Visibility in Random Forests

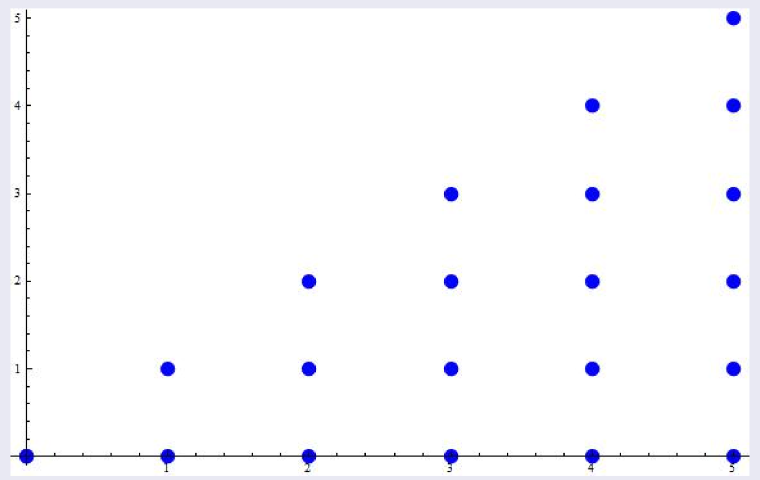
Prof. Jayadev Athreya, Prof. Francesco Cellarosi, Amita Malik, Abigail Turner, Ananya Uppal, Peng Xu

ILLINOIS UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Illinois Geometry Lab

Overview

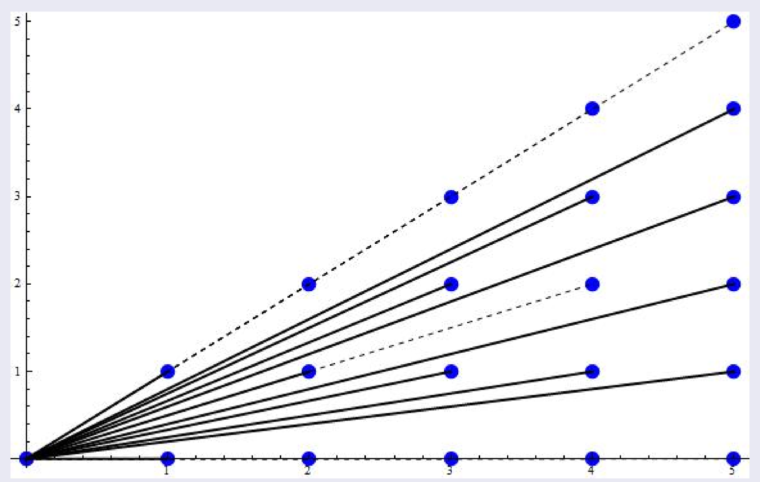
Imagine a tree at every point in the plane with integer coordinates while you stand at the origin.



Only a fraction of the trees in the infinite forest will be visible to you. Surprisingly enough, this fraction is $\frac{6}{\pi^2}$.

Which Trees Are Visible?

- Looking at the above picture, it is evident that we want to say one tree is "behind" another one (i.e., not visible) if there exists another tree that is closer to the origin but has the same slope.
- So, the only trees that are visible to us are the ones that have integer coordinates that are co-prime.



Given that these coordinates are coprime, we can identify each visible tree at coordinate (a, b) with its slope, $\frac{b}{a}$.
The set of all the slopes of visible trees form the set of Farey fractions.

Farey Fractions

The Farey sequence of order n is the set F_n of fractions $\gamma = \frac{a}{q}$ with $1 \leq q \leq n$ in the interval $(0, 1]$, written in lowest terms and arranged in increasing order of magnitude. The elements of this sequence are the Farey fractions.

Considering the Gaps

Assume that only trees within a certain distance of the integer forest are visible to us, i.e. consider only a finite section of this infinite forest. Then,

- We want to look at the distribution of the angular gaps between these trees, i.e. the difference between their slopes.
- By symmetry, we need only consider the trees with slopes less than or equal to 1.

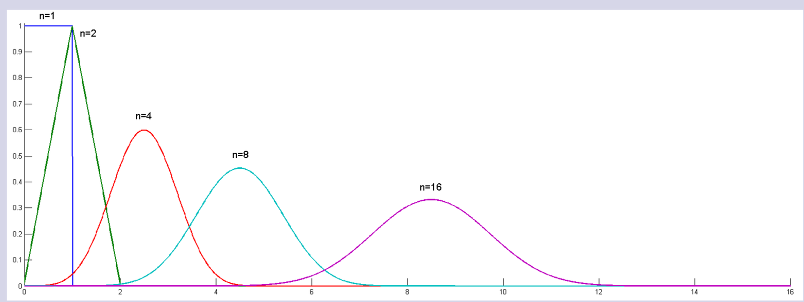
The CLT

- Recall that if $\{X_n\}$ is a sequence of random samples from a distribution that has finite mean μ and variance σ^2, then the random variable

$$\frac{\sum_{k=1}^{n} X_i - n\mu}{\sigma\sqrt{n}}$$

converges in distribution to a random variable that has Standard Normal distribution $\mathcal{N}(0, 1)$.

Uniform Distribution



This demonstrates the convergence of the sum of n Uniform random variables to the Normal distribution as n gets large.

A Generalized CLT

- The CLT does not apply to Hall's distribution due to its infinite variance.
- Let $\{a_n\}, \{b_n\}$ be sequences of constants ($a_n > 0$), and $\{X_n\}$ be a sequence of random variables satisfying

$$\max_{1 \leq i \leq n} P(|X_i| \geq \varepsilon a_n) \to 0 \quad \forall \varepsilon > 0.$$

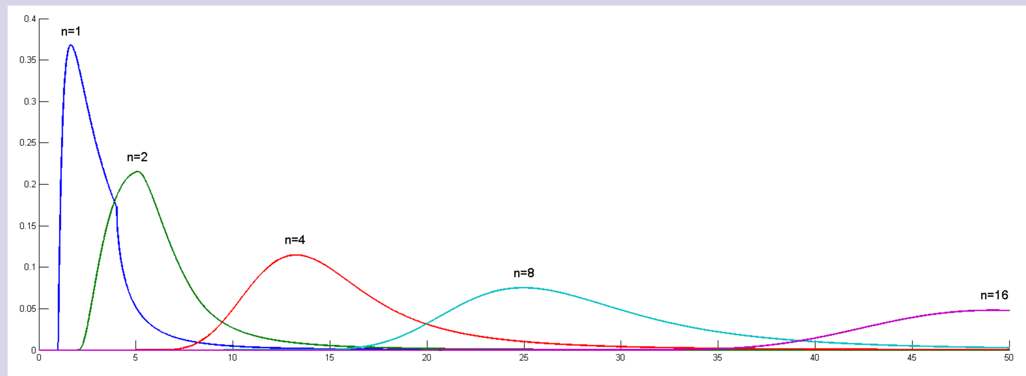
Then the sets of distribution that are limits of distributions of sums

$$\frac{\sum_{k=1}^{n} X_i - b_n}{a_n}$$

of i.i.d. random variables coincides with the set of Stable distribution, whose element has characteristic function in the form

$$\varphi(t) = \exp\left[i\gamma t - c|t|^2\left(1 + i\beta\frac{t}{|t|}\omega(t, \alpha)\right)\right].$$

Hall's Distribution



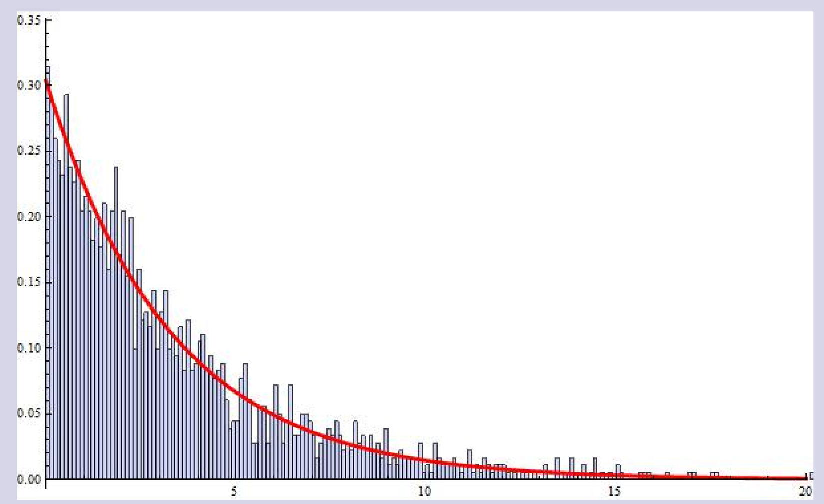
This demonstrates the convergence of the sum of n Hall's random variables to the normal distribution as n gets large.

Killing with Probability $1 - 1/Q$

- Originally, we hypothesized that after "killing" trees with probability $p = 1 - \frac{1}{n}$, the gap distribution would resemble the sum of n Hall's random variables.
- So, with n large we hoped to see a bell curve.
- We looked at the distribution with trees killed with probability $p = 1 - \dfrac{1}{Q}$ as Q gets large (i.e. making the forest sparser).

Exponential Distribution

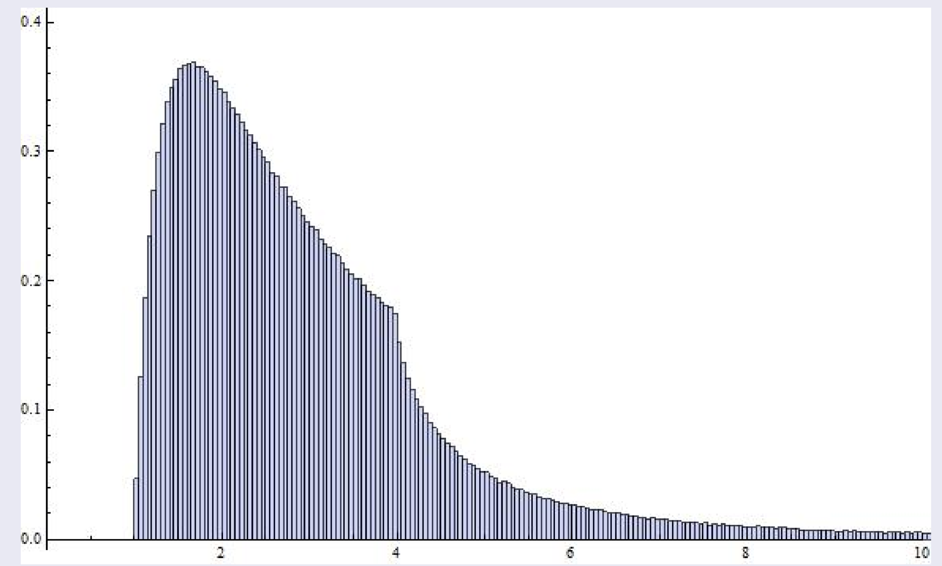
Contrary to our expectation this resembled an exponential distribution with mean $\frac{3}{\pi^2}$.



Note that the mean of the Hall's distribution is $\frac{\pi^2}{3}$.

An Important Distribution

The Hall's distribution, shown below, describes the distribution of gaps between the visible trees i.e. the differences between the Farey fractions.



The Hall's Distribution

Future Directions

- Understand theoretically why this distribution is exponential.
- Study the behaviour of the distributions obtained by merging n consecutive gaps in the gaps between the Farey fractions.

References

- Athreya J. - Gap Distributions and Homogeneous Dynamics, in Proceedings of ICM Satellite Conference on Geometry, Topology, and Dynamics in Negative Curvature.
- Athreya J. and Cheung Y - A Poincare section for horocycle flow on the space of lattices, in International Math Research Notices, 2013.
- Boca F. P. and Zaharescu A. - Farey fractions and two-dimensional tori, in Noncommutative Geometry and Number Theory (C. Consani, M. Marcolli, eds.), Aspects of Mathematics E37, Vieweg Verlag, Wiesbaden, 2006, pp. 57-77.
- Petrov V. V. - Limit Theorems of Probability Theory: Sequences of Independent random variables.
- Hogg R. V., McKean J. W. and Craig A. T. - Introduction to Mathematical Statistics.

This poster was made with support from the Department of Mathematics and the Public Engagement Office at the University of Illinois at Urbana-Champaign