MODEL THEORY AND HYPERGRAPH REGULARITY (MATH 223M, UCLA, FALL 2016)

ARTEM CHERNIKOV

Lecture notes in progress. All comments and corrections are welcome.
Last update: November 23, 2016

Contents

1. Ultrafilters, ultraproducts and ultralimits

1.1. **Filters and ultrafilters.** Let I be a set, and let $\mathcal{P}(I)$ denote the set of all subsets of I. Given a subset $S \subseteq I$, we denote by $\neg S$ the complement of S in I, i.e. $\neg S := I \setminus S$.

Definition 1.1. A *filter on* I is a collection $\mathcal{F}$ of subsets of I such that:

(1) $I \in \mathcal{F}, \emptyset \notin \mathcal{F}$,
(2) $A, B \in \mathcal{F} \implies A \cap B \in \mathcal{F}$,
(3) $A \in \mathcal{F}$ and $A \subseteq B \subseteq I \implies B \in \mathcal{F}$.

It follows from the definition that $I \in \mathcal{F}$ and that the intersection of finitely many sets in $\mathcal{F}$ is also in $\mathcal{F}$. Intuitively, one can think of a filter as a collection of "large" subsets of I.

Example 1.2. (1) Assume that I is an infinite set. Then $\mathcal{F} = \{\neg S : S \subseteq I \text{ finite}\}$ is the *Fréchet* filter on I.
(2) Fix a non-empty set $A \subseteq I$. Then $\mathcal{F} = \{S \subseteq I : A \subseteq S\}$ is the *principal* filter generated by A.

Definition 1.3. We say that a filter $\mathcal{U}$ on I is an *ultrafilter* if for *every* set $S \subseteq I$, either $S \in \mathcal{U}$ or $\neg S \in \mathcal{U}$.

Fact 1.4. *For any filter $\mathcal{F}$ on I there is an ultrafilter $\mathcal{U}$ on I with $\mathcal{F} \subseteq \mathcal{U}$.*

This fact is equivalent (modulo ZFC) to a weak form of the axiom of choice called the *Boolean prime ideal theorem.*

Remark 1.5. (1) Ultrafilters are precisely the maximal filters (under inclusion).
(2) Assume that $\mathcal{U}$ is an ultrafilter on I, $S \in \mathcal{U}$ and $S = S_1 \cup \ldots \cup S_n$. Then $S_i \in \mathcal{U}$ for at least one $1 \leq i \leq n$.
(3) Note that if $a \in I$, then the principal filter generated by $\{a\}$ is an ultrafilter.
(4) An ultrafilter $\mathcal{U}$ on I is non-principal if and only if it extends the Fréchet filter on I. In particular, every infinite set admits a non-principal ultrafilter on it.
(5) In fact, for any infinite set I there are $2^{2^{|I|}}$ different non-principal ultrafilters on it.
(6) For any infinite set $S \subseteq \mathbb{N}$, there is an ultrafilter $\mathcal{U}$ on $\mathbb{N}$ with $S \in \mathcal{U}$.

Non-principal ultrafilters provide a tool for finding a "generic" object associated to an infinite collection of objects. We will need two instances of this idea.

1.2. Ultralimits.

Let (X, d) be a metric space, and let $\mathcal{U}$ be a *non-principal* ultrafilter on $\mathbb{N}$.

Definition 1.6. Let $(x_n)_{n\in\mathbb{N}}$ be a sequence of points in X. The point $x \in X$ is called the *ultralimit of* x_n (relatively to $\mathcal{U}$), denoted $x = \lim_{\mathcal{U}} x_n$, if for every $\varepsilon > 0$ we have $\{n \in \mathbb{N} : d(x_n, x) \leq \varepsilon\} \in \mathcal{U}$.

Remark 1.7. (1) If an ultralimit of a sequence of points exists, then it is unique.
(2) If $x = \lim_{n\to\infty} x_n$ in the usual sense of metric limits, then $x = \lim_{\mathcal{U}} x_n$ (uses that $\mathcal{U}$ is non-principal).

Fact 1.8. *If (X, d) is compact and $\mathcal{U}$ is a non-principal ultrafilter on $\mathbb{N}$, then **any** sequence of points in X has an ultralimit relatively to $\mathcal{U}$.*

Corollary 1.9. *Any* bounded *sequence $(x_n : n \in \mathbb{N})$ of real numbers has a well-defined ultralimit in $\mathbb{R}$ relatively to any non-principal ultrafilter on $\mathcal{U}$ (as closed intervals are compact).*

Of course, this limit depends on the ultrafilter. For example, let $x_n = 0$ if n is even and $x_n = 1$ if n is odd. Then $\lim_{\mathcal{U}} x_n = 0$ for any ultrafilter $\mathcal{U}$ on $\mathbb{N}$ containing the set of even numbers, and $\lim_{\mathcal{U}} x_n = 1$ for any ultrafilter on $\mathbb{N}$ containing the set of odd numbers.

1.3. **Some model-theoretic notation.**

Definition 1.10. A *(first-order) structure*

$$\mathcal{M} = (M, R_1, R_2, \ldots, f_1, f_2, \ldots, c_1, c_2, \ldots)$$

consists of an underlying set M, together with some distinguished relations R_i (subsets of M^{n_i}, $n_i \in \mathbb{N}$), functions $f_i : M^{n_i} \to M$, and constants c_i (distinguished elements of M). We refer to the collection of all these relations, function symbols and constants as *the signature* of $\mathcal{M}$, or the language of $\mathcal{M}$.

Example 1.11. A group can be naturally viewed as a structure $\left(G, \cdot, ^{-1}, 1\right)$, as well as a ring $(R, +, \cdot, 0, 1)$, an ordered set $(X, <)$, a graph (X, E), etc.

Definition 1.12. A *formula* is an expression of the form

$$\psi(y_1, \ldots, y_m) = \forall x_1 \exists x_2 \ldots \forall x_{n-1} \exists x_n \phi(x_1, \ldots, x_n; y_1, \ldots, y_n),$$

where ϕ is given by a boolean combination of (superpositions of) the basic relations and functions (and $y_{1,} \ldots, y_n$ are the *free variables* of ψ).

We denote the set of all formulas by $\mathcal{L}$. We also consider formulas with parameters, i.e. expressions of the form $\psi\left(\bar{y}, \bar{b}\right)$ with $\psi \in \mathcal{L}$ and $\bar{b}$ a tuple of elements in M. Given a set of parameters $B \subseteq M$, we let $\mathcal{L}(B) = \left\{\psi\left(\bar{y}, \bar{b}\right) : \psi \in L, \bar{b} \in B^{\left|\bar{b}\right|}\right\}$. If $\psi(\bar{y}) \in \mathcal{L}(B)$ is satisfied by a tuple $\bar{a}$ of elements of M, we denote it as $\mathcal{M} \models \psi(\bar{a})$ or $a \models \psi(\bar{y})$, and we call $\bar{a}$ a solution of ψ. If $\Psi(\bar{y})$ is a set of formulas, we write $a \models \Psi(\bar{y})$ to denote that $a \models \psi(\bar{y})$ for all $\psi \in \Psi$. Given a set $A \subseteq M^{|x|}$, we denote by $\psi(A)$ the set $\left\{a \in A^{|x|} : \mathcal{M} \models \psi(A)\right\}$ of all solutions of ψ in A. We say that $X \subseteq M^n$ is an A-*definable set* if there is some $\psi(\bar{x}) \in \mathcal{L}(A)$ such that $X = \psi(M^n)$. If ψ has no free variables, then it is called a *sentence*, and it is either true or false in $\mathcal{M}$. By the *theory of* $\mathcal{M}$, or $\mathrm{Th}(\mathcal{M})$, we mean the collection of all sentences that are true in M.

1.4. **Ultraproducts of first-order structures.**

Let $\mathcal{L}$ be a language and I an infinite set. Suppose that $\mathcal{M}_i$ is an $\mathcal{L}$-structure for each $i \in I$. Let $\mathcal{U}$ be an ultrafilter on I. We define a new structure $\mathcal{M} = \prod \mathcal{M}_i / \mathcal{U}$, which we call the *ultraproduct* of the $\mathcal{M}_i$ modulo $\mathcal{U}$.

- Define a relation $\sim$ on $X := \prod_{i \in I} M_i$ by:
  given $a = (a(i) : i \in I), b = (b(i) : i \in I)$ in X, $a \sim b$ if and only if $\{i \in I : a(i) = b(i)\} \in \mathcal{U}$.
- $\sim$ is an equivalence relation on X (using that $\mathcal{U}$ is an ultrafilter), and given a in X, we denote its $\sim$-equivalence class by $[a]$.
- The universe of $\mathcal{M}$ will be $M = X/\sim$, i.e. the set of the equivalence classes relatively to $\sim$.
- If c is a constant symbol of $\mathcal{L}$, let $c^{\mathcal{M}} := \left[\left(c^{\mathcal{M}_i} : i \in I\right)\right]$.
- If $f(x_1, \ldots, x_n)$ is a function symbol in $\mathcal{L}$ and $[a_1], \ldots, [a_n] \in M$, we define $f^{\mathcal{M}}([a_1], \ldots, [a_n]) := \left[f^{\mathcal{M}_i}(a_1(i), \ldots, a_n(i))\right]$.

- If $R(x_1, \ldots, x_n)$ is a relation symbol in $\mathcal{L}$, we define $R^{\mathcal{M}}$ on M^n by saying that $R^{\mathcal{M}}([a_1], \ldots, [a_n])$ holds in $\mathcal{M}$ if and only if
$$\{i \in I : \mathcal{M}_i \models R(a_1(i), \ldots, a_n(i))\} \in \mathcal{U}.$$

Exercise 1.13. Check that this is well-defined using the properties of ultrafilters.

Fact 1.14. *(Łoś theorem) Let $\phi(x_1, \ldots, x_n)$ be an $\mathcal{L}$-formula, and let $\mathcal{M} = \prod_{i \in I} \mathcal{M}_i / \mathcal{U}$. Then for any $[a_1], \ldots, [a_n] \in \mathcal{M}$,*
$$\mathcal{M} \models \phi([a_1], \ldots, [a_n]) \iff \{i \in I : \mathcal{M}_i \models \phi(a_1(i), \ldots, a_n(i))\} \in \mathcal{U}.$$

Hence one can think of $\mathcal{M}$ as a "limit" of the structures $\mathcal{M}_i, i \in I$: a formula holds in $\mathcal{M}$ if it holds in $\mathcal{M}_i$ *for some/any large set of* $i \in I$ (relatively to $\mathcal{U}$).

Corollary 1.15. *For each set of sentences T in $\mathcal{L}$, every ultraproduct of models of T is a model of T.*

Corollary 1.16. *(Compactness theorem of first-order logic) If T is a set of sentences (of arbitrary cardinality) such that every* ***finite*** *subset $T_0 \subseteq T$ is consistent, then T is consistent. (Exercise)*

Example 1.17. Let $\mathcal{U}$ be a non-principal ultrafilter on $\mathbb{N}$. Let $\mathcal{M}_i = (\{0, 1, \ldots, i-1\}, <)$ be a finite linear order on i elements. Let $\mathcal{M} := \prod_{i \in \mathbb{N}} \mathcal{M}_i / \mathcal{U}$, and let $T := \mathrm{Th}(\mathcal{M})$. For any $i \in \mathbb{N}$, $\mathcal{M}_i$ has the first and the last elements, and is a discrete linear order (i.e. every element has immediate successor and predecessor) of size $\geq i$. Each of these properties can be expressed by a first-order sentence. Hence, by Łoś theorem, $\mathcal{M}$ is an *infinite* discrete linear order with endpoints (these properties axiomatize a complete first-order theory, hence determine T). In fact, $\mathcal{M} \cong \mathbb{N} + \sum_{j \in L} \mathbb{Z} + \mathbb{N}^*$, where L is a dense linear order without endpoints and $\mathbb{N}^*$. What is the cardinality of $\mathcal{M}$? We will find out soon.

Definition 1.18. Let $\mathcal{M}$ be an $\mathcal{L}$-structure.

(1) Let A be a set of parameters in M. By a *partial type* $\Phi(x)$ *over* A (where x is an ordered tuple of variables) we mean a collection of $\mathcal{L}$-formulas of the form $\phi(x)$ with parameters from A such that every **finite** subcollection has a common solution in $\mathcal{M}$.
(2) By a *complete type* over A we mean a partial type such that for *every* formula $\phi(x) \in \mathcal{L}(A)$, either $\phi(x)$ or $\neg\phi(x)$ is in it. For $b \in \mathcal{M}$, we denote by $\mathrm{tp}(b/A)$ the *complete type of* b *over* A, i.e.
$$\mathrm{tp}(b/A) = \{\phi(x) : b \models \phi(x), \phi(x) \in \mathcal{L}(A)\}.$$
(3) We say that a (partial) type $\Phi(x)$ is *realized* in $\mathcal{M}$ if there is some $b \in \mathcal{M}$ satisfying simultaneously all of the formulas in Φ.

Example 1.19. Let $\mathcal{M} = (\mathbb{R}, +, \times, 0, 1)$ be the field of real numbers. The partial type $\Phi(x) = \{x < n : n \in \mathbb{N}\}$ over $\emptyset$ is not realized in $\mathbb{R}$ (where $n = \underbrace{1 + \ldots + 1}_{n \text{ times}}$).

Definition 1.20. Let κ be a cardinal. A structure $\mathcal{M}$ is *κ-saturated* if every partial type over a set of parameters of size $< \kappa$ is realized in $\mathcal{M}$.

Consider again $\Phi(x)$ from the previous example. It shows that $\mathbb{R}$ is not $\aleph_0$-saturated. Let $\mathcal{U}$ be a non-principal ultrafilter on $\mathbb{N}$, and let $\mathbb{R}^* := \mathbb{R}^I / \mathcal{U}$. Then $\Phi(x)$ is also a partial type over $\emptyset$ in $\mathbb{R}^*$, and $[(n : n \in \mathbb{N})]$ is an element of $\mathbb{R}^*$ realizing $\Phi(x)$ (using Łoś theorem). More generally:

Proposition 1.21. *Let $\mathcal{L}$ be a countable language, $(\mathcal{M}_i : i \in \mathbb{N})$ a sequence of $\mathcal{L}$-structures and $\mathcal{U}$ a non-principal ultrafilter on $\mathbb{N}$. Then the ultraproduct $\mathcal{M} = \prod_{i\in\mathbb{N}} \mathcal{M}_i/\mathcal{U}$ is $\aleph_1$-saturated (i.e. every partial type over a* ***countable*** *set of parameters is realized in $\mathcal{M}$).*

Proof. Let $\Phi(x)$ be a partial type over a countable set of parameters $A \subseteq M$. As $\mathcal{L}$ is countable, $\Phi(x)$ can be enumerated as $\{\phi_n(x, [a_n]) : n \in \mathbb{N}\}$, $\phi_n(x, [a_n]) \in \mathcal{L}(\mathcal{M})$. Let $X_0 = \mathbb{N}$ and for $1 \leq n \in \mathbb{N}$ let

$$X_n = \{i \in \mathbb{N} : \mathcal{M}_i \models \exists x\, \phi_1(x, a_1(i)) \wedge \ldots \wedge \phi_n(x, a_n(i))\} \cap [n, \infty).$$

As $\Phi(x)$ is a partial type, every finite set of formulas from Φ is realized in $\mathcal{M}$. In particular, $\mathcal{M} \models \exists x\, \phi_1(x, [a_1]) \wedge \ldots \wedge \phi_n(x, [a_n])$ for all $n \in \mathbb{N}$. As $\mathcal{U}$ is non-principal, by Łoś theorem it follows that $X_n \in \mathcal{U}$ for all $n \in \mathbb{N}$. Moreover, $\bigcap_{n\in\mathbb{N}} X_n = \emptyset$ and $X_n \supseteq X_{n+1}$. Hence for every $i \in \mathbb{N}$ there is a greatest $n(i) \in \mathbb{N}$ such that $i \in X_{n(i)}$.

We define a sequence $b = (b(i) : i \in \mathbb{N})$ as follows. If $n(i) = 0$ let $b(i)$ be an arbitrary element in $\mathcal{M}_i$. If $n(i) > 0$, let $b(i)$ be some element in $\mathcal{M}_i$ realizing $\phi_1(x, a_1(i)) \wedge \ldots \wedge \phi_{n(i)}\left(x, a_{n(i)}(i)\right)$.

Now fix any $n > 0$. Then for any $i \in X_n$ we have $n \leq n(i)$, hence $\mathcal{M}_i \models \phi_n(b(i), a_n(i))$. As $X_n \in \mathcal{U}$, it follows that $\mathcal{M} \models \phi_n([b], [a_n])$. As this holds for any n, $[b]$ realizes $\Phi(x)$ in $\mathcal{M}$. □

Note that every infinite κ-saturated structure $\mathcal{M}$ has size at least κ (if $|\mathcal{M}| < \kappa$, then $\{x \neq a : a \in M\}$ is a partial type over a set of size $< \kappa$ which cannot be realized in $\mathcal{M}$). If follows from Proposition 1.21 that any ultraproduct relatively to a non-principal ultrafilter in $\mathbb{N}$ is either finite or of size at least $\aleph_1$. In fact, more is true.

Proposition 1.22. *Let $\mathcal{U}$ be a non-principal ultrafilter on $\mathbb{N}$. Then any ultraproduct $\mathcal{M} = \prod_{i\in\mathbb{N}} \mathcal{M}_i/\mathcal{U}$ is either finite or of cardinality $\geq 2^{\aleph_0}$.*

Proof. Assume that $\mathcal{M}$ is infinite.

Claim 1. There is a family $\mathcal{F}$ of functions $f : \mathbb{N} \to \mathbb{N}$ such that:

(1) $|\mathcal{F}| = 2^{\aleph_0}$,
(2) $f(n) < 2^n$ for any $f \in \mathcal{F}$ and $n \in \mathbb{N}$,
(3) if $f \neq g$ are in $\mathcal{F}$, then $\{n : f(n) = g(n)\}$ is finite.

Proof of Claim 1. For each $A \subseteq \mathbb{N}$, let $f_A : \mathbb{N} \to \mathbb{N}$ be given by $f_A(n) = \sum_{k<n} \mathbf{1}_A(k) 2^k$, where $\mathbf{1}_A$ is the indicator function of A, i.e. $\mathbf{1}_A(k) = 1$ if $k \in A$, and $\mathbf{1}_A(k) = 0$ otherwise. Then $\mathcal{F} = \{f_A : A \subseteq \mathbb{N}\}$ is as needed.

Claim 2. There is a set $S \in \mathcal{U}$ and a partition $S = \bigcup_{n\in\mathbb{N}} A_n$ such that:

(1) $A_n \notin \mathcal{U}$ for all $n \in \mathbb{N}$,
(2) if $i \in A_n$, then $|\mathcal{M}_i| \geq 2^n$.

Proof of Claim 2. Let $S_0 = \{i \in \mathbb{N} : \mathcal{M}_i \text{ is finite}\}$, $S_1 = \{i \in \mathbb{N} : \mathcal{M}_i \text{ is infinite}\}$. As $\mathbb{N} = S_0 \cup S_1$, we have $S_t \in \mathcal{U}$ for some $t \in \{0, 1\}$.

If $S_0 \in \mathcal{U}$, we let $S := S_0$ and let $A_n = \left\{i \in S : 2^n \leq |\mathcal{M}_i| < 2^{n+1}\right\}$. The sets A_n clearly partition $\mathbb{N}$. Assume that $A_n \in \mathcal{U}$ for some n. As having at most 2^{n+1} elements is a property of a structure expressible by a first-order sentence, it would follow by Łoś theorem that $|\mathcal{M}| \leq 2^{n+1}$ — contrary to the assumption. Hence $A_n \notin \mathcal{U}$ for all $n \in \mathbb{N}$.

If $S_1 \in \mathcal{U}$, say $S_1 = \{a_i : i \in \mathbb{N}\}$, we can just take $S = S_1$ and $A_n = \{a_n\}$.

Now for each $i \in A_n$, by Claim 2 let $\{a_{i,j} : j < 2^n\}$ be some 2^n *distinct* elements of $\mathcal{M}_i$. For $f \in \mathcal{F}$ as in Claim 1, define $c_f \in \prod_{i \in \mathbb{N}} \mathcal{M}_i$ by $c_f(i) := a_{i,f(n)}$, where n is such that $i \in A_n$, when $i \in S$, and let $c_f(i)$ be an arbitrary element in $\mathcal{M}_i$ if $i \notin S$.

Note that if $f \neq g$ are in $\mathcal{F}$, then

$$S' := \{i \in S : c_f(i) = c_g(i)\} = \bigcup \{A_n : n \in \mathbb{N}, f(n) = g(n)\}$$

is a finite union of the sets $A_n \notin \mathcal{U}$, hence $S' \notin \mathcal{U}$. But then

$$S \setminus S' = \{i \in S : c_f(i) \neq c_g(i)\} \in \mathcal{U},$$

which implies that $[c_f] \neq [c_g]$. Hence $\{[c_f] : f \in \mathcal{F}\}$ is a subset of $\mathcal{M}$ of size $2^{\aleph_0}$, so $|\mathcal{M}| \geq 2^{\aleph_0}$. □

Corollary 1.23. *Let $\mathcal{U}$ be a non-principal ultrafilter on $\mathbb{N}$, and assume that $\mathcal{M}_i, i \in \mathbb{N}$ is a* ***countable*** *$\mathcal{L}$-structure. Then any ultraproduct $\mathcal{M} = \prod_{i \in \mathbb{N}} \mathcal{M}_i / \mathcal{U}$ is either finite or of size $2^{\aleph_0}$.*

Proof. Obviously $\left|\prod_{i \in \mathbb{N}} \mathcal{M}_i / \mathcal{U}\right| \leq \left|\prod_{i \in \mathbb{N}} \mathcal{M}_i\right| \leq \left|\mathbb{N}^{\mathbb{N}}\right| = 2^{\aleph_0}$, and $|\mathcal{M}| \geq 2^{\aleph_0}$ by Proposition 1.22. □

Example 1.24. Returning to Example 1.17, we now know that $\prod_{i \in \mathbb{N}} (\{0, 1, \ldots, i-1\}, <) / \mathcal{U}$ is a linear order of the form $\mathbb{N} + \sum_{j \in L} \mathbb{Z} + \mathbb{N}^*$, where L is a dense $\aleph_1$-saturated linear order of cardinality $2^{\aleph_0}$.

Exercise 1.25. For $i \in \mathbb{N}$ let $\mathcal{M}_i$ be a graph (undirected, without self-loops) which is a cycle on i vertices (i.e. $\mathcal{M}_i = (\{0, 1, \ldots, i-1\}, E)$ and the edges are $\{j, j+1\}$ for all $j = 0, \ldots, i-2$ and $\{i-1, 0\}$). Let $\mathcal{U}$ be a non-principal ultrafilter on $\mathbb{N}$. Determine $\prod_{i \in \mathbb{N}} \mathcal{M}_i / \mathcal{U}$ (up to isomorphism).

1.5. **References.** See e.g. [10] for a brief survey of further properties of the ultraproduct construction and references for the results in this section.

2. Graph regularity and measures on ultraproducts

2.1. **Szemerédi's regularity lemma.** Szemerédi's regularity lemma is a fundamental result in graph combinatorics with numerous applications in extremal combinatorics, additive number theory, computer science and other areas (see e.g. [11] for a survey). It has many versions and strengthenings, we begin by considering its simplest form.

Roughly speaking, the lemma asserts that every sufficiently large graph can be partitioned into a small number of sets, so that on almost all pairs of those sets the edges are approximately uniformly distributed at random.

More precisely, by a graph $G = (V, E)$ we mean a set G with a symmetric subset $E \subseteq V^2$. For $A, B \subseteq V$ we denote by $E(A, B)$ the set of edges between A and B and by $d_E(A, B) = \frac{|E(A,B)|}{|A||B|}$ the *density of the edges* between A and B. For $n \in \mathbb{N}$, we denote $[n] = \{1, 2, \ldots, n\}$.

Theorem 2.1. *(Szemerédi's regularity lemma) Let $\varepsilon > 0$ be arbitrary. Then there is some $K = K(\varepsilon) \in \mathbb{N}$ such that for every finite graph $G = (V, E)$ with $|V| \geq K$*

there is a partition $V = V_1 \sqcup \cdots \sqcup V_K$ into disjoint sets, real numbers $\delta_{ij}, i, j \in [K]$, and an exceptional set of pairs $\Sigma \subseteq [K] \times [K]$ such that

$$\sum_{(i,j)\in\Sigma} |V_i||V_j| \leq \varepsilon |V|^2$$

and for each $(i, j) \in [K] \times [K] \setminus \Sigma$ we have

$$| \, |E(A, B)| - \delta_{ij}|A||B| \, | < \varepsilon |V_i||V_j|$$

*for **all** $A \subseteq V_i$, $B \subseteq V_j$. We call a pair of sets (V_i, V_j) with $(i, j) \in [K] \times [K] \setminus \Sigma$ an ε-*regular pair.

Exercise 2.2.
(1) We can take $\delta_{ij} = d_E(V_i, V_j) = \frac{|E(V_i,V_j)|}{|V_i||V_j|}$ — the edge density between V_i and V_j (at the price of possibly doubling the error).
(2) The regularity condition can be rephrased as: $|d_E(A, B) - d_E(V_i, V_j)| < \varepsilon$ for all $A \subseteq V_i, B \subseteq V_j$ with $|A| \geq \varepsilon |V_i|, |B| \geq \varepsilon |V_j|$.
(3) Moreover, one can assume that all parts are of almost equal size, i.e. $||V_i| - |V_j|| \leq 1$ for all $i, j \in [K]$. In this case, we say that the partition $V = V_1 \sqcup \ldots \sqcup V_K$ is an *equipartition.*

Remark 2.3. Note that any sufficiently large graph has *some* ε-regular partition, e.g. into parts each of which consists of a single vertex. The crucial point of the theorem is that the size of the partition is bounded only in terms of ε, and independently of the size of G.

Remark 2.4. Regularity lemma doesn't say anything about what happens on the "diagonal" in V^2. Namely, given an ε-regular partition $V_1, \ldots, V_K$ of V, it is possible that all of the pairs on the diagonal $(V_i, V_i), 1 \leq i \leq K$ are exceptional simultaneously. Namely, if Σ is the collection of all bad pairs, we have that $\sum_{(i,j)\in\Sigma} |V_i| |V_j| < \varepsilon |V|^2$. On the other hand, if let's say $(V_i : 1 \leq i \leq K)$ is an equipartition, we have $\sum_{1\leq i\leq K} |V_i|^2 \leq K \frac{|V|^2}{K^2} \leq \frac{1}{K} |V|^2$, which can be smaller than $\varepsilon |V|^2$ when K is sufficiently large.

Exercise 2.5. A *half-graph* on n vertices is $G = (V, E)$ with $V = [n] = \{1, 2, \ldots, n\}$ such that $E = \left\{(i, j) \in [n]^2 : i < j\right\}$. Using half-graphs, show that in Theorem 2.1 one cannot assume in addition that $\Sigma = \emptyset$.

Next we are going to prove Theorem 2.1. Assume that the theorem is false. This means that for some $\varepsilon > 0$ we have a sequence of finite graphs $\mathcal{G}_i = (V_i, E_i)$, $i \in \mathbb{N}$, such that there is no ε-regular partition of V_i into at most i parts (in particular $|V_i| \to \infty$ by Remark 2.3). Let $\mathcal{G} := \prod_{i\in\mathbb{N}} \mathcal{G}_i/\mathcal{U}$, with $\mathcal{U}$ a non-principal ultrafilter on $\mathbb{N}$. We will see that regularity follows from basic measure theory applied to the "limit" of the counting measures on the V_i's.

2.2. **Finitely additive measures.** Let X be a set, and let $\mathcal{B}$ be a *Boolean algebra* of subsets of X, i.e. $\mathcal{B} \subseteq \mathcal{P}(X)$ is such that $\emptyset \in \mathcal{B}, X \in \mathcal{B}$ and if $A, B \in \mathcal{B}$ then $A \cap B \in \mathcal{B}$ and $\neg A \in \mathcal{B}$. Note that this also implies $A \cup B \in \mathcal{B}$.

Definition 2.6. A function $\mu : \mathcal{B} \to \mathbb{R}_{\geq 0}$ is a *finitely additive measure*, or *f.a. measure*, if for every $A, B \in \mathcal{B}$ such that $A \cap B = \emptyset$ we have $\mu(A \cup B) = \mu(A) + \mu(B)$.

Remark 2.7. This implies:

(1) For any *disjoint* $A_1, \ldots, A_n \in \mathcal{B}$, $\mu(A_1 \cup \ldots \cup A_n) = \mu(A_1) + \ldots + \mu(A_n)$.
(2) If $A, B \in \mathcal{B}$, $A \subseteq B$, then $\mu(A) \leq \mu(B)$.
(3) $\mu(\emptyset) = 0$.
(4) For any $A, B \in \mathcal{B}$, $\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B)$.

Definition 2.8. A *finitely additive probability measure*, or *f.a.p. measure*, on $\mathcal{B}$ is f.a. measure on $\mathcal{B}$ such that moreover $\mu(X) = 1$.

Example 2.9. (1) Let X be a *finite* set. The *counting measure* μ on $\mathcal{P}(X)$ is defined by $\mu(Y) = \frac{|Y|}{|X|}$ for all $Y \subseteq X$. Then μ is a f.a.p. measure on $\mathcal{P}(X)$.
(2) Let $\mathcal{U}$ be an ultrafilter on a set X. It may be naturally identified with a f.a.p. measure on the Boolean algebra $\mathcal{P}(X)$ taking values in $\{0,1\}$. Namely, for $Y \subseteq X$, we define $\mu_{\mathcal{U}}(Y) = 1$ if $Y \in \mathcal{U}$, and $\mu_{\mathcal{U}}(Y) = 0$ if $Y \notin \mathcal{U}$. It is easy to check that $\mu_{\mathcal{U}}$ is a f.a.p. measure on $\mathcal{P}$. Conversely, for every f.a.p. measure μ on $\mathcal{P}(X)$ with values in $\{0,1\}$, the set $\{Y \subseteq X : \mu(Y) = 1\}$ is an ultrafilter.

We saw that one can extend ultrafilters using the axiom of choice. The same applies to general f.a.p. measures.

Fact 2.10. *(see e.g.* [12]*) Let X be a set and $\mathcal{B} \subseteq \mathcal{B}' \subseteq \mathcal{P}(X)$ be Boolean algebras. Let μ be a f.a.p. measure on $\mathcal{B}$. Then there is a f.a.p. measure μ' on $\mathcal{B}'$ extending μ. Moreover, for any $S \in \mathcal{B}'$ we can choose μ' with $\mu'(S) = r$ for any r satisfying*

$$\sup\{\mu(A) : A \in \mathcal{B}, A \subseteq S\} \leq r \leq \inf\{\mu(B) : B \in \mathcal{B}, S \subseteq B\}.$$

Another example is given by the limit f.a.p. measure on an ultraproduct of structures each of which is equipped with a f.a.p. measure.

Definition 2.11. Assume we have a fixed sequence of sets V_i, $i \in \mathbb{N}$. For each i, let $\mathcal{B}_i$ be a Boolean algebra of subsets of V_i. Let $\mathcal{U}$ be a non-principal ultrafilter on $\mathbb{N}$, and let $V := \prod_{i \in \mathbb{N}} V_i / \mathcal{U}$.
(1) We call a set $A \subseteq V$ *internal relatively to the $\mathcal{B}_i$'s* if $A = \prod_{i \in \mathbb{N}} A_i / \mathcal{U}$ for some $A_i \in \mathcal{B}_i$ (i.e. $[a] \in X \iff \{i \in \mathbb{N} : a(i) \in A_i\} \in \mathcal{U}$).
(2) We say simply that A is *internal* if it is internal relatively to the Boolean algebras $\mathcal{P}(V_i)$, $i \in \mathbb{N}$.
(3) Let $\mathcal{B}$ be the collection of all subsets of V internal relatively to the $\mathcal{B}_i$'s. It is a Boolean algebra of subsets of V (e.g. by Łoś theorem).

Exercise 2.12. Recall the definition of ultralimit from Definition 1.6. Let (X, d) and (Y, d') be metric spaces, and assume that $f : X \to Y$ is continuous. Then for any sequence $(a_i)_{i \in \mathbb{N}}$ from X and any non-principal ultrafilter $\mathcal{U}$ on $\mathbb{N}$, we have

$$\lim_{\mathcal{U}} a_i = a \implies \lim_{\mathcal{U}} f(a_i) = f(a).$$

Definition 2.13. In the context of Definition 2.11, assume also that μ_i is an f.a.p. measure on $\mathcal{B}_i$, for all $i \in \mathbb{N}$. For any set $A \in \mathcal{B}$, say $A = \prod_{i \in \mathbb{N}} A_i / \mathcal{U}$, define $\mu(A) = \lim_{\mathcal{U}} \mu_i(A_i)$ (ultralimit exists as μ_i take values in $[0,1]$). Then $\mu(X)$ is a f.a.p. measure on $\mathcal{B}$.

(Exercise: check that this is well-defined, i.e. doesn't depend on the choice of the A_i's as above).

Proof. Note that if $a_i, b_i \in [0,1]$, then $\lim_{\mathcal{U}} (a_i + b_i) = \lim_{\mathcal{U}} a_i + \lim_{\mathcal{U}} b_i$ (by Exercise 2.12 applied to $X = [0,1]^2$ and $Y = [0,2]$).

Let now $A = \prod_{i \in \mathbb{N}} A_i/\mathcal{U}, B = \prod_{i \in \mathbb{N}} B_i/\mathcal{U}$ in $\mathcal{B}$ be disjoint. Then there is some $S \in \mathcal{U}$ such that $A_i \cap B_i = \emptyset$ for all $i \in S$. Then for all $i \in S$, we have $\mu_i (A_i \cup B_i) = \mu_i (A_i) + \mu_i (B_i)$. Note that $A \cup B = \prod_{i \in \mathbb{N}} (A_i \cup B_i)/\mathcal{U}$, hence $\mu (A \cup B) = \lim_{\mathcal{U}} \mu_i (A_i \cup B_i) = \lim_{\mathcal{U}} (\mu_i (A_i) + \mu_i (B_i)) = \lim_{\mathcal{U}} \mu_i (A_i) + \lim_{\mathcal{U}} \mu_i (B_i) = \mu (A) + \mu (B)$, as wanted. □

2.3. **Obtaining countable additivity.** We would like to apply some basic theory of integration. Normally it is developed in the context of countably additive measures, rather than f.a.p. measures. We will in fact use the theory of integration for f.a.p. measures (see e.g. [15]), but first we point out how countable additivity can be obtained for free in our setting (the so-called Loeb measure construction).

Definition 2.14. Let X be a set. We say that $\mathcal{E} \subseteq \mathcal{P}(X)$ is a *σ-algebra* on X if $\emptyset \in \mathcal{E}$, $A \in \mathcal{E} \implies \neg A \in \mathcal{E}$, and $A_i \in \mathcal{E}$ for all $i \in \mathbb{N} \implies A = \bigcup_{i \in \mathbb{N}} A_i \in \mathcal{E}$.

This implies: $X \in \mathcal{E}$ and $\mathcal{E}$ is closed under countable intersections. For any $\mathcal{F} \subseteq \mathcal{P}(X)$, there exists a unique smallest (under inclusion) σ-algebra $\sigma\mathcal{F}$ on X with $\mathcal{F} \subseteq \sigma\mathcal{F}$. We call $\sigma\mathcal{F}$ the *σ-algebra generated by* $\mathcal{F}$.

Fact 2.15. *(Carathéodory's extension theorem) Let $\mathcal{B}$ be a Boolean algebra on a set X, and assume that μ is a σ-additive measure defined on $\mathcal{B}$. Then μ extends to the σ-algebra $\sigma\mathcal{B}$ generated by $\mathcal{B}$. Furthermore, if μ is σ-finite (e.g. a probability measure), then this extension is unique.*

Proposition 2.16. *Let $\mathcal{M}$ be an $\aleph_1$-saturated structure, and let $\mathcal{B}$ be a Boolean algebra of* definable *subsets of M^n (with parameters). Let μ be an f.a.p. measure on $\mathcal{B}$. Then it extends in a unique way to a countably additive probability measure μ' on the σ-algebra $\sigma\mathcal{B}$ generated by $\mathcal{B}$.*

Proof. In view of the Carathéodory's theorem, it is enough to check that μ is already σ-additive on $\mathcal{B}$. So assume that $X \in \mathcal{B}$ is a definable set, and assume $X = \bigsqcup_{i \in \mathbb{N}} X_i$ with $X_i \in \mathcal{B}$ definable. We want to show that $\mu (X) = \sum_{i \in \mathbb{N}} \mu (X_i)$. Assume that $X \supsetneq \bigcup_{i<n} X_i$ for all $n \in \mathbb{N}$. But then every finite subset of $\{X\} \cup \{\neg X_i : i \in \mathbb{N}\}$ has a non-empty intersection, so by saturation of $\mathcal{M}$ we must have that $X \cap \bigcap_{i \in \mathbb{N}} \neg X_i \neq \emptyset$ — contradicting the assumption. It follows that $X = \bigsqcup_{i<n} X_i$ for some $n \in \mathbb{N}$, and $X_i = \emptyset$ for $i \geq n$. The conclusion follows from the finite additivity of μ. □

Corollary 2.17. *Let $\mathcal{M}_i, i \in \mathbb{N}$ be $\mathcal{L}$-structures in a countable language $\mathcal{L}$, and let $\mathcal{U}$ be a non-principal ultrafilter on $\mathbb{N}$. Let $n \in \mathbb{N}$ be fixed, and let $\mathcal{B}_i$ be the Boolean algebra of all $\mathcal{L}$-definable subsets of M_i^n. Let μ_i be an f.a.p. measure on $\mathcal{B}_i$, and let μ be the ultralimit measure on $\mathcal{B}$ — the Boolean algebra of all $\mathcal{L}$-definable subsets of M^n. Then μ has a unique extension to a σ-additive measure on the σ-algebra $\sigma\mathcal{B}$.*

Proof. Combine Proposition 2.16 and $\aleph_1$-saturation of the ultraproduct $\mathcal{M}$. □

Exercise 2.18. Let V_i be a sequence of finite sets, and let $V := \prod_{i \in \mathbb{N}} V_i/\mathcal{U}$. Let $\mathcal{B}$ be the Boolean algebra of *all* internal subsets of V (See Definition 2.11(2)). Let μ_i be the counting measure on $\mathcal{P}(V_i)$ Show that the ultralimit of the μ_i's extends to a σ-additive measure on $\sigma\mathcal{B}$.

2.4. Integration for charges (signed f.a. measures).

Definition 2.19. Let $\mathcal{B}$ be a Boolean algebra on a set V. A f.a. *charge* (or a *signed f.a. measure*) μ on $\mathcal{B}$ is a f.a. bounded function $\mu : \mathcal{B} \to \mathbb{R}$.

Hence a f.a. measure is a f.a. charge taking only positive values. The set of all f.a. charges on $\mathcal{B}$ forms a vector space over $\mathbb{R}$.

Definition 2.20. Let $\mathcal{B}_U, \mathcal{B}_V$ be Boolean algebras on the sets U, V, respectively.

(1) Let $\mathcal{B}_U \times \mathcal{B}_V := \{A \times B : A \in \mathcal{B}_U, B \in \mathcal{B}_V\} \subseteq \mathcal{P}(U \times V)$, and let $\mathcal{B}_U \otimes \mathcal{B}_V \subseteq \mathcal{P}(U \times V)$ denote the Boolean algebra generated by $\mathcal{B}_U \times \mathcal{B}_V$.
Note: $X \subseteq \mathcal{B}_U \otimes \mathcal{B}_V$ iff X can be written as a finite (disjoint) union of sets from $\mathcal{B}_U \times \mathcal{B}_V$.

(2) Let μ_U, μ_V be f.a. charges on $\mathcal{B}_U, \mathcal{B}_V$, respectively. Then there is a unique f.a. charge μ on $\mathcal{B}_U \otimes \mathcal{B}_V$ with $\mu(A \times B) = \mu_U(A)\mu_V(B)$ for all $A \in \mathcal{B}_U, B \in \mathcal{B}_V$ (uniqueness follows from finite additivity). We will denote this μ by $\mu_U \times \mu_V$, the product measure on $\mathcal{B}_U \otimes \mathcal{B}_V$.
Note: if both μ_U, μ_V are f.a. (f.a.p.) measures then μ is f.a. (f.a.p.) measure.

Definition 2.21. For an f.a. charge μ, define $\mu^+, \mu^-, |\mu| : \mathcal{B} \to \mathbb{R}$ by

$$\mu^+(X) := \sup\{\mu(Y) : Y \subseteq X, Y \in \mathcal{B}\},$$
$$\mu^-(X) = -\inf\{\mu(Y) : Y \subseteq X, Y \in \mathcal{B}\},$$
$$|\mu|(X) := \mu^+(X) + \mu^-(X)$$

for all $X \in \mathcal{B}$.

Fact 2.22. [15, Theorems 2.2.1 and 2.2.2]

(1) *All of $\mu^+, \mu^-, |\mu|$ are f.a. measures on $\mathcal{B}$, and $\mu = \mu^+ - \mu^-$ and $|\mu| = \mu^+ + \mu^-$.*

(2) *Let μ be an f.a. charge on $\mathcal{B}$. Then for every $X \in \mathcal{B}$ we have*

$$|\mu|(X) = \sup \sum_{Y \in \mathcal{Q}} |\mu(Y)|,$$

where sup *is taken over all finite partitions Q of X with $\mathcal{Q} \subseteq \mathcal{B}$.*

Definition 2.23. For a f.a. charge μ on $\mathcal{B} \subseteq \mathcal{P}(V)$, define $\|\mu\| = |\mu|(V)$.

Exercise 2.24. [15, Theorems 2.2.1 and 2.2.2] $\|\cdot\|$ is a norm on the vector space of f.a. charges on $\mathcal{B}$.

We will use basic theory of integration relatively to f.a. charges.

Fix a set V and a Boolean algebra $\mathcal{B} \subseteq \mathcal{P}(V)$. For a set $X \subseteq V$, $\mathbf{1}_X$ is the indicator function, i.e. $\mathbf{1}_X(a) = 1$ if $a \in X$ and $\mathbf{1}_X(a) = 0$ if $a \notin X$.

Definition 2.25. A function $f : V \to \mathbb{R}$ is *$\mathcal{B}$-simple* (or just *simple* if there is no ambiguity) if

$$f = \sum_{i=1}^{n} r_i \mathbf{1}_{A_i}$$

for some $r_1, \ldots, r_n \in \mathbb{R}$ and $A_1, \ldots, A_n \in \mathcal{B}$.

One may always choose disjoint $A_1, \ldots, A_n$ as above. The set of all $\mathcal{B}$-simple functions forms an $\mathbb{R}$-algebra.

Definition 2.26. For a f.a. charge μ on $\mathcal{B}$ and a simple function $f = \sum_{i=1}^{n} r_i \mathbf{1}_{A_i}$ we define

$$\int_{\Omega} f d\mu := \sum_{i=1}^{n} r_i \mu(A_i).$$

(Exercise: this definition doesn't depend on the specific representation of f as a simple function.)

If $A \subseteq V$, $A \in \mathcal{B}$, then we also define

$$\int_{A} f d\mu := \int_{V} \mathbf{1}_A f d\mu = \sum_{i=1}^{n} r_i \mu(A \cap X_i).$$

Note: for any $A \in \mathcal{B}$, $\mu(A) = \int_V \mathbf{1}_A d\mu$.

Definition 2.27. Let f be a $\mathcal{B}$-simple function. Then the function $\mathcal{B} \to \mathbb{R}$ defined by $A \mapsto \int_A f d\mu$ is a f.a. charge on $\mathcal{B}$. We will denote it by $f d\mu$.

We will need a version of the Radon-Nikodym theorem for f.a. measures. As before, $\mathcal{B}$ is a Boolean algebra on V.

Definition 2.28. Let μ, ν be f.a. charges on $\mathcal{B}$. We say that ν is *absolutely continuous with respect to* μ, and write $\nu \ll \mu$ if for every $\varepsilon > 0$ there is $\delta > 0$ such that $|\mu|(X) < \delta$ implies $|\nu|(X) < \varepsilon$ for every $X \in \mathcal{B}$.

Theorem 2.29. *(Radon-Nikodym for f.a. measures, see* [3]*, or* [15, Theorem 6.3.4]*) Let μ, ν be f.a. charges on $\mathcal{B}$ with $\nu \ll \mu$. Then for every $\varepsilon > 0$ there is a simple function f_ε with $\|\nu - f_\varepsilon d\mu\| < \varepsilon$.*

For f.a. charges μ, ν on $\mathcal{B}$, write $\mu \leq \nu$ if $\mu(X) \leq \nu(X)$ for all $X \in \mathcal{B}$.

Corollary 2.30. *Let μ be a f.a.p. measure on $\mathcal{B}$ and ν a f.a. measure on $\mathcal{B}$ with $\nu \leq \mu$. Then for every ε there is a simple function f_ε such that $\|\nu - f_\varepsilon d\mu\| < \varepsilon$.*

Remark 2.31. Assuming σ-additivity, one finds a $\sigma\mathcal{B}$-measurable function f such that $\nu(A) = \int_A f d\mu$ for all $A \in \sigma\mathcal{B}$ (and this function f can be approximated by simple functions). Moreover, such function f is unique (up to differences on sets of measure zero) and is called the Radon-Nikodym derivative.

2.5. **Measure-theoretic regularity.** We are ready to prove a measure-theoretic (bipartite) form of regularity.

Theorem 2.32. *Let $\mathcal{B}_U, \mathcal{B}_V$ be Boolean algebras on U, V, resp. Let μ_U, μ_V be f.a.p. measures on $\mathcal{B}_U, \mathcal{B}_V$, resp. Let $\mathcal{B}$ be an* ***arbitrary*** *Boolean algebra on $U \times V$ extending $\mathcal{B}_U \otimes \mathcal{B}_V$, and μ a f.a.p. measure on $\mathcal{B}$ extending $\mu_U \times \mu_V$.*

Assume that $E \in \mathcal{B}$. Then for any $\varepsilon > 0$ there are:

1. *a partition $U = U_1 \sqcup \ldots \sqcup U_m$ with $U_i \in \mathcal{B}_U$,*
2. *a partition $V = V_1 \sqcup \ldots \sqcup V_n$ with $V_i \in \mathcal{B}_V$,*
3. *real numbers $\delta_{ij} \in [0,1]$, for $1 \leq i \leq m, 1 \leq j \leq n$,*
4. *an exceptional set of pairs $\Sigma \subseteq [n] \times [m]$*

such that

1. $\sum_{(i,j) \in \Sigma} \mu_U(U_i) \mu_V(V_j) < \varepsilon$,
2. *for every $(i,j) \notin \Sigma$, for any $A \in \mathcal{B}_U, B \in \mathcal{B}_V$ with $A \subseteq U_i, B \subseteq V_j$ we have*
$$|\mu(E \cap (A \times B)) - \delta_{ij} \mu_U(A) \mu_V(B)| < \varepsilon \mu_U(U_i) \mu_V(V_j).$$

Proof. Let $\mathcal{B}_{UV} := \mathcal{B}_U \otimes \mathcal{B}_V$ and $\mu_{UV} := \mu_U \times \mu_V$ — a f.a.p. measure on $\mathcal{B}_{UV}$.

Let $\nu_E : \mathcal{B}_{UV} \to [0,1]$ be defined as $\nu_E(X) := \mu(E \cap X)$ for all $X \in \mathcal{B}_{UV}$. Then ν_E is a f.a. measure on Σ_{UV} with $\nu_E \leq \mu_{UV}$.

By Radon–Nikodym (Corollary 2.30) there is a $\mathcal{B}_{UV}$-simple function f such that $\|\nu_E - f d\mu_{UV}\| < \varepsilon^2$.

As f is simple, there are some partitions $U = U_1 \sqcup \ldots \sqcup U_m$ with $U_i \in \mathcal{B}_U$ and $V = V_1 \sqcup \ldots \sqcup V_n$ with $V_i \in \mathcal{B}_V$, and $\delta_{ij} \in [0,1]$ such that $f = \sum_{(i,j) \in [m] \times [n]} \delta_{ij} \mathbf{1}_{U_i \times V_j}$.

Let Σ be the set of all $(i,j) \in [m] \times [n]$ such that

$$|\nu_E - f d\mu_{UV}| (U_i \times V_j) \geq \varepsilon \mu_{UV} (U_i \times V_j).$$

Since $|\nu_E - f d\mu_{UV}|$ is a f.a. measure on $\mathcal{B}_{UV}$,

$$\varepsilon^2 > |\nu_E - f d\mu_{UV}| (U \times V) \geq \sum_{(i,j) \in \Sigma} \varepsilon \mu_{UV} (U_i \times V_j) = \varepsilon \sum_{(i,j) \in \Sigma} \mu_{UV} (U_i \times V_j).$$

Hence $\sum_{(i,j) \in \Sigma} \mu_{UV} (U_i \times V_j) < \varepsilon$, and conclusion (1) is satisfied.

Let's show (2). Assume $(i,j) \notin \Sigma$, hence

$$|\nu_E - f d\mu_{UV}| (U_i \times V_j) < \varepsilon \mu_U (U_i) \mu_V (V_j).$$

Let $A \in \mathcal{B}_U, B \in \mathcal{B}_V$ with $A \subseteq U_i, B \subseteq V_j$ be arbitrary. Then:

$$|\mu (E \cap (A \times B)) - \delta_{ij} \mu_{UV} (A \times B)| = |\nu_E (A \times B) - f d\mu_{UV} (A \times B)| \leq |\nu_E - f d\mu_{UV}| (A \times B).$$

Since $|\nu_E - f d\mu_{UV}|$ is a f.a. measure and $A \times B \subseteq U_i \times V_j$, we have $|\nu_E - f d\mu_{UV}| (A \times B) \leq |\nu_E - f d\mu_{UV}| (U_i \times V_j)$ — as wanted. □

Exercise 2.33. Give a variant of this proof using σ-additive measures and a standard version of the Radon–Nikodym theorem. (Hint: define a first-order structure $\mathcal{M}$ with two sort U, V in which all elements of $\mathcal{B}_U$, $\mathcal{B}_V$ and $\mathcal{B}$ are named by a predicate. Every structure has an $\aleph_1$-saturated elementary extension — without loss of generality can work in it).

Corollary 2.34. *Szemerédi's regularity lemma for finite graphs, i.e. Theorem 2.1, holds.*

Proof. Assume it doesn't hold. This means that for some fixed $\varepsilon > 0$ we have a sequence of finite graphs $\mathcal{G}_i = (V_i, E_i)$, $i \in \mathbb{N}$, such that there is no ε-regular partition of V_i into at most i parts (in particular $|V_i| \to \infty$ by Remark 2.3). Let $\mathcal{G} := \prod_{i \in \mathbb{N}} \mathcal{G}_i / \mathcal{U}$, with $\mathcal{U}$ a non-principal ultrafilter on $\mathbb{N}$, write $\mathcal{G} = (V, E)$.

Let $\mathcal{B}_i = \mathcal{P}(V_i)$, and let $\mathcal{B}$ be the Boolean algebra of all internal subsets of V.

Let $\mathcal{B}'_i = \mathcal{P}(V^2)$, and let $\mathcal{B}'$ be the Boolean algebra of all internal subsets of V^2.

Finally, let μ_i be the counting measure on V_i, let μ'_i be the counting measure on V_i^2. Then $\mu = \lim_{\mathcal{U}} \mu_i$ is an f.a.p. measure on $\mathcal{B}$, $\mu' = \lim_{\mathcal{U}} \mu'_i$ is an f.a.p. measure on $\mathcal{B}'$, $\mathcal{B}' \supseteq \mathcal{B} \otimes \mathcal{B}$, and μ' is extending $\mu \times \mu$. Moreover, $E \in \mathcal{B}'$.

Applying Theorem 2.32, we obtain an $\frac{\varepsilon}{2}$-regular (relatively to μ) *finite* partition of V into *internal* subsets. But then on a $\mathcal{U}$-large set of indices $i \in \mathbb{N}$ this gives an ε-regular partition of V_i into the same fixed number of pieces — contradicting the choice of the sequence $\mathcal{G}_i$. □

Exercise 2.35. Using the same ultraproduct argument, demonstrate that in Theorem 2.32 a bound on the size of the partition n, m can be chosen depending **only on** ε (so uniformly over *all* Boolean algebras and *all* measures).

2.6. **References.** *** TBA. The use of the finitely additive Radon-Nikodym arose from my work with Sergei Starchenko.

3. Hypergraph removal

3.1. **Removal lemmas.** We first consider the more standard triangle removal for graphs.

Fact 3.1. *(Triangle removal lemma, Ruzsa and Szemerédi) For every $\varepsilon > 0$ there is $\delta > 0$ satisfying the following. If G is a finite graph on n vertices with at most δn^3 triangles, then it may be made triangle-free by removing at most εn^2 edges.*

Proof. We will deduce it from the regularity lemma (Theorem 2.1).

Let $G = (V, E)$ with $|V| = n$. Let $V = V_1 \sqcup \ldots \sqcup V_K$ be an $\frac{\varepsilon}{4}$-regular partition of the vertices of G, where $K = K(\varepsilon)$, i.e.

- $\sum_{(i,j)\in\Sigma} |V_i| |V_j| \leq \frac{\varepsilon}{4} n^2$,
- $|d_E(A, B) - d_E(V_i, V_j)| < \frac{\varepsilon}{4}$ for all $A \subseteq V_i, B \subseteq V_j$ with $|A| \geq \frac{\varepsilon}{4} |V_i|, |B| \geq \frac{\varepsilon}{4} |V_j|$ (see Exercise 2.2).

We remove an edge xy from G if:

(1) $(x, y) \in V_i \times V_j$, where (V_i, V_j) is not an $\frac{\varepsilon}{4}$-regular pair,
(2) $(x, y) \in V_i \times V_j$, where $d_E(V_i, V_j) < \frac{\varepsilon}{2}$,
(3) $x \in V_i$, where $|V_i| \leq \frac{\varepsilon}{4K} n$.

The number of the edges removed in (1) is at most $\sum_{(i,j)\in\Sigma} |V_i| |V_j| \leq \frac{\varepsilon}{4} n^2$, in (2) — clearly at most $\frac{\varepsilon}{2} n^2$, and (3) — at most $Kn \frac{\varepsilon}{4K} n = \frac{\varepsilon}{4} n^2$. Overall, we have removed at most εn^2 edges.

Suppose that some triangle remains in the graph, say xyz, where $x \in V_i, y \in V_j$ and $z \in V_k$. Then the pairs $(V_i, V_j), (V_j, V_k)$ and (V_k, V_i) are all $\frac{\varepsilon}{4}$-regular with density at least $\frac{\varepsilon}{2}$, and $|V_i|, |V_j|, |V_k| \geq \frac{\varepsilon}{4K} n$.

Lemma. Let X, Y, Z be subsets of V such that $(X, Y), (Y, Z), (Z, X)$ are ε-regular with $d(X, Y) = \alpha, d(Y, Z) = \beta, d(Z, X) = \gamma$. Then, provided $\alpha, \beta, \gamma \geq 2\varepsilon$, the number of triangles xyz with $x \in X, y \in Y, z \in Z$ is at least

$$(1 - 2\varepsilon)(\alpha - \varepsilon)(\beta - \varepsilon)(\gamma - \varepsilon) |X| |Y| |Z| .$$

Proof of the lemma. For every $x \in X$, let $d_Y(x)$ and $d_Z(x)$ be the number of neighbors of x in Y and Z, resp.

Let $X' := \{x \in X : d_Y(x) < (\alpha - \varepsilon) |Y|\}$. Then $|X'| \leq \varepsilon |X|$ (if not then $X' \subseteq X$ is of size at least $\varepsilon |X|$ and such that $d_E(X', Y) < \alpha - \varepsilon$ — contradicting regularity).

Let $X'' := \{x : d_Z(x) < (\gamma - \varepsilon) |Z|\}$. Similarly, $|X''| \leq \varepsilon |X|$.

If $d_Y(x) > (\alpha - \varepsilon) |Y|$ and $d_Z(x) > (\gamma - \varepsilon) |Z|$, using that the pair (Y, Z) is ε-regular with density β, the number of edges between $N(x) \cap Y$ and $N(x) \cap Z$ is at least $(\alpha - \varepsilon)(\beta - \varepsilon)(\gamma - \varepsilon) |Y| |Z|$ (hence there are at least as many triangles containing x).

Summing over all $x \in X \setminus (X' \cup X'')$ gives the result.

Applying the lemma to our situation, the number of triangles in G is at least $\left(1 - \frac{\varepsilon}{2}\right) \left(\frac{\varepsilon}{4}\right)^3 \left(\frac{\varepsilon}{4K}\right)^3 n^3$. Taking $\delta = \left(1 - \frac{\varepsilon}{2}\right) \left(\frac{\varepsilon}{4}\right)^3 \left(\frac{\varepsilon}{4K}\right)^3 > 0$ gives a contradiction. □

More recently, this was generalized to hypergraphs.

Definition 3.2. A *k-uniform hypergraph* G on a set of vertices V is any subset $G \subseteq \binom{V}{d}$ of $\binom{V}{d}$.

Theorem 3.3. *(Hypergraph removal lemma, [Gowers] and [Nagle, Rödl, Schacht and Skokan]) For each $k \in \mathbb{N}$, $\varepsilon > 0$ and a finite k-uniform hypergraph (W, F) there is some $\delta > 0$ such that: whenever (V, E) is a k-uniform hypergraph containing at most $\delta |V|^{|W|}$ copies of (W, F), it is possible to remove at most $\varepsilon |V|^k$ edges from it to obtain a hypergraph with no copies of (W, F) at all.*

Again, we will convert it into a more general measure-theoretic statement.

3.2. **Measure-theoretic hypergraph removal.** We introduce some notation.

Fix some sets $V_1, \ldots, V_n$. For every $I \subseteq [n]$, let $V_I = \prod_{i \in I} V_i$. We will write a_I, b_I, c_I, etc. for elements in V_I. Given $a_I \in V_I$ and $J \subseteq I$, we will write $a_J \in V_J$ for the subtuple of a_I given by restricting to the coordinates in J. For any $J \subseteq I \subseteq [n]$, $E \subseteq V_I$ and $b \in V_J$, we write $E_b := \left\{a \in V_{I \setminus J} : (a, b) \in E\right\} \subseteq V_{I \setminus J}$.

Definition 3.4. For every $I \subseteq [n]$, let $\mathcal{B}_I$ be a Boolean algebra of subsets of V_I, such that:

1. for any $I, J \subseteq [n]$ with $I \cap J = \emptyset$, we have $\mathcal{B}_I \otimes \mathcal{B}_J \subseteq \mathcal{B}_{I \cup J}$,
2. for any $I, J \subseteq [n]$ with $I \cap J = \emptyset$, $b \in V_J$ and $E \in \mathcal{B}_{I \cup J}$, the fiber $E_b = \{a \in V_I : (a, b) \in E\}$ is in $\mathcal{B}_I$.

Then we call $(\mathcal{B}_I : I \subseteq [n])$ a *compatible system of b.a.'s* on $(V_i : i \in [n])$.

Example 3.5.
1. Fix a first-order structure $\mathcal{M}$. Fix $n \in \mathbb{N}$, and for each $I \subseteq [n]$ let $\mathcal{B}_I$ be the b.a. of all definable subsets of $M^{|I|}$. Then $(\mathcal{B}_I : I \subseteq [n])$ is a compatible system of b.a.'s on $V_1 = \ldots = V_n = M$.
2. Let $W = \prod_{i \in \mathbb{N}} W_i / \mathcal{U}$, fix n and for $I \subseteq [n]$ let $\mathcal{B}_I$ be the b.a. of all internal subsets of $W^{|I|}$. Then $(\mathcal{B}_I : I \subseteq [n])$ is a compatible system of b.a.'s on $V_1 = \ldots = V_n = W$.

Definition 3.6. (in a compatible system of b.a.'s)

1. For $J \subseteq I \subseteq [n]$, let $\mathcal{B}_{I,J}$ be the b.a. on V_I generated by the sets of the form $\{a_I \in V_I : a_J \in E\}$ for all $E \in \mathcal{B}_J$.
2. If $\mathcal{J} \subseteq \mathcal{P}(I)$, let $\mathcal{B}_{I,\mathcal{J}}$ be the Boolean algebra generated by $\bigcup_{J \in \mathcal{J}} \mathcal{B}_{I,J}$. When $k \leq |I|$, let $\mathcal{B}_{I,k} := \mathcal{B}_{I,\{J \subseteq I : |J| = k\}}$.
3. We write $< I$ for the set of all proper subsets of I, so e.g. $\mathcal{B}_{I,<I} = \mathcal{B}_{I,\{J \subseteq I : J \subsetneq I\}}$.
4. Given $\mathcal{I}, \mathcal{J} \subseteq [n]$, let $\mathcal{I} \wedge \mathcal{J} := \{K : \exists I \in \mathcal{I}, J \in \mathcal{J} \text{ s.t. } K \subseteq I \cap J\}$.
5. We add a superscript $\mathcal{B}^\sigma$ to denote the σ-algebra generated by the b.a. $\mathcal{B}$.

Definition 3.7. Let $(\mathcal{B}_I : I \subseteq [n])$ be a compatible system of b.a.'s on $(V_i : i \in [n])$.

For each $I \subseteq [n]$, let μ_I be a probability measure on $\mathcal{B}_I^\sigma$. Assume moreover that for any $J \subseteq I \subseteq [n]$ we have:

1. μ_I extends the product measure $\mu_J \times \mu_{I \setminus J}$,
2. For each $\mathcal{B}_I^\sigma$-measurable function $f : V_I \to \mathbb{R}$, the function $b \mapsto \int_{V_{I \setminus J}} f\left(x_{I \setminus J}, b\right) d\mu_{I \setminus J}$ from V_J to $\mathbb{R}$ is $\mathcal{B}_J^\sigma$-measurable,
3. (Fubini) For each $\mathcal{B}_I^\sigma$-measurable function $f : V_I \to \mathbb{R}$ and $J \subseteq I$, we have
$$\int_{V_I} f d\mu_I = \int_{V_J} \left(\int_{V_{I \setminus J}} f\left(x_{I \setminus J}, b_J\right) d\mu_{I \setminus J} \right) d\mu_J\left(b_J\right).$$

Then we call $(\mu_I, \mathcal{B}_I : I \subseteq [n])$ a *compatible system of measures* on $(V_i : i \in [n])$.

Remark 3.8.
(1) Note that applying (3) to $I \setminus J$ instead of J, the order of integration in (3) doesn't matter.
(2) In particular we have: for any $E \in \mathcal{B}_I^\sigma$, we have $\mu_I(E) = \int_{V_J} \mu_{I\setminus J}(E_x)\, d\mu_J(x) = \int_{V_{I\setminus J}} \mu_I(E_y)\, d\mu_{I\setminus J}(y)$.

Problem 3.9. Can we recover full (2) and (3) from assuming it only for the indicator functions? I.e., assuming

- For each $E \in \mathcal{B}_I$ and $b \in V_J$, the function $b \mapsto \mu_{I\setminus J}(E_b)$ from V_J to $\mathbb{R}$ is $\mathcal{B}_J$-measurable,
- For any $E \in \mathcal{B}_I$, we have $\mu_I(E) = \int_{V_J} \mu_{I\setminus J}(E_x)\, d\mu_J(x) = \int_{V_{I\setminus J}} \mu_I(E_y)\, d\mu_{I\setminus J}(y)$.

Example 3.10.
(1) Let $V_1, \ldots, V_n$ be finite sets, fix n. For each $I \subseteq [n]$, let $\mathcal{B}_I := \mathcal{P}(V_I)$ and let μ_I be the counting measure on $\mathcal{P}(V_I)$ (i.e. $\mu_I(X) = \frac{|X|}{|V_I|}$ for all $X \subseteq V_I$). Note that $\mathcal{B}_I = \mathcal{B}_I^\sigma$. Then $(\mu_I, \mathcal{B}_I : I \subseteq [n])$ is a compatible system of measures.
(2) In the context of Example 3.5(2), let $\mu_I = \lim_{\mathcal{U}} \mu_i^{|I|}$, where μ_i^k is the counting measure on V_i^k for all $k \in \mathbb{N}$.
Then $(\mu_I, \mathcal{B}_I : I \subseteq [n])$ is a compatible system of measures (**Exercise!** Note that it is obviously satisfied by the counting measures on finite sets, and verify that it transfers to the ultralimit).

Remark 3.11. If $E \in \mathcal{B}_{[n],I}^\sigma$ then $\mu_{[n]}(E) = \mu_I(\pi_I(E))$, where $\pi_I(E) \subseteq V_I$ is the projection of E onto V_I. Indeed, as $E \in \mathcal{B}_{[n],I}^\sigma$, $E = \pi_I(E) \times V_{[n]\setminus I}$, hence by compatibility $\mu_{[n]}(E) = \mu_I(\pi_I(E))\, \mu_{[n]\setminus I}\left(V_{[n]\setminus I}\right) = \mu_I(\pi_I(E))$.

Theorem 3.12. *(Hypergraph removal lemma, measure-theoretic version)*
Let $(\mu_I, \mathcal{B}_I : I \subseteq [n])$ be a compatible system of measures on $(V_i : i \in [n])$. Let $\mathcal{I} \subseteq \binom{[n]}{k}$ and $A_I \in \mathcal{B}_{[n],I}$ for all $I \in \mathcal{I}$.

Suppose there is $\delta > 0$ such that for any $B_I \in \mathcal{B}_{[n],I}(M)$ with $\mu_{[n]}(A_I \setminus B_I) < \delta$ for all $I \in \mathcal{I}$, $\bigcap_{I\in\mathcal{I}} B_I \neq \emptyset$. Then $\mu_{[n]}\left(\bigcap_{I\in\mathcal{I}} A_I\right) > 0$.

Proof. By induction on k.

Base step $k = 1$. If the assumption holds, then necessarily $\mu_{[n]}(A_I) > 0$ for all $I \in \mathcal{I}$ (assume that $\mu_{[n]}(A_{I_0}) = 0$ for some $I_0 \in \mathcal{I}$; taking $B_{I_0} = \emptyset$ and $B_I = A_I$ for all $I \in \mathcal{I} \setminus \{I_0\}$, we would have $\mu_{[n]}(A_I \setminus B_I) = 0$ for all $I \in \mathcal{I}$, yet $\bigcap_{I\in\mathcal{I}} B_I = \emptyset$ — so no $\delta > 0$ as required could be chosen). Hence

$$\mu_{[n]}\left(\bigcap_{I\in\mathcal{I}} A_I\right) = \prod_{I\in\mathcal{I}} \mu_{[n]}(A_I) > 0.$$

Induction step. So we assume that $k > 1$ and that whenever $B_I \in \mathcal{B}_{[n],I}$ with $\mu_{[n]}(A_I \setminus B_I) < \delta$ for all $I \in \mathcal{I}$, then $\bigcap_{I\in\mathcal{I}} B_I \neq \emptyset$.

We prove it in a series of claims.

We saw in the regularity lemma, that the indicator function of a graph can be well-approximated by a simple function on the product Boolean algebra. Similarly, the indicator function of a hypergraph can be "approximated" by a simple function on the Boolean algebra generated by all of the smaller product Boolean algebras — as the following two claims will show.

Fact. Let $(\Omega, \mathcal{B}, \mu)$ be a probability space, and let $\mathcal{A}$ be a σ-subalgebra of $\mathcal{B}$. Given a $\mathcal{B}$-measurable function $f : \Omega \to \mathbb{R}_{\geq 0}$, there is a unique (up to differences on sets of measure zero) $\mathcal{A}$-measurable function $g : \Omega \to \mathbb{R}_{\geq 0}$ with the property that $\int_X f d\mu = \int_X g d\mu$ for every $X \in \mathcal{A}$. Such a g is denoted $\mathbb{E}\left(f|\mathcal{A}\right)$, the *conditional expectation* of f relatively to $\mathcal{A}$. (This is a corollary of the Radon-Nikodym theorem).

Claim 1. For any $I_0 \in \mathcal{I}$,

$$\int_{\bigcap_{I \in \mathcal{I} \setminus \{I_0\}} A_I} \mathbf{1}_{A_{I_0}} = \int_{\bigcap_{I \in \mathcal{I} \setminus \{I_0\}} A_I} \mathbb{E}\left(\mathbf{1}_{A_{I_0}} | \mathcal{B}^{\sigma}_{[n], < I_0}\right) d\mu_{[n]}.$$

Proof.

Let $f : V_{[n]} \to \mathbb{R}$ be the function defined by

$$f := \left(\mathbf{1}_{A_{I_0}} - \mathbb{E}\left(\mathbf{1}_{A_{I_0}} | \mathcal{B}^{\sigma}_{[n], < I_0}\right)\right) \prod_{I \in \mathcal{I} \setminus \{I_0\}} \mathbf{1}_{A_I}.$$

Note that f is $\mathcal{B}_{[n]}$-measurable. By Definition 3.7(3), the function $a \mapsto \int_{V_{I_0}} f\left(x_{I_0}, a\right) d\mu_{I_0}$ from $V_{[n] \setminus I_0}$ to $\mathbb{R}$ is $\mathcal{B}_{[n] \setminus I_0}$-measurable and

$$\int_{V_{[n]}} f\left(x_{[n]}\right) d\mu_{[n]} = \int_{V_{[n] \setminus I_0}} \left(\int_{V_{I_0}} f\left(x_{I_0}, a\right) d\mu_{I_0}\right) d\mu_{[n] \setminus I_0}\left(a\right).$$

Hence it is sufficient to show that $\int_{V_{I_0}} f\left(x_{I_0}, a\right) d\mu_{I_0} = 0$ for all $a \in V_{[n] \setminus I_0}$.

Fix some $a \in V_{[n] \setminus I_0}$. We want to exploit the fact that the sets involved don't depend on the coordinates outside of I_0. For each $I \in \mathcal{I}_0$ we have:

- $A_I = A'_I \times V_{[n] \setminus I}$ for some $A'_I \in \mathcal{B}^{\sigma}_I$,
- $\mathbf{1}_{A_{I_0}}\left(x_{I_0}, a\right) = \mathbf{1}_{A'_{I_0}}\left(x_{I_0}\right)$,
- For any $I \in \mathcal{I} \setminus \{I_0\}$, as $|I_0| = |I| = k$, we have $I \cap I_0 \subsetneq I_0, I$. So $\mathbf{1}_{A_I}\left(x_{I_0}, a\right) = \mathbf{1}_{A'_I}\left(x_{I \cap I_0}, a_{I \setminus I_0}\right)$ — it is $\mathcal{B}^{\sigma}_{I_0, I_0 \cap I}$-measurable by compatibility.
- Hence $\prod_{I \in \mathcal{I} \setminus \{I_0\}} \mathbf{1}_{A_I}\left(x_{I_0}, a\right)$ is $\mathcal{B}_{I_0, < I_0}$-measurable, and
$$C' := \left\{c \in V_{I_0} : \prod_{I \in \mathcal{I} \setminus \{I_0\}} \mathbf{1}_{A_I}\left(x_{I_0}, a\right) > 0\right\} \in \mathcal{B}^{\sigma}_{I_0, < I_0}.$$
- Let $h'\left(x_{I_0}\right) := \mathbb{E}\left(\mathbf{1}_{A'_{I_0}}\left(x_{I_0}\right) | \mathcal{B}^{\sigma}_{I_0, < I_0}\right)$. We claim that $h\left(x_{[n]}\right) := h'\left(x_{I_0}\right)$ gives $\mathbb{E}\left(\mathbf{1}_{A_{I_0}}\left(x_{[n]}\right) | \mathcal{B}^{\sigma}_{[n], < I_0}\right)$.

  To see this, take any $D \in \mathcal{B}^{\sigma}_{[n], < I_0}$. W.m.a. $D = D' \times V_{[n] \setminus I_0}$ for some $D' \in \mathcal{B}^{\sigma}_{I_0, < I_0}$. Then, using compatibility,

$$\int_D h\left(x_{[n]}\right) d\mu_{[n]} = \int_{V_{[n] \setminus I_0}} \left(\int_{D'} h\left(x_{I_0}, x_{[n] \setminus I_0}\right) d\mu_{I_0}\right) d\mu_{[n] \setminus I_0}\left(x_{[n] \setminus I_0}\right) =$$

$$\int_{V_{[n] \setminus I_0}} \left(\int_{D'} h'\left(x_{I_0}\right) d\mu_{I_0}\right) d\mu_{[n] \setminus I_0} = \int_{V_{[n] \setminus I_0}} \left(\int_{D'} \mathbf{1}_{A'_{I_0}}\left(x_{I_0}\right) d\mu_{I_0}\right) d\mu_{[n] \setminus I_0} =$$

$$= \int_{V_{[n] \setminus I_0}} \left(\int_{D'} \mathbf{1}_{A_{I_0}}\left(x_{I_0}, x_{[n] \setminus I}\right) d\mu_{I_0}\right) d\mu_{[n] \setminus I_0} = \int_D \mathbf{1}_{A_{I_0}}\left(x_{[n]}\right) d\mu_{[n]}.$$

Hence, $\mathbb{E}\left(\mathbf{1}_{A'_{I_0}}|\mathcal{B}^{\sigma}_{I_0,<I_0}\right)(x_{I_0})=\mathbb{E}\left(\mathbf{1}_{A_{I_0}}|\mathcal{B}^{\sigma}_{[n],<I_0}\right)(x_{I_0},a)$.

Combining these observations we have

$$\int_{V_{I_0}} f(x_{I_0},a)\,d\mu_{I_0}=\int_{C'}\left(\mathbf{1}_{A'_{I_0}}(x_{I_0})-\mathbb{E}\left(\mathbf{1}_{A'_{I_0}}|\mathcal{B}^{\sigma}_{I_0,<I_0}\right)(x_{I_0})\right)d\mu_{I_0}=0.$$

Claim 2. For any $I_0\in\mathcal{I}$, there is some $A'_{I_0}\in\mathcal{B}^{\sigma}_{[n],<I_0}$ such that:

- If $B_I\in\mathcal{B}_{[n],I}$ for all $I\in\mathcal{I}$ satisfy $\mu_{[n]}(A_I\setminus B_I)<\delta$ for each $I\neq I_0$ and $\mu_{[n]}\left(A'_{I_0}\setminus B_{I_0}\right)<\delta$, then $\bigcap_{I\in\mathcal{I}}B_I\neq\emptyset$,
- If $\mu_{[n]}\left(A'_{I_0}\cap\bigcap_{I\in\mathcal{I}\setminus\{I_0\}}A_I\right)>0$ then $\mu_{[n]}\left(\bigcap_{I\in\mathcal{I}}A_I\right)>0$.

Proof. Define $A'_{I_0}:=\left\{x_{[n]}\in V_{[n]}:\mathbb{E}\left(\mathbf{1}_{A_{I_0}}|\mathcal{B}^{\sigma}_{[n],<I_0}\right)(x_{[n]})>0\right\}$. Note that $A'_{I_0}\in\mathcal{B}^{\sigma}_{[n],<I_0}$ (as $\mathbb{E}\left(\mathbf{1}_{A_{I_0}}|\mathcal{B}^{\sigma}_{[n],<I_0}\right)$ is $\mathcal{B}^{\sigma}_{[n],<I_0}$-measurable).

If $\mu_{[n]}\left(A'_{I_0}\cap\bigcap_{I\in\mathcal{I}\setminus\{I_0\}}A_I\right)>0$ then

$$\int_{\bigcap_{I\in\mathcal{I}\setminus\{I_0\}}A_I}\mathbb{E}\left(\mathbf{1}_{A_{I_0}}|\mathcal{B}^{\sigma}_{[n],<I_0}\right)d\mu_{[n]}>0,$$

and using Claim 1 this implies

$$\mu_{[n]}\left(\bigcap_{I\in\mathcal{I}}A_I\right)=\int_{\bigcap_{I\in\mathcal{I}\setminus\{I_0\}}A_I}\mathbf{1}_{A_{I_0}}d\mu_{[n]}=\int_{\bigcap_{I\in\mathcal{I}\setminus\{I_0\}}A_I}\mathbb{E}\left(\mathbf{1}_{A_{I_0}}|\mathcal{B}^{\sigma}_{[n],<I_0}\right)d\mu_{[n]}>0.$$

Suppose now that for each I, $B_I\in\mathcal{B}_{[n],I}$ with $\mu_{[n]}(A_I\setminus B_I)<\delta$ for $I\in\mathcal{I}\setminus\{I_0\}$ and $\mu_{[n]}\left(A'_{I_0}\setminus B_{I_0}\right)<\delta$. We also have

$$\mu_{[n]}\left(A_{I_0}\setminus A'_{I_0}\right)=\int_{V_{[n]}\setminus A'_{I_0}}\mathbf{1}_{A_{I_0}}d\mu_{[n]}=$$

$$\int_{V_{[n]}\setminus A'_{I_0}}\mathbb{E}\left(\mathbf{1}_{A_{I_0}}|\mathcal{B}_{[n],<I_0}\right)d\mu_{[n]}=0,$$

hence $\mu_{[n]}(A_{I_0}\setminus B_{I_0})<\delta$ as well, therefore $\bigcap_{I\in\mathcal{I}}B_I\neq\emptyset$.

Applying Claim 2 to each $I\in\mathcal{I}$, we may assume for the rest of the proof that $A_I\in\mathcal{B}^{\sigma}_{[n],<I}$ for all $I\in\mathcal{I}$.

Fix some **finite** Boolean algebra $\mathcal{B}\subseteq\mathcal{B}_{[n],k-1}$ (hence $\mathcal{B}^{\sigma}=\mathcal{B}$) so that for every $I\in\mathcal{I}$, $\|\mathbf{1}_{A_I}-\mathbb{E}(\mathbf{1}_{A_I}|\mathcal{B})\|_{L^2(\mu_{[n]})}<\frac{\sqrt{\delta}}{\sqrt{2}(|\mathcal{I}|+1)}$ (such a $\mathcal{B}$ exists because there are finitely many I and each A_I is $\mathcal{B}^{\sigma}_{[n],k-1}$-measurable, and $\mathcal{B}^{\sigma}_{[n],k-1}$ is generated by $\mathcal{B}_{[n],k-1}$). For each $I\in\mathcal{I}$, set $A^*_I:=\left\{a_I:\mathbb{E}(\mathbf{1}_{A_I}|\mathcal{B}^{\sigma})(a_I)>\frac{|\mathcal{I}|}{|\mathcal{I}|+1}\right\}\in\mathcal{B}$.

Claim 3. For each $I\in\mathcal{I}$, $\mu_{[n]}(A_I\setminus A^*_I)\leq\frac{\delta}{2}$.

Proof. Note that

$$A_I\setminus A^*_I=\left\{a\in V_{[n]}:(\mathbf{1}_{A_I}-\mathbb{E}(\mathbf{1}_{A_I}|\mathcal{B}))(a)\geq 1-\frac{|\mathcal{I}|}{|\mathcal{I}|+1}=\frac{1}{|\mathcal{I}|+1}\right\}$$

$$\subseteq\left\{a\in V_{[n]}:|\mathbf{1}_{A_I}-\mathbb{E}(\mathbf{1}_{A_I}|\mathcal{B})|(a)\geq\frac{1}{|\mathcal{I}|+1}\right\}.$$

Recall:

Fact. (Markov's inequality) Let $(\Omega, \mathcal{B}, \mu)$ be a probability space. Given a $\mathcal{B}$-measurable function $f : \Omega \to \mathbb{R}_{\geq 0}$ and $\alpha > 0$, we have

$$\mu(\{a \in \Omega : f \geq \alpha\}) \leq \frac{\int_\Omega (f)\, d\mu}{\alpha}.$$

Applying Markov's inequality to $f := (\mathbf{1}_{A_I} - \mathbb{E}(\mathbf{1}_{A_I}|\mathcal{B}))^2$, we get that $\mu_{[n]}(A_I \setminus A_I^*)$ is at most

$$(|\mathcal{I}|+1)^2 \int_{V_{[n]}} (\mathbf{1}_{A_I} - \mathbb{E}(\mathbf{1}_{A_I}|\mathcal{B}))^2\, d\mu_n = (|\mathcal{I}|+1)^2 \left\|\mathbf{1}_{A_I} - \mathbb{E}(\mathbf{1}_{A_I}|\mathcal{B})\right\|^2_{L^2(\mu_{[n]})} \leq \frac{\delta}{2}.$$

Claim 4. $\mu_{[n]}\left(\bigcap_{I\in\mathcal{I}} A_I\right) \geq \frac{\mu_{[n]}\left(\bigcap_{I\in\mathcal{I}} A_I^*\right)}{|\mathcal{I}|+1}$.

Proof. For each $I_0 \in \mathcal{I}$, as $A_{I_0^*} \cap \bigcap_{I\in\mathcal{I}\setminus\{I_0\}} A_I^* \in \mathcal{B}$, we have

$$\mu_{[n]}\left(\left(A_{I_0}^* \setminus A_{I_0}\right) \cap \bigcap_{I\in\mathcal{I}\setminus\{I_0\}} A_I^*\right) = \int_{A_{I_0^*}\cap\bigcap_{I\in\mathcal{I}\setminus\{I_0\}} A_I^*} \left(1 - \mathbf{1}_{A_{I_0}}\right) d\mu_{[n]} =$$

$$\int_{A_{I_0^*}\cap\bigcap_{I\in\mathcal{I}\setminus\{I_0\}} A_I^*} \left(1 - \mathbb{E}\left(\mathbf{1}_{A_{I_0}}|\mathcal{B}\right)\right) d\mu_{[n]} = \int_{A_{I_0}^*} \left(1 - \mathbb{E}\left(\mathbf{1}_{A_{I_0}}|\mathcal{B}\right)\right) \prod_{I\in\mathcal{I}\setminus\{I_0\}} \mathbf{1}_{A_I^*} d\mu_{[n]}$$

$$\leq \frac{1}{|\mathcal{I}|+1} \int_{V_{[n]}} \prod_{I\in\mathcal{I}} \mathbf{1}_{A_I^*} d\mu_{[n]}$$

$$= \frac{1}{|\mathcal{I}|+1} \mu_{[n]}\left(\bigcap_{I\in\mathcal{I}} A_I^*\right).$$

But then

$$\mu_{[n]}\left(\bigcap_{I\in\mathcal{I}} A_I^* \setminus \bigcap_{I\in\mathcal{I}} A_I\right) \leq \sum_{I_0\in\mathcal{I}} \mu_{[n]}\left(\left(A_{I_0}^* \setminus A_{I_0}\right) \cap \bigcap_{I\in\mathcal{I}\setminus\{I_0\}} A_I^*\right)$$

$$\leq |\mathcal{I}| \frac{1}{|\mathcal{I}|+1} \mu_{[n]}\left(\bigcap_{I\in\mathcal{I}} A_I^*\right),$$

so $\mu\left(\bigcap_{I\in\mathcal{I}} A_I\right) \geq \mu\left(\bigcap_{I\in\mathcal{I}} A_I^*\right) - \mu\left(\bigcap_{I\in\mathcal{I}} A_I^* \setminus \bigcap_{I\in\mathcal{I}} A_I\right) \geq \left(1 - |\mathcal{I}| \frac{1}{|\mathcal{I}|+1}\right) \mu_{[n]}\left(\bigcap_{I\in\mathcal{I}} A_I^*\right) \geq \frac{1}{|\mathcal{I}|+1} \mu_{[n]}\left(\bigcap_{I\in\mathcal{I}} A_I^*\right)$.

- Each $A_I^* \in \mathcal{B}$ can be written in the form $A_I^* = \bigcup_{i \leq r_I} A_{I,i}^*$ for some $r_I \in \mathbb{N}$, where $A_{I,i}^* = \bigcap_{J\in\binom{I}{k-1}} A_{I,i,J}^*$ and $A_{I,i,J}^* \in \mathcal{B}_{[n],J}$, such that if $i \neq i'$ then $A_{I,i}^* \cap A_{I,i'}^* = \emptyset$.
- Then
$$\bigcap_{I\in\mathcal{I}} A_I^* = \bigcup_{\vec{i}\in\prod_{I\in\mathcal{I}}[1,r_I]} \bigcap_{I\in\mathcal{I}} \bigcap_{J\in\binom{I}{k-1}} A_{I,i_I,J}^*.$$
- For each $\vec{i} \in \prod_{I\in\mathcal{I}} [1, r_I]$, let $D_{\vec{i}} := \bigcap_{I\in\mathcal{I}} \bigcap_{J\in\binom{I}{k-1}} A_{I,i_I,J}^*$.
- Each $A_{I,i_I,J}^* \in \mathcal{B}_{[n],J}$, so we may regroup the components and write $D_{\vec{i}} = \bigcap_{J\in\binom{[n]}{k-1}} D_{\vec{i},J}$ where $D_{\vec{i},J} = \bigcap_{J\subsetneq I\in\mathcal{I}} A_{I,i_I,J}^* \in \mathcal{B}_{[n],J}$.

- Suppose, for a contradiction, that $\mu_{[n]}\left(\bigcap_I A_I^*\right) = 0$.
  Then $\mu_{[n]}\left(D_{\vec{i}}\right) = \mu_{[n]}\left(\bigcap_{J \in \binom{[n]}{k-1}} D_{\vec{i},J}\right) = 0$ for all $\vec{i} \in \prod_{I \in \mathcal{I}} [1, r_I]$.
- By the inductive hypothesis applied to each of the $D_{\vec{i}}$, $\vec{i} \in \vec{i} \in \prod_{I \in \mathcal{I}} [1, r_I]$, for each real $\gamma > 0$, there are then some $B_{\vec{i},J} \in \mathcal{B}_{[n],J}$ such that $\mu_{[n]}\left(D_{\vec{i},J} \setminus B_{\vec{i},J}\right) < \gamma$ and $\bigcap_{J \in \binom{[n]}{k-1}} B_{\vec{i},J} = \emptyset$.
  In particular, this holds with $\gamma := \frac{\delta}{2\binom{k}{k-1}\prod_{I \in \mathcal{I}} r_I \max_{I \in \mathcal{I}} r_I}$.
- For each $I \in \mathcal{I}$, $i \leq r_I$, $J \subsetneq I$ define

$$B_{I,i,J}^* = A_{I,i,J}^* \cap \bigcap_{\vec{i}, i_I = i} \left(B_{\vec{i},J} \cup \bigcup_{I' \supsetneq J, I' \neq I} \neg A_{I', i_{I'}, J}^* \right) \in \mathcal{B}_{[n],J}.$$

Claim 5. $\mu_{[n]}\left(A_{I,i,J}^* \setminus B_{I,i,J}^*\right) \leq \frac{\delta}{2\binom{k}{k-1}\max_{I \in \mathcal{I}} r_I}$.

Proof. If $x \in A_{I,i,J}^* \setminus B_{I,i,J}^*$, then for some $\vec{i} \in \prod_{I \in \mathcal{I}} [1, r_I]$ with $i_I = i$ we have

$$x \notin B_{\vec{i},J} \cup \bigcup_{I' \supsetneq J, I' \neq I} \neg A_{I', i_{I'}, J}^*.$$

This means $x \notin B_{\vec{i},J}$ and $x \in \bigcap_{I' \supsetneq J} A_{I', i_{I'}, J}^* = D_{\vec{i},J}$. So

$$\mu_{[n]}\left(A_{I,i,J}^* \setminus B_{I,i,J}^*\right) \leq \sum_{\vec{i} \in \prod_{I \in \mathcal{I}} [1, r_I]} \mu_{[n]}\left(D_{\vec{i},J} \setminus B_{\vec{i},J}\right) \leq \frac{\delta}{2\binom{k}{k-1}\max_{I \in \mathcal{I}} r_I}.$$

Claim 6. Let $B_I^* := \bigcup_{i \leq r_I} \bigcap_{J \in \binom{I}{k-1}} B_{I,i,J}^* \in \mathcal{B}_{[n],<I}$. Then $\mu_{[n]}\left(A_I \setminus B_I^*\right) \leq \delta$.

Proof. As $\mu_{[n]}\left(A_I \setminus A_I^*\right) \leq \frac{\delta}{2}$ by Claim 3, it suffices to show that $\mu_{[n]}\left(A_I^* \setminus B_I^*\right) \leq \frac{\delta}{2}$. We have

$$\begin{aligned}
\mu_{[n]}\left(A_I^* \setminus B_I^*\right) &= \mu_{[n]}\left(A_I^* \setminus \bigcup_{i \leq r_I} \bigcap_{J \in \binom{I}{k-1}} B_{I,i,J}^* \right) \\
&= \mu_{[n]}\left(\bigcup_{i \leq r_I} \bigcap_{J \in \binom{I}{k-1}} A_{I,i,J}^* \setminus \bigcup_{i \leq r_I} \bigcap_{J \in \binom{I}{k-1}} B_{I,i,J}^* \right) \\
&\leq \mu_{[n]}\left(\bigcup_{i \leq r_I} \left(\bigcap_J A_{I,i,J}^* \setminus \bigcap_J B_{I,i,J}^* \right) \right) \\
&\leq \sum_{i \leq r_I} \mu_{[n]}\left(\bigcap_J A_{I,i,J}^* \setminus \bigcap_J B_{I,i,J}^* \right) \\
&\leq \sum_{i \leq r_I} \sum_J \mu_{[n]}\left(A_{I,i,J}^* \setminus B_{I,i,J}^*\right) \\
&\leq r_I \binom{k}{k-1} \frac{\delta}{2\binom{k}{k-1}\max_{I \in \mathcal{I}} r_I} \leq \frac{\delta}{2}
\end{aligned}$$

using Claim 5.

Hence the sets B_I^* satisfy the assumption for all $I \in \mathcal{I}$ by Claim 6, therefore $\bigcap_{I\in\mathcal{I}} B_I^* \neq \emptyset$.

Claim 7. $\bigcap_{I\in\mathcal{I}} B_I^* \subseteq \bigcup_{\vec{i}\in\prod_{I\in\mathcal{I}}[1,r_I]} \bigcap_{J\in\binom{[n]}{k-1}} B_{\vec{i},J}$.

Proof. Suppose $x \in \bigcap_{I\in\mathcal{I}} B_I^* \subseteq \bigcap_{I\in\mathcal{I}} \bigcup_{i\leq r_I} \bigcap_{J\in\binom{I}{k-1}} B_{I,i,J}^*$. Then for each $I \in \mathcal{I}$, there is some $i_I \leq r_I$ such that $x \in \bigcap_{J\in\binom{I}{k-1}} B_{I,i_I,J}^*$, and take $\vec{i}_x := (i_I : I \in \mathcal{I})$.

Since $B_{I,i_I,J}^* \subseteq A_{I,i_I,J}^*$, for each $I \in \mathcal{I}$ and $J \subsetneq I$, $x \in A_{I,i_I,J}^*$.

For any J, let $I \supset J$. Then

$$x \in B_{I,i_I,J}^* = A_{I,i_I,J}^* \cap \bigcap_{\vec{i}',i'_I=i_I} \left(B_{\vec{i}',J} \cup \bigcup_{I'\supseteq J, I'\neq I} \neg A_{I',i_{I'},J}^* \right).$$

In particular, $x \in B_{\vec{i}_x,J} \cup \bigcup_{I'\supseteq J,I'\neq I} \neg A_{I',i_{I'},J}^*$.

Since $x \in A_{I,i_{I'},J}^*$ for each $I' \supset J$, necessarily $x \in B_{\vec{i},J}$. This holds for any J, so $x \in \bigcap_J B_{\vec{i},J}$.

Since $\bigcap_{I\in\mathcal{I}} B_I^* \neq \emptyset$, there is some $\vec{i} \in \prod_{I\in\mathcal{I}}[1,r_I]$ such that $\bigcap_J B_{\vec{i},J} \neq \emptyset$. This is a contradiction to our assumption, hence $\mu_{[n]}\left(\bigcap_I A_I^*\right) > 0$, and therefore, $\mu_{[n]}\left(\bigcap_{I\in\mathcal{I}} A_I\right) \geq \frac{\mu_{[n]}\left(\bigcap_{I\in\mathcal{I}} A_I^*\right)}{|\mathcal{I}|+1} > 0$ by Claim 4. □

Corollary 3.13. *(Hypergraph removal, partitioned version of Theorem 3.3).*

Fix $0 \leq k \leq n$, $\varepsilon > 0$ and $\mathcal{I} \subseteq \binom{[n]}{k}$ a k-uniform hypergraph on $[n]$. Then there is $\delta > 0$ such that the following holds.

Let $(V_i : i \in [n])$ be finite non-empty sets. For each $I \in \mathcal{I}$, let A_I be a subset of $\prod_{i\in I} V_i$. Suppose that

$$\left|\left\{ (x_i)_{i\in[n]} \in \prod_{i\in[n]} V_i : (x_i)_{i\in I} \in A_I \text{ for all } I \in \mathcal{I} \right\}\right| \leq \delta \prod_{i\in[n]} |V_i|.$$

*(i.e. the n-partite hypergraph $G = \left((V_i)_{i\in[n]}, (A_I)_{I\in\mathcal{I}}\right)$ contains at most $\delta \prod_{i\in[n]} |V_i|$ copies of $\mathcal{I}$ - **not induced, just as a subgraph**).*

Then for each $I \in \mathcal{I}$ there exists $B_I \subseteq \prod_{i\in I} V_i$ with $|A_I \setminus B_I| < \varepsilon \prod_{i\in I} |V_i|$ such that

$$\left\{ (x_i)_{i\in[n]} \in \prod_{i\in[n]} V_i : (x_i)_{i\in I} \in B_I \text{ for all } I \in \mathcal{I} \right\} = \emptyset.$$

(i.e. the n-partite hypergraph $G' = \left((V_i)_{i\in[n]}, (B_I)_{I\in\mathcal{I}}\right)$ contains no copies of $\mathcal{I}$ whatsoever).

Proof. Assume not, and let $k, \mathcal{I} \subseteq \binom{[n]}{k}$ and $\varepsilon > 0$ be a counterexample. Since there is no $\delta > 0$ as in the statement of the theorem, for each $m \in \mathbb{N}$ we may choose a k-uniform hypergraph $G_m = \left((V_i^m)_{i\in[n]}, (A_I^m)_{I\in\mathcal{I}}\right)$ such that G_m contains at most $\frac{1}{m}\prod_{i\in[n]} |V_i^m|$ copies of $\mathcal{I}$, but there are no subsets $B_I, I \in \mathcal{I}$ as required.

Then clearly $|V_i^m| \to \infty$ as $m \to \infty$. Let $\mathcal{U}$ be a non-principal ultrafilter on $\mathbb{N}$, and let $G := \prod_{m\in\mathbb{N}} G_m/\mathcal{U}$, $G = \left((V_i)_{i\in[n]}, (A_I)_{I\in\mathcal{I}}\right)$. For each $I \subseteq [n]$, let μ_I^m be the normalized counting measure on V_I^m, and let $\mu_I = \lim_{\mathcal{U}} \mu_I^m$ — a f.a.p. measure on the internal subsets of $V_I = \prod_{i \in [n]} V_i$.

Then $(\mu_I : I \subseteq [n])$ is a compatible system of measures on $(V_i : i \in [n])$ (Exercise 3.10). Note that by assumption $\mu_{[n]}^m \left(\bigcap_{I \in \mathcal{I}} A_I^m \right) < \frac{1}{m}$ for all $m \in \mathbb{N}$, hence $\mu_{[n]} \left(\bigcap_{I \in \mathcal{I}} A_I \right) = 0$. By Theorem 3.12, there are some **internal** sets B_I, $I \in \mathcal{I}$, such that $\mu_{[n]} (A_I \setminus B_I) < \frac{\varepsilon}{2}$ and $\bigcap_{I \in \mathcal{I}} B_I = \emptyset$. Say $B_I = \prod_{m \in \mathbb{N}} B_I^m / \mathcal{U}$. Then for some $S \in \mathcal{U}$ and all $m \in S$, we must have $\mu_{[n]}^m (A_I^m \setminus B_I^m) < \varepsilon$ and $\bigcap_{I \in \mathcal{I}} B_I^m = \emptyset$ — a contradiction to the choice of the G_m's. □

Exercise 3.14. Deduce Theorem 3.3 from Corollary 3.13 (taking $V_i = V$ for all i and constructing the corresponding *partite* hypergraph).

3.3. Szemerédi's theorem on arithmetic progressions.

Theorem 3.15. *(Szemerédi's theorem) For any $\varepsilon > 0$ and $k \in \mathbb{N}$, there is some $n_0 \in \mathbb{N}$ such that for any $n \geq n_0$ and $A \subseteq [1, n]$ with $|A| \geq \varepsilon n$, there exists an a and $d \neq 0$ such that $a, a + d, a + 2d, \ldots, a + (k - 1) d \in A$.*

Proof. Let $\delta > 0$ be as given by Theorem 3.13 for $\varepsilon' := \frac{\varepsilon^k}{2^k k^{2(k-1)}} > 0, k$ and W the complete k-uniform hypergraph on $k + 1$ vertices. Let n_0 be large enough so that $\delta n_0^{k+1} > n_0^k$.

Let $A \subseteq [1, n]$ be given, with $n \geq n_0$. We define a k-uniform $(k + 1)$-partite hypergraph as follows. Let $V_i := [1, n]$ for each $i = 1, \ldots, k$ and let $V_{k+1} := \left[1, k^2 n\right] \subseteq \mathbb{N}$. Given $x_i \in V_i$ for all $i = 1, \ldots, k + 1$, we define:

- $(x_1, \ldots, x_k)$ is an edge iff $\sum_{i \in [1,k]} i x_i \in A$,
- for any $1 \leq i \leq k$, $(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_{k+1})$ is an edge iff $\sum_{j \in [1,k] \setminus \{i\}} j x_j + i \left(x_{k+1} - \sum_{j \in [1,k] \setminus \{i\}} x_j \right) \in A$.

Suppose $(x_1, \ldots, x_{k+1})$ is a copy of the complete k-uniform hypergraph on $k + 1$ vertices with $x_{k+1} \neq \sum_{i \in [1,k]} x_i$. Then let $a := \sum_{i \in [1,k]} i x_i$ and $d = x_{k+1} - \sum_{i \in [1,k]} x_i \neq 0$. Then we have $a \in A$, and for each $i \leq k$ we have

$$a+id = \sum_{j \in [1,k]} j x_j + i \left(x_{k+1} - \sum_{j \in [1,k]} x_j \right) = \sum_{j \in [1,k] \setminus \{i\}} j x_j + i \left(x_{k+1} - \sum_{j \in [1,k] \setminus \{i\}} x_j \right) \in A,$$

as wanted.

On the other hand, for any $a \in A$ and any sequence $(x_1, \ldots, x_k)$ with $a = \sum_{i \in [1,k]} i x_i$, the sequence $\left(x_1, \ldots, x_k, \sum_{i \in [1,k]} x_i \right)$ is also a copy of the complete k-uniform hypergraph on $k + 1$ vertices: clearly $(x_1, \ldots, x_k)$ is an edge, and for any $1 \leq i \leq k$ we have $\left(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_k, \sum_{i \in [1,k]} x_i \right)$ is an edge as $\sum_{j \in [1,k] \setminus \{i\}} j x_j + i \left(\sum_{i \in [1,k]} x_i - \sum_{j \in [1,k] \setminus \{i\}} x_j \right) = \sum_{j \in [1,k] \setminus \{i\}} j x_j + i x_i = \sum_{i \in [1,k]} i x_i = a \in A$. There are at least $\frac{\varepsilon}{2} n$ choices for $a \in A$ with $a \geq \frac{\varepsilon}{2} n$; and for any $a \geq \frac{\varepsilon}{2} n$ and any choice of $x_i \in \left[1, \frac{\varepsilon}{2k^2} n\right]$ for $i = 1, \ldots, k-1$ we have $\sum_{i \in [1,k]} i x_i \leq \sum_{i \in [1,k]} k x_i \leq k^2 \frac{\varepsilon}{2k^2} n \leq \frac{\varepsilon}{2} n$, so there is some $x_k \in [1, n]$ satisfying $\sum_{i \in [1,k]} i x_i = a$. Hence the number of such sequences is at least $\frac{\varepsilon}{2} n \left(\frac{\varepsilon}{2k^2} n \right)^{k-1} = \frac{\varepsilon^k}{2^k k^{2(k-1)}} n^k \geq \varepsilon' n^k$. It is not possible to remove all such sequences by removing $< \varepsilon' n^k$ edges. Hence the hypergraph removal (Corollary 3.13) implies that there must be $> \delta n^{k+1}$ many copies of the complete k-uniform hypergraph on $k + 1$ vertices. But there are at most n^k

sequences of the form $\left(x_1, \ldots, x_k, \sum_{i \leq k} x_i\right)$ and $\delta n^{k+1} > n^k$ by assumption on n, so the remaining copies must correspond to arithmetic progressions. □

3.4. **References.** The proof of Theorem 3.12 presented here follows [19, Section 6], with some clarifications, which in turn is based on the ideas in Tao [17, 18] and others. The deduction of Szemerédi's theorem from hypergraph removal is due to Frankl and Rödl [5], we follow the presentation in [6].

4. Regularity lemma for hypergraphs of finite VC-dimension

4.1. **Bounds in the regularity lemma.** Recall the graph regularity lemma (Theorem 2.1) in the bipartite version.

Theorem. *Let $\varepsilon > 0$ be arbitrary. Then there is some $K = K(\varepsilon) \in \mathbb{N}$ such that for every bipartite finite graph $G = (V, W, E)$ with $|V|, |W| \geq K$ there are partitions $V = V_1 \sqcup \cdots \sqcup V_n$ and $W = W_1 \sqcup \ldots \sqcup W_n$, real numbers $\delta_{ij}, i, j \in [n]$, and an exceptional set of pairs $\Sigma \subseteq [n] \times [n]$ such that:*

(1) *(Bounded size of the partition) $n \leq K$,*
(2) *(Few exceptional pairs) $\sum_{(i,j)\in\Sigma} |V_i||W_j| \leq \varepsilon |V| |W|$,*
(3) *(ε-regularity) for each $(i, j) \in [n] \times [n] \setminus \Sigma$ we have*

$$\big|\, |E(A, B)| - \delta_{ij} |A||B| \,\big| < \varepsilon |V_i||W_j|$$

*for **all** $A \subseteq V_i$, $B \subseteq W_j$. We call a pair of sets (V_i, W_j) with $(i, j) \in [K] \times [K] \setminus \Sigma$ an ε-*regular pair.

Remark 4.1. (1) Exceptional pairs are unavoidable (for large n, let $V = W = [n]$ and let $E \subseteq V \times W$ be defined by $E = \{(i, j) : i, j \in [n], i < j\}$ — there is no way to cover the diagonal by a bounded number of regular pairs).
(2) The densities $\delta_{i,j} \in [0, 1]$ could be arbitrary, e.g. we cannot hope to have $\delta_{i,j} \in \{0, 1\}$ in general (e.g. for large n, take a graph with edges distributed uniformly at random with probability $\frac{1}{2}$).
(3) The size of the partition is unavoidably huge!

Fact 4.2. *(Gowers* [7, 13]*) $K(\varepsilon)$ is at least an exponential tower of 2's of height $O\left(\frac{1}{\varepsilon}\right)^c$ for some $c = \frac{1}{16}$.*

Fact 4.3. *(Fox, Lovás* [4]*) $K(\varepsilon)$ is at least an exponential tower of height $O\left(\frac{1}{\varepsilon}\right)^c$ with $c = 2$, and this bound is tight.*

The graphs witnessing these are constructed using probabilistic methods. Perhaps one can do better for graphs defined "geometrically" or "algebraically"? We discuss improved regularity lemmas for some restricted families of graphs. It turns out that these conditions can be characterized by some model-theoretic notions of "tameness".

4.2. **VC-dimension.** For more details and proofs of the facts in this section, see [2].

Let X be a set (finite or infinite), and let $\mathcal{F}$ be a family of subsets of X. A pair $(X, \mathcal{F})$ is called a *set system.*

Definition 4.4. (1) Given $A \subseteq X$, we say that the family $\mathcal{F}$ *shatters* A if for every $A' \subseteq A$, there is a set $S \in \mathcal{F}$ such that $S \cap A = A'$.

(2) The family $\mathcal{F}$ has *VC-dimension* at most n (written as $\mathrm{VC}(\mathcal{F}) \le n$), if there is no $A \subseteq X$ of cardinality $n+1$ such that $\mathcal{F}$ shatters A. We say that $\mathcal{F}$ is of VC-dimension n if it is of VC-dimension at most n and shatters some subset of size n.
(3) If for every $n \in \mathbb{N}$ we can find a subset of X of cardinality n shattered by $\mathcal{F}$, then we say that $\mathcal{F}$ has infinite VC-dimension ($\mathrm{VC}(\mathcal{F}) = \infty$). If $\mathrm{VC}(\mathcal{F})$ is finite, we say that $\mathcal{F}$ is a *VC-family*. Note that if $\mathcal{F}' \subseteq \mathcal{F}$ then $\mathrm{VC}(\mathcal{F}') \le \mathrm{VC}(\mathcal{F})$.

Example 4.5. (1) Let X be an infinite set and $\mathcal{F} := \mathcal{P}(X)$. Then clearly $\mathrm{VC}(\mathcal{F}) = \infty$. But for $\mathcal{F} = \binom{X}{k}$, $\mathrm{VC}(\mathcal{F}) = k$.
(2) Let $X = \mathbb{R}$ and let $\mathcal{F}$ be the family of all unbounded intervals. Then $\mathcal{F}$ has VC-dimension 2. Clearly any two-element set can be shattered by $\mathcal{F}$. However, if we take any $a < b < c$, then $\{a, b, c\}$ cannot be shattered by $\mathcal{F}$.

Exercise 4.6. (1) Let $X = \mathbb{R}^2$, and let $\mathcal{F}$ be the set of all half-spaces. Show that $\mathrm{VC}(\mathcal{F}) = 3$.
(2) Let $X = \mathbb{R}^2$ and let $\mathcal{F}$ be the set of all convex polygons. Show that $\mathrm{VC}(\mathcal{F}) = \infty$.

Definition 4.7. We define the *shatter function* $\pi_{\mathcal{F}} : \mathbb{N} \to \mathbb{N}$ associated to the family $\mathcal{F}$ as follows. For a set $A \subseteq X$ we let $\mathcal{F} \cap A := \{S \cap A : S \in \mathcal{F}\}$. Then we define $\pi_{\mathcal{F}}(n) := \max\{|\mathcal{F} \cap A| : A \subseteq X, |A| = n\}$.

Note that $\pi_{\mathcal{F}}(n) \le 2^n$, and that $\mathrm{VC}(\mathcal{F}) < n \iff \pi_{\mathcal{F}}(m) < 2^m$ for all $m \ge n$. The following fundamental lemma states that either $\pi_{\mathcal{F}}(n) = 2^n$ for all $n \in \mathbb{N}$, or $\pi_{\mathcal{F}}(n)$ has polynomial growth.

Lemma 4.8. *(Sauer-Shelah lemma) Let $(X, \mathcal{F})$ be a set system of VC-dimension at most k. Then, for all $n \ge k$, we have $\pi_{\mathcal{F}}(n) \le \sum_{i=0}^{k} \binom{n}{i}$.*
In particular, $\pi_{\mathcal{F}}(n) = O(n^k)$.

Remark 4.9. (Boolean operations preserve finite VC-dimension) Let $\mathcal{F}_1, \mathcal{F}_2$ be two families of subsets of X with $\mathrm{VC}(\mathcal{F}_i) = d_i < \infty$. Show that all of the following families have finite VC-dimension:
(1) $\mathcal{F} := \mathcal{F}_1 \cup \mathcal{F}_2$,
(2) $\mathcal{F}_\cap := \{S_1 \cap S_2 : S_i \in \mathcal{F}_i, i = 1, 2\}$ and $\mathrm{VC}(\mathcal{F}_\cap) \le d_1 + d_2 + 1$,
(3) $\mathcal{F}_\cup := \{S_1 \cup S_2 : S_i \in \mathcal{F}_i, i = 1, 2\}, \mathcal{F}_1^c := \{X \setminus S_1 : S_1 \in \mathcal{F}_1\}$ and $\mathrm{VC}(\mathcal{F}_\cup) \le d_1 + d_2 + 1, \mathrm{VC}(\mathcal{F}_1^c) = d_1$,
(4) $\mathcal{F}_1 \times \mathcal{F}_2 := \{S_1 \times S_2 : S_1 \in \mathcal{F}_1, S_2 \in \mathcal{F}_2\}$ — a family of subsets of $X \times X$.
(5) Besides, if X' is an infinite set and $f : X' \to X$ is a map, let $f^{-1}(\mathcal{F}_1) := \{f^{-1}(S) : S \in \mathcal{F}_1\}$. Then $\mathrm{VC}(f^{-1}(\mathcal{F}_1)) \le \mathrm{VC}(\mathcal{F}_1)$.

Recall: by a partitioned formula $\phi(\bar{x}, \bar{y})$ we mean a formula with its free variables partitioned into two groups $\bar{x}$ (object variables) and $\bar{y}$ (parameter variables). Given a partitioned formula $\phi(\bar{x}, \bar{y})$ and $\bar{b} \in M^{|\bar{y}|}$, we let $\phi(M^{|\bar{x}|}, \bar{b})$ be the set of all $\bar{a} \in M^{|\bar{x}|}$ such that $\mathcal{M} \models \phi(\bar{a}, \bar{b})$. Sets of this form are called *definable* (or ϕ-*definable*, in this case). We consider the family $\mathcal{F}_{\phi(\bar{x}, \bar{y})}$ of subsets of $M^{|\bar{x}|}$ defined by $\mathcal{F}_{\phi(\bar{x}, \bar{y})} = \{\phi(M^{|\bar{x}|}, \bar{b}) : \bar{b} \in M^{|\bar{y}|}\}$.

Theorem 4.10. *(Shelah) Let M be a first-order structure. Assume that for every partitioned formula $\phi(x, \bar{y})$ with x a* ***singleton****, the family $\mathcal{F}_\phi$ has finite VC* *dimension. Then for any $\phi(\bar{x}, \bar{y}) \in L$, the corresponding family $\mathcal{F}_\phi$ has finite VC dimension.*

The proof uses Ramsey's theorem, and gives bounds that are quite far from optimal.

In model theory, a partitioned formula $\phi(\bar{x}, \bar{y})$ is called *NIP* (No Independence Property) if the family $\mathcal{F}_\phi$ has finite VC-dimension. A structure $\mathcal{M}$ is NIP if all definable families in it are NIP. Such structures were defined by Shelah around the same time as Vapnik and Chervonenkis have defined their dimension for entirely different purposes, and are currently being actively studied in model theory (see [16] for a survey).

Example 4.11. (Semialgebraic sets of bounded complexity) Recall that a set $X \subseteq \mathbb{R}^n$ is *semialgebraic* if it is given by a Boolean combination of polynomial equalities and inequalities.

We say that the *description complexity* of a semialgebraic set $X \subseteq \mathbb{R}^d$ is bounded by $t \in \mathbb{N}$ if $d \leq t$ and X can be defined as a Boolean combination of at most t polynomial equalities and inequalities, such that all of the polynomials involved have degree at most t. For example, consider the family of all spheres in $\mathbb{R}^n$, or all cubes in $\mathbb{R}^n$, etc.

We claim that for any t, the family $\mathcal{F}_t$ of all semialgebraic sets of complexity $\leq t$ has finite VC-dimension. To see this, consider the field of real numbers as a first-order structure $\mathcal{M} = (\mathbb{R}, +, \times, 0, 1, <)$. Note that $\mathcal{F}_t$ is contained in the union of finitely many families of the form $\left\{\mathcal{F}_{\phi_i(\bar{x},\bar{y})} : i < t'\right\}$ where t' only depends on t (since there are only finitely many different polynomials of degree $\leq t$, up to varying coefficients, and only finitely many different Boolean combinations of size $\leq t$). So it is enough to show that every such family has finite VC-dimension (by Remark 4.9).

By the classical result of Tarski, this structure $\mathcal{M}$ eliminates quantifiers, and so definable sets are precisely the semialgebraic ones. In particular, if we are given a formula of the form $\phi(x, \bar{y})$, for every $b \in M^{|\bar{y}|}$ the set $\phi(M, b)$ is just a union of at most n_ϕ intervals and points, where n_ϕ only depends on ϕ. As the collection of all intervals has finite VC-dimension, in view of Remark 4.9 we have that for all formulas $\phi(x, \bar{y})$ with $|x| = 1$, $\mathcal{F}_\phi$ has finite VC-dimension. By Theorem 4.10 this implies that the same is true for all formulas.

Example 4.12. Definable families in stable structures.

The class of stable structures is well studied in model theory, originating from Morley's theorem and Shelah's work on classification theory. See e.g. [1] for more details. Examples of stable structures:

- $(\mathbb{C}, \times, +, 0, 1)$ (definable sets correspond to the constructible sets, i.e. Boolean combinations of algebraic sets),
- separably closed and differentially closed fields,
- arbitrary planar graphs $G = (V, E)$,
- abelian groups (viewed as structures in the pure group language $(G, \cdot, 1)$),
- [Z. Sela] non-commutative free groups (in the pure group language).

Example 4.13. [8] Let $(G, \cdot, <)$ be an arbitrary ordered abelian group. Then definable families of sets have finite VC-dimension. In particular, all definable families in Presburger arithmetic $(\mathbb{Z}, +, <)$ have finite VC-dimension.

Example 4.14. Let $(\mathbb{Q}_p, \times, +, 0, 1)$ be the field of p-adics. Using the quantifier elimination result of Macintyre in this setting, one can show that again all definable families have finite VC-dimension.

4.3. The VC-theorem, ε-approximations and ε-nets.

Fact 4.15. *(Weak law of large numbers) Let $(\Omega, \mathcal{B}, \mathbb{P})$ be a probability space. Let $A \subseteq \Omega$ be an event and let $\varepsilon > 0$ be fixed. Then for any $n \in \mathbb{N}$ we have:*

$$\mathbb{P}^n \left((\omega_1, \ldots, \omega_n) \in \Omega^n : \left| \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_A (\omega_i) - \mathbb{P}(A) \right| \geq \varepsilon \right) \leq \frac{1}{4n\varepsilon^2}.$$

Note that this probability $\to 0$ as $n \to \infty$. In particular this means that fixing an arbitrary error ε, we can take n large enough so that with high probability the measure of A can be determined up to ε by picking n points at random and counting the proportion of them in A.

The key result in VC-theory is the theorem of Vapnik and Chervonenkis [20] demonstrating that a *uniform* version of the weak law of large numbers holds for *families of events* of finite VC-dimension. That is, with high probability sampling on a sufficiently long random tuple gives a good estimate for the measure of all sets in the family $\mathcal{F}$ simultaneously.

Let us fix some notation. For $S \in \mathcal{F}$ and $(x_1, \ldots, x_n) \in X^n$ we define

$$\mathrm{Av}(x_1, \ldots, x_n; S) := \frac{1}{n} \left| \{1 \leq i \leq n : x_i \in S\} \right|.$$

Theorem 4.16. *(VC-theorem) Let (X, μ) be a **finite** probability space, and $\mathcal{F} \subseteq \mathcal{P}(X)$ a family of subsets of X. Then for every $\varepsilon > 0$ we have*

$$\mu^n \left(\sup_{S \in \mathcal{F}} \left| \mathrm{Av}(x_1, \ldots, x_n; S) - \mu(S) \right| > \varepsilon \right) \leq 8 \pi_{\mathcal{F}}(n) \exp\left(-\frac{n\varepsilon^2}{32} \right).$$

Remark 4.17. Note that if $\mathrm{VC}(\mathcal{F}) = d$, then $\pi_{\mathcal{F}}(n) = O\left(n^d\right)$ and so the right part converges to 0 as n grows. Thus, as long as the VC-dimension of $\mathcal{F}$ is bounded, starting with $\mathcal{F}$ of arbitrary large finite size and an arbitrary measure, we still get an approximation up to an error ε for all sets in $\mathcal{F}$ by sampling on a random tuple of length depending just on d, ε.

Corollary 4.18. *Let $d \in \mathbb{N}$ and $\varepsilon > 0$ be arbitrary. Then there is some $N = N(d, \varepsilon) \in \mathbb{N}$ such that any set system $(X, \mathcal{F})$ on a finite probability space (X, μ) with $\mathrm{VC}(\mathcal{F}) \leq d$ admits an ε-**approximation** of size at most N.*

That is, there is a multi-set $\{x_1, \ldots, x_N\}$ of elements from X (repetitions are allowed) such that for all $S \in \mathcal{F}$ we have

$$\left| \mathrm{Av}(x_1, \ldots, x_N; S) - \mu(S) \right| \leq \varepsilon.$$

Proof. By Remark 4.17, it follows from Theorem 4.16 that for N large enough (with respect to d and ε), with high probability any N-tuple from X works as a ε-approximation (so in particular that is at least one N -tuple with this property). □

Remark 4.19.
1. Note that repetitions among the points $x_1, \ldots, x_n$ are necessary — think of a measure on a finite set, giving certain *different* weights to different points.
2. It is known that one can take $N = C \frac{1}{\varepsilon^2} \log \frac{1}{\varepsilon}$, where $C = C(d)$ is a constant.

Definition 4.20. Let V be a set, $\mathcal{B}$ a b.a. on V and μ a f.a.p. measure on $\mathcal{B}$. Let $\mathcal{F}$ be a family of subsets of V with $\mathcal{F} \subseteq \mathcal{B}$. As usual, for $\varepsilon > 0$ we say that a subset $T \subseteq V$ is an *ε-net for $\mathcal{F}$* if for every $F \in \mathcal{F}$ we have $\mu(F) \geq \varepsilon \Longrightarrow F \cap T \neq \emptyset$.

Note that every ε-approximation is an ε-net. One can get better bounds on the size of an ε-net ($|T| \leq 8d\frac{1}{\varepsilon}\log\frac{1}{\varepsilon}$).

We discuss arbitrary measures (with infinite support) that admit ε-approximations.

Definition 4.21. Let V, W be sets with b.a.'s $\mathcal{B}_V, \mathcal{B}_W$ on them, and let μ be a f.a.p. measure on $\mathcal{B}_V$.

1. Let $\mathcal{F}$ be a family of subsets of V in $\mathcal{B}_V$. We say that μ is *finitely approximable* on $\mathcal{F}$ if for every $\varepsilon > 0$ there are $p_1, \ldots, p_n \in V$ (possibly with repetitions) giving an *ε-approximation* of μ on $\mathcal{F}$.
2. Let $R \subseteq V \times W$ be such that $R_b \in \mathcal{B}_V$ for all $b \in W$. We say that μ is *fin.app. on* R if it is fin.app. on $\mathcal{F}_m$ for all $m \in \mathbb{N}$, where $\mathcal{F}_m$ is the family of all subsets of V given by the Boolean combinations of at most m sets of the form $R_b, b \in W$.

Remark 4.22. In particular, if μ is fin.app. on R, then it is fin.app. on the family $\mathcal{R}^{\Delta} := \{R_b \Delta R_{b'} : b, b' \in W\}$.

Example 4.23. (1) Any measure μon $\mathcal{B}_V$ with a finite support (i.e. there is some finite $B \in \mathcal{B}_V$ with $\mu(B) = 1$) is fin.app. on $\mathcal{B}_V$.

(2) Let $V = \mathbb{R}$, let $\mathcal{B}_V$ be the field generated by all intervals in V, and let $\mathcal{R}$ be the family of all intervals. Let μ be the $0-1$ measure on $\mathcal{B}_V$ such that the measure of a set is 1 if and only if it is unbounded from above. Then there are no finite ε-approximations for μ on $\mathcal{R}$, for any $\varepsilon < 1$, as any finite set can be avoided by some unbounded interval of measure 1. Note that $\mathrm{VC}(\mathcal{R}) < \infty$. This is not a contradiction with the VC-theorem as μ is not finitely supported.

(3) Let λ_n be the Lebesgue measure on the unit cube $[0,1]^n$ in $\mathbb{R}^n$. Let $\mathcal{M}$ be an o-minimal structure expanding $(\mathbb{R}, +, \times, 0, 1)$. If $X \subseteq \mathbb{R}^n$ is definable in $\mathcal{M}$, then $X \cap [0,1]^n$ is Lebesgue measurable (for $n = 1$ this is clear as every definable subset of $\mathbb{R}$ is just a finite union of intervals and points by o-minimality, and for $n > 1$ this follows from the o-minimal cell decomposition). Hence λ_n induces an f.a.p. measure on the b.a. of definable subsets of $\mathbb{R}^n$. This measure is fin.app. on every definable relation (Exercise! E.g. for $n = 1$ and $\varepsilon > 0$, we can take $\left\{\varepsilon i : 1 \leq i \leq \frac{1}{\varepsilon}\right\}$ as an ε-approximation for the family of intervals, etc.).

(4) Similarly, for every prime p, the (additive) Haar measure in $\mathbb{Q}_p$ normalized on a compact ball induces a f.a.p. measure on the b.a. of definable subsets (which are all measurable by the p-adic cell decomposition), and one can check that it is fin.app. on every definable relation.

The following example shows that the class of measures finitely approximable on families of bounded VC-dimension is closed under ultraproducts.

Proposition 4.24. *Let $(\mathcal{M}_i : i \in \mathbb{N})$ be $\mathcal{L}$-structures, let $\mathcal{B}_i$ be a b.a. of definable subsets of M_i and μ_i an f.a.p. measure on $\mathcal{B}_i$. Let $R_i \subseteq M_i \times M_i^k$ be definable with $R_i(x, c) \in \mathcal{B}_i$ for all $c \in M_i^k$. Assume that μ_i is fin.app. on R_i, and assume that*

*$\mathrm{VC}(R_i) \leq d$ for some **fixed** d and all $i \in \mathbb{N}$. Let $\mathcal{U}$ be a non-principal u.f. on $\mathbb{N}$, $\mathcal{M} = \prod_{i\in\mathbb{N}} \mathcal{M}_i/\mathcal{U}$ and $R = \prod_{i\in\mathbb{N}} R_i/\mathcal{U}$. Then $\mu = \lim_{\mathcal{U}} \mu_i$ is fin.app. on R.*

Proof. By Definition 4.21, we have to show that for every $m \in \mathbb{N}$, the family of all Boolean combinations of at most m fibers of R admits a finite ε-approximation. But by Remark 4.9 the VC-dimension of these family is uniformly bounded in terms of d, hence replacing R by the corresponding Boolean combination $R' \subseteq M_i \times M_i^{km}$ if necessary, it is enough to show that $\mathcal{F} := \left\{R(M,c) : c \in M^k\right\}$ admits a finite ε-approximation for every $\varepsilon > 0$.

Fix $\varepsilon > 0$ and $i \in \mathbb{N}$.

Let $\mathcal{F}_i := \left\{R_i(M_i,c) : c \in M_i^k\right\}$. By assumption $\mathrm{VC}(\mathcal{F}_i) \leq d$.

By assumption there is some $n_i \in \mathbb{N}$ and some $\left(a_1^i,\dots,a_{n_i}^i\right) \in M_i^{n_i}$ such that $\mu_i(X) \approx^{\varepsilon} \frac{1}{n_i}\sum_{j=1}^{n_i} \mathbf{1}_X\left(a_j^i\right)$ for all $X \in \mathcal{F}_i$.

Define $\mu_i' : \mathcal{B}_i \to \mathbb{R}$ by $\mu'(X) := \frac{1}{n_i}\sum_{j=1}^{n_i}\mathbf{1}_X\left(a_j^i\right)$ for any $X \in \mathcal{B}_i$. Then clearly μ_i' is a f.a.p. measure on $\mathcal{B}_i$ *supported on a finite set* $A_i := \bigcup_{j=1}^{n_i}\left\{a_j^i\right\}$ and $\mu_i'(X) \approx^{\varepsilon} \mu_i(X)$ for all $X \in \mathcal{F}_i$. By the VC-theorem (Theorem 4.18) there is some $n = n(d,\varepsilon)$ and $\left(b_1^i,\dots,b_{m_i}^i\right) \in A_i^{m_i}$ with $1 \leq m_i \leq n$ such that $\mu_i'(X) \approx^{\varepsilon} \frac{1}{m_i}\sum_{j=1}^{m_i}\mathbf{1}_X\left(b_j^i\right)$ for all $X \in \mathcal{F}_i$. Hence $\mu_i(X) \approx^{2\varepsilon} \frac{1}{m_i}\sum_{j=1}^{m_i}\mathbf{1}_X\left(b_j^i\right)$ for all $X \in \mathcal{F}_i$.

As $\mathcal{U}$ is an ultrafilter, there is some $S_1 \in \mathcal{U}$ and $1 \leq m \leq n$ such that $m_i = m$ for all $i \in S_1$. For $1 \leq j \leq m$, let b_j be an element of $\mathcal{M}$ defined by $b_j := \left(b_j^i : i \in \mathbb{N}\right)/\mathcal{U}$.

Claim. $b_1,\dots,b_m$ is a 3ε-approximation for μ on $\mathcal{F} := \left\{R(M,c) : c \in M^k\right\}$.

Let $c \in M^k$ be arbitrary, say $c = (c_i : i \in \mathbb{N})/\mathcal{U}$. We have:

(1) exists $S_2 \in \mathcal{U}$ such that $\mu(R(M,c)) \approx^{\varepsilon} \mu_i(R_i(M_i,c_i))$ for all $i \in S_2$ (by the definition of the ultralimit measure μ),
(2) exists $S_3 \in \mathcal{U}$ such that $\frac{1}{m}\sum_{j=1}^{m}\mathbf{1}_{R(M,c)}(b_j) = \frac{1}{m}\sum_{j=1}^{m}\mathbf{1}_{R_i(M_i,c_i)}\left(b_j^i\right)$ for all $i \in S_3$ (by Łoś theorem),
(3) $\mu_i(R_i(M_i,c_i)) \approx^{2\varepsilon} \frac{1}{m}\sum_{j=1}^{m}\mathbf{1}_{R_i(M_i,c_i)}\left(b_j^i\right)$ for all $i \in S_1$.

As $S_1 \cap S_2 \cap S_3 \neq \emptyset$, we have $\mu(R(M,c)) \approx^{3\varepsilon} \frac{1}{m}\sum_{j=1}^{m}\mathbf{1}_{R(M,c)}(b_j)$.

As $\varepsilon > 0$ was arbitrary, we can conclude. □

4.4. **Canonical products of finitely approximable measures.** As before, let $\mathcal{B}$ be a Boolean algebra on a set V, and let μ be an f.a.p. measure on $\mathcal{B}$.

Definition 4.25. A function $f : V \to \mathbb{R}$ is $\mathcal{B}$-*integrable* if for all $\varepsilon > 0$ there is a $\mathcal{B}$-simple function g with $|f(x) - g(x)| < \varepsilon$ for all $x \in V$.

Remark 4.26. A function $f : V \to \mathbb{R}$ is $\mathcal{B}$-integrable if and only if for any $\varepsilon > 0$ there are $Y_1,\dots,Y_n \in \mathcal{B}$ covering V such that for any $i \in [n]$ and any $c, c' \in Y_i$ we have $|f(c) - f(c')| < \varepsilon$.

If f is $\mathcal{B}$-integrable and μ is a f.a.p. measure on $\mathcal{B}$, then we define

$$\int_V f d\mu := \lim_{n\to\infty}\int_V g_n d\mu,$$

where $(g_n)_{n\in\mathbb{N}}$ is a sequence of $\mathcal{B}$-simple functions approximating f.

Exercise 4.27. This integral doesn't depend on the choice of a convergent sequence.

Also, for a $\mathcal{B}$-integrable f and a set $A \in \mathcal{B}$ we define

$$\int_A f d\mu := \int_V \mathbf{1}_A f d\mu.$$

Fact 4.28. [15, Theorem 4.4.13]

(1) *If f, g are integrable and $c, d \in \mathbb{R}$, then $cf + dg$ is integrable and for every $X \in \mathcal{B}$,*

$$\int_X (cf + dg)\, d\mu = c \int_X f d\mu + d \int_X g d\mu.$$

(2) *If f is integrable then $|f|$ is integrable and for every $X \in \mathcal{B}$,*

$$\left| \int_X d\mu \right| \leq \int_X |f|\, d\mu.$$

Our aim is, given two fin.app. measures, to define a certain canonical product measure which is fin.app. and forms a compatible system of measures.

For any set $A \in \mathcal{B}_V$, consider the function $h_{R,A} : W \to \mathbb{R}$ given by $h_{R,A}(b) = \mu(R_b \cap A)$.

Proposition 4.29. *Assume that μ is fin.app. on R (or just on $\mathcal{R}^\Delta = \{R_b \Delta R_{b'} : b, b' \in W\}$) and that $R_a \in \mathcal{B}_W$ for all $a \in V$. Then for any set $A \in \mathcal{B}_V$, the function $h_{R,A}$ is $\mathcal{B}_W$-integrable.*

Proof. Let $\varepsilon > 0$. By assumption we can choose $p_1, \ldots, p_n \in V$ such that

$$|\mu(R_b \Delta R_{b'}) - \mathrm{Av}(p_1, \ldots, p_n; R_b \Delta R_{b'})| < \varepsilon$$

for every $b, b' \in W$.

For $I \subseteq [n]$ let $C_I \subseteq W$ be the set $C_I = \{b \in W \colon p_i \in R_b \Leftrightarrow i \in I\} \in \mathcal{B}_W$. Clearly the sets $C_I, I \subseteq [n]$ cover W and for every $I \subseteq [n]$ and $b, b' \in C_I$ we have $\mu(R_b \Delta R_{b'}) < \varepsilon$. Hence, for any $b, b' \in C_I$ we have

$$|h_{R,A}(b) - h_{R,A}(b')| \leq \mu(A \cap (R_b \Delta R_{b'})) \leq \mu(R_b \Delta R_{b'}) < \varepsilon.$$

By Remark 4.26, the function $h_{R,A}$ is $\mathcal{B}_W$-integrable. □

Let now V, W, Z be sets and $R \subseteq V \times W \times Z$. Assume that $\mathcal{R}_V = \{R_{(b,c)} : (b, c) \in W \times Z\} \subseteq \mathcal{B}_V$ and $\mathcal{R}_W = \{R_{(a,c)} : (a, c) \in V \times Z\} \subseteq \mathcal{B}_W$. Let μ be a measure on $\mathcal{B}_V$ which is fin.app. on $R \subseteq V \times (W \times Z)$, and ν a measure on $\mathcal{B}_W$. Note that by assumption and Proposition 4.29 if E is an arbitrary R-definable subset of $V \times W$ (i.e. a Boolean combination of R-fibers) and $A \in \mathcal{B}_V$, then the function $h_{E,A}$ is $\mathcal{B}_W$-integrable. And $h_{E,A}(b) = \int_A \mathbf{1}_E(x, b) d\mu$. Hence the double integral

$$\omega_E(A, B) = \int_B \left(\int_A \mathbf{1}_E(x, y) d\mu \right) d\nu$$

is well defined for any $A \in \mathcal{B}_V, B \in \mathcal{B}_W$.

Let now $\mathcal{B}_{V \times W}$ be the b.a. on $V \times W$ generated by $\mathcal{B}_V \otimes \mathcal{B}_W$ and $\{R_c : c \in Z\}$. Then we have the following.

Proposition 4.30. (1) *There is a unique measure ω on $\mathcal{B}_{V \times W}$ whose restriction to $B_V \otimes B_W$ is $\mu \times \nu$ and such that $\omega(E \cap (A \times B)) = w_E(A, B)$ for every R-definable $E \subseteq V \times W$, $A \in \mathcal{B}_V, B \in \mathcal{B}_W$. We denote this measure by $\mu \ltimes \nu$.*

(2) *If in addition ν is fin.app. on R, then $\mu \ltimes \nu$ is also fin.app. on R and $\mu \ltimes \nu(E) = \nu \ltimes \mu(E)$ for all R-definable sets.*

Proof. (1) It is easy to see that every set Y in $\mathcal{B}_{V \times W}$ is a finite disjoint union of sets of the form $E_i \cap (A_i \times B_i)$ where E_i is an atom of the Boolean algebra of all R-definable subsets of $V \times W$ and $A_i \in \mathcal{B}_V, B_i \in \mathcal{B}_W$. We define $\omega(Y) = \sum \omega_{E_i}(A_i, B_i)$. It is easy to check that ω is well-defined (for all $A' \in \mathcal{B}_V, B' \in \mathcal{B}_W$ and R-definable $E' \subseteq V \times W$, if $(A \times B) \cap E = (A' \times B') \cap E'$, then $w_E(A, B) = w_{E'}(A', B')$) and is a f.a.p. measure on $\mathcal{B}_{V \times W}$ satisfying the requirements. Uniqueness is straightforward from the definition of ω.

(2) It is enough to show that $\mu \ltimes \nu$ is fin.app. on the family of all fibers of any R-definable relation $E \subseteq (V \times W) \times Z$. Fix an arbitrary $\varepsilon > 0$. Let us take $p_1, \dots p_n \in V$ such that $\mu(E_{b,c}) \approx^{\varepsilon} \mathrm{Av}(p_1, \dots, p_n; E_{b,c})$ for all $(b, c) \in W \times Z$, and $q_1, \dots, q_m \in W$ such that $\nu(E_{a,c}) \approx^{\varepsilon} \mathrm{Av}(q_1, \dots, q_m, E_{a,c})$ for all $(a, c) \in V \times Z$.

We claim that the set $\{(p_i, q_j) : 1 \le i < n, 1 \le j < m\}$ gives a 2ε-approximation for $\mu \ltimes \nu(E_c)$, for any $c \in Z$. Namely, using linearity of integration, we have

$$\mu \ltimes \nu(E_c) = \int_W \left(\int_V \mathbf{1}_{E_c}(v, w)\, d\mu \right) d\nu \approx^{\varepsilon}$$
$$\int_W \left(\frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{E_{w,c}}(p_i) \right) d\nu = \frac{1}{n} \sum_{i=1}^{n} \left(\int_W \mathbf{1}_{E_{w,c}}(p_i)\, d\nu \right) =$$
$$\frac{1}{n} \sum_{i=1}^{n} \left(\int_W \mathbf{1}_{E_{p_i,c}}(w) d\nu \right) \approx^{\varepsilon} \frac{1}{n} \sum_{i=1}^{n} \left(\frac{1}{m} \sum_{j=1}^{m} \mathbf{1}_{E_{p_i,c}}(q_j) \right) =$$
$$= \frac{1}{nm} \sum_{1 \le i \le n, 1 \le j \le m} \mathbf{1}_{E_c}(p_i, q_j),$$

so $\mu \ltimes \nu(E_c) \approx^{2\varepsilon} \mathrm{Av}(\{(p_i, q_j) : 1 \le i \le n, 1 \le j \le m\}; E_c)$.

The fact that $\mu \ltimes \nu(E_c) = \nu \ltimes \mu(E_c)$ follows as, by the above, for any $\varepsilon > 0$ we have

$$\mu \ltimes \nu(E_c) \approx^{2\varepsilon} \frac{1}{n} \sum_{i=1}^{n} \left(\frac{1}{m} \sum_{j=1}^{m} \mathbf{1}_{E_{p_i,c}}(q_j) \right) =$$
$$\frac{1}{m} \sum_{j=1}^{m} \left(\frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{E_{q_j,c}}(p_i) \right) \approx^{\varepsilon} \frac{1}{m} \sum_{j=1}^{m} \left(\int_V \mathbf{1}_{E_{q_j,c}}(v)\, d\mu \right) =$$
$$\int_V \left(\frac{1}{m} \sum_{j=1}^{m} \mathbf{1}_{E_{v,c}}(q_j) \right) d\mu \approx^{\varepsilon} \int_V \left(\int_W \mathbf{1}_{E_c}(v, w)\, d\nu \right) d\mu =$$
$$\nu \ltimes \mu(E_c),$$

hence $\mu \ltimes \nu(E_c) \approx^{4\varepsilon} \nu \ltimes \mu(E_c)$ for arbitrary $\varepsilon > 0$. □

It is not hard to see that a product of fin.app. measures satisfies a weak Fubini's property.

Corollary 4.31. *Let V, W be sets, μ a f.a.p. measure on $\mathcal{B}_V$ which is fin. app. on R, ν a f.a.p. measure on $\mathcal{B}_W$. For $\varepsilon > 0$ if $\mu(R_a) < \varepsilon$ for all $a \in W$ then $(\mu_V \ltimes \nu_W)(R) < \varepsilon$.*

We extend products of fap measures to an arbitrary number of sets.

Definition 4.32. Let $V_1, \ldots, V_k$ be sets, $R \subseteq V_1 \times \ldots \times V_k$ and assume that for each $i \in [k]$ we have a field $\mathcal{B}_i$ on V_i and a measure μ_i on $\mathcal{B}_i$ which is fin.app. on R (viewed as a binary relation on $V_i \times V_{[k]\setminus i}$). Then, by induction on k, we define a measure $\mu_1 \ltimes \ldots \ltimes \mu_k = (\mu_1 \ltimes \ldots \ltimes \mu_{k-1}) \ltimes \mu_k$ on $\mathcal{B}_{V_1} \times \ldots \times \mathcal{B}_{V_k}$ (and the order of integration doesn't matter by Proposition 4.30).

4.5. **Measure-theoretic regularity for hypergraphs of finite VC-dimension.**

Definition 4.33.
(1) Let $V_1, \ldots, V_k$ be sets, $R \subseteq V_1 \times \ldots \times V_k$ and $I \subseteq [k]$. We say that a subset $X \subseteq V_I$ is *R-definable over a set $D \subseteq V_{[k]\setminus I}$* if it is a finite Boolean combination of sets of the form R_b with $b \in D$, and say that X is *R-definable* if it is R-definable over $V_{[k]\setminus I}$.
(2) For a set $A \subseteq V_1 \times \ldots \times V_k$ we say that *A is $R_\otimes$-definable* if A can be written as a finite union of sets of the form $X_1 \times \ldots \times X_k$, such that each $X_i \subseteq V_i$ is R-definable.
In addition for a tuple $\vec{D} = (D_1, \ldots, D_k)$ with $D_i \subseteq V_{[k]\setminus i}$ we say that A is $R_\otimes$-definable over $\vec{D}$ if every X_i above is R-definable over D_i. For such a tuple $\vec{D}$ we use notation $\|\vec{D}\| = \max\{|D_i| \colon i \in [k]\}$.

Proposition 4.34. *Let $V, W, R \subseteq V \times W$ be sets, μ a f.a.p. measure on V which is fin.app. on R. Then for any $\varepsilon > 0$ there are R-definable subsets $X_1, \ldots X_m \subseteq W$ partitioning W such that for every $i \in [m]$ and any $a, a' \in X_i$ we have $\mu_V(R_a \Delta R_{a'}) < \varepsilon$.*

In addition, if the family $\mathcal{R} = \{R_a \colon a \in W\}$ has VC-dimension at most d then we can choose $D \subseteq V$ of size at most $320d(\frac{1}{\varepsilon})^2$ such that every X_i is R-definable over D.

Proof. We use the same trick as in the proof of Proposition 4.29.

Let $\mathcal{R}^\Delta = \{R_a \Delta R_{a'} \colon a, a' \in W\}$. Since μ is fin.app. on R, there are $p_1, \ldots p_n \in V$ with $|\mu(F) - \mathrm{Av}(p_1, \ldots, p_n; F)| < \varepsilon$ for any $F \in \mathcal{R}^\Delta$.

For each $I \cap [n]$ let $X_i = \{a \in W \colon p_i \in R_a \Leftrightarrow i \in I\}$. It is easy to see that the sets $X_I, I \subseteq [n]$ partition W, every X_i is R-definable and for every $I \subseteq [n]$ and $a, a' \in X_I$ we have $\mu(R_a \Delta R'_a) < \varepsilon$.

Assume in addition that $\mathcal{R}$ is a VC-family with VC-dimension at most d. As above we choose $p_1, \ldots p_n \in V$ with

$$|\mu(F) - \mathrm{Av}(p_1, \ldots, p_n; F)| < \varepsilon/2$$

for any $F \in \mathcal{R}^\Delta$.

Let ω be a measure on $\mathcal{B}_V$ given by $\omega(X) = \mathrm{Av}(p_1, \ldots, p_n; X)$. Since $\mathcal{R}$ has VC-dimension at most d, the family $\mathcal{R}^\Delta$ had dimension at most $10d$ by Remark 4.9, and by Corollary 4.18 we can choose an $\varepsilon/2$-net D for $\mathcal{R}^\Delta$ and ω with $|D| \leq 80d\frac{2}{\varepsilon}\log\frac{2}{\varepsilon}$. Clearly

$$80d\tfrac{2}{\varepsilon}\log\tfrac{2}{\varepsilon} \leq 80d\big(\tfrac{2}{\varepsilon}\big)^2 = 320d\big(\tfrac{1}{\varepsilon}\big)^2.$$

For each $I \cap D$ let $X_I = \{a \in W \colon R_a \cap D = I\}$. It is easy to see that the sets $X_I, I \subseteq D$, partition W and every X_i is R-definable over D. Let $I \subseteq D$ and $a, a' \in X_I$. Then $R_a \cap D = R_{a'} \cap D$, hence $w(R_a \Delta R_{a'}) \leq \varepsilon/2$, and $\mu(R_a \Delta R_{a'}) < \varepsilon$. □

Definition 4.35. For sets $V_1, \ldots, V_k$ and a set $R \subseteq V_1 \times \ldots \times V_k$ we say that R has *VC-dimension at most d* if for every $I \subseteq [k]$ the family $\{R_a \colon a \in V_{[k]\setminus I}\}$ of subsets of V_I has VC-dimension at most d.

Theorem 4.36. *Let $V_1, \dots V_k$ and $R \subseteq V_1 \times \dots \times V_k$ be sets, and $\mu_1, \dots, \mu_k$ f.a.p. measures on $V_1, \dots, V_k$, respectively, which are all fin. app. on R. Then for every $\varepsilon > 0$ there is an $R_\otimes$-definable $A \subseteq V_1 \times \dots \times V_k$ with*

$$(\mu_1 \ltimes \dots \ltimes \mu_k)(R \Delta A) < \varepsilon.$$

In addition, if R has VC-dimension at most d (see Definition 4.35) then we can choose A to be $R_\otimes$-definable over some $\vec{D}$ with $\|\vec{D}\| \leq C_{k,d}\left(\frac{1}{\varepsilon}\right)^{2(k-1)d}$, where $C_{k,d}$ is a constant that depends on k and d only.

Remark 4.37. Returning to our terminology from Section 3.2, this means in particular that R can be approximated up to measure ε by a set in $\mathcal{B}_{[k],1}$ — a finite union of boxes obtained by products of 1-ary sets.

Proof. We proceed by induction on k.

The case $k = 2$. Let V_1, V_2 and $R \subseteq V_1 \times V_2$ be given. Using Corollary 4.34 we can find R-definable sets $X_1, \dots X_m$ partitioning V_2 such that for every $i \in [m]$ and any $a, a' \in X_i$ we have $\mu_1(R_a \Delta R_{a'}) < \varepsilon$.

For each $i \in [m]$ we pick some $a_i \in X_i$ and let $A = \bigcup_{i \in [m]} R_{a_i} \times X_i$. Obviously A is $R_\otimes$-definable. It is not hard to see that for every $a \in W$ we have $\mu_1(R_a \Delta A_a) < \varepsilon$, hence, by Lemma 4.31, $(\mu_1 \ltimes \nu_2)(R \Delta A) < \varepsilon$.

Assume in addition that R has VC-dimension at most d. Then by Corollary 4.34, we can assume that for some $D_2 \subseteq V_1$ with $|D_2| \leq 320d\left(\frac{1}{\varepsilon}\right)^2$ every X_i is R-definable over D_2. Let $D_1 = \{a_1, \dots, a_m\}$, and $\vec{D} = (D_1, D_2)$. Obviously A is $R_\otimes$-definable over $\vec{D}$. By Sauer-Shelah lemma (Fact 4.8), $m < C_d |D_2|^d$, hence $|D_1| \leq C_d (320d)^d \left(\frac{1}{\varepsilon}\right)^{2d}$. And we can take $C_{2,d} = C_d (320d)^d$.

Inductive step $k+1$. Let $V_1, \dots, V_{k+1}$ and $R \subseteq V_1 \times \dots \times V_{k+1}$ be given.

Viewing $V_1 \times \dots \times V_{k+1}$ as $V_{[k]} \times V_{k+1}$ and using the case of $k = 2$ we obtain R-definable $X_1, \dots X_m$ partitioning V_{k+1} and points $a_i \in X_i$, $i \in [m]$, such that for the set $A' = \bigcup_{i \in [m]} R_{a_i} \times X_i$ we have $(\mu_1 \times \dots \times \mu_{k+1})(R \Delta A') < \varepsilon/2$.

For each $i \in [m]$ let $R^i = R_{a_i}$. It is an R-definable subset of $V_1 \times \dots \times V_k$. It is easy to see that each R^i has VC-dimension at most d. Applying induction hypothesis to each R^i we obtain $R^i_\otimes$-definable sets $A_i \subseteq V_1 \times \dots \times V_k$ such that $(\mu_1 \ltimes \dots \ltimes \mu_k)(R^i \Delta A_i) < \varepsilon/2$. Let $A = \bigcup_{i \in [m]} A_i \times X_i$. It is an $R_\otimes$-definable set and using Lemma 4.31, it is not hard to see that $(\mu_1 \ltimes \dots \ltimes \mu_{k+1})(A' \Delta A) < \varepsilon/2$, hence $(\mu_1 \ltimes \dots \ltimes \mu_{k+1})(R \Delta A) < \varepsilon$, as required.

Assume in addition that R has VC-dimension at most d. As in the case $k = 2$ we can assume that every X_i is R-definable over $D_{k+1} \subseteq V_1, \dots, V_k$ with $|D_{k+1}| \leq 320d\left(\frac{2}{\varepsilon}\right)^2$ and also assume that

$$m \leq C_d |D_{k+1}|^d \leq C_d \left[320d\left(\tfrac{2}{\varepsilon}\right)^2\right]^d = C_d (1280d)^d \left(\tfrac{1}{\varepsilon}\right)^{2d}.$$

Applying induction hypotheses we can assume that each A_i above is $R^i_\otimes$-definable over $\vec{D}^i = (D^i_1, \dots D^i_k)$ with $\|\vec{D}^i\| \leq C_{k,d}\left(\frac{2}{\varepsilon}\right)^{2(k-1)d}$, where $D^i_j \subseteq \prod_{l \in [k] \setminus \{j\}} V_l$.

For each $i \in [m]$ and $j \in [k]$ let $\bar{D}^i_j = \{(c, a_i) \colon c \in D^i_j\}$, $D_j = \bigcup_{i \in [m]} \bar{D}^i_j$, and $\vec{D} = (D_1, \dots, D_{k+1})$.

It is not hard to see that A above is R-definable over $\vec{D}$ and

$$\|\vec{D}\| \leq m C_{k,d}\left(\tfrac{2}{\varepsilon}\right)^{2(k-1)d} \leq C_d (1280d)^d \left(\tfrac{1}{\varepsilon}\right)^{2d} 2^{2(k-1)d} \left(\tfrac{1}{\varepsilon}\right)^{2(k-1)d} =$$

$$= C_{k+1,d}(\tfrac{1}{\varepsilon})^{2kd}.$$

□

Now we apply this product measure decomposition result to deduce a strong regularity lemma.

Definition 4.38. (1) For a k-hypergraph $E \subseteq V_1 \times \ldots \times V_k$ and $A_1 \subseteq V_1, \ldots, A_k \subseteq V_k$ we will denote by $E(A_1, \ldots, A_k)$ the set $E(A_1, \ldots, A_k) = E \cap A_1 \times \ldots \times A_k$

(2) By a *rectangular partition* we mean a k-tuple $\vec{\mathcal{P}} = (\mathcal{P}_1, \ldots, \mathcal{P}_k)$ where each $\mathcal{P}_i$ is a finite partition of V_i. For a rectangular partition $\vec{\mathcal{P}} = (\mathcal{P}_1, \ldots, \mathcal{P}_k)$ we define $\|\vec{\mathcal{P}}\| = \max\{|\mathcal{P}_i| \colon i \in [k]\}$, and for a set $X \subseteq V_1 \times \ldots \times V_k$ we write $X \in \vec{\mathcal{P}}$ if $X = X_1 \times \ldots \times X_k$ for some $X_i \in \mathcal{P}_i, i \in [k]$. We will also write $\Sigma \subseteq \vec{\mathcal{P}}$ to indicate that Σ consists of subsets $X \subseteq V_1 \times \ldots \times V_k$ with $X \in \vec{\mathcal{P}}$.

(3) For $A \subseteq V_1 \times \ldots \times V_k$ and a rectangular partition $\vec{\mathcal{P}} = (\mathcal{P}_1, \ldots, \mathcal{P}_k)$, say that A is *compatible with* $\vec{\mathcal{P}}$ if for any $X \in \vec{\mathcal{P}}$ either $X \subseteq A$ or $X \cap A = \emptyset$. In other words, A is a finite union of sets $X \in \vec{\mathcal{P}}$.

(4) A rectangular partition $\vec{\mathcal{P}}$ is *E-definable* (over $\vec{D} = (D_1, \ldots, D_k)$ as in Definition 4.33) if for each $i \in [k]$, every $X \in \mathcal{P}_i$ is E-definable over D_i.

(5) Let $\mathcal{B}_i$ be a bool. algebra on V_i, and μ_i a f.a.p. measures on $\mathcal{B}_i$ which is fin.app. on E, for all $i \in [k]$. Let $\mu := \mu_1 \ltimes \ldots \ltimes \mu_k$. Given $\varepsilon > 0$, a definable rectangular partition $\vec{\mathcal{P}}$ is *ε-regular with* $0-1$-*densities* if there is $\Sigma \subseteq \vec{\mathcal{P}}$ such that
$$\sum_{X \in \Sigma} \mu(X) \leq \varepsilon,$$
and for every $X_1 \times \ldots \times X_k \in \vec{\mathcal{P}} \setminus \Sigma$ either
$$\mu(Y_1 \times \ldots \times Y_k) - \mu(E(Y_1, \ldots, Y_k)) < \varepsilon\mu(X_1 \times \ldots \times X_k)$$
for all sets $Y_i \in \mathcal{B}_i$, $i = 1, \ldots, k$; or
$$\mu(E(Y_1, \ldots, Y_k)) < \varepsilon\mu(X_1 \times \ldots \times X_k)$$
for all sets $Y_i \in \mathcal{B}_i, i = 1, \ldots, k$.

The next proposition demonstrates how existence of an approximation by rectangular sets (Theorem 4.36) for the product measure can be used to obtain a regular partition.

Proposition 4.39. *(in the context of Definition 4.38) Let $\vec{\mathcal{P}}$ be a definable rectangular partition of $V_1 \times \ldots \times V_k$. If there is $A \subseteq V_1 \times \ldots \times V_k$, an $E_\otimes$-definable set compatible with $\vec{\mathcal{P}}$ with $\mu(A \Delta E) < \varepsilon^2$, then $\vec{\mathcal{P}}$ is ε-regular with $0-1$-densities.*

Proof. Let
$$\Sigma = \{X \in \vec{\mathcal{P}} \colon \mu(X \cap (A\Delta E)) \geq \varepsilon\mu(X)\}.$$
Since $\mu(A\Delta E) < \varepsilon^2$ and μ is finitely additive we obtain that
$$\sum_{X \in \Sigma} \mu(X) \leq \varepsilon.$$

Let $X = X_1 \times \ldots \times X_k \in \vec{\mathcal{P}} \setminus \Sigma$. We have

$$\mu(X \cap (A \Delta E)) < \varepsilon \mu(X).$$

Since A is compatible with $\vec{P}$ either $X \subseteq A$ or $X \cap A = \emptyset$.

Assume first $X \subseteq A$. Let $Y_i \subseteq X_i$ be from$\mathcal{B}_i, i = 1, \ldots, k$, and let $Y = Y_1 \times \ldots \times Y_k$. Since $Y \subseteq X$, by monotonicity of μ we have

$$\mu(Y \cap (A \Delta E)) < \varepsilon \mu(X).$$

As $Y \subseteq A$ we have $Y \cap (A \Delta E) = Y \setminus E(Y_1, \ldots, Y_k)$. Since $E(Y_1, \ldots, Y_k) \subseteq Y$ we also have

$$\mu(Y \setminus E(Y_1, \ldots, Y_k)) = \mu(Y) - \mu(E(Y_1, \ldots, Y_k)),$$

hence

$$\mu(Y_1 \times \ldots \times Y_k) - \mu(E(Y_1, \ldots, Y_k)) \leq \varepsilon \mu(X_1 \times \ldots \times X_k).$$

If $X \cap A = \emptyset$ similar arguments show that

$$\mu(E(Y_1, \ldots, Y_k)) < \varepsilon \mu(X_1, \ldots, X_k).$$

for all $Y_i \subseteq X_i$ from $\mathcal{B}_i, i = 1, \ldots, k$. □

Combining this observation with Theorem 4.36, we obtain a regularity lemma for hypergraphs of finite VC dimension.

Theorem 4.40. *Let $V_1, \ldots, V_k$ and $E \subseteq V_1 \times \ldots \times V_k$ be given, and let $\mu_1, \ldots, \mu_k$ be measures on $V_1, \ldots, V_k$ which are all fin.app. on E. Let $\mu = \mu_1 \ltimes \ldots \ltimes \mu_k$.*

For any $\varepsilon > 0$ there is an E-definable ε-regular partition $\vec{\mathcal{P}}$ with $0-1$-densities.

In addition, if E has VC dimension at most d we can choose $\vec{\mathcal{P}}$ with $\|\vec{\mathcal{P}}\| \leq C_d (C_{k,d})^d \left(\frac{1}{\varepsilon}\right)^{2(k-1)d^2}$, where C_d and $C_{k,d}$ are constants from Fact 4.8 and Theorem 4.36.

Proof. Using Theorem 4.36 there is an $E_\otimes$-definable A with $\mu(A \Delta E) < \varepsilon^2$. Say $A = \cup_{j \in [m]} A_1^j \times \ldots \times A_k^j$ where each $A_i^j \subseteq V_i$ is E-definable. □

For each $I \in [k]$ let $\mathcal{P}_i$ be the set of all atoms in the Boolean algebra generated by $A_i^1, \ldots, A_i^m$. Obviously each $\mathcal{P}_i$ consists of E-definable sets partitioning V_i, and A is compatible with $\vec{\mathcal{P}} = (\mathcal{P}_1, \ldots, \mathcal{P}_k)$. By Proposition 4.39, $\vec{\mathcal{P}}$ is ε-regular with $0-1$-densities.

Assume in addition that E has VC-dimension at most d. Then using Theorem 4.36 we can assume that A is $E_\otimes$-definable over $\vec{D} = (D_1, \ldots, D_k)$ with $|D_i| \leq C_{k,d}\left(\frac{1}{\varepsilon}\right)^{2(k-1)d}$ for $i \in [k]$. For each $i \in [k]$ let $\mathcal{P}_i$ be the set of all atoms in the Boolean algebra generated by E-definable over D_i subsets of V_i. Obviously each $\mathcal{P}_i$ consists of E-definable subsets partitioning V_i and A is compatible with $\vec{\mathcal{P}} = (\mathcal{P}_1, \ldots, \mathcal{P}_k)$. Also, by Sauer-Shelah (Fact 4.8),

$$|\,\mathcal{P}_i\,| \leq C_d |D_i|^d \leq C_d \Big(C_{k,d}\left(\tfrac{1}{\varepsilon}\right)^{2(k-1)d}\Big)^d = C_d (C_{k,d})^d \left(\tfrac{1}{\varepsilon}\right)^{2(k-1)d^2}$$

Remark 4.41. In the case when $V_{[k]}$ is finite the above theorem without the VC part is trivial, since we can take $\mathcal{P}_i$ to be the set of all atoms in the Boolean algebra of **all** E-definable subsets of V_i.

Let $E \subseteq V_1, \ldots, V_k$ be a *finite* k-hypergraph. For each $i \in [k]$ let μ_i be the counting measure on V_i, i.e. $\mu_i(X) = \frac{|X|}{|V_i|}$ and μ be the counting measure on $V_1 \times \ldots \times V_k$. Then all μ_i and μ are fin.app. measures with $\mu = \mu_1 \ltimes \ldots \ltimes \mu_k$. Hence all the results of the previous section can be applied to finite k-hypergraphs with respect to counting measures.

Corollary 4.42. *Assume $E \subseteq V_1 \times \ldots \times V_k$ has VC-dimension at most d.*

Then there are partitions $V_i = V_{i,1} \sqcup \cdots \sqcup V_{i,M}$ for some $M \leq c\left(\frac{1}{\varepsilon}\right)^{c'}$, where $c = c(k,d)$ and $c' = c'(k,d)$, numbers $\delta_{\vec{i}} \in \{0,1\}$ for $\vec{i} \in [M]^k$, and an exceptional set $\Sigma \subseteq [M]^k$ such that

$$\sum_{(i_1,\ldots,i_k,)\in\Sigma} |V_{1,i_1}| \cdots |V_{k,i_k}| \leq \varepsilon |V_1 \times \ldots \times V_k|$$

and for each $\vec{i} = (i_1, \ldots, i_k) \in [M]^k \setminus \Sigma$ we have

$$\big|\, |E(A_1, \ldots, A_k)| - \delta_{\vec{i}} |A_1| \cdots |A_k| \,\big| < \varepsilon |V_{1,i_1}| \cdots |V_{k,i_k}|$$

for all $A_1 \subseteq V_{1,i_1}, \ldots, A_k \subseteq V_{k,i_k}$.

Exercise 4.43. Formulate and show a converse (that this regularity lemma implies finiteness of the VC-dimension of the hypergraph).

4.6. **References.** *** TBA

5. Stable regularity lemma

We work in the same setting as before. Let the sets $V_1, \ldots, V_k$ and $R \subseteq V_1 \ldots \times \ldots V_k$ be given, let $\mathcal{B}_i$ be a b.a. on V_i, and let μ_i be a f.a.p. measure on $\mathcal{B}_i$. Assume moreover that for every $i \in [k]$, $R_b \in \mathcal{B}_i$ for all $b \in V_{[k]\setminus\{i\}}$.

Definition 5.1.
(1) A binary relation $R(x,y) \subseteq V \times W$ is *d-stable* if there is no tree of parameters $(b_\eta : \eta \in 2^{<d})$ in W such that for any $\eta \in 2^d$ there is some $a_\eta \in V$ such that $a_\eta \in R_{b_\nu} \iff \nu \frown 1 \trianglelefteq \eta$ (where $\trianglelefteq$ is the tree order).
(2) A relation $R \subseteq V_1 \times \ldots \times V_k$ is *d-stable* if for every $I \subseteq [k]$, viewed as a binary relation on $V_I \times V_{[k]\setminus I}$ it is d-stable.
(3) A relation R is *stable* if it is d-stable for some d.

Exercise 5.2.
(1) Alternatively, stability of a relation can be defined in terms of the so called *order property*. Namely, $R \subseteq V \times W$ has the d-order property if there are some elements a_i in V and b_i in W, $i = 1, \ldots, d$, such that $a_i \in R_{b_j} \iff i \leq j$ for all $1 \leq i,j \leq d$. Show that R is stable (in the sense of Definition 5.1) if and only if it does not have the d-order property for some d.
(2) Show that if R is d-stable, then $\mathrm{VC}(R) \leq d$.

Lemma 5.3. *Let R be a stable relation. Then any measure μ_i on $\mathcal{B}_i$ is fin.app. on R.*

Proof. Fix $i \in [k]$ and assume that R is d-stable.

Claim 1. For any $\varepsilon > 0$ there is some $m = m(\varepsilon, E)$ and some 0-1 measures $\delta_1, \ldots, \delta_m$ on $\mathcal{B}_i$ (possibly with repetitions) such that $\mu_i(R_c) \approx^{\varepsilon} \frac{1}{m}\sum_{j=1}^{m} \delta_j(R_c)$ for all $c \in V_{[k]\setminus\{i\}}$.

Proof. By Exercise 5.2, $\mathrm{VC}(R) \leq d$. Then the claim follows from the VC-theorem applied on the compact space of 0-1 measures on $\mathcal{B}_i$. See [9, Lemma 4.8] for the details (*** TBA).

Claim 2. Every $0-1$ measure δ on $\mathcal{B}_i$ is fin.app. on E.

Proof. This is a straightforward consequence of the explicit form of the definability of types in local stability. See e.g. the proof of [14, Lemma 2.2]: identifying our measure δ restricted to E with a complete E-type, an ε-approximation of δ on E is given by the $c_1, \ldots, c_m$ constructed in that proof, for any m large enough so that $\frac{N}{m} < \varepsilon$ (*** TBA).

Now, let $\varepsilon > 0$ be arbitrary, and let $\delta_1, \ldots, \delta_m$ be as given by Claim 1. By Claim 2, let A_j be a multiset in V_i giving an ε-approximation for δ_j. It is straightforward to verify that $A = \bigcup_{j=1}^{m} A_j$ is a 2ε-approximation for μ_i. □

In view of this lemma, for $I = \{i_1, \ldots, i_n\} \subseteq [k]$ we have a semi-direct product measure $\mu_I = \mu_{i_1} \ltimes \ldots \ltimes \mu_{i_n}$ on $\mathcal{B}_I = \mathcal{B}_{i_1} \times \ldots \times \mathcal{B}_{i_n}$ (see Definition 4.32) which is fin.app. on R (Proposition 4.30).

Definition 5.4. A set $A \in \mathcal{B}_I$ is ε-good if for any $b \in V_{[k]\setminus I}$, either $\mu_I(A \cap R_b) < \varepsilon\mu_I(A)$ or $\mu_I(A \cap R_b) > (1-\varepsilon)\mu_I(A)$.

Remark 5.5. Notice that if a set is ε-good then it has measure greater than 0.

Lemma 5.6. *Assume that $\mu_{[k]\setminus I}$ is fin.app. on R. For any $\varepsilon > 0$, consider the set*

$$A = \{a \in V_I : \mu_{[k]\setminus I}(R_a) < \varepsilon\}.$$

Then there is an R-definable set $A' \supseteq A$ such that $\mu_{[k]\setminus I}(R_a) < 2\varepsilon$ for all $a \in A'$.

Proof. Let $b_1, \ldots, b_n \in V_{[k]\setminus I}$ be such that $\mu_{[k]\setminus I}(R_a) \approx^{\frac{\varepsilon}{2}} \mathrm{Av}(b_1, \ldots, b_n; R_a)$ for all $a \in V_I$. Let $\mathcal{J} = \{J \subseteq [n] : \frac{|J|}{n} < \frac{3}{2}\varepsilon\}$, and let $A' = \bigcup_{J \in \mathcal{J}} \left(\bigcap_{j \in J} R_{b_j} \cap \bigcap_{j \notin J} R_{b_j}\right)$. It is easy to check that A' satisfies the requirements. □

Lemma 5.7. *Fix some $I \subseteq [k]$ and some $J \subseteq [k] \setminus I$. Let $B \in \mathcal{B}_J$ be an ε-good set, and let $A \in \mathcal{B}_I$ and $c \in V_{[k]\setminus(I\cup J)}$ be arbitrary, such that both A and B are of positive measure. Then (by Definition 5.4) A is a disjoint union of the sets*

$$A^0_{B,c} = \{a \in A : \mu_J(R_{a,c} \cap B) < \varepsilon\mu_J(B)\}$$

and

$$A^1_{B,c} = \{a \in A : \mu_J(R_{a,c} \cap B) > (1-\varepsilon)\mu_J(B)\}.$$

Assume that $\varepsilon < \frac{1}{4}$. Then $A^0_{B,c}, A^1_{B,c} \in \mathcal{B}_I$.

Proof. Indeed, let μ'_I be the restriction of μ_I to A and let μ'_J be the restriction of μ_J to B. As R is stable, by Lemma 5.3 both μ'_I, μ'_J are fin.app. on R. Hence, by Lemma 5.6 applied to μ'_I, μ'_J we can find some R-definable $A'_0 \supseteq A^0_{B,c}, A'_1 \supset A^1_{B,c}$ such that $\mu'_J(R_{a,c}) < 2\varepsilon$ for all $a \in A'_0$ and $\mu'_J(R_{a,c}) > (1-2\varepsilon)$ for all $a \in A'_1$ (here we have applied it to the complement $\neg R$, which is also d-stable). As $\varepsilon < \frac{1}{4}$, it follows that in fact $A^0_{B,c} = A'_0 \cap A, A^1_{B,c} = A'_1 \cap A$. □

In particular, it makes sense to speak of the μ_I-measure of $A^0_{B,c}, A^1_{B,c}$.

Definition 5.8. Let $0 < \varepsilon < \frac{1}{4}$ be arbitrary, and let $I \subseteq [k]$. We say that a set $A \in \mathcal{B}_I$ is ε-*excellent* if it is ε-good and for every $J \subseteq [k] \setminus I$, every ε-good $B \in \mathcal{B}_J$ and every $c \in V_{[k]\setminus(I\cup J)}$, either $\mu_I(A^0_{B,c}) < \varepsilon\mu_I(A)$ or $\mu_I(A^1_{B,c}) < \varepsilon\mu_I(A)$ (in the notation from Lemma 5.7).

References

[1] Artem Chernikov. Lecture notes on stability theory (math 285D). *http://www.math.ucla.edu/~chernikov/teaching/StabilityTheory285D/StabilityNotes.pdf*.

[2] Artem Chernikov. Lecture notes on topics in combinatorics (math 285N). *http://www.math.ucla.edu/~chernikov/teaching/Combinatorics285N/CombinatoricsNotes.pdf*,.

[3] Lester E Dubins. An elementary proof of bochner's finitely additive radon-nikodym theorem. *The American Mathematical Monthly*, 76(5):520–523, 1969.

[4] Jacob Fox and László Miklós Lovász. A tight lower bound for szemerédi?s regularity lemma. *preprint*, 2014.

[5] Peter Frankl and Vojtěch Rödl. Extremal problems on set systems. *Random Structures & Algorithms*, 20(2):131–164, 2002.

[6] Isaac Goldbring and Henry Towsner. An approximate logic for measures. *Israel Journal of Mathematics*, 199(2):867–913, 2014.

[7] William T Gowers. Lower bounds of tower type for szemerédi's uniformity lemma. *Geometric & Functional Analysis GAFA*, 7(2):322–337, 1997.

[8] Yuri Gurevich and Peter H Schmitt. The theory of ordered abelian groups does not have the independence property. *Transactions of the American Mathematical Society*, 284(1):171–182, 1984.

[9] Ehud Hrushovski and Anand Pillay. On nip and invariant measures. *Journal of the European Mathematical Society*, 13(4):1005–1061, 2011.

[10] H Jerome Keisler. The ultraproduct construction. *Ultrafilters Across Mathematics,* **http://www.math.wisc.edu/~keisler/ultraproducts-web-final.pdf**, 530:163–179, 2010.

[11] János Komlós and Miklós Simonovits. Szemerédi's regularity lemma and its applications in graph theory. 1996.

[12] Jerzy Łoś and Edward Marczewski. Extensions of measure. *Fundamenta Mathematicae*, 36(1):267–276, 1949.

[13] Guy Moshkovitz and Asaf Shapira. A short proof of gowers? lower bound for the regularity lemma. *Combinatorica*, pages 1–8, 2013.

[14] Anand Pillay. *Geometric stability theory*. Number 32. Oxford University Press, 1996.

[15] KPS Bhaskara Rao and M Bhaskara Rao. *Theory of charges: a study of finitely additive measures*, volume 109. Academic Press, 1983.

[16] Pierre Simon. *A guide to NIP theories*, volume 44. Cambridge University Press, 2015.

[17] Terence Tao. A variant of the hypergraph removal lemma. *Journal of combinatorial theory, Series A*, 113(7):1257–1280, 2006.

[18] Terence Tao. A correspondence principle between (hyper) graph theory and probability theory, and the (hyper) graph removal lemma. *Journal d'Analyse Mathématique*, 103(1):1–45, 2007.

[19] Henry Towsner. An analytic approach to sparse hypergraphs: hypergraph removal. *Preprint, arXiv:1204.1884*, 2012.

[20] Vladimir N Vapnik and A Ya Chervonenkis. On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability & Its Applications*, 16(2):264–280, 1971.