Density Dichotomy in Random Words

28th Cumberland Conference, USC

Joshua Cooper & Danny Rorabaugh

University of South Carolina

2015 May 15
Word Instances

**Definition**

A *q-ary word* is a string of characters on a *q*-letter alphabet.

**Definition**

Word $W$ is an *instance* of $V$ provided

- $V = x_1 x_2 \cdots x_n$ where each $x_i$ is a letter;
- $W = A_1 A_2 \cdots A_n$ where each $A_i$ is a positive length word and $A_i = A_j$ whenever $x_i = x_j$.

Alternatively definable by a monoid homomorphism $\phi : x_i \mapsto A_i \neq \varepsilon$.

Examples: *oompaloopma* is an instance of *wow*;
- *mississippi* is an instance of *bananas*.
**Definition**

A word $U$ *encounters* word $V$ provided some factor (substring) $W \leq U$ is an instance of $V$.

**Definition**

If $U$ fails to encounter $V$, we say that $U$ *avoids* $V$.

**Example**

Binary words that avoid $xx$:
We see that only seven binary words avoid xx.

However, it has been known for over a century (Thue, 1906) that arbitrarily long ternary words avoid xx.

**Definition**

A word $W$ is *unavoidable* provided: for any finite alphabet, only finitely many words avoid $W$; that is, all sufficiently long words encounter $W$. 
**Unavoidability**

**Definition**
A word $W$ is *unavoidable* provided: for any finite alphabet, only finitely many words avoid $W$; that is, all sufficiently long words encounter $W$.

The unavoidable words were first classified by Bean, Ehrenfeucht, and McNulty (1979) and independently by Zimin (1982).

**Definition**
Define the $n^{th}$ *Zimin word* recursively by $Z_0 = \emptyset$ and $Z_{n+1} = Z_n x_{n+1} Z_n$.

$Z_1 = a, \quad Z_2 = aba, \quad Z_3 = abacaba, \quad Z_4 = abacabadaacaba, \ldots$

**Theorem (A.I. Zimin, 1982)**
A word $W$ with $n$ distinct letters is unavoidable iff $Z_n$ encounters $W$. 
Words, Meet Ramsey

Fix an unavoidable word $Z$. What word-length guarantees a $Z$-encounter?

In 2014, J. Cooper and D. Rorabaugh (USC), J. Tao (CalTech), and W. Rytter and A. Shur (Warsaw U, Ural Federal U) all posted manuscripts to the arXiv with various bounds for the following function:

Define $f(n,q)$ to be the smallest integer $M$ such that every $q$-ary word of length $M$ encounters $Z_n$.

**Theorem (2014+)**

$$q^{2^{(n-1)(1+o(1))}} \leq f(n,q) \leq (2q + 1) \uparrow\uparrow (n - 1).$$
Over the last decade or so, there have been amazing developments in extremal graph theory and the combinatorial limit theory of graphs. Consider:

- Flag algebras: Razborov;
- Graph homomorphisms and graph limits: Lovász, Sós, Szegedy, etc.

It is worth applying these perspectives, such as homomorphism densities, to other discrete structures.

Now a major part of my dissertation is on word densities.
The word *huszonnyolc* encounters *lászló,lászló*:

- *huszonnyolc*,
- *huszonnyolc*.

We say the density of *lászló* in *huszonnyolc*, $\delta(lászló, huszonnyolc)$, is

$$\frac{2}{\#\{\text{substrings in } huszonnyolc\}} = \frac{2}{\binom{12}{2}}.$$
Let $V$ be a nonempty word. Fix alphabet $\Sigma$ with $q \geq 2$ letters, and let $W_n \in \Sigma^n$ be chosen uniformly at random. The following are equivalent:

1. $\lim_{n \to \infty} \mathbb{E}(\delta(V, W_n)) = 0$;
2. $V$ is doubled (that is, every letter in $V$ appears at least twice).

Theorem (... and if $V$ is doubled...)

\[ \frac{1}{n} \ll \mathbb{E}(\delta(V, W_n)) \ll \frac{\log n}{n}; \]

\[ \text{Var}(\delta(V, W_n)) \ll \mathbb{E}(\delta(V, W_n))^2 \frac{(\log n)^3}{n}. \]
Let $V$ be a nondoubled word and $\Sigma$ an alphabet. We know $\mathbb{E}(\delta(V, W_n))$ does not approach 0, for random $W_n \in \Sigma^n$. Q: Does it converge? What to?

Let $I_n(V, q)$ be the probability that a randomly selected $q$-ary word of length $n$ is an instance of $V$. That is,

$$I_n(V, q) = \frac{|\{W \in [q]^n \mid W \text{ is an instance of } V\}|}{q^n}.$$

For example:

- $I_5(\text{lászló}, q) = 0$;
- $I_6(\text{lászló}, q) = \frac{1}{q}$;
- $I_7(\text{lászló}, q) = \frac{2q-1}{q^2}$.
Theorem (Cooper-R. 2015+)

For word $V$, integer $q$, and $W_n \in [q]^n$ chosen uniformly at random,

$$\mathbb{I}(V, q) := \lim_{n \to \infty} \mathbb{I}_n(V, q) = \lim_{n \to \infty} \mathbb{E}(\delta(V, W_n)).$$

Examples:

- Let $V = abcdef \cdots k$ consist of $k$ distinct letters. Then $\mathbb{I}_n(V, q) = 1$ for $n \geq k$, so $\mathbb{I}(V, q) = 1$.

- Recall the Zimin words: $Z_2 = aba$, $Z_3 = abacaba$, ...

<table>
<thead>
<tr>
<th>$q$</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>\ldots</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mathbb{I}(Z_2, q)$</td>
<td>0.7322132</td>
<td>0.4430202</td>
<td>0.3122520</td>
<td>0.2399355</td>
<td>\ldots</td>
</tr>
<tr>
<td>$\mathbb{I}(Z_3, q)$</td>
<td>0.1194437</td>
<td>0.0183514</td>
<td>0.0051925</td>
<td>0.0019974</td>
<td>\ldots</td>
</tr>
</tbody>
</table>

$$\prod_{1 \leq i < n} \left( q^{(2^i-1)} - 1 \right)^{-1} \leq \mathbb{I}(Z_n, q) \leq \prod_{1 \leq i < n} \left( q^{(2^i-1)} - 1 \right)^{-1}.$$
Possible $Z_2$- and $Z_3$-Densities
Possible $Z_2$- and $Z_3$-Densities
Possible $\mathbb{Z}_2$- and $\mathbb{Z}_3$-Densities

\[ \delta(\mathbb{Z}_3, W) \]

\[ \delta(\mathbb{Z}_2, W) \]
Possible $Z_2$- and $Z_3$-Densities
Possible $Z_2$- and $Z_3$-Densities

\[ \delta(Z_3,W) \]

\[ \delta(Z_2,W) \]
Possible $Z_2$- and $Z_3$-Densities
Possible $Z_2$- and $Z_3$-Densities
Possible $Z_2$- and $Z_3$-Densities
Possible $\mathbb{Z}_2$- and $\mathbb{Z}_3$-Densities
Possible $Z_2$- and $Z_3$-Densities
Possible $Z_2$- and $Z_3$-Densities

\[
\delta(Z_3, W) \quad \delta(Z_2, W)
\]
Possible $Z_2$- and $Z_3$-Densities

\[ \delta(Z_3, W) \]

\[ \delta(Z_2, W) \]
Possible $Z_2$- and $Z_3$-Densities
Possible $Z_2$- and $Z_3$-Densities
Possible $Z_2$- and $Z_3$-Densities
Possible $Z_2$- and $Z_3$-Densities
Possible $Z_2$- and $Z_3$-Densities
Possible $\mathbb{Z}_2$- and $\mathbb{Z}_3$-Densities
Possible $Z_2$- and $Z_3$-Densities
\( I_n(V, q) \) for doubled \( V \)

Let \( V \) be doubled with \( k \) distinct letters and \( r \geq 2 \) the minimum letter multiplicity.

We know \( I(V, q) = 0 \).

When \( r \mid (n + |V|) \),

\[
q^{k-|V|/r} q^{-n(1-\frac{1}{r})} \leq I_n(V, q) < n^{k+1} q^{-n(1-\frac{1}{r})}.
\]
Conclusion

For more details, see our papers on the arXiv or my dissertation in August.

Happy Birthday Conference, László Székely!