Department of Computer Science
Faculty of Mathematics, Physics and Informatics
Comenius University

Average degree in the interval graph of a random Boolean function

Master Thesis

Jakub Daubner

Advisor:
Doc. RNDr. Eduard Toman, PhD.

Bratislava, 2008

I hereby declare that I wrote this thesis by myself, only with the help of the referenced literature, under the careful supervision of my thesis advisor.

. . . .

Acknowledgements

I would like to thank to my advisor prof. Eduard Toman for his invaluable guidance and many useful suggestions during my work on this thesis.

A special thank you goes to Katarína Bachratá, Peter Czimmerman and Vojto Bálint for their careful listening to my lectures about this thesis, for their relevant questions, thus for correcting many mistakes. Next special thank you goes to my family and friends for their support and confidence.

Abstract

We consider an n-ary random Boolean function f such that $|\{\tilde{\alpha}; f(\tilde{\alpha}) = 1\}| = m$ and every such Boolean function has the same probability. We study its geometric model, the so called interval graph. The interval graph of a random Boolean function was introduced by Sapozhenko and has been used in constructions of schemes for realizing Boolean functions. Using this model, we estimate the number of maximal intervals intersecting a given maximal interval of a random Boolean function and proove that the asymptotic bound of the number is $n^{(1+\phi(n)) \log_2 \log_{1/p} n}$, where $p = m/2^n$ and $\phi(n) \to 0$ as $n \to \infty$.

We also study the equality of this model of random Boolean function with another one, where $\Pr[f(\tilde{\alpha}) = 1] = p$, for $\tilde{\alpha} \in \{0,1\}^n$. We find the conditions on m under which are these two probabilistic models equivalent, what means that $m/2^n$ can be consider as a equivalent to p.

Finally, we started to study the "structure" of the neighbourhood of the first order for the purposes of estimating the size of neighbourhood of second (or higher) order. We also drop a hint how should be obtained results used in next works on this field.

Keywords. random Boolean function, interval graph

Contents

1 Introduction

Boolean function is a very important subclass of functions for many computer science's areas. The Boolean functions are often represent by their *disjunctive normal forms* (d.n.f). Therefore it is important also to study the minimization of d.n.f.. However this is very complex problem and it is hard to find the general solution of it, which should be sufficiently fast. One special way how to sufficiently fast minimize d.n.f can be done using *local algorithms*.

Local algorithms presents an important subclass of algorithms for constructions of optimal schemes. The main idea of local algorithms is simple: they introduce a metric on the "space" of all elements (building blocks of schemes) and a "measure of quality" of elements. Then for every element of the scheme under construction they analyze its neighbouring elements; searching for better ones; if such elements exist, local algorithm chooses the best of them and substitutes the original element. The whole procedure is repeated until no replacement/improvement can be done.

Zhuravlev [3] studied the use of local algorithms in the minimization of d.n.f.. He introduced the notion of a *conjuction neighbourhood* and proved that the optimal d.n.f. cannot be constructed in general by means of local algorithms based on finite (local) conjuction neighbourhoods. So it is impossible to find an optimal solution by means of local algorithms. Anyway we still can construct a sub-optimal d.n.f. by analyzing neighbourhood of the first order.

The complexity of such local algorithms and the optimality of their results depends on a number of elements in the neighbourhood of the first order. Therefore we use probabilistic methods to obtain the lower and upper bound on this number, where the bounds are satisfied for almost all Boolean functions.

One such probabilistic model was studied by Toman, Olejár and Stanek in [2]. In this thesis we will introduce another probabilistic model and prove the similar results for it.

2 Preliminaries and Notation

We shall use the standard notation of Boolean function theory. The n-ary Boolean function is function $f\colon \{0,1\}^n \to \{0,1\}$. B_n denotes the set of all n-ary Boolean functions. Boolean variables and their negations are called *literals*. The literal of a variable x will be denoted by x^α, ($\alpha \in \{0,1\}$), where

$$x^\alpha = \begin{cases} x & \text{if } \alpha = 1 \\ \neg x & \text{if } \alpha = 0 \end{cases}$$

A conjuction $K = x_{i_1}^{\alpha_{i_1}} \dots x_{i_r}^{\alpha_{i_r}}$ of literals of different variables is called an *elementary conjuction*. The number of literals (r) in a conjuction K is called the *rank* of K. A special case is the conjuction of rank 0; it is called *empty* and its value is set to 1.

A formula $D = K_1 \vee \dots \vee K_m$, the disjunction of distinct elementary conjuctions, is called a *disjunctive normal form*. The parameter m (the number of elementary conjuctions in D) is called the *length* of D. The d.n.f. with $m = 0$ is called *empty* and its value is 0. A d.n.f. D *represents* a Boolean function f if the truth tables of f and D coincide. Let us consider the class of all d.n.f.s representing an n-ary Boolean function f; the d.n.f. with the minimal number of literals in this class is called a *minimal d.n.f. of* f and the d.n.f. with the minimal length (in this class) is called a *shortest d.n.f. of f*.

We use a geometric representation of Boolean functions. The Boolean n-cube B^n is a graph B^n with 2^n vertices $\tilde{\alpha} = (\alpha_1, \dots, \alpha_n); \alpha_i \in \{0,1\}$, where the edges are joining those pairs of vertices which differ in exactly one coordinate. For an n-ary Boolean function f let N_f denote the subset $\{\tilde{\alpha}; f(\tilde{\alpha}) = 1\}$ of all vertices $\tilde{\alpha}$. As can be easily seen, there is a one-to-one correspondence between the sets N_f and Boolean functions f. The subgraph of the Boolean n-cube induced by the set of N_f is called the *graph of* f.

The set of vertices $N_K \subseteq \{0,1\}^n$ corresponding to an elementary conjuction K of rank r is called *interval of order* r. Obviously, to every elementary conjuction $K = x_{i_1}^{\alpha_{i_1}} \dots x_{i_r}^{\alpha_{i_r}}$ corresponds an interval of order r consisting of all vertices $(\beta_1, \dots, \beta_n)$ of B^n, such that $\beta_{i_j} = \alpha_{i_j}$ for $j = 1, \dots, r$ (the vaules of other vertex coordinates can be chosen arbitrarily). Consequently, every vertex of B^n represents an interval of order n and the vertex set of B^n itself corresponds to an interval of order 0. In the geometric model every interval of order r represents $(n-r)$-dimensional subcube of B^n. An interval N_K is called a *maximal interval* of a Boolean function f if $N_K \subseteq N_f$ and there exists no interval $N_{K'} \subseteq N_f$ such that $N_K \subsetneq N_{K'}$. For every elementary conjuction K from the d.n.f. D the *neighbourhood* of K of the first order (with respect to the d.n.f. D) is defined as the set of all elementary conjuctions K_j from

D such that (in algebraic notation) $K \wedge K_j \not\equiv 0$ or (in our geometric model) $N_K \cap N_{K_j} \neq \emptyset$. Since we study mainly the neighbourhood of the first order in this thesis, the notion "neighbourhood" will denote the neighbourhood of the first order. The *interval graph* $\Gamma(f)$ is a graph associated with a Boolean function f; its vertices correspond to maximal intervals of f and the vertices corresponding to intervals N_{K_i} and N_{K_j} are joined by an edge in $\Gamma(f)$ if and only if $K_i \wedge K_j \not\equiv \emptyset$. We study the degree of a vertex in $\Gamma(f)$; namely we estimate the lower and the upper bounds of this parameter.

For an arbitrary Boolean function f and each of its d.n.f.s $K_1 \vee K_2 \vee \cdots \vee K_m$ we have that

$$N_f = \bigcup_{j=1}^{m} N_{K_j}.$$

In other words, every d.n.f. of a Boolean function f correspnods to a covering of N_f by intervals $N_{K_1}, \ldots, N_{K_m}$ such that $N_{K_i} \subseteq N_f$. Conversely, every covering of N_f by intervals $N_{K_1}, \ldots, N_{K_m}$ contained in N_f corresponds to some d.n.f. of f. Using geometric interpretation of d.n.f.s, we can express the "irreducibility" of d.n.f.: the d.n.f. D of a Boolean function f cannot be simplified if and only if every interval N_K of the covering (corresponding to D) contains at least one vertex belonging to just one interval of the covering.

Let r_j denote the order of an interval N_{K_j}. Then the number of literals in d.n.f. is $r = \sum_{j=1}^{m} r_j$ and the construction of the minimal d.n.f. can be formulated in the geometric model as a problem of constructing a covering of N_f by intervals $N_K \subseteq N_f$ with minimal r. On the other hand, the construction of the covering corresponding to a shortest d.n.f. requires to minmize the number of intervals in a covering of N_f.

Various metrical parameters of "typical" Boolean functions have been studied in the context of Boolean functions minimization in the class of d.n.f.s. [5, 6, 7].

More general model of Boolean functions, a concept of *random Boolean function* was studied by Škoviera and Toman, Olejár and Stanek in [1] and [2] (different probabilistic models are studied also in [8]). They used combinatorical-probabilistic methods, considering metric parameters of Boolean functions as random variables, estimated the expectations and variances of these variables and finally they estimated their values by means of Markov's and Chebysev's inequalities. The same approach will be used in this thesis. More about using combinatorical-probabilistic methods can be found in [9, 10, 11].

Let X be a random variable and let symbols $E(X)$ and $Var(X) = E(X - E(X))^2$ denote the expectations and variance of random variable X, respectively. We only use nonnegative random variables in the present thesis.

Theorem 2.1 (Markov's inequality)**.** *If ξ is a non-negative random variable and $\varepsilon > 0$ is a positive real number, then:*

$$P(\xi \geq \varepsilon) \leq \frac{E(\xi)}{\varepsilon}.$$

Theorem 2.2 (Chebyshev's inequality)**.** *For any random variable ξ and $\varepsilon > 0$ the following inequality holds:*

$$P(|\xi - E(\xi)| \geq \varepsilon) \leq \frac{Var(\xi)}{\varepsilon^2}.$$

We also use following notations in this thesis:

Notation 2.1. $a^{\underline{b}}$ (*falling factorial*) denotes

1. $(\forall a \in R, \forall b \in N^+); a^{\underline{b}} = a \cdot (a-1) \cdot \cdots \cdot (a-b+1)$

2. If $b = 0$ and $a \neq 0$, then $a^{\underline{b}} = 1$

Notation 2.2. For functions $f, g \colon R \to R$ we use these asymptotic notations:

- $f \sim g$ means that $\lim_{x\to\infty} \frac{f(x)}{g(x)} = 1$
- $f \lesssim g$ means that $\lim_{x\to\infty} \frac{f(x)}{g(x)} \leq 1$
- $f \gtrsim g$ means that $\lim_{x\to\infty} \frac{f(x)}{g(x)} \geq 1$
- $f = o(g)$ means that $\lim_{x\to\infty} \frac{f(x)}{g(x)} = 0$
- $f = O(g)$ means that $\lim_{x\to\infty} \frac{f(x)}{g(x)} = C$, where C is arbitrary non-negative constant
- $f = \omega(g)$ means that $\lim_{x\to\infty} \frac{f(x)}{g(x)} = \infty$
- $f = \Omega(g)$ means that $\lim_{x\to\infty} \frac{f(x)}{g(x)} > 0$

3 Probabilistic models - Model A and Model B

Definition 3.1 (Model A)**.** *Random boolean function of* n *variables attains the values* 1 *and* 0 *with probability* p *and* $1-p$*, respectively, independently of various points of* B^n*. The probability* p *may depend on* n*, so we label it* p_n*.*

(B_n, P_A) *is a discrete probability space, where* P_A *is a probability measure defined by follows:*

1. $(\forall f \in B_n) P_A(f) = p_n^{|N_f|} \cdot (1-p_n)^{(2^n - |N_f|)}$
2. *For an arbitrary* $S \subset B_n$ *we set* $P_A(S) = \sum_{f \in S} P_A(\{f\})$*.*

Škoviera and Toman, Olejár and Stanek studied Model A in [1] and [2]. In this thesis we will proove similar results for different model - Model B (see next definitions).

Definition 3.2. $B_{m,n}$ *is set of all* n*-ary Boolean functions containing* m *vertices. Formally:*

$$B_{m,n} = \{f \in B_n; |N_f| = m\}.$$

Remark 3.1.

$$|B_{m,n}| = \binom{2^n}{m}$$

Definition 3.3 (Model B)**.** $(B_{m,n}, P_B)$ *is discrete probability space, where* P_B *is probability measure defined by follows:*

1. *Every Boolean function* $f \in B_{m,n}$ *has the same probability, thus*

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

2. *For an arbitrary* $S \subset B_{m,n}$ *we set* $P_B(S) = \sum_{f \in S} P_B(\{f\})$*.*

Definition 3.4. *Let* A *be a certain property that a* n*-ary Boolean function* f *may or may not have. If*

$$\lim_{n\to\infty} \Pr[f \text{ has the property } A] = 1,$$

we say that a random Boolean function has the property A *almost surely. This definition is the same for both Models A and B (*"Pr" *can be* P_A *or* P_B*).*

Lemma 3.1. *Let* $U, V \subset B^n$ *and* $U \cap V = \emptyset$. *Let* $F = \{f \in B_{m,n}; U \subset N_f, V \subset B^n - N_f\}$. *Then*

$$P_B(F) = \frac{m^{\underline{|U|}} \cdot (2^n - m)^{\underline{|V|}}}{(2^n)^{\underline{|U|+|V|}}}.$$

Special case if $|V| = 0$ *then*

$$P_B(F) = \frac{m^{\underline{|U|}}}{(2^n)^{\underline{|U|}}}.$$

Proof. $\sum_{f \in F} \frac{1}{\binom{2^n}{m}} = \frac{\binom{2^n - |U| - |V|}{m - |U|}}{\binom{2^n}{m}} = \frac{(2^n - |U| - |V|)!}{2^n!} \cdot \frac{m!}{(m - |U|)!} \cdot \frac{(2^n - m)!}{(2^n - m - |V|)!} =$

$$\frac{1}{(2^n)^{\underline{|U|+|V|}}} \cdot m^{\underline{|U|}} \cdot (2^n - m)^{\underline{|V|}}$$

□

Notation 3.1.

$$p_{m,n} = \frac{m}{2^n}$$

Our first goal in next chapters will be to proove that $p_{m,n}$ in Model B is equivalent to p_n in Model A.

4 Estimation of dimension of maximal interval in Model B

In this chapter we will proove upper bound on the dimension of maximal intervals of random boolean function in Model B. The upper bound from Corollary 4.4 will be used in next chapter for prooving equality of Model A and Model B. We will obtain the upper bound using Markov's inequality.

Definition 4.1. *Let $i_{m,n,k}$ denote a random variable on $B_{m,n}$ such that $i_{m,n,k}(f)$ is equal to the count of k-dimensional intervals of a function $f \in B_{m,n}$.*

Lemma 4.1.

$$E(i_{m,n,k}) = \binom{n}{k} \cdot 2^{n-k} \cdot \frac{m^{\underline{2^k}}}{(2^n)^{\underline{2^k}}}$$

Proof. For any k-dimensional subcube K of the cube B^n we introduce a random variable η_K (called also an *indicator*) defined by follows:

$$\eta_K(f) = \begin{cases} 1 & \text{if } K \subseteq N_f \\ 0 & \text{otherwise.} \end{cases}$$

Clearly $i_{m,n,k}(f) = \sum_K \eta_K(f)$ where the summation ranges over all k-dimensional subcubes of B^n.

Using Lemma 3.1 we can compute that:

$$E(\eta_K) = P_B(\eta_K = 1) = P_B(K \subseteq N_f) = \frac{m^{\underline{2^k}}}{(2^n)^{\underline{2^k}}}$$

There are $\binom{n}{k} 2^{n-k}$ k-dimensional subcubes in B^n. Thus,

$$E(i_{m,n,k}) = \sum_K E(\eta_K) = \binom{n}{k} 2^{n-k} \frac{m^{\underline{2^k}}}{(2^n)^{\underline{2^k}}}.$$

□

Lemma 4.2. *If $E(i_{m,n,k}) \to 0$ as $n \to \infty$, then $P(i_{m,n,k} = 0) \to 1$ as $n \to \infty$.*

Proof. Lemma is direct consequence of Markov's inequality. □

Lemma 4.3. *Suppose $o(n) = \frac{1}{1-p_{m,n}}$. Suppose $k = \Omega(\lg m)$. Suppose $n = m^{o(1)}$. Then $E(i_{m,n,k}) \to 0$ as $n \to \infty$.*

Proof.
$$E(i_{m,n,k}) \leq 2^{k\cdot(\lg n-1)+n} \cdot \left(\frac{m}{2^n}\right)^{2^k} = U$$

From supposition $k = \Omega(\lg m)$ we have $k \geq c \cdot \lg m$, where c is arbitrary (little) positive constant. If $i_{m,n,k}$ is 0 for some k, then it will be 0 for any greater k, so we can take only minimal value $k = c \cdot \lg m$ and proove that $i_{m,n,k} \to 0$. Then:

$$U = 2^{c\cdot\lg m\cdot(\lg n-1)+n} \cdot \left(\frac{m}{2^n}\right)^{m^c} = 2^{c\cdot\lg m\cdot(\lg n-1)+n} \cdot 2^{m^c\cdot\lg(\frac{m}{2^n})}$$

Using simply facts $\lg m \leq n$ and $\lg n < n$ we obtain:

$$U \lesssim 2^{c\cdot n^2} \cdot 2^{m^c\cdot\lg(\frac{m}{2^n})}.$$

Using supposition $o(n) = \frac{1}{1-p_{m,n}}$, we can write that $p_{m,n} < 1 - \frac{1}{n}$. Thus:

$$U \lesssim 2^{c\cdot n^2+m^c\cdot\lg(1-\frac{1}{n})}.$$

Using Taylor series for $\lg(1-x)$ for $x \to 0^+$ we get $\lg(1-\frac{1}{n}) \leq -\frac{1}{n\cdot\ln 2}$. Thus:

$$U \lesssim 2^{c\cdot n^2-m^c\cdot\frac{1}{n\ln 2}}.$$

Notice that $U \to 0$ when $c \cdot n^2 = o(m^c \cdot n^{-1} \cdot \ln 2)$. This is equivalent to $n^3 = o(m^c)$ (multiplying by any constant has no affect to o-notation). And $n^3 = o(m^c)$ is true, because of supposition $n = m^{o(1)}$ (if we use $n < m^{\frac{c}{4}}$).

Therefore $U \to 0$.

□

Using Lemmas 4.2 and 4.3 we obtain the following corrolary.

Corollary 4.4. *Suppose $o(n) = \frac{1}{1-p_{m,n}}$. Supppose $n = m^{o(1)}$. Then with probability tending to 1 it holds that a random Boolean function $f \in B_{m,n}$ contains only intervals of dimension k, where $k = o(\lg m)$.*

For more overview about supposition $n = m^{o(1)}$ see following Lemma and its Corollary.

Lemma 4.5. *If $m = o(\sqrt{2^n \cdot n^{-1}})$, then P[Every vertex from N_f is isolated] $\to$ 1 as $n \to \infty$.*

Proof. N_f is a graph of boolean function with m vertices. Let $P[1]$ denote the probability that one fixed vertex is in N_f and it is isolated. Let P denote the probability that every vertex from N_f is isolated.

$$
\begin{aligned}
P[1] &= \frac{\binom{2^n-1-n}{m-1}}{\binom{2^n-1}{m-1}} \\
P &\geq \big(P[1]\big)^m = \left[\frac{\binom{2^n-1-n}{m-1}}{\binom{2^n-1}{m-1}}\right]^m = \left[\frac{(2^n-1-n)^{\underline{m-1}}}{(2^n-1)^{\underline{m-1}}}\right]^m \\
&\geq \left[\left(\frac{2^n-n-m+1}{2^n-m+1}\right)^{m-1}\right]^m = \left(1-\frac{n}{2^n-m+1}\right)^{m^2-m} \\
&= \left(1-\frac{1}{\frac{2^n-m+1}{n}}\right)^{\frac{2^n-m+1}{n}\cdot\frac{n\cdot(m^2-m)}{2^n-m+1}}
\end{aligned}
$$

Now we can compute above expression for $m = o(\sqrt{2^n \cdot n^{-1}})$:

$$
\lim_{n\to\infty} P \geq \lim_{n\to\infty} \left(\frac{1}{e}\right)^{\frac{o(2^n)}{2^n-o\left(\sqrt{2^n\cdot n^{-1}}\right)}} = \left(\frac{1}{e}\right)^0 = 1
$$

Because $P \leq 1$ we prooved $P \to 1$ as $n \to \infty$.

□

Corollary 4.6. *If $n = m^{\Omega(1)}$ then P[Every vertex from N_f is isolated] $\to 1$ as $n \to \infty$.*

It means that if we suppose $n = m^{o(1)}$, then we exclude only the cases when N_f contains only isolated vertices. And these cases are not interesting.

5 Comparison of Model A and Model B

In next calculates we will need to use the Corollary 5.2 of following lemma.

Lemma 5.1. *Let k be a function of n. If $k = o(\sqrt{n})$, then $n^{\underline{k}} \sim n^k$.*

Proof. Using Taylor series we get:

$$\ln(1-x) = -\sum_{m=1}^{\infty} \frac{x^m}{m}$$

Using this result we can compute:

$$\frac{n^{\underline{k}}}{n^k} = (1-\frac{1}{n})\cdot(1-\frac{2}{n})\cdots(1-\frac{k-1}{n}) = e^{\sum_{i=1}^{k-1}\ln(1-i\cdot n^{-1})} = e^{-\sum_{i=1}^{k-1}\sum_{m=1}^{\infty}\frac{(i\cdot n^{-1})^m}{m}}$$

Next we use formula for the sum of the integers:

$$\sum_{i=1}^{k-1} i = \frac{k\cdot(k-1)}{2}$$

and the fact that for m fixed:

$$\sum_{i=1}^{k-1} i^m = O(k^{m+1}),$$

where the big-O term is with respect to k, and we find:

$$\frac{n^{\underline{k}}}{n^k} = e^{-\sum_{m=1}^{\infty}\frac{1}{m\cdot n^m}\cdot\sum_{i=1}^{k-1} i^m} \sim e^{o(1)} \sim 1 + o(1),$$

provided $k = o(\sqrt{n})$.

□

Corollary 5.2. *Let k, b and a be functions of n. If $k = o(\sqrt{b})$ and $k = o(\sqrt{a})$, then:*

$$\frac{a^{\underline{k}}}{b^{\underline{k}}} \sim \left(\frac{a}{b}\right)^k$$

If we look carefully at [1], we can notice that only differences between Model A and Model B is in Škoviera's *Proposition 1* (see next theorem) and Lemma 3.1 in this thesis. This is important if we want to proove same results in Model B as Škoviera prooved in Model A. More exactly - Škoviera used the *Proposition 1* in his article only in special case $|V| = 0$ and $|U| = 2^k$ or $|U| = 2^{k+1}$. Here is this special case of Škoviera's *Proposition 1* when $|V| = 0$:

Theorem 5.3 (Škoviera's *Proposition 1*). *Let U be a subset of B^n and $F = \{f \in B_n; U \subset N_f\}$. Then*

$$P_A(F) = p_n^{|U|}.$$

Remark 5.1. From Lemma 3.1 we know that

$$P_B(F) = \frac{m^{\underline{|U|}}}{(2^n)^{\underline{|U|}}}.$$

Lemma 5.4. *Let U be a subset of B^n and $F = \{f \in B_n; U \subset N_f\}$. Let $|U| \leq 2^{k+C}$, where C arbitrary non-negative constant and $k = o(\lg m)$. Then*

$$P_B(F) \sim p_{m,n}^{|U|}.$$

Proof. Using $k = o(\lg m)$ we obtain $k < \lg m^{\frac{1}{3}}$. Thus

$$|U| \leq 2^{k+C} < 2^{\lg m^{\frac{1}{3}}+C} = m^{\frac{1}{3}} \cdot 2^C = o(m^{\frac{1}{2}}).$$

Using Lemma 5.2 with remark 5.1 we get desired result. □

From Corollary 4.4 we know that almost all random Boolean functions in Model B satisfy $k = o(\lg m)$. This means that we can consider $p_{m,n}$ in Model B as equivalent to p_n in Model A. Thus we get all the results for Model B, which computed Škoviera for Model A in [1]. This give us the main result of this chapter: the bounds on a dimension of a maximal interval of a Boolean function from [1] formulated as follows.

Theorem 5.5. *Let $\lim_{n\to\infty} p_{m,n} = p$, where $p \in (0, 1)$. Then with probability tending to 1 as $n \to \infty$ it holds, that dimension k of a maximal interval of a random Boolean function satisfies the following inequalites:*

$$\lg \log_{1/p} n - 1 \leq k \leq \lg \log_{1/p} n + \lg \lg \log_{1/p} n + \varepsilon, \tag{1}$$

where $\varepsilon \to 0$ as $n \to \infty$.

For another view on analogy between Model A and Model B see the rest of this chapter. We will compare the number of vertices of a random Boolean function in Model A with m in Model B.

Definition 5.1. *Let $M(f)$ denote a random variable on B_n such that $M(f)$ is equal to the count of vertices of graph of a function $f \in B_n$ (in Model A). Formally $M(f) = |N_f|$.*

Next we will compute $E(M)$, $Var(M)$ and use Chebyshev's inequality to show that $M \sim 2^n \cdot p_n$.

Lemma 5.6.

$$E(M) = 2^n \cdot p_n$$

Proof. For any vertex $\tilde{\alpha} \in B^n$ we introduce a random variable $\eta_{\tilde{\alpha}}$ (called also an *indicator*) defined by follows:

$$\eta_{\tilde{\alpha}}(f) = \begin{cases} 1 & \text{if } \tilde{\alpha} \in N_f \\ 0 & \text{otherwise.} \end{cases}$$

Clearly $M(f) = \sum_{\tilde{\alpha}} \eta_{\tilde{\alpha}}(f)$ where the summation ranges over all vertices in B^n.

From definition of Model A we know $E(\eta_{\tilde{\alpha}}) = P_A(\eta_{\tilde{\alpha}} = 1) = p_n$.

$|B^n| = 2^n$, therefore

$$E(M) = \sum_{\tilde{\alpha}} E(\eta_{\tilde{\alpha}}) = 2^n \cdot p_n.$$

□

Lemma 5.7.

$$Var(M) = 2^n \cdot p_n \cdot (1 - p_n) \leq 2^n \cdot p_n = E(M)$$

Proof. First recall that $Var(M) = E(M^2) - E^2(M)$. Because of M is expressed as a sum of indicators $\eta_{\tilde{\alpha}}$, then we have

$$M^2 = \Big(\sum_{\tilde{\alpha}} \eta_{\tilde{\alpha}}\Big)^2 = \sum_{(\tilde{\alpha},\tilde{\beta})} \eta_{\tilde{\alpha}} \cdot \eta_{\tilde{\beta}}$$

where last sumamtion ranges over all ordered pair $(\tilde{\alpha}, \tilde{\beta})$ of vertices of B^n.

Let's consider two cases:

1. $\tilde{\alpha} \not\equiv \tilde{\beta}$. The number of such pairs $(\tilde{\alpha}, \tilde{\beta})$ is $2^n \cdot (2^n - 1)$ and $E(\eta_{\tilde{\alpha}} \cdot \eta_{\tilde{\beta}}) = P_A(\eta_{\tilde{\alpha}} \cdot \eta_{\tilde{\beta}} = 1) = p_n^2$.

2. $\tilde{\alpha} \equiv \tilde{\beta}$. The number of such pairs $(\tilde{\alpha}, \tilde{\beta})$ is 2^n and $E(\eta_{\tilde{\alpha}} \cdot \eta_{\tilde{\beta}}) = P_A(\eta_{\tilde{\alpha}} \cdot \eta_{\tilde{\beta}} = 1) = p_n$.

Using this two cases we obtain:

$$E(M^2) = \sum_{\tilde{\alpha},\tilde{\beta}} P_A(\eta_{\tilde{\alpha}} \cdot \eta_{\tilde{\beta}} = 1) = \sum_{\tilde{\alpha} \not\equiv \tilde{\beta}} p_n^2 + \sum_{\tilde{\alpha} \equiv \tilde{\beta}} p_n = 2^n \cdot (2^n - 1) \cdot p_n^2 + 2^n \cdot p_n.$$

Therefore:

$$Var(M) = E(M^2) - E^2(M) = (2^n \cdot p_n)^2 - 2^n \cdot p_n^2 + 2^n \cdot p_n - (2^n \cdot p_n)^2 = 2^n \cdot p_n \cdot (1 - p_n).$$

□

Now we apply Chebyshev's inequality to the random variable M putting $\varepsilon = \phi(n)\sqrt{2^n \cdot p_n}$, where $\phi(n)^{-1} = o(1)$. Using Lemma 5.6 and Lemma 5.7 we obtain:

$$P_A(|M - E(M)| \geq \varepsilon) \leq \frac{Var(M)}{\varepsilon^2} \leq \frac{1}{\phi(n)} \to 0.$$

Hence, $\lim_{n\to\infty} P_A(|M - E(M)| < \varepsilon) = 1$. This gives us result formulated as

Corollary 5.8. *With probability tending to* 1*, as* $n \to \infty$ *for any* $f \in B_n$*, it holds:*

$$2^n \cdot p_n - \phi(n)\sqrt{2^n \cdot p_n} < M < 2^n \cdot p_n - \phi(n)\sqrt{2^n \cdot p_n},$$

where $\phi(n)$ *is an arbitrary function satisfying* $lim_{n\to\infty}\phi(n) = \infty$.

Lemma 5.9 (Estimation of the count of vertices of graph of a random boolean function in Model A)**.** *Suppose that* $p_n = \Omega(\frac{1}{\sqrt{2^n \cdot n}})$*. Then*

$$M \sim E(M) = p_n \cdot 2^n.$$

Proof. Using Corollary 5.8 we obtain:

$$\lim_{n\to\infty} \frac{M}{E(M)} = \lim_{n\to\infty} 1 \pm \sqrt{\frac{1}{p_n \cdot 2^n}} \cdot \phi(n).$$

Now we use supposition $p_n \geq c^2 \cdot \frac{1}{\sqrt{2^n \cdot n}}$, where c is arbitrary positive constant, to show that

$$\lim_{n\to\infty} \sqrt{\frac{1}{p_n \cdot 2^n}} \cdot \phi(n) < \lim_{n\to\infty} \sqrt[4]{\frac{n}{2^n}} \cdot c \cdot \phi(n) = 0.$$

□

Suppositon $p_n = \Omega(\frac{1}{\sqrt{2^n \cdot n}})$ means that random boolean function does not contain only isolated vertices - this can be prooved in Model A analogically to proof of Lemma 4.5.

So if we consider M in Model A as a equivalent to m in Model B, then we get $p_n \cdot 2^n \sim m$, thus $p_n \sim p_{m,n}$.

6 The size of the neighbourhood of given maximal interval in Model B

By the symbol $\Theta(N_K)$ we denote the neighbourhood of the first order of a random maximal interval N_K, that is the set of all maximal intervals of a Boolean function f having a nonempty intersection with N_K.

In this chapter we will proove the lower and upper bound on $|\Theta(N_K)|$. First we will proove that for almost all Boolean functions it holds that all vertices of a random Boolean function are *good* (for exact definition of *good vertices* see Definition 6.5). Then using the inequalities from definition of *good vertices* we will obtain desired upper and lower bound.

This chapter is written accordingly to [2]. We will need to use Theorem 5.5 and therefore we will suppose that

$$\lim_{n\to\infty} p_{m,n} = p, \text{ where } p \in (0,1).$$

Definition 6.1. *$B^{\tilde{\alpha}}_{m,n}$ is set of all n-ary Boolean functions containing a fixed vertex $\tilde{\alpha}$ and another $m-1$ vertices. Formally:*

$$B^{\tilde{\alpha}}_{m,n} = \{f \in B_n; |N_f| = m, \tilde{\alpha} \in N_f\}.$$

Remark 6.1.

$$|B^{\tilde{\alpha}}_{m,n}| = \binom{2^n - 1}{m - 1}$$

Definition 6.2 ($\tilde{\alpha}$-fixed Model B)**.** *$(B^{\tilde{\alpha}}_{m,n}, P^{\tilde{\alpha}}_B)$ is discrete probability space, where $P^{\tilde{\alpha}}_B$ is probability measure defined by follows:*

1. *Every Boolean function $f \in B^{\tilde{\alpha}}_{m,n}$ has the same probability, thus*

$$P^{\tilde{\alpha}}_B(\{f\}) = \frac{1}{\binom{2^n-1}{m-1}}.$$

2. *For an arbitrary $S \subset B^{\tilde{\alpha}}_{m,n}$ we set $P^{\tilde{\alpha}}_B(S) = \sum_{f\in S} P^{\tilde{\alpha}}_B(\{f\})$.*

Lemma 6.1. *Let $\tilde{\alpha} \in B^n$ be a fixed vertex. Let $U \subset B^n$ and $\tilde{\alpha} \in U$ and $F = \{f \in B^{\tilde{\alpha}}_{m,n}; U \subset N_f\}$. Suppose $|U| = m^{o(1)}$. Then*

$$P^{\tilde{\alpha}}_B(F) \sim p^{|U|-1}.$$

Proof. Analogically to Lemma 3.1 we obtain:

$$P^{\tilde{\alpha}}_B(F) = \frac{(m-1)^{\underline{|U|-1}}}{(2^n-1)^{\underline{|U|-1}}}.$$

Analogically to Lemma 5.4 (using $|U| = m^{o(1)}$) we obatin:

$$P_B^{\tilde{\alpha}}(F) \sim \Big(\frac{m-1}{2^n-1}\Big)^{|U|-1}.$$

Now we can show, that this is asimptotically equivalent to $p^{|U|-1}$:

$$\lim_{n\to\infty} \frac{\left(\frac{m-1}{2^n-1}\right)^{|U|-1}}{\left(\frac{m}{2^n}\right)^{|U|-1}} = \lim_{n\to\infty} \left(\frac{1-\frac{1}{m}}{1-\frac{1}{2^n}}\right)^{|U|-1} = \lim_{n\to\infty} \frac{\left(\frac{1}{e}\right)^{\frac{|U|-1}{m}}}{\left(\frac{1}{e}\right)^{\frac{|U|-1}{2^n}}} = 1.$$

The last equation is satisfied with condition $|U| = o(m)$, which can reader clearly obtain from supposition $|U| = m^{o(1)}$. □

Definition 6.3. *Let $X^{\tilde{\alpha}}_{m,n,k}$ denote a random variable on $B^{\tilde{\alpha}}_{m,n}$ such that $X^{\tilde{\alpha}}_{m,n,k}(f)$ is equal to the count of k-dimensional intervals of a function $f \in B^{\tilde{\alpha}}_{m,n}$.*

Lemma 6.2. *Let $k = n^{o(1)}$. Then*

$$E(X^{\tilde{\alpha}}_{m,n,k}) \sim \binom{n}{k} \cdot p^{2^k-1}.$$

Proof. For any k-dimensional subcube K of the cube B^n containing vertex $\tilde{\alpha}$ we introduce a random variable η_K (called also an *indicator*) defined by follows:

$$\eta_K(f) = \begin{cases} 1 & \text{if } K \subseteq N_f \\ 0 & \text{otherwise.} \end{cases}$$

Clearly $i_{m,n,k}(f) = \sum_K \eta_K(f)$ where the summation ranges over all k-dimensional subcubes of B^n.

Using Lemma 6.1 we can compute that:

$$E(\eta_K) = P_B^{\tilde{\alpha}}(\eta_K = 1) = P_B^{\tilde{\alpha}}(K \subseteq N_f) \sim p^{2^k-1}$$

There are $\binom{n}{k}$ k-dimensional subcubes in B^n containing a fixed vertex $\tilde{\alpha}$. Thus,

$$E(X^{\tilde{\alpha}}_{m,n,k}) = \sum_K E(\eta_K) \sim \binom{n}{k} \cdot p^{2^k-1}.$$

□

Lemma 6.3. *Let $k = n^{o(1)}$. Then*

$$Var(X^{\tilde{\alpha}}_{m,n,k}) \lesssim \binom{n}{k}^2 \cdot p^{2^{k+1}-1} \cdot \left[\frac{k^3}{n \cdot p^2} + \frac{k}{p^{2^k} \cdot \binom{n}{k}}\right].$$

Proof. First recall that

$$Var(X^{\tilde{\alpha}}_{m,n,k}) = E((X^{\tilde{\alpha}}_{m,n,k})^2) - E^2(X^{\tilde{\alpha}}_{m,n,k}).$$

Because of $X^{\tilde{\alpha}}_{m,n,k}$ is expressed as a sum of indicators η_K, we have

$$(X^{\tilde{\alpha}}_{m,n,k}(f))^2 = \Big(\sum_K \eta_K(f)\Big)^2 = \sum_{(K,L)} \eta_K(f) \cdot \eta_L(f),$$

where last sumamtion ranges over all ordered pair (K, L) of k-dimensional intervals containig fixed vertex $\tilde{\alpha}$.

The intersection of K and L is nonempty, because they both contain vertex $\tilde{\alpha}$. So $K \cap L$ is j-dimensional interval with $0 \leq j \leq k$ and $|K \cup L| = 2^{k+1} - 2^j$. Thus using Lemma 6.1 we obtain

$$E(\eta_K \cdot \eta_L) = P^{\tilde{\alpha}}_B(\eta_K \cdot \eta_L = 1) \sim p^{2^{k+1}-2^j-1}.$$

The number of such pairs (K, L) with dimension $K \cap L = j$ is

$$\binom{n}{j}\binom{n-j}{k-j}\binom{n-k}{k-j}.$$

Thus

$$E((X^{\tilde{\alpha}}_{m,n,k})^2) = \sum_{K,L} P^{\tilde{\alpha}}_B(\eta_K \cdot \eta_L = 1) \sim \sum_{j=0}^{k} \binom{n}{j}\binom{n-j}{k-j}\binom{n-k}{k-j} p^{2^{k+1}-2^j-1}.$$

Now we can estimate $Var(X^{\tilde{\alpha}}_{m,n,k})$:

$$\begin{aligned}
Var(X^{\tilde{\alpha}}_{m,n,k}) &\sim \sum_{j=0}^{k} \binom{n}{j}\binom{n-j}{k-j}\binom{n-k}{k-j} p^{2^{k+1}-2^j-1} - \binom{n}{k}^2 p^{2^{k+1}-2} \\
&= \binom{n}{k}\binom{n-k}{k} p^{2^{k+1}-2} - \binom{n}{k}^2 p^{2^{k+1}-2} + \\
&\quad + \sum_{j=1}^{k} \binom{n}{j}\binom{n-j}{k-j}\binom{n-k}{k-j} p^{2^{k+1}-2^j-1} \\
&\leq \sum_{j=1}^{k} \binom{n}{j}\binom{n-j}{k-j}\binom{n-k}{k-j} p^{2^{k+1}-2^j-1} \\
&= \binom{n}{k} p^{2^{k+1}-1} \sum_{j=1}^{k} \binom{k}{j}\binom{n-k}{k-j} p^{-2^j}
\end{aligned}$$

We denote $\binom{k}{j}\binom{n-k}{k-j}p^{-2^j}$ by b_j and estimate the ratio

$$\frac{b_{j+1}}{b_j} = \frac{p^{-2^j}(k-j)^2}{(j+1)(n-2k+j+1)} = c_j.$$

Let's consider two cases:

- $j < \lg\log_{1/p} n$
  Using simply fact $0 < j < k$ we obtain:
  $$c_j < \frac{p^{-2^{\lg\log_{1/p} n-1}} \cdot k^2}{2\cdot(n-2k+1)} = \frac{\sqrt{n}\cdot k^2}{2\cdot(n-2k+1)}.$$
  Using supposition $k = n^{o(1)}$ we see that c_j tends to 0 as n tends to ∞.

- $j > \lg\log_{1/p} n$
  Using $j \geq \lg\log_{1/p} n + x$, where x is arbitrary little constant, we obtain:
  $$c_j \geq \frac{p^{-2^j}\cdot 1}{k\cdot(n-k+1)} \geq \frac{n^{2^x}}{k\cdot n} = \frac{n^a}{k},$$
  where $a = 2^x - 1$ is a positive constant.
  Using supposition $k = n^{o(1)}$ we see that c_j tends to ∞ as n tends to ∞.

So the sequence b_j is decreasing till $\lg\log_{1/p} n$ and increasing after. Therefore the maximal value of b_j is b_1 or b_k. Thus

$$\sum_{j=1}^{k} b_j \lesssim k\cdot(b_1+b_k) = k\cdot\left(k\binom{n-k}{k-1}p^{-2} + p^{-2^k}\right).$$

Using simply fact $\binom{n-k}{k-1} \leq \binom{n-1}{k-1} = \binom{n}{k}\cdot\frac{k}{n}$ we obtain:

$$\sum_{j=1}^{k} b_j \lesssim k^3\binom{n}{k}p^{-2} + kp^{-2^k},$$

and

$$Var(X^{\tilde{\alpha}}_{m,n,k}) \lesssim \binom{n}{k}^2 \cdot p^{2^{k+1}-1}\cdot\left[\frac{k^3}{n\cdot p^2} + \frac{k}{p^{2^k}\cdot\binom{n}{k}}\right].$$

□

Corollary 6.4. *Let k be an integer satisfying (1). Then*

$$Var(X^{\tilde{\alpha}}_{m,n,k}) \lesssim E(X^{\tilde{\alpha}}_{m,n,k})^2\cdot\frac{c_1\cdot log_{1/p}n}{n},$$

where c_1 is a positive constant.

Proof. First recall

$$Var(X_{m,n,k}^{\tilde{\alpha}}) \lesssim E(X_{m,n,k}^{\tilde{\alpha}})^2 \cdot p \cdot \left[\frac{k^3}{n \cdot p^2} + \frac{k}{p^{2^k} \cdot \binom{n}{k}}\right].$$

From (1) we have inequality $k \lesssim \sqrt[3]{log_{1/p} n}$ and using it we obtain

$$\frac{k^3}{n \cdot p} \lesssim \frac{\log_{1/p} n}{n \cdot p}. \tag{2}$$

The reader can easily verify that for k satisfying (1) it holds that $p^{2^k} \cdot \binom{n}{k}$ is decreasing for $k < \lg \log_{1/p} n$ and increasing after. Therefore, using $k \lesssim 2 \cdot \lg \log_{1/p} n \lesssim n$ (it holds from (1)), we obtain

$$\frac{k \cdot p}{p^{2^k} \cdot \binom{n}{k}} \lesssim \frac{n}{p^{2^{\lg \log_{1/p} n + 1}} \cdot n^4} \lesssim \frac{n}{\frac{1}{n^2} \cdot n^4} \lesssim \frac{1}{n},$$

so we have

$$\frac{k \cdot p}{p^{2^k} \cdot \binom{n}{k}} \lesssim \frac{\log_{1/p} n}{n}. \tag{3}$$

Using (2) and (3) we get

$$Var(X_{m,n,k}^{\tilde{\alpha}}) \lesssim E(X_{m,n,k}^{\tilde{\alpha}})^2 \cdot \frac{\log_{1/p} n}{n} \cdot (1/p + 1).$$

□

Definition 6.4. *Let $Y_{m,n,k}^{\tilde{\alpha}}$ denote a random variable on $B_{m,n}^{\tilde{\alpha}}$ such that $Y_{m,n,k}^{\tilde{\alpha}}(f)$ is equal to the count of k-dimensional maximal intervals of a function $f \in B_{m,n}^{\tilde{\alpha}}$.*

Lemma 6.5. *Let $k = n^{o(1)}$. Then*

$$E(Y_{m,n,k}^{\tilde{\alpha}}) \sim \binom{n}{k} \cdot p^{2^k - 1} \cdot \left(1 - p^{2^k - 1}\right)^{n-k}$$

Proof. Let $P(N_K)$ denote the probability that a fixed maximal interval N_K containing a vertex $\tilde{\alpha}$ belongs to the set N_f of a random Boolean function f. Without detriment to generality we can assume that $\tilde{\alpha} = (0, 0, \cdots, 0)$ and $N_K = \{(\gamma_1, \gamma_2, \cdots, \gamma_k, 0, 0, \cdots, 0); \gamma_i \in \{0, 1\}, i = 1, 2, \cdots, k\}$. To abbreviate the notation we will use $\star$-notation

$$\begin{aligned} N_K &= (\underbrace{\star, \cdots, \star}_{k}, \underbrace{0, \cdots, 0}_{n-k}) \\ N_{K,i} &= (\underbrace{\star, \cdots, \star}_{k}, \underbrace{0, \cdots, 0}_{i-1}, 1, \underbrace{0, \cdots, 0}_{n-k-i}), \end{aligned}$$

where $N_{K,i}$ is defined for $i = 1, 2, \cdots, n-k$.

Obviously $N_K \cap N_{K,i} = \emptyset$ for each defined i and $N_{K,i} \cap N_{K,j} = \emptyset$ for each defined $i \neq j$ and

$$P(N_K) = P_B^{\tilde{\alpha}}[N_K \subset N_f] \cdot \prod_{i=1}^{n-k} P_B^{\tilde{\alpha}}[N_{K,i} \not\subseteq N_f]$$

Using Lemma 6.1 we can compute that:

$$P(N_K) \sim p^{2^k-1} \cdot \left(1 - p^{2^k-1}\right)^{n-k}.$$

There are $\binom{n}{k}$ k-dimensional subcubes in B^n containing a fixed vertex $\tilde{\alpha}$. Therefore

$$E(Y_{m,n,k}^{\tilde{\alpha}}) = \sum_{N_K} P(N_K) \sim \binom{n}{k} \cdot p^{2^k-1} \cdot \left(1 - p^{2^k-1}\right)^{n-k}.$$

□

Lemma 6.6. *Let k be an integer satisfying (1). Then*

$$Var(Y_{m,n,k}^{\tilde{\alpha}}) \lesssim \frac{1}{n^{c_2}} \cdot E^2(X_{m,n,k}^{\tilde{\alpha}}),$$

where c_2 is an arbitrary constant satisfying $0 < c_2 < 1$.

Proof. Using Lemma 6.5 and the same idea as in proof of Lemma 6.3 we have

$$\begin{aligned} Var(X_{m,n,k}^{\tilde{\alpha}}) &= E((Y_{m,n,k}^{\tilde{\alpha}})^2) - E^2(Y_{m,n,k}^{\tilde{\alpha}}) \\ Var(Y_{m,n,k}^{\tilde{\alpha}}) &= \sum_{j=0}^{k} \binom{n}{j}\binom{n-j}{k-j}\binom{n-k}{k-j} P'_j(N_K, N_L) \\ &\quad - \left[\binom{n}{k} \cdot p^{2^k-1} \cdot \left(1 - p^{2^k-1}\right)^{n-k}\right]^2, \end{aligned}$$

where $P'_j(N_K, N_L)$ denotes the probability that a random Boolean function contains k-dimensional maximal intervals N_K and N_L, both containg a fixed vertex $\tilde{\alpha}$ and $N_K \cap N_L$ is a j-dimensional interval. The probability $P'_j(N_K, N_L)$ can be estimated in the following way

$$P'_j(N_K, N_L) \lesssim P_j(N_K, N_L) = p^{2^{k+1}-2^j-1},$$

where $P_j(N_K, N_L)$ denotes the probability that a random Boolean function contains k-dimensional intervals N_K and N_L (not maximal!), both containg a fixed vertex $\tilde{\alpha}$ and $N_K \cap N_L$ is a j-dimensional interval. Thus we have

$$\begin{aligned} Var(Y^{\tilde{\alpha}}_{m,n,k}) \leq & \binom{n}{k}\binom{n-k}{k} P'_0(N_K, N_L) \\ & + \sum_{j=1}^{k} \binom{n}{j}\binom{n-j}{k-j}\binom{n-k}{k-j} P_j(N_K, N_L) \\ & - \left[\binom{n}{k} \cdot p^{2^k-1} \cdot \left(1-p^{2^k-1}\right)^{n-k}\right]^2. \end{aligned}$$

The expression $\sum_{j=1}^{k} \binom{n}{j}\binom{n-j}{k-j}\binom{n-k}{k-j} P_j(N_K, N_L)$ was already estimate in proof of Lemma 6.3, so we have

$$\begin{aligned} Var(Y^{\tilde{\alpha}}_{m,n,k}) \lesssim & \binom{n}{k}\binom{n-k}{k} P'_0(N_K, N_L) - \left[\binom{n}{k} \cdot p^{2^k-1} \cdot \left(1-p^{2^k-1}\right)^{n-k}\right]^2 \\ & + \frac{c_1 \cdot log_{1/p} n}{n} \cdot E(X^{\tilde{\alpha}}_{m,n,k})^2. \end{aligned}$$

Let's denote

$$Z = \binom{n}{k}\binom{n-k}{k} P'_0(N_K, N_L) - \left[\binom{n}{k} \cdot p^{2^k-1} \cdot \left(1-p^{2^k-1}\right)^{n-k}\right]^2$$

and estimate the expression Z. Using $\binom{n-k}{k} \leq \binom{n}{k}$ and next Lemma 6.7 we have

$$\begin{aligned} Z \lesssim & \binom{n}{k}^2 p^{2^{k+1}-2} \cdot \left(1 - 2p^{2^k-1} + p^{2^{k+1}-3}\right)^{n-2k} \\ & - \left[\binom{n}{k} \cdot p^{2^k-1} \cdot \left(1-p^{2^k-1}\right)^{n-k}\right]^2 \\ = & \left[\binom{n}{k} \cdot p^{2^k-1}\right]^2 \cdot \left[\left(1 - 2p^{2^k-1} + p^{2^{k+1}-3}\right)^{n-2k} - \left(1-p^{2^k-1}\right)^{2(n-k)}\right] \\ \lesssim & E^2(X^{\tilde{\alpha}}_{m,n,k}) \cdot \left(1-p^{2^k-1}\right)^{2(n-k)} \cdot \left[\frac{(1 - 2p^{2^k-1} + p^{2^{k+1}-3})^{n-2k}}{(1-p^{2^k-1})^{2(n-k)}} - 1\right] \end{aligned}$$

Using well-known fact

$$\lim_{n\to\infty} (1 - 1/n)^n = e^{-1}$$

we obtain

$$\begin{aligned} Z &\lesssim E^2(X^{\tilde{\alpha}}_{m,n,k}) \cdot \exp\big(- p^{2^k-1} \cdot 2(n-k)\big) \cdot \\ &\quad \cdot \left[\frac{\exp\big((-2p^{2^k-1} + p^{2^{k+1}-3}) \cdot (n-2k)\big)}{\exp\big(- p^{2^k-1} \cdot 2(n-k)\big)} - 1 \right] \\ &= E^2(X^{\tilde{\alpha}}_{m,n,k}) \cdot \exp\big(- p^{2^k-1} \cdot 2(n-k)\big) \cdot \\ &\quad \cdot \left[\exp\big(n \cdot p^{2^{k+1}} \cdot p^{-3} + 2k \cdot (p^{2^k-1} - p^{2^{k+1}-4})\big) - 1 \right]. \end{aligned}$$

Because k is satisfying (1) then we have $k \cdot p^{2^k} \sim 0$ and $k \cdot p^{2^{k+1}} \sim 0$. Thus

$$Z \lesssim E^2(X^{\tilde{\alpha}}_{m,n,k}) \cdot \exp\big(- 2p^{-1} \cdot np^{2^k}\big) \cdot \left[\exp\big(p^{-3} \cdot np^{2^{k+1}}\big) - 1 \right].$$

We see that for the estimation of Z are important expressions np^{2^k} and $np^{2^{k+1}}$. Let's consider two cases for an arbitrary little positive constant ε:

1. $\lg \log_{1/p} n - 1 \le k \le \lg \log_{1/p} n - \varepsilon$
   Now we have

   $$\exp\left(-2p^{-1} \cdot np^{2^k}\right) \le \exp\left(-2p^{-1} \cdot np^{2^{\varepsilon} \log_{1/p} n}\right) = e^{-2p^{-1} \cdot n \cdot (1 - 1/2^{\varepsilon})} = e^{-n \cdot z_1},$$

   where z_1 is an arbitrary little positive constant, and

   $$\exp\left(p^{-3} \cdot np^{2^{k+1}}\right) - 1 \le \exp\left(p^{-3} \cdot np^{\log_{1/p} n}\right) - 1 = e^{p^{-3}} - 1 = z_2,$$

   where z_2 is an positive constant. Thus

   $$Z \lesssim E^2(X^{\tilde{\alpha}}_{m,n,k}) \cdot a^n, \tag{4}$$

   where a is a positive constant and $a < 1$.

2. $\lg \log_{1/p} n - \varepsilon < k$
   Now we have

   $$\exp\big(- 2p^{-1} \cdot np^{2^k}\big) \le 1$$

   And using simply fact that $e^x - 1 \sim x$ when $x \to 0$ we obtain

   $$\exp\left(p^{-3} \cdot np^{2^{k+1}}\right) - 1 < \exp\left(p^{-3} \cdot np^{2^{1-\varepsilon} \log_{1/p} n}\right) - 1 = e^{p^{-3} \cdot n^{1 - 2^{1-\varepsilon}}} - 1 \lesssim \frac{1}{n^{z_3}},$$

   where z_3 is an arbitrary positive constant satisfying $z_3 < 1$. Thus

   $$Z \lesssim E^2(X^{\tilde{\alpha}}_{m,n,k}) \cdot \frac{1}{n^{z_3}}. \tag{5}$$

From (4) and (5) we have

$$Z \lesssim E^2(X^{\tilde{\alpha}}_{m,n,k}) \cdot \frac{1}{n^{z_4}},$$

where z_4 is an arbitrary positive constant satisfying $z_4 < 1$. Therefore

$$Var(Y^{\tilde{\alpha}}_{m,n,k}) \le \frac{1}{n^{c_2}} \cdot E^2(X^{\tilde{\alpha}}_{m,n,k}).$$

□

Lemma 6.7. *Let k be an integer satisfying (1) and $P'_j(N_K, N_L)$ denotes the probability that a random Boolean function contains k-dimensional maximal intervals N_K and N_L, both containg a fixed vertex $\tilde{\alpha}$ and $N_K \cap N_L$ is a j-dimensional interval. Then*

$$P'_0(N_K, N_L) \sim p^{2^{k+1}-2} \cdot (1 - 2 \cdot p^{2^k - 1} + p^{2^{k+1}-3})^{n-2k}.$$

Proof. Without detriment to generality we can assume that $\tilde{\alpha} = (0, 0, \cdots, 0)$ and

$$\begin{aligned} N_K &= (\underbrace{\star, \cdots, \star}_{k}, \underbrace{0, \cdots, 0}_{k}, \underbrace{0, \cdots, 0}_{n-2k}) \\ N_L &= (\underbrace{0, \cdots, 0}_{k}, \underbrace{\star, \cdots, \star}_{k}, \underbrace{0, \cdots, 0}_{n-2k}). \end{aligned}$$

Next we will use the following notation

$$\begin{aligned} N_{K,r} &= (\underbrace{\star, \cdots, \star}_{k}, \underbrace{0, \cdots, 0}_{r-k-1}, 1, \underbrace{0, \cdots, 0}_{n-r}) \\ N_{L,s} &= (\underbrace{0, \cdots, 0}_{s-1}, 1, \underbrace{0, \cdots, 0}_{k-s}, \underbrace{\star, \cdots, \star}_{k}, \underbrace{0, \cdots, 0}_{n-2k}), \text{ or} \\ N_{L,s} &= (\underbrace{0, \cdots, 0}_{k}, \underbrace{\star, \cdots, \star}_{k}, \underbrace{0, \cdots, 0}_{s-2k-1}, 1, \underbrace{0, \cdots, 0}_{n-s}) \\ \tilde{\alpha}_t &= (\underbrace{0, \cdots, 0}_{t-1}, 1, \underbrace{0, \cdots, 0}_{n-t}) \end{aligned}$$

where $N_{K,r}$ is defined for $r = \overline{k+1, n}$ and $N_{L,s}$ is defined for $s = \overline{1, k}$ or $s = \overline{2k+1, n}$ and $\tilde{\alpha}_t$ is defined for $t = \overline{1, n}$. Obviously

$$P'_0(N_K, N_L) = P[N_K \cup N_L \subset N_f, (\forall r) N_{K,r} \not\subseteq N_f, (\forall s) N_{L,s} \not\subseteq N_f].$$

Using the probability theory about (in)dependent events we have

$$\begin{aligned} P_0'(N_K, N_L) &= P[N_K \cup N_L \subset N_f] \cdot P[(\forall r) N_{K,r} \not\subseteq N_f, \\ &\quad (\forall s) N_{L,s} \not\subseteq N_f | N_K \cup N_L \subset N_f] \\ &\sim p^{2^{k+1}-2} \cdot P[(\forall r = \overline{k+1, 2k}) N_{K,r} - \{\tilde{\alpha}_r\} \not\subseteq N_f, \\ &\quad (\forall r = \overline{2k+1, n}) N_{K,r} \not\subseteq N_f, \\ &\quad (\forall s = \overline{1, k}) N_{L,s} - \{\tilde{\alpha}_s\} \not\subseteq N_f, \\ &\quad (\forall s = \overline{2k+1, n}) N_{L,s} \not\subseteq N_f] \\ &\sim p^{2^{k+1}-2} \cdot P[(\forall r = \overline{k+1, 2k}) N_{K,r} - \{\tilde{\alpha}_r\} \not\subseteq N_f, \\ &\quad (\forall s = \overline{1, k}) N_{L,s} - \{\tilde{\alpha}_s\} \not\subseteq N_f] \\ &\quad \cdot P[(\forall r = \overline{2k+1, n}) N_{K,r} \not\subseteq N_f, \\ &\quad (\forall s = \overline{2k+1, n}) N_{L,s} \not\subseteq N_f]. \end{aligned}$$

Let's estimate above probabilities separately:

1.

$$\begin{aligned} &P[(\forall r = \overline{2k+1, n}) N_{K,r} \not\subseteq N_f, (\forall s = \overline{2k+1, n}) N_{L,s} \not\subseteq N_f] \\ = &\prod_{i=2k+1}^{n} P[N_{K,i} \not\subseteq N_f, N_{L,i} \not\subseteq N_f] \\ = &\prod_{i=2k+1}^{n} \Big(P[\tilde{\alpha}_i \in N_f] \cdot P[N_{K,i} \not\subseteq N_f, N_{L,i} \not\subseteq N_f | \tilde{\alpha}_i \in N_f] \\ &+ P[\tilde{\alpha}_i \notin N_f] \cdot P[N_{K,i} \not\subseteq N_f, N_{L,i} \not\subseteq N_f | \tilde{\alpha}_i \notin N_f] \Big) \\ = &\prod_{i=2k+1}^{n} \Big(P[\tilde{\alpha}_i \in N_f] \cdot P[N_{K,i} - \{\tilde{\alpha}_i\} \not\subseteq N_f] \cdot P[N_{L,i} - \{\tilde{\alpha}_i\} \not\subseteq N_f] + \\ &+ P[\tilde{\alpha}_i \notin N_f] \cdot 1 \Big) \\ \sim &\prod_{i=2k+1}^{n} \Big(p \cdot (1 - p^{2^k-2})^2 + 1 - p \Big) \\ = &(1 - 2 \cdot p^{2^k-1} + p^{2^{k+1}-3})^{n-2k} \end{aligned}$$

2. In following calculates we will use also (1).

$$P[(\forall r = \overline{k+1, 2k}) N_{K,r} - \{\tilde{\alpha}_r\} \not\subseteq N_f, (\forall s = \overline{1, k}) N_{L,s} - \{\tilde{\alpha}_s\} \not\subseteq N_f]$$

$$\begin{aligned}
&\geq \Big(\prod_{r=k+1}^{2k} P[N_{K,r} - \{\tilde{\alpha}_r\} \nsubseteq N_f]\Big) \cdot \Big(\prod_{s=1}^{k} P[N_{L,s} - \{\tilde{\alpha}_s\} \nsubseteq N_f]\Big) \\
&= (1 - p^{2^k - 2})^{2k} \\
&\gtrsim (1 - p^{2^{\lg\log_{1/p} n - 1} - 2})^{4\lg\log_{1/p} n} \\
&= \left(1 - \frac{1}{p^2\sqrt{n}}\right)^{4\lg\log_{1/p} n} \\
&\sim e^{-\frac{4\lg\log_{1/p} n}{p^2\sqrt{n}}} \sim e^0 = 1
\end{aligned}$$

Probability of every event is ≤ 1, thus

$$P[(\forall r = \overline{k+1, 2k}) N_{K,r} - \{\tilde{\alpha}_r\} \nsubseteq N_f, (\forall s = \overline{1,k}) N_{L,s} - \{\tilde{\alpha}_s\} \nsubseteq N_f] \sim 1$$

□

Definition 6.5. *A vertex $\tilde{\alpha} \in N_f$, satisfying the condition*

$$|Y^{\tilde{\alpha}}_{m,n,k} - E(Y^{\tilde{\alpha}}_{m,n,k})| \geq \frac{1}{\log_{1/p} n} \cdot E(X^{\tilde{\alpha}}_{m,n,k})$$

will be called bad vertex of a random Boolean function f, otherwise, the vertex $\tilde{\alpha}$ will be called good vertex of a random Boolean function f.

Lemma 6.8. *Let k be an integer satisfying (1), and let $P_n(\tilde{\alpha})$ be the probability that $\tilde{\alpha}$ is a bad vertex of an n-ary random Boolean function. Then*

$$P_n(\tilde{\alpha}) \lesssim \frac{\log^2_{1/p} n}{n^{c_2}}$$

Proof. From Chebyshev's inequality and Lemma 6.6 we have

$$P_n(\tilde{\alpha}) \leq \frac{Var(Y^{\tilde{\alpha}}_{m,n,k}) \cdot \log^2_{1/p} n}{\big(E(Y^{\tilde{\alpha}}_{m,n,k})\big)^2} \lesssim \frac{\log^2_{1/p} n}{n^{c_2}}$$

□

Now we estimate the number of bad vertices in a maximal interval of a random Boolean function.

Lemma 6.9. *Let k be an integer satisfying (1), and let N_K denote a fixed k-dimensional maximal interval of a random Boolean function f. Let $b_k(f)$ be a random variable expressing the number of bad vertices in N_K. Then probability that*

$$b_k(f) = 0$$

tending to 1 as $n \to \infty$.

Proof. The expectation of number of bad vertices in N_K is

$$E(b_k) = 2^k \cdot P_n(\tilde{\alpha}) \lesssim 2^k \cdot \frac{\log^2_{1/p} n}{n^{c_2}}$$

and consequently (by Markov's inequality)

$$P\left[b_k(f) \leq 2^k \cdot \frac{\log^3_{1/p} n}{n^{c_2}}\right] \geq 1 - P\left[b_k(f) \geq 2^k \cdot \frac{\log^3_{1/p} n}{n^{c_2}}\right] \geq 1 - \frac{E(b_k)}{2^k \cdot \frac{\log^3_{1/p} n}{n^{c_2}}} \gtrsim$$
$$\gtrsim 1 - \frac{1}{\log_{1/p} n},$$

what is tending to 1 as $n \to \infty$. Using (1) we have

$$2^k \cdot \frac{\log^3_{1/p} n}{n^{c_2}} \to 0 \text{ as } n \to \infty,$$

thus

$$P[b_k(f) \to 0] \to 1 \text{ as } n \to \infty.$$

Because $b_k(f)$ is non-negative integer we have

$$P[b_k(f) = 0] \to 1 \text{ as } n \to \infty..$$

□

It means that for almost all Boolean functions every maximal interval N_K contains no bad vertex, so we can suppose that N_K contains only good vertices. Using this supposition we obtain the upper and lower bound of $|\Theta(N_K)|$.

Theorem 6.10. *Let f be an n-ary random boolean function and let N_K induce a maximal interval of f. Then the following inequalities hold with probability tending to 1 as $n \to \infty$:*

$$n^{(1-\varepsilon_n)\lg\log_{1/p} n} \lesssim |\Theta(N_K)| \lesssim n^{(1+\varepsilon'_n)\lg\log_{1/p} n},$$

where ε_n,$\varepsilon'_n \to 0$ as $n \to \infty$.

Proof. We set $k_0 = \lg\log_{1/p} n - 1$ and $k_1 = \lg\log_{1/p} n + \lg\lg\log_{1/p} n + 1$ to abbreviate the notation.

First recall that for almost all random Boolean function k satisfies (1), thus

$$\lceil k_0 \rceil \leq k \leq \lfloor k_1 \rfloor$$

and N_K contains only good vertices.

Upper bound. Using supposition that all vertices in N_K are good we have

$$
\begin{aligned}
|\Theta(N_K)| &\lesssim \sum_{\tilde{\alpha}\in N_K}\sum_{k=\lceil k_0\rceil}^{\lfloor k_1\rfloor}\left[E(Y^{\tilde{\alpha}}_{m,n,k})+\frac{E(X^{\tilde{\alpha}}_{m,n,k})}{\log_{1/p} n}\right]\\
&\lesssim \sum_{\tilde{\alpha}\in N_K}\sum_{k=\lceil k_0\rceil}^{\lfloor k_1\rfloor}\left[E(X^{\tilde{\alpha}}_{m,n,k})+E(X^{\tilde{\alpha}}_{m,n,k})\right]\\
&\lesssim \sum_{k=\lceil k_0\rceil}^{\lfloor k_1\rfloor} 2^k\cdot 2\cdot\binom{n}{k}p^{2^k-1}\\
&\leq (k_1-k_0+1)\cdot 2^{k_1+1}\cdot n^{k_1}\cdot p^{2^{k_0}-1}\\
&\leq (2\cdot\lg\lg\log_{1/p} n)\cdot(\log_{1/p} n\cdot\lg\log_{1/p} n\cdot 4)\cdot n^{k_1}\cdot p^{1/2\log_{1/p} n-1}\\
&= \frac{8}{p}\lg\lg\log_{1/p} n\cdot\log_{1/p} n\cdot\lg\log_{1/p} n\cdot n^{-1/2}\cdot n^{k_1}\\
&\lesssim \log^3_{1/p} n\cdot n^{\lg\log_{1/p} n+\lg\lg\log_{1/p} n+1/2}\\
&\lesssim n^{\lg\log_{1/p} n+2\cdot\lg\lg\log_{1/p} n}\\
&= n^{\left(1+\frac{2\cdot\lg\lg\log_{1/p} n}{\lg\log_{1/p} n}\right)\cdot\lg\log_{1/p} n}.
\end{aligned}
$$

Lower bound. Using supposition that at least one vertex in N_K is good we have

$$
\begin{aligned}
|\Theta(N_K)| &\gtrsim \sum_{k=\lceil k_0\rceil}^{\lfloor k_1\rfloor}\left[E(Y^{\tilde{\alpha}}_{m,n,k})-\frac{E(X^{\tilde{\alpha}}_{m,n,k})}{\log_{1/p} n}\right]\\
&\gtrsim \sum_{k=\lceil k_0\rceil}^{\lfloor k_1\rfloor}\binom{n}{k}p^{2^k-1}\left[\left(1-p^{2^k-1}\right)^{n-k}-\frac{1}{\log_{1/p} n}\right]\\
&\geq \sum_{k=\lceil k_0+2\rceil}^{\lceil k_0+2\rceil}\binom{n}{k}p^{2^k-1}\left[\left(1-p^{2^k-1}\right)^{n-k}-\frac{1}{\log_{1/p} n}\right]\\
&\geq \binom{n}{\lg\log_{1/p} n+1}\cdot p^{2^{\lg\log_{1/p} n+2}-1}\cdot\\
&\quad\cdot\left[\left(1-p^{2^{\lg\log_{1/p} n+1}-1}\right)^{n-\lg\log_{1/p} n-1}-\frac{1}{\log_{1/p} n}\right]\\
&\geq \left(\frac{n}{\lg\log_{1/p} n+1}\right)^{\lg\log_{1/p} n+1}\cdot p^{-1}n^{-4}.
\end{aligned}
$$

$$
\begin{aligned}
&\cdot\Big[\big(1-\frac{1}{pn^2}\big)^{n-\lg\log_{1/p}n-1}-\frac{1}{\log_{1/p}n}\Big]\\
\gtrsim\ & \Big(\frac{n}{\lg\log_{1/p}n+1}\Big)^{\lg\log_{1/p}n+1}\cdot p^{-1}n^{-4}\cdot\Big[\big(\frac{1}{e}\big)^{\frac{n-\lg\log_{1/p}n-1}{pn^2}}-\frac{1}{\log_{1/p}n}\Big]\\
\gtrsim\ & \Big(\frac{n}{\lg\log_{1/p}n+1}\Big)^{\lg\log_{1/p}n+1}\cdot p^{-1}n^{-4}\cdot\Big[\big(1-\frac{1}{\log_{1/p}n}\Big]\\
\gtrsim\ & \Big(\frac{n}{\lg\log_{1/p}n+1}\Big)^{\lg\log_{1/p}n+1}\cdot n^{-5}\\
=\ & n^{\lg\log_{1/p}n-4}\cdot\big(\lg\log_{1/p}n+1\big)^{-(\lg\log_{1/p}n+1)}\\
=\ & n^{\lg\log_{1/p}n-4}\cdot n^{-\log_n(\lg\log_{1/p}n+1)\cdot(\lg\log_{1/p}n+1)}\\
=\ & n^{\lg\log_{1/p}n-4}\cdot n^{-\frac{\lg(\lg\log_{1/p}n+1)}{\lg n}\cdot(\lg\log_{1/p}n+1)}\\
\gtrsim\ & n^{\left(1-\frac{\lg(\lg\log_{1/p}n+1)}{\lg\log_{1/p}n}\right)\cdot\lg\log_{1/p}n}.
\end{aligned}
$$

□

7 The "structure" of neighbourhood

In previous chapters we have demonstrated that computing probabilities in Model B is the same as in Model A (if we set $p_{m,n} = p_n = p \in (0,1)$, then we get asymptotically the same results). Therefore, to abbreviate the notation, we will use in this chapter Model A.

By the meaning of "structure" of neighbourhood we will study in this chapter the following random variable.

Definition 7.1. *Let N_X be a fixed maximal interval with dimension x of a random Boolean function $f \in B_n$. Let $Z_{n,x,k,t}$ denote a random variable on B_n such that $Z_{n,x,k,t}$ is equal to the count of k-dimensional maximal intervals of a function f, which intersect N_X in a t-dimensional interval.*

Notation 7.1. To abbreviate the notation we denote $Z_{n,x,k,t}$ as $Z_{k,t}$ (we will study $Z_{n,x,k,t}$ as a function of parameters k and t).

In next calculates we will need to use the following more generally form of Lemma 6.7.

Lemma 7.1. *Let k, x be integers satisfying (1) and $P_t(N_X, N_K)$ denotes the probability that a random Boolean function contains x-dimensional maximal interval N_X and k-dimensional maximal interval N_K and $N_X \cap N_K$ is a t-dimensional interval. Then*

$$P_t(N_X, N_K) \sim p^{2^x+2^k-2^t} \cdot (1 - p^{2^x} - p^{2^k} + p^{2^x+2^k-2^t})^{n-x-k+t}.$$

Proof. The proof is analogical to proof of Lemma 6.7.

Without detriment to generality we assume

$$\begin{aligned}
N_X &= (\underbrace{\star, \cdots, \star}_{x}, \underbrace{0, \cdots, 0}_{n-x}) \\
N_K &= (\underbrace{0, \cdots, 0}_{x-t}, \underbrace{\star, \cdots, \star}_{k}, \underbrace{0, \cdots, 0}_{n-x-k+t}) \\
N_T &= (\underbrace{0, \cdots, 0}_{x-t}, \underbrace{\star, \cdots, \star}_{t}, \underbrace{0, \cdots, 0}_{n-x}).
\end{aligned}$$

Next we will use the following notation

$$\begin{aligned}
N_{X,i} &= (\underbrace{\star,\cdots,\star}_{x},\underbrace{0,\cdots,0}_{i-x-1},1,\underbrace{0,\cdots,0}_{n-i}) \\
N_{K,i} &= (\underbrace{0,\cdots,0}_{i-1},1,\underbrace{0,\cdots,0}_{x-t-i},\underbrace{\star,\cdots,\star}_{k},\underbrace{0,\cdots,0}_{n-x-k+t}), \text{or} \\
N_{K,i} &= (\underbrace{0,\cdots,0}_{x-t},\underbrace{\star,\cdots,\star}_{k},\underbrace{0,\cdots,0}_{i-x-k+t-1},1,\underbrace{0,\cdots,0}_{n-i}) \\
\tilde{\alpha}_t &= (\underbrace{0,\cdots,0}_{i-1},1,\underbrace{0,\cdots,0}_{n-i}) \\
I &= N_{X,i} \cap N_{K,i}
\end{aligned}$$

where $N_{X,i}$ is defined for $i = \overline{x+1,n}$ and $N_{K,i}$ is defined for $i = \overline{1,x-t}$ or $s = \overline{x+k-t+1,n}$ and $\tilde{\alpha}_i$ is defined for $i = \overline{1,n}$. Obviously

$$P_t(N_X,N_K) = P[N_X \cup N_K \subset N_f, (\forall i)N_{X,i} \nsubseteq N_f, (\forall i)N_{K,i} \nsubseteq N_f].$$

Proof that

$$P[(\forall i = \overline{x+1,x+k-t})N_{X,i}-\{\tilde{\alpha}_i\} \nsubseteq N_f, (\forall i = \overline{1,x-t})N_{K,i}-\{\tilde{\alpha}_i\} \nsubseteq N_f] \sim 1$$

can be done entirely analogically as in proof of Lemma 6.7, so it is omitted. Thus,

$$\begin{aligned}
P_t(N_X,N_K) &\sim P[N_X \cup N_K \subset N_f] \cdot P[(\forall i = \overline{x+k-t+1,n})N_{X,i} \nsubseteq N_f, \\
&\quad (\forall i = \overline{x+k-t+1,n})N_{K,i} \nsubseteq N_f] \\
&= p^{2^x+2^k-2^t} \prod_{i=x+k-t+1}^{n} P[N_{X,i} \nsubseteq N_f, N_{K,i} \nsubseteq N_f] \\
&= p^{2^x+2^k-2^t} \prod_{i=x+k-t+1}^{n} \Big(P[I \subseteq N_f] \cdot P[N_{X,i} \nsubseteq N_f, N_{K,i} \nsubseteq N_f | I \subseteq N_f] \\
&\quad + P[I \nsubseteq N_f] \cdot P[N_{X,i} \nsubseteq N_f, N_{K,i} \nsubseteq N_f | I \nsubseteq N_f]\Big) \\
&= p^{2^x+2^k-2^t} \prod_{i=x+k-t+1}^{n} \Big(P[I \subseteq N_f] \cdot P[N_{X,i} - I \nsubseteq N_f] \cdot \\
&\quad \cdot P[N_{K,i} - I \nsubseteq N_f] + P[I \nsubseteq N_f] \cdot 1\Big) \\
&= p^{2^x+2^k-2^t} \prod_{i=x+k-t+1}^{n} \Big(p^{2^t} \cdot (1-p^{2^x-2^t}) \cdot (1-p^{2^k-2^t}) + 1 - p^{2^t}\Big) \\
&= p^{2^x+2^k-2^t}(1-p^{2^x}-p^{2^k}+p^{2^x+2^k-2^t})^{n-x-k+t}.
\end{aligned}$$

□

Now we are ready to estimate the expectation of $Z_{k,t}$.

Lemma 7.2. *Let k, x be integers satisfying (1).*

$$E(Z_{k,t}) \sim \begin{cases} \phi(n) \to 0 \text{ as } n \to \infty & \text{if } k < \lg\log_{1/p} n \\ p^{2^k-2^t} \cdot \binom{x}{t} \cdot n^{k-t} \cdot e^{-1} & \text{if } k = \lg\log_{1/p} n \\ p^{2^k-2^t} \cdot \binom{x}{t} \cdot n^{k-t} & \text{if } k > \lg\log_{1/p} n \end{cases}$$

Proof. Some routine calculation steps used before are omitted. Here are only the main points of the calculation.

$$\begin{aligned} E(Z_{k,t}) &= \sum_{N_K, N_T} P[N_X \text{ and } N_K \text{ are maximal intervals of } f, \\ &\quad N_X \cap N_K = N_T | N_X \text{ is maximal interval of } f] \\ &= \frac{\binom{x}{t} \cdot \binom{n-x}{k-t} \cdot p^{2^x+2^k-2^t}(1-p^{2^x}-p^{2^k}+p^{2^x+2^k-2^t})^{n-x-k+t}}{p^{2^x} \cdot (1-p^{2^x})^{n-x}} \\ &\sim p^{2^k-2^t} \cdot \binom{x}{t} \cdot n^{k-t} \cdot e^{-n\cdot(p^{2^k}-p^{2^x+2^k-2^t})} \end{aligned}$$

If we consider the value of $e^{-n\cdot(p^{2^k}-p^{2^x+2^k-2^t})}$ with k respect to the $\lg\log_{1/p} n$, then we get desired result. □

Next, we will study $E(Z_{k,t})$ in case that $\lg\log_{1/p} n$ is not an integer, so $k > \lg\log_{1/p} n$. The case $k = \lg\log_{1/p} n$ is very similar, so the reader can easily obtain the same results for it.

Lemma 7.3. *Let k, x be integers satisfying (1) and $\lg\log_{1/p} n$ is not an integer. Let E_k be a such function of k that $E_k = E(Z_{k,t})$. Then E_k is decreasing for $k > \lg\log_{1/p} n$. E_k reach the maximal value for $k = k_m = \lfloor \lg\log_{1/p} n \rfloor + 1$ and moreover*

$$\sum_{k=k_m}^{k_{max}} E_k \sim E_{k_m},$$

where $k_{max} = \lfloor \lg\log_{1/p} n + \lg\lg\log_{1/p} n + \varepsilon \rfloor$.

Proof. Let's compute the ratio E_{k+1}/E_k for $k = \lg\log_{1/p} n + a$, where $a > 0$

$$\begin{aligned} \frac{E_{k+1}}{E_k} &= \frac{p^{2^{k+1}-2^t} \cdot \binom{x}{t} \cdot n^{k+1-t}}{p^{2^k-2^t} \cdot \binom{x}{t} \cdot n^{k-t}} \\ &= n \cdot p^{2^k} \\ &= n \cdot p^{2^{\lg\log_{1/p} n+a}} \\ &= n \cdot \left(\frac{1}{n}\right)^{2^a} \\ &\to 0 \text{ as } n \to \infty. \end{aligned}$$

Moreover $\sum_{k=k_m}^{k_{max}} E_k \geq E_{k_m}$ and

$$\begin{aligned}
\frac{\sum_{k=k_m}^{k_{max}} E_k}{E_{k_m}} &= 1 + \frac{\sum_{k=k_m+1}^{k_{max}} E_k}{E_{k_m}} \\
&\leq 1 + \frac{(k_{max} - k_m) \cdot E_{k_m+1}}{E_{k_m}} \\
&= 1 + (k_{max} - k_m) \cdot p^{2^{k_m}} \cdot n \\
&\leq 1 + (\lg \lg \log_{1/p} n + 1) \cdot p^{2^{k_m}} \cdot n \\
&\sim 1.
\end{aligned}$$

□

Lemma 7.4. *Let k, x be integers satisfying (1) and $\lg \log_{1/p} n$ is not an integer. Let E_t be a such function of t that $E_t = E(Z_{k,t})$. Then E_t is decreasing for $t < \lg \log_{1/p} n$ and increasing for $t > \lg \log_{1/p} n$. Moreover*

$$\sum_{t=0}^{t_{max}} E_t \sim E_0,$$

where $t_{max} = \lfloor \lg \log_{1/p} n \rfloor$.

Proof. Let's compute the ratio E_{t+1}/E_t

$$\begin{aligned}
\frac{E_{t+1}}{E_t} &= \frac{p^{2^k - 2^{t+1}} \cdot \binom{x}{t+1} \cdot n^{k-t-1}}{p^{2^k - 2^t} \cdot \binom{x}{t} \cdot n^{k-t}} \\
&= p^{-2^t} \cdot n^{-1} \cdot \frac{x - t}{t + 1}.
\end{aligned}$$

Obviously for $t < \lg \log_{1/p} n$ the ratio E_{t+1}/E_t tends to 0 as $n \to \infty$ and for $t > \lg \log_{1/p} n$ the ratio E_{t+1}/E_t tends to ∞ as $n \to \infty$. Recall that $\lg \log_{1/p} n$ is not an integer, so $t \neq \lg \log_{1/p} n$. Moreover $\sum_{t=0}^{t_{max}} E_t \geq E_0$ and

$$\begin{aligned}
\frac{\sum_{t=0}^{t_{max}} E_t}{E_0} &= 1 + \frac{\sum_{t=1}^{t_{max}} E_t}{E_0} \\
&\leq 1 + \frac{t_{max} \cdot E_1}{E_0} \\
&= 1 + t_{max} \cdot p^{-1} \cdot x \cdot n^{-1} \\
&\leq 1 + \lg \log_{1/p} n \cdot p^{-1} \cdot x \cdot n^{-1} \\
&\sim 1.
\end{aligned}$$

□

Definition 7.2. *Let N_X be a fixed maximal interval with dimension x of a random Boolean function $f \in B_n$. Let $Z_{n,x}$ denote a random variable on B_n such that $Z_{n,x}$ is equal to the count of all maximal intervals of a function f, which have nonempty intersection with N_X.*

Notation 7.2. To abbreviate the notation we denote $Z_{n,x}$ as Z.

Remark 7.1.

$$Z = \sum_{t=0}^{x-1} \sum_{k=x+1}^{k_{max}} Z_{k,t}$$

Using Lemmas 7.3 and 7.4 we obtain following corollary.

Corollary 7.5. *Let k, x be integers satisfying (1) and $\lg \log_{1/p} n$ is not an integer, then*

$$E(Z) \sim \frac{1}{p} \cdot n^{\lg \log_{1/p} n + c},$$

where $-1 < c < -0.9$.

Proof. Frist recall that for all values of t and k it holds that $t < x$ and $t < k$. Recall that from Lemma 7.3 we have $k = k_m = \lfloor \lg \log_{1/p} n \rfloor + 1$. Thus $t \leq k_m - 1 = \lfloor \lg \log_{1/p} n \rfloor = t_{max}$. So we have for $k_m = \lg \log_{1/p} n + a$, where $0 < a < 1$,

$$\begin{aligned} E(Z) &= \sum_{t=0}^{x-1} \sum_{k=t+1}^{k_{max}} E(Z_{k,t}) \\ &\sim E(Z_{k_m,0}) \\ &\sim p^{2^{\lg \log_{1/p} n+a} - 1} \cdot \binom{x}{0} \cdot n^{\lg \log_{1/p} n + a} \\ &= \frac{1}{p} \cdot n^{\lg \log_{1/p} n + a - 2^a} \\ &= \frac{1}{p} \cdot n^{\lg \log_{1/p} n + c}, \end{aligned}$$

where $-1 < c < -0.9$. □

We see that from the point of view of expectations the neighbourhood of given maximal interval N_X has following "structure". Almost all maximal intervals N_K from neighbourhood of N_X have the dimension equal to $\lfloor \lg \log_{1/p} n \rfloor + 1$ and almost all N_K intersects with N_X only in one vertex. Moreover from the Corollary 7.5 and Markov's inequality we can obtain following upper bound on Z. Recall that Z is the random variable for $|\Theta(N_K)|$ and following upper bound is little bit better than was proved in Theorem 6.10.

Corollary 7.6. *If* $\lg \log_{1/p} n$ *is not an integer, then for almost all random Boolean function it holds that*

$$Z < \phi(n) \cdot n^{\lg \log_{1/p} n + c},$$

where $-1 < c < -0.9$ *and* $\phi(n)$ *is an arbitrary function of* n *which satisfies* $\lim_{n \to \infty} \phi(n) = \infty$.

Proof. It is direct consequence of Corollary 7.5 and Markov's inequality, so the proof is omitted. □

Next we should estimate $Var(Z)$. The best should be to show that $Var(Z_{k,t}) = o(E^2(Z_{k,t}))$, because then we get from Chebyshev's that $Z_{k,t} \sim E(Z_{k,t})$ (what should be very useful result). Moreover we do not have to estimate $Z_{k,t}$ for $k < \lg \log_{1/p} n$ because of following remark.

Remark 7.2. Using the Lemma 7.2 and Markov's inequality, reader can easily obtain that for almost all Boolean functions it holds that $Z_{k,t} = 0$ if $k < \lg \log_{1/p} n$.

Thus, next goal is to estimate $Var(Z_{k,t})$ for $k > \lg \log_{1/p} n$ (and also consider the special case $k = \lg \log_{1/p} n$). However there is no space in this thesis for this goal, so it is left for the next works on this field.

8 Conclusion

We have computed the upper bound on maximal interval in Model B. This result we used to show the equality of Model A and B from the point of view of Škoviera's work in [1]. So we proved for Model B the same as is proved for Model A in [1], the bounds on dimension of a maximal interval of a random Boolean function:

$$\lg \log_{1/p} n - 1 \leq k \leq \lg \log_{1/p} n + \lg \lg \log_{1/p} n + \varepsilon,$$

where $\varepsilon \to 0$ as $n \to \infty$.

We used the above result to find upper and lower bound on the number of maximal intervals intersecting a given maximal interval of a random Boolean function. We proved that the asymptotic bound of the number is $n^{(1+\phi(n)) \log_2 \log_{1/p} n}$, where $\phi(n) \to 0$ as $n \to \infty$.

This result can be used to estimate complexity and the "quality" of outputs of the local algorithms for minimizing d.n.f. (as is shown for example in [2]), which use only neighbourhood of the first order.

Moreover we started to looking for the way how to obtain bounds on the size of the neighbourhood of the second (or higher) order of a given maximal interval of a random Boolean function. Therefore we started to study the "structure" of neighbourhood. Hopefully, that obtained results will be usefull for next works on this field.

References

[1] M. Škoviera, *On the minimization of random Boolean function I.*, Computers and Artificial Intelligence, Vol. 5, No. 4, pp. 321-334, 1986

[2] E. Toman, D. Olejár, M. Stanek, *Average degree in the interval graph of a random Boolean function*, Computing and Informatics, Vol 23., 2006, 1-??, V 2007-Feb-7

[3] J. I. Zhuravlev, *Set theoretical methods in the algebra of logic*, Problemy kibernetiki, Vol. 8, Nauka, Moscow, pp. 5-44, 1962, (in Russian)

[4] A. A. Sapozhenko, *Disjunctive Normal Forms*, Moscow University Press, Moscow, 1975 (in Russian)

[5] E. Toman, *On the size of a neighbourhood of the first rank*, Computers and Artificial Intelligence, Vol. 12, 1993, No. 2, 123-130

[6] A. A. Sapozhenko, *Geometric structure of almost all Boolean functions*, Problemy kibernetiki, Vol. 30, Nauka, Moscow, pp. 227-261, 1975 (in Russian)

[7] E. Toman, J. Tomanová, *Some estimates of the complexity of disjunctive normal forms of a random Boolean function*, Computers and Artificial Intelligence, Vol. 10, No. 4, pp. 327-340, 1991

[8] Edgar M. Palmer, *Graphical evolution*, A Wiley-interscience publication, John Willey and Sons, 1985

[9] A. V. Kostochka, A. A. Sapozhenko and K. Weber, *On random cubical graphs*, In "Fourth Czechoslovakian Symposium on Combinatorics, Graphs and Complexity", J. Nešetřil and M. Fiedler, Eds. Elsevier Science Publishiers, 1992

[10] W. Feller, *An Introduction to Probability Theory and its Application*, J. Willey and Sons, Vol. 1, 3rd Edition, New York, 1970

[11] S. V. Yablonskij, O. B. Lupanov, *Discrete Mathematics and Mathematical Problems of Cybernetics*, Nauka, Moscow, 1974 (in Russian)

Abstrakt

V tejto diplomovej práci uvažujeme o náhodnej Booleanovskej funkcii n premenných, ktorá je splená práve pre m vstupov t.j. $|\{\tilde{\alpha}; f(\tilde{\alpha}) = 1\}| = m$, pričom realizácia každej takejto náhodnej Booleanovskej funkcie je rovnako pravdepodobná. Študujeme jej geometrický model, takzvaný intervalový graf. Pojem intervalového grafu náhodnej Booleanovskej funkcie zadefinoval Sapozhenko a bol využitý v konštrukciách schém realizujúcich Booleanovské funkcie. Použitím tohto modelu odhadneme počet maximálnych intervalov, ktoré majú nenulový prienik s daným maximálnym intervalom náhodnej Booleanovskej funkcie, a dokážeme, že asymptotický odhad tohto čísla je $n^{(1+\phi(n)) \log_2 \log_{1/p} n}$, kde $p = m/2^n$ a $\phi(n) \to 0$ pre $n \to \infty$.

Popri tom sa zaoberáme aj ekvivalenciou tohto pravdepodobnostného modelu náhodnej Booleanovskej funkcie s iným, už predtým študovaným, modelom, kde $\Pr[f(\tilde{\alpha}) = 1] = p$, pre všetky $\tilde{\alpha} \in \{0, 1\}^n$. Nájdeme podmienky, ktoré musí spĺňať m na to, aby tieto dva modely boli ekvivalentné, čo znamená, že $m/2^n$ v jednom modeli môže byť považované za ekvivalent pre p v druhom modeli.

Nakoniec sa snažíme detailnejšie analyzovať "štruktúru" okolia prvého rádu daného maximálneho intervalu náhodnej Booleanovskej funkcie. Naznačíme, ako by sa dali dosiahnuté výsledky využiť v ďalších prácach zaoberajúcich sa odhadom veľkosti okolia druhého resp. n-tého rádu daného maximálneho intervalu náhodnej Booleanovskej funkcie.

Kľúčové slová. náhodná Booleanovská funkcia, intervalový graf