Compacted Binary Trees:

A preliminary work on Implication Boolean formulas represented as Directed Acyclic Graphs

Antoine Genitrini

joint work with Bernhard Gittenberger, Manuel Kauers and Michael Wallner

Sorbonne Université

Laboratoire d'Informatique de Paris 6

May 24th 2018

Statistical properties of simple types

M. MOCZURAD[†], J. TYSZKIEWICZ[‡] and M. ZAIONC[†]

Abstract:

. . .

inhabited type in the set of all types. Under the Curry–Howard isomorphism this means finding the density or asymptotic probability of provable intuitionistic propositional formulas in the set of all formulas. For types with one ground type (formulas with one propositional variable), we prove that the limit exists and is equal to $1/2 + \sqrt{5}/10$, which is approximately 72.36%. This means that a long random type (formula) has this probability of being inhabited (tautology). We also prove that for every finite number k of ground-type

. . .

In Mathematical Structures in Computer Science (2000)

Statistical properties of simple types

M. MOCZURAD[†], J. TYSZKIEWICZ[‡] and M. ZAIONC[†]

Outline

Propositional logic with $\to$ and a single variable α

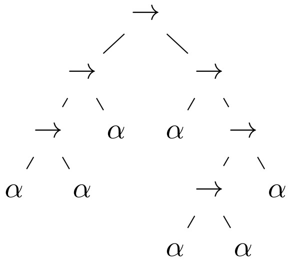
1. Boolean formulas represented by trees
2. The *unranking* method for exhaustive or random generation
3. Boolean formulas represented by circuits

approximately 72.56%. This means that a long random type (formula) has this probability of being inhabited (tautology). We also prove that for every finite number k of ground-type

. . .

Formulas in the *implication* system

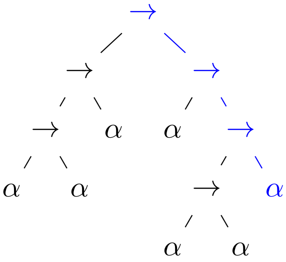
- A Boolean formula is seen as a binary tree, with the internal nodes decorated with the connective $\to$ and the leaves with the single variable α:



$$((\alpha \to \alpha) \to \alpha) \to (\alpha \to ((\alpha \to \alpha) \to \alpha))$$

Formulas in the *implication* system

- A Boolean formula is seen as a binary tree, with internal nodes decorated with the connective $\to$ and the leaves with the single variable α:

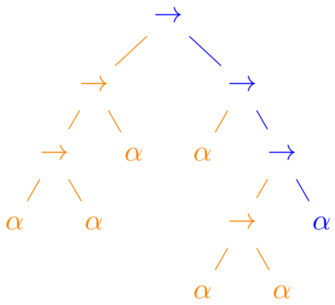


$$((\alpha \to \alpha) \to \alpha) \to (\alpha \to ((\alpha \to \alpha) \to \alpha))$$

- A canonical decomposition along the rightmost branch.

Formulas in the *implication* system

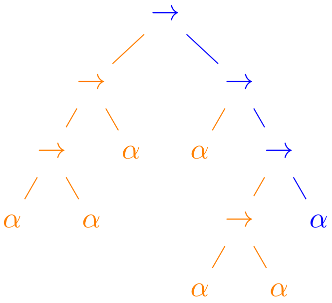
- A Boolean formula is seen as a binary tree, with internal nodes decorated with the connective $\to$ and the leaves with the single variable α:



- A canonical decomposition along the rightmost branch.
- The premises are the left subtrees of the righmost branch.

Formulas in the *implication* system

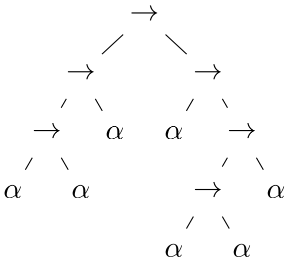
- A Boolean formula is seen as a binary tree, with internal nodes decorated with the connective $\rightarrow$ and the leaves with the single variable α:



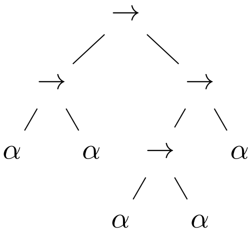
- A canonical decomposition along the rightmost branch.
- The premises are the left subtrees of the righmost branch.
- The *size* of a tree is its number of leaves.

Boolean functions in the *implication* system [S80,MTZ00]

In the *implication* system with one variable α, formulas are representing the function $\alpha \mapsto \texttt{true}$ or the function $\alpha \mapsto \alpha$.
All tautologies are intuitionistic tautologies in this context.



A *tautology*: it represents the function $\alpha \mapsto \texttt{true}$.



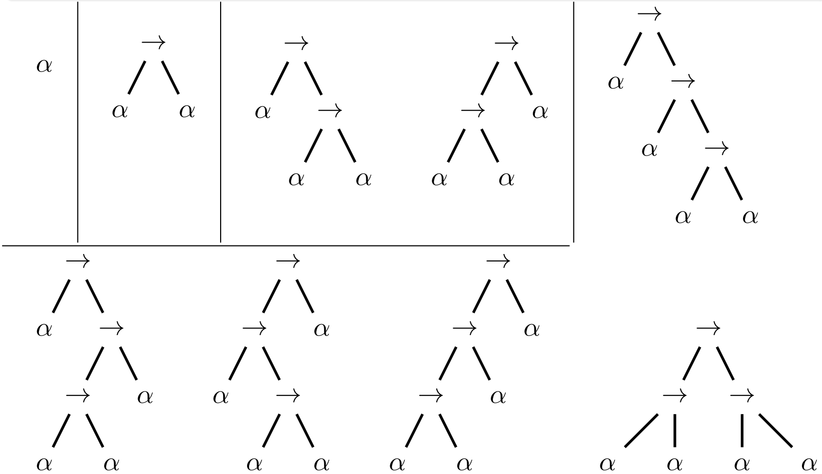
It represents the function $\alpha \mapsto \alpha$.

The *asymptotic ratio* of tautologies is the limit of the proportion of tautologies of size *n* among all formulas of size *n*, when *n* tends to infinity.

Asymptotic ratio of tautologies

Theorem: [MTZ00]

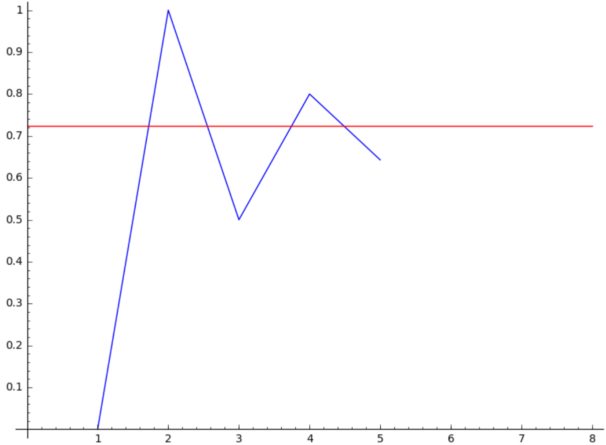
The *asymptotic ratio* of tautologies in the *implication* system with one variable is $\frac{1}{2} + \frac{\sqrt{5}}{10} \approx 0.72361\ldots$



Asymptotic ratio of tautologies

Theorem: [MTZ00]

The *asymptotic ratio* of tautologies in the *implication* system with one variable is $\frac{1}{2} + \frac{\sqrt{5}}{10} \approx 0.72361\ldots$

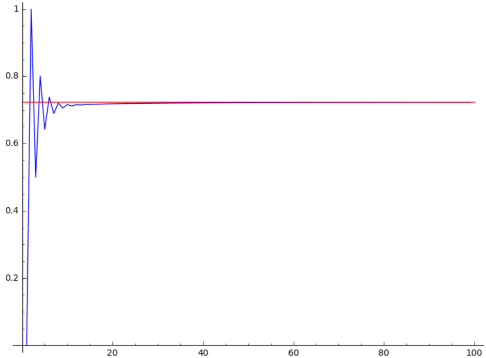


Asymptotic ratio of tautologies

Lemma: [MTZ00]

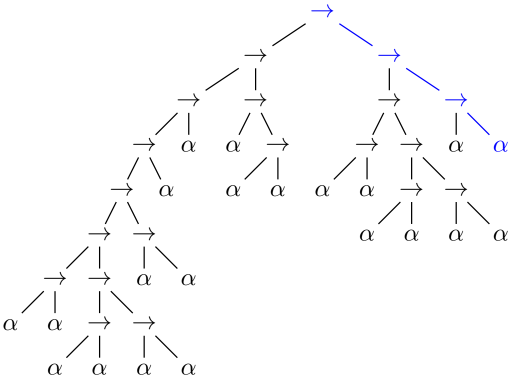
A formula $\tau \to \mu$ is not a tautology iff τ is a tautology and μ is not a tautology. Let Cat_{n-1}, T_n and N_n be respectively the number of formulas (resp. tautologies and non-tautologies) of size n,

$$N_n = \sum_{k=1}^{n-1} T_k \cdot N_{n-k}; \qquad N_1 = 1; \qquad T_n = Cat_{n-1} - N_n.$$



Simple tautologies

A *simple tautology* has one of its premises reduced to α.

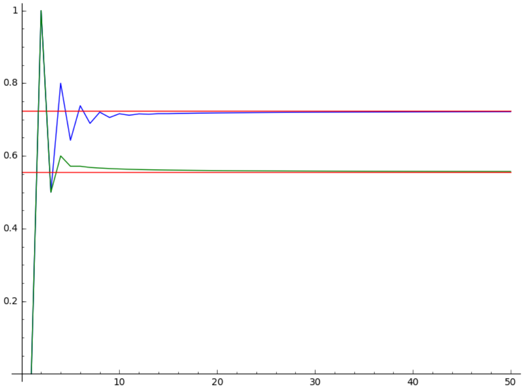


Simple tautologies

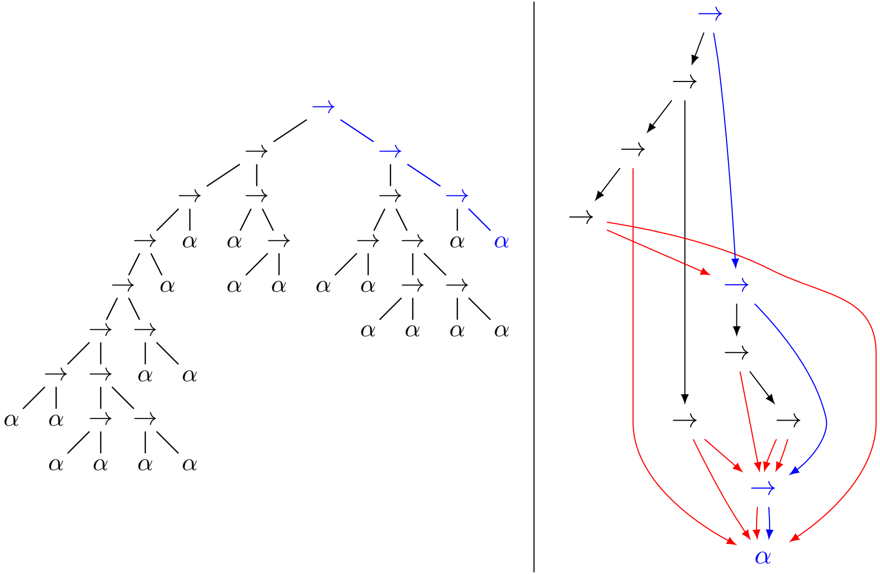
A *simple tautology* has one of its premises reduced to α.

Theorem: [MTZ00]

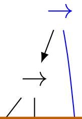
The *asymptotic ratio* of simple tautologies in the *implication* system with one variable is $\frac{5}{9} \approx 0.55556\ldots$



What about if we change the notion of formula ?



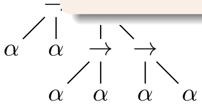
What about if we change the notion of formula ?

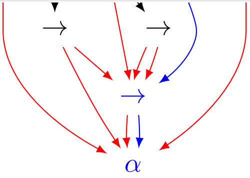


In the compacted context:

- Both children of a node are not of the same nature.
- At each internal node, the pointers in the left substructure are deeply related to the right substructure.
- The distribution of the size is completely disturbed.

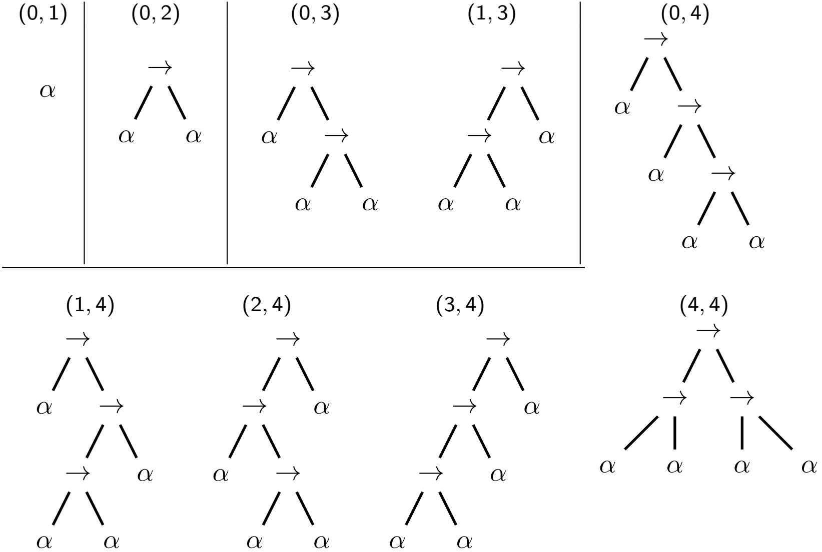
We want an exhaustive study for small compacted trees.





The *Unranking* method for binary trees [NW75,MM03]

1. Defining a total order over trees (of the same size).
2. Constructing the tree only by using its rank number.



Boustrophedon text[FZVC94]

Instead of reading the formula from left to right (increasing k):

$$B(n) = \sum_{k=1}^{n-1} B(k) \cdot B(n-k)$$

We read it as a Boustrophedon text
https://en.wikipedia.org/wiki/Boustrophedon

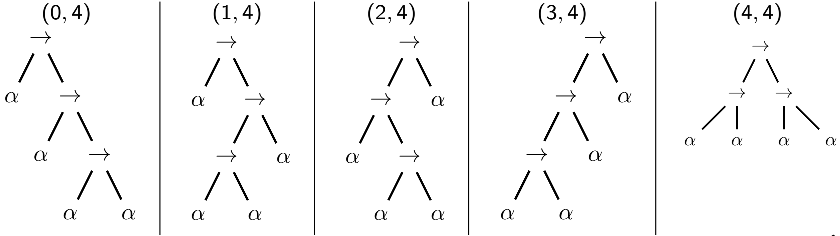
$B(n)$ as :
$B(1)B(n-1)+B(n-1)B(1)+B(2)B(n-2)+B(n-2)B(2)+B(3)B(n-3)+\dots$

Composition of trees

Boustrophedonic order [FZVC94]

Read $B(n)$ as :
$B(1)B(n-1)+B(n-1)B(1)+B(2)B(n-2)+B(n-2)B(2)+B(3)B(n-3)+\dots$

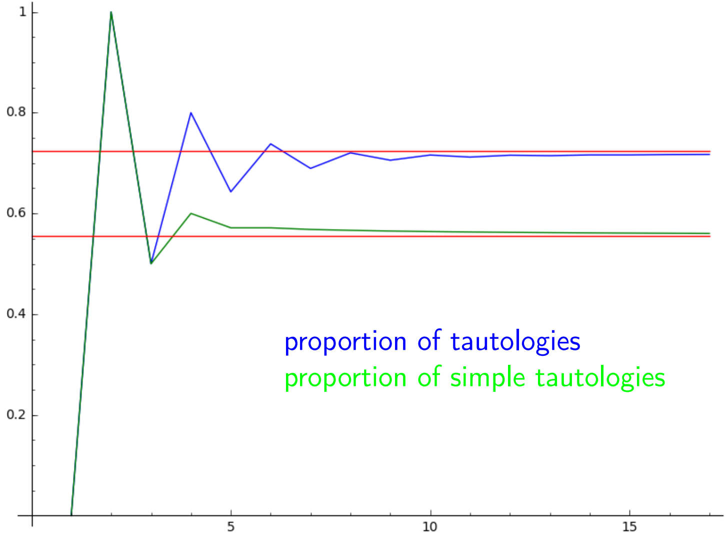
α and β being two trees of sizes resp. $s_\alpha < s_\beta$ then $\alpha < \beta$
α and β being two trees of size n: $\alpha < \beta$ iif α decomposes as $B(i)B(n-i)$ and β as $B(j)B(n-j)$ with
$\mathtt{min}(i, n-i) < \mathtt{min}(j, n-j)$ or
$\mathtt{min}(i, n-i) = \mathtt{min}(j, n-j)$ and $(\alpha_\ell < \beta_\ell$ or $\alpha_\ell = \beta_\ell$ and $\alpha_r < \beta_r)$



After 1 hour computations. . .

An complete description of all trees of size at most 17:

> 48 million objects.



$(0, 50)$ is a tautology (computed without my computer).

$(0, 50)$ is a tautology (computed without my computer).

$(3403667017946520494241856 11, 50)$ is a tautology:

```
n = 340366701794652049424185611
n > B(1)*B(49) ?
True
n = n - B(1)*B(49)

n > B(49)*B(1) ?
True
n = n - B(49)*B(1)

n > B(2)*B(48) + B(48)*B(2) ?
True
n = n - B(2)*B(48) - B(48)*B(2)

n > B(3)*B(47) ?
False
```

$(0, 50)$ is a tautology (computed without my computer).

$(3403667017946520494241856 11, 50)$ is a tautology:

```
n = 340366701794652049424185611
n > B(1)*B(49) ?
True
n = n - B(1)*B(49)

n > B(49)*B(1) ?
True
n = n - B(49)*B(1)

n > B(2)*B(48) + B(48)*B(2) ?
True
n = n - B(2)*B(48) - B(48)*B(2)

n > B(3)*B(47) ?
False
```

size decomposition: $50 \to 3 \times 47$

$(0, 50)$ is a tautology (computed without my computer).

$(340366701794652049424185611, 50)$ is a tautology:

```
n = 340366701794652049424185611
n > B(1)*B(49) ?
True
n = n - B(1)*B(49)

n > B(49)*B(1) ?
True
n = n - B(49)*B(1)

n > B(2)*B(48) + B(48)*B(2) ?
True
n = n - B(2)*B(48) - B(48)*B(2)

n > B(3)*B(47) ?
False
```

size decomposition: $50 \to 3 \times 47$

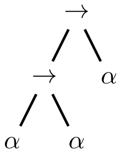
Euclidean division:
define p and r such that

$$n = p * B(47) + r$$

the tree
$(340366701794652049424185611, 50)$
corresponds to

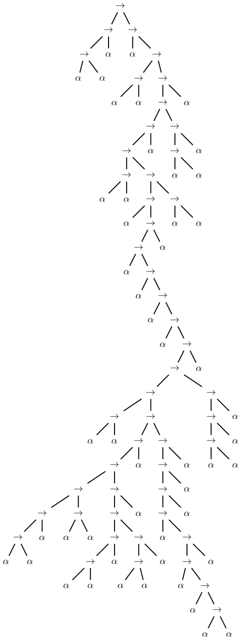
$(1, 3) \times (123302908439805130529651 1, 47)$

$(1, 3)$:



The tree is a tautology.

The tree $(340366701794652049424185611, 50)$



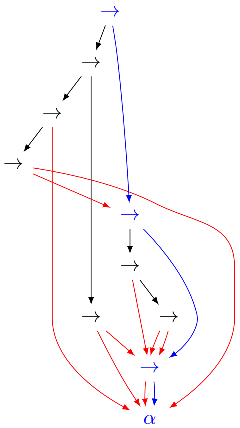
Let us turn to compacted formulas

Implication (relaxed) circuit

An *implication circuit* with one variable is a relaxed compacted binary tree whose internal nodes are decorated with the $\rightarrow$ connective and the leaf by the unique variable α.

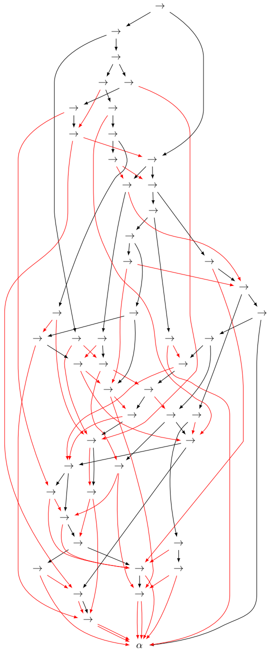
Implication compacted circuit

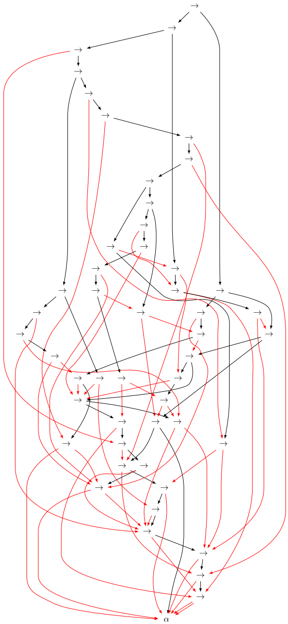
An *implication compacted circuit* with one variable is a compacted binary tree whose internal nodes are decorated with the $\rightarrow$ connective and the leaf by the unique variable α.



The *size* of the circuit is its total number of nodes.

Here two large circuits of size 50 (uniformly sampled)





$\Delta|$ The number of (relaxed) circuits of size n is equal to $\delta_{n,0}$:

Let $n, p \in \mathbb{N}$:

$$\delta_{0,p} = p \quad \text{for } p \geq 0,$$

$$\delta_{1,0} = 1, \qquad \delta_{1,p} = p^2 \qquad \text{for } p \geq 1,$$

$$\delta_{n,p} = \sum_{i=0}^{n-1} \delta_{n-1-i,p+i}\, \delta_{i,p}, \qquad \text{for } n \geq 2.$$

$\delta_{n,0} = 0, 1, 1, 3, 16, 127, 1363, 18628, 311250, 6173791, 142190703, \ldots$

Implication circuits

Δ| The number of (relaxed) circuits of size n is equal to $\delta_{n,0}$:

Let $n, p \in \mathbb{N}$:

$$\delta_{0,p} = p \quad \text{for } p \geq 0,$$

$$\delta_{1,0} = 1, \qquad \delta_{1,p} = p^2 \qquad \text{for } p \geq 1,$$

$$\delta_{n,p} = \sum_{i=0}^{n-1} \delta_{n-1-i,p+i}\ \delta_{i,p}, \qquad \text{for } n \geq 2.$$

$\delta_{n,0} = 0, 1, 1, 3, 16, 127, 1363, 18628, 311250, 6173791, 142190703, \ldots$

Γ| The number of compacted circuits of size n is equal to $\gamma_{n,0}$:

Let $n, p \in \mathbb{N}$:

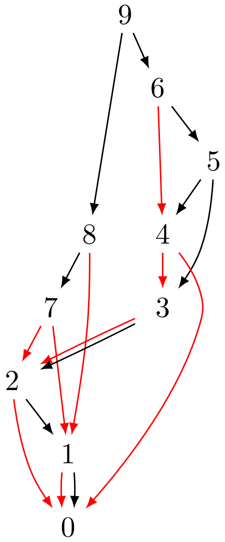
$$\gamma_{0,p} = p \quad \text{for } p \geq 0,$$

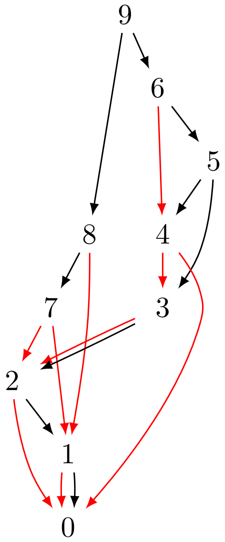
$$\gamma_{1,0} = 1, \qquad \gamma_{1,p} = p^2 - (p-1) \quad \text{for } p \geq 1,$$

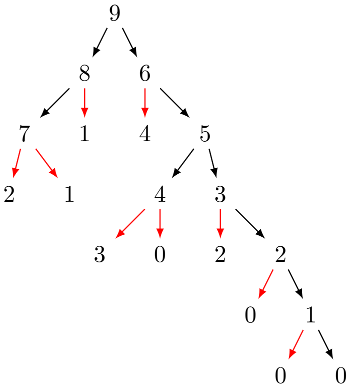
$$\gamma_{n,p} = \sum_{i=0}^{n-1} \gamma_{n-1-i,p+i}\ \gamma_{i,p}, \qquad \text{for } n \geq 2.$$

$\gamma_{n,0} = 0, 1, 1, 3, 15, 111, 1119, 14487, 230943, 4395855, 97608831, \ldots$

Unranking method for circuit building

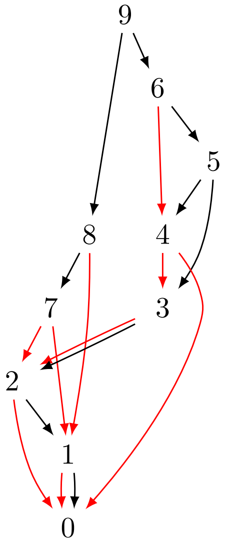


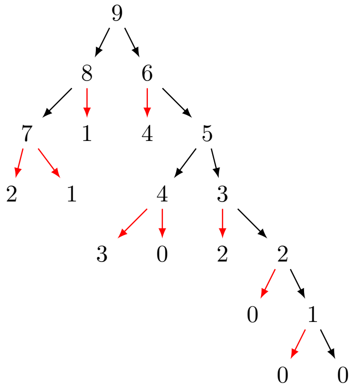




Successive forbidden pairs of children:

Unranking method for circuit building

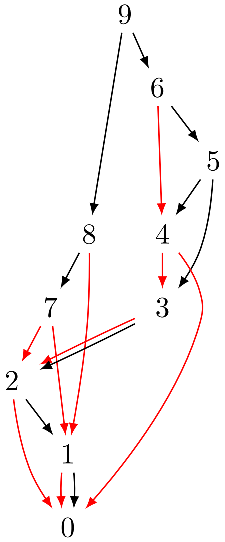


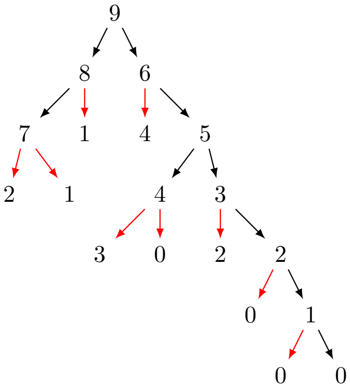


Successive forbidden pairs of children:

(0, 0)

Unranking method for circuit building

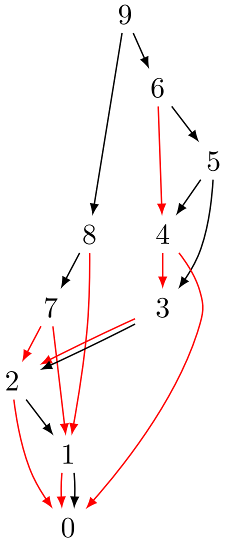


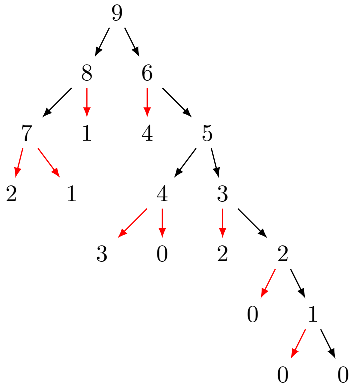


Successive forbidden pairs of children:

$(0, 1)$
$(0, 0)$

Unranking method for circuit building

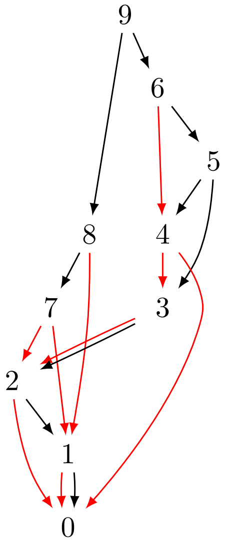


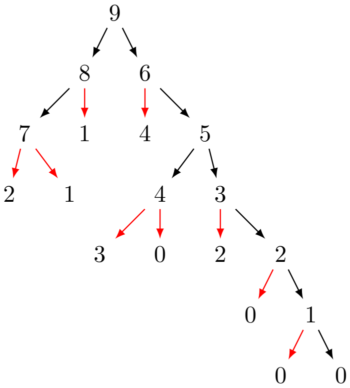


Successive forbidden pairs of children:

$(2, 2)$
$(0, 1)$
$(0, 0)$

Unranking method for circuit building

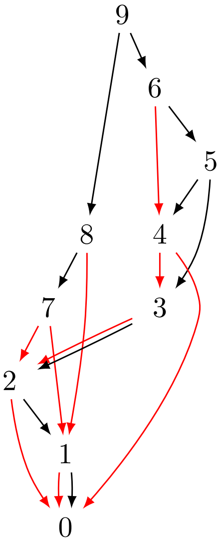


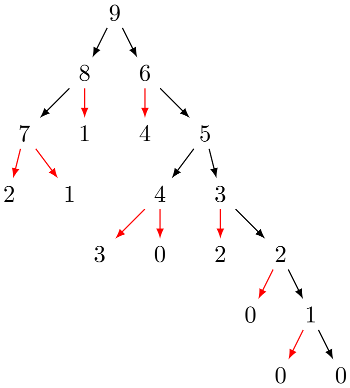


Successive forbidden pairs of children:

(3, 0)
(2, 2)
(0, 1)
(0, 0)

Unranking method for circuit building

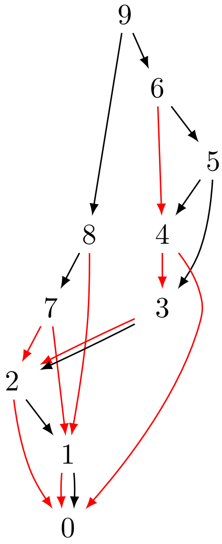


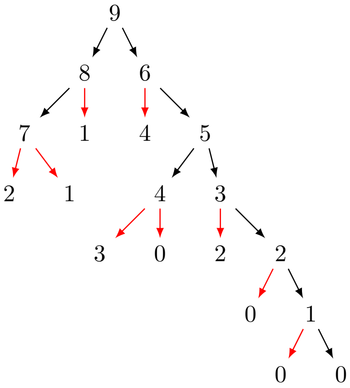


Successive forbidden pairs of children:

$(4, 3)$
$(3, 0)$
$(2, 2)$
$(0, 1)$
$(0, 0)$

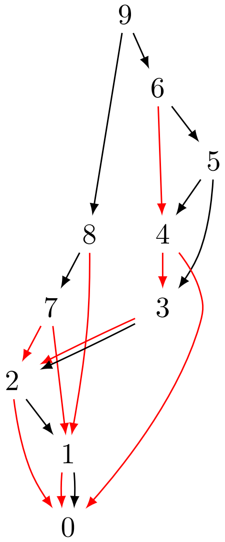
Unranking method for circuit building

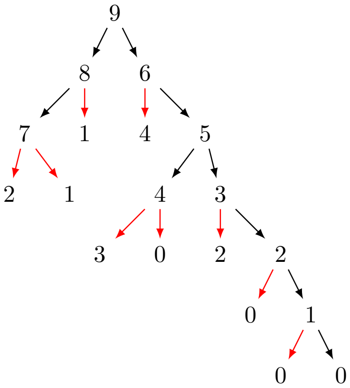




Successive forbidden pairs of children:

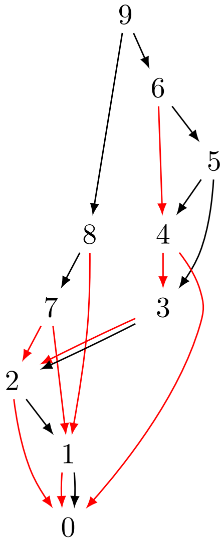
$(4, 5)$
$(4, 3)$
$(3, 0)$
$(2, 2)$
$(0, 1)$
$(0, 0)$

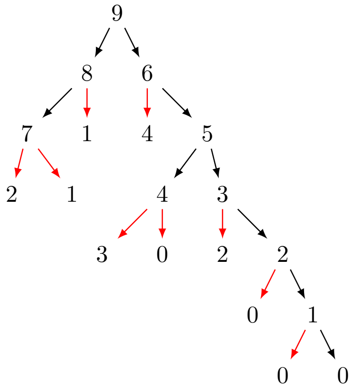




Successive forbidden pairs of children:

$(2, 1)$
$(4, 5)$
$(4, 3)$
$(3, 0)$
$(2, 2)$
$(0, 1)$
$(0, 0)$

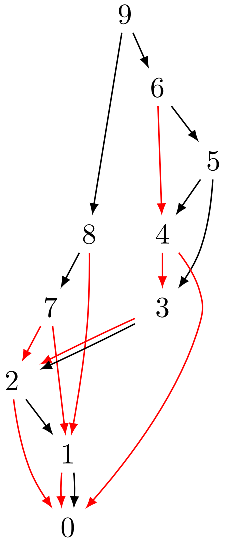


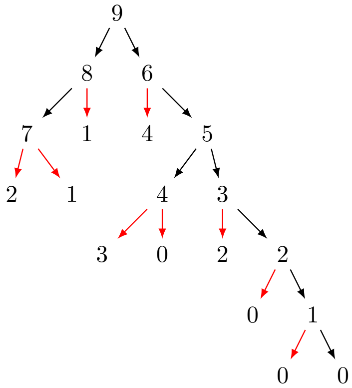


Successive forbidden pairs of children:

(7, 1)
(2, 1)
(4, 5)
(4, 3)
(3, 0)
(2, 2)
(0, 1)
(0, 0)

Unranking method for circuit building

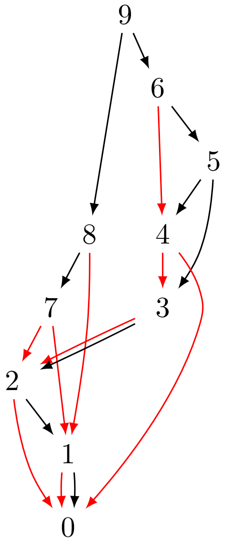


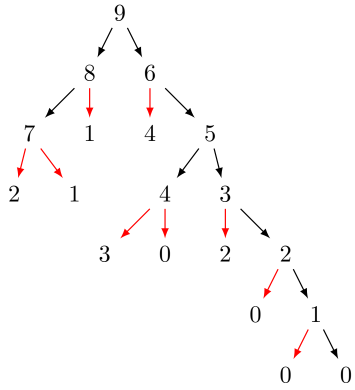


Successive forbidden pairs of children:

(8, 6)
(7, 1)
(2, 1)
(4, 5)
(4, 3)
(3, 0)
(2, 2)
(0, 1)
(0, 0)

Unranking method for circuit building





Successive forbidden pairs of children:

$(8, 6)$
$(7, 1)$
$(2, 1)$
$(4, 5)$
$(4, 3)$
$(3, 0)$
$(2, 2)$
$(0, 1)$
$(0, 0)$

The circuit in Γ decomposes as $\gamma_{2,6}\ \gamma_{6,0}$.

Unranking algorithm

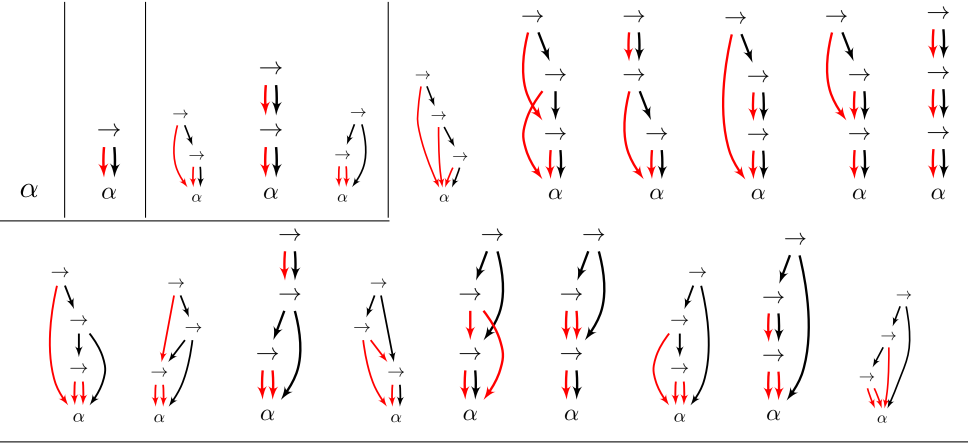
function build(j, n, p, num, F)
  if $n = 0$ **then**
    return $(num, F, \mathbf{point}(j))$
  if $n = 1$ and $p = 0$ **then**
    return $(num + 1, F, \mathbf{node}(j, \emptyset, \emptyset))$
  if $n = 1$ **then**
    $(F', C1, C2) := \mathbf{children}(j, num, F)$
    return $(num + 1, F', \mathbf{node}(j, C1, C2))$

  $(j', i) := \mathbf{decompo}(j, n, p)$
    # $[\gamma_{n-2,1}\gamma_{1,p},\ \gamma_{n-3,2}\gamma_{2,p},\ \gamma_{n-4,3}\gamma_{3,p}, \cdots]$

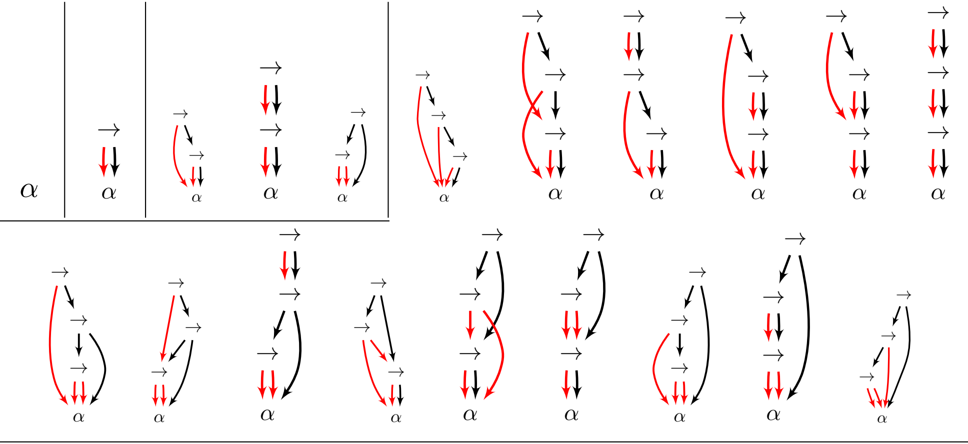
  $(num', F', C2) := \mathbf{build}(j' \ // \ \gamma_{n-1-i,p+i}, i, p, num, F)$
  $(num', F', C1) := \mathbf{build}(j' \ \% \ \gamma_{n-1-i,p+i}, n-1-i, p+i, num', F')$
  $C := \mathbf{node}(num', C1, C2)$
  $F' := \mathbf{insert}(\mathbf{pair}(C1, C2), F')$

  return $(num' + 1, F', C)$

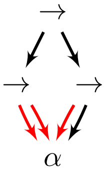
Smallest circuits



Smallest circuits

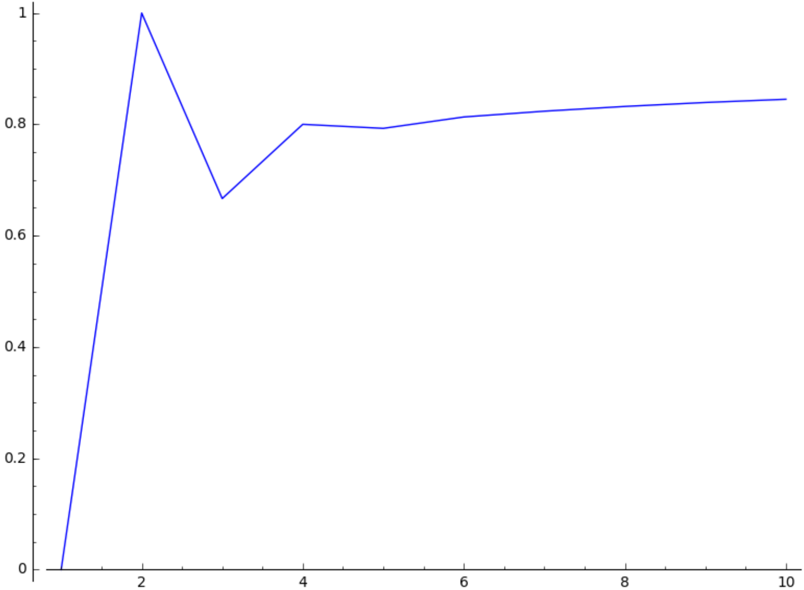


The smallest relaxed circuit that is not a compacted circuit:



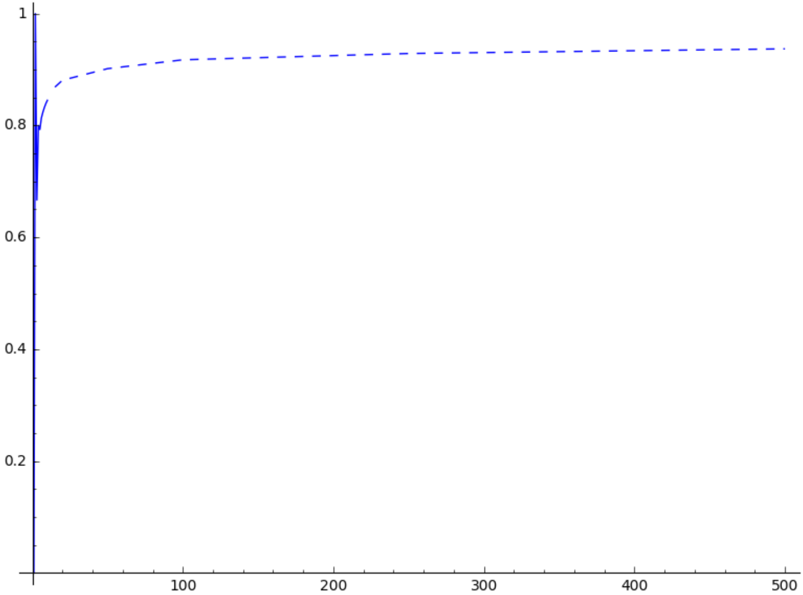
Proportion of tautologies by exhaustive computations

Compacted tautologies ($>$ 102 million objects).



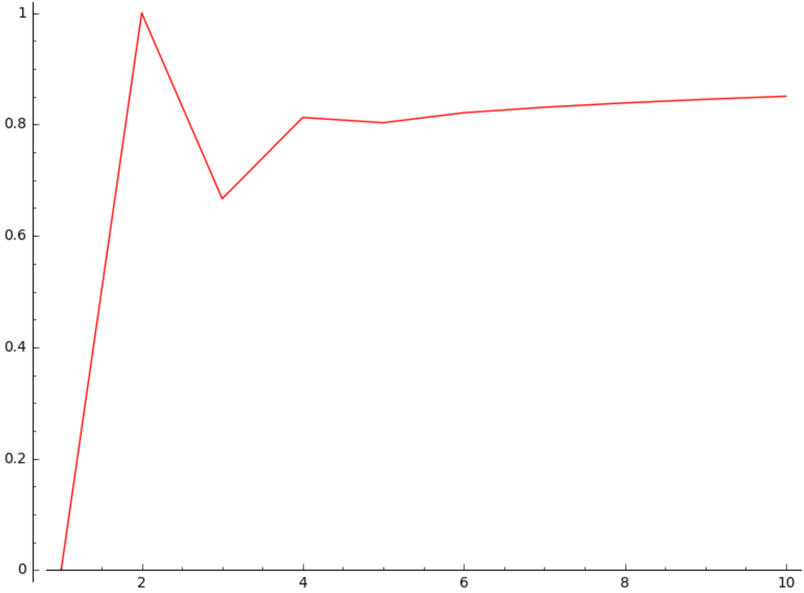
Proportion of compacted tautologies

Unif. sampling of 10 000 objects of sizes: $15, 20, 50, 100, 250, 500$.



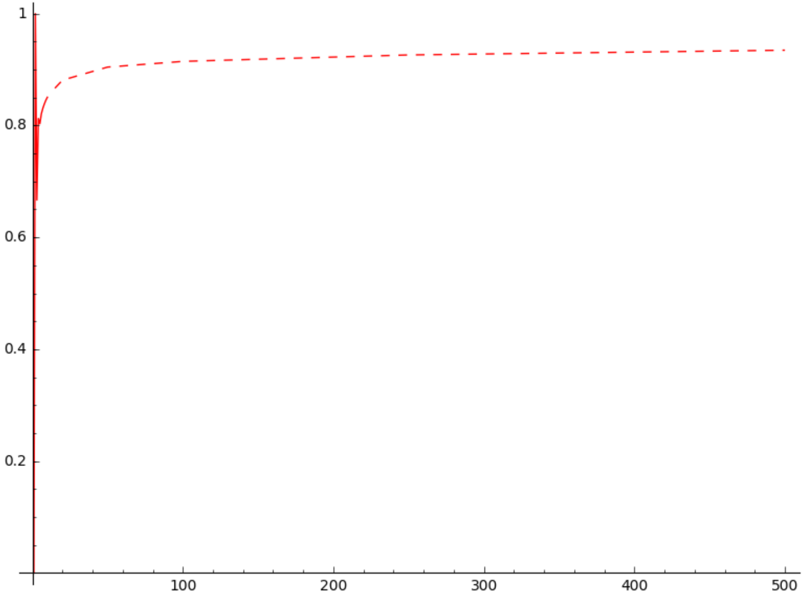
Proportion of tautologies by exhaustive computations

relaxed tautologies (> 148 million objects).



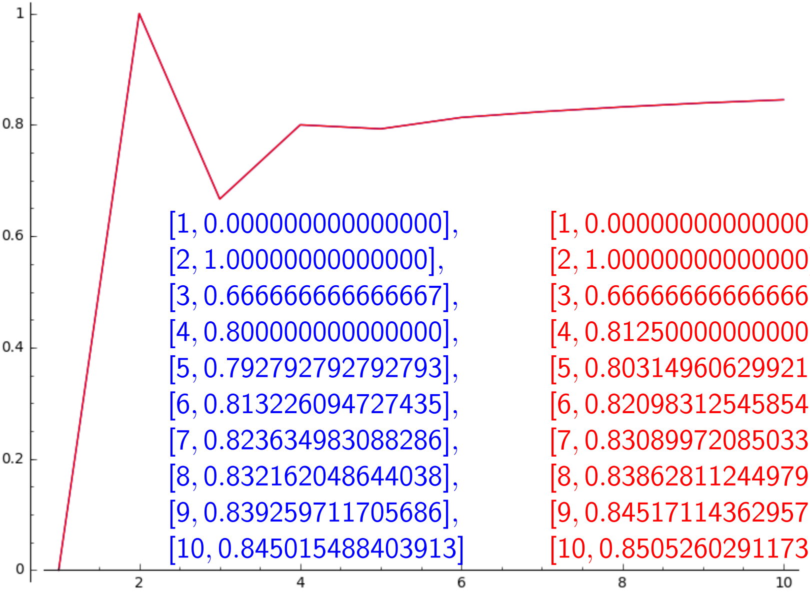
Proportion of compacted tautologies

Unif. sampling of 10 000 objects of sizes: $15, 20, 50, 100, 250, 500$.



Proportion of tautologies by exhaustive computations

compacted tautologies relaxed tautologies

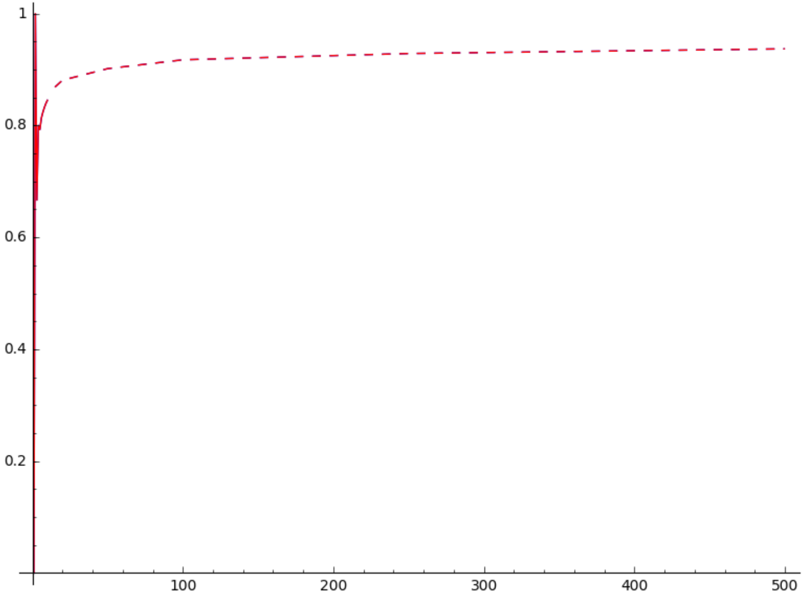


[1, 0.000000000000000],
[2, 1.00000000000000],
[3, 0.666666666666667],
[4, 0.800000000000000],
[5, 0.792792792792793],
[6, 0.813226094727435],
[7, 0.823634983088286],
[8, 0.832162048644038],
[9, 0.839259711705686],
[10, 0.845015488403913]

[1, 0.000000000000000],
[2, 1.00000000000000],
[3, 0.666666666666667],
[4, 0.812500000000000],
[5, 0.803149606299213],
[6, 0.820983125458547],
[7, 0.830899720850333],
[8, 0.838628112449799],
[9, 0.845171143629579],
[10, 0.850526029117389]

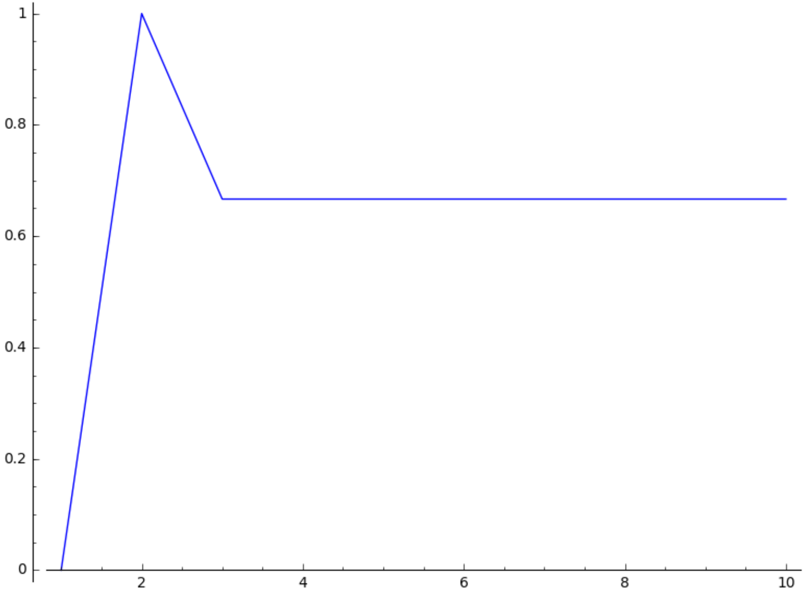
Proportion of compacted tautologies

compacted tautologies relaxed tautologies



Proportion of compacted simple tautologies

Compacted simple tautologies



Enumeration of compacted simple tautologies

The number of compacted simple tautologies circuits of size n is equal to g_n :

Let $n, p \in \mathbb{N}$:

$$g_1 = 1$$

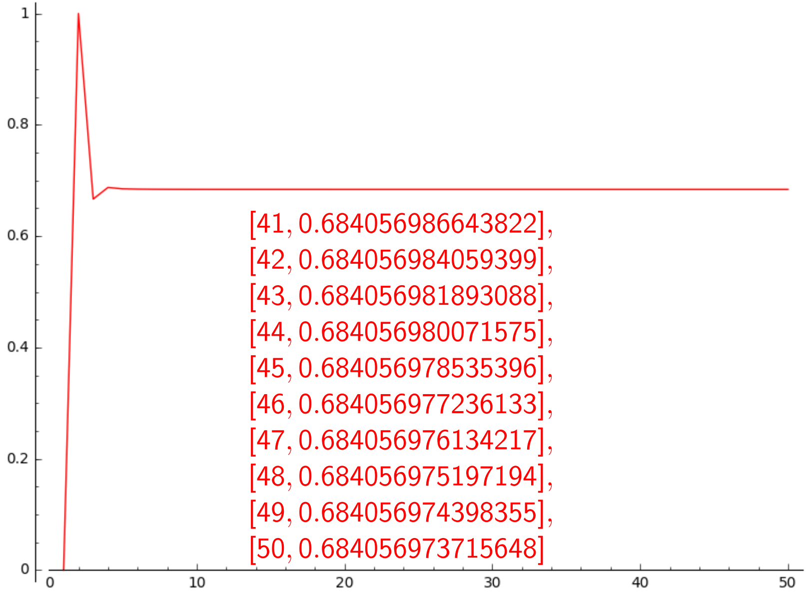
$$g_n = \left(\sum_{i=1}^{n-1} \delta_{n-1-i,i} \; g_i \right) + \gamma_{n-1,0} - g_{n-1}, \qquad \text{for } n \geq 2.$$

Theorem

For $n \geq 3$, the proportion of compacted simple tautologies is $\frac{1}{3}$.

Proportion of relaxed simple tautologies

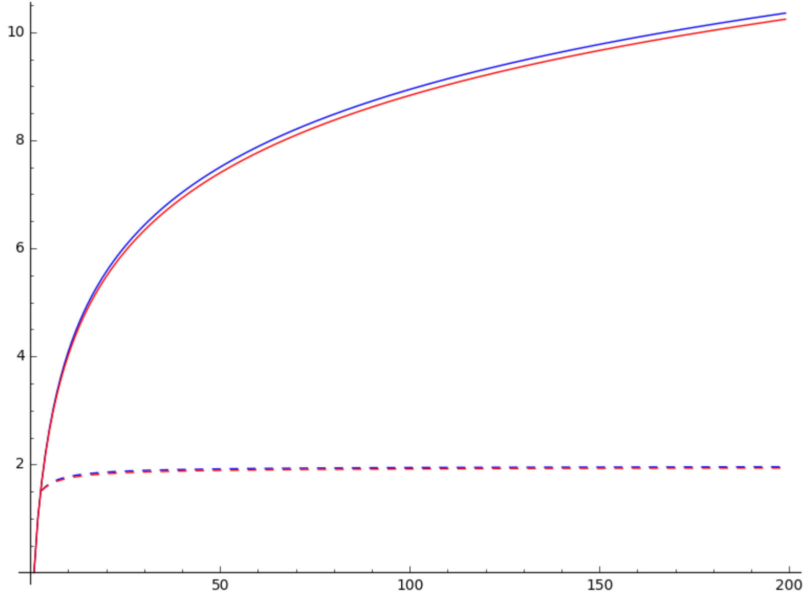
Relaxed simple tautologies



[41, 0.684056986643822],
[42, 0.684056984059399],
[43, 0.684056981893088],
[44, 0.684056980071575],
[45, 0.684056978535396],
[46, 0.684056977236133],
[47, 0.684056976134217],
[48, 0.684056975197194],
[49, 0.684056974398355],
[50, 0.68405697371564 8]

Mean value of the number of premises of circuits

solid line: mean value   dashed line: mean value normalized by log(*size*)



Perspectives

Many properties to prove !!