

Finite Automata and Randomness

Ludwig Staiger

Martin-Luther-Universität Halle-Wittenberg



20th International Conference on Descriptive Complexity of

Formal Systems

Saint Mary's University, Halifax, Nova Scotia, Canada

July 25-27

Outline

1 Notation and preliminaries

Notation

Algorithmic randomness

2 Randomness by martingales

Gambling strategies for automata

Finite-state dimension

3 Incompressibility

Sequential decomposition

4 Automata and measure

Automata on ω -words

Subword complexity

5 Other concepts

Finite-state genericity

Unpredictability

Notation: Strings and Languages

Finite Alphabet $X = \{0, \dots, r-1\}$, cardinality $|X| = r$

Finite strings (words) $w = x_1 \dots x_n \in X^*$, $x_i \in X$

Length $|w| = n$

Languages $W \subseteq X^*$

Infinite strings (ω -words) $\xi = x_1 \dots x_n \dots \in X^\omega$

Prefixes of infinite strings $\xi[0..n] \in X^*$, $|\xi[0..n]| = n$

ω -Languages $F \subseteq X^\omega$

Metric: $\rho(\eta, \xi) := \inf\{r^{-|w|} : w \in \text{pref}(\eta) \cap \text{pref}(\xi)\}$

Balls: $w \cdot X^\omega = \{\eta : w \in \text{pref}(\eta)\} = \{\eta : w \sqsubset \eta\}$

Diameter: $\text{diam } w \cdot X^\omega = r^{-|w|}$

$\text{diam } F = \inf\{r^{-|w|} : F \subseteq w \cdot X^\omega\}$

Open sets: $W \cdot X^\omega = \bigcup_{w \in W} w \cdot X^\omega$

Closure: (Smallest closed set containing F)

$\text{cl}_\rho(F) = \{\xi : \text{pref}(\xi) \subseteq \text{pref}(F)\}$

Fact

$F \subseteq X^\omega$ is closed if and only if $\text{pref}(F) \subseteq \text{pref}(F)$ implies $\xi \in F$.

Algorithmic randomness

measure-theoretic paradigm

An ω -word is random if and only if it is not contained in a constructive null-set.

unpredictability paradigm

An ω -word is random if and only if no constructive predicting strategy can win against it.

incompressibility (complexity-theoretic) paradigm

An ω -word is random if and only if one cannot constructively compress infinitely many of its prefixes.

Measure

Measure on base sets: $\mu(w \cdot X^\omega) := r^{-|w|}$

Constructive null-sets: Unions of ω -languages of the form $\bigcup_{n \in \mathbb{N}} V_n \cdot X^\omega$,

where $V \subseteq \{(v, n) : v \in X^* \wedge n \in \mathbb{N}\}$ is constructive,

$V_n := \{v : (v, n) \in V\}$, and

$\mu(V_n \cdot X^\omega) \leq r^{-n}$.

Definition (Randomness)

$\xi \in X^\omega$ is random if and only if no constructive null-set contains ξ .

Predicting strategy: Gambling

Our model:

- Playing against an ω -word $\xi \in X^\omega$.
- Gambling strategy $\Gamma : X^* \times X \rightarrow [0, 1]$ (bet on outcome $x \in X$)
 $\sum_{x \in X} \Gamma(w, x) \leq 1$ for $w \in X^*$
- yields a (super-)martingale $\mathcal{V}_\Gamma : X^* \rightarrow \mathbb{R}_+$
- $\mathcal{V}_\Gamma(\xi[0..n])$ is the capital after the n th round, that is,

$$\mathcal{V}_\Gamma(\xi[0..n]) = r \cdot \Gamma(\xi[0..n], x) \cdot \mathcal{V}_\Gamma(\xi[0..n-1]), \text{ for } \xi(n) = x$$



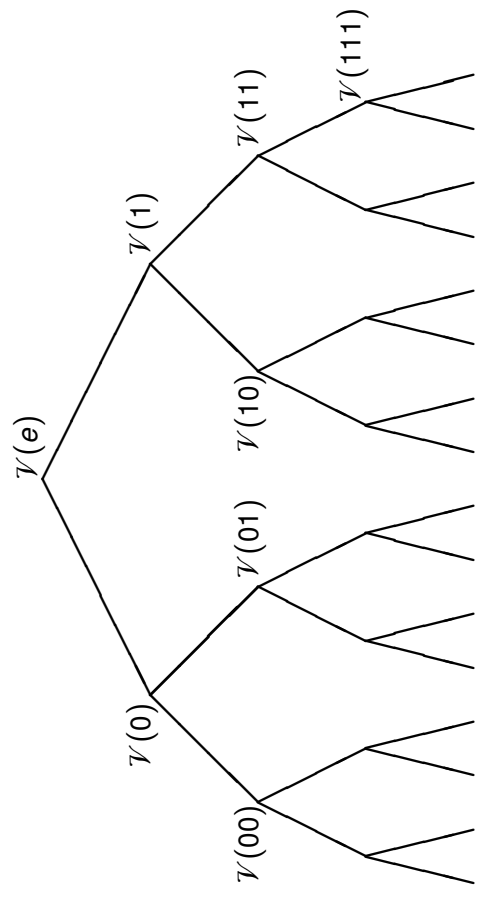
Fact (super-martingale property)

$$\mathcal{V}_\Gamma(w) \geq \frac{1}{r} \cdot \sum_{x \in X} \mathcal{V}_\Gamma(wx)$$

Definition (Randomness)

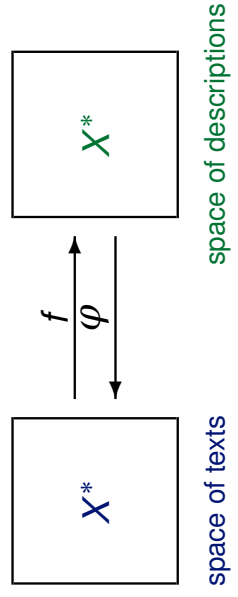
$\xi \in X^\omega$ is random if and only if no constructive gambling strategy Γ can win against ξ , that is, $\limsup_{n \rightarrow \infty} \mathcal{V}_\Gamma(\xi[0..n]) < \infty$.

Gambling strategies: Martingale \mathcal{V}



Compression: The principle of loss-less compression

text $w \in X^*$ description (or program) $\pi \in X^*$



f is injective and $\varphi(f(w)) = w$ for all $w \in X^*$

→ Complexity of w w.r.t. $\varphi: C_\varphi(w) := \inf\{|\pi| : \varphi(\pi) = w\}$

Definition (Randomness = Incompressibility)

$\xi \in X^\omega$ is *random* if and only if all constructive decomposition functions φ satisfy $\exists c \forall n (C_\varphi(\xi[0..n])) \geq n - c$, that is, prefixes of ξ cannot be compressed.

Gambling finite automaton

Definition (Betting automaton)

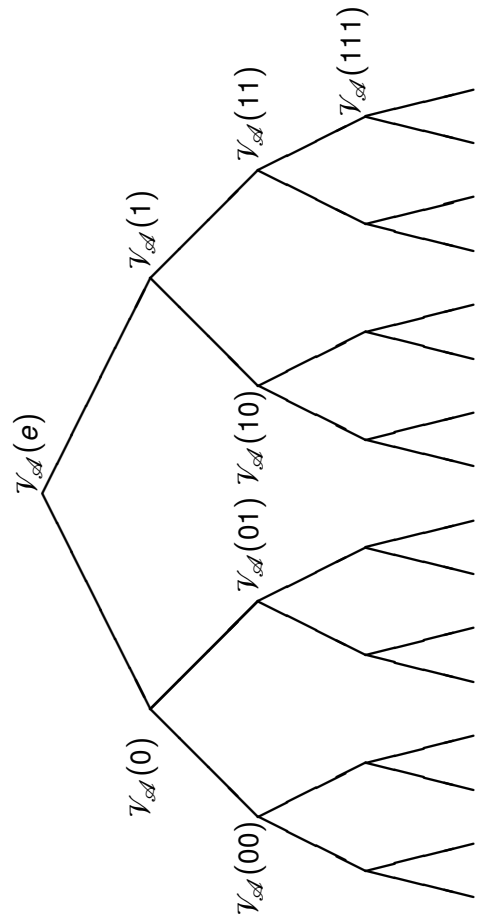
$\mathcal{A} = [X, Q, \mathbb{R}_{\geq 0}, q_0, \delta, \nu]$ is a finite-state betting automaton : \iff

- 1 Q is a finite set (of states), $q_0 \in Q$,
- 2 $\delta : Q \times X \rightarrow Q$,
- 3 $\nu : Q \times X \rightarrow \mathbb{R}_{\geq 0}$ and $\sum_{x \in X} \nu(q, x) \leq 1$, for all $q \in Q$.

Definition (Capital function of \mathcal{A})

$$\begin{aligned} \mathcal{V}_{\mathcal{A}}(e) &:= 1, \text{ and} \\ \mathcal{V}_{\mathcal{A}}(wx) &:= r \cdot \nu(\delta(q_0, w), x) \cdot \mathcal{V}_{\mathcal{A}}(w) \end{aligned}$$

Again: Gambling strategies: Martingale $\mathcal{V} = \mathcal{V}_{\mathcal{A}}(X = \{0, 1\})$



Definition

An ω -word $\xi \in X^\omega$ is **BOREL normal** iff every subword (infix) $w \in X^*$ appears with the same frequency.

$$\forall w \left(\lim_{n \rightarrow \infty} \frac{|\{j : i \leq n \wedge \xi[0..j] \in X^* \cdot w\}|}{n} \right) = r^{-|w|}$$

The theorem of SCHNORR and STIMM

Theorem (SCHNORR/STIMM '72)

If $\xi \in X^\omega$ is BOREL normal then for every finite automaton \mathcal{A} it holds

- 1 $\forall^\infty n (n \in \mathbb{N} \rightarrow \overline{V}_{\mathcal{A}}(\xi[0..n]) = \overline{V}_{\mathcal{A}}(\xi[0..n+1]))$, or
- 2 $\exists \rho (0 \leq \rho < 1 \wedge \forall^\infty n (n \in \mathbb{N} \rightarrow \overline{V}_{\mathcal{A}}(\xi[0..n]) \leq \rho^n)$.



If $\xi \in X^\omega$ is **not** BOREL normal then there are a finite automaton \mathcal{A} and $\gamma > 0$ such that

- 3 $\exists^\infty n (n \in \mathbb{N} \wedge \overline{V}_{\mathcal{A}}(\xi[0..n]) \geq r^{\gamma \cdot n})$.

Partial Randomness: Finite-state dimension [DAI ET AL.'04]

Finite-state dimension tries to measure, for $\xi \in X^\omega$, the largest exponent γ_0 with

$$\overline{V}_{\mathcal{A}}(\xi[0..n]) \approx r^{\gamma_0 \cdot n + o(n)},$$

for some finite automaton \mathcal{A} 'best fitted' to ξ .



More precisely, $\dim_{FS}(\xi) = 1 - \gamma_0 : \iff$

$$\exists \mathcal{A} \left(\limsup_{n \rightarrow \infty} \frac{\overline{V}_{\mathcal{A}}(\xi[0..n])}{r^{\gamma \cdot n}} > 0 \right) \text{ for } \gamma < \gamma_0, \text{ and}$$

$$\forall \mathcal{A} \left(\lim_{n \rightarrow \infty} \frac{\overline{V}_{\mathcal{A}}(\xi[0..n])}{r^{\gamma \cdot n}} = 0 \right) \text{ for } \gamma > \gamma_0.$$



Observe

The higher the dimension $\dim_{FS}(\xi)$ the 'more random' the ω -word.

Corollary

$\dim_{FS}(\xi) = 1$ if and only if ξ is BOREL normal

Uniform extension to $F \subseteq X^\omega$ [DAI ET AL.'04]

Definition (Finite-state Dimension (DAI ET AL.'04))

$$\alpha_{\mathcal{A}}(F) := \inf \left\{ \alpha : \forall \xi (\xi \in F \rightarrow \limsup_{n \rightarrow \infty} \frac{\overline{V}_{\mathcal{A}}(\xi[0..n])}{r^{(1-\alpha) \cdot n}} > 0) \right\}$$

$\dim_{FS}(F) := \sup \{ \alpha_{\mathcal{A}}(F) : \mathcal{A} \text{ is a finite automaton} \}$

Observe

$1 - \alpha$ corresponds to the exponent γ .



Proposition

\dim_{FS} is monotone and stable: $\dim_{FS}(F \cup F') = \max\{\dim_{FS} F, \dim_{FS} F'\}$

Example

\dim_{FS} is not countably stable:

$$\dim_{FS}\{w \cdot v^\omega\} = 0 \text{ and } \dim_{FS}\{w \cdot v^\omega : w, v \in X^*\} = 1.$$

Finite-state dimension: Frequency

Let $h(\alpha) := -\alpha \cdot \log_2 \alpha - (1 - \alpha) \cdot \log_2 (1 - \alpha)$ be the binary SHANNON entropy and let

$$\text{FREQ}(\alpha) := \left\{ \xi : \xi \in \{0, 1\}^\omega \wedge \lim_{n \rightarrow \infty} \frac{|\xi[0..n]|_1}{n} = \alpha \right\}$$

Theorem (DAI ET AL.'04)

Let $\alpha \in [0, 1]$ be rational. Then the following hold.

- 1 There is an ω -word $\xi \in X^\omega$ having $\dim_{FS}(\xi) = \alpha$, and
- 2 $\dim_{FS}(\text{FREQ}(\alpha)) = h(\alpha)$.

Definition

A predicting automaton $\mathcal{A} = [X, Q, q_0, \delta, \lambda]$ weakly predicts $\xi \in X^\omega$ if and only if

- 1 $\lambda(\delta(q_0, \xi[0..n-1])) \in X$ for infinitely many n , and
- 2 if $\lambda(\delta(q_0, \xi[0..n-1])) \in X$ then $\lambda(\delta(q_0, \xi[0..n-1])) \neq \xi(n)$.

**Theorem**

An ω -word ξ is weakly predictable by some automaton $\mathcal{A} = [X, Q, q_0, \delta, \lambda]$ if and only if it is non-disjunctive.

