The Game of Normal Numbers

Ehud Lehrer*†

September 4, 2003

Abstract

We introduce a two-player game where at each period one player, say, Player 2, chooses a distribution and the other player, Player 1, a realization. Player 1 wins the game if the sequence of realized outcomes is normal with respect to the sequence of distributions. We present a pure winning strategy of Player 1 and thereby provide a universal algorithm that generates a normal sequence for any discrete stochastic process. It turns out that to select the n^{th} digit, the algorithm conducts $O(n^2)$ calculations. The proof uses approachability in infinite-dimensional spaces (Lehrer 2002).

*School of Mathematical Sciences, Tel Aviv University, Tel Aviv 69978, Israel; e-mail: lehrer@post.tau.ac.il; home page: www.math.tau.ac.il/∼lehrer. This research was supported by the Israel Science Foundation, Grant no. 178/99.

†I am grateful to Eilon Solan and to the two anonymous referees of *Mathematics of Operations Research* for helpful comments.

1 Introduction

Consider the binary expansion of a number in the interval [0,1]. A number is called normal with respect to the binary expansion, if the relative frequency of any finite string of n zeros and ones converges to 2^{-n}.

The notion of a normal number can easily be extended to any stationary process. A normal number with respect to a stationary process is such that the frequency of any finite string of symbols converges to the corresponding probability.

Borel (1909) concluded from the strong law of large numbers that a number (with respect to any given basis[1]) is normal with probability 1 with respect to the Lebesgue distribution. Nevertheless, the construction of a specific normal number with respect to a specific basis, is not trivial. Champernowne (1933) showed that if one writes down the decimal numbers successively: 1 2 3 4 5 6 7 8 9 10 11... , the result is a decimal normal number. This result follows from the weak law of large numbers. Copeland and Erdős (1946) and Davenport and Erdős (1952) generalized this result.

Another way to define a normal number with respect to the binary expansion is by relative conditional frequencies. A number is normal if for any finite string of zeros and ones, say x^n, the relative frequency of the zeros that appear immediately after x^n is asymptotically $\frac{1}{2}$.

Here we extend the notion of a normal number beyond stationary processes, using relative conditional frequencies. For any discrete stochastic process we define a normal sequence as follows. A sequence is normal if for any finite string, say x^n, the difference between the relative frequency of the appearances of any symbol, say x, after x^n, and the average conditional probability of x diminishes to zero.

[1] That is, with respect to any i.i.d. uniform distribution over a finite set.

Consider, for instance, a Markov chain with two states: 0 and 1. Suppose that the transition probability from 0 to 1 is $\frac{2}{3}$ and from 1 to 0 is $\frac{1}{4}$. In a normal sequence the asymptotic frequency of ones that appear after every history which terminates with 0 is $\frac{2}{3}$ and after every history which terminates with 1 is $\frac{3}{4}$. Although there are many normal numbers (in fact, the probability of the normal numbers, with respect to the distribution induced by the Markov Chain, is 1), the construction of a specific number is not trivial. Smorodinsky and Weiss (1987) gave an explicit construction for the case of ergodic Markov chains which was inspired by Champernowne (1933). Adler, Keane and Smorodinsky (1981) constructed a normal number for another specific distribution.

The notion of normal numbers is strongly related to the notion of calibration (Dawid, 1982). The latter became central to the subject of learning to play equilibrium due to the important paper of Foster and Vohra (1997). Given a distribution over infinite sequences, a calibration test (see Kalai, Lehrer and Smorodinsky, 1999) compares the empirical distribution of a certain event, along an infinite sequence, with its expected value. A sequence passes a calibration test if the empirical distribution and the expected value of the event under consideration are asymptotically equal. Using this terminology, a normal number may be defined as a sequence that passes countably many calibration tests, one for each finite string of digits.

Sandroni and Smorodinsky (2002) were inspired by the notion of normal numbers when they introduced belief-based equilibrium. In a belief-based equilibrium, players optimize against subjective beliefs. Moreover, the realized sequences of actions pass calibration tests like those used to determine whether a number is normal or not.

We introduce a two-player game played over a discrete set of periods. In each period, Player 2 chooses a distribution over a finite or countable set X. After being informed of Player 2's choice, Player 1 selects an element in X (a realization). Player 1 wins the game if the sequence of realizations is normal with respect to the sequence of distributions chosen by Player 2.

It is shown that Player 1 has a *pure* winning strategy, that is, for every strategy of Player 2, Player 1 can construct a normal sequence with respect to Player 2's choices. A winning strategy of Player 1 provides a universal algorithm that generates a normal sequence for any discrete stochastic process.

It turns out that to select the n^{th} digit, the algorithm conducts a quadratic (in n) number of calculations. The proof relies on Lehrer (2002) which extends Blackwell's approachability theorem (1956) to infinite-dimensional spaces.

2 Extended Normal Numbers

Let X be a finite or countable set of digits and let θ be a probability distribution over $X^{I\!N}$, the Cartesian product of X with itself countably many times.[2] For any $x^{I\!N} \in X^{I\!N}$, n and $\ell \in I\!N$, where $\ell < n$, denote $x^{\ell,n} = (x_\ell, x_{\ell+1}, ..., x_n)$ and $x^n = x^{1,n}$. The probability with respect to θ that $z \in X$ will appear after x^n is denoted by $\theta(z|x^n)$.

For any $k = 1, 2, ...$, $z^k = (z_1, ..., z_k) \in X^k$, $x^{I\!N} \in X^{I\!N}$ and $n \in I\!N$, let $I(x^{n-1}, z^k)$ be 1 if $x^{n-k,n-1} = z^k$ and 0 otherwise. In other words, $I(z^k, x^{n-1})$ attains the value 1 if the tail of the word x^{n-1} coincides with the string z^k (i.e., $x^{n-k,n-1} = z^k$) and the value 0 otherwise. If $k = 0$, set $I(x^{n-1}, z^k) = 1$.

For any integer $0 \leq k < n$ denote $\overline{I}(x^{n-1}, z^k) = \sum_{s=k}^{n-1} I(x^s, z^k)$. When $0 < k$, $\overline{I}(x^{n-1}, z^k)$ is the number of times the string z^k appears in the word x^{n-1}. Let $z \in X$ and set

$$Y_\theta(x^n; z^k, z) = \Big(1\!\!1(x_n = z) - \theta\big(z|x^{n-1}\big)\Big) I(x^{n-1}, z^k),$$

where $1\!\!1$ denotes the characteristic function. That is, $1\!\!1(x_n = z) = 1$ if $x_n = z$ and 0 otherwise. $Y_\theta(x^n; z^k, z)$ may attain one of three values: 0 if the word

[2] As usual $I\!N$ denotes the set of natural numbers.

x^{n-1} does not end with the string z^k; $1-\theta\big(z|x^{n-1}\big)$ if the word x^{n-1} ends with the string z^k and the last letter of the word x^n is z; and lastly, $-\theta\big(z|x^{n-1}\big)$ if the word x^{n-1} ends with the string z^k and the last letter of the word x^n is *not* z.

Denote,

$$\overline{Y}_\theta(x^n; z^k, z) = \frac{\sum_{s=k+1}^n Y_\theta(x^s; z^k, z)}{\overline{I}(x^{n-1}), z^k}, \tag{1}$$

where $\frac{0}{0}$ is defined as 0. $\overline{Y}_\theta(x^n; z^k, z)$ is the difference between $\frac{\sum_{s=k+1}^n \mathbf{1}(x_n=z) I(x^{n-1}, z^k)}{\overline{I}(x^{n-1}, z^k)}$ and $\frac{\sum_{s=k+1}^n \theta\big(z|x^{n-1}\big) I(x^{n-1}, z^k)}{\overline{I}(x^{n-1}, z^k)}$. The first item is the relative frequency of z after histories that end with z^k; the second item is the average probability that θ assigns to z after histories that end with z^k. Thus, restricting attention only to the histories that end with z^k, $\overline{Y}_\theta(x^n; z^k, z)$ conveys the difference between the empirical distribution of z and the average of its probabilities.

Definition 1 *A sequence* $x^{\mathbb{N}} = (x_1, x_2, ...) \in X^{\mathbb{N}}$ *is called* θ-normal *(or a* θ-extended normal number *or* normal with respect to θ*) if*

(a) *for any* $n = 0, 1, \dots$, $\theta(x^n) > 0$*; and*

(b) *for any* $k = 0, 1, \dots$, $z^k = (z_1, ..., z_k) \in X^k$ *and every* $x \in X$

$$\lim_{n\to\infty} \overline{Y}_\theta(x^n; z^k, z) = 0,$$

whenever $\overline{I}(x^n, z^k) \to \infty$ *with* n.

Condition (a) states that in order for a sequence $x^{\mathbb{N}}$ to be normal, any prefix of $x^{\mathbb{N}}$ should have a positive probability according to θ.

There exists a normality test that corresponds to every pair (z^k, z). The one corresponding to (z^k, z) examines the difference between the relative frequency of the letter z after the appearances of the word z^k and the average probabilities that θ assigns to z after histories that end with z^k. If the difference is asymptotically zero, then $x^{\mathbb{N}}$ passes the test corresponding to (z^k, z).

Formally, if $\lim_{n\to\infty} \overline{Y}_\theta(x^n; z^k, z) = 0$ whenever the word z^k appears infinitely often in $x^{\mathbb{N}}$, then $x^{\mathbb{N}}$ passes the test corresponding to (z^k, z). Condition (b) states that in order to be normal $x^{\mathbb{N}}$ should pass all the tests that correspond to all (z^k, z).

Remark 1 *When θ is induced by an i.i.d. process, where all the digits in X are equally likely, a θ-normal sequence is known as a normal number of basis $|X|$.*

3 The Game

We define a game played by two players over a sequence of periods. Player 1's stage-action set is always X and Player 2's stage-action set is always $\Delta(X)$, the set of distributions over X. At any stage, Player 2 chooses an action and announces it. Then, Player 1 chooses an action. Both players may condition their choices on the previous actions of their opponents. In other words, Player 1's strategy is a function defined on his set of histories, $\cup_{n=0}^{\infty}\big((\Delta(X) \times X)^n \times \Delta(X)\big)$ that takes its values in his set of actions. Player 2's strategy, on the other hand, is a function defined on his set of histories, $\cup_{n=0}^{\infty}(\Delta(X) \times X)^n$ that takes its values in his set of actions. Note that a strategy of Player 2 in this sequential game provides a distribution over X after every history of Player 1's moves. Thus, it induces a stochastic process with states in X.

Denote Player 2's strategy in the sequential game by θ. Thus, θ indicates the distribution over X that Player 2 chooses after every history of Player 1's moves. The distribution $\theta(x^n)$ can be considered as the conditional distribution over the $(n+1)^{\text{th}}$ state of the process, given the history x^n. By the Kolmogorov extension theorem, θ induces a probability distribution over $X^{\mathbb{N}}$. Call this distribution also θ.

Note that any strategy of Player 1 constructs an infinite sequence of states from X, $x^{\mathbb{N}} = (x_1, x_2, ...)$, while Player 2's strategy induces a distribution

θ over $X^{\mathbb{N}}$. If the sequence $x^{\mathbb{N}}$ is normal with respect to θ, then Player 1 wins; otherwise Player 2 wins. This game is denoted as $G^{\mathbb{N}}$.

4 The Corresponding Random Variable Payoff Game

Let (Ω, μ) be a discrete probability space defined as follows: $\Omega = \cup_{k=0}^{\infty}(X^k \times X)$ and $\mu(z^k, z) = (2|X|)^{-(k+1)}$ for every $(z^k, z) \in \Omega$.[3] Note that $\sum_{(z^k,z)\in\Omega} \mu(z^k, z) = 1$. The space (Ω, μ) can be viewed as the space of the normality tests. Each pair (z^k, z) corresponds to one normality test whose probability is $\mu(z^k, z)$.

The game $G^{\mathbb{N}}$, defined in the previous section, is converted into another game, $\Gamma^{\mathbb{N}}$, with payoffs which are random variables defined over (Ω, μ). The actions of the players in $\Gamma^{\mathbb{N}}$ are identical to those in $G^{\mathbb{N}}$.

The stage payoffs of $\Gamma^{\mathbb{N}}$ depend on the players' actions as follows. Suppose that the history of Player 1's actions at time $n-1$ is x^{n-1}, Player 1's choice at time n is x_n and Player 2's action is $\theta(\cdot|x^{n-1})$ (recall, Player 2's action is a distribution over X). Denote, $x^n = (x^{n-1}, x_n)$. Then, the payoff is defined as the random variable $Y_\theta(x^n; \cdot, \cdot)$ which attains the value $Y_\theta(x^n; z^k, z)$ at the point $(z^k, z) \in \Omega$. Thus, the random variable payoff attains the value 0 at the point (z^k, z) if the word x^{n-1} does not end with the string z^k. Otherwise, it attains the value $1 - \theta\big(z|x^{n-1}\big)$ if $x^n = z$ and the value $-\theta\big(z|x^{n-1}\big)$ if $x^n \neq z$. Note that the stage payoffs not only depend on the stage actions but also on the history of Player 1's actions.

Define $\overline{Y}_\theta(x^n; z^k, z)$ as in Section 2 (see (1)); $\overline{Y}_\theta(x^n) = \overline{Y}_\theta(x^n; \cdot, \cdot)$ is the average (random variable) payoff along the history x^n. Player 1 wins the game $\Gamma^{\mathbb{N}}$ if $\overline{Y}_\theta(x^n)$ converges to 0 (the identically 0 variable)[4] and, moreover,

[3]The set X^0 is a singleton whose element represents the null history.

[4]Since the underlying probability space, (Ω, μ), is discrete, convergence in probability implies convergence in the "almost surely" sense. Thus, when we say that $\overline{Y}_\theta^n$ converges to 0, we mean it in both senses.

every history, x^n, of his actions has a positive probability according to θ. Note that if $\overline{Y}_\theta(x^n)$ converges to 0, then the sequence of Player 1's actions (called the *play path* of Player 1) is a normal number with respect to the distribution induced by Player 2's strategy. In other words, if Player 1 wins $\Gamma^{\mathbb{N}}$, he also wins $G^{\mathbb{N}}$.

The result of this paper is the following

Theorem 1 *Player 1 has a strategy*[5] *such that for any strategy θ of Player 2 the sequence of Player 1's actions, $(x_1, x_2, ...)$, is a θ-normal number. Moreover,*

a. the only information about θ needed to generate the n^{th} digit, x_n, is $\theta(\cdot|x_1, ..., x_s)$, $\quad s < n-1$, and the support of $\theta(\cdot|x_1, ..., x_{n-1})$; and

b. the algorithm is quadratic (i.e., the number of calculations it requires to compute the n^{th} digit is $O(n^2)$).

5 The Intuition behind the Construction

Suppose that Player 2 plays the strategy θ and that the actions that have been chosen by Player 1 up to time n are $(x_1, x_2, ..., x_{n-1})$. Furthermore, suppose that the action of Player 1 at time n is x_n. Denote $x^n = (x_1, x_2, ..., x_n)$. The payoff is the random variable that attains the value $\Big(1-\theta\big(z|x^{n-1}\big)\Big)I(x^{n-1}, z^k)$ at point $(z^k, z) \in \Omega$ if $x_n = z$, and the value $-\theta\big(z|x^{n-1}\big)I(x^{n-1}, z^k)$, otherwise. The average payoff up to time n is the random variable that attains the value $\overline{Y}_\theta(x^n; z^k, z)$ at point (z^k, z).

If the value of $\overline{Y}_\theta(x^n; \cdot, \cdot)$ at point (z^k, z) is positive it means that along x^n, z appears too frequently immediately after the string z^k. In other words, according to θ, z appears more often than expected after z^k. On the other hand, if $\overline{Y}_\theta(x^n; \cdot, \cdot)$ at point (z^k, z) is negative, it means that z appears too

[5] In the game $\Gamma^{\mathbb{N}}$ Player 1 is not allowed to randomize, and thus his winning strategy is *pure*.

rarely after z^k. Thus, the cumulative error that corresponds to the normality test associated with the pair (z^k, z) at time n is expressed by $\overline{Y}_\theta(x^n; z^k, z)$.

The goal of Player 1, who wants to minimize the cumulative error related to (z^k, z), is to bring $\overline{Y}_\theta(x^n; z^k, z)$ to 0. Since the same goal is common to all pairs $(z^k, z) \in \Omega$, the cumulative error at time n is expressed by the random variable $\overline{Y}_\theta(x^n; \cdot, \cdot)$. Roughly speaking, the random variable $\overline{Y}_\theta(x^n; \cdot, \cdot)$ expresses the extent to which x^n is normal.

Suppose that x_{n+1} is the choice of Player 1 at time $n+1$. The payoff at time $n+1$ is then $Y_\theta(x^{n+1}; z^k, z)$, where $x^{n+1} = (x_1, x_2, ..., x_{n+1})$. The contribution of this payoff to the cumulative error at time $n+1$, $\overline{Y}_\theta(x^{n+1}; \cdot, \cdot)$, is $\frac{Y_\theta(x^{n+1}; \cdot, \cdot)}{\overline{I}(x^n, \cdot)}$.

The essence of the construction is to find an action at time $n+1$, say, x_{n+1}, that "corrects" the cumulative error at time n in the sense that the inner product of $\frac{Y_\theta(x^{n+1}; \cdot, \cdot)}{\overline{I}(x^n, \cdot)}$ (recall that this is the contribution of the resulting payoff to the cumulative error) with the cumulative error up to time n, $\overline{Y}_\theta(x^{n+1}; \cdot, \cdot)$, is less than or equal to zero. Geometrically, it means that $\frac{Y_\theta(x^{n+1}; \cdot, \cdot)}{\overline{I}(x^n, \cdot)}$ lies in the half space opposing $\overline{Y}_\theta(x^{n+1}; \cdot, \cdot)$.

It turns out that if Player 1 at any time chooses an action that corrects the cumulative error in this way, then the cumulative error diminishes to zero. This makes his infinite sequence of choices normal with respect to the strategy of Player 2.

6 A Winning Strategy of Player 1: Constructing Normal Numbers

Proof of Theorem 1. We define the strategy of Player 1 inductively. Recall that the choice of Player 1 at any stage may depend on the previous, as well as the current, choices of Player 2. In what follows, θ is a strategy of Player 2.

Let x_1, the action in period 1, be an arbitrary element of X that satisfies $\theta(x) > 0$. Suppose that $x^n = (x_1, x_2, ..., x_n)$, the actions up to period n, have been defined so that $\theta(x_1, x_2, ..., x_n) > 0$. Let $\widehat{X} = \{z \in X;\ \ \theta(z|x_1, x_2, ..., x_n) > 0\}$. Denote $m = |\widehat{X}|$. Thus, there are m objects in X assigned positive probability by the distribution $\theta(\cdot|x_1, x_2, ..., x_n)$.

Suppose that the history of Player 1's actions up to time n is x^n. For every $z^k \in X^k$ and $z \in \widehat{X}$ define the matrix $A(z^k, z)$, with dimension $m \times m$, as follows. The entry in the row corresponding to $x' \in \widehat{X}$ and the column corresponding to $x'' \in \widehat{X}$ is denoted $a_{x',x''}$ and is defined by,

$$a_{x',x''} = I(x^n, z^k)\big(\mathbb{1}(x' = z) - \mathbb{1}(x'' = z)\big).$$

In other words, the entries of the matrix $A(z^k, z)$ are all 0 if x^n does not end with z^k. If x^n ends with z^k, then the entries in the row corresponding to z are 1 (except for the entry in the diagonal), those in the column corresponding to z are -1 (except for the entry in the diagonal), and all the rest are 0 (including the diagonal entries). Note that $A(z^k, z)$ is, in particular, a matrix of the form $(b_i - b_j)_{ij}$ for some $(b_1, ..., b_m) \in I\!R^m$.

Denote by $A(z^k, z)_{x'}$ the row of $A(z^k, z)$ corresponding to x'. That is, $A(z^k, z)_{x'}$ may be one of three types: identically 0 (in the case where x^n does not end with z^k); identically 1 except for the coordinate corresponding to x', which is 0 (in the case where x^n ends with z^k and $x' = z$); or identically 0 except for the coordinate corresponding to z, which is -1 (in the case where x^n ends with z^k and $x' \neq z$).

Recall that $Y_\theta((x^n, x'); \cdot, \cdot)$ is the payoff after the history (x^n, x'). The matrix $A(z^k, z)$ is defined so that for every pair (z^k, z),

$$Y_\theta((x^n, x'); z^k, z) = A(z^k, z)_{x'} \bullet \theta(\cdot|x_1, x_2, ..., x_n), \tag{2}$$

where $\bullet$ denotes the inner product in $I\!R^m$.

Define the matrix $\overline{A}$ as follows:

$$\overline{A} = \sum_{(z^k,z)\in\Omega \text{ and } z\in\widehat{X}} \mu(z^k, z)\frac{\overline{Y}_\theta(x^n; z^k, z)}{\overline{I}(x^n, z^k)} A(z^k, z). \tag{3}$$

The matrix $\overline{A}$ is a linear combination of the matrices $A(z^k, z)$. Recall that $A(z^k, z) = 0$ if x^n does not end with z^k. Thus, only n of the matrices $A(z^k, z)$ are not identically zero. The coefficient of the matrix $A(z^k, z)$ in (3) is the probability of the pair (z^k, z), $\mu(z^k, z)$, multiplied by the cumulative error that corresponds to it, $\overline{Y}_\theta(x^n; z^k, z)$, divided by the number of times z^k appears in the history, $\overline{I}(x^n, z^k)$. Note that the greater (in absolute value) the cumulative error the greater the coefficient, and the smaller the number of appearances of z^k the greater the coefficient.

The matrix $\overline{A} = (\overline{a}_{x',x''})$ is also of the form $(b_i - b_j)_{ij}$, as a linear combination of matrices of the same kind. Thus, there is a row in $\overline{A}$ whose entries are all non-positive[6]. Define Player 1's action at time $n+1$, x_{n+1}, to be an element in $\widehat{X}$ that corresponds to such a row. That is, $\overline{a}_{x_{n+1},x''} \leq 0$ for every x''. In particular, $\theta(x^n, x_{n+1}) > 0$ (since $x_{n+1} \in \widehat{X}$). Therefore, any convex combination of the entries in the row corresponding to x_{n+1} (i.e., a combination of $\overline{a}_{x_{n+1},x''}$, $x'' \in \widehat{X}$) is also non-positive. Specifically, a convex combination taken with respect to $\theta(x''|x_1, x_2, ..., x_n)$ is non-positive. Formally,

$$\sum_{x'' \in \widehat{X}} \overline{a}_{x_{n+1},x''} \theta(x''|x_1, x_2, ..., x_n) =$$

(4)
$$\sum_{x'' \in \widehat{X}} \Big[\sum_{(z^k,z) \in \Omega} \mu(z^k, z) \frac{\overline{Y}_\theta(x^n; z^k, z)}{\overline{I}(x^n, z^k)} a_{x_{n+1},x''}(z^k, z) \Big] \theta(x''|x_1, x_2, ..., x_n) \leq 0,$$

where $a_{x_{n+1},x''}(z^k, z)$ is the $a_{x_{n+1},x''}$ entry of the matrix $A(z^k, z)$. Using (2) we derive from (4) that

$$\sum_{(z^k,z) \in \Omega} \mu(z^k, z) \frac{\overline{Y}_\theta(x^n; z^k, z)}{\overline{I}(x^n, z^k)} Y_\theta(z^k, z, (x^n, x_{n+1})) \leq 0. \qquad (5)$$

[6]In game theoretical terminology $\overline{A}$ is a zero-sum game whose value is 0. Furthermore, $\overline{A}$ has a pure optimal strategy. That is, there is a row in $\overline{A}$ such that by playing it, Player 1 (considered in this context as a minimizer) ensures a non-positive payoff.

To simplify notation, let $Y^{n+1} = Y((x^n, x_{n+1}); z^k, z)$, $\overline{Y}^n = \overline{Y}_\theta(x^n; z^k, z)$ and $\overline{I}^n = \overline{I}(x^n, z^k)$. (Recall that all are random variables defined on (Ω, μ).) Note that Y_θ^{n+1} is bounded between -1 and 1. From (5) we obtain[7]

$$E\Big(\frac{\overline{Y}_\theta^n}{\overline{I}^n} Y_\theta^{n+1}\Big) \leq 0,$$

where the expectation is taken with respect to μ.

Continuing inductively we obtain two sequences, $Y^1, Y^2, ...$ and $\overline{I}^1, \overline{I}^2, ...$ of random variables, defined over Ω, that possess the following properties:

a. Set $\overline{I}^0 = 0$. Then, $\overline{I}^n - \overline{I}^{n-1}$ is either 0 or 1 and moreover, $\overline{I}^n - \overline{I}^{n-1} = 0$ implies $Y^{n+1} = 0$;

b. $Y^1, Y^2, ...$ are uniformly bounded;

c. $\overline{Y}^{n+1} = \frac{\overline{I}^{n-1}\overline{Y}^n + Y^{n+1}}{\overline{I}^n}$;

d. $E\big(\frac{\overline{Y}^n}{\overline{I}^n} Y^{n+1}\big) \leq 0$.

By Theorem 1 of Lehrer (2002), $\overline{Y}^n \to 0$ with θ-probability 1 on the event $\{\overline{I}^n \to \infty\}$. Since $\mu(z^k, z) > 0$ for every (z^k, z), it follows that $\overline{Y}^n = \overline{Y}_\theta(x^n; z^k, z) \to 0$ whenever $\overline{I}^n = \overline{I}(x^n, z^k) \to \infty$. Thus, we have defined a strategy of Player 1 that guarantees a win against any strategy of Player 2. We therefore conclude that if Player 2 uses θ, then $(x_1, x_2, ...)$ is θ-normal.

Note that to determine the n^{th} digit the knowledge about θ needed was the actions of Player 2 up to stage $n-1$ (that is, $\theta(\cdot|x_1, ..., x_s)$ $s < n-1$) and the support of $\theta(\cdot|x_1, ..., x_{n-1})$. If this support is always full (i.e., $\widehat{X} = X$), Player 1 does not need to know the precise action of Player 2 at time n.

During the inductive process we defined the matrix $A(z^k, z)$. This matrix is identically zero unless the word z^k appears at the end of the actual history.

[7] Note that in the following inequality the expectation operator plays the role of the inner product. This inequality expresses the fact that $\frac{Y^{n+1}}{\overline{I}^n}$ corrects the error $\overline{Y}^n$, as explained in the previous section.

Thus, the number of matrices that satisfy $A(z^k, z) \neq 0$ is linear in n. Moreover, for every pair (z^k, z) such that $A(z^k, z) \neq 0$ the number of calculations needed to find the respective values of Y_θ^{n+1}, $\overline{I}^n$ and $\overline{Y}_\theta^{n+1}$ is also linear in n. It means that the number of calculations needed at time $n+1$ is $O(n^2)$ which implies that the total number of calculations needed to calculate $(x_1, ..., x_n)$ is $O(n^3)$. ∎

7 Final Remarks

7.1 Constructing a sequence that satisfies more conditions

The proof method employed can be used to deal with countably many constraints. One, for instance, may use this method to construct a normal number, with respect to any given basis, so that its restriction to any prespecified sequence of times generated by a recursive function is normal. In other words, in addition to the countably many constraints imposed by normality, one may use the proof method to construct a sequence that satisfies more constraints, provided that the total number of constraints is countable.

7.2 About the speed of convergence

The fact that $\overline{Y}_\theta^n$ is bounded by 1 implies, by Lehrer (2002), that $\sum_{n=1}^{\infty} \frac{\|\overline{Y}_\theta^n\|^2}{n} \leq \sum_{n=1}^{\infty} \frac{1}{n^2}$. This provides information about the speed at which $\|\overline{Y}_\theta^n\|$ converges to zero.

8 References

Adler, R., M. Keane and M. Smorodinsky (1981) "A Construction of a Normal Number for the Continued Fraction Transformation," *J. of Number Theory* **13**, 95-105.

Blackwell, D. (1956) "An Analog of the MinMax Theorem For Vector Payoffs," *Pacific J. of Math.* **6**, 1-8.

Borel, E. (1909) "Les probabilités Dénombrables et Leurs Application Aritmétiques," *Rend. Circ. Mat. Palermo* **27**.

Champernowne, D. G. (1933) "The Construction of a Decimal Normal in the Scale of Ten," *J. of London Math. Soc.* **8**, 254-260.

Copeland, A. H. and P. Erdős (1946) "Note on Normal Numbers," *Bull. of Amer. Math. Soc.*, 857-860.

Davenport, H. and P. Erdős (1952) "Note on Normal Decimals," *Canadian J. of Math.* **4**, 58-63.

Dawid, A. P. (1982) "The Well Calibrated Bayesian," *J. of the Amer. Stat. Assoc.* 77(379), 605-613.

Foster, D. and R. Vohra (1997) "Calibrated Learning and Correlated Equilibrium," *Games and Economic Behavior*, 21, 40-55.

Kalai, E., E. Lehrer and R. Smorodinsky (1999) "Calibrated Forecasting and Merging," *Games and Economic Behavior* 29, 151-169.

Lehrer, E. (2002) "Approachability in Infinite Dimensional Spaces," *International Journal of Game Theory*, 31, 253-268.

Sandroni, A. and R. Smorodinsky (2002) "Belief-Based Equilibrium," *Games and Economic Behavior*, forthcoming.

Smorodinsky, M. and B. Weiss (1987) "Normal Sequences for Markov Shifts and Intrinsically Ergodic Subshifts," *Israel J. of Math.* **59**, 225-233.