Estimating Area from a Lattice of Internal Points

John Coffey, Cheshire, UK.

2023

1 The Question

An arbitrary polygon, not necessarily convex, is placed at random on a square lattice. How many lattice nodes are inside the polygon or lie exactly on its boundary? If this number of internal+perimeter nodes is used to estimate the area of the polygon, what is the error likely to be? What if the polygon were replaced by an ellipse or other curved shape?

These questions arose when I was trying to determine the largest 'sofa' which can be moved round a 90° corner in a corridor. My partial solution to this sofa problem is also on *www.mathstudio.co.uk*. To answer the questions above I have looked into errors in a theoretical, probabilistic way for the simplest case of a rectangle placed over a rectangular grid of equally spaced nodes. This described in Section 1 supported by the Appendix. The bulk of the article is about statistics obtained by creating many random polygons and ellipses and placing them over grids. For this it is necessary to know the true area of an arbitrary polygon so §3 describes a historic formula. The computer programs and their results are described in §4.

My conclusion is that the number of nodes, N, inside the closed figure is in most cases a good estimate of the area A of the figure, with a percentage error which falls as the grid is made smaller with respect of the figure, and absolute error $|N - A|$ which is generally less than the area of 5 units cells of the lattice. Under-estimates and over-estimate are equally likely.

2 Counting internal grid points

We lay an arbitrary polygon over a grid of point nodes and count how many are inside the polygon or exactly on its perimeter. This number, N, multiplied by the area of each cell in the lattice, is an estimate of the area A of the polygon. We expect that the finer the grid mesh, the more accurately the number of nodes will represent the area, but seek to quantify this. A rectangular grid is the simplest to use in computing because the x and y directions are independent, which is not the case with a triangular grid. Having decided on rectangular, the grid might as well be square, which is what is illustrated in Figure 1 by the blue nodes. I will regard a cell as associated with the node at its lower left, with the smaller x and y ordinates. In Figure 1 there are 16 columns in x labelled 0 to 15, and 9 in y labelled 0 to 8. In general $N_x + 1$ by $N_y + 1$ nodes define an N_x by N_y grid. In this article the node spacing is 1.

Consider a rectangular polygon aligned with the grid. The smallest such rectangle which could have the full grid inside or on its perimeter is 15 cells by 8. This has the area of 120 cells and $16 \times 9 = 144$ nodes, and in this special case the error in estimating its area by counting nodes would be zero. However, the probability of any given rectangular polygon fitting exactly over the

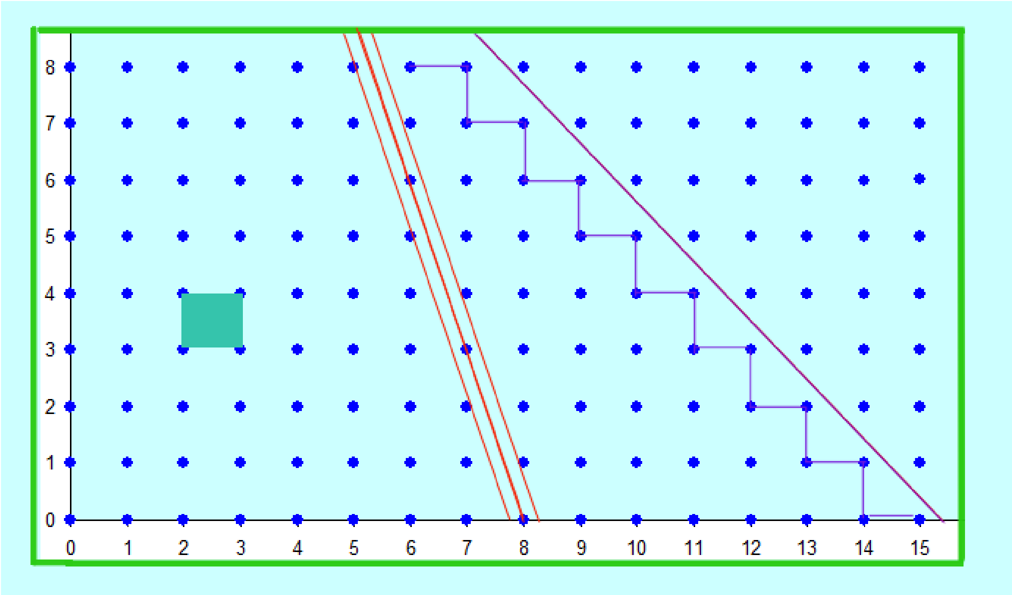

Figure 1: A 16 by 9 grid of equally spaced nodes, with an enclosing rectangles (green) and lines cuts (red, purple) to form trapezia. A single cell is shaded sea green. The stepped line is the boundary of cells internal to the purple cut.

given finite grid is virtually zero. At the other extreme, the largest rectangle which could fit over 144 grid points without touching any other outer nodes would be just less than $17 \times 10 = 170$ units, or $(N_x+2)(N_y+2)$ in general. The green rectangle in Figure 1 illustrates this case. The error would then be an underestimate of 50 units or -30%. However, the probability of a rectangle being of this size and in this precise position is again virtually zero. To know the likely error for a rectangle of random size and placement over the grid we need the probability distribution in area. This can be calculated analytically and some details are given in the Appendix. To complement the theory I have simulated the random placing of rectangles in a computer model and determined the distribution of areas for a number of values of N_x, N_y. The conclusion is the mean error in x is $\frac{1}{2} + \frac{1}{2} = 1$ cell length and the same in y. The error depends on whether we reckon the number of cells or the number of nodes as a measure of the area. Since the mean area of a random rectangle is $(N_x+1)(N_y+1)$ and this is also the number of nodes, the better choice is the number of *nodes*, not the number of cells. To be clear the best estimate of the mean likely area of the superposed rectangle is the (number of enclosed nodes, N)×(area of one cell).

I mention in passing that in the application of finding the largest sofa which can pass round a sharp corner, the sofa had mirror symmetry about the y axis so the calculation was done only on the right half with the grid constrained exactly onto the y axis. In this case the mean error in x can be only $\frac{1}{2}$ a cell length.

The staircase line in Figure 1 shows what happens when the grid tries to represent an oblique edge of a polygon. The three red lines show the span of positions between two oblique lines of nodes. Rather than try an analytical quantification of the errors, I have written a computer program which creates random polygons and calculates both their true areas and the number of nodes inside when placed over a square grid. This is described in §3.1. §3.2 describes an equivalent investigation of figures with curved edges through studying ellipses.

3 Area of a polygon

We need to know the true area of an arbitrary polygon. This can be done by dividing it into triangles by straight lines between selected vertices. Concavities can be treated, for instance, by dividing the polygon into two or more convex ones, or by subtracting the re-entrant part from the enveloping convex polygon. We are given the co-ordinates of the n vertices, $P_1 = (x_1, y_1)$, $P_2 = (x_2, y_2)$, ..., $P_n = (x_n, y_n)$.

To start we need the area A of an arbitrary single triangle as in the left panel in Figure 2. Its area is half the base times its perpendicular height, and can be found from the vertex co-ordinates as follows. Using vector calculus, two adjacent sides are $\mathbf{p}_1 - \mathbf{p}_3$ and $\mathbf{p}_2 - \mathbf{p}_3$, meeting at vertex P_3. Their lengths are $|\mathbf{p}_1 - \mathbf{p}_3|$ and $|\mathbf{p}_2 - \mathbf{p}_3|$, and the angle between them, θ, is given by the dot product

$$\cos\theta \;=\; \frac{(\mathbf{p}_1 - \mathbf{p}_3)\cdot(\mathbf{p}_2 - \mathbf{p}_3)}{|\mathbf{p}_1 - \mathbf{p}_3|\,|\mathbf{p}_2 - \mathbf{p}_3|}.$$

Take side P_3P_2 as the base of the triangle. The perpendicular distance of P_1 from P_3P_2 is $|\mathbf{p}_1 - \mathbf{p}_3|\sin\theta$, and since $\sin^2\theta = 1 - \cos^2\theta$, it is convenient to square these values and calculate $4A^2$. We have

$$4A^2 = |\mathbf{p}_2 - \mathbf{p}_3|^2\,|\mathbf{p}_1 - \mathbf{p}_3|^2\,(1 - \cos^2\theta)$$

$$= |\mathbf{p}_2 - \mathbf{p}_3|^2\,|\mathbf{p}_1 - \mathbf{p}_3|^2 \;-\; [(\mathbf{p}_1 - \mathbf{p}_3)\cdot(\mathbf{p}_2 - \mathbf{p}_3)]^2.$$

Using the definition of the dot product this simplifies to

$$2A \;=\; x_1(y_2 - y_3) \;+\; x_2(y_3 - y_1) \;+\; x_3(y_1 - y_2). \qquad (1a)$$

Note how the indices are in cyclic order. This can be written in matrix form as a sum of determinants:

$$2A \;=\; \begin{vmatrix} x_1 & y_1 \\ x_2 & y_2 \end{vmatrix} + \begin{vmatrix} x_2 & y_2 \\ x_3 & y_3 \end{vmatrix} + \begin{vmatrix} x_3 & y_3 \\ x_1 & y_1 \end{vmatrix} \;=\; \sum_{k=1}^{3} \begin{vmatrix} x_k & y_k \\ x_{k+1} & y_{k+1} \end{vmatrix} \qquad (1b)$$

where $x_4 \equiv x_1,\ y_4 \equiv y_1$.

In planar geometry area is a quantity preserved under rotations and translations. Translation can be checked in Eq 1a by replacing x_j by $x_j + a$ and y_j by $y_j + b$. Rotation is readily checked using Eq

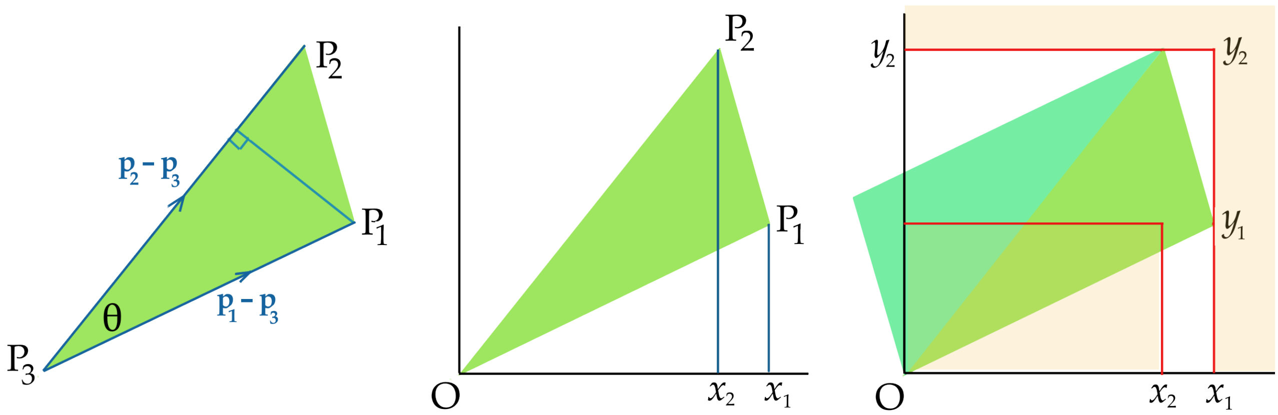

Figure 2: Left: triangle in general position. Centre: placed at origin with trapezium from x_2 to x_1 outlined. Right: two triangles forming a parallelogram of area equal to the Γ-shaped region.

1b by multiplying each matrix on the right by the rotation matrix $\begin{pmatrix} \cos\beta & \sin\beta \\ -\sin\beta & \cos\beta \end{pmatrix}$ whose determinant is +1. Since for two matrices A and B $\det(AB) = \det A \times \det B$, the value of $2A$ is unchanged by rotation through any angle β. Calculated by Eq 1a or b the area is a signed quantity, + or −, depending on the order in which the points are labelled. I find that if the triangle is in the first quadrant and the vertices are ordered anticlockwise as seen from the centroid, the area is positive, and negative if ordered clockwise. For this reason mirroring the figure changes the sign of its area. This is consistent with the fact that interchanging any two rows of a square matrix reverses the sign of its determinant.

Eq 1a can be understood geometrically by reference to the central panel of Figure 2. Here P_3 has been taken to be at the origin, O, making Eq 1 reduce to $2A = x_1y_2 - x_2y_1$. The triangle P_1P_2O is constructed from three figures all with bases on the x-axis: the right-angled triangle with vertex at (x_2, y_2), the right-angled triangle with vertex at (x_1, y_1), and the trapezium between x_2 and x_1. Their areas combine as

$$\tfrac{1}{2}x_2y_2 + (x_1 - x_2)\left(\frac{y_2 + y_1}{2}\right) - \tfrac{1}{2}x_1y_1 = \tfrac{1}{2}(x_1y_2 - x_2y_1).$$

Looking at the right panel of Figure 2, x_2y_1 is the area of the lower left pale orange rectangle, and x_1y_2 the area of the larger rectangle bounded above by the top right pale orange region. The L-shaped region between them has twice the area of the given triangle, equal to that of the green parallelogram in the figure. The area of the L-shaped region is $x_1(y_2 - y_1) + (x_1 - x_2)y_1 = x_1y_2 - x_2y_1$.

The formula for a single triangle can be built up into a similar formula for a general polygon by adding adjacent triangles which meet at O and subtracting ones nearer O which are not part of the polygon. The three panels of Figure 3 illustrate the construction and show how it applies to convex polygons and concave ones of increasing complexity. In the convex case the polygon's area is the sum of the areas of the triangles P_1P_2O, P_2P_3O, P_3P_4O and P_4P_5O, less triangle P_5P_1O. In the left panel of Figure 3 the signs of the contributing edges are marked by + or −. In the form of Eq 1b this is half of

$$\begin{vmatrix} x_1 & y_1 \\ x_2 & y_2 \end{vmatrix} + \begin{vmatrix} x_2 & y_2 \\ x_3 & y_3 \end{vmatrix} + \begin{vmatrix} x_3 & y_3 \\ x_4 & y_4 \end{vmatrix} + \begin{vmatrix} x_4 & y_4 \\ x_5 & y_5 \end{vmatrix} + \begin{vmatrix} x_5 & y_5 \\ x_1 & y_1 \end{vmatrix}. \qquad (2)$$

Observe that we have not had to include a minus sign with the last matrix; its determinant is negative because the vertex ordering in this triangle is clockwise, not anticlockwise. All other triangles are labelled anticlockwise and hence have positive determinants. This reversal in sign for the triangle P_5P_1O is a consequence of labelling the vertices of the polygon consistently in cyclic order. Any polygon in the right quadrant must have at least one edge contributing negative area.

Concave areas are included in the scheme of Eq 2 though it is necessary to check the signs of the associated abutting triangles. Consider the middle panel in Figure 3 and the signs of its six edges. It is clear, for instance, that the order $P_3 \to P_4$ is anticlockwise whilst the ordering $P_4 \to P_5$ is clockwise in their respective triangles with a vertex at O. The magnitude and sign of the contribution from any one triangle $P_jP_{j+1}O$ will depend on the position and orientation of the polygon because each triangle changes size and shape as the polygon is moved, being constrained to have a vertex at O. However, the algebraic sum of these will be the area of the polygon whatever its rotation or translation; this can be checked as above for the triangle. Since area is correctly calculated when the polygon is wholly in the first quadrant, it will be correctly calculated in all other positions and rotations.

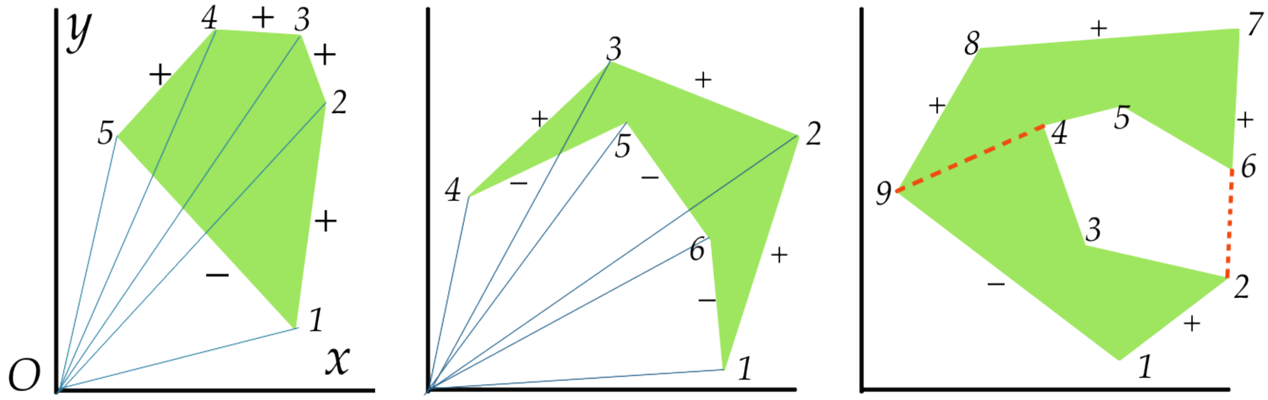

Figure 3: Polygons constructed from triangles with signed areas. Left: convex. Centre: dish-shaped concavity. Right: bottle-shaped concavity.

The right panel is Figure 3 shows a more complicated concave polygon. We could calculate its area in at least three ways:

1. splitting it into two simpler polygons by a cut between P_4 and P_9 such that the signs for each edge are clear and unambiguous,
2. closing the bottle-neck by a new edge from P_2 to P_6, and subtracting the area of the internal empty polygon from the enclosing one,
3. labelling all vertices consistently in an anticlockwise circuit and applying the scheme of Eq 2.

As a numerical example, the right hand polygon in Figure 3 can be created with vertices similar to $P_1 = (7, 1)$, $P_2 = (10, 3)$, ..., $P_9 = (0, 5)$. The full list is given in the three tables below, each corresponding to one of these ways of calculating area $2A$ listed above. The first table lists the co-ordinates for the two sub-polygons and the determinants formed from the row to the left and the one below it. Both the lower and upper sub-polygons are labelled anticlockwise and have areas +40 and +70 units, giving a total of 110.

vertex	x	y	det	vertex	x	y	det
1	7	1	11	4	4	7	−17
2	10	3	25	5	7	8	−46
3	5	4	19	6	11	6	55
4	4	7	20	7	11	11	88
9	0	5	−35	8	2	10	10
1	7	1		9	0	5	−20
				4	4	7	
			40				70

The second table lists the vertices of the enveloping polygon and the smaller cut-out. While the triangles through O associated with the large figure are each labelled anticlockwise, those of the cut-out are clockwise and so have negative signed area. The difference in area is again 110 units.
The third approach labels all vertices in sequence anticlockwise and gives area 110. Note that the signs associated with the edges in the cut-out are the same as those for method 2, and for the same reason – the associated triangles at the origin are taken clockwise.

vertex	x	y	det	vertex	x	y	det
1	7	1	11				
2	10	3	27	2	10	3	25
6	11	6	55	3	5	4	19
7	11	11	88	4	4	7	−17
8	2	10	10	5	7	8	−46
9	0	5	−35	6	11	6	−27
1	7	1		2	10	3	
			156				−46

vertex	x	y	det
1	7	1	11
2	10	3	25
3	5	4	19
4	4	7	−17
5	7	8	−46
6	11	6	55
7	11	11	88
8	2	10	10
9	0	5	−35
1	7	1	
			110

This illustrates that the formula Eq 2 for twice the area can be extended to any number of vertices in any closed polygon. Wikipedia reports that the formula was known to Gauss and others in the 18th century when it was called the 'shoelace' formula because of the repeated criss-crossing in multiplying the sequence of 2 by 2 determinants.

4 Computer simulations

4.1 Polygons

To gain some statistics on the number of enclosed grid points, I have written a computer program to model the situation. It creates a square grid at spacing $\delta x = \delta y = 1$ and places over it a polygon with n vertices at random. These are ordered anticlockwise and joined by straight lines. The area of the polygon is found from the co-ordinates of the vertices as in §3, while the program counts the number of nodes on or within the perimeter by reference to the equations of the line segments which piecewise make up its perimeter.

A particular challenge in writing this code has been to specify whether any given point lies inside, on the perimeter or outside the polygon. The vertices are given in anticlockwise cyclic order as in §3 and the cycle ends at the starting point. The direction of each edge is a signed vector quantity. As we mentally trace a path around the perimeter, the space to the left of every edge is inside and that to the right is outside. I dealt with this by first making the grid only large enough to fit inside the extreme x and and y coordinates of the polygon, for instance, allowing the rightmost vertex with x ordinate $14 \cdot 792$, say, to fall outside the grid by $0 \cdot 792$ units, the grid ending at integer 14 units. For each edge the columns of nodes lying between its two defining vertices were identified, and all nodes in these columns marked if they lay above or below the edge as appropriate. The flag marking their status was used to colour those nodes which were not counted, and so distinguish them from nodes inside the polygon which were counted. Figure 4 gives three examples of the screen display produced by the program. Nodes which are counted are in yellow and those not counted in cyan. The bottom left figure is an example of a spiky concave polygon.

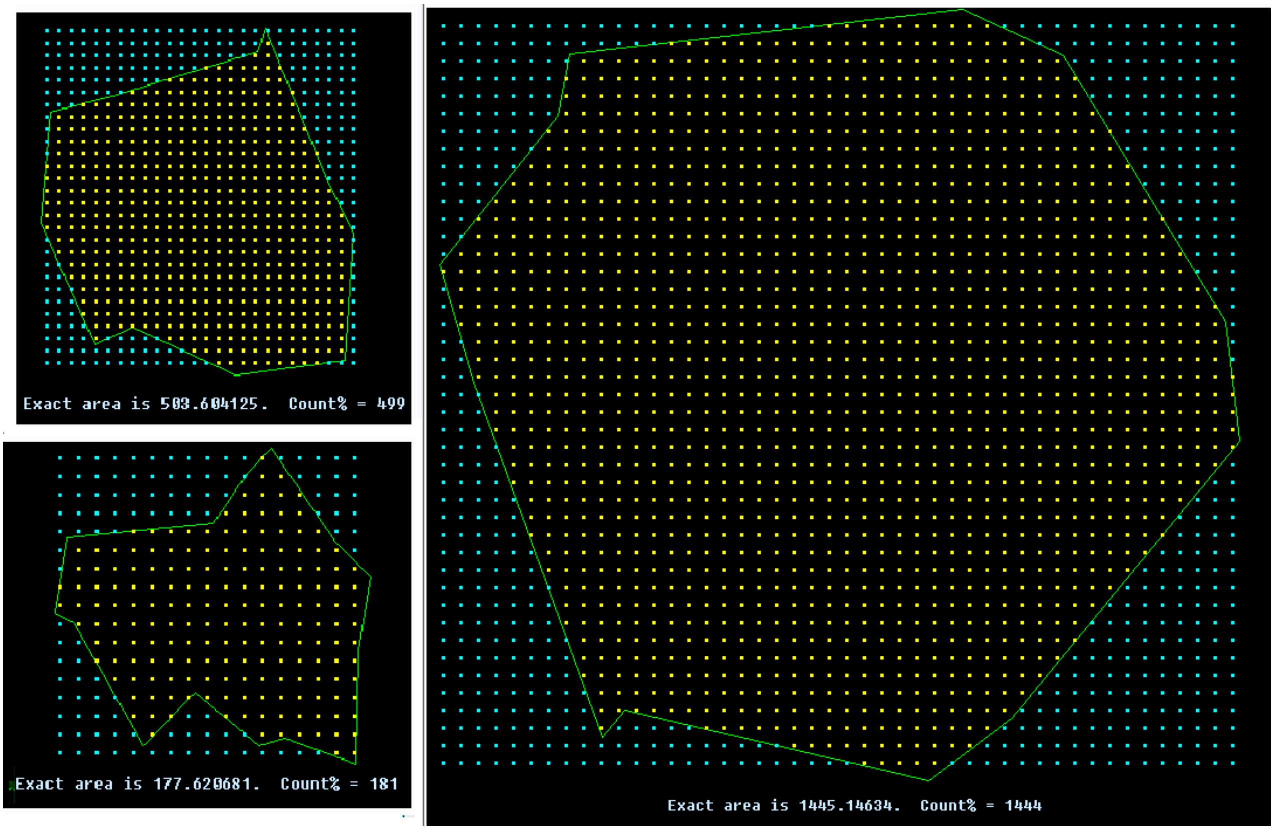

Figure 4: Screen shots of three polygons contributing to the statistics.

The data set for statistics consisted of 330 polygons with between 4 and 17 vertices, and areas between 63 and 6,360 square grid units. 101 had significant concavity, minor indentations not included. The agreement between true area A and the number of internal nodes N is remarkably close. The sum of all areas of all 330 polygons was $398515 \cdot 4$ square units and the sum of internal nodes 398399, only $116 \cdot 4$ an under-estimate smaller than $0 \cdot 03\%$.

Looking at the differences for individual polygons, $N - A$ spanned -18 to $+14$ with mean $-0 \cdot 36$ and standard deviation $3 \cdot 37$. In broad terms the node count is within about 4 square units of the actual area, with over-estimate being almost as likely as under-estimate. The mean absolute error $|N - A|$ was $2 \cdot 4$ nodes, standard deviation also $2 \cdot 4$. I found no significant trend to $N - A$ with size of polygon or with whether they were convex or concave. Consequently the percentage error, $100(N - A)/A$, falls with area. This can be seen in Figure 5. The mean absolute percentage error for polygons with areas between 2000 and 6360 square units was only $0 \cdot 10\%$. I did not make a thorough study for which polygons have the largest difference, but casual observations suggest that those with spikes have the largest.

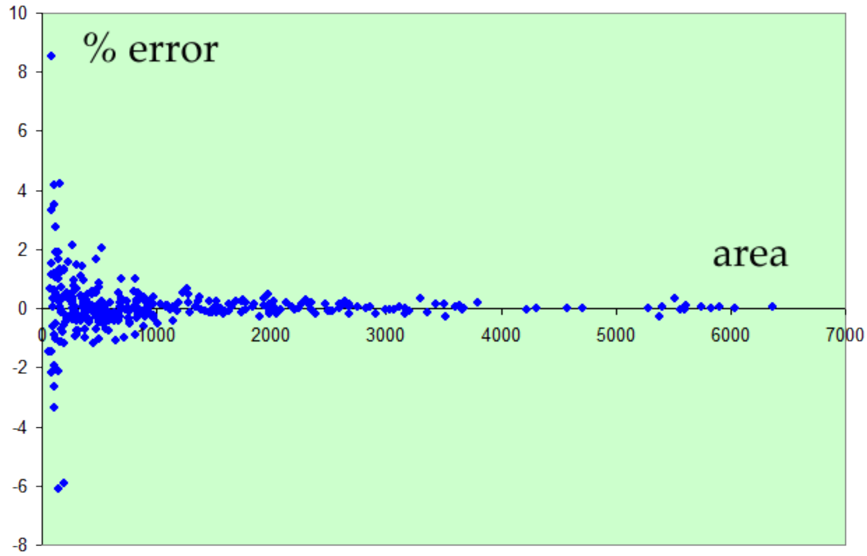

Figure 5: Percentage error $(N - A)/A$ in estimating the area A from number of internal nodes N for 330 random polygons.

4.2 Ellipses

The area of an ellipse is πab where a is the semi-major axis and b the semi-minor. An ellipse rotated anticlockwise by θ has equation

$$\frac{1}{a^2}(xc + ys)^2 + \frac{1}{b^2}(-xs + yc)^2 = 1, \qquad c = \cos\theta,\ s = \sin\theta.$$

Solving for y gives

$$y = \frac{1}{A}\left[-B \pm \sqrt{B^2 - AC}\right],$$

$$A = a^2c^2 + b^2s^2, \quad B = -(a^2 - b^2)csx, \quad C = (a^2s^2 + b^2c^2)x^2 - a^2b^2.$$

Differentiating and solving for where $dy/dx = 0$ gives the x values at the highest and lowest points on the ellipse at

$$x = \frac{\pm(a^2 - b^2)cs}{\sqrt{(a^2 - b^2)s^2 + b^2}}.$$

Taking the reciprocal of dy/dx and solving for zero gives the x positions of the extreme left and right points, where the tangents are vertical, at

$$x = \pm\sqrt{A}, \quad y = \frac{\pm(a^2 - b^2)cs}{\sqrt{A}}.$$

Figure 6 shows examples of ellipses typical of the data set created by a program similar to that for polygons. 124 were created with aspect ratios over the range $0 \cdot 01$ to $1 \cdot 0$ and rotations uniformly over 0 to 90°. Those with very small aspect ratio naturally had small areas – one such is in Figure 6 – and the span of areas was 7 to 3750 square units. The nodes outside the ellipse were determined by again creating a grid only large enough to cover the ellipse without going outside it, and then dividing the ellipse into four curved sections between positions were the tangents were vertical at one end and horizontal at the other. This was necessary to prevent square roots of negative

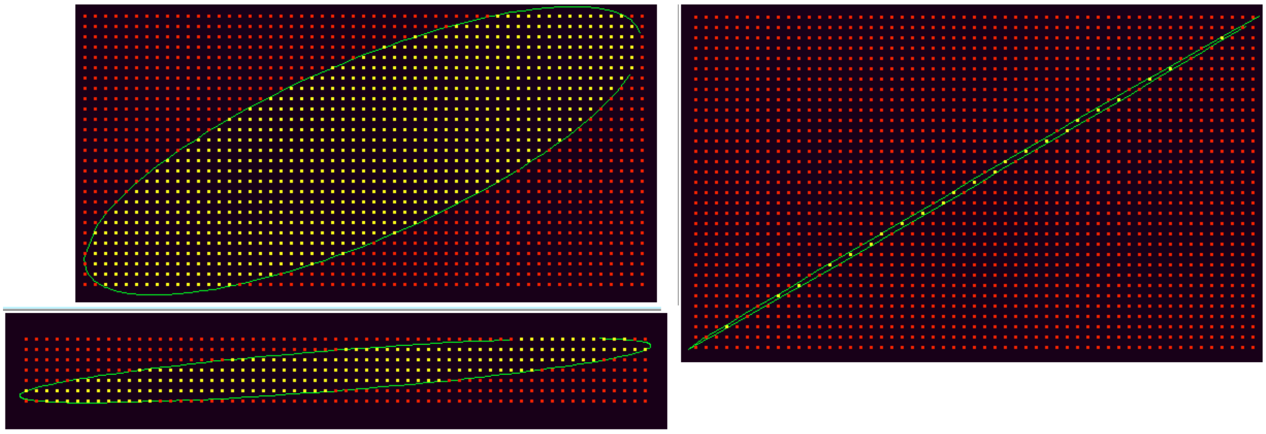

Figure 6: Examples of ellipses from the data set.

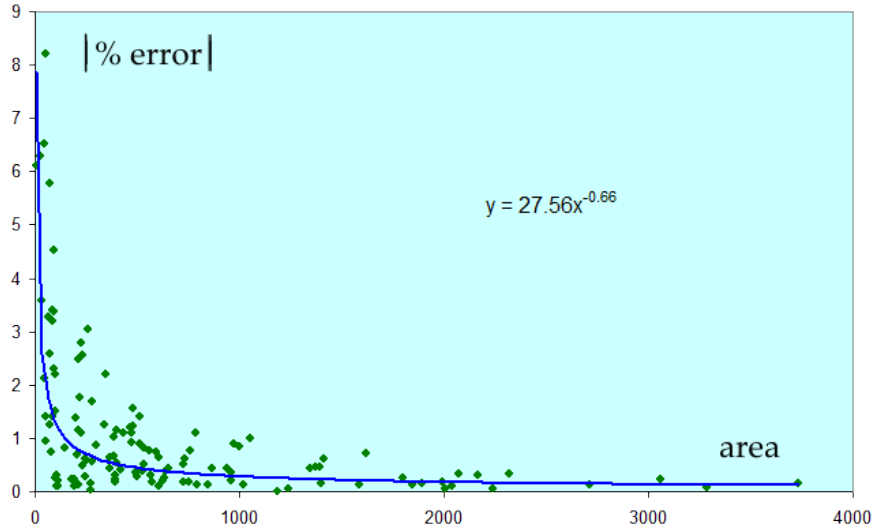

Figure 7: Absolute value of percentage error in area versus area for 124 ellipses.

numbers occurring in the equation of the ellipse. At each section nodes between the curve and the top or bottom row of the grid were flagged as 'outside' and coloured red.

The results were similar to those for the polygons. The differences $N - A$ ranged from -12 to 9, but averaged only $0 \cdot 09$ with standard deviation $4 \cdot 0$. The absolute values $|N - A|$ averaged $3 \cdot 1$, standard deviation $2 \cdot 5$. Because the differences $N - A$ are so similar for polygons and ellipses, I have combined the two data sets and Figure 8 is a histogram of the difference $N - A$ for all 454 figures.

In broad terms we can be fairly confident that the difference $N - A$ will be within ± 5 cell units of area irrespective of the shape or size of the figure unless it has sharp spikes, in which case the difference could be two or three times larger. As a result the percentage error in equating node count N with area A will decrease roughly as $1/A$.

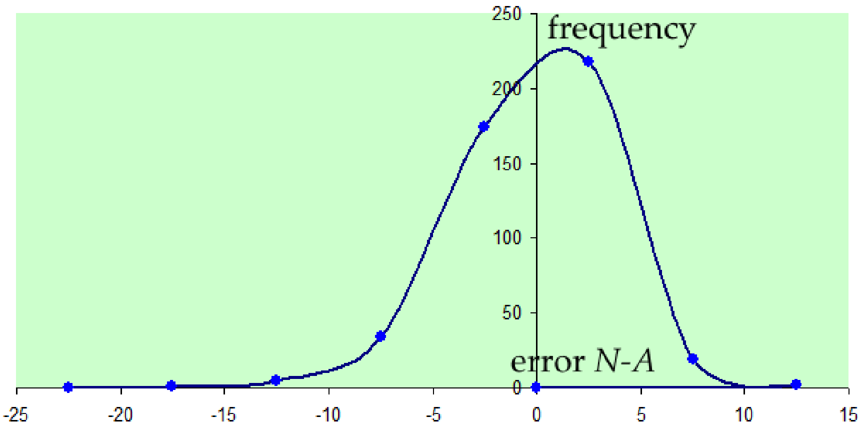

Figure 8: Combined statistics from polygons and ellipses. Histogram of differences $N - A$.

Appendix: pdf of sum and product of random variables

Refer back to Figure 1 showing the rectangular object R being laid over the N_x by N_y grid with the nodes at unit spacing in both x and y. Consider only the length L_x which R could have in the x direction without touching any of the nodes outside of the N_x by N_y block. The left edge could be distance x_1 to the left of node 0, $0 \le x_1 \le 1$. Similarly the right edge could be x_2 to the right of node $N_x - 1$, $0 \le x_2 \le 1$, giving a length $L_x = N_x + x_1 + x_2$. The probability distribution of L_x involves the sum of two independent random variables, each uniformly distributed. A simple illustration may be helpful. The 9 by 6 table of integers and frequency plot in Figure 4 makes clear why the pdf[1] of the sum $x_1 + x_2$ is the convolution of their individual pdfs. The coloured cells running diagonally show the 6 ways in which each of the values 5, 6, 7 and 8 occur. If the table of integers is square, the flat top to the frequency plot reduces to only a single point, making the distribution triangular.
The addition $x_1 + x_2$ is essentially the same. Each variable is uniformly distributed with the same 'top hat' function, $P(x_j) = 1$ for $0 \le x_1 \le 1$ and zero elsewhere. Their convolution is a symmetric triangle between 0 and 2, with peak at $x_1 + x_2 = 1$. The same applies in the y dimension. As numerical evidence Figure 5 plots the results for 10,000 uniformly distributed pseudo-random samples of x_1 and of x_2. The average error, as expected, is 1 grid unit, and the percentage error falls at $1/L_x$. The spread in values is broad; the standard deviation is $1/\sqrt{6} \approx 0 \cdot 41$.

The area of polygon R is $L_x L_y = (N_x + x_1 + x_2)(N_y + y_1 + y_2)$. Its pdf is more difficult to calculate analytically and I will not carry it out in the general case. Expanding the formula, the area

[1] Probability density function.

	0	1	2	3	4	5	6	7	8
0	0	1	2	3	4	5	6	7	8
1	1	2	3	4	5	6	7	8	9
2	2	3	4	5	6	7	8	9	10
3	3	4	5	6	7	8	9	10	11
4	4	5	6	7	8	9	10	11	12
5	5	6	7	8	9	10	11	12	13

adding integers

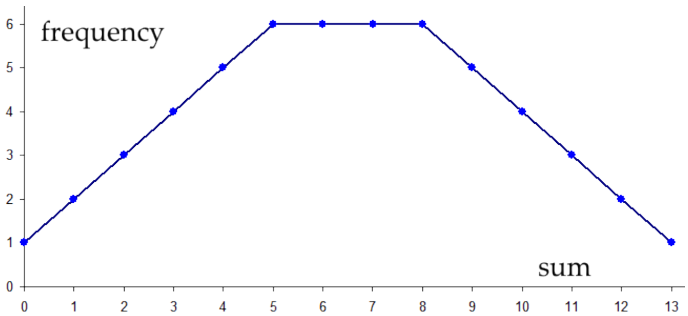

Figure 9: Sum of two uniform random variables and its pdf by convolution.

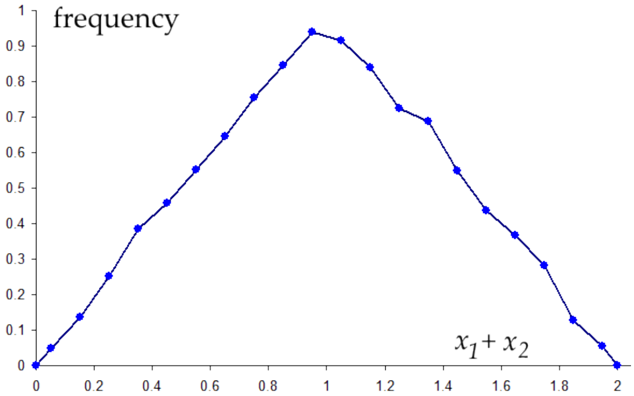

Figure 10: Pdf of sum $x_1 + x_2$ simulated by 10,000 samples of each.

$A = N_x N_y + N_x Z_y + N_y Z_x + Z_x Z_y$, $Z_x = x_1 + x_2$, $Z_y = y_1 + y_2$. This involves the sum, $N_x Z_y + N_y Z_x$, of two random variables with unequal triangular pdfs, and the smaller product $Z_x Z_y$, each factor sharing the same triangular pdf. The pfd of the product term $Z_x Z_y$ can also in principle be calculated analytically, but since its mean value will be about 2, it will make only marginal difference to the area of R. The largest contribution to the variation in area comes from the sides of the grid as $N_x Z_y + N_y Z_x$. Suppose $N_x Z_y$ has values between 0 and $2a$ where $a = N_x Z_y$, and pdf $p_A(u)$, and $N_y Z_x$ ones between 0 and $2b$, where $b = N_y Z_x < a$, with pdf $p_B(u)$. The pdf of their sum is the convolution

$$p_{a+b}(t) = \int_{-\infty}^{+\infty} p_A(u) p_B(t-u)\, du = \int_{-\infty}^{+\infty} p_A(t-u) p_B(u)\, du.$$

Figure 6 shows six relative positions of two such triangular pdfs, $p_A(u)$ in red, $p_B(u)$ in blue. The overlapping regions of the two functions are multiplied together and integrated. Because both parent pdfs are defined piecewise, so is the integral, and Figure 6 shows that the number of contributing sections varies from 1 to 3. The sum's pdf $p_{a+b}(t)$ will be $2a + 2b$ wide and symmetric about its mid-point. Since each of line in the triangles is linear, their product will be quadratic and become cubic upon integration. So $p_{a+b}(t)$ will piecewise sections of cubic curves.

The calculation is less involved if the two pdfs are identical so I have evaluated it for the case $a = 1$ so each pdf is a triangle over [0,2] with peak height of 1 at the central position. The pdf of the sum is

$$p_{A+B}(t) = \begin{cases} \dfrac{t^3}{6} & \text{if } 0 \le t \le 1 \\[2ex] \dfrac{2}{3} + 2t(t-1) - \dfrac{t^3}{2} & \text{if } 1 \le t \le 2 \\[2ex] \dfrac{t^3}{2} - 2t(2t-5) - \dfrac{22}{3} & \text{if } 2 \le t \le 3 \\[2ex] \dfrac{(4-t)^3}{6} & \text{if } 3 \le t \le 4 \end{cases}$$

This corresponds to the polygon being a square placed over a square grid. The standard deviation is $1/\sqrt{3} \approx 0 \cdot 58$.

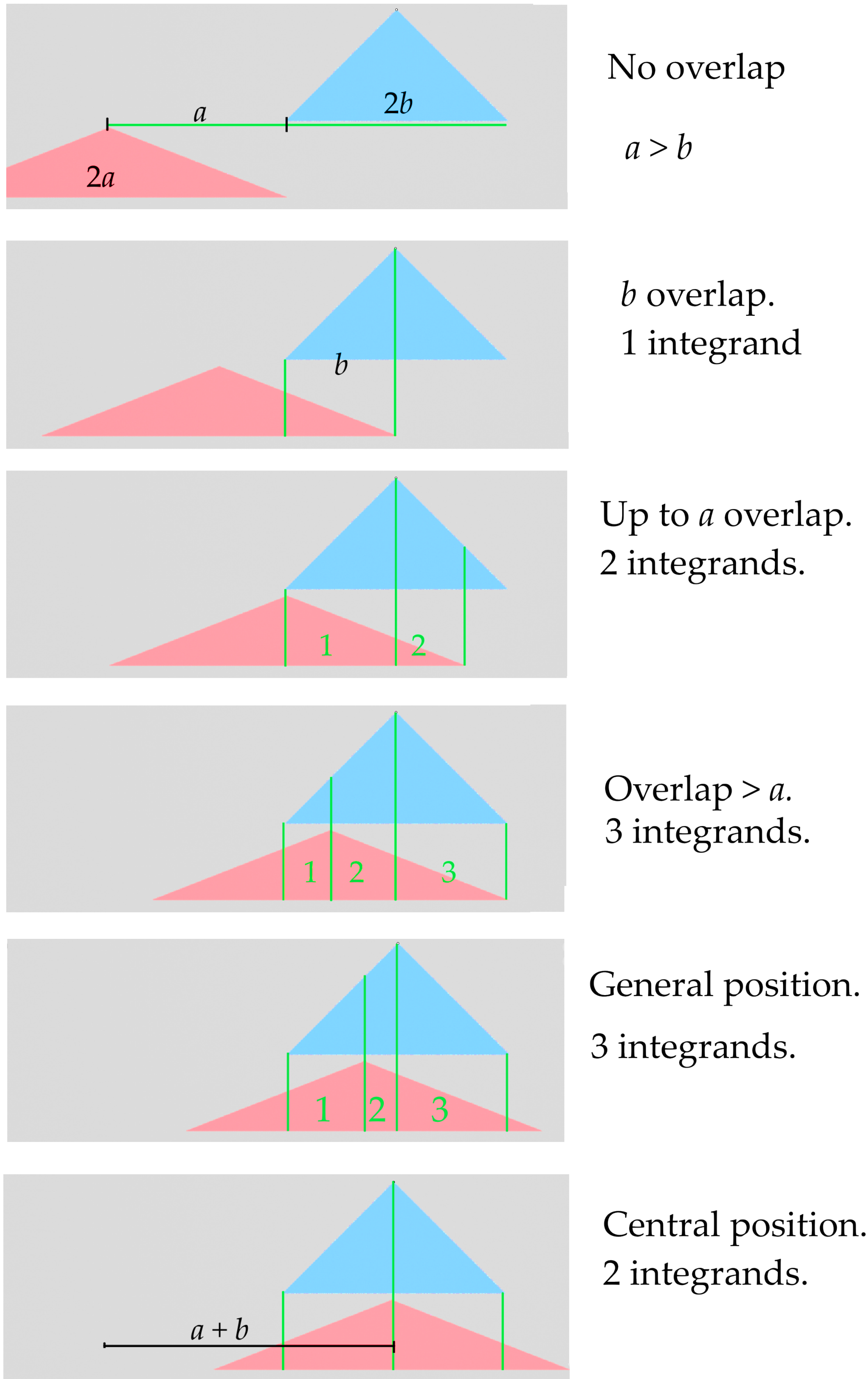

Figure 11: Six panels in the evaluation of the convolution of two unequal triangular pdfs.

To check this formula and bypass the calculational difficulties when the two pdf are different, I wrote a computer program to simulate the area of a rectangle subject to x_1, x_2, y_1, y_2 each being independent uniformly random variables over [0,1) added to a grid N_x by n_y cells in size. The calculation used 10,000 sets of four random displacements, x_1, x_2, y_1, y_2 and applied then to 8 sizes of rectangular grid from $N_x = 10$, $N_y = 10$ to 100 by 100, with several aspect ratios. The difference between areas of rectangle and grid spanned 0 to

$$(Nx + 2)(N_y + 2) - N_x N_y = 2(N_x + N_y) + 2.$$

This was normalised to 4 units to match the theoretical curve for a square. The width 4 represents $2(N_x + N_y)$ in any particular case. I found that for all squares studied, the pdf was also the same and fitted well with the theoretical curve. Figure 7 shows the theoretical pdf curve as a blue line while the red points are the average values obtained from the four squares studied. When the polygon was a rectangle with one pair of sides markedly shorter than the other pair, the distribution retained some of its triangular shape and even showed signs of a flat peak at its central position.

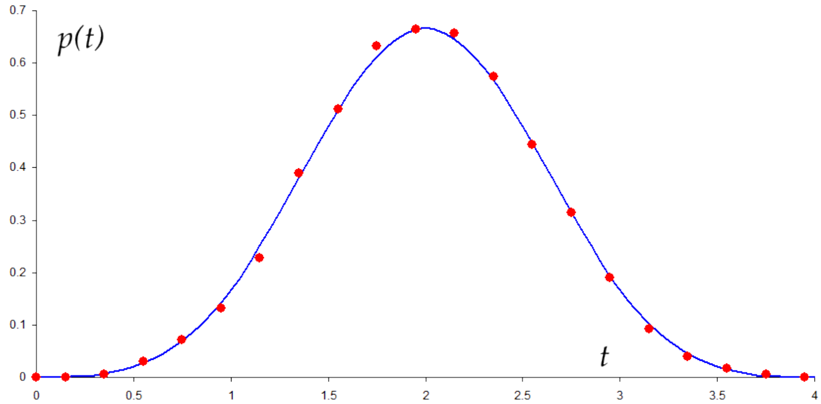

Figure 12: Pdf of sum of two equal triangular distributions each of width 2.

The conclusion from this analysis is that a grid of N_x by N_y cells, with $N_x + 1$ by $N_y + 1$ nodes, can be covered by the rectangle up to almost $N_x + 2$ by $N_y + 2$ in size without additional grid nodes being counted. Conversely, if $(N_x + 1)(N_y + 1) = N_x N_y + N_x + N_y + 1$ nodes are counted, the most likely area of the rectangle is $N_x N_y + (N_x + N_y)$ square units. Therefore the average fractional error in estimating the rectangle's area by counting grid nodes is $1/N_x + 1/N_y$.