On Random Line Segments in the Unit Square

Thomas A. Courtade
Department of Electrical Engineering
University of California
Los Angeles, California 90095
Email: tacourta@ee.ucla.edu

I. Introduction

Let $Q = [0,1] \times [0,1]$ denote the unit square and let $\mathcal{L}_n$ be a set of n line segments in Q. Two line segments are said to be *crossing* if they intersect at any point. A subset of line segments is called *non-crossing* if no two segments in the subset are crossing.

Consider the scenario where the endpoints of the n line segments are randomly distributed, independently and uniformly, in Q. Define $N(\mathcal{L}_n)$ to be the size of the largest non-crossing subset of segments. Formally:

$$N(\mathcal{L}_n) = \max_{\mathcal{U} \subseteq \mathcal{L}_n} \{|\mathcal{U}| : \ell_1, \ell_2 \text{ do not cross for all } \ell_1, \ell_2 \in \mathcal{U}\},$$

where $|\mathcal{U}|$ denotes the number of line segments in the subset $\mathcal{U}$. Further, define $L(\mathcal{L}_n)$ to be the maximum *sum-length* over all non-crossing subsets of $\mathcal{L}_n$. Formally:

$$L(\mathcal{L}_n) = \max_{\mathcal{U} \subseteq \mathcal{L}_n} \left\{ \sum_{\ell \in \mathcal{U}} \|\ell\| : \ell_1, \ell_2 \text{ do not cross for all } \ell_1, \ell_2 \in \mathcal{U} \right\},$$

where $\|\ell\|$ denotes the length of line segment ℓ. In this short note, we prove that $N(\mathcal{L}_n)$ and $L(\mathcal{L}_n)$ are both $\Theta(\sqrt{n})$ asymptotically almost surely. It is possible to extend these results to the case where line segments are required to be separated by some minimum distance. Partial characterizations of this more general case are given in Section II-B.

II. Main Results and Proofs

A. Non-Crossing Line Segments

Our first two results are that $N(\mathcal{L}_n)$ and $L(\mathcal{L}_n)$ both behave roughly like $\sqrt{n}$:

Theorem 1: Asymptotically almost surely, $\sqrt{n}/2 \leq N(\mathcal{L}_n) \leq 15\sqrt{n}$.

Theorem 2: Asymptotically almost surely, $\sqrt{n}/7 \leq L(\mathcal{L}_n) \leq 22\sqrt{n}$.

While we have made no significant efforts to optimize the coefficients of the bounds in Theorems 1 and 2, minor modifications to our arguments can yield somewhat tighter results. Additional techniques are likely required in order to close the gap completely. The case where line segments are required to be separated by some minimum distance is discussed in Section II-B.

Proof of Theorem 1: Throughout the proof, we make the distinction between *left* and *right* endpoints of line segments. This is somewhat arbitrary, but simplifies the argument significantly. Thus, a line segment is generated according to the following process: (Step 1) the left endpoint is chosen uniformly from Q, and (Step 2) the right endpoint is chosen independently and uniformly from Q.

Claim 1: With probability tending to 1 as $n \to \infty$, $N(\mathcal{L}_n) \geq \sqrt{n}/2$.
Partition Q into $\sqrt{n}$ disjoint horizontal strips, each having height $1/\sqrt{n}$ and width 1. Note that if a line segment ℓ is contained in a single strip, then it will not intersect line segments contained in any other strip. Then $N(\mathcal{L}_n) \geq Y$, where Y is the number of strips that contain line segments. Observe that

$$\begin{aligned}\Pr\left[\text{Line } \ell \text{ in strip } j\right] &= \Pr\left[\{\text{left endpoint of } \ell \text{ in strip } j\} \wedge \{\text{right endpoint of } \ell \text{ in strip } j\}\right] \\ &= \Pr\left[\text{left endpoint of } \ell \text{ in strip } j\right] \times \Pr\left[\text{right endpoint of } \ell \text{ in strip } j\right] \\ &= \frac{1}{n}.\end{aligned}$$

Where we used the fact that the probability a given point falls in a particular strip is $1/\sqrt{n}$ and points are chosen independently. Then, the probability that a given strip does not contain any line segments is:

$$(1 - 1/n)^n \approx 1/e.$$

Further, note that:

$$\begin{aligned}&\Pr\left[\{\text{Line } \ell \text{ not in strip } i\} \wedge \{\text{Line } \ell \text{ not in strip } j\}\right] \\ &= 1 - \Pr\left[\{\text{Line } \ell \text{ in strip } i\} \vee \{\text{Line } \ell \text{ in strip } j\}\right] \\ &= 1 - \left(\Pr\left[\text{Line } \ell \text{ in strip } i\right] + \Pr\left[\text{Line } \ell \text{ in strip } j\right]\right) \\ &= 1 - 2/n.\end{aligned}$$

Where we used the fact that the events $\{\text{Line } \ell \text{ in strip } i\}$ and $\{\text{Line } \ell \text{ in strip } j\}$ are disjoint. Then for any pair of strips (i, j), the probability that neither strip i nor strip j contains any line segments is

$$(1 - 2/n)^n \approx 1/e^2.$$

Let X_i be the indicator random variable taking value 1 if strip i contains no line segments and taking the value 0 otherwise. Note that $\mathbb{E}X_i \approx e^{-1}$, Var$(X_i) \approx e^{-1}(1 - e^{-1})$, and

$$\begin{aligned}\text{Cov}(X_i, X_j) &= \mathbb{E}\left[X_i X_j\right] - \mathbb{E}\left[X_i\right]\mathbb{E}\left[X_j\right] \\ &= (1 - 2/n)^n - (1 - 1/n)^{2n} \\ &\leq e^{-2} - e^{-2} + o(1) \\ &= o(1).\end{aligned}$$

Then, letting $X = \sum_{i=1}^{\sqrt{n}} X_i$ be the number of strips that don't contain any line segments, and noting that

$$\begin{aligned}\text{Var}(X) &= \sum_i \text{Var}(X_i) + \sum_{i \neq j} \text{Cov}(X_i, X_j) \\ &\leq \sqrt{n}\left(\frac{1}{e}\left(1 - \frac{1}{e}\right) + o(1)\right) + \sqrt{n}(\sqrt{n} - 1)o(1) \\ &\leq \sqrt{n} + n \cdot o(1),\end{aligned}$$

Chebyshev's inequality yields:

$$\begin{aligned}\Pr\left[|X - \mathbb{E}X| \geq \frac{1}{10}\sqrt{n}\right] &\leq 100 \times \frac{\text{Var}(X)}{n} \\ &\leq 100 \times \frac{\sqrt{n} + n \cdot o(1)}{n} \\ &\to 0.\end{aligned}$$

Therefore, with probability tending to 1,

$$X \leq \left(1 + \frac{1}{10} + o(1)\right) \frac{\sqrt{n}}{e} \leq \frac{\sqrt{n}}{2}.$$

Hence, $Y = \sqrt{n} - X \geq \sqrt{n}/2$ with probability tending to 1. This proves the claim.

We briefly remark that it is possible, with probability tending to 1, to find $\Omega_\epsilon(\sqrt{n})$ non-crossing segments[1] of length greater than $1 - \epsilon$. The basic idea is to modify the above proof so that the left (right) endpoint of each line segment is within distance $\epsilon/2$ of the left (right) edge of Q. We omit the details in the interest of brevity.

Claim 2: With probability tending to 1 as $n \to \infty$, $N(\mathcal{L}_n) \leq 15\sqrt{n}$.

From [1], there exists an absolute constant c such that for any $2k$ points in the plane, the number of non-crossing left-right[2] perfect matchings is upper-bounded by $c \cdot 29^k$. Consider any realization of n line segments in the plane and further consider the $2k$ (k left and k right) endpoints corresponding to any subset S consisting of k line segments. Conditioned on the locations of the left and right endpoints, every left-right perfect matching of these $2k$ points is equally likely, and thus the probability that these k segments are non-crossing is upper bounded by:

$$\Pr\left[S \text{ is non-crossing}\right] \leq \frac{c \cdot 29^k}{k!}$$

since there are $k!$ left-right perfect matchings on the $2k$ endpoints.

Stirling's formula states

$$\lim_{k\to\infty} \frac{k!}{\sqrt{2\pi k}\left(\frac{k}{e}\right)^k} = 1,$$

and thus

$$\frac{c \cdot 29^k}{k!} \leq o(1) \cdot \left(\frac{29 \cdot e}{k}\right)^k.$$

Recalling the crude upper bound $\binom{n}{k} \leq \left(\frac{n \cdot e}{k}\right)^k$, a union bound gives:

$$\begin{aligned}\Pr[\exists\ k \text{ non-crossing segments}] &\leq \binom{n}{k} \frac{c \cdot 29^k}{k!} \\ &\leq o(1) \cdot \left(\frac{29 \cdot n \cdot e^2}{k^2}\right)^k\end{aligned}$$

Letting $k = 15\sqrt{n}$, we have

$$\begin{aligned}\Pr[\exists\ 15\sqrt{n} \text{ non-crossing segments}] &\leq o(1) \cdot \left(\frac{29 \cdot e^2}{15^2}\right)^{15\sqrt{n}} \\ &\leq o(1) \cdot (.96)^{15\sqrt{n}} \\ &\to 0.\end{aligned}$$

This proves the claim and completes the proof of Theorem 1. ■

[1] A function $f(n)$ is said to be $\Omega_\epsilon(g(n))$, if there exists an integer n_0 and a constant c_ϵ depending only on ϵ so that $f(n) \geq c_\epsilon g(n)$ for all $n > n_0$.

[2] A left-right perfect matching distinguishes between left endpoints and right endpoints in edges. In other words, an edge is only allowed to match a left endpoint to a right endpoint.

Proof of Theorem 2:

We claim that, asymptotically almost surely, there exists a subset of non-crossing line segments $\mathcal{U} \subseteq \mathcal{L}_n$ satisfying

$$\sum_{\ell \in \mathcal{U}} \|\ell\| > \frac{1}{7}\sqrt{n}, \tag{1}$$

where $\|\ell\|$ denotes the length of line segment ℓ.

Indeed, consider the non-crossing set of line segments constructed in the proof of the lower bound of Theorem 1. Denote this set $\mathcal{U}$. Let the endpoints of a line segment $\ell \in \mathcal{U}$ be given in Cartesian form: (x_1, y_1) and (x_2, y_2). Note that $\|\ell\| \geq |x_1 - x_2| := L_\ell$, where the random variable L_ℓ is independent from the event $\{\ell \in \mathcal{U}\}$, since this event only depends on y_1 and y_2. Moreover, $\{L_\ell\}_{\ell \in \mathcal{U}}$ is a set of independent identically distributed random variables, each with expectation $1/3$ and finite variance. Therefore, by the weak law of large numbers

$$\Pr\left(\left|\frac{1}{|\mathcal{U}|}\sum_{\ell \in \mathcal{U}} L_\ell - \frac{1}{3}\right| > \frac{1}{21}\right) \to 0.$$

Some basic algebra combined with the fact that $|\mathcal{U}| \geq \sqrt{n}/2$ a.a.s. proves the claim.

The estimate given in (1) is essentially the best possible. Indeed, the upper bound of Theorem 1 combined with the fact that $\|\ell\| \leq \sqrt{2}$ yields $\sum_{\ell \in \mathcal{U}} \|\ell\| < 22\sqrt{n}$ for any non-crossing subset $\mathcal{U}$ a.a.s. This completes the proof of the theorem. ■

B. Extension to d-Disjoint Line Segments

Two apparently stronger results can be deduced via a slight modification of the previous proofs. We state the results and sketch the proofs in this section.

First, define the distance between two line segments $\ell_1, \ell_2 \in \mathcal{L}_n$ as the minimum distance between any point in ℓ_1 and any point in ℓ_2. Formally,

$$d(\ell_1, \ell_2) := \inf_{x \in \ell_1, y \in \ell_2} \|x - y\|,$$

where $\|x - y\|$ is the Euclidean distance between points $x, y \in Q$. A set $\mathcal{U}$ of line segments is said to be d-disjoint if all pairs of line segments in $\mathcal{U}$ are at least distance d apart. Then

$$N_d(\mathcal{L}_n) = \max_{\mathcal{U} \subseteq \mathcal{L}_n} \{|\mathcal{U}| : d(\ell_1, \ell_2) > d \text{ for all } \ell_1, \ell_2 \in \mathcal{U}\}$$

is the size of the largest d-disjoint subset of line segments in $\mathcal{L}_n$. Similarly, define

$$L_d(\mathcal{L}_n) = \max_{\mathcal{U} \subseteq \mathcal{L}_n} \left\{\sum_{\ell \in \mathcal{U}} \|\ell\| : d(\ell_1, \ell_2) > d \text{ for all } \ell_1, \ell_2 \in \mathcal{U}\right\}.$$

Note that $N(\mathcal{L}_n) = N_0(\mathcal{L}_n)$ and $L(\mathcal{L}_n) = L_0(\mathcal{L}_n)$. With these definitions in hand, we state our last two results:

Theorem 3: Asymptotically almost surely, $N_{d(n)}(\mathcal{L}_n) = \Theta(\sqrt{n})$ if $d(n) = O(n^{-1/4})$.

Theorem 4: Asymptotically almost surely:

- $L_{d(n)}(\mathcal{L}_n) = \begin{cases} \Theta(1/d(n)) & \text{if } d(n) = \Omega(1/\sqrt{n}) \\ \Omega(\sqrt{n}) & \text{if } d(n) = o(1/\sqrt{n}) \end{cases}$
- $L_d(\mathcal{L}_n) < 9/d$.

Proof of Theorem 3: Repeat the constructive part in the proof of Theorem 1, except divide Q into squares of size $n^{-1/4} \times n^{-1/4}$. Asymptotically almost surely, this yields at least $\sqrt{n}/2$ disjoint squares containing lines. Since $d(n) = O(n^{-1/4})$, we can find a constant fraction of the boxes containing lines that are pairwise distance $d(n)$ apart. This shows that we can find $\Omega(\sqrt{n})$ line segments which are $d(n)$-disjoint.

Since $N_{d(n)}(\mathcal{L}_n)$ is non-increasing in $d(n)$, the upper bound of Theorem 1 proves the converse part. ∎

Proof of Theorem 4: To prove the lower bounds, repeat the constructive part in the proof of Theorem 1. Assume first that $d(n) = \Omega(1/\sqrt{n})$. Now, we can partition the strips (taking every $(d(n)\sqrt{n})^{th}$ strip) into classes of roughly equal size so that all strips within a class are separated by a distance greater than $d(n)$. By the pigeon-hole principle, at least one of these classes contains at least $\sqrt{n}/(2 \times M_n)$ $d(n)$-disjoint line segments a.a.s., where M_n is the number of classes in the partition. Some simple arithmetic reveals that M_n is roughly $d(n)\sqrt{n}$. Thus, there exists $\Omega(1/d(n))$ $d(n)$-disjoint line segments a.a.s. If $d(n) = o(1/\sqrt{n})$, taking every other strip is sufficient. Hence there exist $\Omega(\sqrt{n})$ $d(n)$-disjoint line segments in this case. In both cases, we can apply the weak law of large numbers (as in the proof of Theorem 2) to give the desired lower estimate for $L_{d(n)}(\mathcal{L}_n)$.

For the converse, let $\mathcal{U}$ be a subset of d-disjoint line segments, and for $\ell \in \mathcal{U}$, generate a rectangle of dimension $d \times \|\ell\|$ by "widening" the segment equally on both sides. Since $\mathcal{U}$ is a subset of d-disjoint line segments, these rectangles must be pairwise disjoint. Since the sum of the areas of the rectangles is at most slightly larger than the area of Q (rectangles can extend beyond the unit square), a simple volumetric argument then shows that $L_d(\mathcal{L}_n) \leq (1+d)^2/d$. Without loss of generality, $d \leq \sqrt{2}$, thus $L_d(\mathcal{L}_n) < 9/d$. ∎

III. Concluding Remarks

We have given several concentration results for n random line segments in the unit square. First, we show that the maximum number of non-crossing segments behaves roughly like $\sqrt{n}$. Second, we show that there is a non-crossing subset of line segments whose lengths sum to roughly $\sqrt{n}$, which is the maximum possible. Finally, if line segments are required to be separated by some minimum distance d, we obtain a partial characterization of the previous two quantities. These results can be applied to the study of user capacity and transport capacity in random geometric graphs.

An interesting direction for further work would be to study $N_{d(n)}(\mathcal{L}_n)$ when $d(n) = \omega(n^{-1/4})$, and $L_{d(n)}(\mathcal{L}_n)$ when $d(n) = o(1/\sqrt{n})$ more thoroughly. Another important direction for future work would be to examine the behavior of a maximal subset of non-crossing (or d-disjoint) line segments that is chosen greedily. We conjecture that a greedily selected non-crossing (or d-disjoint) maximal subset of segments exhibits behavior which is order-identical to an optimally chosen subset. Experimental evidence appears to support this conjecture.

Acknowledgements

This research was performed while the author was with Trellisware Technologies. The author gratefully acknowledges that, in the proof of Claim 1, the idea to divide Q into strips is due to Michael Neely of USC. The author would also like to thank Tom Halford of Trellisware Technologies for the many conversations from which these results transpired.

References

[1] M. Sharir and E. Welzl, On the number of crossing-free matchings, cycles, and partitions, SIAM J. Comput. 36(3) (2006), 695-720.
[2] N. Alon and J. Spencer. The probabilistic method (2ed). New York: Wiley-Interscience (2000). ISBN 0-471-37046-0.