European Journal of Combinatorics 54 (2016) 154–162

Contents lists available at ScienceDirect

European Journal of Combinatorics

journal homepage: www.elsevier.com/locate/ejc

Extractors in Paley graphs: A random model

Rudi Mrazović

Mathematical Institute, Andrew Wiles Building, Radcliffe Observatory Quarter, Woodstock Road, Oxford OX2 6GG, United Kingdom

ARTICLE INFO

Article history:
Received 2 September 2015
Accepted 16 December 2015
Available online 15 January 2016

ABSTRACT

A well-known conjecture in analytic number theory states that for every pair of sets $X, Y \subset \mathbb{Z}/p\mathbb{Z}$, each of size at least $\log^C p$ (for some constant C) we have that the number of pairs $(x, y) \in X \times Y$ such that $x + y$ is a quadratic residue modulo p differs from $\frac{1}{2}|X||Y|$ by $o(|X||Y|)$. We address the probabilistic analogue of this question, that is for every fixed $\delta > 0$, given a finite group G and $A \subset G$ a random subset of density $\frac{1}{2}$, we prove that with high probability for all subsets $|X|, |Y| \geqslant \log^{2+\delta} |G|$, the number of pairs $(x, y) \in X \times Y$ such that $xy \in A$ differs from $\frac{1}{2}|X||Y|$ by $o(|X||Y|)$.

1. Introduction

A folklore result in analytic number theory states that if we let $Q \subset \mathbb{Z}/p\mathbb{Z}$ be the set of quadratic residues modulo a prime p, then for any function $w : \mathbb{N} \to \mathbb{R}$ tending to infinity and any pair of sets $X, Y \subset \mathbb{Z}/p\mathbb{Z}$ of size at least $w(p)\sqrt{p}$, the number of pairs $(x, y) \in X \times Y$ such that $x+y \in Q$ differs from $\frac{1}{2}|X||Y|$ by $o(|X||Y|)$. Here, the rate of convergence implied by the o-notation depends only on w.

The proof of this is relatively simple. We refer the reader to Section 2 for the notation used below. First of all, after setting $1_X * 1_Y(x) = \mathbb{E}_{z\in\mathbb{Z}/p\mathbb{Z}} 1_X(z)1_Y(x - z)$, it is easy to see that the statement one wants to prove is equivalent to requiring that

$$\left| p \sum_{x\in\mathbb{Z}/p\mathbb{Z}} 1_X * 1_Y(x)\chi(x) \right| = o(|X||Y|),$$

where χ is the quadratic character (i.e. $\chi(a) = (\frac{a}{p})$). For $r \in \mathbb{Z}/p\mathbb{Z}$ and a function $f : \mathbb{Z}/p\mathbb{Z} \to \mathbb{R}$, we define the corresponding Fourier coefficient by $\widehat{f}(r) = \mathbb{E}_{x\in\mathbb{Z}/p\mathbb{Z}} f(x)e^{2\pi ixr/p}$. Using standard

E-mail address: Rudi.Mrazovic@maths.ox.ac.uk.

http://dx.doi.org/10.1016/j.ejc.2015.12.009
0195-6698/

formulas from Fourier analysis, the standard estimate for Gauss sums (see e.g. [9, Section 3.5]), and Cauchy–Schwarz inequality, we have

$$\left|p\sum_{x\in\mathbb{Z}/p\mathbb{Z}} 1_X * 1_Y(x)\chi(x)\right| = \left|p^2\sum_{r\in\mathbb{Z}/p\mathbb{Z}} \widehat{1_X}(r)\widehat{1_Y}(r)\bar{\widehat{\chi}}(r)\right| \leqslant p^{3/2}\sum_{r\in\mathbb{Z}/p\mathbb{Z}} |\widehat{1_X}(r)|\,|\widehat{1_Y}(r)|$$
$$\leqslant \sqrt{p}(|X|\,|Y|)^{1/2},$$

and this proves the claim. Although this argument was quite straightforward, no significant improvement (in terms of lower bounds for $|X|$ and $|Y|$) is known, although it is widely believed to be true even for sets X and Y of sizes at least $\log^C p$ for some constant C—a conjecture known in some literature [5] as the Paley graph conjecture.

Given that the set of quadratic residues is believed to have many properties in common with a random set of the same size, it is natural to ask whether the statement above is true if we replace Q with a genuinely random set. In this paper we give the positive answer to this question, that is we prove the following theorem.

Theorem 1. *Let G be a group of size N and $w\colon \mathbb{N}\to\mathbb{R}$ some function that tends to infinity. Let $A\subset G$ be a random subset obtained by putting every element of G into A independently with probability $\frac{1}{2}$. Then the following holds with probability $1-o(1)$: for all sets $X, Y\subset G$, $|X|, |Y|\geqslant w(N)\log^2 N$, the number of pairs $(x,y)\in X\times Y$ such that $xy\in A$ differs from $\frac{1}{2}|X|\,|Y|$ by $o(|X|\,|Y|)$. The rate of convergence implied by the o-notation depends only on w.*

If for a pair of subsets $X, Y\subset G$ we have that the number of pairs $(x,y)\in X\times Y$ such that $xy\in A$ differs from $\frac{1}{2}|X|\,|Y|$ by $\epsilon|X|\,|Y|$, we will say that it is *ϵ-extracted* by the set A. If all the pairs of subsets of size as in the previous theorem are ϵ-extracted by the set A, we will say that A is an *ϵ-extractor*. Theorem 1 shows, in this terminology, that a random subset of G is $o(1)$-extractor with high probability. The reason for this terminology will be explained in Section 6.

A Fourier approach, as above but using Chernoff-type estimates for the Fourier coefficients instead of Gauss sum estimates, suffices to prove Theorem 1 when, say, $|X|, |Y|\geqslant N^{0.51}$. Unfortunately, this argument does not work for sets of size smaller than $\sqrt{N}$.

In Section 3 we will present a different argument that will enable us to prove Theorem 1. A very important part of the argument is Proposition 4, which we prove in Section 4. Section 5 is devoted for proving a bound on the sizes of X and Y for which Theorem 1 does not hold. Namely, we prove that Theorem 1 does not hold if we consider sets X and Y of size at least $C\log N\log\log N$, for arbitrary large constant $C>0$. Finally, in Section 6 we give some open problems left after this paper and explain the connection with randomness extractors.

Before moving on to the proof of the main theorem, let us also mention that one could ask and answer the analogous question in the Erdős–Rényi setting, that is one can easily prove the following folklore theorem.

Theorem 2. *Let $G=(V,E)$ be a graph sampled as in the Erdős–Rényi model $G(N,\frac{1}{2})$, that is G has a vertex set V of size N and contains each possible edge with probability $\frac{1}{2}$, and these choices are all made independently. Let $w\colon\mathbb{N}\to\mathbb{R}$ be some function that tends to infinity. Then the following holds with probability $1-o(1)$: for all sets $X, Y\subset V$, $|X|, |Y|\geqslant w(N)\log N$, $\frac{1}{2}+o(1)$ of all the possible edges connecting an element of X and an element of Y are contained in E. The rate of convergence implied by the o-notation depends only on w.*

2. Notation

Although most of the notation and conventions were implicitly introduced in the previous section, we include them here for the reader's convenience. The logarithm with base 2 will be denoted by $\log_2$, and natural logarithm by log. We will use the standard O-notation. To be concrete, for functions $f, g\colon\mathbb{N}\to\mathbb{R}$ we will write $f(n)=O(g(n))$ and $|f(n)|\ll g(n)$ if $|f(n)|\leqslant Cg(n)$ for some constant