Dark matter among and/or trees

Jakub Kozik

Theoretical Computer Science
Jagiellonian University

6 June 2008
Versailles-Lyon-Chambery-Kraków Workshop on Computational Logic and Applications

and/or trees

- $\mathrm{Var} = \{x_1, x_2, x_3, \ldots\}$ - countable set of variables
- $\mathrm{Var}_k = \{x_1, \ldots, x_k\}$

Definition

$\mathcal{T}_k$ - set of and/or trees with k-variables is a minimal set such that:

1. $x_i, \neg x_i \in \mathcal{T}_k$, for each $x_i \in \mathrm{Var}_k$,
2. $(t_1 \wedge t_2), (t_1 \vee t2)$, for each $t_1, t_2 \in \mathcal{T}_k$.

$\mathcal{T}_k(n)$ - the set of elements of $\mathcal{T}_k$ of size n.

Observation

Every $t \in \mathcal{T}_k$ defines some function $I(t) : \{0,1\}^k \to \{0,1\}$.

$$(I : \mathcal{T}_k \to \{0,1\}^{\{0,1\}^k})$$

and/or trees

- $\mathrm{Var} = \{x_1, x_2, x_3, \dots\}$ - countable set of variables
- $\mathrm{Var}_k = \{x_1, \dots, x_k\}$

Definition

$\mathcal{T}_k$ - set of and/or trees with k-variables is a minimal set such that:

1. $x_i, \neg x_i \in \mathcal{T}_k$, for each $x_i \in \mathrm{Var}_k$,
2. $(t_1 \wedge t_2), (t_1 \vee t2)$, for each $t_1, t_2 \in \mathcal{T}_k$.

$\mathcal{T}_k(n)$ - the set of elements of $\mathcal{T}_k$ of size n.

Observation

Every $t \in \mathcal{T}_k$ defines some function $I(t) : \{0,1\}^k \to \{0,1\}$.

$$(I : \mathcal{T}_k \to \{0,1\}^{\{0,1\}^k})$$

and/or trees

- $\mathrm{Var} = \{x_1, x_2, x_3, \ldots\}$ - countable set of variables
- $\mathrm{Var}_k = \{x_1, \ldots, x_k\}$

Definition

$\mathcal{T}_k$ - set of and/or trees with k-variables is a minimal set such that:

1. $x_i, \neg x_i \in \mathcal{T}_k$, for each $x_i \in \mathrm{Var}_k$,
2. $(t_1 \wedge t_2), (t_1 \vee t2)$, for each $t_1, t_2 \in \mathcal{T}_k$.

$\mathcal{T}_k(n)$ - the set of elements of $\mathcal{T}_k$ of size n.

Observation

Every $t \in \mathcal{T}_k$ defines some function $I(t) : \{0,1\}^k \to \{0,1\}$.

$$(I : \mathcal{T}_k \to \{0,1\}^{\{0,1\}^k})$$

and/or trees

- $\mathrm{Var} = \{x_1, x_2, x_3, \dots\}$ - countable set of variables
- $\mathrm{Var}_k = \{x_1, \dots, x_k\}$

Definition

$\mathcal{T}_k$ - set of and/or trees with k-variables is a minimal set such that:

1. $x_i, \neg x_i \in \mathcal{T}_k$, for each $x_i \in \mathrm{Var}_k$,
2. $(t_1 \wedge t_2), (t_1 \vee t2)$, for each $t_1, t_2 \in \mathcal{T}_k$.

$\mathcal{T}_k(n)$ - the set of elements of $\mathcal{T}_k$ of size n.

Observation

Every $t \in \mathcal{T}_k$ defines some function $I(t) : \{0,1\}^k \to \{0,1\}$.

$$(I : \mathcal{T}_k \to \{0,1\}^{\{0,1\}^k})$$

and/or trees

- $\mathrm{Var} = \{x_1, x_2, x_3, \dots\}$ - countable set of variables
- $\mathrm{Var}_k = \{x_1, \dots, x_k\}$

Definition

$\mathcal{T}_k$ - set of and/or trees with k-variables is a minimal set such that:

1. $x_i, \neg x_i \in \mathcal{T}_k$, for each $x_i \in \mathrm{Var}_k$,
2. $(t_1 \wedge t_2), (t_1 \vee t2)$, for each $t_1, t_2 \in \mathcal{T}_k$.

$\mathcal{T}_k(n)$ - the set of elements of $\mathcal{T}_k$ of size n.

Observation

Every $t \in \mathcal{T}_k$ defines some function $I(t) : \{0,1\}^k \to \{0,1\}$.

$$(I : \mathcal{T}_k \to \{0,1\}^{\{0,1\}^k})$$

and/or trees

- $\mathrm{Var} = \{x_1, x_2, x_3, \dots\}$ - countable set of variables
- $\mathrm{Var}_k = \{x_1, \dots, x_k\}$

Definition

$\mathcal{T}_k$ - set of and/or trees with k-variables is a minimal set such that:

1. $x_i, \neg x_i \in \mathcal{T}_k$, for each $x_i \in \mathrm{Var}_k$,
2. $(t_1 \wedge t_2), (t_1 \vee t2)$, for each $t_1, t_2 \in \mathcal{T}_k$.

$\mathcal{T}_k(n)$ - the set of elements of $\mathcal{T}_k$ of size n.

Observation

Every $t \in \mathcal{T}_k$ *defines some function* $I(t) : \{0,1\}^k \rightarrow \{0,1\}$.

$$(I : \mathcal{T}_k \rightarrow \{0,1\}^{\{0,1\}^k})$$

Distribution $P_k[n]$

Fix $k, n \in \mathbb{N}$.

$u_{\mathcal{T}_k(n)}$ - uniform probability distribution on the (finite) set $\mathcal{T}_k(n)$.

FOR

$$f : 0,1^k \to 0,1$$

PUT

$$P_k[n](f) = u_{\mathcal{T}_k(n)}(I^{-1}(f)).$$

$P_k[n]$ is a probability distribution on $\{0,1\}^{\{0,1\}^k}$.

Distribution $P_k[n]$

Fix $k, n \in \mathbb{N}$.

$u_{\mathcal{T}_k(n)}$ - uniform probability distribution on the (finite) set $\mathcal{T}_k(n)$.

FOR

$$f : 0, 1^k \to 0, 1$$

PUT

$$P_k[n](f) = u_{\mathcal{T}_k(n)}(I^{-1}(f)).$$

$P_k[n]$ is a probability distribution on $\{0,1\}^{\{0,1\}^k}$.

Distribution $P_k[n]$

Fix $k, n \in \mathbb{N}$.

$u_{\mathcal{T}_k(n)}$ - uniform probability distribution on the (finite) set $\mathcal{T}_k(n)$.

FOR

$$f : 0, 1^k \to 0, 1$$

PUT

$$P_k[n](f) = u_{\mathcal{T}_k(n)}(I^{-1}(f)).$$

$P_k[n]$ is a probability distribution on $\{0, 1\}^{\{0,1\}^k}$.

Distribution $P_k[n]$

Fix $k, n \in \mathbb{N}$.

$u_{\mathcal{T}_k(n)}$ - uniform probability distribution on the (finite) set $\mathcal{T}_k(n)$.

FOR

$$f : 0, 1^k \to 0, 1$$

PUT

$$P_k[n](f) = u_{\mathcal{T}_k(n)}(I^{-1}(f)).$$

$P_k[n]$ is a probability distribution on $\{0, 1\}^{\{0,1\}^k}$.

Theorem (Lefmann,Savicki 1997)

There exists distribution P_k *on the set* $\{0,1\}^{\{0,1\}^k}$ *such that:*

$$P_k[n](f) \rightarrow_{n\rightarrow\infty} P_k(f)$$

for every $f \in \{0,1\}^{\{0,1\}^k}$.

Equivalently

$$P_k(f) = \lim_{n\rightarrow\infty} \frac{Card((I^{-1}(f) \cap \mathcal{T}_k(n)))}{Card(\mathcal{T}_k(n))}$$

Theorem (Lefmann,Savicki 1997)

There exists distribution P_k *on the set* $\{0,1\}^{\{0,1\}^k}$ *such that:*

$$P_k[n](f) \to_{n\to\infty} P_k(f)$$

for every $f \in \{0,1\}^{\{0,1\}^k}$.

Equivalently

$$P_k(f) = \lim_{n\to\infty} \frac{Card((I^{-1}(f) \cap \mathcal{T}_k(n)))}{Card(\mathcal{T}_k(n))}$$

Bounds for distribution P_k

Non constant f.

Upper bound - B. Chauvin, P. Flajolet, D. Gardy, and B. Gittenberger 2004

$$P_k(f) \leq (1 + O(1/k))exp(-c\frac{L(f)}{k^2})$$

for some constant $c > 0$.

Lower bound - Lefmann,Savicki 1997

$$\tfrac{1}{4}(8k)^{-L(f)-1} \leq P_k(f)$$

Improved lower bound - D. Gady, A. Woods 2005

$$\frac{4m(f)}{(16k)^{L(f)+1}} \leq P_k(f)$$

where $m(f)$ the number of minimal trees for f.

Bounds for distribution P_k

Non constant f.

Upper bound - B. Chauvin, P. Flajolet, D. Gardy, and B. Gittenberger 2004

$$P_k(f) \leq (1 + O(1/k))exp(-c\frac{L(f)}{k^2})$$

for some constant $c > 0$.

Lower bound - Lefmann,Savicki 1997

$$\frac{1}{4}(8k)^{-L(f)-1} \leq P_k(f)$$

Improved lower bound - D. Gady, A. Woods 2005

$$\frac{4m(f)}{(16k)^{L(f)+1}} \leq P_k(f)$$

where $m(f)$ the number of minimal trees for f.

Bounds for distribution P_k

Non constant f.

Upper bound - B. Chauvin, P. Flajolet, D. Gardy, and B. Gittenberger 2004

$$P_k(f) \leq (1 + O(1/k))exp(-c\frac{L(f)}{k^2})$$

for some constant $c > 0$.

Lower bound - Lefmann,Savicki 1997

$$\tfrac{1}{4}(8k)^{-L(f)-1} \leq P_k(f)$$

Improved lower bound - D. Gady, A. Woods 2005

$$\frac{4m(f)}{(16k)^{L(f)+1}} \leq P_k(f)$$

where $m(f)$ the number of minimal trees for f.

Bounds for distribution P_k

Non constant f.

Upper bound - B. Chauvin, P. Flajolet, D. Gardy, and B. Gittenberger 2004

$$P_k(f) \leq (1 + O(1/k))exp(-c\frac{L(f)}{k^2})$$

for some constant $c > 0$.

Lower bound - Lefmann,Savicki 1997

$$\frac{1}{4}(8k)^{-L(f)-1} \leq P_k(f)$$

Improved lower bound - D. Gady, A. Woods 2005

$$\frac{4m(f)}{(16k)^{L(f)+1}} \leq P_k(f)$$

where $m(f)$ the number of minimal trees for f.

Focus on low complexities.

Theorem (J.K. 2007)

f - non-constant function. There exists positive constant B_f such that

$$P_k(f) \sim_k B_f k^{-L(f)-1}$$

Corollary

Lower bound is quite good for the functions with very low complexity.

D. Gardy, A. Woods conjecture

Focus on low complexities.

Theorem (J.K. 2007)

f - non-constant function. There exists positive constant B_f such that

$$P_k(f) \sim_k B_f k^{-L(f)-1}$$

Corollary

Lower bound is quite good for the functions with very low complexity.

D. Gardy, A. Woods conjecture

Focus on low complexities.

Theorem (J.K. 2007)

f - non-constant function. There exists positive constant B_f such that

$$P_k(f) \sim_k B_f k^{-L(f)-1}$$

Corollary

Lower bound is quite good for the functions with very low complexity.

What next? Big complexity ?

Uniform counting result

A fraction tending to 1 (as $k \to \infty$) of boolean functions in k variables have tree complexity at least $2^k/logk$.

Guess

The set of functions with high complexity should have a big measure $P_k(\_)$.

What next? Big complexity ?

Uniform counting result

A fraction tending to 1 (as $k \to \infty$) of boolean functions in k variables have tree complexity at least $2^k/logk$.

Guess

The set of functions with high complexity should have a big measure $P_k(\_)$.

What next? Big complexity ?

Uniform counting result

A fraction tending to 1 (as $k \to \infty$) of boolean functions in k variables have tree complexity at least $2^k / logk$.

Guess

The set of functions with high complexity should have a big measure $P_k(\_)$.

But... Why the result for low complexities hold?

Definition (Negative pattern)

$$N \to N \vee N \,\Big|\, \square \wedge N \,\Big|\, \bullet$$

Definition (Positive pattern)

$$P \to \square \vee P \,\Big|\, P \wedge P \,\Big|\, \bullet$$

But... Why the result for low complexities hold?

Definition (Negative pattern)

$$N \to N \vee N \,\Big|\, \Box \wedge N \,\Big|\, \bullet$$

Definition (Positive pattern)

$$P \to \Box \vee P \,\Big|\, P \wedge P \,\Big|\, \bullet$$

But... Why the result for low complexities hold?

Definition (Negative pattern)

$$N \rightarrow N \vee N \Big| \Box \wedge N \Big| \bullet$$

Definition (Positive pattern)

$$P \rightarrow \Box \vee P \Big| P \wedge P \Big| \bullet$$

Positive leaves number distribution.

Distribution of numbers of positive leaves

$$Prob(\text{ tree of size } n \text{ has } m \text{ positive leaves }) \to_{n\to\infty} C_{m-1}(2m-1)(\frac{2}{9})^m$$

Corollary

Most of trees have small number od positive leaves.

Corollary

In most of and/or trees the variables in the positive leaves are all different.

Positive leaves number distribution.

Distribution of numbers of positive leaves

$$Prob(\text{ tree of size } n \text{ has } m \text{ positive leaves }) \to_{n\to\infty} C_{m-1}(2m-1)(\frac{2}{9})^m$$

Corollary

Most of trees have small number od positive leaves.

Corollary

In most of and/or trees the variables in the positive leaves are all different.

Positive leaves number distribution.

Distribution of numbers of positive leaves

$$Prob(\text{ tree of size } n \text{ has } m \text{ positive leaves }) \to_{n\to\infty} C_{m-1}(2m-1)(\frac{2}{9})^m$$

Corollary

Most of trees have small number od positive leaves.

Corollary

In most of and/or trees the variables in the positive leaves are all different.

On the other hand...

IF a tree t has :

- m positive leaves
- all variables in positive leaves different

THEN

the table for $I(t)$ contains hypercube of dimension $n - m$ filled with 1.

Observation

Big costant hypercubes are quite rare among boolean functions.

On the other hand...

IF a tree t has :

- m positive leaves
- all variables in positive leaves different

THEN

the table for $I(t)$ contains hypercube of dimension $n - m$ filled with 1.

Observation

Big costant hypercubes are quite rare among boolean functions.

On the other hand...

IF a tree t has :

- m positive leaves
- all variables in positive leaves different

THEN

the table for $I(t)$ contains hypercube of dimension $n - m$ filled with 1.

Observation

Big costant hypercubes are quite rare among boolean functions.

The family

F_m - all the functions definable by trees with at most m positive leaves.

Consider

$$G_k = F_{\log k} \cap \{0,1\}^{\{0,1\}^k}$$

THEN

$$P(G_{\log k}) \to_{k \to \infty} 1$$

The family

F_m - all the functions definable by trees with at most m positive leaves.

Consider

$$G_k = F_{\log k} \cap \{0,1\}^{\{0,1\}^k}$$

THEN

$$P(G_{\log k}) \to_{k\to\infty} 1$$

The family

F_m - all the functions definable by trees with at most m positive leaves.

Consider

$$G_k = F_{\log k} \cap \{0,1\}^{\{0,1\}^k}$$

THEN

$$P(G_{\log k}) \to_{k \to \infty} 1$$

Functions from $F_{\log k}$ can not have maximal possible complexity.

Constribution from the functions with low complexity is negligible.

Functions from $F_{\log k}$ can not have maximal possible complexity.

Constribution from the functions with low complexity is negligible.

Corollary

There exist functions whose complexity is neither small nor heigh which are dominant for the distributrion P.

What functions are they?

What complexity do they have?

Corollary

There exist functions whose complexity is neither small nor heigh which are dominant for the distributrion P.

What functions are they?

What complexity do they have?

Corollary

There exist functions whose complexity is neither small nor heigh which are dominant for the distributtion P.

What functions are they?

What complexity do they have?