Rauzy dimension and entropy

Verónica Becher[1] Olivier Carton[2] Santiago Figueira[1]

[1]Universidad de Buenos Aires & CONICET

[2]IRIF, Université Paris Cité & CNRS & IUF

[1,2]Supported by LIA SINFIN

Uniform Distribution of Sequences – ESI

Outline

Rauzy dimensions

Measure-theoretic entropy

Results

$\beta(x)$ vs $\gamma(x)$

Preservation of normality by addition

Theorem (Rauzy, 1976)

For $\beta \in [0,1)$, the following conditions are equivalent:

- *for each α normal, $\alpha + \beta$ is still normal,*
- *β has Rauzy dimension 0.*

Theorem (Rauzy, 1976)

For $\beta \in [0,1)$, the following conditions are equivalent:

- *β is normal,*
- *β has Rauzy dimension $(\#A - 1)/\#A$.*

Outline

Rauzy dimensions

Measure-theoretic entropy

Results

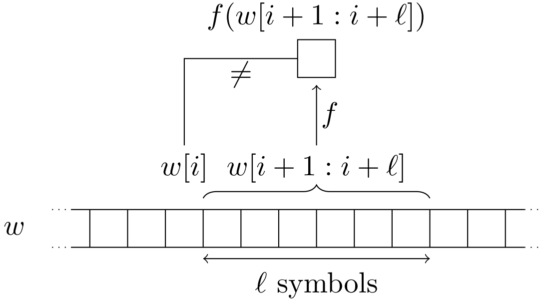
$\beta(x)$ vs $\gamma(x)$

Mismaches of prediction

Suppose that $\ell \geqslant 0$ and $f : A^\ell \to A$ are given.
For $w \in A^*$,

$$\#\{i : w[i] \neq f(w[i+1 : i+\ell])\}$$

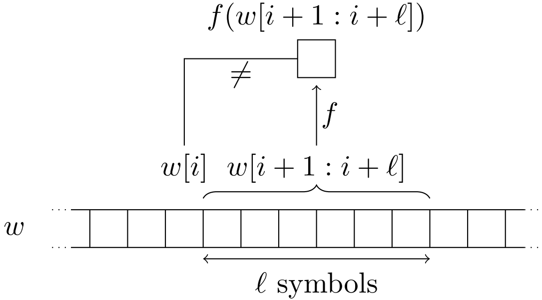
is the number of mismatches between $w[i]$ and $f(w[i+1 : i+\ell])$.



Rauzy dimensions

For $w \in A^*$ and $0 \leqslant \ell \leqslant |w|$,

$$\beta_\ell(w) := \min_{f : A^\ell \to A} \frac{\#\{i : w[i] \neq f(w[i+1 : i+\ell])\}}{|w|}$$

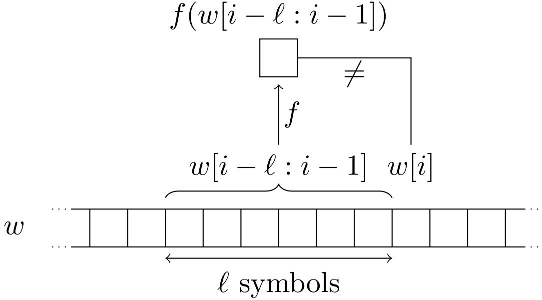


Rauzy dimensions

For $w \in A^*$ and $0 \leqslant \ell \leqslant |w|$,

$$\beta_\ell(w) := \min_{f:A^\ell \to A} \frac{\#\{i : w[i] \neq f(w[i+1 : i+\ell])\}}{|w|}$$

$$\gamma_\ell(w) := \min_{f:A^\ell \to A} \frac{\#\{i : w[i] \neq f(w[i-\ell : i-1])\}}{|w|}$$



Rauzy dimensions

For $w \in A^*$ and $0 \leqslant \ell \leqslant |w|$,

$$\beta_\ell(w) := \min_{f:A^\ell \to A} \frac{\#\{i : w[i] \neq f(w[i+1 : i+\ell])\}}{|w|}$$
$$\gamma_\ell(w) := \min_{f:A^\ell \to A} \frac{\#\{i : w[i] \neq f(w[i-\ell : i-1])\}}{|w|}$$

For $x \in A^{\mathbb{N}}$ and $\ell \geqslant 0$,

$$\underline{\beta}_\ell(x) := \liminf_{n\to\infty} \beta_\ell(x[1:n]) \quad \text{and} \quad \overline{\beta}_\ell(x) := \limsup_{n\to\infty} \beta_\ell(x[1:n])$$
$$\underline{\beta}(x) := \lim_{\ell\to\infty} \underline{\beta}_\ell(x) \quad \text{and} \quad \overline{\beta}(x) := \lim_{\ell\to\infty} \overline{\beta}_\ell(x)$$

$\underline{\gamma}(x)$ and $\overline{\gamma}(x)$ are defined similarly using γ_ℓ instead of β_ℓ.

Examples of Rauzy dimensions

Suppose that x is ultimately periodic, that is $x = uv^{\mathbb{N}}$, like

$$00(011)^{\mathbb{N}} = 00011011011011 \cdots$$

For $\ell \geqslant |uv|$, $\overline{\beta}_\ell(x) = \overline{\gamma}_\ell(x) = 0$ and thus $\overline{\beta}(x) = \overline{\gamma}(x) = 0$.

Suppose that x is Sturmian like

$$0100101001001 \cdots$$

For each $\ell \geqslant 0$, there are exactly $\ell + 1$ factors of length ℓ with only one with two extensions to the left (to the right resp.). Mismatches can only occur with this special factor.

$$\overline{\beta}_\ell(x) \leqslant \text{frequency of the left special factor}$$
$$\overline{\gamma}_\ell(x) \leqslant \text{frequency of the right special factor}$$

$$\overline{\beta}(x) = \overline{\gamma}(x) = 0$$

Questions

What about polynomial factor complexity, that is sequences x such that

$$\#(\mathrm{Fact}(x) \cap A^\ell) \leqslant P(\ell)$$

for some polynomial $P(x)$.

What about sub-exponential factor complexity, that is sequences x such that, for instance,

$$\#(\mathrm{Fact}(x) \cap A^\ell) \leqslant e^{\sqrt{\ell}}$$

Outline

Measure-theoretic entropy

Recall that $|w|_u$ is the number of occurrences of u in w.

Normalized ℓ-entropy of $w \in A^*$:

$$h_\ell(w) := -\frac{1}{\ell} \sum_{u \in A^\ell} f_u \log f_u \quad \text{where} \quad f_u := \frac{|w|_u}{|w| - |u| + 1}$$

Note that

$$\sum_{u \in A^\ell} f_u = 1 \quad \text{and} \quad 0 \leqslant h_\ell(w) \leqslant 1.$$

Normalized entropy of $x \in A^{\mathbb{N}}$:

$$\underline{h}(x) := \liminf_{\ell \to \infty} \underline{h}_\ell(x) \quad \text{where} \quad \underline{h}_\ell(x) := \liminf_{n \to \infty} h_\ell(x[1:n])$$

$$\overline{h}(x) := \liminf_{\ell \to \infty} \overline{h}_\ell(x) \quad \text{where} \quad \overline{h}_\ell(x) := \limsup_{n \to \infty} h_\ell(x[1:n])$$

Other definitions of entropy

The entropy h has several equivalent definitions using either

- Finite state compressibility, or
- Finite state predictors/martingales.

Outline

Rauzy dimensions

Measure-theoretic entropy

Results

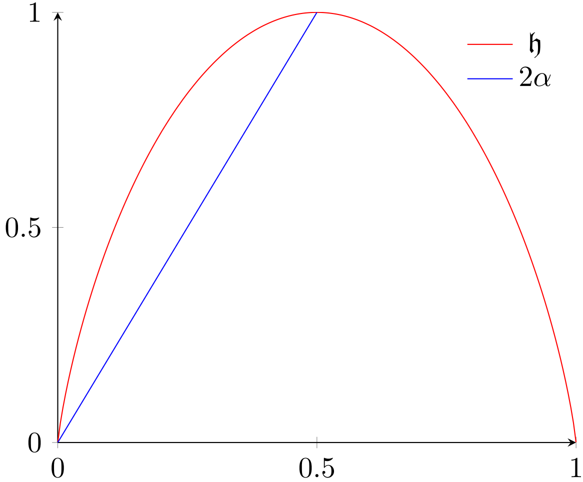
$\beta(x)$ vs $\gamma(x)$

Function $\mathfrak{h}$

Let $\mathfrak{h} : [0,1] \to [0,1]$ be the classical entropy function

$$\mathfrak{h}(\alpha) := -\alpha \log_2 \alpha - (1-\alpha) \log_2 (1-\alpha)$$

whose graph is



Results

Here, we suppose $\#A = 2$.

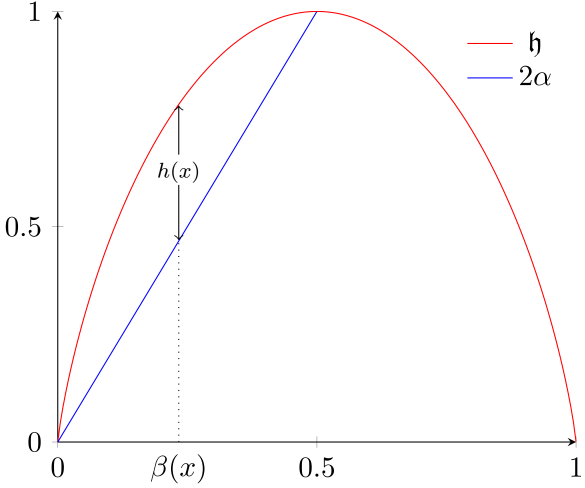
Theorem

- *For every* $x \in A^{\mathbb{N}}$,

$$\begin{aligned} 2\underline{\gamma}(x) &\leqslant \underline{h}(x) \leqslant \mathfrak{h}(\underline{\gamma}(x)), \\ 2\overline{\gamma}(x) &\leqslant \overline{h}(x) \leqslant \mathfrak{h}(\overline{\gamma}(x)), \\ 2\underline{\beta}(x) &\leqslant \underline{h}(x) \leqslant \mathfrak{h}(\underline{\beta}(x)), \\ 2\overline{\beta}(x) &\leqslant \overline{h}(x) \leqslant \mathfrak{h}(\overline{\beta}(x)). \end{aligned}$$

- *These inequalities are sharp.*

Visualizing the result



Results

Reformulation of one implication of second Rauzy's theorem:

Theorem (Rauzy, 1976)

If $\underline{h}(x) = 1$ *and* $\overline{h}(y) = 0$*, then* $\underline{h}(x+y) = 1$*.*

Theorem

For every $x, y \in A^{\mathbb{N}}$,

$$\underline{h}(x) - \overline{h}(y) \leqslant \underline{h}(x+y) \leqslant \underline{h}(x) + \overline{h}(y),$$
$$\overline{h}(x) - \overline{h}(y) \leqslant \overline{h}(x+y) \leqslant \overline{h}(x) + \overline{h}(y).$$

Outline

Rauzy dimensions

Measure-theoretic entropy

Results

$\beta(x)$ vs $\gamma(x)$

From the previous theorem

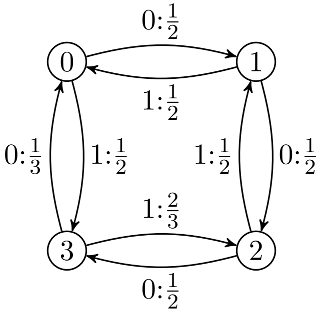
For every $x \in A^{\mathbb{N}}$,

$$\begin{aligned}
\underline{h}(x) = 0 &\iff \underline{\beta}(x) = 0 \iff \underline{\gamma}(x) = 0 \\
\overline{h}(x) = 0 &\iff \overline{\beta}(x) = 0 \iff \overline{\gamma}(x) = 0 \\
\underline{h}(x) = 1 &\iff \underline{\beta}(x) = \tfrac{1}{2} \iff \underline{\gamma}(x) = \tfrac{1}{2} \\
\overline{h}(x) = 1 &\iff \overline{\beta}(x) = \tfrac{1}{2} \iff \overline{\gamma}(x) = \tfrac{1}{2}
\end{aligned}$$

Quoting Rauzy, are the function β and γ similar ?

Counter-example

Let x be a *generic* sequence for the following Markov chain.
$\underline{\beta}(x) = \overline{\beta}(x) < \underline{\gamma}(x) = \overline{\gamma}(x) = \frac{11}{24}$.



A lower bound and a conjecture

Proposition

Let x be a "generic" sequence of a Markov chain with stationary distribution π. Let $\theta_{i,b}$ be the probability of having symbol b after state i. Then

$$\gamma(x) \geqslant \sum_{i \in Q} \pi_i \min(\theta_{i,0}, \theta_{i,1}).$$

We conjecture that

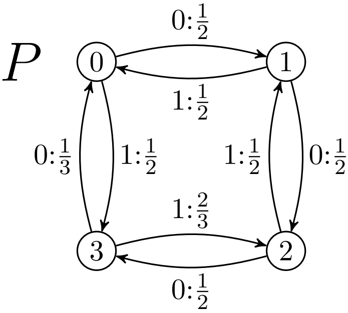
$$\gamma(x) = \sum_{i \in Q} \pi_i \min(\theta_{i,0}, \theta_{i,1}).$$

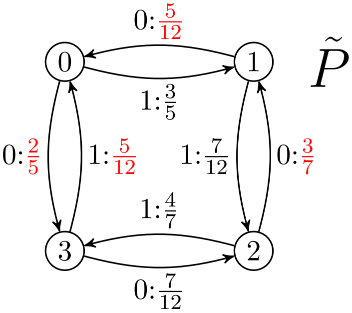
The stationary distribution π of the previous Markov chain is the vector $\pi = [\frac{5}{24}, \frac{1}{4}, \frac{7}{24}, \frac{1}{4}]$ and

$$\frac{11}{24} = \frac{5}{24} \cdot \frac{1}{2} + \frac{1}{4} \cdot \frac{1}{2} + \frac{7}{24} \cdot \frac{1}{2} + \frac{1}{4} \cdot \frac{1}{3}$$

Reversing time

Reversing time in a Markov chain gives another Markov chain whose matrix $\tilde{P}$ is given by $\pi_i \tilde{P}_{i,j} = \pi_j P_{j,i}$. where π is the stationnary distribution of P.





The stationnary distribution is the same: $\pi = [\frac{5}{24}, \frac{1}{4}, \frac{7}{24}, \frac{1}{4}]$.

$$\frac{5}{12} = \frac{5}{24} \cdot \frac{2}{5} + \frac{1}{4} \cdot \frac{5}{12} + \frac{7}{24} \cdot \frac{3}{7} + \frac{1}{4} \cdot \frac{5}{12}$$

Computing $\beta_\ell(x)$ and $\gamma_\ell(x)$

We suppose here that $A = \{0, 1\}$.

Lemma

Let $\ell \geqslant 0$ and let $x \in A^{\mathbb{N}}$ such that the frequency of u in x is equal to α_u for each $u \in A^{\ell+1}$. Then

$$\beta_\ell(x) = \sum_{w \in A^\ell} \min(\alpha_{0w}, \alpha_{1w}) \quad \textit{and} \quad \gamma_\ell(x) = \sum_{w \in A^\ell} \min(\alpha_{w0}, \alpha_{w1}).$$

ℓ	$\beta_\ell(x)$	$\gamma_\ell(x)$
1	$\frac{11}{24}$	$\frac{11}{24}$
2	$\frac{11}{24}$	$\frac{11}{24}$
3	$\frac{11}{24}$	$\frac{11}{24}$
4	$\frac{11}{24}$	$\frac{11}{24}$
5	$\frac{11}{24}$	$\frac{11}{24}$
6	$\frac{9503}{20736}$	$\frac{11}{24}$