

Number of Regions Created by Random Chords in the Circle

Shi Feng

University of Washington, sf599@cornell.edu

Follow this and additional works at: <https://scholar.rose-hulman.edu/rhumj>



Part of the [Probability Commons](#)

Recommended Citation

Feng, Shi (2023) "Number of Regions Created by Random Chords in the Circle," *Rose-Hulman Undergraduate Mathematics Journal*: Vol. 24: Iss. 2, Article 4.

Available at: <https://scholar.rose-hulman.edu/rhumj/vol24/iss2/4>

Number of Regions Created by Random Chords in the Circle

Cover Page Footnote

This paper would not be possible without the help from my mentor Professor Soumik Pal, who has provided the main idea of Stein's method for the problem and commented on every sentence of this paper. I am grateful to Professor David Aldous for suggesting references and giving comments for this paper. I am thankful to my friend Ami Oka for providing help on the computer simulation. I also thank an anonymous referee for a careful review of this work and for providing several helpful comments.

Number of regions created by random chords in the circle

By *Shi Feng*

Abstract. In this paper we discuss the number of regions in a unit circle after drawing n i.i.d. random chords in the circle according to a particular family of distributions. We find that as n goes to infinity, the distribution of the number of regions, properly shifted and scaled, converges to the standard normal distribution and the error can be bounded by Stein's method for proving Central Limit Theorem.

1 Introduction

1.1 Literature review

Geometric probability has an old history starting 17th century (Kalousová, 2011). However, we usually consider Buffon to be the starter for geometric probability in 18th century, who raised the "Buffon's Needle Problem" (Buffon, 1777). We focus on one of the branches in geometric probability called the random chord. One of the most famous problem about random chords is the Bertrand paradox (Bertrand, 1889). It asks what is the probability for a random chord in a unit circle to have a length larger than $\sqrt{3}$. Then in 1964, David and Fix started to investigate random chords rigorously to study random mechanism of chromosomes in cell (David & Fix, 1964), where five randomization models are described. The result is repeated in the book *Geometric Probability* (Solomon, 1978), from where we will borrow the definitions for our problem.

Consider the unit circle. A chord can be identified by two parameters, the distance from the origin and the direction of the chord with respect to the origin. To create a random chord, we use the half the angle subtended by the chord (θ , as shown in Figure 1) to measure the distance from the origin and use the position of the left endpoint from the perspective of the origin (shown as the point C in Figure 1) to measure direction. We assume θ is a continuous random variable with a probability density function f (we call it distance distribution) in domain $[0, \frac{\pi}{2}]$, and the left endpoint is uniformly distributed on the circumference of the circle, and they are independent. With these two parameters, we can define every chord in the circle.

Mathematics Subject Classification. 60D05

Keywords. random chords, Stein's method, central limit theorem

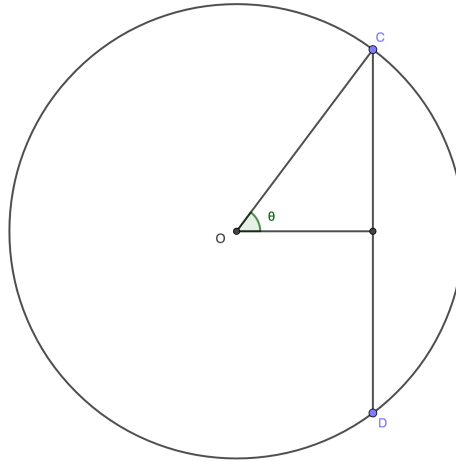


Figure 1: \overline{CD} is the random chord and θ is the angle that represents the chord.

This paper focus on the number of regions in the circle created by those random chords. After drawing n chords, the unit disc is partitioned into disjoint regions through which no chord passes. For example, if no chord is drawn, there is only one region, which is the area enclosed by the circle. After drawing one random chord, the circle is cut into two parts and there are two regions. If two random chords are drawn, there will be two cases. If the two chords have an intersection, there will be 4 regions. On the other hand, if the two chords have no intersection, there will only be 3 regions. Our question is the distribution of the number of regions after drawing n random chords when n is large. For example, in Figure 2, 4 chords are drawn and there are 8 regions in the circle.

Closely related to our paper is the Poisson line process (PLP), which is derived from Poisson point process (PPP). A planar (2D) line can be identified by two parameter (Dhillon & Chetlur, 2020) "(i) the perpendicular distance y of the line from the origin, and (ii) the angle x subtended by the perpendicular dropped onto the line from the origin with respect to the positive x -axis in counterclockwise direction." A planar PLP can be mapped from a PPP in a 2D plane where x is in range $[0, 2\pi]$ and y is in range $[0, \infty)$. The investigation of PLP starts since 1940s by S. Goudsmit (Goudsmit, 1945). The following research includes topics about PLP hitting a convex region and polygons created by PLP by Miller (Miles, 1964). Now, Poisson line processes have been applied in various fields, "including material science, geology, image processing, transportation, localization, and wired and wireless communications" (Dhillon & Chetlur, 2020).

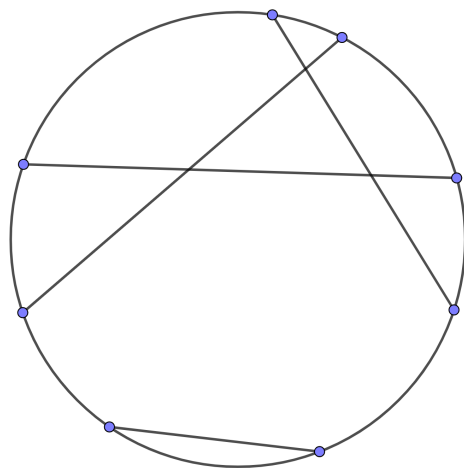


Figure 2: 4 chords in the circle create 8 regions

1.2 Computer simulation

Let's do a computer simulation to see what does the distribution of the number of regions look like when n i.i.d. chords are drawn from the above model for large n . Consider the probability density function

$$f(\theta) = \begin{cases} \sin(\theta), & \text{if } 0 \leq \theta \leq \frac{\pi}{2} \\ 0, & \text{else.} \end{cases}$$

Again, we will assume the left endpoint is uniformly distributed on the circumference of the circle. We draw 100 i.i.d chords from $f(\theta)$ in the circle and count the number of regions. The process is repeated 1000 times and we plot the result into a histogram, which is shown in Figure 3.

The histogram appears to be a normal distribution. To test this hypothesis, we do a Kolmogorov–Smirnov test ([Massey Jr, 1951](#)) on the distribution. To find the expected value and standard deviation of the regions, we need to borrow the results from equation (12) and (13) in the following sections. In this case, the expected value is 2576 and the standard deviation is 144.3. After normalizing the distribution and putting it into the (KS) test, we get statistic = 0.03 and p-value = 0.29, which is larger than 0.05. Therefore, the null hypothesis cannot be rejected. The cumulative distributions of the number of regions and normal distribution are shown in Figure 3.

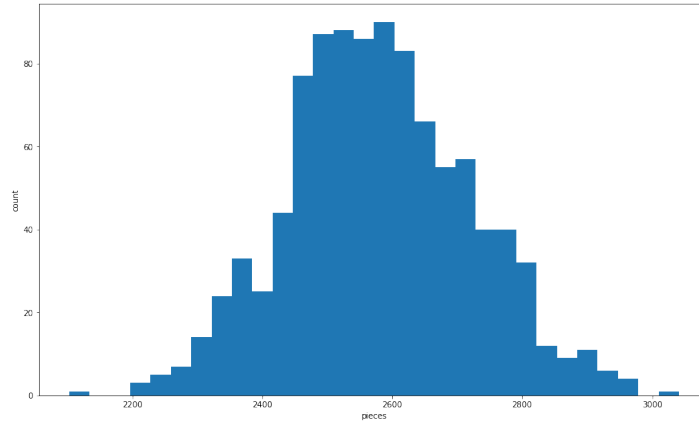


Figure 3: A histogram of 1000 simulations of the number of regions created when 100 random chords are drawn according to the distance distribution $f(\theta) = \sin \theta$ and left endpoint uniformly distributed on the circumference of the circle.

To strengthen this hypothesis, we do simulation and test on various numbers of random chords, and the result is showing in the following table.

	number of chords	repetition	statistic	p-value
0	10.0	100.0	0.09	0.36
1	20.0	100.0	0.12	0.08
2	40.0	100.0	0.07	0.74
3	80.0	100.0	0.07	0.73
4	160.0	100.0	0.11	0.14

We can see all the simulations pass the test. Now we prove that the distribution converges to normal distribution using Stein's method for Central Limit Theorem. Here is the main result of this paper. It will be repeated and explained in the following sections.

For a function $f : \mathbb{R} \rightarrow \mathbb{R}$, denote

$$\|f\| = \sup_x |f(x)|.$$

Suppose h is a differential function that $\|h\|$ and $\|h'\|$ are finite. Then we denote Φh to be $\mathbb{E}[h(Z)]$ and $\Phi(w)$ to be $P(Z < w)$ where Z is a standard normal random variable. We also define notation Θ such that $a_n \sim \Theta(n^i)$ if there exist $k_2 > k_1 > 0$ and n_0 , and for all $n > n_0$, $k_1 \cdot n^i \leq a_n \leq k_2 \cdot n^i$.

Theorem 1.1. Suppose F_n is the number of regions after drawing n random chords. Then, its expectation, $\mathbb{E}[F_n]$ is $\sim \Theta(n^2)$ and its standard deviation, $\sigma_n \sim \Theta(n^{3/2})$. Moreover, for

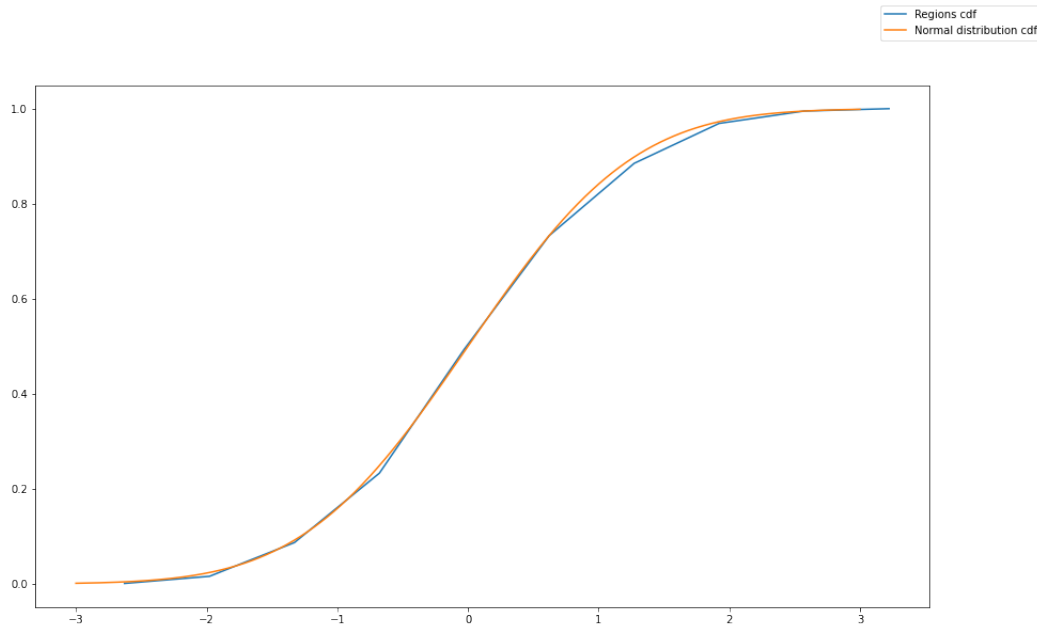


Figure 4: The comparison of the cumulative distribution between the regions and normal distribution.

$n > 5$,

$$\sup_{\|h\| \leq 1, \|h'\| \leq 1} \left(\left| \mathbb{E} \left[h \left(\frac{F_n - \mathbb{E}[F_n]}{\sigma_n} \right) \right] - \Phi(h) \right| \right) \leq \frac{6n^{5/2}}{\sigma_n^2} + \frac{2n^4}{\sigma_n^3} \sim \Theta(n^{-1/2})$$

and

$$\sup_w \left(\left| \mathbb{P} \left(\frac{F_n - \mathbb{E}[F_n]}{\sigma_n} < w \right) - \Phi(w) \right| \right) \leq 14 \frac{n^4}{\sigma_n^3} \sim \Theta(n^{-1/2}).$$

$\mathbb{E}[F_n]$ and σ_n will be calculated in section 2, where Theorem 1.1 is proved in two parts.

2 Proving Central Limit Theorem by Stein's method

In the following proof, we will first express the number of regions in terms of the number of "chord i crosses chord j " events. These events satisfy an "only locally dependent" property. We can then apply known theorems which implement the underlying Stein's method in order to give explicit general bounds for Normal approximation for such "only locally dependent" events. The contribution of this paper is to evaluate these general bounds in the specific context of our random chords model.

Suppose we draw n i.i.d. random chords (from $f(\theta)$ and left endpoint uniformly distributed on the circumference of the circle) in the circle. Let R_n be the total number of intersections after drawing n i.i.d random chords. Let A_{ij} be the indicator function that chords i and j cross each other

$$A_{ij} = \begin{cases} 1, & \text{if } i \text{ and } j \text{ intersect} \\ 0, & \text{else.} \end{cases}$$

Then according to (Solomon, 1978) page 134-137,

$$\mathbb{E}[R_n] = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \mathbb{E}[A_{ij}] = \binom{n}{2} \mathbb{E}[A_{12}], \quad (1)$$

and

$$\mathbb{E}[R_n^2] = \binom{n}{2} \mathbb{E}[A_{12}] + 6 \binom{n}{3} \mathbb{E}[A_{12}A_{13}] + 6 \binom{n}{4} (\mathbb{E}[A_{12}])^2. \quad (2)$$

As for $\mathbb{E}[A_{12}]$ and $\mathbb{E}[A_{12}A_{13}]$ in (1) and (2), we can bring $i = 1, j = 2, k = 3$ into the following equations (Solomon, 1978):

$$\mathbb{E}[A_{ij}] = \frac{4}{\pi} \int_0^{\frac{\pi}{2}} \int_0^{\theta_i} \theta_j f(\theta_j) f(\theta_i) d\theta_j d\theta_i, \quad (3)$$

and

$$\mathbb{E}[A_{ij}A_{ik}] = \frac{4}{\pi^2} \int_0^{\frac{\pi}{2}} \left[\int_0^{\theta_i} \theta_j f(\theta_j) d\theta_j + \theta_i \int_{\theta_i}^{\frac{\pi}{2}} f(\theta_j) d\theta_j \right]^2 f(\theta_i) d\theta_i. \quad (4)$$

With a closer observation of (3) and (4), we find the Proposition 2.1, which will come useful in the following calculation.

Proposition 2.1. $(\mathbb{E}[A_{ij}])^2 < \mathbb{E}[A_{ij}A_{ik}]$ regardless of our choice of probability distribution function f if θ is a continuous random variable. The inequality also implies A_{ij} and A_{ik} are dependent.

Proof. By symmetry,

$$\begin{aligned} 2\mathbb{E}[A_{ij}] &= \frac{4}{\pi} \int_0^{\frac{\pi}{2}} \int_0^{\theta_i} \theta_j f(\theta_j) f(\theta_i) d\theta_j d\theta_i + \frac{4}{\pi} \int_0^{\frac{\pi}{2}} \int_0^{\theta_j} \theta_i f(\theta_i) f(\theta_j) d\theta_i d\theta_j \\ &= \frac{4}{\pi} \int_0^{\frac{\pi}{2}} \left[\int_0^{\theta_i} \theta_j f(\theta_j) d\theta_j + \theta_i \int_{\theta_i}^{\frac{\pi}{2}} f(\theta_j) d\theta_j \right] f(\theta_i) d\theta_i. \end{aligned}$$

Suppose θ is a random variable with density f and $g(\theta) = \int_0^\theta \theta_j f(\theta_j) d\theta_j + \theta \int_\theta^{\pi/2} f(\theta_j) d\theta_j$. Thus,

$$(\mathbb{E}[A_{ij}])^2 = \frac{4}{\pi^2} (\mathbb{E}[g(\theta)])^2$$

and

$$\mathbb{E}[A_{ij}A_{ik}] = \frac{4}{\pi^2} \mathbb{E}[g(\theta)^2].$$

Therefore, it is always the case that

$$\frac{4}{\pi^2} \text{Var}(g(\theta)) = \mathbb{E}[A_{ij}A_{ik}] - (\mathbb{E}[A_{ij}])^2 \geq 0$$

and

$$\mathbb{E}[A_{ij}A_{ik}] - (\mathbb{E}[A_{ij}])^2 = 0$$

if and only if $g(\theta)$ is a constant. Suppose $F(\theta)$ is the cumulative distribution of θ and $F(0) = 0$, $F(\frac{\pi}{2}) = 1$. Since $f(\theta)$ is the density function, by integrating by parts,, we get

$$\begin{aligned} g(\theta) &= \theta_j F(\theta_j) \Big|_0^\theta - \int_0^\theta F(\theta_j) d\theta_j + \theta(1 - F(\theta)) \\ &= \theta - \int_0^\theta F(\theta_j) d\theta_j \end{aligned}$$

Suppose $g(\theta)$ is a non-random constant c . Then, $\int_0^\theta 1 - F(\theta_j) d\theta_j = c$. The only solution for $F(\theta_j)$ would be $F(\theta_j) = 1$, Lebesgue almost everywhere on $[0, \pi/2]$. Since F is continuous as it has a density f , the previous statement implies that $F(\theta_j) = 1$ everywhere. However, this contradicts the fact that $F(0)$ must be zero. Thus a contradiction arises and $g(\theta)$ cannot be a constant random variable if $f(\theta)$ is a continuous function. As a result, $(\mathbb{E}[A_{ij}])^2 < \mathbb{E}[A_{ij}A_{ik}]$ is $f(\theta)$ is a density distribution. Since $\text{Cov}(A_{ij}A_{ik}) = \mathbb{E}[A_{ij}A_{ik}] - (\mathbb{E}[A_{ij}])^2 > 0$, A_{ij} and A_{ik} are dependent. \square

When we draw n chords on the circle we can create a graph whose vertices are any points of intersections between two chords or an intersection of the chord and the circle. Then Euler's characteristic ([Euler, 1758](#)) for planar graphs tells us that

$$V_n - E_n + F_n = 2,$$

where after drawing n chords, V_n means the number of vertices; E_n means the number of edges; F_n means the number of regions. Since we only care about the regions inside the circle, we need to exclude the one region that is outside the circle. Therefore, we

redefine F_n to be the number of regions inside the circle and an updated formula would be

$$V_n - E_n + F_n = 1. \quad (5)$$

For example, in Figure 2, $V_n = 11$, $E_n = 18$, $F_n = 8$. Now our target is to find the number of regions. Suppose n i.i.d. random chords are drawn and R_n intersections are created by chords only. Each vertex of the above graph corresponds to the intersection of two chords, or to an intersection of a chord with the circle. By assumption there are R_n vertices of the first type. Since each chord intersects the circle twice, and the probability for two random chords to intersect the circle at the same point is zero as the left endpoint has a continuous distribution on the circumference of the circle, there are $2n$ vertices of the latter type. Therefore,

$$V_n = R_n + 2n. \quad (6)$$

As for edges, the degree of the vertices created by the intersection of two chords is 4. The degree of the vertices created by the intersection one chord and the circle is 3. Every edge is adjacent to two vertices. Therefore,

$$E_n = \frac{4 \cdot R_n + 3 \cdot 2n}{2} = 2R_n + 3n. \quad (7)$$

Thus, bringing (6) and (7) to (5), we get

$$F_n = R_n + n + 1.$$

Recall by definition of R_n and A_{ij} , $R_n = \sum_{i=1}^{n-1} \sum_{j=i+1}^n A_{ij}$. Thus, after expanding R_n , the number of regions can be written as

$$F_n = \sum_{i=1}^{n-1} \sum_{j=i+1}^n A_{ij} + n + 1. \quad (8)$$

With an easy observation of number of regions (F_n), we find that it is the summation of dependent Bernoulli random variables (A_{ij}). We first calculate its expected value and standard deviation. From (8), we get

$$\mathbb{E}[F_n] = \frac{1}{2}n(n-1)\mathbb{E}[A_{12}] + n + 1 \sim \Theta(n^2). \quad (9)$$

From (1), (2) and Proposition 2.1, we get

$$\begin{aligned} \text{Var}(R_n) &= \mathbb{E}[R_n^2] - (\mathbb{E}[R_n])^2 \\ &= \frac{1}{4} [2n(n-1)\mathbb{E}[A_{12}] + 4n(n-1)(n-2)\mathbb{E}[A_{12}A_{13}] - n(n-1)(4n-6)(\mathbb{E}[A_{12}])^2] \sim \Theta(n^3). \end{aligned} \quad (10)$$

Thus,

$$\begin{aligned}\sigma_n &= \sqrt{\text{Var}(F_n)} = \sqrt{\text{Var}(F_n - n - 1)} = \sqrt{\text{Var}(R_n)} \\ &= \frac{1}{2} \left[2n(n-1)\mathbb{E}[A_{12}] + 4n(n-1)(n-2)\mathbb{E}[A_{12}A_{13}] - n(n-1)(4n-6)(\mathbb{E}[A_{12}]^2) \right]^{\frac{1}{2}} \sim \Theta(n^{3/2}).\end{aligned}\quad (11)$$

Before we get any deeper, recall in the simulation part we calculate the expectation and standard deviation of F_n when $f(\theta) = \sin(\theta)$ and $n = 100$. Let's do the calculation here. According to (3) and (4), $\mathbb{E}[A_{12}] = \frac{1}{2}$ and $\mathbb{E}[A_{12}A_{13}] = \frac{8}{3\pi^2}$. Then, we bring them into (9) and (11) and get

$$\mathbb{E}[F_n] = \frac{100 \cdot 99}{2} \cdot \frac{1}{2} + 100 + 1 = 2576 \quad (12)$$

$$\sigma_n = \frac{1}{2} \left[2 \cdot 100 \cdot 99 \cdot \frac{1}{2} + 4 \cdot 100 \cdot 99 \cdot 98 \cdot \frac{8}{3\pi^2} - 100 \cdot 99 \cdot 394 \cdot \frac{1}{4} \right]^{\frac{1}{2}} = 144.3 \quad (13)$$

With the expected value and standard deviation in (9) and (11), we can shift and scale the distribution to make it have a mean of 0 and a variance of 1 as mentioned in the introduction. Therefore, we would like to prove after shifting and scaling, the distribution would converges to a standard normal distribution.

To prove F_n converges to normal distribution, we would want to show for all function h such that $\|h\|$ and $\|h'\|$ are finite,

$$\left| \mathbb{E} \left[h \left(\frac{F_n - \mathbb{E}(F_n)}{\sigma_n} \right) \right] - \Phi h \right|$$

is small when n is large. Stein's method is first developed by Stein in his paper (Stein, 1972) and explained in detail in (Ross, 2011). We will apply the method to our problem by borrowing the following Lemma 2.2 from (Rinott & Rotar, 2000) on page 21 Theorem 2.1.

Lemma 2.1. *Suppose Y_1, Y_2, \dots, Y_n are random variables whose expected value is 0 and $\text{Var}[\sum_{i=1}^n Y_i] = \sigma$. Let M_i denote the dependent neighborhood of Y_i , which means Y_i and Y_j are dependent if and only if $j \in M_i$. Then*

$$\left| \mathbb{E} \left[h \left(\frac{\sum_{i=1}^n Y_i}{\sigma} \right) \right] - \Phi h \right| \leq \frac{2}{\sigma^2} \|h\| \sqrt{\mathbb{E} \left[\left(\sum_{i=1}^n \sum_{j \in M_i} (Y_i Y_j - \mathbb{E}[Y_i Y_j]) \right)^2 \right]} + \frac{1}{\sigma^3} \|h'\| \mathbb{E} \left[\sum_{i=1}^n |Y_i| \left(\sum_{j \in M_i} Y_j \right)^2 \right]. \quad (14)$$

Now, we can prove the following theorem for our problem.

Theorem 2.2. For $n > 5$,

$$\left| \mathbb{E} \left[h \left(\frac{F_n - \mathbb{E}(F_n)}{\sigma_n} \right) \right] - \Phi h \right| \leq \frac{6n^{5/2}}{\sigma_n^2} \|h\| + \frac{2n^4}{\sigma_n^3} \|h'\| \sim \Theta(n^{-1/2})$$

Proof. To apply (14) on our problem, we first need to reduce the expected value of random variables to 0. Suppose,

$$Y_{i,j} = A_{ij} - \mathbb{E}[A_{ij}].$$

We denote $M_{\{i,j\}}$ to be the dependent neighborhood for $Y_{i,j}$. According to (Solomon, 1978) page 135, A_{ij} and A_{kl} are independent if $\{i, j\} \cap \{k, l\} = \emptyset$. And according to Proposition 2.1, A_{ij} and A_{kl} are dependent if $\{i, j\} \cap \{k, l\} \neq \emptyset$. We have

$$\{k, l\} \notin M_{\{i,j\}} \text{ if and only if } \{i, j\} \cap \{k, l\} = \emptyset. \quad (15)$$

According to (8) and (14), we get

$$\begin{aligned} \left| \mathbb{E} \left[h \left(\frac{F_n - \mathbb{E}(F_n)}{\sigma_n} \right) \right] - \Phi h \right| &= \left| \mathbb{E} \left[h \left(\frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^n Y_{i,j}}{\sigma_n} \right) \right] - \Phi h \right| \\ &\leq \frac{2}{\sigma_n^2} \|h\| \sqrt{\mathbb{E} \left[\left(\sum_{i=1}^{n-1} \sum_{j=i+1}^n \sum_{\{k,l\} \in M_{\{i,j\}}} (Y_{i,j} Y_{k,l} - \mathbb{E}[Y_{i,j} Y_{k,l}]) \right)^2 \right]} \\ &\quad + \frac{1}{\sigma_n^3} \|h'\| \mathbb{E} \left[\sum_{i=1}^{n-1} \sum_{j=i+1}^n |Y_{i,j}| \left(\sum_{\{k,l\} \in M_{\{i,j\}}} Y_{k,l} \right)^2 \right]. \end{aligned} \quad (16)$$

Therefore, our next step is to bound the right hand side of (16). We start with the first part,

$$\begin{aligned} &\mathbb{E} \left[\left(\sum_{i=1}^{n-1} \sum_{j=i+1}^n \sum_{\{k,l\} \in M_{\{i,j\}}} Y_{i,j} Y_{k,l} - \mathbb{E}[Y_{i,j} Y_{k,l}] \right)^2 \right] \\ &= \mathbb{E} \left[\sum_{i=1}^{n-1} \sum_{j=i+1}^n \sum_{\{k,l\} \in M_{\{i,j\}}} \sum_{i'=1}^{n-1} \sum_{j'=i+1}^n \sum_{\{k',l'\} \in M_{\{i',j'\}}} (Y_{i,j} Y_{k,l} - \mathbb{E}[Y_{i,j} Y_{k,l}]) (Y_{i',j'} Y_{k',l'} - \mathbb{E}[Y_{i',j'} Y_{k',l'}]) \right] \end{aligned}$$

The number of terms in $\sum_{i'=1}^{n-1} \sum_{j'=i+1}^n \sum_{\{k',l'\} \in M_{\{i',j'\}}}$ that $|\{i', j'\} \cap \{k', l'\}| = 1$ is $\frac{n(n-1)}{2} \cdot 2(n-2)$. And the number of elements that $|\{i', j'\} \cap \{k', l'\}| = 2$ is $\frac{n(n-1)}{2}$. When $|\{i', j'\} \cap \{k', l'\}| = 0$, $[k', l'] \notin M_{\{i', j'\}}$ and we do not need to consider such case because of (15). By

symmetry we can assume $Y_{i',j'}Y_{k',l'}$ to be $Y_{1,2}Y_{1,3}$ when $|\{i', j'\} \cap \{k', l'\}| = 1$ and $Y_{i',j'}Y_{k',l'}$ to be $Y_{1,2}^2$ when $|\{i', j'\} \cap \{k', l'\}| = 2$. Thus,

$$\begin{aligned} & \mathbb{E} \left[\left(\sum_{i=1}^{n-1} \sum_{j=i+1}^n \sum_{\{k,l\} \in \mathcal{M}_{\{i,j\}}} Y_{i,j}Y_{k,l} - \mathbb{E}[Y_{i,j}Y_{k,l}] \right)^2 \right] \\ &= \frac{n(n-1)}{2} \cdot 2(n-2) \mathbb{E} \left[\sum_{i=1}^{n-1} \sum_{j=i+1}^n \sum_{\{k,l\} \in \mathcal{M}_{\{i,j\}}} (Y_{i,j}Y_{k,l} - \mathbb{E}[Y_{i,j}Y_{k,l}])(Y_{1,2}Y_{1,3} - \mathbb{E}[Y_{1,2}Y_{1,3}]) \right] \\ &+ \frac{n(n-1)}{2} \cdot \mathbb{E} \left[\sum_{i=1}^{n-1} \sum_{j=i+1}^n \sum_{\{k,l\} \in \mathcal{M}_{\{i,j\}}} (Y_{i,j}Y_{k,l} - \mathbb{E}[Y_{i,j}Y_{k,l}])(Y_{1,2}^2 - \mathbb{E}[Y_{1,2}^2]) \right]. \end{aligned}$$

Note that $Y_{i,j}Y_{k,l} - \mathbb{E}[Y_{i,j}Y_{k,l}]$ and $Y_{i',j'}Y_{k',l'} - \mathbb{E}[Y_{i',j'}Y_{k',l'}]$ are independent if $\{i, j, k, l\} \cap \{i', j', k', l'\} = \emptyset$ as every chord is independent of each other. In such case, we can break the multiplication inside $\mathbb{E}[\cdot]$ apart and get

$$\begin{aligned} \mathbb{E}[(Y_{i,j}Y_{k,l} - \mathbb{E}[Y_{i,j}Y_{k,l}])(Y_{1,2}Y_{1,3} - \mathbb{E}[Y_{1,2}Y_{1,3}])] &= \mathbb{E}[Y_{i,j}Y_{k,l} - \mathbb{E}[Y_{i,j}Y_{k,l}]] \cdot \mathbb{E}[Y_{1,2}Y_{1,3} - \mathbb{E}[Y_{1,2}Y_{1,3}]] \\ &= 0 \text{ if } \{i, j, k, l\} \cap \{1, 2, 3\} = \emptyset. \end{aligned}$$

$$\begin{aligned} \mathbb{E}[(Y_{i,j}Y_{k,l} - \mathbb{E}[Y_{i,j}Y_{k,l}])(Y_{1,2}^2 - \mathbb{E}[Y_{1,2}^2])] &= \mathbb{E}[Y_{i,j}Y_{k,l} - \mathbb{E}[Y_{i,j}Y_{k,l}]] \cdot \mathbb{E}[Y_{1,2}^2 - \mathbb{E}[Y_{1,2}^2]] \\ &= 0 \text{ if } \{i, j, k, l\} \cap \{1, 2\} = \emptyset. \end{aligned}$$

Thus, let $I(\cdot)$ be the indicator function of event \cdot ,

$$\begin{aligned} & \mathbb{E} \left[\left(\sum_{i=1}^{n-1} \sum_{j=i+1}^n \sum_{\{k,l\} \in \mathcal{M}_{\{i,j\}}} Y_{i,j}Y_{k,l} - \mathbb{E}[Y_{i,j}Y_{k,l}] \right)^2 \right] \\ &\leq \frac{n(n-1)}{2} \cdot 2(n-2) \mathbb{E} \left[\sum_{i=1}^{n-1} \sum_{j=i+1}^n \sum_{\{k,l\} \in \mathcal{M}_{\{i,j\}}} I(\{i, j, k, l\} \cap \{1, 2, 3\} \neq \emptyset) \right] \\ &+ \frac{n(n-1)}{2} \cdot \mathbb{E} \left[\sum_{i=1}^{n-1} \sum_{j=i+1}^n \sum_{\{k,l\} \in \mathcal{M}_{\{i,j\}}} I(\{i, j, k, l\} \cap \{1, 2\} \neq \emptyset) \right] \\ &= n(n-1)(n-2) \cdot (9n^2 - 24n) + \frac{n(n-1)}{2} \cdot (6n^2 - 10n - 3) \\ &= 9n^5 - 48n^4 + 82n^3 - \frac{89n^2}{2} + \frac{3n}{2}. \end{aligned}$$

Since

$$\begin{aligned} -48n^4 + 82n^3 - \frac{89n^2}{2} + \frac{3n}{2} &= -(48n^4 - 82n^3 + 15n^2) - \left(\frac{59n^2}{2} - \frac{3n}{2}\right) \\ &= -n^2(2n-3)(24n-5) - \frac{n}{2}(59n-3) \\ &< 0 \text{ if } n > 5, \end{aligned}$$

we have

$$\mathbb{E} \left[\left(\sum_{i=1}^{n-1} \sum_{j=i+1}^n \sum_{\{k,l\} \in \mathcal{M}_{\{i,j\}}} Y_{i,j} Y_{k,l} - \mathbb{E}[Y_{i,j} Y_{k,l}] \right)^2 \right] \leq 9n^5 \text{ for all } n > 5.$$

Bring it back to (16). For all $n > 5$,

$$\begin{aligned} \frac{2}{\sigma_n^2} \|h\| \sqrt{\mathbb{E} \left[\left(\sum_{i=1}^{n-1} \sum_{j=i+1}^n \sum_{\{k,l\} \in \mathcal{M}_{\{i,j\}}} (Y_{i,j} Y_{k,l} - \mathbb{E}[Y_{i,j} Y_{k,l}]) \right)^2 \right]} &\leq \frac{2}{\sigma_n^2} \|h\| \sqrt{9n^5} \\ &= \frac{6n^{5/2}}{\sigma_n^2} \|h\|. \end{aligned}$$

As for the second part of (16),

$$\begin{aligned} \frac{1}{\sigma_n^3} \|h'\| \mathbb{E} \left\{ \sum_{i=1}^{n-1} \sum_{j=i+1}^n |Y_{i,j}| \left(\sum_{\{k,l\} \in \mathcal{M}_{\{i,j\}}} Y_{k,l} \right)^2 \right\} &\leq \frac{1}{\sigma_n^3} \|h'\| \mathbb{E} \left\{ \sum_{i=1}^{n-1} \sum_{j=i+1}^n 1 \left(\sum_{\{k,l\} \in \mathcal{M}_{\{i,j\}}} 1 \right)^2 \right\} \\ &= \frac{1}{\sigma_n^3} \|h'\| \left(\frac{n(n-1)}{2} (2n-3)^2 \right) \\ &= \frac{1}{\sigma_n^3} \|h'\| \left(2n^4 - 8n^3 + \frac{21n^2}{2} - \frac{9n}{2} \right). \end{aligned}$$

Since $-8n^3 + \frac{21n^2}{2} - \frac{9n}{2} = -\frac{n}{2}(16n-5)(n-1) - 2n < 0$ for all $n > 5$,

$$\frac{1}{\sigma_n^3} \|h'\| \mathbb{E} \left\{ \sum_{i=1}^{n-1} \sum_{j=i+1}^n |Y_{i,j}| \left(\sum_{\{k,l\} \in \mathcal{M}_{\{i,j\}}} Y_{k,l} \right)^2 \right\} \leq \frac{2n^4}{\sigma_n^3} \|h'\|.$$

Since $\sigma_n \sim \Theta(n^{3/2})$, from (16) we get

$$\left| \mathbb{E} \left[h \left(\frac{F_n - \mathbb{E}(F_n)}{\sigma_n} \right) \right] - \Phi h \right| \leq \frac{6n^{5/2}}{\sigma_n^2} \|h\| + \frac{2n^4}{\sigma_n^3} \|h'\| \sim \Theta(n^{-1/2}).$$

□

Another way to bound the difference between the distribution of F_n and the normal distribution is to measure the difference of c.d.f between them. In other word,

$$\left| \mathbb{P} \left(\frac{F_n - \mathbb{E}[F_n]}{\sigma_n} < w \right) - \Phi(w) \right| \tag{17}$$

needs to be small when n is large. To bound this term, we need to use the following Lemma 2.4, which restates Theorem 2.2 from (Rinott & Rotar, 2000). The detailed proof can be found in (Dembo & Rinott, 1996).

Lemma 2.2. Let Y_1, \dots, Y_n be random variables satisfying $|Y_i - \mathbb{E}[Y_i]| \leq B$ a.s., $i = 1, \dots, n$, $\mathbb{E}[\sum_{i=1}^n Y_i] = \lambda$, $\text{Var}[\sum_{i=1}^n Y_i] = \sigma^2 > 0$ and $\frac{1}{n}\mathbb{E}[\sum_{i=1}^n |Y_i - \mathbb{E}[Y_i]|] = \mu$. Let $M_i \subset \{1, \dots, n\}$ be such that $j \in M_i$ if and only if $i \in M_j$ and (Y_i, Y_j) is independent of $\{Y_k\}_{k \notin M_i \cup M_j}$ for $i, j = 1, \dots, n$, and set $D = \max_{1 \leq i \leq n} |M_i|$. Then

$$\left| \mathbb{P}\left(\frac{\sum_{i=1}^n Y_i - \lambda}{\sigma} \leq w\right) - \Phi(w) \right| \leq 7 \frac{n\mu}{\sigma^3} (DB)^2.$$

The definition of M_i might be ambiguous in lemma 2.4, but it basically means the set of random variables in Y_1, \dots, Y_n that is dependent of Y_i . With the help of this inequality, we can bound (17) in the following theorem.

Theorem 2.3.

$$\left| \mathbb{P}\left(\frac{F_n - \mathbb{E}[F_n]}{\sigma_n} < w\right) - \Phi(w) \right| \leq 14 \frac{n^4}{\sigma_n^3} \sim \Theta(n^{-1/2})$$

Proof. For our problem, after eliminating the constant, the target random variable is $\sum_{i=1}^{n-1} \sum_{j=i+1}^n A_{ij}$ according to (8). $\mathbb{E}[F_n]$ can be found in (9). The standard deviation $\sigma = \sigma_n$ according to (11). The total number of random variables is $\binom{n}{2} = \frac{n(n-1)}{2}$. Since A_{ij} is Bernoulli random variable, $|A_{ij} - \mathbb{E}[A_{ij}]| \leq 1$. Therefore, $B \leq 1$ and

$$\begin{aligned} \mu &= \frac{2}{n(n-1)} \mathbb{E} \left[\sum_{i=1}^{n-1} \sum_{j=i+1}^n |A_{ij} - \mathbb{E}[A_{ij}]| \right] \\ &\leq \frac{2}{n(n-1)} \mathbb{E} \left[\sum_{i=1}^{n-1} \sum_{j=i+1}^n 1 \right] = 1. \end{aligned}$$

A_{ij} is dependent on A_{kl} if and only if $\{|i, j\} \cap \{k, l\} = 1$ or 2 , so $D = 2n - 3$. Finally, we can bring everything into Lemma 2.4.

$$\begin{aligned} \left| \mathbb{P}\left(\frac{F_n - \mathbb{E}[F_n]}{\sigma_n} < \epsilon\right) - \Phi(\epsilon) \right| &\leq 7 \frac{\frac{n(n-1)}{2} \cdot 1}{\sigma_n^3} ((2n-3) \cdot 1)^2 \\ &\leq 7 \frac{n^2}{2\sigma_n^3} (2n)^2 \\ &= 14 \frac{n^4}{\sigma_n^3}. \end{aligned}$$

Since $\sigma_n \sim \Theta(n^{3/2})$,

$$\left| \mathbb{P}\left(\frac{F_n - \mathbb{E}[F_n]}{\sigma_n} < \epsilon\right) - \Phi(\epsilon) \right| \leq 14 \frac{n^4}{\sigma_n^3} \sim \Theta(n^{-1/2}).$$

□

Note that the random chords for our problem have a uniformly distributed left endpoint on the circumference of the circle. The reason for this condition is that there is no convenient formula to calculate $\mathbb{E}[A_{ij}]$ and $\mathbb{E}[A_{ij}A_{ik}]$ if the distribution of left endpoint is nonuniform or dependent of distance distribution $f(\theta)$. Thus, we cannot calculate the expectation and the variance of F_n . However, we can still prove the following corollary using the same reasoning.

Theorem 2.4. Suppose there are n i.i.d. random chords in the circle. For every chord, the left endpoint may not uniformly distributed on the circumference of the circle and may be dependent on the half angle distribution f . However, each chord is constructed in a way such that the probability for three or more points to create an intersection at the same point or two or more points create an intersection with the circle at the same point is zero. Suppose $N = 7 \frac{n\mu}{\sigma_n^2} (DB)^2$ as in the RHS of the formula in Lemma 2.4. If $0 < (\mathbb{E}[A_{12}])^2 < \mathbb{E}[A_{12}A_{13}] < 1$,

$$\left| \mathbb{P} \left(\frac{F_n - \mathbb{E}[F_n]}{\sigma_n} < w \right) - \Phi(w) \right| \leq N \sim \Theta(n^{-1/2}).$$

If $0 < (\mathbb{E}[A_{12}])^2 = \mathbb{E}[A_{12}A_{13}] < 1$,

$$\left| \mathbb{P} \left(\frac{F_n - \mathbb{E}[F_n]}{\sigma_n} < w \right) - \Phi(w) \right| \leq N \sim \Theta(n^{-1}).$$

Proof. Since there is no overlapping point, the reasoning about Euler's characteristic still holds and

$$F_n = \sum_{i=1}^{n-1} \sum_{j=i+1}^n A_{ij} + n + 1.$$

F_n is still the sum of Bernoulli random variables, so our next step is to bring everything into Lemma 2.4. We first consider the case where $0 < (\mathbb{E}[A_{12}])^2 < \mathbb{E}[A_{12}A_{13}] < 1$. According to (11), $\sigma_n \sim \Theta(n^{3/2})$. By only considering the order, we have $\mu \sim \Theta(1)$, $B \sim \Theta(1)$. Since $(\mathbb{E}[A_{12}])^2 < \mathbb{E}[A_{12}A_{13}]$, A_{ij} is dependent of A_{ik} and $D \sim \Theta(n^1)$. Finally, the total number of random variables is $\frac{n(n-1)}{2} \sim \Theta(n^2)$. Thus,

$$\left| \mathbb{P} \left(\frac{F_n - \mathbb{E}[F_n]}{\sigma_n} < w \right) - \Phi(w) \right| \leq N \sim \frac{\Theta(n^2) \cdot \Theta(n^1)^2}{\Theta(n^{3/2})^3} \sim \Theta(n^{-1/2}).$$

When $0 < (\mathbb{E}[A_{12}])^2 = \mathbb{E}[A_{12}A_{13}] < 1$, according to (11), $\sigma_n \sim \Theta(n^1)$. Again, $\mu \sim \Theta(1)$, $B \sim \Theta(1)$, and the total number of random variables is $\frac{n(n-1)}{2} \sim \Theta(n^2)$. Since $(\mathbb{E}[A_{12}])^2 = \mathbb{E}[A_{12}A_{13}]$, A_{ij} is independent of A_{ik} and $D \sim \Theta(1)$. Thus,

$$\left| \mathbb{P} \left(\frac{F_n - \mathbb{E}[F_n]}{\sigma_n} < w \right) - \Phi(w) \right| \leq N \sim \frac{\Theta(n^2) \cdot \Theta(1)^2}{\Theta(n^1)^3} \sim \Theta(n^{-1}).$$

□

A natural question one would ask from Theorem 2.6 is what will happen if $(\mathbb{E}[A_{12}])^2 > \mathbb{E}[A_{12}A_{13}]$ as Proposition 2.1 is no longer useful. The case is that when we bring such situation into (10), we find $\text{Var}(R_n) < 0$ as n goes to infinity. Therefore, $(\mathbb{E}[A_{12}])^2 \leq \mathbb{E}[A_{12}A_{13}]$ at all times.

3 Acknowledgments

This paper would not be possible without the help from my mentor Professor Soumik Pal, who has provided the main idea of Stein's method for the problem and commented on every sentence of this paper. I am grateful to Professor David Aldous for suggesting references and giving comments for this paper. I am thankful to my friend Ami Oka for providing help on the computer simulation. I also thank an anonymous referee for a careful review of this work and for providing several helpful comments.

References

- Bertrand, J. (1889). *Calcul des probabilités*. Gauthier-Villars.
- Buffon, G. (1777). Essai d'arithmétique morale. *Supplémenta l'Histoire naturelle*, 4, 1777.
- David, F., & Fix, E. (1964). Intersections of random chords of a circle. *Biometrika*, 51(3/4), 373–379.
- Dembo, A., & Rinott, Y. (1996). Some examples of normal approximations by Stein's method. In *Random discrete structures* (pp. 25–44). Springer.
- Dhillon, H. S., & Chetlur, V. V. (2020). Poisson line Cox process: Foundations and applications to vehicular networks. *Synthesis Lectures on Learning, Networks, and Algorithms*, 1(1), 1–149.
- Euler, L. (1758). Elementa doctrinae solidorum. *Novi commentarii academiae scientiarum Petropolitanae*, 109–140.
- Goudsmit, S. (1945). Random distribution of lines in a plane. *Reviews of Modern Physics*, 17(2-3), 321.
- Kalousová, A. (2011). The origins of the geometric probability in England. In *WDS* (Vol. 8, pp. 7–12).
- Massey Jr, F. J. (1951). The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American statistical Association*, 46(253), 68–78.
- Miles, R. E. (1964). Random polygons determined by random lines in a plane, ii. *Proceedings of the National Academy of Sciences of the United States of America*, 52(5), 1157.
- Rinott, Y., & Rotar, V. (2000). Normal approximations by Stein's method. *Decisions in Economics and Finance*, 23(1), 15–29.
- Ross, N. (2011). Fundamentals of Stein's method. *Probability Surveys*, 8, 210–293.

Solomon, H. (1978). *Geometric probability*. SIAM.

Stein, C. (1972). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proceedings of the sixth Berkeley symposium on mathematical statistics and probability, volume 2: Probability theory* (pp. 583–602).

Shi Feng

University of Washington, Seattle

fengshi@uw.edu