Height-biased trees

joint work with L. Addario-Berry, N. Maitra, and M. Ünel

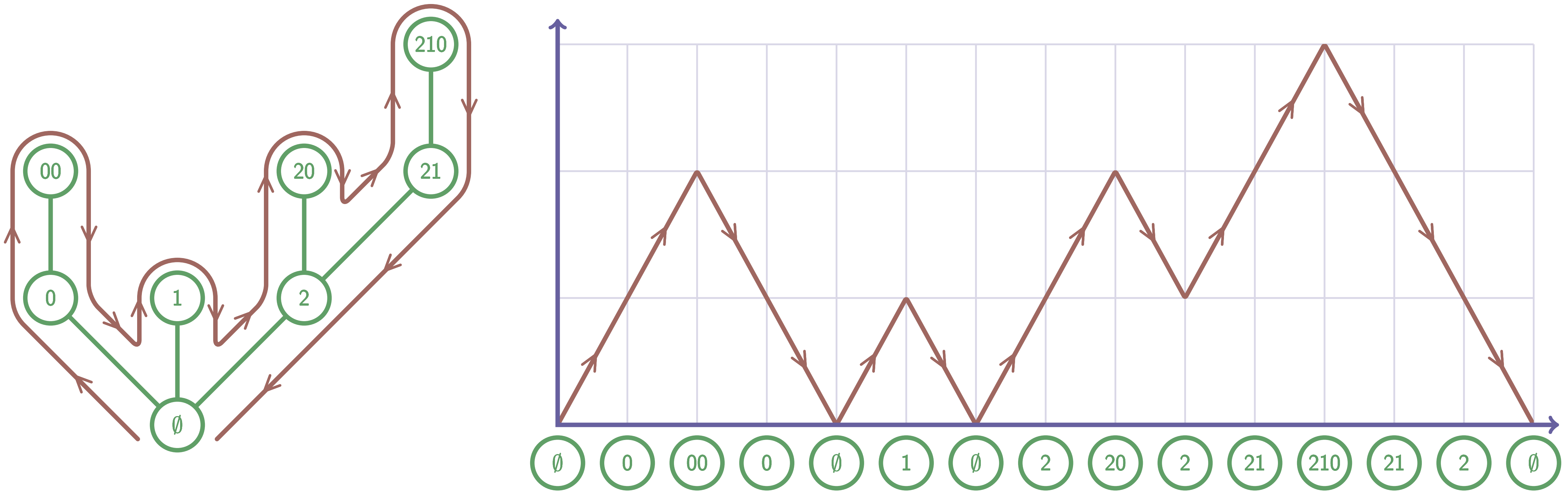
Benoît Corsini

LOUCCOUM Research School

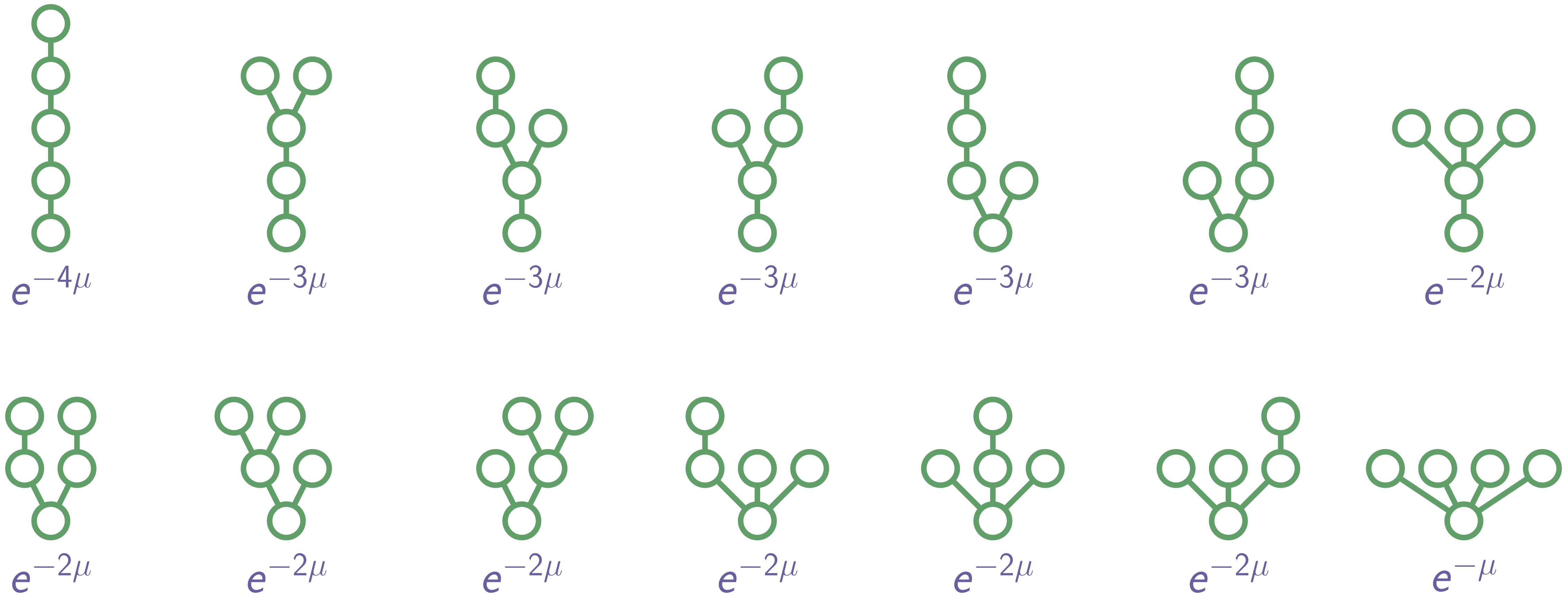
Table of contents

Table of contents

Rooted planar trees



Height-biased trees



Height-biased trees

A *height-biased tree* of size n and parameter μ, denoted $T_{n,\mu}$, is a random tree whose distribution is given by

$$\mathbb{P}\big(T_{n,\mu} = T\big) = \frac{e^{-\mu h(T)}}{Z_{n,\mu}},$$

where $h(T)$ is the height of the tree and

$$Z_{n,\mu} = \sum_{|T|=n} e^{-\mu h(T)}$$

is a normalizing constant (also called the partition function).

→ The first work on these trees considered $\mu \in \mathbb{R}$ arbitrary but fixed as $n \to \infty$.[DU23]

→ For this presentation, we consider the case $\mu = \mu_n \geq 0$.

Main result

Theorem (Addario-Berry, , Maitra, & Ünel, 2025+)

The asymptotic height of a height-biased tree $T_{n,\mu}$ can be summarized by

$$h(T_{n,\mu}) \simeq \left(\frac{2\pi^2 n}{\mu}\right)^{1/3} + O_{\mathbb{P}}\left(\left(\frac{n}{\mu^4}\right)^{1/6}\right) .$$

Table of contents

Interpretation

The height of a height-biased tree $T_{n,\mu}$ satisfies

$$h(T_{n,\mu}) \simeq \left(\frac{2\pi^2 n}{\mu}\right)^{1/3} + O_{\mathbb{P}}\left(\left(\frac{n}{\mu^4}\right)^{1/6}\right) .$$

- When $\mu \to 0$ very quickly, the height seems to explode.
- When $\mu = 0$, the tree behaves like a CRT, with a height of order $\sqrt{n}$.
- The right hand side only applies as long as it is smaller than $\sqrt{n}$.

$\to$ This corresponds to the condition $\mu \gg 1/\sqrt{n}$.

$\to$ This is also exactly the condition for the second term to be less than the first one.

The Brownian regime

$$h(T_{n,\mu}) \simeq \left(\frac{2\pi^2 n}{\mu}\right)^{1/3} + O_{\mathbb{P}}\left(\left(\frac{n}{\mu^4}\right)^{1/6}\right)$$

Theorem (Addario-Berry, 👤, Maitra, & Ünel, 2025)

Assume that $\mu \sim \alpha/\sqrt{n}$ for some $\alpha \in \mathbb{R}$. Then the tree $T_{n,\mu}$ behaves like a α-height-biased Continuum Random Tree.

- In that regime, $h(T_{n,\mu})/\sqrt{n}$ converges to a random variable.
- Considering the case $\alpha \neq 0$, the global formula is still relevant as it becomes

$$h(T_{n,\mu}) \simeq \left(\frac{2\pi^2}{\alpha}\right)^{1/3} \sqrt{n} + O_{\mathbb{P}}\left(\frac{\sqrt{n}}{\alpha^{2/3}}\right) = \Theta_{\mathbb{P}}\left(\sqrt{n}\right) .$$

Interpretation

The height of a height-biased tree $T_{n,\mu}$ satisfies (when $\mu \gg 1/\sqrt{n}$)

$$h(T_{n,\mu}) \simeq \left(\frac{2\pi^2 n}{\mu}\right)^{1/3} + O_{\mathbb{P}}\left(\left(\frac{n}{\mu^4}\right)^{1/6}\right).$$

- When $\mu \ll n^{1/4}$ the two terms diverge; we have a standard central limit theorem.
- When $\mu \sim \alpha n^{1/4}$, the error term is constant; we have a discrete central limit theorem.
- When $\mu \gg n^{1/4}$, the error term goes to 0; we have localized Bernoulli jumps.
  - When $\mu \ll n$, the height still diverges.
  - When $\mu \sim \alpha n$, the height is of constant order.
  - When $\mu \gg n$, the tree is almost-surely a star.

Low-intermediate regime

$$h(T_{n,\mu}) \simeq \left(\frac{2\pi^2 n}{\mu}\right)^{1/3} + O_{\mathbb{P}}\left(\left(\frac{n}{\mu^4}\right)^{1/6}\right)$$

Theorem (Addario-Berry, 👤, Maitra, & Ünel, 2025)
Assume that $1/\sqrt{n} \ll \mu \ll n^{1/4}$. Then the height of $T_{n,\mu}$ satisfies

$$\frac{h(T_{n,\mu}) - (2\pi^2 n/\mu)^{1/3}}{(2\pi^2 n/\mu^4)^{1/6}/\sqrt{3}} \longrightarrow \mathcal{N}(0,1)\,.$$

- This result is the most similar to that of the global formula.

Mid-intermediate regime

$$h(T_{n,\mu}) \simeq \left(\frac{2\pi^2 n}{\mu}\right)^{1/3} + O_{\mathbb{P}}\left(\left(\frac{n}{\mu^4}\right)^{1/6}\right)$$

Theorem (Addario-Berry, 🧍, Maitra, & Ünel, 2025)

Assume that $\mu \sim \alpha n^{1/4}$ for some $\alpha \in (0, \infty)$. Then the random variable $h(T_{n,\mu}) - \lfloor (2\pi^2 n/\mu)^{1/3} \rfloor$ converges to the discrete analogous of a normal random variable.

- The discrete analogous of a continuous random variable is simply obtained by using the probability density function as the probability mass function.
- The exact statement does not exactly use $\lfloor (2\pi^2 n/\mu)^{1/3} \rfloor$ and the parameters of the limit require more definitions to be clearly stated.

High-intermediate regime

$$h(T_{n,\mu}) \simeq \left(\frac{2\pi^2 n}{\mu}\right)^{1/3} + O_{\mathbb{P}}\left(\left(\frac{n}{\mu^4}\right)^{1/6}\right)$$

Theorem (Addario-Berry, 👤, Maitra, & Ünel, 2025)
Assume that $n^{1/4} \ll \mu \ll n$. Then the random variable $h(T_{n,\mu}) - \lfloor (2\pi^2 n/\mu)^{1/3} \rfloor$ converges to a discrete random variable with support of size at most 2.

- In "most cases", the limit is actually deterministic (support of size 1).
- The support has size 2 for specific values of μ.

High regime

$$h(T_{n,\mu}) \simeq \left(\frac{2\pi^2 n}{\mu}\right)^{1/3} + O_{\mathbb{P}}\left(\left(\frac{n}{\mu^4}\right)^{1/6}\right)$$

Theorem (Addario-Berry, , Maitra, & Ünel, 2025)

Let $\mu/n \to \alpha > 0$ and $t_{n,\mu}$ minimizes $t \mapsto t\mu - 2n \log \cos(\pi/t)$. Assume that

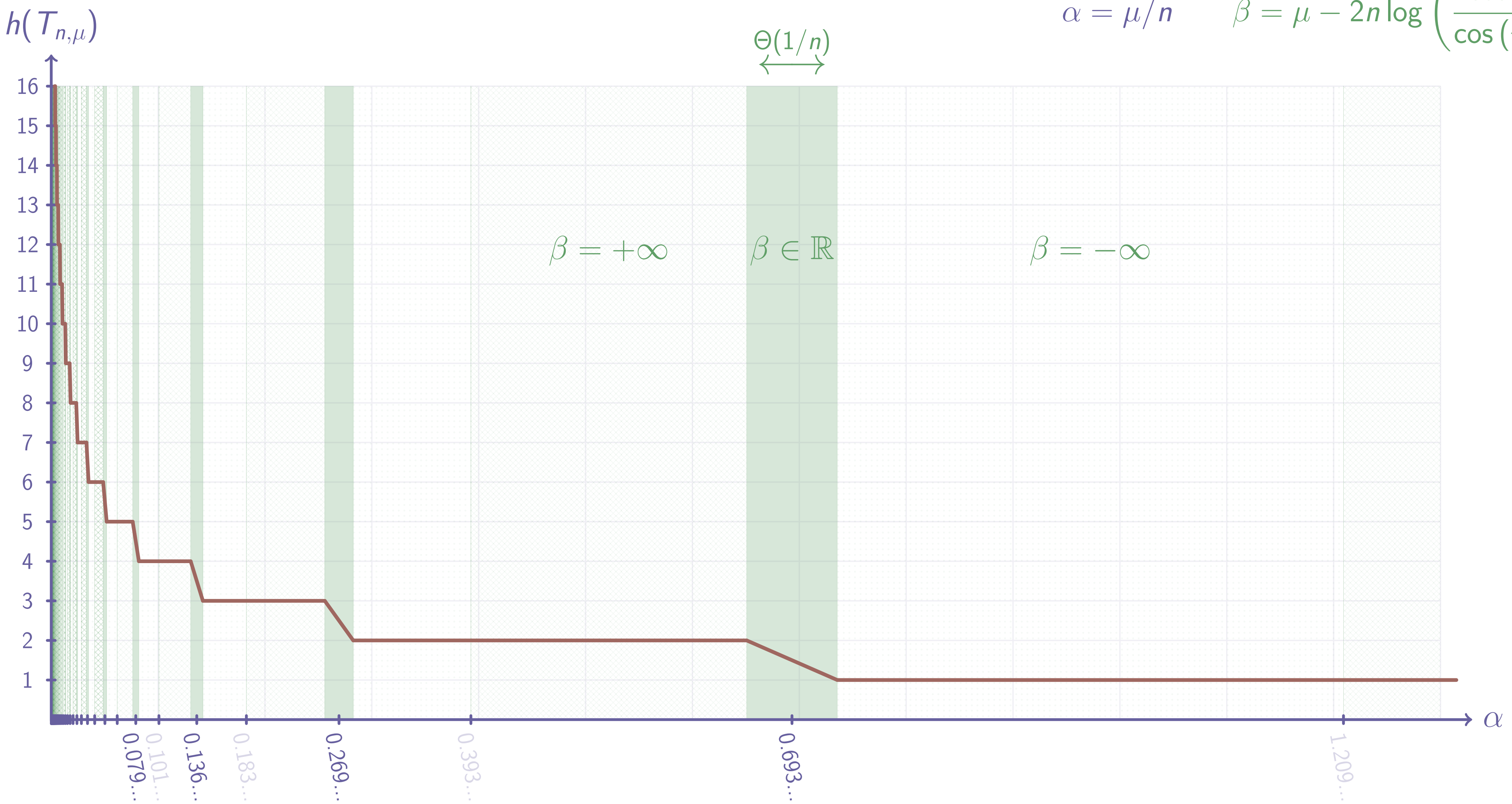
$$\mu - 2n \log\left(\frac{\cos\left(\pi/\lceil t_{n,\mu}\rceil\right)}{\cos\left(\pi/\lfloor t_{n,\mu}\rfloor_{3+}\right)}\right) \to \beta \in [-\infty, \infty] .$$

Then $h(T_{n,\mu}) - \lfloor t_{n,\mu}\rfloor_{3+} + 2$ converges to a Bernoulli variable with parameter $p_{\alpha,\beta}$.

- β is finite only when α belongs to specific set of values.
- $p_{\alpha,-\infty} = 1$ and $p_{\alpha,+\infty} = 0$; in both cases, $h(T_{n,\mu})$ is deterministic.

High regime

$\alpha = \mu/n \qquad \beta = \mu - 2n\log\left(\frac{\cos(\pi/\lceil t_{n,\mu}\rceil)}{\cos(\pi/\lfloor t_{n,\mu}\rfloor_{3+})}\right)$



Highest threshold

Theorem (Addario-Berry, , Maitra, & Ünel, 2025)

Assume that $\mu - n \log 2 \to \beta \in [-\infty, \infty]$. Then the probability that the height of $T_{n,\mu}$ is 1 (meaning that $T_{n,\mu}$ is a star) satisfies

$$\mathbb{P}\big(h(T_{n,\mu}) = 1\big) = \frac{4}{4 + e^{-\beta}}.$$

- As soon as $\mu - n \log 2$ diverges to $+\infty$, the tree is almost-surely the star.

Summary

We have a single expression summarizing all cases:

$$h(T_{n,\mu}) \simeq \left(\frac{2\pi^2 n}{\mu}\right)^{1/3} + O_{\mathbb{P}}\left(\left(\frac{n}{\mu^4}\right)^{1/6}\right).$$

- When $\mu \sim \alpha/\sqrt{n}$, then $T_{n,\mu}$ behaves as a biased CRT.
- When $1/\sqrt{n} \ll \mu \ll n^{1/4}$, then $h(T_{n,\mu})$ satisfies a CLT.
- When $\mu \sim \alpha n^{1/4}$, then $h(T_{n,\mu})$ satisfies a discrete CLT.
- When $n^{1/4} \ll \mu \ll n$, then $h(T_{n,\mu})$ diverges but is mostly deterministic.
- When $\mu \sim \alpha n$, then $h(T_{n,\mu})$ converges to a fixed value, with small window jumps.
- When $\mu \gg n$, then $T_{n,\mu}$ is a star with height 1 (true as soon as $\mu - n \log 2 \to +\infty$).

Table of contents

Height-bounded trees

Consider the number of trees of height less than m:

$$H_{n,m} = \left|\left\{ T : |T| = n \text{ and } h(T) < m \right\}\right| .$$

Whenever $m \geq n$, this is the total number of trees, corresponding to the Catalan number:

$$H_{n,m} = C_{n-1} = \frac{1}{n}\binom{2n-2}{n-1} .$$

When $2 \leq m < n$, we can still find a closed formula:

$$H_{n,m} = \frac{4^n}{m+1} \sum_{k=1}^{\lfloor m/2 \rfloor} \sin^2\left(\frac{\pi k}{m+1}\right) \cos^{2n-2}\left(\frac{\pi k}{m+1}\right) .$$

Height-bounded trees

We can approximate $H_{n,m}$ under some assumptions on n and m using the previous formula:

$$H_{n,m} = \frac{4^n}{m+1} \sum_{k=1}^{\lfloor m/2 \rfloor} \sin^2\left(\frac{\pi k}{m+1}\right) \cos^{2n-2}\left(\frac{\pi k}{m+1}\right) .$$

- When n diverges, the cos term will have more impact than the sin term.
- We can approximate it as $\cos^{2n-2}(\pi k/(m+1)) \simeq e^{-\pi^2 k^2 n/m^2}$.
- When $n/m^2 \to \infty$, we can neglect the terms $k \geq 2$.

$\rightarrow$ Whenever $m \ll \sqrt{n}$, we can approximate $H_{n,m}$ only using the term $k = 1$:

$$H_{n,m} \sim \frac{4^n}{m+1} \sin^2\left(\frac{\pi}{m+1}\right) \cos^{2n-2}\left(\frac{\pi}{m+1}\right) \simeq \frac{4^n \pi^2}{m^3} e^{-\pi^2 n/m^2} .$$

Height-bounded trees and range-bounded walks

The number of trees of height less than m satisfies the formula (for $m \ll \sqrt{n}$)

$$H_{n,m} \simeq \frac{4^n \pi^2}{m^3} e^{-\pi^2 n/m^2} .$$

From this, we see that a uniform tree T_n of size n satisfies (for $m \ll \sqrt{n}$)

$$\mathbb{P}\big(h(T_n) < m\big) \simeq 4\pi^{5/2} \left(\frac{n}{m^2}\right)^{3/2} e^{-\pi^2 n/m^2} .$$

This directly relates to a standard result on the range of a random walk (for $b - a \ll \sqrt{n}$):

$$\frac{(b-a)^2}{n} \log \mathbb{P}\big(a \leq S_k \leq b, \forall k \leq n\big) \longrightarrow -\frac{\pi^2}{2} .$$

Partition function

Let us now try to approximate the partition function using the previous result:

$$Z_{n,\mu} = \sum_{|T|=n} e^{-\mu h(T)} \,.$$

By combining the tree with the same height and re-organizing the terms, this becomes

$$Z_{n,\mu} = (e^{\mu} - 1) \sum_{m=2}^{n-1} H_{n,m} e^{-\mu m} + C_{n-1} e^{-\mu(n-1)} \,.$$

The final term corresponds to the tree of height $n-1$ and can be ignored whenever $\mu \geq 0$:

$$Z_{n,\mu} \sim (e^{\mu} - 1) \sum_{m=2}^{n-1} H_{n,m} e^{-\mu m} \,.$$

Partition function

Recall the formula we had for the number of trees of bounded height ($m \ll \sqrt{n}$):

$$H_{n,m} \simeq \frac{4^n \pi^2}{m^3} e^{-\pi^2 n/m^2} .$$

Using this formula for all m and plugging it into the partition function, we obtain

$$Z_{n,\mu} \sim (e^{\mu} - 1) \sum_{m=2}^{n-1} H_{n,m} e^{-\mu m} \simeq (e^{\mu} - 1) \sum_{m=2}^{n-1} \left(\frac{4^n \pi^2}{m^3} e^{-\pi^2 n/m^2} \right) e^{-\mu m} .$$

Rewriting the formula, we get this new form:

$$Z_{n,\mu} \simeq 4^n \pi^2 (e^{\mu} - 1) \sum_{m=2}^{n-1} \frac{e^{-\pi^2 n/m^2 - \mu m}}{m^3} .$$

Partition function

The partition function can be approximated by

$$Z_{n,\mu} = \sum_{|T|=n} e^{-\mu h(T)} \simeq 4^n \pi^2 (e^{\mu} - 1) \sum_{m=2}^{n-1} \frac{e^{-\pi^2 n/m^2 - \mu m}}{m^3}\,.$$

Ignoring the $1/m^3$ term, the maximum in the sum is obtained for m satisfying

$$\frac{2\pi^2 n}{m^3} - \mu = 0 \iff m = \left(\frac{2\pi^2 n}{\mu}\right)^{1/3}$$

If we further let $m = (2\pi^2 n/\mu)^{1/3} + x(n/\mu^4)^{1/6} = (2\pi^2 n/\mu)^{1/3}[1 + x/(n\mu^2)^{1/6}]$, we obtain

$$\frac{\pi^2 n}{m^2} + \mu m = 3\left(\frac{\pi^2 \mu^2 n}{\mu}\right)^{1/3} + 3\left(\frac{\pi}{2}\right)^{2/3} x^2\,.$$

Partition function

In the partition function, the only trees that contribute asymptotically satisfy

$$h(T) \simeq \left(\frac{2\pi^2 n}{\mu}\right)^{1/3} + x\left(\frac{n}{\mu^4}\right)^{1/6}.$$

Theorem (Addario-Berry, , Maitra, & Ünel, 2025)

The asymptotic height of a height-biased tree $T_{n,\mu}$ can be summarized by

$$h(T_{n,\mu}) \simeq \left(\frac{2\pi^2 n}{\mu}\right)^{1/3} + O_{\mathbb{P}}\left(\left(\frac{n}{\mu^4}\right)^{1/6}\right).$$

Table of contents

Other results

Structure around the root.

- When $\mu = O(1)$, the tree admits a local limit (varies whether $\mu \ll 1$ or not).
- When $1 \ll \mu \ll n^{1/4}$, the root degree satisfies a CLT.

Width of the tree.

- When $\mu \sim \alpha/\sqrt{n}$ for some $\alpha \in \mathbb{R}$, the width is related to the biased CRT.
- When $\mu \gg (\log n)^{3/2}/\sqrt{n}$, the width is of order $(\mu n^2)^{1/3}$ (without exact constant).

Q: What happens outside of the covered ranges?

Q: Can we identify the constant for the width?

Relation to random walks

Our results apply to trees biased by their height.

- Thanks to the contour walk, this directly applies to bridges biased by their range.
- A similar result studied the case of (unconstrained) walks biased by their range[B24].
  - When $(\log n)^{3/2}/\sqrt{n} \ll \mu \ll n^{1/4}$, the range satisfies a CLT.
  - When $\mu \sim \alpha n^{1/4}$ for some $\alpha > 0$, the range satisfies a discrete CLT.

→ The condition $\mu \gg (\log n)^{3/2}/\sqrt{n}$ appears again here.
→ This is the threshold when $e^{-\mu h(T)}$ with $h(T) \simeq (n/\mu)^{1/3}$ becomes polynomial.

Q: Can we connect these results?

Negative parameter (positive bias)

In the original work, the authors consider $\mu \geq 0$ as well as $\mu < 0$ (but fixed).

- Due to the difficulty already arising when $\mu = \mu_n \geq 0$, we started with this case.
- The methods to study $\mu > 0$ and $\mu < 0$ are very different.
- Both models (tree and walk) when $\mu < 0$ has rarely been studied.

Q: Can we extend the original results when $\mu < 0$ to $\mu = \mu_n < 0$?

Q: What happens for the random walk when $\mu < 0$?

References

- L. Addario-Berry, B. Corsini, N. Maitra, & M. Ünel (2025). **On exponentially height-penalized random trees**. *arXiv:2512.17747.*.[AB+25]
- N. Bouchot (2024). **Scaling limits for the random walk penalized by its range in dimension one**. *ALEA, 21, 791-813.*[B24]
- B. Durhuus, & M. Ünel (2023). **Trees with exponential height dependent weight**. *Probability Theory and Related Fields, 186(3), 999-1043.*.[DU23]

Thank you! Thank you! Thank you!

Thank you! Thank you! Thank you!