ON THE NUMBER OF BOOLEAN FUNCTIONS IN A GIVEN NEIGHBOURHOOD OF THE AFFINE FUNCTIONS SET

A.M. Zubkov, A.A. Serov
Steklov Mathematical Institute of RAS
Moscow
e-mail: `zubkov@mi.ras.ru`

Let $V_n = (\mathrm{GF}(2))^n$. Denote by $\mathbb{F}_2^{V_n}$ the set of all Boolean functions and by $\mathbb{A}_n$ the set of all affine functions of n Boolean variables. Let $\mathbb{F}_2^{V_n}(r) \subseteq \mathbb{F}_2^{V_n}$ be the set of all Boolean functions with Hamming distance to $\mathbb{A}_n$ not exceeding r. It is known [1] that $\mathbb{F}_2^{V_n}(r) = \mathbb{F}_2^{V_n}$ if $r \geq 2^{n-1} - 2^{n/2-1}$. Let's define $x_n(r)$ by the equation

$$r = 2^{n-1} - \sqrt{2^{n-1}\left(n\ln 2 - \tfrac{1}{2}\,\ln(4\pi n\ln 2) + x_n(r)\right)}.$$

According to a theorem from [2] for the set $\mathbb{F}_2^{V_n,0}(r)$ of all Boolean functions with Hamming distance to the set of linear functions not exceeding r for $n, r \to \infty$, $x_n(r) \to x \in \mathbb{R}$ we have

$$|\,\mathbb{F}_2^{V_n,0}(r)| = 2^{2^n}\left(B(x) + o(1)\right), \quad B(x) \stackrel{\mathrm{def}}{=} 1 - e^{-e^{-x}}, \tag{1}$$

that is for almost all Boolean functions of n variables their distances to the set of linear functions have the form

$$2^{n-1} - \sqrt{2^{n-1}\left(n\ln 2 - \tfrac{1}{2}\ln n\right)} + O\left(\sqrt{2^n/n}\right),\ n \to \infty.$$

We find two-sided estimates for $|\,\mathbb{F}_2^{V_n}(r)|$ which are valid for all nonnegative $r < 2^{n-1} - 2^{n/2-1}$. Similar estimates for $|\,\mathbb{F}_2^{V_n,0}(r)|$ show that for $n, r \to \infty$ the expression $2^{2^n} B(x_n(r))$ from (1) isn't a correct asymptotics for $|\,\mathbb{F}_2^{V_n,0}(r)|$ if $x_n(r) \to \infty$.

Let

$$N_1(n,r) = \sum_{m=0}^{r} C_{2^n}^m, \quad N_2(n,r) = \sum_{m_0=0}^{r-2^{n-2}} C_{2^{n-1}}^{m_0} \sum_{m_1=2^{n-1}-(r-m_0)}^{r-m_0} C_{2^{n-1}}^{m_1},$$

$$N_3(n,r) = \sum_{v=0}^{r-2^{n-2}} \sum_{u=2^{n-1}-r+2v}^{r} C_{2^{n-2}}^{v}\, C_{2^{n-2}}^{u-v}\, S(r-u, r+u-2v-2^{n-1}),$$

where

$$S(a,b) = \sum_{\substack{g,h \geqslant 0:\\ g+h \leqslant a,\ |g-h| \leqslant b}} C_{2^{n-2}}^{g}\, C_{2^{n-2}}^{h}.$$

Theorem 1. a) *If* $0 \leqslant r < 2^{n-1}$ *then* $|\,\mathbb{F}_2^{V_n}(r)| \leqslant 2^{n+1} N_1(n,r)$ *and*

$$2^{n+1}N_1(n,r) - 4\,C_{2^n}^2 N_2(n,r) \leqslant |\,\mathbb{F}_2^{V_n}(r)| \leqslant$$

$$\leqslant 2^{n+1}N_1(n,r) - 4\,C_{2^n}^2 N_2(n,r) + 8\,C_{2^n}^3 N_3(n,r). \tag{2}$$

b) *If* $0 < r < 2^{n-2}$ *then* $|\,\mathbb{F}_2^{V_n}(r)| = 2^{n+1}N_1(n,r)$.
c) *If* $2^{n-2} \leqslant r < 2^{n-2} + 2^{n-4}$ *then* $|\,\mathbb{F}_2^{V_n}(r)|$ *equals to the right part of* (2).

Analogous statements (with natural changes) are valid for $|\,\mathbb{F}_2^{V_n,0}(r)|$, for example,

$$2^n N_1(n,r) - C_{2^n}^2 N_2(n,r) \leqslant |\,\mathbb{F}_2^{V_n,0}(r)| \leqslant$$

$$\leqslant 2^n N_1(n,r) - C_{2^n}^2 N_2(n,r) + C_{2^n}^3 N_3(n,r). \tag{2$'$}$$

By means of results of [3] we prove that for $0 < r < 2^{n-1} - 2^{n/2-1}$

$$\binom{2^n}{r}^r \left(\frac{2^n}{2^n - r}\right)^{2^n - r} \frac{1}{\sqrt{2^{n+1}\pi V\left(1 - \frac{r}{2^{n-1}}\right)}} \left(1 - \frac{1}{2^n V\left(1 - \frac{r}{2^{n-1}}\right)}\right) <$$

$$< N_1(n,r) < \left(\frac{2^n}{r+1}\right)^{r+1} \left(\frac{2^n}{2^n - r - 1}\right)^{2^n - r - 1} \frac{1}{\sqrt{2^{n+1}\pi V\left(1 - \frac{r+1}{2^{n-1}}\right)}},$$

where $V(z) = (1-z)\ln(1-z) + (1+z)\ln(1+z)$.

As the obvious corollary of Theorem 1 we obtain inequalities

$$2^{n+1}N_1(n,r)(1 - Q(n,r)) \leqslant |\,\mathbb{F}_2^{V_n}(r)| \leqslant 2^{n+1}N_1(n,r),$$

where

$$Q(n,r) = \frac{4\,C_{2^n}^2 N_2(n,r)}{2^{n+1}N_1(n,r)} < \frac{2^n N_2(n,r)}{N_1(n,r)}.$$

Theorem 2. *If* $n \geqslant 10$, $r > 2^{n-2}$ *and* $y = 2^{n-1} - r > 0$ *then*

$$Q(n,r) < \frac{2}{5} \cdot 2^{n/2} \left(\frac{2^{n-2}}{y} + 1\right)^2 \exp\left\{-\frac{y^2}{2^{n-1}}\left(1 - \frac{3y}{2^n}\right)\right\}.$$

Corollary. *If* $n \geqslant 2$, $c > 1$ *then*

$$Q\left(n, 2^{n-1} - \sqrt{cn2^{n-1}}\right) < \frac{1}{2}\,2^{3n/2} \exp\left\{-cn\left(1 - \frac{3\sqrt{cn}}{2^{(n+1)/2}}\right)\right\}.$$

The last inequality shows that for any $c > \frac{3}{2}\ln 2$

$$Q\left(n, 2^{n-1} - \sqrt{cn2^{n-1}}\right) \to 0,\ n \to \infty,$$

i.e. the left and right parts of (2) and (2$'$) are asymptotically equivalent when $n \to \infty$, $\frac{2^{n-1}-r}{\sqrt{n2^{n-1}}} > \sqrt{\frac{3}{2}\ln 2}$. According to [2] this domain of values (n,r) is close to the domain containing almost all Boolean functions when $n \to \infty$.

From (1) and (2′) it follows that if $n, r_n \to \infty$ and $x_n(r_n) \to \infty$ then

$$\frac{|\,\mathbb{F}_2^{V_n,0}(r_n)|}{2^{2^n} B(x_n(r_n))} < \frac{2^n N_1(n, r_n)}{2^{2^n} B(x_n(r_n))} < \frac{e\sqrt{2^{n-1} n \ln 2}}{2^{n-1} - r_n - 1},$$

i. e. for $2^{n-1} - r_n - 1 > 3\sqrt{2^{n-1} n \ln 2}$ the "main" term of right hand side in (1) overestimates $|\,\mathbb{F}_2^{V_n,0}(r_n)|$.

Let $f = f(x_1, \ldots, x_n) \in \mathbb{F}_2^{V_n}$. For the partition $\{1, \ldots, n\} = I_n^s \cup J_n^{n-s}$, $I_n^s = \{i_1, \ldots, i_s\}$, $J_n^{n-s} = \{j_1, \ldots, j_{n-s}\}$ and for the collection of constants $C_{n-s} = \{c_1, \ldots, c_{n-s} \in \mathbb{F}_2\}$ we define the *subfunction* $g(J_n^{n-s}, C_{n-s}; y_1, \ldots, y_s) \in \mathbb{F}_2^{V_s}$ of f as the function obtained from $f(x_1, \ldots, x_n)$ by the following change of variables: $x_{j_k} = c_k$ $(k = 1, \ldots, n-s)$, and $x_{i_m} = y_m$ $(m = 1, \ldots, s)$.

Let $\nu(f, s, r)$ be the number of subfunctions $g(J_n^{n-s}, C_{n-s}; y_1, \ldots, y_s) \in \mathbb{F}_2^{V_s}$ with distance from $\mathbb{A}_s$ not exceeding r, i. e. the number of pairs (J_n^{n-s}, C_{n-s}) such that $g(J_n^{n-s}, C_{n-s}) \in \mathbb{F}_2^s(r)$.

Theorem 3. *If $\varphi(x_1, \ldots, x_n)$ is a random boolean function with the uniform distribution on $\mathbb{F}_2^{V_n}$ then for $r < 2^{s-2}$*

$$\mathbf{E}\,\nu(\varphi, s, r) = C_n^s\, 2^{n-2^s+1} \sum_{j=0}^{r} C_{2^s}^j,$$

$$2^{n-2^s+1} C_n^s C_{2^s}^r \leqslant \mathbf{E}\,\nu(\varphi, s, r) \leqslant 2^{n+1}\, C_n^s \left(\tfrac{2}{3}\right)^{2^{s-2}}.$$

Note that the left hand side tends to infinity if $s \leqslant \log_2 n$, $n \to \infty$, and the right hand side tends to 0 if $s \geqslant \log_2 n + 3$, $n \to \infty$; thus a "threshold" value of s belongs to the range from $\log_2 n$ to $\log_2 n + 3$ when $n \to \infty$.

Acknowledgements
This work was supported by RFBR, grant 08-01-00078.

References

[1] Logachev O. A., Salnikov A. A., Yashchenko V. V. Boolean functions in coding theory and cryptology (in Russian) — M.: MCCME, 2004.

[2] Ryasanov B. V. Probabilistic methods in the theory of approximation of discrete functions. — Probabilistic Methods in Discr. Math.: Proceedings of the Third International Petrozavodsk Conference, Moscow: TVP/VSP, 1993, p. 403–412.

[3] Alfers D., Dinges H. A normal approximation for Beta and Gamma tail probabilities. — Z. Wahrscheinlichkeitstheor. verw. Gebiete, 1984, v. 65, p. 399-420.