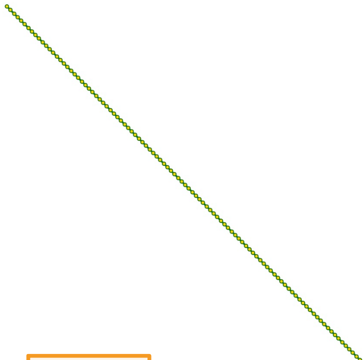
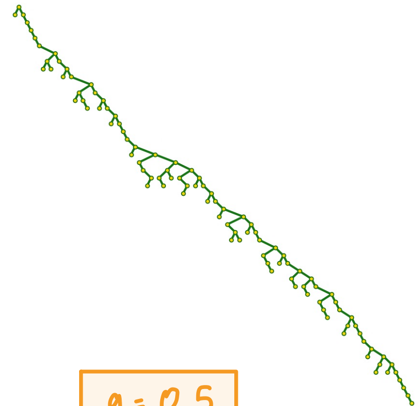

The height
of
Mallows trees

Example of Mallows trees (n,q):

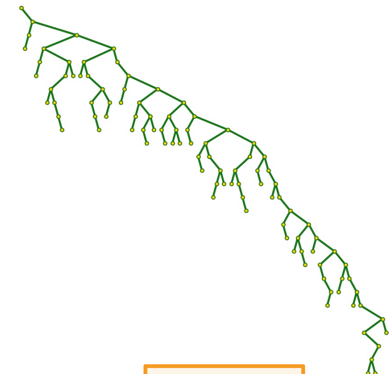
$n = 100$



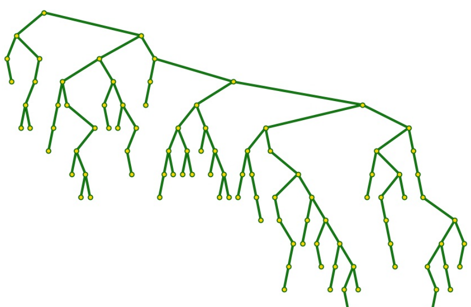
$q = 0$



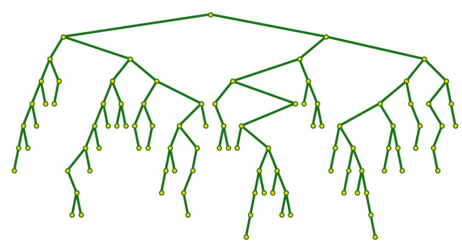
$q = 0.5$



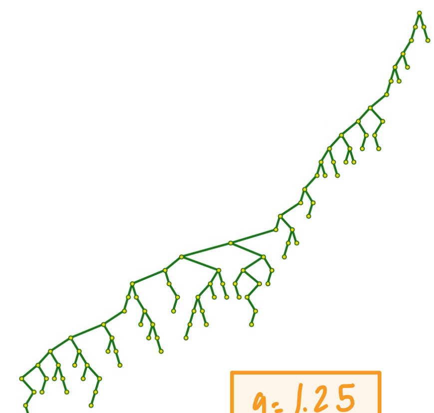
$q = 0.8$



$q = 0.9$



$q = 1$



$q = 1.25$

Selling pitch:

- Mallows trees are random binary search trees.
- Binary search trees are common structures in computer science, related to sorting and accessing data.
- The height of binary search trees is related to the time/difficulty to access data.

* Mallows trees would be terrible sorting structures for data...

1. Mallows trees
2. Right path
3. Links with Math 447

1 Mallows trees

- Imagine a matrix $B=(B_{ij})_{i,j\geq 1}$ with Bernoulli$(1-q)$ entries:

$$\mathbb{P}[B_{ij}=0]=q$$

- Consider the 1st non-zero entry in the 1st line:

  j_1 such that $B_{1j_1}=1$ and $B_{ij}=0$ for $j<j_1$

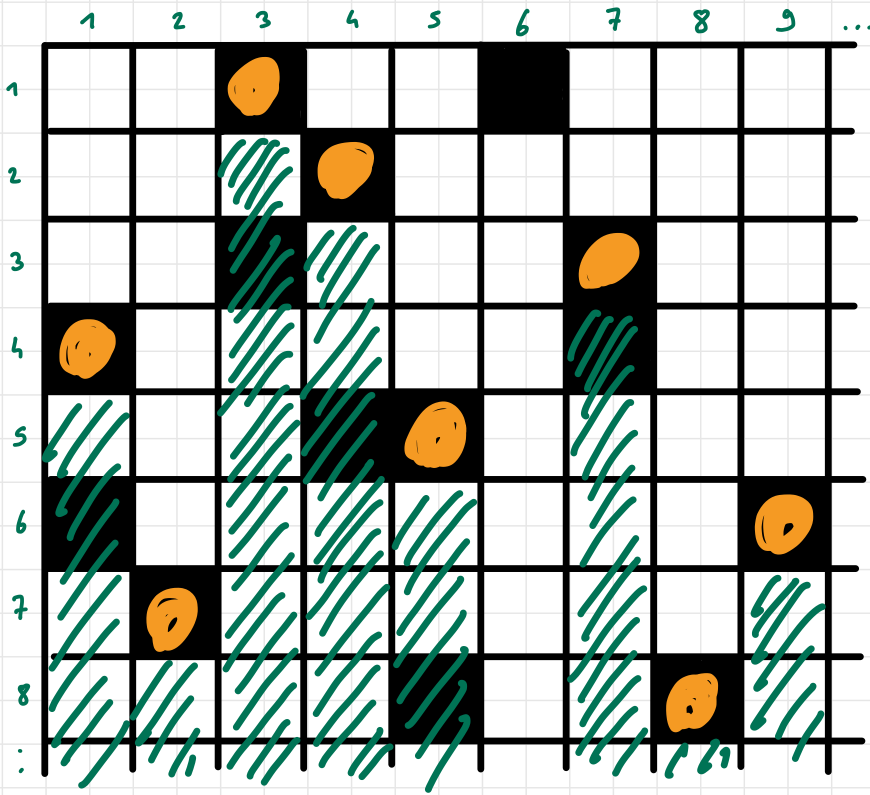
- With j_1 defined, consider the 1st non-zero entry in the 2nd line different than j_1:

  j_2 such that $B_{2j_2}=1$ and $B_{ij}=0$ for $j<j_2$, $j\neq j_1$

- Repeat...

Mallows trees

$B =$



$j_1 = 3$
$j_2 = 4$
$j_3 = 7$
$j_4 = 1$
$j_5 = 5$
$j_6 = 9$
$j_7 = 2$
$j_8 = 8$

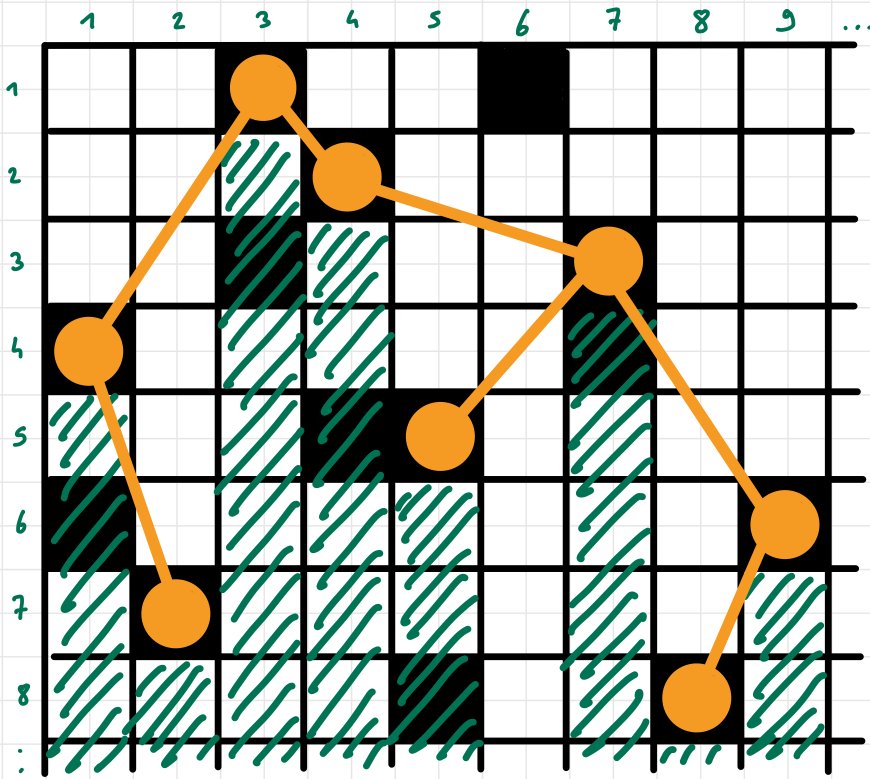
□ : 0 ■ : 1

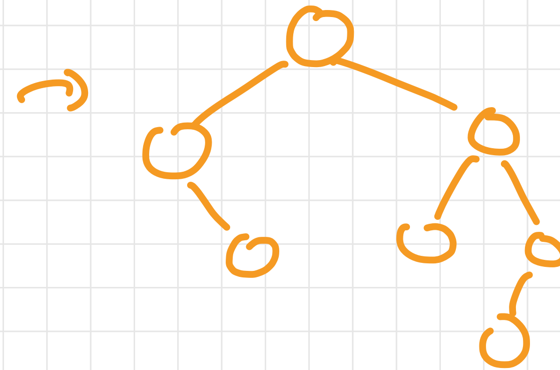
1 Mallows trees

With this drawing, imagine that the orange dots appear one after the other, starting from the top.

We create a tree by connecting each new dot to the most recent existing dot, while not going over green lines.

Mallows trees





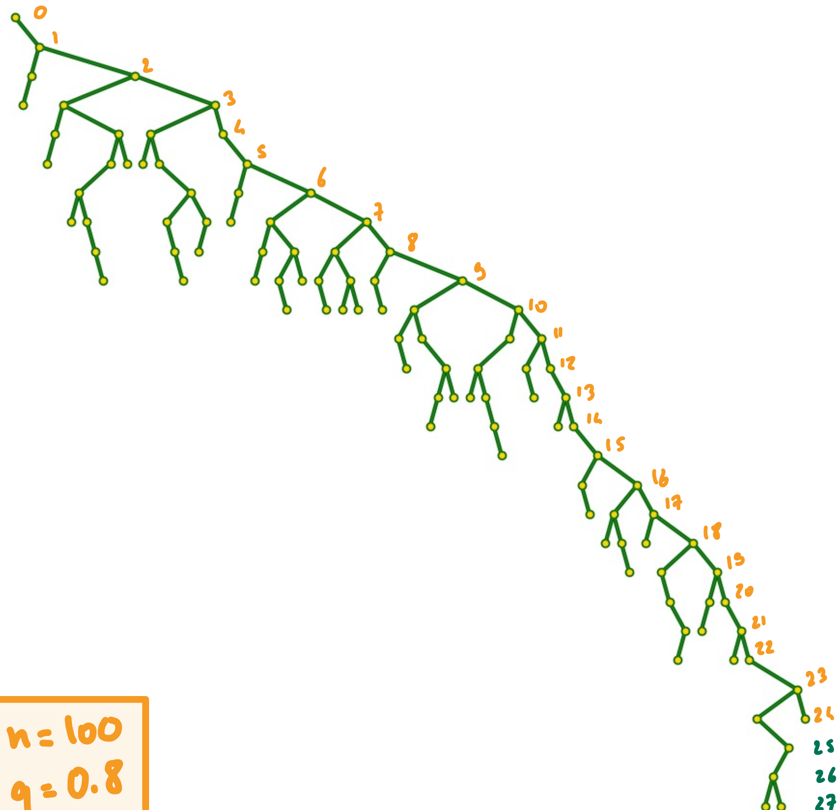
1 Mallows trees

Exact definition

- Mallows tree are a form of random binary search trees.
- Binary search trees are labelled tree with the label of a node being larger (smaller) than the labels of its left (right) subtree.
- Mallows trees are obtained by inserting a Mallows permutation in a binary search tree.

→ $\mathbb{P}[\text{Mallows permutation} = \sigma] \sim q^{\mathrm{Inv}(\sigma)}$ inversion number of σ

2 Right path



$n = 100$
$q = 0.8$

Going to the left:

1st path = 3

2nd path = 5

3rd path = 7

Going to the right:

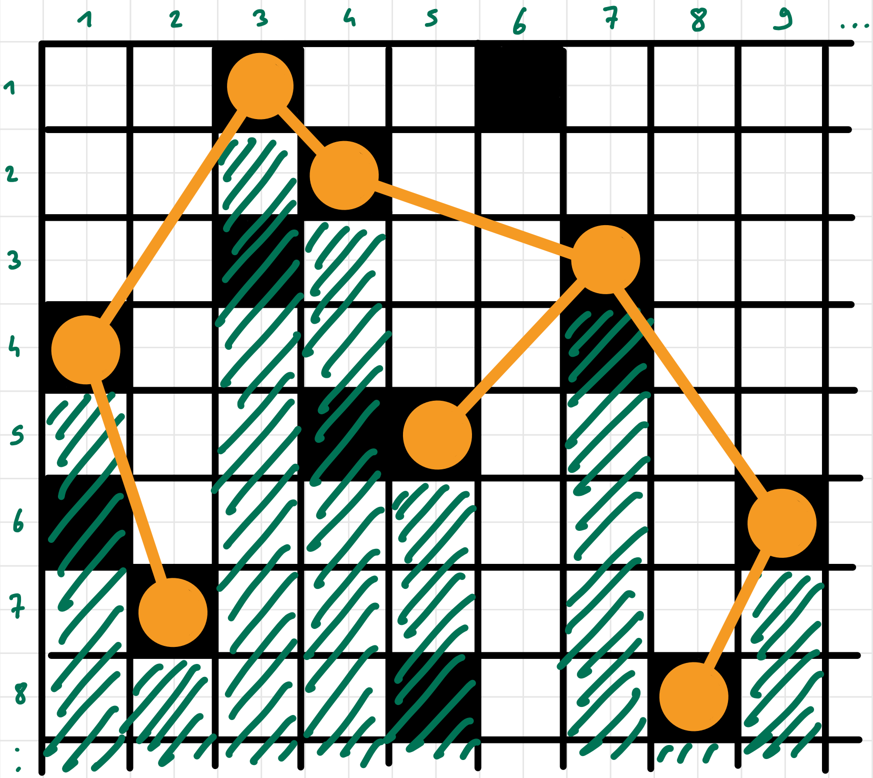
1st path = 24

2nd path = 27

Height : 27

2 Right path

R_n = length of right path after n dots



- $R_1 = 0$
- $R_2 = 1$
- $R_3 = 2$
- $R_4 = 2$
- $R_5 = \cancel{2}$
- $R_6 = 3$
- $R_7 = 3$
- $R_8 = \cancel{3}$

$\text{height}_1 = 0$

$\text{height}_2 = 1$

$\text{height}_3 = 2$

$\text{height}_4 = 2$

$\text{height}_5 = \cancel{3}$

$\text{height}_6 = 3$

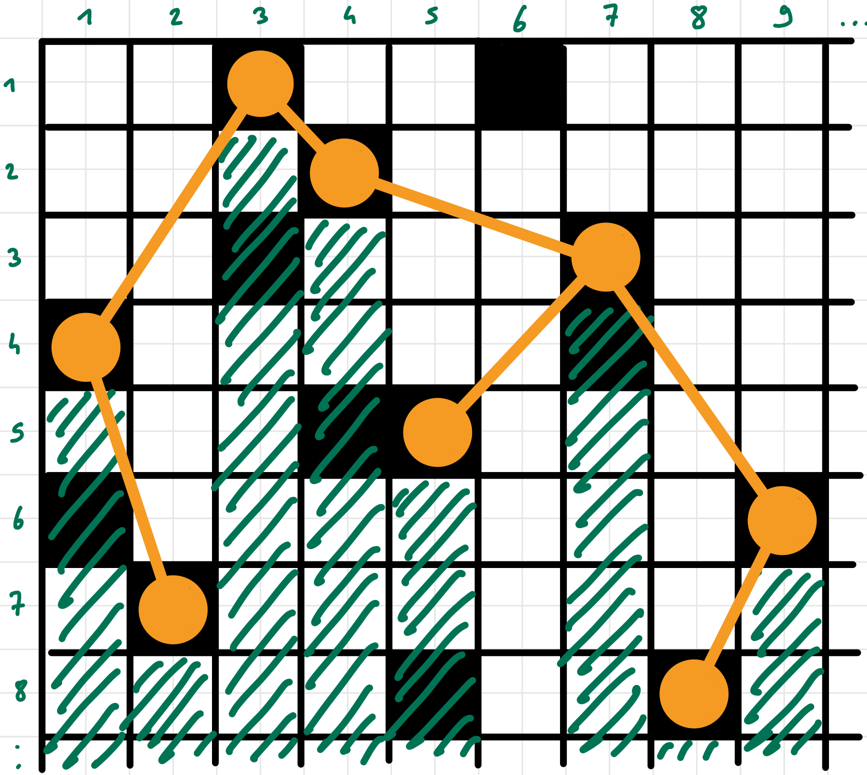
$\text{height}_7 = 3$

$\text{height}_8 = \cancel{4}$

Hope: $R_n \simeq \text{height}_n$

2 Right path

M_n = position of the rightmost of the first n dots



$R_1 = 0$	$M_1 = 3$	$height_1 = 0$
$R_2 = 1$	$M_2 = 4$	$height_2 = 1$
$R_3 = 2$	$M_3 = 7$	$height_3 = 2$
$R_4 = 2$	$M_4 = 7$	$height_4 = 2$
$R_5 = 2$	$M_5 = 7$	$height_5 = 3$
$R_6 = 3$	$M_6 = 9$	$height_6 = 3$
$R_7 = 3$	$M_7 = 9$	$height_7 = 3$
$R_8 = 3$	$M_8 = 9$	$height_8 = 4$

Hope: M_n helps finding R_n

3 Links with Math 447

- Where is the Markov chain?
- Can you find the hidden Geometric and why their memoryless property is important?
- Generating functions are easily computed for what variable?
- Are Mallows trees branching processes?
- Where is the coupling and why are we using it?
- Bonus: Markov's inequality is also widely used with which previous point? $\longrightarrow$ Markov's ineq. + Gen. fun. = Chernoff's bound

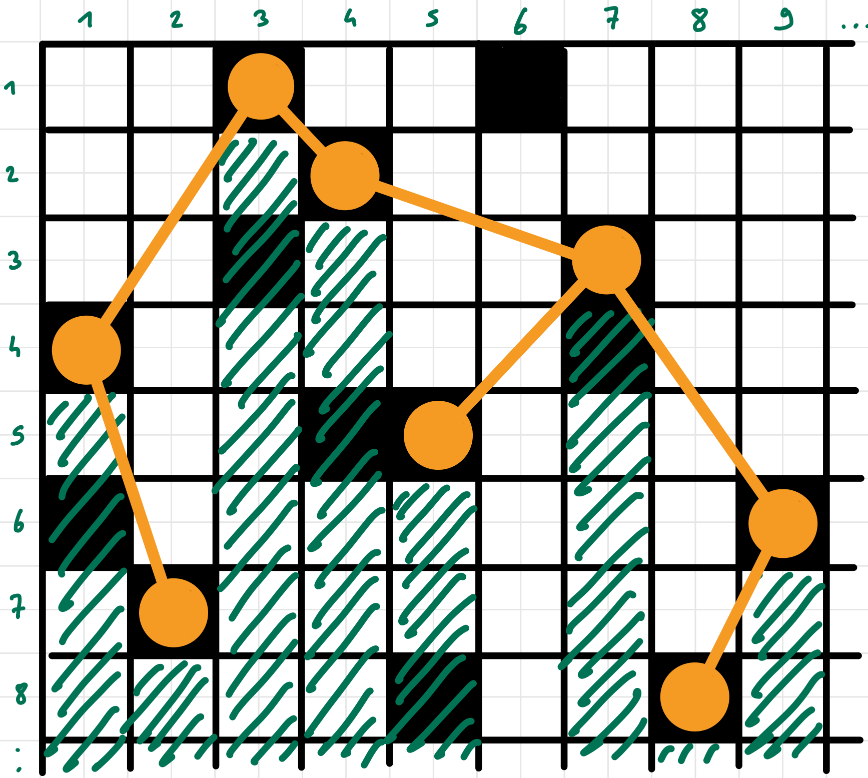
3 Links with Math 447

Markov chains:

- (M_n) is a Markov chain.
  - it is not time homogeneous.
  - it is transient and converges to ∞ (it is increasing).
- (R_n, M_n) is also a Markov chain.
- (R_n) is probably not a Markov chain.

The interesting object is $M_n - n$

Links with Math 447



$M_1 = 3 \longrightarrow M_1 - 1 = 2$

$M_2 = 4 \longrightarrow M_2 - 2 = 2$

$M_3 = 7 \longrightarrow M_3 - 3 = 4$

$M_4 = 7 \longrightarrow M_4 - 4 = 3$

$M_5 = 7 \longrightarrow M_5 - 5 = 2$

$M_6 = 9 \longrightarrow M_6 - 6 = 3$

$M_7 = 9 \longrightarrow M_7 - 7 = 2$

$M_8 = 9 \longrightarrow M_8 - 8 = 1$

3 Links with Math 447

Geometric:

→ $(M_n \cdot n)$ is a positive recurrent, time homogeneous M C.

→ What is P_{ij} ?

→ If a spot is empty with probability $1-q$, how far is the first non-empty one : Geometric $(1-q)$

→ It follows that :

$$P_{ij} = \begin{cases} \mathbb{P}[\text{Geometric}(1-q) \leq i\ ] & \text{if } j = i-1 \\ \mathbb{P}[\text{Geometric}(1-q) = j+1] & \text{o.w.} \end{cases}$$

3 Links with Math 447

Generating function:

→ It was already known that

$$f(s) = \mathbb{E}[s^{R_n}] = \prod_{1<k\leq n} \left(1 + (s-1)\frac{1-q}{1-q^k}\right)$$

→ $\mathbb{P}[R_n = r]$ is quite complicated.

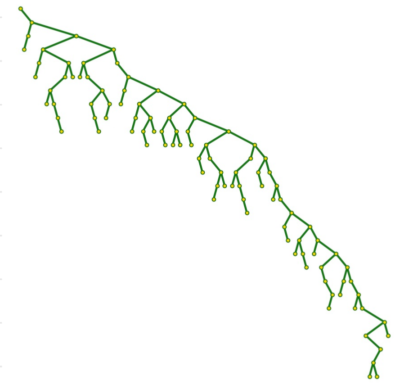
→ $\mathbb{E}[R_n] = f'(1) = \sum_{1<k\leq n} \frac{1-q}{1-q^k}$

→ We extended to (R_n, M_n):

$$\mathbb{E}[s^{R_n} \cdot t^{M_n}] = \frac{t^n}{s} \cdot \prod_{1\leq k\leq n} \frac{q+(1-q)s-q^k}{1-t\cdot q^k}$$

bivariate generating function

3 Links with Math 447



Branching process:

→ Not a Galton-Watson tree.

→ Why?

→ Possibly a branching process?

→ Related to branching random walks:

* Galton-Watson trees with positions for the individuals, following random walks.

3 Links with Math 447

Coupling:

→ The interest: connecting two (or more) random variables

→ Mallows trees are given with two parameters:
n and q

→ There is no obvious reasons why they should be connections between the different parameters.

⇝ Using a single matrix, we can cover all $n \in \mathbb{N}$.

This is a coupling.

3 Links with Math 447

Coupling:

Why is this interesting?

⇝ Relation between n-th tree and $(n+1)$-th tree:

- Proofs by induction.
- Recursif Mallows trees generating models.
- Direct extraction of properties: height increases with n.

⇝ Relation between q and q' trees.

$\boxed{\text{OPEN QUESTION:}}$ height decreases with q?


Thank you!