High-dimensional random arrays. Structural decompositions and concentration. Part II

Pandelis Dodos

Athens, 22 January 2021

Functional Analysis and Operator Algebras Seminar

Joint work with Kostas Tyros and Petros Valettas

1. Random arrays

Definition (Random arrays, and their subarrays)

Let d be a positive integer, and let I be a set with $|I| \geqslant d$.

A *d-dimensional random array on I* is a stochastic process $\boldsymbol{X} = \langle X_s : s \in \binom{I}{d} \rangle$ indexed by the set $\binom{I}{d}$ of all d-element subsets of I.

If J is a subset of I with $|J| \geqslant d$, then the *subarray of $\boldsymbol{X}$ determined by J* is the d-dimensional random array $\boldsymbol{X}_J := \langle X_s : s \in \binom{J}{d} \rangle$.

2.a. Notions of symmetry

• A d-dimensional random array $\boldsymbol{X} = \langle X_s : s \in \binom{I}{d} \rangle$ on I is called *exchangeable* if for every (finite) permutation π of I, the random arrays $\boldsymbol{X}$ and $\boldsymbol{X}_\pi := \langle X_{\pi(s)} : s \in \binom{I}{d} \rangle$ have the same distribution.

• A d-dimensional random array $\boldsymbol{X}$ on I is called *spreadable* if for every pair J, K of finite subsets of I with $|J| = |K| \geqslant d$, the subarrays $\boldsymbol{X}_J$ and $\boldsymbol{X}_K$ have the same distribution.

exhangeability $\Rightarrow$ spreadability

2.b. Notions of symmetry

Definition (Approximate spreadability)

Let $\boldsymbol{X}$ be a d-dimensional random array on a (possibly infinite) set I, and let $\eta \geqslant 0$. We say that $\boldsymbol{X}$ is *η-spreadable* provided that for every pair J, K of finite subsets of I with $|J| = |K| \geqslant d$ we have

$$\rho_{\mathrm{TV}}(P_J, P_K) \leqslant \eta$$

where P_J and P_K denote the laws of the random subarrays $\boldsymbol{X}_J$ and $\boldsymbol{X}_K$ respectively, and ρ_{TV} stands for the total variation distance.

2.c. Notions of symmetry

Fact

For every triple m, n, d of positive integers with $n \geqslant d$, and every $\eta > 0$, there exists an integer $N \geqslant n$ with the following property. If $\mathcal{X}$ is a set with $|\mathcal{X}| = m$ and $\boldsymbol{X}$ is an $\mathcal{X}$-valued, d-dimensional random array on a set I with $|I| \geqslant N$, then there exists a subset J of I with $|J| = n$ such that the random array $\boldsymbol{X}_J$ is η-spreadable.

3. Main goal

- Our main goal is to describe the structure of **finite**, finite-valued, approximately spreadable, high-dimensional random arrays.

- We will also discuss the relation between the structure theorems and the concentration results presented the previous week.

4. Infinite random arrays

The infinitary branch of the theory was developed in a series of foundational papers by Aldous (1981), Hoover (1979) and Kallenberg (1992), with important earlier contributions by Fremlin and Talagrand (1985).

5.a. Aldous–Hoover–Kallenberg theorem: two-dimensional case

Let $\mathcal{X}$ be a Polish space, and let $\boldsymbol{X} = \langle X_s : s \in \binom{\mathbb{N}}{2} \rangle$ be an $\mathcal{X}$-valued, spreadable, two-dimensional random array on $\mathbb{N}$. Then there exists a Borel function $f\colon [0,1]^4 \to \mathcal{X}$ with the following property.

Define an $\mathcal{X}$-valued, spreadable, two-dimensional random array $\boldsymbol{X}_f = \langle X^f_s : s \in \binom{\mathbb{N}}{2} \rangle$ by setting for every $s = \{i < j\} \in \binom{\mathbb{N}}{2}$,

$$X^f_s := f(\xi_\emptyset, \xi_i, \xi_j, \xi_{\{i,j\}})$$

where $\xi_\emptyset, (\xi_i)_{i\in\mathbb{N}}, (\xi_s)_{s\in\binom{\mathbb{N}}{2}}$ are i.i.d. $\mathrm{Unif}[0,1]$.

Then we have

$$\boldsymbol{X} \overset{d}{=} \boldsymbol{X}_f.$$

(If $\boldsymbol{X}$ is exchangeable, then f is “middle-symmetric”, that is, $f(x,y,z,w) = f(x,z,y,w)$.)

5.b. Aldous–Hoover–Kallenberg theorem: three-dimensional case

Let $\mathcal{X}$ be a Polish space, and let $\boldsymbol{X} = \langle X_s : s \in \binom{\mathbb{N}}{3} \rangle$ be an $\mathcal{X}$-valued, spreadable, three-dimensional random array on $\mathbb{N}$. Then there exists a Borel function $f\colon [0,1]^8 \to \mathcal{X}$ with the following property.

Define an $\mathcal{X}$-valued, spreadable, three-dimensional random array $\boldsymbol{X}_f = \langle X^f_s : s \in \binom{\mathbb{N}}{3} \rangle$ by setting for every $s = \{i < j < k\} \in \binom{\mathbb{N}}{3}$,

$$X^f_s := f(\xi_\emptyset, \xi_i, \xi_j, \xi_k, \xi_{\{i,j\}}, \xi_{\{i,k\}}, \xi_{\{j,k\}}, \xi_{\{i,j,k\}})$$

where $\xi_\emptyset, (\xi_i)_{i\in\mathbb{N}}, (\xi_t)_{t\in\binom{\mathbb{N}}{2}}, (\xi_s)_{s\in\binom{\mathbb{N}}{3}}$ are i.i.d. $\mathrm{Unif}[0,1]$.

Then we have $\boldsymbol{X} \overset{d}{=} \boldsymbol{X}_f$.

5.c. Aldous–Hoover–Kallenberg theorem: general case

In general, a d-dimensional spreadable random arrays is represented by a Borel function of 2^d variables. (It is known that the use of 2^d variables is necessary, even for finite-valued random arrays.)

6.a. The Fremlin–Talagrand decomposition

- Let $\mathcal{X}$ be a finite set with $|\mathcal{X}| \geqslant 2$, and let d be a positive integer.

- Let (Ω, Σ, μ) be a probability space, and let Ω^d be equipped with the product measure. We say that a collection $\mathcal{H} = \langle h^a : a \in \mathcal{X} \rangle$ of $[0,1]$-valued random variables on Ω^d is an *$\mathcal{X}$-partition of unity* if $\mathbf{1}_{\Omega^d} = \sum_{a \in \mathcal{X}} h^a$ almost surely.

- With every $\mathcal{X}$-partition of unity $\mathcal{H}$ we associate an $\mathcal{X}$-valued, spreadable, d-dimensional random array $\boldsymbol{X}_{\mathcal{H}} = \langle X_s^{\mathcal{H}} : s \in \binom{\mathbb{N}}{d} \rangle$ on $\mathbb{N}$ whose distribution satisfies the following: for every nonempty finite subset $\mathcal{F}$ of $\binom{\mathbb{N}}{d}$ and every collection $(a_s)_{s \in \mathcal{F}}$ of elements of $\mathcal{X}$, we have

$$(*) \qquad \mathbb{P}\Big(\bigcap_{s \in \mathcal{F}} [X_s^{\mathcal{H}} = a_s] \Big) = \int \prod_{s \in \mathcal{F}} h^{a_s}(\boldsymbol{\omega}_s) \, d\boldsymbol{\mu}(\boldsymbol{\omega})$$

where $\boldsymbol{\mu}$ stands for the product measure on $\Omega^{\mathbb{N}}$ and $\boldsymbol{\omega}_s$ denotes the restriction of $\boldsymbol{\omega}$ on the coordinates determined by s.

6.b. The Fremlin–Talagrand decomposition

• These distributions were considered by Fremlin and Talagrand who showed that if "$d=2$" and "$\mathcal{X}=\{0,1\}$", then they are precisely the extreme points of the compact convex set of all distributions of boolean, spreadable, two-dimensional random arrays on $\mathbb{N}$.

• This fact together with Choquet's representation theorem yield that the distribution of an arbitrary boolean, spreadable, two-dimensional random array on $\mathbb{N}$ is a mixture of distributions of the form $(*)$.

6.c. The Fremlin–Talagrand decomposition

- Instead of mixtures we will consider finite convex combinations. Specifically, let J be a nonempty finite index set, let $\boldsymbol{\lambda} = \langle \lambda_j : j \in J \rangle$ be convex coefficients, and let $\boldsymbol{\mathcal{H}} = \langle \mathcal{H}_j : j \in J \rangle$ be $\mathcal{X}$-partitions of unity.

- Given these data, we define an $\mathcal{X}$-valued, spreadable, d-dimensional random array $\boldsymbol{X}_{\boldsymbol{\lambda},\boldsymbol{\mathcal{H}}} = \langle X_s^{\boldsymbol{\lambda},\boldsymbol{\mathcal{H}}} : s \in \binom{\mathbb{N}}{d} \rangle$ on $\mathbb{N}$ whose distribution satisfies

$$(*)' \qquad \mathbb{P}\Big(\bigcap_{s\in\mathcal{F}} [X_s^{\boldsymbol{\lambda},\boldsymbol{\mathcal{H}}} = a_s] \Big) = \sum_{j\in J} \lambda_j \int \prod_{s\in\mathcal{F}} h_j^{a_s}(\boldsymbol{\omega}_s)\, d\boldsymbol{\mu}_j(\boldsymbol{\omega})$$

for every nonempty finite subset $\mathcal{F}$ of $\binom{\mathbb{N}}{d}$ and every collection $(a_s)_{s\in\mathcal{F}}$ of elements of $\mathcal{X}$.

7.a. The main results — Part I

Theorem (D, Tyros, Valettas–2020; distributional decomposition)

Let d, m, k be positive integers with $m \geqslant 2$ and $k \geqslant d$, let $0 < \varepsilon \leqslant 1$, and set

$$C = C(d, m, k, \varepsilon) := \exp^{(2d)}\Big(\frac{2^8\, m^{7k^d}}{\varepsilon^2}\Big)$$

where for every positive integer ℓ by $\exp^{(\ell)}(\cdot)$ we denote the ℓ-th iterated exponential.

Also let $n \geqslant C$ be an integer, let $\mathcal{X}$ be a set with $|\mathcal{X}| = m$, and let $\boldsymbol{X} = \langle X_s : s \in \binom{[n]}{d}\rangle$ be an $\mathcal{X}$-valued, $(1/C)$-spreadable, d-dimensional random array on $[n]$.

7.b. The main results — Part I

Theorem (distributional decomposition; cont'd)

Then there exist

- *two nonempty finite sets J and Ω with $|J|, |\Omega| \leqslant C$,*
- *convex coefficients $\boldsymbol{\lambda} = \langle \lambda_j : j \in J \rangle$, and*
- *for every $j \in J$ a probability measure μ_j on the set Ω and an $\mathcal{X}$-partition of unity $\mathcal{H}_j = \langle h_j^a : a \in \mathcal{X} \rangle$ defined on Ω^d*

such that, setting $\boldsymbol{\mathcal{H}} := \langle \mathcal{H}_j : j \in J \rangle$ and letting $\boldsymbol{X}_{\boldsymbol{\lambda},\boldsymbol{\mathcal{H}}}$ be as in $(*)'$, the following holds. If L is a subset of $[n]$ with $|L| = k$, and P_L and Q_L denote the laws of the subarrays of $\boldsymbol{X}$ and $\boldsymbol{X}_{\boldsymbol{\lambda},\boldsymbol{\mathcal{H}}}$ determined by L respectively, then we have

$$\rho_{\mathrm{TV}}(P_L, Q_L) \leqslant \varepsilon.$$

7.c. The main results — Part II

Theorem (D, Tyros, Valettas–2020)

Let the parameters d, m, k, ε and the constant C be as in the previous theorem. Also let n, $\mathcal{X}$, $\boldsymbol{X}$ be as in the previous theorem.

Then there exists a Borel measurable function $f\colon [0,1]^{d+1} \to \mathcal{X}$ with the following property. Let $\boldsymbol{X}_f = \langle X^f_s : s \in \binom{\mathbb{N}}{d} \rangle$ be the $\mathcal{X}$-valued, spreadable,d-dimensional random array on $\mathbb{N}$ defined by setting for every $s = \{i_1 < \cdots < i_d\} \in \binom{\mathbb{N}}{d}$,

$$X^f_s = f(\zeta, \xi_{i_1}, \dots, \xi_{i_d})$$

where $(\zeta, \xi_1, \dots)$ are i.i.d. $\mathrm{Unif}[0,1]$.

Then, for every subset L of $[n]$ with $|L| = k$, denoting by P_L and Q_L the laws of the subarrays of $\boldsymbol{X}$ and $\boldsymbol{X}_f$ determined by L respectively, we have $\rho_{\mathrm{TV}}(P_L, Q_L) \leqslant \varepsilon$.

7.d. The main results — Part II

- Of course, the previous result is akin to the Aldous–Hoover–Kallenberg representation theorem. The main difference is that the number of variables which are needed in order to represent the random array $\boldsymbol{X}$ is $d+1$, while the corresponding number of variables required by the Aldous–Hoover–Kallenberg theorem is 2^d.

- This particular information is a genuinely finitary phenomenon, and it is important for the results related to concentration which we will discuss shortly.

7.e. The main results — Part III

For finite, spreadable, high-dimensional random arrays with *square integrable* entries we have a physical decomposition which is in the spirit of the classical Hoeffding/Efron–Stein decomposition.

It is less informative than the previous results, but this is offset by the fact that it applies to a fairly large class of distributions (including bounded, gaussian, subgaussian, etc.).

8.a. Ideas of the proof

Both proofs proceed by induction on d. We actually prove a slightly stronger result which encompasses the previous theorems and it is more amenable to an inductive scheme. There are two basic steps in the proof.

Step 1. We approximate, in distribution, any finite-valued, approximately spreadable random array by a random array of "lower-complexity". A similar approximation is used in the proof of the Aldous–Hoover theorem. However, our argument is technically different since we work with approximately spreadable, instead of exchangeable, random arrays.

The notion of "complexity" which appears in this context is related to the notion of "complexity" which appears in hypergraph regularity (that is, in the development of the regularity method for uniform hypergraphs). The relation was first pointed out by Austin/Tao.

8.b. Ideas of the proof

Example (two-dimensional, boolean case)

We find a “large” subset L of $[n]$, and a collection $\langle \mathcal{A}_i : i \in L\rangle$ of σ-algebras with the following property.

Define $\boldsymbol{Y}_L = \langle Y_s : s \in \binom{L}{2}\rangle$ by setting for every $s = \{i < j\} \in \binom{L}{2}$,

$$Y_s := \mathbb{E}[X_s \,|\, \mathcal{A}_i \vee \mathcal{A}_j].$$

Then we have $\boldsymbol{X}_L \overset{d}{\approx} \boldsymbol{Y}_L$.

8.c. Ideas of the proof

Example (three-dimensional, boolean case)

We find a "large" subset L of $[n]$, and a collection $\langle \mathcal{A}_t : t \in \binom{L}{2} \rangle$ of σ-algebras with the following property.

Define $\boldsymbol{Y}_L = \langle Y_s : s \in \binom{L}{3} \rangle$ by setting for every $s = \{i < j < k\} \in \binom{L}{3}$,

$$Y_s := \mathbb{E}[X_s \,|\, \mathcal{A}_{\{i,j\}} \vee \mathcal{A}_{\{i,k\}} \vee \mathcal{A}_{\{j,k\}}].$$

Then we have $\boldsymbol{X}_L \overset{d}{\approx} \boldsymbol{Y}_L$.

8.d. Ideas of the proof

Step 2. We show that the laws of finite subarrays of the form $(*)$ can be approximated, with arbitrary accuracy, by the laws of subarrays of distributions of the form $(*)$ which are generated by genuine partitions instead of partitions of unity.

More precisely, given an $\mathcal{X}$-partition of unity $\mathcal{H} = \langle h^a : a \in \mathcal{X} \rangle$ on a finite probability space (Ω, μ), a positive integer κ and $\varepsilon > 0$, we find a finite probability space (Y, λ) and a partition $\mathcal{E} = \langle E^a : a \in \mathcal{X} \rangle$ of Y^d such that

$$\Big| \int \prod_{s \in \mathcal{F}} h^{a_s}(\boldsymbol{\omega}_s)\, d\boldsymbol{\mu}(\boldsymbol{\omega}) - \int \prod_{s \in \mathcal{F}} \mathbf{1}_{E^{a_s}}(\boldsymbol{y}_s)\, d\boldsymbol{\lambda}(\boldsymbol{y}) \Big| \leqslant \varepsilon$$

for every nonempty subset $\mathcal{F}$ of $\binom{\mathbb{N}}{d}$ with $|\mathcal{F}| \leqslant \kappa$ and every collection $(a_s)_{s \in \mathcal{F}}$ of elements of $\mathcal{X}$.

8.e. Ideas of the proof

The proof of this step is based on a random selection of uniform hypergraphs and basic properties of the box norms introduced by Gowers.

(Recall that if $d \geqslant 2$ is an integer and (Ω, Σ, μ) is a probability space, then for every $h\colon \Omega^d \to \mathbb{R}$ we define its *box norm* $\|h\|_\square$ by setting

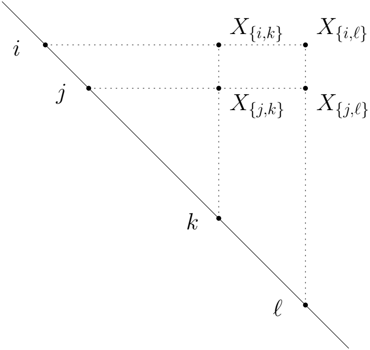
$$\|h\|_\square := \Big(\int \prod_{\epsilon \in \{0,1\}^d} h(\boldsymbol{\omega}_\epsilon)\, d\boldsymbol{\mu}(\boldsymbol{\omega}) \Big)^{1/2^d}$$

where $\boldsymbol{\mu}$ denotes the product measure on Ω^{2d} and, for every $\boldsymbol{\omega} = (\omega_1^0, \omega_1^1, \dots, \omega_d^0, \omega_d^1) \in \Omega^{2d}$ and every $\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_d) \in \{0,1\}^d$ we have $\boldsymbol{\omega}_{\boldsymbol{\epsilon}} := (\omega_1^{\epsilon_1}, \dots, \omega_d^{\epsilon_d}) \in \Omega^d$.)

9.a. Connection with concentration

Box independence condition: if $\boldsymbol{X} = \langle X_s : s \in \binom{[n]}{2} \rangle$, then for every $i, j, k, \ell \in [n]$ with $i < j < k < \ell$ we have

$$\Big|\mathbb{E}[X_{\{i,k\}} X_{\{i,\ell\}} X_{\{j,k\}} X_{\{j,\ell\}}] - \\ - \mathbb{E}[X_{\{i,k\}}]\,\mathbb{E}[X_{\{i,\ell\}}]\,\mathbb{E}[X_{\{j,k\}}]\,\mathbb{E}[X_{\{j,\ell\}}]\Big| \leqslant \frac{6}{C}.$$



9.b. Connection with concentration

On the other hand, from the representation theorem, for every integer $k \geqslant 4$ and every $\varepsilon > 0$ there exist

- two nonempty finite sets J, Ω,
- convex coefficients $\boldsymbol{\lambda} = \langle \lambda_j : j \in J \rangle$, and
- for every $j \in J$ a probability measure μ_j on Ω, and a function $h_j \colon \Omega \times \Omega \to [0,1]$

such that for every nonempty subset $\mathcal{F}$ of $\binom{[n]}{2}$ with $|\mathcal{F}| \leqslant 4$,

$$\Big| \mathbb{E}\Big[\prod_{s \in \mathcal{F}} X_s \Big] - \sum_{j \in J} \lambda_j \int \prod_{s \in \mathcal{F}} h_j(\boldsymbol{\omega}_s)\, d\boldsymbol{\mu}_j(\boldsymbol{\omega}) \Big| \leqslant \varepsilon.$$

9.c. Connection with concentration

Set $\delta := \mathbb{E}[X_{\{1,2\}}]$. Then the following are equivalent.

- $\boldsymbol{X}$ satisfies the box independence condition.
- For "almost every" $j \in J$ we have
  (i) $\mathbb{E}[h_j] \approx \delta$, and
  (ii) h_j is *box uniform*, that is, $\big\| h_j - \mathbb{E}[h_j] \big\|_\square \approx 0$.

(Here, $\| \cdot \|_\square$ denotes the corresponding box norm.)

Thanks again for listening!