INSTITUTE OF PHYSICS PUBLISHING NONLINEARITY

Nonlinearity **17** (2004) 1–10 PII: S0951-7715(04)61150-9

A version of Maurer's conjecture for stationary ψ-mixing processes

Miguel Abadi and Antonio Galves

Instituto de Matemática e Estatística, Universidade de São Paulo, Rua do Matão 1010, São Paulo SP 05508-090, Brazil

E-mail: abadi@ime.usp.br and galves@ime.usp.br

Received 19 March 2003, in final form 17 February 2004
Published
Online at stacks.iop.org/Non/17

Recommended by G Morriss

Abstract

For a stationary source with finite alphabet, let $\overline{R_n}$ be the number of non-overlapping n-blocks of symbols, occurring before the initial n-block reappears. When the source is ψ-mixing, we prove that the difference between the expectation of $\log \overline{R_n}$ and the entropy of n-blocks converges to the constant of Euler divided by $-\ln(2)$. This can be considered the correct version of a conjecture presented in Maurer (1992 *J. Cryptol.* **5** 89–105). Our theorem generalizes recent results presented in Coron and Naccache (1999 *Lecture Notes in Computer Science* vol 1556, pp 51–71), Choe and Kim (2000 *Coll. Math.* **84** 159–71) and Wegenkittl (2001 *IEEE Trans. Inform. Theory* **47** 2480–9), in the context of Markov chains. We also prove that the difference between the variance of $\log \overline{R_n}$ and the variance of the probability of n-blocks converges to an explicit constant as n diverges. The basic ingredient of the proofs is an upper-bound for the exponential approximation of the distribution of the number of non-overlapping n-blocks until a fixed but otherwise arbitrary n-block reappears. This is a new result that is interesting by itself.

Mathematics Subject Classification: 60F05, 60G10, 37A50

1. Introduction

Let $\overline{R_n}$ be the number of non-overlapping blocks of n symbols produced by a stationary ergodic source occurring before the initial n-block reappears. When the source is binary and iid, Maurer (1992) proves that

$$\lim_{n\to+\infty} \mathbb{E}(\log \overline{R_n}) - nh = \frac{-\gamma}{\ln 2},$$

0951-7715/04/010001+10$30.00 Printed in the UK

where γ is Euler's constant and h is the entropy of the source. In the same paper, he made the conjecture that this result should hold for a binary symmetric stationary ergodic source. However, an example showing that this conjecture was not true even for Markov chains has been presented in Coron and Naccache (1999).

Recently, Choe and Kim (2000) and Wegenkittl (2001) proved that a modified version of the conjecture is true for ergodic Markov chains on a finite alphabet. The modification consisted of replacing the entropy of the source by the entropy of blocks of length n.

In this paper, we prove that this corrected version of Maurer's conjecture holds in the much more general setting of ψ-mixing stationary processes on a finite alphabet. We also prove that the difference between the variance of $\log \overline{R_n}$ and the variance of the probability of n-blocks converges to an explicit constant as n diverges. Explicit upper bounds for the rates of convergence are provided for both limits. Moreover, we show that the convergence takes place exponentially fast when the rate of ψ-mixing is exponential. In particular, we show that for processes having one-dimensional marginals that are distributed uniformly on the alphabet, the expectation of $\log \overline{R_n}$ converges to a constant. Again this generalizes Maurer's result for independent binary symmetric random variables.

The main tool of our proof is an upper-bound for the exponential approximation of the distribution of $\bar{\tau}_A$, which is the number of non-overlapping n-blocks occurring until a fixed but otherwise arbitrary n-block A reappears. This is a new result that is interesting by itself. For recent results on the exponential approximation for mixing processes, we refer the reader to Galves and Schmitt (1997), Collet *et al* (1999), Haydn (1999), Hirata *et al* (1999), Abadi (2004a) and Abadi (2004b), among others. A review of the field can be found in Abadi and Galves (2001).

This paper is organized as follows. In section 2, we present the notation and the definitions and state the results. Section 3 contains the proof of the exponential approximation for $\bar{\tau}_A$. Section 4 contains the proof of the exponential approximation for $\bar{\tau}_A$ when the process is conditioned to start with the n-block A. The proof of the main result is given in section 5.

2. Notation, definitions and main result

Let $(X_m)_{m\in\mathbb{Z}}$ be a stationary stochastic process taking values on a finite alphabet $\mathcal{E}$ and defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We denote by $\mathcal{F}_I$ the σ-algebra generated by the cylinders with coordinates in I, $I \subseteq \mathbb{Z}$.

Let $\psi = (\psi(l))_{l\geqslant 0}$ be a sequence decreasing to zero. The process $(X_m)_{m\in\mathbb{Z}}$ is called *ψ-mixing* if

$$\sup_{n\in\mathbb{N}, B\in\mathcal{F}_{\{0,..,n\}}, C\in\mathcal{F}_{\{n+l+j,\, j\geqslant 1\}}} \frac{|\mathbb{P}(B\cap C) - \mathbb{P}(B)\mathbb{P}(C)|}{\mathbb{P}(B)\mathbb{P}(C)} = \psi(l),$$

where the supremum is taken only for sets B and C, such that $\mathbb{P}(B)\mathbb{P}(C) > 0$. The sequence $\psi(l)$ is called the mixing rate of the process.

Let n be a fixed positive integer. The n-block process $(\bar{X}_k)_{k\in\mathbb{Z}}$, taking values on the set of n-blocks $\mathcal{E}^n$, is defined by

$$\bar{X}_k := (X_{kn}, \ldots, X_{(k+1)n-1}).$$

We remark that if the process $(X_k)_{k\in\mathbb{Z}}$ is ψ-mixing, then the n-block process $(\bar{X}_k)_{k\in\mathbb{Z}}$ is also ψ-mixing, with mixing rate $\overline{\psi}(\ell) = \psi(\ell n)$.

Given a n-block $A = (a_0, \ldots, a_{n-1}) \in \mathcal{E}^n$, we define $\bar{\tau}_A$, the first *hitting time* of A after the origin, as follows:

$$\bar{\tau}_A = \inf\{k \geqslant 1 : \bar{X}_k = A\}\,.$$

We recall that mixing (at any rate) implies ergodicity, and ergodicity implies that $\bar{\tau}_A$ is almost surely finite (see, e.g. Cornfeld *et al* (1982)).

We define $\overline{R_n}$ as the first time after the origin that the first n-block reappears, namely,

$$\overline{R_n} = \inf\{k \geqslant 1 : \bar{X}_k = \bar{X}_0\}.$$

We define the entropy of the n-blocks as

$$H_n = -\sum_{A \in \mathcal{E}^n} \mathbb{P}\{\bar{X}_0 = A\} \log \mathbb{P}\{\bar{X}_0 = A\}.$$

The entropy of the process is defined as the limit

$$h = \lim_{n \to +\infty} \frac{1}{n} H_n.$$

The following quantities will appear in the statement of our main theorem. We define

$$\Xi = \lim_{\epsilon \to \infty} \int_{\epsilon}^{\infty} \mathrm{e}^{-t} \log t \, \mathrm{d}t$$

and

$$\Delta = \lim_{\epsilon \to 0} [\omega(\epsilon) - (\log \epsilon)^2 + 2\Xi \log \epsilon],$$

where

$$\omega(\epsilon) = \epsilon \sum_{k \geqslant 1} (\log k)^2 \, \mathrm{e}^{-\epsilon(k-1)}.$$

We will call $\mathbb{P}\{\bar{X}_0\}$ the random variable that takes the value $\mathbb{P}\{\bar{X}_0 = A\}$ on the set $\{\bar{X}_0 = A\}$. We can now state the main result of this paper.

Theorem 1. *Let $(X_m)_{m \in \mathbb{Z}}$ be a ψ-mixing process. Then the following inequality holds for all $n \in \mathbb{N}$.*

$$|\mathbb{E}(\log \overline{R_n}) - H_n - \Xi| \leqslant C\varepsilon_n, \tag{1}$$

where $\varepsilon_n = \max\{\sqrt{\psi(n)}, \, \mathrm{e}^{-cn}\}$, and $C > 0, \; c > 0$ are suitable constants. Moreover,

$$|\mathrm{Var}(\log \overline{R_n}) - \mathrm{Var}(\log \mathbb{P}\{\bar{X}_0\}) - \Delta + \Xi^2| \leqslant \varepsilon_n.$$

Remark. In Maurer (1992) and Wegenkittl (2001) the constants Ξ and Δ were denoted C and D, respectively. We found it more comfortable to use C as a generic constant in different computations and to use Ξ and Δ to denote the specific constants appearing in theorem 1. In the same papers, $\omega(\epsilon)$ was defined as

$$\omega(\epsilon) = \epsilon \sum_{k \geqslant 1} (\log k)^2 (1 - \epsilon)^{k-1}.$$

It is a simple exercise to pass from this definition to ours. Our definition is natural since the basis of our proof is an exponential approximation. The same remark explains the definition of the function $\nu(\epsilon)$, which will appear in the proof of theorem 1.

Theorem 1 implies the following corollary for Markov chains.

Corollary 2. *Let $(X_m)_{m \in \mathbb{Z}}$ be an ergodic Markov chain taking values on $\mathcal{E}$ with entropy h. Then there exist two positive constants C, c such that*

$$|\mathbb{E}(\log \overline{R_n}) - (n-1)h - H_1 - \Xi| \leqslant C \, \mathrm{e}^{-cn}.$$

The main ingredients of the proof of theorem 1 are theorems 3 and 4. Theorem 3 gives an exponential approximation for the first time after the origin the process hits a fixed n-block. Theorem 4 proves the same result for the process conditioned to start in the fixed n-block. We emphasize that both theorems hold for *any* fixed n-block. In the rest of the paper, for shorthand notation and whenever it is clear, we will write $\rho(A)$ instead of $\mathbb{P}\{\bar{X}_j = A\}$ for any $j \in \mathbb{N}$. This is justified since the process is stationary.

Theorem 3. *Let the process $(X_m)_{m\in\mathbb{Z}}$ be ψ-mixing and let A be an n-block, with n a positive integer. Then, the following inequality holds.*

$$|\mathbb{P}\{\bar{\tau}_A > t\} - \mathrm{e}^{-\rho(A)t}| \leqslant C\ \varepsilon(A)\ f(A,t),$$

where C is a positive constant, $\varepsilon(A) = \rho(A)^\beta$, $0 < \beta = \beta(d)$, if $\psi(\ell) \leqslant 1/\ell^d$, for all $\ell \geqslant \ell_0$, with $d > 0$, and $\varepsilon(A) = \sqrt{\psi(n)}$, otherwise. Moreover, $f(A,t) = \mathrm{e}^{-C_1\rho(A)t}$, where C_1 is another positive constant.

Given an n-block A such that $\rho(A) > 0$, we denote $\mathbb{P}_A$ the conditional probability given the event $\{\bar{X}_0 = A\}$.

Theorem 4. *Let the process $(X_m)_{m\in\mathbb{Z}}$ be ψ-mixing. Then, for all $n \in \mathbb{N}$, all A and all $t > 0$, the following inequality holds*

$$|\mathbb{P}_A\{\bar{\tau}_A > t\} - \mathrm{e}^{-\rho(A)t}| \leqslant C\ [\varepsilon(A) + \psi(n)]\ f(A,t),$$

where $f(A,t)$ and $\varepsilon(A)$ are the same as in the preceding theorem.

Remark. We will refer to $P\{\bar{\tau}_A > t\}$ has the hitting time $\bar{\tau}_A$ and to $Q\{\bar{\tau}_A > t\}$ has the return time $\bar{\tau}_A$. Theorem 3 states that the hitting time $\bar{\tau}_A$ is exponentially distributed with parameter $\rho(A)$ for *every* n-block A. Theorem 4 states the same result for the *return time*. In the literature (cf Abadi and Galves (2001) and references therein), it is considered a similar quantity: the hitting time

$$\tau_A = \inf\{k \geqslant 1 : (X_k, \ldots, X_{k+n-1}) = A\},$$

that is, the first occurrence of the (fixed) n-block A over the *whole* process and not *block by block* as $\bar{\tau}_A$ is defined. Similarly, the return time is defined over the whole process, when the process is conditioned to start with the fixed n-block A. We remark that for the hitting time τ_A, the exponential approximation holds but with parameter $\lambda(A)\rho(A)$, where $\lambda(A)$ is a positive uniformly bounded factor. See Galves and Schmitt (1997) for the first result on this subject. Abadi (2001) provides an example for which $\lambda(A) \neq 1$. He also gives an explicit computable expression for it. Moreover, Abadi (2004b) shows that the return time, τ_A, can have even a law that is not exponential. Thus far, our results establish the different behaviour between the quantities $\bar{\tau}_A$ and τ_A.

3. Poof of theorem 3

Since the n-block process $(\bar{X}_k)_{k\in\mathbb{Z}}$ is ψ-mixing, it follows from theorem 1 in Abadi (2004a) that for all $t > 0$ the following inequality holds:

$$|\mathbb{P}\{\bar{\tau}_A > t\} - \mathrm{e}^{-\overline{\xi_A}\rho(A)t}| \leqslant C\delta(A)\,\mathrm{e}^{-(\Xi/2)\rho(A)t}, \tag{2}$$

where $\delta(A) \to 0$ when $n \to 0$ and $\overline{\xi_A} \in [\Xi_1, \Xi_2]$ and $0 < \Xi_1 < 1 \leqslant \Xi_2 < +\infty$ are constants independent of A, n and t. The same theorem implies that $\delta(A) < \varepsilon(A)$.

We remark that, by the mean value theorem,

$$|\mathrm{e}^{-\rho(A)t} - \mathrm{e}^{-\overline{\xi_A}\rho(A)t}| \leqslant |1 - \overline{\xi_A}|\rho(A)t\,\mathrm{e}^{-\Xi_1\rho(A)t}.$$

Therefore, to prove theorem 3, it is enough to show that $|1 - \overline{\xi_A}| \leqslant \varepsilon(A)$. Let us define

$$\overline{\lambda_A} = \frac{-\log \mathbb{P}\{\bar{\tau}_A > f_A\}}{f_A \rho(A)} \qquad \text{with} \quad f_A = \frac{1}{\rho(A)^\alpha},$$

for a suitable $\alpha \in (0, 1)$. It has been proved in Galves and Schmitt (1997), for a summable ψ, and in Abadi (2001), for any ψ, that the following uniform bound holds:

$$\sup_t |\mathbb{P}\{\bar{\tau}_A > t\} - \mathrm{e}^{-\overline{\lambda_A}\rho(A)t}| \leqslant C\varepsilon(A), \tag{3}$$

where $\overline{\lambda_A} \in [\Lambda_1, \Lambda_2] \subset (0, \infty)$, and $\Lambda_1 < 1 \leqslant \Lambda_2$ are constants independent of A, n and t. Inequalities (2) and (3) together imply that

$$\sup_t |\mathrm{e}^{-\overline{\xi_A}\rho(A)t} - \mathrm{e}^{-\overline{\lambda_A}\rho(A)t}| \leqslant C\varepsilon(A).$$

Taking $t = 1/\rho(A)$, we conclude that $|\overline{\xi_A} - \overline{\lambda_A}| \leqslant C\varepsilon(A)$. Taking $C_1 = \Xi_1/2$, it is enough to show that $|\overline{\lambda_A} - 1| \leqslant C\rho(A)^\beta$ to conclude the proof. This is done in the next lemma.

Lemma 5. *Let the process $(X_m)_{m\in\mathbb{Z}}$ be ψ-mixing. Then there exist positive constants C and β such that for all $n \in \mathbb{N}$ and all $A \in \mathcal{C}_n$ the following inequality holds:*

$$|\overline{\lambda_A} - 1| \leqslant C\rho(A)^\beta.$$

Proof. We will first prove that

$$\left| \frac{\mathbb{P}\{\bar{\tau}_A \leqslant f_A\}}{f_A \rho(A)} - \overline{\lambda_A} \right| \leqslant C\rho(A)^\beta. \tag{4}$$

By Taylor's expansion, we have the relation that

$$1 - \mathrm{e}^{-\theta} \leqslant \theta \leqslant 1 - \mathrm{e}^{-\theta} + 2(1 - \mathrm{e}^{-\theta})^2 \tag{5}$$

for all $\theta \in [0, 1]$. Put

$$\theta = -\log \mathbb{P}\{\bar{\tau}_A > f_A\}. \tag{6}$$

Clearly, $\theta \geqslant 0$. Moreover, $\theta \leqslant 1$ if and only if $\mathbb{P}\{\bar{\tau}_A \leqslant f_A\} \leqslant 1 - e^{-1}$. To check this last inequality, we observe that

$$\mathbb{P}\{\bar{\tau}_A \leqslant f_A\} \leqslant \sum_{i=1}^{[f_A]} \mathbb{P}\{\bar{X}_i = A\} \leqslant f_A \rho(A).$$

Since $\rho(A)$ converges to 0 as n diverges, it follows that $f_A\rho(A) \leqslant 1 - e^{-1}$ for all n large enough. Therefore, replacing (6) in inequalities (5), we deduce expression (4).

We will now prove that

$$\left| 1 - \frac{\mathbb{P}\{\bar{\tau}_A \leqslant f_A\}}{f_A \rho(A)} \right| \leqslant C\rho(A)^\beta. \tag{7}$$

For any integer $j \geqslant 1$, write

$$\mathbb{P}\{\bar{\tau}_A = j\} = \mathbb{P}\{\bar{\tau}_A > j - 1\} - \mathbb{P}\{\bar{\tau}_A > j\}. \tag{8}$$

By stationarity,

$$\mathbb{P}\{\bar{\tau}_A > j - 1\} = \mathbb{P}\{\bar{X}_i \neq A;\ 2 \leqslant i \leqslant j\}. \tag{9}$$

Expressions (8) and (9) together imply that

$$\mathbb{P}\{\bar{\tau}_A = j\} = \mathbb{P}\{\bar{X}_0 = A,\ \bar{\tau}_A > j - 1\}$$

for all positive integer j. Using the above equality, we have the relation that for any $j \geqslant 1$

$$\rho(A) - \mathbb{P}\{\bar{\tau}_A = j\} = \mathbb{P}\{\bar{X}_0 = A, \bar{\tau}_A \leqslant j-1\} \leqslant \sum_{i=1}^{j-1} \mathbb{P}\{\bar{X}_0 = A, \bar{X}_i = A\}.$$

Using the ψ-mixing property of the process, this last expression is bounded above by

$$\sum_{i=0}^{j-1} \rho(A)^2(1+\psi(in)).$$

Using this upper-bound we obtain immediately the inequality

$$|\,[f_A]\rho(A) - \mathbb{P}\{\bar{\tau}_A \leqslant f_A\}| \leqslant \sum_{j=1}^{[f_A]} \sum_{i=0}^{j-1} \rho(A)^2(1+\psi(in)),$$

where $[x]$ stands for the integer part of $x \in \Re$. This last expression is bounded trivially above by $\leqslant C {f_A}^2 \rho(A)^2$. From this and the trivial fact that $1/[f_A] - 1/f_A \leqslant 1/f_A$, inequality (7) follows. This concludes the proof of the lemma.

4. Proof of theorem 4

By the triangle inequality we have

$$|\mathbb{P}_A\{\bar{\tau}_A > t\} - \mathrm{e}^{-\rho(A)t}| \leqslant |\mathbb{P}_A\{\bar{\tau}_A > t\} - \mathbb{P}\{\bar{\tau}_A > t\}| + |\mathbb{P}\{\bar{\tau}_A > t\} - \mathrm{e}^{-\rho(A)t}|. \tag{10}$$

To obtain an upper-bound for the first term of the right-hand side of the above inequality, we use the following triangle inequality, which holds for any $t \geqslant 2$:

$$\begin{aligned}
&|\mathbb{P}_A\{\bar{\tau}_A > t\} - \mathbb{P}\{\bar{\tau}_A > t\}| \\
&\qquad = |\mathbb{P}_A\{\bar{X}_k \neq A\ ;\ 1 \leqslant k \leqslant t\} - \mathbb{P}\{\bar{X}_k \neq A\ ;\ 1 \leqslant k \leqslant t\}| \\
&\qquad \leqslant |\mathbb{P}_A\{\bar{X}_k \neq A\ ;\ 1 \leqslant k \leqslant t\} - \mathbb{P}_A\{\bar{X}_k \neq A\ ;\ 2 \leqslant k \leqslant t\}| && (11)\\
&\qquad\quad + |\mathbb{P}_A\{\bar{X}_k \neq A\ ;\ 2 \leqslant k \leqslant t\} - \mathbb{P}\{\bar{X}_k \neq A\ ;\ 2 \leqslant k \leqslant t\}| && (12)\\
&\qquad\quad + |\mathbb{P}\{\bar{X}_k \neq A\ ;\ 2 \leqslant k \leqslant t\} - \mathbb{P}\{\bar{X}_k \neq A\ ;\ 1 \leqslant k \leqslant t\}|. && (13)
\end{aligned}$$

Term (11) is bounded using twice the mixing property by

$$\begin{aligned}
&\mathbb{P}_A\{\bar{X}_1 = A, \bar{X}_k \neq A; 2 \leqslant k \leqslant t\} \leqslant \rho(A)\psi(0)(1+\psi(0))\mathbb{P}\{\bar{X}_k \neq A; 2 \leqslant k \leqslant t\} \\
&\qquad = C\rho(A)\mathbb{P}\{\bar{\tau}_A > t-1\}. && (14)
\end{aligned}$$

Term (12) is also bounded using the mixing property by

$$\psi(n)\mathbb{P}\{\bar{\tau}_A > t-1\}. \tag{15}$$

As for (11), term (13) can be bounded by

$$\begin{aligned}
&\mathbb{P}\{\bar{X}_1 = A, \bar{X}_k \neq A; 2 \leqslant k \leqslant t\} \leqslant \rho(A)(1+\psi(0))\mathbb{P}\{\bar{X}_k \neq A; 2 \leqslant k \leqslant t\} \\
&\qquad = C\rho(A)\mathbb{P}\{\bar{\tau}_A > t-1\}. && (16)
\end{aligned}$$

A similar upper-bound holds for $t = 1$. To prove it, we first observe that

$$|\mathbb{P}_A\{\bar{\tau}_A > 1\} - \mathbb{P}\{\bar{\tau}_A > 1\}| = |\mathbb{P}_A\{\bar{\tau}_A = 1\} - \mathbb{P}\{\bar{\tau}_A = 1\}|.$$

Then we use the the mixing property to conclude that

$$|\mathbb{P}_A\{\bar{\tau}_A = 1\} - \mathbb{P}\{\bar{\tau}_A = 1\}| \leqslant \psi(0)\rho(A). \tag{17}$$

Since $\mathbb{P}\{\bar{\tau}_A > 0\} = 1$, this last inequality is the upper-bound we wanted for $t = 1$. Inequalities (14), (15), (16) and (17) together imply that

$$|\mathbb{P}_A\{\bar{\tau}_A > t\} - \mathrm{e}^{-\rho(A)t}| \leqslant C(\psi(n) + \rho(A))\mathbb{P}\{\bar{\tau}_A > t-1\},$$

where C is a positive constant. To conclude the proof of theorem 4, it is enough to apply theorem 3 to the right-hand side of inequality (10).

5. Proof of theorem 1

By definition,

$$\mathbb{E}(\log \overline{R_n}) = \sum_{t\geqslant 1} \log t \; \mathbb{P}\{\overline{R_n} = t\} = \sum_{t\geqslant 1} (\log(t+1) - \log t)\mathbb{P}\{\overline{R_n} > t\}.$$

Also, by definition

$$\mathbb{P}\{\overline{R_n} > t\} = \sum_{A\in\mathcal{C}_n} \rho(A)\mathbb{P}_A\{\bar{\tau}_A > t\}.$$

Thus

$$\mathbb{E}(\log \overline{R_n}) = \sum_{A} \rho(A) \sum_{t\geqslant 1} (\log(t+1) - \log t)\mathbb{P}_A\{\bar{\tau}_A > t\}. \tag{18}$$

We define $\nu(\epsilon) = \epsilon \sum_{k\geqslant 1} \log k \, \mathrm{e}^{-\epsilon(k-1)}$. We have the relation that $\Xi = \lim_{\epsilon\to\infty}[\nu(\epsilon) + \log\epsilon]$. Therefore,

$$\sum_{A} \rho(A) \sum_{t\geqslant 1} (\log(t+1) - \log t)\, \mathrm{e}^{-\rho(A)t} = \sum_{A} (1 - \mathrm{e}^{-\rho(A)})\nu(\rho(A)). \tag{19}$$

By the triangle inequality,

$$|\mathbb{E}(\log \overline{R_n}) - H_n - \Xi| \leqslant \mathrm{I} + \mathrm{II} + \mathrm{III},$$

where the terms I, II and III are defined below.

In view of (18), we define I. We use the mean value theorem and theorem 4 to get an upper bound for it. Define $\varepsilon_n = \max_{A\in\mathcal{E}} \varepsilon(A)$. Therefore,

$$\begin{aligned}
\mathrm{I} &:= \left| \sum_{A} \rho(A) \sum_{t\geqslant 1} (\log(t+1) - \log t)[\mathbb{P}_A\{\bar{\tau}_A > t\} - \mathrm{e}^{-\rho(A)t}] \right| \\
&\leqslant \sum_{A} \rho(A) \sum_{t\geqslant 1} \frac{1}{t} \, |\mathbb{P}_A\{\bar{\tau}_A > t\} - \mathrm{e}^{-\rho(A)t}| \\
&\leqslant \sum_{A} \rho(A) \sum_{t\geqslant 1} \frac{1}{t} (\varepsilon(A) + \psi(n)) f(A,t) \\
&\leqslant C \; (\varepsilon_n + \psi(n)).
\end{aligned} \tag{20}$$

To define II, we use identity (19). An upper-bound is obtained with straightforward computations.

$$\begin{aligned}
\mathrm{II} &:= \left| \sum_{A} \rho(A) \left[\frac{1 - \mathrm{e}^{-\rho(A)t}}{\rho(A)} \; \nu(\rho(A)) \; - \nu(\rho(A)) \right] \right| \\
&\leqslant \sum_{A} \rho(A) \left| \frac{1 - \mathrm{e}^{-\rho(A)t} - \rho(A)}{\rho(A)} \; \nu(\rho(A)) \right| \\
&\leqslant \sum_{A} \rho(A) \; 2\rho(A)\nu(\rho(A)) \\
&\leqslant C\varepsilon_n.
\end{aligned} \tag{21}$$

Finally,

$$\begin{aligned}\mathrm{III} &:= \left|\sum_A \rho(A)\nu(\rho(A)) - H_n - \Xi\right| \\ &= \left|\sum_A \rho(A)[\nu(\rho(A)) + \log\rho(A) - \Xi]\right| \\ &\leqslant \max_A |\nu(\rho(A)) + \log\rho(A) - \Xi| \\ &\leqslant C\,\mathrm{e}^{-cn}.\end{aligned}$$

Thus, we have established (1). Further,

$$\begin{aligned}|\mathrm{Var}(\log\overline{R_n}) - \mathrm{Var}(\log\mathbb{P}\{X_1^n\}) - \Delta + \Xi^2| &\leqslant \left|\mathbb{E}((\log\overline{R_n})^2) - \sum_A \rho(A)\omega(\rho(A))\right| \\ &+ \left|\sum_A \rho(A)\omega(\rho(A)) - \sum_A \rho(A)(\log\rho(A))^2 - 2\Xi H_n - \Delta\right| \\ &+ |H_n^2 + 2\Xi H_n + \Xi^2 - (\mathbb{E}(\log\overline{R_n}))^2|\,.\end{aligned}$$

In the right-hand side of the above inequality, the first term is bounded by the same arguments that were used in (20) and (21). The bound for the second one is obtained using the definition of Δ. The upper-bound of the third one is a consequence of inequality (1). This ends the proof of the theorem.

Proof of corollary 2. We follow the ideas of Wegenkittl's paper. Let $A = (a_0, \ldots, a_{n-1}) \in \mathcal{E}^n$. We first recall that for Markov chains one has $\psi(n) \leqslant C\,\mathrm{e}^{-cn}$. Second, we observe that

$$\rho(A) = \mathbb{P}\{X_0 = a_0\}\prod_{i=1}^{n-1}\mathbb{P}\{X_i = a_i | X_{i-1} = a_{i-1}, \ldots, X_0 = a_0\}.$$

Define

$$\begin{aligned}F_n &= H_n - H_{n-1} \\ &= \sum_{a_{n-1},\ldots,a_0} \mathbb{P}\{X_{n-1} = a_{n-1}, \ldots, X_0 = a_0\}\log\mathbb{P}\{X_{n-1} = a_{n-1} | X_{n-2} = a_{n-2}, \ldots, X_0 = a_0\}\end{aligned}$$

for all positive integers n, with $H_0 = 0$. Then it is enough to observe that inequality (1) can be rewritten as

$$\left|\mathbb{E}(\log\overline{R_n}) - \sum_{i=1}^n F_i - \Xi\right| \leqslant \varepsilon_n$$

to conclude the proof.

6. Final remarks

Remark 1. Maurer's conjecture was formulated originally using logarithms of base 2, which is a natural choice when we consider binary sources. We recall that using base 2 for the logarithms, the values of the constants considered in this paper are $\Xi = -0.832\,746\cdots = -\gamma/\ln 2$, where γ is Euler's constant and $\Delta = 4.117\,181\ldots$.

Remark 2. The return time, $\overline{R_n}$, is defined considering the non-overlapping *adjacent* observations $(X_{kn}, \ldots, X_{kn+n-1}); k \geqslant 1$. Now, consider n-blocks that are non-overlapping but are not necessarily adjacent. Let us fix a positive integer M and define the process

$$(X_k^M)_{k\in\mathbb{N}} \qquad \text{with } X_k^M = (X_{kMn}, \ldots, X_{(kMn)+n-1})$$

with the corresponding return time

$$R_n^M = \inf\{k \geqslant 1 : (X_{kMn}, \ldots, X_{(kMn+n-1)} = (X_0, \ldots, X_{n-1})\}.$$

As with to $\bar{\tau}_A$, we can define τ_A^M. Clearly, theorems 3 and 4 hold for τ_A^M and so does theorem 1 too. The larger M is, the more mixing is the M-process $(X_k^M)_{k\in\mathbb{N}}$, and the limits in theorem 1 (and theorems 3 and 4) take place faster. Statistically speaking, this means that considering R_n^M has the disadvantage that we need a larger sample for observing the repetition time, but when this is not a problem, it has the advantage that the convergence given by theorem 1 for R_n^M is faster than the convergence for $\overline{R_n}$.

Remark 3. An important result in ergodic theory is the famous Kac's lemma (e.g. Kac (1947)), which states that for an ergodic system the expected *return* time to a measurable set with positive measure is the inverse of the measure of the event. We present in the next corollary an estimation for all the moments of the non-overlapping hitting and return times. This a consequence of the bounds given in theorems 3 and 4. Roughly speaking, if the hitting (or return) time distribution is close to an exponential distribution with parameter $\rho(A)$, we have the relation that $k!/\rho(A)^k \approx \mathbb{E}(\tau_A^k)$.

Corollary 6. *Let $(X_m)_{m\in\mathbb{Z}}$ be a ψ-mixing process. Let $k \in \mathbb{N}$. Then the following inequalities hold:*

$$|\rho(A)^k \mathbb{E}(\bar{\tau}_A^k) - k!| \leqslant C_k \varepsilon(A)$$

and

$$|\rho(A)^k \mathbb{E}_A(\bar{\tau}_A^k) - k!| \leqslant C_k\, [\varepsilon(A) + \psi(n)],$$

where $\varepsilon(A)$ is the same as in theorem 3, and C_k is a positive constant that only depends on k.

Proof. We recall that for a random variable X with exponential distribution with parameter λ, we have $\mathbb{E}(X^k) = k!/\lambda^k$. Now we use the inequality

$$\begin{aligned} |\mathbb{E}(X^k) - \mathbb{E}(Y^k)| &= \left| \sum_{t\geqslant 0} ((t+1)^k - t^k)\mathbb{P}\{X > t\} - \sum_{t\geqslant 0} ((t+1)^k - t^k)\mathbb{P}\{Y > t\} \right| \\ &\leqslant \sum_{t\geqslant 0} k(t+1)^{k-1} |\mathbb{P}\{X > t\} - \mathbb{P}\{Y > t\}| \end{aligned}$$

for any pair of positive random variables X, Y. Now, apply the above inequality with X exponentially distributed and $Y = \bar{\tau}_A$. The exponential decay of the function $f(A, t)$ in the error term in theorem 3 ends the proof of the corollary.

Acknowledgments

We are indebted to Pablo Ferrari and Davide Gabrielli for many interesting discussions on the subject. Special thanks to Davide Gabrielli for calling our attention to Wegenkittl's paper. We

thank Pierre Collet and Bernard Schmitt, who pointed out a mistake in a previous version of theorem 1. The authors thank FAPESP, FINEP, CNPq and a USP-COFECUB agreement for support. MA thanks the Centre de Physique Théorique, CNRS, Luminy and Université de Toulon et du Var for the hospitality during the preparation of this paper.

Work done within the Projeto Temático 'Rhythmic patterns, parameter setting and language change', supported by FAPESP, grant 98/3382-0, and as part of the activities of the Núcleo de Excelência 'Critical phenomena in probability and stochastic processes' grant 41.96.0923.00. This research was supported partially by the USP-COFECUB agreement. MA is supported by FAPESP. AG is supported partially by CNPq, grant 301301/79.

References

Abadi M 2001 Exponential approximation for hitting times in mixing processes *Math. Phys. Electr. J.* **7**

Abadi M 2004a Sharp error terms and necessary conditions for exponential hitting times in mixing processes *Ann. Probab.* at press http://www.ime.usp.br/~abadi

Abadi M 2004b Statistics and error terms of occurrence times in mixing processes submitted http://www.ime.usp.br/~abadi

Abadi M and Galves A 2001 Inequalities for the occurrence times of rare events in mixing processes. The state of the art *Markov Proc. Relat. Fields.* **7** 97–112

Choe G and Kim D 2000 Average convergence rate of the first return time *Coll. Math.* **84** 159–71

Collet P, Galves A and Schmitt B 1999 Repetition times for Gibbsian sources *Nonlinearity* **12** 1225–37

Cornfeld I, Fomin S and Sinai Y 1982 *Ergodic theory, Grundlehren der Mathematischen Wissenschaften* vol 245 (New York: Springer)

Coron J and Naccache D 1999 An accurate evaluation of Maurer's universal test. Selected areas in cryptography *Lecture Notes in Computer Science* vol 1556, pp 51–71

Galves A and Schmitt B 1997 Inequalities for hitting times in mixing dynamical systems *Random Comput. Dyn.* **5** 337–48

Haydn N 1999 The distribution of the first return time for rational maps *J. Stat. Phys.* **94** 5–6

Haydn N 1999 The distribution of the first return time for rational maps *J. Stat. Phys.* **94** 1027–36

Hirata M, Saussol B and Vaienti S 1999 Statistics of return times: a general framework and new applications *Commun. Math. Phys.* **206** 33–55

Kac M 1947 On the notion of recurrence in discrete stochastic processes *Bull. Am. Math. Soc.* **53** 1002–10

Maurer U 1992 A universal test for random bit generators *J. Cryptol.* **5** 89–105

Wegenkittl S 2001 Entropy estimators and serial tests for ergodic chains *IEEE Trans. Inform. Theory* **47** 2480–9

Summary of Comments on non161150

Page: 4

Sequence number: 1
Author:
Date: 4/12/2004 12:28:15 PM
Type: Highlight
We will refer to

Sequence number: 2
Author:
Date: 4/12/2004 12:28:28 PM
Type: Note
Au: Please check if this sentence needs rephrasing

Page: 10

Sequence number: 1
Author:
Date: 4/12/2004 1:24:41 PM
Type: Highlight
The authors thank

Sequence number: 2
Author:
Date: 4/12/2004 1:25:06 PM
Type: Note
Au: Editing changes have been made to this sentence, please ensure the meaning is still as intended

Sequence number: 3
Author:
Date: 4/12/2004 1:26:22 PM
Type: Highlight
Abadi M

Sequence number: 4
Author:
Date: 4/12/2004 1:26:40 PM
Type: Note
Au: Please provide page range

Sequence number: 5
Author:
Date: 4/12/2004 1:28:33 PM
Type: Highlight
Abadi M 2004b

Sequence number: 6
Author:
Date: 4/12/2004 1:28:41 PM
Type: Note
Au: Please update