Asymptotyczne zachowanie się logik zdaniowych *

Marek Zaionc
Instytut Informatyki Uniwersytet Jagielloński,
Nawojki 11, 30-072 Kraków
zaionc@ii.uj.edu.pl

21 listopada 2001 roku

Streszczenie

W pracy badane są logiki zdaniowe pod względem ilościowym. Formuły pewnego rachunku logicznego dzielimy na frakcje w zależności od ich długości. Dla frakcji formuł o długości n analizujemy stosunek ilości formuł prawdziwych do ilości wszystkich formuł. Gdy przy n dążącym do nieskończoności ułamek ten posiada granicę nazywamy ją asymptotyczną gęstością prawdy. Istnienie asymptotycznej gęstości świadczy o jednostajnym rozkładzie formuł prawdziwych w zbiorze wszystkich formuł i dają obraz częstości ich występowania. Pozwala także na analizę pewnych efektów typowo logicznych metodami analitycznymi. Technika zliczania obiektów skończonych, przedstawiana w pracy, bazuje na typowym aparacie kombinatorycznym. Techniki dowodowe oparte są głównie na analizie osobliwości funkcji analitycznych jednej zmiennej zespolonej. Pomimo że dyskutowane tutaj pojęcia mają w znakomitej większości charakter typowo matematyczny wydaje się że analiza gęstości prawdy może mieć również pewien ważki aspekt filozoficzny będąc próbą odpowiedzi na pytanie na ile pojęcie prawdy jest zjawiskiem częstym w matematyce.

1 Zliczanie formuł

Nasze rozważania rozpoczniemy od logiki implikacji. Bardziej formalnie niech język $F^{\{\to\}}$ nad k zmiennymi zdaniowymi składa się z tychże zmiennych $\{a_1, \ldots, a_k\}$ i będzie domknięty na tworzenie implikacji.

Po pierwsze musimy ustalić sposób w jaki mierzona będzie długość formuły.

*Badania opisywane w tej pracy są wspomagane przez grant Komitetu Badań Naukowych numer 7T11C 022 21.

Przez $\|\phi\|$ rozumieć będziemy ilość wystąpień zmiennych zdaniowych w formule ϕ.

Mając miarę złożoności formuł $\|\phi\|$ można przyporządkowywać, klasom formuł $A \subset F^{\{\rightarrow\}}$, asymptotyczną gęstość rozważając ciąg ułamków w których liczniki podają ilość formuł o długości n z klasy A mianowniki natomiast są ilością wszystkich formuł o długości n. Przyjmijmy zatem jaka gęstość zbioru A w zbiorze $F^{\{\rightarrow\}}$ granicę, o ile taka istnieje, następującego ciągu ułamków.

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

Oczywiście nie jest trudno znaleźć takie klasy formuł które nie posiadają tak zdefiniowanej gęstości. Przykładowo klasa formuł o długościach parzystych jest właśnie takim przykładem.

Ażeby policzyć granice trzeba najpierw policzyć ilości formuł. Przez F_n^k będziemy rozumieli ilość formuł z naszego języka $F^{\{\rightarrow\}}$ która posiada długość n czyli:

$$F_n^k = \#\{\phi \in F^{\{\rightarrow\}} : \|\phi\| = n\} \tag{2}$$

Jak łatwo będzie zauważyć liczby F_n^k są zadane następującym schematem rekursji ze względu na n:

$$F_0^k = 0, \quad F_1^k = k, \tag{3}$$

$$F_n^k = \sum_{i=1}^{n-1} F_i^k F_{n-i}^k \tag{4}$$

Łatwo również zauważyć że $F_n^k = k^n C_n$ gdzie C_n jest n-tym wyrazem znanego w kombinatoryce ciągu Catalana który to ciąg można zdefiniować przy pomocy bardzo podobnego schematu rekursji.

$$C_0 = 0, \quad C_1 = 1, \tag{5}$$

$$C_n = \sum_{i=1}^{n-1} C_i C_{n-i} \tag{6}$$

Przy pomocy prostej indukcji łatwo dowiedziemy że $F_n^k = k^n C_n$. Oczywiście dla formuł zbudowanych z jednej zmiennej zdaniowej będziemy mieli $F_n^1 = C_n$.

Bardziej zaawansowane rozważania na temat ciągu Catalana można znaleźć w [5, str. 43–44] lub [3, str. 171]. W szczególności należy wspomnieć o dobrze znanej nierekurencyjnej formule na C_n.

$$C_n = \frac{1}{n}\binom{2n-2}{n-1}, \tag{7}$$

W naszych rozważaniach dotyczących zliczania formuł wprowadzimy jeszcze klasę formuł uzależnioną od pewnego parametru liczbowego $p \in N$. Będzie to klasa formuł posiadających p przesłanek. W dalszej części pracy postaramy się wyliczyć jaka jest *gęstość* takiej klasy w zależności od parametru p a co za ty idzie znaleźć rozkład prawdopodobieństwa.

Zatem przez formuła o p przesłankach rozumiemy formułę postaci $\tau = \tau_1 \to (\ldots \to (\tau_p \to \alpha))$, gdzie α jest już zmienną zdaniową. Niech $F_n^k(p)$ oznacza liczbę takich formuł o długości n w języku posiadającym k zmienne zdaniowe. Aby być w zgodzie z notacją poprzednio wprowadzoną niech liczba $C_n(p)$ oznacza $F_n^1(p)$. Łatwo zauważyć że $F_n^k(p) = k^n C_n(p)$. Spróbujmy zdefiniować kolejne liczby $F_n^k(p)$ przy pomocy schematu rekursji ze względu na p. Wygląda on następująco.

$$F_n^k(0) = \begin{cases} k & \text{if } n = 1 \\ 0 & \text{if } n \neq 1 \end{cases} \tag{8}$$

$$F_n^k(1) = \begin{cases} 0 & \text{if } n = 0 \\ kF_{n-1}^k & \text{if } n > 0 \end{cases} \tag{9}$$

$$F_n^k(p) = \sum_{i=1}^{n-p} F_i^k F_{n-i}^k(p-1). \tag{10}$$

Dowód przypominający uzasadnienie faktu (4) pominiemy.

Do naszych rozważań potrzebna nam będzie jeszcze jedna grupa formuł których ilości dla danego n będziemy starali się policzyć. Przez prostą tautologię rozumiemy formułę postaci $\tau_1 \to (\tau_2 \to (\ldots(\tau_n \to \alpha)\ldots))$ gdzie α jest już zmienną zdaniową w dodatku taką że przynajmniej jedna podformuła τ_i jest identyczna z α. Jest zupełnie oczywiste że prosta tautologia jest tautologią rachunku zdań. Niech więc zatem G_n^k będzie liczbą takich prostych tautologii o długości n. Analogicznie jak poprzednio niech również $G_n^k(p)$ będzie liczbą prostych tautologii o długości n które mają p przesłanek. Naszym celem będzie zmierzenie gęstości tak dobranych zbiorów. Zobaczymy że zbiór prostych tautologii będzie asymptotyczne *duży* w zbiorze wszystkich tautologii. Zobaczymy również że rozkład prawdopodobieństwa ilości przesłanek dla prostych tautologii jest diametralnie różny od tego samego rozkładu dla wszystkich formuł.

Zauważmy że liczba G_n^k jest zadana przez następujący schemat rekursji ze względu na n

$$\begin{aligned} G_1^k &= 0 && (11)\\ G_2^k &= k && (12)\\ G_n^k &= \sum_{i=2}^{n-1} F_{n-i}^k G_i^k + (F_{n-1}^k - G_{n-1}^k) && (13) \end{aligned}$$

Takiej samej obróbce kombinatorycznej poddamy liczby $G_n^k(p)$ i otrzymamy:

$$G_n^k(0) = \begin{cases} k & \text{if } n = 1 \\ 0 & \text{if } n \neq 1 \end{cases} \tag{14}$$

$$G_n^k(p+1) = \begin{cases} 0 & \text{if } n \leqslant p \\ \sum_{i=2}^{n-1} F_{n-i}^k G_i^k(p) + (F_{n-1}^k(p) - G_{n-1}^k(p)) & \text{if } n > p \end{cases} \tag{15}$$

2 Kombinatoryka i lemat Szegö

Podstawowym narzędziem z którym będziemy mieli do czynienia w badaniach nad asymptotycznym zachowaniem się logik zdaniowych są *funkcje tworzące*. Naszym podstawowym zadaniem będzie wyliczanie granic pewnych ciągów zadanych rekurencyjnie. Kombinatoryka wytworzyła niezwykle praktyczne i potężne narzędzie do tych celów w formie *funkcji tworzących* i *szeregow tworzących*

Niech $A = (A_0, A_1, A_2, \ldots)$ będzie ciągiem liczb rzeczywistych. Szereg formalny $\sum_{n=0}^{\infty} A_n z^n$ będzie nazywany szeregiem tworzącym dla ciągu A. Z drugiej strony patrząc na z jako na zmienna zespoloną, można analizować funkcje zespoloną $f_A(z)$ zakładając że szereg ten jest jednostajnie zbieżny w pewnym otwartym dysku $\{z \in C : |z| < R\}$, gdzie $R \geqslant 0$ jest nazywany promieniem zbieżności. Łatwo zauważyć odpowiedniość miedzy ciągami a funkcjami tworzącymi (chyba że $R = 0$). Mając funkcje tworząca dla danego ciągu A możemy odzyskać elementy tego ciągu biorąc

$$A_n = \frac{1}{n!} \frac{d^n f}{dz^n}(z_0). \tag{16}$$

Wiele problemów dotyczących asymptotycznego zachowania się ciągów może być rozwiązywanych przy pomocy analizy funkcji tworzących dla tych ciągów.

Kluczowym narzędziem dla naszych rozważań będzie następujący rezultat zwany lematem Szegö ([4, Thm. 8.4] porównaj również [5, Thm. 5.3.2]). Dla naszych potrzeb będziemy stosowali jedynie uproszczoną jego wersje. W twierdzeniu tym symbol $[z^n]\{F\}$ oznacza współczynnik stojący przy z^n w rozwinięciu funkcji F. Uogólniony symbol Newtona $\binom{a}{n}$ dla dowolnej liczby zespolonej a we wzorze (18) oznacza $a(a-1)\dots(a-(n-1))/n!$.

Lemat. 1 *Niech $v(z)$ będzie funkcją analityczną na okręgu jednostkowym $|z|<1$ posiadającą na okręgu $|z|=1$ jedyna osobliwość w punkcie $z=1$. Jeżeli funkcja $v(z)$ w otoczeniu $z=1$ posiada rozwinięcie w postaci*

$$v(z)=\sum_{p\geqslant 0} v_p(1-z)^{p/2}, \tag{17}$$

to

$$[z^n]\{v(z)\}=v_1\binom{1/2}{n}(-1)^n+O(n^{-2}). \tag{18}$$

Lemat Szegö daje nam potężne narzędzie do badania zachowania się ciągów, zważywszy że dowolny element ciągu może być wyrażony i oszacowany z dokładnością $O(n^{-2})$ przy pomocy elementu pierwszego rozwinięcia funkcji tworzącej dla tego ciągu. Jeżeli zatem będziemy mieli taka sytuacje że mamy dwa ciągi $\{D_n\}$ i $\{F_n\}$ których odpowiednie funkcje tworzące spełniają lemat Szegö to:

$$\frac{D_n}{F_n}=\frac{\left(v_1\binom{1/2}{n}(-1)^n+O(n^{-2})\right)}{\left(w_1\binom{1/2}{n}(-1)^n+O(n^{-2})\right)}=\frac{v_1}{w_1}(1+o(1)) \tag{19}$$

Gdzie v_1 i w_1 są właśnie pierwszymi wyrazami rozwinięcia funkcji tworzących dla ciągów $\{D_n\}$ i $\{F_n\}$. W takim przypadku jesteśmy w stanie stwierdzić że granica ułamka $\frac{D_n}{F_n}$ istnieje oraz jest następująca:

$$\lim_{n\to\infty}\frac{D_n}{F_n}=\frac{v_1}{w_1} \tag{20}$$

3 Gęstość prawdy w logice klasycznej

Rozważymy poniżej logikę klasyczną w której występuje jedynie jedna zmienna zdaniowa oraz funktor implikacji. Mamy zatem C_n formuł o długości n. Pierwszą rzeczą którą należy zrobić to znaleźć funkcję tworzącą f_C dla tego ciągu. Jest to łatwe. Zauważmy mianowicie, że rekurencja zadana przez $\sum_{i=1}^{n-1} C_i C_{n-i}$ odpowiada mnożeniu szeregów formalnych. Otrzymujemy zatem następujące równanie:

$$f_C(z) = f_C(z)^2 - z \tag{21}$$

Część liniowa $-z$ w tym równaniu jest związana z początkowym elementem rekursji $C_1 = 1$. Rozwiązaniem równania kwadratowego wraz z warunkiem brzegowym $f_C(0) = 0$ (który odpowiada zerowemu elementowi ciągu $C_0 = 0$) daje jedno rozwiązanie

$$f_C(z) = \frac{1}{2} - \frac{1}{2}\sqrt{1 - 4\,z}. \tag{22}$$

Następnie znajdujemy rekurencyjny opis ilości tautologii omawianej logiki. Do tego celu rozwarstwimy zbiór wszystkich formuł na dwa podzbiory. Niech $C_n = T_n + N_n$ gdzie T_n jest ilością tautologii a N_n ilość nietautologii o długości n. Rekurencyjny opis ilości tautologi i nietautologii może zostać łatwo odczytany z następującego prostego lematu

Lemat. 2 *$\varphi \to \psi$ nie jest tautologią wtw φ jest tautologią i ψ nie jest tautologią*

Dowód tego faktu wynika prosto z analizy tabelki implikacji. Należy tutaj dodać, że lemat powyższy jest wyłącznie prawdziwy dla logiki z jedną zmienną zdaniową. Przy większej ilości zmiennych tak nie jest. Lemat 2 łatwo daje się przełożyć na rekurencyjną zależność pomiędzy T_n, N_n oraz wcześniej zdefiniowane C_n.

$$N_0 = 0, \;\; N_1 = 1, \;\; N_n = \sum_{i=1}^{n-1} T_i N_{n-i} \tag{23}$$

$$T_0 = 0, \;\; T_1 = 0, \;\; T_n = C_n - N_n \tag{24}$$

Jeżeli zatem przez f_N oraz f_T oznaczymy odpowiednio funkcje tworzące dla zbioru *nietautologii* i *tautologi* otrzymamy układ równań w którym f_C jest znaną już tworzącą dla liczb Catalana:

$$f_N(z) = f_N(z)(f_C(z) - f_N(z)) + z \tag{25}$$

$$f_T(z) = f_C(z) - f_N(z). \tag{26}$$

Rozwiązując ten układ z warunkiem brzegowym $f_N(0) = 0$ dostajemy

$$
\begin{aligned}
f_N(z) &= -\tfrac{1}{4} - \tfrac{1}{4}\sqrt{1-4\,z} + \tfrac{1}{4}\sqrt{2+2\sqrt{1-4\,z}+12\,z} && (27)\\
f_T(z) &= \tfrac{3}{4} - \tfrac{1}{4}\sqrt{1-4\,z} - \tfrac{1}{4}\sqrt{2+2\sqrt{1-4\,z}+12\,z}, && (28)
\end{aligned}
$$

Funkcje f_C oraz f_T, f_N maja osobliwość w punkcie $z = 1/4$. Aby funkcje te spełniały założenia lematu Szegö (1) należy je rozkalibrować przez podstawienie $t = z/4$ tak, aby najbliższy zeru punk osobliwości znalazł się w pozycji $z = 1$. Dlatego też po tym podstawieniu otrzymujemy.

$$
\begin{aligned}
f_C(z/4) &= \frac{1}{2} - \frac{1}{2}\sqrt{1-z} && (29)\\
f_T(z/4) &= \frac{3}{4} - \frac{1}{4}\sqrt{1-z} - \frac{1}{4}\sqrt{2+2\sqrt{1-z}+3\,z}. && (30)
\end{aligned}
$$

Funkcje te mają rzeczywiście najbliższą ososbliwość w punkcie $z = 1$. Ponadto funkcje $f_T(z/4)$ i $f_N(z/4)$ mają jeszcze osobliwość poza kołem jednostkowym w punkcie $z = -16/9$. Rozwijając $f_T(z/4)$ w otoczeniu osobliwości $z = 1$ otrzymujemy:

$$
f_T(t/4) = \left(\frac{3}{4} - \frac{1}{4}\sqrt{5}\right) + \left(-\frac{1}{4} - \frac{1}{20}\sqrt{5}\right)\sqrt{1-t} + \ldots \tag{31}
$$

Zgodnie zatem z lematem Szegö otrzymujemy

$$
\begin{aligned}
C_n &= \left(-\frac{1}{2}(-1)^n \binom{1/2}{n} + O(n^{-2})\right) 4^n && (32)\\
T_n &= \left(\left(-\frac{1}{4} - \frac{\sqrt{5}}{20}\right)(-1)^n \binom{1/2}{n} + O(n^{-2})\right) 4^n && (33)
\end{aligned}
$$

Czynniki 4^n na końcu każdego wyrażenia są efektem zastosowania naszego podstawienia. Dlatego też dzięki obserwacjom uczynionym w poprzednim rozdziale i dzięki wzorowi (20) łatwo teraz uzyskać analityczny wzór na interesującą nas granice.

$$
\lim_{n\to\infty} \frac{T_n}{C_n} = \frac{\left(-\frac{1}{4} - \frac{1}{20}\sqrt{5}\right)}{-\frac{1}{2}}(1+o(1)) = \frac{1}{2} + \frac{\sqrt{5}}{10} \approx 0.7236067978\ldots \tag{34}
$$

Jak widać gęstość prawdy w tym przypadku jest zadziwiająco wysoka. Mając zatem długą losowo wybraną formułę tej logiki mamy olbrzymie szanse przekraczając 72% na to by była ona prawdziwa. Co ciekawsze identyczną granice uzyskamy dla logiki intuicjonistycznej implikacji z jedną zmienną z czego daje się łatwo pokazać że właśnie dla jednej zmiennej logiki te są identyczne.

Można również udowodnić, aczkolwiek jest to trudne, że w logice implikacji zwiększając ilość zmiennych zdaniowych zawsze uzyskamy prawo asymptotycznej zbieżności dla tautologii w sensie logiki klasycznej.

Twierdzenie. 3 *W klasycznej logice implikacyjnej dla dowolnego k będącego ilością zmiennych zdaniowych zbiór tautologii ma prawo asymptotycznej zbieżności. Co więcej dla każdego k asymptotyczna gęstość prawdy jest dodatnia natomiast ciąg tych gęstości dąży do 0 wraz k dążącym do nieskończoności.*

Rozważmy teraz logikę klasyczną opartą na implikacji i negacji dla jednej zmiennej zdaniowej. Wyniki tej części są zaczerpnięte głownie z pracy [6]. Tym razem funkcje tworzące dla odpowiednich ciągów są znacznie bardziej skomplikowane. Dlatego też poniżej przytoczymy jedynie wyniki umożliwiające porównanie z wynikami poprzedniego przykładu. W szczególności pokazujemy tutaj, że gęstość prawdy w logice z negacją jest istotnie mniejsza niż logiki czysto implikacyjnej.

Twierdzenie. 4 *W logice klasycznej jednej zmiennej zdaniowej z implikacją i negacją zbiór tautologii ma prawo zbieżności. Gęstość prawdy wynosi:*

$$\begin{aligned} \lim_{n\to\infty} \frac{T_n}{F_n^{\{\to,\neg\}}} &= \frac{1}{(4\sqrt{13})} + \frac{1}{(4\sqrt{17})} + \\ &+ \frac{1}{2\sqrt{2\left(\sqrt{221}-9\right)}} + \frac{15}{2\sqrt{442\left(\sqrt{221}-9\right)}} \\ &= 0.423238538401941... \end{aligned} \tag{35}$$

Jak widać asymptotyczna zawartość formuł prawdziwych w logice $F^{\{\to,\neg\}}$ jest duża, wynosząca około 42% ale znacząco mniejsza niż w logice $F^{\{\to\}}$. Jest to spowodowane dodaniem funktora negacji który obniżył zawartość prawdy w logice.

Jednocześnie jesteśmy w stanie pokazać dwa twierdzenia opisujące relacje ilościowe omawianych logik.

Twierdzenie. 5 *Implikacyjny fragment logiki $F^{\{\to,\neg\}}$ jest asyptotycznie pusty.*

Twierdzenie. 6 *Prawdopodobieństwo warunkowe znalezienia implikacyjnej tautologii pod warunkiem że jest to tautologia implikacyjno-negacyjna wynosi 0.*

4 Rozkłady prawdopodobieństw, typowe formuły, typowe formuły prawdziwe.

W tej części przytoczymy wyniki dotyczące wyglądu typowej formuły implikacyjnej w zależności od pewnych jej cech ilościowych jak również pokażemy jak wyglądają typowe formuły prawdziwe w rozkładzie względem tej samej cechy. Rozważamy tutaj implikacyjne formuły zbudowane z k zmiennych.

Następujące twierdzenie ustala asymptotyczną wielkość frakcji formuł o p przesłankach w stosunku do wszystkich formuł.

Twierdzenie. 7 *Niech $p > 0$ i $k > 0$ będą ustalone.*

$$\lim_{n \to \infty} \frac{F_n^k(p)}{F_n^k} = \frac{p}{2^{p+1}}$$

Konsekwencją tego twierdzenia jest ustalenie rozkładu prawdopodobieństwa wystąpienia formuły o p przesłankach. Niech X będzie zmienną losową która przypisuje formule ilość jej przesłanek. Zmienna ta ma następujący rozkład:

Twierdzenie. 8

$$X(p) = \begin{cases} 0 & \text{\textit{if} } p = 0 \\ \frac{p}{2^{p+1}} & \text{\textit{if} } p > 1 \end{cases}$$

Jak łatwo policzyć wartość oczekiwana zmiennej X wynosi

$$E(X) = \sum_{p=1}^{\infty} p \frac{p}{2^{p+1}} = 3,$$

wariancja zaś

$$D^2(X) = E((X - E(X))^2) = E(X^2) - (E(X))^2 = \sum_{p=1}^{\infty} p^2 \frac{p}{2^{p+1}} - 9 = 4,$$

w takim razie odchylenie standardowe X jest 2.

Mamy zatem dowód na to, że typowe formuły implikacyjne powinny mieć 3 przesłanki z odchyleniem standardowym ± 2. Asymptotyczna ilość formuł których ilość przeslanek zawiera się miedzy 1 a 5 wynosi jak łatwo policzyć z rozkładu X tyle co $57/64$ czyli ponad 89%.

Zupełnie inaczej wygląda podobna analiza dla formuł prawdziwych. W następnej części tego rozdziału podamy podobne wyniki dla pewnej znaczącej grupy tautologii nazwanej wcześniej prostymi tautologiami. Klasa ta jest jednak na tyle liczna że asymptotyczne wyniki jej dotyczące będą reprezentatywne dla całego zbioru tautologii. Pokażemy poniżej jaki jest rozkład prawdopodobieństwa ilości przesłanek w przypadku prostych tautologii. Rozkład ten będzie istotnie różny od rozkładu dla wszystkich formuł. Dowodzi to tego, że tak prosta cecha jak ilość przesłanek w losowo wybranej formule, może świadczyć o jej prawdziwości. Poniższe twierdzenia ze względu na złożoność i długość podane zostaną bez dowodu. Większość poniższego materiału została zaadoptowana z pracy [7]. Dowód twierdzenia 9 jest również konsekwencją lematu Szegö w wersji dokładnie takiej jak w lemacie 1. Następujące twierdzenie ustala asymptotyczną wielkość frakcji prostych tautologii o p przesłankach w stosunku do wszystkich prostych tautologii w implikacyjnej logice o k zmiennych zdaniowych.

Twierdzenie. 9 *Niech $p > 0$ i $k > 0$ będą ustalone. Granica $\frac{G_n^k(p)}{G_n^k}$ istnieje oraz:*

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

Tak jak i w przypadku zbioru formuł opisanym twierdzeniami 7 i 8 i tutaj możemy policzyć wartość oczekiwaną rozkładu ilości przesłanek dla prostych tautologii.

Twierdzenie. 10 *Niech X będzie zmienną losową o rozkładzie takim jak w twierdzeniu 9. Wartość oczekiwana zmiennej X wynosi*

$$E(X) = \sum_{p=1}^{\infty} p\frac{(2k+1)^2}{4k+1}\left(\frac{p}{2^{p+1}} - p\frac{(2k-1)^{p-1}}{4^p k^{p-1}}\right) = \frac{40k^2+18k+3}{(2k+1)(4k+1)}$$

Warto zwrócić uwagę że $\lim_{k\to\infty} \frac{40k^2+18k+3}{(2k+1)(4k+1)} = 5$ co pokazuje że w odniesieniu do prostych tautologii zbudowanych z dużej liczby zmiennych zdaniowych otrzymujemy wartość oczekiwaną ilości przesłanek bliską 5 co jest zdecydowanie różne od tej która przysługuje wszystkim formułom. (patrz twierdzenie 8). Dokładne badania dotyczące granicznego rozkładu tego prawdopodobieństwa są zawarte w pracy [7].

Zestawiając twierdzenie 9 ze wzorem na asymptotyczną gęstość prostych tautologii z pracy [2] (patrz Theorem 6.3 strona 590) który stanowi że

Twierdzenie. 11

$$\lim_{n\to\infty} \frac{G_n^k}{F_n^k} = \frac{4k+1}{(2k+1)^2}$$

oraz z faktem opisywanym w twierdzeniu 7 otrzymujemy bardzo elegancki opis wielkości frakcji prostych tautologii o p przesłankach w stosunku do wszystkich formuł o p przesłankach.

Twierdzenie. 12 *Niech $p > 1$ i $k > 0$ będą ustalone. Granica $\frac{G_n^k(p)}{F_n^k(p)}$ istnieje oraz:*

$$\lim_{n\to\infty} \frac{G_n^k(p)}{F_n^k(p)} = 1 - \left(\frac{2k-1}{2k}\right)^{p-1}$$

Łatwo zauważyć że zależność powyższa jest niezwykle pomocna przy probabilistycznym ocenianiu szans tautologiczności losowo wybranej formuły. Łatwo można sobie wyobrazić że sprawdzenie prawdziwości formuł bardzo długich może być obliczeniowo ekstremalnie trudne. Wiadomo bowiem że problem tautologiczności formuł leży w klasie złożoności $coNP$ a zatem możliwości automatyzowania testu na tautologiczność są bardzo ograniczone. Dlatego też warto stosować algorytm probabilistyczny którego podstawą może być twierdzeniem 12. Otóż, algorytm ten dla zadanej formuły zlicza jedynie jej przesłanki. Można to zrobić bardzo szybko w czasie logarytmicznie zależnym od długości formuły. Dla długich formuł szansa na to że formuła jest tautologią jest bliska $1 - \left(\frac{2k-1}{2k}\right)^{p-1}$. Należy zwrócić uwagę na to że wyrażenie to przy ustalonym k zbliża się do 1 wraz z tym jak ilość otrzymanych przesłanek dąży do nieskończoności. Dla długich formuł o dużej ilości przesłanek możemy być bardzo bliscy pewności. Innymi słowy dla każdego $\varepsilon > 0$ istnieje efektywnie takie p że wśród formuł o p przesłankach prawie wszystkie formuły, za wyjątkiem niewielkiej frakcji o rozmiarze ε, są tautologiami. Możemy mieć zatem tak dużo tautologii jak tylko chcemy w miarę wzrostu p.

Literatura

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

[2] M. Moczurad, J. Tyszkiewicz, M. Zaionc "Statistical properties of simple types " Mathematical Structures in Computer Science, vol 10 (2000), pp 575-594.

[3] Lipski W, Marek W, (1986) Analiza Kombinatoryczna *PWN*, Warszawa.

[4] Szegö, G. (1975) Orthogonal polynomials. Fourth edition.*AMS, Colloquium Publications*, **23**, Providence.

[5] Wilf, H.S. (1994) Generatingfunctionology. Second edition.*Academic Press*, Boston.

[6] M. Zaionc (2001), Negative impact of negation on the density of truth, w druku.

[7] M. Zaionc (2001), Probability distribution for simple tautologies, w przygotowniu