Metrické vlastnosti boolovských funkcií s daným počtom jednotiek

Magdaléna Kovalíková

2006

Metrické vlastnosti boolovských funkcií s daným počtom jednotiek

DIPLOMOVÁ PRÁCA

Magdaléna Kovalíková

UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY
KATEDRA INFORMATIKY

Dipl. vedúci: doc. RNDr. Eduard Toman, CSc.

BRATISLAVA 2006

Týmto prehlasujem, že som diplomovú prácu vypracovala samostatne s odbornou pomocou školiteľa a s použitím uvedenej literatúry.

Bratislava, máj 2006 Magdaléna Kovalíková

Abstrakt

Skúmame náhodné boolovské funkcie n premenných, ktorých počet jednotiek je daný funkciou $m(n)$, $0 \leq m(n) \leq 2^n$. Táto práca prezentuje asymptotické odhady dĺžky hlavnej a skrátenej disjunktívnej normálnej formy náhodnej boolovskej funkcie. Naviac sú detailne skúmané rády prostých implikantov.

Kľúčové slová: Náhodné boolovské funkcie, disjunktívne normálne formy, minimalizácia, pravdepodobnostné metódy.

Obsah

Kapitola 1

Úvod

Každú boolovskú (logickú) funkciu $f \neq 0$ možno realizovať pomocou tzv. disjunktívnej normálnej formy (DNF). Disjunktívna normálna forma je výraz, ktorý vznikne disjunkciou rôznych elementárnych konjunkcií nad abecedou boolovských premených $\{x_1, \ldots, x_n\}$, pričom v jednej elementárnej konjunkcii sú všetky písmená rôzne a dĺžka konjunkcie sa nazýva rádom elementárnej konjunkcie. Elementárnu konjunkciu nazývame tiež implikant.
Funkcia ale môže byť vyjadrená vo forme DNF mnohými spôsobmi. Len pre ilustráciu, počet DNF nad abecedou z n písmen je rovný 2^{3^n}, ale počet rôznych boolovských funkcií nad n-prvkovou abecedou je len 2^{2^n}. Pre boolovské funkcie teda má zmysel hľadať DNF, ktoré majú minimálnu zložitosť. Disjunktívnu normálnu formu charakterizuje index jednoduchosti, ktorý obvykle vyjadruje buď počet písmen alebo počet elementárnych konjunkcií v zápise DNF. Minimálnu zložitosť má teda DNF, ktorá je minimálna vzhľadom na nejaký index jednoduchosti.
Skúmanie v oblasti minimalizácie je stále aktuálne, avšak počiatky siahajú

do rokov päťdesiatym a existuje aj niekoľko skorších prác.

Cieľom mojej práce bude sformulovať a dokázať tvrdenia o aymptotických odhadoch zložitostí rôznych disjunktívnych normálnych foriem náhodnej boolovskej funkcie. Náhodné boolovské funkcie môžu závisieť od parametrov, v našom prípade tým parametrom bude počet jednotiek $m(n)$ a budú tvoriť špeciálnu triedu.[1] Predovšetkým budeme odhadovať dĺžku hlavnej DNF a skrátenej DNF náhodnej boolovskej funkcie.

Práca by mohla pokračovať asymptotickými odhadmi dĺžky a počtu iredundantných DNF, odhadom počtu regulárnych vrcholov NBF vzhľadom na pokrytie množiny N_f. Vrcholy a vrcholové pokrytia náhodnej boolovskej funkcie súvisia s jej geometrickou reprezentáciou ako náhodného grafu. Ďalej by sa dalo pokračovať odhadom veľkosti jadra - počtu jadrových vrcholov NBF a určovaním prahových funkcií pre ich výskyt.

Náhodná boolovská funkcia n premených je zobrazenie $D^n \to D$, $D = \{0, 1\}$. V triede boolovských funkcií F_m nadobúda hodnoty 0 a 1 v závislosti od daného parametra m tak, že počet jednotiek je práve m spomedzi všetkých bodov (n-tíc) D^n. Pritom parameter m môže byť konštantný údaj, ale obvykle predpokladáme, že je závislý od n.

Známe minimalizačné algoritmy pracujú vo viacerých krokoch. Najskôr sa skonštruuje skrátená DNF využijúc nejakú reprezentáciu boolovskej funkcie, napr. tabuľku alebo úplnú DNF. Zmazaním nadbytočných implikantov zo skrátených DNF rôznymi spôsobmi získame iredundantné DNF. Nakoniec hľadaním medzi všetkými iredundantnými DNF nájdeme DNF s minimálnou zložitosťou (vzhľadom na nejaký index jednoduchosti).

[1] Ak m bude konštanta, bude to explicitne povedané, inak $m = m(n)$.

Vyšetrovanie hraníc parametrov nám umožňuje rozhodnúť, aké zložité sú jednotlivé kroky minimalizácie pre náhodný prípad.

Vo všeobecnom modeli, sa odhadujú zložitosti na celej množine B_n boolovských funkcií n premenných a náhodný výber boolovskej funkcie závisí od parametra p_n. Ten predstavuje pravdepodobnosť, že vrchol patrí generovanej náhodnej boolovskej funkcii, t.j., že má hodnotu 1. Obvyklý prípad našich výsledkov, keď $p_n = \frac{1}{2}$ získali Glagolev a Sapošenko. Všeobecný tvar tohoto problému bol preskúmaný Weberom. Podstatne budeme využívať výsledky docenta Škovieru.

V súvislosti s náhodnými boolovskými funkciami bolo napísané značné množstvo článkov.

Kapitola 2

Predbežné kroky

Geometrická formulácia

Uprednostňujeme geometrický prístup k minimalizačnému problému pred algebraickým, pretože je názornejší.

Na množinu všetkých binárnych n-tíc D^n sa pozeráme ako na množinu všetkých vrcholov n-rozmernej jednotkovej kocky Q_n. Boolovskej funkcii f priradíme podmnožinu N_f vrcholov tejto kocky . Dva vrcholy $\alpha, \beta \in D^n$ sú susedné v Q_n práve vtedy, keď sa líšia v práve jednej súradnici. Teda Q_n má 2^n vrcholov a $n * 2^{n-1}$ hrán. Funkcia f je teda identifikovaná s podgrafom indukovaným na množine $f^{-1}(1) = \{\alpha \in D^n; f(\alpha) = 1\} = N_f$.

Ľahko sa ukáže, že pre každý implikant (tiež nazývaný elementárna konjunkcia) $h = x_{i_1}^{\sigma_1} \wedge x_{i_2}^{\sigma_2} \wedge \ldots \wedge x_{i_k}^{\sigma_k}$ rádu k s $i_1 < i_2 < \ldots < i_k$ a $0 \leq k \leq n$. Množina N_h indukuje $(n-k)$-rozmernú podkocku Q_n a naopak. Tiež ak h_1, h_2 sú elementárne konjunkcie, potom $N(h_1 \vee h_2) = N_{H_1} \cup N_{H_2}$. Preto existuje priama súvislosť medzi disjunktnými normálnymi formami funkcie

f a vrcholovými pokrytiami množiny N_f podkockami obsiahnutými v N_f. Voláme ich intervaly f.

Z tohoto pohľadu je štúdium náhodných boolovských funkcií úzko spojené so štúdiom podkockových pokrytí náhodných indukovaných podgrafov v n-kocke Q_n.

Základné pojmy

Nech $K \subseteq N_f$ je interval f. Hovoríme, že K je maximálny, ak neexistuje žiadny interval L taký, že $K \subset L \subseteq N_f$. Elementárna konjunkcia odpovedajúca maximálnemu intervalu sa nazýva prostý implikant.

Teraz je už ľahké podať presné definície základných typov DNF boolovskej funkcie f. Hlavná (alebo úplná) DNF zodpovedá pokrytiu N_f 0-rozmernými intervalmi, t.j. jednoprvkovými množinami. Jej dĺžka, čiže počet implikantov, je rovná $|N_f|$.

Skrátená DNF odpovedá pokrytiu N_f všetkými maximálnymi intervalmi f. Keďže môže byť medzi nimi veľa zbytočných implikantov, motivuje nás to k nasledujúcej definícii: Pokrytie U množiny N_f sa bude nazývať iredundantné, ak žiadna vlastná podmnožina U nepokrýva N_f. Iredundantná DNF je tá, ktorá odpovedá iredundantnému pokrytiu.

Pre pojmy, týkajúce sa boolovských funkcií, disjunktívnych normálnych foriem a minimalizačného problému, odkazujeme na zdroj [2].

Trieda F_m

Skúmaným modelom je trieda boolovských funkcií $F_m = \{f \in B_n, |N_f| = m\}$, kde B_n je množina boolovských funkcií n premenných $f : \{0,1\}^n \to \{0,1\}$. Vyberáme m-ticu vrcholov, ktorým bude priradená jednotka. Preto počet funkcií v triede F_m je $\binom{2^n}{m}$ Použijeme niektoré základné fakty z teórie pravdepodobnosti. Náhodný výber funkcie $f \in F_m$ si môžeme predstaviť ako náhodné rozdelenie prvkov z D^n do dvoch disjunktných podmnožín - N_f a $D^n - N_f$, pričom obe majú vždy konštantný počet prvkov pre dané m. Keďže pravdepodobnosť výberu každej funkcie $f \in F_m$ je rovnaká, výber nastáva s pravdepodobnosťou

$$\frac{1}{\binom{2^n}{|N_f|}} = \frac{1}{\binom{2^n}{m}} = P(\{f\}).$$

Následne pre ľubovoľnú podmnožinu boolovských funkcií $A \subseteq F_m$ priradíme $P(A) = \sum_{f \in A} P(\{f\})$.
Použijúc binomickú vetu ľahko získame jednoduchý, ale užitočný výsledok.

Tvrdenie 1 *Nech U a V sú disjunktné podmnožiny $D^n = \{0,1\}^n$, $|U| \leq m$, $|V| \leq 2^n - m$ a $F = \{f \in F_m; U \subseteq N_f; V \subseteq D^n - N_f\}$. Potom*

$$P(F) = P(U \subseteq N_f \wedge V \subseteq D^n - N_f) = \frac{\binom{2^n - (|U|+|V|)}{m-|U|}}{\binom{2^n}{m}}.$$

Dôkaz. Počet funkcií v množine F je rovný počtu rôznych priradení jednotiek vrcholom D^n. Platí, že $|U|+|V|$ hodnôt je fixovaných a ostáva priradiť ešte $m-|U|$ jednotiek. To je možno urobiť $\binom{2^n-(|U|+|V|)}{m-|U|}$ spôsobmi. Správnosť je zrejmá.

Z toho dôvodu, ak K je k-rozmerná podkocka D^n, tak $P(K \subseteq N_f) = \frac{\binom{2^n - 2^k}{m - 2^k}}{\binom{2^n}{m}}$.

Priestor (F_m, P)

Pravdepodobnostný priestor (F_m, P) má pre nás kľúčový význam. Totiž, že v tomto priestore vyšetrujeme rôzne vlastnosti boolovských funkcií a všetky parametre boolovských funkcií sú uvažované ako náhodné premenné na (F_m, P). Predpokladáme, že všetky tu použité náhodné premenné sú celočíselné a nezáporné.

Nech S je určitá vlastnosť. Ak $\lim_{n\to\infty} P(f$ má vlastnosť $S) = 1$, hovoríme, že náhodná boolovská funkcia má vlastnosť S, alebo, že boolovská funkcia spĺňa vlastnosť S skoro isto.

Aj v triede F_m môžme označiť ako $p(n)$pravdepodobnosť, že vrchol patrí do N_f. Nie je to ale pevne dané číslo alebo funkcia ako v B_n, ale iba asymptotické ohraničenie podielu funkcií $\frac{\binom{2^n-1}{m(n)-1}}{\binom{2^n}{m}} \leq \frac{m(n)}{2^n}$. Zjavne závisí od m a n a získame ho použitím Formuly 1, ktorú využívame vo veľkom v Kapitole 3. A predsa je rozumné predpokladať, že postupnosť $(\frac{m}{2^n})_{n=1}^{\infty}$ konverguje k nejakému číslu p. Ak $p = 0$ alebo $p = 1$, musíme dať nejaké obmedzenia na rast postupností $\frac{1}{\frac{m}{2^n}}$ a $\frac{1}{\frac{m}{2^n}}$.

Pravdepodobnostné metódy

$E(X)$ a $D(X) = E(X - E(X))^2$ označuje strednú hodnotu a disperziu náhodnej premennej X. Dobre známa Markovova a Čebyševova nerovnosť využívajúce $E(X)$ a $D(X)$ nám ponúkajú najdôležitejší nástroj na získanie asymptotických odhadov.

Tvrdenie 2 (Markovova nerovnosť) *Ak $X \geq 0$ je nezáporná náhodná premenná a $\varepsilon > 0$ je kladné reálne číslo, potom*

$$P(X \geq \varepsilon) \leq \frac{E(X)}{\varepsilon}. \tag{2.1}$$

Tvrdenie 3 (Čebyševova nerovnosť) *Pre každú náhodnú premennú X a $\varepsilon > 0$ platí nasledujúca nerovnosť:*

$$P(|X - E(X)| \geq \varepsilon) \leq \frac{D(X)}{\varepsilon^2}. \tag{2.2}$$

Všetky limity a asymptoty uvažujeme pre $n \to \infty$, preto tento symbol vynechávame. Používame O-notáciu. Symbol $o(a_n)$ znamená výraz, ktorý ide k 0, keď sa delí a_n a $O(a_n)$ výraz, ktorý zostáva ohraničený, ak sa delí a_n. Postupnosti (a_n) a (b_n) sú asymptoticky ekvivalentné, ak $\lim \frac{a_n}{b_n} = 1$. Obvyklé označenie je $a_n \sim b_n$.

Symbol $\log_a x$ označuje logaritmus pri základe a. Ak $a = 2$, označenie vynechávame.

Kapitola 3

Intervaly

Aby sme získali odhady zložitostí hlavných a skrátených DNF a neskôr podobné výsledky pre iredundantné DNF, začneme štúdiom štruktúry intervalov v náhodnej boolovskej funkcii.

Nech $i_{n,k}$ označuje náhodnú premennú na F_m takú, že $i_{n,k}(f)$ je rovné počtu k-rozmerných intervalov funkcie $f \in F_m$. Prvý krok spočíva vo výpočte a odhade strednej hodnoty a disperzie.

Tvrdenie 4 $Ei_{n,k} = \binom{n}{k}.2^{n-k}.\frac{\binom{2^n-2^k}{m-2^k}}{\binom{2^n}{m}}$

Dôkaz. Pre každú k-rozmernú podkocku K kocky D^n zavedieme náhodnú premennú η_K (niekedy ju voláme aj indikátor) definovanú neasledovne:

$$\eta_K(f) = \begin{cases} 1, & \text{ak } K \subseteq N_f; \\ 0, & \text{inak.} \end{cases}$$

Zrejme $i_{n,k} = \sum_K \eta_K$, pričom sumácia ide cez všetky k-rozmerné podkocky D^n.

Z Tvrdenia 1 vieme, že

$$\frac{\binom{2^n-2^k}{m-2^k}}{\binom{2^n}{m}} = P(K \subseteq N_f) = P(\eta_K = 1) = E\eta_K$$

Existuje $\binom{n}{k}.2^{n-k}$ k-rozmerných podkociek D^n. Teda

$$Ei_{n,k} = \sum_K \eta_K = \binom{n}{k}.2^{n-k}.\frac{\binom{2^n-2^k}{m-2^k}}{\binom{2^n}{m}}.$$

Tvrdenie 5 $Di_{n,k} \leq \binom{n}{k}^2.2^{n-k}.\left(\frac{m}{2^n}\right)^{2^k}$

Dôkaz. Na dôkaz tohoto horného ohraničenia si najskôr pripomenieme, že $Di_{n,k} = E(i_{n,k})^2 - (Ei_{n,k})^2$. Keďže $i_{n,k}$ je vyjadrené ako suma indikátorov η_K, máme $(i_{n,k})^2 = (\sum_K \eta_K)^2 = \sum_{(K,L)} \eta_K.\eta_L$, kde sumácia sa berie cez všetky usporiadané dvojice (K, L) k-rozmerných podkociek D^n.

Ak $K \cap L \neq \emptyset$, potom prienikom je j-rozmerná podkocka s $0 \leq j \leq k$ a $|K \cup L| = 2^{k+1} - 2^j$. Teda

$$P(\eta_k.\eta_L = 1) = \frac{\binom{2^n-(2^{k+1}-2^j)}{m-(2^{k+1}-2^j)}}{\binom{2^n}{m}} = E(\eta_K.\eta_L). \tag{3.1}$$

Ľahko sa dokáže, že počet dvojíc s $\dim K \cap L = j$ je

$$2^{n-j}\binom{n}{j}\binom{n-j}{k-j}\binom{n-k}{k-j}. \tag{3.2}$$

Ak $K \cap L = \emptyset$, potom $|K \cup L| = 2^{k+1}$. Teraz

$$P(\eta_k.\eta_L = 1) = \frac{\binom{2^n-2^{k+1}}{m-2^{k+1}}}{\binom{2^n}{m}} = E(\eta_K.\eta_L). \tag{3.3}$$

Počet takých párov (K, L) je

$$\left(\binom{n}{k}2^{n-k}\right)^2 - \sum_{j=0}^{k} 2^{n-j}\binom{n}{j}\binom{n-j}{k-j}\binom{n-k}{k-j}. \tag{3.4}$$

Použijúc (3.1)-(3.4) dostaneme

$$E(i_{n,k})^2 = \sum_{j=0}^{k} \binom{n}{k} 2^{n-j} \binom{n-j}{k-j} \binom{n-k}{k-j} \frac{\binom{2^n-(2^{k+1}-2^j)}{m-(2^{k+1}-2^j)}}{\binom{2^n}{m}} +$$

$$+ \left[\binom{n}{k}^2 2^{2(n-k)} - \sum_{j=0}^{k} \binom{n}{k} 2^{n-j} \binom{n-j}{k-j} \binom{n-k}{k-j} \right] \frac{\binom{2^n-2^{k+1}}{m-2^{k+1}}}{\binom{2^n}{m}}. \tag{3.5}$$

Po odčítaní $(Ei_{n,k})^2$ od (3.5) je teda disperzia $Di_{n,k}$ je rovná

$$Di_{n,k} = E(i_{n,k})^2 - (Ei_{n,k})^2 = \sum_{j=0}^{k} \binom{n}{k} 2^{n-j} \binom{n-j}{k-j} \binom{n-k}{k-j} \frac{\binom{2^n-(2^{k+1}-2^j)}{m-(2^{k+1}-2^j)}}{\binom{2^n}{m}} +$$

$$+ \left[\binom{n}{k}^2 2^{2(n-k)} - \sum_{j=0}^{k} \binom{n}{k} 2^{n-j} \binom{n-j}{k-j} \binom{n-k}{k-j} \right] \frac{\binom{2^n-2^{k+1}}{m-2^{k+1}}}{\binom{2^n}{m}} -$$

$$-\binom{n}{k}^2 2^{2(n-k)} \left(\frac{\binom{2^n-2^k}{m-2^k}}{\binom{2^n}{m}} \right)^2 = 2^n \binom{n}{k} \left[\sum_{j=0}^{k} \binom{k}{j} \binom{n-k}{k-j} 2^{-j} \frac{\binom{2^n-(2^{k+1}-2^j)}{m-(2^{k+1}-2^j)}}{\binom{2^n}{m}} \right] +$$

$$+2^n \binom{n}{k} \left[\binom{n}{k} 2^{n-2k} - \sum_{j=0}^{k} \binom{k}{j} \binom{n-k}{k-j} 2^{-j} \right] \frac{\binom{2^n-2^{k+1}}{m-2^{k+1}}}{\binom{2^n}{m}} -$$

$$-\binom{n}{k}^2 2^{2(n-k)} \left(\frac{\binom{2^n-2^k}{m-2^k}}{\binom{2^n}{m}} \right)^2 =$$

$$= 2^n \binom{n}{k} \sum_{j=0}^{k} \binom{k}{j} \binom{n-k}{k-j} 2^{-j} \left(\frac{\binom{2^n-(2^{k+1}-2^j)}{m-(2^{k+1}-2^j)}}{\binom{2^n}{m}} - \frac{\binom{2^n-2^{k+1}}{m-2^{k+1}}}{\binom{2^n}{m}} \right) +$$

$$+\binom{n}{k}^2 2^{2(n-k)} \left(\frac{\binom{2^n-2^{k+1}}{m-2^{k+1}}}{\binom{2^n}{m}} - \left(\frac{\binom{2^n-2^k}{m-2^k}}{\binom{2^n}{m}} \right)^2 \right) \tag{3.6}$$

Použijeme jednu užitočnú formulu.

Formula 1 ([4]: Appendix III) *Pre* $k \leq b \leq a$*:*

$$\frac{\binom{a-k}{b-k}}{\binom{a}{b}} = \frac{(b)_k}{(a)_k} \leq \left(\frac{b}{a} \right)^k.$$

Disperziu $Di_{n,k}$ tak môžme zhora ohraničiť

$$2^n \binom{n}{k} \sum_{j=0}^{k} \binom{k}{j} \binom{n-k}{k-j} 2^{-j} \left(\left(\frac{m}{2^n}\right)^{2^{k+1}-2^j} - \left(\frac{m}{2^n}\right)^{2^{k+1}} \right), \qquad (3.7)$$

pretože po použití Formuly 1 sa ukazuje, že druhý sčítanec
$\binom{n}{k} 2^{n-2k} \left(\left(\frac{m}{2^n}\right)^{2^{k+1}} - \left(\frac{m}{2^n}\right)^{2^{k+1}} \right)$ môžme zanedbať.
Rutinné výpočty ukazujú, že postupnosť $a_j = 2^{-j} \left(\left(\frac{m}{2^n}\right)^{-2^j} - 1 \right)$ je neklesajúca. Teda (3.7) má ďalšie horné ohraničenie rovné

$$2^n \binom{n}{k} \left(\frac{m}{2^n}\right)^{2^{k+1}} .a_k. \sum_{j=0}^{k} \binom{k}{j} \binom{n-k}{k-j} =$$

$$= \binom{n}{k}^2 2^{n-k} \left(\frac{m}{2^n}\right)^{2^{k+1}} \left(\left(\frac{m}{2^n}\right)^{-2^k} - 1 \right) \leq \binom{n}{k}^2 2^{n-k} \left(\frac{m}{2^n}\right)^{2^k}. \qquad (3.8)$$

Podobne ako disperziu môžme použitím Formuly 1 ohraničiť aj strednú hodnotu

$$Ei_{n,k} \leq \binom{n}{k} 2^{n-k} \left(\frac{m}{2^n}\right)^{2^k}. \qquad (3.9)$$

Teraz použijeme Čebyševovu nerovnosť (Tvrdenie 3) na náhodnú premennú $i_{n,k}$ priradiac $\varepsilon = \varphi(n)\binom{n}{k}\sqrt{2^{n-k}\left(\frac{m}{2^n}\right)^{2^k}}$, kde $\frac{1}{\varphi(n)} = o(1)$. Použijúc predchádzajúce tvrdenia získame

$$P(|i_{n,k} - Ei_{n,k}| \geq \varepsilon) \leq \frac{Di_{n,k}}{\varepsilon^2} = \frac{1}{\varphi(n)} \to 0.$$

Teda $\lim P(|i_{n,k} - Ei_{n,k}| < \varepsilon) = 1$. To nám dáva výsledok, ktorý formulujeme nasledovne.

Tvrdenie 6 *S pravdepodobnosťou idúcou k* 1 *pre* $n \to \infty$ *pre každú* $f \in F_m$

platí:

$$\binom{n}{k}(2^{n-k}\left(\frac{m}{2^n}\right)^{2^k} - \varphi(n)\sqrt{2^{n-k}\left(\frac{m}{2^n}\right)^{2^k}}) < i_{n,k}(f) <$$

$$< \binom{n}{k}(2^{n-k}\left(\frac{m}{2^n}\right)^{2^k} + \varphi(n)\sqrt{2^{n-k}\left(\frac{m}{2^n}\right)^{2^k}}),$$

kde $\lim \varphi(n) = \infty$.

Pre nás je teda dôležitý podiel $\frac{m}{2^n} = \frac{m(n)}{2^n}$ a to pre $n \to \infty$. Ak je m konštanta, potom $\frac{m}{2^n} \to 0$. Ak $m(n) = 2^n$, potom $\frac{m}{2^n} \to 1$ pre $n \to \infty$.

Odteraz budeme predpokladať, že podiel $\frac{m}{2^n}$ konverguje k nejakému číslu $p \in (0, 1)$; špeciálny prípad bude, keď $p = \frac{1}{2}$ a to je vtedy, keď počet jednotiek pre každé n je práve polovica všetkých vrcholov D^n.

Nech teda $\frac{m(n)}{2^n}$ konverguje a $\frac{1}{\frac{m}{2^n}} = o(n) = \frac{1}{1-\frac{m}{2^n}}$, čiže $\frac{2^n}{m} = o(n) = \frac{2^n}{2^n - m}$.

Vyššieuvedené tvrdenie ukazuje možnosť získať asymptotický odhad $i_{n,k}$ pre k obmedzené medzi určitými hodnotami. O tom bude nasledujúca veta.

Veta 1 *Nech* $\frac{2^n}{m} = o(n) = \frac{2^n}{2^n - m}$. *S pravdepodobnosťou idúcou k* 1 *platí:*

(i) náhodná boolovská funkcia neobsahuje intervaly dimenzie väčšej ako $\mu = \lfloor \log n - \log\log \frac{1}{\frac{m}{2^n}} \rfloor + 1$. *Navyše*

(ii) pre $k \leq \lfloor \log n - \log\log \frac{1}{\frac{m}{2^n}} \rfloor - 1 = \mu - 2$ *počet* $i_{n,k}$ *k-rozmerných intervalov náhodnej boolovskej funkcie je asymptoticky ekvivalentný* $\binom{n}{k} 2^{n-k}\left(\frac{m}{2^n}\right)^{2^k}$, *čiže*

$$i_{n,k} \sim \binom{n}{k} 2^{n-k}\left(\frac{m}{2^n}\right)^{2^k}.$$

Dôkaz. Najskôr ukážeme, že vždy, keď $k > \mu$, horný odhad $i_{n,k}$ daný Tvrdením 6, ide k 0 pre $n \to \infty$. Tým dokážeme časť (i) Vety 1.

Nech $k=\mu+r$, kde $r\geq 1$. Keďže $\lim\binom{n}{\mu+r}\varphi(n)\sqrt{2^{n-\mu-r}.\frac{m}{2^n}^{2^{\mu+r}}}=0$ impliku-je, že $\lim\binom{n}{\mu+r}2^{n-\mu-r}.\frac{m}{2^n}^{2^{\mu+r}}=0$, stačí dokázať, že $\lim\binom{n}{\mu+r}\varphi(n)^2 2^{n-\mu-r}.\frac{m}{2^n}^{2^{\mu+r}}=0$. Vidiac toto úspešne získame

$$\varphi(n)^2\binom{n}{\mu+r}2^{n-\mu-r}\frac{m}{2^n}^{2^{\mu+r}}\leq\varphi(n)^2\frac{2^{2(\mu+r).\log n}.2^n}{\left(\frac{1}{\frac{m}{2^n}}\right)^{2^{\mu+r}}}\leq$$

$$\leq\varphi(n)^2\frac{2^{2(\log n-\log\log 1/\frac{m}{2^n}+r+1).\log n}}{2^{(2^r-1).n}}=X_1(n).$$

Ak $\lim\frac{m}{2^n}=p\in(0,1)$, potom obyčajne $\lim X_1(n)=0$. Ak $\frac{m}{2^n}\to 0$, potom $\log\log\frac{1}{\frac{m}{2^n}}\to\infty$, teda aj $X_1(n)$ ide k 0. Konečne ak $\frac{m}{2^n}\to 1$, potom $\log\log\frac{1}{\frac{m}{2^n}}\to-\infty$. Hoci

$$\lim\frac{-\log\log\frac{1}{\frac{m}{2^n}}.\log n}{n}=\lim\frac{1}{n}.\log\left(\frac{1}{\log\frac{1}{\frac{m}{2^n}}}\right).\log n=$$

$$\lim\frac{1}{n}.\log\left(1/\log\left(1+\frac{1-\frac{m}{2^n}}{\frac{m}{2^n}}\right)\right).\log n=\lim\frac{1}{n}.\log\frac{1}{1-\frac{m}{2^n}}.\log n=0\,,$$

čo naopak implikuje, že $\lim X_1(n)=0$ dokonca aj keď $\lim\frac{m}{2^n}=1$ a $\frac{1}{1-\frac{m}{2^n}}=o(n)$. Časť (i) je dokázaná.

Aby sme dokázali časť (ii), musíme overiť, že $\varphi(n)\binom{n}{k}\sqrt{2^{n-k}\left(\frac{m}{2^n}\right)^{2^k}}=o\left(\binom{n}{k}2^{n-k}\left(\frac{m}{2^n}\right)^{2^k}\right)$, čo znamená, že $\lim\frac{\varphi(n)}{\sqrt{2^{n-k}\left(\frac{m}{2^n}\right)^{2^k}}}=0$ pre $k\leq\mu-2$ a vhodné $\varphi(n)$. Na to stačí ukázať, že $\lim 2^{n-k}\left(\frac{m}{2^n}\right)^{2^k}=\infty$ vždy, keď $k\leq\mu-2$. Teraz máme

$$2^{n-k}\left(\frac{m}{2^n}\right)^{2^k}=2^{(1/2)n-\log n+\log\log 1/\frac{m}{2^n}+1}=X_2(n).$$

Podobné argumenty ako sú vyššie ukazujú, že $\lim X_2(n)=\infty$. Tu končí dôkaz.

Prípad, že $k=0$, si zaslúži špeciálnu pozornosť. Naozaj $i_{n,0}=|N_f|$, čo je rovné dĺžke hlavnej DNF funkcie f.

Veta 2 *Dĺžka hlavnej DNF funkcie $f \in F_m$ - počet bodov v N_f - je asymptoticky ekvivalentná $2^n.p(n)$ a zároveň rovná m:*

$$m =_{def} |N_f| = i_{n,0}(f) \sim 2^n.p(n) = 2^n.\frac{m}{2^n} = m.$$

Kapitola 4

Maximálne intervaly

Najskôr uvedieme pár formálnych definícií.

Definícia 1 *Interval N_K, obsiahnutý v N_f (prislúchajúci elementárnej konjunkcii K), sa nazýva maximálny (vzhľadom na pokrytie N_f), ak neexistuje interval $N_{K'}$ taký, že*
1. $N_K \subseteq N_{K'} \subseteq N_f$*;*
2. *Rád intervalu $N_{K'}$ je menší ako rád intervalu N_K.*

Definícia 2 *Konjunkcia K, prislúchajúca maximálnemu intervalu N_K množiny N_f, sa nazýva prostý implikant funkcie f.*

Platí, že $N_K \subseteq N_{K'}$ práve vtedy, keď všetky činitele z K' sú obsiahnuté v K. Z prostého implikanta K funkcie f nemožno vynechať žiaden činiteľ, pretože po vynechaní by sme dostali konjunkciu K', pre ktorú $N_{K'} \subset N_f$.

Zložitosť skrátenej DNF boolovskej funkcie je rovná počtu maximálnych intervalov (Veta 1(i)). Ako ďalší krok vyšetrujeme jeho strednú hodnotu.

Nech $I_{n,k}(f)$ je počet k-rozmerných maximálnych intervalov funkcie $f \in F_m$ a $s(f)$ zložitosť jej skrátenej DNF. Zrejme $s(f) = \sum_{k=0}^{n} I_{n,k}(f)$.

Tvrdenie 7

$$EI_{n,k} = \binom{n}{k} 2^{n-k}. \left(2^k\right)^{n-k} \frac{\binom{2^n-(2^k+(n-k))}{m-2^k}}{\binom{2^n}{m}}.$$

Dôkaz. Pre ľubovoľnú k-rozmernú podkocku $K \subseteq D^n$ definujeme indikátor Θ_K nasledovne:

$$\Theta_K = \begin{cases} 1, & \text{ak } K \text{ je maximálny interval } f; \\ 0, & \text{inak,} \end{cases}$$

$f \in F_m$.

Teraz $I_{n,k} = \sum_K \Theta_K$, kde K ide cez všetky k-rozmerné podkocky D^n. Pripomíname, že k-rozmerných podkociek v D^n je $\binom{n}{k} 2^{n-k}$. Preto stačí vypočítať $EI_{n,k} = P(\Theta_K(f) = 1)$ pre fixnú K.

k-rozmerná podkocka je teda maximálna, ak neexistuje žiadna $k+1$-rozmerná podkocka, ktorá by ju obsahovala; alebo inak, neexistuje iná k-rozmerná podkocka, s ktorou by spolu tvorili $k+1$-rozmernú.

Rozložením funkcie f vzhľadom na premenné x_i, $i = 1, \ldots, n-k$ získame

$$f(x_1, \ldots, x_n) = \bigvee_{\sigma=(\sigma_1,\ldots,\sigma_{n-k}) \in D^{n-k}} x_1^{\sigma_1} \wedge x_2^{\sigma_2} \wedge \ldots \wedge x_{n-k}^{\sigma_{n-k}} \wedge f(\sigma, x_{n-k+1}, \ldots, x_n) =$$

$$= x_1 \wedge x_2 \wedge \ldots \wedge x_{n-k} \wedge f_0(x_{n-k+1}, \ldots, x_n) \vee$$

$$\vee x_1' \wedge x_2 \wedge \ldots \wedge x_{n-k} \wedge f_1(x_{n-k+1}, \ldots, x_n) \vee \ldots$$

$$\ldots \vee x_1 \wedge x_2 \wedge \ldots \wedge x_{n-k}' \wedge f_{n-k}(x_{n-k+1}, \ldots, x_n) \vee \ldots$$

Pripomíname, že $x^\sigma = \begin{cases} \text{x}, & \text{ak } \sigma = 1; \\ \text{x'}, & \text{ak } \sigma = 0. \end{cases}$

Z predchádzajúcej rovnice vyplýva, že K je maximálny interval f práve vtedy, keď

(a) $f_0 = 1$

(b) $f_i \neq 1$ pre $i = 1, 2, \ldots, n-k$.

Teda ak v zápise elementárnej konjunkcie K negujeme jednu premennú, získame elementárnu konjunkciu prislúchajúcu k-rozmernému intervalu. Jeho existenciu chceme vylúčiť, pretože spolu s pôvodným tvoria $k+1$-rozmerný interval.

Opäť sa nám naskytla príležitosť, kedy môžme úspešne použiť Tvrdenie 1 z druhej kapitoly.

Chceme určiť pravdepodobnosť, že náhodná boolovská funkcia $f \in F_m$ bude obsahovať fixovanú maximálnu k-rozmernú podkocku K. Koľko je teda takých funkcií? Keďže 2^k jednotiek je týmto priradených vrcholom K, ostáva ešte vybrať zvyšných $m - 2^k$ jednotiek.

Potenciálnych k-rozmerných intevalov, ktoré by mohli narušiť maximálnosť K, je práve $n-k$, t.j. toľko, koľko premenných je v zápise K. Aby neohrozili maximálnosť K, priradíme napevno v každom, z týchto $n-k$ intervalov, jednému vrcholu nulu. To možno urobiť

$$\left(2^k\right)^{n-k} \tag{4.1}$$

spôsobmi.

Ak je teda v indukovanom grafe funkcie f fixovaných 2^k jednotiek k-rozmerného maximálneho intervalu K a vyššie uvedenými spôsobmi fixujeme aj nuly, počet funkcií, ktoré obsahujú maximálnu podkocku K je rovný

$$\left(2^k\right)^{n-k} \frac{\binom{2^n-(2^k+(n-k))}{m-2^k}}{\binom{2^n}{m}}. \tag{4.2}$$

Pravdepodobnosť, že f obsahuje k-rozmernú maximálnu podkocku, je teda

$$EI_{n,k}(f) = \binom{n}{k} 2^{n-k} \left(2^k\right)^{n-k} \frac{\binom{2^n-(2^k+(n-k))}{m-2^k}}{\binom{2^n}{m}}.$$

Odhad dĺžky skrátenej DNF náhodnej boolovskej funkcie budeme robiť v sérii nasledujúcich tvrdení.

Pre každé k potrebujeme spočítať očakávaný počet k-rozmerných maximálnych intervalo funkcie f.

Tvrdenie 8 *Nech* $\frac{m}{2^n}$ *ide k nejakému* p *pre* $n \to \infty$. *Tvrdíme, že* $Es \leq n^{(1+\varepsilon_1(n)) \log\log 1/\frac{m}{2^n} n}.2^n$, $\varepsilon_1(n) \to 0$. *Navyše ak* $\lim \frac{m}{2^n} = p \in (0,1)$, *potom* $\varepsilon_1(n) = O\left(\frac{1}{\log\log_{1/\frac{m}{2^n}} n}\right)$.

Dôkaz. Z toho, že $s(f) = \sum_{k=0}^{n} I_{n,k}(f)$, nás môže ako prvé zaujímať $\max_{0 \leq k \leq n} EI_{n,k}$. Uvažujme pomer

$$\frac{EI_{n,k+1}}{EI_{n,k}} = \frac{n-k}{k+1} 2^{n-2k-2}. \frac{\binom{2^n-2^{k+1}-(n-k-1)}{m-2^{k+1}}}{\binom{2^n-2^k-(n-k)}{m-2^k}} =$$

$$= \frac{n-k}{k+1} 2^{n-2k-2}. \frac{\binom{2^n-2^k-(n-k)+1-2^k}{m-2^k-2^k}}{\binom{2^n-2^k-(n-k)}{m-2^k}} =$$

$$= \frac{(n-k)2^{n-k}}{2(k+1)2^{k+1}}. \frac{2^n-2^{k+1-(n-k)-1}}{2^n-m-(n-k)-1} \left(\frac{\binom{2^n-2^k-(n-k)-2^k}{m-2^k-2^k}}{\binom{2^n-2^k-(n-k)}{m-2^k}} \right)^{2^k} = X_3(n,k).$$

Použijeme nasledujúcu formulu, ktorá je podobná Formule 1.

Formula 2 ([4]: Appendix III) *Ak* k, b *a* a *sú funkcie* n, *potom*

$$\frac{\binom{a-k}{b-k}}{\binom{a}{b}} \sim \left(\frac{b}{a}\right)^k,$$

za predpokladu, že $k = o(b^{1/2})$ *a* $k = o(a^{1/2})$.

Platí teda, že ak $m \leq 2^n - (n-k)$, potom

$$X_3(n,k) \sim \frac{(n-k)2^{n-k}}{2(k+1)2^{k+1}}.\frac{2^n - 2^{k+1-(n-k)-1}}{2^n - m - (n-k) - 1}\left(\frac{m - 2^k}{2^n - 2^k - (n-k)}\right)^{2^k}$$

Vidíme, že $X_3(n,k)$ zrejme závisí od n a počtu jednotiek m daného funkciou. Keďže predpokladáme, že $\lim \frac{m}{2^n}$ ide k $p \in (0,1)$ pre $n \to \infty$, možno tvrdiť, že $X3(n,k)$ aj rozmer maximálnych intervalov závisí od n a p. V skutočnosti ani nemôže závisieť od iného parametra.

Ukážeme, že pre $k \leq \log\log_{1/p} n - 1 = \log\log n - \log\log\frac{1}{p} - 1$ a všetky dostatočne veľké hodnoty n platí, že $X_3(n,k) > 1$. Naozaj

$$X_3(n,k) \geq \frac{n - \log\log n + \log\log\frac{1}{p} + 1}{2\log\log n - 2\log\log\frac{1}{p}}.\frac{2^n}{(\log\log n - \log\log\frac{1}{p})^2}.n.$$

S využitím predchádzajúcich faktov sa ľahko presvedčíme, že vo všetkých prípadoch ($\frac{m}{2^n} \to 0$, $\frac{m}{2^n} \to p \in (0,1)$, $\frac{m}{2^n} \to 1$) máme

$$\lim \frac{n - \log\log n + \log\log\frac{1}{p} + 1}{2\left(\log\log n - \log\log\frac{1}{p}\right)}.\frac{2^n}{(\log\log n - \log\log\frac{1}{p})^2} = \infty = \lim n$$

a odtiaľ $\lim X_3(n,k) = \infty$ vždy, keď $k \leq \log\log n - \log\log\frac{1}{p} - 1$. Teda $\frac{EI_{n,k+1}}{EI_{n,k}} > 1$ pre dostatočne veľké n.

Na druhú stranu, ak $k > \log\log n - \log\log\frac{1}{p}$, získame

$$0 \leq X_3(n,k) \leq \frac{n - \log\log n + \log\log\frac{1}{p}}{2\left(\log\log n - \log\log\frac{1}{p} + 1\right)}.\frac{1}{n-1}.\left(1 + \frac{1}{n}\right)^{n - \log\log n + \log\log 1/p - 1}.$$

Teraz máme

$$\lim \frac{n - \log\log n + \log\log\frac{1}{p}}{2\left(\log\log n - \log\log\frac{1}{p} + 1\right).(n+1)} = 0$$

$$a \qquad \lim\left(1 + \frac{1}{n}\right)^{n - \log\log n + \log\log 1/p - 1} = e.$$

Teda $\lim X_3(n,k) = 0$ vždy, keď $k > \log\log n - \log\log\frac{1}{p}$ a $\frac{EI_{n,k+1}}{EI_{n,k}} < 1$ pre dostatočne veľké n.

Z toho vyplýva, že $EI_{n,k}$ ako funkcia $k$nadobúda maximum buď, keď k je rovné $\lfloor\log\log n - \log\log\frac{1}{p}\rfloor = \lambda$ alebo$\lfloor\log\log n - \log\log\frac{1}{p}\rfloor + 1 = \lambda + 1$. Ďalej ešte,

$$\max_k EI_{n,k} \leq \binom{n}{\lambda+1} 2^{n-\lambda} p^{2^\lambda} (1 - p^{2^{\lambda+1}})^{n-\lambda} \leq 2^n . n^{(1+\varepsilon_1'(n))\log\log 1/pn},$$

kde $\varepsilon_1'(n) = \frac{1+(1/2\log n - \log\log n + \log\frac{1}{p})}{\log n\left(\log\log n - \log\log\frac{1}{p}\right)} \to 0$. Zrejme ak $p \to p \in (0,1)$, potom $\varepsilon_1'(n) = O\left(\frac{1}{\log\log_{1/p} n}\right)$.

Vrátením sa na začiatok dôkazu zistíme, že

$$Es \leq (n+1) max_k EI_{n,k} \leq$$
$$\leq n^{\log(n+1)/\log n} . n^{(1+\varepsilon_1'(n))\log\log 1/pn} . 2^n = n^{(1+\varepsilon_1(n))\log\log 1/pn} . 2^n.$$

Výsledok je jasný.

Tvrdenie 9 *S pravdepodobnosťou idúcou k 1 pre $n \to \infty$ platí, že $s(f) \geq \frac{i_{n,\lambda}(f)}{\binom{\mu}{\lambda}2^{\mu-\lambda}}$, kde $\mu = \lfloor\log n - \log\log\frac{1}{p}\rfloor + 1$ a $\lambda = \lfloor\log\log n - \log\log\frac{1}{p}\rfloor$.*

Dôkaz. Podľa Vety 1(i) dimenzia maximálnych intervalov neprekročí μ skoro isto. Teda ľubovoľný maximálny interval f obsahuje najviac $\binom{\mu}{\lambda}2^{\mu-\lambda}$ λ-rozmerných podintervalov.

Na druhej strane, každý λ-rozmerný podinterval je obsiahnutý v prinajmenšom jednom maximálnom intervale. Z toho $i_{n,\lambda}(f) \leq s(f).\binom{\mu}{\lambda}2^{\mu-\lambda}$ skoro isto.

Čitateľ môže hádať, že tvrdenia 8 a 9 sa použijú na dôkaz horných a dolných ohraničení pre náhodnú premennú $s(f)$. Keď sú raz tieto ohraničenia dokázané, vieme len, že $s(f) \leq$horná hranica (H.H.) skoro isto a $s(f) \geq$ dolná hranica (D.H.) skoro isto.

Avšak my chceme tvrdenie, že $D.H. \leq s(f) \leq H.H.$ skoro isto. Aby sme to videli, označme $\{f \in B_m;\ s(f) \leq H.H.\}$ a $\{f \in F_m;\ s(f) \geq D.H.\}$ ako F_n resp. G_n. Potom $\lim P(F_n) = 1 = \lim P(G_n)$. Teraz

$$1 \geq \lim P(F_n \cap G_n) = 1 - \lim P(F_m - (F_n \cap G_n)) \geq$$
$$\geq 1 - \lim P(F_m - F_n) - \lim P(F_m - G_n) = 1.$$

Ale, že $\lim P(F_n \cap G_n) = 1$ je ekvivalentné tomu, že hovoríme, že $D.H. \leq s(f) \leq H.H.$ skoro isto.

Kombinovanie skoro istých udalostí analogicky ako hore nám umožňuje nasledujúca lema. Umožňuje obemedziť naše konštrukcie na podmnožiny F_m, ktorých pravdepodobnosť ide k 1 pre $n \to \infty$, ktoré niekedy môžu byť výhodné alebo dokonca nutné.

Lema 1 *Nech* $(X_n, S_n, Q_n)_{n=1}^{\infty}$ *je postupnosť pravdepodobnostných priestorov. Predpokladajme, že* $Y_n, Z_n \in S_n$ *sú udalosti také, že* $\lim Q_n(Y_n) = 1 = \lim Q_n(Z_n)$ *a nech* $A_n \in S$. *Potom*

(i) $\lim Q_n(Y_n \cap Z_n) = 1$

(ii) $\lim Q_n(A_n|Y_n) = 1$ *implikuje* $\lim Q_n(A_n) = 1$.

Dôkaz je ponechaný na čitateľa.

Veta 3 *Nech* $\frac{1}{p} = o(n) = \frac{1}{1-p_n}$. *Potom s pravdepodobnosťou idúcou k* 1 *pre* $n \to \infty$ *platí, že*

$$n^{(1-\varepsilon_2(n))\log\log 1/pn}.2^n \leq s(f) \leq n^{(1+\varepsilon_3(n))\log\log 1/pn}.2^n;\ \varepsilon_2(n),\ \varepsilon_3(n) \to 0.$$

Ak $p \to p \in (0,1)$, *potom* $\varepsilon_2(n) = O\left(\frac{1}{\log\log_{1/p} n}\right) = \varepsilon_3(n)$.

Dôkaz. Horné ohraničenie. Z Markovovej nerovnosti (Tvrdenie 2) pre $\varepsilon = n$ máme

$$s(f) \leq n.Es \qquad \textit{skoro isto.} \tag{4.3}$$

Skombinovanie (4.3) a Tvrdenia 8 dáva požadované horné ohraničenie. Tu $\varepsilon_2(n) = \varepsilon_1(n) + \frac{1}{\log\log_{1/p} n} \to 0$.
Zrejme $\varepsilon_3(n) = O\left(\frac{1}{\log\log_{1/p} n}\right)$ vždy, keď $\lim \frac{m}{2^n} p \in (0,1)$.
Dolné ohraničenie. Najskôr odhadneme $i_{n,\lambda}(f)$. Tvrdenie 6 implikuje, že

$$i_{n,\lambda}(f) > \binom{n}{\lambda}(2_{n-\lambda} - \varphi(n)\sqrt{2^{n-\lambda}p^{2^\lambda}}) \qquad \textit{skoro isto.} \tag{4.4}$$

Použitím $\binom{n}{\lambda} > \left(\frac{n}{\lambda}\right)^\lambda$ a nahradením λ v (4.4) získame $i_{n,k}(f) > n^{(1+\varepsilon_2'(n))\log\log 1/pn}.2^n$, kde

$$\varepsilon_2'(n) = \frac{2\log n + \log\log_{1/p} n.\log\log\log_{1/p} n - \log(1 - \sqrt{(n^3/2^n).\log_{1/p} n})}{\log n.\log\log_{1/p} n} \to 0.$$

Z použitia Tvrdenia 9 vyplýva

$$s(f) \geq \frac{i_{n,\lambda}}{\binom{\mu}{\lambda}2^{\mu-\lambda}} \geq \frac{n^{(1+\varepsilon_2'(n)).\log\log 1/pn}.2^n}{\mu^\lambda 2^{\mu-\lambda}} \geq n^{(1+\varepsilon_2(n))\log\log 1/pn}.2^n,$$

kde $\varepsilon_2(n) = \varepsilon_2'(n) + \frac{\log\left(\log n - \log\log\frac{1}{p} + 1\right)}{\log n} \to 0$. Znovu ľahký výpočet ukazuje, že $\varepsilon_2(n)$ má charakter $O\left(\frac{1}{\log\log_{1/p} n}\right)$ za predpokladu, že $\lim p(n) = p$ je rôzna od 0 a 1.

Kapitola 5

Dimenzia maximálnych intervalov

Ako vyplýva z Vety 1(i), rozmer žiadneho intervalu náhodnej boolovskej funkcie neprekročí $\lfloor \log n - \log\log \frac{1}{p} \rfloor + 1 = \mu$ skoro isto. Zároveň je μ aj horné ohraničenie rozmeru maximálnych intervalov. Na druhej strane, vyšetrovanie realizované v postupe dôkazu Tvrdenia 8 ukazuje, že kvantita $EI_{n,k}$, na ktorú sa pozeráme ako na funkciu k, rastie, kým k nie je približne rovné $\log\log n - \log\log \frac{1}{p}$ a potom, po dosiahnutí maxima, klesá. Tak môžme očakávať, že väčšia časť maximálnych intervalov má rozmer blízko $\log\log n - \log\log \frac{1}{p}$. Potvrdíme toto očakávanie vo vete nižšie.
Uveďme niektoré označenia použité v nasledujúcej vete.
$I^+_{n,k} = \sum_{l>k} I_{n,k}$
$I^-_{n,k} = \sum_{l<k} I_{n,k}$
$H^+_{n,k}$ ($H^-_{n,k}$) - počet bodov pokrytých intervalmi dimenzie väčšej (menšej)

Veta 4 *Nech* $\frac{2^n}{m} = o(n) = \frac{2^n}{2^n - m}$ *a* $\lambda_1 = \log\log 1/\frac{m}{2^n} n - 1,$

$$\lambda_2 = \log\log \frac{1}{\frac{m}{2^n}} n + \log\log\log \frac{1}{\frac{m}{2^n}} n + \varepsilon,\ \varepsilon > 0.$$

Potom s pravdepodobnosťou idúcou k 1 pre $n \to \infty$ *platí*

$$I^-_{n,\lambda_1}(f) = o(2^n) = H^-_{n,\lambda_1}(f),$$
$$I^+_{n,\lambda_2}(f) = o(2^n) = H^+_{n,\lambda_2}(f).$$

Dôkaz. Začneme so skúmaním toho, že boolovská funkcia f spĺňa systém nerovností

$$I_{n,k}(f) < n.EI_{n,k},\ k = 0, 1, \ldots, \mu,\ \left(\mu = \lfloor \log n - \log\log \frac{1}{p} \rfloor + 1\right) \tag{5.1}$$

skoro isto. Najskôr označíme $M_{n,k}$ udalosť, že konkrétna nerovnosť $I_{n,k}(f) < n.\ EI_{n,k}$ zodpovedá pevnému k a $M_n = \bigcap_{0 \le k \le \mu} M_{n,k}$. Potom

$$P(F_m - M_n) = P(\cup(F_m - M_{n,k})) \le \sum_{k=0}^{\mu} P(F_m - M_{n,k}) =$$
$$= \sum_{k=0}^{\mu} P(I_{n,k}(f) \ge n.EI_{n,k}) \le \frac{\mu + 1}{n},$$

z Markovovej nerovnosti. Keďže $\frac{\mu+1}{n} \to 0$, potvrdili sme, že $\lim P(M_n) = 1$. Ako vyplýva z našej Lemy, v nasledujúcich výpočtoch sa môžme obmedziť na podmnožinu $Y_n \subseteq F_m$ takých boolovských funkcií, že

(i) rozmer maximálnych intervalov f neprekročí μ;

(ii) f vyhovuje systému nerovností (5.1).

Terazsa vráťme k odhadom. Na dokázanie $I^+_{n,\lambda_2}(f) = o(2^n)$ a $H^+_{n,\lambda_2}(f) = o(2^n)$ budeme postupovať simultánne najskôr podaním spoločného horného ohraničenia pre I^+_{n,λ_2} a H^+_{n,λ_2}.

Platí skoro isto, že

$$I^+_{n,\lambda_2}(f) = \sum_{\lambda_2<k\leq\mu} I_{n,k}(f) < \sum_{\lambda_2<k\leq\mu} I_{n,k}(f).2^k,$$
$$H^+_{n,\lambda_2}(f) \leq \sum_{\lambda_2<k\leq\mu} I_{n,k}(f).2^k.$$

Z (5.1) a Tvrdenia 7 máme

$$\sum_{\lambda_2<k\leq\mu} I_{n,k}(f).2^k < \sum_{\lambda_2<k\leq\mu} n.EI_{n,k}.2^k =$$
$$= \sum_{\lambda_2<k\leq\mu} n.\binom{n}{k}2^{n-k}.\left(2^k\right)^{n-k}\frac{\binom{2^n-(2^k+(n-k))}{m-2^k}}{\binom{2^n}{m}}.2^k \leq$$
$$\leq \sum_{\lambda_2<k\leq\mu} n^{k+1}\left(\frac{m}{2^n}\right)^{2^k}$$

skoro isto.

Ak $k > \log\log_{1/p} n$, postupnosť $a(k) = n^{k+1}p^{2^k} = n^{k+1}\left(\frac{m}{2^n}\right)^{2^k}$ je neklesajúca.

Kedže $\lambda_2 > \log\log_{1/p} n$, dostávame

$$\sum_{\lambda_2<k\leq\mu} I_{n,k}(f).2^k < \sum_{\lambda_2<k\leq\mu} n^{k+1}p^{2^k} \leq$$
$$\leq 22^n\mu n^{\lambda_2+1}p^{2^{\lambda_2+1}} = 2^n o(1) = o(2^n).$$

Analogicky ako vyššie

$$I^-_{n,\lambda_1}(f) = \sum_{0\leq k<\lambda_1} I_{n,k}(f) < \sum_{0\leq k<\lambda_1} I_{n,k}(f).2^k,$$
$$H^-_{n,\lambda_1}(f) \leq \sum_{0\leq k<\lambda_1} I_{n,k}(f).2^k \qquad \textit{skoro isto.}$$

Tvrdenie 5 a (5.1) implikuje, že

$$\sum_{0\le k<\lambda_1} I_{n,k}(f).2^k < \sum_{0\le k<\lambda_1} n.EI_{n,k}.2^k =$$

$$= \sum_{0\le k<\lambda_1} n.\binom{n}{k}2^{n-k}.\left(2^k\right)^{n-k}\frac{\binom{2^n-(2^k+(n-k))}{m-2^k}}{\binom{2^n}{m}}.2^k \tag{5.2}$$

V dôkaze Tvrdenia 8 sme zistili, že $EI_{n,k}$, $\exp_p 2^{\lambda_1}$ ako funkcia k, je rastúca vždy, keď $k \le \lambda_1$. Teda nahradením λ_1 v (5.2) získame

$$\sum_{0\le k<\lambda_1} I_{n,k}(f)2^k < (\lambda_1+1)n.n^{\lambda_1}\exp_p 2^{\lambda_1-1}(1-\exp_p 2^{\lambda_1})^{n-\lambda_1} \le$$

$$\le 2^n n^{\log\log 1/pn}.\left(1-\frac{1}{\sqrt{n}}\right)^{n-\log\log 1/pn} = 2^n o(1) = o(2^n).$$

Týmto končí dôkaz.

Poslednú vetu možno interpretovať tým spôsobom, že počet maximálnych intervalov s dimenziou menšou ako λ_1 a väčšou ako λ_2 je malý v porovnaní s počtom všetkých maximálnych intervalov náhodnej boolovskej funkcie. Teda podstatnú časť skrátenej DNF tvoria prosté implikanty odpovedajúce maximálnym intervalom s dimenziou medzi λ_1 a λ_2. Okrem toho, počet vrcholov, nepokrytých maximálnymi intervalmi, vnútri týchto obmedzení je malý v porovnaní s počtom vrcholov v n-kocke. Vezmúc do úvahy fakt, že náhodná boolovská funkcia má asymptoticky $p2^n$ vrcholov, usudzujeme, že s pravdepodobnosťou idúcou k 1 pre $n \to \infty$,

$$\lim \frac{H^-_{n,\lambda_1}(f)}{|N_f|} = \lim \frac{H^-_{n,\lambda_1}(f)/p.2^n}{|N_f|/p.2^n} = \lim \frac{1}{p}.\frac{o(1)}{1} = \lim \frac{H^+_{n,\lambda_2}(f)}{|N_f|}.$$

To vedie k nasledujúcemu výsledku.

Veta 5 *Nech* $\frac{1}{p(n)} = o(n) = \frac{1}{1-p(n)}$. *Potom*

$$I^-_{n,\lambda_1}(f) = o(s(f)) = I^+_{n,\lambda_2}(f).$$

Ak navyše $\lim p(n) = p > 0$, *potom*

$$H^{-}_{n,\lambda_1}(f) = o(|N_f|) = H^{+}_{n,\lambda_2}(f).$$

Podmienka $\frac{1}{1-p(n)} = o(n)$ je splnená, ak $\lim p(n) > 0$.

Kapitola 6

Záver

Na záver nám ostáva zhrnúť dosiahnuté výsledky.

V tretej kapitole sme vypočítali a odhadli počet k-rozmerných podkociek. Dokázali sme, že boolovská funkcia neobsahuje intervaly rozmeru väčšieho ako $\mu = \lfloor \log n - \log\log \frac{1}{\frac{m}{2^n}} \rfloor + 1$, t.j. určili sme rozmer k ako funkciu n a m, pre ktorý počet k-podkociek ide k 0 pre $n \to \infty$.

Pre rozmer menší alebo rovný $\mu - 2$ sme ukázali, že počet k-rozmerných intervalov NBF je ekvivalentný $\binom{n}{k} 2^{n-k} \left(\frac{m}{2^n}\right)^{2^k}$

V ďalšom sme vypočítali a ohraničili počet k-rozmerných maximálnych podkociek a zároveň tým aj súvisiacu priemernú dĺžku skrátenej DNF náhodnej boolovskej funkcie.

V poslednej časti o dimenziách maximálnych intervalov sme odhadovali počet maximálnych intervalov, ktorých rozmer je väčší ako nejaká dolná hranica λ_2 a menší ako horná hranica λ_1

Problematika náhodných boolovských funkcií a minimalizácie disjunktívnych normálnych funkcií zďaleka nie je vyčerpaná.

Takisto aj v triede boolovských funkcií s daným počtom jednotiek F_m ostalo mnoho nezodpovedaných a nevypovedaných otázok, ktorým som sa z nedostatku času nemohla venovať.

Téma určovania metrických vlastností boolovských funkcií pomocou pravdepodobnostných a kombinatorických metód je zaujímavá, ako pokračovanie v mojej práci by stáli za preskúmanie, okrem iného, iredundantné disjunktívne normálne formy a ich implikanty. Taktiež regulárne a jadrové vrcholy vo vrcholových pokrytiach náhodných boolovských funkcií.

Literatúra

[1] M. Škoviera: On the Minimization of Random Boolean Functions, Part 1, Computers and Artificial Intelligence, Bratislava 1986.

[2] S. V. Yablonskii,O. B. (Eds.) Lupanov: Discrete Mathematics and Mathematical Problems of Cybernetics, Nauka, Moscow 1974.

[3] B. Bollobas: Random Graphs, Academic Press, London 1985.

[4] E. M. Palmer: Graphical Evolution: Appendixes, New York 1985.

[5] W. Feller: An Introduction to Probability Theory and Its Applications. Vol. 1, 3rd Ed., J. Wiley and Sons, New York 1970.

[6] J. C. Bioch: Modular Decomposition of Boolean Functions, ERIM Report Series, Rotterdam 2002.