Densities of Unavoidable Words

Danny Rorabaugh

Queen's University, Department of Mathematics & Statistics

rorabaugh@mast.queensu.ca

Free words over a finite alphabet–that is, elements of a free monoid with a finite generating set–are simple discrete objects with a natural substructure relation arising from homomorphic embedding. This has traditionally been studied in terms of pattern avoidance: determining when no substring of a word is a homomorphic image of a pattern word. Given recent developments in extremal graph theory and combinatorial limit theory, we introduce and explore asymptotic word densities.

This is joint work with Joshua Cooper, University of South Carolina.

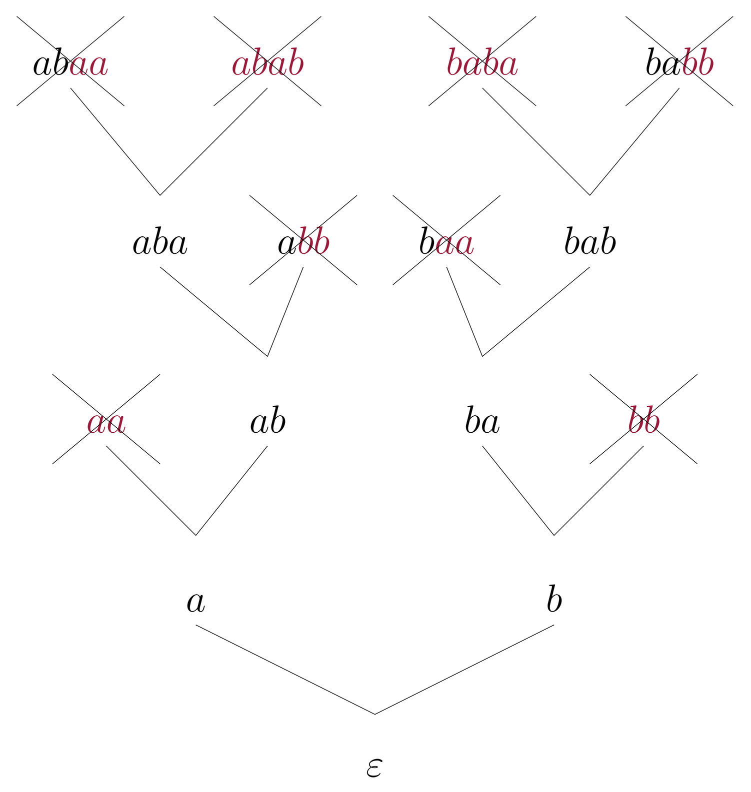
Terminology

$mumbojumbo$ is an **instance** of $uhoh$.

$mississippi$ is an instance of $bananas$.

$S\tilde{a}oPauloSchool$ **encounters** $stats$.

$bananasplit$ encounters xxy.



$yahoo$ **avoids** xxy.

$haha$ avoids xxy.

Only seven binary words avoid xx.

We've known for over a century (Thue, 1906) that arbitrarily long ternary words avoid xx.

A word W is **unavoidable** provided over any finite alphabet, only finitely many words avoid W.

The word $postdoctoral$ encounters WOW in 4 ways:

$$postdoctoral,\quad postdoctoral,\quad postdoctoral,\quad postdoctoral.$$

The **density** of WOW in $postdoctoral$ is thus

$$\delta(WOW, postdoctoral) = \frac{4}{\#\{\text{substrings in } postdoctoral\}} = \frac{4}{\binom{12+1}{2}}.$$

Denote by $\mathbb{E}(V,q,n)$ the expected density $\mathbb{E}(\delta(V,W_n))$ with $W_n \in [q]^n$ chosen uniformly at random.

Results

Cooper-R. (2016+)

Let V be a nonempty word. Fix q. The limit expectation $\delta(V,q) = \lim_{n\to\infty} \mathbb{E}(V,q,n)$ exists.
In particular, for $q \geq 2$, the following are equivalent:

- $\delta(V,q) = 0$;
- V is doubled (every letter in V appears at least twice).

Moreover, if V is doubled, then as $n \to \infty$,

$$\mathbb{E}(V,q,n) \sim \frac{1}{n}; \quad \mathrm{Var}(V,q,n) = O\left(\mathbb{E}(V,q,n)^2 \frac{(\log n)^3}{n}\right).$$

Define the **Zimin words** recursively as follows:

$$Z_0 = \varepsilon,\ Z_1 = \mathbf{a},\ Z_2 = a\mathbf{b}a,\ Z_3 = aba\mathbf{c}aba, Z_4 = abacaba\mathbf{d}abacaba,\ \ldots.$$

Zimin (1982)

Word W with n distinct letters is unavoidable if and only if Z_n encounters W.

As Zimin words are the maximal unavoidable patterns, we considered them worthy of special investigation. The follows are some results on asymptotic densities of Zimin words over a fixed alphabet.

Cooper-R. (2016)

Expected densities in random words:

q	2	3	4	5	6	$\cdots$
$\delta(Z_2,q)$	0.7322132	0.4430202	0.3122520	0.2399355	0.1944229	$\cdots$
$\delta(Z_3,q)$	0.1194437	0.0183514	0.0051925	0.0019974	0.0009253	$\cdots$

$$\prod_{1\leq i<n} \frac{1}{q^{(2^i-1)}} \leq \delta(Z_n,q) \leq \prod_{1\leq i<n} \frac{1}{q^{(2^i-1)}-1}.$$

Minimum densities over all words with fixed q:

$$\liminf_{\substack{W\in[q]^n\\ n\to\infty}} \delta(Z_2,W) = \frac{1}{q}.$$

$$\liminf_{\substack{W\in[q]^n\\ n\to\infty}} \delta(Z_3,W) \geq \frac{1}{(2q-1)^2 q! 2^q}.$$

$$\liminf_{\substack{W\in[2]^n\\ n\to\infty}} \delta(Z_3,W) \geq \frac{1}{54}.$$

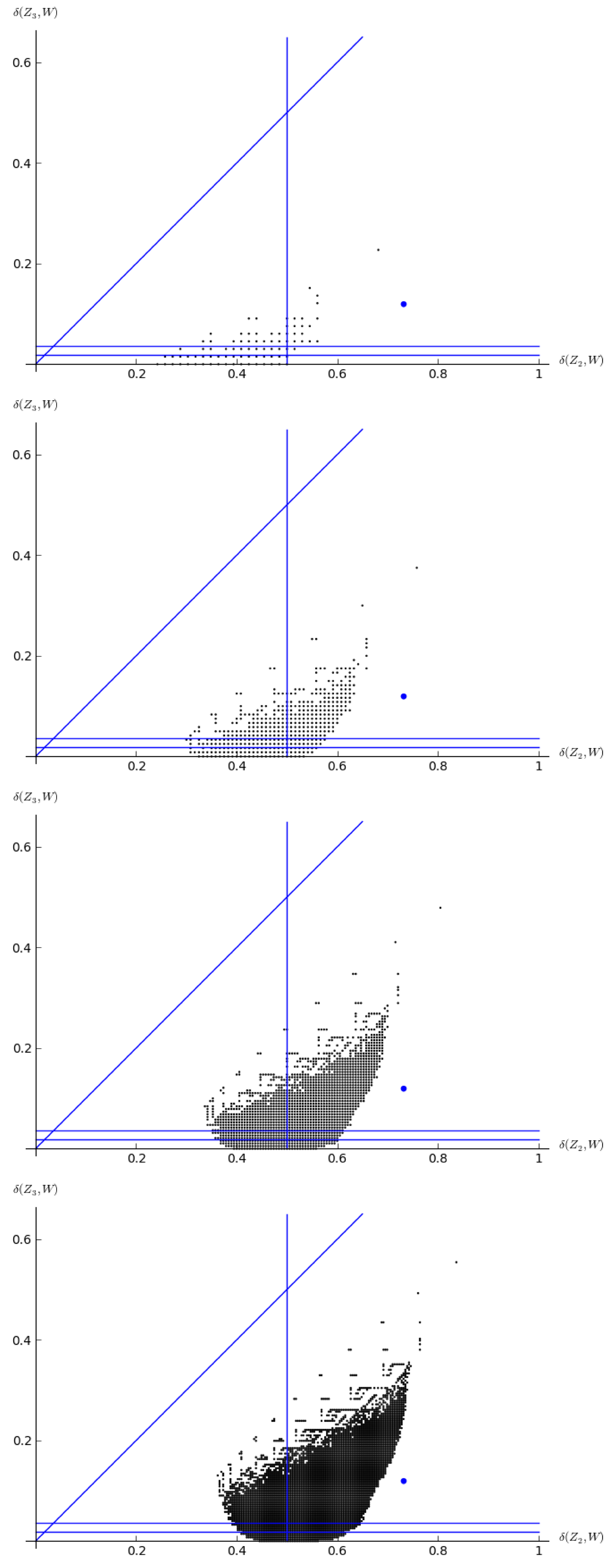


Figure 1: Z_2 and Z_3 densities in all binary words of length $11, 15, 19, 23$.

Parting Questions

- What is the asymptotic minumum density of Z_3?
  One computations approach suggests at least 0.102.
  Does it equal the asymptotic expected density (≈ 0.119)?
- What is lower-bounding curve for Z_3-density in terms of Z_2-density?

References

[1] J. Cooper and D. Rorabaugh. Density dichotomy in random words.

[2] J. Cooper and D. Rorabaugh. Asymptotic Density of Zimin Words. *Discrete Mathematics & Theoretical Computer Science*, 2016.

[3] A. Thue. *Über unendliche Zeichenreihen*, volume 7 of *Norske Vid. Skrifter I Mat.-Nat. Kl.* Kristiania, 1906.

[4] A. I. Zimin. Blokirujushhie mnozhestva termov. *Mat. Sb.*, 119:363–375, 1982.

[5] A. I. Zimin. Blocking sets of terms. *Math. USSR-Sb.*, 47:353–364, 1984.