Mixing in Product Spaces

Elchanan Mossel

Poincaré Recurrence Theorem

Theorem (Poincaré, 1890)

Let $f : X \to X$ be a measure preserving transformation. Let $E \subset X$ measurable. Then

$$P[x \in E : f^n(x) \notin E, n > N(x)] = 0$$

Poincaré Recurrence Theorem

Theorem (Poincaré, 1890)

Let $f : X \to X$ be a measure preserving transformation. Let $E \subset X$ measurable. Then

$$P[x \in E : f^n(x) \notin E, n > N(x)] = 0$$

- One of the first results in Ergodic Theory.
- Long term mixing.
- This talk is about short term mixing.

Finite Markov Chains

- As a first example consider a *Finite Markov chain*.
- Let M be a $k \times k$ *doubly stochastic symmetric matrix*.
- Pick X^0 uniformly at random from $1, \ldots, k$.
- Given $X^i = a$, let $X^{i+1} = b$ with probability $M_{a,b}$.

Finite Markov Chains

- As a first example consider a *Finite Markov chain*.
- Let M be a $k \times k$ *doubly stochastic symmetric matrix*.
- Pick X^0 uniformly at random from $1, \ldots, k$.
- Given $X^i = a$, let $X^{i+1} = b$ with probability $M_{a,b}$.

Theorem (Long Term Mixing for Markov Chains)

Suppose that other than 1, all eigenvalues λ_i of M satisfy $|\lambda_i| \leq \lambda < 1$. Then for any two sets $A, B \subset [k]$, it holds that

$$\Big| P[X^0 \in A, X^t \in B] - P[A]P[B] \Big| \leq \lambda^t$$

Short Term Mixing for Markov Chains

Theorem

$\left| P[X^0 \in A, X^1 \in B] - P[A]P[B] \right|$ *is upper bounded by*

$$\lambda\sqrt{P[A](1-P[A])P[B](1-P[B])}$$

- Shows: mixing in one step for *large* sets.

Short Term Mixing for Markov Chains

Theorem

$\left|P[X^0 \in A, X^1 \in B] - P[A]P[B]\right|$ *is upper bounded by*

$$\lambda\sqrt{P[A](1-P[A])P[B](1-P[B])}$$

- Shows: mixing in one step for *large* sets.
- Proof: $1_A = P[A]\mathbf{1} + f, 1_B = P[B]\mathbf{1} + g$, where $f, g \perp 1$

Short Term Mixing for Markov Chains

Theorem

$\left|P[X^0 \in A, X^1 \in B] - P[A]P[B]\right|$ *is upper bounded by*

$$\lambda\sqrt{P[A](1-P[A])P[B](1-P[B])}$$

- Shows: mixing in one step for *large* sets.
- Proof: $1_A = P[A]\mathbf{1} + f, 1_B = P[B]\mathbf{1} + g$, where $f, g \perp 1$
-

$$P[X^0 \in A, X^1 \in B] = \frac{1}{k}(P[A]\mathbf{1} + f)^t M(P[B]\mathbf{1} + g)$$
$$= P[A]P[B] + \frac{1}{k}f^t Mg,$$

$$\frac{1}{k}|f^t Mg| \leq \lambda\|f\|_2\|g\|_2 = \lambda\sqrt{P[A](1-P[A])P[B](1-P[B])}$$

Short Term Mixing for Markov Chains

Theorem

$\left| P[X^0 \in A, X^1 \in B] - P[A]P[B] \right|$ *is upper bounded by*

$$\lambda \sqrt{P[A](1-P[A])P[B](1-P[B])}$$

- Shows: mixing in one step for *large* sets.
- Proof: $1_A = P[A]\mathbf{1} + f, 1_B = P[B]\mathbf{1} + g$, where $f, g \perp 1$
- $$P[X^0 \in A, X^1 \in B] = \frac{1}{k}(P[A]\mathbf{1} + f)^t M (P[B]\mathbf{1} + g) = P[A]P[B] + \frac{1}{k} f^t M g,$$

$$\frac{1}{k}|f^t M g| \leq \lambda \|f\|_2 \|g\|_2 = \lambda \sqrt{P[A](1-P[A])P[B](1-P[B])}$$

- Also called *Expander Mixing Lemma*.
- Used a lot in computer science, e.g. in (de)randomization.

The tensor property

- Consider $(Y_1, Z_1), \ldots, (Y_n, Z_n)$ which are drawn independently from the distribution of (X^0, X^1).
- Equivalently, the transition matrix from $Y = (Y_1, \ldots, Y_n)$ to $Z = (Z_1, \ldots, Z_n)$ is $M^{\otimes n}$.

The tensor property

- Consider $(Y_1, Z_1), \ldots, (Y_n, Z_n)$ which are drawn independently from the distribution of (X^0, X^1).
- Equivalently, the transition matrix from $Y = (Y_1, \ldots, Y_n)$ to $Z = (Z_1, \ldots, Z_n)$ is $M^{\otimes n}$.
- Thm $\implies$ that for any sets $A, B \subset [k]^n$:

$$\Big| P[Y \in A, Z \in B] - P[A]P[B] \Big| \leq \lambda \sqrt{P[A](1 - P[A])P[B](1 - P[B])}$$

The tensor property

- Consider $(Y_1, Z_1), \ldots, (Y_n, Z_n)$ which are drawn independently from the distribution of (X^0, X^1).
- Equivalently, the transition matrix from $Y = (Y_1, \ldots, Y_n)$ to $Z = (Z_1, \ldots, Z_n)$ is $M^{\otimes n}$.
- Thm $\implies$ that for any sets $A, B \subset [k]^n$:

$$\Big| P[Y \in A, Z \in B] - P[A]P[B] \Big| \leq \lambda \sqrt{P[A](1 - P[A])P[B](1 - P[B])}$$

- Follows immediately from tensorization of the spectrum.

Log Sobolev inequalities

Entropy, Log Sobolev and hyper-contraction

A similar story could be told using more sophisticated analytic tools. Easier to work with Markov semi-groups $T_t = e^{-tL}$.

Log Sobolev inequalities

Entropy, Log Sobolev and hyper-contraction

A similar story could be told using more sophisticated analytic tools. Easier to work with Markov semi-groups $T_t = e^{-tL}$.

Entropy, Dirchelet Form

$$Ent(f) = \mathbb{E}(f \log f) - \mathbb{E}f \cdot \log \mathbb{E}f$$

$$\mathcal{E}(f, g) = \mathbb{E}(fLg) = \mathbb{E}(gLf) = \mathcal{E}(g, f) = -\frac{d}{dt}\mathbb{E}fT_t g\Big|_{t=0}.$$

Log Sobolev inequalities

Entropy, Log Sobolev and hyper-contraction

A similar story could be told using more sophisticated analytic tools. Easier to work with Markov semi-groups $T_t = e^{-tL}$.

Entropy, Dirchelet Form

$$Ent(f) = \mathbb{E}(f \log f) - \mathbb{E}f \cdot \log \mathbb{E}f$$

$$\mathcal{E}(f,g) = \mathbb{E}(fLg) = \mathbb{E}(gLf) = \mathcal{E}(g,f) = -\frac{d}{dt}\mathbb{E}fT_t g\Big|_{t=0}.$$

Definition of Log-Sob

$$p\text{-logSob(C)} \iff \forall f, Ent(f^p) \le \frac{Cp^2}{4(p-1)}\mathcal{E}(f^{p-1}, f)\ (p \neq 0, 1)$$

$$1\text{-logSob(C)} \iff \forall f, Ent(f) \le \frac{C}{4}\mathcal{E}(f, \log f)$$

$$0\text{-logSob(C)} \iff \forall f, Var(\log f) \le -\frac{C}{2}\mathcal{E}(f, 1/f)$$

Log Sob. Inequalities and Hyper-Contraction

Hyper-Contraction (Gross, Nelson 1960 …)

r-logSob with constant C *implies*

$$\|T_t f\|_p \leq \|f\|_q, \quad t \geq \frac{C}{4} \log \frac{p-1}{q-1}, \quad 1 < p < q < r \text{ or } r' < q < p$$

$$\implies |\mathbb{E}[g(X_0)f(X_t)]| = |E[gT_t f| \leq \|g\|_{p'} \|Tf\|_p \leq \|g\|_{p'} \|f\|_q$$

If $f = 1_A$ and $g = 1_B$, get:

$$P[X_0 \in A, X_t \in B] \leq \|1_A\|_q \|1_B\|_{p'} = P[A]^{1/q} P[B]^{1/p'},$$

Now optimize over norms to get a better bound than CS.

Reverse-Hyper-Contraction

Log-Sobolev and Rev. Hyper-Contraction(M-Oleszkiewicz-Sen-13)

Let $T_t = e^{-tL}$ be a general Markov semi-group satisfying

- 2-Logsob with constant C or
- 1-Logsob inequality with constant C.

Then for all $q < p < 1$, all positive f, g and all $t \geq \frac{C}{4} \log \frac{1-q}{1-p}$ it holds that

$$\|T_t f\|_q \geq \|f\|_p \implies$$

$$\mathbb{E}[g(X_0)f(X_t)] = E[g T_t f] \geq \|g\|_{q'} \|f\|_p$$

Short-Time Implications

Theorem (M-Oleszkiewicz-Sen-13 ; Short-Time Implications)

Let $T_t = e^{-tL}$, *where* L *satisfy* 1 *or* 2*-LogSob inequality with constant* C. *Let* $A, B \subset \Omega^n$ *with* $\mathbb{P}[A] \geq \epsilon$ *and* $\mathbb{P}[B] \geq \epsilon$. *Then:*

$$\mathbb{P}[X(0) \in A, X(t) \in B] \geq \epsilon^{\frac{2}{1-e^{-2t/C}}}$$

Short-Time Implications

Theorem (M-Oleszkiewicz-Sen-13 ; Short-Time Implications)

Let $T_t = e^{-tL}$*, where* L *satisfy* 1 *or* 2*-LogSob inequality with constant* C*. Let* $A, B \subset \Omega^n$ *with* $\mathbb{P}[A] \geq \epsilon$ *and* $\mathbb{P}[B] \geq \epsilon$*. Then:*

$$\mathbb{P}[X(0) \in A, X(t) \in B] \geq \epsilon^{\frac{2}{1-e^{-2t/C}}}$$

Comments

1. Works for small sets too.
2. Tensorizes.
3. Some examples where it is (almost) tight.
4. Uses in social choice analysis, queuing theory.

Comment: typical application MCMC

Long Time Behavior

Log Sobolev inequalities play a major role in analyzing *long term* mixing of Markov chains, in particular in analysis of mixing times (Diaconis, Saloff-Coste etc.)

Long Time Behavior

The ε-total variation mixing time of a finite Markov chain is bounded by:

$$\frac{1}{\lambda}\left(\log(1/\pi^*) + \log(1/\epsilon)\right)$$

$$\frac{1}{C}\left(\log\log(1/\pi^*) + \log(1/\epsilon)\right)$$

for a continuous time Markov chain with spectral gap λ and 2-LogSob C.

Comment: typical application MCMC

Long Time Behavior

Log Sobolev inequalities play a major role in analyzing *long term* mixing of Markov chains, in particular in analysis of mixing times (Diaconis, Saloff-Coste etc.)

Long Time Behavior

The ε-total variation mixing time of a finite Markov chain is bounded by:

$$\frac{1}{\lambda}\left(\log(1/\pi^*) + \log(1/\epsilon)\right)$$

$$\frac{1}{C}\left(\log\log(1/\pi^*) + \log(1/\epsilon)\right)$$

for a continuous time Markov chain with spectral gap λ and 2-LogSob C.

What are these lectures about?

High Dimensional Phenomena

High dimensional mixing: mixing of product processes on product spaces Ω^n with n large.

Tight bounds

- For which processes, given measures a and b can we find *precise* upper/lower bounds for

$$\sup\left(P[X_0 \in A, X_t \in B] : P[A] = a, P[B] = b\right)$$

- Interested in product space/processes of dimension n and answers as $n \to \infty$.
- Most important examples / techniques from probability / analysis.

What are these lectures about?

Mulit-step prcoesses

- How to bound $P[X_0 \in A_0, X_1 \in A_1, \ldots, X_k \in A_k]$ for processes $X_0, \ldots, X_k$?
- Interested in product space/processes of dimension n and answers as $n \to \infty$.
- Most important examples / techniques from additive combinatorics.

What are these lectures about?

And more

- Theory that does both?
- Applications?

Today: tight bounds

- Borell's result.
- Open Problem: The Boolean cube.
- The state of affairs - partition into 3 parts or more.

Two Examples: Gaussian, Boolean

Correlated pairs (M-O'Donnell-Regev-Steif-Sudakov-05):

- Let $x, y \in \{-1,1\}^n$ be e^{-t} *correlated*:
- x is chosen uniformly and y is T_t correlated version.
- i.e. $\mathbb{E}[x_i y_i] = e^{-t}$ for all i independently
- Let $A, B \subset \{-1,1\}^n_{1/2}$ with $\mathbb{P}[A] \geq \epsilon$ and $\mathbb{P}[B] \geq \epsilon$
- Then: $\mathbb{P}[x \in A, y \in B] \geq \epsilon^{\frac{2}{1-e^{-t}}}$
- Easy to prove when $A = B$ ...

Two Examples: Gaussian, Boolean

Correlated pairs (M-O'Donnell-Regev-Steif-Sudakov-05):

- Let $x, y \in \{-1,1\}^n$ be e^{-t} *correlated*:
- x is chosen uniformly and y is T_t correlated version.
- i.e. $\mathbb{E}[x_i y_i] = e^{-t}$ for all i independently
- Let $A, B \subset \{-1,1\}^n_{1/2}$ with $\mathbb{P}[A] \geq \epsilon$ and $\mathbb{P}[B] \geq \epsilon$
- Then: $\mathbb{P}[x \in A, y \in B] \geq \epsilon^{\frac{2}{1-e^{-t}}}$
- Easy to prove when $A = B$ ...

Gaussian Version

- Let $x, y \in \mathbb{R}^n$ two Gaussian vectors:
- $x \sim N(0,1), y \sim N(0,1), E[x_i y_j] = e^{-t}\delta_{i,j}$
- Let $A, B \subset \mathbb{R}^n$ with $\mathbb{P}[A] \geq \epsilon$ and $\mathbb{P}[B] \geq \epsilon$
- Then: $\mathbb{P}[x \in A, y \in B] \geq \epsilon^{\frac{2}{1-e^{-t}}}$

Borell's Result and Open Problems

- Borell (85): In Gaussian case the maximum and minimum of $\mathbb{P}[x \in A, y \in B]$ as a function of $P[A]$ and $P[B]$ is obtained for parallel half-spaces.

Borell's Result and Open Problems

- Borell (85): In Gaussian case the maximum and minimum of $\mathbb{P}[x \in A, y \in B]$ as a function of $P[A]$ and $P[B]$ is obtained for parallel half-spaces.
- Do not know what is the optimum in $\{-1, 1\}^n$. In particular:

Borell's Result and Open Problems

- Borell (85): In Gaussian case the maximum and minimum of $\mathbb{P}[x \in A, y \in B]$ as a function of $P[A]$ and $P[B]$ is obtained for parallel half-spaces.
- Do not know what is the optimum in $\{-1,1\}^n$. In particular:
- Open Problem:

$$\lim_{n \to \infty} \min(P[X \in A, Y \in B] : A, B \subset \{-1,1\}^n, P[A] = P[B] = 1/4)$$

and similarly for max.

Borell's Result and Open Problems

- Borell (85): In Gaussian case the maximum and minimum of $\mathbb{P}[x \in A, y \in B]$ as a function of $P[A]$ and $P[B]$ is obtained for parallel half-spaces.
- Do not know what is the optimum in $\{-1,1\}^n$. In particular:
- Open Problem:

  $$\lim_{n\to\infty} \min(P[X \in A, Y \in B] : A, B \subset \{-1,1\}^n, P[A] = P[B] = 1/4)$$

  and similarly for max.
- Partition to 3 or more parts even in Gaussian space.

If there is time before the break ...

- A cute proof of a special case of Borell's result.
- Connections to social choice Theory.

Simple Example 1

- Cosmic coin problem(M-O'Donnell-05):
- $x \in \{-1, 1\}^n$ uniform.
- $(y^i)_1^m$ conditionally independent given x.
- Each pair (x, y^i) is ρ-correlated.
- Problem: What is the largest $P[y^1 \in A, \ldots y^m \in A]$ can be?

Simple Example 2

- $(y^{i,j})_{1 \leq i < j \leq m}$ is an exchangeable collection of vectors in $\{-1,1\}^n$.
- If $|I \cap J| = 1$ then y_I, y_J are $-1/3$ correlated.
- Otherwise independent.
- Why?
- If n voters rank alternatives uniformly at random, the pairwise preferences between alternatives will be given by the collection y.

Full support finite Ω using hyper-contraction

Thm: More General Reverse Hypercontractivity Theorem (M-Oleszkiewicz-Sen-13)

Let a the measure Ψ over a finite Ω^k satisfy $\min_{x_1,\ldots,x_k \in \Omega} \Pr[X_1 = x, \ldots, X_k = x_k] = \alpha > 0$ and have equal marginals.

Full support finite Ω using hyper-contraction

Thm: More General Reverse Hypercontractivity Theorem (M-Oleszkiewicz-Sen-13)

Let a the measure Ψ over a finite Ω^k satisfy $\min_{x_1,\ldots,x_k \in \Omega} \Pr[X_1 = x, \ldots, X_k = x_k] = \alpha > 0$ and have equal marginals.
Consider the distribution Ψ^n and let $A_1, \ldots, A_k \subseteq \Omega^n$, $\mu(A_i) \geq \mu$.
Then:

$$\Pr[X_1 \in A_1, \ldots X_k \in A_k] \geq \mu^{O\left(\frac{1}{\alpha}\right)},$$

where $(X_1(i), \ldots, X_k(i))$ are i.i.d. according to Ψ.

Note

This is a key tool of analyzing the examples above as well as many others.

Notation

Distributed according to $\underline{\mathcal{P}} := \mathcal{P}^n$.

Tuples $\overline{X}_i$ are i.i.d. according to $\mathcal{P}$. The marginals of $\mathcal{P}$ are π_j.

Vectors $\underline{X}^{(j)}$ are distributed according to $\underline{\pi_j} := \pi_j^n$.

$\overline{X}$	$\overline{X}_1$	$\overline{X}_2$	$\cdots$	$\overline{X}_i$	$\cdots$	$\overline{X}_n$
$\underline{X}^{(1)}$	$X_1^{(1)}$	$X_2^{(1)}$	$\cdots$	$X_i^{(1)}$	$\cdots$	$X_n^{(1)}$
$\underline{X}^{(2)}$	$X_1^{(2)}$	$X_2^{(2)}$	$\cdots$	$X_i^{(2)}$	$\cdots$	$X_n^{(2)}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$
$\underline{X}^{(j)}$	$X_1^{(j)}$	$X_2^{(j)}$	$\cdots$	$X_i^{(j)}$	$\cdots$	$X_n^{(j)}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$
$\underline{X}^{(\ell)}$	$X_1^{(\ell)}$	$X_2^{(\ell)}$	$\cdots$	$X_i^{(\ell)}$	$\cdots$	$X_n^{(\ell)}$

Lower Bounds

We are mostly interested in two types of lower bounds:

- *Set hitting:* Lower bounds on

$$P[X^1 \in A_1, \ldots, X^k \in A_k]$$

  in terms of $P[A_1], \ldots, P[A_k]$
- *Same set hitting:* Lower bounds on

$$P[X^1 \in A, \ldots, X^k \in A]$$

  in terms of $P[A]$.
- Set hitting will require something ... - e.g. $X^1 = X^2 = \ldots = X^k$.

Gaussian Bounds

- Borell (85) $k = 2$ - parallel half-spaces are optimal (also Isaksson-Mossel, Neeman)
- By a *Reverse Brascamp-Lieb inq.* (Ledoux, Chen-Dafnis-Paouris 14-15) for $A, \dots C \subset R^n$:

$$P[U \in A, \dots, Z \in C] \geq (P[A] \cdots P[C])^{1/(1-\rho^2)},$$

  where ρ is the second eigenvalue of Σ.
- Doesn't require independence of coordinates

Full Support Case

Thm: More General Reverse Hypercontractivity Theorem (M-Oleszkiewicz-Sen-13)

Let a the measure Ψ over a finite Ω^k satisfy $\min_{x_1,\ldots,x_k \in \Omega} \Pr[X_1 = x, \ldots, X_k = x_k] = \alpha > 0$ and have equal marginals.

Full Support Case

Thm: More General Reverse Hypercontractivity Theorem (M-Oleszkiewicz-Sen-13)

Let a the measure Ψ over a finite Ω^k satisfy $\min_{x_1,\ldots,x_k \in \Omega} \Pr[X_1 = x, \ldots, X_k = x_k] = \alpha > 0$ and have equal marginals.
Then:

$$\Pr[X_1 \in A_1, \ldots X_k \in A_k] \geq \mu^{O(\frac{1}{\alpha})},$$

where $(X_1(i), \ldots, X_k(i))$ are i.i.d. according to Ψ.

Non full support?

- What if the support of Ω is not full?
- Do we care?

Non full support?

- What if the support of Ω is not full?
- Do we care?
- Maybe: This is what *additive combinatorics* is all about.
- In particular: finite cominatorics in finite field models (Green-04 ...).
- Many other applications in combinatorics and computer science.

Additive combinatorics perspective

Example:

Theorem (Finite Field Roth Theorem)

- *Y, R be chosen uniformly at random at F_3^n.*
- *Then for every $\mu > 0$ there exists $c(\mu) > 0, N(\mu)$ such that if $n \geq N(\mu)$ and*
- *$A \subset F_3^n$ satisfies $P[A] \geq \mu$, then:*

$$P[Y \in A, Y + R \in A, Y + 2R \in A] \geq c(\mu).$$

Additive combinatorics perspective

Example:

Theorem (Finite Field Roth Theorem)

- *Y, R be chosen uniformly at random at F_3^n.*
- *Then for every $\mu > 0$ there exists $c(\mu) > 0, N(\mu)$ such that if $n \geq N(\mu)$ and*
- *$A \subset F_3^n$ satisfies $P[A] \geq \mu$, then:*

$$P[Y \in A, Y + R \in A, Y + 2R \in A] \geq c(\mu).$$

Why is this true?

Fourier Obstructions

Theorem (Finite Field Roth Theorem - Analysis)

Let Y, R be chosen uniformly at random at F_3^n. Let $A, B, C \subset F_3^n$ then

$$|P[Y \in A, Y + R \in B, Y + 2R \in C] - P[A]P[B]P[C]| \leq \|\hat{A}\|_\infty$$

Fourier Obstructions

Theorem (Finite Field Roth Theorem - Analysis)

Let Y, R be chosen uniformly at random at F_3^n. Let $A, B, C \subset F_3^n$ then

$$|P[Y \in A, Y + R \in B, Y + 2R \in C] - P[A]P[B]P[C]| \leq \|\hat{A}\|_\infty$$

Only obstruction to uniformity is linear structure

Fourier Obstructions

Theorem (Finite Field Roth Theorem - Analysis)

Let Y, R be chosen uniformly at random at F_3^n. Let $A, B, C \subset F_3^n$ then

$$|P[Y \in A, Y + R \in B, Y + 2R \in C] - P[A]P[B]P[C]| \leq \|\hat{A}\|_\infty$$

Only obstruction to uniformity is linear structure If $A = B = C$, high Fourier coefficient $\implies$ can restrict to linear subspace with higher denisty

Fourier Obstructions

Theorem (Finite Field Roth Theorem - Analysis)

Let Y, R be chosen uniformly at random at F_3^n. Let $A, B, C \subset F_3^n$ then

$$|P[Y \in A, Y + R \in B, Y + 2R \in C] - P[A]P[B]P[C]| \leq \|\hat{A}\|_\infty$$

Only obstruction to uniformity is linear structure If $A = B = C$, high Fourier coefficient $\implies$ can restrict to linear subspace with higher denisty Density increase arguments ...

Higher Order Arithmetic Obstructions

- Furstenberg-Weiss (80s): For longer arithmetic progressions, obstructions other than Fourier.
- Gowers: Obstructions can be identified using the Gowers norms.
- Again - use obstruction to your benefit.
- Thm: (Gowers 08; Rodel and Skokan 04,06):
  - If q is prime and $\ell \leq q$ then
  - for every $\mu > 0$ there exists $c(\mu) > 0, N(\mu)$ such that if $n \geq N(\mu)$ and
  - $A \subset F_q^n$ satisfies $P[A] \geq \mu$, then:

$$P[Y \in A, Y + R \in A, \ldots, Y + (\ell - 1)R \in A] \geq c(\mu),$$

  - where $A, R \in F_q^n$ are chosen uniformly at random.

Higher Order Arithmetic Obstructions

- Furstenberg-Weiss (80s): For longer arithmetic progressions, obstructions other than Fourier.
- Gowers: Obstructions can be identified using the Gowers norms.
- Again - use obstruction to your benefit.
- Thm: (Gowers 08; Rodel and Skokan 04,06):
  - If q is prime and $\ell \leq q$ then
  - for every $\mu > 0$ there exists $c(\mu) > 0, N(\mu)$ such that if $n \geq N(\mu)$ and
  - $A \subset F_q^n$ satisfies $P[A] \geq \mu$, then:

$$P[Y \in A, Y + R \in A, \dots, Y + (\ell - 1)R \in A] \geq c(\mu),$$

  - where $A, R \in F_q^n$ are chosen uniformly at random.
- Question: Is the additive structure necessary?

Obstruction to Chaos

- Consider the support of Ω as a graph G with vertex $V =$ all atoms with non-zero weight and edges between any two atoms that differ in one coordinate.
- We say that $\rho < 1$ if the graph G is connected.
- More formally:

Definition

$$\rho(\mathcal{P}, S, T) := \sup\Big\{ Cov[f(X^{(S)}), g(X^{(T)})] \,\Big|\, f : \Omega^{(S)} \to \mathbb{R}, g : \Omega^{(T)} \to \mathbb{R}, \\ Var[f(X^{(S)})] = Var[g(X^{(T)})] = 1 \Big\} .$$

The correlation of $\mathcal{P}$ is $\rho(\mathcal{P}) := \max_{j \in [\ell]} \rho\left(\mathcal{P}, \{j\}, [\ell] \setminus \{j\}\right)$.

The quest for a unifying theory

Is there one theory that explains both the noisy examples and the additive theory?

Example

- Let X be uniform in F_3^n.
- Let $Y_i = X_i$ or $X_i + 1$ with probability $1/2$ independently for each coordinate.
- Theorem $\implies P[X \in A, Y \in a] \geq c(P[A])$.
- Motivation from understanding "parallel repetition".
- Does not follow from hyper-contraction nor does it follow from additive techniques ...

A General Result

Theorem (+ Hazla, Holenstein)

Suppose (X, Y) is distributed in a finite Ω^2 such that:

- $\alpha = min_a P[X = Y = a] > 0$.
- $P[X = a] = P[Y = a]$ *for all a.*

Then for any set $A \subset \Omega^n$ with $P_{X^{\otimes n}}[A] = P_{Y^{\otimes n}}[A] \geq \mu$ it holds that

$$P[X \in A, Y \in A] \geq c(\alpha, \mu) > 0$$

- Our c is pretty bad:

$$c = 1/\exp(\exp(\exp(1/(\mu)^D))), \quad D = D(\alpha)$$

A General Result

Theorem (+ Hazla, Holenstein)

Suppose (X, Y) is distributed in a finite Ω^2 such that:

- $\alpha = min_a P[X = Y = a] > 0$.
- $P[X = a] = P[Y = a]$ *for all* a.

Then for any set $A \subset \Omega^n$ with $P_{X^{\otimes n}}[A] = P_{Y^{\otimes n}}[A] \geq \mu$ it holds that

$$P[X \in A, Y \in A] \geq c(\alpha, \mu) > 0$$

- Our c is pretty bad:

$$c = 1/\exp(\exp(\exp(1/(\mu)^D))), \quad D = D(\alpha)$$

- Related to the fact that the proof is interesting:
  1. Lose in "Regularity Lemma" type arguments.
  2. Lose in "Invariance" transforming the problem to a Gaussian problem.

A Markov Chain Theorem and a general process theorem

Theorem[+Hazla, Holenstein]

$X_i, Y_i, Z_i, \ldots, W_i$ be a Markov chain over Ω with $\min_{x \in \Omega} \Pr[X_i = Y_i = Z_i = \ldots W_i = x] = \beta > 0$ and uniform marginals.
Let $A \subseteq \Omega^n$, $\mu(A) = \mu > 0$.

$$\Pr[X \in A \wedge Y \in A \wedge Z \in A, \ldots, \wedge W \in A] \geq f(\mu, \beta) > 0 \, .$$

Theorem[+Hazla, Holenstein]

$X_i, Y_i, Z_i, \ldots, W_i$ be distributed over Ω^k with $\min_{x \in \Omega} \Pr[X_i = Y_i = Z_i = \ldots W_i = x] = \beta > 0$ and uniform marginals. *Suppose further that* $\rho(X_i, Y_i, \ldots, W_i) < 1$. Let $A \subseteq \Omega^n$, $\mu(A) = \mu > 0$.

$$\Pr[X \in A \wedge Y \in A \wedge Z \in A, \ldots, \wedge W \in A] \geq f(\mu, \beta) > 0 \, .$$

The condition $\rho < 1$

- Weaker than full support.

The condition $\rho < 1$

- Weaker than full support.
- Does not hold in arithmetic setups.

The condition $\rho < 1$

- Weaker than full support.
- Does not hold in arithmetic setups.
- $\rho = 1$ **iff** the support of Ψ is connected with respect to changing one coordinate at a time.

The condition $\rho < 1$

- Weaker than full support.
- Does not hold in arithmetic setups.
- $\rho = 1$ **iff** the support of Ψ is connected with respect to changing one coordinate at a time.
- Example: $(x, y) \in F_3^2$ where $y = x, x + 1$ has $\rho < 1$ but not full support.

Open Problems

- Still searching for unified theory.
- Concrete Example: Suppose Ψ is uniform over

$$\{(0,0,0),(1,1,1),(2,2,2),(0,1,2),(1,2,0),(2,0,1)\}$$

- $\rho = 1$ but not arithmetic.
- Do not understand.

Questions??

Questions??

Thank you!