Limits of permutation sequences [☆],[☆☆],[☆☆☆]

Carlos Hoppen [a], Yoshiharu Kohayakawa [b], Carlos Gustavo Moreira [c], Balázs Ráth [d], Rudini Menezes Sampaio [e]

[a] *Instituto de Matemática, UFRGS – Avenida Bento Gonçalves, 9500, 91509-900 Porto Alegre, RS, Brazil*
[b] *Instituto de Matemática e Estatística, USP – Rua do Matão 1010, 05508-090 São Paulo, SP, Brazil*
[c] *IMPA – Estrada Dona Castorina 110, 22460-320 Rio de Janeiro, RJ, Brazil*
[d] *Department of Mathematics, University of British Columbia, 1984 Mathematics Road, V6T1Z2 Vancouver, BC, Canada*
[e] *Departamento de Computação, Centro de Ciências, UFC – Campus do Pici, Bloco 910, 60451-760 Fortaleza, CE, Brazil*

ARTICLE INFO

ABSTRACT

A permutation sequence $(\sigma_n)_{n \in \mathbb{N}}$ is said to be convergent if, for every fixed permutation τ, the density of occurrences of τ in the elements of the sequence converges. We prove that such a convergent sequence has a natural limit object, namely a Lebesgue measurable function $Z : [0, 1]^2 \to [0, 1]$ with the additional properties that, for every fixed $x \in [0, 1]$, the restriction $Z(x, \cdot)$ is a cumulative distribution function and, for every $y \in [0, 1]$, the restriction $Z(\cdot, y)$ satisfies a "mass" condition. This limit process is well-behaved: every function in the class of limit objects is a limit of some permutation sequence, and two of these functions are limits of the same sequence if and only if they are equal almost everywhere. An ingredient in the proofs is a new model of random permutations, which generalizes previous models and might be interesting for its own sake.

© 2012 Elsevier Inc. All rights reserved.

1. Introduction

As usual, a *permutation* of a finite set X is a bijective function of X into itself. We shall focus on permutations σ on the set $X = \{1, \ldots, n\} = [n]$, where n is a positive integer, called the *length* of σ, and is denoted by $|\sigma|$. In this work, a permutation σ on $[n]$ is represented by $\sigma = (\sigma(1), \ldots, \sigma(n))$, and the set of all permutations on $[n]$ is denoted by S_n. We denote by $\mathcal{S} = \bigcup_{i=1}^{\infty} S_n$ the set of all finite permutations. A *graph* $G = (V, E)$ is given by its *vertex set* V and its *edge set* $E \subseteq \{\{u, v\} \subset V \colon u \neq v\}$.

The main goal of this paper is to introduce a notion of convergence of a permutation sequence $(\sigma_n)_{n \in \mathbb{N}}$ and to identify a natural limit object for such a convergent sequence whose associated sequence of lengths $(|\sigma_n|)_{n \in \mathbb{N}}$ tends to infinity. Lovász and Szegedy [21] were concerned with these questions in the case of graph sequences $(G_n)_{n \in \mathbb{N}}$. This has been further investigated by Borgs et al. in [5] and [6], where, among other things, limits of graph sequences were used to characterize the testability of graph parameters. The convergence of sequences of combinatorial objects has also been addressed in other structures. For instance, graphs with degrees bounded by a constant have been addressed in recent work of Benjamini and Schramm [2] and of Elek [10,11]. Elek and Szegedy [12] studied this problem for hypergraphs. See Lovász [19] for a comprehensive survey of this area.

Currently, the main application of our results in this paper is in property testing of permutations. Roughly speaking, the objective of testing is to decide whether a combinatorial structure satisfies some property, or to estimate the value of some numerical function associated with this combinatorial structure, by considering only a randomly chosen substructure of sufficiently large, but constant size. These problems are called *property testing* and *parameter testing*, respectively; a property or parameter is said to be *testable* if it can be estimated accurately in this way. The algorithmic appeal of testability is evident, as, conditional on sampling, this leads to reliable constant-time randomized estimators for the said properties or parameters. In [15] four of the present authors address these questions through the prism of subpermutations. Among their main results are a permutation result in the direction of Alon and Shapira's [1] work on the testability of hereditary graph properties, and a permutation counterpart of the characterization of testable parameters by Borgs et al. [4].

Given the similarity of our results with the ones obtained in [21], we briefly describe that work. Central in the arguments is the notion of a *homomorphism* of a graph F into a graph G, a function $\phi \colon V(F) \to V(G)$ that maps the vertex set $V(F)$ of F into the vertex set $V(G)$ of G with the property that, for every edge $\{u, v\}$ in F, the pair $\{\phi(u), \phi(v)\}$ is an edge in G. The number of homomorphisms of F into G is denoted by $\hom(F, G)$, while the *homomorphism density* of F into G is given by the probability that a uniformly chosen ϕ is a homomorphism:

$$t(F, G) = \frac{\hom(F, G)}{|V(G)|^{|V(F)|}}. \tag{1}$$

It is natural to measure the similarity between two graphs G and G' by comparing the homomorphism density of different graphs F into them. This suggests defining a graph sequence $(G_n)_{n \in \mathbb{N}}$ as being *convergent* if, for every (simple) graph F, the sequence of real numbers $(t(F, G_n))_{n \in \mathbb{N}}$ converges. Lovász and Szegedy identify a class of natural limit objects for such convergent sequences, which they call *graphons*, in the form of symmetric Lebesgue measurable functions $W \colon [0, 1]^2 \to [0, 1]$. One might heuristically imagine the adjacency matrix of a simple graph as a black-and-white television screen (a white pixel at position (i, j) represents an edge between vertices i and j), a graph sequence as a sequence of TV sets with higher and higher resolution, and the limiting graphon W as the "perfect TV" where each point $(x, y) \in [0, 1]^2$ is a "pixel of infinitesimal size" and the color of each infinitesimal pixel can be any shade of grey between black and white.

An important feature of this limit object is that it may be used to generate random graphs: given a graphon $W \colon [0, 1]^2 \to [0, 1]$ and a positive integer n, a W-*random* graph $G(n, W)$ with vertex set $[n]$ is generated as follows. First, n real numbers $X_1, \ldots, X_n$ are generated independently according to the uniform probability distribution on the interval $[0, 1]$. Then, for every pair of distinct vertices i and j in $[n]$, the pair $\{i, j\}$ is added to the edge set of the graph independently with probability $W(X_i, X_j)$. Heuristically, the adjacency matrix of $G(n, W)$ looks like the graphon W if $1 \ll n$: we approximate the perfect TV screen by choosing the coordinates of infinitesimal pixels of W at random and declaring a pixel of $G(n, W)$ white with a probability proportional to the greyness of the corresponding

infinitesimal pixel of W. It is important to point out that this model of random graphs generalizes the random graph model $G(n, H)$ (see Lovász and Sós [20]), further generalizing the classical model $G_{n,p}$ due to Erdős and Rényi [13] and Gilbert [14].

With this, Lovász and Szegedy define the *homomorphism density* of a k-vertex graph F into a graphon W as the probability $t(F, W)$ that F is a subgraph of the W-random graph $G(k, W)$, which can be calculated as follows given the above definition of $G(k, W)$:

$$t(F, W) = \int_0^1 \dots \int_0^1 \prod_{\{i, j\} \in E(F)} W(x_i, x_j) \, dx_1 \dots dx_k. \tag{2}$$

They use this to prove that, with each convergent graph sequence $(G_n)_{n \in \mathbb{N}}$, one may associate a graphon W such that, for every fixed F,

$$\lim_{n \to \infty} t(F, G_n) = t(F, W).$$

The graphon W is said to be a *limit* to $(G_n)_{n \in \mathbb{N}}$. Conversely, Lovász and Szegedy show that, for any fixed graphon W, the randomized sequence $(G(n, W))_{n \in \mathbb{N}}$ converges to W with probability one. Hence, given a graphon W, there exists a graph sequence $(G_n(W))_{n \in \mathbb{N}}$ converging to W.

In our paper, a similar path is traced for permutation sequences $(\sigma_n)_{n \in \mathbb{N}}$. The rôle of the homomorphism density $t(F, G_n)$ of a fixed graph F into G_n is played here by the subpermutation density $t(\tau, \sigma_n)$ of a fixed permutation τ into σ_n, which we now define. By $[n]^m_<$ we mean the set of m-tuples in $[n]$ whose elements are in strictly increasing order.

Definition 1.1 *(Subpermutation density).* For positive integers $k, n \in \mathbb{N}$, let $\tau \in S_k$ and $\pi \in S_n$. The *number of occurrences* $\Lambda(\tau, \pi)$ of the permutation τ in π is the number of k-tuples $(x_1, x_2, \dots, x_k) \in [n]^k_<$ such that $\pi(x_i) < \pi(x_j)$ if and only if $\tau(i) < \tau(j)$. The *density* of the permutation τ as a *subpermutation* of π is given by

$$t(\tau, \pi) = \begin{cases} \binom{n}{k}^{-1} \Lambda(\tau, \sigma) & \text{if } k \leqslant n, \\ 0 & \text{if } k > n. \end{cases} \tag{3}$$

As an illustration, the permutation $\tau = (3, 1, 4, 2)$ occurs in $\pi = (5, 6, 2, 4, 7, 1, 3)$, since π maps the index set $(1, 3, 5, 7)$ onto $(5, 2, 7, 3)$, which appears in the relative order given by τ. This concept may be used to define a convergent permutation sequence in a natural way.

Definition 1.2 *(Convergence of a permutation sequence).* A permutation sequence $(\sigma_n)_{n \in \mathbb{N}}$ is *convergent* if, for every fixed permutation τ, the sequence of real numbers $(t(\tau, \sigma_n))_{n \in \mathbb{N}}$ converges.

The interesting case occurs when the sequence of lengths $(|\sigma_n|)_{n \in \mathbb{N}}$ tends to infinity, since, as we shall see, every convergent permutation sequence $(\sigma_n)_{n \in \mathbb{N}}$ is otherwise eventually constant. We prove that, when $|\sigma_n| \to \infty$, any convergent permutation sequence has a natural limit object, called a *limit permutation* (all asymptotics in this paper are with respect to $n \to \infty$). This limit object consists of a family of *cumulative distribution functions*, or *cdf* for short. We say that a function $F : [0, 1] \to [0, 1]$ is a cdf if

$$F \text{ is a non-decreasing and right-continuous function with } F(0) \geqslant 0 \text{ and } F(1) = 1. \tag{4}$$

Note that F is a cdf if and only if there is a $[0, 1]$-valued random variable Y such that, for every $y \in [0, 1]$, we have $F(y) = \mathbf{P}(Y \leqslant y)$.

Definition 1.3 *(Limit permutation).* A *limit permutation* is a Lebesgue measurable function $Z : [0, 1]^2 \to [0, 1]$ satisfying the following conditions:

(a) for every $x \in [0, 1]$, the function $Z(x, \cdot)$ is a cdf, i.e., (4) holds;
(b) for every $y \in [0, 1]$, the function $Z(\cdot, y)$ satisfies $\int_0^1 Z(x, y) \, dx = y$.

We denote the set of limit permutations by $\mathcal{Z}$.

In fact, limit permutations are in one-to-one correspondence with probability measures on the unit square whose marginals are uniformly distributed in $[0, 1]$, as discussed below. The heuristic picture can again be imagined using TV screens: a permutation $n \in S_n$ might be represented by the square matrix $A_\sigma = (\delta_{\sigma(i),j})_{i,j=1}^n$, where $\delta_{i,j}$ denotes Dirac's delta function. Again, if $A(i, j) = 1$, we say that the corresponding pixel is white. A TV screen obtained from a permutation has the property that the brightness of each row and each column is the same, since each row and column has a unique white pixel in it. Now imagine a convergent sequence of permutations as a sequence of TV screens showing the same image with higher and higher resolution, and the limit permutation as the "perfect TV". The brightness distribution of the limit object can be represented by a probability measure μ on $[0, 1]^2$ and inherits the property that the total brightness of each "row" and each "column" is the same, i.e., the marginal distributions of μ are uniform on $[0, 1]$. Consider the $[0, 1]^2$-valued random variable (X, Y) with distribution μ. Denote by $Z : [0, 1]^2 \to [0, 1]$ the *regular conditional distribution function* of Y given X (for a formal definition of this concept, see Shiryaev [23] and Lemma 2.2 below; heuristically, we may think of $Z(x, \cdot)$ as the cdf of Y under the condition $X = x$). The function Z will satisfy Definition 1.3(b) because X and Y are uniformly distributed on $[0, 1]$. The fact that we may think of limit permutations as regular conditional distribution functions Z, or as probability distributions μ, or as random variables (X, Y) as above will be useful in what follows.

As with graphs, limit permutations may be used to define a model of random permutations.

Definition 1.4 (*Z-random permutation*). Given a limit permutation Z we generate the *Z-random permutation* $\sigma(n, Z)$ using the following. A sequence of n real numbers $X_1, \ldots, X_n$ is generated independently and uniformly on $[0, 1]$. Conditional on $(X_1, \ldots, X_n)$, we generate n real numbers $Y_1, \ldots, Y_n$, with each Y_i generated independently according to the cdf $Z(X_i, \cdot)$. Let $(X_1^*, \ldots X_k^*)$ and $(Y_1^*, \ldots Y_k^*)$ denote the values $X_1, \ldots, X_k$ and $Y_1, \ldots, Y_k$, respectively, rearranged in increasing order. The *Z-random permutation* $\sigma = \sigma(n, Z)$ is given by $\sigma(i) = j$ if and only if $X_i^* = X_\ell$ and $Y_j^* = Y_\ell$ for some $\ell \in [n]$.

In other words, the Z-random permutation $\sigma = \sigma(n, Z)$ is given by the relative order of the vertical coordinates of the points $(X_1, Y_1), \ldots, (X_n, Y_n)$ with respect to their horizontal coordinates (we shall see later that, with probability one, $X_i \neq X_j$ and $Y_i \neq Y_j$ if $i \neq j \in [n]$).

This new model of random permutation generalizes the classical random permutation model, in which a permutation is selected uniformly at random from all permutations on $[n]$. Indeed, a classical random permutation may be obtained as a Z-random permutation for the *uniform limit permutation* Z_u, where $Z_u(x, y) = y$ for all $(x, y) \in [0, 1]^2$. The distribution of the corresponding (X, Y) is uniform on $[0, 1]^2$, i.e., X and Y are *independent* and uniform on $[0, 1]$.

Again inspired by the graph case, given a limit permutation Z, we may define the subpermutation density $t(\tau, Z)$ of a permutation τ on $[k]$ in Z as the probability that the Z-random permutation $\sigma(k, Z)$ is equal to τ.

Definition 1.5. Let $(\sigma_n)_{n \in \mathbb{N}}$ be a sequence of permutations such that $|\sigma_n| \to \infty$. Let $Z \in \mathcal{Z}$. We say that $(\sigma_n)_{n \in \mathbb{N}}$ converges to Z, or briefly write $\sigma_n \to Z$, if

$$\forall \tau \in \mathcal{S}: \quad \lim_{n \to \infty} t(\tau, \sigma_n) = t(\tau, Z). \tag{5}$$

Note that the assumption $|\sigma_n| \to \infty$ is quite natural: we prove in Claim 2.4 that if $(\sigma_n)_{n \in \mathbb{N}}$ converges and $(|\sigma_n|)_{n \in \mathbb{N}}$ has a bounded subsequence then the sequence $(\sigma_n)_{n \in \mathbb{N}}$ is eventually constant.

Theorem 1.6 *(Main result).*

(i) *Given a convergent permutation sequence $(\sigma_n)_{n\in\mathbb{N}}$ for which $|\sigma_n| \to \infty$, there exists a limit permutation Z such that $\sigma_n \to Z$ holds.*

(ii) *Conversely, every $Z \in \mathcal{Z}$ is a limit of a convergent permutation sequence, i.e., there is a sequence $(\sigma_n)_{n\in\mathbb{N}}$ such that $\sigma_n \to Z$ holds.*

Theorem 1.6 naturally raises the question of characterizing the limit permutations that are limits to the same given convergent permutation sequence $(\sigma_n)_{n\in\mathbb{N}}$. It is clear that this limit is not unique in a strict sense, because, if $\sigma_n \to Z$ and A is a measurable subset of $[0, 1]$ with measure zero, then any limit permutation obtained through the replacement of the cdf $Z(x, \cdot)$ by a cdf $Z^*(x, \cdot)$ for every $x \in A$ is also a limit of $(\sigma_n)_{n\in\mathbb{N}}$, as the value of the probabilities corresponding to $t(\tau, Z)$ and $t(\tau, Z^*)$ will be identical. This question has also been raised in the case of graph limits in [5], where uniqueness was captured by an equivalence relation induced by a pseudometric $d_\square$ between graphons. Two graphons W and W^* were proved to be limits of the same graph sequence if and only if $d_\square(W, W^*) = 0$. The case of permutations is simpler.

Theorem 1.7. *Let $Z_1, Z_2 \in \mathcal{Z}$. Let $(\sigma_n)_{n\in\mathbb{N}}$ be a convergent permutation sequence. Then $\sigma_n \to Z_1$ and $\sigma_n \to Z_2$ if and only if the set $\{x: Z_1(x, \cdot) \not\equiv Z_2(x, \cdot)\}$ has Lebesgue measure zero.*

Recall that limit permutations Z may be viewed as certain probability distributions μ or as certain random variables (X, Y) (see the discussion just after Definition 1.3). Theorem 1.7 implies that limit permutations are unique when viewed as such probability distributions or random variables.

Based on a previous concept by Cooper [8], we may introduce a distance $d_\square$ between permutations (see (31)) and, more generally, between limit permutations, which is a permutation counterpart of the graph pseudometric discussed in the previous paragraph. In particular, we may characterize our notion of convergence of permutation sequences in terms of this metric. As usual, a sequence $(\sigma_n)_{n\in\mathbb{N}}$ is said to be a *Cauchy sequence* with respect to the metric $d_\square$ if, for every $\varepsilon > 0$, there exits $n_0 = n_0(\varepsilon)$ such that $d_\square(\sigma_n, \sigma_m) < \varepsilon$ for every $n, m \geqslant n_0$.

Theorem 1.8. *A permutation sequence $(\sigma_n)_{n\in\mathbb{N}}$ converges if and only if it is a Cauchy sequence with respect to the metric $d_\square$.*

Also in analogy with the work for graphs by Lovász and Szegedy, the theory in this paper can be considered in terms of the discrete metric space $(\mathcal{S}, d_\square)$, where $\mathcal{S} = \bigcup_{i=1}^{\infty} S_n$ is the set of all finite permutations and $d_\square$ is the metric of the previous paragraph. By a standard diagonalization argument, every permutation sequence can be shown to have a convergent subsequence (see Lemma 2.5). As a consequence, the metric space $(\mathcal{S}, d_\square)$ can be enlarged to a compact metric space $(\mathcal{Z}/_\sim, d_\square)$ by adding limit permutations, where we identify limit permutations that are equal almost everywhere. By Theorem 1.6 the subspace of permutations is dense in $\mathcal{Z}/_\sim$; moreover, it is discrete, since a sequence cannot converge to a permutation without being eventually constant (Claim 2.4). Finally, Theorem 1.8 tells us that, when restricted to permutations, convergence in this metric space coincides with the concept of convergence in Definition 1.2.

We have found two essentially different paths for establishing the main results in this paper. One of them starkly resembles the work of Lovász and Szegedy [21] for graph sequences, which relies on Szemerédi-type regularity arguments. However, several difficulties of technical nature arise, as the limit objects here are more constrained than in the graph case. For a detailed account of this approach, we refer the reader to [16], while most of its permutation regularity ingredients may be found in [17].

In this paper, we have opted for an alternative approach, with a distinctive probabilistic flavor, which has the advantage of being both more compact and more direct, since several of the technicalities of the first approach can be avoided.

The remainder of this paper is structured as follows. In Section 2 we collect some preliminary results from probability theory. For instance, we discuss the relation between limit permutations and

probability measures on the unit square, and we recall some facts about weak convergence of probability measures in this context. Moreover, we prove some simple facts about the convergence of permutation sequences. Section 3 is devoted to a discussion of Z-random permutations and of the probabilistic meaning of subpermutation densities. Section 4 deals with the rectangular distance on $\mathcal{S}$ and $\mathcal{Z}$, and we prove that a large Z-random permutation is close to Z in the $d_\square$-distance with high probability. In Section 5 we define three natural, different notions of convergence on $\mathcal{Z}$, and prove that they are all equivalent. This is then used to prove Theorems 1.6, 1.7 and 1.8. For completeness, we provide Appendix A containing the proof of an auxiliary probabilistic lemma that turns out to be important in our proofs, but which may be proven by standard measure-theoretic arguments.

2. Preliminaries

The present section introduces the main concepts of probability theory that are important in this paper, and gives the proofs of two simple remarks about the convergence of permutation sequences. It is organized as follows. Subsection 2.1 deals with relevant properties of probability distributions on the unit square. In Subsection 2.2 we recall useful facts about weak convergence of probability measures on compact subspaces of $\mathbb{R}^d$, while in Subsection 2.3 we relate limit permutations to probability distributions on the unit square using the notion of regular conditional probability distributions. The remarks about the convergence of permutation sequences are addressed in Subsection 2.4.

2.1. Probability distributions on the unit square

If $A \subseteq \Omega$ is an event in a probability space, we denote by $\mathbb{1}[A]$ the indicator of the event, i.e., the random variable which takes value 1 if A occurs and value 0 if A does not occur.

We shall work with random variables (X, Y) that take values in the unit square $[0, 1]^2$: X is the horizontal and Y is the vertical coordinate of the random point (X, Y). The joint distribution of (X, Y) can be represented in many different ways.

We might define a probability measure μ on $[0, 1]^2$ by defining $\mu(B) = \mathbf{P}((X, Y) \in B)$ for every Borel set $B \subseteq [0, 1]^2$, but since the σ-algebra of Borel sets is generated by the collection of rectangles of the form $B = [0, x] \times [0, y]$, it is enough to specify $\mu(B)$ for sets of this form to define μ uniquely. Thus we define the *joint probability distribution function F* of (X, Y) by

$$F(x, y) = \mathbf{P}(X \leqslant x, Y \leqslant y) = \mu([0, x] \times [0, y]), \quad x, y \in [0, 1].$$

We call the distribution of X and Y the *first* and *second marginal distributions* of (X, Y), respectively, which satisfy $\mathbf{P}(X \leqslant x) = F(x, 1)$ and $\mathbf{P}(Y \leqslant y) = F(1, y)$.

If $x_1 < x_2$ and $y_1 < y_2$ then

$$\mathbf{P}(X \in (x_1, x_2], Y \in (y_1, y_2]) = F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + F(x_1, y_1). \tag{6}$$

The $[0, 1]^2$-valued random variables that correspond to limit permutations will always have the property that both X and Y are uniformly distributed on $[0, 1]$, which we denote by $X, Y \sim U[0, 1]$. This happens if and only if we have $F(x, 1) = x$, $F(1, y) = y$ for any $x, y \in [0, 1]$.

We shall make use of the following simple inequality many times:

$$X, Y \sim U[0, 1]$$
$$\implies \quad \forall x_1, x_2, y_1, y_2 \in [0, 1]: \quad \big|F(x_2, y_2) - F(x_1, y_1)\big| \leqslant |x_2 - x_1| + |y_2 - y_1|. \tag{7}$$

A direct consequence of the fact that X and Y are both uniform is that, for any rectangle $R = [x_1, x_2] \times [y_1, y_2] \subseteq [0, 1]^2$, we have

$$\mathbf{P}((X, Y) \in \partial R) \leqslant \mathbf{P}(X = x_1) + \mathbf{P}(X = x_2) + \mathbf{P}(Y = y_1) + \mathbf{P}(Y = y_2) = 0, \tag{8}$$

where ∂R denotes the boundary of R. In particular we have

$$F(x, y) = \mathbf{P}(X < x, Y < y) = \mathbf{P}(X \leqslant x, Y \leqslant y).$$

If (X_1, Y_1) and (X_2, Y_2) are both $[0, 1]^2$-valued random variables then we say that they have the same distribution or briefly write $(X_1, Y_1) \sim (X_2, Y_2)$ if their respective probability measures μ_1 and μ_2 agree on the Borel sets of $[0, 1]^2$, or, equivalently, if their respective joint distribution functions are the same: $F_1(x, y) = F_2(x, y)$ for all $x, y \in [0, 1]$.

2.2. Weak convergence of probability measures

Now we recall some well-known facts about weak convergence of probability measures (for details, see [3]).

Let Ω be a complete, separable metric space and let $\mathcal{B}$ denote the σ-algebra of Borel sets. If $\mu_1, \mu_2, \ldots$ and μ are probability measures on the measurable space $(\Omega, \mathcal{B})$, then we say that the sequence $(\mu_n)_{n=1}^{\infty}$ converges weakly to μ (which we denote $\mu_n \Rightarrow \mu$) if for all bounded, continuous functions $f : \Omega \to \mathbb{R}$ we have

$$\lim_{n \to \infty} \int_{\Omega} f(\omega) \, d\mu_n(\omega) = \int_{\Omega} f(\omega) \, d\mu(\omega). \tag{9}$$

We will make use of the following consequence of Prokhorov's theorem (see Chapter 1, Section 5 of [3]) characterizing compact subsets of the space of probability measures on Ω.

If Ω is compact then every sequence $(\mu_n)_{n=1}^{\infty}$ has a weakly convergent subsequence. $\qquad$ (10)

For $\Omega = \mathbb{R}^d$, if we denote by $(X^1, \ldots, X^d)$ the $\mathbb{R}^d$-valued random variable with distribution μ, and, similarly, we let $(X_n^1, \ldots, X_n^d)$ be the random variable with distribution μ_n, then we say that $(X_n^1, \ldots, X_n^d)$ converges in distribution to $(X^1, \ldots, X^d)$ (or briefly write $(X_n^1, \ldots, X_n^d) \xrightarrow{d} (X^1, \ldots, X^d)$) if $\mu_n \Rightarrow \mu$.

It easily follows from (9) that

$$\left(X_n^1, \ldots, X_n^d \right) \xrightarrow{d} \left(X^1, \ldots, X^d \right) \quad \Longrightarrow \quad \forall i \in [d] : X_n^i \xrightarrow{d} X^i. \tag{11}$$

In other words, weak convergence of the joint d-dimensional distributions implies weak convergence of the marginal distributions. As a consequence we obtain the following useful fact about weak convergence of $[0, 1]^2$-valued random variables with uniform marginals:

$$\forall n \in \mathbb{R}: \ X_n, Y_n \sim U[0, 1], \quad \text{and} \quad (X_n, Y_n) \xrightarrow{d} (X, Y) \quad \Longrightarrow \quad X, Y \sim U[0, 1]. \tag{12}$$

The converse implication of (11) does not hold in general, but it does hold in the special case when the marginals are independent.

Let Ω^i denote a metric space for all $i \in [d]$, and denote by $\Omega = \Omega^1 \times \cdots \times \Omega^d$ the product space (we may define the distance between points of Ω to be the L_{∞}-distance, that is, the maximum of the distance of each pair of coordinates). If μ_n^i, μ^i are probability measures on Ω^i then

$$\forall i \in [d]: \ \mu_n^i \Rightarrow \mu^i \quad \Longrightarrow \quad \mu_n^1 \times \cdots \times \mu_n^d \Rightarrow \mu^1 \times \cdots \times \mu^d, \tag{13}$$

where $\mu^1 \times \cdots \times \mu^d$ denotes the product measure of the measures $\mu^1, \ldots, \mu^d$ on the product space Ω. For the proof of this fact, see [3, Theorem 2.8(ii)].

We say that $B \in \mathcal{B}$ is a continuity set of the measure μ if $\mu(\partial B) = 0$. The following equivalent characterization of weak convergence is a consequence of the *Portmanteau theorem* (see [3, Theorem 2.1]):

$$\mu_n \Rightarrow \mu \quad \Longleftrightarrow \quad \mu_n(B) \to \mu(B) \quad \text{for all continuity sets } B \text{ of } \mu. \tag{14}$$

Lemma 2.1. *Let $(\mu_n)_{n=1}^{\infty}$ and μ be probability measures on $[0, 1]^2$. Denote by (X_n, Y_n) and (X, Y) the corresponding $[0, 1]^2$-valued random variables and by $F_n(x, y)$ and $F(x, y)$ the corresponding joint probability distribution functions. Assume that for all $n \in \mathbb{N}$, we have $X_n, Y_n \sim U[0, 1]$. Then*

$$(X_n, Y_n) \xrightarrow{d} (X, Y) \quad \Longleftrightarrow \quad \|F_n - F\|_{\infty} = \sup_{x, y \in [0, 1]} \left| F_n(x, y) - F(x, y) \right| \to 0.$$

Proof. In order to prove that $\|F_n - F\|_\infty \to 0$ implies $(X_n, Y_n) \xrightarrow{d} (X, Y)$ we need to show (9). It is enough to prove that, for all $\varepsilon > 0$ and for every continuous function $f : [0, 1]^2 \to \mathbb{R}$, we have

$$\limsup_{n\to\infty} \left| \int_0^1 \int_0^1 f(x, y) \, d\mu_n(x, y) - \int_0^1 \int_0^1 f(x, y) \, d\mu(x, y) \right| \leqslant \varepsilon. \tag{15}$$

Since $[0, 1]^2$ is compact, f is uniformly continuous, so we can choose $k \in \mathbb{N}$ such that if we define $B_k(i, j) = [\frac{i-1}{k}, \frac{i}{k}] \times [\frac{j-1}{k}, \frac{j}{k}]$ then we have

$$\forall i, j \in [k]: \quad \max\{f(x, y): (x, y) \in B_k(i, j)\} - \min\{f(x, y): (x, y) \in B_k(i, j)\} \leqslant \varepsilon.$$

We bound the integral on the l.h.s. of (15):

$$\int_0^1 \int_0^1 f(x, y) \big(d\mu_n(x, y) - d\mu(x, y)\big)$$

$$\leqslant \sum_{i,j=1}^k \left(\mu_n\big(B_k(i, j)\big) \cdot \max_{B_k(i,j)} f - \mu\big(B_k(i, j)\big) \cdot \min_{B_k(i,j)} f \right)$$

$$\leqslant \sum_{i,j=1}^k \varepsilon \cdot \mu_n\big(B_k(i, j)\big) + \sum_{i,j=1}^k \|f\|_\infty \cdot \big|\mu_n\big(B_k(i, j)\big) - \mu\big(B_k(i, j)\big)\big|$$

$$\overset{(6)}{\leqslant} \varepsilon + k^2 \|f\|_\infty \cdot 4\|F_n - F\|_\infty.$$

With the same arguments, we may prove that

$$\int_0^1 \int_0^1 f(x, y) \big(d\mu(x, y) - d\mu_n(x, y)\big) \leqslant \varepsilon + k^2 \|f\|_\infty \cdot 4\|F_n - F\|_\infty,$$

so that

$$\left| \int_0^1 \int_0^1 f(x, y) \big(d\mu_n(x, y) - d\mu(x, y)\big) \right| \leqslant \varepsilon + k^2 \|f\|_\infty \cdot 4\|F_n - F\|_\infty.$$

Now (15) follows from the above inequality and $\|F_n - F\|_\infty \to 0$.

In order to prove that $(X_n, Y_n) \xrightarrow{d} (X, Y)$ implies $\|F_n - F\|_\infty \to 0$, we will have to make use of our assumption that $X_n, Y_n \sim U[0, 1]$ for every n (this implication does not hold for general sequences of weakly convergent probability measures on $[0, 1]^2$).

We are going to show that

$$\forall k \in \mathbb{R} \; \exists n_0 \in \mathbb{N} \; \forall n \geqslant n_0: \quad \|F_n - F\|_\infty = \sup_{x, y \in [0,1]} |F_n(x, y) - F(x, y)| \leqslant \frac{5}{k}. \tag{16}$$

Since $X_n, Y_n \sim U[0, 1]$ for every n and $(X_n, Y_n) \xrightarrow{d} (X, Y)$, we know from (12) that $X, Y \sim U[0, 1]$. By (8), it follows that, for every rectangle $R \in [0, 1]^2$, $\mathbf{P}((X, Y) \in \partial R) = 0$, so that rectangles are continuity sets with respect to the associated measure μ. We may then apply (14) to deduce that $F_n(x, y) \to F(x, y)$ for all $x, y \in [0, 1]$. Thus there exists $n_0 \in \mathbb{N}$ such that for all $n \geqslant n_0$ and $i, j \in \{0, 1, \ldots, k\}$ we have

$$\left| F_n\left(\frac{i}{k}, \frac{j}{k}\right) - F\left(\frac{i}{k}, \frac{j}{k}\right) \right| \leqslant \frac{1}{k}. \tag{17}$$

In order to deduce (16) from this, pick $x, y \in [0, 1]$ and let $i := \lfloor k \cdot x \rfloor$ and $j := \lfloor k \cdot y \rfloor$. Then

$$
\begin{aligned}
&\left| F_n(x, y) - F(x, y) \right| \\
&\leqslant \left| F_n(x, y) - F_n\left(\frac{i}{k}, \frac{j}{k}\right) \right| + \left| F_n\left(\frac{i}{k}, \frac{j}{k}\right) - F\left(\frac{i}{k}, \frac{j}{k}\right) \right| + \left| F(x, y) - F\left(\frac{i}{k}, \frac{j}{k}\right) \right| \\
&\overset{(7)}{\leqslant} \frac{2}{k} + \left| F_n\left(\frac{i}{k}, \frac{j}{k}\right) - F\left(\frac{i}{k}, \frac{j}{k}\right) \right| + \frac{2}{k} \overset{(17)}{\leqslant} \frac{5}{k},
\end{aligned}
\tag{18}
$$

which concludes the proof. $\quad\square$

2.3. Regular conditional probabilities

The aim of this subsection is to show that limit permutations are (essentially) in one-to-one correspondence with probability measures on the unit square with $U[0, 1]$ marginals: Z is the *regular conditional distribution function* of Y with respect to X. Heuristically we may think about Z as the function satisfying $Z(x, y) = \mathbf{P}(Y \leqslant y \mid X = x)$, although the condition $X = x$ has zero probability.

Denote by $\mathcal{B}[0, 1]$ the σ-algebra of Borel sets of the unit interval $[0, 1]$ and recall the definition of limit permutation, given in Definition 1.3.

Lemma 2.2. *Let (X, Y) be a $[0, 1]^2$-valued random variable satisfying $X, Y \sim U[0, 1]$.*

(a) *Then there exists a limit permutation Z such that*

$$
\forall B \in \mathcal{B}[0, 1], \ \forall y \in [0, 1]: \quad \mathbf{P}(X \in B, Y \leqslant y) = \int_0^1 Z(x, y) \mathbb{1}[x \in B] \, dx.
\tag{19}
$$

(b) *Moreover, if $\widehat{Z}$ is another limit permutation satisfying (19) then*

$$
\int_0^1 \mathbb{1}\big[\exists y \in [0, 1]: \ Z(x, y) \neq \widehat{Z}(x, y)\big] \, dx = 0.
\tag{20}
$$

The proof of Lemma 2.2 is a standard measure-theoretic argument that is almost identical to that of Theorem 4 of Chapter II, Section 7 of [23] about the existence of *regular conditional distribution functions*. Nevertheless we give a full proof of Lemma 2.2 in Appendix A in order to keep the paper self-contained.

Lemma 2.2 tells us how to obtain a limit permutation Z from a $[0, 1]^2$-valued random variable (X, Y) satisfying $X, Y \sim U[0, 1]$. Conversely, with a limit permutation Z, we may associate such a random variable (X, Y) as follows.

Definition 2.3. Let $Z \in \mathcal{Z}$. We generate the $[0, 1]^2$-valued random variable (X, Y) *associated with Z* in the following way. We first pick $X \sim U[0, 1]$, and, given X, the random variable Y is generated according to the cdf $Z(X, \cdot)$.

It is easy to check that with this definition Z is indeed the regular conditional distribution function of Y with respect to X in the sense of Lemma 2.2.

The joint distribution function of (X, Y) associated with Z can be expressed as

$$
F(x, y) = \mathbf{P}(X \leqslant x, Y \leqslant y) \overset{(19)}{=} \int_0^x Z(x, y) \, dx.
\tag{21}
$$

By Definition 1.3(b) we get $F(1, y) = y$ which is equivalent to $Y \sim U[0, 1]$.

2.4. Two simple remarks about permutation convergence

In this subsection, we prove two simple facts about the convergence of permutation sequences that were mentioned in the introduction. We show that every convergent sequence $(\sigma_n)_{n\in\mathbb{N}}$ such that $|\sigma_n| \nrightarrow \infty$ must be eventually constant, and we establish that every permutation sequence contains a convergent subsequence.

For the first, we remind the reader that Theorem 1.6 is only stated for permutation sequences $(\sigma_n)_{n\in\mathbb{N}}$ whose lengths tend to infinity, as we claimed that every other convergent sequence is eventually constant. In light of this, we prove this claim prior to addressing the main results. As before, if $\sigma \in S_n$ we write $|\sigma| = n$. Recall the notion of convergence of permutation sequences from Definition 1.2.

Claim 2.4. *Let $(\sigma_n)_{n\in\mathbb{N}}$ be a convergent permutation sequence such that $|\sigma_n| \nrightarrow \infty$. Then the sequence $(\sigma_n)_{n\in\mathbb{N}}$ is eventually constant, that is, there is a permutation σ and an $n_0 \in \mathbb{N}$ such that $n \geqslant n_0$ implies $\sigma_n = \sigma$.*

Proof. It follows from (3) that $\sum_{\tau\in S_k} t(\tau, \pi) = \mathbb{1}[k \leqslant |\pi|]$ for any $k \in \mathbb{N}$ and any permutation π. By Definition 1.2 we get that for any fixed $k \in \mathbb{N}$ the limit $\lim_{n\to\infty} \sum_{\tau\in S_k} t(\tau, \sigma_n) = \lim_{n\to\infty} \mathbb{1}[k \leqslant |\sigma_n|]$ exists and must be equal to 0 or 1.

From this and our assumption that $\liminf_{n\to\infty} |\sigma_n| < \infty$ one deduces that $\liminf_{n\to\infty} |\sigma_n| = \limsup_{n\to\infty} |\sigma_n|$; thus there is some $m \in \mathbb{N}$ such that $|\sigma_n| = m$ if n is large enough.

Now if $\tau, \pi \in S_m$ then $t(\tau, \pi) = \mathbb{1}[\tau = \pi]$, and hence $\lim_{n\to\infty} t(\tau, \sigma_n)$ must be equal to 0 or 1 for all $\tau \in S_m$. From this and $|S_m| = m! < \infty$ it is straightforward to deduce that the sequence $(\sigma_n)_{n\in\mathbb{N}}$ is eventually constant. ☐

To conclude this section, we show that permutation sequences always contain convergent subsequences.

Lemma 2.5. *Every permutation sequence has a convergent subsequence.*

Proof. Let $(\sigma_n)_{n\in\mathbb{N}}$ be a permutation sequence. We shall find a convergent subsequence of $(\sigma_n)_{n\in\mathbb{N}}$ by a standard diagonalization argument. Since S_n is finite for every $n \geqslant 1$, the set of $S = \bigcup_{n=1}^{\infty} S_n$ of all finite permutations is countable, say $S = (\tau_m)_{m\in\mathbb{N}}$.

If $(\sigma_n)_{n\in\mathbb{N}}$ does not converge, starting with τ_1, we let $(\sigma_n^1)_{n\in\mathbb{N}}$ be a subsequence of $(\sigma_n)_{n\in\mathbb{N}}$ for which the bounded real sequence $(t(\tau_1, \sigma_n^1))_{n\in\mathbb{N}}$ converges. Inductively, for $m \geqslant 2$, we let $(\sigma_n^m)_{n\in\mathbb{N}}$ be a subsequence of $(\sigma_n^{m-1})_{n\in\mathbb{N}}$ such that $(t(\tau_m, \sigma_n^{m-1}))_{n\in\mathbb{N}}$ converges. It is now easy to see that the diagonal sequence $(\sigma_n^n)_{n\in\mathbb{N}}$ is such that, for every positive integer m, the sequence $(t(\tau_m, \sigma_n^n))_{n\in\mathbb{N}}$ converges. In other words, the sequence $(\sigma_n^n)_{n\in\mathbb{N}}$ is a convergent subsequence of (σ_n). ☐

3. *Z*-random permutations and subpermutation densities

In this section we define the concept of a random subpermutation $\sigma(k, \pi)$ of length k of a permutation π and relate the subpermutation densities $t(\tau, \pi)$ to the distribution of $\sigma(k, \pi)$. Analogously, given a limit permutation Z we define the Z-random permutation $\sigma(k, Z)$ of length k and the subpermutation densities $t(\tau, Z)$. In order to treat permutations and limit permutations in a unified way we assign a limit permutation Z_σ to every permutation σ in such a way that their subpermutation densities are close to each other.

First we recall Definition 1.1, the definition of the density $t(\tau, \pi)$ of $\tau \in S_k$ in $\pi \in S_n$. Now we give a probabilistic interpretation of this quantity.

Definition 3.1. Let $\pi \in S_n$. For $k \leqslant n$, the *random subpermutation* $\sigma(k, \pi)$ of π of *length* k is the random element of S_k generated in the following way. Choose an element $(\mathcal{X}_1^*, \ldots, \mathcal{X}_k^*) \in [n]_<^k$ uniformly from the $\binom{n}{k} = |[n]_<^k|$ possibilities and let $\sigma = \sigma(k, \pi)$ be the permutation of $[k]$ satisfying

$$\forall i, j \in [k]: \quad \pi(\mathcal{X}_i^*) < \pi(\mathcal{X}_j^*) \quad \Longleftrightarrow \quad \sigma(i) < \sigma(j).$$

In plain words: $\sigma(k, \pi)$ encodes the relative order of $(\pi(\mathcal{X}_1^*), \ldots, \pi(\mathcal{X}_k^*))$. It is obvious from Definition 1.1 that we have

$$\forall \tau \in S_k: \quad \mathbf{P}\big(\sigma(k, \pi) = \tau\big) = t(\tau, \pi). \tag{22}$$

Definition 3.2 *(Z-random permutation (Version 2)).* Given a limit permutation Z and a positive integer n, a *Z-random permutation* $\sigma(n, Z)$ is a permutation of $[n]$ generated as follows. Recall Definition 2.3, and let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be independent and identically distributed $[0, 1]^2$-valued random variables with distribution associated with Z. These pairs define the permutations $R, S \in S_n$, where

$$R(i) = \big|\{j: X_j \leqslant X_i\}\big|, \qquad S(i) = \big|\{j: Y_j \leqslant Y_i\}\big|. \tag{23}$$

The random permutation $\sigma = \sigma(n, Z)$ is given by $\sigma(n, Z) = S \circ R^{-1}$, that is, $\sigma(i) = S(R^{-1}(i))$ for every $i \in [n]$.

Note that Z-random permutations are well defined with probability one, because the probability that either $(X_1, \ldots, X_n)$ or $(Y_1, \ldots, Y_n)$ has repeated elements (i.e., S or R is not a well-defined permutation) is zero. This follows directly from the fact that both $(X_1, \ldots, X_n)$ and $(Y_1, \ldots, Y_n)$ are independent and uniformly distributed on $[0, 1]$. It is easy to see that $\sigma(n, Z)$ in Definition 3.2 is equivalent to the one given in Definition 1.4, since the X_i^* and Y_j^* defined in the latter may be related to X and Y by $X_{R^{-1}(i)} = X_i^*$ and $Y_j = Y_{S(j)}^*$. Moreover, observe that one could alternatively define a random permutation with the same distribution as $\sigma(n, Z)$ by first generating a sequence $(X_1^*, \ldots, X_n^*)$ uniformly distributed on $[0, 1]_<^n$ and then drawing each $\widehat{Y}_i$ in $[0, 1]$ independently according to the probability distribution induced by $Z(X_i^*, \cdot)$. The random permutation $\sigma(n, Z)$ is given by the order of the elements in $(\widehat{Y}_1, \ldots, \widehat{Y}_n)$. This definition of $\sigma(n, Z)$ resembles Definition 3.1.

Definition 3.3. Given a limit permutation Z and $\tau \in S_k$ the *density* of τ in Z is given by

$$t(\tau, Z) = \mathbf{P}\big(\sigma(k, Z) = \tau\big). \tag{24}$$

It will be convenient for us to associate a limit permutation Z_σ with every permutation σ in such a way that $t(\tau, Z_\sigma)$ of (24) is close to $t(\tau, \sigma)$ of (22).

Definition 3.4. Given a permutation $\sigma \in S_n$ define $Z_\sigma \in \mathcal{Z}$ by defining the associated $[0, 1]^2$-valued random (X_σ, Y_σ) to have joint density function

$$f_\sigma(x, y) = n \cdot \mathbb{1}\big[\sigma\big(\lceil n \cdot x\rceil\big) = \lceil n \cdot y\rceil\big]. \tag{25}$$

Note that the support of the density function $f_\sigma(x, y)$ "looks like" the matrix $A_\sigma = (\delta_{\sigma(i), j})_{i, j=1}^n$ of the permutation σ, the row and column sums of which are all equal to 1. From this it easily follows that we have $X_\sigma, Y_\sigma \sim U[0, 1]$; thus, by Lemma 2.2 we can define Z_σ almost surely uniquely. Indeed,

$$Z_\sigma(x, y) = \int_0^y f_\sigma(x, \tilde{y}) \, d\tilde{y}. \tag{26}$$

Lemma 3.5. *Let $\tau \in S_k$, $\sigma \in S_n$, $k \leqslant n$. Then*

$$\big|t(\tau, \sigma) - t(\tau, Z_\sigma)\big| \leqslant \frac{1}{n}\binom{k}{2}. \tag{27}$$

Proof. If A, B are events in a probability space then it is easy to check that

$$\left| \mathbf{P}(A) - \mathbf{P}(A \mid B) \right| \leqslant 1 - \mathbf{P}(B) = \mathbf{P}\left(B^c\right). \tag{28}$$

We are going to apply this inequality to prove (27). To this end, let (X_i, Y_i), $i = 1, \ldots, k$, be i.i.d. with the same joint distribution as (X_σ, Y_σ) of Definition 3.4. Then X_i, $i = 1, \ldots, k$, are i.i.d. and uniform on $[0, 1]$.

Let A be the event that the relative order of the vertical coordinates of (X_i, Y_i), $i = 1, \ldots, k$ is τ with respect to their ordered horizontal coordinates, i.e., with the notation of (23), let $A = \{\tau = S \circ R^{-1}\}$.

Define $\mathcal{X}_i := \lceil n \cdot X_i \rceil$ for each $i \in [k]$. We define the event

$$B = \{\forall 1 \leqslant i < j \leqslant k: \ \mathcal{X}_i \neq \mathcal{X}_j\} \tag{29}$$

and let $(\mathcal{X}_1^*, \mathcal{X}_2^*, \ldots, \mathcal{X}_k^*)$ denote the k-tuple that we get by arranging $(\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_k)$ in increasing order. Note that $\lceil n \cdot Y_i \rceil = \sigma(\mathcal{X}_i)$ by (25), and that under the condition that the event B occurs the random k-tuple $(\mathcal{X}_1^*, \mathcal{X}_2^*, \ldots, \mathcal{X}_k^*)$ is uniformly distributed on $[n]_<^k$. Thus we have

$$\mathbf{P}(A) \overset{(24)}{=} t(\tau, Z_\sigma), \qquad \mathbf{P}(A \mid B) \overset{(22)}{=} t(\tau, \sigma) \quad \text{and} \quad \mathbf{P}\left(B^c\right) \leqslant \frac{1}{n} \binom{k}{2},$$

and our result follows from (28). $\quad\square$

4. Rectangular distance

In this section we introduce the notion of rectangular distance $d_\square(\sigma_1, \sigma_2)$ between two permutations and also give the analogous definition $d_\square(Z_1, Z_2)$ of rectangular distance between two limit permutations. Using the notion of the limit permutation Z_σ assigned to a permutation σ, we make a connection between these two definitions of $d_\square$.

In Subsection 4.2 we prove that the rectangular distance between Z and the random subpermutation $\sigma(k, Z)$ is small with high probability if $1 \ll k$. In other words, we show that Z can be recovered from the sample $\sigma(k, Z)$ with small error. From this we derive an analogous statement which quantifies how small $d_\square(\pi, \sigma(k, \pi))$ is. This will prove to be useful in the characterization of testability in [15].

4.1. Preliminary facts about the rectangular distance

Let $I[n]$ be the set of all *intervals* in $[n]$, that is, the set of all subsets of the form $\{x \in [n]: a \leqslant x < b\}$, where $a, b \in [n + 1]$ are called the *endpoints* of the interval. Given a permutation σ on $[n]$, Cooper [7] defines the *discrepancy of* σ as

$$D(\sigma) = \max_{S, T \in I[n]} \left| \left|\sigma(S) \cap T\right| - \frac{|S||T|}{n} \right|. \tag{30}$$

This is used to measure the "randomness" of a permutation. Indeed, sequences with low discrepancy, i.e., for which $D(\sigma) = o(n)$, are said to be *quasi-random*. We use a normalized version of the same concept to introduce a distance between permutations.

Definition 4.1. Given permutations $\sigma_1, \sigma_2 \in S_n$, the *rectangular distance* between σ_1 and σ_2 is given by

$$d_\square(\sigma_1, \sigma_2) = \frac{1}{n} \max_{S, T \in I[n]} \left| \left|\sigma_1(S) \cap T\right| - \left|\sigma_2(S) \cap T\right| \right|. \tag{31}$$

An analogous metric $d_\square$ may be defined to measure the distance between limit permutations. Given $Z_1, Z_2 \in \mathcal{Z}$, and the associated $[0, 1]^2$-valued random variables (X_1, Y_1) and (X_2, Y_2) (see Definition 2.3), the *rectangular distance* between Z_1 and Z_2 is defined by

$$d_\square(Z_1, Z_2)$$

$$= \sup_{\substack{x_1 < x_2 \in [0,1] \\ y_1 < y_2 \in [0,1]}} \left| \int_{x_1}^{x_2} \big(Z_1(x, y_2) - Z_1(x, y_1)\big) \, dx - \int_{x_1}^{x_2} \big(Z_2(x, y_2) - Z_2(x, y_1)\big) \, dx \right|$$

$$\overset{(19),(8)}{=} \sup_{\substack{x_1 < x_2 \in [0,1] \\ y_1 < y_2 \in [0,1]}} \left| \mathbf{P}\big(X_1 \in [x_1, x_2], Y_1 \in [y_1, y_2]\big) - \mathbf{P}\big(X_2 \in [x_1, x_2], Y_2 \in [y_1, y_2]\big) \right|. \tag{32}$$

Denote by F_1 and F_2 the joint probability distribution function (see (21)) of (X_1, Y_1) and (X_2, Y_2), respectively. We define

$$d_\infty(Z_1, Z_2) = \|F_1 - F_2\|_\infty = \sup_{x,y \in [0,1]} \left| \mathbf{P}(X_1 \leqslant x, Y_1 \leqslant y) - \mathbf{P}(X_2 \leqslant x, Y_2 \leqslant y) \right|. \tag{33}$$

Using identity (6), it is easy to deduce that

$$d_\infty(Z_1, Z_2) \leqslant d_\square(Z_1, Z_2) \leqslant 4 \cdot d_\infty(Z_1, Z_2). \tag{34}$$

In the sequel we will use the simpler d_∞ in our proofs rather than $d_\square$. Note that we have

$$d_\square(Z_1, Z_2) = 0 \quad \Longleftrightarrow \quad d_\infty(Z_1, Z_2) = 0 \quad \Longleftrightarrow \quad F_1 \equiv F_2 \quad \Longleftrightarrow \quad (X_1, Y_1) \sim (X_2, Y_2)$$

$$\overset{(20)}{\Longleftrightarrow} \quad \int_0^1 \mathbb{1}\big[Z_1(x, \cdot) \equiv Z_2(x, \cdot)\big] \, dx = 1. \tag{35}$$

Recalling Definition 3.4 it is not hard to see that, for permutations $\sigma_1, \sigma_2 \in S_n$, we have $d_\square(\sigma_1, \sigma_2) = d_\square(Z_{\sigma_1}, Z_{\sigma_2})$, which allows us to extend the definition of rectangular distance to permutations on different sets of integers. Indeed, we may define $d_\square(\sigma, \pi) := d_\square(Z_\sigma, Z_\pi)$ for every pair of permutations $\sigma, \pi \in \mathcal{S}$.

Similarly, we define the rectangular distance of a permutation σ and a limit permutation Z by

$$d_\square(\sigma, Z) := d_\square(Z_\sigma, Z). \tag{36}$$

With this definition we may express the discrepancy in (30) as $D(\sigma) = n \cdot d_\square(\sigma, Z_u)$, where Z_u is the uniform limit permutation defined on page 96.

4.2. Rectangular distance and subpermutations

The main objective of this subsection is to show that the rectangular distance between a limit permutation and a large constant-size random subpermutation $\sigma(k, Z)$ of it is small.

Lemma 4.2. *If $k \in \mathbb{N}$ is a sufficiently large integer and $Z \in \mathcal{Z}$, then*

$$\mathbf{P}\big(d_\square\big(Z, \sigma(k, Z)\big) \leqslant 16k^{-1/4}\big) \geqslant 1 - \frac{1}{2}e^{-\sqrt{k}}. \tag{37}$$

Proof. As in (21), denote by F the joint probability distribution function of the random variable (X, Y) associated with the limit permutation Z (see Definition 2.3). Let $\sigma_k = \sigma(k, Z)$ and, recalling (25), we define the random functions f_k and F_k by

$$f_k(x, y) = n \cdot \mathbb{1}\big[\sigma_k(\lceil n \cdot x \rceil) = \lceil n \cdot y \rceil\big], \qquad F_k(x, y) = \int_0^y \int_0^x f_k(u, v) \, dv \, du. \tag{38}$$

By (36), (33) and (34), we only need to prove that if k is large enough then

$$\mathbf{P}\left(\sup_{x,y\in[0,1]}\left|F(x,y)-F_k(x,y)\right|>4k^{-1/4}\right)\leqslant\frac{1}{2}e^{-\sqrt{k}}. \tag{39}$$

As in (18), for $i=\lfloor k\cdot x\rfloor$ and $j=\lfloor k\cdot y\rfloor$, we have $|F(x,y)-F_k(x,y)|\leqslant\frac{4}{k}+|F(\frac{i}{k},\frac{j}{k})-F_k(\frac{i}{k},\frac{j}{k})|$. We claim that, in order to show (39), we only need to show that for large k we have

$$\mathbf{P}\left(\max_{i,j\in[k]}\left|F\left(\frac{i}{k},\frac{j}{k}\right)-F_k\left(\frac{i}{k},\frac{j}{k}\right)\right|>3k^{-1/4}\right)<6e^{-2\sqrt{k}}. \tag{40}$$

This is because $4/k<k^{-1/4}$ for large k and

$$\begin{aligned}
&\mathbf{P}\left(\max_{i,j\in[k]}\left|F\left(\frac{i}{k},\frac{j}{k}\right)-F_k\left(\frac{i}{k},\frac{j}{k}\right)\right|>3k^{-1/4}\right)\\
&=\mathbf{P}\left(\bigcup_{i,j\in[k]}\left\{\left|F\left(\frac{i}{k},\frac{j}{k}\right)-F_k\left(\frac{i}{k},\frac{j}{k}\right)\right|>3k^{-1/4}\right\}\right)\\
&\leqslant\sum_{i,j\in[k]}\mathbf{P}\left(\left|F\left(\frac{i}{k},\frac{j}{k}\right)-F_k\left(\frac{i}{k},\frac{j}{k}\right)\right|>3k^{-1/4}\right)\\
&\overset{(40)}{\leqslant}k^2\cdot 6e^{-2\sqrt{k}}=\left(12k^2e^{-\sqrt{k}}\right)\cdot\left(\frac{1}{2}e^{-\sqrt{k}}\right).
\end{aligned}$$

Now $12k^2e^{-\sqrt{k}}\leqslant 1$ if k is large enough, and we get (39).

To establish (40), let (X_1,Y_1), ..., (X_k,Y_k) be i.i.d. $[0,1]^2$-valued random variables with distribution function $F(x,y)$. Denote by $(X_1^*,\ldots X_k^*)$ and $(Y_1^*,\ldots Y_k^*)$ the values of $\{X_1,\ldots,X_k\}$ and $\{Y_1,\ldots,Y_k\}$, respectively, rearranged in increasing order.

Then we have

$$F_k\left(\frac{i}{k},\frac{j}{k}\right)\overset{(38)}{=}\frac{1}{k}\sum_{\ell=1}^{i}\mathbb{1}\left[\sigma_k(\ell)\leqslant j\right]=\frac{1}{k}\sum_{\ell=1}^{k}\mathbb{1}\left[X_\ell\leqslant X_i^*,Y_\ell\leqslant Y_j^*\right].$$

In order to prove (40) we first show that

$$\forall i,j\in[k]:\mathbf{P}\left(\frac{1}{k}\sum_{l=1}^{k}\mathbb{1}\left[X_l\leqslant X_i^*,Y_l\leqslant Y_j^*\right]>F\left(\frac{i}{k},\frac{j}{k}\right)+3k^{-1/4}\right)<3e^{-2\sqrt{k}}. \tag{41}$$

Set $\varepsilon=k^{-1/4}$. By (7) we get

$$F\left(\frac{i}{k}+\varepsilon,\frac{j}{k}+\varepsilon\right)+\varepsilon\leqslant F\left(\frac{i}{k},\frac{j}{k}\right)+3\varepsilon. \tag{42}$$

Let us define the event

$$A:=\left\{\frac{1}{k}\sum_{l=1}^{k}\mathbb{1}\left[X_l\leqslant X_i^*,Y_l\leqslant Y_j^*\right]>F\left(\frac{i}{k}+\varepsilon,\frac{j}{k}+\varepsilon\right)+\varepsilon\right\}.$$

Then by (42) the probability on the left-hand side of (41) is less then or equal to $\mathbf{P}(A)$. On the other hand,

$$\begin{aligned}
&\mathbf{P}(A)\\
&=\mathbf{P}\left(A\cap\left\{X_i^*<\frac{i}{k}+\varepsilon\right\}\cap\left\{Y_j^*<\frac{j}{k}+\varepsilon\right\}\right)+\mathbf{P}\left(A\cap\left(\left\{X_i^*\geqslant\frac{i}{k}+\varepsilon\right\}\cup\left\{Y_j^*\geqslant\frac{j}{k}+\varepsilon\right\}\right)\right)\\
&\leqslant\mathbf{P}\left(A\cap\left\{X_i^*<\frac{i}{k}+\varepsilon\right\}\cap\left\{Y_j^*<\frac{j}{k}+\varepsilon\right\}\right)+\mathbf{P}\left(X_i^*>\frac{i}{k}+\varepsilon\right)+\mathbf{P}\left(Y_j^*>\frac{j}{k}+\varepsilon\right). \tag{43}
\end{aligned}$$

In order to bound these probabilities we are going to use the following large deviation estimate on binomial random variables, which is a consequence of Hoeffding's inequality (see McDiarmid [22]).

$$
S \sim \mathrm{BIN}(k, p) \quad \Longrightarrow \quad
\begin{cases}
\mathbf{P}\left(\dfrac{1}{k}S > p + \varepsilon\right) \leqslant \exp(-2k\varepsilon^2), \\[3mm]
\mathbf{P}\left(\dfrac{1}{k}S < p - \varepsilon\right) \leqslant \exp(-2k\varepsilon^2).
\end{cases}
\tag{44}
$$

This inequality will be used to bound the three terms on the r.h.s. of (43),

$$
\mathbf{P}\left(X_i^* > \frac{i}{k} + \varepsilon\right) = \mathbf{P}\left(\frac{1}{k}\sum_{\ell=1}^{k} \mathbb{1}\left[X_\ell \leqslant \frac{i}{k} + \varepsilon\right] < \frac{i}{k}\right) \overset{(44)}{\leqslant} \exp(-2k\varepsilon^2)
$$

and, similarly, it holds that $\mathbf{P}(Y_j^* > \frac{j}{k} + \varepsilon) \leqslant \exp(-2k\varepsilon^2)$. Thus it only remains to notice that

$$
\mathbf{P}\left(A \cap \left\{X_i^* < \frac{i}{k} + \varepsilon\right\} \cap \left\{Y_j^* < \frac{j}{k} + \varepsilon\right\}\right)
$$
$$
\leqslant \mathbf{P}\left(\frac{1}{k}\sum_{l=1}^{k} \mathbb{1}\left[X_l \leqslant \frac{i}{k} + \varepsilon, Y_l \leqslant \frac{j}{k} + \varepsilon\right] > F\left(\frac{i}{k} + \varepsilon, \frac{j}{k} + \varepsilon\right) + \varepsilon\right) \overset{(44)}{\leqslant} \exp(-2k\varepsilon^2).
$$

Thus by (43) and the above applications of (44) we get (41), namely

$$
\mathbf{P}(A) \leqslant 3\exp(-2k\varepsilon^2) = 3\exp\left(-2k\left(k^{-1/4}\right)^2\right) = 3e^{-2\sqrt{k}}.
$$

The inequality

$$
\forall i, j \in [k] : \mathbf{P}\left(\frac{1}{k}\sum_{\ell=1}^{k} \mathbb{1}\left[X_\ell \leqslant X_i^*, Y_\ell \leqslant Y_j^*\right] < F\left(\frac{i}{k}, \frac{j}{k}\right) - 3k^{-1/4}\right) < 3e^{-2\sqrt{k}}
$$

can be proven analogously. Putting this and (41) together we get (40). □

Lemma 4.3. *Let k be a sufficiently large positive integer. Let $n > e^{2\sqrt{k}}$ and $\pi \in S_n$. Then we have*

$$
\mathbf{P}\left(d_\square\left(\pi, \sigma(k, \pi)\right) \leqslant 16k^{-1/4}\right) \geqslant 1 - e^{-\sqrt{k}}.
\tag{45}
$$

Proof. The proof uses ideas similar to the ones in the proof of Lemma 3.5.

Given the permutation π, define Z_π by Definition 3.4. The random permutations $\sigma(k, \pi)$ and $\sigma(k, Z_\pi)$ do not have the same distribution, but, if $(X_1, Y_1), \ldots, (X_k, Y_k)$ are i.i.d. $[0, 1]^2$-valued random variables associated with Z_π (see Definition 2.3), then

(i) the distribution of $\sigma(k, Z_\pi)$ is the same as that of $S \circ R^{-1}$ (defined for Z_π in (23)),
(ii) the distribution of $\sigma(k, \pi)$ is the same as that of $S \circ R^{-1}$ under the condition that the event B defined in (29) occurs.

It follows from (28) that for any event A we have

$$
\left|\mathbf{P}(A) - \mathbf{P}(A \mid B)\right| \leqslant \mathbf{P}(B^c) \leqslant \frac{1}{n}\binom{k}{2} \leqslant e^{-2\sqrt{k}}\binom{k}{2} \overset{(*)}{\leqslant} \frac{1}{2}e^{-\sqrt{k}},
$$

where the inequality $(*)$ holds if k is large enough. If we let $A = \{d_\square(Z_\pi, S \circ R^{-1}) \leqslant 16k^{-1/4}\}$ then we have $\mathbf{P}(A) \geqslant 1 - e^{-\sqrt{k}}/2$ by Lemma 4.2. Now, for k large enough, (45) follows from

$$
\mathbf{P}\left(d_\square\left(\pi, \sigma(k, \pi)\right) \leqslant 16k^{-1/4}\right) = \mathbf{P}(A \mid B) \geqslant \mathbf{P}(A) - \mathbf{P}(B^c)
$$
$$
\geqslant 1 - \frac{1}{2}e^{-\sqrt{k}} - \frac{1}{2}e^{-\sqrt{k}} \geqslant 1 - e^{-\sqrt{k}}. \quad \square
$$

5. Limits of permutation sequences

The objective of this section is to use the machinery developed in the previous sections to prove our main results. In Subsection 5.1 we define three different types of convergence on the space of limit permutations $\mathcal{Z}$ (weak convergence, $d_\square$-convergence and convergence of subpermutation densities) and prove that they are all equivalent. This is then used in Subsection 5.2 to prove the theorems stated in the introduction.

5.1. Equivalence of different notions of convergence of limit permutations

First we prove that a limit permutation is uniquely determined by the collection of its subpermutation densities (up to the almost sure equivalence in (35)). Recall that we denote by $\mathcal{S} = \bigcup_{i=1}^{\infty} S_n$ the set of all finite permutations.

Lemma 5.1. *Let $Z, \widehat{Z} \in \mathcal{Z}$. Denote by (X, Y) and $(\widehat{X}, \widehat{Y})$ the $[0,1]^2$-valued random variables associated with Z and $\widehat{Z}$ (see Definition 2.3). Then we have*

$$\forall \tau \in \mathcal{S}: \quad t(\tau, Z) = t(\tau, \widehat{Z}) \quad \Longrightarrow \quad (X, Y) \sim (\widehat{X}, \widehat{Y}).$$

Proof. Denote by F the joint probability density function of (X, Y) (see (21)). By (35) it suffices to prove that, if we know $(t(\tau, Z))_{\tau \in \mathcal{S}}$, then we also know the value of $F(x, y)$ for every $x, y \in [0, 1]$.

As a matter of fact, given $(t(\tau, Z))_{\tau \in \mathcal{S}}$ we know the distribution of the random permutation $\sigma(k, Z)$ for every $k \in \mathbb{N}$ by (24), and, from $\sigma(k, Z)$, we can define the random function F_k as in (38). Thus we can calculate the expected value $\mathbf{E}(F_k(x, y))$ given $(t(\tau, Z))_{\tau \in S_k}$. As in the proof of Lemma 4.2, it follows that (39) holds, which implies that for every $x, y \in [0, 1]$ we have $\lim_{k \to \infty} \mathbf{E}(F_k(x, y)) = F(x, y)$.

In other words, we have shown that, given $(t(\tau, Z))_{\tau \in \mathcal{S}}$, the value of $F(x, y)$ can be recovered for every $x, y \in [0, 1]$. $\quad\square$

Recall the definition of $\overset{d}{\longrightarrow}$ from Subsection 2.2 and the definition of $d_\square(Z_1, Z_2)$ from (32).

Definition 5.2. Let $Z, Z_1, Z_2, \ldots$ be limit permutations. Denote by (X_n, Y_n) and (X, Y) the $[0, 1]^2$-valued random variables associated with Z_n and Z, respectively. We say that

1. $Z_n \Rightarrow Z$ if $(X_n, Y_n) \overset{d}{\longrightarrow} (X, Y)$ holds;
2. $Z_n \overset{\square}{\longrightarrow} Z$ if $\lim_{n \to \infty} d_\square(Z_n, Z) = 0$;
3. $Z_n \overset{t}{\longrightarrow} Z$ if $\forall \tau \in \mathcal{S}: \lim_{n \to \infty} t(\tau, Z_n) = t(\tau, Z)$.

Lemma 5.3. *Let $Z, Z_1, Z_2, \ldots$ be limit permutations. The three notions of convergence of Definition 5.2 are equivalent, that is we have*

$$Z_n \Rightarrow Z \quad \Longleftrightarrow \quad Z_n \overset{\square}{\longrightarrow} Z \quad \Longleftrightarrow \quad Z_n \overset{t}{\longrightarrow} Z.$$

Proof. As before, denote by F_n and F the respective joint probability distribution functions of (X_n, Y_n) and (X, Y), following (21).

We first address the equivalence of $Z_n \Rightarrow Z$ and $Z_n \overset{\square}{\longrightarrow} Z$. On the one hand, $Z_n \Rightarrow Z$ and $(X_n, Y_n) \overset{d}{\longrightarrow} (X, Y)$ are equivalent by definition, and it holds that $X_n, Y_n \sim U[0, 1]$. On the other hand, by (33) and (34), we have that $Z_n \overset{\square}{\longrightarrow} Z$ is equivalent to $\|F_n - F\|_\infty \to 0$. The equivalence of $Z_n \Rightarrow Z$ and $Z_n \overset{\square}{\longrightarrow} Z$ is now a direct consequence of Lemma 2.1.

Now we are going to prove that $Z_n \Rightarrow Z$ and $Z_n \overset{t}{\longrightarrow} Z$ are equivalent.

First we assume that $Z_n \Rightarrow Z$ holds. In order to deduce $Z_n \overset{t}{\longrightarrow} Z$ from this, we need to show that

$$\forall \tau \in \mathcal{S}: \quad \lim_{n \to \infty} t(\tau, Z_n) = t(\tau, Z).$$

Fix k and $\tau \in S_k$. Let (X_n^i, Y_n^i), $i = 1, \ldots, k$, be i.i.d. with the same joint distribution as (X_n, Y_n), while (X^i, Y^i), $i = 1, \ldots, k$, are i.i.d. with the same joint distribution as (X, Y).

It follows from $Z_n \Rightarrow Z$ and (13) that

$$\left((X_n^i, Y_n^i) \right)_{i \in [k]} \overset{d}{\longrightarrow} \left((X^i, Y^i) \right)_{i \in [k]} \quad \text{as } n \to \infty, \tag{46}$$

i.e., the sequence of $[0, 1]^{2k}$-valued random variables $((X_n^i, Y_n^i))_{i \in [k]}$, $n = 1, 2, \ldots$, converges in distribution to the $[0, 1]^{2k}$-valued random variable $((X^i, Y^i))_{i \in [k]}$.

Let $A_\tau \subseteq [0, 1]^{2k}$ denote the event that the relative order of the vertical coordinates of (x^i, y^i), $i = 1, \ldots, k$, is τ with respect to their ordered horizontal coordinates. Then

$$\mathbf{P}\left(\left((X_n^i, Y_n^i) \right)_{i \in [k]} \in A_\tau \right) = t(\tau, Z_n) \quad \text{and} \quad \mathbf{P}\left(\left((X^i, Y^i) \right)_{i \in [k]} \in A_\tau \right) = t(\tau, Z). \tag{47}$$

In order to deduce $t(\tau, Z_n) \to t(\tau, Z)$ from (47) we apply (14). We only need to show that the boundary of A_τ has measure 0 with respect to the distribution of $((X^i, Y^i))_{i \in [k]}$, which is indeed true because, on the boundary of A_τ, either $x^i = x^j$ or $y^i = y^j$ holds for some $i \neq j$, but this has zero probability since both $(X^1, \ldots, X^k)$ and $(Y^1, \ldots, Y^k)$ are i.i.d. and $U[0, 1]$-distributed. Thus we have proven that $Z_n \Rightarrow Z$ implies $Z_n \overset{t}{\longrightarrow} Z$.

Now we assume that $Z_n \overset{t}{\longrightarrow} Z$ holds and deduce from this that $Z_n \Rightarrow Z$ holds. By (9) we only need to show that for any bounded, continuous function $f : [0, 1]^2 \to \mathbb{R}$ we have

$$\mathbf{E}\left(f(X_n, Y_n) \right) \to \mathbf{E}\left(f(X, Y) \right). \tag{48}$$

We are going to prove this by contradiction: assume that there is some f such that (48) does not hold. Since $|\mathbf{E}(f(X_n, Y_n))| \leqslant \|f\|_\infty$, we can choose a subsequence $(n(m))_{m=1}^\infty$ such that $\lim_{m \to \infty} \mathbf{E}(f(X_{n(m)}, Y_{n(m)})) = a$ for some $a \neq \mathbf{E}(f(X, Y))$. Now by (10) we can find a second subsequence $(m(\ell))_{\ell=1}^\infty$ and a $[0, 1]^2$-valued random variable $(\widehat{X}, \widehat{Y})$ such that if we define $(\widehat{X}_\ell, \widehat{Y}_\ell) := (X_{n(m(\ell))}, Y_{n(m(\ell))})$ then $(\widehat{X}_\ell, \widehat{Y}_\ell) \overset{d}{\longrightarrow} (\widehat{X}, \widehat{Y})$ as $\ell \to \infty$. In particular, we have $\mathbf{E}(f(\widehat{X}, \widehat{Y})) = a$.

We now look at the properties of $\widehat{X}$ and $\widehat{Y}$. By (12) we know that $\widehat{X}, \widehat{Y} \sim U[0, 1]$. Using Lemma 2.2, we define $\widehat{Z}$ to be the limit permutation associated with $(\widehat{X}, \widehat{Y})$. By definition, we now have $\widehat{Z}_l \Rightarrow \widehat{Z}$ where $\widehat{Z}_\ell := Z_{n(m(\ell))}$, and hence $\widehat{Z}_\ell \overset{t}{\longrightarrow} \widehat{Z}$ follows from our previous argument. On the other hand, from the assumption $Z_n \overset{t}{\longrightarrow} Z$, we obtain $\widehat{Z}_\ell \overset{t}{\longrightarrow} Z$.

However, from $\widehat{Z}_l \overset{t}{\longrightarrow} Z$, $\widehat{Z}_l \overset{t}{\longrightarrow} \widehat{Z}$ and Lemma 5.1, we obtain $(X, Y) \sim (\widehat{X}, \widehat{Y})$ and, in particular, $\mathbf{E}(f(\widehat{X}, \widehat{Y})) = \mathbf{E}(f(X, Y))$ which contradicts $\mathbf{E}(f(\widehat{X}, \widehat{Y})) = a \neq \mathbf{E}(f(X, Y))$. Thus we have proven that $Z_n \Rightarrow Z$ and $Z_n \overset{t}{\longrightarrow} Z$ are equivalent, which completes the proof of the lemma. $\quad\square$

5.2. Proof of the main theorems

In this section, we shall focus on the proofs of the main results of this work, which have been stated in the introduction. We start with a brief overview of these results. Theorem 1.6 establishes the correspondence between convergent permutation sequences and limit permutations, in the sense that there is a limit permutation associated with every convergent sequence and vice-versa. The fact that the limit of a permutation sequence is essentially unique is the subject of Theorem 1.7, while Theorem 1.8 relates the original concept of convergence of a permutation sequence to convergence with respect to the metric $d_\square$.

As a consequence of Claim 2.4, we shall henceforth restrict our attention to permutation sequences $(\sigma_n)_{n \in \mathbb{N}}$ such that $|\sigma_n| \to \infty$. Moreover, as a useful auxiliary fact, we point out that it follows from Definitions 1.5 and 3.4 and from Lemma 3.5 that we have

$$\sigma_n \to Z \quad \Longleftrightarrow \quad Z_{\sigma_n} \overset{t}{\longrightarrow} Z. \tag{49}$$

Proof of Theorem 1.6(i). Let us define $Z_n := Z_{\sigma_n}$. As usual, we denote by (X_n, Y_n) the $[0,1]^2$-valued random variables associated with Z_n (see Definition 2.3), which satisfy $X_n, Y_n \sim U[0,1]$ for every n. By (10) there is a $[0,1]^2$-valued random variable (X, Y) and a subsequence $(n(m))_{m=1}^\infty$ such that $(X_{n(m)}, Y_{n(m)}) \xrightarrow{d} (X, Y)$ as $m \to \infty$. Moreover, it follows from (12) that we have $X, Y \sim U[0,1]$. Using Lemma 2.2 we define Z to be the limit permutation associated with (X, Y). With these definitions we have $Z_{n(m)} \Rightarrow Z$ as $m \to \infty$, from which $Z_{n(m)} \xrightarrow{t} Z$ follows by Lemma 5.3. By (49) we obtain $t(\tau, \sigma_{n(m)}) \to t(\tau, Z)$ as $m \to \infty$ for every τ. Since we have assumed that $t(\tau, \sigma_n)$ is convergent (see Definition 1.2) as $n \to \infty$ we obtain $t(\tau, \sigma_n) \to t(\tau, Z)$ as $n \to \infty$, i.e., $\sigma_n \to Z$. $\quad\square$

Proof of Theorem 1.6(ii). The basic idea of the proof comes from [9], where the authors suggest that the analogous theorem for graphs and graphons can be proven using reverse martingales.

We jointly define the random permutations $(\sigma_n)_{n \in \mathbb{N}}$ in the following way: let (X_n, Y_n), $n \in \mathbb{N}$ be i.i.d. $[0,1]^2$-valued random variables associated with Z (see Definition 2.3). Let σ_n be the relative order of the vertical coordinates of (X_i, Y_i), $i = 1, \ldots, n$ with respect to their ordered horizontal coordinates (see Definition 3.2). Note that, with this definition, σ_n is almost surely a subpermutation of σ_{n+1}.

We are going to show that $\mathbf{P}(\sigma_n \to Z) = 1$, i.e., that the random sequence of permutations $(\sigma_n)_{n \in \mathbb{N}}$ almost surely converges to Z. It is enough to show that, for all $\tau \in \mathcal{S}$, we have $\mathbf{P}(t(\tau, \sigma_n) \to t(\tau, Z)) = 1$. Fix $\tau \in S_k$ and let $A_\tau \subseteq [0,1]^{2k}$ denote the event that the relative order of the vertical coordinates of (x^i, y^i), $i = 1, \ldots, k$, is τ with respect to their ordered horizontal coordinates. If $k \leqslant n$ then by (3) the random variable $t(\tau, \sigma_n)$ can be expressed as

$$ t(\tau, \sigma_n) = \frac{1}{\binom{n}{k}} \sum_{(i_1, \ldots, i_k) \in [n]^k_<} \mathbb{1}\big[\big((X_{i_j}, Y_{i_j}) \big)_{j \in [k]} \in A_\tau \big]. $$

From this formula and (47), linearity of expectation leads to $\mathbf{E}(t(\tau, \sigma_n)) = t(\tau, Z)$.

Now $\mathbf{P}(t(\tau, \sigma_n) \to t(\tau, Z)) = 1$ follows from the strong law of large numbers for U-statistics (see Section 3.4 of [18]). $\quad\square$

We observe that there is an alternative, more self-contained proof of $\mathbf{P}(\sigma_n \to Z) = 1$, with (σ_n) as in the proof above. In fact, we may use the Borel–Cantelli lemma to derive $\mathbf{P}(d_\square(\sigma_n, Z) \to 0) = 1$ from Lemma 4.2 and then apply Lemma 5.3 to derive $\mathbf{P}(Z_{\sigma_n} \xrightarrow{t} Z) = 1$ from this and finally use (49) to arrive at $\mathbf{P}(\sigma_n \to Z) = 1$.

Proof of Theorem 1.7. We want to show that $\sigma_n \to Z_1$ and $\sigma_n \to Z_2$ implies that the set $\{x: Z_1(x, \cdot) \not\equiv Z_2(x, \cdot)\}$ has Lebesgue measure zero. In light of the equivalence of the statements in (35), this follows from Lemma 5.1. $\quad\square$

Proof of Theorem 1.8. By Claim 2.4 we may assume that $|\sigma_n| \to \infty$.

Let $Z_n := Z_{\sigma_n}$. It suffices to show that $(\sigma_n)_{n \in \mathbb{N}}$ is a Cauchy sequence w.r.t. $d_\square$ if and only if $Z_n \xrightarrow{\square} Z$ for some $Z \in \mathcal{Z}$ (i.e., the completion of $(\mathcal{S}, d_\square)$ is $(\mathcal{Z}/_\sim, d_\square)$, where $\sim$ is the equivalence relation in (35)). Indeed, assuming that this is true, Lemma 5.3 implies that the fact that $(\sigma_n)_{n \in \mathbb{N}}$ is a Cauchy sequence with respect to $d_\square$ is equivalent to $Z_n \xrightarrow{t} Z$ for some limit permutation Z, which is equivalent to $\sigma_n \to Z$ by (49), as required.

We now prove the above assertion. It is clear that, if $Z_n \xrightarrow{\square} Z$ for some $Z \in \mathcal{Z}$, then σ_n is a Cauchy sequence w.r.t. $d_\square$. For the converse, let σ_n be such a Cauchy sequence and denote by (X_n, Y_n) the $[0,1]^2$-valued random variables associated with Z_n. Denote by F_n the joint probability distribution function of (X_n, Y_n). By (33) and (34) we get that F_n is a Cauchy sequence w.r.t. the L_∞-norm. Thus by the completeness of $L_\infty[0,1]^2$, we have $F_n \xrightarrow{L_\infty} F$ for some $F: [0,1]^2 \to [0,1]$. Now we show that F itself is the joint probability distribution function of $[0,1]^2$-valued random variable with uniform marginals.

By (10) there is an (X, Y) and a subsequence $(n(m))_{m=1}^{\infty}$ such that $(X_{n(m)}, Y_{n(m)}) \xrightarrow{\text{d}} (X, Y)$ as $m \to \infty$. By (12) we have $X, Y \sim U[0, 1]$. Denote by $\widehat{F}$ the joint probability distribution function of (X, Y). By Lemma 2.1 we have $F_{n(m)} \xrightarrow{L_{\infty}} \widehat{F}$. Comparing this to $F_n \xrightarrow{L_{\infty}} F$ we get that $F = \widehat{F}$. In other words, (Z_n) converges to some $Z \in \mathcal{Z}$ w.r.t. the L_{∞}-norm, and hence $Z_n \xrightarrow{\square} Z$ by (35), as required. $\square$

Acknowledgments

Appendix A. Proof of Lemma 2.2

In this appendix we prove Lemma 2.2. The proof is quite similar to that of Theorem 4 of Chapter II, Section 7 of [23]. It relies on Kolmogorov's notion of conditional expectation, about which we recall some useful facts now.

Given a measurable space $(\Omega, \mathcal{F})$ and an event $A \in \mathcal{F}$ we denote by $\mathbb{1}_A = \mathbb{1}[A]$ the random variable which takes value 1 if the event A occurs and 0 if A does not occur. The following theorem states the existence and almost sure uniqueness of conditional expectation. For a proof, see Section 7 of Chapter 2 of [23].

Theorem A.1. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $\mathcal{G} \subseteq \mathcal{F}$ be a σ-algebra. Consider an $\mathcal{F}$-measurable, real-valued random variable Y with $\mathbf{E}(|Y|) < \infty$.*

 (i) *There exists a $\mathcal{G}$-measurable, real-valued random variable W such that $\mathbf{E}(|W|) < \infty$ and*

$$\forall A \in \mathcal{G}: \quad \mathbf{E}(Y \cdot \mathbb{1}_A) = \mathbf{E}(W \cdot \mathbb{1}_A). \tag{50}$$

 (ii) *If $\widehat{W}$ is another $\mathcal{G}$-measurable, real-valued random variable such that $\mathbf{E}(|\widehat{W}|) < \infty$ and*

$$\forall A \in \mathcal{G}: \quad \mathbf{E}(Y \cdot \mathbb{1}_A) = \mathbf{E}(\widehat{W} \cdot \mathbb{1}_A)$$

 then we have $\mathbf{P}(W = \widehat{W}) = 1$.

We call any W that satisfies Theorem A.1(i) the *conditional expectation of Y with respect to $\mathcal{G}$* and let $\mathbf{E}(Y \mid \mathcal{G}) = W$. Note that Theorem A.1(ii) states that the conditional expectation of Y w.r.t. $\mathcal{G}$ is almost surely uniquely defined.

If X is another real-valued $\mathcal{F}$-measurable random variable and $\sigma(X)$ is the σ-algebra generated by X then we let $\mathbf{E}(Y \mid X) = \mathbf{E}(Y \mid \sigma(X))$. Since $\mathbf{E}(Y \mid X)$ is a $\sigma(X)$-measurable random variable, it follows from Theorem 3 of Section 4 of Chapter 2 of [23] that there is a Borel-measurable function $\varphi : \mathbb{R} \to \mathbb{R}$ such that

$$\mathbf{E}(Y \mid X) = \varphi(X). \tag{51}$$

For $A \in \mathcal{F}$, let $\mathbf{P}(A \mid X) := \mathbf{E}(\mathbb{1}_A \mid X)$.

The proofs of the following facts about conditional expectation can be found in Section 7 of Chapter 2 of [23]:

$$Y_1 \leqslant Y_2 \quad \Longrightarrow \quad \mathbf{P}\big(\mathbf{E}(Y_1 \mid \mathcal{G}) \leqslant \mathbf{E}(Y_2 \mid \mathcal{G})\big) = 1 \tag{52}$$

$$Y_n \leqslant Y^*, \quad \mathbf{E}(Y^*) < \infty, \quad Y_n \searrow Y \quad \Longrightarrow \quad \mathbf{P}\big(\mathbf{E}(Y_n \mid \mathcal{G}) \searrow \mathbf{E}(Y \mid \mathcal{G})\big) = 1, \tag{53}$$

where the notation $Z_n \searrow Z$ indicates that (Z_n) is nonincreasing and converges to Z. Recall that we write $\mathcal{B}[0, 1]$ for the σ-algebra of Borel sets of the unit interval $[0, 1]$. Denote by λ the Lebesgue measure on $[0, 1]$. We are now ready to give the proof of Lemma 2.2.

Proof of Lemma 2.2. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the probability space on which our $[0, 1]^2$-valued random variable (X, Y) lives. Recall that we assume $X, Y \sim U[0, 1]$ and, therefore, for any $B \in \mathcal{B}[0, 1]$, we have $\mathbf{P}(X \in B) = \mathbf{P}(Y \in B) = \lambda(B)$. Denote by $\tilde{\mathbb{Q}} = \mathbb{Q} \cap [0, 1]$ the set of rational numbers in $[0, 1]$. Thus, $\tilde{\mathbb{Q}}$ is a countable, dense subset of $[0, 1]$. Throughout the proof, $\tilde{y}, \tilde{y}_1$ and $\tilde{y}_2$ denote elements of $\tilde{\mathbb{Q}}$, while x and y denote elements of $[0, 1]$.

For $x \in [0, 1]$ and $\tilde{y} \in \tilde{\mathbb{Q}}$ define $\tilde{Z}(x, \tilde{y})$ to be the function for which

$$\tilde{Z}(X, \tilde{y}) \overset{(51)}{:=} \mathbf{E}\big(\mathbb{1}[Y \leqslant \tilde{y}] \mid X\big) = \mathbf{P}(Y \leqslant \tilde{y} \mid X). \tag{54}$$

Note that, given $\tilde{y}$, the function $\tilde{Z}(\cdot, \tilde{y})$ is only defined almost surely uniquely with respect to the distribution of X (i.e., the Lebesgue measure λ). We may assume $\tilde{Z}(x, 1) = 1$.

For $\tilde{y} \in \tilde{\mathbb{Q}}$ and $\tilde{y}_1 \leqslant \tilde{y}_2 \in \tilde{\mathbb{Q}}$, define $A_{\tilde{y}_1, \tilde{y}_2} \in \mathcal{B}[0, 1]$ and $C_{\tilde{y}} \in \mathcal{B}[0, 1]$ by

$$A_{\tilde{y}_1, \tilde{y}_2} = \big\{x\colon \tilde{Z}(x, \tilde{y}_1) \leqslant \tilde{Z}(x, \tilde{y}_2)\big\}, \qquad C_{\tilde{y}} = \Big\{x\colon \lim_{n \to \infty} \tilde{Z}\Big(x, \tilde{y} + \frac{1}{n}\Big) = \tilde{Z}(x, \tilde{y})\Big\}. \tag{55}$$

It follows from $\mathbb{1}[Y \leqslant \tilde{y}_1] \leqslant \mathbb{1}[Y \leqslant \tilde{y}_2]$ and (52) that $\mathbf{P}(\tilde{Z}(X, \tilde{y}_1) \leqslant \tilde{Z}(X, \tilde{y}_2)) = 1$, which implies $\lambda(A_{\tilde{y}_1, \tilde{y}_2}) = 1$. Moreover, it follows from $\mathbb{1}[Y \leqslant \tilde{y} + \frac{1}{n}] \searrow \mathbb{1}[Y \leqslant \tilde{y}]$ and (53) that $\mathbf{P}(\tilde{Z}(X, \tilde{y} + \frac{1}{n}) \searrow \tilde{Z}(X, \tilde{y})) = 1$, which implies $\lambda(C_{\tilde{y}}) = 1$.

Define $A \in \mathcal{B}[0, 1]$ by

$$A := \Big(\bigcap_{\tilde{y}_1 < \tilde{y}_2 \in \tilde{\mathbb{Q}}} A_{\tilde{y}_1, \tilde{y}_2}\Big) \cap \Big(\bigcap_{\tilde{y} \in \tilde{\mathbb{Q}}} C_{\tilde{y}}\Big).$$

If $x \in A$ then $\tilde{Z}(x, \cdot)|_{\tilde{\mathbb{Q}}}$ is a monotone increasing, right continuous function. By the countable additivity of the measure λ we have $\lambda(A) = 1$.

Now we construct $Z : [0, 1]^2 \to [0, 1]$ by defining $Z(x, 1) = 1$ for all x and, for $y \in [0, 1)$, letting

$$Z(x, y) := \begin{cases} \inf_{y < \tilde{y}} \tilde{Z}(x, \tilde{y}) & \text{if } x \in A, \\ y & \text{if } x \notin A. \end{cases} \tag{56}$$

Now we show that this Z is a limit permutation (see Definition 1.3) and that it satisfies Lemma 2.2(a). Note first that it follows from (54) and (56) that Z is measurable.

It follows directly from (56) that, for all $x \in [0, 1]$, the function $Z(x, \cdot)$ is a cdf (see (4)), and thus Z satisfies Definition 1.3(a). For all $x \in A$ and $\tilde{y} \in \tilde{\mathbb{Q}}$ we have $Z(x, \tilde{y}) = \tilde{Z}(x, \tilde{y})$.

Next we show that for any $B \in \mathcal{B}[0, 1]$ and $y \in [0, 1]$ we have (19):

$$\int_0^1 Z(x, y)\mathbb{1}[x \in B]\,dx \overset{\lambda(A) = 1}{=} \int_0^1 Z(x, y)\mathbb{1}[x \in B \cap A]\,dx$$

$$\overset{(56)}{=} \int_0^1 \inf_{y < \tilde{y}} \tilde{Z}(x, \tilde{y})\mathbb{1}[x \in B \cap A]\,dx \overset{(*)}{=} \inf_{y < \tilde{y}} \int_0^1 \tilde{Z}(x, \tilde{y})\mathbb{1}[x \in B \cap A]\,dx$$

$$= \inf_{y < \tilde{y}} \mathbf{E}\big(\tilde{Z}(X, \tilde{y})\mathbb{1}[X \in B \cap A]\big) \overset{\lambda(A) = 1}{=} \inf_{y < \tilde{y}} \mathbf{E}\big(\tilde{Z}(X, \tilde{y})\mathbb{1}[X \in B]\big)$$

$$\overset{(54),(50)}{=} \inf_{y < \tilde{y}} \mathbf{E}\big(\mathbb{1}[Y \leqslant \tilde{y}] \cdot \mathbb{1}[X \in B]\big) \overset{(*)}{=} \mathbf{E}\big(\mathbb{1}[X \in B, Y \leqslant y]\big)$$

$$= \mathbf{P}(X \in B, Y \leqslant y).$$

The equalities marked by $(*)$ hold because of the monotone convergence theorem. Our assumption $Y \sim U[0, 1]$ and an application of (19) with $B = [0, 1]$ implies that Definition 1.3(b) holds, i.e., Z is indeed a limit permutation. This finishes the proof of Lemma 2.2(a).

It only remains to prove Lemma 2.2(b). Assume that $\widehat{Z}$ is another limit permutation that satisfies (19). For any given $\tilde{y} \in \tilde{\mathbb{Q}}$, define

$$D_{\tilde{y}} = \left\{ x \colon Z(x, \tilde{y}) = \widehat{Z}(x, \tilde{y}) \right\}.$$

From the above proof it follows that both $Z(X, \tilde{y})$ and $\widehat{Z}(X, \tilde{y})$ satisfy the definition of $\mathbf{E}(\mathbb{1}[Y \leqslant \tilde{y}] \mid X)$. Thus by Theorem A.1(ii) we obtain $\mathbf{P}(Z(X, \tilde{y}) = \widehat{Z}(X, \tilde{y})) = 1$, which implies $\lambda(D_{\tilde{y}}) = 1$. If we define $D = \bigcap_{\tilde{y}} D_{\tilde{y}}$ then we have $\lambda(D) = 1$ by countable additivity. By Definition 1.3(a) both $Z(x, \cdot)$ and $\widehat{Z}(x, \cdot)$ are right continuous functions for any x, and if $x \in D$ then these two functions agree on $\tilde{\mathbb{Q}}$ and, therefore, they agree on the whole unit interval. This proves that $\lambda(\{x \colon Z(x, \cdot) \equiv \widehat{Z}(x, \cdot)\}) = 1$, which is a reformulation of (20), completing the proof of Lemma 2.2(b). $\square$

References

[1] N. Alon, A. Shapira, A characterization of the (natural) graph properties testable with one-sided error, SIAM J. Comput. 37 (6) (2008) 1703–1727.
[2] I. Benjamini, O. Schramm, Recurrence of distributional limits of finite planar graphs, Electron. J. Probab. 6 (23) (2001), 13 pp.
[3] P. Billingsley, Convergence of Probability Measures, second ed., Wiley Ser. Probab. Stat., John Wiley & Sons Inc., A Wiley–Interscience Publication, New York, 1999.
[4] C. Borgs, J.T. Chayes, L. Lovász, V.T. Sós, B. Szegedy, K. Vesztergombi, Graph limits and parameter testing, in: STOC, 2006, pp. 261–270.
[5] C. Borgs, J.T. Chayes, L. Lovász, V.T. Sós, K. Vesztergombi, Convergent sequences of dense graphs. I. Subgraph frequencies, metric properties and testing, Adv. Math. 219 (6) (2008) 1801–1851.
[6] C. Borgs, J.T. Chayes, L. Lovász, V.T. Sós, K. Vesztergombi, Convergent sequences of dense graphs. II. Multiway cuts and statistical physics, Ann. of Math. 176 (1) (2012) 151–219.
[7] J.N. Cooper, Quasirandom permutations, J. Combin. Theory Ser. A 106 (1) (2004) 123–143.
[8] J.N. Cooper, A permutation regularity lemma, Electron. J. Combin. 13 (1) (2006), Research Paper 22, 20 pp.
[9] P. Diaconis, S. Janson, Graph limits and exchangeable random graphs, Rend. Mat. Appl. (7) 28 (1) (2008) 33–61.
[10] G. Elek, On limits of finite graphs, Combinatorica 27 (4) (2007) 503–507.
[11] G. Elek, On the limit of large girth graph sequences, Combinatorica 30 (5) (2010) 553–563.
[12] G. Elek, B. Szegedy, Limits of hypergraphs, removal and regularity lemmas. A non-standard approach, arXiv:0705.2179, 2007.
[13] P. Erdős, A. Rényi, On random graphs. I, Publ. Math. Debrecen 6 (1959) 290–297.
[14] E.N. Gilbert, Random graphs, Ann. Math. Statist. 30 (1959) 1141–1144.
[15] C. Hoppen, Y. Kohayakawa, C.G. Moreira, R.M. Sampaio, Testing permutation properties through subpermutations, Theoret. Comput. Sci. 412 (29) (2011) 3555–3567.
[16] C. Hoppen, Y. Kohayakawa, C.G.T.d.A. Moreira, R.M. Sampaio, Limits of permutation sequences through permutation regularity, arXiv:1106.1663, 2011.
[17] C. Hoppen, Y. Kohayakawa, R.M. Sampaio, A note on permutation regularity, Discrete Appl. Math., http://dx.doi.org/10.1016/j.dam.2011.06.002, in press.
[18] A.J. Lee, U-statistics, in: Statistics: Theory and Practice, in: Textb. Monogr., vol. 110, Marcel Dekker Inc., New York, 1990.
[19] L. Lovász, Very large graphs, in: Current Developments in Mathematics, 2008, Int. Press, Somerville, MA, 2009, pp. 67–128.
[20] L. Lovász, V.T. Sós, Generalized quasirandom graphs, J. Combin. Theory Ser. B 98 (1) (2008) 146–163.
[21] L. Lovász, B. Szegedy, Limits of dense graph sequences, J. Combin. Theory Ser. B 96 (6) (2006) 933–957.
[22] C. McDiarmid, Concentration, in: Probabilistic Methods for Algorithmic Discrete Mathematics, in: Algorithms Combin., vol. 16, Springer-Verlag, Berlin, 1998, pp. 195–248.
[23] A.N. Shiryaev, Probability, second ed., Grad. Texts in Math., vol. 95, Springer-Verlag, New York, 1996, translated from the first (1980) Russian edition by R.P. Boas.