Departamento de Computación, Facultad de Ciencias Exactas y Naturales, UBA

Azar y Autómatas

Clase 5: Algoritmo eficiente para computar un número absolutamente normal

A fast construction of absolutely normal numbers

Theorem 1 (Becher, Heiber and Slaman in 2013)

There is an algorithm that computes an absolutely normal number x in nearly quadratic time complexity: the first n digits in the expansion of x in base 2 are obtained by performing $O\left(n^2 \sqrt[4]{\log n}\right)$ mathematical operations.

Recall the definition of absolute normality

For the construction, the most convenient definition of absolute normality is:

Recall the definition of absolute normality

For the construction, the most convenient definition of absolute normality is:

A real number x is absolutely normal if it is simply normal to
all integer bases b greater than or equal to 2.

Let x be a real in the unit interval, and let x_b be its expansion in base b.

Recall the definition of absolute normality

For the construction, the most convenient definition of absolute normality is:

A real number x is absolutely normal if it is simply normal to all integer bases b greater than or equal to 2.

Let x be a real in the unit interval, and let x_b be its expansion in base b.

We define

$$\Delta_N(x_b) = \max_{d \in \{0,\ldots,b-1\}} \left| \frac{|x_b[1\ldots N]|_d}{N} - \frac{1}{b} \right|$$

Recall the definition of absolute normality

For the construction, the most convenient definition of absolute normality is:

A real number x is absolutely normal if it is simply normal to all integer bases b greater than or equal to 2.

Let x be a real in the unit interval, and let x_b be its expansion in base b.

We define

$$\Delta_N(x_b) = \max_{d \in \{0,\ldots,b-1\}} \left| \frac{|x_b[1\ldots N]|_d}{N} - \frac{1}{b} \right|$$

Then, x is simply normal to base b if

$$\lim_{N\to\infty} \Delta_N(x_b) = 0$$

Recall the definition of absolute normality

For the construction, the most convenient definition of absolute normality is:

A real number x is absolutely normal if it is simply normal to all integer bases b greater than or equal to 2.

Let x be a real in the unit interval, and let x_b be its expansion in base b.

We define

$$\Delta_N(x_b) = \max_{d \in \{0,\ldots,b-1\}} \left| \frac{|x_b[1\ldots N]|_d}{N} - \frac{1}{b} \right|$$

Then, x is simply normal to base b if

$$\lim_{N\to\infty} \Delta_N(x_b) = 0$$

If w is a block of digits in base b we just write $\Delta(w)$ instead of $\Delta_{|w|}(w)$.

The following two lemmas are not hard to prove.

Lemma 2 (Lemma 3.1 BHS 2013)

Let u and v be blocks and let a positive real ε.

1. *If $\Delta(u) < \varepsilon$ and $\Delta(v) < \varepsilon$ then $\Delta(uv) < \varepsilon$.*
2. *If $\Delta(u) < \varepsilon$, $v = a_1 \ldots a_{|v|}$ and $|v|/|u| < \varepsilon$ then $\Delta(vu) < 2\varepsilon$, and for every ℓ such that $1 \le \ell \le |v|$, $\Delta(ua_1 a_2 \ldots a_\ell) < 2\varepsilon$.*

Lemma 3 (Lemma 3.4 BHS2013)

For any interval I and any base b, there is a b-ary subinterval J such that $\mu J \ge \mu I/(2b)$.

The next two definitions are the core of the construction.

Definition 4

A t-sequence $\overrightarrow{\sigma}$ is a sequence of intervals $(\sigma_2, \ldots, \sigma_t)$ such that

for each base $b = 2, \ldots, t$, σ_b is b-ary,

for each base $b = 3, \ldots, t$, $\sigma_b \subset \sigma_{b-1}$ and $\mu\sigma_b \geq \mu\sigma_{b-1}/(2b)$.

Observe that the definition implies $\mu\sigma_t \geq (\mu\sigma_2)/(2^t t!)$.

Definition 5

A t-sequence $\overrightarrow{\tau} = (\tau_2, \ldots, \tau_t)$ *refines* a t'-sequence $\overrightarrow{\sigma} = (\sigma_2, \ldots, \sigma_{t'})$ if $t' \leq t$ and $\tau_b \subset \sigma_b$ for each $b = 2, \ldots, t'$. A refinement has *discrepancy less than* ε if for each $b = 2, ..t'$ there are words u, v such that $\sigma_b = I_u$, $\tau_b = I_{uv}$ and $\Delta(v) < \varepsilon$.

We say that an interval is *b*-ary of *order n* if it is of the form

$$\left(\frac{a}{b^n}, \frac{a+1}{b^n}\right)$$

for some integer a such that $0 \leq a < b^n$. If σ_b and τ_b are *b*-ary intervals, and $\tau_b \subseteq \sigma_b$ we say that the *relative order* of τ_b with respect to σ_b is the *order* of τ_b minus the *order* of σ_b.

Lemma 6

Let t be an integer greater than or equal to 2, let t' be equal to t or to $t+1$, and let ε be a positive real less than $1/t$. Then, any t-sequence $\overrightarrow{\sigma} = (\sigma_2, \ldots, \sigma_t)$ admits a refinement $\overrightarrow{\tau} = (\tau_2, \ldots, \tau_{t'})$ with discrepancy less than ε. The relative order of τ_2 can be any integer greater than or equal to $\max(6/\varepsilon, 24(\log_2 t)(\log(t!))/\varepsilon^2)$.

Proof of Lemma 6

First assume $t' = t$. We must pick a t-sequence $(\tau_2, \ldots, \tau_t)$ that refines $(\sigma_2, \ldots, \sigma_t)$ in a zone of low discrepancy. This is possible because the measure of the zones of large discrepancy decreases at an exponential rate in the order of the interval. To prove the lemma we need to determine the relative order N of τ_2 such that the measure of the union of the bad zones inside σ_2 for the bases $b = 2, \ldots t$ is strictly less than the measure of the set all the possible t-ary subintervals τ_t of σ_2.

Let L be the largest binary subinterval in σ_t.
Partition of L in 2^N binary intervals τ_2 of equal length.
For each τ_2 apply iteratively Lemma 3 to define $\tau_3, \ldots, \tau_{t_n}$.

Thus, we have defined 2^N many t_n-sequences $(\tau_2, \ldots \tau_t)$. Let S be the union of the set of all possible intervals τ_t over these 2^N many t_n-sequences. Hence, by the definition of t-sequence,

$$\mu S \geq \mu L/(2^t t!).$$

By Lemma 3,

$$\mu L \geq \mu\sigma_t/4.$$

And by the definition of t-sequence again,

$$\mu\sigma_t \geq \mu\sigma_2/(2^t t!).$$

Combining inequalities we obtain,

$$\mu S \geq \mu\sigma_2/(2^t t!\ 4\ 2^t t!)$$

Now consider the bad zones inside σ_2. For each $b = 2, \dots t$, for a length N and a real value ε consider the the following set of intervals of relative order $\lceil N/\log_2 b \rceil$ with respect to σ_2,

$$B_{b,\lceil N/\log_2 b\rceil,\varepsilon} = \bigcup_{\substack{u \in \{0,\dots,b-1\}^{\lceil N/\log_2 b\rceil} \\ \Delta(u) \geq \varepsilon}} I_u.$$

Thus, the actual measure of the bad zones is

$$\mu\sigma_2 \ \mu\Big(\bigcup_{b=2,..,t} \mu B_{b,\lceil N/\log_2 b\rceil,\varepsilon} \Big)$$

Then, N must be such that

$$\mu\sigma_2 \ \mu\Big(\bigcup_{b=2,..,t} B_{b,\lceil N/\log_2 b\rceil,\varepsilon} \Big) < \mu S.$$

Using Lemma **??** on the left and the inequality above for μS on the right it suffices that N be greater than $6/\varepsilon$ and also N be such that

$$2t^2 \cdot e^{-\varepsilon^2 (N/3\log_2 t)} < \frac{1}{2^t t!} \frac{1}{4} \frac{1}{2^t t!}.$$

We can take N greater than or equal to $\max(6/\varepsilon, 24(\log_2 t)(\log(t!))/\varepsilon^2)$.

The case $t' = t + 1$ follows easily by taking first a t-sequence $\overrightarrow{\tau}$ refining $\overrightarrow{\sigma}$ with discrepancy less than ϵ. Definition 5 does not require any discrepancy considerations for τ_{t+1}. Take τ_{t+1} the largest $(t+1)$-ary subinterval of τ_t. By Lemma 3, $\mu\tau_{t+1} \geq (\mu\tau_t)/(2(t+1))$. This completes the proof of the lemma.

El algoritmo BHS 2013

The algorithm considers three functions of the step number n:

t_n is the maximum base to be considered at step n,
ε_n is the maximum discrepancy tolerated at step n, and
N_n is the number of digits in base 2 added at step n.

The algorithm constructs $\overrightarrow{\sigma}_0, \overrightarrow{\sigma}_1, \overrightarrow{\sigma}_2, \ldots$ such that $\overrightarrow{\sigma}_0 = (0, 1)$ and for each $n \geq 1$, $\overrightarrow{\sigma}_n$ is t_n-sequence that refines $\overrightarrow{\sigma}_{n-1}$ with discrepancy ε_n and such that the order of $\sigma_{n,2}$ is N_n plus the order of $\sigma_{n-1,2}$.

Definition 7

Define the following functions of n,

$$\begin{aligned} t_n &= \max(2, \lfloor \sqrt[4]{\log n} \rfloor), \\ \varepsilon_n &= 1/t_n, \\ N_n &= \lfloor \log n \rfloor + n_{start}, \end{aligned}$$

where n_{start} is the minimum integer such that that validates the condition in Lemma 6. Thus we require that for every positive n,

$$\begin{aligned} &\lfloor \log n \rfloor + n_{start} \geq 6/\varepsilon_n \qquad \text{and} \\ &\lfloor \log n \rfloor + n_{start} \geq 24(\log_2 t_n)(\log(t_n!))/\varepsilon_n^2. \end{aligned}$$

Algorithm BHS

Output: $y_1 y_2 y_3 \ldots$ the symbols in the base 2 expansion of an absolutely normal number.

Initial step, $n = 1$. $\overrightarrow{\sigma}_1 = (\sigma_2)$, with $\sigma_2 = (0, 1)$.

Recursive step, $n > 1$. Assume $\overrightarrow{\sigma}_{n-1} = (\sigma_2, \ldots, \sigma_{t_{n-1}})$.
Take $\overrightarrow{\sigma}_n = (\tau_2, \ldots, \tau_{t_n})$ the leftmost t_n-sequence that refines $\overrightarrow{\sigma}_{n-1}$ with discrepancy less than ε_n and such that if $\sigma_2 = I_u$ then $\tau_2 = I_{uv}$ with $|v| = N_n$.
Set $y_{M_n+1} \ldots y_{M_n+N_n} = v$, where $M_n = \sum_{j=1}^{n} N_n$.

Proof of Theorem 1

The existence of the sequence $\overrightarrow{\sigma}_1, \overrightarrow{\sigma}_2, \ldots$ is guaranteed by Lemma 6. We have to prove that the real number x defined by the intersection of all the intervals in the sequence is absolutely normal. We pick a base b and show that x is simply normal to base b. Let $\tilde{\varepsilon} > 0$. Choose n_0 so that $t_{n_0} \geq b$ and $\varepsilon_{n_0} \leq \tilde{\varepsilon}/4$. At each step n after n_0 the expansion of x in base b was constructed by appending blocks u_n such that $\Delta(u_n) < \varepsilon_{n_0}$. Thus, by Lemma 2 (item 1) for any $n > n_0$,

$$\Delta(u_{n_0} \ldots u_n) < \varepsilon_{n_0}.$$

Applying Lemma 2 (item 2a), we obtain n_1 such that for any $n > n_1$

$$\Delta(u_1 \ldots u_n) < 2\epsilon_{n_0}.$$

Let $N_n^{(b)}$ be the relative order of τ_b with respect to σ_b. By Lemma 3,

$$\frac{N_n}{\log_2 b} \leq N_n^{(b)} \leq \frac{N_n + 1}{\log_2 b} + 1.$$

Since $N_n = \lfloor \log n \rfloor + n_{start}$, N_n grows logarithmically and so does $N_n^{(b)}$ for each base b. Then, for n sufficiently large,

$$N_n^{(b)} \leq \frac{N_n + 1}{\log_2 b} + 1 \leq 2\epsilon_{n_0} \sum_{j=1}^{n-1} \frac{N_j}{\log_2 b} \leq 2\epsilon_{n_0} \sum_{j=1}^{n-1} N_j^{(b)}.$$

By Lemma 2 (item 2b) we conclude that for n sufficiently large, if $u_n = a_1 \dots a_{|u_n|}$ then for every ℓ such that $1 \leq \ell \leq |u_n|$,

$$\Delta_\ell(u_1 \dots u_{n-1} a_1 \dots a_\ell) < \ 4\epsilon_{n_0} < \tilde{\epsilon}.$$

So, x is simply normal to base b for every $b \geq 2$.

We now analyze the computational complexity of the algorithm. Lemma 6 ensures the existence of the wanted t-sequence at each step n. To effectively find it we proceed as follows. Divide the interval σ_2 into

$$2^{N_n}$$

equal binary intervals. In the worst case, for each of them, we need to check if it allocates a t_n-sequence $(\tau_2, \ldots, \tau_{t_n})$ that refines $(\sigma_2 \ldots, \sigma_{t_{n-1}})$ with discrepancy less than ε_n. Since we are just counting the number of mathematical operations ignoring the precision, at step n the algorithm performs

$$O\big(2^{N_n} t_n\big)$$

many mathematical operations. Since N_n is logarithmic in n and t_n is a rational power of $\log(n)$ we conclude that at step n the algorithm performs

$$O(n\sqrt[4]{\log n})$$

mathematical operations. Finally, in the first k steps the algorithm will output at lest k many digits of the binary expansion of the computed number having performed

$$O(k^2\sqrt[4]{\log k})$$

many mathematical operations. This completes the proof of Theorem 1.