



On the covering by small random intervals

Ai-Hua Fan^{a,b,1}, Jun Wu^{a,b,2}

^a Department of Mathematics, Wuhan University, Wuhan, Hubei, 430072, PR China

^b LAMFA, CNRS, UMR 6140, University of Picardie Jules Verne, 33, rue Saint Leu, 80039, Amiens, France

Received 15 May 2002; accepted 15 July 2003

Abstract

Consider the random intervals $I_n = \omega_n + (0, \ell_n)$ (modulo 1) with their left points ω_n independently and uniformly distributed over the interval $[0, 1) = \mathbb{R}/\mathbb{Z}$ and with their lengths decreasing to zero. We prove that the Hausdorff dimension of the set $\overline{\lim}_n I_n$ of points covered infinitely often is almost surely equal to $1/\alpha$ when $\ell_n = a/n^\alpha$ for some $a > 0$ and $\alpha > 1$.

© 2003 Elsevier SAS. All rights reserved.

Résumé

Considérons des intervalles aléatoires $I_n = \omega_n + (0, \ell_n)$ (modulo 1) dont les extrémités gauches ω_n sont indépendantes et uniformément réparties sur l'intervalle $[0, 1) = \mathbb{R}/\mathbb{Z}$ et dont les longueurs décroissent vers zéro. Nous montrons que la dimension de Hausdorff de l'ensemble $\overline{\lim}_n I_n$ des points infiniment recouverts est presque sûrement égale à $1/\alpha$ quand $\ell_n = a/n^\alpha$ avec $a > 0$ et $\alpha > 1$.

© 2003 Elsevier SAS. All rights reserved.

MSC: 60D05; 52C17; 28A80

Keywords: Random covering; Uniform sample spacing; Hausdorff dimension

1. Introduction

Let $\{\ell_n\}_{n \geq 1}$ be a sequence of positive real numbers which is decreasing to zero and let $I_n(\omega) = (\omega_n, \omega_n + \ell_n)$ (modulo 1) be a random interval where $\{\omega_n\}_{n \geq 1}$ is a sequence of independent random variables uniformly distributed over the unit interval $I = [0, 1)$ which is identified with the circle \mathbb{R}/\mathbb{Z} . We consider the set $E_\infty(\omega) = \overline{\lim}_n I_n$ of those points which are covered infinitely often.

It is easy to see that $E_\infty(\omega)$ is almost surely (a.s. for short) a set of Lebesgue measure 0 or 1 according to $\sum_{n=1}^{\infty} \ell_n < \infty$ or $\sum_{n=1}^{\infty} \ell_n = \infty$. A. Dvoretzky [2] asked the question when $E_\infty(\omega) = [0, 1)$ a.s. or not. There was a series of contributions (for references and related topics see [5] for information before 1985 and [6]

E-mail addresses: ai-hua.fan@u-picardie.fr (A.-H. Fan), wujunyu@public.wh.hb.cn (J. Wu).

¹ The author was partially supported by the Grant of Key Laboratory—Optimal Control and Discrete Mathematics.

² The author is partly supported by the Special Funds for Major State Basic Research Projects of China.

for recent information). A complete answer was obtained by L. Shepp [7]: $E_\infty(\omega) = [0, 1)$ a.s. if and only if $\sum_{n=1}^{\infty} (1/n^2) \exp(\ell_1 + \ell_2 + \dots + \ell_n) = \infty$. It is the case for $\ell_n = a/n$ if and only if $a \geq 1$.

We address to the study of $E_\infty(\omega)$ when the covering intervals I_n are small in the sense $\sum_{n=1}^{\infty} \ell_n < \infty$. As we mentioned above, in this case the set $E_\infty(\omega)$ is of Lebesgue measure zero. However, it is a.s. of second category in Baire sense (see [5, p. 55]). We will determine the Hausdorff dimension of $E_\infty(\omega)$ in the case $\ell_n = a/n^\alpha$ with $a > 0, \alpha > 1$.

Theorem. Suppose $\ell_n = a/n^\alpha$ for some $a > 0$ and $\alpha > 1$. Then

$$\dim E_\infty(\omega) = \frac{1}{\alpha} \quad \text{a.s.}$$

As we shall see from the proof of the theorem, $\ell_n = O(n^{-\alpha})$ implies $\dim E_\infty(\omega) \leq 1/\alpha$ and $n^{-\alpha} = O(\ell_n)$ implies $\dim E_\infty(\omega) \geq 1/\alpha$. It follows that $\dim E_\infty(\omega) = 1/\alpha$ a.s. when the following limit exists:

$$1 < \alpha = \lim_{n \rightarrow \infty} \frac{-\log \ell_n}{\log n}.$$

We point out that a similar result holds for random coverings on trees [3].

Back to the theorem. The inequality $\dim E_\infty(\omega) \leq 1/\alpha$ is easy to see. It even holds for every ω . Because $\{I_n(\omega)\}_{n \geq N}$ is a δ -cover of $E_\infty(\omega)$ with $\delta = \ell_N$ and for any $\varepsilon > 0$

$$\sum_{n=N}^{\infty} |I_n|^{1/\alpha+\varepsilon} = a^{1/\alpha+\varepsilon} \sum_{n=N}^{\infty} n^{-1-\varepsilon\alpha} < \infty.$$

In order to prove the inverse inequality, we will construct a random Cantor subset of $E_\infty(\omega)$ by using known results due to D.A. Darling on random spacings of uniform random samples. Before our proof of the theorem, let us give some preliminaries including Darling's results and a construction of Cantor set.

2. Preliminaries

Let X_1, X_2, \dots, X_n ($n \geq 2$) be a set of independent random variables uniformly distributed over the unit interval $I = [0, 1)$. We call it a random sample of size n . Reordering the n points X_1, X_2, \dots, X_n in their natural order from left to right, we get n new random variables which will be denoted by $X_{(1)}, X_{(2)}, \dots, X_{(n)}$. The intervals $[X_{(k)}, X_{(k+1)})$, $0 \leq k \leq n$, are called the subspacings and their lengths are denoted by L_k , $0 \leq k \leq n$ (by convention, $X_{(0)} = 0$ and $X_{(n+1)} = 1$). There is a vast literature on the distributions of (L_0, L_1, \dots, L_n) and related statistics. We will only need the following results among others due to D.A. Darling [1]. Suppose $h: I \rightarrow \mathbf{R}$. Let

$$W_n = \sum_{j=0}^n h(L_j).$$

The first two moments of W_n are expressed by the following Darling formulas:

$$\mathbb{E}W_n = n(n+1) \int_0^1 (1-t)^{n-1} h(t) dt. \quad (1)$$

$$\mathbb{E}W_n^2 = n(n+1) \int_0^1 (1-t)^{n-1} h^2(t) dt + n^2(n^2-1) \iint_D (1-x-y)^{n-2} h(x)h(y) dx dy, \quad (2)$$

where $D = \{(x, y): x \geq 0, y \geq 0, x + y \leq 1\}$. We need to know how many subspacings with given length fall into a fixed subinterval. Let $J \subset I$ be a subinterval of length ℓ and let $0 < s_1 < s_2 < 1$. We denote by $M = M_n(\ell, s_1, s_2)$ the number of subspacings in J whose lengths are between s_1 and s_2 . Using the Darling formulas, J. Hawkes [4] obtained explicit expressions of the first two moments of M :

$$\mathbb{E}M = n\ell \left(\left(1 - \frac{s_1}{\ell}\right)(1 - s_1)^{n-1} - \left(1 - \frac{s_2}{\ell}\right)(1 - s_2)^{n-1} \right) + (1 - s_1)^n - (1 - s_2)^n. \tag{3}$$

$$\mathbb{E}M^2 = \mathbb{E}M + S(n, \ell, s_1) + S(n, \ell, s_2) - 2S(n, \ell, (s_1 + s_2)/2) \tag{4}$$

where

$$S(n, \ell, s) = n(n - 1)\ell^2 \left(1 - \frac{2s}{\ell}\right)^2 (1 - 2s)^{n-2} + 2n\ell \left(1 - \frac{2s}{\ell}\right)(1 - 2s)^{n-1}.$$

Proposition 1. *Suppose $0 < c_1 \leq c_2 < 1/2$, $0 < \ell < 1$ and $n \geq 3$. Let $J \subset [0, 1)$ be a subinterval of length ℓ and let n be the sample size. Denote by M the number of subspacings in J having length in $[\frac{c_1 \log n}{n}, \frac{c_2 \log n}{n}]$. Then there exist constants γ and C only depending on c_1 and c_2 (independent of ℓ and n) such that*

$$\mathbf{P}(M < \gamma \ell n^{1-c_1}) \leq \frac{C}{n^{c_2-c_1}}$$

for all n such that $n^{1-2(c_2-c_1)} \ell \geq \log^4 n$.

Proof. We claim that

$$\mathbb{E}M = \ell n^{1-c_1} + O(\ell n^{1-c_2}) \tag{5}$$

$$\mathbb{E}M^2 = \mathbb{E}M + \ell^2 n^{2(1-c_1)} + O(\ell^2 n^{2(1-c_2)}), \tag{6}$$

where, and in the sequel, the constants involved in $O(1)$ depend only on c_1 and c_2 and is independent with ℓ and n . First notice that

$$\left(1 - \frac{c \log n}{n}\right)^n = \frac{1}{n^c} \left(1 + O\left(\frac{\log^2 n}{n}\right)\right), \tag{7}$$

$$\frac{1 - \frac{c \log n}{\ell n}}{1 - \frac{c \log n}{n}} = 1 + O\left(\frac{\log n}{\ell n}\right), \tag{8}$$

$$\left(\frac{1 - \frac{c \log n}{\ell n}}{1 - \frac{c \log n}{n}}\right)^2 = 1 + O\left(\frac{\log n}{\ell^2 n}\right). \tag{9}$$

The equalities (8) and (9) hold under the condition $n\ell \geq c \log n$ which is ensured by the hypothesis made in the proposition. Let $s = \frac{c \log n}{n}$. Using (7) and (8), we get

$$\left(1 - \frac{s}{\ell}\right)(1 - s)^{n-1} = \frac{1}{n^c} \left(1 + O\left(\frac{\log^2 n}{\ell n}\right)\right).$$

Then, by the formula (3), we obtain

$$\mathbb{E}M = \ell n^{1-c_1} - \ell n^{1-c_2} + O(n^{-c_1} \log^2 n) = \ell n^{1-c_1} + O(\ell n^{1-c_2}).$$

Thus we have proved (5). Using (7) and (9), we get

$$\left(1 - \frac{2s}{\ell}\right)^2 (1 - 2s)^{n-2} = \frac{1}{n^{2c}} \left(1 + O\left(\frac{\log^4 n}{\ell^2 n}\right)\right).$$

Then

$$S(n, \ell, s) = n^{2(1-c)} \ell^2 + O(\ell n^{1-2c} \log^4 n).$$

Notice that $\ell n^{1-2c} \log^4 n$ is dominated by $n^{2(1-c)} \ell^2$ if $n\ell \geq \log^4 n$. So the main term in $S(n, \ell, s)$ is $n^{2(1-c)} \ell^2$. Also notice that $\ell n^{1-2c_1} \log^4 n$ is dominated by $n^{2(1-c_2)} \ell^2$ if $\ell n^{1-2(c_2-c_1)} \geq \log^4 n$ (this is the hypothesis). So we get (6).

As a consequence of (5) and (6), we have the following estimate of the variance of M :

$$\text{Var}(M) = \mathbb{E}M + O(\ell^2 n^{2-(c_1+c_2)}) = O(\ell^2 n^{2-(c_1+c_2)}).$$

By Chebyshev inequality,

$$\mathbf{P}\left(M \leq \frac{\mathbb{E}M}{2}\right) \leq \mathbf{P}\left(|M - \mathbb{E}M| > \frac{\mathbb{E}M}{2}\right) \leq \frac{4\text{Var}(M)}{(\mathbb{E}M)^2} = O\left(\frac{1}{n^{c_2-c_1}}\right). \quad \square$$

Consider now a construction of generalized Cantor sets on $[0, 1)$. Let $\{n_k\}_{k \geq 1}$ be a sequence of integers satisfying $n_k \geq 2$. Let $\{\rho_k\}_{k \geq 1}$ and $\{d_k\}_{k \geq 1}$ be two sequences of positive real numbers. Assume that for any $k \geq 1$, we have a collection \mathcal{J}_k of closed subintervals of $[0, 1)$. Each interval in \mathcal{J}_k is called a k -interval. Suppose

- (1) Each k -interval is of length ρ_k and contains n_{k+1} intervals;
- (2) Each $(k + 1)$ -interval is contained in some k -interval;
- (3) The gap between any two k -intervals is at least d_k .

Let $C_n = \bigcup_{J \in \mathcal{J}_k} J$ and $C_\infty = \bigcap_{k=1}^\infty C_n$. We call C_∞ a generalized Cantor set.

Proposition 2. Consider the generalized Cantor set C_∞ constructed above. Suppose that there is a number $a \geq 1$ such that $n_{k+1}d_{k+1} \geq \rho_k^a$ ($\forall k \geq 1$). Then we have

$$\dim C_\infty \geq \liminf_{k \rightarrow \infty} \frac{\log(n_1 n_2 \cdots n_k)}{-a \log \rho_k}.$$

Proof. Define a probability measure μ on $[0, 1)$ (concentrated on C_∞) by

$$\mu(J_k) = \frac{1}{n_1 n_2 \cdots n_k},$$

where J_k represents an arbitrary k -interval contained in C_k . Let s be the lim inf. Since $n_1 n_2 \cdots n_k \rho_k \leq 1$, we have $s \leq 1/a \leq 1$. Suppose $s > 0$. By the Frostman lemma, we have only to prove that for any $0 < t < s$ and any open interval U we have $\mu(U) \leq 2|U|^t$ ($|U|$ denotes the length of U). Without loss of generality, we assume that $n_1 n_2 \cdots n_k \rho_k^{at} \geq 1$ for all $k \geq 1$. Choose k_0 such that $\rho_{k_0+1} \leq |U| < \rho_{k_0}$. We distinguish two cases:

- (a) The case $|U| < d_{k_0+1}$. Then U intersects with at most one $(k_0 + 1)$ -interval. So

$$\mu(U) \leq \frac{1}{n_1 n_2 \cdots n_{k_0+1}} \leq \rho_{k_0+1}^{at} \leq |U|^{at} \leq |U|^t.$$

- (b) The case $|U| \geq d_{k_0+1}$. Then U intersects with at most $\min(n_{k_0+1}, \frac{2|U|}{d_{k_0+1}})$ $(k_0 + 1)$ -intervals. So

$$\begin{aligned} \mu(U) &\leq \frac{1}{n_1 n_2 \cdots n_{k_0+1}} \min\left(n_{k_0+1}, \frac{2|U|}{d_{k_0+1}}\right) \\ &\leq \frac{1}{n_1 n_2 \cdots n_{k_0+1}} n_{k_0+1}^{1-t} \left(\frac{2|U|}{d_{k_0+1}}\right)^t \\ &\leq \frac{2}{n_1 n_2 \cdots n_{k_0} (n_{k_0+1} d_{k_0+1})^t} |U|^t \leq \frac{2}{n_1 n_2 \cdots n_{k_0} \rho_{k_0}^{at}} |U|^t \leq 2|U|^t. \quad \square \end{aligned}$$

3. Proof of theorem

We only consider the case $\ell_n = a/n^\alpha$ with $a = 1$. As we shall see, only the order α of n^α plays the role. So, we may also assume that I_n is the closed interval $\omega_n + [0, \ell_n]$.

Fix two constants $0 < c_1 < c_2 < 1/2$ verifying the condition of Proposition 1. Take a large integer Δ . Define $m_k = \Delta^k$ ($k = 1, 2, \dots$). For $k \geq 1$, let

$$\begin{aligned} \rho_k &= \ell_{2^{m_k+1}}, \quad d_k = \frac{c_1 m_k \log 2}{2^{m_k}}, \\ n_k &= \left\lfloor \frac{\gamma \ell_{\rho_{k-1}} 2^{m_k(1-c_1)}}{2} \right\rfloor - 1 \quad (\rho_0 = 1), \\ q_k &= 1 - C \frac{\prod_{j=1}^{k-1} n_j}{2^{m_k(c_2-c_1)}} \end{aligned}$$

where $[x]$ denotes the integral part of a real number x and the constants γ and C are those in Proposition 1.

Consider the random sample of size 2^{m_1} from the uniform distribution over $[0, 1)$: $\omega_{2^{m_1}}, \omega_{2^{m_1+1}}, \dots, \omega_{2^{m_1+1}-1}$. Applying Proposition 1 with $n = 2^{m_1}$ and $\ell = 1$. Proposition 1 is applicable if Δ is large enough so that $2^{m_1(1-2(c_2-c_1))} \geq \log^4 2^{m_1}$. We assume that $\gamma 2^{m_1(1-c_1)} > 7$ so that $n_1 \geq 2$. Let \mathcal{L}_1 be the set of left points of subsampling intervals contained in $J = [0, 1)$ having length in $[\frac{c_1 m_1 \log 2}{2^{m_1}}, \frac{c_2 m_1 \log 2}{2^{m_1}}]$. By Proposition 1, we have

$$\mathbf{P}(\#\mathcal{L}_1 \geq \gamma 2^{m_1(1-c_1)}) > 1 - \frac{C}{2^{m_1(c_2-c_1)}} = q_1.$$

Thus with probability q_1 we can find a set $\mathcal{L}_1^* \subset \mathcal{L}_1$ with n_1 points such that for each point in \mathcal{L}_1^* there is on its right side a point in $\mathcal{L}_1 \setminus \mathcal{L}_1^*$. So, any two points in \mathcal{L}_1^* has a distance at least $2d_1$. Define

$$C_1 = \bigcup_{\omega \in \mathcal{L}_1^*} [\omega, \omega + \rho_1].$$

Notice that there are n_1 (≥ 2) intervals in C_1 each of which has length ρ_1 and that these intervals are separated by a distance at least d_1 .

Suppose that with probability $q_1 q_2 \cdots q_k$ we have successively constructed a nested sequence of sets $C_1 \supset C_2 \supset \cdots \supset C_k$ such that

- (i) every C_j ($1 \leq j < k$) is a union of disjoint closed intervals and each such interval in C_j is of length ρ_j and contains n_{j+1} intervals contained in C_{j+1} , and every interval contained in C_{j+1} is a subset of C_j ;
- (ii) the gap between two intervals contained in C_{j+1} is at least d_{j+1} .

We now construct C_{k+1} . Consider the random sample of size $2^{m_{k+1}}$: $\omega_{2^{m_{k+1}}}, \omega_{2^{m_{k+1}+1}}, \dots, \omega_{2^{m_{k+1}+1}-1}$. This sample is independent of all preceding random samples in the construction of C_1, C_2, \dots, C_k since $2^{m_{k+1}} - 1 < 2^{m_k}$. Apply Proposition 1 to each interval J contained in C_k with $n = 2^{m_{k+1}}$ and $\ell = \rho_k = \ell_{2^{m_{k+1}}}$. Notice that

$$2^{m_{k+1}(1-2(c_2-c_1))} \ell_{2^{m_{k+1}}} = 2^{-\alpha+\Delta^k(\Delta((1-2(c_2-c_1))-\alpha))} \geq \log^4 2^{m_{k+1}}$$

if Δ is large enough. So we can really apply Proposition 1. Thus if $\mathcal{L}_{k+1,J}$ denote the set of left points of subspacings contained in J having length in $[\frac{c_1 m_{k+1} \log 2}{2^{m_{k+1}}}, \frac{c_2 m_{k+1} \log 2}{2^{m_{k+1}}}]$, we have

$$\mathbf{P}(\#\mathcal{L}_{k+1,J} \leq \gamma \rho_k 2^{m_{k+1}(1-c_1)} \text{ for some } J \subset C_k) \leq C \cdot \frac{\prod_{j=1}^k n_j}{2^{m_{k+1}(c_2-c_1)}}.$$

In other words,

$$\mathbf{P}(\#\mathcal{L}_{k+1,J} > \gamma \rho_k 2^{m_{k+1}(1-c_1)} \text{ for all } J \subset C_k) \geq q_{k+1}$$

where J denotes a typical interval in C_k . For each J in C_k , take a set $\mathcal{L}_{k+1,J}^*$ of n_{k+1} points from $\mathcal{L}_{k+1,J}$ such that for each point in $\mathcal{L}_{k+1,J}^*$ there is on its right side a point in $\mathcal{L}_{k+1,J} \setminus \mathcal{L}_{k+1,J}^*$. Then construct

$$C_{k+1} = \bigcup_{J \subset C_k} \bigcup_{\omega \in \mathcal{L}_{k+1,J}^*} [\omega, \omega + \rho_{k+1}]$$

where $J \subset C_k$ means that J is a component of C_k . Thus with probability $q_1 q_2 \dots q_{k+1}$ we have constructed a nested sequence of sets $C_1 \supset C_2 \supset \dots \supset C_{k+1}$ which have the properties described by (i) and (ii) (see above, k being replaced by $k + 1$). Thus by induction we get an infinite sequence of nested sets C_k and we can construct a Cantor set $C_\infty = \bigcap_{k=1}^\infty C_k$ with probability

$$p = \prod_{k=1}^\infty q_k > 0.$$

The positivity of this probability is the consequence of

$$\sum_{k=1}^\infty (1 - q_k) \leq C \sum_{k=1}^\infty \frac{\prod_{j=1}^k n_j}{2^{m_{k+1}(c_2-c_1)}} < \infty,$$

because the general term of the series is bounded by

$$\gamma^k \left(\prod_{i=1}^{k-1} \rho_i \right) 2^{(m_1+m_2+\dots+m_k)(1-c_1)-m_{k+1}(c_2-c_1)} = O\left(\gamma^k 2^{\Delta^k(\frac{1-c_1}{\Delta-1}-(c_2-c_1))}\right).$$

By the construction, with probability $p > 0$ we have $C_\infty \subset E_\infty(\omega)$. Actually C_∞ is infinitely covered by those intervals I_n with $2^{m_k} \leq n < 2 \cdot 2^{m_k}$ for some $k \geq 1$.

Let us apply Proposition 2 to estimate the Hausdorff dimension of C_∞ from below. Notice that $\rho_k = 2^{-\alpha(\Delta^k+1)}$ and

$$n_{k+1} d_{k+1} \approx m_{k+1} 2^{-\alpha m_{k+1}(1-c_1)-m_{k+1}} = \Delta^{k+1} 2^{-\Delta^k(\alpha+c_1\Delta)}.$$

For any $a > 1$ and small $c_1 > 0$ so that $c_1 \Delta$ is small, the condition $n_{k+1} d_{k+1} \geq \rho_k^a$ is satisfied. Also notice that

$$\lim_{k \rightarrow \infty} \frac{\log(n_1 n_2 \dots n_k)}{-\log \rho_k} = \frac{1}{\alpha} (1 - c_1) \frac{\Delta}{\Delta - 1} - \frac{1}{\Delta - 1}.$$

Thus with probability $p > 0$,

$$\dim E_\infty(\omega) \geq \dim C_\infty \geq \frac{1}{\alpha} (1 - c_1) \frac{\Delta}{\Delta - 1} - \frac{1}{\Delta - 1}.$$

Since $E_\infty(\omega)$ is a tail event, we have with probability one

$$\dim E_\infty(\omega) \geq \frac{1}{a\alpha}(1 - c_1) \frac{\Delta}{\Delta - 1} - \frac{1}{\Delta - 1}.$$

Let $c_1 \rightarrow 0$, $\Delta \rightarrow \infty$ and then $a \rightarrow 1$, we get $\dim E_\infty(\omega) \geq 1/\alpha$ a.s.

References

- [1] D.A. Darling, On a class of problems related to the random division of an interval, *Ann. Math. Statistics* 24 (1953) 239–253.
- [2] A. Dvoretzky, On covering a circle by randomly placed arcs, *Proc. Nat. Acad. Sci. USA* 42 (1956) 199–203.
- [3] A.H. Fan, Oscillations of tree-indexed walks, LAMFA Prepublication, University of Picardie, 2000.
- [4] J. Hawkes, On the asymptotic behaviour of sample spacings, *Math. Proc. Cambridge Philos. Soc.* 90 (2) (1981) 293–303.
- [5] J.P. Kahane, *Some Random Series of Functions*, second ed., in: *Cambridge Stud. Adv. Math.*, vol. 5, Cambridge Univ. Press, Cambridge, 1985.
- [6] J. Kahane, Random coverings and multiplicative processes, in: *Fractal Geometry and Stochastics, II (Greifswald/Koserow, 1998)*, in: *Progr. Probab.*, vol. 46, Birkhäuser, Basel, 2000, pp. 125–146.
- [7] L. Shepp, Covering the circle with random arcs, *Israel J. Math.* 11 (1972) 328–345.