The permuton tree

Benoît Corsini

joint work with Victor Dubach
and Valentin Féray

TU/e

Table of Content

Binary search trees

Permutons

Our results

The conditions

Table of Content

Binary search trees

Permutons

Our results

The conditions

Motivation

Motivation

Motivation

‘ ’

Motivation

‘’

Motivation

‘ ’

‘0’

Motivation

‘’

‘0’

Motivation

‘’

‘0’

‘1’

Motivation

‘’

‘0’

‘1’

‘00’

Motivation

‘’

‘0’

‘1’

‘00’

‘01’

Motivation

‘’

‘0’

‘1’

‘00’

‘01’

‘10’

Motivation

‘’

‘0’

‘1’

‘00’

‘01’

‘10’

‘11’

Motivation

Motivation

58m

51m

116m

108m

99m

3357m

330m

Motivation

58m

51m

116m

108m

99m

3357m

330m

Motivation

58m

‘’

51m

116m

108m

99m

3357m

330m

Motivation

58m

‘ ’

51m

‘0’

116m

108m

99m

3357m

330m

58m

‘ ’

51m

‘0’

116m

‘1’

108m

99m

3357m

330m

Motivation

58m

‘’

51m

‘0’

116m

‘1’

108m

‘10’

99m

3357m

330m

Motivation

58m

‘’

51m

‘0’

116m

‘1’

108m

‘10’

99m

‘100’

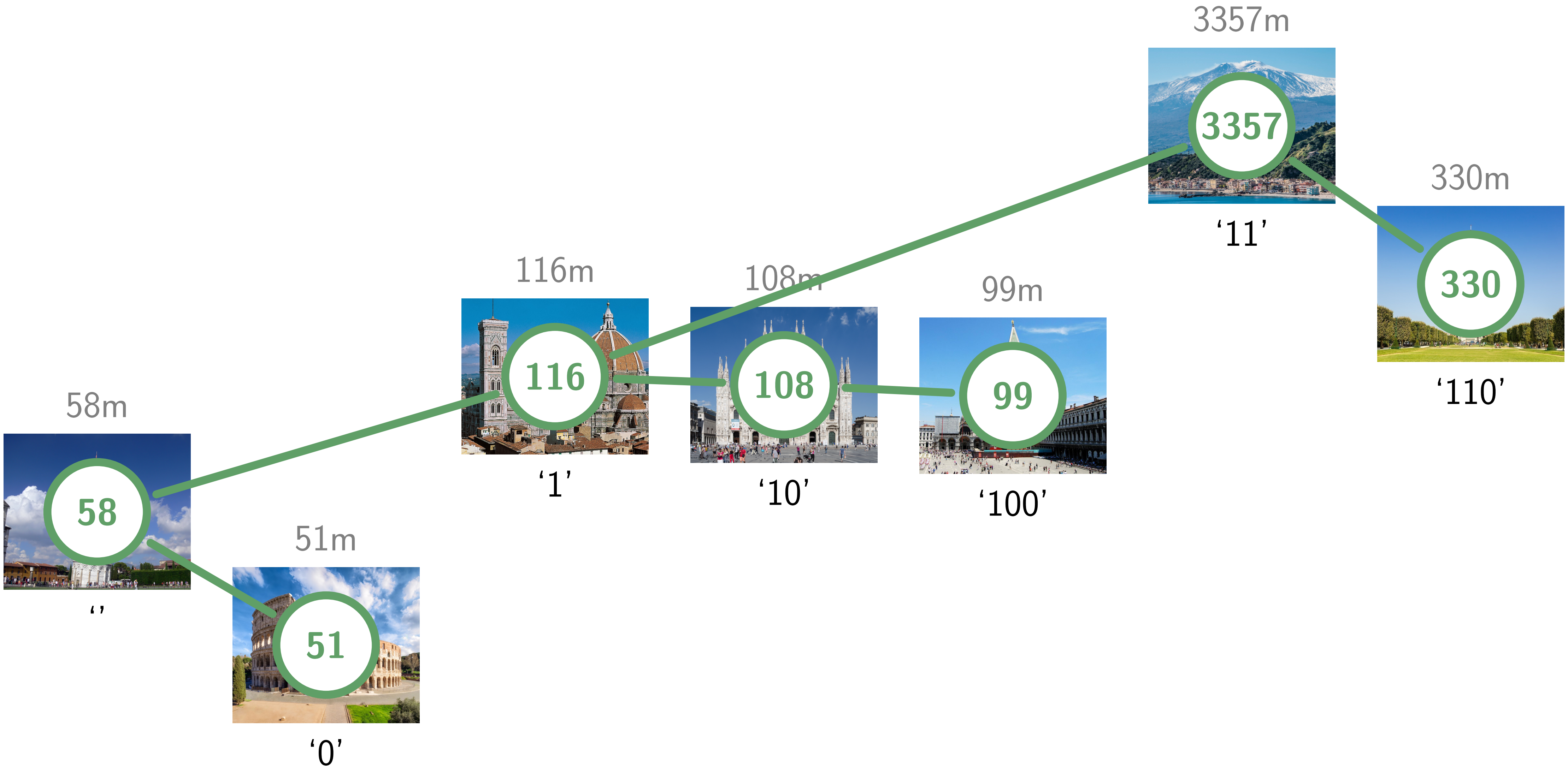
3357m

‘11’

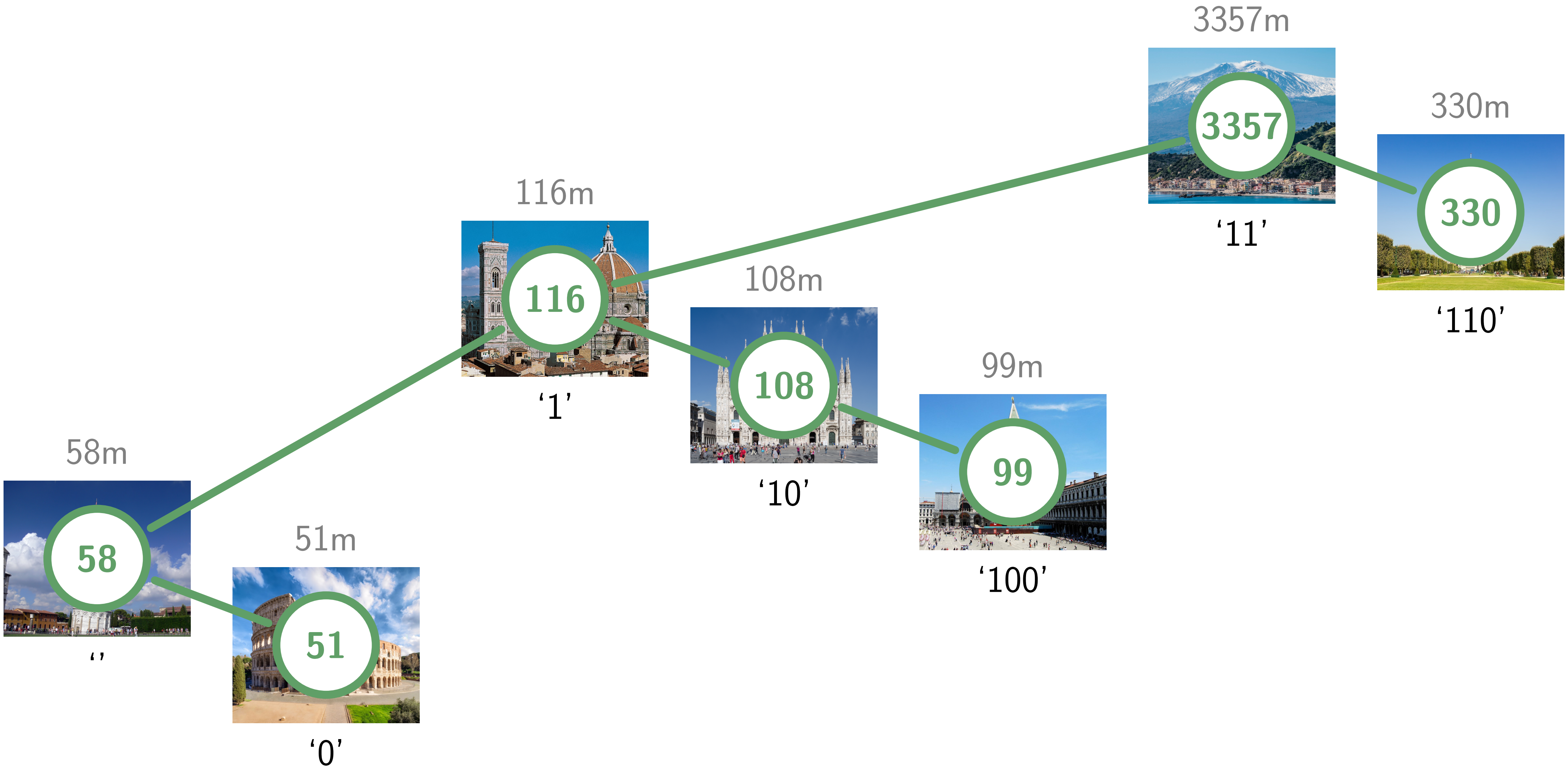
330m

‘110’

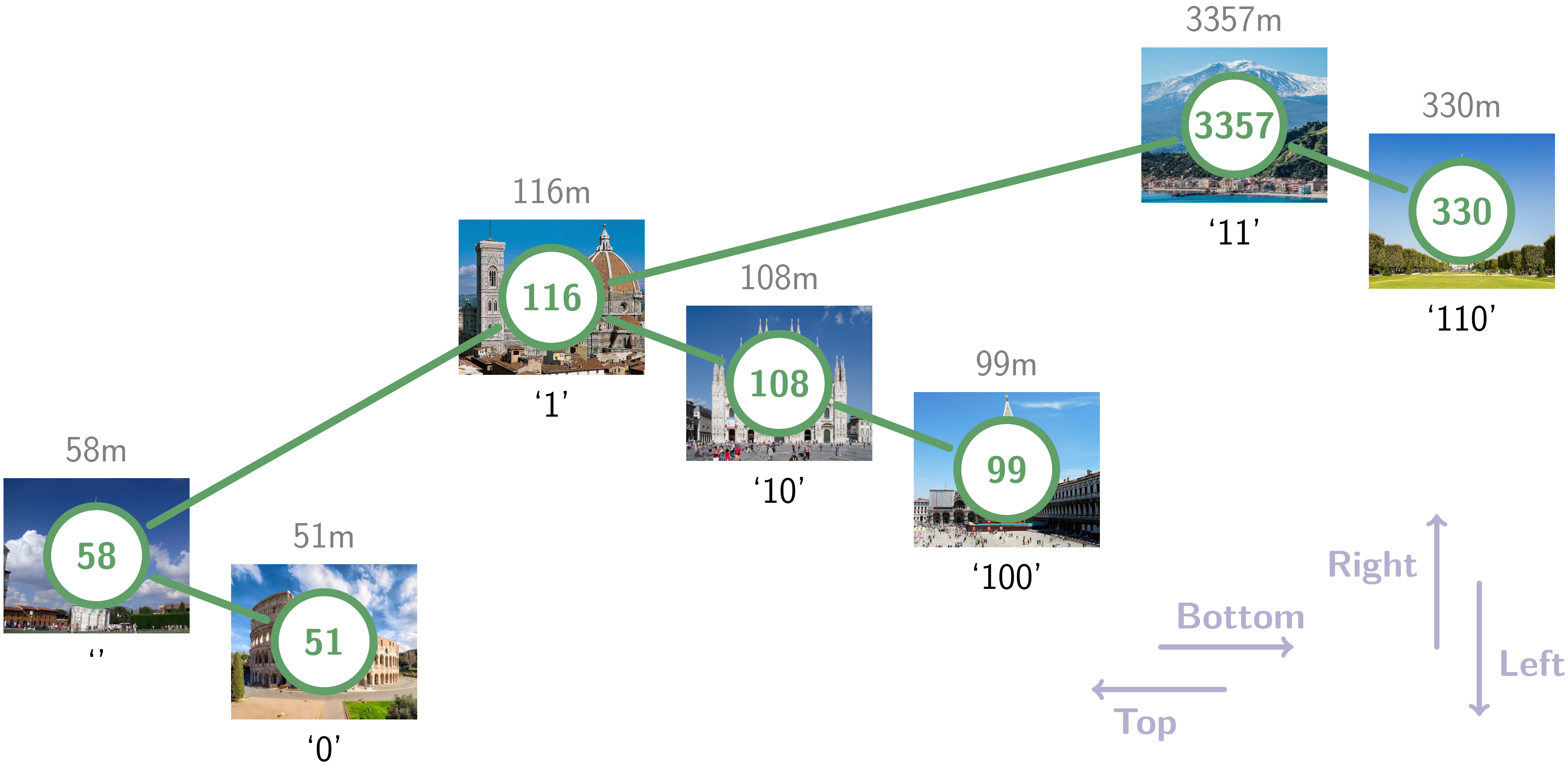
Motivation


3357m
3357
'11'
330m
330
'110'
116m
116
'1'
108m
108
'10'
99m
99
'100'
58m
58
''
51m
51
'0'

Motivation


3357m
3357
'11'
330m
330
'110'
116m
116
'1'
108m
108
'10'
99m
99
'100'
58m
58
''
51m
51
'0'

Motivation



3357m
3357
‘11’
330m
330
‘110’
116m
116
‘1’
108m
108
‘10’
99m
99
‘100’
58m
58
‘’
51m
51
‘0’
Right
Bottom
Left
Top

Binary Search Trees

Binary Search Trees

Let $\mathcal{P} = \{(x_i, y_i) : 1 \leq i \leq n\}$ be a set of points with distinct coordinates. We construct the binary search tree from $\mathcal{P}$ as follows.

Binary Search Trees

Let $\mathcal{P} = \{(x_i, y_i) : 1 \leq i \leq n\}$ be a set of points with distinct coordinates. We construct the binary search tree from $\mathcal{P}$ as follows.

- Order the points $(x_{(1)}, y_{(1)}), \ldots, (x_{(n)}, y_{(n)})$ using their x-coordinates: $x_{(1)} < \ldots < x_{(n)}$.

Binary Search Trees

Let $\mathcal{P} = \{(x_i, y_i) : 1 \leq i \leq n\}$ be a set of points with distinct coordinates. We construct the binary search tree from $\mathcal{P}$ as follows.

- Order the points $(x_{(1)}, y_{(1)}), \ldots, (x_{(n)}, y_{(n)})$ using their x-coordinates: $x_{(1)} < \ldots < x_{(n)}$.
- Insert $y_{(1)}, \ldots, y_{(n)}$ in that order into a binary search tree, that is:

Binary Search Trees

Let $\mathcal{P} = \{(x_i, y_i) : 1 \leq i \leq n\}$ be a set of points with distinct coordinates. We construct the binary search tree from $\mathcal{P}$ as follows.

- Order the points $(x_{(1)}, y_{(1)}), \dots, (x_{(n)}, y_{(n)})$ using their x-coordinates: $x_{(1)} < \dots < x_{(n)}$.
- Insert $y_{(1)}, \dots, y_{(n)}$ in that order into a binary search tree, that is:
    - $y_{(1)}$ is placed at the root.

Binary Search Trees

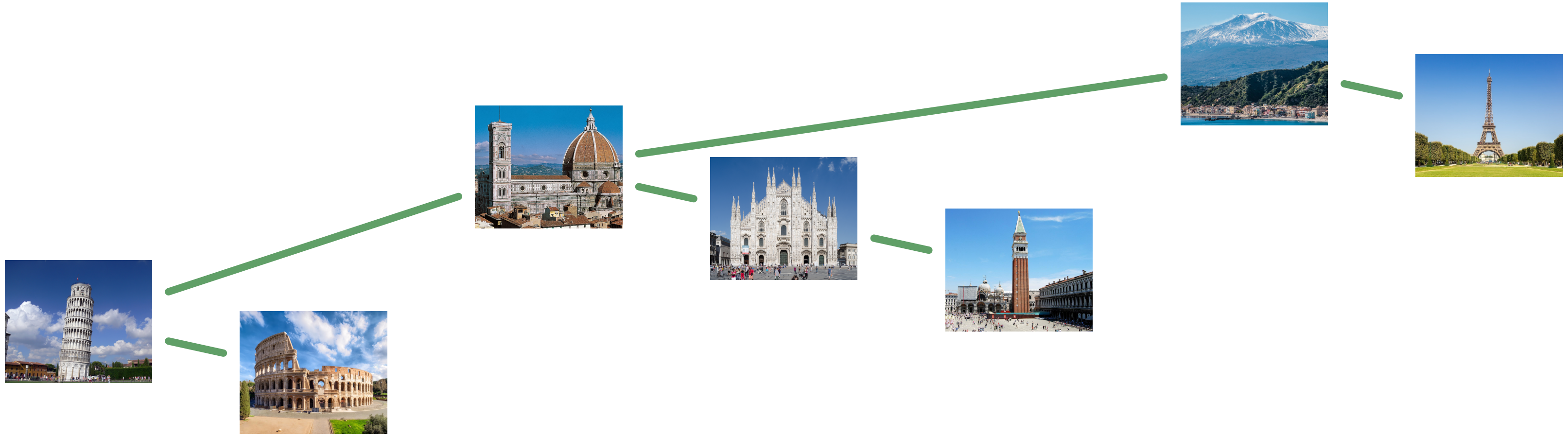
Let $\mathcal{P} = \{(x_i, y_i) : 1 \leq i \leq n\}$ be a set of points with distinct coordinates. We construct the binary search tree from $\mathcal{P}$ as follows.

- Order the points $(x_{(1)}, y_{(1)}), \dots, (x_{(n)}, y_{(n)})$ using their x-coordinates: $x_{(1)} < \dots < x_{(n)}$.
- Insert $y_{(1)}, \dots, y_{(n)}$ in that order into a binary search tree, that is:
  - $y_{(1)}$ is placed at the root.
  - Given that the tree of $y_{(1)}, \dots, y_{(i)}$ is constructed, $y_{(i+1)}$ is inserted down the tree until there is a free space, going right if it is larger than the current value and left otherwise.

Binary Search Trees

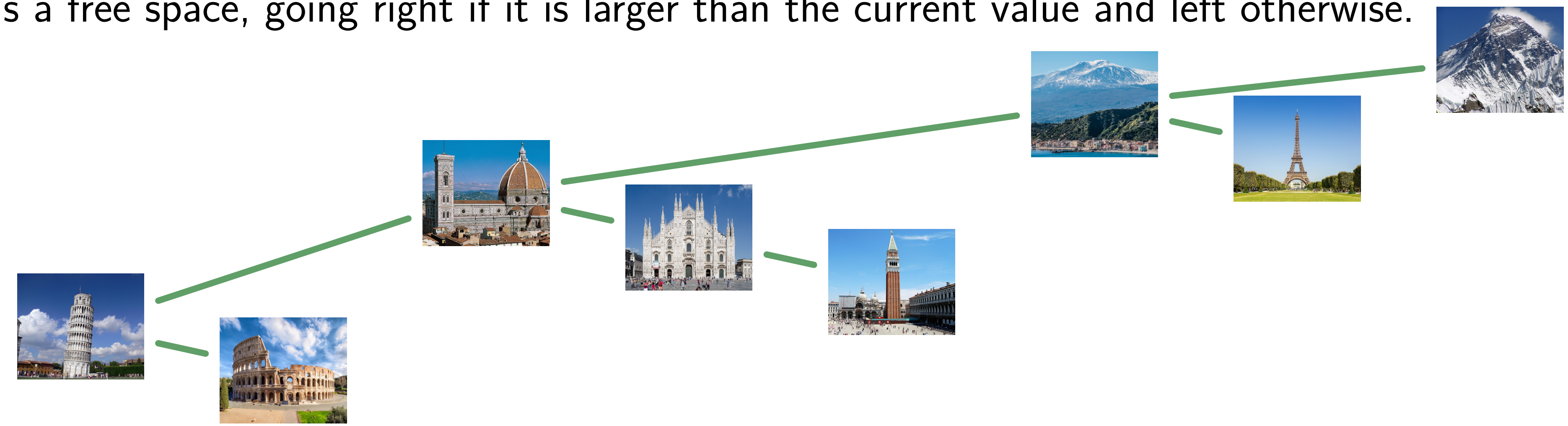
Let $\mathcal{P} = \{(x_i, y_i) : 1 \leq i \leq n\}$ be a set of points with distinct coordinates. We construct the binary search tree from $\mathcal{P}$ as follows.

- Order the points $(x_{(1)}, y_{(1)}), \ldots, (x_{(n)}, y_{(n)})$ using their x-coordinates: $x_{(1)} < \ldots < x_{(n)}$.
- Insert $y_{(1)}, \ldots, y_{(n)}$ in that order into a binary search tree, that is:
    - $y_{(1)}$ is placed at the root.
    - Given that the tree of $y_{(1)}, \ldots, y_{(i)}$ is constructed, $y_{(i+1)}$ is inserted down the tree until there is a free space, going right if it is larger than the current value and left otherwise.



Let $\mathcal{P} = \{(x_i, y_i) : 1 \leq i \leq n\}$ be a set of points with distinct coordinates. We construct the binary search tree from $\mathcal{P}$ as follows.

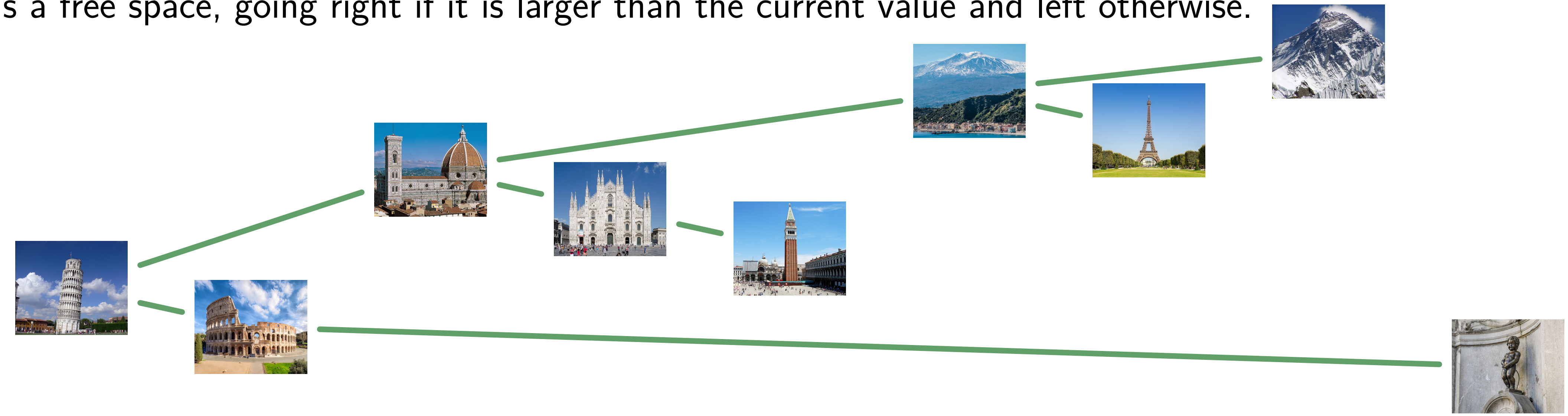
- Order the points $(x_{(1)}, y_{(1)}), \ldots, (x_{(n)}, y_{(n)})$ using their x-coordinates: $x_{(1)} < \ldots < x_{(n)}$.
- Insert $y_{(1)}, \ldots, y_{(n)}$ in that order into a binary search tree, that is:
  - $y_{(1)}$ is placed at the root.
  - Given that the tree of $y_{(1)}, \ldots, y_{(i)}$ is constructed, $y_{(i+1)}$ is inserted down the tree until there is a free space, going right if it is larger than the current value and left otherwise.



Binary Search Trees

Let $\mathcal{P} = \{(x_i, y_i) : 1 \leq i \leq n\}$ be a set of points with distinct coordinates. We construct the binary search tree from $\mathcal{P}$ as follows.

- Order the points $(x_{(1)}, y_{(1)}), \ldots, (x_{(n)}, y_{(n)})$ using their x-coordinates: $x_{(1)} < \ldots < x_{(n)}$.
- Insert $y_{(1)}, \ldots, y_{(n)}$ in that order into a binary search tree, that is:
  - $y_{(1)}$ is placed at the root.
  - Given that the tree of $y_{(1)}, \ldots, y_{(i)}$ is constructed, $y_{(i+1)}$ is inserted down the tree until there is a free space, going right if it is larger than the current value and left otherwise.



Binary Search Trees

Let $\mathcal{P} = \{(x_i, y_i) : 1 \leq i \leq n\}$ be a set of points with distinct coordinates. We construct the binary search tree from $\mathcal{P}$ as follows.

- Order the points $(x_{(1)}, y_{(1)}), \ldots, (x_{(n)}, y_{(n)})$ using their x-coordinates: $x_{(1)} < \ldots < x_{(n)}$.
- Insert $y_{(1)}, \ldots, y_{(n)}$ in that order into a binary search tree, that is:
    - $y_{(1)}$ is placed at the root.
    - Given that the tree of $y_{(1)}, \ldots, y_{(i)}$ is constructed, $y_{(i+1)}$ is inserted down the tree until there is a free space, going right if it is larger than the current value and left otherwise.

→ Seen as storing devices, we are often interested in the "worst case retrieval time", corresponding to the height of the binary search tree.

Random Binary Search Tree

Random Binary Search Tree

It is natural to wonder what happens in the following two scenarios.

Random Binary Search Tree

It is natural to wonder what happens in the following two scenarios.

- The values of the entries are iid ($x_i = i$ and $(y_i)_{1\leq i\leq n}$ iid).

Random Binary Search Tree

It is natural to wonder what happens in the following two scenarios.

- The values of the entries are iid ($x_i = i$ and $(y_i)_{1\leq i\leq n}$ iid).
- The x (or y)-coordinates are randomized ($\mathcal{P}' = \{(x_{\sigma(i)}, y_i)\}$ for a uniform permutation σ).

Random Binary Search Tree

It is natural to wonder what happens in the following two scenarios.

- The values of the entries are iid ($x_i = i$ and $(y_i)_{1\leq i\leq n}$ iid).
- The x (or y)-coordinates are randomized ($\mathcal{P}' = \{(x_{\sigma(i)}, y_i)\}$ for a uniform permutation σ).

$\rightarrow$ Since the structure of the tree only depends on the relative positions of the points, both trees are identically distributed and referred to as *random binary search trees* (RBST).

It is natural to wonder what happens in the following two scenarios.

- The values of the entries are iid ($x_i = i$ and $(y_i)_{1\leq i\leq n}$ iid).
- The x (or y)-coordinates are randomized ($\mathcal{P}' = \{(x_{\sigma(i)}, y_i)\}$ for a uniform permutation σ).

→ Since the structure of the tree only depends on the relative positions of the points, both trees are identically distributed and referred to as *random binary search trees* (RBST).

Devroye (1986)

Let H_n be the height of a RBST of size n. Then there exists a constant $c^* \simeq 4.311\ldots$ such that

$$\frac{H_n}{c^* \log n} \longrightarrow 1\,.$$

Table of Content

Binary search trees

Permutons

Our results

The conditions

Permuton Samples

Permuton Samples

From the previous case, we understand the behaviour of the height of the binary search tree when $\{(x_i, y_i)\}$ and $\{x_{\sigma(i)}, y_i\}$ are identically distributed for all σ.

Permuton Samples

From the previous case, we understand the behaviour of the height of the binary search tree when $\{(x_i, y_i)\}$ and $\{x_{\sigma(i)}, y_i\}$ are identically distributed for all σ.

$\rightarrow$ Can we study more general point processes?

Permuton Samples

From the previous case, we understand the behaviour of the height of the binary search tree when $\{(x_i, y_i)\}$ and $\{x_{\sigma(i)}, y_i\}$ are identically distributed for all σ.

$\rightarrow$ Can we study more general point processes?

Pick your favourite "non-atomic" distribution on $\mathbb{R}^2$ and sample n points accordingly.

Permuton Samples

From the previous case, we understand the behaviour of the height of the binary search tree when $\{(x_i, y_i)\}$ and $\{x_{\sigma(i)}, y_i\}$ are identically distributed for all σ.

$\rightarrow$ Can we study more general point processes?

Pick your favourite "non-atomic" distribution on $\mathbb{R}^2$ and sample n points accordingly.

$\rightarrow$ What happens for the height of the tree as $n \to \infty$?

Permuton Samples

From the previous case, we understand the behaviour of the height of the binary search tree when $\{(x_i, y_i)\}$ and $\{x_{\sigma(i)}, y_i\}$ are identically distributed for all σ.

$\rightarrow$ Can we study more general point processes?

Pick your favourite "non-atomic" distribution on $\mathbb{R}^2$ and sample n points accordingly.

$\rightarrow$ What happens for the height of the tree as $n \to \infty$?

$\rightarrow$ Under what hypothesis do we still get the height of a RBST (i.e. $c^* \log n$)?

Permuton Samples

From the previous case, we understand the behaviour of the height of the binary search tree when $\{(x_i, y_i)\}$ and $\{x_{\sigma(i)}, y_i\}$ are identically distributed for all σ.

→ Can we study more general point processes?

Pick your favourite "non-atomic" distribution on $\mathbb{R}^2$ and sample n points accordingly.

→ What happens for the height of the tree as $n \to \infty$?

→ Under what hypothesis do we still get the height of a RBST (i.e. $c^* \log n$)?

Since the structure of the tree is only define by the relative positions of the points, we can add two extra assumptions on this distribution:

Permuton Samples

From the previous case, we understand the behaviour of the height of the binary search tree when $\{(x_i, y_i)\}$ and $\{x_{\sigma(i)}, y_i\}$ are identically distributed for all σ.

→ Can we study more general point processes?

Pick your favourite "non-atomic" distribution on $\mathbb{R}^2$ and sample n points accordingly.

→ What happens for the height of the tree as $n \to \infty$?

→ Under what hypothesis do we still get the height of a RBST (i.e. $c^* \log n$)?

Since the structure of the tree is only define by the relative positions of the points, we can add two extra assumptions on this distribution:

- Its support is $[0, 1]^2$.

Permuton Samples

From the previous case, we understand the behaviour of the height of the binary search tree when $\{(x_i, y_i)\}$ and $\{x_{\sigma(i)}, y_i\}$ are identically distributed for all σ.

$\rightarrow$ Can we study more general point processes?

Pick your favourite "non-atomic" distribution on $\mathbb{R}^2$ and sample n points accordingly.

$\rightarrow$ What happens for the height of the tree as $n \to \infty$?

$\rightarrow$ Under what hypothesis do we still get the height of a RBST (i.e. $c^* \log n$)?

Since the structure of the tree is only define by the relative positions of the points, we can add two extra assumptions on this distribution:

- Its support is $[0, 1]^2$.
- It has uniform marginals.

Permuton Samples

From the previous case, we understand the behaviour of the height of the binary search tree when $\{(x_i, y_i)\}$ and $\{x_{\sigma(i)}, y_i\}$ are identically distributed for all σ.

→ Can we study more general point processes?

Pick your favourite "non-atomic" distribution on $\mathbb{R}^2$ and sample n points accordingly.

→ What happens for the height of the tree as $n \to \infty$?

→ Under what hypothesis do we still get the height of a RBST (i.e. $c^* \log n$)?

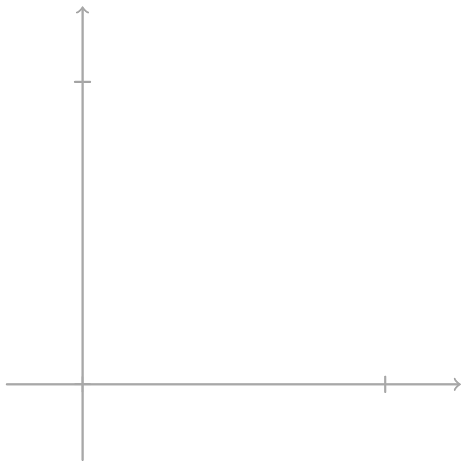
Since the structure of the tree is only define by the relative positions of the points, we can add two extra assumptions on this distribution:

- Its support is $[0, 1]^2$.
- It has uniform marginals.

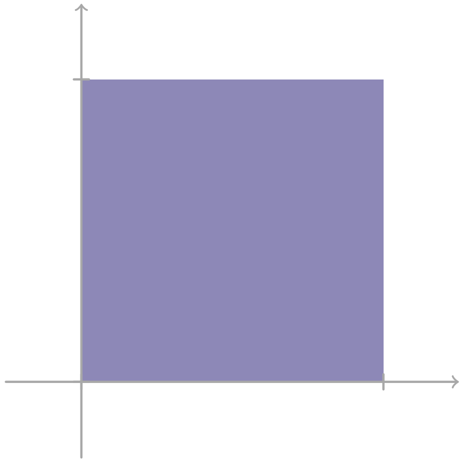
→ We call such a distribution a *permuton*.

Examples

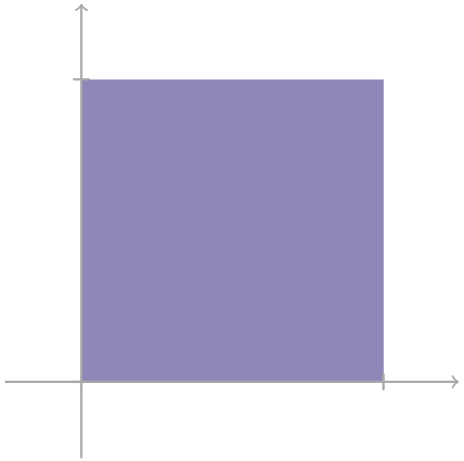
Examples

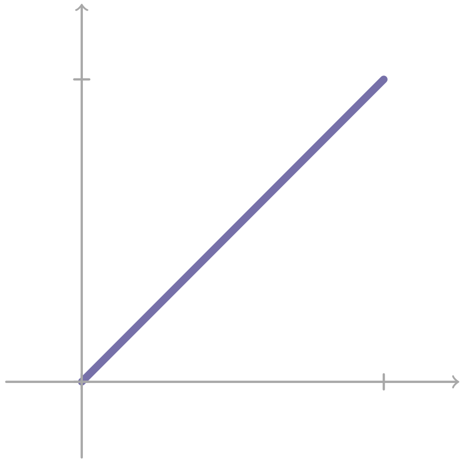


Examples

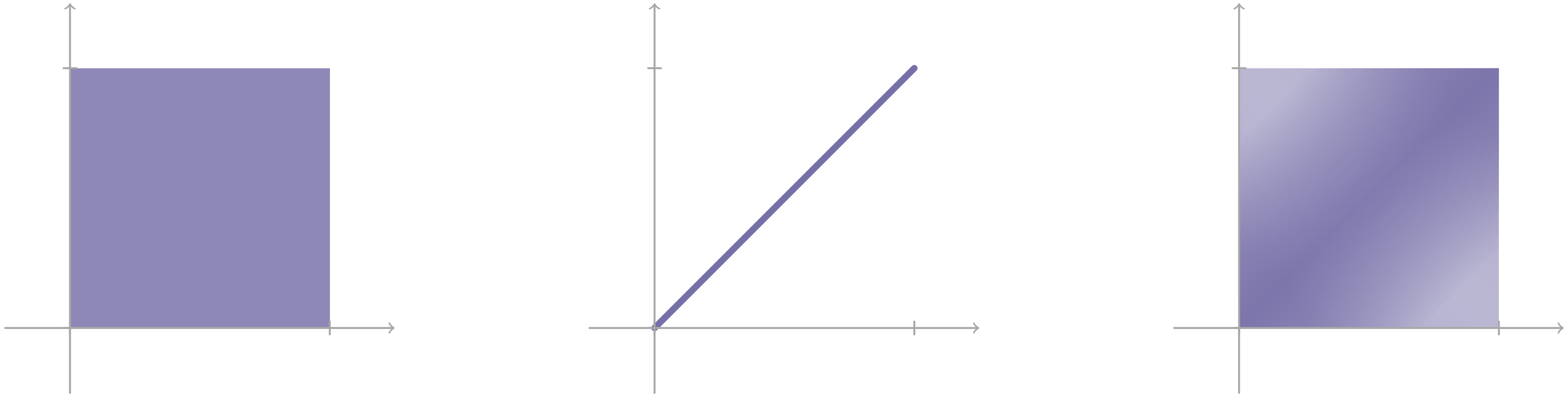


Examples

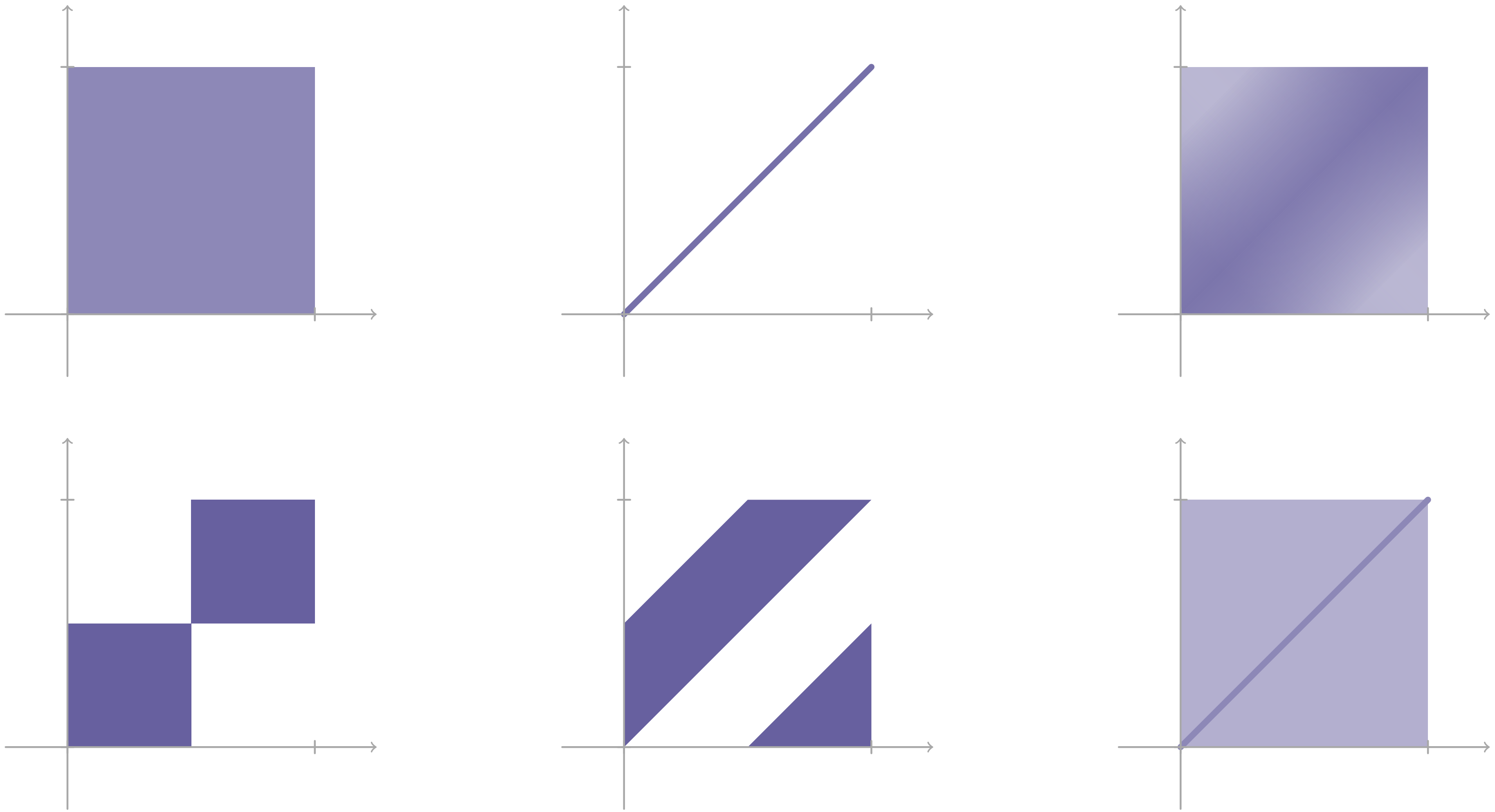




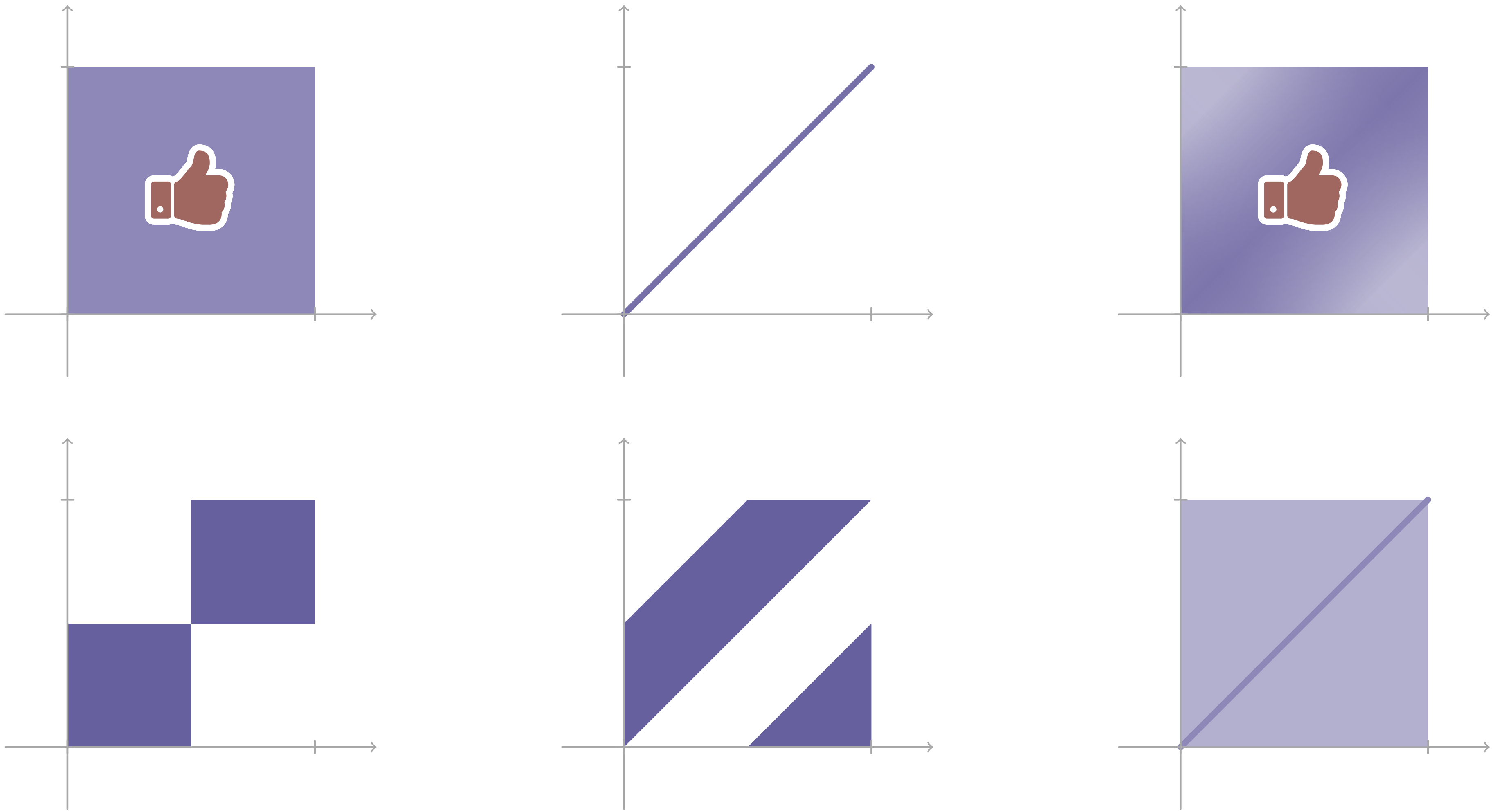
Examples



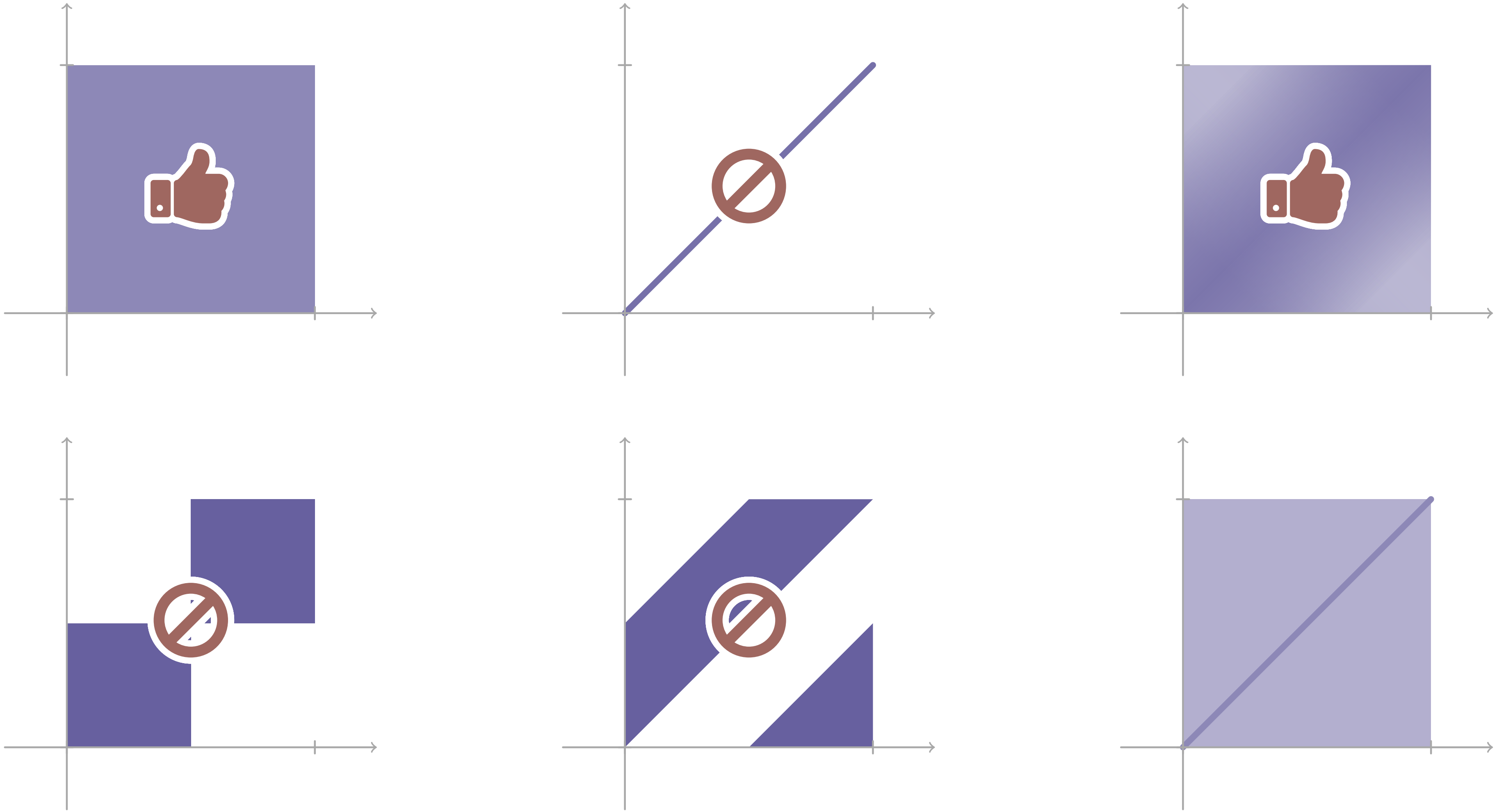
Examples



Examples



Examples



Examples

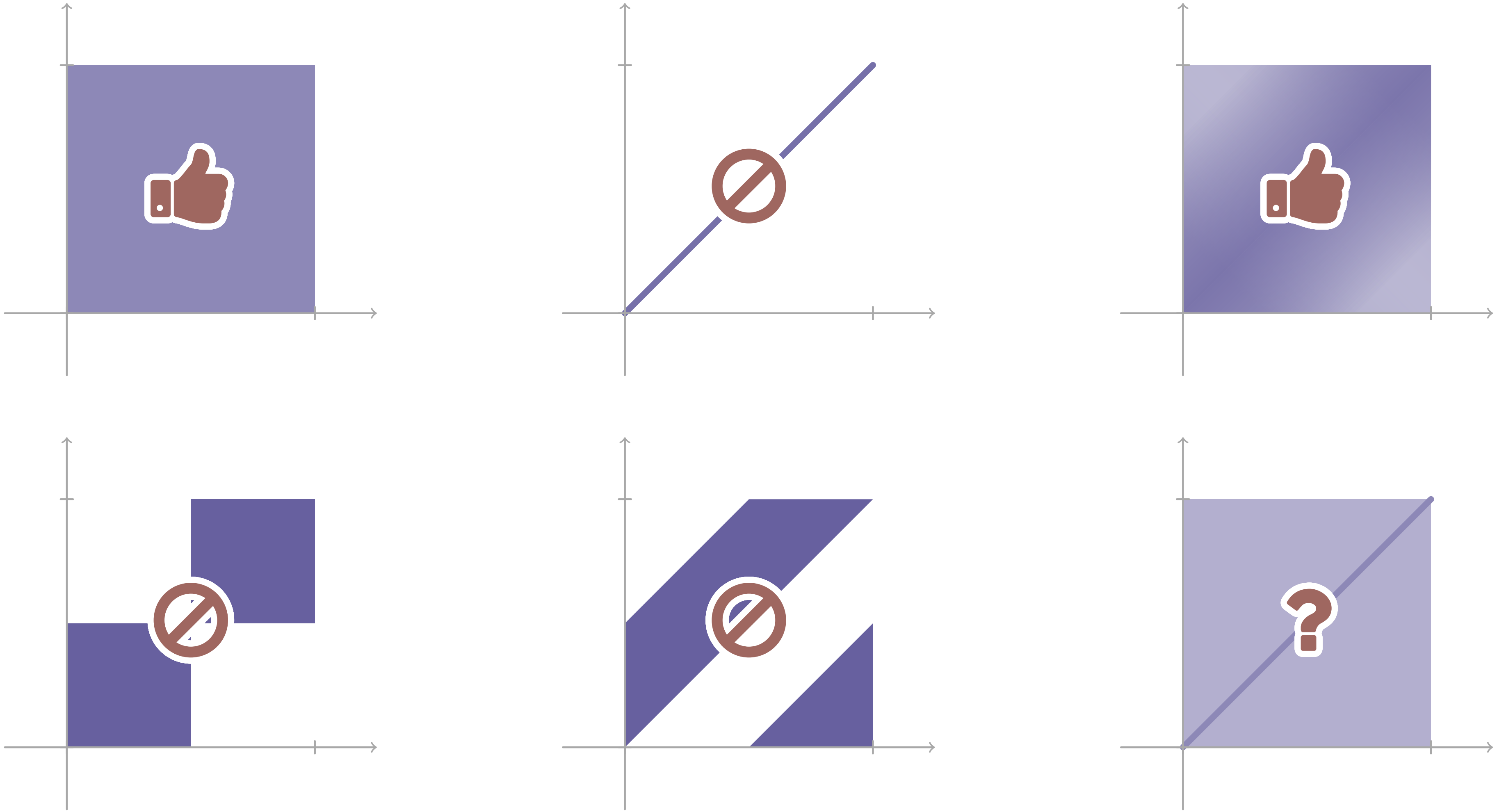


Table of Content

Binary search trees

Permutons

Our results

The conditions

A simple case

A simple case

Thanks to Devroye's result, we know that:

A simple case

Thanks to Devroye's result, we know that:

- the uniform permuton on $[0,1]^2$ leads to a tree of height $c^* \log n$; and

A simple case

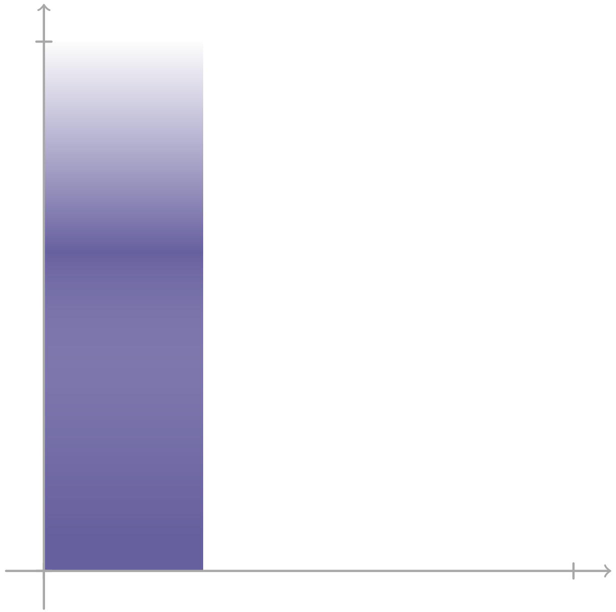
Thanks to Devroye's result, we know that:

- the uniform permuton on $[0,1]^2$ leads to a tree of height $c^* \log n$; and
- this result also extends to any distribution on a rectangle of $\mathbb{R}^2$ independent of x or y.

A simple case

Thanks to Devroye's result, we know that:

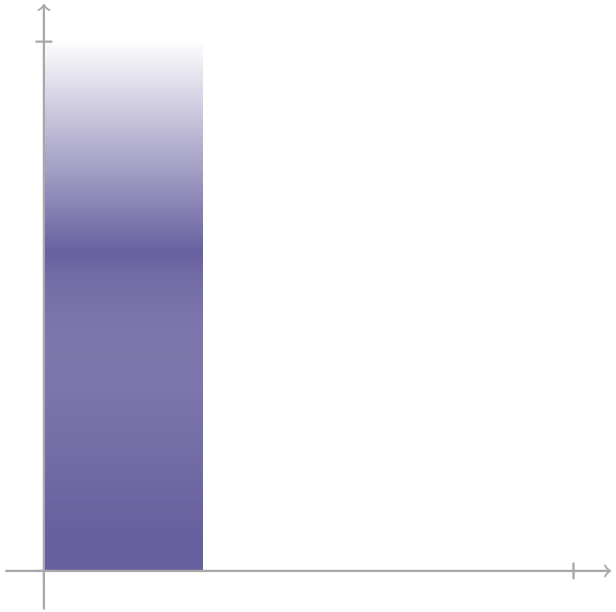
- the uniform permuton on $[0,1]^2$ leads to a tree of height $c^* \log n$; and
- this result also extends to any distribution on a rectangle of $\mathbb{R}^2$ independent of x or y.

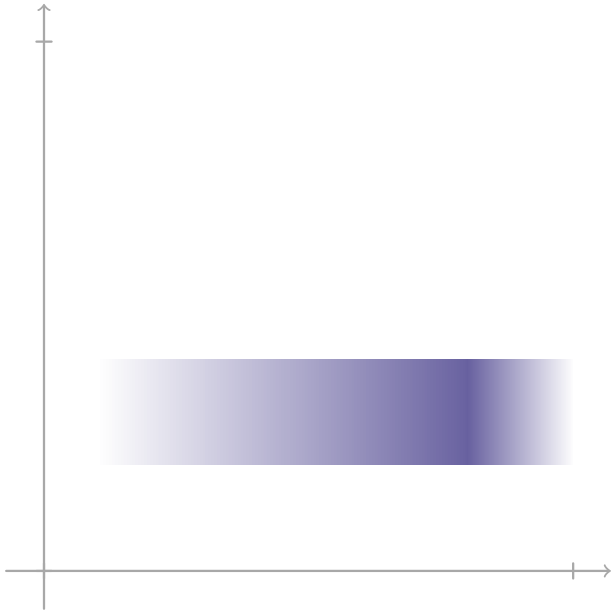


A simple case

Thanks to Devroye's result, we know that:

- the uniform permuton on $[0,1]^2$ leads to a tree of height $c^* \log n$; and
- this result also extends to any distribution on a rectangle of $\mathbb{R}^2$ independent of x or y.



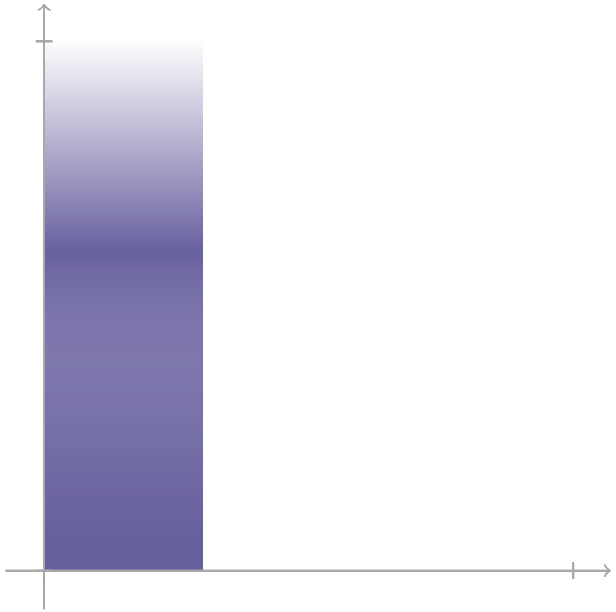


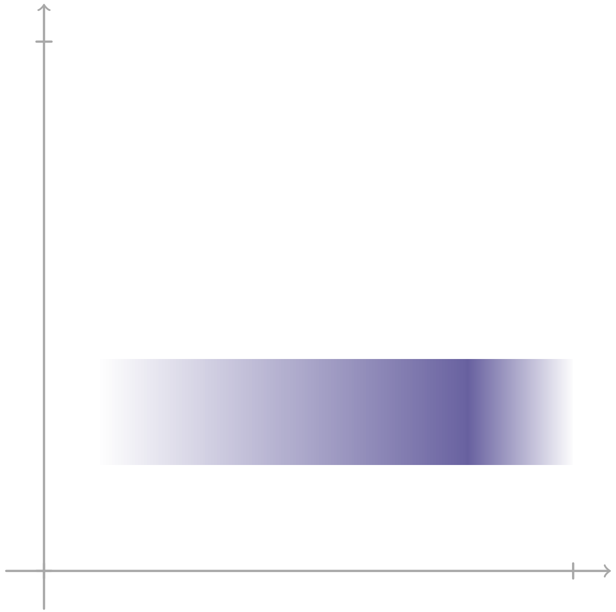
A simple case

Thanks to Devroye's result, we know that:

- the uniform permuton on $[0,1]^2$ leads to a tree of height $c^* \log n$; and
- this result also extends to any distribution on a rectangle of $\mathbb{R}^2$ independent of x or y.

$\rightarrow$ What happens in general when the distribution admits a density?





A simple case

Thanks to Devroye's result, we know that:

- the uniform permuton on $[0,1]^2$ leads to a tree of height $c^* \log n$; and
- this result also extends to any distribution on a rectangle of $\mathbb{R}^2$ independent of x or y.

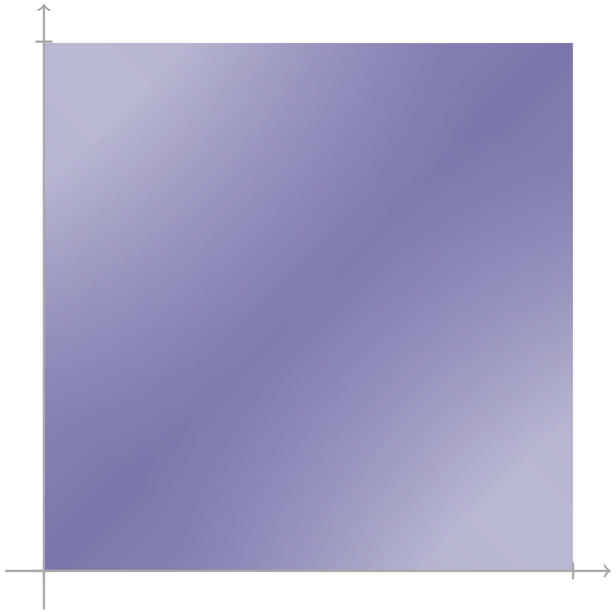
→ What happens in general when the distribution admits a density?

A simple case

Thanks to Devroye's result, we know that:

- the uniform permuton on $[0,1]^2$ leads to a tree of height $c^* \log n$; and
- this result also extends to any distribution on a rectangle of $\mathbb{R}^2$ independent of x or y.

$\rightarrow$ What happens in general when the distribution admits a density?

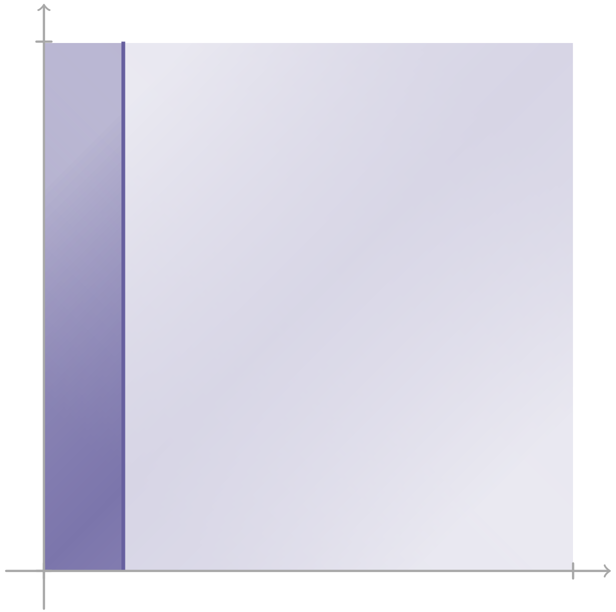


A simple case

Thanks to Devroye's result, we know that:

- the uniform permuton on $[0,1]^2$ leads to a tree of height $c^* \log n$; and
- this result also extends to any distribution on a rectangle of $\mathbb{R}^2$ independent of x or y.

$\rightarrow$ What happens in general when the distribution admits a density?

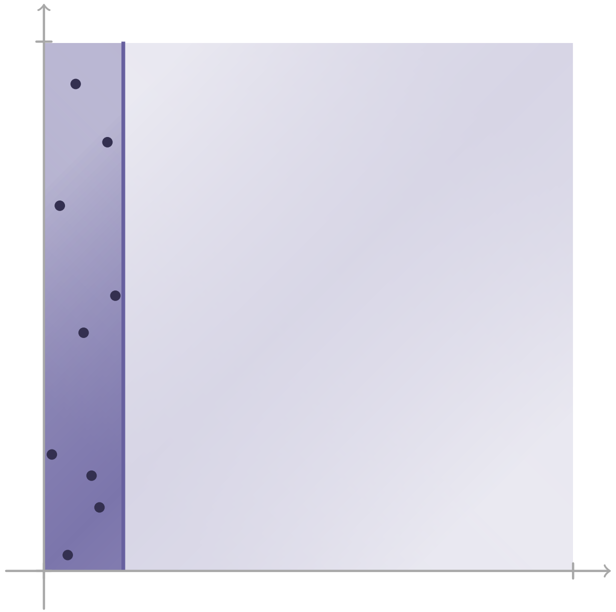


A simple case

Thanks to Devroye's result, we know that:

- the uniform permuton on $[0,1]^2$ leads to a tree of height $c^* \log n$; and
- this result also extends to any distribution on a rectangle of $\mathbb{R}^2$ independent of x or y.

→ What happens in general when the distribution admits a density?

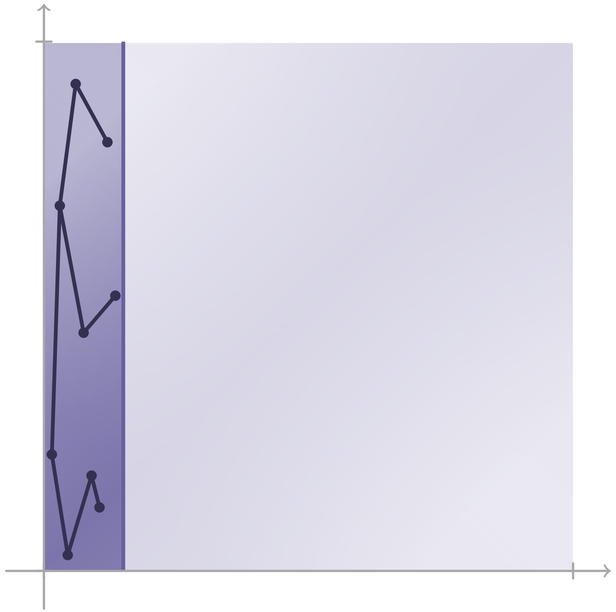


A simple case

Thanks to Devroye's result, we know that:

- the uniform permuton on $[0,1]^2$ leads to a tree of height $c^* \log n$; and
- this result also extends to any distribution on a rectangle of $\mathbb{R}^2$ independent of x or y.

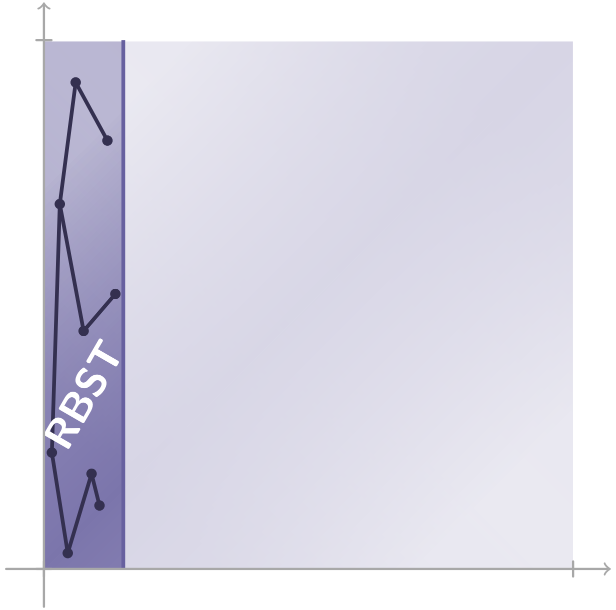
→ What happens in general when the distribution admits a density?



Thanks to Devroye's result, we know that:

- the uniform permuton on $[0,1]^2$ leads to a tree of height $c^* \log n$; and
- this result also extends to any distribution on a rectangle of $\mathbb{R}^2$ independent of x or y.

$\rightarrow$ What happens in general when the distribution admits a density?

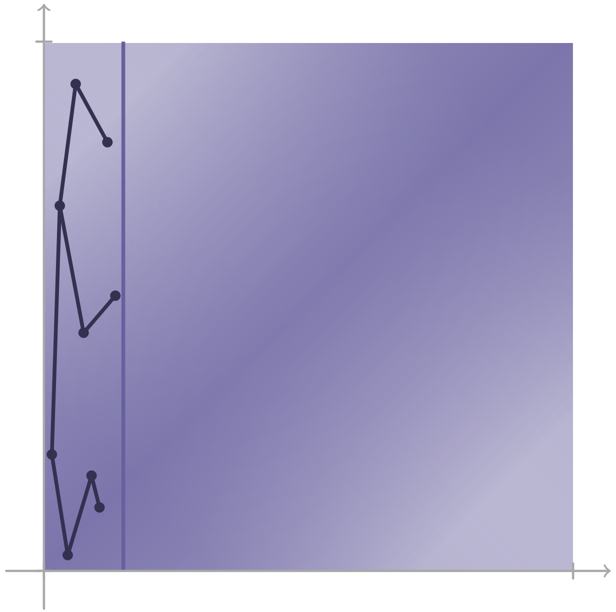


A simple case

Thanks to Devroye's result, we know that:

- the uniform permuton on $[0,1]^2$ leads to a tree of height $c^* \log n$; and
- this result also extends to any distribution on a rectangle of $\mathbb{R}^2$ independent of x or y.

→ What happens in general when the distribution admits a density?

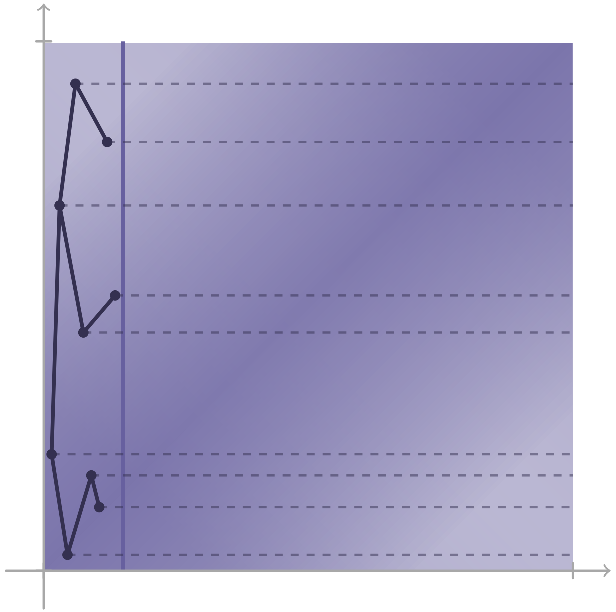


A simple case

Thanks to Devroye's result, we know that:

- the uniform permuton on $[0,1]^2$ leads to a tree of height $c^* \log n$; and
- this result also extends to any distribution on a rectangle of $\mathbb{R}^2$ independent of x or y.

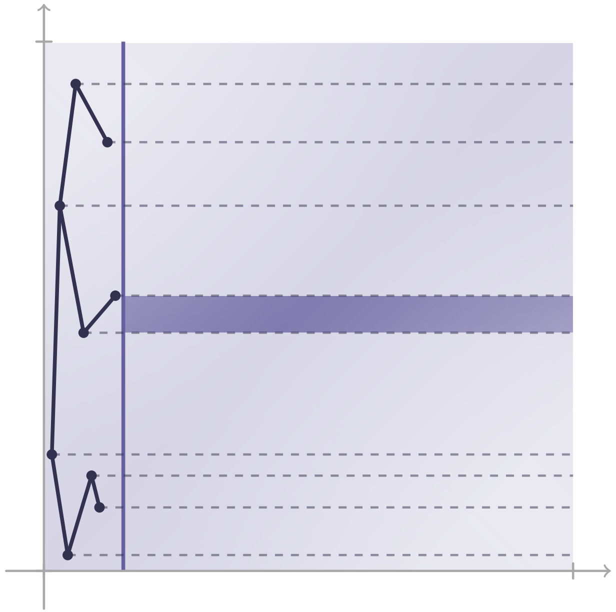
$\rightarrow$ What happens in general when the distribution admits a density?



Thanks to Devroye's result, we know that:

- the uniform permuton on $[0,1]^2$ leads to a tree of height $c^* \log n$; and
- this result also extends to any distribution on a rectangle of $\mathbb{R}^2$ independent of x or y.

$\rightarrow$ What happens in general when the distribution admits a density?

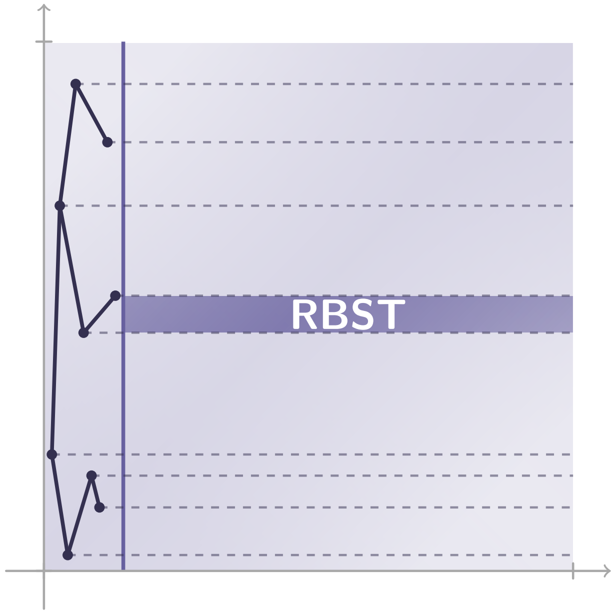


A simple case

Thanks to Devroye's result, we know that:

- the uniform permuton on $[0,1]^2$ leads to a tree of height $c^* \log n$; and
- this result also extends to any distribution on a rectangle of $\mathbb{R}^2$ independent of x or y.

→ What happens in general when the distribution admits a density?

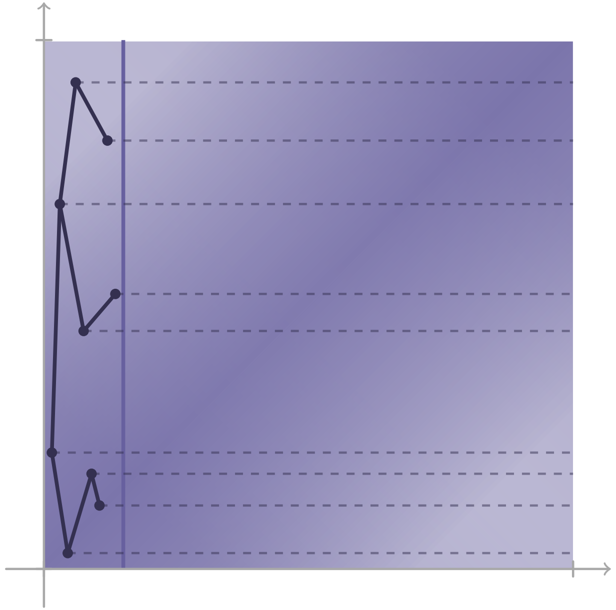


A simple case

Thanks to Devroye's result, we know that:

- the uniform permuton on $[0,1]^2$ leads to a tree of height $c^* \log n$; and
- this result also extends to any distribution on a rectangle of $\mathbb{R}^2$ independent of x or y.

$\rightarrow$ What happens in general when the distribution admits a density?

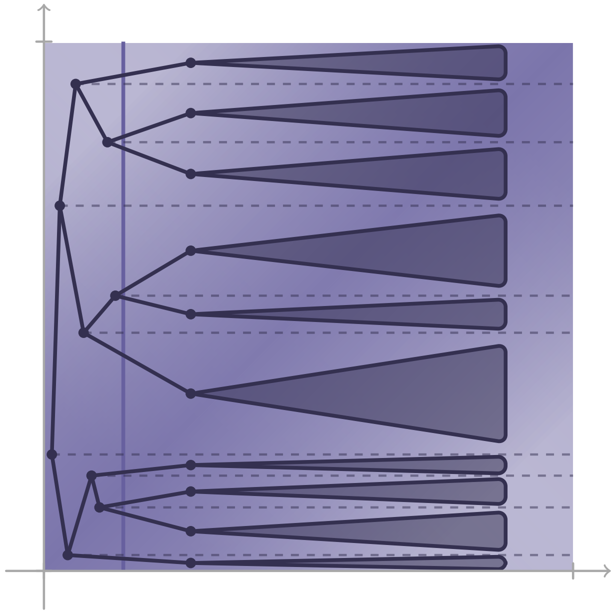


A simple case

Thanks to Devroye's result, we know that:

- the uniform permuton on $[0,1]^2$ leads to a tree of height $c^* \log n$; and
- this result also extends to any distribution on a rectangle of $\mathbb{R}^2$ independent of x or y.

→ What happens in general when the distribution admits a density?

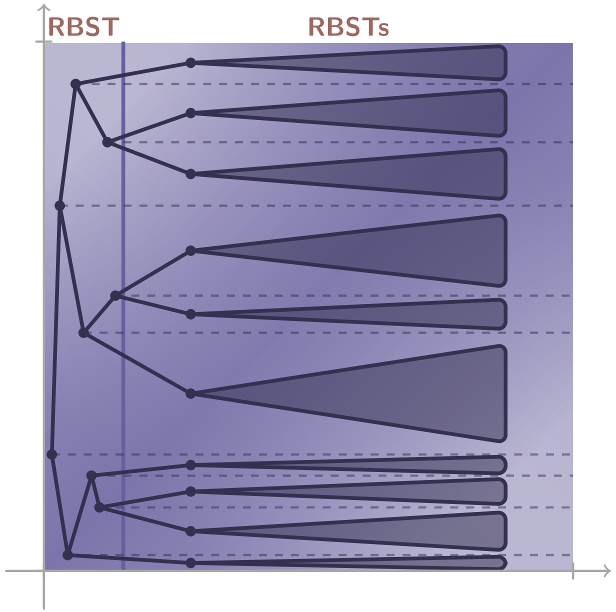


A simple case

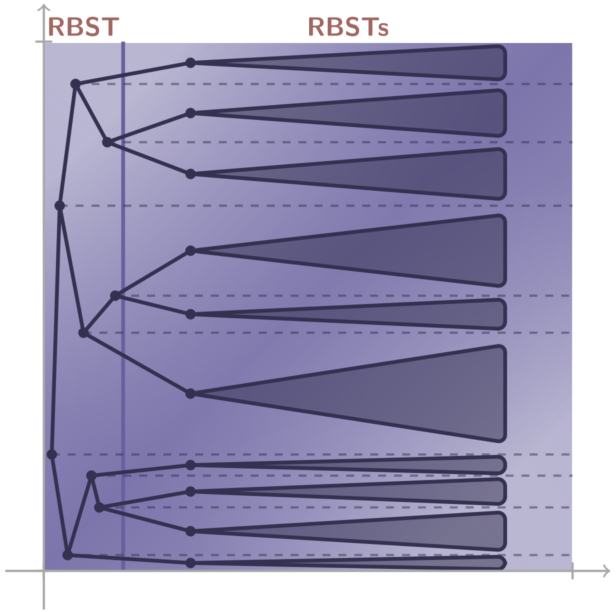
Thanks to Devroye's result, we know that:

- the uniform permuton on $[0,1]^2$ leads to a tree of height $c^* \log n$; and
- this result also extends to any distribution on a rectangle of $\mathbb{R}^2$ independent of x or y.

→ What happens in general when the distribution admits a density?

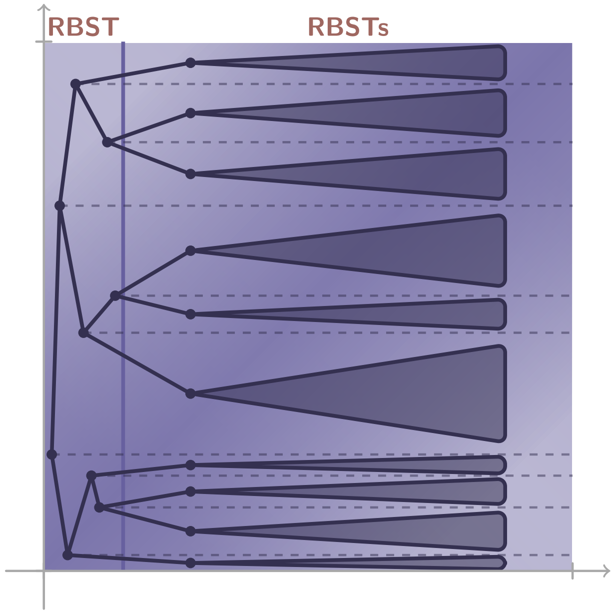


A simple case



A simple case

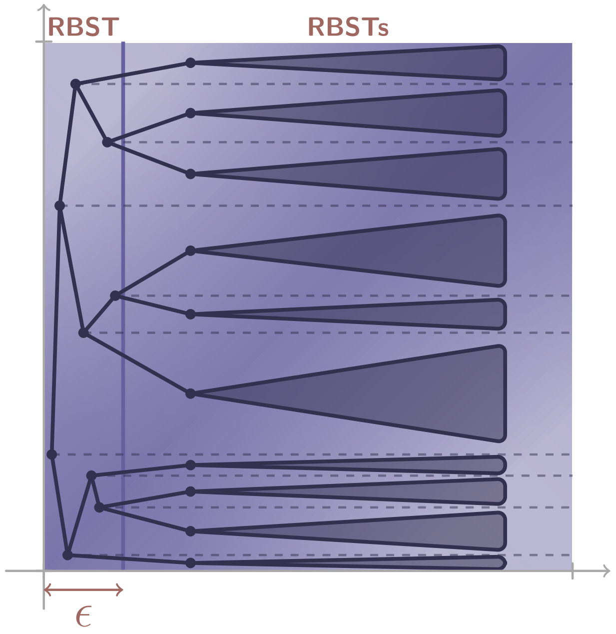
Consider the following approximate argument.



A simple case

Consider the following approximate argument.

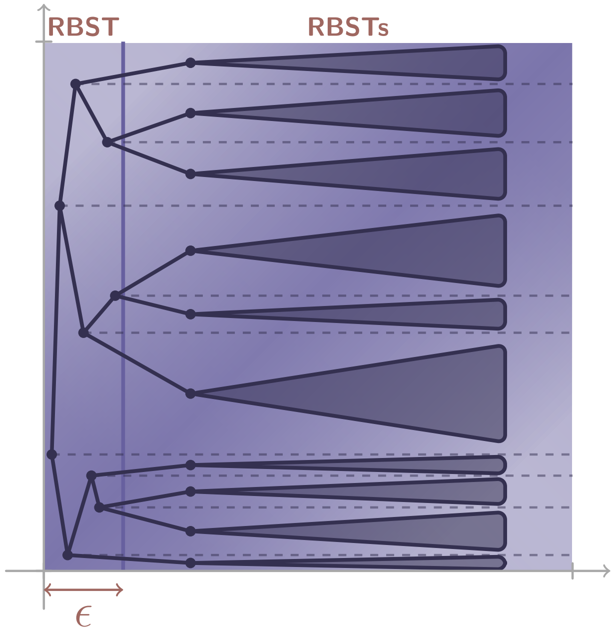
- If the left band has width ϵ, there are ϵn points in it.



A simple case

Consider the following approximate argument.

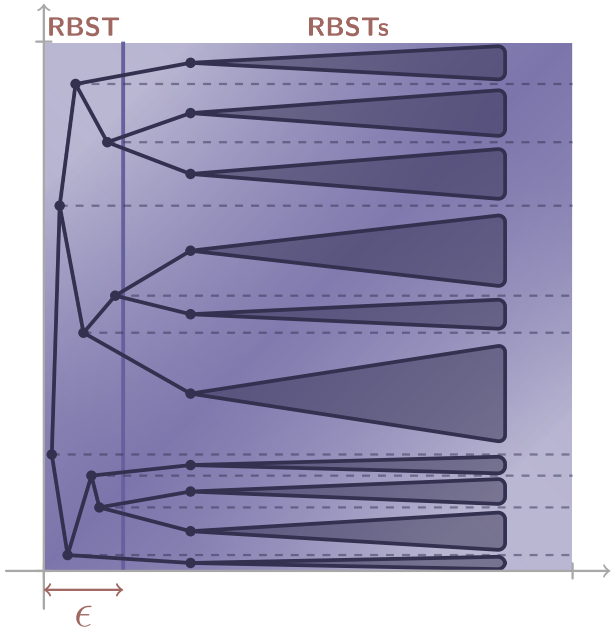
- If the left band has width ϵ, there are ϵn points in it.
- The tree of these first ϵn points is a RBST.



A simple case

Consider the following approximate argument.

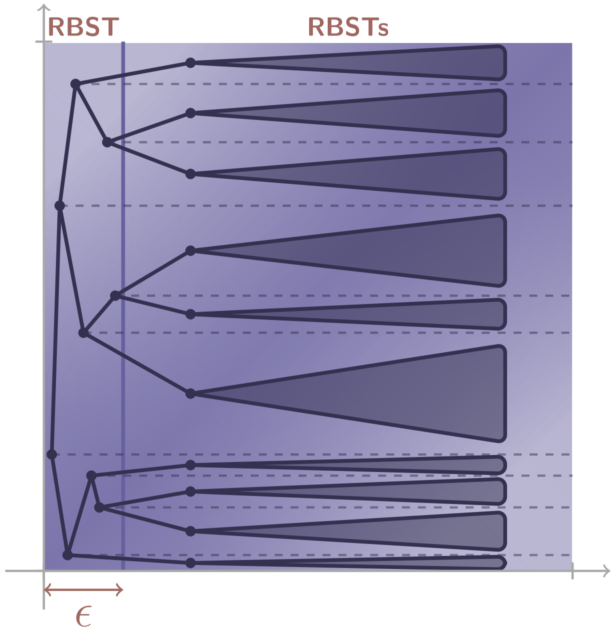
- If the left band has width ϵ, there are ϵn points in it.
- The tree of these first ϵn points is a RBST.
  - The height of this top (or left) tree is $c^* \log(\epsilon n)$.



A simple case

Consider the following approximate argument.

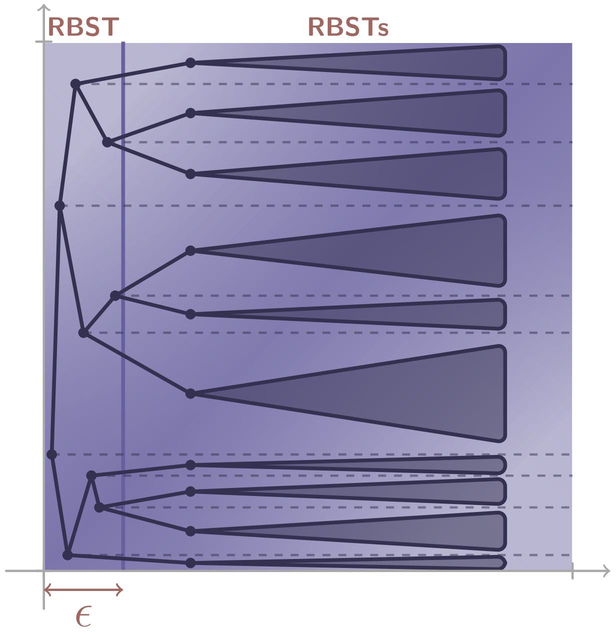
- If the left band has width ϵ, there are ϵn points in it.
- The tree of these first ϵn points is a RBST.
  - The height of this top (or left) tree is $c^* \log(\epsilon n)$.
- The vertical distance between two points in the left band is $1/\epsilon n$.



A simple case

Consider the following approximate argument.

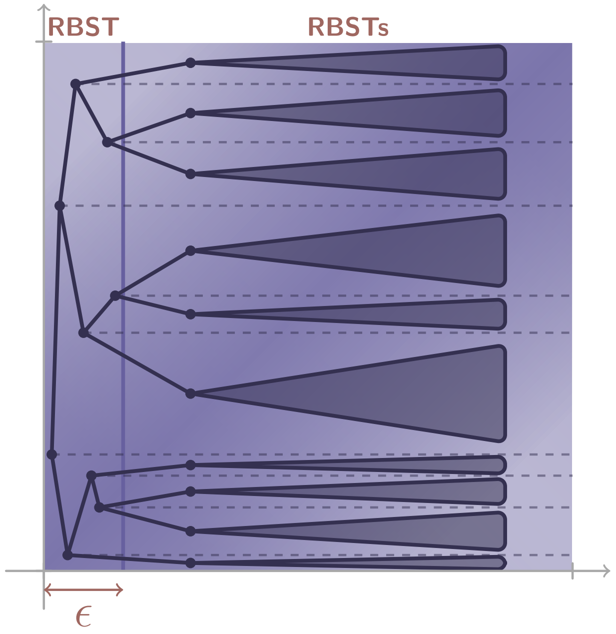
- If the left band has width ϵ, there are ϵn points in it.
- The tree of these first ϵn points is a RBST.
  - The height of this top (or left) tree is $c^* \log(\epsilon n)$.
- The vertical distance between two points in the left band is $1/\epsilon n$.
- The number of points in each band on the right is $(1-\epsilon)/\epsilon$.



A simple case

Consider the following approximate argument.

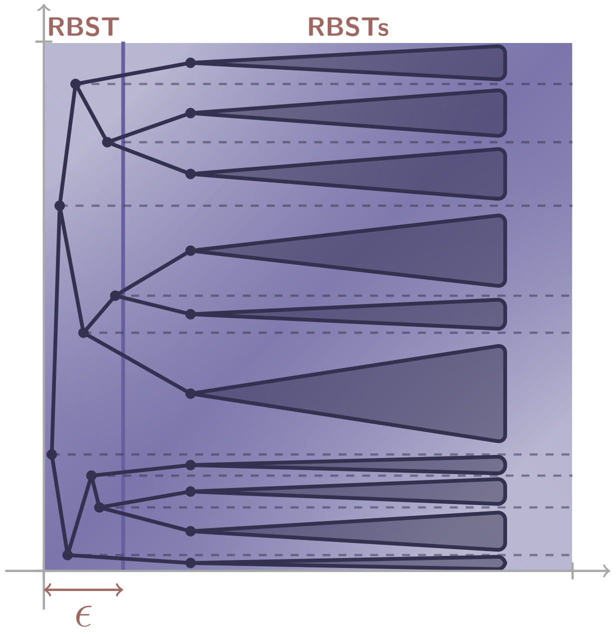
- If the left band has width ϵ, there are ϵn points in it.
- The tree of these first ϵn points is a RBST.
  - The height of this top (or left) tree is $c^* \log(\epsilon n)$.
- The vertical distance between two points in the left band is $1/\epsilon n$.
- The number of points in each band on the right is $(1-\epsilon)/\epsilon$.
- The subtrees on the right are RBSTs.



A simple case

Consider the following approximate argument.

- If the left band has width ϵ, there are ϵn points in it.
- The tree of these first ϵn points is a RBST.
  - The height of this top (or left) tree is $c^* \log(\epsilon n)$.
- The vertical distance between two points in the left band is $1/\epsilon n$.
- The number of points in each band on the right is $(1-\epsilon)/\epsilon$.
- The subtrees on the right are RBSTs.
  - The height of the right subtrees is $c^* \log((1-\epsilon)/\epsilon)$.

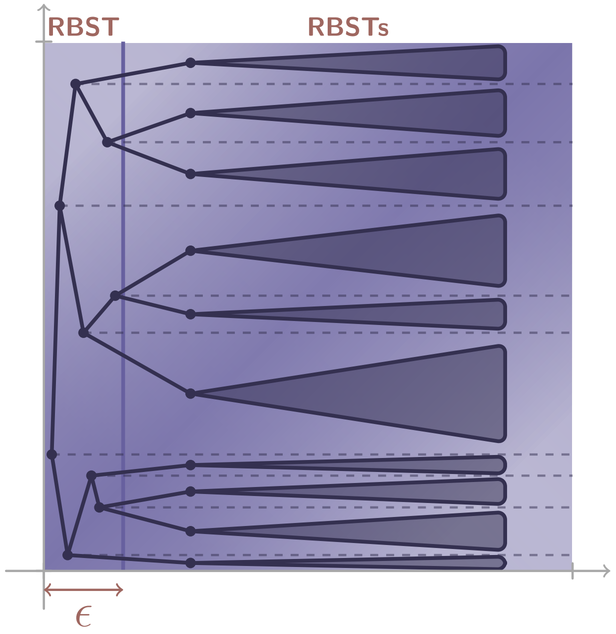


A simple case

Consider the following approximate argument.

- If the left band has width ϵ, there are ϵn points in it.
- The tree of these first ϵn points is a RBST.
  - The height of this top (or left) tree is $c^* \log(\epsilon n)$.
- The vertical distance between two points in the left band is $1/\epsilon n$.
- The number of points in each band on the right is $(1-\epsilon)/\epsilon$.
- The subtrees on the right are RBSTs.
  - The height of the right subtrees is $c^* \log((1-\epsilon)/\epsilon)$.

→ The height of the whole tree is $c^* \log n$.



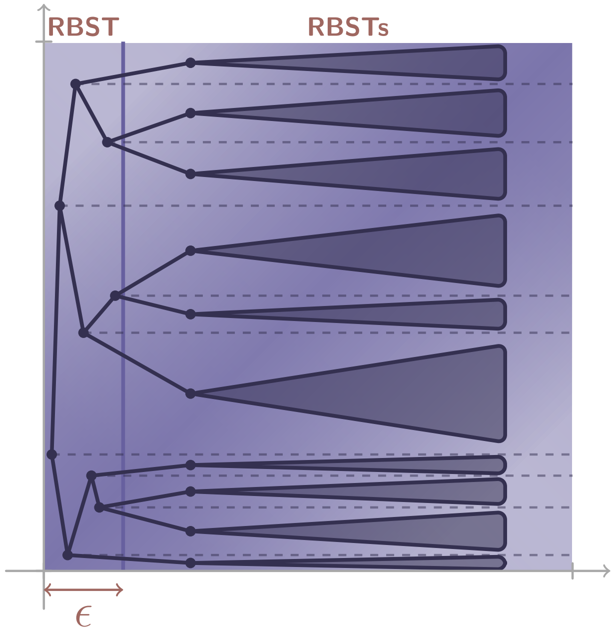
A simple case

Consider the following approximate argument.

- If the left band has width ϵ, there are ϵn points in it.
- The tree of these first ϵn points is a RBST.
  - The height of this top (or left) tree is $c^* \log(\epsilon n)$.
- The vertical distance between two points in the left band is $1/\epsilon n$.
- The number of points in each band on the right is $(1-\epsilon)/\epsilon$.
- The subtrees on the right are RBSTs.
  - The height of the right subtrees is $c^* \log((1-\epsilon)/\epsilon)$.

→ The height of the whole tree is $c^* \log n$.

⚠ While the average width of a band is $1/\epsilon n$, the maximal width is actually of order $\log n/n$.



Our result

Our result

, Dubach, Féray (2024)

, Dubach, Féray (2024)

Let ρ be a permuton with a density.

Our result

, Dubach, Féray (2024)

Let ρ be a permuton with a density. Further assume that this density is bounded on $[0,1]^2$ and positive continuous on a neighbourhood of $\{0\} \times [0,1]$ (the left boundary).

Our result

, Dubach, Féray (2024)

Let ρ be a permuton with a density. Further assume that this density is bounded on $[0,1]^2$ and positive continuous on a neighbourhood of $\{0\} \times [0,1]$ (the left boundary). Then, by letting $H_n(\rho)$ be the height of a binary search tree built from a sample of n points acording to ρ, we have

, Dubach, Féray (2024)

Let ρ be a permuton with a density. Further assume that this density is bounded on $[0,1]^2$ and positive continuous on a neighbourhood of $\{0\} \times [0,1]$ (the left boundary). Then, by letting $H_n(\rho)$ be the height of a binary search tree built from a sample of n points acording to ρ, we have

$$\frac{H_n(\rho)}{c^* \log n} \longrightarrow 1\,,$$

, Dubach, Féray (2024)

Let ρ be a permuton with a density. Further assume that this density is bounded on $[0,1]^2$ and positive continuous on a neighbourhood of $\{0\} \times [0,1]$ (the left boundary). Then, by letting $H_n(\rho)$ be the height of a binary search tree built from a sample of n points acording to ρ, we have

$$\frac{H_n(\rho)}{c^* \log n} \longrightarrow 1\,,$$

where the convergence occurs in probability and in $\mathbb{L}_p$, for any $p \geq 1$.

Our result

Our result

The conditions of the theorem can be understood as follows.

Our result

The conditions of the theorem can be understood as follows.

- Close to the left border of $[0,1]^2$, we need:

Our result

The conditions of the theorem can be understood as follows.

- Close to the left border of $[0,1]^2$, we need:
  - a density,

Our result

The conditions of the theorem can be understood as follows.

- Close to the left border of $[0,1]^2$, we need:
  - a density,
  - positive,

Our result

The conditions of the theorem can be understood as follows.

- Close to the left border of $[0,1]^2$, we need:
  - a density,
  - positive, and
  - continuous.

Our result

The conditions of the theorem can be understood as follows.

- Close to the left border of $[0,1]^2$, we need:
  - a density,
  - positive, and
  - continuous.
- On the rest of the square to the right, we need:

Our result

The conditions of the theorem can be understood as follows.

- Close to the left border of $[0,1]^2$, we need:
  - a density,
  - positive, and
  - continuous.
- On the rest of the square to the right, we need:
  - a bounded density.

Our result

The conditions of the theorem can be understood as follows.

- Close to the left border of $[0,1]^2$, we need:
  - a density,
  - positive, and
  - continuous.
- On the rest of the square to the right, we need:
  - a bounded density.

→ We will now see why these conditions are necessary.

References

- Corsini, B., Dubach, V., & Féray, V. (2024). **Binary search trees of permuton samples.** *arXiv preprint arXiv:2403.03151.*
- Devroye, L. (1986). **A note on the height of binary search trees.** *Journal of the ACM (JACM), vol. 33, no 3, p. 489-498.*

This project has received funding from the Future Leader Program of the LUE (Lorraine Université d'Excellence) initiative and from the European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie Grant Agreement No. 101034253 .

Thank you!

Thank you!

Thank you!

Thank you!

Table of Content

Binary search trees

Permutons

Our results

The conditions

Condition: left density

Condition: left density

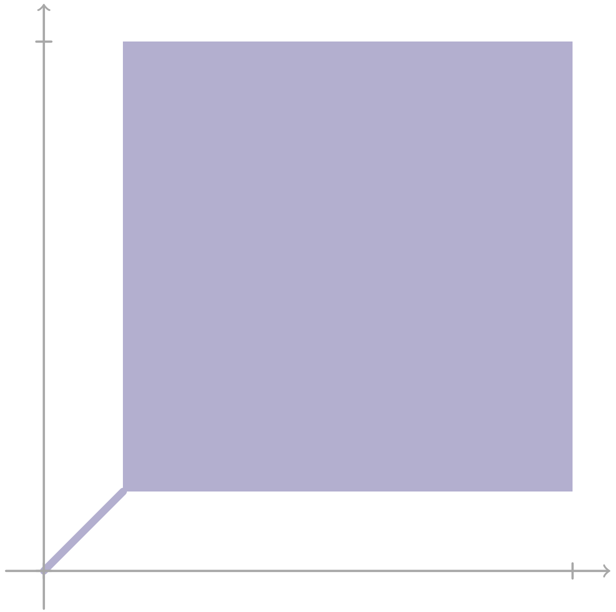
Consider the following distribution:

$$\rho(A) = \Big|\{x \in [0,\epsilon] : (x,x) \in A\}\Big|_1 + \frac{1}{1-\epsilon}\Big|A \cap [\epsilon,1]^2\Big|_2 .$$

Condition: left density

Consider the following distribution:

$$\rho(A) = \Big|\{x \in [0,\epsilon] : (x,x) \in A\}\Big|_1 + \frac{1}{1-\epsilon}\Big|A \cap [\epsilon,1]^2\Big|_2 .$$

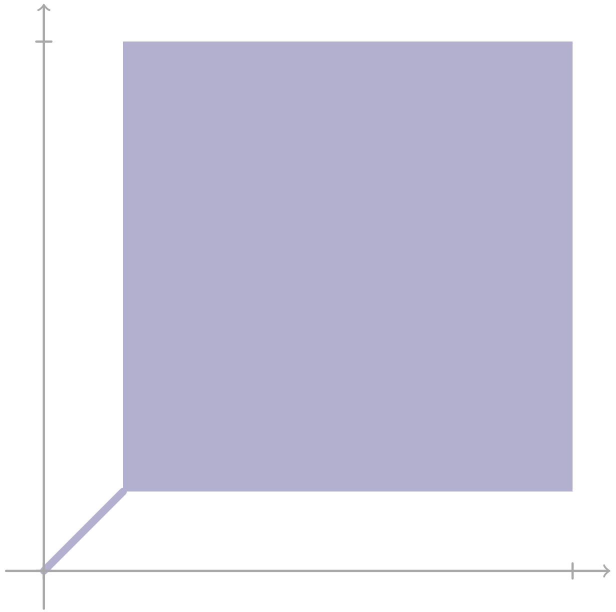


Condition: left density

Consider the following distribution:

$$\rho(A) = \Big|\{x \in [0,\epsilon] : (x,x) \in A\}\Big|_1 + \frac{1}{1-\epsilon}\Big|A \cap [\epsilon,1]^2\Big|_2 .$$

In that case, the tree looks like:

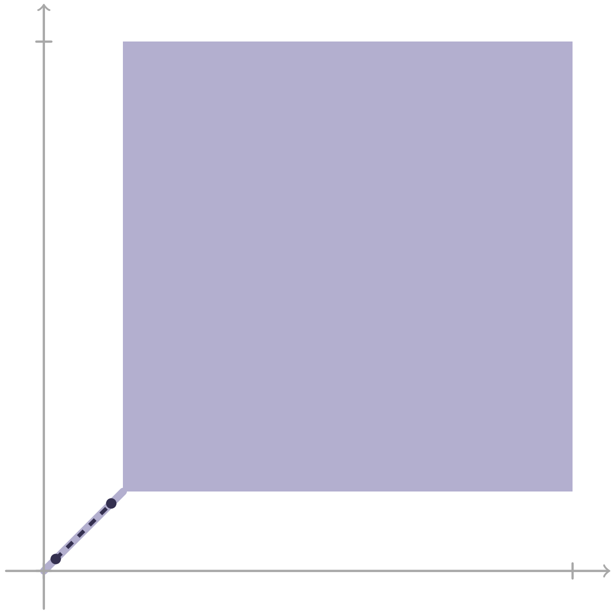


Consider the following distribution:

$$\rho(A) = \left|\left\{x \in [0,\epsilon] : (x,x) \in A\right\}\right|_1 + \frac{1}{1-\epsilon}\left|A \cap [\epsilon,1]^2\right|_2 .$$

In that case, the tree looks like:

- a right path composed of ϵn points,



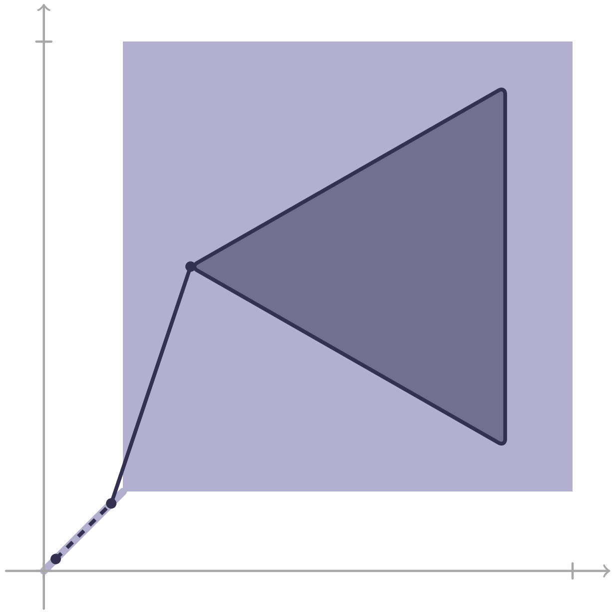
Condition: left density

Consider the following distribution:

$$\rho(A) = \left|\{x \in [0,\epsilon] : (x,x) \in A\}\right|_1 + \frac{1}{1-\epsilon}\left|A \cap [\epsilon,1]^2\right|_2 .$$

In that case, the tree looks like:

- a right path composed of ϵn points,
- to which a RBST on $(1-\epsilon)n$ points is attached.



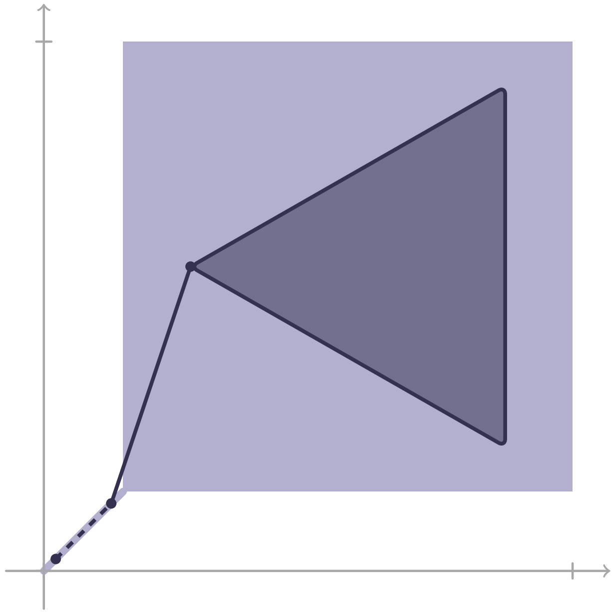
Condition: left density

Consider the following distribution:

$$\rho(A) = \Big|\{x \in [0,\epsilon] : (x,x) \in A\}\Big|_1 + \frac{1}{1-\epsilon}\Big|A \cap [\epsilon,1]^2\Big|_2 .$$

In that case, the tree looks like:

- a right path composed of ϵn points,
- to which a RBST on $(1-\epsilon)n$ points is attached.



→ The height is $\epsilon n + c^* \log(1-\epsilon)n \sim \epsilon n$.

Condition: positive left density

Condition: positive left density

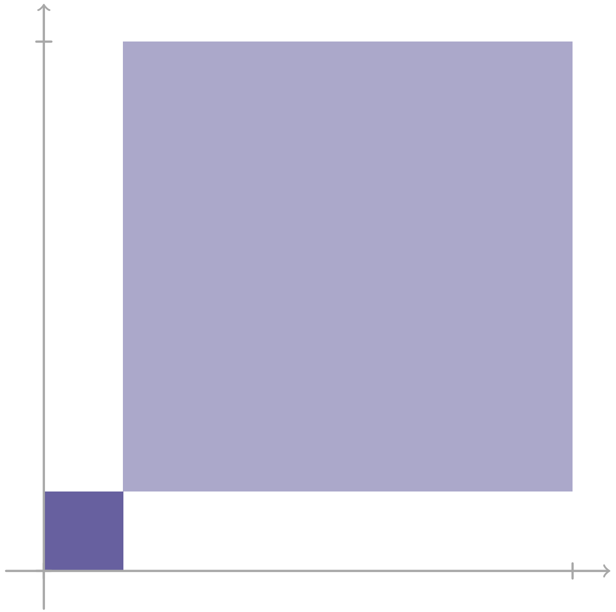
Consider the following distribution:

$$\rho(A) = \frac{1}{\epsilon}\Big|A \cap [0,\epsilon]^2\Big|_2 + \frac{1}{1-\epsilon}\Big|A \cap [\epsilon,1]^2\Big|_2 .$$

Condition: positive left density

Consider the following distribution:

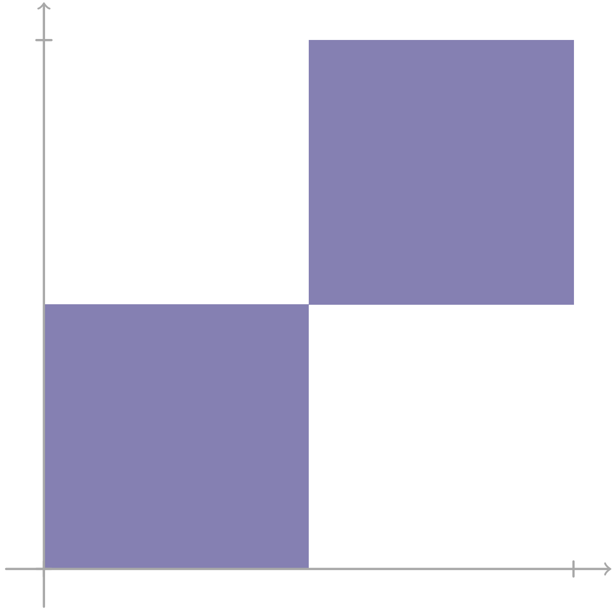
$$\rho(A) = \frac{1}{\epsilon}\Big|A \cap [0,\epsilon]^2\Big|_2 + \frac{1}{1-\epsilon}\Big|A \cap [\epsilon,1]^2\Big|_2 .$$



Condition: positive left density

Consider the following distribution:

$$\rho(A) = 2\left[\left|A \cap [0,1/2]^2\right|_2 + \left|A \cap [1/2,1]^2\right|_2\right].$$

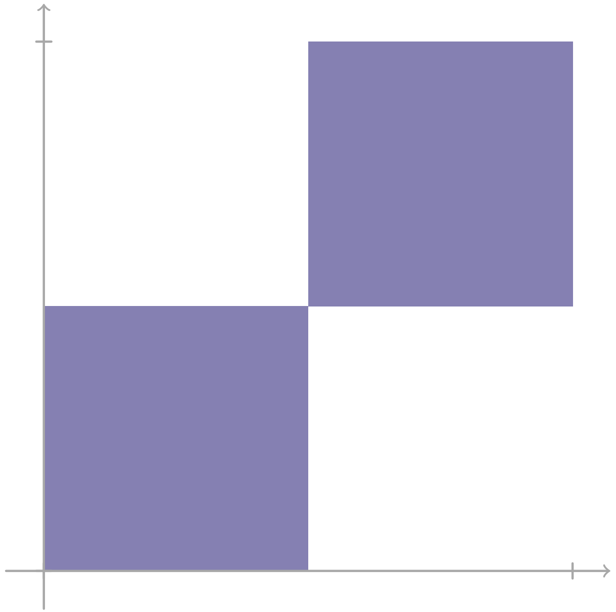


Condition: positive left density

Consider the following distribution:

$$\rho(A) = 2\left[\left|A \cap [0,1/2]^2\right|_2 + \left|A \cap [1/2,1]^2\right|_2\right].$$

In that case, the tree looks like:



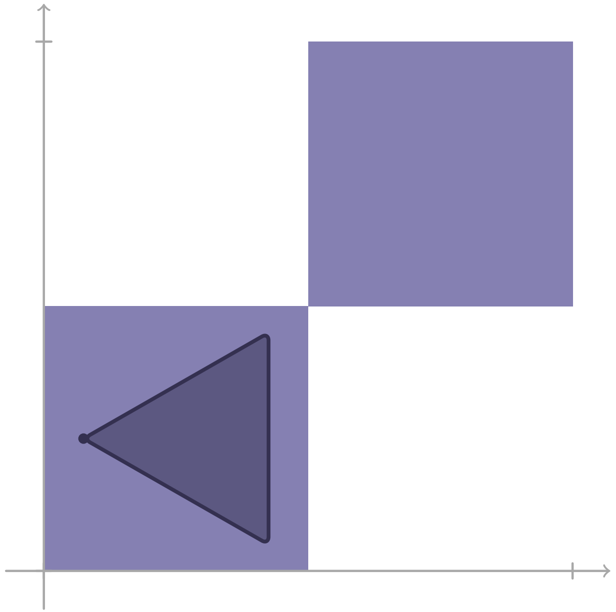
Condition: positive left density

Consider the following distribution:

$$\rho(A) = 2\left[\left|A \cap [0,1/2]^2\right|_2 + \left|A \cap [1/2,1]^2\right|_2\right].$$

In that case, the tree looks like:

- a RBST on $n/2$ points,



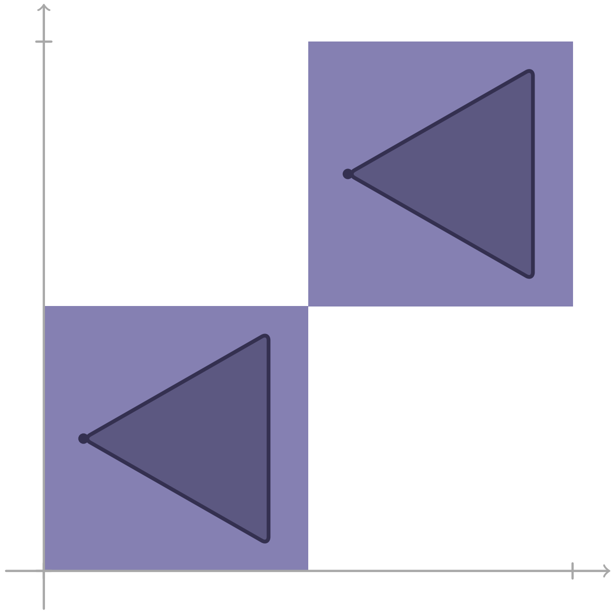
Condition: positive left density

Consider the following distribution:

$$\rho(A) = 2\left[\left|A \cap [0,1/2]^2\right|_2 + \left|A \cap [1/2,1]^2\right|_2\right].$$

In that case, the tree looks like:

- a RBST on $n/2$ points,
- to which a RBST on $n/2$ points is attached;



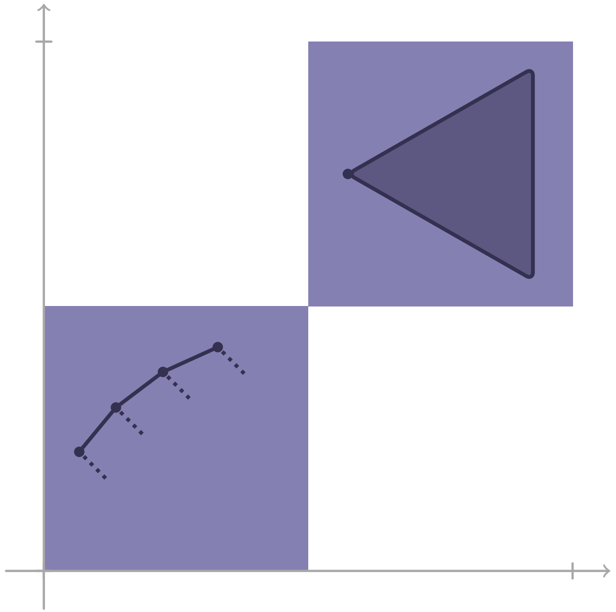
Condition: positive left density

Consider the following distribution:

$$\rho(A) = 2\left[\left|A \cap [0,1/2]^2\right|_2 + \left|A \cap [1/2,1]^2\right|_2\right].$$

In that case, the tree looks like:

- a RBST on $n/2$ points,
- to which a RBST on $n/2$ points is attached;



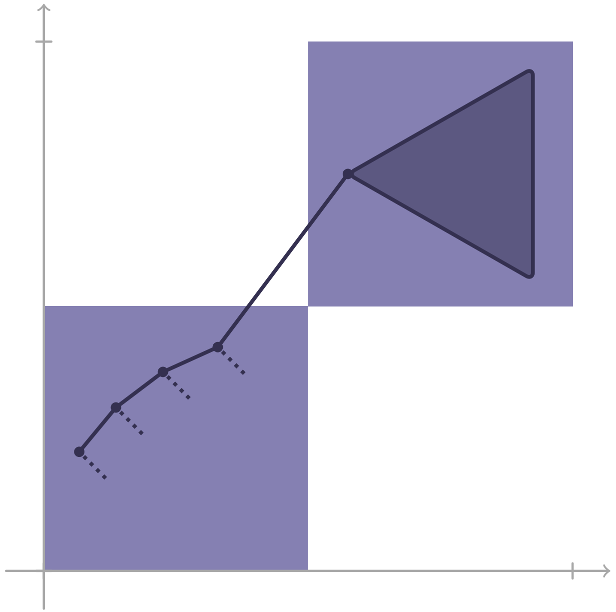
Condition: positive left density

Consider the following distribution:

$$\rho(A) = 2\left[\left|A \cap [0,1/2]^2\right|_2 + \left|A \cap [1/2,1]^2\right|_2\right].$$

In that case, the tree looks like:

- a RBST on $n/2$ points,
- to which a RBST on $n/2$ points is attached;
- this attachment is at the end of the rightmost path.



Condition: positive left density

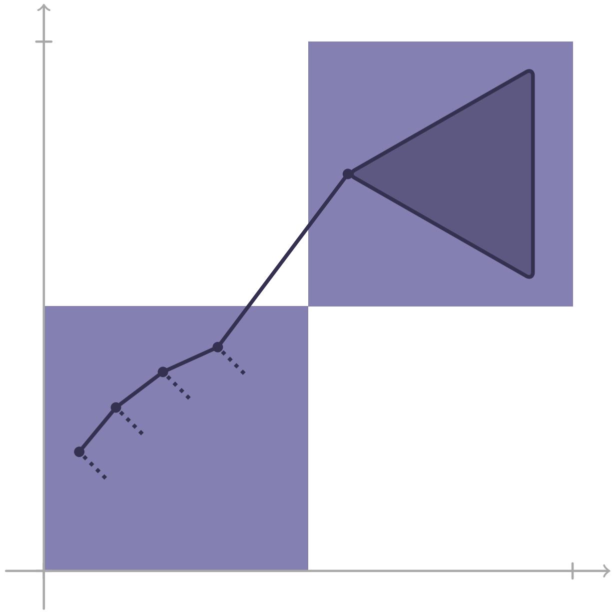
Consider the following distribution:

$$\rho(A) = 2\left[\left|A \cap [0,1/2]^2\right|_2 + \left|A \cap [1/2,1]^2\right|_2\right].$$

In that case, the tree looks like:

- a RBST on $n/2$ points,
- to which a RBST on $n/2$ points is attached;
- this attachment is at the end of the rightmost path.

The rightmost path of a RBST has length $\log n$.



Condition: positive left density

Consider the following distribution:

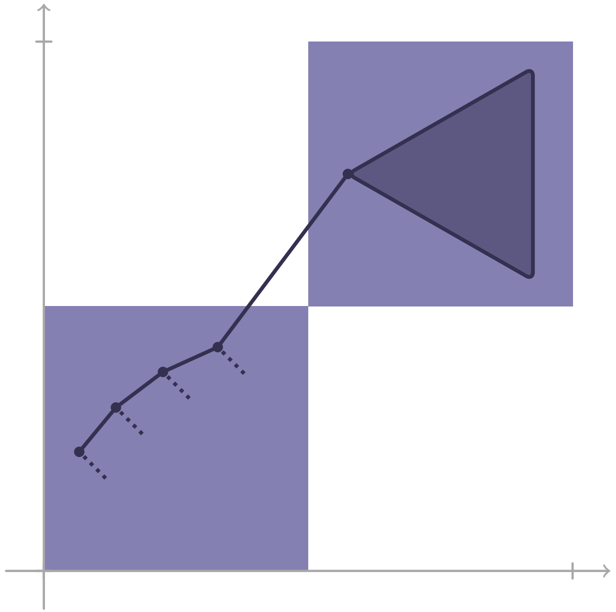
$$\rho(A) = 2\left[\left|A \cap [0,1/2]^2\right|_2 + \left|A \cap [1/2,1]^2\right|_2\right].$$

In that case, the tree looks like:

- a RBST on $n/2$ points,
- to which a RBST on $n/2$ points is attached;
- this attachment is at the end of the rightmost path.

The rightmost path of a RBST has length $\log n$.

→ The height is $\log(n/2) + c^* \log(n/2) \sim (c^* + 1)\log n$.



Condition: positive continuous left density

Condition: positive continuous left density

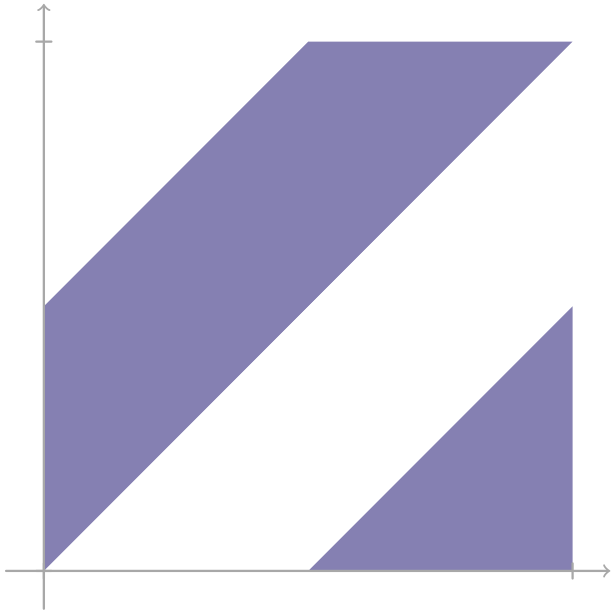
Consider the following distribution:

$$\rho(A) = 2\left[\left|A \cap \left\{(x,y) \in [0,1]^2 : y - x \in [0,1/2] \bmod 1\right\}\right|_2\right].$$

Condition: positive continuous left density

Consider the following distribution:

$$\rho(A) = 2\left[\left|A \cap \left\{(x,y) \in [0,1]^2 : y - x \in [0,1/2] \bmod 1\right\}\right|_2\right].$$

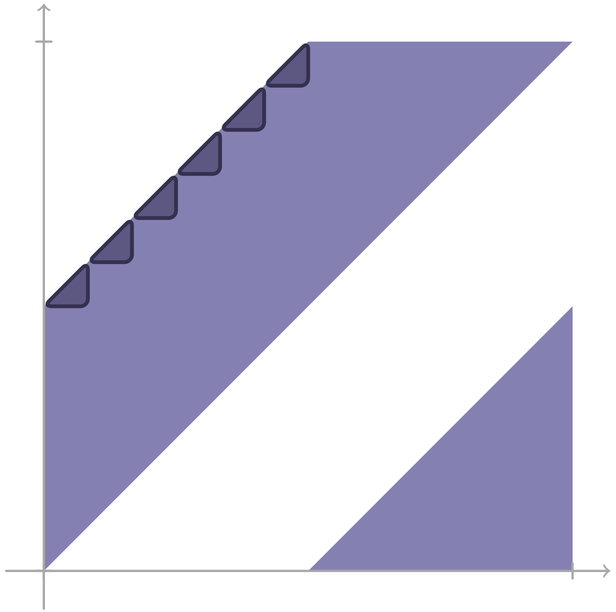


Condition: positive continuous left density

Consider the following distribution:

$$\rho(A) = 2\left[\left|A \cap \left\{(x,y) \in [0,1]^2 : y - x \in [0,1/2] \bmod 1\right\}\right|_2\right].$$

In that case, we consider the top $\sqrt{n}/2$ triangles.



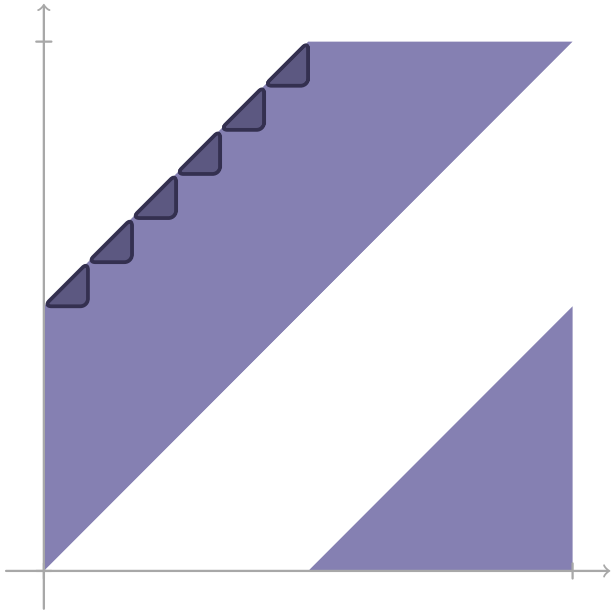
Condition: positive continuous left density

Consider the following distribution:

$$\rho(A) = 2\left[\left|A \cap \left\{(x,y) \in [0,1]^2 : y - x \in [0,1/2] \bmod 1\right\}\right|_2\right].$$

In that case, we consider the top $\sqrt{n}/2$ triangles.

- Each such triangle contains 1 points on average.



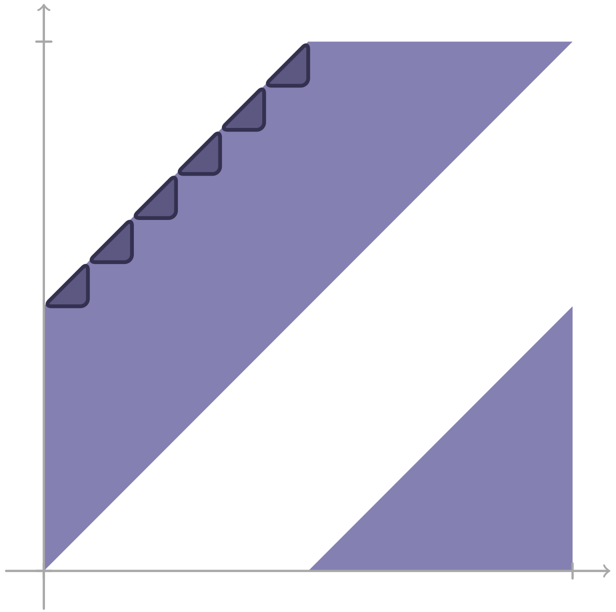
Condition: positive continuous left density

Consider the following distribution:

$$\rho(A) = 2\left[\left|A \cap \{(x,y) \in [0,1]^2 : y - x \in [0,1/2] \bmod 1\}\right|_2\right].$$

In that case, we consider the top $\sqrt{n}/2$ triangles.

- Each such triangle contains 1 points on average.
- They form part of the rightmost path of the tree.



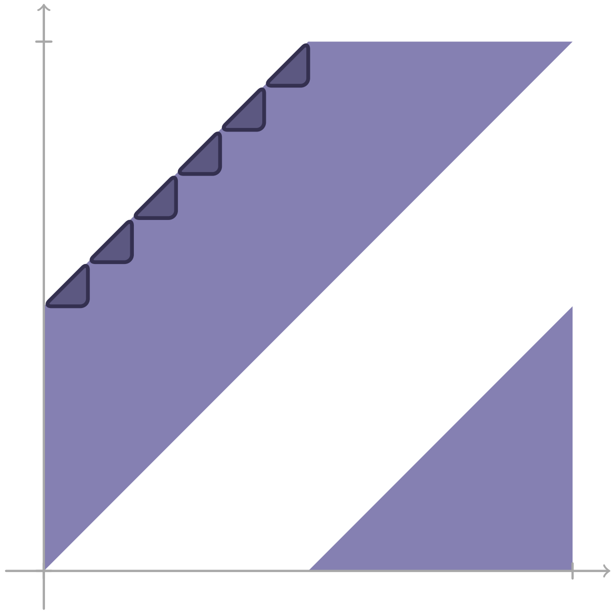
Condition: positive continuous left density

Consider the following distribution:

$$\rho(A) = 2\left[\left|A \cap \left\{(x,y) \in [0,1]^2 : y - x \in [0, 1/2] \bmod 1\right\}\right|_2\right].$$

In that case, we consider the top $\sqrt{n}/2$ triangles.

- Each such triangle contains 1 points on average.
- They form part of the rightmost path of the tree.
- The height is at least the number of triangles with 1 point.



Condition: positive continuous left density

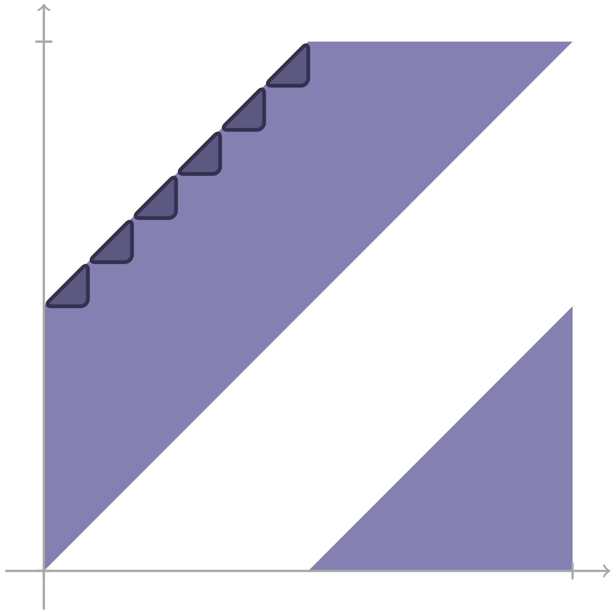
Consider the following distribution:

$$\rho(A) = 2\left[\left|A \cap \left\{(x,y) \in [0,1]^2 : y - x \in [0,1/2] \bmod 1\right\}\right|_2\right].$$

In that case, we consider the top $\sqrt{n}/2$ triangles.

- Each such triangle contains 1 points on average.
- They form part of the rightmost path of the tree.
- The height is at least the number of triangles with 1 point.

$\rightarrow$ The height is at least of order $\sqrt{n}$.



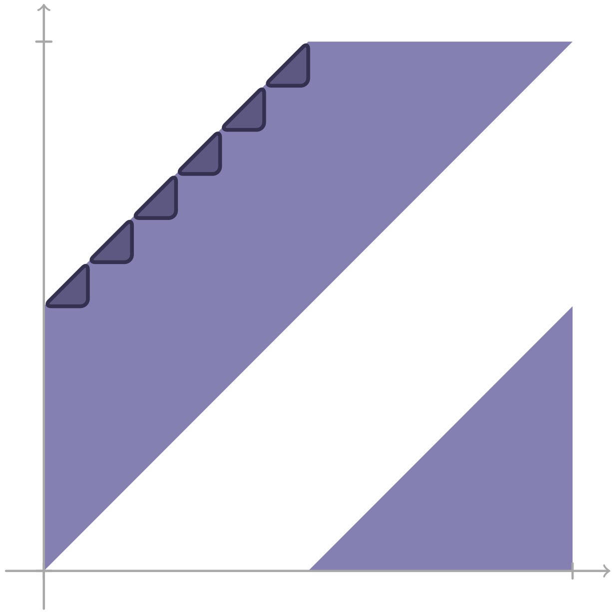
Condition: positive continuous left density

Consider the following distribution:

$$\rho(A) = 2\left[\left|A \cap \{(x,y) \in [0,1]^2 : y - x \in [0,1/2] \bmod 1\}\right|_2\right].$$

In that case, we consider the top $\sqrt{n}/2$ triangles.

- Each such triangle contains 1 points on average.
- They form part of the rightmost path of the tree.
- The height is at least the number of triangles with 1 point.



→ The height is at least of order $\sqrt{n}$.

→ By "smoothing" the density, we can get a height of order $n^{1/2-\delta}$.

Condition: right density

Condition: right density

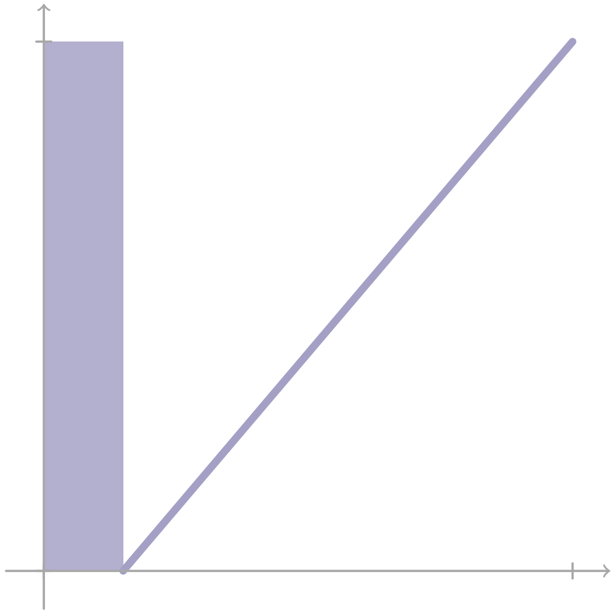
Consider the following distribution:

$$\rho(A) = \Big|A \cap [0,\epsilon] \times [0,1]\Big|_2 + \Big|\big\{x \in [\epsilon,1] : \big(x, (x-\epsilon)/(1-\epsilon)\big) \in A\big\}\Big|_1 .$$

Condition: right density

Consider the following distribution:

$$\rho(A) = \Big|A \cap [0,\epsilon] \times [0,1]\Big|_2 + \Big|\{x \in [\epsilon, 1] : \big(x, (x-\epsilon)/(1-\epsilon)\big) \in A\}\Big|_1 .$$

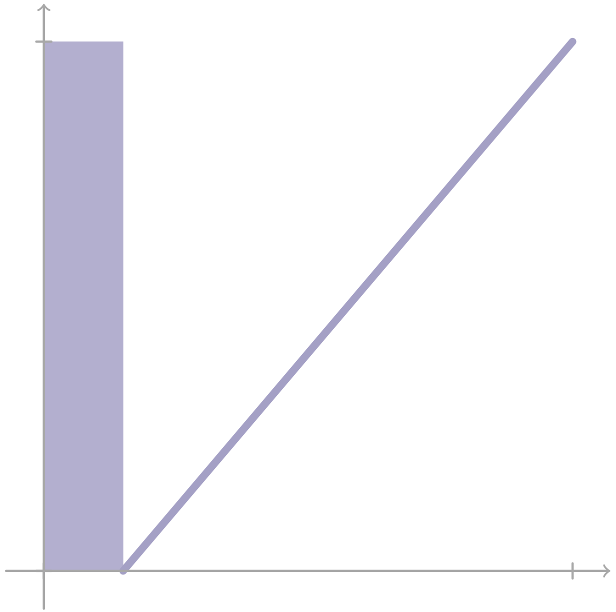


Condition: right density

Consider the following distribution:

$$\rho(A) = \Big|A \cap [0,\epsilon] \times [0,1]\Big|_2 + \Big|\{x \in [\epsilon,1] : \big(x, (x-\epsilon)/(1-\epsilon)\big) \in A\}\Big|_1 .$$

In that case, the tree looks like:



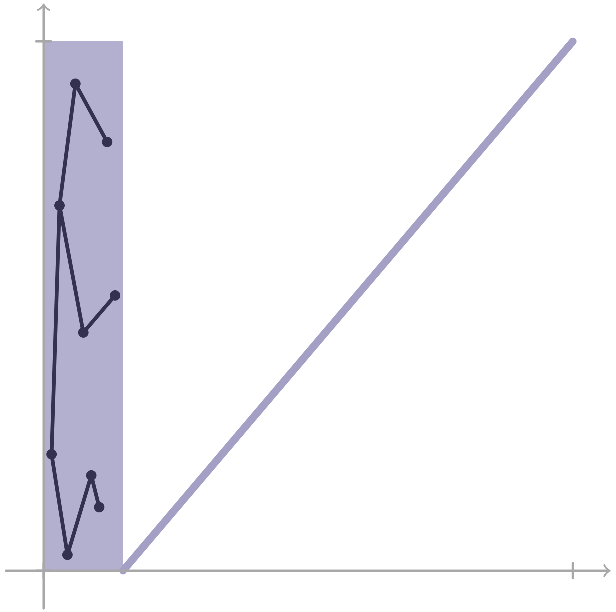
Condition: right density

Consider the following distribution:

$$\rho(A) = \left| A \cap [0,\epsilon] \times [0,1] \right|_2 + \left| \left\{ x \in [\epsilon, 1] : \big(x, (x-\epsilon)/(1-\epsilon)\big) \in A \right\} \right|_1 .$$

In that case, the tree looks like:

- a RBST on ϵn nodes at the top,



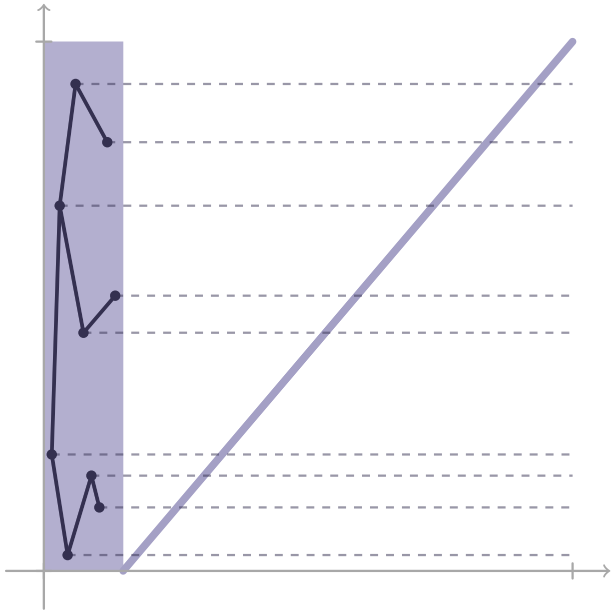
Condition: right density

Consider the following distribution:

$$\rho(A) = \left| A \cap [0, \epsilon] \times [0, 1] \right|_2 + \left| \left\{ x \in [\epsilon, 1] : \big(x, (x - \epsilon)/(1 - \epsilon) \big) \in A \right\} \right|_1 .$$

In that case, the tree looks like:

- a RBST on ϵn nodes at the top,



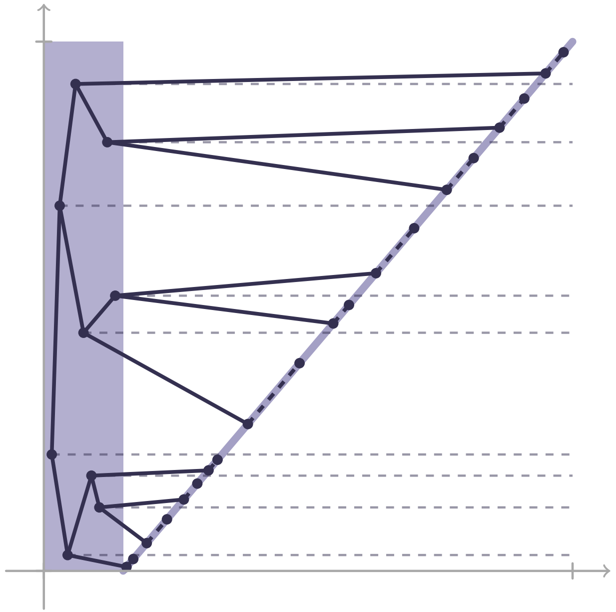
Condition: right density

Consider the following distribution:

$$\rho(A) = \Big|A \cap [0,\epsilon] \times [0,1]\Big|_2 + \Big|\{x \in [\epsilon,1] : \big(x,(x-\epsilon)/(1-\epsilon)\big) \in A\}\Big|_1 .$$

In that case, the tree looks like:

- a RBST on ϵn nodes at the top,
- on which we attach linear subtrees.



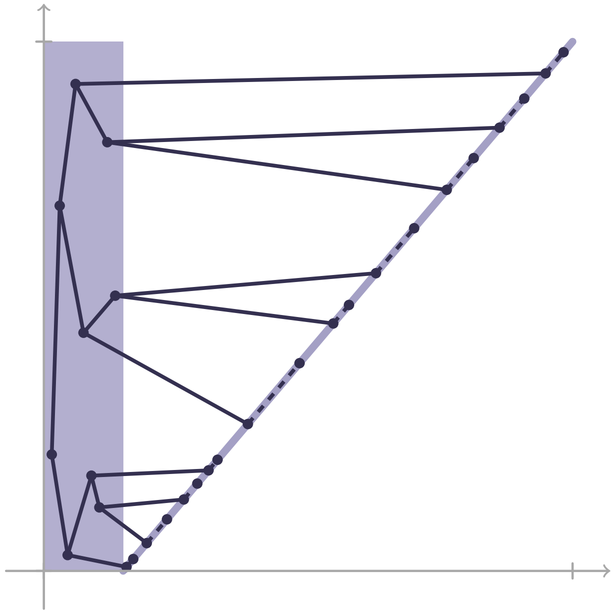
Condition: right density

Consider the following distribution:

$$\rho(A) = \Big| A \cap [0,\epsilon] \times [0,1] \Big|_2 + \Big| \{ x \in [\epsilon, 1] : \big(x, (x-\epsilon)/(1-\epsilon)\big) \in A \} \Big|_1 .$$

In that case, the tree looks like:

- a RBST on ϵn nodes at the top,
- on which we attach linear subtrees;



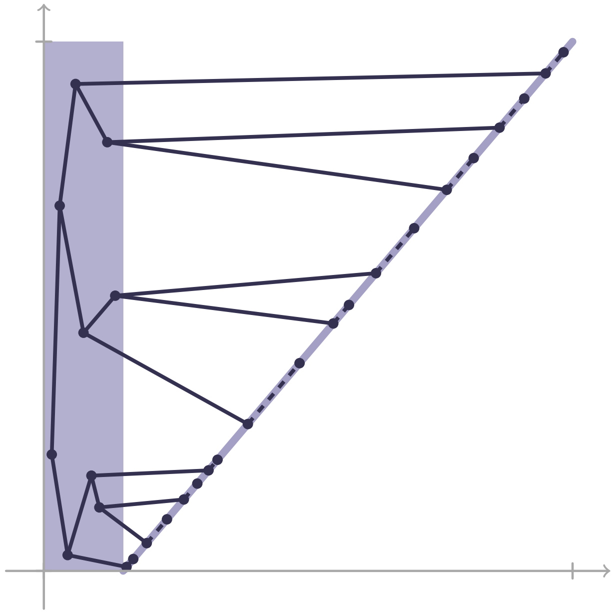
Condition: right density

Consider the following distribution:

$$\rho(A) = \Big|A \cap [0,\epsilon] \times [0,1]\Big|_2 + \Big|\{x \in [\epsilon,1] : \big(x, (x-\epsilon)/(1-\epsilon)\big) \in A\}\Big|_1 .$$

In that case, the tree looks like:

- a RBST on ϵn nodes at the top,
- on which we attach linear subtrees;
- these subtrees have a bounded average size,

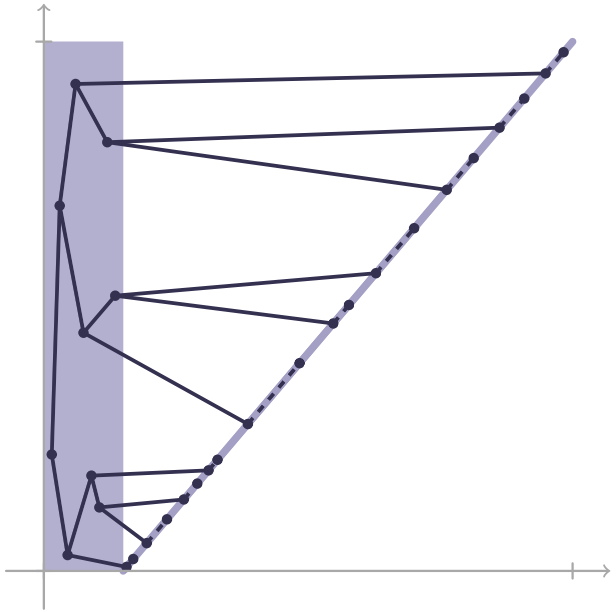


Consider the following distribution:

$$\rho(A) = \Big|A \cap [0,\epsilon] \times [0,1]\Big|_2 + \Big|\big\{x \in [\epsilon, 1] : \big(x, (x-\epsilon)/(1-\epsilon)\big) \in A\big\}\Big|_1 .$$

In that case, the tree looks like:

- a RBST on ϵn nodes at the top,
- on which we attach linear subtrees;
- these subtrees have a bounded average size,
- but the largest such subtree has size $(1-\epsilon)\epsilon^{-1} \log n$.



Condition: right density

Consider the following distribution:

$$\rho(A) = \Big|A \cap [0,\epsilon] \times [0,1]\Big|_2 + \Big|\big\{x \in [\epsilon, 1] : \big(x, (x-\epsilon)/(1-\epsilon)\big) \in A\big\}\Big|_1 .$$

In that case, the tree looks like:

- a RBST on ϵn nodes at the top,
- on which we attach linear subtrees;
- these subtrees have a bounded average size,
- but the largest such subtree has size $(1-\epsilon)\epsilon^{-1}\log n$.

$\rightarrow$ The height is at least $(1-\epsilon)\epsilon^{-1}\log n$.