EXPECTED AREA OF THREE POINTS ON A RANDOM WALK IN 2 DIMENSIONS

by

BRIAN R. LAWLER

(Under the Direction of Sybilla K. Beckmann-Kazez)

ABSTRACT

Motivated by the mesmerizing qualities of three animated points in a computer microworld, this thesis explores the average triangular area determined three particles on a random walk on $\mathbb{Z}^2$. Intermediate steps include the study of distance between two walking points in one and two dimensions, and the average area of three random points on $\mathbb{Z}^2$. The distance functions are shown to grow on the order $\sqrt{n}$, while area appears to grow linearly. Relationships between the distance results and area investigation are explored while some curious contradictions between an unbounded and bounded region arise in data collected through modeling the situations. Numerous questions for further study are presented.

INDEX WORDS: Random Walk, Expectation, Geometer's Sketchpad

EXPECTED AREA OF THREE POINTS ON A RANDOM WALK IN 2 DIMENSIONS

by

BRIAN R. LAWLER

B. S. Mathematics, Colorado State University, 1992

M. A. Education, California State University – Dominguez Hills, 1999

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment

of the Requirements for the Degree

MASTER OF ARTS IN MATHEMATICS

ATHENS, GEORGIA

2006

EXPECTED AREA OF THREE POINTS ON A RANDOM WALK IN 2 DIMENSIONS

by

BRIAN R. LAWLER

Major Professor:	Sybilla K. Beckmann-Kazez
Committee:	Malcolm R. Adams Robert Varley

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
August 2006

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION TO THE PROBLEM

Computer microworlds have established themselves in the growing body of mathematics knowledge as well as in mathematical research. While powerful computing tools, such as the user-friendly Maple (Maplesoft, 2005) programming language, or something living much closer to machine-coding, have allowed for tremendous increase in potential for exploration, hypothesis building, etc., another sort of software has allowed for spatial—or geometric—investigations. Chief among these computer microworlds for ease of use is The Geometer's Sketchpad (Jackiw, 2004), an interactive environment used from elementary through secondary schools, and into post-secondary education. This microworld is not driven by line-by-line programming, but instead the user commands the space in the same way a geometer would create with a pencil and paper, straightedge, and compass.

The Geometer's Sketchpad (Jackiw, 2004) is one in a class of mathematics software referred to as an Interactive Geometry software or a Dynamic Geometry Environment. These are computer programs that allow one to create and then manipulate geometric constructions, primarily in plane geometry. This dynamic nature is what sets the software apart from the static experience of paper and pencil, or a simple computer assisted drawing tool. According to its publishers, Geometer's Sketchpad 4 allows the user to create, explore and analyze a wide range of mathematics. One can construct interactive mathematical models ranging from basic investigations about shape and number to advanced, animated illustrations of complex systems.

Technology, and in particular computer software such as that exemplified by this dynamic geometry microworld, has played a significant role in modern mathematics. There is a new potential for computing power that was simply not available in previous centuries. This power for computation has also opened possibility for solving, or often approximating—with a great deal of accuracy—equations that were previously too complex for established algebraic algorithms, or too cumbersome to pursue. The "Monte Carlo method"[1] has long been applied to approximate solutions to a variety of mathematical problems. This name refers to any sort of method which solves a problem by generating appropriate random numbers and then observing what fraction of those numbers obeys a particular property or properties. Modern uses of this method utilize a computer to perform these sampling experiments. Interestingly, the method is applied to problems with no probabilistic content as well as to those with inherent probabilistic structure. For example, Monte Carlo Integration can be used to find the area of a complicated domain *D*. The method picks some random points out of some simple superset *D'* of *D*, checks to see if the point is in *D*, and then estimates the area of *D* as the area of *D'* multiplied by the fraction of points falling within *D*.

Of course, such changes to techniques has led the field to re-examine what constitutes mathematical thought, argument and proof. (Chang, 2004; Knight, 2005; Peterson, 2004). Dr. Thomas C. Hales of the University of Pittsburgh proved the Kepler Conjecture—stated simply that the best way to stack oranges is as a pyramid, with the next higher layer being placed in the hollow of the layer below—by relying on a complex series of computer calculations. These computations were too tedious to be checked by hand, and ultimately the proof required faith that the computations were performed flawlessly. It is widely accepted that a Dynamic Geometry

---

[1] A nice history at http://www.geocities.com/CollegePark/Quad/2435/history.html.

System (DGS) can engage and enhance exploration and aid in the potential to create mental pictures—all necessary to create proof (Aarnes & Knudston, 2003). At odds among mathematicians is whether or not a DGS is limited to only demonstration that something is true (or seems to be true) or is in fact a proof, including the *why* of the apparent truisms (Hanna, 2001).

Although the potentials created by this newly burgeoning computing power in mathematics has created some stir while the field attempts to keep tight reigns on the defining characteristics of what is mathematics?—or more particularly, what is proof?—few argue the potential for technology to open the doorway to new conjectures and to raise new questions. It is from this spirit of exploration that the topic of this paper arose. Although I was familiar with previous iterations of the software, the dynamic environment of version *The Geometer's Sketchpad 4* (Jackiw, 2004) (I will use the common acronym GSP) added some particular elements that were quite new. In particular, a whole new element of exploration was opened by the new implementation of point animation. Previously in GSP, a point could be animated along a path, such as a segment or circle. The point would travel along the path, either constantly in one direction, or bidirectionally. This newest version of GSP allowed for a point (and other objects) to be animated free from a well-defined path (Figure 1), following some kind of random walk.

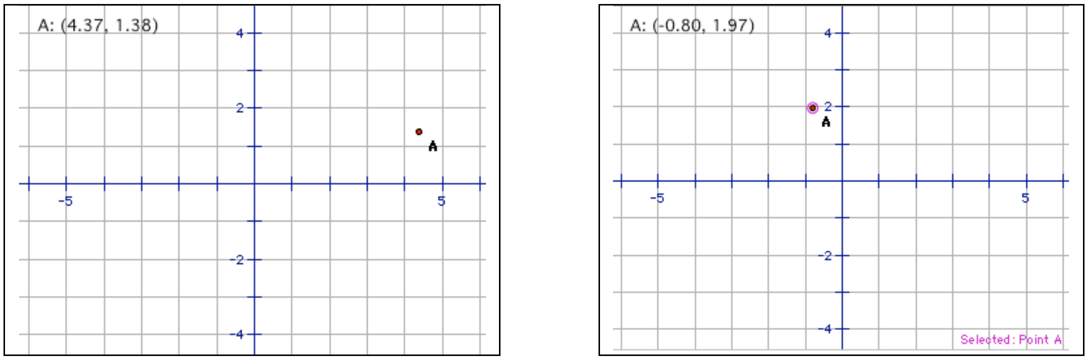


Figure 1. An animated point in Geometer's Sketchpad.

A large number of these points looked like the busy activity of a disturbed fire ant colony, the mixing of two types of molecules in a jar, or the unfurling distribution of an invading party of secret agents into a public gathering. (Figure 2) . Toying around with animated points in this dynamic environment creates a great deal of wonder, but also inspires many mathematical questions. For me, the questions became about three points on this animated walk, and the triangle formed by them. (Figure 3)

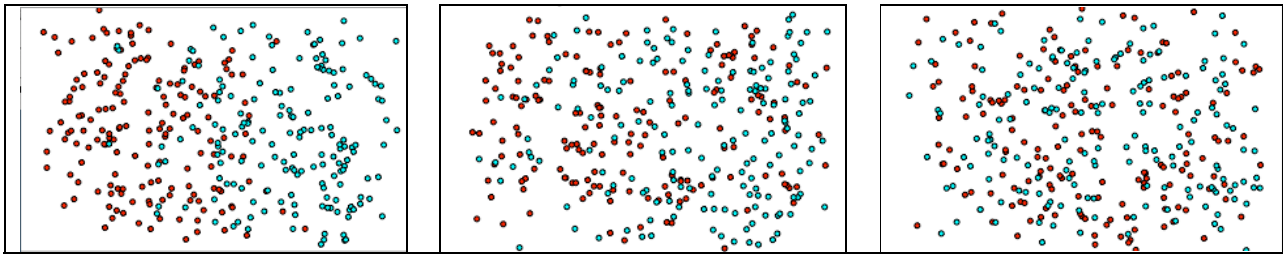


Figure 2. A time-series picture of several animated points.

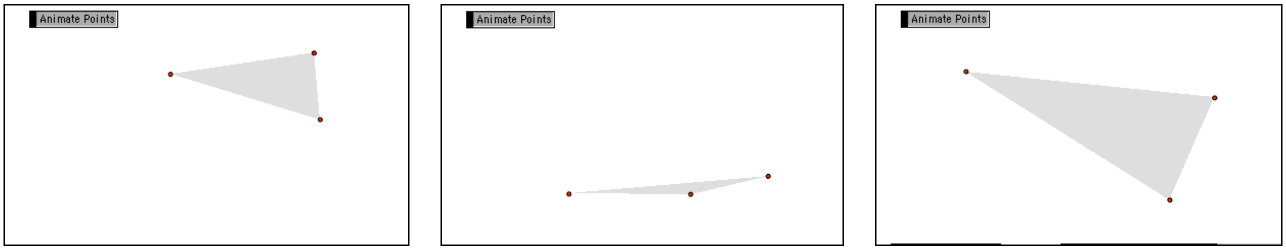


Figure 3. Triangle formed by 3 animated points.

For me, this animated triangle appeared to be a rigid, two-dimensional triangle dancing in 3-space, turning, twisting, wandering away, and drawing closer. It was only perspective that seemingly changed its shape and size. Of course, all points remained in the constrained two-dimensional plane of the GSP window, and the animation yielded this mesmerizing optical trick. But some immediate observations began to emerge. The points could align, even if momentarily, and the triangle would disappear to nothingness. It was actually surprising how often this occurred—in fact, every time one point crossed between the other two. What is the probability that this could occur?—apparently greater than my gut reaction seemed to predict. I also wondered about the size of the triangle. I stretched it in my mind, putting all points at the edges of the GSP window, and realized the triangle could be at most ½ the area of the computer window that I watched these points dance on. I quickly realized that any one of the points could be dragged along an edge of the window and this triangle would remain the same area (Figure 4). However, a proof that the maximum area of the triangle is ½ the area of the window is not immediate obvious. Notice that by changing one perspective, the length of the base could actually be greater than the horizontal dimension of the window. However, that the height of the triangle with this new perspective necessarily diminishes with this increased base, making the observation hold true.

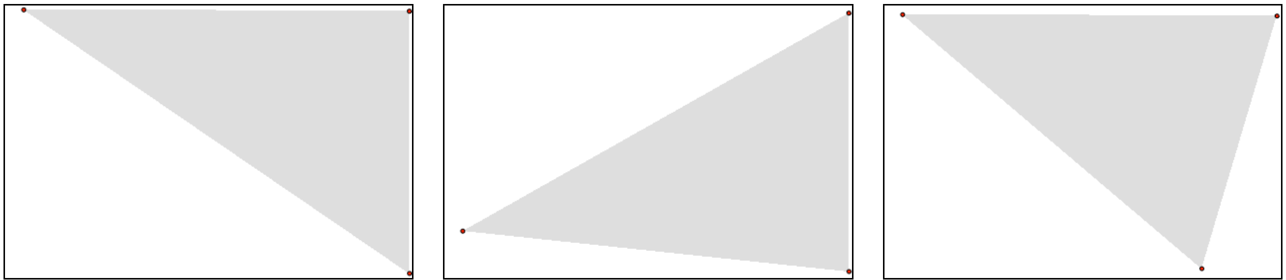


Figure 4. Observations on the animated triangles maximum area.[2]

Proof Area of Random Triangle is One-Half Area of Rectangle

I present a proof of this area relationship that will demonstrate the insight provided by the dynamic environment of GSP. What follows is a geometric proof that the area of a triangle determined by three points *p*, *q*, and *r* randomly chosen inside (or on) an $m \times n$ rectangle must be less than or equal to one-half the area of the $m \times n$ rectangle. Key to the proof is that similarly colored areas are equal (see Figure 5). That is,

$$\text{Area}_{qEA''} = \text{Area}_{qEA'},\ \text{Area}_{qpA} = \text{Area}_{qpA'},\ \text{and Area}_{prE'} = \text{Area}_{prE}.$$

---

[2] Note: The GSP software design makes it difficult to move the points to the window boundary.

The three darker shades ($\text{Area}_{qA'p}$, $\text{Area}_{qA'p}$, and Area_{pEr}) add to equal the area of the triangle Area_{pqr}, while all 6 colors sum to less than the area of the rectangle:

$$\text{Area}_{qEA''} + \text{Area}_{qEA'} + \text{Area}_{qpA} + \text{Area}_{qpA'} + \text{Area}_{prE'} + \text{Area}_{prE} \leq \text{Area}_{ABCD}\,.$$

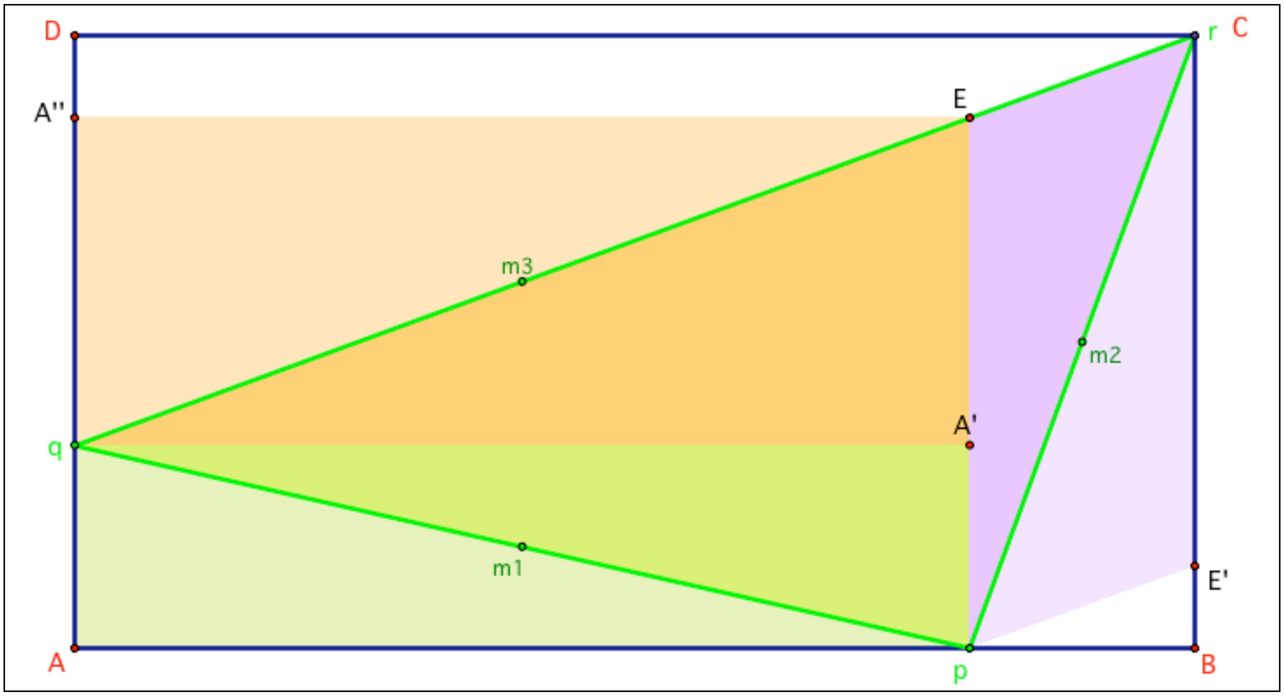


Figure 5. Diagram showing area of triangle less than half bounding rectangle.

Proof: Consider an $m \times n$ rectangle on the $\mathbb{Z}^2$ lattice, in which the upper right vertex is at (m, n), a point in Quadrant I of $\mathbb{Z}^2$ and the lower left vertex is at the origin. Name this region $R_{m\times n}$ for simplicity. Select 3 random points on $R_{m\times n}$, allowing for the possibility any of these points to be on the boundary. Define the rectangle bounding these three points to be *ABCD*, where *A* is at the origin, and the letters proceed alphabetically counterclockwise. Note that $\text{Area}_{ABCD} \leq \text{Area}_{R_{m\times n}}$.

Next I will establish a naming and orientation convention for the triangle formed by the 3 random points in order to simplify the proof. First I will state a method by which, without loss of generality, to rotate, reflect, and/or translate the triangle formed by the three random points until this triangle is oriented in such a manner that

- the point with the largest *y*-coordinate is furthest right, i.e. has the greatest *x*-coordinate;
- the *x*-coordinate of the point with the smallest *y*-coordinate is between the *x*-coordinates of the other two points;
- leaving the point with the middle *y*-coordinate to have the smallest *x*-coordinate.

This process will involve reflecting the triangle across vertical and/or horizontal bisectors of *ABCD*, or rotating about the point of intersection of the diagonals of *ABCD*. If the vertex with the middle x-coordinate lies above the line joining the left-most and right-most vertices, reflect across the horizontal bisector of the bounding rectangle so that the middle vertex now lies below. If the left-most vertex has a higher *y*-coordinate than the right-most vertex, reflect across the vertical bisector of the bounding rectangle so as to reverse that situation. Now the left-most vertex has a lower (or equal) *y*-coordinate than the right-most vertex and the other vertex lies below the line connecting them. If the resulting triangle formed by these three points does not satisfy the 3 bulleted goals, the triangle must be entirely within the right triangle formed by cutting the bounding rectangle diagonally, an area one-half the area of the original bounding rectangle. In this case, the triangle area is already less than or equal to one-half the area of it's boundary rectangle and no further argument is needed.

Throughout this process of transformation, any equality of coordinates is allowable, provided remainder of these orientation rules hold true. An a final note, keep in mind that these transformation rules keep the triangle inside $R_{m\times n}$. This transformed triangle remains congruent and hence of equal area. And this (possibly) new bounding rectangle keeps an area equal to Area_{ABCD}.

Name the point with the smallest *y*-coordinate (and middle *x*-coordinate in case of a tie) to be point *p*. Travel clockwise to name point *q* and then *r*. If the bounding rectangle has changed orientation, rename vertices *ABCD* as originally done with A being the bottom-left coordinate and the other points named counterclockwise. (See Figure X for naming convention.)

Construct *A'* as a 180° rotation of *A* about midpoint *m1* of $\overline{pq}$. *A'* must be on the interior of $\triangle pqr$ since *p* has the middle *x*-coordinate of the three points *p*, *q*, *r* and *q* has the middle *y*-coordinate.

$\triangle pqA' \approx \triangle qpA$ by the side-side-side (SSS) triangle congruence property.

Next, construct *E* as the point at the intersection of $\overline{qr}$ and $\overrightarrow{pA'}$. *E* must exist because the *x*-coordinate of *p* is situated between the *x*-coordinates of *q* and *r*. Construct *E'* as a 180° rotation of *E* about midpoint *m2* of $\overline{pr}$. Since $\overleftrightarrow{pE}$ and $\overleftrightarrow{Br}$ are parallel lines cut by transversal $\overline{pr}$, *E'* must lie on $\overleftrightarrow{Br}$, with $\overline{EE'}$ the minor diagonal of parallelogram *EpE'r*.[3] Therefore $\triangle prE \approx \triangle rpE'$.

Finally, construct *A''* as a 180° rotation of *A'* about midpoint *m3* of $\overline{qE}$. Since $\overleftrightarrow{pE}$ and $\overleftrightarrow{Aq}$ are parallel lines cut by transversal $\overline{qE}$, *A''* must lie on $\overleftrightarrow{Aq}$, with $\overline{A'A''}$ a diagonal of parallelogram *A'EA''q*. Therefore $\triangle qEA' \approx \triangle EqA''$.

To summarize so far, the following congruences are established: $\triangle pqA' \approx \triangle qpA$, $\triangle prE \approx \triangle rpE'$, and $\triangle qEA' \approx \triangle EqA''$. Therefore

$$\text{Area}_{qEA''} = \text{Area}_{qEA'},\ \text{Area}_{qpA} = \text{Area}_{qpA'},\ \text{and Area}_{prE'} = \text{Area}_{prE}.$$

---

[3] Note: I will briefly argue $\overline{EE'}$ must be the minor diagonal by demonstrating $m\angle pE'r \geq 90°$. $m\angle pBR = 90°$ and

The final step to show

$$\text{Area}_{qEA''} + \text{Area}_{qEA'} + \text{Area}_{qpA} + \text{Area}_{qpA'} + \text{Area}_{prE'} + \text{Area}_{prE} \leq \text{Area}_{ABCD}$$

remains. The areas do not overlap, by construction. Further, the areas are restricted to the rectangle, since three areas reside inside $\triangle pqr$ and the other three are defined by points *A*, *E'*, and *A''* on the rectangle. Hence, the sum of the areas of the six triangles must be less than or equal to the area of the bordering rectangle.

By the congruences established, $2 \cdot \text{Area}_{pqr} \leq \text{Area}_{ABCD}$ which is equivalent to $\text{Area}_{pqr} \leq \frac{1}{2}\text{Area}_{ABCD}$. Since $\text{Area}_{ABCD} \leq \text{Area}_{R_{m\times n}}$, the proof is complete.

This observation that the maximum area of the triangle must be ½ the area of the window began to shape the question I pursue in this thesis: What is the average area of that triangle? Does it stay closer to ½, or closer to zero—the previously observed instance of one point crossing the segment created by the other two. As I thought about this question, I began setting some parameters for answering, and noticed additional properties of the situation. First, I realized that I was imagining the question on a general level—that is thinking generally about 3 animated points in two-dimensions. But this GSP microworld environment is constrained by the borders of the window. My observations reflect that: “The maximum area is ½ the area of the window.” What would be my unit of measure for area? Square pixels seemed reasonable. Thinking of the area as some portion of whatever the window size seemed appropriate as well.

As I further explored the possibilities here, I attended a lecture by the software’s author, Nicholas Jackiw (June 2002, Mills College, Oakland, CA). He was speaking about the new features of the 4th version of *Geometer’s Sketchpad*, and in particular about this animation feature. He noted that in their initial programming, these dancing points would occasionally

---

since $rE' \leq rB$, $\angle pE'r$ must be obtuse (or right).

wander off the screen; sometimes not returning. The programmers felt this may be disheartening to a user, and that objects constructed from such points become less interesting when the points leave the screen. They adjusted their purely random algorithm so that as points neared the boundary, they would be discouraged from leaving the screen. A senior programmer in this project, Scott Steketee, replied to my inquiry into this design with the following comments:

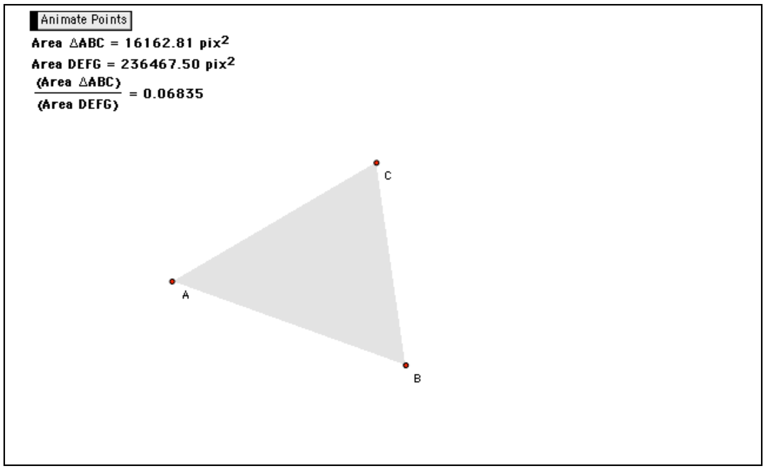
> Our design considerations included several desiderata:
>
> 1. The random motion of an independent point should be purely random over most of the screen.
> 2. Points, and objects constructed based on them, tend to become much less interesting when they leave the screen. Therefore points in random motion should be discouraged from leaving the screen. If they have left the screen they should be encouraged to return.
>
> The simplest way to accomplish these two objectives is to define a “force” that pushes points toward the center of the screen if they are off-screen or near the boundary. The strength of the force is zero when points are safely on-screen, increases gradually when they near the edge, and continues to increase if they have gone off-screen. Once the point is a certain distance outside the boundary, the force reaches and stays at its maximum strength. This fairly simple model works reasonably well to keep the motion observable, though at the cost of true (pseudo-) randomness. (Steketee, May 2, 2006, via email)

Clearly the points were not moving with a true random motion in this computer environment. So my observations were skewed from the theoretical question I had posed.

I felt that I could create a sketch in GSP that would allow me to explore the question further from an empirical standpoint prior to developing a theory-based analysis. I hoped the exploration would provide some insight and direction, as well as a better sense of the problem through which I could consider the reasonableness of theoretical conclusions.

Design of the Computer Testing Grounds

I knew I could use GSP to provide numerical data alongside the visual data from merely watching the dance of the triangle. I programmed GSP to compute the area of the window, in square pixels, the area of the triangle formed by the three points, and the ratio of these areas (Figure 6). Recall that given the software designer’s goal to keep all points onscreen, this ratio of the areas must remain between zero and 0.5. As I watched the ratio adjust in real time during an animation, it appeared to me that 0.11 seemed a fairly reasonable estimate for the expected area of this triangle, expressed as a portion of the enclosing window.



Area $\triangle$ABC = 16162.81 pix^2

Area DEFG = 236467.50 pix^2

$\frac{(\text{Area } \triangle ABC)}{(\text{Area DEFG})} = 0.06835$

Figure 6. Initial investigation of the animated triangle in the GSP microworld.

Other than having established some grounding for the task and a rough estimate as to reasonable values for the area, this exploration yielded the push into considering the mathematics that could be engaged to make greater sense of the problem. These animated points could be considered to be traveling in a process commonly known as a *random walk*.

The Mathematical Theory

As a fun, and simplistic introduction to the notion of a random walk, I will draw upon the imagery of a drunken man, stumbling to and fro. As he travels, imagine that he lives in a city on a square grid. As he arrives at any intersection, he could turn in any direction, north, south, east, or west, and continue his journey. The path of his stumbling may look something like the picture in Figure 7. It is a curious question to ask, will the drunken man ever arrive home?

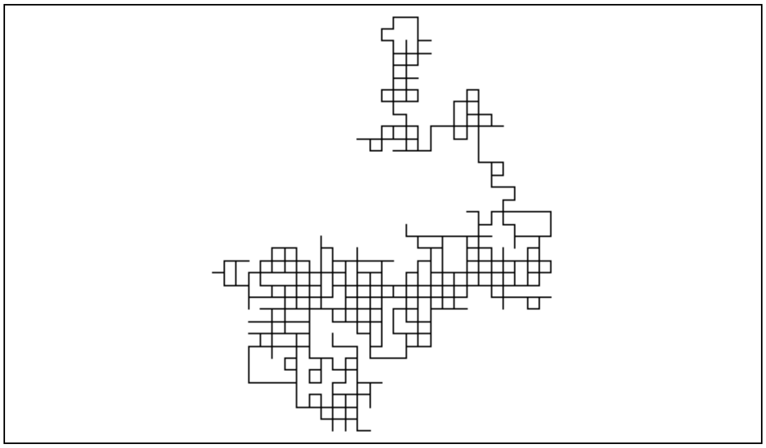


Figure 7. A possible path of the drunken man on a city grid.

In a one-dimensional case, the simplest random walk is a path constructed according to a similar set of rules. The path begins at some point. The step-distance from one point to the next is some constant. And the direction from one point to the next is chosen at random, each with equal probability. In one dimension, this direction is either +1 or –1 unit of distance. An interesting result of such a set of conditions is approached by the question, how many times will this random walk cross some particular distance from the origin? The surprising answer is found in the theorem: every point in the domain of the random walk will be visited an infinite number of times almost surely (Adams & Guillemin, 1996; Spitzer, 1964). A parallel to this theory tells us that the drunk will return home! In fact, this amazing property ends after two dimensions. If the drunk could venture into a 3-dimensional (or greater) world, he may never return home. In fact, it is highly unlikely, of probability zero, that he returns home. In formal terms, all walks on one- and two-dimensions are *recurrent*, while in dimension 3 and above, the walks are *transient* (Spitzer, 1964). Polya proved this in 1921.

It is an interesting note that the theory of the random walk, although possibly less generalized than a theory of Markov chains, is far more complete (Spitzer, 1964). Furthermore, it has tendrils of thought drawing from and leading to number theory, probability theory, measure theory, topology, Fourier analysis, differential equations, analytic functions, and differential and integral operators. This notion of a random walk bridges, and connects, many realms of mathematics.

It will be sufficient for this inquiry to utilize an analytic definition for a *random walk*, the simplest definition. The definition will be motivated by probabilistic ideas, but the definition itself will involve neither probability theory nor measure theory. The problem of the triangle formed by three animated points is a discrete random walk of three points on $\mathbb{Z} \times \mathbb{Z} = \mathbb{Z}^2$, a 2-dimensional set of lattice points—the integer lattice in the plane. Define $\boldsymbol{x} \in \mathbb{Z}^2$ as $\boldsymbol{x} = (x^1, x^2)$. The problem posed is a discrete time parameter stochastic process $\{\boldsymbol{x}_n, n = 0,1,2,...\}$ with $\boldsymbol{x}_n \in \mathbb{Z}^2$, the *state space*. $\mathbb{Z}^2$ is the state space of the stochastic processes discussed in this thesis.

For each pair ***x***, ***y*** in $\mathbb{Z}^2$ I'll define a real number $P(\boldsymbol{x},\boldsymbol{y})$, the *transition function* of the random walk. The value of $P(\boldsymbol{x},\boldsymbol{y})$ is the probability a point starting at ***x*** will move to ***y*** in one time step. $P(\boldsymbol{x},\boldsymbol{y})$ must have the properties

$$\begin{gathered} 0 \le P(\boldsymbol{x},\boldsymbol{y}) = P(\mathbf{0},\boldsymbol{y}-\boldsymbol{x}), \\ \sum_{x\in\mathbb{Z}^2} P(\mathbf{0},\boldsymbol{x}) = 1 \end{gathered} \quad . \tag{1}$$

It is in the most restrictive of the properties above that $P(\boldsymbol{x},\boldsymbol{y}) = P(\mathbf{0},\boldsymbol{y}-\boldsymbol{x})$, where $\boldsymbol{y}-\boldsymbol{x}$ is the point in $\mathbb{Z}^2$ with the coordinates $\boldsymbol{y}^i - \boldsymbol{x}^i$, $i = 1,2$. This property of spatial homogeneity—that spatial properties are the same at every point[4]—shows that the transition function $P(\boldsymbol{x},\boldsymbol{y})$ is really determined by a single function $p(\boldsymbol{x}) = P(\mathbf{0},\boldsymbol{x})$ on $\mathbb{Z}^2$ with the properties

$$\begin{gathered} 0 \le p(\boldsymbol{x}), \\ \sum_{x\in\mathbb{Z}^2} p(\boldsymbol{x}) = 1 \end{gathered} \, .$$

In other words, naming a transition function is equivalent to specifying a probability measure on $\mathbb{Z}^2$ —a non-negative function $p(\boldsymbol{x})$ whose sum over $\mathbb{Z}^2$ is one.

Informally, the transition function $P(\boldsymbol{x},\boldsymbol{y})$ is a random walk. Or more carefully, define a *random walk as a function* $P(\boldsymbol{x},\boldsymbol{y})$ *possessing property* (1) *defined for all pairs* ***x***, ***y*** *in a space of lattice points* $\mathbb{Z}^d$, *with d = 1,2,3,….* A random walk in $\mathbb{Z}^d$ is said to be *d-dimensional.*

Again returning to the specifics of my problem, mine is a particular class of random walks: the *simple random walk*. For the *d*-dimensional state space (that is $\mathbb{Z}^d$), let

$$|\boldsymbol{x}| = \left[\sum_{i=1}^{d} (x^i)^2\right]^{\frac{1}{2}}$$

---

[4] To maintain clarity, the animated points in GSP do not maintain this spatial homogeneity. Their behaviors change as they near the boundary of the window.

denote the Euclidean distance of the point x to the origin. Then $P(\mathbf{0}, \boldsymbol{x})$ is a *d-dimensional simple random walk* when

$$P(0,\boldsymbol{x}) = \begin{Bmatrix} \dfrac{1}{2d} & \text{when } |\boldsymbol{x}| = 1 \\ 0 & \text{when } |\boldsymbol{x}| \neq 1 \end{Bmatrix}.$$

Consider a function that would be the probability of an n-step transition from **0** to ***x*.** For this case, define

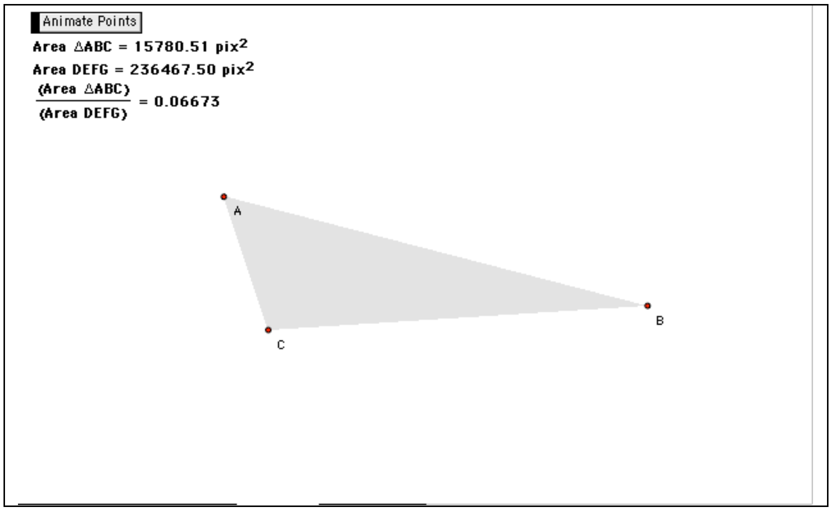
$$P_n(\mathbf{0},\boldsymbol{x}) = \begin{Bmatrix} \left(\dfrac{1}{2}\right)^{(n+x)/2} \left(\dfrac{1}{2}\right)^{(n-x)/2} \begin{pmatrix} n \\ \frac{n+x}{2} \end{pmatrix}, & \text{when } n + x \text{ is even and } |x| \leq n \\ 0 & \text{otherwise} \end{Bmatrix}.$$

The probability interpretation of this function P_n can be thought of as *the probability that a particle on a random walk, starting at the point* ***x*** *at time =0, will be at the point* ***y*** *at time = n.*

In my problem, the 1-dimensional equivalent is not quite equivalent to this simple random walk, but a minor variant. Imagine a particle placed at the origin of the real line. The toss of a coin would dictate its travels to the left or right; 1-unit to the right if a head is tossed, one unit left if tails. This is a 1-dimensional case, that is $d = 1$ with state space $\mathbb{Z}$, of the random walk—a case belonging to a class somewhat wider than the simple random walk. Note that now $P(0,1) = ½$ and $P(0,–1) = ½$. In this way, the erratic travels of this particle can also be considered to be a *Bernoulli sequence* – a sequence of independent events each with probability p or $1 - p$. As a point of clarification, it is not the random path, but the sequence of movements that constitute the random walk. This $P(\boldsymbol{x},\boldsymbol{y})$ is the transition function of a *Bernoulli random walk.* In this study of the topic, I will return to this case of the one dimensional random walk in Chapter 2.

New Questions Emerging from the Discord between Experience and Theory

In pursuit of investigating the original question—What is the average area of the triangle defined by three points on a random walk in the unrestricted plane?—I designed a GSP workspace that would help collect additional experimental data about the triangle area. I decided to focus on the area of the triangle as a portion of the desktop window with the idea that such data may be more readily comparable between different approaches to the problem, as opposed to relying on square pixels, square units, etc. as varying measures of area. Purely by watching the ratio change over time, I predicted that the ratio—and thus the answer to my question—was typically around 0.11 (see Figure 8), as was mentioned earlier. So on average, the area of the triangle formed by three points on a random walk appeared to average at most 0.11 times the area of the enclosing rectangular window of the GSP microworld.



Area $\triangle ABC = 15780.51$ pix^2

Area DEFG $= 236467.50$ pix^2

$\frac{(\text{Area } \triangle ABC)}{(\text{Area DEFG})} = 0.06673$

Figure 8. GSP microworld to observe triangle area ratio.

If a point on a random walk in $\mathbb{Z}^2$ moves with simple rules, that is it consistently takes steps of size 1 unit and is equally likely to move any one of the 4 directions N, S, E, and W, such a point will on average, be at it's starting location. This proof is rather obvious, by the symmetry of the probabilistic walk. However, this observation creates an interesting question: Is the average area of this random walking triangle dependent on the starting locations of the 3 points?

In other words, if the three points start out very near one another, would the triangle tend to have a small area? Versus the situation where the triangle area began at a much larger ratio of the rectangle… would this triangle tend to have a larger area? Remember, on average each vertex of the triangle will be at it's starting point. Again, I did some empirical exploration (see Figure 9). I simply created three different start sizes for the triangle, area portions 0.00006, 0.48485, and 0.06466. I allowed the triangles to "walk" for several minutes, taking measures of their areas at fixed intervals during this walk. The averages of these measures were 0.04133, 0.06472, and 0.08627. These averages cluster around 0.065 without a consistent pairing of value to triangle start size. This and similar empirical investigations suggest the average area is not affected by the starting size of the triangle. For now, this question remains a puzzler. However, this more careful data collection did reduce my personal estimate for the triangle's average area portion from 0.11 to 0.065.

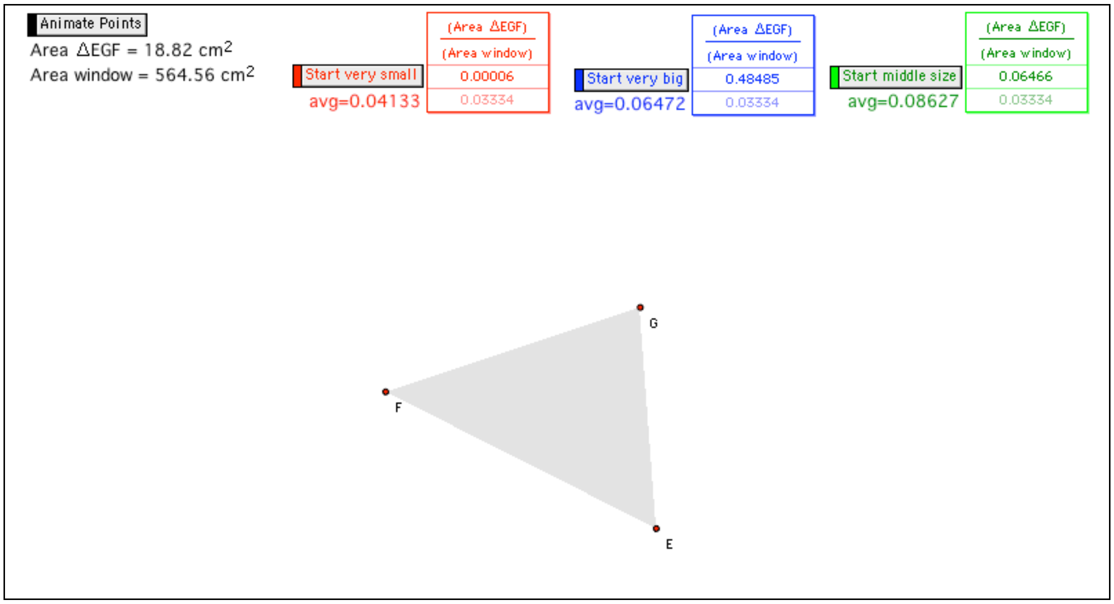


Figure 9. Results for the average area of walking triangles with 3 different initial areas.

The Refined Goal of the Investigation

It appears as though the average area of a triangle on a random walk is rather small, when watching such an event in the GSP environment. However, several assumptions must be more carefully ferreted out before relying too greatly on this empirical evidence. Recall that a true random walk would allow the point to travel beyond the boundary of the GSP window. However, the animated points in GSP are controlled by an algorithm that discourages them from leaving the screen. Another important consideration may be that the initial placement of the starting points of the triangle will impact the average area of the walking triangle.

To return to the original question and frame the remainder of this thesis, I will clarify the primary question of this study: What is average area of the triangle determined by 3 points on a simple random walk on the integer lattice $\mathbb{Z}^2$. To pursue this question, I intend to study:

- A probabilistic model of the area determined by 3 points on a simple random walk in 2-space.
- An analysis of this model that reports the expectation of this probability function, that is, what is the average area of such a triangle?
- A correction to the probabilistic model that better matches the constraints imposed by *Geometer's Sketchpad* on the random walks of the three points.
- How well does the mathematical theory assuming a true, simple random walk, coincide with the empirical results of a pseudo-random walk of the *Geometer's Sketchpad* environment?
- Finally, how does the answer to the average area of the triangle formed by three walking points constrained to a *Geometer's Sketchpad* window compare to the ratio

> of the area of the triangle determined by 3 points randomly chosen on the interior of some fixed-size rectangle?

Chapter 2 of the thesis will bring the context to 1-dimension, considering the distance between two points on a line without a GSP-like constraint. Chapter 3 will take the exploration back into 2-dimensions. Chapter 4 will present and discuss the unexplored (or underexplored) questions that have emerged during the work on this thesis, and Chapter 5 will serve to summarize and conclude the investigation.

CHAPTER 2

INVESTIGATION IN ONE DIMENSION

To initiate the investigation, I simplified the problem to consider 2 points (or particles) on a random walk on the line, and the distance between them. The somewhat equivalent question I pursued was: What is the average distance between 2 particles, *p* and *q*, on a random walk on state space $\mathbb{Z}$. In particular, consider

$p_n, q_n \in \mathbb{Z},$ with $n \in \{0,1,...\}$ indicating the step number of the random walk,
$p_0 = q_0 = 0,$ and
distance $d(p_n, q_n) = |p_n - q_n|$

These particles, *p* and *q*, are equally likely to take one step to the right or to the left on the integer line. That is $P(0,1) = ½$ and $P(0,–1) = ½$, Where $P(\mathbf{0}, \boldsymbol{x})$ is the transition function.

I first present some initial observations.

After an odd number of steps (*n* odd), $p_n, q_n \in$ odd integers , and

After an even number of steps (*n* even), $p_n, q_n \in$ even integers.

This behavior shows that $d(p_n, q_n) \in$ even integers since the distances between odd integers is always even, as is the distance between even integers. To specify the exact set of even integers that could be the distance after *n* steps, consider the furthest right a particle could travel in *n* steps is the point *n*. And furthest left is point –*n*. The distance between them would be at most 2*n*. All even distances up to this are possible. So, after *n* steps,

$$d(p_n, q_n) \in \{0, 2, ...2n\}$$

To figure the probability of any of these distances occurring, consider the initial steps of the walk. When n = 1, $(p_1, q_1) \in \{(-1,-1),(-1,1),(1,-1),(1,1)\}$. So

$$P(d(p_1,q_1)=0)=2\cdot\tfrac{1}{2}\cdot\tfrac{1}{2}=\tfrac{1}{2}$$
$$P(d(p_1,q_1)=2)=2\cdot\tfrac{1}{2}\cdot\tfrac{1}{2}=\tfrac{1}{2}$$

Continuing to n = 2, first note that p_2 and q_2 could arrive at point –2 with a ¼ probability, since $P(0,-1)\cdot P(0,-1)=\tfrac{1}{2}\cdot\tfrac{1}{2}=\tfrac{1}{4}$. Symmetrically, p_2 and q_2 could arrive at point +2 with a ¼ probability. But either point could arrive at point 0 after 2 steps by a move right then left, $P(0,1)\cdot P(0,-1)=\tfrac{1}{2}\cdot\tfrac{1}{2}=\tfrac{1}{4}$, or by a move left then right, $P(0,-1)\cdot P(0,1)=\tfrac{1}{2}\cdot\tfrac{1}{2}=\tfrac{1}{4}$. Thus (and summarizing),

$$P(d(p_2,q_2)=-2)=1\cdot\tfrac{1}{2}\cdot\tfrac{1}{2}=\tfrac{1}{4}$$
$$P(d(p_2,q_2)=0)=2\cdot\tfrac{1}{2}\cdot\tfrac{1}{2}=\tfrac{1}{2}$$
$$P(d(p_2,q_2)=2)=1\cdot\tfrac{1}{2}\cdot\tfrac{1}{2}=\tfrac{1}{4}$$

Distance Probabilities: From Examples to Generalization

The structure of the computation at n = 2 reveals that to figure the probability of each distance, I must count the number of ways the pair of points can end up that distance apart, and multiply by the probability any one of the necessary sequence of steps occurs. This probability is rather simple, it is $\tfrac{1}{2^n}$ because each step has an equal probability of occurring, ½. Counting the number of ways each distance can occur is more of a challenge. I began considering this question by asking, “How many ways can p arrive at {-n, -n + 2, …, n – 2, n} after n steps?” For $p_n = -n$, each step must be P(0, -1). At each step of the sequence, there is only one choice. The multiplication principle for counting says that the total number of ways such a move can occur would be $1 \cdot 1 \cdot ... \cdot 1$, n times which is equal to $1^n = \binom{n}{0}$. Considering the possibility that $p_n = -n + 2$, I realized that at any one step, the move must be $P(0, 1)$. For an n step walk, this makes n total options for taking the $P(0, 1)$ step, which is also equal to $\binom{n}{1}$. This logic continues for $n > 1$. To end at $p_n = -n + 4$, two steps must be $P(0, 1)$ with all others being $P(0, -1)$. So, of the n steps, select 2 to be $P(0, 1)$. There are $\binom{n}{2}$ ways for this to occur. Extending this argument through all n steps creates the following set of probabilities for each p and q ending an n-step walk at a given point:

$$P(p_n = -n + 2i)_{i=0,1,...n} = \binom{n}{i} \cdot \frac{1}{2^n}$$

(Note: the $-n + 2i$ keeps the domain of possible locations for p_n either the odds or the evens, as appropriate to the value of p_n.)

Notating this computation was not too complicated. Figuring the probabilities of both p_n and q_n being in some particular location is also not much of a challenge. To answer what the probability of $p_n = r$ and $q_n = s$, with $r, s \in \{-n + 2i\}_{i=0,1,...,n}$ requires a simple multiplication

$$P(p_n = r) \cdot P(q_n = s) = \left(\binom{n}{\frac{r+n}{2}} \cdot \frac{1}{2^n} \right) \cdot \left(\binom{n}{\frac{s+n}{2}} \cdot \frac{1}{2^n} \right) = \frac{1}{2^{2n}} \cdot \binom{n}{\frac{r+n}{2}} \cdot \binom{n}{\frac{s+n}{2}}$$

However, this is only one situation in which $d(p_n, q_n) = |r - s|$. For example, the same distance occurs when $p_n = r + 1$ and $q_n = s + 1$, provided $r+1, s+1 \in \{-n + 2i\}_{i=0,1,...,n}$. To count all possible, first consider the largest possible distance, $|p_n - q_n| = 2n$. This only can occur when $p_n = -n$ and $q_n = n$, or vice-versa. So there are 2 ways. The next largest distance $|p_n - q_n| = 2n - 2$

can occur when $p_n = -n$ and $q_n = n - 2$, or $p_n = -n + 2$ and $q_n = n$. The same "vice-versa" applies, so there are 4 ways. Each 2-unit decrease in distance adds two possible pairings. Generalizing, for the distance $|p_n - q_n| = 2n - 2j$ with $\mathrm{j} = 0,1,\ldots,n-1$, there will be $2 + 2j$ possible pairings of p_n and q_n. When considering $|p_n - q_n| = 0$, the "vice-versa" achieved by switching the positions of the p_n and q_n no longer create new possible outcomes. So, the distance $|p_n - q_n| = 0$ when $2n - 2i$ with $j = n$ occurs $1 + j$ times.

Gathering these results together with the previous computations for $P(p_n = r)\cdot P(q_n = s)$ for the case when $|p_n - q_n| = 0$ yields,

$$(2) \qquad P(d(p_n,q_n)=0)=\frac{1}{2^{2n}}\cdot\binom{n}{0}\cdot\binom{n}{0}+\frac{1}{2^{2n}}\cdot\binom{n}{1}\cdot\binom{n}{1}+\ldots+\frac{1}{2^{2n}}\cdot\binom{n}{n}\cdot\binom{n}{n}=\frac{1}{2^{2n}}\sum_{i=0}^{n}\binom{n}{i}^2 .$$

For the case $|p_n - q_n| = 2n - 2j$ with $\mathrm{j} = 0,1,\ldots,n-1$,

$$\begin{aligned} P(d(p_n,q_n)=2n-2j) &= 2\cdot\frac{1}{2^{2n}}\cdot\binom{n}{0}\cdot\binom{n}{n-j}+2\cdot\frac{1}{2^{2n}}\cdot\binom{n}{1}\cdot\binom{n}{n-j+1}+\ldots+2\cdot\frac{1}{2^{2n}}\cdot\binom{n}{j}\cdot\binom{n}{n} \\ &= 2\cdot\frac{1}{2^{2n}}\sum_{i=0}^{j}\binom{n}{i}\binom{n}{n-j+i}. \end{aligned}$$

To simplify, rename j in order to consider $d(p_n,q_n) = 2j$ with $j = 0, 1, \ldots, n$—that is j now refers to half the distance between particles p and q at step n:

$$P(d(p_n,q_n)=2j)_{j=0\ldots n}=\begin{cases} \dfrac{1}{2^{2n}}\displaystyle\sum_{i=0}^{n-j}\binom{n}{i}^2 & \text{when } j=0 \\ 2\cdot\dfrac{1}{2^{2n}}\displaystyle\sum_{i=0}^{n-j}\binom{n}{i}\binom{n}{j+i} & \text{when } j=1\ldots n \end{cases} .$$

Expectation for Distance Between Two Walking Points, One Dimension

To determine the average distance between these points requires one further algebraic step. This average distance is figured as the expectation of the random variable—in this case our distance, a discrete random variable. Expectation, denoted E(***X***) for random variable ***X*** is defined by

$$E(\boldsymbol{X}) = \sum_{\boldsymbol{x}:p(\boldsymbol{x})>0} \boldsymbol{x}p(\boldsymbol{x}).$$

More simply put, expectation of ***X*** is a weighted average of the possible values ***X*** can take on., each being weighted by the probability that ***X*** assumes it. To sum these weighted values results in a simplified computation since the case when the distance = 0 drops out the first part of the piecewise formula. The second part becomes $2j$ times the probability, yielding

$$E_n(d(p_n,q_n)) = \sum_{j=1}^{n}\left[2j\cdot 2\cdot \frac{1}{2^{2n}}\sum_{i=0}^{n-j}\binom{n}{i}\binom{n}{j+i}\right].$$

For me, this formula is powerful yet limited as it stands now. It doesn't mean too much; it would make better sense by considering the values it takes on for different *n*'s, while wondering $\lim_{n\to\infty} E_n(d(p_n,q_n))$. Using Maple to run the computations, I found the expectation for several values of *n,* reported in Table 1. It is apparent that as *n* increases, so does the expectation. A plot of this data resembles a power function, $f(x) = x^a$, manner of growth with $0 < a < 1$ (see Figure 10). The limit appears to be unbounded. Allowing a TI-84 to complete a power regression on the data [except for point (0,0)] yielded the following function:

$$E(n) = 1.08688693n^{0.506738910} \qquad r = 0.999873253,$$

where the *r* is a correlation coefficient. More elegantly, this expectation curve resembles a square root curve, with $a = 0.5$.

Table 1

Average Distance (Expectation) Between 2 Points on an n-Step Random Walk

n number of steps	*E* expectation	*n* number of steps	*E* expectation
0	0	20	5.014827505
1	1.0	30	6.154690380
2	1.500000000	40	7.114230302
3	1.875000000	50	7.958923739
4	2.187500000	60	8.722197469
5	2.460937500	70	9.423854237
6	2.707031250	80	10.07677293
7	2.932617188	90	10.68988739
8	3.142089844	100	11.26969580
9	3.338470459	200	15.94772079
10	3.523941040	300	19.53595880
11	3.700138092	400	22.56053208
12	3.868326187	500	25.22501818
13	4.029506445	600	27.63377432
14	4.184487462	700	29.84877599
15	4.333933443	800	31.91039604
16	4.478397891	900	33.84667376
17	4.618347825	1000	35.67802229
18	4.754181585	2000	50.45949662
19	4.886242184	5000	79.78646139

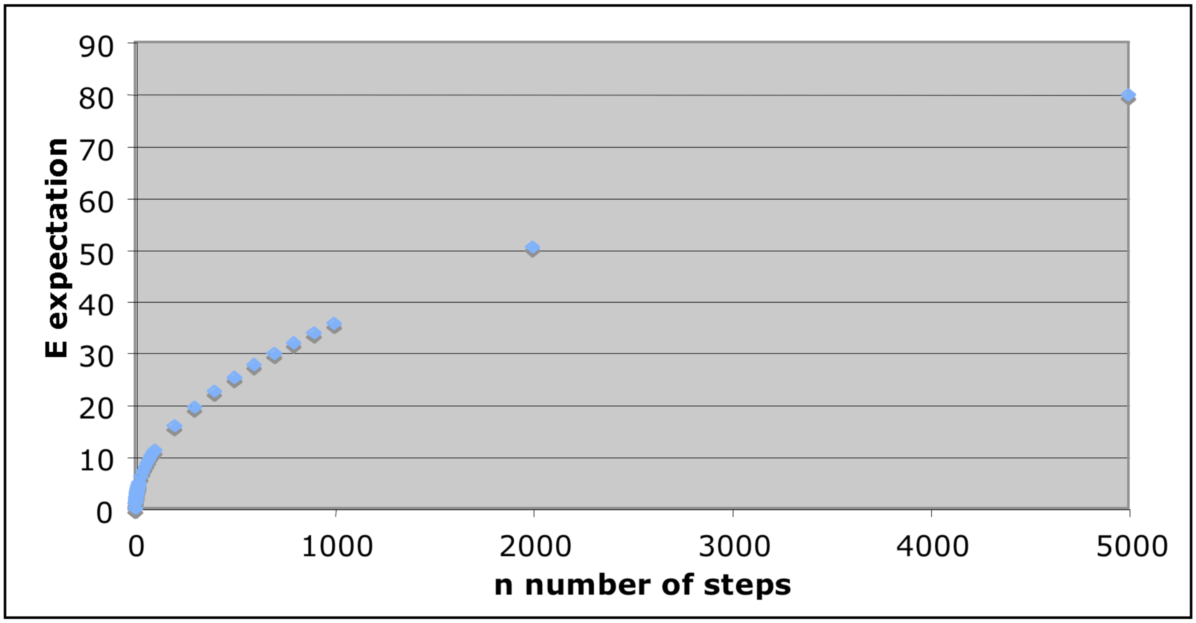


Figure 10. Distance expectation between 2 walking points in 1 dimension.

Further inquiry into this approximate square root behavior yielded a finding on Wikipedia that suggested “The average straight-line distance between start and finish points of a random walk of *n* steps is on the order of $\sqrt{n}$, or more precisely, its converges asymptotically to $\sqrt{\frac{2n}{\pi}} \approx 0.8\sqrt{n}$ ” [online at http://en.wikipedia.org/wiki/Random_walk, 06/08/06]. Note this refers to the random walk of a single point. This claim opens yet to be explored questions, discussed further in Chapter 4. However, it did prompt me to consider the end behavior of this function.

End Behavior of Distance Expectation

In my first effort to study the end behavior of the distance expectation, I sought a closed form of the summation created above. Recall:

$$E_n(d(p_n,q_n)) = \sum_{j=1}^{n}\left[2j\cdot 2\cdot\frac{1}{2^{2n}}\sum_{i=0}^{n-j}\binom{n}{i}\binom{n}{j+i}\right].$$

Simplifying,

$$E_n(d(p_n,q_n)) = \frac{4}{2^{2n}}\sum_{j=1}^{n}\left[j\sum_{i=0}^{n-j}\binom{n}{i}\binom{n}{j+i}\right].$$

Next I will establish the following relationship, useful several times in this analysis.

Lemma (A): $\sum_{i=0}^{n}\binom{n}{i}^2 = \binom{2n}{n}$.

Proof: First I will argue that

$$\binom{n+m}{r} = \binom{n}{0}\binom{m}{r} + \binom{n}{1}\binom{m}{r-1} + \ldots + \binom{n}{r}\binom{m}{0}.$$

The RHS says: “For every way I can choose 0 items from *n*, I can choose *r* items from *m*. Multiply this to count the number of ways to choose 0 items from *n* and *r* from *m*. Next consider how many ways to get 1 item from *n* and *r*-1 items from *m*. Multiply again to know total possible. And continue until you've counted all the ways to choose all *r* items from *n* and 0 from *m*. Add the results." Thus, this RHS sum yields the total ways to choose *r* items from separated batches *m* and *n* first taking 0 from *n* (the rest from *m*), then 1 from *n*, …, until all *r* are from *n*. This is equivalent to counting the number of ways to select *r* items from *n* + *m* items, which is the LHS $\binom{n+m}{r}$.

Now that this equality is established, let $m = n$ and $r = n$ and rewrite.

$$\binom{n+n}{n} = \binom{n}{0}\binom{n}{n} + \binom{n}{1}\binom{n}{n-1} + \ldots + \binom{n}{n}\binom{n}{0}.$$

This is equivalent to

$$\binom{2n}{n} = \binom{n}{0}\binom{n}{0} + \binom{n}{1}\binom{n}{1} + \ldots + \binom{n}{n}\binom{n}{n} = \sum_{i=0}^{n}\binom{n}{i}^2$$

since $\binom{n}{r} = \binom{n}{n-r}$, the vertical symmetry seen in Pascal's Triangle.

Next I will use Lemma (A) to demonstrate the following relationship:

Lemma (B): $\sum_{i=0}^{n-j}\binom{n}{i}\binom{n}{j+i} = \binom{2n}{n+j}$.

Proof: I have established above that $\binom{n+m}{r} = \binom{n}{0}\binom{m}{r} + \binom{n}{1}\binom{m}{r-1} + \ldots + \binom{n}{r}\binom{m}{0}$.

Substitute $m = n$, $r = n - j$ to obtain

$$\binom{n+n}{n-j} = \binom{2n}{n-j} = \binom{n}{0}\binom{n}{n-j} + \binom{n}{1}\binom{n}{n-j-1} + \ldots + \binom{n}{n-j}\binom{n}{0}.$$

Because $\binom{u}{v} = \binom{u}{u-v}$, $\binom{2n}{n-j} = \binom{2n}{2n-(n-j)} = \binom{2n}{n+j}$. Thus,

$$\binom{n}{0}\binom{n}{n-j}+\binom{n}{1}\binom{n}{n-j-1}+\ldots+\binom{n}{n-j}\binom{n}{0}=\binom{2n}{n+j}.$$

Again utilizing $\binom{u}{v}=\binom{u}{u-v}$, rewrite the above by changing the second binomial coefficient in each addend of the LHS:

$$\binom{n}{0}\binom{n}{j}+\binom{n}{1}\binom{n}{j+1}+\ldots+\binom{n}{n-j}\binom{n}{n}=\binom{2n}{n+j}.$$

Rewriting the LHS in summation notation yields the stated form of Lemma (B):

$$\sum_{i=0}^{n-j}\binom{n}{i}\binom{n}{j+i}=\binom{2n}{n+j}.$$

Since Lemma (B) is true,

$$\frac{4}{2^{2n}}\sum_{j=1}^{n}\left[j\sum_{i=0}^{n-j}\binom{n}{i}\binom{n}{j+i}\right]=\frac{1}{2^{2n-2}}\sum_{j=1}^{n}\left[j\binom{2n}{n+j}\right] \tag{3}$$

Adding and then subtracting the same term gives

$$\sum_{j=1}^{n}j\binom{2n}{n+j}=\sum_{j=1}^{n}(n+j)\binom{2n}{n+j}-\sum_{j=1}^{n}n\binom{2n}{n+j}.$$

Using the property $y\binom{x}{y}=x\binom{x-1}{y-1}$,

$$\begin{aligned}\sum_{j=1}^{n}(n+j)\binom{2n}{n+j}-\sum_{j=1}^{n}n\binom{2n}{n+j}&=\sum_{j=1}^{n}2n\binom{2n-1}{n+j-1}-\sum_{j=1}^{n}n\binom{2n}{n+j}\\&=2n\sum_{j=1}^{n}\binom{2n-1}{n+j-1}-n\sum_{j=1}^{n}\binom{2n}{n+j}\end{aligned}. \tag{4}$$

Careful examination shows the first sum in expression (4) to be the summation of the right half of an odd row of Pascal's Triangle. Since $\sum_{i=0}^{n}\binom{n}{i}=2^n$ and odd numbered rows of Pascal's Triangle have an even number of entries, $\sum_{j=1}^{n}\binom{2n-1}{n+j-1}=\frac{2^{2n-1}}{2}$. The second sum in expression (4) is the sum of the right "half" of an even row—not including the middle term. Note that even rows of Pascal's triangle have an odd number of entries, with the middle entry $\binom{2n}{n}$. So $\sum_{j=1}^{n}\binom{2n}{n+j}=\frac{1}{2}\left(2^{2n}-\binom{2n}{n}\right)$. Restating:

$$\begin{aligned}2n\sum_{j=1}^{n}\binom{2n-1}{n+j-1}-n\sum_{j=1}^{n}\binom{2n}{n+j}&=2n\frac{2^{2n-1}}{2}-n\frac{1}{2}\left(2^{2n}-\binom{2n}{n}\right)\\&=\frac{2n\cdot 2^{2n-1}-n\cdot 2^{2n}+n\binom{2n}{n}}{2}\\&=n\cdot 2^{2n-1}-n\cdot 2^{2n-1}+\frac{n}{2}\binom{2n}{n}\\&=\frac{n}{2}\binom{2n}{n}\end{aligned}.$$

Substituting into (3),

$$\begin{aligned}\frac{1}{2^{2n-2}}\sum_{j=1}^{n}\left[j\binom{2n}{n+j}\right]&=\frac{1}{2^{2n-2}}\frac{n}{2}\binom{2n}{n}\\&=\frac{n}{2^{2n-1}}\binom{2n}{n}\end{aligned}. \tag{5}$$

Stirling's Formula states that, for k large, $k!\sim k^{k}e^{-k}\sqrt{2\pi k}$. So,

$$\binom{2n}{n}=\frac{2n!}{(2n-n)!(n)!}=\frac{2n!}{(n)!(n)!}\sim\frac{(2n)^{2n}e^{-2n}\sqrt{2\pi 2n}}{\left(n^{n}e^{-n}\sqrt{2\pi n}\right)^{2}}=\frac{(2n)^{2n}e^{-2n}2\sqrt{\pi n}}{n^{2n}e^{-2n}2\pi n}=\frac{2^{2n}}{\sqrt{\pi n}}.$$

Rewriting (5) with this approximation,

$$\frac{n}{2^{2n-1}}\binom{2n}{n}\sim\frac{n}{2^{2n-1}}\cdot\frac{2^{2n}}{\sqrt{\pi n}}=\frac{2n}{\sqrt{\pi n}}=\frac{2}{\sqrt{\pi}}\sqrt{n}\approx 1.1284\sqrt{n}$$

This result agrees with the near square-root like behavior suggested by the regression equation determined above.

In a second effort to analyze the end behavior for the distance expectation of two walking particles in one dimension, I returned to an analysis of the expectation summation considering the effect from step *n* to step *n* +1. The regression above suggests this growth is of a $\sqrt{n}$ type. If it is apparent, for some sequence A_n, that $A_{n+1} = A_n + k \cdot \frac{1}{2\sqrt{n}}$ for some constant *k*, then the sequence A_n grows on the order $\sqrt{n}$. This expectation, that is for the distance between two walking points in one dimension, behaves in this manner by both my proof in the first effort and in the data regression. In particular, I claim $\lim\limits_{n\to\infty} E_n(d(p_n, q_n)) = \frac{2}{\sqrt{\pi}}\sqrt{n}$.

Proof: Let $S_n = 2^{2n} \cdot E_n$, where E_n is the expected distance at step *n*. Then S_n is the sum across even distances *j* from 0 to 2*n* of the distance *j* times the numbers of ways to get distance *j*, at step *n*. Let $P_{n,j}$ be the number of ways to get distance *j*, at step *n*. Then restating,

$S_n = \sum\limits_{j \text{ from } 0 \text{ to } 2n \text{ by } 2} j \cdot P_{n,j}$, and $S_{n+1} = \sum\limits_{j \text{ from } 0 \text{ to } 2n+2 \text{ by } 2} j \cdot P_{n+1,j}$.

Lemma (C): $S_{n+1} = 4S_n + 4P_{n,0}$.

Proof: Every path $P_{n,j}$ of the sequence S_n contributes 4 of the paths composing S_{n+1} because each of the particles p_n and q_n have two options for the next step.

For any $P_{n,j}$, two of these options contribute a $P_{n+1,j}$, one contributes a $P_{n+1,j-2}$, and one contributes to $P_{n+1,j+2}$. Except $P_{n,0}$ which contributes two $P_{n+1,j}$ and two $P_{n+1,j+2}$ (that is two $P_{n+1,0}$ and two $P_{n+1,2}$). Each of these contributions are predicated on the notion that *p* and *q* could both move forward together, backward together, apart, or closer. Except when they begin at the same place (i.e. distance 0), whence they cannot move closer but instead apart in two ways: by *p* moving right or by *q* moving right.

Next, I account for the distances each of these contributing paths influences.

Recall that by definition, $S_n = 0P_{n,0} + 2P_{n,2} + 4P_{n,4} + ... + (2n-2)P_{n,2n-2} + 2nP_{n,2n}$.

All that follows is by algebraic simplification, beginning with a restatement of the definition of S_{n+1} followed by a substitution based upon the arguments above:

$$\begin{aligned} S_{n+1} &= 0P_{n+1,0} + 2P_{n+1,2} + 4P_{n+1,4} + 6P_{n+1,6} + 8P_{n+1,8} + ... + \\ &\quad (2n-4)P_{n+1,2n-4} + (2n-2)P_{n+1,2n-2} + (2n)P_{n+1,2n} + (2n+2)P_{n+1,2n+2} \\ &= 2(2P_{n,0} + 2P_{n,2} + P_{n,4}) + 4(P_{n,2} + 2P_{n,4} + P_{n,6}) + 6(P_{n,4} + 2P_{n,6} + P_{n,8}) + \\ &\quad 8(P_{n,6} + 2P_{n,8} + P_{n,10}) + ... + (2n-4)(P_{n,2n-6} + 2P_{n,2n-4} + P_{n,2n-2}) + \\ &\quad (2n-2)(P_{n,2n-4} + 2P_{n,2n-2} + P_{n,2n}) + (2n)(P_{n,2n-2} + 2P_{n,2n}) + (2n+2)(P_{n,2n}) \end{aligned}$$

Distributing the distance coefficients,

$$\begin{aligned} S_{n+1} &= [4P_{n,0} + 4P_{n,2} + 2P_{n,4}] + [4P_{n,2} + 8P_{n,4} + 4P_{n,6}] + [6P_{n,4} + 12P_{n,6} + 6P_{n,8}] + \\ &\quad [8P_{n,6} + 16P_{n,8} + 8P_{n,10}] + ... + [(2n-4)P_{n,2n-6} + (4n-8)P_{n,2n-4} + \\ &\quad (2n-4)P_{n,2n-2}] + [(2n-2)P_{n,2n-4} + (4n-4)P_{n,2n-2} + (2n-2)P_{n,2n}] + \\ &\quad [(2n)P_{n,2n-2} + (4n)P_{n,2n}] + [(2n+2)P_{n,2n}] \end{aligned}$$

Next step is to simplify by combining like terms:

$$S_{n+1} = 4P_{n,0} + 8P_{n,2} + 16P_{n,4} + 24P_{n,6} + ... + (8n-8)P_{n,2n-2} + (8n)P_{n,2n}$$

And by regrouping,

$$S_{n+1} = 4P_{n,0} + 4(2P_{n,2} + 4P_{n,4} + 6P_{n,6} + ... + (2n-2)P_{n,2n-2} + (2n)P_{n,2n})$$

Finally, recalling the definition of S_n,

$$S_{n+1} = 4P_{n,0} + 4S_n$$

Returning to the effort to show E_n grows on order $\sqrt{n}$, recall from equation (2) and Lemma (A), $P_{n,0} = \sum_{i=0}^{n} \binom{n}{i}^2 = \binom{2n}{n}$. Therefore $S_{n+1} = 4S_n + 4\binom{2n}{n}$. Since $S_n = 2^{2n}E_n$,

Rewrite as

$$2^{2n+2}E_{n+1} = 4 \cdot 2^{2n} E_n + 4\binom{2n}{n}$$

$$2^{2n+2}E_{n+1} = 2^{2n+2} E_n + 4\binom{2n}{n}$$

$$E_{n+1} = E_n + \frac{1}{2^{2n}}\binom{2n}{n}$$

Recall Stirling's Formula states that, for k large, $k! \sim k^k e^{-k}\sqrt{2\pi k}$. So,

$$\binom{2n}{n} = \frac{2n!}{(2n-n)!(n)!} = \frac{2n!}{(n)!(n)!} \sim \frac{(2n)^{2n} e^{-2n}\sqrt{2\pi 2n}}{\left(n^n e^{-n}\sqrt{2\pi n}\right)^2} = \frac{(2n)^{2n} e^{-2n} 2\sqrt{\pi n}}{n^{2n}e^{-2n}2\pi n} = \frac{2^{2n}}{\sqrt{\pi n}}$$

Using this result to rewrite E_{n+1},

$$E_{n+1} = E_n + \frac{1}{2^{2n}} \cdot \frac{2^{2n}}{\sqrt{\pi n}} = E_n + \frac{1}{\sqrt{\pi n}} = E_n + \frac{2}{\sqrt{\pi}} \cdot \frac{1}{2\sqrt{n}} .$$

So, in fact the sequence E_n grows on order $\sqrt{n}$, with $k = \frac{2}{\sqrt{\pi}} \approx 1.1284$. This approach may provide an alternative to thinking of these types of problems: 2 or 3 walking points and their expected distances or areas.

CHAPTER 3

RETURN TO 2 DIMENSIONS

The investigation of the average distance between 2 points on a random walk in 1 dimension grounded some ideas for considering the problem of the average area of the triangle determined by 3 points in 2 dimensions. I realized my computations would involve several components. These were:

- Accounting for the possible ending locations of a particle on a 2D random walk after *n* steps (this is the set of possible points after *n* steps);
- Determining the number of ways that particle could arrive at that location;
- Accounting for the sets of locations where 3 (or 2) particles could arrive after *n* steps;
- Determining the probability of each set occurring—namely, how many ways can each set occur (this is simply an application of the multiplication principle, using results of bullet 2);
- Figuring the area (or distance) determined by the 3 (or 2) points in each of the sets mentioned above; and
- Accumulating all these areas (or distances), giving care to count the number of ways each occurs, in order to determine an average.

Bullets 1, 2, 3, and 6 presented the mathematical challenge of this task. First, the entire space $\mathbb{Z}$ or $\mathbb{Z}^2$ are not all possible ending locations. Developing a simple way to name the set of ending locations for any *n* would aid this task. Next, to count the number of ways a particle could travel to any of these end locations was not directly obvious. Third, a system of accounting for all possible pairs of points where the particles could end demanded attention. Should this be done systematically via a simple convention such as: {AA, AB, AC, … BA, BB,…} or would another manner ease computation? The final challenge is to develop a systematic and efficient way to account for each distance or area and its associated probability, and accumulate these in a manner allowing for the computation of the expected value. Rather than begin tackling these challenges with 3 walking points on the plane, I began with 2 points.

Distance Between 2 Walking Points on $\mathbb{Z}^2$

In pursuance of the expectation for the distance between two points on a random walk on $\mathbb{Z}^2$ that began at the origin, I follow the same outline above. Once again the expectation is computed

$$E(\boldsymbol{X}) = \sum_{\boldsymbol{x}:p(\boldsymbol{x})>0} \boldsymbol{x}p(\boldsymbol{x}).$$

with $\boldsymbol{x}$ being the possible distances between 2 points in $\mathbb{Z}^2$, and $p(\boldsymbol{x})$ being the probability of the associated distance occurring. $p(\boldsymbol{x})$ will be computed by figuring

$$P(d(p_n,q_n)) = \sum(P \cdot Q)$$

where P and Q are the probabilities of $\boldsymbol{p}_n$ and $\boldsymbol{q}_n$ arriving at a pair of points a particular distance apart. The summation is to indicate that the probabilities of all possible pairings $\boldsymbol{p}_n$, $\boldsymbol{q}_n$ yielding this distance must be summed.

For clarification, I will demonstrate some simpler cases. At step $n = 0$, both $\boldsymbol{p}_n$ and $\boldsymbol{q}_n$ are at the origin. Thus, $d(\boldsymbol{p}_n, \boldsymbol{q}_n) = 0$ and $P(d(\boldsymbol{p}_n, \boldsymbol{q}_n) = 0) = 1$. So $E_0(\boldsymbol{X}) = (0)(1) = 0$.

Continuing beyond this trivial case, consider step $n = 1$. After one step, $\boldsymbol{p}_n$ and $\boldsymbol{q}_n$ are equally likely to be at any of the four points {(1,0), (0,1), (-1,0), (0,-1)}. This results in the set of possible distances to be

$$d(\boldsymbol{p}_n, \boldsymbol{q}_n) \in \{0, \sqrt{2}, 2\}.$$

The distance is 0 when each point ends at the same point. There are 4 ways this can occur. The distance is $\sqrt{2}$ when the second point is a diagonal neighbor of the first, $4 \cdot 2 = 8$ possibilities. And finally, the distance is 2 when the second point is furthest from the first point—one place for $\boldsymbol{q}_n$ for each choice of $\boldsymbol{p}_n$—4 possibilities. So after one step, the probabilities are

$$\begin{aligned} P(d(\boldsymbol{p}_1, \boldsymbol{q}_1) = 0) &= 4 \cdot \tfrac{1}{4} \cdot \tfrac{1}{4} = \tfrac{4}{16} \\ P(d(\boldsymbol{p}_1, \boldsymbol{q}_1) = \sqrt{2}) &= 8 \cdot \tfrac{1}{4} \cdot \tfrac{1}{4} = \tfrac{8}{16} \\ P(d(\boldsymbol{p}_1, \boldsymbol{q}_1) = 2) &= 4 \cdot \tfrac{1}{4} \cdot \tfrac{1}{4} = \tfrac{4}{16} \end{aligned}$$

Compute the expectation by multiplying distances by probabilities and summing:

$$E_1(\boldsymbol{X}) = 0 \cdot \tfrac{4}{16} + \sqrt{2} \cdot \tfrac{8}{16} + 2 \cdot \tfrac{4}{16} = \tfrac{\sqrt{2}+1}{2} \approx 1.2071.$$

These first two examples are quite trivial. When $n = 2$, the computation becomes a significant challenge. The number of possible locations for $\boldsymbol{p}_2$ or $\boldsymbol{q}_2$ is 9 (see Figure 11, points **A**-**I**), yielding 6 possible distances, $d(\boldsymbol{p}_2, \boldsymbol{q}_2) \in \{0, \sqrt{2}, 2, 2\sqrt{2}, \sqrt{10}, 4\}$. Counting the number of ways each of these distances can occur is not immediately obvious. Not only must I count the number of combinations that yield a particular distance, I also must count the number of ways a particle can arrive at a point. For example, particle $\boldsymbol{p}_2$ could arrive at point **C** along the path a-**C** or b-**C**.[5] I will demonstrate this for the case $d(\boldsymbol{p}_2, \boldsymbol{q}_2) = 0$. In this case, particles $\boldsymbol{p}_2$ and $\boldsymbol{q}_2$ must both end their walk at the same point. Consider point **A**. There are 4 ways for a particle to travel to **A** in 2 steps, the paths suggested by a-**A**, b-**A**, c-**A**, and d-**A**, each with probability $\tfrac{1}{4} \cdot \tfrac{1}{4} = \tfrac{1}{16}$. So, the probability that $\boldsymbol{p}_2$ and $\boldsymbol{q}_2$ both end at **A** is $P((\boldsymbol{p}_2, \boldsymbol{q}_2) = \boldsymbol{AA}) = 4 \cdot \tfrac{1}{16} \cdot 4 \cdot \tfrac{1}{16} = \tfrac{16}{256}$. There is only one path to **B**, so $P((\boldsymbol{p}_2, \boldsymbol{q}_2) = \boldsymbol{BB}) = 1 \cdot \tfrac{1}{16} \cdot 1 \cdot \tfrac{1}{16} = \tfrac{1}{256}$. And there are 2 paths to **C**. Thus,

---

[5] Notational note: I will indicate paths using a dash, "-" between the point names along the path.

$P((\boldsymbol{p}_2,\boldsymbol{q}_2)=\boldsymbol{CC})=2\cdot\frac{1}{16}\cdot 2\cdot\frac{1}{16}=\frac{4}{256}$. Next I invoke the symmetry of the dihedral group symmetry of the integer lattice $\mathbb{Z}^2$ to note that

$$P((\boldsymbol{p}_2,\boldsymbol{q}_2)=\boldsymbol{BB})=P((\boldsymbol{p}_2,\boldsymbol{q}_2)=\boldsymbol{DD})=P((\boldsymbol{p}_2,\boldsymbol{q}_2)=\boldsymbol{FF})=P((\boldsymbol{p}_2,\boldsymbol{q}_2)=\boldsymbol{HH}),$$

and

$$P((\boldsymbol{p}_2,\boldsymbol{q}_2)=\boldsymbol{CC})=P((\boldsymbol{p}_2,\boldsymbol{q}_2)=\boldsymbol{EE})=P((\boldsymbol{p}_2,\boldsymbol{q}_2)=\boldsymbol{GG})=P((\boldsymbol{p}_2,\boldsymbol{q}_2)=\boldsymbol{II}).$$[6]

Summing these results yields the probability of the two particles ending at the same point.

$$P(d(\boldsymbol{p}_2,\boldsymbol{q}_2)=0)=\frac{16}{256}+4\cdot\frac{1}{256}+4\cdot\frac{4}{256}=\frac{36}{256}.$$

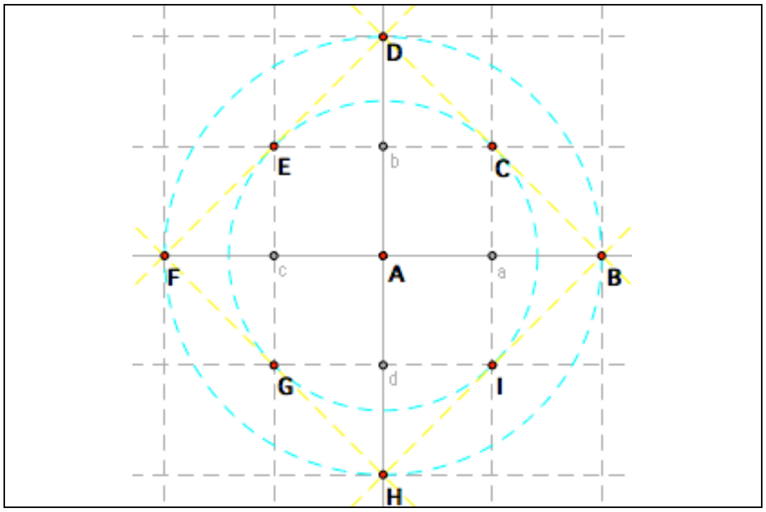


Figure 11. Possible end locations for a walking point in 2 dimensions at $n = 2$.

Drawing upon this symmetry created an algorithm for which to count the number of ways a particular distance could occur—the numerator of the fraction above. The algorithm for case $n = 2$ is as follows:

1. Count the number of paths from the origin to each point in the state space. Note that because of the symmetry, once all points in the first 45° of Quadrant 1 are counted, all other points in the space are equivalent to one of these. (Note: Step 1 does not need to be repeated for each distance).

---

[6] Notational note: To indicate end locations for a pair of walking points, I will juxtapose letters.

2. For each point in the first 45° of Quadrant I (including **A**, **B**, and **C**), identify a second point anywhere on the state space for the current number of steps that is this distance away. Here, a special note must be considered for utilizing the symmetry in the next step. When identifying the second point from **A**, only look for points inside this 45° range. That pair will rotate with the symmetry.[7] For example, **AC**[8] will rotate to **AE**, **AG**, **AI**. But when using any other point inside the 45° range as the first point, consider points in the entire state space. For example, **CA** rotates to **EA**, **GA**, and **IA**.
3. For each pair noted in step 2, multiply the number of paths from the origin to the first point times the number of paths from the origin to the second point.
4. Sum the results of 3 and multiple by 4 for the symmetry to find the total number of ways the particular distance can occur.

I will demonstrate this algorithm for $d(\boldsymbol{p}_2, \boldsymbol{q}_2) = \sqrt{2}$.

1. There are 4 paths to **A**; 1 path to **B**, **D**, **F**, and **H**; and 2 paths to **C**, **E**, **G**, and **I**.
2. **A**: **AC**

   B: **BC**, **BI**

   C: **CA**, **CB**, **CD**
3. **AC**: (4)(2) = 8

   BC: (1)(2) = 2 **BI**: (1)(2) = 2

   CA: (2)(4) = 8 **CB**: (2)(1) = 2 **CD**: (2)(1) = 2
4. (8+2+2+8+2+2)(4) = 96

---

[7] Note: For state space after step 2, this will also require accounting for a reflection. This will be addressed later.

[8] For clarification, I am continuing the notation begun above for example when both p_n and q_n would end their walk at **A**, I recorded **AA**. In this case **AC**, p_n ends at **A** and q_n ends at **C**.

Hence, there are 96 different pairs of paths for $\boldsymbol{p}_2$ and $\boldsymbol{q}_2$ that will result in a $d(\boldsymbol{p}_2,\boldsymbol{q}_2)=\sqrt{2}$

To generate a list of all possible distances at a given step *n*, consider that the two particles always lie on the hypotenuse of a right triangle (possibly degenerate) in which the side lengths are 0, 2, … 2*n*. Therefore, by the Pythagorean Theorem,

$$d(\boldsymbol{p}_n,\boldsymbol{q}_n)\in\{D_n;D_n=\sqrt{a^2+b^2}\text{ for all }a,b\in\mathbb{N}\cup\{0\}\text{ such that }a+b=2c,\text{ with }c=0...n\}.$$

For *n* = 2, this yields pairs $(a,b)\in\{(0,0),(1,1),(0,2),(2,2),(1,3),(0,4)\}$ giving distances

$$\sqrt{0^2+0^2}=0,\sqrt{1^2+1^2}=\sqrt{2},\sqrt{0^2+2^2}=2,\sqrt{2^2+2^2}=2\sqrt{2},\sqrt{1^2+3^2}=\sqrt{10},\sqrt{0^2+4^2}=4.$$

The probabilities associated with each distance at step *n* =2 can be figured following the number of paths algorithm. Consider that there is always a $\frac{1}{4^n}$ chance that any one particle travels along any one path. Thus, the probabilities for the possible distances at step *n* =2 are:

$$P\left(d(p_2,q_2)=\begin{Bmatrix}0\\\sqrt{2}\\2\\2\sqrt{2}\\\sqrt{10}\\4\end{Bmatrix}\right)=\begin{Bmatrix}36\cdot\frac{1}{4^2}\cdot\frac{1}{4^2}\\96\cdot\frac{1}{4^2}\cdot\frac{1}{4^2}\\64\cdot\frac{1}{4^2}\cdot\frac{1}{4^2}\\24\cdot\frac{1}{4^2}\cdot\frac{1}{4^2}\\32\cdot\frac{1}{4^2}\cdot\frac{1}{4^2}\\4\cdot\frac{1}{4^2}\cdot\frac{1}{4^2}\end{Bmatrix}=\begin{Bmatrix}\frac{36}{256}\\\frac{96}{256}\\\frac{64}{256}\\\frac{24}{256}\\\frac{32}{256}\\\frac{4}{256}\end{Bmatrix}=\begin{Bmatrix}0.1406\\0.3750\\0.2500\\0.0938\\0.1250\\0.0156\end{Bmatrix}$$

For one sort of confirmation, see that $\frac{36}{256}+\frac{96}{256}+\frac{64}{256}+\frac{24}{256}+\frac{32}{256}+\frac{4}{256}=\frac{256}{256}=1$.

Compute the expectation for step *n* = 2 by multiplying distances by probabilities and summing:

$$E_2(\boldsymbol{X})=0\cdot\frac{36}{256}+\sqrt{2}\cdot\frac{96}{256}+2\cdot\frac{64}{256}+2\sqrt{2}\cdot\frac{24}{256}+\sqrt{10}\cdot\frac{32}{256}+4\cdot\frac{4}{256}=\frac{2\sqrt{10}+9(\sqrt{2}+1)}{16}\approx1.7533.$$

For the purposes of working toward a generalized algorithm, I will demonstrate the steps necessary to figure the expectation for the distance at step 3. First, consider the possible ending locations of a particle and the number of paths to each ending location. There are 16,[9] labeled A through P in Figure 12.

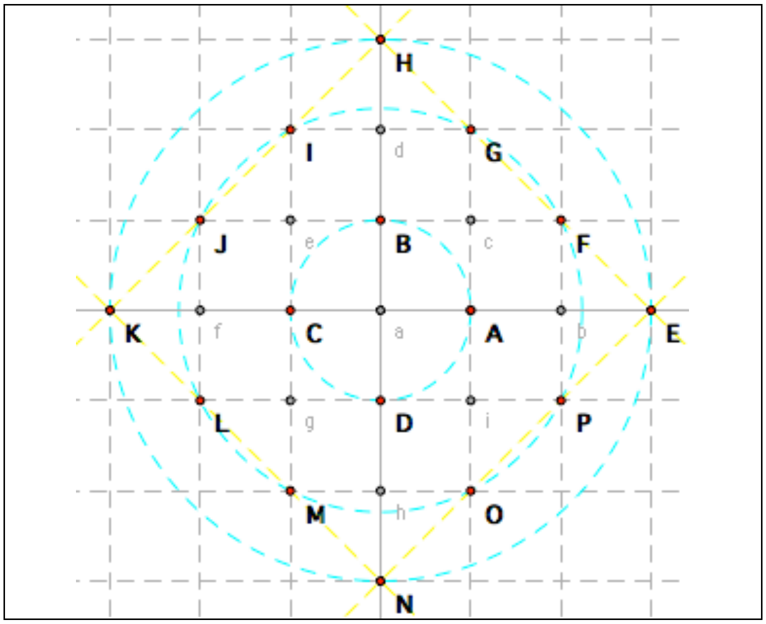


Figure 12. Possible end locations for a walking point in 2 dimensions at $n = 3$.

Next will be to count the number of paths (summarized in Table 2) to each of these possible ending locations. Again, this only must be done for the point in the first 45° of Quadrant I. This results in 9 paths to point **A**, 1 path to **E**, and 3 paths to **F**. Now is an appropriate time to more carefully discuss the symmetry of the state space $\mathbb{Z}^2$. Of course, if a possible ending point is the origin, this point is not repeated in the symmetry. Any point on the positive horizontal axis or the positive 45° ray repeats itself 4 times in the symmetry—all by the 90° rotation. Point **E** is an example in the $n = 3$ case (see Figure 11), while points **B** and **C** are examples in the $n = 2$ case (see Figure 10). All other points repeat themselves 8 times as they flip across the 45° ray in addition to being rotated 90°. Point **F** is an example of this in the $n = 3$ case. **F** flips to **G**, then both **F** and **G** rotate to **I** and **J**, **L** and **M**, and **O** and **P**. These repeats save some by-hand counting by careful multiplication of paths counted in the first 45° of Quadrant I.

---

[9] For n steps, there will always be $(n + 1)^2$ possible ending locations.

Table 2

Number of Paths from Origin to Points in State Space $\boldsymbol{X}_3$

point	# paths	example list								
A (**B C D**)	9	**A-a-A**	**A-b-A**	**A-c-A**	**A-i-A**	**B-a-A**	**B-c-A**	**C-a-A**	**D-a-A**	**D-i-A**
E (**H K N**)	1	**A-b-E**								
F (**G I J L M O P**)	3	**A-b-F**	**A-c-F**	**B-c-F**						

The possible distances after $n = 3$ steps are

$$d(\boldsymbol{p}_3, \boldsymbol{q}_3) \in \{D_3; D_3 = \sqrt{a^2 + b^2} \text{ for all } a, b \in \mathbb{N} \cup \{0\} \text{ such that } a + b = 2c, \text{ with } c = 0...3\}.$$

$a + b = 2c$, with $c = 0..3$ yields the pairs

$$(a,b) \in \{(0,0),(1,1),(0,2),(2,2),(1,3),(0,4),(3,3),(2,4),(1,5),(0,6)\}$$

giving distances

$$d(\boldsymbol{p}_3, \boldsymbol{q}_3) \in \{0, \sqrt{2}, 2, 2\sqrt{2}, \sqrt{10}, 4, 3\sqrt{2}, 2\sqrt{5}, \sqrt{26}, 6\}.$$[10]

Distance zero occurs in the first 45° of Quadrant I when the two particles end at **AA**, **EE**, or **FF**. There are (9)(9) = 81 pairs of paths for particles $\boldsymbol{p}_3$ and $\boldsymbol{q}_3$ to arrive at **AA**; (1)(1) = 1 for **EE**; and (3)(3) = 9 for **FF**. Multiply **AA** and **EE** by 4 and **FF** by 8 to account for symmetries, gathering all possible pairs of paths that yield a distance of 0.

$$(4)(81) + (4)(1) + (8)(9) = 400$$

Table 3 summarizes the similar work necessary to sum the total pairs of paths yielding a distance of $\sqrt{2}$. Similar work must be carried out to compute the total pairs of paths yielding each distance. The resulting probabilities for all possible distances after $n = 3$ steps are

---

[10] Another interesting note, the number of possible distances at step n is the $(n + 1)^{\text{th}}$ triangular number.

$$P\left(d(\boldsymbol{p}_3,\boldsymbol{q}_3)=\left\{\begin{bmatrix} 0 \\ \sqrt{2} \\ 2 \\ 2\sqrt{2} \\ \sqrt{10} \\ 4 \\ 3\sqrt{2} \\ 2\sqrt{5} \\ \sqrt{26} \\ 6 \end{bmatrix}\right\} \right) = \left\{\begin{bmatrix} \frac{400}{4096} \\ \frac{1200}{4096} \\ \frac{900}{4096} \\ \frac{480}{4096} \\ \frac{720}{4096} \\ \frac{144}{4096} \\ \frac{80}{4096} \\ \frac{120}{4096} \\ \frac{48}{4096} \\ \frac{4}{4096} \end{bmatrix}\right\} \approx \left\{\begin{bmatrix} 0.0977 \\ 0.2930 \\ 0.2197 \\ 0.1172 \\ 0.1758 \\ 0.0352 \\ 0.0195 \\ 0.0293 \\ 0.0117 \\ 0.0010 \end{bmatrix}\right\}.$$

Using these results leads to an expectation

$$E_3(\boldsymbol{X}) = \frac{2400+2400\sqrt{2}+240\sqrt{5}+720\sqrt{10}+48\sqrt{26}}{4096} \approx 2.1612$$

Table 3

Total 2-Point Paths Yielding Distance $\sqrt{2}$

ending points	# paths to first endpoint	# paths to second endpoint	symmetry family multiplier (first point determines)	result
AB	9	9	4	324
AD	9	9	4	324
AF	9	3	4	108
AP	9	3	4	108
EF	1	3	4	12
EP	1	3	4	12
FA	3	9	8	216
FE	3	1	8	24
FG	3	3	8	72
TOTAL				1200

The same computational process resulted in the following probabilities for time step 4, with results summarized in Table 4.

$$P\left(d(\boldsymbol{p}_4, \boldsymbol{q}_4) = \left\{ \begin{matrix} 0 \\ \sqrt{2} \\ 2 \\ 2\sqrt{2} \\ \sqrt{10} \\ 4 \\ 3\sqrt{2} \\ 2\sqrt{5} \\ \sqrt{26} \\ 6 \\ 4\sqrt{2} \\ \sqrt{34} \\ 2\sqrt{10} \\ \sqrt{50} \\ 8 \end{matrix} \right\} \right) = \left\{ \begin{matrix} \frac{4900}{65536} \\ \frac{15680}{65536} \\ \frac{12544}{65536} \\ \frac{7840}{65536} \\ \frac{12544}{65536} \\ \frac{3136}{65536} \\ \frac{2240}{65536} \\ \frac{3584}{65536} \\ \frac{1792}{65536} \\ \frac{256}{65536} \\ \frac{280}{65536} \\ \frac{448}{65536} \\ \frac{224}{65536} \\ \frac{64}{65536} \\ \frac{4}{65536} \end{matrix} \right\} \approx \left\{ \begin{matrix} 0.075768 \\ 0.239258 \\ 0.191406 \\ 0.119629 \\ 0.191406 \\ 0.047852 \\ 0.034180 \\ 0.054688 \\ 0.027344 \\ 0.003906 \\ 0.004272 \\ 0.006836 \\ 0.003418 \\ 0.000977 \\ 0.000061 \end{matrix} \right\}, \text{ with } E_4(\boldsymbol{X}) \approx 2.5017.$$[11]

Table 4

Expectation After N-Steps for the Distance between Two Walking Points on $\mathbb{Z}^2$

n number of steps	E_n expectation
0	0
1	1.2071
2	1.7533
3	2.1612
4	2.5017

It is apparent that as n increases, so does the expectation. Again, creating a plot of this data reveals a power function, $f(x) = x^a$, manner of growth with $0 < a < 1$ (see Figure 13).

---

[11] It may be worth noting these distances are no longer recorded in numerical order, but continue to be recorded by subtending the new $n + 1$ distances for each step n.

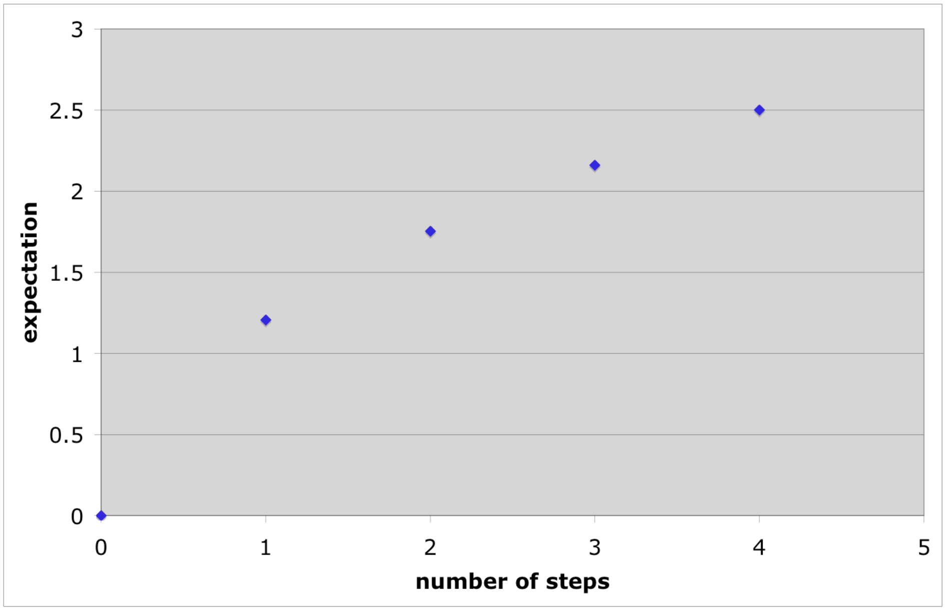


Figure 13. Distance expectation between 2 walking points in 2 dimensions.

This expectation curve again resembles more precisely a square root curve, with $a \approx 0.5$. Allowing a TI-84 to complete a power regression on the data [except for point (0,0)] yielded the following function:

$$E(n) = 1.210752n^{0.526230} \qquad r = 0.999908,$$

where the r is a correlation coefficient. The behavior of the expectation function for distance between two points in 2 dimensions is, not surprisingly, very similar to the expectation function for the distance between two points in 1 dimension. As I collected data for larger n's, the value of a more nearly approximated 0.5. I suspect the same would occur here. The growth behavior between 1 and 2 dimensions is likely to be on the same order, only the coefficient should be different. It is possibly surprising how close the values of the average distances in 2 dimensions are to their corresponding steps in the 1 dimensional case.

The algorithm to determine the probability for each possible distance, and then ultimately the expectation, for any step n is summarized as follows:

1. Determine all possible points after n steps in the first 45° of Quadrant I.
2. Determine the number of paths to each of these points.
3. Calculate all possible distances between the ending points identified in step 1, and all other ending points created both by the reflection of the points about $y = x$, and then all of these Quadrant I points rotated 90° counterclockwise three times.
4. For each distance, do as in Table 3 (summarized here):
   a. List all pairs of points, where $\boldsymbol{p}_n$—the first point—remains in the first 45° of Quadrant I.
   b. Multiply the number of paths to the first point by the number of paths to the second point by the symmetry family of the first point.
   c. Sum 4b for each 4a. This totals the number of ways this distance can occur.
5. Figure the probability for each distance to occur is equal to the number of ways it can occur (result of 4c above) divided by 4^{2n}.
6. The expectation for this time step is the sum of the products of the distances and their probabilities.

I have yet to devise a direct method to complete steps 2 and 4a, leaving a closed-form formula for this process yet to be created.

Triangle Area Determined by 3 Walking Points on $\mathbb{Z}^2$

I took the same approach used for the distance between two points to the situation of attempting to determine the triangular area determined by 3 points on 2 dimensions. I will introduce a heuristic that has guided my reasoning through this portion of the investigation. This heuristic is founded basically in the algorithm to compute the expectation for the average distance between two points in $\mathbb{Z}^2$, and in particular in the work evident in Table 3. First, recall the expectation is

$$E(\boldsymbol{X}) = \sum \boldsymbol{x} p(\boldsymbol{x})$$

Adding details for the purpose of the heuristic,

$$E(\boldsymbol{X}) = \sum_{\substack{\text{all possible areas} \\ \text{determined by } \boldsymbol{p}_n, \boldsymbol{q}_n, \boldsymbol{r}_n}} A \cdot \left[\frac{\displaystyle\sum_{\text{all } \boldsymbol{p}_n, \boldsymbol{q}_n, \boldsymbol{r}_n \text{ with this } A} \left[\text{paths}_{\boldsymbol{p}} \cdot \text{paths}_{\boldsymbol{q}} \cdot \text{paths}_{\boldsymbol{r}} \cdot S \right]}{4^{3n}} \right].$$

To begin to clarify meaning in this heuristic, *A*, area, is our measure, and what lies inside the outer brackets is the computation for probability for each of these possible measures. $\boldsymbol{p}_n$, $\boldsymbol{q}_n$, $\boldsymbol{r}_n$ refers to the location of the three points ***p***, ***q***, ***r*** after *n* steps. *A* is the area determined by $\boldsymbol{p}_n$, $\boldsymbol{q}_n$, $\boldsymbol{r}_n$. *A* is a rather straight forward set at any step *n*. "$\text{paths}_{\boldsymbol{p}}$", "$\text{paths}_{\boldsymbol{q}}$", and "$\text{paths}_{\boldsymbol{r}}$" are the number of possible paths from the origin to the point where particle ***p***, ***q***, or ***r*** end after *n* steps. This is identical to the computation for the number of paths to any possible point after *n* steps as determined in the 2 points in $\mathbb{Z}^2$ investigation. *S* is intended to be the symmetry family multiplier. This is a construct that may aid in the actual counting of the triangles, relying on the symmetry of the dihedral group structure of $\mathbb{Z}^2$. 4^{3n} counts the total number of paths possible for ***p***, ***q***, and ***r*** after *n* steps (each particle has 4^n possible paths).

The set of areas after *n* steps is possible to figure. Considering the triangle formed by the three endpoints of particles ***p***, ***q***, and ***r***, the area can be computed as one-half the cross product of any two of the sides. That is,

$$A = 0.5 \cdot \left\| \begin{matrix} \boldsymbol{r}_x - \boldsymbol{p}_x & \boldsymbol{r}_y - \boldsymbol{p}_y \\ \boldsymbol{q}_x - \boldsymbol{p}_x & \boldsymbol{q}_y - \boldsymbol{p}_y \end{matrix} \right\|.$$

It is next to be done in this investigation to define the set of possible areas formed by the three particles after n steps. However, this has proven an elusive question. My initial conjecture is that all (and only) integral areas $\{0,1,\ldots,n^2\}$ are possible for step n. This held true for steps $n = 0,1,2$. However at step 3, the conjecture suggests that all integral areas {0,1,2,3,4,5,6,7,8,9} should be possible, yet I have not found a triangle on the $n = 3$ state space with areas 6, 7, or 8.[12] Although I have yet to explore this further, I have come to an even greater hurdle: How to count the number of possible triangles formed by the placement of particles $\boldsymbol{p}_n$, $\boldsymbol{q}_n$, $\boldsymbol{r}_n$ on $\boldsymbol{X}_n$.

Area of the Random Triangle on $\mathbb{Z}^2$

As I arrived at the apparent uncrossable hurdle with the problem above, I wondered what relation my original question has to the question, "What is the average area of the triangle determined by three random points selected on $\mathbb{Z}^2$?" I explored this question by narrowing it to ask the average area of a triangle determined by three random points selected on a finite $m \times n$ rectangle in $\mathbb{Z}^2$. Again, the average area—that is the expected value—would be computed by summing the product of areas and the probability that each of these areas occurs:

$$E(\boldsymbol{X}) = \sum \boldsymbol{x} p(\boldsymbol{x}),$$

where $\boldsymbol{x}$ is an element of the set of possible areas determined by $m \times n$, and $p(\boldsymbol{x})$ is the probability this element occurs. For example, for $m = 3$ and $n = 2$, $\boldsymbol{x} \in \{0,\tfrac{1}{2},1,\tfrac{3}{2},2,\tfrac{5}{2},3\}$.[13] But again, the task to count the probabilities that each area occur became very difficult. In particular, counting the number of ways three randomly selected points on an $m \times n$ lattice create a triangle

---

[12] I have yet to explore this further, but my current conjecture is that for step size n, the possible areas are $\{0,1,\ldots,\frac{n+1}{2},n\}$, adjusting the $\frac{n-1}{2}$ when n is even in some way. I suspect that it is not possible for this area set to include anything but integers. However, this is also an unproven conjecture.

[13] Open question: are all possible areas the numbers from 0 to the greater of m or n, counting by ½'s?

of some particular area was the challenge. As I bumped into this roadblock, I realized the possibility of an alternative strategy for computing the expectation—a brute force method employing a programmable computing tool.

I could create a program that would step each of the three points, one at a time, through each possible location on the $m \times n$ lattice, computing the area of the triangle formed at every step. To figure the average area of the randomly selected triangle on the $m \times n$ lattice would be a matter of summing the product of each resulting area and its associated probability—a much easier computation. Notationally, the average area would be computed:

$$E(\boldsymbol{X}) = \sum_{\substack{\text{all possible} \\ \text{combinations } \boldsymbol{p},\boldsymbol{q},\boldsymbol{r}}} A_{\boldsymbol{p},\boldsymbol{q},\boldsymbol{r}} \cdot P(A_{\boldsymbol{p},\boldsymbol{q},\boldsymbol{r}}) \text{ with } \boldsymbol{p},\boldsymbol{q},\boldsymbol{r} \text{ points in the } m \times n \text{ rectangle.}$$

To begin to enact this heuristic, consider the $m \times n$ lattice placed in the 1st quadrant of the coordinate plane, from the origin to the point (m, n) with $m, n \in \mathbb{N} \cup \{0\}$. Name ($\boldsymbol{p}_x$, $\boldsymbol{p}_y$), ($\boldsymbol{q}_x$, $\boldsymbol{q}_y$), ($\boldsymbol{r}_x$, $\boldsymbol{r}_y$) to be the coordinates of points $\boldsymbol{p}$, $\boldsymbol{q}$, $\boldsymbol{r}$ in or on the boundary of this $m \times n$ lattice, respectively. $A_{p,q,r} = 0.5 \cdot \left\| \begin{matrix} \boldsymbol{r}_x - \boldsymbol{p}_x & \boldsymbol{r}_y - \boldsymbol{p}_y \\ \boldsymbol{q}_x - \boldsymbol{p}_x & \boldsymbol{q}_y - \boldsymbol{p}_y \end{matrix} \right\|$. And $P(A_{p,q,r}) = \frac{1}{(m+1)^3 (n+1)^3}$. The Maple (Maplesoft, 2005) code for this brute force program follows as Figure 14. I ran this program to obtain several results for rather small rectangles. With small m, n the program does not take long to run. However, as evident in the degree 6 = 3 + 3 denominator, the number of possible outcomes that must be tested increases cubically for each increase in m or n. This makes for a heavy drain on computing resources. Notice that a 10-fold increase in the m and in the n creates a $10^3 \cdot 10^3 = 1{,}000{,}000$-fold increase in computing time. As a result, I was very limited, for now, in use of this program. Table 5 reports the ratio of the expected area to the area of the whole rectangle for the small set of possible m, n pairs that I have currently computed.

```
randtri := proc( m,n )
  local area, expect, probab, M, px, py, qx, qy, rx, ry;
  probab := 1.0/(((m+1)^3)*(n+1)^3);
  area := 0;
  expect := 0;
  for px from 0 to m do
    for py from 0 to n do
      for qx from 0 to m do
        for qy from 0 to n do
          for rx from 0 to m do
            for ry from 0 to n do
              M := Matrix ([[rx-px,ry-py],[qx-px,qy-py]]);
              area:=abs(.5*Determinant(M));
              expect := expect+probab*area;
            end do;
          end do;
        end do;
      end do;
    end do;
  end do;
  print(expect/(m*n));
end proc:
```

Figure 14. Maple code to compute average area of 3 points on $m \times n$ lattice.

Table 5

Average Area Ratio of a Random Triangle an $m \times n$ *Rectangle*

m	n	$\frac{A_{p,q,r}}{m \cdot n}$	m	n	$\frac{A_{p,q,r}}{m \cdot n}$	m	n	$\frac{A_{p,q,r}}{m \cdot n}$
1	1	0.1875	3	1	0.1563	5	1	0.1458
1	2	0.1667	3	2	0.1343	5	2	0.1247
1	3	0.1563	3	3	0.1243	5	3	0.1152
1	4	0.1500	3	4	0.1188	5	4	0.1097
1	5	0.1458	3	5	0.1152	5	5	0.1063
1	6	0.1429	3	6	0.1126	5	6	0.1038
2	1	0.1667	4	1	0.1500	6	1	0.1429
2	2	0.1440	4	2	0.1284	6	2	0.1220
2	3	0.1343	4	3	0.1188	6	3	0.1126
2	4	0.1284	4	4	0.1132	6	4	0.1073
2	5	0.1247	4	5	0.1097	6	5	0.1038
2	6	0.1220	4	5	0.1073	6	6	0.1015

Of course the area ratio entry at (3, 4) is equal to the entry at (4, 3). An interesting reorganization of the table found by looking at increasing values of $m \cdot n$ is presented in Table 6. Observing the behaviors in this table, I noted that the area ratio continues to decline, with a few bumps along the way. Interestingly these bumps, i.e. increases in average area, occur is when the next rectangle is significantly more thin. This feels counterintuitive because there seems to be such a larger portion of possible triangles of area = 0 or almost 0 in the thinner $m \times n$ triangles. As a quick further investigation, I ran the program for a 1×12 rectangle (area = 12). The expectation ratio was 0.1346—following in line with this observation. And finally, while the area ratio is decreasing, does it approach some number other than zero as $m \cdot n$ increases? Although this data leads me to suspect it may, it is a very small data set to draw conclusions for large *m*, *n*.

Table 6

Average Area Ratio of a Random Triangle an $m \times n$ *Rectangle, by Increasing* $m \cdot n$

m	n	$m \cdot n$	$\frac{A_{p,q,r}}{m \cdot n}$	m	n	$m \cdot n$	$\frac{A_{p,q,r}}{m \cdot n}$
1	1	1	0.1875	2	5	10	0.1247
1	2	2	0.1667	2	6	12	0.1220
1	3	3	0.1563	3	4	12	0.1188
1	4	4	0.1500	3	5	15	0.1152
2	2	4	0.1440	4	4	16	0.1132
1	5	5	0.1458	3	6	18	0.1126
1	6	6	0.1429	4	5	20	0.1097
2	3	6	0.1343	4	6	24	0.1073
2	4	8	0.1284	5	5	25	0.1063
3	3	9	0.1243	6	5	30	0.1038
2	5	10	0.1247	6	6	36	0.1015
2	6	12	0.1220				

Although the question that drove this inquiry begins to stray from my original about three particles on a random walk, several aspects connect to my original situation. I am considering the area determined by three points, on a bounded 2-dimensional integer lattice. More importantly though, is that the computational method employed here suggests additional options for figuring the average area of the three particles on the random walk. If I could develop an algorithm to compute the expectation of the area determined by three particles on a random walk in 2 dimensions and translate it into a program, I could explore further the original problem.

Area of the Random Triangle on $\boldsymbol{X}_n$

Next I turned my attention to a task near to the original question posed in this research. I decided to adapt the previous question, "What is the average area of the triangle determined by three random points selected on $\mathbb{Z}^2$?", to ask: "What is the average area of the triangle determined by three points randomly selected on the state space $\boldsymbol{X}_n$?" $\boldsymbol{X}_n$ is the subset of $\mathbb{Z}^2$ where a particle beginning at the origin could end after an *n*-step random walk. As I have assumed until now, this random walk is defined by the characteristic that at each step, the particle has an equal likelihood of traveling one unit N, S, E, or W. In other words, the transition function $P(\boldsymbol{x},\boldsymbol{y})$ for $\boldsymbol{x},\boldsymbol{y} \in \mathbb{Z}^2$ would be such that $\boldsymbol{y} \in \{\boldsymbol{x}+(1,0), \boldsymbol{x}+(0,1), \boldsymbol{x}+(-1,0), \boldsymbol{x}+(0,-1)\}$. As I did in the previous situation, I adopted my heuristic for the area expectation so as to compute and sum the areas for every possible placement of particles *p*, *q*, and *r*. Notationally, I am again thinking to compute the average area by:

$$E(\boldsymbol{X}) = \sum_{\substack{\text{all possible} \\ \text{combinations } \boldsymbol{p},\boldsymbol{q},\boldsymbol{r}}} A_{\boldsymbol{p},\boldsymbol{q},\boldsymbol{r}} \cdot P(A_{\boldsymbol{p},\boldsymbol{q},\boldsymbol{r}}) \text{ with } \boldsymbol{p},\boldsymbol{q},\boldsymbol{r} \text{ points in the state space } \boldsymbol{X}_n.$$

This state space, $\boldsymbol{X}_n$ is not the simple integer lattice $\mathbb{Z}^2$, but instead a diamond bounded at step *n* by the lines $y_1 = -n + x,\ y_2 = n + x,\ y_3 = -n - x,$ and $y_4 = n - x$. The points in $\boldsymbol{X}_n$ are determined by the intersections of two sets of parallel, diagonal lines:

$$\{y = -n + x, y = (-n+2) + x, \ldots, y = n + x\} \cap \{y = -n - x, y = (-n+2) - x, \ldots, y = n - x\}.$$

A few observations about the state space $\boldsymbol{X}_n$: the number of points in the set is equal to $(n+1)^2$, and the origin will be an element of $\boldsymbol{X}_n$ only when *n* is even.

To enact this heuristic as a programmed algorithm, I imagined moving each particle beginning at the western corner (-*n*, 0) along the northeast line. When this particle reached the northernmost point (0, *n*), I restarted it one diagonal lower (-*n*+1, -1) and again moved in to the northeast. This continued until that particle began at the southernmost point (0, -n) and ended at the eastern corner (n, 0). Of course, the algorithm required that each particle make this journey with each of the other particles at every point along the journey. At each instance for the possible location of these three particles, $\boldsymbol{p}$, $\boldsymbol{q}$, and $\boldsymbol{r}$, I computed the area and kept a running total for a partial expectation. Again, name $(\boldsymbol{p}_x, \boldsymbol{p}_y)$, $(\boldsymbol{q}_x, \boldsymbol{q}_y)$, $(\boldsymbol{r}_x, \boldsymbol{r}_y)$ to be the coordinates of points $\boldsymbol{p}$, $\boldsymbol{q}$, $\boldsymbol{r}$ in $\boldsymbol{X}_n$. $A_{\boldsymbol{p},\boldsymbol{q},\boldsymbol{r}} = 0.5 \cdot \left\| \begin{matrix} \boldsymbol{r}_x - \boldsymbol{p}_x & \boldsymbol{r}_y - \boldsymbol{p}_y \\ \boldsymbol{q}_x - \boldsymbol{p}_x & \boldsymbol{q}_y - \boldsymbol{p}_y \end{matrix} \right\|$. And $P(A_{\boldsymbol{p},\boldsymbol{q},\boldsymbol{r}}) = \frac{1}{(n+1)^6}$. The Maple (Maplesoft, 2005) code to enact this algorithm is reported in Figure 15, with results for step sizes n = 1..10 in Table 7.

```
reset;
with(LinearAlgebra):
randwalktri := proc( n )
 local area, expect, probab, M, px, py, qx, qy, rx, ry, i, j, k;
 probab := 1/(n+1)^6;
 area := 0;
 expect := 0;
 for i from 0 to n do
  for px from (-n+i) to (0+i) do
   for j from 0 to n do
    for qx from (-n+j) to (0+j) do
     for k from 0 to n do
      for rx from (-n+k) to (0+k) do
       py := (n-2*i) + px;
       qy := (n-2*j) + qx;
       ry := (n-2*k) + rx;
       M := Matrix ([[rx-px,ry-py],[qx-px,qy-py]]);
       area:=abs(.5*Determinant(M));
       expect := expect+probab*area;
      end do;
     end do;
    end do;
   end do;
  end do;
 end do;
 print(expect);
end proc:
```

Figure 15. Maple code to compute average area of 3 random points on $\boldsymbol{X}_n$.

Table 7

Average Area of All Possible Triangles on $\boldsymbol{X}_n$

step size n	Expectation $E(\boldsymbol{X}_n)$
1	0.3750
2	1.1523
3	2.2383
4	3.6234
5	5.3133
6	7.3047
7	9.6017
8	12.2027
9	15.1086
10	18.3193

The most immediately obvious result is that the expectation is growing, apparently without bound. In fact, this observation is an important distinction from what I suspect will occur when the computation is adjusted to not make each point of $\boldsymbol{X}_n$ equally likely to occur, but rather respect the possibilities determined by the structure of the random walk, where points are more likely to be found closer to the particle's origination.

Return to "Triangle Area Determined by 3 Points on $\mathbb{Z}^2$"

Finally I returned to the context of the original problem: What would be the average area of 3 particles on a random walk? However, I will not yet account for the lack of spatial homogeneity of the GSP microworld. The work on the random triangle in $\mathbf{X}_n$ led me to develop a manner by which to analyze the behaviors of the three points on the $\boldsymbol{X}_n$ state space. However, the probabilities associated with the location of the particles were not true to the random walk. In the previous effort, the particles were considered to occur at any point of $\boldsymbol{X}_n$ with equal probability. In the random walk, the behavior of the particles is not exactly this. Imagining a sort of probability cloud over the state space, this cloud would be most dense near the particle's origin.

To model this behavior, I considered 3 randomly selected points. In this initial exploration, I restricted the possible location of the initial location of the particles to a rectangularly-bounded planar region. Once these points were selected, each particle followed a random walk for *s* steps. I computed the triangular area determined by the particles at each step, kept a running total, and computed the expectation after some finite number (*s*) of steps. The Maple (Maplesoft, 2005) code is presented in Figure 16.

```
reset;
with(LinearAlgebra): with(RandomTools):
randwalkpts := proc(m,n,s)
 local area, expect, expectratio, M, px, py, qx, qy, rx, ry, i, j, k, l;
 area := 0;
 expect := 0;
 px := Generate(integer(range=0..m));
 py := Generate(integer(range=0..n));
 qx := Generate(integer(range=0..m));
 qy := Generate(integer(range=0..n));
 rx := Generate(integer(range=0..m));
 ry := Generate(integer(range=0..n));
 for i from 1 to s do
  j := Generate(integer(range=(1..4)));
  if j=1 then py := py + 1
   elif j=2 then px := px + 1
   elif j=3 then py := py - 1
   elif j=4 then px := px - 1
  end if:
  k := Generate(integer(range=(1..4)));
  if k=1 then qy := qy + 1
   elif k=2 then qx := qx + 1
   elif k=3 then qy := qy - 1
   elif k=4 then qx := qx - 1
  end if:
  l := Generate(integer(range=(1..4)));
  if l=1 then ry := ry + 1
   elif l=2 then rx := rx + 1
   elif l=3 then ry := ry - 1
   elif l=4 then rx := rx - 1
  end if:
  M := Matrix ([[rx-px,ry-py],[qx-px,qy-py]]);
  area:=area + abs(.5*Determinant(M));
 end do;
 expect := area/s;
 expectratio := expect/(m*n);
 print(expectratio);
end proc:
```

Figure 16. Maple code to compute average area of 3 walking points on $\mathbb{Z}^2$.

Table 8 summarizes the results of running the program, across some variations on initial window size and number of steps run. I began with a $120{,}000 \times 100{,}000$ window because I knew my algorithm would allow the points to wander off the screen. I imagined such a large starting window would in some ways minimize this effect. My next trials were on a window much more similar to the window size used in my initial playing with this problem in the GSP screen. I estimated the GSP window to be pixels. Initially I ran this 10 times at 1000 steps each. Because the run did not take a very long time, I increased the number of steps to 10,000 in order to get a more accurate estimate.

Table 8

Average Area Ratio of a Walking Triangle with Random Initial Vertices, Small n

Window: $120{,}000 \times 100{,}000$ Steps 1000		Window: 800×600 Steps 1000		Window: 800×600 Steps 10,000	
Trial Number	Expectation $E(\boldsymbol{X}_n)$	Trial Number	Expectation $E(\boldsymbol{X}_n)$	Trial Number	Expectation $E(\boldsymbol{X}_n)$
1	0.0424	1	0.1720	1	0.1204
2	0.1651	2	0.1417	2	0.1417
3	0.0994	3	0.0459	3	0.0166
4	0.0202	4	0.0437	4	0.0888
5	0.0563	5	0.0047	5	0.0085
6	0.0661	6	0.0031	6	0.0397
7	0.1164	7	0.0109	7	0.2870
8	0.0786	8	0.0292	8	0.1213
9	0.0779	9	0.1080	9	0.0301
10	0.0382	10	0.0107	10	0.1551
average	0.0761	average	0.0570	average	0.1009

The results of this experimental approach to the problem, reported in Table 8, support the initial estimates provided by GSP that this area seems to hover between 5 and 10% of the bounding window. This approach does allow for following the triangular area determined by 3 randomly chosen points on a plane. It will compute the average area over any finite number of steps. This program does limit the initial random choice of the points to be confined to an $m \times n$

window. Further, this window size is used to compute the area ratio—that is, what portion of this rectangular window area is the triangle. However, this restriction causes some mathematical “errors”. In particular, this ratio is a bit misleading. Because the points can freely walk outside the bounds of the window, the triangle area can meander outside this rectangle. So at best, the ratio is more a “relation to” than a “part of”. But that this program does not yet encourage the points to remain within a bounded window also fails to emulate the initial GSP program as well. In the GSP environment, the points are pushed back into the $m \times n$ window as they near the borders. In that case, the triangle area should remain smaller, decreasing the ratio.

CHAPTER 4

EMERGENT QUESTIONS

Several, in fact many questions emerged through the course of this inquiry. For example, in Chapter 1 I asked what is the probability of selecting 3 points in $\mathbb{Z}^2$, 2 of which are collinear? When considering all of $\mathbb{Z}^2$, I suspect the answer to be that the probability is zero. But when restricting to a finite subset of $\mathbb{Z}^2$, as is done in much of this investigation, the probability measure becomes significant (i.e. non-zero). The number of triangles of area 0 on a finite subset of $\mathbb{Z}^2$ is certainly a countable phenomenon. I relegated some of these sorts emerging questions and initial ideas to footnotes in the discussions in previous chapters. Other questions are cut out a bit more clearly, and occasionally pursued slightly, below.

One of the first questions opened in this exploration pointed to what seems to be a contradiction. If it is true that the average center of a point on a random walk on 2 dimensions remains it's starting point, than wouldn't the average area of the triangle determined by 3 randomly walking points be closely related (if not equal) to the area determined by the initial locations of these three points? One reason this seems contradictory to this point in my investigation is from the data collected in the GSP environment.[14] However, this data may be unreliable because it is likely that the "force" programmed near the boundaries of the GSP

---

[14] I ran several instances of the Maple (Maplesoft, 2005) program code in order to report the initial area ratio and final average area ratio. These results seem to contradict the observations from the GSP environment. Note, the Maple code suggests that the initial area and average areas are quite closely related. This data is presented in Chapter 5 where I take this up in a related question.

window has a much greater effect on this question about expectation than I originally guessed. Exploring this oddity further: Again, accepting that the average location of a point on a simple random walk in one dimension is its starting location, one might conjecture that the distance between any two points would remain 0. However, as proven in this thesis, that distance grows on the order $\sqrt{n}$, with coefficient $\frac{2}{\sqrt{\pi}}$.

Taking this observation into 3 walking points on 2 dimensions yields an interesting conjecture. If any two points are on average *k* apart at step *n*, pairwise, 3 points would each average this distance—forming a sort of equilateral triangle with side lengths *k*. Such a triangle has area $\frac{1}{2}\cdot(k)\cdot\left(\frac{\sqrt{3}}{2}k\right)=\frac{\sqrt{3}}{4}k^2$.

A third idea for later pursuit: In the case of the distance between 2 walking points on one dimension investigation, Wikipedia reports that "the average straight-line distance between start and finish points of a random walk of *n* steps is on the order of $\sqrt{n}$, or more precisely, its asymptote converges to $\sqrt{\frac{2n}{\pi}} \approx 0.8\sqrt{n}$". Although this does not disagree with my proof that the expectation for the distance between two points on a random walk is $2\sqrt{\frac{n}{\pi}}$. An interesting question is why these two findings relate by exactly a factor of $\sqrt{2}$? The first may be thought of as the average distance of one point from the origin while the second is the average distance between 2 points (see Figure 17). How must these expectations be related? This is to remain an open question, for now.

Figure 17. One versus two walking points.

Fourth: I suggested in Chapter 3 that there is a definable set of possible areas on the state space $\boldsymbol{X}_n$, the lattice created by the potential ending locations of a point on a random walk after *n* steps. I do suspect a few constraints for this set, but have yet to prove any of these. My current conjecture is that for step size *n*, the possible areas are $\{0,1,\ldots,\frac{n+1}{2},n\}$, adjusting the $\frac{n-1}{2}$ when *n* is even in some way. I also suspect that it is not possible for this set of areas to include anything but integers. These unproven conjectures remain, for now, an interesting curiosity.

Along a similar vein, I found a similar open question residing in the set of areas possible when randomly selecting 3 points on an $m \times n$ lattice. I currently hypothesize that all the possible areas are $\{0,\frac{1}{2},1,\frac{3}{2},2,\ldots,k\}$ where *k* is the larger of the two values *m*, *n*. In particular, are all the "½-areas" on this set possible? Also left to prove is whether there are areas in this set other than integral and these ½-areas.

A fifth open question is about the rate at which the average areas reported in Table 7 increase. Because they are areas, a quadratic increase seems plausible. A quadratic regression yields the best-fit function $E(n) = 0.1517n^2 + 0.3244n - 0.1015$ with $r = 0.9999999$. A more general power regression yields the best-fit function $E(n) = 0.3646 \cdot n^{1.6735}$ with $r = 0.9998271$. I believe that either more data or another sort of analysis should be able to confirm that this growth behavior is likely quadratic.

And finally, the last open question is about adjusting the program code to keep in accord with the GSP design team's goal that

> Points, and objects constructed based on them, tend to become much less interesting when they leave the screen. Therefore points in random motion should be discouraged from leaving the screen. If they have left the screen they should be encouraged to return. (Steketee, May 2, 2006, via email)

The effect this has is to maintain a truly random quality to N, S, E, or W decision-making of the walk. That is, unless the point is very near or outside the boundary of the viewing window. This could be done by creating a new set of rules for a point that walks into a particular zone—possibly the outer 1% (by length) and beyond region of the actual initial $m \times n$ rectangle, as sketched in grey of Figure 18.

Figure 18. Zone (grey) in which a walking point is encouraged back toward window.

If a point enters this zone, the random walk takes on a new set of rules. Instead of being equally likely to step in any one of the 4 directions, the point might have a 0.40 chance to step back toward the center of the $m \times n$ rectangle and 0.10 chance to travel further outwards. The side-step probabilities could remain at 0.25. There is no particular reason for the probabilities I assigned. They could be easily modifiable based on the actual algorithm used by the GSP programmers, if that could be known.

Further, the rules could be a bit more complex, acting like a gravitational force rather than a sudden change. The programmers seemed to have a force of this nature in mind when designing the microworld.

> The simplest way to accomplish these two objectives is to define a “force” that pushes points toward the center of the screen if they are off-screen or near the boundary. The strength of the force is zero when points are safely on-screen,

> increases gradually when they near the edge, and continues to increase if they have gone off-screen. Once the point is a certain distance outside the boundary, the force reaches and stays at its maximum strength. (Steketee, May 2, 2006, via email)

Again, once a point got to some region near or beyond the initial $m \times n$ boundary, the probabilities could adjust as a function based on the distance from the center. The benefits of exploring these sorts of adaptations to the model are to more closely fit the design of the program and the intentions of the author. Or, said another way, to more closely reflect the situation that drove the initial question in this investigation.

CHAPTER 5

SUMMARY AND CONCLUDING REMARKS

This thesis was initiated by an experience in a computer-based geometry microworld, using software called *The Geometer's Sketchpad* (GSP). Upon seeing the point animation feature of the software, I began wondering about the behaviors of a triangle determined by three animated points. Applying the mathematics of geometry, probability theory, counting, and random walks and utilizing a powerful mathematical programming environment led me to some fascinating insights into the problem and modifications of the problem.

To pursue this question, I studied the following notions. Although the order and significance of each changed during the course of the study, they remained my guide while I worked. I set out to develop:

- A probabilistic model of the area determined by 3 points on a simple random walk in 2-space.
- An analysis of this model that reports the expectation of this probability function, that is, what is the average area of such a triangle?
- A correction to the probabilistic model that better matches the constraints imposed by *Geometer's Sketchpad* on the random walks of the three points.
- How well does the mathematical theory assuming a true, simple random walk, coincide with the empirical results of a pseudo-random walk of the *Geometer's Sketchpad* environment?

- Finally, how does the answer to the average area of the triangle formed by three walking points constrained to a *Geometer's Sketchpad* window compare to the ratio of the area of the triangle determined by 3 points randomly chosen on the interior of some fixed-size rectangle?

I was able to develop a probabilistic model of the area determined by 3 points on a simple random walk in 2-space. I did this by designing a program. The program also reported the expectation that this model predicted. To this point with my investigation, I have not developed a closed form function for the expectation. However, the work concluding chapter 2, especially that of the iterative analysis, yields potential to create a closed form function, or at least a manner to determine the end behavior of this expectation function.

Although I did not complete a correction to this model of 3 walking points, I proposed how such a correction could be made with simple adjustment to the probability rules for the next move for each point on the walk. And finally, the remainder of this thesis will be dedicated to comparing results of the variety of investigations into the average area determined by 3 walking points in 2 dimensions. I will compare the idea of these points on an unbounded $\mathbb{Z}^2$ to the random selection of three points on a very large bounded region as well as the observations initially made in the GSP microworld.

The final results (Table 8) indicate that the average area (divided by the area of the region within which the initial points were chosen) of 3 walking points in 2 dimensions seems to be bounded above by approximately 0.10. However, my work earlier found that the average distance between 2 walking points grows on the order $\sqrt{n}$, corroborated by a theorem stating that the average distance between a walking point and its origin is on the order $\sqrt{n}$. If this 2-point distance grows, why would the area be bounded (approximately by 0.10)? This caused me to wonder what is the average area of 3 points on a random walk beginning at the origin? How does considering *n* steps vs. an “over the long run” (i.e. limit) way of thinking affect this problem?

Minor changes to my program code allowed for this investigation. I began by studying the initial area of three random points to their average area after a large number of steps. Table 9 shows this initial area and average area figured as a ratio of the initial window. The data reported in Table 9 confirms the notion that the average area of the walking triangle would equal the initial area. This is reasonable given the fact that the average location of a particle on a random walk is its initial location. This finding suggests that if all three points began at the origin, the average area of the triangle would be zero. And of course, this makes the GSP problem a bit more interesting, especially if seeking a sort of closed-form formula that will describe the average area of the points on the GSP window. So it still appears that the area of the triangle after *n* steps will approach the initial area, as *n* gets large.

Table 9

Average Area Ratio of Walking Triangle with Random Initial Vertices, Large n

Window: 800 × 600 Steps 10,000			Window: 800000 × 600000 Steps 100,000		
Trial Number	Initial Area	Expectation E(X_n)	Trial Number	Initial Area	Expectation E(X_n)
1	0.19600	0.19473	1	0.032004	0.031998
2	0.01597	0.03000	2	0.061159	0.061155
3	0.10561	0.15290	3	0.163339	0.163438
4	0.02378	0.07902	4	0.113654	0.113636
5	0.16640	0.18709	5	0.026232	0.026402
6	0.10271	0.12688	6	0.032866	0.032731
7	0.02039	0.07438	7	0.105850	0.105784
8	0.04946	0.02815	8	0.058225	0.058191
9	0.00267	0.01373	9	0.140737	0.140778
10	0.04151	0.03899	10	0.107044	0.107091

The results of Table 9 imply a contradiction to earlier findings. The conclusion suggested by Table 9 is that if three points are chosen at random, and then walk for *n* number of steps, the average of the triangular areas during these *n* steps will be equal to the area of the triangle determined by the initial points. This seems to be reasonable given the theorem that that average location of a point on a random walk is its starting location. However, the contradiction comes with the observation (and similar theorem for one point and the origin) that two points, beginning in the same location, have an average distance between them after *n* steps on the order of $\sqrt{n}$. The average location is affected by direction, that is positive and negative distances cancel each other out. I would imagine the same sort of "canceling" behavior works for 3 points beginning at some distance from each other, yields the behavior that is appearing—that the average areas during the walk remain equal to the starting area. But if the particles begin at the same point, as in the case of the distance between 2 points beginning at the origin, it appears as though the points want to separate from one another—in other words not maintain an average distance of zero. So, is there a steady state that is achieved once the points are some distance apart?[15]

I made one final inquiry into this odd behavior; I wanted to watch the average area of three particles that begin at the origin as they progress step-wise through *n* stages. I adapted the program in Figure 16 (see Figure 19) to initiate all particles ***p***, ***q***, ***r*** at the origin. I followed the same, very simple random walk behavior. I asked for the average area of the triangle determined by the 3 particles as they walked over the course of some *n* (variable "*s*" in the program) number of steps. I did not divide by the ratio of any sort of window, as was done in the investigation reported in Table 8. This would be an unnecessary division, and if I allowed *n* steps to become

---

[15] As a point of clarification, since the area function is dependant on distance, I predict the tendencies in the behaviors to be basically equivalent. That is since both are always positive measures, there is not the opportunity for "cancellation" as in the case of the location. I would expect a quadratic relationship of sorts between distance and area, yielding an initial conjecture that the average area of 3 points after *n* steps grows on an order of *n*—linearly.

sufficiently large, would suggest a boundary that the walking particles did not accept. I ran this computation several times (variable "*t*" in the program) and averaged the results. The results, reported in Table 10, once again suggest that the average area of the triangle, when all three particles begin at the origin, increases with the number of steps.

```
with(LinearAlgebra): with(RandomTools):
randwalkpts := proc(s,t)
  local area, expect, sumexpect, M, px, py, qx, qy, rx, ry, a, i, j, k, l;
  area := 0; expect := 0; sumexpect := 0;
  px := 0; py := 0; qx := 0; qy := 0; rx := 0; ry := 0;
  for a from 1 to t do
    area := 0; expect := 0;
    px := 0; py := 0; qx := 0; qy := 0; rx := 0; ry := 0;
    for i from 1 to s do
      j := Generate(integer(range=(1..4)));
      if j=1 then py := py + 1
        elif j=2 then px := px + 1
        elif j=3 then py := py - 1
        elif j=4 then px := px - 1
      end if:
      k := Generate(integer(range=(1..4)));
      if k=1 then qy := qy + 1
        elif k=2 then qx := qx + 1
        elif k=3 then qy := qy - 1
        elif k=4 then qx := qx - 1
      end if:
      l := Generate(integer(range=(1..4)));
      if l=1 then ry := ry + 1
        elif l=2 then rx := rx + 1
        elif l=3 then ry := ry - 1
        elif l=4 then rx := rx - 1
      end if:
      M := Matrix ([[rx-px,ry-py],[qx-px,qy-py]]);
      area:=area + abs(.5*Determinant(M));
    end do;
    expect := area/s;
    sumexpect := sumexpect + expect;
  end do;
  print(sumexpect/t);
end proc:
```

Figure 19. Program code for walking triangle initialized at origin.

Table 10 not only suggests that the average area of the triangle increases with the step size, but that it increases linearly, that is of order *n* (see Figure 20 for visual confirmation). A quick linear regression[16] yields the best-fit function to be $E(n) = 0.225009n - 59.0472$, with r = 0.99135. This result suggests two different phenomena are occurring when compared to earlier findings when selecting random initial points for the triangle vertices.

Table 10

Average Area Expectation Over 50 Trials of an n-Step 3-Point Random Walk

n number of steps	50-trial average of the expectation		n number of steps	50-trial average of the expectation
1	0.356		20,000	4534.640
2	0.648		30,000	6894.988
3	0.816		40,000	7748.730
4	0.885		50,000	13429.081
5	1.295		60,000	13355.082
10	2.464		70,000	13158.324
100	19.072		80,000	16249.082
1000	223.126		90,000	21299.567
10,000	2118.074		100,000	23909.145

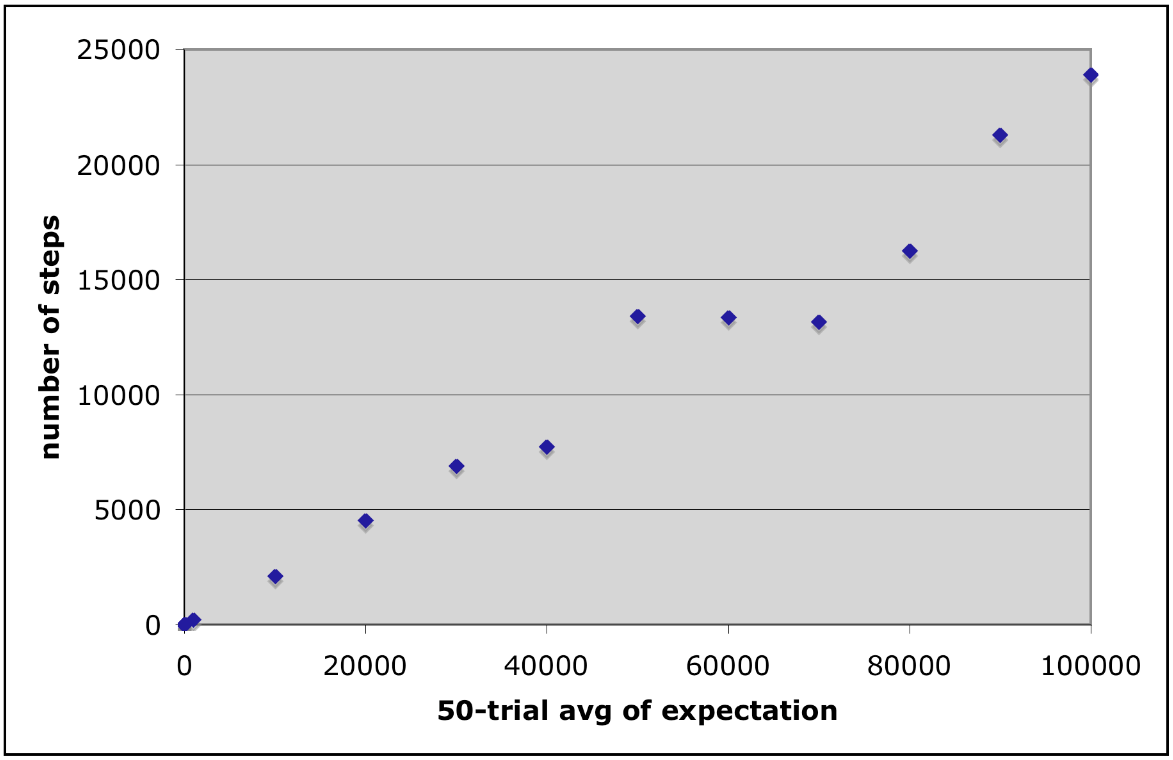


Figure 20. Average area expectation over 50 trials of an *n*-step 3-point random walk.

---

[16] Regression done on a Texas Instruments TI-84 graphing calculator.

This unresolved, surprisingly different behavior, along with other questions raised through the course of this study, open the doorway to a great deal of potential further research. Some initial thoughts consider that the scale of the changes are not observable in my data because of the relationships between step-size and boundary window—compounded by the division by the area of this window. It may be—and seemingly should be—that there truly is a difference between the initial area of three points randomly chosen on $\mathbb{Z}^2$ and the average area after step n. This problem could be adapted for undergraduate or further graduate work. I hope this initial foray has cleared the way for some initial navigation of the problem, in which worthwhile mathematical challenges have emerged, and more direct results have fallen out. The problem itself could be useful to motivate textbook problems in probability, combinatorics, algebra, and geometry. It could also create a rich, problem-based investigation for high school mathematics students. The problem has a simple level for entry and initial investigation does not draw on high-level mathematics.

A final note on what this exploration into 3 walking points seems to have demonstrated. Three walking points, beginning at the origin, create a triangle that grows in area in a linear fashion. However, three walking points in the GSP environment cannot meet this growth pattern. So it must be that the “force-field” built into the walls of the GSP environment pushed these points into the window in such a way that they seem to average an area of 10% the size of the window.[17]

---

[17] Because this thesis would seem incomplete if left on any note other than a new question, in what way does this area percentage depend on the size of the window? If the GSP force-field only takes place near the borders, than the average area should grow with the size of the window.

REFERENCES

Aarnes, J.F., and Knudtzon, S. H. (2003). Conjecture and discovery in geometry: A dialogue between exploring with dynamic geometric software (DGS) and mathematical reasoning. *Växjö*, May 9 2003.

Adams, M., and Guillemin, V. (1996). *Measure Theory and Probability*. Boston: Birkhäuser.

Chang, K. (2004). In math, computers don't lie. Or do they? *New York Times* [viewed online at http://www.nytimes.com/2004/04/06/science/06MATH.html on 5.2.06]

Cohen, J. W. (1992). *Analysis of random walks*. Amsterdam: IOS Press.

Hanna, G. (2001). Proof, explanation, and exploration: An overview. *Educational Studies in Mathematics 44*(1-2): 5-23.

Jackiw, N. (2004). *The geometer's sketchpad 4*. Berkeley: Key Curriculum Press.

Knight, W. (2005). Computer generates verifiable mathematics proof. *NewScientist.com 18*(27) [viewed online at http://www.newscientist.com/article.ns?id=dn7286 on 5.2.06]

Maplesoft. (2005). *Maple 10*. Waterloo, Canada: Author.

Peterson, I. (2004). Proof by computer. *Science News 166*(14) [viewed online at http://www.sciencenews.org/articles/20041002/mathtrek.asp on 5.2.06]

Spitzer, F. (1964). *Principles of random walk*. Princeton, NJ: D. Van Nostrand Company.