The expected size of the convex hull of random points in a convex is increasing

Barbados gang

Saturday February 5th

Abstract

Keywords: extreme points,

1 Introduction

Given a set of n points evenly distributed in a domain D, studying $f(n)$ the expected number of extreme points (convex hull vertices) is a natural question. It is well known that $f(n) = \Theta(k \log n)$ if $\mathcal{D}$ is a convex k-gon in the plane and $\Theta\left(n^{\frac{d-1}{d+1}}\right)$ if it is a ball in $\mathbb{R}^d$ [3, 4]. Although the global behavior of $f(n)$ is precisely known, its local behavior is more uncertain. We prove in this paper that $f(n)$ is actually increasing when n is large enough and D is a smooth comvex, although this property seems quite natural, its proof require some technicity. We also remark that the some hypotheses on D are needed, since we have examples of non-smooth non-convex domain where $f(n)$ is not an increasing function.

Previous results

If D is a convex domain with boundary 2 times differentiable in the plane an equivalent of $f(n)$ is known.

$$f(n) \sim ?????n^{\frac{1}{3}},$$

where????

If D is a planar convex $k-$gon even more on the asymptotic behavior is known

$$f(n) = ???klogn + ??? + o(1),$$

where ????

In higher dimension???

$$f(n) = \Theta\left(n^{\frac{d-1}{d+1}}\right)$$

Contribution

Notations

S is set of n points in $D \subset \mathbb{R}^2$.

$f(S)$ is the number of vertices of $CH(S)$. Since we are in 2D, it is also the expected number of edges. $f_1(S)$ is the number of oriented 1-sets defined by S; $(p,q,r) \in S^3$ is an oriented one set if line pq separate r from $S \setminus \{p,q,r\}$, r being on its right. Edges of the convex hull are also called 0-sets.

$f(n)$ is the expected value of $f(S)$ when S is a random subset of D. Similarly $f_1(n)$ is the expected value of $f_1(S)$.

$\mathbb{1}_X$ is the characteristic function of event X, e.g. $\mathbb{1}_{p \in CH(S)}$ is 1 if $p \in CH(S)$ and 0 otherwise.

2 Convex hull size does not increase much

We first remark that $f(n)$ cannot increase too much, whatever the domain D is. When we remove a random point p, we have that $f(S) \leq f(S \setminus \{p\}) + 1$, summing over all p we get:

$$nf(n) \leq nf(n-1) + f(n)$$

thus we get that $\frac{f(n)}{n}$ is decreasing.

3 Convex hull and 1-sets

3.1 Removing one point

First, we observe what happen when one point is removed. A 0-set of S is either a 0-set of $S \setminus \{p\}$ or one of the two edges of the convex hull incident to p if p is on the convex hull. A 0-set of $S \setminus \{p\}$ which is not a 0-set of S is a 1-set *hidden by p*. Thus we get:

$$f(S) = f(S \setminus \{p\}) + 2\ \mathbb{1}_{p \text{ vertex of } CH(S)} - \sharp 1 - \text{sets hidden by } p,$$

summing over p and averaging on the choice of S yields

$$nf(n) = nf(n-1) + 2f(n) - f_1(n),$$

thus we get that f is increasing if f_1 is two times smaller than f:

$$f(n) \nearrow \Longleftrightarrow f_1(n) \leq 2f(n).$$

3.2 Random sampling

Following Clarkson and Shor[1] we take R a random sample of size r of S and look at the probability that a 0-set of S survives in R and that a 1-set of S becomes a 0-set in R. Bounding the size of $CH(R)$ by such sets we get:

$$\frac{\binom{n-2}{r-2}}{\binom{n}{r}} f(S) + \frac{\binom{n-3}{r-2}}{\binom{n}{r}} f_1(S) \quad \leq \quad f(R)$$

Then averaging on S (notice that a random sample of size r of a set S of n random points is just a set of r random points)

$$\begin{aligned}
\frac{\binom{n-2}{r-2}}{\binom{n}{r}} f(n) + \frac{\binom{n-3}{r-2}}{\binom{n}{r}} f_1(n) &\leq f(r) \\
\frac{r(r-1)}{n(n-1)} f(n) + \frac{r(r-1)(n-r)}{n(n-1)(n-2)} f_1(n) &\leq f(r) \\
\frac{r(r-1)}{n(n-1)} f(n) + \frac{r(r-1)}{n(n-1)} \left(1 - \frac{r-2}{n-2}\right) f_1(n) &\leq f(r) \\
p^2 f(n) + p^2(1-p) f_1(n) &\leq f(r)
\end{aligned}$$

with $p = \frac{r-1}{n-1}$.

By Renyi and Sulanke result [4], we have

$$An^{\frac{1}{3}} \leq f(n) \leq A(1+\epsilon)n^{\frac{1}{3}}$$

for $A =$??? any ϵ positive and n big enough. Then

$$\begin{aligned}
p^2 f(n) + p^2(1-p) f_1(n) \leq f(np) &\leq A(1+\epsilon)p^{\frac{1}{3}} n^{\frac{1}{3}} \\
p^2(1-p) f_1(n) &\leq An^{\frac{1}{3}} \left(p^{\frac{1}{3}}(1+\epsilon)\right) - p^2 f(n) \\
&\leq f(n) \left(p^{\frac{1}{3}}(1+\epsilon) - p^2\right) \\
f_1(n) &\leq f(n) \frac{p^{\frac{1}{3}}(1+\epsilon) - p^2}{p^2(1-p)}
\end{aligned}$$

3.3 Wrapping up

Combining these two facts we get that $f(n)$ is increasing if there exists values of ϵ and p such that:

$$\frac{p^{\frac{1}{3}}(1+\epsilon) - p^2}{p^2(1-p)} \leq 2,$$

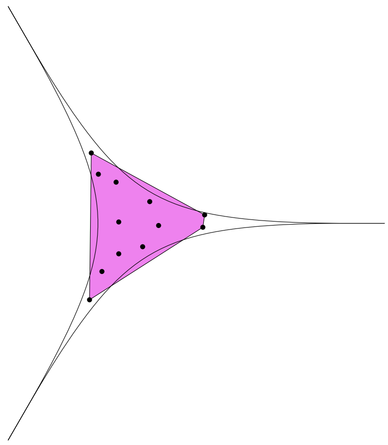


Figure 1: $f(n)$ goes to 3.

using $\epsilon = 0.01$ and $p = 0.93$ gives 1.998.

4 Convex hulls and 2-sets

The above approach does not work with square because the different asymptotic behavior yields to a different function of ϵ and p which is never smaller than 2.

4.1 Removing two points

We get

$$\begin{aligned} f_0(S) &= f_0(S \setminus \{p, q\}) - \sharp\text{1-sets}_{S\setminus\{p\}}(q) - \sharp\text{1-sets}_{S\setminus\{q\}}(p) + \sharp\text{2-sets}_S(q) \\ &\quad +2 \cdot \mathbb{1}_{p \text{ vertex of } CH(S)} + 2 \cdot \mathbb{1}_{q \text{ vertex of } CH(S)} - \mathbb{1}_{pq \text{ edge of } CH(S)} \end{aligned}$$

where $\sharp\text{1-sets}_S(p)$ is the number of 1-sets in S that separates p. Summing over $p \neq q \in S$ yields:

$$\begin{aligned} n(n-1)f_0(S) &= f_0(S \setminus \{p, q\}) - \sharp\text{1-sets}_{S\setminus\{p\}}(q) - \sharp\text{1-sets}_{S\setminus\{q\}}(p) + \sharp\text{2-sets}_S(q) \\ &\quad +2 \cdot \mathbb{1}_{p \text{ vertex of } CH(S)} + 2 \cdot \mathbb{1}_{q \text{ vertex of } CH(S)} - \mathbb{1}_{pq \text{ edge of } CH(S)} \end{aligned}$$

5 It does not increase at cusps

Given a real A, let S be a set a points containing n random points in the domain defined by $0 \leq x \leq A$ and $|y| \leq \lambda(x) = \frac{5(A-x)^4}{2x^5}$, and the points $T = (0, \lambda(0))$ and $B = (0, -\lambda(0))$.

We have $\Lambda(x) = \int_0^t 2\lambda(x)dx = 1 - \frac{(A-t)^5}{t^5}$ is the area of the domain to the left of line $x = t$.

Then the probability that there is only 1 point in between B and T on the convex hull is minorized by $n\text{Prob}(x_0 \in [t, t+dt], \forall i \neq 0 x_i < t-h) \geq n \int_{x;h\geq 0}^{A} \lambda(t)\Lambda(t-h)^{n-1}dt$

where h depends on t and guarantee that x_0 is the only point on the convex hull. We have $\frac{\lambda(0)+\lambda(t)}{t} = \frac{\lambda(0)-\lambda(t-h)}{t-h}$ that is $h = \frac{\lambda(t)(t-h)+\lambda(t-h)t}{\lambda(0)} \sim 2\frac{t\lambda(t)}{\lambda(0)}$.

Then numeric computations show that the probability goes to 1 when n goes to infinty. Since the number of CH points between B and T cannot be smaller than 1, it cannot be an increasing function.

6 The french way

Let $f(n)$ be the expected number of extreme points amongst n random points in domain D. Assume wlog $area(D) = 1$.

$$\begin{aligned} f(n) &= n(n-1)\text{Prob}(X_0X_1 \text{ is an ccw edge of the convex hull}) \\ &= n(n-1)\int_{x\in D^2} g(x)^{n-2}dx \end{aligned}$$

where $g(x)$ is the area of the intersection of the half plane to the left of line x_0x_1 and D.

$$\begin{aligned} f(n+1)-f(n) &= \int_{x\in D^2} g(x)^{n-2}\left(n(n+1)g(x) - n(n-1)\right)dx \\ &= n(n-1)\int_{t=0}^{1} t^{n-2}\left(\frac{n+1}{n-1}t - 1\right)\text{Prob}\left(g(x)\in[t,t+dt]\right) \\ &= n(n-1)\int_{t=\frac{n-1}{n+1}}^{1} t^{n-2}\left(\frac{n+1}{n-1}t - 1\right)\text{Prob}\left(g(x)\in[t,t+dt]\right) \\ & \quad -n(n-1)\int_{t=0}^{\frac{n-1}{n+1}} t^{n-2}\left(1-\frac{n+1}{n-1}t\right)\text{Prob}\left(g(x)\in[t,t+dt]\right) \end{aligned}$$

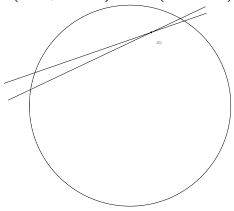
where the both integrals are positive

6.1 Disk

Let D be the disk of area 1.

We have to minorate $\text{Prob}\left(g(x)\in[t,t+dt]\right)$ when $t \geq \frac{n-1}{n+1}$ and to majorate it when $t \leq \frac{n-1}{n+1}$.

The height of lune of area $1-t$ is h such that $h\sqrt(h) = \Theta(1-t)$. Thus if x_0 is at a distance bigger than h of the boundary, then a lune define by x_0 have an area necessarily bigger than $1-t$. When x_0 is choosen, then the place to choose x_1 to induce lune of area between t and $t+dt$ has area bigger than $\Theta(dt)$ if x_0 is not in the middle of the chord which is granted if $dist(x_0, \partial D) > (1-t)^{\frac{2}{3}}$ (with coefficient ???).



$$\begin{aligned}
\text{Prob}\left(g(x) \in [t, t+dt]\right) &\geq \int_{\lambda=0}^{(1-t)^{\frac{2}{3}}} dt\text{Prob}(dist(x_0, \partial D) \in [\lambda, \lambda + d\lambda]) \\
&= (1-t)^{\frac{2}{3}} dt
\end{aligned}$$

then

$$\begin{aligned}
&\int_{t=\frac{n-1}{n+1}}^{1} t^{n-2}\left(\frac{n+1}{n-1}t - 1\right) \text{Prob}\left(g(x) \in [t, t+dt]\right) \\
&\geq \int_{t=\frac{n}{n+1}}^{1} t^{n-2}\left(\frac{n+1}{n-1}t - 1\right)(1-t)^{\frac{2}{3}} dt \\
&\geq \int_{t=\frac{n}{n+1}}^{1} (1-\frac{1}{n})^n \frac{1}{n}(1-t)^{\frac{2}{3}} dt \\
&\geq \frac{1}{ne}\int_{t=\frac{n}{n+1}}^{1} (1-t)^{\frac{2}{3}} dt \\
&\geq \frac{1}{ne}\frac{5}{3}\frac{1}{n^{\frac{5}{3}}} = n^{-\frac{8}{3}}
\end{aligned}$$

The other part has to be majorated

$$\begin{aligned}
\text{Prob}\left(g(x) \in [t, t+dt]\right) &\leq \int_{\lambda=0}^{(1-t)^{\frac{2}{3}}} dt\text{Prob}(dist(x_0, \partial D) \in [\lambda, \lambda + d\lambda]) \\
&\leq ????
\end{aligned}$$

$$
\begin{aligned}
&\int_{t=0}^{\frac{n-1}{n+1}} t^{n-2}\left(1-\frac{n+1}{n-1}t\right)\operatorname{Prob}\left(g(x)\in[t,t+dt]\right)\\
&\leq \int_{t=0}^{\frac{n-1}{n+1}} t^{n-2}\left(1-\frac{n+1}{n-1}t\right)(1-t)^{\frac{2}{3}}dt\\
&\leq n^{-\frac{8}{3}}
\end{aligned}
$$

References

[1] K. L. Clarkson and P. W. Shor. Applications of random sampling in computational geometry, II. *Discrete Comput. Geom.*, 4:387–421, 1989.

[2] Sariel Har-Peled. On the expected complexity of random convex hulls, 1997. http://valis.cs.uiuc.edu/~sariel/research/papers/notes/rand_hull.html, http://valis.cs.uiuc.edu/~sariel/research/papers/notes/rand_hull.html.

[3] H. Raynaud. Sur l'enveloppe convexe des nuages de points aléatoires dans $\mathbb{R}^n$. *J. Appl. Probab.*, 7:35–48, 1970.

[4] A. Rényi and R. Sulanke. Über die konvexe Hülle von n zufällig gerwähten Punkten I. *Z. Wahrsch. Verw. Gebiete*, 2:75–84, 1963.