isid/ms/2003/13
May 02, 2003
http://www.isid.ac.in/~statmath/eprints

On A Random Directed Spanning Tree

Abhay G. Bhatt
Rahul Roy

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

On A Random Directed Spanning Tree

Abhay G. Bhatt and Rahul Roy
Indian Statistical Institute,
7 SJS Sansanwal Marg, New Delhi 110016, India.

Abstract We study the asymptotic properties of a minimal spanning tree formed by n points uniformly distributed in the unit square, where the minimality is amongst all rooted spanning trees with a direction of growth. We show that the number of branches from the root of this tree, the total length of these branches, and the length of the longest branch each converge weakly. This model is related to the study of record values in the theory of extreme value statistics and this relation is used to obtain our results. The results also hold when the tree is formed from a Poisson point process of intensity n in the unit square.

Keywords minimal spanning tree, record values, weak convergence.
AMS 1991 Classification Primary 60D05, 60G70, Secondary 05C05, 90C27.

On A Random Directed Spanning Tree

Abhay G. Bhatt and Rahul Roy
Indian Statistical Institute, New Delhi.

1 Introduction

In this paper we introduce a notion of a minimal directed spanning tree. To illustrate this consider the following model of transmission of radio waves by transmitters, receivers and amplifiers. Suppose a source transmitter located at the origin transmits messages which can be received only by receivers placed in the positive quadrant. These receivers in turn amplify the message and transmit it to the receivers lying in the positive quadrant with respect to these amplifiers. In this way each receiver lying in the positive quadrant with respect to the transmitter receives the message. The graph which represents this model of transmission may be viewed as a directed spanning tree. For transmission of radio waves in this model, the required strength of the source transmitter is clearly related to the number of receivers which receive the message directly from the transmitter at the origin, as well as the sum of the distances of these receivers from the origin and the distance of the receiver farthest away from the origin which receives the message directly from the origin. In this paper we investigate these three factors when the receivers are located at random in the unit square (say) and we study the asymptotics as the number of transmitters go to infinity.

This model of transmission of radio waves is in contrast to that introduced by Gilbert [1961] where the radio waves can travel in any direction from the point of their origin. However they may be received only by receivers located within a certain fixed distance from the source of transmission.

In the context of transmission of information through wireless networks, there has been quite a lot of work done in recent years, see *e.g.*, Gupta and Kumar [1998]. The object of interest in these studies is the *throughput*, which is the rate of transmission of information. This throughput is related to the strength of the transmitter and its subsequent reception by the receivers. It is in this perspective that the questions we discuss in this paper are important. In particular, the total length of the transmitting network could determine the throughput.

The model we propose could also be viewed as the catchment area of a river. In particular, various mountain gorges drain into a river. The amount of water collected in the river depends on the lengths of the gorges. This should be viewed in relation to the lattice model of such a catchment area. (See *e.g.* Rodriguez-Iturbe and Rinaldo [1997].)

Before we define the random graphs which is the subject of this paper, we need to introduce some notation. Given two points (a_1, b_1) and (a_2, b_2) in $\mathbb{R}^2$, we write $(a_1, b_1) \preceq (a_2, b_2)$ if $a_1 \leq a_2$ and $b_1 \leq b_2$.

Given a vertex set A of $k+1$ vertices $(a_0, b_0), (a_1, b_1), \ldots, (a_k, b_k)$ in $[0,1] \times [0,1]$ satisfying $(a_0, b_0) \preceq (a_i, b_i)$ for every $1 \leq i \leq k$, let E be the set of all directed edges $e_{ij} := <$

$(a_i, b_i), (a_j, b_j) >$ between vertices (a_i, b_i) and (a_j, b_j) satisfying $(a_i, b_i) \preceq (a_j, b_j)$ for all $0 \leq i \neq j \leq k$. Let $\mathcal{G}$ be the collection of all possible graphs G with vertex set $G_V = A$ and edge set G_E a subset of E, such that given any vertex (a_j, b_j), there exist vertices (a_{i_l}, b_{i_l}), $0 \leq l \leq m$ for some $m \geq 1$, such that

1. $(a_{i_0}, b_{i_0}) = (a_0, b_0), \quad (a_{i_m}, b_{i_m}) = (a_j, b_j)$,
2. $< (a_{i_l}, b_{i_l}), (a_{i_{l+1}}, b_{i_{l+1}}) > \in G_E$ for all $0 \leq l \leq m-1$.

Let T denote a graph in $\mathcal{G}$ such that $\sum_{e \in T_E} |e| = \min_{G \in \mathcal{G}} \sum_{e \in G_E} |e|$, where $|e|$ denotes the Euclidean length of the edge e. Clearly T need not be unique and T must necessarily be a tree. This is the *directed minimal spanning tree* we consider in this article.

To construct such a random tree, let $\xi_1, \xi_2, \ldots$ and $\zeta_1, \zeta_2, \ldots$ be i.i.d. random variables, each being uniformly distributed on the interval $[0,1]$. Let $V_n = \{(\xi_1, \zeta_1), \ldots, (\xi_n, \zeta_n)\}$ be the vertex set of n points on the unit square $[0,1] \times [0,1]$. Let T_n be the minimal directed spanning tree obtained with vertex set $V_n \cup (0,0)$. Clearly, T_n is almost surely unique, in the sense that the set of all realisations under each of which there are two or more distinct minimal directed spanning trees has probability 0.

In contrast to the Euclidean minimal spanning tree T'_n on $V_n \cup (0,0)$ the minimal directed spanning tree that we consider has quite a few distinctive properties. In particular for the Euclidean minimal spanning tree T'_n there is a vast literature describing its properties, results on the total length of the tree, degree of a fixed vertex etc. (see e.g. Beardwood, Halton and Hammersley [1959], Steele [1988], Aldous and Steele [1992], Ketsen and Lee [1996]). A property of the Euclidean minimal spanning tree which is quite central to its study (in the case when the weight function is monotone) is that the degree of any vertex is bounded by a constant which depends only on the dimension of the underlying space.

In the minimal directed spanning tree we do not have the above property. Let L_n denote the subgraph of T_n with vertex set $\{(0,0)\} \cup \{(a_i, b_i) : \langle (0,0), (a_i, b_i) \rangle \in T_n \text{ for } 1 \leq i \leq n\}$ and edge set $\{\langle (0,0), (a_i, b_i) \rangle : \langle (0,0), (a_i, b_i) \rangle \in T_n \text{ for } 1 \leq i \leq n\}$. The degree $\delta(n)$ of the vertex $(0,0)$ is then equal to the number of edges of L_n and we have

Theorem 1.1 *As $n \to \infty$, we have*
(i) $\frac{\delta(n)}{\log n}$ *converges almost surely to* 1,
(ii) $(\log n)^{-1/2}(\delta(n) - \log n)$ *converges in distribution to a standard normal random variable,*
(iii) $\limsup_{n\to\infty} \dfrac{\delta(n) - \log n}{\sqrt{(2\log n)(\log\log\log n)}} = 1$ *almost surely, and*
(iv) $\liminf_{n\to\infty} \dfrac{\delta(n) - \log n}{\sqrt{(2\log n)(\log\log\log n)}} = -1$ *almost surely.*

As in the study of random minimal spanning trees, it would be interesting to obtain asymptotic behaviour of the total length and other properties of T_n. In this paper, we do not have such an ambitious goal, rather, we study L_n, the subgraph of T_n which consists of the edges adjacent to the root.

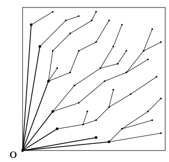

Figure 1: A minimal directed spanning tree. The subgraph formed by the bold edges is L_n.

In the next section we will exhibit the connection between our model of the minimal directed spanning tree and the theory of record values in extreme value statistics. Theorem 1.1 will then be shown to be a restatement of a theorem of Rényi [1976].

Finally we study the sum $l(n)$ of the lengths of the edges and the length $h(n)$ of the longest edge of the subgraph L_n. We will show that

Theorem 1.2 *As $n \to \infty$, $l(n)$ converges weakly to a random variable whose mean and variance are 2 and 1 respectively.*

Theorem 1.3 *As $n \to \infty$, $h(n)$ converges weakly to* $\max\{U_1, U_2\}$, *where U_1 and U_2 are i.i.d. uniform random variables on* $[0, 1]$.

The proofs of the above theorems crucially depend on a 'reflection principle' which we discuss in Section 3. This is the observation that the distribution of the vertices in the subgraph L_n is probabilistically invariant under reflection along the line $x = y$. The proofs are given in Section 4. In the last section, Section 5, we identify the moments of the random variable which is the limit of $l(n)$.

Finally all our results above remain true if the minimal directed spanning tree T_n was constructed from a vertex set V_n' consisting of points of a Poisson point process of intensity n in the unit square $[0, 1] \times [0, 1]$, instead of the vertex set V_n.

2 Records

Consider the vertex set V_n described earlier and assume that
(i) no points of the vertex set V_n lie on the boundary of $[0,1] \times [0,1]$,
(ii) $\xi_i \neq \xi_j$ for all $1 \leq i \neq j \leq n$,
(iii) $\zeta_i \neq \zeta_j$ for all $1 \leq i \neq j \leq n$;
– an event which occurs with probability 1.

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be a permutation of $(\xi_1, \zeta_1), \ldots, (\xi_n, \zeta_n)$, the vertices of V_n such that $0 < X_1 < X_2 < \cdots < X_n < 1$. Thus $(X_1, \ldots, X_n)$ is the order statistic obtained from $(\xi_1, \ldots \xi_n)$. Hence $\big((X_1, Y_1), \ldots, (X_n, Y_n)\big)$ has the same distribution as $\big((O_1, U_1), \ldots, (O_n, U_n)\big)$, where $O_1 < O_2 < \cdots < O_n$ is the order statistic generated from a sample of n i.i.d. uniform random variables on $[0,1]$ and $U_1, \ldots, U_n$ are i.i.d. random variables each being uniformly distributed on $[0,1]$ and independent of all random variables considered so far.

Let R_i denote the i th lower record time of $Y_1, \ldots, Y_n$. In other words, R_i's are random variables defined as follows:

$$R_1 = 1,$$

and, for $i > 1$,

$$R_i = \begin{cases} \infty & \text{if } Y_j > Y_{R_{i-1}} \text{ for all } j > R_{i-1} \\ & \quad \text{or if } R_{i-1} \geq n \\ \min\{j > R_{i-1} : Y_j < Y_{R_{i-1}}\} & \text{otherwise.} \end{cases}$$

Let $k = k(n)$ (random) be such that $R_{k+1} = \infty$ and $1 = R_1 < R_2 < \cdots < R_k \leq n$.

Lemma 2.1 $\{(0,0)\} \cup \{(X_{R_1}, Y_{R_1}), \ldots, (X_{R_k}, Y_{R_k})\}$ *is the vertex set of the subgraph* L_n.

Proof : For $j < R_i$ we have $Y_j > Y_{R_i}$, while for $j > R_i$ we have $X_j > X_{R_i}$; thus there does not exist any (X_j, Y_j), $j \neq R_i$, with $(X_j, Y_j) \preceq (X_{R_i}, Y_{R_i})$. Hence (X_{R_i}, Y_{R_i}) belongs to the vertex set of L_n.

Moreover, for $R_i < j < R_{i+1}$, we have $X_j > X_{R_i}$ and $Y_j > Y_{R_i}$; thus $(X_{R_i}, Y_{R_i}) \preceq (X_j, Y_j)$. Hence (X_j, Y_j) does not belong to the vertex set of L_n. □

From Lemma 2.1, we see that the degree of $(0,0)$ in L_n is exactly $k(n)$. Theorem 1.1 follows immediately from the following theorem by Rényi [1976].

Theorem 2.1 *For an i.i.d. sequence* $W_1, W_2, \ldots$ *of random variables uniformly distributed on* $[0,1]$*, the number of records* $k(n)$ *in* $W_1, \ldots, W_n$ *satisfies*
(i) $P\Big(\lim_{n\to\infty} \dfrac{k(n)}{\log n} = 1\Big) = 1,$
(ii) $\lim_{n\to\infty} P\big((\log n)^{-1/2}(k(n) - \log n) \leq x\big) = \Phi(x)$, *where* Φ *is the standard normal distribution function,*
(iii) $P\Big(\limsup_{n\to\infty} \dfrac{k(n) - \log n}{\sqrt{(2\log n)(\log\log\log n)}} = 1\Big) = 1$, *and*
(iv) $P\Big(\liminf_{n\to\infty} \dfrac{k(n) - \log n}{\sqrt{(2\log n)(\log\log\log n)}} = -1\Big) = 1.$

Let $\eta_1, \ldots, \eta_n$ be defined as

$$\eta_j = \begin{cases} 1 & \text{if } R_i = j \text{ for some } 1 \le i \le k(n) \\ 0 & \text{otherwise.} \end{cases}$$

Clearly,

$$\eta_1 + \cdots + \eta_n = k(n). \tag{2.1}$$

Lemma 2.2 *$\eta_1, \ldots, \eta_n$ are independent random variables with $P(\eta_j = 1) = 1 - P(\eta_j = 0) = j^{-1}$ for $1 \le j \le n$.*

Proof : Note that for $1 \le j \le n$,

$$\begin{aligned} P(\eta_j = 1) &= P(Y_j < \min\{Y_1, Y_2, \ldots, Y_{j-1}\} < 1, 0 < Y_j < 1) \\ &= \int_0^1 \Big(\prod_{i=1}^{j-1} \int_{y_j}^1 dy_i\Big) dy_j \\ &= \int_0^1 (1 - y_j)^{j-1} dy_j = 1/j. \end{aligned}$$

Since η_i's are Bernoulli random variables, to check independence it suffices to show that for $1 \le i_1 < i_2 < \ldots < i_p \le n, 2 \le p \le n$,

$$P(\eta_{i_1} = \eta_{i_2} = \ldots = \eta_{i_p} = 1) = \prod_{j=1}^{p} P(\eta_{i_j} = 1).$$

This equality is easily checked by noting that

$$\begin{aligned} &\big\{\eta_{i_1} = \eta_{i_2} = \ldots = \eta_{i_p} = 1\big\} = \\ &\quad \big\{Y_{i_1} < \min\{Y_1, Y_2, \ldots, Y_{i_1-1}\} < 1; Y_{i_2} < \min\{Y_{i_1}, Y_{i_1+1}, \ldots Y_{i_2-1}\} < 1; \\ &\qquad \ldots; Y_{i_p} < \min\{Y_{i_{p-1}}, Y_{i_{p-1}+1}, \ldots, Y_{i_p-1}\} < 1\big\}. \end{aligned}$$

□

Lemma 2.3 $\lim_{n\to\infty} \frac{1}{\sqrt{n}} E[k(n)]^l = 0$ *for all* $l \ge 1$.

Proof : Using (2.1) and Lemma 2.2 we get, for any real number t

$$\begin{aligned} E[e^{tk(n)}] &= Ee^{t(\eta_1+\cdots+\eta_n)} \\ &= \prod_{i=1}^{n} Ee^{t\eta_i} \\ &= \prod_{i=1}^{n}(\frac{e^t}{i}+\frac{i-1}{i}) \\ &= \frac{1}{2}e^t(1+e^t)\prod_{i=3}^{n}\frac{1}{i}((i-1)+e^t) \\ &= \frac{1}{n}e^t(1+e^t)\prod_{i=2}^{n-1}(1+\frac{e^t}{i}). \end{aligned}$$

Now $E[k(n)]^l$ is the coefficient of $t^l/l!$ in the power series expansion of $E[e^{tk(n)}]$. Moreover

$$0 \leq \frac{E[k(n)]^l t^l}{l!} \leq E[e^{tk(n)}] \qquad \forall t \geq 0.$$

Thus to prove the lemma it suffices to show that

$$\frac{1}{\sqrt{n}}\Big(\frac{1}{n}\prod_{i=2}^{n-1}(1+\frac{e^t}{i})\Big) \to 0 \text{ for some } t \text{ as } n \to \infty. \tag{2.2}$$

Fix $0 < t < \log(3/2)$. Note that

$$\begin{aligned} \log \frac{1}{n\sqrt{n}}\prod_{i=2}^{n-1}\left(1+\frac{e^t}{i}\right) &= \sum_{i=2}^{n-1}\log\left(1+\frac{e^t}{i}\right) - \frac{3}{2}\log n \\ &\leq \sum_{i=2}^{n-1}\frac{e^t}{i} - \frac{3}{2}\log n \\ &\to -\infty \text{ as } n \to \infty. \end{aligned}$$

This proves (2.2) and hence completes the proof of the lemma. □

A question which does not arise naturally in the study of record values and as such has not been considered in its study is the computation of the moments of $\sum_{i=1}^{k(n)} R_i$, the sum of record times. Since we need it in our study we present the following proposition.

Proposition 2.1 *For every $l \geq 1$, we have*

$$E\left(\sum_{i=1}^{k(n)} R_i\right)^l \leq n^l l^l.$$

Proof : First observe that $\sum_{i=1}^{k(n)} R_i = \sum_{i=1}^{n} i\eta_i$. Thus, using the notation $\sum_\alpha$ to denote the sum over $\alpha_1, \ldots, \alpha_j$ such that $1 \leq \alpha_1, \ldots, \alpha_j \leq l$ and $\alpha_1 + \cdots + \alpha_j = l$, we get from Lemma 2.2

$$E\left(\sum_{i=1}^{k(n)} R_i\right)^l = E\left(\sum_{i=1}^{n} i\eta_i\right)^l$$

$$
\begin{aligned}
&= \sum_{j=1}^{l} \sum_{1\le i_1<\cdots<i_j\le n} \sum_{\alpha} \frac{l!}{\alpha_1!\ldots\alpha_j!} i_1^{\alpha_1}\ldots i_j^{\alpha_j} \frac{1}{i_1\ldots i_j} \\
&\le \sum_{j=1}^{l} \sum_{\alpha} \frac{l!}{\alpha_1!\ldots\alpha_j!} \sum_{i_1=1}^{n}\cdots\sum_{i_j=1}^{n} i_1^{\alpha_1-1}\ldots i_j^{\alpha_j-1} \\
&\le \sum_{j=1}^{l} \sum_{\alpha} \frac{l!}{\alpha_1!\ldots\alpha_j!} n^{\alpha_1+\cdots+\alpha_j} \\
&= n^l \sum_{j=1}^{l} \sum_{\alpha} \frac{l!}{\alpha_1!\cdots\alpha_j!} \\
&= n^l l^l.
\end{aligned}
$$

□

In particular we have,

$$
E\left(\sum_{i=1}^{k(n)} R_i\right) = \sum_{j=1}^{n} E(j\eta_j) = n, \tag{2.3}
$$

and

$$
\begin{aligned}
E\left(\sum_{i=1}^{k(n)} R_i\right)^2 &= \sum_{i=1}^{n} i + \sum_{1\le i_1<i_2\le n} \sum_{\substack{1\le\alpha_1,\alpha_2\le 2\\ \alpha_1+\alpha_2=2}} \frac{2!}{\alpha_1!\alpha_2!} i_1^{\alpha_1-1} i_2^{\alpha_2-1} \\
&= \sum_{i=1}^{n} i + \sum_{i_1=1}^{n-1} \sum_{i_2=i_1+1}^{n} 2 \\
&= \frac{n(3n-1)}{2}.
\end{aligned} \tag{2.4}
$$

3 A Reflection Principle

We now revisit the vertex set $V_n = \{(\xi_1,\zeta_1),\ldots,(\xi_n,\zeta_n)\}$ introduced in Section 1. In particular $(\xi_1,\xi_2,\ldots\xi_n)$ and $(\zeta_1,\zeta_2,\ldots\zeta_n)$ are two independent sets of i.i.d. random variables, each being uniformly distributed on the interval $[0,1]$. Earlier we considered a permutation $\{(X_1,Y_1),\ldots,(X_n,Y_n)\}$ of the set V_n which together with the lower record times $\{R_i : 1 \le i \le n\}$ of the random variables $Y_1,\ldots,Y_n$ provided a representation of the subgraph L_n (see Lemma 2.1).

Alternately, we may consider the order statistic of $\zeta_1,\zeta_2,\ldots,\zeta_n$ and then look at the record values in the other co-ordinates. Let $Y_1',\ldots,Y_n'$, with $Y_1' < \cdots < Y_n'$, be the order statistic of $\zeta_1,\ldots,\zeta_n$ and let $X_1',\ldots,X_n'$ be the corresponding values of ξ's. *i.e.* $\{(X_1',Y_1'),\ldots,(X_n',Y_n')\}$ is a permutation of V_n. Let $R_1',\ldots,R_{m(n)}'$ be the lower record times of $X_1',\ldots,X_n'$. Arguing exactly as in Lemma 2.1 we get that the vertex set of the subgraph L_n is $\{(0,0)\} \cup$

$\{(X'_{R'_1}, Y'_{R'_1}), \ldots, (X'_{R'_{m(n)}}, Y'_{R'_{m(n)}})\}$. We note here that

$$Y'_{R'_1} < \cdots < Y'_{R'_{m(n)}} \text{ and } X'_{R'_1} > \cdots > X'_{R'_{m(n)}} \text{ a.s.} \tag{3.1}$$

Since either of the two constructions lead to the same set, viz., the vertex set of L_n, we have

$$\{(X_{R_1}, Y_{R_1}), \ldots, (X_{R_{k(n)}}, Y_{R_{k(n)}})\} = \{(X'_{R'_1}, Y'_{R'_1}), \ldots, (X'_{R'_{m(n)}}, Y'_{R'_{m(n)}})\}, \tag{3.2}$$

and so $k(n) = m(n)$. However, contrary to (3.1) we have

$$X_{R_1} < \cdots < X_{R_{k(n)}} \text{ and } Y_{R_1} > \cdots > Y_{R_{k(n)}} \text{ a.s.} \tag{3.3}$$

Thus we have

$$X_{R_l} = X'_{R'_{k(n)+1-l}} \text{ and } Y_{R_l} = Y'_{R'_{k(n)+1-l}} \text{ for every } l = 1, \ldots, k(n). \tag{3.4}$$

Now note that the random vectors $(Y_1, \ldots, Y_n)$ and $(X'_1, \ldots, X'_n)$ are identically distributed, each being a vector of n i.i.d. uniform $[0, 1]$ random variables. Hence the record times $(R_1, \ldots, R_{k(n)})$ and $(R'_1, \ldots, R'_{k(n)})$ are also identically distributed. Also the random vectors $(X_1, \ldots, X_n)$ and $(Y'_1, \ldots, Y'_n)$ are identically distributed, each being the order statistic obtained from a sample of n i.i.d. uniform $[0, 1]$ random variables. Further $(X_1, \ldots, X_n)$ and $(R_1, \ldots, R_{k(n)})$ are independent since $(R_1, \ldots, R_{k(n)})$ is obtained from $(Y_1, \ldots, Y_n)$ which is independent of $(X_1, \ldots, X_n)$. Similarly $(Y'_n, \ldots, Y'_1)$ and $(R'_{k(n)}, \ldots, R'_1)$ are independent. Thus the random vectors $(X_{R_1}, \ldots, X_{R_{k(n)}})$ and $(Y'_{R'_{k(n)}}, \ldots, Y'_{R'_1})$ are identically distributed.

In combination with our observation (3.4) we have

$$(X_{R_1}, \ldots, X_{R_{k(n)}}) \text{ and } (Y_{R_{k(n)}}, \ldots, Y_{R_1}) \text{ are identically distributed.} \tag{3.5}$$

4 Asymptotic length of L_n

First we state a result which we will be using quite often in this section.

Theorem 4.1 *Let $Z_1, Z_2, \ldots$ be random variables on a probability space such that $E|Z_n^r| < \infty$ for every $n \geq 1$ and every $r \geq 1$. Let $m_r(n) = EZ_n^r$ and suppose that $\lim_{n\to\infty} m_r(n) = m_r$ for every $r \geq 1$. If $\{m_r, r \geq 1\}$ is such that m_r is finite for every $r \geq 1$ and $\sum_{r=1}^{\infty} m_{2r}^{-\frac{1}{2r}} = \infty$, then there exists a random variable Z such that Z_n converges weakly to Z as $n \to \infty$ and $E(Z^r) = m^r$ for every $r \geq 1$.*

The proof of this theorem follows from combining Theorem 4.5.5 of Chung (1974) and Theorem 1.10 of Shohat and Tamarkin (1960).

To show the weak convergence of the sum $\sum_{i=1}^{k(n)} \sqrt{X_{R_i}^2 + Y_{R_i}^2}$ of the lengths of the edges of L_n as $n \to \infty$, we first show the weak convergence of the sum $\sum_{i=1}^{k(n)} X_{R_i}$ as $n \to \infty$. Towards this end we will show

$$E\left(\sum_{i=1}^{k(n)} X_{R_i}\right)^l \to \mu_l \text{ (say) for every } l \text{ as } n \to \infty, \tag{4.1}$$

and

$$\sum_{l=1}^{\infty} \mu_{2l}^{-1/2l} = \infty, \tag{4.2}$$

which, from Theorem 4.1 guarantees the weak convergence of $\sum_{i=1}^{k(n)} X_{R_i}$ as $n \to \infty$.

The product moments of the order statistics $0 \le O_1 \le O_2 \le \cdots \le O_n \le 1$ obtained from a sample of i.i.d. uniform random variables are as follows:
Let $1 \le i_1 < i_2 < \ldots < i_m \le n$, and let $\alpha_1, \alpha_2, \ldots, \alpha_m$ be non-negative integers with $\sum_{i=1}^m \alpha_i = \alpha$. Then (see David [1970], page 28)

$$\begin{aligned} E\Big[O_{i_1}^{\alpha_1} O_{i_2}^{\alpha_2} \ldots O_{i_m}^{\alpha_m}\Big] &= \frac{n!}{(n + \sum_{j=1}^m \alpha_j)!} \prod_{j=1}^m \frac{(i_j - 1 + \sum_{k=1}^j \alpha_k)!}{(i_j - 1 + \sum_{k=1}^{j-1} \alpha_k)!} \\ &= \frac{\prod_{p=1}^m \prod_{q=0}^{\alpha_p - 1} \big(i_p + \alpha_1 + \cdots + \alpha_{p-1} + q\big)}{\prod_{j=1}^{\alpha} (n+j)}, \end{aligned} \tag{4.3}$$

where in the last term we have used the convention that

$$\prod_{q=0}^{\alpha_p - 1} \big(i_p + \alpha_1 + \cdots + \alpha_{p-1} + q\big) = 1, \text{ whenever } \alpha_p = 0. \tag{4.4}$$

Proposition 4.1 *For every $l \ge 1$, as $n \to \infty$,*

$$E\Big(\sum_{i=1}^{k(n)} X_{R_i}\Big)^l = \frac{E\Big(\sum_{i=1}^{k(n)} R_i\Big)^l}{\prod_{j=1}^l (n+j)} + o(1). \tag{4.5}$$

Proof : Fix $l \ge 1$. Note that

$$\Big(\sum_{i=1}^{k(n)} X_{R_i}\Big)^l = \sum_{\alpha} \frac{l!}{\alpha_1! \ldots \alpha_{k(n)}!} X_{R_1}^{\alpha_1} \cdots X_{R_{k(n)}}^{\alpha_{k(n)}} \tag{4.6}$$

where $\sum_\alpha$ stands for the summation over all non-negative integers $\alpha_1, \ldots, \alpha_{k(n)}$ such that $\sum_{p=1}^{k(n)} \alpha_p = l$.

We also note that the number of records $k(n)$ and the record times $R_1, \ldots, R_{k(n)}$ depend only on the random variables $Y_1, \ldots, Y_n$ and thus, as noted in Section 2, are independent of the random variables $X_1, \ldots, X_n$. Using this fact, (4.3) and a conditioning argument, we get

$$\begin{aligned} E\Big(\sum_{i=1}^{k(n)} X_{R_i}\Big)^l &= E\Big(\sum_{\alpha} \frac{l!}{\alpha_1! \ldots \alpha_{k(n)}!} E\left(X_{R_1}^{\alpha_1} \cdots X_{R_{k(n)}}^{\alpha_{k(n)}} | Y_1, \ldots, Y_n\right)\Big) \\ &= E\left(\sum_{\alpha} \frac{l!}{\alpha_1! \ldots \alpha_{k(n)}!} \frac{\Big[\prod_{p=1}^{k(n)} \prod_{q=0}^{\alpha_p - 1} \big(R_p + \alpha_1 + \ldots + \alpha_{p-1} + q\big)\Big]}{\prod_{j=1}^l (n+j)}\right), \end{aligned} \tag{4.7}$$

where $\sum_\alpha$ is as defined in (4.6). Following the convention (4.4) we use the notation $\prod_P$ below to mean product over all $p = 1, 2, \ldots, k(n)$ for which $\alpha_p > 0$. Now using the fact that $R_p \le n$

we observe that

$$\begin{aligned}
\prod_{p=1}^{k(n)}\prod_{q=0}^{\alpha_p-1}\big(R_p+\alpha_1+\ldots+\alpha_{p-1}+q\big) &= \prod_{P}\prod_{q=0}^{\alpha_p-1}\big(R_p+\alpha_1+\ldots+\alpha_{p-1}+q\big)\\
&\le \prod_{P}\prod_{q=0}^{\alpha_p-1}\big(R_p+l\big) = \prod_{P}\big(R_p+l\big)^{\alpha_p}\\
&= \prod_{P}\sum_{j=0}^{\alpha_p}\binom{\alpha_p}{j} l^j R_p^{\alpha_p-j}\\
&= \prod_{P}\Big[R_p^{\alpha_p}+\sum_{j=1}^{\alpha_p}\binom{\alpha_p}{j} l^j R_p^{\alpha_p-j}\Big]\\
&\le \prod_{P}\Big[R_p^{\alpha_p}+n^{\alpha_p-1}\sum_{j=1}^{\alpha_p}\binom{\alpha_p}{j} l^j\Big]\\
&\le \prod_{P}\Big[R_p^{\alpha_p}+n^{\alpha_p-1}(l+1)^{\alpha_p}\Big]\\
&= \Big[\prod_{P}R_p^{\alpha_p}\Big]+\Big[\sum_{\substack{p=1\\ \alpha(p)\ge 1}}^{k(n)}(l+1)^{\alpha_p}\,n^{\alpha_p-1}\prod_{\substack{j=1\\ j\neq p}}^{k(n)}\beta_j\Big]
\end{aligned}$$

where for each j, β_j is either $R_j^{\alpha_j}$ or $(l+1)^{\alpha_j}n^{\alpha_j-1}$. Since $R_j \le n$ for every j and $\sum_{\substack{j=1\\ j\neq p}}^{k(n)}\alpha_j = l-\alpha_p$, we have

$$\sum_{\substack{p=1\\ \alpha(p)\ge 1}}^{k(n)}(l+1)^{\alpha_p}n^{\alpha_p-1}\prod_{\substack{j=1\\ j\neq p}}^{k(n)}\beta_j \le \sum_{p=1}^{k(n)}(l+1)^l n^{l-1}.$$

Thus we get

$$\prod_{p=1}^{k(n)}\prod_{q=0}^{\alpha_p-1}\big(R_p+\alpha_1+\ldots+\alpha_{p-1}+q\big) \le R_1^{\alpha_1}\ldots R_{k(n)}^{\alpha_{k(n)}}+k(n)(l+1)^l n^{l-1}. \tag{4.8}$$

It follows from (4.7) and (4.8) that

$$\begin{aligned}
E\Big(\sum_{i=1}^{k(n)}X_{R_i}\Big)^l \le \prod_{j=1}^{l}(n+j)^{-1}E\Big(\sum_{\alpha}\frac{l!}{\alpha_1!\ldots\alpha_{k(n)}!}R_1^{\alpha_1}\ldots R_{k(n)}^{\alpha_{k(n)}}\Big)\\
+\prod_{j=1}^{l}(n+j)^{-1}E\Big(\sum_{\alpha}\frac{l!}{\alpha_1!\ldots\alpha_{k(n)}!}k(n)(l+1)^l n^{l-1}\Big).
\end{aligned} \tag{4.9}$$

Clearly,

$$\sum_{\alpha}\frac{l!}{\alpha_1!\ldots\alpha_{k(n)}!}R_1^{\alpha_1}\ldots R_{k(n)}^{\alpha_{k(n)}} = \Big(\sum_{i=1}^{k(n)}R_i\Big)^l. \tag{4.10}$$

Moreover,

$$\sum_{\alpha}\frac{l!}{\alpha_1!\ldots\alpha_{k(n)}!}k(n)(l+1)^l n^{l-1} = k(n)(l+1)^l n^{l-1}\sum_{\alpha}\frac{l!}{\alpha_1!\ldots\alpha_{k(n)}!}$$

$$
\begin{aligned}
&= k(n)(l+1)^l n^{l-1} k(n)^l. \\
&= (l+1)^l n^{l-1} k(n)^{l+1}. \qquad (4.11)
\end{aligned}
$$

Thus the second term in (4.9) becomes

$$
\begin{aligned}
&\prod_{j=1}^{l}(n+j)^{-1} E\Big(\sum_{\alpha} \frac{l!}{\alpha_1! \dots \alpha_{k(n)}!} k(n)(l+1)^l n^{l-1}\Big) \\
&\qquad = \prod_{j=1}^{l}(n+j)^{-1}(l+1)^l n^{l-1} E\left(k(n)^{l+1}\right) \\
&\qquad \leq \frac{(l+1)^l E\left(k(n)^{l+1}\right)}{n} \to 0 \text{ as } n \to \infty, \qquad (4.12)
\end{aligned}
$$

where the last implication follows from Lemma 2.3. The proposition now follows from (4.9), (4.10) and (4.12). □

We now proceed to prove (4.1). Let $\{(X_1', Y_1'), \dots, (X_{n+1}', Y_{n+1}')\}$ be the random vectors used in the construction of the graph T_{n+1}. Here we assume that $X_1' < X_2' < \cdots < X_{n+1}'$ is a permutation of $n+1$ i.i.d. uniform random variables on $[0,1]$ and $Y_1', Y_2', \dots, Y_{n+1}'$ is an independent sequence of $n+1$ i.i.d. uniform random variables on $[0,1]$. Let U and V be two independent uniform random variables on $[0,1]$. Consider the random variables

$$
\tilde{Y}_i = \begin{cases} Y_i & \text{if } i \leq n \\ V & \text{if } i = n+1. \end{cases} \qquad (4.13)
$$

Let M be the random variable defined by

$$
\{M(\omega) = m\} = \{X_m(\omega) < U(\omega) < X_{m+1}(\omega)\}, \quad m = 0, 1, 2, \dots, n
$$

where $X_0 = 0$ and $X_{n+1} = 1$. Let

$$
\tilde{X}_i = \begin{cases} X_i & \text{if } i \leq M \\ U & \text{if } i = M+1 \\ X_{i-1} & \text{if } i \geq M+2 \end{cases} \qquad (4.14)
$$

Note that $\{(X_1', Y_1'), \dots, (X_{n+1}', Y_{n+1}')\}$ and $\{(\tilde{X}_1, \tilde{Y}_1), \dots, (\tilde{X}_{n+1}, \tilde{Y}_{n+1})\}$ are identically distributed. Thus if $\sigma_1, \sigma_2, \dots, \sigma_{k'(n+1)}$ are the lower record times obtained from $Y_1', Y_2', \dots, Y_{n+1}'$, then

$$
\left(\sum_{i=1}^{k'(n+1)} X_{\sigma_i}', \sum_{i=1}^{k'(n+1)} Y_{\sigma_i}'\right) \stackrel{d}{=} \left(\sum_{i=1}^{\tilde{k}(n+1)} \tilde{X}_{S_i}, \sum_{i=1}^{\tilde{k}'(n+1)} \tilde{Y}_{S_i}\right) \qquad (4.15)
$$

where $\{S_1, \dots, S_{\tilde{k}(n+1)}\}$ are the record times obtained from $\tilde{Y}_1, \dots, \tilde{Y}_{n+1}$ and where $\stackrel{d}{=}$ means same in distribution.

Define a random variable $t(U)$ as follows. Let $t(U) = 0$ if $M = 0$ and for $M > 0$ let $t(U)$ be such that $X_{R_{t(U)}} \leq X_M$ and $X_{R_{t(U)+1}} > X_M$. In the following we use the usual convention that $\sum_1^0 = 0$. From (4.13) and (4.14) we have

$$\begin{aligned}\sum_{i=1}^{\tilde{k}(n+1)} \tilde{X}_{S_i} &= \sum_{i=1}^{t(U)} X_{R_i} + U 1_{\{R_{t(U)+1}=M+1\}} + X_{R_{t(U)+1}-1} 1_{\{R_{t(U)+1}\neq M+1\}} \\ &\quad + \sum_{i=t(U)+2}^{k(n)} X_{R_i-1} + X_n 1_{\{V<\min\{Y_1,\ldots,Y_n\}\}} \qquad (4.16)\end{aligned}$$

and

$$\sum_{i=1}^{\tilde{k}(n+1)} \tilde{Y}_{S_i} = \sum_{i=1}^{k(n)} Y_{R_i} + V 1_{\{V<\min\{Y_1,\ldots,Y_n\}\}}. \qquad (4.17)$$

Theorem 4.2 $\sum_{i=1}^{k(n)} X_{R_i}$ *converges weakly as* $n \to \infty$.

Proof : From Propositions 2.1 and 4.1, it follows that for all n large enough and for every integer $l \geq 1$

$$E[\sum_{i=1}^{k(n)} X_{R_i}]^l \leq l^l. \qquad (4.18)$$

Also, the reflection principle (3.5), (4.15) along with (4.17) yields

$$E[\sum_{i=1}^{k(n)} X_{R_i}]^l = E[\sum_{i=1}^{k(n)} Y_{R_i}]^l \leq E[\sum_{i=1}^{k'(n+1)} Y'_{\sigma_i}]^l = E[\sum_{i=1}^{k'(n+1)} X'_{\sigma_i}]^l$$

for every $l \geq 1$. (4.18) now implies that for every $l \geq 1$

$$\mu_l := \lim_{n\to\infty} E[\sum_{i=1}^{k(n)} X_{R_i}]^l \text{ exists and } \mu_l \leq l^l. \qquad (4.19)$$

Furthermore

$$\sum_{l=1}^{\infty} \frac{1}{(\mu_{2l})^{1/2l}} \geq \sum_{l=1}^{\infty} \frac{1}{2l} = \infty. \qquad (4.20)$$

Theorem 4.1 now completes the proof of the theorem. □

Now for every fixed n, consider

$$d(n) = \sum_{i=1}^{k(n)} (X_{R_i} + Y_{R_i}), \qquad (4.21)$$

which is the sum of the Manhattan distance (L_1 distance) of the vertices of L_n from the origin.

First we have

$$E[(\sum_{i=1}^{k(n+1)} \tilde{X}_{S_i} + \sum_{i=1}^{k(n+1)} \tilde{Y}_{S_i})^l]$$

$$= \sum_{j=0}^{l} \binom{l}{j} E\left[\left(\sum_{i=1}^{k(n+1)} \tilde{X}_{S_i}\right)^{j} \left(\sum_{i=1}^{k(n+1)} \tilde{Y}_{S_i}\right)^{l-j}\right]. \tag{4.22}$$

From (4.16) and (4.17) we get

$$\begin{aligned}
&\Big(\sum_{i=1}^{k(n+1)} \tilde{X}_{S_i}\Big)^j \Big(\sum_{i=1}^{k(n+1)} \tilde{Y}_{S_i}\Big)^{l-j} \\
=&\Big(\sum_{i=1}^{t(U)} X_{R_i} + \sum_{i=t(U)+2}^{k(n)} X_{R_i-1} + U 1_{\{R_{t(U)+1}=M+1\}} + X_{R_{t(U)+1}-1} 1_{\{R_{t(U)+1}\neq M+1\}} \\
&\quad + X_n 1_{\{V<\min\{Y_1,\dots,Y_n\}\}}\Big)^j \Big(\sum_{i=1}^{k(n)} Y_{R_i} + V 1_{\{V<\min\{Y_1,\dots,Y_n\}\}}\Big)^{l-j} \\
\geq&\Big(\sum_{i=1}^{k(n)} X_{R_i} - \sum_{i=t(U)+2}^{k(n)} (X_{R_i} - X_{R_i-1}) - (X_{R_{t(U)+1}} - U) 1_{\{R_{t(U)+1}=M+1\}} \\
&\quad - (X_{R_{t(U)+1}} - X_{R_{t(U)+1}-1}) 1_{\{R_{t(U)+1}\neq M+1\}}\Big)^j \Big(\sum_{i=1}^{k(n)} Y_{R_i}\Big)^{l-j} \\
\geq&\Big(\sum_{i=1}^{k(n)} X_{R_i} - (k(n)+1)(\max\{X_{j+1}-X_j : j=0,\dots,n\})\Big)^j \Big(\sum_{i=1}^{k(n)} Y_{R_i}\Big)^{l-j}.
\end{aligned} \tag{4.23}$$

Let $\Delta_n(X) = \max\{X_{j+1} - X_j : j = 0, \dots, n\}$. Then we have

$$\begin{aligned}
&\Big(\sum_{i=1}^{k(n)} X_{R_i} - (k(n)+1)(\Delta_n(X))\Big)^j \Big(\sum_{i=1}^{k(n)} Y_{R_i}\Big)^{l-j} - \Big(\Big(\sum_{i=1}^{k(n)} X_{R_i}\Big)^j \Big(\sum_{i=1}^{k(n)} Y_{R_i}\Big)^{l-j}\Big) \\
&\quad = \sum_{k=1}^{j} \binom{j}{k} (-1)^k \Big(\sum_{i=1}^{k(n)} X_{R_i}\Big)^{j-k} \big((k(n)+1)(\Delta_n(X))\big)^k \Big(\sum_{i=1}^{k(n)} Y_{R_i}\Big)^{l-j} \\
&\quad \geq - \sum_{k \text{ odd},\, k=1}^{j} \binom{j}{k} \Big(\sum_{i=1}^{k(n)} X_{R_i}\Big)^{j-k} \big((k(n)+1)(\Delta_n(X))\big)^k \Big(\sum_{i=1}^{k(n)} Y_{R_i}\Big)^{l-j} \\
&\quad \geq - \sum_{k \text{ odd},\, k=1}^{j} \binom{j}{k} (k(n)+1)^l (\Delta_n(X))
\end{aligned} \tag{4.24}$$

where the last inequality holds because $X_{R_i} \leq 1$ and $Y_{R_i} \leq 1$ for every i and $0 \leq X_{j+1} - X_j \leq 1$ for every j.

Moreover,

$$\begin{aligned}
&E\Big(\sum_{i=1}^{k(n)} X_{R_i}\Big)^{j-m} \big((k(n)+1)(\Delta_n(X))\big)^m \Big(\sum_{i=1}^{k(n)} Y_{R_i}\Big)^{l-j} \\
&\qquad \leq E(k(n))^{j-m} \big((k(n)+1)(\Delta_n(X))\big)^m (k(n))^{l-j} \\
&\qquad \leq E(k(n)+1)^l E(\Delta_n(X)),
\end{aligned} \tag{4.25}$$

where the last equality follows from the independence of $k(n)$ and the sequence $\{X_1, \ldots, X_n\}$ and the fact that $(\Delta_n(X))^m \leq \Delta_n(X)$. We are now ready to prove the following theorem.

Theorem 4.3 *$d(n)$ converges weakly as $n \to \infty$.*

Proof : From (4.22)–(4.25) we get

$$\begin{aligned} E(d(n+1))^l &\geq \sum_{j=0}^{l} \binom{l}{j} \Big[E[(\sum_{i=1}^{k(n)} X_{R_i})^j (\sum_{i=1}^{k(n)} Y_{R_i})^{l-j})] \\ &\quad - \sum_{k \text{ odd},\, k=1}^{j} \binom{j}{k} E(k(n)+1)^l E(\Delta_n(X)) \Big] \\ &\geq E(d(n))^l - 2^{2l} E(k(n)+1)^l E(\Delta_n(X)). \end{aligned} \tag{4.26}$$

To bound $E(\Delta_n(X))$, observe that for $\max_{j=0,\ldots,n}\{X_{j+1} - X_j\} \geq 4/\sqrt{n}$ to occur at least one of the intervals $\{(j/\sqrt{n}, (j+1)/\sqrt{n}] : j = 0, \ldots, \lfloor \sqrt{n} \rfloor\}$ must not contain any point from $\{X_1, \ldots, X_n\}$ – an event which occurs with probability at most $(\sqrt{n}+1)\big(1 - \frac{1}{\sqrt{n}}\big)^n$. Thus

$$\begin{aligned} E(\Delta_n(X)) &= E(\max_{j=0,\ldots,n}\{X_{j+1} - X_j\}) \\ &\leq \frac{4}{\sqrt{n}} P\{\max_{j=0,\ldots,n}\{X_{j+1} - X_j\} \leq 4/\sqrt{n}\} + (\sqrt{n}+1)\left(1 - \frac{1}{\sqrt{n}}\right)^n \\ &\leq \frac{5}{\sqrt{n}} \text{ for large } n. \end{aligned} \tag{4.27}$$

Combining (4.26), (4.27) and Lemma 2.3 we have, for every $l \geq 1$

$$E(d(n+1))^l \geq E(d(n))^l - c(n,l)$$

where $c(n,l) \to 0$ as $n \to \infty$. Also, from (4.18) and the reflection principle (3.5) we have for n large enough,

$$\|d(n)\|_l := [E(d(n))^l]^{1/l} \leq \| \sum_{i=1}^{k(n)} X_{R_i} \|_l + \| \sum_{i=1}^{k(n)} Y_{R_i} \|_l \leq 2l, \qquad l \geq 1.$$

Thus for every $l \geq 1$, as $n \to \infty$, $E(d(n))^l$ converges to $\hat{\mu}_l$ (say) with $\hat{\mu}_l \leq 2^l \mu_l \leq (2l)^l$. Now $\sum_{l=1}^{\infty} \frac{1}{(\hat{\mu}_{2l})^{1/2l}} \geq \sum_{l=1}^{\infty} \frac{1}{4l} = \infty$ and Theorem 4.1 yields the desired weak convergence. □

We now show that the Manhattan distance $d(n)$ is a good approximation of the Euclidean distance $l(n)$. Let (X_c, Y_c) $(= (X_{c(n)}, Y_{c(n)}))$ be the vertex of L_n closest to the origin with respect to the Euclidean distance, i.e.,

$$l_c = \sqrt{X_c^2 + Y_c^2} = \min_{1 \leq i \leq k(n)} \sqrt{X_{R_i}^2 + Y_{R_i}^2}. \tag{4.28}$$

We first get a bound on the expected value of l_c.

Lemma 4.1 $E[l_c] \leq \frac{Constant}{\sqrt{n}} + \frac{1}{n+1}$.

Proof : Note that $\{l_c > a\}$ is the event that none of the n independent uniformly distributed points lie in the ball with radius a and with the origin at $(0,0)$ (intersected with $[0,1]^2$). Thus

$$P(l_c > a) = \begin{cases} \left(1 - \frac{\pi a^2}{4}\right)^n & \text{for } 0 \leq a \leq 1 \\ \left[1 - \left(\sqrt{a^2-1} + \frac{a^2}{2}(\frac{\pi}{2} - 2\cos^{-1}(\frac{1}{a}))\right)\right]^n & \text{for } 1 \leq a \leq \sqrt{2} \\ 0 & \text{otherwise.} \end{cases}$$

Also note that $\frac{\pi}{2} - 2\cos^{-1}(\frac{1}{a}) \geq 0$ for $1 \leq a \leq \sqrt{2}$. Thus

$$P(l_c > a) \leq \left[1 - \sqrt{a^2-1}\right]^n \qquad \text{for } 1 \leq a \leq \sqrt{2}.$$

This implies that

$$\begin{aligned} E[l_c] &= \int_0^\infty P(l_c > a)da \\ &\leq \int_0^1 \left(1 - \frac{\pi a^2}{4}\right)^n da + \int_1^{\sqrt{2}} \left[1 - \sqrt{a^2-1}\right]^n da \\ &= I_1 + I_2 \text{ (respectively, say).} \end{aligned} \tag{4.29}$$

Substituting $a^2 - 1 = u^2$ in I_2 we get

$$I_2 = \int_0^1 (1-u)^n \frac{u}{\sqrt{u^2+1}} du \leq \int_0^1 (1-u)^n du = \frac{1}{n+1}.$$

Similarly, in I_1 substituting $a = 2\cos\theta/\sqrt{\pi}$ we get,

$$\begin{aligned} I_1 &= \frac{2}{\sqrt{\pi}} \int_{\cos^{-1}(\sqrt{\pi/4})}^{\pi/2} (\sin\theta)^{2n+1} d\theta \\ &\leq \frac{2}{\sqrt{\pi}} \int_0^{\pi/2} (\sin\theta)^{2n+1} d\theta \\ &= \frac{2}{\sqrt{\pi}} \frac{2^{2n}(n!)^2}{(2n+1)!}. \end{aligned}$$

For the last equality above see *e.g.* Gradshteyn and Ryzhik (1980), page 369. Using Stirling's approximation (see Feller [1978] pg. 52) we get, for every $n \geq 1$,

$$\begin{aligned} \frac{2^{2n+1}(n!)^2}{\sqrt{\pi}(2n+1)!} &\leq \frac{(2\pi)2^{2n+1}n^{2n+1}e^{-2n}\exp(1/(6n))}{\pi\sqrt{2}(2n+1)^{2n+(3/2)}e^{-2n-1}\exp(1/(24n+13))} \\ &\leq \frac{en^{2n+1}\exp(1/6)}{(n+1/2)^{2n+(3/2)}} \end{aligned}$$

$$\leq \quad \text{Constant}/\sqrt{n}.$$

The lemma now follows. ☐

Theorem 4.4 *$|l(n) - d(n)|$ converges to zero in probability as $n \to \infty$.*

Proof : Note that if (X_r, Y_r) is a vertex of L_n, other than the origin, then $X_r \leq X_c$ if and only if $Y_r \geq Y_c$. The sets $A := \{(X_r, Y_r) \in L_n \backslash \{0, 0\} : X_r \leq X_c\}$ and $B := \{(X_r, Y_r) \in L_n \backslash \{0, 0\} : Y_r < Y_c\}$, are disjoint and the vertex set of L_n may be written as $A \cup B \cup \{(0, 0)\}$. Also, for $(X_r, Y_r) \in A$,

$$Y_r \leq \sqrt{X_r^2 + Y_r^2} \leq X_r + Y_r \leq X_c + Y_r.$$

Summing all records in A, we get

$$\sum_{(X_r,Y_r)\in A} Y_r \leq l^A(n) \leq d^A(n) \leq k^A(n)X_c + \sum_{(X_r,Y_r)\in A} Y_r \tag{4.30}$$

where $l^A(n)$ and $d^A(n)$ are the respective Euclidean and Manhattan distances when the graph is restricted to the set A and $k^A(n) = \max\{r : (X_r, Y_r) \in L_n \backslash \{0, 0\} : X_r \leq X_c\}$. For the set B, we may define $l^B(n), d^B(n)$ similarly and $k^B(n) = k(n) - k^A(n)$. We obtain

$$\sum_{(X_r,Y_r)\in B} X_r \leq l^B(n) \leq d^B(n) \leq \sum_{(X_r,Y_r)\in B} X_r + k^B(n)Y_c. \tag{4.31}$$

Adding (4.30) and (4.31) we get

$$\sum_{(X_r,Y_r)\in A} Y_r + \sum_{(X_r,Y_r)\in B} X_r \leq l(n) \leq d(n) \leq \sum_{(X_r,Y_r)\in A} Y_r + \sum_{(X_r,Y_r)\in B} X_r + k(n)(X_c + Y_c).$$

This implies

$$0 \leq d(n) - l(n) \leq k(n)(X_c + Y_c).$$

Thus, for $\varepsilon > 0$,

$$\begin{aligned} P\{d(n) - l(n) > \varepsilon\} &\leq P\{k(n)(X_c + Y_c) > \varepsilon\} \\ &\leq \frac{1}{\varepsilon} E[k(n)(X_c + Y_c)] \\ &\leq \frac{1}{\varepsilon} \sqrt{E[k(n)^2] E[(X_c + Y_c)^2]} \\ &\leq \frac{2}{\varepsilon} \sqrt{E[k(n)^2] E[l_c]}, \end{aligned} \tag{4.32}$$

where the last inequality follows because $E(X_c+Y_c)^2 \leq 2E(X_c^2+Y_c^2) \leq 2E(X_c+Y_c) \leq 4E(l_c)$. Now using Lemma 2.3 and Lemma 4.1 we get,

$$P\{d(n) - l(n) > \varepsilon\} \to 0 \qquad \text{as } n \to \infty.$$

☐

Proof of Theorem 1.2 : From Theorems 4.3 and 4.4 it follows that, as $n \to \infty$, $l(n)$ converges weakly and $E(l(n)^j) - E(d(n)^j) \to 0$ for every $j \geq 1$. Thus we need to evaluate $\lim_{n\to\infty} E(d(n))$ and $\lim_{n\to\infty} \mathrm{Var}(d(n))$ to complete the proof of the theorem. Now, from the reflection principle (3.5), (2.3), Proposition 4.1 and (4.19) we have

$$\begin{aligned} E(d(n)) &= E[(\sum_{i=1}^{k(n)}(X_{R_i} + Y_{R_i})] \\ &= 2E(\sum_{i=1}^{k(n)} X_{R_i}) \\ &\to 2\mu_1 = 2 \text{ as } n \to \infty. \end{aligned}$$

Similarly

$$\begin{aligned} E(d(n)^2) &= E[(\sum_{i=1}^{k(n)}(X_{R_i} + Y_{R_i}))^2] \\ &= E[(\sum_{i=1}^{k(n)} X_{R_i})^2] + E[(\sum_{i=1}^{k(n)} Y_{R_i})^2] + 2E[(\sum_{i=1}^{k(n)} X_{R_i})(\sum_{i=1}^{k(n)} Y_{R_i})]. \end{aligned}$$

It follows from the reflection principle (3.5), (2.4) and (4.19) that, as $n \to \infty$, $E[(\sum_{i=1}^{k(n)} X_{R_i})^2] + E[(\sum_{i=1}^{k(n)} Y_{R_i})^2] \to 2\mu_2 = 3$. Also observe that

$$\begin{aligned} E[(\sum_{i=1}^{k(n)} X_{R_i})(\sum_{i=1}^{k(n)} Y_{R_i})] &= E[(\sum_{i=1}^{n} X_i\eta_i)(\sum_{j=1}^{n} Y_j\eta_j)] \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} E(X_i\eta_i Y_j\eta_j) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} E(X_i)E(\eta_i Y_j\eta_j), \end{aligned}$$

where the last equality follows from the independence properties discussed in Section 2.

Next, for $0 < y < 1$

$$\begin{aligned} P(Y_i\eta_i > y) &= P(\min\{Y_1, \dots Y_{i-1}\} > Y_i > y) \\ &= \int_y^1 (\int_{y_i}^1 dv)^{i-1} dy_i \\ &= \int_y^1 (1 - y_i)^{i-1} dy_i, \end{aligned}$$

so,

$$\begin{aligned} E(Y_i\eta_i) &= \int_0^1 y(1-y)^{i-1} dy \\ &= \frac{1}{i} - \frac{1}{i+1}. \end{aligned}$$

Also, for $i < j$, since η_i and η_j are independent (see Lemma 2.2) and η_i, being dependent only on $Y_1, \ldots, Y_i$, is independent of Y_j, we have

$$\begin{aligned} E(\eta_i Y_j \eta_j) &= E(\eta_i) E(Y_j \eta_j) \\ &= \frac{1}{i}(\frac{1}{j} - \frac{1}{j+1}). \end{aligned}$$

While, for $j < i$, we see that, for $0 < y < 1$,

$$\begin{aligned} P(\eta_i Y_j \eta_j \geq y) &= P(\min\{Y_1, \ldots, Y_{j-1}\} > Y_j > y \text{ and } \min\{Y_j, \ldots, Y_{i-1}\} > Y_i) \\ &= \int_y^1 (\int_0^{y_j} (\int_{y_j}^1 dv)^{j-1} (\int_{y_i}^1 dv)^{i-1-j} dy_i) dy_j \\ &= \int_y^1 (i-j)^{-1}[(1-y_j)^{j-1} - (1-y_j)^{i-1}] dy_j, \end{aligned}$$

so,

$$\begin{aligned} E(\eta_i Y_j \eta_j) &= \int_0^1 y(i-j)^{-1}[(1-y)^{j-1} - (1-y)^{i-1}] dy \\ &= \frac{1}{i-j}(\frac{1}{i} - \frac{1}{j} - \frac{1}{i+1} + \frac{1}{j+1}) \\ &= \frac{i+j+1}{ij(i+1)(j+1)}. \end{aligned}$$

Combining the above we have

$$\begin{aligned} \sum_{i=1}^n \sum_{j=1}^n E(X_i) E(\eta_i Y_j \eta_j) &= \sum_{i=1}^n \sum_{j=1}^{i-1} E(X_i) E(\eta_i Y_j \eta_j) + \sum_{i=1}^n E(X_i) E(Y_i \eta_i) \\ &\quad + \sum_{i=1}^n \sum_{j=i+1}^n E(X_i) E(\eta_i Y_j \eta_j) \\ &= \sum_{i=1}^n E(X_i) \sum_{j=1}^{i-1} \frac{i+j+1}{ij(i+1)(j+1)} \\ &\quad + \sum_{i=1}^n E(X_i)(\frac{1}{i} - \frac{1}{i+1}) \\ &\quad + \sum_{i=1}^n E(X_i) \sum_{j=i+1}^n \frac{1}{i}(\frac{1}{j} - \frac{1}{j+1}). \end{aligned}$$

Since $E(X_i) = \frac{i}{n+1}$, we have

$$\begin{aligned} \sum_{i=1}^n E(X_i) \sum_{j=1}^{i-1} \frac{i+j+1}{ij(i+1)(j+1)} &= \sum_{i=1}^n \frac{1}{n+1} \sum_{j=1}^{i-1} (\frac{1}{j(j+1)} + \frac{1}{(i+1)(j+1)}) \\ &= \frac{1}{n+1}[\sum_{i=1}^n (1 - \frac{1}{i}) + \sum_{i=1}^n \frac{1}{i+1} \sum_{j=1}^{i-1} \frac{1}{j+1}] \end{aligned}$$

$$\to \quad 1 \text{ as } n \to \infty,$$

where we have used the fact that

$$0 \le \frac{1}{n+1}\sum_{i=1}^{n}\frac{1}{i+1}\sum_{j=1}^{i-1}\frac{1}{j+1} \le \frac{1}{n+1}\sum_{i=1}^{n}\frac{1}{i+1}\sum_{j=1}^{n}\frac{1}{j+1} \le \frac{(1+\log n)^2}{n+1}.$$

Also,

$$\begin{aligned}\sum_{i=1}^{n} E(X_i)E(Y_i\eta_i) &= \sum_{i=1}^{n}\frac{i}{n+1}(\frac{1}{i}-\frac{1}{i+1})\\ &= \frac{1}{n+1}\sum_{i=1}^{n}\frac{1}{i+1}\\ &\to 0 \text{ as } n\to\infty,\end{aligned}$$

and,

$$\begin{aligned}\sum_{i=1}^{n}E(X_i)\sum_{j=i+1}^{n}\frac{1}{i}(\frac{1}{j}-\frac{1}{j+1}) &= \sum_{i=1}^{n}\frac{i}{n+1}\sum_{j=i+1}^{n}\frac{1}{i}(\frac{1}{j}-\frac{1}{j+1})\\ &= \frac{1}{n+1}\sum_{i=1}^{n}\sum_{j=i+1}^{n}(\frac{1}{j}-\frac{1}{j+1})\\ &= \frac{1}{n+1}\sum_{i=1}^{n}(\frac{1}{i+1}-\frac{1}{n+1})\\ &\to 0 \text{ as } n\to\infty.\end{aligned}$$

Thus we have

$$E[(\sum_{i=1}^{k(n)}X_{R_i})(\sum_{i=1}^{k(n)}Y_{R_i})] \to 1 \text{ as } n\to\infty.$$

This now yields $\mathrm{Var}(l(n)) \to 1$ and completes the proof of the theorem. □

Proof of Theorem 1.3 : To prove Theorem 1.3 let (X'_f, Y'_f) and (X_f, Y_f) be the a.s. unique points (not necessarily distinct) in L_n which are farthest from the origin in terms of Euclidean distance and Manhattan distance respectively, i.e.,

$$\sqrt{(X'_f)^2+(Y'_f)^2} = \max\left\{\sqrt{X_{R_i}^2+Y_{R_i}^2}\right\} \text{ and } X_f+Y_f = \max\{X_{R_i}+Y_{R_i}\}.$$

Let

$$h(n) = \sqrt{(X'_f)^2+(Y'_f)^2}.$$

It follows from Theorem 4.4 that, for $\varepsilon > 0$,

$$\begin{aligned}P\left(|h(n)-(X_f+Y_f)| > \varepsilon\right) &\le P(|d(n)-l(n)|>\varepsilon)\\ &\to 0 \qquad \text{as } n\to\infty. \qquad (4.33)\end{aligned}$$

Now note that, for $\varepsilon > 0$,

$$
\begin{aligned}
&P\left((X_f + Y_f) - \max\{X_1 + Y_1, X_{k(n)} + Y_{k(n)}\} > \varepsilon\right) \\
&\quad = \quad P\left((X_f + Y_f) - (X_{k(n)} + Y_{k(n)}) > \varepsilon, X_1 + Y_1 < X_{k(n)} + Y_{k(n)}\right) \\
&\qquad\qquad + P\left((X_f + Y_f) - (X_1 + Y_1) > \varepsilon, X_1 + Y_1 > X_{k(n)} + Y_{k(n)}\right) \\
&\quad = \quad E_1 + E_2 \text{ (say)},
\end{aligned}
$$

and

$$
\begin{aligned}
E_1 \quad = & \quad P\big((X_f + Y_f) - (X_{k(n)} + Y_{k(n)}) > \varepsilon, X_1 + Y_1 < X_{k(n)} + Y_{k(n)}, \\
& \qquad\qquad X_f \leq X_c\big) \\
+ & \quad P\big((X_f + Y_f) - (X_{k(n)} + Y_{k(n)}) > \varepsilon, X_1 + Y_1 < X_{k(n)} + Y_{k(n)}, \\
& \qquad\qquad X_f > X_c\big) \\
= & \quad E_{11} + E_{12} \text{ (say)},
\end{aligned}
$$

where (X_c, Y_c) is as defined in (4.28).

To evaluate E_{11}, note that if $(X_f + Y_f) - (X_{k(n)} + Y_{k(n)}) > \varepsilon$ then either $Y_f - X_{k(n)} > \varepsilon/2$ or $X_f - Y_{k(n)} > \varepsilon/2$. Along with the observation that $Y_f \leq Y_1$ and $Y_{k(n)} \leq Y_c$ we get

$$
\begin{aligned}
E_{11} \quad \leq & \quad P\left(Y_f - X_{k(n)} > \frac{\varepsilon}{2}, X_1 + Y_1 < X_{k(n)} + Y_{k(n)}, X_f \leq X_c\right) \\
& \quad + P\left(X_f - Y_{k(n)} > \frac{\varepsilon}{2}, X_1 + Y_1 < X_{k(n)} + Y_{k(n)}, X_f \leq X_c\right) \\
\leq & \quad P\left(Y_1 - X_{k(n)} > \frac{\varepsilon}{2}, X_1 + Y_1 < X_{k(n)} + Y_{k(n)}, X_f \leq X_c\right) \\
& \qquad\qquad + P\left(X_c + Y_c > \frac{\varepsilon}{2}\right) \\
\leq & \quad P\left(Y_{k(n)} - X_1 > \frac{\varepsilon}{2}\right) + P\left(X_c + Y_c > \frac{\varepsilon}{2}\right) \\
\leq & \quad 2P\left(X_c + Y_c > \frac{\varepsilon}{2}\right) \\
\to & \quad 0 \qquad \text{as } n \to \infty.
\end{aligned}
$$

For E_{12} observe that $X_f > X_c$ if and only if $Y_f < Y_c$ and using calculations similar to that above we have E_{12} and hence E_1 tend to zero as $n \to \infty$.

Also, similar arguments show that E_2 tends to zero as $n \to \infty$, which implies that

$$
P\left((X_f + Y_f) - \max\{X_1 + Y_1, X_{k(n)} + Y_{k(n)}\} > \varepsilon\right) \to 0 \tag{4.34}
$$

as $n \to \infty$.

Now observe that, for any $\varepsilon > 0$,

$$
P\left(\max\{X_1 + Y_1, X_{k(n)} + Y_{k(n)}\} - \max\{Y_1, X_{k(n)}\} > \varepsilon\right) \to 0 \tag{4.35}
$$

as $n \to \infty$. Indeed this follows from the observation that $X_1 \leq X_c$ and $Y_{k(n)} \leq Y_c$ and that Lemma 4.1 implies that both X_c and Y_c converge in probability to 0 as $n \to \infty$.

Recall that the sequences $(X_1, Y_1), \ldots, (X_n, Y_n)$ arose as a permutation of the vectors $(\xi_1, \zeta_1), \ldots, (\xi_n, \zeta_n)$ such that $X_1 = \xi_l$ where $\xi_l = \min\{\xi_1, \ldots, \xi_n\}$ and $Y_{k(n)} = \zeta_m$ where $\zeta_m = \min\{\zeta_1, \ldots, \zeta_n\}$. Thus (4.33), (4.34) and (4.35) along with the following lemma completes the proof of Theorem 1.3.

Lemma 4.2 *For two independent sequences $\xi_1, \xi_2, \ldots$ and $\zeta_1, \zeta_2, \ldots$ of i.i.d. random variables, each random variable being uniformly distributed on $[0,1]$, define $\mu(n)$ and $\nu(n)$ by*

$$\xi_{\mu(n)} = \min\{\xi_1, \ldots, \xi_n\} \text{ and } \zeta_{\nu(n)} = \min\{\zeta_1, \ldots, \zeta_n\}.$$

Then $\max\{\xi_{\nu(n)}, \zeta_{\mu(n)}\}$ converges in distribution to $\max\{U_1, U_2\}$, where U_1 and U_2 are i.i.d. random variables, each distributed uniformly on $[0,1]$.

Proof : For $\alpha \in I\!R$,

$$\begin{aligned} & P\left\{\max\{\xi_{\nu(n)}, \zeta_{\mu(n)}\} < \alpha\right\} \\ = & \sum_{l=1}^{n}\sum_{m=1}^{n} P\{\xi_m < \alpha, \zeta_l < \alpha, \mu(n) = l, \nu(n) = m\} \\ = & \sum_{l=1}^{n}\sum_{m=1}^{n} P\{\xi_m < \alpha, \mu(n) = l\} P\{\zeta_l < \alpha, \nu(n) = m\}, \end{aligned} \tag{4.36}$$

by the independence of the two sequences of random variables.

Now, for $0 < \alpha < 1$ and $l \neq m$,

$$\begin{aligned} & P\{\xi_m < \alpha, \mu(n) = l\} \\ = & P\{\mu(n) = l\} - P\{\xi_m > \alpha, \xi_l < \xi_j \text{ for all } j = 1, \ldots, l-1, l+1, \ldots, n\} \\ = & \frac{1}{n} - \int_0^1 dy_l \int_{\max\{\alpha, y_l\}}^1 dy_m \prod_{j \neq l,m} \left(\int_{y_l}^1 dy_j\right) \\ = & \frac{1}{n} - \left(\frac{1-\alpha}{n-1} - \frac{(1-\alpha)^n}{n(n-1)}\right). \end{aligned}$$

Similarly, for $0 < \alpha < 1$

$$P\{\zeta_m < \alpha, \nu(n) = l\} = \frac{1}{n} - \left(\frac{1-\alpha}{n-1} - \frac{(1-\alpha)^n}{n(n-1)}\right) \text{ for } l \neq m.$$

For $l = m$ and $0 < \alpha < 1$, similar calculations as above yield

$$P\{\xi_l < \alpha, \mu(n) = l\} = \frac{1}{n} - \frac{(1-\alpha)^n}{n}.$$

Thus, for $0 < \alpha < 1$, from (4.36), we have

$$\begin{aligned} & P\left\{\max\{\xi_{\nu(n)}, \zeta_{\mu(n)}\} < \alpha\right\} \\ = & \sum_{l=1}^{n}\sum_{l \neq m=1}^{n} \frac{1}{n^2}\left(1 - (\frac{n(1-\alpha)}{n-1} - \frac{(1-\alpha)^n}{n-1})\right)^2 + \sum_{l=1}^{n} \frac{1}{n^2}\left(1 - (1-\alpha)^n\right)^2 \\ \to & \ \alpha^2 \text{ as } n \to \infty. \end{aligned}$$

This proves the lemma. □

5 Moments

We now present a calculation to obtain $\lim_{n\to\infty} E[\sum_{i=1}^{k(n)} X_{R_i}]^l$. Note that this limit may be used instead of Propositions 2.1 and 4.1 to obtain Theorem 1.2.

Let $a(n,l) := E[\sum_{i=1}^{k(n)} X_{R_i}]^l$. In the sequel we use the notation

$$\sum_{\mathbf{l}_p=q} := \sum_{l_1+\cdots+l_p=q} \quad\text{and}\quad \sum_{1\le \mathbf{i}_p\le q} := \sum_{1\le i_1<\cdots<i_p\le q}.$$

By (4.3) we have

$$\begin{aligned}
a(n+1,l) &= \sum_{k=1}^{n+1}\sum_{\mathbf{l}_k=l}\sum_{1\le \mathbf{i}_k\le n+1} \frac{l!}{l_1!\cdots l_k!}\frac{1}{i_1\dots i_k}\frac{(n+1)!}{(n+1+l)!} \\
&\qquad\times\prod_{j=1}^{k}\frac{(i_j-1+\sum_{s=1}^{j} l_s)!}{(i_j-1+\sum_{s=1}^{j-1} l_s)!} \\
&= \sum_{k=1}^{n+1}\sum_{t=1}^{l}\sum_{\mathbf{l}_{k-1}=l-t}\sum_{1\le \mathbf{i}_{k-1}\le n} \frac{l!}{l_1!\cdots l_{k-1}!t!}\frac{1}{i_1\dots i_{k-1}(n+1)} \\
&\qquad\times\frac{(n+1)!}{(n+1+l)!}\prod_{j=1}^{k-1}\frac{(i_j-1+\sum_{s=1}^{j} l_s)!}{(i_j-1+\sum_{s=1}^{j-1} l_s)!}\frac{(n+l)!}{(n+l-t)!} \\
&\quad+\sum_{k=1}^{n}\sum_{\mathbf{l}_k=l}\sum_{1\le \mathbf{i}_k\le n} \frac{l!}{l_1!\cdots l_k!}\frac{1}{i_1\dots i_k}\frac{(n+1)!}{(n+1+l)!} \\
&\qquad\times\prod_{j=1}^{k}\frac{(i_j-1+\sum_{s=1}^{j} l_s)!}{(i_j-1+\sum_{s=1}^{j-1} l_s)!} \\
&= T_1+T_2 \text{ (say)}.
\end{aligned}$$

From (4.3) we see that

$$T_2 = \frac{n+1}{n+1+l}a(n,l).$$

We write the term T_1 as

$$\begin{aligned}
T_1 &= \sum_{t=1}^{l-1}\frac{l!}{t!}\frac{1}{n+1}\frac{(n+1)!}{(n+1+l)!}\frac{(n+l)!}{(n+l-t)!}\sum_{k=1}^{n}\sum_{\mathbf{l}_k=l-t}\sum_{1\le \mathbf{i}_k\le n}\frac{1}{l_1!\cdots l_k!}\frac{1}{i_1\dots i_k} \\
&\qquad\times\prod_{j=1}^{k}\frac{(i_j-1+\sum_{s=1}^{j} l_s)!}{(i_j-1+\sum_{s=1}^{j-1} l_s)!} \\
&\quad+\frac{1}{n+1}\frac{(n+1)!}{(n+1+l)!}\frac{(n+l)!}{n!} \\
&= T_{11}+\frac{1}{n+1+l}\text{ (say)}.
\end{aligned}$$

The term T_{11} may now be simplified as

$$T_{11} = \sum_{t=1}^{l-1}\binom{l}{t}\frac{1}{n+1+l}\sum_{k=1}^{n}\sum_{\mathbf{l}_k=l-t}\sum_{1\le \mathbf{i}_k\le n}\frac{(l-t)!}{l_1!\cdots l_k!}\frac{1}{i_1\dots i_k}\frac{n!}{(n+l-t)!}$$

$$\times \prod_{j=1}^{k} \frac{(i_j - 1 + \sum_{s=1}^{j} l_s)!}{(i_j - 1 + \sum_{s=1}^{j-1} l_s)!}$$

$$= \sum_{t=1}^{l-1} \binom{l}{t} \frac{1}{n+1+l} a(n, l-t).$$

Combining the above and noting that $a(n,0) = 1$ for all n, we have that

$$a(n+1,l) = \frac{1}{n+1+l}\left[(n+1)a(n,l) + \sum_{t=1}^{l} \binom{l}{t} a(n,l-t)\right],$$

from which, by a change of variables, we have

$$(n+1+l)a(n+1,l) = (n+1)a(n,l) + \sum_{t=0}^{l-1} \binom{l}{t} a(n,l-t). \tag{5.1}$$

For $l \geq 0$, define a sequence γ_l by

$$\gamma_0 = 1,\ \gamma_1 = 1 \text{ and, for } l \geq 2,\ \gamma_l = \frac{1}{l} \sum_{t=0}^{l-1} \gamma_t \binom{l}{t}. \tag{5.2}$$

Then $a(n,l) = \gamma_l$ for all n, is a solution of equation (5.1) with $a(n,0) = 1$.

Now let $b(n,l)$ be a general solution of (5.1). Then we claim $b(n,l) - \gamma_l$ tends to zero as $n \to \infty$. If $b(n,l)$ is a solution of (5.1), then $d(n,l) = b(n,l) - \gamma_l$ is also a solution of (5.1), with $d(n,0) = 0$. We show, by induction, that $d(n,l)$ tends to zero as $n \to \infty$.

Indeed, for $l = 1$,

$$\begin{aligned}(n+2)d(n+1,1) &= (n+1)d(n,1) + \sum_{t=0}^{l-1} \binom{l}{t} d(n,0) \\ &= (n+1)d(n,1).\end{aligned}$$

Hence $(n+1)d(n,1)$ is a constant function and hence $d(n,1)$ tends to 0 as $n \to \infty$. In fact, $d(n,1) = O(n^{-1})$ as $n \to \infty$. More generally assume $d(n,r) = O(n^{-1})$ as $n \to \infty$, for $r = 0, 1, 2, \ldots, l-1$.

From (5.1),

$$\begin{aligned}(n+l+1)d(n+1,l) - (n+1)d(n,l) &= \sum_{t=0}^{l-1} \binom{l}{t} d(n,l-t) \\ &= O(n^{-1}) \text{ as } n \to \infty.\end{aligned}$$

Now the left side above is not telescopic, however multiplying by $(n+2)\cdots(n+l)$, we have

$$\frac{(n+l+1)!}{(n+1)!} d(n+1,l) - \frac{(n+l)!}{n!} d(n,l) = O(n^{l-2}) \text{ as } n \to \infty.$$

Summing both sides for $n = 1, \cdots, N-1$,

$$\frac{(N+l+1)!}{(N+1)!} d(N+1,l) - l!d(1,l) = O(N^{l-2}) \text{ as } N \to \infty.$$

Hence, as $N \to \infty$, $\frac{(N+l+1)!}{(N+1)!} d(N,l) = O(N^{l-2})$, i.e., $d(N,l) = O(N^{-1})$. This proves the claim and establishes

$$\lim_{n\to\infty} E[\sum_{i=1}^{k(n)} X_{R_i}]^l = \gamma_l,$$

where γ_l is as given in (5.2).

Acknowledgements

We thank K. R. Parthasarathy, R. Balasubramanian for helpful suggestions and an anonymous referee for pointing out several mistakes.

References

1. Aldous, D. and J.M. Steele, [1992]. Asymptotics for Euclidean minimal spanning trees on random points, *Probab. Th. Rel. Fields*, **92**, 247 – 258.
2. Beardwood, J., J.H. Halton and J.M. Hammersley, [1959]. The shortest path through many points, *Proc. Camb. Phil. Soc.*, **55**, 299 – 327.
3. Chung, K.L., [1974]. *A course in probability theory* (2nd. edition), Academic Press, New York.
4. David H.A., [1970] *Order statistics*, Wiley, New York.
5. Feller, W., [1978]. *An introduction to probability theory and its applications*, Wiley, New York.
6. Gilbert, E.N., [1961]. Random plane networks, *J. Soc. Indust. Appl.*, **9**, 533 – 543.
7. Gradshteyn, I.S. and I.M. Ryzhik, [1980]. *Table of integrals, series and products* Academic Press, New York.
8. Gupta, P and P.R. Kumar, [1998]. Critical power for asymptotic connectivity in wireless networks, *Stochastic analysis, control, optimization and applications: A volume in honor of W.H. Fleming*, W.M. McEneaney, G. Yin and Q. Zhang (edited), Birkhauser, Boston.
9. Kesten, H. and S. Lee, [1996]. The central limit theorem for weighted minimal spanning trees on random points, *Ann. Appl. Probab.*, **6**, 495 – 527.
10. Rényi, A., [1976]. *On the extreme elements of observations*, in Selected papers of Alfred Rényi, Vol. 3, Akadémiai Kiadó, 50 – 65.
11. Rodriguez-Iturbe, I. and A. Rinaldo, [1997]. *Fractal river basins, chance and self-organization* Cambridge University Press, New York.
12. Shohat, J.A. and J.D. Tamarkin, [1960], *The problem of moments*, Mathematical Surveys No. 1, American Mathematical Society, Providence.
13. Steele, J.M., [1988], Growth rates of Euclidean minimal spanning trees with power weighted edges, *Ann. Probab.*, **16**, 1767 – 1787.