Acta Math. Univ. Comenianae
Vol. LXXIX, 2(2010), pp. 151–164

VERTEX DEGREE IN THE INTERVAL GRAPH OF A RANDOM BOOLEAN FUNCTION

J. DAUBNER and E. TOMAN

Abstract. In the present paper we obtain an asymptotic estimate of vertex degree in the interval graph of a random Boolean function. This substantially improves the known upper and lower bounds of this parameter. Till now only lower and upper bounds of this parameter were known.

1. Introduction

The concept of the *interval graph* of a Boolean function was introduced by Sapozhenko in [**5**]. He obtained results about the size and the number of connected components, and estimated the radius and diameter of this graph. These results are directly related to so called local algorithms for minimization of disjunctive normal forms of Boolean functions, described by Zhuravlev in [**12**]. Toman [**9**] employed the method of good and bad vertices of a Boolean function to estimate the vertex degree of the interval graph. This method has been applied by Toman, Olejár, and Stanek in [**8**] where they have obtained an upper and a lower bound for the average vertex degree in the interval graph of a random Boolean function.

In the present paper we analyse the probability of each edge in the interval graph of a random Boolean function. We determine which edges have the largest probability and which have negligible probability. By using this method we asymptotically estimate the vertex degree in the interval graph of a random Boolean function. As a corollary we obtain the following simplified estimation of the vertex degree

$$n^{\lg \log_{1/p} n + c_n + o(1)},$$

where $c_n = a_n - 2^{a_n}$, where $a_n = \lceil \lg \log_{1/p} n \rceil - \lg \log_{1/p} n$. Notice that $a_n \in \langle 0, 1)$ and therefore $-1 \leq c_n < -0.9$.

2. Preliminaries and Notation

We use the standard notation of the Boolean function theory. An n-ary Boolean function is a function $f\colon \{0,1\}^n \to \{0,1\}$. The symbol Bool_n denotes the set of all

Received August 5, 2008; revised November 4, 2009.
2001 *Mathematics Subject Classification.* Primary 05C80, 60C05, 06E30, 68R05.
Key words and phrases. random Boolean function; interval graph.
Supported by VEGA grant No 1/0406/09.

n-ary Boolean functions. Boolean variables and their negations are called *literals*. A literal of a variable x is denoted by x^α, where $\alpha \in \{0,1\}$, and we set

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

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

A formula $D = K_1 \vee \dots \vee K_m$, the disjunction of distinct elementary conjunctions, is called a *disjunctive normal form* (briefly *d.n.f.*). The parameter m (the number of elementary conjunctions in D) is called the *length* of D. A 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; a d.n.f. with a minimal number of literals in this class is called a *minimal d.n.f. of f* and one with 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 is a graph B^n with 2^n vertices $\tilde{\alpha} = (\alpha_1, \dots, \alpha_n); \alpha_i \in \{0,1\}$, in which the edges join 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 $\tilde{\alpha}$. Notice that 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* and is denoted by $G(f)$.

The set of vertices $N_K \subseteq \{0,1\}^n$ corresponding to an elementary conjunction K of rank r is called an *interval of rank r*. Notice that to every elementary conjunction $K = x_{i_1}^{\alpha_{i_1}} \dots x_{i_r}^{\alpha_{i_r}}$ there corresponds an interval of rank 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 values of other vertex coordinates can be chosen arbitrarily. In the present paper we will often work with intervals corresponding to elementary conjunctions. To abbreviate notation we will use the following *$\star$-notation*.

Notation 2.1. Let $K = x_{i_1}^{\alpha_{i_1}} \dots x_{i_r}^{\alpha_{i_r}}$ be an elementary conjunction of rank r and let N_K be an interval of rank r corresponding to K. Then we denote

$$N_K = \{(\beta_1, \beta_2, \dots, \beta_n) | (\forall i) \beta_i \in \{0,1\} \text{ and } \beta_{i_1} = \alpha_{i_1}, \beta_{i_2} = \alpha_{i_2}, \dots, \beta_{i_r} = \alpha_{i_r}\}$$

briefly as

$$N_K = (\underbrace{\underbrace{\underbrace{\star, \dots, \star}_{i_1-1}, \alpha_{i_1}, \star, \dots, \star}_{i_2-1}, \alpha_{i_2}, \star, \dots, \star}_{i_r-1}, \alpha_{i_r}, \star, \dots, \star).$$

In the geometric model, every interval of rank r represents an $(n-r)$-dimensional subcube of B^n. So we call an interval of rank r also an $(n-r)$-dimensional interval.

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 conjunction K from the d.n.f., D the *neighbourhood of the first order* of K (with respect to the d.n.f. D) is defined as the set of all elementary conjunctions 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 mainly study the neighbourhood of the first order in this paper, the term *neighbourhood* in the present paper means the neighbourhood of the first order. The *interval graph* $\Gamma(f)$ is a graph associated with a Boolean function f as follows: 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 vertex degree in $\Gamma(f)$ and give an asymptotic estimation of this parameter. Note that the degree of the vertex corresponding to a maximal interval N_K is equal to the number of elements in the neighbourhood of N_K.

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

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

In other words, every d.n.f. of a Boolean function f corresponds 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 a certain d.n.f. of f d.n.f. of f. Using the geometric interpretation of d.n.f.s, we can express the "irreducibility" of a d.n.f.: a 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 the d.n.f. is $r = \sum_{j=1}^{m} r_j$ and the construction of a 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 a covering corresponding to a shortest d.n.f. requires to minimize the number of intervals in covering of N_f.

Various parameters of "typical" Boolean functions have been studied in the context of minimization of Boolean functions in the class of d.n.f.s. [**5, 6, 7, 10**]. We use a more general model of Boolean functions, the concept of a *random Boolean function*. A random Boolean function is defined on the vertices of the Boolean n-cube in the following way

$$f(\alpha_1, \alpha_2, \ldots, \alpha_n) = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1-p, \end{cases}$$

where the value $f(\tilde{\alpha})$ does not depend on the values which the Boolean function f takess on other vertices. Recall that $G(f)$ is the graph of f. The probability that the graph $G(f)$ of a random Boolean function f coincides with a subgraph G of the Boolean n-cube is

$$\Pr[G(f) = G] = p^m \cdot (1-p)^{2^n - m},$$

where m denotes the number of vertices in G. Škoviera in [**6**] studied this probabilistic model. A result from his work appears later in our paper as Theorem 3.3.

Let A be a property that a Boolean function may or may not have. If

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

we say that a random Boolean function f has the property A almost surely; equivalently we say that almost all random Boolean functions have property A.

For a random variable Z let the symbols $E(Z)$ and $\mathrm{Var}(Z) = E(Z - E(Z))^2$ denote the *expectation* and the *variance* of Z, respectively. In the present paper we only use nonnegative random variables.

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

$$\Pr(Z \geq \varepsilon) \leq \frac{E(Z)}{\varepsilon}.$$

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

$$\Pr(|Z - E(Z)| \geq \varepsilon) \leq \frac{\mathrm{Var}(Z)}{\varepsilon^2}.$$

Notation 2.2. For functions $f, g\colon \mathbb{R} \to \mathbb{R}$ we use the following asymptotic notations:

- $f \sim g$ means that $\lim\limits_{x\to\infty} \dfrac{f(x)}{g(x)} = 1$
- $f \lesssim g$ means that $\lim\limits_{x\to\infty} \dfrac{f(x)}{g(x)} \leq 1$
- $f \gtrsim g$ means that $\lim\limits_{x\to\infty} \dfrac{f(x)}{g(x)} \geq 1$
- $f = o(g)$ means that $\lim\limits_{x\to\infty} \dfrac{f(x)}{g(x)} = 0$.

Note that $f \sim g$ means that $f = (1 + o(1))g$ and that all asymptotic notations in this paper are used with respect to the dimension n of a random n-ary Boolean function in question.

To estimate one falling factorial we will often use the following lemma.

Lemma 2.3. *Let f and g be functions of n. If $f = o(\sqrt{g})$, then*

$$g^{\underline{f}} \equiv g \cdot (g - 1) \dots (g - f + 1) \sim g^f.$$

Proof. See, for example, [**3**]. □

The rest of this paper has the following structure. First, we asymptotically estimate the probability that a random Boolean function contains a fixed x-dimensional maximal interval and a fixed k-dimensional maximal interval whose intersection is a t-dimensional interval. This is done in Lemma 3.5. Using this result, we asymptotically estimate the expectation value of the random variable

$Z^{n,x}_{k,t}$ (see Definition 3.1). This is done in Lemma 3.6. Then we analyse the expectation value $E(Z^{n,x}_{k,t})$ as a function of k and t to show that only one special value ($E(Z^{n,x}_{m_n,0})$) is asymptotically significant. This is done in Corollaries 3.8, 3.9, and 3.10. As a direct consequence of these corollaries and Markov's inequality we obtain in Lemma 3.11 that $Z^{n,x}_{k,t} = Z^{n,x}_{m_n,0} + o(E(Z^{n,x}_{m_n,0}))$. Then we asymptotically estimate the variance of the random variable $Z^{n,x}_{m_n,0}$. This is done in Lemma 3.12. Using this lemma and Chebyshev's inequality, we show that the random variable $Z^{n,x}_{m_n,0}$ is asymptotically equal to its expectation. This is done in Corollary 3.13. Finally, in Theorem 3.14, we use Lemma 3.11 and Corollary 3.13 to show that the random variable $Z_{n,x}$ (see Definition 3.2) is asymptotically equal to $E(Z^{n,x}_{m_n,0})$.

3. Size and structure of the neighbourhood of a maximal interval

We describe the size and the structure of a neighbourhood by the following random variables.

Definition 3.1. *Let N_X be a fixed x-dimensional maximal interval of a random Boolean function $f \in \mathrm{Bool}_n$. Let $Z^{n,x}_{k,t}$ denote the random variable on* Bool_n *such that $Z^{n,x}_{k,t}$ is equal to the number of k-dimensional maximal intervals of f which intersect N_X in a t-dimensional interval.*

Definition 3.2. *Let N_X be a fixed x-dimensional maximal interval of a random Boolean function $f \in \mathrm{Bool}_n$. Let $Z_{n,x}$ denote the random variable on* Bool_n *such that $Z_{n,x}$ is equal to the number of all maximal intervals of f which have a nonempty intersection with N_X.*

Notice that the random variable $Z_{n,x}$ is equal to the degree of the vertex ν_x in the interval graph, where ν_x corresponds to a fixed maximal interval N_X. Notice that the random variable $Z^{n,x}_{k,t}$ is equal to the number of edges (ν_x, ν_k) in the interval graph, where ν_x corresponds to a fixed maximal interval N_X and ν_k corresponds to any k-dimensional maximal interval N_K such that $N_X \cap N_K = N_T$ and N_T is a t-dimensional interval.

Theorem 3.3. *Let $p \in (0,1)$. Then with probability tending to 1 as $n \to \infty$, the dimension k of a maximal interval of a random Boolean function satisfies the following inequalities*

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

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

Proof. See [**6**]. □

Notation 3.1.

$$k_{\min} = \lceil \lg \log_{1/p} n - 1 \rceil$$
$$k_{\max} = \lfloor \lg \log_{1/p} n + \lg \lg \log_{1/p} n + \varepsilon \rfloor$$

A consequence of (1) is that for almost all random Boolean functions, the dimension k of a maximal interval satisfies following inequality

$$k < 2\lg\log_{1/p} n. \tag{2}$$

Corollary 3.4. *For almost all random Boolean functions we have*

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

Proof. This is a direct consequence of Theorem 3.3 and definitions 3.1 and 3.2. □

Next, we asymptotically estimate the probability that a random Boolean function contains a fixed x-dimensional maximal interval and a fixed k-dimensional maximal interval whose intersection is a t-dimensional interval.

Lemma 3.5. *Let k and x be integers satisfying* (1) *and let $P_t(N_X, N_K)$ denote the probability that a random Boolean function contains an x-dimensional maximal interval N_X and a k-dimensional maximal interval N_K such that $N_X \cap N_K$ 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. Without loss of generality we assume that

$$N_X = (\underbrace{\star, \dots, \star}_{x}, \underbrace{0, \dots, 0}_{n-x})$$
$$N_K = (\underbrace{0, \dots, 0}_{x-t}, \underbrace{\star, \dots, \star}_{k}, \underbrace{0, \dots, 0}_{n-x-k+t})$$
$$N_T = (\underbrace{0, \dots, 0}_{x-t}, \underbrace{\star, \dots, \star}_{t}, \underbrace{0, \dots, 0}_{n-x}).$$

Next, we use the following notation

$$N_{X,i} = (\underbrace{\star, \dots, \star}_{x}, \underbrace{0, \dots, 0}_{i-x-1}, 1, \underbrace{0, \dots, 0}_{n-i}), \qquad \text{for } i = x+1, \dots, n$$

$$N_{K,i} = \begin{cases} (\underbrace{0, \dots, 0}_{i-1}, 1, \underbrace{0, \dots, 0}_{x-t-i}, \underbrace{\star, \dots, \star}_{k}, \underbrace{0, \dots, 0}_{n-x-k+t}), & \text{for } i = 1, \dots, x-t \\ (\underbrace{0, \dots, 0}_{x-t}, \underbrace{\star, \dots, \star}_{k}, \underbrace{0, \dots, 0}_{i-x-k+t-1}, 1, \underbrace{0, \dots, 0}_{n-i}), & \text{for } i = x+k-t+1, \dots, n \end{cases}$$

$$\tilde{\alpha}_i = (\underbrace{0, \dots, 0}_{i-1}, 1, \underbrace{0, \dots, 0}_{n-i}), \text{ for } i = 1, \dots, n$$

$$I_i = N_{X,i} \cap N_{K,i}, \qquad \text{for } i = x+k-t+1, \dots, n.$$

From the definition of a maximal interval we obtain

$$P_t(N_X, N_K) = \Pr[N_X \cup N_K \subset N_f, (\forall i) N_{X,i} \not\subseteq N_f, (\forall i) N_{K,i} \not\subseteq N_f]. \tag{3}$$

First, we look at the case when $i \leq x+k-t$. Let P_1 denote the probability that N_f does not contain any $N_{X,i}$ for all $i = x+1, \dots, x+k-t$ and it does not contain any $N_{K,i}$ for all $i = 1, \dots, x-t$. Formally,

$$\begin{aligned} P_1 = \ & \Pr[\text{for each } i = x+1, \dots, x+k-t\text{: } N_{X,i} - \{\tilde{\alpha}_i\} \nsubseteq N_f \\ & \quad \text{and for each } i = 1, \dots, x-t\text{: } N_{K,i} - \{\tilde{\alpha}_i\} \nsubseteq N_f]. \end{aligned}$$

Using inequalities (1) and (2) for dimensions k and x, we obtain

$$\begin{aligned} P_1 &\geq \prod_{i=x+1}^{x+k-t} \Pr[N_{X,i} - \{\tilde{\alpha}_i\} \nsubseteq N_f] \cdot \prod_{i=1}^{x-t} \Pr[N_{K,i} - \{\tilde{\alpha}_i\} \nsubseteq N_f] \\ &= (1 - p^{2^x - 1})^{k-t} \cdot (1 - p^{2^k - 1})^{x-t} \\ &\gtrsim (1 - p^{2^{\lg \log_{1/p} n - 1} - 1})^{2 \lg \log_{1/p} n} \cdot (1 - p^{2^{\lg \log_{1/p} n - 1} - 1})^{2 \lg \log_{1/p} n} \\ &= (1 - p^{\frac{1}{2} \log_{1/p} n - 1})^{4 \lg \log_{1/p} n} = \left(1 - \frac{1}{p\sqrt{n}}\right)^{4 \lg \log_{1/p} n} \sim 1, \end{aligned}$$

and hence

$$P_1 \sim 1.$$

So, if we calculate $P_t(N_X, N_K)$ according to (3), then we can omit all cases where $i \leq x+k-t$. If $i > x+k-t$, then some of the events are independent, therefore,

$$\begin{aligned} & P_t(N_X, N_K) \\ \sim\ & \Pr[N_X \cup N_K \subset N_f] \cdot \Pr[(\forall i > x+k-t) N_{X,i} \nsubseteq N_f \text{ and } N_{K,i} \nsubseteq N_f] \\ =\ & p^{2^x + 2^k - 2^t} \prod_{i=x+k-t+1}^{n} \Pr[N_{X,i} \nsubseteq N_f \text{ and } N_{K,i} \nsubseteq N_f] \\ =\ & p^{2^x + 2^k - 2^t} \prod_{i=x+k-t+1}^{n} \Big(1 - \Pr[N_{X,i} \subseteq N_f \text{ or } N_{K,i} \subseteq N_f]\Big) \\ =\ & p^{2^x + 2^k - 2^t} \prod_{i=x+k-t+1}^{n} \Big(1 - \Pr[N_{X,i} \subseteq N_f] - \Pr[N_{K,i} \subseteq N_f] + \Pr[I_i \subseteq N_f]\Big) \\ =\ & p^{2^x + 2^k - 2^t} (1 - p^{2^x} - p^{2^k} + p^{2^x + 2^k - 2^t})^{n-x-k+t}. \end{aligned}$$

□

Next, we evaluate the expectation value of $Z_{k,t}^{n,x}$.

Lemma 3.6. *For almost all random Boolean functions we have*

- *if* $k < \lg \log_{1/p} n$, *then*

$$E(Z_{k,t}^{n,x}) \lesssim c^{n^b}$$

where c, b *are constants satisfying* $c < 1$ *and* $b > 0$,

- *if* $k = \lg \log_{1/p} n$*, then*

$$E(Z_{k,t}^{n,x}) \sim n^{k-t} \cdot p^{2^k - 2^t} \cdot 2^{x-t} \binom{x}{t} \frac{1}{(k-t)!} \cdot \mathrm{e}^{-1},$$

- *if* $k > \lg \log_{1/p} n$*, then*

$$E(Z_{k,t}^{n,x}) \sim n^{k-t} \cdot p^{2^k - 2^t} \cdot 2^{x-t} \binom{x}{t} \frac{1}{(k-t)!}.$$

Proof. First, let us recall that $0 \le t < x$ and $t < k$ and that for almost all random Boolean functions, k and x satisfy (1) and (2).

Let f be an n-ary random Boolean function. Let N_X be a maximal interval of N_f. For every k-dimensional interval N_K of the n-cube B^n we introduce the random variable η_K (also called an *indicator*) defined as follows

$$\eta_K(f) = \begin{cases} 1 & \text{if } N_K \text{ is maximal interval of } N_f \\ & \text{and } N_K \cap N_X \text{ is } t\text{-dimensional interval} \\ 0 & \text{otherwise.} \end{cases}$$

Obviously, the random variable $Z_{k,t}^{n,x}$ is the sum of all indicators η_K

$$Z_{k,t}^{n,x} = \sum_{N_K} \eta_K(f),$$

where the summation extends over all k-dimensional intervals of B^n.

Next, for every k-dimensional interval N_K and every t-dimensional interval N_T of B^n, we introduce the random variable $\eta_{K,T}$ defined as follows

$$\eta_{K,T}(f) = \begin{cases} 1 & \text{if } N_K \text{ is maximal interval of } N_f \text{ and } N_K \cap N_X = N_T \\ 0 & \text{otherwise.} \end{cases}$$

Obviously, the indicator η_K is the sum of all indicators $\eta_{K,T}$

$$\eta_K(f) = \sum_{N_T \subset N_X} \eta_{K,T}(f),$$

where the summation extends over all t-dimensional intervals of N_X. Thus

$$Z_{k,t}^{n,x} = \sum_{N_K, N_T} \eta_{K,T}(f),$$

$$E(Z_{k,t}^{n,x}) = \sum_{N_K, N_T} E(\eta_{K,T}).$$

There are $2^{x-t} \cdot \binom{x}{t}$ t-dimensional intervals of N_X and for each such N_T there are $\binom{n-x}{k-t}$ k-dimensional intervals of N_f which intersect N_X in N_T. Thus

$$E(Z_{k,t}^{n,x}) = \binom{n-x}{k-t} 2^{x-t} \binom{x}{t} E(\eta_{K,T}).$$

By applying Lemma 2.3 and inequalities (1) and (2) (for k and x) to $\binom{n-x}{k-t}$, we obtain

$$\binom{n-x}{k-t} \sim \frac{(n-x)^{k-t}}{(k-t)!} \sim \frac{n^{k-t}}{(k-t)!}.$$

Thus

$$E(Z_{k,t}^{n,x}) \sim \frac{n^{k-t}}{(k-t)!} 2^{x-t} \binom{x}{t} E(\eta_{K,T}).$$

Now we use Lemma 3.5 to calculate the expectation value of $\eta_{K,T}$. We also use the conditional probability equation $\Pr[A|B] = \Pr[A \cap B] / \Pr[B]$.

$$\begin{aligned} E(\eta_{K,T}) = {} & \Pr[N_K \text{ is a maximal interval of } N_f \text{ and } N_X \cap N_K = N_T \\ & |N_X \text{ is a maximal interval of } N_f] \\ = {} & \frac{\Pr[N_K \text{ and } N_X \text{ are maximal intervals of } N_f \text{ and } N_X \cap N_K = N_T]}{\Pr[N_X \text{ is a maximal interval of } N_f]} \\ \sim {} & \frac{p^{2^x+2^k-2^t}(1-p^{2^x}-p^{2^k}+p^{2^x+2^k-2^t})^{n-x-k+t}}{p^{2^x}(1-p^{2^x})^{n-x}} \end{aligned}$$

Using inequalities (1) and (2) for the dimensions k and x of maximal intervals, we obtain

$$\begin{aligned} & (1-p^{2^x}-p^{2^k}+p^{2^x+2^k-2^t})^{-x-k+t} \\ \leq {} & (1-p^{2^x})^{-x-k} \lesssim (1-p^{2^{-1+\lg\log_{1/p} n}})^{-4\lg\log_{1/p} n} = (1-\frac{1}{\sqrt{n}})^{-4\lg\log_{1/p} n} \sim 1, \end{aligned}$$

and because $(1-p^{2^x}-p^{2^k}+p^{2^x+2^k-2^t})^{-x-k+t} \geq 1$, we get

$$(1-p^{2^x}-p^{2^k}+p^{2^x+2^k-2^t})^{-x-k+t} \sim 1. \tag{4}$$

Using the same technique, we obtain

$$(1-p^{2^x})^{-x} \sim 1. \tag{5}$$

By applying equations (4) and (5) to $E(\eta_{K,T})$, we obtain

$$\begin{aligned} E(\eta_{K,T}) \sim {} & \frac{p^{2^x+2^k-2^t}(1-p^{2^x}-p^{2^k}+p^{2^x+2^k-2^t})^n}{p^{2^x}(1-p^{2^x})^n} \\ & = p^{2^k-2^t}\left(1-p^{2^k}\cdot\frac{1-p^{2^x-2^t}}{1-p^{2^x}}\right)^n = p^{2^k-2^t}(1-p^{2^k}\cdot(1+o(1)))^n \\ & \sim p^{2^k-2^t}e^{-n\cdot p^{2^k}\cdot(1+o(1))}. \end{aligned}$$

Thus

$$E(Z_{k,t}^{n,x}) \sim \frac{n^{k-t}}{(k-t)!} 2^{x-t} \binom{x}{t} p^{2^k-2^t} e^{-n\cdot p^{2^k}\cdot(1+o(1))}.$$

Finally, if we compare the value of k to $\lg\log_{1/p} n$ in last the expression $-n\cdot p^{2^k}$, we get the desired result. $\square$

Remark. The following expression is a negligible part of $E(Z_{k,t}^{n,x})$ because

$$2^{x-t}\binom{x}{t}\frac{1}{(k-t)!} = n^{o(1)}.$$

Corollary 3.7. *If* $k < \lg \log_{1/p} n$*, then for almost all random Boolean functions we have*

$$Z_{k,t}^{n,x} = 0.$$

Proof. From Lemma 3.6 we obtain

$$\lim_{n\to\infty} E(Z_{k,t}^{n,x}) = 0$$

and by using Markov's inequality we obtain

$$\lim_{n\to\infty} \Pr[Z_{k,t}^{n,x} = 0] = 1 - \lim_{n\to\infty} \Pr\left[Z_{k,t}^{n,x} \geq \frac{1}{2}\right] \geq 1 - \lim_{n\to\infty} \frac{E(Z_{k,t}^{n,x})}{\frac{1}{2}} = 1$$

□

We see that $Z_{k,t}^{n,x}$ adds nothing to $Z_{n,x}$ if $k < \lg \log_{1/p} n$. So, next we will analyse $Z_{k,t}^{n,x}$ only when $k \geq \lg \log_{1/p} n$.

Corollary 3.8. *Suppose that* $t > \lg \log_{1/p} n$ *and* $k \geq \lg \log_{1/p} n$*. Then for almost all random Boolean functions we have* $Z_{k,t}^{n,x} \lesssim n^c$ *where* $c < 0$ *is a constant.*

Proof. Let us write k as $t + y$, where $y \geq 1$. Then

$$E(Z_{k,t}^{n,x}) \lesssim p^{2^t \cdot (2^y - 1)} n^y \cdot n^{o(1)}.$$

By substituting $t = \lg \log_{1/p} n + \tau$, where $\tau > 0$, we obtain

$$E(Z_{k,t}^{n,x}) \lesssim n^{2^\tau \cdot (2^y - 1)} n^y \cdot n^{o(1)} \lesssim n^{-2^\tau}$$

and by using Markov's inequality we get the desired result. □

Next, we analyse $Z_{n,x}$ only in the case that $\lg \log_{1/p} n$ is not an integer, so for $E(Z_{k,t}^{n,x})$, we can suppose that $k > \lg \log_{1/p} n$. The case $k = \lg \log_{1/p} n$ is very similar, so similar results can be obtained in this case with the same technique as in the case $k > \lg \log_{1/p} n$. We will mention these results at the end of this paper.

Notation 3.2. Set

$$m_n = \lceil \lg \log_{1/p} n \rceil = \lg \log_{1/p} n + a_n,$$

where a_n is a number satisfying $0 < a_n < 1$.

If we analyse $E(Z_{k,t}^{n,x})$ from Lemma 3.6 as a function of k or t, then we obtain two following corollaries.

Corollary 3.9. *Suppose that* $\lg \log_{1/p} n$ *is not an integer. Set* $E_k = E(Z_{k,t}^{n,x})$*. Then* E_k *is decreasing for* $k > \lg \log_{1/p} n$*.* E_k *reaches the maximal value for* $k = m_n$*. Moreover, for all integers* k_1*,* k_2 *satisfying the inequalities* $m_n \leq k_1 \leq k_2 \leq k_{\max}$*, we have*

$$\sum_{k=k_1}^{k_2} E_k \sim E_{k_1}.$$

Proof. Let us consider the ratio E_{k+1}/E_k for $k = \lg\log_{1/p} n + a_n$, where $a_n > 0$. We obtain

$$\frac{E_{k+1}}{E_k} = n \cdot \left(\frac{1}{n}\right)^{2^{a_n}} \cdot \frac{1}{k+1-t} = n^{-\tau+o(1)},$$

where $\tau > 0$. Thus

$$\begin{aligned}\sum_{k=k_1}^{k_2} E_k &\leq E_{k_1} + \sum_{k=k_1+1}^{k_{\max}} E_k \\ &\lesssim E_{k_1} + (k_{\max} - k_1 - 1) \cdot n^{-\tau+o(1)} \cdot E_{k_1} \\ &\lesssim E_{k_1} + \lg\lg\log_{1/p} n \cdot n^{-\tau+o(1)} \cdot E_{k_1} \\ &= E_{k_1} + o(E_{k_1}).\end{aligned}$$

By using the trivial fact that $\sum_{k=k_1}^{k_2} E_k \geq E_{k_1}$, we obtain

$$\sum_{k=k_1}^{k_2} E_k \sim E_{k_1}.$$

□

Corollary 3.10. *Suppose that* $\lg\log_{1/p} n$ *is not an integer. Set* $E_t = E(Z_{k,t}^{n,x})$. *Then* E_t *is decreasing for* $t \leq \lg\log_{1/p} n$. *Moreover, for integers* t_1, t_2 *satisfying the inequalities* $0 \leq t_1 \leq t_2 \leq \lg\log_{1/p} n$, *we have*

$$\sum_{t=t_1}^{t_2} E_t \sim E_{t_1}.$$

Proof. We get the desired result (using the same technique as in Corollary 3.9), by considering the ratio E_{t+1}/E_t

$$\frac{E_{t+1}}{E_t} = p^{-2^t} \cdot n^{-1} \cdot 2^{-1} \frac{x-t}{t+1}(k-t) = n^{-\tau+o(1)},$$

where $\tau > 0$. □

Lemma 3.11. *Suppose that* $\lg\log_{1/p} n$ *is not an integer. Then for almost all random Boolean functions, we have*

$$Z_{n,x} = Z_{m_n,0}^{n,x} + o(E(Z_{m_n,0}^{n,x})).$$

Proof. From Corollary 3.4 we obtain

$$\begin{aligned}Z_{n,x} &= \sum_{t=0}^{x-1} \sum_{k=t+1}^{k_{\max}} Z_{k,t}^{n,x} \\ &= Z_{m_n,0}^{n,x} + \sum_{t=0}^{0} \sum_{k=m_n+1}^{k_{\max}} Z_{k,t}^{n,x} + \sum_{t=1}^{m_n-1} \sum_{k=m_n}^{k_{\max}} Z_{k,t}^{n,x} + \sum_{t=m_n}^{x-1} \sum_{k=t+1}^{k_{\max}} Z_{k,t}^{n,x} + o(1)\end{aligned}$$

and by using Markov's inequality and all the previous corollaries of Lemma 3.6, we obtain that all the sums in the previous expression are $o(E(Z_{m_n,0}^{n,x}))$. □

Next, we estimate $\mathrm{Var}(Z_{m_n,0}^{n,x})$ to show that $Z_{n,x} \sim E(Z_{m_n,0}^{n,x})$.

Lemma 3.12. *Suppose that* $\lg\log_{1/p} n$ *is not an integer. Then for almost all random Boolean functions, we have*

$$\mathrm{Var}(Z^{n,x}_{m_n,0}) = o(E^2(Z^{n,x}_{m_n,0})).$$

Proof. First, let us recall that

$$\mathrm{Var}(Z^{n,x}_{m_n,0}) = E\big((Z^{n,x}_{m_n,0})^2\big) - E^2(Z^{n,x}_{m_n,0}).$$

We can estimate $E\big((Z^{n,x}_{m_n,0})^2\big)$ as follows. Let N_{K_1} and N_{K_2} be m_n-dimensional intervals of N_f. Let $P_{\max}(N_{K_1}, N_{K_2})$ denote the conditional probability that N_{K_1} and N_{K_2} are maximal intervals of N_f intersecting N_X in only one vertex under the condition that N_X is a maximal interval of N_f. Let $P(N_{K_1}, N_{K_2})$ denote the conditional probability that N_{K_1} and N_{K_2} are intervals of N_f intersecting N_X in only one vertex under the condition that N_X is a maximal interval of N_f. Obviously,

$$P_{\max}(N_{K_1}, N_{K_2}) \le P(N_{K_1}, N_{K_2}).$$

Thus
$$E\big((Z^{n,x}_{m_n,0})^2\big) = \sum_{N_{K_1},N_{K_2}} P_{\max}(N_{K_1}, N_{K_2}) \le \sum_{N_{K_1},N_{K_2}} P(N_{K_1}, N_{K_2}).$$

Set $\{\tilde{\alpha}_1\} = N_{K_1} \cap N_X$ and $\{\tilde{\alpha}_2\} = N_{K_2} \cap N_X$. Let us consider the following two cases:

1. $\tilde{\alpha}_1 \equiv \tilde{\alpha}_2$. Let $N_{K_1} \cap N_{K_2}$ be denoted by an u-dimensional interval N_U. Then, for a fixed x-dimensional interval N_X, there are $2^x\binom{n-x}{m_n}$ m_n-dimensional intervals N_{K_1} which intersect N_X in just one vertex. For each such N_X and N_{K_1} and a fixed u, there are $\binom{m_n}{u}$ u-dimensional intervals N_U such that $N_U \subseteq N_{K_1}$ and $N_{K_1} \cap N_X \subseteq N_U$. Finally, for such N_X and N_{K_1} and N_U, there are $\binom{n-x-m_n}{m_n-u}$ m_n-dimensional intervals N_{K_2} such that $N_{K_2} \cap N_{K_1} = N_U$ and N_{K_2} intersects N_X in just one vertex. The probability that $(N_{K_1} \cap N_{K_2}) - N_X \subseteq N_f$ is $p^{|N_{K_1}|-1} \cdot p^{|N_{K_2}|-|N_{K_1}\cap N_{K_2}|}$. Thus

$$\begin{aligned}\sum P(N_{K_1}, N_{K_2}) &\lesssim \sum_{u=0}^{m_n} 2^x \binom{n-x}{m_n} p^{2^{m_n}-1} \cdot \binom{m_n}{u}\binom{n-x-m_n}{m_n-u} p^{2^{m_n}-2^u} \\ &= o(E^2(Z^{n,x}_{m_n,0})).\end{aligned}$$

2. $\tilde{\alpha}_1 \not\equiv \tilde{\alpha}_2$. Then $N_{K_1} \cap N_{K_2} = \emptyset$ and we obtain

$$\begin{aligned}\sum P(N_{K_1}, N_{K_2}) &\lesssim 2^x \binom{n-x}{m_n} p^{2^{m_n}-1} \cdot (2^x - 1)\binom{n-x}{m_n} p^{2^{m_n}-1} \\ &\lesssim E^2(Z^{n,x}_{m_n,0}).\end{aligned}$$

By combining these two cases, we obtain

$$E\big((Z^{n,x}_{m_n,0})^2\big) \le E^2(Z^{n,x}_{m_n,0}) + o(E^2(Z^{n,x}_{m_n,0})).$$

Thus
$$\mathrm{Var}(Z^{n,x}_{m_n,0}) = o(E^2(Z^{n,x}_{m_n,0})).$$

□

Since the following corollary is a direct consequence of Chebyshev's inequality and Lemma 3.12, we omit the proof.

Corollary 3.13. *Suppose* $\lg\log_{1/p} n$ *is not an integer. Then for almost all random Boolean functions we have*

$$Z^{n,x}_{m_n,0} \sim E(Z^{n,x}_{m_n,0}).$$

Thus, we can estimate $Z_{n,x}$ as follows.

Theorem 3.14. *Suppose that* $\lg\log_{1/p} n$ *is not an integer. Then for almost all random Boolean functions we have*

$$Z_{n,x} \sim E(Z^{n,x}_{m_n,0}) \sim n^{\lg\log_{1/p} n + c_n} \cdot 2^x \cdot \frac{p^{-1}}{\lceil \lg\log_{1/p} n\rceil !},$$

where $c_n = a_n - 2^{a_n}$.

Proof. As a direct consequence of Lemma 3.11 and Corollary 3.13 we obtain

$$Z_{n,x} \sim E(Z^{n,x}_{m_n,0}).$$

Next, from Lemma 3.6 we obtain

$$\begin{aligned} E(Z^{n,x}_{m_n,0}) &\sim n^{m_n} \cdot p^{2^{m_n}} \cdot 2^x \cdot \frac{p^{-1}}{m_n!} \\ &= n^{\lg\log_{1/p} n + a_n - 2^{a_n}} \cdot 2^x \cdot \frac{p^{-1}}{\lceil \lg\log_{1/p} n\rceil !}. \end{aligned}$$

□

Corollary 3.15. *Suppose* $\lg\log_{1/p} n$ *is not an integer. Then for almost all random Boolean functions we have*

$$Z_{n,x} = n^{\lg\log_{1/p} n + c_n + o(1)},$$

where $c_n = a_n - 2^{a_n}$.

Remark. For $\lg\log_{1/p} n$ an integer we get the following similar results

$$\begin{aligned} Z_{n,x} &\sim E(Z^{n,x}_{\lg\log_{1/p} n,0}) \\ &\sim n^{\lg\log_{1/p} n - 1} \cdot 2^x \cdot \frac{p^{-1}e^{-1}}{(\lg\log_{1/p} n)!}. \end{aligned}$$

We also get $Z_{n,x} = n^{\lg\log_{1/p} n - 1 + o(1)}$.

The above results show that the neighbourhood of a given maximal interval N_X has the following structure. Almost all maximal intervals N_K from the neighbourhood of N_X have dimension $\lceil \lg\log_{1/p} n\rceil$ and almost all N_K intersect with N_X in only one vertex and the number of all such N_K's is

$$n^{\lg\log_{1/p} n + c_n} \cdot 2^x \cdot \frac{p^{-1}}{\lceil \lg\log_{1/p} n\rceil !},$$

where $c_n = a_n - 2^{a_n}$, where $a_n = \lceil \lg\log_{1/p} n\rceil - \lg\log_{1/p} n$.

4. Conclusion

In the present paper we have estimated the size of the neighbourhood of the first order. This result can be used for analysing the complexity (and other properties) of local algorithms, that use the neighbourhood of the first order to find the minimal or shortest d.n.f. of Boolean function.

There also exist local algorithms that use neighbourhoods of the second or higher orders. Such algorithms can be found, for example, in [**12**]. The results from this paper can be also used for analysing these neighbourhoods.

References

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

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

3. Palmer E. M., *Graphical evolution*, A Wiley-interscience publication, John Willey and Sons, 1985.

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

5. ______, *Geometric structure of almost all Boolean functions*, Problemy kibernetiki, **30** (1975), 227–261, (in Russian).

6. M. Škoviera, *On the minimization of random Boolean function I.*, Computers and Artificial Intelligence, **5(4)** (1986), 321–334

7. ______, *On the minimization of random Boolean function II.*, Computers and Artifical Intelligence, **5(6)** (1986), 493–509.

8. Toman E., Olejár D. and Stanek M., *Average degree in the interval graph of a random Boolean function*, Computing and Informatics, **27** (2008), 627–638.

9. Toman E., *On the size of a neighbourhood of the first rank*, Computers and Artificial Intelligence, **12(2)** (1993), 123–130.

10. Toman E. and Tomanová J., *Some estimates of the complexity of disjunctive normal forms of a random Boolean function*, Computers and Artificial Intelligence, 10(4) (1991), 327–340.

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

12. Zhuravlev J. I., *Set theoretical methods in the algebra of logic*, Problemy kibernetiki, **8** (1962), 5–44, (in Russian).

J. Daubner, University of Žilina, Faculty of Management Science and Informatics, Department of Info Communication Networks, Univerzitná 8215/1, 01026 Žilina, Slovakia,
e-mail: `jdaubner@gmail.com`

E. Toman, Comenius University, Faculty of Mathematics, Physics and Informatics, Department of Computer Science, Mlynská dolina, 842 48 Bratislava, Slovak Republic,
e-mail: `toman@dcs.fmph.uniba.sk`