Available online at www.sciencedirect.com
SCIENCE DIRECT®

Operations Research Letters 34 (2006) 17–21

Operations Research Letters
www.elsevier.com/locate/orl

Coordination through De Bruijn sequences

Olivier Gossner[a,*], Penélope Hernández[b,1]

[a]Paris-Jourdan Sciences Économiques, UMR CNRS - EHESS - ENPC - ENS 8545, 48 boulevard Jourdan, 75014 Paris, France
[b]Departmento Fundamentos del Análisis Económico, Universidad de Alicante, Carretera de San Vicente s/n, 03080 San Vicente, Alicante, Spain

Received 5 October 2004; accepted 17 January 2005
Available online 17 March 2005

Abstract

Let (x_t) be an n-periodic sequence in which the first n elements are drawn i.i.d. according to some rational distribution. We prove there exists a constant C such that whenever $m \ln m \geqslant Cn$, with probability close to 1, there exists an automaton of size m that matches the sequence at almost all stages.

Keywords: Coordination; Complexity; De Bruijn sequences; Automata

1. Introduction

A consequence of Myhill–Nerode's classical theorem on the theory of regular languages (see [4] 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 [1,5]. 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 [5] proves that, if $x_1, \ldots, x_n$ 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 exists no automaton of size m that achieves non-negligible correlation with the sequence $x_1, \ldots, x_n, x_1, \ldots$. 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 $m \ln m \geqslant Cn$, 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 $C = \overline{p}/(1-\overline{p}) \ln 1/\underline{p}$, 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.

* Corresponding author.
E-mail addresses: Olivier.Gossner@enpc.fr (O. Gossner), penelope@merlin.fae.ua.es (P. Hernández).
[1] This author thanks the financial support from the Spanish Ministry of Education under project SEJ 2004-02172/ECON and the Instituto Valenciano de Investigaciones Económicas (Ivie).

0167-6377/$ - see front matter
doi:10.1016/j.orl.2005.01.006

In a previous article [3], we prove a similar result when μ is the counting measure. For a given sequence, the construction of an automaton in [3] 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 $y_1, \dots, y_\ell$ 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 deviation properties. This approach simplifies the computations in [3] that relies on counting arguments, and improves the constant C when μ is uniform over a set X $(1/(|X|-1)\ln|X|$ 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 $(z-1 < \lfloor z \rfloor \leqslant z$ and $z \leqslant \lceil z \rceil < z+1)$. The cardinality of a finite set Z, is denoted by $|Z|$. 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 $M = \langle Q, q^*, f, g \rangle$, 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 $(q_1, y_1, q_2, y_2, \dots)$, where $q_1 = q^*$, $y_1 = f(q^*)$, and for $t \geqslant 2$, $q_t = g(q_{t-1}, x_{t-1})$, $y_t = f(q_t)$. The corresponding sequence of actions $(y_t)_{t \geqslant 1}$ chosen by the automaton is denoted by $y(x, M)$. 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 $M \in FA(m)$ as

$$\rho(x^n, M) = \lim_{T\to\infty} \frac{1}{T} \left| \left\{ 1 \leqslant t \leqslant T : y_t(x^n, M) = x_t^n \right\} \right|$$

$\rho(x^n, M)$ 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 w.l.o.g. $\underline{p} > 0$. Pr represents the induced probability on the sets Θ_n. Neyman [5] proved the following:

Theorem 1 (*Neyman [5]*). *For a sequence* $(m(n))_n$ *of positive integers, if*

$$\lim_{n\to\infty} \frac{m(n) \ln m(n)}{n} = 0$$

then

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

This result provides an asymptotic condition on m and n, namely $(m \ln m)/n \to 0$, 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 $(m \ln m)/n$ is asymptotically larger than C, then automata of size m 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*

$$\lim_{n\to\infty} \frac{m(n) \ln m(n)}{n} > C,$$

then

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

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 these 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 $\ell \leqslant n$. 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 $1 \leqslant j \leqslant \lfloor n/\ell \rfloor$, we write $r_j = (x_{\ell(j-1)+1}, \ldots, x_{\ell j})$ and $r' = (x_{\lfloor n/\ell \rfloor \ell+1}, \ldots, x_{n-1}, x_n)$. This way, x can be expressed as the concatenation of the words $r_1, \ldots, r_{\lfloor n/\ell \rfloor}$ and of $r' \in \Theta^{n-\ell\lfloor n/\ell \rfloor}$. Let x^* be the concatenation of $r_1, \ldots, r_{\lfloor n/\ell \rfloor}$. The number of times that a word r appears in x^* is

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

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, $S(x^*, r) \leqslant \alpha (n/\ell) \Pr(r)$.

3.2. Construction of an automaton for regular sequences

Proposition 3. *Let $x \in R_\ell(n, \alpha)$. With*

$$m = \left\lceil \alpha \frac{\overline{p}}{1 - \overline{p}} \frac{n}{\ell \Phi^\ell} \right\rceil \Phi^\ell + \ell, \quad \rho^m(x) \geqslant 1 - \frac{1}{\ell}.$$

The proof of the proposition is constructive.

3.2.1. Proof of Proposition 3

We present the construction of an automaton $M = \langle Q, q^*, f, g \rangle \in FA(m)$ 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 frequency 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:

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

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 [6, Chapter 8, p. 56]). The empirical frequency $EF(\tilde{s}, \tilde{r})$ of each $\tilde{r} \in \Phi^\ell$ is then $1/\Phi^\ell$.

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 by π. For $r \in \Theta^\ell$, it is straightforward that $EF(s, r) = \Pr(r)$. □

Let

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

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 $f(q) = x_{\lfloor n/\ell \rfloor \ell + q}$ 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 $1 \leqslant i < \ell$ 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* $\{1, \ldots, \lfloor n/\ell \rfloor\}$ *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, $S(x^*, r) \leqslant \overline{T}(r, Q_1)$. On the one hand, $S(x^*, r) \leqslant \alpha(n/\ell)\Pr(r)$ since x is regular. On the other hand,

$$\overline{T}(r, Q_1) = \left\lceil \alpha \frac{\overline{p}}{1-\overline{p}} \frac{n}{\ell \Phi^\ell} \right\rceil \Phi^\ell \Pr(\overline{C}_r) \geqslant \left(\alpha \frac{\overline{p}}{1-\overline{p}} \frac{n}{\ell \Phi^\ell} \right) \Phi^\ell \Pr(r) \frac{1-\overline{p}}{\overline{p}}.$$

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 $1 \leqslant t < n - \lfloor n/\ell \rfloor \ell$, $g(t, f(t)) = t+1$
  - $g(n - \lfloor n/\ell \rfloor \ell, f(n - \lfloor n/\ell \rfloor \ell)) = q^*$.

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

- If $q = \beta(t) + \ell - 1$ for some $1 \leqslant t \leqslant \lfloor n/\ell \rfloor$, this t is then unique since β is injective.
  - If $t < \lfloor n/\ell \rfloor$, let $g(q, a) = \beta(t+1)$ for all $a \neq f(q)$.
  - If $t = \lfloor n/\ell \rfloor \neq n/\ell$, let $g(q, a) = 1 \in Q_2$ for all $a \neq f(q)$.
  - If $t = \lfloor n/\ell \rfloor = n/\ell$, 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. $\rho(x, M) \geqslant 1 - 1/\ell$.

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 $t=0, \ldots, \lfloor n/l \rfloor$, $q_{\ell t+1} = \beta(t+1)$. This property is verified for $t=0$ since $q^* = \beta(r_1)$. Assume it is true for some $t < \lfloor n/\ell \rfloor$. 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 $\lfloor n/\ell \rfloor \ell + 1$ to n is $f(1), \ldots, f(n - \lfloor n/\ell \rfloor \ell) = r'$, 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:

$$\rho(x, M) = \frac{n - \lfloor n/\ell \rfloor}{n} \geqslant 1 - \frac{1}{\ell}. \qquad \square$$

Since the number of states of M is not larger than

$$\left\lceil \alpha \frac{\overline{p}}{1-\overline{p}} \frac{n}{\ell \Phi^\ell} \right\rceil \Phi^\ell + \ell,$$

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:

$$\Pr(R_\ell(n, \alpha)) \geqslant 1 - \Theta^\ell \exp\left\{ -C(\alpha) \left\lfloor \frac{n}{\ell} \right\rfloor \underline{p}^\ell \right\}.$$

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

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

From Azuma's inequality (see e.g. [2]), there exists $C = C(\alpha)$ such that:

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

Summing up all possible values of r,

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

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 $x^3(1/\underline{p})^x = n$ 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 $\Pr(x \notin R_\ell(n,\alpha)) \leqslant |\Theta|^\ell \exp\{-C\lfloor n/\ell\rfloor \underline{p}^\ell\}$. We compute the limit of ln $\Pr(x \notin R_\ell(n,\alpha))$.

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

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

Lemma 9. *For n large enough*,

$$m \geqslant \left\lceil \alpha \frac{\overline{p}}{1-\overline{p}} \frac{n}{\ell\Phi^\ell} \right\rceil \Phi^\ell + \ell.$$

Proof. Let

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

Then

$$\limsup \frac{m'\ln m'}{n} \leqslant \alpha \frac{\overline{p}}{1-\overline{p}} \ln\frac{1}{\underline{p}} < \lim \frac{m\ln m}{n}. \qquad \square$$

References

[1] D. Abreu, A. Rubinstein, The structure of Nash equilibrium in repeated games with finite automata, Econometrica 56 (1988) 1259–1281.

[2] N. Alon, J. Spencer, The probabilistic method, Interscience Series in Discrete Mathematics and Optimization, second ed., Wiley, New York, 2000.

[3] O. Gossner, P. Hernández, On the complexity of coordination, Math. Oper. Res. 28 (2003) 127–141.

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

[5] A. Neyman, Cooperation, repetition, and automata, in: S. Hart, A. Mas Colell (Eds.), Cooperation: Game-Theoretic Approaches, NATO ASI Series F, vol. 155. Springer, Berlin, 1997, pp. 233–255.

[6] J.H. van Lint, R.M. Wilson, A Course in Combinatorics, Cambridge University Press, Cambridge, 2001.