Asymptotic probability of Boolean functions over implication

Danièle Gardy, Univ. Versailles

with

Hervé Fournier and Antoine Genitrini, Univ. Versailles

Bernhard Gittenberger, T.U. Wien

April 2008
Outline

- Boolean expressions and trees
- A restricted propositional calculus
- Tautologies
- Probability and complexity of a Boolean function
- Main result: sketch of proof
- Extensions and open questions
Boolean expressions

$\left( (x \lor \bar{x}) \land x \right) \land \left( \bar{x} \lor (x \lor \bar{x}) \right)$

$\left( x \lor (y \land \bar{x}) \right) \lor \left( ((z \land \bar{y}) \lor (x \lor \bar{u})) \land (x \lor y) \right)$
Boolean expressions

\[
((x \lor \bar{x}) \land x) \land (\bar{x} \lor (x \lor \bar{x})) \\
(x \lor (y \land \bar{x})) \lor (((z \land \bar{y}) \lor (x \lor \bar{u})) \land (x \lor y))
\]

Probability that a “random” expression on \( n \) boolean variables is a tautology (always true)?
Boolean expressions

\[ ((x \lor \bar{x}) \land x) \land (\bar{x} \lor (x \lor \bar{x})) \]
\[ (x \lor (y \land \bar{x})) \lor (((z \land \bar{y}) \lor (x \lor \bar{u})) \land (x \lor y)) \]

Probability that a “random” expression on \( n \) boolean variables is a tautology (always true)?

- \( n = 1 \): 4 boolean functions; \( \text{Proba}(\text{True}) = 0.2886 \)
- \( n = 2 \): 16 boolean functions; \( \text{Proba}(\text{True}) = 0.209 \)
- \( n = 3 \): 256 boolean functions; \( \text{Proba}(\text{True}) = 0.165 \)
Boolean expressions

\[(x \lor \bar{x}) \land (\bar{x} \lor (x \lor \bar{x}))\]
\[(x \lor (y \land \bar{x})) \lor (((z \land \bar{y}) \lor (x \lor \bar{u})) \land (x \lor y))\]

Probability that a “random” expression on \(n\) boolean variables is a tautology (always true)?

- \(n = 1\): 4 boolean functions; \(\text{Proba}(True) = 0.2886\)
- \(n = 2\): 16 boolean functions; \(\text{Proba}(True) = 0.209\)
- \(n = 3\): 256 boolean functions; \(\text{Proba}(True) = 0.165\)
- \(n \to +\infty\): \(2^n\) boolean functions

\(\text{Proba}(True) \sim?\)
Boolean expressions

\[
((x \lor \overline{x}) \land x) \land (\overline{x} \lor (x \lor \overline{x})) \\
(x \lor (y \land \overline{x})) \lor (((z \land \overline{y}) \lor (x \lor \overline{u})) \land (x \lor y))
\]

Probability that a “random” expression on \( n \) boolean variables is a tautology (always true)?

- \( n = 1 \): 4 boolean functions; \( \text{Proba}(True) = 0.2886 \)
- \( n = 2 \): 16 boolean functions; \( \text{Proba}(True) = 0.209 \)
- \( n = 3 \): 256 boolean functions; \( \text{Proba}(True) = 0.165 \)
- \( n \rightarrow +\infty \): \( 2^{2^n} \) boolean functions

\[ \text{Proba}(True) \sim? \]

\( \text{Proba}(f) \) for any boolean function \( f \)?
Boolean expressions and trees

\[(x \lor \bar{x}) \land x) \land (\bar{x} \lor (x \lor \bar{x}))\]

Consider a well-formed boolean expression

- Choose set of logical connectors, with arities
  \(\rightarrow\) Choose labels and arities for internal nodes

- Choose set of boolean literals for the leaves
  \(\rightarrow\) Choose labels for leaves
Boolean expressions and trees

- Expression $\sim$ labelled tree
- Random expression $\sim$ random labelled tree
- What notion of randomness on trees?
  - Choose size $m$ of the tree; assume all trees of same size are equiprob.
    Then let $m \to +\infty$
  - Choose tree at random (e.g., by a branching process): size is also random. Then label tree at random.
Boolean expressions and trees

- Expression $\sim$ labelled tree
- Random expression $\sim$ random labelled tree
- Two notions of randomness on trees/boolean expressions
- Each boolean expression computes a boolean function
- A boolean function is represented by an infinite number of expressions
- Can we use random boolean expressions to define a probability distribution on boolean functions?
Former work: And/Or trees

- One of the most studied models for random boolean expressions
- Binary trees; no simple node
- Internal nodes are labelled by $\lor$ or $\land$
- Leaves are labelled by the literals: $x_1, \ldots, x_n, \overline{x}_1, \ldots, \overline{x}_n$
And/Or trees

- Paris et al. 94: first definition of a tree distribution on boolean functions
- Lefman and Savicky 97:
  - Proof of existence of a tree distribution (by pruning)
  - Tree complexity of $f$: $L(f) = \text{size of smallest tree that computes } f$
  - $\frac{1}{4} \left(\frac{1}{8n}\right)^{L(f)} \leq P(f) \leq e^{-cL(f)/n^3} \left(1 + O(1/n)\right)$
- Chauvin et al. 04: alternative definition of probability by generating functions; improvement on upper bound: $P(f) \leq e^{-cL(f)/n^2} \left(1 + O(1/n)\right)$
- For tautologies:
  - Woods 05: Asymptotic probability $P(True) \sim 1/4n$ and probable shape of tautologies: $l \lor \ldots \lor \overline{l} \lor \ldots$
  - Kozik 08: Alternative derivation of asymptotic probability and shape
And/Or trees: probability and complexity

To sum up:

- definition of a tree-induced probability distribution on boolean functions
- probability of constant functions $True$ and $False$: known
- probability of a non-constant function:
  - lower bound $(1/4) (8n)^{-L(f)}$ (not that bad; order looks right)
  - upper bound $e^{-cL(f)/n^2} (1 + O(1/n))$ (probably not tight)
And/Or trees: probability and complexity

To sum up:

• definition of a tree-induced probability distribution on boolean functions
• probability of constant functions *True* and *False*: known
• probability of a non-constant function:
  – lower bound \((1/4) (8n)^{-L(f)}\) (not that bad; order looks right)
  – upper bound \(e^{-cL(f)/n^2} (1 + O(1/n))\) (probably not tight)

• Partial results. Can we go further?
And/Or trees: probability and complexity

To sum up:

- definition of a tree-induced probability distribution on boolean functions
- probability of constant functions *True* and *False*: known
- probability of a non-constant function:
  - lower bound $(1/4) (8n)^{-L(f)}$ (not that bad; order looks right)
  - upper bound $e^{-cL(f)/n^2} (1 + O(1/n))$ (probably not tight)

- Partial results. Can we go further?
- Consider a simpler system
A restricted propositional calculus

- Finite number of boolean variables: \( x_1, x_2, \ldots, x_n \); no negative literals.
- A single connector \( \rightarrow (x_1 \rightarrow x_2 \text{ is also } \overline{x_1} \lor x_2) \).
- Expressions are binary trees: \( (x \rightarrow y) \rightarrow (x \rightarrow (z \rightarrow u) \rightarrow t) \)
A restricted propositional calculus

- Finite number of boolean variables: \(x_1, x_2, \ldots, x_n\); no negative literals.
- A single connector \(\rightarrow (x_1 \rightarrow x_2\) is also \(\overline{x_1} \lor x_2\).
- Expressions are binary trees: \((x \rightarrow y) \rightarrow (x \rightarrow (z \rightarrow u) \rightarrow t)\)

An expression is a (possibly empty) sequence of expressions: premises, followed by a variable: goal.
A restricted propositional calculus

• Finite number of boolean variables: $x_1, x_2, \ldots, x_n$; no negative literals.

• A single connector $\rightarrow$

• “Simple” system: may hope for a detailed study of random expressions and boolean functions.

• Relevance to intuitionistic logic:

  Tautology $\sim$ proof of a goal from premises
Boolean functions and expressions

An expression (a tree) computes a boolean function on $k$ variables.

- What is the set of boolean functions that can be computed?

$\Rightarrow$ Post set $S_0 = \{ x \lor g(x_1, \ldots, x_k) \}$
Boolean functions and expressions

An expression (a tree) computes a boolean function on \( k \) variables.

- What is the set of boolean functions that can be computed?

  \[ S_0 = \{ x \lor g(x_1, \ldots, x_k) \} \]

- Many different expressions compute the same boolean function.

  Probability that a “random” expression computes a specific function?
Probability of a boolean function

- Informally, it is the ratio of trees that compute $f$ to the total number of trees (assuming this ratio can be defined).
- Define the size of a formula (tree) as the number of variable occurrences (leaves).
- Define $A_m = \{\text{trees of size } m\}$; $A_m(f) = \{\text{trees in } A_m \text{ that compute } f\}$. Assume a uniform distribution on $A_m$.
- Probability that a tree of size $m$ computes $f$:

$$P_m(f) = \frac{|A_m(f)|}{|A_m|}$$

- For any boolean function $f$, $\lim_{m \to +\infty} P_m(f)$ exists?
Probability of a boolean function

Existence of a limit $P(f) = \lim_{m \to +\infty} P_m(f)$?

- Enumerate trees by size: g.f. $\Phi(z) = \sum_m |A_m|z^m = (1 - \sqrt{1 - 4nz})/2$
- Enumerate the set $A(f)$ of trees computing a specific function $f$:
  Generating function $\phi_f(z)$?
  Consider all boolean functions
  $A(f) = \bigcup_{g,h} (A(g), \to, A(h)) \Rightarrow \phi_f = \sum_{g,h} \phi_g \phi_h$
  $\Rightarrow$ write a system of algebraic equations for the enumerating functions
  $\Rightarrow$ Drmota-Lalley-Woods theorem gives asymptotics of $[z^m]\phi_f(z)$

- Putting all this together proves the existence of the prob. distribution $P$

For any boolean function $f$, we compute

$$P(f) = \lim_{m \to +\infty} \frac{[z^m]\phi_f(z)}{[z^m]\Phi(z)}$$
Probability of a boolean function

- We have proved the existence of $P(f)$ for any $f$
  
  $(f \not\in S_0: P(f) = 0)$

- *Can we compute explicitly the probability of a boolean function?*
Probability of a boolean function

- We have proved the existence of \( P(f) \) for any \( f \)
  
  \[ (f \notin S_0: P(f) = 0) \]

- Can we compute explicitly the probability of a boolean function?

- The complexity of a function \( f \) is the smallest size of a tree that computes \( f \).

- What is the relation between the complexity and the probability of a boolean function?

- What is the typical shape of a tree that computes a specific function?

- What is the average complexity of a random boolean function?
Tautologies

We begin with the simplest function: the constant $True$

- **Simple** tautology: a premise is equal to the goal.
- We know the probability of simple tautologies:
  \[
  \frac{4n + 1}{(2n + 1)^2} \sim \frac{1}{n}
  \]

- Almost all tautologies are simple (Fournier et al. 07)
- Hence $P(True) \sim 1/n$

- Consequence: almost all tautologies in the system of implication and positive literals are intuitionnistic tautologies.
Probability of boolean functions

We know a.s. the shape of a random tautology.

We can compute the probability of True.

*Can we extend this to a non-constant boolean function $f$?*
Probability of boolean functions

- True: $1/n + O(1/n^2)$
- Literal $x$: $1/2n^2 + O(1/n^3)$
- Function $x \rightarrow y$: $9/16n^3 + O(1/n^4)$
- For all $f \in S_0 \setminus \{1\}$:

$$P(f) = \frac{\lambda(f)}{4^{L(f)}n^{L(f)+1}} (1 + O(1/n))$$

- $\lambda(f)$ is related to the minimal trees for $f$
- The trees of $A(f)$ are simple: a.s. obtained from a minimal tree by a single expansion
Sketch of proof

- Start from the set of minimal trees that compute $f$.
- Define extension rules: we obtain a larger (infinite) set of trees, still computing $f$; we can compute the probability of this set.
- Probability of this new set is related to the sizes of the initial trees, i.e. to the tree complexity of $f$.
- Do we obtain a.s. all the trees that compute $f$?
- If so, we know the probability of $f$, and we can express it in terms of its complexity.
Extensions of minimal trees

Consider a tree $A$ that computes $f$, and a node of $A$

When can we expand a node of $A$, and still get a tree that computes $f$?
Extensions of minimal trees: example

$f = x_1 \rightarrow x_2$ has a unique minimal tree $A_{min}$:

- $E$ is a tautology
- $E$ has goal $x_1$
- $E$ has a premise $x_2$
Extensions of minimal trees: example

$f = x_1 \rightarrow x_2$ has a unique minimal tree $A_{\text{min}}$:

- $E$ is a tautology
- $E$ has goal $x_2$
- $E$ has a premise $x_1$
Extensions of minimal trees: example

\( f = x_1 \rightarrow x_2 \) has a unique minimal tree \( A_{min} \):

\[
\begin{array}{c}
\rightarrow \\
\wedge \\
\ \ x_1 \\
\ \ \ x_2 \\
\Rightarrow \\
\rightarrow \\
\wedge \\
\ \ x_1 \\
\ \ \ E \\
\ \ \ x_2
\end{array}
\]

- \( E \) is a tautology
- \( E \) has goal \( x_1 \)
- \( E \) has a premise \( x_2 \)
Extensions of minimal trees: example

$f = x_1 \rightarrow x_2$ has a unique minimal tree $A_{\text{min}}$

• Nine possible types of expansion ⇒ set $\mathcal{E}(A_{\text{min}})$ of trees computing $f$
• We can compute the probability of $\mathcal{E}(A_{\text{min}})$:

$$\frac{9}{16n^3} + O \left( \frac{1}{n^4} \right)$$

• This is the probability of $f$
Extensions of minimal trees

- Define extensions for minimal trees
- Compute probability of the set $\mathcal{E}(f)$ obtained by one extension
- Compute probability of the set $\mathcal{E}^+(f)$ obtained by a finite number of extensions
- Compute probability of $A(f) \setminus \mathcal{E}^+(f)$:
  - Define pruning rules: inverses of expansion rules
  - Any tree of $A(f)$ can be pruned into an irreducible tree
  - $\{\text{Minimal trees}\} \subset \{\text{Irreducible trees}\}$
  - Almost all trees of $f$ can be pruned into irreducible trees.
Probability of a boolean function \( f \)

- Expression of the probability

\[
P(f) = \frac{\lambda(f)}{4L(f)nL(f)+1} \left(1 + O(1/n)\right)
\]

- We obtain almost all the trees by a single expansion of a minimal tree

\[
P(f) = Proba(\mathcal{E}(f)) \left(1 + o(1)\right)
\]

- The number of possible expansions is related to properties of minimal trees:

  - \( m \) = number of minimal trees for \( f \)
  - \( e \) = number of essential variables of \( f \)

Then

\[
2(2m - 1)L(f) \leq \lambda(f) \leq (1 + 2e)(2L(f) - 1)m
\]
Possible extensions

- Computation of the constant factor $\lambda(f)$?
  
  Done for read-once functions; for other functions?

- Result can be adapted when trees are obtained by a growing process

- What if we allow negative literals?

- What if we choose a different set of connectors?
Average complexity of a boolean function

- For a uniform distribution on boolean functions, maximal and average tree complexity is $2^k / \log k$ (Shannon, Lupanov...)

- What if the distribution is not uniform? for example, a tree distribution?

- We have computed the probability of a boolean function of known (hence, “fixed, small” and independent of $k$) complexity.

- What about the probability of a function of “large” (dependent on $k$) complexity?