A discusión

COORDINATION THROUGH DE BRUIJN SEQUENCES*

Olivier Gossner and Penélope Hernández**

WP-AD 2005-05

Corresponding author: O. Gossner. Paris-Jourdan Sciences Économiques, UMR CNRS - EHESS - ENPC - ENS. 48 boulevard Jourdan. 75014 Paris. Tel.: +33 4 43 13 63 80 / E-mail: Olivier.Gossner@enpc.fr.

Editor: Instituto Valenciano de Investigaciones Económicas, S.A.
Primera Edición Febrero 2005
Depósito Legal: V-998-2005

---

* P. Hernández thanks the financial support from the Spanish Ministry of Education under project SEJ 2004-02172/ECON and the Instituto Valenciano de Investigaciones Económicas (Ivie).

** O. Gossner: Paris-Jourdan Sciences Économiques. P. Hernández: Universidad de Alicante. Dpto. Fundamentos del Análisis Económico.

COORDINATION THROUGH DE BRUIJN SEQUENCES

Olivier Gossner and Penélope Hernández

ABSTRACT

Let μ be a rational distribution over a finite alphabet, and (c_t) be a n-periodic sequences which first n elements are drawn i.i.d. according to μ. We consider automata of bounded size that input c_{t-1} and output y_t at stage t. We prove the existence of a constant C such that, whenever $m \ln m \geq Cn$, with probability close to 1 there exists an automaton of size m such that the empirical frequency of stages such that $y_t = c_t$ is close to 1. In particular, one can take $C = \frac{\overline{p}}{1-\overline{p}} \ln \frac{1}{\underline{p}}$, where $\overline{p} = \max_{q \in \Theta} m(q)$ and $\underline{p} = \min_{q \in \Theta} m(q)$.

Key words: Coordination, complexity, De Bruijn sequences, automata

1 Introduction

A consequence of a Myhill-Nerode's classical theorem on the theory of regular languages (see [HMU01] for instance) is that the size of any automaton that implements a sequence of least period n must be at least n. This result has been used to measure the complexity of strategies in repeated games played by finite automata e.g. by [AR88], [Ney97]. More generally, these games lead to study the complexity of coordination between a periodic sequence (x_t) and an automaton that inputs x_{t-1} at stage t.

Neyman [Ney97] proves that, if $x_1, \dots, x_n$ are drawn i.i.d. according to any probability distribution μ over an alphabet Θ, whenever $m \ln m \ll n$, with probability close to 1 there exist no automaton of size m that achieves non-negligible correlation with the sequence $x_1, \dots x_n, x_1, \dots$. This implies that in a repeated zero-sum game, there exists a sequence of size n (and thus an automaton of size n) that guarantees the value of the stage game against all automata of size m of the opponent if $m \ln m \ll n$.

In this article we prove that if μ is rational, there exists a constant C such that, whenever $m \ln m \geq Cn$, with probability close to 1 there exists an automaton of size m that matches the sequence at almost every stage. In particular, one can take $C = \frac{\overline{p}}{1-\overline{p}} \ln \frac{1}{\underline{p}}$, where $\overline{p} = \max_{\theta \in \Theta} \mu(\theta)$ and $\underline{p} = \min_{\theta \in \Theta} \mu(\theta)$. This implies that the condition $m \ln m \ll n$ in Neyman's result is (almost) tight when μ is rational.

In a previous article [GH03], we prove a similar result when μ is the counting measure. For a given sequence, the construction of an automaton in [GH03] relies on sequences for which the frequencies of all words $y_1, \dots, y_\ell$ of length ℓ

are the same (De Bruijn sequences). In the present work, we rely on generalized De Bruijn sequences, in which the empirical frequency of a word $y_1, \ldots, y_\ell$ of length ℓ is $\Pi_{k=1}^{\ell} \mu(y_k)$. The assumption that μ is rational is needed for the existence of these sequences. The construction of the automaton depends on a statistical condition on the n periodic sequence that we call *regularity*. We prove that the probability of the set of such regular sequences goes to 1 as n goes to infinity using large deviations properties. This approach simplifies the computations in [GH03] that relies on counting arguments, and improves the constant C when μ is uniform over a set X ($\frac{1}{|X|-1} \ln |X|$ instead of $e|X| \ln |X|$).

We present the model in Section 2, and state and prove the main result in Section 3.

2 Model

For $z \in \mathbb{R}$, we let $\lfloor z \rfloor$ and $\lceil z \rceil$ denote the integer part and the superior integer part of z respectively ($z-1 < \lfloor z \rfloor \leq z$ and $z \leq \lceil z \rceil < z+1$). The cardinality of a finite set Z, is denoted $|Z|$. Let Θ be a finite alphabet, and let Θ_n represent the set of n-periodic sequences of elements of Θ.

A *(finite) automaton* $M \in FA(m)$ of size m with inputs and outputs in Θ is a tuple $M = <Q, q^*, f, g>$, where Q s.t. $|Q| = m$ is the finite set of states, $q^* \in Q$ is the initial state, $f : Q \to \Theta$ is the action function, and $g : Q \times \Theta \to Q$ is the transition function.

An automaton $M \in FA(m)$ and a sequence $x = (x_t)_t \in \Theta^{\mathbb{N}}$ induce a sequence of states and actions $(q_1, y_1, q_2, y_2, \ldots)$, where $q_1 = q^*$, $y_1 = f(q^*)$, and for $t \geq 2$, $q_t = g(q_{t-1}, x_{t-1})$, $y_t = f(q_t)$. The corresponding sequence of actions $(y_t)_{t\geq 1}$ chosen by the automaton is denoted $y(x, M)$. If $x^n \in \Theta_n$, then $(x_t, y_t(x^n, M))_t$ is periodic of period at most mn after a finite number of stages.

We define the *ratio of coincidences* between $x^n \in \Theta_n$ and $M \in FA(m)$ is:

$$\rho(x^n, M) = \lim_{T\to\infty} \frac{1}{T} |\{1 \leq t \leq T \colon y_t(x^n, M) = x^n_t\}|$$

$\rho(x^n, M)$ is the average proportion of stages for which M predicts correctly the sequence x^n. Given x^n, the best ratio of coincidences that an automaton of size m can achieve with x^n is $\rho^m(x^n) = \max_{M\in FA(m)} \rho(x^n, M)$.

3 Asymptotic properties

We are concerned with asymptotic properties of the distribution of $\rho^m(x^n)$ when the first n elements of x^n are drawn i.i.d. according to some rational distribution μ in $\Delta(\Theta)$. Let Φ be a common denominator of $(p_i)_{i\in\Theta}$, and denote $\overline{p} = \max_i p_i$, $\underline{p} = \min_i p_i$. We assume wlog. $\underline{p} > 0$. Pr represents the induced probability on the sets Θ_n. Neyman [Ney97] proved the following:

Theorem 1 (Neyman 97) *For a sequence $(m(n))_n$ of positive integers, if $\lim_{n\to\infty} \frac{m(n)\ln m(n)}{n} = 0$ then:*

$$\forall \varepsilon > 0, \lim_{n\to\infty} \Pr(\rho^m(x^n) < \overline{p} + \varepsilon) = 1$$

This result provides an asymptotic condition on m and n, namely $\frac{m \ln m}{n} \to 0$, under which automata of size m cannot achieve coordination ratios larger than $\overline{p}$ with probability close to 1. Our main result shows the existence of a constant C such that if $\frac{m \ln m}{n}$ is asymptotically larger than C, then automata of size n can achieve coordination ratios arbitrarily close to 1 with a set of periodic sequences of probability close to 1.

Theorem 2 *There exists a constant C such that for any sequence of positive integers $(m(n))_{m\in\mathbb{N}}$ with $\lim_{n\to\infty} \frac{m(n)\ln m(n)}{n} > C$,*

$$\forall\varepsilon,\ \Pr(\rho^m(x^n) > 1-\varepsilon) \longrightarrow 1$$

In particular, one can take $C = \frac{\overline{p}}{1-\overline{p}} \ln \frac{1}{\underline{p}}$.

To prove this, we define in Section 3.1 a subset of Θ_n of sequences verifying a statistical regularity condition. We call those sequences *regular*. Then, in Section 3.2, for each regular sequence x^n, we construct an automaton in $FA(m)$ that achieves a large ratio of coincidences with x^n. We estimate the probability of regular sequences in Section 3.3, and conclude the proof in Section 3.4.

3.1 Regularity

In this section we define the statistical regularity condition that ensures a large ratio of coincidences. Let $x = x^n = (x_1, x_2, \ldots) \in \Theta_n$ and $\ell \leq n$. We call *word* an element of Θ^ℓ. We identify x to its n first elements, thus making the abuse of notation $x \in \Theta^n$. For $1 \leq j \leq \left\lfloor \frac{n}{\ell} \right\rfloor$, we write $r_j = (x_{\ell(j-1)+1}, \ldots, x_{\ell j})$ and $r' = (x_{\lfloor \frac{n}{\ell} \rfloor \ell+1}, \ldots, x_{n-1}, x_n)$. This way, x can be expressed as the concatenation of the words $r_1, \ldots, r_{\lfloor \frac{n}{\ell} \rfloor}$ and of $r' \in \Theta^{n-\ell\lfloor \frac{n}{\ell} \rfloor}$. Let x^* be the concatenation of $r_1, \ldots, r_{\lfloor \frac{n}{\ell} \rfloor}$. The number of times that a word r appears in x^* is

$$S(x^*, r) = \left| \left\{ 0 \leq j \leq \left\lfloor \frac{n}{\ell} \right\rfloor : r_j = r \right\} \right|.$$

For $\alpha > 1$, we define the set of (α, ℓ)-regular (or regular for short) sequences $R_\ell(n, \alpha)$ as the subset of elements x of Θ_n such that for each word r, $S(x^*, r) \leq$

$\alpha \frac{n}{\ell} \Pr(r)$.

3.2 Construction of an automaton for regular sequences

Proposition 3 *Let $x \in R_\ell(n, \alpha)$. With $m = \lceil \alpha \frac{\overline{p}}{1-\overline{p}} \frac{n}{\ell \Phi^\ell} \rceil \Phi^\ell + \ell$, $\rho^m(x) \geq 1 - \frac{1}{\ell}$.*

The proof of the proposition is constructive.

3.2.1 Proof of Proposition 3

We present the construction of an automaton $M = <Q, q^*, f, g> \in FA(m)$ that ensures a sufficient coincidence ratio with $x \in R_\ell(n, \alpha)$. First, we design Q and f, second we define q^* and g. Finally we check that M achieves the desired ratio of coincidences with x.

3.2.1.1 Construction of the state space and action function The state space and action function we design depend only on μ, α, n and ℓ, they are independent of the particular element x of $R_\ell(n, \alpha)$. Our construction relies on a sequence of elements of Θ such that the empirical frequencies of each word coincides with its probability under Pr. To construct this sequence, we first construct a sequence over an alphabet of size Φ of minimal length Φ^ℓ in which each subsequence of length ℓ appears once.

The empirical frequency of a word r in a sequence $s \in \Theta_L$ is:

$$EF(s, r) = \frac{1}{L} |\{1 \leq j \leq L \colon (s_j, s_{j+1}, \ldots, s_{j+\ell-1}) = r\}|$$

Lemma 4 *There exists a sequence $s \in \Theta_{\Phi^\ell}$ such that $EF(s, r) = \Pr(r)$ for every word r.*

Proof. Let $\Phi = \{1, \ldots, \Phi\}$, and $\tilde{s} \in \Phi_{\Phi^\ell}$ be a De Bruijn sequence of length Φ^ℓ over Φ (cf. for instance [vLW01], chapter 8, p. 56). The empirical frequency $EF(\tilde{s}, \tilde{r})$ of each $\tilde{r} \in \Phi^\ell$ is then $\frac{1}{\Phi^\ell}$.

Let $\pi : \Phi \to \Theta$ be such that for every $i \in \Theta$, $|\pi^{-1}(i)| = p_i\Phi$, and let $s = (\pi(\tilde{s}_t))_t$. The application from Φ^ℓ to Θ^ℓ canonically induced by π is also denoted π. For $r \in \Theta^\ell$, it is straight forward that $EF(s, r) = \Pr(r)$. ■

Let $Q = Q_1 \bigcup Q_2$ with $Q_1 = \{1, \ldots, \lceil \alpha \frac{n}{\ell \Phi^\ell} \frac{\overline{p}}{1-\overline{p}} \rceil\} \times \{1, \ldots, \Phi^\ell\}$ and $Q_2 = \{1, \ldots, n - \lfloor \frac{n}{\ell} \rfloor \ell\}$.

We let $(s_1, \ldots, s_{\Phi^\ell}) \in \Phi^\ell$ be the first elements of a sequence as in Lemma 4, and define f by $f(q) = s_t$ if $q = (k, t) \in Q_1$ and $f(q) = x_{\lfloor \frac{n}{\ell} \rfloor \ell + q}$ if $q \in Q_2$

3.2.1.2 Construction of the transition function and initial state

For $q = (k, t) \in Q_1$ and $c \in \mathbb{N}$ we let $q + c = (k, t + c \bmod \Phi^\ell)$. Given a word $r \in \Theta^\ell$, let $\overline{C_r}$ be the set of $\overline{r} \in \Theta^\ell$ such that $\overline{r}_i = r_i$ for $1 \leq i < \ell$ and $\overline{r}_\ell \neq r_\ell$. Notice that the cardinality of $\overline{C_r}$ equals $|\Theta| - 1$.

The crucial element of the construction is the existence of a map between the index of the words r_t to Q, as stated by the following lemma.

Lemma 5 *There exists an injective map β from $\{1, \ldots, \left\lfloor \frac{n}{\ell} \right\rfloor\}$ to Q_1 such that*

$$(f(\beta(t)), \ldots, f(\beta(t) + \ell)) \in \overline{C}_{r_t}$$

Proof. Let $T(\overline{r}, Q_1) = \{q \in Q_1, (f(q)), \ldots, f(q + l)) = \overline{r}\}$ and $\overline{T}(r, Q_1) = \sum_{\overline{r} \in \overline{C}_r} |T(r, Q_1)|$. It is enough to prove that for every r, $S(x^*, r) \leq \overline{T}(r, Q_1)$. On the one hand, $S(x^*, r) \leq \alpha \frac{n}{\ell} \Pr(r)$ since x is regular. On the other hand, $\overline{T}(r, Q_1) = \lceil \alpha \frac{\overline{p}}{1-\overline{p}} \frac{n}{\ell \Phi^\ell} \rceil \Phi^\ell \Pr(\overline{C}_r) \geq (\alpha \frac{\overline{p}}{1-\overline{p}} \frac{n}{\ell \Phi^\ell}) \Phi^\ell \Pr(r) \frac{1-\overline{p}}{\overline{p}}$. Hence the result. ■

Let the initial state be $q^* = \beta(1)$. We first define the transition function when M matches the sequence.

- For $q \in Q_1$, $g(q, f(q)) = q + 1$
- For $q \in Q_2$
  - · For $1 \leq t < n - \lfloor \frac{n}{\ell} \rfloor \ell$, $g(t, f(t)) = t + 1$
  - · $g(n - \lfloor \frac{n}{\ell} \rfloor \ell, f(n - \lfloor \frac{n}{\ell} \rfloor \ell)) = q^*$.

We now define $g(q, a)$ for $a \neq f(q)$.

- If $q = \beta(t) + \ell - 1$ for some $1 \leq t \leq \left\lfloor \frac{n}{\ell} \right\rfloor$, this t is then unique since β is injective.
  - · If $t < \left\lfloor \frac{n}{\ell} \right\rfloor$, let $g(q, a) = \beta(t + 1)$ for all $a \neq f(q)$.
  - · If $t = \left\lfloor \frac{n}{\ell} \right\rfloor \neq \frac{n}{\ell}$, let $g(q, a) = 1 \in Q_2$ for all $a \neq f(q)$.
  - · If $t = \left\lfloor \frac{n}{\ell} \right\rfloor = \frac{n}{\ell}$, let $g(q, a) = q^*$ for all $a \neq f(q)$.
- If there exists no t such that $q = \beta(t) + \ell - 1$ we let $g(q, a)$ when $a \neq f(q)$ arbitrary.

3.2.1.3 The induced sequence of actions and states We now check that M has sufficient ratio of coincidences with x.

Lemma 6 $\rho(x, M) \geq 1 - \frac{1}{\ell}$

Proof. Let $(q^*, y_1, q_2, \ldots)$ be the sequence of states and actions induced by M and x. We prove by induction that for $t = 0, \ldots, \left\lfloor \frac{n}{l} \right\rfloor$, $q_{\ell t+1} = \beta(t + 1)$. This property is verified for $t = 0$ since $q^* = \beta(r_1)$. Assume it is true for some $t < \left\lfloor \frac{n}{\ell} \right\rfloor$. From the definition of β, the sequence of actions played by M coincides with r_t at stages $\ell t + 1, \ldots, \ell(t + 1) - 1$ and differs at stage $\ell(t + 1)$. Hence the property.

Furthermore, we have proved that $(y_{\ell t+1}, \ldots, y_{(\ell+1)t}) \in \overline{C}_{r_t}$ for those t. The sequence of actions and states from stage $\left\lfloor \frac{n}{\ell} \right\rfloor \ell{+}1$ to n is $f(1), \ldots, f(n - \lfloor \frac{n}{\ell} \rfloor \ell) = r'$, and at stage $n{+}1$, M reaches the state $q_{n+1} = q^*$, which implies that $y(M, x)$ is n-periodic.

The ratio of coincidences between x and M is then: $\rho(x, M) = \frac{n - \lfloor \frac{n}{\ell} \rfloor}{n} \geq 1 - \frac{1}{\ell}$

■

Since the number of states of M is not larger than $\lceil \alpha \frac{\overline{p}}{1-\overline{p}} \frac{n}{\ell \Phi^\ell} \rceil \Phi^\ell + \ell$, this proves Proposition 3.

3.3 Probability of regular sequences

We estimate the probability of the set $R_\ell(n, \alpha)$ of regular sequences.

Lemma 7 *For every $\alpha > 1$, there exists $C = C(\alpha)$ such that for every ℓ, n:*

$$\Pr(R_\ell(n, \alpha)) \geq 1 - \Theta^\ell \exp\{-C(\alpha) \frac{n}{\ell} \underline{p}^\ell\}$$

Proof. For a given word r, $S(x^*, r)$ is the sum of $\lfloor \frac{n}{\ell} \rfloor$ independent indicator random variables, and the expected number of occurrences of r is

$$\mathbf{E} S(x^*, r) = \lfloor \frac{n}{\ell} \rfloor \Pr(r).$$

From Azuma's inequality (see e.g. [AS00]), there exists $C = C(\alpha)$ such that:

$$\Pr(S(x^*, r) > \alpha \lfloor \frac{n}{\ell} \rfloor \Pr(r)) \leq \exp\{-C(\alpha) \lfloor \frac{n}{\ell} \rfloor \Pr(r)\} \leq \exp\{-C(\alpha) \lfloor \frac{n}{\ell} \rfloor \underline{p}^\ell\}$$

Summing over all possible values of r,

$$\Pr(x \notin R_\ell(n, \alpha)) \leq \sum_{r \in \Theta^\ell} \Pr(S(x^*, r) > \alpha \lfloor \frac{n}{\ell} \rfloor P(r)) \leq |\Theta|^\ell \exp\{-C(\alpha) \lfloor \frac{n}{\ell} \rfloor \underline{p}^\ell\}$$

■

3.4 Proof of Theorem 2

Consider a sequence $m(n)$ such that $\lim \frac{m(n)\ln(m(n))}{n} > \frac{\overline{p}}{1-\overline{p}} \ln \frac{1}{\underline{p}}$, and let $\alpha > 1$ such that for n sufficiently large, $\frac{m(n)\ln(m(n))}{n} > \alpha \frac{\overline{p}}{1-\overline{p}} \ln \frac{1}{\underline{p}}$.

Let $\ell_0(n)$ be the unique solution of the equation $x^3(\frac{1}{\underline{p}})^x = n$ and $\ell(n) = \lceil \ell_0(n) \rceil$. We denote $m(n)$ by m, and similarly for ℓ. The next lemma states that the probability of regular sequences $R_\ell(n,\alpha)$ tends to 1 as n goes to infinity.

Lemma 8

$$\lim_{n\to\infty} \Pr(R_\ell(n,\alpha)) = 1$$

Proof. From Lemma 7, there exists $C > 0$ such that $\Pr(x \notin R_\ell(n,\alpha)) < |\Theta|^\ell \exp\{-C\lfloor \frac{n}{\ell} \rfloor \underline{p}^\ell\}$. We compute the limit of $\ln \Pr(x \notin R_\ell(n,\alpha))$.

$$\lim_{n\to\infty} \ln(|\Theta|^\ell \exp\{-C\lfloor \frac{n}{\ell} \rfloor \underline{p}^\ell\}) = \lim_{n\to\infty} \ell \ln |\Theta| - C\lfloor \frac{n}{\ell} \rfloor \underline{p}^\ell = -\infty$$

■

The next lemma shows that the automaton constructed in Proposition 3 belongs to $FA(m)$.

Lemma 9 *For n large enough, $m \geq \lceil \alpha \frac{\overline{p}}{1-\overline{p}} \frac{n}{\ell \Phi^\ell} \rceil \Phi^\ell + \ell$.*

Proof. Let $m' = \lceil \alpha \frac{\overline{p}}{1-\overline{p}} \frac{n}{\ell \Phi^\ell} \rceil \Phi^\ell + \ell$.

$$\limsup \frac{m' \ln m'}{n} \leq \alpha \frac{\overline{p}}{1-\overline{p}} \ln \frac{1}{\underline{p}} < \lim \frac{m \ln m}{n}$$

■

References

[AR88] D. Abreu and A. Rubinstein. The structure of nash equilibrium in repeated games with finite automata. *Econometrica*, 56:1259–1281, 1988.

[AS00] N. Alon and J. Spencer. *The probabilistic method*. Interscience Series in Discrete Mathematics and Optimization. John Wiley, New York, second edition, 2000.

[GH03] O. Gossner and P. Hernández. On the complexity of coordination. *Mathematics of Operations Research*, 28:127–141, 2003.

[HMU01] J. Hopcroft, R. Motwani, and J. Ullman. *Introduction to Automata Theory, Languages, and Computation*. Addison-Wesley, Amsterdam, second edition, 2001.

[Ney97] A. Neyman. Cooperation, repetition, and automata. In S. Hart and A. Mas-Colell, editors, *Cooperation: Game-Theoretic Approaches*, volume 155 of *NATO ASI Series F*, pages 233–255. Springer-Verlag, 1997.

[vLW01] J.H. van Lint and R.M. Wilson. *A course in combinatorics*. Cambridge University Press, Cambridge, second edition, 2001.