FINAL DEGREE PROJECT

TFG TITLE: Monte Carlo Methods for the Rectilinear Crossing Number

DEGREE: Air-Navigation Engineering Degree

AUTHOR: Shutao Zheng

ADVISOR: Clemens Huemer

DATE: July 18, 2016
Resum

Aquesta tesi pretén trobar un algorisme Monte Carlo ràpid pel càlcul del nombre d’encreuament rectilini $cr(S)$ d’un conjunt de punts $S$ en un pla, on $cr(S)$ és el nombre d’interseccions de tots els segments rectilinis que connecten parelles de punts del conjunt.

El nombre d’encreuament rectilini és un tema central de recerca en l’àrea de la Geometria Discreta i Computacional. Es coneix un algorisme amb complexitat quadràtica per calcular $cr(S)$, però aquest necessita molt de temps si l’input és un nombre gran de punts. Llavors proposem algorismes Monte Carlo ràpids per produir solucions aproximades. Creiem que no s’han fet servir mètodes Monte Carlo en aquest context abans.

Hi ha una relació entre la precisió del nombre aproximat d’encreuament rectilini i el temps d’execució de l’algorisme Monte Carlo. Com que els outputs de l’algorisme Monte Carlo segueixen una distribució normal quan les mostres són independents, l’exactitud de l’aproximació de Monte Carlo està relacionada amb la variància. Si la variància és més petita, el nombre de mostres requerides per arribar a una precisió predefinida de la solució (amb alta probabilitat) serà més petit i l’exactitud serà millor. En aquesta tesi, hem introduït sis mètodes de Monte Carlo pel càlcul del nombre d’encreuament rectilini, i hem estudiat les variàncies i les mides de les mostres necessàries. Amb experiments computacionals hem confirmat els resultats teòrics obtinguts.

També hem aplicat algunes tècniques de reducció de variància, com ara Importance Sampling, Antithetic Variates i Control Variates, per millorar el rendiment dels mètodes Monte Carlo desenvolupats. La millor tècnica de reducció de variància en aquesta tesi és Control Variates que redueix el nombre de les mostres necessàries de manera significativa en comparació amb el mètode original.
Overview

The thesis is dedicated to find a fast Monte Carlo algorithm for the calculation of the rectilinear crossing number $Cr(S)$ of a point set $S$ in the plane, where $Cr(S)$ is the number of intersections of all the straight line segments which connect pairs of points of the set.

Crossing numbers are a central topic of research in the area of Discrete and Computational Geometry. A quadratic time algorithm to calculate $Cr(S)$ is known, which, for large input size, is very time-consuming. We propose fast Monte Carlo algorithms to produce approximate solutions. To our knowledge, Monte Carlo methods have not been applied before in this setting.

There is a trade-off between the precision of the approximated crossing number and the running time of the Monte Carlo algorithm. Since the outputs of Monte Carlo methods follow a normal distribution when the samples are independent, the exactitude of a Monte Carlo method is related with the variance. If the variance is smaller, the required sample size to reach a predefined precision of the solution (with high probability) will be smaller and the exactitude will be higher. In this thesis we introduce six Monte Carlo methods for the calculation of the rectilinear crossing number, and study their variances and the required sample sizes. Computational experiments confirm the obtained theoretical results.

Also, we apply some variance reduction techniques, such as importance sampling, antithetic variates, and control variates, to enhance the performance of the developed Monte Carlo methods. The best reduction technique in this thesis is the control variates technique which reduces the required sample size significantly compared to the original method.
I wish to express my sincere gratitude to Mr. Huemer Clemens for his guidance and encouragement in carrying out this project work. And also thank him to offer all the resources I need and resolve the questions which I obtained during this project.
## CONTENTS

Introduction ........................................................................... 1

CHAPTER 1. Concept ................................................................. 3
  1.1. Monte Carlo Method ....................................................... 3
  1.2. Rectilinear Crossing Number .......................................... 3
  1.3. Combination of Rectilinear Crossing Number and Monte Carlo Method ......................................................... 5
  1.4. Outline ........................................................................... 6

CHAPTER 2. Monte Carlo Methods in Rectilinear Crossing Number 9
  2.1. Method I — Points .......................................................... 9
    2.1.1. Base Case - One Crossing Determination ......................... 9
    2.1.2. Sample Size Determination ...................................... 10
    2.1.3. Mean and Variance ............................................... 11
  2.2. Method II — Segments .................................................... 11
    2.2.1. Mean and Variance ............................................... 14
    2.2.2. Sample Size Determination ...................................... 14
  2.3. Method III — Circles ....................................................... 15
    2.3.1. Mean and Variance ............................................... 16
    2.3.2. Sample Size Determination ...................................... 17
  2.4. Method IV — Downsizing .............................................. 18
    2.4.1. Mean and Variance ............................................... 19
  2.5. Method V — j-edges ...................................................... 19
    2.5.1. Mean and Variance ............................................... 21
  2.6. Method VI — k-InsideCircle .......................................... 22
    2.6.1. Mean and Variance ............................................... 22
    2.6.2. Sample Size Determination ...................................... 23

CHAPTER 3. Experimental Results ........................................... 25
  3.1. Method I — Points ........................................................ 25
  3.2. Method II — Segments .................................................... 28
### 3.3. Method III — Circles ................................. 28

### 3.4. Method IV — Downsizing ............................. 29

### 3.5. Method V — $j$-edges ................................. 30

### 3.6. Method VI — $k$-InsideCircle ......................... 30

### 3.7. Comparison ........................................... 31

### CHAPTER 4. Improvements ............................... 33

#### 4.1. Importance Sampling .............................. 33

#### 4.2. Antithetic Variates ................................. 34

   4.2.1. Implementation .................................. 34
   4.2.2. New Sample Size ................................. 36
   4.2.3. Experimental Data .............................. 37

#### 4.3. Control Variates .................................. 38

   4.3.1. Implementation .................................. 38
   4.3.2. New Sample Size ................................. 40
   4.3.3. Experimental Data .............................. 40

### Conclusions ............................................. 45

### Bibliography ............................................. 47

### APPENDIX A. Special Case — Method VI Improvement With Control Variates ................................. 51

### APPENDIX B. Point Sets Used for Experimental Results ............................................. 55

### APPENDIX C. Table of Normal Distribution $Q(x)$ ............................................. 59

### APPENDIX D. Variances of the Methods ................................. 61

   D.0.1. Method I — Points ................................ 61
   D.0.2. Method II — Segments ............................ 61
   D.0.3. Method III — Circles ............................. 62
   D.0.4. Method V — $j$-edges ............................ 62
   D.0.5. Method VI — $k$-InsideCircle .................... 63
   D.0.6. Method II with Antithetic Variates ............... 63
   D.0.7. Method I with Control Variates .................. 64
LIST OF FIGURES

1.1 Approximate the value of π with 30000 points situated randomly in the square [11]. ................................................................. 4
1.2 Convex quadrilateral with one crossing S (left) and non-convex (concave) quadrilateral without crossing (right) ......................................................... 4
1.3 A 100-point set with 2739011 rectilinear crossings .................................................. 5

2.1 Relation between probability error (ε_P), required sample size (K) and crossing probability (P_c) ............................................................. 12
2.2 Relation between required sample size (K) and crossing probability (P_c) when ε_P = 0.01 ..................................................................... 13
2.3 Required sample size K - crossing probability P_c of Method II when ε_P = 0.01 ............................................................... 15
2.4 Circles determined by a convex quadrilateral (left) and by a non-convex quadrilateral (right) ................................................................. 16
2.5 Required sample size K - crossing probability P_c of Method III when ε_P = 0.01 ............................................................... 18
2.6 A 3-edge leaves 3 points on one side and 4 points on the other side in a 9-point set ..................................................................... 20

3.1 Four examples of point sets used for the experiments ............................................................................. 26
3.2 Experimental variances for different point sets by extracting the mean of 5000 experiments ................................................................. 27
3.3 Percentage of ε_P ≤ 0.01 in function of the sample size K for 5000 experiments ............................................................................. 27
3.4 Normal distribution of occurrence of P_cm for 5000 experiments of a 100-point set with 1463457 crossings under different sample sizes ......................... 28
3.5 Reliability percentage for 5000 experiments of Method II ............................................................................. 29
3.6 Reliability percentage for 5000 experiments of Method III ............................................................................. 29
3.7 The variances and reliability percentages of different η with K = 9450 ............................................................................. 30
3.8 Reliability percentage for 5000 experiments of Method V ............................................................................. 31
3.9 Reliability percentage for 5000 experiments of Method VI ............................................................................. 32

4.1 Crossing probability of each point of a 100-point set from [12] ............................................................................. 34
4.2 Points A,B are on one side of segment CD and vice versa (left) whereas a non-convex quadrilateral (right) cannot meet this condition ............................................................................. 35
4.3 Comparison between \( Var(X) \) and \( Cov(X,Y) \) according to \( P_c \) ............................................................................. 36
4.4 Comparison of sample size K of Method II ............................................................................. 37
4.5 New reliability percentages of Method II by applying antithetic variates ............................................................................. 37
4.6 Reductions of variance by using control variates ............................................................................. 42
4.7 New relation between probability error (ε_P), required sample size (K) and crossing probability (P_c) of Method I by applying control variates with variance reduction R1 ............................................................................. 43
4.8 New relation between required sample size (K) and crossing probability (P_c) of Method I when ε_P = 0.01 by applying control variates (Y = (\( i \))) ............................................................................. 44
4.9 New reliability percentages of Method I by applying control variates ............................................................................. 44

D.1 Mean variance for 5000 experiments of Method I ............................................................................. 61
LIST OF TABLES

1.1 Notations and meanings ............................................. 7

2.1 An example of the values of $e_j$ ................................. 20
2.2 Global estimation of the values $e_j$ of the example Table 2.1 ............................................. 21

3.1 $\lambda$ and $\eta$ of $K = 9450$ ........................................ 30
3.2 Variance of crossing probability, reliability of $\varepsilon_{P_c} \leq 0.01$, running time in seconds of 5000 experiments in MATLAB of each method for $K = 9000$ and the required sample sizes to reach the approximation exactitude ........................ 32

C.1 Table of Normal Distribution $Q(x) = P(N(0,1) > x)$ ................. 59
INTRODUCTION

In a rectilinear drawing of a graph $G$, the vertices of $G$ are represented by points, and an edge joining two vertices is represented by the straight segment joining the corresponding two points. Edges are allowed to cross, but an edge cannot contain a vertex other than its endpoints. The rectilinear crossing number $Cr(G)$ of a graph $G$ is the minimum number of pairwise crossings of edges in a rectilinear drawing of $G$ in the plane. We consider the rectilinear crossing number of the complete graph $K_n$, whose vertices are drawn on a set $S$ of $n$ points in the plane. In this graph, each pair of points is connected by an edge. We therefore also talk of the rectilinear crossing number $Cr(S)$ of a point set $S$. We also only consider point sets $S$ in general position, meaning that no three points of $S$ are collinear. The minimum of $Cr(S)$ among all $n$-point sets $S$ is denoted $Cr(n)$.

The topic of the rectilinear crossing number has been studied by many researchers. The current best lower bound for the minimum rectilinear crossing $Cr(n)$ among all $n$-point sets, is by Ábrego, Cetina, Fernández-Merchant, Leaños, and Salazar [2, 4], stating that $\lim_{n \to \infty} \frac{Cr(n)}{{n \choose 4}} > 0.37997$. Around the year 2000, a team of researchers led by Aichholzer undertook the task of building databases with all the distinct $n$-point configurations in general position (no three points collinear), for $n \leq 10$. The raw knowledge of all possible $n$-point configurations put Aichholzer and his collaborators in a position to explore in depth several classical combinatorial geometry problems. In particular, it allowed for the exact calculation of $Cr(n)$ for small values of $n$. The Rectilinear Crossing Number project, led by Aichholzer [12], has been a fruitful source of inspiration as well as an invaluable tool for establishing results and testing conjectures [3].

However, $Cr(n)$ is not known for large values of $n$. This motivates to have a fast algorithm to calculate the crossing number of an $n$-point set. The calculation of the rectilinear crossing number of $n$-point sets, for large values of $n$, is quite time-consuming, though there is a known algorithm with quadratic time complexity [8, 16].

With this context, this thesis focuses on seeking another way for rectilinear crossing number calculation. With a computational algorithm we will calculate the rectilinear crossing number approximately instead of calculating it exactly. This is done by merging Monte Carlo approximation methods with techniques for the calculation of the rectilinear crossing number. Since all the outputs of Monte Carlo methods follow a normal distribution when the samples are independent, the exactitude of the approximated crossing number depends on the variance of the used Monte Carlo method. Obviously, when the variance is smaller, the exactitude is higher. This thesis contains six Monte Carlo methods. For each one we study the variance and the required sample size to guarantee that the approximated crossing number is very close to the real crossing number of the point set with high probability (95% confidence intervals are used).

For the purpose of a quick understanding, the necessary concepts will be introduced in the beginning of Chapter 1. In Chapter 2, we will present the underlying principles of the methods that we have thought during the work. And also, the variance of each method will be estimated. After this theoretical part, the experimental results of each method will be
shown in Chapter 3. Due to the page limitation, some experimental graphics will be shown in the Appendix. The basic Monte Carlo methods can still be optimized. Three variance reduction techniques for possible improvements of the developed Monte Carlo methods are discussed in Chapter 4. In particular, the control variate technique allows to enhance the performance by reducing the variances significantly.

Finally, we will compare the exactitudes of all the discussed Monte Carlo methods. We conclude with some further open problems and unresolved issues which arose during this work and which are left for future investigations.
CHAPTER 1. CONCEPT

In this chapter we present the Monte Carlo method and the definition of the rectilinear crossing number. This thesis focuses on how to combine these concepts to calculate the rectilinear crossing number approximately in a fast way.

1.1. Monte Carlo Method

There are two famous computational algorithms that rely on repeated random sampling to obtain numerical results. One is the Las Vegas algorithm which is a randomized algorithm that always gives the correct result if it finishes, but it may not produce an output. And another one is the Monte Carlo algorithm, which is a randomized algorithm that may produce incorrect results, but with bounded error probability. This method is useful for obtaining numerical solutions to problems which are too complicated to solve analytically and it is explained in detail in the book *Monte Carlo* by George Fishman [9]. Las Vegas algorithms can be contrasted with Monte Carlo algorithms, in which the resources used are bounded and the answer is not guaranteed to always be correct. In this thesis, a Monte Carlo algorithm is used to find an approximate solution of a problem in a quick way.

A Monte Carlo method is a computation process that uses random numbers to produce an output. Instead of having fixed inputs, probability distributions are assigned to some or all of the inputs. This will generate a probability distribution for the output after the simulation is run.

A famous example of applying this algorithm is to estimate the value of \( \pi \). The amount of area within a quarter of a disk of radius 1 depends on the value of \( \pi \). The probability that a randomly-chosen point of the unit square will lie in that quarter of a disk depends on the area of the disk. If points are placed randomly in a square with sides of length 1, the percentage of points that fall within a quarter-circle of radius 1 and center the origin, will depend on the value of \( \pi \). A Monte Carlo algorithm would randomly place points in the square like in Figure 1.1, and use the percentage of points falling inside of the disk to estimate the value of \( \pi \). This is an effective way for making approximations of \( \pi \).

1.2. Rectilinear Crossing Number

The crossing number \( Cr(G) \) of a simple graph \( G \) is the minimum number of edge crossings in any drawing of \( G \) in the plane, where each edge is a simple curve. The rectilinear crossing number \( Cr(G) \) is the minimum number of edge crossings when \( G \) is drawn in the plane using straight segments as edges [3].

In this paper we focus on the determination of the rectilinear crossing number, \( Cr(K_n) \), where \( K_n \) denotes the complete graph on \( n \) vertices which are \( n \) points in the plane, we also talk of the rectilinear crossing number \( Cr(S) \) of a point set \( S \). All the point sets we consider are in general position in the plane which means there are not three points of the
Figure 1.1: Approximate the value of $\pi$ with 30000 points situated randomly in the square $[11]$. 

set that lie on a straight line.

The minimum number of points to form a rectilinear crossing is four and the polygon formed by four points must be a convex quadrilateral. In this case, the crossing is the intersection of two diagonals. If the polygon is formed by four points which are not in convex position, there is no crossing, as shown in the figure below. An example of a random 100-point set

Figure 1.2: Convex quadrilateral with one crossing $S$ (left) and non-convex (concave) quadrilateral without crossing (right)

is shown in Figure 1.3.
1.3. Combination of Rectilinear Crossing Number and Monte Carlo Method

Four points in convex position is the base to determine if there is a crossing (as shown in Figure 1.2). If the point set is formed by $n$ points, to calculate the total rectilinear crossing number we should find all the combinations of four points which can form a convex quadrilateral. The total number of combinations of four points for a point set of $n$ points is $\binom{n}{4}$. To compute the crossing number exactly, a brute force algorithm would test all the $\binom{n}{4}$ combinations. Whenever four points form a convex quadrilateral, it sums one crossing.

The main disadvantage of this algorithm is that it needs a lot of time. Other algorithms are known, with quadratic time complexity [8, 16]. Still, these algorithms are very time consuming if the input point set has a large number of points. Thence, this project is dedicated to find a faster method than the conventional method by using a Monte Carlo algorithm with a limited approximation inexactitude. Consequently, there is a tradeoff between precision of the solution and running time of the algorithm.

The test whether four random points of a point set constitute a crossing is a Bernoulli random variable, which has two possible outcomes: 0 or 1. It takes the value 1 with success probability of $p$ and the value 0 with failure probability of $q = 1 - p$. In this case, when four random points constitute a crossing, the result is 1, otherwise the result is 0. The Monte Carlo experiment takes $N$ combinations of four random points. The resulting number of crossing will follow a binominal distribution since the binominal distribution is a sum of independent and identically distributed Bernoulli random variables. That is to say, if we repeat a Bernoulli experiments $B(p)$ $n$ times and count the number $X$ of successes, the distribution of $X$ is called the Binomial $B(n, p)$ random variable.
Where the mean $E(X)$ and variance $Var(X)$ will be:

$$E(X) = n \cdot p = \mu$$  \hspace{1cm} (1.1)

$$Var(X) = n \cdot p \cdot (1 - p) = \sigma^2$$  \hspace{1cm} (1.2)

For example, if the tested point set has $n$ points with a real crossing number $Cr$, then the probability of crossing $P_c$ of four random points from this point set can be calculated: dividing the crossing number by all combinations of four points which is

$$P_c = \frac{Cr}{\binom{n}{4}}$$  \hspace{1cm} (1.3)

The mean and variance of the approximated rectilinear crossing number $Cm$ after testing $K$ combinations of four points will be:

$$E(Cm) = K \cdot P_c$$  \hspace{1cm} (1.4)

$$Var(Cm) = K \cdot P_c \cdot (1 - P_c)$$  \hspace{1cm} (1.5)

Remarkably, when $n$, $np$ and $nq$ are large, then the binomial distribution is well approximated by the normal distribution $N(\mu, \sigma)$ according to the central limit theorem. It states that the arithmetic mean of a sufficiently large number of iterates of independent random variables, each with a well-defined expected value and well-defined variance, will be approximately normally distributed, regardless of the underlying distribution [17].

We will denote the approximation error of the presented Monte Carlo algorithms with epsilon ($\epsilon$), here we have $\epsilon_{P_c} = |P_{cm} - P_c|$. The quality of a Monte Carlo method can be estimated by its variance and the variance is related with the error. Obviously, when the variance of a method is minor, the standard deviation is smaller and the probability to get the correct answer is higher. That is to say, the required sample size will be smaller to get the same exactitude comparing to another Monte Carlo method's sample size whose variance is bigger.

### 1.4. Outline

The next chapters are organized in four parts:

- Part I: Theoretical study of Monte Carlo methods. We will introduce six Monte Carlo methods for the calculation of the rectilinear crossing number. The mean and variance of each method will be calculated and also the required sample size to reach a certain exactitude.

- Part II: Experimental part of Monte Carlo methods. In Chapter 3, there are experimental results for each method that we have mentioned in Chapter 2, which confirm the obtained results of Chapter 2.

- Part III: Improvements or variance reduction technics. In Chaper 4, we will apply some known techniques of variance reduction to the presented Monte Carlo algorithms and the
experimental data.

- Part IV: Conclusion. This part consists of a comparison of the applied methods and further possible problems.

Due to space limitation, some figures and text are deferred to the Appendix. To be quickly understandable here is the table of notations and their meanings:

<table>
<thead>
<tr>
<th>Notation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n )</td>
<td>Number of points in a point set</td>
</tr>
<tr>
<td>( C_r )</td>
<td>Real rectilinear crossing number of a point set</td>
</tr>
<tr>
<td>( C_m )</td>
<td>Rectilinear crossing number obtained by Monte Carlo method</td>
</tr>
<tr>
<td>( P_c )</td>
<td>Real probability of rectilinear crossing ([C_r/n^4])</td>
</tr>
<tr>
<td>( K )</td>
<td>Monte Carlo sample size</td>
</tr>
<tr>
<td>( H )</td>
<td>Number of obtained rectilinear crossings for ( K ) random samples</td>
</tr>
<tr>
<td>( P_m )</td>
<td>Probability of rectilinear crossing obtained by Monte Carlo method ( (H/K) )</td>
</tr>
<tr>
<td>( \varepsilon_P )</td>
<td>Error of rectilinear crossing probability ( (P_m - P_c) )</td>
</tr>
<tr>
<td>( \varepsilon_P )</td>
<td>Error of probability that two points lie on the same side of a segment ( (P_m - P_s) )</td>
</tr>
<tr>
<td>( Q(x) )</td>
<td>Probability that Normal distribution ( N(0, 1) &gt; x )</td>
</tr>
<tr>
<td>( P_c )</td>
<td>Probability that four points form a convex quadrilateral ( (1 - P_o) )</td>
</tr>
<tr>
<td>( P_o )</td>
<td>Probability that the circle formed by three points contains the fourth point inside</td>
</tr>
<tr>
<td>( P_m )</td>
<td>Probability that two points lie on the same side of a segment</td>
</tr>
<tr>
<td>( P_m )</td>
<td>Probability that two points lie on the same side of a segment obtained by Monte Carlo method</td>
</tr>
<tr>
<td>( P_o )</td>
<td>Probability that two points lie on the same side of a segment formed by another two points and vice versa</td>
</tr>
<tr>
<td>( P_m )</td>
<td>Probability that the circle formed by three points contains the fourth point inside</td>
</tr>
<tr>
<td>( P_m )</td>
<td>Probability that the circle formed by three points contains the fourth point inside obtained by Monte Carlo method</td>
</tr>
<tr>
<td>( P_o )</td>
<td>Probability that these four points form a convex quadrilateral</td>
</tr>
<tr>
<td>( P_o )</td>
<td>Probability that these four points form a concave quadrilateral</td>
</tr>
<tr>
<td>( P_o )</td>
<td>Probability that these four points form a convex quadrilateral</td>
</tr>
<tr>
<td>( P_o )</td>
<td>Probability that these four points form a concave quadrilateral</td>
</tr>
<tr>
<td>( \varepsilon_j )</td>
<td>Crossing probability of a random ( n )-point set selected from the original point set</td>
</tr>
<tr>
<td>( J )</td>
<td>Number of segments which have ( j ) points in one side</td>
</tr>
<tr>
<td>( k )</td>
<td>Number of points inside a circle defined by three random points</td>
</tr>
</tbody>
</table>

Table 1.1: Notations and meanings
CHAPTER 2. MONTE CARLO METHODS IN RECTILINEAR CROSSING NUMBER

In this chapter we will introduce the working principles of the different Monte Carlo methods we have obtained and study the sample size they need to obtain a given exactitude. Method I, II, III and IV use the geometric property of four points to form a crossing whereas Method V and VI take advantage of current research results about the rectilinear crossing number. In Chapter 3, the experimental data and the required sample sizes (Table 3.2) of each method are then given.

2.1. Method I — Points

This method imitates the brute force method, which calculates the exact crossing number of an \( n \)-point set by computing all \( \binom{n}{4} \) combinations. Nevertheless, this method calculates the crossing number with a certain inexactitude by choosing only \( K \) combinations at random, with a \( K \) much smaller than \( \binom{n}{4} \).

If there are \( H \) crossings for \( K \) combinations, the crossing number \( Cr \) will be approximated in this way:

\[
Cr \approx Cm = \frac{H}{K} \cdot \binom{n}{4} = P_{cm} \cdot \binom{n}{4}
\]  

(2.1)

Where \( P_{cm} \) is the crossing probability obtained by the Monte Carlo method.

To obtain the total crossing number \( H \) for \( K \) combinations, the first step will be how we can determine whether only four points form a rectilinear crossing.

2.1.1. Base Case - One Crossing Determination

As shown in Figure 1.2, there will be a rectilinear crossing if and only if four points form a convex quadrilateral. Therefore, the problem is converted to determine whether or not the polygon defined by four points is a convex quadrilateral. There are more than one ways to solve the case, but the method that we use to determine if a polygon is a convex quadrilateral has the following explanation.

According to the geometric property of a convex quadrilateral, when all the straight lines between two vertices are drawn, see Figure 1.2 (left), it is easy to see that the two diagonals \( AC, BD \) have one of the remaining two points of the quadrilateral on each side and the four exterior edges \( AB, BC, CD, AD \) have both remaining points of the quadrilateral on the same side. On the other side, for a non-convex quadrilateral, see Figure 1.2 (right), the three diagonals \( AC, BC, CD \) have one of the remaining two points of the quadrilateral on each side and the three exterior edges \( AB, AD, BD \) have both remaining points of the quadrilateral on the same side.

Hence, the problem has been simplified even more - to determine whether a point \( C \) lies
on the left or on the right of the straight-line segment $AB$. Thereto it is sufficient to calculate the determinant to know the relative position. For instance, if we want to determine the position of $C$ with respect to the directed segment $AB$ we have to calculate the determinant:

$$\text{det} = \begin{vmatrix} A_x & B_x & C_x \\ A_y & B_y & C_y \\ 1 & 1 & 1 \end{vmatrix}$$ (2.2)

If $\text{det} > 0$, point $C$ lies on the left side of the directed segment $AB$, otherwise $C$ lies on the right side. Note that we discard the possibility of three aligned points, in which case the determinant will be 0, because of the general position assumption.

Method I is hence to randomly choose four points, $K$ times, and count the relative frequency of crossings. Note that our random variable is not the number of crossings, but the crossing probability $P_{cm}$. In order to ensure the correctness of the answer, the next step must be to find the sample size $K$ to reach a deterministic exactitude. We define the error range of crossing probability ($\varepsilon_{P_c}$) and its confidence interval: we propose a reliability of 95% that the crossing probability error $\varepsilon_{P_c}$ is less than 0.01. This means, the sample size $K$ is chosen in a way that the 95% confidence interval is $P_c \in [P_{cm} - 0.01, P_{cm} + 0.01]$.

### 2.1.2. Sample Size Determination

As we mentioned in Chapter 1, the calculated crossing number follows a binomial distribution $B(n, p)$ which is approximated by the normal distribution $N(np, \sqrt{np(1-p)})$. In this case we repeat $K$ times a Bernoulli experiment with a real probability of crossing $P_c$, thence, the resulting distribution of the number of crossings will be $B(K, P_c)$ with normal approximation:

$$N(KP_c, \sqrt{KP_c(1-P_c)})$$

We now estimate the required number of samples $K$ using the normal approximation for the binomial distribution following the steps explained in [10], page 287. First, we have to standardize the normal distribution:

$$Z_n = \frac{N(\mu, \sigma) - \mu}{\sigma} \Rightarrow N(0, 1)$$

In this case:

$$Z_n = \frac{N(\mu, \sigma) - KP_c}{\sqrt{KP_c(1-P_c)}}$$

Let $P_{cm}$ be the relative frequency of crossings in $K$ Bernoulli trials, which is the crossing probability calculated by Monte Carlo method. Then

$$Z_n = \frac{KP_{cm} - KP_c}{\sqrt{KP_c(1-P_c)}} = \frac{P_{cm} - P_c}{\sqrt{P_c(1-P_c)}}$$ (2.3)
Then $Z_n$ has zero mean and unit variance, and is approximately normal for $K$ sufficiently large. The probability of interest is

$$P[|P_{cm} - P_c| < \varepsilon P_c] \simeq P \left[ |Z_n| < \frac{\varepsilon P_c \cdot \sqrt{K}}{\sqrt{P_c(1 - P_c)}} \right] = 1 - 2Q \left( \frac{\varepsilon P_c \cdot \sqrt{K}}{\sqrt{P_c(1 - P_c)}} \right)$$  \hspace{1cm} (2.4)

Where $Q(x)$ is the probability that $N(0,1) > x$ and $\varepsilon P_c$ is the error of the crossing probability.

We want the above probability (2.4) to equal 95%. This implies that

$$Q \left( \frac{\varepsilon P_c \cdot \sqrt{K}}{\sqrt{P_c(1 - P_c)}} \right) = \left( 1 - 0.95 \right) / 2 = 0.025.$$  \hspace{1cm} (1 - 0.95)/2 = 0.025. From Table C of the appendix, we see that the argument of $Q(x)$ should be approximately 1.95, thus

$$\frac{\varepsilon P_c \cdot \sqrt{K}}{\sqrt{P_c(1 - P_c)}} = 1.95$$  \hspace{1cm} (2.5)

The equation has three variables: error of crossing probability ($\varepsilon P_c$), required number of samples ($K$) and crossing probability ($P_c$). The relation between these terms is shown in Figure 2.1. As we proposed a probability error $\varepsilon P_c \leq 0.01$, the maximum required sample size to ensure a 95% confidence interval is also shown in the plane for $\varepsilon P_c = 0.01$ (Figure 2.2). We see that a sample size of $K = 9506$ always is sufficient to reach an error $\varepsilon P_c$ of at most 0.01 with reliability of 95%.

### 2.1.3. Mean and Variance

Since the sample size depends on the probability of crossings $P_c$ of the point set, we have defined the approximated crossing probability $P_{cm}$ instead of the approximated crossing number $Cm$, as the evaluation criterion of the Monte Carlo algorithm. On the other side, this choice allows us to compare the behaviour of the algorithm for point sets of different sizes, since the size $n$ does not carry weight with the crossing probability $P_c$. Thus, the mean and the variance change- if we test $K$ samples of four random points of a point set with crossing probability $P_c$, the expected number of crossing and the variance will be

$$E(X') = \frac{E(X)}{K} = \frac{KP_c}{K} = P_c \hspace{1cm} (2.6)$$

$$Var(X') = \frac{Var(X)}{K^2} = \frac{KP_c(1 - P_c)}{K^2} = \frac{P_c(1 - P_c)}{K} \hspace{1cm} (2.7)$$

### 2.2. Method II — Segments

We now use a geometric property of a convex quadrilateral, similar as done in Section 1.2.. We choose randomly 2 points of the point set to form a segment and choose randomly another 2 points. Cases when two points lie on the same side of a segment:

- for a convex quadrilateral, see Figure 1.2 (left):
Figure 2.1: Relation between probability error ($\epsilon_{P_c}$), required sample size ($K$) and crossing probability ($P_c$)

1. segment $AB$ with points $C$ and $D$
2. segment $BC$ with points $D$ and $A$
3. segment $CD$ with points $A$ and $B$
4. segment $DA$ with points $B$ and $C$
There are four segments out of six with the other two points lying on the same side. Hence, the conditional probability that these 2 points lie on the same side of the segment, if the four points form a convex quadrilateral, is $P(\Box) = \frac{4}{6}$.

- for a non-convex quadrilateral, see Figure 1.2 (right):
  1. segment $AB$ with points $C$ and $D$
  2. segment $BD$ with points $C$ and $A$
  3. segment $DA$ with points $C$ and $B$

There are three segments out of six with the other two points lying on the same side. Hence the conditional probability is $P(\triangledown) = \frac{3}{6}$.

Note that $P(\cdot)$ is the probability of one random segment with two random points on the same side. $P(\Box)$ is the probability that four randomly selected points form a convex quadrilateral and $P(\triangledown)$ is the probability that four randomly selected points form a non-convex quadrilateral which can be expressed as $1 - P(\Box)$.

In conclusion, the total probability $P(\cdot)$ is the sum of these two cases:

$$P(\cdot) = P(\Box) \cdot P(\Box) + P(\triangledown) \cdot P(\triangledown)$$

$$= P(\Box) \cdot P(\Box) + P(\triangledown) \cdot (1 - P(\Box))$$

$$= \frac{4}{6} \cdot P(\Box) + \frac{3}{6} \cdot (1 - P(\Box))$$

$$= \frac{1}{2} + \frac{1}{6} \cdot P(\Box)$$

Finally we get the expression of crossing probability for Method-II:

$$P_s = \frac{1}{2} + \frac{1}{6} \cdot P_c \Rightarrow P_c = 6 \cdot \left( P_s - \frac{1}{2} \right)$$

(2.8)

Where $P_c = P(\Box)$ and $P_s = P(\cdot)$. 
2.2.1. Mean and Variance

We calculate mean and variance of this method as described in Section 2.1.3., consider the random variable \( X = P_{sm} \) which is the probability that two random points lie on the same side of a random segment. Replace the probability \( P_{cm} \) with \( P_{sm} \) and the number of samples \( K \) into the equations of mean (Eq. 2.6) and variance (Eq. 2.7):

\[
E(X) = P_s \tag{2.9}
\]

\[
Var(X) = \frac{P_s (1 - P_s)}{K} \tag{2.10}
\]

Replace \( P_s \) with \( P_c \) into Eq. 2.8, the variance of \( X \) is equal to:

\[
Var(X) = \frac{\left( \frac{1}{2} + \frac{1}{6} \cdot P_c \right) \cdot \left( 1 - \left( \frac{1}{2} + \frac{1}{6} \cdot P_c \right) \right)}{K} = \frac{1}{4} - \frac{1}{36} \cdot P_c^2 \tag{2.11}
\]

Once we get the mean and variance of \( P_{sm} \), the mean and variance of \( X' = P_{cm} \) can be estimated by substituting in Eq. 2.8:

\[
E(X') = P_c = 6 \cdot \left( P_s - \frac{1}{2} \right) \tag{2.12}
\]

\[
Var(X') = Var\left( 6 \cdot \left( P_{sm} - \frac{1}{2} \right) \right) = 6^2 \cdot Var(P_{sm}) = 36 \cdot Var(X) = \frac{9 - P_c^2}{K} \tag{2.13}
\]

2.2.2. Sample Size Determination

A set of \( n \) points determines \( \binom{n}{2} \) segments and for each segment, there are \( \binom{n-2}{2} \) pairs of other points. This means, the number of combinations to be tested to get the exact solution is:

\[
\binom{n}{2} \cdot \binom{n-2}{2} = \frac{n \cdot (n-1)}{2!} \cdot \frac{(n-2) \cdot (n-3)}{2!} = \binom{n}{4} \cdot 6 \tag{2.14}
\]

To estimate the real crossing number of an \( n \)-point set it will cost six times more than to run the brute force algorithm which performs \( \binom{n}{4} \) tests. In this method we repeat \( K \) times a Bernoulli experiment with a probability \( P_{(i)} \), thence, the resulting distribution will be \( B(K, P_i) \) with normal approximation: \( N(KP_s, \sqrt{KP_s(1-P_s)}) \). Following the procedures of Section 2.1.2. of this chapter we get the same sample size \( K = 9506 \) to reach the precision \( P(\varepsilon_{P_s} \leq 0.01) = 95\% \). By using Eq. 2.8 the exactitude of approximation can also be obtained from this sample size: \( P(\varepsilon_{P_c} \leq 0.06) = 95\% \). That means, if we want to get \( P(\varepsilon_{P_c} \leq 0.01) = 95\% \), \( K \) must be chosen such that with 95% reliability we have \( \varepsilon_{P_s} \leq \frac{0.01}{6} \). Then \( K \) can be estimated by Eq. 2.5 by replacing \( P_c \) with \( P_s \):

\[
\frac{\varepsilon_{P_s} \sqrt{K}}{\sqrt{P_s(1-P_s)}} = 1.95 \Rightarrow \frac{\varepsilon_{P_c} \sqrt{K}}{\sqrt{P_s(1-P_s)}} = 1.95 \Rightarrow \frac{0.01 \sqrt{K}}{6 \sqrt{P_s(1-P_s)}} = 1.95 \tag{2.15}
\]
Or another way to calculate the required sample size $K$ to reach $P(\varepsilon_P \leq 0.01) = 95\%$:

follow the steps described in Section 2.1.2, but with the variance that we calculated before $\text{Var}(X = P_{cm}) = \frac{9 - P_c^2}{K}$. We will get the expression of $K$ in relation with $P_c$:

$$\frac{\varepsilon_P \sqrt{K}}{\sqrt{9 - P_c^2}} = 1.95 \Rightarrow \frac{0.01 \sqrt{K}}{\sqrt{9 - P_c^2}} = 1.95$$

(2.16)

Both methods are available to get the required sample size, see Figure 2.3. Method II has a larger $K$ than Method I, it means, we need to test much more combinations to get the same precision. However, its computation velocity is quicker than the first method since it only determines the position of two points with respect to a given segment.

### 2.3. Method III — Circles

This method has a similar mechanism as Method II which takes into account the geometric properties of convex quadrilaterals. The difference is that this method uses the probability that the remaining point falls inside the circle formed by the other three points $P(O)$ instead of $P(\|)$. We pick randomly four points, the probabilities $P(O)$ that the disk formed by the first three points contains the fourth points inside are not the same in convex quadrilaterals and in non-convex quadrilaterals. Figure 2.4 shows plainly the difference.

The points on the left of Figure 2.4 form a convex quadrilateral and the points on the right do not. A circle is defined by three points, hence every quadrilateral has $\binom{4}{3} = 4$ circles. The red circles contain the fourth point inside and the black circles do not. The probability of containing a point inside a circle formed by another three points can be derived as follows:

- for a convex quadrilateral, see Figure 2.4 (left):
1. circle $ABC$ contains point $D$

2. circle $CDA$ contains point $B$

There are two circles out of four containing the fourth point inside. Hence, the conditional probability is $P(⨀|□) = \frac{2}{4}$.

- for a non-convex quadrilateral, see Figure 2.4 (right):

1. circle $CDA$ contains point $B$

There is one circle out of four containing the fourth point inside. Hence, the conditional probability in this case is $P(⨀|⋗) = \frac{1}{4}$.

The total probability $P(⨀)$ is the sum of these two cases:

$$P(⨀) = P(⨀|□) \cdot P(□) + P(⨀|⋗) \cdot P(⋗)$$

$$= P(⨀|□) \cdot P(□) + P(⨀|⋗) \cdot (1 - P(□))$$

$$= \frac{2}{4} \cdot P(□) + \frac{1}{4} \cdot (1 - P(□))$$

$$= \frac{1}{4} + \frac{1}{4} \cdot P(□)$$

Finally we get the expression for the crossing probability:

$$P_o = \frac{1}{4} + \frac{1}{4} \cdot P_c \Rightarrow P_c = 4 \cdot \left( P_o - \frac{1}{4} \right)$$

(2.17)

Where $P_c = P(□)$ and $P_o = P(⨀)$.

### 2.3.1. Mean and Variance

In the same way as in Section 2.1.3., consider the random variable $X = P_{om}$ which is the probability that the circle formed by three random points contains another random
point inside. Replace the probability $P_{cm}$ with $P_{om}$ and the number of samples $K$ into the equations of mean (Eq. 2.6) and variance (Eq. 2.7):

$$E(X) = P_o$$

(2.18)

$$Var(X) = \frac{P_o(1 - P_o)}{K}$$

(2.19)

Replace $P_o$ with $P_c$ into Eq. 2.17, the variance of $X$ is equal to:

$$Var(X) = \left(\frac{1}{4} + \frac{1}{4} \cdot P_c\right) \left(1 - \left(\frac{1}{4} + \frac{1}{4} \cdot P_c\right)\right) = \frac{(3 - P_c)(P_c + 1)}{16K}$$

(2.20)

Once we get the mean and variance of $P_{om}$, the mean and variance of $X' = P_{cm}$ can be estimated by substituting in Eq. 2.17:

$$E(X') = P_c = 4 \cdot \left(P_o - \frac{1}{4}\right)$$

(2.21)

$$Var(X') = Var\left(4 \cdot (P_{om} - \frac{1}{4})\right) = 4^2 \cdot Var(P_{om}) = 16 \cdot Var(X) = \frac{(3 - P_c)(P_c + 1)}{K}$$

(2.22)

### 2.3.2. Sample Size Determination

For an $n$-point set, there are $\binom{n}{3}$ circles and for each circle we can select one of $(n - 3)$ points. The total number of combinations is:

$$\binom{n}{3} \cdot (n - 3) = \frac{n \cdot (n - 1) \cdot (n - 2)}{3!} \cdot (n - 3) = \binom{n}{4} \cdot 4$$

(2.23)

That is, to obtain the real crossing number of an $n$-point set it will cost four times more tests than in the brute force algorithm, which tests $\binom{n}{4}$ combinations. In this method we repeat $K$ times a Bernoulli experiment with a probability $P_{(O)}$, therefore, the resulting distribution will be $B(K, P_o)$ with normal approximation: $N(KP_o, \sqrt{KP_o(1 - P_o)})$. Following the procedures of Section 2.1.2. of this chapter we get the same sample size (9506) to reach the precision $P(\epsilon_{P_o} \leq 0.01) = 95\%$. The exactitude of crossing probability $P_c$ can also be determined by using the Eq. 2.17 : $P(\epsilon_{P_c} \leq 0.04) = 95\%$.

As explained in Method II Section 2.2.2., here also are two ways to calculate the required sample size $K$ for the condition that $P(\epsilon_{P_c} \leq 0.01) = 95\%$: one is to use $P(\epsilon_{P_o} \leq \frac{0.01}{4}) = 95\%$ and another one is to use the variance of this method that we computed before: $Var(X = P_{cm}) = \frac{(3 - P_c)(P_c + 1)}{K}$. The first one has an equation related with $P_o$:

$$\frac{0.01 \sqrt{K}}{4 \sqrt{P_o(1 - P_o)}} = 1.95$$

(2.24)

And the second one is related with $P_c$:

$$\frac{0.01 \sqrt{K}}{\sqrt{(3 - P_c)(P_c + 1)}} = 1.95$$

(2.25)
The relation between sample size $K$ and crossing probability $P_c$ is shown in Figure 2.5: We can find that the sample size is smaller than in Method II but larger than in Method I. The computation time is also between these two methods. It means, this method is not the most effective one.

2.4. Method IV — Downsizing

As the name indicates, this method reduces a point-set to a smaller one which is easier to be computed. For instance, a 100-point set requires $\binom{100}{4}$ calculation steps. If we choose randomly 20 points which belong to the original 100-point set to form a new set of 20 points, to estimate the exact crossing number $Cr_{20}$ it will only take $\binom{20}{4}$ steps, the Monte Carlo crossing number for the original 100-point set $Cm_{100}$ will be computed proportionally following this model:

$$Cm_{100} = \frac{Cr_{20}}{\binom{20}{4}} \cdot \binom{100}{4} = P_{c20} \cdot \binom{100}{4} \approx Cr$$

To make the results more precise, we can select randomly $\lambda$ times 20 points to form a new point set and retain the crossing number of each set ($Cr_{20_1}, Cr_{20_2}...Cr_{20_\lambda}$). The final crossing number will be calculated by the mean crossing number of these point sets:

$$Cm_{100} = \frac{\hat{Cr}_{20}}{\binom{20}{4}} \cdot \binom{100}{4} = \frac{\sum_{i=1}^{\lambda} Cr_{20_i}}{\binom{20}{4}} \cdot \binom{100}{4} = \bar{P}_{c20} \cdot \binom{n}{4}$$

(2.26)

where

$$\bar{P}_{c20} = \frac{\hat{Cr}_{20}}{\binom{20}{4}}$$
Instead of using subsets of 20 points, we could also use subsets of \( \eta \) points, for \( 4 \leq \eta \leq n \). Then the number of combinations \( K \) which are tested in total is \( K = \binom{\eta}{4} \cdot \lambda \). Obviously if \( K \) is bigger, the final result will be closer to the correct answer, however, the problem is: for the same \( K \), which \( \eta \) and \( \lambda \) should be choosen to get the best accuracy? We will find out the answer experimentally.

### 2.4.1. Mean and Variance

Define \( X' = \bar{P}_{c\eta m} \) as the random variable which is the mean crossing probability of \( \lambda \) subsets of \( \eta \) random points. Hence

\[
E(X') = P_c 
\]  

(2.27)

Where \( P_c \) is the crossing probability of the \( n \)-point set.

For each of \( \binom{n}{4} \) experiments, the variance of the crossing probability is the same as the variance explained in Method I - we test \( \binom{n}{4} \) samples of four points extracted from a random \( \eta \)-point set with crossing probability \( P_{c\eta} \), the variance of \( X = P_{c\eta m} \) will be Eq. 2.7.

\[
Var(X) = \frac{P_{c\eta}(1 - P_{c\eta})}{\binom{n}{4}} \]  

(2.28)

Repeat the experiment \( \lambda \) times, the average crossing probability \( \bar{X} \) is calculated by:

\[
\bar{X} = \bar{P}_{c\eta m} = \frac{\sum_{i=1}^{\lambda} P_{c\eta m_i}}{\lambda} 
\]  

(2.29)

The random variables \( P_{c\eta m_i} \) are independent and different, so that \( \sum_{i=1}^{\lambda} Var(P_{c\eta m_i}) \) cannot be simplified, thus the variance of \( X' = \bar{P}_{c\eta m} \) is:

\[
Var(X') = \frac{\sum_{i=1}^{\lambda} Var(P_{c\eta m_i})}{\lambda^2} \neq \frac{\lambda \cdot Var(P_{c\eta m_1})}{\lambda^2} 
\]  

(2.30)

Since we cannot calculate the variance exactly we will find the answer to the previous question out with experimental data in Chapter 3.

### 2.5. Method V — \( j \)-edges

This method is inspired by a work published by Lovász, Vesztergombi, Wagner and Welzl [15]. They prove that the minimum number of convex quadrilaterals determined by \( n \) points in general position in the plane, or in other words, the rectilinear crossing number of the complete graph \( K_n \), is at least \( \left( \frac{3}{8} + 10^{-5} \right) \binom{n}{4} + O(n^{-3}) \), and for every set of \( n \) points in the plane in general position the number of convex quadrilaterals (denoted as \( \square \)) is equivalent to:

\[
\square = \frac{1}{2} \left( \sum_{j=0}^{n-2} e_j \left( \frac{(n-2)(n-3)}{4} - j(n-j-2) \right) \right) 
\]  

(2.31)
Here \( e_j \) is number of directed segments defined by points of the given \( n \)-point set, which have \( j \) points on one side and \( n - j - 2 \) on the other side. An example is shown in Figure 2.6:

![9-point set diagram](image)

Figure 2.6: A 3-edge leaves 3 points on one side and 4 points on the other side in a 9-point set.

The above figure illustrates how a segment divides the point set into two parts, the upper part contains 4 points and the lower part contains 3 points, hence this segment is a 3-edge and a 4-edge.

For each segment we have to find \( j \) such that the segment is a \( j \)-edge, this means we have to test \( n - 2 \) points with respect to the segment, and there are \( \binom{n}{2} \) segments. Thus, by applying Eq. 2.31 to get the real crossing number, it will take \( \binom{n}{2} \cdot (n - 2) = 3 \cdot \binom{n}{3} \) calculation steps.

To apply this equation in the Monte Carlo method, the main point is to find the values of \( e_j \) for \( j \in \{0, 1, \cdots, n - 2\} \). Instead of computing the values of \( e_j \) from all the \( \binom{n}{2} \) segments we extract \( K \) segments randomly and count number of the points on each side of the \( K \) segments. Then we count the number of \( j \)-edges for each \( j = 0, 1, \cdots, n - 2 \), note that \( e_j = e_{n-j-2} \) holds for all \( j \in \{0, \cdots, n-2\} \). Finally we get a table of \( j \)-edges, see Table 2.1. Once we get the table of \( j \)-edges, we can get the global estimation of \( j \)-edges

<table>
<thead>
<tr>
<th>( j )-edge</th>
<th>Number ( e_j )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-edge</td>
<td>6</td>
</tr>
<tr>
<td>2-edge</td>
<td>5</td>
</tr>
<tr>
<td>3-edge</td>
<td>9</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>( \binom{n-2}{2} )-edge</td>
<td>4</td>
</tr>
<tr>
<td>Total</td>
<td>( K )</td>
</tr>
</tbody>
</table>

Table 2.1: An example of the values of \( e_j \)

proportionally. The global values of \( e_j \) are which we use in the Eq. 2.31 to compute the approximate crossing number. For example, we test \( K \) segments in an \( n \)-point set and
there are 6 segment which leave 1 point on one side and \( n - 2 - 1 \) points on the other side (6 out of \( K \) segments are type \( e_1 \)), hence the global estimated number of \( e_1 \) is \( \frac{6}{K} \binom{n}{2} \). See Table 2.2:

<table>
<thead>
<tr>
<th>( j )-edge</th>
<th>Number ( e_j )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-edge</td>
<td>( \frac{6}{K} \cdot \binom{n}{2} )</td>
</tr>
<tr>
<td>2-edge</td>
<td>( \frac{5}{K} \cdot \binom{n}{2} )</td>
</tr>
<tr>
<td>3-edge</td>
<td>( \frac{9}{K} \cdot \binom{n}{2} )</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>( \frac{n-2}{2} )-edge</td>
<td>( \frac{4}{K} \cdot \binom{n}{2} )</td>
</tr>
<tr>
<td>Total</td>
<td>( \binom{n}{2} )</td>
</tr>
</tbody>
</table>

Table 2.2: Global estimation of the values \( e_j \) of the example Table 2.1

2.5.1. Mean and Variance

The Eq. 2.31 shows the principal relation between \( e_j \) and the crossing number.

\[
\square = \frac{1}{2} \left( \sum_{j=0}^{n-2} e_j \frac{(n-2)(n-3)}{4} - \sum_{j=0}^{n-2} j e_j (n-j-2) \right) \tag{2.32}
\]

Where \( \square \) is the number of convex quadrilaterals. We can simplify the equation by using that \( \sum_{j=0}^{n-2} e_j = n(n-1) \) and \( e_j = e_{n-2-j} \) which gives us \( \sum_{j=0}^{n-2} j e_j = \binom{n}{2} \):)

\[
\square = 3 \cdot \binom{n}{4} - \frac{1}{2} \sum_{j=0}^{n-2} j e_j (n-j-2) \\
= 3 \cdot \binom{n}{4} - \frac{1}{2} \left( \sum_{j=0}^{n-2} j e_j (n-2) - \sum_{j=0}^{n-2} j^2 e_j \right) \\
= 3 \cdot \binom{n}{4} - \frac{1}{2} \left( (n-2)^2 \binom{n}{2} - \sum_{j=0}^{n-2} j^2 e_j \right)
\]

We define the random variable \( J \), where \( J \) is the number of points lying on one side of a random segment. The mean of \( J \) and \( J^2 \) can be expressed as

\[
E(J) = \sum_{j=0}^{n-2} j \frac{e_j}{2 \binom{n}{2}} = \frac{(n-2) \binom{5}{2}}{2 \binom{n}{2}} = \frac{n-2}{2} \tag{2.33}
\]

\[
E(J^2) = \sum_{j=0}^{n-2} j^2 \frac{e_j}{2 \binom{n}{2}}
\]

Hence, the crossing number \( \square \) is

\[
\square = Cr = 3 \binom{n}{4} - \frac{3}{2} (n-2) \binom{n}{3} + \binom{n}{2} E(J^2) \tag{2.34}
\]
Define the approximated crossing number \( C_{m} \) as the random variable \( C_{m} = 3 \binom{n}{4} - \frac{3}{2} (n - 2) \binom{n}{3} + \binom{n}{2} J^2 \) which has \( E(C_{m}) = Cr \), and

\[
Var(C_{m}) = \left( \binom{n}{2} \right)^2 \cdot Var(J^2)
\]  

(2.35)

\( Var(J^2) \) can be calculated by \( E(J^4) - E(J^2)^2 \). Due to the lack of numerical expressions of \( E(J^2) \) and \( E(J^4) \), \( Var(J^2) \) cannot be estimated in this case, neither the sample size. Thus, we deduce the variance of crossing probability from its experimental data in Chapter 3.

2.6. Method VI — \( k \)-InsideCircle

This method uses the research results published by Fabila-Monroy, Huemer and Tramuns [7] and by Urrutia [18]. They discovered that the rectilinear crossing number of a point set is related with the variance of the number of points inside the disk formed by three random points which belong to the point set. Let \( S \) be a plane point set of \( n \) points with no three of them collinear and no four cocircular. Consider the circle passing through three points of \( S \) chosen uniformly at random and let \( k \) be the random variable that counts the number of points of \( S \) inside this circle. Urrutia showed that the expectation \( E(k) \) only depends on the rectilinear crossing number \( Cr \) of \( S \). Fabila-Monroy et al. prove that this also holds for the variance \( Var(k) \). More precisely,

\[
E(k) = \frac{Cr}{\binom{n}{3}} + \frac{n - 3}{4} \Rightarrow C_{m} = \left( k - \frac{n - 3}{4} \right) \cdot \binom{n}{3}
\]  

(2.36)

where \( C_{m} \) is the approximated crossing number which has the mean \( E(C_{m}) = Cr \). Then

\[
Var(k) = (n - 3)^2 \left( \frac{Cr}{8 \binom{n}{4}} - \frac{Cr^2}{16 \binom{n}{4}} + \frac{1}{80} + \frac{1}{5(n - 3)} \right)
\]  

(2.37)

In this method, we use Eq. 2.36 to estimate the Monte Carlo crossing number \( C_{m} \) from \( k \). The same methodology as in other Monte Carlo methods is applied: pick randomly three points to from a circle, then count the number of points \( k \) inside the disk to estimate \( Cr \). Repeat this experiment \( K \) times and finally we take the average of \( K \) estimations of the crossing number:

\[
\bar{C}_{m} = \frac{\sum_{i=1}^{K} C_{m_i}}{K}
\]  

(2.38)

Note that \( Cr \approx \bar{C}_{m} \).

2.6.1. Mean and Variance

Although the above mentioned equations are related with the crossing number \( Cr \), we define its probability as the random variable \( X = P_{cm} = \binom{Cr}{\binom{n}{4}} \). Thus, the mean is the same as the mean of other methods, \( E(X) = \bar{C}_{c} \), but the variance has changed:
\[
Var(X) = \frac{Var\left(\binom{n}{a}\right)}{(\binom{n}{4})^2} = \frac{Var(Cm)}{(\binom{n}{4})^2}
\]

Where \(Var(Cm)\) is known by Eq. 2.36: \(Var(Cm) = Var(k)^2\). Then

\[
\frac{Var(Cm)}{(\binom{n}{4})^2} = \frac{Var(k)^2}{(\binom{n}{4})^2} = \frac{16}{(n - 3)^2} Var(k)
\]

Replacing \(Var(k)\) of Eq. 2.37:

\[
Var(X) = 16 \left( \frac{Cr}{8(\binom{n}{4})^2} - \frac{Cr^2}{16(\binom{n}{4})^2} - \frac{1}{80} + \frac{1}{5(n - 3)} \right) = 2P_c - P_c^2 - \frac{1}{5} + \frac{16}{5(n - 3)} \quad (2.39)
\]

The result shows that the variance of the crossing probability for one experiment depends on the real crossing probability \(P_c\) and the number of points \(n\) of the chosen point set. We repeat the experiment \(K\) times (all the experiments are independent) and take the average, so that \(\bar{X} = \frac{\sum_{i=1}^{K} X_i}{K}\). The variance will be:

\[
Var(\bar{X}) = Var\left(\frac{\sum_{i=1}^{K} X_i}{K}\right) = \frac{\sum_{i=1}^{K} Var(X_i)}{K^2}
\]

For a point set, its crossing probability and number of points are fixed. Thus, \(Var(X_i)\) is always the same:

\[
Var(\bar{X}) = \frac{K \cdot Var(X)}{K^2} = \frac{Var(X)}{K} = 2P_c - P_c^2 - \frac{1}{5} + \frac{16}{5(n - 3)}
\]

(2.40)

### 2.6.2. Sample Size Determination

The required sample size to ensure that the probability error is small, depends on the variance of the normal distribution with the expected crossing probability \(E(X = P_{cm}) = P_c\) and standard deviation \(\sigma = \sqrt{Var(\bar{X})} \Rightarrow N(P_c, \sqrt{Var(\bar{X})})\). Follow the same steps that we have explained in Section 2.1.2., but using the new variance \(Var(\bar{X}) = \frac{2P_c - P_c^2 - \frac{1}{5} + \frac{16}{5(n - 3)}}{K}\).

The new sample size can be calculated from the normal distribution (Eq. 2.3):

\[
N(0, 1) \Rightarrow Z_n = \frac{N(\mu, \sigma) - P_c}{\sqrt{Var(\bar{X})}} = \frac{P_{cm} - P_c}{\sqrt{2P_c - P_c^2 - \frac{1}{5} + \frac{16}{5(n - 3)}}}
\]

(2.41)

The new version of Eq. 2.5 has now four variables: crossing probability \(P_c\), error of crossing probability \(\varepsilon_{P_c}\), number of points \(n\) and required sample size \(K\) to reach 95% reliability (the reliability can also be defined as the fifth variable). We know for 95% reliability \(P(N(0, 1) > x) = Q(x) = 0.025 \Rightarrow x \approx 1.95\):

\[
\frac{\varepsilon_{P_c} \sqrt{K}}{\sqrt{2P_c - P_c^2 - \frac{1}{5} + \frac{16}{5(n - 3)}}} = 1.95
\]

(2.42)
Consider, for example, a 100-point set with crossing number $Cr = 1463457$: $n=100, P_c = 1463457 \binom{100}{4} = 0.3732$. To ensure a 95% confidence interval with error $\varepsilon_{P_c} = 0.01$, that is $P_c \in [P_{cm} - 0.01, P_{cm} + 0.01]$ happens with probability 95%, the required sample size $K$ is given by

$$
\frac{0.01 \sqrt{K}}{\sqrt{2 \cdot 0.3732 - 0.3732^2 - \frac{1}{5} + \frac{16}{5(100-3)}}} = 1.95 \Rightarrow K = 16736
$$
CHAPTER 3. EXPERIMENTAL RESULTS

In this chapter we will present the experimental data of the methods we explained in Chapter 2. We use MATLAB 2013a to compile the codes (which are omitted in this thesis). We use point sets which have different crossing probabilities and different number of points to compare the methods and detect differences. The point sets which have the minimum known crossing probability \( P_c \) are extracted from Oswin Aichholzer’s homepage [12]. A trick to obtain point sets with different crossing numbers is to use the double circle sets, which are formed by two circles with the same center and with different radius, and the same number of points on each circle (see Figure 3.1(d)). Since the crossing probability of a double circle point set depends on the radius of each circle and no three points of this configuration are collinear, we can get easily point sets with distinct crossing probabilities. Some examples that we used for this part: a 25-point set (Figure 3.1(a)) and a 100-point set (Figure 3.1(b)) with minimum known crossing number and a 100-point set with higher crossing number (Figure 3.1(c)) and another special case of a double circle (Figure 3.1(d)) which has a crossing probability \( P_c \approx 1 \). For each method we repeat the experiment 5000 times and take the mean value to get the approximated solutions to compare with the theoretical results.

3.1. Method I — Points

In the theoretical part of this method we have drawn the conclusion that the variance of the crossing probability \( P_{cm} \) for \( K \) combinations depends on the real crossing probability \( P_c \), not the number of points \( n \) of the point set (Eq.2.7):

\[
Var(P_{cm}) = \frac{P_c(1-P_c)}{K}
\]

We test four point sets:

- **A**) a 25-point set with 4430 crossings, \( P_{cA} = \frac{4430}{\binom{25}{4}} = 0.3502 \)
- **B**) a 100-point set with 1463457 crossings, \( P_{cB} = 0.3732 \)
- **C**) a 100-point set with 2739011 crossings, \( P_{cC} = 0.6985 \)
- **D**) a 100-point set (double circle-DC) with 3879575 crossings, \( P_{cD} = 0.9894 \)

From Figure 3.2 we can see clearly the variance difference between A, B, C compared with D. The 25-point set (A) has similar crossing probability as the 100-point set (B). Their variances are almost the same despite the difference of the number of points. However, though B, C and D have the same number of points, they behave quite differently due to the difference of crossing probability.

The required sample size to reach a certain precision in this case is also related to the crossing probability and it is proportional to the variance. Obviously, when the variance is less, the required sample size is less. The exact sample size \( K \) can be estimated by apply-
Monte Carlo Methods for the Rectilinear Crossing Number

Figure 3.1: Four examples of point sets used for the experiments

Due to the lack of information of the crossing probability before calculation, we can ensure that the calculated crossing probability has a 95% reliability that the crossing probability error is \( \epsilon_{P_c} \leq 0.01 \) by using the maximum sample size \( K = 9506 \), see Figure 2.2.

Before testing the mentioned point sets we can estimate the expected sample size to have \( P(\epsilon_{P_c} \leq 0.01) = 95\% \) of each one, as explained in Section 2.1.2.

\[
\frac{0.01 \sqrt{K}}{\sqrt{P_c(1-P_c)}} = 1.95 \Rightarrow
\]

- A) a 25-point set with 4430 crossings, \( P_{cA} = 0.3502 \Rightarrow K \approx 8653 \)
- B) a 100-point set with 1463457 crossings, \( P_{cB} = 0.3732 \Rightarrow K \approx 8894 \)
- C) a 100-point set with 2739011 crossings, \( P_{cC} = 0.6985 \Rightarrow K \approx 8007 \)
- D) a 100-point set with 3879575 crossings (double circle with \( r_1=99, r_2=100 \)), \( P_{cD} = 0.9894 \Rightarrow K \approx 399 \)
CHAPTER 3. EXPERIMENTAL RESULTS

27

(a) 25-point set with \( P_c = 0.3502 \) and 100-point set with \( P_c = 0.3732 \)

(b) Variances of all the point sets

Figure 3.2: Experimental variances for different point sets by extracting the mean of 5000 experiments

After testing these four point sets, in Figure 3.3 we get the experimental evolution of exactitude- the percentage of reliability that \( \varepsilon_{P_c} \leq 0.01 \). For instance, the value 0.95 on the \( y \)-axis indicates that in 95% out of 5000 experiments, the error \( |P_{cm} - P_c| \leq 0.01 \)

Figure 3.3: Percentage of \( \varepsilon_{P_c} \leq 0.01 \) in function of the sample size \( K \) for 5000 experiments

As we can see when \( K \approx 8700 \) all the point sets have accomplished the condition. The exactitude for point set \( D \) is sharply rising at the beginning and after \( K \approx 1000 \), the percentage of \( \varepsilon_{P_c} \leq 0.01 \) is hundred percent. The growth for point set \( C \) is between \( A/B \) and \( D \), which corresponds to the theoretical result. Figure 3.4 shows the difference of the resulting distributions under different sample sizes.
Monte Carlo Methods for the Rectilinear Crossing Number

3.2. Method II — Segments

To prove the method works, we choose three point sets as the examples:

A) a 25-point set with 4430 crossings, \( P_{cA} = \frac{4430}{\binom{25}{4}} = 0.3502 \)

B) a 25-point set with 8832 crossings, \( P_{cB} = 0.6982 \)

C) a 100-point set with 1463457 crossings, \( P_{cC} = 0.3732 \)

The graphics of reliability are shown in Figure 3.5 and the variance evolution is shown in Appendix Figure D.2

There are only little differences in the variances and in the percentages since \( \text{Var}(P_{cm}) = \frac{9 - P_c^2}{K} \) with \( P_c \in (0, 1) \) does not change a lot. The reliability percentage of \( \varepsilon_{P_c} \leq 0.06 \) tends to 95% when \( K \) tends to 9506. However, the percentage of \( \varepsilon_{P_c} \leq 0.01 \) is growing slowly as we explained in the theoretical part.

3.3. Method III — Circles

Using the same idea as for Method II, we tested the mentioned three point sets and get the variance data and the percentage of reliability (note that in this case we cannot use the double circle point sets since there are four cocircular points). The results are similar because \( P_c \in (0, 1) \) has small influence on \( \text{Var}(P_{cm}) = \frac{(3 - P_c)(1 + P_c)}{K} \) when \( K \) is the same for the three point sets, see Figure 3.6 and Figure D.3 in Appendix.
The three point sets have more or less the same growth. With $K \approx 9506$ all the point sets reach the 95% reliability of $\varepsilon_{P_c} \leq 0.04$. But for $\varepsilon_{P_c} \leq 0.01$ it takes much more time to reach the same precision.

### 3.4. Method IV — Downsizing

To find the best combination of $\lambda$ and $\eta$ for the same number of experiments $K = \binom{n}{3} \lambda$, we choose a sample size $K = 9450$, then we get Table 3.1.
Table 3.1: \( \lambda \) and \( \eta \) of \( K = 9450 \)

<table>
<thead>
<tr>
<th>( \eta )</th>
<th>10</th>
<th>9</th>
<th>8</th>
<th>7</th>
<th>6</th>
<th>5</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \lambda )</td>
<td>45</td>
<td>75</td>
<td>135</td>
<td>270</td>
<td>630</td>
<td>1890</td>
<td>9450</td>
</tr>
</tbody>
</table>

For different combinations of \( \lambda \) and \( \eta \) we get the Figure 3.7

![Figure 3.7: The variances and reliability percentages of different \( \eta \) with \( K = 9450 \)](image)

The result shows that when \( \eta \) is smaller, the precision is higher. Hence, this effect affirms that Method I is the most optimal one for Downsizing because it uses \( \eta = 4 \) which is the lowest value.

3.5. Method V — \( j \)-edges

From the theoretical part we know the variance of crossing number is related with \( \text{Var}(J^2) \), see Eq. 2.35. The variance can be affected by \( j \) and \( e_j \). We tested two 25-point sets with different crossing probabilities and two 100-point sets which have the same crossing probability as one of the 25-point sets. All the point sets have distinct values of \( e_j \).

From Figure 3.8 and Figure D.4 of appendix we can see that the two 25-point sets have a similar performance and so do the two 100-point sets. We can interpret this as the point sets with same number of points have the same variance and exactitude in this method and the crossing probability does not affect the performance. Another surmise from the figure is that when the number of points is larger, the variance is also larger.

3.6. Method VI — \( k \)-InsideCircle

We showed in Section 2.6.1. that the variance and corresponding sample size not only depend on the crossing probability \( P_c \), but also the number of points \( n \). According to the
CHAPTER 3. EXPERIMENTAL RESULTS

Figure 3.8: Reliability percentage for 5000 experiments of Method V

The experimental result agree with the previous analysis. And the reliability of $\varepsilon_{P_c} \leq 0.01$ is inversely proportional with the variance, which is logical- when the variance is smaller, the reliability is higher.

3.7. Comparison

We can compare all the variances and the reliability of crossing probability after obtaining the experimental results of all the methods. To give an example, when $K = 9000$, the average variance, reliability and running time in seconds in MATLAB (with the codes of the explained equations of each method), of the 100-point set with 1463457 crossings by each method after taking the average value of 5000 experiments are shown in Table 3.2.
Table 3.2: Variance of crossing probability, reliability of $\varepsilon_{P_c} \leq 0.01$, running time in seconds of 5000 experiments in MATLAB of each method for $K = 9000$ and the required sample sizes to reach the approximation exactitude

*The estimation of the required sample size of Method V is deduced from experimental data. The approximate polynomial function is: $f(K) = 10^{-12} K^3 - 3 \cdot 10^{-8} K^2 + 0.0002K + 0.1229$, $R^2 = 0.9932$, where $f(K)$ is the reliability percentage and $R^2$ is the coefficient of determination, is the percentage of the variance of the outcome variable explained by the predictor variable. $R^2$ tells you how well the line fits your points. It is 0 when it is the worst case and in the contrary, the best case is $R^2 = 1$.

We can see the best option to test this point set is Method I which has the least variance and highest reliability. However, we cannot say it is the most optimal one because this method only depends on the real crossing probability whereas Method V and Method VI have other factors that can affect the outcome. The running times of Method V and Method VI are quite larger than the others since for the $j$-edges we have to count all the points for each random segment and the same happens in the $k$-InsideCircle for each random circle. This is the main disadvantage of these two methods. Method II and Method III are less time-consuming, however, they require a large number of samples to reach the predetermined approximation exactitude.
CHAPTER 4. IMPROVEMENTS

In this chapter we apply some variance reduction techniques to enhance the performance of the Monte Carlos methods mentioned in Chapter 2. Different techniques admit distinct reductions and also, the required sample size for each method will change due to the variance reduction.

4.1. Importance Sampling

As explained in [13], importance sampling is a variance reduction technique that can be used in the Monte Carlo method. The idea behind importance sampling is that certain values of the input random variables in a simulation have more impact on the parameter being estimated than others. If these "important" values are emphasized by sampling more frequently, then the estimator variance can be reduced. Hence, the basic methodology in importance sampling is to choose a distribution which gives prominence to the important values.

The fundamental issue in implementing importance sampling simulation is the choice of the biased distribution which encourages the important regions of the input variables. Choosing or designing a good biased distribution is the "art" of importance sampling. The rewards for a good distribution can be huge run-time savings; the penalty for a bad distribution can be longer run times than for a general Monte Carlo simulation without importance sampling, [13].

In our case, in Method I we tested four random points with equal probability $K$ times. Applying importance sampling, we can assign four random points a weight: for example, $\frac{1}{4}$ for each point if there exists a crossing formed by them. If they do not form a crossing, the points do not have weight (the assigned weight is 0). After testing $K$ times we get $X$ crossings, and obtain the vector of weight of an $n$-point $P_p$. The entries of the vector $P_p$ are the weights of the corresponding points, thus, the crossing probability of each point is the corresponding entry in the vector of weight divided by the crossing number:

$$P_u = \frac{P_p}{X}$$  \hspace{1cm} (4.1)

$P_u$ is the estimated vector of the crossing probability or 'importance' of each point. Once we got $P_u$, we do the test again with the obtained probability distribution. The calculated crossing number should be closer to the real crossing number. Nevertheless, due to the large number of points and crossings, the importance difference between each point is almost negligible, see Figure 4.1. In conclusion, importance sampling does not give an improvement in this case although, in general, it is the most common and effective variance reduction technique.
4.2. Antithetic Variates

Another variance reduction technique is Antithetic Variates [6]. Suppose that we would like to estimate

$$\theta = E(X) = E(Y) \quad (4.2)$$

Where $X$ and $Y$ are identically distributed random variables with mean $\theta$. For that we have generated two samples $X$ and $Y$, an unbiased estimator of $\theta$ is given by

$$\hat{\theta} = \frac{\hat{\theta}_1 + \hat{\theta}_2}{2} \quad (4.3)$$

with estimator of $X$: $\hat{\theta}_1$ and estimator of $Y$: $\hat{\theta}_2$. And

$$\text{Var}(\hat{\theta}) = \frac{\text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X,Y)}{4} \quad (4.4)$$

In the case where $X$ and $Y$ are independent and identically distributed, the covariance is zero and $\text{Var}(X) = \text{Var}(Y)$ therefore

$$\text{Var}(\hat{\theta}) = \frac{\text{Var}(X) + \text{Var}(Y)}{2} = \frac{\text{Var}(Y)}{2} \quad (4.5)$$

The antithetic variates technique consists in this case of choosing the second sample in such a way that $X$ and $Y$ are not independent and identically distributed anymore and $\text{Cov}(X,Y)$ is negative. As a result, $\text{Var}(\hat{\theta})$ is reduced and is smaller than the previous variance $\text{Var}(X)$, [6].

4.2.1. Implementation

In our case, we used this technique to reduce the variance of Method II – Segments by defining $X$ and $Y$ as Bernoulli random variables:

$$X = \begin{cases} 
1 & \text{if points } A \text{ and } B \text{ are on the same side of segment } CD \\
0 & \text{otherwise}
\end{cases}$$
\[ Y = \begin{cases} 1 & \text{if points } C \text{ and } D \text{ are on the same side of segment } AB \\ 0 & \text{otherwise} \end{cases} \]

Both \( X \) and \( Y \) have the same mean (Eq. 2.8):

\[ E(X) = E(Y) = P_s = \frac{1}{2} + \frac{1}{6} \cdot P_c \tag{4.6} \]

and the same variance

\[ Var(X) = Var(Y) = E(X^2) - E(X)^2 = \frac{1}{2} + \frac{1}{6} \cdot P_c - \left( \frac{1}{2} + \frac{1}{6} \cdot P_c \right)^2 \tag{4.7} \]

and the covariance is

\[ Cov(X,Y) = E(XY) - E(X)E(Y) = \frac{4}{6}P_c - \left( \frac{1}{2} + \frac{1}{6} \cdot P_c \right)^2 \tag{4.8} \]

where \( E(XY) = 1 \cdot P_{(X=1\&Y=1)} = P_{(0)}. \) Note that \( P_{(0)} \) is the probability that the following condition holds: two points which compose a segment are on the same side with respect to other segment which is composed by the remaining two points and vice versa. This condition only can be archived when the polygon is a convex quadrilateral. Figure 4.2 shows that 4 out of 6 segments of a convex quadrilateral meet the condition whereas a concave quadrilateral does not fulfill it.

![Figure 4.2](image.png)

Figure 4.2: Points \( A,B \) are on one side of segment \( CD \) and vice versa (left) whereas a non-convex quadrilateral (right) cannot meet this condition

Consider the random variable \( X' = \frac{X + Y}{2} \). Then we have

\[ E(X') = E(X) \tag{4.9} \]

\[ Var(X') = \frac{Var(X) + Var(Y) + 2Cov(X,Y)}{4} = \frac{2Var(X) + 2Cov(X,Y)}{4} \tag{4.10} \]

if \( Cov(X,Y) < Var(X) \), then \( Var(X') \) will be smaller than \( Var(X) \). Figure 4.3 shows that this is the case. Replacing \( Var(X) \) and \( Cov(X,Y) \) in Eq. 4.10 we get the new expression for the variance of \( X' \):
36 Monte Carlo Methods for the Rectilinear Crossing Number

Figure 4.3: Comparison between $\text{Var}(X)$ and $\text{Cov}(X,Y)$ according to $P_c$

$$Var(X') = \frac{1}{4} P_c \left( 1 - \frac{1}{9} P_c \right)$$

$X'$ only considers four random points.

Now repeat $K$ independent experiments of four random points, $P_{sm}$ is the approximated probability that two randomly chosen points lie on the same side of a segment formed by another two randomly chosen points obtained by Monte Carlo method. Then the variance of the random variable $X = P_{sm}$ is

$$Var(X) = \frac{Var(X')}{K} = \frac{1}{4} P_c \left( 1 - \frac{1}{9} P_c \right) \left( \frac{1}{K} \right)$$

(4.11)

Consider the random variable $P_{cm}$, apply the Eq. 2.8: $P_c = 6 \cdot (P_s - \frac{1}{2})$

$$Var(P_{cm}) = 36 \cdot Var(P_{sm}) = \frac{P_c (9 - P_c)}{K}$$

(4.12)

4.2.2. New Sample Size

The new required sample size can be estimated in replacing the term of the former variance by the new one in Eq. 2.5 (the omitted process will be explained in Section 4.3.2.):

$$\frac{\varepsilon_{P_c} \sqrt{K}}{\sqrt{P_c(1 - P_c)}} = 1.95 \Rightarrow \frac{\varepsilon_{P_c} \sqrt{K}}{\sqrt{P_c(36 - 4P_c)}} = 1.95$$

(4.13)

To reach the 95% reliability of error $\varepsilon_{P_c} \leq 0.01$ the sample size $K$ depends on the crossing probability $P_c$.

$$\frac{0.01 \sqrt{K}}{\sqrt{P_c(9 - P_c)}} = 1.95$$

We can obtain the new sample size of Method II as a function of $P_c$, see Figure 4.4.
4.2.3. Experimental Data

We use the antithetic variates technique to enhance the performance of Method II. Now we can compare the experimental data of the original method and this method combined with the antithetic variates technique, see Figure 4.5 and Figure D.6 of Appendix. Compare this figure to Figure 3.5. The improvement is clearly shown - the variances have decreased about 30% and the reliability of each point set has been increased significantly according to their crossing probabilities.
4.3. Control Variates

The principle of control variates consists in using the estimation error of known quantities, to improve the estimation of unknown quantities \[1\]. Assume the desired simulation quantity is \( \theta = E[X] \); there is another simulation random variable \( Y \) with known expectation \( \mu_Y = E[Y] \). For any \( c \), the new random variable \( Z = X + c(Y - \mu_Y) \), is an unbiased estimator of \( \theta \), because \( E[Z] = E[X] + c(E[Y] - \mu_Y) = \theta \). Now

\[
\text{Var}(Z) = \text{Var}(X + cY) = \text{Var}(X) + c^2 \text{Var}(Y) + 2c \text{Cov}(X, Y)
\]

(4.14)

\( \text{Var}(Z) \) is minimized when \( c = c^* = -\frac{\text{Cov}(X, Y)}{\text{Var}(Y)} \), and

\[
\text{Var}(X + c^*Y) = \text{Var}(X) - \frac{\text{Cov}(X, Y)^2}{\text{Var}(Y)}
\]

(4.15)

\( Y \) is called a control variate for \( X \). In order to reduce variance, choose a \( Y \) correlated with \( X \), \[1\].

4.3.1. Implementation

Given \( K \) samples \( X_1, \ldots, X_K \) and \( K \) samples \( Y_1, \ldots, Y_K \), with means \( \bar{X} \) and \( \bar{Y} \). \( \text{Cov}(X, Y) \) and \( \text{Var}(Y) \) can be estimated from the data:

\[
\text{Cov}(X, Y) \approx \frac{1}{K-1} \sum_{i=1}^{K} (X_i - \bar{X})(Y_i - \bar{Y})
\]

(4.16)

\[
\text{Var}(Y) \approx \frac{1}{K-1} \sum_{i=1}^{K} (Y_i - \bar{Y})^2
\]

(4.17)

so

\[
c^* \approx -\frac{\sum_{i=1}^{K} (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{K} (Y_i - \bar{Y})}
\]

(4.18)

We used this technique to improve the Method I and obtain less variance. Define the random variable \( X = P_{cm} \) and look for a correlated \( Y \) that can reduce the variance as much as possible. There are three variables mentioned in this thesis, which can be defined as \( Y \):

1. \((\ast)\): Choose four points at random, then the Bernoulli random variable \( Y = 1 \) if two of the selected four points lie on the same side with respect to a segment composed by the remaining two points and vice versa.

2. \((\dagger)\): Choose four points at random, then the Bernoulli random variable \( Y = 1 \) if two of the selected four points lie on the same side with respect to a segment composed by the remaining two points.

3. \((\bigcirc)\): Choose four points at random, then the Bernoulli random variable \( Y = 1 \) if the formed disk by three of the selected four points contains the fourth point inside.

The reduction of variance according to Eq. 4.15 is the term \( \text{Cov}(X, Y)^2 / \text{Var}(Y) \). Define \( R = \text{Cov}(X, Y)^2 / \text{Var}(Y) \), thus for the above mentioned three cases we have to compare their reduction values \( R \).
In first case, we define $Y = (\text{(i)})$, hence

\[
X = \begin{cases} 
1 & \text{if the chosen four points form a convex quadrilateral} \\
0 & \text{otherwise}
\end{cases}
\]

\[
Y = \begin{cases} 
1 & \text{if two points lie on the same side of the line joining two other points, vice versa} \\
0 & \text{otherwise}
\end{cases}
\]

The mean of $X$ is:

\[
E(X) = P(\square) = P_c \quad (4.19)
\]

and the mean of $Y$ can be calculated as follows:

\[
E(Y) = 1 \cdot P(\text{(i)}) + 0 \cdot (1 - P(\text{(i)})) \\
= 1 \cdot [P(\square) \cdot P(\square) + P(\text{(i)} \triangleright) \cdot P(\triangleright)] \\
= 1 \cdot \left[ \frac{4}{6} \cdot P(\square) + 0 \cdot P(\triangleright) \right] \\
= \frac{4}{6} \cdot P(\square)
\]

Hence we get the final expression for the mean:

\[
E(Y) = \frac{2}{3} P_c = P_y \quad (4.20)
\]

Now we can estimate the covariance $Cov(X,Y)$:

\[
Cov(X,Y) = E(XY) - E(X)E(Y) \\
= 1 \cdot 1 \cdot P(X = 1, Y = 1) - P_c \cdot P_y \\
= \frac{4}{6} P_c - P_c \cdot \frac{2}{3} P_c
\]

The final expression of the covariance of the first case:

\[
Cov(X,Y) = \frac{2}{3} P_c - \frac{2}{3} P_c^2 = \frac{2}{3} P_c (1 - P_c) \quad (4.21)
\]

To get the reduction term, we still need to calculate $Var(Y)$:

\[
Var(Y) = E(Y^2) - [E(Y)]^2 = 1^2 \cdot P_y - P_y^2 = \frac{2}{3} P_c - \frac{4}{9} P_c^2 = \frac{2}{3} P_c \left( 1 - \frac{2}{3} P_c \right) \quad (4.22)
\]

Now the variance reduction $R_1$ has been found:

\[
R_1 = \frac{Cov(X,Y)^2}{Var(Y)} = \frac{\left( \frac{2}{3} P_c (1 - P_c) \right)^2}{\frac{2}{3} P_c \left( 1 - \frac{2}{3} P_c \right)} = \frac{2}{3} P_c (1 - P_c)^2 \quad (4.23)
\]

Repeat these calculation steps to obtain $R_2$ and $R_3$ for the other two ways of defining $Y$:

\[
R_2 = \frac{P_c^2 - 2P_c^3 + P_c^4}{9 - P_c^2} \quad (4.24)
\]
\[ R_3 = \frac{(P_c - P_c^2)^2}{3 + 2P_c - P_c^2} \] (4.25)

All the reductions depend on the crossing probability as shown in Figure 4.6.

### 4.3.2. New Sample Size

Because of the reduction of variance, the required sample size to reach the same exactitude (95% reliability, with error \( \varepsilon_{P_c} \leq 0.01 \)) will be smaller. Previously, the variance of crossing probability of Method I for one experiment was Eq. 2.7: \( \text{Var}(X)_{\text{old}} = P_c(1 - P_c) \).

The new variance of crossing probability for one experiment is the old one minus the variance reduction \( \text{Var}(X)_{\text{new}} = \text{Var}(X)_{\text{old}} - R \), for example, in the case of \( R_1 \):

\[ \text{Var}(X_{\text{new}}) = P_c(1 - P_c) - \frac{2}{3}P_c(1 - P_c)^2 \] (4.26)

The new required sample size can be estimated in the same way as Section 2.1.2. in Chapter 2. The only difference is in replacing the term of the old variance by the new one in Eq. 2.5:

\[ \frac{\varepsilon_{P_c} \sqrt{K}}{\sqrt{P_c(1 - P_c)}} = 1.95 \Rightarrow \frac{\varepsilon_{P_c} \sqrt{K}}{P_c(1 - P_c) - \frac{2}{3}P_c(1 - P_c)^2} = 1.95 \] (4.27)

Thus, the relation between \( \varepsilon_{P_c} \), \( P_c \), and \( K \) has changed. The new graphics which depict this relation are shown in Figure 4.7. We can compare this figure with the old one - Figure 2.1. We can see clearly the maximum sample size has reduced asymmetrically according to \( P_c \).

In Method I Figure 2.2 we found that the maximum required sample size to ensure a 95% confidence interval for \( \varepsilon_{P_c} = 0.01 \) is 9506, which is attained when \( P_c = 0.5 \). After applying control variates, the new sample size is \( K = 5094 \) which is achieved when \( P_c = 0.634 \), see Figure 4.8.

### 4.3.3. Experimental Data

To use the control variates technique, the mean of \( Y \): \( \mu_Y = E[Y] \) must be known. We will experiment the above-mentioned case: Method I- with variance reduction \( R_1 \) so that \( E(X) = P_c \) and \( E(Y) = P_{(i)} \). The expectation of \( Y \) is unknown at the beginning, hence, we will use the same methodology as explained Method I- do 9506 tests to get the approximate \( P_{(i)} \). We can ensure that the obtained value lies in the confidence interval \([P_{(i)} - 0.01, P_{(i)} + 0.01]\) with confidence level 95%. By applying this approximate value of \( E(Y) \) there is a significant enhancement of Method I, see Figure 4.9 and Figure D.7 of Appendix. The new variances of the point sets that we tested before have decreased almost 50% and now, to reach with reliability 95% a maximum error \( \varepsilon_{P_c} \) of at most 0.01, a
sample size $K \approx 5200$ is enough. The sample size for Method I has been decreased by using this technique, however, we have to test 9506 samples firstly to get the expectation $E(Y)$. Although this is a weak point of this improvement, it does not take much time to test 9506 samples of $Y^{(1)}$ compared to Method I.
Figure 4.6: Reductions of variance by using control variates

(a) Variance reduction of $R_1$

(b) Variance reduction of $R_2$

(c) Variance reduction of $R_3$
Figure 4.7: New relation between probability error ($\varepsilon_{P_c}$), required sample size ($K$) and crossing probability ($P_c$) of Method I by applying control variates with variance reduction R1.
Figure 4.8: New relation between required sample size ($K$) and crossing probability ($P_c$) of Method I when $\varepsilon_{P_c} = 0.01$ by applying control variates ($Y = (ii)$).

Figure 4.9: New reliability percentages of Method I by applying control variates.
CONCLUSIONS

Due to the large time complexity of the rectilinear crossing number calculation problem, this thesis is dedicated to find a new way to calculate the rectilinear crossing number with a new idea- not the exact number, but an approximate value whose range of error is controllable. Thence we thought of Monte Carlo methods, whose outputs follow a normal distribution so that the exactitude can be determined by the variance. Monte Carlo methods are often used in physical and mathematical problems. Their essential idea is using randomness to solve problems that might be deterministic in principle. The found algorithms might be useful to find new large point sets with small rectilinear crossing number.

In this thesis we have introduced six Monte Carlo methods to calculate an approximate value of the rectilinear crossing number. In order to find the best one we compared the variance and the required sample size of each method, and also the necessary running time to arrive at the solution. That is why we put the table of comparison at the end of Chapter 3 with these factors (Table 3.2).

For each method we have deduced the variance and the sample size which is related to the predetermined error range, they are interconnected, which is logical: when the variance is smaller, the required sample size is also smaller. There are two variables related with the exactitude of the approximation: reliability percentage and crossing probability error. Although we have always used 95% reliability and $\varepsilon_{Pc} \leq 0.01$ for this project, it is easy to adapt the calculations for different reliability and error $\varepsilon_{Pc}$. The best one of the six methods is Method I – Points, which requires at most 9506 samples to reach the predetermined precision for any point set.

The required sample sizes of Method II and Method III are much higher than the others although the running time is faster. It turns out that for Method IV, the best size of a downsizing point set is a 4-point set which affirms that Method I gives the most optimal result compared to other downsizing methods. Method V and Method VI are quite time-consuming. The variance of Method VI depends on the crossing probability, and the number of points when the point set is small. The variance of Method V depends on the crossing probability of the point set and on $E(J^2)$ which we cannot deduce so far.

After applying three variance reduction techniques (importance sampling, antithetic variates and control variates) that we used for some methods to enhance the performances, the control variates technique is the most optimal one. Though we have to do one more test (to estimate $E(Y)$, which can be done fastly) before using this technique, it has reduced the required sample size for Method I. The second technique- antithetic variates can be used in Method II which can reduce variance about 30%. However, the required sample size is still large compared to Method I. The importance sampling technique is the most common variance reduction technique for samples which have distinct ‘importance’. In the case of crossing probability, there is few importance discrimination among the points of a point set (each point roughly contributes the same weight for the crossing number of the set). Thus, importance sampling does not improve the performance significantly.

There are several unresolved problems in this thesis which could be studied in the fu-
ture. First, the variance and the sample size of Method V – j-edges. The mean \( E(J^2) \) is unknown so far. Second, the variance and the sample size of the improvement of Method VI – k-InsideCircle by using control variates technique are unknown, which is mentioned in Appendix A. We did not get the numerical expression of the term \( kC_k \) for \( k \in \left[ \frac{n-3}{2} - A, \frac{n-3}{2} + A \right] \). Moreover, the influence of \( A \) and \( n \) to the variance cannot be estimated as yet.

We do not rule out that there exist other possible ways to apply control variates for the presented methods and also other possible Monte Carlo methods that could be applied for crossing number approximation. But, the best known method for calculation of an approximate value of the rectilinear crossing number in this thesis, is Method I – Points combined with a control variate-two segments. It required only 5094 samples, after precalculation of \( E(Y) \), to meet the condition 95% reliability when \( \varepsilon_{P_c} \leq 0.01 \).
BIBLIOGRAPHY


APPENDICES
APPENDIX A. SPECIAL CASE — METHOD VI
IMPROVEMENT WITH CONTROL VARIATES

Another example of applying control variates is reduction of variance of Method VI. Using circles defined by three points, we find an interesting pattern, see [7]: A point set \( S \) in the plane is in general position if no three points of the set lie on a common line, and no four points of the set lie on a common circle. Throughout, all considered point sets will be in general position in the plane and the number of points \(|S| = n\). We denote circles defined by three points of a set \( S \) as circumcircles. Denote with \( C_k \) the number of circumcircles containing exactly \( k \) points of \( S \) in its interior. Then, see [5, 14]

\[
C_k + C_{n-k-3} = 2(k+1)(n-2-k) \tag{A.1}
\]

Note that this expression only depends on \( n \) but not on the positions of the points of \( S \).

Remember the random variable \( X = k \), where \( k \) is the number of points inside the selected circle defined by three random points. Then \( 0 \leq k \leq n-3 \). According to this property of \( C_k \) we propose a random variable \( Y \):

\[
Y = \begin{cases} 
1 & \text{if } \frac{n-3}{2} - A \leq X \leq \frac{n-3}{2} + A \\
0 & \text{otherwise}
\end{cases}
\]

For some fixed number \( A \in \left[0, \frac{n-3}{2}\right] \) to be determined.

The mean of \( Y \) can be calculated thanks to the Eq. A.1.

\[
E(Y) = 1 \cdot P(Y = 1) + 0 \cdot P(Y = 0) = 1 \cdot P(\text{circle contains between } \frac{n-3}{2} - A \text{ and } \frac{n-3}{2} + A \text{ points})
\]

This probability can be estimated, and hence:

\[
E(Y) = 1 \cdot \sum_{k=\left\lfloor \frac{n-3}{2} \right\rfloor}^{\left\lceil \frac{n-3}{2} \right\rceil + A} C_k \binom{n}{3} \tag{A.2}
\]

- When \( n \) is an odd number, \( \frac{n-3}{2} \) will be a natural number. Note that when \( k = \frac{n-3}{2} \) then \( n - k - 3 = k \), thus the Eq. A.2 can be rewritten by substituting Eq. A.1:

\[
E(Y) = \sum_{k=\left\lfloor \frac{n-3}{2} \right\rfloor}^{\left\lceil \frac{n-3}{2} \right\rceil + 1} \frac{2(k+1)(n-2-k)}{\binom{n}{3}} + \frac{C_{n-3}}{\binom{n}{3}}
\]

Where for \( k = \frac{n-3}{2} \)
\[ C_k = C_{n-3-k} = (k+1)(n-2-k) = \frac{1}{4}(n-1)^2 \Rightarrow \]

\[ E(Y) = \sum_{k=n-3}^{n-3-1} \frac{2(k+1)(n-2-k)}{\binom{n}{3}} + \frac{1}{4}(n-1)^2 \]

Hence, for an odd \( n \)

\[ E(Y) = \frac{A(-4A^2 - 6A + 3n^2 - 6n + 1)}{(n-2)(n-1)n} + \frac{3(n-1)}{2(n-2)n} \]  

(A.3)

- When \( n \) is an even number, \( \frac{n-3}{2} \) will not be a natural number, the range from \( \frac{n-3}{2} - A \)

to \( \frac{n-3}{2} + A \) actually is to rounded to \( \left[ \frac{n-3}{2} - A + 0.5, \frac{n-3}{2} + A - 0.5 \right] \), which is equivalent to \( \left[ \frac{n-2}{2} - A, \frac{n-4}{2} + A \right] \). Thus, for an even \( n \)

\[ E(Y) = \sum_{k=\frac{n-2}{2}-A}^{\frac{n-4}{2}+A} \frac{C_k}{\binom{n}{3}} \]

Substituting the Eq. A.1 we obtain the expectation of \( Y \):

\[ E(Y) = \sum_{k=\frac{n-2}{2}-A}^{\frac{n-4}{2}+A} \frac{2(k+1)(n-2-k)}{\binom{n}{3}} = \frac{A(-4A^2 + 3n^2 - 6n + 4)}{(n-2)(n-1)n} \]  

(A.4)

and the variance of \( Y \) is

\[ \text{Var}(Y) = E(Y)(1 - E(Y)) \]  

(A.5)

In this case, the mean of \( Y \) is a function \( E(Y) = f(A,n) \). Due to this reason, it is not necessary to do any expriment to get \( E(Y) \), it can be easily calculated if we know \( A \) and \( n \). The covariance \( \text{Cov}(X,Y) \) in this case is not so easy to compute:

\[ \text{Cov}(X,Y) = E(X,Y) - E(X)E(Y) \]

We have calculated \( E(Y) \) in both cases, for even and for odd \( n \), \( E(X) \) has also been estimated, see Method VI, Eq. 2.36. However it is difficult to compute \( E(X,Y) \), because

\[ E(X,Y) = k \cdot 1 \cdot P(X = k, Y = 1) = \sum_{k=\frac{n-3}{2}-A}^{\frac{n-3}{2}+A} k \cdot \frac{C_k}{\binom{n}{3}} \]  

(A.6)

Where \( C_k \) is the number of disks which contain \( k \) points inside. The covariance now can
be expressed as

\[ \text{Cov}(X, Y) = E(X, Y) - E(X)E(Y) = \sum_{k=\frac{n-3}{2}}^{n-3+\Delta} k \cdot \frac{C_k}{n^3} - E(X)E(Y) \quad (A.7) \]

Define \( X' = R \)

\[ \text{Var}(X') = \text{Var}(X) - \frac{\text{Cov}(X, Y)^2}{\text{Var}(Y)} \quad (A.8) \]

Assume our main random variable is \( Z = P_{cm} \). According to Section 2.6.1. of Method VI of Chapter 2. Then the variance of \( Z \) is

\[ \text{Var}(Z) = \frac{16}{(n-3)^2} \text{Var}(X') \quad (A.9) \]

To calculate the covariance, \( \sum_{k=\frac{n-3}{2}}^{n-3+\Delta} k \cdot C_k \) has to be known. Nevertheless, we don’t have this number previously. Thus we can not estimate \( \text{Var}(Z) \) theoretically with the current information. It would be possible to give theoretical upper and lower bounds for \( \text{Var}(Z) \).
APPENDIX B. POINT SETS USED FOR EXPERIMENTAL RESULTS

A) 25-point set with 4430 rectilinear crossings:
points=[59 152;83 73;131 133;82 146;49;178 107;36 218;49 21;243 109;27 228;43 11;279 102;0 289;39 3;301 98;5 278;38 0;274 103;7 274];

B) 25-point set with 8832 rectilinear crossings:
points=[3276 1092;2315 3474;2786 1229;3713 3167;2570 4797;604 3629;1339 4843;4618 3806;2943 2622;2692 2230;4677 3186;1276 1533;4549 2212;2779 3318;3684 4100;3412 1591;697 3275;4851 2519;2116 785;380 1520;538 524;4841 4335;2840 3380;2408 4116;3210 4708];

C) 100-point set with 1463457 rectilinear crossings:
points=[989381 1837095;1187324 1969875;824102 2483181;1732933 1906588;1262132 1874303;1110706 1853439;587631 2834785;2352711 1931324;937028 2187921;945051 2168383;888728 1533737;728240 1304843;1095943 2086966;991271 1840950;895553 1835801;766256 1526442;3220082 1935216;1045617 1978636;157837 274086;225786 3453867;2774503 1933237;2117441 1933174;1074723 2101479;580062 1004391;1263376 1875725;2931426 1934032;366143 634696;1086898 2095208;390953 3170588;463072 804589;801227 2516946;961930 1618711;473953 760570;938791 2184671;2317467 1930012;1263027 1875418;465206 80320;789997 2533805;2125669 1933130;1260710 1873423;1278429 1848417;853262 2440128;3815182 1937663;1045601 1978595;156331;1938026;1397112 1957730;239732 1933006;2906 38247;841328 1918477;878713 2329744;948277 2160481;482972 1874423;591683 2826289;1110531;1853318;946320 2165284;989948 1832844;1086887 2095178;1702187 1904620;1408682 1956095;1393696 1958135;989128 1830668;567693 983304;885898 2321616;593341 2823445;987803;1814694;1410763 1956030;2133242 1933130;1110681 1853346;1795427 1910833;1794335 1910760;584643 1012284;988295 1660302;1187072 1970194;1371928 1960753;2340058;1930843;1090900 2093443;888768 1533783;567747 2867354;3835022 1937735;495512;2996564;785862 1421803;963195 1628342;523066 2931470;32319 3785689;444093 3079441;426967 74100;557663 966163;761355 1371943;2173089 1932791;337120 3262918;0 0;455166;3061781;730125 1308388;33305 57754;2556868 1933141;1074828 2101425;1090882 2093451];

D) 100-point set with 2739011 rectilinear crossings:
points=[8223 5832;4484 5354;2845 9020;4177 4017;6393 2783;5702 811;9301 6166;2983 6052;9481 707;6100 5934;6463 3608;7203 62;7500 3915;6750 3731;1815 1004;6963 1205;5725 8577;5486 3173;9499 2090;2962 8919;4013 2521;3913 738;2400 177;4210 3434;275 5060;8207 1851;8301 8567;3585 6242;7106 1489;4591 1974;7695 8518;893 2823;3445 1532;4739 4215;6990 4197;3571 783;8184 2230;7548 4124;810 8986;6271 698;2918 2187;7559 4548;9749 5766;2580 5760;2439 4712;1836 990;4659 5697;5010 8630;8991 2377;5169 50;9034 7415;9849 4033;5017 3368;4808 4739;7427 801;3482 8194;1870 9643;672 5197;5412 9239;1393 6942;166 9820;3349 6549;1422 4314;2695 2491;3225 8913;2552 2641;1428 8102;5389 3402;4041 6312;7135 9367;6208 2556;8072 5791;4972 1341;5183 675;201 3946;6960 4966;7947 9223;200 5403;8791 790;555 4253;2539]
\[ E \) 100-point set with 2740640 rectilinear crossings: points=[29062 17670;3531 41947;18712 6695;1680 34906;20306 9517;32895 19839;28759 16946;48791 46129;22079 41914;38885 31471;28468 43266;1055 10813;25608 3245;20373 33892;10989 26749;11520 25860;14212 829;18182 23727;24956 47187;40659 22921;36209 37679;44677 33513;38351 46932;10329 24227;2163 25871;7668 39408;35867 28520;6424 38272;11024 45944;22646 10501;23230 30362;4483 23253;25299 30790;3820 10140;8314 33829;37328 865;16703 11313;15229 45652;40879 38361;49023 26623;48370 3414;28784 9963;11655 600;14581 48247;33631 41539;28638 22227;9968 49438;47825 39644;16778 24387;7470 16389 10593;5645 41479;38524 16159;39553 36869;29298 7979;3952 36503;9627 2680;38694 35649];

\[ F \) Double Circle:

- Matlab function codes:

```matlab
function points= DoubleCircle( r1, r2, NumPoints )

%% Input
% r1: radius of circle 1;
% r2: radius of circle 2;
% NumPoints: number of total points of 2 circles

%% Output
% points: the vector of all the points of 2 circles.

i=NumPoints/2;
for j=1:i
    a(j)=(j-1)*2*pi/i;
    b(j)=2*pi/NumPoints+(j-1)*2*pi/i;
    x1(j)=r1*cos(a(j));
    y1(j)=r1*sin(a(j));
    x2(j)=r2*cos(b(j));
    y2(j)=r2*sin(b(j));
end
points1(:,1)=x1;
points1(:,2)=y1;
points2(:,1)=x2;
points2(:,2)=y2;
for j=1:i
    points(j,1)= points1(j,1);
    points(j,2)= points1(j,2);
end
```
for j=i+1:NumPoints
    points(j,1)= points2(j-i,1);
    points(j,2)= points2(j-i,2);
end
### APPENDIX C. TABLE OF NORMAL DISTRIBUTION Q(X)

<table>
<thead>
<tr>
<th>x</th>
<th>Q(x)</th>
<th>x</th>
<th>Q(x)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>5.00E-01</td>
<td>2.7</td>
<td>3.47E-03</td>
</tr>
<tr>
<td>0.1</td>
<td>4.60E-01</td>
<td>2.8</td>
<td>2.56E-01</td>
</tr>
<tr>
<td>0.2</td>
<td>4.21E-01</td>
<td>2.9</td>
<td>1.87E-03</td>
</tr>
<tr>
<td>0.3</td>
<td>3.82E-01</td>
<td>3.0</td>
<td>1.35E-03</td>
</tr>
<tr>
<td>0.4</td>
<td>3.45E-01</td>
<td>3.1</td>
<td>9.68E-04</td>
</tr>
<tr>
<td>0.5</td>
<td>3.09E-01</td>
<td>3.2</td>
<td>6.87E-04</td>
</tr>
<tr>
<td>0.6</td>
<td>2.74E-01</td>
<td>3.3</td>
<td>4.83E-04</td>
</tr>
<tr>
<td>0.7</td>
<td>2.42E-01</td>
<td>3.4</td>
<td>3.37E-04</td>
</tr>
<tr>
<td>0.8</td>
<td>2.12E-01</td>
<td>3.5</td>
<td>2.33E-04</td>
</tr>
<tr>
<td>0.9</td>
<td>1.84E-01</td>
<td>3.6</td>
<td>1.59E-04</td>
</tr>
<tr>
<td>1.0</td>
<td>1.59E-01</td>
<td>3.7</td>
<td>1.08E-04</td>
</tr>
<tr>
<td>1.1</td>
<td>1.36E-01</td>
<td>3.8</td>
<td>7.24E-05</td>
</tr>
<tr>
<td>1.2</td>
<td>1.15E-01</td>
<td>3.9</td>
<td>4.81E-05</td>
</tr>
<tr>
<td>1.3</td>
<td>9.68E-02</td>
<td>4.0</td>
<td>3.17E-05</td>
</tr>
<tr>
<td>1.4</td>
<td>8.08E-02</td>
<td>4.5</td>
<td>3.40E-06</td>
</tr>
<tr>
<td>1.5</td>
<td>6.68E-02</td>
<td>5.0</td>
<td>2.87E-07</td>
</tr>
<tr>
<td>1.6</td>
<td>5.48E-02</td>
<td>5.5</td>
<td>1.90E-08</td>
</tr>
<tr>
<td>1.7</td>
<td>4.46E-02</td>
<td>6.0</td>
<td>9.87E-10</td>
</tr>
<tr>
<td>1.8</td>
<td>3.59E-02</td>
<td>6.05</td>
<td>4.02E-11</td>
</tr>
<tr>
<td>1.9</td>
<td>2.87E-02</td>
<td>7.0</td>
<td>1.28E-12</td>
</tr>
<tr>
<td>2.0</td>
<td>2.28E-02</td>
<td>7.5</td>
<td>3.19E-14</td>
</tr>
<tr>
<td>2.1</td>
<td>1.79E-02</td>
<td>8.0</td>
<td>6.22E-16</td>
</tr>
<tr>
<td>2.2</td>
<td>1.39E-02</td>
<td>8.5</td>
<td>9.48E-18</td>
</tr>
<tr>
<td>2.3</td>
<td>1.07E-02</td>
<td>9.0</td>
<td>1.13E-19</td>
</tr>
<tr>
<td>2.4</td>
<td>8.20E-03</td>
<td>9.5</td>
<td>1.05E-21</td>
</tr>
<tr>
<td>2.5</td>
<td>6.21E-03</td>
<td>10.0</td>
<td>7.62E-24</td>
</tr>
</tbody>
</table>

Table C.1: Table of Normal Distribution \( Q(x) = P(N(0, 1) > x) \)
APPENDIX D. VARIANCES OF THE METHODS

D.0.1. Method I — Points

Figure D.1: Mean variance for 5000 experiments of Method I

D.0.2. Method II — Segments

Figure D.2: Mean variance for 5000 experiments of Method II
D.0.3. Method III — Circles

![Experimental Variance of Method III-Circles](image)

Figure D.3: Mean variance for 5000 experiments of Method III

D.0.4. Method V — j-edges

![Experimental Variance of Method V- j-edges](image)

Figure D.4: Mean variance for 5000 experiments of Method V
D.0.5. **Method VI — $k$-InsideCircle**

Figure D.5: Mean variance for 5000 experiments of Method VI

D.0.6. **Method II with Antithetic Variates**

Figure D.6: Mean variance for 5000 experiments of Method VI by applying antithetic variates technique
D.0.7. Method I with Control Variates

Figure D.7: Mean variance for 5000 experiments of Method I by applying control variates technique