LYNDON TREE & BINARY SEARCH TREE

LUCAS MERCIER

&

PHILIPPE CHASSAING

INSTITUT ELIE CARTAN

Séminaire PHILIPPE FLAJOLET , 31 Janvier 2013

GLOSSARY

- Alphabet
- n-letters long words
- Language
- u is a factor of w
- u is a Prefix of w
- u is a Suffix of w
- Rotation
- Necklace, circular word
- Primitive word

GLOSSARY

- Alphabet $\mathcal{A} = \{a_1 < a_2 < \cdots < a_k < \cdots\}$
- n-letters long words
- Language
- u is a factor of w
- u is a Prefix of w
- u is a Suffix of w
- Rotation
- Necklace, circular word
- Primitive word

GLOSSARY

- Alphabet $\mathcal{A} = \{a_1 < a_2 < \cdots < a_k < \cdots\}$
- n-letters long words $\omega = \omega_1\omega_2\ldots\omega_n,\ \omega_i \in \mathcal{A},\ |\omega| = n,\ \omega \in \mathcal{A}^n$
- Language
- u is a factor of w
- u is a Prefix of w
- u is a Suffix of w
- Rotation
- Necklace, circular word
- Primitive word

GLOSSARY

- Alphabet $\mathcal{A} = \{a_1 < a_2 < \cdots < a_k < \cdots\}$
- n-letters long words $\omega = \omega_1\omega_2 \ldots \omega_n,\ \omega_i \in \mathcal{A},\ |\omega| = n,\ \omega \in \mathcal{A}^n$
- Language $\mathcal{A}^* = \{\emptyset\} \cup \mathcal{A}^1 \cup \mathcal{A}^2 \cup \mathcal{A}^3 \cup \ldots$
- u is a factor of w
- u is a Prefix of w
- u is a Suffix of w
- Rotation
- Necklace, circular word
- Primitive word

GLOSSARY

- Alphabet $\mathcal{A} = \{a_1 < a_2 < \cdots < a_k < \cdots\}$
- n-letters long words $\omega = \omega_1\omega_2\ldots\omega_n,\ \omega_i \in \mathcal{A},\ |\omega| = n,\ \omega \in \mathcal{A}^n$
- Language $\mathcal{A}^* = \{\emptyset\} \cup \mathcal{A}^1 \cup \mathcal{A}^2 \cup \mathcal{A}^3 \cup \ldots$
- u is a factor of w $\exists\, r, s \in \mathcal{A}^\star$ such that $w = rus$
- u is a Prefix of w
- u is a Suffix of w
- Rotation
- Necklace, circular word
- Primitive word

GLOSSARY

- Alphabet $\mathcal{A} = \{a_1 < a_2 < \cdots < a_k < \cdots\}$
- n-letters long words $\omega = \omega_1\omega_2\ldots\omega_n,\ \omega_i \in \mathcal{A},\ |\omega| = n,\ \omega \in \mathcal{A}^n$
- Language $\mathcal{A}^* = \{\emptyset\} \cup \mathcal{A}^1 \cup \mathcal{A}^2 \cup \mathcal{A}^3 \cup \ldots$
- u is a factor of w $\exists\, r, s \in \mathcal{A}^\star$ such that $w = rus$
- u is a Prefix of w $r = \emptyset$
- u is a Suffix of w
- Rotation
- Necklace, circular word
- Primitive word

GLOSSARY

- Alphabet $\mathcal{A} = \{a_1 < a_2 < \cdots < a_k < \cdots\}$
- n-letters long words $\omega = \omega_1\omega_2\ldots\omega_n,\ \omega_i \in \mathcal{A},\ |\omega| = n,\ \omega \in \mathcal{A}^n$
- Language $\mathcal{A}^* = \{\emptyset\} \cup \mathcal{A}^1 \cup \mathcal{A}^2 \cup \mathcal{A}^3 \cup \ldots$
- u is a factor of w $\exists\, r, s \in \mathcal{A}^\star$ such that $w = rus$
- u is a Prefix of w $r = \emptyset$
- u is a Suffix of w $s = \emptyset$
- Rotation
- Necklace, circular word
- Primitive word

GLOSSARY

- Alphabet $\mathcal{A} = \{a_1 < a_2 < \cdots < a_k < \cdots\}$
- n-letters long words $\omega = \omega_1\omega_2\ldots\omega_n,\ \omega_i \in \mathcal{A},\ |\omega| = n,\ \omega \in \mathcal{A}^n$
- Language $\mathcal{A}^* = \{\emptyset\} \cup \mathcal{A}^1 \cup \mathcal{A}^2 \cup \mathcal{A}^3 \cup \ldots$
- u is a factor of w $\exists\, r, s \in \mathcal{A}^\star$ such that $w = rus$
- u is a Prefix of w $r = \emptyset$
- u is a Suffix of w $s = \emptyset$
- Rotation $\omega = \omega_1\omega_2\ldots\omega_n \longrightarrow \tau\omega = \omega_2\omega_3\ldots\omega_n\omega_1$
- Necklace, circular word
- Primitive word

GLOSSARY

- Alphabet $\mathcal{A} = \{a_1 < a_2 < \cdots < a_k < \cdots\}$
- n-letters long words $\omega = \omega_1\omega_2\ldots\omega_n,\ \omega_i \in \mathcal{A},\ |\omega| = n,\ \omega \in \mathcal{A}^n$
- Language $\mathcal{A}^* = \{\emptyset\} \cup \mathcal{A}^1 \cup \mathcal{A}^2 \cup \mathcal{A}^3 \cup \ldots$
- u is a factor of w $\exists\, r, s \in \mathcal{A}^\star$ such that $w = rus$
- u is a Prefix of w $r = \emptyset$
- u is a Suffix of w $s = \emptyset$
- Rotation $\omega = \omega_1\omega_2\ldots\omega_n \rightarrow \tau\omega = \omega_2\omega_3\ldots\omega_n\omega_1$
- Necklace, circular word $\langle\omega\rangle = \left\{\tau^k\omega \mid k \in \mathbb{Z}\right\}$
- Primitive word

GLOSSARY

- Alphabet $\mathcal{A} = \{a_1 < a_2 < \cdots < a_k < \cdots\}$
- n-letters long words $\omega = \omega_1\omega_2\ldots\omega_n,\ \omega_i \in \mathcal{A},\ |\omega| = n,\ \omega \in \mathcal{A}^n$
- Language $\mathcal{A}^* = \{\emptyset\} \cup \mathcal{A}^1 \cup \mathcal{A}^2 \cup \mathcal{A}^3 \cup \ldots$
- u is a factor of w $\exists\, r, s \in \mathcal{A}^\star$ such that $w = rus$
- u is a Prefix of w $r = \emptyset$
- u is a Suffix of w $s = \emptyset$
- Rotation $\omega = \omega_1\omega_2\ldots\omega_n \rightarrow \tau\omega = \omega_2\omega_3\ldots\omega_n\omega_1$
- Necklace, circular word $\langle\omega\rangle = \{\tau^k\omega \mid k \in \mathbb{Z}\}$
- Primitive word $\#\langle\omega\rangle = |\omega|$

GLOSSARY

- Alphabet $\mathcal{A} = \{a_1 < a_2 < \cdots < a_k < \cdots\}$
- n-letters long words $\omega = \omega_1\omega_2\ldots\omega_n,\ \omega_i \in \mathcal{A},\ |\omega| = n,\ \omega \in \mathcal{A}^n$
- Language $\mathcal{A}^* = \{\emptyset\} \cup \mathcal{A}^1 \cup \mathcal{A}^2 \cup \mathcal{A}^3 \cup \ldots$
- u is a factor of w $\exists\, r, s \in \mathcal{A}^\star$ such that $w = rus$
- u is a Prefix of w $r = \emptyset$
- u is a Suffix of w $s = \emptyset$
- Rotation $\omega = \omega_1\omega_2\ldots\omega_n \to \tau\omega = \omega_2\omega_3\ldots\omega_n\omega_1$
- Necklace, circular word $\langle\omega\rangle = \{\tau^k\omega \mid k \in \mathbb{Z}\}$
- Primitive word $\#\langle\omega\rangle = |\omega|$

LYNDON WORDS

- Lexicographic Order

LYNDON WORDS

- Lexicographic Order

$\mathbf{a} \prec \mathbf{b}$

$$\text{if} \quad \begin{cases} \text{either } \exists\, p, \alpha, \beta \in \mathcal{A}^\star, a_i, a_j \in \mathcal{A} \text{ s.t. } \; i < j, \; \begin{cases} \mathbf{a} = p a_i \alpha, \\ \mathbf{b} = p a_j \beta, \end{cases} \\ \\ \text{or } \mathbf{a} \text{ is a prefix of } \mathbf{b} \end{cases}$$

LYNDON WORDS

- Lexicographic Order

$\mathbf{a} \prec \mathbf{b}$

$$\text{if} \quad \begin{cases} \text{either } \exists\, p, \alpha, \beta \in \mathcal{A}^\star, a_i, a_j \in \mathcal{A} \text{ s.t. } \; i < j, \; \begin{cases} \mathbf{a} = p a_i \alpha, \\ \mathbf{b} = p a_j \beta, \end{cases} \\ \\ \text{or } \mathbf{a} \text{ is a prefix of } \mathbf{b} \end{cases}$$

- w is a Lyndon word if w is primitive, and is the smallest word in its necklace

LYNDON WORDS

- Lexicographic Order

$\mathbf{a} \prec \mathbf{b}$

$$\text{if} \quad \begin{cases} \text{either } \exists\, p, \alpha, \beta \in \mathcal{A}^\star, a_i, a_j \in \mathcal{A} \text{ s.t. } \; i < j, \; \begin{cases} \mathbf{a} = p a_i \alpha, \\ \mathbf{b} = p a_j \beta, \end{cases} \\ \\ \text{or } \mathbf{a} \text{ is a prefix of } \mathbf{b} \end{cases}$$

- w is a Lyndon word if w is primitive, and is the smallest word in its necklace

- cbaa, baac, aacb, acba: aacb is a Lyndon word,

LYNDON WORDS

- Lexicographic Order

$\mathbf{a} \prec \mathbf{b}$

$$\text{if} \quad \begin{cases} \text{either } \exists\, p, \alpha, \beta \in \mathcal{A}^\star, a_i, a_j \in \mathcal{A} \text{ s.t. } \; i < j, \; \begin{cases} \mathbf{a} = p a_i \alpha, \\ \mathbf{b} = p a_j \beta, \end{cases} \\ \\ \text{or } \mathbf{a} \text{ is a prefix of } \mathbf{b} \end{cases}$$

- w is a Lyndon word if w is primitive, and is the smallest word in its necklace

  - cbaa, baac, aacb, acba: aacb is a Lyndon word,
  - aabaab, baac are not

FACTORIZATIONS

- The standard right factor v of a word w is its smallest proper suffix.

FACTORIZATIONS

- The standard right factor v of a word w is its smallest proper suffix.
- The related factorization $w=uv$ is often called the standard factorization of w.

FACTORIZATIONS

- The standard right factor v of a word w is its smallest proper suffix.
- The related factorization w=uv is often called the standard factorization of w.
  - w=abaabbabaabb u=abaabbab v=aabb

FACTORIZATIONS

- The standard right factor v of a word w is its smallest proper suffix.
- The related factorization w=uv is often called the standard factorization of w.
  - w=abaabbabaabb   u=abaabbab   v=aabb
  - w=abaabbabaabb   u'=ab v'=aabbabaabb   v<v'

FACTORIZATIONS

- The standard right factor v of a word w is its smallest proper suffix.
- The related factorization $w=uv$ is often called the standard factorization of w.
  - $w=abaabbabaabb$ $u=abaabbab$ $v=aabb$
  - $w=abaabbabaabb$ $u'=ab$ $v'=aabbabaabb$ $v<v'$
- Theorem (Lyndon, 1954) Any word w may be written uniquely as a non-increasing product of Lyndon words (by iteration of a variant of the standard factorization).

FACTORIZATIONS

- The standard right factor v of a word w is its smallest proper suffix.
- The related factorization $w=uv$ is often called the standard factorization of w.
  - $w=abaabbabaabb$ $u=abaabbab$ $v=aabb$
  - $w=abaabbabaabb$ $u'=ab$ $v'=aabbabaabb$ $v<v'$
- Theorem (Lyndon, 1954) Any word w may be written uniquely as a non-increasing product of Lyndon words (by iteration of a variant of the standard factorization).

$$aabbaaababbaaabaa$$

FACTORIZATIONS

- The standard right factor v of a word w is its smallest proper suffix.
- The related factorization w=uv is often called the standard factorization of w.
  - w=abaabbabaabb u=abaabbab v=aabb
  - w=abaabbabaabb u'=ab v'=aabbabaabb v<v'
- Theorem (Lyndon, 1954) Any word w may be written uniquely as a non-increasing product of Lyndon words (by iteration of a variant of the standard factorization).

$$aabbaaababbaaabaa$$

FACTORIZATIONS

- The standard right factor v of a word w is its smallest proper suffix.
- The related factorization $w=uv$ is often called the standard factorization of w.
  - $w=abaabbabaabb$ $u=abaabbab$ $v=aabb$
  - $w=abaabbabaabb$ $u'=ab$ $v'=aabbabaabb$ $v<v'$
- Theorem (Lyndon, 1954) Any word w may be written uniquely as a non-increasing product of Lyndon words (by iteration of a variant of the standard factorization).

$$aabbaaababbaaabaa$$

FACTORIZATIONS

- The standard right factor v of a word w is its smallest proper suffix.
- The related factorization w=uv is often called the standard factorization of w.
  - w=abaabbabaabb u=abaabbab v=aabb
  - w=abaabbabaabb u′=ab v′=aabbabaabb v<v′
- Theorem (Lyndon, 1954) Any word w may be written uniquely as a non-increasing product of Lyndon words (by iteration of a variant of the standard factorization).

$$aabbaaababbaaabaa$$

FACTORIZATIONS

- The standard right factor v of a word w is its smallest proper suffix.
- The related factorization w=uv is often called the standard factorization of w.
  - w=abaabbabaabb u=abaabbab v=aabb
  - w=abaabbabaabb u'=ab v'=aabbabaabb v<v'
- Theorem (Lyndon, 1954) Any word w may be written uniquely as a non-increasing product of Lyndon words (by iteration of a variant of the standard factorization).

$$aabbaaababbaaabaa$$

FACTORIZATIONS

- The standard right factor v of a word w is its smallest proper suffix.
- The related factorization $w=uv$ is often called the standard factorization of w.
  - $w=abaabbabaabb$ $u=abaabbab$ $v=aabb$
  - $w=abaabbabaabb$ $u'=ab$ $v'=aabbabaabb$ $v<v'$
- **Theorem** (Lyndon, 1954) Any word w may be written uniquely as a non-increasing product of Lyndon words (by iteration of a variant of the standard factorization).

$$aabbaaababbaaabaa$$

FACTORIZATIONS

- The standard right factor v of a word w is its smallest proper suffix.
- The related factorization w=uv is often called the standard factorization of w.
  - w=abaabbabaabb u=abaabbab v=aabb
  - w=abaabbabaabb u'=ab v'=aabbabaabb v<v'
- **Theorem** (Lyndon, 1954) Any word w may be written uniquely as a non-increasing product of Lyndon words (by iteration of a variant of the standard factorization).

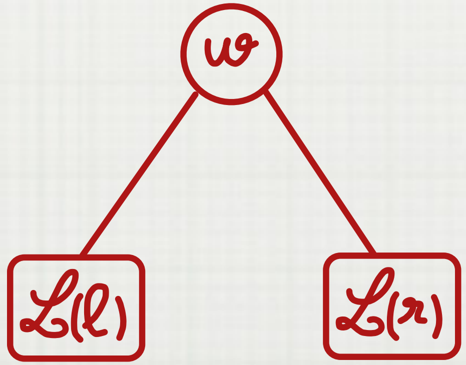
$$aabbaaababbaaabaa$$

- The standard factorization of a Lyndon word is the first step in the construction of some basis of the free Lie algebra over A

Lyndon Tree

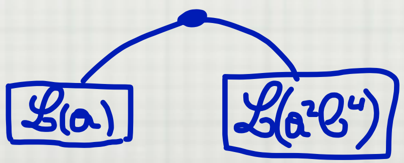
* If $w = \ell r$ is the standard decomposition of w, then the Lyndon tree of w, $\mathcal{L}(w)$, is given by

$\mathcal{L}(w) =$



* Examples

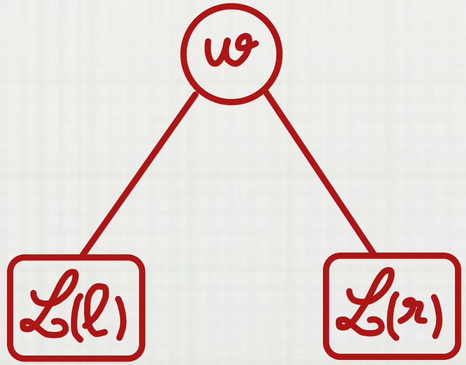
$\mathcal{L}(a^3b^4) =$



Lyndon tree

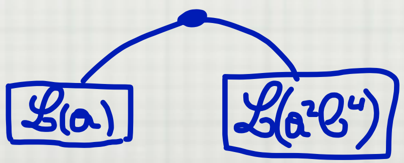
* If $w = \ell r$ is the standard decomposition of w, then the Lyndon tree of w, $\mathcal{L}(w)$, is given by

$\mathcal{L}(w) =$

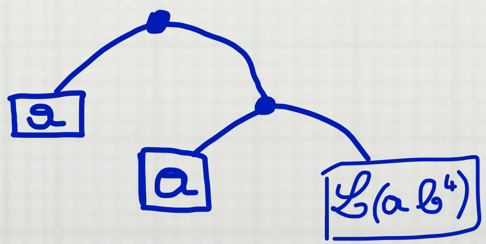


* Examples

$\mathcal{L}(a^3b^4) =$



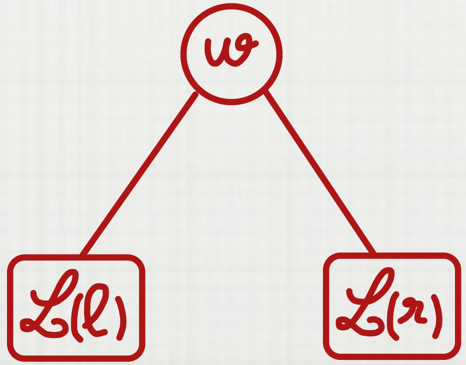
$=$



Lyndon tree

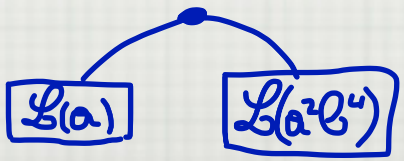
* If $w = \ell r$ is the standard decomposition of w, then the Lyndon tree of w, $\mathcal{L}(w)$, is given by

$\mathcal{L}(w) =$

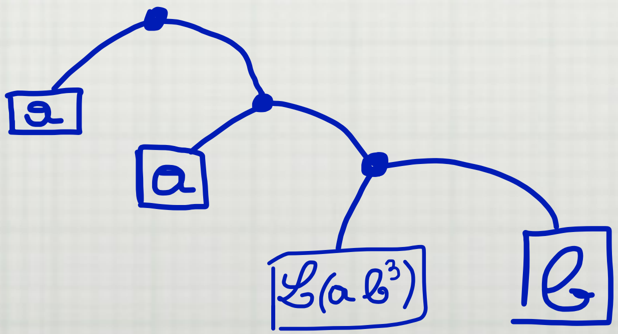


* Examples

$\mathcal{L}(a^3b^4) =$



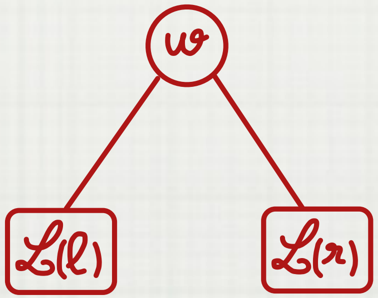
$=$



Lyndon tree

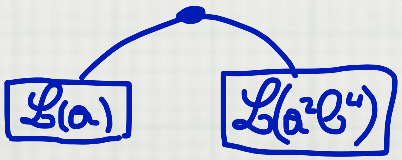
* If $w = \ell r$ is the standard decomposition of w, then the Lyndon tree of w, $\mathcal{L}(w)$, is given by

$\mathcal{L}(w) =$

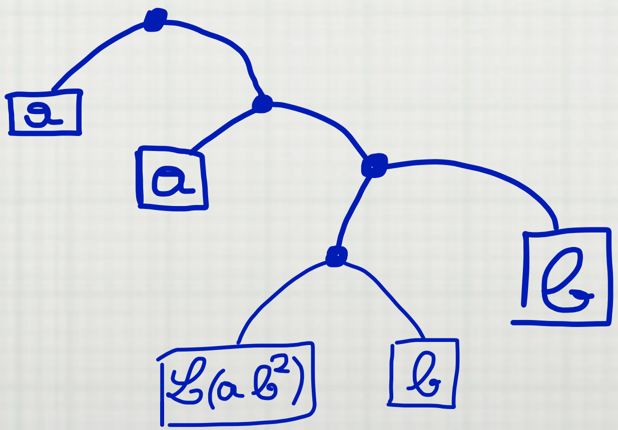


* Examples

$\mathcal{L}(a^3b^4) =$

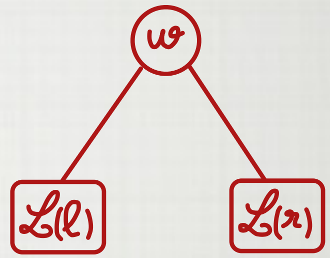


$=$

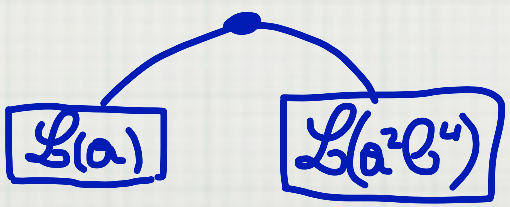


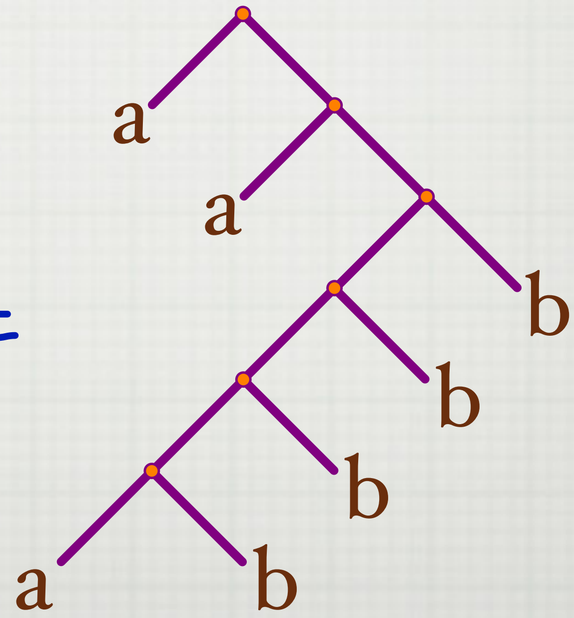
Lyndon tree

- If $w = \ell\, r$ is the standard decomposition of w, then the Lyndon tree of w, $\mathcal{L}(w)$, is given by

$\mathcal{L}(w) =$ 

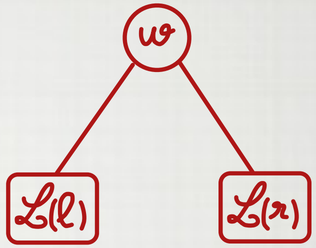
- Examples

$\mathcal{L}(a^3b^4) =$ 

$=$ 

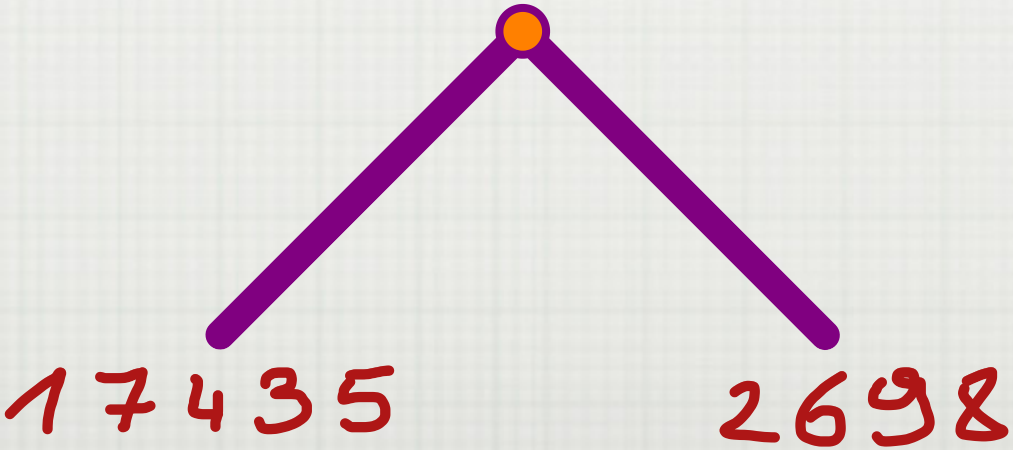
Lyndon tree

* If $w = \ell r$ is the standard decomposition of w, then the Lyndon tree of w, $\mathcal{L}(w)$, is given by $\mathcal{L}(w) =$



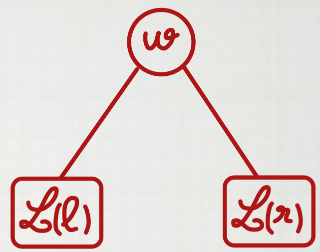
* Examples σ a random permutation

$\sigma = 174352698$



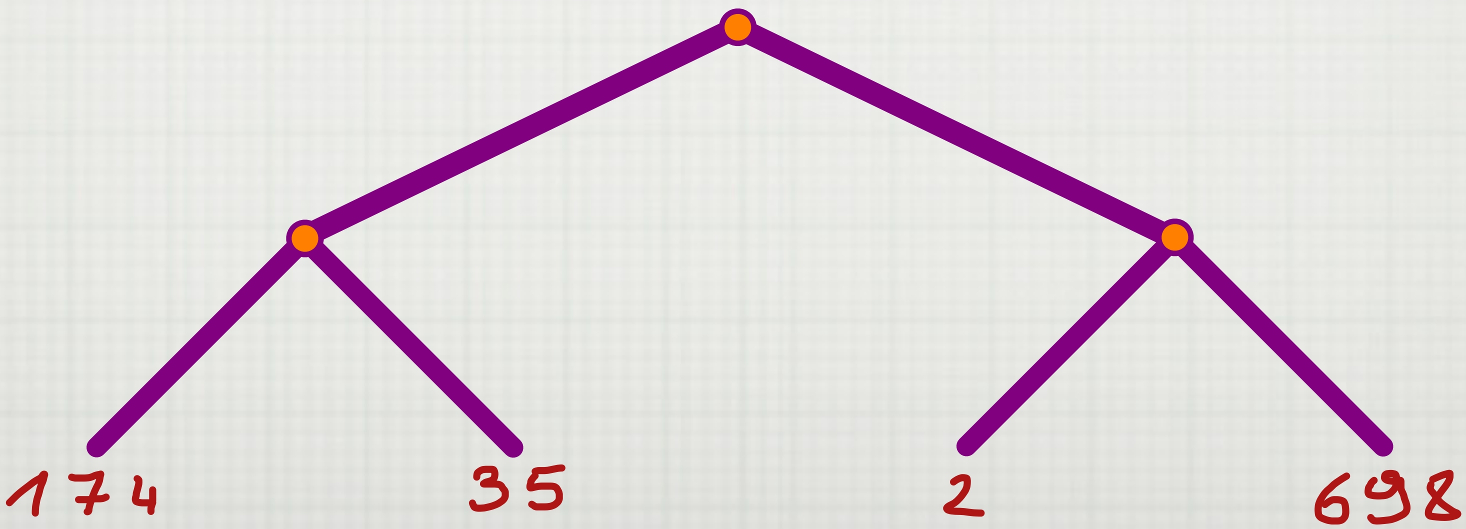
Lyndon tree

* If $w = \ell r$ is the standard decomposition of w, then the Lyndon tree of w, $\mathcal{L}(w)$, is given by $\mathcal{L}(w) =$



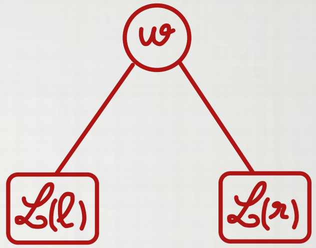
* Examples σ a random permutation

$\sigma = 1\,7\,4\,3\,5\,2\,6\,9\,8$



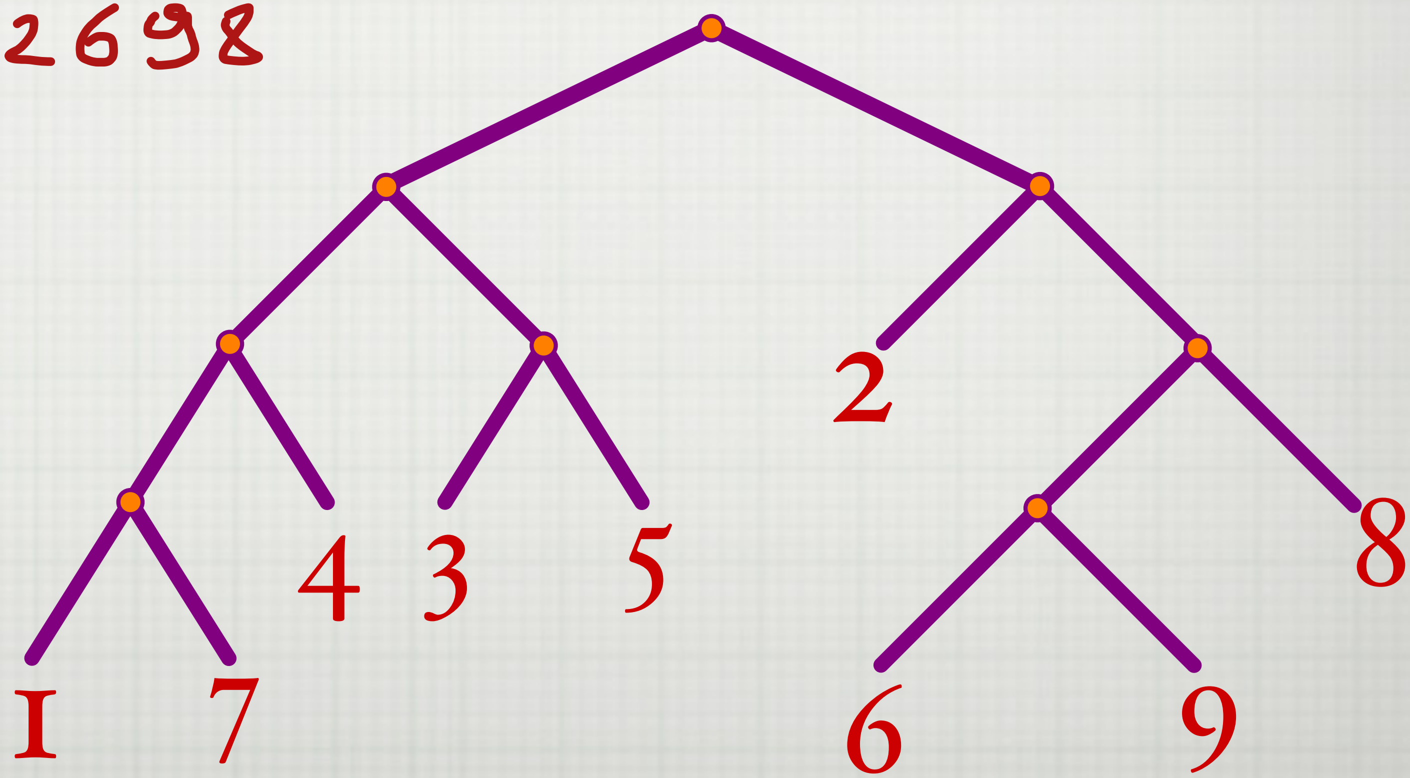
Lyndon tree

* If $w = \ell r$ is the standard decomposition of w, then the Lyndon tree of w, $\mathcal{L}(w)$, is given by $\mathcal{L}(w) =$



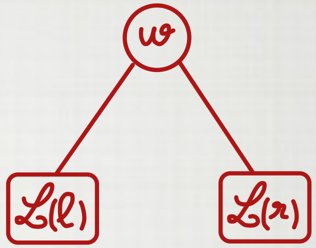
* Examples σ a random permutation

$\sigma = 1\,7\,4\,3\,5\,2\,6\,9\,8$



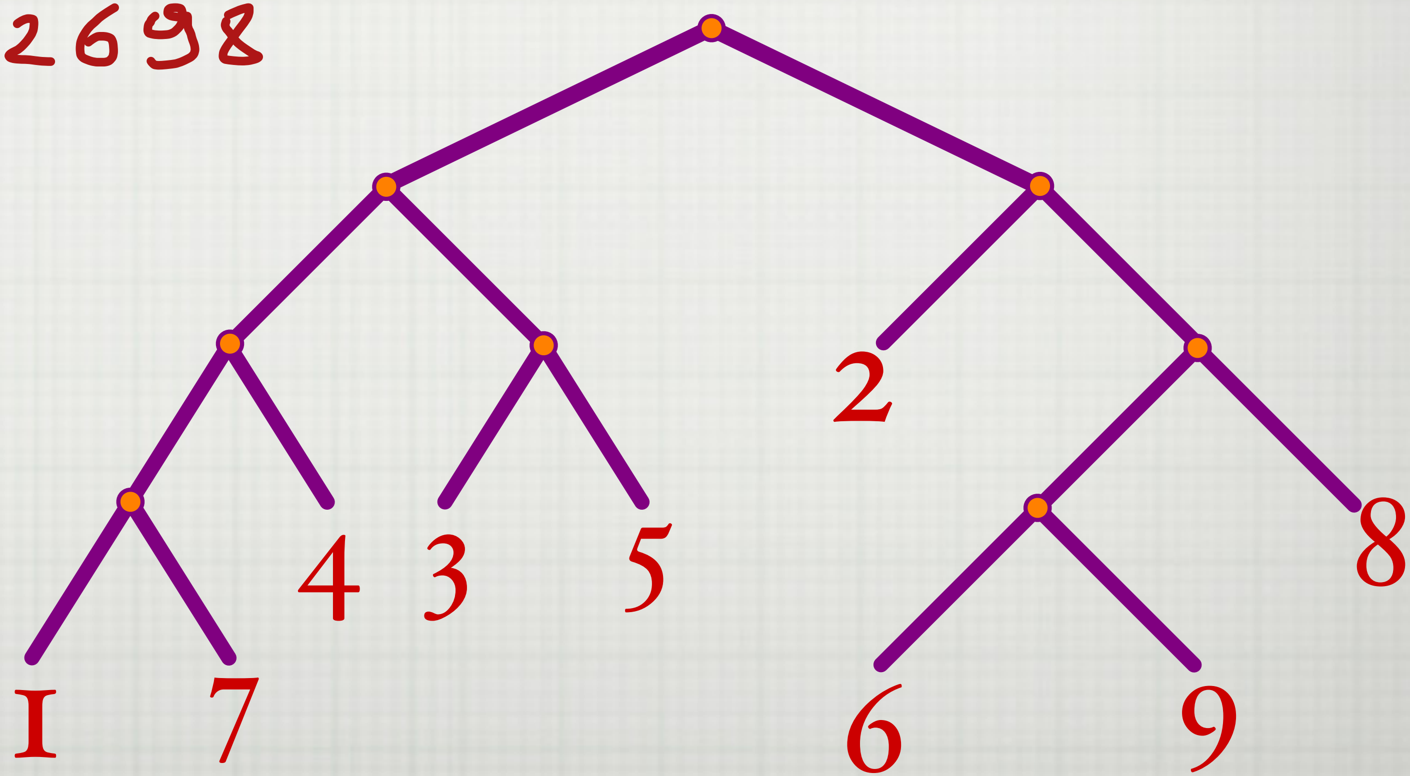
Lyndon tree

* If $w = \ell r$ is the standard decomposition of w, then the Lyndon tree of w, $\mathcal{L}(w)$, is given by $\mathcal{L}(w) =$



* Examples σ a random permutation

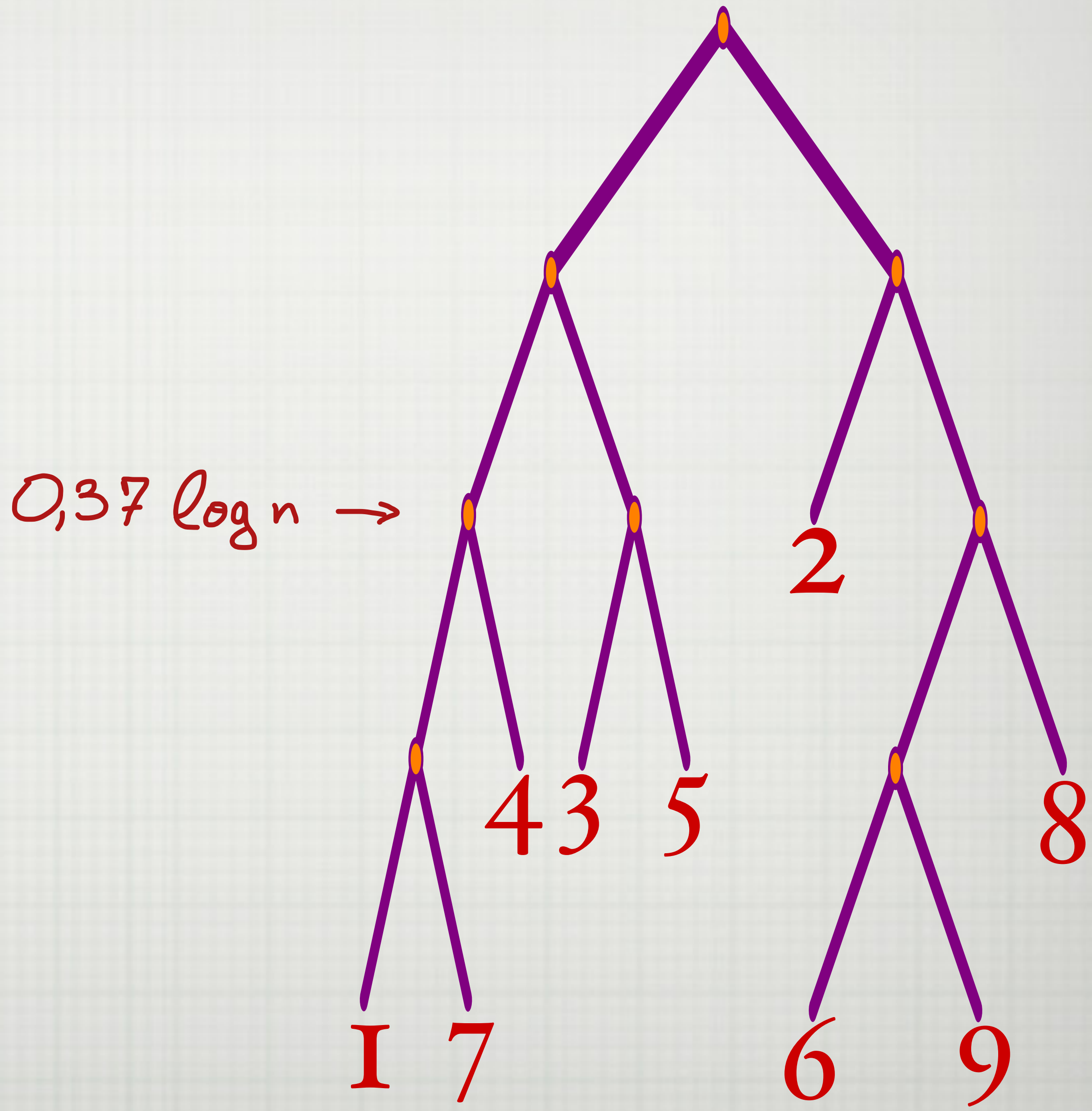
$\sigma = 174352698$



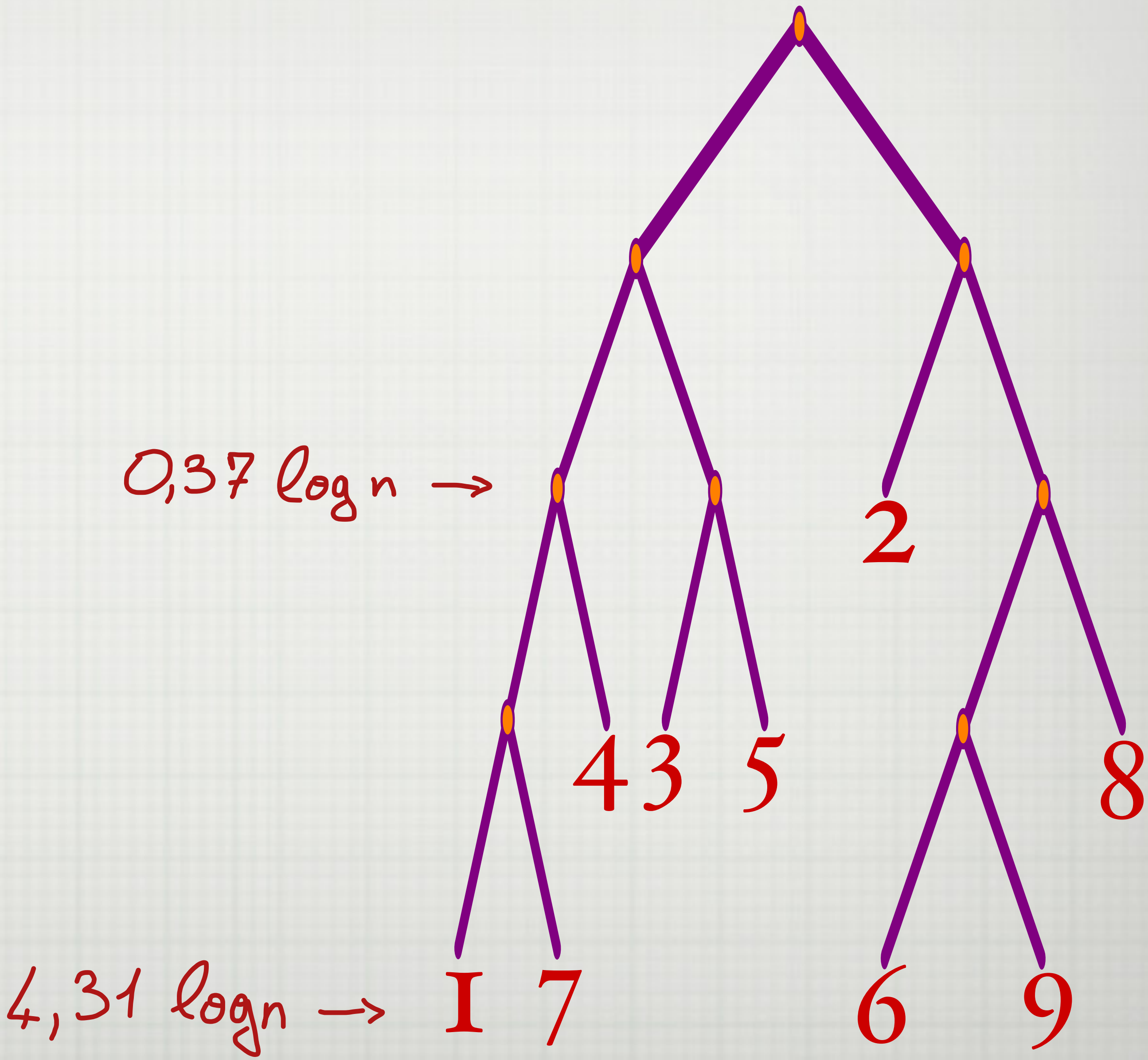
Binary search tree!

(BST)

Binary search tree (BST)



Binary search tree (BST)



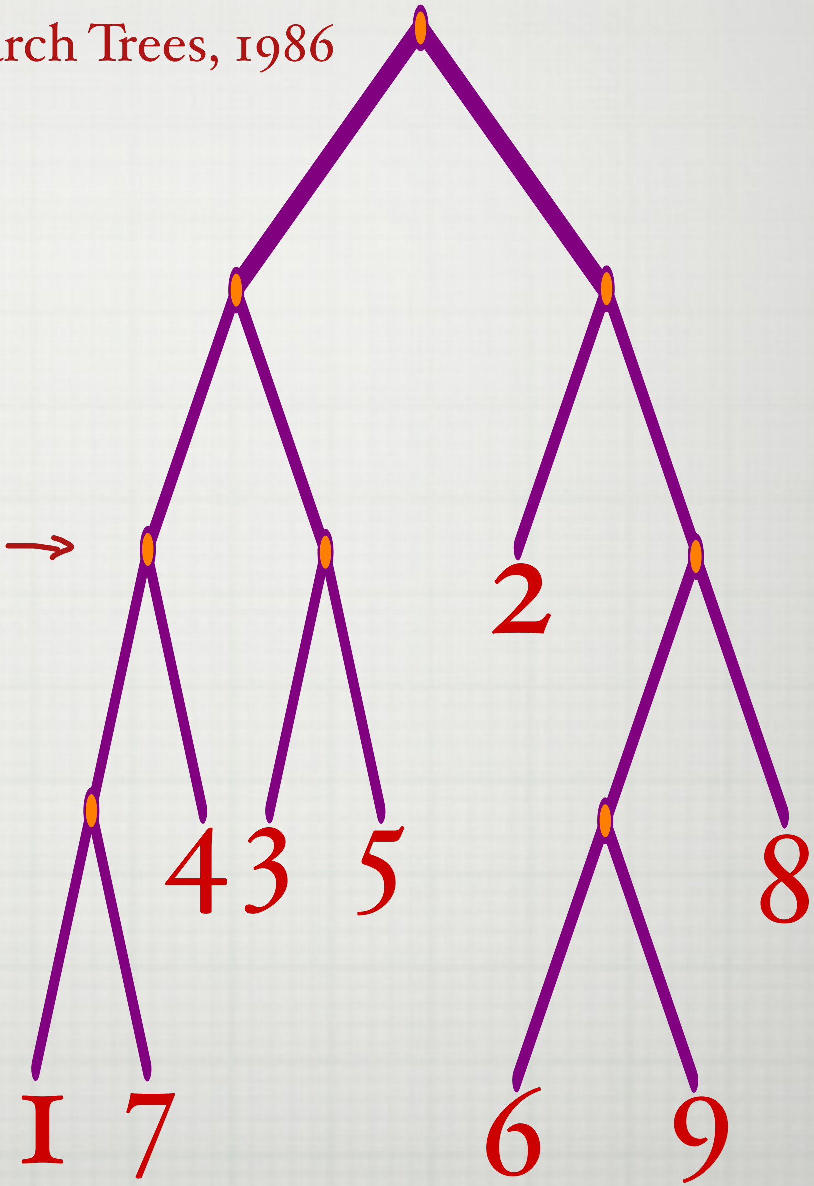
Binary search tree (BST)

LUC DEVROYE, A Note on the Height of Binary Search Trees, 1986

Journal of the Association for Computing Machinery, Vol. 33, No. 3, July 1986, pp. 489-498

$0{,}37 \log n \rightarrow$

$4{,}31 \log n \rightarrow$



Binary search tree (BST)

LUC DEVROYE, A Note on the Height of Binary Search Trees, 1986

Journal of the Association for Computing Machinery, Vol. 33, No. 3, July 1986, pp. 489-498

η_λ: Cramer transform of the Poisson distribution of parameter **λ**

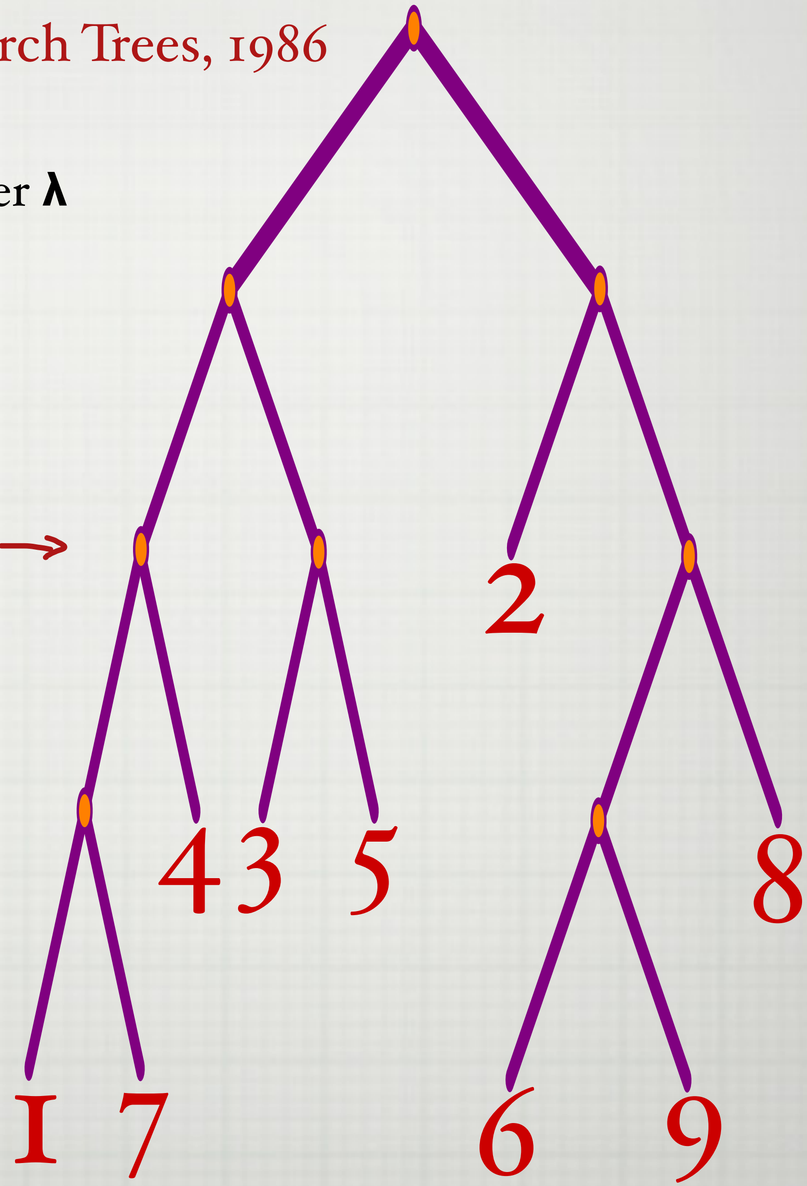
c and c' are the roots of $\eta_2(x) = 1$

$$\eta_\lambda(x) = x \ln\left(\frac{x}{\lambda}\right) - x + \lambda$$

$\underbrace{c' = 0{,}37}\ \log n \rightarrow$

see also Pittel 1984

$\underbrace{c = 4{,}31}\ \log n \rightarrow$



Binary search tree (BST)

LUC DEVROYE, A Note on the Height of Binary Search Trees, 1986

Journal of the Association for Computing Machinery, Vol. 33, No. 3, July 1986, pp. 489-498

B. CHAUVIN, M. DRMOTA, J. JABBOUR-HATTAB,
The Profile of Binary Search Trees, 2001

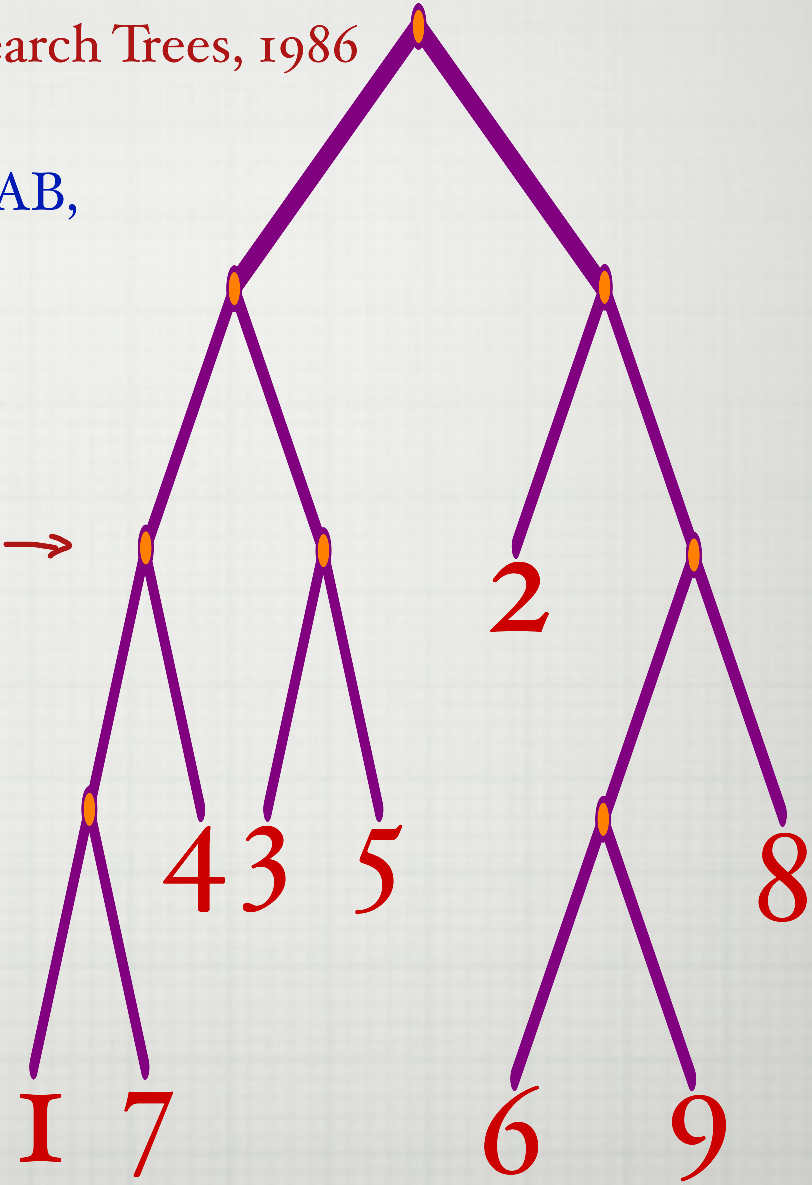
Ann. Appl. Prob., 11:1042–1062.

$0{,}37 \log n \rightarrow$

$2 \log n$ $\longrightarrow$

$+\sqrt{\log n}\,\mathcal{N}(0,\sigma)$

$4{,}31 \log n \rightarrow$



Binary search tree (BST)

LUC DEVROYE, A Note on the Height of Binary Search Trees, 1986

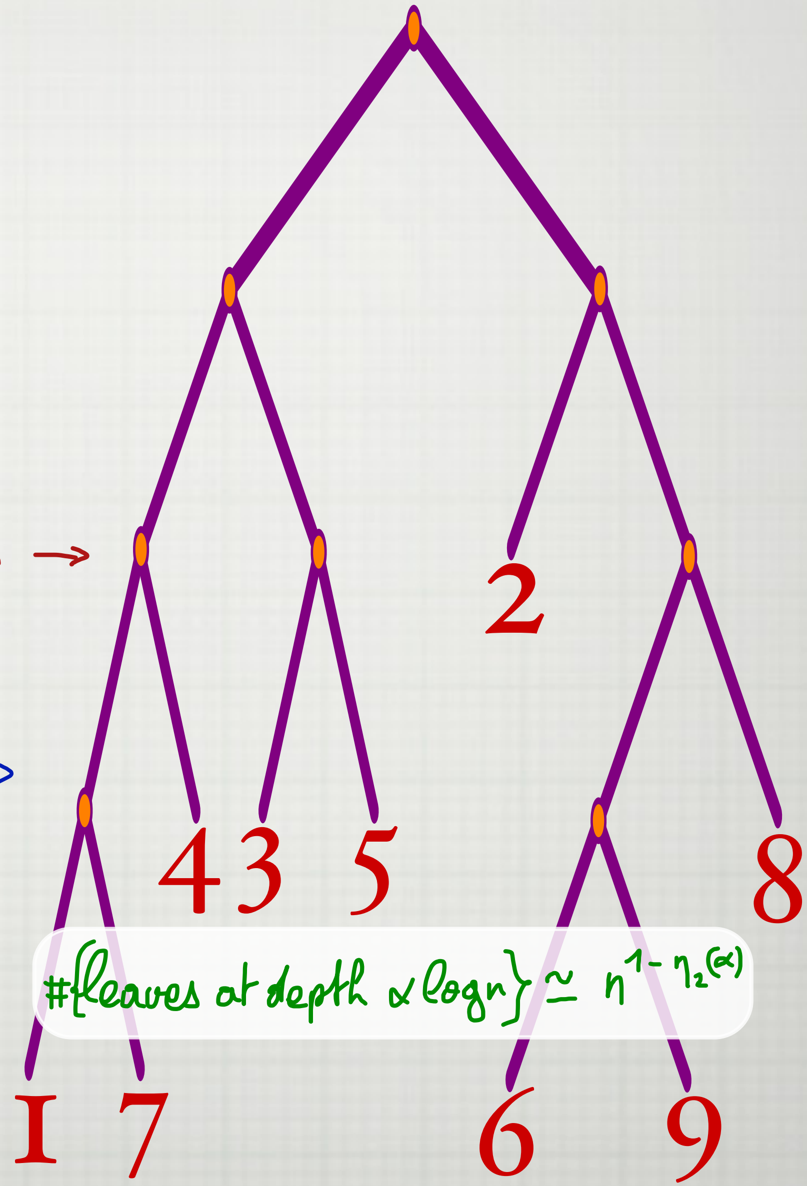
Journal of the Association for Computing Machinery, Vol. 33, No. 3, July 1986, pp. 489-498

B. CHAUVIN, M. DRMOTA, J. JABBOUR-HATTAB,
The Profile of Binary Search Trees, 2001

Ann. Appl. Prob., 11:1042–1062.

J. JABBOUR-HATTAB,
Martingales and large deviations for binary search trees, 2001

Random Structure and Algorithms, 19:112–127.



$0{,}37 \log n \rightarrow$

$2 \log n + \sqrt{\log n}\, \mathcal{N}(0,\sigma) \longrightarrow$

$\alpha \log n \longrightarrow$ $\#\{\text{leaves at depth } \alpha \log n\} \simeq n^{1-\eta_2(\alpha)}$

$4{,}31 \log n \rightarrow$

Binary search tree (BST)

LUC DEVROYE, A Note on the Height of Binary Search Trees, 1986

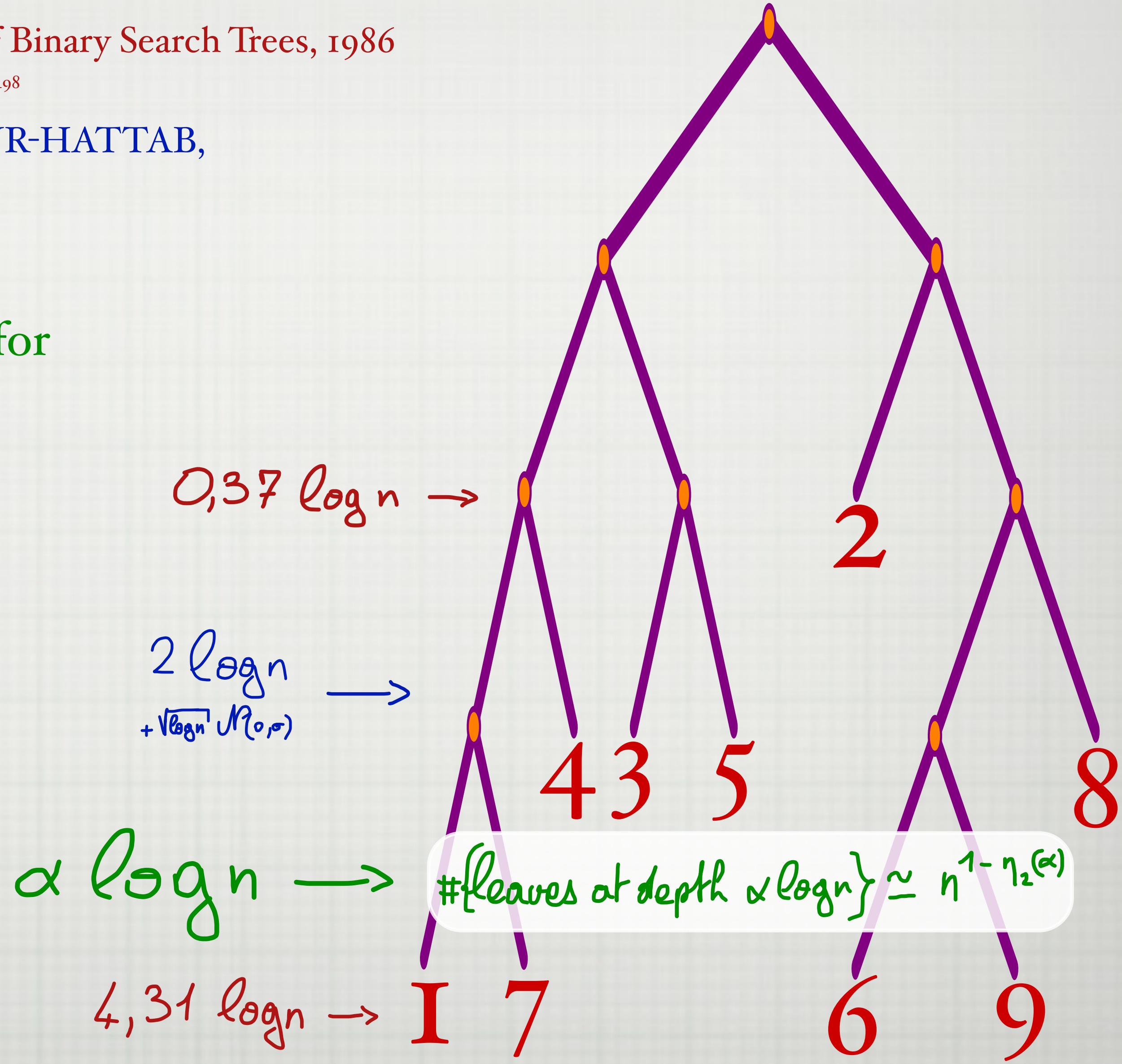
Journal of the Association for Computing Machinery, Vol. 33, No. 3, July 1986, pp. 489-498

B. CHAUVIN, M. DRMOTA, J. JABBOUR-HATTAB,
The Profile of Binary Search Trees, 2001

Ann. Appl. Prob., 11:1042–1062.

J. JABBOUR-HATTAB,
Martingales and large deviations for binary search trees, 2001

Random Structure and Algorithms, 19:112–127.



fine study of the top level
see Matthew Roberts, 2010

Lyndon trees on the alphabet $\{a, b\}$

$\mathcal{L}_n$ set of n letters long Lyndon words, uniform probability distribution $\mathbb{L}_n$

Lyndon trees on the alphabet $\{a, b\}$

$\mathcal{L}_n$ set of n letters long Lyndon words, uniform probability distribution $\mathbb{L}_n$

Thm Under $\mathbb{L}_n(dw)$, $\frac{|\ell|}{n} \xrightarrow{(d)} \frac{1}{2}\delta_0 + \frac{1}{2} U_{[0,1]}$ Marchand Zohoorian

Lyndon trees on the alphabet $\{a, b\}$

$\mathcal{L}_n$ set of n letters long Lyndon words, uniform probability distribution $\mathbb{L}_n$

Thm Under $\mathbb{L}_n(dw)$, $\frac{|\ell|}{n} \xrightarrow{(d)} \frac{1}{2}\delta_0 + \frac{1}{2}U_{[0,1]}$ Marchand Zohoorian

∞^{te} alphabet, Poisson Dirichlet dist^n $\to$ Chassaing Zohoorian

Lyndon trees on the alphabet $\{a, b\}$

$\mathcal{L}_n$ set of n letters long Lyndon words, uniform probability distribution $\mathbb{L}_n$

Thm Under $\mathbb{L}_n(dw)$, $\frac{|\ell|}{n} \xrightarrow{(d)} \frac{1}{2}\delta_0 + \frac{1}{2}\mathcal{U}_{[0,1]}$ Marchand Zohoorian

∞^{te} alphabet, Poisson Dirichlet dist^n $\to$ Chassaing Zohoorian

Idea the positions of the $\log_2 n$ long runs of "a" are uniform on $[\![1, n]\!]$, approximately. They give the sizes of the factors (= subtrees)

Lyndon trees on the alphabet $\{a, b\}$

$\mathcal{L}_n$ set of n letters long Lyndon words, uniform probability distribution $\mathbb{L}_n$

<u>Thm</u> Under $\mathbb{L}_n(dw)$, $\frac{|\ell|}{n} \xrightarrow{(d)} \frac{1}{2}\delta_0 + \frac{1}{2} U_{[0,1]}$ Marchand Zohoorian

∞^{te} alphabet, Poisson Dirichlet distⁿ → Chassaing Zohoorian

Idea the positions of the $\log_2 n$ long runs of "a" are uniform on $[\![1, n]\!]$, approximately they give the sizes of the factors (= subtrees)

<u>Problems</u> * # $\mathcal{L}_n$ and $\mathbb{L}_n$ unfriendly

Lyndon trees on the alphabet $\{a, b\}$

$\mathcal{L}_n$ set of n letters long Lyndon words, uniform probability distribution $\mathbb{L}_n$

<u>Thm</u> Under $\mathbb{L}_n(dw)$, $\frac{|\ell|}{n} \xrightarrow{(d)} \frac{1}{2}\delta_0 + \frac{1}{2} U_{[0,1]}$ Marchand Zohoorian

∞^{te} alphabet, Poisson Dirichlet dist^n $\to$ Chassaing Zohoorian

Idea the positions of the $\log_2 n$ long runs of "a" are uniform on $[\![1, n]\!]$, approximately. They give the sizes of the factors (= subtrees)

<u>Problems</u>

* $\# \mathcal{L}_n$ and $\mathbb{L}_n$ unfriendly
* lexicographic order not easy to handle

Lyndon trees on the alphabet $\{a, b\}$

$\mathcal{L}_n$ set of n letters long Lyndon words, uniform probability distribution $\mathbb{L}_n$

<u>Thm</u> Under $\mathbb{L}_n(dw)$, $\frac{|\ell|}{n} \xrightarrow{(d)} \frac{1}{2}\delta_0 + \frac{1}{2} U_{[0,1]}$ Marchand Zohoorian

∞^{te} alphabet, Poisson Dirichlet distⁿ → Chassaing Zohoorian

Idea the positions of the $\log_2 n$ long runs of "a" are uniform on $[\![1, n]\!]$, approximately they give the sizes of the factors (= subtrees)

<u>Problems</u>

* $\# \mathcal{L}_n$ and $\mathbb{L}_n$ unfriendly
* lexicographic order not easy to handle
* many tied for the title of longest run to break the tie, one has to look at the suffix

Lyndon trees on the alphabet $\{a,b\}$

$\mathcal{L}_n$ set of n letters long Lyndon words, uniform probability distribution $\mathbb{L}_n$

<u>Thm</u> Under $\mathbb{L}_n(dw)$, $\frac{|l|}{n} \xrightarrow{(d)} \frac{1}{2}\delta_0 + \frac{1}{2} U_{[0,1]}$ Marchand Zohoorian

∞^{te} alphabet, Poisson Dirichlet distⁿ → Chassaing Zohoorian

Idea the positions of the $\log_2 n$ long runs of "a" are uniform on $[\![1,n]\!]$, approximately they give the sizes of the factors (= subtrees)

<u>Problems</u>
- $\#\mathcal{L}_n$ and $\mathbb{L}_n$ unfriendly
- lexicographic order not easy to handle
- many tied for the title of longest run to break the tie, one has to look at the suffix

$$\mathbb{L}_n(k \text{ tied}) \simeq \sum_{\ell\in\mathbb{Z}} e^{-2^{\alpha+\ell}} \frac{(2^{\alpha+\ell-1})^k}{k!} \quad \text{if } \{\log_2 n\} \simeq \alpha$$

The height H_n of the Lyndon tree

<u>Thm</u> Under $\mathbb{L}_n$, $\dfrac{H_n}{\log n} \xrightarrow{(P)} 5{,}09$.

The height H_n of the Lyndon tree

Thm Under $\mathbb{L}_n$, $\dfrac{H_n}{\log n} \xrightarrow{(P)} 5{,}09$.

Ideas * to find a BST somewhere

The height H_n of the Lyndon tree

Thm Under $\mathbb{L}_n$, $\dfrac{H_n}{\log n} \xrightarrow{(P)} 5{,}09$.

Ideas

* to find a BST somewhere

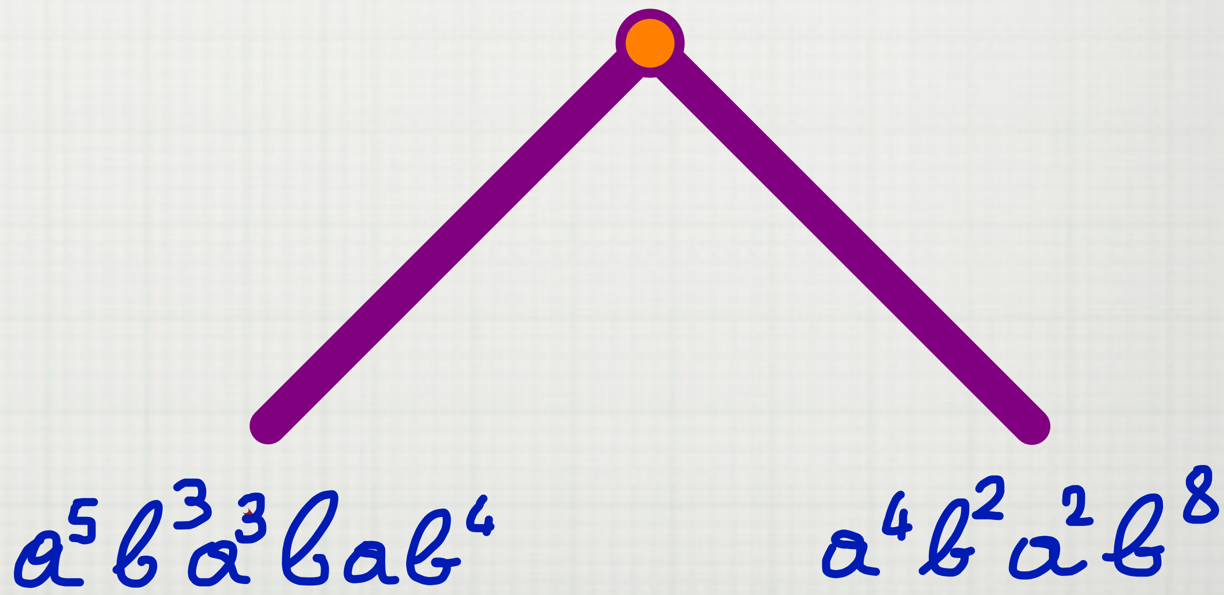
* to use Jablour LD results

The height H_n of the Lyndon tree

<u>Thm</u> Under $\mathbb{L}_n$, $\dfrac{H_n}{\log n} \xrightarrow{(P)} 5{,}09$.

Ideas
* to find a BST somewhere
* to use Jabbour LD results

$w = a^5 b^3 a^3 b a b^4 a^4 b^2 a^2 b^8$

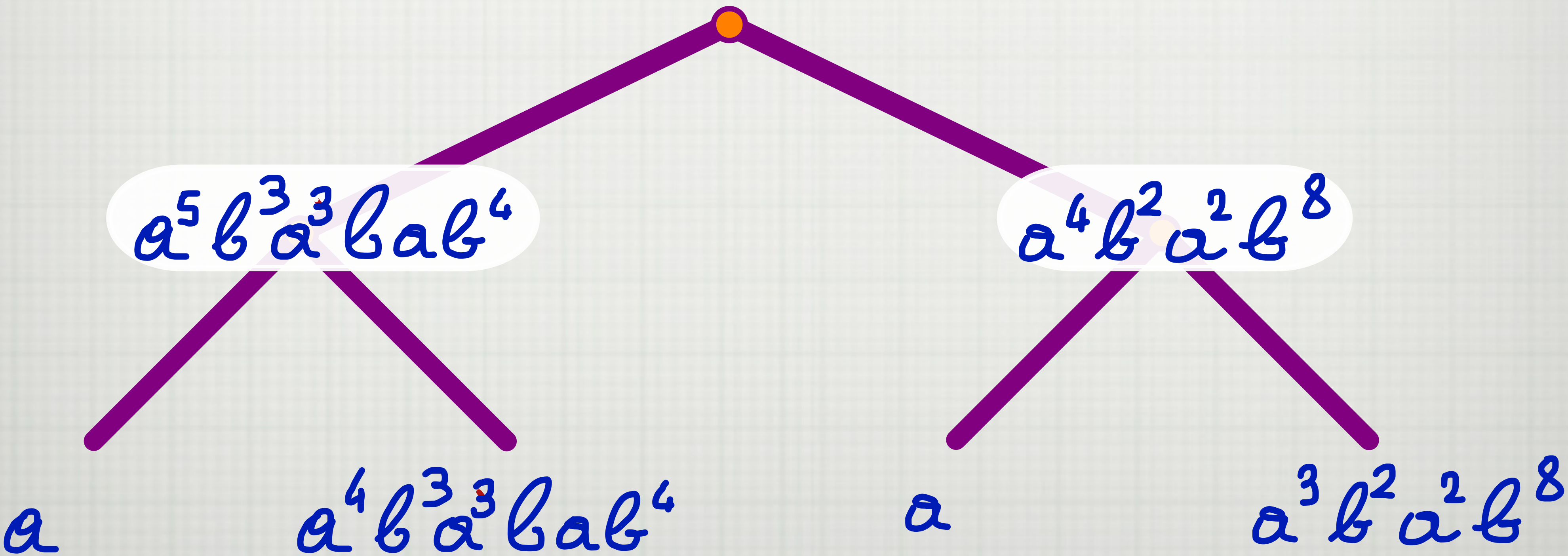


The height H_n of the Lyndon tree

<u>Thm</u> Under $\mathbb{L}_n$, $\dfrac{H_n}{\log n} \xrightarrow{(P)} 5{,}09$.

Ideas
- to find a BST somewhere
- to use Jabbour LD results

$w = a^5 b^3 a^3 b a b^4 a^4 b^2 a^2 b^8$

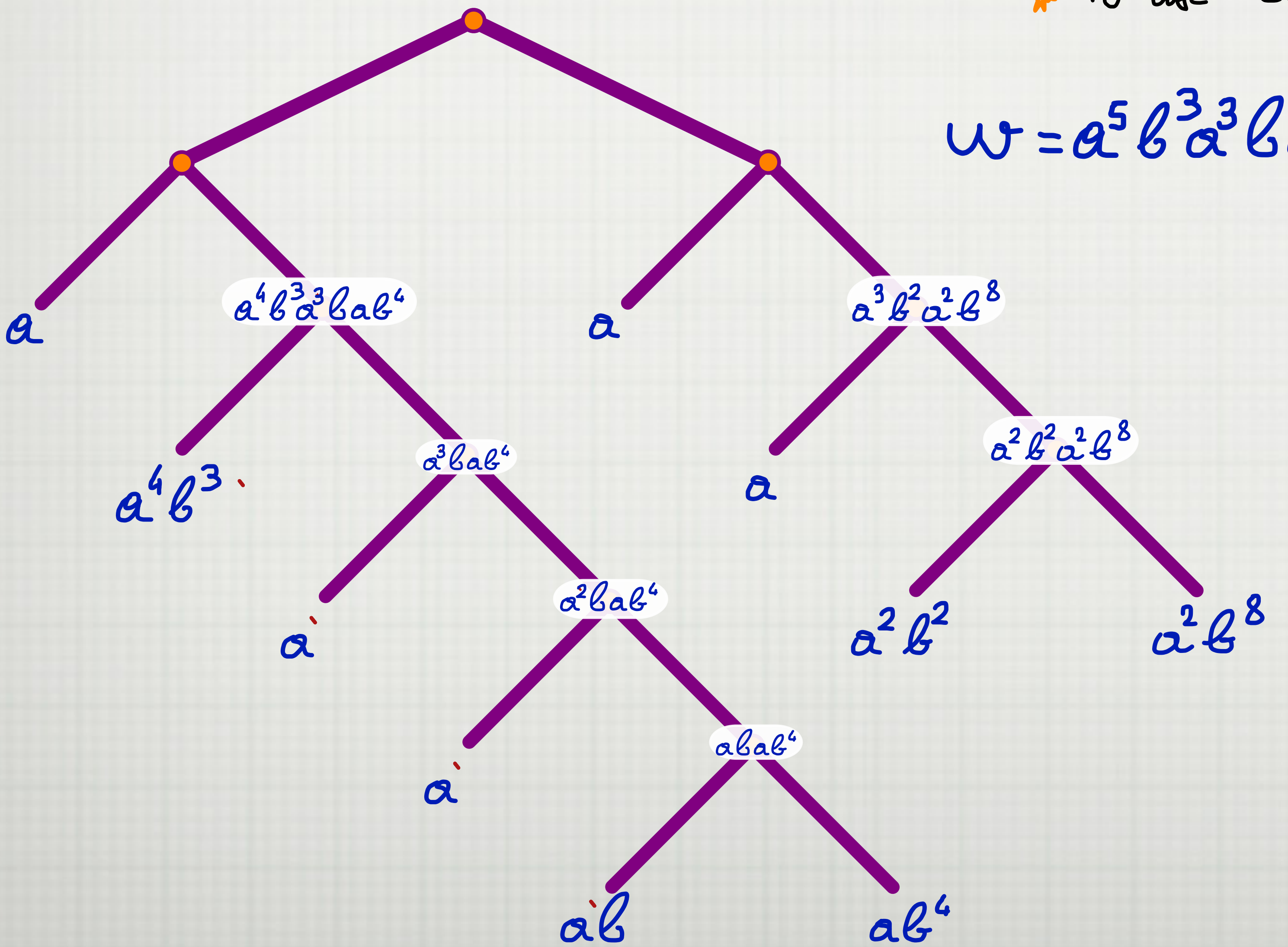


The height H_n of the Lyndon tree

<u>Thm</u> Under $\mathbb{L}_n$, $\dfrac{H_n}{\log n} \xrightarrow{(P)} 5{,}09$.

Ideas
* to find a BST somewhere
* to use Jabbour LD results

$$w = a^5 b^3 a^3 b a b^4 a^4 b^2 a^2 b^8$$

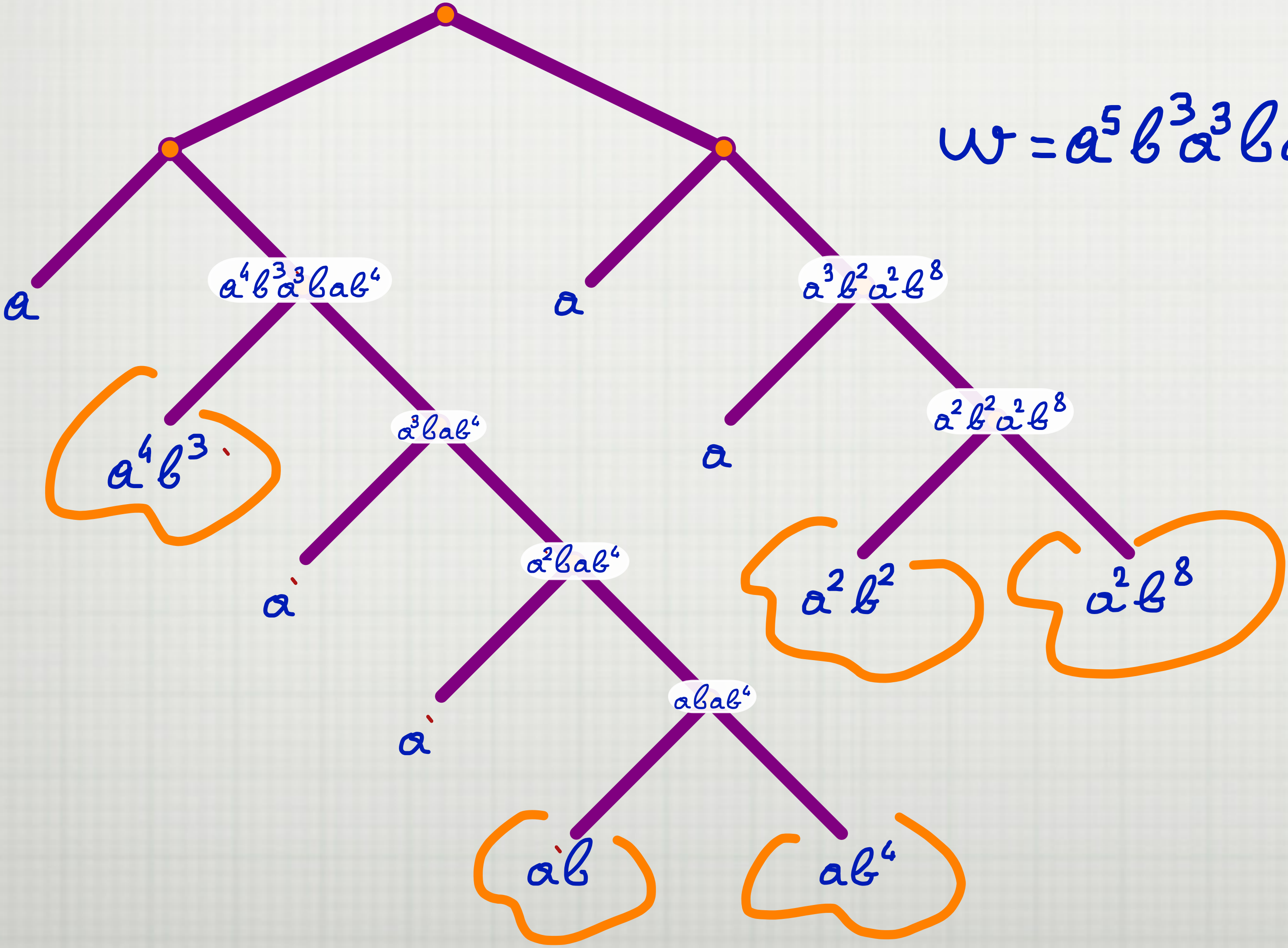


The height H_n of the Lyndon tree

Thm Under $\mathbb{L}_n$, $\dfrac{H_n}{\log n} \xrightarrow{(P)} 5{,}09$.

Ideas
* to find a BST somewhere
* to use Jabbour LD results

$$w = a^5 b^3 a^3 b a b^4 a^4 b^2 a^2 b^8$$

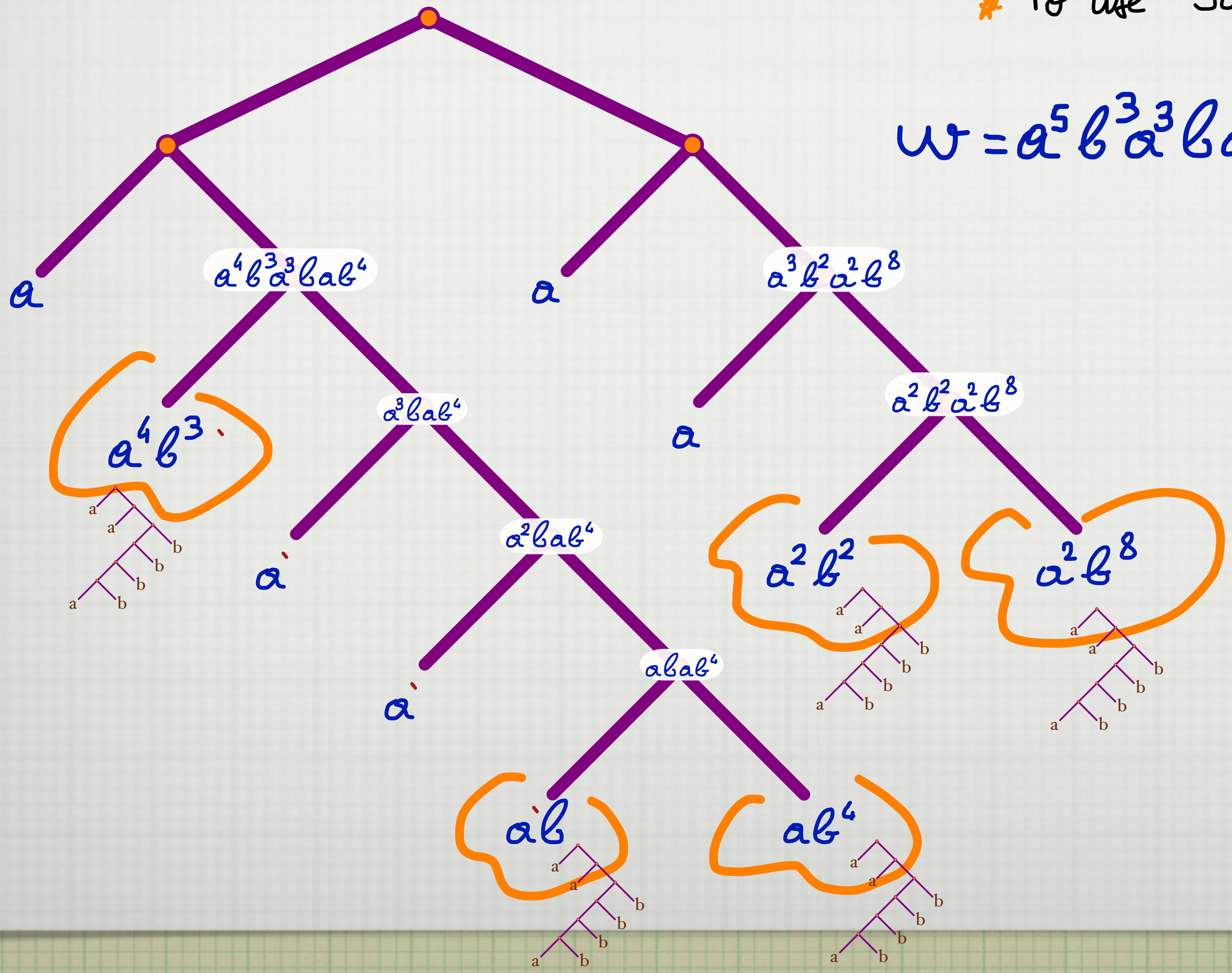


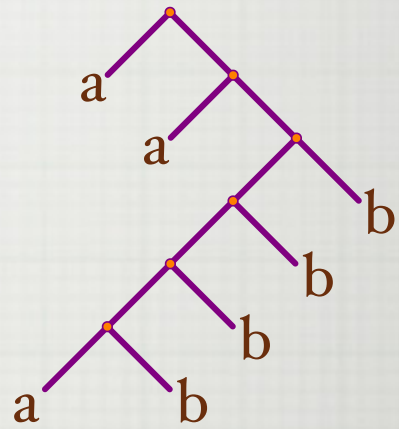
The height H_n of the Lyndon tree

<u>Thm</u> Under $\mathbb{L}_n$, $\dfrac{H_n}{\log n} \xrightarrow{(P)} 5{,}09$.

Ideas
* to find a BST somewhere
* to use Jabbour LD results

$$w = a^5b^3a^3bab^4a^4b^2a^2b^8$$



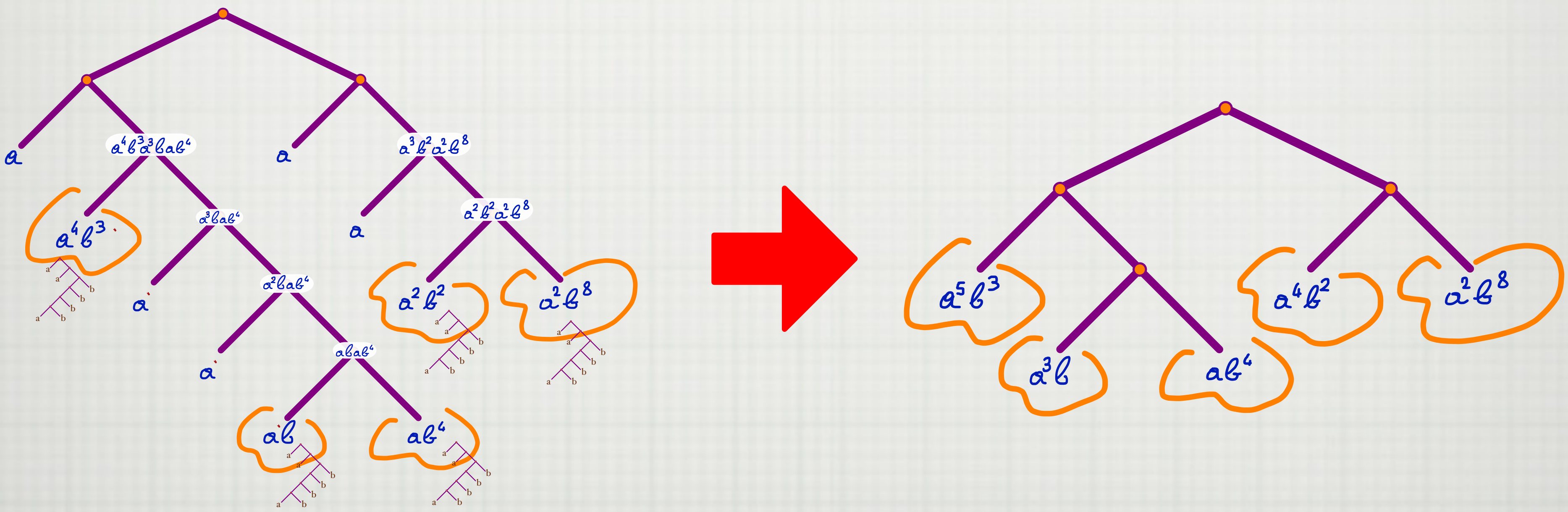


Lyndon tree & BST

Thm Under $\mathbb{L}_n$, $\frac{H_n}{\log n} \xrightarrow{(P)} 5{,}09$. Ideas

* to find a BST somewhere
* to use Jabbour LD results

$$w = a^5 b^3 a^3 b a b^4 a^4 b^2 a^2 b^8$$

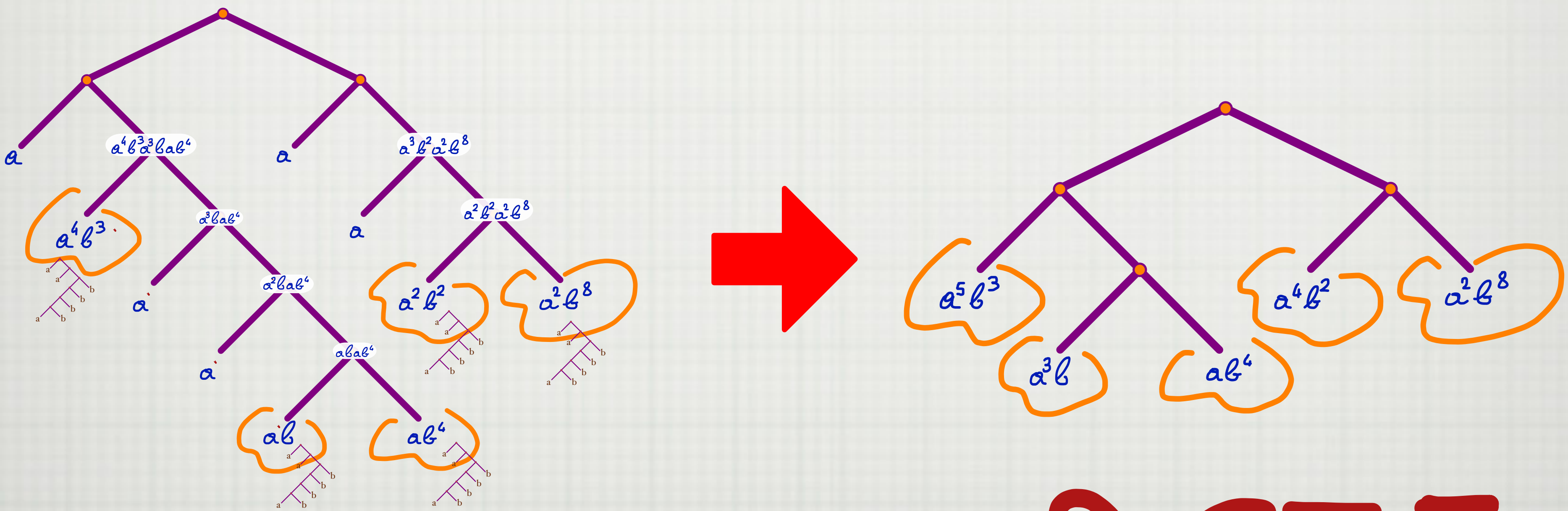


Lyndon tree & BST

Thm Under $\mathbb{L}_n$, $\dfrac{H_n}{\log n} \xrightarrow{(P)} 5{,}09$.

Ideas
* To find a BST somewhere
* To use Jabbour LD results

$$w = a^5 b^3 a^3 b a b^4 a^4 b^2 a^2 b^8$$



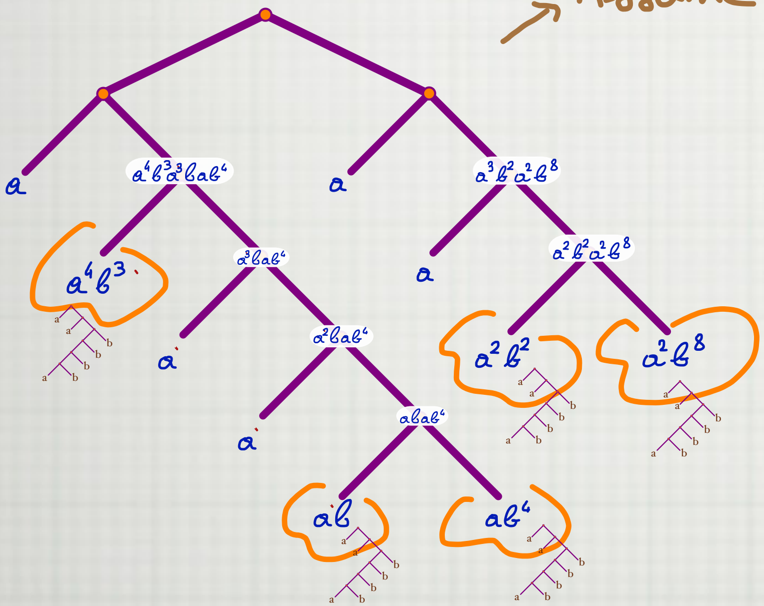
BST!

Lyndon tree & BST

<u>Thm</u> Under $\mathbb{L}_n$, $\dfrac{H_n}{\log n} \xrightarrow{(P)} 5{,}09$.

Ideas
* to find a BST somewhere
* to use Jabbour LD results

$w = a^5 b^3 a^3 b a b^4 a^4 b^2 a^2 b^8$



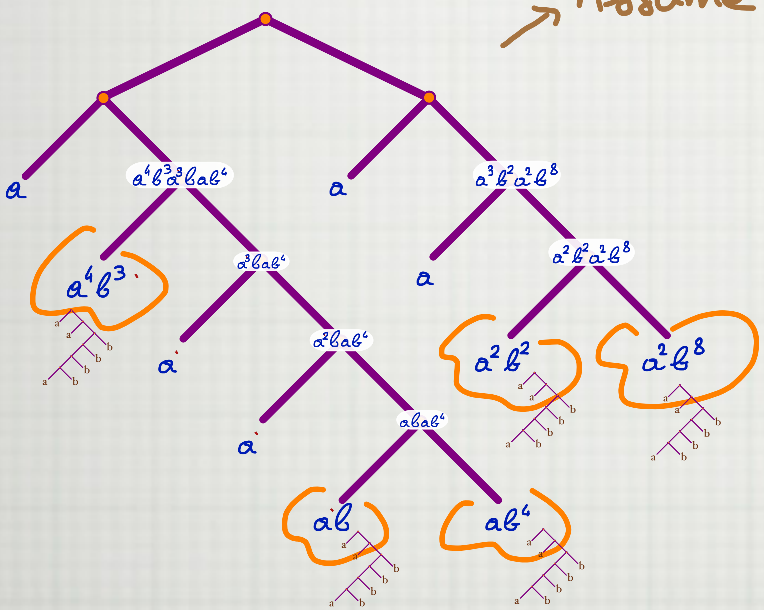
Assume n leaves of type

Lyndon tree & BST

Thm Under $\mathbb{L}_n$, $\dfrac{H_n}{\log n} \xrightarrow{(P)} 5,09$. Ideas

* to find a BST somewhere
* to use Jabbour LD results

$w = a^5b^3a^3bab^4a^4b^2a^2b^8$



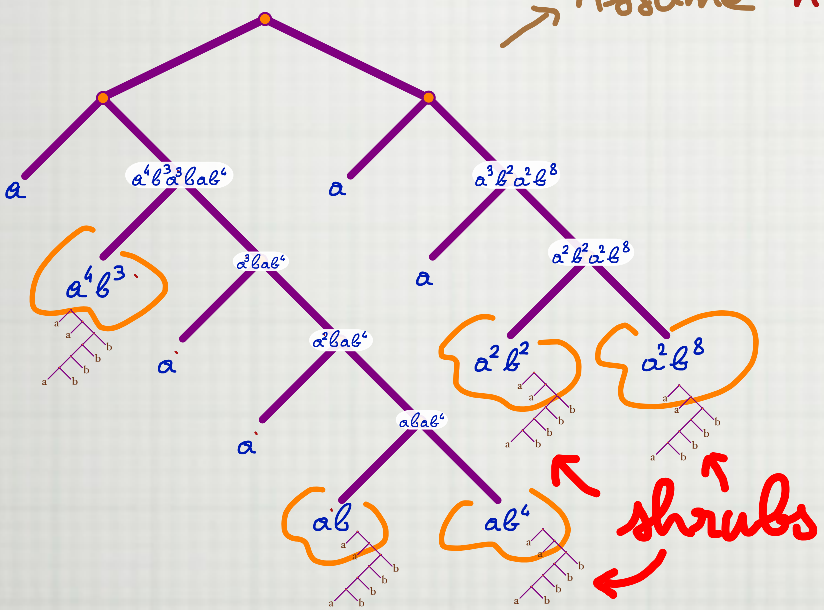
Assume n leaves of type $\{\omega\}$ → Jabbour $n^{1-\eta_2(\alpha)}$ of them at level $\alpha \log n$

Lyndon tree & BST

Thm Under $\mathbb{L}_n$, $\dfrac{H_n}{\log n} \xrightarrow{(P)} 5{,}09$.

Ideas
* to find a BST somewhere
* to use Jabbour LD results

$w = a^5 b^3 a^3 b a b^4 a^4 b^2 a^2 b^8$



Assume n leaves of type {ω} → Jabbour $n^{1-\eta_2(\alpha)}$ of them at level $\alpha \log n$

↓

$n^{1-\eta_2(\alpha)}$ shrubs behaving like i.i.d geometric 1/2

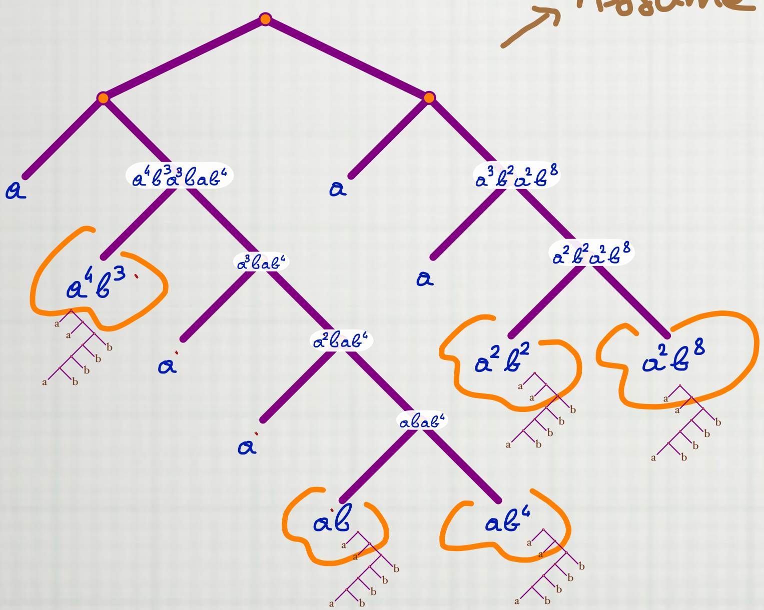
the highest is $(1-\eta_2(\alpha)) \log_2 n$ high

Lyndon tree & BST

Thm Under $\mathbb{L}_n$, $\frac{H_n}{\log n} \xrightarrow{(P)} 5{,}09$.

Ideas
* to find a BST somewhere
* to use Jabbour LD results

$w = a^5 b^3 a^3 b a b^4 a^4 b^2 a^2 b^8$



Assume n leaves of type $\{\omega\}$ $\to$ Jabbour $n^{1-\eta_2(\alpha)}$ of them at level $\alpha \log n$

$\downarrow$

$n^{1-\eta_2(\alpha)}$ shrubs behaving like i.i.d geometric 1/2

the highest is $(1-\eta_2(\alpha)) \log_2 n$ high

$\downarrow$

Contribution to H_n:

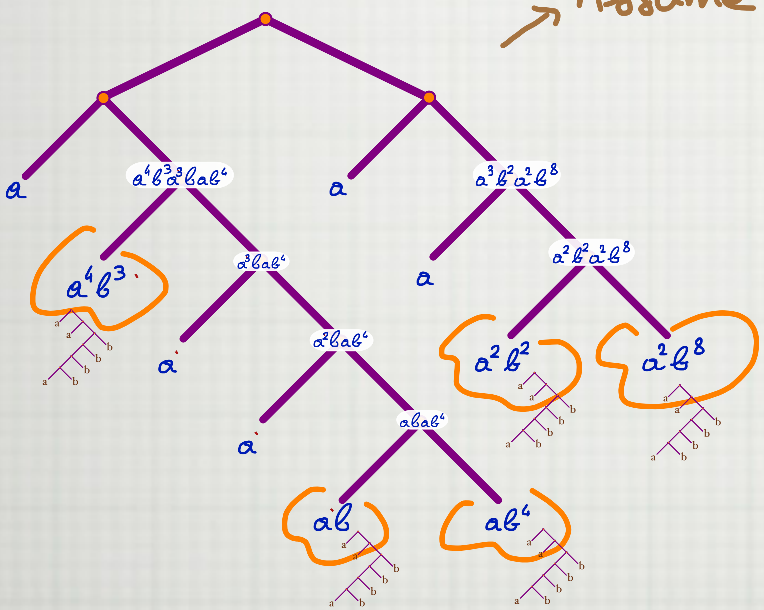
$$\left(\alpha + \frac{1-\eta_2(\alpha)}{\ln 2}\right) \times \log n .$$

Lyndon tree & BST

Thm Under $\mathbb{L}_n$, $\dfrac{H_n}{\log n} \xrightarrow{(P)} 5{,}09$.

Ideas
* to find a BST somewhere
* to use Jabbour LD results

$w = a^5 b^3 a^3 b a b^4 a^4 b^2 a^2 b^8$



Assume n leaves of type {ω} → Jabbour $n^{1-\eta_2(\alpha)}$ of them at level $\alpha \log n$

↓

$n^{1-\eta_2(\alpha)}$ shrubs behaving like i.i.d geometric 1/2

the highest is $(1-\eta_2(\alpha)) \log_2 n$ high

↓

Contribution to H_n:

$$\left(\alpha + \frac{1-\eta_2(\alpha)}{\ln 2}\right) \times \log n .$$

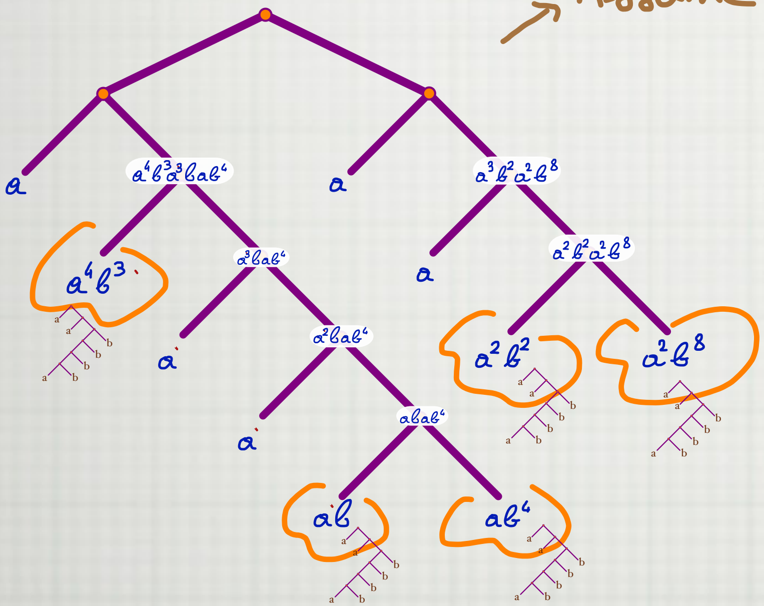
$$\sup_{\alpha} \left\{ \alpha + \frac{1-\eta_2(\alpha)}{\ln 2} \right\} \simeq 5{,}09 \quad ?$$

Lyndon tree & BST

Thm Under $\mathbb{L}_n$, $\frac{H_n}{\log n} \xrightarrow{(P)} 5{,}09$.

Ideas
* to find a BST somewhere
* to use Jabbour LD results

$w = a^5 b^3 a^3 b a b^4 a^4 b^2 a^2 b^8$



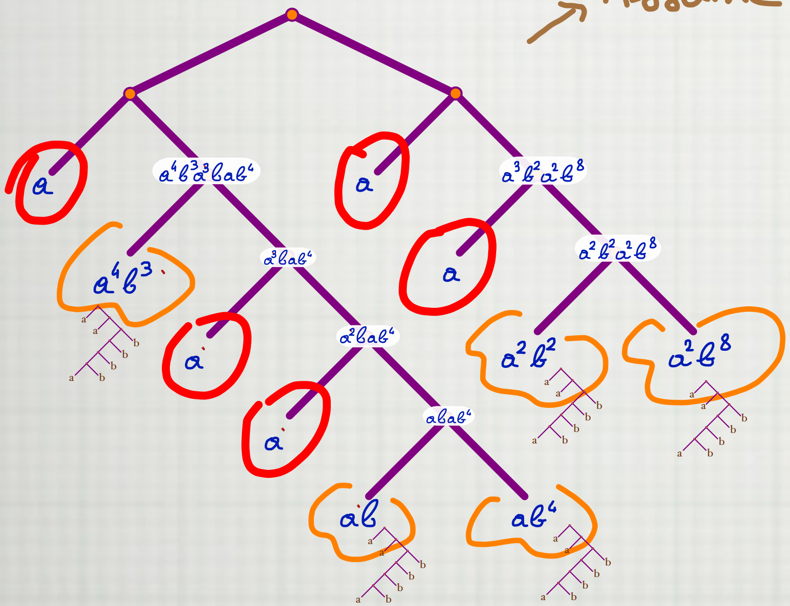
Assume n leaves of type $\{\omega\}$ → ~~Jabbour~~ $n^{1-\eta_2(\alpha)}$ ~~of them at level~~ $\alpha \log n$

Lyndon tree & BST

<u>Thm</u> Under $\mathbb{L}_n$, $\dfrac{H_n}{\log n} \xrightarrow{(P)} 5{,}09$.

Ideas
* to find a BST somewhere
* to use Jabbour LD results

$w = a^5 b^3 a^3 b a b^4 a^4 b^2 a^2 b^8$



Assume n leaves of type $\{w\}$ $\to$ ~~Jabbour~~ $n^{1-\eta_2(\alpha)}$ of them ~~at level~~ $\alpha \log n$

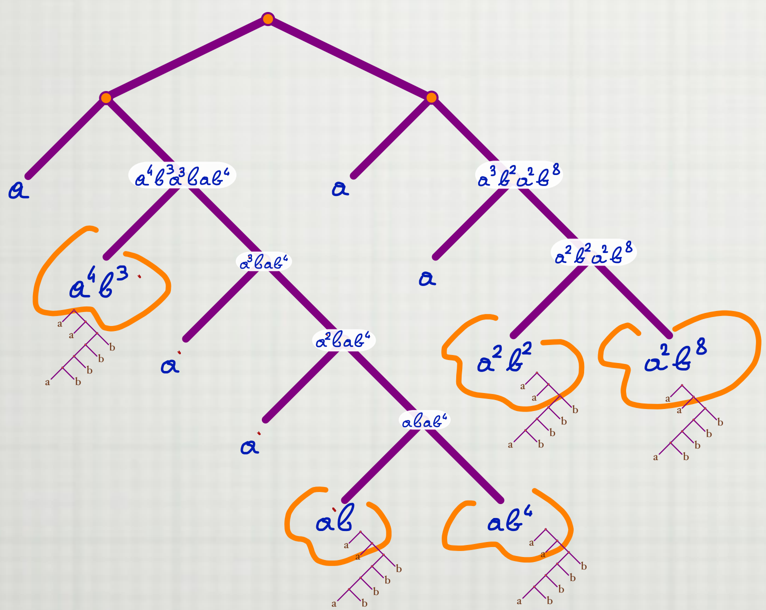
Lyndon tree & BST

Thm Under $\mathbb{L}_n$, $\frac{M_n}{\log n} \xrightarrow{(P)} 5{,}09$.

Ideas
- * to find a BST somewhere
- * to use Jabbour LD results

Problem find factors $a^k b^\ell$

$w = a^5 b^3 a^3 b a b^4 a^4 b^2 a^2 b^8$



Lyndon Tree & BST

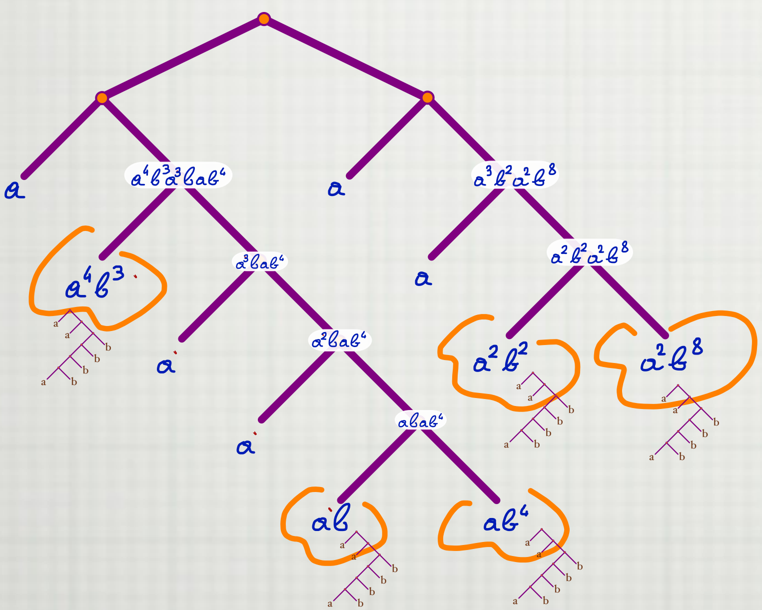
Thm Under $\mathbb{L}_n$, $\dfrac{H_n}{\log n} \xrightarrow{(P)} 5,09$.

Ideas
* to find a BST somewhere
* to use Jabbour LD results

Problem tied factors $a^k b^l$

Solution $a^{k_i} b^{l_i} \to a^{k_i} b^{l_i} w_i$,
in which w_i ∞^{te} iid binary words

$w = a^5 b^3 a^3 b a b^4 a^4 b^2 a^2 b^8$

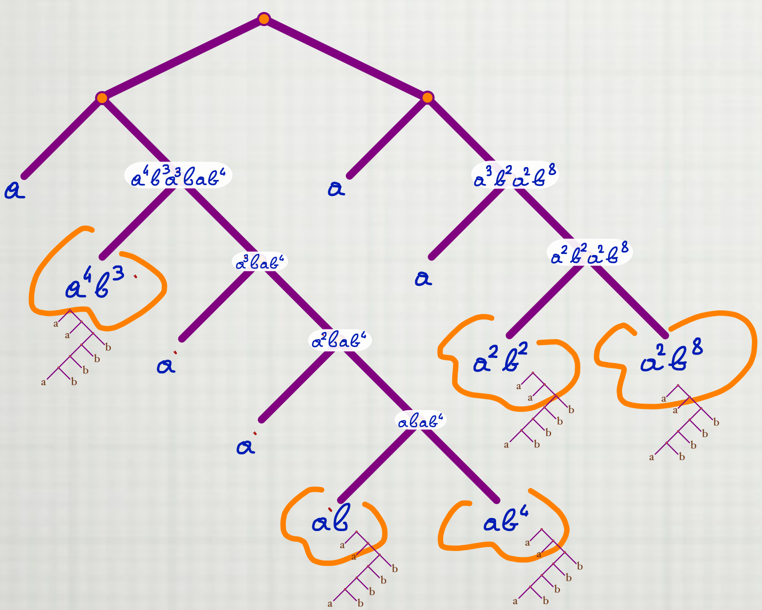


Lyndon tree & BST

<u>Thm</u> Under $\mathbb{L}_n$, $\dfrac{H_n}{\log n} \xrightarrow{(P)} 5{,}09$. Ideas

* to find a BST somewhere
* to use Jabbour LD results

$w = a^5 b^3 a^3 b a b^4 a^4 b^2 a^2 b^8$



Problem tied factors $a^k b^\ell$

Solution $a^{k_i} b^{\ell_i} \to a^{k_i} b^{\ell_i} w_i$,
in which w_i ∞^{te} iid binary words

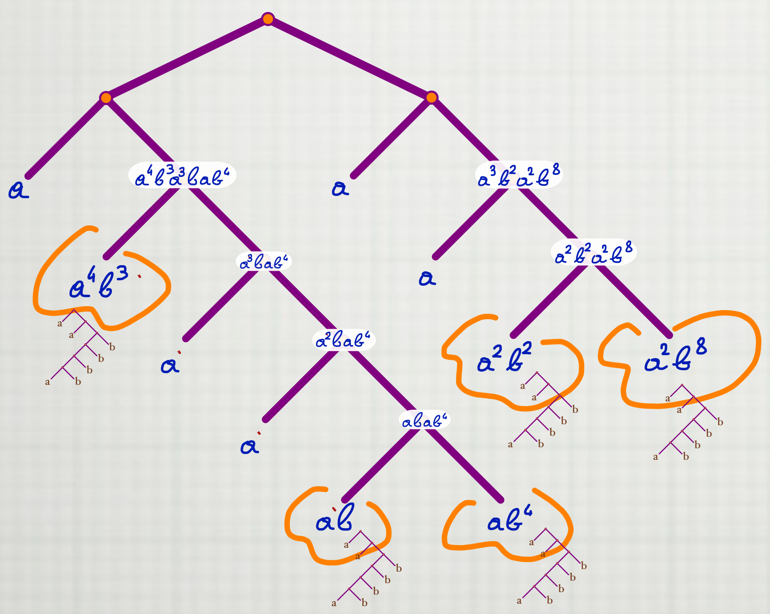
Problem For $w \in \mathcal{L}_n$, the # of factors is not easy to handle

Lyndon tree & BST

Thm Under $\mathbb{L}_n$, $\quad \dfrac{H_n}{\log n} \xrightarrow{(P)} 5{,}09$.

Ideas
- * to find a BST somewhere
- * to use Jabbour LD results

$w = a^5 b^3 a^3 b a b^4 a^4 b^2 a^2 b^8$



Problem tied factors $a^k b^\ell$

Solution $a^{k_i} b^{\ell_i} \to a^{k_i} b^{\ell_i} w_i$,
in which w_i ∞te iid binary words

Problem For $w \in \mathcal{L}_n$, the #
of factors is not easy to handle

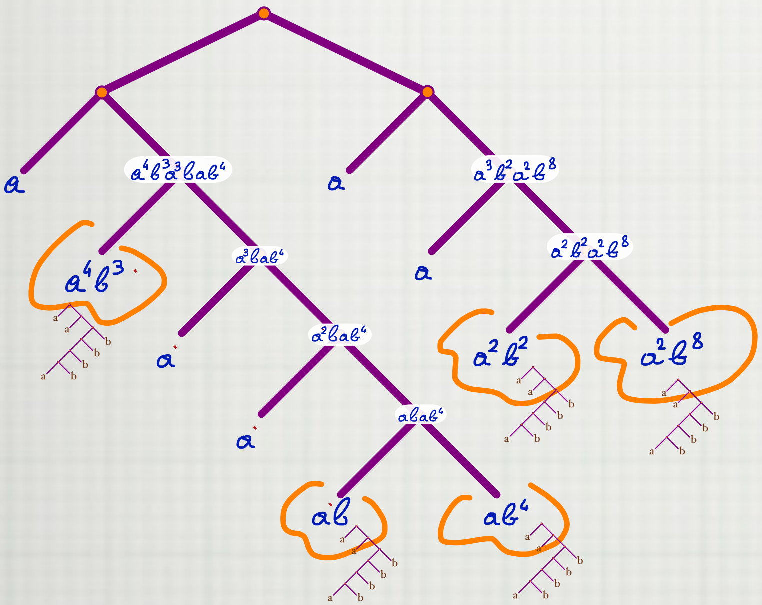
Solution Consider W_ℓ the random
∞te binary word truncated after the
1st occurence of a^ℓ, then reversed
(to be Lyndon)

Lyndon tree & BST

<u>Thm</u> Under $\mathbb{L}_n$, $\frac{H_n}{\log n} \xrightarrow{(P)} 5{,}09$. Ideas

- to find a BST somewhere
- to use Jabbour LD results

$w = a^5 b^3 a^3 b a b^4 a^4 b^2 a^2 b^8$



Problem For $w \in \mathcal{L}_n$, the # of factors is not easy to handle

Solution Consider W_ℓ the random a^∞ binary word truncated after the 1st occurence of a^ℓ, then reversed (to be Lyndon)

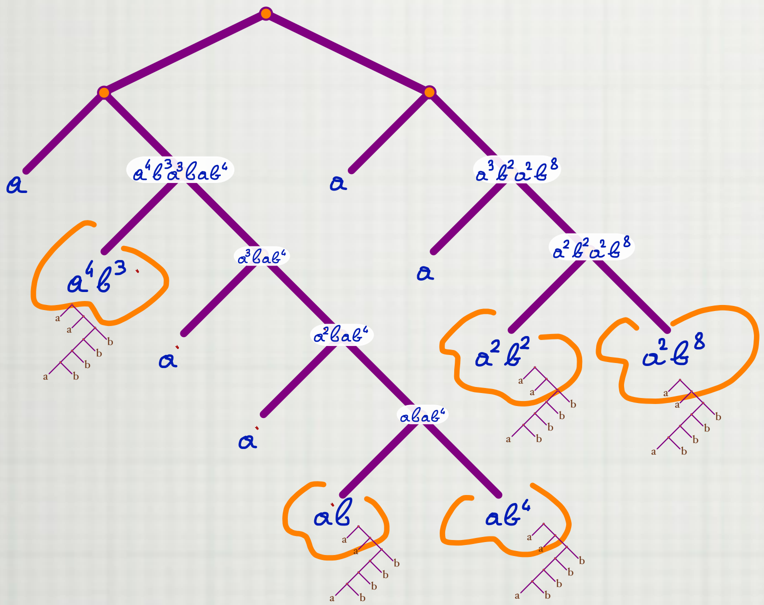
Assumption $|W_\ell| \simeq 2^\ell$ and $H(\mathcal{L}(W_\ell)) \simeq \alpha \ell$

Lyndon tree & BST

<u>Thm</u> Under $\mathbb{L}_n$, $\frac{H_n}{\log n} \xrightarrow{(P)} 5{,}09$.

Ideas
* to find a BST somewhere
* to use Jabbour LD results

$w = a^5 b^3 a^3 b a b^4 a^4 b^2 a^2 b^8$



Problem For $w \in \mathcal{L}_n$, the # of factors is not easy to handle

Solution Consider W_ℓ the random ate binary word truncated after the 1st occurence of a^ℓ, then reversed (to be Lyndon)

Assumption $|W_\ell| \simeq 2^\ell$ and $H(\mathcal{L}(W_\ell)) \simeq \alpha \ell$

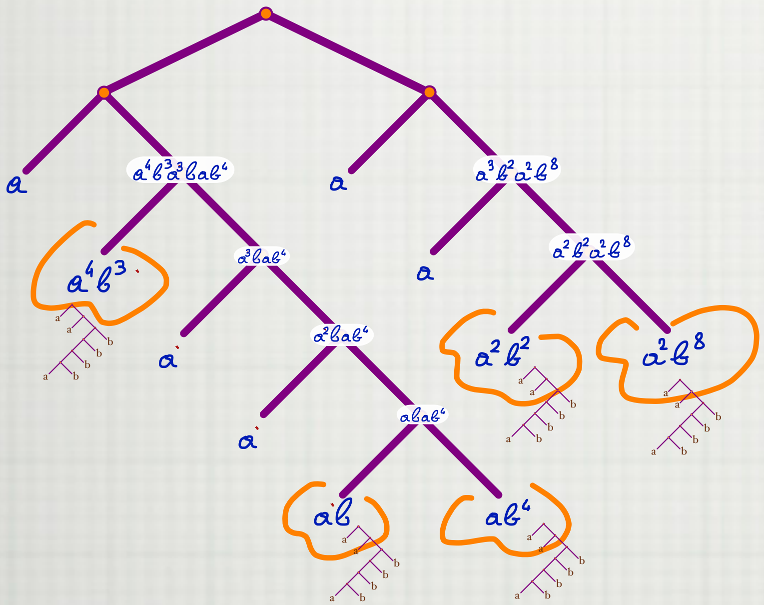
$\Rightarrow H_n \simeq \alpha \log_2 n$

Lyndon tree & BST

Thm Under $\mathbb{L}_n$, $\dfrac{H_n}{\log n} \xrightarrow{(P)} 5,09$.

Ideas
- to find a BST somewhere
- to use Jabbour LD results

$w = a^5 b^3 a^3 b a b^4 a^4 b^2 a^2 b^8$



Problem For $w \in \mathcal{L}_n$, the # of factors is not easy to handle

Solution Consider W_ℓ the random ∞^{te} binary word truncated after the 1^{st} occurence of a^ℓ, then reversed (to be Lyndon)

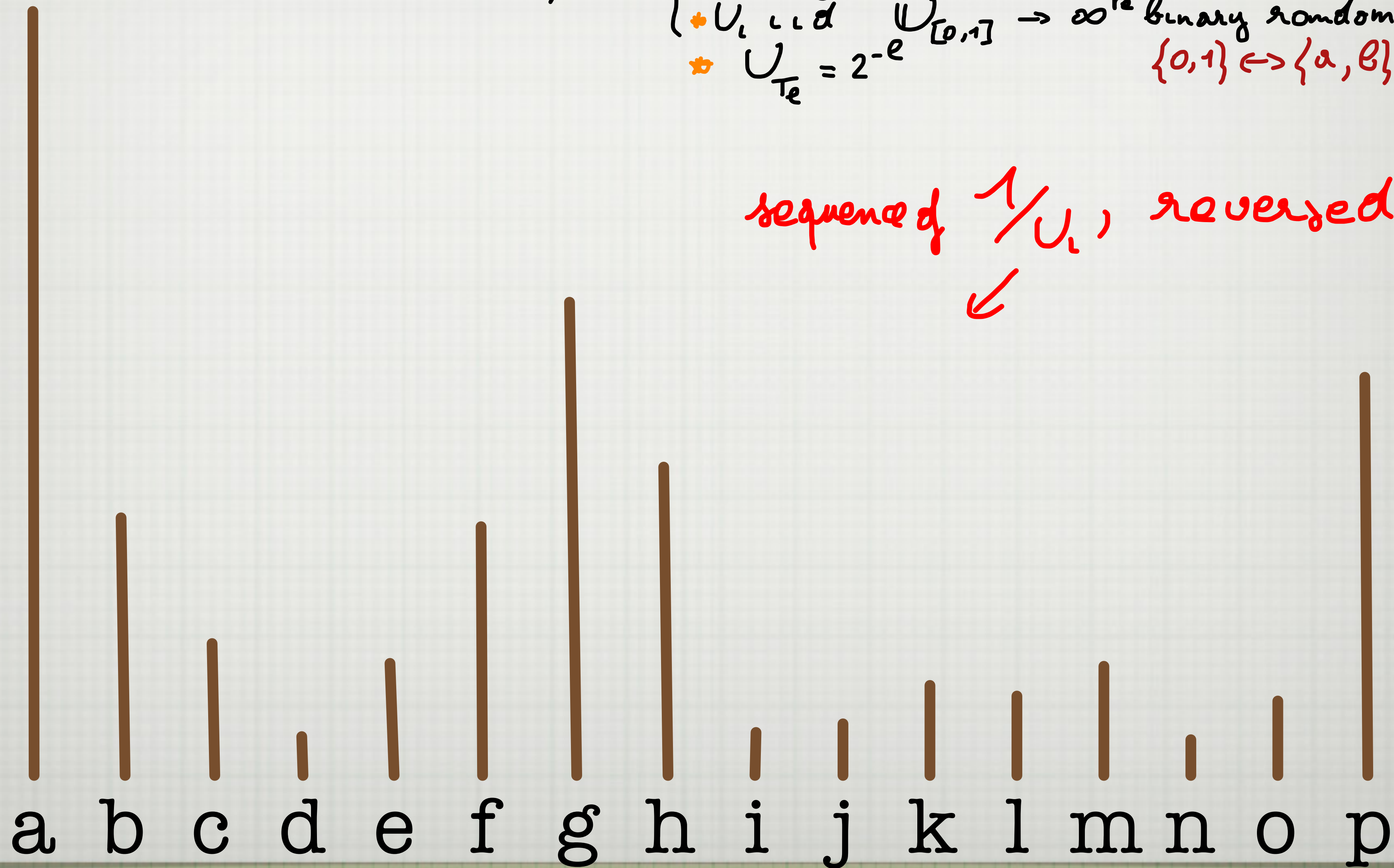
Assumption $|W_\ell| \simeq 2^\ell$ and $H(\mathcal{L}(W_\ell)) \simeq \alpha \ell$

$\Rightarrow H_n \simeq \alpha \log_2 n \Rightarrow \dfrac{\alpha}{\ln 2} = 5,09$

Lyndon Tree & Yule

$U^{(\ell)} = (U_1, U_2, \dots, U_{T_\ell}) \to$
- $T_\ell = \inf\{ k \mid U_k < 2^{-\ell} \}$
- U_i iid $U_{[0,1]}$ $\to$ ∞te binary random words
  $\{0,1\} \leftrightarrow \{a,b\}$
- $U_{T_\ell} = 2^{-\ell}$

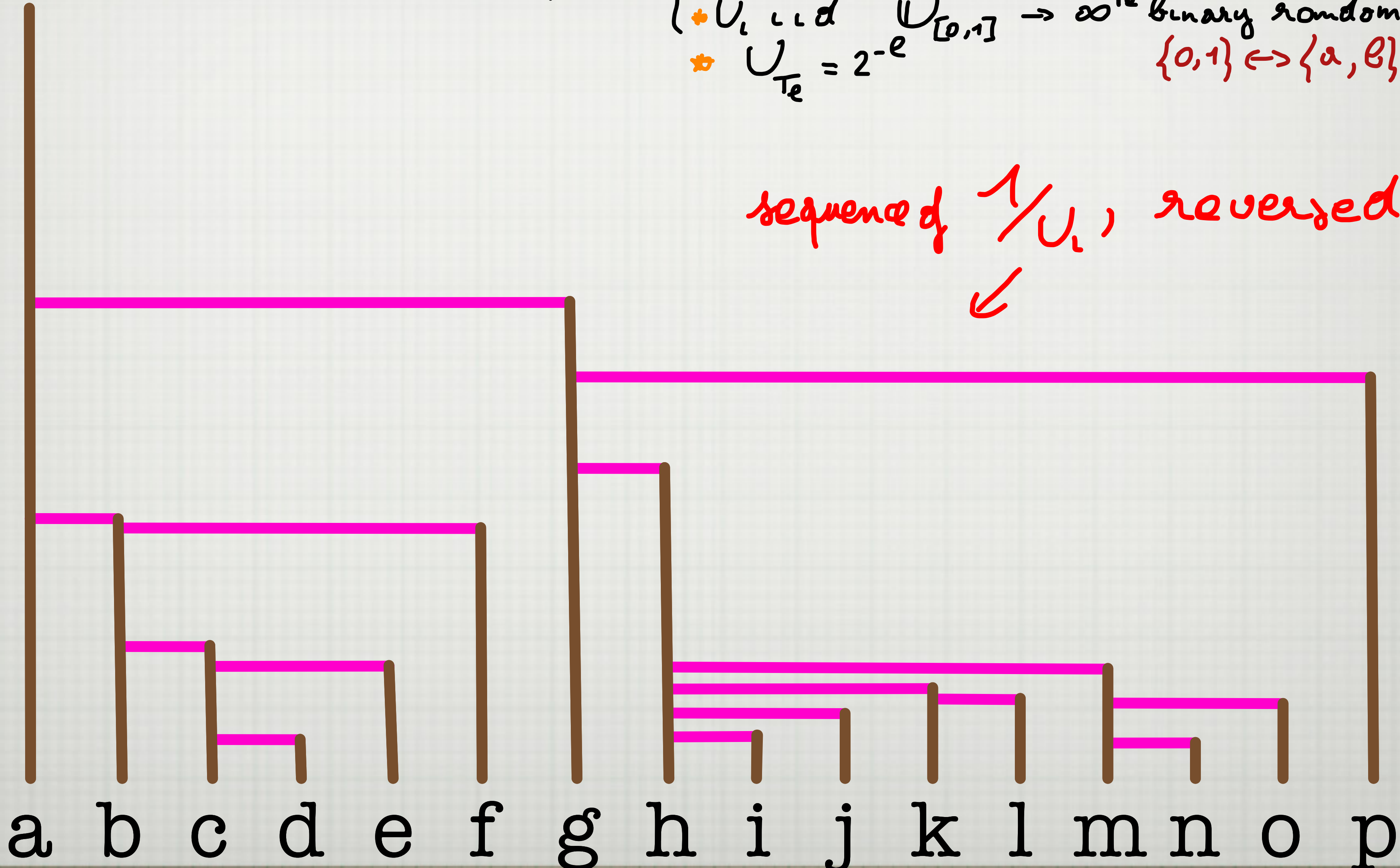
sequence of $1/U_i$, reversed



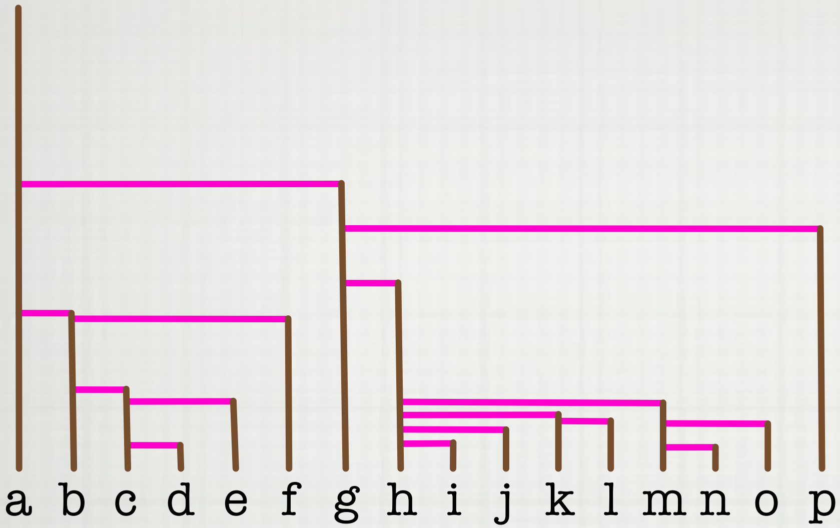
Lyndon tree & Yule

$U^{(\ell)} = (U_1, U_2, \dots, U_{T_\ell}) \rightarrow$

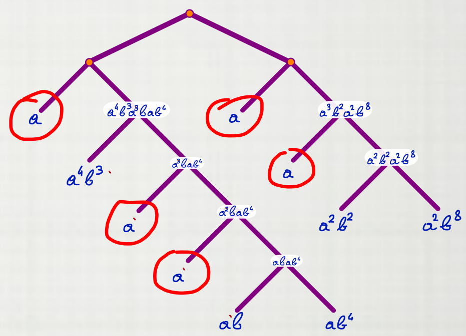
- $T_\ell = \inf\{k \mid U_k < 2^{-\ell}\}$
- U_i iid $U_{[0,1]}$ $\rightarrow$ ∞ te binary random words $\{0,1\} \leftrightarrow \{a,b\}$
- $U_{T_\ell} = 2^{-\ell}$

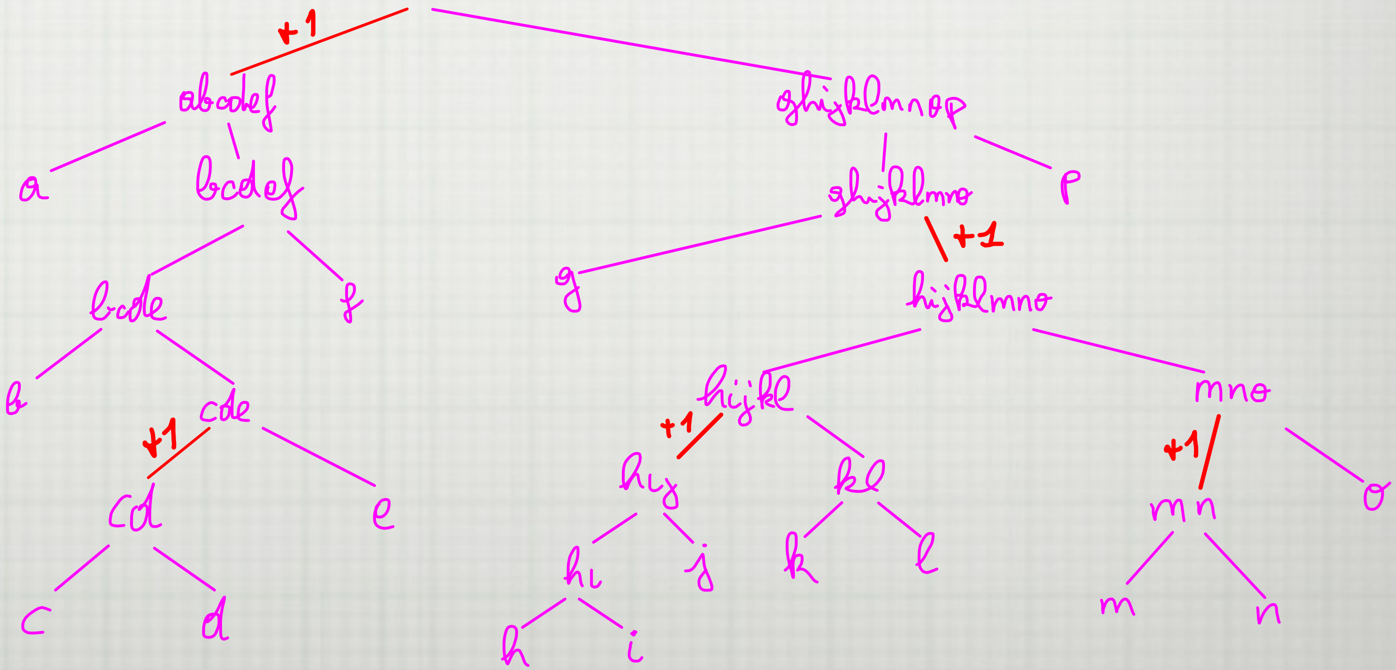


Lyndon tree & Yule

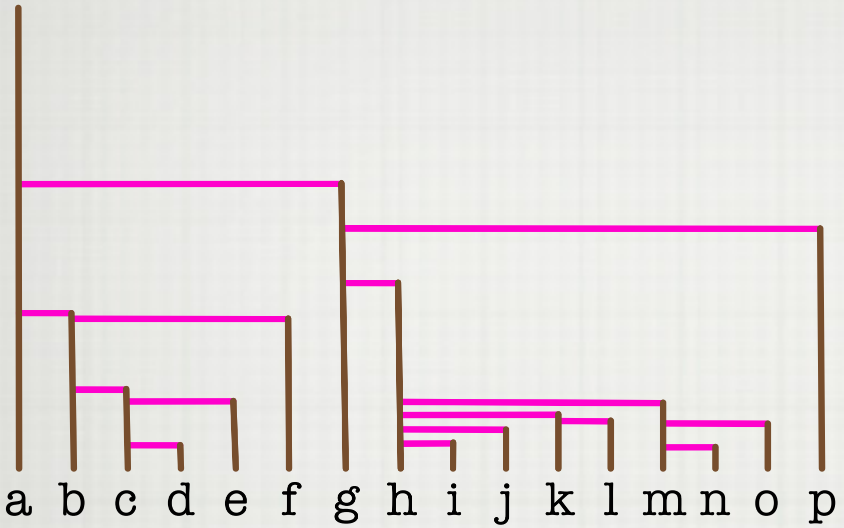


$$\frac{1}{U_x} > 2\,\frac{1}{U_y} \Rightarrow +1 \;(\,a\,)$$





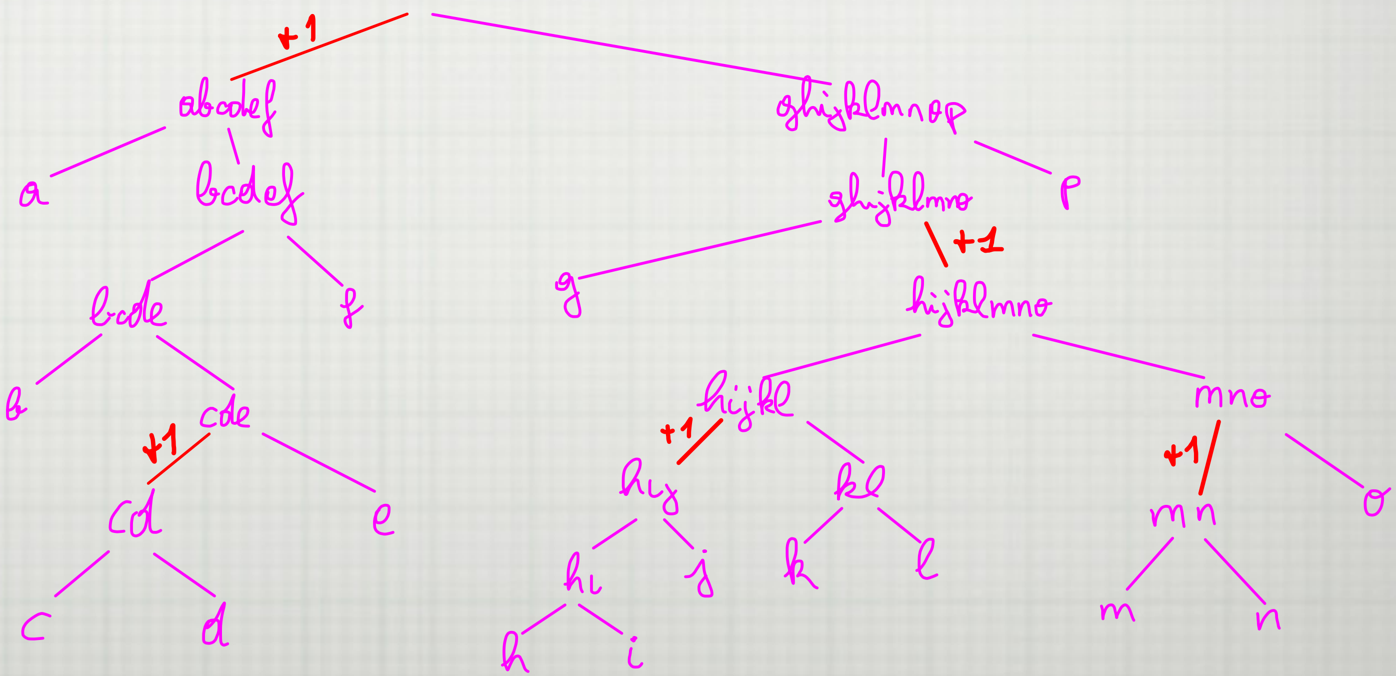
Lyndon tree & Yule



$$-\log_2 U_x > -\log_2 U_y + 1 \Rightarrow +1 \;(a)$$

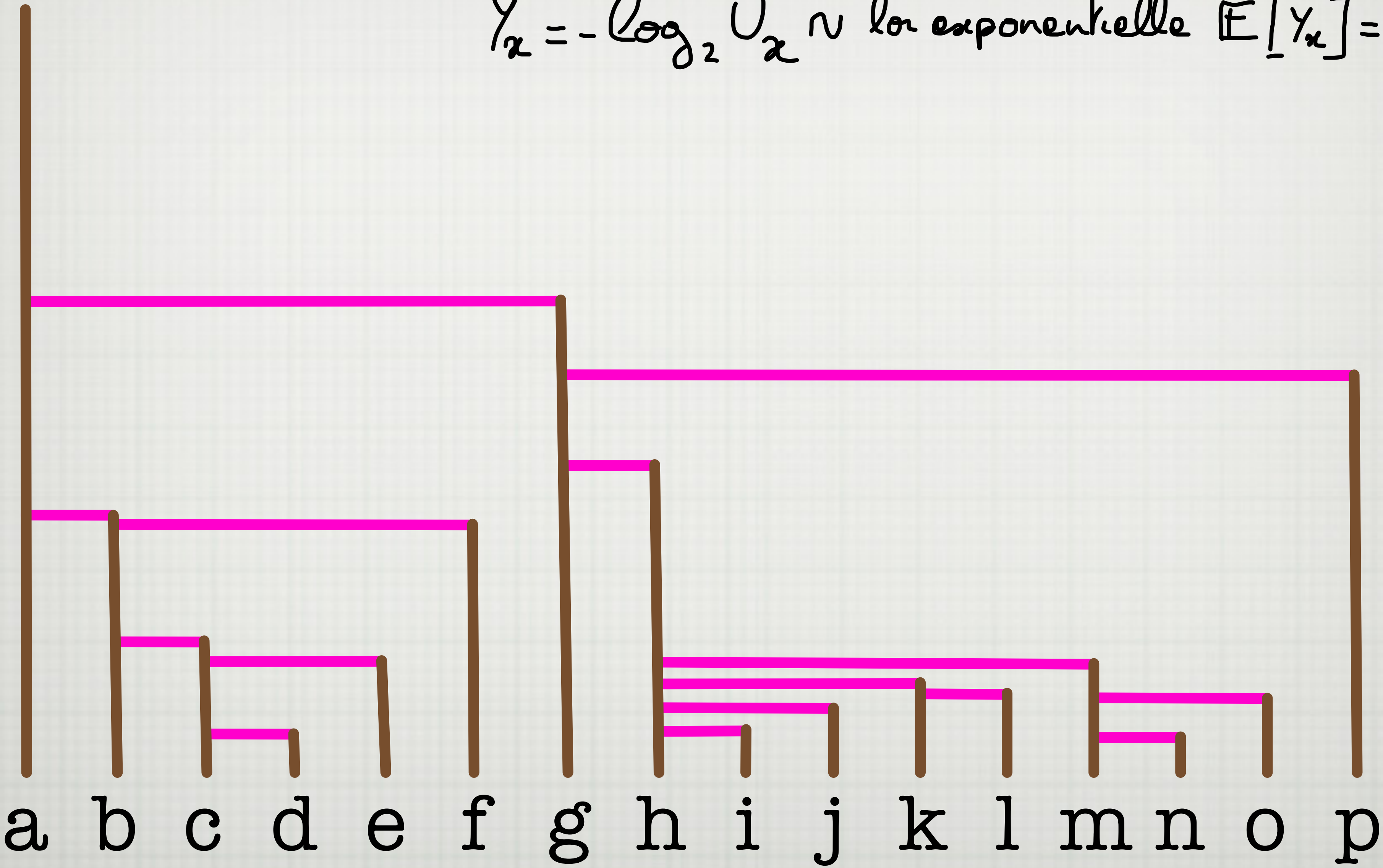
$$Y_x = -\log_2 U_x \sim \text{loi exponentielle}$$

$$\mathbb{E}[Y_x] = \frac{1}{\ln 2}$$



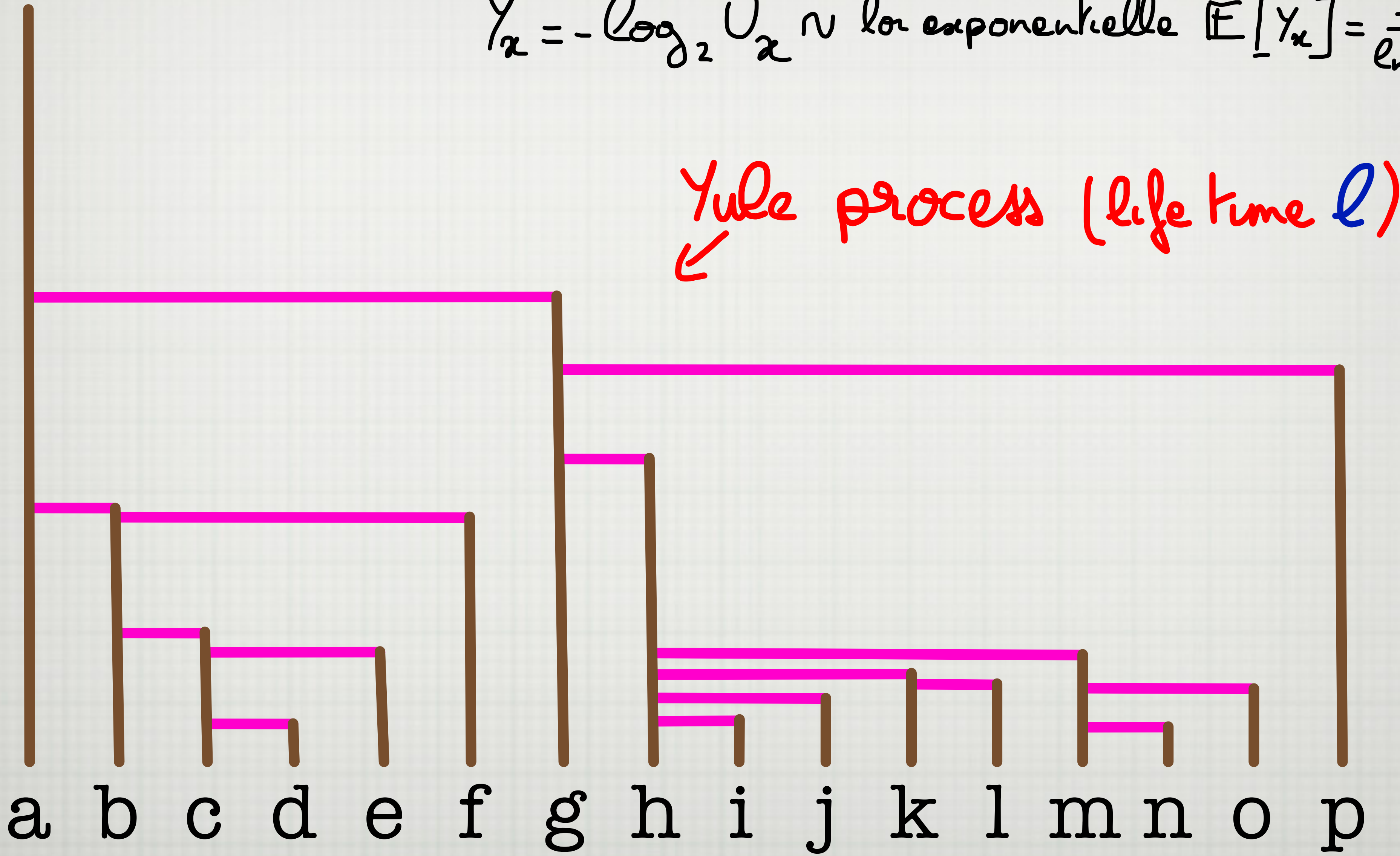
Lyndon tree & Yule

$Y_x = -\log_2 U_x \sim$ loi exponentielle $\mathbb{E}[Y_x] = \frac{1}{\ln 2}$

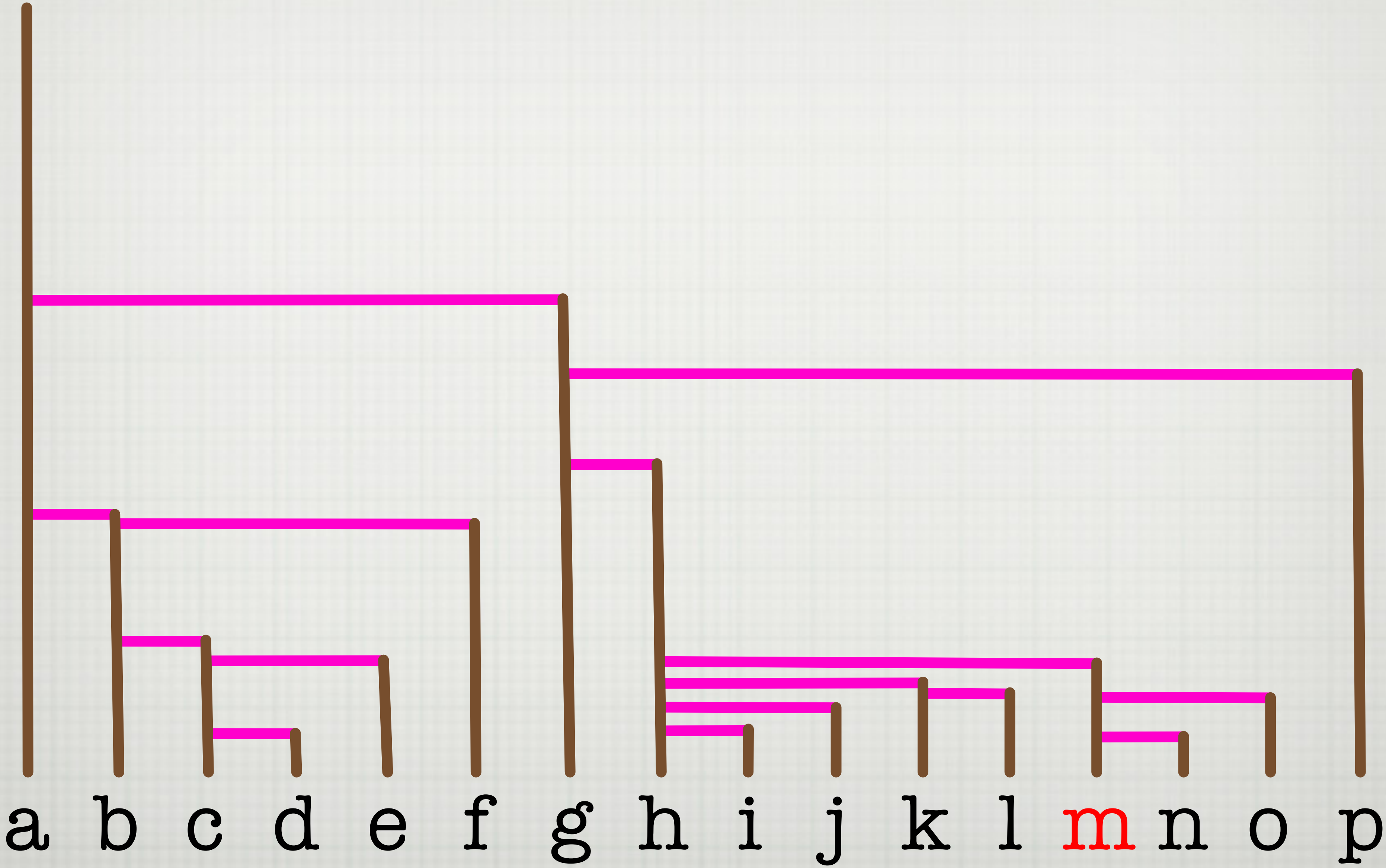


Lyndon tree & Yule

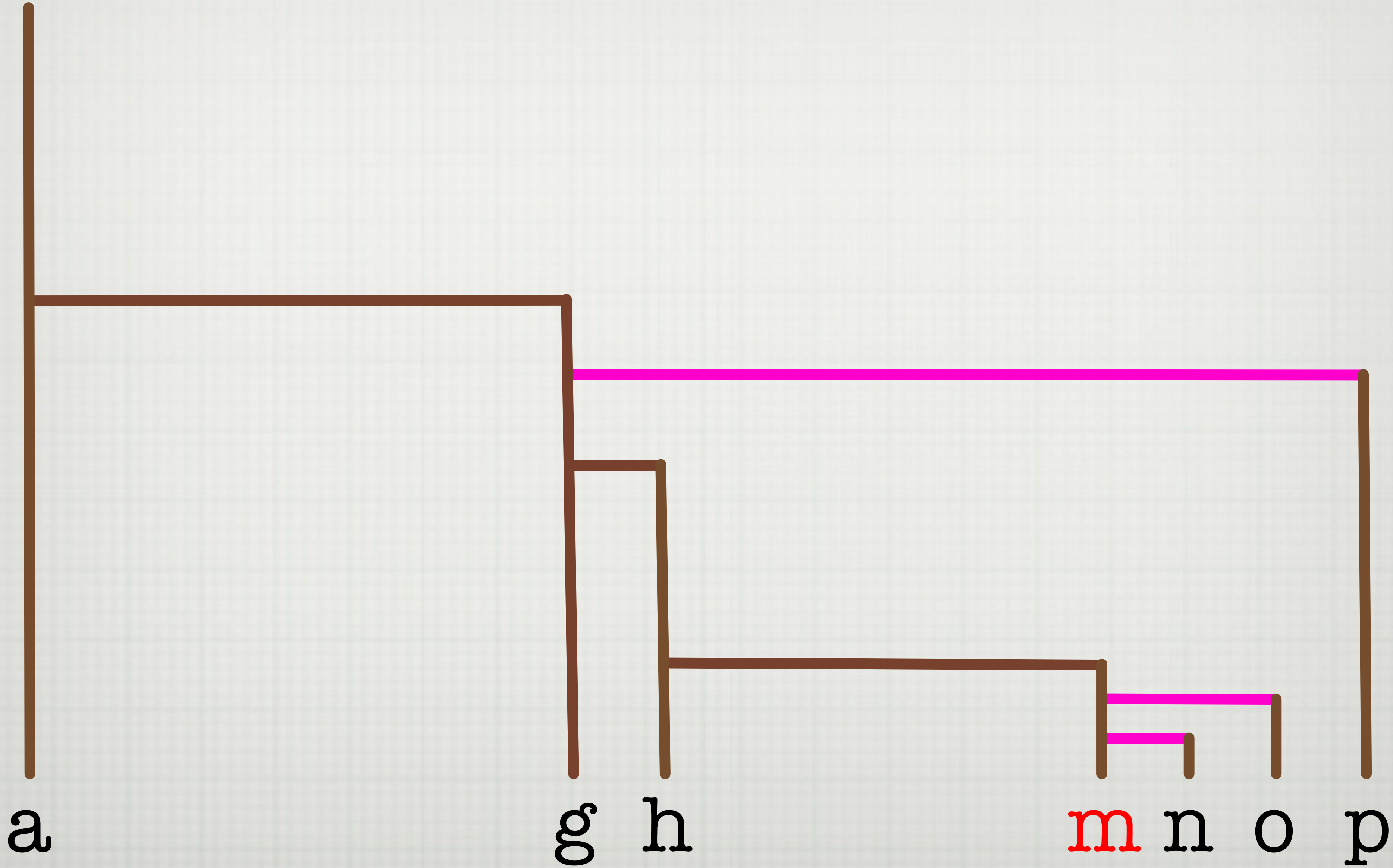
$Y_x = -\log_2 U_x \sim$ loi exponentielle $\mathbb{E}[Y_x] = \frac{1}{\ln 2}$



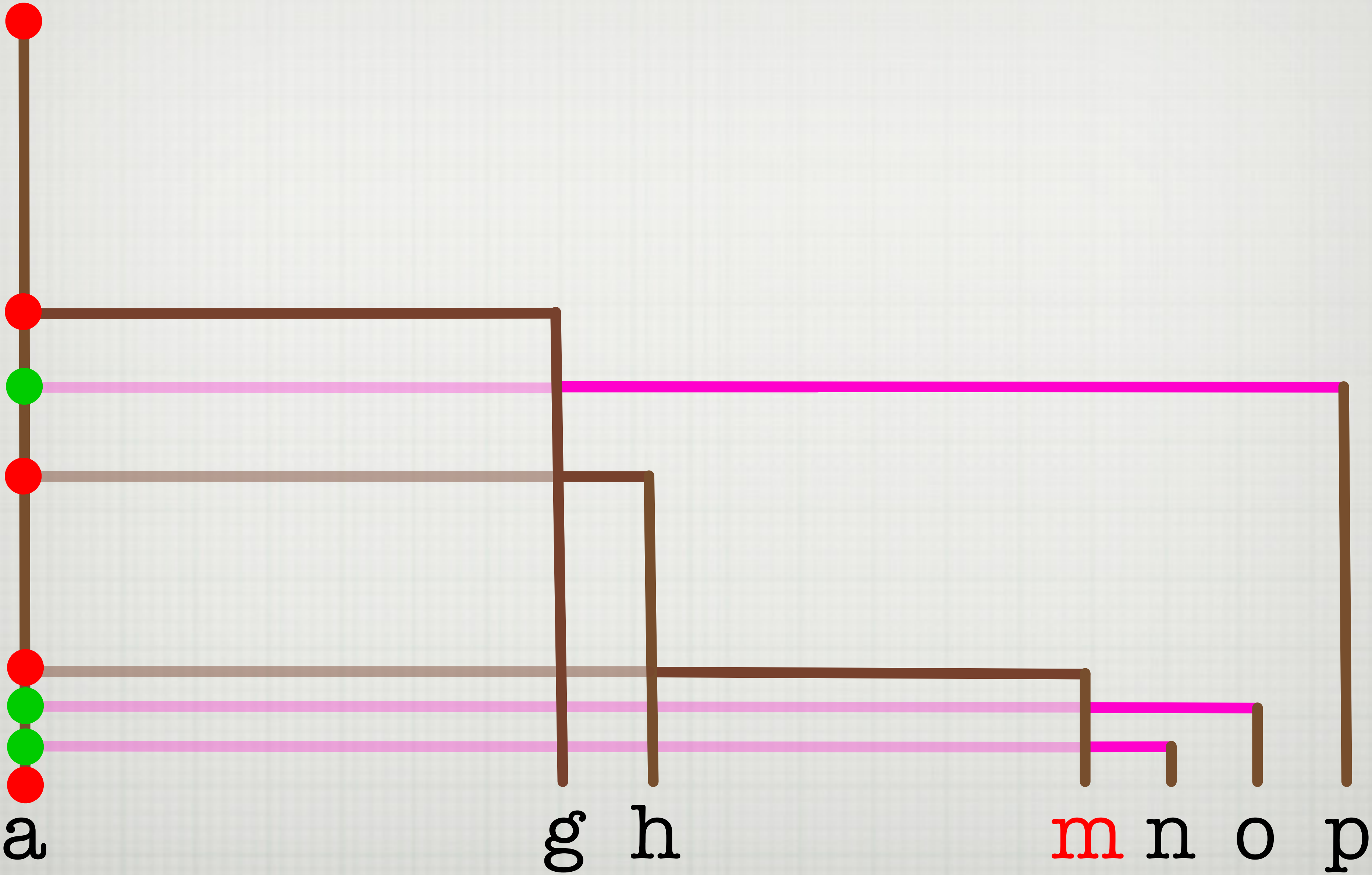
Lyndon Tree & Yule



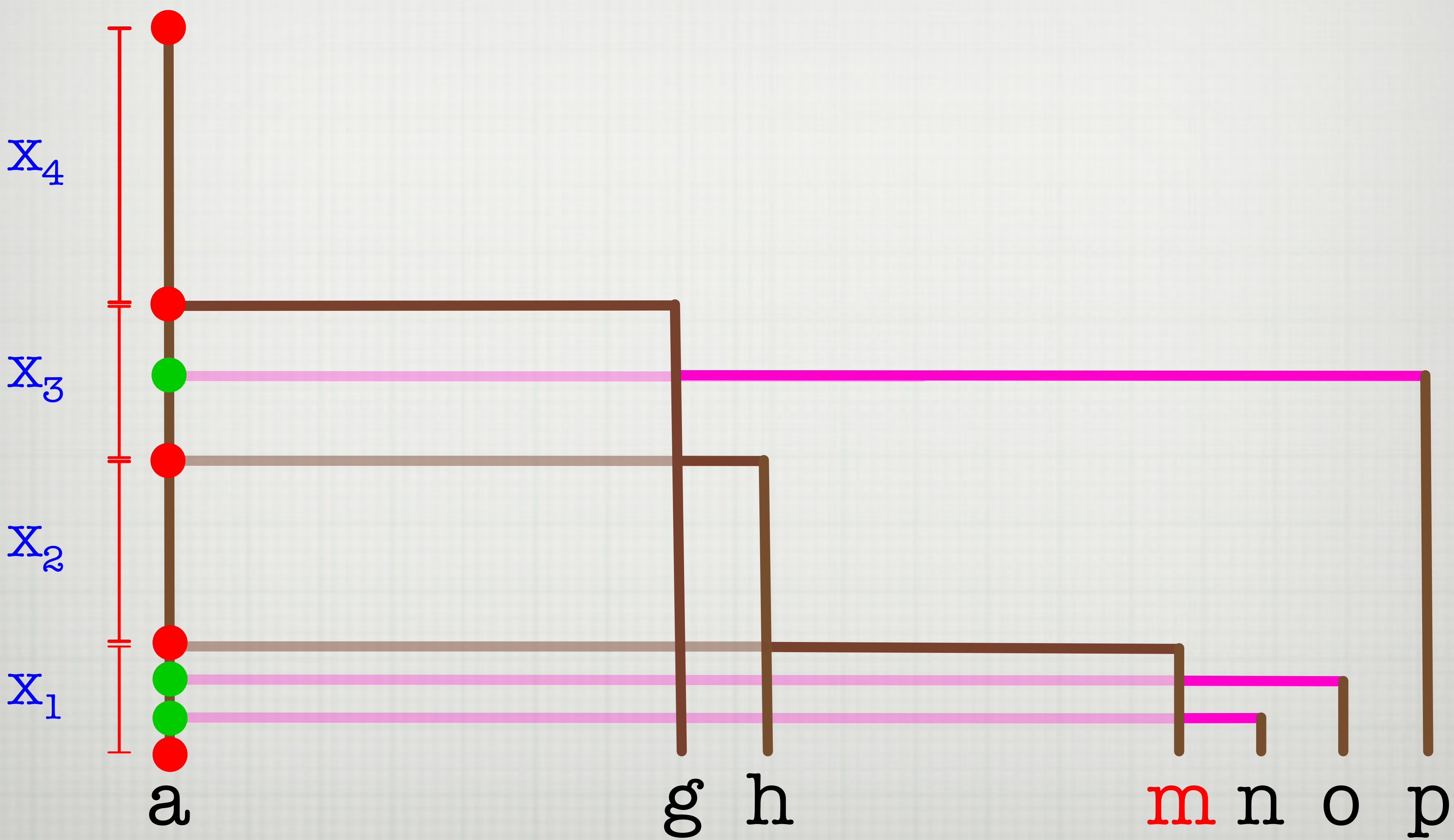
Lyndon tree & Yule



Lyndon Tree & Yule

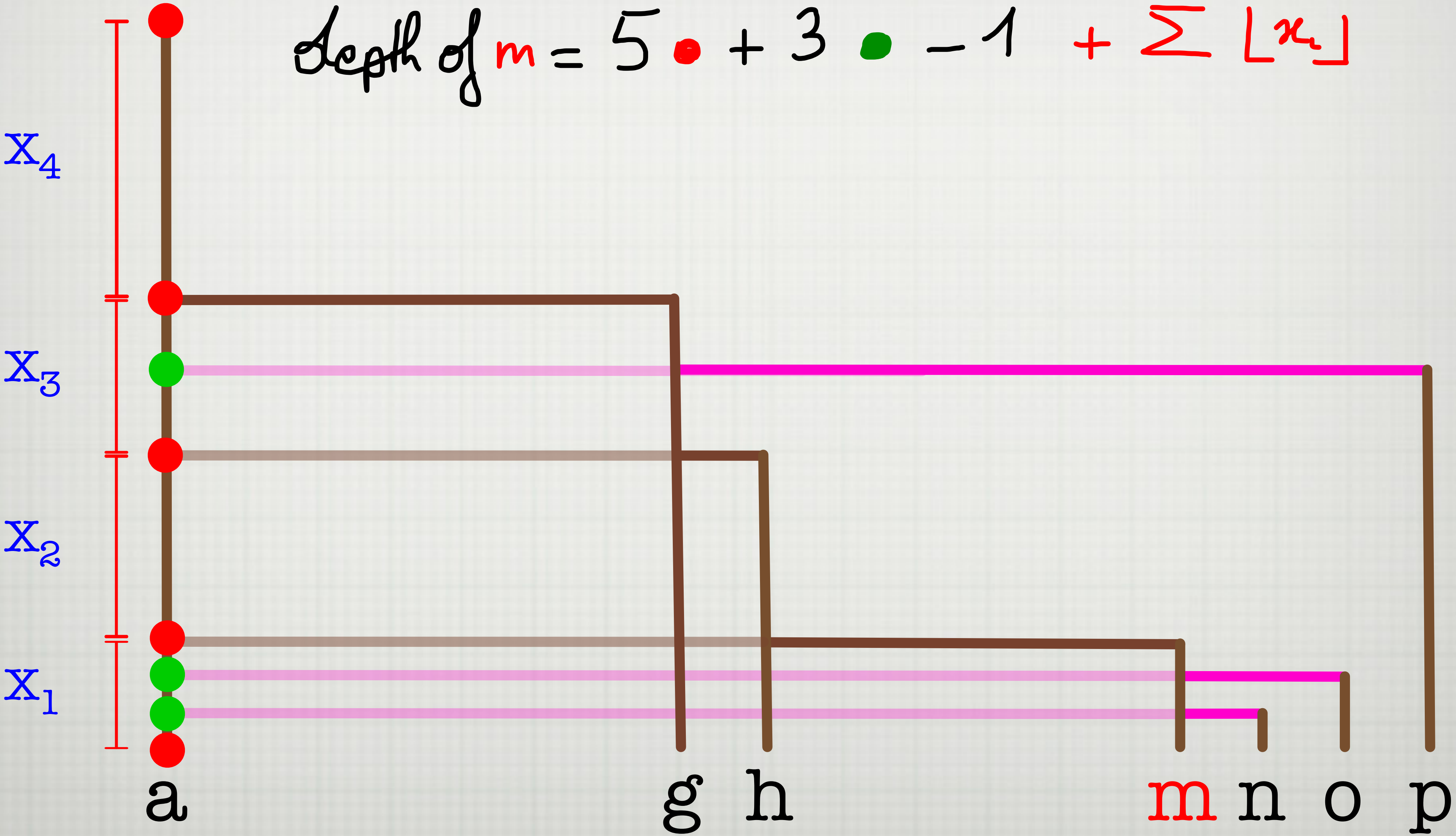


Lyndon tree & Yule



Lyndon tree & Yule

$$\text{depth of } m = 5 \, {\color{red}\bullet} + 3 \, {\color{green}\bullet} - 1 \; {\color{red} + \sum \lfloor x_i \rfloor}$$



Lyndon tree & Yule

$$\text{depth of } m = 5 \bullet + 3 \bullet - 1 + \sum \lfloor x_i \rfloor$$

$\Pi_m^{\bullet}$ red point process \ but the endpoints $\Pi_m^{\bullet}$ green point process

$$\sum \lfloor x_i \rfloor = G(\Pi_m^{\bullet})$$

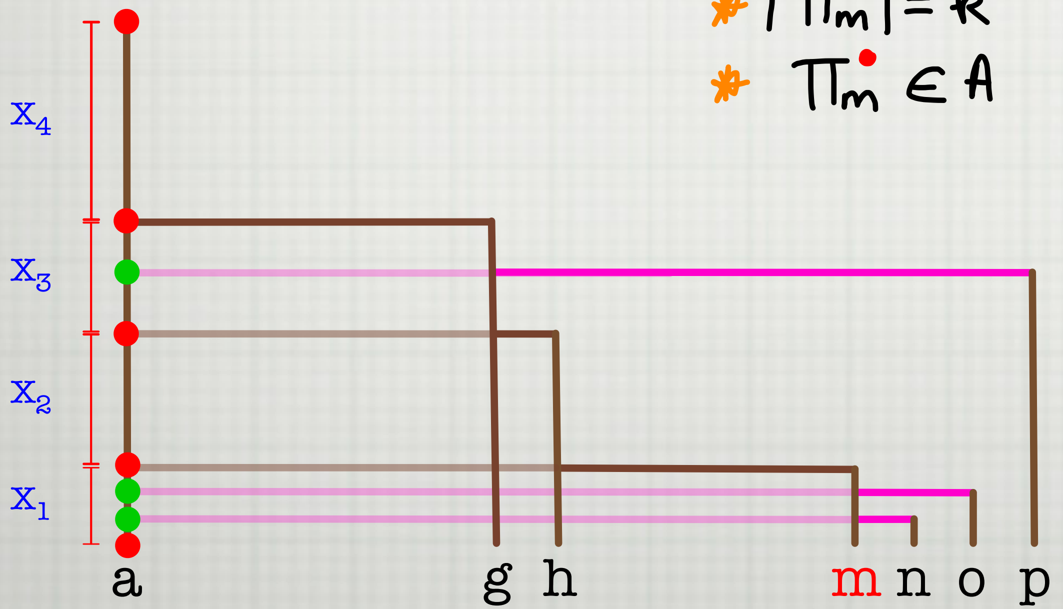
m is of type (n, k, A) if

* $|\Pi_m^{\bullet}| = n$
* $|\Pi_m^{\bullet}| = k$
* $\Pi_m^{\bullet} \in A$

$$\mathbb{E}[\# \text{ of leaves } (n,k,A)]$$

$$= L_{\ell,n,k,A}$$

$$= \frac{(\ln 2\ \ell)^n 2^{-\ell}}{n!} \frac{(\ln 2\ \ell)^k 2^{-\ell}}{k!} \mathbb{U}_{n,\ell}(A)$$



Lyndon tree & Yule

$$\text{depth of } m = |\Pi_m^{\bullet}| + |\Pi_m^{\bullet}| + 1 + G(\Pi_m^{\bullet}) \qquad \Sigma \lfloor x_i \rfloor = G(\Pi_m^{\bullet})$$

$$L_{\ell,n,k,A} = \mathbb{E}[\#\text{ of leaves }(n,k,A)] = \frac{(\ln 2\ e)^n 2^{-\ell}}{n!} \frac{(\ln 2\ e)^k 2^{-\ell}}{k!} U_{n,\ell}(A)$$

$$\frac{1}{n} \ln L_{\lambda n, n, \nu n, G=\mu n} \longrightarrow \psi(\lambda,\mu,\nu)$$

$$\sup_{\lambda\mu\nu} \frac{1}{\lambda \ln 2}\left(1+\mu+\nu+\frac{\psi(\lambda,\mu,\nu)}{\ln 2}\right) = 5{,}09 \quad !!$$