Comm. Korean Math. Soc. **15** (2000), No. 4, pp. 683–689

ENTROPY AND THE RANDOMNESS OF THE DIGITS OF PI

GEON HO CHOE AND DONG HAN KIM

ABSTRACT. The convergence rate of the expectation of the logarithm of the first return time R_n with block length n has been investigated for Bernoulli processes. This idea is applied to check the randomness of the digits of the decimal expansion of π, e and $\sqrt{2}$.

1. Introduction

Let (X, μ) be a probability space. A measurable transformation $T : (X, \mu) \to (X, \mu)$ is called measure preserving if $\mu(E) = \mu(T^{-1}E)$ for every measurable set E. Sometimes we say that T preserves μ or μ is T-invariant. It is called ergodic if a measurable set E satisfies $\mu(T^{-1}E \triangle E) = 0$ then $\mu(E) = 0$ or $\mu(E) = 1$ where $\triangle$ denotes the symmetric difference of two sets. The Birkhoff Ergodic Theorem states that if T is measure preserving and ergodic, then

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k x) = \int_X f d\mu$$

for almost every (a.e.) x.

Let $1_E(x)$ denote the characteristic function of a measurable subset $E \subset X$. Choose $f(x) = 1_E(x)$ in the Birkhoff Ergodic Theorem. Then we see that if T is ergodic then the average number of times that the points $T^k x$ visit E is equal to the size of the subset E. In other words, ergodicity implies the uniform distribution.

An example of an ergodic invariant map is given by a translation $T(x) = \{2x\}$ on $X = [0, 1)$ where $\{t\}$ is the fractional part of a real number t. The Lebesgue measure is the invariant measure. By applying

Received January 4, 2000. Revised September 25, 2000.
2000 Mathematics Subject Classification: Primary 37M25, secondary 94A17.
Key words and phrases: the first return time, entropy, digits of π, ergodicity.

the Birkhoff Ergodic Theorem to this example we obtain Borel's Normal Number Theorem.

EXAMPLE 1.1 (Borel's Normal Number Theorem). Take $X = [0,1)$, $Tx = \{2x\}$ and $E = [\frac{1}{2}, 1)$. A Fourier series expansion argument shows that any invariant function is constant, hence T is ergodic. Let $x = 0.a_1a_2a_3\ldots$, $a_k \in \{0,1\}$, be the binary expansion of x. Note that $1_E(T^{k-1}x) = 1$ if and only if $a_k = 1$. Using the Birkhoff Ergodic Theorem one can show that the average number of 1's (or 0's) in the binary expansion of a.e. x is the measure of E, which equals $\frac{1}{2}$.

In the rest of this paper we consider $X = \prod_{i=1}^{\infty} \mathcal{A}$ be the product of a finite set $\mathcal{A} = \{a_1, \ldots, a_n\}$ and T is the shift transformation given by $(Tx)_n = x_{n+1}$ for every n. An example of the T-invariant or stationary measure is Bernoulli measure which is the product measure of a finite measure on $\mathcal{A}$. It is known to be ergodic.

Entropy was first formulated by C. Shannon[10] in 1948. It was used in information theory to estimate the compression rate of given data. Mathematically, it is equivalent to the entropy of a measure preserving transformation. More precisely, first we define entropy for a partition of a probability measure space as follows: Let $\mathcal{P} = \{B_1, \ldots, B_r\}$ be a finite partition of a probability space (X, μ). Then the entropy of the partition $\mathcal{P}$ is defined by

$$H(\mathcal{P}) = -\sum_{i=1}^{r} \mu(B_i) \log \mu(B_i).$$

And the entropy of T with respect to $\mathcal{P}$ is defined by

$$h(T, \mathcal{P}) = \lim_{n\to\infty} H(\mathcal{P} \vee \cdots \vee T^{-(n-1)}\mathcal{P})/n,$$

where $\mathcal{P}_1 \vee \cdots \vee \mathcal{P}_n$ consists of intersections of subsets from each partition. It is known that the limit exists. Finally the entropy of T is defined by

$$h(T) = \sup_{\mathcal{P}} h(T, \mathcal{P}).$$

Consider $X = \prod_{i=1}^{\infty}\{a_1, \ldots, a_n\}$ with a shift invariant (or stationary) ergodic probability measure. In this case the transformation under consideration is given by the shift $T : (x_1x_2x_3\cdots) \mapsto (x_2x_3x_4\cdots)$. For a Bernoulli shift transformation on X, the entropy is equal to $h = -\sum_{i=1}^{n} p_i \log p_i$, where the Bernoulli measure is given by $\mu(\{a_i\}) = p_i$.

Recently in relation to data compression methods such as Ziv-Lempel algorithms[15], convergence of the logarithm of the first return time normalized by the block length n has been investigated. For a typical sequence $x = (x_1, x_2, \dots)$ from an ergodic stationary information source X, i.e., a probability space of the form $X = \prod_{i=1}^{\infty}\{a_1, \dots, a_n\}$ with a shift invariant ergodic measure, define $P_n(x)$ to be the probability of the initial n-block in x, i.e., $P_n(x) = \Pr(x_1, \dots, x_n)$. The Shannon-McMillan-Breiman Theorem states that $-\log P_n(x)/n$ converges to the entropy of T in L^1 and almost surely.

DEFINITION 1.2. 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[4] states that $E[R_n(x) \mid x_1 \dots x_n = a_1 \dots a_n]$ is equal to $1/\Pr(a_1 \dots a_n)$ for an ergodic stationary source. 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[14] and almost surely by Ornstein and Weiss[9]. For a sharp estimation of the convergence of the average of $(\log R_n)/n$, see [3]. For a comprehensive introduction to the subject consult Shields[11] and the references therein. Kim[6] showed that for Bernoulli shift the expectation of the return time $E[\log R_n]/n$ asymptotically goes to $h - \gamma/n$ where γ is Euler's constant. In this article we apply this relation between first return time and entropy to π and some other numbers.

The property of $\pi = 3.141592\dots$ has been investigated for many centuries. A real number is normal in base b if in its representation in base b every combination of digits occurs with proper frequency. If $a_1 a_2 \dots a_k$ is any combination of k digits and $N(t)$ is the number of times this combination occurs among the first t digits, the condition is that $\lim_{t\to\infty} N(t)/t = 1/b^k$ for the base b expansion. By the Birkhoff's ergodic theorem, it is easily shown that almost every number is normal. It is not known whether π is normal[12], although the first 29 million digits are very uniformly distributed[1]. More related results are collected in [2]. Many people believe that the decimal expansion of π is uniformly distributed.

It is interesting to investigate if the average of the logarithm of the first return time follows the theoretical prediction. See the formulas 1 and 2. Y. Kanada[5] at Tokyo University has computed digits of π for many years and has obtained the first $68,719,470,000$ decimal digits of π, which is the world record as of April 1999. We use the first $33,554,434$ decimal digits of π obtained from the computer program written by D. Takahashi who is a member of the research team led by Kanada. For other interesting numbers such as e and $\sqrt{2}$ we use R. Nemiroff and J. Bonnell's results[8].

2. Overlapping case

Overlapping algorithm is harder to analyze than nonoverlapping algorithm. The latter is studied in the next section.

FACT 2.1 ([6]). *For Bernoulli processes*

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

Since $E[\log R_n] - n \cdot h$ is close to $-\gamma$ for sufficiently large n it is recommended that we should approximate the entropy by the formula

$$h_{\text{approx.}} \equiv \frac{E[\log R_n] + \gamma}{n}. \tag{1}$$

Let x be the sequence from decimal expansion of π, $x = (314159\ldots)$. We estimate $E[\log R_n]$ by taking average at 10^6 sample paths $x, T^n x, \ldots, T^{(10^6-1)n}x$, which are obtained by shifting n times to reduce the correlation among each the sample values of $\log R_n$. Here the sample size is rather large to demonstrate the accuracy of the theoretical prediction and in practical applications a sample of small size will do.

The test results with sample size $1,000,000$ are given in Table 2, where Ave denotes the sample average and S.E.M. denotes the standard error of the mean of h_{approx}. The base of the logarithm used in these tests is 10. We do not have sufficient digits to simulate R_n for $n = 7$. We have similar simulations for e and $\sqrt{2}$ and the test results with sample size 100,000 are in Tables 2 and 3.

TABLE 1. Test result for π with sample size 10^6

n	Ave$[\log_{10} R_n]$	Ave$[\log_{10} R_n]/n$	$h_{\text{approx.}}$	$h_{\text{approx.}} - h$	S.E.M.
2	1.75612	0.88059	1.00593	0.00593	0.00026
3	2.75099	0.91700	1.00056	0.00056	0.00018
4	3.74932	0.93733	1.00000	-4.7×10^{-7}	0.00014
5	4.74885	0.94977	0.99991	−0.00009	0.00011
6	5.74906	0.95818	0.99996	−0.00004	0.00009

TABLE 2. Test result for e with sample size 10^5

n	Ave$[\log_{10} R_n]$	Ave$[\log_{10} R_n]/n$	$h_{\text{approx.}}$	$h_{\text{approx.}} - h$	S.E.M.
2	1.75980	0.87990	1.00524	0.00524	0.00083
3	2.75251	0.91750	1.00106	0.00106	0.00058
4	3.74795	0.93699	0.99966	−0.00034	0.00044
5	4.74795	0.94959	0.99973	−0.00027	0.00035

TABLE 3. Test result for $\sqrt{2}$ with sample size 10^5

n	Ave$[\log_{10} R_n]$	Ave$[\log_{10} R_n]/n$	$h_{\text{approx.}}$	$h_{\text{approx.}} - h$	S.E.M.
2	1.75996	0.87998	1.00532	0.00532	0.00083
3	2.75115	0.91705	1.00061	0.00061	0.00058
4	3.74880	0.93720	0.99987	−0.00013	0.00044
5	4.74789	0.94958	0.99971	−0.00029	0.00035

3. Nonoverlapping case

The non-overlapping case for Bernoulli processes was studied by Maurer[7] to test pseudorandom numbers. His algorithm corresponds to the *non-overlapping first return time*

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

TABLE 4. Non-overlapping test result for π with sample size 10^6

n	Ave$[\log_{10} R'_n]$	Ave$[\log_{10} R'_n]/n$	$h_{\text{approx.}}$	$h_{\text{approx.}} - h$	S.E.M.
2	1.76095	0.88048	1.00582	0.00582	0.00026
3	2.75131	0.91710	1.00066	0.00066	0.00018
4	3.74920	0.93730	0.99997	−0.00003	0.00014
5	4.74874	0.94975	0.99988	−0.00012	0.00011

Then for the Bernoulli process

$$\begin{aligned}
&\mu\{R'_n(x) = i | x_1 \cdots x_n = a_1 \cdots a_n\} \\
&= \prod_{j=1}^{i-1} \mu\{x_{jn+1} \cdots x_{jn+n} \neq a_1 \cdots a_n\} \cdot \mu\{x_{in+1} \cdots x_{in+n} = a_1 \cdots a_n\} \\
&= \mu\{a_1 \cdots a_n\}(1 - \mu\{a_1 \cdots a_n\})^{i-1}.
\end{aligned}$$

Let $v(r) \equiv r \sum_{i=1}^{\infty} (1-r)^{i-1} \log i$, $0 < r < 1$. Then the expectation of $\log R'_n$ equals $v(10^{-n})$ in case of the decimal Bernoulli process. Note that

$$\begin{aligned}
\lim_{r \to 0+} [v(r) + \log r] &= \lim_{s \to 1-} [v(1-s) + \log(1-s)] \\
&= \sum_{i=1}^{\infty} \left(\log \frac{i+1}{i} - \frac{1}{i} \right) \\
&= -\gamma \\
&= -0.577216 \cdots,
\end{aligned}$$

where $\gamma = \lim_{n \to \infty} (\sum_{i=1}^{n} (1/i) - \log n)$ is Euler's constant. Hence the expectation of $\log R'_n$ is approximately equal to $n - \gamma$ for large n. Like the overlapping case we use the formula

$$h_{\text{approx.}} \equiv \frac{E[\log R'_n] + \gamma}{n}. \tag{2}$$

By the same procedure we simulate $E[\log R'_n]$ on the decimal expansion of π. The test result with sample size $1,000,000$ is given in Table 3. As in the overlapping case the estimation is also good. But we could not test it for $n = 6$ in non-overlapping case since the method needs n times more digits of π than overlapping algorithm.

References

[1] D. Bailey, *The computation of π to 29,360,000 decimal digits using Borweins' quartically convergent algorithm*, Math. Comp. **50** (1988), 283–296.

[2] L. Berggren, J. Borwein and P. Borwein, *Pi: A Source Book*, Springer-Verlag, New York, 1997.

[3] G. H. Choe and D. H. Kim, *Average convergence rate of the first return time*, Colloq. Math. **84** (2000), 159–171.

[4] M. Kac, *On the notion of recurrence in discrete stochastic processes*, Bull. Amer. Math. Soc. **53** (1947), 1002-1010.

[5] Y. Kanda, `http://pi2.cc.u-tokyo.ac.jp`.

[6] D. H. Kim, *The recurrence of blocks for Bernoulli processes*, submitted.

[7] U. Maurer, *A universal statistical test for random bit generators*, J. Cryptology **5** (1992), 89–105.

[8] R. Nemiroff, `http://antwrp.gsfc.nasa.gov/htmltest/rjn.html`.

[9] D. Ornstein and B. Weiss, *Entropy and data compression schemes*, IEEE Trans. Inform. Theory **39** (1993), 78–83.

[10] C. Shannon, *The mathematical theory of communication*, Bell Sys. Tech. J. **27** (1948), 379–423 and 623–656.

[11] P. C. Shields, *The Ergodic Theory of Discrete Sample Paths (Graduate Studies in Math. Vol.* **13***)*, Amer. Math. Soc., 1996.

[12] S. Wagon, *Is π normal?*, Math. Intelligencer **7** (1985), 65–67.

[13] P. Walters, *An Introduction to Ergodic Theory, 2nd ed.*, Springer-Verlag, New York, 1981.

[14] 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.

[15] J. Ziv and A. Lempel, *A universal algorithm for sequential data compression*, IEEE Trans. Inform. Theory 23 (1977), 337–343.

Department of Mathematics
Korea Advanced Institute of Science and Technology
Taejon
E-mail: choe@euclid.kaist.ac.kr, kim@euclid.kaist.ac.kr