Computing Absolutely Normal Numbers in Nearly Linear Time

Jack H. Lutz and Elvira Mayordomo

Iowa State University, Universidad de Zaragoza

Continuity, Computability, Constructivity 2017
Loria, Nancy

- A base is an integer $b \geq 2$.

- A base is an integer $b \geq 2$.
- A real number α is normal in base b if any two non-empty strings of equal length appear equally often (asymptotically) in the base-b expansion of the fractional part $\{\alpha\} = \alpha \mod 1$ of α.

- A base is an integer $b \geq 2$.
- A real number α is normal in base b if any two non-empty strings of equal length appear equally often (asymptotically) in the base-b expansion of the fractional part $\{\alpha\} = \alpha \mod 1$ of α.
- A real number α is absolutely normal if it is normal in every base.

- Theorem (Borel, 1909). Almost every real number is absolutely normal.

- Theorem (Borel, 1909). Almost every real number is absolutely normal.
- Theorem (Turing, late 1930s). There is an algorithm that computes an absolutely normal number.

- Theorem (Borel, 1909). Almost every real number is absolutely normal.
- Theorem (Turing, late 1930s). There is an algorithm that computes an absolutely normal number.
- Theorem (Becher, Heiber, and Slaman, 2013). There is an algorithm that computes an absolutely normal number α in polynomial time. (It computes the successive bits of the binary expansion of α, with the n^{th} bit appearing in time polynomial in n.)

- Our result today: An algorithm that computes an absolutely normal number α in nearly linear time.

- Our result today: An algorithm that computes an absolutely normal number α in nearly linear time.
- It computes the successive bits of the binary expansion of α, with the n^{th} bit appearing within $n(\log n)^{O(1)}$ steps.

- Our result today: An algorithm that computes an absolutely normal number α in nearly linear time.
- It computes the successive bits of the binary expansion of α, with the n^{th} bit appearing within $n(\log n)^{O(1)}$ steps.
- This was called **nearly linear time** by Gurevich and Shelah (1989), who proved that nearly linear time – unlike linear time! – is model robust.

Outline

1. Martingales (and why)
2. Lempel-Ziv martingales (and why)
3. Savings Accounts
4. Base Change
5. Absolutely Normal Numbers
6. Open Problem (if time)

Martingales

- $\Sigma_b = \{0, \ldots, b-1\}$ the base b alphabet
- Σ_b^* are finite sequences, Σ_b^∞ infinite. sequences
- $x \upharpoonright n$ is the length-n prefix of x.
- A **martingale** is a function $d : \Sigma_b^* \to [0..\infty)$ with the fairness property that, for every finite sequence w,
$$d(w) = \frac{\sum_{i \in \Sigma_b} d(wi)}{b}.$$
- A martingale d succeeds on an infinite sequence $x \in \Sigma_b^\infty$ if
$$limsup_n d(x \upharpoonright n) = \infty$$
(x can be predicted by d).
- Lebesgue measure can be defined in terms of martingales (a set has measure 0 if there is a martingale succeeding on every element of the set).
- And you **have** to use martingales to have a useful measure on small complexity classes ...
- ... because they aggregate a lot of information!

Martingales

But how fast do they succeed?
Let $g : \Sigma_b^* \to [0, \infty)$ (may or may not be a martingale) and $S \in \Sigma_b^\infty$.

- g **succeeds** on S ($S \in S^\infty[g]$) if $\limsup_{n\to\infty} g(S \restriction n) = \infty$.
- g $f(n)$-**succeeds** on S ($S \in S^{f(n)}[g]$) if $\limsup_{n\to\infty} \frac{\log g(S\restriction n)}{\log f(n)} > 1$.
- g **succeeds exponentially** on S ($S \in S^{\exp}[g]$) if $\exists \epsilon > 0$ $S \in S^{2^{\epsilon n}}[g]$.

Lempel-Ziv martingales

Schnorr and Stimm (1972) implicitly defined **finite-state martingales** and proved that every sequence $S \in \Sigma_b^\infty$ obeys the following dichotomy:

1. If S is b-normal, then no finite-state base-b martingale succeeds on S. (In fact, every finite-state base-b martingale decays exponentially on S.)
2. If S is not b-normal, then some finite-state base-b martingale succeeds exponentially on S.

Lempel-Ziv martingales

Feder (1991) implicitly defined the **base-b Lempel-Ziv martingale** $d_{\mathrm{LZ}(b)}$ and proved that it is at least as successful **on every sequence** as every finite-state martingale.

Lempel-Ziv martingales

Feder (1991) implicitly defined the **base-b Lempel-Ziv martingale** $d_{\mathrm{LZ}(b)}$ and proved that it is at least as successful **on every sequence** as every finite-state martingale.
$\therefore$ if $S \in \Sigma_b^\infty$ is not normal, then $S \in S^{\exp}[d_{\mathrm{LZ}(b)}]$.

Lempel-Ziv martingales

Feder (1991) implicitly defined the **base-b Lempel-Ziv martingale** $d_{\mathrm{LZ}(b)}$ and proved that it is at least as successful **on every sequence** as every finite-state martingale.
$\therefore$ if $S \in \Sigma_b^\infty$ is not normal, then $S \in S^{\exp}[d_{\mathrm{LZ}(b)}]$.
$\therefore$ $x \in (0,1)$ is absolutely normal if none of the martingales $d_{\mathrm{LZ}(b)}$ succeed exponentially on the base-b expansion of x.

Lempel-Ziv martingales

Feder (1991) implicitly defined the **base-b Lempel-Ziv martingale** $d_{\mathrm{LZ}(b)}$ and proved that it is at least as successful **on every sequence** as every finite-state martingale.
$\therefore$ if $S \in \Sigma_b^\infty$ is not normal, then $S \in S^{\exp}[d_{\mathrm{LZ}(b)}]$.
$\therefore$ $x \in (0,1)$ is absolutely normal if none of the martingales $d_{\mathrm{LZ}(b)}$ succeed exponentially on the base-b expansion of x.
Moreover, $d_{\mathrm{LZ}(b)}$ is fast and has a beautiful theory.

Lempel-Ziv martingales

How $d_{\mathrm{LZ}(b)}$ works:
Parse $w \in \Sigma_b^*$ into distinct **phrases**, using a growing tree whose leaves are all of the previous phrases.
At each step, bet on the next digit in proportion to the number of leaves below each of the b options.

Savings Accounts

- The value of Lempel-Ziv martingale $d_{\mathrm{LZ}(b)}$ on a certain infinite string S can fluctuate a lot.
- This makes base change more complicated (and time consuming).
- We use the notion of "savings account" here. That is, we construct an alternative martingale that **keeps money aside for the bad times to come**
- This is a (refinement of a) technique known since the 1970s.

Savings Accounts

- The value of Lempel-Ziv martingale $d_{\mathrm{LZ}(b)}$ on a certain infinite string S can fluctuate a lot.
- This makes base change more complicated (and time consuming).
- We use the notion of "savings account" here. That is, we construct an alternative martingale that **keeps money aside for the bad times to come**
- This is a (refinement of a) technique known since the 1970s.

Definition

A **savings account** *for a martingale* $d : \Sigma_b^* \to [0, \infty)$ *is a nondecreasing function* $g : \Sigma_b^* \to [0, \infty)$ *such that* $d(w) \geq g(w)$ *for every* w*.*

Savings Accounts

- We construct a new martingale d'_b with a savings account g'_b that is a conservative version of $d_{\mathrm{LZ}(b)}$.
- g'_b succeeds at least on non-b-normal sequences.

Savings Accounts

- We construct a new martingale d'_b with a savings account g'_b that is a conservative version of $d_{\mathrm{LZ}(b)}$.
- g'_b succeeds at least on non-b-normal sequences.
- Both d'_b and g'_b can be computed in nearly linear time.
- If $S \notin S^\infty[g'_b]$ then S is b-normal.

Base Change

- We want an absolutely normal real number α, that is, the base b representation $seq_b(\alpha)$ is not in $S^\infty[d'_b]$.
- For this we convert d'_b into a base-2 martingale $d_b^{(2)}$ succeeding on the **base-2 representations of the reals with base-b representation in $S^\infty[d'_b]$.**
- Again, $d_b^{(2)}$ succeeds on $seq_2(real(S^\infty[d'_b]))$.

Base Change

- We want an absolutely normal real number α, that is, the base b representation $seq_b(\alpha)$ is not in $S^\infty[d'_b]$.
- For this we convert d'_b into a base-2 martingale $d_b^{(2)}$ succeeding on the **base-2 representations of the reals with base-b representation in $S^\infty[d'_b]$.**
- Again, $d_b^{(2)}$ succeeds on $seq_2(real(S^\infty[d'_b]))$.
- We use Carathéodory construction to define measures.
- Computing in nearly linear time is also delicate.

Base Change

- We want an absolutely normal real number α, that is, the base b representation $seq_b(\alpha)$ is not in $S^\infty[d'_b]$.
- For this we convert d'_b into a base-2 martingale $d_b^{(2)}$ succeeding on the **base-2 representations of the reals with base-b representation in $S^\infty[d'_b]$.**
- Again, $d_b^{(2)}$ succeeds on $seq_2(real(S^\infty[d'_b]))$.
- We use Carathéodory construction to define measures.
- Computing in nearly linear time is also delicate.
- In fact our computation $\widehat{d_b^{(2)}}$ approximates $d_b^{(2)}$ slowly

$$|\widehat{d_b^{(2)}}(y) - d_b^{(2)}(y)| \leq \frac{1}{|y|^3}$$

.

Absolutely Normal Numbers

- From previous steps we have a family of martingales $(d_b^{(2)})_b$ so that $d_b^{(2)}$ **succeeds on base-2 representations of non-*b*-normal sequences**.
- For each b we have a nearly linear time computation $\widehat{d_b^{(2)}}$.

Absolutely Normal Numbers

- From previous steps we have a family of martingales $(d_b^{(2)})_b$ so that $d_b^{(2)}$ **succeeds on base-2 representations of non-b-normal sequences**.
- For each b we have a nearly linear time computation $\widehat{d_b^{(2)}}$.
- We want to construct $S \notin S^\infty[d_b^{(2)}]$ for every b.
- Nearly linear time makes it painful to construct a martingale d for the union of $S^\infty[d_b^{(2)}]$.

Absolutely Normal Numbers

- From previous steps we have a family of martingales $(d_b^{(2)})_b$ so that $d_b^{(2)}$ **succeeds on base-2 representations of non-*b*-normal sequences**.
- For each b we have a nearly linear time computation $\widehat{d_b^{(2)}}$.
- We want to construct $S \notin S^\infty[d_b^{(2)}]$ for every b.
- Nearly linear time makes it painful to construct a martingale d for the union of $S^\infty[d_b^{(2)}]$.
- Then we diagonalize over d to construct S.

Time bounds ...

- All the steps were performed in **online** nearly linear time on a common **time bound independent of base** b.
- Many technical details were simplified in this presentation ... please read our paper.

Thank you!