Kim, D.-H.
Osaka J. Math.
40 (2003), 171–186

THE RECURRENCE OF BLOCKS FOR BERNOULLI PROCESSES

DONG HAN KIM

(Received July 25, 2000)

1. Introduction

Convergence of the logarithm of the first return time normalized by the block length n has been investigated in relation to data compression methods such as Ziv-Lempel algorithms [11]. For each sample sequence $x = (x_1, x_2, \ldots)$ from an ergodic stationary information source, define $P_n(x)$ to be the probability of the initial n-block in x, i.e., $P_n(x) = \Pr(x_1 \cdots x_n)$. The classical Shannon-Breiman-McMillan Theorem states that $-(\log P_n)/n$ converges to entropy in L^1 and almost surely. Throughout the article, log denotes the logarithm with respect to base 2 and ln denotes the natural logarithm.

DEFINITION 1.1. Given a block size n, the first return time R_n is defined by

$$R_n(x) = \min\{j \geq 1 : x_1 \cdots x_n = x_{j+1} \cdots x_{j+n}\}.$$

Kac's Lemma [3] states that $E[R_n \mid x_1 \ldots x_n = B] = 1/\Pr(B)$. This suggests that $R_n(x)$ is close to $1/P_n(x)$, hence we expect that $(\log R_n)/n$ converges to entropy h in a suitable sense. It was proved that $(\log R_n)/n$ converges to entropy in probability by Wyner and Ziv [8] and almost surely by Ornstein and Weiss [6]. For a comprehensive introduction to the subject consult Shields [7] and the references therein. For the application to the testing pseudorandom numbers, see [2]. Recently several interesting results have been obtained regarding convergence rates by other investigators for R_n and related concepts such as the longest match-length, the waiting time and the redundancy rate, etc. See [4], [10]. In this article we investigate the relation between first return time and entropy for a Bernoulli process. Since the formula contains a correction terms, it approximates the entropy very well. See the last section for simulations.

In his Ph.D thesis [9] A.J. Wyner discovered that for a stationary aperiodic Markov chain with entropy h we have a second order limit law:

$$\lim_{n\to\infty} \Pr\left(\frac{\log R_n - nh}{\sigma\sqrt{n}} \leq \alpha\right) = \Phi(\alpha)$$

2000 *Mathematics Subject Classification* : Primary 94A17, secondary 37M25.

where

$$\Phi(\alpha) = \int_{-\infty}^{\alpha} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx$$

and

$$\sigma^2 = \lim_{n\to\infty} \frac{\mathrm{Var}(-\log P_n(x))}{n}.$$

Kontoyiannis ([4, Corollary 1]) showed that for any $\beta > 0$

$$\log[R_n(x)P_n(x)] = o(n^\beta)$$

almost surely for mixing Markov chains. Later A.J. Wyner ([10, Corollary B5]) proved that for any $\epsilon > 0$

$$-(1+\epsilon)\cdot \log n \le \log[R_n(x)P_n(x)] \le \log\log n$$

eventually, almost surely for mixing Markov chains. Hence we have

$$-(1+\epsilon)\cdot \log n \le E[\log R_n P_n] \le \log\log n$$

approximately for large n. In this paper we investigate the speed of convergence of the average of $\log R_n$ to entropy. Now we state the main theorem.

Theorem 1.2. *For a Bernoulli process with entropy h, we have*

$$\lim_{n\to\infty} E[\log(R_n P_n)] = -\frac{\gamma}{\ln 2}$$

and

$$\lim_{n\to\infty} E[\log R_n] - n\cdot h = -\frac{\gamma}{\ln 2}.$$

Maurer [5] studied the *non-overlapping first return time*

$$R'_n(x) \equiv \min\{j \ge 1 : x_1 \cdots x_n = x_{jn+1}\cdots x_{jn+n}\}.$$

In computing $R'_n(x)$ we need approximately n times more digits of x than $R_n(x)$. So the overlapping algorithm is efficient compared to the non-overlapping one.

DEFINITION 1.3. For $0 < r < 1$, define

$$v(r) \equiv r\sum_{i=1}^{\infty}(1-r)^{i-1}\log i.$$

Put $r = 2^{-n}$. Then the expectation of $\log R'_n$ equals $v(r)$ in case of the Bernoulli $(1/2, 1/2)$-process. Note that

$$\begin{aligned}\lim_{r\to 0+}[v(r) + \log r] &= \lim_{s\to 1-}[v(1-s) + \log(1-s)] \\ &= \sum_{i=1}^{\infty}\left(\ln\frac{i+1}{i} - \frac{1}{i}\right) \Big/ \ln 2 \\ &= -\frac{\gamma}{\ln 2} \\ &= -0.832746\cdots,\end{aligned}$$

where $\gamma = \lim_{n\to\infty}\left(\sum_{i=1}^{n}(1/i) - \ln n\right)$ is Euler's constant. Hence the expectation of $\log R'_n$ is approximately equal to $n - \gamma/\ln 2$ for large n. In Markov case a similar result is obtained in [1].

In Section 2 we prove Theorem 1.2 and we propose a practical formula for entropy approximation in Section 3.

2. Proof of Theorem 1.2

An *alphabet* is a finite set $\mathcal{A}$ and we call each element of $\mathcal{A}$ a *symbol*. A *block* is a finite sequence of symbols, and an *n-block* is a block of length n. Let $|B|$ be the length of the block B. For an n-block $B = a_1 a_2 \cdots a_n$ we write $B_{[i,j]} = a_i a_{i+1} \cdots a_j$, $1 \le i \le j \le n$.

DEFINITION 2.1. Let B be an n-block. Suppose m satisfies $1 \le m < n$ and

$$(B_{[1,m]} B_{[1,m]} \cdots B_{[1,m]})_{[1,n]} = B,$$

for some $1 \le j \le m$. The smallest such m is denoted by $\lambda_1(B)$, and the next smallest such m that is not a multiple of $\lambda_1(B)$ is $\lambda_2(B)$, and we can define $\lambda_k(B)$ by the smallest such m which is not a multiple of $\lambda_i(B)$ for every $i < k$.

Let $\Lambda(B) = \{\lambda_1(B), \lambda_2(B), \ldots\}$ and if B has no such m, we write $\Lambda(B) = \emptyset$.

EXAMPLE 2.2. Consider the case of binary blocks, in other words, the symbols are 0 and 1. The number of different binary 4-blocks is 16. By symmetry we only have to examine 8 different blocks having '0' as the first symbol. We have

$$\begin{aligned}&\Lambda(0000) = \{1\}, && \Lambda(0010) = \Lambda(0100) = \Lambda(0110) = \{3\},\\ &\Lambda(0101) = \{2\}, && \Lambda(0001) = \Lambda(0011) = \Lambda(0111) = \emptyset.\end{aligned}$$

If we consider the 5-block $B = 00100$, we have $\Lambda(B) = \{3, 4\}$ and $B_{[1,\lambda_1(B)]} = 001$, $B_{[1,\lambda_2(B)]} = 0010$.

We classify n-blocks into the following sets

$$\mathcal{T}(n) = \left\{|B| = n : \lambda_1(B) > \frac{n}{2} \text{ or } \Lambda(B) = \emptyset\right\},$$
$$\mathcal{R}(n) = \left\{|B| = n : \lambda_1(B) \le \frac{n}{2}\right\}.$$

Lemma 2.3. *For a Bernoulli process,* $\Pr(x_1 \cdots x_n \in \mathcal{R}(n))$ *converges to* 0 *exponentially as* $n \to \infty$.

Proof. Let d be the maximal probability of the symbols. Then for $i < n$ we have

$$\Pr(i \in \Lambda(x_1 \cdots x_n)) \le d^{n-i}$$

and

$$\Pr(x_1 \cdots x_n \in \mathcal{R}(n)) \le \sum_{i=1}^{[n/2]} d^{n-i} = \frac{d^{n-[n/2]} - d^n}{1-d},$$

where $[t]$ is the greatest integer that does not exceed t. □

Lemma 2.4. *Let B be an n-block.*
(i) *If $B = (CB)_{[1,n]}$ for some m-block C, $1 \le m < n$, then $m \in \Lambda(B)$ or m is a multiple of some $\lambda \in \Lambda(B)$.*

(ii) *If $B = B_{[m+1,n]}B_{[1,m]}$ for some $1 \le m < n$, then there is $\lambda \in \Lambda(B)$ such that λ divides n and m.*

Proof. (i) is directly derived from the definition of $\Lambda(B)$.
(ii) Let $m' = \gcd(m, n)$ and $n = hm'$, $m = lm'$. Put $B_i = B_{[(i-1)m'+1, im']}$. Then $B_1 \cdots B_h = B_{l+1} \cdots B_h B_1 \cdots B_l$. So we have $B_i = B_j$ if $i \equiv j + l \pmod h$. Since l and h are relatively prime, B_i's are identical for every i.
□

Definition 2.5. (i) For an n-block B and $k \ge n$ let $\mathcal{F}(B, k)$ be the set of all k-blocks C such that BC of length $k+n$ does not contain B except for the first B, in other words,

$$\mathcal{F}(B, k) = \{C : (BC)_{[i,i+n-1]} \ne B \text{ for any } i > 1\}.$$

For $1 \le k < n$, let $\mathcal{F}(B, k)$ be the set of all k-blocks.
(ii) Let $\mathcal{S}(B, k)$ be the set of k-blocks C, $k \ge 1$, such that BCB of length $k+2n$ does not contain B except for the first and the last B's, or equivalently

$$\mathcal{S}(B, k) = \{C : (BCB)_{[i,i+n-1]} \ne B \text{ for any } i,\ 1 < i \le k+n\}.$$

Clearly, we have $\mathcal{S}(B,k) \subset \mathcal{F}(B,k)$.

EXAMPLE 2.6. Take $B = 010$ and $k = 3$. The 3-blocks '001' is not in $\mathcal{S}(010,3)$ but in $\mathcal{F}(010,3)$, since the 6-block '010 001 010' has three '010' blocks (e.g. 010 001 010). Now we have

$$\mathcal{F}(010,3) = \{000, 001, 011, 110, 111\},$$
$$\mathcal{S}(010,3) = \{000, 011, 110, 111\}.$$

The following shows the relation between $\mathcal{F}(B,k)$ and $\mathcal{S}(B,k)$.

Lemma 2.7. *For $B \in \mathcal{T}(n)$ and $k > n$ we have a pairwise disjoint union*

$$\begin{aligned}\mathcal{S}(B,k) &= \mathcal{F}(B,k) \setminus \bigcup_{\lambda\in\Lambda(B)} \{C \in \mathcal{F}(B,k) : (BC)_{[k+n-\lambda+1,k+n]} = B_{[1,\lambda]}\} \\ &= \mathcal{F}(B,k) \setminus \bigcup_{\lambda\in\Lambda(B)} \{CB_{[1,\lambda]} : C \in \mathcal{S}(B, k-\lambda)\}.\end{aligned}$$

Proof. Take a k-block $C \in \mathcal{F}(B,k)$. If $(BCB)_{[s,s+n-1]} = B$ for some s, then $s > k+1$ and

$$B = (BCB)_{[s,s+n-1]} = (BC)_{[s,k+n]}B_{[1,s-k-1]}.$$

Put $\lambda = k+n-s+1$. Then by Lemma 2.4(i) $\lambda \in \Lambda(B)$ and $(BC)_{[s,k+n]} = B_{[1,\lambda]}$. Hence we have

$$\begin{aligned}\mathcal{S}(B,k) &= \{C : (BCB)_{[i,i+n-1]} \neq B \text{ for any } i,\ \ i < i \le k+n\} \\ &= \{C \in \mathcal{F}(B,k) : (BC)_{[k+n-\lambda+1,k+n]} \neq B_{[1,\lambda]} \text{ for any } \lambda \in \Lambda(B)\} \\ &= \mathcal{F}(B,k) \setminus \bigcup_{\lambda\in\Lambda(B)} \{C \in \mathcal{F}(B,k) : (BC)_{[k+n-\lambda+1,k+n]} = B_{[1,\lambda]}\}.\end{aligned}$$

Suppose that there exists $C \in \mathcal{F}(B,k)$ such that $C_{[k+n-\lambda+1,k+n]} = B_{[1,\lambda]}$ and $C_{[k+n-\lambda'+1,k+n]} = B_{[1,\lambda']}$ for some λ, $\lambda' \in \Lambda(B)$ with $\lambda < \lambda'$. Then $B_{[1,\lambda]} = B_{[\lambda'-\lambda+1,\lambda']}$ and

$$B_{[1,\lambda']} = (B_{[1,\lambda]}B_{[1,\lambda]})_{[1,\lambda']} = B_{[1,\lambda]}B_{[1,\lambda'-\lambda]} = B_{[\lambda'-\lambda+1,\lambda']}B_{[1,\lambda'-\lambda]}.$$

By Lemma 2.4(ii) $B_{[1,\lambda']} = B_{[1,\lambda'-\lambda]} \cdots B_{[1,\lambda'-\lambda]}$ and this contradicts $\lambda' \in \Lambda(B)$. Hence the sets $\{C \in \mathcal{F}(B,k) : C_{[k+n-\lambda+1,k+n]} = B_{[1,\lambda]}\}$ are disjoint.

For every $C \in \mathcal{S}(B,k-\lambda)$ we have $CB_{[1,\lambda]} \in \mathcal{F}(B,k)$ obviously. Put $C \in \mathcal{F}(B,k)$ with $C_{[k-\lambda+1,k]} = B_{[1,\lambda]}$. If $C_{[1,k-\lambda]} \notin \mathcal{S}(B,k-\lambda)$, then there is λ' with $\lambda' < n-\lambda$ such that $C_{[1,k-\lambda-\lambda']}B = (C_{[1,k-\lambda]}B)_{[1,k+n-\lambda-\lambda']}$ or $B = (C_{[k-\lambda-\lambda'+1,k-\lambda]}B)_{[1,n]}$. So by Lemma 2.4(i) $\lambda' \in \Lambda(B)$ or λ' is a multiple of $\lambda_1(B)$. Since $\lambda' < n-\lambda < n/2$, this contradicts $B \in \mathcal{T}(n)$. Hence we have

$$\{C \in \mathcal{F}(B,k) : (BC)_{[k+n-\lambda+1,k+n]} = B_{[1,\lambda]}\} = \{CB_{[1,\lambda]} : C \in \mathcal{S}(B,k-\lambda)\}. \qquad \square$$

DEFINITION 2.8. For a given n-block B, define

$$\begin{aligned} p_i(B) &= \Pr(R_n(x) = i \mid x_1 \ldots x_n = B,\ R_n(x) \geq i), \\ r_i(B) &= \Pr(x_{n+1} \cdots x_{n+i} \in \mathcal{F}(B,i) \mid x_1 \cdots x_n = B), \\ s_i(B) &= \Pr(x_{n+1} \cdots x_{n+i} \in \mathcal{S}(B,i) \mid x_1 \cdots x_n = B). \end{aligned}$$

We have $r_i(B) \geq s_i(B)$. Put $r_0(B) = 1$, $s_0(B) = 1$.

Proposition 2.9. *For Bernoulli processes we have*

$$\begin{aligned} \Pr(R_n(x) > i \mid x_1 \cdots x_n = B) &= r_i(B), \\ \Pr(R_n(x) = i + n \mid x_1 \cdots x_n = B) &= s_i(B) \Pr(B). \end{aligned}$$

Proof. Let $x_1 \cdots x_n = B$. Since $R_n(x) > i$ if and only if $x_1 \cdots x_{i+n} = BC$ for some $C \in \mathcal{F}(B,i)$, we have

$$\Pr(R_n(x) > i \mid x_1 \cdots x_n = B) = \Pr(x_{n+1} \cdots x_{i+n} \in \mathcal{F}(B,i)) = r_i(B).$$

And since $R_n(x) = i + n$ if and only if $x_1 \cdots x_{i+2n} = BCB$ for some $C \in \mathcal{S}(B,i)$, we have

$$\begin{aligned} &\Pr(R_n(x) = i + n \mid x_1 \cdots x_n = B) \\ &\quad = \Pr(x_{n+1} \cdots x_{i+n} \in \mathcal{S}(B,i)) \cdot \Pr(x_{i+n+1} \cdots x_{i+2n} = B) \\ &\quad = s_i(B) \cdot \Pr(B). \end{aligned}$$

Since $r_0(B) = s_0(B) = 1$, the equations hold for $i = 0$. $\square$

Now we find the recurrence relations between $r_k(B)$ and $s_k(B)$.

Proposition 2.10. *For a Bernoulli process with $B \in \mathcal{T}(n)$, if $i \geq n$,*

$$\begin{aligned} r_i(B) &= r_{i-1}(B) - s_{i-n}(B) \Pr(B), \\ s_i(B) &= r_i(B) - \sum_{\lambda \in \Lambda(B)} \Pr(B_{[1,\lambda]}) s_{i-\lambda}(B). \end{aligned}$$

Proof. From Proposition 2.9 we have

$$\begin{aligned} r_i(B) &= \Pr(R_n(x) > i \mid x_1 \cdots x_n = B) \\ &= \Pr(R_n(x) > i-1 \mid x_1 \cdots x_n = B) - \Pr(R_n(x) = i \mid x_1 \cdots x_n = B) \\ &= r_{i-1}(B) - s_{i-n}(B)\Pr(B). \end{aligned}$$

By Lemma 2.7 we have

$$\begin{aligned} s_i(B) &= \Pr(x_{n+1} \cdots x_{n+i} \in \mathcal{F}(B,i)) \\ &\quad - \sum_{\lambda \in \Lambda(B)} \Pr(x_{n+1} \cdots x_{n+i-\lambda} \in \mathcal{S}(B, i-\lambda),\ x_{n+i-\lambda+1} \cdots x_{n+i} = B_{[1,\lambda]}) \\ &= r_i(B) - \sum_{\lambda \in \Lambda(B)} s_{i-\lambda}(B)\Pr(B_{[1,\lambda]}). \end{aligned}$$

□

Lemma 2.11. *For a Bernoulli process with $B \in \mathcal{T}(n)$, if $i \geq n$,*

$$p_i(B) = \frac{\Pr(B)}{(1-p_{i-1}(B)) \cdots (1-p_{i-n+1}(B))} - \sum_{\lambda \in \Lambda(B)} \frac{\Pr(B_{[1,\lambda]})p_{i-\lambda}(B)}{(1-p_{i-1}(B)) \cdots (1-p_{i-\lambda}(B))}.$$

Proof. From Proposition 2.10 we have for $i \geq n$

$$s_i(B) = r_i(B) - \sum_{\lambda \in \Lambda(B)} \frac{r_{i+n-\lambda-1}(B) - r_{i+n-\lambda}(B)}{\Pr(B_{[\lambda+1,n]})}$$

and

$$r_{i-1}(B) - r_i(B) = \Pr(B)r_{i-n}(B) - \sum_{\lambda \in \Lambda(B)} \Pr(B_{[1,\lambda]})(r_{i-\lambda-1}(B) - r_{i-\lambda}(B)).$$

Since $r_i(B) = (1-p_1(B))(1-p_2(B)) \cdots (1-p_i(B))$, we conclude the lemma. □

REMARK 2.12. (i) From the definition of $\Lambda(B)$ we have for $1 \leq i < n = |B|$

$$p_i(B) = \begin{cases} 0 & \text{if } i \notin \Lambda(B), \\ \dfrac{\Pr(B_{[1,i]})}{\prod_{j<i}(1-p_j(B))} & \text{if } i \in \Lambda(B) \end{cases}$$

and

$$(1-p_1(B)) \cdots (1-p_{n-1}(B)) = \begin{cases} 1 & \text{if } \Lambda(B) = \emptyset, \\ 1 - \sum_{\lambda \in \Lambda(B)} \Pr(B_{[1,\lambda]}) & \text{if } \Lambda(B) \neq \emptyset. \end{cases}$$

(ii) For $B \in \mathcal{T}(n)$ and $\lambda \in \Lambda(B)$ we have

$$\Pr(B) = \Pr(B_{[1,\lambda]}) \Pr(B_{[1,n-\lambda]}) = \Pr(B_{[1,n-\lambda]})^2 \Pr(B_{[n-\lambda+1,\lambda]}),$$

and

$$\Pr(B) < \Pr(B_{[1,\lambda]}) < \sqrt{\Pr(B)}.$$

Lemma 2.13. *For each Bernoulli process there is N such that for every block $B \in \mathcal{T}(n)$ with $n > N$ we have $p_i(B) < \sqrt{\Pr(B)}$ for each $i \geq 1$.*

Proof. Let d be the maximal probability of the symbols. Choose N so that $\sqrt{d^n} < 1/n$ and $n^{1-2/n}/(n-1) < 1$ for all $n > N$ and $N \geq 2$. Put $\lambda_i = \lambda_i(B)$. If $p_{\lambda_j}(B) < \sqrt{\Pr(B)}$ for all $j < i$, then

$$\begin{aligned}
p_{\lambda_i}(B) &= \frac{\Pr(B_{[1,\lambda_i]})}{(1 - p_{\lambda_1}(B)) \cdots (1 - p_{\lambda_{i-1}}(B))} \\
&= \Pr(B_{[1,\lambda_1]}) \frac{\Pr(B_{[\lambda_1+1,\lambda_2]})}{1 - p_{\lambda_1}(B)} \cdots \frac{\Pr(B_{[\lambda_{i-1}+1,\lambda_i]})}{1 - p_{\lambda_{i-1}}(B)} \\
&< \sqrt{\Pr(B)} \left(\frac{d}{1 - \sqrt{\Pr(B)}} \right)^{i-1} < \sqrt{\Pr(B)} \left(\frac{n^{-2/n}}{1 - \sqrt{d^n}} \right)^{i-1} \\
&< \sqrt{\Pr(B)} \left(\frac{n^{1-2/n}}{n-1} \right)^{i-1} < \sqrt{\Pr(B)}.
\end{aligned}$$

Since

$$p_{\lambda_1}(B) = \Pr(B_{[1,\lambda_1]}) < \sqrt{\Pr(B)},$$

by induction rule we have $p_\lambda(B) < \sqrt{\Pr(B)}$ for all $\lambda \in \Lambda(B)$. and $p_i(B) < \sqrt{\Pr(B)}$ for all $i < n$.

If $p_i(B) < \sqrt{\Pr(B)}$ for all $i < j$, Then from Lemma 2.11 we have

$$\begin{aligned}
p_j(B) &< \frac{\Pr(B)}{(1 - \sqrt{\Pr(B)})^{n-1}} \\
&\leq \sqrt{\Pr(B)} \frac{\sqrt{d^n}}{(1 - \sqrt{d^n})^{n-1}} < \sqrt{\Pr(B)} \frac{\sqrt{d^n}}{1 - (n-1)\sqrt{d^n}} \\
&= \sqrt{\Pr(B)} \frac{\sqrt{d^n}}{1 - n\sqrt{d^n} + \sqrt{d^n}} \leq \sqrt{\Pr(B)}.
\end{aligned}$$

□

Lemma 2.14. *If we put*

$$\alpha_n \equiv \frac{n\sqrt{d^n}}{1 - n\sqrt{d^n}},$$

where d is the maximal probability of the symbols, then for $i \geq n$ we have

$$p_i(B) < \Pr(B) + \Pr(B)\alpha_n.$$

Proof. By Lemma 2.11 and 2.13 we have

$$\begin{aligned} p_i(B) &\leq \frac{\Pr(B)}{(1 - p_{i-1}(B)) \cdots (1 - p_{i-n+1}(B))} \\ &< \frac{\Pr(B)}{(1 - \sqrt{\Pr(B)}\,)^{n-1}} \leq \frac{\Pr(B)}{(1 - \sqrt{d^n}\,)^{n-1}} \end{aligned}$$

and

$$\begin{aligned} p_i(B) - \Pr(B) &< \Pr(B) \left(\frac{1 - (1 - \sqrt{d^n}\,)^{n-1}}{(1 - \sqrt{d^n}\,)^{n-1}} \right) \\ &< \Pr(B) \left(\frac{(n-1)\sqrt{d^n}}{1 - (n-1)\sqrt{d^n}} \right) < \Pr(B)\alpha_n. \end{aligned}$$

□

Lemma 2.15. *For sufficiently large n if $B \in \mathcal{T}(n)$ and $i \geq n$, then we have*

$$p_i(B) > \Pr(B) - \Pr(B)\alpha_n.$$

Proof. By Lemma 2.11, 2.13 and 2.14 we have

$$\begin{aligned} p_i(B) &\geq \Pr(B) - \sum_{\lambda} \frac{\Pr(B_{[1,\lambda]}) p_{i-\lambda}(B)}{(1 - p_{i-1}(B)) \cdots (1 - p_{i-\lambda}(B))} \\ &> \Pr(B) - \frac{n}{2} \cdot \frac{\Pr(B)(1 + \alpha_n)\sqrt{\Pr(B)}}{(1 - \sqrt{\Pr(B)}\,)^{n-1}} \\ &\geq \Pr(B) - \frac{\Pr(B)(1 + \alpha_n) n \sqrt{d^n}}{2(1 - \sqrt{d^n}\,)^{n-1}} \end{aligned}$$

and

$$\begin{aligned} p_i(B) - \Pr(B) &> -\Pr(B) \frac{(1 + \alpha_n) n \sqrt{d^n}}{2(1 - \sqrt{d^n}\,)^{n-1}} \\ &> -\Pr(B) \frac{\alpha_n (1 + \alpha_n)}{2} > -\Pr(B)\alpha_n \end{aligned}$$

for sufficiently large n. □

Lemma 2.16. *For any sequence $0 \leq c_1 < c_2 < \cdots$, let*

$$F(x_1, x_2, \ldots) = \sum_{i=1}^{\infty} (1 - x_1) \cdots (1 - x_{i-1}) x_i c_i$$

be a function of $x_1, x_2, \ldots$ *under the conditions of*
(1) $0 \le x_i \le 1$, $i = 1, 2, \ldots,$
(2) $\sum_{i=1}^{\infty} x_i$ *diverges.*
Then F *is monotonously decreasing in* $x_1, x_2, \ldots$.

Proof. For a fixed k and $\delta > 0$, we have

$$\begin{aligned} &F(x_1, \ldots, x_k, \ldots) - F(x_1, \ldots, x_k + \delta, \ldots) \\ &=(1 - x_1) \cdots (1 - x_{k-1})\delta \left(-c_k + \sum_{i=k+1}^{\infty} (1 - x_{k+1}) \cdots (1 - x_{i-1}) x_i c_i \right) \\ &\ge (1 - x_1) \cdots (1 - x_{k-1})\delta \left(-c_k + c_{k+1} \sum_{i=k+1}^{\infty} (1 - x_{k+1}) \cdots (1 - x_{i-1}) x_i \right). \end{aligned}$$

Since for $m \ge k + 1$,

$$\prod_{i=k+1}^{m} (1 - x_i) = 1 - \sum_{i=k+1}^{m} (1 - x_{k+1}) \cdots (1 - x_{i-1}) x_i,$$

we have

$$\sum_{i=k+1}^{\infty} (1 - x_{k+1}) \cdots (1 - x_{i-1}) x_i = 1 - \prod_{i=k+1}^{\infty} (1 - x_i).$$

From the condition (2)

$$\log \left(\prod_{i=k+1}^{\infty} (1 - x_i) \right) = \sum_{i=k+1}^{\infty} \log(1 - x_i) \le - \sum_{i=k+1}^{\infty} x_i = -\infty,$$

and

$$\sum_{i=k+1}^{\infty} (1 - x_{k+1}) \cdots (1 - x_{i-1}) x_i = 1.$$

Hence we have

$$\begin{aligned} &F(x_1, \ldots, x_k, \ldots) - F(x_1, \ldots, x_k + \delta, \ldots) \\ &\qquad \ge (1 - x_1) \cdots (1 - x_{k-1})\delta(-c_k + c_{k+1}) \ge 0. \end{aligned}$$ □

Proof of Theorem 1.2. Since

$$\Pr(R_n(x) = i \mid x_1 \ldots x_n = B) = (1 - p_1(B)) \cdots (1 - p_{i-1}(B)) p_i(B),$$

we have from Lemma 2.14, 2.15 and 2.16

$$
\begin{aligned}
E[\log R_n \mid x_1\cdots x_n = B] &= \sum_{i=1}^{\infty}(1-p_1(B))\cdots(1-p_{i-1}(B))p_i(B)\log i\\
&\geq \sum_{i=1}^{n-1}(1-p_1(B))\cdots(1-p_{i-1}(B))p_i(B)\log i\\
&\quad + (1-p_1(B))\cdots(1-p_{n-1}(B))\sum_{i=n}^{\infty}(1-\Pr(B)(1+\alpha_n))^{i-n}\Pr(B)(1+\alpha_n)\log i\\
&= \sum_{i=1}^{n-1}(1-p_1)\cdots(1-p_{i-1})p_i\log i + \frac{(1-p_1)\cdots(1-p_{n-1})}{(1-\Pr(B)(1+\alpha_n))^{n-1}}v(\Pr(B)(1+\alpha_n))\\
&\quad - (1-p_1)\cdots(1-p_{n-1})\sum_{i=1}^{n-1}(1-\Pr(B)(1+\alpha_n))^{i-n}\Pr(B)(1+\alpha_n)\log i
\end{aligned}
$$

and

$$
\begin{aligned}
&E[\log R_n \mid x_1\cdots x_n = B]\\
&\quad\leq \sum_{i=1}^{n-1}(1-p_1)\cdots(1-p_{i-1})p_i\log i + \frac{(1-p_1)\cdots(1-p_{n-1})}{(1-\Pr(B)(1-\alpha_n))^{n-1}}v(\Pr(B)(1-\alpha_n))\\
&\qquad - (1-p_1)\cdots(1-p_{n-1})\sum_{i=1}^{n-1}(1-\Pr(B)(1-\alpha_n))^{i-n}\Pr(B)(1-\alpha_n)\log i,
\end{aligned}
$$

where v is the function in Definition 1.3. Put

$$
\begin{aligned}
\Sigma_n^{\pm}(B) \equiv\ &\sum_{i=1}^{n-1}(1-p_1(B))\cdots(1-p_{i-1}(B))p_i(B)\log i\\
&- (1-p_1(B))\cdots(1-p_{n-1}(B))\\
&\quad\sum_{i=1}^{n-1}(1-\Pr(B)(1\pm\alpha_n))^{i-n}\Pr(B)(1\pm\alpha_n)\log i.
\end{aligned}
$$

Then for all $B \in \mathcal{T}(n)$ we have

$$
\begin{aligned}
&\frac{(1-p_1(B))\cdots(1-p_{n-1}(B))}{(1-\Pr(B)(1+\alpha_n))^{n-1}}v(\Pr(B)(1+\alpha_n)) + \Sigma_n^{+}(B)\\
&\quad< E[\log R_n \mid x_1\cdots x_n = B]\\
&\quad< \frac{(1-p_1(B))\cdots(1-p_{n-1}(B))}{(1-\Pr(B)(1-\alpha_n))^{n-1}}v(\Pr(B)(1-\alpha_n)) + \Sigma_n^{-}(B).
\end{aligned}
$$

From Remark 2.12 we have for $B \in \mathcal{T}(n)$

$$\begin{aligned}\frac{1-\sum_{\lambda\in\Lambda(B)}\Pr(B_{[1,\lambda]})}{(1-\Pr(B))^{n-1}}&\upsilon(\Pr(B)(1+\alpha_n))+\Sigma_n^+(B)\\ &<E[\log R_n \mid x_1\cdots x_n=B]\\ &<\frac{1-\sum_{\lambda\in\Lambda(B)}\Pr(B_{[1,\lambda]})}{(1-\Pr(B))^{n-1}}\upsilon(\Pr(B)(1-\alpha_n))+\Sigma_n^-(B)\end{aligned}$$

and

$$\begin{aligned}\frac{1-\sum_{\lambda\in\Lambda(B)}\Pr(B_{[1,\lambda]})}{(1-\Pr(B))^{n-1}}&\upsilon(\Pr(B)(1+\alpha_n))-\upsilon(\Pr(B))+\Sigma_n^+(B)\\ &<E[\log R_n-\upsilon(\Pr(x_1\cdots x_n))\mid x_1\cdots x_n=B]\\ &<\frac{1-\sum_{\lambda\in\Lambda(B)}\Pr(B_{[1,\lambda]})}{(1-\Pr(B))^{n-1}}\upsilon(\Pr(B)(1-\alpha_n))-\upsilon(\Pr(B))+\Sigma_n^-(B).\end{aligned}$$

By Lemma 2.17 given below we have

$$\lim_{n\to\infty} E\left[\log R_n(x)-\upsilon(\Pr(x_1\cdots x_n))\mid x_1\cdots x_n\in\mathcal{T}(n)\right]=0.$$

and

$$\lim_{n\to\infty} E\left[\log R_n(x)+\log P_n(x)\mid x_1\cdots x_n\in\mathcal{T}(n)\right]=-\frac{\gamma}{\ln 2}.$$

By Jensen's inequality and Kac's lemma, for any B we have

$$E\left[\log R_n+\log P_n\mid x_1\cdots x_n=B\right]\le\log E\left[R_nP_n\mid x_1\cdots x_n=B\right]\le 0$$

and if we let $\overline{d}$ be the minimal probability of a symbol,

$$E\left[\log R_n+\log P_n\mid x_1\cdots x_n=B\right]\ge E\left[\log P_n\mid x_1\cdots x_n=B\right]\ge n\log\overline{d}.$$

Hence by Lemma 2.3 we have

$$\lim_{n\to\infty} E\left[\log R_n+\log P_n\right]=-\frac{\gamma}{\ln 2}.$$ □

Lemma 2.17. *For sufficiently large n, if $B\in\mathcal{T}(B)$, then*

$$\left|\Sigma_n^{\pm}(B)\right|<(n-1)\sqrt{d^n}\,\log(n-1)+(n-1)d^n(1+\alpha_n)\log(n-1)$$

and

$$\begin{aligned}&\left|\frac{1-\sum_{\lambda\in\Lambda(B)}\Pr(B_{[1,\lambda]})}{(1-\Pr(B))^{n-1}}\upsilon(\Pr(B)(1\pm\alpha_n))-\upsilon(\Pr(B))\right|\\ &\quad<\frac{\eta(d^n(1+\alpha_n))}{(1-d^n)^{n-1}}-\frac{\log(1-\alpha_n)}{(1-d^n)^{n-1}}-\frac{\log d}{2}n^2d^{n/2},\end{aligned}$$

where $\eta(x) = \upsilon(x) + \log(x) + \gamma/\ln 2$.

Proof. By Remark 2.12 for $1 \le i < n$ we have

$$(1-p_1(B))\cdots(1-p_{i-1}(B))p_i(B) = \begin{cases} 0 & \text{if } i \notin \Lambda(B), \\ \Pr(B_{[1,i]}) & \text{if } i \in \Lambda(B). \end{cases}$$

Since if $B \in \mathcal{T}(n)$, $\Pr(B_{[1,\lambda]}) < \sqrt{\Pr(B)}$ for $\lambda \in \Lambda(B)$, we have

$$\begin{aligned} |\Sigma_n^{\pm}(B)| \le & \sum_{i=1}^{n-1}(1-p_1)\cdots(1-p_{i-1})p_i \log i \\ & + (1-p_1)\cdots(1-p_{n-1})\sum_{i=1}^{n-1}(1-\Pr(B)(1\pm\alpha_n))^{i-n}\Pr(B)(1\pm\alpha_n)\log i \\ < & (n-1)\sqrt{\Pr(B)}\log(n-1) + (n-1)\Pr(B)(1+\alpha_n)\log(n-1) \\ \le & (n-1)\sqrt{d^n}\log(n-1) + (n-1)d^n(1+\alpha_n)\log(n-1). \end{aligned}$$

Now consider the blocks of $\mathcal{T}(n)$. Since $\Pr(B_{[1,\lambda]}) < \sqrt{\Pr(B)}$, we have

$$1 - n\sqrt{\Pr(B)} < \frac{1-\sum_{\lambda\in\Lambda(B)}\Pr(B_{[1,\lambda]})}{(1-\Pr(B))^{n-1}} < \frac{1}{(1-\Pr(B))^{n-1}}.$$

The function $\eta(x) = \upsilon(x) + \log(x) + \gamma/\ln 2$ of x is monotonically increasing with $\lim_{x\to 0}\eta(x) = 0$. Hence

$$\begin{aligned} & \frac{1-\sum_{\lambda\in\Lambda(B)}\Pr(B_{[1,\lambda]})}{(1-\Pr(B))^{n-1}}\upsilon(\Pr(B)(1\pm\alpha_n)) - \upsilon(\Pr(B)) \\ & < \frac{\upsilon(\Pr(B)(1-\alpha_n))}{(1-\Pr(B))^{n-1}} - \upsilon(\Pr(B)) \\ & = \frac{\eta(\Pr(B)(1-\alpha_n))}{(1-\Pr(B))^{n-1}} - \eta(\Pr(B)) - \frac{\log(1-\alpha_n)}{(1-\Pr(B))^{n-1}} \\ & \quad - \left(\frac{1}{(1-\Pr(B))^{n-1}} - 1\right)\left(\log\Pr(B) + \frac{\gamma}{\ln 2}\right) \\ & < \frac{\eta(\Pr(B)(1-\alpha_n))}{(1-\Pr(B))^{n-1}} - \frac{\log(1-\alpha_n)}{(1-\Pr(B))^{n-1}} - \frac{(n-1)\Pr(B)}{1-(n-1)\Pr(B)}\log\Pr(B) \\ & < \frac{\eta(d^n(1-\alpha_n))}{(1-d^n)^{n-1}} - \frac{\log(1-\alpha_n)}{(1-d^n)^{n-1}} - \frac{n(n-1)d^n}{1-(n-1)d^n}\log d \\ & < \frac{\eta(d^n(1-\alpha_n))}{(1-d^n)^{n-1}} - \frac{\log(1-\alpha_n)}{(1-d^n)^{n-1}} - n^2 d^n \log d \end{aligned}$$

and

$$\begin{aligned}
&\frac{1-\sum_{\lambda\in\Lambda(B)}\Pr(B_{[1,\lambda]})}{(1-\Pr(B))^{n-1}}\upsilon(\Pr(B)(1\pm\alpha_n))-\upsilon(\Pr(B))\\
&\quad>\left(1-\sum_{\lambda}\Pr(B_{[1,\lambda]})\right)\upsilon(\Pr(B)(1+\alpha_n))-\upsilon(\Pr(B))\\
&\quad=\left(1-\sum_{\lambda}\Pr(B_{[1,\lambda]})\right)\eta(\Pr(B)(1+\alpha_n))-\eta(\Pr(B))\\
&\qquad-\left(1-\sum_{\lambda}\Pr(B_{[1,\lambda]})\right)\log(1+\alpha_n)+\sum_{\lambda}\Pr(B_{[1,\lambda]})\left(\log\Pr(B)+\frac{\gamma}{\ln 2}\right)\\
&\quad>-\eta(\Pr(B))-\log(1+\alpha_n)+\frac{n}{2}\sqrt{\Pr(B)}\log\Pr(B)\\
&\quad>-\eta(d^n)-\alpha_n+\frac{\log d}{2}n^2d^{n/2}.
\end{aligned}$$

Hence we have

$$\left|\frac{1-\sum_{\lambda\in\Lambda(B)}\Pr(B_{[1,\lambda]})}{(1-\Pr(B))^{n-1}}\upsilon(\Pr(B)(1\pm\alpha_n))-\upsilon(\Pr(B))\right|<A_n$$

where

$$A_n=\frac{\eta(d^n(1+\alpha_n))}{(1-d^n)^{n-1}}-\frac{\log(1-\alpha_n)}{(1-d^n)^{n-1}}-\frac{\log d}{2}n^2d^{n/2}.$$

□

3. Estimation of entropy

From Theorem 1.2 $(E[\log R_n]+\gamma/\ln 2)/n$ is close to the entropy for sufficiently large n. If T is the left-shift defined by $(Tx)_k=x_{k+1}$, then by the ergodicity we have $(1/M)\sum_{0\le i\le 1}\log R_n(T^ix)$ converges to $E[\log R_n]$ almost surely as $M\to\infty$. Hence we approximate the entropy by

$$H(n,M)\equiv\frac{1}{n}\left(\frac{1}{M}\sum_{0\le i\le M-1}\log R_n(T^ix)+\frac{\gamma}{\ln 2}\right).$$

The conventional formula with no correction term is given by

$$H'(n,M)\equiv\frac{1}{n}\frac{1}{M}\sum_{0\le i\le M-1}\log R_n(T^ix).$$

In the following we compare the effectiveness of $H(n,M)$ and $H'(n,M)$.

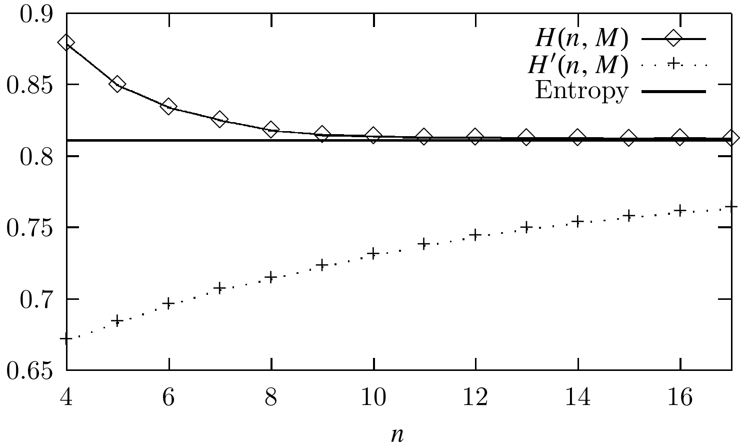

Fig. 1. Test result for Example 3.1 for $M = 10000$.

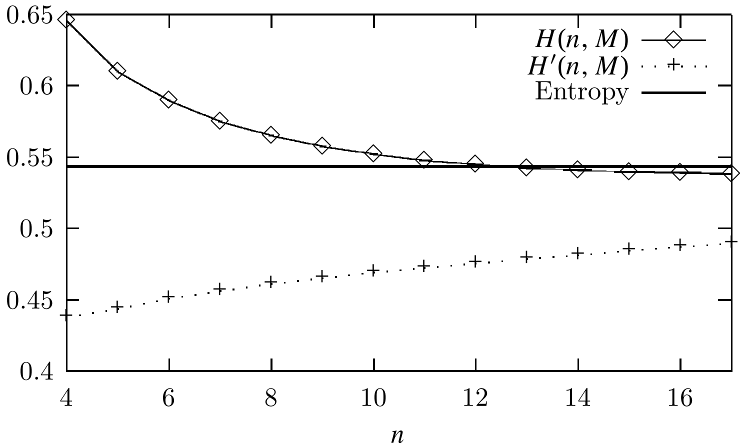

Fig. 2. Test result for Example 3.2 for $M = 10000$.

EXAMPLE 3.1. Consider the Bernoulli process associated with the $(1/4, 3/4)$ product measure. Note that $h = -1/4\log_2(1/4) - 3/4\log_2(3/4) = 0.811278\cdots$. For generating the typical point of Bernoulli process x, we use the random number generator employed in Fortran 90. Here $M = 10000$ is rather large to demonstrate the accuracy of the theoretical prediction and in practical applications a sample of small size will do. The test result is given in Fig. 1.

EXAMPLE 3.2. Consider the Bernoulli process associated with the $(1/8, 7/8)$ product measure. Note that $h = -1/8\log_2(1/8) - 7/8\log_2(7/8) = 0.543564\cdots$. We test this example by the same method as before. The test result is given in Fig. 2.

ACKNOWLEDGEMENT. The author thank Prof. Geon Ho Choe, Prof. S. Kotani and the referee for many helpful suggestions and comments.

References

[1] G.H. Choe and D.H. Kim: *Average convergence rate of the first return time*, Colloq. Math. **84/85** (2000), 159–171.
[2] G.H. Choe and D.H. Kim: *The first return time test of pseudorandom numbers*, J. Comput. Appl. Math. **143** (2002), 263–247.
[3] M. Kac: *On the notion of recurrence in discrete stochastic processes*, Bull. Amer. Math. Soc. **53** (1947), 1002–1010.
[4] I. Kontoyiannis: *Asymptotic recurrence and waiting times for stationary processes*, J. Theor. Prob. **11** (1998), 795–811.
[5] U. Maurer: *A universal statistical test for random bit generators*, J. Cryptology, **5** (1992), 89–105.
[6] D. Ornstein and B. Weiss: *Entropy and data compression schemes*, IEEE Trans. Inform. Theory, **39** (1993), 78–83.
[7] P. Shields: *The interactions between ergodic theory and information theory*, IEEE Trans. Inform. Theory, **44** (1998), 2079–2093, Information theory: 1948–1998.
[8] A.D. Wyner and J. Ziv: *Some asymptotic properties of the entropy of stationary ergodic data source with applications to data compression*, IEEE Trans. Inform. Theory, **35** (1989), 1250–1258.
[9] A.J. Wyner: Strong matching theorems and applications to data compression and statistics, Ph.D. thesis, Stanford Univ., 1993.
[10] A.J. Wyner: *More on recurrence and waiting times*, Ann. Appl. Probab. **9** (1999), 780–796.
[11] J. Ziv and A. Lempel: *A universal algorithm for sequential data compression*, IEEE Trans. Inform. Theory, **23** (1977), 337–343.

School of Mathematics
Korea Institute for Advanced Study
Seoul 130-722 Korea
e-mail: kimdh@kias.re.kr