Preservation of normality by selection

Olivier Carton

IRIF
Université de Paris & CNRS

Based on join works with V. Becher, P. Heiber and J. Vandehey
(Universidad de Buenos Aires & CONICET)
(University of Texas at Tyler)

2020 – Journées du GT SDA2

Outline

Outline

Normality

Selection

Extensions

Selectors

Shifts of finite type

Normal sequences

A normal sequence is a sequence such that all finite blocks of the same length occur in it with the same limiting frequency.

If $x \in A^{\mathbb{N}}$ and $w \in A^*$, the frequency of w in x is defined by

$$\mathrm{freq}(x, w) = \lim_{n \to \infty} \frac{|x[1..n]|_w}{n}.$$

where $|z|_w$ denotes the number of occurrences of w in z.

A sequence $x \in A^{\mathbb{N}}$ is normal if for each $w \in A^*$:

$$\mathrm{freq}(x, w) = \frac{1}{(\#A)^{|w|}}$$

where
- $\#A$ is the cardinality of the alphabet A
- $|w|$ is the length of w.

Normal sequences (continued)

Theorem (Borel, 1909)

The decimal expansion of almost every real number in $[0,1)$ *is a normal sequence in the alphabet* $\{0,1,...,9\}$.

Nevertheless, not so many examples have been proved normal. Some of them are:

- Champernowne 1933 (natural numbers):

  $$12345678910111213141516171819202122232425\cdots$$

- Besicovitch 1935 (squares):

  $$149162536496481100121144169196225256289324\cdots$$

- Copeland and Erdős 1946 (primes):

  $$235711131719232931374143475359616771737983\cdots$$

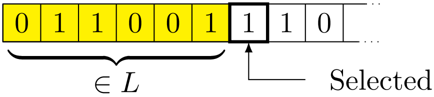
Outline

Selection rules

- If $x = a_1 a_2 a_3 \cdots$ is a normal sequence, then so is $x' = a_2 a_3 a_4 \cdots$ made of digits at all positions but the first one.
- If $x = a_1 a_2 a_3 \cdots$ is normal sequence, then so is $x' = a_2 a_4 a_6 \cdots$ made of digits at even positions.
- What about selecting digits at positions 2^n ?
- What about selecting digits at prime positions ?
- What about selecting digits coming right after 1 ?
- What about selecting digits coming before 1 ?

Oblivious prefix selection

Let $L \subseteq A^*$ be a set of finite words and $x = a_1 a_2 a_3 \cdots \in A^{\mathbb{N}}$.
The oblivious prefix selection of x by L is the sequence
$x \upharpoonright L = a_{i_1} a_{i_2} a_{i_3} \cdots$ where
$\{i_1 < i_2 < i_3 < \cdots\} = \{i : a_1 a_2 \cdots a_{i-1} \in L\}$.



Example (Letters coming right after 1)

If $L = \{0,1\}^*1$, then $i_1 - 1, i_2 - 1, i_3 - 1$ are the positions of 1 in x and $x \upharpoonright L$ is made of the digits coming right after 1.

Theorem (Agafonov 1968)

Prefix selection by a regular set of finite words preserves normality.

The same result does not hold for sets accepted by deterministic one-counter automata or one-turn pushdown automata.

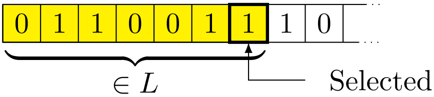
Examples

- $L = (A^k)^* A^r$: Selection along an arithmetic progression (Wall's theorem)
- $L = A^* 1$: Selection of digits after each occurrence of 1
- $L = A^* w$: Selection of digits after each occurrence of w
- $L = (0 + 10^* 1)^*$: Selection after an even number of 1.
- $L = A^* 0(11)^*$: Selection after an even number of 1 since the last 0

Outline

Non-oblivious prefix selection

Let $L \subseteq A^*$ be a set of finite words and $x = a_1 a_2 a_3 \cdots \in A^{\mathbb{N}}$.
The non-oblivious prefix selection of x by L is the sequence
$x \upharpoonright L = a_{i_1} a_{i_2} a_{i_3} \cdots$ where
$\{i_1 < i_2 < i_3 < \cdots\} = \{i : a_1 a_2 \cdots a_{i-1} a_i \in L\}$.



Non-oblivious prefix selection does not preserve normality in general. If $L = A^*1$, then $x \upharpoonright L = 1^{\mathbb{N}} = 111\cdots$.

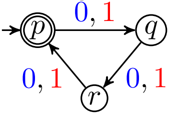
A regular set is called a group set if it is accepted by an automaton where each digit induces a permutation of the states.

Theorem (C., Vandehey)

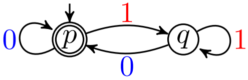
Non-oblivious prefix selection by a group regular set of finite words preserves normality.

Examples

- $L = (A^k)^* A^r$: Selection along an arithmetic progression (Wall's theorem)
  Group set

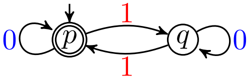


- $L = A^* 1$: Selection of all 1
  Not a group set

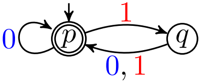


- $L = A^* w$: Selection of the last digit of w
  Not a group set

- $L = (0 + 10^*1)^*$: Selection after an even number of 1 (including the digit itself)
  Group set

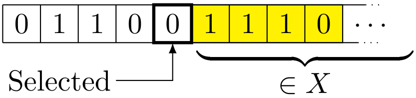


- $L = A^* 0 (11)^*$: Selection after an even number of 1 since the last 0
  Not a group set



Oblivious suffix selection

Let $X \subseteq A^{\mathbb{N}}$ be a set of sequences and $x = a_1 a_2 a_3 \cdots \in A^{\mathbb{N}}$.
The oblivious suffix selection of x by X is the sequence
$x \upharpoonleft X = a_{i_1} a_{i_2} a_{i_3} \cdots$ where
$\{i_1 < i_2 < i_3 < \cdots\} = \{i : a_{i+1} a_{i+2} a_{i+3} \cdots \in X\}$.



Example (Letters coming before 1)

If $X = 1\{0,1\}^{\mathbb{N}}$, then $i_1 + 1, i_2 + 1, i_3 + 1$ are the positions of 1 in x and $x \upharpoonleft X$ is made of the digits coming before 1.

Theorem (Becher, C., Heiber 2018)

Oblivious suffix selection by a regular set of sequences preserves normality.

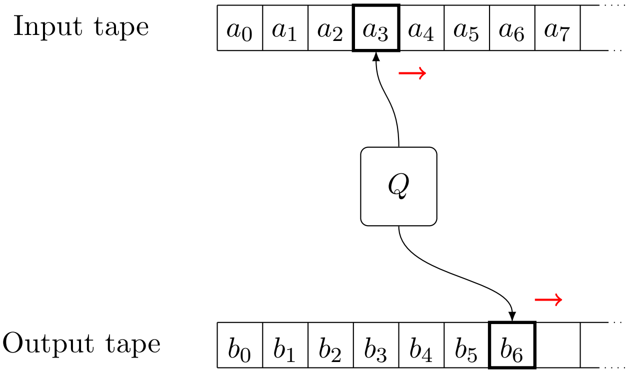
Combining prefix and suffix does not preserve normality in general. Selecting digits having a 1 just before and just after them does not preserve normality.

Examples

- $X = 1A^{\mathbb{N}}$: Selection of digits before each occurrence of 1
- $X = wA^{\mathbb{N}}$: Selection of digits before each occurrence of w (for a finite word w)
- $X = (11)^*0A^{\mathbb{N}}$: Selection before an even number of 1 until the next 0

Outline

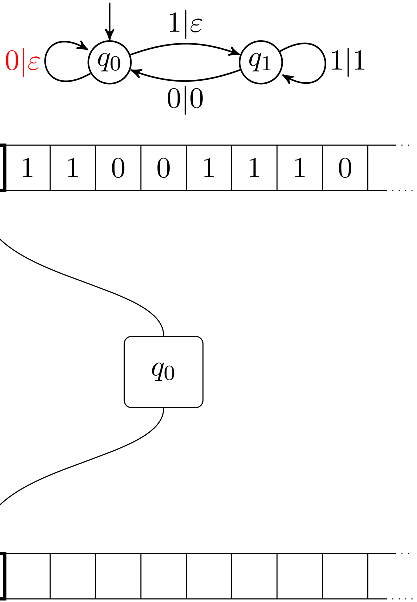
Selectors (special finite state machines)



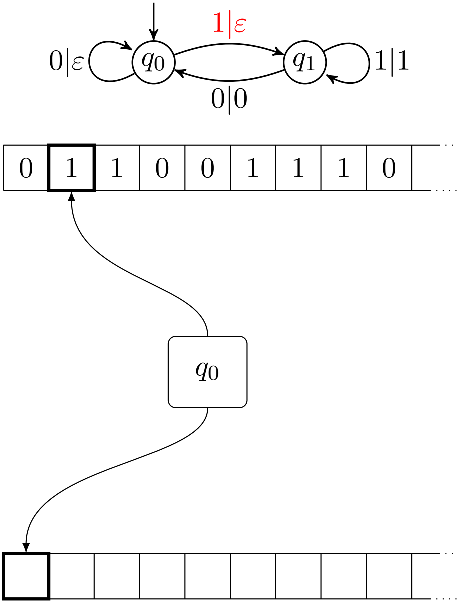
Transitions of the form

- $p \xrightarrow{a|a} q$ for $a \in A$ (copy transition)
- $p \xrightarrow{a|\varepsilon} q$ for $a \in A$ (discard transition)

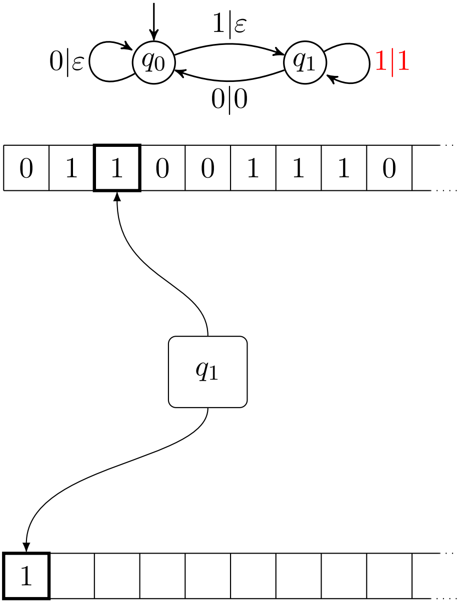
Selecting each digit coming right after 1



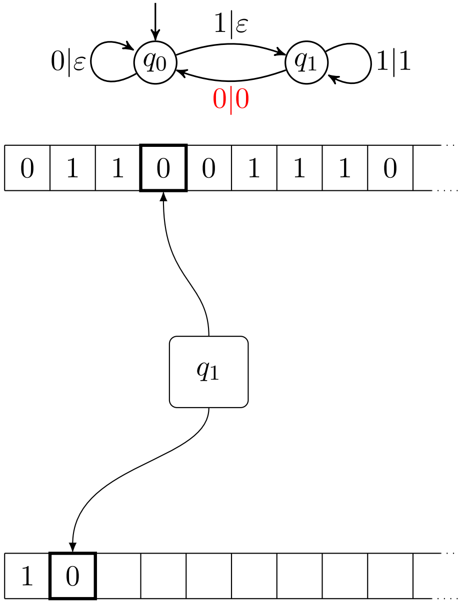
Selecting each digit coming right after 1



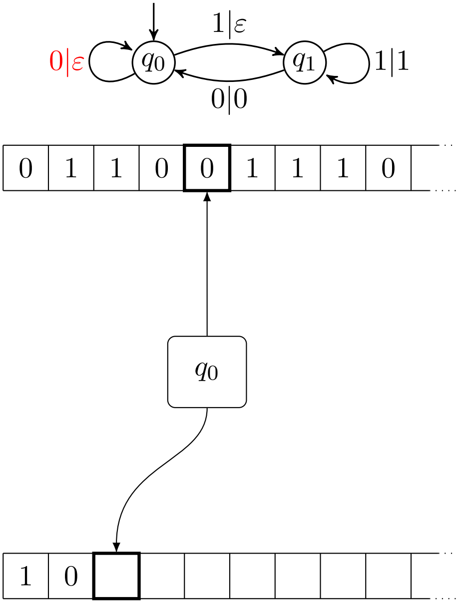
Selecting each digit coming right after 1



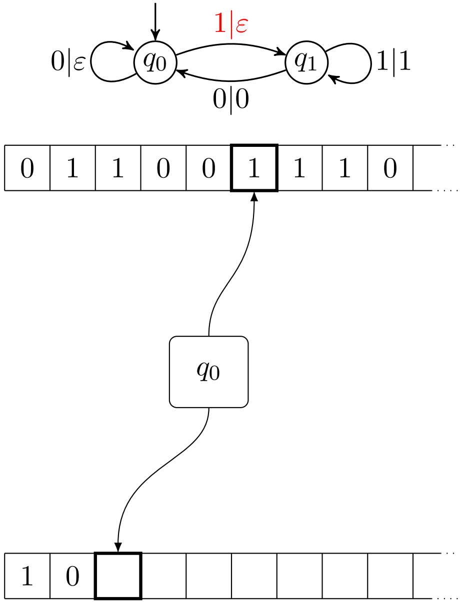
Selecting each digit coming right after 1



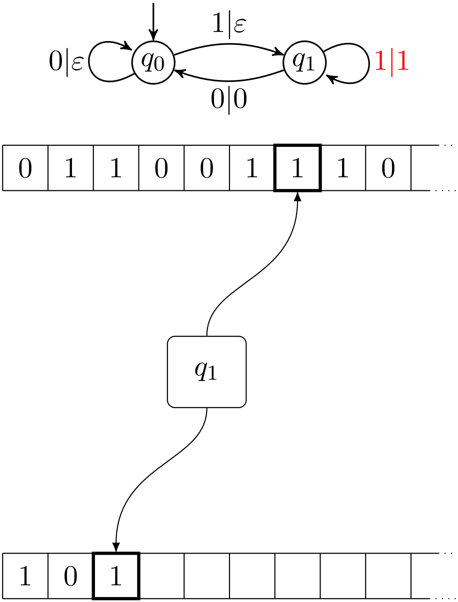
Selecting each digit coming right after 1



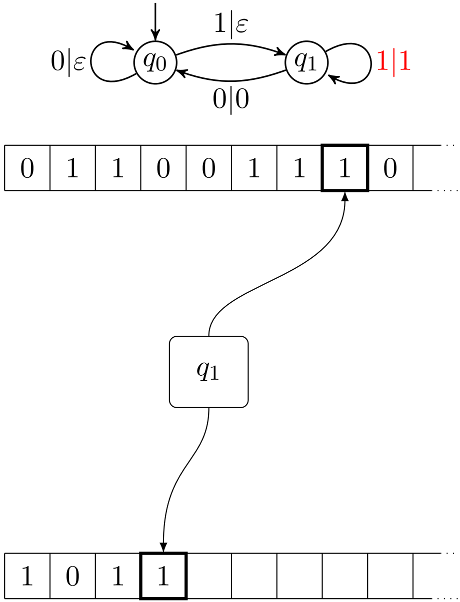
Selecting each digit coming right after 1



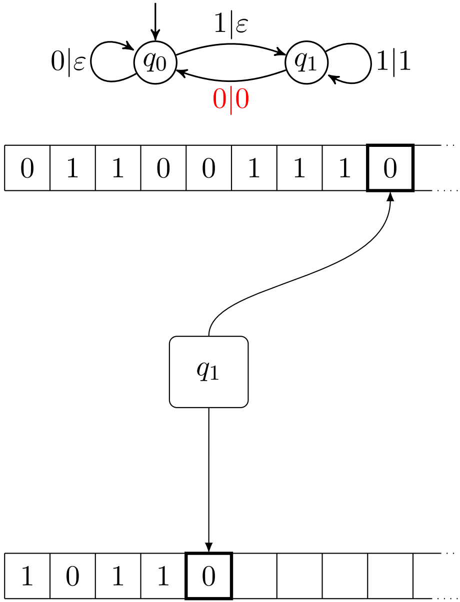
Selecting each digit coming right after 1



Selecting each digit coming right after 1



Selecting each digit coming right after 1



Outline

Normality

Selection

Extensions

Selectors

Shifts of finite type

Shifts of finite type

The shift σ maps each sequence $(x_n)_{n\geqslant 1}$ to $(x_{n+1})_{n\geqslant 1}$.
A shift (space) X is a closed subset $A^{\mathbb{N}}$ such that $\sigma(X) = X$.

Example (Golden mean shift)

The golden mean shift $X = \{0, 10\}^{\mathbb{N}}$ is the set of sequences with no consecutive 1s.

Example (Even shift)

The even shift is the set of sequences where each maximal block of 1 has even length.

For $F \subset A^*$, X_F is the shift of sequences with no block in F.
A shift X is of finite type if $X = X_F$ for some finite set F.

Example (Golden mean shift)

The golden mean shift X is of finite type because $X = X_F$ where $F = \{11\}$.

Invariant measures and the Parry measure

An invariant probability measure on A^* is a function $\mu : A^* \to [0,1]$ such that $\mu(\varepsilon) = 1$ and

$$\sum_{a \in A} \mu(wa) = \sum_{a \in A} \mu(aw) = \mu(w)$$

Its support is the shift X_F where $F = \mu^{-1}(0)$.
If μ is Markovian, its support is of finite type.
The Parry measure of a shift X of finite type is the invariant measure of maximal entropy whose support is X.

A sequence x is μ-generic if for each word $w \in A^*$,

$$\mathrm{freq}(x, w) = \mu(w)$$

Normality in a shift of finite type means genericity for the Parry measure.

Selecting in a shift: problems

Le X be the golden mean shift and μ its Parry measure.
Suppose that x is in X and that it is μ-generic.

- The block 010100101 occurs in x with a non-zero frequency: Selecting along even positions yields the block 11 with a non-zero frequency $\rightsquigarrow$ not in X
- Selecting each digit in x after a 1 yields the sequence $0^{\mathbb{N}} = 000\cdots$. $\rightsquigarrow$ in X but μ-genericity is lost

Selectors compatible with a shift

Let $X = X_F$ be a shift of finite type where $F \subset A^2$.
The selector "knows" the last read and selected digits.
It is compatible with X if there are two function ι and η from its state set Q to A such that:

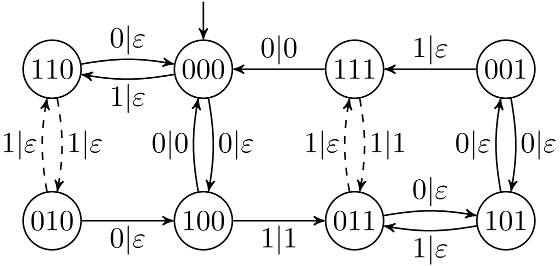
i) If $p \xrightarrow{a|a} q$, then $\iota(p)a \notin F$, $\iota(q) = \eta(q) = a$, and $\eta(p) = \iota(p)$

ii) If $p \xrightarrow{a|\varepsilon} q$, then $\iota(p)a \notin F$, $\iota(q) = a$ and $\eta(q) = \eta(p)$.

- Condition $\iota(p)a \notin F$ means that the selector only reads inputs from X.
- Equality $\iota(q) = a$ guaranties that the last read digit is correct.
- Equalities $\eta(q) = a$ and $\eta(q) = \eta(p)$ guaranty that the last selected digit is correct.
- Equality $\eta(p) = \iota(p)$ only allows the selector to select when the last read and the last selected digits coincide.

A selector compatible with the golden mean shift

It selects if the number of read digits is odd and if the last read and selected digits coincide.



[Dashed transitions are useless if the input is in the golden mean shift]

Each state is labelled by prs where

- $p \in \{0, 1\}$ is the **p**arity of the number of read digits,
- $r \in \{0, 1\}$ is the last **r**ead digit,
- $s \in \{0, 1\}$ is the last **s**elected digit.

Functions ι and η can be defined by $\iota(prs) = r$ and $\eta(prs) = s$.

Preservation of genericity

Theorem (C.)

If X is a shift of finite type and μ is a Markovian measure with support X, selecting with a selector compatible with X preserves μ-genericity.

Since the Parry measure of a shift of finite type is Markovian, normality in shifts of finite type is preserved by compatible selection.

Open questions

- Can we enlarge the class of selectors preserving normality in shifts of finite type ?
- What about sofic shifts (like the even shift) ?

Thank you