

A probabilistic approach to the density of tautologies

Taehyun Eom

GIST

2026.10.29

Propositional Logic System

Language	Logic
Syntax Semantics Syntactics	Well-formed formulae Valuations Proofs
Semantic Truth Syntactic Entailment	Tautologies Theorems

Well-formed formulae

- Variables

$x_0, x_1, \dots,$

$\overline{x_0}, \overline{x_1}, \dots,$

$\perp, \top$

$x, x1, x11, x111, \dots$

Well-formed formulae

- Variables

$x_0, x_1, \dots,$

$\overline{x_0}, \overline{x_1}, \dots,$

$\perp, \top$

$x, x1, x11, x111, \dots$

Well-formed formulae

- Variables

$x_0, x_1, \dots,$

$\overline{x_0}, \overline{x_1}, \dots,$

$\perp, \top$

$x, x1, x11, x111, \dots$

- Logical operators (Connectives)

$\neg$

$\rightarrow, \vee, \wedge, |, \dots$

Well-formed formulae

- Variables

$x_0, x_1, \dots,$

$\overline{x_0}, \overline{x_1}, \dots,$

$\perp, \top$

$x, x1, x11, x111, \dots$

- Logical operators (Connectives)

$\neg$

$\rightarrow, \vee, \wedge, |, \dots$

Number of Well-formed formulae

If we have m variables, $\neg$, $\rightarrow$, there exists

$\neg\neg\neg\neg X$ type : m

$\neg\neg[X \rightarrow Y]$ type : m^2

$\neg[\neg X \rightarrow Y]$ type : m^2

$\neg[X \rightarrow \neg Y]$ type : m^2

$[\neg X \rightarrow \neg Y]$ type : m^2

$[\neg\neg X \rightarrow Y]$ type : m^2

$[X \rightarrow \neg\neg Y]$ type : m^2

$X \rightarrow [Y \rightarrow Z]$ type : m^3

$[X \rightarrow Y] \rightarrow Z$ type : m^3

$2m^3 + 6m^2 + m$ well-formed formulae of length 5.

Valuations

Truth table

ϕ	ψ	$\perp$	$\top$	$\neg\phi$	$\phi \rightarrow \psi$	$\phi \vee \psi$	$\phi \wedge \psi$	$\phi \mid \psi$	$\dots$
F	F	F	T		T	F	F	T	
	T			T	T	T	F	T	
T	F				F	T	F	T	
	T			F	T	T	T	F	

T/F assignment for variables determines the valuation of well-formed formulae in a propositional logic system.

Number of tautologies

If we have m variables, $\neg$, $\rightarrow$, there exists

$\neg\neg\neg\neg X$ type

$\neg\neg[X \rightarrow X]$ type : m

$\neg[\neg X \rightarrow Y]$ type

$\neg[X \rightarrow \neg Y]$ type

$[\neg X \rightarrow \neg X]$ type : m

$[\neg\neg X \rightarrow X]$ type : m

$[X \rightarrow \neg\neg X]$ type : m

$X \rightarrow [Y \rightarrow X]$ type : m^2

$X \rightarrow [Y \rightarrow Y]$ type : m^2

$X \rightarrow [X \rightarrow X]$ type : m

$[X \rightarrow Y] \rightarrow Z$ type

$2m^2 + 3m$ tautologies of length 5.

Density of tautologies

Let $p_{m,n}$ be the probability of tautologies among length n well-formed formulae with m variables.

Density of tautologies

Let $p_{m,n}$ be the probability of tautologies among length n well-formed formulae with m variables.

Existence of the density

For any fixed number m of variables, the density of tautologies $\tau_m := \lim_{n \rightarrow \infty} p_{m,n}$ exists.

This is a straightforward result of the Drmota-Lalley-Woods Theorem.

Density of tautologies

Let $p_{m,n}$ be the probability of tautologies among length n well-formed formulae with m variables.

Existence of the density

For any fixed number m of variables, the density of tautologies $\tau_m := \lim_{n \rightarrow \infty} p_{m,n}$ exists.

This is a straightforward result of the Drmota-Lalley-Woods Theorem.

Question

What is the asymptotic behavior of τ_m ?

Known results

Fournier, Gardy, Genitrini, and Zaionc, 2010

For the logic system with the implication, m variables and their negative literals, both asymptotic behavior of the density of tautologies and the density of simple tautologies are

$$\frac{7}{8m} + O\left(\frac{1}{m^2}\right)$$

E., *preprint*

For the logic system with the implication and negation, the asymptotic behavior of the density of simple tautologies is

$$\frac{1}{m} + O\left(\frac{1}{m\sqrt{m}}\right)$$

New result

The density of tautologies

For the logic system with the implication and negation, the asymptotic behavior of the density of tautologies is

$$\frac{1}{m} + O\left(\frac{1}{m\sqrt{m}}\right)$$

Expected valuation

If we have a well-formed formula which is expected to be true/false, then we can define expected valuations.

$\neg\phi$ is True	ϕ is False
$\neg\phi$ is False	ϕ is True
$\phi \rightarrow \psi$ is False	ϕ is True, ψ is False
$\phi \rightarrow \psi$ is True	ϕ is False or ψ is True

Expected valuation

If we have a well-formed formula which is expected to be true/false, then we can define expected valuations.

$\neg\phi$ is True	ϕ is False
$\neg\phi$ is False	ϕ is True
$\phi \rightarrow \psi$ is False	ϕ is True, ψ is False
$\phi \rightarrow \psi$ is True	ϕ is False or ψ is True

We can repeat this process until we reach the variables.

Backward valuation

A well-formed formula ϕ is a tautology if and only if every possible choice of expected valuations for “ ϕ is False” has a contradiction.

Backward valuation

A well-formed formula ϕ is a tautology if and only if every possible choice of expected valuations for “ ϕ is False” has a contradiction.

$$[[\neg X \rightarrow \neg Y] \rightarrow [Y \rightarrow X]]:F$$

Backward valuation

A well-formed formula ϕ is a tautology if and only if every possible choice of expected valuations for “ ϕ is False” has a contradiction.

$$[\neg X \rightarrow \neg Y]:T, [Y \rightarrow X]:F$$

Backward valuation

A well-formed formula ϕ is a tautology if and only if every possible choice of expected valuations for “ ϕ is False” has a contradiction.

$$(\neg X:F \text{ or } \neg Y:T), Y:T, X:F$$

Backward valuation

A well-formed formula ϕ is a tautology if and only if every possible choice of expected valuations for “ ϕ is False” has a contradiction.

$$(X:T, Y:T, X:F) \text{ or } (Y:F, Y:T, X:F)$$

Shape

A shape is a template for generating well-formed formulae.

$$[\neg X_1 \rightarrow \neg X_2] \rightarrow [X_3 \rightarrow X_4]$$

Here, $X_1, \dots, X_4$ are placeholder for variables.

Shape

A shape is a template for generating well-formed formulae.

$$[\neg X_1 \rightarrow \neg X_2] \rightarrow [X_3 \rightarrow X_4]$$

Here, $X_1, \dots, X_4$ are placeholder for variables.

As we observed, this shape is a tautology if $X_1 = X_4$ and $X_2 = X_3$.

L-valuation and R-valuation

For a falsely expected shape, we have two special expected valuations.

- If $[\phi \rightarrow \psi]:T$, then always choose $\phi:F$.
- If $[\phi \rightarrow \psi]:T$, then always choose $\psi:T$.

We will call the first one as the L-valuation, the later one as the R-valuation, and the common expected valuation as the C-valuation.

Generating functions

Generating function for well-formed formulae

The generating function $W(z)$ for well-formed formulae is defined as

$$W(z) = \sum_{\phi} z^{\text{length}(\phi)}$$

Then,

$$W(z) = mz + zW(z) + zW(z)^2$$

Here,

- mz refers to variables.
- $zW(z)$ refers to $\neg\phi$ form well-formed formulae.
- $zW(z)^2$ refers to $\phi \rightarrow \psi$ form well-formed formulae.

Generating functions

For a fixed shape,

- ν_F be the number of falsely expected placeholders for the C-valuation.
- ν_T be the number of truly expected placeholders for the C-valuation.

Generating function for C-valuations.

The generating function $F(z, s, t)$ is defined as

$$F(z, s, t) = \sum_{\phi} s^{\nu_F} t^{\nu_T} z^{\text{length}(\phi)}$$

Then,

$$F(z, s, t) = mzs + z(1 + F(z, s, t))(mzt + zF(z, s, t) + zW(z)^2).$$

Probability estimation

For a fixed shape, let

- ν_L be the number of placeholders for the L-valuation.
- ν_R be the number of placeholders for the R-valuation.

Probability of tautologies with a fixed shape

For a fixed shape, probability of tautologies with given shape is bounded by

$$\frac{\nu_F \nu_T}{m} + \frac{\nu_L^2 \nu_R^2}{16m^2}$$

Probability estimation

Probability of tautologies

The probability of tautologies of the length n is bounded by

$$\frac{1}{m} \frac{1}{[z^n]W(z)} \left. \frac{\partial}{\partial s} \frac{\partial}{\partial t} [z^n]F(z, s, t) \right|_{(1,1)} + O(m^{-3/2}).$$

This comes from $\mathbb{E}[\mathbb{E}[A|B]] = \mathbb{E}[A]$.

Probability estimation

Probability of tautologies

The probability of tautologies of the length n is bounded by

$$\frac{1}{m} \frac{1}{[z^n]W(z)} \left. \frac{\partial}{\partial s} \frac{\partial}{\partial t} [z^n]F(z, s, t) \right|_{(1,1)} + O(m^{-3/2}).$$

This comes from $\mathbb{E}[\mathbb{E}[A|B]] = \mathbb{E}[A]$.

Szegő lemma

If W_* is a generating function on a subset of well-formed formula,

$$\lim_{n \rightarrow \infty} \frac{[z^n]W_*(z)}{[z^n]W(z)} = \lim_{z \rightarrow \frac{1}{1+2\sqrt{m}}} \frac{W'_*(z)}{W'(z)}.$$

if both exist.

The density of tautologies

Result

The density of tautologies τ_m among well-formed formulae with m variables is

$$\tau_m = \frac{1}{m} + O\left(\frac{1}{m\sqrt{m}}\right)$$

Application

Algorithm

Suppose that we consider the following simple probabilistic algorithm:

- For k random valuations, check whether ϕ is true for all of them.
- If then, assume ϕ as a tautology.

Application

Algorithm

Suppose that we consider the following simple probabilistic algorithm:

- For k random valuations, check whether ϕ is true for all of them.
- If then, assume ϕ as a tautology.

With some more detailed computations, there is a constant C such that if $k \geq \frac{C}{\epsilon} m(\log m)^2$, then the algorithm is correct with probability at least $1 - \epsilon$.

Thank you.