Distribution of the absolute indicator of random Boolean functions

Florian Caullery, François Rodier

Noname manuscript No.
(will be inserted by the editor)

Distribution of the absolute indicator of random Boolean functions

Florian Caullery · François Rodier

Received: date / Accepted: date

Abstract The absolute indicator is one of the measures used to determine the resistance offered by a Boolean function when used in the design of a symmetric cryptosystem. It was proposed along with the sum of square indicator to evaluate the quality of the diffusion property of block ciphers and hash functions. While the behaviour of the sum of square of random Boolean functions was already known, what remained was the study of the comportment of the absolute indicator of random Boolean functions. As an application, we show that the absolute indicator can distinguish a nonrandom binary sequence from a random one.

Keywords absolute indicator · random Boolean function · autocorrelation · nonlinearity · finite field.

Mathematics Subject Classification (2000) 94A60 · 11T712 · 14G50

1 Introduction

Let $\mathbb{F}_2^n$ be the n-dimensional vector space over the finite field of 2 elements. The Boolean functions are the functions from $\mathbb{F}_2^n$ to $\mathbb{F}_2$. They are used in cryptosystems as they are a convenient way to describe S-Boxes. We refer to [Car10] for a global survey on the cryptographic applications of Boolean

This work was partially done while the first author was funded by the Instituto Nacional de Matematica Pura e Aplicada (IMPA), Rio de Janeiro, RJ - Brazil and the foundation of Coordenaçao de Aperfeiccoamento de Pessoal de Nivel Superior (Capes) of the Brazilian ministry of education

F. Caullery
DarkMatter LLC, Abu Dhabi, United Arab Emirates
E-mail: florian.caullery@darkmatter.ae

F. Rodier
Institut Mathématiques de Marseille, France
E-mail: francois.rodier@univ-amu.fr

functions. There exist several ways to measure the resistance offered by a Boolean function against specific cryptanalysis. Among them, we should mention the *nonlinearity*, which is the Hamming distance of the function to the set of affine functions, the *absolute indicator* and the *sum of squares*, which are usually grouped into the term of Global Avalanche Criterion. The two latter were introduced in [ZZ96] to measure the capacity of a Boolean function to ensure the propagation property of a cryptosystem. The relations between the absolute indicator and the other cryptographic measures have been extensively studied as well as the distribution of the absolute indicator of certain specific classes of Boolean functions (sometimes under the name of auto-correlation value), see for example [Zho, ZXX09, CCC+00, TKB01, ZZ01].

Another possible use of these measures was proposed by one of the the authors during a workshop bringing in industrial representatives and academics (see the online report in [CGM+14]). The problem set out was to determine if a short binary sequence could be pseudo-random. We proposed to see a binary sequence of length 2^n as the truth table of a Boolean function, compute its nonlinearity and absolute indicator and compare it to the expected values of random Boolean functions. The idea came from the fact that it was proved by Schmidt in [Sch15], finalising the work of Rodier [Rod06], Dib [Dib10, Dib14] and Lytsin and Shpunt [LS09], that the nonlinearity of random Boolean functions is concentrated around its expected value. Also, the same kind of result exists for the fourth moment of the nonlinearity of random Boolean functions, which is actually the sum of squares (see [Rod03]).

However, there did not exist a study of the distribution of the absolute indicator of random Boolean functions, we fill the gap with our result. The difficulty of our case arises from the fact that we are not dealing only with independent random variables. Indeed, by writing the truth table of a Boolean function as a binary sequence, one can see the absolute indicator as the correlation of order 2 between the sequence and its circular shift (or rotation). Hence, we cannot straightforwardly apply estimates on sums of independent variables. We overcome the problem by carefully analysing the dependencies between the random variables. Also, our techniques allow us to keep the proofs simple and only based on combinatorics. As an example, we show that a short binary sequence would be detected to be non random with respect to the absolute indicator while it would pass the test with nonlinearity.

We begin by defining the absolute indicator of a Boolean function and set the formal frame for the study of the distribution of the absolute indicator.

Definition 1 Let $B_n = \{f_n : \mathbb{F}_2^n \to \mathbb{F}_2\}$ be the set of the Boolean functions of n variables and let $f_n \in B_n$. For all $u \in \mathbb{F}_2^n$, write

$$\Delta_{f_n}(u) := \sum_{x \in \mathbb{F}_2^n} (-1)^{f_n(x)+f_n(x+u)}.$$

The absolute indicator of f_n is defined by

$$\Delta(f_n) := \max_{u \in \mathbb{F}_2^n - \{0\}} |\Delta_{f_n}(u)|.$$

Our goal is to show the following theorem on the distribution of the absolute indicator of random Boolean functions. We denote by $\mathcal{E}(X)$ the expectation of a random variable X and by $f_n : \mathbb{F}_2^n \to \mathbb{F}_2$ a random Boolean function running over the set B_n provided with equiprobability.

Theorem 1 *The expectation of the absolute indicator has the following limit:*

$$\frac{\mathcal{E}[\Delta(f_n)]}{\sqrt{2^n \log 2^n}} \to 2$$

as $n \to \infty$. Moreover,

$$\mathbb{P}\left[\left|\frac{\Delta(f_n)}{2\sqrt{2^n \log 2^n}} - 1\right| > \epsilon\right] \to 0$$

for all $\epsilon > 0$ as $n \to \infty$.

The strategy to prove this result is based on ideas developed in [Sch14] and in [Rod03] with suitable alteration and on an idea in [AS00]. The main difference in our case is that the estimation of the expected value of the absolute indicator involves dealing with non-independent random variables. By separating the dependent and independent parts in the expectation, we are able to apply classical results from martingales theory and then derive the best estimation possible. To sum up, we first prove that, for all $\epsilon > 0$,

$$\mathbb{P}\left[\frac{\mathcal{E}[\Delta(f_n)]}{\sqrt{2^n \log 2^n}} > 2 + \epsilon\right] \to 0$$

as $n \to \infty$ and then show that, for all $\delta > 0$, the set

$$N(\delta) = \left\{n > 1 : \frac{\mathcal{E}[\Delta(f_n)]}{\sqrt{2^n \log 2^n}} < 2 - \delta\right\}$$

is finite. Moreover, in the last section, we prove that the absolute indicator of a random Boolean function converges almost surely towards $2\sqrt{2^n \log 2^n}$.

This article is an expansion of a paper which was presented at WCC17 [CR17].

We begin with preliminary lemmata which shall be used in the fourth section to prove Theorem 1.

2 Preliminary lemmata

From now on we set $l = 2^n$.

Lemma 1 *For all $\epsilon > 0$, as $n \to \infty$,*

$$\mathbb{P}\left[\frac{\Delta(f_n)}{\sqrt{l \log l}} > 2 + \epsilon\right] \to 0.$$

Proof Write $\mu_l = (2+\epsilon)\sqrt{l \log l}$. The union bound gives:

$$\mathbb{P}(\Delta(f_n) > \mu_l) \leq \sum_{u \in \mathbb{F}_2^n - \{0\}} \mathbb{P}(|\Delta_{f_n}(u)| > \mu_l).$$

Now note the trivial fact that $f_n(x) + f_n(x+u) = f_n(x+u) + f_n(x+u+u)$. Hence, choosing one subset $C_u \subset \mathbb{F}_2^n$ of maximal cardinality such that if $x \in C_u$, then $x + u \notin C_u$, we can write

$$\Delta_{f_n}(u) = 2 \sum_{x \in C_u} X_{x,u},$$

where $X_{x,u} = (-1)^{f_n(x)+f_n(x+u)}$. Since f_n is drawn at random, we know from proposition 1.1 of [Mer06] that the $X_{x,u}$'s are independent random variables equally likely to take the value -1 or $+1$. We can now apply Corollary A.1.2 of [AS00] with $k = \sharp C_u = l/2$ to obtain

$$\begin{aligned} \sum_{u \in \mathbb{F}_2^n - \{0\}} \mathbb{P}(|\Delta_{f_n}(u)| > \mu_l) &= \sum_{u \in \mathbb{F}_2^n - \{0\}} \mathbb{P}\left[\Big| \sum_{x \in C_u} X_{x,u} \Big| > \mu_l/2 \right] \\ &\leq 2l e^{-\mu_l^2/4l} \\ &< 2l^{-\epsilon} \end{aligned}$$

which tends to 0 as $n \to \infty$. □

We now prove a lower bound on $\mathbb{P}(\Delta(f) > \lambda)$. We need the following refinement of the central limit theorem.

Lemma 2 ([Cra38], Thm. 2) *Let $X_0, X_1, \ldots$ be i.i.d random variables satisfying $\mathcal{E}[X_0] = 0$ and $\mathcal{E}[X_0^2] = 1$ and suppose that there exists $T > 0$ such that $\mathcal{E}[e^{tX_0}] < \infty$ for all $|t| < T$. Write $Y_k = X_0 + X_1 + \ldots + X_{k-1}$ and let Φ be the distribution function of a normal random variable with zero mean and unit variance. If $\theta_k > 1$ and $\theta_k / k^{1/6} \to 0$ as $n \to \infty$, then*

$$\frac{\mathbb{P}\left[|Y_k| \geq \theta_k \sqrt{k}\right]}{2\Phi(-\theta_k)} \to 1.$$

We can now apply lemma 2 to obtain the following proposition.

Proposition 1 *For all n sufficiently large,*

$$\mathbb{P}\left[|\Delta_{f_n}(u)| \geq 2\sqrt{l \log l}\right] \geq \frac{1}{2l\sqrt{\log l}}.$$

Proof From the proof of lemma 1, we know that $\Delta_{f_n}(u)$ is the double of a sum of $l/2$ mutually independent variables equally likely to take the value -1 or 1. Notice that $\mathcal{E}[e^{t(-1)^{f_n(0)+f_n(u)}}] = \cosh(t)$ and set $\xi_l = \sqrt{2 \log l}$. We can check that $\xi_l / (l/2)^{1/6} \to 0$ as $n \to \infty$ and we can now apply lemma 2 to obtain

$$\mathbb{P}\left[|\Delta_{f_n}(u)| \geq 2\sqrt{l \log l}\right] = \mathbb{P}\left[\Big| \sum_{x \in C_u} (-1)^{f_n(x)+f_n(x+u)} \Big| \geq \sqrt{l \log l} \right] \sim 2\Phi(-\xi_l)$$

with Φ as in lemma 2. Now use the fact that

$$\frac{1}{\sqrt{2\pi}z}\left(1-\frac{1}{z^2}\right)e^{-z^2/2} \leq \Phi(-z) \leq \frac{1}{\sqrt{2\pi}z}e^{-z^2/2} \quad \text{for } z>0.$$

So, as $l \to \infty$,

$$2\Phi(-\xi_n) \sim \frac{1}{l\sqrt{\pi \log l}},$$

from which the lemma follows. □

3 Upper bound on $\mathbb{P}\left[|\Delta_{f_n}(u)| \geq \lambda_l \cap |\Delta_{f_n}(v)| \geq \lambda_l\right]$

We will proceed by first estimating the expected value of $(\Delta_{f_n}(u)\Delta_{f_n}(v))^{2p}$ and then use Markov's inequality.

Let $0 < r < l$ and choose r elements $x_1, \ldots, x_r$ in $\mathbb{F}_2^n$, let f_n be a random function in B_n, and write $\tilde{f} = (-1)^{f_n}$. First remark that the following properties trivially hold:

- $\mathcal{E}[\tilde{f}(x_1)\tilde{f}(x_2)\ldots\tilde{f}(x_r)] = \mathcal{E}[\tilde{f}(x_3)\tilde{f}(x_4)\ldots\tilde{f}(x_r)]$ if $x_1 = x_2$
- $\mathcal{E}[\tilde{f}(x_1)\tilde{f}(x_2)\ldots\tilde{f}(x_r)] = 0$ or 1
- $\mathcal{E}[\tilde{f}(x_1)\tilde{f}(x_2)\ldots\tilde{f}(x_r)] = 1$ if and only if for every $y \in \mathbb{F}_2^n$ the set of the x_i's equals to y is of even cardinality.

Choose r elements $a_1, \ldots, a_r$ of $\mathbb{F}_2^n$ and define

$$E[a_1, \ldots, a_r] = \sum_{(x_1,\ldots,x_r)\in\mathbb{F}_2^n\times\mathbb{F}_2^n\times\ldots\times\mathbb{F}_2^n} \mathcal{E}\left[\tilde{f}(x_1)\tilde{f}(x_1+a_1)\ldots\tilde{f}(x_r)\tilde{f}(x_r+a_r)\right].$$

Lemma 3 *For $a \in \mathbb{F}_2^n, a \neq 0$, the following inequality is true:*

$$E[a, a_1, \ldots, a_r] \leq 2\sum_{1<i\leq r} E[a_1, \ldots, a_i + a, \ldots, a_r].$$

Proof (see [Rod03]) From the previously stated properties, we can write

$$\begin{aligned}
&E[a, a_1, \ldots, a_r] \\
&= \sum_{(x,x_1,\ldots,x_r)} \mathcal{E}\left[\tilde{f}(x)\tilde{f}(x+a)\tilde{f}(x_1)\tilde{f}(x_1+a_1)\ldots\tilde{f}(x_r)\tilde{f}(x_r+a_r)\right] \\
&= \sum_{(x_1,\ldots,x_r)} \sum \mathcal{E}\left[\tilde{f}(x)\tilde{f}(x+a)\tilde{f}(x_1)\tilde{f}(x_1+a_1)\ldots\tilde{f}(x_r)\tilde{f}(x_r+a_r)\right],
\end{aligned}$$

where the last summand is taken over $x \in \{x_1, x_1+a_1, \ldots, x_r, x_r+a_r\}$. If $x = x_1$,

$$\begin{aligned}
&\mathcal{E}\left[\tilde{f}(x)\tilde{f}(x+a)\tilde{f}(x_1)\tilde{f}(x_1+a_1)\ldots\tilde{f}(x_r)\tilde{f}(x_r+a_r)\right] \\
&= \mathcal{E}\left[\tilde{f}(x_1+a)\tilde{f}(x_1+a_1)\ldots\tilde{f}(x_r)\tilde{f}(x_r+a_r)\right] \\
&= \mathcal{E}\left[\tilde{f}(t)\tilde{f}(t+a_1+a)\tilde{f}(x_1)\tilde{f}(x_1+a_1)\ldots\tilde{f}(x_r)\tilde{f}(x_r+a_r)\right],
\end{aligned}$$

putting $t = x_1 + a_1$. In the same way, if $x = x_1 + a_1$

$$\mathcal{E}\left[\tilde{f}(x)\tilde{f}(x+a)\tilde{f}(x_1)\tilde{f}(x_1+a_1)\dots\tilde{f}(x_r)\tilde{f}(x_r+a_r)\right]$$
$$= \mathcal{E}\left[\tilde{f}(x_1)\tilde{f}(x_1+a+a_1)\dots\tilde{f}(x_r)\tilde{f}(x_r+a_r)\right].$$

□

With this lemma we will show the following.

Lemma 4 *Let r and $s \geq 2$ be integers, a and b be distinct elements of $\mathbb{F}_2^n - \{0\}$ and define $S_{r,s} = E[\underbrace{a,\dots,a}_{r},\underbrace{b,\dots,b}_{s}]$. We have*

$$S_{r,s} \leq 2(s-1)(l+2r)S_{r,s-2} + 4r(r-1)S_{r-2,s}.$$

Proof By successive application of the inequality stated in lemma 3, we obtain

$$\begin{aligned} S_{r,s} &\leq 2(s-1)lS_{r,s-2} + 2rE[\underbrace{a+b,a,\dots,a}_{r},\underbrace{b,\dots,b}_{s}] \\ &\leq 2(s-1)lS_{r,s-2} + 4r(r-1)E(\underbrace{b,a,\dots,a}_{r-1},\underbrace{b,b,\dots,b}_{s-1}) \\ &\quad + 4r(s-1)E(\underbrace{a,\dots,a}_{r-1},\underbrace{a,b,\dots,b}_{s-1}) \\ &\leq 2(s-1)(l+2r)S_{r,s-2} + 4r(r-1)S_{r-2,s}. \end{aligned}$$

□

In the case where $r = 0$ and s is even, the lemma gives

$$S_{0,s} \leq (2(s-1)l)\mathcal{E}[\underbrace{a,\dots,a}_{s-2}] \leq (2(s-1)l)(2(s-3)l)\dots((2.3l)(2l)) \leq l^{s/2}\frac{s!}{(s/2)!}.$$

We write $M_r = l^{r/2}r!/(r/2)!$, therefore $S_{0,s} \leq M_s$. We get from the preceding Lemma, assuming that $r, s \geq 2$:

$$\frac{S_{r,s}}{M_rM_s} \leq (1 + rl^{-1})\frac{S_{r,s-2}}{M_rM_{s-2}} + 2rl^{-1}\frac{S_{r-2,s}}{M_{r-2}M_s}.$$

Lemma 5 *For $2 \leq r \leq s$, r and s both even and putting $t = (r+s)/2$, we get*

$$\frac{S_{r,s}}{M_rM_s} \leq \left(1 + \frac{2rs}{l(t-1)}\right)^{(t-1)}. \tag{1}$$

Proof We will proceed by induction. The inequality is clearly true for $r = s = 2$. Hence we can suppose $2 \leq r \leq s$ and $t > 2$. Write $a_{r,s} = 1 + \frac{2rs}{l(t-1)}$

with $t = (r+s)/2$. Suppose now that the relation (1) is true for $\frac{S_{r,s-2}}{M_r M_{s-2}}$ and $\frac{S_{r-2,s}}{M_{r-2} M_s}$. It implies

$$\begin{aligned}\frac{S_{r,s}}{M_r M_s} &\leq (1+rl^{-1})\frac{S_{r,s-2}}{M_r M_{s-2}} + 2rl^{-1}\frac{S_{r-2,s}}{M_{r-2}M_s} \\ &= \left(1+2rl^{-1}\right)\left(a_{r,s-2}\right)^{(t-2)} + 2rl^{-1}(a_{r-2,s})^{(t-2)} \\ &= \left(a_{r,s-2}\right)^{(t-1)} + 2rl^{-1}\left(1-\frac{s-2}{t-2}\right)\left(a_{r,s-2}\right)^{(t-2)} + 2rl^{-1}\left(a_{r-2,s}\right)^{(t-2)}.\end{aligned}$$

Observe that $1-\frac{s-2}{t-2} = -\frac{s-r}{2(t-2)}$. So we want to show that the sum

$$\left(a_{r,s}\right)^{(t-1)} - \left(a_{r,s-2}\right)^{(t-1)} + 2rl^{-1}\frac{s-r}{2(t-2)}\left(a_{r,s-2}\right)^{(t-2)} - 2rl^{-1}\left(a_{r-2,s}\right)^{(t-2)} \quad (2)$$

is positive. One has $a_{r,s} - a_{r,s-2} = \frac{2r(s-2)}{l(t-2)(t-1)} \geq 0$, from which we deduce

$$\left(a_{r,s}\right)^{t-1} - \left(a_{r,s-2}\right)^{(t-1)} = \frac{2r(r-2)}{l(t-1)(t-2)}\sum_{i=0}^{t-2}\left(a_{r,s}\right)^{i}\left(a_{r,s-2}\right)^{(t-2-i)}.$$

By dividing the sum (2) by $2r/l$ and using $2 \leq r \leq s$ we get:

$$\begin{aligned}\frac{(r-2)}{(t-1)(t-2)}\sum_{i=0}^{t-2}\left(a_{r,s}\right)^{i}\left(a_{r,s-2}\right)^{(t-2-i)} + \frac{s-r}{2(t-2)}\left(a_{r,s-2}\right)^{t-2} - \left(a_{r-2,s}\right)^{t-2} \\ \geq \left(a_{r-2,s}\right)^{t-2}\left(\frac{r-2}{t-2} + \frac{s-r}{2(t-2)} - 1\right).\end{aligned}$$

because $2r(s-2) - 2s(r-2) = 4(s-r) \geq 0$ and consequently

$$a_{r,s} \geq a_{r,s-2} \geq a_{r-2,s}.$$

Finally, we verify that $\frac{r-2}{t-2} + \frac{s-r}{2(t-2)} - 1 = 0$. □

Now, we can conclude that

$$S_{2p,2p} \leq l^{2p}\left(\frac{(2p)!}{p!}\right)^2\left(1+\frac{8p^2}{l(2p-1)}\right)^{2p-1},$$

which will be used to prove the following:

Proposition 2 *Write $\lambda_l = 2\sqrt{l\log l}$. If f_n runs over the set B_n then, for distinct $u, v \in \mathbb{F}_2^n - \{0\}$ and n sufficiently large,*

$$\mathbb{P}\left[|\Delta_{f_n}(u)| \geq \lambda_l \cap |\Delta_{f_n}(v)| \geq \lambda_l\right] < 4l^{-2}.$$

Proof With the notation of previous lemmas and applying Markov's inequality:

$$\begin{aligned}
&\mathbb{P}\Big(\Big(\sum_{x\in\mathbb{F}_2^n}\tilde{f}(x)\tilde{f}(x+u)\geq\theta_1\Big)\cap\Big(\sum_{x\in\mathbb{F}_2^n}\tilde{f}(x)\tilde{f}(x+v)\geq\theta_2\Big)\Big)\\
&\leq\mathcal{E}\Big[\Big(\sum_{x\in\mathbb{F}_2^n}\tilde{f}(x)\tilde{f}(x+u)\sum_{x\in\mathbb{F}_2^n}\tilde{f}(x)\tilde{f}(x+v)\Big)^{2p}\Big]\Big/(\theta_1\theta_2)^{2p}\\
&\leq S_{2p,2p}/(\theta_1\theta_2)^{2p}\leq\left(1+\frac{8p^2}{l(2p-1)}\right)^{2p-1}l^{2p}\left(\frac{(2p)!}{p!}\right)^2\Big/(\theta_1\theta_2)^{2p}\quad.\quad(3)
\end{aligned}$$

If we take $\theta_1=\theta_2=\lambda_l$ and $p=n$, we have

$$\begin{aligned}
&\mathbb{P}\Big(\Big(\sum_{x\in\mathbb{F}_2^n}\tilde{f}(x)\tilde{f}(x+u)\geq\lambda_l\Big)\cap\Big(\sum_{x\in\mathbb{F}_2^n}\tilde{f}(x)\tilde{f}(x+v)\geq\lambda_l\Big)\Big)\\
&\qquad\leq\left(1+\frac{8n^2}{l(2n-1)}\right)^{2n-1}\left(\frac{(2n)!}{n!}\right)^2\Big/(4n)^{2n}.
\end{aligned}$$

By Stirling's approximation, $\sqrt{2\pi k}k^ke^{-k}\leq k!\leq\sqrt{3\pi k}k^ke^{-k}$, we get

$$\frac{(2n)!}{n!}\leq\frac{\sqrt{3\pi 2n}(2n)^{2n}e^{-2n}}{\sqrt{2\pi n}n^ne-n}\leq\sqrt{3}.2^{2n}n^ne^{-n}.$$

Plugging it into inequality 3, we get:

$$\begin{aligned}
&\mathbb{P}\Big(\Big(\sum_{x\in\mathbb{F}_2^n}\tilde{f}(x)\tilde{f}(x+u)\geq\lambda_l\Big)\cap\Big(\sum_{x\in\mathbb{F}_2^n}\tilde{f}(x)\tilde{f}(x+v)\geq\lambda_l\Big)\Big)\\
&\qquad\leq 3\left(1+\frac{8n^2}{l(2n-1)}\right)^{2n-1}e^{-2n}<l^{-2}
\end{aligned}$$

for $n\geq 7$. The proposition is now straightforward.

□

4 Proof of theorem 1

We first recall the following inequality from martingales theory:

Lemma 6 ([McD89]) *Let $X_0,\ldots,X_{l-1}$ be mutually independent random variables taking values in a set S. Let $g:S^l\to\mathbb{R}$ be a measurable function and suppose that*

$$|g(x)-g(y)|\leq c$$

whenever x and y differ only in one coordinate. Define the random variable $Y=g(X_0,\ldots,X_{l-1})$. Then, for all $\theta\geq 0$,

$$\mathbb{P}\left[|Y-\mathcal{E}[Y]|\geq\theta\right]\leq 2\exp\Big(-\frac{2\theta^2}{c^2l}\Big).$$

Now let $(x_i)_{0\leq i\leq l-1}$ be a bijection between the set of nonnegative numbers strictly smaller than l and the set $\mathbb{F}_2^n$. Let σ_u for $u \in \mathbb{F}_2^n$ be the substitution in the set $\{0,\ldots,l-1\}$ such that $x_{\sigma_u(i)} = x_i + u$ and let g be the function

$$\begin{aligned} g : \{0,1\}^l &\to \mathbb{R} \\ (X_0,\ldots,X_{l-1}) &\mapsto \max_{u\in\mathbb{F}_2^n-\{0\}} \sum_{i=0}^{l-1} (-1)^{X_i+X_{\sigma_u(i)}}. \end{aligned}$$

Clearly $g(f(x_0),\ldots,f(x_{l-1})) = \Delta(f)$ and we can apply lemma 6 with $c = 4$ to obtain the following corollary.

Corollary 1 *For $\theta \geq 0$,*

$$\mathbb{P}\left[|\Delta(f_n) - \mathcal{E}[\Delta(f_n)]| \geq \theta\right] \leq 2\exp\left(-\frac{\theta^2}{8l}\right).$$

The first part of Theorem 1 is proven by the following result.

Theorem 2 *The following limit holds when $n \to \infty$,*

$$\frac{\mathcal{E}\left[\Delta(f_n)\right]}{\sqrt{l\log l}} \to 2.$$

Proof By the union bound and triangle inequality, we have, for all $\epsilon > 0$

$$\begin{aligned} &\mathbb{P}\left[\frac{\mathcal{E}\left[\Delta(f_n)\right]}{\sqrt{l\log l}} - 2 > \epsilon\right] \\ &\qquad \leq \mathbb{P}\left[\frac{\mathcal{E}\left[\Delta(f_n)\right]}{\sqrt{l\log l}} - \frac{\Delta(f_n)}{\sqrt{l\log l}} > \frac{1}{2}\epsilon\right] + \mathbb{P}\left[\frac{\Delta(f_n)}{\sqrt{l\log l}} - 2 > \frac{1}{2}\epsilon\right]. \end{aligned}$$

The right hand side of the last inequality goes to zero as $n \to \infty$ by corollary 1 and proposition 1. So we conclude

$$\limsup_{n\to\infty} \frac{\mathcal{E}\left[\Delta(f_n)\right]}{\sqrt{l\log l}} \leq 2.$$

The proof of the claim is based on an idea in [LS09]: to bound by below $\frac{\mathcal{E}[\Delta(f_n)]}{\sqrt{l\log l}}$, we will prove that the following set is finite. Let $\delta > 0$ and define

$$N(\delta) = \left\{n > 1 : \frac{\mathcal{E}\left[\Delta(f_n)\right]}{\sqrt{l\log l}} < 2 - \delta\right\}.$$

Now set $\lambda_l = 2\sqrt{l\log l}$ and, for each $n \geq 7$, chose a subset $W \subset \mathbb{F}_2^n - \{0\}$ of size $\lceil l/\log l\rceil$. Hence

$$\begin{aligned} \mathbb{P}\left[\Delta(f_n) \geq \lambda_l\right] &\geq \mathbb{P}\Big[\max_{u\in W} |\Delta_{f_n}(u)| \geq \lambda_l\Big] \\ &\geq \sum_{u\in W} \mathbb{P}\left[|\Delta_{f_n}(u)| \geq \lambda_l\right] - \sum_{u,v\in W, u\neq v} \mathbb{P}\left[|\Delta_{f_n}(u)| \geq \lambda_l \cap |\Delta_{f_n}(v)| \geq \lambda_l\right] \end{aligned}$$

by Bonferroni inequality. Propositions 1 and 2 give, for l big enough:

$$\begin{aligned}\mathbb{P}\left[\Delta(f_n) \geq \lambda_l\right] &\geq |W|\frac{1}{2l\sqrt{\log l}} - 4\frac{|W|^2}{l^2} \\ &\geq \frac{1}{10(\log l)^{3/2}}. \end{aligned} \tag{4}$$

By definition of $N(\delta)$, $\lambda_l \geq \mathcal{E}[\Delta(f_n)]$ so we can apply corollary 1 with $\lambda_l - \mathcal{E}[\Delta(f_n)]$ so that for all $n \in N(\delta)$,

$$\mathbb{P}\left[\Delta(f_n) \geq \lambda_l\right] \leq 2\exp\Big(-\frac{1}{8l}(\lambda_l - \mathcal{E}[\Delta(f_n)])^2\Big).$$

Comparison with (4) implies

$$\frac{\mathcal{E}[\Delta(f_n)]}{\sqrt{l\log l}} \geq 2 - \sqrt{\frac{12\log\log l + 8\log 20}{\log l}},$$

which means in view of its definition that $N(\delta)$ is finite for all $\delta > 0$. □

We can now easily deduce the second part of Theorem 1.

Corollary 2 *As $n \to \infty$,*

$$\frac{\Delta(f_n)}{\sqrt{l\log l}} \to 2 \quad \textit{in probability.}$$

Proof By the triangle inequality

$$\begin{aligned}&\mathbb{P}\left[\left|\frac{\Delta(f_n)}{\sqrt{l\log l}} - 2\right| > \epsilon\right] \\ &\qquad \leq \mathbb{P}\left[\left|\frac{\Delta(f_n)}{\sqrt{l\log l}} - \frac{\mathcal{E}\left[\Delta(f_n)\right]}{\sqrt{l\log l}}\right| > \frac{\epsilon}{2}\right] + \mathbb{P}\left[\left|\frac{\mathcal{E}\left[\Delta(f_n)\right]}{\sqrt{l\log l}} - 2\right| > \frac{\epsilon}{2}\right]. \end{aligned} \tag{5}$$

By Theorem 2 and Corollary 1, the two terms on the right-hand side go to 0 as $n \to \infty$ which proves the corollary.

□

The combination of these two last theorems gives the proof of Theorem 1.

5 Stronger convergence

The goal of this section is to use the Borel-Cantelli Lemma to prove the following Theorem:

Theorem 3 *Denote by Ω the set of infinite sequences of elements of $\mathbb{F}_2$ and by B the space of functions from Ω to $\mathbb{F}_2$. For every $f \in B$, we denote its restriction to its n first coordinates by f_n which is in B_n. We define the following probability measure on B:*

$$\mathbb{P}[f \in B : f_n = g \in B_n] = 2^{-2^n}$$

for all $g \in B_n$. Now let f be a random element in B. Then

$$\lim_{n\to\infty} \frac{\Delta(f_n)}{\sqrt{l \log l}} = 2 \text{ almost surely.}$$

Proof Let f be a random element in B and f_n its restriction. We know from the proof of Theorem 1 that the inequality (5) is true. By Theorem 2, the second term of the right hand side in (5) is null for n sufficiently large. From Corollary 1, the first term is bounded by $2l^{-\frac{\epsilon^2}{32}} = 2^{\frac{-n\epsilon^2+32}{32}}$. Therefore, for $\epsilon > 0$:

$$\sum_{n=1}^{\infty} \mathbb{P}\left[\left|\frac{\Delta(f_n)}{\sqrt{l \log l}} - 2\right| > \epsilon\right] < \infty.$$

Now applying the Borel-Cantelli Lemma gives the theorem. □

6 The non-linearity vs the autocorrelation

6.1 Application: a test on Boolean functions

The article [CGM + 14] proposes as application of the absolute indicator a test on the Boolean functions to know if a random Boolean function (for example given by a True Random Number Generator) would be reliable (see the Introduction). We will show here that autocorrelation can sometimes detect a function that cannot pass the test whereas with non-linearity it does.

Indeed a common weakness in random function generators is that a function can always loop, that is, it can be periodic. We will see that a function with two periods can pass the test of non-linearity, but not that of autocorrelation.

6.2 Two periods function

Let a Boolean function on $\mathbb{F}_2^n$ be defined by

$$\begin{aligned} f : \mathbb{F}_2^n = \mathbb{F}_2^{n-1} \times \mathbb{F}_2 &\longrightarrow \mathbb{F}_2 \\ (x, 0) &\longmapsto g(x) \\ (x, 1) &\longmapsto g(x). \end{aligned}$$

where g is a Boolean function on $\mathbb{F}_2^{n-1}$.

6.2.1 The nonlinearity

We want to compute the nonlinearity of f based on that of g.

Proposition 3 *The nonlinearity of f (computed in $\mathbb{F}_2^n$) is the double of the nonlinearity of g (computed in $\mathbb{F}_2^{n-1}$).*

Proof Using the formulas

$$NL(f) = 2^{n-1} - 1/2 \max_{u \in \mathbb{F}_2^n} |\widehat{f}(u)|$$

and

$$\widehat{f}(u) = \sum_{x \in \mathbb{F}_2^n} (-1)^{f(x)+x.u}$$

and by taking $u = (v, a)$ for $v \in \mathbb{F}_2^{n-1}$ and $a \in \mathbb{F}_2$, it is an easy matter to show that $\widehat{f}(v, 0) = 2\widehat{g}(v)$, $\widehat{f}(v, 1) = 0$. Therefore $NL(f) = 2NL(g)$. □

To perform the test we need to compare this nonlinearity to the one of random Boolean functions. Using the calculation of Litsyn and Shpunt [LS09], we see that with a probability of $O(1/n^4)$ most of the function with n variables will lie in $[lowNL_n,\ highNL_n]$ with $lowNL_n = 2^{n-1} - \sqrt{2^{n-1}(n \log 2 + 3.5 \log(n \log 2) + 0.125)}$ and $highNL_n = 2^{n-1} - \sqrt{2^{n-1}(n \log 2 - 4.5 \log(n \log 2))}$. As we want to determine the pseudo-randomness of only short sequences (say $n < 20$) we have

$$2lowNL_{n-1} < lowNL_n < 2highNL_{n-1} < highNL_n.$$

This means that the segment supporting the nonlinearity of most functions with a non-linearity close to $2NL(g)$ intersect the segment supporting the nonlinearity of most random Boolean functions with n variables. So the function f is likely to pass the test.

6.2.2 The autocorrelation

Proposition 4 *The autocorrelation of f is 2^n.*

Proof If we take $u = (0, \ldots, 0, 1) = (\mathbf{0}, 1) \in \mathbb{F}_2^{n-1} \times \mathbb{F}_2$ it is straigthforward to check that

$$\Delta(f) = \Delta_f(u) = 2^n.$$

□

This value is quite different from the value expected for the autocorrelation of most of the functions which is $2\sqrt{2^n \log(2^n)}$. Hence the function f will not pass the test with autocorrelation, as with nonlinearity it may.

6.3 Disturbed two period function

We want to stress the fact that this phenomenon still exist if the function is slightly disturbed (by a noise for instance).

Define inductively a sequence $(f_i)_{0\leq i\leq r}$ such that $f_0 = f$, and f_{i+1} is is obtained from f_i by choosing a random value i in the truth table of f_i and changing the sign of $f_i(u_i)$. So the Boolean function f_r is equal to f, except for a set E of r points.

6.3.1 The autocorrelation

Proposition 5 *The autocorrelation of f_r fulfills*

$$\Delta(f_r) \geq 2^n - 4r$$

where r is the number of errors in f_r with respect to f.

Proof Let us take $u = (0, \ldots, 0, 1) = (\mathbf{0}, 1) \in \mathbb{F}_2^{n-1} \times \mathbb{F}_2$. Then

$$\begin{aligned}\Delta_{f_r}(\mathbf{0},1) &= \sum_{x\in\mathbb{F}_2^n, x\notin E, x+u\notin E} (-1)^{f_r(x)+f_r(x+u)} + \sum_{x\in\mathbb{F}_2^n, x\in E\cup(E+u)} (-1)^{f_r(x)+f_r(x+u)} \\ &\geq \sum_{x\in\mathbb{F}_2^n, x\notin E, x+u\notin E} (-1)^{f(x)+f(x+u)} - 2r \\ &\geq 2^n - 4r\end{aligned}$$

because $f_r(x+u) = f_r(x)$ if $x \in \mathbb{F}_2^n, x \notin E, x+u \notin E$. Therefore

$$\Delta(f_r) = \sup_u \Delta_{f_r}(u) \geq \Delta_{f_r}(\mathbf{0}, 1) \geq 2^n - 4r.$$

□

6.3.2 The nonlinerarity

Proposition 6 *The nonlinearity of f_r is such that*

$$NL(f) - r \leq NL(f_r) \leq NL(f) + r.$$

Proof The nonlinearity of a Boolean function is the minimum of the number of bits that you need to change in the truth table of this function to get an affine function. To get f_r from f you have to change at most r bits. So $NL(f) \leq NL(f_r) + r$. To get f from f_r you have to change at most r bits. So $NL(f_r) \leq NL(f) + r$. The conclusion follows. □

One can even get a more precise estimate of the nonlinearity.

Proposition 7 *The following inequality is true:*

$$P(|NL(f_r)) - NL(f)| > s) \leq 2e^{-s^2/2r}.$$

Proof Let we define a random variable X_i for $1 \leq i \leq r$ by $X_i = NL(f_i) - NL(f_{i-1})$ so that $NL(f_r) - NL(f) = \sum_1^r X_i$. These perturbation are independent, so Chernoff bound gives the following result:

$$P\left[\left|\sum_1^r X_i\right| > s\right] < 2e^{-s^2/2r}.$$

Whence the result. □

This proposition shows that the nonlinearity of f_r does not deviate too much from the nonlinearity of f. Hence again a function slightly disturbed will pass the test for nonlinearity, as it will not for autocorrelation.

7 Conclusion

We proved that the absolute indicator of most of the Boolean functions is close to a small value. Thus we draw the following conclusions:

- One should not consider the absolute indicator of a Boolean function as a primary criterion in the design of symmetric cryptographical primitives but focus on other properties relevant to his desired application. The attention should be also given to, for example, simplicity of the algebraic expression (to allow easy bitsliced or efficient hardware implementation) or nonlinearity. Once a Boolean function is selected, the designer should only verify that the absolute indicator is not too far from the expected value.
- Regarding the application of the absolute indicator proposed in [CGM+14], one can say that this test would have an clearly favour type I errors against type II errors, i.e., a string not passing the test is certainly not random while we cannot guaranty that a string passing the test is "truly" random.
- As an example of this test, we have shown that some short binary sequence would be detected to be non random with the absolute indicator while it would pass the test with nonlinearity.

References

[AS00] Noga Alon and Joel H. Spencer, The probabilistic method, Wiley & Sons, Hoboken, NJ, USA (2000).

[CCC+00] Anne Canteaut, Claude Carlet, Pascale Charpin, et al., Propagation characteristics and correlation-immunity of highly nonlinear Boolean functions. In: Advances in Cryptology EUROCRYPT 2000, Springer, pp. 507-522 (2000).

[Car10] Carlet, Claude, Boolean Models and Methods in Mathematics, Computer Science, and Engineering, Cambridge University Press, New York, NY, USA, pp 398-469, (2010).

[CGM+14] Florian Caullery, Alexander Getmanenko, Vito Mandorino, et al. Semaine d'Etude Mathématiques et Entreprises 9 : Testing the reliability of a true random generator at run time, working paper or preprint, url: `https://hal.archives-ouvertes.fr/hal-01021026` (Apr. 2014).

[CR17] Florian Caullery, François Rodier, Distribution of the absolute indicator of random Boolean functions, Proceedings of The Tenth International Workshop on Coding and Cryptography, http://wcc2017.suai.ru/proceedings.html, (2017).

[Cra38] H. Cramer, Sur un nouveau théorème limite de la théorie des probabilités. In: Actualités Sci. Indust. 736 , pp. 5-23 (1938)

[Dib10] Stéphanie Dib, Distribution of Boolean Functions According to the Second-Order Nonlinearity. In: Arithmetic of Finite Fields: Third International Workshop, WAIFI 2010, Istanbul, Turkey, June 27-30, pp. 86-96 (2010).

[Dib14] Stéphanie Dib, Asymptotic Nonlinearity of Vectorial Boolean Functions, In: Cryptography Commun, 6.2, pp. 103-115 (2014).

[LS09] Simon Litsyn and Alexander Shpunt, On the Distribution of Boolean Function Nonlinearity. In: SIAM Journal on Discrete Mathematics 23.1, pp. 79-95 (2009).

[McD89] Colin McDarmid, On the method of bounded differences. In: Surveys in Combinatorics, Ed. by J. Siemons. Vol. 141, London Mathematical Society Lectures Notes, Cambridge Univ. Press, Cambridge, pp. 148-188 (1989).

[Mer06] Idris David Mercer. Autocorrelations of random binary sequences. In: Combinatorics, Probability and Computing 15.05, pp. 663- 671 (2006).

[Rod03] François Rodier, On the nonlinearity of Boolean functions. In: Proceedings of WCC2003, Workshop on coding and cryptography, pp. 397-405 (2003).

[Rod06] François Rodier, Asymptotic nonlinearity of Boolean functions. In: Des. Codes Cryptogr. 40.1, pp. 59-70 (2006).

[Sch14] Kai-Uwe Schmidt, The peak sidelobe level of random binary sequences. In: Bulletin of the London Mathematical Society (2014),

[Sch15] Kai-Uwe Schmidt, Nonlinearity measures of random Boolean functions. In: Cryptography and Communications, pp. 1-9 (2015).

[TKB01] Yuriy Tarannikov, Peter Korolev, and Anton Botev, Advances in Cryptology—ASIACRYPT 2001 Proceedings, In: Chap. Autocorrelation Coefficients and Correlation Immunity of Boolean Functions, pp. 460-479 (2001).

[ZZ96] Xian-Mo Zhang and Yuliang Zheng, GACthe criterion for global avalanche characteristics of cryptographic functions. In: J. UCS The Journal of Universal Computer Science, Springer, pp. 320-337 (1996).

[ZZ01] Xian-Mo Zhang and Yuliang Zheng, Information Security and Cryptology, ICISC 2000 Proceedings. In: Chap. New Results on Correlation Immunity, pp. 49-63 (2001).

[Zho] Y. Zhou, On the distribution of auto-correlation value of balanced Boolean functions. In: Advances in Mathematics of Communications 7.3, pp. 335-347 (2013)

[ZXX09] Y. Zhou, M. Xie, and G. Xiao. On cross-correlation properties of Boolean functions. In: Communications and Networking in China, ChinaCOM 2009, Fourth International Conference on Communications and Networking in China, pp. 1-5 (2009).