A structure theorem for stochastic processes indexed by the discrete hypercube

Pandelis Dodos

University of Athens

Bertinoro, July 15–20, 2018

1.a. Motivation/Overview

Let X and Y be two (say) bounded real-valued random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then

- $\mathrm{X} \approx \mathrm{Y}$ in distribution, *provided that*
- $\mathbb{E}[\mathrm{X}^m] \approx \mathbb{E}[\mathrm{Y}^m]$ for every positive integer m.

We will apply this basic fact to the case

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

where

(1) I is a finite index set,

(2) $\langle E_i : i \in I \rangle$ and $\langle D_i : i \in I \rangle$ are measurable events with $\mathbb{P}(E_i) = \mathbb{P}(D_i) = \varepsilon > 0$ for every $i \in I$, and

(3) $\langle E_i : i \in I \rangle$ are independent.

1.b. Motivation/Overview

By expanding the product, for every positive integer m we have $\mathbb{E}[\mathrm{X}^m] = \sum_{j=0}^{|I|} c_{j,m} \sum_{F\in\binom{I}{j}} \mathbb{P}(\bigcap_{i\in F} E_i)$ for some nonnegative coefficients $c_{0,m},\dots,c_{|I|,m}$, and similarly for Y.

Thus, assuming that X and Y are **not** close in distribution, then one is led to the following problem.

Problem

Let $F \subseteq I$ be nonempty, and assume that

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

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

Here, $\sigma > 0$ is a parameter that measures the deviation of the joint probability of $\langle D_i : i \in F\rangle$ from the expected value.

1.c. Motivation/Overview

We will look at this problem when the index set I is a discrete hypercube.

Let A be a finite set (alphabet) with $|A| \geqslant 2$, let n be a positive integer, and let A^n denote the **discrete n-dimensional hypercube**, that is,

$$A^n := \underbrace{A \times \cdots \times A}_{n-\text{times}}.$$

Thus, elements of A^n are strings (finite sequences) of length n having values in A.

Convention: as we shall see, for our purposes the nature of the set A is irrelevant. Consequently, if $|A| = k$, then it is convenient to identify A with the discrete interval $[k] := \{1, \ldots, k\}$.

2.a. Combinatorial background: the density Hales–Jewett theorem

Let A be a finite set with $|A| \geqslant 2$, and let n be a positive integer. We fix a letter $x \notin A$ which we view as a variable.

• A **variable word over A of length** n is a finite sequence of length n having values in $A \cup \{x\}$ such that the letter x appears at least once. If v is a variable word and $\alpha \in A$, then $v(\alpha)$ denotes the unique element of A^n obtained by replacing all appearances of x in v with α.
E.g., if $v = (1, x, 3, 5, x, 2, 1)$, then $v(2) = (1, \mathbf{2}, 3, 5, \mathbf{2}, 2, 1)$.

• 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.

2.b. Combinatorial background: the density Hales–Jewett theorem

The following result is known as the *density Hales–Jewett theorem*.

Theorem (Furstenberg & Katznelson—1991)

For every integer $k \geqslant 2$ and every $0 < \varepsilon \leqslant 1$ there exists a positive integer $\mathrm{DHJ}(k,\varepsilon)$ with the following property.
If A is a set with $|A| = k$ and $n \geqslant \mathrm{DHJ}(k,\varepsilon)$, then every $D \subseteq A^n$ with $|D| \geqslant \varepsilon |A^n|$ contains a combinatorial line of A^n.

The best known upper bounds for the numbers $\mathrm{DHJ}(k,\varepsilon)$ have an Ackermann-type dependence with respect to k. (We will come back on this issue later on.)

2.c. Combinatorial background: the density Hales–Jewett theorem

The density Hales–Jewett theorem has a number of significant consequences, including:

- Szemerédi's theorem (1975);
- the multidimensional Szemerédi theorem (Furstenberg & Katznelson, 1978);
- the density version of the affine Ramsey theorem (Furstenberg & Katznelson, 1985);
- Szemerédi's theorem for abelian groups (Furstenberg & Katznelson, 1985);
- the IP_r-Szemerédi theorem (Furstenberg & Katznelson, 1985).

3. From dense sets to stochastic processes

Theorem (density Hales–Jewett theorem—reformulation)

For every integer $k \geqslant 2$ *and every* $0 < \varepsilon \leqslant 1$ *there exists a positive integer* $\mathrm{PHJ}(k, \varepsilon)$ *with the following property.*
If A *is a set with* $|A| = k$ *and* $n \geqslant \mathrm{PHJ}(k, \varepsilon)$, *then for every family* $\langle D_t : t \in A^n \rangle$ *of measurable events in a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *satisfying* $\mathbb{P}(D_t) \geqslant \varepsilon$ *for every* $t \in A^n$, *there exists a combinatorial line* L *of* A^n *such that*

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

This result is straightforward if $\langle D_t : t \in A^n \rangle$ are independent. So, the core of the theorem is to understand what happens when the events are not "behaving" as if they were independent.

4.a. From dense sets to stochastic processes: concentration

Let $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1), \dots, (\Omega_n, \mathcal{F}_n, \mathbb{P}_n)$ be a finite sequence of probability spaces, and let $(\boldsymbol{\Omega}, \boldsymbol{\mathcal{F}}, \boldsymbol{P})$ denote their product. More generally, for every nonempty $I \subseteq [n]$ by $(\boldsymbol{\Omega}_I, \boldsymbol{\mathcal{F}}_I, \boldsymbol{P}_I)$ we denote the product of the spaces $\langle (\Omega_i, \mathcal{F}_i, \mathbb{P}_i) : i \in I \rangle$.

Let $I \subseteq [n]$ be such that I and $I^{\mathrm{c}} := [n] \setminus I$ are nonempty. For every integrable random variable $f \colon \boldsymbol{\Omega} \to \mathbb{R}$ and every $\mathbf{x} \in \boldsymbol{\Omega}_I$ let $f_{\mathbf{x}} \colon \boldsymbol{\Omega}_{I^{\mathrm{c}}} \to \mathbb{R}$ be the section of f at $\mathbf{x}$, that is, $f_{\mathbf{x}}(\mathbf{y}) = f(\mathbf{x}, \mathbf{y})$ for every $\mathbf{y} \in \boldsymbol{\Omega}_{I^{\mathrm{c}}}$. Fubini's theorem asserts that the random variable $\mathbf{x} \mapsto \mathbb{E}(f_{\mathbf{x}})$ is integrable and satisfies

$$\int \mathbb{E}(f_{\mathbf{x}}) \, d\boldsymbol{P}_I = \mathbb{E}(f).$$

4.b. From dense sets to stochastic processes: concentration

Theorem (D, Kanellopoulos, Tyros—2014)

Let $0 < \eta \leqslant 1$ *and* $1 < p \leqslant 2$*, and set*

$$c(\eta, p) = \frac{1}{4}\,\eta^{\frac{2(p+1)}{p}}(p-1).$$

Also let n *be a positive integer with* $n \geqslant c(\eta, p)^{-1}$ *and let* $(\boldsymbol{\Omega}, \boldsymbol{\mathcal{F}}, \boldsymbol{P})$ *be the product of a finite sequence* $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1), \ldots, (\Omega_n, \mathcal{F}_n, \mathbb{P}_n)$ *of probability spaces. Then for every* $f \in L_p(\boldsymbol{\Omega}, \boldsymbol{\mathcal{F}}, \boldsymbol{P})$ *with* $\|f\|_{L_p} \leqslant 1$ *there exists an interval* $J \subseteq [n]$ *with* $J^{\mathrm{c}} \neq \emptyset$ *and* $|J| \geqslant c(\eta, p)n$ *such that for every nonempty* $I \subseteq J$ *we have*

$$\boldsymbol{P}_I\big(\{\mathbf{x} \in \boldsymbol{\Omega}_I : |\mathbb{E}(f_{\mathbf{x}}) - \mathbb{E}(f)| \leqslant \eta\}\big) \geqslant 1 - \eta.$$

4.c. From dense sets to stochastic processes: concentration

Theorem ("geometric" formulation)

Let $0 < \eta \leqslant 1$ and $1 < p \leqslant 2$. If $n \geqslant c(\eta, p)^{-1}$ and $(\boldsymbol{\Omega}, \mathcal{F}, \boldsymbol{P})$ is the product of a finite sequence $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1), \dots, (\Omega_n, \mathcal{F}_n, \mathbb{P}_n)$ of probability spaces, then for every $A \in \mathcal{F}$ there exists an interval $J \subseteq [n]$ with $J^{\mathsf{c}} \neq \emptyset$ and $|J| \geqslant c(\eta, p)n$ such that for every nonempty $I \subseteq J$ we have

$$\boldsymbol{P}_I\Big(\big\{\mathbf{x} \in \boldsymbol{\Omega}_I : |\boldsymbol{P}_{I^{\mathsf{c}}}(A_{\mathbf{x}}) - \boldsymbol{P}(A)| \leqslant \eta \boldsymbol{P}(A)^{1/p}\big\}\Big) \geqslant 1 - \eta.$$

This result does not hold true for $p = 1$ (thus, the range of p in the previous theorem is optimal).

4.d. From dense sets to stochastic processes: concentration

Corollary

Let k, m be positive integers with $k \geqslant 2$ and $0 < \eta \leqslant 1$. Also let A be a set with $|A| = k$, and let n be a positive integer with

$$n \geqslant \frac{16\, m\, k^{3m}}{\eta^3}.$$

*Then for every $D \subseteq A^n$ there exists an interval $I \subseteq [n]$ with $|I| = m$ such that for **every** $t \in A^I$ we have*

$$|\mathbb{P}_{A^{I^c}}(D_t) - \mathbb{P}(D)| \leqslant \eta$$

where $D_t = \{s \in A^{I^c} : (t, s) \in D\}$ is the section of D at t.

(Here, all measures are uniform probability measures.)

5.a. Back to the main problem: examples

Example

Here, $A = \{1, 2, 3\}$. Let n be an arbitrary positive integer. We start with a family $\langle E_x : x \in \{1, 2\}^n \rangle$ of independent events in a probability space with equal probability $\varepsilon > 0$. Given $t \in \{1, 2, 3\}^n$ we "project" it into $\{1, 2\}^n$ as follows: let $t^{3\to 1}$ denote the unique element of $\{1, 2\}^n$ obtained by replacing all appearances of 3 in t with 1.
E.g., if $t = (1, 3, 2, 3, 1, 2)$, then $t^{3\to 1} = (1, \mathbf{1}, 2, \mathbf{1}, 1, 2)$.

Define $D_t := E_{t^{3\to 1}}$ for every $t \in \{1, 2, 3\}^n$.

5.b. Back to the main problem: examples

Properties:

• For every $t \in \{1,2,3\}^n$ we have $\mathbb{P}(D_t) = \varepsilon$.

• For every combinatorial line L of $\{1,2,3\}^n$ we have

$$\mathbb{P}\Big(\bigcap_{t\in L} D_t\Big) = \varepsilon^2.$$

Thus, we have significant deviation from what is expected.

• The stochastic process $\langle D_t : t \in \{1,2,3\}^n\rangle$ is $(1,3)$-**insensitive** in the following sense. If $t, s \in \{1,2,3\}^n$ differ only in the coordinates taking values in $\{1,3\}$, then $D_t = D_s$.

$$\textit{E.g.,} \qquad \left.\begin{array}{l} t = (\mathbf{1},2,3,\mathbf{1},2,\mathbf{3}) \\ s = (\mathbf{3},2,3,\mathbf{3},2,\mathbf{1}) \end{array}\right\} \Rightarrow D_t = D_s.$$

5.c. Back to the main problem: examples

Example (cont'd)

Again, let $A = \{1,2,3\}$, let n be an arbitrary positive integer, and let $\langle E_x : x \in \{1,2\}^n \rangle$ be independent events in a probability space with equal probability $\varepsilon > 0$. Given $t \in \{1,2,3\}^n$ let $t^{3\to 1}$ and $t^{3\to 2}$ denote the two "projections" of t into $\{1,2\}^n$.

Define $D_t := E_{t^{3\to 1}} \cap E_{t^{3\to 2}}$ for every $t \in \{1,2,3\}^n$.

- For "almost every" $t \in \{1,2,3\}^n$ we have $\mathbb{P}(D_t) = \varepsilon^2$.
- For "almost every" combinatorial line L of $\{1,2,3\}^n$ we have

$$\mathbb{P}\Big(\bigcap_{t\in L} D_t\Big) = \varepsilon^4$$

.

- Here, D_t is the intersection of "insensitive" events.

6.a. Stationarity

Definition

Let A be a finite set with $|A| \geqslant 2$, let n be a positive integer, and let $0 < \eta \leqslant 1$. We say that a stochastic process $\langle D_t : t \in A^n \rangle$ is η**-stationary** if for every nonempty $B \subseteq A$ and every pair v_1, v_2 of variable words over A of length n we have

$$\Big| \mathbb{P}\Big(\bigcap_{\alpha \in B} D_{v_1(\alpha)} \Big) - \mathbb{P}\Big(\bigcap_{\alpha \in B} D_{v_2(\alpha)} \Big) \Big| \leqslant \eta.$$

Notice, in particular, that if $\langle D_t : t \in A^n \rangle$ is η-stationary, then for every pair L_1, L_2 of combinatorial lines of A^n we have

$$\Big| \mathbb{P}\Big(\bigcap_{t \in L_1} D_t \Big) - \mathbb{P}\Big(\bigcap_{t \in L_2} D_t \Big) \Big| \leqslant \eta.$$

6.b. Stationarity

Stationarity is a mild condition. Specifically, we have the following fact which follows from a classical result due to Graham & Rothschild (1971).

Fact

If n is large enough compared with $|A|$ and η, then for every stochastic process $\langle D_t : t \in A^n \rangle$ one can find a large-dimensional "sub-cube" of A^n such that the restriction of the process on the "sub-cube" is η-stationary.

7.a. The main result

Theorem (D, Tyros—2018)

Let $k \geqslant 2$ be an integer, and let $\varepsilon, \sigma, \eta$ be positive reals with

$$0 < \eta \ll \sigma \ll \varepsilon \leqslant 1 - \frac{1}{2k}.$$

Let A be a finite set with $|A| = k$, let $n \geqslant k$ be an integer, and let $\langle D_t : t \in A^n \rangle$ be an η-stationary process such that $\varepsilon - \eta \leqslant \mathbb{P}(D_t) \leqslant \varepsilon + \eta$ 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*

$$\Big| \mathbb{P}\Big(\bigcap_{t \in G} D_t \Big) - \varepsilon^{|G|} \Big| \leqslant \sigma,$$

7.b. The main result

Theorem (cont'd)

(ii) *or* $\langle D_t : t \in A^n \rangle$ *correlates with a "structured" stochastic process* $\langle S_t : t \in A^n \rangle$*, that is,*

(ii.1) S_t *is the intersection of insensitive events; precisely, there exist nonempty* $B \subseteq A$ *and* $\alpha \in A \setminus B$ *such that* $S_t = \bigcap_{\beta \in B} E_t^\beta$ *where for every* $\beta \in B$ *the stochastic process* $\langle E_t^\beta : t \in A^n \rangle$ *is* (α, β)*-insensitive;*

(ii.2) *for every* $t \in A^n$ *which takes the value* α *(thus, for "almost every"* $t \in A^n$*) we have*

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

8. Comments

• A similar theorem holds true if (instead of combinatorial lines) we look at correlations over an arbitrary nonempty subset F of A^n. Of course, the "structured" process $\langle S_t : t \in A^n \rangle$ depends upon the "geometry" of F (*type*).

• The previous dichotomy yields a new proof of the density Hales–Jewett theorem; in fact, it is a step towards obtaining primitive recursive upper bounds for the density Hales–Jewett numbers (belonging to the class $\mathcal{E}^7$ of Grzegorczyk's hierarchy, or slightly higher).

• Because we assume stationarity, our theorem is "local" in nature. It would be much more desirable if we had a "global" structure theorem. Formulating and proving a useful "global" theorem (with quantitative aspects comparable to our "local" version) might lead to upper bounds for the density Hales–Jewett numbers which are of tower-type, or even better.

9. What about lower bounds?

• For alphabets with two letters the density Hales–Jewett numbers are understood rather well:

$$\frac{1}{\varepsilon} \leqslant \mathrm{DHJ}(2,\varepsilon) \leqslant 4\Big(\frac{1}{\varepsilon}\Big)^2.$$

• However, the case $k \geqslant 3$ is quite different. Specifically, by transferring Behrend's classical construction of a 3AP-free set, one obtains a quasi-polynomial lower bound:

$$2^{O\left((\log \frac{1}{\varepsilon})^{\ell}\right)} \leqslant \mathrm{DHJ}(k,\varepsilon)$$

with $\ell = \Theta(\log k)$.

10. A problem

"Perhaps the Hales–Jewett numbers are exponential."
—József Beck, Combinatorial Games: Tic-Tac-Toe Theory.

Problem

Is it true that

$$\mathrm{DHJ}(k, \varepsilon) \leqslant 2^{\left(\frac{1}{\varepsilon}\right)^{O_k(1)}}$$

with a "reasonable" implied constant?

Thanks for listening!