THE AREA OF A RANDOM CONVEX POLYGON.

JOHAN PHILIP

ABSTRACT. We consider the area of the convex hull of \( n \) points with random positions in a square. We give the distribution function of the area for three and four random points. Our method is applicable for a larger number of points but the number of cases to consider becomes unmanageable. We also present some results on the number of vertices of the convex hull. Results from Monte Carlo tests with larger \( n \) are presented and compared with asymptotic estimates.

1. Introduction

Random objects are used when testing motion algorithms in computer vision. The typical object is a set of points in a square image. The performance of such an algorithm depends on several parameters and one of them is the size of the object. This raised our interest in the area of the polygon spanned by \( n \) randomly generated points in a square.

It is easily seen that the square can be dilated in one direction to a rectangle without changing the fraction that the polygon takes up of the area. In fact, any affine transformation will preserve the fraction of the area that is covered by the polygon. This follows from the fact that the area scaling is constant for an affine transformation. The scale equals the determinant of the homogeneous part of the transformation. This means that our results hold for any parallelogram.

Various aspects of our problem have been considered in the field of geometric probability, see e.g. [7]. J. J. Sylvester considered the problem of a random triangle \( T \) in an arbitrary convex set \( K \) and posed the following problem: Determine the shape of \( K \) for which \( \kappa = \text{the average of the ratio area}(T)/\text{area}(K) \) is maximal and minimal. A first attempt to solve the problem was published by M. W. Crofton in 1885. Wilhelm Blaschke [3] proved in 1917 that \( \frac{35}{48\pi} \leq \kappa \leq \frac{1}{12} \), where the minimum is attained only when \( K \) is an ellipse and the maximum only when \( K \) is a triangle. The upper and lower bounds of \( \kappa \) only differ by about 13%. It has been shown, [2], and we shall show that \( \kappa = \frac{11}{144} \) for the square.

A. Rényi and R. Sulanke, [5] and [6], consider the area ratio when the triangle \( T \) is replaced by the convex hull of \( n \) random points. They

Date: June 21, 2004.
obtain asymptotic estimates of the expected area ratio for large \( n \) and for various convex \( K \). R. E. Miles \cite{4} generalizes these asymptotic estimates for \( K \) a circle to higher dimensions.

H. A. Alikoski \cite{2} has given expressions for the expected area ratio for \( n = 4 \) and \( K \) a regular \( r \)-polygon. Our contribution is that we give the whole probability distribution for \( n = 3, n = 4 \) and \( K \) a parallelogram. We also give some asymptotic estimates in the spirit of Rényi and Sulanke. All theoretical results are compared with Monte Carlo tests.

2. Notation and formulation.

Consider points in the unit square \((0 \leq x \leq 1, 0 \leq y \leq 1)\). Use a constant probability density in the square for generating \( n \) random points in the square. Let \( A_n \) be the area of the convex hull of the \( n \) points. We shall determine the probability distribution of the random variable \( A_n \) for \( n = 3 \) and \( n = 4 \). Our method is applicable for bigger \( n \) but the number of cases that must be considered increases like \( n! \) and becomes unmanageable. Numerically, the distributions can be determined for any \( n \) by Monte Carlo tests. Since the unit square has area one, \( A_n \) can be interpreted as the fraction of the unit square that is covered.

The coordinates of the random points will be denoted \((x_k, y_k)\) for \( 1 \leq k \leq n \). Each \( x_k \) and \( y_k \) is evenly distributed in \((0, 1)\) and they are independent.

There is an area formula for a polygon with \( r \) vertices:

\[
(1) \quad A = \pm \frac{1}{2} (x_1 y_2 + x_2 y_3 + \ldots + x_r y_1 - x_2 y_1 - x_3 y_2 - \ldots - x_1 y_r).
\]

This formula is valid also for nonconvex polygons. The points must be ordered around the boundary. Counter-clockwise ordering gives the plus sign.

3. The area of a triangle.

For \( n = 3 \) the convex hull of the points is a triangle having maximal area \( = 1/2 \).

Even though any numbering of the points around the boundary will be consecutive for \( n = 3 \), formula (1) cannot be used directly because of the \( \pm \)-sign. Another difficulty with formula (1) is that the terms are dependent variables. We tackle the situation by ordering the point coordinates after size. Let \( p_x \) be a function that permutes \( x \)-indices, in this case the numbers 1, 2, and 3. Let \( p_y \) be a similar permutation of the \( y \)-indices so that

\[
(2) \quad x_{p_x(1)} \leq x_{p_x(2)} \leq x_{p_x(3)} \quad y_{p_y(1)} \leq y_{p_y(2)} \leq y_{p_y(3)}.
\]

There are \( 3! = 6 \) such \( p_x \) and equally many \( p_y \), so the combined number of permutations amount to 36. Without loss of generality, we
can assume that the triangle vertices are numbered so that \( x_1 \leq x_2 \leq x_3 \). This means that \( p_x \) can be dispensed with and we are left with 6 possible permutations which potentially correspond to geometrical configurations having different area formulae. These 6 permutations of the y-coordinates are equally probable to occur. However, a reversal of the y index sequence corresponds to turning the triangle upside down, which doesn’t affect its area. This leaves us with \( 3!/2 = 3 \) sets of permutations. Each set has a characteristic geometry and all sets have the same probability of occurring. Such sets will be called cases. As we shall see, two of these cases have the same kind of geometry, so we are left with two cases needing consideration and these are depicted in Figures 1 and 2.

The triangle spans a ‘big’ rectangle with sides of length \( x_3 - x_1 \) and \( y_{p_y(3)} - y_{p_y(1)} \). The coordinates \( x_2 \) and \( y_{p_y(2)} \) define a splitting of the ‘big’ rectangle into four sub-rectangles. Figure 1 depicts the y-permutation \( \{1,3,2\} \). Changing to \( \{3,1,2\} \) will flip the figure left-right, which preserves the geometry. This means that the case depicted in Figure 1 occurs twice as often as that in Figure 2, which corresponds to the permutation is \( \{1,2,3\} \).

It takes some time to realize that the area of the triangle in Figure 1 is half the sum of the areas of the three subrectangles marked with a +. In Figure 2, the area of the triangle is half the difference of the two subrectangles marked with + and -.

The distribution function for the \( k \)-th ordered variable among \( n \) variables with distribution function \( F(x) \) is
Figure 2. Triangle with point 2 in the interior of the 'big' subrectangle (Case 2). The difference of the areas of the small (dashed) subrectangles marked with a + and - equals twice the triangle area.

\[ F_{n,k}(x) = \frac{n!}{(k-1)!(n-k)!} \int_0^x F(t)^{k-1}(1 - F(t))^{n-k} dF(t). \]

Here, \( F(x) = x \). For \( n = 3 \) the distribution functions for the smallest, middle, and largest variables are

\[ F_{3,1}(x) = 1 - (1 - F(x))^3 = 1 - (1 - x)^3, \]
\[ F_{3,2}(x) = 6 \int_0^x t(1 - t) \, dt = 3x^2 - 2x^3, \]
\[ F_{3,3}(x) = (F(x))^3 = x^3. \]

The lengths of the intervals \((0, x_1), (x_1, x_2), (x_2, x_3), \) and \((x_3, 1)\) all have the same probability distribution, namely \( F_{3,1}(x) \). The lengths of the intervals \((0, x_2), (x_1, x_3), \) and \((x_2, 1)\) all have the same probability distribution, namely \( F_{3,2}(x) \). This implies that the sides of the 'big' rectangle in Figures 1 and 2 have the probability distributions \( F_{3,2}(x) \) and \( F_{3,2}(y) \). Let \( G_3(z) \) be the distribution function for the area of the 'big' rectangle. We have \( G_3(z) = \text{Prob}(area = xy \leq z) = 1 - \text{Prob}(xy > z) = 1 - \text{Prob}(y > z/x) \times \text{Prob}(x) \). We get \(^1\)

\(^1\)We are indebted to Maple for helping us calculate the integrals of this paper.
\[(7) \quad G_3(z) = 1 - \int_z^1 (1 - F_{3,2}(z/x)) \, dF_{3,2}(x) = 28z^3 - 27z^2 - 6z^2(3 + 2z) \log(z), \quad 0 \leq z \leq 1.\]

In order to determine the distribution of the triangle area, we shall study the fraction \( f \) that the triangle area takes up of its circumscribed 'big' rectangle. To avoid the factor \( 1/2 \) of the area formula for a triangle, or the factor 2, in numerous places below, we will work with twice the fraction \( f \) and twice the triangle area. Let \( H_{3,1}(u) = \text{Prob}(2f \leq u) \) in Case 1 depicted in Figure 1. Let \( H_{3,2}(u) \) be the corresponding probability for Case 2 of Figure 2.

When the 'big' rectangle is fixed, the conditional density distribution of \((x_2, y_{2p}(z))\) is constant in the 'big' rectangle. For the calculation of the fraction, we dilate the 'big' rectangle so that it fills the unit square. This doesn’t affect fraction. In Case 1, we have \( 2f = 1 - x_2y_3 \leq u \), so we shall integrate over \( x_2y_3 \geq 1 - u \), we get

\[(8) \quad H_{3,1}(u) = \int_{1-u}^1 dx_2 \int_{(1-u)/x_2}^1 dy_3 = u + (1 - u) \log(1 - u), \quad 0 \leq u \leq 1.\]

In Case 2, we have \( 2f = |x_2(1-y_2)-(1-x_2)y_2| = |x_2-y_2| \). Assuming that point 2 is below the diagonal, we have \( 2f = x_2 - y_2 \leq u \) and get

\[(9) \quad H_{3,2}(u) = 1 - 2 \int_u^1 dx_2 \int_0^{x_2-u} dy_2 = 2u - u^2, \quad 0 \leq u \leq 1.\]

Since Case 1 occurs twice as often as Case 2, we get the distribution function for \( 2f \) irrespective of case as

\[(10) \quad H_3(u) = \frac{2}{3}H_{3,1}(u) + \frac{1}{3}H_{3,2}(u) = \frac{1}{3}(4u - u^2 + 2(1 - u) \log(1 - u)), \quad 0 \leq u \leq 1.\]

Let \( F_3(x) \) be the distribution function for the triangle area \((= T)\). We have \( T \leq x \) if the area of the 'big' subrectangle \((= z)\) times the fraction \( f \) is less than or equal to \( x \), i.e. if \( 2fz \leq 2x \). Putting \( 2x = y \), we get
\[
F_3(y/2) = \int_0^1 G_3(y/u) \, dH_3(u) = \\
= [G_3(y/u)H_3(u)]_0^1 - \int_y^1 H_3(u) \frac{d}{du}G_3(y/u) \, du = \\
= G_3(y) - \int_y^1 H_3(u) \frac{d}{du}G_3(y/u) \, du, \quad 0 \leq y \leq 1.
\]

The partial integration in (11) is used to avoid integrating to the lower bound \( u = 0 \). The result cannot be written with elementary functions. Let us define the function \( \nu(x) \) which is not elementary but well defined.

\[
(12) \quad \nu(x) = \text{dilog}(x) + \log(x) \log |1 - x| = \int_x^1 \frac{\log(t)}{t-1} \, dt + \log(x) \log |1 - x|.
\]

Some properties of \( \nu(x) \) are given in Appendix A.

We will not carry out the integration (11) in detail, but will just give the result

\[
(13) \quad F_3(x) = \frac{4}{3}x(10 - 17x) - \frac{16}{3}x^3(17 - 3 \log(2x)) \log(2x) \\
+ \frac{2}{3}(1 - 16x - 68x^2)(1 - 2x) \log(1 - 2x) - 16x^3(3 + 2x)\nu(2x) \\
0 \leq x \leq 1/2.
\]

The density function \( dF_3/dx \) is shown in Figure 3.

The first moments and the standard deviation of the triangle area are

\[
(14) \quad \alpha_1 = \int_0^{1/2} x \, dF_3(x) = \frac{11}{144} \approx .076389,
\]

\[
(15) \quad \alpha_2 = \int_0^{1/2} x^2 \, dF_3(x) = \frac{1}{96},
\]

\[
(16) \quad \sigma = \sqrt{\alpha_2 - \alpha_1^2} = \frac{\sqrt{95}}{144} \approx .067686.
\]

4. **The area of the convex hull of four points.**

The convex hull of four points can be either a triangle or a quadrangle. The triangle case occurs if one point is generated inside the triangle spanned by the other three points. The probability for this to happen equals four times the expected size of the triangle = \( 4\alpha_1 = \frac{11}{36} \).

To find the distribution function for the area of \( A_4 \), we shall go about in the same way as above and number the points so that
Figure 3. Density function for the area of an arbitrary triangle.

\[ x_1 \leq x_2 \leq x_3 \leq x_4. \]

The corresponding \( y_k \) can be permuted in \( 4! = 24 \) ways. We form 12 sets each consisting of a permutation and its reversed permutation. All these sets are equally probable to occur. The permutations, but not the reversed ones, are listed in Table 1. The 13 cases in the table each correspond to a geometrical configuration. Cases 1 - 5 are triangles, cases 6 - 13 are quadrangles.

Like the case with three points, the generated convex hull spans a 'big' rectangle with sides \( x_4 - x_1 \) and \( y_{p_k(4)} - y_{p_k(1)} \). The distributions of these sidelengths are \( F_{4,3}(x) \) and \( F_{4,3}(y) \) where

\[
F_{4,3}(x) = 4x^3 - 3x^4.
\]

In analogy with (7), we get the distribution function for the 'big' rectangle

\[
G_4(z) = 1 - \int_z^1 (1 - F_{4,3}(z/x)) \, dF_{4,3}(x) =
81z^4 - 80z^3 - 12z^3(4 + 3z) \log(z), \quad 0 \leq z \leq 1.
\]

Now, we turn to determining the distribution function \( H_4 \) for the fraction of the 'big' rectangle that is covered by the triangle or quadrangle that spans it. We will have to determine a \( H_{4,k} \) for each of the 13 cases of Table 1. We must also determine the probabilities for these cases to occur, which are given in Table 1. Each of these cases requires the evaluation of an integral in four-space over a set bounded by linear and nonlinear inequalities. Having limited geometrical intuition in
Table 1. Probabilities of geometrical cases for each permutation. Cases 1 - 5 are triangles. Cases 6 - 13 are quadrangles. The Σ-row is the sum of the probabilities above. The $E\{u\}$-row shows the expectation of $u$ for each case.

four-space, each integral evaluation has been a challenge. We are not going to describe all these calculations in detail here but shall carry out the calculations for only one case in the text. Two more cases are done in appendices B and C. The author can provide the interested reader with Maple files describing the remaining cases.

Consider the possible geometric configurations that can occur for the $y$-permutation 1234. These are shown in Figure 4.

In these figures, we don’t show the whole unit square but only the ‘big’ rectangle. This rectangle has been enlarged by an affine transformation so that it fills a unit square. This doesn’t affect the area fraction that we are studying.

The triangle Case 4 depicted in Figure 4 is in fact the subcase of Case 4 having point 2 as a vertex below the diagonal and point 3 as an interior point. Case 4 also includes the situations with point 2 above the diagonal and point 3 a vertex below and above the diagonal. These four subcases all have the same probability and the same type of geometry. We shall do the case in Figure 4.
For fixed values of $x_2$ and $y_2$, the conditional probability density for point 3 is constant and equals $1/(1 - x_2)(1 - y_2)$ in the rectangle $x_2 \leq x_3 \leq 1$, $y_2 \leq y_3 \leq 1$. We have the conditional probability that point 3 is interior in the triangle spanned by 1, 2, and 4:

$$p(x_2, y_2) = \int_{x_2}^{1} \frac{dx_3}{1 - x_2} \int_{1-(1-y_2)(1-x_3)/(1-x_2)}^{x_3} \frac{dy_3}{1 - y_2}.$$

This probability shall be integrated over all positions of point 2. The density for point 2 is the same as for the smaller of two ordered points in the unit interval, i.e. $2(1 - x_2)dx_2 \times 2(1 - y_2)dy_2$. Notice that the factors $1 - x_2$ and $1 - y_2$ of the density cancel the same factors in (19). We get
\[
\text{Prob(triangle under diagonal with vertex in point 2) =}
\]
\[
= \int_0^1 2(1 - x_2) \, dx_2 \int_0^{x_2} 2(1 - y_2)p(x_2, y_2) \, dy_2 =
\]
\[
(20)
\]
\[
= 4 \int_0^1 dx_2 \int_0^{x_2} dy_2 \int_0^{x_3} dx_3 \int_{1-(1-y_2)(1-x_3)/(1-x_2)}^{x_3} dy_3 =
\]
\[
= \frac{1}{12}.
\]

The total probability for Case 4 is \(\frac{4}{12} = \frac{1}{3}\), which is inserted in Table 1.

Case 13 is a quadrangle with points 2 and 3 on the same side of the diagonal, see Figure 4. In analogy with (20), we get

\[
\begin{align*}
\text{Prob(quadrangle under diagonal)} &= \\
(21) &= 4 \int_0^1 dx_2 \int_0^{x_2} dy_2 \int_0^{x_3} dx_3 \int_{1-(1-y_2)(1-x_3)/(1-x_2)}^{x_3} dy_3 = \\
&= \frac{1}{6}.
\end{align*}
\]

Case 13 also includes a quadrangle above the diagonal. This gives the probability \(\frac{2}{6} = \frac{1}{3}\) given in Table 1.

Case 12 is a quadrangle with one vertex above and one below the diagonal, see Figure 4. It has the remaining probability = \(\frac{1}{3}\).

Returning to the distribution function for Case 4, we note that the difference between the areas of the rectangles marked + and - in Figure 2 equals twice the triangle area. The condition: area fraction of 'big' rectangle = \(\frac{1}{2} \cdot |x_2(1-y_2) - y_2(1-x_2)| \leq u/2\) reduces to \(x_2 - y_2 \leq u\) when point 2 is below the diagonal. Since this inequality is linear, we happen to have one of the simplest cases. The integral below is for the complementary event.

\[
(22)
\]
\[
H_{4,4}(u) = \text{Prob(triangle area fraction \leq u/2)} =
\]
\[
= 1 - 48 \int_0^1 dx_2 \int_0^{x_2-u} dy_2 \int_0^{x_3} dx_3 \int_{1-(1-y_2)(1-x_3)/(1-x_2)}^{x_3} dy_3 =
\]
\[
= 3u^4 - 8u^3 + 6u^2, \quad 0 \leq u \leq 1.
\]

The factor 48 is \(4 \cdot (1/12)^{-1}\), where \(1/12\) is the probability calculated in (20).
Figure 5. Density function $dH_4/du$ for twice the fraction that the four point object takes up of the 'big' rectangle. The singularity at $u = 1$ stems from Cases 3, 6, and 7.

$H_{4,8}(u)$ and $H_{4,13}(u)$ are calculated in appendices B and C. We picture all the cases and list their $H_{4,k}(u)$ in Appendix D. The total fraction probability distribution function denoted $H_4$ is obtained by weighting the $H_{4,k}(u)$ together with $1/12$ times the weights in the $\Sigma$-row of Table 1. It is

\[
H_4(u) = \begin{cases} 
\frac{9}{72}(24 + 84u - 16u^2 + u^3) + \\
+ \frac{1}{3}(1 - u^2) \log(1 - u) - \frac{2u}{18}, & 0 \leq u \leq 1, \\
\frac{1}{72}(-56 + 212u - 54u^2 - 8u^3 - u^4) + \\
+ \frac{1}{6}(9u - 7)(u - 1) \log(u - 1) - \frac{2}{3}u^2 \nu(u) - \frac{2u^2}{18}, & 1 \leq u \leq 2.
\end{cases}
\]

The density $dH_4/du$ is shown in Figure 5.

In analogy with (11), we can now form

\[
F_4(y/2) = \int_0^y G_4(y/u) \, dH_4(u) = \\
= [G_4(y/u)H_4(u)]_0^y - \int_y^2 H_4(u) \frac{d}{du} G_4(y/u) \, du = \\
= G_4(y/2) - \int_y^2 H_4(u) \frac{d}{du} G_4(y/u) \, du, \quad 0 \leq y \leq 2.
\]

Substituting back $y = 2x$, we have the distribution function for the area spanned by four points.
The density function \( dF_4/dx \) for area spanned by four points is shown in Figure 6. The first moments and the standard deviation of the four point area are

\[
\begin{align*}
\alpha_1 &= \int_0^1 x \, dF_4(x) = \frac{11}{72} \approx .15278, \\
\alpha_2 &= \int_0^1 x^2 \, dF_4(x) = \frac{859}{27000}, \\
\alpha_3 &= \int_0^1 x^3 \, dF_4(x) = \frac{73}{9000}, \\
\sigma &= \sqrt{\alpha_2 - \alpha_1^2} = \frac{17\sqrt{95}}{1800} \approx .09205.
\end{align*}
\]

The density function \( dF_4/dx \) is given by

\[
F_4(x) = \begin{cases}
\frac{8}{3}(6 + 1302x - 856x^2 + 1574x^3) - \\
\frac{2}{3}x^2(x + 1)(x + 3)\pi^2 - \\
\frac{8}{3}x^4(235 - 43 \log(2x)) \log(2x) + \\
\frac{1}{3}(1 + 2x - 140x^2 - 792x^3)(1 - 2x) \log(1 - 2x) - \\
-64x^3(3x + 4)\nu(2x), & 0 \leq x \leq 1/2 \\
-\frac{1}{3}(7 - 149x - 1365x^2 - 76x^3 + 1574x^4) + \\
-\frac{1}{6}(7 - 114x - 660x^2 - 296x^3)(2x - 1) \log(2x - 1) - \\
-32x^2(x + 1)(x + 3)(\nu(2x) + \frac{\pi^2}{12}), & 1/2 \leq x \leq 1.
\end{cases}
\]
5. The area of the convex hull of \( n \geq 5 \) points.

We have not found it meaningful to carry out detailed calculations for more than 4 points. Already with 5 points, we would have 60 permutations to consider. We shall present results from Monte Carlo tests for \( n \geq 5 \) below, but will start with some theoretical considerations.

The distribution function \( G_n(z) \) for the size of the big rectangle can be found explicitely. Using \( F_{n,n-1}(x) \) from (3), we get in analogy with (7) that

\[
(27) \quad G_n(z) = 1 - \int_z^1 (1 - F_{n,n-1}(z/x)) \, dF_{n,n-1}(x) = \\
= (n-1)^2(2n+1)z^n + n^2(3-2n)z^{n-1} - \\
- n(n-1)(nz^{n-1} + (n-1)z^n) \log(z), \quad 0 \leq z \leq 1.
\]

The expression for the density \( dG_n(z)/dz \) is more appealing

\[
(28) \quad \frac{dG_n(z)}{dz} = -n^2(n-1)^2z^{n-2}[2(1-z) + (1+z) \log(z)].
\]

It follows from the discussion preceding (7) that the expectation of \( z \) is \( \alpha_{1G} = \left(\frac{n-1}{n+1}\right)^2 \). The standard deviation can be evaluated to \( \sigma_G = \frac{2(n-1)}{(n+2)\sqrt{n+1}} \).

The density \( dG_n/dz \) for \( n = 5, 10, 20, 100 \) is shown in Figure 7. The fraction distribution \( H_n \) is hard to derive. Like for \( G_n \), the expectation of the fraction tends to 1 with increasing \( n \).

The densities \( dH_n/dt \), where \( t = u/2 \), for \( n = 5, 10, 20, 100 \) obtained from Monte Carlo tests are shown in Figure 8. The pronounced peak of the density at \( t = 1/2 \) present for \( n = 4 \) exists also for \( n = 5 \) but for higher \( n \), the maximum is more flat.

The results of the Monte Carlo tests with the convex polygon area density \( dF_n/dx \) for \( n = 5, 10, 20, 100 \) are shown in Figure 9.

6. The number of vertices of the convex hull.

Our detailed study of the convex hull of 3 and 4 points permits us to calculate some expectations for the number of vertices of a random convex polygon in a square.

For \( k \leq n \), define

\[
q_n(k) = \text{Prob}( \text{n points generate a convex polygon with k vertices}).
\]

Of course, \( q_3(3) = 1 \). We noted in the beginning of section 4 that \( q_4(3) = \frac{1}{36} \) implying \( q_4(4) = \frac{2}{36} \). These two probabilities can also be deducted from the \( \Sigma \)-row of Table 1 as follows. The sum of the elements of the first five cases, which are triangle cases, divided by 12 equals \( q_4(3) \).
Figure 7. Density functions $dG_n(x)/dx$.

Figure 8. Density functions $dH_n(x)/dx$ from 10000 Monte Carlo tests.
Figure 9. Density functions $dF_n(x)/dx$ for area spanned by $n$ points from 10000 Monte Carlo tests for each $n$.

The knowledge of $F_3(x)$ permits us to calculate $q_n(3)$ for all $n$. In fact, we get a triangle if points 4 through $n$ are generated inside the triangle generated by the first three points. The probability $\pi(m)$ that $m$ points fall inside an area of size $x$ distributed according to $F_3(x)$ is

$$\pi(m) = \int_0^{1/2} x^m dF_3(x).$$

Taking into account the number of ways the three points forming the triangle can be chosen, we get

$$q_4(3) = \binom{n}{3} \pi(n - 3) = \binom{n}{3} \int_0^{1/2} x^{n-3} dF_3(x).$$

Some values are

$$q_4(3) = \frac{11}{36} \approx .3056, \quad q_5(3) = \frac{5}{48} \approx .1042, \quad q_6(3) = \frac{137}{3600} \approx .0381,$$

$$q_7(3) = \frac{7}{480} \approx .0146, \quad q_8(3) = \frac{363}{62720} \approx .0058.$$

For $k = 4$, we need the conditional probability that points 5 through $n$ are generated inside the area generated by the first four points, provided these four points span a quadrangle. Cases 6 - 13 are quadrangles,
so summing the $H_{4,k}$ for $6 \leq k \leq 13$ multiplied by the weights in Table 1 will give us the wanted conditional distribution function $H_{4q}(u)$. Combining $H_{4q}$ with $G_4$ will give the conditional distribution function $F_{4q}(x)$. Including the probability $\frac{25}{36}$ of getting a quadrangle, we get

\begin{equation}
q_n(4) = \frac{25}{36} \binom{n}{4} \int_0^1 x^{n-4} dF_{4q}(x).
\end{equation}

Some values are

\begin{align*}
q_5(4) &= \frac{5}{9}, \\
q_6(4) &= \frac{1307}{3600} \approx 0.3631, \\
q_7(4) &= \frac{203}{900} \approx 0.2256.
\end{align*}

From the above, we can deduce $q_5(5) = 1 - \frac{5}{9} - \frac{5}{48} = \frac{49}{144} \approx 0.3402$.

These numbers can be compared with those of Table 2, which contains results from 10000 Monte Carlo tests for each $n$.

7. ASYMPTOTIC ESTIMATES

Rényi and Sulanke [6] consider the convex hull $H_n$ of $n$ random points generated inside a convex set $K$. For large $n$, we have the following formulas for the expected value $E_{H_n}(K)$ of $\text{area}(H_n)/\text{area}(K)$. The formulas for $K$ a square and a circle are from [6], while the formula for $K$ a triangle is deduced in Appendix E.

\begin{align*}
E_{H_n}(\text{triangle}) &= 1 - 2 \frac{\log(n) + \gamma}{n} + O \left( \frac{1}{n^2} \right), \\
E_{H_n}(\text{square}) &= 1 - \frac{8 \log(n)}{3} + O \left( \frac{1}{n} \right), \\
E_{H_n}(\text{circle}) &= 1 - \frac{(24 \pi)^{2/3} \Gamma(8/3)}{10n^{2/3}} + O \left( \frac{1}{n} \right) \approx 1 - \frac{3.3832}{n^{2/3}},
\end{align*}

where $\gamma \approx 0.5772$ is Euler’s constant.

These theoretical expectations are plotted in Figure 10 together with the results from Monte Carlo tests.

Closely related to the above are the formulas for the expected number of vertices $E_n$ of the convex hull of $n$ random points inside various convex polygons $K$. For $K$ a triangle, a square, and a circle and for large $n$, we have
<table>
<thead>
<tr>
<th>$k$</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>12</th>
<th>15</th>
<th>20</th>
<th>40</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>3068</td>
<td>1060</td>
<td>374</td>
<td>147</td>
<td>64</td>
<td>30</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>6932</td>
<td>5525</td>
<td>3627</td>
<td>2253</td>
<td>1379</td>
<td>876</td>
<td>562</td>
<td>236</td>
<td>70</td>
<td>27</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>3415</td>
<td>4731</td>
<td>4739</td>
<td>4185</td>
<td>3424</td>
<td>2732</td>
<td>1624</td>
<td>787</td>
<td>311</td>
<td>19</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1268</td>
<td>2528</td>
<td>3407</td>
<td>3805</td>
<td>3944</td>
<td>3534</td>
<td>2448</td>
<td>1340</td>
<td>169</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>333</td>
<td>882</td>
<td>1599</td>
<td>2176</td>
<td>3066</td>
<td>3447</td>
<td>2677</td>
<td>715</td>
<td>52</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>83</td>
<td>254</td>
<td>538</td>
<td>1260</td>
<td>2256</td>
<td>2902</td>
<td>1700</td>
<td>237</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>12</td>
<td>42</td>
<td>246</td>
<td>777</td>
<td>1836</td>
<td>2379</td>
<td>645</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>2</td>
<td>27</td>
<td>190</td>
<td>700</td>
<td>2312</td>
<td>1286</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td>3</td>
<td>23</td>
<td>175</td>
<td>1621</td>
<td>1920</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
<td>2</td>
<td>30</td>
<td>714</td>
<td>2009</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td></td>
<td>2</td>
<td>287</td>
<td>1785</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td></td>
<td></td>
<td>58</td>
<td>1084</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td></td>
<td></td>
<td>20</td>
<td>602</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td>267</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>81</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td></td>
<td>19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6</td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$E_q$</th>
<th>3.69</th>
<th>4.24</th>
<th>4.68</th>
<th>5.06</th>
<th>5.39</th>
<th>5.69</th>
<th>5.95</th>
<th>6.43</th>
<th>6.99</th>
<th>7.72</th>
<th>9.56</th>
<th>11.98</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_q$</td>
<td>.46</td>
<td>.63</td>
<td>.74</td>
<td>.81</td>
<td>.89</td>
<td>.94</td>
<td>.99</td>
<td>1.08</td>
<td>1.18</td>
<td>1.30</td>
<td>1.60</td>
<td>1.95</td>
</tr>
</tbody>
</table>

**Table 2.** Number of vertices $k$ for various number of generated points $n$.

The presented data are the actual outcome of 10000 Monte Carlo tests for each $n$ (= for each column). Notice that the largest number of vertices observed for $n = 100$ is 20. The last two rows give the average and standard deviation for each $n$. 
Figure 10. The fraction of the area that the convex hull of \( n \) random points takes up inside a triangle (top), square (middle), or circle (bottom curves). The thicker lines are theoretical from [6] and Appendix E, the thinner are from Monte Carlo tests.

\[
\begin{align*}
E_n(\text{triangle}) &= 2(\log(n/2) + \gamma) + o(1), \\
E_n(\text{square}) &= \frac{8}{3}(\log(n/2) + \gamma) + o(1), \\
E_n(\text{circle}) &= \frac{4}{3} \left( \frac{3\pi}{2} \right)^{2/3} \Gamma \left( \frac{5}{3} \right) n^{1/3} - \frac{4}{15} \left( \frac{3\pi}{2} \right)^{4/3} \Gamma \left( \frac{7}{3} \right) n^{-1/3} + O(n^{-2/3}) \\
&\approx 3.3832 n^{1/3} - 2.5084 n^{-1/3}.
\end{align*}
\]

The formulas for the triangle and the square and the first term for the circle are given in [5]. The second term for the circle is deduced in Appendix F. This term is needed to get the accuracy for the circle comparable to the other two.

The theoretical \( E_n \) and the result of the Monte Carlo tests are shown in Figure 11. The formulas in (32) conform so well with the Monte Carlo tests all the way from \( n = 3 \) so that the curves hardly can be distinguished in Figure 11. We have \( E_3(K) = 3 \) for all \( K \). The formulas in (32) give \( E_3(\text{triangle}) = 3.52 \), \( E_3(\text{square}) = 2.62 \), and \( E_3(\text{circle}) = 3.14 \).
$E_n$(square) can be compared with the averages in the last row of Table 2, which holds the numbers

\[(33) \quad E_q(n) = \sum_{k=3}^{n} k q_n(k).\]

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure.png}
\caption{Expected number of vertices of the convex hull of $n$ random points inside a triangle, square, or circle. The thicker lines are theoretical, the thinner are from Monte Carlo tests. The difference between theory and tests is hardly discernable.}
\end{figure}
The function $\nu(x)$ is defined

$$\nu(x) = \text{dilog}(x) + \log(x) \log |1 - x| =$$

$$= \int_x^1 \frac{\log(t)}{t-1} dt + \log(x) \log |1 - x| = \int_1^x \frac{\log |1 - s|}{s} ds.$$  

(34)

Figure 12. The function $\nu(x)$.

The dilog function and some of its properties are described in [1], page 1004. The dilog function has a series expansion in a unit circle centered at $x = 1$. It follows from the last integral in (34) that the $\nu$ function is defined and is real on the whole real axis. We use this function for $0 \leq x \leq 2$.

$\nu(x)$ is decreasing from $\nu(0) = \pi^2/6$ via $\nu(1) = 0$ to $\nu(2) = -\pi^2/12$.

The integrals involving $\nu(x)$ needed for calculating the moments of various distributions take rational values like

$$\int_0^1 x \, d\nu(x) = -1, \quad \int_1^2 x \, d\nu(x) = -1,$$

$$\int_0^1 x^2 \, d\nu(x) = -\frac{3}{4}, \quad \int_1^2 x^2 \, d\nu(x) = -\frac{5}{4},$$

$$\int_0^1 x^3 \, d\nu(x) = -\frac{11}{18}, \quad \int_1^2 x^3 \, d\nu(x) = -\frac{29}{18}.$$
We shall show in detail how the distribution function \( H_{4,13}(u) \) is calculated in Case 13. The quadrangle of this case is shown in Figure 4.

In analogy with the reasoning for triangle Case 2, we get that the fraction of the big rectangle is

\[
(35) \quad f_{13} = \frac{(1 - y_2)x_3 - (1 - x_2)y_3}{2}.
\]

We have

\[
(36) \quad H_{4,13}(u) = \text{Prob}(2f_{13} \leq u) = \text{Prob}((1 - y_2)x_3 - (1 - x_2)y_3 \leq u).
\]

We shall integrate over the set used in evaluating (21) though with the supplementary bound \((1 - y_2)x_3 - (1 - x_2)y_3 \leq u\). When \(x_2 - y_2 \leq u \leq 1 - y_2/x_2\), we shall integrate over the trapezoid in the \(x_3, y_3\)-plane marked in Figure 13a.

We get the trapezoid area

\[
(37) \quad A(x_2, y_2, u) = \frac{x_2(u - x_2 + y_2)}{2(1 - x_2)(x_2 - y_2)} + \frac{(u - x_2 + y_2)(x_2 - y_2 - ux_2)}{(1 - x_2)(x_2 - y_2)}.
\]

The area is zero for \(u \leq x_2 - y_2\). For \(u \geq 1 - y_2/x_2\), the area is the same as for \(u = 1 - y_2/x_2\), namely

\[
(38) \quad B(x_2, y_2, u) = \frac{(x_2 - y_2)(1 - x_2)}{2x_2}.
\]

These two expressions shall be integrated over the triangles marked in Figure 13b.

The obtained distribution function is
The density function \( dH_{4,13}/du \) is shown in Figure 14.

\[
H_{4,13}(u) = u(6 + 3u - 8u^2) + 6(1 - 3u)(1 - u) \log(1 - u) - 12u^2 \nu(u),
0 \leq u \leq 1.
\]

The corresponding density function \( dH_{4,13}/du \) is shown in Figure 14.

**APPENDIX C**

We shall show in detail how to calculate the distribution function \( H_{4,8}(u) \) for the area fraction in Case 8. The quadrangle of this case is shown in Figure 18. For ease of computation, we shall carry out the calculations for this quadrangle turned upside down, namely for the permutation 3214 which is the reverse of the one in Figure 18. Then, the conditions describing Case 8 are

\[
\frac{x_2}{x_3} + \frac{y_2}{y_1} \leq 1, \quad x_2 \leq x_3, \quad y_2 \leq y_1, \quad x_3(1 - y_2) + y_1(1 - x_2) \leq u.
\]

These inequalities describe the trapetsoidal domain in the \( x_2 \)\( y_2 \)-space marked in Figure 15a. For \( (1 - x_3)(1 - y_1) \geq 1 - u \) and \( x_3 + y_1 \geq u \), the area of the domain is

\[
A(x_3, y_1, u) = \frac{x_3 y_1}{2} - \frac{(x_3 + y_1 - u)^2}{2x_3 y_1}.
\]

The area is zero for \( (1 - x_3)(1 - y_1) \leq 1 - u \). For \( x_3 + y_1 \leq u \), the area is the the same as for \( u = x_3 + y_1 \), namely

\[
B(x_3, y_1, u) = \frac{x_3 y_1}{2}.
\]

For \( 0 \leq u \leq 1 \), these two expressions shall be integrated over the areas marked in Figure 15b. For \( 1 \leq u \leq 2 \) the areas are shown in Figure 15c. We get
Figure 15. Areas to integrate over in Case 8. (a) Trapezoid to integrate over in $x_2y_2$-plane. (b) Areas to integrate $A$ and $B$ over in $x_3y_1$-plane when $0 \leq u \leq 1$. (c) Areas to integrate $A$ and $B$ over in $x_3y_1$-plane when $1 \leq u \leq 2$.

\[
H_{4,8} = \begin{cases} 
-6u + 13u^2 - 2(3 - 5u)(1 - u)\log(1 - u) - \\
-4u^2\text{dilog}(1 - u), & 0 \leq u \leq 1, \\
-15 + 36u - 14u^2 - \pi^2u^2/3 + \\
+4(3u - 1)(u - 1)\log(u - 1) - 8u^2\nu(u), & 1 \leq u \leq 2. 
\end{cases}
\]

The corresponding density function $dH_{4,8}/du$ is shown in Figure 16.

Appendix D

The $H_{4,k}$-functions for the four point problem are given below. If no specification is given, the functions are defined for $0 \leq u \leq 1$. The cases are shown in Figures 17 and 18.
Figure 16. Density function $dH_{4,8}/du$ for two times the fraction that the quadrangle takes up of the 'big' rectangle in Case 8. The peak value is $16 - 4\pi^2/3$.

\[ H_{4,1} = 1 + (1 - u)^2(7 - 4u + 6 \log(1 - u)). \]
\[ H_{4,2} = 1 - (1 - u)(1 + 7u - 2u^2 + 6(1 - u) \log(1 - u)). \]
\[ H_{4,3} = u + (5u + 2(3 - u)(1 - u) \log(1 - u)). \]
\[ H_{4,4} = 6u^2 - 8u^3 + 3u^4. \]
\[ H_{4,5} = 4u^3 - 3u^4. \]
\[ H_{4,6} = (u - 1)(5u - 9 - 2(u + 1) \log(u - 1)), \quad 1 \leq u \leq 2. \]
\[ H_{4,7} = u + (5u + 2(3 - u)(1 - u) \log(1 - u)). \]
\[ H_{4,8} = \begin{cases} -6u + 13u^2 - 2(3 - 5u)(1 - u) \log(1 - u) - 4u^2 \text{dilog}(1 - u), & 0 \leq u \leq 1, \\ -15 + 36u - 14u^2 - \pi^2u^2/3 + 4(3u - 1)(u - 1) \log(u - 1) - 8u^2 \nu(u), & 1 \leq u \leq 2. \end{cases} \]
\[ H_{4,9} = 6u - 9u^2 + 4u^3 + 6(1 - u)^2 \log(1 - u). \]
\[ H_{4,10} = \begin{cases} -3u + 9u^2/2 - u^3 - 3(1 - u)^2 \log(1 - u), & 0 \leq u \leq 1, \\ -3 + 3u + 3u^2/2 - u^3 + 3(u - 1)^2 \log(u - 1), & 1 \leq u \leq 2. \end{cases} \]
\[ H_{4,11} = \begin{cases} 2u^3 - 5u^4/4, & 0 \leq u \leq 1, \\ -3 + 8u - 6u^2 + 2u^3 - u^4/4, & 1 \leq u \leq 2. \end{cases} \]
\[ H_{4,12} = 6u^2 - 8u^3 + 3u^4. \]
\[ H_{4,13} = u(6 + 3u - 8u^2) + 6(1 - 3u)(1 - u) \log(1 - u) - 12u^2 \nu(u). \]
Figure 17. Cases 1 - 6.
Case 7
perm = 3412
\[ u = 1 - (x_3 - x_2)(y_1 - y_4) \]

Case 8
perm = 4123
\[ u = (1 - x_2)(1 - y_1) + y_2x_3 \]

Case 9
perm = 1243
\[ u = 1 - y_2 - y_4(x_3 - x_2) \]

Case 10
perm = 1423
\[ u = 1 - x_2 + x_3(y_2 - y_4) \]

Case 11
perm = 1324
\[ u = x_3 - x_2 + y_2 - y_3 \]

Case 12
perm = 1234
\[ u = x_2 - x_3 + y_3 - y_2 \]

Figure 18. Cases 7 - 12. Case 13 is in Figure 4.
We shall deduce the limit formula for large \( n \) for the expected area fraction of a triangle

\[
E_{H_n}(\text{triangle}) = 1 - 2 \frac{\log(n) + \gamma}{n},
\]

where \( \gamma \) is Euler’s constant.

We shall use the technique of Rényi and Sulanke, [6] and start with their formula

\[
(42) \quad E_{H_n}(\text{triangle}) = \frac{\binom{n}{2}}{12F^3} \int_0^{2\pi} \int_0^{p(\phi)} (1 - f/F)^{n-2} l^4 p \, dp \, d\phi,
\]

where \( F \) is the area of the triangle, \( f \) is the smaller of the two areas that the triangle is divided into by the line \( x\cos(\phi) + y\sin(\phi) = p \) and \( l \) is the length of this dividing cut. The origin is supposed to be inside the triangle and we shall use the equilateral triangle depicted in Figure 19, which has the area \( F = 3\sqrt{3}/4 \). \( p(\phi) \) is the largest \( p \) for which the line intersects the triangle. By the symmetry of the triangle, the \( \phi \)-integral can be replaced by 6 times the integral from 0 to \( \pi/3 \). We have two cases: (i) the line intersects the top and bottom sides of the triangle or (ii) the line intersects the top and right hand sides of the triangle. We have case (i) when

\[
0 \leq p \leq p_1(\phi) = \frac{\cos(\phi) - \sqrt{3}\sin(\phi)}{2}.
\]
In case (i), we have
\[ l_1 = 2\sqrt{3}\frac{\cos(\phi) + p}{4\cos(\phi)^2 - 1} \quad \text{and} \quad r_1 = 1 - \frac{f}{F} = \frac{4\cos(\phi)^2 + 2p\cos(\phi) + p^2}{3 \cdot 4\cos(\phi)^2 - 1}. \]

We have case (ii) when
\[ p_1(\phi) \leq p \leq p(\phi) = \frac{\cos(\phi) + \sqrt{3}\sin(\phi)}{2}. \]

In case (ii), we have
\[ l_2 = \frac{p(\phi) - p}{\sin(\phi)\sqrt{1 + 2\cos(\phi)(\cos(\phi) + \sqrt{3}\sin(\phi))}} \]

and
\[ r_2 = 1 - \frac{f}{F} = 1 - \frac{(p(\phi) - p)^2}{3\sqrt{3} \left(1 - \cos(\phi)(\cos(\phi) - \sqrt{3}\sin(\phi))\right)}. \]

The integrals arising from the two cases will be denoted \( J_1 \) and \( J_2 \), so that
\[ E_{H_n}(\text{triangle}) = J_1 + J_2, \]

where
\[ J_1 = \frac{16n(n - 1)}{81\sqrt{3}} \int_0^{\pi/6} \int_0^{p_1(\phi)} r_1^{n-2} l_1^4 p \, dp \, d\phi \]

and
\[ J_2 = \frac{16n(n - 1)}{81\sqrt{3}} \int_0^{\pi/3} \int_{\max(p_1(\phi),0)}^{p(\phi)} r_2^{n-2} l_2^4 p \, dp \, d\phi. \]

Here, we integrate over \( 0 \leq \phi \leq \pi/6 \) in \( J_1 \) because \( p_1 \leq 0 \) for \( \phi \geq \pi/6 \). For the evaluation of \( J_1 \), we substitute \( p \) by \( x \) by putting \( p = x\cos(\phi) \). The integral over \( p \) then transforms to
\[ \int_0^{(1 - \sqrt{3}\tan(\phi))/2} (1 + x)^{2n} x \, dx = \left[ \frac{(1 + x)^{2n+1}}{2(n + 1)} \left( x - \frac{1}{2n + 1} \right) \right]_0^{(1 - \sqrt{3}\tan(\phi))/2} \]

After inserting the bounds, we have a remaining \( \phi \)-integral in which we substitute \( \phi \) by \( s \) by putting \( \tan(\phi) = \sqrt{3}(1 - s)/(1 + s) \). After skipping terms of the order \( 9^{-n} \), we get.

\[ J_1 \approx \frac{n(n - 1)}{3(n + 1)(2n + 1)} \int_{1/2}^{1} \left[ (4n + 1)s^n - 2(n + 1)s^{n-1} \right] ds = \]

\[ = \frac{n(n - 1)}{3(n + 1)(2n + 1)} \left[ \frac{(4n + 1)s^{n+1}}{n + 1} - \frac{2(n + 1)s^n}{n} \right]_{1/2}^{1}. \]

We get for large \( n \)
\[ J_1 \approx \frac{1}{3} - \frac{5}{3n} + O \left( \frac{1}{n^2} \right). \]
For the evaluation of $J_2$, we substitute $p$ by $x$ and $\phi$ by $t$ by putting

$$p = p(\phi) - \sqrt{\frac{3\sin(\phi)(\sin(\phi) + \sqrt{3}\cos(\phi))}{2n}} \quad \text{and} \quad \phi = \arctan\left(\frac{\sqrt{3}t}{2n-t}\right).$$

After tedious manipulations, we get

$$J_2 = \frac{n}{3n^2} \int_0^n dt \int_0^{x_1(t)} \left(1 - \frac{x}{n}\right)^{n-2} x^{3/2} (nt^{-3/2} + t^{-1/2} - 3x^{1/2}t^{-1}) \, dx,$$

where

$$x_1(t) = \begin{cases} t, & 0 \leq t \leq n/2 \\ \frac{(n+t)^2}{9t}, & n/2 \leq t \leq n. \end{cases}$$

For $n/2 \leq t \leq n$ we have $4n/9 \leq (n+t)^2/9t \leq n/2$, meaning that $(1 - \frac{x}{n})^{n-2}$ is smaller than $(5/9)^{n-2}$ for these $t$. By putting $x_1(t) = n/2$ for $n/2 \leq t \leq n$ and reversing the order of integration, we have with good approximation for large $n$

$$J_2 \approx \frac{n}{3n^2} \int_0^{n/2} \left(1 - \frac{x}{n}\right)^{n-2} x^{3/2} dx \int_x^n (nt^{-3/2} + t^{-1/2} - 3x^{1/2}t^{-1}) \, dt =$$

$$= \frac{n}{3n^2} \int_0^{n/2} \left(1 - \frac{x}{n}\right)^{n-2} \left[2(n-x)x - 3(\log(n) - \log(x))x^2\right] \, dx.$$

We make an error of at most the order $n^22^{-n/2}$ by extending the upper limit of integration to infinity. We have the following expansions

$$\int_0^\infty \left(1 - \frac{x}{n}\right)^{n-2} x \, dx = 1 + \frac{1}{n} - \frac{18}{n^2} + O\left(\frac{1}{n^3}\right),$$

$$\int_0^\infty \left(1 - \frac{x}{n}\right)^{n-2} x^2 \, dx = 2 - \frac{40}{n^2} + O\left(\frac{1}{n^3}\right),$$

$$\int_0^\infty \left(1 - \frac{x}{n}\right)^{n-2} x^2 \log(x) \, dx = 3 - 2\gamma + O\left(\frac{1}{n}\right).$$

Using these expansions for the above integral, we get

$$J_2 \approx \frac{2}{3} - \frac{2\log n + 2\gamma + 5/3}{n} + O\left(\frac{1}{n^2}\right).$$

Altogether, we have

$$E_{H_n}(\text{triangle}) = J_1 + J_2 \approx 1 - \frac{2\log(n) + \gamma}{n} + O\left(\frac{1}{n^2}\right),$$

where $\gamma$ is Euler’s constant.
We shall deduce the formula for the expected number of vertices of the convex hull of $n$ points inside a circle given in (32). We use the method of [5] and are led to the evaluation of the following integral, similar to (42)

\[
E_n(\text{circle}) = \frac{(\pi/2)^2}{3\pi} \int_0^{\pi/2} \int_0^{p(\phi)} (1 - f/F)^{n-2} dp \, d\phi,
\]

where $F$ is the area of the circle. We use a circle with radius one centered at the origin. The cutting line is $x = p$. We put $p = \cos(\alpha)$ and get the length of the cut $l = 2\sin(\alpha)$. The area of the circle segment is $f = \alpha - \sin(\alpha) \cos(\alpha)$. We get

\[
E_n = \frac{8n(n-1)}{3\pi} \int_0^{\pi/2} \left(1 - \frac{\alpha - \sin(\alpha) \cos(\alpha)}{\pi}\right)^{n-2} \sin(\alpha)^4 \, d\alpha.
\]

Series expansion of the trigonometric functions gives

\[
E_n \approx \frac{8n(n-1)}{3\pi} \int_0^{\pi/2} \left(1 - \frac{1}{\pi} \left(\frac{2}{3} \alpha^3 + \frac{2}{15} \alpha^5 + O(\alpha^7)\right)\right)^{n-2} \cdot \alpha^4 \left(1 - \frac{\alpha^2}{6}\right)^4 \, d\alpha.
\]

Use the substitution $\frac{2\alpha^3}{3\pi} = \frac{\pi}{n}$ and get

\[
E_n \approx 2(n^{1/3} - n^{-2/3}) \lambda \int_0^{\pi/2} \left(1 - \frac{x}{n} + \mu \left(\frac{x}{n}\right)^{5/3} + O\left(\left(\frac{x}{n}\right)^{7/3}\right)\right)^{n-2} \cdot \left(1 - \lambda \left(\frac{x}{n}\right)^{2/3}\right)^4 \frac{x^{2/3}}{x^{2/3}} \, dx,
\]

where $\lambda = \frac{2}{3} \left(\frac{3\pi}{2}\right)^{2/3}$ and $\mu = \frac{1}{5} \left(\frac{3\pi}{2}\right)^{2/3}$.

We have

\[
\left(1 - \frac{x}{n} + \mu \left(\frac{x}{n}\right)^{5/3} + O\left(\left(\frac{x}{n}\right)^{7/3}\right)\right)^{n-2} = e^{-x} \left(1 + \mu x^{5/3} - x^{2/3} - 2x/n + 2x/n + O\left(n^{-5/3}\right)\right) = e^{-x} \left(1 + \mu x^{5/3} - x^{2/3} - 2x/n + 2x/n + O\left(n^{-5/3}\right)\right).
\]

Inserting this in the integral, we get

\[
E_n \approx n^{1/3} \frac{2\lambda}{2} \int_0^{\pi/2} e^{-x} x^{2/3} \, dx - n^{-1/3} 2\lambda^2 \int_0^{\pi/2} e^{-x} x^{4/3} \, dx +
\]

\[
+ n^{-1/3} 2\lambda \mu \int_0^{\pi/2} e^{-x} x^{7/3} \, dx + O(n^{-2/3}).
\]

We make an error of negative exponential order by extending the upper limits of the integrals to infinity and get
\[ E_n(\text{circle}) \approx \frac{4}{3} \left( \frac{3\pi}{2} \right)^{2/3} n^{1/3} - \frac{4}{15} \left( \frac{3\pi}{2} \right)^{4/3} n^{-1/3} + O(n^{-2/3}) \approx 3.3832 \, n^{1/3} - 2.5084 \, n^{-1/3}. \]

REFERENCES


DEPARTMENT OF MATHEMATICS, ROYAL INSTITUTE OF TECHNOLOGY, S-10044 STOCKHOLM SWEDEN

E-mail address: johanph@math.kth.se