Let $a_1, \ldots, a_n$ be independent random gaussian points in the plane with variance 1. Today lecture is devoted to the following question: what is the expected size of their convex hull?

**Theorem 1 (Renyi-Salanke).**

$$E_{a_1, \ldots, a_n} [\text{size of C.H.}] = \Theta(\sqrt{\log n}).$$

In this lecture we will prove a weaker bound $O(\log^2 n)$. First we notice, that

$$\Pr[\bar{o} \not\in \text{C.H.}] \leq \left(\frac{3}{4}\right)^{\frac{n}{3}}.$$

This is because the probability that $\bar{o} \not\in \text{C.H.}$ of three points is exactly $3/4$, so we can divide all points into $n/3$ groups of 3, and each group covers $o$ with probability $1/4$. Thus, with exponentially high probability $\bar{o} \in \text{C.H.}$ so we can assume for the rest of the lecture that this is always the case.

For a vector $z$ consider the edge $(a_j, a_k)$ of the convex hull that crosses $z$ clockwise. Denote

$$P_z(\epsilon) = \Pr[\text{ang}(z\bar{o}a_k) < \epsilon].$$

Then clearly

$$E[\text{size of C.H.}] \leq \lim_{\epsilon \to 0} \frac{2\pi}{\epsilon} P_z(\epsilon) + n \left(\frac{3}{4}\right)^{\frac{n}{3}}.$$

Denote by $CH_{j,k}$ the event that $(a_j, a_k)$ is an edge of $CH(a_1, \ldots, a_n)$ and other points lie on the origin side of the line $a_ja_k$. For a fixed vector $z$, $\text{Cross}_{j,k}$ is the event that the edge $(a_j, a_k)$ crosses $z$ clockwise.

$$P_z(\epsilon) = \sum_{j,k} \Pr[CH_{j,k} \land \text{Cross}_{j,k}] \cdot \Pr[\text{ang}(z\bar{o}a_k) < \epsilon | CH_{j,k} \land \text{Cross}_{j,k}].$$
Figure 1:

\[ \Pr \left[ \text{ang}(zak) < \epsilon | CH_{j,k} \land \text{Cross}_{j,k} \right] \]

for any choice of \( j \) and \( k \) (the latter equation follows because of the symmetry). Assume that \((a_j, a_k)\) crosses \( z \). It will be convenient to choose the coordinates \( \theta, t, l, r \) instead of \( a_j, a_k \) (see figure 1). Let \( \alpha = \text{ang}(zak) \). Then the probability \( P_z(\epsilon) \) can be expressed as:

\[
\int_{t, \theta} \left( \int_{l \neq j, k} [CH_{j,k}] \right) \cdot \int_{l, r \geq 0} \left( [\alpha < \epsilon] \right) (l + r) \sin(\theta) \mu(a_j) \mu(a_k) \, d\theta \, dt \, dl \, dr
\]

We need the following claim that estimates the maximal norm of \( n \) gaussian points in the plane.

**Claim 2.**

\[
\Pr \left[ \max_i \|a_i\| > \sqrt{8 \lg n} \right] < \frac{1}{n}
\]

In the assumption of the claim, we can bound \( t \leq \sqrt{8 \lg n}; r, l \leq 2 \sqrt{8 \lg n} \). Once again we can assume that this is always the case (it can change the expectation at most by 1). When \( \alpha \) is sufficiently small,

\[
\alpha > \frac{1}{2} \tan(\alpha) = \frac{1}{2} \cdot \frac{r \sin(\theta)}{t + r \cos(\theta)} \geq \frac{r \sin(\theta)}{6 \sqrt{8 \lg n}}
\]

Thus
We estimate the latter probability using the Combination Lemma from Lecture 19. Namely, we show that

\[ \Pr[r < \epsilon] < O(\sqrt{\lg n} \cdot \epsilon) \]

\[ \Pr[\sin(\theta) < \epsilon] = O(\lg n \cdot \epsilon^2). \]

Thus, the following two lemmas imply the theorem:

**Lemma 1.** \( \forall t \leq \sqrt{8 \lg n} \)

\[
\int_{r \geq 0}^{r \leq t} \frac{|r < \epsilon|(l + r)\mu(a_k)\, dr}{\int_{r \geq 0}^{r < t}(l + r)\mu(a_k)\, dr} \leq O(\sqrt{\lg n} \cdot \epsilon) \quad (1)
\]

**Lemma 2.** \( \forall t \leq \sqrt{8 \lg n}, l, r \leq 2\sqrt{8 \lg n} \)

\[
\int_{\theta}^{|\sin(\theta) \leq \epsilon|} \left( \int_{i \neq j, k} \left| CH_{j,k} \right| \sin(\theta)\mu(a_j)\mu(a_k) \right) \leq O((\lg n \cdot \epsilon^2)^2) \quad (2)
\]

**Proof of Lemma 1.**

Let \( u_\theta \) be a unit vector along \( a_ja_k \) (and assume that \( z \) is a unit vector). Look at

\[ \mu(a_k) = \mu(tz + ru_\theta) = e^{-\frac{d^2}{2}}e^{-\frac{(s+r)^2}{2}}, \]

where \( d \) and \( s \) are defined at figure 2.

**Proposition 1.** For \( s > 1, r \leq \frac{1}{s} \)

\[
\frac{e^{-\frac{r^2}{2}}}{e^{-\frac{(s+r)^2}{2}}} \leq e^2
\]

As a corollary, for \( 0 < r_1 < r_2 < \frac{1}{\sqrt{8 \lg n}} \) holds

\[
\frac{\mu(tz + r_1u_\theta)(l + r_1)}{\mu(tz + r_2u_\theta)(l + r_2)} \leq e^2.
\]
Proposition 2. Let $f$ be s.t. for $0 < x_1 < x_2 < K$ and $\frac{f(x_1)}{f(x_2)} < c$. Then

$$\frac{\int_0^\varepsilon f(x) \, dx}{\int_0^K f(x) \, dx} \leq \frac{ec}{K}.$$  

Proof. One can split the interval $[0, K]$ into $\frac{K}{\varepsilon}$ subintervals of length $\varepsilon$. The integral of $f$ on each subinterval is lower bounded by $c^{-1} \int_0^\varepsilon f(x) \, dx$, thus $\int_0^K f(x) \, dx \geq \frac{K}{\varepsilon} c^{-1} \int_0^\varepsilon f(x)$. □

It is sufficient to choose $K = 1/\sqrt{8 \lg n}$ to finish the proof of Lemma 1. □

Proof of Lemma 2.

Let

$$g(\theta) = \left( \int_a [CH_{j,k}] \prod \mu(a_i) \right) \left( \sin(\theta) \mu(tz + ru_\theta)\mu(tz - lu_\theta) \right)_{g_2(\theta)}$$

In order to estimate the ratio (2) it will be sufficient to confine ourselves to $0 < \theta < \frac{1}{16 \lg n}$ in the denominator. For $0 < \theta_1 < \theta_2 < \frac{\pi}{2}$ holds

$$\frac{g_1(\theta_1)}{g_1(\theta_2)} \leq 1.$$
To see this, notice that $g_1(\theta)$ is the probability of the rest of the points $a_i$ lying on the origin side of the line $a_ja_k$. This probability decreases when the distance of the line $a_ja_k$ to $o$ decreases, thus $g_1(\theta)$ is monotone.

As in Proposition 1, for $t < \sqrt{\frac{1}{\log n}}$, $l, r < 2\sqrt{\frac{1}{\log n}}$, $0 < \theta_1 < \theta_2 < \frac{1}{16 \log n}$ holds

$$\frac{\mu(tz + ru_{\theta_1})}{\mu(tz + ru_{\theta_2})} < e^2,$$

which implies for $0 < \theta_1 < \theta_2 < \frac{1}{16 \log n}$

$$\frac{g_2(\theta_1)}{g_2(\theta_2)} < e^2.$$

Finally, for small values of $\theta$, $\sin(\theta) \sim \theta$, so for $0 < \theta_1 < \theta_2 < \frac{1}{16 \log n}$

$$\frac{g(\theta_1)}{g(\theta_2)} \leq 2e^4 \frac{\theta_1}{\theta_2}.$$

The following fact is the analog of Proposition 2:

**Proposition 3.** If for $x_1 < x_2 \frac{f(x_1)}{f(x_2)} \leq e \frac{x_1}{x_2}$ then

$$\frac{\int_0^\epsilon f(x) \, dx}{\int_0^K f(x) \, dx} \leq 4e \left( \frac{\epsilon}{K} \right)^2$$

It is left to set $K = \frac{1}{16 \log n}$. The lemma is proven.

At the end we justify the change of variables $(a_j, a_k) \rightarrow (l, r, t, \theta)$ that we made in the proof and compute the Jacobian of this transform. Let $a = a_j$ and $b = a_k$ be two points in $\mathbb{R}^2$, specified by four parameters $l, r, h, \theta$ as shown on figure 1. By the straightforward calculation,

$$a_x = l \sin(\theta),$$
$$a_y = t - l \cos(\theta),$$
$$b_x = r \sin(\theta),$$
$$b_y = t + r \cos(\theta).$$
The Jacobi matrix

\[
J = \begin{pmatrix}
\partial a_x & \partial a_y & \partial b_x & \partial b_y \\
0 & 0 & \sin(\theta) & \cos(\theta) \\
\sin(\theta) & -\cos(\theta) & 0 & 0 \\
l \cos(\theta) & l \sin(\theta) & r \cos(\theta) & -r \sin(\theta) \\
0 & 1 & 0 & 1
\end{pmatrix}
\begin{align*}
\frac{\partial r}{\partial r} \\
\frac{\partial l}{\partial l} \\
\frac{\partial \theta}{\partial \theta} \\
\frac{\partial t}{\partial t}
\end{align*}
\]

and the Jacobian

\[|\det J| = (l + r) \sin(\theta),\]

hence

\[da \, db = (l + r) \sin(\theta) \, dr \, dl \, d\theta \, dt.\]