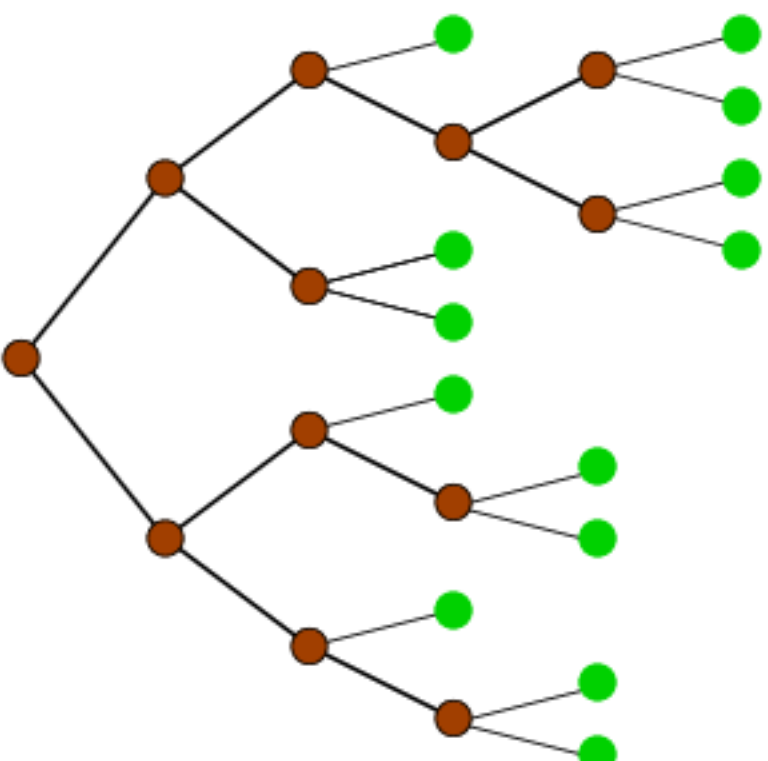


# On the shape of binary trees

Mireille Bousquet-Mélou, CNRS, LABRI, Bordeaux



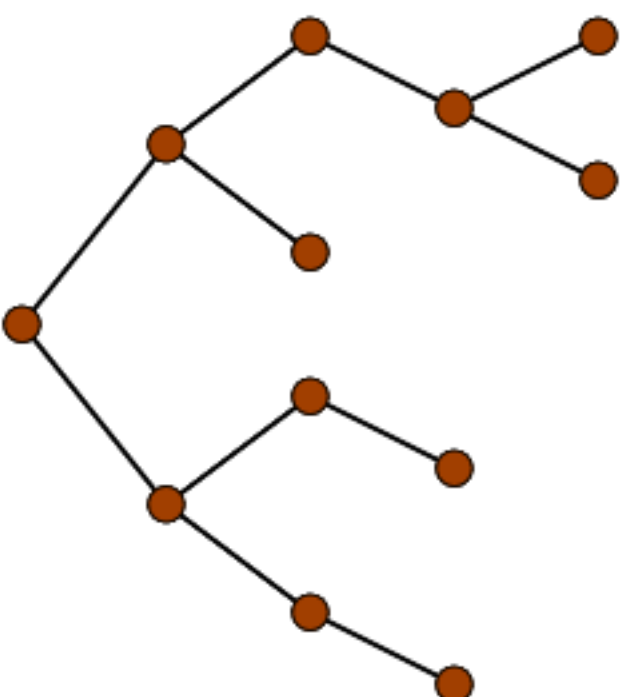
## A (complete) binary tree



$n$  internal vertices, called nodes (size  $n$ )

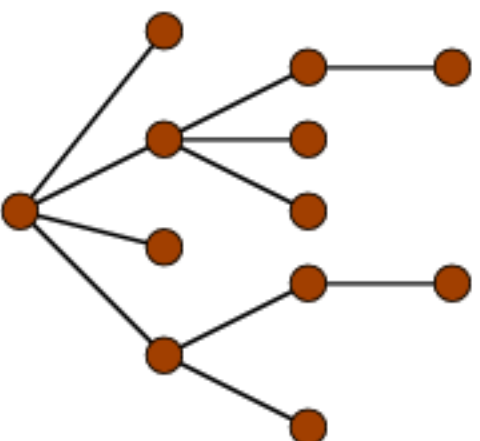
$n + 1$  external vertices, called leaves

## An incomplete binary tree



$n$  nodes (size  $n$ )

## A plane tree

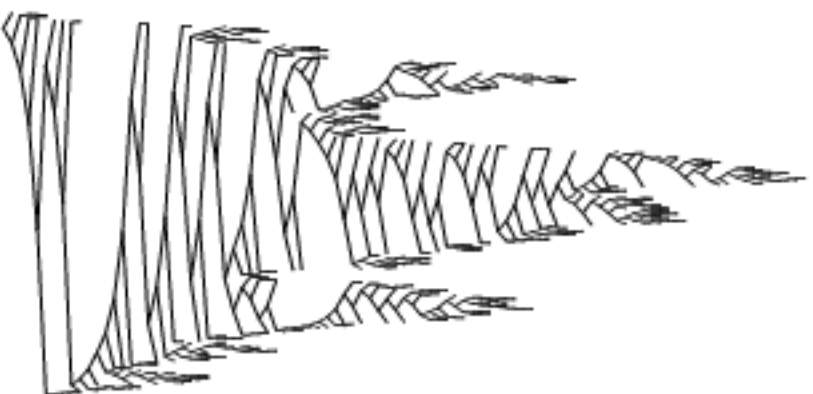


Each node has a (possibly empty) **ordered sequence of children**

$n$  edges,  $n + 1$  nodes (size  $n$ )

**What is the shape of a *large* binary tree?**

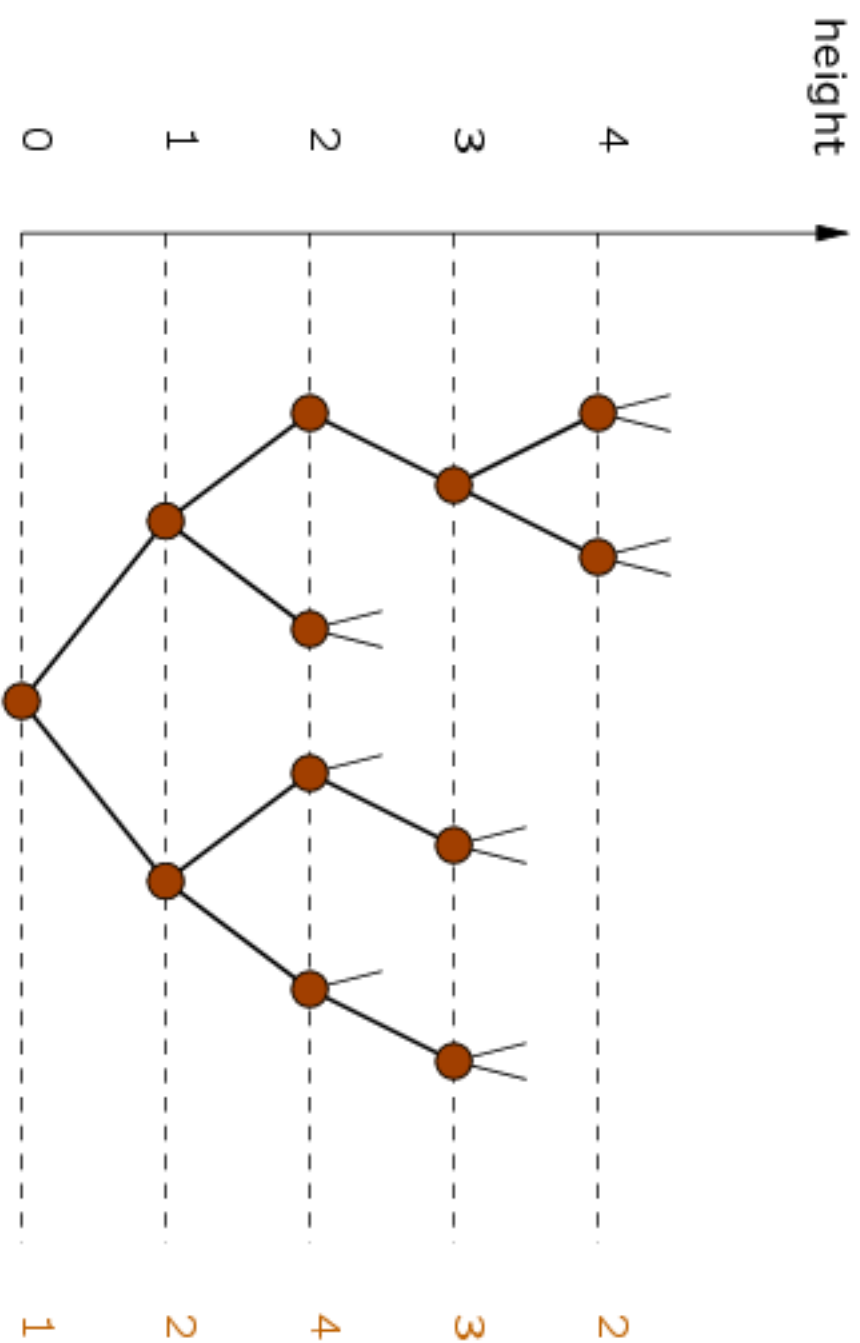
**What is the shape of a *large* binary tree?**



A random complete binary tree with 985 vertices

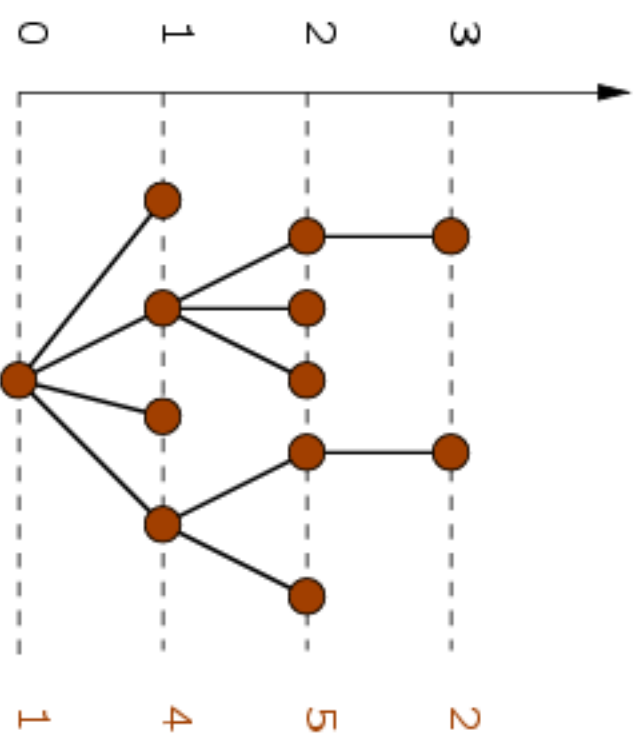
Random generation in linear time

## The (horizontal) profile of a binary tree



A binary tree of height 4 (or 5...)

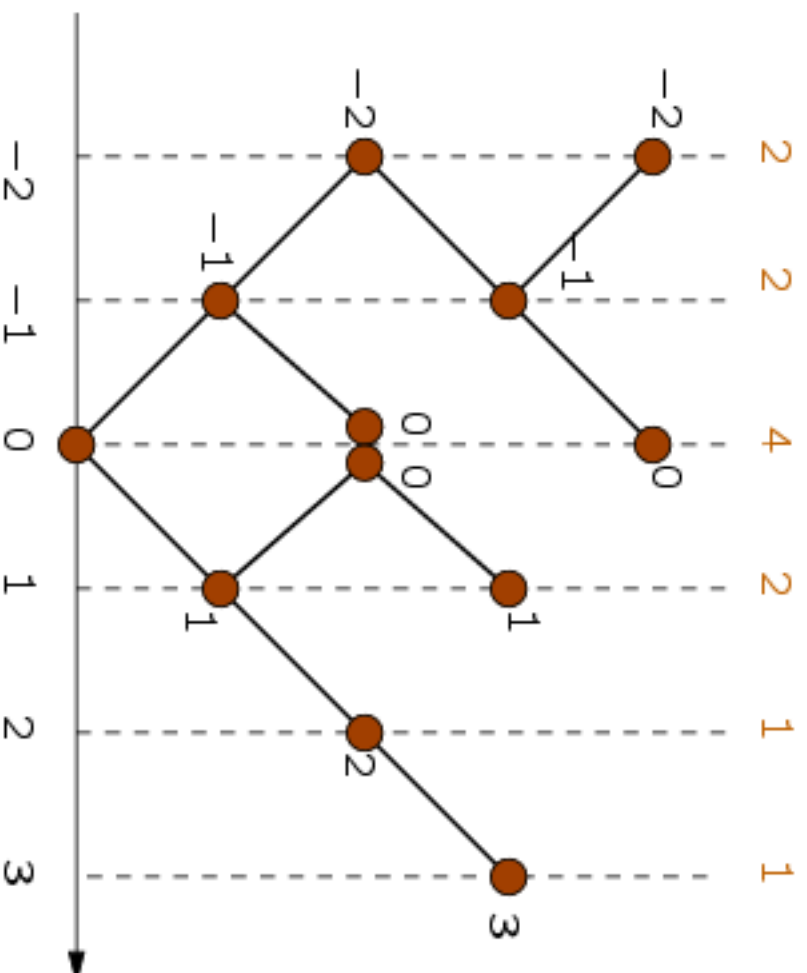
## The horizontal profile of a plane tree



A plane tree of height 3.



## NEW! The vertical profile of a binary tree



LABEL  
=  
ABSCISSA

A binary tree of **right width 3**,  
and vertical profile **[2, 2; 4, 2, 1, 1]**.

## Limit results on the shape of trees: general approach

- **Enumerative combinatorics** (decompositions of trees, recurrence relations, generating functions...)
- **Singularity analysis** [Flajolet-Odlyzko 90]: extract from the generating functions the asymptotic behaviour of their coefficients

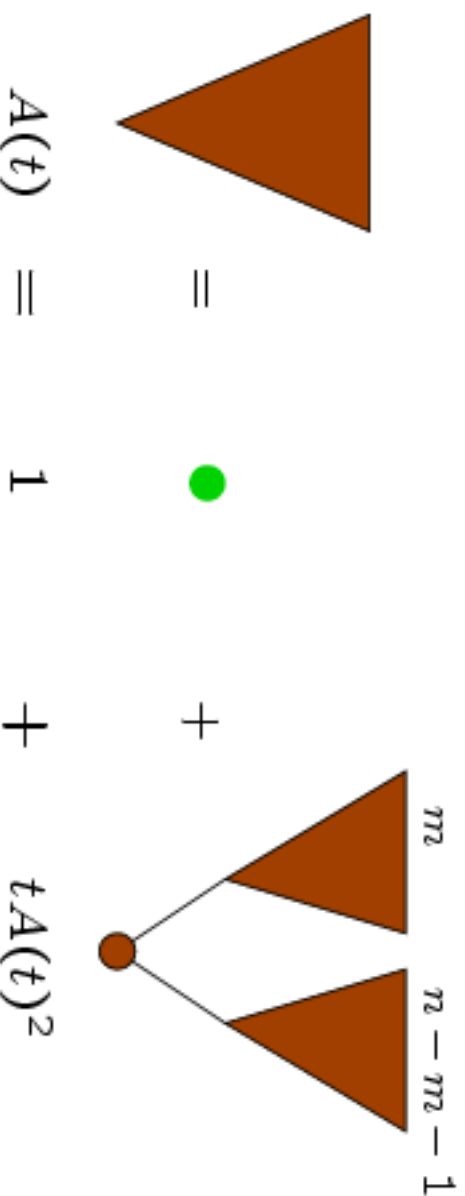
## Decomposition and enumeration of binary trees

Let  $a_n$  be the number of binary trees with  $n$  nodes (internal vertices):

$$\begin{cases} a_0 = 1 \\ a_n = \sum_{m=0}^{n-1} a_m a_{n-m-1} \end{cases}$$

Let  $A(t) := \sum_{n \geq 0} a_n t^n$  be the associated **generating function**:

$$A(t) = 1 + tA(t)^2.$$



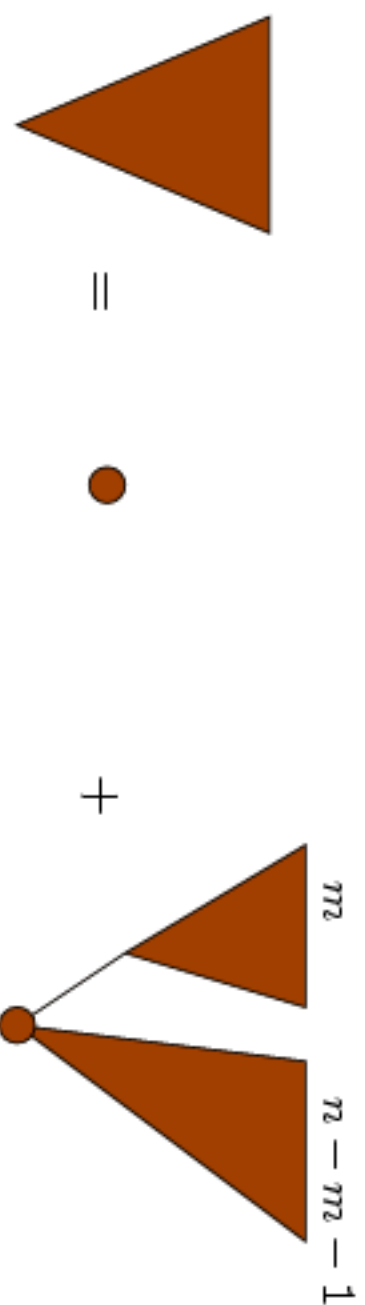
## Decomposition and enumeration of plane trees

Let  $a_n^*$  be the number of plane trees with  $n$  edges ( $n + 1$  vertices):

$$\begin{cases} a_0^* = 1 \\ a_n^* = \sum_{m=0}^{n-1} a_m^* a_{n-m-1}^* \end{cases}$$

Let  $A^*(t) := \sum_{n \geq 0} a_n^* t^n$  be the associated **generating function**:

$$A^*(t) = 1 + tA^*(t)^2.$$



$$A^*(t) = 1 + tA^*(t)^2$$

## The Catalan numbers

There are as many **binary trees** with  $n$  nodes as **plane trees** with  $n$  edges:

$$a_n = a_n^*$$

The associated generating function is

$$A(t) = A^*(t) = \sum_n a_n t^n$$

It satisfies

$$A(t) = 1 + tA(t)^2 \Rightarrow A(t) = \frac{1 - \sqrt{1 - 4t}}{2t} = \sum_{n \geq 0} \frac{1}{n+1} \binom{2n}{n} t^n.$$

Hence

$$a_n = a_n^* = \frac{1}{n+1} \binom{2n}{n}$$

is the  $n$ th Catalan number.

**The height of binary trees: Experiments and observations**

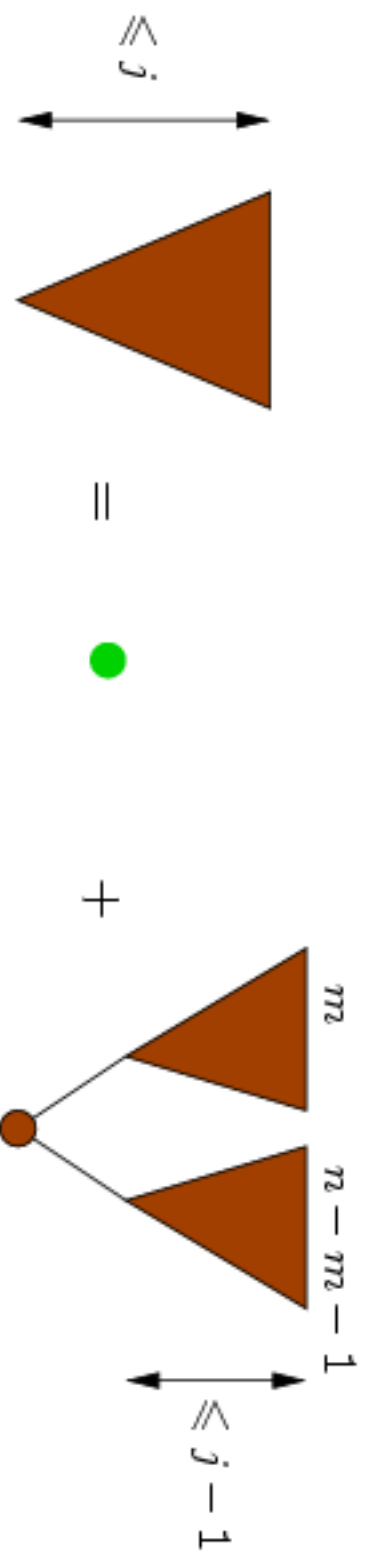
## The height of binary trees: recurrence relations

Let  $h_{n,j}$  be the number of binary trees of size  $n$  and height  $j$ . Let  $h_{n,\leq j}$  be the number of binary trees of size  $n$  and height at most  $j$ . Then

$$h_{n,j} = h_{n,\leq j} - h_{n,\leq j-1}$$

and

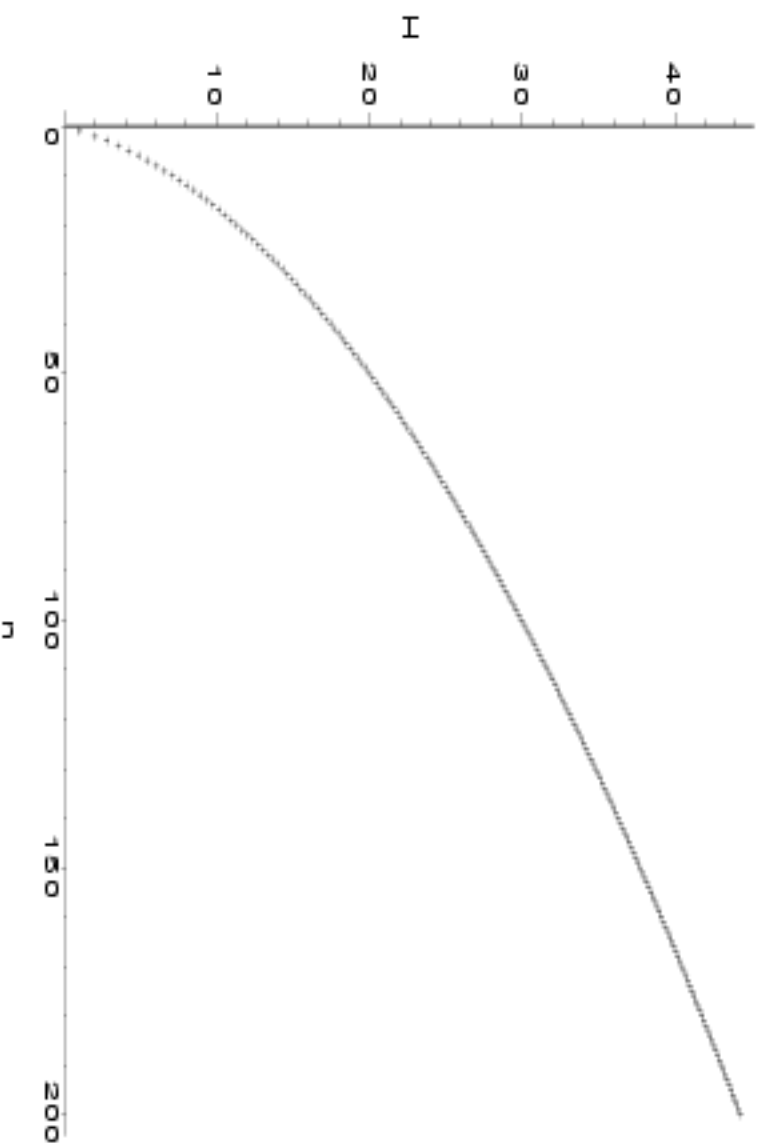
$$\begin{cases} h_{0,\leq j} = 1 \\ h_{n,\leq 0} = \delta_{n,0} \end{cases} = \begin{cases} h_{n,\leq j} = \sum_{m=0}^{n-1} h_{m,\leq j-1} h_{n-m-1,\leq j-1}. \end{cases}$$



## The average height of binary trees

Consider the **average height** of binary trees with  $n$  nodes:

$$\mathbb{E}(H_n) := \frac{1}{a_n} \sum_{|\tau|=n} h(\tau) = \frac{1}{a_n} \sum_{j \geq 0} j h_{n,j}.$$





## The average height of trees

Consider the **average height** of binary trees with  $n$  nodes:

$$\mathbb{E}(H_n) := \frac{1}{a_n} \sum_{|\tau|=n} h(\tau) = \frac{1}{a_n} \sum_{j \geq 0} j h_{n,j}.$$

**Theorem:** As  $n \rightarrow \infty$ ,

$$\frac{\mathbb{E}(H_n)}{\sqrt{n}} \rightarrow 2\sqrt{\pi}.$$

[Flajolet-Odlyzko 82], [Brown-Schubert 84]

Similarly, the average height of plane trees satisfies:

$$\frac{\mathbb{E}(H_n^*)}{\sqrt{n}} \rightarrow \sqrt{\pi}.$$

[de Bruijn, Knuth, Rice 72]

**Convergence of the mean of the (normalized) height**

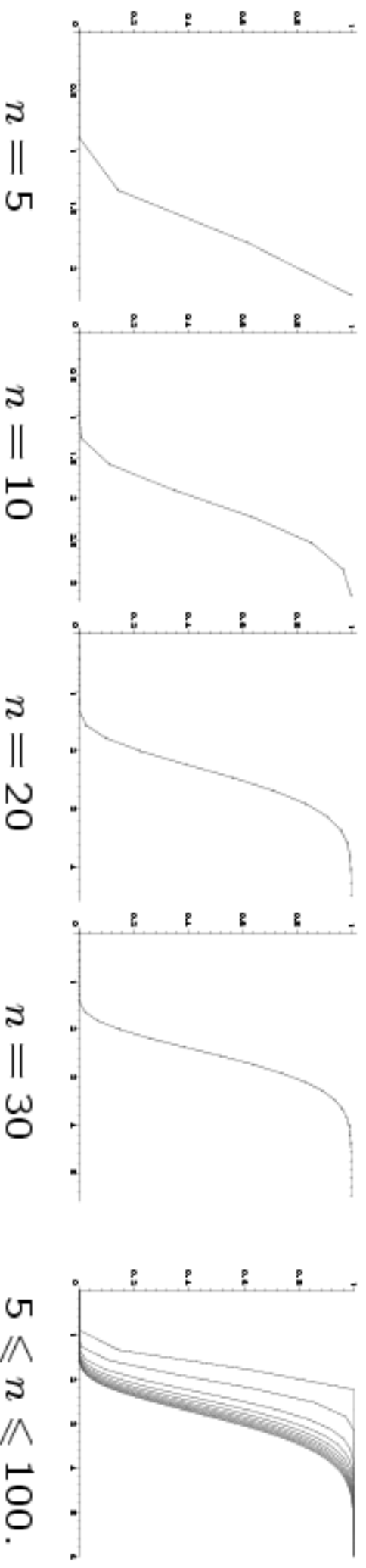
## The distribution function of the height

The number of trees of size  $n$  and height at most  $j$  is  $h_{n, \leq j}$ . For  $x \geq 0$ , consider

$$F_n(x) := \mathbb{P} \left( \frac{H_n}{\sqrt{n}} \leq x \right) = \frac{1}{a_n} h_{n, \leq x\sqrt{n}},$$

the **proportion** of trees of size  $n$  having height at most  $x\sqrt{n}$ . It is the **probability** that a random tree of size  $n$  has height at most  $x\sqrt{n}$ .

Graph of  $F_n(x)$  :



## The distribution function of the height

Let  $F_n(x)$  be the probability that a random tree of size  $n$  has height at most  $x\sqrt{n}$ .

**Theorem:** As  $n \rightarrow \infty$ ,

$$F_n(x) = \mathbb{P}\left(\frac{H_n}{\sqrt{n}} \leq x\right) \rightarrow F(x/2).$$

where

$$F(x) = \sum_{k=-\infty}^{\infty} e^{-k^2 x^2} (1 - 2k^2 x^2).$$

[Kemp 79], [Flajolet-Odlyzko 82]

Similarly, the height of plane trees satisfies:

$$\mathbb{P}\left(\frac{H_n^*}{\sqrt{n}} \leq x\right) \rightarrow F(x).$$

[Flajolet-Odlyzko 82], [Brown-Schubert 84]

**Convergence in law of the (normalized) height**

## Limit results on the shape of trees: general approach

- **Enumerative combinatorics**: Using a recursive decomposition of trees, write equations for the relevant generating functions and **solve** them...

- Use the results (or the technique) of **singularity analysis** [Flajolet-Odlyzko 90] to extract from these series the asymptotic behaviour of their coefficients
- Results: “If  $B(t) = \sum_n b_n t^n$  behaves like this in the neighborhood of its dominant singularity  $t_s$ , then its coefficients  $b_n$  behave asymptotically like that.”

- Technique: Cauchy’s formula

$$b_n = \frac{1}{2i\pi} \int_{C_n} B(t) \frac{dt}{t^{n+1}}$$

for a carefully chosen contour  $C_n$ .

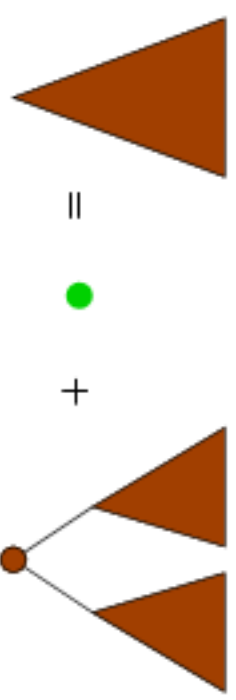
## Generating functions of trees of bounded height

Let  $H_{\leq j}(t) \equiv H_{\leq j}^*$  be the generating function of binary trees of height at most  $j$ :

$$H_{\leq j} = \sum_{n \geq 0} h_{n, \leq j} t^n$$

$$H_{\leq 0} = 1$$

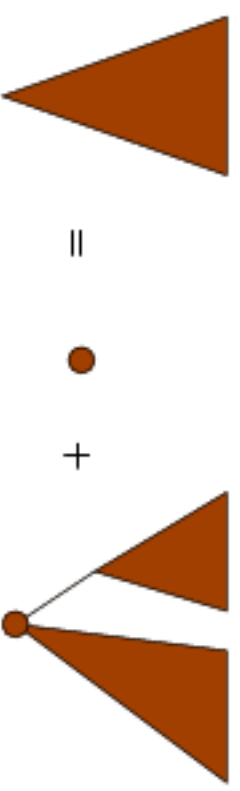
$$H_{\leq j} = 1 + tH_{\leq j-1}^2$$



Let  $H_{\leq j}^*(t) \equiv H_{\leq j}^*$  be the generating function of plane trees of height at most  $j$ , where  $t$  counts edges.

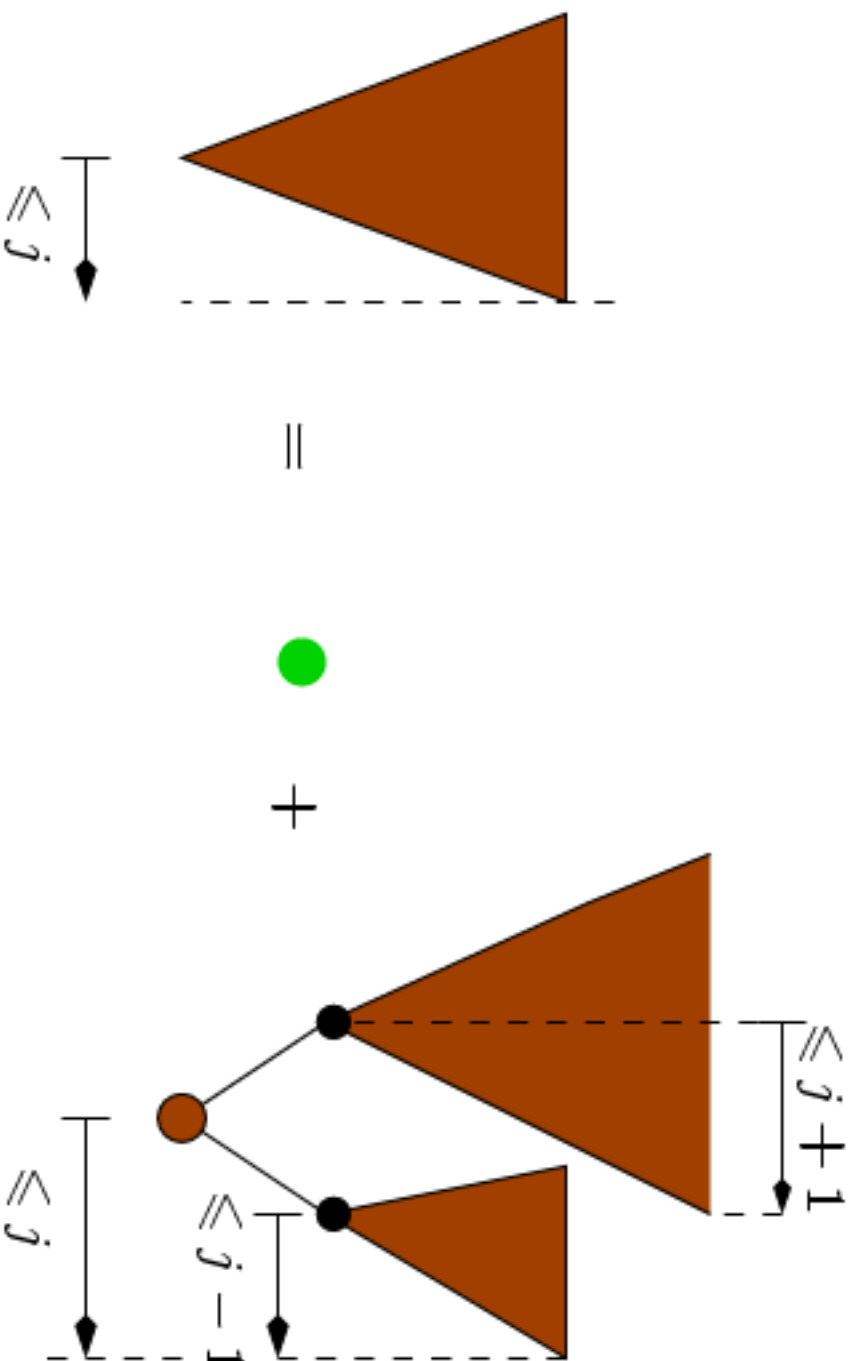
$$H_{\leq 0}^* = 1$$

$$H_{\leq j}^* = 1 + tH_{\leq j-1}^* H_{\leq j}^*$$



## The (right) width of binary trees

Let  $W_{\leq j}(t) \equiv W_{\leq j}$  be the generating function of binary trees of right width at most  $j$ .

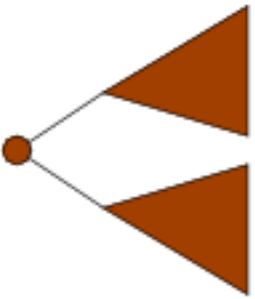
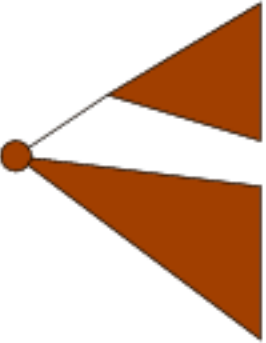


$$W_{\leq -1} = 1$$

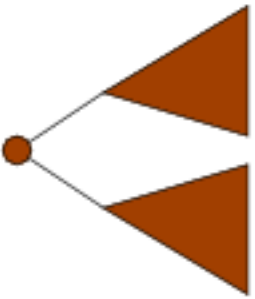
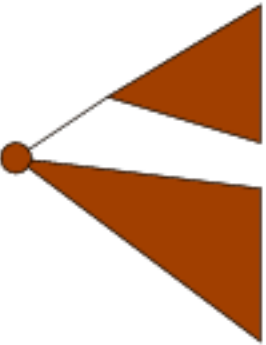
For  $j \geq 0$ ,

$$W_{\leq j} = 1 + tW_{\leq j+1}W_{\leq j-1}$$

## Trees of bounded height or bounded width: functional equations

Family of trees	Bounded height	Bounded (right) width
	$H_{\leq 0} = 1$ $H_{\leq j} = 1 + tH_{\leq j-1}^2$	$W_{\leq -1} = 1$ $W_{\leq j} = 1 + tW_{\leq j+1}W_{\leq j-1}$ <p style="text-align: center;">(<math>j \geq 0</math>)</p>
	$H_{\leq 0} = 1$ $H_{\leq j} = 1 + tH_{\leq j-1}^*H_{\leq j}^*$	

## Trees of bounded height or bounded width: functional equations

Family of trees	Bounded height	Bounded (right) width
	$H_{\leq 0} = 1$ $H_{\leq j} = 1 + tH_{\leq j-1}^2$	$W_{\leq -1} = 1$ $W_{\leq j} = 1 + tW_{\leq j+1}W_{\leq j-1}$ <p style="text-align: center;">(<math>j \geq 0</math>)</p>
	$H_{\leq 0} = 1$ $H_{\leq j} = 1 + tH_{\leq j-1}^*$	

Induction on  $j$



## Generating functions for trees of bounded height / width

**Proposition:** The generating function of plane trees of height  $\leq j$  is:

$$H_{\leq j}^* = A \frac{1 - Q^{j+1}}{1 - Q^{j+2}}$$

where  $A$  counts plane trees and  $Q = A - 1$  [de Bruijn, Knuth, Rice 72].

## Generating functions for trees of bounded height / width

**Proposition:** The generating function of plane trees of height  $\leq j$  is:

$$H_{\leq j}^* = A \frac{1 - Q^{j+1}}{1 - Q^{j+2}}$$

where  $A$  counts plane trees and  $Q = A - 1$  [de Bruijn, Knuth, Rice 72].

◁ ◁ ◊ ▷ ▷ ▷

**Proposition** [mbm 05]: The generating function of binary trees of right width  $\leq j$  is:

$$W_{\leq j} = A \frac{(1 - Z^{j+2})(1 - Z^{j+7})}{(1 - Z^{j+4})(1 - Z^{j+5})},$$

where  $A$  counts binary trees and  $Z \equiv Z(t)$  is the unique series in  $t$  such that

$$Z = t \frac{(1 + Z^2)^2}{1 - Z + Z^2} \quad \text{and} \quad Z(0) = 0.$$

But. . . why?

[Bouttier et al. 03]

## Limit results for the height of trees

**Convergence of the moments:** Let  $H_n^*$  be the (random) height of a plane tree with  $n$  edges. As  $n \rightarrow \infty$ ,

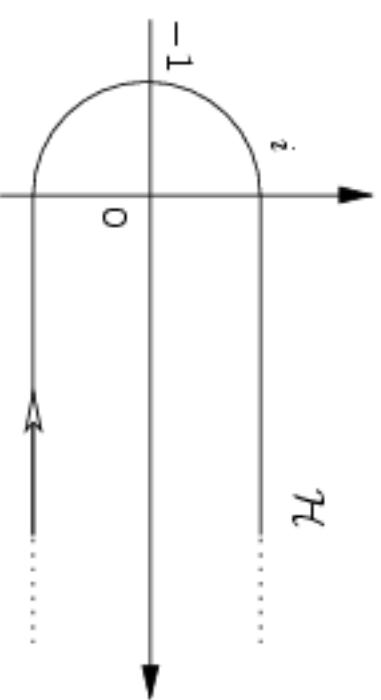
$$\frac{\mathbb{E}(H_n^*)}{\sqrt{n}} \rightarrow \sqrt{\pi}, \quad \mathbb{E} \left( \left( \frac{H_n^*}{\sqrt{n}} \right)^k \right) \rightarrow k(k-1)\Gamma(k/2)\zeta(k) \quad \text{for } k \geq 2.$$

**Convergence in law:**

$$\mathbb{P} \left( \frac{H_n^*}{\sqrt{n}} \leq x \right) \rightarrow F(x),$$

where

$$\begin{aligned} F(x) &= -\frac{1}{i\sqrt{\pi}} \int_{\mathcal{H}} \coth(x\sqrt{-z})\sqrt{-z}e^{-z} dz \\ &= \sum_{k=-\infty}^{+\infty} e^{-k^2 x^2} (1 - 2k^2 x^2) \end{aligned}$$



[Flajolet-Odlyzko 82], [Brown-Schubert 84]

+ similar results for binary trees

## Limit results for the right width of binary trees [Imbm 05]

**Convergence of the moments:** Let  $W_n$  be the (random) right width of a binary tree with  $n$  nodes. Then, as  $n \rightarrow \infty$ ,

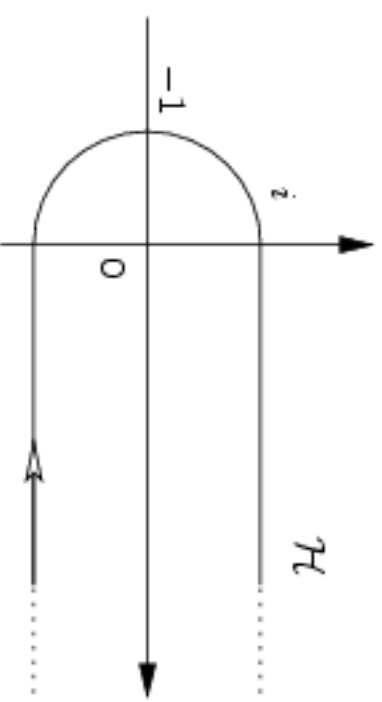
$$\frac{\mathbb{E}(W_n)}{n^{1/4}} \rightarrow \frac{3\sqrt{\pi}}{\sqrt{2}\Gamma(3/4)}, \quad \mathbb{E}\left(\left(\frac{W_n}{n^{1/4}}\right)^2\right) \rightarrow 6\sqrt{\pi},$$

$$\mathbb{E}\left(\left(\frac{W_n}{n^{1/4}}\right)^k\right) = \frac{24\sqrt{\pi}k!\zeta(k-1)}{\sqrt{2}^k\Gamma((k-2)/4)} \quad \text{for } k \geq 3.$$

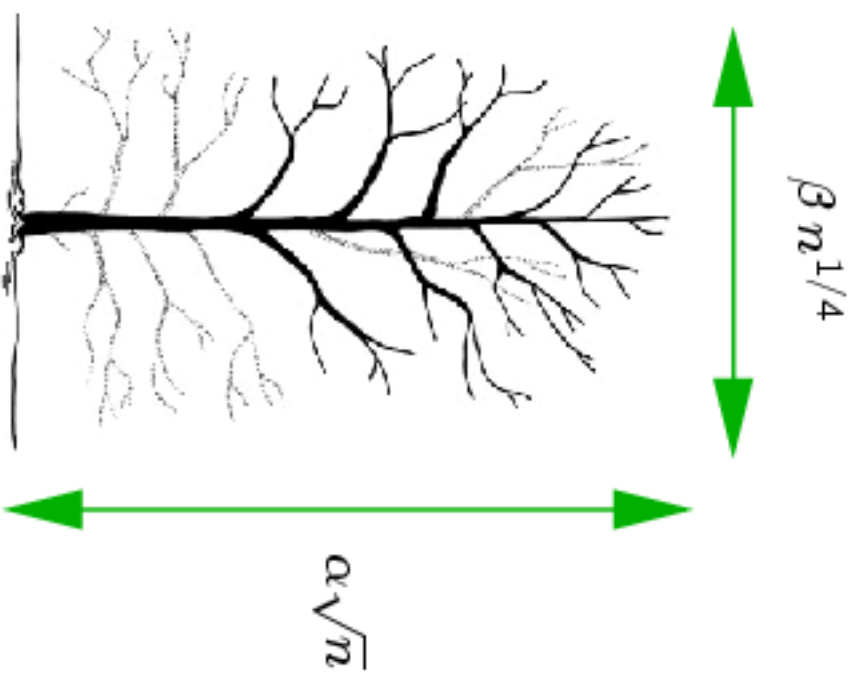
**Convergence in law:**

$$\mathbb{P}\left(\frac{W_n}{n^{1/4}} \geq x\right) \rightarrow G(x)$$

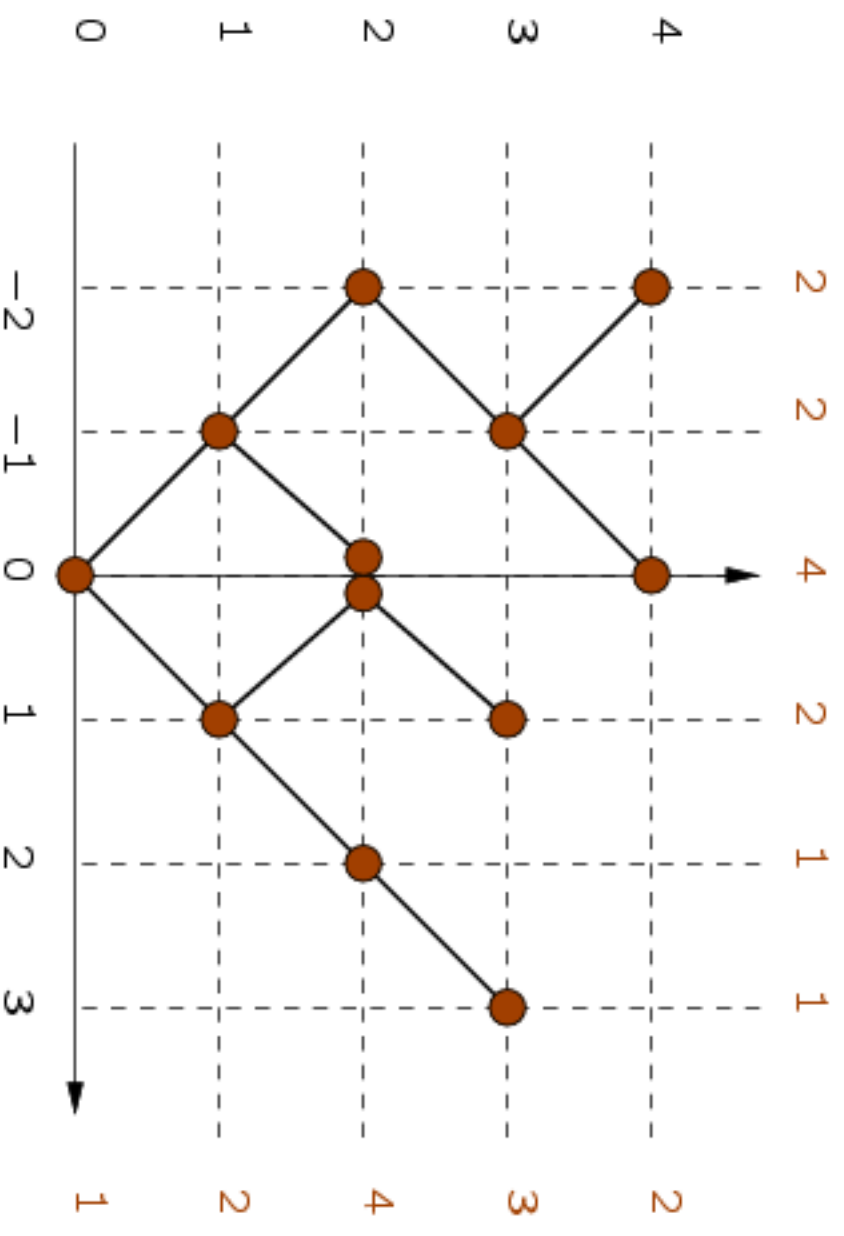
$$G(x) = \frac{3}{i\sqrt{\pi}} \int_{\mathcal{H}} \frac{\sqrt{-ze^{-z}}}{\sinh^2(x(-z^{1/4}/\sqrt{2}))} dz$$



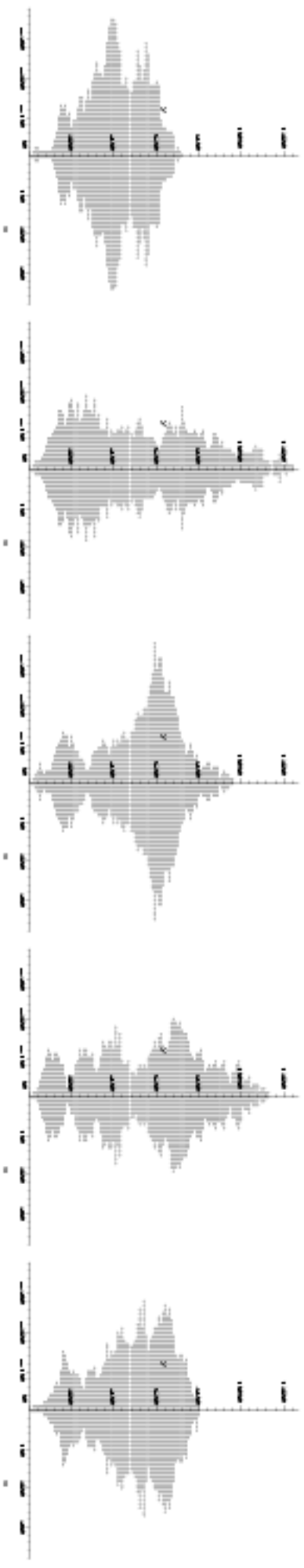
## Height and width of binary trees



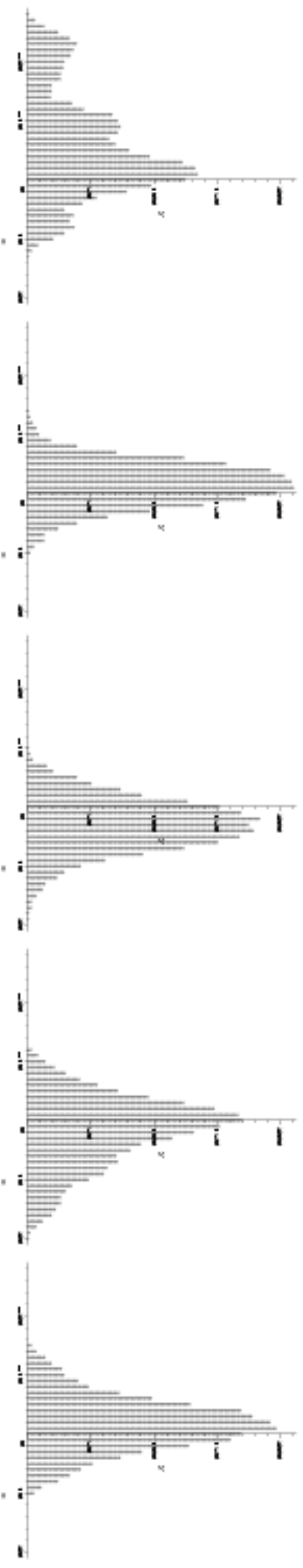
## What about the profiles?



## Gallery of (horizontal and vertical) profiles



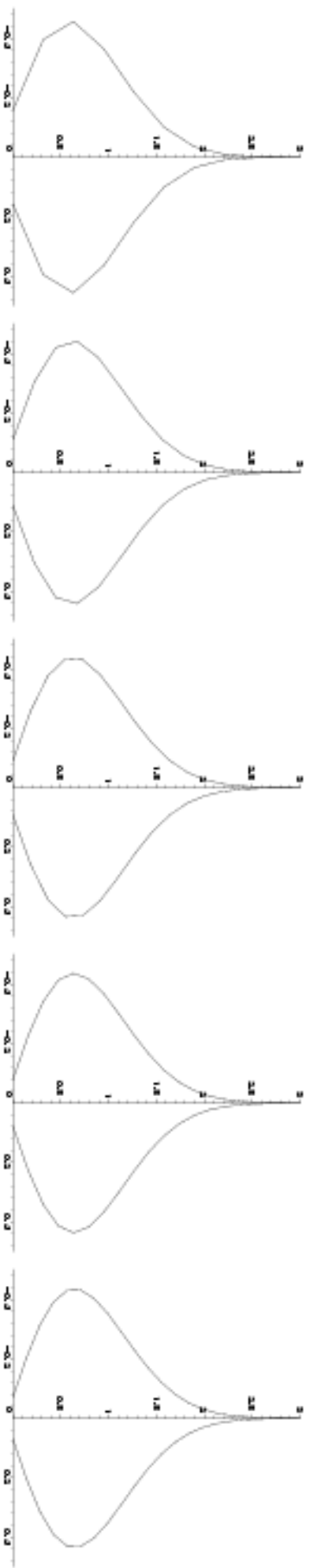
Horizontal profiles of random binary trees with 1000 nodes.



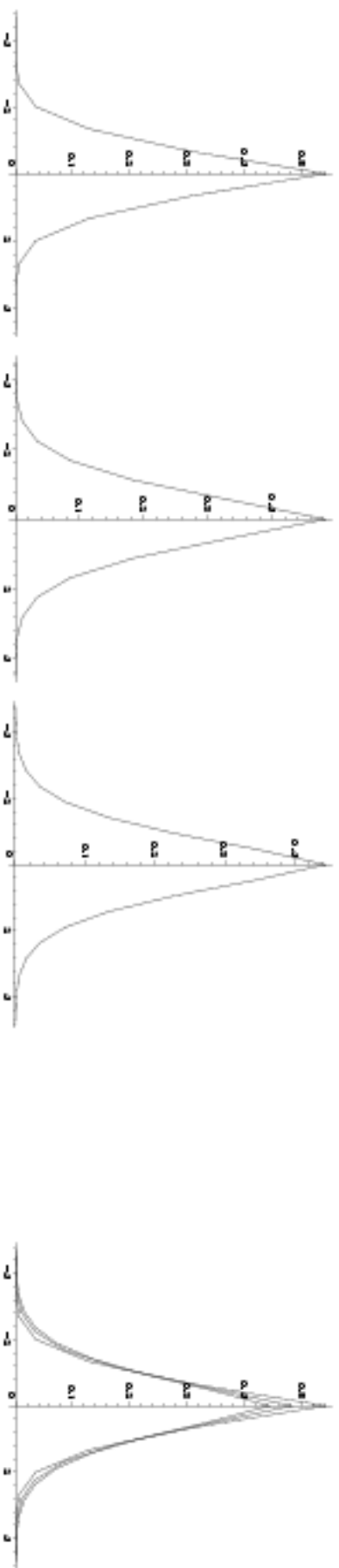
Vertical profiles of random binary trees with 1000 nodes.

## The average (horizontal and vertical) profiles

Average horizontal profile of plane trees of size 10, 20, 30, 40, 50:



Average vertical profile of binary trees of size 5, 10, 20:





## The horizontal profile

A tree of size  $n$  has, on average, height  $O(\sqrt{n})$ . How many nodes are there at a given (horizontal) level?

???

## The horizontal profile

A tree of size  $n$  has, on average, height  $O(\sqrt{n})$ . How many nodes are there at a given (horizontal) level?

About  $\sqrt{n}$ .

## The horizontal profile

A tree of size  $n$  has, on average, height  $O(\sqrt{n})$ . How many nodes are there at a given (horizontal) level?

About  $\sqrt{n}$ .

Let  $X_{n^*}(j)$  be the number of nodes located at height  $j$  in a random plane tree with  $n$  edges.

We study the quantity

$$\frac{X_{n^*}(\lfloor \lambda \sqrt{n} \rfloor)}{\sqrt{n}}.$$

## The vertical profile

A binary tree of size  $n$  has, on average, width  $O(n^{1/4})$ . How many nodes are there on a given (vertical) layer?

About  $n^{3/4}$ .

Let  $Y_n(j)$  be the number of nodes located at abscissa  $j$  in a random binary tree with  $n$  nodes.

We study the quantity

$$\frac{Y_n(\lfloor \lambda n^{1/4} \rfloor)}{n^{3/4}}.$$

## One new ingredient: bivariate generating functions

Before: Given  $j$ , how many plane trees of size  $n$  have height at most  $j$ ?

$$\Rightarrow \text{series } H_{\leq j}^*(t) = \sum_n h_{n, \leq j} t^n$$

Now: Given  $j$ , how many plane trees of size  $n$  have exactly  $k$  nodes at height  $j$ ?

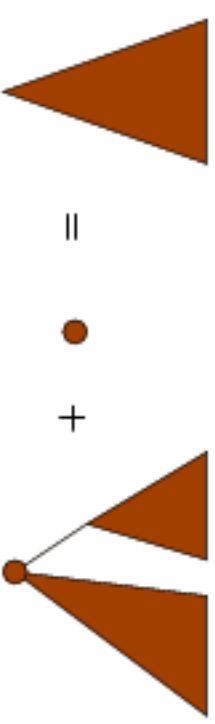
$$\Rightarrow \text{series } P_j^*(t, u) = \sum_{n, k} p_{n, k, j} t^n u^k$$

## The number of nodes at height $j$ (plane trees)

Let  $P_j^* \equiv P_j^*(t, u)$  be the generating function of plane trees, counted by edges (variable  $t$ ) and by the number of nodes at height  $j$  (variable  $u$ ):

$$P_0^* = uA(t)$$

$$P_j^* = 1 + tP_{j-1}^*P_j^*$$



**Proposition** [???, mbm 04]:

$$P_j^* = A \frac{1 - MQ^j}{1 - MQ^{j+1}}$$

where  $A$  counts plane trees,  $Q = A - 1$  and

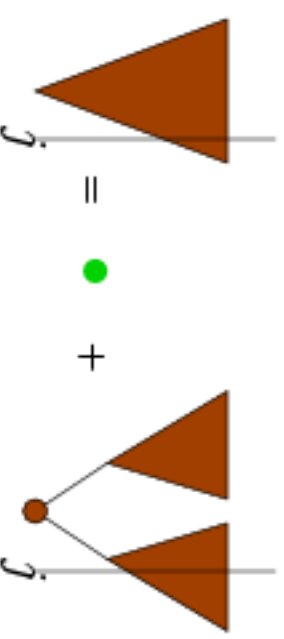
$$M = \frac{A - u - tuA^2}{u + A(1 - u) + tuA^2(1 - A)}.$$

## The number of nodes at abscissa $j$ (binary trees)

Let  $V_j \equiv V_j(t, u)$  be the generating function of binary trees, counted by nodes ( $t$ ) and by the number of nodes at abscissa  $j$  (variable  $u$ ).

$$V_0 = 1 + utV_1^2$$

$$\text{For } j \geq 0, \quad V_j = 1 + tV_{j+1}V_{j-1}$$



**Proposition [mbm 05]:** the series  $V_j$  are algebraic. Moreover,

$$V_j = A \frac{(1 + MZ^j)(1 + MZ^{j+5})}{(1 + MZ^{j+2})(1 + MZ^{j+3})}$$

where  $A$  counts binary trees,

$$Z = t \frac{(1 + Z^2)^2}{1 - Z + Z^2},$$

and  $M \equiv M(t, u)$  is the unique power series in  $t$  such that

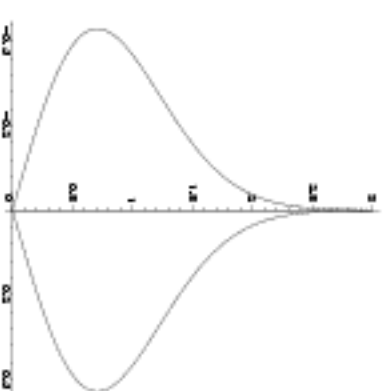
$$M = (u - 1) \frac{Z(1 + MZ)^2(1 + MZ^2)(1 + MZ^6)}{(1 + Z)^2(1 + Z + Z^2)(1 - Z)^3(1 - M^2Z^5)}.$$

## The average horizontal profile of plane trees

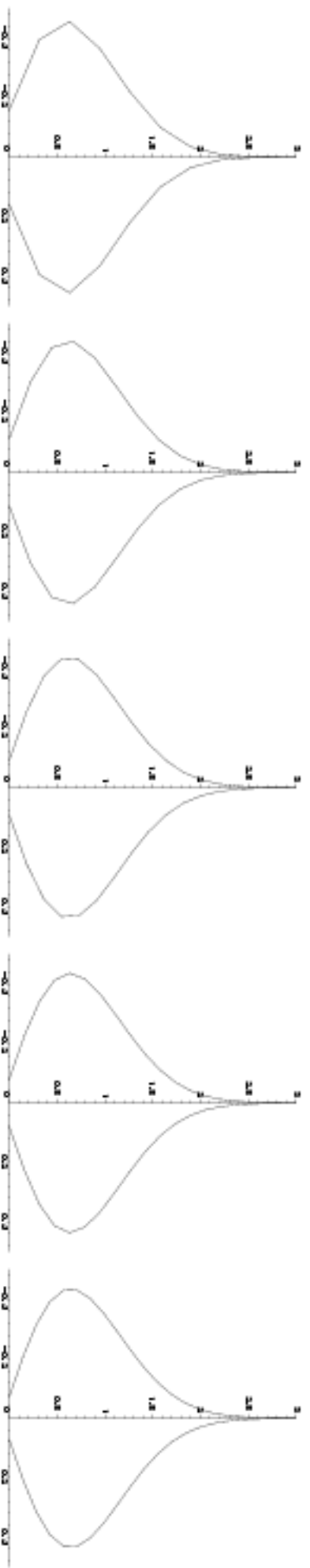
Proposition :

$$\mathbb{E} \left( \frac{X_n^* (\lfloor \lambda \sqrt{n} \rfloor)}{\sqrt{n}} \right) \rightarrow 2\lambda e^{-\lambda^2}.$$

On average, there are about  $2\lambda e^{-\lambda^2} \sqrt{n}$  nodes at height  $\lfloor \lambda \sqrt{n} \rfloor$  in a plane tree having  $n$  edges.



Average profile of plane trees of size 10, 20, 30, 40, 50:



+ Convergence in law (and as a process) of the horizontal profile [Drmotá-Gittenberger 97]

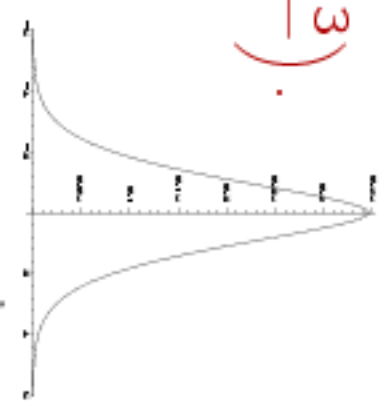


## The average vertical profile of binary trees

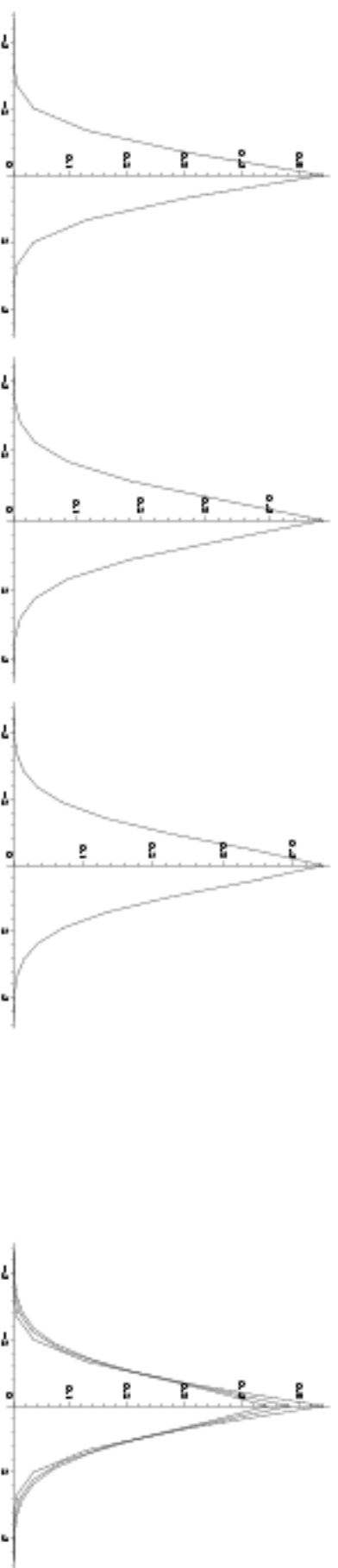
**Proposition [mbm 05]:** Let  $Y_n(j)$  be the (random) number of nodes at abscissa  $j$  in a binary tree having  $n$  nodes. Then

$$\mathbb{E} \left( \frac{Y_n(\lfloor \lambda n^{1/4} \rfloor)}{n^{3/4}} \right) \rightarrow \frac{1}{\sqrt{2\pi}} \sum_{m \geq 0} \frac{(-\sqrt{2}|\lambda|)^m}{m!} \cos \frac{(m+1)\pi}{4} \Gamma \left( \frac{m+3}{4} \right).$$

This gives the average number of nodes at abscissa  $\lfloor \lambda n^{1/4} \rfloor$  in a random binary tree having  $n$  nodes.

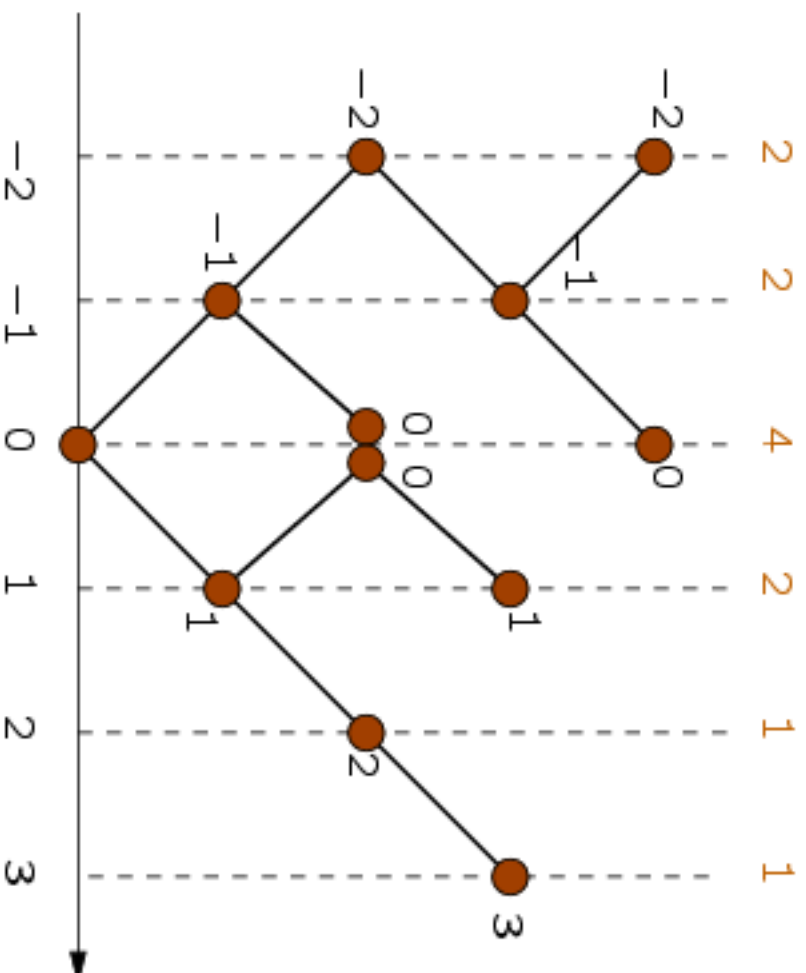


Average vertical profile of binary trees of size 5, 10, 20 :



+ Convergence in law (and as a process [mbm-Janson 05])

## Why study the vertical profile?

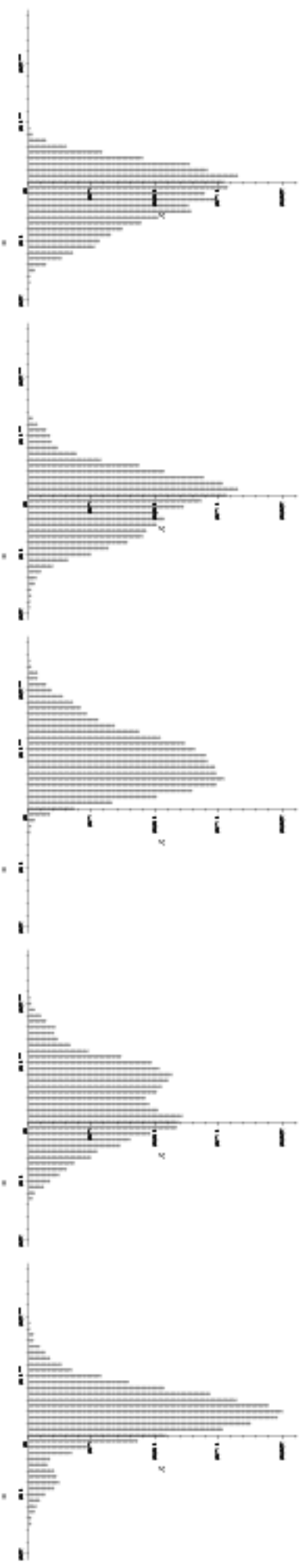


LABEL  
=  
ABSCISSA

A binary tree of **right width 3**,  
and vertical profile **[2, 2; 4, 2, 1, 1]**.

## Why study the vertical profile?

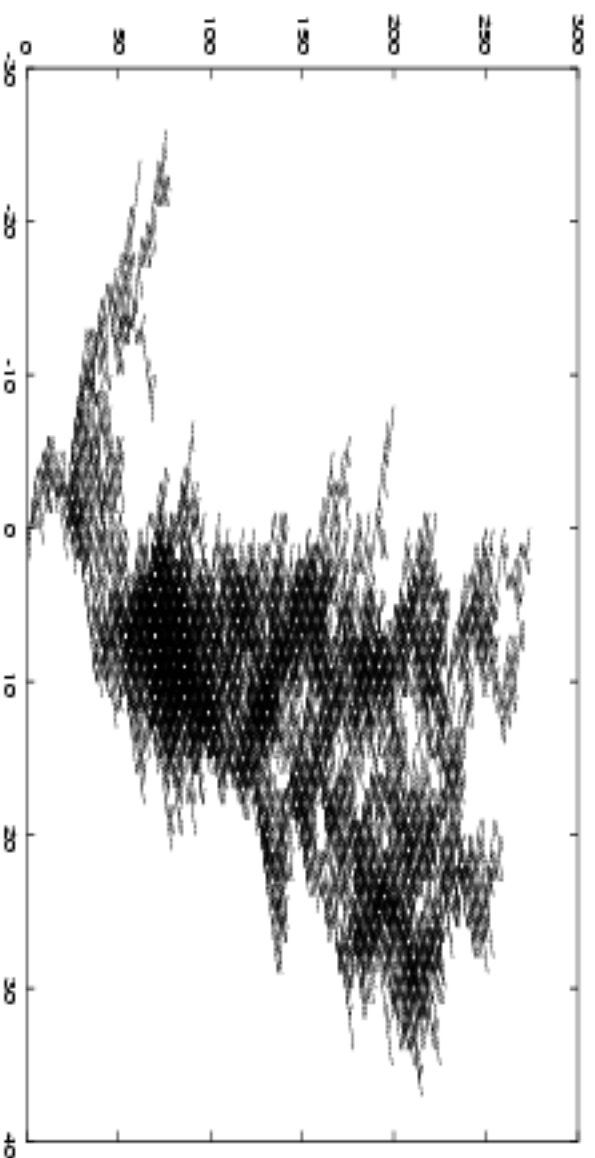
Because the vertical profile of (random) binary trees, seen as the histogram of a (random) probability distribution, converges to the [ISE](#).



## Did you say “ISE”?

### The Integrated SuperBrownian Excursion!

The ISE is a (random) probability distribution that occurs (almost) everywhere. At least, as soon as a branching structure (tree) is combined with an embedding of the nodes in the space ( $\mathbb{Z}$  here). The ISE describes how the space is occupied by the nodes [Aldous 93], [Marckert-Mokkadem 03].



**That's it!**



## Convergence in law: A special case

**Theorem [mbm 05]** : The random variable

$$\frac{Y_n(0)}{n^{3/4}},$$

which gives the (normalized) number of nodes at abscissa 0, converges in law to a variable  $Y(0)$  such that

$$E(Y(0)^k) = \left(\frac{\sqrt{2}}{3}\right)^k \frac{\Gamma(1 + 3k/4)}{\Gamma(1 + k/2)}.$$

Hence

$$Y(0) = \frac{\sqrt{2}}{3\sqrt{T_{2/3}}},$$

where  $T_{2/3}$  follows a unilateral stable law of parameter  $2/3$ .

$$E(e^{-aT_{2/3}}) = e^{-a^{2/3}} \text{ pour } a \geq 0.$$

Merci **Alain Rouault** !

Did you say "ISE"?

