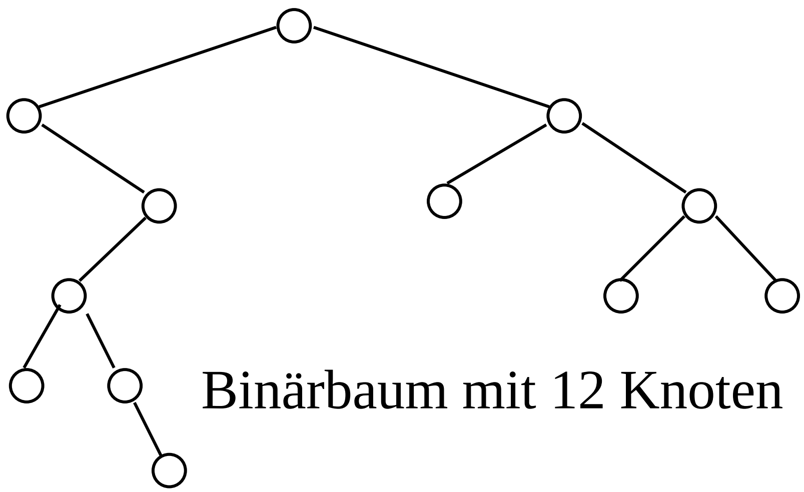
Zufällige diskrete Strukturen

Ralph Neininger
Institut für Mathematik
J. W. Goethe-Universität Frankfurt a.M.

3. Alumni Treffen des Instituts für Mathematik

27. November 2010

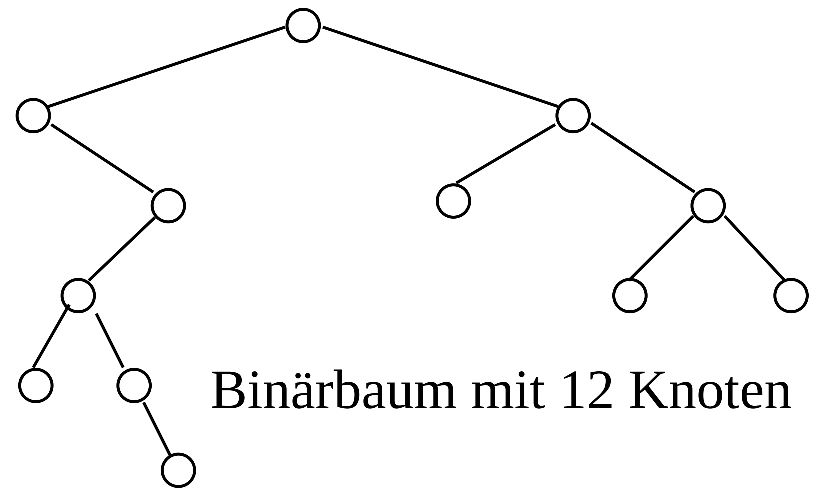
Binäre Bäume



Binärbaum mit 12 Knoten

Binäre Bäume

Stochastische Modelle:

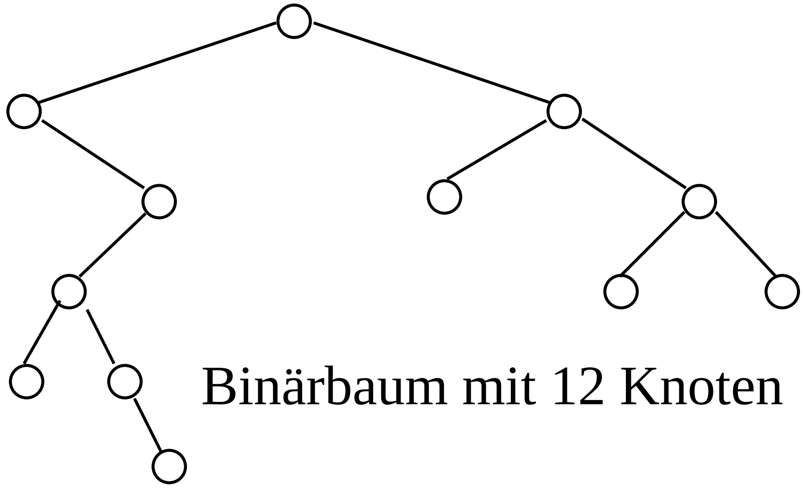


Binärbaum mit 12 Knoten

Binäre Bäume

Stochastische Modelle:

— Catalan-Modell



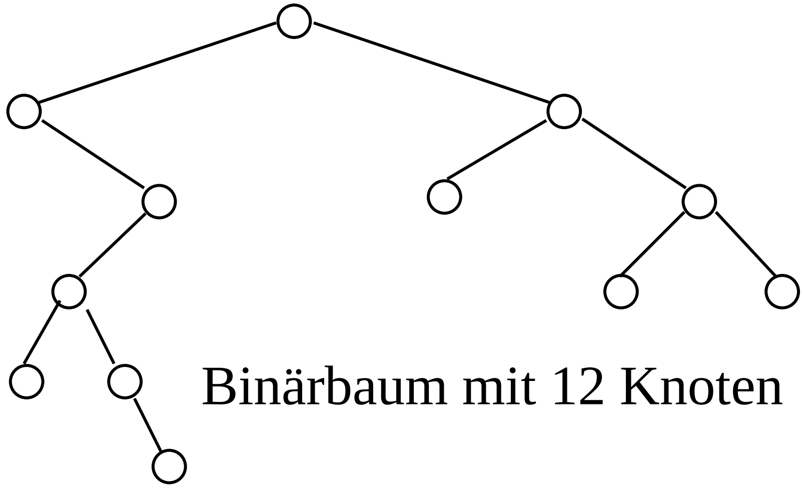
Binärbaum mit 12 Knoten

Binäre Bäume

Stochastische Modelle:

— Catalan-Modell

— Permutationsmodell
(Zufälliger Binärsuchbaum)



Binärbaum mit 12 Knoten

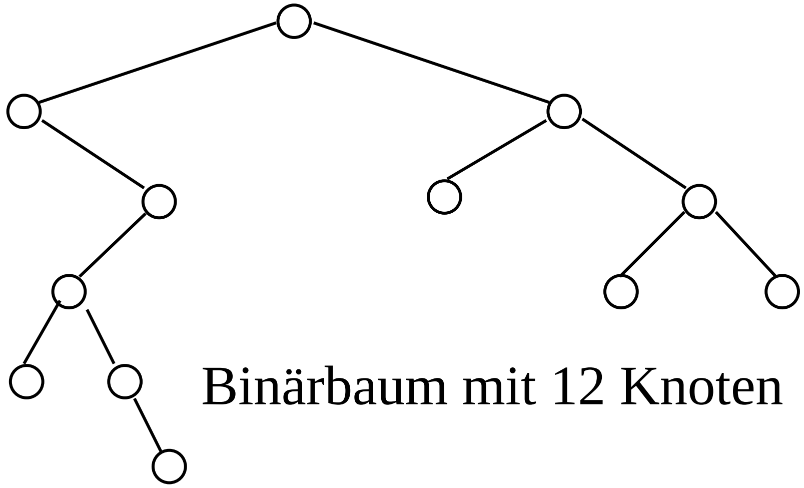
Binäre Bäume

Stochastische Modelle:

— Catalan-Modell

— Permutationsmodell
(Zufälliger Binärsuchbaum)

— Zufälliger Suffixsuchbaum



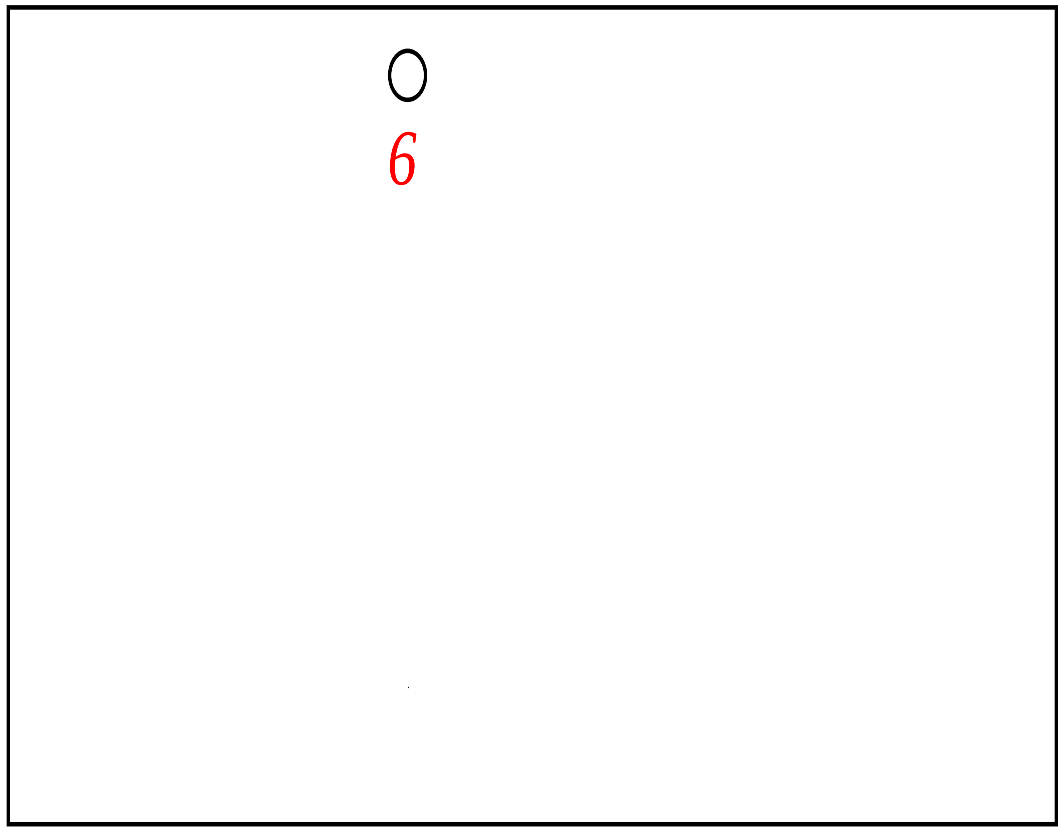
Binärbaum mit 12 Knoten

Binärsuchbaum

Liste der Daten: 6, 1, 8, 7, 5, 3, 10, 2, 11, 4, 9.

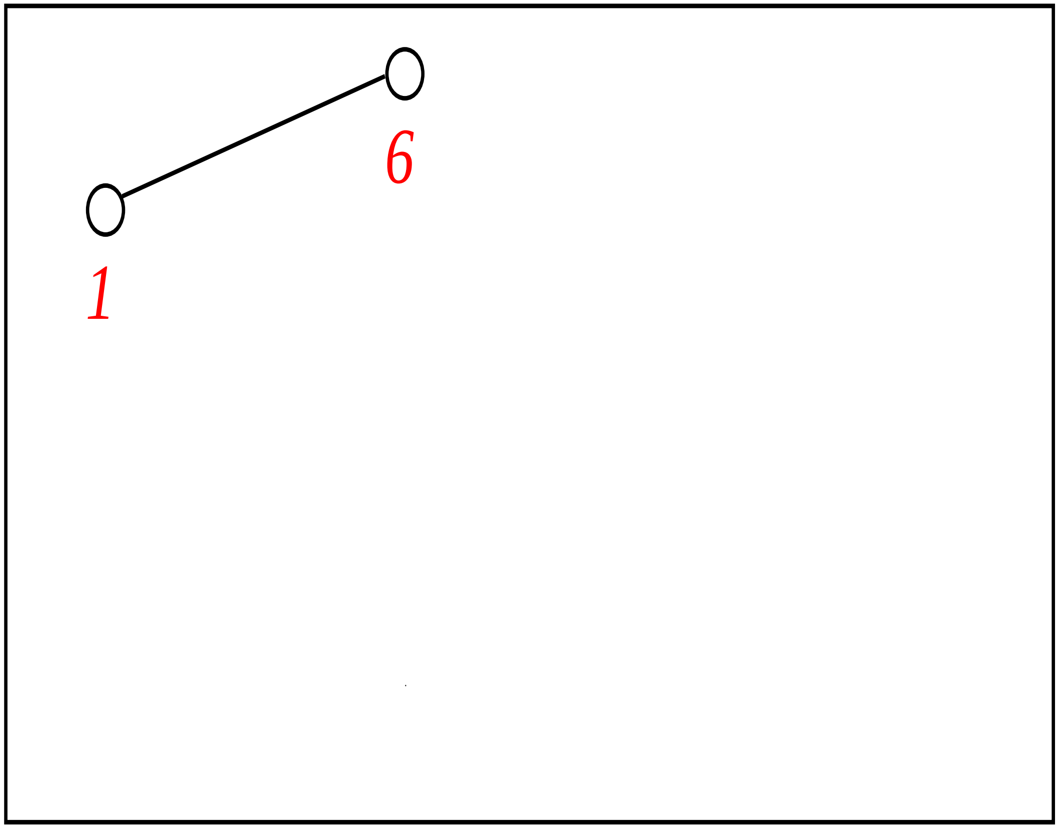
Binärsuchbaum

Liste der Daten: 6, 1, 8, 7, 5, 3, 10, 2, 11, 4, 9.



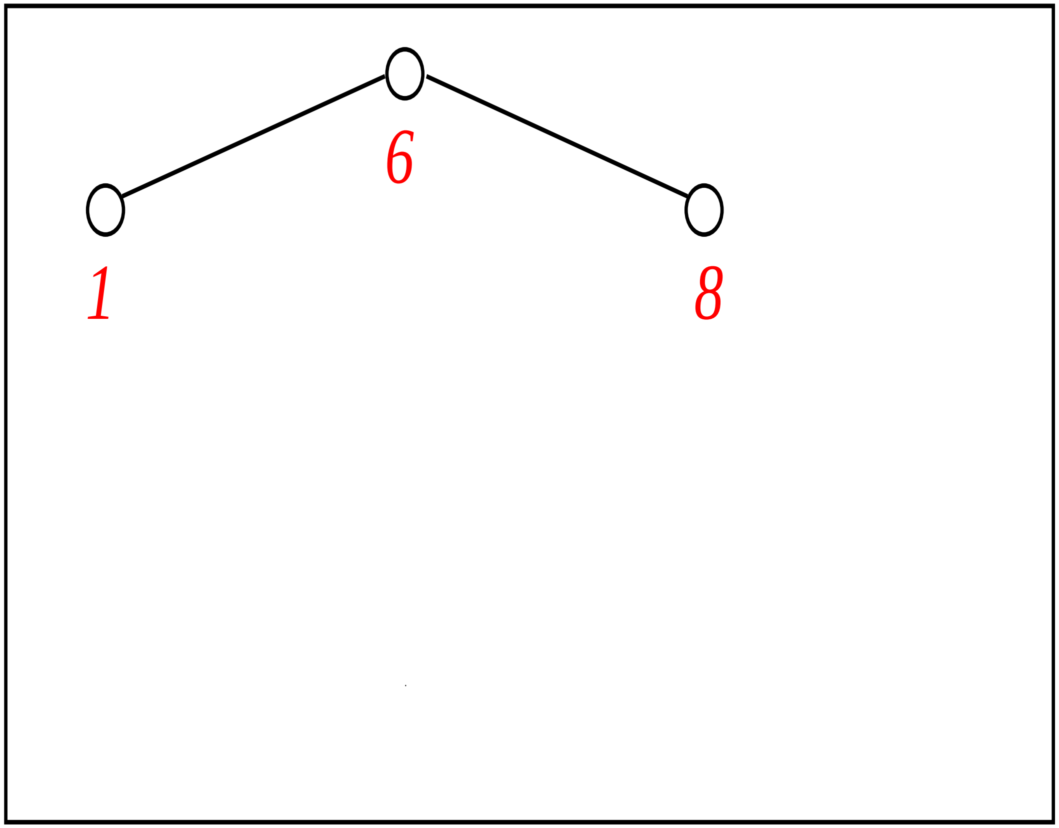
Binärsuchbaum

Liste der Daten: 6, 1, 8, 7, 5, 3, 10, 2, 11, 4, 9.



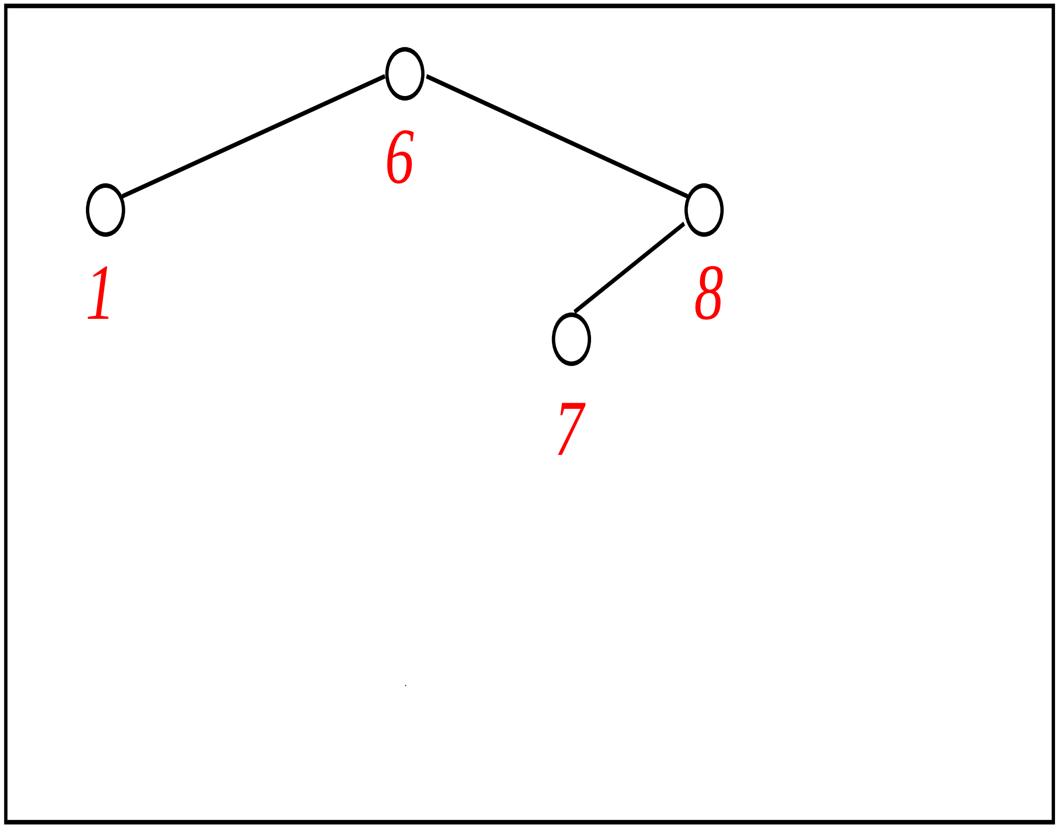
Binärsuchbaum

Liste der Daten: 6, 1, 8, 7, 5, 3, 10, 2, 11, 4, 9.



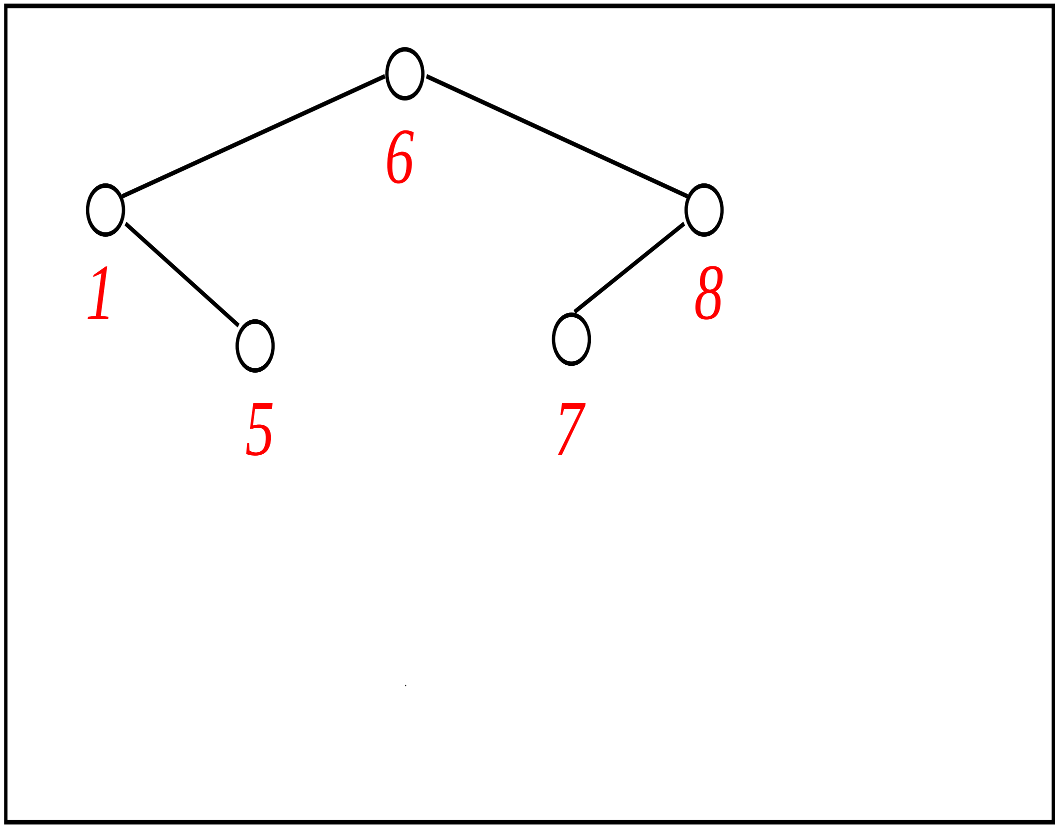
Binärsuchbaum

Liste der Daten: 6, 1, 8, 7, 5, 3, 10, 2, 11, 4, 9.



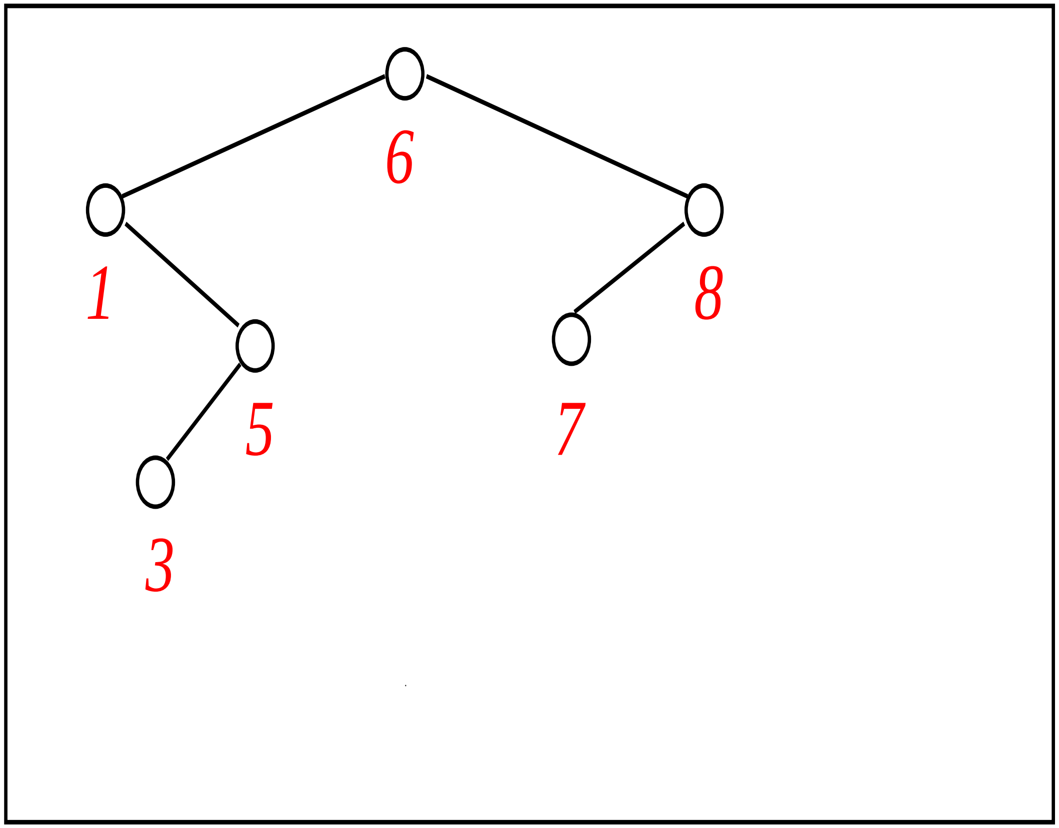
Binärsuchbaum

Liste der Daten: $6, 1, 8, 7, 5, 3, 10, 2, 11, 4, 9.$



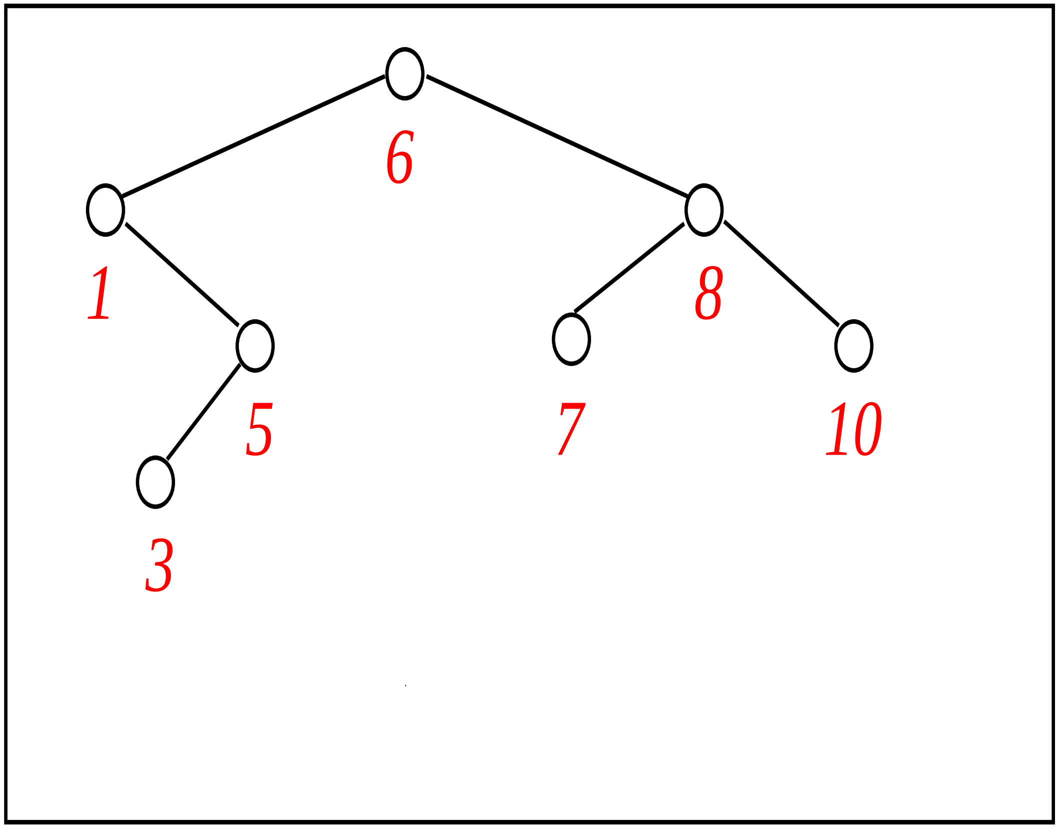
Binärsuchbaum

Liste der Daten: 6, 1, 8, 7, 5, 3, 10, 2, 11, 4, 9.



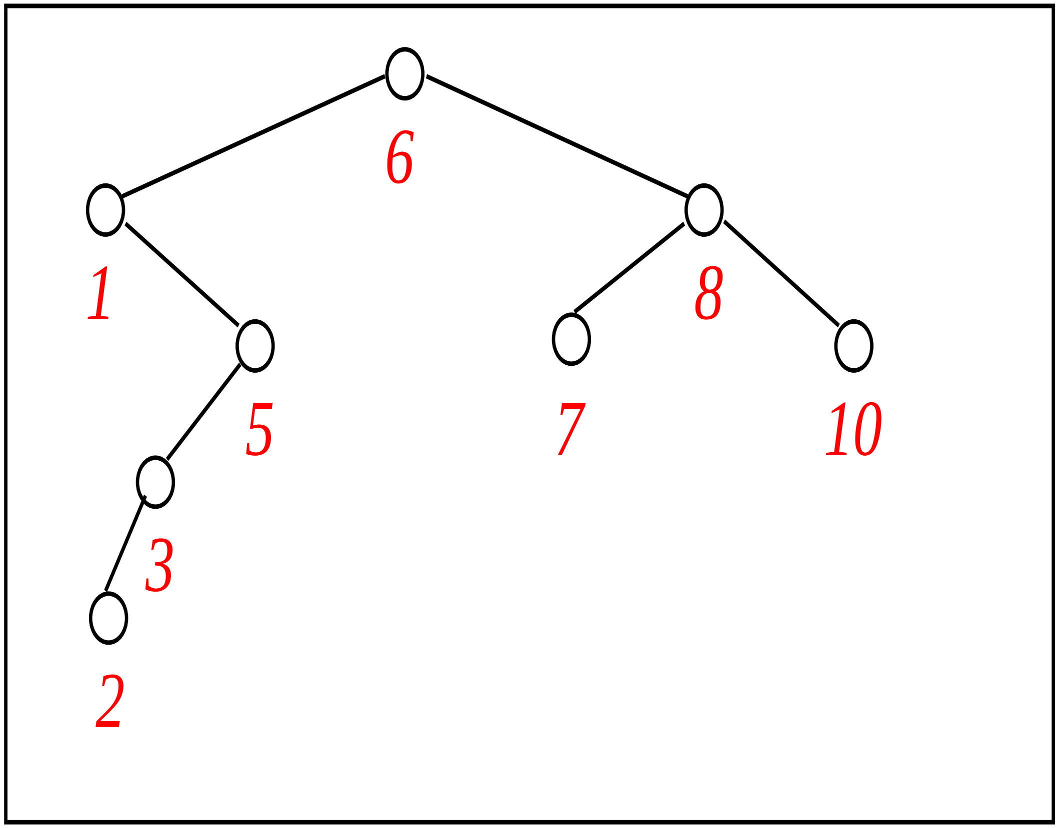
Binärsuchbaum

Liste der Daten: 6, 1, 8, 7, 5, 3, 10, 2, 11, 4, 9.



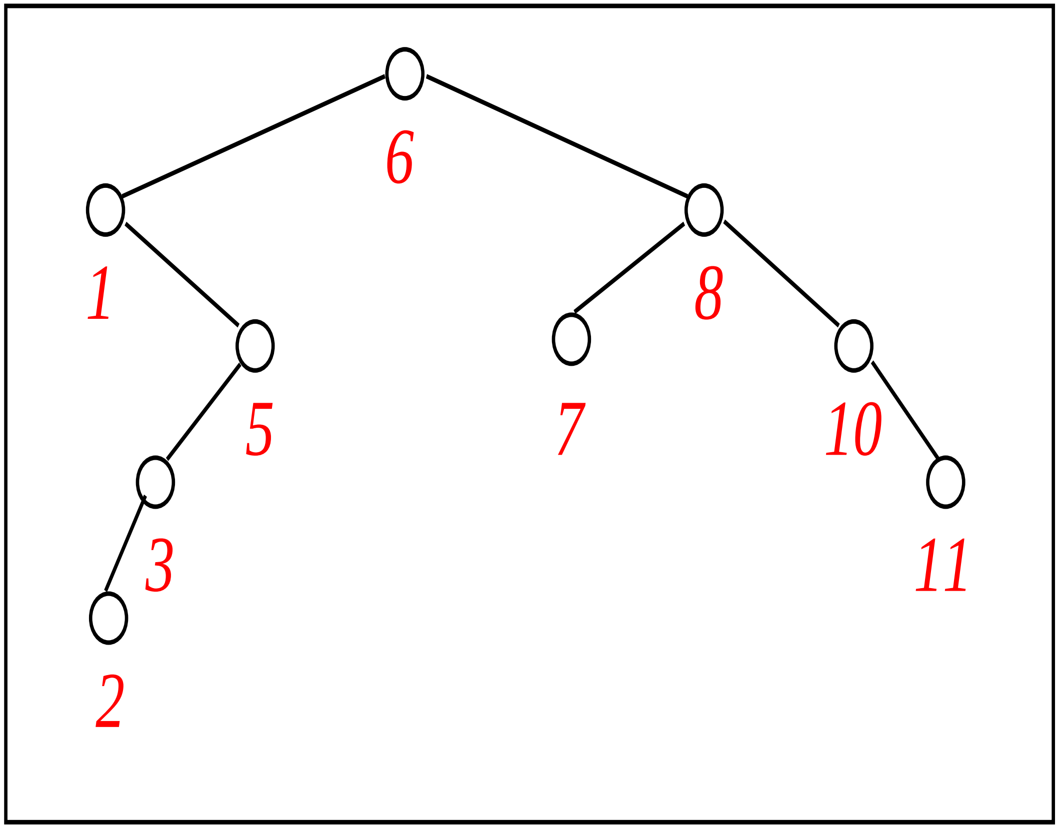
Binärsuchbaum

Liste der Daten: 6, 1, 8, 7, 5, 3, 10, 2, 11, 4, 9.



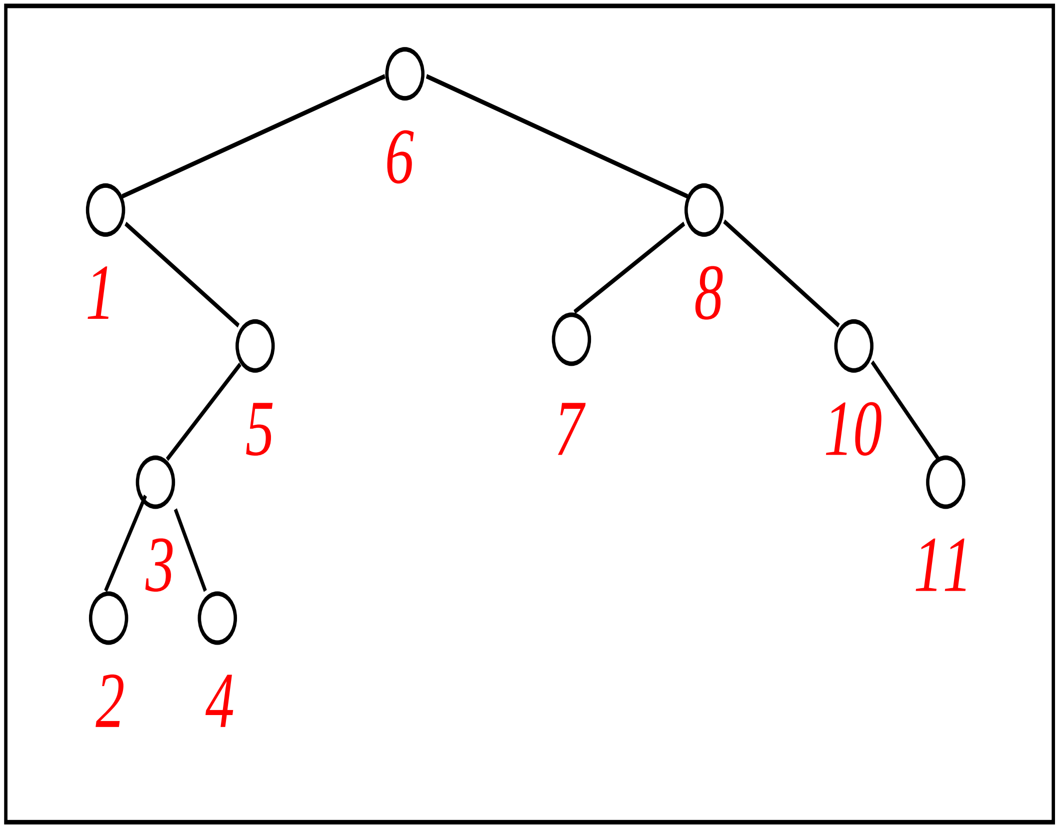
Binärsuchbaum

Liste der Daten: 6, 1, 8, 7, 5, 3, 10, 2, 11, 4, 9.



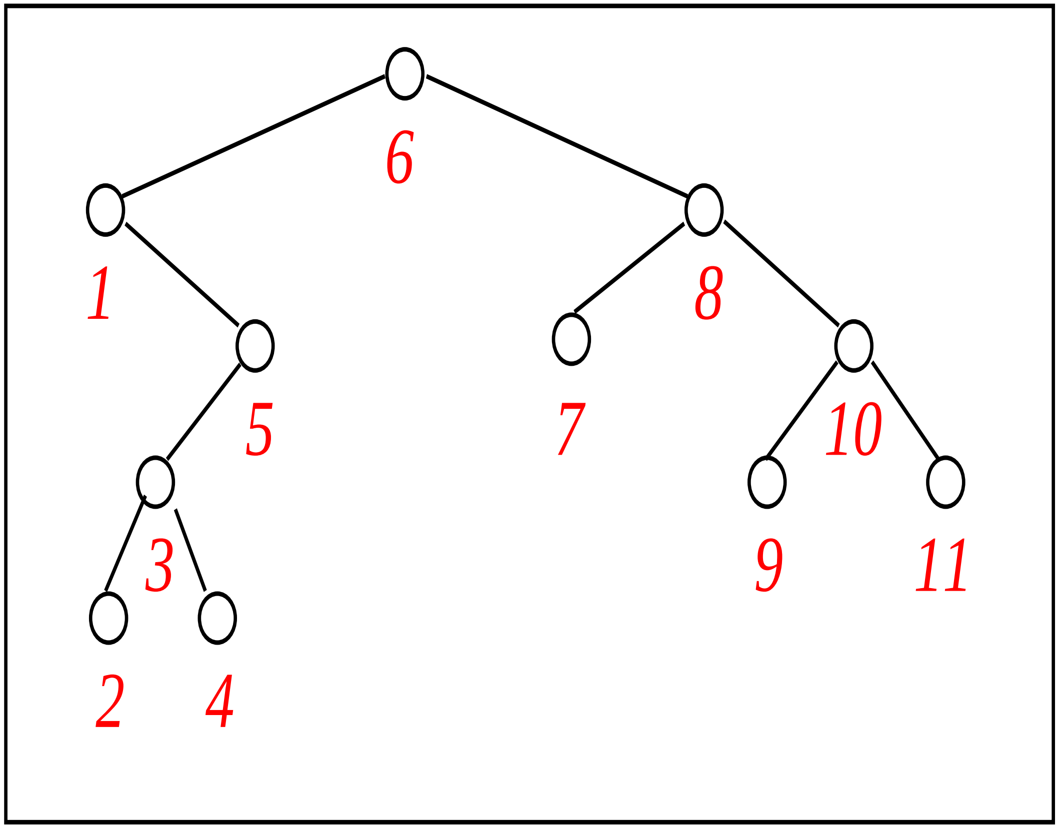
Binärsuchbaum

Liste der Daten: 6, 1, 8, 7, 5, 3, 10, 2, 11, 4, 9.

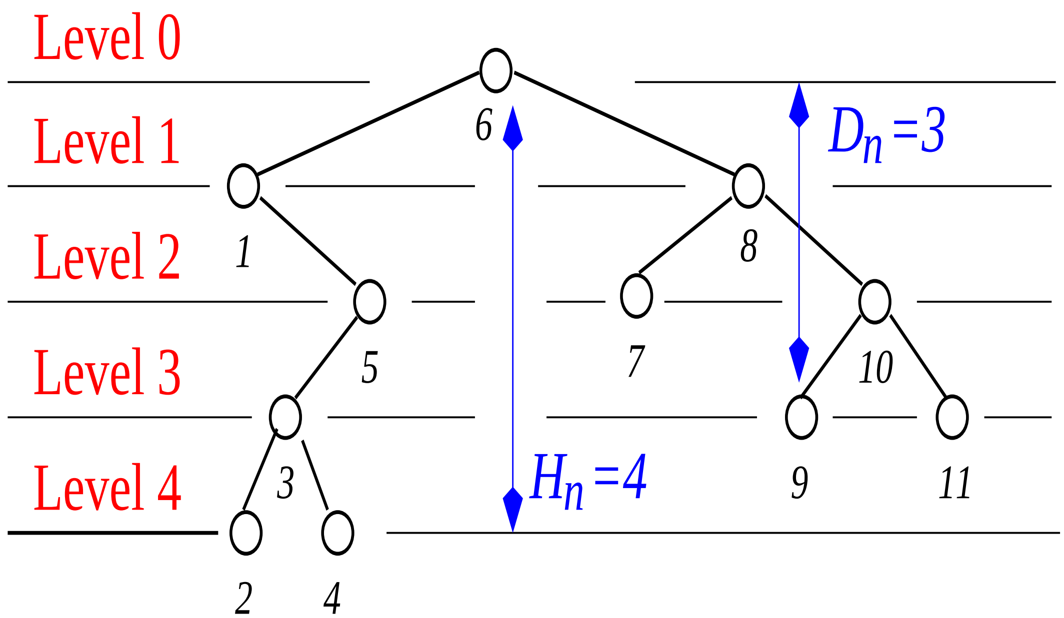


Binärsuchbaum

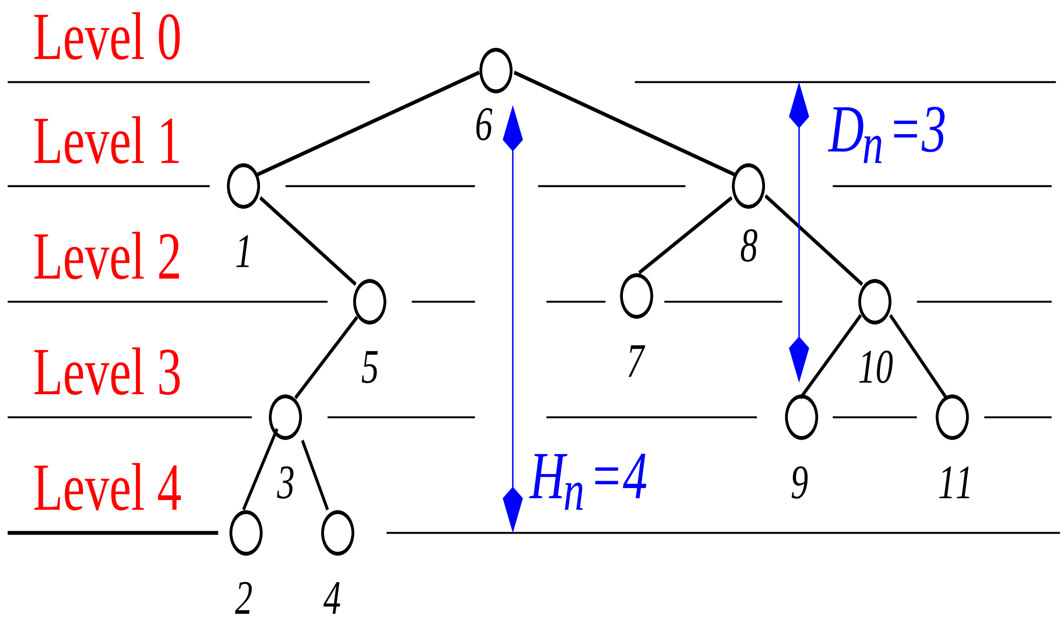
Liste der Daten: 6, 1, 8, 7, 5, 3, 10, 2, 11, 4, 9.



Größen im BSB



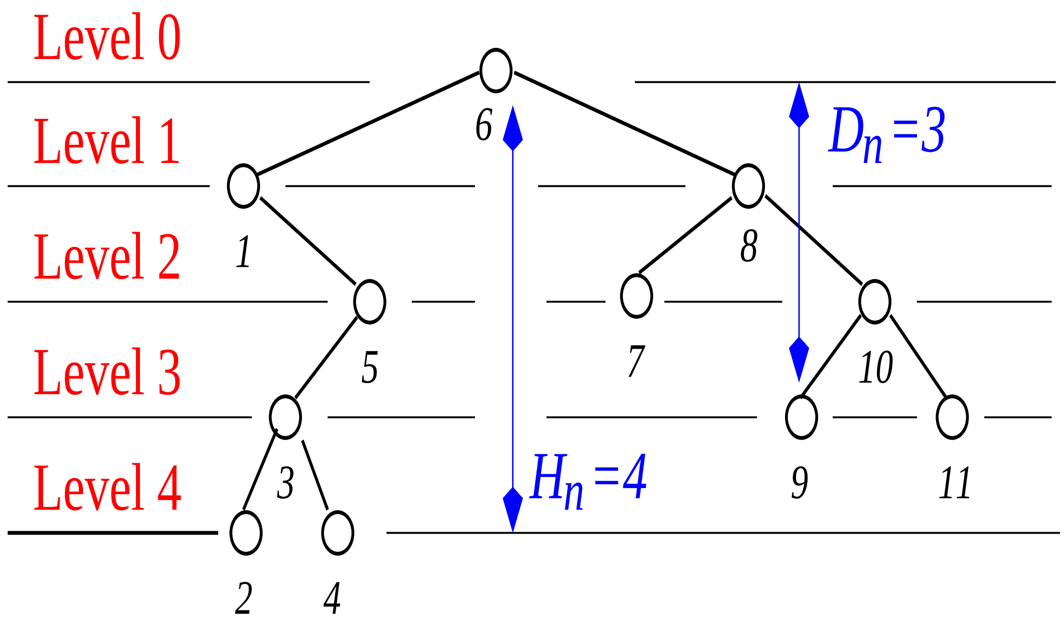
Größen im BSB



D_n — Tiefe des n-ten Knotens

= Abstand zwischen Wurzel und n-tem eingefügten Knoten

Größen im BSB



D_n — Tiefe des n-ten Knotens
= Abstand zwischen Wurzel und n-tem eingefügten Knoten

$H_n = \max_{1 \le j \le n} D_j$ — Höhe des Baumes

Permutationsmodell

Stochastisches Modell:

Alle Permutationen von $1, \ldots, n$ gleich wahrscheinlich.

Äquivalent: $U_1, \ldots, U_n$ i.i.d. unif$[0, 1]$.

Simulation

Binäre Bäume

Anzahl binärer Bäume mit n Knoten:

Binäre Bäume

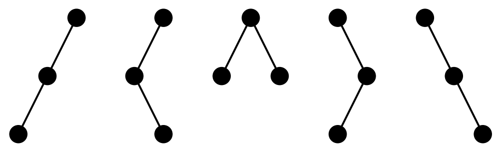
Anzahl binärer Bäume mit n Knoten:

Mit 3 Knoten:

Binäre Bäume

Anzahl binärer Bäume mit n Knoten:

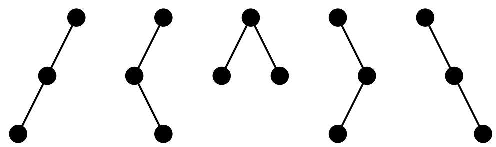
Mit 3 Knoten:



Binäre Bäume

Anzahl binärer Bäume mit n Knoten:

Mit 3 Knoten:

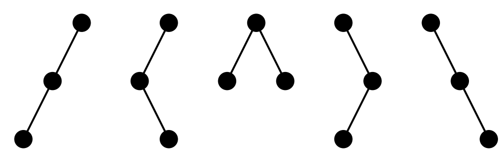


Mit 4 Knoten:

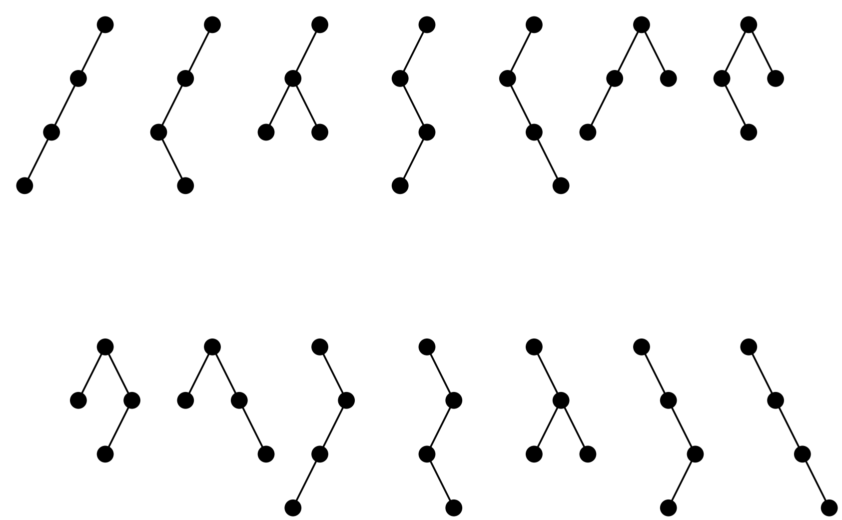
Binäre Bäume

Anzahl binärer Bäume mit n Knoten:

Mit 3 Knoten:



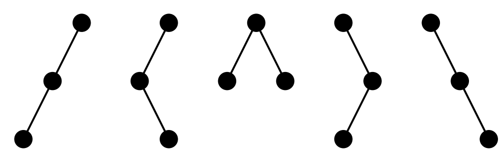
Mit 4 Knoten:



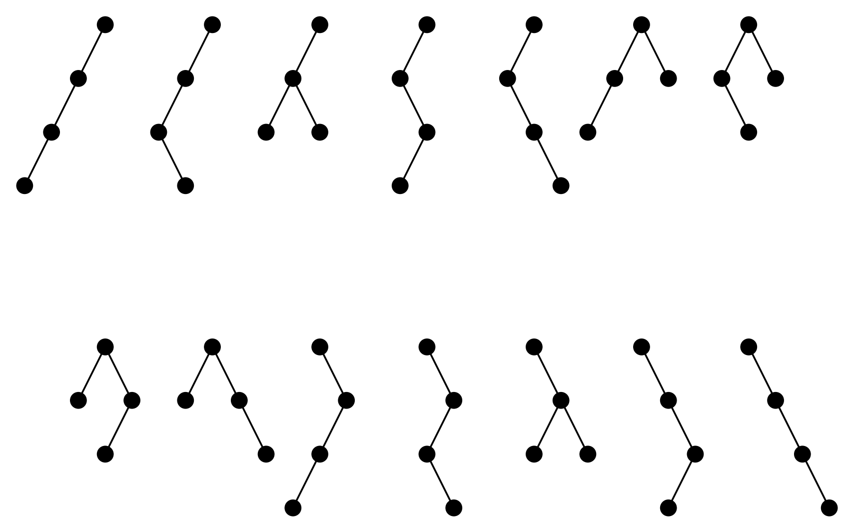
Binäre Bäume

Anzahl binärer Bäume mit n Knoten: $\frac{1}{n+1}\binom{2n}{n}$ (Catalan Zahlen)

Mit 3 Knoten:



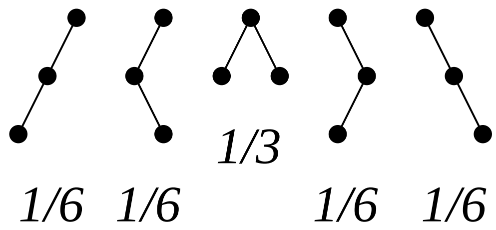
Mit 4 Knoten:



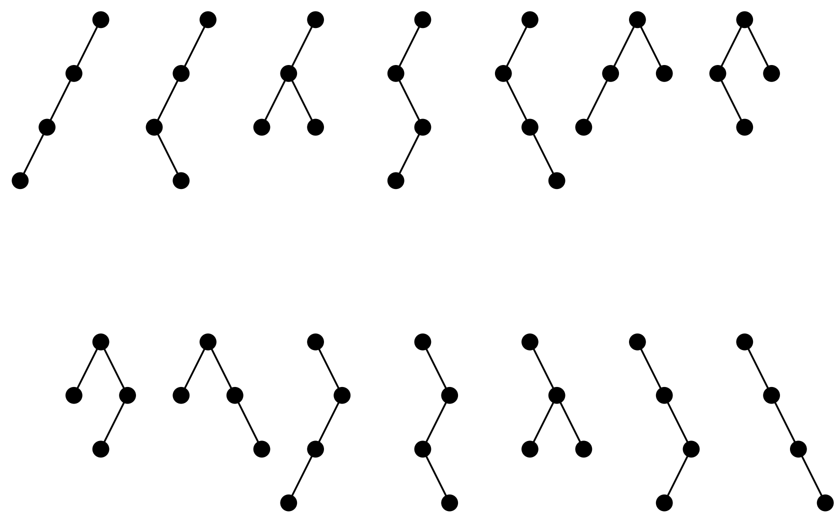
Binäre Bäume

Anzahl binärer Bäume mit n Knoten: $\frac{1}{n+1}\binom{2n}{n}$ (Catalan Zahlen)

Mit 3 Knoten:



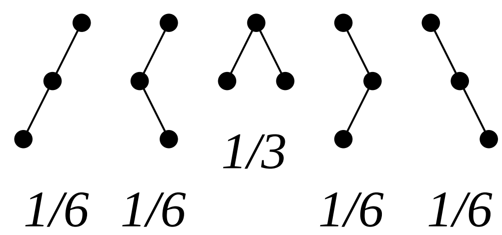
Mit 4 Knoten:



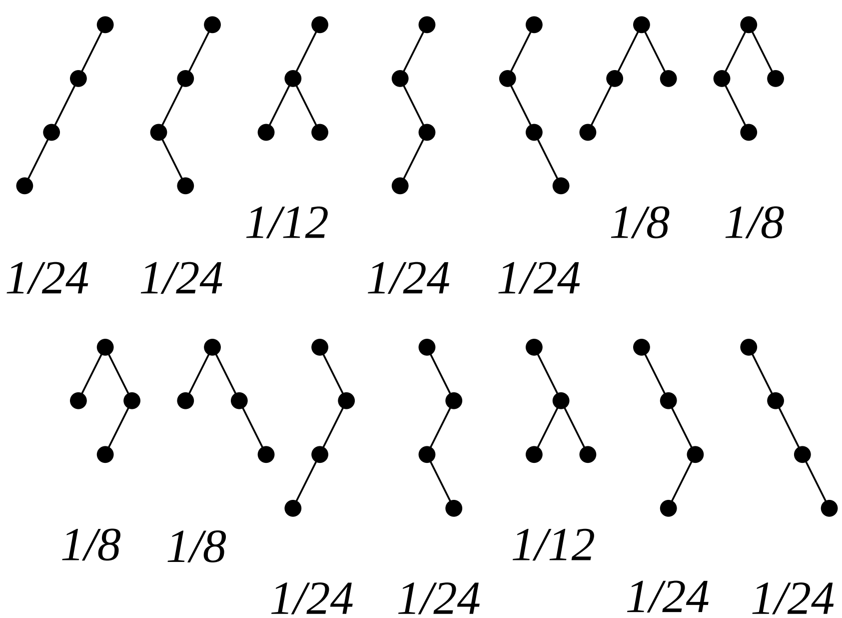
Binäre Bäume

Anzahl binärer Bäume mit n Knoten: $\frac{1}{n+1}\binom{2n}{n}$ (Catalan Zahlen)

Mit 3 Knoten:



Mit 4 Knoten:

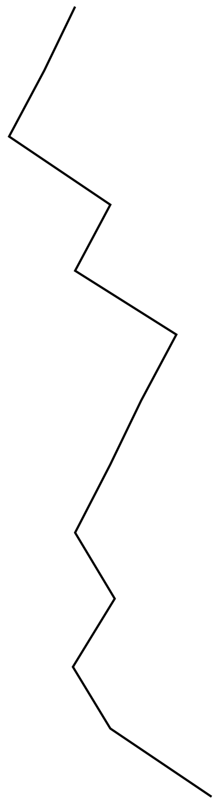


Catalan- versus Permutationsmodell

Catalan-Modell	Permutationsmodell

Catalan- versus Permutationsmodell

Catalan-Modell

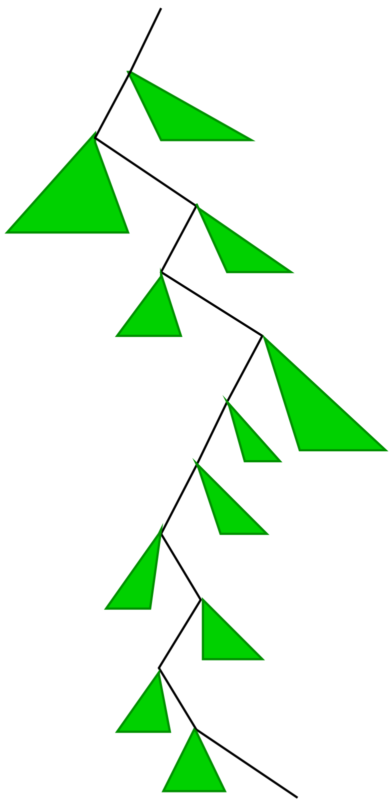


Permutationsmodell

Catalan- versus Permutationsmodell

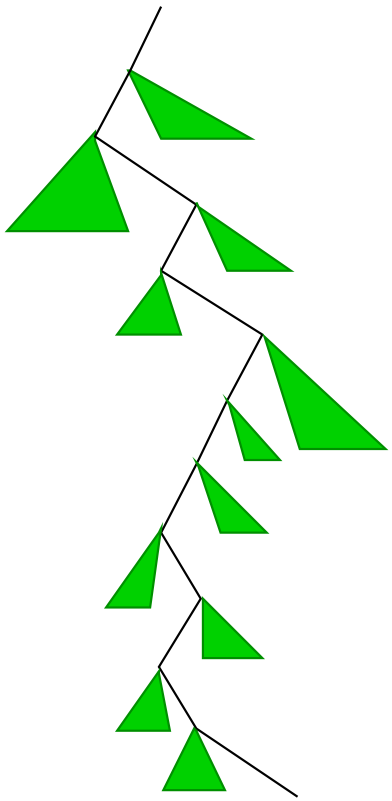
Catalan-Modell

Permutationsmodell

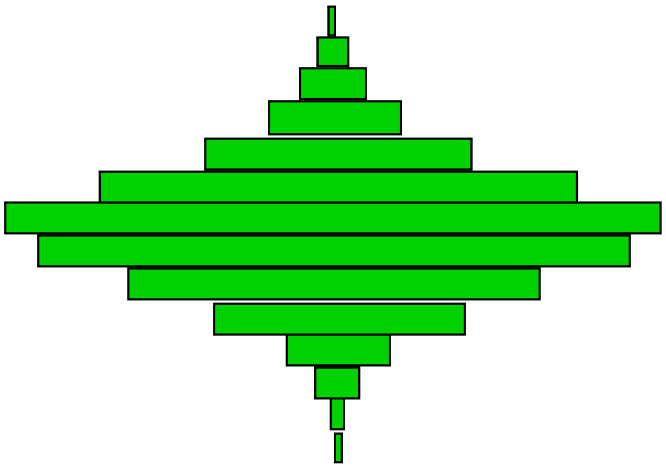


Catalan- versus Permutationsmodell

Catalan-Modell

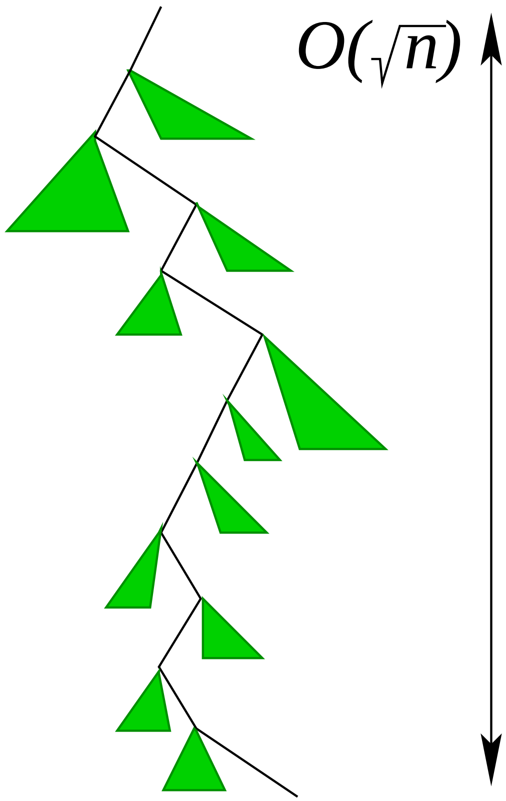


Permutationsmodell

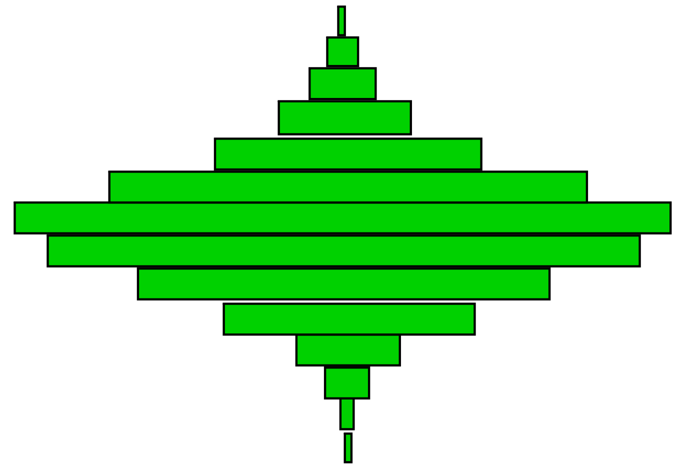


Catalan- versus Permutationsmodell

Catalan-Modell

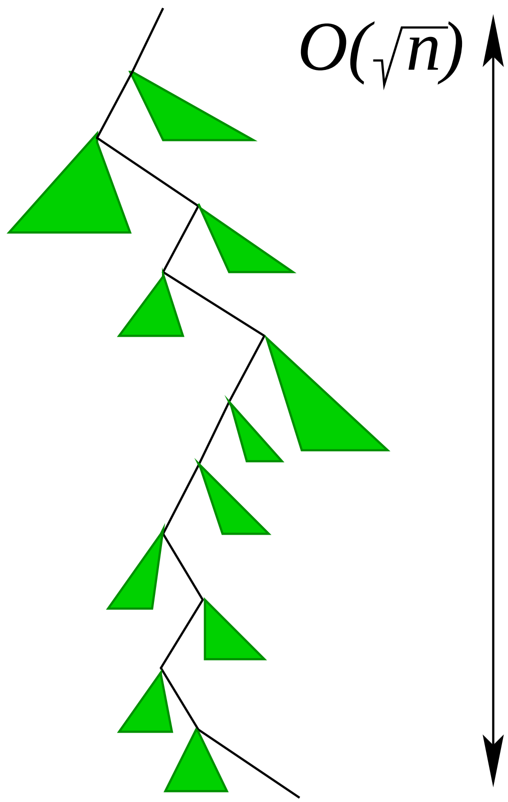


Permutationsmodell

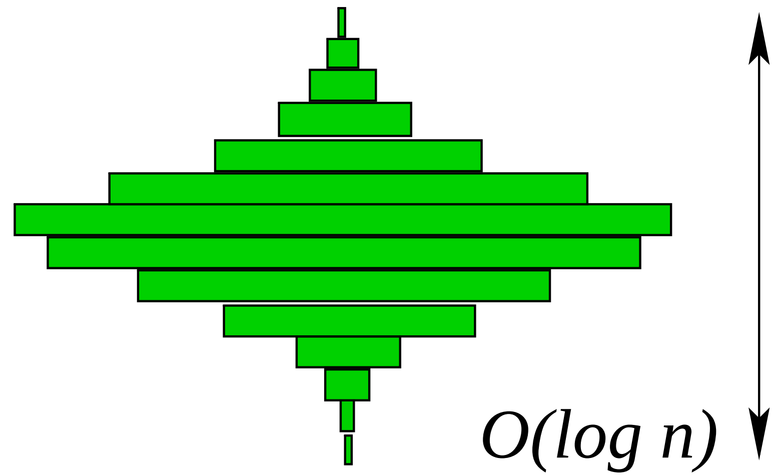


Catalan- versus Permutationsmodell

Catalan-Modell

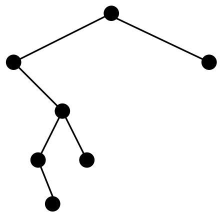


Permutationsmodell

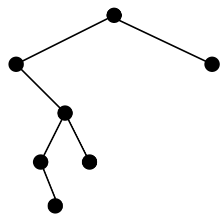


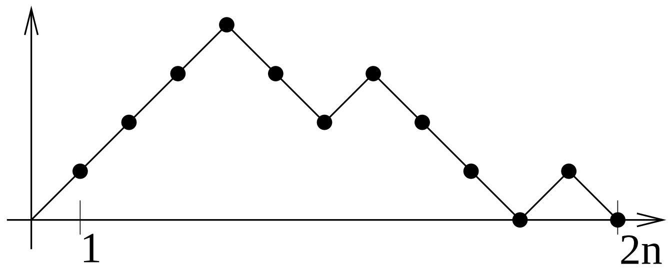
Gestalt im Catalan-Modell

Gestalt im Catalan-Modell

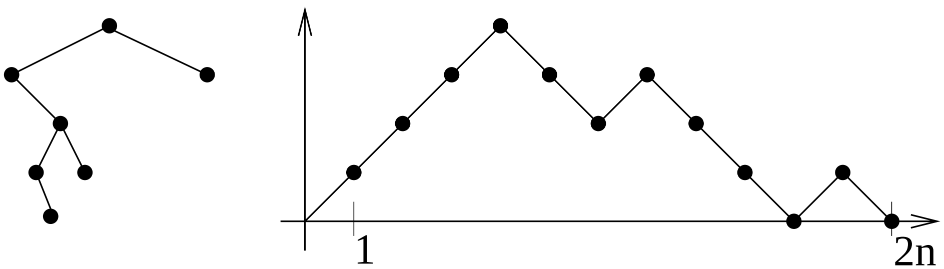


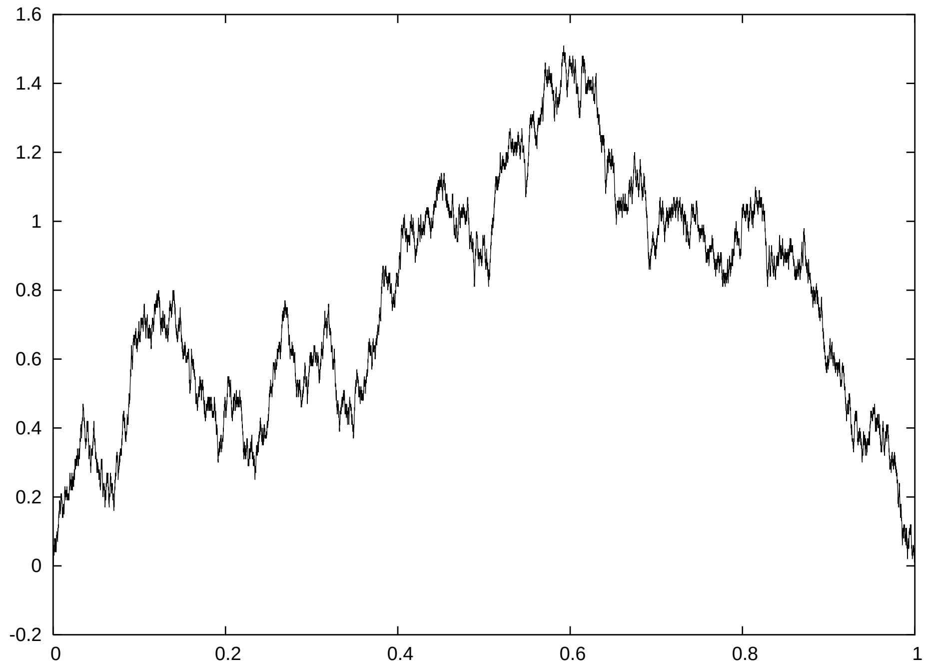
Gestalt im Catalan-Modell



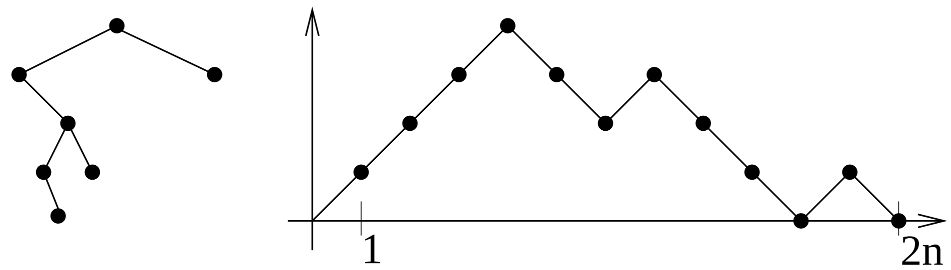


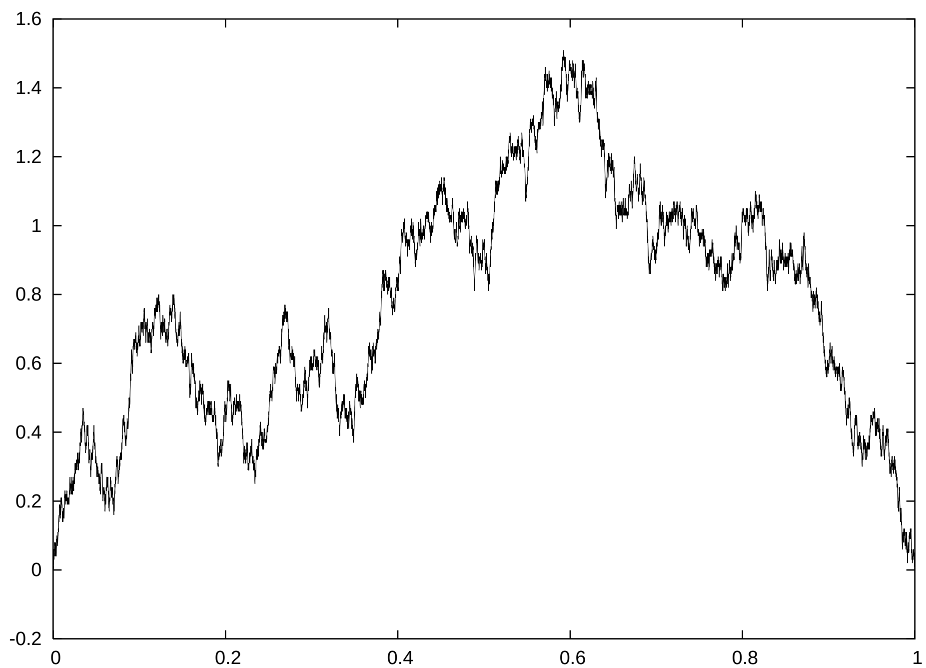
Gestalt im Catalan-Modell





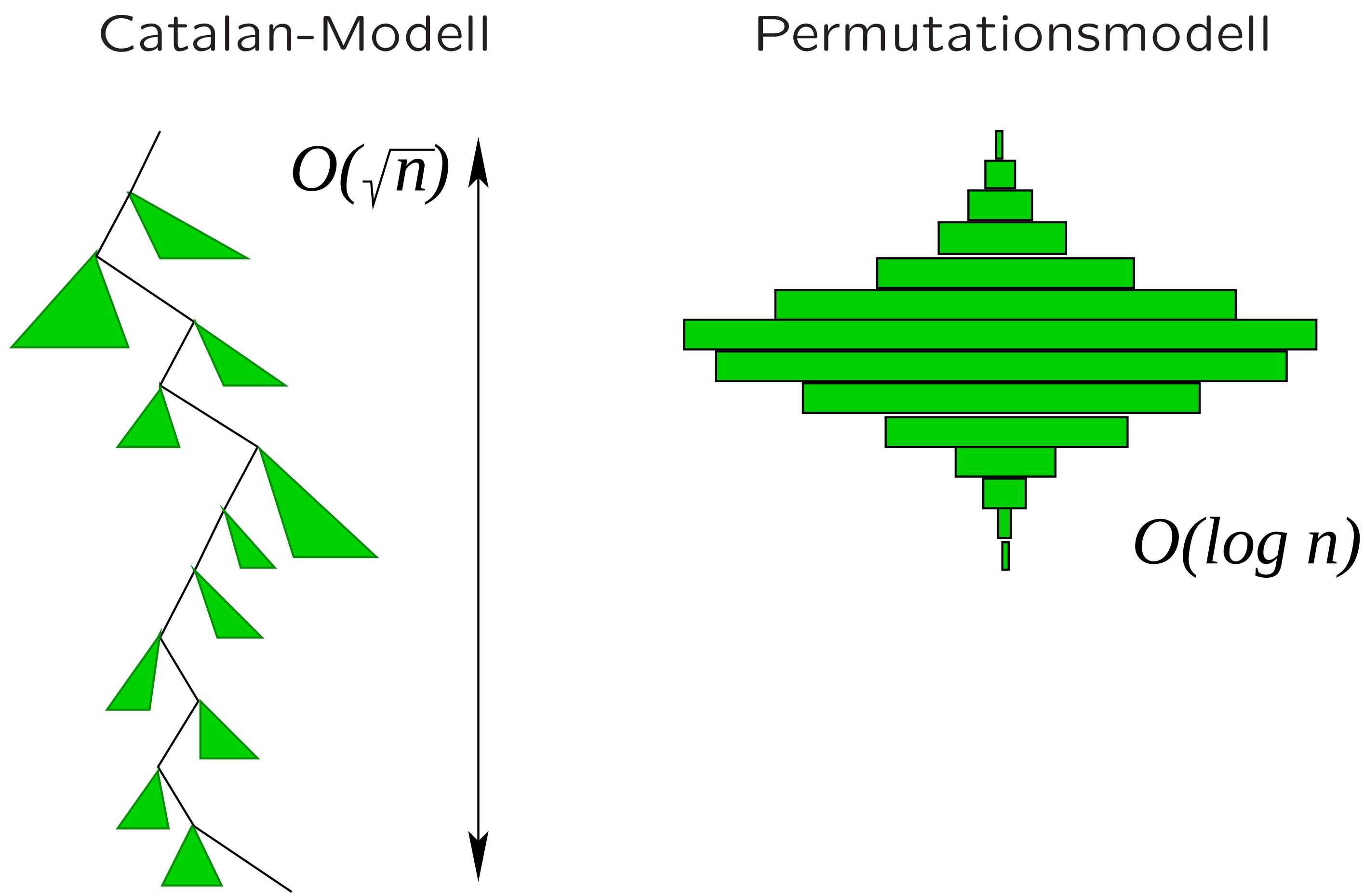
Gestalt im Catalan-Modell



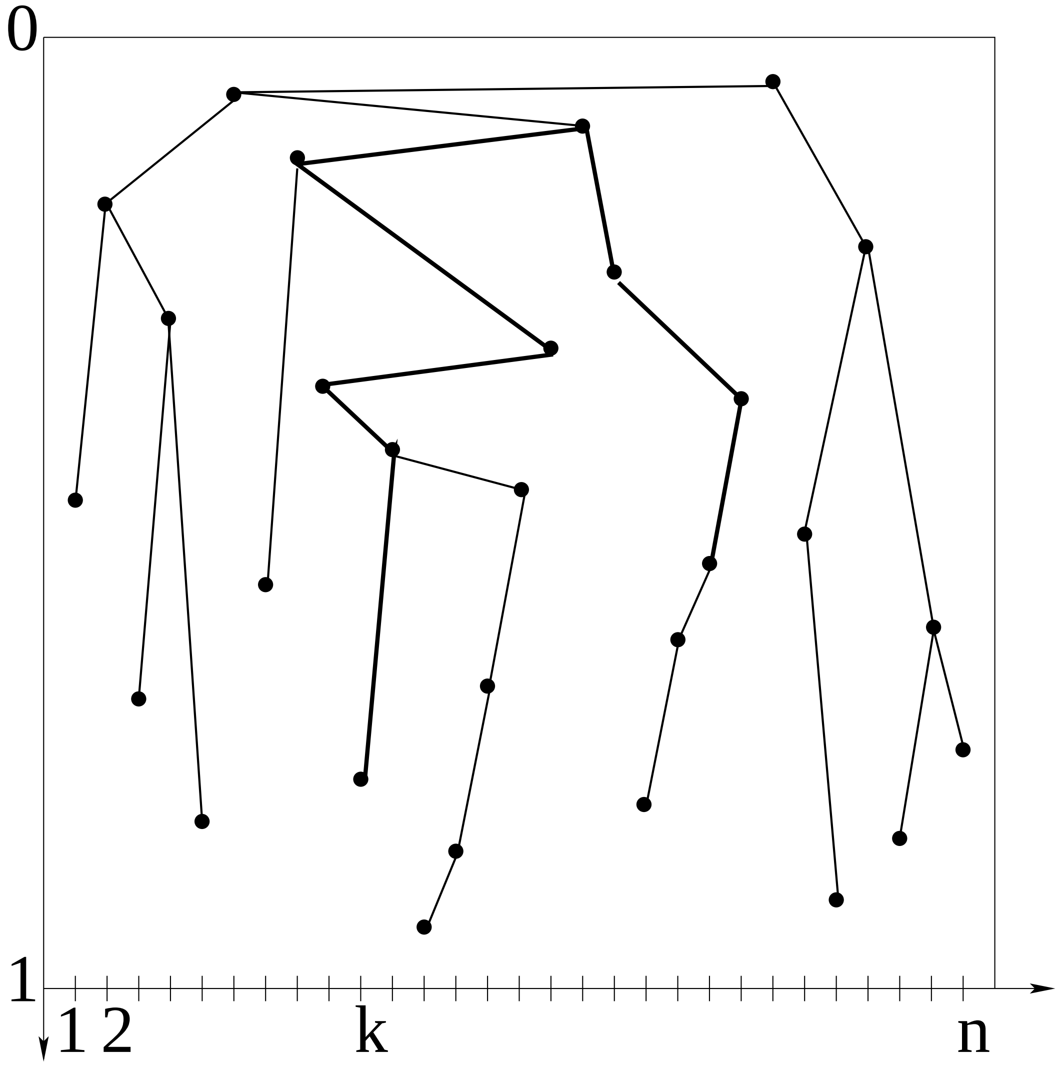


$$\left(\frac{H_n(2nt)}{\sqrt{2n}}\right)_{t\in[0,1]} \xrightarrow{d} (e(t))_{t\in[0,1]}$$ e : Brownsche Exkursion

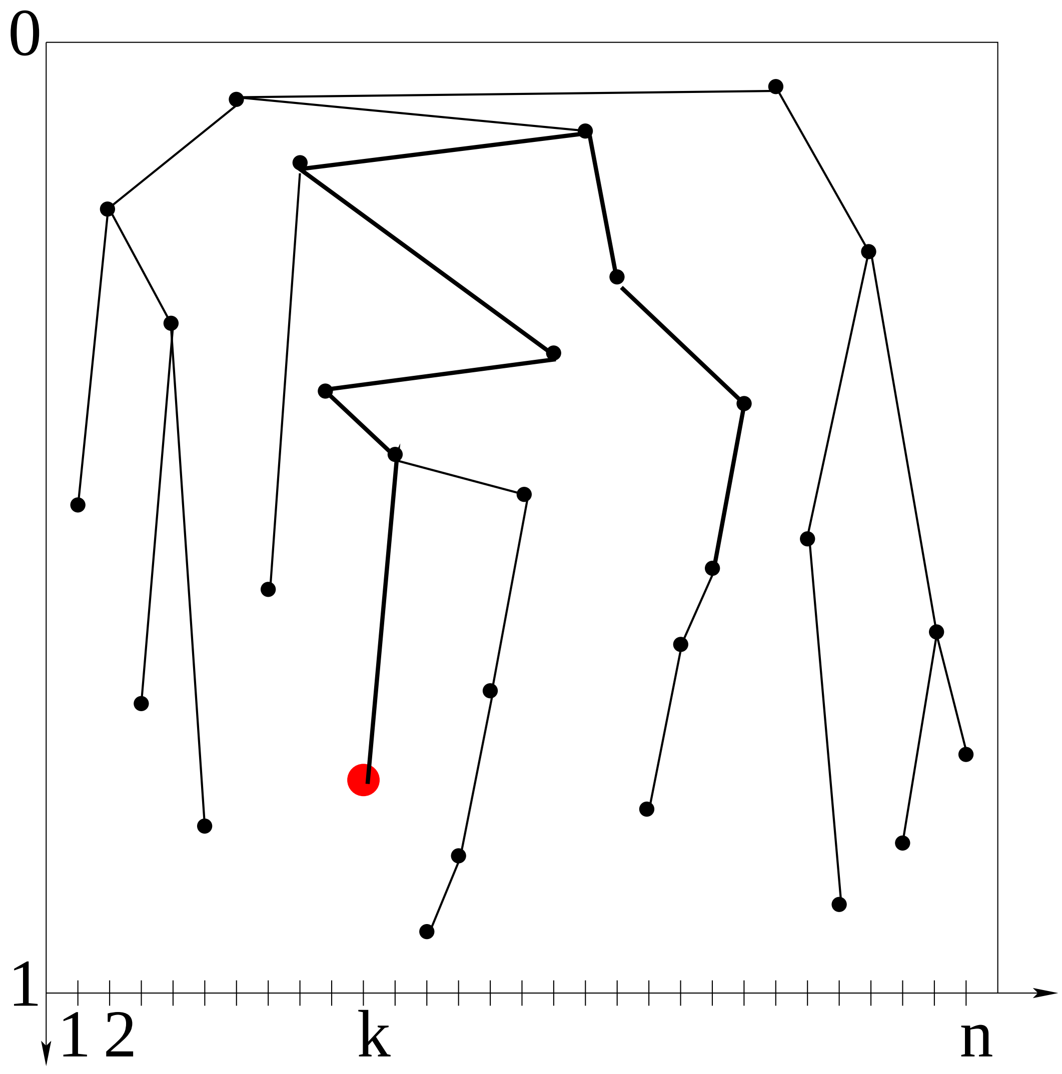
Catalan- versus Permutationsmodell



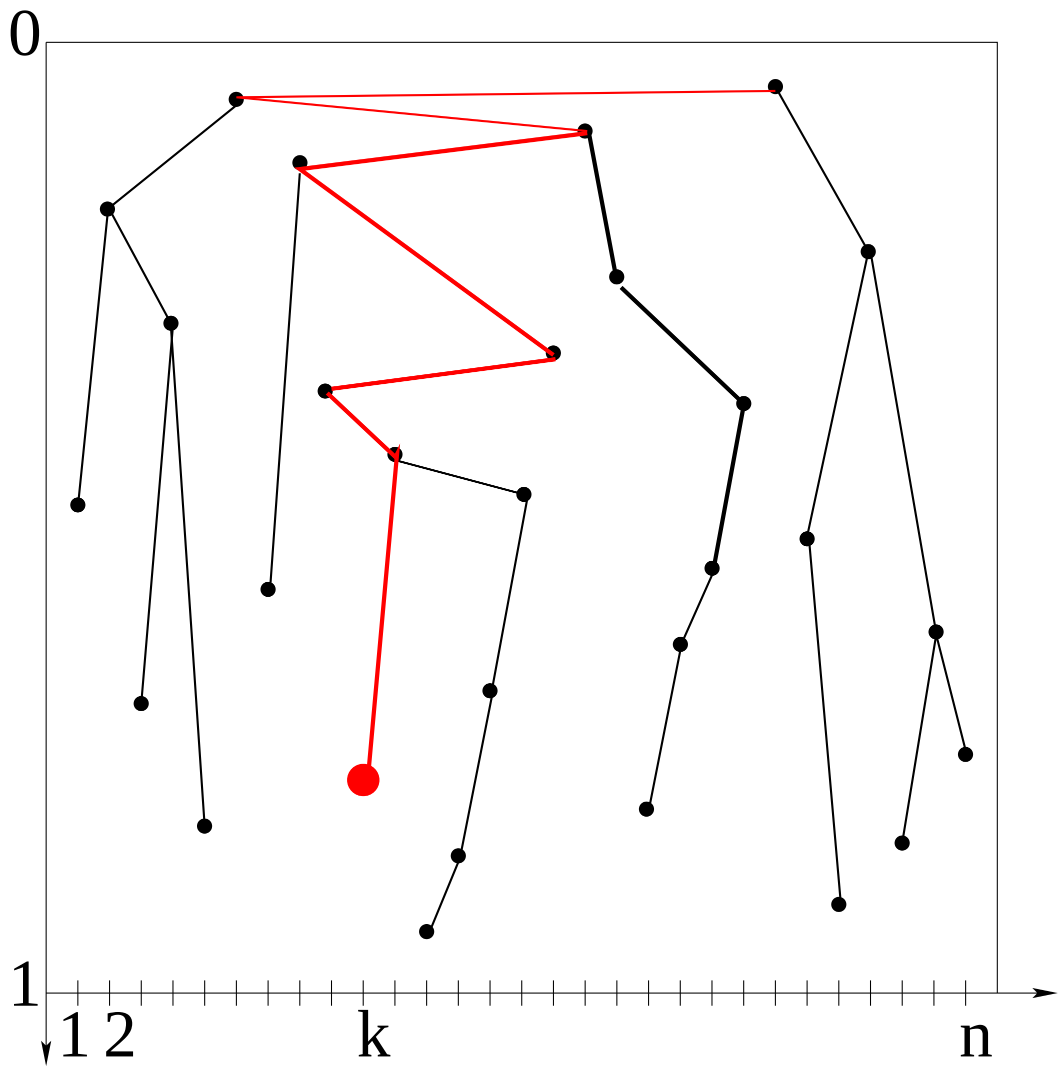
Stochastische Analyse der Tiefe D_n



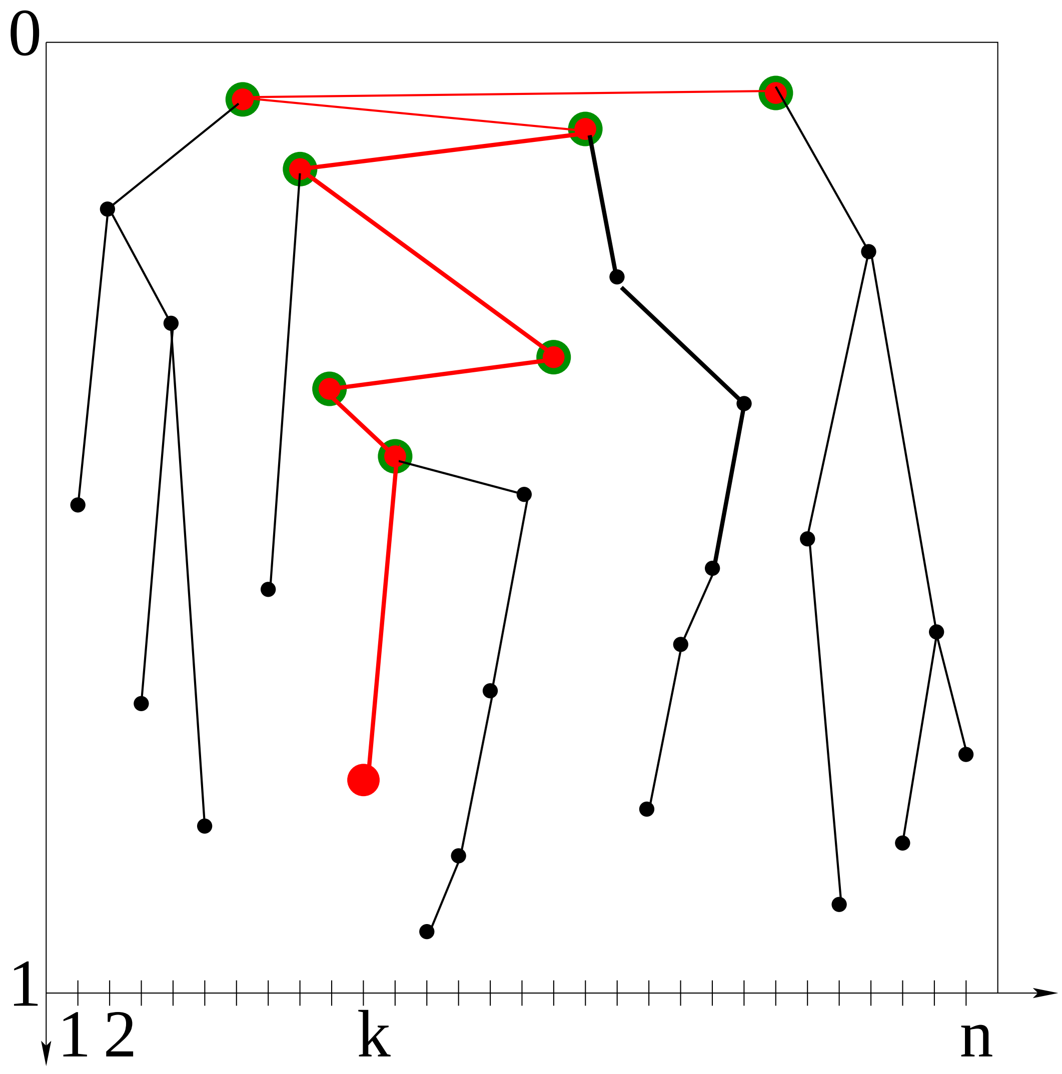
Stochastische Analyse der Tiefe D_n



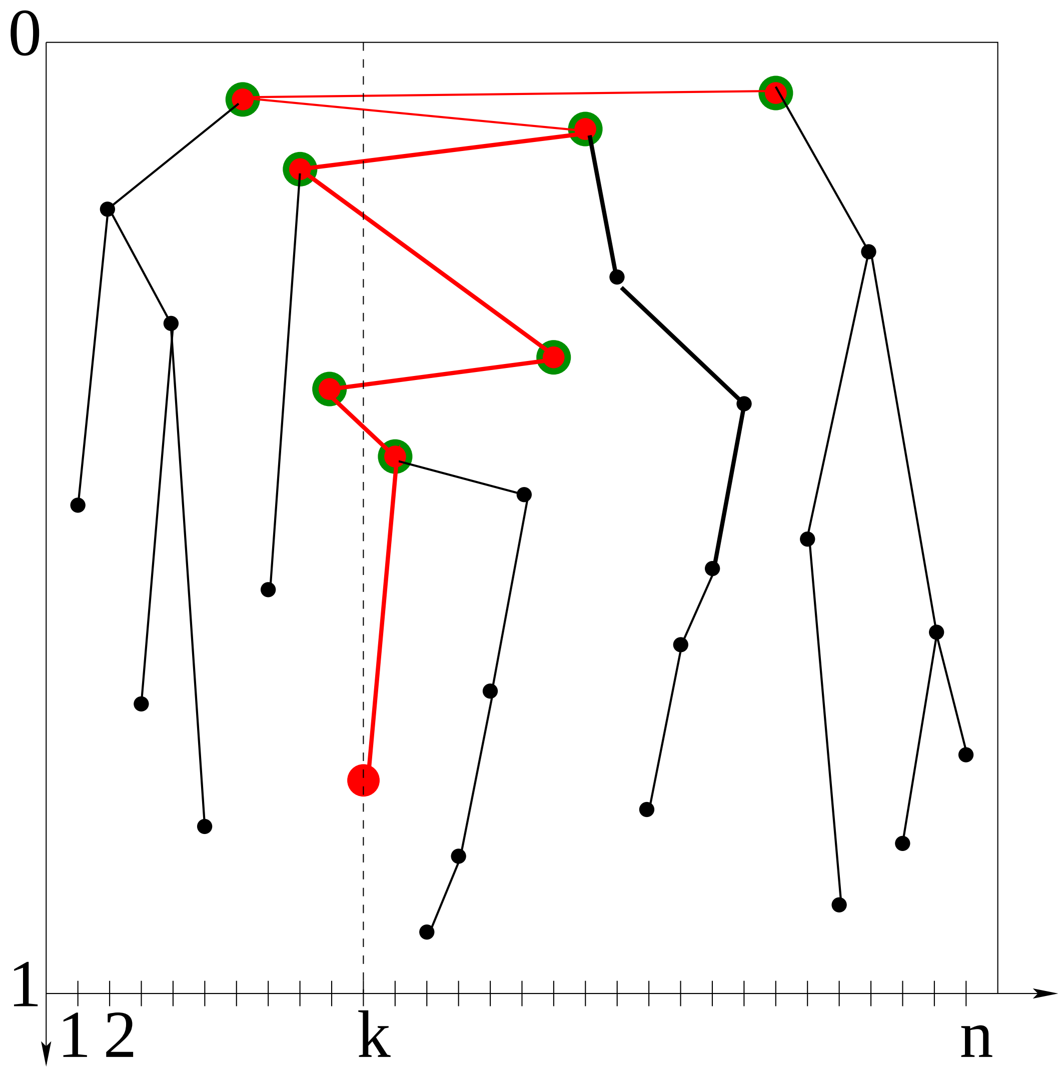
Stochastische Analyse der Tiefe D_n



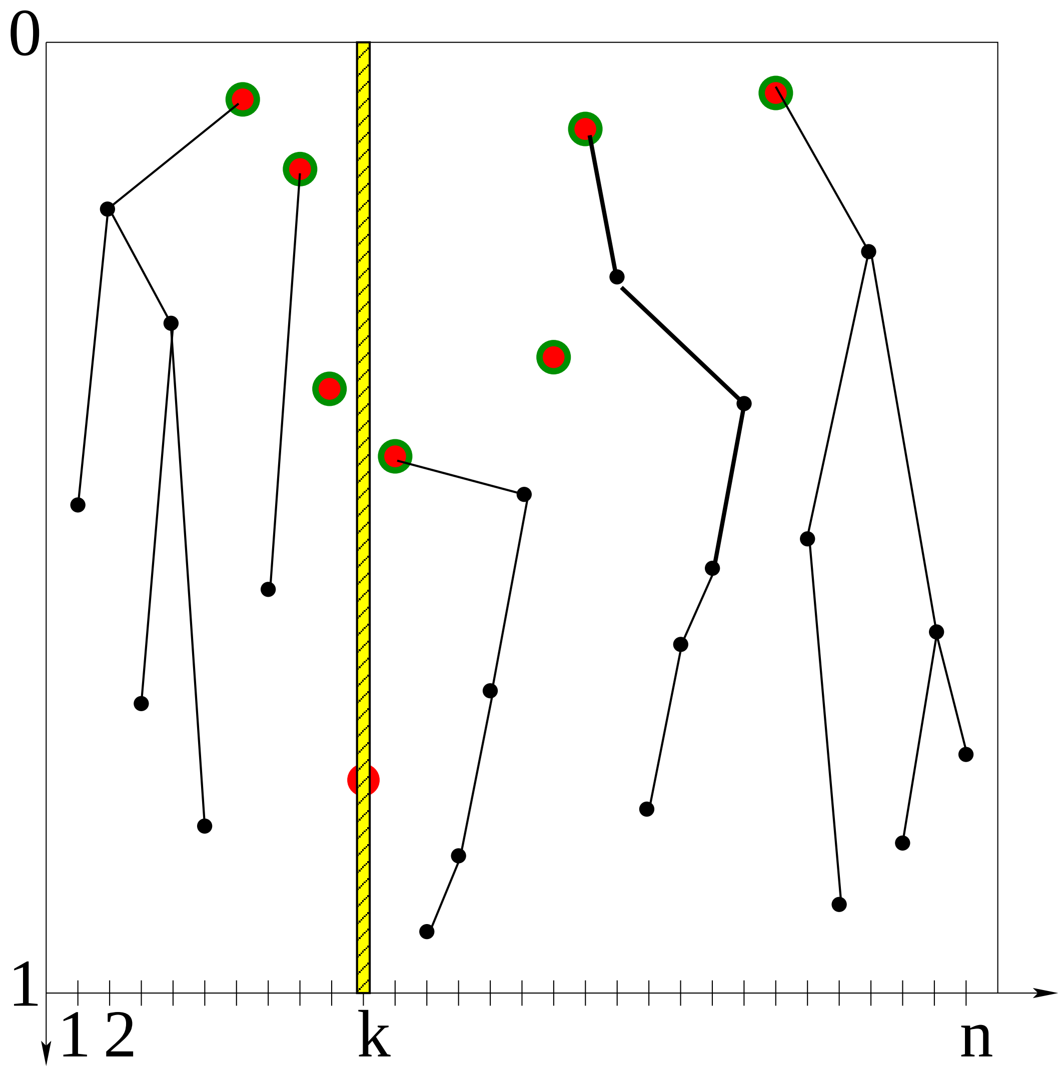
Stochastische Analyse der Tiefe D_n



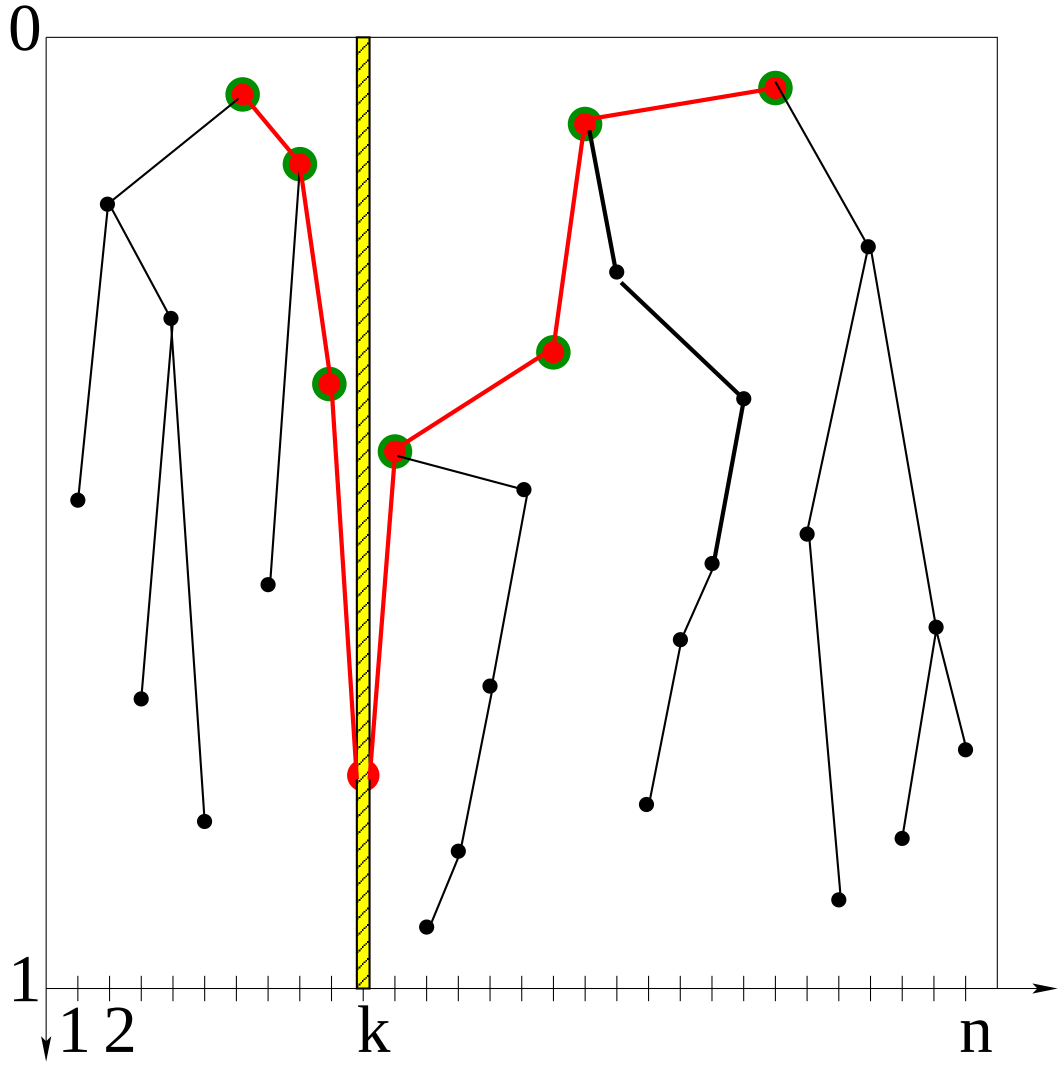
Stochastische Analyse der Tiefe D_n



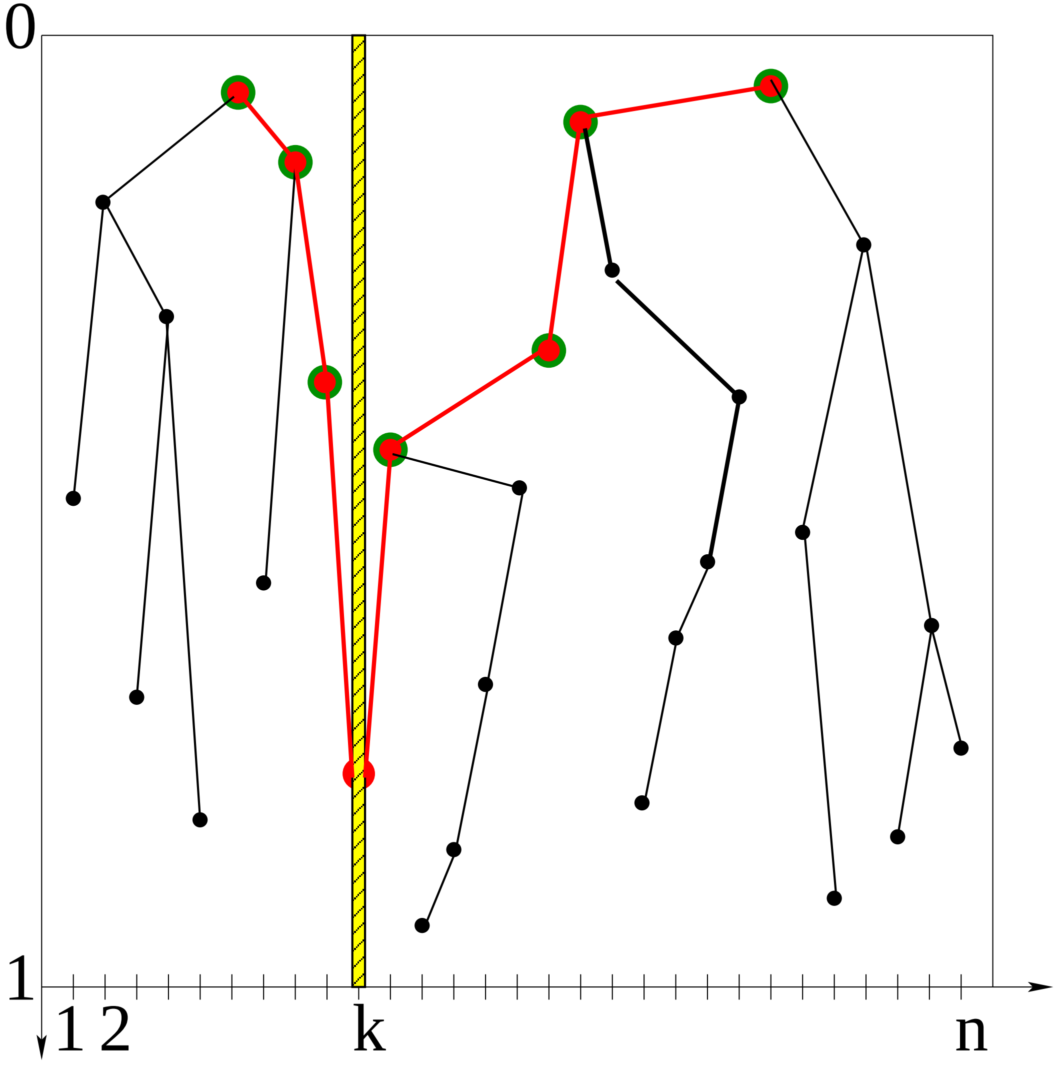
Stochastische Analyse der Tiefe D_n



Stochastische Analyse der Tiefe D_n

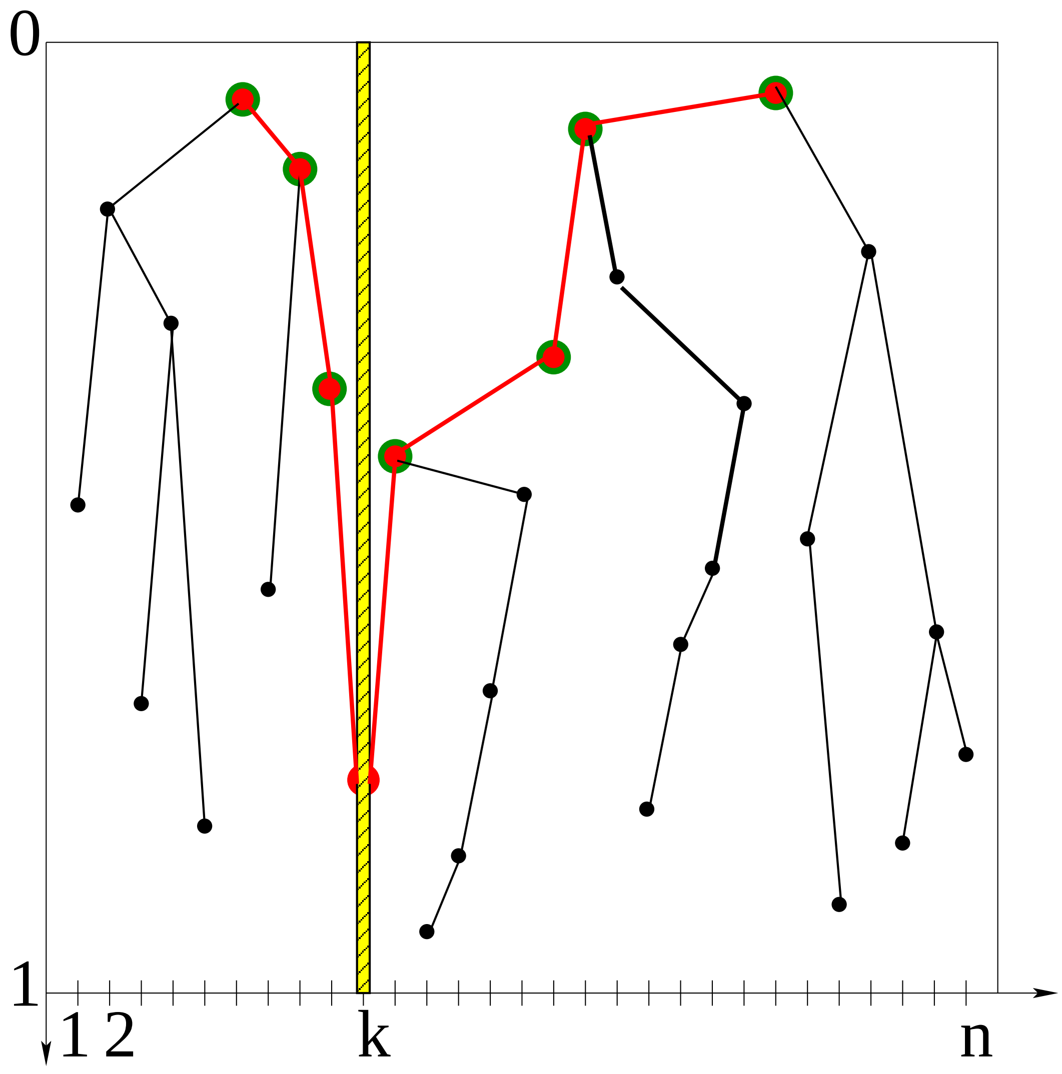


Stochastische Analyse der Tiefe D_n



Auf- bzw. Ab-Rekorde
sind unabhängig!

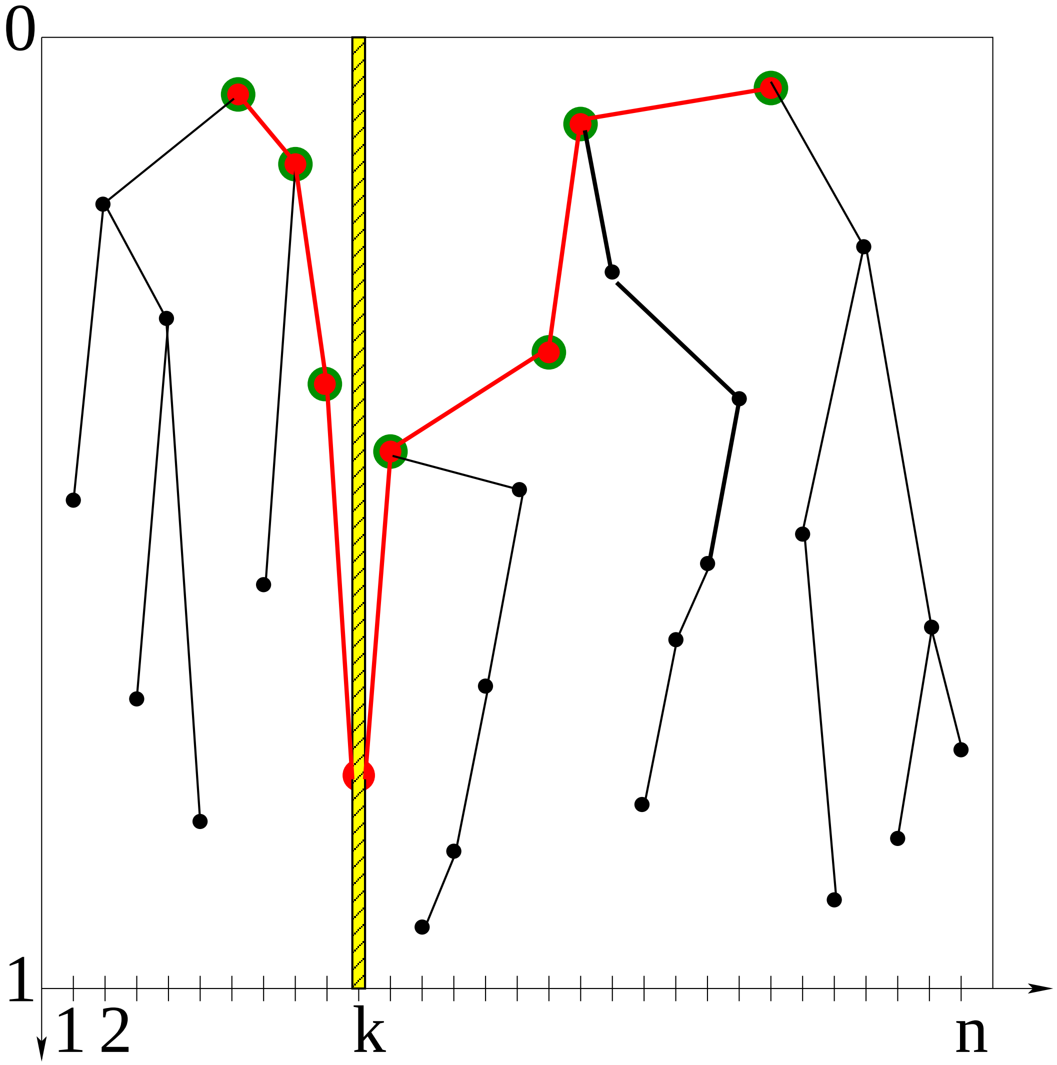
Stochastische Analyse der Tiefe D_n



Auf- bzw. Ab-Rekorde
sind unabhängig!

$$D_n \stackrel{d}{=} \sum_{j=2}^{n} Z_j$$

Stochastische Analyse der Tiefe D_n

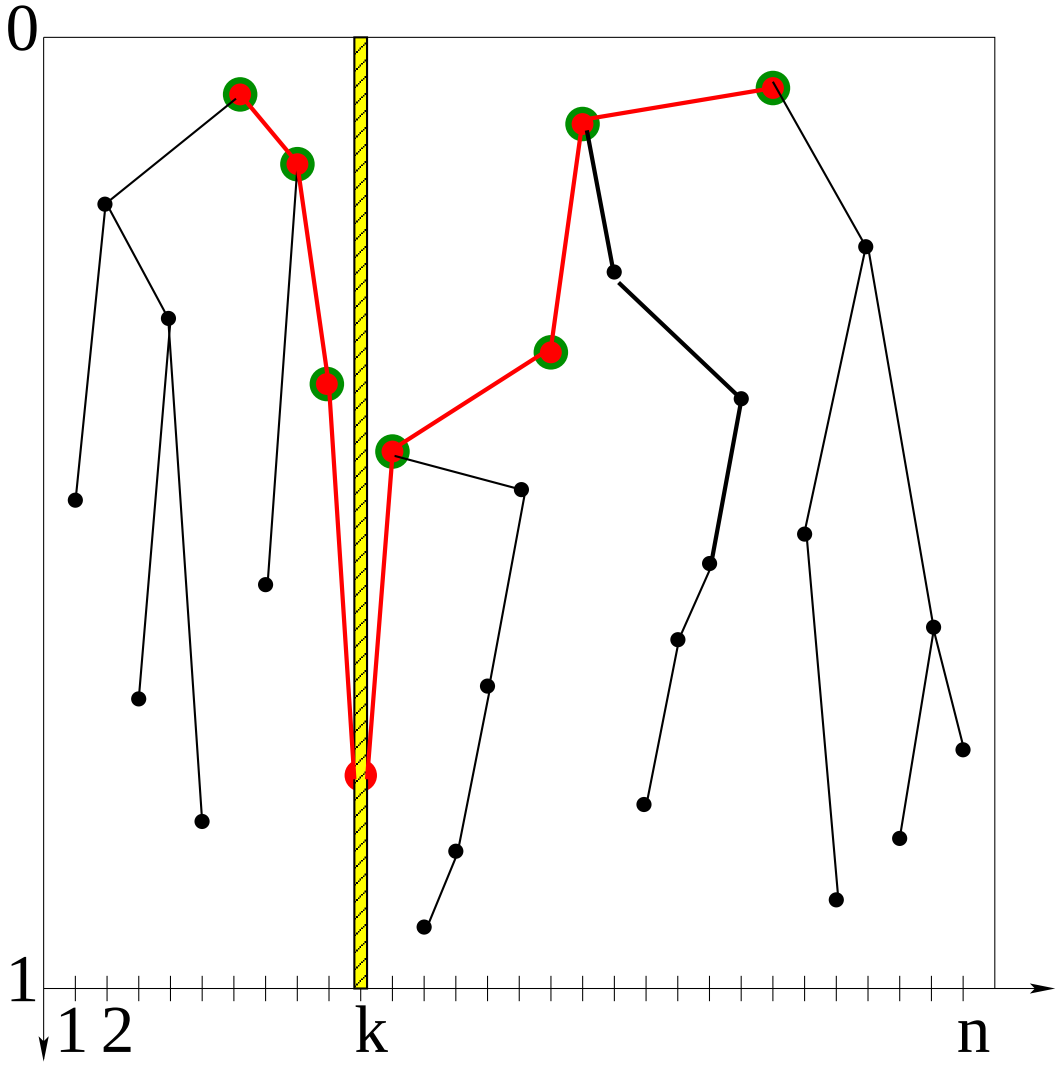


Auf- bzw. Ab-Rekorde sind unabhängig!

$$D_n \stackrel{d}{=} \sum_{j=2}^{n} Z_j$$

Z_j unabhängig,

Stochastische Analyse der Tiefe D_n



Auf- bzw. Ab-Rekorde sind unabhängig!

$$D_n \stackrel{d}{=} \sum_{j=2}^{n} Z_j$$

Z_j unabhängig,

$Z_j \stackrel{d}{=} \mathrm{Ber}(2/j)$.

Asymptotisches Verhalten der Tiefe

$$D_n \stackrel{d}{=} \sum_{j=2}^{n} \mathrm{Ber}(2/j)$$

Asymptotisches Verhalten der Tiefe

$$D_n \stackrel{d}{=} \sum_{j=2}^{n} \mathrm{Ber}(2/j)$$

Asymptotiken:

Asymptotisches Verhalten der Tiefe

$$D_n \stackrel{d}{=} \sum_{j=2}^{n} \mathrm{Ber}(2/j)$$

Asymptotiken:

$$\mathbb{E}\, D_n = 2 \log n + O(1),$$

Asymptotisches Verhalten der Tiefe

$$D_n \stackrel{d}{=} \sum_{j=2}^{n} \mathrm{Ber}(2/j)$$

Asymptotiken:

$$\mathbb{E}\, D_n = 2 \log n + O(1), \quad \mathrm{Var}(D_n) = 2 \log n + O(1)$$

Asymptotisches Verhalten der Tiefe

$$D_n \stackrel{d}{=} \sum_{j=2}^{n} \mathrm{Ber}(2/j)$$

Asymptotiken:

$$\mathbb{E}\, D_n = 2\log n + O(1), \qquad \mathrm{Var}(D_n) = 2\log n + O(1)$$

$$\varrho\left(\frac{D_n - 2\log n}{\sqrt{2\log n}}, \mathcal{N}(0,1)\right)$$

Asymptotisches Verhalten der Tiefe

$$D_n \stackrel{d}{=} \sum_{j=2}^{n} \mathrm{Ber}(2/j)$$

Asymptotiken:

$$\mathbb{E}\, D_n = 2\log n + O(1), \qquad \mathrm{Var}(D_n) = 2\log n + O(1)$$

$$\varrho\left(\frac{D_n - 2\log n}{\sqrt{2\log n}}, \mathcal{N}(0,1)\right) = O\left(\frac{1}{\sqrt{\log n}}\right)$$

Asymptotisches Verhalten der Tiefe

$$D_n \overset{d}{=} \sum_{j=2}^{n} \mathrm{Ber}(2/j)$$

Asymptotiken:

$$\mathbb{E}\, D_n = 2\log n + O(1), \qquad \mathrm{Var}(D_n) = 2\log n + O(1)$$

$$\varrho\left(\frac{D_n - 2\log n}{\sqrt{2\log n}}, \mathcal{N}(0,1)\right) = O\left(\frac{1}{\sqrt{\log n}}\right)$$

$$d_{\mathrm{TV}}(D_n, \Pi(\mathbb{E}\, D_n))$$

Asymptotisches Verhalten der Tiefe

$$D_n \stackrel{d}{=} \sum_{j=2}^{n} \mathrm{Ber}(2/j)$$

Asymptotiken:

$$\mathbb{E}\, D_n = 2\log n + O(1), \quad \mathrm{Var}(D_n) = 2\log n + O(1)$$

$$\varrho\left(\frac{D_n - 2\log n}{\sqrt{2\log n}}, \mathcal{N}(0,1)\right) = O\left(\frac{1}{\sqrt{\log n}}\right)$$

$$d_{\mathrm{TV}}(D_n, \Pi(\mathbb{E}\, D_n)) = O\left(\frac{1}{\log n}\right)$$

Asymptotisches Verhalten der Tiefe

$$D_n \stackrel{d}{=} \sum_{j=2}^{n} \mathrm{Ber}(2/j)$$

Asymptotiken:

$$\mathbb{E}\, D_n = 2\log n + O(1), \qquad \mathrm{Var}(D_n) = 2\log n + O(1)$$

$$\varrho\left(\frac{D_n - 2\log n}{\sqrt{2\log n}}, \mathcal{N}(0,1)\right) = O\left(\frac{1}{\sqrt{\log n}}\right)$$

$$d_{\mathrm{TV}}(D_n, \Pi(\mathbb{E}\, D_n)) = O\left(\frac{1}{\log n}\right)$$

$$\mathbb{P}\Big(|D_n - \mathbb{E}\, D_n| > \varepsilon \mathbb{E}\, D_n\Big) \le C n^{-\frac{\varepsilon^2}{2+\varepsilon}}$$

Pattern matching

Problem: Suche kurze Textabschnitte in großen Texten

Pattern matching

Problem: Suche kurze Textabschnitte in großen Texten

- Internet: Google

Pattern matching

Problem: Suche kurze Textabschnitte in großen Texten

- Internet: Google
- DNA

Pattern matching

Problem: Suche kurze Textabschnitte in großen Texten

- Internet: Google
- DNA

Großer Text: $B_1, \ldots, B_n$ (Folge von Bits)

Pattern matching

Problem: Suche kurze Textabschnitte in großen Texten

- Internet: Google
- DNA

Großer Text: $B_1, \ldots, B_n$ (Folge von Bits)

$$X_1 \;:=\; 0.B_1B_2B_3B_4B_5\ldots$$

Pattern matching

Problem: Suche kurze Textabschnitte in großen Texten

- Internet: Google
- DNA

Großer Text: $B_1, \ldots, B_n$ (Folge von Bits)

$$X_1 \;:=\; 0.B_1 B_2 B_3 B_4 B_5 \ldots$$

Pattern matching

Problem: Suche kurze Textabschnitte in großen Texten

- Internet: Google
- DNA

Großer Text: $B_1, \ldots, B_n$ (Folge von Bits)

$$X_1 := 0.B_1 B_2 B_3 B_4 B_5 \ldots$$

$$X_2 := 0.B_2 B_3 B_4 B_5 B_6 \ldots$$

Pattern matching

Problem: Suche kurze Textabschnitte in großen Texten

- Internet: Google
- DNA

Großer Text: $B_1, \ldots, B_n$ (Folge von Bits)

$$
\begin{aligned}
X_1 &:= 0.B_1 B_2 B_3 B_4 B_5 \ldots \\
X_2 &:= 0.B_2 B_3 B_4 B_5 B_6 \ldots \\
X_3 &:= 0.B_3 B_4 B_5 B_6 B_7 \ldots
\end{aligned}
$$

Pattern matching

Problem: Suche kurze Textabschnitte in großen Texten

- Internet: Google
- DNA

Großer Text: $B_1, \ldots, B_n$ (Folge von Bits)

$$\begin{aligned} X_1 &:= 0.B_1 B_2 B_3 B_4 B_5 \ldots \\ X_2 &:= 0.B_2 B_3 B_4 B_5 B_6 \ldots \\ X_3 &:= 0.B_3 B_4 B_5 B_6 B_7 \ldots \\ X_4 &:= 0.B_4 B_5 B_6 B_7 B_8 \ldots \end{aligned}$$

Pattern matching

Problem: Suche kurze Textabschnitte in großen Texten

- Internet: Google
- DNA

Großer Text: $B_1, \ldots, B_n$ (Folge von Bits)

$$
\begin{aligned}
X_1 &:= 0.B_1 B_2 B_3 B_4 B_5 \ldots \\
X_2 &:= 0.B_2 B_3 B_4 B_5 B_6 \ldots \\
X_3 &:= 0.B_3 B_4 B_5 B_6 B_7 \ldots \\
X_4 &:= 0.B_4 B_5 B_6 B_7 B_8 \ldots \\
X_5 &:= 0.B_5 B_6 B_7 B_8 B_9 \ldots \\
&\vdots
\end{aligned}
$$

Suffixsuchbaum

Binärsuchbaum für Folge der Suffixe $X_1, X_2, \ldots$

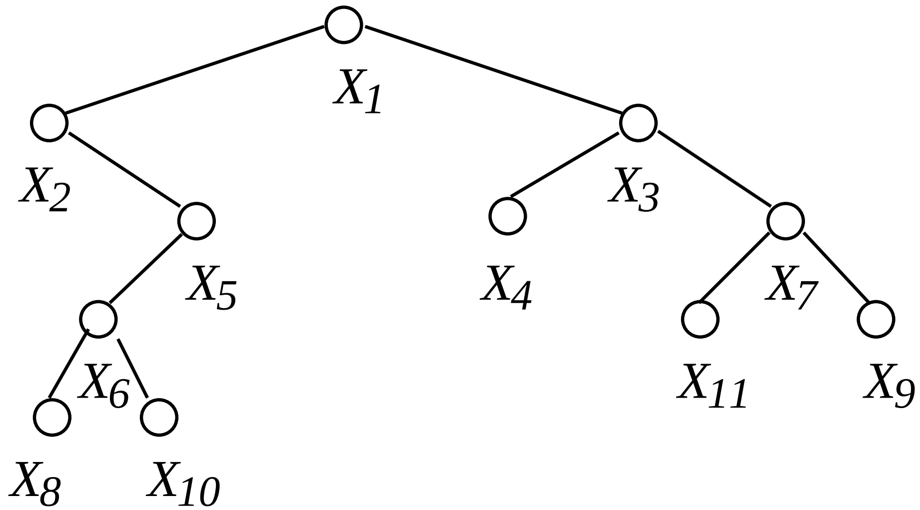
Suffixsuchbaum

Binärsuchbaum für Folge der Suffixe $X_1, X_2, \ldots$

$0.B_1\, B_2\, B_3\, B_4\, B_5\, B_6\, B_7\, B_8\, B_9\, B_{10}\, B_{11} \ldots$

Suffixsuchbaum

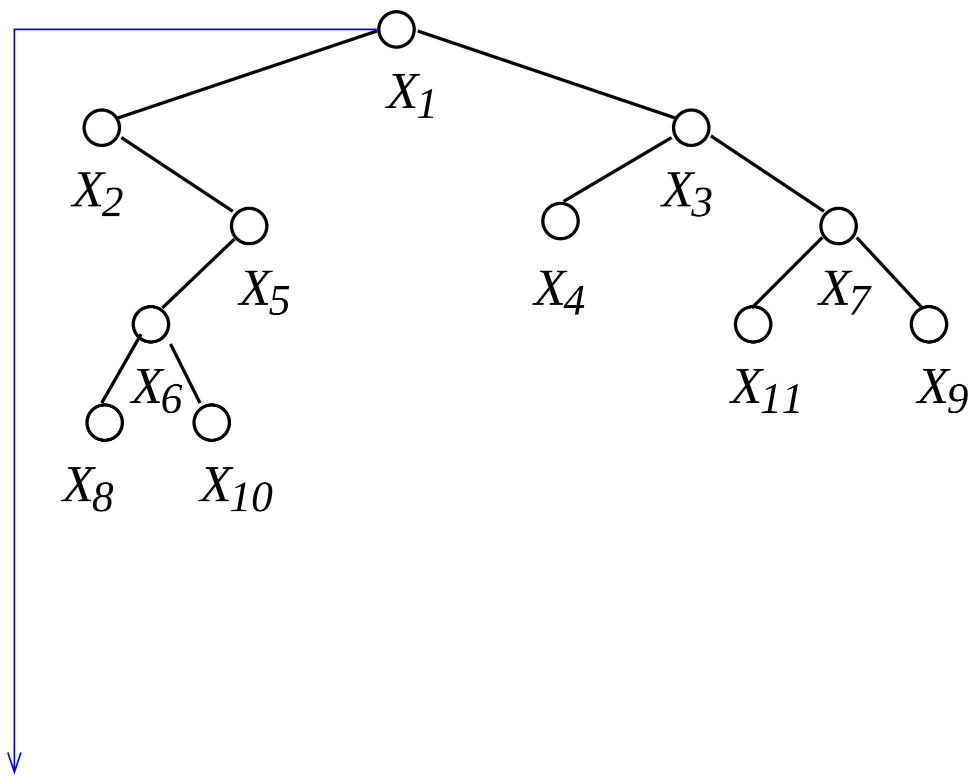
Binärsuchbaum für Folge der Suffixe $X_1, X_2, \ldots$



$0.B_1 B_2 B_3 B_4 B_5 B_6 B_7 B_8 B_9 B_{10} B_{11} \ldots$

Suffixsuchbaum

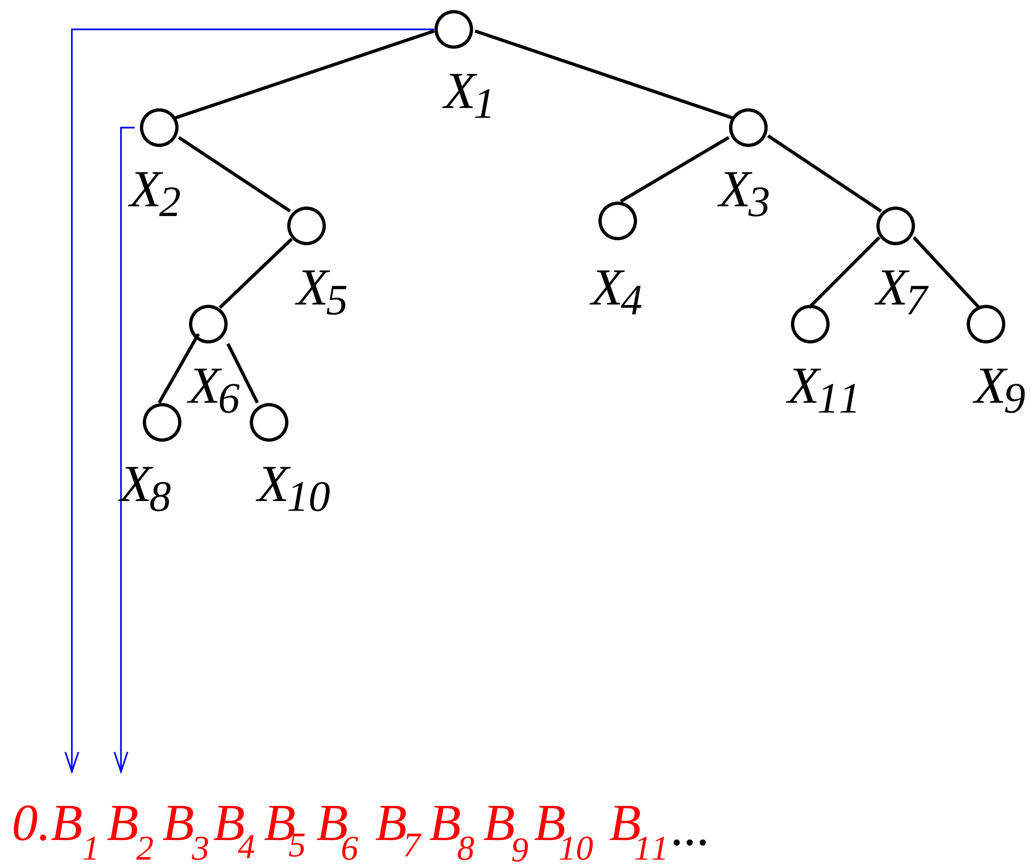
Binärsuchbaum für Folge der Suffixe $X_1, X_2, \ldots$



$0.B_1 B_2 B_3 B_4 B_5 B_6 B_7 B_8 B_9 B_{10} B_{11} \ldots$

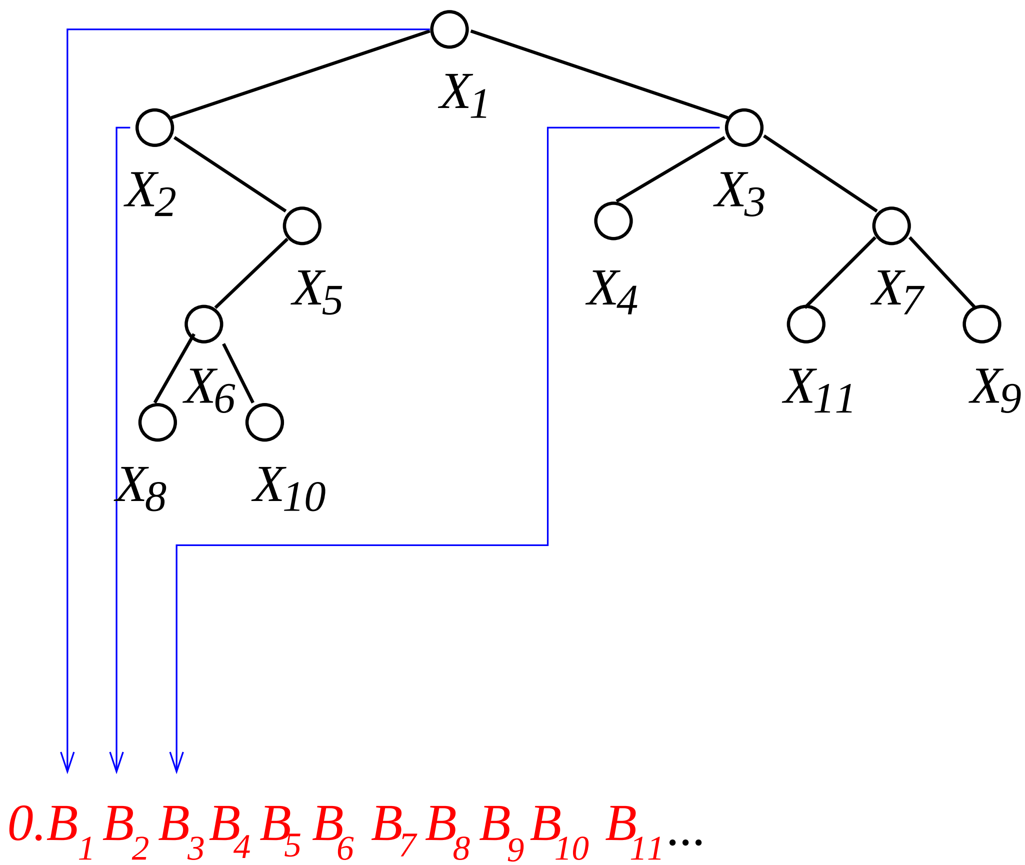
Suffixsuchbaum

Binärsuchbaum für Folge der Suffixe $X_1, X_2, \ldots$



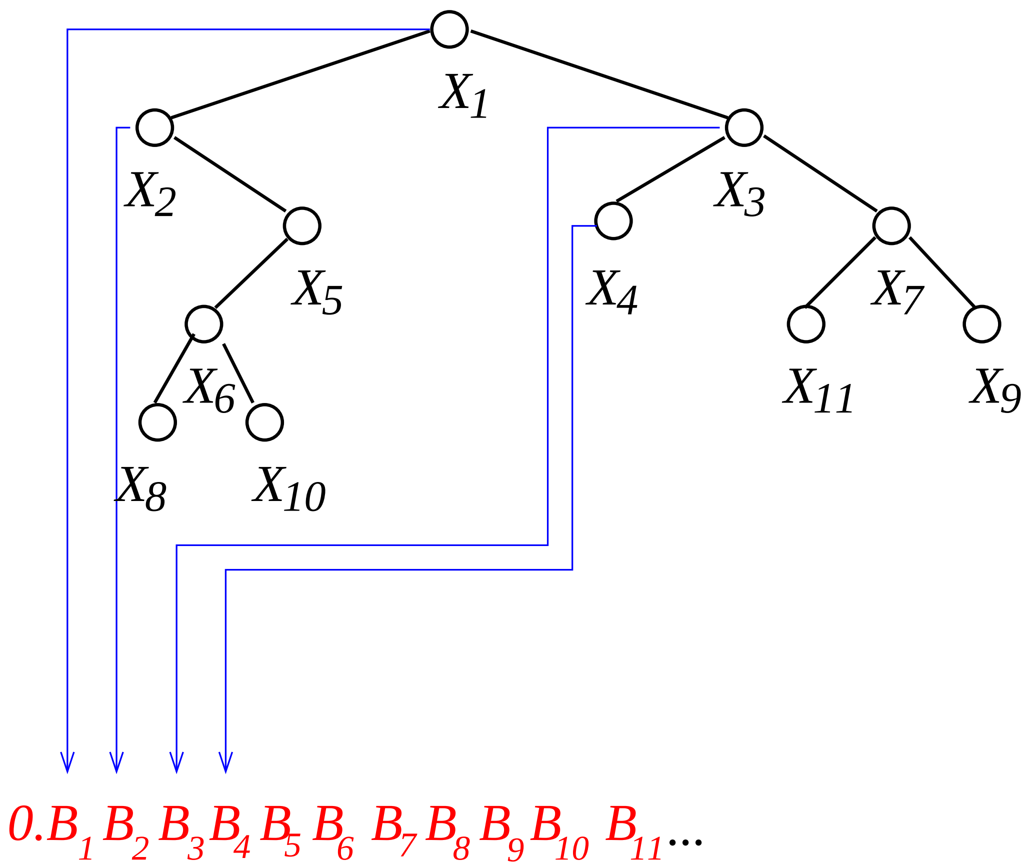
Suffixsuchbaum

Binärsuchbaum für Folge der Suffixe $X_1, X_2, \ldots$



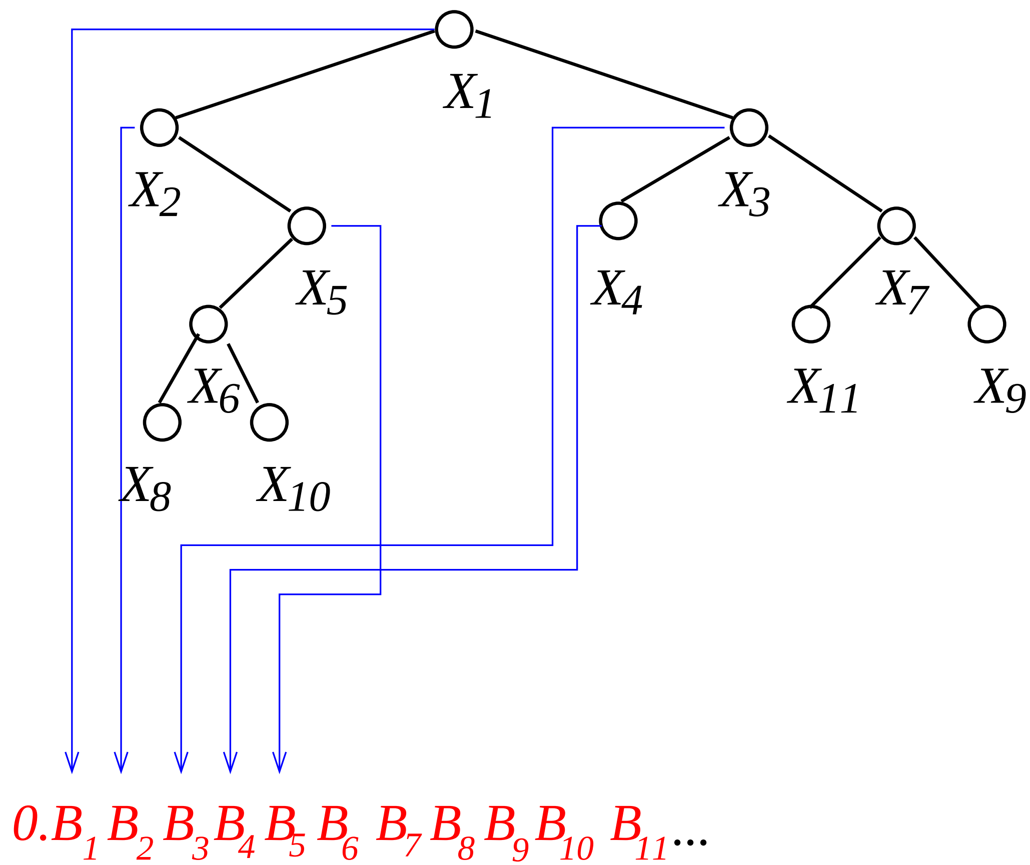
Suffixsuchbaum

Binärsuchbaum für Folge der Suffixe $X_1, X_2, \ldots$



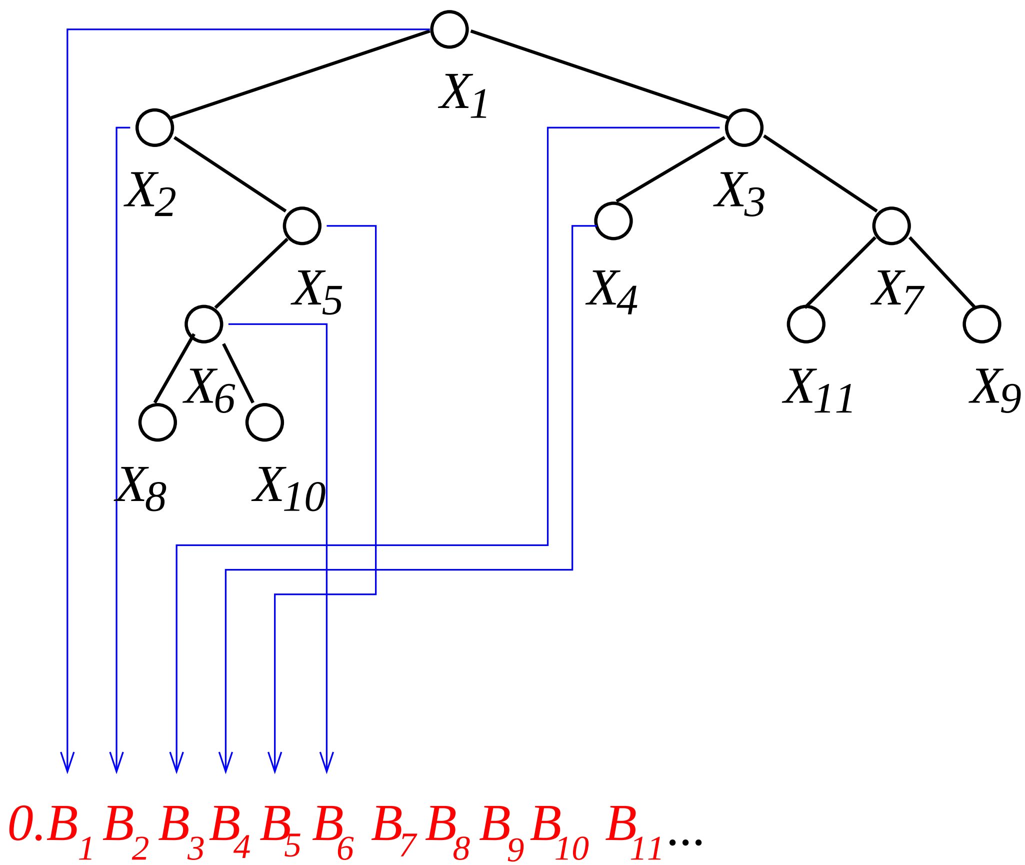
Suffixsuchbaum

Binärsuchbaum für Folge der Suffixe $X_1, X_2, \ldots$



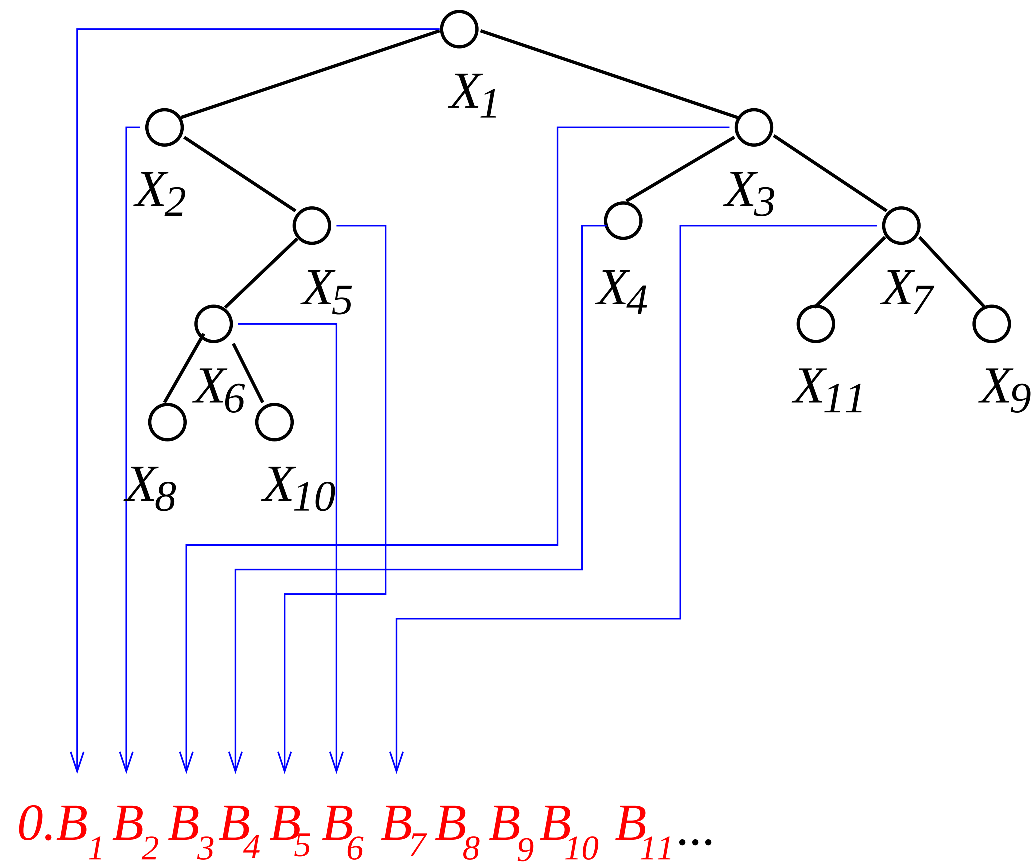
Suffixsuchbaum

Binärsuchbaum für Folge der Suffixe $X_1, X_2, \ldots$



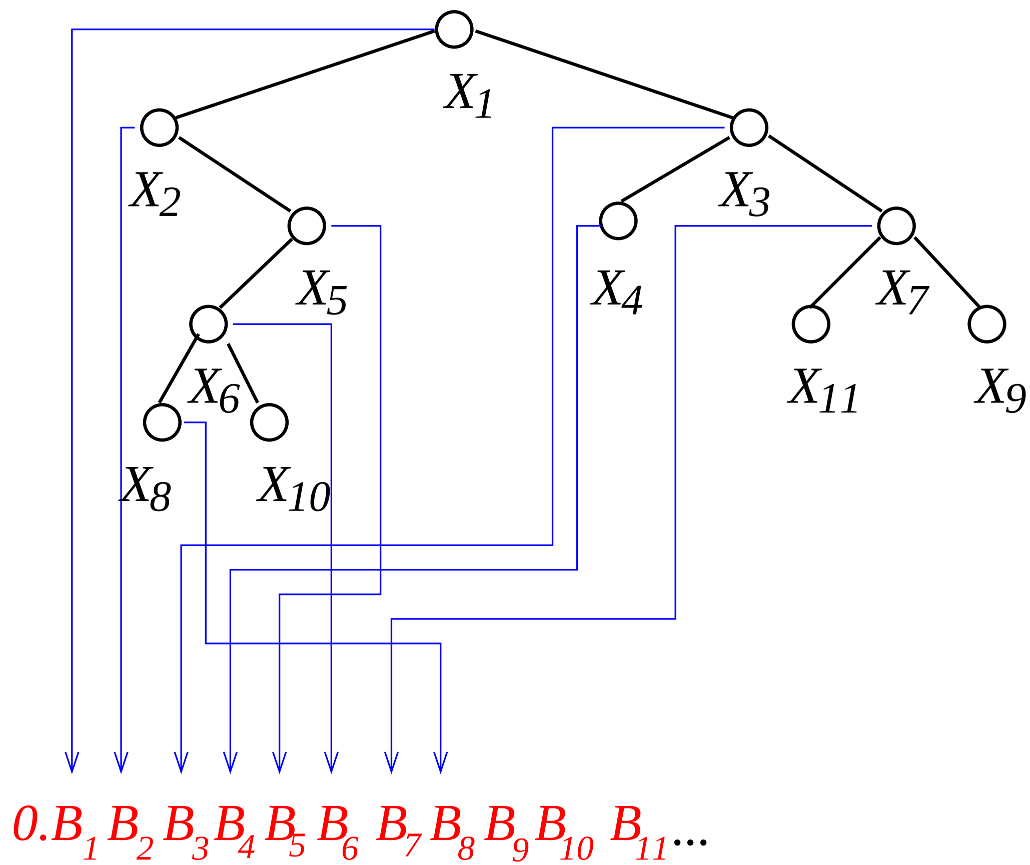
Suffixsuchbaum

Binärsuchbaum für Folge der Suffixe $X_1, X_2, \ldots$



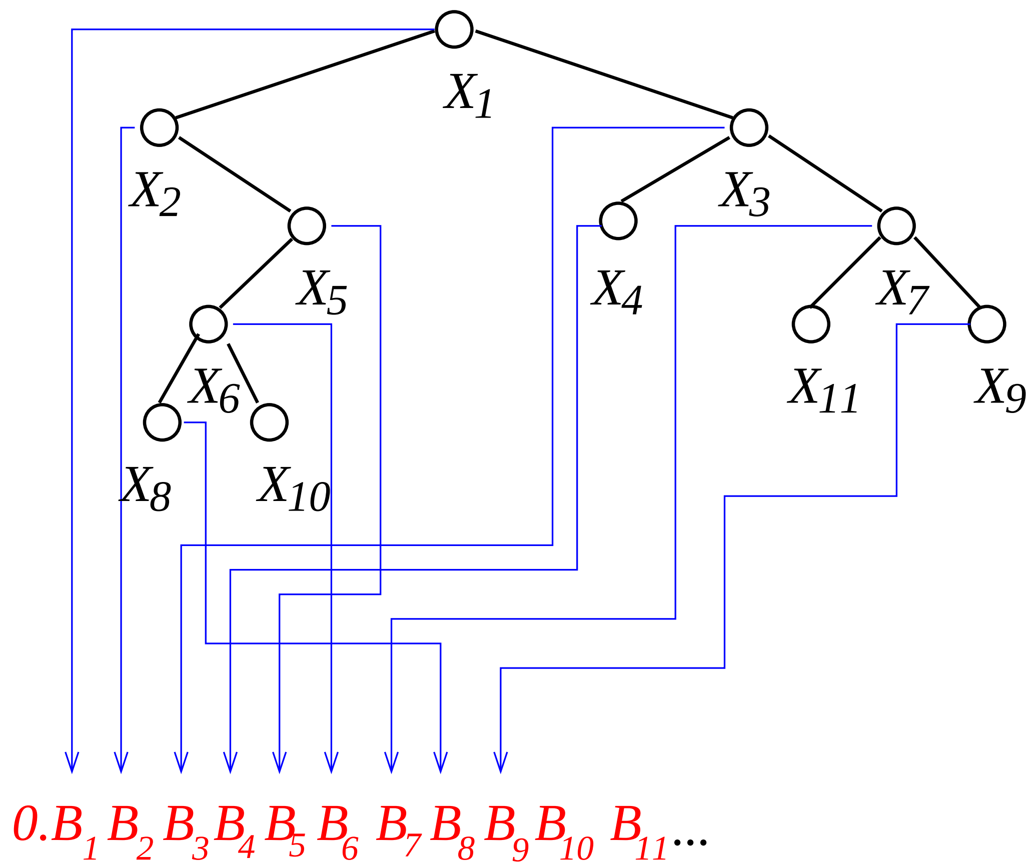
Suffixsuchbaum

Binärsuchbaum für Folge der Suffixe $X_1, X_2, \ldots$



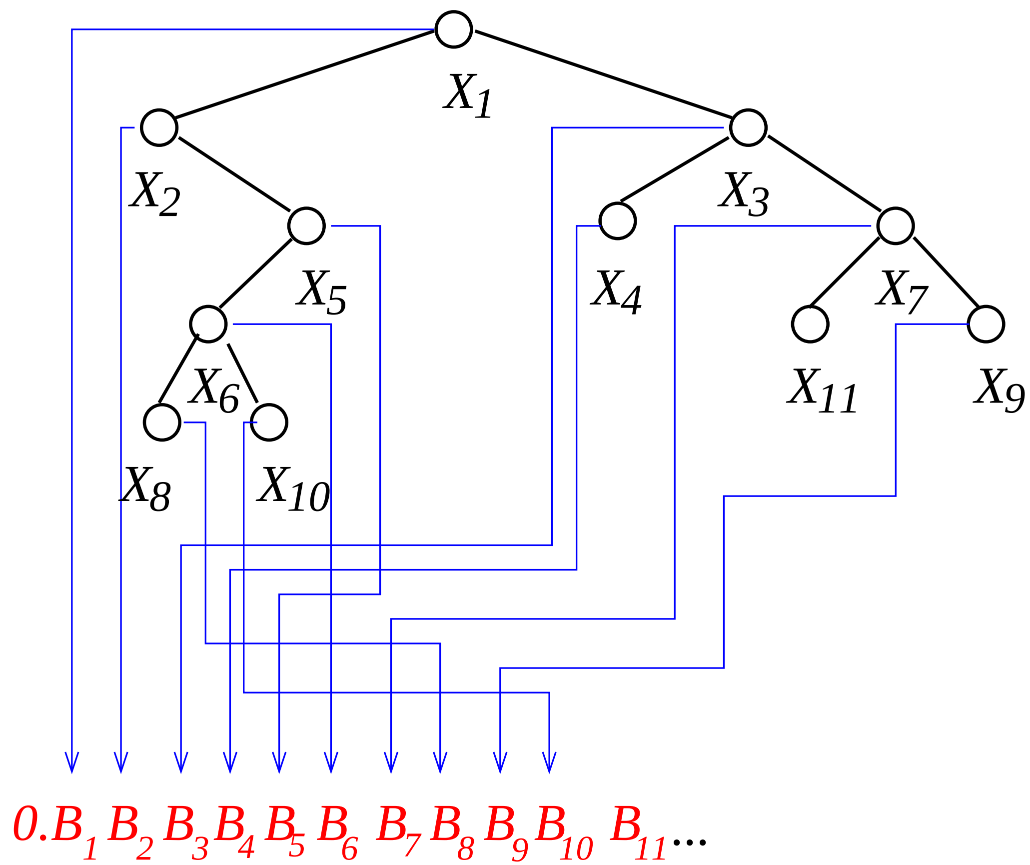
Suffixsuchbaum

Binärsuchbaum für Folge der Suffixe $X_1, X_2, \ldots$



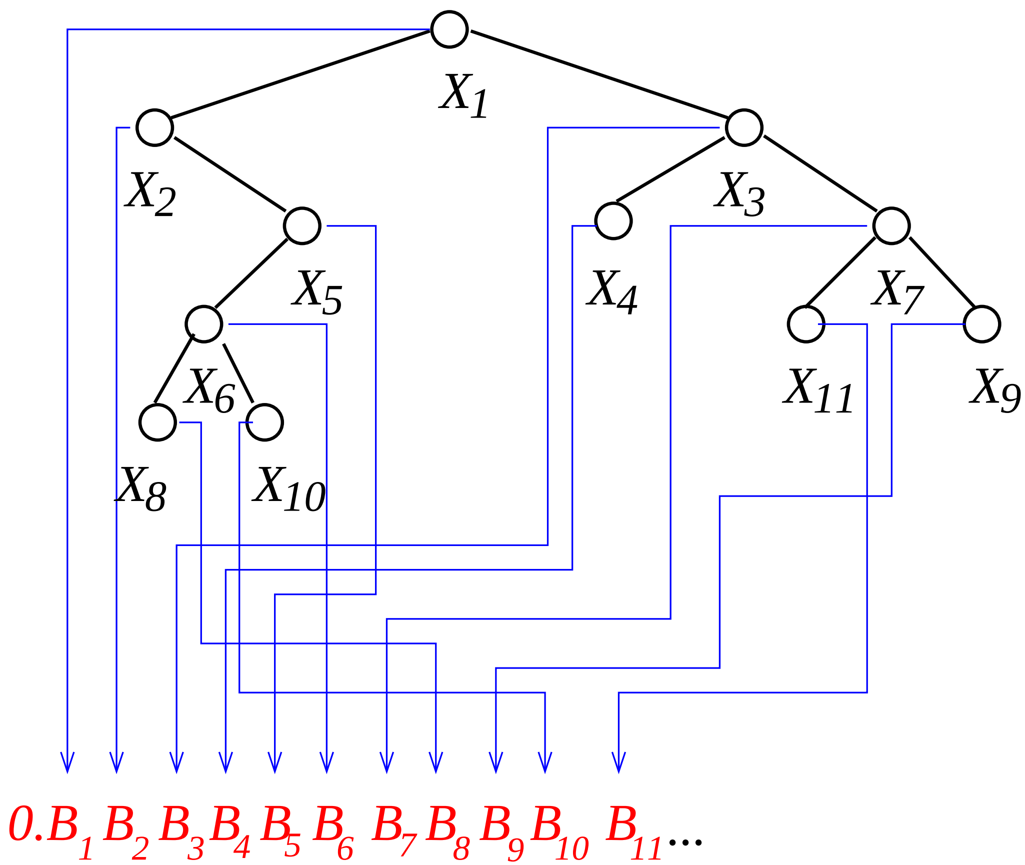
Suffixsuchbaum

Binärsuchbaum für Folge der Suffixe $X_1, X_2, \ldots$



Suffixsuchbaum

Binärsuchbaum für Folge der Suffixe $X_1, X_2, \ldots$



Pattern matching: Modell

Pattern matching: Modell

$$X_1 = 0.B_1 B_2 B_3 B_4 B_5 \ldots$$

Pattern matching: Modell

$$X_1 = 0.B_1 B_2 B_3 B_4 B_5 \ldots$$

Modell:

Pattern matching: Modell

$$X_1 = 0.B_1 B_2 B_3 B_4 B_5 \ldots$$

Modell: $B_1, B_2, \ldots$ i.i.d. Bernoulli(1/2).

Pattern matching: Modell

$$X_1 = 0.B_1 B_2 B_3 B_4 B_5 \ldots$$

Modell: $B_1, B_2, \ldots$ i.i.d. Bernoulli(1/2).

Äquivalente Darstellung:

Pattern matching: Modell

$$X_1 = 0.B_1 B_2 B_3 B_4 B_5 \ldots$$

Modell: $B_1, B_2, \ldots$ i.i.d. Bernoulli(1/2).

Äquivalente Darstellung:

$$X_1 \;=\; U, \qquad \text{mit} \qquad U \stackrel{d}{=} \text{unif}\,[0,1]$$

Pattern matching: Modell

$$X_1 = 0.B_1 B_2 B_3 B_4 B_5 \ldots$$

Modell: $B_1, B_2, \ldots$ i.i.d. Bernoulli(1/2).

Äquivalente Darstellung:

$$\begin{aligned} X_1 &= U, \qquad \text{mit} \qquad U \overset{d}{=} \text{unif}\,[0,1] \\ X_{n+1} &= 2X_n \bmod 1, \qquad n \geq 1. \end{aligned}$$

Satz (Devroye und N.)

Satz (Devroye und N.)

Im zufälligen Suffixsuchbaum gilt für die Tiefe D_n:

Satz (Devroye und N.)

Im zufälligen Suffixsuchbaum gilt für die Tiefe D_n:

$$\mathbb{E}\, D_n = 2 \log n +$$

Satz (Devroye und N.)

Im zufälligen Suffixsuchbaum gilt für die Tiefe D_n:

$$\mathbb{E}\, D_n \quad = \quad 2 \log n + O(\log^2 \log n),$$

Satz (Devroye und N.)

Im zufälligen Suffixsuchbaum gilt für die Tiefe D_n:

$$\mathbb{E}\,D_n = 2\log n + O(\log^2 \log n),$$

$$\frac{D_n}{\mathbb{E}\,D_n} \xrightarrow{\mathbb{P}} 1.$$

Asymptotisches Verhalten: Höhe

Höhe H_n:

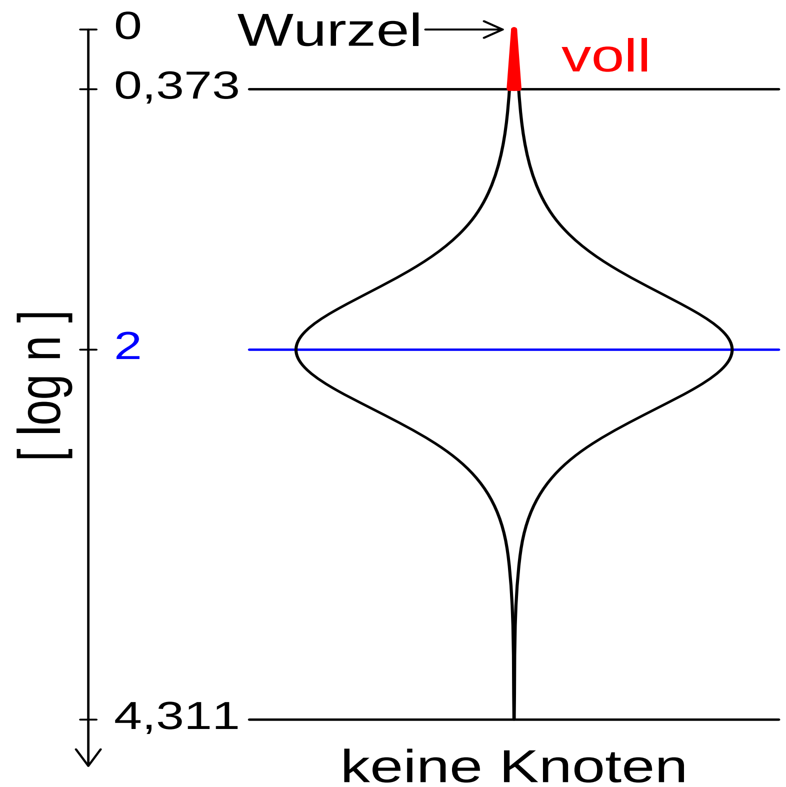
$$\mathbb{E}\, H_n = \alpha \log n + O(\log \log n),$$

$$\mathrm{Var}(H_n) = O(1), \qquad \frac{H_n}{\log n} \xrightarrow{\mathbb{P}} \alpha.$$

Sättigungslevel S_n:

$$\mathbb{E}\, S_n \sim \alpha_- \log n, \qquad \mathrm{Var}(S_n) = O(1),$$

$$\frac{S_n}{\log n} \xrightarrow{\mathbb{P}} \alpha_-$$



Asymptotisches Verhalten: Höhe

Höhe H_n:

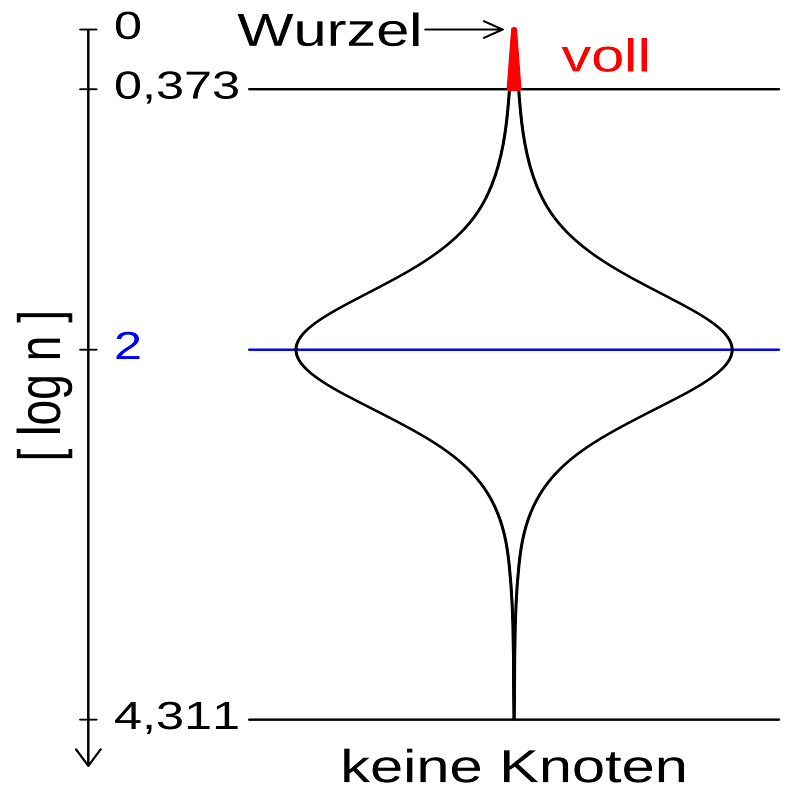
$$\mathbb{E}\, H_n = \alpha \log n + O(\log\log n),$$

$$\mathrm{Var}(H_n) = O(1), \qquad \frac{H_n}{\log n} \xrightarrow{\mathbb{P}} \alpha.$$

Sättigungslevel S_n:

$$\mathbb{E}\, S_n \sim \alpha_- \log n, \qquad \mathrm{Var}(S_n) = O(1),$$

$$\frac{S_n}{\log n} \xrightarrow{\mathbb{P}} \alpha_-$$



Dabei ist α die in $(2, \infty)$ eind. Lösung von

$$\alpha \log\left(\frac{2e}{\alpha}\right) = 1, \qquad \alpha \doteq 4,311$$

Pittel ('84), Devroye ('86), Reed ('03), Drmota ('03),. . .