Aalto University
School of Science

Linear and Statistical Independence of Linear Approximations and their Correlations

Kaisa Nyberg

Aalto University School of Science
kaisa.nyberg@aalto.fi

Boolean Functions and their Applications
Os, Norway, July 2017

Outline

Outline

Independence

- Random variables $Z_1, \ldots, Z_k$ are statistically independent if

$$\Pr(Z_1 = z_1, \ldots, Z_k = z_k) = \Pr(Z_1 = z_1) \cdots \Pr(Z_k = z_k)$$

  for all $z_1, \ldots, z_k$ in the value spaces
- If random variables $Z_1, \ldots, Z_k$ are statistically independent then

$$\mathbb{E}(Z_1 \cdots Z_k) = \mathbb{E}(Z_1) \cdots \mathbb{E}(Z_k)$$

- Binary random variables $X_1, \ldots, X_k$ are linearly independent if

$$\lambda_1 X_1 + \cdots + \lambda_k X_k \neq 0$$

  for every choice of $\lambda_1, \ldots, \lambda_k$, not all zero, in $\mathbb{F}_2$.

Clearly, linear dependence (of non-zero variables) implies statistical dependence.

In general, the converse statement is not true.

This talk: For a certain class of binary random variables linear independence guarantees statistical independence.

Background

- Biryukov et al. 2004: Model for multiple linear cryptanalysis developed under the assumption that the linear approximations are statistically independent, and hence, they must be linearly independent
- Linear independence often seen as hurdle preventing from using the best approximations
- Hermelin et al. 2009 presented multidimensional linear cryptanalysis to overcome the assumption of statistical independence: linear approximations form a linear space.
  - Disadvantage: also weak linear approximations included
- In practice, multiple linear approximations (derived from the cipher) have been found to follow the model even if they are not linearly independent and the independence assumption is often ignored.

Motivation

- Distinguishing attack (the basis for key recovery in iterated block ciphers) uses a statistical model for the practical cipher, and the alternative object is modelled to follow random behavior
- Independence assumptions, if required by the model, should be satisfied, in partcular, for the random case
- It is too easy to distinguish from random something coming from the cipher that is not random even in the random world
- To satisfy statistical independence the linear approximations must be linearly independent.
- Is linear independence enough? No, not in general.
- Yes, in a linear space of pairwise independent variables.

Outline

Xiao-Massey Lemma

Presented by Xiao and Massey 1988 in the context of correlation-immune functions. A short proof was presented by Brynielsson in 1989 (both in IEEE Trans of IT).

Lemma

(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 $\mathbb{F}_2$.

Outline

Main Result

Theorem

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 $\mathcal{X}$.

Outline of the Proof

By induction. Main step:

Lemma

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 $X_1, \ldots, X_k, Y$ are statistically dependent $\Rightarrow$
Y is dependent of $X_1, \ldots, X_k$.
Then Xiao-Massey lemma $\Rightarrow$
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.
Both Y and the sum are in $\mathcal{X} \Rightarrow Y = \lambda_1 X_1 + \cdots + \lambda_k X_k$. □

Statistical Independence of Correlations

Correlation of X

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

Proposition

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

$$\mathbb{E}(\mathrm{cor}(X)) = 0 \textit{ and } \mathbb{E}(\mathrm{cor}(X)^2) \neq 0$$

for all $X \in A$. If then the correlations of random variables in A are statistically independent, the variables are statistically independent and hence also linearly independent.

That is, we cannot have independence of correlations unless the variables are linearly independent.

Proof is based on the piling-up lemma.

Summarizing

Corollary

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.*

Outline

Applications

$\mathcal{X}$ the linear space of linear approximations

A subset in $\mathcal{X}$ used in an attack

χ^2 distinguisher

- builds statistical models: one for random and one for cipher, and
- defines a χ^2 test statistic
  1. by summing (non-trivial) empirical squared correlations over a linear subspace of linear approximations (multidimensional)
  2. by summing independent empirical squared correlations of individual linear approximations
  3. by combination of independent, type 1 and/or type 2, χ^2 statistics, e.g., from direct sums of linear spaces of linear approximations related to parallel S-boxes.

Checking Validity of Assumptions

Random

Two different linear approximations of a random permutation are statistically independent

$\mathbb{E}(\text{cor}(X)) = 0$ and $\mathbb{E}(\text{cor}(X)^2) = 2^{-n} \neq 0$

Long-key Cipher

Iterated block ciphers with independent round keys are pairwise independent, and

$\mathbb{E}(\text{cor}(X)) = 0$ and $\mathbb{E}(\text{cor}(X)^2) = ELP \neq 0$

Other Ciphers

Assumptions to be checked and tested on reduced versions

Outline

Conclusions

- Natural necessary and sufficient conditions under which correlations of linear approximations are statistically independent
- For example, correlations of linear approximations of a random cipher are statistically independent if and only if the linear approximations are linearly independent
- Our observations are particularly useful for getting the model for the random cipher correct.