Random models of binary search trees

TU/e

Benoît Corsini

Content Table

Content Table

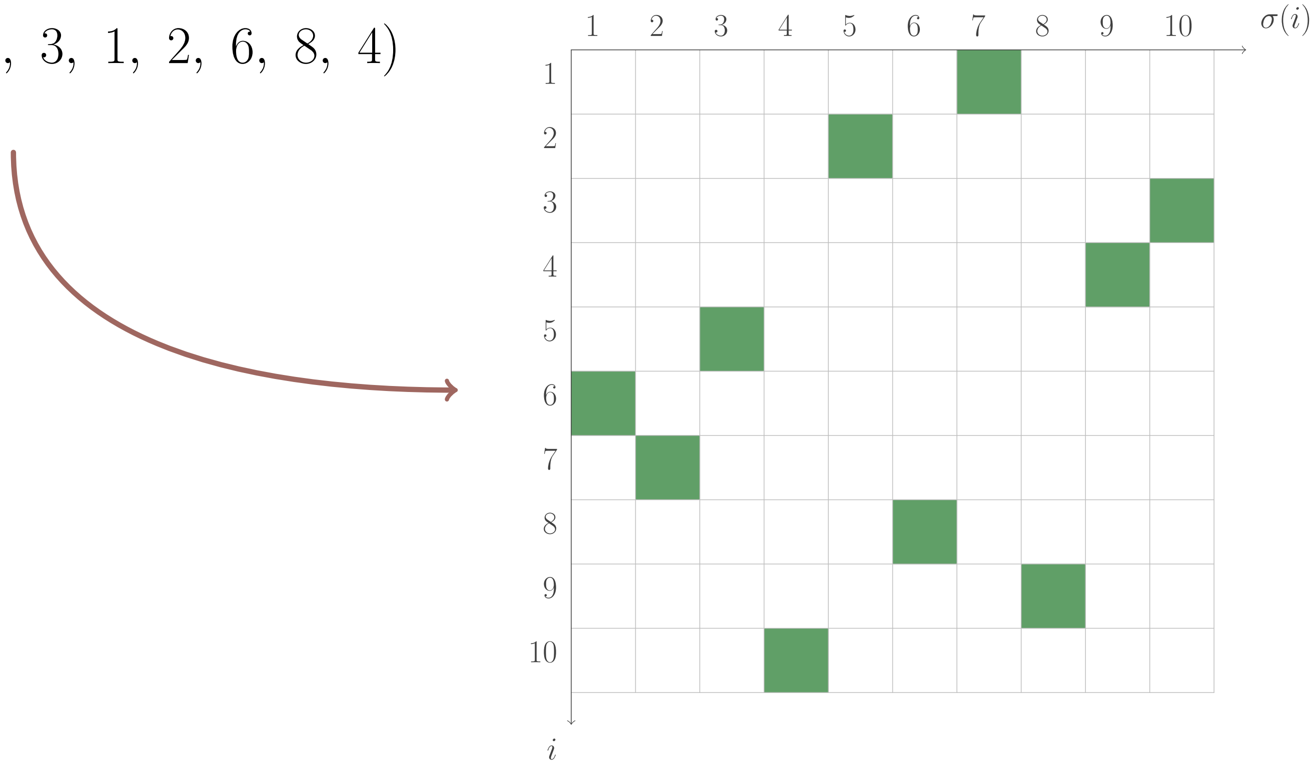
Representation

Representation

$$\sigma = (7,\ 5,\ 10,\ 9,\ 3,\ 1,\ 2,\ 6,\ 8,\ 4)$$

Representation

$$\sigma = (7,\ 5,\ 10,\ 9,\ 3,\ 1,\ 2,\ 6,\ 8,\ 4)$$



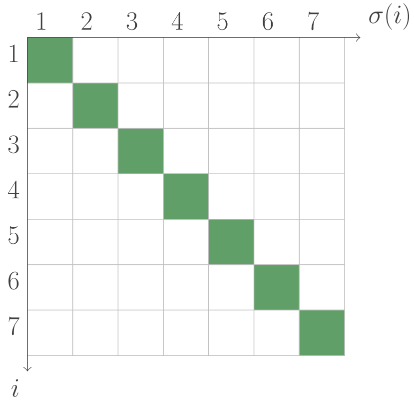
More examples

More examples

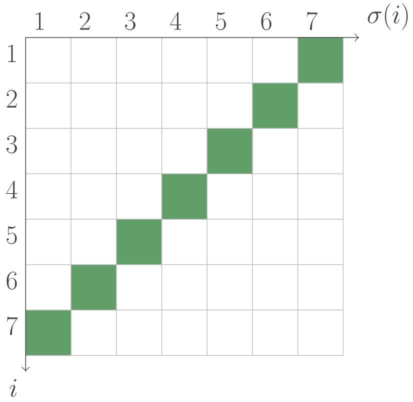
$(1, 2, 3, 4, 5, 6, 7)$ $(7, 6, 5, 4, 3, 2, 1)$ $(4, 2, 6, 1, 3, 5, 7)$ $(4, 6, 2, 7, 5, 3, 1)$

More examples

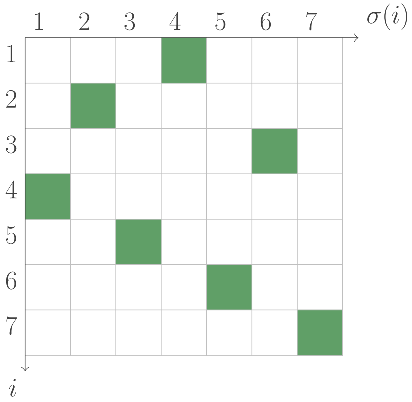
$(1,2,3,4,5,6,7)$



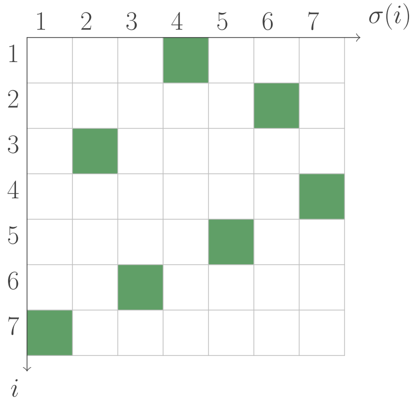
$(7,6,5,4,3,2,1)$



$(4,2,6,1,3,5,7)$

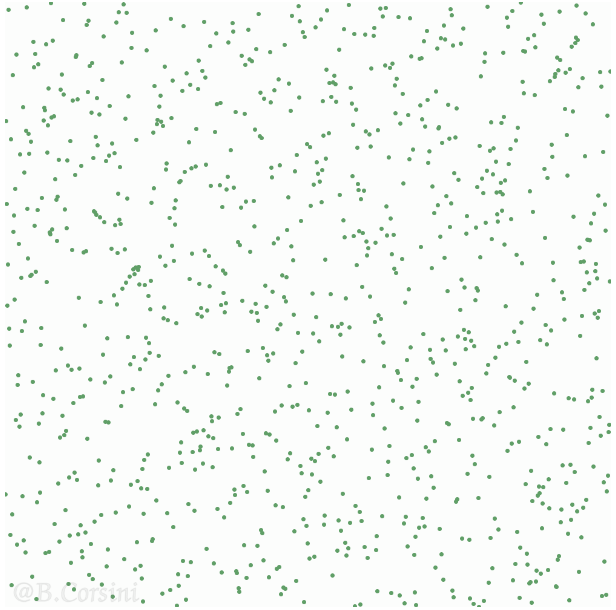


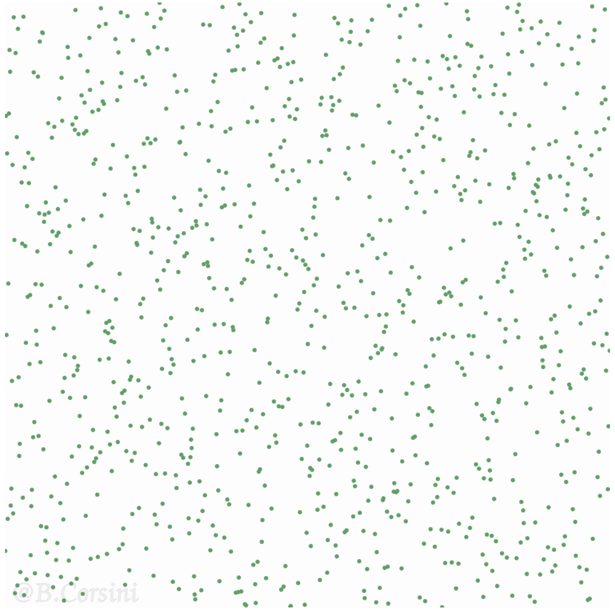
$(4,6,2,7,5,3,1)$

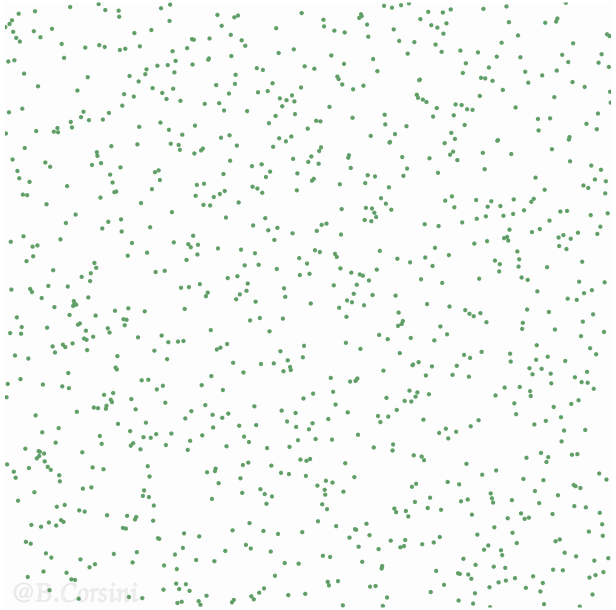


Images (uniform permutations)

Images (uniform permutations)







Random models of permutations

Random models of permutations

Definition (Mallows permutations)

A random Mallows permutation $X_{n,\lambda}$ with parameters $n \in \mathbb{N}$ and $\lambda \in [0, \infty)$ is defined by

$$\mathbb{P}[X_{n,\lambda} = \sigma] = \frac{\lambda^{\mathrm{Inv}(\sigma)}}{Z_{n,\lambda}},$$

where $\mathrm{Inv}(\sigma) = |\{i < j : \sigma(i) > \sigma(j)\}|$ is the number of inversions of σ and $Z_{n,\lambda} = \Sigma_{\sigma \in S_n} \lambda^{\mathrm{Inv}(\sigma)}$ is a normalizing constant.

Random models of permutations

Definition (Record-biased permutations)

A random record-biased permutation $X_{n,\lambda}$ with parameters $n \in \mathbb{N}$ and $\lambda \in [0, \infty)$ is defined by

$$\mathbb{P}[X_{n,\lambda} = \sigma] = \frac{\lambda^{\mathrm{Rec}(\sigma)}}{W_{n,\lambda}},$$

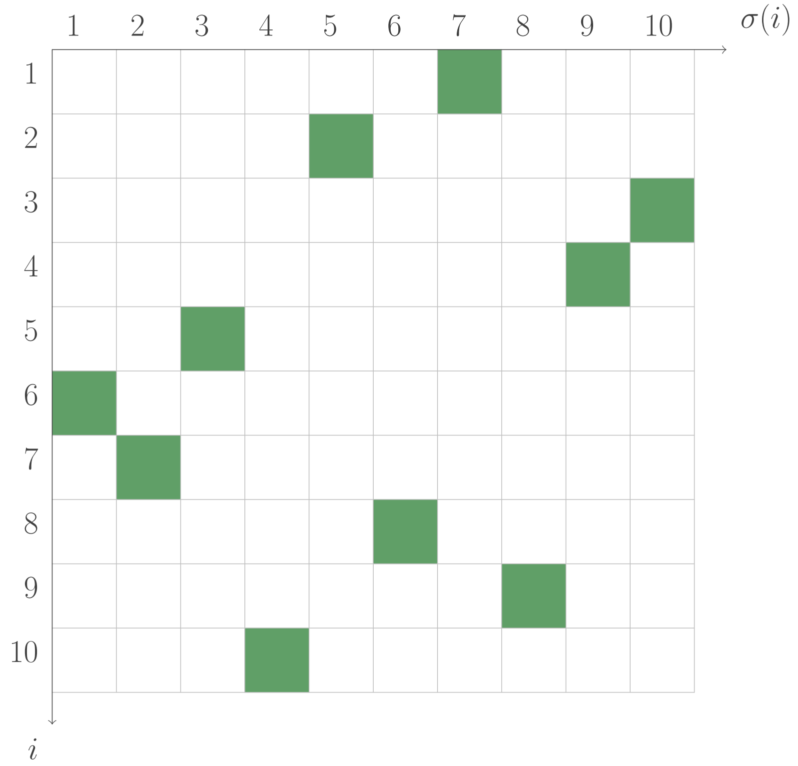
where $\mathrm{Rec}(\sigma) = |\{i : \forall j < i, \sigma(i) > \sigma(j)\}|$ is the number of records of σ and $W_{n,\lambda} = \Sigma_{\sigma \in S_n} \lambda^{\mathrm{Rec}(\sigma)}$ is a normalizing constant.

Summary

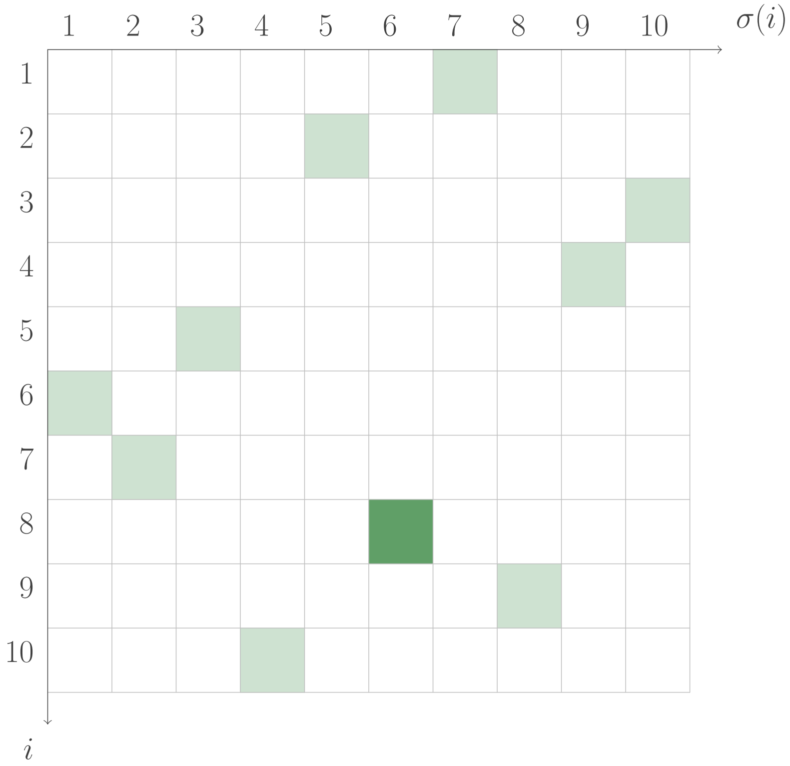
- A Mallows permutation X depends on $\mathrm{Inv}(\sigma) = |\{i < j : \sigma(i) > \sigma(j)\}|$ as follows
$$\mathbb{P}\big[X = \sigma\big] \propto \lambda^{\mathrm{Inv}(\sigma)}\,.$$
- A record-biased permutation X depends on $\mathrm{Rec}(\sigma) = |\{i : \forall j < i, \sigma(i) > \sigma(j)\}|$ as follows
$$\mathbb{P}\big[X = \sigma\big] \propto \lambda^{\mathrm{Rec}(\sigma)}\,.$$

Inversions and records

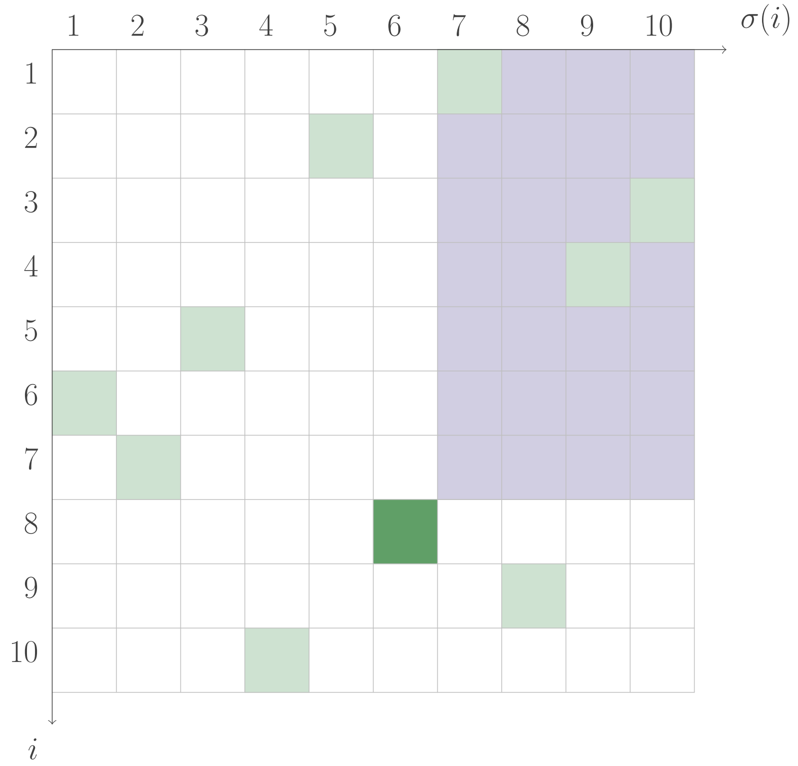
Inversions and records



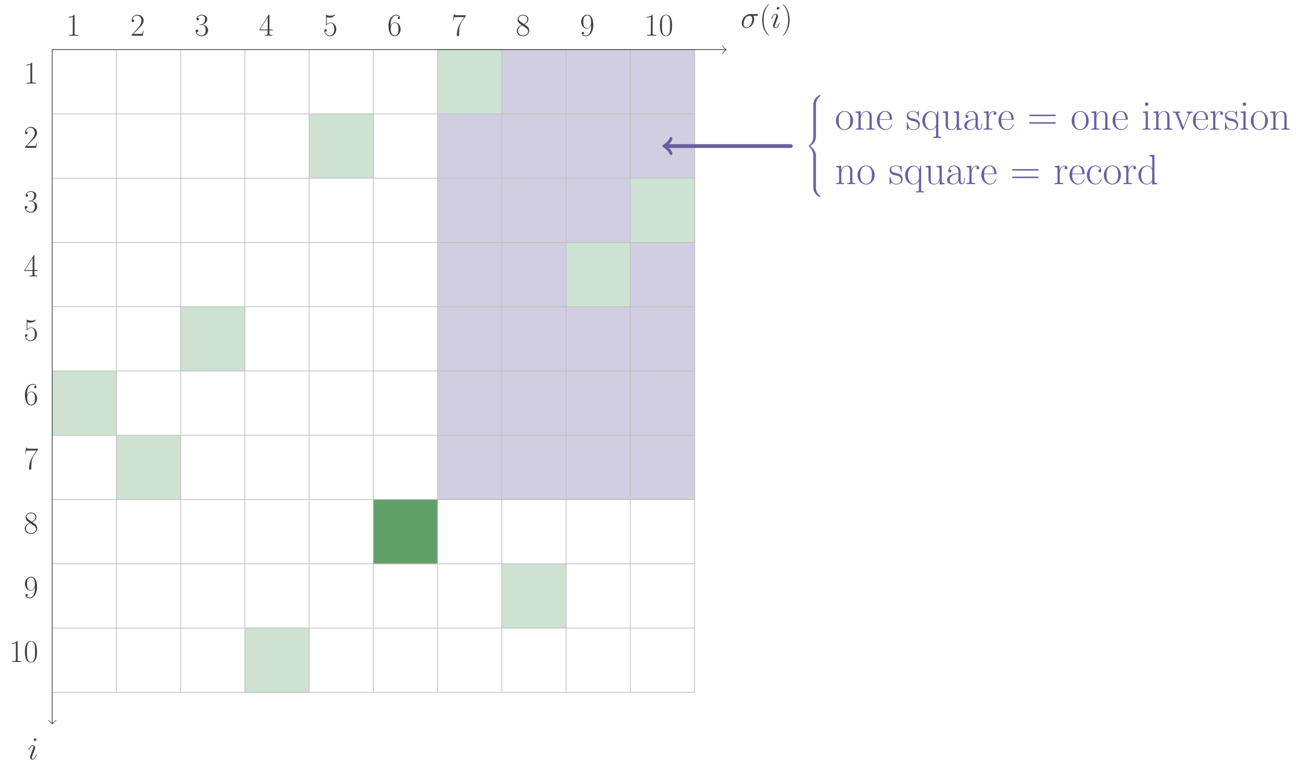
Inversions and records



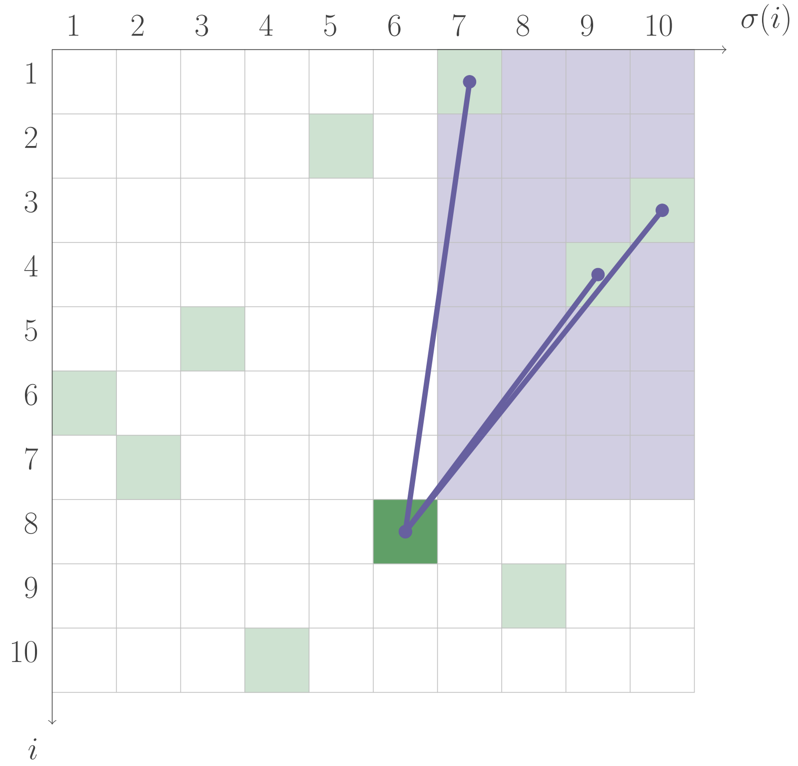
Inversions and records



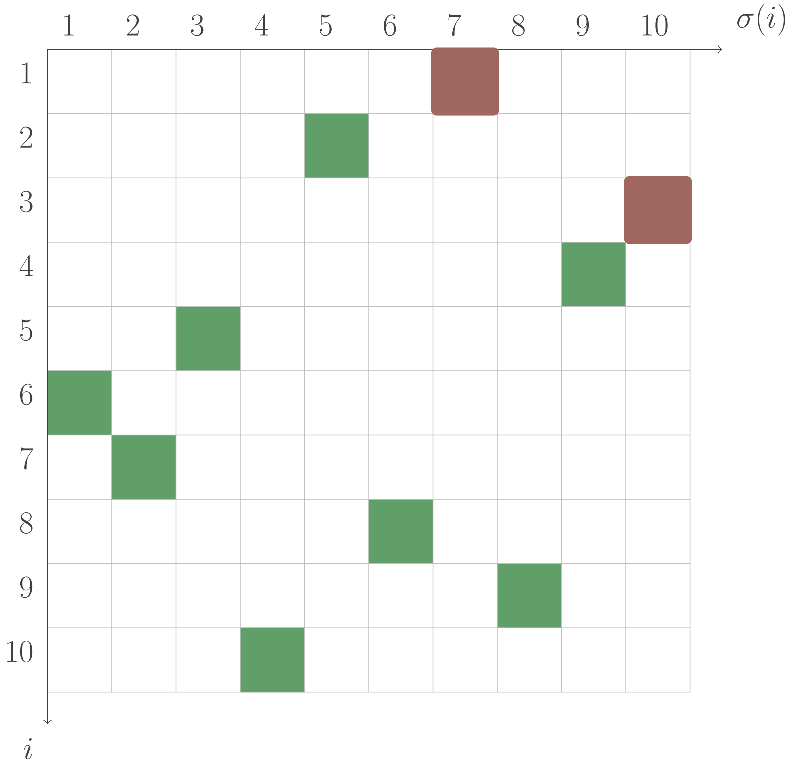
Inversions and records



Inversions and records

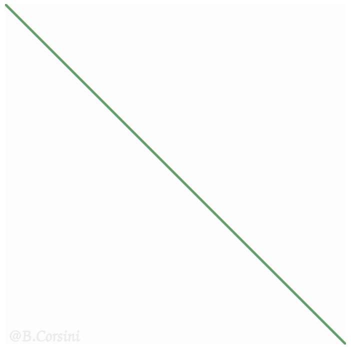


Inversions and records

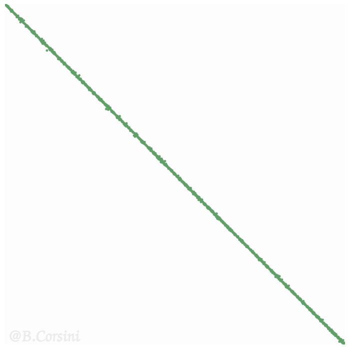


Images (Mallows permutations) $\mathbb{P}[X=\sigma] \propto \lambda^{|\{i<j:\sigma(i)>\sigma(j)\}|}$

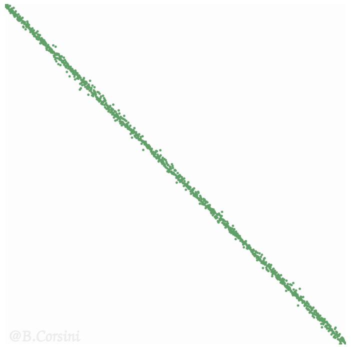
Images (Mallows permutations) $\mathbb{P}[X=\sigma] \propto \lambda^{|\{i<j:\sigma(i)>\sigma(j)\}|}$



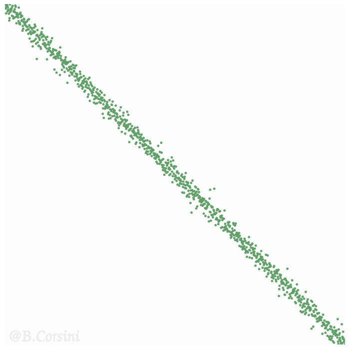
$\lambda = 0$



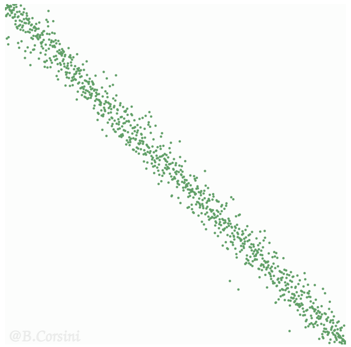
$\lambda = 0.5$



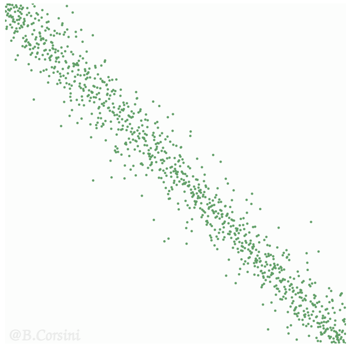
$\lambda = 0.8$



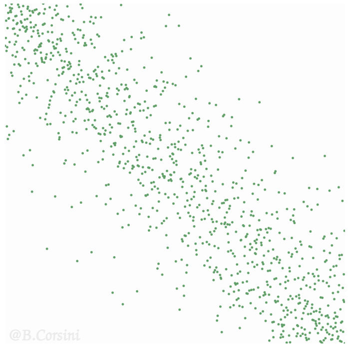
$\lambda = 0.9$



$\lambda = 0.95$

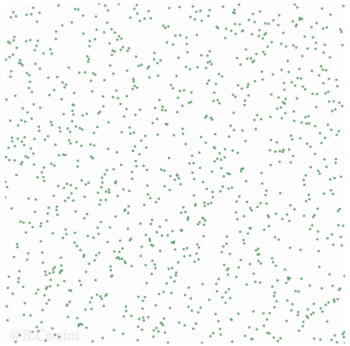


$\lambda = 0.97$

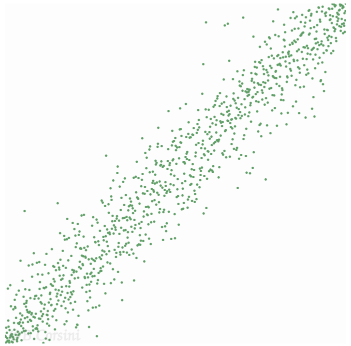


$\lambda = 0.99$

$\lambda = 0.997$



$\lambda = 1$

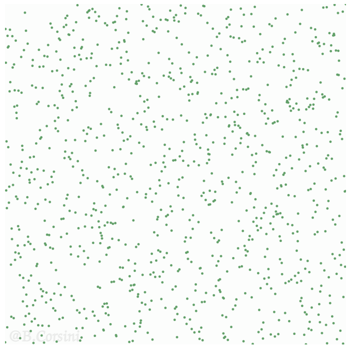


$\lambda = 1.02$

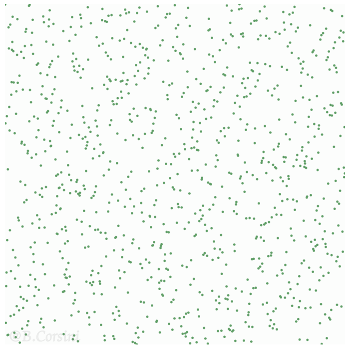
Images (record-biased permutations) $\mathbb{P}[X=\sigma] \propto \lambda^{|\{i : \forall j<i, \sigma(i)>\sigma(j)\}|}$

Images (record-biased permutations)

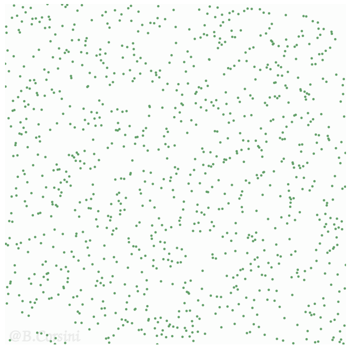
$\mathbb{P}[X = \sigma] \propto \lambda^{|\{i : \forall j < i, \sigma(i) > \sigma(j)\}|}$



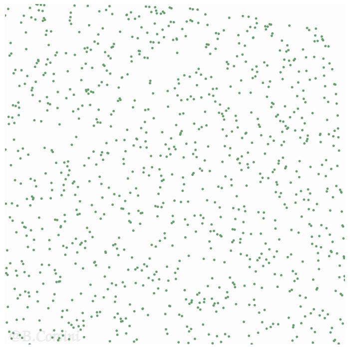
$\lambda = 0$



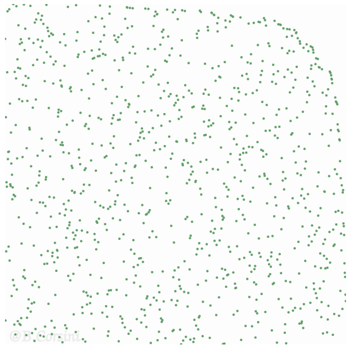
$\lambda = 1$



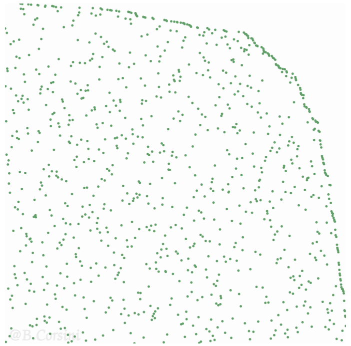
$\lambda = 5$



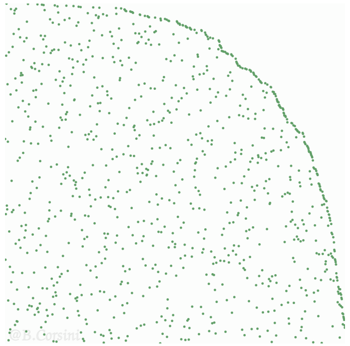
$\lambda = 10$



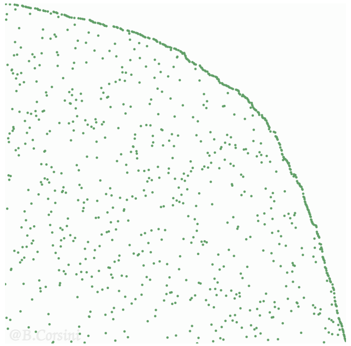
$\lambda = 20$



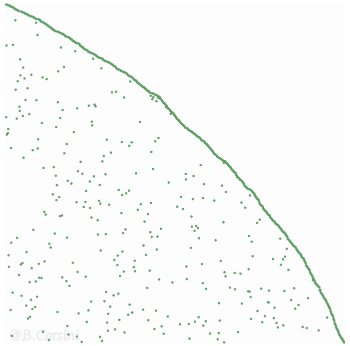
$\lambda = 50$



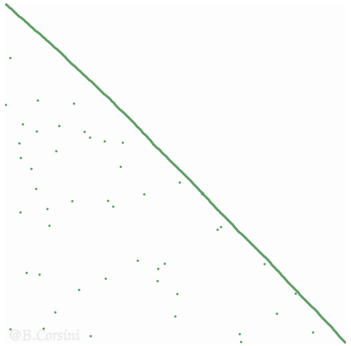
$\lambda = 100$



$\lambda = 200$



$\lambda = 1000$



$\lambda = 10000$

Content Table

Binary search trees

Binary search trees

Definition

A binary search tree is a binary labelled tree such that the label of each node is larger than the labels of the nodes in its left subtree and smaller than the labels of the nodes in its right subtree.

Definition

A binary search tree is a binary labelled tree such that the label of each node is larger than the labels of the nodes in its left subtree and smaller than the labels of the nodes in its right subtree.

Given a binary search tree, any new value has a unique position where it can be inserted in the tree.

Binary search trees

Definition

A binary search tree is a binary labelled tree such that the label of each node is larger than the labels of the nodes in its left subtree and smaller than the labels of the nodes in its right subtree.

Given a binary search tree, any new value has a unique position where it can be inserted in the tree.

→ Any sequence of distinct values corresponds to a unique binary search tree.

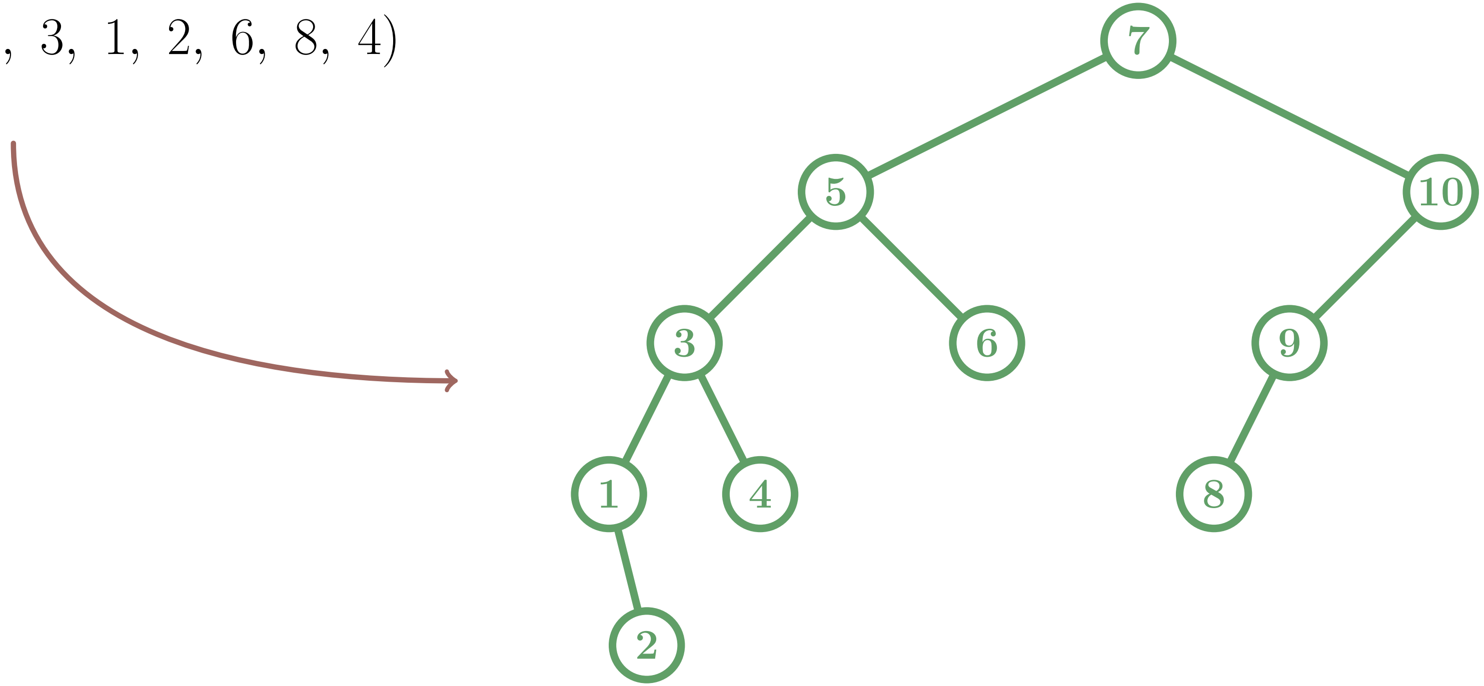
An example

An example

$$\sigma = (7,\ 5,\ 10,\ 9,\ 3,\ 1,\ 2,\ 6,\ 8,\ 4)$$

An example

$$\sigma = (7,\ 5,\ 10,\ 9,\ 3,\ 1,\ 2,\ 6,\ 8,\ 4)$$



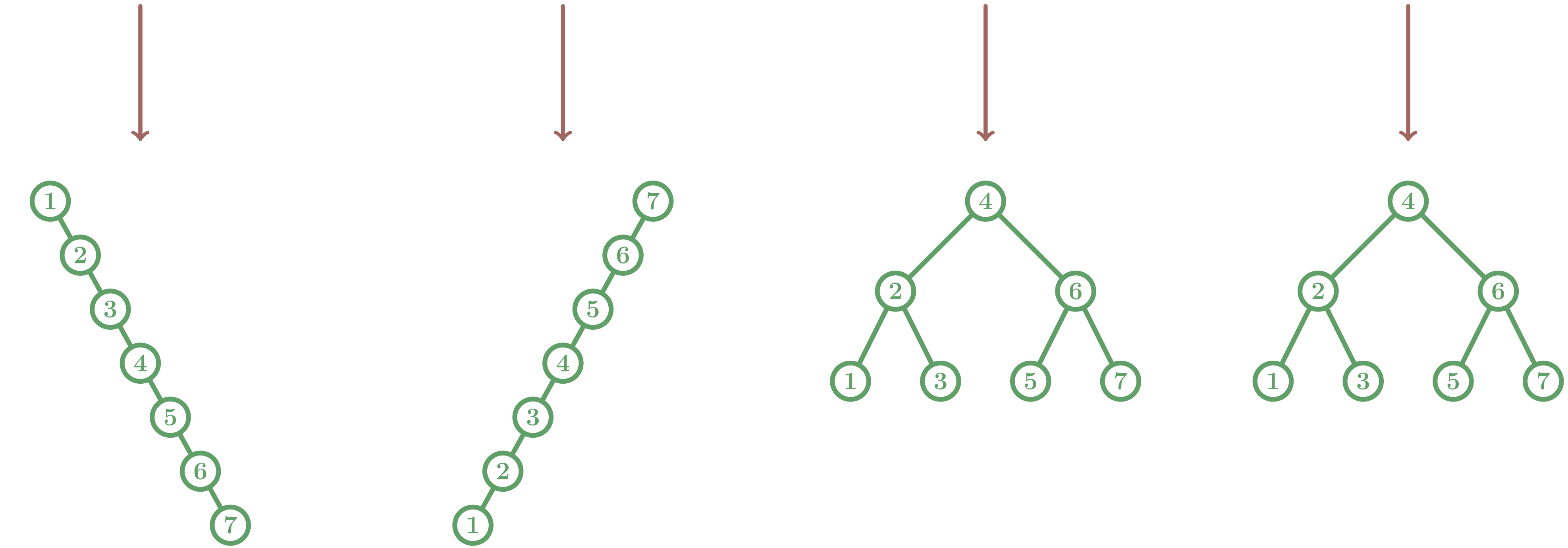
More examples

More examples

$(1, 2, 3, 4, 5, 6, 7)$ $(7, 6, 5, 4, 3, 2, 1)$ $(4, 2, 6, 1, 3, 5, 7)$ $(4, 6, 2, 7, 5, 3, 1)$

More examples

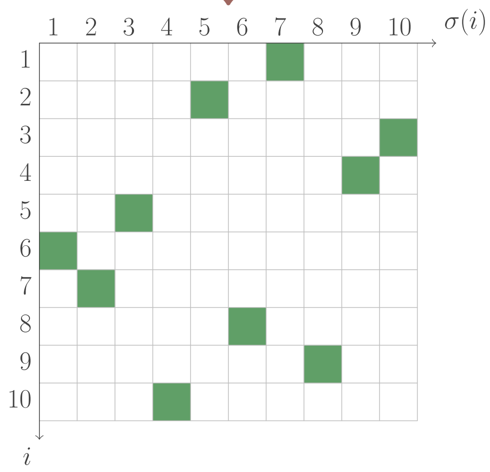
$(1, 2, 3, 4, 5, 6, 7)$ $(7, 6, 5, 4, 3, 2, 1)$ $(4, 2, 6, 1, 3, 5, 7)$ $(4, 6, 2, 7, 5, 3, 1)$

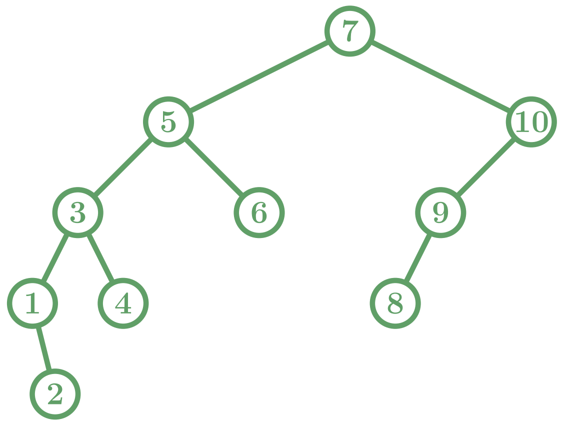


$$\sigma = (7,\ 5,\ 10,\ 9,\ 3,\ 1,\ 2,\ 6,\ 8,\ 4)$$

Binary search trees and matrix representation of permutations

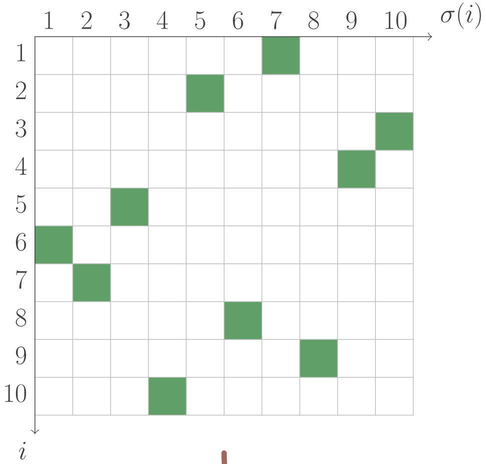
$$\sigma = (7,\ 5,\ 10,\ 9,\ 3,\ 1,\ 2,\ 6,\ 8,\ 4)$$

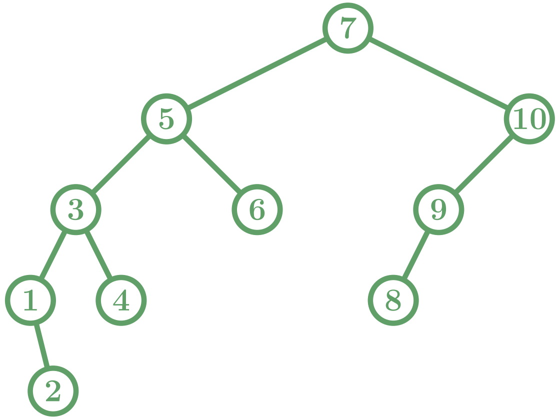


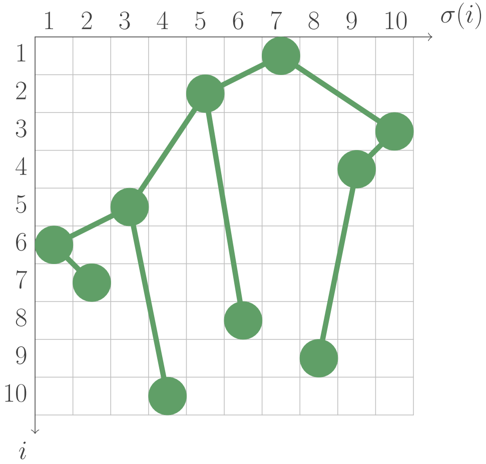


Binary search trees and matrix representation of permutations

$$\sigma = (7,\ 5,\ 10,\ 9,\ 3,\ 1,\ 2,\ 6,\ 8,\ 4)$$

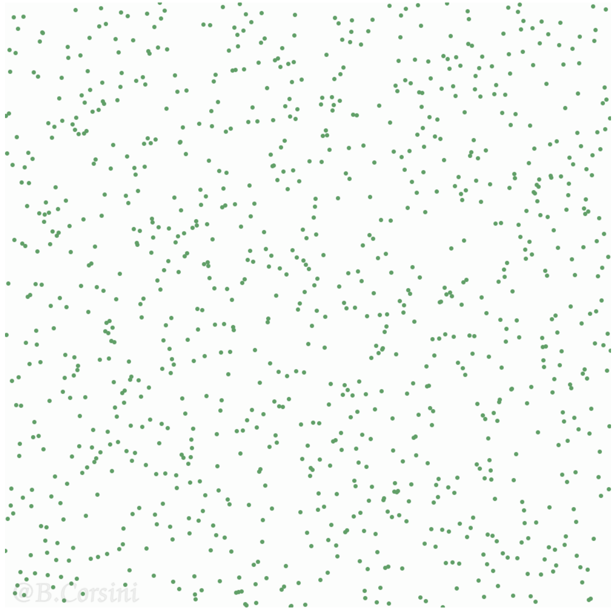




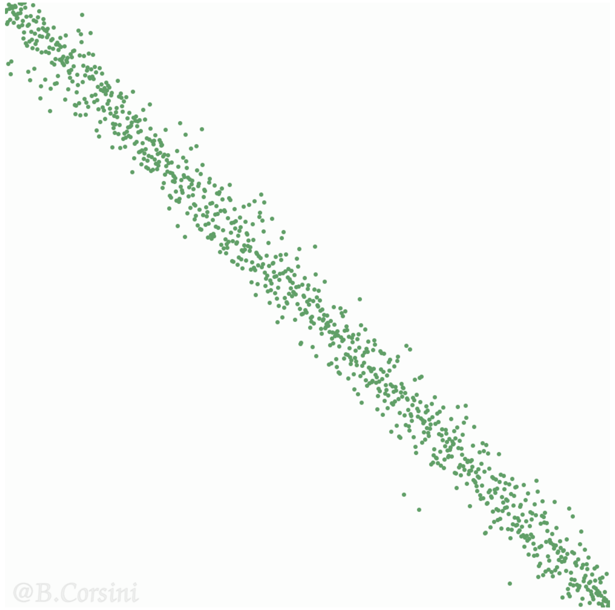


Images

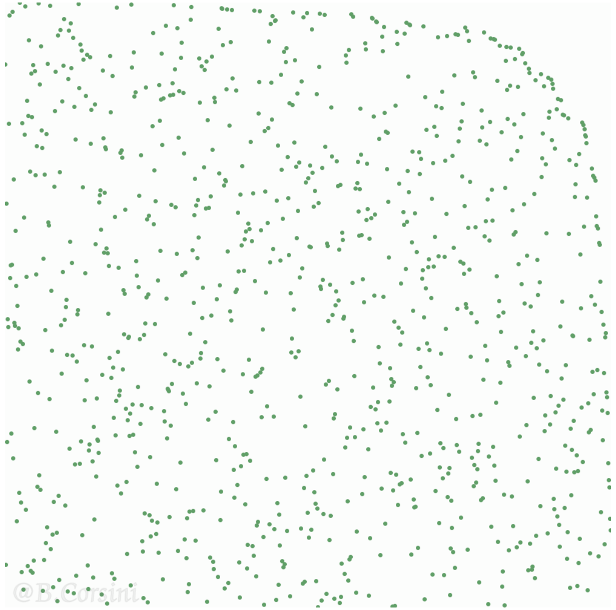
Images



Uniform permutation

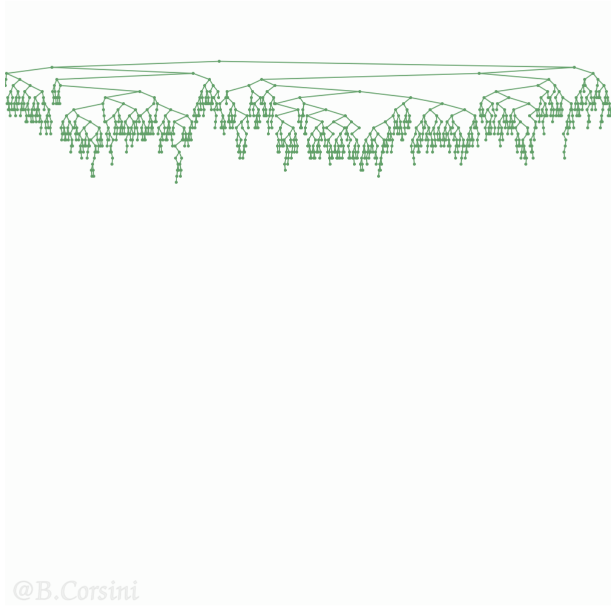


Mallows permutation

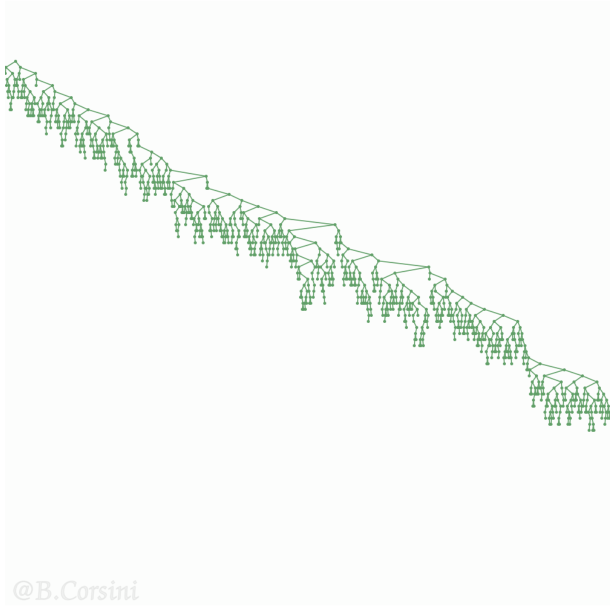


Record-biased permutation

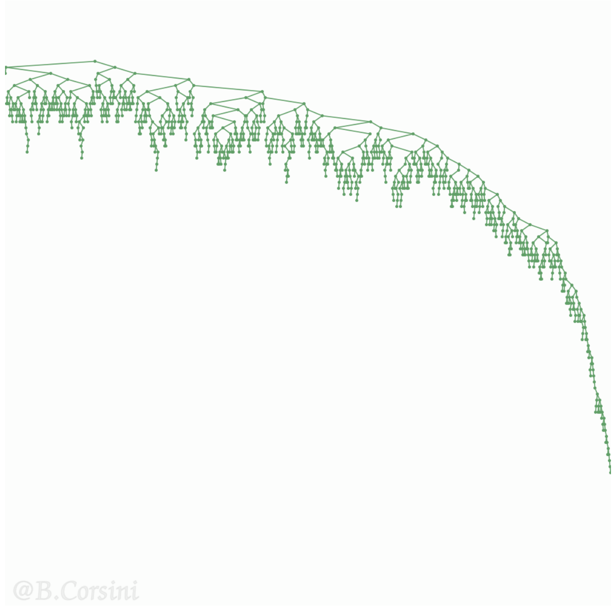
Images



Uniform permutation



Mallows permutation



Record-biased permutation

A fun property!

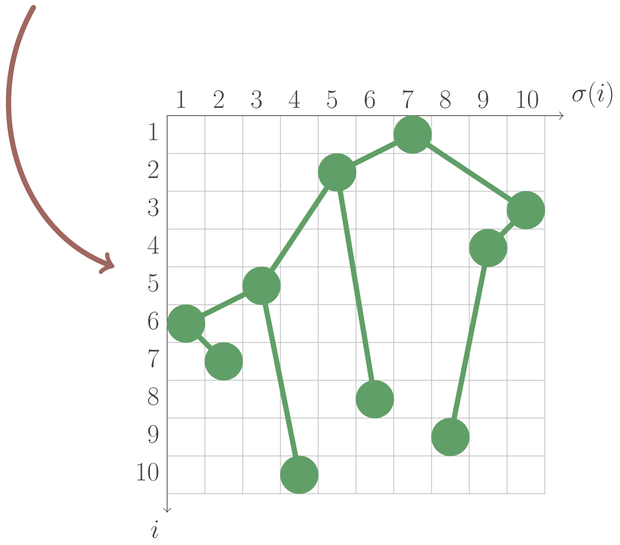
A fun property!

Why can we restrict λ to $[0, 1]$ in the case of Mallows trees ?

A fun property!

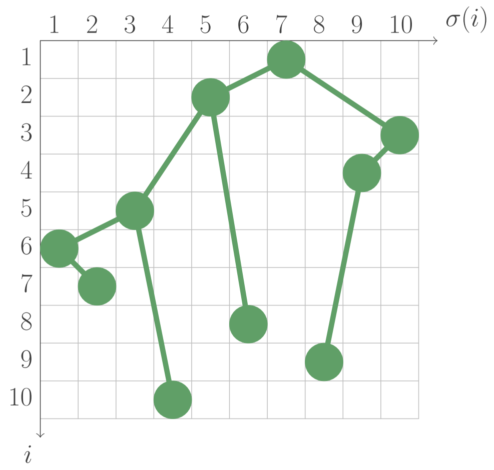
A fun property!

$$\sigma = (7, 5, 10, 9, 3, 1, 2, 6, 8, 4)$$

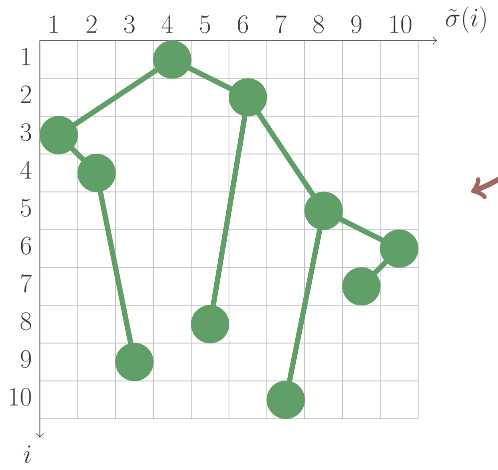


A fun property!

$$\sigma = (7, 5, 10, 9, 3, 1, 2, 6, 8, 4)$$

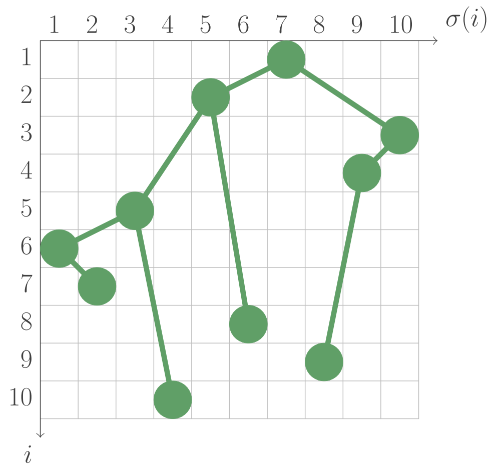


$$\tilde{\sigma} = (4, 6, 1, 2, 8, 10, 9, 5, 3, 7)$$



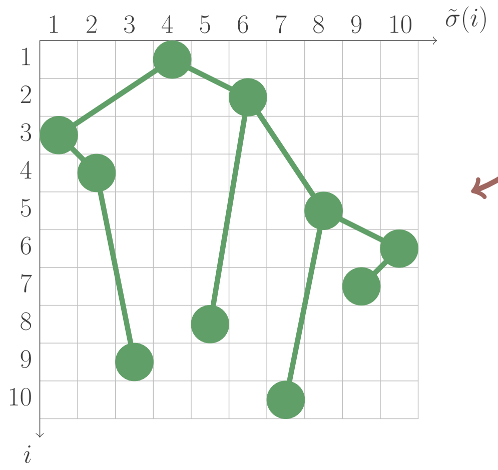
A fun property!

$$\sigma = (7, 5, 10, 9, 3, 1, 2, 6, 8, 4)$$



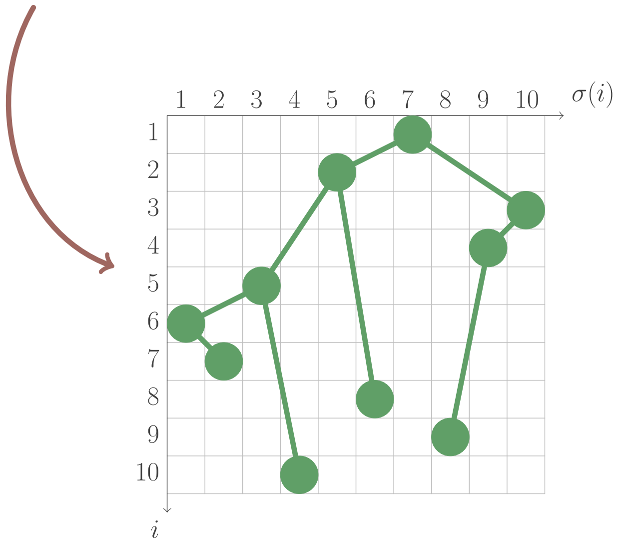
$$\mathbb{P}[X = \sigma] \propto \lambda^{\mathrm{Inv}(\sigma)}$$

$$\tilde{\sigma} = (4, 6, 1, 2, 8, 10, 9, 5, 3, 7)$$



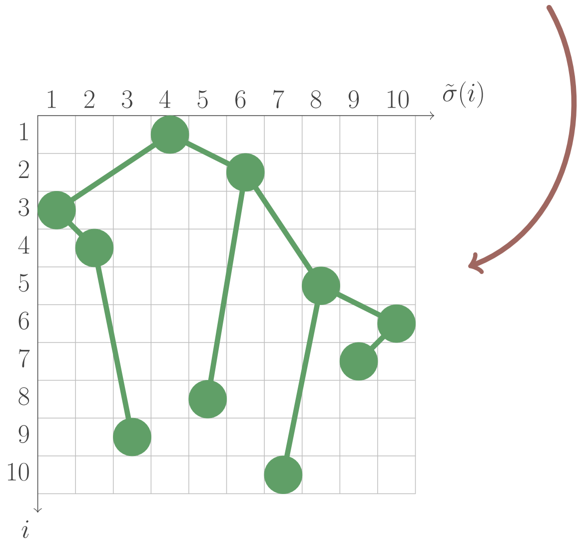
A fun property!

$$\sigma = (7, 5, 10, 9, 3, 1, 2, 6, 8, 4)$$



$$\mathbb{P}[X = \sigma] \propto \lambda^{\mathrm{Inv}(\sigma)}$$

$$\tilde{\sigma} = (4, 6, 1, 2, 8, 10, 9, 5, 3, 7)$$



$$\mathbb{P}[\tilde{X} = \sigma] \propto \lambda^{\mathrm{Inv}(\tilde{\sigma})}$$

$$\propto \lambda^{\binom{n}{2} - \mathrm{Inv}(\sigma)}$$

$$\propto (1/\lambda)^{\mathrm{Inv}(\sigma)}$$

Content Table

Height of random binary search trees

Theorem (Devroye [1986])

Write H_n for the height of a binary search tree of drawn from a random uniform permutation of size n. Then

$$\frac{H_n}{c^* \log n} \xrightarrow{\mathbb{P}} 1\,,$$

where $c^* = 4.311...$ is the unique solution to $c\log(2e/c) = 1$ with $c \geq 2$.

Height of random binary search trees

Theorem (Devroye [1986])

Write H_n for the height of a binary search tree of drawn from a random uniform permutation of size n. Then

$$\frac{H_n}{c^* \log n} \xrightarrow{\mathbb{P}} 1\,,$$

where $c^* = 4.311...$ is the unique solution to $c \log(2e/c) = 1$ with $c \geq 2$.

- The asymptotic behaviour is close to the optimal height, $\lceil \log_2 n \rceil$.

Height of random binary search trees

Theorem (Devroye [1986])

Write H_n for the height of a binary search tree of drawn from a random uniform permutation of size n. Then

$$\frac{H_n}{c^* \log n} \xrightarrow{\mathbb{P}} 1\,,$$

where $c^* = 4.311...$ is the unique solution to $c \log(2e/c) = 1$ with $c \geq 2$.

- The asymptotic behaviour is close to the optimal height, $\lceil \log_2 n \rceil$.
- c^* relates to properties of branching processes.

Height of Mallows trees

Height of Mallows trees

Theorem (Addario-Berry and [2021])

Write $H_{n,\lambda}$ for the height of a Mallows tree with parameters n and $\lambda \in [0,1]$. Then

$$\frac{H_{n,\lambda}}{n(1-\lambda)+c^* \log n} \xrightarrow{\mathbb{P}} 1\,.$$

Height of Mallows trees

Theorem (Addario-Berry and [2021])

Write $H_{n,\lambda}$ for the height of a Mallows tree with parameters n and $\lambda \in [0,1]$. Then

$$\frac{H_{n,\lambda}}{n(1-\lambda)+c^*\log n} \xrightarrow{\mathbb{P}} 1\,.$$

- When $1-\lambda \ll \log n/n$, the height looks like a RBST and behaves as $c^*\log n$.

Theorem (Addario-Berry and 👤 [2021])

Write $H_{n,\lambda}$ for the height of a Mallows tree with parameters n and $\lambda \in [0,1]$. Then

$$\frac{H_{n,\lambda}}{n(1-\lambda) + c^* \log n} \xrightarrow{\mathbb{P}} 1\,.$$

- When $1 - \lambda \ll \log n/n$, the height looks like a RBST and behaves as $c^* \log n$.
- When $1 - \lambda \gg \log n/n$, the height looks "linear" and behaves as $n(1-\lambda)$.

Height of Mallows trees

Theorem (Addario-Berry and 🧍 [2021])

Write $H_{n,\lambda}$ for the height of a Mallows tree with parameters n and $\lambda \in [0,1]$. Then

$$\frac{H_{n,\lambda}}{n(1-\lambda) + c^* \log n} \xrightarrow{\mathbb{P}} 1\,.$$

- When $1-\lambda \ll \log n/n$, the height looks like a RBST and behaves as $c^* \log n$.
- When $1-\lambda \gg \log n/n$, the height looks "linear" and behaves as $n(1-\lambda)$.
- When $1-\lambda \simeq \log n/n$, both terms contribute.

Height of record-biased trees

Theorem ([2023+])

Write $H_{n,\lambda}$ for the height of a record-biased tree with parameters n and $\lambda \in [0, \infty)$. Then

$$\frac{H_{n,\lambda}}{\max\left\{c^* \log n, \lambda \log\left(1 + \frac{n}{\lambda}\right)\right\}} \xrightarrow{\mathbb{P}} 1\,.$$

Height of record-biased trees

Theorem (👤 [2023+])

Write $H_{n,\lambda}$ for the height of a record-biased tree with parameters n and $\lambda \in [0, \infty)$. Then

$$\frac{H_{n,\lambda}}{\max\left\{c^* \log n, \lambda \log\left(1 + \frac{n}{\lambda}\right)\right\}} \xrightarrow{\mathbb{P}} 1\,.$$

- When $\lambda \leq c^*$, the height looks like a RBST and behaves as $c^* \log n$.

Height of record-biased trees

Theorem (👤 [2023+])

Write $H_{n,\lambda}$ for the height of a record-biased tree with parameters n and $\lambda \in [0, \infty)$. Then

$$\frac{H_{n,\lambda}}{\max\left\{c^* \log n, \lambda \log\left(1 + \frac{n}{\lambda}\right)\right\}} \xrightarrow{\mathbb{P}} 1\,.$$

- When $\lambda \leq c^*$, the height looks like a RBST and behaves as $c^* \log n$.
- When $\lambda \geq c^*$, the height behaves as $\lambda \log(1 + n/\lambda)$ and corresponds to the number of records.

Theorem (🧍 [2023+])

Write $H_{n,\lambda}$ for the height of a record-biased tree with parameters n and $\lambda \in [0, \infty)$. Then

$$\frac{H_{n,\lambda}}{\max\left\{c^* \log n, \lambda \log\left(1 + \frac{n}{\lambda}\right)\right\}} \xrightarrow{\mathbb{P}} 1\,.$$

- When $\lambda \leq c^*$, the height looks like a RBST and behaves as $c^* \log n$.
- When $\lambda \geq c^*$, the height behaves as $\lambda \log(1+n/\lambda)$ and corresponds to the number of records.
- When λ is fixed, the height behaves as $\max\{c^*, \lambda\} \log n$.

Height of random models of binary search trees

Summary

For random binary search trees

$$H_n \simeq c^* \log n\,.$$

For Mallows trees

$$H_{n,\lambda} \simeq n(1-\lambda) + c^* \log n\,.$$

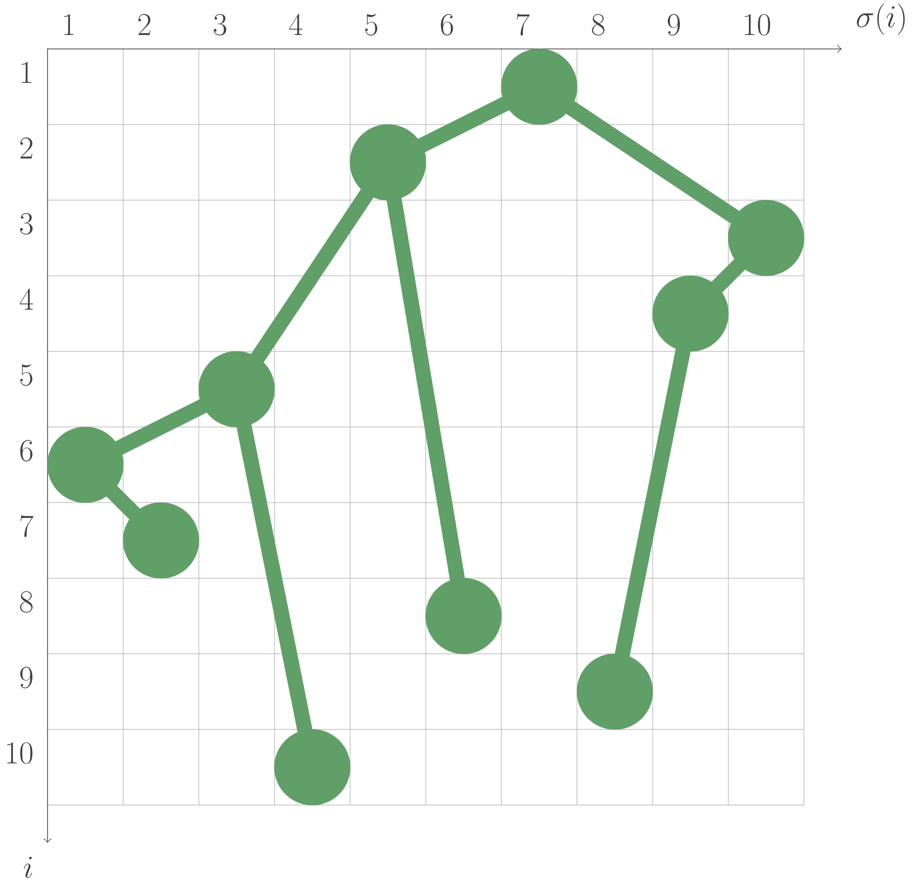
For record-biased trees

$$H_{n,\lambda} \simeq \max\left\{c^* \log n, \lambda \log\left(1 + n/\lambda\right)\right\} \simeq \max\{c^*, \lambda\} \log n\,.$$

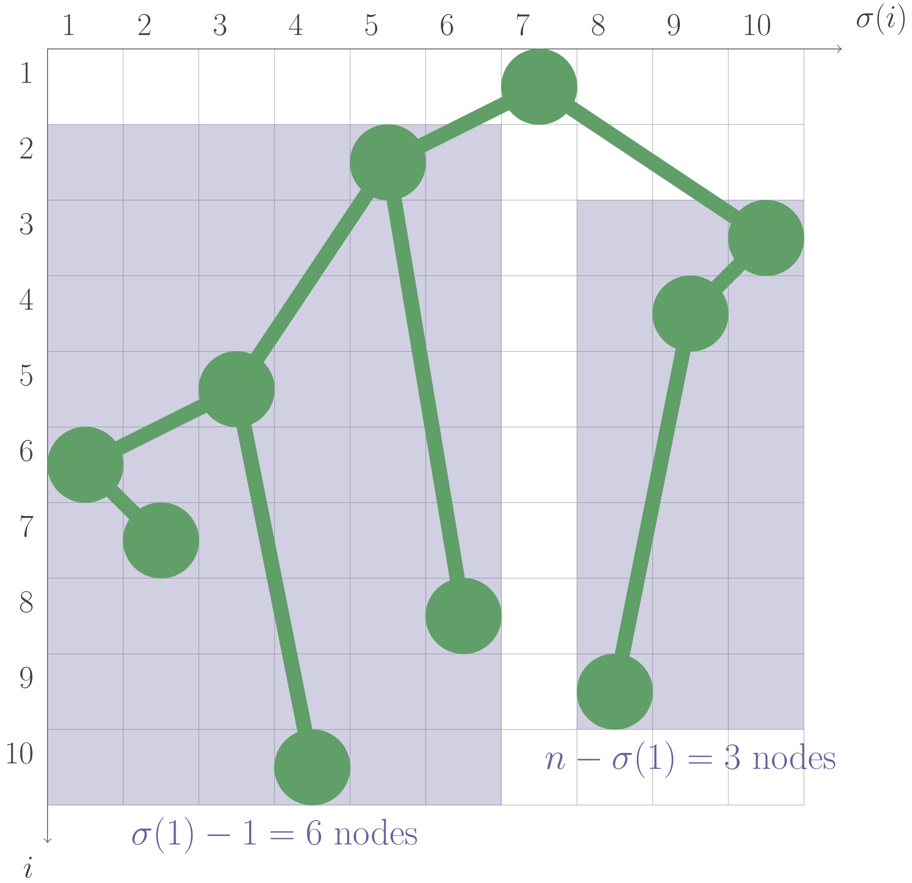
Content Table

An important remark

An important remark

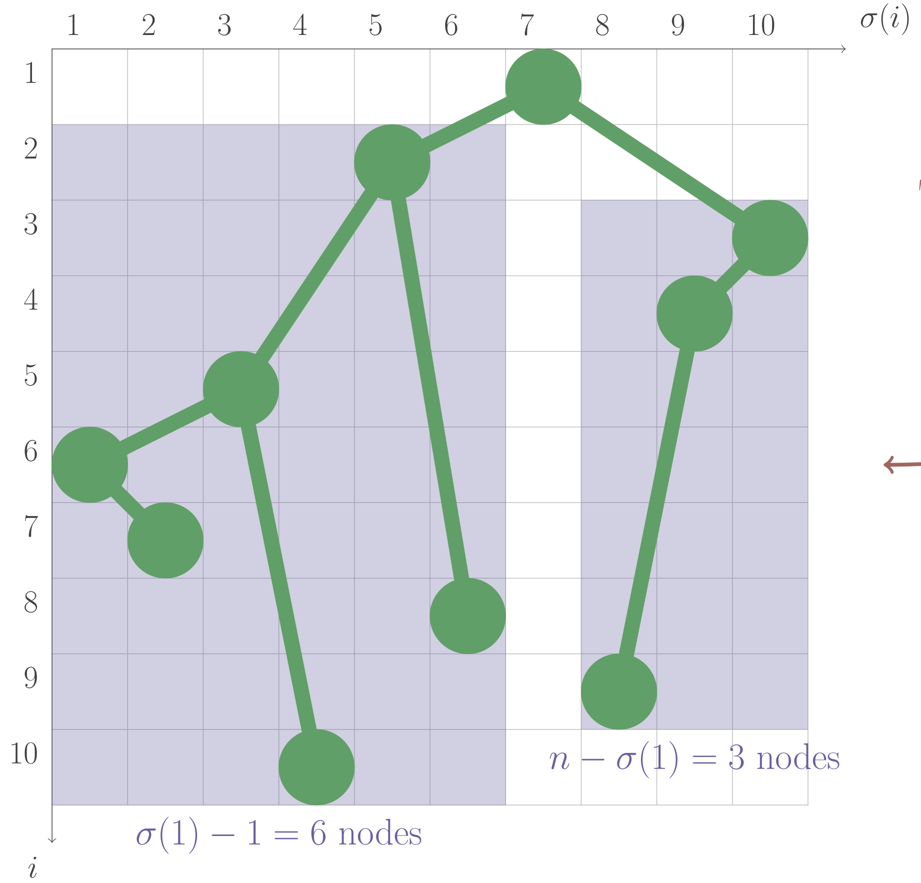


An important remark

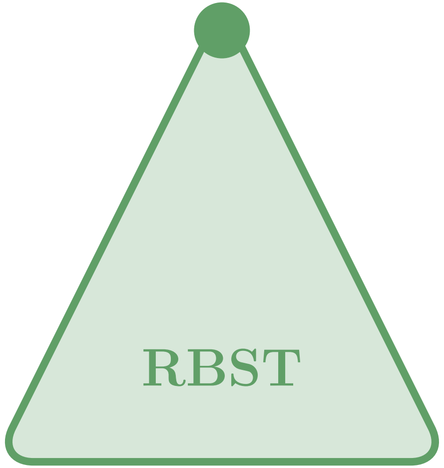


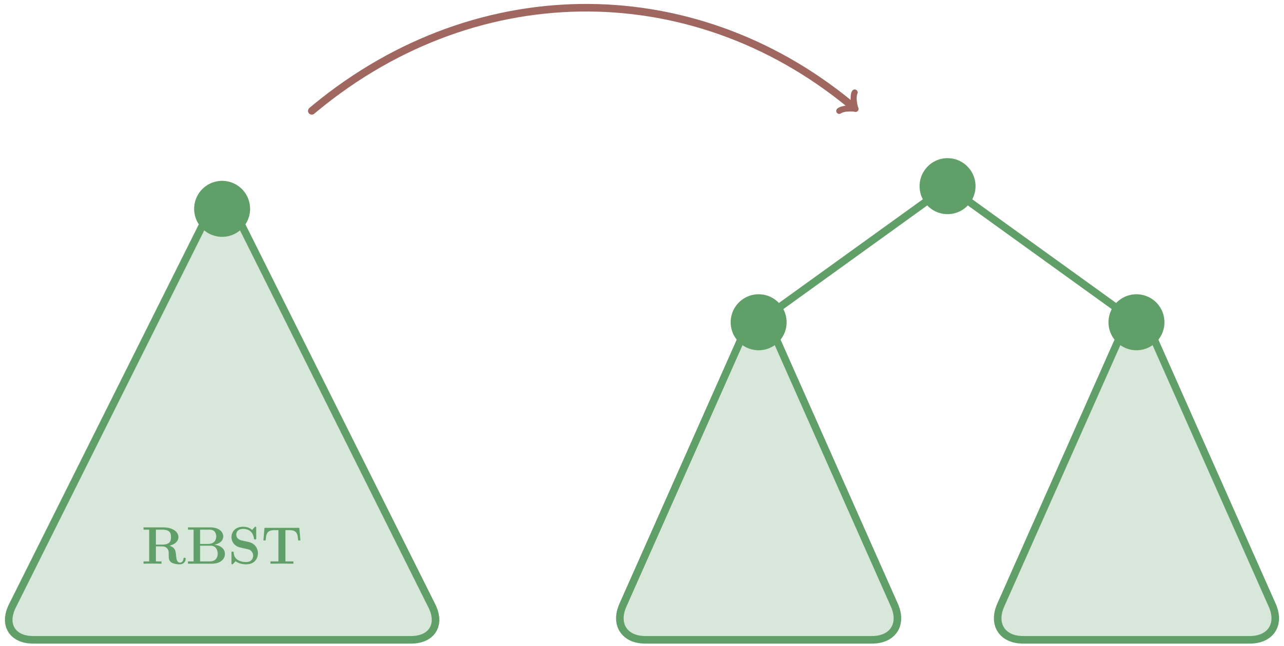
An important remark

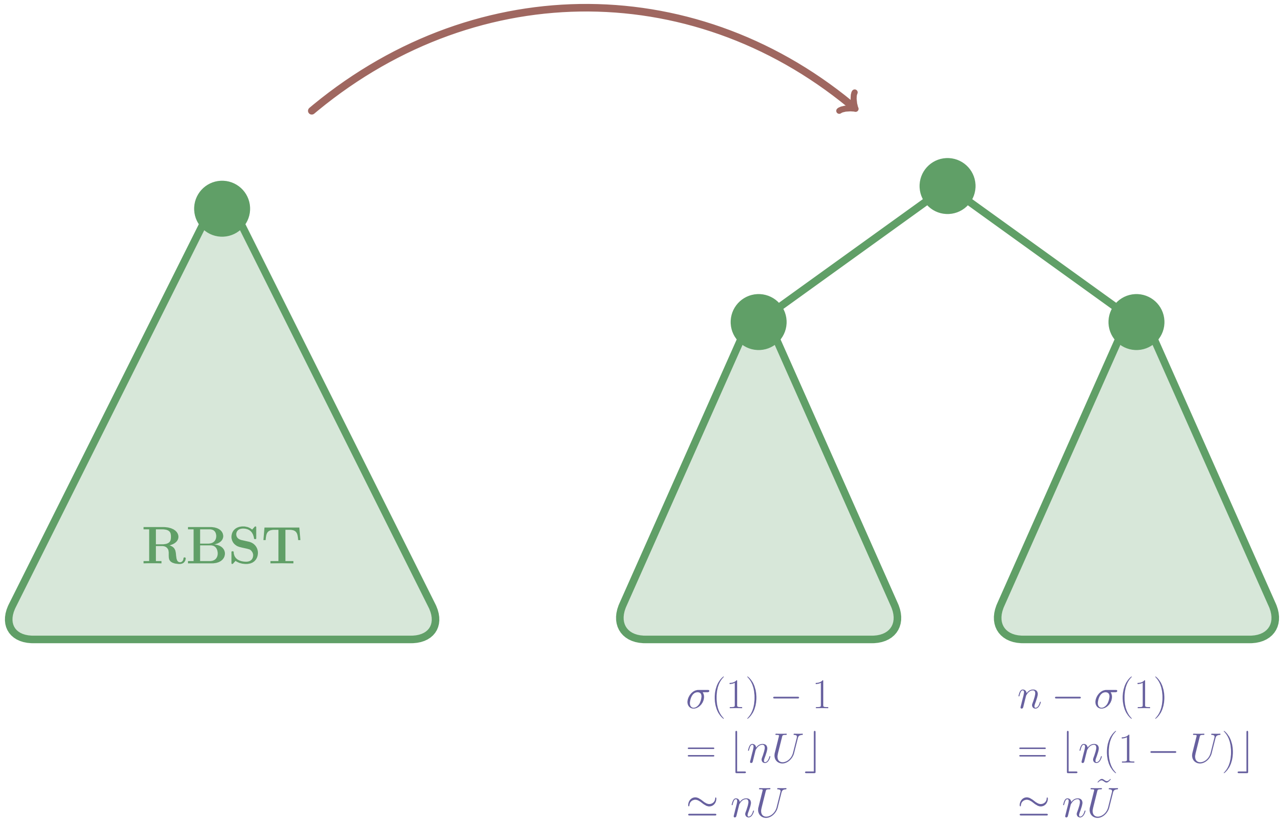
Tree of $\sigma_- = \big(\sigma(i) : \sigma(i) < \sigma(1)\big)$



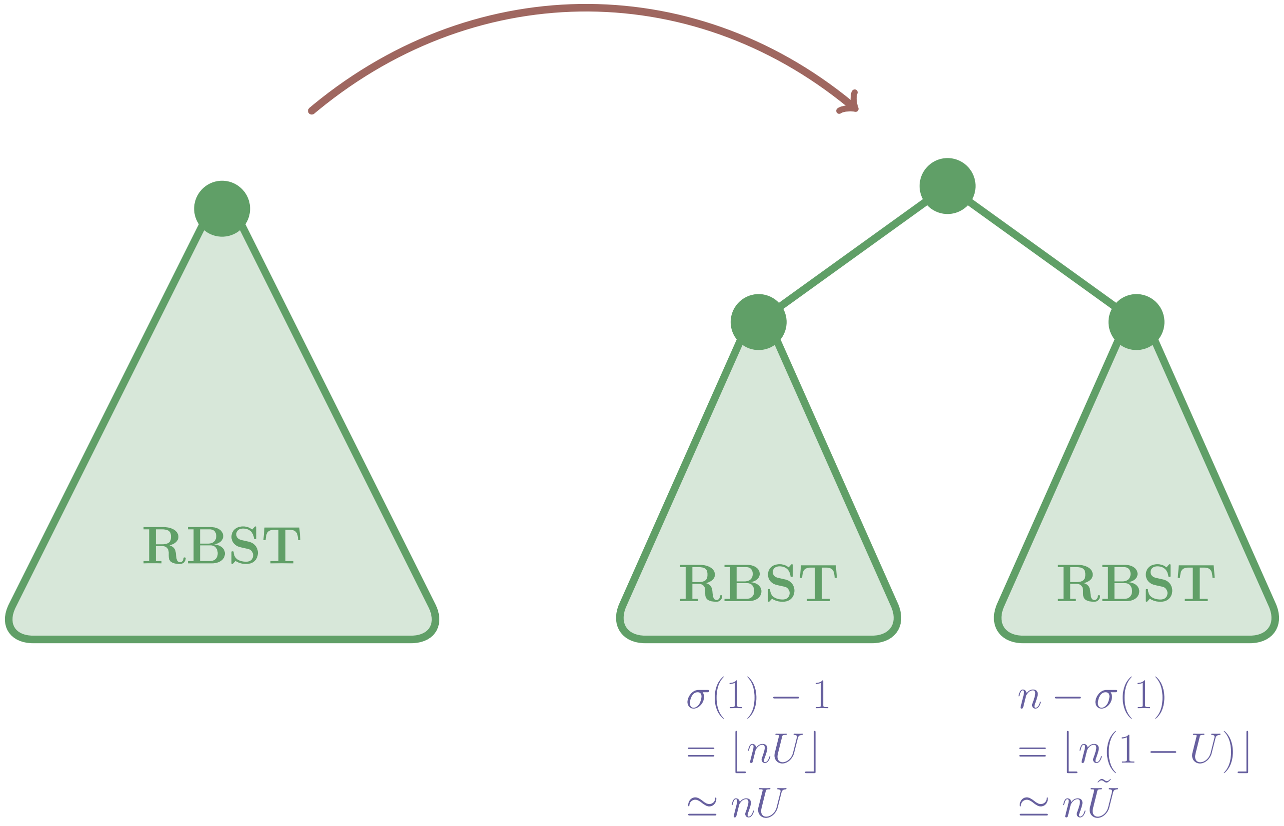
Tree of $\sigma_+ = \big(\sigma(i) : \sigma(i) > \sigma(1)\big)$


RBST


RBST

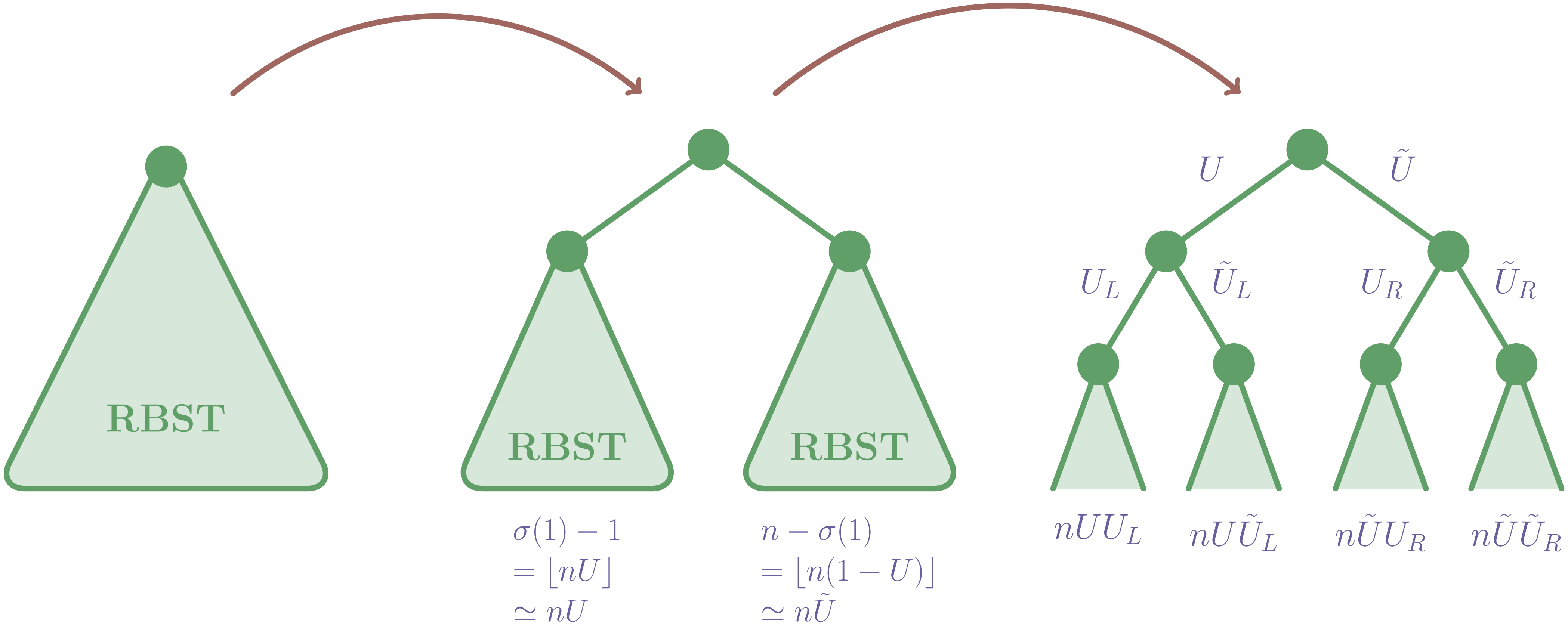


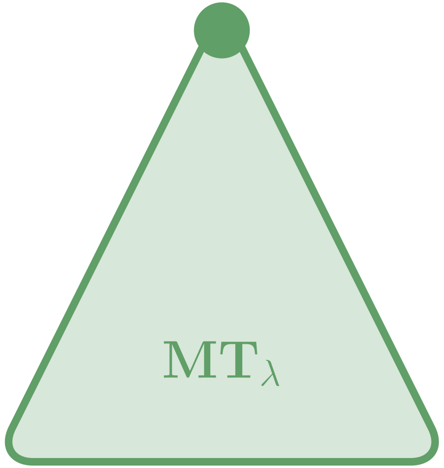
RBST
$\sigma(1) - 1$
$= \lfloor nU \rfloor$
$\simeq nU$
$n - \sigma(1)$
$= \lfloor n(1-U) \rfloor$
$\simeq n\tilde{U}$

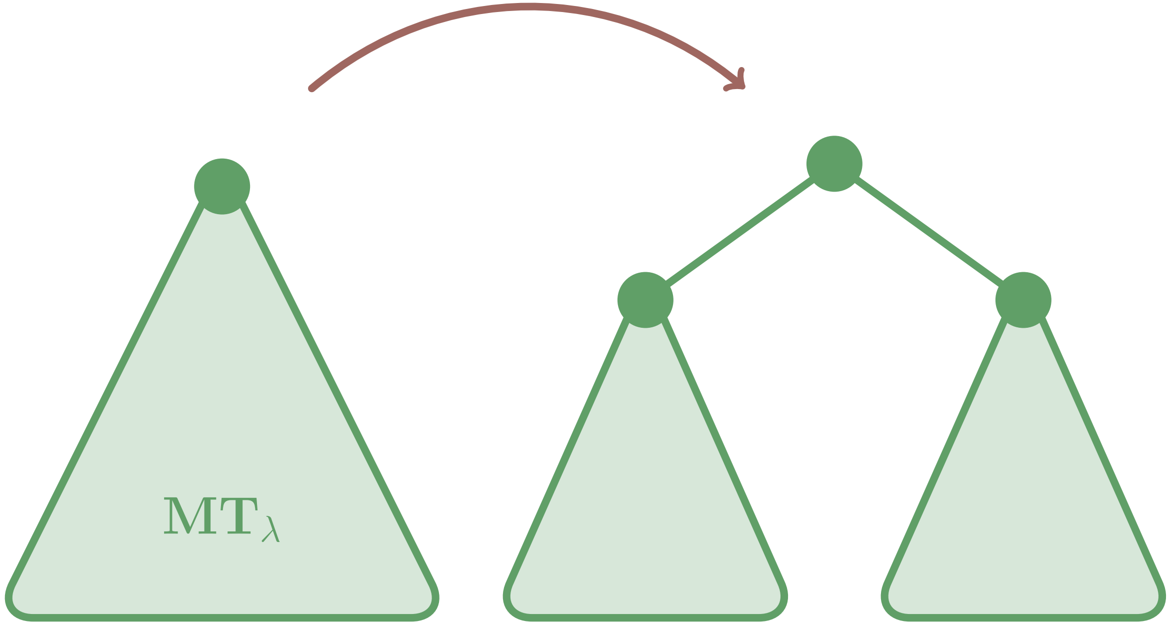

RBST
RBST
RBST
$\sigma(1) - 1$
$= \lfloor nU \rfloor$
$\simeq nU$
$n - \sigma(1)$
$= \lfloor n(1-U) \rfloor$
$\simeq n\tilde{U}$

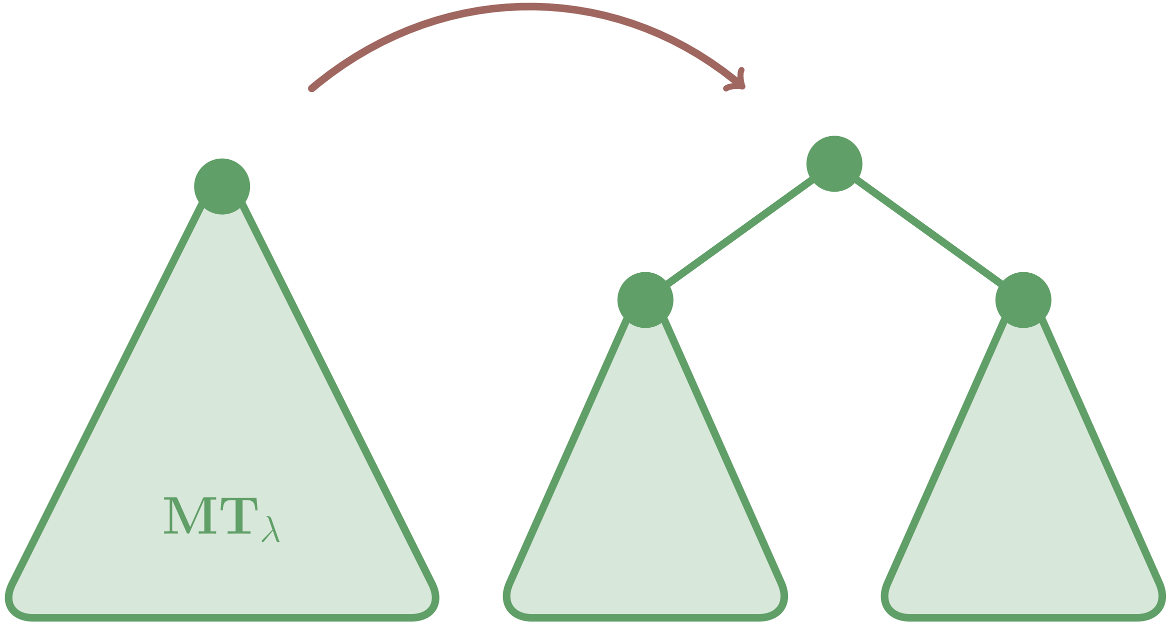
Random binary search trees

$H_n \simeq c^* \log n$




$\mathbf{MT}_\lambda$

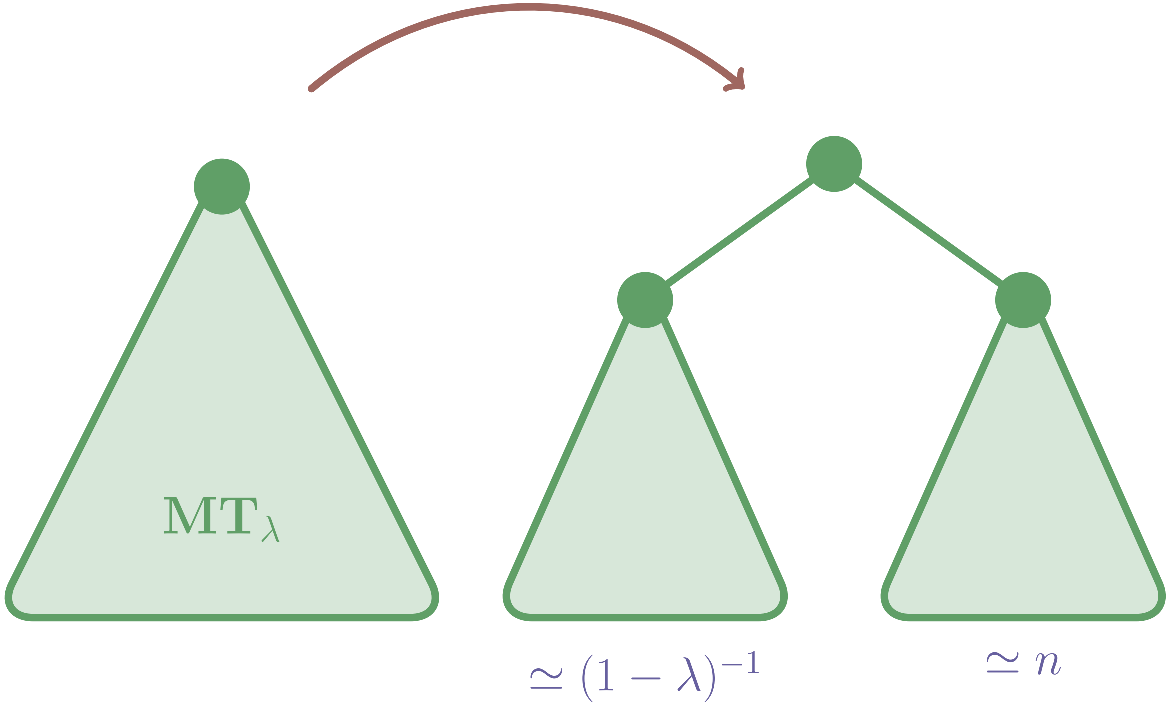

$\mathbf{MT}_\lambda$



$$\begin{aligned}\sigma(1) &= \mathrm{Geom}(\lambda \mid n) \\ &\simeq n \wedge (1-\lambda)^{-1}\end{aligned}$$

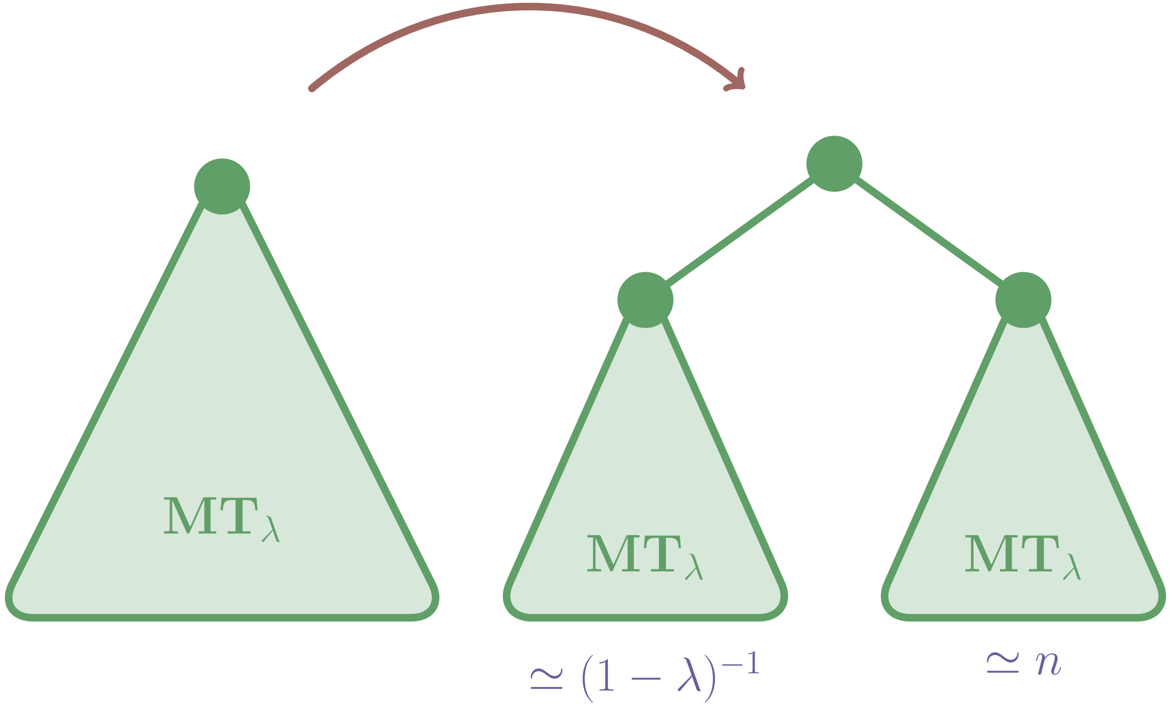
Mallows trees

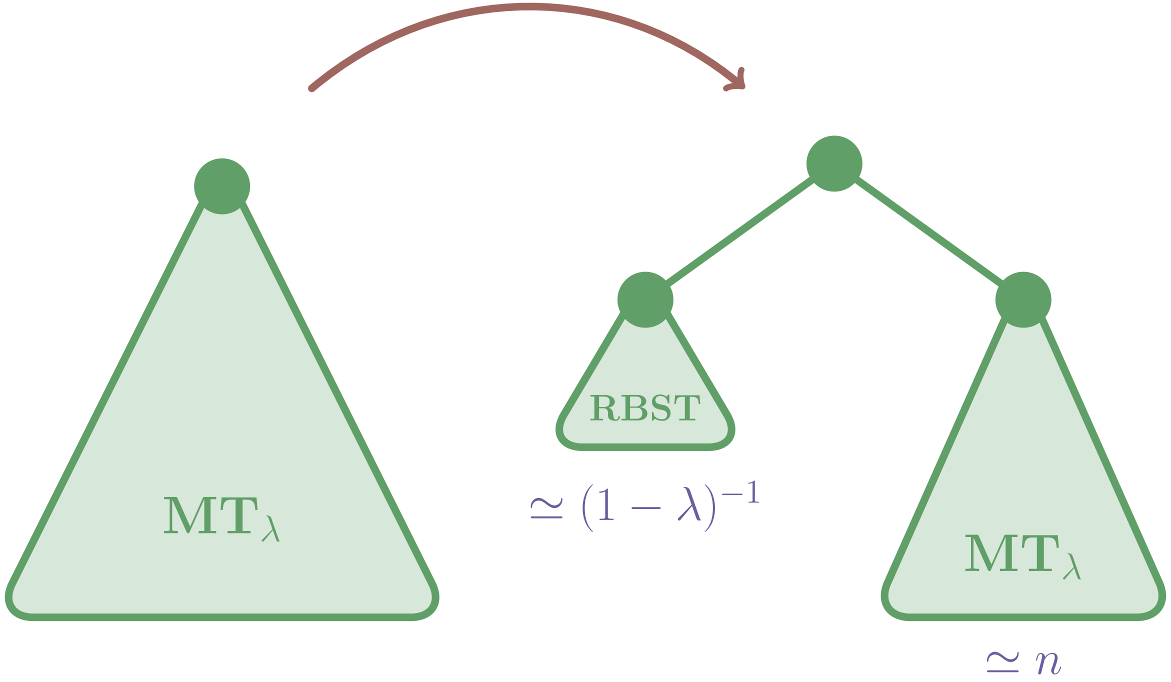
$H_{n,\lambda} \simeq n(1-\lambda) + c^* \log n$



Mallows trees

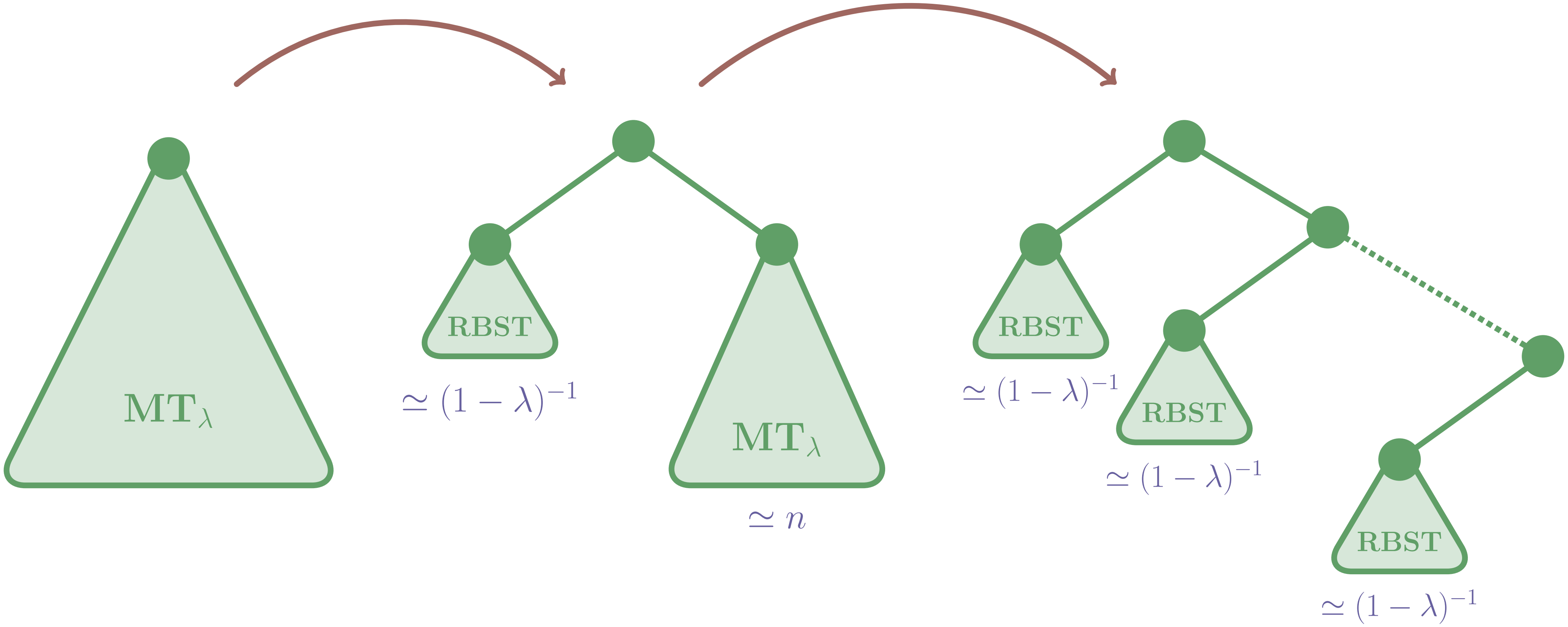
$H_{n,\lambda} \simeq n(1-\lambda) + c^* \log n$

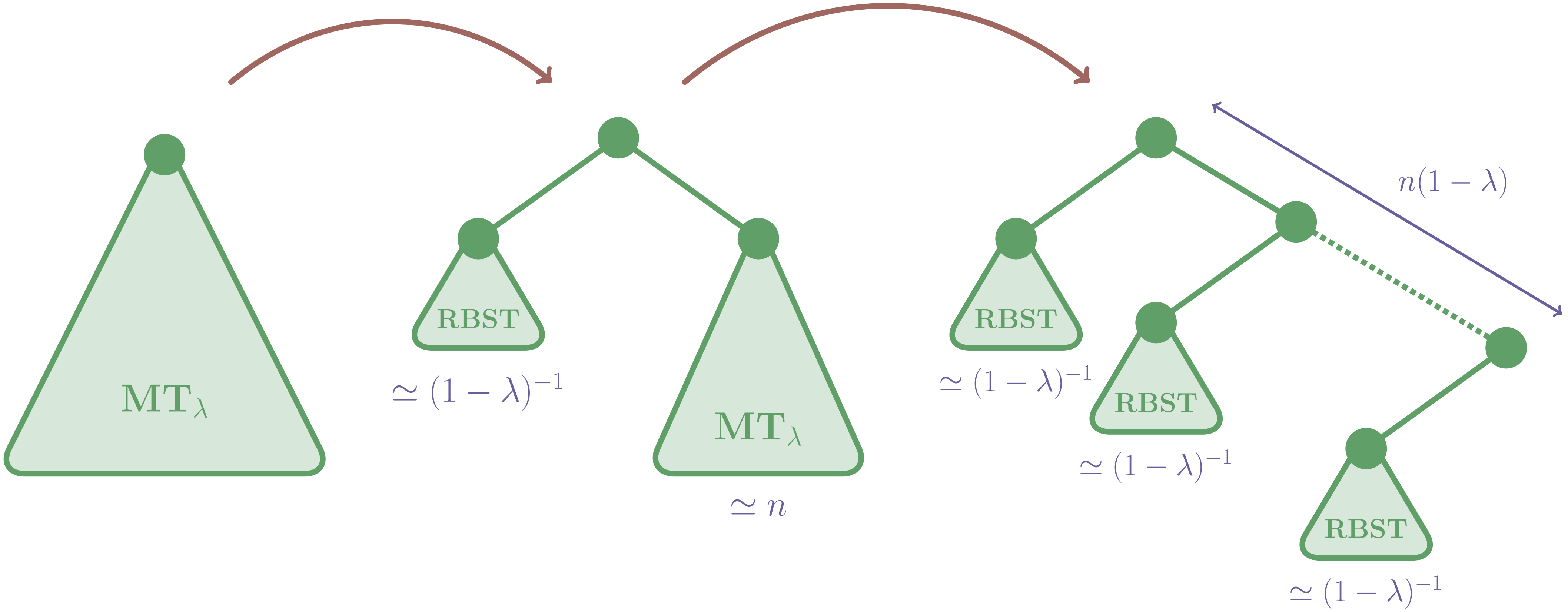



MT$_\lambda$
RBST
$\simeq (1-\lambda)^{-1}$
MT$_\lambda$
$\simeq n$

Mallows trees

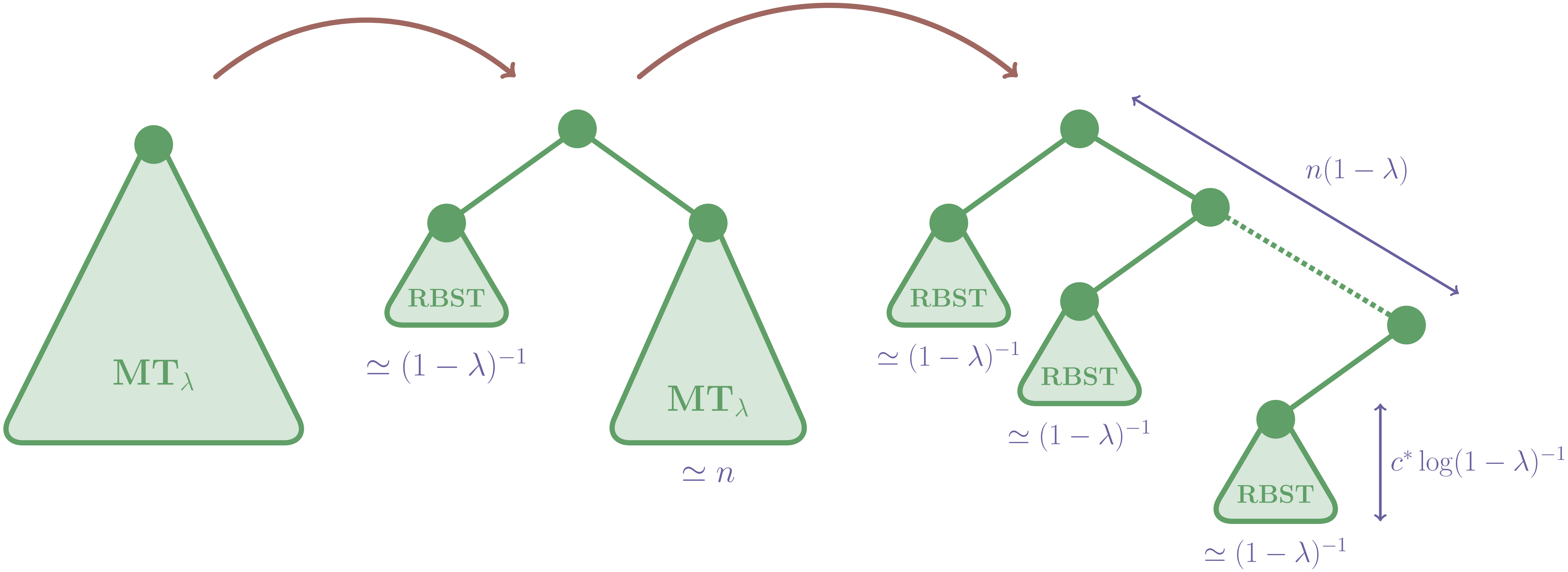
$H_{n,\lambda} \simeq n(1-\lambda) + c^* \log n$




MT$_\lambda$
RBST
$\simeq (1-\lambda)^{-1}$
MT$_\lambda$
$\simeq n$
$n(1-\lambda)$
RBST
$\simeq (1-\lambda)^{-1}$
RBST
$\simeq (1-\lambda)^{-1}$
RBST
$\simeq (1-\lambda)^{-1}$

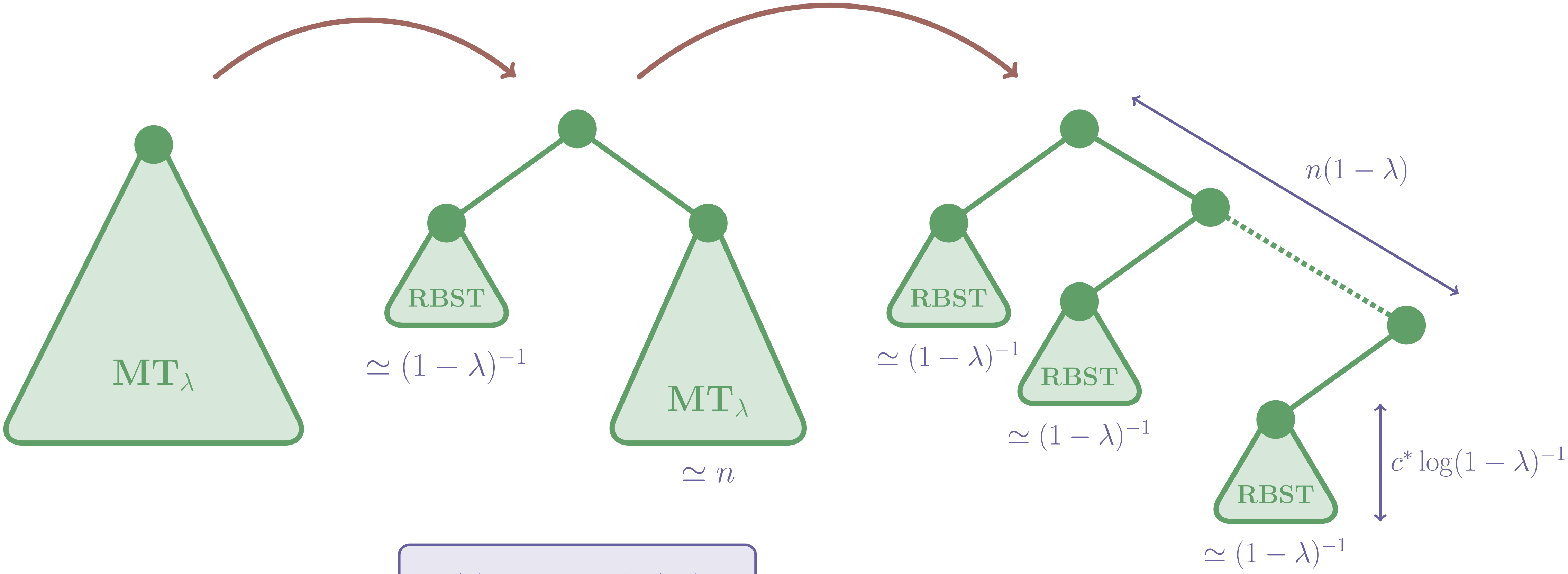
Mallows trees

$H_{n,\lambda} \simeq n(1-\lambda) + c^* \log n$

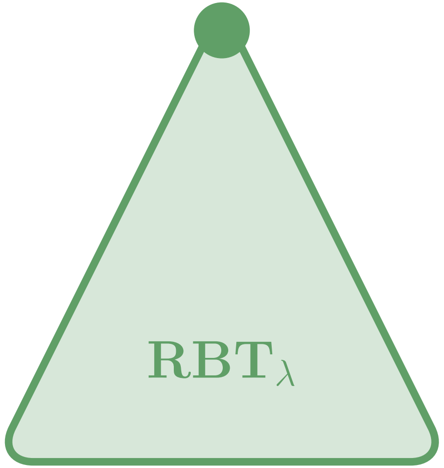


Mallows trees

$$H_{n,\lambda} \simeq n(1-\lambda) + c^* \log n$$

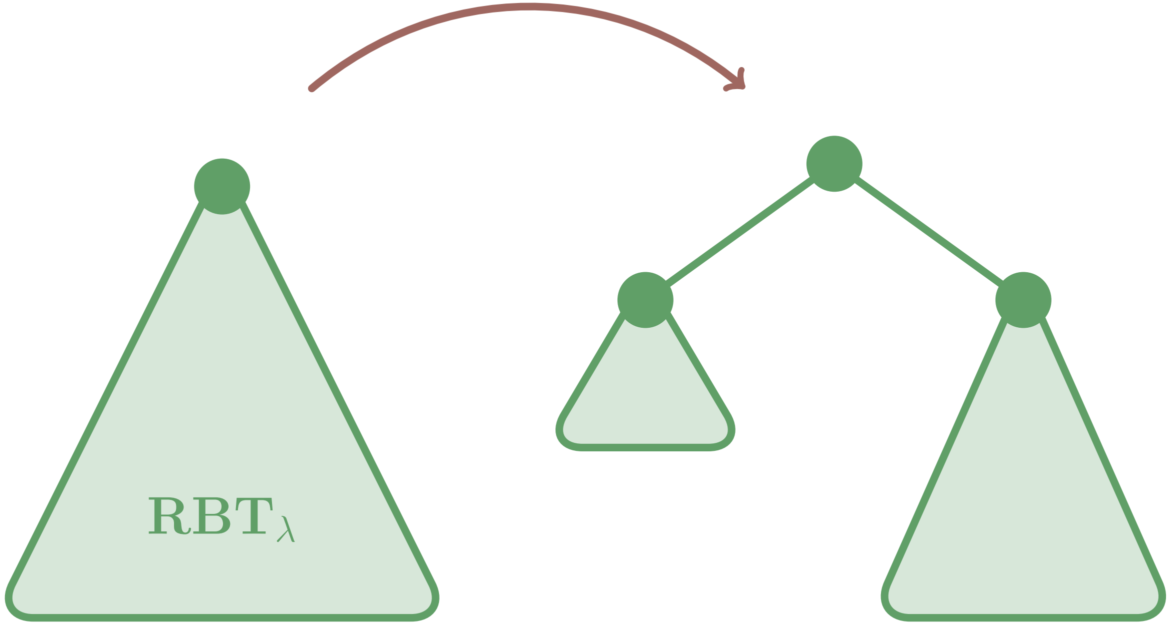


$$\begin{aligned}\sigma(1) &= \mathrm{Geom}(\lambda \mid n) \\ &\simeq n \wedge (1-\lambda)^{-1}\end{aligned}$$

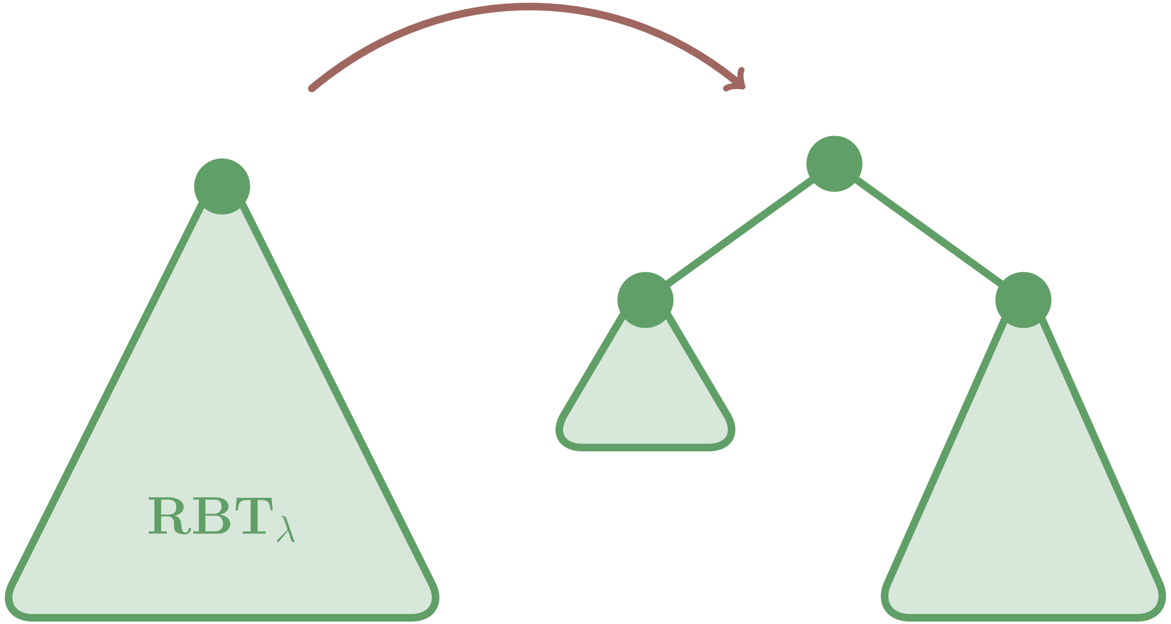

$\mathbf{RBT}_\lambda$

Record-biased trees

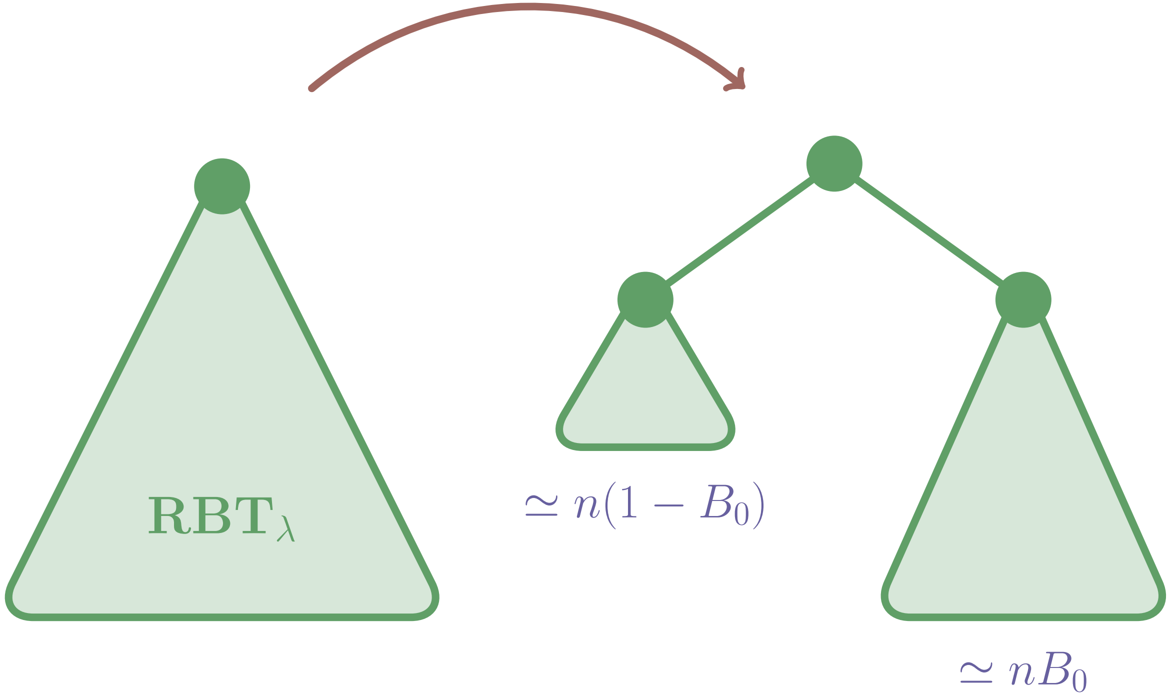
$H_{n,\lambda} \simeq \max\{c^*, \lambda\} \log n$



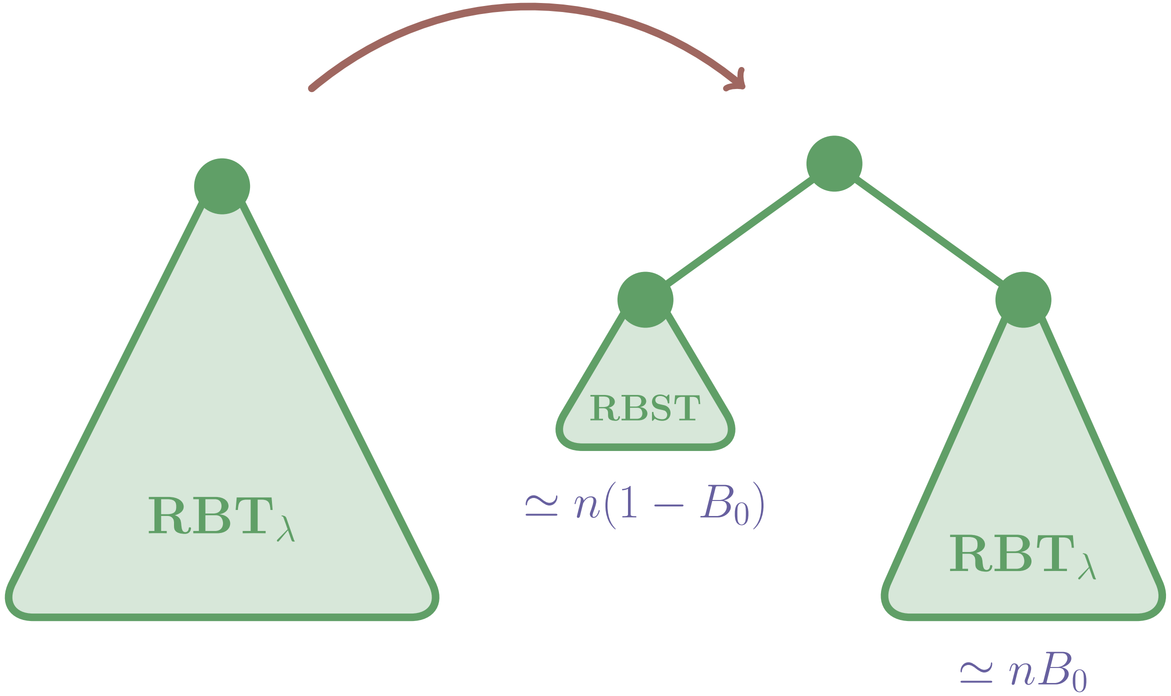
Record-biased trees



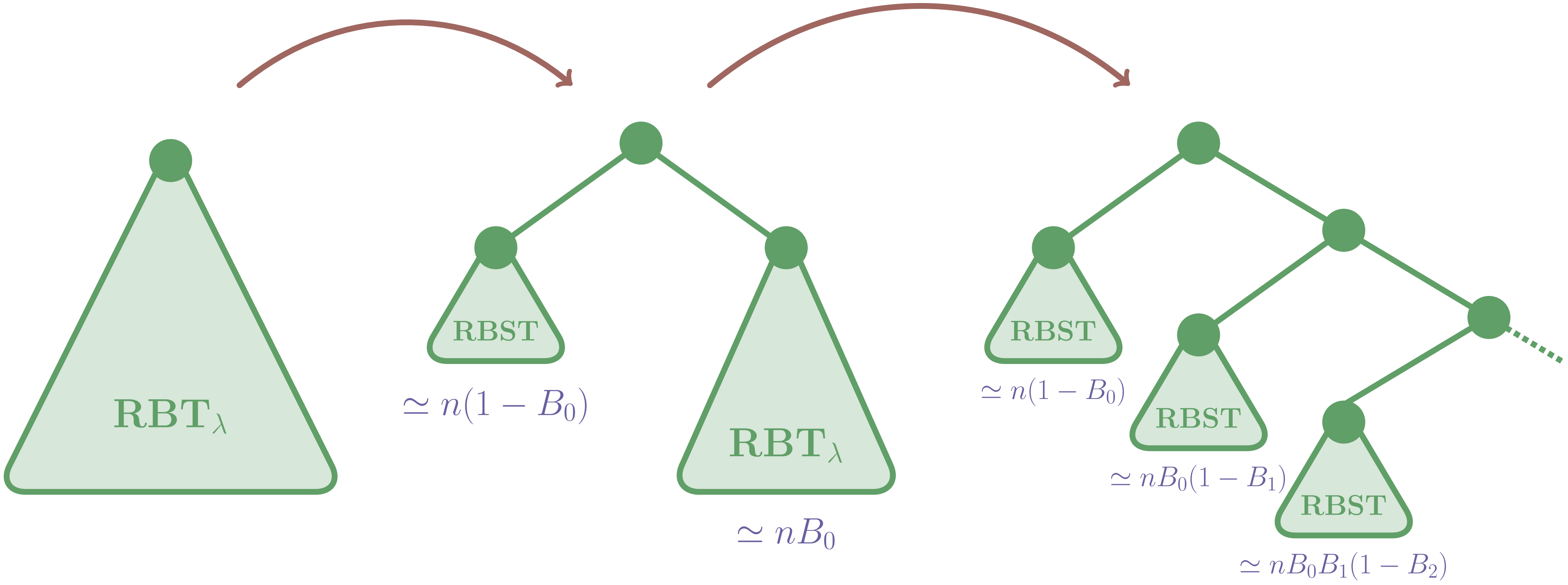
$$\sigma(1) \simeq n \cdot \mathrm{Beta}(1, \lambda) = n \cdot \big(1 - \mathrm{Beta}(\lambda, 1)\big)$$


$\mathbf{RBT}_\lambda$
$\simeq n(1 - B_0)$
$\simeq nB_0$

Record-biased trees

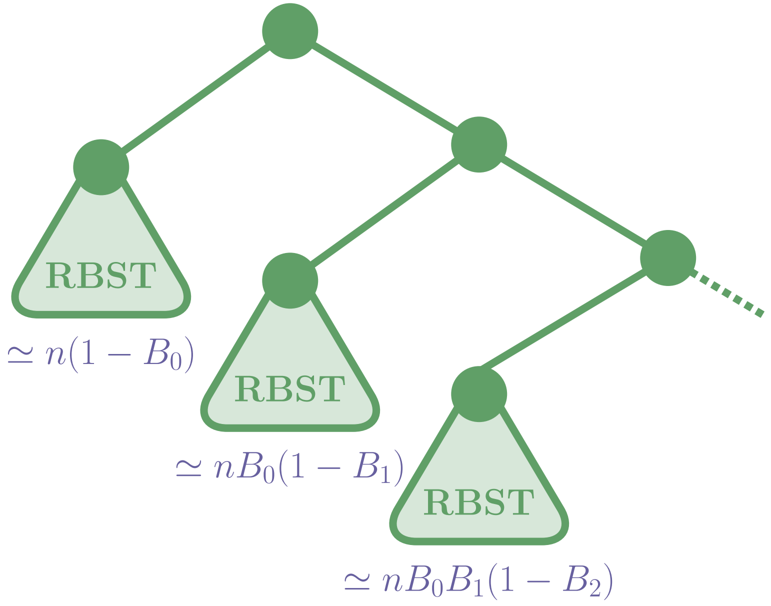


Record-biased trees



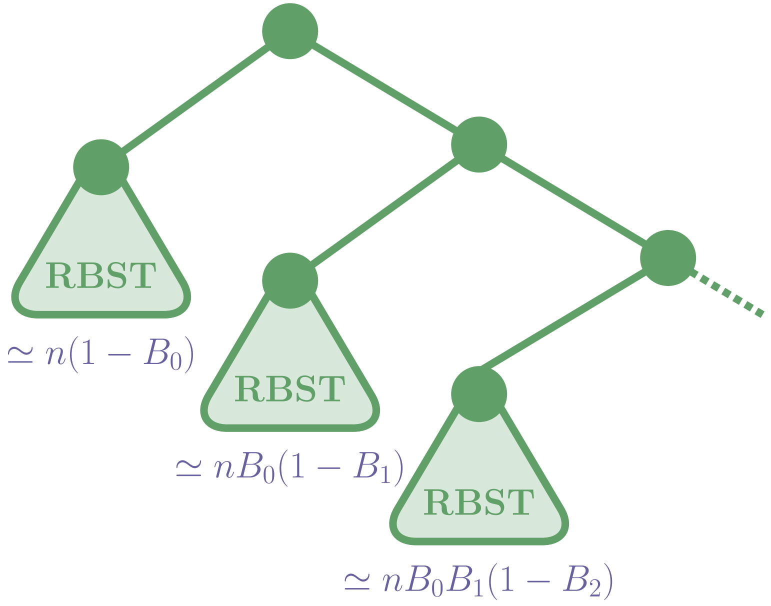
Record-biased trees

$H_{n,\lambda} \simeq \max\{c^*, \lambda\} \log n$



- The contribution to the height of the i-th left subtree is

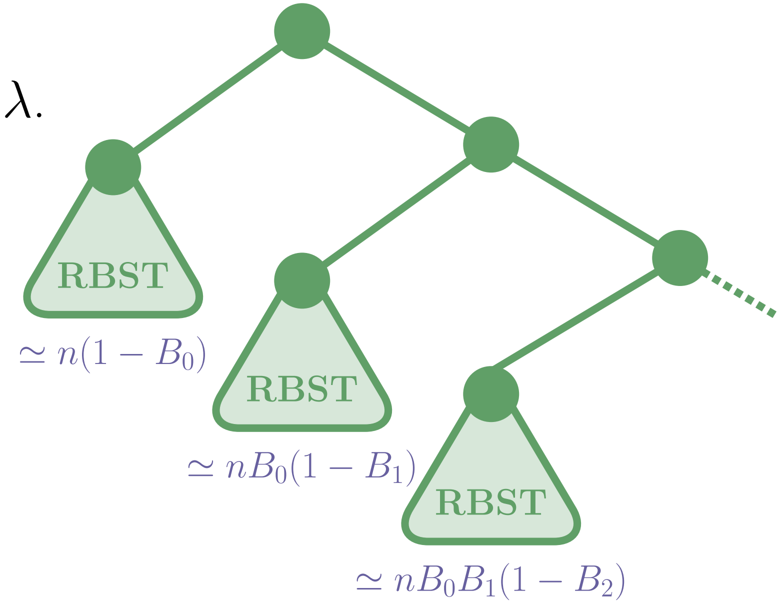
$$i + 1 + c^* \log \left(n(1 - B_i)\Pi_{j<i} B_j \right) \simeq i + c^* \log n + c^* \Sigma_{j<i} \log B_j \,.$$



- The contribution to the height of the i-th left subtree is

$$i + 1 + c^* \log \left(n(1 - B_i)\Pi_{j<i} B_j\right) \simeq i + c^* \log n + c^* \Sigma_{j<i} \log B_j \, .$$

- We have $\mathbb{E}[\log B_j] = -1/\lambda$ and thus $\Sigma_{j<i} \log B_j \simeq -i/\lambda$.



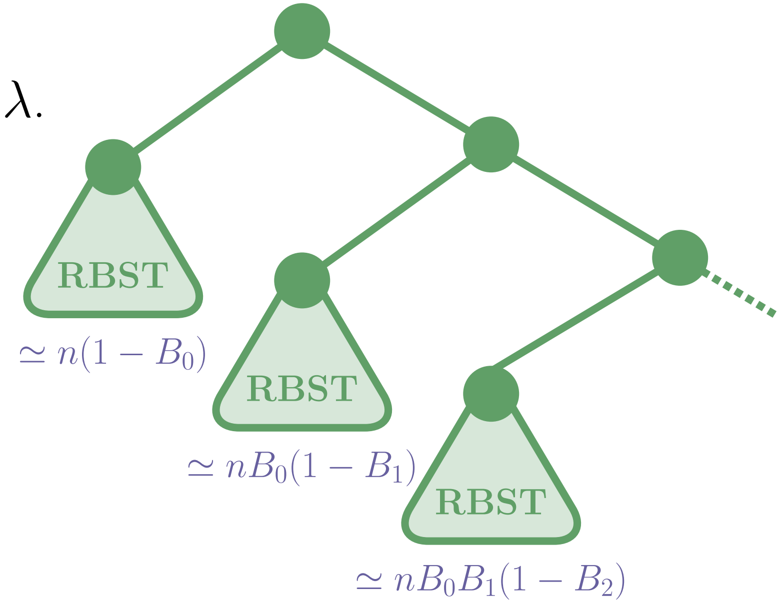
- The contribution to the height of the i-th left subtree is

$$i + 1 + c^* \log \Big(n(1 - B_i)\Pi_{j<i} B_j \Big) \simeq i + c^* \log n + c^* \Sigma_{j<i} \log B_j \,.$$

- We have $\mathbb{E}[\log B_j] = -1/\lambda$ and thus $\Sigma_{j<i} \log B_j \simeq -i/\lambda$.

- The length k of the right path satisfies

$$n\Pi_{j<k} B_j \simeq 1 \;\Leftrightarrow\; \log n - k/\lambda \simeq 0$$
$$\Leftrightarrow\; k \simeq \lambda \log n \,.$$



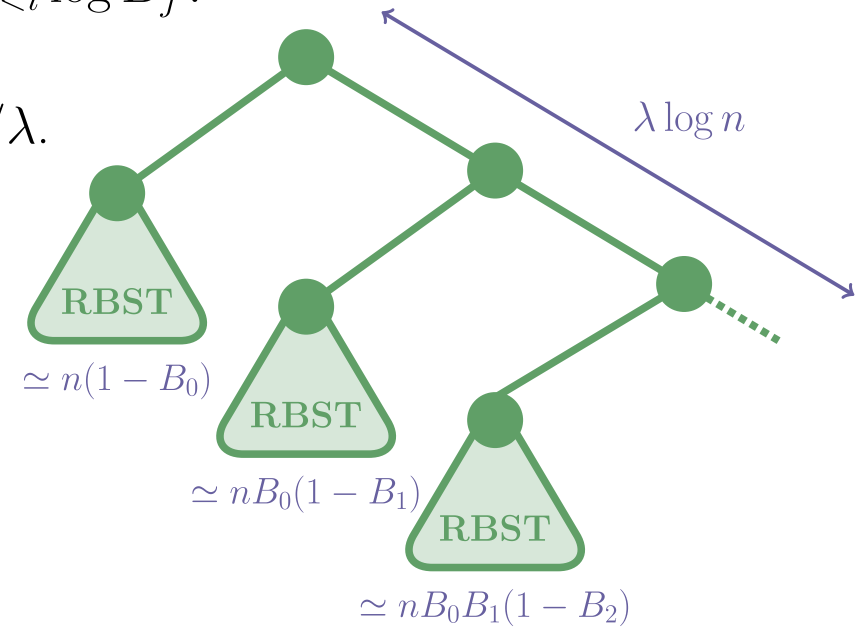
- The contribution to the height of the i-th left subtree is

$$i + 1 + c^* \log \left(n(1 - B_i)\Pi_{j<i} B_j \right) \simeq i + c^* \log n + c^* \Sigma_{j<i} \log B_j \,.$$

- We have $\mathbb{E}[\log B_j] = -1/\lambda$ and thus $\Sigma_{j<i} \log B_j \simeq -i/\lambda$.

- The length k of the right path satisfies

$$n\Pi_{j<k} B_j \simeq 1 \Leftrightarrow \log n - k/\lambda \simeq 0$$
$$\Leftrightarrow k \simeq \lambda \log n \,.$$

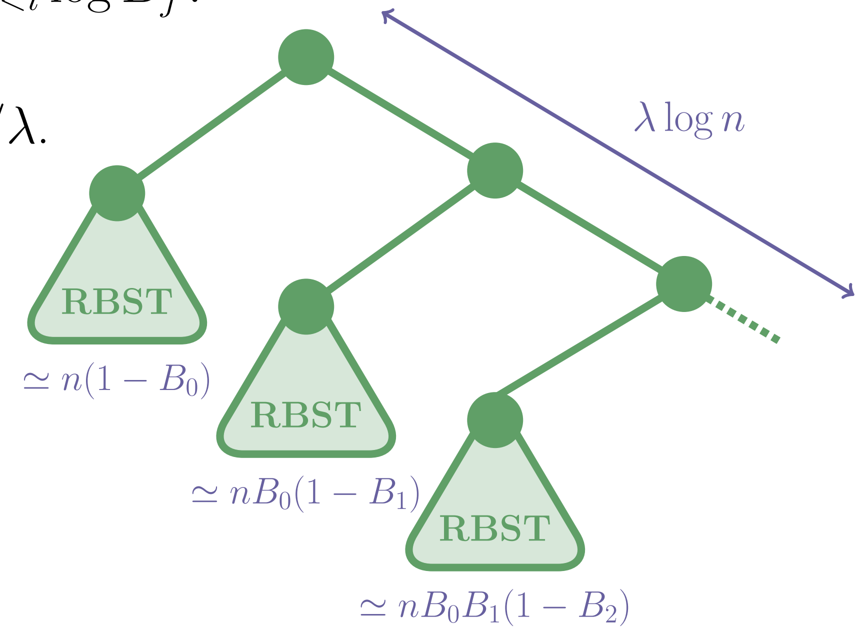


- The contribution to the height of the i-th left subtree is

$$i + 1 + c^* \log \left(n(1 - B_i)\Pi_{j<i}B_j\right) \simeq i + c^* \log n + c^*\Sigma_{j<i} \log B_j \,.$$

- We have $\mathbb{E}[\log B_j] = -1/\lambda$ and thus $\Sigma_{j<i} \log B_j \simeq -i/\lambda$.
- The length k of the right path satisfies

$$n\Pi_{j<k}B_j \simeq 1 \;\Leftrightarrow\; \log n - k/\lambda \simeq 0$$
$$\Leftrightarrow\; k \simeq \lambda \log n \,.$$



$$\rightarrow \quad H_{n,\lambda} \simeq \max_{i \le \lambda \log n} \left\{c^* \log n + i(1 - c^*/\lambda)\right\} \simeq \max\{c^*, \lambda\} \log n \,.$$

Content Table

Open questions

Open questions

- Second order behaviour?

Open questions

- Second order behaviour?
  - Known under some assumptions on λ for Mallows trees.

Open questions

- Second order behaviour?
  - Known under some assumptions on λ for Mallows trees.
  - Unknown for record-biased trees.

Open questions

- Second order behaviour?
  - Known under some assumptions on λ for Mallows trees.
  - Unknown for record-biased trees.
- Other models of permutations?

Open questions

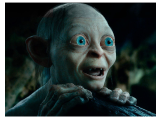
- Second order behaviour?
  - Known under some assumptions on λ for Mallows trees.
  - Unknown for record-biased trees.
- Other models of permutations?
  - Ewens permutations.

Open questions

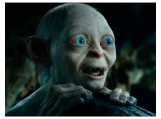
- Second order behaviour?
  - Known under some assumptions on λ for Mallows trees.
  - Unknown for record-biased trees.
- Other models of permutations?
  - Ewens permutations.
  - Random regenerative permutations.

Open questions

- Second order behaviour?
  - Known under some assumptions on λ for Mallows trees.
  - Unknown for record-biased trees.
- Other models of permutations?
  - Ewens permutations.
  - Random regenerative permutations.



- Second order behaviour?
  - Known under some assumptions on λ for Mallows trees.
  - Unknown for record-biased trees.
- Other models of permutations?
  - Ewens permutations.
  - Random regenerative permutations.



Since we know that $H_{n,\lambda} \simeq n(1-\lambda) + c^* \log n$, can we prove that the height of a Mallows tree is stochastically decreasing with respect to $\lambda \in [0,1]$?

References

- **Addario-Berry, L., & Corsini, B. (2021).** The height of Mallows trees. *The Annals of Probability, 49(5), 2220-2271.*
- **Auger, N., Bouvel, M., Nicaud, C., & Pivoteau, C. (2016).** Analysis of algorithms for permutations biased by their number of records. *arXiv preprint arXiv:1605.02905.*
- **Corsini, B. (2021).** The height of record-biased trees. *Random Structures & Algorithms.*
- **Devroye, L. (1986)**. A note on the height of binary search trees. *Journal of the ACM (JACM), 33(3), 489-498.*
- **Ewens, W. J. (1972).** The sampling theory of selectively neutral alleles. *Theoretical population biology, 3(1), 87-112.*
- **Mallows, C. L. (1957).** Non-null ranking models. I. *Biometrika, 44(1/2), 114-130.*
- **Pitman, J., & Tang, W. (2019).** Regenerative random permutations of integers. *The Annals of Probability, 47(3), 1378-1416.*

Thank you!

Thank you!
Thank you!
Thank you!