

RESEARCH ARTICLE

A structure theorem for stochastic processes indexed by the discrete hypercube

Pandelis Dodos¹ and Konstantinos Tyros²

¹Department of Mathematics, University of Athens, Panepistimiopolis 157 84, Athens, Greece; E-mail: pdodos@math.uoa.gr.

²Department of Mathematics, University of Athens, Panepistimiopolis 157 84, Athens, Greece; E-mail: ktyros@math.uoa.gr.

Received: 3 March 2019; Revised: 27 November 2020; Accepted: 30 December 2020

2020 Mathematics Subject Classification: Primary – 05D10, 05D40; Secondary – 60G99

Abstract

Let A be a finite set with $|A| \geq 2$, let n be a positive integer, and let A^n denote the discrete n -dimensional hypercube (that is, A^n is the Cartesian product of n many copies of A). Given a family $\langle D_t : t \in A^n \rangle$ of measurable events in a probability space (a stochastic process), what structural information can be obtained assuming that the events $\langle D_t : t \in A^n \rangle$ are not behaving as if they were independent? We obtain an answer to this problem (in a strong quantitative sense) subject to a mild ‘stationarity’ condition. Our result has a number of combinatorial consequences, including a new (and the most informative so far) proof of the density Hales-Jewett theorem.

1. Introduction

1.1. Motivation/Overview

Let I be a nonempty finite set and let $\langle E_i : i \in I \rangle$ and $\langle D_i : i \in I \rangle$ be stochastic processes (families of measurable events) in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with $\mathbb{P}(E_i) = \mathbb{P}(D_i) = \varepsilon > 0$ for all $i \in I$, and assume that the events $\langle E_i : i \in I \rangle$ are independent. We wish to compare the distributions of the random variables

$$X = \sum_{i \in I} \mathbf{1}_{E_i} \quad \text{and} \quad Y = \sum_{i \in I} \mathbf{1}_{D_i},$$

with the main goal being here that of transferring information from the distribution of X (which we understand very well) to the distribution of Y , an object on which we have a priori no control. A classical method for doing so is by comparing the moments of X and Y (see, e.g., [Du]), a task that essentially reduces¹ to that of comparing the joint probability of $\langle D_i : i \in F \rangle$ with the expected value $\varepsilon^{|F|}$ as F varies over all nonempty subsets of the index set I . Thus, assuming that the random variables X and Y are *not* close in distribution, one is led to the following problem.

Problem 1. Let $F \subseteq I$ be nonempty, let $\sigma > 0$ and assume that

$$\left| \mathbb{P} \left(\bigcap_{i \in F} D_i \right) - \varepsilon^{|F|} \right| \geq \sigma.$$

What structural information can be obtained for the process $\langle D_i : i \in I \rangle$?

¹This is because X and Y are both sums of indicator functions.

1.1.1. The combinatorial content

We will study Problem 1 in the case where the index set I is a *discrete hypercube*; that is, a set of the form

$$A^n := \underbrace{A \times \cdots \times A}_{n\text{-times}}, \tag{1.1}$$

where A is a finite set with $|A| \geq 2$ and n is a positive integer that is commonly referred to as the *dimension* of the hypercube A^n . This choice of the index set is by no means arbitrary and it is ultimately related to the density Hales-Jewett theorem, a deep result due to Furstenberg and Katznelson [FK2] with numerous consequences in combinatorics, number theory and theoretical computer science.

In order to properly discuss this relation, we need to recall some basic definitions. Let A and n be as above and fix a letter $x \notin A$ that we view as a variable. A *variable word over A of length n* is a finite sequence of length n having values in $A \cup \{x\}$ where the letter x appears at least once. If v is a variable word over A of length n and $\alpha \in A$, then let $v(\alpha)$ denote the unique element of A^n that is obtained by replacing every appearance of the letter x in v with α . (For instance, if $A = \{\alpha, \beta, \gamma\}$ and $v = (\alpha, x, \gamma, \beta, x)$, then $v(\beta) = (\alpha, \beta, \gamma, \beta, \beta)$.) A *combinatorial line of A^n* is a set of the form $\{v(\alpha) : \alpha \in A\}$ where v is a variable word over A of length n (see [GRS, HJ]).

We are now in a position to recall the density Hales-Jewett theorem. We will state a probabilistic version – see, for example, [FK2, Proposition 2.1] – which is closer in spirit to our discussion. The relation between this probabilistic version and the more well-known combinatorial form that refers to dense subsets of discrete hypercubes will be discussed in Section 4.

Theorem 1. *For every integer $k \geq 2$ and every $0 < \varepsilon \leq 1$ there exists a positive integer $\text{PHJ}(k, \varepsilon)$ with the following property. Let A be a set with $|A| = k$, let $n \geq \text{PHJ}(k, \varepsilon)$ be an integer and let $\langle D_t : t \in A^n \rangle$ be a stochastic process in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that $\mathbb{P}(D_t) \geq \varepsilon$ for every $t \in A^n$. Then there exists a combinatorial line L of A^n such that*

$$\mathbb{P}\left(\bigcap_{t \in L} D_t\right) > 0.$$

Of course, Theorem 1 is straightforward if the events $\langle D_t : t \in A^n \rangle$ are independent. Thus, the core of the theorem is to understand what happens when the events are not behaving as if they were independent, which is clearly an instance of Problem 1.

1.1.2. Deviating from the expected value: Examples

To gain insight on the kind of structure one expects to obtain in Problem 1, it is useful to give examples of stochastic processes that exhibit nonindependent behaviour. Here and in the rest of this introduction, we will restrict our discussion to correlations over combinatorial lines. This is mainly because of the combinatorial importance of this case but also because it is already quite representative of the behaviour of correlations over more complicated sets.

Example 1. For concreteness we will work with the set $\{1, 2, 3\}$, but the argument can also be applied for any finite set A with $|A| \geq 2$. Let n be an arbitrary positive integer. We start with a family $\langle E_s : s \in \{1, 2\}^n \rangle$ of independent events in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with equal probability $\varepsilon > 0$. Given $t \in \{1, 2, 3\}^n$ there are two natural ways to ‘project’ it into $\{1, 2\}^n$. Specifically, let $t^{3 \rightarrow 1}$ and $t^{3 \rightarrow 2}$ denote the unique elements of $\{1, 2\}^n$ that are obtained by replacing every appearance of 3 in t with 1 and 2 respectively. (For example, if $t = (3, 2, 1, 3, 1) \in \{1, 2, 3\}^5$, then $t^{3 \rightarrow 1} = (1, 2, 1, 1, 1)$ and $t^{3 \rightarrow 2} = (2, 2, 1, 2, 1)$.) Then let $\langle D_t : t \in \{1, 2, 3\}^n \rangle$ be defined by setting $D_t := E_{t^{3 \rightarrow 1}} \cap E_{t^{3 \rightarrow 2}}$ for every $t \in \{1, 2, 3\}^n$.

Although the process $\langle D_t : t \in \{1, 2, 3\}^n \rangle$ in Example 1 is, arguably, quite easy to define, the analysis of its properties requires some work.

1.1.2.1

We first observe that for every $t \in \{1, 2, 3\}^n$ that contains 3 we have

$$\mathbb{P}(D_t) = \varepsilon^2. \tag{1.2}$$

Because the density of the set of all elements of $\{1, 2, 3\}^n$ that do not contain 3 decreases exponentially with respect to the dimension n , we see that (1.2) holds true for ‘almost every’ t .

1.1.2.2

The second basic property of the process $\langle D_t : t \in \{1, 2, 3\}^n \rangle$ concerns its correlations over combinatorial lines. Specifically, let $L = \{v(1), v(2), v(3)\}$ be a combinatorial line of $\{1, 2, 3\}^n$ where v is a variable word over $\{1, 2, 3\}$ of length n that contains 3. Then we have

$$\mathbb{P}\left(\bigcap_{t \in L} D_t\right) = \varepsilon^4, \tag{1.3}$$

which implies that $\langle D_t : t \in \{1, 2, 3\}^n \rangle$ exhibits nonindependent² behaviour.

However, identity (1.3) shows yet another important property of this process. More precisely, if v_1, v_2 are variable words over $\{1, 2, 3\}$ of length n that both contain 3, then

$$\mathbb{P}(D_{v_1(1)} \cap D_{v_1(2)} \cap D_{v_1(3)}) = \mathbb{P}(D_{v_2(1)} \cap D_{v_2(2)} \cap D_{v_2(3)}). \tag{1.4}$$

In other words, the correlations of $\langle D_t : t \in \{1, 2, 3\}^n \rangle$ over combinatorial lines are essentially constant. This property is abstracted in the following definition that originates³ in the work of Furstenberg and Katznelson [FK2].

Definition 1 (Stationarity). Let A be a finite set with $|A| \geq 2$, let n be a positive integer, let $\eta > 0$ and let $\langle D_t : t \in A^n \rangle$ be a stochastic process in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We say that $\langle D_t : t \in A^n \rangle$ is η -stationary (with respect to combinatorial lines) if for every nonempty $\Gamma \subseteq A$ and every pair v_1, v_2 of variable words over A of length n we have

$$\left| \mathbb{P}\left(\bigcap_{\alpha \in \Gamma} D_{v_1(\alpha)}\right) - \mathbb{P}\left(\bigcap_{\alpha \in \Gamma} D_{v_2(\alpha)}\right) \right| \leq \eta. \tag{1.5}$$

(In particular, if $\langle D_t : t \in A^n \rangle$ is an η -stationary process, then for every pair L_1, L_2 of combinatorial lines of A^n we have $|\mathbb{P}(\bigcap_{t \in L_1} D_t) - \mathbb{P}(\bigcap_{t \in L_2} D_t)| \leq \eta$.)

Besides being very natural in this context,⁴ stationarity is not a particularly restrictive condition. Indeed, it follows from a classical result due to Graham and Rothschild [GR] that stationary processes are the building blocks of arbitrary processes. (See Fact 1 in the main text.)

1.1.2.3

The last, and most significant, property of the process $\langle D_t : t \in \{1, 2, 3\}^n \rangle$ is its hidden arithmetic structure, which is described in the following definition.

Definition 2 (Insensitivity). Let A be a finite set with $|A| \geq 2$, let n be a positive integer and let $\alpha, \beta \in A$ with $\alpha \neq \beta$.

- (1) Let $s, t \in A^n$ and write $s = (s_1, \dots, s_n)$ and $t = (t_1, \dots, t_n)$. We say that s, t are (α, β) -equivalent if for every $i \in \{1, \dots, n\}$ and every $\gamma \in A \setminus \{\alpha, \beta\}$ we have that $s_i = \gamma$ if and only if $t_i = \gamma$. (Namely, s, t are (α, β) -equivalent if they possibly differ only in the coordinates taking values in $\{\alpha, \beta\}$.)

²Specifically, by (1.2), the expected probability in (1.3) is ε^6 .

³The framework in [FK2] is somewhat different, but the essential content of Definition 1 is present in that work.

⁴In particular, note that, without assuming stationarity, one should instead study the behaviour of an appropriately weighted average, over all possible combinatorial lines, of the corresponding correlations of a process.

(2) We say that a stochastic process $\langle D_t : t \in A^n \rangle$ in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is (α, β) -insensitive provided that $D_s = D_t$ for every $s, t \in A^n$ that are (α, β) -equivalent.

The notion of insensitivity was introduced by Shelah [Sh] in his proof of the Hales-Jewett theorem [HJ], though it was originally referring to subsets of discrete hypercubes and not to stochastic processes (see Definition 5 in the main text); the difference, however, between the two frameworks is minor. In the discrete setting, insensitivity is the analogue⁵ of the concept of a (discrete) Hilbert cube that is ubiquitous in additive combinatorics and arithmetic Ramsey theory (see, e.g., [GRS, TV]). Insensitive processes have also been considered in [Au, Definition 5.1].

Now, taking into account the definition of $t^{3 \rightarrow 1}$ and $t^{3 \rightarrow 2}$ in Example 1, it is easy to see that the processes $\langle E_{t^{3 \rightarrow 1}} : t \in \{1, 2, 3\}^n \rangle$ and $\langle E_{t^{3 \rightarrow 2}} : t \in \{1, 2, 3\}^n \rangle$ are (1, 3)- and (2, 3)-insensitive, respectively. This property by itself yields that for every variable word v over $\{1, 2, 3\}$ of length n we have

$$D_{v(1)} \cap D_{v(2)} \cap D_{v(3)} = D_{v(1)} \cap D_{v(2)}. \tag{1.6}$$

Note that identity (1.6) implies, in a rather extreme way, that the events $D_{v(1)}, D_{v(2)}$ and $D_{v(3)}$ cannot be independent. Thus, we have a structural explanation of the fact that $\langle D_t : t \in \{1, 2, 3\}^n \rangle$ exhibits nonindependent behaviour: It is the intersection of insensitive processes.

1.2. The main result

The following theorem (which is one of the main results of this article and is proved in Section 3) shows that the example presented above is essentially the only example of a stationary process whose correlations over combinatorial lines deviate from what is expected.

Theorem 2. *Let $k \geq 2$ be an integer, and let $\varepsilon, \sigma, \eta > 0$ be such that*

$$\varepsilon \leq 1 - \frac{1}{2k}, \quad \sigma \leq \frac{\varepsilon^{k-1}}{2k} \quad \text{and} \quad \eta \leq \frac{\sigma}{4^{k-1}}. \tag{1.7}$$

Also let A be a set with $|A| = k$, let $n \geq k$ be an integer and let $\langle D_t : t \in A^n \rangle$ be an η -stationary process in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that $|\mathbb{P}(D_t) - \varepsilon| \leq \eta$ for every $t \in A^n$. Then, either

(i) *for every combinatorial line L of A^n and every nonempty $G \subseteq L$ we have*

$$\left| \mathbb{P}\left(\bigcap_{t \in G} D_t\right) - \varepsilon^{|G|} \right| \leq \sigma, \tag{1.8}$$

(ii) *or $\langle D_t : t \in A^n \rangle$ correlates with a ‘structured’ stochastic process; precisely, there exist a nonempty subset Γ of A , $\beta \in A \setminus \Gamma$ and a stochastic process $\langle S_t : t \in A^n \rangle$ in $(\Omega, \mathcal{F}, \mathbb{P})$ such that the following are satisfied:*

- (a) *For every $t \in A^n$ we have $S_t = \bigcap_{\alpha \in \Gamma} E_t^\alpha$ where for every $\alpha \in \Gamma$ the process $\langle E_t^\alpha : t \in A^n \rangle$ is (α, β) -insensitive.*
- (b) *For every $t \in A^n$ that contains β we have*

$$\mathbb{P}(S_t) \geq \frac{\varepsilon^{k-1}}{4k} \quad \text{and} \quad \mathbb{P}(D_t | S_t) \geq \varepsilon + \frac{\sigma}{4^{k-1}}. \tag{1.9}$$

Theorem 2 is a new result whose most surprising feature is perhaps the fact that the conditional probability $\mathbb{P}(D_t | S_t)$ depends linearly on the parameter σ . As is expected by Theorem 1, this information can in turn be used to prove the density Hales-Jewett theorem. We present this proof and we discuss

⁵This can be seen by identifying any nonempty finite set A with the interval $\{1, \dots, |A|\}$ and then projecting the hypercube A^n into the integers via the map $(\alpha_1, \dots, \alpha_n) \mapsto \sum_{i=1}^n \alpha_i |A|^{i-1}$.

in detail its quantitative aspects in Section 4. At this point we simply mention that it is a step towards obtaining primitive recursive bounds for the density Hales-Jewett numbers.

We also note that an infinitary structural result for stationary processes has been obtained by Austin in [Au, Theorem 6.2]; it is somewhat different than Theorem 2, but it provides additional information for a certain subclass of stationary processes.⁶ (See also [Tao] for a finitary and effective version of Austin’s approach.)

1.3. Correlations over arbitrary sets

Beyond its combinatorial consequences, Theorem 2 is also the starting point of the analysis of correlations of stochastic processes over arbitrary nonempty subsets of discrete hypercubes. This analysis leads to an answer to Problem 1, and it is presented in the second part of this article⁷ consisting of Sections 5 to 8. It can be seen as a natural – though not quite straightforward – generalisation of the study of correlations over combinatorial lines. Specifically, there are two notable differences.

Firstly, the argument relies on the notion of the *type*, a Ramsey-theoretic invariant that was introduced in [DKT2] and encodes the ‘geometry’ of a nonempty subset of a discrete hypercube. The definition of this invariant is recalled in Section 5, and it is crucially used in order to extend the notion of stationarity in this more general context (Definition 6 in the main text).

Secondly, the ‘structured’ process that appears in part (ii.a) of Theorem 2 depends on the type of the set G one is looking at in part (i). This dependence is controlled by another invariant – the *separation index* – which is introduced in Section 6. In particular, for correlations over sets that have the smallest possible separation index, we have the exact analogue of Theorem 2 (Theorem 5 in the main text); however, the analogy breaks down at this point and the ‘structured’ process that appears in part (ii.a) becomes more involved as the separation index increases (see Theorem 6 in the main text).

1.4. Outline of the argument

The proof of Theorem 2 proceeds in two steps. In the first step and assuming that part (i) does not hold true, we select a subset B of A such that for every variable word v over A of length n and every nonempty proper subset Σ of B the events $\langle D_{v(\alpha)} : \alpha \in \Sigma \rangle$ are essentially independent, yet the joint probability of $\langle D_{v(\alpha)} : \alpha \in B \rangle$ deviates from the expected value. We emphasise that this selection is possible because the process $\langle D_t : t \in A^n \rangle$ is stationary. The second step, which is the combinatorial heart of the matter, is to convert the irregularity of the correlations of $\langle D_t : t \in A^n \rangle$ into correlation with a single structured process. This is achieved by taking advantage of the uniform behaviour of $\langle D_{v(\alpha)} : \alpha \in B \rangle$ as v varies over all variable words over A of length n and by carefully using the ‘projections’ $t^{3 \rightarrow 1}$ and $t^{3 \rightarrow 2}$ described in Example 1 as well as their natural generalisations.

The argument for the case of correlations over arbitrary sets follows the same outline, though the details are – as expected – more complicated. We comment on the differences of the proof of the general case in Sections 7 and 8.

2. Combinatorial background

2.1

By $\mathbb{N} = \{0, 1, 2, \dots\}$ we denote the set of all natural numbers, and for every positive integer n we set $[n] := \{1, \dots, n\}$. For every set X we denote by $|X|$ its cardinality; moreover, for every subset A of X by A^C we denote the complement of A ; that is, $A^C := \{x \in X : x \notin A\}$.

⁶This is related to the notion of *satedness*; see [Au] for details.

⁷The two parts are largely independent of each other and can be read separately.

2.2. Definitions

Let A denote a finite set with $|A| \geq 2$.

2.2.1

As in (1.1), for every positive integer n by A^n we denote the Cartesian product of n many copies of A ; we view A^n as the set of all sequences of length n having values in A . Also let \emptyset denote the empty sequence, set $A^0 := \{\emptyset\}$ and let

$$A^{<\mathbb{N}} := \bigcup_{n \in \mathbb{N}} A^n \tag{2.1}$$

denote the set of all finite (possibly empty) sequences in A . For every $t, s \in A^{<\mathbb{N}}$ by $t \frown s$ we denote the concatenation of t and s ; notice, in particular, that if $t \in A^n$ and $s \in A^m$ for some $n, m \in \mathbb{N}$, then $t \frown s \in A^{n+m}$.

2.2.2. Variable words

Let n, m be positive integers, and fix a set $\{x_1, \dots, x_m\}$ that is disjoint from A ; we view $\{x_1, \dots, x_m\}$ as a set of variables. An m -variable word over A of length n is a finite sequence v of length n having values in $A \cup \{x_1, \dots, x_m\}$ such that (1) for every $i \in [m]$ the letter x_i appears in v at least once and (2) if $m \geq 2$, then for every $i, j \in [m]$ with $i < j$ all appearances of x_i in v precede all appearances of x_j . If v is an m -variable word over A of length n and $\alpha_1, \dots, \alpha_m \in A$, then by $v(\alpha_1, \dots, \alpha_m)$ we denote the unique element of A^n that is obtained by replacing every appearance of x_i in v with α_i for every $i \in [m]$. (For example, if $A = \{\alpha, \beta, \gamma\}$ and $v = (x_1, \gamma, x_2, x_2, \beta, x_3)$, then $v(\beta, \alpha, \gamma) = (\beta, \gamma, \alpha, \alpha, \beta, \gamma)$.)

2.2.3. Combinatorial spaces and canonical isomorphisms

A combinatorial space of $A^{<\mathbb{N}}$ is a subset V of $A^{<\mathbb{N}}$ of the form

$$V = \{v(\alpha_1, \dots, \alpha_m) : \alpha_1, \dots, \alpha_m \in A\}, \tag{2.2}$$

where m is a positive integer and v is an m -variable word over A of length n for some positive integer n (in particular, we have $V \subseteq A^n$.) Notice that m, v and n are unique because $|A| \geq 2$; the (unique) positive integer m is called the dimension of V and is denoted by $\dim(V)$. Also observe that the 1-dimensional combinatorial spaces are precisely the combinatorial lines already mentioned in the introduction. Finally, if V_1 and V_2 are two combinatorial spaces of $A^{<\mathbb{N}}$, then we say that V_1 is a combinatorial subspace of V_2 provided that $V_1 \subseteq V_2$.

We view an m -dimensional combinatorial space V as a ‘copy’ of A^m inside $A^{<\mathbb{N}}$, and we will identify V with A^m for most practical purposes. To this end, we introduce the following definition.

Definition 3. Let A be a finite set with $|A| \geq 2$, and let V be a combinatorial space of $A^{<\mathbb{N}}$. Set $m := \dim(V)$ and let v be the unique m -variable word over A that generates V via formula (2.2). The canonical isomorphism associated with V is the bijection $I_V : A^m \rightarrow V$ defined by the rule

$$I_V((\alpha_1, \dots, \alpha_m)) = v(\alpha_1, \dots, \alpha_m) \tag{2.3}$$

for every $(\alpha_1, \dots, \alpha_m) \in A^m$.

Note that canonical isomorphisms preserve combinatorial subspaces and their dimension; precisely, if V is an m -dimensional combinatorial space of $A^{<\mathbb{N}}$ and $W \subseteq A^m$, then W is a combinatorial subspace of A^m with $\dim(W) = \ell$ if and only if $I_V(W)$ is a combinatorial subspace of V with $\dim(I_V(W)) = \ell$. For an exposition of the properties of canonical isomorphisms we refer to [DK, Section 1.3].

2.3. Colorings of combinatorial lines

We will need the following special case⁸ of the Graham-Rothschild theorem [GR]. The corresponding primitive recursive bounds are taken from [Ty].

Proposition 1. *For every triple k, m, r of positive integers with $k \geq 2$ there exists a positive integer N with the following property. For every set A with $|A| = k$, every combinatorial space V of $A^{<\mathbb{N}}$ with $\dim(V) \geq n$ and every r -coloring of the set of all combinatorial lines of V there exists an m -dimensional combinatorial subspace W of V such that the set of all combinatorial lines of W is monochromatic. The least positive integer N with this property is denoted by $\text{GRL}(k, m, r)$.*

Moreover, the numbers $\text{GRL}(k, m, r)$ are upper bounded by a primitive recursive function belonging to the class \mathcal{E}^5 of Grzegorzcyk's hierarchy.

For a discussion of Grzegorzcyk's hierarchy of primitive recursive functions and its role in analyzing the bounds associated with various results in Ramsey theory, we refer to [DK, Appendix A].

3. Correlations over combinatorial lines

In this section we give the proof of Theorem 2. As we have noted in the Introduction, the argument (which also pertains to the proofs of Theorems 5 and 6) can be roughly summarised by saying that higher order correlations of a process can be converted into correlation with a single structured process. Perhaps the most transparent instance of this fact is the proof of Proposition 2.

We begin with some preliminary steps, including a discussion of some basic properties of stationary processes.

3.1. Stationarity

We have already noted that the Graham-Rothschild theorem (more precisely, Proposition 1) implies that stationary processes are the building blocks of arbitrary processes. In particular, we have the following fact. The proof is straightforward.

Fact 1. *Let $k \geq 2$ be an integer, and let A be a set with $|A| = k$. Also let $0 < \eta \leq 1$, and let n, m be positive integers such that*

$$n \geq \text{GRL}(k, m, \lceil 1/\eta \rceil^{2^k - 1}). \tag{3.1}$$

Then for every stochastic process $\langle D_t : t \in A^n \rangle$ in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ there exists an m -dimensional combinatorial subspace V of A^n such that the process $\langle D_{\Gamma_V(s)} : s \in A^m \rangle$ (namely, the restriction of $\langle D_t : t \in A^n \rangle$ to V) is η -stationary.

The following lemma shows that one can upgrade the estimate in (1.5) and stabilise the joint distribution of certain Boolean combinations of the events of a stationary processes. (Here, and in the rest of this article, we follow that convention that the intersection of an empty family of events of a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is equal to the sample space Ω .)

Lemma 1. *Let A be a finite set with $|A| \geq 2$, let n be a positive integer, let $\eta > 0$ and let $\langle D_t : t \in A^n \rangle$ be an η -stationary stochastic process in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then for every pair $\Gamma_1, \Gamma_2 \subseteq A$ with $\Gamma_1 \cap \Gamma_2 = \emptyset$ and every pair v_1, v_2 of variable words over A of length n , we have*

$$\left| \mathbb{P} \left(\bigcap_{\alpha \in \Gamma_1} D_{v_1(\alpha)} \cap \bigcap_{\alpha \in \Gamma_2} D_{v_1(\alpha)}^c \right) - \mathbb{P} \left(\bigcap_{\alpha \in \Gamma_1} D_{v_2(\alpha)} \cap \bigcap_{\alpha \in \Gamma_2} D_{v_2(\alpha)}^c \right) \right| \leq 2^{|\Gamma_2|} \eta. \tag{3.2}$$

⁸Actually, the Graham-Rothschild theorem refers to *parameter words*, a concept that is slightly different from the notion of a variable word. However, for colorings of combinatorial lines the difference between the two frameworks is minor.

Proof. Let $\Gamma_1, \Gamma_2, \nu_1, \nu_2$ be as in the statement of the lemma. Then, using the inclusion–exclusion formula, we have

$$\begin{aligned} & \left| \mathbb{P} \left(\bigcap_{\alpha \in \Gamma_1} D_{\nu_1(\alpha)} \cap \bigcap_{\alpha \in \Gamma_2} D_{\nu_1(\alpha)}^c \right) - \mathbb{P} \left(\bigcap_{\alpha \in \Gamma_1} D_{\nu_2(\alpha)} \cap \bigcap_{\alpha \in \Gamma_2} D_{\nu_2(\alpha)}^c \right) \right| \\ &= \left| \mathbb{P} \left(\bigcap_{\alpha \in \Gamma_1} D_{\nu_1(\alpha)} \cap \left(\bigcup_{\alpha \in \Gamma_2} D_{\nu_1(\alpha)} \right)^c \right) - \mathbb{P} \left(\bigcap_{\alpha \in \Gamma_1} D_{\nu_2(\alpha)} \cap \left(\bigcup_{\alpha \in \Gamma_2} D_{\nu_2(\alpha)} \right)^c \right) \right| \\ &\leq \sum_{\Gamma \subseteq \Gamma_2} \left| \mathbb{P} \left(\bigcap_{\alpha \in \Gamma_1} D_{\nu_1(\alpha)} \cap \bigcap_{\alpha \in \Gamma} D_{\nu_1(\alpha)} \right) - \mathbb{P} \left(\bigcap_{\alpha \in \Gamma_1} D_{\nu_2(\alpha)} \cap \bigcap_{\alpha \in \Gamma} D_{\nu_2(\alpha)} \right) \right| \stackrel{(1.5)}{\leq} 2^{|\Gamma_2|} \eta \end{aligned}$$

and the proof is completed. □

Remark 1. We notice that the assumption in Theorem 2 that $|\mathbb{P}(D_t) - \varepsilon| \leq \eta$ for every $t \in A^n$ follows from η -stationarity provided that the dimension n is sufficiently large. Indeed, let A, n and $\langle D_t : t \in A^n \rangle$ be as in Theorem 2; clearly, we have $n \geq |A|$. We select $t_0 \in A^n$ such that every $\alpha \in A$ appears in t_0 at least once (this selection is possible because $n \geq |A|$), and we set $\varepsilon := \max\{\mathbb{P}(D_{t_0}), \eta\} > 0$. Note that for every $t \in A^n$ there exist two variable words ν_1, ν_2 over A of length n and $\alpha \in A$ such that $t = \nu_1(\alpha)$ and $t_0 = \nu_2(\alpha)$. Invoking (1.5), we conclude that $|\mathbb{P}(D_t) - \varepsilon| \leq \eta$.

3.2. Insensitivity

Let A be a finite set with $|A| \geq 2$, let n be a positive integer and let $\alpha, \beta \in A$ with $\alpha \neq \beta$. As in Example 1, for every $t \in A^n$ let $t^{\beta \rightarrow \alpha}$ denote the unique element of $(A \setminus \{\beta\})^n$ that is obtained by replacing every appearance of β in t with α . We will use this operation in order to produce insensitive processes. To this end, we will need the following elementary (though crucial) fact. Its proof is straightforward.

Fact 2. *Let A be a finite set with $|A| \geq 2$, let n be a positive integer and let $\alpha, \beta \in A$ with $\alpha \neq \beta$. Then the map $A^n \ni t \mapsto t^{\beta \rightarrow \alpha} \in (A \setminus \{\beta\})^n$ is a projection; that is, for every $t \in A^n$ that does not contain β we have that $t^{\beta \rightarrow \alpha} = t$. Moreover, if $t, s \in A^n$ are (α, β) -equivalent, then $t^{\beta \rightarrow \alpha} = s^{\beta \rightarrow \alpha}$.*

3.3. Pseudorandomness, supercorrelation, subcorrelation

Let E_1, \dots, E_ℓ be measurable events in a probability space with equal probability $\varepsilon > 0$. Notice that the joint probability of E_1, \dots, E_ℓ can be naturally categorised according to whether it is greater than, less than, or almost equal to the expected value ε^ℓ . As expected, our analysis depends on this trichotomy and, as such, it is convenient to introduce the following definition.

Definition 4. Let A be a finite set with $|A| \geq 2$, let $n \geq |A|$ be an integer, let $0 < \eta, \varepsilon \leq 1$ and let $\langle D_t : t \in A^n \rangle$ be an η -stationary process in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that $|\mathbb{P}(D_t) - \varepsilon| \leq \eta$ for every $t \in A^n$. Also let $\Gamma \subseteq A$ be nonempty, and let $\theta \geq 0$.

- (1) (*Pseudorandomness*) We say that $\langle D_t : t \in A^n \rangle$ is (Γ, θ) -pseudorandom if $|\mathbb{P}(\bigcap_{\alpha \in \Gamma} D_{\nu(\alpha)}) - \varepsilon^{|\Gamma|}| \leq \theta$ for every variable word ν over A of length n .
- (2) (*Supercorrelation*) We say that $\langle D_t : t \in A^n \rangle$ is (Γ, θ) -supercorrelated if $\mathbb{P}(\bigcap_{\alpha \in \Gamma} D_{\nu(\alpha)}) \geq \varepsilon^{|\Gamma|} + \theta$ for every variable word ν over A of length n .
- (3) (*Subcorrelation*) We say that $\langle D_t : t \in A^n \rangle$ is (Γ, θ) -subcorrelated if $\mathbb{P}(\bigcap_{\alpha \in \Gamma} D_{\nu(\alpha)}) \leq \varepsilon^{|\Gamma|} - \theta$ for every variable word ν over A of length n .

We have the following fact.

Fact 3. *Let A, n, η, ε and $\langle D_t : t \in A^n \rangle$ be as in Definition 4. Also let $\Gamma \subseteq A$ be nonempty, and let $\theta \geq \eta$. Then one of the following holds true.*

- (i) *The process $\langle D_t : t \in A^n \rangle$ is (Γ, θ) -pseudorandom.*

- (ii) The process $\langle D_t : t \in A^n \rangle$ is $(\Gamma, \theta - \eta)$ -supercorrelated.
- (iii) The process $\langle D_t : t \in A^n \rangle$ is $(\Gamma, \theta - \eta)$ -subcorrelated.

Proof. Assume that (i) does not hold true; that is, there is a variable word v over A of length n such that either $\mathbb{P}(\bigcap_{\alpha \in \Gamma} D_{v(\alpha)}) \geq \varepsilon^{|\Gamma|} + \theta$ or $\mathbb{P}(\bigcap_{\alpha \in \Gamma} D_{v(\alpha)}) \leq \varepsilon^{|\Gamma|} - \theta$. Invoking the η -stationarity of $\langle D_t : t \in A^n \rangle$, we see that the first alternative yields part (ii), and the second alternative yields part (iii). □

We are ready to state the main result in this subsection.

Proposition 2. *Let A be a finite set with $|A| \geq 2$, and let $n \geq |A|$ be an integer. Also let $0 < \eta, \varepsilon \leq 1$, and let $\langle D_t : t \in A^n \rangle$ be an η -stationary stochastic process in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that $|\mathbb{P}(D_t) - \varepsilon| \leq \eta$ for every $t \in A^n$. Finally, let $\theta, \sigma \geq 0$, let $\Gamma \subseteq A$ be nonempty and let $\beta \in A \setminus \Gamma$. Assume that $\langle D_t : t \in A^n \rangle$ is (Γ, θ) -pseudorandom, and set $p := |\Gamma|$. Then there exists a stochastic process $\langle S_t : t \in A^n \rangle$ in $(\Omega, \mathcal{F}, \mathbb{P})$ with the following properties.*

- (i) For every $t \in A^n$ we have $S_t = \bigcap_{\alpha \in \Gamma} E_t^\alpha$ where for every $\alpha \in \Gamma$ the process $\langle E_t^\alpha : t \in A^n \rangle$ is (α, β) -insensitive.
- (ii) For every $t \in A^n$ that does not contain β and every $\alpha \in \Gamma$ we have $E_t^\alpha = D_t$. (Thus, $S_t = D_t$ for every $t \in A^n$ that does not contain β .)
- (iii) For every $t \in A^n$ that contains β we have $|\mathbb{P}(S_t) - \varepsilon^p| \leq \theta$.
- (iv) If $\langle D_t : t \in A^n \rangle$ is $(\Gamma \cup \{\beta\}, \sigma)$ -supercorrelated, then for every $t \in A^n$ that contains β we have

$$\mathbb{P}(D_t | S_t) \geq \varepsilon \left(1 + \frac{\sigma \varepsilon^{-1} - \theta}{\varepsilon^p + \theta} \right). \tag{3.3}$$

- (v) If $\langle D_t : t \in A^n \rangle$ is $(\Gamma \cup \{\beta\}, \sigma)$ -subcorrelated, then for every $t \in A^n$ that contains β we have

$$\mathbb{P}(D_t | S_t) \leq \varepsilon \left(1 - \frac{\sigma \varepsilon^{-1} - \theta}{\varepsilon^p - \theta} \right). \tag{3.4}$$

Proof. We first observe that the conditions in parts (i) and (ii) completely determine the stochastic process $\langle E_t^\alpha : t \in A^n \rangle$ for every $\alpha \in \Gamma$. However, it is possible to give an alternative (and more intrinsic) definition of these processes that facilitates the proofs of parts (iii)–(v) and it is easier to generalise when considering correlations over more complicated sets (see, in particular, Sections 7 and 8). More precisely, notice that, by Fact 2, we have $E_t^\alpha = D_{t\beta \rightarrow \alpha}$ for every $t \in A^n$ and every $\alpha \in \Gamma$. We will also need the following important property of this construction. For every $t \in A^n$ that contains β let v_t denote the unique variable word over $A \setminus \{\beta\}$ of length n that is obtained by replacing every appearance of β in t with the variable x and note that $t = v_t(\beta)$ and $t^{\beta \rightarrow \alpha} = v_t(\alpha)$ for every $\alpha \in \Gamma$. Consequently, for every $t \in A^n$ that contains β we have

$$S_t = \bigcap_{\alpha \in \Gamma} D_{v_t(\alpha)} \quad \text{and} \quad D_t \cap S_t = \bigcap_{\alpha \in \Gamma \cup \{\beta\}} D_{v_t(\alpha)}. \tag{3.5}$$

After this preliminary discussion, we are ready to proceed to the rest of the proof. Part (iii) follows immediately by the first identity in (3.5) and our assumption that the process $\langle D_t : t \in A^n \rangle$ is (Γ, θ) -pseudorandom.

For part (iv), assume that $\langle D_t : t \in A^n \rangle$ is $(\Gamma \cup \{\beta\}, \sigma)$ -supercorrelated. Fix $t \in A^n$ that contains β . By the second identity in (3.5) and the supercorrelation assumption, we see that $\mathbb{P}(D_t \cap S_t) \geq \varepsilon^{p+1} + \sigma$; on the other hand, by part (iii), we have $\mathbb{P}(S_t) \leq \varepsilon^p + \theta$. Therefore,

$$\mathbb{P}(D_t | S_t) = \frac{\mathbb{P}(D_t \cap S_t)}{\mathbb{P}(S_t)} \geq \frac{\varepsilon^{p+1} + \sigma}{\varepsilon^p + \theta} = \varepsilon \left(1 + \frac{\sigma \varepsilon^{-1} - \theta}{\varepsilon^p + \theta} \right)$$

as desired.

Finally, assume that $\langle D_t : t \in A^n \rangle$ is $(\Gamma \cup \{\beta\}, \sigma)$ -subcorrelated, and fix $t \in A^n$ that contains β . As above, using the second identity in (3.5) and the subcorrelation assumption, we obtain that $\mathbb{P}(D_t \cap S_t) \leq \varepsilon^{p+1} - \sigma$. By part (iii), we have $\mathbb{P}(S_t) \geq \varepsilon^p - \theta$, and so

$$\mathbb{P}(D_t | S_t) = \frac{\mathbb{P}(D_t \cap S_t)}{\mathbb{P}(S_t)} \leq \frac{\varepsilon^{p+1} - \sigma}{\varepsilon^p - \theta} = \varepsilon \left(1 - \frac{\sigma \varepsilon^{-1} - \theta}{\varepsilon^p - \theta} \right).$$

The proof is completed. □

Remark 2. Observe that the variable word v_t defined in the proof of Proposition 2 is not typical because it does not contain β . Nevertheless, because stationarity is a *global* property, it is possible to have information for the correlation of the events $\langle D_{v_t(\alpha)} : \alpha \in \Gamma \rangle$. This fact (namely, the necessity to understand the correlations of $\langle D_t : t \in A^n \rangle$ over sparse sets of combinatorial lines) is rather subtle and appears to be a genuine obstacle for extending Theorem 2 to not necessarily stationary processes.

Remark 3 (Extreme cases). Note that the extreme cases in Proposition 2 are (a) ‘ $\theta = 0$ ’ and ‘ $\sigma = \varepsilon^p - \varepsilon^{p+1}$ ’, if the stochastic process $\langle D_t : t \in A^n \rangle$ is supercorrelated and (b) ‘ $\theta = 0$ ’ and ‘ $\sigma = \varepsilon^{p+1}$ ’, if $\langle D_t : t \in A^n \rangle$ is subcorrelated. In the first case we have that $\mathbb{P}(D_t | S_t) = 1$ for every $t \in A^n$ containing β , which is clearly equivalent to saying that $S_t \subseteq D_t$. Examples of stochastic processes of this form can be obtained by modifying (in a straightforward way) Example 1. At the other extreme, we see that $\mathbb{P}(D_t | S_t) = 0$ for every $t \in A^n$ that contains β . In contrast to the previous case, this phenomenon cannot occur if the dimension n is sufficiently large; this is a consequence of Theorem 1.

3.4. Proof of Theorem 2

We begin by introducing a finite sequence $(\theta_p)_{p=0}^k$ of positive reals defined by the rule

$$\begin{cases} \theta_0 = 0, \theta_1 = \eta, \\ \theta_p = 4^{p-k} \sigma \text{ if } p \in \{2, \dots, k\}. \end{cases} \tag{3.6}$$

(Note that, by (1.7), the sequence $(\theta_p)_{p=0}^k$ is increasing.) Next observe that if for every nonempty $\Gamma \subseteq A$ the process $\langle D_t : t \in A^n \rangle$ is $(\Gamma, \theta_{|\Gamma|})$ -pseudorandom, then part (i) of the theorem holds true. Therefore, we may assume that there exists nonempty $\Delta \subseteq A$ such that $\langle D_t : t \in A^n \rangle$ is *not* $(\Delta, \theta_{|\Delta|})$ -pseudorandom. We fix a nonempty subset Γ_0 of A that satisfies this property and with minimal cardinality. (Notice, in particular, that if Σ is a nonempty proper subset of Γ_0 , then $\langle D_t : t \in A^n \rangle$ is $(\Sigma, \theta_{|\Sigma|})$ -pseudorandom.) Because the process $\langle D_t : t \in A^n \rangle$ is η -stationary and $\theta_1 = \eta$, we see that $|\Gamma_0| \geq 2$. We select $\beta \in \Gamma_0$, and we set $\Gamma := \Gamma_0 \setminus \{\beta\}$ and $p := |\Gamma|$; observe that $1 \leq p \leq k - 1$. Set $\theta := \theta_p$ and $\Theta := \theta_{p+1}$.

By Fact 3 and our assumption that the stochastic process $\langle D_t : t \in A^n \rangle$ is not $(\Gamma \cup \{\beta\}, \Theta)$ -pseudorandom, we see that either

- (A1) $\langle D_t : t \in A^n \rangle$ is $(\Gamma \cup \{\beta\}, \Theta - \eta)$ -supercorrelated or
- (A2) $\langle D_t : t \in A^n \rangle$ is $(\Gamma \cup \{\beta\}, \Theta - \eta)$ -subcorrelated.

We will show that in both cases part (ii) of the theorem holds true.

CASE 1: $\langle D_t : t \in A^n \rangle$ is $(\Gamma \cup \{\beta\}, \Theta - \eta)$ -supercorrelated. By Proposition 2 applied for ‘ $\sigma = \Theta - \eta$ ’, there exists a process $\langle S_t : t \in A^n \rangle$ that satisfies part (ii.a) of the theorem such that for every $t \in A^n$ that contains β we have

- (a) $|\mathbb{P}(S_t) - \varepsilon^p| \leq \theta$ and
- (b) $\mathbb{P}(D_t | S_t) \geq \varepsilon \left(1 + \frac{\Theta \varepsilon^{-1} - \eta \varepsilon^{-1} - \theta}{\varepsilon^p + \theta} \right)$.

Therefore, by (a) above and the fact that $\theta \leq \sigma/4$, for every $t \in A^n$ that contains β we have

$$\mathbb{P}(S_t) \geq \varepsilon^p - \theta \geq \varepsilon^{k-1} - \frac{\sigma}{4} \stackrel{(1.7)}{\geq} \frac{\varepsilon^{k-1}}{4k}$$

and, by (b) and the fact that $\eta, \theta \leq \Theta/4$,

$$\begin{aligned} \mathbb{P}(D_t | S_t) &\geq \varepsilon + \frac{\Theta - \eta - \varepsilon\theta}{\varepsilon^p + \theta} \geq \varepsilon + \frac{1}{2}(\Theta - \eta - \theta) \\ &\geq \varepsilon + \frac{\Theta}{4} \geq \varepsilon + \frac{\theta_2}{4} = \varepsilon + \frac{\sigma}{4^{k-1}}. \end{aligned}$$

Thus, part (ii.b) of the theorem is also satisfied, as desired.

CASE 2: $\langle D_t : t \in A^n \rangle$ is $(\Gamma \cup \{\beta\}, \Theta - \eta)$ -subcorrelated. For every $t \in A^n$ and every $\alpha \in \Gamma$ set $\mathcal{E}_t^\alpha := D_{t^{\beta \rightarrow \alpha}}$. (Recall that $t^{\beta \rightarrow \alpha}$ denotes the unique element of $(A \setminus \{\beta\})^n$ that is obtained by replacing every appearance of β in t with α .) We select $\gamma \in \Gamma$, we set $B := \Gamma \setminus \{\gamma\}$ and for every $t \in A^n$ we define

$$S_t := \left(\bigcap_{\alpha \in B} \mathcal{E}_t^\alpha \right) \cap (\mathcal{E}_t^\gamma)^c. \tag{3.7}$$

(Recall that, by convention, $\bigcap_{\alpha \in B} \mathcal{E}_t^\alpha = \Omega$ if $B = \emptyset$.) Clearly, the stochastic process $\langle S_t : t \in A^n \rangle$ satisfies part (ii.a) of the theorem. As in the proof of Proposition 2, for every $t \in A^n$ that contains β by v_t we denote the unique variable word over $A \setminus \{\beta\}$ of length n that is obtained by replacing every appearance of β in t with the variable x ; recall that $t = v_t(\beta)$ and $t^{\beta \rightarrow \alpha} = v_t(\alpha)$ for every $\alpha \in \Gamma$. Consequently, for every $t \in A^n$ that contains β we have

$$S_t = \left(\bigcap_{\alpha \in B} D_{v_t(\alpha)} \right) \setminus \left(\bigcap_{\alpha \in \Gamma} D_{v_t(\alpha)} \right) \tag{3.8}$$

and

$$D_t \cap S_t = \left(\bigcap_{\alpha \in B \cup \{\beta\}} D_{v_t(\alpha)} \right) \setminus \left(\bigcap_{\alpha \in \Gamma \cup \{\beta\}} D_{v_t(\alpha)} \right). \tag{3.9}$$

Because $\langle D_t : t \in A^n \rangle$ is

- $(\Gamma \cup \{\beta\}, \Theta - \eta)$ -subcorrelated,
- (B, θ_{p-1}) -pseudorandom if $p > 1$ (if $p = 1$, then this is superfluous) and
- (Γ, θ) -pseudorandom and $(B \cup \{\beta\}, \theta)$ -pseudorandom,

for every $t \in A^n$ that contains β we have

$$\mathbb{P}(S_t) \leq (\varepsilon^{p-1} + \theta_{p-1}) - (\varepsilon^p - \theta) \quad \text{and} \quad \mathbb{P}(D_t \cap S_t) \geq (\varepsilon^p - \theta) - (\varepsilon^{p+1} - \Theta + \eta).$$

Moreover, by (1.7) and (3.6), we have $\theta + \theta_{p-1} \leq \varepsilon^p$, $\theta \leq \Theta/4$ and $\theta_{p-1} + \eta \leq \Theta/4$. Therefore, for every $t \in A^n$ that contains β ,

$$\begin{aligned} \mathbb{P}(D_t | S_t) &\geq \frac{\varepsilon^p - \varepsilon^{p+1} + \Theta - \theta - \eta}{\varepsilon^{p-1} - \varepsilon^p + \theta + \theta_{p-1}} \geq \varepsilon + \frac{\Theta - 2\theta - \theta_{p-1} - \eta}{\varepsilon^{p-1} - \varepsilon^p + \theta + \theta_{p-1}} \\ &\geq \varepsilon + \frac{\Theta}{4} \geq \varepsilon + \frac{\sigma}{4^{k-1}}. \end{aligned}$$

Finally, by (1.7), (3.8) and the fact that $\langle D_t : t \in A^n \rangle$ is (B, θ_{p-1}) -pseudorandom and (Γ, θ) -pseudorandom, we conclude that

$$\mathbb{P}(S_t) \geq \varepsilon^{p-1} - \varepsilon^p - \theta - \theta_{p-1} \geq \varepsilon^{k-1}(1 - \varepsilon) - 2\theta \geq \frac{\varepsilon^{k-1}}{2k} - 2\theta \geq \frac{\varepsilon^{k-1}}{4k}$$

for every $t \in A^n$ that contains β . The proof is completed.

4. Proof of the density Hales-Jewett theorem

4.1

In this section we give a proof of the density Hales-Jewett theorem that is based on Theorem 2. We begin by recalling the combinatorial version of the density Hales-Jewett theorem. (The reader is advised to compare this version with Theorem 1 stated in the Introduction.)

Theorem 3. *For every integer $k \geq 2$ and every $0 < \delta \leq 1$ there exists a positive integer $\text{DHJ}(k, \delta)$ with the following property. Let A be a set with $|A| = k$, and let $n \geq \text{DHJ}(k, \delta)$ be an integer. Then every $D \subseteq A^n$ with $|D| \geq \delta|A^n|$ contains a combinatorial line of A^n .*

There are several effective proofs⁹ of Theorem 3; see [DKT1, P2, Tao]. Despite this progress, the understanding of the behaviour of the density Hales-Jewett numbers $\text{DHJ}(k, \delta)$ is rather poor. Indeed, the best known upper bounds are obtained in [P2] and have an Ackermann-type dependence with respect to k .

The proof of Theorem 3 given below is based on a density increment strategy (a method introduced by Roth [Ro]) and follows the general scheme developed in [P2]. Its most important feature is the quantitative improvement of a crucial step that appears (in various forms) in all known combinatorial proofs of the density Hales-Jewett theorem. (We discuss this particular feature in Remark 6.) The driving force behind this improvement is Theorem 2.

4.1.1. Step 1: From dense subsets of discrete hypercubes to stochastic processes

Strictly speaking, this step is not an internal part of the proof of Theorem 3. However, it is conceptually significant because it enables us to pass from dense sets to stochastic processes. This is essentially the content of the following simple lemma, whose proof can be found in [DKT1, Lemma 4].

Lemma 2. *Let k, m be positive integers with $k \geq 2$, let $0 < \eta \leq 1$, let A be a set with $|A| = k$ and let n be a positive integer such that*

$$n \geq \frac{k^m m}{\eta}. \tag{4.1}$$

Then for every $D \subseteq A^n$ there exist $\ell \in \{m, \dots, n - 1\}$ and an m -dimensional combinatorial subspace V of A^ℓ such that for every $t \in V$ we have

$$\frac{|D_t|}{|A^{n-\ell}|} \geq \frac{|D|}{|A^n|} - \eta, \tag{4.2}$$

where $D_t = \{s \in A^{n-\ell} : t \hat{\cap} s \in D\}$ denotes the section of D at t .

Remark 4. There is a more powerful probabilistic version of Lemma 2 that can be stated as a concentration inequality and relies on properties of martingale difference sequences; see [DKT3, Theorem 1]. See also [DK, Chapter 6] for a discussion on the role of this result in density Ramsey theory.

Remark 5. Lemma 2 can be used to relate the numerical invariants $\text{PHJ}(k, \varepsilon)$ and $\text{DHJ}(k, \delta)$ associated with the two versions of the density Hales-Jewett theorem. Indeed, notice that for every integer $k \geq 2$ and every $0 < \theta < \varepsilon \leq 1$ we have

$$\text{PHJ}(k, \varepsilon) \leq \text{DHJ}(k, \varepsilon) \leq (\varepsilon - \theta)^{-1} \cdot \text{PHJ}(k, \theta) \cdot k^{\text{PHJ}(k, \theta)}. \tag{4.3}$$

4.1.2. Step 2: Obtaining correlation with an insensitive set

We start by introducing the combinatorial analogue of the notion of an insensitive process. We note that this combinatorial analogue in fact predates Definition 2.

⁹Another ergodic-theoretic proof was given in [Au].

Definition 5. Let A be a finite set with $|A| \geq 2$, let n be a positive integer and let $\alpha, \beta \in A$ with $\alpha \neq \beta$.

- (1) We say that a subset \mathcal{E} of A^n is (α, β) -insensitive if for every $s, t \in A^n$ that are (α, β) -equivalent we have that $t \in \mathcal{E}$ if and only if $s \in \mathcal{E}$.
- (2) We say that a subset \mathcal{E} of an n -dimensional combinatorial space V of $A^{<\mathbb{N}}$ is (α, β) -insensitive in V if $\Gamma_V^{-1}(\mathcal{E})$ is (α, β) -insensitive, where $\Gamma_V : A^n \rightarrow V$ denotes the canonical isomorphism associated with V .

The following lemma is the second step of the proof of Theorem 3. It is precisely in the proof of this step that Theorem 2 is applied.

Lemma 3. Let $m \geq k \geq 2$ be integers, and let $0 < \delta \leq 1$. Set

$$N = \text{GRL}(k + 1, m + 1, \lceil 2(k + 1)4^k \delta^{-(k+1)} \rceil^{(2^{k+1}-1)}) \tag{4.4}$$

and let n be a positive integer such that

$$n \geq \frac{2(k + 1)4^k}{\delta^{k+1}} (k + 1)^N N. \tag{4.5}$$

Let A be a set with $|A| = k + 1$, and let $D \subseteq A^n$ with $|D| \geq \delta|A^n|$. Then, either

- (i) D contains a combinatorial line of A^n or
- (ii) there exist $\beta \in A$, an m -dimensional combinatorial subspace V of A^n and a subset \mathcal{S} of V with the following properties:
 - (a) We have $\mathcal{S} = \bigcap_{\alpha \in A \setminus \{\beta\}} \mathcal{E}^\alpha$ where for every $\alpha \in A \setminus \{\beta\}$ the set \mathcal{E}^α is (α, β) -insensitive in V .
 - (b) We have

$$\frac{|\mathcal{S}|}{|V|} \geq \frac{\delta^{2k+1}}{(k + 1)^2 4^{k+2}} \quad \text{and} \quad \frac{|D \cap \mathcal{S}|}{|\mathcal{S}|} \geq \delta + \frac{\delta^{k+1}}{(k + 1) 4^{k+1}}. \tag{4.6}$$

Remark 6. Lemma 3 improves upon two important quantitative aspects of what was known before. Firstly, by Proposition 1, the threshold on the dimension n appearing in (4.5) is bounded by a primitive recursive function that belongs to the class \mathcal{E}^5 of Grzegorzczuk’s hierarchy; in particular, it is independent of the numbers $\text{DHJ}(k, \delta)$. Secondly, the increment of the density of the set D obtained in the second part of (4.6) depends polynomially on δ ; in order to appreciate this particular improvement, we recall that all previous proofs yield a density increment that has an Ackermann-type dependence with respect to k . We also note that this quantity controls the number of iterations needed to be performed in order to prove Theorem 3 and, as such, it has significant impact on the behaviour of the density Hales-Jewett numbers.

Proof of Lemma 3. We set $\eta := \frac{\delta^{k+1}}{2(k+1)4^k}$. By Lemma 2 and (4.5), there exist $\ell \in \{m, \dots, n - 1\}$ and an N -dimensional combinatorial subspace V_1 of A^ℓ such that for every $t \in V_1$ we have

$$\frac{|D_t|}{|A^{n-\ell}|} \geq \frac{|D|}{|A^n|} - \eta. \tag{4.7}$$

We view the set $A^{n-\ell}$ as a discrete probability measure equipped with the uniform probability measure that we shall denote by \mathbb{P}_1 . By Fact 1 and (4.4), there exists an $(m + 1)$ -dimensional combinatorial subspace V_2 of V_1 such that the process $\langle D_{\Gamma_{V_2}(t)} : t \in A^{m+1} \rangle$ is η -stationary; consequently, by Remark 1, (4.7) and the fact that $|D| \geq \delta|A^n|$, there exists $\varepsilon \geq \delta$ such that $|\mathbb{P}_1(D_t) - \varepsilon| \leq \eta$ for every $t \in V_2$.

Now assume that part (i) does not hold true; that is, the set D contains no combinatorial line of A^n . This in turn implies that $\bigcap_{t \in L} D_t = \emptyset$ for every combinatorial line L of V_2 ; in particular, $\varepsilon \leq 1 - \frac{1}{2(k+1)}$.

Next, set $\sigma := \frac{\varepsilon^{k+1}}{2(k+1)}$ and notice that $\eta \leq \sigma/4^k$. Thus, by Theorem 2, there exist a nonempty subset Γ

of $A, \beta \in A \setminus \Gamma$ and a stochastic process $\langle S_{I_{V_2}(t)} : t \in A^{m+1} \rangle$ consisting of subsets of $A^{n-\ell}$ such that the following are satisfied:

- (a) For every $t \in V_2$ we have $S_t = \bigcap_{\alpha \in \Gamma} E_t^\alpha$ where for every $\alpha \in \Gamma$ the stochastic process $\langle E_{I_{V_2}(t)}^\alpha : t \in A^{m+1} \rangle$ is (α, β) -insensitive.
- (b) For every $t \in V_2$ such that $I_{V_2}^{-1}(t)$ contains β we have

$$\mathbb{P}_1(S_t) \geq \frac{\varepsilon^k}{4(k+1)} \quad \text{and} \quad \mathbb{P}_1(D_t | S_t) \geq \varepsilon + \frac{\sigma}{4k}. \tag{4.8}$$

By setting $E_t^\alpha = A^{n-\ell}$ for every $t \in V_2$ and every $\alpha \in A \setminus (\Gamma \cup \{\beta\})$, we may assume that $\Gamma = A \setminus \{\beta\}$. Next, let V_3 denote the set of all $t \in V_2$ such that $I_{V_2}^{-1}(t)$ starts with β and notice that V_3 is an m -dimensional combinatorial subspace of V_2 . Also observe that property (a) above and (4.8) hold true for every $t \in V_3$.

With the process $\langle S_t : t \in V_3 \rangle$ at our disposal, the rest of the proof follows by a double counting argument and an application of the first moment method. Indeed, let \mathbb{P}_2 and \mathbb{P}_3 denote the uniform probability measures on V_3 and $V_3 \times A^{n-\ell}$, respectively. Set $S := \bigcup_{t \in V_3} \{t\} \times S_t \subseteq V_3 \times A^{n-\ell}$ and notice that, by (4.8),

$$\mathbb{P}_3(S) \geq \frac{\varepsilon^k}{4(k+1)} \quad \text{and} \quad \mathbb{P}_3(D | S) \geq \varepsilon + \frac{\sigma}{4k}. \tag{4.9}$$

For every $s \in A^{n-\ell}$ let $S^s = \{t \in V_3 : t \wedge s \in S\}$ and $D^s = \{t \in V_3 : t \wedge s \in D\}$ denote the sections of S and D at s respectively, and set

$$B := \left\{ s \in A^{n-\ell} : \mathbb{P}_2(S^s) \leq \frac{\varepsilon^k \sigma}{2(k+1)4^{k+1}} \right\} \quad \text{and} \quad C := \bigcup_{s \in B} S^s \times \{s\} \subseteq S.$$

Noticing that $\mathbb{P}_3(C) \leq (\varepsilon^k \sigma) / (2(k+1)4^{k+1})$, by (4.9), we obtain that

$$\mathbb{P}_3(C | S) \leq \frac{\sigma}{2 \cdot 4^k}. \tag{4.10}$$

We thus have

$$\begin{aligned} \mathbb{P}_3(D | S \setminus C) &= \frac{\mathbb{P}_3(D \cap (S \setminus C))}{\mathbb{P}_3(S \setminus C)} \geq \frac{\mathbb{P}_3(D \cap (S \setminus C))}{\mathbb{P}_3(S)} \\ &\geq \mathbb{P}_3(D | S) - \mathbb{P}_3(C | S) \stackrel{(4.9), (4.10)}{\geq} \varepsilon + \frac{\sigma}{2 \cdot 4^k}. \end{aligned}$$

Because

$$\mathbb{P}_3(D | S \setminus C) = \sum_{s \in A^{n-\ell} \setminus B} \mathbb{P}_2(D^s | S^s) \cdot \mathbb{P}_3(S^s \times \{s\} | S \setminus C)$$

and $\sum_{s \in A^{n-\ell} \setminus B} \mathbb{P}_3(S^s \times \{s\} | S \setminus C) = 1$, there exists $s \in A^{n-\ell} \setminus B$ such that $\mathbb{P}_2(D^s | S^s) \geq \varepsilon + \sigma / (2 \cdot 4^k)$. We set

$$V := V_3 \times \{s\}, \quad \mathcal{S} := S \cap V \quad \text{and} \quad \mathcal{E}^\alpha := \left(\bigcup_{t \in V_3} \{t\} \times E_t^\alpha \right) \cap V \quad \text{for every } \alpha \in A \setminus \{\beta\}.$$

It is easy to see that with these choices the second part of the lemma is satisfied. The proof is completed. □

4.1.3. Step 3: Partitioning the insensitive set into combinatorial subspaces

The following lemma, which is proved in [P2, Lemma 8.2], is the last step of the proof of Theorem 3.

Lemma 4. *Let $k \geq 2$ be an integer, and assume that for every $0 < \delta \leq 1$ the number $\text{DHJ}(k, \delta)$ has been defined.*

Then for every positive integer m and every $0 < \eta \leq 1$ there exists a positive integer $\text{Til}(k, m, \eta)$ – which depends on the numbers $\text{DHJ}(k, \delta)$ – satisfying the following property. Let A be a set with $|A| = k+1$, let $n \geq \text{Til}(k, m, \eta)$ be an integer and let $\beta \in A$. Also let V be an n -dimensional combinatorial subspace of $A^{<\mathbb{N}}$ and let $S \subseteq V$, which is of the form $S = \bigcap_{\alpha \in A \setminus \{\beta\}} \mathcal{E}^\alpha$ where \mathcal{E}^α is (α, β) -insensitive in V for every $\alpha \in A \setminus \{\beta\}$. Then there exists a (possibly empty) collection \mathcal{W} of pairwise disjoint m -dimensional combinatorial subspaces of V with $\bigcup \mathcal{W} \subseteq S$ and such that $|S \setminus \bigcup \mathcal{W}| \leq \eta|V|$.

Although the proof of Lemma 4 given in [P2] is quite natural, unfortunately it leads to a very bad dependence of the numbers $\text{Til}(k, m, \eta)$ on the numbers $\text{DHJ}(k, \delta)$; see, for example, [P2, Section 9] for a discussion on this issue.

4.1.4. Completion of the proof of Theorem 3

Given Lemmas 3 and 4, the proof of Theorem 3 follows easily by induction on k . (The base case ‘ $k = 2$ ’ is a consequence of the classical Sperner theorem [Sp].) See, for example, [DK, Chapter 8] or [P2] for detailed expositions.

4.2. Comments

As alluded to earlier, Lemma 3 is a step towards obtaining primitive recursive bounds for the numbers $\text{DHJ}(k, \delta)$. It is clear that what is missing at this point is a quantitatively not wasteful proof of Lemma 4 (or a related variant). Although this will certainly require new ideas, it is likely that this program will eventually lead to primitive recursive bounds for the numbers $\text{DHJ}(k, \delta)$ belonging to the class \mathcal{E}^7 of Grzegorzczuk’s hierarchy or slightly higher.

A disadvantage of this approach is that it relies on an analysis that is ‘local’ in nature because we assume stationarity. It would be much more desirable if we had a ‘global’ structure theorem. Formulating and proving a ‘global’ theorem with quantitative aspects comparable to that of Theorem 2 might lead to upper bounds for the numbers $\text{DHJ}(k, \delta)$ that are of tower-type; note that this would also improve the longstanding upper bounds for the coloring version of the Hales-Jewett theorem obtained by Shelah [Sh].

However, even tower-type upper bounds are rather unlikely to be anywhere close to optimal. Indeed, the best known lower bounds for the numbers $\text{DHJ}(k, \delta)$ are merely quasi-polynomial with respect to δ^{-1} (see [P1, Theorem 1.3]).

5. The type of a subset of a discrete hypercube

This is the first section of the second part of this article, which is devoted to the study of correlations of stochastic processes over arbitrary nonempty subsets of discrete hypercubes. As we have pointed out in the Introduction, the analysis of these correlations relies, in an essentially way, on the notion of the *type* of a nonempty subset of A^n . This Ramsey-theoretic invariant was introduced in [DKT2], though it can be traced¹⁰ to [FK1]. We point out that for technical reasons (that will become transparent in Sections 6, 7 and 8), we will work with nonempty tuples of distinct elements of hypercubes instead of nonempty finite sets. This is an equivalent framework, but it does have some impact on our exposition when compared with that of [DKT2]. With this machinery at our disposal, it is straightforward to extend the notions of stationarity, pseudorandomness, supercorrelation and subcorrelation introduced in Definitions 1 and 4, respectively; these extensions are presented in Subsection 5.4.

¹⁰More precisely, the results in [FK1] concern colorings of variable words; this is a similar, but not identical, setting.

5.1. The type of a nonempty tuple

Let A be a finite set with $|A| \geq 2$, and let n, p be positive integers with $p \leq |A|^n$. Let $\mathbf{t} = (t_1, \dots, t_p)$ be a nonempty tuple (a nonempty finite sequence) of distinct elements of A^n .

5.1.1

If $p = 1$, then we define the type $\tau(\mathbf{t})$ of \mathbf{t} to be the empty sequence.

5.1.2

If $p \geq 2$, then we define $\tau(\mathbf{t})$ as follows. Let $R = (r_{ij}) \in A^{n \times p}$ denote the $n \times p$ matrix whose (i, j) th entry r_{ij} is the i th coordinate of t_j . (More precisely, writing $t_j = (t_{1,j}, \dots, t_{n,j})$ for every $j \in [p]$, we have $r_{ij} = t_{i,j}$.) Next, let E denote the matrix that is obtained by first erasing all rows of R with constant entries and then shrinking all consecutive appearances of identical rows to single rows; note that E is nonempty because $p \geq 2$. Let m denote the numbers of rows of E , and let s_1, \dots, s_p denote its columns (in particular, we have that $s_j \in A^m$ for every $j \in [p]$). We define the type $\tau(\mathbf{t})$ of \mathbf{t} by the rule

$$\tau(\mathbf{t}) = (s_1, \dots, s_p) \tag{5.1}$$

and we call the positive integer m the *dimension* of $\tau(\mathbf{t})$. (Thus, $\tau(\mathbf{t})$ is a p -tuple of distinct elements of A^m .)

Example 2. Let $A = [4]$, $n = 5$, $p = 5$ and

$$\mathbf{t} = ((2, 1, 3, 2, 3), (3, 1, 4, 2, 4), (4, 1, 3, 2, 3), (1, 1, 4, 2, 4), (4, 1, 2, 2, 2)).$$

Then we have

$$R = \begin{bmatrix} 2 & 3 & 4 & 1 & 4 \\ 1 & 1 & 1 & 1 & 1 \\ 3 & 4 & 3 & 4 & 2 \\ 2 & 2 & 2 & 2 & 2 \\ 3 & 4 & 3 & 4 & 2 \end{bmatrix} \quad \text{and} \quad E = \begin{bmatrix} 2 & 3 & 4 & 1 & 4 \\ 3 & 4 & 3 & 4 & 2 \end{bmatrix}$$

and, consequently, $m = 2$ and $\tau(\mathbf{t}) = ((2, 3), (3, 4), (4, 3), (1, 4), (4, 2))$.

Example 3. Let A be a finite set with $|A| \geq 2$, let $\Gamma \subseteq A$ be nonempty, set $p := |\Gamma|$ and let $(\gamma_1, \dots, \gamma_p)$ be an enumeration of the set Γ . Also, let n be an arbitrary positive integer. Then for every variable word v over A of length n we have $\tau((v(\gamma_1), \dots, v(\gamma_p))) = (\gamma_1, \dots, \gamma_p)$.

We isolate, for future use, two basic properties of types that are both straightforward consequences of the definition. The first property shows that the type is an isomorphic invariant.

Fact 4. Let A be a finite set with $|A| \geq 2$, let n, p be positive integers with $2 \leq p \leq |A|^n$ and let (t_1, \dots, t_p) be a nonempty tuple of distinct elements of A^n . Then for every n -dimensional combinatorial space V of $A^{\leq \mathbb{N}}$ we have

$$\tau((t_1, \dots, t_p)) = \tau((I_V(t_1), \dots, I_V(t_p))),$$

where $I_V : A^n \rightarrow V$ denotes the canonical isomorphism associated with V .

The second property is the permutation invariance of types.

Fact 5. Let A, n and p be as in Fact 4. Let (t_1, \dots, t_p) be a nonempty tuple of distinct elements of A^n and write $\tau((t_1, \dots, t_p)) = (s_1, \dots, s_p)$. Then for every permutation $\pi \in S_p$ we have $\tau((t_{\pi(1)}, \dots, t_{\pi(p)})) = (s_{\pi(1)}, \dots, s_{\pi(p)})$.

5.2. The type of a nonempty finite set

Let A be a finite set with $|A| \geq 2$, and let n be a positive integer. Let $G \subseteq A^n$ be nonempty. Set $p := |G|$ and fix an enumeration (t_1, \dots, t_p) of G . If $p = 1$ (that is, if G is a singleton), then we define the type $\tau(G)$ of G to be the empty set. Otherwise, if $p \geq 2$, then write $\tau((t_1, \dots, t_p)) = (s_1, \dots, s_p)$ and define the type $\tau(G)$ of G by setting

$$\tau(G) = \{s_1, \dots, s_p\}. \tag{5.2}$$

Note that, by Fact 5, $\tau(G)$ is well defined and independent of the enumeration of G , and observe that $\tau(G)$ is a subset of A^m of cardinality $|G|$ where m denotes the dimension of $\tau((t_1, \dots, t_p))$. In a slight abuse of the previous terminology, we will call this positive integer m the *dimension* of $\tau(G)$. (Note that the dimension of $\tau(G)$ bounds its cardinality; specifically, we have $|\tau(G)| \leq |A|^m$.) We set

$$\text{Type}(A) := \{\tau(G) : G \text{ is a nonempty subset of } A^n \text{ for some integer } n \geq 1\} \tag{5.3}$$

and we call an element of $\text{Type}(A)$ a *type over A* . We also observe the following analogue of Fact 4. (As before, the proof is straightforward.)

Fact 6. *Let A, n and V be as in Fact 4. Then for every nonempty $G \subseteq A^n$ we have $\tau(G) = \tau(I_V(G))$.*

5.3. Types and the Ramsey property

The most important property of types is that they can be used to classify all partition regular families of subsets of discrete hypercubes. To motivate this classification, we start by observing that there is no analogue of Ramsey’s classical theorem for colorings of subsets of combinatorial spaces of a fixed cardinality. Indeed, let A be a finite set with $|A| \geq 2$, and let $d, \ell \in \mathbb{N}$ with $|A|^d \geq \ell \geq 2$. Also, let V be a combinatorial space of $A^{<\mathbb{N}}$ of dimension at least $d + 1$, and define a coloring c of the set $\{G \subseteq V : |G| = \ell\}$ as follows. Let $G \subseteq V$ with $|G| = \ell$, and set $c(G) = \tau(G)$ if the dimension of the type of G is at most d ; otherwise, set $c(G) = 0$. Regardless of how large the dimension of V is, using Fact 4 it is easy to see that for every $(d + 1)$ -dimensional combinatorial subspace W of V the set $\{G \subseteq W : |G| = \ell\}$ is not monochromatic.

However, colorings that depend on the type are the only obstacles to the Ramsey property. Specifically, we have the following theorem whose proof can be found in [DK, Theorem 5.5] and that relies on the Graham-Rothschild theorem [GR].

Theorem 4. *For every triple k, m, r of positive integers with $k \geq 2$ there exists a positive integer N with the following property. For every integer $n \geq N$, every set A with $|A| = k$ and every r -coloring of the powerset of A^n there exists an m -dimensional combinatorial subspace V of A^n such that every pair of nonempty subsets of V with the same type is monochromatic. The least positive integer N with this property is denoted by $\text{RamSp}(k, m, r)$.*

Moreover, the numbers $\text{RamSp}(k, m, r)$ are upper bounded by a primitive recursive function belonging to the class \mathcal{E}^6 of Grzegorzczuk’s hierarchy.

5.4. Stochastic processes and types: Stationarity, pseudorandomness, supercorrelation, subcorrelation

Our next goal is to extend Definitions 1 and 4. We begin by generalising the notion of stationarity.

Definition 6. Let A be a finite set with $|A| \geq 2$, let n be a positive integer, let $\eta > 0$ and let $\langle D_t : t \in A^n \rangle$ be a stochastic process in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We say that $\langle D_t : t \in A^n \rangle$ is η -stationary if for

every pair of nonempty sets $G_1, G_2 \subseteq A^n$ with $\tau(G_1) = \tau(G_2)$ we have

$$\left| \mathbb{P}\left(\bigcap_{t \in G_1} D_t\right) - \mathbb{P}\left(\bigcap_{t \in G_2} D_t\right) \right| \leq \eta. \tag{5.4}$$

(In particular, by Example 3, if a process $\langle D_t : t \in A^n \rangle$ is η -stationary, then it is also η -stationary with respect to combinatorial lines.)

The following fact, which extends Fact 1, is an immediate consequence of Theorem 4.

Fact 7. *Let $k \geq 2$ be an integer, and let A be a set with $|A| = k$. Also, let $0 < \eta \leq 1$, and let n, m be positive integers such that*

$$n \geq \text{RamSp}(k, m, \lceil 1/\eta \rceil^{2k-1}). \tag{5.5}$$

Then for every stochastic process $\langle D_t : t \in A^n \rangle$ in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ there exists an m -dimensional combinatorial subspace V of A^n such that the process $\langle D_{I_V(s)} : s \in A^m \rangle$ (that is, the restriction of $\langle D_t : t \in A^n \rangle$ to V) is η -stationary.

We also have the following analogue of Lemma 1 whose proof is identical to that of Lemma 1.

Lemma 5. *Let A, n, η and $\langle D_t : t \in A^n \rangle$ be as in Definition 6. Then the following are satisfied.*

- (i) *For every $t_1, t_2 \in A^n$ we have $|\mathbb{P}(D_{t_1}) - \mathbb{P}(D_{t_2})| \leq \eta$. Thus, for every $t \in A^n$ we have $|\mathbb{P}(D_t) - \varepsilon| \leq \eta$ where $\varepsilon := \max\{\max\{\mathbb{P}(D_t) : t \in A^n\}, \eta\} > 0$.*
- (ii) *Let $m \in [n]$, and let $\tau \in \text{Type}(A)$ be a type over A of dimension m and with $|\tau| \geq 2$. Then for every $Q \subseteq \tau$ and every pair V_1, V_2 of m -dimensional combinatorial subspaces of A^n we have*

$$\left| \mathbb{P}\left(\bigcap_{t \in I_{V_1}(Q)} D_t \cap \bigcap_{t \in I_{V_1}(\tau \setminus Q)} D_t^c\right) - \mathbb{P}\left(\bigcap_{t \in I_{V_2}(Q)} D_t \cap \bigcap_{t \in I_{V_2}(\tau \setminus Q)} D_t^c\right) \right| \leq 2^{|Q|} \eta.$$

We proceed by generalising the notions of pseudorandomness, supercorrelation and subcorrelation introduced in Definition 4.

Definition 7. Let A be a finite set with $|A| \geq 2$, let n be a positive integer, let $0 < \eta, \varepsilon \leq 1$ and let $\langle D_t : t \in A^n \rangle$ be an η -stationary process in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that $|\mathbb{P}(D_t) - \varepsilon| \leq \eta$ for every $t \in A^n$. Also, let $\tau \in \text{Type}(A)$ be a type over A of dimension at most n , and let $\theta \geq 0$.

- (1) (*Pseudorandomness*) We say that $\langle D_t : t \in A^n \rangle$ is (τ, θ) -pseudorandom if $|\mathbb{P}(\bigcap_{t \in G} D_t) - \varepsilon^{|G|}| \leq \theta$ for every $G \subseteq A^n$ with $\tau(G) = \tau$.
- (2) (*Supercorrelation*) We say that $\langle D_t : t \in A^n \rangle$ is (τ, θ) -supercorrelated if $\mathbb{P}(\bigcap_{t \in G} D_t) \geq \varepsilon^{|G|} + \theta$ for every $G \subseteq A^n$ with $\tau(G) = \tau$.
- (3) (*Subcorrelation*) We say that $\langle D_t : t \in A^n \rangle$ is (τ, θ) -subcorrelated if $\mathbb{P}(\bigcap_{t \in G} D_t) \leq \varepsilon^{|G|} - \theta$ for every $G \subseteq A^n$ with $\tau(G) = \tau$.

We close this section with the following analogue of Fact 3. (Its simple proof is left to the interested reader.)

Fact 8. *Let A, n, η, ε and $\langle D_t : t \in A^n \rangle$ be as in Definition 7. Also, let $\tau \in \text{Type}(A)$ be a type over A of dimension at most n , and let $\theta \geq \eta$. Then one of the following holds true:*

- (i) *The process $\langle D_t : t \in A^n \rangle$ is (τ, θ) -pseudorandom.*
- (ii) *The process $\langle D_t : t \in A^n \rangle$ is $(\tau, \theta - \eta)$ -supercorrelated.*
- (iii) *The process $\langle D_t : t \in A^n \rangle$ is $(\tau, \theta - \eta)$ -subcorrelated.*

6. The separation index

This section, like Section 5, also contains preparatory material that is needed for the analysis of arbitrary correlations of stationary stochastic processes. Our aim is to define another isomorphic invariant of nonempty subsets of discrete hypercubes – the *separation index* – that is coarser than the type and measures how ‘well-distributed’ a subset is. Specifically, we have the following definition.

Definition 8. Let A be a finite set with $|A| \geq 2$, and let n be a positive integer.

- (1) Let $\mathbf{t} = (t_1, \dots, t_p)$ be a nonempty tuple of distinct elements of A^n , and let ℓ be a positive integer. We say that \mathbf{t} is ℓ -separated if for every $j \in [p]$ with $j \geq 2$ there exists $I \subseteq [n]$ (depending, possibly, on j) with $|I| = \ell$ and satisfying the following property: For every $q \in \{1, \dots, j - 1\}$ there exists $i \in I$ such that $t_j(i) \neq t_q(i)$. (Namely, the i th coordinate $t_j(i)$ of t_j is different from the i th coordinate $t_q(i)$ of t_q .) We define the *separation index* $s(\mathbf{t})$ of \mathbf{t} to be the least positive integer ℓ such that \mathbf{t} is ℓ -separated.
- (2) Let $G \subseteq A^n$ be nonempty, and set $p := |G|$. We define the *separation index* $s(G)$ of G by the rule

$$s(G) := \min\{s(\mathbf{t}) : \mathbf{t} = (t_1, \dots, t_p) \text{ is an enumeration of } G\}, \tag{6.1}$$

and we say that G is ℓ -separated if $s(G) = \ell$.

Remark 7. Note that the separation index of a nonempty finite set may be strictly smaller than the separation index of one of its enumerations. For instance, let $G = \{(0, 0), (1, 0), (0, 1)\} \subseteq \{0, 1\}^2$ and $\mathbf{t} = ((1, 0), (0, 1), (0, 0))$. Then we have $s(\mathbf{t}) = 2$, but $s(G) = 1$ as witnessed by the tuple $\mathbf{s} = ((0, 0), (1, 0), (0, 1))$.

In the following fact we state two basic properties of the separation index that were mentioned above, namely, that it is preserved under canonical isomorphisms and that it is coarser than the type. The proof follows from the relevant definitions and is left to the reader.

Fact 9. Let A be a finite set with $|A| \geq 2$, let n be a positive integer and let V be an n -dimensional combinatorial space of $A^{<\mathbb{N}}$. If (t_1, \dots, t_p) is a nonempty tuple of distinct elements of A^n , then $s((t_1, \dots, t_p)) = s((I_V(t_1), \dots, I_V(t_p)))$. Respectively, if $G \subseteq A^n$ is nonempty, then $s(G) = s(I_V(G))$; consequently, if $H \subseteq A^l$ for some positive integer l with $\tau(H) = \tau(G)$, then $s(H) = s(G)$.

We proceed by determining the separation index of some concrete examples of sets that are important from a combinatorial perspective.

Example 4 (Combinatorial lines). Let A and n be as in Definition 8. Let $\Gamma \subseteq A$ be nonempty, and set $p := |\Gamma|$. Also, let v be a variable word over A of length n . Then, by Fact 9, for every enumeration $(\gamma_1, \dots, \gamma_p)$ of Γ we have $s((v(\gamma_1), \dots, v(\gamma_p))) = s((\gamma_1, \dots, \gamma_p)) = 1$. In particular, every combinatorial line L of A^n is 1-separated.

Example 5 (Shelah lines). As above, let A be a finite set with $|A| \geq 2$. For every $\alpha \in A$ and every positive integer m , let $\alpha^m = (\alpha, \dots, \alpha)$ denote the sequence of length m taking the constant value α ; also, let α^0 denote the empty sequence.

Now let n be a positive integer, let $\alpha, \beta \in A$ with $\alpha \neq \beta$ and define the *Shelah line*¹¹ with parameters α, β by rule

$$S = \{\alpha^{n-m} \beta^m : m \in \{0, \dots, n\}\} \subseteq A^n. \tag{6.2}$$

Clearly, we have $|S| = n + 1$, and it is easy to see that the set S is 1-separated.

Example 5 implies, in particular, that there exist 1-separated sets of arbitrarily large cardinality. More generally, we have the following lemma.

¹¹These sets play a crucial role in Shelah’s proof of the Hales-Jewett theorem.

Lemma 6 (Random tuples of small size are 1-separated). *Let k, n, p be positive integers with $k \geq 2$ and $2 \leq p \leq k^n$. Let A be a set with $|A| = k$, and let \mathbb{P} denote the uniform probability measure on $(A^n)^p$. (That is, $(A^n)^p$ is the Cartesian product of p many copies of A^n .) Then we have*

$$\mathbb{P}(\mathbf{t} \text{ is 1-separated}) \geq 1 - pe^{-n(\frac{k-1}{k})^p}. \tag{6.3}$$

In particular, if $p \leq \log(n)$, then $\mathbb{P}(\mathbf{t} \text{ is 1-separated}) = 1 - o_{n \rightarrow \infty; k}(1)$.

Proof. Set $S := \{\mathbf{t} \in (A^n)^p : \mathbf{t} \text{ is 1-separated}\}$, and let S^c denote the complement of S . Note that for every $i \in [n]$ and every $j \in [p]$ with $j \geq 2$ the set of all $\mathbf{t} = (t_1, \dots, t_p) \in (A^n)^p$ such that $t_j(i) \notin \{t_1(i), \dots, t_{j-1}(i)\}$ has probability $\frac{k(k-1)^{j-1}}{k^j} \geq (\frac{k-1}{k})^p$. (Here, $t_q(i)$ denotes the i th coordinate of t_q for every $q \in [j]$.) Therefore, for every $j \in [p]$ with $j \geq 2$ the set of all $\mathbf{t} = (t_1, \dots, t_p) \in (A^n)^p$ such that t_j fails to satisfy the condition of being 1-separated has probability at most $(1 - (\frac{k-1}{k})^p)^n$. Using the fact that $(1 - \frac{r}{n})^n \leq e^{-r}$ for every $r > 0$ and every positive integer n , we thus have

$$\mathbb{P}(S^c) \leq p \left(1 - \left(\frac{k-1}{k}\right)^p\right)^n \leq pe^{-n(\frac{k-1}{k})^p}, \tag{6.4}$$

which is equivalent to (6.3).

Next assume that $p \leq \log(n)$. Because the function $f(x) = xe^{-nr^x}$ is increasing for every $r \in (0, 1)$ and every positive integer n , by (6.4) we obtain that

$$\mathbb{P}(S^c) \leq \log(n) e^{-n(\frac{k-1}{k})^{\log(n)}} = \log(n) e^{-n^{1+\log(\frac{k-1}{k})}}.$$

Therefore, $\mathbb{P}(S) = 1 - o_{n \rightarrow \infty; k}(1)$ as desired. □

The last example in this section provides us with a representative example of an n -separated set.

Example 6 (Combinatorial subspaces). Let A and n be as in Definition 8 and notice that for every nonempty $G \subseteq A^n$ we have $s(G) \leq n$. On the other hand, it is easy to verify that $s(A^n) = n$. Using this observation and Fact 9, we see that every n -dimensional combinatorial space of $A^{<\mathbb{N}}$ is n -separated.

7. Correlations over 1-separated sets

7.1. The main result

We begin by introducing the analogue of insensitivity for processes indexed by combinatorial spaces.

Definition 9. Let A, n, α and β be as in Definition 2, let V be an n -dimensional combinatorial space of $A^{<\mathbb{N}}$ and let $I_V : A^n \rightarrow V$ denote the canonical isomorphism associated with V . We say that a stochastic process $\langle D_t : t \in V \rangle$ in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is (α, β) -insensitive in V if $\langle D_{I_V(s)} : s \in A^n \rangle$ is (α, β) -insensitive in the sense of Definition 2. (That is, if $D_{I_V(s)} = D_{I_V(t)}$ for every $s, t \in A^n$ that are (α, β) -equivalent.)

The main result of this section is the following extension of Theorem 2, which concerns correlations of stationary processes over 1-separated sets. (We recall that the notion of stationarity in this more general context is given in Definition 6.)

Theorem 5. *Let k, κ, m be positive integers with $k, \kappa \geq 2$ and $\kappa \leq k^m$, and let $\varepsilon, \sigma, \eta > 0$ such that*

$$\varepsilon \leq 1 - \frac{1}{2\kappa}, \quad \sigma \leq \frac{\varepsilon^{\kappa-1}}{2\kappa} \quad \text{and} \quad \eta \leq \frac{\sigma}{4^{\kappa-1}}. \tag{7.1}$$

Also, let A be a set with $|A| = k$, let $n > m$ be an integer and let $\langle D_t : t \in A^n \rangle$ be an η -stationary process in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that $|\mathbb{P}(D_t) - \varepsilon| \leq \eta$ for every $t \in A^n$. Then, either

(i) for every nonempty 1-separated $G \subseteq A^n$ with cardinality at most κ and whose type $\tau(G)$ has dimension at most m we have

$$\left| \mathbb{P}\left(\bigcap_{t \in G} D_t\right) - \varepsilon^{|G|} \right| \leq \sigma, \tag{7.2}$$

(ii) or $\langle D_t : t \in A^n \rangle$ correlates with a ‘structured’ stochastic process when restricted to a large subspace; precisely, there exist a combinatorial subspace V of A^n with $\dim(V) \geq n - m$, a nonempty subset Γ of A , $\beta \in A \setminus \Gamma$ and a stochastic process $\langle S_t : t \in V \rangle$ in $(\Omega, \mathcal{F}, \mathbb{P})$ with the following properties:

- (a) For every $t \in V$ we have $S_t = \bigcap_{\alpha \in \Gamma} E_t^\alpha$ where for every $\alpha \in \Gamma$ the process $\langle E_t^\alpha : t \in V \rangle$ is (α, β) -insensitive in V .
- (b) For every $t \in V$ we have

$$\mathbb{P}(S_t) \geq \frac{\varepsilon^{\kappa-1}}{4\kappa} \quad \text{and} \quad \mathbb{P}(D_t | S_t) \geq \varepsilon + \frac{\sigma}{4\kappa-1}. \tag{7.3}$$

Theorem 5 shows that stationary processes that exhibit nonindependent behaviour over 1-separated sets are essentially characterised – in the strong quantitative sense described in (7.3) – by their correlation with insensitive processes. Note, however, that in contrast to Theorem 2, this correlation is ‘local’ in nature; that is, we need to pass to a subspace in order to verify it. We present an example in Subsection 7.2 that elucidates the necessity of this restriction.

The proof of Theorem 5 is given in Subsection 7.5. It relies on the following analogue of Proposition 2, whose proof is given in Subsection 7.4. (The concepts of pseudorandomness, supercorrelation and subcorrelation that appear below are introduced in Definition 7.)

Proposition 3. *Let A be a finite set with $|A| \geq 2$, let n, p be positive integers with $p + 1 \leq |A|^n$, let $0 < \eta, \varepsilon \leq 1$ and let $\langle D_t : t \in A^n \rangle$ be an η -stationary process in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that $|\mathbb{P}(D_t) - \varepsilon| \leq \eta$ for every $t \in A^n$. Let $\mathbf{t} = (t_1, \dots, t_{p+1})$ be an 1-separated tuple consisting of distinct elements of A^n , set $G := \{t_1, \dots, t_{p+1}\}$ and $H := \{t_1, \dots, t_p\}$ and let d denote the dimension of $\tau(G)$. Finally, let $0 < \theta, \sigma \leq 1$, and assume that the process $\langle D_t : t \in A^n \rangle$ is $(\tau(H), \theta)$ -pseudorandom. Then there exist an $(n - d)$ -dimensional combinatorial subspace V of A^n , a nonempty subset Γ of A , $\beta \in A \setminus \Gamma$ and a process $\langle S_t : t \in V \rangle$ in $(\Omega, \mathcal{F}, \mathbb{P})$ with the following properties:*

- (i) For every $t \in V$ we have $S_t = \bigcap_{\alpha \in \Gamma} E_t^\alpha$ where for every $\alpha \in \Gamma$ the process $\langle E_t^\alpha : t \in V \rangle$ is (α, β) -insensitive in V .
- (ii) For every $t \in V$ we have $|\mathbb{P}(S_t) - \varepsilon^p| \leq \theta$.
- (iii) If $\langle D_t : t \in A^n \rangle$ is $(\tau(G), \sigma)$ -supercorrelated, then for every $t \in V$ we have

$$\mathbb{P}(D_t | S_t) \geq \varepsilon \left(1 + \frac{\sigma \varepsilon^{-1} - \theta}{\varepsilon^p + \theta} \right). \tag{7.4}$$

(iv) If $\langle D_t : t \in A^n \rangle$ is $(\tau(G), \sigma)$ -subcorrelated, then for every $t \in V$ we have

$$\mathbb{P}(D_t | S_t) \leq \varepsilon \left(1 - \frac{\sigma \varepsilon^{-1} - \theta}{\varepsilon^p - \theta} \right). \tag{7.5}$$

7.2. Correlations over 1-separated sets: Example

We are about to present an example of a process that exhibits nonindependent behaviour when we look at its correlations over 1-separated sets whose type is rather simple but not quite similar to that of combinatorial lines. As in Example 1, for concreteness we will work with the set $A = \{1, 2, 3\}$ and the 1-separated type

$$\tau = \{(1, 2, 1), (2, 1, 2), (2, 2, 3)\} \in \text{Type}(\{1, 2, 3\}). \tag{7.6}$$

Let $\varepsilon > 0$, let $n \geq 5$ be an integer and let

$$\langle E_{y^{\frown} s} : y \in \{1, 2, 3\}^3 \setminus \{(2, 2, 3)\} \text{ and } s \in \{1, 2\}^{n-2} \rangle$$

be a family of independent events in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with equal probability $\sqrt{\varepsilon}$. We define $\langle D_t : t \in \{1, 2, 3\}^n \rangle$ by setting for every $z \in \{1, 2, 3\}^{n-3}$

- (a) $D_{(2,2,3)^{\frown} z} := E_{(1,2,1)^{\frown} z^3 \rightarrow 1} \cap E_{(2,1,2)^{\frown} z^3 \rightarrow 2}$ and
- (b) $D_{y^{\frown} z} := E_{y^{\frown} z^3 \rightarrow 1} \cap E_{y^{\frown} z^3 \rightarrow 2}$ if $y \in \{1, 2, 3\}^3 \setminus \{(2, 2, 3)\}$.

Note the difference between the definition in (a) and the definition in Example 1: Given $t \in A^n$, first we change a short initial segment of t and then we ‘project’ the rest of the sequence. This manoeuvre will be generalised in the next subsection.

The analysis of the correlations of $\langle D_t : t \in \{1, 2, 3\}^n \rangle$ is fairly straightforward. Specifically, notice that for every $t \in \{1, 2, 3\}^n$ we have $\mathbb{P}(D_t) = \varepsilon$. Moreover, for every $z \in \{1, 2, 3\}^{n-3}$ set

$$G_z := \{(1, 2, 1)^{\frown} (z^3 \rightarrow 1), (2, 1, 2)^{\frown} (z^3 \rightarrow 2), (2, 2, 3)^{\frown} z\} \subseteq \{1, 2, 3\}^n$$

and observe that $\tau(G_z) = \tau$ and $\mathbb{P}(\bigcap_{t \in G_z} D_t) = \varepsilon^2$, which deviates, of course, from the expected value ε^3 .

Finally, note that $\langle D_t : t \in \{1, 2, 3\}^n \rangle$ cannot be written as the intersection of insensitive processes but only barely so. Indeed, set

$$V := \{(2, 2, 3)^{\frown} z : z \in \{1, 2, 3\}^{n-3}\} \tag{7.7}$$

and observe that V is an $(n-3)$ -dimensional combinatorial subspace of $\{1, 2, 3\}^n$. Clearly, by (a) above, the restriction of $\langle D_t : t \in \{1, 2, 3\}^n \rangle$ to V is the intersection of two processes that are $(1, 3)$ - and $(2, 3)$ -insensitive in V , respectively.

7.3. Definitions/Notation

Let A, n and p be as in Proposition 3. Let

$$\mathbf{t} = (t_1, \dots, t_{p+1}) \tag{7.8}$$

be an 1-separated tuple consisting of distinct elements of A^n , let $\tau = \tau(\mathbf{t})$ denote its type and let d denote the dimension of τ . We will define

- an $(n-d)$ -dimensional combinatorial subspace V of A^n ,
- a nonempty subset Γ of A ,
- $\beta \in A \setminus \Gamma$,
- an integer $\iota \in [d]$ and
- for every $j \in [p]$ a map $T_j : V \rightarrow A^n$.

These data will be used in the proofs of Theorem 5 and Proposition 3; in fact, they constitute the combinatorial heart of the argument. We also note that V, Γ, β, ι and $\langle T_j : j \in [p] \rangle$ will essentially depend on the type τ of \mathbf{t} and not on the tuple \mathbf{t} itself; however, it is technically easier to work with \mathbf{t} .

7.3.1. Defining ι, β and Γ and splitting the type τ

We write $\tau = (s_1, \dots, s_{p+1})$ where $s_j = (s_j(1), \dots, s_j(d)) \in A^d$ for every $j \in [p+1]$. By Fact 9 and our assumption that \mathbf{t} is 1-separated, we see that τ is also 1-separated. Taking into account this remark, we define

$$\iota := \min \{i \in [d] : s_{p+1}(i) \neq s_j(i) \text{ for every } j \in [p]\} \tag{7.9}$$

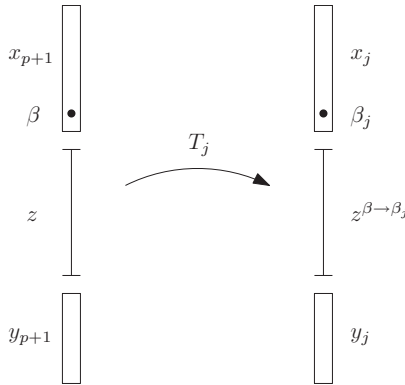


Figure 1. The map T_j acting on V .

and

$$\beta := s_{p+1}(t), \beta_j := s_j(t) \text{ for every } j \in [p], \text{ and } \Gamma := \{\beta_1, \dots, \beta_p\}. \tag{7.10}$$

(In particular, we have $\beta \notin \Gamma$; also note that $|\Gamma| \leq p$ because the elements β_1, \dots, β_p are not necessarily distinct.) Moreover, for every $j \in [p + 1]$ set

$$x_j = (s_j(1), \dots, s_j(t)) \text{ and } y_j = (s_j(t + 1), \dots, s_j(n)) \tag{7.11}$$

with the convention that y_j is the empty sequence if $t = d$; note that $s_j = x_j \hat{\ } y_j$.

7.3.2. Defining V and the maps $\langle T_j : j \in [p] \rangle$

Next, set

$$V := \{x_{p+1} \hat{\ } z \hat{\ } y_{p+1} : z \in A^{n-d}\} \tag{7.12}$$

and observe that V is an $(n - d)$ -dimensional combinatorial subspace¹² of A^n . Finally, for every $j \in [p]$ we define $T_j : V \rightarrow A^n$ by the rule

$$T_j(x_{p+1} \hat{\ } z \hat{\ } y_{p+1}) = x_j \hat{\ } (z^{\beta \rightarrow \beta_j}) \hat{\ } y_j. \tag{7.13}$$

7.3.3. Basic properties

We close this subsection by observing the following two elementary (though important) properties of the previous constructions.

Fact 10. Let $t, V, \beta, \beta_1, \dots, \beta_p$ and $\langle T_j : j \in [p] \rangle$ be as above.

(i) For every $t \in V$ and every $1 \leq i_1 < \dots < i_q \leq p$ we have

$$\tau((T_{i_1}(t), \dots, T_{i_q}(t), t)) = \tau((t_{i_1}, \dots, t_{i_q}, t_{p+1})) \tag{7.14}$$

and

$$\tau((T_{i_1}(t), \dots, T_{i_q}(t))) = \tau((t_{i_1}, \dots, t_{i_q})). \tag{7.15}$$

(ii) Let $\langle D_t : t \in A^n \rangle$ be a stochastic process in a probability space (Ω, Σ, μ) . Then for every $j \in [p]$ the process $\langle D_{T_j(t)} : t \in V \rangle$ is (β_j, β) -insensitive in V .

¹²Notice that the subspace V is of very special form; in particular, the canonical isomorphism associated with V is the map $A^{n-d} \ni z \mapsto x_{p+1} \hat{\ } z \hat{\ } y_{p+1} \in V$.

7.4. Proof of Proposition 3

Let $V, \beta, \beta_1, \dots, \beta_p, \Gamma$ and $\langle T_j : j \in [p] \rangle$ be the data obtained in Subsection 7.3 for the 1-separated tuple $\mathbf{t} = (t_1, \dots, t_{p+1})$.

For every $t \in V$ define $S_t = \bigcap_{j=1}^p D_{T_j(t)}$ and notice that, by part (ii) of Fact 10, the process $\langle S_t : t \in V \rangle$ satisfies part (i) of the theorem. On the other hand, because the process $\langle D_t : t \in A^n \rangle$ is $(\tau(H), \theta)$ -pseudorandom, by (7.15), for every $t \in V$ we have

$$|\mathbb{P}(S_t) - \varepsilon^p| \leq \theta; \tag{7.16}$$

that is, $\langle S_t : t \in V \rangle$ satisfies part (ii) of the theorem.

For part (iii) assume that $\langle D_t : t \in A^n \rangle$ is $(\tau(G), \sigma)$ -super-correlated, and let $t \in V$ be arbitrary. By (7.14) and the super-correlation assumption, we have that $\mathbb{P}(D_t \cap S_t) \geq \varepsilon^p + \sigma$. On the other hand, by (7.16), we see that $\mathbb{P}(S_t) \leq \varepsilon^p + \theta$. Therefore, $\mathbb{P}(D_t | S_t) \geq \varepsilon(1 + \frac{\sigma\varepsilon^{-1}-\theta}{\varepsilon^p+\theta})$ as desired.

Finally, for part (iv) assume that $\langle D_t : t \in A^n \rangle$ is $(\tau(G), \sigma)$ -sub-correlated. Fix $t \in V$. Using again (7.14), the sub-correlation assumption and (7.16), we obtain that $\mathbb{P}(D_t \cap S_t) \leq \varepsilon^p - \sigma$ and $\mathbb{P}(S_t) \geq \varepsilon^p - \theta$, which implies that $\mathbb{P}(D_t | S_t) \leq \varepsilon(1 - \frac{\sigma\varepsilon^{-1}-\theta}{\varepsilon^p-\theta})$. The proof is completed.

7.5. Proof of Theorem 5

It is similar to the proof of Theorem 2 with the main new ingredients being Proposition 3 and the material in Subsection 7.3. We shall describe in detail the necessary changes, because this proof will also serve as a model for the proof of Theorem 6 in Section 8.

Let $(\theta_p)_{p=0}^\kappa$ be the finite sequence defined in (3.6) – that is, $\theta_0 = 0, \theta_1 = \eta$, and $\theta_p = 4^{p-\kappa}\sigma$ if $p \in \{2, \dots, \kappa\}$ – and recall that $(\theta_p)_{p=0}^\kappa$ is increasing. Assume that part (i) of the theorem does not hold true and fix an 1-separated set $G \subseteq A^n$ of cardinality at most κ whose type $\tau(G)$ has dimension at most m and such that (a) the process $\langle D_t : t \in A^n \rangle$ is not $(\tau(G), \theta_{|G|})$ -pseudorandom and (b) G has the minimal cardinal among all sets with these properties. (Note that $|G| \geq 2$.) Let $\mathbf{t} = (t_1, \dots, t_{|G|})$ be an enumeration of G such that the tuple \mathbf{t} is 1-separated, let d denote the dimension of $\tau(G)$ and set $H := \{t_1, \dots, t_{|G|-1}\}$ and $p := |H|$; notice that $1 \leq p \leq \kappa - 1$ and $1 \leq d \leq m$. Also, observe that for every nonempty proper subset Σ of G the process $\langle D_t : t \in A^n \rangle$ is $(\tau(\Sigma), \theta_{|\Sigma|})$ -pseudorandom.

We set $\theta := \theta_p$ and $\Theta := \theta_{p+1}$. Because $\langle D_t : t \in A^n \rangle$ is not $(\tau(G), \Theta)$ -pseudorandom, by Fact 8 we see that either

- (A1) the process $\langle D_t : t \in A^n \rangle$ is $(\tau(G), \Theta - \eta)$ -super-correlated or
- (A2) the process $\langle D_t : t \in A^n \rangle$ is $(\tau(G), \Theta - \eta)$ -sub-correlated.

If the first case holds true, then, arguing precisely as in the proof of Theorem 2 and using Proposition 3 instead of Proposition 2, it is easy to verify that part (ii) of the theorem is satisfied.

So assume that the process $\langle D_t : t \in A^n \rangle$ is $(\tau(G), \Theta - \eta)$ -sub-correlated, and let V and $\langle T_j : j \in [p] \rangle$ be the combinatorial space and the maps obtained in Subsection 7.3 for the 1-separated tuple \mathbf{t} . For every $t \in V$ we set

$$S_t := \left(\bigcap_{j=1}^{p-1} D_{T_j(t)} \right) \cap D_{T_p(t)}^c = \left(\bigcap_{j=1}^{p-1} D_{T_j(t)} \right) \setminus \left(\bigcap_{j=1}^p D_{T_j(t)} \right). \tag{7.17}$$

(Recall that, by convention, $\bigcap_{j=1}^{p-1} D_{T_j(t)} = \Omega$ if $p = 1$.) Notice that, by part (ii) of Fact 10, the process $\langle S_t : t \in V \rangle$ satisfies part (ii.a) of the theorem. Next, we set $F := \{t_1, \dots, t_{p-1}\}$ (observe that F may be empty). Because $\langle D_t : t \in A^n \rangle$ is $(\tau(F), \theta_{p-1})$ -pseudorandom if $p > 1$ (if $p = 1$, then this is superfluous) and $(\tau(H), \theta)$ -pseudorandom, by (7.15) for every $t \in V$ we have

$$|\mathbb{P}(S_t) - \varepsilon^{p-1}(1 - \varepsilon)| \leq \theta + \theta_{p-1}. \tag{7.18}$$

Using (7.14) and the fact that the process $\langle D_t : t \in A^n \rangle$ is $(\tau(G), \Theta - \eta)$ -subcorrelated and $(\tau(F \cup \{t_{p+1}\}), \theta)$ -pseudorandom, for every $t \in V$ we have

$$\mathbb{P}(D_t \cap S_t) \geq (\varepsilon^p - \theta) - (\varepsilon^{p+1} - \Theta + \eta). \tag{7.19}$$

Moreover, by (7.1) and the definition of $(\theta_j)_{j=1}^k$, we see that $\theta + \theta_{p-1} \leq \varepsilon^p$, $\theta \leq \Theta/4$ and $\theta_{p-1} + \eta \leq \Theta/4$. Therefore, by (7.18) and (7.19), for every $t \in V$

$$\begin{aligned} \mathbb{P}(D_t | S_t) &\geq \frac{\varepsilon^p - \varepsilon^{p+1} + \Theta - \theta - \eta}{\varepsilon^{p-1} - \varepsilon^p + \theta + \theta_{p-1}} \geq \varepsilon + \frac{\Theta - 2\theta - \theta_{p-1} - \eta}{\varepsilon^{p-1} - \varepsilon^p + \theta + \theta_{p-1}} \\ &\geq \varepsilon + \frac{\Theta}{4} \geq \varepsilon + \frac{\sigma}{4^{\kappa-1}}. \end{aligned}$$

Finally, by (7.1) and (7.18), we conclude that

$$\mathbb{P}(S_t) \geq \varepsilon^{p-1} - \varepsilon^p - \theta - \theta_{p-1} \geq \varepsilon^{\kappa-1}(1 - \varepsilon) - 2\theta \geq \frac{\varepsilon^{\kappa-1}}{2\kappa} - 2\theta \geq \frac{\varepsilon^{\kappa-1}}{4\kappa}$$

for every $t \in V$. The proof is completed.

8. Correlations over ℓ -separated sets

8.1. Obstructions to independence: Simplicial processes

We are about to begin our analysis of arbitrary correlations of stationary processes. As we have noted, the main – and perhaps the most interesting – difference lies in the fact that insensitive processes are not enough to characterise nonindependent behaviour. Our goal in this subsection is to discuss this phenomenon and introduce the ‘structured’ processes that appear in this more general context.

To this end, we need to define a ‘local’ version of insensitivity. To motivate this ‘local’ version, let A be a finite set with $|A| \geq 2$, let $n, \ell, r_1, \dots, r_\ell$ be positive integers with $n, \ell \geq 2$ and $n = r_1 + \dots + r_\ell$ and note that we may identify the hypercube A^n with the product $A^{r_1} \times \dots \times A^{r_\ell}$ via the map

$$A^{r_1} \times \dots \times A^{r_\ell} \ni (t_1, \dots, t_\ell) \mapsto t_1 \frown \dots \frown t_\ell \in A^n.$$

Having in mind this identification, we may consider subsets of A^n that are insensitive only in one of the factors $A^{r_1}, \dots, A^{r_\ell}$. This is, essentially, the content of the following definition.

Definition 10 (Local insensitivity). Let A be a finite set with $|A| \geq 2$, let n be a positive integer, let $\alpha, \beta \in A$ with $\alpha \neq \beta$ and let $I \subseteq [n]$ be nonempty. Also, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space.

- (1) Let $t, s \in A^n$ and write $t = (t_1, \dots, t_n)$ and $s = (s_1, \dots, s_n)$. We say that t, s are (α, β, I) -equivalent if for every $i \in [n] \setminus I$ we have $t_i = s_i$ and, moreover, for every $i \in I$ and every $\gamma \in A \setminus \{\alpha, \beta\}$ we have $t_i = \gamma$ if and only if $s_i = \gamma$.
- (2) We say that a process $\langle E_t : t \in A^n \rangle$ in $(\Omega, \mathcal{F}, \mathbb{P})$ is (α, β, I) -insensitive if $E_t = E_s$ for every $t, s \in A^n$ that are (α, β, I) -equivalent.
- (3) Let V be an n -dimensional combinatorial space of $A^{\leq n}$. We say that a process $\langle E_t : t \in V \rangle$ in $(\Omega, \mathcal{F}, \mathbb{P})$ is (α, β, I) -insensitive in V provided that $\langle E_{I_V(t)} : t \in A^n \rangle$ is (α, β, I) -insensitive where $I_V : A^n \rightarrow V$ denotes the canonical isomorphism associated with V .

We proceed with the following example, which shows the need to extend the notion of a ‘structured’ process.

Example 7. As in Example 1, we will work with the set $A = \{1, 2, 3\}$, and we will focus on correlations over 2-dimensional combinatorial spaces of $\{1, 2, 3\}^{\leq n}$. Notice that, by Fact 6 and Example 6, all

2-dimensional combinatorial spaces are 2-separated and are of type

$$\tau = \{(1, 1), (1, 2), (1, 3), (2, 1), (2, 2), (2, 3), (3, 1), (3, 2), (3, 3)\} \in \text{Type}(\{1, 2, 3\}).$$

Now let n be an arbitrary positive integer, and fix a family

$$\langle E_{t \frown s} : t \frown s \in (\{1, 2\}^n \times \{1, 2, 3\}^n) \cup (\{1, 2, 3\}^n \times \{1, 2\}^n) \rangle$$

of independent events in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with equal probability $\varepsilon > 0$. We define a process $\langle D_z : z \in \{1, 2, 3\}^{2n} \rangle$ by setting

$$D_{t \frown s} := S_{t \frown s}^1 \cap S_{t \frown s}^2, \tag{8.1}$$

where $S_{t \frown s}^1 := E_{t^3 \rightarrow 1 \frown s^3 \rightarrow 1} \cap E_{t^3 \rightarrow 1 \frown s^3 \rightarrow 2} \cap E_{t^3 \rightarrow 1 \frown s} \cap E_{t^3 \rightarrow 2 \frown s^3 \rightarrow 1} \cap E_{t^3 \rightarrow 2 \frown s^3 \rightarrow 2} \cap E_{t^3 \rightarrow 2 \frown s}$ and $S_{t \frown s}^2 := E_{t \frown s^3 \rightarrow 1} \cap E_{t \frown s^3 \rightarrow 2}$ for every $t, s \in \{1, 2, 3\}^n$. Note that

$$\mathbb{P}(D_{t \frown s}) = \begin{cases} \varepsilon^8 & \text{if both } t, s \text{ contain } 3, \\ \varepsilon^3 & \text{if exactly one of } t, s \text{ contains } 3, \\ \varepsilon & \text{if both } t, s \text{ do not contain } 3. \end{cases} \tag{8.2}$$

Next, for every $t, s \in \{1, 2, 3\}^n$ that contain 3, set

$$G_{t,s} := \{t^3 \rightarrow 1 \frown s^3 \rightarrow 1, t^3 \rightarrow 1 \frown s^3 \rightarrow 2, t^3 \rightarrow 1 \frown s, t^3 \rightarrow 2 \frown s^3 \rightarrow 1, t^3 \rightarrow 2 \frown s^3 \rightarrow 2, t^3 \rightarrow 2 \frown s, t \frown s^3 \rightarrow 1, t \frown s^3 \rightarrow 2, t \frown s\}$$

and observe that $G_{t,s}$ is a 2-dimensional combinatorial subspace of $\{1, 2, 3\}^{2n}$; also notice that $D_{t \frown s} = \bigcap_{z \in G_{t,s}} D_z$ and, therefore, $\mathbb{P}(\bigcap_{z \in G_{t,s}} D_z) = \varepsilon^8$, which deviates from the expected value ε^{24} . In other words, the process $\langle D_z : z \in \{1, 2, 3\}^{2n} \rangle$ exhibits nonindependent behaviour when we look at its correlations over 2-dimensional combinatorial subspaces.

Note, however, that $\langle D_z : z \in \{1, 2, 3\}^{2n} \rangle$ cannot be written as the intersection of insensitive processes even if we restrict it to subspaces of very small dimension. (This is a consequence of the fact that the processes $\langle S_{t \frown s}^1 : t, s \in \{1, 2, 3\}^n \rangle$ and $\langle S_{t \frown s}^2 : t, s \in \{1, 2, 3\}^n \rangle$ depend nontrivially on the parameters s and t , respectively.) Nevertheless, the process $\langle D_z : z \in \{1, 2, 3\}^{2n} \rangle$ is not random at all: It is obtained from $\langle S_{t \frown s}^1 : t, s \in \{1, 2, 3\}^n \rangle$ and $\langle S_{t \frown s}^2 : t, s \in \{1, 2, 3\}^n \rangle$, which are both the intersection of locally insensitive processes but for disjoint domains of insensitivity. This less restrictive form of structurability is abstracted in the following definition.

Definition 11 (Simplicial processes). Let A be a finite set with $|A| \geq 2$, let n, ℓ be positive integers with $n \geq \ell$, let $\mathbf{r} = (r_1, \dots, r_\ell)$ be an ℓ -tuple of positive integers such that $n = r_1 + \dots + r_\ell$ and let $I_1^{\mathbf{r}}, \dots, I_\ell^{\mathbf{r}}$ denote the unique successive intervals of $[n]$ such that $|I_l^{\mathbf{r}}| = r_l$ for every $l \in [\ell]$. We say that a process $\langle S_t : t \in A^n \rangle$ in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is (ℓ, \mathbf{r}) -simplicial if there exist

- $\beta_1, \dots, \beta_\ell \in A$,
- for every $l \in [\ell]$ a nonempty subset Γ_l of $A \setminus \{\beta_l\}$ and
- for every $l \in [\ell]$ and $\alpha \in \Gamma_l$ an $(\alpha, \beta_l, I_l^{\mathbf{r}})$ -insensitive process $\langle E_t^{l,\alpha} : t \in A^n \rangle$,

such that for every $t \in A^n$ we have

$$S_t = \bigcap_{l=1}^{\ell} \bigcap_{\alpha \in \Gamma_l} E_t^{l,\alpha}. \tag{8.3}$$

More generally, let V be an n -dimensional combinatorial space of $A^{<\mathbb{N}}$, and let $\langle S_t : t \in V \rangle$ be a process in $(\Omega, \mathcal{F}, \mathbb{P})$. We say that $\langle S_t : t \in V \rangle$ is (ℓ, \mathbf{r}) -simplicial in V if the process $\langle S_{I_V(t)} : t \in A^n \rangle$ is (ℓ, \mathbf{r}) -simplicial.

Remark 8. In order to see the relevance of simplicial processes in this context, note that if A, n, ℓ and \mathbf{r} are as in Definition 11 and $\langle S_t : t \in A^n \rangle$ is an arbitrary (ℓ, \mathbf{r}) -simplicial process, then there exist nonempty $G \subseteq A^n$ and $x \in A^n \setminus G$ such that the set $G \cup \{x\}$ is ℓ -separated and, moreover,

$$\bigcap_{t \in G \cup \{x\}} S_t = \bigcap_{t \in G} S_t.$$

In particular, the events $\langle S_t : t \in A^n \rangle$ cannot be independent.

8.2. The main result

The following theorem – which is the main result in this section – complements Theorems 2 and 5 and completes the analysis of correlations of stationary processes. (We recall that the notion of stationarity for arbitrary correlations is given in Definition 6.)

Theorem 6. *Let k, κ, m be positive integers with $k, \kappa \geq 2$ and $\kappa \leq k^m$, and let $\varepsilon, \sigma, \eta > 0$ such that*

$$\varepsilon \leq 1 - \frac{1}{2\kappa}, \quad \sigma \leq \frac{\varepsilon^{\kappa-1}}{2\kappa} \quad \text{and} \quad \eta \leq \frac{\sigma}{4^{\kappa-1}}. \tag{8.4}$$

Also, let A be a set with $|A| = k$, let $n > m$ be an integer and let $\langle D_t : t \in A^n \rangle$ be an η -stationary process in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that $|\mathbb{P}(D_t) - \varepsilon| \leq \eta$ for every $t \in A^n$. Then, either

- (i) for every nonempty $G \subseteq A^n$ with cardinality at most κ and whose type $\tau(G)$ has dimension¹³ at most m we have

$$\left| \mathbb{P}\left(\bigcap_{t \in G} D_t\right) - \varepsilon^{|G|} \right| \leq \sigma \tag{8.5}$$

- (ii) or $\langle D_t : t \in A^n \rangle$ correlates with a simplicial process when restricted to a large subspace; precisely, there exist $\ell \in [m]$ with the following property. If $\mathbf{r} = (r_1, \dots, r_\ell)$ is an ℓ -tuple of positive integers with $r := \sum_{l=1}^\ell r_l \leq n - m$, then there exist an r -dimensional combinatorial subspace V of A^n and a process $\langle S_t : t \in V \rangle$ in $(\Omega, \mathcal{F}, \mathbb{P})$ that is (ℓ, \mathbf{r}) -simplicial in V such that for every $t \in V$ we have

$$\mathbb{P}(S_t) \geq \frac{\varepsilon^{\kappa-1}}{4\kappa} \quad \text{and} \quad \mathbb{P}(D_t | S_t) \geq \varepsilon + \frac{\sigma}{4^{\kappa-1}}. \tag{8.6}$$

We have already pointed out that the proof of Theorem 6 is conceptually similar to the proofs of Theorems 2 and 5. More precisely, it relies on the following version of Propositions 2 and 3, which, in turn, is based on the higher-dimensional extensions of the archetypical ‘projection’ $t^{\beta \rightarrow \alpha}$. These extensions are presented in Subsection 8.3. (See Definition 7 for the notions of pseudorandomness, supercorrelation and subcorrelation that appear below.)

Proposition 4. *Let A be a finite set with $|A| \geq 2$, let n, p, ℓ be positive integers with $p + 1 \leq |A|^n$, let $0 < \eta, \varepsilon \leq 1$, and let $\langle D_t : t \in A^n \rangle$ be an η -stationary process in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that $|\mathbb{P}(D_t) - \varepsilon| \leq \eta$ for every $t \in A^n$. Let $\mathbf{t} = (t_1, \dots, t_{p+1})$ be a tuple consisting of distinct elements of A^n with $s(\mathbf{t}) = \ell$, set $G := \{t_1, \dots, t_{p+1}\}$ and $H := \{t_1, \dots, t_p\}$ and let d denote the dimension of $\tau(G)$. Finally, let $0 < \theta, \sigma \leq 1$, let $\mathbf{r} = (r_1, \dots, r_\ell)$ be an ℓ -tuple of positive integers such that $r := \sum_{l=1}^\ell r_l \leq n - d$, and assume that the process $\langle D_t : t \in A^n \rangle$ is $(\tau(H), \theta)$ -pseudorandom. Then there exist an r -dimensional combinatorial subspace V of A^n and a process $\langle S_t : t \in V \rangle$ in $(\Omega, \mathcal{F}, \mathbb{P})$ that is (ℓ, \mathbf{r}) -simplicial in V with the following properties:*

- (i) For every $t \in V$ we have $|\mathbb{P}(S_t) - \varepsilon^p| \leq \theta$.

¹³Note that, by Fact 9, if the type $\tau(G)$ of a nonempty set $G \subseteq A^n$ has dimension at most m , then G is ℓ -separated for some $\ell \leq m$.

(ii) If $\langle D_t : t \in A^n \rangle$ is $(\tau(G), \sigma)$ -supercorrelated, then for every $t \in V$ we have

$$\mathbb{P}(D_t | S_t) \geq \varepsilon \left(1 + \frac{\sigma \varepsilon^{-1} - \theta}{\varepsilon^p + \theta} \right). \tag{8.7}$$

(iii) If $\langle D_t : t \in A^n \rangle$ is $(\tau(G), \sigma)$ -subcorrelated, then for every $t \in V$ we have

$$\mathbb{P}(D_t | S_t) \leq \varepsilon \left(1 - \frac{\sigma \varepsilon^{-1} - \theta}{\varepsilon^p - \theta} \right). \tag{8.8}$$

8.3. Definitions/Notation

This subsection is the analogue of Subsection 7.3. More precisely, let A, n, p and ℓ be as in Proposition 4. Let

$$\mathbf{t} = (t_1, \dots, t_{p+1}) \tag{8.9}$$

be an ℓ -separated tuple consisting of distinct elements of A^n , and let d be the dimension of $\tau := \tau(\mathbf{t})$. Also, let $\mathbf{r} = (r_1, \dots, r_\ell)$ be a tuple of positive integers such that $r := \sum_{l=1}^\ell r_l \leq n - d$. We will define

- a set $J \subseteq [d]$ with $|J| = \ell$,
- $\beta_1, \dots, \beta_\ell \in A$,
- an r -dimensional combinatorial subspace V of A^n and
- for every $j \in [p]$ a map $T_j : V \rightarrow A^n$.

These data are the combinatorial core of Theorem 6 and Proposition 4.

8.3.1. Defining J and $\beta_1, \dots, \beta_\ell$

We write the type $\tau(\mathbf{t}) = (s_1, \dots, s_{p+1})$ where $s_j = (s_j(1), \dots, s_j(d)) \in A^d$ for every $j \in [p + 1]$. Because \mathbf{t} is ℓ -separated, by Fact 9 we have $s(\tau) = \ell$. Therefore, there exists $I \subseteq [d]$ with $|I| = \ell$ such that for every $j \in [p]$ there exists $i \in I$ satisfying $s_j(i) \neq s_{p+1}(i)$; let J be the lexicographically least set with this property, write $J = \{\iota_1 < \dots < \iota_\ell\}$ and set

$$\beta_1 := s_{p+1}(\iota_1), \dots, \beta_\ell := s_{p+1}(\iota_\ell). \tag{8.10}$$

Moreover, for every $j \in [p + 1]$ and every $l \in [\ell + 1]$ set

$$y_j^l = \begin{cases} (s_j(\iota_{l-1} + 1), \dots, s_j(\iota_l)) & \text{if } l \in [\ell], \\ (s_j(\iota_\ell + 1), \dots, s_j(d)) & \text{if } l = \ell + 1, \end{cases} \tag{8.11}$$

where $i_0 = 0$ and with the convention that $y_j^{\ell+1}$ is the empty sequence if $\iota_\ell = d$. Notice that $s_j = y_j^1 \frown \dots \frown y_j^{\ell+1}$ for every $j \in [p + 1]$.

8.3.2. Defining V and the maps $\langle T_j : j \in [p] \rangle$

Set

$$V := \{y_{p+1}^1 \frown z_1 \frown \dots \frown y_{p+1}^\ell \frown z_\ell \frown y_{p+1}^{\ell+1} : z_1 \in A^{r_1}, \dots, z_\ell \in A^{r_\ell}\} \tag{8.12}$$

and observe that V is an r -dimensional combinatorial subspace of A^n . Finally, for every $j \in [p]$ we define $T_j : V \rightarrow A^n$ by the rule

$$T_j(y_{p+1}^1 \frown z_1 \frown \dots \frown y_{p+1}^\ell \frown z_\ell \frown y_{p+1}^{\ell+1}) = y_j^1 \frown z_1^{\beta_1 \rightarrow s_j(\iota_1)} \frown \dots \frown y_j^\ell \frown z_\ell^{\beta_\ell \rightarrow s_j(\iota_\ell)} \frown y_j^{\ell+1} \tag{8.13}$$

with the convention $t^{\alpha \rightarrow \alpha} = t$ for every $t \in A^{<\mathbb{N}}$ and every $\alpha \in A$.

8.3.3. Basic properties

We close this subsection with the following analogue of Fact 10. The proof is straightforward.

Fact 11. *Let ℓ, \mathbf{r}, V and $\langle T_j : j \in [p] \rangle$ be as above.*

(i) *For every $t \in V$ and every $1 \leq i_1 < \dots < i_q \leq p$ we have*

$$\tau((T_{i_1}(t), \dots, T_{i_q}(t), t)) = \tau((t_{i_1}, \dots, t_{i_q}, t_{p+1})) \tag{8.14}$$

and

$$\tau((T_{i_1}(t), \dots, T_{i_q}(t))) = \tau((t_{i_1}, \dots, t_{i_q})). \tag{8.15}$$

(ii) *Let $\langle D_t : t \in A^n \rangle$ be a stochastic process in a probability space (Ω, Σ, μ) . Let $\langle S_t : t \in V \rangle$ be a process of the form $S_t = \bigcap_{j=1}^p E_t^j$ where for every $j \in [p]$ either $\langle E_t^j : t \in V \rangle = \langle D_{T_j(t)} : t \in V \rangle$ or $\langle E_t^j : t \in V \rangle = \langle D_{T_j(t)}^c : t \in V \rangle$. Then the process $\langle S_t : t \in V \rangle$ is (ℓ, \mathbf{r}) -simplicial in V .*

8.4. Proof of Proposition 4

Let V and $\langle T_j : j \in [p] \rangle$ be the data obtained in Subsection 8.3 for the ℓ -separated tuple \mathbf{t} and the tuple \mathbf{r} . We define $\langle S_t : t \in V \rangle$ by setting $S_t = \bigcap_{j=1}^p D_{T_j(t)}$ for every $t \in V$. By part (ii) of Fact 11, we see that the process $\langle S_t : t \in V \rangle$ is (ℓ, \mathbf{r}) -simplicial. Moreover, by (8.15) and our assumption that $\langle D_t : t \in A^n \rangle$ is $(\tau(H), \theta)$ -pseudorandom, we have

$$|\mathbb{P}(S_t) - \varepsilon^p| \leq \theta \tag{8.16}$$

for every $t \in V$; that is, part (i) of the theorem holds true. The rest of the proof is identical to that of Proposition 3.

8.5. Proof of Theorem 6

The proof follows arguing precisely as in the proof of Theorem 5 using Proposition 4 and the material in Subsection 8.3.

Acknowledgements. We thank the anonymous reviewers for carefully reading the article and for several helpful suggestions.

Conflict of Interest: None.

References

[Au] T. Austin, ‘Deducing the density Hales–Jewett theorem from an infinitary removal lemma’, *J. Theor. Probab.* **24** (2011), 615–633.
 [DK] P. Dodos and V. Kanellopoulos, *Ramsey Theory for Product Spaces, Mathematical Surveys and Monographs*, 212 (American Mathematical Society, Providence, RI, 2016).
 [DKT1] P. Dodos, V. Kanellopoulos and K. Tyros, ‘A simple proof of the density Hales–Jewett theorem’, *Int. Math. Res. Not.* **12** (2014), 3340–3352.
 [DKT2] P. Dodos, V. Kanellopoulos and K. Tyros, ‘Measurable events indexed by words’, *J. Comb. Theory, Ser. A* **127** (2014), 176–223.
 [DKT3] P. Dodos, V. Kanellopoulos and K. Tyros, ‘A concentration inequality for product spaces’, *J. Funct. Anal.* **270** (2016), 609–620.
 [Du] R. Durrett, *Probability: Theory and Examples, Cambridge Series in Statistical and Probabilistic Mathematics*, fourth edition (Cambridge University Press, Cambridge, 2010).
 [FK1] H. Furstenberg and Y. Katznelson, ‘Idempotents in compact semigroups and Ramsey theory’, *Israel J. Math.* **68** (1989), 257–270.
 [FK2] H. Furstenberg and Y. Katznelson, ‘A density version of the Hales–Jewett theorem’, *J. Anal. Math.* **57** (1991), 64–119.
 [GR] R. L. Graham and B. L. Rothschild, ‘Ramsey’s theorem for n -parameter sets’, *Trans. Amer. Math. Soc.* **159** (1971), 257–292.

- [GRS] R. L. Graham, B. L. Rothschild and J. H. Spencer, *Ramsey Theory*, second edition (John Wiley & Sons, Hoboken, NJ, 1980).
- [HJ] A. H. Hales and R. I. Jewett, 'Regularity and positional games', *Trans. Amer. Math. Soc.* **106** (1963), 222–229.
- [P1] D. H. J. Polymath, 'Density Hales-Jewett and Moser numbers', in *An Irregular Mind. Szemerédi Is 70*, Bolyai Society Mathematical Studies, 21 (Springer, Berlin, 2010), 689–753.
- [P2] D. H. J. Polymath, 'A new proof of the density Hales-Jewett theorem', *Ann. Math.* **175** (2012), 1283–1327.
- [Ro] K. F. Roth, 'On certain sets of integers', *J. London Math. Soc.* **28** (1953), 104–109.
- [Sh] S. Shelah, 'Primitive recursive bounds for van der Waerden numbers', *J. Amer. Math. Soc.* **1** (1988), 683–697.
- [Sp] E. Sperner, 'Ein Satz über Untermengen einer endlichen Menge', *Math. Z.* **27** (1928) 544–548.
- [Tao] T. Tao, *An Epsilon of Room, II: Pages from Year Three of a Mathematical Blog* (American Mathematical Society, Providence, RI, 2011).
- [TV] T. Tao and V. Vu, *Additive Combinatorics, Cambridge Studies in Advanced Mathematics*, 105 (Cambridge University Press, Cambridge, 2006).
- [Ty] K. Tyros, 'On colorings of variable words', *Discr. Math.* **338** (2015), 1025–1028.