SCHEDAE INFORMATICAE
VOLUME 12 2003

Asymptotic Properties of Logics [1]

Marek Zaionc
Institute of Computer Science, Jagiellonian University, Nawojki 11, Kraków, 30-072, Poland
e-mail: zaionc@ii.uj.edu.pl

Abstract. This paper presents the number of results concerning problems of asymptotic densities in the variety of propositional logics. We investigate, for propositional formulas, the proportion of tautologies of the given length n against the number of all formulas of length n. We are specially interested in asymptotic behavior of this fraction. We show what the relation between a number of premises of an implicational formula and asymptotic probability of finding a formula with this number of premises is. Furthermore we investigate the distribution of this asymptotic probabilities. Distribution for all formulas is contrasted with the same distribution for tautologies only.

Keywords: Propositional logic, asymptotic density of tautologies, probabilistic methods in logic.

1. Introduction

The research described in this paper is a part of the project[1] of quantitative investigations in logic. This paper summarizes the research in which we develop methods of finding the asymptotic probability in some propositional

[1]Supported by the State Committee for Scientific Research (KBN), research grant 7T11C 022 21.

logics. We investigate, for propositional formulas, the proportion between the number of valid formulas of the given length n against the number of all formulas of length n for propositional formulas. Our interest lays in finding limit of that fraction when $n \to \infty$. If the limit exists it represents the real number which we may call *the density of truth* for the logic investigated. In general we are interested in finding also the *"density"* of some other classes of formulas.

We assume that the set of formulas $\mathcal{F}$ of a given propositional calculus is equipped with norm $\|.\|$ which is a function $\|.\| : \mathcal{F} \mapsto \mathbb{N}$. Moreover, we assume that for any n the set of formulas $\{\phi \in \mathcal{F} : \|\phi\| = n\}$ is finite. Typical norms are presented in Definitions 5 and 8. In the Definition 5 norm $\|\phi\|$ means the total number of appearances of propositional variables in the formula ϕ while in the Definition 8 $\|\phi\|$ is the number of characters in formula ϕ without parentheses.

In the whole paper we present some properties of numbers characterizing the amount of formulas in different classes defined in our language and we are concerned with the asymptotic behavior of those numbers. The main tools we use for dealing with asymptotics of sequences of numbers are known in combinatorics as *generating functions.* A nice exposition of the method can be found in [4] and [1]. Also see papers [5, 6, 2, 3] for the presentation of this method in logics.

Definition 1. *We associate the density $\mu(\mathcal{A})$ with a subset $\mathcal{A}$ of formulas as:*

$$\mu(\mathcal{A}) = \lim_{n\to\infty} \frac{\#\{t \in \mathcal{A} : \|t\| = n\}}{\#\{t \in \mathcal{F} : \|t\| = n\}} \tag{1}$$

if the appropriate limit exists.

The number $\mu(\mathcal{A})$, if exists, is an asymptotic probability of finding a formula from the set $\mathcal{A}$ among all formulas. It also may be interpreted as the asymptotic density of the set $\mathcal{A}$. It can be seen immediately that the density μ is finitely additive so if $\mathcal{A}$ and $\mathcal{B}$ are disjoined classes of formulas such that $\mu(\mathcal{A})$ and $\mu(\mathcal{B})$ exist, then $\mu(\mathcal{A} \cup \mathcal{B})$ exists also and

$$\mu(\mathcal{A} \cup \mathcal{B}) = \mu(\mathcal{A}) + \mu(\mathcal{B}). \tag{2}$$

It is straightforward to observe that for any finite set $\mathcal{A}$ the density $\mu(\mathcal{A})$ exists and is 0, and dually for co-finite sets $\mathcal{A}$ the density $\mu(\mathcal{A}) = 1$. Unfortunately, the density μ is not countably additive so in the general formula (3) below

$$\mu\left(\bigcup_{i=0}^{\infty}\mathcal{A}_i\right)=\sum_{i=0}^{\infty}\mu\left(\mathcal{A}_i\right) \tag{3}$$

it is not true for all pairwise disjoint classes of sets $\{\mathcal{A}_i\}_{i\in\mathbb{N}}$. The good counterexample for the equation (3) is to take as $\mathcal{A}_i$ the singleton consisting of i-th formula from our language. On the left hand side of (3) we get 1 but on the right hand side $\mu\left(\mathcal{A}_i\right)=0$ for all $i\in\mathbb{N}$. In the paper we also discuss the distribution of densities with respect to some numerical property of formulas.

Definition 2. *By a random variable X we understand the function*

$$X:\mathcal{F}\mapsto\mathbb{N}$$

which assigns a number $n\in\mathbb{N}$ to the formula in such a way that for any n the density $\mu\left(\{\phi : X(\phi)=n\}\right)$ exists and moreover

$$\sum_{n=0}^{\infty}\mu\left(\{\phi : X(\phi)=n\}\right)=1.$$

Definition 3. *By the distribution of a random variable X we mean the function $\overline{X}:\mathbb{N}\mapsto\mathbb{R}$ defined by:*

$$\overline{X}\ :\ \mathbb{N}\ni n\mapsto\mu\left(\{\phi : X(\phi)=n\}\right)\in\mathbb{R}.$$

Definition 4. *The expected value, variance and standard deviation are defined in the conventional way by:*

$$\begin{aligned} E(X) &= \sum_{p=0}^{\infty}p\,\overline{X}(p), &(4)\\ D^2(X) &= E((X-E(X))^2)=E(X^2)-(E(X))^2 &(5)\\ &= \sum_{p=0}^{\infty}p^2\overline{X}(p)-(E(X))^2 \end{aligned}$$

so the standard deviation of X is $\sqrt{\mathcal{D}^2(X)}$.

2. Densities in logics

In this section we present some results obtained in [2, 3, 5] and [6] characterizing the density of tautologies in some propositional languages.

Definition 5. *The set of formulas* $\mathcal{F}_k^{\rightarrow}$ *over a* k *propositional variables is a minimal set consisting of those variables and closed for implication only. In this definition the norm* $\|.\|$ *measure the total number of appearances of propositional variables in the formula.*

The theorem bellow describes the asymptotic density of the set of tautologies in the simplest possible language of implicational formulas over one propositional variable (see [3]). The natural inspiration for research comes from the typed lambda calculus in which the set of simple types can be identified with formulas under so called Curry-Howard isomorphism. Under this isomorphism the class of provable formulas can be understood as the class of inhabited types. Notice that the density of provable formulas in this language is surpassingly hight. Notice also that the classical tautologies in this language coincide with intuitionistic ones. The *proof by counting* is sketched in Theorem 7.

Theorem 6. *(see [3] page 592) For* $k = 1$ *the asymptotic density of the set of intuitionisticaly provable formulas* $\mathcal{T}_1^{\rightarrow}$ *exists and is:*

$$\mu(\mathcal{T}_1^{\rightarrow}) = \frac{1}{2} + \frac{\sqrt{5}}{10} \approx 0.7236067978\ldots$$

Theorem 7. *The implicational classical and intuitionistic logic of one variable are identical.*

Proof by counting can be found in [6]. Proof is based on the fact that generating functions for classical and intuitionistic logics are identical.

In the next two theorems we change the language. We consider formulas built by means of implication and negation from one variable. We can see that adding negation has a negative impact on the density of tautologies. Moreover, as a result of paper [2] we are able to find the exact density of an intuitionistic fragment of the classical logic for this language. We can also see by Theorem 12 that within the reacher language with negation the density of purely implicational tautologies is 0 in the class of all tautologies.

Definition 8. *The set* $\mathcal{F}_k^{\rightarrow,\neg}$ *over a k propositional variables is a minimal set consisting of those variables and closed for implication and negation. In this definition the norm* $\|\phi\|$ *means the total number of characters in formula* ϕ *without parentheses.*

Theorem 9. *(see Zaionc [6]) For* $k = 1$ *the asymptotic density of the set of classical tautologies* $Cl_1^{\rightarrow,\neg}$ *exists and is:*

$$\begin{aligned} \mu(Cl_1^{\rightarrow,\neg}) &= \frac{1}{(4\sqrt{13})} + \frac{1}{(4\sqrt{17})} + \frac{1}{2\sqrt{2\left(\sqrt{221}-9\right)}} + \frac{15}{2\sqrt{442\left(\sqrt{221}-9\right)}} \\ &= 0.423238538401941... \end{aligned}$$

Theorem 10. *(see Kostrzycka, Zaionc in [2]) For* $k = 1$ *the asymptotic density of the set of intuitionisticaly provable formulas* $I_1^{\rightarrow,\neg}$ *exists and is:*

$$\mu(I_1^{\rightarrow,\neg}) \approx 0.395205.$$

In the paper [2] the reader can find the analytical formula for $\mu(I_1^{\rightarrow,\neg})$. Putting together Theorems 9 and 10 we obtain:

Theorem 11. *[Relative density (see Kostrzycka, Zaionc in [2])]*
The relative density of intuitionistic tautologies among classical ones in the language $\mathcal{F}_1^{\rightarrow,\neg}$ *is more than 93 %.*

$$\mu((I_1^{\rightarrow,\neg})/(Cl_1^{\rightarrow,\neg})) \approx 0.93.$$

In the paper [2] the reader can find the analytical formula for $\mu((I_1^{\rightarrow,\neg})/(Cl_1^{\rightarrow,\neg}))$. We may also compare classical fragments of the languages $\mathcal{F}_1^{\rightarrow,\neg}$ and $\mathcal{F}_1^{\rightarrow}$.

Theorem 12. *Probability of finding an implicational tautology among implicational, negational tautologies is* 0 *(in the sense of the norm in Definition 8).*

3. Probability distribution, typical formulas, typical tautologies

In this section we will discuss the questions concerning probability distribution of formulas written in the implicational language $\mathcal{F}_k^{\rightarrow}$ (see Definition 5) equipped with the norm $\|\cdot\|$ measuring the total number of appearances of propositional variables in the formula

Definition 13. *By* $\mathcal{F}_k^{\rightarrow}(p)$ *we mean the set of formulas having* p *premises, i.e. formulas which are of the form:* $\tau = \tau_1 \rightarrow (\ldots \rightarrow (\tau_p \rightarrow \alpha))$, *where* α *is a propositional variable.*

Definition 14. *A* simple tautology *is a formula* $\tau \in \mathcal{F}_k^{\rightarrow}$ *on the form*

$$\tau = \tau_1 \rightarrow (\ldots \rightarrow (\tau_p \rightarrow \alpha)),$$

such that there is at least one component τ_i *identical with* α*. Let* $\mathcal{G}_k$ *be the set of all simple tautologies in* $\mathcal{F}_k^{\rightarrow}$ *and* $\mathcal{G}_k(p)$ *be the set of simple tautologies with* p *premises*

Evidently, a simple tautology is a tautology. Our goal is to find how big asymptotically is the fragment of simple tautologies within the set of all formulas and also how big is the fragment of simple tautologies with p premises in the set of all simple tautologies.

Definition 15. *Let us define a random variable* $X : \mathcal{F}_k^{\rightarrow} \mapsto \mathbb{N}$ *(see Definition 2) which assigns a number of premisses to an implicational formula.*

In Theorem 17 we check the correctness of the definition above since for any n the density $\mu\left(\{\phi : X(\phi) = n\}\right)$ exists and moreover

$$\sum_{n=0}^{\infty} \mu\left(\{\phi : X(\phi) = n\}\right) = 1.$$

We wish to answer two questions:

Question 1: *What is a probability that a randomly chosen implicational formula admits* p *premises.*

Question 2: *What is a probability that a randomly chosen implicational simple tautology admits* p *premises.*

LEMMA 16. *(see [3]) The asymptotic density of the set of all formulas with p premisses $\mathcal{F}_k^{\rightarrow}(p)$ exists and is:*

$$\mu(\mathcal{F}_k^{\rightarrow}(p)) = \frac{p}{2^{p+1}}. \tag{6}$$

THEOREM 17. *The random variable X has the following distribution (see Definition 3):*

$$\overline{X}(p) = \frac{p}{2^{p+1}}.$$

Expected value $E(X) = 3$, variance $D^2(X) = 4$. The standard deviation of X is 2.

From the whole discussion we can see that surprisingly a typical implicational formula suppose to have exactly 3 premisses. For example, the amount of formulas with number of premises laying between 1 and 5, i.e. which are typical ± standard deviation is 57/64 which is about 89%.

Now we will start to answer the second question. We will see the difference between distribution of the number of premises for all formulas contrasted with the same distribution for simple tautologies only.

DEFINITION 18. *For every $k \geq 1$ separately let us define a random variable Y_k which assigns to an implicational simple tautology in the language $\mathcal{F}_k^{\rightarrow}$ the number of its premises.*

THEOREM 19. *(Zaionc [5]) The random variable Y_k has the following distribution:*

$$\overline{Y_k}(p) = \left(\frac{(2k+1)^2}{4k+1}\right)\left(\frac{p}{2^{p+1}} - p\frac{(2k-1)^{p-1}}{4^p k^{p-1}}\right).$$

The natural question is how the distribution of true sentences looks like for very large numbers k or if there exists an uniform asymptotic distribution when k, the number of propositional variables in the logic, tends to infinity. The answers are following:

LEMMA 20. *(Zaionc [5]) In this lemma the number of premises $p \geq 0$ is fixed.*

$$\lim_{k\to\infty}\left(\frac{(2k+1)^2}{4k+1}\right)\left(\frac{p}{2^{p+1}} - p\frac{(2k-1)^{p-1}}{4^p k^{p-1}}\right) = \frac{p(p-1)}{2^{p+2}}. \tag{7}$$

Let us name the limit distribution by $Y_\infty(p) = \dfrac{p(p-1)}{2^{p+2}}$ since:

$$\sum_{p=0}^{\infty} \frac{p(p-1)}{2^{p+2}} = 1. \tag{8}$$

Expected value of Y_∞ is:

$$E(Y_\infty) = \sum_{p=0}^{\infty} p\frac{p(p-1)}{2^{p+2}} = 5. \tag{9}$$

The variance of Y_∞ is:

$$D^2(Y_\infty) \;\; = \;\; \sum_{p=0}^{\infty} p^2\frac{p(p-1)}{2^{p+2}} - 25 = 31 - 25 = 6.$$

Comparing this with the distribution $\overline{X}(p)$ the reader can easily check that starting with $k = 1$ the expected value of the number of premises for simple tautologies is substantially greater then 3 and is growing asymptotically to 5 and

$$\lim_{k\to\infty} E(Y_k) = 5. \tag{10}$$

Also asymptotical behavior of $D^2(Y_k)$ is

$$\lim_{k\to\infty} D^2(Y_k) = 6. \tag{11}$$

So it is clear now that:

$$\forall p \geq 0 \;\; \lim_{k\to\infty} \overline{Y_k}(p) \;\; = \;\; \overline{Y_\infty}(p), \tag{12}$$

$$\lim_{k\to\infty} E(Y_k) \;\; = \;\; E(Y_\infty), \tag{13}$$

$$\lim_{k\to\infty} D^2(Y_k) \;\; = \;\; D^2(Y_\infty). \tag{14}$$

The componentwise convergence presented in Lemma 20 and summarized by the formula (12) can be extended to the much stronger uniform convergence. Therefore, in fact the distribution $\overline{Y_\infty}$ can be treated as a good model of distribution for simple tautologies for the language $\mathcal{F}$ when the number k of atomic propositional variables is large.

THEOREM 21. *(Zaionc [5]) The sequence of distributions $\overline{Y_k}$ uniformly converges to the distribution $\overline{Y_\infty}$.*

We can also see surprising result

THEOREM 22. *(Zaionc [5]) For fixed $k > 0$ and $p > 0$*

$$\mu[(\mathcal{G}_k)/(\mathcal{F}_k^{\rightarrow}(p))] = 1 - \left(\frac{2k-1}{2k}\right)^{p-1}. \tag{15}$$

This result is somehow intriguing. It shows how asymptotically big the size of the fraction of simple tautologies with p premises among all formulas of p premises is. We can see that with p growing this fraction becomes closer and closer to 1. Of course, the fraction of all, not only simple, tautologies with p premises is even larger. So the *"density of truth"* within the classes of formulas of p premises can be as big as we wish. For every $\varepsilon > 0$ we can effectively find p such that among formulas with p premises almost all formulas (except the tiny fraction of the size ε) asymptotically are tautologies. This should be contrasted with the results proved in Theorem 6.3 and Corollary 6.10 page 587 at [3]. It shows that *the density of truth* for all $p's$ together is always of the size $O(1/k)$. The result for every p treated separately is very different.

4. References

[1] Comtet L.; *Advanced combinatorics. The art of finite and infinite expansions.*, Revised and enlarged edition, Reidel, Dordrecht 1974.

[2] Kostrzycka Z., Zaionc M.: *Statistics of intuitionistic versus classical logics,* to appear in Studia Logica.

[3] Moczurad M., Tyszkiewicz J., Zaionc M.; *Statistical properties of simple types,* Mathematical Structures in Computer Science, Vol. 10, 2000, pp. 575–594.

[4] Wilf H.S.; *Generatingfunctionology*, Second edition, *Academic Press*, Boston 1994.

[5] Zaionc M.; *Statistics of implicational logic,* to appear in: Electronic Notes in Theoretical Computer Science, Vol. 84.

[6] Zaionc M.; *On the asymptotic density of tautologies in logic of implication and negation*, to appear in: Reports on Mathematical Logic.

Received November 8, 2002