axioms to ensure that $R$ or $<1, <2$ are as desired.

We define the TOTO formulas

$$\text{succ}_1(x, y) := (x <_1 y) \land \neg\exists w((x <_1 w) \land (w <_1 y)),$$

$$\text{succ}_2(x, y) := (x <_2 y) \land \neg\exists w((x <_2 w) \land (w <_2 y)).$$

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

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

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

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

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

We will denote $\mathcal{L}(V_{\text{TOTO}})$ by TOTO and $\mathcal{L}(V_{\text{TOOB}})$ by TOOB.

3.3.2 Relativizing sentences in TOTO

Recall the definition of $\text{rk}(x_1, \ldots, x_n)$ for a sequence $(x_1, \ldots, x_n)$ of distinct numbers as given in [7].

Lemma 3.24. Let $\varphi(\bar{x}) \in \text{TOTO}$ be a formula. There exists a formula $\varphi^\leq(y, \bar{z}) \in \text{TOTO}$ such that for all $j \in [n]$ and all $\bar{i} \in [j]^{|\text{Free}(\varphi)|}$ we have

$$(\pi, j, \bar{i}) \models \varphi^\leq(y, \bar{z}) \text{ if and only if } (\pi_j, \bar{i}) \models \varphi(\bar{x}),$$

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

Remark 3.25. We are not concerned with what happens in the case that $\bar{i} \notin [j]^{|\text{Free}(\varphi)|}$.

Proof. We proceed by induction on the production rules of formulas, starting with the atomic formulas. Every such an atomic formula is of the form $\varphi(\bar{x}) = R(x_1, x_2)$, for $R \in \{=, <_1, <_2\}$. We then define $\varphi^\leq(y, \bar{z})$ simply as $R(x_1, x_2)$. Two $i_1, i_2 \in [j]$ satisfy in $\pi$ exactly the same relations as in $\pi_j$ which finishes the base case. For the production rule $\varphi \to \neg \psi$ it suffices to define $\neg \varphi^\leq$ as $\neg(\varphi^\leq)$ as is straightforward to see. We define $\varphi^\leq \lor \psi^\leq$ as $\varphi^\leq \lor \psi^\leq$ and similarly
Now, $\pi$ in (10). Moreover, atomic formulas we let As in the proof Lemma 3.24, we use induction on the production rules of formulas. For Proof. Fix some $\phi$ formula, we define $\phi$ be the unique element such that $\phi \leq \phi$ and similarly for quantification over one of the other free variables $x_1, \ldots, x_{k-1}$. Here $x_k \leq 1 y$ is shorthand for $(x_k < 1 y) \lor (x_k = y)$. Then, for all $j$ we have by induction that $$(\pi, i_1, \ldots, i_{k-1}) \models \forall x_k \phi \iff (\pi, i_1, \ldots, i_{k-1}, i_k) \models \phi \quad \text{for all } i_k \in [j]$$ $$\iff (\pi, j, i_1, \ldots, i_k) \models \phi \leq \phi \quad \text{for all } i_k \in [j]$$ $$\iff (\pi, j, i_1, \ldots, i_k) \models \forall x_k ((x_k \leq 1 y) \to \phi \leq \phi(y, x_1, \ldots, x_k)).$$ The final case can be handled by e.g. $(\exists x \psi) \leq := (\forall x (\neg \psi)) \leq$.

We will call $\phi \leq$ the relativization of $\phi$. Let us remark that this is a specific application of a more general procedure known as the relativization of formulas, see e.g. Theorem 4.2.1 in [35].

Lemma 3.26. Let $\xi(x)$ be a formula with one free variable such that for some $\pi \in S_n$ there is a unique $j \in [n]$ with $\pi \models \xi[j]$. Let $\phi(x)$ be a formula with $k$ free variables. Then

$$(\pi, \xi) \models \exists z \left( \xi(z) \land \left( \bigwedge_{\ell=1}^k i_\ell \leq 1 z \right) \land \phi \leq \phi(z, x) \right)$$

if and only if $i_\ell \leq j$ for $\ell \in [k]$ and $(\pi_j, \xi) \models \phi(x)$ where $\pi_j = \text{rk}(\pi(1), \ldots, \pi(j))$.

Proof. Fix some $\pi \in S_n$, $i \in [n]^{\text{Free}(\phi)}$ and let $j$ be the unique element such that $(\pi, j) \models \xi$. Suppose first that (10) holds. Then certainly $i_\ell \leq j$ for $\ell \in [k]$ by the first two conjunctions in (10). Moreover, $\pi(j, i_1, \ldots, i_k) \models \phi \leq \phi$ so that by Lemma 3.24 we have $(\pi_j, \xi) \models \models \phi(x)$. Suppose now that $i_\ell \leq j$ for $\ell \in [k]$ and that $(\pi_j, \xi) \models \phi(x)$. By Lemma 3.24 this implies that $(\pi, j, \xi) \models \phi \leq \phi(y, x)$. Then clearly (10) is satisfied.

Lemma 3.27. Let $\phi$ be a TOTO formula with $k \geq 0$ free variables. There exists a TOTO formula $\phi^{\text{reverse}}$ such that for all $n \geq 1$, all $\pi \in S_n$ and all $\xi \in [n]^k$ we have

$$(\pi, \xi) \models \phi \iff (r_n \circ \pi, \xi) \models \phi^{\text{reverse}}.$$ 

Proof. As in the proof Lemma 3.24 we use induction on the production rules of formulas. For atomic formulas we let

$$(x = y)^{\text{reverse}} = (x = y),$$

$$(x <_1 y)^{\text{reverse}} = (x <_1 y),$$

$$(x <_2 y)^{\text{reverse}} = (y <_2 x).$$

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

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

The other two relations are straightforward to check. The result then follows by a simply induction argument on the production rules of the formulas. 

14
3.4 Two rapidly growing functions

The tower function is defined as

\[ T(n) = 2^{2^{\cdots^{2}}} \]

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

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

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. [29]. The tower function is sometimes known under the name tetration and the wowzer function under the name pentation. Knuth also describes these functions in [40] 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

\[ \log^{**} n := \min\{k \in \mathbb{Z}_{\geq 0} : W(k) \geq n\}. \quad (11) \]

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:

**Lemma 3.28.** 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 (11) we immediately obtain that \( W^{(i)}(x) < \log^{**} y \), and the result follows by induction. \( \square \)

Given a function \( f \) we use the usual notation \( f^{(1)} = f \) and \( f^{(t+1)} = f \circ f^{(t)} \) for \( t \geq 1 \).

**Lemma 3.29.** 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

\[ W(m) + 3 < W(m + 1) = T(W(m)). \quad (12) \]

But \( T(x) > x + 3 \) for all \( x > 2 \). As \( W(m) > 2 \) for \( m > 2 \), (12) holds. \( \square \)
4 Proof of Theorem 1.1

In this section we develop and apply the necessary tools to prove Theorem 1.1. In this section we are regarding TOTO, the first–order language with a single binary predicate.

We start by providing some more definitions and results from model theory. We follow the exposition in Sections 6.2 and 6.3 of [37].

For a structure $\mathfrak{A}$ 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 $(a,b) \in A^2$ such that $a$ and $b$ occur together in an element of at least one relation $R_{\mathfrak{A}}$. This graph may contain self loops. 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 $d$. In TOTO the degree of a permutation $\pi \in S_n$ is always equal to $n – 1$. Permutations in TOTO however all have degree at most 2, and this is the setting we are concerned with in the present section.

For a structure $\mathfrak{A}$ and $X \subseteq A$ we define $\mathfrak{A} \mid X$ to be the restriction of $\mathfrak{A}$ to $X$. That is $\mathfrak{A} \mid X$ has domain $X$ and for every relation symbol $R_i$, $R_{\mathfrak{A}}^{|X} = R^{|X}_{\mathfrak{A}} \cap X^{a_i}$, where $a_i$ is the arity of $R_i$. Note that for a permutation $\sigma$ and $X$ a subset of its domain, it is not in general true that $\sigma \mid X$ defines a permutation as the relation $R_{\sigma}^{|X}$ may not encode a bijection.

For $r \geq 0$ and a structure $\mathfrak{A}$, the $r$–type of an element $x \in A$ is the isomorphism class of $\mathfrak{A} \mid N(x,r)$. 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 be found in for instance [37] Theorem 6.27.

**Theorem 4.1** (Bounded–Degree Hanf Theorem). Let $r$ and $d$ be fixed. Then there is an integer $s$ such that for all structures $\mathfrak{A}$ and $\mathfrak{B}$ of degree at most $r$, if $\mathfrak{A}$ and $\mathfrak{B}$ are $(2^d, s)$–equivalent, then $\mathfrak{A} \equiv_d \mathfrak{B}$.

For convenience we cast this result into a form appropriate for permutations in TOTO.

**Theorem 4.2**. Let $d$ be fixed. There is an integer $s$ such that for all permutations $\pi$ and $\sigma$, if $\pi$ and $\sigma$ are $(2^d, s)$–equivalent, then $\pi \equiv_d \sigma$.

For every $t$ and $r$, all elements in a $t$–cycle of a permutation have the same $r$–type. This immediately gives the following basic corollary.

**Corollary 4.3.** If $\pi$ and $\sigma$ are such that $C_i(\pi) = C_i(\sigma)$ for all $i \geq 1$, then $\pi \equiv_d \sigma$ for all $d \geq 1$.

It is intuitively clear that if two permutations consist only of very long cycles compared to some $r \geq 1$, then the $r$–type of every element will be a path, and there will be very many such elements. We formalize this idea in the next lemma.

**Lemma 4.4.** Let $d \geq 1$. There exists an $h = h(d)$ such that if $\pi$ and $\sigma$ are two permutations with non–empty domains consisting only of cycles of length at least $h$, then $\pi \equiv_d \sigma$. We can select this $h$ such that $h \geq s(d)$ from Theorem 4.2.

**Proof.** Let $h = h(d)$ be such that $h(d) \geq s(d)$ from Theorem 4.2 and such that the $2^d$–neighborhood of any point in a $h$–cycle is a path. Such paths are necessarily all isomorphic. If $\pi$ and $\sigma$ have non–empty domains and contain only cycles of length at least $h$, the $2^d$–type of any of their elements is a path, and they both have at least $h \geq s$ such elements. So they are $(2^d, s)$–equivalent and $\pi \equiv_d \sigma$ by Theorem 4.2. ■

**Lemma 4.5.** Let $d,t \geq 1$ and let $h(d)$ be as in Lemma 4.4. If $\pi$ and $\sigma$ both are disjoint unions of at least $h(d)$ cycles of length $t$, then $\pi \equiv_d \sigma$.
Proof. The two permutations are \((2^d, s)\) equivalent by \(h(d) \geq s(d)\). The result follows from Theorem 4.2. ■

Lemma 4.6. Let \(d \geq 1\) and let \(h(d)\) be as in Lemma 4.4. Let \(\pi\) and \(\sigma\) be two permutations satisfying

\[
C_i(\pi) = C_i(\sigma), \quad \text{for all } i \in [h - 1].
\]

Suppose moreover that \(\pi\) and \(\sigma\) both contain at least one cycle of length at least \(h\). Then \(\pi \equiv_d \sigma\).

Proof. Let \(\kappa_1\) be a permutation such that

\[
C_i(\kappa_1) = \begin{cases} 
C_i(\pi) & \text{if } i \leq h - 1, \\
0 & \text{otherwise}.
\end{cases}
\]

Let \(\kappa_2\) be an \(h\)–cycle. By Corollary 4.3 and 3.23 and Lemma 4.4, both \(\pi\) and \(\sigma\) are \(d\)–equivalent to \(\kappa_1 \oplus \kappa_2\). As \(\equiv_d\) is an equivalence relation we also have \(\pi \equiv_d \sigma\). ■

We now have the necessary tools to prove Theorem 1.1. Part (i) has already been shown by Compton [16], so we will only supply proofs for the other two parts.

Proof of Theorem 1.1 Part (ii). Let \(\varphi\) have quantifier depth \(d\), and let \(h = h(d)\) be as in Lemma 4.4. Let \(\kappa\) be a permutation with

\[
C_i(\kappa) = \begin{cases} 
\kappa & \text{if } 1 \leq h - 1, \\
1 & \text{if } i = h, \\
0 & \text{otherwise}.
\end{cases}
\]

We will show that \(\Pi_n\) agrees with \(\pi\) on \(\varphi\) a.a.s. as \(n \to \infty\). By Corollary 3.23 together with Lemmas 4.4 and 4.5, \(\Pi_n\) agrees with \(\kappa\) if

\[
C_i(\Pi_n) \geq \begin{cases} 
h & \text{if } i < h, \\
1 & \text{if } i = h.
\end{cases}
\]

By Theorem 3.11, the above inequalities hold a.a.s. as \(n \to \infty\). Thus

\[
\lim_{n \to \infty} \mathbb{P}(\Pi_n \models \varphi) = \mathbb{P}(\kappa \models \varphi).
\]

Proof of Theorem 1.1 Part (iii). We will use Theorem 3.12 which says that for \(q > 1\), \(\Pi_n \sim \text{Mallows}(n, q)\) and \(\Sigma \sim \text{Mallows}(\Sigma, 1/q)\) we have

\[
(C_1(\Pi_{2n+3}), C_3(\Pi_{2n+1}), \ldots) \xrightarrow{d_{n \to \infty}} (C_1(\rho \circ \Sigma), C_3(\rho \circ \Sigma), \ldots),
\]

\[
(C_1(\Pi_{2n}), C_3(\Pi_{2n}), \ldots) \xrightarrow{d_{n \to \infty}} (C_1(\rho \circ \Sigma), C_3(\rho \circ \Sigma), \ldots),
\]

the limit distributions being distinct for all \(q > 1\).

As per Remark 3.13, with probability one both \((C_1(\rho \circ \Sigma), C_3(\rho \circ \Sigma), \ldots)\) and \((C_1(\rho \circ \Sigma), C_3(\rho \circ \Sigma), \ldots)\) contain finitely many nonzero values. So the limit distributions being distinct implies that there exists a sequence \(c_1, c_3, \ldots\) with finitely many nonzero values such that

\[
\mathbb{P}((C_1(\Pi_{2n+1}), C_3(\Pi_{2n+1}), \ldots) = (c_1, c_3, \ldots))
\]

and

\[
\mathbb{P}((C_1(\Pi_{2n}), C_3(\Pi_{2n}), \ldots) = (c_1, c_3, \ldots))
\]

both converge as \(n \to \infty\), but to different numbers. Fixing such a sequence \(c_1, c_3, \ldots\) and letting \(c_{2k+1}\) be its last nonzero element, we conclude that

\[
\lim_{n \to \infty} \mathbb{P}((C_1(\Pi_n), C_3(\Pi_n), \ldots, C_{2k+1}(\Pi_n)) = (c_1, c_3, \ldots, c_{2k+1}))
\]

does not exist. This event can clearly be queried by a sentence in \text{TOOB}. ■
5 Proof of Theorem 1.2 Part (i)

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 $\pi \equiv_d \sigma$ means that $\pi$ and $\sigma$ agree on all TOTO sentences of depth at most $d$. As per Lemma 3.20, there are finitely many equivalence classes for the equivalence relation $\equiv_d$. Given some fixed $d$, we denote these classes by $E_1, \ldots, E_\ell$, where $\ell = \ell(d)$. For a permutation $\pi$ we let $[\pi]_d$ be the equivalence class containing $\pi$.

We define $id_n$ as the unique increasing permutation on $[n]$, i.e., $id_n : i \mapsto i$ for $i \in [n]$. For a $\pi \in S_k$ and $\sigma_1, \ldots, \sigma_k$, the authors of [2] define the permutation $\pi[\sigma_1, \ldots, \sigma_k]$ called the inflation of $\pi$ by $\sigma_1, \ldots, \sigma_k$. We do not need the full definition; it will suffice to know that $id_2[\sigma_1, \sigma_2] = \sigma_1 \oplus \sigma_2$, where $\oplus$ is as defined in [9].

**Proposition 5.1** ([2], Proposition 26). Let $n, m$ and $d$ be positive integers with $n, m \geq 2^d - 1$. Then we have $id_m \equiv_d id_n$.

**Proposition 5.2** ([2], Proposition 28). Let $\alpha \in S_n$ and for $1 \leq i \leq n$ suppose that $\sigma_i \equiv \kappa_i$. Then $\alpha[\sigma_1, \ldots, \sigma_n] \equiv_d \alpha[\kappa_1, \ldots, \kappa_n]$.

**Corollary 5.3.** Let $\pi$ be any permutation and $n = 2^d - 1$. Then $\pi \circ id_n \equiv_d \pi \circ id_{n+1}$.

Until further notice we let $q < 1$ be fixed. We generate a sequence $\Pi_n \in S_n$ as follows: We let $Z_1, Z_2, \ldots$ be i.i.d. Geo$(1 - q)$ random variables, and sample $\Pi$ according to them as described in Section 3.2.2. Then for $n \geq 1$ we define $\Pi_n = \rk(\Pi_1, \ldots, \Pi(n))$, which implies that $\Pi_n \sim$ Mallows$(n, q)$ by Lemma 3.18. Given such a permutation $\Pi$ we define the sequence

$$
T_0 = 0,
T_{i+1} = \inf\{t > T_i : \Pi([t]) = [t]\} \quad \text{for } i \geq 0.
$$

Then $\Pi \upharpoonright [T_{i+1}] \setminus [T_i]$ is a bijection for all $i \geq 0$. Defining $\Pi^* : \mathbb{N} \to \mathbb{N}$ by

$$
\Pi^*(i) = \Pi(T_i + i) - T_1 \quad \text{for } i \in \mathbb{N},
$$

we have that $\Pi^*$ is a bijection of $\mathbb{N}$ to $\mathbb{N}$. Moreover, as is clear from the definition of $p$-shifted random permutations given in Section 3.2.2, $\Pi^* \equiv_d$ Mallows$(\mathbb{N}, q)$. This implies in particular that $T_1 \leq T_{i+1} - T_i$ for all $i \geq 0$.

**Lemma 5.4** ([7], Lemma 4.1 and Lemma 4.5). $\mathbb{E}T_1^2 < \infty$.

What follows is a reformulation of Proposition 3.16 in terms of the $T_i$, it also follows directly from the previous lemma.

**Lemma 5.5.** The $T_i$ are all finite with probability 1.

The approach we will take to prove Theorem 1.2 Part (i) is inspired by the approach taken by Lynch in [46] 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 $M_n$ that we will define is more complicated than the one used in [46], mainly because it is defined on a countably infinite state space.

For $\Pi \sim$ Mallows$(\mathbb{N}, q)$ we define

$$
R_n := \max\{0 \leq t \leq n : \Pi([t]) = [t]\} = \max\{T_i : T_i \leq n\},
$$
where the $T_i$ are as defined earlier. Note that if $T_1 > n$ then $R_n = T_0 = 0$, so that $R_n$ is well-defined. For a permutation $\Pi$ generated by $Z_1, Z_2, \ldots$ we define the sequence $M_n = M_n(\Pi)$ by

$$M_n = ([\Pi R_n]_d, (Z_{R_n+1}, \ldots, Z_n)).$$

The process $M_n$ takes values in the set

$$\mathcal{S} \subseteq \{\mathcal{E}_1, \ldots, \mathcal{E}_t\} \times \left(\emptyset \cup \bigcup_{i=1}^{\infty} \mathbb{N}^i\right),$$

where we denote the empty sequence by $\emptyset$. Note the strict inclusion as $M_n$ will for instance never be of the form $(\mathcal{E}, (1))$. The state space $\mathcal{S}$ is a (subset of a) countable union of countable sets, and thus itself countable. We have for example that $M_n(\Pi) = ([\Pi \iota]_d, \emptyset)$ for all $j \geq 0$. A key property of $M_n$ is that if $M_n(\Pi) = (\mathcal{E}, (z_1, \ldots, z_k))$ then for any $\pi \in \mathcal{E}$ we have by Proposition 5.2 that

$$\Pi_n \equiv_d \pi \oplus \text{rk}(z_1, \ldots, z_k).$$

Ergo, to know the equivalence class of $\text{rk}(\Pi(1), \ldots, \Pi(n))$ under $\equiv_d$ it is enough to know the value of $M_n(\Pi)$.

We take a dynamic viewpoint, where we determine the $Z_1, Z_2, \ldots$ one by one and observe the evolution of $M_n$.

**Lemma 5.6.** Let $\pi$ be a permutation of $\mathbb{N}$ and $n \in \mathbb{N}$ be such that $M_n(\pi) = (\mathcal{E}, (z_1, \ldots, z_k))$. Then there exists an $\mathcal{E}'$ such that

$$(\mathcal{E}, (z_1, \ldots, z_k)) \to (\mathcal{E}', \emptyset).$$

**Proof.** Let $\Pi \sim \text{Mallows}(\mathbb{N}, q)$ and consider the event $A = \{\Pi \upharpoonright [n] = \pi \upharpoonright [n]\}$; by Remark 3.17 we have $\mathbb{P}(A) > 0$. Let $B$ be the event

$$B := \bigcup_{j > n} \{M_j(\Pi) \in \{\mathcal{E}_1, \ldots, \mathcal{E}_t\} \times \{\emptyset\}\}.$$ 

To conclude the proof it suffices to show that $\mathbb{P}(B \mid A) > 0$. By Lemma 5.5 and $\mathbb{P}(A) > 0$ we conclude that

$$\mathbb{P}(B \mid A) \geq \mathbb{P}\left(\bigcup_{i=0}^{\infty} \{T_i < \infty\} \mid A\right) = 1.$$ 

We define also the reduced chain $L_i := [\Pi \iota]_d$ for $i \geq 0$. This is also a Markov chain as

$$\mathbb{P}(L_{i+1} = [\pi \oplus \sigma]_d \mid L_i = [\pi]_d) = \sum_{\sigma' \equiv_d \sigma} \mathbb{P}(\Pi \iota = \sigma').$$

Note also that $L_i$ is defined on the finite state state $\{\mathcal{E}_1, \ldots, \mathcal{E}_t\}$, so that every recurrent class of $L_i$ is positive recurrent. We also have that $M_n(\Pi) = (L_j, \emptyset)$. This implies in particular that $\mathcal{E}_j$ is a transient state of $L_i$ if and only if $(\mathcal{E}_j, \emptyset)$ is a transient state of $M_n$. As $L_i$ has a finite state space, it has at least one recurrent state by Lemma 3.6, so that $M_n$ also contains at least one recurrent state.

The next lemma contains all the important results about the chain $M_n$.

**Lemma 5.7.** (i) Every recurrent communicating class of $M_n$ contains an element of the form $(\mathcal{E}, \emptyset)$, where $\mathcal{E}$ is also a recurrent state for $L_i$;

(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 $(\mathcal{E},(z_1,\ldots,z_k))$ is contained in some recurrent communicating class. Then Lemma 5.6 implies that this communicating class contains an element of the form $(\mathcal{E}',\emptyset)$. 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 of the form $(\mathcal{E},\emptyset)$ such that $\mathcal{E}$ is also a recurrent state of $L_i$. Let $(\mathcal{E}',\emptyset)$ be an element in the same class as $(\mathcal{E},\emptyset)$. Define

$$\tau = \inf\{n \geq 1 : M_n = (\mathcal{E}',\emptyset)\},$$

$$\tilde{\tau} = \inf\{i \geq 1 : L_i = \mathcal{E}'\}.$$  

We may assume that $\mathcal{E}$ and $\mathcal{E}'$ are such that $E(\mathcal{E},\emptyset)\tau$ is maximized over all $\mathcal{E}, \mathcal{E}'$ occurring together in a recurrent class of $M_n$. We do not a priori assume that this expectation is finite. We may have that $\mathcal{E} = \mathcal{E}'$; this will not be a problem. Then $\mathcal{E}'$ is also in the same recurrent class as $\mathcal{E}$ for the chain $L_i$. We have $\tilde{\tau} = T_\tau$. If $\tilde{\tau} \geq N$ for some $N \geq 1$, then the expected hitting time of $(\mathcal{E}',\emptyset)$ given $M_0 = (\mathcal{E},\emptyset)$ is at most $T_N + E(\mathcal{E},\emptyset)\tau$, by our choice of $\mathcal{E}$ and $\mathcal{E}'$. Thus

$$E(\mathcal{E},\emptyset)\tau \leq ET_N + \mathbb{P}_\mathcal{E}(\tilde{\tau} \geq N) E(\mathcal{E},\emptyset)\tau.$$  

If $\mathbb{P}_\mathcal{E}(\tilde{\tau} \geq N) < 1$, this may be rearranged to

$$E(\mathcal{E},\emptyset)\tau \leq \frac{N ET_1}{1 - \mathbb{P}_\mathcal{E}(\tilde{\tau} \geq N)},$$

As the chain $L_i$ has a finite state space and $\mathcal{E}$ and $\mathcal{E}'$ are elements of a recurrent class, $E(\mathcal{E},\emptyset)\tau < \infty$ by Lemma 3.6. So there is some some $N$ such that $\mathbb{P}_\mathcal{E}(\tilde{\tau} \geq N) < 1$. By Lemma 5.4 this shows that $E(\mathcal{E},\emptyset)\tau < \infty$.

For the fourth claim, consider a recurrent class containing $(\mathcal{E},\emptyset)$. For $k \geq 1$ let $\kappa_k$ be the permutation in $S_k$ consisting of $k$ fixed points. Then for $\pi \in \mathcal{E}$ we have for all $s \geq 1$ that

$$\mathbb{P}(M_S = ([\pi + \kappa_s]_d,\emptyset) | M_0 = (\mathcal{E},\emptyset)) = (\mathbb{P}(\text{Geo}(1 - q) = 1))^s > 0.$$  

In particular the recurrent class contains $([\pi + \kappa_{2d-1}]_d,\emptyset)$ and $([\pi + \kappa_{2d}]_d,\emptyset)$, where $M_n$ may transition from the former to the latter in one step with positive probability. By Lemma 5.3 we have $[\pi + \kappa_{2d-1}]_d = [\pi + \kappa_{2d}]_d$ so that the class is indeed aperiodic.

The probability that $M_n$ is never equal to an element $(\mathcal{E},\emptyset)$ of a recurrent class is exactly the probability that $L_i$ is never equal to a recurrent state $\mathcal{E}$. This probability is zero by Lemma 5.7.

Let the recurrent classes of $M_n$ be denoted $\mathcal{M}_1,\ldots,\mathcal{M}_k$; there are finitely many of them by Lemma 5.7 Part (ii). Define

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

Proof of Theorem 1.3 Part (i). Recall that we fixed $q < 1$. Fix some $\varphi \in \text{TOTO}$ with quantifier depth $d$. By Lemma 5.7 Part (v) we have

$$\mathbb{P} (\Pi_n \models \varphi) = \sum_{i=1}^k \mathbb{P} (\Pi_n \models \varphi \mid \tau_{\mathcal{M}_i} < \infty) \mathbb{P} (\tau_{\mathcal{M}_i} < \infty).$$
We have seen that the equivalence class under \( \equiv_d \) of \( \Pi_n \) can be recovered from \( M_n(\Pi) \), so that conditional on \( M_n \in M_1 \), we have \( \Pi_n \models \varphi \) if and only if \( M_n(\Pi) \in M_i^\varphi \) for some subset \( M_i^\varphi \subseteq M_i \). That is, there are subsets \( M_i^\varphi \subseteq M_1, \ldots, M_i^\varphi \subseteq M_k \) such that

\[
P(\Pi_n \models \varphi) = \sum_{i=1}^k P(M_n \in M_i^\varphi | \tau_{M_i} < \infty) P(\tau_{M_i} < \infty).
\] (13)

By Lemma 5.7 Parts (iii) and (iv) and Lemma 3.9 each of the summands in (13) has a limit as \( n \to \infty \). As the sum is finite, we conclude that \( \lim_{n \to \infty} P(\Pi_n \models \varphi) \) exists. Examining the proof of Lemma 3.9 gives the intuitive expression

\[
\lim_{n \to \infty} P(\Pi_n \models \varphi) = \sum_{i=1}^k \pi_i(M_i^\varphi) P(\tau_{M_i} < \infty),
\]

where the \( \pi_i \) are the unique invariant probability distributions over the \( M_i \) provided by Theorem 3.8.

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

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

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

\[
\lim_{n \to \infty} P(\Pi_n \models \varphi) = \lim_{n \to \infty} P(r_n \circ \Pi_n \models \varphi^{-\text{reverse}}),
\]

where \( \varphi^{-\text{reverse}} \) is as in Lemma 3.24. As \( r_n \circ \Pi_n \overset{d}{=} \text{Mallows}(n, 1/q) \) 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 \( \Pi_n(1) = n \).

Remark 5.8. The proof of the convergence law in Theorem 1.2 Part (i) only uses the positive recurrence of the Mallows\((\mathbb{N}, q)\) distribution, i.e., Lemma 5.5. But by Proposition 3.16, this property holds in general for \( p \)-shifted random permutations of \( \mathbb{N} \) as long as \( \sum_i ip(i) < \infty \) and \( p(1) > 0 \). Therefore the convergence law in Theorem 1.2 Part (i) holds in general for \( \Pi_n^* := \text{rk}(\Pi^*(1), \ldots, \Pi^*(n)) \) where \( \Pi^* \) is a \( p \)-shifted permutation of \( \mathbb{N} \) with \( \sum_i ip(i) < \infty \) and \( p(1) > 0 \). The zero–one law may also hold in this case: Take for example \( p(1) = 1 \), then, deterministically, \( \Pi_n^* \) will consist of \( n \) fixed points and thus always converges to the same equivalence class by Proposition 5.1.

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 (iii).

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 \( \varphi \in \text{TOTO} \) such that for \( \Pi_n \sim \text{Mallows}(n, 1) \) and \( n \) satisfying \( W(2)(\log^{**} \log^{**} n - 1) < \log \log n \), we have

\[
P(\Pi_n \models \varphi) = \begin{cases} 1 - O(n^{-100}) & \text{if } \log^{**} \log^{**} n \text{ is even}, \\ O(n^{-100}) & \text{if } \log^{**} \log^{**} n \text{ is odd}. \end{cases}
\]
Proposition 6.1 is an explicit version of a result given in [24] by Foy and Woods who showed that there is a TOTO sentence $\psi$ such that $P(\text{Mallows}(n, 1) \models \psi)$ does not have a limit. Our plan of attack will follow in broad lines the proof given by Shelah and Spencer in [58] to show non-convergence in first-order logic for graphs for the Erdős–Rényi random graph $G(n, p)$ with $p$ near $n^{-1/7}$. See also Chapter 8 of the excellent monograph [60] by Spencer for a similar argument for $p$ near $n^{-1/3}$.

6.1 Arithmetic on sets in second-order logic

In this section we consider finite sets $S$ having a total order $\prec$. Let $\eta(x)$ be a formula (in some logic to specified later on) and define $A = \{s \in S : S \models \eta[s]\}$, the set of all elements satisfying $\eta$ in $S$. The total order on $S$ induces a total order on $A$, which we may use to uniquely determine the smallest element in $A$ and call it 1, the second smallest element in $A$ and call it 2, and so on. Recall that second-order sentences may also take relations as free variables, and allow quantification over relations of any arity. Our aim will be to determine a second-order sentence $\text{Parity}$ such that $S \models \text{Parity}$ if and only if $\log^* \log^* |A|$ is even.

Our sentence $\text{Parity}$ will begin with the existential quantification over four binary relations $R_D, R_E, R_T$ and $R_W$. Then we want to demand that $R_D$ is such that $R_D(i, j)$ if and only if $j = 2i$. In Section 3.3.1 we defined the formulas $\text{succ}_1^{(k)}$ using the $<_1$ relation in TOTO. In the current context we also assume that we have such a total ordering, namely $\prec$. So we can again use the formulas $\text{succ}_i^{(j)}$ with $<_1$ replaced by $\prec$. A formula expressing that the relation $R_D$ is as desired can then be written as

\[ \psi_D(R_D) := \forall i, j \in A : \left( (R_D(1, j) \leftrightarrow (j = 2)) \land (R_D(i, j) \land (i \neq 1)) \leftrightarrow \exists j', j'' \in A : \left( \text{succ}_1(i', i) \land \text{succ}_1^{(2)}(j', j) \land R_D(i', j) \right) \right). \]

Here $\forall i \in A : \psi(i)$ is shorthand for $\forall i(\eta(i) \rightarrow \psi(i))$ and $\exists i \in A : \psi(i)$ is shorthand for $\exists i(\eta(i) \land \psi(i))$. The relation $R_D$ is asymmetric. Recalling the definition of $T(i)$ and $W(i)$ from Section 3.4 we further want our sentence $\text{Parity}$ to demand that

\[ R_E(i, j) \iff j = 2^i, \]
\[ R_T(i, j) \iff j = T(i), \]
\[ R_W(i, j) \iff j = W(i). \]

For $R_E$, (15) can be expressed by a formula closely resembling that in (14):

\[ \psi_E(R_E) := \forall i, j \in A : \left( (R_E(1, j) \leftrightarrow (j = 2)) \land (R_E(i, j) \land (i \neq 1)) \leftrightarrow \exists j', j'' \in A : \left( \text{succ}_1(i', i) \land R_D(i', j) \land R_E(i', j') \right) \right). \]

This says that $j = 2^i$ if and only if $j = 2$, and that $j = 2^i$ for $i \geq 2$ if and only if there are $i', j'$ such that $i' + 1 = i$, $j = 2j'$ and $j' = 2^j$. To handle the relation $R_T$ we replace in the above $R_D$ and $R_E$ by $R_E$ and $R_T$, respectively. Similarly, to express that $R_W$ is as desired, we replace in the above $R_D$ and $R_E$ by $R_T$ and $R_W$, respectively. The conjunction $\bigwedge_{s \in \{E, D, T, W\}} \psi_s(R_s)$ gives a second-order formula $\text{Arith}(R_D, R_E, R_T, R_W)$ satisfied by the relations if and only if they indeed correctly encode the arithmetic operations as described.

In any set with a total order there are always four relations satisfying $\text{Arith}$. Given such relations we now want to express that $\log^* \log^* n$ is even, where $n = |A|$. Suppose that there is an element $x \in A$ for which there exists $y, z \in A$ 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 \( \exists w \in A : R_D(w, z) \). If instead \( x < n \), then \( \log^*(\log^* n) = z + 1 \), and we can again query whether this is even by \( \neg \exists w \in A : R_D(w, z) \). All of this can be formalized in a second–order formula \( \text{EvenSize}(R_D, R_E, R_T, R_W) \) in a straightforward manner. So a second–order sentence \( \text{Parity} \) of the form

\[
\text{Parity} := \exists R_D, R_E, R_T, R_W : \text{Arith}(R_D, R_E, R_T, R_W) \land \text{EvenSize}(R_D, R_E, R_T, R_W)
\]

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

We now look at structures which are directed graphs on domains having a total order \( < \). Suppose that we have four such directed graphs \( G_D, G_E, G_T \) and \( G_W \) all defined on the same domain \( T \) as defined above. Denote the arc relation of \( G_D \) as \( E_D \subseteq A \times A \), and similarly for \( G_E, G_T \) and \( G_W \). If \( \text{Arith}(E_D, E_E, E_T, E_W) \) holds, then \( \text{EvenSize}(E_D, E_E, E_T, E_W) \) holds if and only if \( \log^* \log^* |A| \) is even. We leave open for the moment how we will define such graphs, but having them dispenses of the need for second–order quantifiers in \( (16) \).

In broad lines we will now do the following: On \( \Pi_n \sim \text{Mallows}(n, 1) \) we will define directed graphs. These directed graphs will be so abundant that we can a.a.s. find four such graphs defined on the same ordered set that together do indeed encode the correct relations \( R_D, R_E, R_T \) and \( R_W \) on this set (i.e., the four arc relations satisfy \( \text{Arith} \)). Moreover, we can do so on a set of size roughly \( \log \log n \). Then \( \log^* \log^* \log \log n \) oscillates between being even and odd indefinitely, giving the desired non–convergence.

### 6.2 Defining directed graphs on permutations in TOTO

We will now show how we can define directed graphs on \( \Pi_n \) 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 \( \{a, \ldots, b\} \) can be checked by \( (a \leq_1 i) \land (i \leq_1 b) \), where the meaning of \( \leq \) should be clear.

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

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

For \( I, J \subseteq [n] \) with \( S(I, J) \neq \emptyset \), we define
Thus $y(I, J)$ is the element of $J$ that gets mapped to $\Pi_n(x(I, J)) + 1$ by $\Pi_n$. If $S(I, J) = \emptyset$, 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 $\text{TOTO}$ if $I$ and $J$ are intervals using the $<_2$ 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 $\text{TOTO}$ when $I$ and $J$ are intervals.

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

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

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

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

$$V(H(I; J)) := \{I_1, \ldots, I_N\},$$

$$E(H(I; J)) := \{e(I; J_\ell) : \ell = 1, \ldots, M\}.$$

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

For $A \subseteq [n]$ we define

$$W_k(A) := \{i : \{\Pi_n(i), \Pi_n(i) + 1, \ldots, \Pi_n(i) + k - 1\} \subseteq \Pi_n[A]\},$$

$$w_k(A) := |W_k(A)|. \quad (17)$$

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 $\Pi_n$ 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 $\text{succ}^{(j)}_2 (i, c)$ for some $1 \leq j \leq k - 1$, where $\text{succ}^{(j)}_2$ is as defined in Section 3.3.1. This may be expressed in $\text{TOTO}$ as

$$\neg \exists c \left( ((c <_1 a) \lor (b <_1 c)) \land \bigvee_{j=1}^{k-1} \text{succ}^{(j)}_2 (i, c) \right), \quad (18)$$

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 $i_1 \prec \cdots \prec i_L$, the sequence $I_k(J)$ of the minimal intervals between points of $W_k(J)$ is defined as:

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

(If $W_k(J) \leq 1$ then $I_k(J)$ is the “empty sequence”.) We will sometimes abuse notation and consider $I_k$ to be a set instead of a sequence.

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

$$H(I_k(I); I_k(J)) \quad \text{or} \quad H(\langle I_k(I_1), I_k(I_2)\rangle; I_k(J)),$$
Figure 6.2: In the figure we give an example of a permutation $\pi$ and intervals $I$ and $J$ such that $H(\mathcal{I}_2(I); \mathcal{I}_2(J))$ is a directed cherry. Take $n = 22$, $I = \{1, \ldots, 12\}$, $J = \{13, \ldots, 22\}$ 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 $i \in W_2(J)$.

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

We need one more ingredient in order to describe the final non–convergent sentence $\phi$ in Proposition 6.1. Given two disjoint sets $A, B$ we say that a directed graph $G$ with vertex set $A \cup B$ 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 $\mathcal{I}_k(I) = (I_1, \ldots, I_N)$, respectively $\mathcal{I}_k(I') = (I'_1, \ldots, I'_{N'})$. We will need to determine if $N > N'$. We construct a formula $Bigger(I, I')$ 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;
Lemma 6.3. For every two disjoint intervals $I, J$, there exists an interval $K$ such that $H(I, J) = K$, and moreover $w_{k+1}(I) = 0$ and $w_{k-1}(J) > 0$ for all $I, J$. 

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

6.3 Proof of Proposition 6.1

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

We will apply the ideas developed in Section 6.1 to sets of the form $I_k(I)$. Such sets are subsets of the collection of all intervals in $[n]$, and (given $I$) we can determine membership of an interval $\{a, \ldots, b\}$ in $I_k(I)$ by a TOTO formula as described in the previous section. This formula will serve as the formula $\eta$ in Section 6.1. Any four intervals $J_D, J_E, J_T, J_W$ induce four directed graph structures on $I_k(I)$, and we may check whether these graphs satisfy Arith as given in Section 6.1. If so, then EvenSize($J_D, J_E, J_T, J_W$) holds if and only if $\log^{**} \log^{**} |I_k(I)|$ 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 $\log^{**} \log^{**} |I_k(I)|$.

We now describe the sentence $\varphi$ appearing in Proposition 6.1. It will be of the form $\varphi = \exists x_1, y_1, \ldots, x_5, y_5 (\varphi_0 \land \forall x_6, y_6, \ldots, x_{10}, y_{10} (\varphi_1 \rightarrow \varphi_2))$. (19)

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 $I = (x_1, y_1), \ldots, J_W' = (x_{10}, y_{10})$. We define $\varphi_0(x_1, y_1, \ldots, x_5, y_5)$ as the TOTO formula such that \Pi_n, i_1, j_1, \ldots, i_5, j_5 \models \varphi_0 if and only if

- $w_{k+1}(I) = 0$ and $w_{k-1}(\hat{I}) > 0$ for all $\hat{I} \in I_k(I)$;
- Arith($I, J_D, J_E, J_T, J_W$) is satisfied;
• EvenSize($I, J_D, J_E, J_T, J_W$) is satisfied.

We define the formula \( \varphi_1(x_6, y_6, \ldots, x_{10}, y_{10}) \) that holds if and only if the first two conditions above are satisfied but EvenSize is not satisfied in the induced structure. Finally, we let \( \varphi_2(x_1, y_1, x_6, y_6) := \text{Bigger}(I, I') \).

**Proof of Proposition 6.7.** We show that with \( \varphi \) as defined in [19], \( \Pi_n \sim \text{Mallows}(n, 1) \) and \( n \) satisfying \( W(2)(\log^* \log^* n - 1) < \log \log n \) we have

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

In both cases we have by \( W(2)(\log^* \log^* n - 1) < \log \log n \) and Lemma 3.28 that

\[
\log^* \log^* n - 1 < \log \log^* [\log \log n] \leq \log^* \log^* n.
\]

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

We let \( k \) be such that the conclusion of Lemma 5.3 holds.

Consider first the case \( \log^* \log^* n \) even. By Lemma 6.2 and \( \log^* \log^* m \) being even for all \( \log \log n \leq m \leq n \), with probability \( 1 - e^{-n^2(1)} \) 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 \( \log \log n < |I_k(I)| < 2 \log \log n \). 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 \( N' := |I_k(I')| \) 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 \( H((I_k(I), I_k(I')), J) \) is a directed matching between \( I_k(I) \) and \( I_k(I') \) that saturates the vertices of \( I_k(I') \) but not the vertices of \( I_k(I) \). That is, \( \text{Bigger}(I, I') \) 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 \( |I_k(I)| < \log \log n \). As before, with probability \( 1 - e^{-n^2(1)} \) 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 \( \log \log n < |I_k(I')| < 2 \log \log n \). But then \( |I_k(I')| > |I_k(I)| \), so \( \text{Bigger}(I, I') \) cannot be satisfied. ■

### 6.4 The probabilistic component

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

\[
\mathbb{P}_\tau(.) := \mathbb{P}(., | \Pi_n(i) = \tau(i) \text{ for all } i \in B), \quad \mathbb{E}_\tau(.) := \mathbb{E}(., | \Pi_n(i) = \tau(i) \text{ for all } i \in B).
\]

Recall that we defined

\[
w_k(A) := \{|i: \{\Pi_n(i), \Pi_n(i) + 1, \ldots, \Pi_n(i) + k - 1\} \subseteq \Pi_n(A)\}.
\]

We define for the remainder of this section

\[
Y_i(A) := 1_{\{i, \ldots, i+k-1\} \subseteq \Pi_n(A)}.
\]

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 \( \tau \).)

Proof. Let us denote by \( J \) the set of all \( 1 \leq j \leq n - k + 1 \) for which \( \{j, \ldots, j + k - 1\} \cap \text{Im}(\tau) = \emptyset \). We note that
\[
\mathbb{E}_\tau Y_j = \begin{cases} \frac{|A|(|A| - 1) \ldots (|A| - k + 1)}{(n - |B|) \ldots (n - |B| - k + 1)} & \text{if } j \in J, \\ 0 & \text{otherwise}. \end{cases}
\]
Hence
\[
\mathbb{E}_\tau w_k(A) = |J| \cdot \frac{|A|(|A| - 1) \ldots (|A| - k + 1)}{(n - |B|) \ldots (n - |B| - k + 1)} = (1 + o(1)) n^{1-k} |A|^k,
\]
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
\[
\mathbb{E}_\tau Y_i Y_j = \begin{cases} \frac{|A|(|A| - 1) \ldots (|A| - 2k + 1)}{(n - |B|) \ldots (n - |B| - 2k + 1)} & \text{if } i, j \in J \text{ and } |i - j| > k, \\ 0 & \text{if } i \notin J \text{ or } j \notin J, \\ \leq \mathbb{E}_\tau Y_i & \text{if } |i - j| \leq k. \end{cases}
\]
and we also remark that
\[
\frac{|A|(|A| - 1) \ldots (|A| - 2k + 1)}{(n - |B|) \ldots (n - |B| - 2k + 1)} \leq \left( \frac{|A|(|A| - 1) \ldots (|A| - k + 1)}{(n - |B|) \ldots (n - |B| - k + 1)} \right)^2,
\]
since \( 1 \ll |A|, |B| \ll n \) so that \( (|A| - x)/(n - |B| - x) \leq (|A| - x - 1)/(n - |B| - x - 1) \) for all \( x = 1, \ldots, 2k - 1 \).

We see that
\[
\mathbb{E}_\tau w_k(A)^2 = \sum_i \sum_j \mathbb{E}_\tau Y_i Y_j \\
\leq |J|^2 \frac{|A|(|A| - 1) \ldots (|A| - 2k + 1)}{(n - |B|) \ldots (n - |B| - 2k + 1)} + (2k - 1) \sum_i \mathbb{E}_\tau Y_i \\
\leq (\mathbb{E}_\tau w_k(A))^2 + (2k - 1) \cdot \mathbb{E}_\tau w_k(A).
\]

The following lemmas establish that if \( w_k(I) > 0 \) then \( I \) is large in some suitable sense and if \( w_k(I) = 0 \) then \( I \) is 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 \( w_k(I') = 0 \) for all \( I' \subseteq I \). By Markov’s inequality and Lemma 6.4 with \( B = \emptyset \) we have for any given interval \( I \subseteq [n] \) with \( |I| = \lfloor n^{1-1000/k} \rfloor \) that
\[
\mathbb{P}(w_k(I) > 0) \leq \mathbb{E} w_k(I) = (1 + o(1)) n^{1-k} |I|^k \leq (1 + o(1)) n^{1-k} n^{k-1000} = O(n^{-999}).
\]
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 $1 - e^{-\Omega(n^{1-ke})}$, every interval $I \subseteq [n]$ of length at least $[n^{1-\varepsilon}]$ 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 $[n^{1-\varepsilon}]$ we have $\mathbb{P}(w_k(I) = 0) \leq e^{-\Omega(n^{1-ke})}$.

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

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

Denote by $E$ the event $\{|J| = |I|\}$. By Lemma 3.3 we have $\mathbb{P}(E) \geq 1/(n+1)$. Recall the definition of $Y_j(I)$ in (20), define for $1 \leq j \leq n - k + 1$ also the random variable

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

If $E$ holds then $J \equiv \Pi_n(I)$, namely, both are uniformly selected random subsets of $[n]$ of size $|I|$ if $E$ occurs. Thus

$$\mathbb{P}(w_k(I) = 0) = \mathbb{P}\left(\bigcap_j Y_j \right) = \mathbb{P}\left(\bigcap_j X_j \mid E\right) = \frac{\mathbb{P}\left(\bigcap_j X_j \cap E\right)}{\mathbb{P}(E)} \leq (n+1)\mathbb{P}\left(\bigcap_j X_j\right).$$

Let $S = \{ki : i \in \mathbb{N}\} \cap [n-k+1]$. The $X_j$ are independent for $j \in S$, and if $X_j$ is to occur for no $j \in [n-k+1]$ then certainly $X_j$ should not occur for any $j \in S$. Therefore

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

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

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

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

(i) $(1-\varepsilon)(\log \log n)^2 < w_k(I) < (1+\varepsilon)(\log \log n)^3$, and;

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

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

(iv) $|I'| > n^{1/7}$ 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 $\tau$.

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

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

where

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

29
Thus
\[
\mathbb{P}_\tau \left( |w_k(I) - (1 + o(1))(\log \log n)^3| > \frac{\varepsilon}{2}(\log \log n)^3 \right) < \frac{8k}{\varepsilon^2(\log \log n)^3} = o(1),
\]
showing Part \((i)\).

A simple application of Markov’s inequality together with Lemma 6.4 gives
\[
\mathbb{P}_\tau(w_{k+1}(I) > 0) \leq E_{\tau}(w_{k+1}(I) = (1 + o(1))n^{-k}k(\log \log n)^3)^{k+1} = o(1),
\]
giving Part \((ii)\).

We now look at Parts \((iii)\) and \((iv)\). We first show the stronger claim that for any fixed \(\delta > 1/k\) we have
\[
\mathbb{P}(I' \geq n^{1-\delta} \text{ for all } I' \in \mathcal{I}_k(I)) = 1 - o(1).
\]
Then Part \((iii)\) will follow from Lemma 6.6 with e.g. \(\delta = \frac{1}{2}\left(\frac{1}{k} + \frac{1}{k-1}\right)\) satisfying \(\frac{1}{k} < \delta < \frac{1}{k-1}\), and Part \((iv)\) will follow by setting \(\delta = \frac{3}{2}\). Note that \(w_k(I)\) depends only on the image \(\Pi_n[I]\). Conditioning on \(\Pi_n[I] = I'\) for some \(I' \subseteq [n]\), \(\Pi_n[I] = I'\) is distributed uniformly over all bijections \(I \to I'\). So 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 \(\{w_k(I) < 2(\log \log n)^3\}\). 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 \(2(\log \log n)^3\) points selected uniformly at random from \(I\), there is a pair \(i, j\) with \(|i - j| \leq n^{1-\delta}\). There are at most \(|I|n^{1-\delta}\) candidates for such \(i, j\). Thus
\[
\mathbb{P}(\exists i, j \in W_k(I) \text{ st. } |i - j| \leq n^{1-\delta} \mid F) \leq |I|n^{-\delta}O\left(\frac{4(\log \log n)^6}{|I|^2}\right)
\]
\[
= \frac{n^{1-\delta + o(1)}}{|I|} = n^{-\delta + \frac{1}{k} + o(1)}.
\]
For \(\delta > \frac{1}{k}\) this is indeed \(o(1)\). We finish the proof by noting that by Part \((i)\) \(\mathbb{P}(F) = 1 - o(1)\).

**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 \(S \subseteq \{I_i\}\), we will look at the probability that \(E_i\) does not hold for some \(i \in S\). Let \(g(n)\) be a function going to zero such that \(\mathbb{P}_\tau(E_i) \geq 1 - g(n)\) for any \(i \in S\), \(B \subseteq \bigcup_{j \in S} I_j \setminus \{I_i\}\) and any injection \(\tau : B \to [n]\). Such a \(g\) exists by Lemma 6.7 and \(\sum |I_j| \ll n/\log n\). Then
\[
\mathbb{P}(\bigcap_{i \in S} E_i) = \prod_{i \in S} \mathbb{P}(E_i | \bigcap_{j \in S, j < i} E_j).
\]
The event \(\bigcap_{j \in S, j < i} E_j\) depends only on the restriction of \(\Pi_n\) to \(\bigcup_{j \in S, j < i} I_j\). But for any such \(\tau : \bigcup_{j \in S, j < i} I_j \to [n]\) we have \(\mathbb{P}_\tau(\overline{E}_i) \leq g(n)\). Thus \(\mathbb{P}(\bigcup_{i \in S} \overline{E}_i) \leq g(n)^{|S|}\). By Theorem 3.2 with \(\gamma = \frac{1}{2} > g(n)\) we have
\[
\mathbb{P}(\sum_{i=1}^N E_i \geq \frac{N}{2}) \leq \exp\left\{-2N\left(\frac{1}{2} - g(n)\right)^2\right\} = \exp\left\{-\frac{(1 - o(1))N^\varepsilon}{2}\right\}.
\]
Lemma 6.9. Let $I$ be an interval satisfying Conditions (ii), (iv) of Lemma 6.7. Let $J$ be obtained from $I$ by deleting its largest element. Then $J$ satisfies Conditions (ii), (iv) of Lemma 6.7 and
\[ |I_k(I)| - k \leq |I_k(J)| \leq |I_k(I)|. \] (21)

Proof. Note that $W_t(J) \subseteq W_t(I)$ for all $t \geq 1$, so that $w_{k+1}(I) = 0$ implies $w_{k+1}(J) = 0$. Now let $J' \in I_k(J)$, then $J' = \{a+1, \ldots, b-1\}$ with $a, b \in W_{k-1}(J) \subseteq W_{k-1}(I)$. Then either $J' \in I_k(I)$ or $J' \supseteq I'$ for some $I' \in I_k(I)$. In the latter case, such $I'$ satisfies Conditions (iii) and (iv) of Lemma 6.7 so $J'$ does as well.

We now show (21). Define $j$ to be the largest element in $I$. 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 $|W_k(I) \setminus W_k(J)| \leq k$. $\blacksquare$

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 $1 - e^{-n^{O(1)}}$, for every two disjoint intervals $I_1, I_2 \subseteq [n]$ of length at least $n^{1/2 + \varepsilon}$, 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 $|I_1|$, resp. $|I_2|$, 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 $p_i := |I_i|/n$ 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
\[ E := \{|J'_1| = |I_1|, |J'_2| = |I_2|\}, \]
the pair $(J'_1, J'_2)$ has the same distribution as $(J_1, J_2)$, namely uniform over all pairs of disjoint sets of the correct cardinalities $|I_1|$ and $|I_2|$. Lemma 3.3 gives the crude bound
\[ \mathbb{P}(E) = \mathbb{P}(|J'_1| = |I_1|) \mathbb{P}(|J'_2| = |I_2|) \geq \frac{1}{(n+1)^2}. \] (22)

Next, consider the size of
\[ S' := \{j : j \in J'_1, j + 1 \in J'_2\}. \]

Write $X_j := 1_{\{j \in S'\}}$, so that $|S'| = X_1 + \cdots + X_{n-1}$. We have $\mathbb{E}X_j = p_1p_2 = p$ and
\[ \mathbb{E}|S'| = (n-1)p = \frac{|I_1| \cdot |I_2|}{n} - p. \] (23)

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 $X_{\text{even}} \overset{d}{=} \text{Bi}([(n - 1)/2], p)$, $X_{\text{odd}} \overset{d}{=} \text{Bi}([(n - 1)/2], p)$. By the Chernoff bound we have

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

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

$$
P\left(||S'| - \mathbb{E}|S'|| \geq \frac{\delta}{100} \mathbb{E}|S'|\right) \leq e^{-n^{\Omega(1)}}. \quad (24)$$

We see that

$$
P\left(|S - |I_1| |I_2| \rangle > \frac{\delta}{n} |I_1| |I_2| \right) \leq P\left(|S - \mathbb{E}S| > \frac{\delta}{100} \mathbb{E}|S|\right) = P\left(|S' - \mathbb{E}S'| > \frac{\delta}{100} \mathbb{E}|S'|\right) = \frac{P\left(|S' - \mathbb{E}S'\rangle \cap E\right)}{P(E)} \leq \frac{P\left(|S' - \mathbb{E}S'| > \frac{\delta}{100} \mathbb{E}|S'|\right)}{P(E)} \leq (n + 1)^2 \cdot e^{-n^{\Omega(1)}} = e^{-n^{\Omega(1)}}.
$$

(Here the first inequality holds for $n$ sufficiently large and uses (23). In the second line we use the earlier observation that $(S'|E) \overset{d}{=} S$ and in the fifth line we use (22) and (24).) ■

**Lemma 6.11.** For every fixed $\varepsilon > 0$ and $k \geq 3$, with probability $1 - e^{-n^{\Omega(1)}}$, for every sequence $I = (I_1, \ldots, I_N)$ 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

$$H(I; J_k(J)) = G.$$  

**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 $n^{1-1/k} (\log \log n)^{3/k}$.

For $\tau : I_1 \cup \cdots \cup I_N \to [n]$ an injection and $L_1, \ldots, L_M \geq 0$ and $A_{11}, \ldots, A_{1L_1} \subseteq J_1$, ..., $A_{M1}, \ldots, A_{ML_M} \subseteq J_M$ 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

$$F_{\tau, A, B} := \left\{ J_k(J_\ell) = (A_{\ell 1}, \ldots, A_{\ell L_\ell}) \text{ for all } \ell = 1, \ldots, M, \quad \Pi_n|A_\ell|' = B_\ell' \text{ for all } \ell = 1, \ldots, M \text{ and } \ell' = 1, \ldots, L_\ell, \quad \Pi_n(x) = \tau(x) \text{ for all } x \in I_1 \cup \cdots \cup I_n. \right\}$$

We observe that, conditional on $F_{\tau, A, B}$, 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

$$\mathbb{P}(\Pi_n | A_{ij} \equiv \tau_{ij} \text{ for all } i, j | F_{\tau, A, B}) = \frac{1}{|A_{11}|! \cdots |A_{ML_M}|!}.$$
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 $S(I_\ell, J') > 0$ 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 $\ell$ is good.) Note also that whether $\ell$ is good or not is determined completely by $F_{\tau, A_B}$, because the sets $S(I_\ell, A_{ij})$ depend only on $B_{ij}$, not on the choice of the specific bijection $A_{ij} \rightarrow B_{ij}$.

Note that if $\ell$ is good and we randomly choose bijections $\tau_{\ell} : A_{\ell} \rightarrow B_{\ell}$ then the points $y(I_1, A_{\ell}), \ldots, y(I_N, A_{\ell})$ are chosen uniformly without replacement from $A_{\ell}$. In particular, defining $P := \{(I_1, I_2) : i \neq j\}$, for each ordered pair $uv \in P$ the probability that $e(I, A_{\ell}) = uv$ is $\frac{1}{|P|} = \frac{1}{N^2 - 1}$.

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

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

Since $(1 - \varepsilon)(\log \log n)^3 n > N^2$, there are necessarily some duplicate arcs, this will not be a problem. Then $H(\mathcal{I}, (A_{\ell_1}, \ldots, A_{\ell_{L_k}})) = G$ is implied by the event $\{e(I, A_{\ell}) = e_i \text{ for all } 1 \leq i \leq L_\ell\}$.

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 $F_{\tau, A_B}$, and

$$\mathbb{P}(E | F_{\tau, A_B}) \geq 1 - \left(1 - \left(\frac{1}{N^2}\right)^{(1 + \varepsilon)(\log \log n)^3}\right)^Z \geq 1 - \exp \left(-Z \cdot \left(\frac{1}{N^2}\right)^{(1 + \varepsilon)(\log \log n)^3}\right) = 1 - \exp \left(-Z \cdot n^{o(1)}\right).$$

By Corollary 6.8 and Lemma 6.10 we have that

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

It follows that

$$\mathbb{P}(E) = \sum_{\tau, A_B} \mathbb{P}(E | F_{\tau, A_B}) \cdot \mathbb{P}(F_{\tau, A_B}) \geq 1 - \exp[-M/2] - \mathbb{P}(Z < M/2) = 1 - e^{-n^{\Omega(1)}}.$$

This proves the statement of the lemma. \hfill \blacksquare

**Proof of Lemma 6.3.** Let $I_1, \ldots, I_N$ be $\lceil n^{1/2k} \rceil$ intervals of length $\lfloor n^{-1/k} \cdot (\log n)^{3/k} \rfloor$. By Corollary 6.8, with probability $1 - e^{-n^{\Omega(1/2k)}}$ 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 the maximum 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 $|I'| > n^{1/7}$ for all $I' \in \mathcal{I}_k(I_1)$, and by $|I_1| \leq n^{1/2k}$ we also have $I' \leq n^{1/2k}$ 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 $1 - e^{-n^{\Omega(1)}}$. \hfill \blacksquare

**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 $I \in$
Define \( \alpha = \frac{1}{2(k+1)} \). 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. \( \square \)

### 7 Proof of Theorem 1.2 Part (ii)

Here we extend the case \( q = 1 \) to \( |1 - q| < 1/\log^* n \). We will make frequent use of the sampling algorithm for a Mallows\((n, q)\) distribution detailed in Section 3.2.1. Recall also the definition of TruncGeo\((n, p)\) random variables whose probability mass function is given in [1] and is valid in the range \( p \in (-\infty, 0) \cup (0, 1) \). In particular, TruncGeo\((n, 1 - q)\) random variables are well defined for \( q \in \mathbb{R}_{>0} \setminus \{1\} \).

#### 7.1 Extending the \( q = 1 \) case to \( q = 1 \pm O(n^{-4}) \)

We first extend Proposition 6.1 to the following:

**Proposition 7.1.** There exists a \( \varphi \in \text{TOTO} \) such that for \( \Pi_n \sim \text{Mallows}(n, q) \), all \( n \) satisfying \( W(2)(\log^* \log^* n - 1) < \log \log n \) and all \( c > 0 \) we have

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

for all sequences \( q = q(n) \) satisfying \( 0 \leq |1 - q| \leq cn^{-4} \).

Proposition 6.1 will follow by showing that for \( q = 1 - O(n^{-4}) \), a Mallows\((n, q)\) distribution is in some sense very close to the uniform distribution over \( S_n \).

**Lemma 7.2.** Let \( q \) satisfy \( 0 \neq |1 - q| \leq cn^{-4} \) for some \( c > 0 \). For \( 1 \leq i \leq n \) let \( Z_i \sim \text{TruncGeo}(n - i + 1, 1 - q) \) and \( X_i \sim \text{Unif}([n - i + 1]) \) be independent. Then

\[
d_{\text{TV}}(Z_i, X_i) \leq O(cn^{-3}), \quad \text{as } n \to \infty.
\]

**Proof.** Let \( m := n - i + 1 \). By [3] we may write

\[
d_{\text{TV}}(Z_i, X_i) = \frac{1}{2} \sum_{j=1}^{m} \left| \frac{(1 - q)q^{j-1} - 1}{1 - q^m} \right| = \frac{1}{2} \sum_{j=1}^{m} \left| \frac{m(1 - q)q^{j-1} - (1 - q^m)}{m|1 - q^m|} \right|. \tag{25}
\]

Define \( \alpha := 1 - q \). We may bound the numerator in the summands of (25) via

\[
\left| m(1 - q)q^{j-1} - (1 - q^m) \right| = \left| ma(1 - \alpha)^{j-1} - (1 - (1 - \alpha)^m) \right|
= \left| ma(1 - ja + O(j^2 \alpha^2)) - (1 - (1 - ma + O(m^2 \alpha^2))) \right|
= jma^2 + O(m^2 \alpha^2)
= O(m^2 \alpha^2).
\]

34
We then restrict to the permutation \( \Pi \). So the summands on the right-hand of (25) are all \( O(m\alpha^2)/|1-q|^m \). Now,

\[
\frac{O(m\alpha^2)}{|1-(1-\alpha)^m|} = \frac{O(m\alpha^2)}{m\alpha + O(m^2\alpha^2)} = \frac{O(\alpha)}{1 + O(\alpha)} = O(\alpha)(1 + O(\alpha)) = O(cn^{-4}).
\]

Thus \( d_{TV}(Z_i, X_i) = O(cnm^{-4}) = O(cn^{-3}) \).  

**Lemma 7.3.** Let \( q \) satisfy \(|1-q| \leq cn^{-4} \) for some \( c > 0 \). Let \( \Pi_n \sim Mallows(n, q) \) and \( \Pi_n^* \sim Mallows(n, 1) \). There is a coupling of \( \Pi_n \) and \( \Pi_n^* \) such that \( P(\Pi_n \neq \Pi_n^*) = O(cn^{-2}) \).

**Proof.** If \( q = 1 \), then \( \Pi_n = \Pi_n^* \) and the coupling may be achieved for instance by sampling \( \Pi_n \) and setting \( \Pi_n^* = \Pi_n \). So we now assume that \( q \neq 1 \).

Let \( \Pi_n \) be generated by a sequence \( Z_1, \ldots, Z_n \) of independent random variables with \( Z_i \sim TruncGeo(n - i + 1, 1 - q) \) for \( i \in [n] \) (see Section 3.2.1). Let \( \Pi_n^* \) be generated in the same manner by a sequence \( X_1, \ldots, X_n \) of independent random variables where \( X_i \sim Unif([n-i+1]) \) for \( i \in [n] \). By Lemma 7.2 and 3.4 we may couple \( Z_i \) and \( X_i \) for each \( i \) such that

\[
P(Z_i \neq X_i) \leq O(cn^{-3}), \quad \text{as } n \to \infty.
\]

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

\[
P(\Pi_n = \Pi_n^*) \geq 1 - O(cn^{-2}).
\]

**Proof of Proposition 7.1.** Let \( q \) satisfy \(|1-q| \leq cn^{-4} \) for some \( c > 0 \). Let \( \Pi_n \sim Mallows(n, q) \) and \( \Pi_n^* \sim Mallows(n, 1) \) be coupled such that

\[
P(\Pi_n \neq \Pi_n^*) = O(cn^{-2}),
\]

This coupling exists by Lemma 7.3. Let \( \varphi \) be the TOTO sentence from Proposition 6.1. Then \( \Pi_n \) and \( \Pi_n^* \) agree on \( \varphi \) with probability \( 1 - O(cn^{-2}) \), giving the desired result.

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

The strategy of the current section is as follows. We will show that there is a formula \( \psi(y) \) such that for \( \Pi_n \sim Mallows(n, 1 - O(n^{-1})) \), with probability \( 1 - \varepsilon \) there is a unique \( j \) such that \( \Pi_n \models \psi[j] \), and where also \( j = \Theta(n^{1/4}) \), see Lemma 7.12. We can then restrict to the permutation \( \Pi_j := \rk(\Pi_n(1), \ldots, \Pi_n(j)) \) by Lemma 3.26. For such \( q \) and \( j \) we will then have the bound

\[
|1 - q| = O(n^{-1}) = O(j^{-4}),
\]

so we may subsequently apply Proposition 7.1 to conclude the proof. To find this formula \( \psi \) with the desired property, we first look for the value \( J_1 \) which we define as the smallest \( j \) such that there is an \( i \in [n] \) with \( \{\Pi_n(i), \Pi_n(i) + 1\} \subseteq \Pi_n[[j]] \). For the current range of \( q \), \( \Pi_n \) is still similar enough to a uniform permutation that we roughly have

\[
\mathbb{E}[\{i : \{\Pi_n(i), \Pi_n(i) + 1\} \subseteq \Pi_n[[j]]\}] \approx j \cdot j/n.
\]

Consequently we will be able to conclude that \( J_1 = \Theta(\sqrt{n}) \) with appropriately high probability. We then restrict to the permutation \( \rk(\Pi_n(1), \ldots, \Pi_n(J_1)) \) and repeat this procedure to find a \( K_1 \) satisfying \( K_1 = \Theta(n^{-1/4}) \) again with at least the right probability for our purposes. We now fill in all the details.

We first show that \( q = 1 - O(1/n) \) is sufficiently close to 1 that \( TruncGeo(n, 1 - q) \) random variables behave somewhat like uniform random variables. This is made explicit in the next lemma. Recall the definition of the TruncGeo \( (n, p) \) distribution as given in (3), valid for \( p \in (-\infty, 0) \cup (0, 1) \).
**Lemma 7.4.** Let $\alpha = \alpha(n)$ satisfy $|\alpha| \leq c/n$ for some constant $c > 0$ and $q = 1 - \alpha$. Then there exist $c_U = c_U(c) > 0$ and $c_L = c_L(c) > 0$ such that for $0 \leq i \leq 2n/3$, $J \subseteq [n-i]$ and $n \geq 4c$ we have

$$
c_L \frac{|J|}{n} \leq \mathbb{P}(\text{TruncGeo}(n-i,1-q) \in J) \leq c_U \frac{|J|}{n}.
$$

Before we prove Lemma 7.4, we need the following simple result.

**Lemma 7.5.** If $0 \leq x \leq 1/4$ then $1 - x \geq e^{-2x}$.

**Proof.** If $0 \leq x \leq 1/4$ then $\frac{1}{2} \leq 4^{-2x} \leq e^{-2x}$. So for $0 \leq x \leq 1/4$ we have

$$
e^{-2x} = 1 - \int_0^x 2e^{-2z}dz \leq 1 - \int_0^x dz = 1 - x.
$$

\[
\square
\]

**Proof of Lemma 7.4.** We consider first the case that $q < 1$. Let $k \in J$, and define $m := n-i \geq n/3$. Let $Z_i \sim \text{TruncGeo}(m,1-q)$ so that

$$
\mathbb{P}(Z_i = k) = \frac{(1-q)q^{k-1}}{1-q^m} \quad \text{for } k = 1, \ldots, m.
$$

(26)

By $0 < q < 1$ this is strictly decreasing in $k$. In particular, as $Z_i$ has support $[m]$ we have $(1-q)/(1-q^m) > 1/m$. Thus

$$
\mathbb{P}(Z_i = k) > \frac{q^k}{m} = \frac{(1-\alpha)^{k-1}}{m}.
$$

By $n \geq 4c$ we have $\alpha < \frac{1}{3}$. Lemma 7.5 thus gives $\mathbb{P}(Z_i = k) \geq e^{-2k\alpha}/m$. By $k \leq m \leq n$ and $\alpha \leq c/n$ this is at least $e^{-2c}/n$. Thus

$$
\mathbb{P}(Z_i \in J) \geq e^{-2c} \frac{|J|}{n}.
$$

For the upper bound, note that $(1-\alpha)^m < e^{-m\alpha} \leq e^{-n\alpha/3}$. Thus

$$
\frac{(1-q)q^{k-1}}{1-q^m} < \frac{\alpha q^{k-1}}{1-e^{-n\alpha/3}} \leq \frac{\alpha}{1-e^{-n\alpha/3}}.
$$

(27)

To find an upper-bound for this expression, we define $g(x) := x/(1 - e^{-nx/3})$. We have

$$
g'(x) = \frac{1-e^{-nx/3} - \frac{nx}{3}e^{-nx/3}}{(1-e^{-nx/3})^2} = \frac{e^{-nx/3}(e^{nx/3} - 1 - \frac{nx}{3})}{(1-e^{-nx/3})^2}
$$

By $e^{nx/3} > 1 + nx/3$ for all $x > 0$, $g'(x)$ is positive for all $x > 0$. So the maximum of the right-hand side of (27) over $0 < \alpha \leq c/n$ is attained at $c/n$. Therefore

$$
\mathbb{P}(Z_i = k) \leq \frac{c}{1-e^{-c/3}} \frac{1}{n},
$$

and

$$
\mathbb{P}(Z_i \in J) \leq \frac{c}{1-e^{-c/3}} \frac{|J|}{n}.
$$

We now consider the case $q > 1$ implying $\alpha < 0$. The probability mass function in (26) is now strictly increasing in $q$, so that $(1-q)/(1-q^m) < 1/m$ in this case. We thus have the bound

$$
\mathbb{P}(Z_i = k) < \frac{(1+|\alpha|)^{k-1}}{m} \leq \frac{e^{|\alpha|}}{m}.
$$
By $k|\alpha| \leq c$ and $m \geq n/3$ this is at most $3e^c/n$. Therefore

$$\mathbb{P}(Z_i \in J) \leq 3e^c \frac{|J|}{n}.$$  

For the lower bound, observe that by $(1 + |\alpha|)^m \leq e^{m|\alpha|}$ we have

$$\mathbb{P}(Z_i \in J) \geq |J| \cdot \frac{q - 1}{q^m - 1} = |J| \cdot \frac{|\alpha|}{(1 + |\alpha|)^m - 1} \geq |J| \cdot \frac{|\alpha|}{e^{m|\alpha|} - 1}.$$  

(28)

Let $h(x) := x/(e^{mx} - 1)$, then

$$h'(x) = \frac{e^{mx} - 1 - mx e^{mx}}{(e^{mx} - 1)^2} = \frac{e^{mx}(1 - mx - e^{-mx})}{(e^{mx} - 1)^2}.$$  

As $1 - mx < e^{-mx}$ for $x > 0$, the above is negative for all $x > 0$. So (28) is minimized at $\alpha = c/n$ giving

$$\mathbb{P}(Z_i \in J) \geq \frac{c}{e^{cm/n} - 1} \frac{|J|}{n} \geq \frac{c}{e^c - 1} \frac{|J|}{n}.$$  

\[\blacksquare\]

For $\Pi_n \in S_n$, recall the definition of $w_k(A)$ as given in (17). We will be examining the quantity $w_2([j])$ for $j \in [n]$. For $\Pi_n \in S_n$, $m \in [n - 1]$ and $j \in \mathbb{N}$ we define

$$A_m(j) := |\{i \leq m : \Pi_n(i) \leq j\}|,$$

$$Y_m(j) := 1_{\{\Pi_n(m), \Pi_n(m + 1) \leq j\}}.$$  

When the permutation under consideration is ambiguous we will write $w_2([j], \Pi_n)$, $A_m(j, \Pi_n)$ and $Y_m(j, \Pi_n)$. We note that $Y_m(j) = 1$ cannot be queried in TOTO, as we cannot make comparisons of the form $\Pi(m) \leq j$. However, we do have

$$w_2([j], \Pi_n) = |\{i : \{\Pi_n(i), \Pi_n(i + 1) \leq \Pi_n([j])\}|$$

$$= |\{m : \{\Pi_n^{-1}(m), \Pi_n^{-1}(m + 1) \leq [j]\}|$$

$$\overset{\Delta}{=} Y_1(j, \Pi_n) + \ldots + Y_{n-1}(j, \Pi_n).$$  

Here we used the fact that $\Pi_n^{-1} \overset{\Delta}{=} \Pi_n$ if $\Pi_n \sim \text{Mallows}(n, q)$, by (6).

For the following proofs we will make use of the sampling algorithm for $\Pi_n \sim \text{Mallows}(n, q)$ using a sequence $Z_1, \ldots, Z_n$ of independent random variables with $Z_i \sim \text{TruncGeo}(n-i+1, 1-q)$, see Section 3.2.1. In this procedure, during step $i$ we select $\Pi_n(i)$ as the $Z_i$-th smallest number in the set $[n] \setminus \{\Pi_n(1), \ldots, \Pi_n(i-1)\}$. In particular, when determining $\Pi_n(i)$ we have $\Pi_n(i) \leq j$ only if $Z_i \leq j$. Moreover, if $A_m < j/3$ and $Z_m \leq j/2$ then certainly $\Pi_n(m) \leq j$. We will put these facts to work in the proofs of the following lemmas.

**Lemma 7.6.** Let $\alpha = \alpha(n)$ satisfy $|\alpha| \leq c/n$ for some constant $c > 0$. Let $q = 1 - \alpha$, $\Pi_n \sim \text{Mallows}(n, q)$ and $1 \leq j \leq n$. Then, with $cV$ as in Lemma 7.4, we have

$$\mathbb{E}Y_i(j, \Pi_n) \leq cV \frac{j^2}{n^2} \text{ for } i \in [n - 1].$$  

**Proof.** If $Z_i > j$, then certainly $\Pi_n(i) > j$. Thus, for $i < n$ we have

$$\mathbb{E}Y_i = \mathbb{P}(Y_i = 1) \leq \mathbb{P}(Z_i \leq j) \mathbb{P}(Z_{i+1} \leq j).$$  

For $i < 2n/3 - 1$, by Lemma 7.4 this is at most $cV j^2/n^2$.  

37
Now consider the case $i \geq 2n/3$. Recall that we defined the reversal permutation $r_n \in S_n$ by $r_n : i \mapsto n - i + 1$. Let $\Pi_n := r_n \circ \Pi_n \circ r_n$ and $i' := r_n(i) \leq n/3 + 1$. We observe that
\[
P(\Pi_n(i), \Pi_n(i + 1) \leq j) = P(\Pi_n \circ r_n(i'), \Pi_n \circ r_n(i' - 1) \leq j) = P(r_n \circ \Pi_n \circ r_n(i'), r_n \circ \Pi_n \circ r_n(i' - 1) \geq r_n(j)).
\] (29)
As $r_n \circ \Pi_n \circ r_n \not\equiv \Pi_n$, (29) equals $P(\Pi_n(i'), \Pi_n(i' - 1) \geq r_n(j))$. Similarly to before, we have $\Pi_n(i') \geq n - j + 1$ only if $Z_{i'} \geq n - i' + 1 - j$. So for $i' \geq 2n/3$ we have
\[
P(Y_i = 1) \leq P(Z_{i'} \geq n - i' + 1 - j) P(Z_{i' - 1} \geq n - i' + 2 - j).
\]
By Lemma 7.4 we again conclude that $P(Y_i = 1) \leq c_U j^2/n^2$.

**Corollary 7.7.** Let $\alpha = \alpha(n)$ satisfy $|\alpha| \leq c/n$ for some constant $c > 0$. Let $q = 1 - \alpha$, $\Pi_n \sim \text{Mallows}(n, q)$ and $1 \leq j \leq n$. Then, with $c_U$ as in Lemma 7.4, we have
\[
Ew_2([j], \Pi_n) \leq c_U j^2/n.
\]

**Lemma 7.8.** Let $\alpha = \alpha(n)$ satisfy $|\alpha| \leq c/n$ for some constant $c > 0$ and $q = 1 - \alpha$. Let $\Pi_n \sim \text{Mallows}(n, q)$. Then there exist $\gamma = \gamma(c) > 0$ and $K = K(c) > 0$ such that for every $n$ and every $1 \leq j \leq n$
\[
P\left(A_m(j, \Pi_n) < \frac{j}{3} \text{ for all } m < \gamma n\right) \geq 1 - e^{-Kj}.
\]

**Proof.** Let $m = \lceil \gamma n \rceil$ for $\gamma = \gamma(c) > 0$ to be determined. Let $Z_1, \ldots, Z_n$ be independent random variables generating $\Pi_n \sim \text{Mallows}(n, q)$ with $Z_i \sim \text{TruncGeo}(n - i + 1, 1 - q)$. If less than $j/3$ of the $Z_1, \ldots, Z_m$ are at most $j$ then indeed $A_m < j/3$. Letting $c_L = c_L(c)$ and $c_U = c_U(c)$ be as in Lemma 7.4, we have for $i \leq m$ that
\[
c_L \frac{j}{n} \leq P(Z_i \leq j) \leq c_U \frac{j}{n},
\]
provided that $\gamma < 2/3$. Let $X_m$ be the the number of $Z_i$ with $i \leq m$ such that $Z_i \leq j$, so that $A_m \leq X_m$. Then
\[
\mu_m := EX_m \geq c_L j \frac{j}{n} \geq c_L j,
\]
and
\[
\mu_m \leq c_U j \frac{j}{n} < c_U (\gamma n + 1) \frac{j}{n}.
\]
By Chernoff’s inequality (Theorem 3.1) with $\epsilon = 1/2$ we have
\[
P\left(X_m \geq \frac{3}{2} c_U (\gamma n + 1) \frac{j}{n}\right) \leq P\left(X_m \geq \frac{3}{2} \mu_m\right) \leq \exp\left\{-\frac{\mu_m}{12}\right\} \leq \exp\left\{-\frac{c_L j}{12}\right\}.
\] (30)
Suppose that $n \geq 6c_U$ and set $\gamma = \frac{1}{12c_U} \leq \frac{1}{6c_U} - \frac{1}{n}$. Then
\[
\frac{3}{2} c_U (\gamma n + 1) \frac{j}{n} \leq \frac{3}{2} c_U \left(\frac{n}{6c_U}\right) \frac{j}{n} < \frac{j}{3}.
\]
Using this $\gamma$ in the right–hand side of (30) we then obtain
\[
P\left(A_m \geq \frac{j}{3}\right) \leq P\left(X_m \geq \frac{j}{3}\right) \leq \exp\left\{-\frac{c_L j}{144c_U}\right\}.
\]
We also have $A_1 \leq A_2 \leq \ldots \leq A_n$, so the result follows. We are left with the case $n < 6c_U$, but then $\gamma n < 1$ so that there is in fact no $m < \gamma n$ if $n < 6c_U$ in which case the result is vacuously true. ■
For the permutation $\Pi$ depicted we have $J_1(\Pi) = 5$. Indeed the set $W_2([5]) = \{i : \{\Pi(i), \Pi(i) + 1\} \subseteq \Pi([5])\}$ contains exactly $i = 5$ (depicted by red lines) and $i = 2$ (depicted by dashed lines), whereas $W_2([4])$ is empty.

The permutation $\Pi_{J_1} = \text{rk}(\Pi(1), \ldots, \Pi(5))$. We have $K_1(\Pi) = J_1(\Pi_{J_1}) = 3$ as witnessed by the red lines.

Figure 7.1: An illustration of $J_1$ and $K_1$
For the other bound we will again use the fact that $w_2([j], \Pi_n) \leq Y_1(j, \Pi_n) + \ldots + Y_{n-1}(j, \Pi_n)$. We claim that if $A_m < j/3$ and $Z_m, Z_{m+1} \leq j/2$, then $Y_m(j, \Pi_n) = 1_{\{\Pi_n(m), \Pi_n(m+1) \leq j\}} = 1$. To see this, note that $A_m < j/3$ means that

$|j| \setminus \{\Pi_n(1), \ldots, \Pi_n(m)\}| \geq \left\lceil \frac{2j}{3} \right\rceil$.

In this case, if $Z_m \leq j/2$ then $\Pi_n(m) \leq \left\lfloor \frac{j}{2} \right\rfloor + \frac{j}{2} < j$ and if additionally $Z_{m+1} \leq j/2$ then $\Pi_n(m + 1) \leq \left\lfloor \frac{j}{2} \right\rfloor + \frac{j}{2} + 1 \leq j$.

So let $X_i$ be the indicator random variable for the event $\{Z_i, Z_{i+1} \leq \frac{j}{2}\}$. If $A_m \leq j/3$ holds then $Y_1 + \ldots + Y_m \geq X_1 + \ldots + X_m$. Let $m = \lfloor \gamma n \rfloor - 1$ with $\gamma = \gamma(c)$ as in Lemma 7.8. Let $E$ be the set of even numbers in $[m]$ and $O$ be the set of odd numbers in $[m]$. For $n$ large enough, depending on $c$, we have $m/3 < |E|, |O| < 2m/3$. Moreover, the $X_i$ for $i \in E$ are mutually independent, as are the $X_i$ for $i \in O$. By Lemma 7.4 we also have

$$\mu_E := \mathbb{E} \left( \sum_{i \in E} X_i \right) \geq |E| c_L^2 \frac{j^2}{4n^2} \geq mc_L^2 \frac{j^2}{12n^2},$$

and the same bound for $\mu_O := \mathbb{E} \left( \sum_{i \in O} X_i \right)$. By Chernoff’s bound (Theorem 3.1 with $\varepsilon = \frac{1}{2}$) we have

$$\mathbb{P} \left( \sum_{i \in E} X_i \leq mc_L^2 \frac{j^2}{24n^2} \right) \leq \mathbb{P} \left( \sum_{i \in E} X_i \leq \frac{\mu_E}{2} \right) \leq \exp \left( -\frac{\mu_E}{8} \right) \leq \exp \left( -mc_L^2 \frac{j^2}{96n^2} \right).$$

We have the same bound replacing $E$ by $O$ in the sums in (31), as we only used the bounds on $|E|$ and the mutual independence of $X_i$ for $i \in E$. Let $b_2$ be for the moment undetermined and set $j = \lfloor b_2 \sqrt{n} \rfloor$. Then

$$\frac{mj^2}{n^2} \geq \frac{(\gamma n - 1)(b_2 \sqrt{n} - 1)^2}{n^2} = (1 + o(1)) \frac{\gamma b_2^2 n^2}{n^2}.$$

So we can select $b_2$ so large that $mc_L^2 \frac{j^2}{24n^2} > 1$ and such that the right–hand side of (31) is less than $\varepsilon/8$. By $|m| = E \cup O$ we have

$$\mathbb{P} \left( w_2([b_2 \sqrt{n}], \Pi_n) \leq 1 \mid A_m \leq \frac{j}{3} \right) \leq \mathbb{P} \left( \sum_{i \in [m]} X_i \leq 1 \mid A_m \leq \frac{j}{3} \right) \leq \frac{\varepsilon}{4} \mathbb{P} \left( A_m \leq \frac{j}{3} \right).$$

By Lemma 7.8 and our choice of $m$ we have

$$\mathbb{P} \left( w_2([b_2 \sqrt{n}], \Pi_n) \leq 1 \right) \leq \frac{\varepsilon}{4} + e^{-Kj} = \frac{\varepsilon}{4} + e^{-K[b_2 \sqrt{n}]},$$

where $K = K(c)$. For $n$ large enough, depending on $\varepsilon$ and $c$, this is at most $\varepsilon/2$. We note that $\mathbb{P} (J_1 > [b_2 \sqrt{n}])$ only if $w_2([b_2 \sqrt{n}], \Pi_n) = 0$, giving the desired bound. \hfill \blacksquare

We now have an appropriate probabilistic bound on $J_1$ that we wish to extend to $K_1$. However, we cannot simply examine the permutation $\Pi_{J_1}$, as exposing the value of $J_1$ imposes some constraints on $\Pi_{J_1}$ so that the latter is no longer necessarily distributed as a Mallows permutation. The next lemma will show that $J_1(\Pi_x)$ is monotonically increasing in $x$. Now, without exposing $J_1$ we have $J_1(\Pi_{[b_2 \sqrt{n}]}) = \Theta(n^{-1/4})$ and $J_1(\Pi_{[b_2 \sqrt{n}]}) = \Theta(n^{-1/4})$ with probability at least $1 - 2\varepsilon$. By the mentioned monotonicity of $J_1(\Pi_x)$ we can thus sandwich $K_1$ within these bounds with probability at least $1 - 3\varepsilon$ by the previous lemma. The next two lemmas provide all details.
Lemma 7.10. For every $n \geq 3$, $2 \leq k \leq n-1$ and every $\pi \in S_n$ we have

$$J_1(\pi_k) \leq J_1(\pi_{k+1}),$$

where $\pi_k = \text{rk}(\pi(1), \ldots, \pi(k))$ and $\pi_{k+1} = \text{rk}(\pi(1), \ldots, \pi(k+1))$.

Proof. As $J_1(\pi_k) \leq k$, we need only consider the case $J_1(\pi_{k+1}) = a < k$. Then $w_2([a], \pi_{k+1}) \geq 1$ so there is some $i \in [k+1]$ such that $\pi_{k+1}(i)$ and $\pi_{k+1}(i) + 1$ are both contained in $\pi_{k+1}([a])$. This implies that $i \in [a] \subseteq [k-1]$ and $\pi_{k+1}(i) + 1 = \pi_{k+1}(a_0)$ for some $a_0 \in [a]$, which in turn implies $1_{\{\pi(k+1) < \pi(i)\}} = 1_{\{\pi(k+1) < \pi(a_0)\}}$. Thus

$$\pi_{k}(i) + 1 = \pi_{k+1}(i) + 1 - 1_{\{\pi(k+1) < \pi(i)\}} = \pi_{k+1}(a_0) - 1_{\{\pi(k+1) < \pi(a_0)\}} = \pi_{k}(a_0).$$

Thus $\pi_{k}(i) \in \pi_k([a])$ and $\pi_{k}(i) + 1 = \pi_{k}(a_0) \in \pi_k([a])$. So $J_1(\pi_k) \leq a$.

Lemma 7.11. Let $\alpha = \alpha(n)$ satisfy $|\alpha| \leq c/n$ for some constant $c > 0$. Let $q = 1 - \alpha$ and $\Pi_n \sim \text{Mallows}(n, q)$. Then for every $\varepsilon > 0$ there exist positive constants $b'_1, b'_2$ such that

$$\mathbb{P} \left( b'_1 n^{1/4} \leq K_1(\Pi_n) \leq b'_2 n^{1/4} \right) > 1 - \varepsilon,$$

provided that $n \geq n_0$ for some $n_0$ depending on $\varepsilon$ and $c$.

Proof. By Lemma 7.9 there exists $0 < b_1 < b_2$ depending on $\varepsilon$ and $c$ such that

$$\mathbb{P} \left( b_1 \sqrt{n} \leq J_1(\Pi_n) \leq b_2 \sqrt{n} \right) \geq 1 - \frac{\varepsilon}{3};$$

Denote the event above by $E$. Define $k_1 = [b_1 \sqrt{n}]$, $k_2 = [b_2 \sqrt{n}]$ and $\Pi_{k_1} = \text{rk}(\Pi_n(1), \ldots, \Pi_n(k_1))$ and similarly $\Pi_{k_2}$. Conditional on $E$, by $k_1 \leq k_2$ and Lemma 7.10 we have

$$J_1(\Pi_{k_1}) \leq K_1(\Pi_n) \leq J_1(\Pi_{k_2}). \quad (32)$$

By Lemma 3.10 we have $\Pi_{k_1} \sim \text{Mallows}(k_1, q)$ where

$$|1 - q| \leq \frac{c}{n} \leq \frac{cb_1}{k_1^2}.$$  

Similarly $\Pi_{k_2} \sim \text{Mallows}(k_2, q)$ with $|1 - q| \leq \frac{c b_2}{k_2^2}$. Lemma 7.9 says that there are $b''_1, b''_2$ depending on $\varepsilon$ and $c$ (and on $b_1$ and $b_2$ which themselves depend on $\varepsilon$ and $c$) such that

$$\mathbb{P} \left( b''_1 \sqrt{k_1} \leq J_1(\Pi_{k_1}) \right) \geq 1 - \frac{\varepsilon}{3},$$

$$\mathbb{P} \left( b''_2 \sqrt{k_2} \geq J_1(\Pi_{k_2}) \right) \geq 1 - \frac{\varepsilon}{3},$$

provided that $k_1$ and $k_2$ both exceed some lower bound depending on $\varepsilon$ and $c$. These bounds together with (32) and $\mathbb{P}(E) > 1 - \varepsilon/3$ give

$$\mathbb{P} \left( b'_1 \sqrt{b_1 n^{1/4}} \leq K_1(\Pi_n) \leq b'_2 \sqrt{b_2 n^{1/4}} \right) \geq 1 - \varepsilon.$$

Lemma 7.12. There exists a TOTOT formula $\psi$ such that the following holds: Let $\alpha = \alpha(n)$ satisfy $|\alpha| \leq c/n$ for some constant $c > 0$, and $\Pi_n \sim \text{Mallows}(n, 1 - \alpha)$. Then for every $\varepsilon > 0$ there exist positive constants $b'_1, b'_2$ such that with probability at least $1 - \varepsilon$, there is a unique $j$ such that $\Pi_n \models \psi[j]$, and this unique $j$ satisfies the bounds

$$b'_1 n^{1/4} \leq j \leq b'_2 n^{1/4}, \quad (33)$$

provided that $n \geq n_0$ for some $n_0$ depending on $\varepsilon$ and $c$.  

41
Proof. We have already found a formula that determines whether \( i \in W_2([j], \Pi_n) \) in (18), call this formula \( \zeta(i, j) \). Then \( J_1 = j_1 \in [n] \) if and only if \( \Pi_n \models \zeta[j_1] \) where
\[
\zeta(x) := \exists z (\zeta(z, x)) \land \forall w \, (w < 1 \, x \rightarrow \neg \exists z \, (\zeta(z, w))).
\]

We now define the formula
\[
\psi(x) := \exists z (\xi(z) \land (x \leq z) \land \xi\hat{=} (z, x)),
\]
where \( \xi\hat{=} \) is the relativization of \( \xi \), see Lemma 3.24. By Lemma 3.26 \( K_1 = j \in [n] \) if and only if \( \Pi_n \models \psi[j] \), and such a unique \( j \) always exists (as long as \( n \geq 2 \)). By Lemma 7.11 there are \( b_1', b_2' \) depending on \( \varepsilon \) and \( c \) such that this \( j \) also satisfies (33) with probability at least \( 1 - \varepsilon \). ■

**Proposition 7.13.** Let \( \alpha = \alpha(n) \) satisfy \( |\alpha| \leq c/n \) for some constant \( c > 0 \). Let \( q = 1 - \alpha \) and \( \Pi_n \sim \text{Mallows}(n, q) \). Then there is a sentence \( \omega \in \text{TOTO} \) such that for all \( n \) satisfying \( W(2)(\log^* \log^* n - 1) < \log^* n \) we have\[
P(\Pi_n \models \omega) = \begin{cases} 
> \frac{n}{m} - O(n^{-1/4}) & \text{if } \log^* \log^* n \text{ is even}, \\
< \frac{n}{m} + O(n^{-1/4}) & \text{if } \log^* \log^* n \text{ is odd},
\end{cases}
\]
where the implicit constants in the \( O(n^{-1/4}) \) terms depend on \( c \).

Proof. Let \( \psi \) be as in Lemma 7.12 and let \( \varphi \) be the sentence provided by Proposition 7.1. We define the sentence
\[
\omega := \exists z (\psi(z) \land \varphi\hat{=} (z)),
\]
where \( \varphi\hat{=} \) is the relativization of \( \varphi \) as in Lemma 3.24.

Set \( \varepsilon = \frac{1}{2n} \). Let \( 0 < b_1' < b_2' \), both depending on \( c \), be as in Lemma 7.12 Let \( E \) be the event that there is a unique \( j \) such that \( \Pi_n \models \psi[j] \) and that this unique \( j \) satisfies (33). We have \( P(E) > \frac{n}{10} \) by Lemma 7.12. Letting \( \Pi_j = \text{rk}(\Pi_n(1), \ldots, \Pi_n(j)) \) if \( E \) holds, we have by Lemma 3.26 that
\[
P(\Pi_n \models \omega) \geq \frac{9}{10} \cdot P(\Pi_j = \varphi \mid E),
P(\Pi_n \not\models \omega) \geq \frac{9}{10} \cdot P(\Pi_j \neq \varphi \mid E).
\]

We will first show that conditional on \( E \), the condition \( W(2)(\log^* \log^* n - 1) < \log^* n \) implies that \( \log^* \log^* n = \log^* \log^* j \) and \( W(2)(\log^* \log^* j - 1) < \log \log j \), at least for \( n \) larger than some lower bound depending on \( b_1' \) and \( b_2' \). Note that \( W(2)(\log^* \log^* n - 1) < \log^* n \) implies
\[
\log^* \log^* n - 1 < \log^* \log^* \log^* n \leq \log^* \log^* n.
\]
This forces the second inequality to be an equality. If \( n \geq j \geq b_1' n^{1/4} > \log^* n \), then conditional on \( E \) we indeed have that \( \log^* \log^* j = \log^* \log^* n \). For \( n \) exceeding some lower bound depending on \( b_1' \), we also have \( \log^* n < \log \log b_1' n^{1/4} \), so that in this case the event \( E \) implies
\[
W(2)(\log^* \log^* n - 1) = W(2)(\log^* \log^* j - 1) < \log^* n < \log \log b_1' n^{1/4} \leq \log \log j.
\]
Consider then the case \( \log^* \log^* n \) even and let \( \ell := [b_1' n^{1/4}] \) and \( u := [b_2' n^{1/4}] \). For every \( \ell \leq m \leq u \) we have \( \Pi_m := \text{rk}(\Pi_n(1), \ldots, \Pi_n(m)) \sim \text{Mallows}(m, q) \), where
\[
|1 - q| \leq \frac{c}{n} \leq \frac{cb_2'}{m^4}.
\]
Thus, Proposition 7.1 yields
\[
P(\Pi_j \models \varphi \mid E) \geq P \left( \bigwedge_{m=\ell}^u \Pi_m \models \varphi \right) \geq 1 - \sum_{m=\ell}^u P(\Pi_m \not\models \varphi) \geq 1 - u O(c b_2' \ell^{-2}). \tag{34}
\]
We conclude that
\[ P\left(\Pi_n \models \omega\right) \geq \frac{9}{10} - O(n^{-1/4}). \]

We consider now the case \( \log^* \log^* n \) odd. Again, conditioning on \( E \), we are in the second case of Proposition 7.1 for \( n \) large enough. We again have \( P\left(\Pi_n \models \omega\right) \geq \frac{9}{10} P\left(\Pi_n \models \varphi \mid E\right) \). The same calculations as in (34) show that in this case \( P\left(\Pi_n \models \omega\right) > \frac{9}{10} - O(n^{-1/4}). \)

\[ \Box \]

7.3 Lifting to \(|1 - q| < 1/\log^* n\)

In the section we finalize the proof of Theorem 1.2 Part (ii) by extending the results of the previous sections to \( 1 - \frac{1}{\log^* n} < q < 1 + \frac{1}{\log^* n} \). We will proceed as follows: We partition the interval \((1 - \frac{1}{\log^* n}, 1 + \frac{1}{\log^* n})\) into the parts

\[ D_1 := \left(1 - \frac{1}{\log^* n}, 1 - \frac{c}{n}\right), \quad D_2 := \left[1 - \frac{c}{n}, 1 + \frac{c}{n}\right], \quad D_3 := \left(1 + \frac{c}{n}, 1 + \frac{1}{\log^* n}\right). \]

With foresight we set \( c = 50 \). We will find TOTO sentences \( \rho, \xi_1 \) and \( \xi_2 \) and disjoint infinite sets \( A^-, A^+ \subseteq \mathbb{N} \) such that all bounds in Table 1 hold. Before we show the existence of such sentences with the desired properties, we show how they imply Theorem 1.2 Part (ii).

| \( n > 50 \) | \( \mathbb{P}(\Pi_n \models \rho) \) | \( |1 - q| \leq \frac{50}{n} \) | \( 1 - \frac{50}{n} < q < 1 - \frac{1}{\log^* n} \) | \( 1 + \frac{50}{n} < q < 1 + \frac{1}{\log^* n} \) |
|---|---|---|---|---|
| \( n \in A^+ \) | \( \mathbb{P}(\Pi_n \models \xi_1) \) | \( \geq 0.8 \) | \( \geq 0.8 \) | \( \geq 0.8 \) |
| \( \mathbb{P}(\Pi_n \models \xi_2) \) | \( \geq 0.8 \) | \( \geq 0.8 \) | \( \geq 0.8 \) |
| \( n \in A^- \) | \( \mathbb{P}(\Pi_n \models \xi_1) \) | \( \leq 0.2 \) | \( \leq 0.2 \) | \( \leq 0.2 \) |
| \( \mathbb{P}(\Pi_n \models \xi_2) \) | \( \leq 0.2 \) | \( \leq 0.2 \) | \( \leq 0.2 \) |

**Table 1**: We find TOTO sentences \( \rho, \xi_1 \) and \( \xi_2 \) and two disjoint infinite sets \( A^-, A^+ \subseteq \mathbb{N} \) such that all bounds in Table 1 hold for all \( n \in A^- \cup A^+ \). Let

\[ \varphi := (\rho \rightarrow \xi_2) \land ((\neg \rho) \rightarrow \xi_1). \]

As \( \Pi_n \) satisfies exactly one of \( \rho \) and \( \neg \rho \), any of the three events

\[ E_1 := \{\Pi_n \models (\neg \rho) \land \xi_1\}, \quad E_2 := \{\Pi_n \models (\xi_1 \land \xi_2)\}, \quad \text{or} \quad E_3 := \{\Pi_n \models \rho \land \xi_2\} \]

imply that \( \Pi_n \models \varphi \). For \( n \in A^+ \), by the bounds in Table 1 we have that

- if \( q \in D_1 \) then \( \mathbb{P}(E_1) \geq 1 - 0.1 - 0.2 = 0.7 \),
- if \( q \in D_2 \) then \( \mathbb{P}(E_2) \geq 1 - 0.2 - 0.2 = 0.6 \),
- if \( q \in D_3 \) then \( \mathbb{P}(E_3) \geq 1 - 0.1 - 0.2 = 0.7 \).

In general, for \( n \in A^+ \) and \( q \in (1 - \frac{1}{\log^* n}, 1 + \frac{1}{\log^* n}) = D_1 \cup D_2 \cup D_3 \) we have \( \mathbb{P}(\Pi_n \models \varphi) \geq 0.6 \).

We now consider \( n \in A^- \). Note that the three events

\[ F_1 := \{\Pi_n \models \rho \land \Pi_n \models \xi_1\}, \quad F_2 := \{\Pi_n \models \xi_1 \lor \xi_2\} \quad \text{and} \quad F_3 := \{\Pi_n \models \rho \land \Pi_n \models \xi_2\} \]

now each imply that \( \Pi_n \models \varphi \). For \( n \in A^- \) the bounds in Table 1 now give that

- if \( q \in D_1 \) then \( \mathbb{P}(F_1) \geq 1 - 0.1 - 0.2 = 0.7 \),
- if \( q \in D_2 \) then \( \mathbb{P}(F_2) \geq 1 - 0.2 - 0.2 = 0.6 \),
- if \( q \in D_3 \) then \( \mathbb{P}(F_3) \geq 1 - 0.1 - 0.2 = 0.7 \).

43
So for $n \in A^-$ and $q \in (1 - 1/\log^* n, 1 + 1/\log^* n)$ we have $\mathbb{P}(\Pi_n \not\models \varphi) \geq 0.6$.

As both $A^-$ and $A^+$ have infinite cardinality, we conclude that $\mathbb{P}(\Pi_n \models \varphi)$ does not have a limit as $n \to \infty$ for any sequence $q = q(n)$ satisfying $1 - 1/\log^* n < q < 1 + 1/\log^* n$. \hfill \blacksquare

So it remains to find $\rho, \xi_1, \xi_2, A^-$ and $A^+$ and prove the bounds in Table $1$. We start by defining the TOTO sentence $\rho$ as

$\rho = \exists xy \left( \neg \exists w(\text{succ}_1(w, x) \vee \text{succ}_1(y, w)) \land (y <_2 x) \right)$.

If $n \geq 1$, then there always exist unique $x$ and $y$ such that $\neg \exists w(\text{succ}_1(w, x) \vee \text{succ}_1(y, w))$, namely $x = 1$ and $y = n$ the largest element in $[n]$. Then $\Pi_n \models \rho$ precisely when $\Pi_n(1) > \Pi_n(n)$.

**Lemma 7.14.** Let $\Pi_n \sim \text{Mallows}(n, q)$. For every $n > 50$ we have

$$
\begin{align*}
\mathbb{P}(\Pi_n \not\models \rho) &\geq 0.9 & \text{if } q < 1 - \frac{50}{n}, \\
\mathbb{P}(\Pi_n \models \rho) &\geq 0.9 & \text{if } q > 1 + \frac{50}{n}.
\end{align*}
$$

**Proof.** Recall once more the definition of $\text{TruncGeo}(n, p)$ random variables as in [4], valid for $p \in (-\infty, 0) \cup (0, 1)$.

We first consider the case $q > 1 + 50/n$. Recall that we defined the mapping $r_n : [n] \to [n]$ by $r_n(i) = i - n - i + 1$. By $r_n(1) = n$ we may estimate

$$
\mathbb{P}(\Pi_n(1) \geq \frac{n}{3} + 1) = \mathbb{P}\left((r_n \circ \Pi_n \circ r_n)(1) \leq n - \frac{n}{3}\right) = \mathbb{P}\left(\Pi_n(1) \leq \frac{2n}{3}\right),
$$

where we used the fact that $r_n \circ \Pi_n \circ r_n \overset{d}{=} \Pi_n$. Using this estimate we have

$$
\mathbb{P}(\Pi_n \models \rho) \geq \mathbb{P}\left(\Pi_n(1) > \frac{2n}{3} \text{ and } \Pi_n(n) < \frac{n}{3} + 1\right) \geq 1 - 2 \cdot \mathbb{P}\left(\Pi_n(1) \leq \frac{2n}{3}\right).
$$

The value $\Pi_n(1)$ follows a $\text{TruncGeo}(n, 1 - q)$ distribution, so that

$$
\begin{align*}
\mathbb{P}(\Pi_n \models \rho) &> 1 - 2 \cdot \frac{q - 1}{q^n - 1} \sum_{i=0}^{\lfloor 2n/3 \rfloor} q^i \\
&= 1 - 2 \cdot \frac{q^{\lfloor 2n/3 \rfloor + 1} - 1}{q^n - 1} \\
&> 1 - 2 \cdot \frac{q^{3n/4} - 1}{q^n - 1}.
\end{align*}
$$

The last inequality holds as $n > 50$ implies that $\lfloor 2n/3 \rfloor + 1 < 3n/4$. Now, for all $1 < x < y$ it holds that $\frac{y-1}{y-1} < \frac{x}{y}$, so that $1 < q < q^n$ implies

$$
\mathbb{P}(\Pi_n \models \rho) > 1 - 2q^{-n/4} > 1 - 2 \left(1 + \frac{50}{n}\right)^{-n/4}.
$$

By $n > 50$, the latter quantity is at least $1 - 2 \cdot 2^{-n/4} > 1 - 2 \cdot 2^{-12.5} > 0.9$.

Now suppose that $q < 1 - 50/n$ and that $n > 50$. Then

$$
\begin{align*}
\frac{1}{q} > &\frac{1}{1 - \frac{50}{n}} = \sum_{i=0}^{\left\lfloor \frac{50}{n} \right\rfloor} \left(\frac{50}{n}\right)^i > 1 + \frac{50}{n}.
\end{align*}
$$

Recalling that for $\Pi_n \sim \text{Mallows}(n, q)$ we have $r_n \circ \Pi_n \overset{d}{=} \text{Mallows}(n, 1/q)$ gives

$$
\mathbb{P}(\Pi_n \not\models \rho) = \mathbb{P}(r_n \circ \Pi_n \models \rho) > 0.9.
$$

\hfill \blacksquare

44
We now build up a collection of results that will allow us to find $\xi_1$ and $\xi_2$ with the desired properties. For the next few lemmas we will consider $c = 100$ instead of 50 to simplify some of the manipulations later on.

**Lemma 7.15.** Let $q = q(n)$ be such that $1 - 100/n \le q \le 1 + 100/n$. Then there exists an absolute constant $c_1 > 0$ such that for $\Pi_n \sim \text{Mallows}(n,q)$ and $n \ge 400$ we have

$$\mathbb{P}(c_1 n \le \Pi_n^{-1}(1)) > 0.99.$$  

*Proof.* If $q = 1$, then $\Pi_n \sim \text{Mallows}(n,q)$ is uniformly distributed over $S_n$. Then $\Pi_n^{-1}(1)$ is uniformly distributed over $[n]$ and such $c_1$ clearly exists.

So let $q \neq 1$. As before, we note that $\Pi_n^{-1}(1) \triangleq \Pi_n$ so that also $\Pi_n^{-1}(1) \triangleq \Pi_n(1) \triangleq Z_1$ where $Z_1 \sim \text{TruncGeo}(n, 1 - q)$. By Lemma 7.4 there is a $c_U > 0$ such that for $n \ge 400$ we have

$$\mathbb{P}(Z_1 < c_1 n) \le c_U \left(\frac{c_1 n}{n}\right).$$

For $c_1$ small enough this is less than 0.01. ■

**Lemma 7.16.** There exist constants $0 < c_1 < c_2$ and $n_0$ such that for $n \ge n_0$ and any sequence $q = q(n)$ satisfying $1 - 1/\log^* n < q < 1 - 100/n$, the random permutation $\Pi_n \sim \text{Mallows}(n,q)$ satisfies

$$\mathbb{P}\left(\frac{c_1}{1 - q} \le \Pi_n^{-1}(1) \le \frac{c_2}{1 - q}\right) > 0.99.$$  

*Proof.* We again have $\Pi_n^{-1}(1) \triangleq Z_1$ where $Z_1 \sim \text{TruncGeo}(n, 1 - q)$. Therefore the inequalities $q < 1 - 100/n < e^{-100}/n < 1$ imply that

$$\mathbb{P}\left(\Pi_n^{-1}(1) < \frac{c_1}{1 - q}\right) \le \frac{1 - q}{1 - q^n} \sum_{i=1}^{\left\lfloor \frac{c_1}{q} \right\rfloor} q^{i-1} < \frac{c_1}{1 - q^n} < \frac{c_1}{1 - e^{-100}}.$$  

Clearly the above is less than 0.005 for $c_1$ small enough.

For the remaining bound, Markov’s inequality gives

$$\mathbb{P}\left(\Pi^{-1}(1) > \frac{c_2}{1 - q}\right) = \mathbb{P}\left(\Pi^{-1}(1) - 1 > \frac{c_2 - 1 + q}{1 - q}\right) \le \frac{(1 - q)\mathbb{E}(\Pi_n^{-1}(1) - 1)}{c_2 - 1 + q}.$$  

Theorem 3.15 implies that $\mathbb{E}(\Pi_n^{-1}(1) - 1) \le \min\left(\frac{2q}{1 - q}, n - 1\right) \le \frac{2}{1 - q}$. So

$$\mathbb{P}\left(\Pi^{-1}(1) > \frac{c_2}{1 - q}\right) \le \frac{2}{c_2 - 1 + q} < \frac{2}{c_2 - \frac{2}{\log^* n}}.$$  

For $n$ and $c_2$ large enough the above is less than 0.005, finishing the proof. ■

We now combine the previous two lemmas into a form that will be convenient for us.

**Lemma 7.17.** There exist absolute constants $n_1, c_1, c_2$ and a TOTO formula $\lambda(y)$ with one free variable $y$ such that the following holds: For $1 - 1/\log^* n < q < 1 + 100/n$ and $\Pi_n \sim \text{Mallows}(n,q)$, if $n \ge n_1$, then

(i) $\Pi_n \models \lambda[x]$ if and only if $\Pi_n(x + 1) = 1$;

(ii) $\mathbb{P}\left(\exists x \text{ s.t. } \Pi_n(x + 1) = 1 \text{ and } x \ge c_1 \log^* n \text{ and } |1 - q| \le \frac{c_2}{x}\right) \ge 0.99. \quad (35)$
Proof. Let \( \lambda(y) \) be the formula
\[
\lambda(y) := \exists z \ (\text{succ}_1(y, z) \land \neg \exists w \ (w <_2 z)).
\]
Then \( \Pi_n \models \lambda[N] \) if and only if the successor under \( <_1 \) of \( N \) is the smallest number under the ordering \( <_2 \), that is, if and only if \( \Pi_n(N + 1) = 1 \). So let \( N = \Pi_n^{-1}(1) - 1 \).

We consider two cases. First suppose that \( 1 - 100/n < q < 1 - 100/n \). By Lemma 7.17 there is a \( c_1 \) such that for \( n \) large enough we have \( N \geq c_1 n - 1 > c_1 \log^* n \) with probability at least 0.99. Moreover, by \( N \leq n \) we have \( |1 - q| \leq 100/n \leq 100/N \). This finishes the first case.

Suppose then that \( 1 - 1/\log^* n < q < 1 - 100/n \). By Lemma 7.16 there are absolute \( c_1, c_2 \) such that with probability at least 0.99 we have
\[
\frac{c_1}{1 - q} \leq N + 1 \leq \frac{c_2}{1 - q}.
\]
The first inequality gives \( N \geq \frac{c_1}{1 - q} - 1 > c_1 \log^* n - 1 \). For \( n \) large enough (depending on the absolute constant \( c_1 \)) this is at least, say, \( \frac{c_1}{2} \log^* n \). The second inequality gives \( 0 < 1 - q \leq c_2/N \).
This finishes the second case as well.

We will need the following algebraic lemma before we can turn to the existence of \( \xi_1, \xi_2 \) as advertised at the start of this section. Recall the definitions of \( W(\cdot) \) and \( T(\cdot) \) from Section 3.4

**Lemma 7.18.** Let \( c_1 \) be as given in Lemma 7.17. Then there exists an \( m \in \mathbb{N} \) such that the following holds: Let \( t \geq m \), set \( n := W^{(2)}(t) \) and let \( N \) satisfy \( c_1 \log^* n \leq N \leq n \). Then
\[
W^{(2)}(\log^{**} \log^* N - 1) < \log^* N,
\]
and
\[
\log^{**} \log^* N = t.
\]

**Proof.** First, let \( m > 3 \) so that by Lemma 3.29 the inequality
\[
T^{(3)} \circ W^{(2)}(t - 1) < W^{(2)}(t)
\]
holds for all \( t \geq m \). We may also select \( m \) so large that for all \( t \geq m \)
\[
c_1 T^{(2)} \circ W^{(2)}(t - 1) > T \circ W^{(2)}(t - 1).
\]
We claim that any \( m \) satisfying these two conditions is as desired. To see this, let \( t \geq m \), \( n = W^{(2)}(t) \) and \( c_1 \log^* n \leq N \leq n \). The inequalities (37) and (36) imply that
\[
N \geq c_1 \log^* W^{(2)}(t) > c_1 \log^* (T^{(3)} \circ W^{(2)}(t - 1)) > T \circ W^{(2)}(t - 1).
\]
Thus
\[
\log^* W^{(2)}(t - 1) = W^{(2)}(\log^* \log^* N - 1).
\]
As we trivially have \( \log^{**} \log^* N \leq \log^* \log^* n = t \), we conclude that \( \log^{**} \log^* N = t \). Moreover, (38) also implies that
\[
\log^* N > W^{(2)}(t - 1) = W^{(2)}(\log^{**} \log^* N - 1).
\]

For what follows, we let \( S_0 \) be the set containing only the empty permutation; that is, the \text{TOTO} structure with empty domain. We let Mallows(0,q) be the degenerate distribution with support \( S_0 \),

46
Lemma 7.19. Let $0 < q \leq 1$ and $\Pi_n \sim \text{Mallows}(n,q)$. If $N := \Pi_n^{-1}(1) - 1$ and $\Pi_N := \text{rk}(\Pi_n(1),\ldots,\Pi_n(N))$, then

$$\Pi_N \overset{d}{=} \text{Mallows}(N,q),$$

(I.e., we generate a Mallows distribution with $n$ random.)

Proof. Let $0 \leq m < n$ and $\pi \in S_m$, we will determine the distribution of $\Pi_N$ conditional on $N = m$. If $m = 0$, then we trivially have $\Pi_m \sim \text{Mallows}(m,q)$, so we may consider $m \geq 1$. We have for each $\pi \in S_m$

$$\mathbb{P}(\Pi_m = \pi \mid N = m) = \frac{\mathbb{P}(\Pi_m = \pi \text{ and } N = m)}{\mathbb{P}(N = m)}.$$

If $N = m$, then $\Pi(m + 1) = 1$. So the two events $\{\Pi_m = \pi\}$ and $\{N = m\}$ imply that $(i,m)$ is an inversion of $\Pi_{m+1}$ for every $i \leq m$. That is, conditioning on these two events we have $\text{inv}(\Pi_{m+1}) = \text{inv}(\pi) + m$. We obtain that

$$\mathbb{P}(\Pi_m = \pi \mid N = m) = \frac{q^{\text{inv}(\pi)}}{F(m,q)},$$

for some function $F(m,q)$. In fact, summing the above over all $\pi \in S_m$ gives one, so that $F(m,q)$ necessarily equals $\sum_{\pi \in S_m} q^{\text{inv}(\pi)}$. \hfill \blacksquare

We are now ready to show the existence of $\xi_1$ satisfying the bounds in Table 1.

Lemma 7.20. There exists a sentence $\xi_1 \in \text{TOTO}$ and two infinite disjoint sets $A^+, A^- \subseteq \mathbb{N}_{>100}$ such that for

$$1 - \frac{1}{\log^* n} \leq q < 1 + \frac{100}{n}$$

and $\Pi_n \sim \text{Mallows}(n,q)$ we have

$$\mathbb{P}(\Pi_n \models \xi_1) > 0.8, \quad \text{if } n \in A^+,$$

$$\mathbb{P}(\Pi_n \models \xi_1) < 0.2, \quad \text{if } n \in A^-.$$

Proof. Let $n_1, c_1, c_2$ and $\lambda(y)$ be as provided by Lemma 7.17, and let the event $E$ be as in (35). Let $m$ be as given in Lemma 7.18. Without loss of generality, we may assume that $m$ is so large that $W(2)(m) > n_1$. Define the TOTO sentence $\xi_1$ as

$$\xi_1 = \exists z (\lambda(z) \wedge \omega^0(z)),$$

where $\omega$ is the sentence $\omega(c_1)$ as given in Proposition 7.13 (using the absolute constant $c_1$), and $\omega^0$ is its relativization as in Lemma 3.24. Then by Lemma 3.26

$$\mathbb{P}(\Pi_n \models \xi_1 \mid E) = \mathbb{P}(\Pi_N \models \omega \mid E).$$

Let $t \geq m$ be even and set $n = W(2)(t)$. Then for all $c_1 \log^* n \leq N < n$ we have by Lemma 7.18 that $N$ satisfies the conditions of the first case in Proposition 7.13 so that Lemma 7.19 yields

$$\mathbb{P}(\Pi_N \models \omega \mid E) > \frac{9}{10} - O \left( (\log^* n)^{-1/4} \right).$$

For $n$ large enough (i.e., $t$ large enough), this is at least 0.81. As $\mathbb{P}(E) > 0.99$ we have that $\mathbb{P}(\Pi_n \models \xi_1) > 0.8$.

Completely analogously, when $t$ is odd, then $\mathbb{P}(\Pi_N \models \omega \mid E) < 0.19$. Using $\mathbb{P}(E) > 0.99$ we again conclude that there are also infinitely many $n = W(2)(t)$ with $t$ now odd, such that $\mathbb{P}(\Pi_n \models \xi_1) < 0.2$. \hfill \blacksquare
Remark 7.21. As alluded to earlier, Lemma 7.20 is slightly stronger than needed: we only require the bounds for the smaller domain $P (\Pi_n \models \xi_1)$ for $1 - 1/\log^* n < q < 1 + 50/n$.

The last lemma of this section gives the existence of $\xi_2$ with the desired specifications, finishing the proof of Theorem 1.2 Part (ii).

Lemma 7.22. Let $A^-$ and $A^+$ be as in Lemma 7.20. There exists a TOTO sentence $\xi_2$ such that for

$$1 - \frac{50}{n} \leq q < 1 + \frac{1}{\log^* n}$$

and $\Pi_n \sim \text{Mallows}(n, q)$ we have

$$P (\Pi_n \models \xi_2) > 0.8, \quad \text{if } n \in A^+,$$

$$P (\Pi_n \models \xi_2) < 0.2, \quad \text{if } n \in A^-.$$

Proof. Let $\xi_2 := \xi_1^{\text{reverse}}$, where $\xi_1$ is as provided by Lemma 7.20. Let $1 - \frac{50}{n} \leq q < 1 + \frac{1}{\log^* n}$.

Then

$$\frac{1}{q} \leq \frac{1}{1 - \frac{50}{n}} = 1 + \frac{50}{n - 50}.$$

If $n \geq 100$ then this is at most $1 + 100/n$, as is simple to verify. Moreover,

$$\frac{1}{q} > 1 + \frac{1}{\log^* n} = 1 - \frac{1}{\log^* n + 1} > 1 - \frac{1}{\log^* n}.$$

Recall from Lemma 7.20 that $A^-$ and $A^+$ are subsets of $N_{>100}$. By Lemma 3.27 and Lemma 7.20 we have for $\Pi_n \sim \text{Mallows}(n, q)$ that

$$P (\Pi_n \models \xi_2) = P (r_n \circ \Pi_n \models \xi_1) = \begin{cases} > 0.8 & \text{if } n \in A^+, \\ < 0.2 & \text{if } n \in A^-. \end{cases}$$

Acknowledgements

We thank Kevin Compton, Valentin Féray, Tomasz Łuczak, Marc Noy and Joel Spencer for helpful discussions and pointers to the literature.

References


