Statistical and Linear Independence of Binary Random Variables

Kaisa Nyberg

Department of Computer Science,
Aalto University School of Science, Finland
kaisa.nyberg@aalto.fi

January 30, 2018

Abstract. Linear cryptanalysis makes use of statistical models that consider linear approximations over practical and ideal block ciphers as binary random variables. Recently, more complex models have been proposed that take also into account the statistical behavior of correlations of linear approximations over the key space of the cipher and over the randomness of the ideal cipher. The goal of this ongoing work is to investigate independence properties of linear approximations and their relationships. In this third revised version we show that the assumptions of Proposition 1 of the previous version are contradictory and hence renders that result useless. In particular, we prove that linear and statistical independence of binary random variables are equivalent properties in a vector space of variables if and only if all non-zero variables in this vector space are balanced, that is, correspond to components of a permutation. This study is motivated by finding reasonable wrong-key hypotheses for linear cryptanalysis and its generalizations which will also be discussed.

Keywords: Xiao-Massey lemma, block cipher, linear cryptanalysis, linear approximation, random Boolean function, random vectorial Boolean function, multidimensional linear cryptanalysis, wrong-key hypothesis

1 Introduction

Linear cryptanalysis is a method that is used for distinguishing a block cipher from a random permutation and can be extended to key recovery attacks in practical applications.

To this end, a cryptanalyst builds a statistical model of the linear approximations over the cipher, on the one hand, and over a random permutation, on the other hand. Sometimes only the latter is used.

The goal of this ongoing work is to investigate the relationship between linear and statistical independence of linear approximations seen as random binary variables over the text space. It is clear that linearly dependent linear approximations cannot be statistically independent. On the other hand, it would be important to know what kind of independence assumptions are required for a set of linear approximations that is used in multiple linear cryptanalysis.

Recently, more complex models have been proposed that take also into account the statistical behavior of correlations of linear approximations over the key space of the cipher and over the randomness of the ideal cipher. In the previous version of this report, it was stated in Proposition 1 that under the assumption of pairwise independence of the linear approximations and non-zero variance of their correlations over the key space, statistical independence of correlations is equivalent to linear independence of the approximations. Unfortunately, there is no such world, since pairwise statistical independence of linear approximations imply that all nontrivial correlations are equal to zero and do not have non-zero variance.

After proving this result, we discuss only informally some aspects of statistical independence of correlations and give some partial results.

In the context of linear cryptanalysis, a linear approximation of a transformation F from $\mathbb{F}_2^n$ to $\mathbb{F}_2^m$ is a Boolean function in $\mathbb{F}_2^n$ defined by two vectors a, $b \in \mathbb{F}_2^n$ as follows

$$x \mapsto a \cdot x + b \cdot F(x).$$

In the statistical setting, a linear approximation is considered as a binary random variable X over the given space of transformations with a probability density function defined by

$$\Pr(X = 0) = \#\{x \in \mathbb{F}_2^n \,|\, a \cdot x + b \cdot F(x) = 0\}.$$

So we can write $X = a{\cdot}x{+}b{\cdot}F(x)$. In the algebraic setting, a linear approximation $a \cdot x + b \cdot F(x)$ is identified with the vector (a, b) in the linear space $\mathbb{F}_2^n \times \mathbb{F}_2^n$ over $\mathbb{F}_2$.

2 Independence of Binary Random Variables

In this section, we consider binary random variables X, which form a linear space $\mathcal{X}$ over $\mathbb{F}_2$, and their statistical and linear independence. We show that under the condition of pairwise statistical independence of all variables, random variables in any subset of $\mathcal{X}$ are statistically independent if and only if they are linearly independent.

We first recall the classical Xiao-Massey lemma [6]. For a short proof, see [3].

Lemma 1. *(Xiao-Massey lemma) A binary random variable Y is independent of the set of k independent binary variables $X_1, \ldots, X_k$ if and only if Y is independent of the linear combination $\lambda_1 X_1 + \cdots + \lambda_k X_k$ for every choice of $\lambda_1, \ldots, \lambda_k$, not all zero, in $\mathbb{F}_2$.*

Let us now state the main result.

Theorem 1. *Let $\mathcal{X}$ be a linear space of binary random variables over $\mathbb{F}_2$ such that any two different variables in $\mathcal{X}$ are statistically independent. Then linearly independent random variables in $\mathcal{X}$ are also statistically independent. The converse holds for nonzero random variables in $\mathcal{X}$.*

The proof of the theorem goes by induction, where the main step is given by the following lemma.

Lemma 2. *Let $\mathcal{X}$ be a linear space of binary random variables over $\mathbb{F}_2$ such that any two different variables in $\mathcal{X}$ are statistically independent. Assume that the binary random variables $X_1, \dots, X_k$ in $\mathcal{X}$ are linearly and statistically independent. If given $Y \in \mathcal{X}$ the variables $X_1, \dots, X_k, Y$ are linearly independent, then they are also statistically independent.*

Proof. Assume that $X_1, \dots, X_k, Y$ are statistically dependent. Since $X_1, \dots, X_k$ are independent, it means that Y is dependent of the set $X_1, \dots, X_k$. By the Xiao-Massey lemma, this can happen only if there exist $\lambda_1, \dots, \lambda_k$ not all zero in $\mathbb{F}_2$ such that Y and $\lambda_1 X_1 + \dots + \lambda_k X_k$ are statistically dependent. Since both of these variables are in $\mathcal{X}$ it follows that Y and $\lambda_1 X_1 + \dots + \lambda_k X_k$ must be equal, and therefore $X_1, \dots, X_k, Y$ are linearly dependent. □

Proof. (Proof of Theorem 1) Assume first that the variables $X_1, \dots, X_m$ in $\mathcal{X}$ are linearly independent. For $2 \leq k < m$ let us state the induction hypothesis as follows: If $X_1, \dots, X_k$ are linearly independent, then they are statistically independent. Since linear independence of any two of them implies that they are different, they are also statistically independent by the assumption. Hence the induction hypothesis holds for $k = 2$.

Let us assume that the induction hypothesis holds for k, and let $X_1, \dots, X_{k+1}$ be linearly independent. Then $X_1, \dots, X_k$ are linearly independent and hence by the induction hypothesis also statistically independent. By Lemma 2 it follows that $X_1, \dots, X_{k+1}$ are statistically independent.

Assume then that the variables $X_1, \dots, X_m$ are nonzero and linearly dependent. W.l.o.g it can be assumed that there exist a relation

$$X_1 = X_2 + \dots + X_k$$

where $X_2, \dots, X_k$ are linearly independent and $k \leq m$. By the first part of the proof it then follows that $X_2, \dots, X_k$ are statistically independent. Now by the Xiao-Massey lemma, Lemma 1, the variable X_1 must be statistically dependent of $X_2, \dots, X_k$. Hence $X_1, \dots, X_m$ are not statistically independent. □

3 Linear Spaces of Binary Variables

Let us first recall the piling-up lemma. We state it here for two variables. A proof for an arbitrary number of variables can be found in [5].

Lemma 3. Piling-up Lemma. *Let X_1 and X_2 be binary random variables. If X_1 and X_2 are independent then*

$$\Pr(X_1 + X_2 = 0) - \frac{1}{2} = 2(\Pr(X_1 = 0) - \frac{1}{2})(\Pr(X_2 = 0) - \frac{1}{2}). \tag{1}$$

The aim of this section is to show that pairwise statistical independence in a vector space of binary random variables is quite a stringent condition and essentially implies balancedness of all non-constant random variables. The next result is an easy consequence of the piling-up lemma, and can be found in textbooks.

Lemma 4. *If binary random variables X_1 and X_2 are statistically independent, then X_1 and $X_1 + X_2$ are statistically independent only if X_1 is constant or $\Pr(X_2 = 0) = \frac{1}{2}$.*

Proof. If X_1 and $X_1 + X_2$ are statistically independent, then by the piling-up lemma

$$\Pr(X_2 = 0) - \frac{1}{2} = 2(\Pr(X_1 = 0) - \frac{1}{2})(\Pr(X_1 + X_2 = 0) - \frac{1}{2}).$$

On the other hand, by the independence of X_1 and X_2 Equation (1) holds. By substituting it to the above equation, one obtains

$$\Pr(X_2 = 0) - \frac{1}{2} = 4\left(\Pr(X_1 = 0) - \frac{1}{2}\right)^2 \left(\Pr(X_2 = 0) - \frac{1}{2}\right).$$

This equation holds only if X_2 is balanced or X_1 is constant.

Given an arbitrary binary random variable X_1 let us pick another X_2. Assuming pairwise independence of both pairs X_1 and X_1+X_2 and X_2 and X_1+X_2 we get the following result. In particular, the always-one variable is excluded, since X_1 and $X_2 = X_1 + 1$ are not statistically independent.

Theorem 2. *If in a vector space $\mathcal{X}$ of binary random variables all elements are pairwise statistically independent, then all non-zero elements in $\mathcal{X}$ are balanced.*

Next we show the converse, that is, that in a vector space of binary variables, balancedness implies pairwise independence. For this, we need the converse of the piling-up lemma.

Lemma 5. Converse of Piling-up Lemma. *Let X_1 and X_2 be binary random variables. If 1 holds, then X_1 and X_2 are independent.*

Proof. Let us observe that for all $(t_1, t_2) \in \{0, 1\}^2$ the following holds.

$$\begin{aligned}
&\Pr(X_1 = t_1, X_2 = t_2) \\
&= \frac{1}{4} + \frac{1}{2}(-1)^{t_1}(\Pr(X_1 = 0) - \frac{1}{2}) + \frac{1}{2}(-1)^{t_2}(\Pr(X_2 = 0) - \frac{1}{2}) \\
&+ \frac{1}{2}(-1)^{t_1+t_2}(\Pr(X_1 + X_2 = 0) - \frac{1}{2}). \qquad (2)
\end{aligned}$$

This can be verified exhaustively for all four values of (t_1, t_2), or by the general formula

$$\Pr(X = t) = 2^{-n} \sum_{a \in \mathbb{F}_2^n} (-1)^{a \cdot t}(2\Pr(a \cdot X = 0) - 1)$$

where X is a random variable in $\mathbb{F}_2^n$, $t \in \mathbb{F}_2^n$, and n any positive integer. Substituting (1) in (2) gives

$$\begin{aligned}&\Pr(X_1 = t_1, X_2 = t_2)\\ &= \left(\frac{1}{2} + (-1)^{t_1}(\Pr(X_1 = 0) - \frac{1}{2})\right)\left(\frac{1}{2} + (-1)^{t_2}(\Pr(X_2 = 0) - \frac{1}{2})\right)\\ &= \Pr(X_1 = t_1)\Pr(X_2 = t_2).\end{aligned}$$

Let us now state the converse of Theorem 2.

Theorem 3. *If in a vector space $\mathcal{X}$ of binary random variables all non-zero elements are balanced, then any two variables X_1 and X_2, $X_1 \neq X_2$, are statistically independent.*

Proof. Take X_1, $X_2 \in \mathcal{X}$ such that $X_1 \neq X_2$. By the assumption of balancedness we have

$$\Pr(X_1 + X_2 = 0) - \frac{1}{2} = 0 = 2\left(\Pr(X_1 = 0) - \frac{1}{2}\right)\left(\Pr(X_2 = 0) - \frac{1}{2}\right). \quad (3)$$

By the converse of the piling-up lemma the claim holds in this case.

4 Boolean Functions

As an application let us now consider vectorial Boolean functions F from $\mathbb{F}_2^n$ to $\mathbb{F}_2^m$ where n and m are positive integers. Let $a \in \mathbb{F}_2^n$. We call the Boolean function $a \cdot F$ the component of F. The components of F form a vector space of dimension m and can be identified with binary random variables over the input space $\mathbb{F}_2^n$. By Theorem 2 the components of a vectorial Boolean function are pairwise statistically independent only if the non-zero components are balanced. Hence F is a truncated permutation. Conversely, it follows from Theorem 3 that the components of a (truncated) permutation are pairwise statistically independent.

It follows that the only vectorial Boolean functions with pairwise independent components are the truncated permutations. We state the following corollary of Theorem 1.

Corollary 1. *A set of non-zero components of a permutation are statistically independent if and only if they are linearly independent.*

On the other hand, the condition that linearly independent components are statistically independent implies pairwise independence.

Corollary 2. *Let F be a vectorial Boolean function with the property that any set of linearly independent components are statistically independent. Then F is a truncated permutation.*

5 Statistical Independence of Correlations

Given a binary random variable X its correlation $\mathrm{cor}(X)$ is defined as

$$\mathrm{cor}(X) = \Pr(X = 0) - \Pr(X = 1) = 2\Pr(X = 0) - 1.$$

The correlations of linear approximations $X = a \cdot x + b \cdot F(x)$ depend on the function F. By randomizing the function space, the correlations become random variables.

Next we investigate whether there is any relationship between linear dependence of linear approximations and statistical independence of their correlations.

From the discussion of permutations in the previous section, we see that the correlations of any set of linear components of permutations are equal to zero and hence statistically independent. Hence linear independence is not a necessary condition for correlations to be statistically independent.

Linear approximations of functions are not balanced in general and their behavior is different from the components of permutations as demonstrated by the following two examples. First, we show that linear approximations are not necessarily pairwise statistically independent.

Example 1. Let F be a permutation, and $a \cdot x + b_1 \cdot F(x)$ and $a \cdot x + b_2 \cdot F(x)$ two linear approximations that share the same input mask. If their correlations are non-zero, they are not independent. This follows by the piling-up lemma by observing

$$\mathrm{cor}(a \cdot x + b_1 \cdot F(x))\mathrm{cor}(a \cdot x + b_2 \cdot F(x)) \neq 0 = \mathrm{cor}((b_1 + b_2) \cdot F(x)).$$

The following example illustrates a situation where correlations of linearly dependent linear approximations of the ideal cipher cannot be statistically independent.

Example 2. Let X_1 and X_2 be binary random variables related to linear approximations of random permutation. Assume that they are statistically independent for all F and that $X_1 + X_2$ has non-zero ELP, that is,

$$\mathrm{Exp}(\mathrm{cor}(X_1 + X_2)^2) \neq 0.$$

By the piling-up lemma

$$\mathrm{Exp}(\mathrm{cor}(X_1 + X_2)\mathrm{cor}(X_1)\mathrm{cor}(X_2)) \neq 0.$$

On the other hand, $\mathrm{Exp}\,\mathrm{cor}(X_1) = 0$, from where it follows that $\mathrm{cor}(X_1)$, $\mathrm{cor}(X_2)$, and $\mathrm{cor}(X_1 + X_2)$ are not statistically independent.

However, it is not clear that such X_1 and X_2 exist. More generally, possible independence of correlations of linearly dependent linear approximations remains an open question.

6 Wrong-Key Hypotheses

In previous literature on linear cryptanalysis, single linear approximations of random permutations (ideal ciphers) were modeled as balanced random Boolean functions until it was observed in [2] that they are not strictly balanced, and should be more accurately modeled as random Boolean functions. Then for large input size n, the probability distribution of the correlation of a linear approximation can be approximated by a normal distribution with mean equal to zero and variance 2^{-n}.

Distinction between balanced and general random Boolean functions is also important for the so-called zero-correlation linear cryptanalysis [1]. If such distinction is not made the linear approximations of a cipher having zero-correlation linear approximation (for all keys) cannot be distinguished from linear approximations of a random permutation.

The distinction between random Boolean functions and balanced Boolean functions has been studied in a general setting of distinguishing between random vectorial Boolean functions and truncated Boolean permutations. It was shown in [4] that the distinguishing advantage is upperbounded by

$$\frac{q}{2^{n-\frac{m}{2}}}$$

where n is the input size, m is the output size, and q is the number of queries. This bound is also tight for larger number of queries, see [4] for details. The advantage grows exponentially as the output size m grows, which suggests that using truncated permutations to model multidimensional linear approximations, which are not permutations, is not a valid approach as the sample size (number of queries) exceeds $2^{n-\frac{m}{2}}$.

In multiple linear cryptanalysis, the information extracted from the (ideal) cipher is given in the form of correlations of a number of distinct linear approximation. A single linear approximation behaves as a random Boolean function ($m = 1$) and almost the full codebook of data is needed to distinguish it from a balanced Boolean function. It is not known what is the effect of using multiple correlations simultanously to the distinguishing advantage between a permutation and a random function. Nevertheless, modeling linear approximations of an ideal cipher as components of a random vectorial Boolean function rather than those of a permutation seems like a reasonable approach also in this case. In general, such linear approximations (similarly as components of random functions) may have pairwise statistical dependencies.

7 Conclusion

We showed that permutations are the only functions which have components such that their linear and statistical independence are equivalent concepts. We also demonstrated that different behavior can be found in the vector space of linear approximations. Further, some arguments were elaborated to support the idea that random Boolean functions provide a natural setting for modeling linear approximations of random permutation. For single linear approximations this has been known to be the case. The problem arises if more than one linear approximations are to be used simultaneously. In multidimensional linear cryptanalysis we propose to use the statistical properties of random vectorial Boolean functions to model correlations in multidimensional linear approximation of random permutations.

References

1. Andrey Bogdanov and Vincent Rijmen. Linear hulls with correlation zero and linear cryptanalysis of block ciphers. *Designs, Codes and Cryptography*, 70(3):369–383, 2014.
2. Andrey Bogdanov and Elmar Tischhauser. On the wrong key randomisation and key equivalence hypotheses in Matsui's Algorithm 2. In Shiho Moriai, editor, *Fast Software Encryption - 20th International Workshop, FSE 2013, Singapore, March 11-13, 2013. Revised Selected Papers*, volume 8424 of *Lecture Notes in Computer Science*, pages 19–38. Springer, 2013.
3. Lennart Brynielsson. A short proof of the Xiao-Massey lemma. *IEEE Trans. Inform. Theory*, IT-35(6):1344, 1989.
4. Shoni Gilboa, Shay Gueron, and Ben Morris. How many queries are needed to distinguish a truncated random permutation from a random function? *Journal of Cryptology*, 31:162–171, 2018.
5. Mitsuru Matsui. Linear Cryptanalysis Method for DES Cipher. In Tor Helleseth, editor, *Advances in Cryptology - EUROCRYPT '93, Workshop on the Theory and Application of of Cryptographic Techniques, Lofthus, Norway, May 23-27, 1993, Proceedings*, volume 765 of *Lecture Notes in Computer Science*, pages 386–397. Springer, 1993.
6. G. Z. Xiao and J. L. Massey. A spectral characterization of correlation-immune combining functions. *IEEE Trans. Inform. Theory*, IT-34(3):569–571, 1988.