Statistics and Probability Letters 82 (2012) 2103–2107

Contents lists available at SciVerse ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

The connectivity threshold of random geometric graphs with Cantor distributed vertices

Antar Bandyopadhyay *, Farkhondeh Sajadi

Theoretical Statistics and Mathematics Unit, Indian Statistical Institute, Delhi Centre, 7 S. J. S. Sansanwal Marg, New Delhi 110016, India

ARTICLE INFO

Article history:
Received 2 April 2012
Received in revised form 24 July 2012
Accepted 24 July 2012
Available online 31 July 2012

MSC:
primary 60D05
05C80
secondary 60F15
60F25
60G70

Keywords:
Cantor distribution
Connectivity threshold
Random geometric graph
Singular distributions

ABSTRACT

For the connectivity of *random geometric graphs*, where there is no density for the underlying distribution of the vertices, we consider n i.i.d. *Cantor* distributed points on $[0, 1]$. We show that for such a random geometric graph, the connectivity threshold, R_n, converges almost surely to a constant $1 - 2\phi$ where $0 < \phi < 1/2$, which for the standard Cantor distribution is $1/3$. We also show that $\|R_n - (1 - 2\phi)\|_1 \sim 2\,C(\phi)\,n^{-1/d_\phi}$ where $C(\phi) > 0$ is a constant and $d_\phi := -\log 2/\log\phi$ is the *Hausdorff dimension* of the generalized Cantor set with parameter ϕ.

1. Introduction

1.1. Background and motivation

A *random geometric graph* consists of a set of vertices, distributed randomly over some metric space, in which two distinct such vertices are joined by an edge if the distance between them is sufficiently small. More precisely, let V_n be a set of n points in $\mathbb{R}^d$, distributed independently according to some distribution F on $\mathbb{R}^d$. Let r be a fixed positive real number. Then, the random geometric graph $\mathcal{G} = \mathcal{G}(V_n, r)$ is a graph with vertex set V_n where two vertices $\mathbf{v} = (v_1, \ldots, v_d)$ and $\mathbf{u} = (u_1, \ldots, u_d)$ in V_n are adjacent if and only if $\|\mathbf{v} - \mathbf{u}\| \le r$ where $\|.\|$ is some norm on $\mathbb{R}^d$.

A considerable amount of work has been done on the *connectivity threshold* defined as

$$R_n = \inf\{r > 0 \mid \mathcal{G}(V_n, r) \text{ is connected}\}. \tag{1}$$

For the case where the vertices are assumed to be uniformly distributed in $[0, 1]^d$, Appel and Russo (2002) showed that with probability 1,

$$\lim_{n\to\infty} \frac{n}{\log n} R_n^d = \begin{cases} 1 & \text{for } d = 1, \\ \dfrac{1}{2d} & \text{for } d \ge 2 \end{cases} \tag{2}$$

* Corresponding author. Tel.: +91 11 4149 3932; fax: +91 11 4149 3981.
E-mail addresses: antar@isid.ac.in (A. Bandyopadhyay), farkhondeh.sajadi@gmail.com (F. Sajadi).
URL: http://www.isid.ac.in/~antar/ (A. Bandyopadhyay).

0167-7152/$ – see front matter
doi:10.1016/j.spl.2012.07.015

when the norm $\|.\|$ is taken to be the $\mathcal{L}_\infty$ or the sup norm. Later, Penrose (2003) showed that the limit in (2) holds but with different constants for any $\mathcal{L}_p$ norm for $1 \le p \le \infty$. Penrose (1999) considered the case where the distribution F has a continuous density f with respect to the Lebesgue measure which remains bounded away from 0 on the support of F. Under certain technical assumptions such as that of a smooth boundary for the support, he showed that with probability 1,

$$\lim_{n\to\infty} \frac{n}{\log n} R_n^d = C$$

where C is an explicit constant which depends on the dimension d and the essential infimum of f and its value on the boundary of the support. Recently, Sarkar and Saurabh (2010) (a personal communication), studied a case where the density f of the underlying distribution may have minimum zero. They proved, in particular, that when the support of f is [0, 1] and f is bounded below on any compact subset not containing the origin but is regularly varying at the origin, then $R_n/F^{-1}(1/n)$ has a weak limit.

The proof by Sarkar and Saurabh (2010) can easily be generalized to the case where the density is zero at finitely many points. A question then naturally arises: that of what happens to the case where the distribution function is flat on some intervals, that is, if a density exists then it will be zero on some intervals. Also, the question arises of what happens in the somewhat extreme case, where the density may not exist even though the distribution function is continuous and has flat parts. To consider these questions, in this work we study the connectivity of random geometric graphs where the underlying distribution of the vertices has no mass and is also singular with respect to the Lebesgue measure, that is, it has no density. For this purpose, we consider the *generalized Cantor distribution* with parameter ϕ denoted by Cantor(ϕ) as the underlying distribution of the vertices of the graph. The distribution function is then flat on infinitely many intervals. We will show that the connectivity threshold converges almost surely to the length of the longest flat part of the distribution function and we also provide some finer asymptotics for the same case.

1.2. Preliminaries

In this subsection, we discuss the *Cantor set* and the *Cantor distribution* which is defined on it.

1.2.1. The Cantor set

The Cantor set was first discovered by Smith (1875) but became popular after Cantor (1883). The standard Cantor set is constructed on the interval [0, 1] as follows. One successively removes the open middle third of each subinterval of the previous set. More precisely, starting with $C_0 := [0, 1]$, we inductively define

$$C_{n+1} := \bigcup_{k=1}^{2^n} \left(\left[a_{n,k}, a_{n,k} + \frac{b_{n,k} - a_{n,k}}{3} \right] \cup \left[b_{n,k} - \frac{b_{n,k} - a_{n,k}}{3}, b_{n,k} \right] \right)$$

where $C_n := \cup_{k=1}^{2^n} \left[a_{n,k}, b_{n,k} \right]$. The standard Cantor set is then defined as $C = \bigcap_{n=o}^{\infty} C_n$. It is known that the Hausdorff dimension of the standard Cantor set is $\frac{\log 2}{\log 3}$ (see Theorem 2.1 of Chapter 7 of Stein and Shakarchi, 2005).

For constructing the generalized Cantor set, we start with the unit interval [0, 1] and at the first stage, we delete the interval $(\phi, 1-\phi)$ where $0 < \phi < 1/2$. Then, this procedure is reiterated with the two segments $[0, \phi]$ and $[1-\phi, 1]$. We continue ad infinitum. The Hausdorff dimension of this set is given by $d_\phi := -\frac{\log 2}{\log \phi}$ (see Exercise 8 of Chapter 7 of Stein and Shakarchi, 2005). Note that the standard Cantor set is a special case when $\phi = 1/3$.

1.2.2. The Cantor distribution

The *Cantor distribution* with parameter ϕ where $0 < \phi < 1/2$ is the distribution of a random variable X defined by

$$X = \sum_{i=1}^{\infty} \phi^{i-1} Z_i \tag{3}$$

where the Z_i are i.i.d. with $\mathbb{P}[Z_i = 0] = \mathbb{P}[Z_i = 1-\phi] = 1/2$. If a random variable X admits a representation of the form (3) then we will say that X has a Cantor distribution with parameter ϕ, and write $X \sim$ Cantor(ϕ). Observe that Cantor (ϕ) is self-similar, in the sense that

$$X \stackrel{d}{=} \begin{cases} \phi X & \text{with probability } 1/2 \\ \phi X + 1 - \phi & \text{with probability } 1/2. \end{cases} \tag{4}$$

This follows easily by conditioning on Z_1. It is worth noting here that if $X \sim$ Cantor(ϕ) then the same is true of $1 - X$.

Note that for $\phi = 1/3$ we obtain the *standard Cantor distribution*.

2. The main results

Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed random variables with the Cantor(ϕ) distribution on [0, 1]. Given the graph $\mathcal{G} = \mathcal{G}(V_n, r)$, where $V_n = \{X_1, X_2, \ldots, X_n\}$, let R_n be defined as in (1).

Theorem 1. *For any* $0 < \phi < 1/2$, *as* $n \longrightarrow \infty$ *we have*

$$R_n \longrightarrow 1 - 2\phi \quad a.s. \tag{5}$$

Our next theorem gives finer asymptotics but before we state the theorem, we provide here some basic notation and facts. Let $m_n := \min\{X_1, X_2, \ldots, X_n\}$. Using (4) we get

$$m_n \stackrel{d}{=} \begin{cases} \phi m_k & \text{with probability } 2^{-n}\binom{n}{k} \text{ for } k = 1, 2, \ldots, n \\ \phi m_n + 1 - \phi & \text{with probability } 2^{-n}. \end{cases} \tag{6}$$

Let $a_n := \mathbb{E}[m_n]$. Using (6), Hosking (1994) derived the following recursion formula for the sequence (a_n):

$$\left(2^n - 2\phi\right) a_n = 1 - \phi + \phi \sum_{k=1}^{n-1} \binom{n}{k} a_k, \quad n \geq 1. \tag{7}$$

Moreover, Knopfmacher and Prodinger (1996) showed that whenever $0 < \phi < 1/2$, then as $n \to \infty$,

$$\frac{a_n}{n^{-\frac{1}{d_\phi}}} \longrightarrow C(\phi), \tag{8}$$

where

$$C(\phi) := \frac{(1-\phi)(1-2\phi)}{\phi \log 2} \Gamma(-\log_2 \phi) \zeta(-\log_2 \phi), \tag{9}$$

and $d_\phi = -\frac{\log 2}{\log \phi}$ is the Hausdorff dimension of the generalized Cantor set. Here $\Gamma(\cdot)$ and $\zeta(\cdot)$ are the Gamma and Riemann zeta functions, respectively.

Our next theorem gives the rate convergence of R_n to $(1 - 2\phi)$ in terms of the $\mathcal{L}_1$ norm.

Theorem 2. *For any* $0 < \phi < 1/2$, *as* $n \longrightarrow \infty$ *we have*

$$\frac{\|R_n - (1-2\phi)\|_1}{n^{-\frac{1}{d_\phi}}} \longrightarrow 2C(\phi), \tag{10}$$

where $C(\phi)$ *is as in Eq.* (9) *and* $\|\cdot\|_1$ *is the* $\mathcal{L}_1$ *norm.*

3. Proofs of the theorems

3.1. Proof of Theorem 1

We draw a sample of size n from Cantor(ϕ) on [0, 1]. Let N_n be the number of elements falling in the subinterval $[0, \phi]$ and $n - N_n$ the number of elements falling in $[1 - \phi, 1]$. From the construction, we have $N_n \sim \text{Bin}\left(n, \frac{1}{2}\right)$. In selecting this sample of size n, there are three cases which may happen. Some of these points may fall in the interval $[0, \phi]$ and the rest in the interval $[1 - \phi, 1]$. That means that $N_n \notin \{0, n\}$. In this case, the distance between the points in $[0, \phi]$ and $[1 - \phi, 1]$ is at least $1 - 2\phi$. The other cases are those where all points fall in $[0, \phi]$ or all fall in $[1 - \phi, 1]$, which in this case are those with $N_n = n$ or $N_n = 0$. Let $m_n = \min_{1 \leq i \leq n} X_i$, $M_n = \max_{1 \leq i \leq n} X_i$ and define

$$L_n := \max\{X_i | 1 \leq i \leq n \text{ and } X_i \in [0, \phi]\} \tag{11}$$

and

$$U_n := \min\{X_i | 1 \leq i \leq n \text{ and } X_i \in [1 - \phi, 1]\}. \tag{12}$$

We will take $L_n = 0$ (and similarly $U_n = 0$) if the corresponding set is empty.

Now find a $K \equiv K(\phi)$ such that $\phi^K < \frac{1}{2}(1-\phi)(1-2\phi)$. Note that such a $K < \infty$ exists since $0 < \phi < 1$. Let $I_1, I_2, \ldots, I_{2^K}$ be the 2^K subintervals of length ϕ^K which are part of the Kth stage of the "removal of the middle interval" for obtaining the generalized Cantor set with parameter ϕ. For $1 \leq j \leq 2^K$ define $N_j := \sum_{i=1}^{n} \mathbf{1}\left(X_i \in I_j\right)$, which is the number of sample points in the subinterval I_j. From the construction of the generalized Cantor distribution with parameter ϕ, it follows that

$$\boldsymbol{N}_K := \left(N_1, N_2, \ldots, N_{2^K}\right) \sim \text{Multinomial}\left(n; \left(\frac{1}{2^K}, \frac{1}{2^K}, \ldots, \frac{1}{2^K}\right)\right), \tag{13}$$

and $N_n = \sum_{I_j \subseteq [0,\phi]} N_j$. Consider the event $E_n := \cap_{j=1}^{2^k} \left[N_j \geq 1\right]$. Observe that for the event E_n the maximum distance between two points in $[0, \phi]$ as well as in $[1-\phi, 1]$ is at most $2\phi^K + \phi(1-2\phi) < 1-2\phi$ by the choice of K. Thus for E_n we must have $R_n = U_n - L_n$ and so we can write

$$R_n = (U_n - L_n)\ \mathbf{1}_{E_n} + R_n^* \mathbf{1}_{E_n^c} \tag{14}$$

where R_n^* is a random variable such that $0 < R_n^* < \phi$ a.s.

Observe that conditioned on $\left[N_1 = r_1, N_2 = r_2, \ldots, N_{2^k} = r_{2^k}\right]$, we have $U_n \stackrel{d}{=} 1-\phi+\phi m_{n-k}$ and $L_n \stackrel{d}{=} \phi M_k$, and $N_n = k$ where $k = \sum_{I_j \subseteq [0,\phi]} r_j$. More generally,

$$\left(\left(L_n, U_n\right), \boldsymbol{N}_K\right)_{n\geq 1} \stackrel{d}{=} \left(\left(\phi M_{N_n}, 1-\phi+\phi m_{n-N_n}\right), \boldsymbol{N}_K\right)_{n\geq 1}. \tag{15}$$

Note that for technical correctness, we define $M_0 = m_0 = 0$.

Now it is easy to see that $m_n \longrightarrow 0$ and $M_n \longrightarrow 1$ a.s. But by the SLLN, $N_n/n \longrightarrow 1/2$ a.s.; thus (N_n) and $(n - N_n)$ are two subsequences which are converging to infinity a.s. Moreover,

$$\mathbb{P}\left(E_n^c\right) \leq \sum_{j=1}^{2^K} \mathbb{P}\left(N_j = 0\right) = 2^K \left(1 - \frac{1}{2^K}\right)^n = 2^K \exp\left(-\alpha_K n\right), \tag{16}$$

where $\alpha_K = -\log\left(1 - \frac{1}{2^K}\right) > 0$. Thus $\sum_{n=1}^{\infty} \mathbb{P}\left(E_n^c\right) < \infty$, so by the first Borel–Cantelli lemma we have

$$\mathbb{P}\left(E_n^c \text{ infinitely often}\right) = 0 \Rightarrow \mathbb{P}\left(E_n \text{ eventually}\right) = 1.$$

In other words $\mathbf{1}_{E_n} \longrightarrow 1$ a.s. and $\mathbf{1}_{E_n^c} \longrightarrow 0$ a.s. Finally observing that $0 \leq R_n^* \leq \phi$ we get from Eqs. (14) and (15)

$$R_n \longrightarrow (1-2\phi). \quad \square$$

3.2. Proof of Theorem 2

We start by observing that

$$\begin{aligned}
\mathbb{E}\left[|R_n - (1-2\phi)|\right] &= \mathbb{E}\left[(R_n - (1-2\phi))\,\mathbf{1}_{E_n}\right] + \mathbb{E}\left[\left|R_n^* - (1-2\phi)\right|\,\mathbf{1}_{E_n^c}\right] \\
&= \mathbb{E}\left[(U_n - L_n - (1-2\phi))\,\mathbf{1}_{E_n}\mathbf{1}_{K\leq N_n\leq n-K}\right] + \mathbb{E}\left[\left|R_n^* - (1-2\phi)\right|\,\mathbf{1}_{E_n^c}\right] \\
&= \mathbb{E}\left[(U_n - L_n - (1-2\phi))\,\mathbf{1}_{K\leq N_n\leq n-K}\right] \\
&\quad - \mathbb{E}\left[(U_n - L_n - (1-2\phi))\,\mathbf{1}_{E_n^c}\mathbf{1}_{K\leq N_n\leq n-K}\right] + \mathbb{E}\left[\left|R_n^* - (1-2\phi)\right|\,\mathbf{1}_{E_n^c}\right] \\
&= \mathbb{E}\left[(U_n - L_n - (1-2\phi))\,\mathbf{1}_{1\leq N_n\leq n-1}\right] \\
&\quad - \mathbb{E}\left[(U_n - L_n - (1-2\phi))\,\mathbf{1}_{E_n^c}\mathbf{1}_{1\leq N_n\leq n-1}\right] + \mathbb{E}\left[\left|R_n^* - (1-2\phi)\right|\,\mathbf{1}_{E_n^c}\right].
\end{aligned} \tag{17}$$

In the above the first equality holds because of (14) and the fact that for the event E_n we must have $R_n > 1-2\phi$. The second, third and fourth equalities follow from the simple fact that $E_n \subseteq [K \leq N_n \leq n-K]$.

Now recall that $a_n = \mathbb{E}[m_n]$, so for the first part of the right-hand side of the Eq. (17) we can write

$$\begin{aligned}
\mathbb{E}\left[(U_n - L_n - (1-2\phi))\,\mathbf{1}_{1\leq N_n\leq n-1}\right] &= \frac{\phi}{2^n}\sum_{k=1}^{n-1}\binom{n}{k}\left(a_{n-k} + a_k\right) \\
&= \frac{1}{2^{n-1}}\left[\left(2^n - 2\phi\right)a_n - (1-\phi)\right],
\end{aligned} \tag{18}$$

where the last equality follows from (7). The other two parts of the right-hand side of the Eq. (17) are bounded in absolute value by

$$\mathbb{P}\left(E_n^c\right) \leq 2^K \exp\left(-\alpha_K n\right)$$

because of (16). Now observe that from Eq. (8) we get that $a_n \sim C(\phi)\, n^{-\frac{1}{d_\phi}}$ where $d_\phi = -\frac{\log 2}{\log \phi}$ is the Hausdorff dimension of the generalized Cantor set. Thus using (17) and (18) we conclude that

$$\frac{\mathbb{E}\left[|R_n - (1-2\phi)|\right]}{a_n} \longrightarrow 2 \quad \text{a.s. } n \longrightarrow \infty.$$

This completes the proof using (8). $\square$

4. Final remarks

It is worth noting here that our proofs depend on the recursive nature of the generalized Cantor distribution (see Eq. (4)). Thus, unfortunately, they do not have obvious extensions to other singular distributions. It will be interesting to derive a version of Theorem 1 for a general singular distribution with no mass and flat parts. Intuitively it seems that the final limit should be the length of the longest flat part. It will be more interesting if Theorem 2 can also be generalized to general singular distributions with no mass and flat parts, where $(1 - 2\phi)$ is replaced by the length of the longest flat part and d_ϕ is replaced by the Hausdorff dimension of the support.

Acknowledgment

The authors would like to thank the anonymous referee for pointing out a glitch in the first proof of Theorem 1 which was later corrected and also for his/her remarks which have led to improvement of the exposition.

References

Appel, M.J.B., Russo, R.P., 2002. The connectivity of a graph on uniform points on $[0, 1]^d$. Statist. Probab. Lett. 60 (4), 351–357.
Cantor, G., 1883. Über unendliche, lineare punktmannigfaltigkeiten v, [on infinite, linear point-manifolds (sets)]. Math. Ann. 21, 545–591.
Hosking, J.R.M., 1994. Moments of order statistics of the Cantor distribution. Statist. Probab. Lett. 19 (2), 161–165.
Knopfmacher, A., Prodinger, H., 1996. Explicit and asymptotic formulae for the expected values of the order statistics of the Cantor distribution. Statist. Probab. Lett. 27 (2), 189–194.
Penrose, M.D., 1999. A strong law for the largest nearest-neighbour link between random points. J. Lond. Math. Soc. (2) 60 (3), 951–960.
Penrose, M., 2003. Random Geometric Graphs. In: Oxford Studies in Probability, vol. 5. Oxford University Press, Oxford.
Sarkar, A., Saurabh, B., 2010. Connectivity distance in random geometric graph. Unpublished results.
Smith, H.J., 1875. On the integration of discontinuous functions. Proc. Lond. Math. Soc. 6, 140–153.
Stein, E.M., Shakarchi, R., 2005. Real Analysis. In: Princeton Lectures in Analysis, vol. III. Princeton University Press, Princeton, NJ.