Intuitionistic vs. Classical Tautologies, Quantitative Comparison

Antoine Genitrini[1], Jakub Kozik[2], and Marek Zaionc[2]

[1] PRiSM, CNRS UMR 8144, Université de Versailles
Saint-Quentin en Yvelines, 45 av. des États-Unis, 78035 Versailles cedex, France
antoine.genitrini@prism.uvsq.fr
[2] Theoretical Computer Science, Jagiellonian University,
Gronostajowa 3, Kraków, Poland
[jkozik,zaionc]@tcs.uj.edu.pl

Abstract. We consider propositional formulas built on implication. The size of a formula is the number of occurrences of variables in it. We assume that two formulas which differ only in the naming of variables are identical. For every $n \in \mathbb{N}$, there is a finite number of different formulas of size n. For every n we consider the proportion between the number of intuitionistic tautologies of size n compared with the number of classical tautologies of size n. We prove that the limit of that fraction is 1 when n tends to infinity[1].

1 Introduction

In the present paper we consider propositional formulas built on implication only. In particular we do not use logical constant $\perp$. The size of a formula is the number of occurrences of variables in it. We assume that two formulas which differs only in the naming of variables are identical. For every $n \in \mathbb{N}$, there is finite number of different formulas of size n, we denote that number by $F(n)$. Consequently there is also finite number of classical tautologies and of intuitionistic tautologies of that size. These numbers are denoted by $Cl(n)$ and $Int(n)$ respectively. We are going to prove that:

$$\lim_{n\to\infty} \frac{Int(n)}{Cl(n)} = 1.$$

This work is a part of the research in which the asymptotic likelihood of truth is estimated. We refer to Gardy [4] for a survey on probability distribution on Boolean functions induced by random Boolean expressions. For the purely implicational logic of one variable the exact value of the density of truth was computed in the paper [11] of Moczurad, Tyszkiewicz and Zaionc. It is well known that under Curry-Howard isomorphism this result answered the question

[1] Research described in this paper was partially supported by POLONIUM grant *Quantitative research in logic and functional languages*, cooperation between Jagiellonian University of Krakow, L' École Normale Supérieure de Lyon and Universite de Versailles Saint-Quentin, contract number 7087/R07/R08.

M. Miculan, I. Scagnetto, and F. Honsell (Eds.): TYPES 2008, LNCS 4941, pp. 100–109, 2008.

of finding the "density" of inhabited types in the set of all types. The classical logic of one variable and the two connectors – implication and negation – was studied in Zaionc [13]. Over the same language, the exact proportion between intuitionistic and classical logics has been determined in Kostrzycka and Zaionc [7]. Some variants involving formulas with other logical connectors have also been considered. The case of and/or connectors received much attention – see Lefmann and Savický [8], Chauvin, Flajolet, Gardy and Gittenberger [1] and Gardy and Woods [5]. Matecki [9] considered the case of the equivalence connector.

In the latest paper [6] of Fournier, Gardy, Genitrini and Zaionc, the proportion between intuitionistic and classical logics when the overall number of variables is finite was studied. In this paper, the methods (and moreover the partition of formulas into several classes) developed in [6] are used in the different case, when the number of variables is arbitrary and not fixed. But formally this paper and [6] are incomparable in the sense that each one is not an extension of the other.

2 Basic Facts

2.1 Catalan Numbers

The n-th Catalan number is the number of binary trees with n internal nodes or, equivalently, $n+1$ leaves. For our exposition it will be convenient to focus on leaves, therefore we denote by $C(n)$ the $(n-1)$-th Catalan number. Its (ordinary) generating function is

$$c(z) = \sum_{n\in\mathbb{N}} C(n)z^n = \frac{1-\sqrt{1-4z}}{2}.$$

That function fulfills the following property

$$c(z) = c(z)c(z) + z. \tag{1}$$

The radius of convergence of $c(z)$ is $\frac{1}{4}$ and $\lim_{z\to_{\mathbb{R}}\frac{1}{4}} c(z) = \frac{1}{2}$. We have also the following property

$$\lim_{n\to\infty} \frac{C(n-1)}{C(n)} = \frac{1}{4}.$$

2.2 Algebraic Asymptotics

Lemma 1. *Let $f, g \in \mathbb{Z}[[z]]$ be two algebraic generating functions, having (as complex analytic functions) unique dominating singularities in $\rho \in \mathbb{R}_+$. Suppose that these functions have Puiseux expansions around ρ of the form*

$$f(z) = c_f + d_f(z-\rho)^{\frac{1}{2}} + o((z-\rho)^{\frac{1}{2}})$$

$$g(z) = c_g + d_g(z-\rho)^{\frac{1}{2}} + o((z-\rho)^{\frac{1}{2}}).$$

Then

$$\lim_{n\to\infty} \frac{[z^n]f(z)}{[z^n]g(z)} = \lim_{z\to_{\mathbb{R}}\rho^-} \frac{f'(z)}{g'(z)}.$$

By the singularity analysis for algebraic generating functions (see e.g. Theorem 8.12 from [2]) we obtain that:

$$\lim_{n\to\infty} \frac{[z^n]f(z)}{[z^n]g(z)} = \frac{d_f}{d_g}.$$

On the other hand it can be easily calculated that $\lim_{z\to_{\mathbb{R}}\rho^-} \frac{f'(z)}{g'(z)} = \frac{d_f}{d_g}$. The analogous argument can be derived from the Szegö Lemma (see [12]).

2.3 Bell Numbers

The n-th Bell number, denoted by $B(n)$, is the number of equivalence relations which can be defined on some fixed set of size n. We use the following property, which can be derived from the asymptotic formula for Bell numbers by Moser and Wyman ([10], see also [3]).

$$\frac{B(n-1)}{B(n)} \sim \frac{e\log(n)}{n}.$$

2.4 Formulas

Implicational formulas can be represented by binary trees, suitably labeled: their internal nodes are labeled by the connector $\to$ and their leaves by some variables. By $\|\phi\|$ we mean the *size* of formula ϕ which we define to be the total number of occurrences of propositional variables in the formula (or leaves in the tree representation of the formula). Parentheses (which are sometimes necessary) and the implication sign itself are not included in the size of expressions. Formally,

$$\|x_i\| = 1 \text{ and } \|\phi \to \psi\| = \|\phi\| + \|\psi\|.$$

We denote by $F(n)$ the number of implicational formulas of size n. Let T be a formula (tree). It can be decomposed with respect to its right branch. Hence it is of the form $A_1 \to (A_2 \to (\ldots \to (A_p \to r(T)))\ldots)$ where $r(T)$ is a variable. We shall write it as

$$T = A_1, \ldots, A_p \to r(T).$$

The formulas A_i are called the *premises* of T and $r(T)$, the rightmost leaf of the tree, is called the *goal* of T.

2.5 Counting up to Names

It is easy to observe, that the number of different formulas of size n is

$$F(n) = B(n)C(n).$$

$C(n)$ corresponds to the shapes of formulas represented by trees, and $B(n)$ to all possible distributions of variables in that shape.

3 Simple Tautologies

We follow notation from [11], [14] and [6]. We are going to prove a theorem analogous to the main theorem of [6].

Definition 1. *G is the set of* simple tautologies *i.e. expressions that can be written as*

$$T = A_1, \ldots, A_p \to r(T),$$

where there exists i such that A_i is a variable equal to $r(T)$. We let $(G(n))_{n\in\mathbb{N}}$ be the sequence of the numbers of simple tautologies of size n.

It is easy to prove that simple tautologies are indeed intuitionistic tautologies. The asymptotic equivalence of classical and intuitionistic logic is a direct consequence of the following theorem.

Theorem 1. *Asymptotically, all the classical tautologies are simple.*

We are going to prove the theorem in three steps. First, we estimate the number of simple tautologies. Then in two steps we show that the number of remaining tautologies is asymptotically negligible.

Lemma 2. *The fraction of simple tautologies among all formulas of size n is asymptotically equal to $\frac{e\log(n)}{n}$.*

Proof. First, we enumerate all the shapes of trees which can be labelled to be simple tautologies. The set of such trees will be denoted by GT. A tree belongs to GT if and only if it has at least one premise which is a leaf. Let $GT(n,l)$ denote the number of trees of size n, whose l premises are leaves. We define bivariate generating function $gt(x,z) = \sum_{n,l\in\mathbb{N}\setminus\{0\}} GT(n,l)z^n x^l$. We use standard unlabeled constructions (see [3]) to obtain the explicit expression for $gt(x,z)$. Clearly, for every tree t from GT there is a premise which is a leaf. The last such premise decomposes uniquely the sequence of premises of t into two sequences. The first consists of arbitrary trees, while the second – of trees which are not leaves. Note that $c^2(z)$ is the generating function for trees which are not leaves. Corresponding constructions on generating function yields:

$$gt(x,z) = \frac{1}{1-c^2(z)-xz} \cdot xz \cdot \frac{1}{1-c^2(z)} \cdot z. \tag{2}$$

In the expression above the term $\frac{1}{1-c^2(z)-xz}$ corresponds to the sequence of trees which are either leaves (xz) or not ($c^2(z)$). The second term xz corresponds to the last premise that is a leaf. The third term $\frac{1}{1-c^2(z)}$ corresponds to the remaining sequence of premises which are not leaves. The last occurrence of z corresponds to the goal.

Let us fix $l \in \mathbb{N}\setminus\{0\}$. Let $G_l(n)$ denote the number of simple tautologies of size n in which l premises are leaves. ¿From the inclusion-exclusion principle we obtain

$$GT(n,l)\cdot l\cdot B(n-1) > G_l(n) > GT(n,l)\cdot l\cdot B(n-1) - GT(n,l)\frac{l(l-1)}{2}B(n-2).$$

The first inequality comes from the fact that in every simple tautology there is at least one premise which is equal to goal. Hence $GT(n,l)$ corresponds to the shape of the tree, l corresponds to the possibilities of choice of the premise, $B(n-1)$ corresponds to the all possible labeling of variables ($n-1$ since one premise is chosen to be equal to the goal). Of course, the formulas in which many premises are equal to the goal are counted more than once. The second inequality comes from subtracting all the formulas in which at least two premises are equal to the goal (again some formulas are subtracted many times). We have

$$\sum_{l\in\mathbb{N}\setminus\{0\}} GT(n,l)\cdot l\cdot B(n-1) > \sum_{l\in\mathbb{N}\setminus\{0\}} G_l(n)$$

$$\sum_{l\in\mathbb{N}\setminus\{0\}} G_l(n) > \sum_{l\in\mathbb{N}\setminus\{0\}} \left(GT(n,l)\cdot l\cdot B(n-1) - GT(n,l)\frac{l(l-1)}{2}B(n-2)\right)$$

therefore for every $n\in\mathbb{N}$

$$\frac{\sum_{l\in\mathbb{N}\setminus\{0\}} GT(n,l)\cdot l\cdot B(n-1)}{F(n)} > \frac{\sum_{l\in\mathbb{N}\setminus\{0\}} G_l(n)}{F(n)} \tag{3}$$

and

$$\frac{\sum_{l\in\mathbb{N}\setminus\{0\}} G_l(n)}{F(n)} > \frac{\sum_{l\in\mathbb{N}\setminus\{0\}}(GT(n,l)\cdot l\cdot B(n-1) - GT(n,l)\frac{l(l-1)}{2}B(n-2))}{F(n)} \tag{4}$$

We are going to find a succinct formula for the generating function $\overline{gt}(z) = \sum_{n\in\mathbb{N}} z^n \sum_{l\in\mathbb{N}\setminus\{0\}} l\cdot GT(n,l)$. Taking the derivative of the function $gt(x,z)$ with respect to x we obtain the generating function of the sequence

$$gt'_x(x,z) = \sum_{n,l\in\mathbb{N}\setminus\{0\}} GT(n,l)\cdot l\cdot x^{l-1}z^n.$$

It remains to substitute 1 for x to obtain the sought generating function. We can write that function explicitly by applying those operations to the explicit formula for $gt(x,z)$. We obtain:

$$\overline{gt}(z) = \left(\frac{z}{1-c(z)}\right)^2 = c(z)^2,$$

the last equality results from (1). We encourage the reader to find a direct interpretation in terms of trees of the obtained expression for $\overline{gt}(z)$.

By differentiating $gt(x,z)$ twice with respect to x, substituting 1 for x and multiplying by $\frac{1}{2}$, we analogously obtain the generating function:

$$\overline{\overline{gt}}(z) = \sum_{n\in\mathbb{N}\setminus\{0\}} z^n \sum_{l\in\mathbb{N}\setminus\{0\}} GT(n,l)\frac{l(l-1)}{2}$$

Hence

$$\overline{\overline{gt}}(z) = gt''_x(1, z) = c(z)(c(z) - z).$$

Both generating functions $\overline{gt}(z)$ and $\overline{\overline{gt}}(z)$ are algebraic generating functions with unique dominating singularities in $\frac{1}{4}$. Hence by the Lemma 1:

$$\lim_{n\to\infty} \frac{[z^n]\overline{gt}(z)}{[z^n]c(z)} = \lim_{z\to_{\mathbb{R}}\frac{1}{4}^-} \frac{\overline{gt}'(z)}{c'(z)} = 1$$

and

$$\lim_{n\to\infty} \frac{[z^n]\overline{\overline{gt}}(z)}{[z^n]c(z)} = \lim_{z\to_{\mathbb{R}}\frac{1}{4}^-} \frac{\overline{\overline{gt}}'(z)}{c'(z)} = \frac{3}{4}$$

Therefore

$$\frac{\sum_{l\in\mathbb{N}} GT(n,l)\cdot l\cdot B(n-1)}{C(n)B(n)} = \frac{([z^n]\overline{gt}(z))}{C(n)}\frac{B(n-1)}{B(n)} \sim \frac{e\log(n)}{n}$$

and

$$\frac{\sum_{l\in\mathbb{N}} GT(n,l)\frac{l(l-1)}{2}B(n-2)}{C(n)B(n)} = \frac{([z^n]\overline{\overline{gt}}(z))}{C(n)}\frac{B(n-2)}{B(n)} \sim \frac{3}{8}\left(\frac{e\log(n)}{n}\right)^2$$

Finally from (3) and (4) we obtain

$$\frac{G(n)}{F(n)} = \frac{\sum_{l\in\mathbb{N}} G_l(n)}{F(n)} \sim \frac{e\log(n)}{n} \tag{5}$$

□

It remains to estimate the number of tautologies which are not simple. Those have to be found among formulas which are neither simple tautologies nor simple nontautologies (a formula T is simple nontautology if the goal of T does not occur as a goal of any premise of T). That means that in every such formula T there is at least one premise which is not a variable, and whose goal is equal to the goal of T. First we will show that the number of formulas $A_1, \ldots, A_k \to x$ in which x is a goal of at least two premises is negligible, the set of such formulas will be denoted by MP and the number of such formulas of size n by $MP(n)$.

Lemma 3. *The fraction of formulas $A_1, \ldots, A_k \to x$ in which x is a goal of at least two premises among all formulas of size n is $o(\frac{e\log(n)}{n})$.*

Proof. Let $P(n,l)$ denote the number of trees of size n in which l premises are not leaves. Let $p(x,z) = \sum_{n,l\in\mathbb{N}} P(n,l)z^n x^l$ be its generating function. From the equation (1) we know that

$$c(z) = \frac{z}{1 - c^2(z) - z}$$

That equation can be interpreted in terms of combinatorial constructions. Every tree is a sequence of premises, followed by the goal. That translates to the expression $c(z) = \frac{1}{1-c(z)} \cdot z$. Every tree is either a leaf or it consists of two subtrees, therefore we can substitute $c^2(z) + z$ for $c(z)$ in the last equation. We add a formal parameter x for every premise which is not a leaf to obtain the generating function $p(x, z)$:

$$p(x, z) = \frac{z}{1 - xc^2(z) - z}$$

Every formula $T \in MP$ has two premises with the goal equal to the goal of T, and those premises are not leaves. Therefore

$$MP(n) \leq B(n-2) \sum_{l \in \mathbb{N} \setminus \{0\}} P(n, l) \frac{l(l-1)}{2}$$

Note that

$$\sum_{n \in \mathbb{N}} z^n \sum_{l \in \mathbb{N}} P(n, l) \frac{l(l-1)}{2} = \tfrac{1}{2} p''_x(1, z) = \frac{c^5(z)}{(1 - c(z))^2} = \frac{c^7(z)}{z^2}$$

Hence

$$\lim_{n \to \infty} \frac{[z^n] \frac{c^5(z)}{(1-c(z))^2}}{[z^n] c(z)} = \lim_{z \to_{\mathbb{R}} \frac{1}{4}^-} \frac{(\frac{c^7(z)}{z^2})'}{c'(z)} = \frac{7}{4}$$

It follows that

$$\frac{MP(n)}{F(n)} \leq \frac{B(n-2) \sum_{l \in \mathbb{N}_1} P(n, l) \frac{l(l-1)}{2}}{C(n) B(n)} \sim \frac{7}{4} \left(\frac{e \log(n)}{n} \right)^2$$

and comparing to (5), formulas from MP are negligible. □

Finally, we estimate the number of tautologies which have exactly one premise with goal equal to the goal of the whole formula (compare the part about less simple non-tautologies in [6]).

Let T be such a formula, and C be that premise. Let D be the first premise of C (C is not a variable), and $r(D)$ be the goal of D (see Figure 3).

A necessary condition for the formula T to be a tautology is that either $r(D)$ is a goal of at least one premise of T or D, or $r(T)$ is a goal of some premises of D. We estimate the number of such formulas. Let LT be the set of formulas T for which both of the following conditions hold:

1. T has exactly one premise C whose goal is $r(T)$, and that premise is not a variable (this implies that T is not a simple tautology nor a simple nontautology).
2. Let D be the first premise of C. At least one of the following conditions holds:
   (a) there is a premise of D with goal $r(T)$,
   (b) there is a premise of D with goal $r(D)$,
   (c) $r(D) \neq r(T)$ and there is a premise of T with goal $r(D)$.

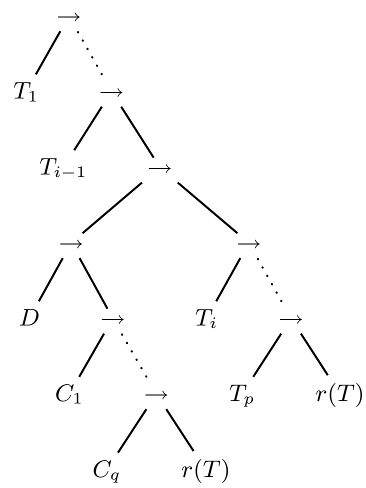

Fig. 1. Tautologies from LT

Lemma 4. *The fraction of tautologies which are not simple among all formulas of size n is $o\left(\frac{e\log(n)}{n}\right)$.*

Proof. Clearly all the tautologies, which are not simple and not in MP, belong to LT. Let $LT(n)$ denote the number of formulas from LT of size n. Since $MP(n)$ is $o\left(\frac{e\log(n)}{n}\right)$ it is enough to prove the estimation for $LT(n)$.

Let $LTT(n,m,l,k)$ denote the number of trees of size n which have $l+1$ premises, and in which the first premise of the m-th premise has exactly k premises. Let $LTT(n,l,k) = \sum_{m\in\mathbb{N}} LTT(n,m,l,k)$. Every such tree can be turned into formula from LT by the appropriate assignment of variables, and every formula from LT can be constructed in this way. Therefore

$$LT(n) \leq \sum_{l,k\in\mathbb{N}} (l+k)\cdot LTT(n,l,k)\cdot B(n-2) + \sum_{l,k\in\mathbb{N}} k\cdot LTT(n,l,k)\cdot B(n-2)$$

The first sum corresponds to the situation, where $r(D)$ occurs as a goal in some premises of D or T. The second one – to the the situation, where $r(T)$ occurs as a goal of some premises of D (these situations are not disjoint). Clearly

$$LT(n) \leq 2\sum_{l,k\in\mathbb{N}} (l+k)\cdot LTT(n,l,k)\cdot B(n-2).$$

Let $ltt(x,y,z) = \sum_{l,k,n\in\mathbb{N}} x^l y^k z^n LTT(n,l,k)$. We have

$$ltt(x,y,z) = \frac{1}{1-xc(z)}\cdot\frac{z}{1-yc(z)}\cdot c(z)\cdot\frac{1}{1-xc(z)}\cdot z.$$

The first term $\frac{1}{1-xc(z)}$ corresponds to the sequence of premises preceding the distinguished premise C. The component $\frac{z}{1-yc(z)}\cdot c(z)$ corresponds to the premise

C (formal parameter y counts the premises of the subtree corresponding to D). The last component $\frac{1}{1-xc(z)} \cdot z$ corresponds to the remaining premises of the main tree and the leaf (goal).

Let $ltt(x,z) = ltt(x,x,z)$, then $ltt(x,z) = \sum_{l,k,n\in\mathbb{N}} x^{(l+k)} z^n LTT(n,l,k)$, and $ltt'_x(1,z) = \sum_{l,k,n\in\mathbb{N}} z^n (l+k) LTT(n,l,k)$. We denote the last function by $ltt(z)$, it is the generating function for the sequence which majorizes $(\frac{LT(n)}{2B(n-2)})_{n\in\mathbb{N}}$. We can write it explicitly as

$$ltt(z) = \frac{3c^6(z)}{z^2}.$$

Then we have

$$\lim_{n\to\infty} \frac{LTT(n)}{C(n)} = \lim_{n\to\infty} \frac{[z^n]\frac{3c^6(z)}{z^2}}{[z^n]c(z)} = \lim_{z\to_{\mathbb{R}}\frac{1}{4}^-} \frac{\left(\frac{3c^6(z)}{z^2}\right)'}{c'(z)} = \frac{18}{2} = 9$$

Therefore

$$\frac{LT(n)}{F(n)} \leq \frac{2LTT(n)B(n-2)}{B(n)C(n)} \sim 18\frac{B(n-2)}{B(n)} \sim 18\left(\frac{e\log(n)}{n}\right)^2$$

□

The Theorem 1 is a direct consequence of Lemmas 2,3,4.

4 Discussion

Actually, in this paper much more is proved. The result obtained is not related only to intuitionistic tautologies, but it holds also in any logic which is able to prove simple tautologies. Indeed, all formulas of the form of simple tautologies are tautologies of every reasonable logic with this syntax. Therefore results comparing densities of any logic between minimal and classical one are the same. So the theorem proved may be applied as well to minimal, intuitionistic, and any intermediate logic. It shows, in fact, that a randomly chosen theorem has a proof which is a projection and statistically all true statements are the trivial ones.

In the paper only implicational fragment is taken in to consideration. Right now we do not know the analogous result for more complex syntax. But based on our experience we believe that the similar theorems holds for more complex syntaxes, including full propositional logic. At the moment, these are just expectations, but certainly it is worth to look in this direction.

Despite of the fact that all discussed problems and methods are solved by mathematical means, the paper, as was suggested by referees may have some philosophical interpretation and impact. However, the paper is purely technical and we are not ready to comment on these philosophical issues.

Acknowledgements. We are very grateful to all three anonymous referees who suggested many improvements to the presentation of our paper.

References

1. Chauvin, B., Flajolet, P., Gardy, D., Gittenberger, B.: And/Or trees revisited. Combinatorics, Probability and Computing 13(4-5), 475–497 (2004)
2. Flajolet, P., Sedgewick, R.: Analytic combinatorics: functional equations, rational and algebraic functions. In: INRIA, vol. 4103 (2001)
3. Flajolet, P., Sedgewick, R.: Analytic combinatorics. Book in preparation (2007), available at: `http://algo.inria.fr/flajolet/Publications/books.html`
4. Gardy, D.: Random Boolean expressions. In: Colloquium on Computational Logic and Applications. Proceedings in DMTCS, Chambery (France), June 2005, pp. 1–36 (2006)
5. Gardy, D., Woods, A.: And/or tree probabilities of Boolean function. Discrete Mathematics and Theoretical Computer Science, 139–146 (2005)
6. Fournier, H., Gardy, D., Genitrini, A., Zaionc, M.: Classical and intuitionistic logic are asymptotically identical. In: Duparc, J., Henzinger, T.A. (eds.) CSL 2007. LNCS, vol. 4646, pp. 177–193. Springer, Heidelberg (2007)
7. Kostrzycka, Z., Zaionc, M.: Statistics of intuitionistic versus classical logic. Studia Logica 76(3), 307–328 (2004)
8. Lefmann, H., Savický, P.: Some typical properties of large And/Or Boolean formulas. Random Structures and Algorithms 10, 337–351 (1997)
9. Matecki, G.: Asymptotic density for equivalence. Electronic Notes in Theoretical Computer Science 140, 81–91 (2005)
10. Moser, L., Wyman, M.: An asymptotic formula for the Bell numbers, Transactions of the Royal Society of Canada, XLIX (1955)
11. Moczurad, M., Tyszkiewicz, J., Zaionc, M.: Statistical properties of simple types. Mathematical Structures in Computer Science 10(5), 575–594 (2000)
12. Wilf, H.: generatingfunctionology, 3rd edn. A K Peters Publishers (2006)
13. Zaionc, M.: On the asymptotic density of tautologies in logic of implication and negation. Reports on Mathematical Logic 39, 67–87 (2005)
14. Zaionc, M.: Probability distribution for simple tautologies. Theoretical Computer Science 355(2), 243–260 (2006)