Statistical and Linear Independence of Binary Random Variables

Kaisa Nyberg

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

May 19, 2017

Abstract. Linear cryptanalysis makes use of statistical models that consider linear approximations over block cipher and random permutation as binary random variables. In this note we show that linear and statistical independence are equivalent properties for linear approximations of the random permutations and the block ciphers with independent pre- and post-whitening keys.
Keywords: Xiao-Massey lemma, block cipher, linear cryptanalysis, linear approximation, random permutation, multiple linear cryptanalysis

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, the cryptanalyst builds a statistical model of the correlations of 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 purpose of this short note is to clarify the issues between linear and statistical independence of linear approximations of block ciphers and random permutations. It is clear that linearly dependent linear approximations cannot be statistically independent. On the other hand, it would be important to know, in particular for random permutations, what kind of independence assumptions hold for a set of linear approximations that is used in multiple linear cryptanalysis.

In the context of linear cryptanalysis, a linear approximation of a permutation F in $\mathbb{F}_2^n$ 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 permutations with a probability density function defined by

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

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 [5]. For a short proof, see [1].

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 $X_1, \ldots, X_k$ in $\mathcal{X}$ are linearly and statistically independent. If given $Y \in \mathcal{X}$ the variables $X_1, \ldots, X_k, Y$ are linearly independent, then they are also statistically independent.*

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

Proof. (Proof of Theorem 1) Assume first that the variables $X_1, \ldots, X_m$ in $\mathcal{X}$ are linearly independent. For $2 \leq k < m$ let us state the induction hypothesis as follows: If $X_1, \ldots, X_k$ 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 $X_1, \ldots, X_{k+1}$ be linearly independent. Then $X_1, \ldots, X_k$ are linearly independent and hence by the induction hypothesis also statistically independent. By Lemma 2 it follows that $X_1, \ldots, X_{k+1}$ are statistically independent.

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

$$X_1 = X_2 + \cdots + X_k$$

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

3 Independence of Correlations

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

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

It is clear that independence of variables implies independence of their correlations. The converse statement is not necessarily true. Next we show that in the setting of Theorem 1 also the converse holds. Moreover, we prove equivalence between linear independence of pairwise independent binary random variables and statistical independence of their correlations. We start by proving the following result.

Proposition 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. Let A be a set of elements in $\mathcal{X}$ such that $\mathbb{E}(\mathrm{cor}(X)) = 0$ and $\mathbb{E}(\mathrm{cor}(X)^2) \neq 0$ for all $X \in A$. If then the correlations of random variables in A are statistically independent, the variables are linearly independent.*

Proof. Let $A = \{X_1, \ldots, X_m\}$. To prove that the variables $X_1, \ldots, X_m$ are linearly independent, let us assume the contrary. Since their expected correlation is equal to zero, they are all nonzero. Then, as in the proof of Theorem 1 it can be assumed w.l.o.g that there exist a relation

$$X_1 = X_2 + \cdots + X_k$$

where $X_2, \ldots, X_k$ are linearly independent and $k \leq m$. Then we can use Theorem 1 to obtain that the variables $X_2, \ldots, X_k$ are statistically independent. By the Piling-up lemma [3], we then have

$$\mathrm{cor}(X_2 + \cdots + X_k) = \mathrm{cor}(X_2) \cdots \mathrm{cor}(X_k).$$

Now we consider the expectation of the product of correlations of $X_1, \ldots, X_k$ and get

$$\mathbb{E}(\mathrm{cor}(X_1) \cdots \mathrm{cor}(X_k)) = \mathbb{E}(\mathrm{cor}(X_1)^2) \neq 0.$$

On the other hand,

$$\mathbb{E}(\mathrm{cor}(X_1)) \cdots \mathbb{E}(\mathrm{cor}(X_k)) = 0,$$

from where it follows that the correlations $\mathrm{cor}(X_1), \ldots, \mathrm{cor}(X_k)$, and hence the correlations $\mathrm{cor}(X_1), \ldots, \mathrm{cor}(X_m)$, are not statistically independent. □

By combining the results of Proposition 1 with Theorem 1 we get the following corollary.

Corollary 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. Let A be a subset in $\mathcal{X}$ such that $\mathbb{E}(\mathrm{cor}(X)) = 0$ and $\mathbb{E}(\mathrm{cor}(X)^2) \neq 0$ for all $X \in A$. Then the following three conditions are equivalent.*

(i) *The variables in A are statistically independent.*
(ii) *The correlations of variables in A are statistically independent.*
(iii) *The variables in A are linearly independent.*

4 Applications

The natural applications of these results are the sets of linear approximations of block ciphers and random permutations.

It is well known that the correlation of a non-trivial linear approximation of a random permutation in $\mathbb{F}_2^n$ is normally distributed with mean 0 and variance 2^{-n} [2,4]. Moreover, it is easy to see that the linear approximations of a random permutation are pairwise independent over the space of random permutations. Hence Corollary 1 applies and we obtain that the correlations of a set of linear approximations are independent if and only if the linear approximations are linearly independent.

Corollary 2. *Let A be a set of non-trivial linear approximations of a random permutation. Then the following conditions are equivalent:*

(i) *The correlations of the linear approximations are independent.*
(ii) *The correlations of the linear approximations in A are jointly normally distributed.*
(iii) *The linear approximations in A are linearly independent.*

The equivalence of (i) and (ii) follow from the fact that normally distributed pairwise independent random variables are jointly normally distributed if and only if they are independent.

It is straightforward to show that block ciphers that have independent prewhitening and post-whitening keys have the property that the linear approximations are pairwise independent over the key space. Moreover, their correlations have zero mean and nonzero variance. A special case of such ciphers are the so-called long-key ciphers, which are iterated key-alternating block ciphers with independent round keys. By Theorem 1, the correlations of a set of linear approximations over such a cipher are independent over the key space if and only if these linear approximations are linearly independent.

References

1. Lennart Brynielsson. A short proof of the xiao-massey lemma. *IEEE Trans. Inform. Theory*, IT-35(6):1344, 1989.
2. Joan Daemen and Vincent Rijmen. Probability distributions of correlation and differentials in block ciphers. *J. Mathematical Cryptology*, 1(3):221–242, 2007.
3. 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.
4. Luke O'Connor. Properties of linear approximation tables. In Bart Preneel, editor, *FSE 1994*, volume 1008 of *LNCS*, pages 131–136. Springer, Heidelberg, 1995.
5. 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.