The number of empty four-gons in random point sets

R. Fabila-Monroy [1,3]

Departamento de Matemáticas
Centro de Investigación y de Estudios Avanzados del IPN
México

C. Huemer [2,4]

Dept. de Matemàtica Aplicada IV
Universitat Politècnica de Catalunya, BarcelonaTech
Catalonia

D. Mitsche [5]

Laboratoire J.A.Dieudonné
Université de Nice Sophia-Antipolis
France

Abstract

Let $\mathcal{S}$ be a set of n points distributed uniformly and independently in the unit square. Then the expected number of empty four-gons with vertices from $\mathcal{S}$ is $\Theta(n^2 \log n)$. A four-gon is empty if it contains no points of $\mathcal{S}$ in its interior.

Keywords: random point set, four-gon, empty polygon, geometric probability.

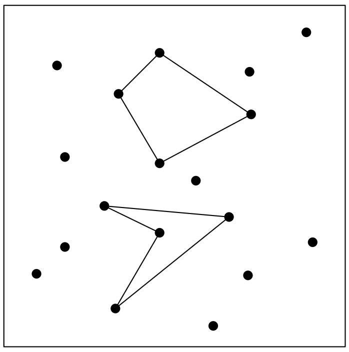

Figure 1. An empty convex four-gon and an empty non-convex four-gon in $\mathcal{S}$.

Throughout this paper let $\mathcal{S}$ be a set of n points distributed uniformly and independently in the unit square. Since with probability 1 the n points will be in general position (no three points are collinear), we may and will assume this throughout the paper. All asymptotics in this paper are w.r.t. the number of points n, that is, when $n \to \infty$. As our results are asymptotic, we may ignore also rounding issues throughout the paper, that is, if for some constant $c > 0$, cn is not an integer, depending on the context, we may and will consider $\lfloor cn \rfloor$ or $\lceil cn \rceil$ without changing the results. A four-gon whose vertices are from $\mathcal{S}$ is empty if it contains no other point from $\mathcal{S}$ in its interior. A four-gon can be convex or non-convex, see Figure 1.

Denote by N_4 the random variable that counts the number of empty non-convex four-gons with vertices from $\mathcal{S}$. Our main result is the following:

Theorem 0.1 $\mathbb{E}[N_4] = \Theta(n^2 \log n)$.

Denote by C_4 the random variable that counts the number of empty convex four-gons with vertices from $\mathcal{S}$. Complementing Theorem 0.1, we obtain the following result, which might be known, but we only found a proof for the lower bound [6]. There, also another construction with $O(n^2)$ empty convex four-gons is given.

Theorem 0.2 $\mathbb{E}[C_4] = \Theta(n^2)$.

[1] Supported by CONACyT of Mexico grant 153984.
[2] Supported by projects MEC MTM2012-30951 and Gen. Cat. DGR 2009SGR1040, 2014SGR46 and ESF EUROCORES programme EuroGIGA, CRP ComPoSe: grant EUI-EURC-2011-4306.
[3] Email: `ruyfabila@math.cinvestav.edu.mx`
[4] Email: `clemens.huemer@upc.edu`
[5] Email: `dmitsche@unice.fr`

We omit the proof of Theorem 0.2 in this abstract due to lack of space. [6] We also remark that Theorems 0.1 and 0.2 also hold for other convex, bounded sets, not only the square. In related works, for point sets of n points distributed uniformly and independently in a convex, bounded set, Valtr [13] proved that the expected number of empty triangles is at most $2n^2 - 2n$ and at least $2n^2 - o(n^2)$, and Bárány and Füredi [6] proved that, in R^d, the number of empty simplices is at most $K\binom{n}{d}$, for some constant K. Balogh et al. [5] showed that the expected number of vertices of the largest empty convex polygon in $\mathcal{S}$ (and in any convex, bounded set in the plane) is $\Theta(\frac{\log n}{\log \log n})$. A lot of research has been done to determine the minimum number $f_k(n)$ of empty convex k-gons among all sets of n points in general position in the plane (not only random point sets). For the case of empty triangles, Katchalski and Meir [11] showed that $f_3(n)$ is of order $\Theta(n^2)$. Later, this bound has been refined [2,6,7,8,9,13]; the currently best bounds are $n^2 - \frac{32n}{7} + \frac{22}{7} \leq f_3(n) \leq 1.6196n^2 + o(n^2)$. Concering empty convex four-gons, Bárány and Füredi [6] established that $f_4(n)$ is of order $\Theta(n^2)$, and the currently best bounds on $f_4(n)$ are $\frac{n^2}{2} - \frac{9}{4}n - o(n) \leq f_4(n) \leq 1.9397n^2 + o(n^2)$, see [2,7]. Research mainly focussed on empty convex polygons. Only recently the number of convex and non-covex polygons in point sets has been studied [1,3,4]. In [1] it is shown that every set of n points in general position in the plane determines at least $\frac{5n^2}{2} - \Theta(n)$ empty four-gons and a point set with only $O(n^{5/2}\log n)$ empty four-gons is given. Our result improves this bound to $O(n^2 \log n)$.

1 Proof of Theorem 0.1

The proof of Theorem 0.1 is implied by the following lemmas, for which we need some definitions. Fix three points p_a, p_b, p_c and denote by $\Delta(p_a, p_b, p_c)$ the triangle spanned by them. Let $\mathcal{P}$ be a set of $k \geq 1$ points distributed uniformly and independently in $\Delta(p_a, p_b, p_c)$. Denote by $\mathcal{E}^{\mathcal{P}}_{p_ap_b,p_bp_c}$ the event that $\mathcal{P} \cup \{p_a, p_b, p_c\}$ contains an empty non-convex four-gon with p_ap_b and p_bp_c among its edges.

Lemma 1.1 $\mathbb{P}(\mathcal{E}^{\mathcal{P}}_{p_ap_b,p_bp_c}) = \frac{2}{k+1}$.

Proof First observe that the points $\{p_a, p_b, p_c\}$ together with a fourth point $p_d \in \mathcal{P}$ form an empty non-convex four-gon with p_ap_b and p_bp_c among its edges if and only if the triangle $\Delta(p_a, p_d, p_c)$ contains $\mathcal{P}\backslash\{p_d\}$ in its interior. We now determine the distribution of the height h_d, which is the distance from

[6] The full version of this work has been submitted for publication in journal, 2014.

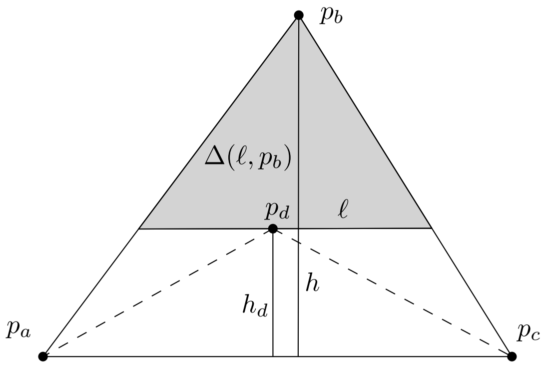

Figure 2. The triangles $\Delta(p_a, p_b, p_c)$, $\Delta(p_a, p_d, p_c)$ and $\Delta(\ell, p_b)$.

p_d to the segment $p_a p_c$. Denote by h the height of $\Delta(p_a, p_b, p_c)$ with respect to the edge $p_a p_c$. Let ℓ be the segment parallel to the edge $p_a p_c$, at distance h_d from this edge, and with endpoints on the edges $p_a p_b$ and $p_b p_c$ respectively. Assume w.l.o.g. that ℓ is a horizontal line with p_a and p_c below it. Define then $\Delta(\ell, p_b)$ to be the triangle coming from the intersection of $\Delta(p_a, p_b, p_c)$ and all points lying on or above ℓ, as shown in Figure 2. Since only relative heights between h and h_d matter, we may assume w.l.o.g. that $h = 1$. By the intercept theorem we have $\frac{|p_a p_c|}{h} = \frac{|\ell|}{h - h_d}$, where $|p_a p_c|$ and $|\ell|$ are the lengths of the segments $p_a p_c$ and ℓ. It follows that

$$\frac{\text{area}(\Delta(\ell, p_b))}{\text{area}(\Delta(p_a, p_b, p_c))} = \frac{\frac{|\ell|(h-h_d)}{2}}{\frac{|p_a p_c| h}{2}} = (1 - h_d)^2.$$

Hence, the distribution function F_{h_d} for the height h_d satisfies $F_{h_d}(x) = \mathbb{P}(h_d \leq x) = 1 - (1-x)^2$ and the density of the height h_d is $f_{h_d}(x) = 2 - 2x$ for $x \in [0, 1]$.

Fix any $p \in \mathcal{P} \setminus \{p_d\}$. Since the points are distributed uniformly at random inside $\Delta(p_a, p_b, p_c)$, we have

$$\mathbb{P}(p \in \Delta(p_a, p_d, p_c)) = \frac{\text{area}(\Delta(p_a, p_d, p_c))}{\text{area}(\Delta(p_a, p_b, p_c))} = h_d.$$

Therefore, integrating over all possible heights $0 \leq h_d \leq 1$,

$$\mathbb{P}(\mathcal{P} \setminus \{p_d\} \in \Delta(p_a, p_d, p_c)) = \int_0^1 x^{k-1}(2 - 2x)\ dx = \frac{2}{k(k+1)}.$$

As there are k choices for the point p_d, by a union bound, we have

$$\mathbb{P}(\mathcal{E}^{\mathcal{P}}_{p_a p_b, p_b p_c}) \leq \frac{2}{k+1}.$$

On the other hand, there is at most one point $p_d \in \mathcal{P}$ such that $\mathcal{P} \backslash \{p_d\} \in \Delta(p_a, p_d, p_c)$: indeed, if this were true for another point $p_e \neq p_d$, then $p_d \notin \Delta(p_a, p_e, p_c)$, contradicting the assumption. Hence,

$$\mathbb{P}((\mathcal{P} \backslash \{p_d\} \in \Delta(p_a, p_d, p_c)) \wedge (\mathcal{P} \backslash \{p_e\} \in \Delta(p_a, p_e, p_c))) = 0,$$

and thus

$$\mathbb{P}(\mathcal{E}^{\mathcal{P}}_{p_a p_b, p_b p_c}) = \bigcup_{p_d \in \mathcal{P}} \mathbb{P}(\mathcal{P} \backslash \{p_d\} \in \Delta(p_a, p_d, p_c))$$

$$= \sum_{p_d \in \mathcal{P}} \mathbb{P}(\mathcal{P} \backslash \{p_d\} \in \Delta(p_a, p_d, p_c)) = \frac{2}{k+1}.$$

□

For the next lemma, we need one more definition. Let T_k denote the random variable that counts the number of triangles with vertices in $\mathcal{S}$ containing exactly $k \geq 0$ points from $\mathcal{S}$ in its interior.

Lemma 1.2 *For any $k = k(n) \geq 0$, $\mathbb{E}[T_k] \leq 2n^2 - 2n$.*

Proof The density function $f_{area(\Delta)}(v)$ for the area v of a triangle Δ formed by three points chosen uniformly and independently in the unit square is given in [10], and also in [12]. From the results there one can see that for any area $v \geq 0, f_{area(\Delta)}(v) \leq 12$. Fix now three points $p_a, p_b, p_c \in \mathcal{S}$, let as before $\Delta(p_a, p_b, p_c)$ be the triangle spanned by them, and let $int(\Delta(p_a, p_b, p_c))$ denote the interior of this triangle. Denote also for $x, y > 0$ by $\beta(x, y) = \int_0^1 t^{x-1}(1-t)^{y-1} dt$ the beta function of x and y. Integrating over all possible areas v of the triangle $\Delta(p_a, p_b, p_c)$, we obtain

$$
\begin{aligned}
\mathbb{P}(|int(\Delta(p_a,p_b,p_c))\cap\mathcal{S}|=k) &= \int_0^{1/2}\binom{n-3}{k}v^k(1-v)^{n-3-k}f_{area(\Delta)}(v)\ dv\\
&\le 12\int_0^{1/2}\binom{n-3}{k}v^k(1-v)^{n-3-k}\ dv\\
&\le 12\int_0^{1}\binom{n-3}{k}v^k(1-v)^{n-3-k}\ dv\\
&= 12\binom{n-3}{k}\beta(k+1,n-k-2)\\
&= 12\binom{n-3}{k}\frac{k!(n-k-3)!}{(n-2)!}=\frac{12}{n-2}.
\end{aligned}
$$

Hence, by linearity of expectation, for any $k=k(n)\ge 0$,

$$
\mathbb{E}\left[T_k\right]\le\binom{n}{3}\frac{12}{n-2}=2n^2-2n.
$$

$\square$

Remark. The special case $k=0$ of Lemma 1.2 was also proved by Valtr in [13].

We are now ready to prove the upper bound of Theorem 0.1, which is encapsulated in the following lemma.

Lemma 1.3 $\mathbb{E}\left[N_4\right]=O(n^2\log n)$.

Proof Note that each triangle $\Delta(p_a,p_b,p_c)$ with vertices $p_a,p_b,p_c\in\mathcal{S}$ determines at most three empty non-convex four-gons such that p_a,p_b,p_c are the vertices on the boundary of the convex hull of the quadrilateral: indeed, any pair of edges from $\{p_ap_b,p_bp_c,p_ap_c\}$ can be chosen and might possibly give rise to an empty non-convex four-gon. Let $\mathcal{P}\subseteq\mathcal{S}$ denote the set of points in the interior of $\Delta(p_a,p_b,p_c)$, and let $\mathcal{E}^{\mathcal{P}}_{p_a,p_b,p_c}$ be the event that any of the three pairs of edges gives rise to an empty non-convex four-gon. By a union bound,

$$
\mathbb{P}(\mathcal{E}^{\mathcal{P}}_{p_a,p_b,p_c})\le\mathbb{P}(\mathcal{E}^{\mathcal{P}}_{p_ap_b,p_bp_c})+\mathbb{P}(\mathcal{E}^{\mathcal{P}}_{p_ap_c,p_bp_c})+\mathbb{P}(\mathcal{E}^{\mathcal{P}}_{p_ap_b,p_ap_c}),
$$

and thus, by Lemma 1.1, since the points of $\mathcal{P}$ are uniformly distributed in $\Delta(p_a,p_b,p_c)$ (if a point is distributed uniformly at random in the unit square, then conditional under knowing that it is inside a subarea of that square, it is still uniform in this subarea),

$$
\mathbb{P}(\mathcal{E}^{\mathcal{P}}_{p_a,p_b,p_c}\mid|\mathcal{P}|=k)\le\frac{6}{k+1}.
$$

By Lemma 1.2, for each $k \geq 0$ in expectation there are at most $2n^2$ triangles with k interior points and since once again, conditioned on having k interior points, their distribution is uniform inside the triangle, we obtain

$$\begin{aligned}\mathbb{E}[N_4] &\leq \sum_{k=0}^{n-3} \mathbb{E}[T_k]\, \mathbb{P}(\mathcal{E}^{\mathcal{P}}_{p_a,p_b,p_c} \mid |\mathcal{P}| = k) \\ &\leq 2n^2 \sum_{k=0}^{n-3} \frac{6}{k+1} = O(n^2 \log n).\end{aligned}$$

□

We now proceed to prove the corresponding lower bound of Theorem 0.1. We first prove the following lower bound on T_k.

Lemma 1.4 *For every $\epsilon > 0$ there exists some $\alpha = \alpha(\epsilon) > 0$ such that $\mathbb{E}[T_k] \geq (2-\epsilon)n^2$ for any $k = 0, 1, \ldots, \alpha n$.*

Proof By [10,12], the density function $f_{area(\Delta)}(v)$ for the area of the triangle $\Delta = \Delta(p_a, p_b, p_c)$, formed by three points p_a, p_b, p_c from $\mathcal{S}$, satisfies $f_{area(\Delta)}(0) = 12$ and is then strictly monotonically decreasing. In particular, for every small $\epsilon > 0$ there exists $v_0 > 0$ such that $f_{area(\Delta)}(v_0) = 12 - \epsilon$. We define $\alpha = 0.6 v_0$.

$$\begin{aligned}\mathbb{P}(|int(\Delta) \cap \mathcal{S}| = k) &= \int_0^{1/2} \binom{n-3}{k} v^k (1-v)^{n-3-k} f_{area(\Delta)}(v)\; dv \\ &\geq (12-\epsilon) \int_0^{v_0} \binom{n-3}{k} v^k (1-v)^{n-3-k}\; dv.\end{aligned}$$

As in the proof of Lemma 1.2 we have

$$\int_0^1 \binom{n-3}{k} v^k (1-v)^{n-3-k}\; dv = \frac{1}{n-2}.$$

We will show that $\mathbb{P}(|int(\Delta) \cap \mathcal{S}| = k) \geq \frac{12-\epsilon}{n-2} - o(\frac{1}{n})$, for $k \leq \alpha n$. To this end, it is sufficient to show that

$$\int_{v_0}^1 \binom{n-3}{k} v^k (1-v)^{n-3-k}\; dv = o(\frac{1}{n}). \tag{1}$$

Computing the derivative of the function $g(v) := v^k(1-v)^{n-3-k}$ in $[v_0, 1]$, we see that $g'(v) \leq 0$, implying that in $[v_0, 1]$, $g(v)$ is maximized at $v = v_0$. Thus,

$$\int_{v_0}^1 \binom{n-3}{k} v^k (1-v)^{n-3-k}\; dv \leq \binom{n-3}{k} {v_0}^k (1-v_0)^{n-3-k} (1-v_0).$$

It is easily verified that

$$\binom{n-3}{k} {v_0}^k (1-v_0)^{n-3-k} < \binom{n-3}{k+1} {v_0}^{k+1} (1-v_0)^{n-3-k-1}$$

holds for $k < v_0(n-2) - 1$. If we can show that

$$\binom{n-3}{k} {v_0}^k (1-v_0)^{n-3-k} (1-v_0) = o(\frac{1}{n})$$

holds for $k = \alpha n$, then it holds for all smaller values of k as well, and then also (1) holds for all $k = 0, 1, \ldots, \alpha n$. Using $\binom{n}{k} \leq \left(\frac{ne}{k}\right)^k$, we get for some constant $C > 0$

$$\begin{aligned}
\binom{n-3}{\alpha n} {v_0}^{\alpha n} (1-v_0)^{n-3-\alpha n} (1-v_0) &\leq C \left(\frac{e}{0.6 v_0}\right)^{0.6 v_0 n} v_0^{0.6 v_0 n} (1-v_0)^{(1-0.6 v_0) n} \\
&= C \left((e/0.6)^{0.6 v_0} (1-v_0)^{1-0.6 v_0}\right)^n \\
&= o(\frac{1}{n}),
\end{aligned}$$

where the last line follows from the fact that

$$f(v_0) := (\frac{e}{0.6})^{0.6 v_0} (1-v_0)^{1-0.6 v_0}$$

is monotone decreasing for $v_0 \in [0,1]$, $v_0 > 0$, and that $f(0) = 1$. Thus,

$$\begin{aligned}
\mathbb{P}(|int(\Delta) \cap \mathcal{S}| = k) &\geq (12-\epsilon) \int_0^{v_0} \binom{n-3}{k} v^k (1-v)^{n-3-k} \, dv \\
&\geq \frac{12-\epsilon}{n-2} - o(\frac{1}{n}).
\end{aligned}$$

As before, by linearity of expectation,

$$\mathbb{E}\left[T_k\right] \geq \binom{n}{3} \left(\frac{12-\epsilon}{n-2} - o(\frac{1}{n})\right) \geq (2-\epsilon) n^2$$

for any $k = 0, 1, \ldots, \alpha n$, thus concluding the proof. □

The lower bound of Theorem 0.1 now also follows easily.

Lemma 1.5 $\mathbb{E}\left[N_4\right] = \Omega(n^2 \log n)$.

Proof As before, define for three points $p_a, p_b, p_c \in \mathcal{S}$ by $\Delta(p_a, p_b, p_c)$ the triangle with vertices p_a, p_b, p_c. Let $\mathcal{P} \subseteq \mathcal{S}$ denote the set of points in the interior of $\Delta(p_a, p_b, p_c)$. Using the notation of Lemma 1.3, it is clear that

$$\mathbb{P}(\mathcal{E}^{\mathcal{P}}_{p_a,p_b,p_c}) \geq \mathbb{P}(\mathcal{E}^{\mathcal{P}}_{p_a p_b, p_b p_c}).$$

Hence, by Lemma 1.1,

$$\mathbb{P}(\mathcal{E}^{\mathcal{P}}_{p_a,p_b,p_c}) \geq \frac{2}{k+1}.$$

By Lemma 1.4, for $k = 0, 1, \ldots \alpha n$, in expectation there are at least $(2-\epsilon)n^2$ triangles with k interior points and since once again, conditioned on having k interior points, their distribution is uniform inside the triangle, we obtain

$$\begin{aligned} \mathbb{E}\left[N_4\right] &\geq \sum_{k=0}^{\alpha n} \mathbb{E}\left[T_k\right] \mathbb{P}(\mathcal{E}^{\mathcal{P}}_{p_a,p_b,p_c} \mid |\mathcal{P}| = k) \\ &\geq (2-\epsilon)n^2 \sum_{k=0}^{\alpha n} \frac{2}{k+1} = \Omega(n^2 \log n). \end{aligned}$$

□

References

[1] O. Aichholzer, R. Fabila-Monroy, H. González-Aguilar, T. Hackl, M.A. Heredia, C. Huemer, J. Urrutia, and B. Vogtenhuber, *4-holes in point sets*, Comput. Geom. **47(6)** (2014) 644-650.

[2] O. Aichholzer, R. Fabila Monroy, T. Hackl, C. Huemer, A. Pilz, and B. Vogtenhuber, *Lower bounds for the number of small convex k-holes*, Comput. Geom. **47(5)** (2014) 605–613.

[3] O. Aichholzer, R. Fabila-Monroy, H- González-Aguilar, T. Hackl, M.A. Heredia, C. Huemer, J. Urrutia, P. Valtr, and B. Vogtenhuber, *On k-gons and k-holes in point sets*, 23rd Annual Canadian Conference on Computational Geometry CCCG (2011), 21–26, Toronto, Canada.

[4] O. Aichholzer, T. Hackl, and B. Vogtenhuber, *On 5-gons and 5-holes*, XIV Encuentros de Geometría Computacional ECG2011, Lecture Notes in Comput. Sci. **7579** (2012) 1–13, Springer-Verlag, Berlin.

[5] J. Balogh, H. González-Aguilar, and G. Salazar, *Large convex holes in random point sets*, Comput. Geom. **46(6)** (2013) 725–733.

[6] I. Bárány and Z. Füredi, *Empty simplices in Euclidean space*, Canad. Math. Bull. **30(4)** (1987) 436–445.

[7] I. Bárány and P. Valtr, *Planar point sets with a small number of empty convex polygons*, Stud. Sci. Math. Hung. **41** (2004) 243–266.

[8] A. Dumitrescu, *Planar sets with few empty convex polygons*, Stud. Sci. Math. Hung. **36** (2000) 93–109.

[9] A. García, *A note on the number of empty triangles*, XIV Encuentros de Geometría Computacional ECG2011, Lecture Notes in Comput. Sci. **7579** (2012) 249–257, Springer-Verlag, Berlin.

[10] N. Henze, *Random triangles in convex regions*, J. Appl. Prob. **20** (1983) 111–125.

[11] M. Katchalski and A. Meir, *On empty triangles determined by points in the plane*, Acta Math. Hung. **51** (1988) 323-328.

[12] J. Philip, *The area of a random convex polygon*, http://www.math.kth.se/~johanph/area12.pdf, 2004.

[13] P. Valtr, *On the minimum number of empty polygons in planar point sets*, Stud. Sci. Math. Hungar. **30** (1995) 155-163.