Publication VII

Miia Hermelin and Kaisa Nyberg. 2010. Dependent linear approximations: The algorithm of Biryukov and others revisited. In: Josef Pieprzyk (editor). Topics in Cryptology. Proceedings of the Cryptographers' Track at the RSA Conference 2010 (CT-RSA 2010). San Francisco, California, USA. 1-5 March 2010. Berlin, Heidelberg, Germany. Springer. Lecture Notes in Computer Science, volume 5985, pages 318-333. ISBN 978-3-642-11924-8.

http://www.springerlink.com/openurl.asp?genre=article&id=doi:10.1007/978-3-642-11925-5_22

Dependent Linear Approximations: The Algorithm of Biryukov and Others Revisited

Miia Hermelin[1] and Kaisa Nyberg[1,2]

[1] Aalto University, School of Science and Technology
[2] Nokia, Finland

Abstract. Biryukov, et al., showed how it is possible to extend Matsui's Algorithm 1 to find several bits of information about the secret key of a block cipher. Instead of just one linear approximation, they used several linearly independent approximations that were assumed to be statistically independent. Biryukov, et al., also suggested a heuristic enhancement to their method by adding more linearly and statistically dependent approximations.

We study this enhancement and show that if all linearly dependent approximations with non-negligible correlations are used, the method of Biryukov, et al., is the same as the convolution method presented in this paper. The data complexity of the convolution method can be derived without the assumption of statistical independence. Moreover, we compare the convolution method with the optimal ranking statistic log-likelihood ratio, and show that their data complexities have the same order of magnitude in practice. On the other hand, we show that the time complexity of the convolution method is smaller than for the other two methods.

Keywords: Matsui's Algorithm 1, linear cryptanalysis, multidimensional cryptanalysis, method of Biryukov, convolution method.

1 Introduction

Linear cryptanalysis of block ciphers makes use of probabilistic relations between the plaintext and ciphertext data and the secret key. Such a relation is called a linear approximation of the block cipher. Given a sufficient amount of data derived from the cipher, Matsui's Algorithm 1 [1] can be used in recovering one bit of information about the secret key.

First, Kaliski and Robshaw [2] showed that by using multiple linear approximations, the data complexity can be reduced and later, Biryukov, et al., [3] that multiple bits of information about the secret key can be obtained. However, these methods rely on the assumption that the linear approximations used in the attack are statistically independent. Murphy noted that this is not true in general [4]. Hermelin, et al., investigated this problem in practice using a reduced round Serpent and showed that strong linear approximations are not usually statistically independent [5].

It was observed already in [3] that including more strong linear approximations seemed only to improve the results even if the used approximations were neither linearly nor statistically independent. The practical experiments performed in [5] also showed that when using multiple linear approximations the larger the number of strong approximations was in the method of Biryukov, et al., the closer the observed data complexity became to the data complexities of the methods based on χ^2 and the Kullback-Leibler distance [5].

J. Pieprzyk (Ed.): CT-RSA 2010, LNCS 5985, pp. 318–333, 2010.

These observations suggest that the assumption about statistical independence of the linear approximations could and should be relaxed when applying in practice the method presented Biryukov, et al., which we will call the Biryukov method, for brevity. In this paper we give theoretical justification that this is really the case. For this purpose, we investigate the Biryukov method in the case, where the set of linear approximations is the full linear span of the given set of linear approximations. Completed in this manner the method can be shown to be equivalent to a new method, which we will call the convolution method. The convolution method is interesting, first because it does not rely on the assumption about statistical independence. Secondly, it has the same time complexity as the Biryukov method would have if only the linearly independent approximations are used. Thirdly, the data complexity of the convolution method is at most the same as the data complexity of the Biryukov method.

Previously, the log-likelihood ratio (LLR) was used in [6] for realising another Algorithm 1 type linear attack. In this work we also compare the convolution method and the LLR-method in theory by modelling the problem of finding the correct key information bit as a multiple hypothesis testing problem. While the LLR is the optimal solution with the smallest data complexity, the data complexity of the convolution method is of the same order of magnitude. The key ranking problem in the Algorithm 1 type attacks is also investigated and the existing approaches are compared.

The structure of this paper is as follows: In Sect. 2, some basic notation is given. The linear approximation of a block cipher and the basic Biryukov method is studied in Sect. 3. Section 3.3 studies the completed Biryukov method and presents the convolution method. Statistical analysis of the convolution method is done in Sect. 5. It is shown that the convolution method or the completed Biryukov method do not require the assumption about statistical independence. Section 6 studies the data, time and memory complexities for convolution method, the completed Biryukov method Biryukov and LLR-method.

2 Probability Distributions and Boolean Functions

The space of n-dimensional binary vectors is denoted by $\mathbb{Z}_2^n$. The sum modulo 2 is denoted by $\oplus$. The inner product for $a = (a^1, \dots, a^n), b = (b^1, \dots, b^n) \in \mathbb{Z}_2^n$ is defined as $a \cdot b = a^1 b^1 \oplus \cdots \oplus a^n b^n$. Then the vector a is called the (linear) mask of b. The Hamming weight w_H of a binary vector $a \in \mathbb{Z}_2^n$ is $w_H(a) = \#\{i = 1, \dots, n : a^i = 1\}$, the number of non-zero components in a.

A function $f : \mathbb{Z}_2^n \mapsto \mathbb{Z}_2$ is called a Boolean function. A linear Boolean function is a mapping $x \mapsto u \cdot x$. A function $f : \mathbb{Z}_2^n \mapsto \mathbb{Z}_2^m$ with $f = (f_1, \dots, f_m)$, where f_i are Boolean functions, is called a vector Boolean function of dimension m. A linear Boolean function from $\mathbb{Z}_2^n$ to $\mathbb{Z}_2^m$ is represented by an $m \times n$ binary matrix U. The m rows of U are denoted by $u_1, \dots, u_m$, where each u_i is a linear mask.

The correlation between a Boolean function $f : \mathbb{Z}_2^n \mapsto \mathbb{Z}_2$ and zero is

$$c(f) = c(f, 0) = 2^{-n} \left(\#\{x \in \mathbb{Z}_2^n : f(x) = 0\} - \#\{x \in \mathbb{Z}_2^n : f(x) \neq 0\}\right)$$

and it is also called the correlation of f.

We denote random variables $\mathbf{X}, \mathbf{Y}, \ldots$ by capital boldface letters, their domains by $\mathcal{X}, \mathcal{Y}, \ldots$ and their realisations $x \in \mathcal{X}, y \in \mathcal{Y}, \ldots$ by small letters. Let $\mathbf{X}$ be a random variable taking on values in $\mathcal{X} = \{0, 1, \ldots, M\}$. The discrete probability distribution (p.d.) of $\mathbf{X}$ is vector a $p = (p_0, \ldots, p_M)$ if $\Pr(\mathbf{X} = \eta) = p_\eta$, for all $\eta \in \mathcal{X}$. Then we denote $\mathbf{X} \sim p$. We denote the uniform p.d. by θ.

Let $f : \mathbb{Z}_2^n \mapsto \mathbb{Z}_2^m$ and $\mathbf{X} \sim \theta$, where $\mathbf{X}$ takes on values in $\mathbb{Z}_2^n$. If $\mathbf{Y} = f(\mathbf{X})$, then the p.d. of $\mathbf{Y}$ is called the p.d. of f and we say that the random variable $\mathbf{Y}$ is associated with f. Let $f_1, \ldots, f_m : \mathbb{Z}_2^n \mapsto \mathbb{Z}_2^m$ be Boolean functions and for each f_i the associated random variable is $\mathbf{Y}_i$. Then we say that the Boolean functions $f_1, \ldots, f_m$, are statistically independent (s.i.), if the random variables $\mathbf{Y}_1, \ldots \mathbf{Y}_m$, are s.i.

3 Multidimensional Matsui's Algorithm 1

3.1 Linear Approximation of a Block Cipher

Let f be an encryption function of a block cipher with block size n. We denote by x the plaintext, by K the expanded key, that is, a vector consisting of all (fixed) round key bits and by $y = f(x, K)$ the ciphertext. Then an m-dimensional linear approximation of the block cipher is a vector Boolean function

$$\mathbb{Z}_2^n \times \mathbb{Z}_2^n \to \mathbb{Z}_2^m,\ (x, y) \mapsto Ux \oplus Wy \oplus VK, \tag{1}$$

where U and W are $m \times n$ binary matrices and the modulo 2 addition $\oplus$ is calculated component-wise for the vectors. The matrix V has also m rows and it divides the expanded keys, and therefore also the keys, to 2^m equivalence classes $z = VK \in \mathbb{Z}_2^m$. The task is to find the right inner key class, denoted by z_0.

The most complex task in linear cryptanalysis is to determine the p.d. p of the Boolean function (1). A method for determining an approximation p given the biases of $2^m - 1$ one-dimensional linear approximations related to (1) was presented in [5]. We will henceforth assume that a good approximation of the p.d. p of (1) is available.

We make the usual assumption that the plaintexts $x_1, \ldots, x_N$, are the realised values of N independent and identically distributed (i.i.d.) random variables, each following the uniform distribution. Then for all $t = 1, \ldots, N$, the observed values $Ux_t \oplus Wy_t \oplus z$, $z \in \mathbb{Z}_2^m$, are realisations of i.i.d. random variables following p. Hence, for each $z \in \mathbb{Z}_2^m$, the values $Ux_t \oplus Wy_t$, $t = 1, \ldots, N$, are the realisations of i.i.d. random variables following p^z, a fixed permutation of p determined by z. Then all the p.d.'s p^z, $z \in \mathbb{Z}_2^m$, are each other's permutations, and in particular,

$$p^z_{\eta \oplus a} = p^{z \oplus a}_\eta, \quad \text{for all} \quad z, \eta, a \in \mathbb{Z}_2^m. \tag{2}$$

The goal of Alg. 1. is to determine z_0 using the empirical data of N plaintext-ciphertext pairs (x_t, y_t), $t = 1, \ldots, N$. For each key $z \in \mathbb{Z}_2^m$ we give a mark defined by $F(z) = T((x_1, y_1), \ldots, (x_N, y_N); z)$, where T is a suitable ranking statistics with data as the variable [7] [8]. The key z is a parameter of T. Given the data, the keys are ordered in increasing or decreasing order according to their marks $F(z)$. The key z' with the highest mark is chosen to be the right key candidate. The error probability $\Pr(z' \neq z_0)$ should decrease if the amount of data N is increased. The best statistics gives the smallest error for a given N. The ranking statistic proposed by Biryukov, et al., is described in the next section.

3.2 Method of Biryukov, et al.

The basic version of the Biryukov method uses m linearly independent approximations $u_i \cdot x \oplus w_i \cdot y \oplus v_i \cdot K, i = 1, \ldots, m$, where the ith approximation has a non-negligible correlation c_i. Biryukov, et al., assumed that the approximations are s.i., that is, if $\mathbf{X}_i$ is a binary random variable associated with the ith approximation $u_i \cdot x \oplus w_i \cdot y \oplus z$, then the random variables $\mathbf{X}_1, \ldots, \mathbf{X}_m$, are s.i.

For each $i = 1, \ldots, m$, let ρ_i denote the empirical correlation of the ith approximation calculated using the data (x_t, y_t), $t = 1, \ldots, N$ as follows:

$$\rho_i = 2N^{-1}\{t = 1, \ldots, N : u_i \cdot x_t \oplus w_i \cdot y_t = 0\} - 1.$$

Denote $z = (z^1, \ldots, z^m)$ such that z^i is the ith bit of the key z. Denote the theoretical and empirical correlation vectors by $\mathbf{c}_z = ((-1)^{z^1} c_1, \ldots, (-1)^{z^m} c_m)$ and $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_m)$, respectively. The mark for each $z \in \mathbb{Z}_2^m$ is given by the ℓ_2 distance between the two correlation vectors:

$$b(z) = ||\mathbf{c}_z - \boldsymbol{\rho}||_2^2 .$$

The key z' minimising $b(z)$ is chosen to be the right key.

Later Murphy noted that the assumption about statistical independence of the linear approximations does not hold in general [4]. In particular, linearly dependent approximations are also statistically dependent. Murphy also suggested to use the traditional measure of covariance of two linear approximations in verifying the assumption about linear independence. This method has been subsequently used by other researchers, for example in [9]. The most natural way is to use the converse of the Piling Up lemma [1], which we give in the Appendix 7.

Biryukov, et al., proposed a heuristic enhancement to their method [3]. They added approximations that were linearly dependent of the m original approximations. Ultimately, they could use all $2^m - 1$ one-dimensional approximations in the span of the original approximations. We call this method the full Biryukov method and we will study it in the next section.

3.3 The Full Biryukov Method

In this method, the empirical correlation $\rho(a)$ for each $a \in \mathbb{Z}_2^m$ is calculated using the data (x_t, y_t), $t = 1, \ldots N$ as follows:

$$\rho(a) = 2N^{-1}\{t = 1, \ldots, N : Ux_t \oplus Wy_t \oplus = 0\} - 1$$

The ηth component of the theoretical correlation vector $\mathbf{c}_z$ is now $(-1)^{\eta \cdot z} c(\eta)$ and the vector of empirical correlations is $\boldsymbol{\rho} = (\rho(0), \ldots, \rho(2^m - 1))$. Similarly to the basic version, the mark is given by

$$B(z) = ||\mathbf{c}^z - \boldsymbol{\rho}||_2^2 = \sum_{a \in \mathbb{Z}_2^m} ((-1)^{a \cdot z} c(a) - \rho(a))^2$$

and the key z' that minimises $B(z)$ is chosen to be the right key.

Next we analyse this full method. Our analysis is based on the observation that there exists another statistic which is equivalent to the $B(z)$ statistic, in the sense that both will produce exactly the same key ranking. Moreover, this equivalent statistic gives a more efficient way of ranking the candidate keys, and in particular, to determine the most likely key candidate.

4 Convolution Method

We now show how to make the full Biryukov method more efficient in practice. We obtain the empirical distribution $q = (q_0, \dots, q_{2^m-1})$ of the multidimensional approximation $Ux \oplus Wy \oplus VK$ by computing

$$q_\eta = N^{-1}\#\{t = 1, \dots, N : Ux_t \oplus Wy_t = \eta\}, \quad \text{for all} \quad \eta \in \mathbb{Z}_2^m. \tag{3}$$

The mark $B(z)$ of the full Biryukov method can also be written as

$$B(z) = -2 \sum_{a \in \mathbb{Z}_2^m} (-1)^{a \cdot z} c(a)\rho(a) + \sum_{a \in \mathbb{Z}_2^m} (\rho(a)^2 + c(a)^2),$$

where the latter sum does not depend on z. On the other hand, by equation (3) in [10], we have

$$c(a) = \sum_{\eta \in \mathbb{Z}_2^m} (-1)^{a \cdot \eta} p_\eta \quad \text{and} \quad \rho(a) = \sum_{\eta \in \mathbb{Z}_2^m} (-1)^{a \cdot \eta} q_\eta.$$

Using the previous formulas for correlations we have

$$\sum_{a \in \mathbb{Z}_2^m} (-1)^{a \cdot z} c(a)\rho(a) = 2^m \sum_{\eta \in \mathbb{Z}_2^m} q_\eta p_{\eta \oplus z}. \tag{4}$$

But the sum is just the zth component of the convolution $q * p$ of the p.d.'s p and q. Hence, finding the minimum of $B(z)$ is equivalent to finding the maximum of the zth component of the convolution of q and p, that is, z is the mode of the p.d. $q * p$. We now propose the following mark

$$G(z) = (p * q)_z, \tag{5}$$

and the key z' that maximises $G(z)$ is chosen to be the right key. We call this new method based on $G(z)$ the convolution method. We have the following result.

Theorem 1. *The key z' minimises $B(z)$ if and only if it maximises $G(z)$. Hence, the full Biryukov method and the convolution method are equivalent.*

Both methods are also equivalent to the maximum likelihood decoding. The problem is to decode the code where the channel has error probability distribution p and the original message is $z \in \mathbb{Z}_2^m$. The message is sent N times over the channel with noise $Ux_t \oplus Wy_t \sim p^z$, at each time $t = 1, \dots, N$. The receiver obtains sequence $z \oplus Ux_t \oplus Wy_t, t = 1, \dots, N$, with observed empirical p.d. q that should approximate p. Then $q * p^z$ gives an empirical p.d. for $z = (Ux_t \oplus Wy_t) \oplus (Ux_t \oplus Wy_t \oplus z)$ and the key candidate z is given as the mode of the p.d. $q * p^z$.

While the two methods have the same data complexities, the convolution method has smaller time complexity. The basic and full Biryukov methods have time complexities $m2^m$ and 2^{2m}, respectively. This is because we have to compute the rank $b(z)$ or $B(z)$, respectively, for each $z \in \mathbb{Z}_2^m$. In the convolution method we do not have to consider each key or p.d. p^z separately. It suffices to compute only one convolution $p * q$ and determine its mode. The convolution is computed using FFT with time complexity $m2^m$. Hence, with the same data the convolution method outputs the same key class as the full Biryukov method, but the time complexity for the convolution method is the same as for the basic Biryukov method. In [6] Hermelin, et al., studied the optimal method based on the LLR-statistic. We prove in the next section that the data complexities of the convolution method and the LLR-method are approximately equal.

More accurate descriptions for the algorithms for the different methods are given in Section 6.2. In the next section, we study the statistical properties of the convolution method.

5 Statistical Analysis

Finding the right key z_0 is actually a multiple hypothesis testing problem. Section 5.2 studies the problem and how to solve it. The next section gives some necessary theory about discrete random variables and multinomial probability distributions needed in multiple hypothesis testing problems.

5.1 Multinomial Distribution

Let $\mathbf{X}_1, \ldots, \mathbf{X}_N$, be i.i.d. random variables drawn from space $\mathcal{X} = \{0, 1, \ldots, M\}$ by a discrete p.d. $s = (s_0, \ldots, s_M)$, where M is some positive integer. Let $\mathbf{Q} = (\mathbf{Q}_0, \ldots, \mathbf{Q}_M)$ be a vector of random variables where for each $\eta \in \mathcal{X}$,

$$\mathbf{Q}_\eta = N^{-1}\#\{i = 1, \ldots, N : X_i = \eta\}. \tag{6}$$

Hence, $\mathbf{Q}$ is a vector of relative frequencies of the elements of the sample space $\mathcal{X}$. The sample space $\mathcal{Q}$ of $\mathbf{Q}$ consists of vectors $q = (q_0, \ldots, q_M)$, where $q_0, \ldots, q_M \in N^{-1}\{0, 1, \ldots, N\}$ and $q_0 + \cdots + q_M = 1$. The random vector $\mathbf{Q}$ follows the multinomial distribution $\mathrm{Multi}(N, s)$, with probabilities

$$\Pr(\mathbf{Q} = q) = \frac{N!}{\prod_{\eta=0}^{M}(q_\eta N)!}\prod_{\eta=0}^{M} s_\eta^{Nq_\eta}, \quad \text{for all} \quad q \in \mathcal{Q}. \tag{7}$$

Since for each $z \in \mathbb{Z}_2^m$, the observed values $Ux_t \oplus Wy_t$, $t = 1, \ldots, N$ are realisations of i.i.d. random variables following p^z, the empirical p.d. q calculated using (3) is a realisation of a random vector $\mathbf{Q}$ that has multinomial distribution $\mathrm{Multi}(N, p^z)$. Using (2), we have for all $z \in \mathbb{Z}_2^m$,

$$(p * \mathbf{Q})_z = \sum_{\eta \in \mathbb{Z}_2^m} p_{\eta \oplus z}\mathbf{Q}_\eta = \sum_{\eta \in \mathbb{Z}_2^m} p_\eta^z \mathbf{Q}_\eta. \tag{8}$$

Hence, maximising $G(z)$ in (5) is equivalent to finding $z' \in \mathbb{Z}_2^m$ that maximises

$$\sum_{\eta \in \mathbb{Z}_2^m} p_\eta^z q_\eta. \tag{9}$$

By (8) the convolution method has the same statistical behaviour as the method using (9). The next lemma gives the distribution of (8).

Lemma 1. *Let $\lambda_0, \ldots, \lambda_M$ be any real numbers and $\mathbf{Q} = (\mathbf{Q}_0, \ldots, \mathbf{Q}_M)$ be a multinomially distributed random vector with distribution* $\mathrm{Multi}(N, s)$. *Then the linear combination $N \sum_{\eta=0}^{M} \lambda_\eta \mathbf{Q}_\eta$ is asymptotically normal with mean and variance given by*

$$\mu = N \sum_{\eta=0}^{M} \lambda_\eta s_\eta \qquad \sigma^2 = N \sum_{\eta=0}^{M} \lambda_\eta^2 s_\eta - \mu^2.$$

The proof is given in Appendix 7. Since the lemma does not require the assumption about statistical independence, the assumption is also not needed when using full Biryukov or convolution method.

The concept of capacity was introduced in [5] and it was used in simplifying the formulas of the data complexities:

Definition 1. *The capacity between two p.d.'s $p = (p_0, \ldots, p_M)$ and $q = (q_0, \ldots, q_M)$ is defined by*

$$C(p, q) = \sum_{\eta=0}^{M} (p_\eta - q_\eta)^2 q_\eta^{-1}.$$

If q is the uniform distribution, we denote $C(p, q) = C(p)$.

5.2 Multiple Hypothesis Testing Problems

Let $\mathbf{X}_1, \ldots, \mathbf{X}_N$, be a sequence of i.i.d. random variables drawn from sample space $\mathcal{X} = \{0, 1, \ldots, M\}$, where M is a positive integer, and let $x_1, \ldots, x_N$, be the corresponding realisations. Assume $d \geq 2$ simple hypotheses, where each hypothesis H_i states that the sample is drawn according to a p.d. $p^i = (p_0^i, \ldots, p_M^i)$, $i = 1, \ldots, d$, and $p^i \neq p^j$, if $i \neq j$. Equivalently, each hypothesis H_i states that the vector $\mathbf{Q}$ defined by (6) is multinomial distributed as $\mathrm{Multi}(N, p^i)$.

The simple d-ary hypothesis testing problem is to determine which hypothesis is correct. Hence, one hypothesis is accepted and the others are rejected. In Bayesian statistics, each hypothesis is given an *a priori* probability $\Pr(H_i)$ for all $i = 1, \ldots, d$. We assume that the *a priori* probabilities are equal.

Let $q = (q_0, \ldots, q_M)$ be the empirical p.d. calculated from the observed values $x_1, \ldots, x_N$, by

$$q_\eta = N^{-1} \#\{t = 1, \ldots, N : x_t = \eta\}, \quad \text{for all} \quad \eta \in \mathcal{X}.$$

A distinguisher is a rule that based on the observed data $x_1, \ldots, x_N$, or, equivalently, q, outputs which hypotheses is accepted:

$$\delta(x_1, \ldots, x_N) = \delta(q) = i, \quad \text{if} \quad H_i \quad \text{is accepted, for} \quad i = 1, \ldots, d$$

The distinguisher is defined using a suitable test statistic $T(q; p^i)$, where p^i (or i) is considered as the parameter and q is the variable.

Let $f(i) = T(q; p^i)$ be a function of the parameter i for given empirical data q. The distinguisher outputs j if it gives the maximum (or minimum) of $f(i)$, for given q. The statistic T should be easy to compute in practice and accurate such that the total error

$$P_e = \sum_{i=1}^{d} \Pr(H_i) \Pr(\delta(\mathbf{Q}) \neq i \mid H_i) \tag{10}$$

is as small as possible. An optimal distinguisher minimising the error probability exists for simple hypotheses testing problems.

Consider first the simple binary hypothesis testing problem with $d = 2$. By Neyman-Pearson lemma in classical statistics and Chernoff's theorem in Bayesian statistics [11], the optimal distinguisher for distinguishing between H_1 and H_2, or p^1 and $p^2 \neq p^1$, equivalently, is given by the log-likelihood ratio (LLR) test statistic

$$\mathrm{LLR}(q; p^1, p^2) = \sum_{\eta \in \mathcal{X}} N q_\eta \log \frac{p^1_\eta}{p^2_\eta}.$$

The distinguisher accepts H_1, that is, outputs p^1 (or accepts H_2 and outputs p^2, respectively) if $\mathrm{LLR}(q; p^1, p^2) \geq \tau$ ($< \tau$) where τ is the threshold that depends on P_e. Obviously, using LLR is the same as finding for given q the maximum of the function

$$l(i) = \sum_{\eta \in \mathcal{X}} q_\eta \log p^i_\eta,\ i = 1, 2.$$

If $p^1, p^2 \neq \theta$ this is equivalent to finding the maximum of

$$L(i) = l(i) + \log(M + 1) = \mathrm{LLR}(q, p^i, \theta),\ i = 1, 2.$$

In Bayesian theory Chernoff's theorem [11] states that $P_e = \mathcal{O}\left(2^{-ND^*(p^1,p^2)}\right)$, where $D^*(p^1, p^2)$ is the Chernoff information between p^1 and p^2 given by

$$D^*(p^1, p^2) = - \min_{0 \leq \lambda \leq 1} \log \left(\sum_{\eta=0}^{M} (p^1_\eta)^\lambda (p^2_\eta)^{1-\lambda} \right). \tag{11}$$

Assume now a d-ary hypothesis testing problem with $d \geq 3$ simple hypotheses. Moreover, assume that $p^i \neq \theta$ for all $i = 1, \ldots, d$. The optimal distinguisher that minimises P_e chooses the hypothesis with the largest conditional probability $\Pr(H_i \mid \mathbf{Q} = q)$, see [12]. Equivalently, by Bayes' theorem, the distinguisher chooses the hypothesis that maximises $\Pr(\mathbf{Q} = q \mid H_i)$.

Consider the likelihood function $\mathcal{L}(p^i) = \Pr(\mathbf{Q} = q \mid H_i)$ that should reach its maximum for the right p.d. p^i, given data q. Using the formula (7) of the p.d. of the multinomial distribution the likelihood function can be written as

$$\mathcal{L}(p^i) = \frac{N!}{\prod_{\eta=0}^{M} (q_\eta N)!} \prod_{\eta=0}^{M} (p^i_\eta)^{N q_\eta}.$$

Taking logarithm and omitting the terms not depending on p^i gives an equivalent test statistics

$$L(i) = \sum_{\eta \in \mathcal{X}} q_\eta \log p^i_\eta + \log(M+1) = N^{-1} \operatorname{LLR}(q, p^i, \theta). \tag{12}$$

Hence, LLR-statistics gives the optimal distinguisher for a multiple hypothesis testing problem for $d \geq 3$, also. The LLR measures whether the data is drawn from p^i or the uniform distribution. High values imply that the data q is closer to p^i than θ. Hence, we have a theoretical justification for the heuristic LLR-method presented in [6].

Both convolution method and the LLR have the form of a general linear method [7] using the statistic

$$T(\mathbf{Q}; z) = N \sum_{\eta \in \mathcal{X}} \lambda^z_\eta \mathbf{Q}_\eta,$$

where the coefficients $\lambda^z_0, \ldots, \lambda_M$, depend on the parameter z. Comparing the coefficients in the formulas (9) and (12) shows that the LLR-method and convolution method are not equivalent. Hence, the convolution method is not optimal in theory.

Consider the definition (10) of the error probability when distinguishing $d \geq 3$ hypothesis. Each term $\Pr(\delta(\mathbf{Q}) \neq i \mid H_i)$ in the sum is equal to

$$\Pr(\delta(\mathbf{Q}) \neq i \mid H_i) = \sum_{j \neq i, j=1,\ldots,d} \Pr(\delta(\mathbf{Q}) = j \mid H_i).$$

But each probability $\Pr(\delta(\mathbf{Q}) = j \mid H_i)$ corresponds to the binary hypothesis testing problem of distinguishing parameter i from $j \neq i$. Hence, if for given P_e two distinguishers have same data complexity for the binary hypothesis testing problem, then they are also equally efficient in the multiple hypothesis testing setting.

It remains to show that for a given error probability, if the p.d. p is nearly uniform (but not uniform), then the data complexity of the convolution method is of the same order of magnitude as the data complexity of the LLR-method. We study the complexities in the next section.

6 Complexity Analysis

6.1 Data Complexity

To compare the LLR and convolution methods, we have to calculate the data complexity N for given error probability P_e. We know by Sect. 5.2 that the LLR-method is optimal, i.e., for given P_e it has the smallest data complexity. However, based on the tests made in [5] and [13], we suspect that the data complexities of the convolution method and the LLR-method are practically the same as long as the p.d.'s do not variate much from the uniform distribution. More accurately, we assume that there exists ϵ, $0 < \epsilon < 0.5$ such that each p.d. p^z, $z \in \mathbb{Z}_2^m$, satisfies the following conditions:

$$\begin{aligned} &|p^z_\eta - 2^{-m}| \leq \epsilon 2^{-m} \quad \text{for all} \quad z, \eta \in \mathbb{Z}_2^m \quad \text{and} \\ &|p^{z_1}_\eta - p^{z_2}_\eta| \leq \epsilon p^{z_2}_\eta \quad \text{for all} \quad z_1 \neq z_2 \quad \text{and} \quad z_1, z_2, \eta \in \mathbb{Z}_2^m. \end{aligned} \tag{13}$$

Then for all $z, z_1, z_2 \in \mathbb{Z}_2^m$ the capacities $C(p^z) = C(p) = \epsilon^2 < 1$ and $C(p^{z_1}, p^{z_2}) = \epsilon^2 < 1$, if $z_1 \neq z_2$. The condition (13) holds for all practical ciphers. For example the experiments with reduced round Serpent in [8] showed that for $m \leq 12$, the condition held with the parameter value $\epsilon \approx 1/150$. In general, the value ϵ should be so small that it is possible to approximate the Chernoff information $D^*(p^{z_1}, p^{z_1})$ between two distinct distributions p^{z_1} and p^{z_2} using their capacity: $D^*(p^{z_1}, p^{z_2}) \approx (8 \ln 2)^{-1} C(p^{z_1}, p^{z_2})$, see Theorem 7 in [14].

As noted in the previous section, we only have to consider the distinguishing between two keys z_1 and $z_2 \neq z_1$. Denote for simplicity $p = p^{z_1}$ and $s = p^{z_2}$. If the p.d.'s satisfy condition (13), then by definition (11), the data complexity of the LLR-method is proportional to

$$N = C(p, s)^{-1}. \tag{14}$$

See also [15] for another proof. We now show that (14) holds also for the convolution method, provided that the distributions p and s satisfy condition (13).

The cumulative distribution function of the normed, normal distribution is

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt\,.$$

By Lemma 1 we obtain that the probability of choosing $z_2 \neq z_1$ when H_{z_1} is true is

$$\Pr(\delta(\mathbf{Q}) = y \mid H_{z_1}) = \Pr(T(\mathbf{Q}; y) > T(\mathbf{Q}; z) \mid H_{z_1}) = \Phi\left(\sqrt{N}\frac{\mu}{\sigma}\right),$$

where the expected value μ and variance σ^2 are given by

$$\mu = \sum_{\eta \in \mathbb{Z}_2^m} (p_\eta - s_\eta) p_\eta \qquad \sigma^2 = \sum_{\eta \in \mathbb{Z}_2^m} (p_\eta - s_\eta)^2 p_\eta - \mu^2.$$

The mean μ can be approximated by

$$\mu \approx 2^{-m} \sum_{\eta \in \mathbb{Z}_2^m} (p_\eta - s_\eta) \frac{p_\eta}{s_\eta} = 2^{-m} \sum_{\eta \in \mathbb{Z}_2^m} \left((p_\eta - s_\eta) \frac{p_\eta}{s_\eta} - (p_\eta - s_\eta) \right) = 2^{-m} C(p, s).$$

Moreover,

$$\sum_{\eta \in \mathbb{Z}_2^m} (p_\eta - s_\eta)^2 p_\eta = \sum_{\eta \in \mathbb{Z}_2^m} \frac{(p_\eta - s_\eta)^2}{s_\eta} p_\eta s_\eta \approx 2^{-2m} C(p, s). \tag{15}$$

As $C(p, s) < 1$, the dominating term of σ^2 is given by (15). Hence, $\sigma^2 \approx 2^{-2m} C(p, s)$ and the data complexity is proportional to

$$N = \frac{2^{-2m} C(p, s)}{2^{-2m} C(p, s)^2} = C(p, s)^{-1}.$$

As the number of hypotheses grows, the data complexity N is increased in both cases [5]. For $d = 2^m$ it is proportional to $m / C_{\min}(p)$, where $C_{\min}(p) = \min_{z_1 \neq z_2} C(p^{z_1}, p^{z_2})$.

In [3] efficiency of key ranking was also discussed and the measure *gain* to quantify success in key ranking as a function of data complexity was introduced. Later, in [6] it was proposed to use the measure *advantage*. While Biryukov, et al., need the assumption about statistical independence of the linear approximations in all their theoretical derivations, Hermelin, et al., can do without it, but instead, must make another unrealistic assumption that the ranking statistics for each key candidate are statistically independent. This assumption can be fulfilled if, for each key candidate value, new fresh data is generated to compute the ranking statistic, which will result in overestimating the data complexity. Hence, it is not known exactly in the general case, what the success probabilities of key ranking are for Algorithm 1. Nevertheless, the above analysis applies to key ranking also, and we can conclude that the LLR method and the convolution method have practically the same advantage.

6.2 Time and Memory Complexities

In [8] the Alg. 2 was divided to two phases: the on-line phase and the off-line phase. We follow the division in this paper. The on-line phase is independent of the statistics used in the attack and its sole purpose is to obtain the empirical p.d. q from the N plaintext-ciphertext pairs. The time complexity is Nm and memory complexity is 2^m. We now assume that given data N, we have obtained the empirical p.d. q.

Figures 1, 2 and 3 depict the off-line phase for the full Biryukov method, LLR-method and convolution method, respectively.

```
Input: empirical correlation vector $\boldsymbol{\rho} = (\rho(0), \ldots, \rho(2^m - 1))$ and theoretical
        correlations $c(0), \ldots, c(2^m - 1)$, of the linear approximation (1) ;
Output: the best key candidate;
for $z = 0, \ldots, 2^m - 1$ do
    compute $B(z) = \sum_{a \in \mathbb{Z}_2^m} ((-1)^{a \cdot z} c(a) - \rho(a))^2$;
end
find $z'$ that maximises $B(z)$;
output $z'$;
```

Fig. 1. Off-line phase of Alg. 1 using full Biryukov method

```
Input: empirical p.d. $q$ and theoretical p.d. $p$ of the linear approximation (1) ;
Output: the best key candidate;
for $z = 0, \ldots, 2^m - 1$ do
    compute $p^z$, a permutation of $p$;
    compute $L(z) = \mathrm{LLR}(q, p^z, \theta)$;
end
find $z'$ that maximises $L(z)$;
output $z'$;
```

Fig. 2. Off-line phase of Alg. 1 using LLR-method

```
Input: empirical p.d. $q$ and theoretical p.d. $p$ of the linear approximation (1) ;
Output: the best key candidate;
compute $p * q$ using FFT;
find mode $z'$ of $p * q$;
output $z'$;
```

Fig. 3. Off-line phase of Alg. 1 using convolution method

For each $z \in \mathbb{Z}_2^m$, both full Biryukov method and LLR-method take time 2^m to evaluate. Hence, the time complexity of both the full Biryukov and the LLR-method is 2^{2m}.

In the convolution method the computation of the convolution $p*q$ is done only once. Using FFT, that is, left hand side of (4), it takes time $m2^m$. Hence, the convolution method is much faster than the LLR or the full Biryukov, while all three methods have the same data complexities.

If all the correlations $c(a)$, $a \in \mathbb{Z}_2^m$, are non-negligible, then all three methods have the same memory complexity 2^m. In practice the full linear span of the linear approximations contains many approximations with zero or negligible correlations. Such approximations do not contribute to the capacity and hence are discarded. This has certain effect to the complexities of the algorithms.

Let l be the number of linear approximations used, $m \leq l \leq 2^m$. Then the memory requirement of the off-line phase of the Biryukov method will be reduced from 2^m to l and the time complexity becomes $2^m l$. Since the convolution is computed using the correlations by (4), the same reduction of memory is possible also for the convolution method if we use the correlations instead of the distribution in evaluating the statistic.

We run some experiments on the four-round Serpent, see [16] for an accurate description for the cipher. The test settings were the same as in [6]. To compare the LLR and convolution method in practice, we measured the advantage by Selçuk [17]. In Figure 4 we have plotted the empirical advantage as a function of the data complexity, for $m = 7$. The curves are indistinguishable.

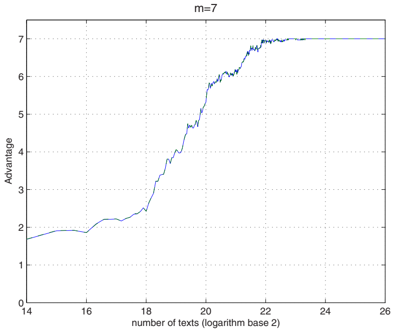
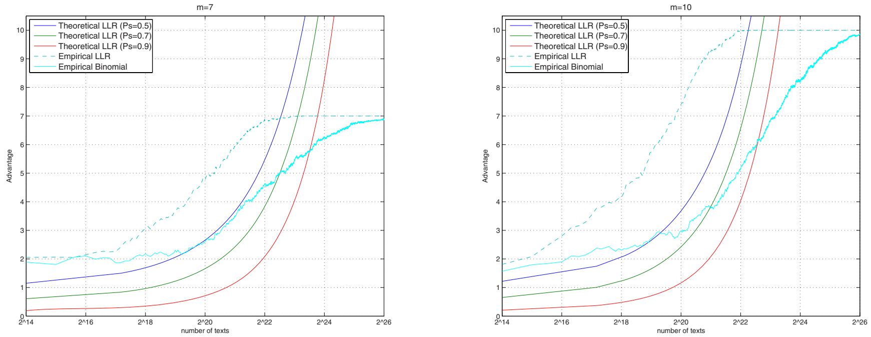


Fig. 4. The empirical advantage as a function of data complexity using LLR and convolution method with $m = 7$ for 4-round Serpent. The curves are equal.



(a) $m = 7$ (b) $m = 10$

Fig. 5. The theoretical and empirical advantage as a function of data complexity using LLR-method for the 4-round Serpent

Figure 5 shows the empirical and theoretical advantage of the LLR-method for $m = 7$ and $m = 10$. The convolution method gives exactly the same results. The theoretical prediction is slightly more pessimistic than the empirical results. However, they are still consistent.

7 Conclusions

We proposed a new method, which we call the convolution method, to perform multidimensional linear attacks. The convolution method is expected to give the same result with the same data complexity as the Biryukov method in case the set of linear approximations is completed to contain all approximations with significant correlations within the linear span of the set.

In the convolution method we form the convolution between the empirical and the theoretical p.d related to the multidimensional linear approximation of a block cipher. The right key class is determined as the mode of the resulting p.d. The data complexities of both LLR and the convolution method are of the same magnitude. Moreover, the LLR-method and full Biryukov method require time 2^{2m}, where m is the dimension of the approximation, whereas the convolution method only needs time $m2^m$. Hence, the convolution method is the most efficient in practice. Also, there is no need to assume statistical independence.

In [3] the measure gain and in [6] the measure advantage was used in studying the success of key ranking. The gain requires the assumption of statistical independence of base approximations whereas the advantage requires that the ranking statistics corresponding to different keys should be statistically independent. The latter condition can be satisfied for Alg. 2 [8] but seems to result in an unrealistic and unnecessary increase of the data complexity for Alg. 1. However, the efficiency of the convolution or LLR-methods is not affected by the assumption that is needed in calculating the advantage. If needed, the advantage can be determined approximately also for the Alg.1. The calculations in that case are the same as in [6] and the convolution method remains the most efficient method in practice.

Acknowledgements

We thank Joo Yeon Cho for producing the experiments.

References

1. Matsui, M.: Linear Cryptanalysis Method for DES Cipher. In: Helleseth, T. (ed.) EUROCRYPT 1993. LNCS, vol. 765, pp. 386–397. Springer, Heidelberg (1994)
2. Burton, S., Kaliski, J., Robshaw, M.J.B.: Linear Cryptanalysis Using Multiple Approximations. In: Desmedt, Y.G. (ed.) CRYPTO 1994. LNCS, vol. 839, pp. 26–39. Springer, Heidelberg (1994)
3. Biryukov, A., Cannière, C.D., Quisquater, M.: On Multiple Linear Approximations. In: Franklin, M. (ed.) CRYPTO 2004. LNCS, vol. 3152, pp. 1–22. Springer, Heidelberg (2004)
4. Murphy, S.: The Independence of Linear Approximations in Symmetric Cryptology. IEEE Transactions on Information Theory 52(12), 5510–5518 (2006)
5. Hermelin, M., Nyberg, K., Cho, J.Y.: Multidimensional Linear Cryptanalysis of Reduced Round Serpent. In: Mu, Y., Susilo, W., Seberry, J. (eds.) ACISP 2008. LNCS, vol. 5107, pp. 203–215. Springer, Heidelberg (2008)
6. Hermelin, M., Cho, J.Y., Nyberg, K.: Statistical Tests for Key Recovery Using Multidimensional Extension of Matsui's Algorithm 1. In: Joux, A. (ed.) EUROCRYPT 2009 - POSTER SESSION. LNCS, vol. 5479. Springer, Heidelberg (2009)
7. Vaudenay, S.: An experiment on DES statistical cryptanalysis. In: CCS 1996: Proceedings of the 3rd ACM conference on Computer and communications security, pp. 139–147. ACM, New York (1996)
8. Hermelin, M., Cho, J.Y., Nyberg, K.: Multidimensional Extension of Matsui's Algorithm 2. In: Dunkelman, O. (ed.) Fast Software Encryption. LNCS, vol. 5665, pp. 209–227. Springer, Heidelberg (2009)
9. Gérard, B., Tillich, J.: On linear cryptanalysis with many linear approximations (2009)
10. Hermelin, M., Nyberg, K.: Multidimensional Linear Distinguishing Attacks and Boolean Functions. In: Fourth International Workshop on Boolean Functions: Cryptography and Applications (2008)
11. Cover, T.M., Thomas, J.A.: 11. Wiley Series in Telecommunications and Signal Processing. In: Elements of Information Theory, 2nd edn. Wiley Interscience, Hoboken (2006)
12. McDonough, R.N., Whalen, A.D.: 5. In: Detection of Signals in Noise, 2nd edn. Academic Press, London (1995)
13. Collard, B., Standaert, F.X., Quisquater, J.J.: Experiments on the Multiple Linear Cryptanalysis of Reduced Round Serpent. In: Nyberg, K. (ed.) FSE 2008. LNCS, vol. 5086, pp. 382–397. Springer, Heidelberg (2008)
14. Baignères, T., Vaudenay, S.: The Complexity of Distinguishing Distributions (Invited Talk). In: Safavi-Naini, R. (ed.) ICITS 2008. LNCS, vol. 5155, pp. 210–222. Springer, Heidelberg (2008)
15. Baignères, T., Junod, P., Vaudenay, S.: How Far Can We Go Beyond Linear Cryptanalysis? In: Lee, P.J. (ed.) ASIACRYPT 2004. LNCS, vol. 3329, pp. 432–450. Springer, Heidelberg (2004)

16. Biham, E., Anderson, R., Knudsen, L.: Serpent: A New Block Cipher Proposal. In: Vaudenay, S. (ed.) FSE 1998. LNCS, vol. 1372, pp. 222–238. Springer, Heidelberg (1998)
17. Selçuk, A.A.: On probability of success in linear and differential cryptanalysis. Journal of Cryptology 21(1), 131–147 (2008)
18. Xiao, G.Z., Massey, J.L.: A Spectral Characterization of Correlation-Immune Combining Functions. IEEE Transactions on Information Theory 34(3), 569–571 (1988)
19. Rohatgi, V.K.: 6.7. Wiley Series in Probability and Mathematical Statistics. In: Statistical Inference, 1st edn. John Wiley & Sons, New York (1984)

Appendix

A Proof of Theorem 2

The Piling Up lemma [1] that has traditionally been used in calculating correlations of linear combinations of statistically independent linear approximations has a converse. This converse of the Piling Up lemma offers a natural criterion for verifying statistical independence of linear approximations. Given a set of linear approximations it is not sufficient to verify that all linear approximations in the set are pairwise statistically independent. We must also verify that the correlations (or imbalances [3] [4])

$$c(a) = c(a \cdot (Ux \oplus Wy \oplus VK)),\ a \in \mathbb{Z}_2^m.$$

of all linear combinations of the linear approximations must be of certain small magnitude as given by the following theorem.

Theorem 2. *Let $m \geq 2$ be an integer. The binary random variables $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_m$, with correlations $c_i = c(\mathbf{X}_i)$, $i = 1, \ldots, m$ are statistically independent, if and only if for all index sets $I \subset \{1, 2, \ldots, m\}$,*

$$c(\bigoplus_{i \in I} \mathbf{X}_i) = \prod_{i \in I} c_i. \tag{16}$$

The *only if* part follows from the Piling Up lemma. The proof of the *if* part, that is, the converse of the Piling Up lemma, is given below, using the Xiao-Massey lemma [18]:

Lemma 2 (Xiao-Massey lemma). *The discrete random variable $\mathbf{Z}$ is independent of the m independent binary random variables $\mathbf{X}_1, \ldots, \mathbf{X}_m$ if and only if $\mathbf{Z}$ is independent of the sum $b_1\mathbf{X}_1 \oplus \cdots \oplus b_m\mathbf{X}_m$, for every choice of $b_1, \ldots, b_m \in \{0, 1\}$, and not all coefficient b_i is zero.*

Proof (Converse of the Piling Up lemma). We assume that the random variables $\mathbf{X}_1, \ldots, \mathbf{X}_m$ satisfy condition (16). We do the proof with induction on m. Let $m = 2$. We assume $c(\mathbf{X}_1 \oplus \mathbf{X}_2) = c_1 c_2$ and we have to prove that for all pairs $t = (t_1, t_2) \in \{0, 1\} \times \{0, 1\}$, the probability $\Pr(\mathbf{X}_1 = t_1, \mathbf{X}_2 = t_2) = \Pr(\mathbf{X}_1 = t_1) \Pr(\mathbf{X}_2 = t_2)$.

Denote $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$. Using the definition of the correlation we have

$$\begin{aligned}
&\Pr(\mathbf{X}_1 = t_1)\Pr(\mathbf{X}_2 = t_2)\\
&= (1/2 + (-1)^{t_1} c_1)(1/2 + (-1)^{t_2} c_2)\\
&= 1/4 + (-1)^{(0,1)\cdot t} c_1 + (-1)^{(1,0)\cdot t} c_2 + (-1)^{(1,1)\cdot t} c_1 c_2\\
&= c((0,0)\cdot \mathbf{X}) + (-1)^{(0,1)\cdot t} c((1,0)\cdot \mathbf{X}) + (-1)^{(1,0)\cdot t} c((1,0)\cdot \mathbf{X}) + (-1)^{(1,1)\cdot t} c((1,1)\cdot \mathbf{X})\\
&= \sum_{a \in \mathbb{Z}_2^2} (-1)^{a\cdot t} c(a \cdot \mathbf{X}).
\end{aligned}$$

But by Lemma 2.1 in [10], the last sum is equal to $\Pr(\mathbf{X} = t) = \Pr(\mathbf{X}_1 = t_1, \mathbf{X}_2 = t_2)$.

Assume now that the claim holds for $2, \ldots, m-1$ binary random variables and let $\mathbf{X}_1, \ldots, \mathbf{X}_m$, satisfy condition (16). By the induction assumption random variables $\mathbf{X}_2, \ldots, \mathbf{X}_m$, are s.i. Hence, it suffices to show that $\mathbf{X}_1$ is s.i. of the $m-1$ random variables $\mathbf{X}_2, \ldots, \mathbf{X}_m$.

Choose any binary coefficients $b_2, \ldots, b_m \in \{0, 1\}$, not all zero, and let $I = \{i = 2, \ldots, m : b_i = 1\}$ be the index set of non-zero coefficients b_i. Denote $\mathbf{Z}_I = b_2\mathbf{X}_2 \oplus \cdots \oplus b_m\mathbf{X}_m$. By the Xiao-Massey lemma, we must show that the random variable $\mathbf{X}_1$ is s.i. of $\mathbf{Z}_I$ for all index sets $I \subset \{2, 3, \ldots, m\}$. By the induction assumption and Xiao-Massey lemma, the claim holds already for all $I \neq \{2, 3, \ldots, m\}$ and we only have to consider the set $J = \{2, 3, \ldots, m\}$. By the condition (16), the correlation $c(\mathbf{Z}_J) = \prod_{i=2}^{m} c_i$ and $c(\mathbf{X}_1 \oplus \cdots \oplus \mathbf{X}_m) = \prod_{i=1}^{m} c_i$. Hence, the random variables $\mathbf{X}_1$ and $\mathbf{Z}_J$ satisfy

$$c(\mathbf{X}_1 \oplus \mathbf{Z}_J) = \prod_{i=1}^{m} c_i = c_1 c(\mathbf{Z}_J).$$

But since the theorem holds for $m = 2$, the random variables $\mathbf{X}_1$ and $\mathbf{Z}_J$ must be s.i.

□

B Proof of Lemma 1

Proof. The expected values, variances and covariances of elements of $\mathbf{Q}$ are [19]

$$E(\mathbf{Q}_\eta) = s_\eta \quad \mathrm{Var}(\mathbf{Q}_\eta) = s_\eta(1 - s_\eta) \quad \mathrm{Cov}(\mathbf{Q}_\eta, \mathbf{Q}_\nu) = -s_\eta s_\nu, \tag{17}$$

for all $\eta, \nu = 0, 1, \ldots, M$ and $\nu \neq \eta$. The normality follows from the law of large numbers. The expected value follows from linearity and (17). The variance is obtained by

$$\begin{aligned}
\sigma^2 &= \sum_{\eta=0}^{M} \mathrm{Var}(\lambda_\eta \mathbf{Q}_\eta) + \sum_{\eta,\nu=0,\nu\neq\eta}^{M} \mathrm{Cov}(\lambda_\eta \mathbf{Q}_\eta, \lambda_\nu \mathbf{Q}_\nu)\\
&= \sum_{\eta=0} \lambda_\eta^2 s_\eta (1 - s_\eta) - \sum_{\eta,\nu=0,\nu\neq\eta} \lambda_\eta \lambda_\nu s_\eta s_\nu = \sum_{\eta=0}^{M} \lambda_\eta^2 s_\eta - \mu^2.
\end{aligned}$$

□