Associative and commutative tree representation for Boolean functions

A. Genitrini, B. Gittenberger, V. Kraus & C. Mailler

Université P. et M. Curie,
Technische Universität Wien &
Université de Versailles St-Quentin-en-Yvelines.

17th. June 2011

1. Tree representation for Boolean functions
   - Plane binary model
   - Commutative or associative models

2. Probability distributions on Boolean functions
   - Relation between complexity and probability
   - Key ideas of the proof

3. Conclusion and perspectives

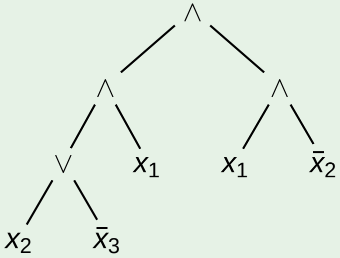
Tree representation for Boolean functions

Boolean formulas

- A set of literals: $\{x_1, \bar{x}_1, \ldots, x_n, \bar{x}_n\}$
- A set of connectors: $\{\wedge, \vee\}$

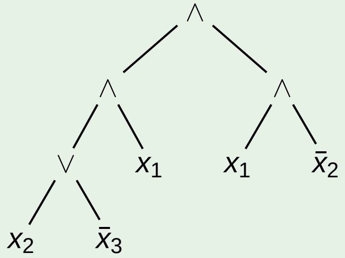
Boolean formulas

- A set of literals: $\{x_1, \bar{x}_1, \dots, x_n, \bar{x}_n\}$
- A set of connectors: $\{\wedge, \vee\}$
- Plane binary tree representation of $((x_2 \vee \bar{x}_3) \wedge x_1) \wedge (x_1 \wedge \bar{x}_2)$



Boolean formulas

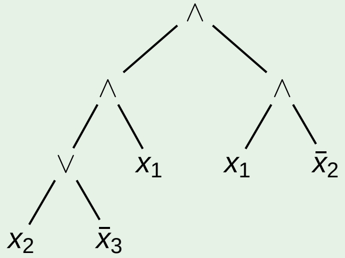
- A set of literals: $\{x_1, \bar{x}_1, \ldots, x_n, \bar{x}_n\}$
- A set of connectors: $\{\wedge, \vee\}$
- Plane binary tree representation of $((x_2 \vee \bar{x}_3) \wedge x_1) \wedge (x_1 \wedge \bar{x}_2)$



- The set of all plane binary and labelled trees: $\mathcal{T}_n$

Boolean formulas

- A set of literals: $\{x_1, \bar{x}_1, \ldots, x_n, \bar{x}_n\}$
- A set of connectors: $\{\wedge, \vee\}$
- Plane binary tree representation of $((x_2 \vee \bar{x}_3) \wedge x_1) \wedge (x_1 \wedge \bar{x}_2)$



- The set of all plane binary and labelled trees: $\mathcal{T}_n$
- The **size** of a tree corresponds to its number of leaves

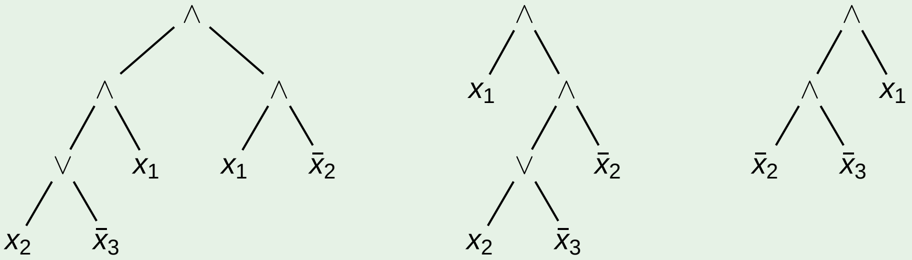
Boolean functions

- Each formula represents one function

Boolean functions

- Each formula represents one function
- There are **infinitely** many formulas for each function

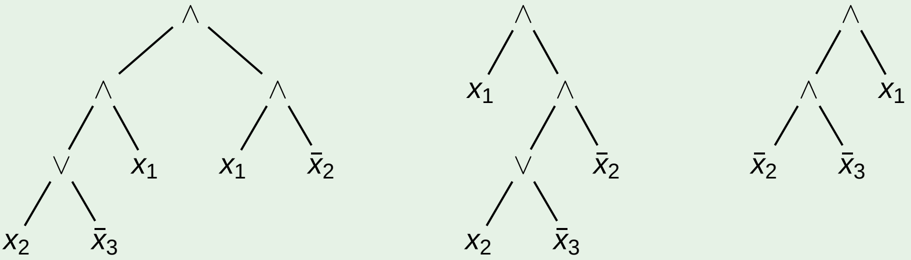
$((x_2 \vee \bar{x}_3) \wedge x_1) \wedge (x_1 \wedge \bar{x}_2) \quad \sim \quad x_1 \wedge ((x_2 \vee \bar{x}_3) \wedge \bar{x}_2) \quad \sim \quad (\bar{x}_2 \wedge \bar{x}_3) \wedge x_1$



Boolean functions

- Each formula represents one function
- There are **infinitely** many formulas for each function

$$((x_2 \vee \bar{x}_3) \wedge x_1) \wedge (x_1 \wedge \bar{x}_2) \quad \sim \quad x_1 \wedge ((x_2 \vee \bar{x}_3) \wedge \bar{x}_2) \quad \sim \quad (\bar{x}_2 \wedge \bar{x}_3) \wedge x_1$$

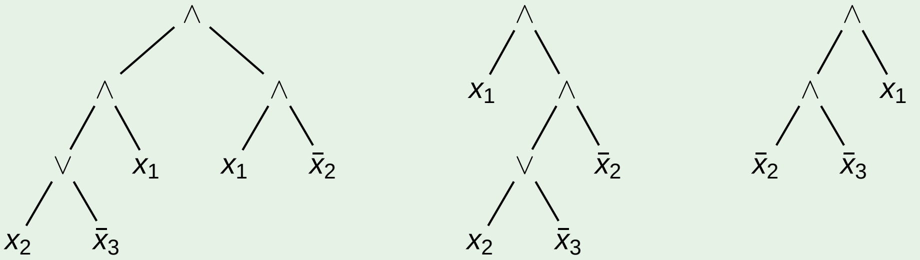


- The set of all Boolean functions on n variables: $\mathcal{F}_n$

Boolean functions

- Each formula represents one function
- There are **infinitely** many formulas for each function

$$((x_2 \vee \bar{x}_3) \wedge x_1) \wedge (x_1 \wedge \bar{x}_2) \quad \sim \quad x_1 \wedge ((x_2 \vee \bar{x}_3) \wedge \bar{x}_2) \quad \sim \quad (\bar{x}_2 \wedge \bar{x}_3) \wedge x_1$$



- The set of all Boolean functions on n variables: $\mathcal{F}_n$
- The **complexity** of a function is the size of its smallest trees

Probability distribution

- $\mathcal{A}$ a subset of $\mathcal{T}_n$
- $\mathcal{A}(m)$ the set of all formulas of size m in $\mathcal{A}$

[Lefman, Savický. *Some typical properties of large And / Or boolean formulas*, 1997]

[Chauvin, Flajolet, Gardy, Gittenberger. *And / Or trees revisited*, 2004]

Probability distribution

- $\mathcal{A}$ a subset of $\mathcal{T}_n$
- $\mathcal{A}(m)$ the set of all formulas of size m in $\mathcal{A}$
- **Limiting ratio of $\mathcal{A}$:**

$$\mu_n(\mathcal{A}) = \lim_{m\to\infty} \frac{\#\mathcal{A}(m)}{\#\mathcal{T}_n(m)}, \text{ if this limit exists.}$$

[Lefman, Savický. *Some typical properties of large And/Or boolean formulas*, 1997]
[Chauvin, Flajolet, Gardy, Gittenberger. *And/Or trees revisited*, 2004]

Probability distribution

- $\mathcal{A}$ a subset of $\mathcal{T}_n$
- $\mathcal{A}(m)$ the set of all formulas of size m in $\mathcal{A}$
- **Limiting ratio of $\mathcal{A}$:**

$$\mu_n(\mathcal{A}) = \lim_{m\to\infty} \frac{\#\mathcal{A}(m)}{\#\mathcal{T}_n(m)}, \text{ if this limit exists.}$$

- Let $f \in \mathcal{F}_n$ and $\mathcal{A}_f$ the set of all its formulas:

$$\mathbb{P}_n(f) = \mu_n(A_f).$$

[Lefman, Savický. *Some typical properties of large And/Or boolean formulas*, 1997]

[Chauvin, Flajolet, Gardy, Gittenberger. *And/Or trees revisited*, 2004]

Probability distribution

- $\mathcal{A}$ a subset of $\mathcal{T}_n$
- $\mathcal{A}(m)$ the set of all formulas of size m in $\mathcal{A}$
- **Limiting ratio of $\mathcal{A}$:**

$$\mu_n(\mathcal{A}) = \lim_{m\to\infty} \frac{\#\mathcal{A}(m)}{\#\mathcal{T}_n(m)}, \text{ if this limit exists.}$$

- Let $f \in \mathcal{F}_n$ and $\mathcal{A}_f$ the set of all its formulas:

$$\mathbb{P}_n(f) = \mu_n(A_f).$$

- Using Drmota-Lalley-Woods Theorem:

$\mathbb{P}_n(\cdot)$ is a probability distribution on $\mathcal{F}_n$.

[Lefman, Savický. *Some typical properties of large And/Or boolean formulas*, 1997]

[Chauvin, Flajolet, Gardy, Gittenberger. *And/Or trees revisited*, 2004]

Previous results

Let $f \in \mathcal{F}_n$ be a function of complexity $L(f)$.

And/Or plane binary trees

$$\mathbb{P}_n(f) = \Theta\left(\frac{1}{n^{L(f)+1}}\right), \text{ when } n \text{ tends to infinity.}$$

[Kozik, 2008. *Subcritical pattern languages for and/or trees.*]

Previous results

Let $f \in \mathcal{F}_n$ be a function of complexity $L(f)$.

And/Or plane binary trees

$$\mathbb{P}_n(f) = \Theta\left(\frac{1}{n^{L(f)+1}}\right),$$ when n tends to infinity.

[Kozik, 2008. *Subcritical pattern languages for and/or trees.*]

Implication plane binary trees

$$\mathbb{P}_n(f) = \Theta\left(\frac{1}{n^{L(f)+1}}\right),$$ when n tends to infinity.

[Fournier, Gardy, G., Gittenberger, 2008. *Complexity and Limiting Ratio of Boolean Functions over Implication.*]

Previous results

In the Implication model,
G. and Gittenberger proved that functions of quadratic complexity (in n) have a non-negligible probability, when $n \to \infty$. [AofA'10]

Previous results

In the Implication model,
G. and Gittenberger proved that functions of quadratic complexity (in n) have a non-negligible probability, when $n \to \infty$. [AofA'10]

Theorem: [Lefman, Savický, 97]

And/Or plane binary trees:
Almost all functions are of polynomial complexity in n, when $n \to \infty$.

Theorem: [Shannon 42; Lupanov 62]

Uniform distribution on Boolean functions:
Almost all functions are of exponential complexity in n, when $n \to \infty$.

Previous results

In the Implication model,
G. and Gittenberger proved that functions of quadratic complexity (in n) have a non-negligible probability, when $n \to \infty$. [AofA'10]

Theorem: [Lefman, Savický, 97]

And/Or plane binary trees:
Almost all functions are of polynomial complexity in n, when $n \to \infty$.

Theorem: [Shannon 42; Lupanov 62]

Uniform distribution on Boolean functions:
Almost all functions are of exponential complexity in n, when $n \to \infty$.

Can we define a probability distribution that is neither entirely biased on functions of low complexity nor on functions on very large complexity ?

Previous results

In the Implication model,
G. and Gittenberger proved that functions of quadratic complexity (in n) have a non-negligible probability, when $n \to \infty$. [AofA'10]

Theorem: [Lefman, Savický, 97]

And/Or plane binary trees:
Almost all functions are of polynomial complexity in n, when $n \to \infty$.

Theorem: [Shannon 42; Lupanov 62]

Uniform distribution on Boolean functions:
Almost all functions are of exponential complexity in n, when $n \to \infty$.

Can we define a probability distribution that is neither entirely biased on functions of low complexity nor on functions on very large complexity ?

Remark: Formulas with symmetries have a smaller limiting ratio than others.

Natural properties of the connectors

Let us take into account some properties of the operators.
What about:

associativity or

commutativity of connectors *And* and *Or*.

Gardy introduced these models in 2006 and gave the probability of the function *True* when $n = 1$.

Natural properties of the connectors

Let us take into account some properties of the operators.
What about:

associativity or

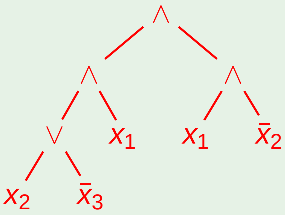
commutativity of connectors *And* and *Or*.

Gardy introduced these models in 2006 and gave the probability of the function *True* when $n = 1$.

- How does the tree structure evolve with such properties?
- Which behaviour for the new distributions on Boolean functions?

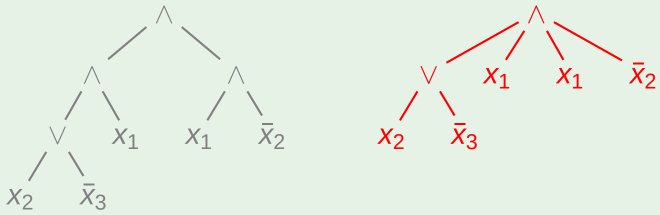
Tree models

connectors without associativity and commutativity	plane binary trees



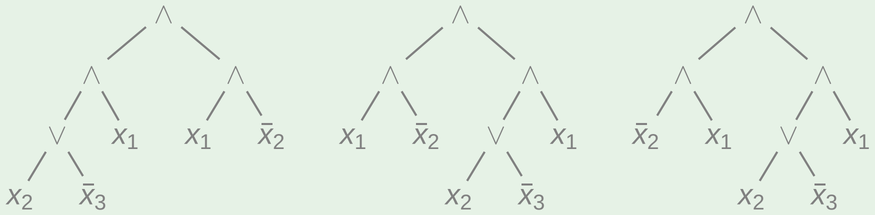
Tree models

connectors without associativity and commutativity	plane binary trees
associative connectors	plane non-binary trees



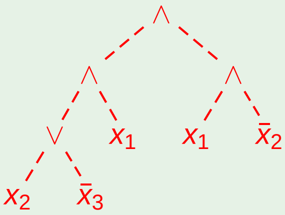
Tree models

connectors without associativity and commutativity	plane binary trees
associative connectors	plane non-binary trees
commutative connectors	non-plane binary trees



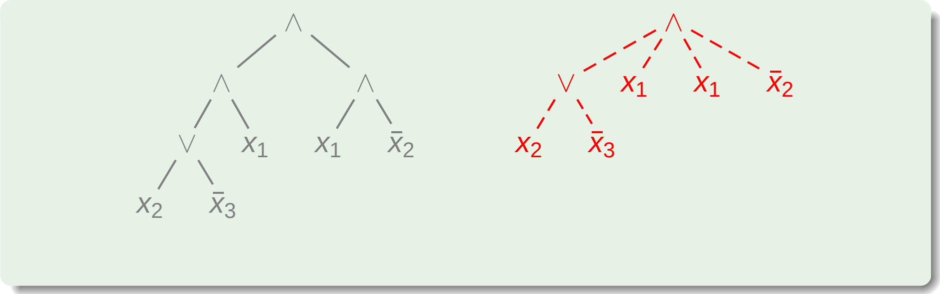
Tree models

connectors without associativity and commutativity	plane binary trees
associative connectors	plane non-binary trees
commutative connectors	non-plane binary trees



Tree models

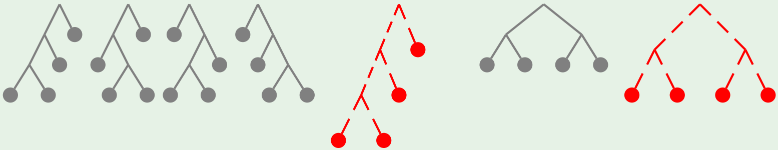
connectors without associativity and commutativity	plane binary trees
associative connectors	plane non-binary trees
commutative connectors	non-plane binary trees
associative commutative connectors	non-plane non-binary trees



Tree models

connectors without associativity and commutativity	plane binary trees
associative connectors	plane non-binary trees
commutative connectors	non-plane binary trees
associative commutative connectors	non-plane non-binary trees

Remark: The commutative models increase the limiting ratio of symmetric formulas.



Probability distributions on Boolean functions

Results

Let $f \in \mathcal{F}_n$ be a Boolean function.

In the distinct models (with or without associativity or commutativity),

- the complexity $L(f)$ is invariant

Results

Let $f \in \mathcal{F}_n$ be a Boolean function.

In the distinct models (with or without associativity or commutativity),

- the complexity $L(f)$ is invariant
- the probability of f exists

Results

Let $f \in \mathcal{F}_n$ be a Boolean function.

In the distinct models (with or without associativity or commutativity),

- the complexity $L(f)$ is invariant
- the probability of f exists

Theorem

Let us fix one of the four tree models, then

$$\mathbb{P}_n(f) = \frac{\lambda_f}{n^{L(f)+1}} + \mathcal{O}\left(\frac{1}{n^{L(f)+2}}\right), \text{ when } n \text{ tends to infinity.}$$

The constant λ_f is not the same from one model to another.

Results

Let $f \in \mathcal{F}_n$ be a Boolean function.

In the distinct models (with or without associativity or commutativity),

- the complexity $L(f)$ is invariant
- the probability of f exists

Theorem

Let us fix one of the four tree models, then

$$\mathbb{P}_n(f) = \frac{\lambda_f}{n^{L(f)+1}} + \mathcal{O}\left(\frac{1}{n^{L(f)+2}}\right), \text{ when } n \text{ tends to infinity.}$$

The constant λ_f is not the same from one model to another.

Corollary: Most trees computing f have a simple structure.

Key ideas [Kozik; Fournier *et al*]

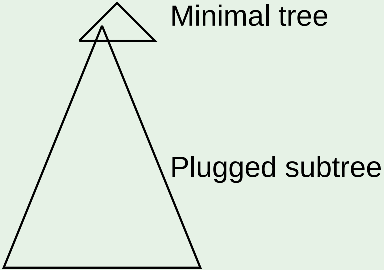
1. Definition of **simple family** of trees for f

Key ideas [Kozik; Fournier *et al*]

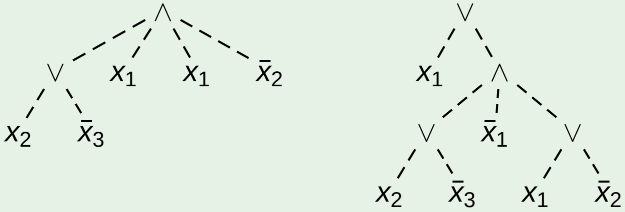
1. Definition of **simple family** of trees for f
2. All other trees for f:
   i. some “non-necessary” restrictions in their structures
   ii. their quantity is negligible

Key ideas [Kozik; Fournier *et al*]

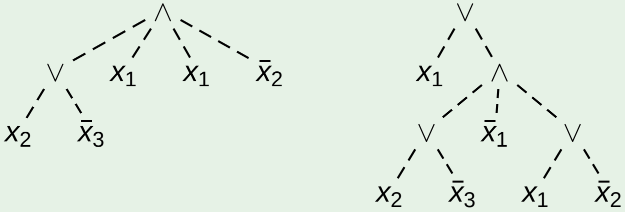
1. Definition of **simple family** of trees for f
2. All other trees for f:
   i. some “non-necessary” restrictions in their structures
   ii. their quantity is negligible
3. **Theorem:** Almost all trees computing the function f are obtained by plugging a special subtree in a minimal tree of f.



The associative and commutative model

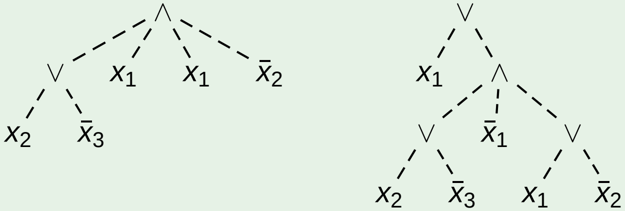


The associative and commutative model



$$\mathcal{T}^{\vee} = \mathcal{X} \mid \vee\, MSET^{\geq 2}(\mathcal{T}^{\wedge})$$

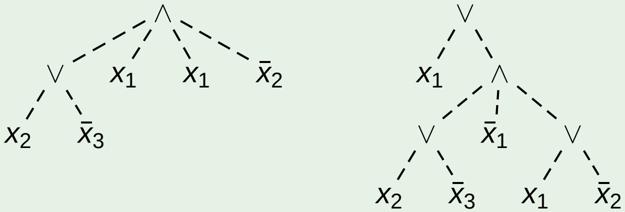
The associative and commutative model



$$\mathcal{T}^{\vee} = \mathcal{X} \mid \vee\, MSET^{\geq 2}(\mathcal{T}^{\wedge})$$

$$T^{\vee}(z) = 2nz + \exp\left(\sum_{i\geq 1}\frac{T^{\wedge}(z^i)}{i}\right) - 1 - T^{\wedge}(z),$$

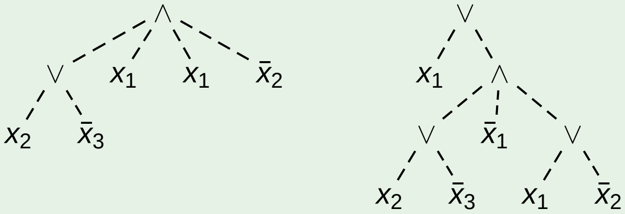
The associative and commutative model



$$\mathcal{T}^{\vee} = \mathcal{X} \mid \vee\, MSET^{\geq 2}(\mathcal{T}^{\wedge})$$

$$T^{\vee}(z) = 2nz + \exp\left(\sum_{i \geq 1} \frac{T^{\vee}(z^i)}{i}\right) - 1 - T^{\vee}(z),$$

The associative and commutative model



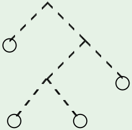
$$\mathcal{T}^{\vee} = \mathcal{X} \mid \vee\, MSET^{\geq 2}(\mathcal{T}^{\wedge})$$

$$T^{\vee}(z) = 2nz + \exp\left(\sum_{i\geq 1} \frac{T^{\vee}(z^i)}{i}\right) - 1 - T^{\vee}(z),$$

$$T(z) = 2T^{\vee}(z) - 2nz.$$

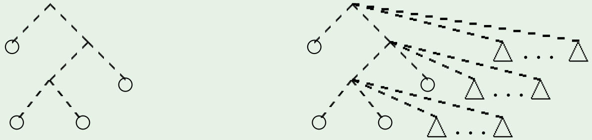
Main lemma

Let $\mathcal{V}$ be a fixed subset of literals and
A a tree whose leaves are labelled by $\mathcal{V}$:



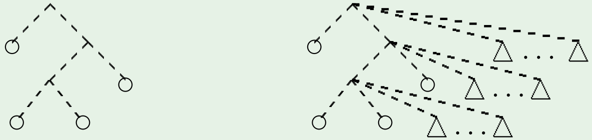
Main lemma

Let $\mathcal{V}$ be a fixed subset of literals and
A a tree whose leaves are labelled by $\mathcal{V}$:



Main lemma

Let $\mathcal{V}$ be a fixed subset of literals and
A a tree whose leaves are labelled by $\mathcal{V}$:



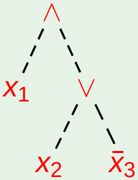
Lemma

$$\mu_n(\mathcal{F}_A) = \Theta\left(\frac{1}{n^{|A|}}\right),$$ when n tends to infinity.

The proof is adapted from Kozik's proof on plane binary trees.

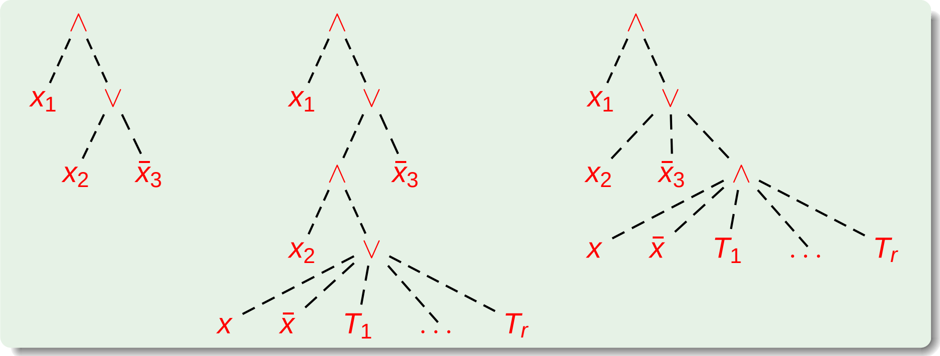
Valid expansions

Let f be the function computed by $x_1 \wedge (x_2 \vee x_3)$: $L(f) = 3$.



Valid expansions

Let f be the function computed by $x_1 \wedge (x_2 \vee x_3)$: $L(f) = 3$.



Bounds of the probability

Let f be the function computed by $x_1 \wedge (x_2 \vee x_3)$: $L(f) = 3$.

Four types of valid expansions applied on minimal trees of f give:

$$\mathbb{P}_n(f) \gtrsim_{n\to\infty} \frac{\lambda_f}{n^4}.$$

Bounds of the probability

Let f be the function computed by $x_1 \wedge (x_2 \vee x_3)$: $L(f) = 3$.

Four types of valid expansions applied on minimal trees of f give:

$$\mathbb{P}_n(f) \gtrsim_{n\to\infty} \frac{\lambda_f}{n^4}.$$

All other trees computing f belong to families containing
at least 5 restrictions.

Conclusion and perspectives

Conclusion

- Without closed forms for the generating functions of the families of trees we consider, we have obtained the same kind of results than with plane binary trees.

Conclusion

- Without closed forms for the generating functions of the families of trees we consider, we have obtained the same kind of results than with plane binary trees.
- Although the commutative models give more importance to symmetric formulas, it is not sufficient for modifying the asymptotic order of the probability of fixed functions.

Conclusion

- Without closed forms for the generating functions of the families of trees we consider, we have obtained the same kind of results than with plane binary trees.
- Although the commutative models give more importance to symmetric formulas, it is not sufficient for modifying the asymptotic order of the probability of fixed functions.
- According to these results on probabilities on fixed functions, we conjecture to have the same behaviour:
  Almost all functions are of low complexity.

Perspectives

To obtain other kind of probability distributions:

- Changing the size notion for associative model,

Perspectives

To obtain other kind of probability distributions:

- Changing the size notion for associative model,
- Using direct acyclic graphs instead of trees.