Two arithmetical sources and their associated tries

Valérie Berthé¹, Eda Cesaratto², Frédéric Paccaut³, Pablo Rotondo⁴, Martín D. Safe⁵ and Brigitte Vallée⁶

¹IRIF, CNRS and Université de Paris, France
²CONICET and Instituto del Desarrollo Humano, Univ. Nac. de General Sarmiento, Argentina
³LAMFA CNRS UMR 7352, Université de Picardie Jules Verne, France
⁴LITIS, Université de Rouen, France
⁵INMABB, Depto. de Matemática, Universidad Nacional del Sur-CONICET, Argentina
⁶GREYC, CNRS and Université de Caen, France

AofA2020
Introduction

Analytic Combinatorics

- Information Theory
- Number Theory
- Data Structures
Introduction

Source and entropy

Data Structures

Analytic Combinatorics

Number Theory
Introduction

Source and entropy

Data Structures

Analytic Combinatorics

Farey and Stern-Brocot partitions
Introduction

Source and entropy

Dirichlet generating functions

Tries

Farey and Stern-Brocot partitions
Sources:
- We consider sources defined by two classical related partitions:
  - Sturm source (Farey partitions)
  - Stern-Brocot source (Stern-Brocot partitions).
- The Sturm source produces characteristic Sturmian words.
- These two partitions have zero Shannon entropy ($h = 0$).

Tries:
- We use tries to differentiate these two sources.
- For good sources, with entropy $h > 0$,
  
  \[
  \text{the average trie depth is } \sim \frac{\log n}{h}.
  \]

Our main result is the following:

**Theorem**

The average trie depths for these sources are:

- Sturm source: $\sim \frac{24}{\pi^{3/2}} \sqrt{n}$
- Stern-Brocot source: $\sim \frac{3}{\pi^2} \log^2 n$
Overview

Sources, partitions, and expressions for DGFs

Tries, average depth, and the Rice method

Analytical study of the DGFs
Overview

Sources, partitions, and expressions for DGFs

Tries, average depth, and the Rice method

Analytical study of the DGFs
Let $\Sigma = \{0, 1\}$ and let $\{\mathcal{P}_k\}_{k \geq 0}$ be a family of partitions of $[0, 1]$ $\mathcal{P}_k = \{J_w : w \in \Sigma^k\}$ where $\max_{w \in \Sigma^k} |J_w| \rightarrow 0$ $(k \rightarrow \infty)$ consisting of closed intervals such that:

- $J_e = [0, 1]$
- For $k \geq 0$ and $w \in \Sigma^k$, $J_{w0}$ and $J_{w1}$ partition $J_w$ (left to right).
Dirichlet generating functions (DGFs)

Such a family of partition defines a probabilistic source where

\[ p_w := |I_w| \]

is the probability of emitting prefix \( w \).

The Dirichlet generating function (DGF) of this source is defined as:

\[
\Lambda(s) := \sum_{w \in \Sigma^*} p_w^s = \sum_{k \geq 0} \sum_{w \in \Sigma^k} p_w^s.
\]
Sturm and Stern-Brocot sources

Sturm source is defined by Farey partitions $S_k$

$S_0 := [0/1, 1/1]$

$S_k$ arises from $S_{k-1}$ by dividing each interval $[a/c, b/d]$ by its mediant $(a + b)/(c + d)$ only if $c + d \leq k + 1$

Stern-Brocot source is defined by Stern-Brocot partitions $B_k$

$B_0 := [0/1, 1/1]$

$B_k$ arises from $B_{k-1}$ by dividing each interval $[a/c, b/d]$ by its mediant $(a + b)/(c + d)$ (always)
Sturm and Stern-Brocot sources

Sturm source is defined by Farey partitions $S_k$

$S_0 := [0/1, 1/1]$

$S_k$ arises from $S_{k-1}$ by dividing each interval $[a/c, b/d]$ by its mediant $(a + b)/(c + d)$ only if $c + d \leq k + 1$

Stern-Brocot source is defined by Stern-Brocot partitions $B_k$

$B_0 := [0/1, 1/1]$

$B_k$ arises from $B_{k-1}$ by dividing each interval $[a/c, b/d]$ by its mediant $(a + b)/(c + d)$ (always)
DGF for the Sturm source

If \( I_w = \left[ \frac{a}{c}, \frac{b}{d} \right] \in S_k \), then \( |I_w| = \frac{1}{cd} \) and \( \gcd(c, d) = 1 \).

The set of pairs \((c, d)\) for which some \( \left[ \frac{a}{c}, \frac{b}{d} \right] \in S_k \) is

\[ C_k = \{(c, d) : \max(c, d) \leq k + 1, \ c + d > k + 1, \ \text{and} \ \gcd(c, d) = 1\} \]

Since \( |\{k \geq 0 : (c, d) \in C_k\}| = \min(c, d) \) whenever \( \gcd(c, d) = 1 \),

\[
\Lambda(s) = \sum_{k \geq 0} \sum_{(c, d) \in C_k} \left( \frac{1}{cd} \right)^s = \sum_{c \geq 1, d \geq 1 \atop \gcd(c, d) = 1} \frac{\min(c, d)}{(cd)^s}.
\]

Proposition

The DGF \( \Lambda \) for the Sturm source is:

\[
\Lambda(s) = 1 + 2 \frac{\zeta(s, s - 1)}{\zeta(2s - 1)} \quad \text{where} \quad \zeta(\alpha, \beta) := \sum_{c \geq 1} \frac{1}{c^\beta} \sum_{d : d > c} \frac{1}{d^\alpha}.
\]

Here:

- the denominator \( \zeta(2s - 1) \) arises from the \( \gcd(c, d) = 1 \) condition,
- the numerator \( \zeta(s, s - 1) \) arises from the \( \min(c, d) \) term.
DGF for the Stern-Brocot source

The dynamical system associated with the Farey map $T : [0, 1] \to [0, 1]$

$$
\chi \mapsto \begin{cases} 
\frac{\chi}{1 - \chi} & \text{if } \chi \in [0, 1/2] \\
\frac{(1 - \chi)}{\chi} & \text{if } \chi \in [1/2, 1]
\end{cases}
$$
generates the Stern-Brocot partitions $\mathcal{B}_k$.

Let:
- $a$ be the inverse of the left branch of $T$,
- $b$ be the inverse of the right branch of $T$,
and $\mathcal{H} = \{a, b\}$.

$\mathcal{H}^k := \{a, b\}^k$ is the set of inverses branches of $T^k$ and generates $\mathcal{B}_k$:

$$
\mathcal{B}_k = \{h([0, 1]) : h \in \mathcal{H}^k\}.
$$
DGF for the Stern-Brocot source

Since

$$\mathcal{B}_k = \{h([0, 1]) : h \in \mathcal{H}^k\},$$

for each $k \geq 0$ we have that

$$\Lambda_k(s) = \sum_{w \in \Sigma^k} p_w^s = \sum_{h \in \mathcal{H}^k} |h(1) - h(0)|^s = \mathbf{H}_s^k[1](0, 1)$$

where $\mathbf{H}_s$ is a variant of the transfer operator for the Farey map.

Hence, if $\mathcal{H}^* = \bigcup_{k \geq 0} \mathcal{H}^k$, then

$$\Lambda(s) = \sum_{w \in \Sigma^*} p_w^s = \sum_{h \in \mathcal{H}^*} |h(1) - h(0)|^s = (I - \mathbf{H}_s)^{-1}[1](0, 1).$$
The induced map of the Farey map (associated with Stern-Brocot partitions) is the **Gauss map** (associated with continued fractions).

In fact, \((a^{m-1} \circ b)(x) = \frac{1}{m + x}\) (an inverse branch of the Gauss map).

Note that \(\mathcal{H}^* = \{a, b\}^* = (a^* \circ b)^* \circ a^*\).

**Proposition**

For the DGF of the Stern-Brocot source,

\[
\Lambda(s) = (I - H_s)^{-1}[1](0, 1) = (I - A_s)^{-1}(I - G_s)^{-1}[1](0, 1)
\]

where

- \(A_s\) is the part of \(H_s\) corresponding to inverse branch \(a\), and
- \(G_s\) is a variant of the transfer operator of the Gauss map.
Overview

Sources, partitions, and expressions for DGFs

Tries, average depth, and the Rice method

Analytical study of the DGFs
Tries from words emitted by a source

\begin{align*}
\chi_1 &= \texttt{cacc}a\texttt{aacc}\ldots \\
\chi_2 &= \texttt{aaac}a\texttt{c}b\texttt{a}c\ldots \\
\chi_3 &= \texttt{aa}ab\texttt{c}b\texttt{bb}ca\ldots \\
\chi_4 &= \texttt{bacba}a\texttt{aab}c\ldots \\
\chi_5 &= \texttt{bc}ab\texttt{bbbb}b\texttt{c}\ldots \\
\chi_6 &= \texttt{ba}c\texttt{a}ab\texttt{bb}ba\ldots \\
\chi_7 &= \texttt{cc}b\texttt{a}bc\texttt{bcb}b\ldots
\end{align*}

\begin{itemize}
\item Let $\mathcal{T}$ be a trie on $n$ words independently drawn from a source.
\item We perform a probabilistic analysis of the shape of $\mathcal{T}$.
\item We focus on trie depth $D_n$, the depth of a random branch.
\item If $D_n^{(i)}$ is the depth of the branch corresponding to the $i$-th word,
\end{itemize}

\[
\Pr[D_n \geq k + 1] = \frac{1}{n} \sum_{i=1}^{n} \Pr[D_n^{(i)} \geq k + 1] = \sum_{w \in \Sigma^k} p_w [1 - (1 - p_w)^{n-1}].
\]
Average trie depth and DGFs

Proposition

(i) If $\Lambda(s)$ is well-defined for $s \geq 2$, then

$$\mathbb{E}[D_n] = \frac{1}{n} \sum_{\ell=2}^{n} (-1)^\ell \binom{n}{\ell} \ell \Lambda(\ell).$$

(ii) If $\exists a, A > 0$ such that $\forall k \geq 1, \exists w \in \Sigma^k$ such that $p_w \geq Ak^{-a}$,

$$\mathbb{E}[D^r_n] = \infty, \quad \forall r \geq 2.$$ 

Remarks

- Both statements (i) and (ii) apply to our two sources.
- Statement (ii) applies to our sources with $w = 000\ldots0$ ($k$ times).
- In order to find estimates for $\mathbb{E}[D_n]$, we will study

$$f(n) := n\mathbb{E}[D_n] = \sum_{\ell=2}^{n} (-1)^\ell \binom{n}{\ell} \ell \Lambda(\ell).$$

$$=: p(\ell).$$
Rice method: Step 1 – Integral representation

It applies to binomial sums

\[ f(n) = \sum_{\ell=a}^{\infty} (-1)^\ell \binom{n}{\ell} p(\ell) \]

for some \( \ell \mapsto p(\ell) \) for \( \ell \geq a \) (where \( a \geq 0 \)).

---

Rice method: Step 1

If \( \psi(s) \) is a lifting of \( p(\ell) \) and \( \exists c \in ]a - 1, a[ \) such that:

\( \psi(s) \) is analytic and of polynomial growth in the half-plane \( \Re s > c \),

then \( \forall b \in ]c, a[ \) and sufficiently large \( n \):

\[ f(n) = \frac{1}{2\pi i} \int_{b-i\infty}^{b+i\infty} L_n(s) \cdot \psi(s) \, ds \]

where \( L_n(s) = \frac{\Gamma(n + 1)\Gamma(-s)}{\Gamma(n + 1 - s)} \).
Rice method: Step 2 – shifting to the left

The integral representation is shifted to the left, using the *tameness* of $\psi$ to the left (meromorphic and of polynomial growth).

### Rice method: Step 2

If for some $\delta_0 > 0$, the lifting $\psi$ satisfies that:

- it is *meromorphic* on $\Re s > c - \delta_0$,
- its only pole on $\Re s > c - \delta_0$ is at $s = c$, and
- $\forall \delta < \delta_0$, $\psi(s)$ is of *polynomial growth* on $\Re s \geq c - \delta$ as $|\Im s| \to \infty$,

then $\forall \delta < \delta_0$

$$f(n) = \text{Res}[L_n(s) \cdot \psi(s); s = c] + O(n^{c-\delta}) \quad (n \to \infty).$$
Rice method: Step 3 – estimating the residue

The residue

$$\text{Res}[L_n(s) \cdot \psi(s); s = c] = n^c \cdot P(\log n) \left[ 1 + O(1/n) \right]$$

where $P$ is a polynomial determined by the singular expression

$$\psi(s) \underset{\sim}{=} a_d \frac{1}{(s-c)^d} + \cdots + a_1 \frac{1}{s-c} + a_0,$$

according to the following two cases:

1. If $c$ is not an integer, then

   $P$ has degree $d - 1$ and leading coefficient $\frac{\Gamma(-c)}{\Gamma(d)} |a_d|$.

2. If $c$ is an integer, then

   $P$ has degree $d$ and leading coefficient $\frac{1}{\Gamma(d+1)} |a_d|$.

   If $c = 1$, the factor $[1 + O(1/n)]$ equals 1.
Overview

Sources, partitions, and expressions for DGFs

Tries, average depth, and the Rice method

Analytical study of the DGFs
Singularity analysis for the Sturm source

Recall that for Sturm source: \( \Lambda(s) = 1 + 2 \frac{\zeta(s, s - 1)}{\zeta(2s - 1)} \).

Because of well-known properties of the zeta function:

**Proposition**

For any \( a_0 > 0 \), the DGF \( \Lambda(s) \) of the Sturm source satisfies:

- it is meromorphic on \( \Re s > 1 + a_0 \),
- its only pole is a simple pole at \( s = 3/2 \),
- \( \forall a > a_0 \), it is of polynomial growth on \( \Re s \geq 1 + a \).

Moreover, as \( s \to 3/2 \),

\[
\begin{align*}
    s\Lambda(s) &\sim \frac{36}{\pi^2} \left( \frac{1}{2s - 3} \right) \\
    \Gamma(-s) \cdot s\Lambda(s) &\sim \frac{36}{\pi^2} \Gamma\left( -\frac{3}{2} \right) \left( \frac{1}{2s - 3} \right).
\end{align*}
\]

Since \( \Gamma(-3/2) = (4/3)\sqrt{\pi} \), in Step 3 of the Rice method we have:

The DGF \( \Lambda(s) \) of the Sturm source satisfies:

\[
\text{Res} \left[ L_n(s) \cdot s\Lambda(s); s = \frac{3}{2} \right] = \frac{24}{\pi^{3/2} n^{3/2}} \left( 1 + O \left( \frac{1}{n} \right) \right).
\]
Singularity analysis for the Stern-Brocot source

Recall that for Stern-Brocot: $\Lambda(s) = (I - A_s)^{-1}(I - G_s)^{-1}[1](0, 1)$. Because of deep properties of the quasi-inverse of $G_s$:

**Proposition**

The DGF $\Lambda(s)$ of the Stern-Brocot source satisfies:

- it is meromorphic on $\Re s > 1 - \delta_0$ (for some $\delta_0 > 0$),
- its only pole is at $s = 1$ and is of order 2,
- $\forall \delta < \delta_0$, it is of polynomial growth on $\Re s \geq 1 - \delta$.

Moreover, as $s \to 1$,

$$s\Lambda(s) \sim \frac{1}{\zeta(2)} \left( \frac{1}{s - 1} \right)^2$$

and

$$\Gamma(-s) \cdot s\Lambda(s) \sim \frac{6}{\pi^2} \left( \frac{1}{s - 1} \right)^3.$$

Hence, for Step 3 of the Rice method we have:

The DGF $\Lambda(s)$ of the Stern-Brocot source satisfies:

$$\text{Res}[L_n(s) \cdot s\Lambda(s); s = 1] = n \left( \frac{3}{\pi^2} \log^2 n + b_1 \log n + b_0 \right)$$

for some constants $b_1$ and $b_0$. 
Main results

We are ready to apply the Rice method to

\[ f(n) := n \mathbb{E}[D_n] = \sum_{\ell=2}^{n} (-1)^\ell \binom{n}{\ell} \ell \Lambda(\ell) \]

for our sources. We show that the two sources behave very differently.

**Theorem**

For each source, consider a trie built on \( n \) words independently drawn from the source. Then, the mean values of the trie depths are:

- For the Sturm source:
  \[ \mathbb{E}[D_n] = \frac{24}{\pi^{3/2}} n^{1/2} + O(n^a) \quad \text{for any } a > 0. \]

- For the Stern-Brocot source:
  \[ \mathbb{E}[D_n] = \frac{3}{\pi^2} \log^2 n + b_1 \log n + b_0 + O(n^{-\delta}) \quad \text{for some } \delta > 0 \]

and some constants \( b_1 \) and \( b_0 \).
Concluding remarks and future work

These are instances of tries built on seven words emitted from the Sturm source (on the left), the Stern-Brocot source (in the middle), and the Bernoulli source with $p = 1/2$ (on the right).

As the value $n = 7$ is small, and the moments $\mathbb{E}[D_n^2]$ for the Sturm source and the Stern-Brocot are infinite, there does not really exist a “typical trie” for these sources.
Concluding remarks and future work

This work appears as (one of) the first study on sources of zero Shannon entropy via Analytic Combinatorics tools.

1. Rényi entropy:
   - The Rényi entropy for our two sources are very similar
   - Known via Number Theory arguments
   - TO DO: Derive these results using Analytic Combinatorics.

2. The VLMC (Variable Length Markov Chain):
   - Are the simplest source where dependency from the past is unbounded
   - The depth of the associated suffix tries has been studied before on a special class of VLMC sources
   - TO DO: Analyze the trie depth in an intermittent subclass

3. Trie built on the Farey dynamical source
   - Its invariant measure is $1/t$ which has infinite mass
   - Not clearly related fundamental probabilities with Stern-Brocot source
   - Strongly different from absolutely continuous invariant measure
   - TO DO: We wish to analyze its trie depth
Thank you very much for your attention!