ERGODIC TRANSFERENCE THEOREM

TERENCE TAO

Abstract. We present Furstenberg's correspondence principle in the context of additive, and not-quite-additive, weights on the cube $\{0,1\}^{\mathbf{Z}}$, and show how this can be used to prove a weak version of the transference principle used in the establishment of long arithmetic progressions in the primes. It turns out that one can deduce this principle without any appeal to the axiom of choice.

1. The doubly infinite cube

Consider the doubly infinite discrete cube $\{0,1\}^{\mathbf{Z}}$. One can think of $\{0,1\}^{\mathbf{Z}}$ as the space of all doubly infinite sequences $(a_n)_{n\in\mathbf{Z}}$ which are binary (i.e. they take values in $\{0,1\}$). For any disjoint finite subsets $A, B \subset \mathbf{Z}$, define the cylinder $C(A,B) \subset \{0,1\}^{\mathbf{Z}}$ to be the space of all sequences $(a_n)_{n\in\mathbf{Z}}$ which equal 1 on A and 0 on B; thus for instance

$$C(\emptyset, \emptyset) = \{0,1\}^{\mathbf{Z}} \tag{1}$$

and we have the decomposition identity

$$C(A,B) = C(A \cup \{n\}, B) \uplus C(A, B \cup \{n\}) \tag{2}$$

for all $n \notin A \cup B$, where $\uplus$ denotes disjoint union. Also observe that the intersection of two cylinders is again a cylinder or the empty set. Furthermore, if $A \cup B \subset A' \cup B'$, then $C(A', B')$ is either disjoint from, or contained in $C(A,B)$.

The cube $\{0,1\}^{\mathbf{Z}}$ has the usual product topology generated by the cylinders (thus an open set is nothing more than an arbitrary union of cylinders), and then it has the Borel σ-algebra $\mathcal{B}$ generated by the open sets. Inside $\mathcal{B}$, we also have $\mathcal{A}$, the set of finite unions of cylinders; this is an algebra but not a σ-algebra. Let us call the elements in $\mathcal{A}$ *elementary sets*; these are the sets which can be described using only finitely many elements of the sequence $(a_n)_{n\in\mathbf{Z}}$. Every elementary set in $\mathcal{A}$ is both open and closed. Every element is also compact:

Theorem 1.1 (Tychonoff's theorem for $\{0,1\}^{\mathbf{Z}}$)**.** *Suppose that an elementary set $E \subset \mathcal{A}$ is covered by a family $\mathcal{E} \subset \mathcal{A}$ of other elementary sets. Then it is in fact covered by a finite sub-family of $\mathcal{E}$.*

Proof This follows immediately from the general Tychonoff's theorem, but we give a proof here to emphasize that for this specific application of Tychonoff, we do

1991 *Mathematics Subject Classification.* ???

not need the axiom of choice[1]. Since E is the finite union of cylinders, it suffices to verify the claim when E is a cylinder, say $E = C(A_0, B_0)$. Similarly, by breaking up all the elementary sets in $\mathcal{E}$ into cylinders, it suffices to verify the claim when $\mathcal{E}$ also consists entirely of cylinders[2]. For any integer $N \geq 1$, let E_N be the set

$$E_N := E \backslash \bigcup \{C(A,B) \in \mathcal{E} : A \cup B \subseteq \{-N, \dots, N\}\}$$

thus E_N is the set formed by deleting all the cylinders arising solely from the positions between $-N, \dots, N$. Then we have

$$E \supseteq E_1 \supseteq E_2 \supseteq \dots$$

and $\bigcap_{N=1}^{\infty} E_N = \emptyset$, by hypothesis.

Call a cylinder $C(A,B)$ *intersective* if it has a non-empty intersection with infinitely many E_N. If the set $E = C(A_0, B_0)$ is not intersective, then at least one of the E_N is empty, which means that E can be covered by finitely many cylinders from $\mathcal{E}$, as desired. Thus we may assume for sake of contradiction that $C(A_0, B_0)$ is intersective. Next, observe from (2) that if $C(A,B)$ is intersective and $n \notin A \cup B$, then at least one of $C(A \cup \{n\}, B)$ or $C(A, B \cup \{n\})$ is also intersective. Thus by a recursive construction, we can construct a nested sequence of intersective cylinders

$$C(A_0, B_0) \supset C(A_1, B_1) \supset C(A_2, B_2) \supset \dots$$

where each $C(A_j, B_j)$ is formed from $C(A_{j-1}, B_{j-1})$ by adding another integer n to either[3] A_{j-1} or B_{j-1}. We can also ensure (again without axiom of choice) that every integer n gets eventually added to either A_j or B_j, thus the sets $A_\infty := \bigcup_j A_j$ and $B_\infty := \bigcup_j B_j$ partition $\mathbf{Z}$. Now let $a = (a_n)_{n \in \mathbf{Z}}$ be the sequence which equals 1 on A_∞ and 0 on B_∞, thus a lies inside all the intersective cylinders $C(A_j, B_j)$ and in particular lies in E. Since $\mathcal{E}$ covers E, there must be a cylinder $C(A', B')$ in $\mathcal{E}$ which contains a. But then this entire cylinder will lie outside E_N for all sufficiently large N. On the other hand, for sufficiently large j, $A_j \cup B_j$ will contain $A' \cup B'$, and so $C(A_j, B_j)$, since it shares the common element a with $C(A', B')$, must be entirely contained in $C(A', B')$. But this means that $C(A_j, B_j)$ is not intersective, a contradiction. ■

We now consider how the cube $\{0,1\}^{\mathbf{Z}}$ can be interpreted as a dynamical system. We already have a Borel σ-algebra $\mathcal{B}$ on the cube, but we also need a notion of a shift $T : \{0,1\}^{\mathbf{Z}} \to \{0,1\}^{\mathbf{Z}}$, and some sort of measure (or something resembling a measure) on $\mathcal{B}$. The shift is easy to define, we shall take the right-shift $T : \{0,1\}^{\mathbf{Z}} \to \{0,1\}^{\mathbf{Z}}$, defined by $T(a_n)_{n \in \mathbf{Z}} := (a_{n-1})_{n \in \mathbf{Z}}$. This right-shift induces a map on the σ-algebra $\mathcal{B}$, for instance we have

$$TC(A,B) = C(A+1, B+1). \tag{3}$$

[1]The point is that we have an obvious way to choose between 0 and 1; to paraphrase Bertrand Russell, we have here infinitely many pairs of shoes, as opposed to infinitely many pairs of socks.

[2]It is easy to well-order the space of elementary sets without using the axiom of choice, so one can assign to each cylinder in the new $\mathcal{E}$, an elementary set in the original $\mathcal{E}$ which contains the cylinder, without using choice.

[3]Again, one can do this without the axiom of choice, for instance by always choosing $C(A \cup \{n\}, B)$ in favour of $C(A, B \cup \{n\})$ whenever there is a choice.

We can define integer powers T^h of T in the usual manner. Note that the σ-algebra $\mathcal{B}$ can now be generated by a single cylinder, e.g. the standard cylinder $E_0 := C(\{0\}, \emptyset)$, together with the shift T, indeed we have

$$C(A,B) = \bigcap_{h \in A} T^h E_0 \cap \bigcap_{l \in B} T^h E_0^c \quad (4)$$

where of course we use E_0^c to denote the complement of E_0.

The remaining question is how to build a measure. It turns out to be convenient to have a rather weak notion of measure.

Definition 1.2. A *weight* on the cube $\{0,1\}^{\mathbf{Z}}$ is any function $\mu : C(A,B) \to \mathbf{R}^+$ from the cylinders to the non-negative reals. We define a *sequence of weights* to be any collection $(\mu_N)_{N \in I}$, where N ranges over an infinite set I of natural numbers. We give the space of weights the pointwise (i.e. product) topology, thus μ_N converges to μ if and only if $\mu_N(C(A,B))$ converges to $\mu(C(A,B))$ for all cylinders $C(A,B)$.

- We say that a weight μ is *additive* if we have
$$\mu(C(A,B)) = \mu(C(A \cup \{n\}, B)) + \mu(C(A, B \cup \{n\}) \quad (5)$$
for all disjoint finite $A, B \in \mathbf{Z}$ and all $n \notin A \cup B$.
- We say that a weight μ is *shift-invariant* if we have
$$\mu(C(A+1, B+1)) = \mu(C(A,B)) \quad (6)$$
for all cylinders $C(A,B)$.
- We say that a sequence $(\mu_N)_{N \in I}$ of weights is *asymptotically finite* if we have $\limsup_{N \to \infty} \mu_N(C(A,B)) < \infty$ for all cylinders $C(A,B)$. In this case, we define the weight $\overline{\mu} = \limsup_{N \to \infty} \mu_N$ by the formula $\overline{\mu}(C(A,B)) := \limsup_{N \to \infty} \mu_N(C(A,B))$.
- We say that a sequence $(\mu_N)_{N \in I}$ of weights is *asymptotically additive* if we have
$$\lim_{N \to \infty} [\mu_N(C(A,B)) - \mu_N(C(A \cup \{n\}, B)) - \mu_N(C(A, B \cup \{n\})] = 0 \quad (7)$$
for all disjoint finite $A, B \in \mathbf{Z}$ and all $n \notin A \cup B$.
- We say that a sequence $(\mu_N)_{N \in I}$ of weights is *asymptotically shift-invariant* if we have
$$\lim_{N \to \infty} [\mu_N(C(A+1, B+1)) - \mu_N(C(A,B))] = 0. \quad (8)$$

Example 1.3. Let $(X', \mathcal{B}', \mu')$ be a finite measure space, let $T' : X' \to X'$ be a measurable map whose inverse $(T')^{-1}$ is also measurable, and let E' be a measurable set in X'. Then we can define the additive weight

$$\mu(C(A,B)) := \mu'(\bigcap_{h \in A} (T')^h E' \cap \bigcap_{l \in B} (T')^h (E')^c).$$

Observe that if the shift T' preserves the measure μ', then the weight μ is also shift-invariant.

Example 1.4. Let E be a set of integers. Then we can define the sequence of weights

$$\mu_N(C(A,B)) := \frac{1}{[-N,N]}|\bigcap_{h\in A}(E+h)\cap\bigcap_{l\in B}(E^c+l)\cap[-N,N]|$$

where $[-N,N] := \{-N,\dots,N\}$. This sequence is asymptotically finite, asymptotically additive, and asymptotically shift-invariant. Observe that

$$\overline{\mu}(C(A,B)) = \overline{d}(\bigcap_{h\in A}(E+h)\cap\bigcap_{l\in B}(E^c+l))$$

where $\overline{d}(E)$ is the upper density of a set E. One can of course generalize this example to other groups than $\mathbf{Z}$, with Folner sequences playing the role of the intervals $[-N,N]$.

Example 1.5. Suppose that we have a sequence N of integers going to infinity, and for each N we have a set $E_N \subset \mathbf{Z}/N\mathbf{Z}$. We can then define the sequence of weights

$$\mu_N(C(A,B)) := \frac{1}{N}|\bigcap_{h\in A}(E_N+h)\cap\bigcap_{l\in B}(E_N^c+l)|.$$

This sequence is asymptotically finite, asymptotically additive, and asymptotically shift-invariant (in fact each individual μ_N is additive and shift-invariant.

Example 1.6. Suppose that we have a sequence N of integers going to infinity, and for each N we have functions $f_N, g_N : \mathbf{Z}/N\mathbf{Z} \to \mathbf{R}^+$. We can then define the sequence of weights

$$\mu_N(C(A,B)) := \mathbf{E}((\prod_{h\in A} f_N(x+h))(\prod_{l\in B} g_N(x+l))|x \in \mathbf{Z}_N);$$

this generalizes the preceding example, which corresponds to the case $f_N := 1_{E_N}$ and $g_N := 1 - 1_{E_N}$. Here we are using the averaging notation $\mathbf{E}(f(x)|x \in A) := \frac{1}{|A|}\sum_{x\in A} f(x)$ from [3]. Later we will consider the case $f_N := 1_{E_N}\nu_N$ and $g_N := (1-1_{E_N})\nu_N$, where $\nu_N : \mathbf{Z}/N\mathbf{Z} \to \mathbf{R}^+$ is a pseudorandom measure. This sequence of weights is asymptotically shift-invariant (indeed each individual μ_N is shift-invariant), and if in addition we have $f_N + g_N = 1$ then it is also asymptotically additive and asymptotically finite. Later we shall see that we still can recover asymptotic additivity and finiteness (after a dilation) if $f_N + g_N$ is not equal to 1, but instead is equal to a sufficiently pseudorandom measure.

Observe that if μ_N converges to μ, then the μ_N are asymptotically additive if and only if μ is additive, and the μ_N are asymptotically shift-invariant if and only if μ is shift-invariant. They are also automatically asymptotically finite. Of course, not all sequences converge; however, we have

Lemma 1.7 (Arzela-Ascoli theorem)**.** *Every asymptotically finite sequence μ_N contains a convergent subsequence.*

Proof For any given cylinder $C(A,B)$, the sequence $\mu_N(C(A,B))$ of non-negative real numbers is bounded and thus has a convergent subsequence. Since the number of cylinders is countable, the claim then follows from the standard Arzela-Ascoli argument. Note that this argument is completely constructive and does not require the axiom of choice; one could also proceed using choice-dependent tools such as ultrafilters or the Banach-Alaoglu theorem. ∎

We also observe that asymptotic finiteness is easy to verify if one has asymptotic additivity:

Lemma 1.8. *Let μ_N be asympotically additive. Then μ_N is asymptotically finite if and only if $\limsup_{N\to\infty} \mu_N(C(\emptyset,\emptyset))$ is finite.*

Proof From multiple applications of asymptotic additivity and positivity we see that

$$\limsup_{N\to\infty} \mu_N(C(A,B)) \leq \limsup_{N\to\infty} \mu_N(C(\emptyset,\emptyset))$$

for all cylinders $C(A,B)$, and the claim follows. ∎

If a weight is shift-invariant, then it is easy to see that one has the more general statement

$$\mu(C(A+h,B+h)) = \mu(C(A,B))$$

for all $h \in \mathbf{Z}$. Similarly, if a weight is additive, one sees that

$$\sum_{C(A,B)\in\mathcal{C}} \mu(C(A,B)) = \sum_{C(A,B)\in\mathcal{C}'} \mu(C(A,B)) \tag{9}$$

whenever $\mathcal{C}$ and $\mathcal{C}'$ are two partitions of the same elementary set E into cylinders; this is easiest to see by comparing both $\mathcal{C}$ and $\mathcal{C}'$ to some common refinement $\mathcal{C}''$. One can thus extend an additive weight uniquely to elementary sets E. By Theorem 1.1 this weight is not only additive, it is a premeasure (i.e. we have $\mu(E) = \sum_n \mu(E_n)$ whenever the elementary set E is partitioned into at most countably many elementary sets E_n). Applying Carathéodory's theorem, we thus conclude

Theorem 1.9 (Kolmogorov extension theorem)**.** *Let μ be a weight. Then μ extends to a finite Borel measure on $\{0,1\}^{\mathbf{Z}}$ if and only if μ is additive. Furthermore this extension is unique, and will be shift-invariant if and only if the original weight μ was shift invariant.*

Note that Carathéodory's theorem is completely constructive, and thus this theorem does not require the axiom of choice. Combining the Kolmogorov extension theorem with the Arzela-Ascoli theorem, we obtain

Theorem 1.10 (Furstenberg correspondence principle)**.** *Let $(\mu_N)_{N\in I}$ be an asymptotically finite, asymptotically additive, asymptotically shift-invariant sequence of weights. Let $\overline{\mu} := \limsup_{N\to\infty} \mu_N$ be the limit superior of the μ_N. Then there exists a shift-invariant finite Borel measure μ on $\{0,1\}^{\mathbf{Z}}$ such that*

$$\mu(C(\{0\},\emptyset)) = \overline{\mu}(C(\{0\},\emptyset))$$

and

$$\mu(C(A,B)) \leq \overline{\mu}(C(A,B))$$

for all cylinders $C(A,B)$.

Proof By passing to a subsequence of N if necessary, we may reduce to the case where $\mu_N(C(\{0\},\emptyset))$ actually converges to the limit superior $\overline{\mu}(C(\{0\},\emptyset))$ (note that this passage to a subsequence may lower some of the other limit superiors, but this will not harm us). Applying the Arzela-Ascoli theorem, we may then assume that the μ_N converge to a weight μ, which is then additive and shift-invariant. The claim then follows from the Kolmogorov extension theorem. ■

Again, observe that the Furstenberg correspondence principle does not actually require the axiom of choice (though it is often proved using this axiom in the literature). Next, we recall the deep multiple recurrence theorem of Furstenberg:

Theorem 1.11 (Furstenberg recurrence theorem)**.** [1], [2] *Let μ be a shift-invariant finite Borel measure on $\{0,1\}^{\mathbf{Z}}$ such that $\mu(C(0,\emptyset)) > 0$, and let $k \geq 1$. Then we have*

$$\mu(C(\{0,h,\dots,(k-1)h\},\emptyset)) > 0$$

for all h in a set of integers of positive upper density.

This theorem is only stated for the cube $\{0,1\}^{\mathbf{Z}}$, but extends automatically to all other measure-preserving systems thanks to Example 1.3. Again, while the proof in [2] uses the axiom of choice, the original proof in [1] does not.

Combining this with the correspondence principle and Lemma 1.8, we obtain

Theorem 1.12 (Szemerédi's theorem)**.** *Let $(\mu_N)_{N\in I}$ be an asymptotically additive, asymptotically shift-invariant sequence of weights with limit superior $\overline{\mu}$ such that*

$$\overline{\mu}(C(\emptyset,\emptyset)) < \infty$$

and

$$\overline{\mu}(C(\{0\},\emptyset)) > 0,$$

and let $k \geq 1$. Then we have

$$\overline{\mu}(C(\{0,h,\dots,(k-1)h\},\emptyset)) > 0$$

for all h in a set of integers of positive upper density.

Combining this with Example 1.4 we obtain the standard formulation of Szemerédi's theorem, namely that any subset of integers of positive density contains arbitrarily long arithmetic progressions (and that these progressions in fact have "positive density" in a certain sense). However this is not the only class of asymptotically additive, asymptotically shift-invariant weights available, and we will later apply this theorem to another setting involving a pseudorandom measure.

The Furstenberg recurrence theorem has many extensions, generalizations, and refinements, which can then be converted via the correspondence principle to Szemerédi type theorems, which we will not discuss here.

2. The Varnavides averaging trick

In the ergodic approach to recurrence and Szemerédi theorems, the shift h is often held fixed while the discretization parameter N goes to infinity, and then only after taking limits as $N \to \infty$ does one allow h to go to infinity. Thus, the ergodic regime occurs when N is extremely large compared with h. In contrast, in the combinatorial and Fourier-analytic approach to Szemerédi's theorem is typically located in the cyclic group $\mathbf{Z}/N\mathbf{Z}$, and the shift h is often of comparable size with N.

Thus the two asymptotic regimes of the ergodic approach and the combinatorial approach are genuinely different. Fortunately, there is a simple averaging argument of Varnavides [4] which (partially) connects the two. The idea is that if an ensemble of weights has a nice property on the average, then one can select a representative of that weight which exhibits that property without averaging. We give an abstract formulation of this argument as follows.

Lemma 2.1 (Abstract Varnavides argument)**.** *Let Σ be a countable index set, I be a set of positive integers, and for each $N \in I$ let Λ_N be a finite non-empty index set. Suppose for each $\sigma \in \Sigma$, $N \in I$, and $\lambda \in \Lambda_N$ we are given a non-negative number $c_{\sigma,N,\lambda}$, which is asymptotically zero on the average in the sense that*

$$\lim_{N\to\infty} \mathbf{E}(c_{\sigma,N,\lambda}|\lambda \in \Lambda_N) = 0 \text{ for all } \sigma \in \Sigma. \tag{10}$$

Then we can choose a $\lambda_N \in \Lambda_N$ for all $N \in I$ such that

$$\lim_{N\to\infty} c_{\sigma,N,\lambda_N} = 0 \text{ for all } \sigma \in \Sigma. \tag{11}$$

Proof Without loss of generality we may take Σ to be the natural numbers $\Sigma = \{1, 2, \dots\}$. From (10) and linearity of expectation we see that

$$\lim_{N\to\infty} \mathbf{E}(\sum_{N\le w} c_{\sigma,N,\lambda}|\lambda \in \Lambda_N) = 0$$

for all $w \ge 0$. Thus if we choose $w(N)$ to be an integer-valued function of N which goes to infinity sufficiently slowly, we have

$$\lim_{N\to\infty} \mathbf{E}(\sum_{\sigma\le w(N)} c_{\sigma,N,\lambda}|\lambda \in \Lambda_N) = 0.$$

By the pigeonhole principle[4] we can thus find a λ_N for each N such that

$$\lim_{N\to\infty} \sum_{\sigma\le w(N)} c_{\sigma,N,\lambda_N} = 0.$$

Since the c_{σ,N,λ_N} are non-negative, and $w(N)$ will eventually exceed any given σ for N large enough, we obtain (11). ∎

As an immediate corollary of this abstract principle, we obtain

[4] Note that one will not need the axiom of choice here, if each Λ_N comes with a well-ordering, which will be the case for our application.

Corollary 2.2. *Let I be a set of numbers, and for each $N \in I$ and each $\lambda \in \mathbf{Z}/N\mathbf{Z}$ let $\mu_{N,\lambda}$ be a weight. Suppose also that the $\mu_{N,\lambda}$ are asymptotically shift-invariant on the average, in the sense that*

$$\lim_{N\to\infty} \mathbf{E}(|\mu_{N,\lambda}(C(A+1,B+1)) - \mu_{N,\lambda}(C(A,B))|\big|\lambda \in \mathbf{Z}/N\mathbf{Z}) = 0$$

for all cylinders $C(A,B)$, and also asymptotically additive on the average, in the sense that

$$\lim_{N\to\infty} \mathbf{E}(|\mu_{N,\lambda}(C(A,B)) - \mu_{N,\lambda}(C(A\cup\{n\},B) - \mu_{N,\lambda}(C(A,B\cup\{n\}))|\big|\lambda \in \mathbf{Z}/N\mathbf{Z}) = 0$$

Suppose furthermore that there were a collection Ω of cylinders which were asymptotically measure zero on the average, in the sense that

$$\lim_{N\to\infty} \mathbf{E}(|\mu_{N,\lambda}(C(A,B))|\big|\lambda \in \mathbf{Z}/N\mathbf{Z}) = 0$$

for all $C(A,B) \in \Omega$. Then there is an element $\lambda_N \in \mathbf{Z}/N\mathbf{Z}$ for each N with the property that the sequence (μ_{N,λ_N}) is asymptotically shift-invariant and asymptotically additive. Furthermore, we have the property

$$\lim_{N\to\infty} \mu_{N,\lambda_N}(C(A,B)) = 0$$

for all $C(A,B) \in \Omega$.

Indeed, there are only countably many properties that the λ_N need to satisfy, and so Lemma 2.1 can indeed be invoked.

Let us now give a concrete application of this corollary (which was essentially Varnavides' original application).

Theorem 2.3 (Finitary Szemerédi theorem)**.** *For every $k \geq 1$ and $\delta > 0$ there exists $c(k,\delta) > 0$ such that*

$$\mathbf{E}(f(x)f(x+r)\dots f(x+(k-1)r)|x,r \in \mathbf{Z}/N\mathbf{Z}) \geq c(k,\delta) - o_{N\to\infty;k,\delta}(1)$$

for all prime numbers N and all functions $f : \mathbf{Z}/N\mathbf{Z} \to [0,1]$ such that $\mathbf{E}(f(x)|x \in \mathbf{Z}/N\mathbf{Z}) \geq \delta$.

Proof Fix $k \geq 1$ and $\delta > 0$, and suppose for contradiction that the theorem failed. Then we could find a sequence of prime numbers N tending to infinity, together with functions $f_N : \mathbf{Z}/N\mathbf{Z} \to [0,1]$ such that $\mathbf{E}(f_N(x)|x \in \mathbf{Z}/N\mathbf{Z}) \geq \delta$ but

$$\lim_{N\to\infty} \mathbf{E}(f_N(x)f_N(x+r)\dots f_N(x+(k-1)r)|x,r \in \mathbf{Z}/N\mathbf{Z}) = 0.$$

For any $h > 0$, we may then make the change of variables $r = \lambda h$ if N is larger than h, and conclude

$$\lim_{N\to\infty} \mathbf{E}(f_N(x)f_N(x+\lambda h)\dots f_N(x+(k-1)\lambda h)|x \in \mathbf{Z}/N\mathbf{Z}; \lambda \in \mathbf{Z}/N\mathbf{Z}) = 0.$$

If we now set $g_N := 1 - f_N$ and let $\mu_{N,\lambda}$ be the weight

$$\mu_{N,\lambda}(C(A,B)) := \mathbf{E}((\prod_{h\in A} f_N(x+h\lambda))(\prod_{l\in B} g_N(x+l\lambda))|x \in \mathbf{Z}_N)$$

then we see that

$$\lim_{N\to\infty} \mathbf{E}(\mu_{N,\lambda}(\{0,h,\dots,(k-1)h\},\emptyset)|\lambda \in \mathbf{Z}/N\mathbf{Z}) = 0.$$

Also, by construction each $\mu_{N,\lambda}$ is shift-invariant and additive. Applying Corollary 2.2 we can then find a sequence (μ_{N,λ_N}) which is asymptotically shift-invariant, asymptotically additive, and whose limit superior $\overline{\mu}$ is such that

$$\overline{\mu}(\{0, h, \dots, (k-1)h\}, \emptyset) = 0.$$

On the other hand, observe that

$$\overline{\mu}(C(\emptyset, \emptyset)) = \limsup_{N\to\infty} \mu_{N,\lambda_N}(C(\emptyset, \emptyset)) = 1$$

so $(\mu_{N,\lambda_N}$ is asymptotically finite by Lemma 1.8. Similarly

$$\begin{aligned}\overline{\mu}(C(\{0\}, \emptyset)) &= \limsup_{N\to\infty} \mu_{N,\lambda_N}(C(\{0\}, \emptyset)) \\ &= \mathbf{E}(f_N(x)|x \in \mathbf{Z}/N\mathbf{Z}) \\ &\geq \delta.\end{aligned}$$

Since $\delta > 0$, we contradict Szemerédi's theorem. The claim follows. ■

A simple extension of the above argument gives

Theorem 2.4 (Relative finitary Szemerédi theorem)**.** *For every $k \geq 1$ and $\delta > 0$ there exists $c(k, \delta) > 0$ such that*

$$\mathbf{E}(f(x)f(x+r)\dots f(x+(k-1)r)|x, r \in \mathbf{Z}/N\mathbf{Z}) \geq c(k, \delta) - o_{N\to\infty;k,\delta}(1)$$

for all prime numbers N and all functions $\mathbf{E}(f(x)|x \in \mathbf{Z}/N\mathbf{Z}) \geq \delta$ obeying the pointwise bound $0 \leq f \leq \nu$, whenever $\nu : \mathbf{Z}/N\mathbf{Z} \to \mathbf{R}^+$ obeys a generalized von Neumann theorem

$$\mathbf{E}((\nu(x)-1)(\nu(y)-1)\prod_{h\in A} F_h(x+h\lambda)F_h(y+h\lambda)\big|x, y, \lambda \in \mathbf{Z}/N\mathbf{Z}) = o_{N\to\infty;A}(1) \tag{12}$$

for all finite sets $A \in \mathbf{Z}\backslash\{0\}$, and all functions F_h bounded in magnitude by ν.

Proof Fix k, δ, and suppose for contradiction that the theorem failed. Then as before we can find a sequence of primes N and functions $f_N : \mathbf{Z}/N\mathbf{Z} \to \mathbf{R}^+$ bounded by functions ν_N obeying (12). This property can be rewritten as

$$\mathbf{E}(\left|\mathbf{E}((\nu_N(x)-1)\prod_{h\in A} F_h(x+h\lambda)|x \in \mathbf{Z}/N\mathbf{Z})\right|^2 |\lambda \in \mathbf{Z}/N\mathbf{Z}) = o_{N\to\infty;A}(1)$$

and hence by Cauchy-Schwarz

$$\mathbf{E}(\left|\mathbf{E}((\nu_N(x)-1)\prod_{h\in A} F_h(x+h\lambda)|x \in \mathbf{Z}/N\mathbf{Z})\right| |\lambda \in \mathbf{Z}/N\mathbf{Z}) = o_{N\to\infty;A}(1).$$

Translating x by $n\lambda$, we thus have

$$\mathbf{E}(\left|\mathbf{E}((\nu_N(x+n\lambda)-1)\prod_{h\in A} F_h(x+h\lambda)|x \in \mathbf{Z}/N\mathbf{Z})\right| |\lambda \in \mathbf{Z}/N\mathbf{Z}) = o_{N\to\infty;A}(1)$$

whenever A is a finite subset of $\mathbf{Z}$ and $n \notin A$.

If we write $g_N := \nu_N - f_N$, and let $\mu_{N,\lambda}$ be the weight

$$\mu_{N,\lambda}(C(A,B)) := \mathbf{E}((\prod_{h\in A} f_N(x+h\lambda))(\prod_{l\in B} g_N(x+l\lambda))|x \in \mathbf{Z}_N)$$

as before, then we have

$$\begin{aligned}&\mu_{N,\lambda}(C(A,B)) - \mu_{N,\lambda}(C(A\cup\{n\},B)) - \mu_{N,\lambda}(C(A,B\cup\{n\}))\\&\quad = \mathbf{E}((1-\nu_N(x+n\lambda))\prod_{h\in A} f_N(x+h\lambda)\prod_{h\in A} g_N(x_l\lambda)|x \in \mathbf{Z}/N\mathbf{Z}).\end{aligned}$$

By the preceding discussion, we thus see that the $\mu_{N,\lambda}$ are asymptotically additive on the average:

$$\lim_{N\to\infty} |\mu_{N,\lambda}(C(A,B)) - \mu_{N,\lambda}(C(A\cup\{n\},B)) - \mu_{N,\lambda}(C(A,B\cup\{n\}))| = 0.$$

The $\mu_{N,\lambda}$ are also shift-invariant. We can now argue exactly as in Theorem 2.3 to obtain the desired contradiction. ■

Of course, to apply this theorem one would need to verify (12). In the language of [3], this condition can be verified assuming that ν obeys the linear forms condition to arbitrary order; this can be achieved by a repetition of the proof of the generalized von Neumann theorem in [3, Proposition 5.3], which is a tediously large number of applications of the Cauchy-Schwarz inequality.

This can be used for instance to deduce that if one chooses a random model E for the primes by selecting any integer N to lie in E with independent probability $1/\log N$, then one almost surely has a relative Szemerédi theorem for this model, in that every subset of E of positive relative density will contain arbitrarily long arithmetic progressions. However, it is unfortunately not quite strong enough to obtain arbitrarily long progressions in the primes. The reason is that we only know how to contain the primes inside a weight ν (concentrated on almost primes) which obeys the linear forms condition to any specified *finite* order, but if one increases the order of correlations that one desires, the density of the primes inside the almost primes will decay to zero. This setting was enough to make the more complicated transference arguments in [3] go through, but if one uses that hypothesis here, one only ends up (after using Arzela-Ascoli to take limits) with a weight which is shift-invariant, but is only additive up to a certain point, or more precisely it is only additive as long as the complexity $|A| + |B|$ of the cylinders $C(A,B)$ are all bounded by some universal bound $K < \infty$, which has to be selected in advance. It is not clear to what extent the Furstenberg recurrence theorem can be salvaged in such a restrictive setting; a major difficulty is that K is not permitted to depend on the density $\delta = \mu(C(\{0\},\emptyset))$. It may be that one may also be able to obtain some other cases of (9), or perhaps some monotonicity properties (e.g. analogues of the Bonferroni inequalities) which may serve as some sort of substitute. We do not know how to proceed beyond this point, but would be very interested to hear any ideas on this topic.

We thank Bryna Kra for useful discussions regarding this note and for some corrections, and Nikos Frantzinakis for pointing out that the Kolmogorov extension theorem does not require the axiom of choice.

References

[1] H. Furstenberg, *Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions*, J. Analyse Math. **31** (1977), 204–256.
[2] H. Furstenberg, Y. Katznelson and D. Ornstein, *The ergodic-theoretical proof of Szemerédi's theorem,* Bull. Amer. Math. Soc. **7** (1982), 527–552.
[3] B. Green, T. Tao, *The primes contain arbitrarily long arithmetic progressions*, preprint.
[4] P. Varnavides, *On certain sets of positive density,* J. London Math. Soc. **34** (1959) 358–360.

Department of Mathematics, UCLA, Los Angeles CA 90095-1555

E-mail address: tao@math.ucla.edu

URL: http://www.math.ucla.edu/∼tao