isid/ms/2012/06
April 3, 2012
http://www.isid.ac.in/~statmath/eprints

Connectivity Threshold of Random Geometric Graphs with Cantor Distributed Vertices

Antar Bandyopadhyay
Farkhondeh Sajadi

Indian Statistical Institute, Delhi Centre
7, SJSS Marg, New Delhi–110 016, India

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

April 2, 2012

Abstract

For connectivity of *random geometric graphs*, where there is no density for underlying distribution of the vertices, we consider n i.i.d. *Cantor* distributed points on $[0, 1]$. We show that for this random geometric graph, the connectivity threshold R_n, converges almost surely to a constant $1-2\phi$ where $0 < \phi < 1/2$, which for 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 a the *Hausdorff dimension* of the generalized Cantor set with parameter ϕ.

Keywords: Cantor distribution, connectivity threshold, random geometric graph, singular distributions.

AMS 2010 Subject Classification: Primary: 60D05, 05C80; Secondary: 60F15, 60F25, 60G70

*E-Mail: `antar@isid.ac.in`
†E-Mail: `sajadi7r@isid.ac.in`

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 preciously, 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, 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 $\|\mathbf{v} - \mathbf{u}\|_2 \le r$ where $\|.\|_2$ is a standard Euclidean distance.

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

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

The case when the vertices are assumed to be uniformly distributed in $[0,1]^d$, Appel and Russo [1] showed that with probability one

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

Penrose [2] considered the case when the distribution F has a continuous density f with respect to the Lebesgue measure which remain bounded away from 0 on the support of F. Under certain technical assumption such as smooth boundary for the support he showed that with probability one,

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

where C is an explicit constant which depends on the dimension d and essential infimum of f and its value on the boundary of the support. Recently, Sarkar and Saurabh [3] [personal communication], studied a case when the density f of underlying the distribution may have minimum zero. They in particular, proved that when the support of f is $[0,1]$ and f is bounded below on any compact subset not containing zero but it is regularly varying at the origin, then $R_n/F^{-1}(1/n)$ has a weak limit.

In this paper we study the connectivity of random geometric graphs where the underlying distribution of the vertices has no mass but is singular with respect to the Lebesgue measure, that is, it has no density. For that, we consider the *generalized Cantor distribution* with parameter ϕ denoted by $Cantor(\phi)$ as the underlying distribution of veritices of the graph.

1.2 Preliminaries

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

1.2.1 Cantor Set

The Cantor set was first discovered by Smith [4] but became popular after Cantor [5]. 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 $C_n := \overset{2^n}{\underset{k=1}{\cup}} [a_{n,k}, b_{n,k}]$. 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 [6]).

For constructing the generalized Cantor set, we start with unit interval $[0,1]$ and at 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 [6]). Note that the standard Cantor set is a special case when $\phi = 1/3$.

1.2.2 Cantor distribution

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

$$X = \sum_{i=1}^{\infty} \phi^{i-1} Z_i \tag{2}$$

where 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 (2) then we will say that X has a Cantor distribution with parameter ϕ, and write $X \sim \text{Cantor}(\phi)$. Observe that Cantor (ϕ) is self-similar, in the sense that,

$$X \overset{d}{=} \begin{cases} \phi X & \text{with probability } 1/2 \\ \phi X + 1 - \phi & \text{with probability } 1/2 \end{cases} \tag{3}$$

This follows easily by conditioning on Z_1.

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

2 Main results

Let $X_1, X_2, \ldots, X_n$ be independent and identical distribution from Cantor(ϕ) on $[0,1]$. Given 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*

$$R_n \longrightarrow 1 - 2\phi \ a.s. \tag{4}$$

Our next theorem gives finer asymptotics but before we state the theorem we provide here some basic notations and facts. Let $X_{1:n} := \min\{X_1, X_2, \dots, X_n\}$. Using (3) we get

$$X_{1:n} \stackrel{d}{=} \begin{cases} \phi X_{1:k} & \text{with probability} \quad 2^{-n}\binom{n}{k} \text{ for } k = 1, 2, ..., n \\ \phi X_{1:n} + 1 - \phi & \text{with probability} \quad 2^{-n} \end{cases} \tag{5}$$

Let $a_n := \mathbb{E}[X_{1:n}]$. Using (5) Hosking [7] derived the following recursion formulae for the sequence (a_n)

$$(2^n - 2\phi)\, a_n = 1 - \phi + \phi \sum_{k=1}^{n-1} \binom{n}{k} a_k, \quad n \geq 1 \tag{6}$$

Moreover Knopfmacher and Prodinger [8] showed that whenever $0 < \phi < 1/2$ then as $n \to \infty$,

$$\frac{a_n}{n^{-\frac{1}{d_\phi}}} \longrightarrow C(\phi)\,, \tag{7}$$

where

$$C(\phi) := \frac{(1-\phi)(1-2\phi)}{\phi \log 2} \Gamma(-\log_2 \phi) \zeta(-\log_2 \phi)\,, \tag{8}$$

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*

$$\frac{\|R_n - (1-2\phi)\|_1}{n^{-\frac{1}{d_\phi}}} \longrightarrow 2C(\phi)\,, \tag{9}$$

where $C(\phi)$ is as in equation (8) and $\|\cdot\|_1$ is the $\mathcal{L}_1$ norm.

3 Proof of the theorems

3.1 Proof of Theorem 1

We draw a sample of size n from $Cantor(\phi)$ on $[0, 1]$. Therefore this sample will be divided with N_n elements falling in the subinterval $[0, \phi]$ and $n - k$ in $[1 - \phi, 1]$. From the construction $N_n \sim \text{Bin}\left(n, \frac{1}{2}\right)$. In selecting this sample of size n, there are three cases which may happen. Some of these points may fall in interval $[0, \phi]$ and rest in interval $[1 - \phi, 1]$. That means

$N_n \notin \{0, n\}$. In this case the distance between the points in $[0, \phi]$ and $[1 - \phi, 1]$ is at least ϕ. The other cases are when all points fall in $[0, \phi]$ or all fall in $[1 - \phi, 1]$, which in this case $N_n = n$ or $N_n = 0$. Let $m_n = \min_{1 \le i \le n} X_i$, $M_n = \max_{1 \le i \le n} X_i$ and we define

$$L_n := \max \{X_i | \ 1 \le i \le n \text{ and } X_i \in [0, \phi]\} \tag{10}$$

and

$$U_n := \min \{X_i | \ 1 \le i \le n \text{ and } X_i \in [1 - \phi, 1]\} \ . \tag{11}$$

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

$$R_n = (U_n - L_n) \mathbf{1} (N_n \notin \{0, n\}) + R_n^* \mathbf{1}(N_n \in \{0, n\}) \tag{12}$$

where $R_n^* \stackrel{d}{=} \phi R_n$.

Observe that condition on $N_n = k$ where $1 \le k \le n - 1$, we have $U_n \stackrel{d}{=} 1 - \phi + \phi m_{n-k}$ and $L_n \stackrel{d}{=} \phi M_k$. More generally

$$((L_n, U_n), N_n)_{n \ge 1} \stackrel{d}{=} ((\phi M_{N_n}, 1 - \phi + \phi m_{n-N_n}), N_n)_{n \ge 1} \ . \tag{13}$$

Note that to be technically correct 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 SLLN, $N_n/n \longrightarrow 1/2$ a.s., thus $\mathbf{1}(N_n \in \{0, n\}) \longrightarrow 0$ a.s., moreover both (N_n) and $(n - N_n)$ are two subsequences which are converging to infinity a.s. Finally observing that $0 \le R_n^* \le \phi$ a.s. we get from equation (12) and (13)

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

□

3.2 Proof of Theorem 2

Recall that $a_n = \mathbb{E}[m_n]$. From (12) we have

$$\begin{aligned} \mathbb{E}[R_n] &= (1 - 2\phi)(1 - \frac{1}{2^{n-1}}) + \frac{\phi}{2^n} \sum_{k=1}^{n-1} \binom{n}{k} (a_{n-k} + a_k) + \mathbb{E}[R_n^* \mathbf{1}(N_n \in \{0, n\})] \\ &= (1 - 2\phi)(1 - \frac{1}{2^{n-1}}) + \frac{2\phi}{2^n} (\frac{2^n - 2\phi}{\phi} a_n - \frac{1 - \phi}{\phi}) + \mathbb{E}[R_n^* \mathbf{1}(N_n \in \{0, n\})] \, . \end{aligned}$$

The last equality follows from (6). Therefore

$$\frac{\mathbb{E}[R_n - (1 - 2\phi)]}{a_n} = \frac{-1}{a_n 2^{n-1}} + \frac{2^n - \phi}{2^{n-1}} - \frac{1 - \phi}{a_n 2^{n-1}} + \frac{\mathbb{E}[R_n^* \mathbf{1}(N_n \in \{0, n\})]}{a_n} \, . \tag{14}$$

Now observe that

$$0 \le \frac{\mathbb{E}[R_n^* \mathbf{1}(N_n \in \{0, n\})]}{a_n} \le \frac{\phi}{a_n 2^{n-1}} \, ,$$

and by equation (7) we get that $a_n 2^{n-1} \longrightarrow \infty$ as $n \to \infty$, so using (14) we conclude that

$$\frac{\mathbb{E}[R_n - (1-2\phi)]}{a_n} \longrightarrow 2 \quad \text{as} \quad n \longrightarrow \infty$$

Finally note that by definition of R_n, almost surely $R_n \geq 1 - 2\phi$. Using (7) this completes the proof. □

References

[1] Martin J. B. Appel and Ralph P. Russo. The connectivity of a graph on uniform points on $[0,1]^d$. *Statist. Probab. Lett.*, 60(4):351–357, 2002.

[2] Mathew D. Penrose. A strong law for the largest nearest-neighbour link between random points. *J. London Math. Soc. (2)*, 60(3):951–960, 1999.

[3] A. Sarkar and B. Saurabh. Connectivity distance in random geometric graph. *Unpublished results*, 2010.

[4] Henry J.S. Smith. On the integration of discontinuous functions. *Proc. Lond. Math. Soc.*, 6:140–153, 1875.

[5] G. Cantor. Über unendliche, lineare punktmannigfaltigkeiten v, [on infinite, linear point-manifolds (sets)]. *Math. Ann.*, 21:545–591, 1883.

[6] Elias M. Stein and Rami Shakarchi. *Real analysis.* Princeton Lectures in Analysis, III. Princeton University Press, Princeton, NJ, 2005. Measure theory, integration, and Hilbert spaces.

[7] J. R. M. Hosking. Moments of order statistics of the Cantor distribution. *Statist. Probab. Lett.*, 19(2):161–165, 1994.

[8] Arnold Knopfmacher and Helmut Prodinger. Explicit and asymptotic formulae for the expected values of the order statistics of the Cantor distribution. *Statist. Probab. Lett.*, 27(2):189–194, 1996.