Random Convex Hulls: A Variance Revisited

Steven Finch* and Irene Hueter†

Abstract

An exact expression is determined for the asymptotic constant c_2 in the limit theorem of Groeneboom [4] (1988) which states that the number of vertices of the convex hull of a uniform sample of n random points from a circular disk satisfies a central limit theorem as $n \to \infty$ with asymptotic variance $2\pi c_2 n^{1/3}$.

1 Introduction

The convex hull of random point sets that arises in the statistical analysis of multivariate data, as explained elsewhere a great number of times, forms a vital and widely studied subject in stochastic geometry. An alternative point of view looks at the convex hull of a random point set as a graph on random points in which the vertices of the convex hull are connected. A multitude of questions, especially in three and higher dimensions and in the non-uniform case, remain open.

Consider the convex hull K_n of n independent and uniformly distributed points, drawn from a fixed circular disk K in $\mathbf{R}^2$. Let N_n and D_n denote the number of vertices of the random polygon K_n and the area of $K \setminus K_n$, the area in K outside the convex hull K_n, respectively. Groeneboom [4] in 1988 made substantial progress in our understanding of N_n. He introduced an elegant method to derive its asymptotic k-th moments for $k \geq 1$ and proved the central limit theorem that as $n \to \infty$,

$$\frac{N_n - 2\pi\, c_1\, n^{1/3}}{\{2\pi c_2\, n^{1/3}\}^{1/2}} \xrightarrow{d} Z \sim \mathcal{N}(0,1), \tag{1.1}$$

where "$\xrightarrow{d}$" denotes convergence in distribution, $\mathcal{N}(0,1)$ stands for the standard normal distribution, $c_1 = (3\pi/2)^{-1/3}\,\Gamma(\frac{5}{3}) \approx 0.5384576135...$ and c_2 is expressed in terms of two rather complicated double integrals whose approximate values are obtained numerically in his paper. (We will see shortly that his stated numerical value 0.33503... for c_2 is incorrect.)

*Clay Mathematics Institute. This research was partially supported by a Clay Book Fellowship.

†City University of New York, Baruch College, and University of Massachusetts Boston. Partially supported by PSC-CUNY grant #60032-33-34.

Date: March 17, 2004; revised August 20, 2004.

AMS 2000 Math. Subject Classification: *Primary:* 52A22. *Secondary:* 60D05.

Key words and phrases: convex hull, random point set, random polygon, variance.

Another result in [4] says that, when K is replaced by a fixed convex polygon with $r \geq 3$ vertices in the plane, then N_n as $n \to \infty$ satisfies a central limit theorem with asymptotic mean $(2/3)\, r \log n$ and variance $(10/27)\, r \log n$ (HUETER [6] proves a central limit theorem for N_n when the sample is normal rather than uniform). While the exact value of c_1 (in the case of a circular disk) goes back to the work of RÉNYI AND SULANKE [7] in 1963, to the best of the authors' knowledge, it is an open problem to determine the exact value for c_2. The purpose of this note is to do so. BUCHTA [2] points out at the conclusion of Section 5.2 that finding an exact expression for c_2 may also identify the variance in a related question: HSING [5] in 1994 demonstrated that

$$\lim_{n\to\infty} n^{5/3}\, \mathrm{Var}(D_n) = \sigma^2,$$

and as $n \to \infty$,

$$n^{5/6}\,(D_n - E(D_n)) \xrightarrow{d} \tilde{Z} \sim \mathcal{N}(0, \sigma^2) \tag{1.2}$$

for a finite nonnegative number σ^2 (BUCHTA refers to σ^2 as Hsing's constant). The parameter σ^2 is in terms of an integral. How to evaluate this integral is not obvious nor is the strict inequality $\sigma^2 > 0$ apparent. Let area(K) denote the area of the disk K. BUCHTA [2] explains that, if $\mathrm{Var}(N_n) \sim 2\pi\, c_2\, n^{1/3}$ as $n \to \infty$, then

$$\sigma^2 = 2\pi(\frac{1}{3}\, c_1 + c_2)\,(\mathrm{area}(K))^2 = \theta\,(\mathrm{area}(K))^2. \tag{1.3}$$

We remark that it is stated (without proof) in GROENEBOOM [4], page 328, that $\mathrm{Var}(N_n) \sim 2\pi\, c_2\, n^{1/3}$ as $n \to \infty$, in other words, the variance of the asymptotic normal distribution of N_n equals the asymptotic variance of N_n. Under the assumption that the relationship in (1.3) between σ^2 and c_2 holds, applying the Efron identity (EFRON [3], 1965)

$$\frac{ED_n}{\mathrm{area}(K)} = \frac{EN_{n+1}}{n+1}$$

in combination with the result of RÉNYI AND SULANKE on EN_n, we may restate the limit theorem in (1.2) as

$$\frac{(\mathrm{area}(K))^{-1}\, D_n - 2\pi c_1\, n^{-2/3}}{\{2\pi(\frac{1}{3}\, c_1 + c_2)\, n^{-5/3}\}^{1/2}} \xrightarrow{d} Z \sim \mathcal{N}(0, 1) \tag{1.4}$$

as $n \to \infty$. Our contribution is the following

THEOREM **1** *The asymptotic constant c_2 is given by*

$$c_2 = \frac{16\pi^2\, \Gamma(\frac{2}{3})^{-3} - 57}{27}\, c_1 \approx 0.1316029298..., \tag{1.5}$$

where $c_1 = (3\pi/2)^{-1/3}\, \Gamma(\frac{5}{3}) \approx 0.5384576135...$, and the constant θ is

$$\theta = 2\pi(\frac{1}{3}\, c_1 + c_2) = \frac{32\pi(\pi^2\, \Gamma(\frac{2}{3})^{-3} - 3)}{27}\, c_1 \approx 1.9546285835.... \tag{1.6}$$

Remarks. (1) We see from (1.5) that the limiting ratio c_2/c_1 of the variance and the expectation of the asymptotic normal distribution of N_n as $n \to \infty$ is 0.2444072226..., which is smaller than 5/9, the analogous ratio in the case when a convex polygon with $r \geq 3$ vertices in the plane replaces the disk K.
(2) The discrepancy in the numerical value for c_2 in [4] might be explainable by the factor 2 in front of the double integral in the formula after display (3.35) not having been included in the numerical evaluation.

2 Proof of Theorem 1

In order to determine an exact expression for c_2 in Theorem 3.4 in [4], it suffices to do so for c_2 in Theorem 3.3(i) in [4] by deriving an asymptotic expression for $\mathrm{Var}(N(a))$ as $a \to \infty$. For convenience, we recall a couple of definitions from [4] that lead up to the one of $N(a)$.

Let $\{T(a) = (R(a), S(a)) : a \in \mathbf{R}\}$ denote the stationary jump process that arises from the jump process $\{Z(a) = (X(a), Y(a)) : a \in \mathbf{R}\}$ (see Definition 3.3) upon the transformations in (3.18) and (3.19) in [4], p. 356. The process $\{Z(a) : a \in \mathbf{R}\}$ relies on the approximation of the sample point process of size n from the unit disk by a Poisson point process on $\mathbf{R}^2$ with intensity $(n/\pi)\times$ Lebesgue measure (see Lemma 3.2). With these defined, $N(a)$ denotes the number of jumps of the process $\{T(c) : c \geq 0\}$ on the time interval $[0, a]$ (p. 357, [4]). We refer the interested reader to [4] for the full development that concludes with the limit law for N_n.

Lemma 3.7 lies at the core of the main results in [4]. Since the former result has a few misprints at crucial places in such a fashion that it is rather challenging to follow the exposition, starting on page 360 and running into page 361 of [4], we give a corrected version of that lemma: For $u, v > 0$, define

$$\begin{aligned} A_1(u,v) &= \sqrt{u} + \frac{1}{2}(u - v + 1) \\ A_2(u,v) &= \sqrt{v} + \frac{1}{2}(v - u + 1) \\ A_3(u,v) &= \frac{2}{3}(u^{3/2} + v^{3/2}) + \frac{1}{4}(u-v)^2 + \frac{1}{2}(u+v) - \frac{1}{12} \end{aligned} \tag{2.7}$$

and let $\mathcal{K}$ be the region $\mathcal{K} = \{(u,v) \in \mathbf{R}^2_+ : |1 - v^{1/2}| < u^{1/2} < 1 + v^{1/2}\}$. Then for $c > 0$,

$$\begin{aligned} E[(S(0)^{1/2} + R(0))^2 \cdot ((S(c)^{1/2} - R(c))^2] &= \\ &= \frac{c^{10}}{36\pi^2} \iint_{\mathcal{K}} A_1(u,v)^3\, A_2(u,v)^3\, e^{-A_3(u,v)c^3/(2\pi)}\, du\, dv \\ &+ \frac{c^7}{4\pi} \iint_{\mathcal{K}} A_1(u,v)^2\, A_2(u,v)^2\, e^{-A_3(u,v)c^3/(2\pi)}\, du\, dv \\ &+ \frac{16}{9\pi^2} \iint_{c-\sqrt{v}>\sqrt{u}} u^{3/2}\, v^{3/2}\, e^{-2(u^{3/2}+v^{3/2})/(3\pi)}\, du\, dv. \end{aligned} \tag{2.8}$$

(The factor 1/4 in $A_3(u,v)$ in (2.7) was 1/2 and the ratio $16/(9\pi^2)$ in (2.8) was $4/(9\pi^2)$ in [4].) Proceeding along the arguments relied on after display (3.35) in [4], as $a \to \infty$,

$$\begin{aligned}
\mathrm{Var}(N(a)) &\sim c_1 a + 2\int_0^a db \int_0^b \frac{1}{(2\pi)^2} E[(S(0)^{1/2}+R(0))^2(S(c)^{1/2}-R(c))^2]\, dc - (c_1 a)^2 \\
&= c_1 a + \frac{1}{2\pi^2}\int_0^a db \int_0^b [\frac{c^{10}}{36\pi^2}\iint_{\mathcal{K}} A_1(u,v)^3\, A_2(u,v)^3\, e^{-A_3(u,v)c^3/(2\pi)}\, du\, dv \\
&\quad + \frac{c^7}{4\pi}\iint_{\mathcal{K}} A_1(u,v)^2\, A_2(u,v)^2\, e^{-A_3(u,v)c^3/(2\pi)}\, du\, dv \\
&\quad + \frac{16}{9\pi^2}\{\iint_{c-\sqrt{v}>\sqrt{u}} u^{3/2}\, v^{3/2}\, e^{-2(u^{3/2}+v^{3/2})/(3\pi)}\, du\, dv \\
&\quad - \int_0^\infty u^{3/2}\, e^{-2u^{3/2}/(3\pi)}\, du \int_0^\infty v^{3/2}\, e^{-2v^{3/2}/(3\pi)}\, dv\ \}]\, dc \\
&= c_1 a + \frac{1}{2\pi^2}\int_0^a db\, [\frac{1}{36\pi^2}\iint_{\mathcal{K}} A_1(u,v)^3\, A_2(u,v)^3\, B_1(u,v)\, du\, dv \\
&\quad + \frac{1}{4\pi}\iint_{\mathcal{K}} A_1(u,v)^2\, A_2(u,v)^2\, B_2(u,v)\, du\, dv \\
&\quad - \frac{16}{9\pi^2}\iint_{c-\sqrt{v}<\sqrt{u}} u^{3/2}\, v^{3/2}\, e^{-2(u^{3/2}+v^{3/2})/(3\pi)}\, du\, dv] \\
&\sim a[c_1 + \frac{1}{2\pi^2}\{\frac{1}{36\pi^2}\,\frac{1}{3}\,(2\pi)^{11/3}\,\Gamma(11/3)\, I_1 + \frac{1}{4\pi}\,\frac{1}{3}\,(2\pi)^{8/3}\,\Gamma(8/3)\, I_2 - \frac{16}{9\pi^2} I_3\}],
\end{aligned} \tag{2.9}$$

where $B_1(u,v)$ and $B_2(u,v)$ are given at the bottom of page 361 in [4], I_1 and I_2 are the two double integrals

$$\begin{aligned}
I_1 &= \iint_{\mathcal{K}} A_1(u,v)^3\, A_2(u,v)^3\, A_3(u,v)^{-11/3}\, du\, dv, \\
I_2 &= \iint_{\mathcal{K}} A_1(u,v)^2\, A_2(u,v)^2\, A_3(u,v)^{-8/3}\, du\, dv,
\end{aligned}$$

and the integral I_3 is evaluated in [4], p. 362, and comes out as $I_3 = \frac{8}{9}(3\pi/2)^{11/3}\,\Gamma(5/3)$.

Let us now turn to the evaluation of I_1 and I_2. For this purpose, we make the substitution $u = (x+y)^2$ and $v = (x-y)^2$, equivalently, $x = (u^{1/2}+v^{1/2})/2$ and $y = (u^{1/2}-v^{1/2})/2$. This transforms the region $\mathcal{K}$ into the semi-infinite strip $\tilde{\mathcal{K}} = \{(x,y) : 1/2 \le x < \infty, -1/2 \le y \le 1/2\}$. Also, we obtain $du\, dv = 8(x^2-y^2)dx\, dy$, while $A_1(u,v)A_2(u,v)$ and $A_3(u,v)$ become $A_1(x,y)A_2(x,y) = (2x+1)^2(1-4y^2)/4$ and $A_3(x,y) = (2x+1)^2(\frac{x}{3}+y^2-\frac{1}{12})$. Consequently,

$$\begin{aligned}
I_1 &= \frac{1}{8}\iint_{\tilde{\mathcal{K}}}(x^2-y^2)\,(2x+1)^{-4/3}\,(1-4y^2)^3\,(\frac{x}{3}+y^2-\frac{1}{12})^{-11/3}\, dx\, dy \\
I_2 &= \frac{1}{2}\iint_{\tilde{\mathcal{K}}}(x^2-y^2)\,(2x+1)^{-4/3}\,(1-4y^2)^2\,(\frac{x}{3}+y^2-\frac{1}{12})^{-8/3}\, dx\, dy.
\end{aligned} \tag{2.10}$$

Some computations that we report in the Appendix provide

$$I_1 = -(9/8)6^{2/3} + (27/5)2^{-1/3}\,3^{7/6}\, I_4 = 4.1603038934..., \tag{2.11}$$

where $I_4 = \int_{4^{-1/3}}^{1} (4t^3 - 1)^{-1/2}\, dt$. A glimpse at formulas (18.13.17) and (18.13.24) of ABRAMOWITZ AND STEGUN [1] reveals that, alternatively, $I_4 = \Gamma(1/3)^3/(12\pi)$. Similarly, in a couple of steps, one derives an expression for I_2, which simplifies to

$$I_2 = 6^{5/3}/5 = 3.9623126986.... \tag{2.12}$$

Furthermore, one shows that I_1 and I_2 are related by

$$I_1 = \frac{8\pi^2\,\Gamma(2/3)^{-3} - 15}{16}\, I_2. \tag{2.13}$$

Finally, plugging I_1, I_2, and I_3 into the expression for $\mathrm{Var}(N(a))$ in (2.9), appealing to the identity $\Gamma(w+1) = w\,\Gamma(w)$ repeatedly as necessary, and recalling the formula for c_1, after a sequence of elementary steps, we arrive at, as $a \to \infty$,

$$\mathrm{Var}(N(a)) \quad \sim \quad \frac{a\,c_1}{27}\,[16\pi^2\,\Gamma(2/3)^{-3} - 57].$$

This completes the proof of the statement in (1.5). Since the assertion in (1.6) is readily verified, this finishes our proof of Theorem 1.

A Appendix: Identity in (2.11)

We use the shorthand notation $\alpha = y^2 - 1/12$, $\beta = 1 - 6\alpha$ and $\gamma = x + 3\alpha$ and write

$$\begin{aligned}
g_1(x,y) &= (\gamma/3)^{1/3}(2x+1)^{2/3}\{(-243/8)\beta^{-2}\,\gamma^{-3} + (1701/10)\beta^{-3}\,\gamma^{-2} \\
&\quad - (9963/10)\beta^{-4}\,\gamma^{-1} - 1944\beta^{-4}\,(2x+1)^{-1}\} \\
g_2(x,y) &= (\gamma/3)^{1/3}\,(2x+1)^{2/3}\{(-2187/8)\alpha^2\beta^{-2}\,\gamma^{-3} + (729/10)\alpha(4-3\alpha)\,\beta^{-3}\,\gamma^{-2} \\
&\quad + (243/10)(-5 - 36\alpha + 27\alpha^2)\,\beta^{-4}\,\gamma^{-1} - 486\beta^{-4}\,(2x+1)^{-1}\} \\
g_3(y) &= -(27/20)\,6^{2/3}\,(7/4 + y^2) \\
g_4(y) &= -54 \cdot 3^{2/3}\,(1+12y^2)^{-8/3}\,(1/5 + 4y^2/5 + 144y^4/5 + 64y^6) \\
G_3(y) &= -(9/20)\,6^{2/3}\,y\,(21/4 + y^2) \\
G_4(y) &= -(27/5)\,3^{2/3}\,y\,(1+12y^2)^{-5/3}(1 + 4y^2 + 32y^4).
\end{aligned}$$

Integrating with respect to x, we obtain

$$\begin{aligned}
I_1 &= \frac{1}{8}\int_{-1/2}^{1/2} [\,g_2(x,y) - y^2\,g_1(x,y)\,|_{1/2}^{\infty}\,] \cdot (1-4y^2)^3\,dy \\
&= \int_{-1/2}^{1/2} (g_3(y) - g_4(y))\,dy \\
&= (G_3(y) - G_4(y))|_{-1/2}^{1/2} + (27/5)\,3^{2/3}\int_{-1/2}^{1/2} (1+12y^2)^{-2/3}\,dy \\
&= -(9/8)\,6^{2/3} + (27/5)\,3^{2/3}\,2^{-1/3}\,3^{1/2}\int_{4^{-1/3}}^{1} (4t^3-1)^{-1/2} dt \qquad (3.1) \\
&= -(9/8)\,6^{2/3} + (27/5)\,2^{-1/3}\,3^{7/6}\,I_4,
\end{aligned}$$

where the substitution $t = 4^{-1/3}(1 + 12y^2)^{1/3}$ provides the Weierstrass elliptic integral in (3.1).

Acknowledgements.
We are grateful to JOE KEANE for his expertise in evaluating the integrals I_1 and I_2, to CHRISTIAN BUCHTA for providing us with a copy of his unpublished manuscript [2], and to a referee for useful comments.

References

[1] ABRAMOWITZ, M. AND STEGUN, I.A. (1972). *Handbook of Mathematical Functions,* 9th printing. Dover Publications, Inc., New York.

[2] BUCHTA, C. (2002). An identity relating moments of functionals of convex hulls. Unpublished manuscript.

[3] EFRON, B. (1965). The convex hull of a random set of points. *Biometrika* **52**, 331–343.

[4] GROENEBOOM, P. (1988). Limit theorems for convex hulls. *Probability Theory and Related Fields* **79**, 327–368.

[5] HSING, T. (1994). On the asymptotic distribution of the area outside a random convex hull in a disk. *Annals of Applied Probability* **4**, 478–493.

[6] HUETER, I. (1994). The convex hull of a normal sample. *Advances in Applied Probability* **26**, 855–875.

[7] RÉNYI, A. AND SULANKE, R. (1963). Über die konvexe Hülle von n zufällig gewählten Punkten. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **2**, 75–84.

STEVEN FINCH, CLAY MATHEMATICS INSTITUTE,
ONE BOW STREET, CAMBRIDGE, MA 02138
Steven.Finch@inria.fr
http://pauillac.inria.fr/algo/bsolve/

IRENE HUETER, DEPARTMENT OF MATHEMATICS, BARUCH COLLEGE-CUNY,
BOX B6-230, ONE BERNARD BARUCH WAY, NEW YORK, NY 10010
irene_hueter@baruch.cuny.edu
http://www.math.baruch.cuny.edu/~ihueter/