Thin groups and superstrong approximation
MSRI Publications
Volume **61**, 2013

How random are word maps?

MICHAEL LARSEN

Any word w in the free group on d generators determines a function $G^d \to G$ for every group G. If w is a fixed nontrivial word and G ranges over the finite simple groups, the resulting sequence of functions can be expected to enjoy some properties of a random sequence of functions. In this paper, we review the current state of the art, emphasizing open problems.

1. Introduction

Let F_d denote the free group on d generators $x_1, \ldots, x_d$ and $w \in F_d$ a nontrivial element. For every group G, w defines a *word map* $f_{w,G}\colon G^d \to G$ obtained by substituting for $x_1, \ldots, x_d$ respectively the coordinates $g_1, \ldots, g_d$ of a given element of G^d. A number of authors have examined the behavior of $f_{w,G}$ when w is fixed and G ranges over some set of groups, especially the set of all (nonabelian) finite simple groups. A unifying theme behind a good deal of recent work is this question: for a given word w, do the maps $f_{w,G}$ behave like random functions $G^d \to G$? The answer depends partly on the choice of w but also on what properties of random functions are desired. This paper examines some recent progress in understanding basic randomness properties of word maps, with an emphasis on open questions.

Given a positive integer d and an infinite sequence $X_1, X_2, X_3, \ldots$ of finite sets, we consider sequences $f_1, f_2, f_3, \ldots$ of functions $f_i\colon X_i^d \to X_i$ chosen uniformly and independently. We are interested in properties of such sequences of functions which hold with probability 1. We will limit ourselves to the case that

$$\sum_{n=1}^{\infty} \frac{1}{|X_n|} < \infty. \tag{1-1}$$

It is an easy consequence of classification that the finite simple groups satisfy this property. The following theorem gives some typical examples of properties which hold for almost all sequences of random functions for X_i satisfying (1-1).

Partially supported by the National Science Foundation.

Theorem 1.1. *Let d be a positive integer. Let $X_1, X_2, X_3, \ldots$ denote an infinite sequence of finite sets satisfying* (1-1)*. Then for almost every sequence of functions $f_i\colon X_i^d \to X_i$, the following conditions hold*:

(1) (Image size.)

(a) *If $d = 1$,*

$$\lim_{n\to\infty} \frac{|f_n(X_n)|}{|X_n|} = 1 - \frac{1}{e}.$$

(b) *If $d > 1$, then $f_n(X_n) = X_n$ for all but finitely many n.*

(2) (Measure preservation in the limit.) *If $d > 1$,*

$$\lim_{n\to\infty} \sup_{S\subset X_n} \left| \frac{|S|}{|X_n|} - \frac{|f_n^{-1}(S)|}{|X_n^d|} \right| = 0.$$

(3) (Fiber size.) *If $d > 1$ and $a > (d-1)/2$, then for all $n \gg 0$ and all $x \in X_n$,*

$$\left| |f_n^{-1}(x)| - |X_n|^{d-1} \right| < |X_n|^a.$$

Note that item (3) implies (1b) and (2). The proof of this theorem is given in the Appendix. In the body of the paper, we will examine whether these properties of random sequences of maps in fact hold for the $f_{w,G}$ as G ranges over the finite simple groups. Because they are all limiting properties, we can omit the sporadic groups, restricting our attention to alternating groups and groups of Lie type.

If $\underline{G}$ is an algebraic group, the word map $f_{w,\underline{G}}$ is a morphism of algebraic varieties. A theorem of Borel [1983] asserts that if $\underline{G}$ is a semisimple algebraic group and w is nontrivial, then $f_{w,\underline{G}}$ is a dominant morphism, so that its image contains a nonempty open subset of $\underline{G}$. (Note that in general $f_{w,\underline{G}}$ need not be surjective; for example, the matrix

$$\begin{pmatrix} -1 & 1 \\ 0 & -1 \end{pmatrix}$$

does not lie in the image of $f_{x_1^2,\mathrm{SL}_2}$.) Information about $f_{w,\underline{G}}$ can translate into information about $f_{w,G}$ when G is of Lie type, especially when G is a large group of small rank. This point of view is especially useful in dealing with groups of exceptional Lie type, such as $E_8(q)$ but of course breaks down completely in dealing with alternating groups.

2. Image size

In this section, we consider the images $w(G) := f_{w,G}(G^d)$ of word maps. We consider first the case $d = 1$. In this case $w = x_1^m$ for some nonzero integer m.

Of course the cases $m = \pm 1$ are special, with $f_{w,G}$ surjective for every G. For $|m| \geq 2$, we have [Larsen 1997]

$$|w(S_n)| \sim c_{|m|} n^{-\phi(m)/m} n!,$$

where ϕ is Euler's function, so

$$\lim_{n\to\infty} \frac{|w(A_n)|}{|A_n|} = 0.$$

The factor $n^{\phi(m)/m}$ grows very slowly compared to $n!$, so we might hope for some result that says that in general $|w(G)|$ is not much smaller than $|G|$. Borel's theorem implies that if G ranges over finite simple groups of bounded rank, $|w(G)|/|G|$ is bounded below by a positive constant. Starting from this point and systematically exploiting the fact that $w(G)$ contains $w(H)$ for every subgroup $H \subset G$, one can prove the following theorem:

Theorem 2.1 [Larsen 2004]. *For w fixed, we have*

$$\lim_{|G|\to\infty} \frac{\log|w(G)|}{\log|G|} = 1.$$

One might ask for something stronger:

Question 2.2. *For w fixed, do we have*

$$|w(G)| \geq \frac{|G|}{\log|G|},$$

for all G sufficiently large?

Here is a result in this direction, by Aner Shalev and the author:

Theorem 2.3 [Larsen and Shalev 2009]. *For w and $\epsilon > 0$ fixed, we have*

$$|w(G)| \geq \frac{|G|}{\log^{29/9+\epsilon}|G|},$$

for all sufficiently large finite simple groups, except possibly those of the form $\mathrm{PSL}_n(\mathbb{F}_q)$ *and* $\mathrm{PSU}_n(\mathbb{F}_q)$.

For $d \geq 2$, randomness would imply that $w(G) = G$ for all sufficiently large G. In special cases, this is known to be true. We mention two such results, the first due to Martin Liebeck, Eamonn O'Brien, Shalev and Pham Tiep, and the second by Shalev, Tiep, and the author:

Theorem 2.4 (Ore's conjecture [Liebeck et al. 2010]). *The commutator map is surjective for all finite simple groups.*

Theorem 2.5 [Larsen et al. 2011]. *If $w = w_1 w_2$, where w_1 and w_2 are nontrivial words such that no generator x_i of F_d appears in both w_1 and w_2, then $w(G) = G$ for all G sufficiently large.*

Words of the type described in Theorem 2.5 are particularly easy to handle because $w(G) = w_1(G)w_2(G)$. By analogy with the classical Waring problem in analytic number theory, we say that words of the form $w = x_1^n x_2^n$, and more generally, all words of the form $w = w_1 w_2$, where the generators appearing in w_1 and w_2 are disjoint, are of *Waring type*.

These special examples might raise the hope that if $d \geq 2$, (1b) holds for all G sufficiently large, but this cannot be true if w is of the form w_0^m for $m \geq 2$. Words of this type will be called *power words*; among all words, they seem in some sense to be the ones whose word maps are the farthest from random.

Recent examples due to Sebastian Jambor, Liebeck, and O'Brien [Jambor et al. 2012] show that even if w is not a power word, there may be arbitrarily large G for which $w(G) \neq G$. There seems to be a general expectation that if w is not a power word and G has sufficiently high rank, then $w(G) = G$. In a related direction, one might ask the following question:

Question 2.6. *If w is a nonpower word, is it true that*

$$\lim_{|G|\to\infty} \frac{|w(G)|}{|G|} = 1? \tag{2-1}$$

3. Measure preservation in the limit

For any finite set X, let μ_X denote the uniform probability measure on X. A function $f: X \to Y$ is *measure preserving* if $f_*\mu_X = \mu_Y$, or equivalently, if $|f^{-1}(y)|$ is constant on Y. The L^1 norm of $f_*\mu_X - \mu_Y$ therefore quantifies the failure of f to preserve measure. (To date, nobody seems to have considered the behavior of $\|f_*\mu_X - \mu_Y\|_p$ except in the cases $p = 1$ and $p = \infty$, the latter to be discussed in §4.) It is easy to see that

$$\|f_*\mu_X - \mu_Y\|_1 = 2 \sup_{S\subset Y} |\mu_X(f^{-1}(S)) - \mu_Y(S)|,$$

so that (2) can be reformulated as

$$\lim_{n\to\infty} \|f_{n*}\mu_{X_n^d} - \mu_{X_n}\|_1 = 0.$$

Question 3.1. *If w is a nonpower word, is it true that*

$$\lim_{|G|\to\infty} \|f_{w,G*}\mu_{G^d} - \mu_G\|_1 = 0? \tag{3-1}$$

Note that for any word w, (3-1) implies (2-1), so an affirmative answer to Question 3.1 implies an affirmative answer to Question 2.6.

Here is a geometric analogue of Question 3.1:

Question 3.2. *If w is a nonpower word, is it true that for all simply connected almost simple algebraic groups $\underline{G}$, the generic fiber of the morphism $f_{w,\underline{G}}$ is geometrically irreducible*?

One can ask equivalently whether $\underline{G}$ has a nonempty open set of points whose inverse images are geometrically irreducible of dimension $(d-1)\dim \underline{G}$. If the answer is affirmative, then by the Lang–Weil bound, almost all the fibers of $f_{w,\underline{G}(\mathbb{F}_q)}$ have approximate size $q^{(d-1)\dim \underline{G}}$ when q is large, and this implies that almost all fibers of $f_{w,G}$ have approximate size $|G|^{d-1}$ where G is the (simple) quotient of $\underline{G}(\mathbb{F}_q)$ by its center. It follows that an affirmative answer to Question 3.2 leads to an affirmative answer to Question 3.1 for bounded rank.

At present, little is known about these questions. We mention two results; the first is due to Shelly Garion and Shalev.

Theorem 3.3 [Garion and Shalev 2009]. *If $w = x_1x_2x_1^{-1}x_2^{-1}$ then* (3-1) *holds.*

Theorem 3.4 [Larsen and Shalev 2013]. *If $w = x_1^{n_1}x_2^{n_2}$ for integers n_1, n_2, then* (3-1) *holds.*

It seems likely that Theorem 3.4 should generalize at least to all words of Waring type. In this setting, the answer to Question 3.2 is known to be yes [Larsen and Shalev 2009].

4. Fiber size

Any nontrivial lower bound on fiber size immediately implies the surjectivity of $f_{w,G}$ for large G, so at present, we can only hope to prove such bounds for very special words. For example, by a theorem of Liebeck and Shalev [2005], for the "surface" words

$$w = \prod_{i=1}^{g} x_{2i-1}x_{2i}x_{2i-1}^{-1}x_{2i}^{-1}$$

of genus $g \geq 2$ we have

$$|f_{w,G}^{-1}(e)| = |G|^{2g-1} + o(|G|^{2g-1}),$$

and the proof gives the same estimate uniformly for all fibers of $f_{w,G}$. This, of course, is much weaker than (3), but it does give both lower and upper bounds for fiber size.

The best behaved words, in some sense, are the *primitive* words. Recall that an element $w_1 \in F_d$ is primitive if and only if there exist $w_2, \ldots, w_d$ such that $w_1, \ldots, w_d$ generate F_d, in which case they do so freely. There is a one-to-one correspondence between d-tuples $(g_1, \ldots, g_d) \in G^d$ and homomorphisms

$\phi\colon F_d \to G$ given by substituting $x_1 = g_1, \dots x_d = g_d$ and likewise a one-to-one correspondence between homomorphisms ϕ and d-tuples $(h_1, \dots, h_d)$ obtained by setting $h_1 = \phi(w_1), \dots, h_d = \phi(w_d)$. Thus $|f_{w,G}^{-1}(h_1)| = |G|^{d-1}$ for all $h_1 \in G$. Thus (3) holds for primitive words.

The situation for upper bounds is only a little better. Some words which behave fairly well with respect to other randomness properties can have quite large fibers. For instance, if $w = x_1 x_2 x_1^{-1} x_2^{-1}$ and $e \in G$ is the identity, then $|f_{w,G}^{-1}(e)| = |G| \cdot |G^\natural|$, where $G^\natural$ denotes the set of conjugacy classes of G. For G of Lie type of fixed rank r, $|G^\natural|$ grows like q^r. So, for instance, if $G = \mathrm{PSL}_2(\mathbb{F}_q)$, then $|f_{w,G}^{-1}(e)|$ grows like $|G|^{4/3}$.

For general words, we have:

Theorem 4.1 [Larsen and Shalev 2012]. *For every nontrivial word $w \in F_d$, there exists $\epsilon > 0$ such that for all sufficiently large finite simple groups G and all $g \in G$,*

$$|f_{w,G}^{-1}(g)| < |G|^{d-\epsilon}.$$

When $w = x_1^m$ for $m > 1$ a fixed odd integer, and $G = \mathrm{PSL}_m(\mathbb{F}_q)$ for $q \equiv 1 \pmod{m}$, the identity fiber $f_{w,G}^{-1}(e)$ contains a regular semisimple element, namely the image $\bar{\delta}$ in G of any diagonal element $\delta \in \mathrm{SL}_m(\mathbb{F}_q)$ whose diagonal entries are all the m-th roots of unity in $\mathbb{F}_q$. Thus, the identity fiber contains the image in G of the conjugacy class of δ in $\mathrm{SL}_m(\mathbb{F}_q)$, so

$$|F_{w,G}^{-1}(e)| \geq \frac{1}{m} \frac{|\mathrm{SL}_m(\mathbb{F}_q)|}{(q-1)^{m-1}} \geq \frac{q^{m^2-m}}{m}.$$

Taking the limit $q \to \infty$, we see that ϵ must be taken no greater than $1/(m+1)$ in Theorem 4.1. Taking the limit $m \to \infty$, we see that in some sense Theorem 4.1 is as strong as possible, though it remains an interesting question how ϵ depends on w, for example on the length of w as a word. One might also ask whether a stronger bound holds for nonpower words.

It turns out that, except when w is primitive, condition (3) *never* holds. Alexandru Nica [1994] proved that given a word w, there exists a rational function $R(x)$ such that for n sufficiently large, the average number of fixed points of $f_{w,S_n}(g_1, \dots, g_d)$ is $R(1/n)$. The argument holds for every sufficiently transitive permutation group and therefore for alternating groups A_n if n is sufficiently large. Doron Puder and Ori Parzanchevski [2012; 2013] proved that $R(x)$ is identically 1 if and only if w is primitive. Otherwise, it admits a power series expansion in x. On the other hand, if every fiber of f_{w,A_n} has order $|A_n|^{d-1} + |A_n|^a$ and $a < d-1$, then the expected number E_n of fixed points of $f_{w,A_n}(g_1, \dots, g_d)$ is $1 + O(n!^{a-d})$, so for every k, $|E_n - 1| < n^{-k}$ for large n. It follows that $R(x) = 1$, and w is primitive.

In conclusion, the only word maps which satisfy (1), (2), and (3) are those associated to primitive elements. However, even the primitive elements cannot boast true randomness. Indeed, if $f_n: X_n^d \to X_n$ is a random sequence, with probability 1, there exists an arbitrarily large n and a pair of elements $x_n, y_n \in X_n$ such that $|f_n^{-1}(x_n)| \neq |f_n^{-1}(y_n)|$.

Appendix: Proof of Theorem 1.1

Let X be a finite set and $n = |X|$. Let Y denote the set of functions $X \to X$. Define $f: Y \to \mathbb{Z}$ by $f(g) := |g(X)|$. Let $F: X \times Y \to \{0, 1\}$ be the function such that $F(x, g) = 1$ if and only if $g^{-1}(x) = \varnothing$. Thus

$$n - f(g) = \sum_{x \in X} F(x, g),$$

and

$$(n-f(g))^2 = \sum_{x_1, x_2 \in X} F(x_1, g)F(x_2, g) = n - f(g) + \sum_{x_1 \neq x_2} F(x_1, g)F(x_2, g).$$

Thus

$$\begin{aligned}\sum_{g \in Y}\left(\frac{1}{e} - \left(1 - \frac{f(g)}{n}\right)\right)^2 &= e^{-2}|Y| - 2e^{-1}n^{-1}\sum_{x \in X}\sum_{g \in Y} F(x, g) \\ &\quad + n^{-2}\sum_{x \in X}\sum_{g \in Y} F(x, g) \\ &\quad + n^{-2}\sum_{x_1 \neq x_2}\sum_{g \in Y} F(x_1, g)F(x_2, g) \\ &= e^{-2}n^n + \left(\frac{1}{n} - \frac{2}{e}\right)(n-1)^n + \frac{n-1}{n}(n-2)^n. \quad (4\text{-}1)\end{aligned}$$

As

$$\frac{(n-k)^n}{n^n} = e^{-k} + O(1/n),$$

we conclude that the average value of $(e^{-1} - (1 - f(g)/n))^2$ on Y is $O(1/n)$. Thus, with respect to the uniform probability distribution on Y, the measure of the set of functions $g: X \to X$ such that

$$|g(X)| \notin [(1 - e^{-1} - \epsilon)n, [(1 - e^{-1} + \epsilon)n]$$

is bounded above by a constant multiple of $1/\epsilon^2 n$. By (1-1), the Borel–Cantelli lemma implies (1a).

For (3) (and therefore also (1b) and (2)), we observe that the probability that

$$|f^{-1}(x) - n^{d-1}| \geq n^a$$

for any particular $x \in X$ is bounded above by

$$n^d \sup_{\{k:|k-n^{d-1}|\geq n^a\}} \binom{n^d}{k}\left(\frac{1}{n}\right)^k \left(\frac{n-1}{n}\right)^{n^d-k}.$$

Now,

$$\begin{aligned}\binom{n^d}{k}\left(\frac{1}{n}\right)^k \left(\frac{n-1}{n}\right)^{n^d-k} &\leq \frac{\binom{n^d}{k}\left(\frac{1}{n}\right)^k \left(\frac{n-1}{n}\right)^{n^d-k}}{\binom{n^d}{n^{d-1}}\left(\frac{1}{n}\right)^{n^{d-1}} \left(\frac{n-1}{n}\right)^{n^d-n^{d-1}}} \\ &= \frac{n^{d-1}!(n^d-n^{d-1})!}{(n-1)^{k-n^{d-1}}k!(n^d-k)!}. \end{aligned} \tag{4-2}$$

Writing $k = n^{d-1} + j$, if j is positive (and therefore $j \geq n^a$), the right hand side of (4-2) equals

$$\prod_{i=1}^{j} \frac{(n-1)n^{d-1}-i}{(n-1)(n^{d-1}+i)} \leq \left(\frac{(n-1)n^{d-1}-\lfloor\frac{n^a}{2}\rfloor}{(n-1)(n^{d-1}+\lfloor\frac{n^a}{2}\rfloor)}\right)^{\lfloor\frac{n^a}{2}\rfloor} \ll e^{-\frac{n^{2a-(d-1)}}{2+\epsilon}}.$$

for every $\epsilon > 0$. If j is negative (and therefore $j \leq -n^a$), the right hand side of (4-2) equals

$$\prod_{i=1}^{-j} \frac{(n-1)(n^{d-1}-i)}{(n-1)n^{d-1}+i} \leq \left(\frac{(n-1)(n^{d-1}-\lfloor\frac{n^a}{2}\rfloor)}{(n-1)n^{d-1}+\lfloor\frac{n^a}{2}\rfloor}\right)^{\lfloor\frac{n^a}{2}\rfloor} \ll e^{-\frac{n^{2a-(d-1)}}{2+\epsilon}}.$$

Since $2a > d-1$, we can apply Borel–Cantelli to obtain (3). □

References

[Borel 1983] A. Borel, "On free subgroups of semisimple groups", *Enseign. Math.* (2) **29**:1-2 (1983), 151–164.

[Garion and Shalev 2009] S. Garion and A. Shalev, "Commutator maps, measure preservation, and T–systems", *Trans. Amer. Math. Soc.* **361**:9 (2009), 4631–4651.

[Jambor et al. 2012] S. Jambor, M. W. Liebeck, and E. A. O'Brien, "Some word maps that are non-surjective on infinitely many finite simple groups", preprint, 2012. arXiv 1205.1952

[Larsen 1997] M. Larsen, "How often is a permutation an nth power?", preprint, 1997. arXiv math/9712223

[Larsen 2004] M. Larsen, "Word maps have large image", *Israel J. Math.* **139** (2004), 149–156.

[Larsen and Shalev 2009] M. Larsen and A. Shalev, "Word maps and Waring type problems", *J. Amer. Math. Soc.* **22**:2 (2009), 437–466.

[Larsen and Shalev 2012] M. Larsen and A. Shalev, "Fibers of word maps and some applications", *J. Algebra* **354** (2012), 36–48.

[Larsen and Shalev 2013] M. Larsen and A. Shalev, "On the distribution of values of certain word maps", preprint, 2013. arXiv 1308.1286

[Larsen et al. 2011] M. Larsen, A. Shalev, and P. H. Tiep, "The Waring problem for finite simple groups", *Ann. of Math.* (2) **174**:3 (2011), 1885–1950.

[Liebeck and Shalev 2005] M. W. Liebeck and A. Shalev, "Fuchsian groups, finite simple groups and representation varieties", *Invent. Math.* **159**:2 (2005), 317–367.

[Liebeck et al. 2010] M. W. Liebeck, E. A. O'Brien, A. Shalev, and P. H. Tiep, "The Ore conjecture", *J. Eur. Math. Soc.* (*JEMS*) **12**:4 (2010), 939–1008.

[Nica 1994] A. Nica, "On the number of cycles of given length of a free word in several random permutations", *Random Structures Algorithms* **5**:5 (1994), 703–730.

[Puder 2012] D. Puder, "Primitive words, free factors and measure preservation", preprint, 2012. arXiv 1104.3991

[Puder and Parzanchevski 2013] D. Puder and O. Parzanchevski, "Measure preserving words are primitive", preprint, 2013. arXiv 1202.3269

mjlarsen@indiana.edu *Department of Mathematics, Indiana University, Bloomington, IN 47405 United States*